VACCINE

Visual Analytics for Command, Control and Interoperability Environments
A U.S. Department of Homeland Security
Science and Technology Center of Excellence

VACCINE ANNUAL REPORT – YEAR 6
Addendum B - Presentations
JULY 1, 2014 – JUNE 30, 2015

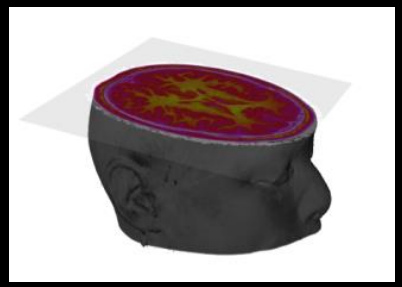
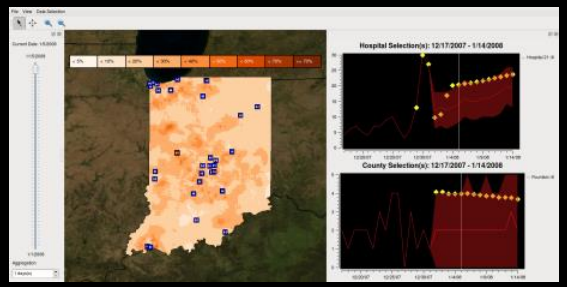
Cooperative Agreement No. 2009-ST-061-CI0001

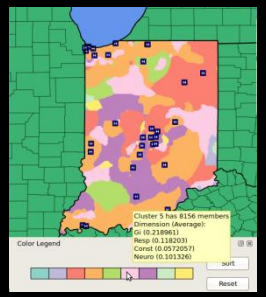
PURDUE UNIVERSITY

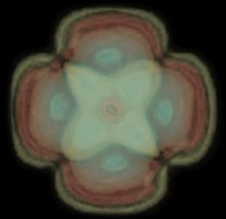
Georgia Tech
SFU SIMON FRASER UNIVERSITY THINKING OF THE WORLD
Florida International University
Stanford University
Jackson State University
The Pennsylvania State University
UNIVERSITY of WASHINGTON
UNCCHARLOTTE
MORGAN STATE UNIVERSITY
OAK RIDGE National Laboratory
Grambling State University
ASU ARIZONA STATE UNIVERSITY
Carleton UNIVERSITY
VICTORIA UNIVERSITY
UNIVERSITY OF CALGARY
DALHOUSIE UNIVERSITY Inspiring Minds
UNIVERSITY OF MANITOBA
Universität Stuttgart
The University of North Carolina at Charlotte
UNIVERSITY OF ONTARIO INSTITUTE OF TECHNOLOGY
Prairie View A&M University
UNIVERSITY OF NOTRE DAME
UNIVERSITY OF OXFORD
THE UNIVERSITY OF TEXAS AT AUSTIN
Virginia Tech
UNIVERSITY OF MARYLAND

U.S. DEPARTMENT OF HOMELAND SECURITY

HOMELAND SECURITY UNIVERSITY PROGRAMS
TODAY'S RESEARCH & EDUCATION, TOMORROW'S SECURITY



Integrating Visual Analytics Curriculum Into Your Own Classroom

Ross Maciejewski



Ross Maciejewski
E-mail: rmacieje@asu.edu
http://rmaciejewski.faculty.asu.edu/
http://lab.vader.asu.edu



The Ongoing Data Deluge

- Since 2003, digital information has accounted for 90% of all information produced[1]
- In 2009, drones from Iraq and Afghanistan recorded 24 years of video footage
- In 2010, the amount of information added annually to the digital universe was estimated to be nearly 1 ZB
- Wal-mart process > 1 million transactions per hour
- By 2013 Cisco estimates the annual internet traffic will be 667 EBs

1 – C. R. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans, and T. S. Yoo, Eds., *NIH-NSF Visualization Research Challenges Report*, IEEE Press, ISBN 0-7695-2733-7, 2006

The Ongoing Data Deluge

- FedEx's ships more than 8.5 million packages per day (http://www.fedex.com/us/about/today/companies/corporation/facts.html)
- Consumers carry more than 1 billion Visa cards worldwide. More than 450 million of those cards are in the United States (http://www.creditcards.com/credit-card-news/credit-card-industry-facts-personal-debt-statistics-1276.php)

Data Overload

- Opportunity: Huge amounts of data available in digital form and ready for analysis!
  - But, how can we make sense of this data?
  - How can we harness this data in the decision making process?
  - How do we avoid being overwhelmed by all of this data?



Having Fun In the Data Deluge

- More people are injured on toilets than by skiing or snowboarding.
- More households have dogs as pets than cats, but cat lovers are more likely to have multiple pets.
- Americans are eating more peanuts and drinking less coffee.
- More Americans belong to a fantasy sports league (10.6 million) than to book clubs (5.7 million).
- Book club members are outnumbered by avid bird-watchers (5.8 million).

Facts from the Statistical Abstract of the United States, 2013

Why Is This Important to Educators?

- In 2011, a report by the McKinsey Global Institute projected that the US needs 140,000-190,000 works with "deep analytical" expertise[1]
- These are not just computer science jobs!
- Harvard Business Review calls data science "the sexiest job in the 21st century"[2]
- We need to train the next generation of scientists for data analytics!

1 - http://www.mckinsey.com/Features/Big_Data

2 - http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html?pagewanted=all

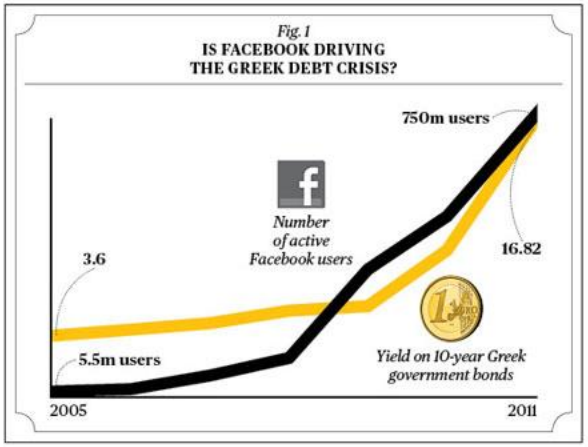
3 - http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all&_r=0

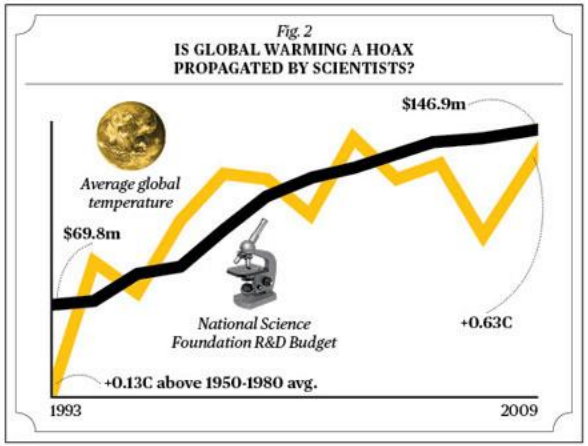
What Goes Into an Degree in “Data Science”?

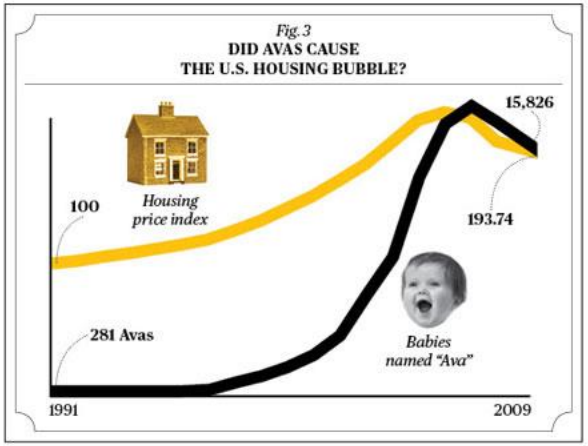
- Typically Data Science is a Masters Program
- Basic programming
- Data cleaning and scraping
- Predictive statistics (regression modeling)
- Ethics classes on data privacy implications
- What about communicating findings, what other skills do we need?

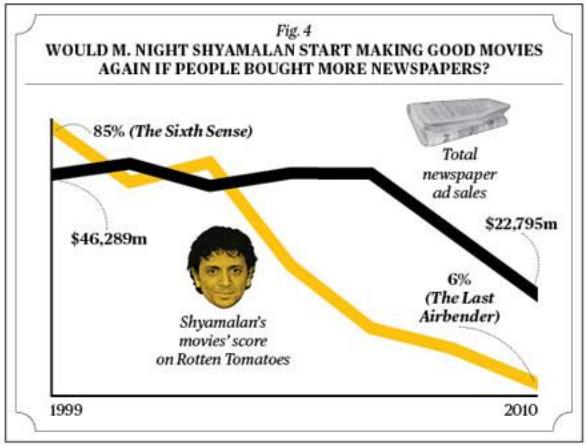
The Power of Visualization

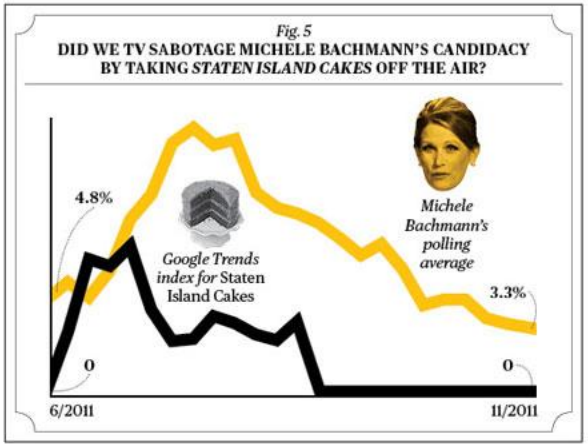
- The goal is to take all of this *data* and transform it into *information*
- How many terabytes of data we have collected doesn't matter, it's how many petaflops of *insights* we can generate from this data
- "*The purpose of computing is insight, not numbers.*" Dedication to his book: RW Hamming (1971). *Introduction to Applied Numerical Analysis*. McGraw Hill.
- We need to make the data understandable to people and a key way of doing this is through *visualization*

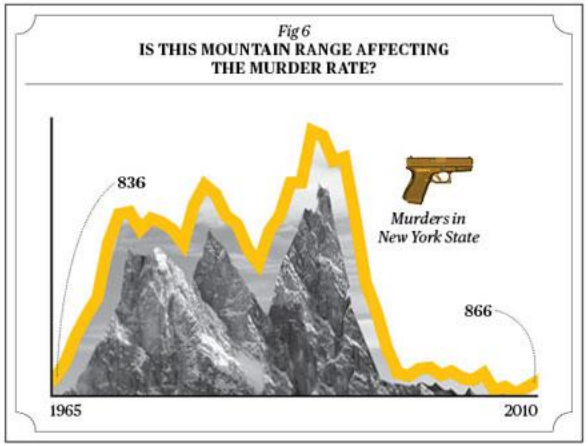












ALAMY(3); BLOOMBERG(1); GETTY IMAGES(7). DATA: FIG 1: FACEBOOK, BLOOMBERG; FIG 2: NASA, NATIONAL SCIENCE FOUNDATION; FIG 3: U.S. SOCIAL SECURITY ADMINISTRATION, FEDERAL HOUSING FINANCE AGENCY; FIG 4: ROTTEN TOMATOES, NEWSPAPER ASSOCIATION OF AMERICA; FIG 5: GOOGLE, REAL CLEAR POLITICS; FIG 6: NEW YORK LAW ENFORCEMENT AGENCY

What Skills Are Students Missing?

Non-Data Components of Graphs

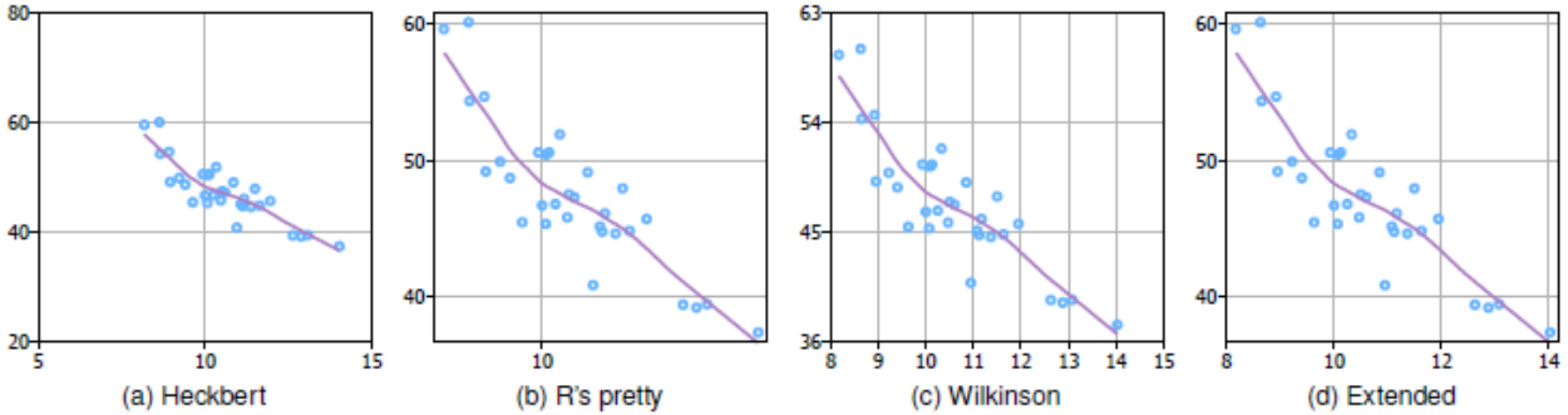
- Axes and legends can often be as important as the data themselves
- Poor axis choices and label choices can lead to confusing visualizations
- Axis tick labels provide cognitive context for most basic plot types
- They support estimation and contribute to the overall appearance of the graphic
- Cleveland suggests choosing the scales so that the data rectangle fills up as much of the scale-line rectangle as possible

W. S. Cleveland. The Elements of Graphing Data. Wadsworth Publ. Co., Belmont, CA, USA, 1985.

Extension of Wilkinson's Algorithm

- Utilizes four components for optimizing tick placement
  - Simplicity $= 1 - \frac{i-1}{|Q|-1}$-j+v
  - Density $= 2 - max\left(\frac{\rho}{\rho_t}, \frac{\rho_t}{\rho}\right)$

	Name	Description
Input	Q	preference-ordered list of nice step sizes
	(d_{min}, d_{max})	data range
	ρ_t	target label density
	fs_t	target font size
	Formats	set of label formats
		(defaults shown in Table 3)
Other	q	element of Q
	j	skip amount
	i	index of q in Q
	(l_{min}, l_{max})	labeling sequence range
	v	equals 1 if labeling includes 0, 0 otherwise



1 – J. Talbot, S. Lin, P. Hanrahan, "**An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes**," IEEE Transactions on Visualization and Computer Graphics 16(6): 1036-1043, 2010

2 - L. Wilkinson. The Grammar of Graphics (Statistics and Computing). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

Graph Aspect Ratios

- Our ability to perceive trends and patterns in a given display is heavily influenced by the aspect ratio
- Aspect ratio affects densities, relative distances and orientations
- Several methods have been proposed for automatically selecting the aspect ratio
- Aspect ratio: a = width/height

1 – J. Talbot, J. Gerth, P. Hanrahan, “Arc Length-based Aspect Ratio Selection,” IEEE Transactions on Visualization and Computer Graphics 17(12): 2276-2282, 2011

How Can We Integrate Data Analytics Into Our Classes?

National Health and Nutrition Examination Survey (NHANES)

Designed to assess health and nutritional status of adults and children in the US

- *Combines interviews and physical examinations*
- *Exams a nationally representative sample of about 5,000 persons per year*
- *Includes demographic, socioeconomic, dietary and health-related questions*

Healthy Eating Index (HEI)

Measure of diet quality that assesses conformance to federal dietary guidance

USDA uses HEI to monitor diet quality of the US population

HEI-2005 uses recommendations of the MyPyramid guidance as the baseline

1 - Guenther, P. M., Reedy, J., Krebs-Smith, S. M., Reeve, B. B., "Evaluation of the Healthy Eating Index-2005," *Journal of the American Dietetic Association*, 108:1854-1896, 2008.

2 – Freedman, L. S., Guenther, P. M., Krebs-Smith, S. M., Kott, P. S., "A Population's Mean Healthy Eating Index-2005 Scores are Best Estimated by the Score of the Population Ratio when One 24-Hour Recall is Available," Journal of Nutrition, 138:1725-1729, 2008.

Healthy Eating Index–2005 components and standards for scoring[1]

Component		Maximum points	Standard for maximum score	Standard for minimum score of zero
Total Fruit (includes 100% juice)	(HEI-1)	5	≥0.8 cup equiv. per 1,000 kcal	No Fruit
Whole Fruit (not juice)	(HEI-2)	5	≥0.4 cup equiv. per 1,000 kcal	No Whole Fruit
Total Vegetables	(HEI-3)	5	≥1.1 cup equiv. per 1,000 kcal	No Vegetables
Dark Green and Orange Vegetables and Legumes[2]	(HEI-4)	5	≥0.4 cup equiv. per 1,000 kcal	No Dark Green or Orange Vegetables or Legumes
Total Grains	(HEI-5)	5	≥3.0 oz equiv. per 1,000 kcal	No Grains
Whole Grains	(HEI-6)	5	≥1.5 oz equiv. per 1,000 kcal	No Whole Grains
Milk[3]	(HEI-7)	10	≥1.3 cup equiv. per 1,000 kcal	No Milk
Meat and Beans	(HEI-8)	10	≥2.5 oz equiv. per 1,000 kcal	No Meat or Beans
Oils[4]	(HEI-9)	10	≥12 grams per 1,000 kcal	No Oil
Saturated Fat	(HEI-10)	10	≤7% of energy[5]	≥15% of energy
Sodium	(HEI-11)	10	≤0.7 gram per 1,000 kcal[5]	≥2.0 grams per 1,000 kcal
Calories from Solid Fats, Alcoholic beverages, and Added Sugars (SoFAAS)	(HEI-12)	20	≤20% of energy	≥50% of energy

[1]Intakes between the minimum and maximum levels are scored proportionately, except for Saturated Fat and Sodium (see note 5).
[2]Legumes counted as vegetables only after Meat and Beans standard is met.
[3]Includes all milk products, such as fluid milk, yogurt, and cheese, and soy beverages.
[4]Includes nonhydrogenated vegetable oils and oils in fish, nuts, and seeds.
[5]Saturated Fat and Sodium get a score of 8 for the intake levels that reflect the 2005 Dietary Guidelines, <10% of calories from saturated fat and 1.1 grams of sodium/1,000 kcal, respectively.

http://www.cnpp.usda.gov/Publications/HEI/healthyeatingindex2005factsheet.pdf

Playing Data Detective

So, we've collected all of these surveys, now what?

Well, how about Exploratory Data Analysis where we visually inspect the data and ask questions about it to help form hypotheses.

What are some easy questions to ask?

- *What is the average … (total fruit consumption)?*
- *What is the average … (milk consumption)?*

1 – J. W. Tukey. Exploratory Data Analysis

NHANES

Utilizing Survey Cycle 2-3

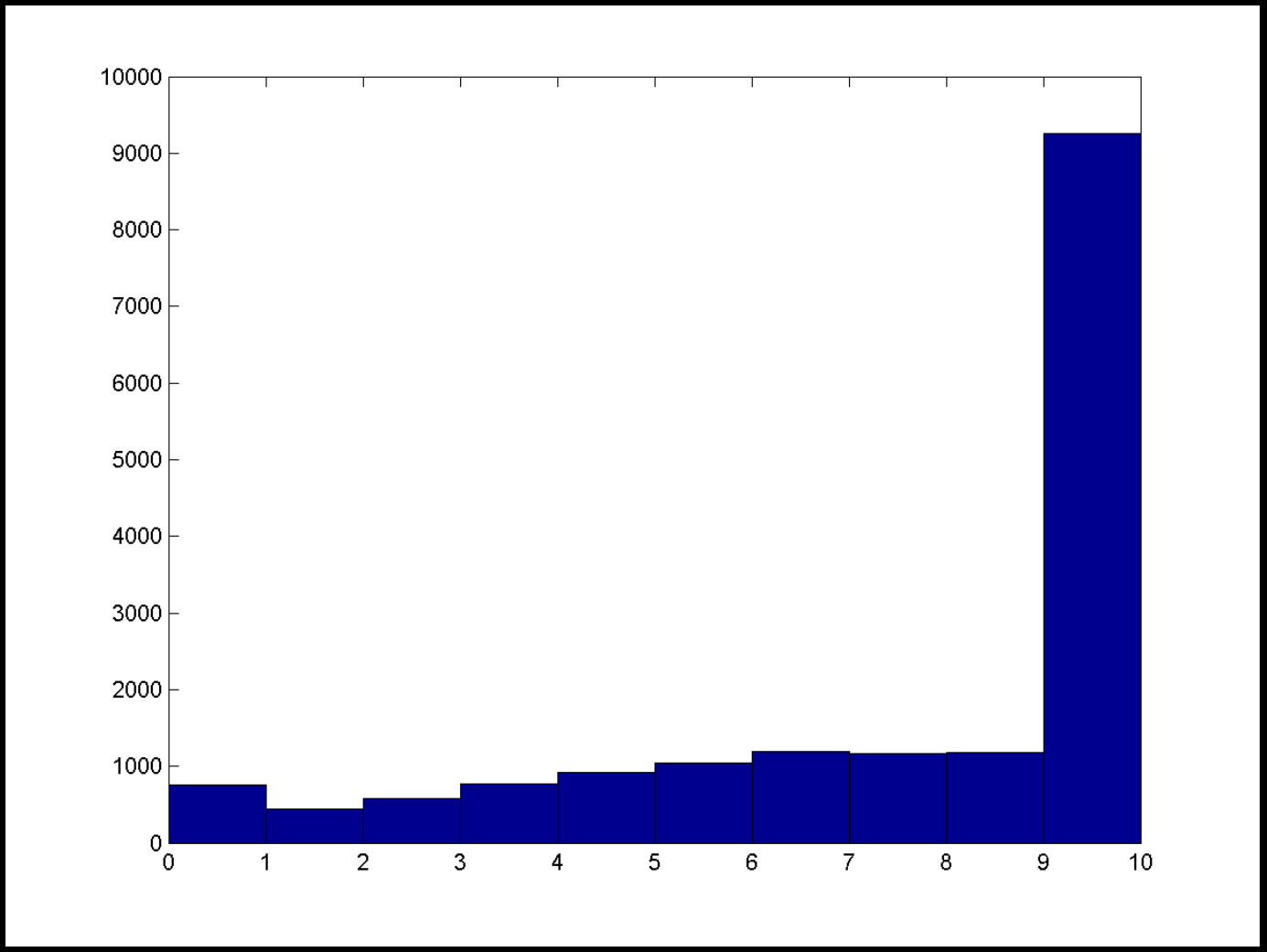
- ***Cycle 2: 2001-2002***
- ***Cycle 3: 2003-2004***
- ***Sample size of ~15000***

Other variables:

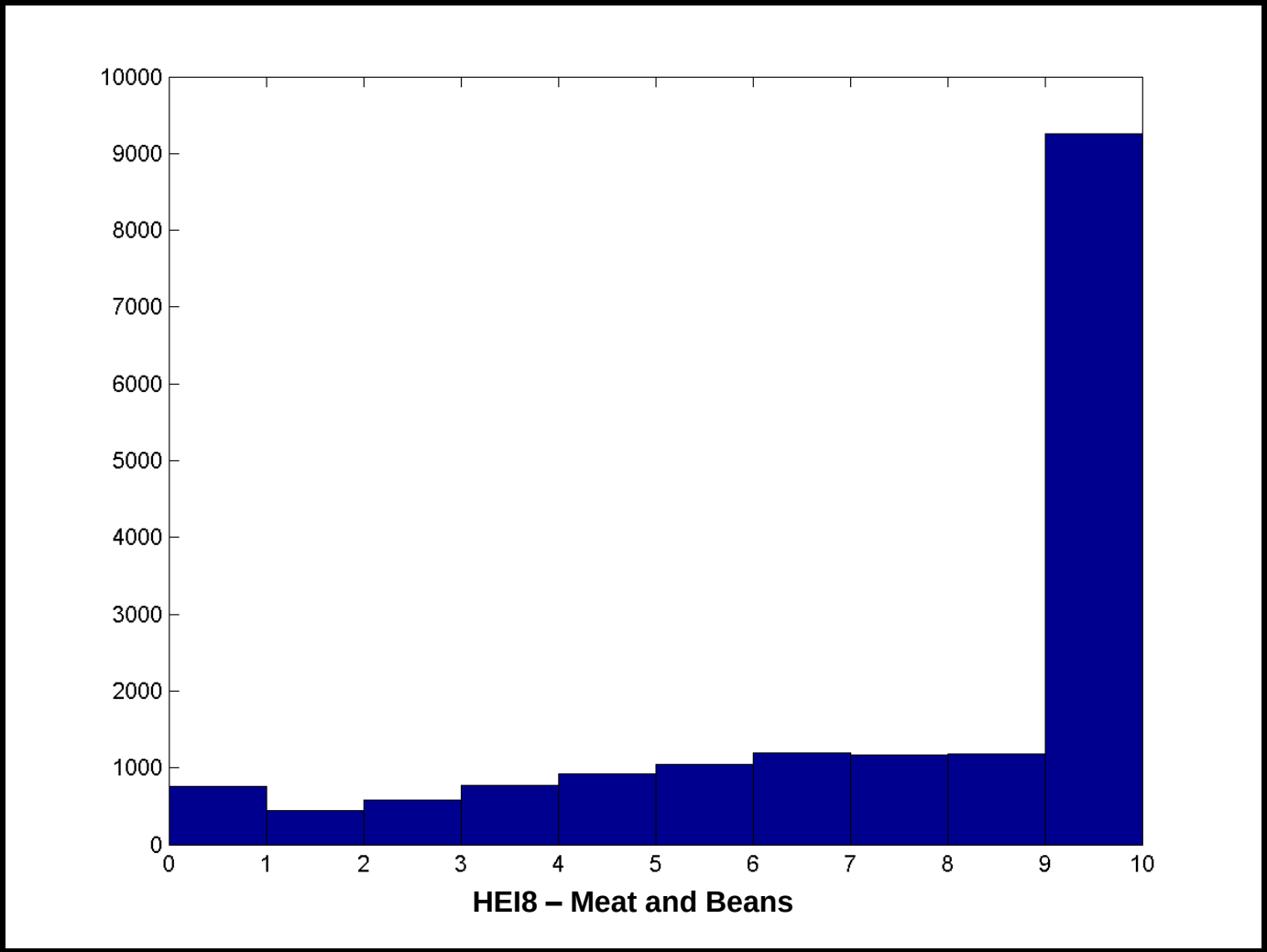
- **Gender**
- **Age**
- **Race**
- **Total energy for the day (kcal)**
- **Total saturated fatty acid (gm)**
- **BMI**

So with more data, maybe we can think of more questions!

What is the average … ?



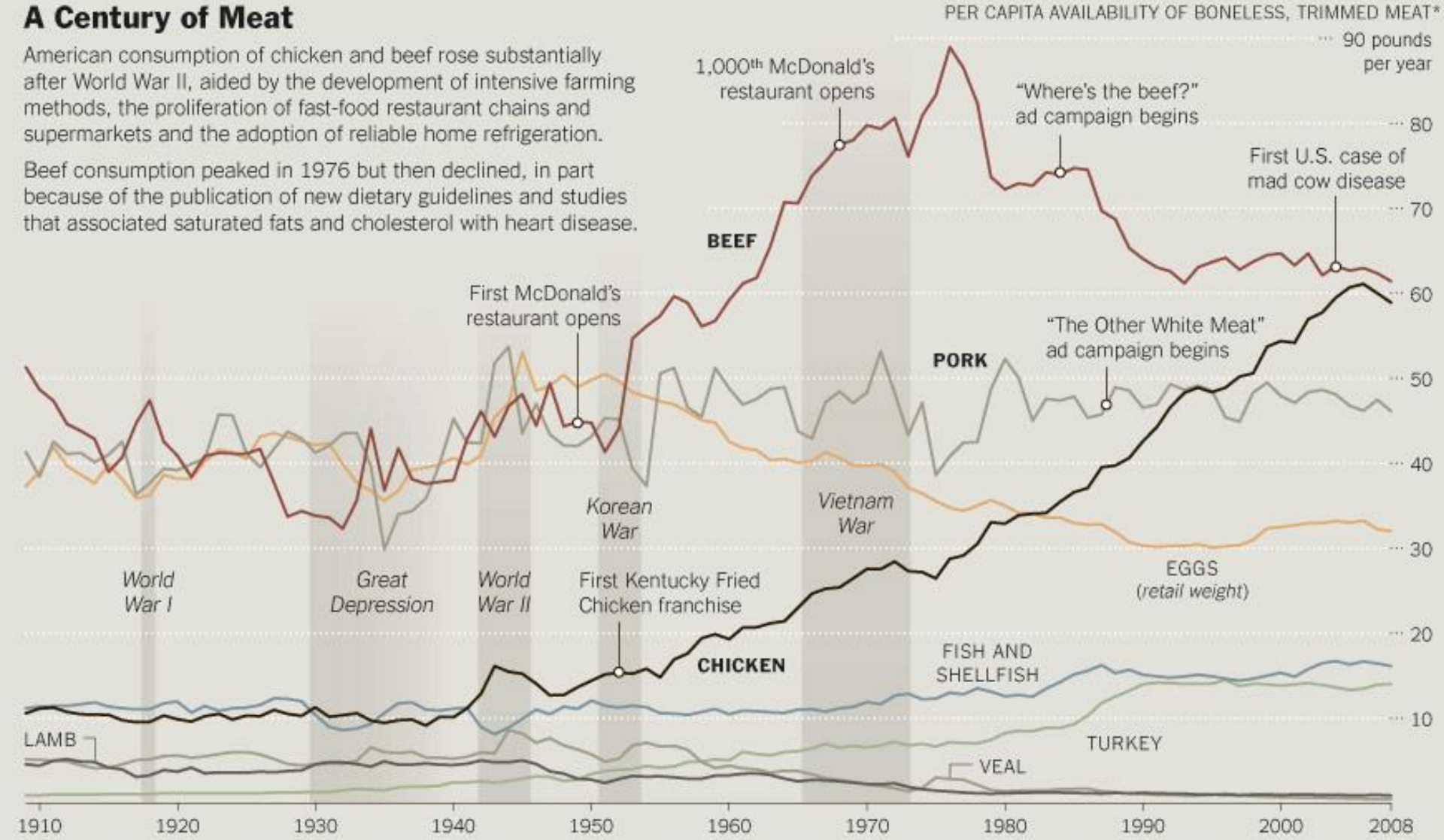
What is the average … ? Meat Consumption



A Century of Meat

American consumption of chicken and beef rose substantially after World War II, aided by the development of intensive farming methods, the proliferation of fast-food restaurant chains and supermarkets and the adoption of reliable home refrigeration.

Beef consumption peaked in 1976 but then declined, in part because of the publication of new dietary guidelines and studies that associated saturated fats and cholesterol with heart disease.



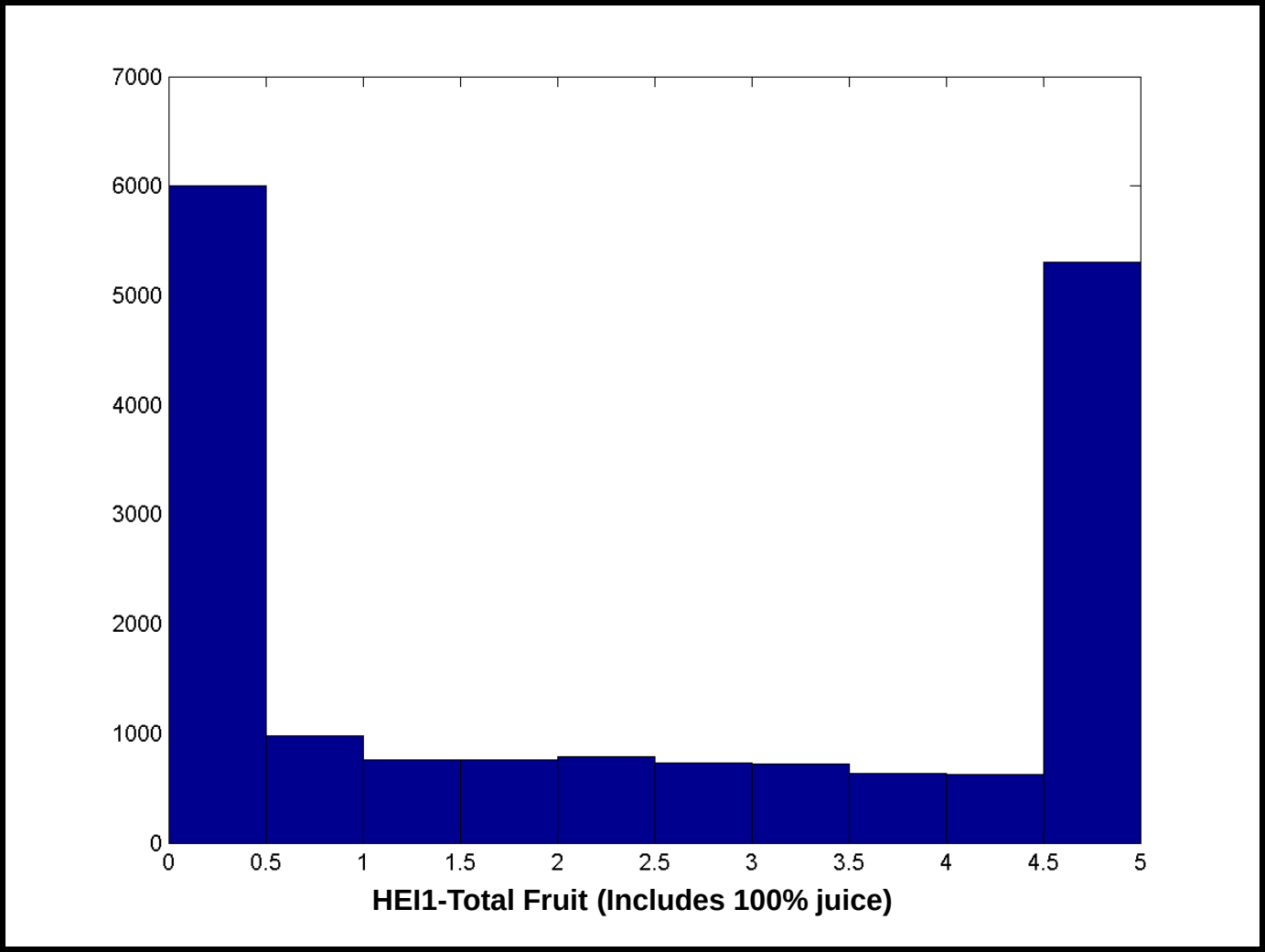
*Note: per capita availability of boneless meat is a proxy for human consumption, and is lower than retail weight or carcass weight. Bones, offal and game are excluded.

Sources: U.S. Department of Agriculture (data); news and company reports; "Putting Meat on the American Table," by Roger Horowitz

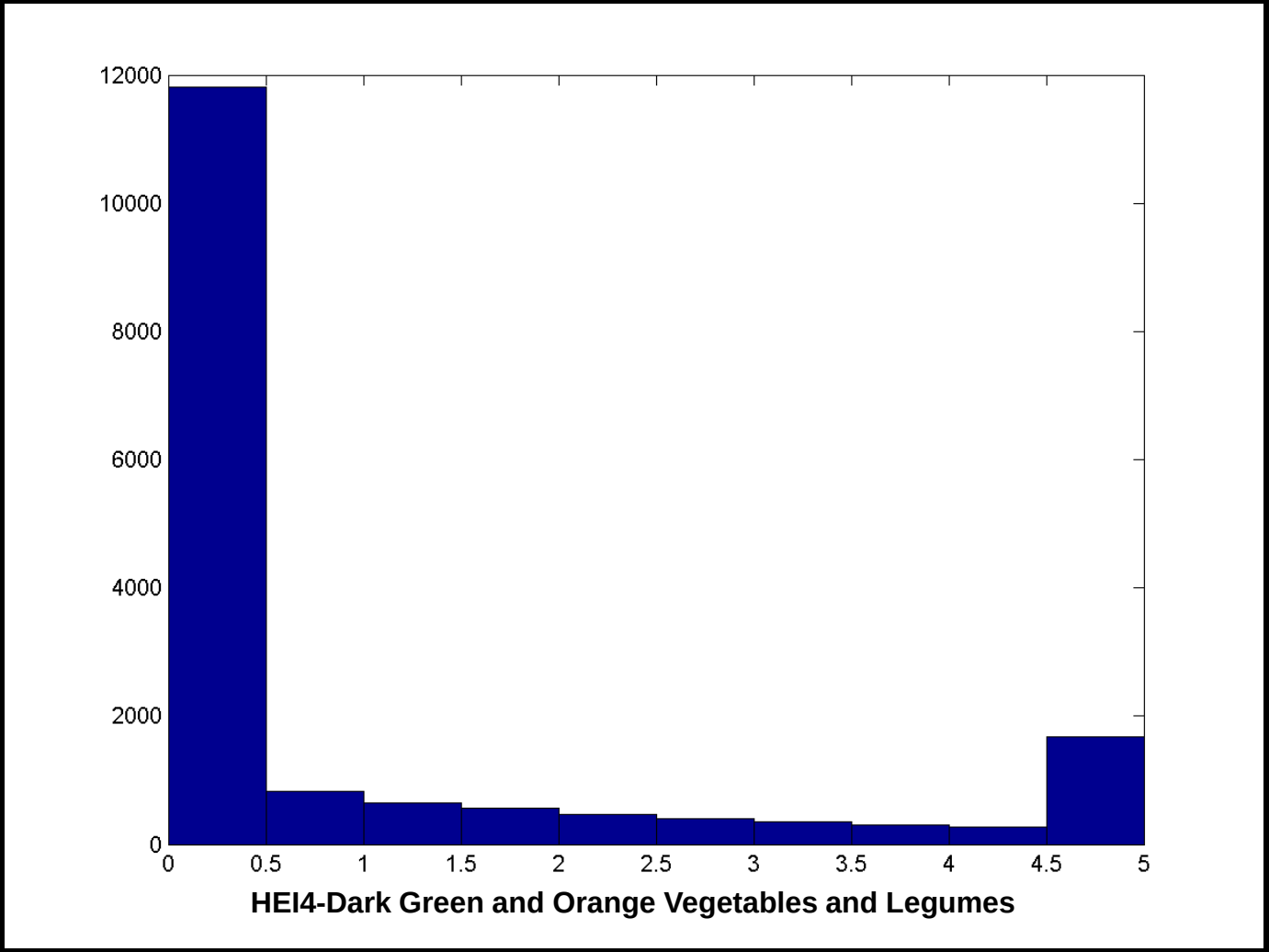
JONATHAN CORUM/THE NEW YORK TIMES

http://www.nytimes.com/imagepages/2011/03/15/science/15food_graphic.html

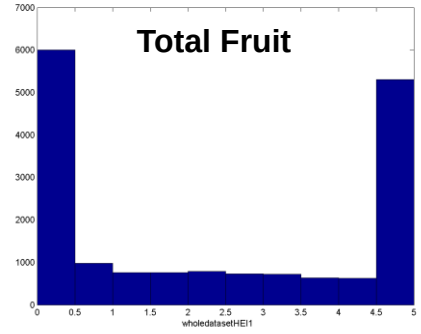
What is the average … ? Total Fruit Consumption

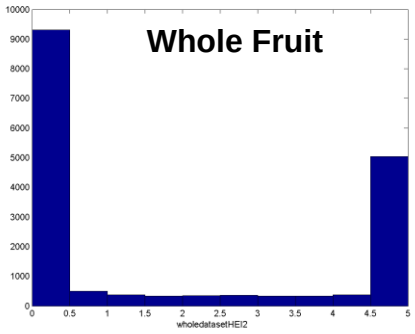


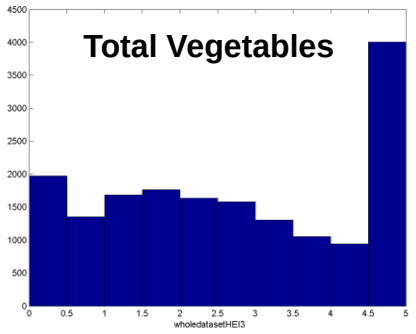
What is the average … ? Dark Green and Orange Vegetable Consumption

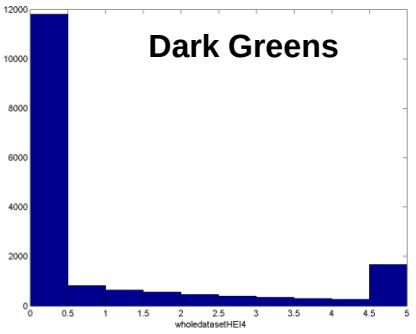


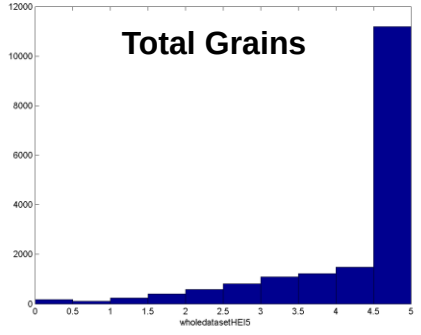
How Do These Compare?

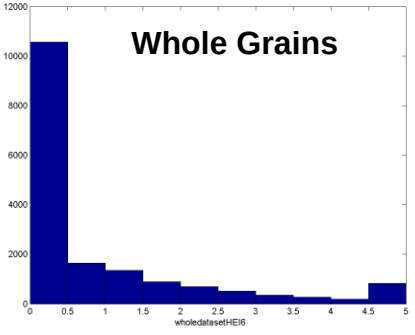


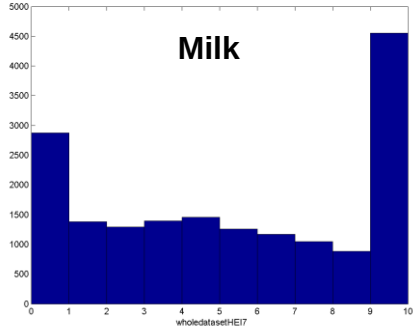


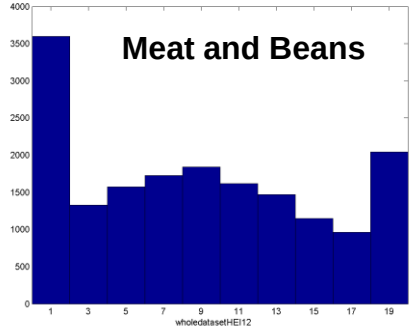


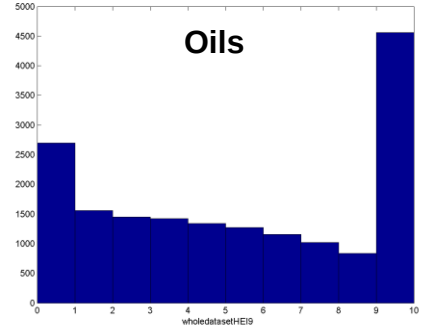


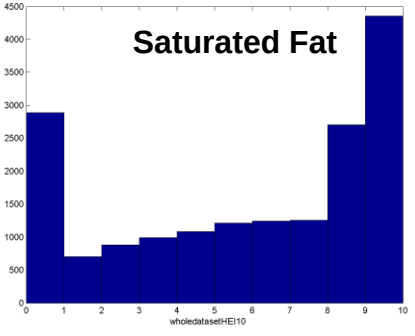


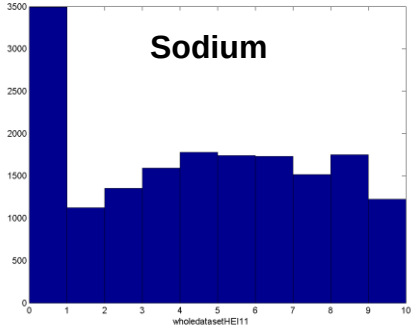


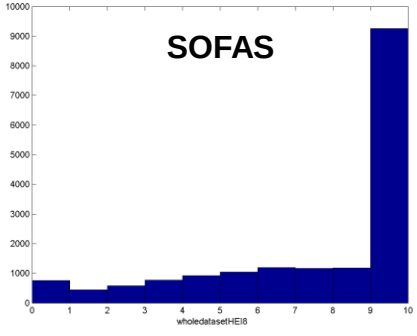












So, What Can I say About the Population as a Whole?

- *We consume lots of SOFAs*
- *We are not good at eating our vegetables*
- *We eat lots of grains …*
- *But not lots of whole grains*

- *But what does this say about someone like you?*

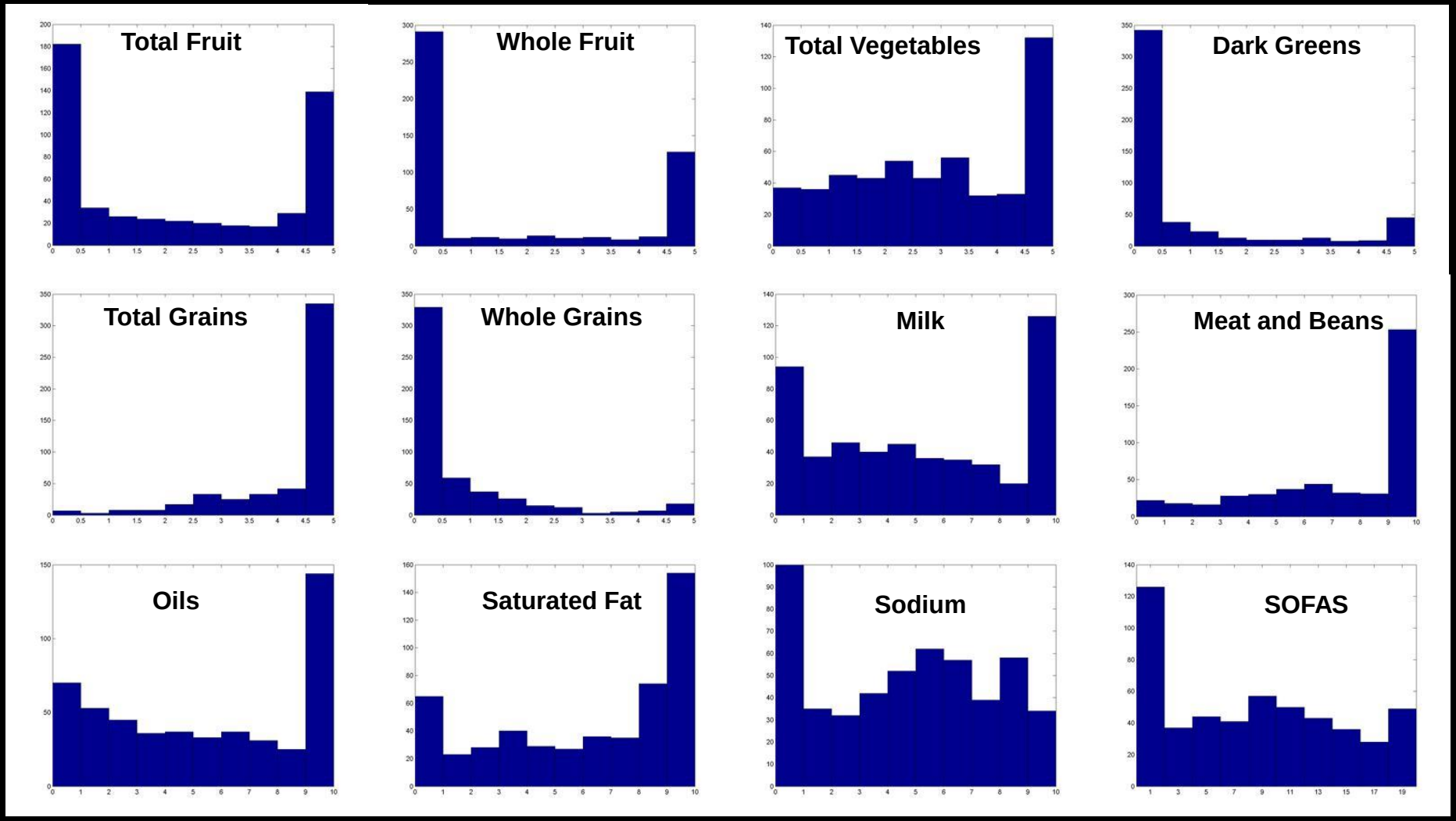
What if I further segment the data? Filter by Age, Gender, Etc.?

Let's look at Females, Age 19-30

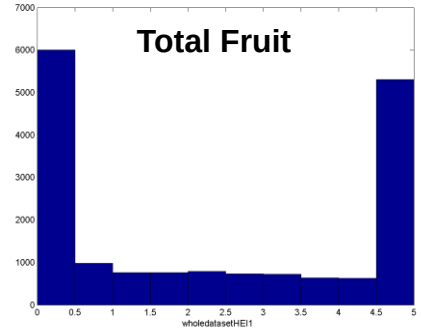
18.5 < BMI < 25.0

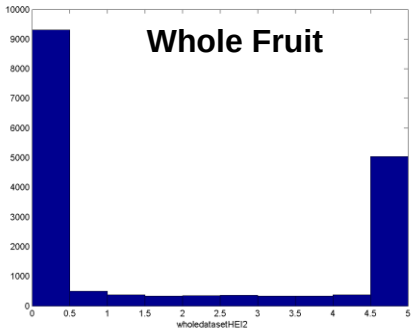
Total population in our sample = 511

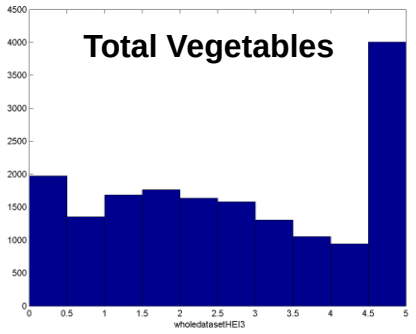
NHANES in Females Age 19-30, Average BMI

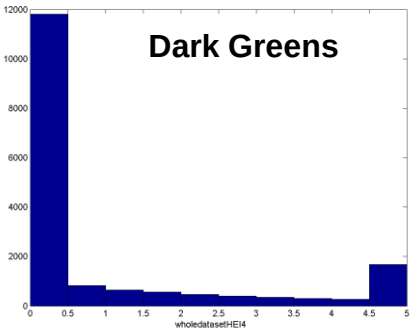


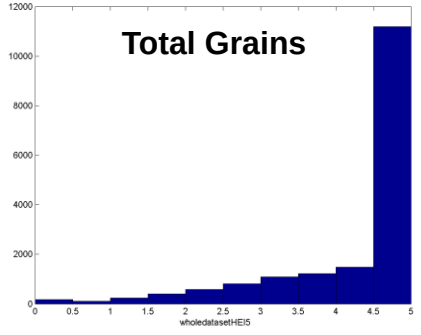
Back to the Entire Population!

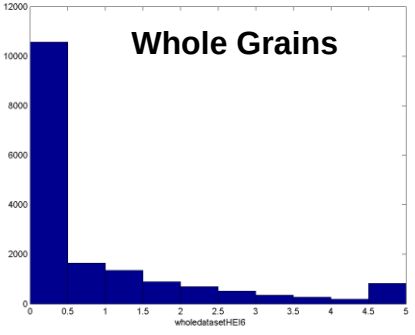


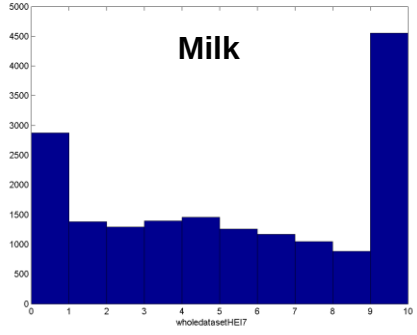


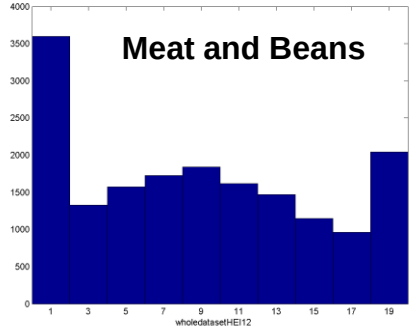


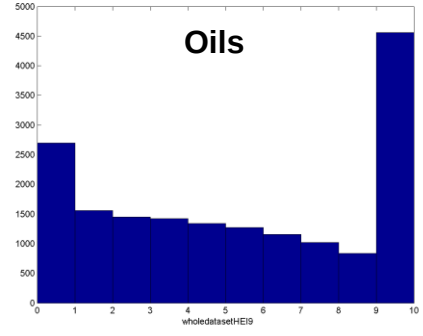


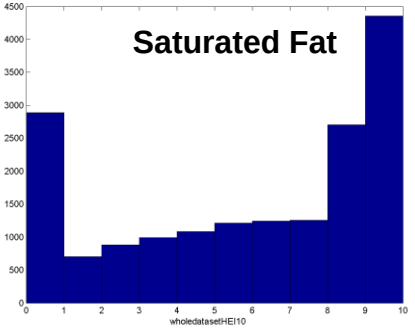


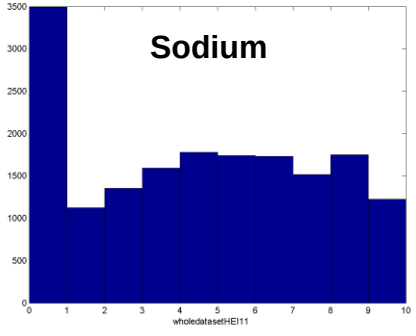


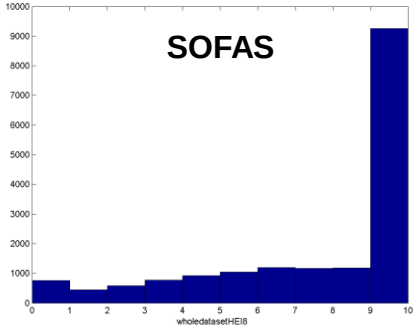












Forming Hypotheses

What do we think we see when comparing this group of females to the entire population?

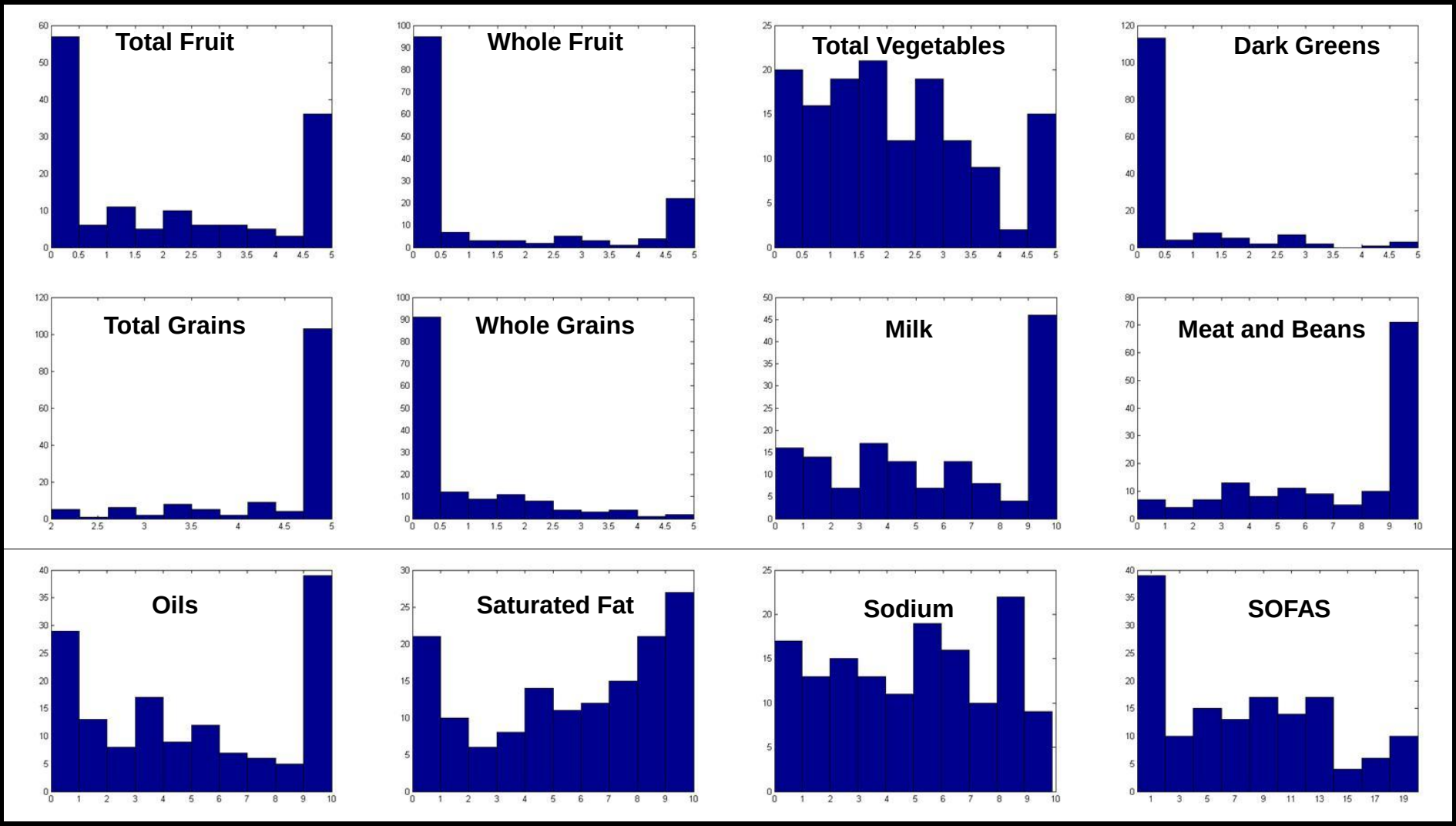
- ***Their SOFA consumption is more evenly distributed***
- ***Their meat and bean consumption is more skewed***

What if I further segment the data? Filter by Age, Gender, Etc.?

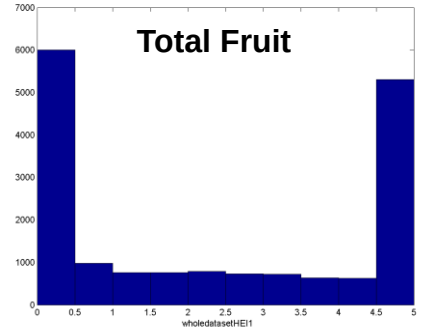
Let's look at Males, Age 14-18, BMI < 18.5

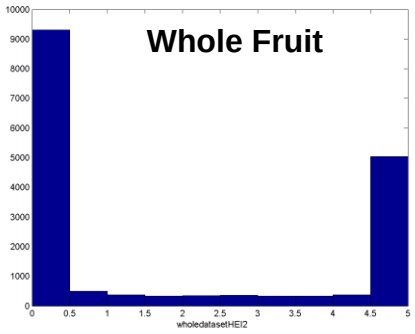
Total population in our sample = 145

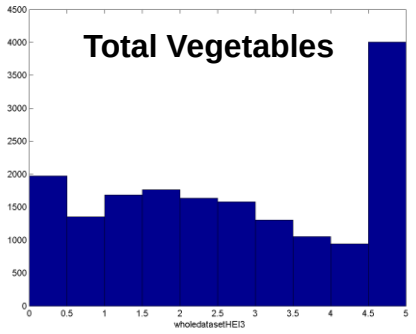
NHANES in Adolescent Underweight Males

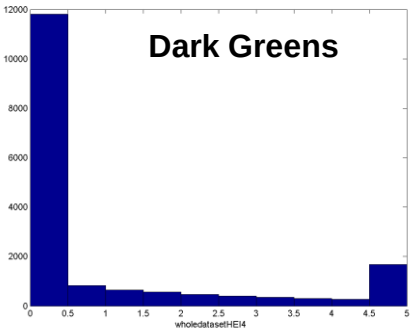


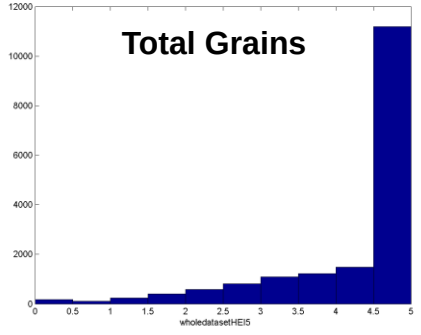
Back to the Entire Population!

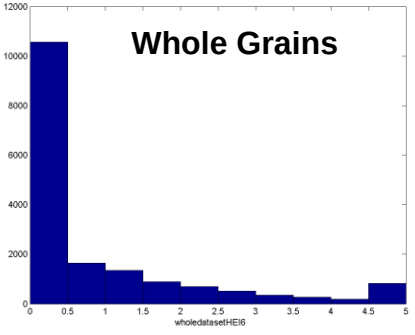


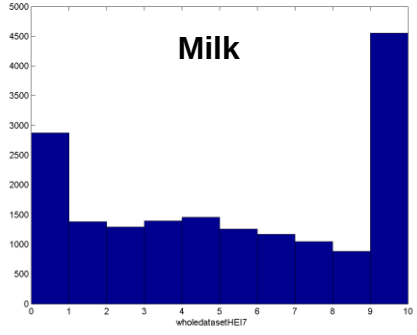


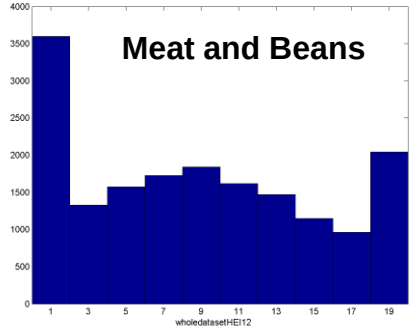


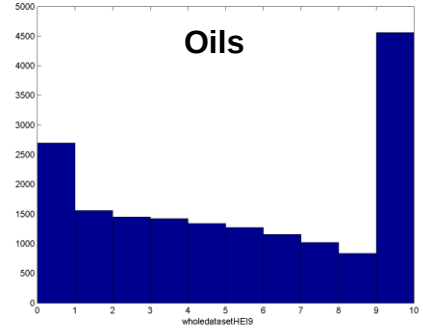


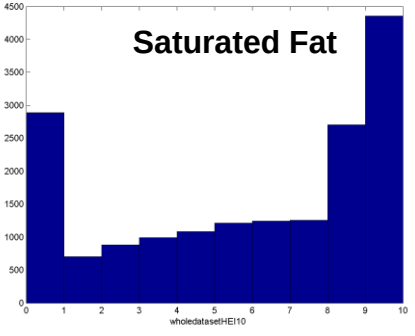


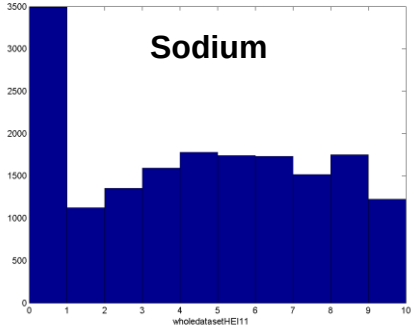


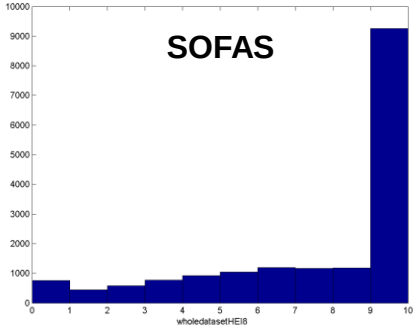












Forming Hypotheses

What do we think we see when comparing this group of females to the entire population?

- ***Their vegetable consumption is more evenly distributed***
- ***Their meat and bean consumption is more skewed***
- ***Their SOFAs are more evenly distributed***

What if I want to ask: How is … related to … ?

From our histograms, we came up with interesting insights into our data

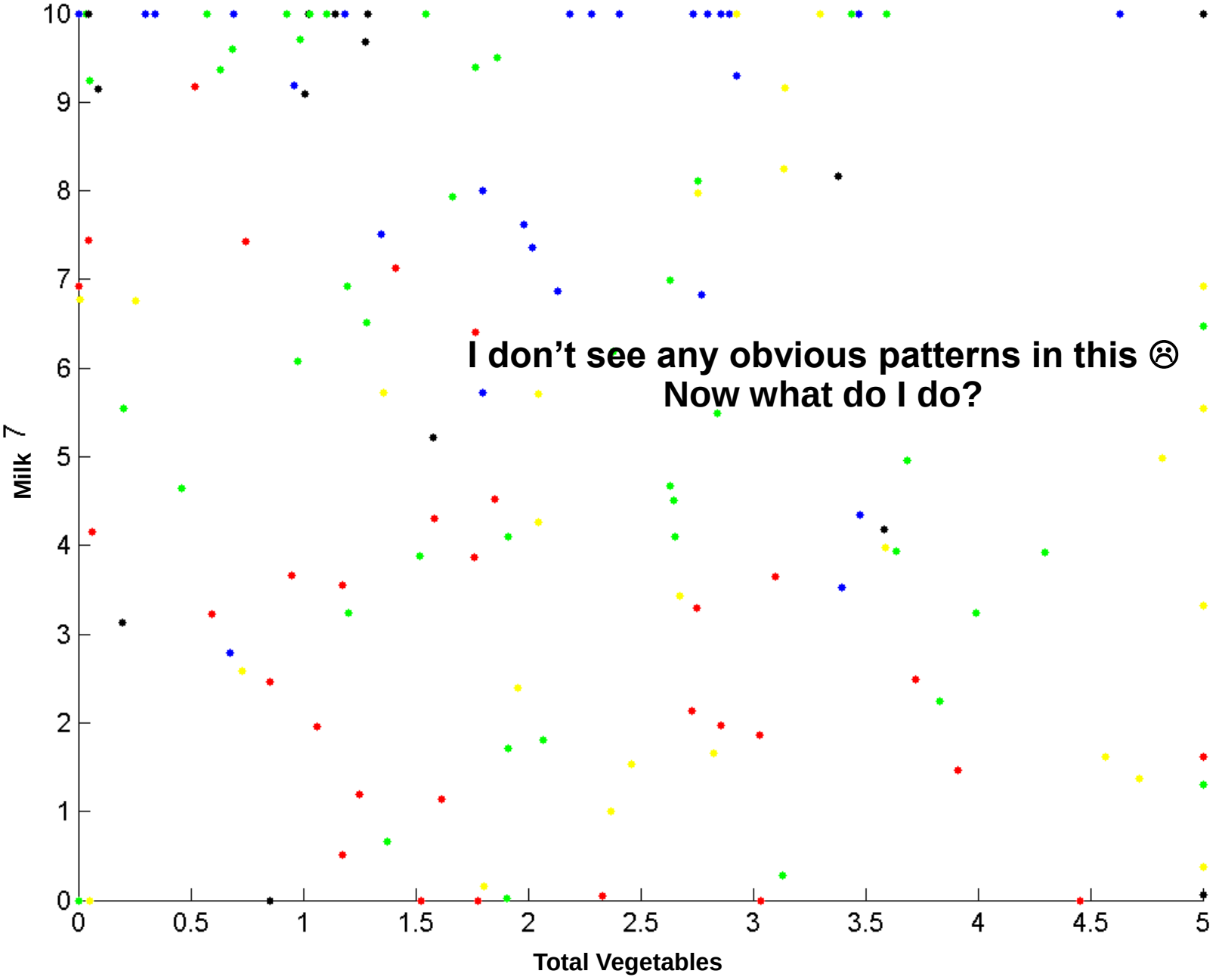
But what if I want to know about relationships between the data?



More like:
Do people that consume more milk consume more vegetables?

We can try a scatterplot!

Group: Underweight Adolescent Males



What if I further question the data? All relationships by Age, Gender?

How many of these did you really expect me to do?

These histograms were made in Excel

One can also use SAS, JMP, MatLab

In all of those programs we have to write scripts, modify variables, etc.

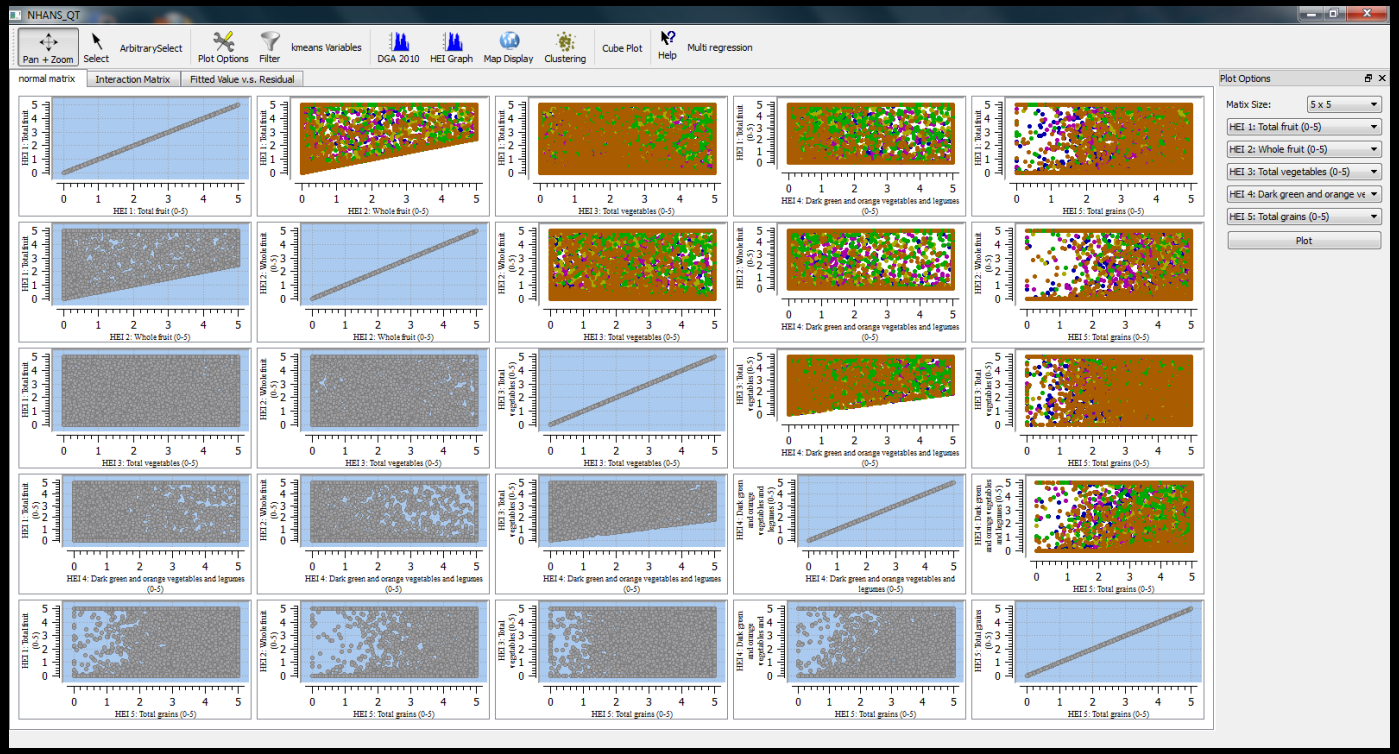
We can't directly manipulate our variables

So, being a data detective can be very tedious

And these were just simple comparisons!

So what do we do now?

We can utilize interactive graphics combined with underlying analytical processes to help us be better detectives!



The Curse of Dimensionality

A term coined by Bellman in 1961

Refers to the problems associated with multivariate data analysis as the dimensionality increases the available data becomes sparse

Sparsity is a problem for any method that requires statistical significance

Sometimes data dimensions are redundant and can be reduced with minimal information loss

In visualization we are also limited with screen space and the number of available visual variables, so choosing the most appropriate dimensions is key

Richard Ernest Bellman; Rand Corporation (1957). *Dynamic programming*. Princeton University Press.

The Curse of Dimensionality

So, what can we do?

- We can incorporate prior knowledge of the data
- We can smooth the target function
- We can reduce the dimensionality

The Practicality of the Curse of Dimensionality

For a given sample size, there is a maximum number of features above which the performance of classifying samples will degrade rather than improve

In most cases, the additional information that is lost when discarding some features is compensated by a more accurate mapping in the lower-dimensional space

So, how do we know what features we can throw away?

For visualization this implies that there are some features of a dataset that will be better to visualize (contain more information) than others!

Curses! I'm Still Overwhelmed!

Probably so, we are creating a lot of data in the world

Overall, NHANES is actually a relatively modest set of data

As our datasets get more variables we need ways to question the data

What I really want to know is are there relationships/groups/clusters in the data, how can I see these when we have so many dimensions?

One way is to just cluster the data and draw these clusters in an arbitrary space (we saw this in the demo)

Another is to try dimensional reduction!

Dimensional Reduction

Two approaches are available to reduce dimensionality

- **Feature extraction:** creating a subset of new features by combinations of the existing features
- **Feature Selection:** choosing a subset of all the features
- Given a feature space $x_i \in R^N$ find a mapping $y = f(x): R^N \rightarrow R^M$ with M<N such that the transformed feature vector $y \in R^M$ preserves (most of) the information or structure in R^N
- An optimal mapping is one that does not increase error

What Does Dimensional Reduction Mean For You?

Feature Selection: choosing a subset of all the features

- Well, that's what we were already doing
- We use our own expert knowledge to select features of interest
- Note we could also use machine learning and statistics here too if we were so inclined

What Does Dimensional Reduction Mean For You?

Feature extraction: creating a subset of new features by combinations of the existing features

- You can think of this in terms of the NHANES data
- Let's say I do an analysis and I find a correlation between SOFA consumption and sodium consumption (the more SOFAs, the more sodium)
- I could combine SOFAs and sodium into a new variable – SSOFAS
- I make up some formula to do this:
  - $SSOFAS = \alpha * SOFAs + \beta * sodium$
- So basically we just combine variables that are correlated

Principle Components Analysis

One of the most commonly applied dimension reduction techniques

PCA is a deterministic analytical procedure that utilizes an orthogonal transformation to reduce a set of sample observations with potentially correlated variables into a set of uncorrelated variables called principal components

The number of principal components will always be less than or equal to the original number of variables in the sample set

. T. Jolliffe, “Principal Component Analysis," *Springer Series in Statistics*, Springer, NY, 2002.

Principle Component Analysis

The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector

- PCA simple performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance
- There is no guarantee that the directions of maximum variance will contain the most interesting/important features

That sounded really hard!

- *Yes, but we have software to help us!*
- *In JMP Principal components can be accessed through the Principal Components command on the Analyze > Multivariate Methods menu.*

- *But, how does principal component analysis help me?*
- *What can principal component analysis tell me?*

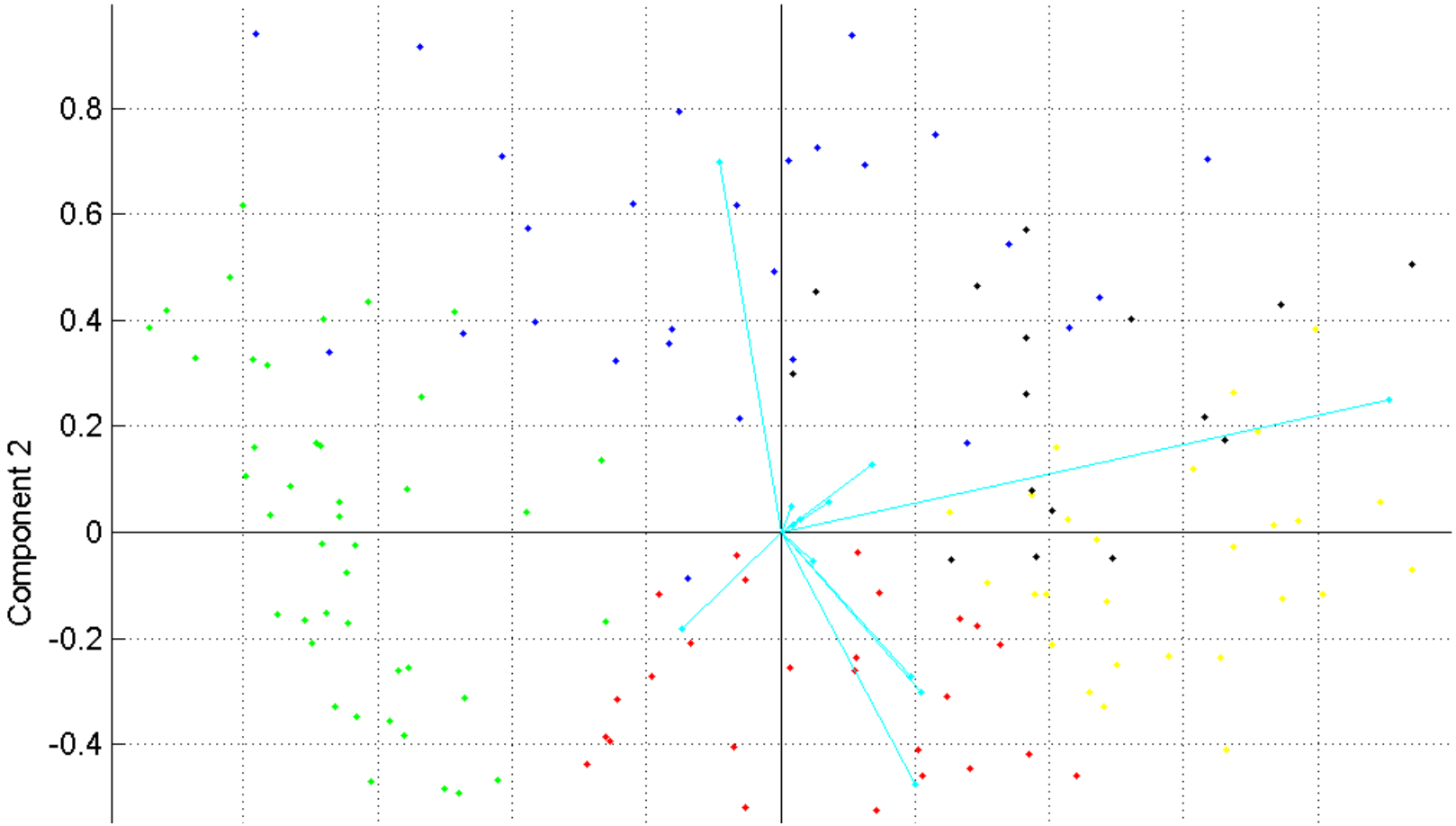
Principal Component Analysis

- *Let's go back to NHANES for an example and focus on the HEI values.*
- *Each person in my data set has 12 HEI values, so you can think of each person as a point in 12-dimensional space*
- *It is hard to understand all the relationships in this space, PCA tries to figure out what dimensions to combine for you automatically*
- *After you run PCA you will get a set of what you can think of as new variables*
- *Since there were 12 variables in our NHANES data, we will get at most 12 principal components (but hopefully less!)*

Principal Component Analysis

- *Ok, you've convinced me to run PCA on NHANES and I get 8 components, but now I'm confused!*
- *The output is such that your first component accounts for as much of the variability in the data set as possible*
- *The second component accounts for the second most variability, etc.*
- *Typically we will plot the data using the first two principal components as axes*
- *We can also write out each component to analyze it*

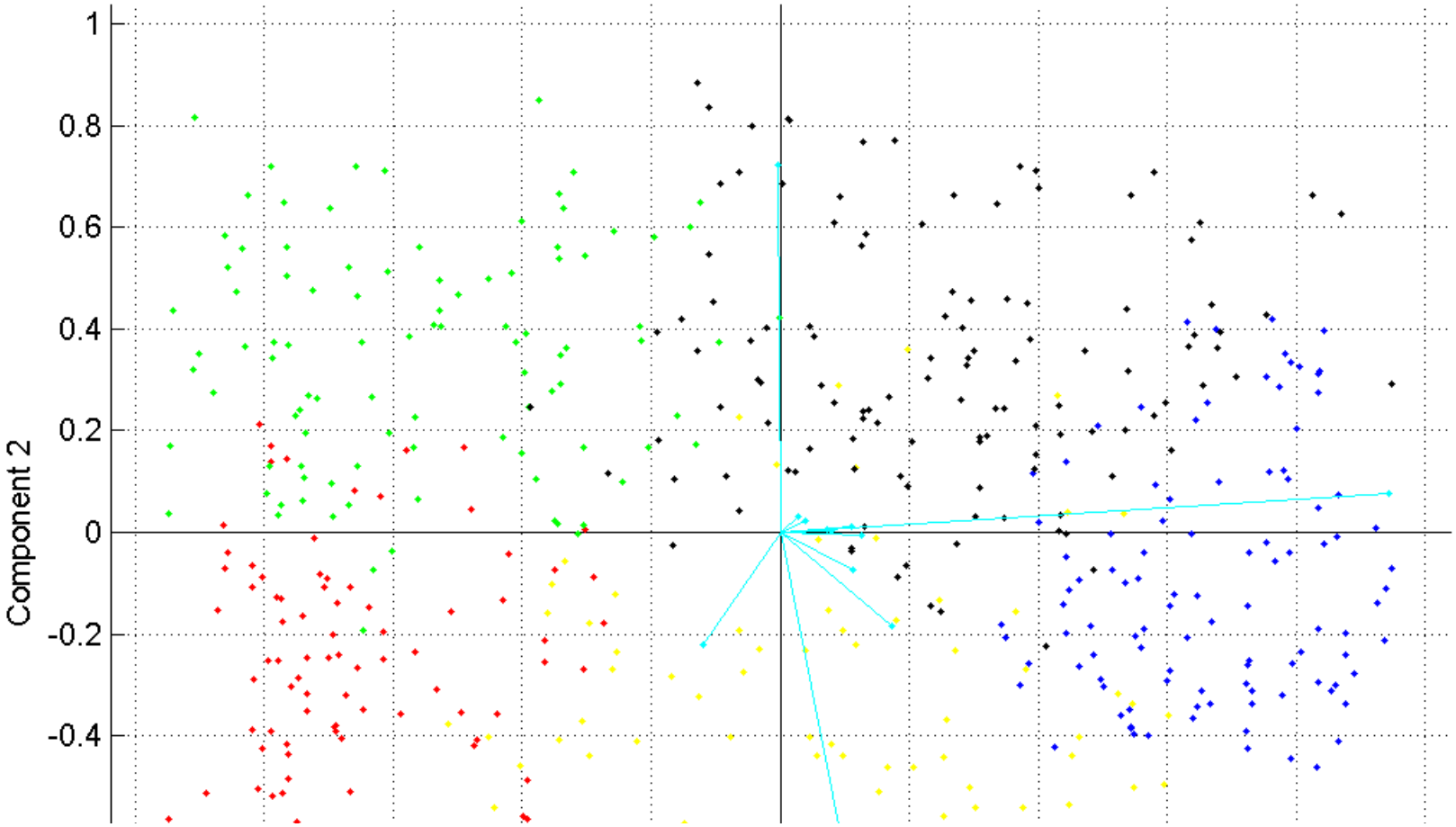
Underweight Adolescent Males



FV1 ≈ 0.14 HEI1 + 0.07 HEI2 +0.02 HEI5 + 0.02 HEI6 + -0.09 HEI7 + 0.19 HEI8 +0.05 HEI3 + 0.03 HEI4 + 0.20 HEI9 + 0.20 HEI10 + -0.15 HEI11 + 0.91 HEI12

FV2 ≈ 0.13 HEI1 + 0.06 HEI2 + 0.05 HEI5 + 0.01 HEI6 + 0.70 HEI7 + -0.27 HEI8 + -0.05 HEI3 + 0.02 HEI4 + -0.48 HEI9 + -0.30 HEI10 + -0.18 HEI11 + 0.25 HEI12

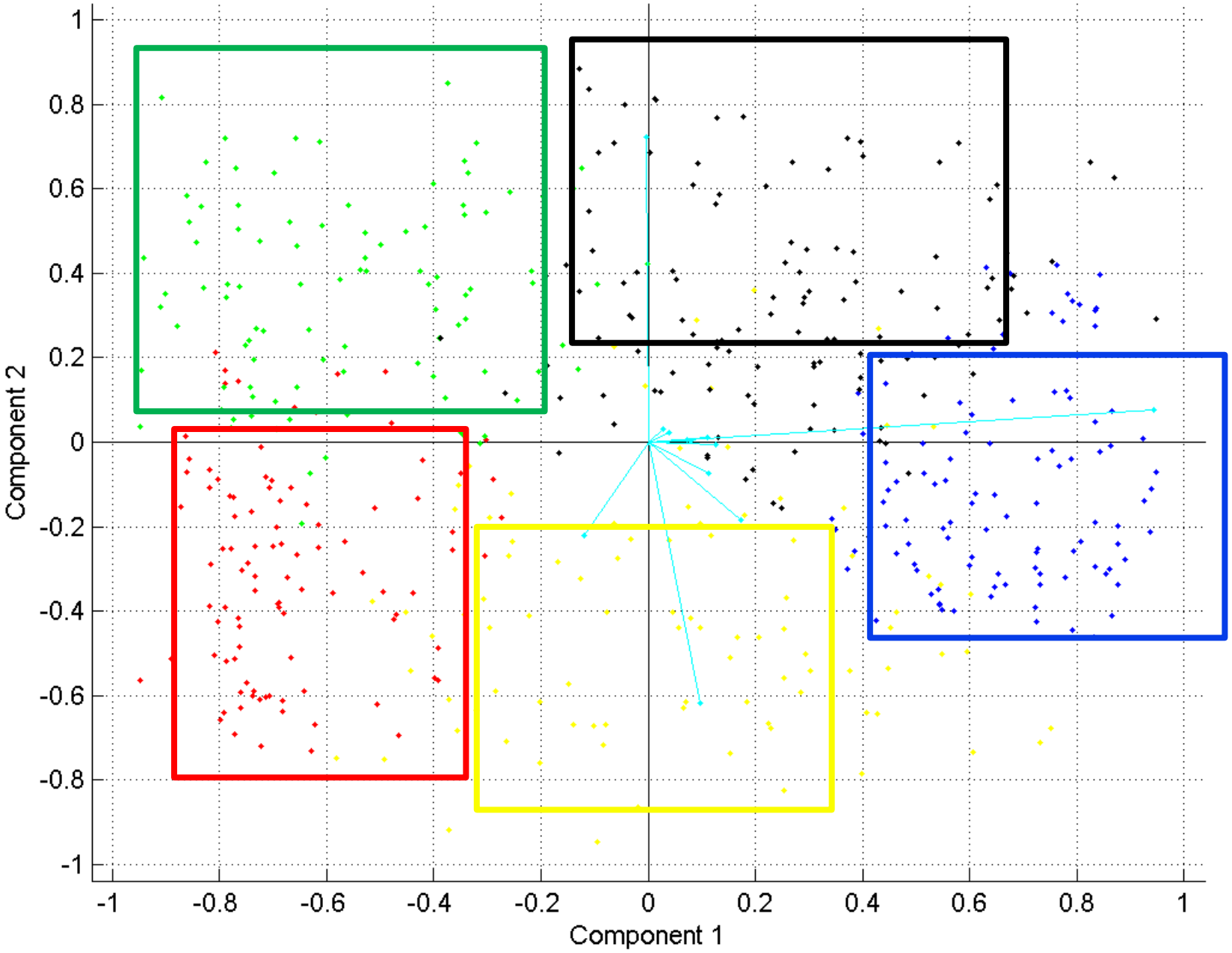
Females Age 19-30, Average BMI



FV1 ≈ 0.13 HEI1 +0.11 HEI2 +0.04 HEI5 + 0.03 HEI6 + -0.00 HEI7 + 0.11 HEI8 +0.08 HEI3 + 0.07 HEI4 + 0.17 HEI9 + 0.10 HEI10 +-0.12 HEI11 + 0.94 HEI12

FV2 ≈-0.01 HEI1 + 0.01 HEI2 +0.02 HEI5 + 0.03 HEI6 + 0.72 HEI7 + -0.07 HEI8 + 0.00 HEI3 + 0.00 HEI4 + -0.18 HEI9 +-0.62 HEI10 +-0.22 HEI11 +0.08 HEI12

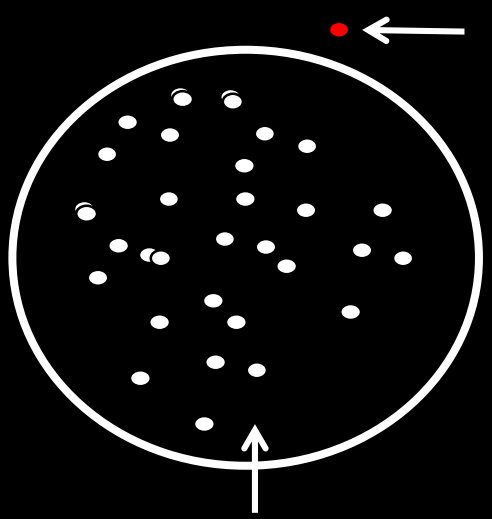
Females Age 19-30, Average BMI



Putting it all together

- ***If I know things about you …***
  - Age, gender, BMI
- ***Can I guess your typical daily diet?***
  - Probably not very accurately
- ***But, if I ask you a few questions, maybe I can get closer!***

Putting it all together



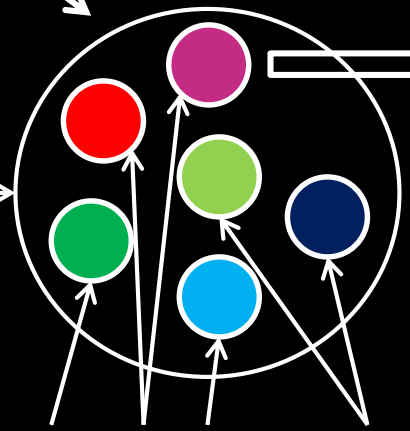
This is you.
Maybe you tell me your age, gender, BMI?
(Inform me maybe?)

This is the NHANES data
I know everyone's diet

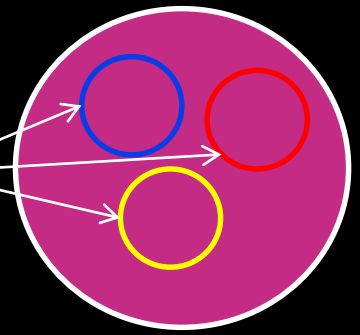
I can first split the data by demographics

I can stick you in one of these groups based on your demographics

Total Population



Split by demographics
(Learned or *a prior*)



Now my demographics groups having groups based on diet!

Now I know your diet!
(Well sort of)

This is still you.
I know you're in this large group
But I want to know your subgroup.

If I know how many sodas you drink, maybe I can narrow it down from three groups to two

Maybe if you ate a salad today tells me more!

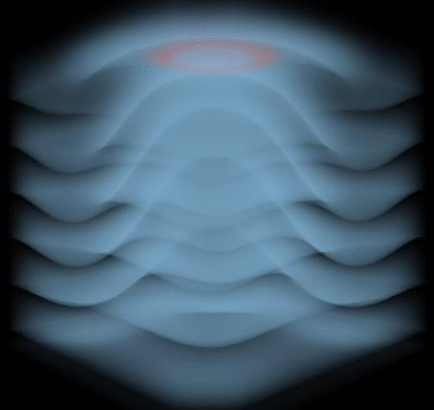
?

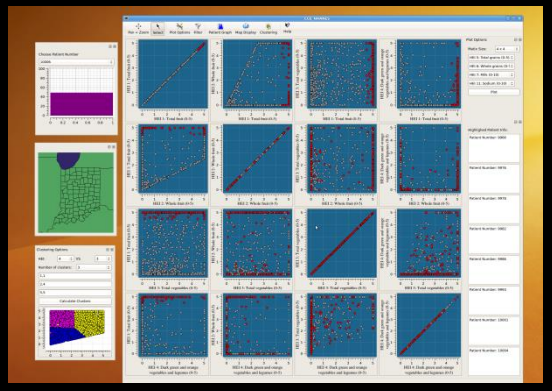
?

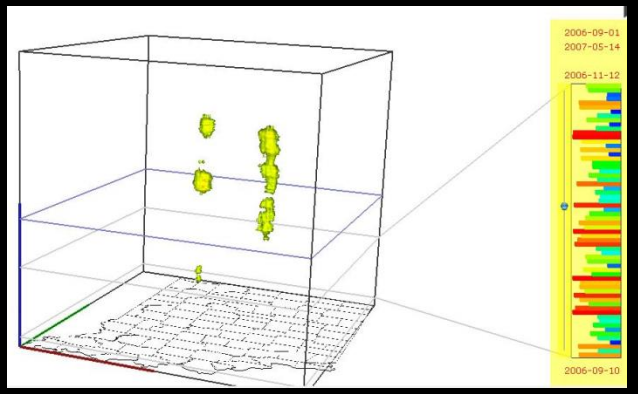
Summary

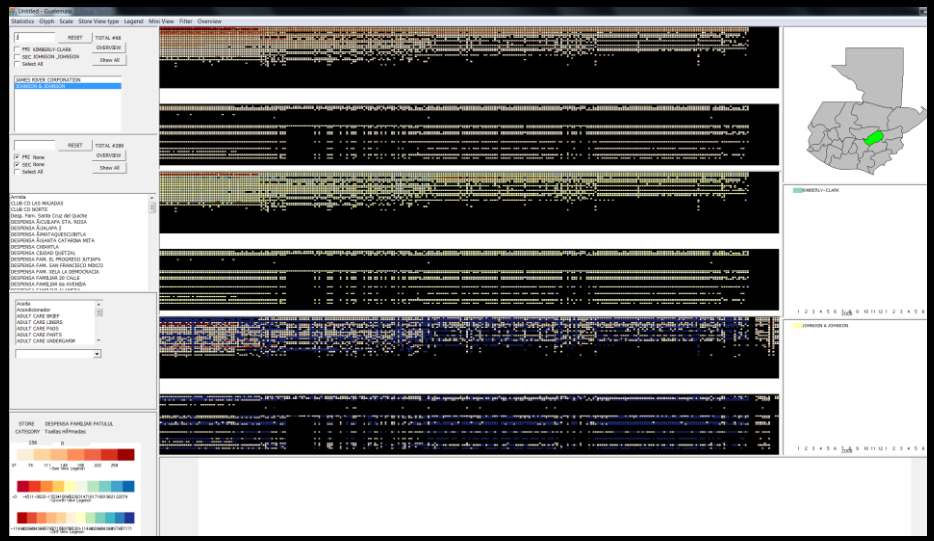
- ***Data analysis is hard!***
- ***We need better tools to be data detectives***
- ***By being data detectives, we can ask questions and maybe create predictions about unknown data based on available data (we make a model of the data)***
- ***Visual analytics helps us in the data detective work!***
- ***But we still need to understand what the analyses do for us so that we can put them into practice.***

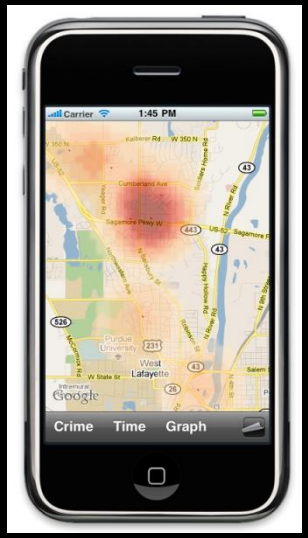
Questions? (http://rmaciejewski.faculty.asu.edu)











VizBi '15 Keynote
March 25, 2015

Data Visualization Principles (Some Things I've Learned)

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu



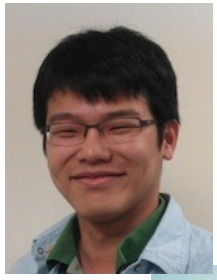


Georgia Tech College of Computing
School of Interactive Computing

Three themes

1. Understanding purpose
2. Embracing interaction
3. Identifying value

Talented students



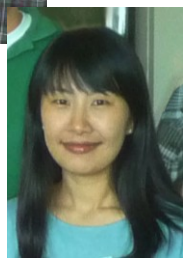
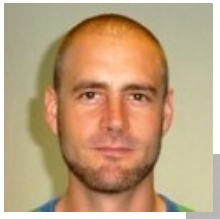
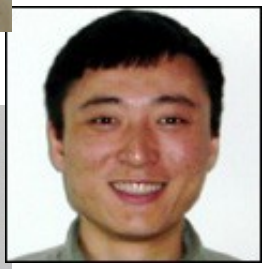




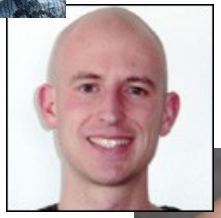


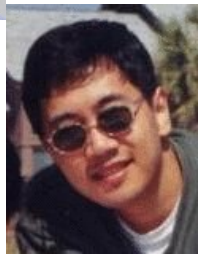


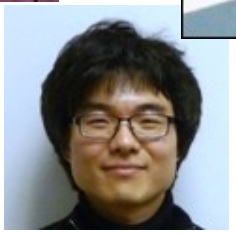
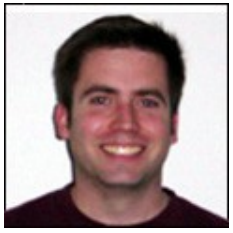







Why visualize?

Applications of Visualization

Presentation	(Explanatory)
Analysis	(Exploratory)

1. Presentation

Communicate data and ideas

Explain and inform

Provide evidence and support

Influence and persuade



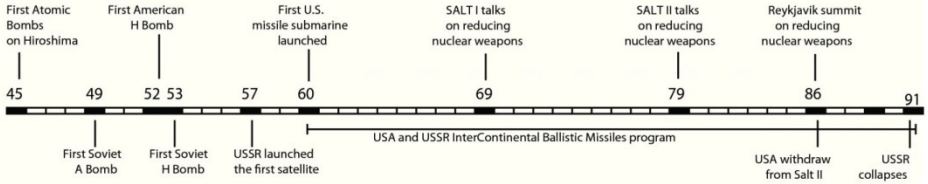
Infographics

All the rage…

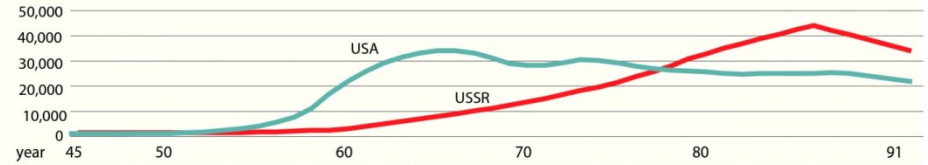
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



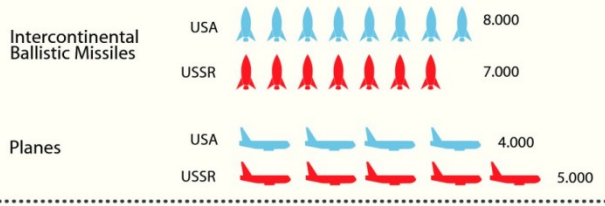
NUCLEAR STOCKPILE



ROCKET MODELS



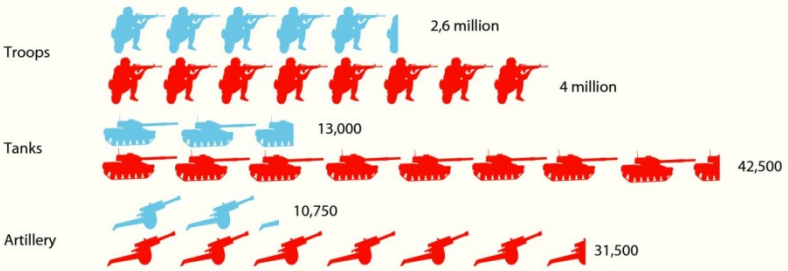
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.





Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

NY Times
2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

Bonds was injured last season. He played 14 games and hit 6 homers.

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

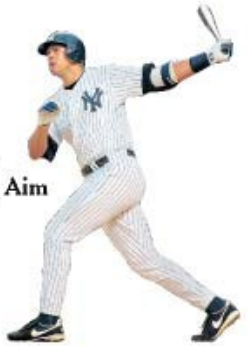
Aaron
Actual homers slightly outpace projected homers for five seasons.

Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

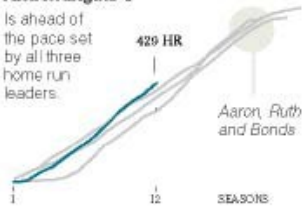
Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

Note: Ages as of July 1 of each season.

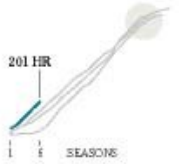
Others Taking Aim



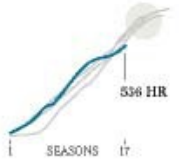
Alex Rodriguez
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

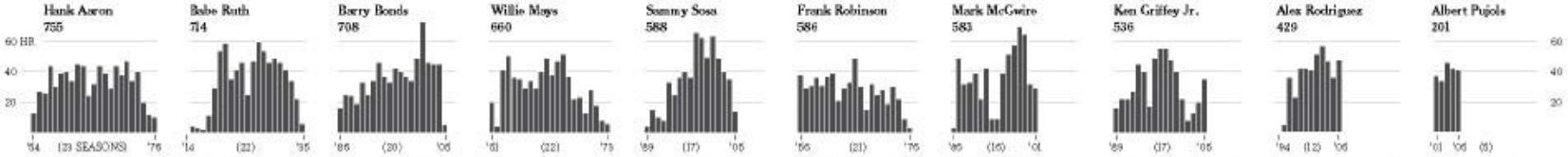


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

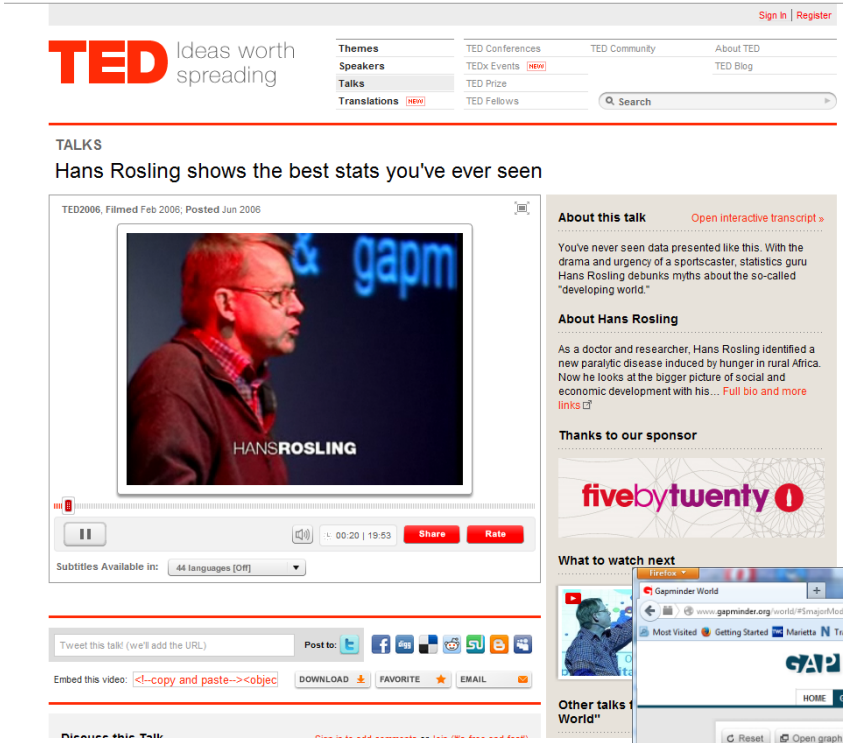


Differing Paths to the Top of the Charts

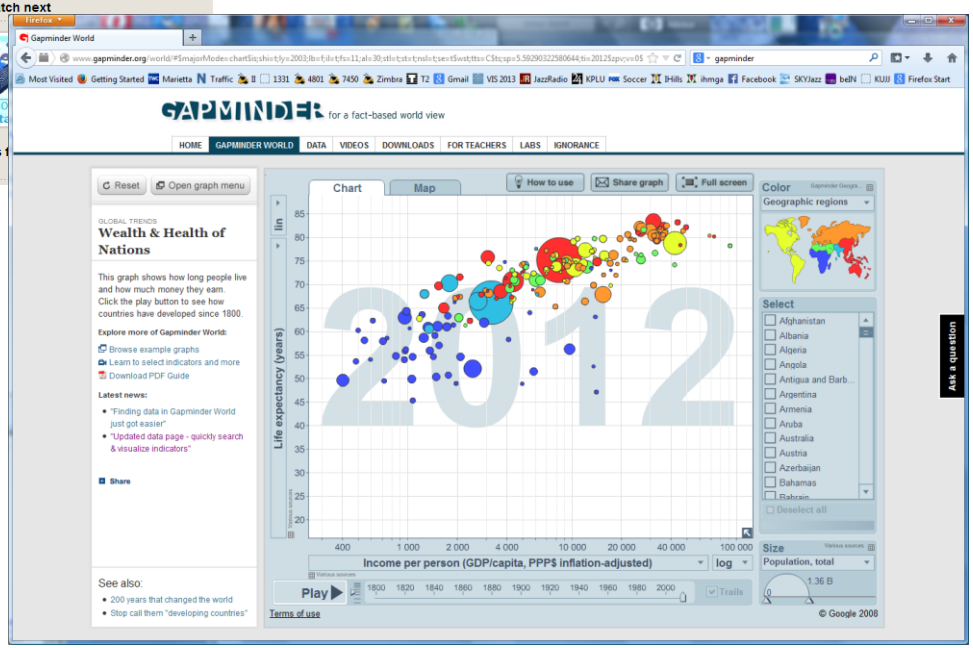
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



Amanda Cox and Joe Ward/The New York Times



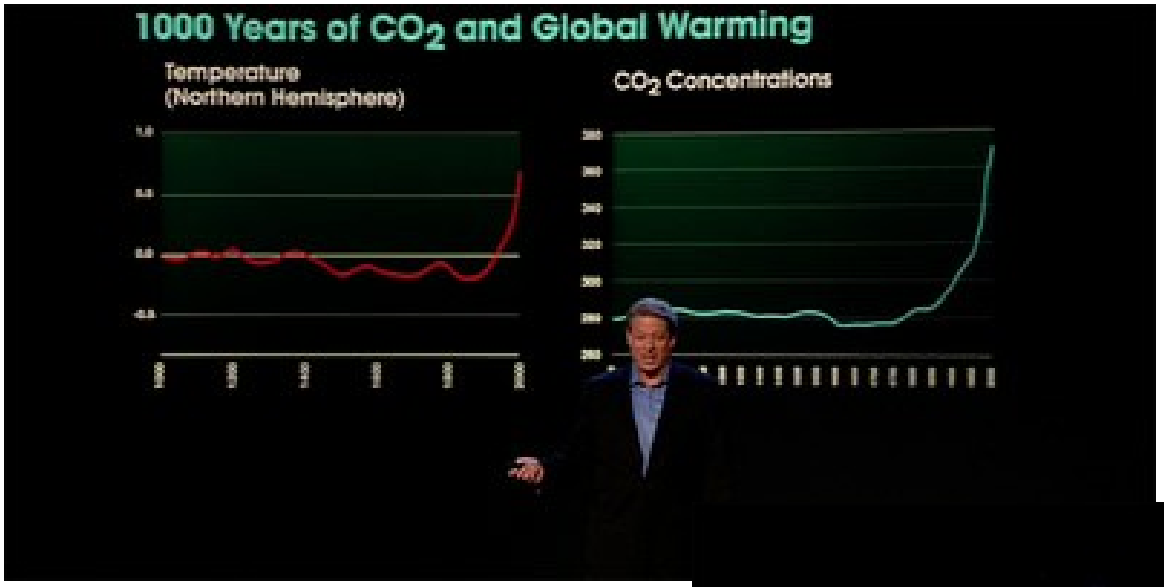
Hans Rosling
Gapminder



2006

http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

An Inconvenient Truth





Gore made extensive use of data graphics

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

Persuade

May just show a few variables
and/or a subset of the data cases



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





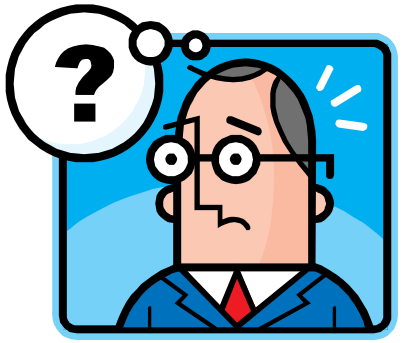
Thought

If you can articulate very precisely what you're seeking, visualization likely isn't your best approach

OK, so what is visualization good for?

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Frequent analysis goals

Show many/all variables

Illustrate overview and detail

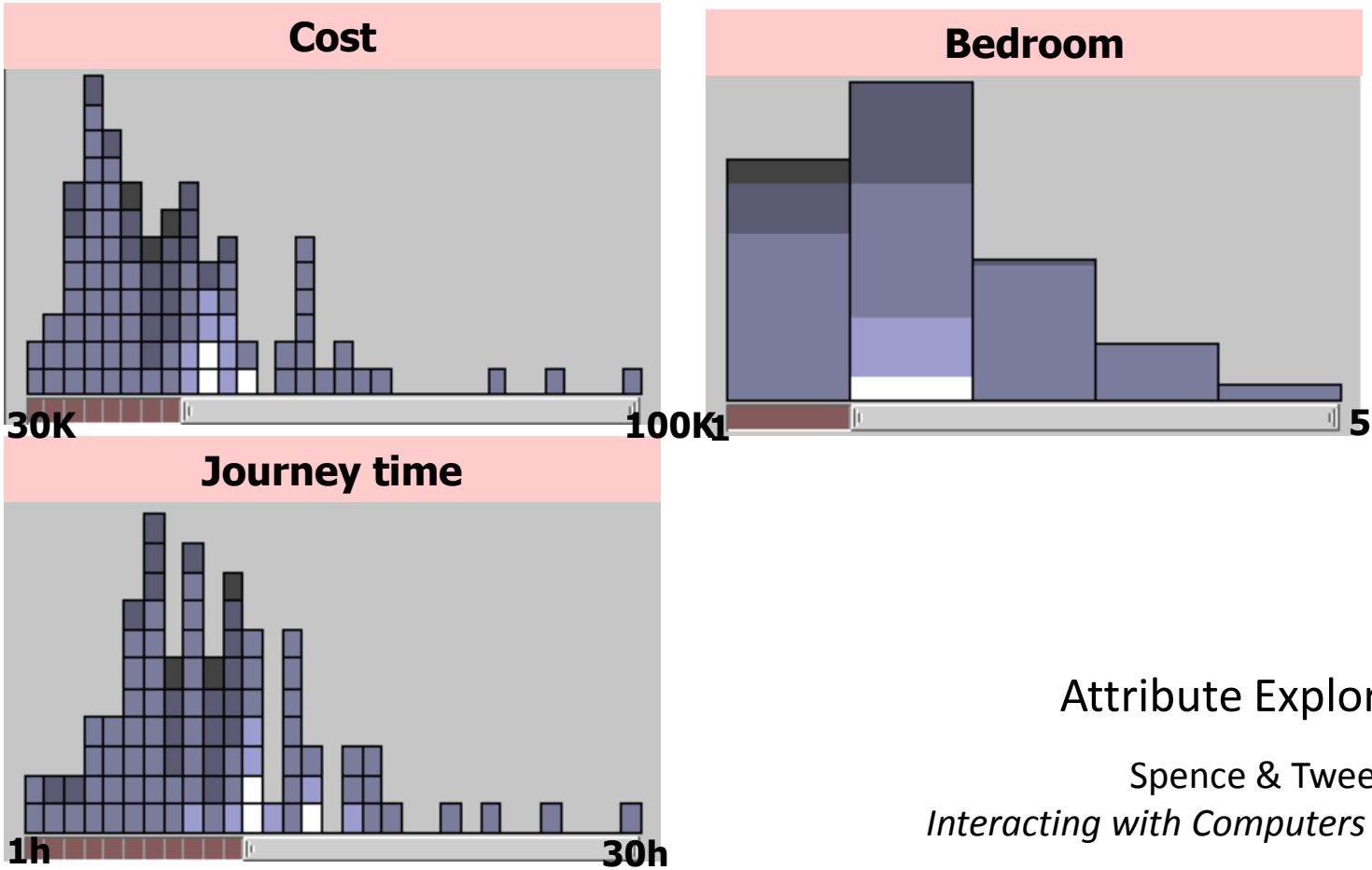
Facilitate comparison

Display may not be easy to interpret at first

At Its Heart: Uncertainty

Nothing is perfect

Humans trade off choices, compromise



Attribute Explorer

Spence & Tweedie

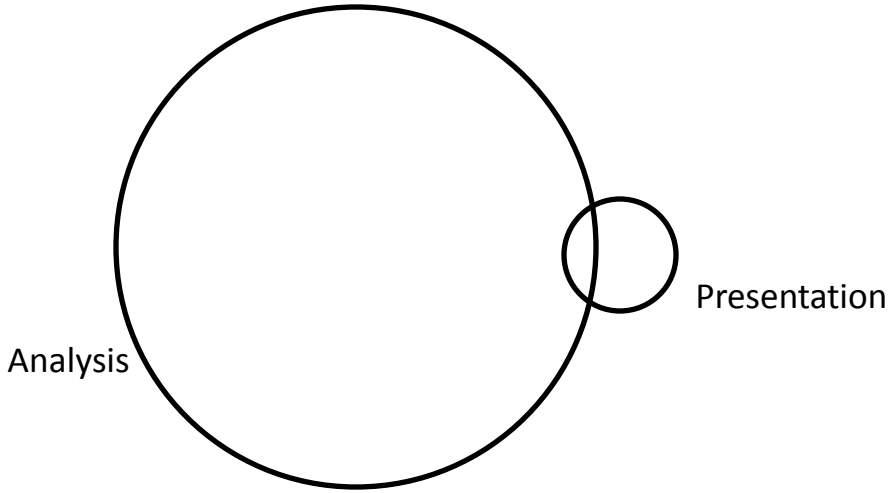
Interacting with Computers '98

Takeaway 1

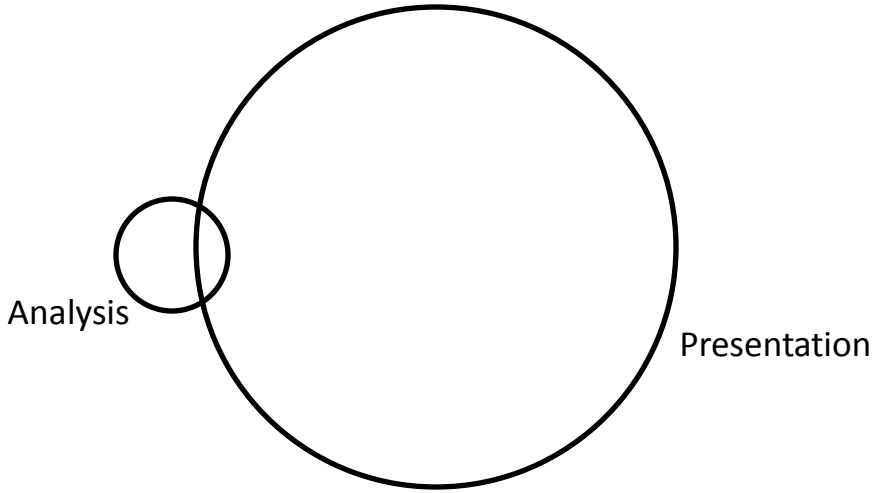
Analysis $\neq$ Presentation

Views of the World

Vis researchers



Everyone else





Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these "data stories" differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and messaging, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the interface) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 Introduction

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reporters use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent feature in The Economist [6] explores the proliferation of digital data and notes that visualization designers are "*melding the skills of computer science, statistics, artistic design and storytelling.*"

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful "data stories" requires a diverse set of skills. Gershon and Page [12] note that effective story-telling "*require[s] skills like those familiar to movie directors, beyond a technical expert's knowledge of computer engineering and science.*" While techniques from oration, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data similarly can be organized in a linear sequence, they can also be interactive, inviting verification, new questions, and alternative explanations.

Currently, most sophisticated visualization tools focus on data exploration and analysis. Applications such as spreadsheets and visualization tools support an array of analysis routines and visual encodings, but beyond exporting images for presentation typically provide scant support for crafting stories with analysis results. As such, they provide powerful vehicles for discovering "stories", but do little to aid narrative communication of these findings to others. As tools mature and more richly integrate with the web (e.g., Many Eyes [25], Tableau Public [22], GeoTime Stories [8]), they are enabling the publication of dynamic graphics with variably constrained levels of interactivity. It remains an open question how the design of such tools might be evolved to support richer and more diverse forms of storytelling.

In this paper, we investigate the design of narrative visualizations and identify techniques for telling stories with data graphics. We take an empirical approach, analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. After reviewing related work, we share five selected case studies which highlight varied design strategies and illustrate our analytic approach. We then formulate a design space constructed from an analysis of 58 examples. Our analysis identifies salient dimensions of visual storytelling, including how graphical techniques and interactivity can enforce various levels of structure and narrative flow. We describe seven genres of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and video. These genres can be combined with interactivity and messaging to produce varying balances of author-driven and reader-driven experiences. Finally, we discuss the implications of our framework, noting recurring design strategies, promising yet under-utilized approaches to integrating visualization with other media, and the potential for improved user interfaces for crafting data stories. By focusing on the graphical and interactive elements of narrative visualization, our approach gives less attention to the cognitive and emotional experience of the reader. We recognize the importance of these elements, however, and describe directions for future reader-centric research in our conclusion.

2 Related Work

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as "*the world's second-oldest profession*" [12]. Without summarizing millennia of achievement, we describe a few of the key concepts informing narrative visualization.

2.1 Narrative Structure

The Oxford English Dictionary defines *narrative* as "an account of a series of events, facts, etc., given in order and with the establishing of connections between them." Central to this definition is the notion of a chain of causally related events. Stories of this form often have a beginning, middle, and end [3, 24]: an introduction to the situation, a series of events often involving tension or conflict, and a resolution.

Since ancient times, people have tried to understand and formalize the elements of storytelling. For example, writers (e.g., [5, 19, 21]) have developed typologies of dramatic situations and identified plot lines common to many narratives, such as the "hero's journey" [5]. This research typically distinguishes between the content of the story and the form in which it is told. While stories often concern interacting characters, they may also present a sequence of facts and observations linked together by a unifying theme or argument.

Storytelling strategies vary among media and genre. For instance, stories told through writing have access to a different set of formal mechanisms and narrative structures (e.g., stream of consciousness) than stories told through film (e.g., split-screen sequences [3]). Blundell [2] describes narrative devices for journalism such as the *anecdo-*

• *The authors are with Stanford University, Stanford, CA 94305. E-mail: {esegel, jheer}@stanford.edu.*

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

1077-2626/10/$26.00 © 2010 IEEE Published by the IEEE Computer Society

Jessica Hullman
Nick Diakopoulos
…

Static ok, Interactive better

Data Visualization 101



Problem:
You have a lot of data (& attributes) to understand

Do you?

- Pack all the data into one complex representation
- Spread the data into multiple coordinated views
- Use interaction to reveal different subsets of the data

Constituents

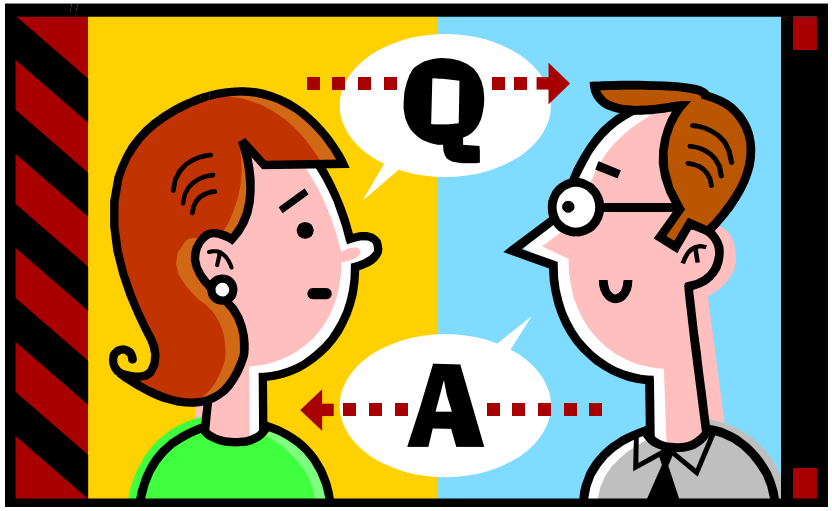
Two key aspects of data visualization

- Representation
- Interaction

> "The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."
>
> S. Few, *Now you see it*

Interaction is Vital

Engage in a dialog with your data



Interaction

Why interact?

Intent

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko & Jacko
TVCG (InfoVis) '07

Chad Stolper



Interaction

How manifested today?

Tooltips & selection

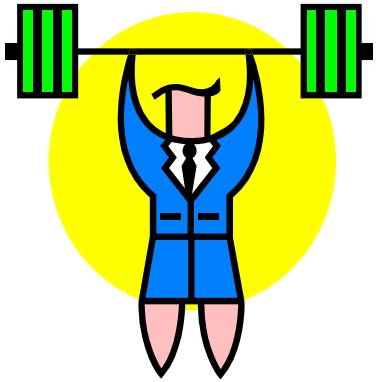
Get details

Navigation

Brushing & linking

Interaction

Can we do more?

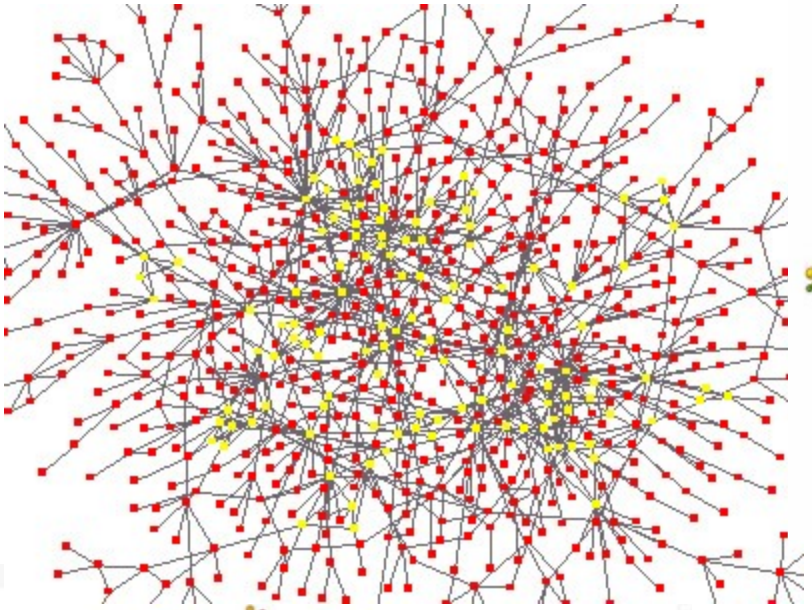


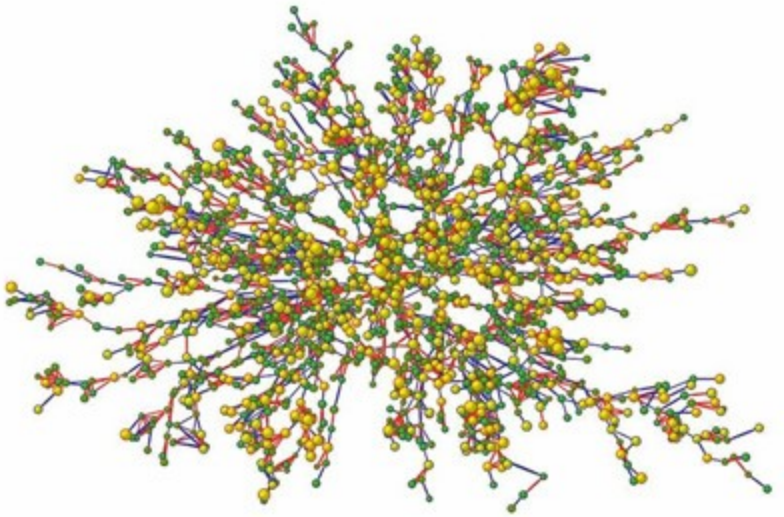
Employ interaction in a more fundamental manner to strengthen the power of visualization

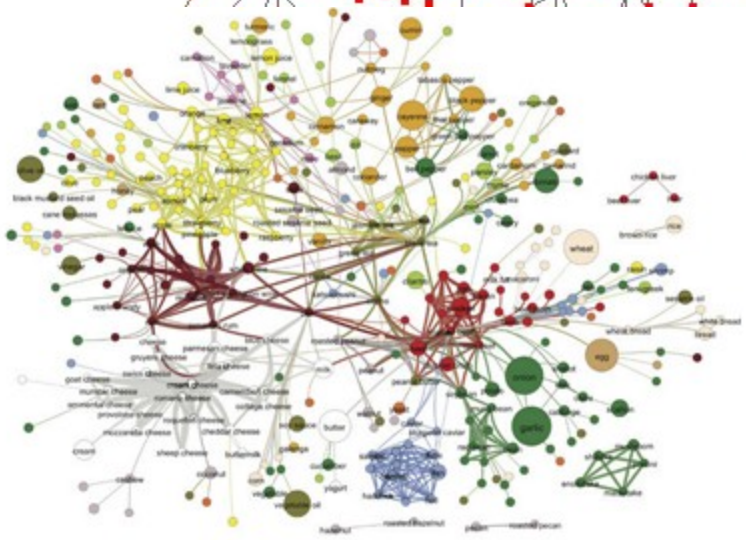
Some examples…

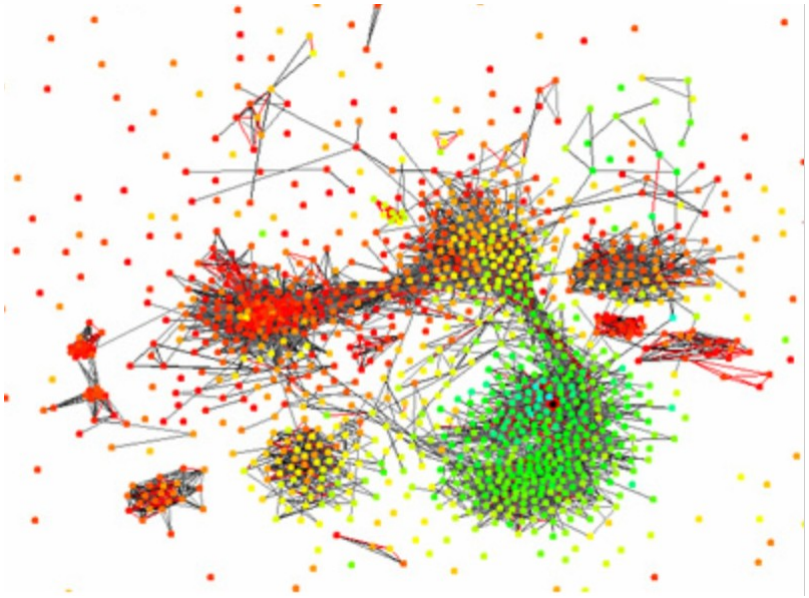
Node-link graphs

http://www.visualcomplexity.com/vc/



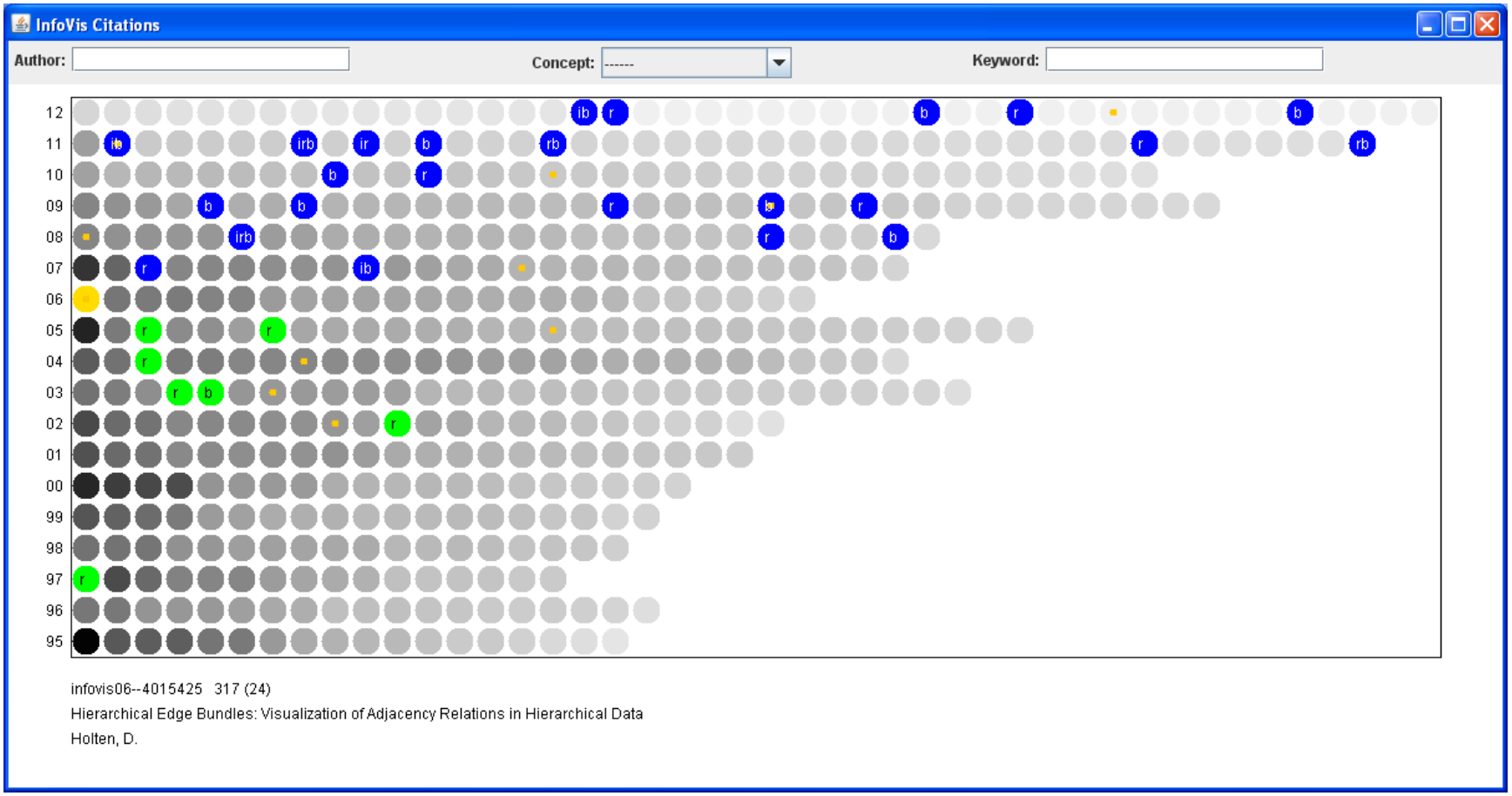






CiteVis

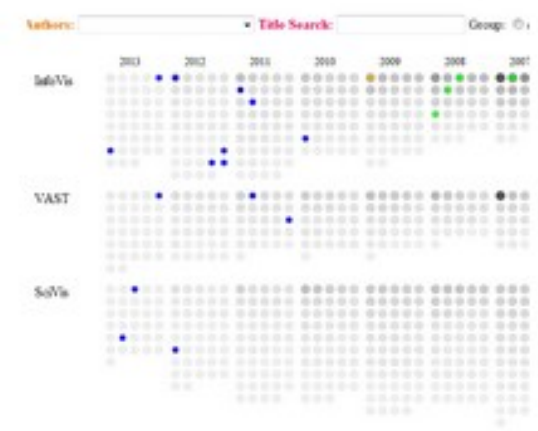
http://www.cc.gatech.edu/gvu/ii/citevis



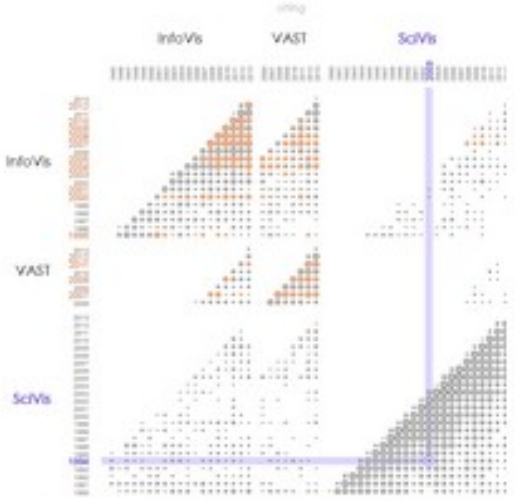
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

VIS25

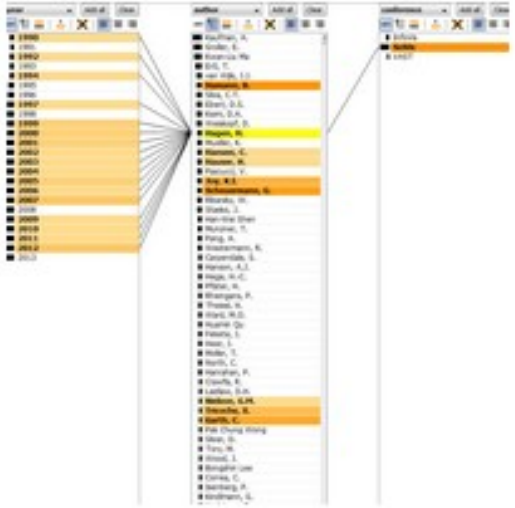
http://www.cc.gatech.edu/gvu/ii/citevis/VIS25/



CiteVis2



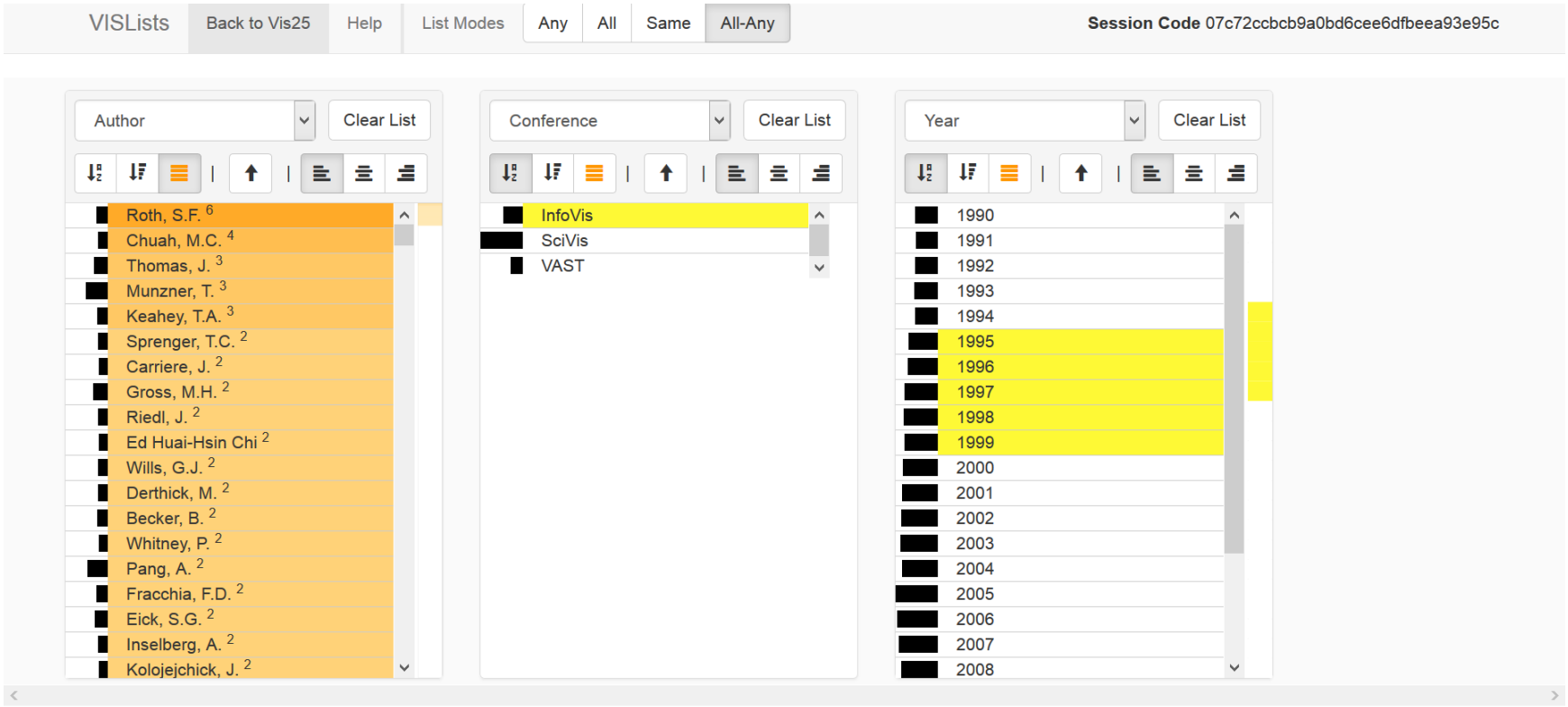
CiteMatrix



VISLists

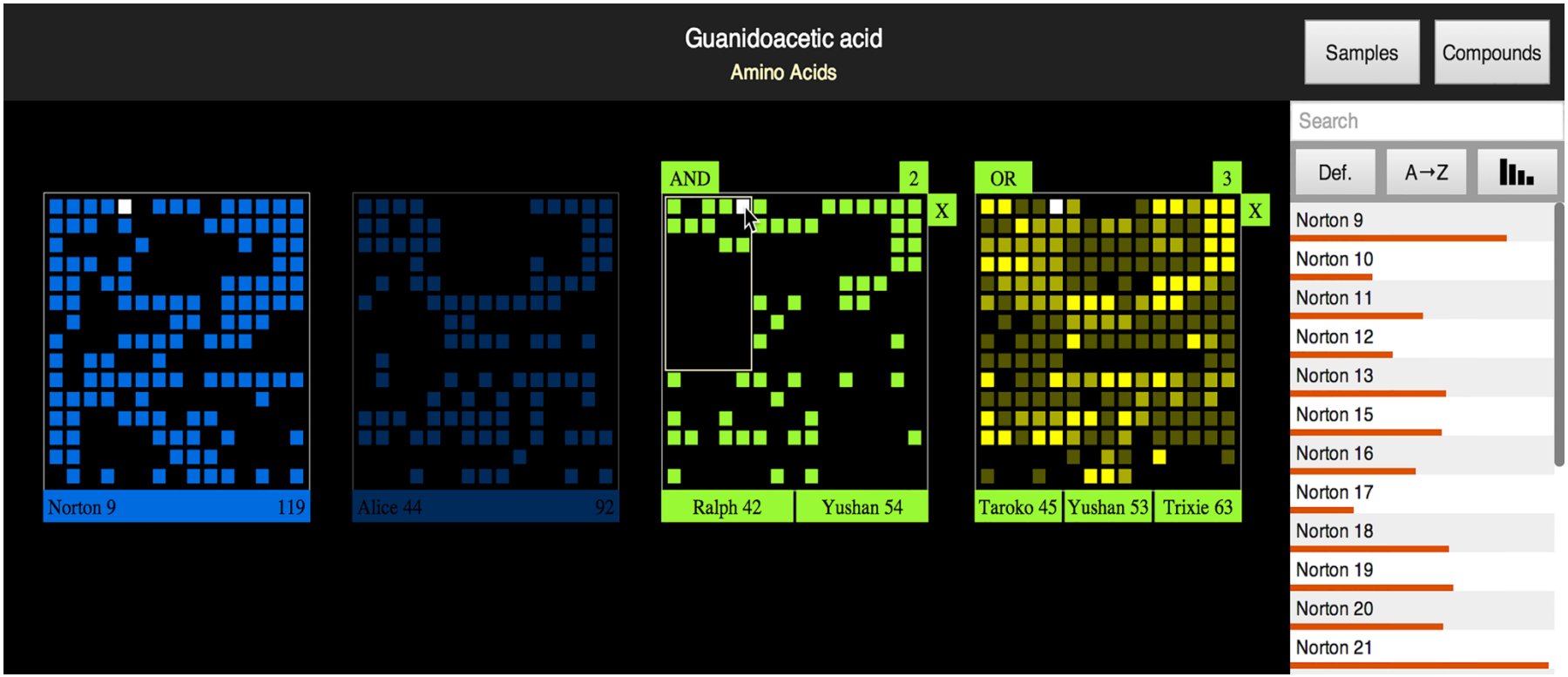
Visualization design: Panpan Xu, Chad Stolper, Anand Sainath

http://vislists.anandsainath.com/list/vis25



OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/



Demo

Sadana, Major, Dove, & Stasko
TVCG (InfoVis) '14

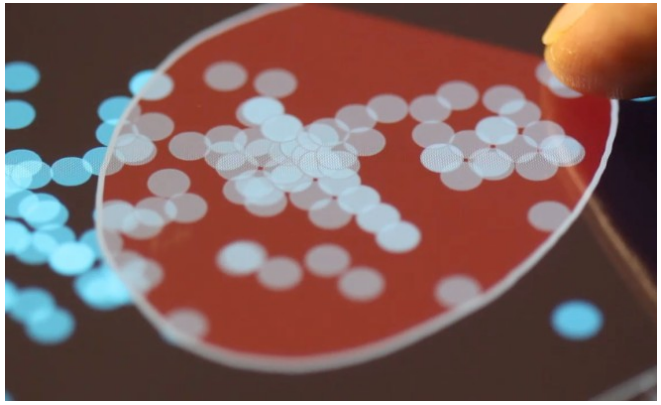
Interaction

What are the tools of interaction?

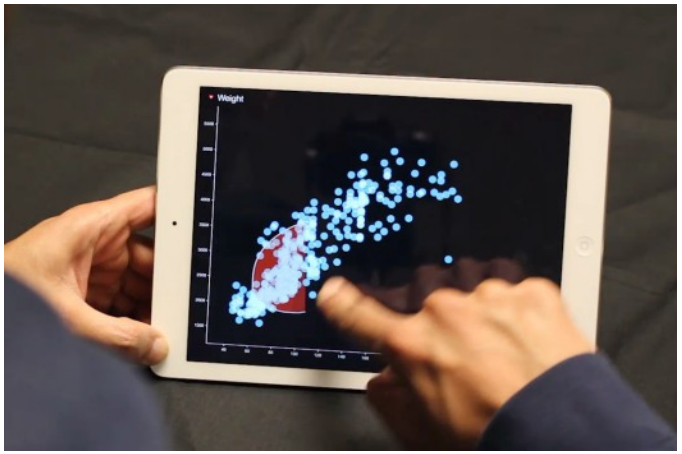
Traditional – Desktop: keyboard, mouse

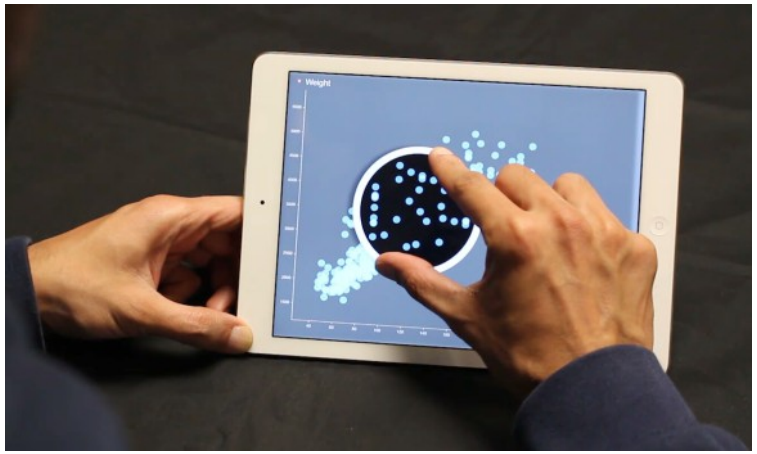
New – Tablet: fingers and multi-touch

Moving to Tablets



Scatterplot





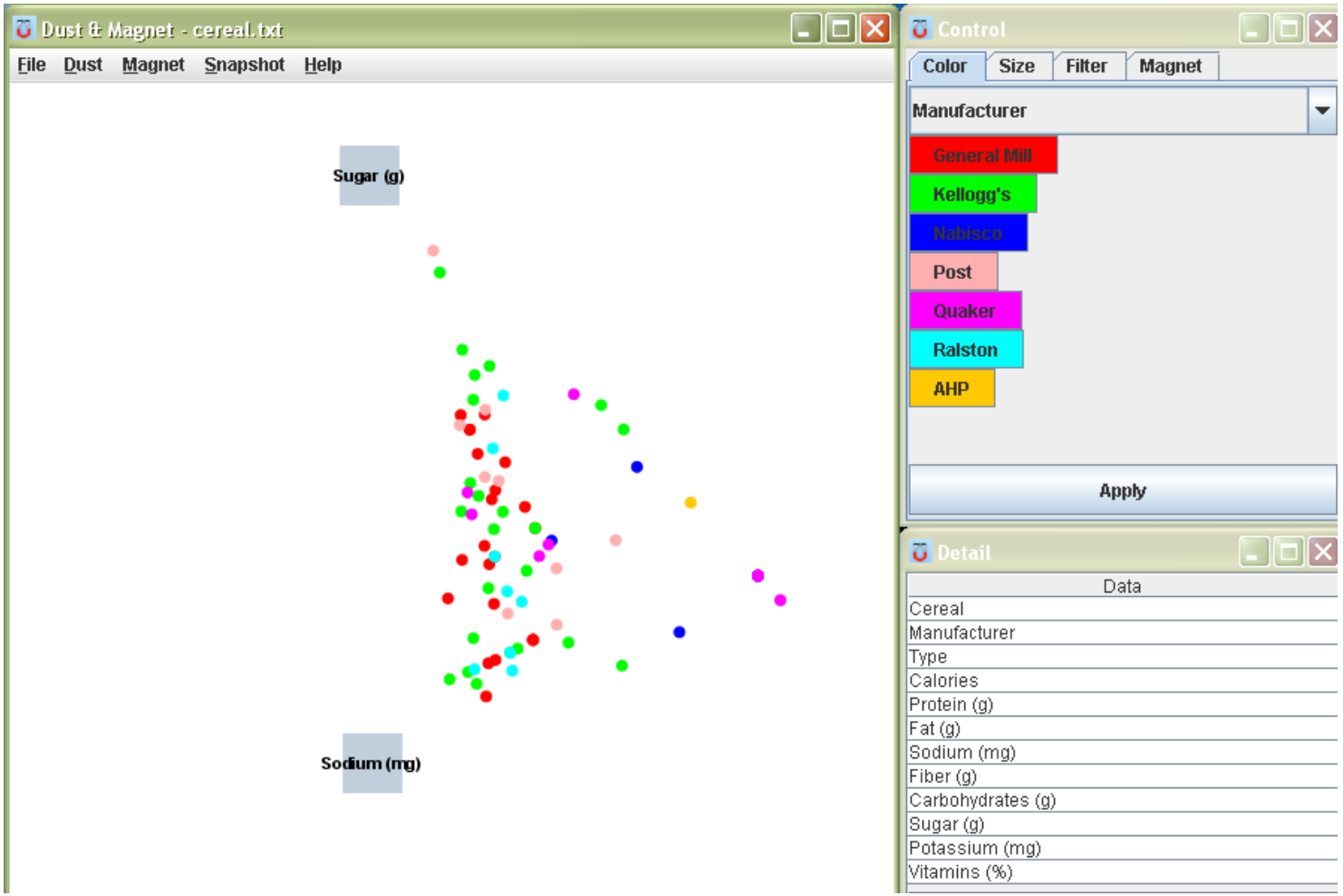


Video

Sadana & Stasko
AVI ‘14

Scatterplot touch video

Dust and Magnet



Yi, Melton, Stasko & Jacko
Information Visualization '05

Pushing it Further

Perceptive Pixel D n M

Takeaway 2

Take advantage of interaction



Visualization's Value

The Value of Information Visualization

Jean-Daniel Fekete[1], Jarke J. van Wijk[2], John T. Stasko[3], and Chris North[4]

[1] Université Paris-Sud, INRIA, Bât 490,
F-91405 Orsay Cedex, France,
Jean-Daniel.Fekete@inria.fr,
http://www.aviz.fr/~fekete/
[2] Department of Mathematics and Computing Science,
Eindhoven University of Technology, P.O. Box 513,
5600 MB EINDHOVEN, The Netherlands,
vanwijk@win.tue.nl,
http://www.win.tue.nl/~vanwijk/
[3] School of Interactive Computing, College of Computing & GVU Center,
Georgia Institute of Technology, 85 5th St., NW,
Atlanta, GA 30332-0760, USA,
stasko@cc.gatech.edu,
http://www.cc.gatech.edu/~john.stasko
[4] Dept of Computer Science, 2202 Kraft Drive,
Virginia Tech, Blacksburg, VA 24061-0106, USA,
north@vt.edu,
http://people.cs.vt.edu/~north/

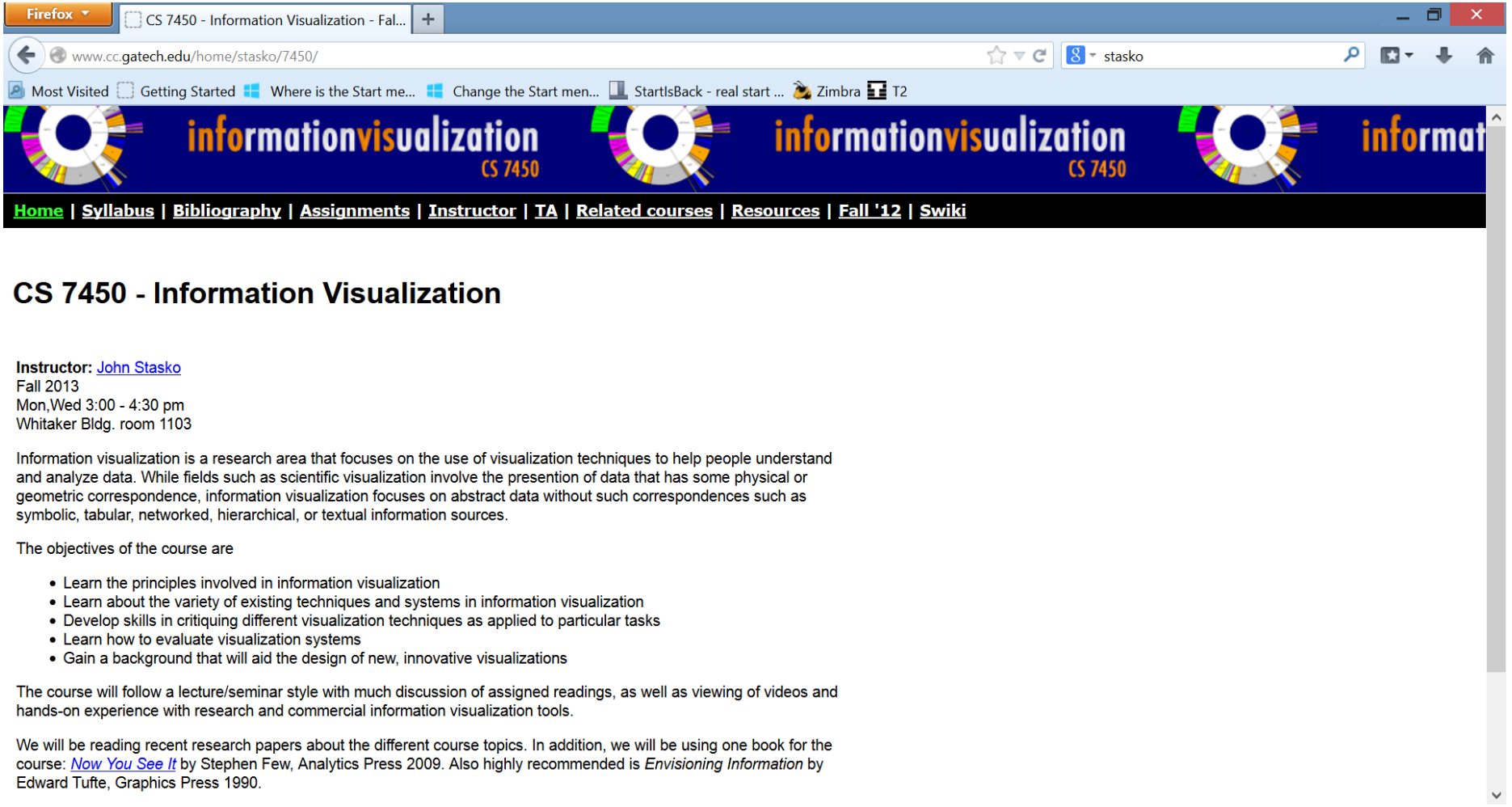
Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic

Information Visualization: Human-Centered Issues and Perspectives
Springer, 2008

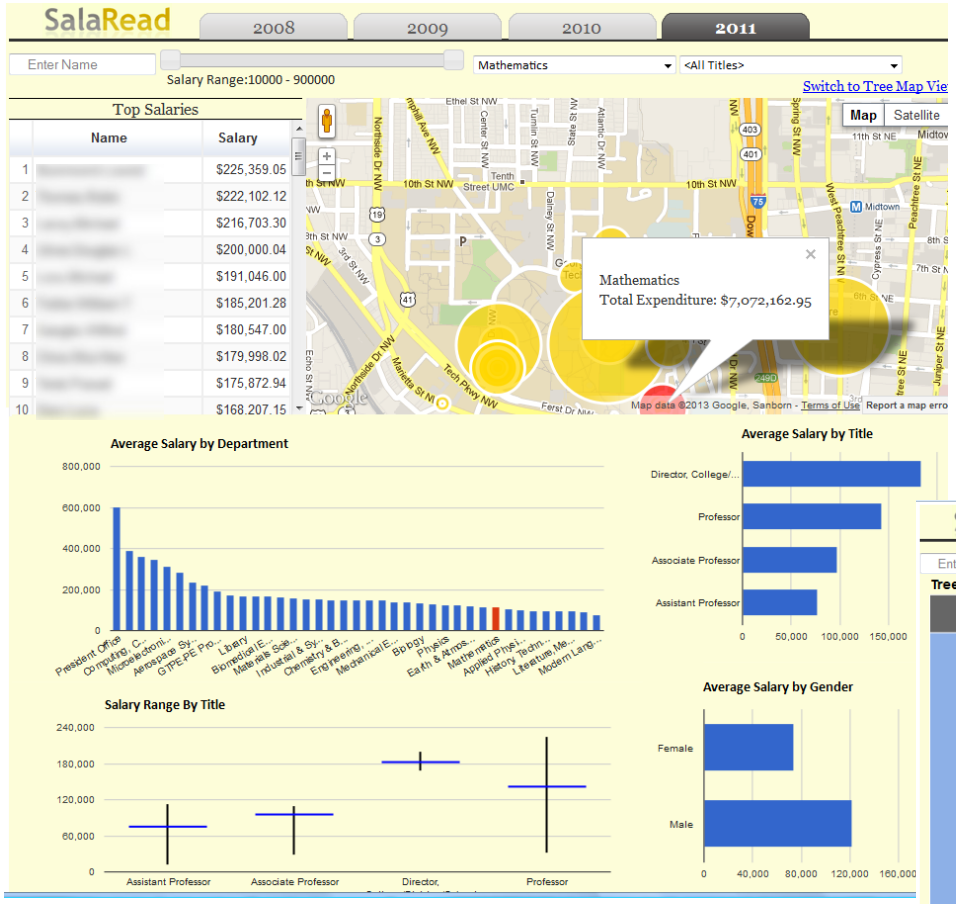
Simply presenting data ***visually*** can have a profound impact

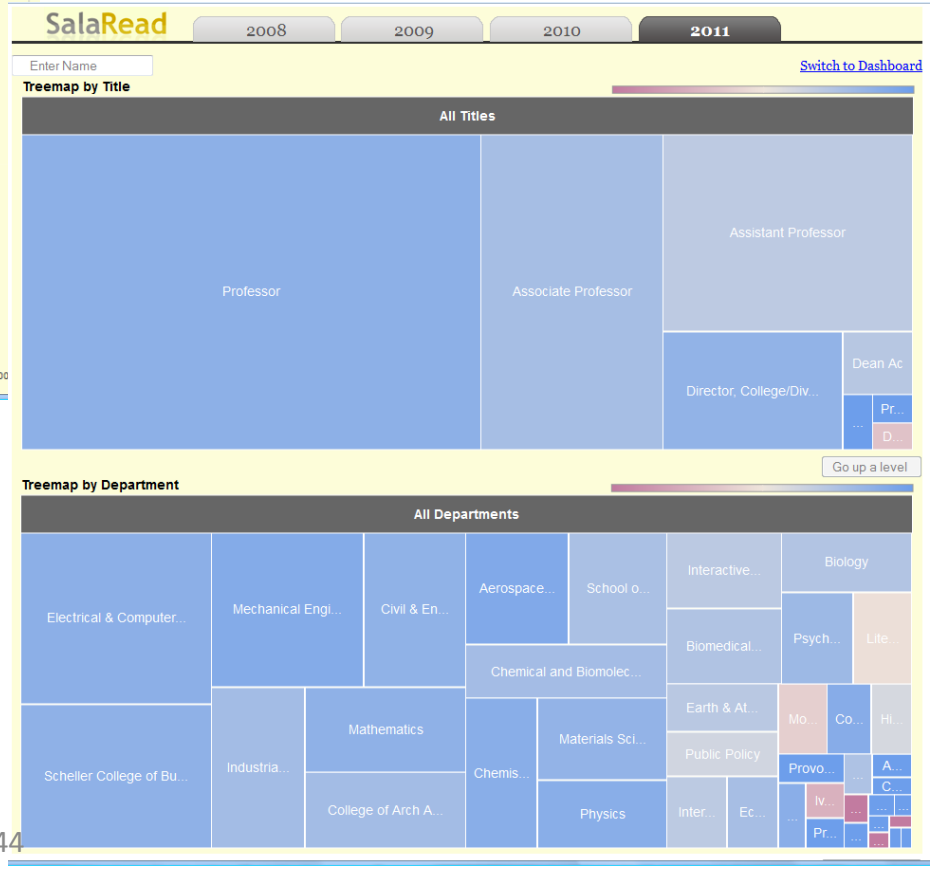
My Class

http://www.cc.gatech.edu/~stasko/7450



CS 7450
Fall ‘12
Design project

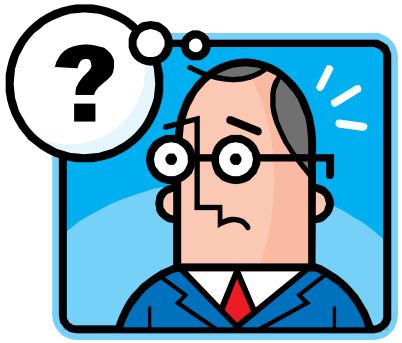




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Visualization's Value?

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Stasko
ACM BELIV '14

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)

What kinds of questions?

"Low-level" tasks

Retrieve value
Filter
Compute derived value
Find extremum
Sort
Determine range
Characterize distribution
Find anomalies
Cluster
Correlate

Amar, Eagan & Stasko
InfoVis '95

Visualization's Value?

$$V_{alue} = T + \mathbf{I} + E + C$$

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)

What is Insight?

An individual observation about the data by the participant, a unit of discovery

Complex
Deep
Qualitative
Relevant
Unexpected

Saraiya, North, & Duca
TVCG '05

North
IEEE CG&A '06

What is Insight?

Sudden grasp of new relationships that are necessary to solve a problem and that were not learned in the past

Bernstein, Penner, Clarke-Stewart & Roy
Psychology, 6th edition

What is Insight?

Is not spontaneous “aha!” moments (eg, in cognitive science)
Is knowledge-building and model-confirmation
    Like a substance that people acquire with the aid of systems

Chang, Ziemkiewicz, Green, & Ribarsky
IEEE CG&A ‘09

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to convey an overall **essence** or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)

Overview and detail

Focus + context

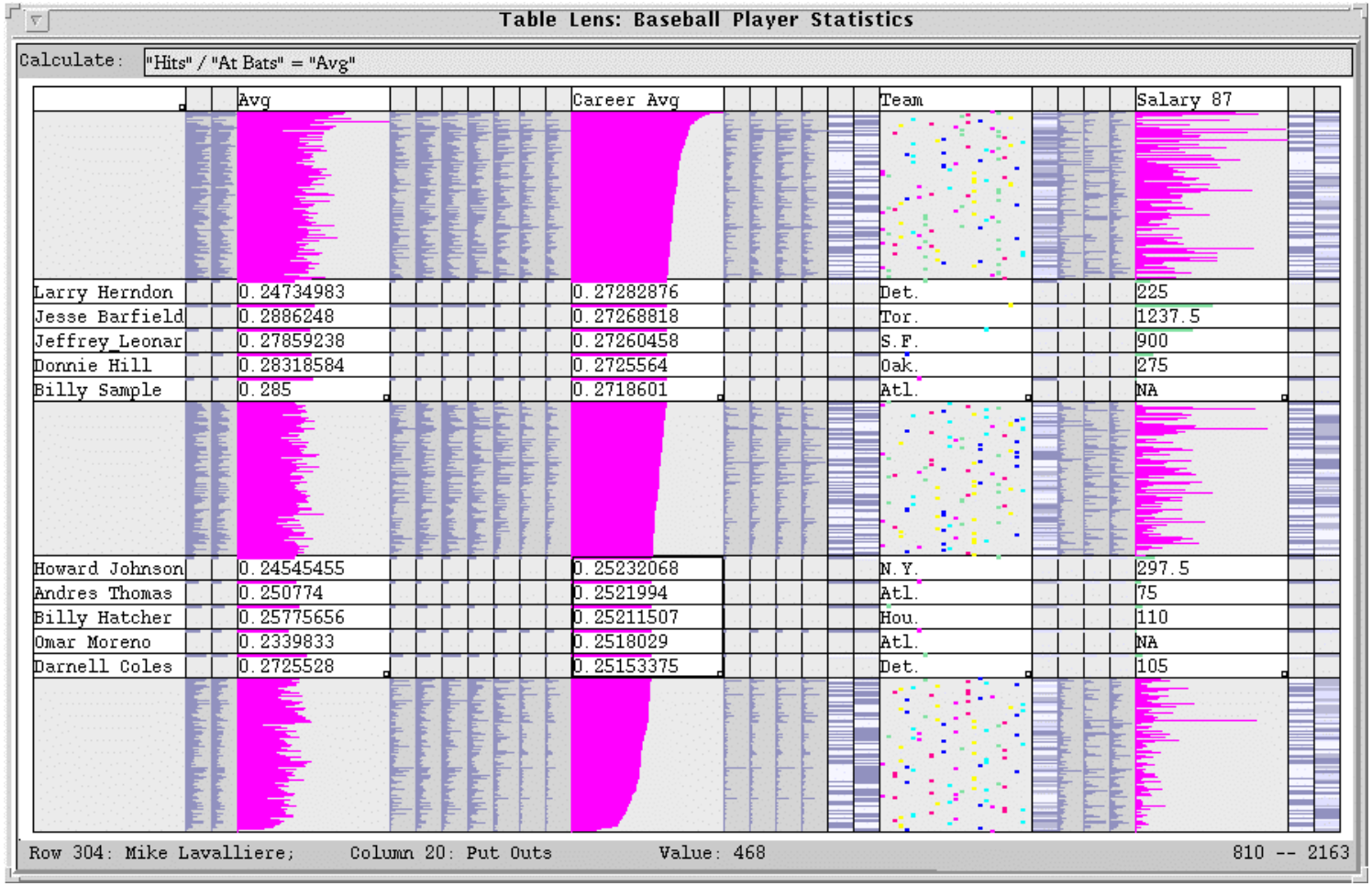


Table Lens

Rao & Card
CHI '94

Visualization's Value?

$$V_{alue} = T + I + E + C$$

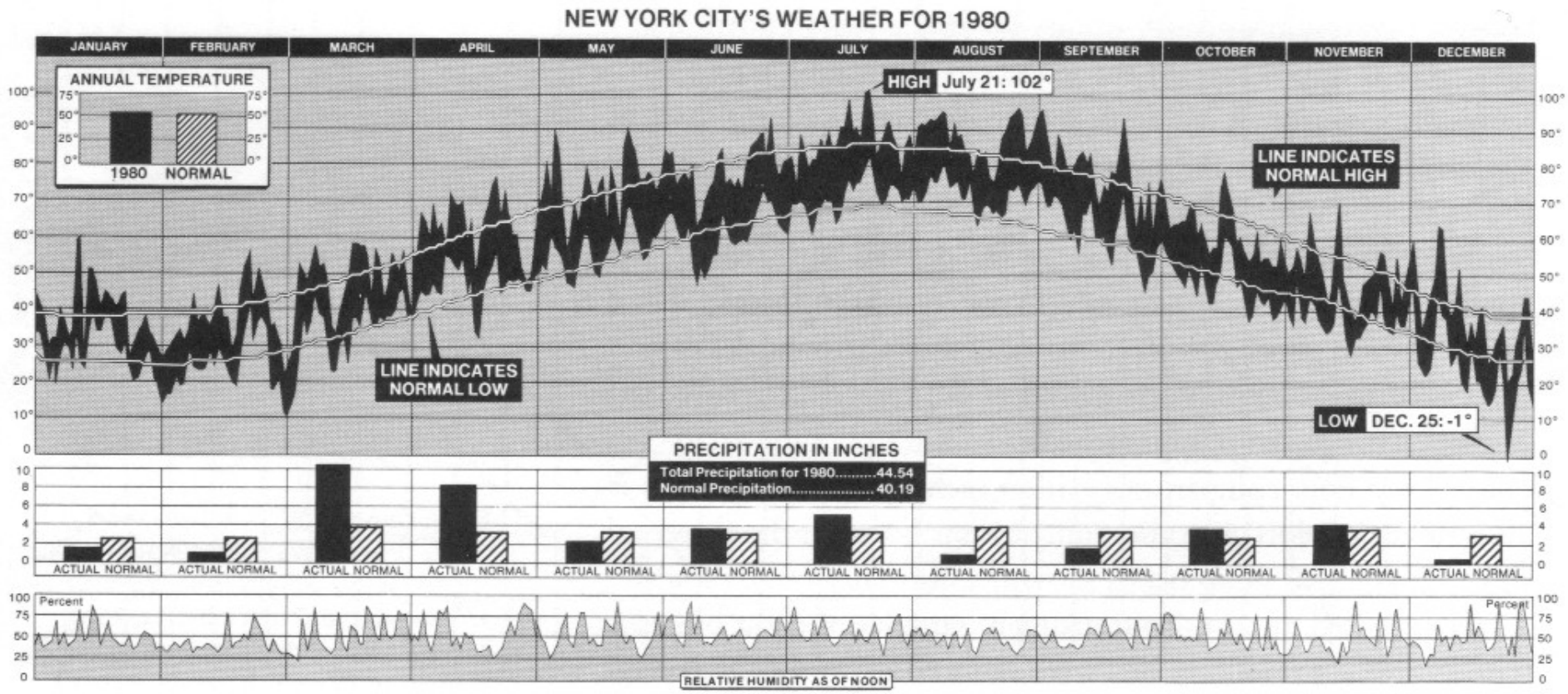
Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)

Some examples

1.

2220 numbers



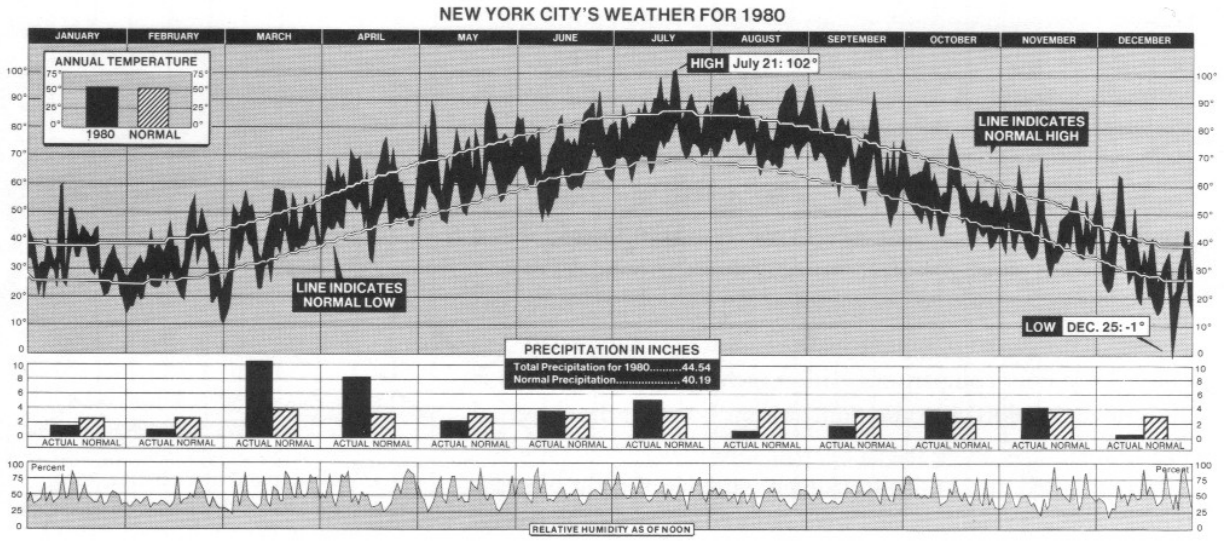
New York Times, January 11, 1981, p. 32.

E. Tufte
The Visual Display of Quantitative Information
1983

Data Values

365	High temp for each day
365	Low temp for each day
365	Avg high temp for each day
365	Avg low temp for each day
365	Precipitation for each day
365	Humidity for each day
12	Precipitation for each month
12	Avg precipitation for each month
1	Precipitation for the year
1	Avg precipitation per year
1	Highest temp (& day) for the year
1	Lowest temp (&day) for the year
1	Avg daily temp for the year
1	Avg daily temp per year

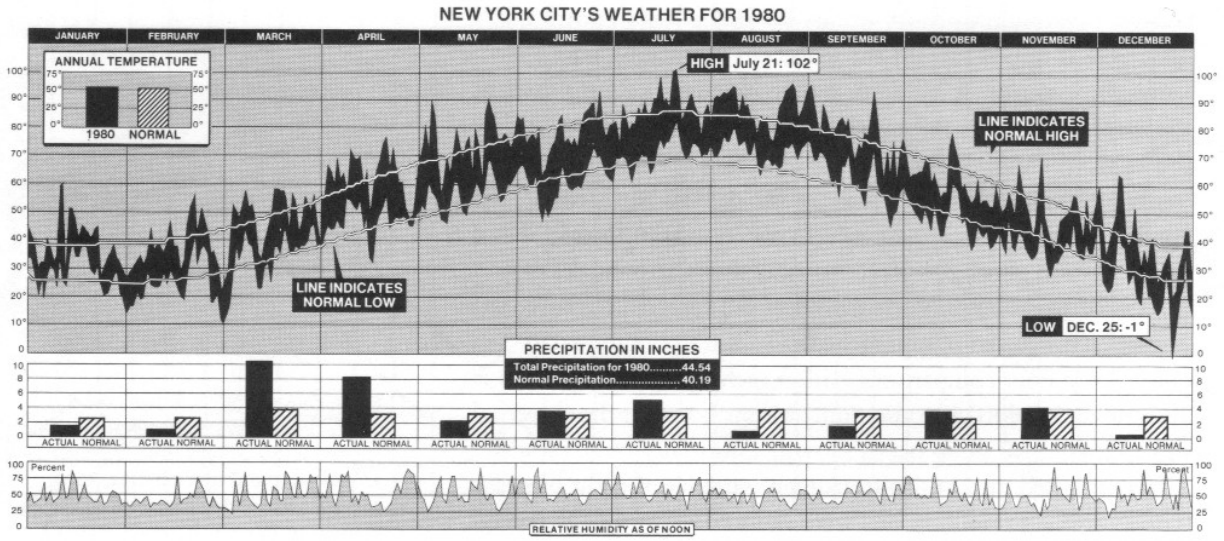
Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

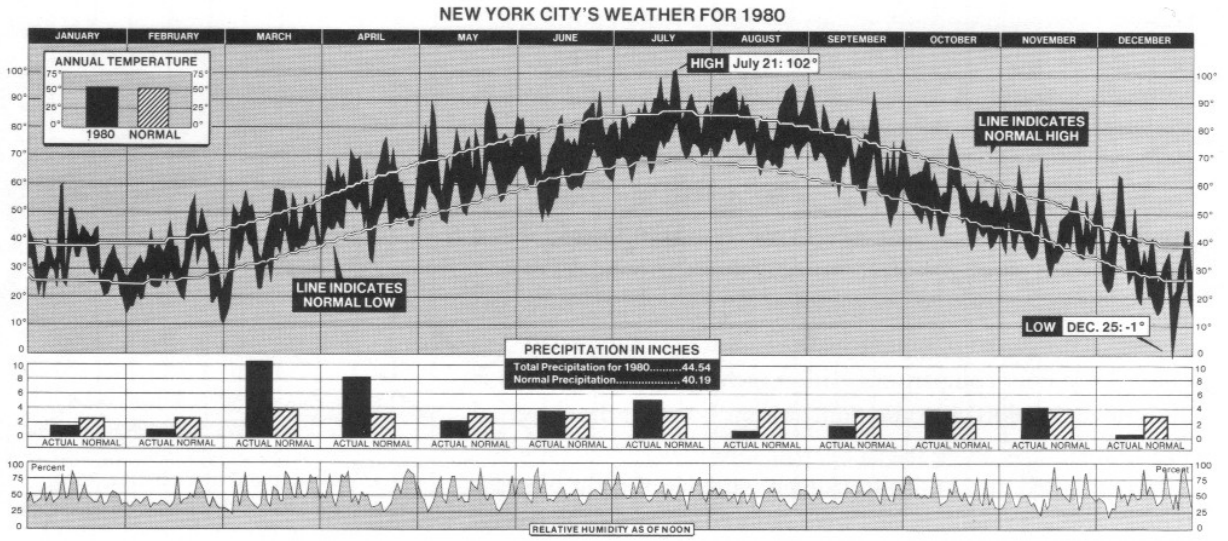
Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

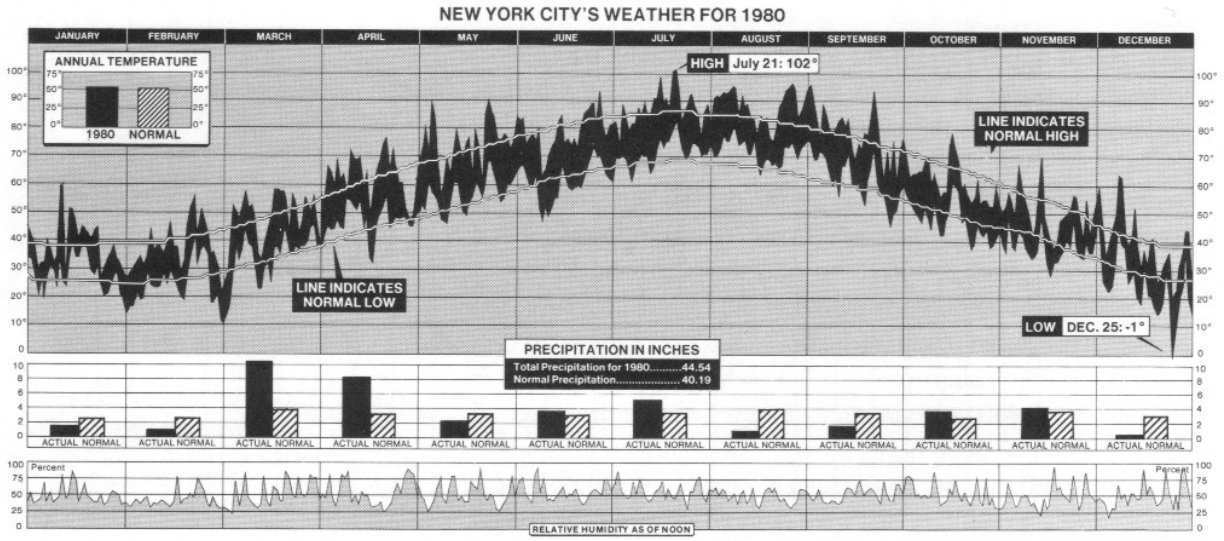
Visualization's Value



New York Times, January 11, 1981, p. 32.

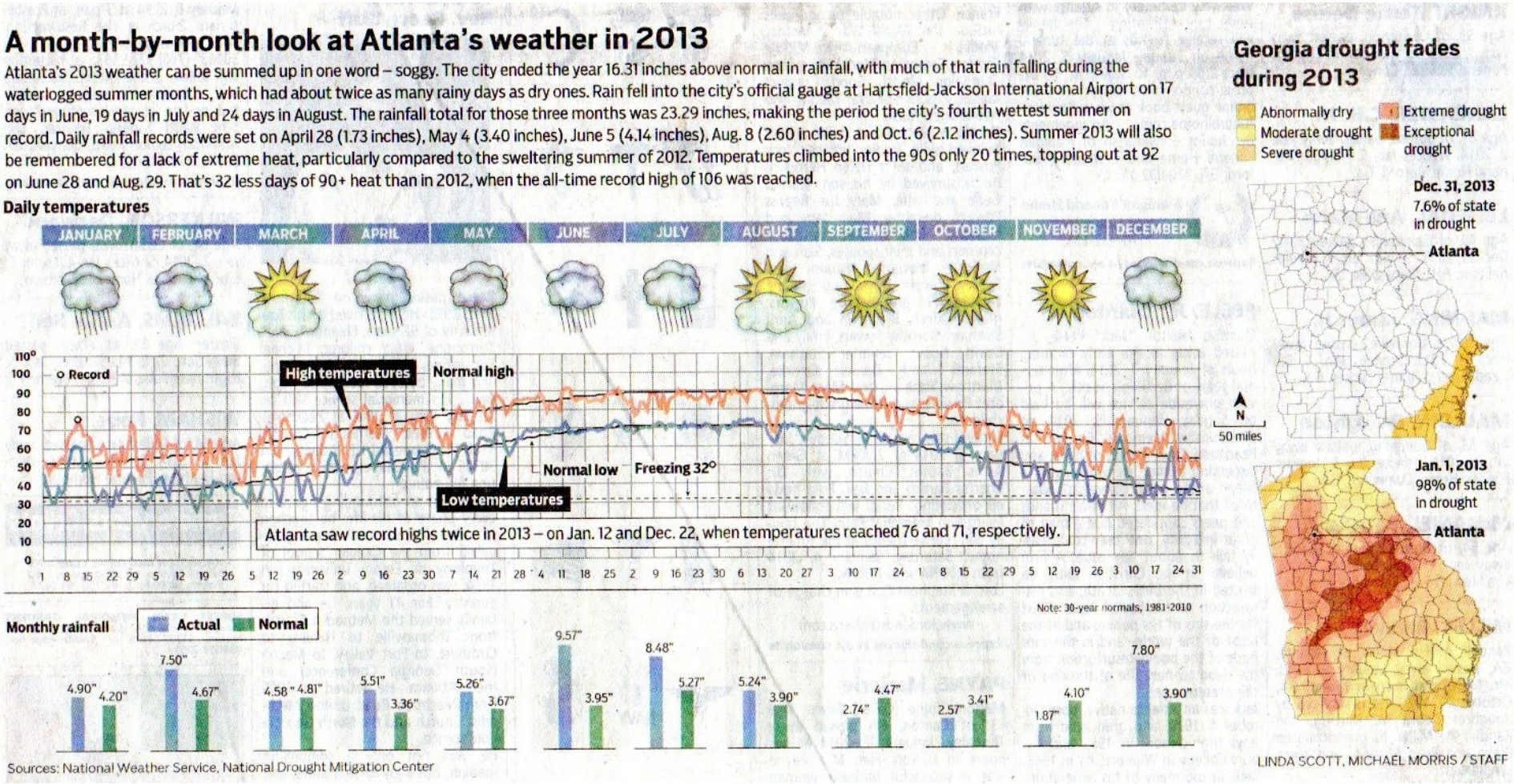
$$V_{alue} = T + I + E + C$$

Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

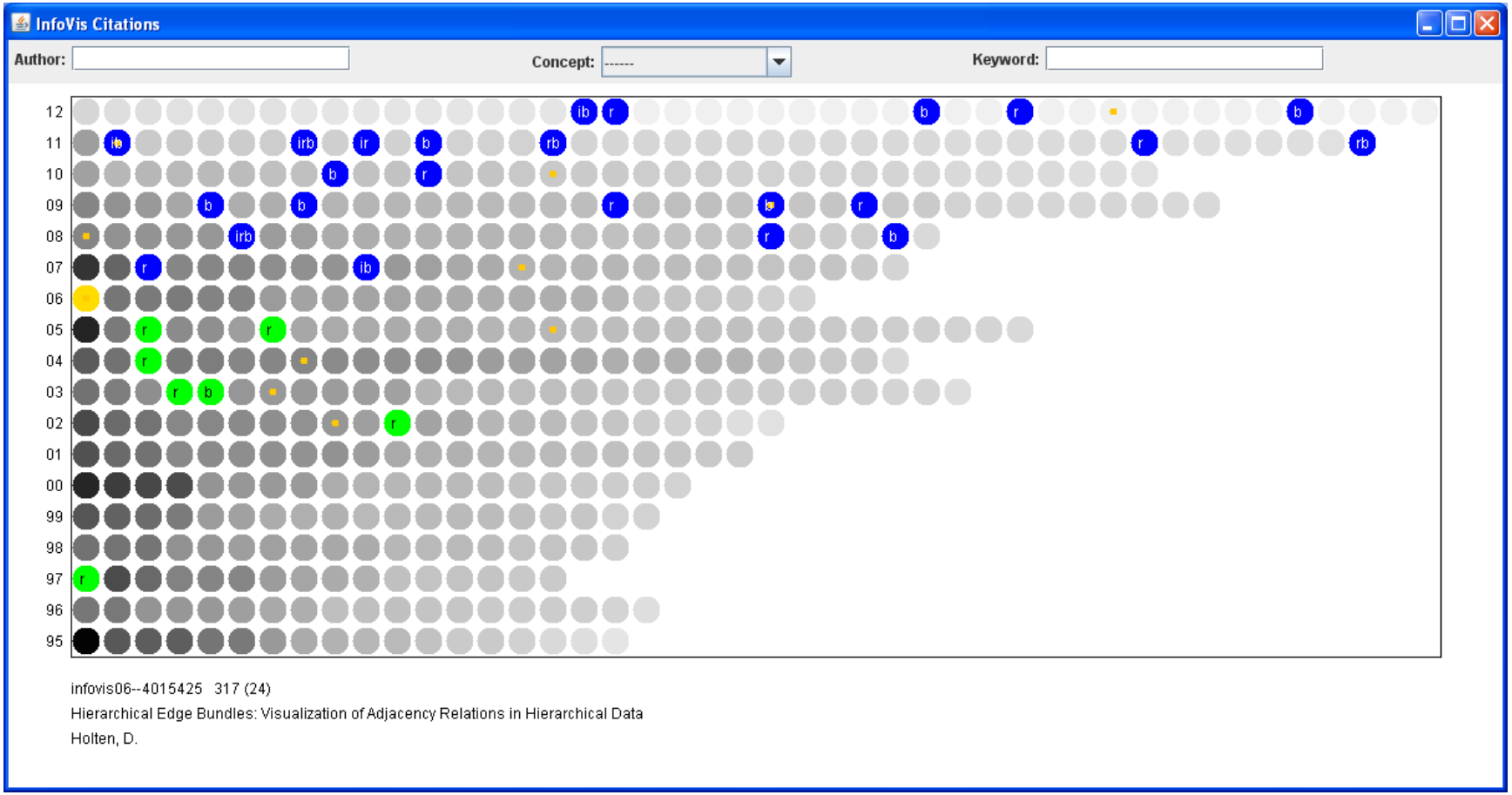


Atlanta Journal Constitution
Jan. 3, 2014

2.

http://www.cc.gatech.edu/gvu/ii/citevis

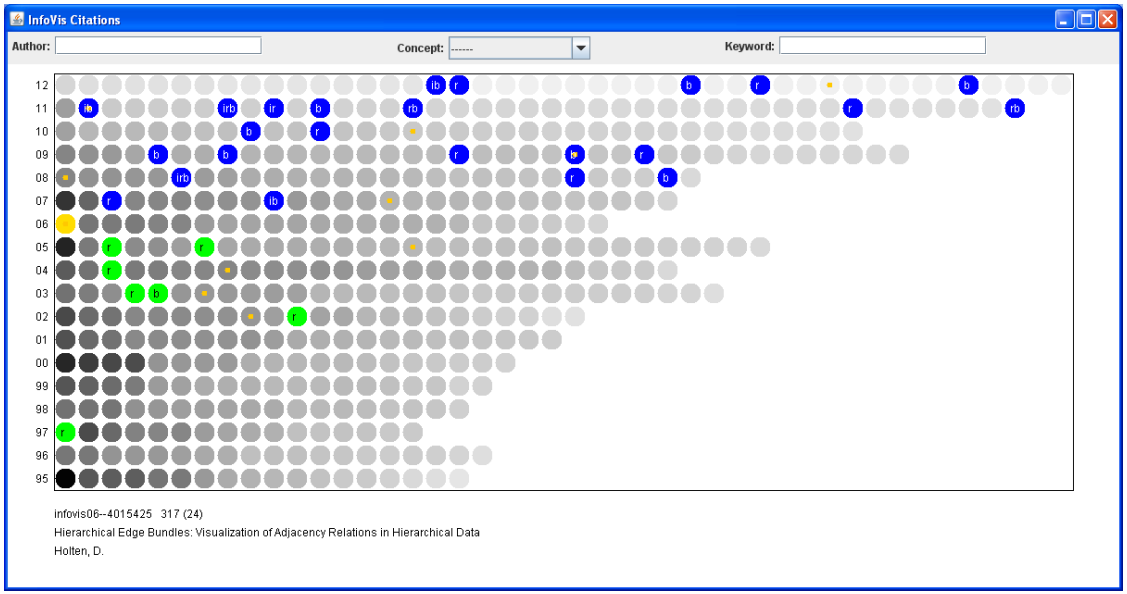
CiteVis



Demo

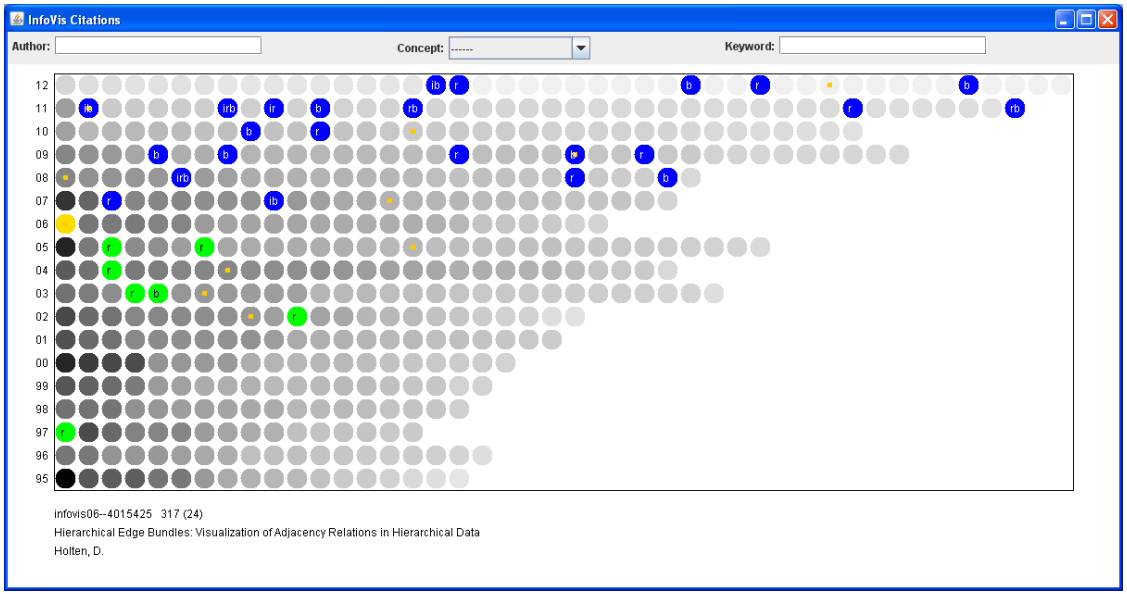
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

Visualization's Value



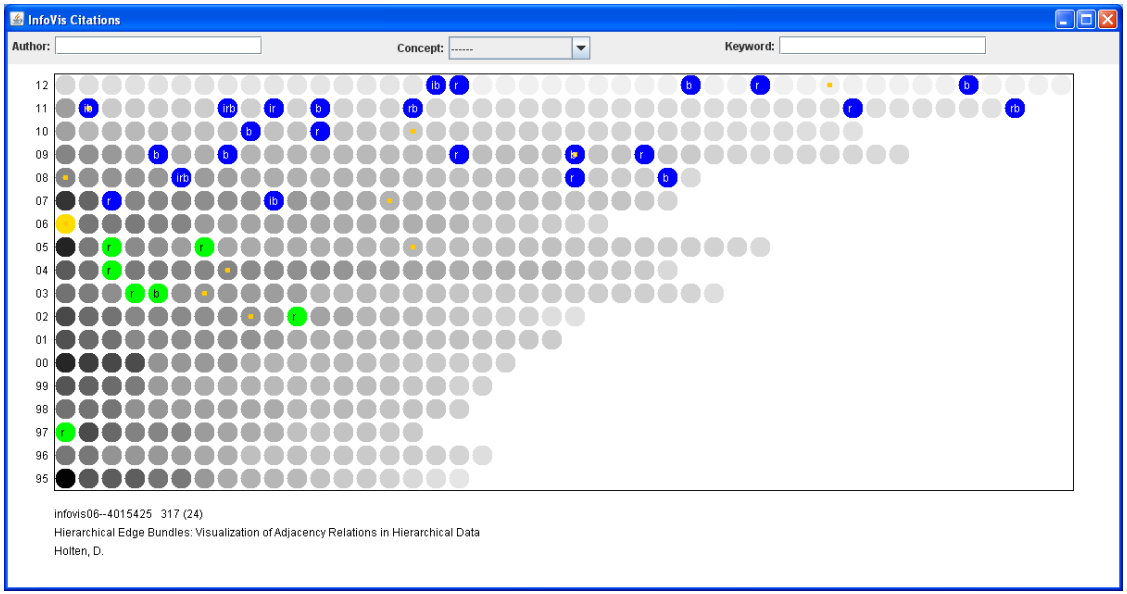
$$V_{alue} = T + I + E + C$$

Visualization's Value



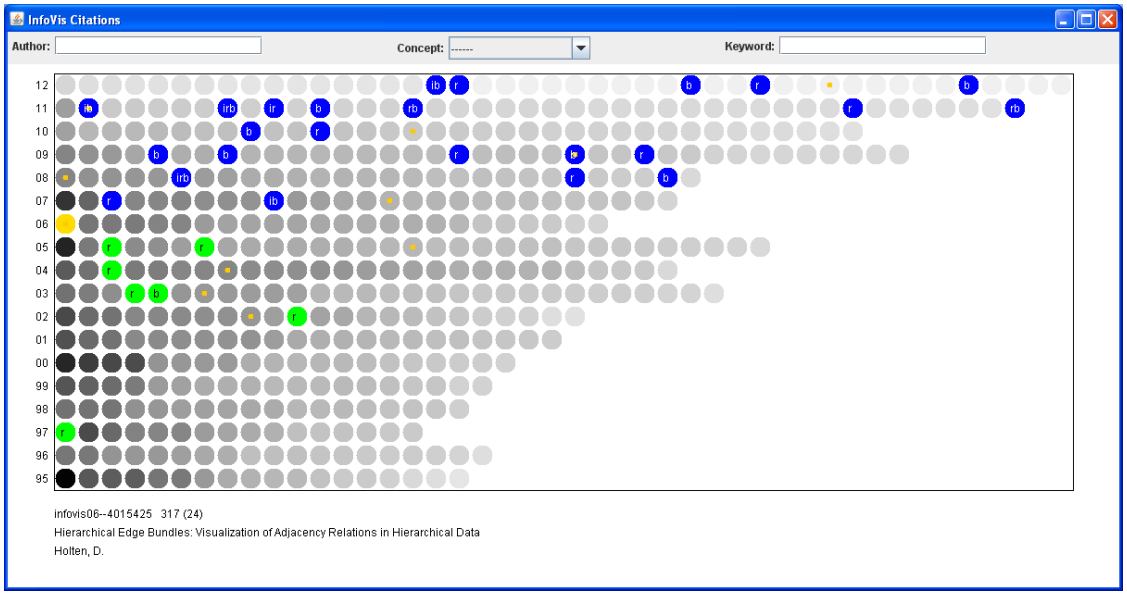
$$V_{alue} = T + I + E + C$$

Visualization's Value



$$V_{alue} = T + I + E + C$$

Visualization's Value



$$V_{alue} = T + I + E + C$$

Value

Not evaluation in traditional sense

User study with 12 grad students not so helpful

Takeaway 3

Understand & communicate the value

Key Open Problems in Visualization

Key Open Problems in Visualization

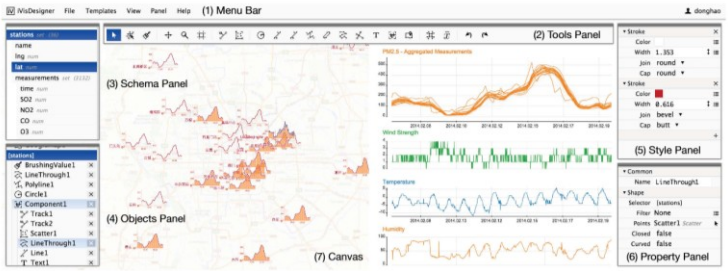
1. Assess and communicate **value**

$$V_{alue} = T + I + E + C$$

Key Open Problems in Visualization

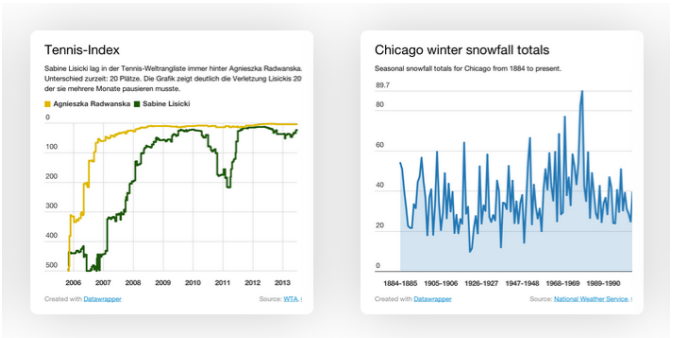
1. Assess and communicate value

2. Make the construction of visualizations easier



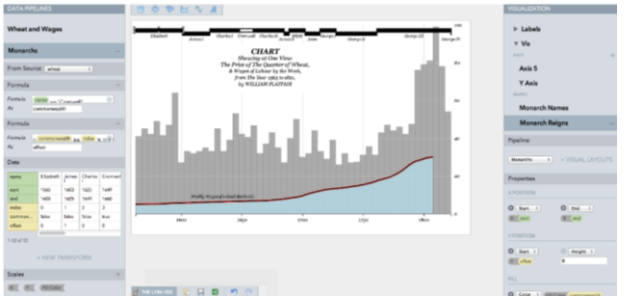
iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) '14

DataWrapper
https://datawrapper.de/





SketchStory
Lee, Kazi & Smith
TVCG (InfoVis) '13



Lyra
Satyanarayan & Heer
EuroVis '14

Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier
3. Address real world “big” problems









Take Aways

Understand the purpose

Presentation & analysis, related but different

Interaction provides power, use it

Explain the value (T + I + E + C)

Thanks for your attention

Acknowledgments

- Supported by DARPA's XDATA program

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)





- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



August 28, 2014

The Value of Visualization for Exploring and Understanding Data

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu

Boeing Data Analytics Community of Excellence



information interfaces

Georgia Tech College of Computing
School of Interactive Computing


Data

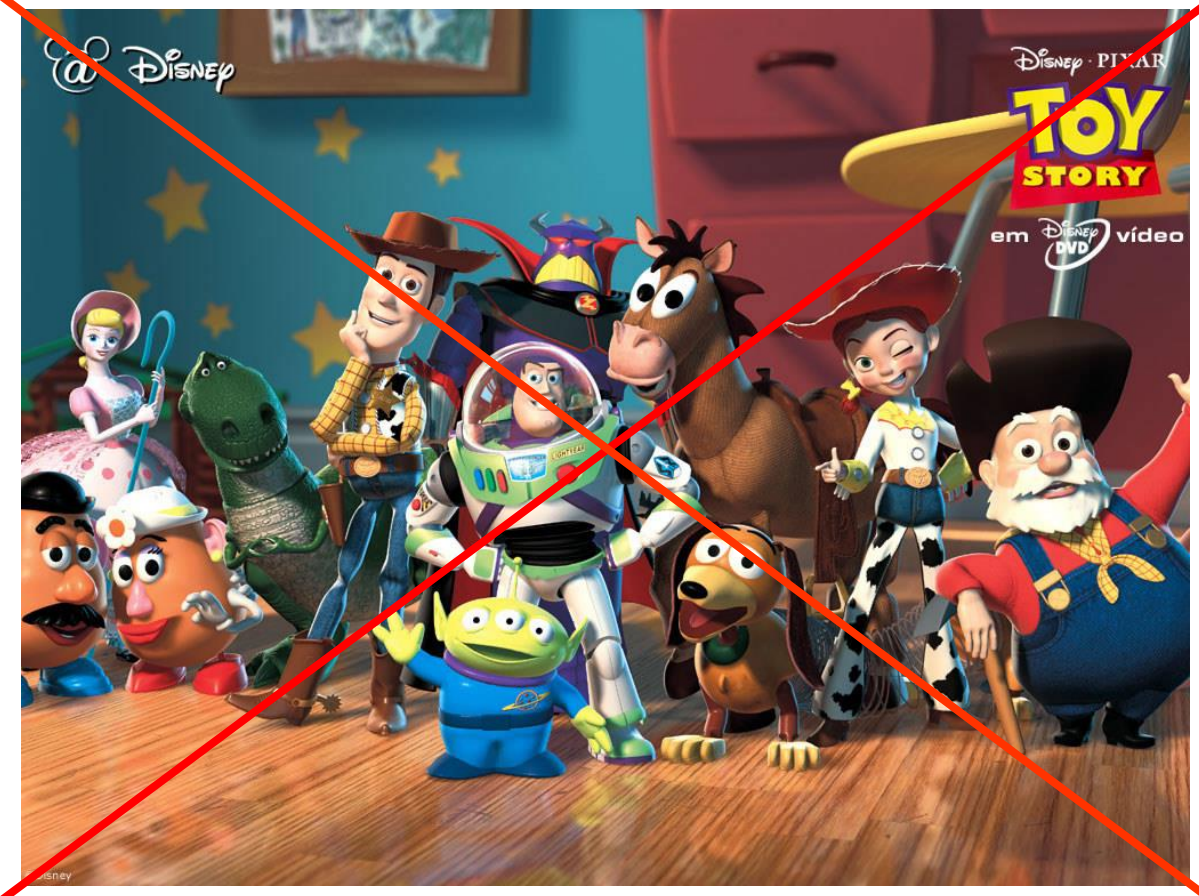
Data Visualization

Making pretty pictures?



Data Visualization

~~Making pretty pictures~~



Data Visualization

A cognitive process

Gain an understanding



Data Visualization

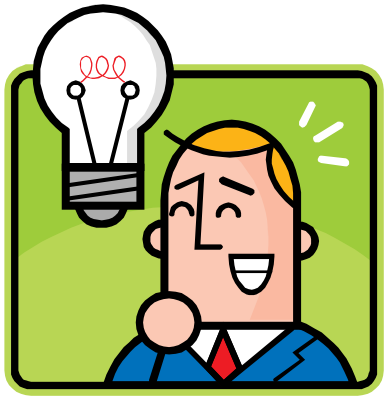
A cognitive process

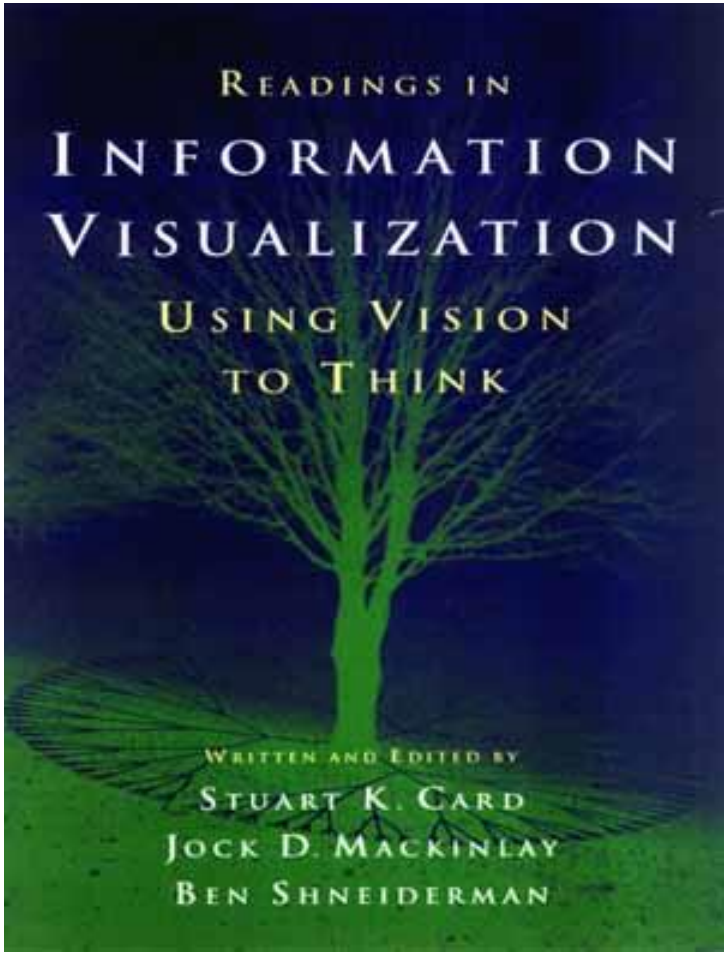
Gain an **understanding**



Challenge

Transform the *data* into *information* (understanding, insight) thus making it useful to people





Visualization

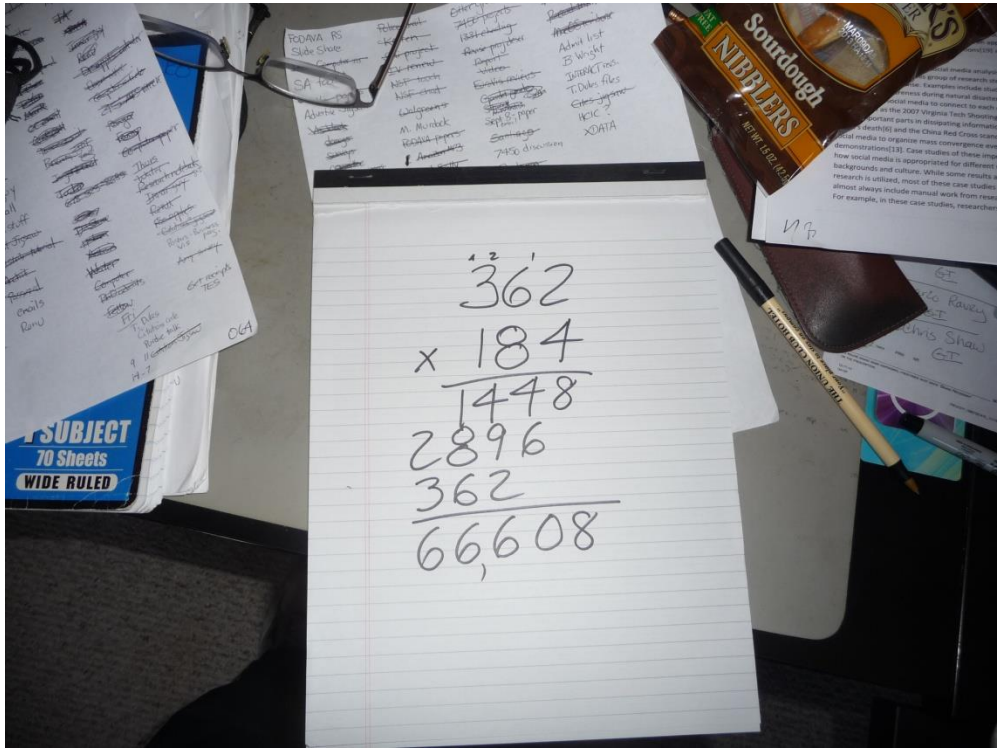
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provides a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Let Me Illustrate

An example…





Questions:

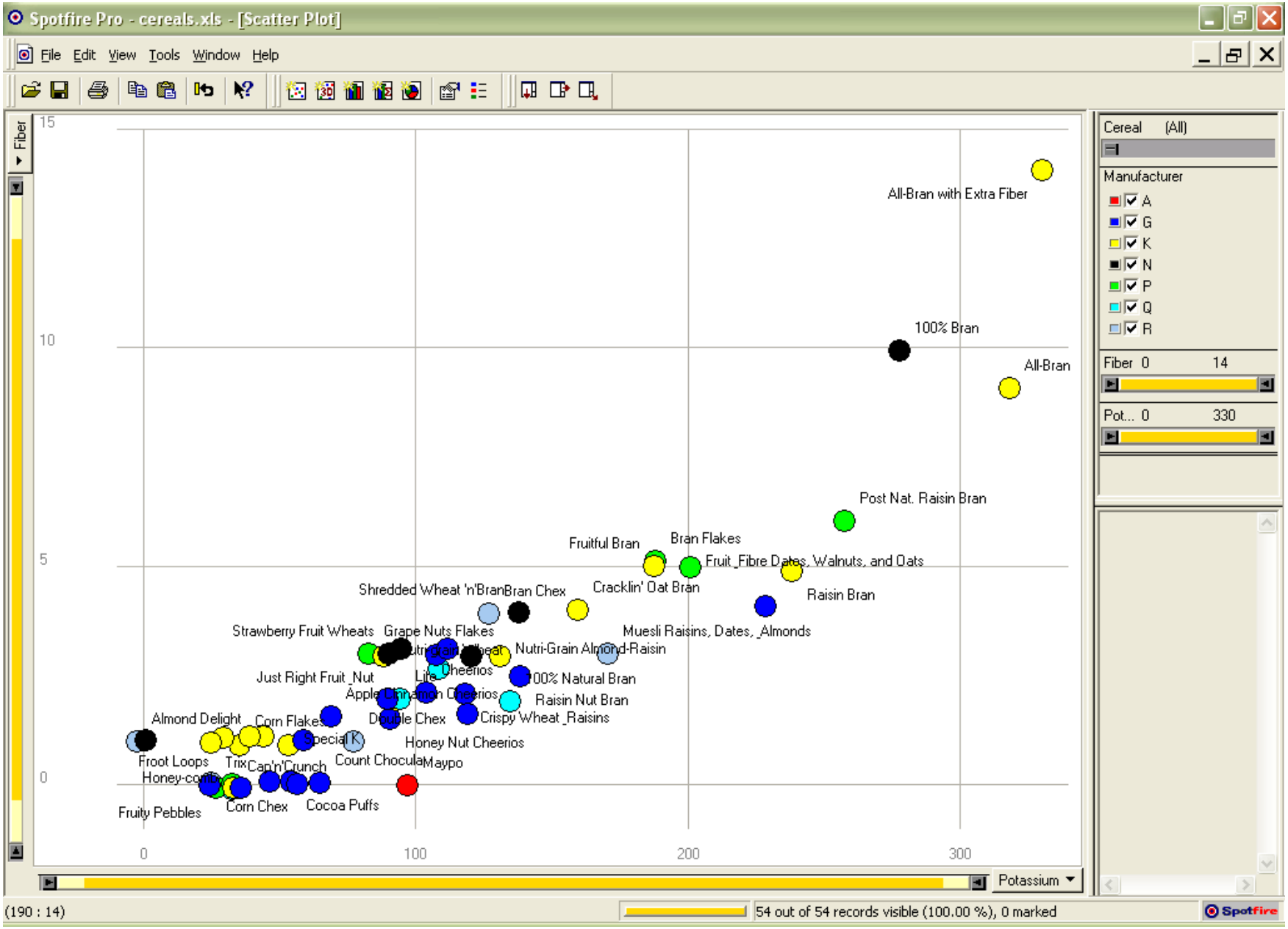
Which cereal has the most/least potassium?

Is there a relationship between potassium and fiber?

If so, are there any outliers?

Which manufacturer makes the healthiest cereals?

	A	B	C	D
1	Cereal	Manufacturer	Fiber	Potassium
2	100% Bran	N	10	280
3	100% Natural Bran	Q	2	135
4	All-Bran	K	9	320
5	All-Bran with Extra Fiber	K	14	330
6	Almond Delight	R	1	0
7	Apple Cinnamon Cheeric	G	1.5	70
8	Bran Chex	R	4	125
9	Bran Flakes	P	5	190
10	Cap'n'Crunch	Q	0	35
11	Cheerios	G	2	105
12	Cocoa Puffs	G	0	55
13	Corn Chex	R	0	25
14	Corn Flakes	K	1	35
15	Count Chocula	G	0	65
16	Cracklin' Oat Bran	K	4	160
17	Cream of Wheat (Quick)	N	1	0
18	Crispy Wheat & Raisins	G	2	120
19	Double Chex	R	1	80
20	Froot Loops	K	1	30
21	Frosted Flakes	K	1	25
22	Fruit & Fibre Dates, Wal	P	5	200
23	Fruitful Bran	K	5	190
24	Fruity Pebbles	P	0	25
25	Golden Grahams	G	0	45
26	Grape Nuts Flakes	P	3	85
27	Honey Nut Cheerios	G	1.5	90
28	Honey-comb	P	0	35
29	Just Right Fruit & Nut	K	2	95
30	Life	Q	2	95
31	Lucky Charms	G	0	55
32	Maypo	A	0	95
33	Muesli Raisins, Dates, &	R	3	170
34	Multi-Grain Cheerios	G	2	90
35	Nutri-Grain Almond-Rais	K	3	130
36	Nutri-grain Wheat	K	3	90
37	Oatmeal Raisin Crisp	G	1.5	120
38	Post Nat. Raisin Bran	P	6	260
39	Product 19	K	1	45
40	Quaker Oatmeal	Q	2.7	110
41	Raisin Bran	K	5	240
42	Raisin Nut Bran	G	2.5	140
43	Rice Krispies	K	0	35
44	Shredded Wheat	N	3	95
45	Shredded Wheat 'n'Bran	N	4	140
46	Shredded Wheat spoon	N	3	120
47	Smacks	K	1	40
48	Special K	K	1	55
49	Strawberry Fruit Wheats	N	3	90
50	Total Corn Flakes	G	0	35
51	Total Raisin Bran	G	4	230
52	Total Whole Grain	G	3	110
53	Trix	G	0	25
54	Wheaties	G	3	110
55	Wheaties Honey Gold	G	1	60

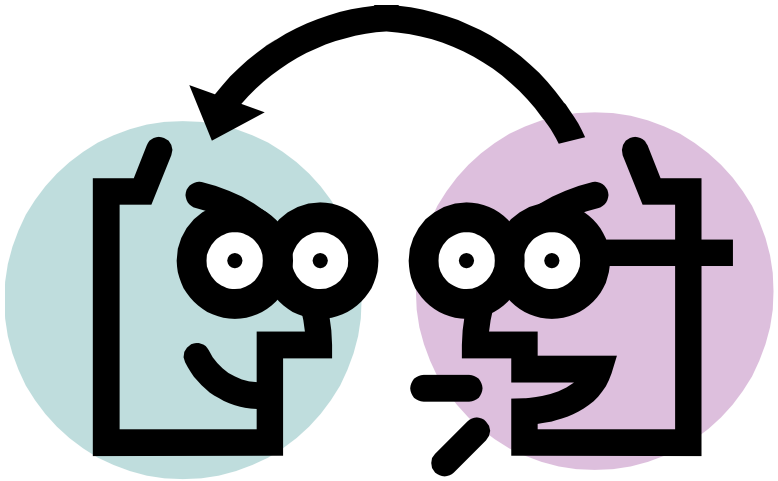


Fiber

Potassium

Even Tougher?

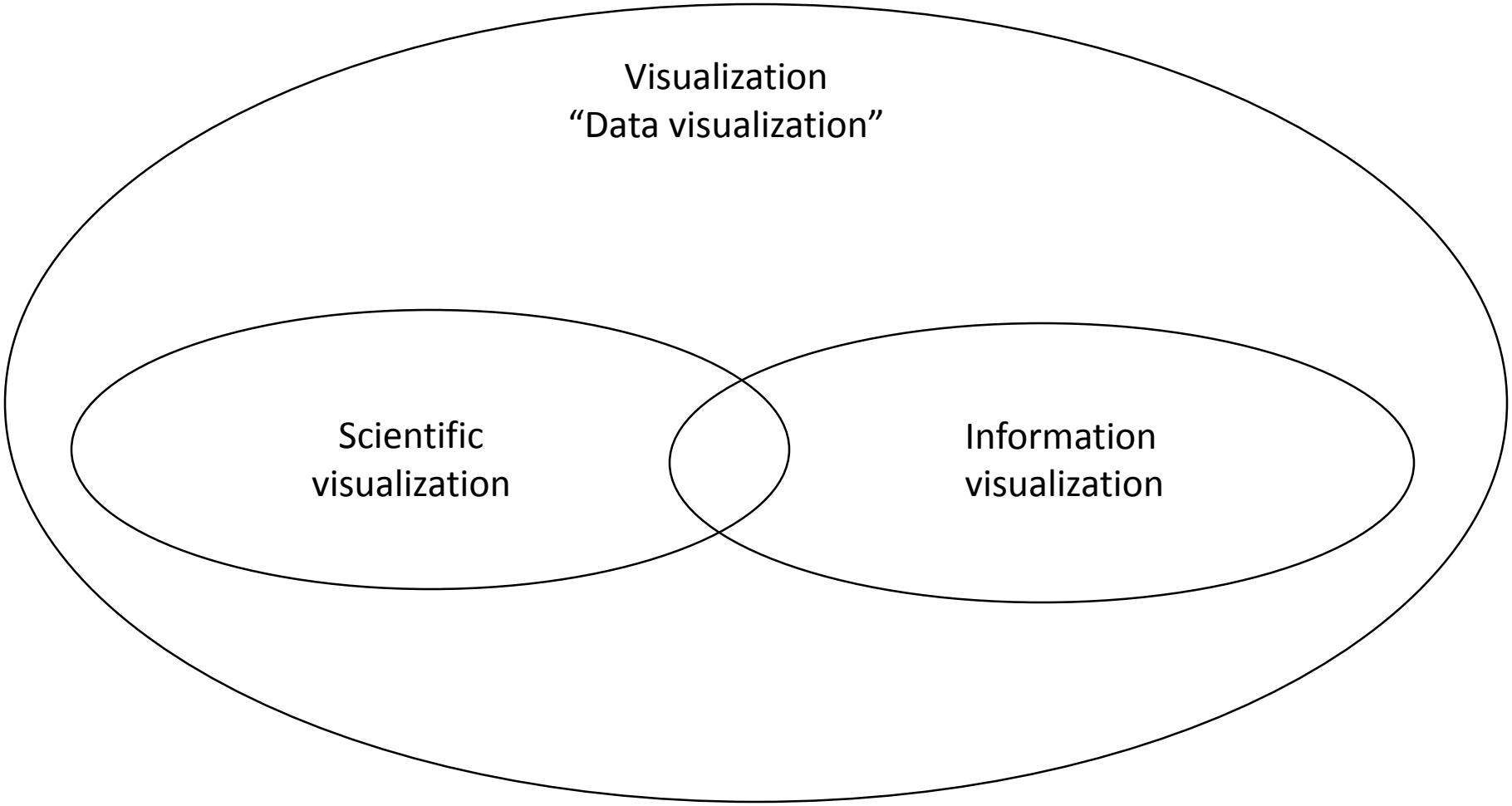
What if you could only see one cereal's data at a time?
(e.g. some websites)
What if I read the data to you?



Part of our Culture

“I see what you’re saying”
“Seeing is believing”
“A picture is worth a thousand words”

Academic Communities



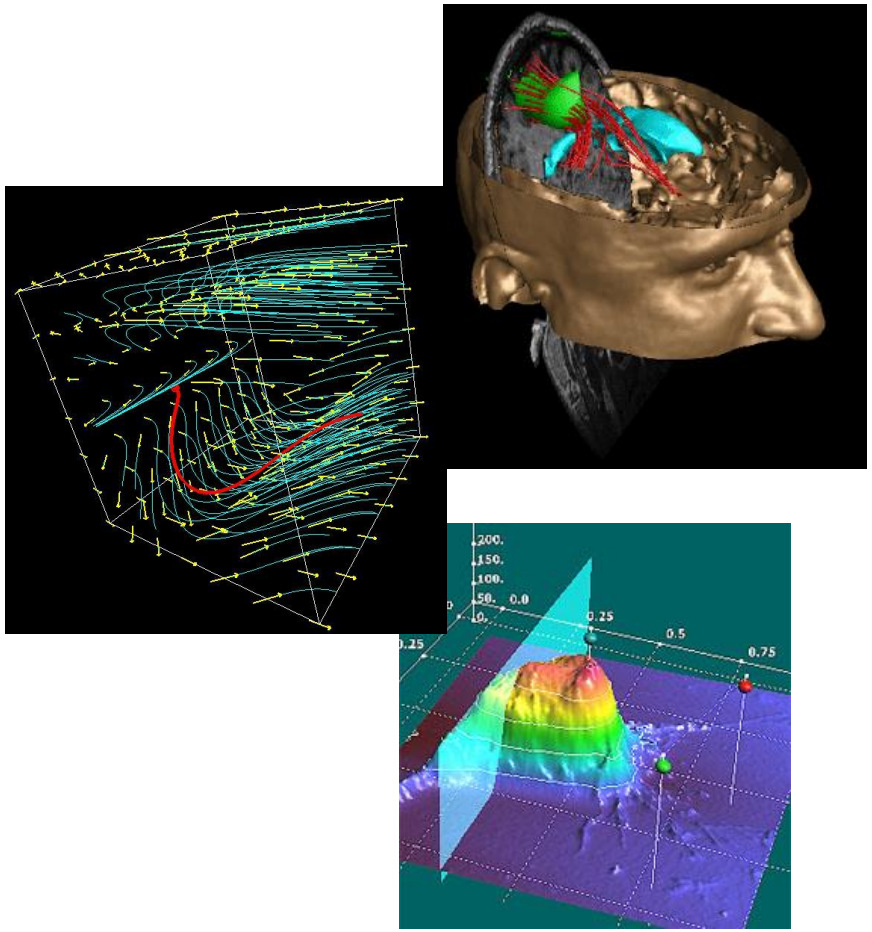
Scientific Visualization (SciVis)

Primarily relates to and represents something spatial, physical or geometric

Often 3-D

Examples

- Air flow over a wing
- Stresses on a girder
- Torrents inside a tornado
- Organs in the human body
- Molecular bonding



Not my research focus

Information Visualization (InfoVis)

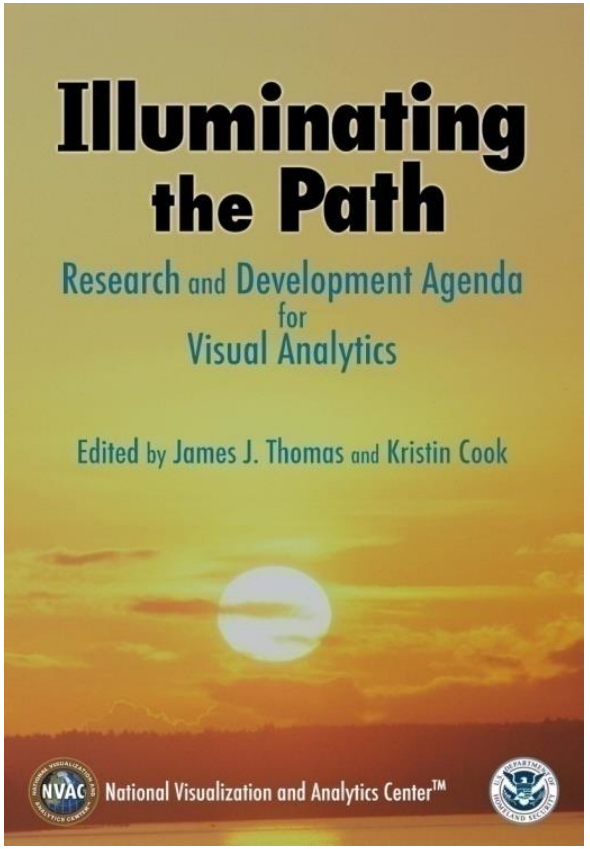
Visualizing *non-spatial* data: Items, entities, things which do not have a direct physical correspondence

- Notion of *abstractness* of the entities is important too
- Examples: baseball statistics, stock trends, connections between criminals, car attributes...

> "A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art."
>
> http://conferences.computer.org/infovis/

Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces



Available at http://nvac.pnl.gov/ in PDF form

Applications of Visualization

Presentation

Analysis

Explanatory

Exploratory

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support



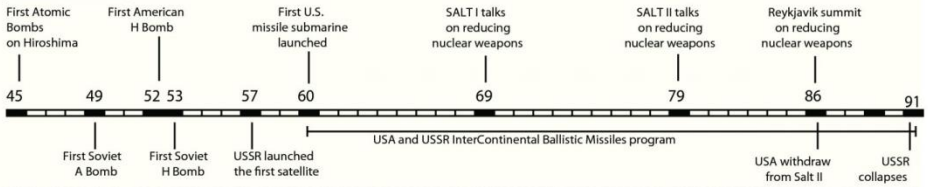
Infographics

All the rage…

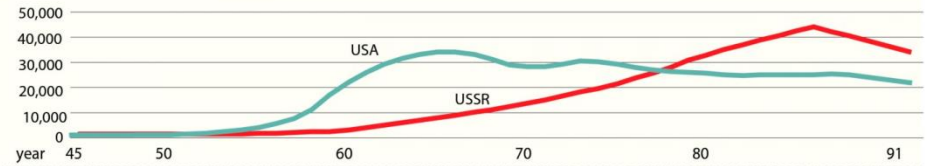
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



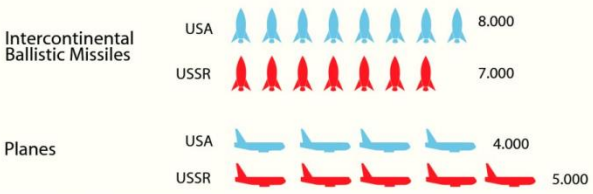
NUCLEAR STOCKPILE



ROCKET MODELS



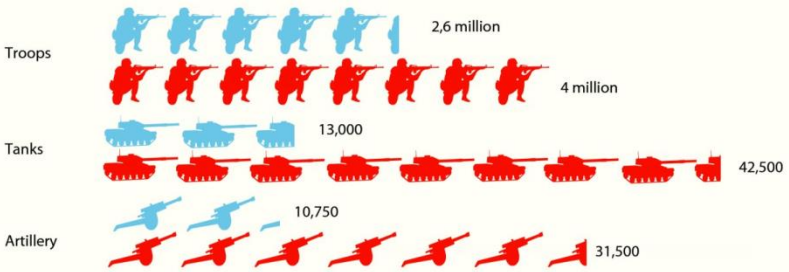
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



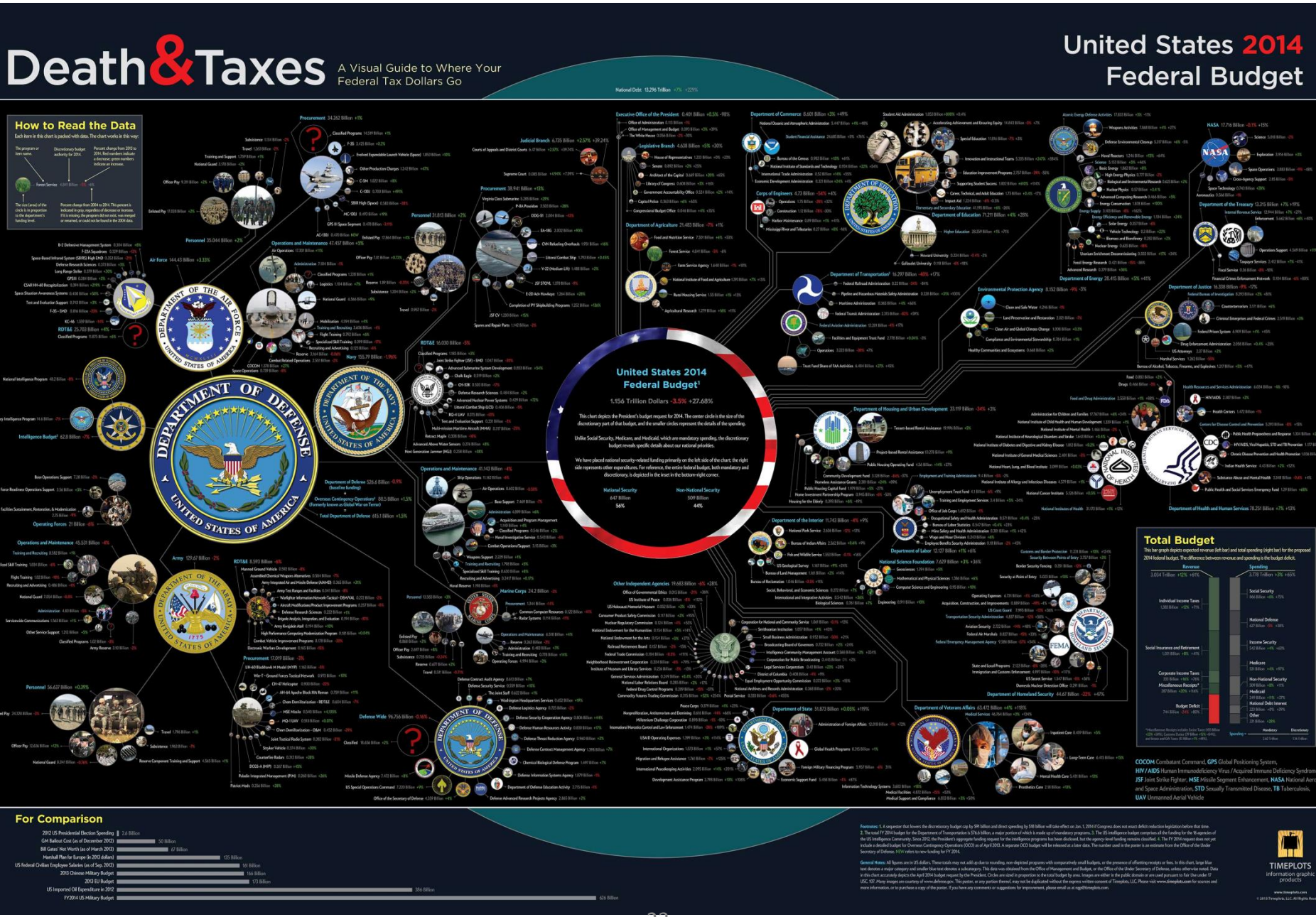


Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget



NY Times 2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons
Bonds was injured last season. He played 14 games and hit 5 homers.

600
400
200
0
14th season
5 seasons
10
15

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

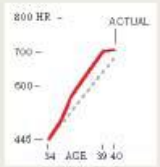
Aaron
Actual homers slightly outpace projected homers for five seasons.



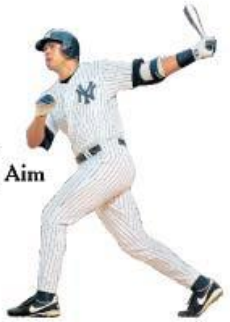
Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.



Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

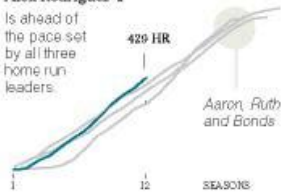


Note: Ages as of July 1 of each season.

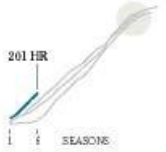


Others Taking Aim

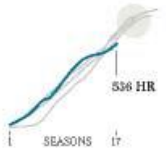
Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

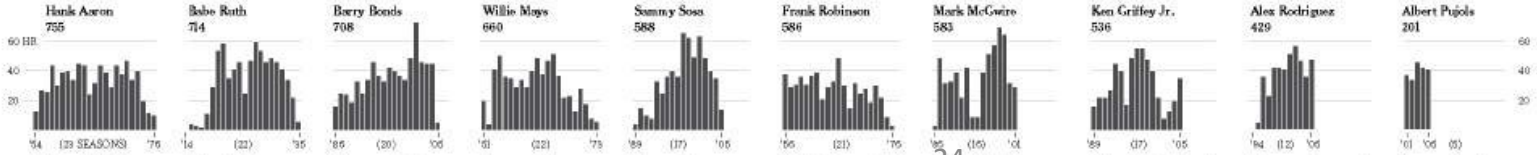


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

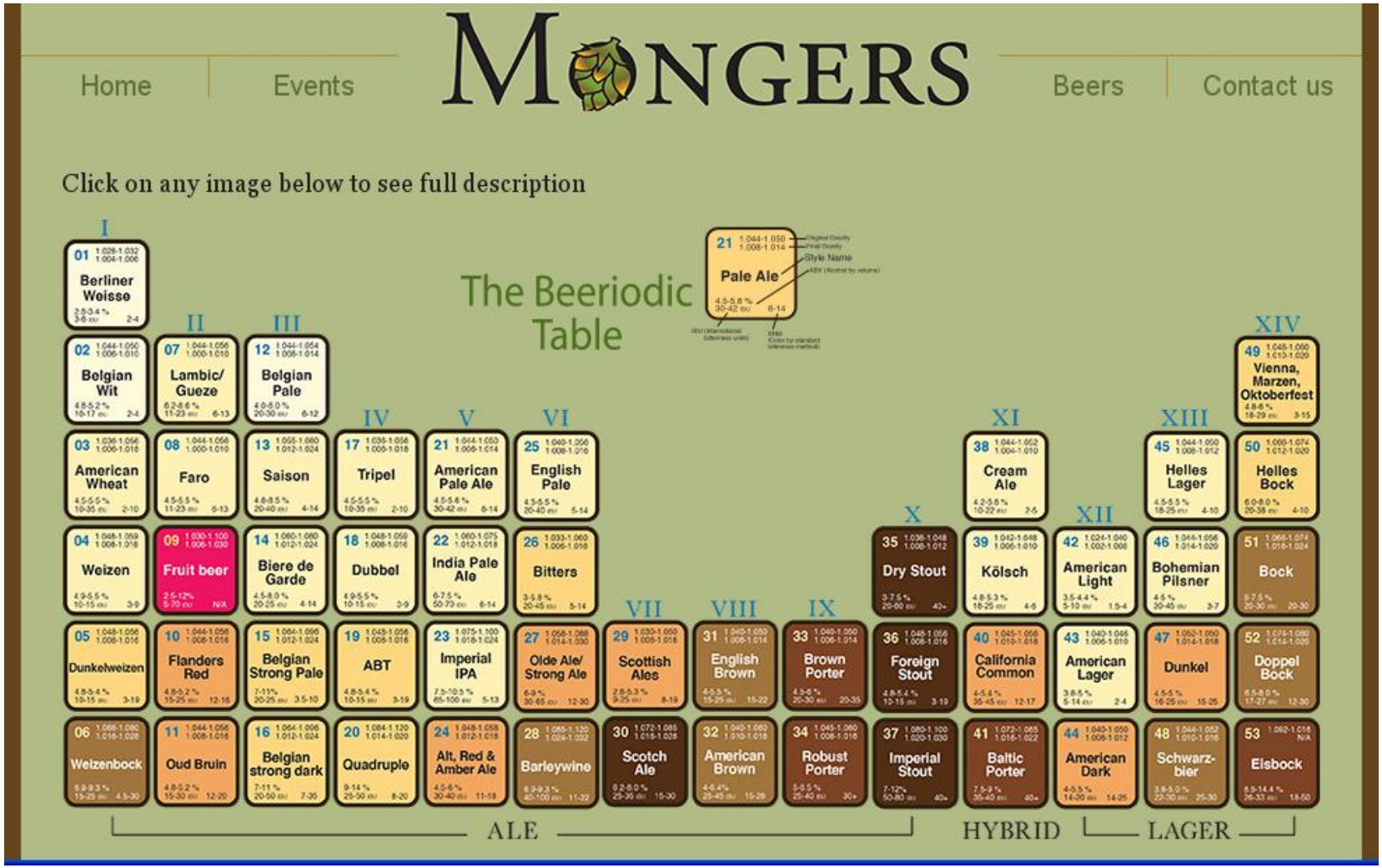


Differing Paths to the Top of the Charts

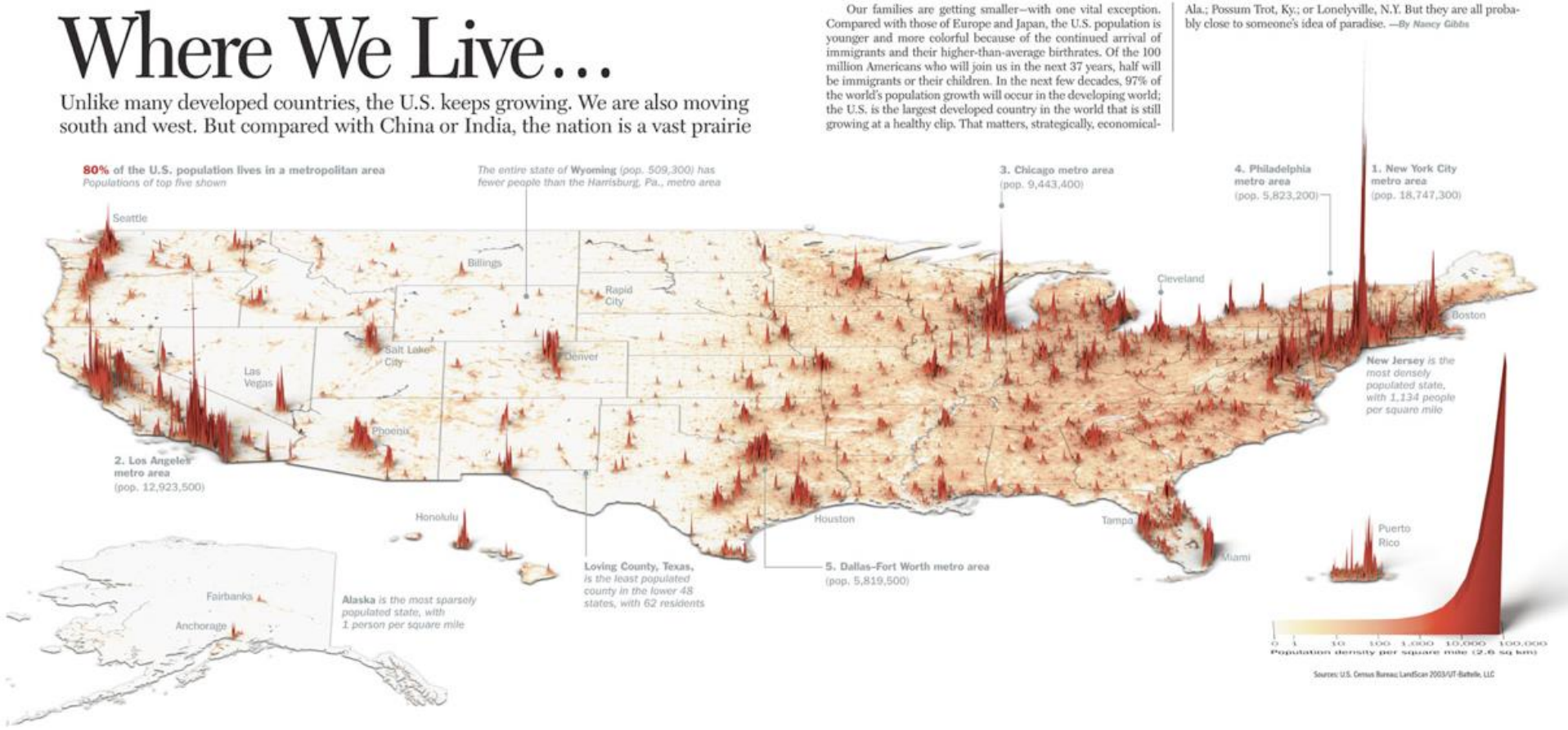
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



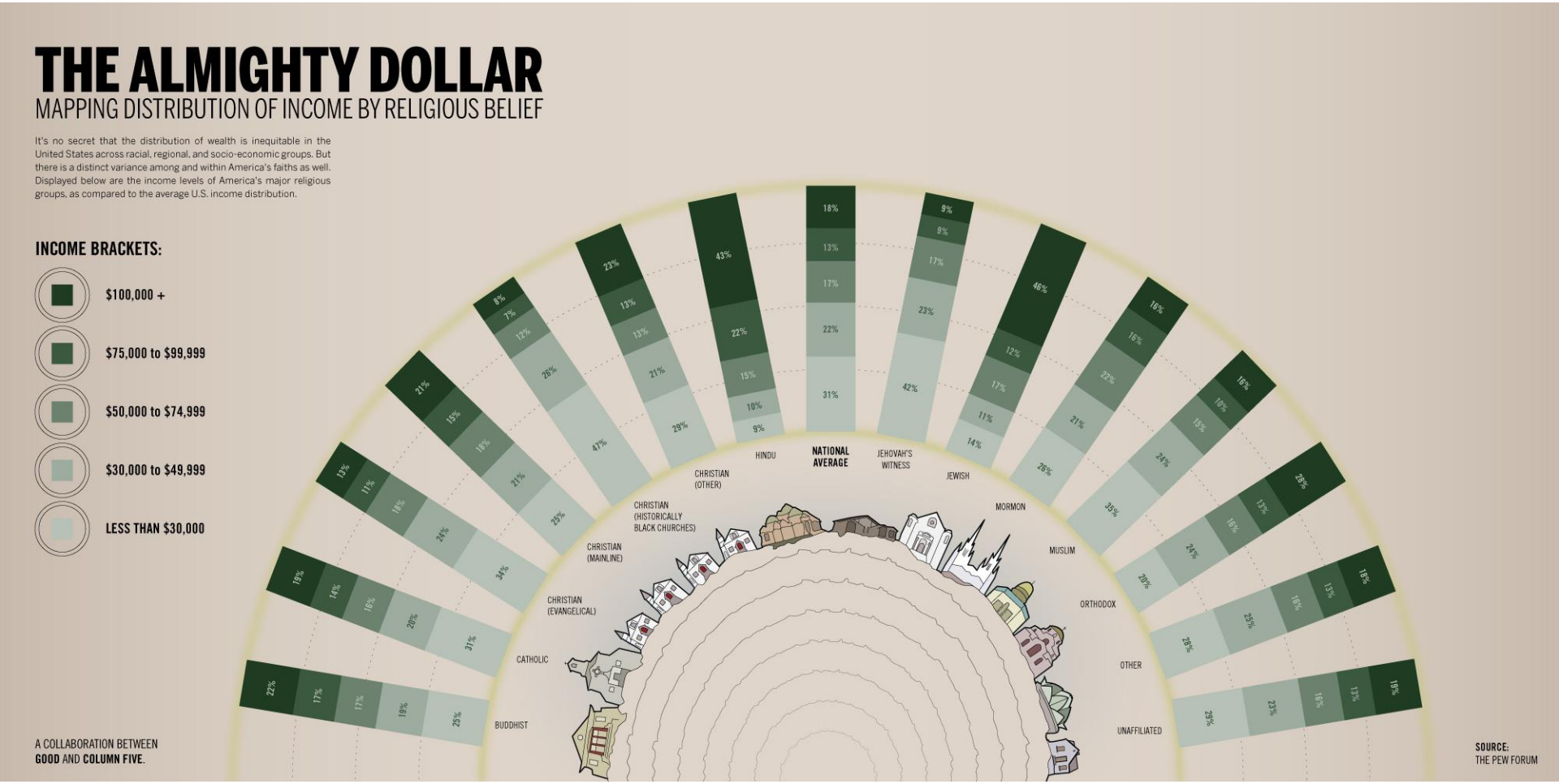
Amanda Cox and Joe Ward/The New York Times



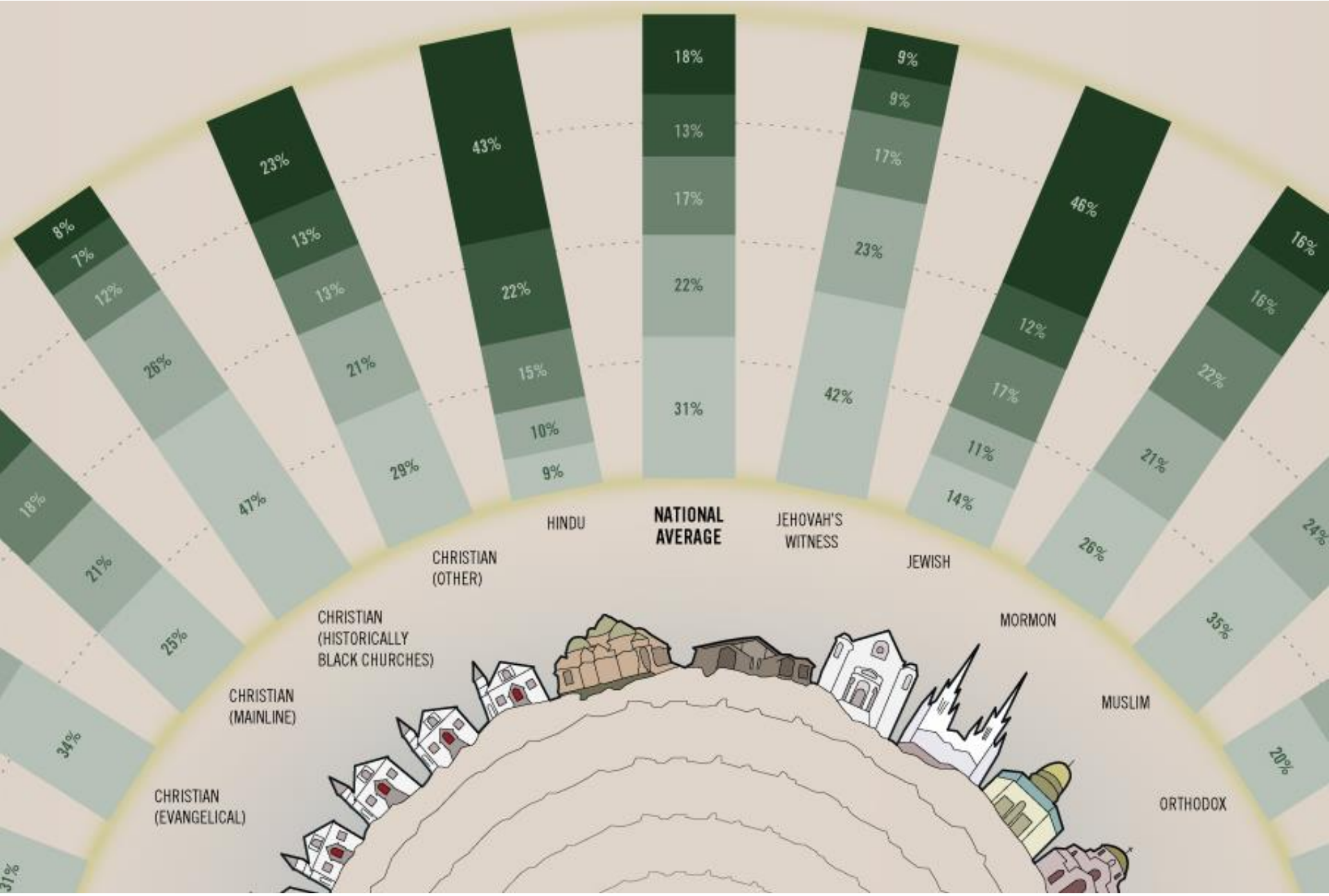
http://thebeermongers.com/beers/



http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html



http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg


8%
7%
12%
26%
41%
CHRISTIAN (EVANGELICAL)
18%
21%
25%
34%
23%
13%
13%
21%
29%
CHRISTIAN (MAINLINE)
43%
22%
15%
10%
9%
CHRISTIAN (HISTORICALLY BLACK CHURCHES)
18%
13%
17%
22%
31%
NATIONAL AVERAGE
9%
9%
17%
23%
42%
HINDU
46%
12%
17%
11%
14%
CHRISTIAN (OTHER)
16%
16%
22%
21%
26%
JEWISH
24%
35%
20%
ORTHODOX
MORMON
MUSLIM
JEHOVAH'S WITNESS

Frequent presentation goals

Clarify

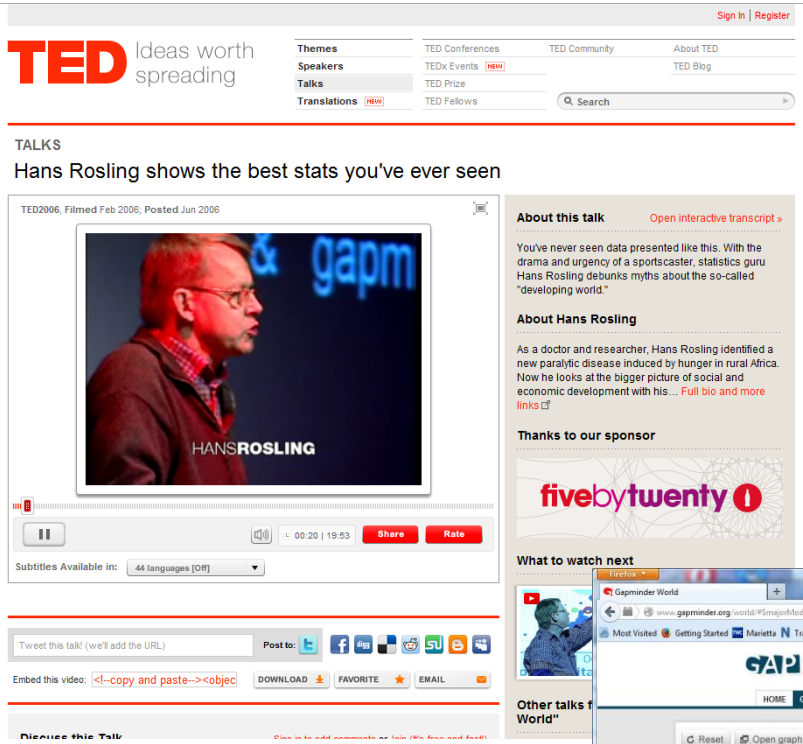
Focus

Highlight

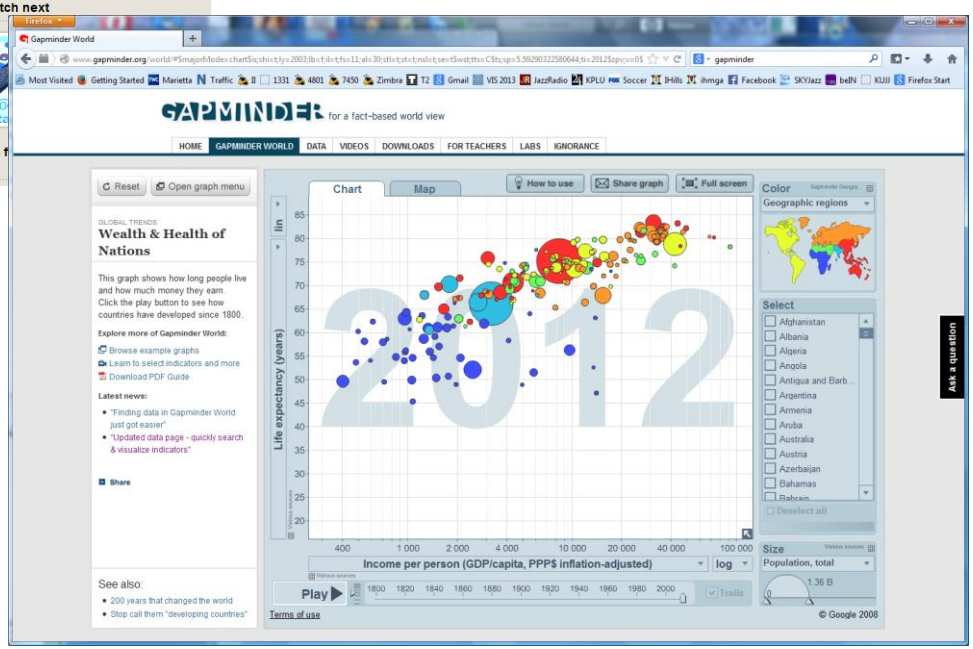
Simplify

Persuade

May just show a few variables
and/or a subset of the data cases



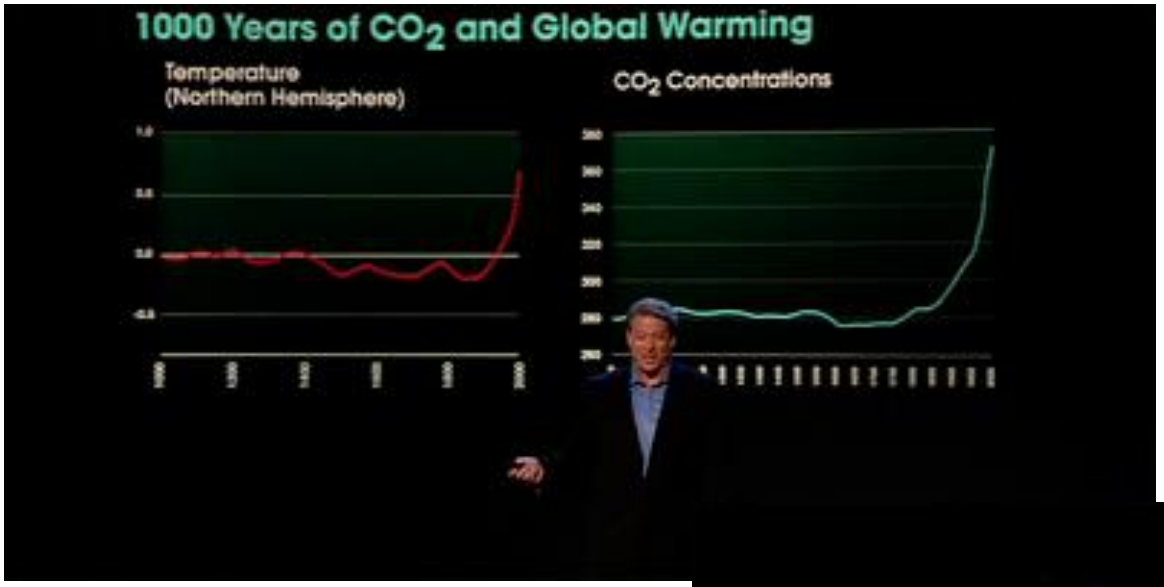
Hans Rosling
Gapminder

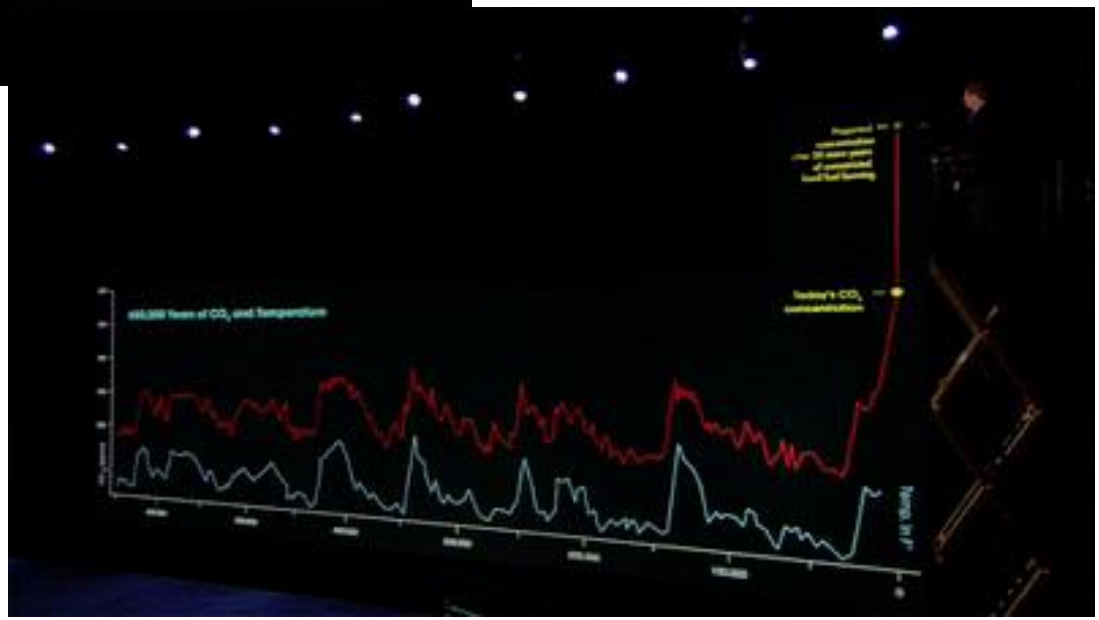


2006

http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

An Inconvenient Truth





Gore made extensive use of data graphics



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

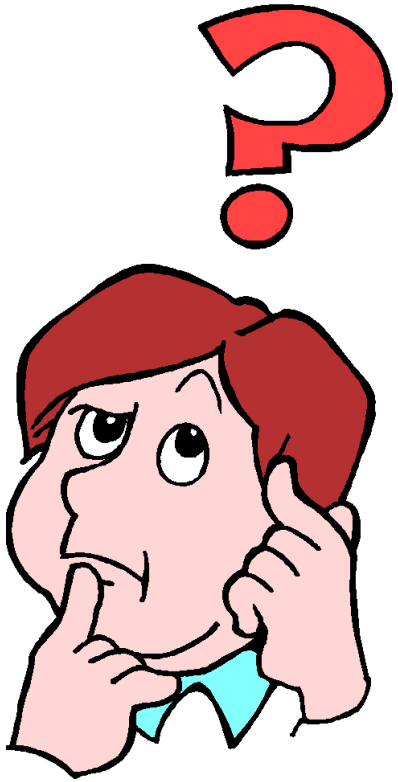
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context



Investigative analysis



COVER FEATURE

Visual Analytics Support for Intelligence Analysis

Carsten Görg, *University of Colorado*
Younah Kang, *Google*
Zhicheng Liu, *Stanford University*
John Stasko, *Georgia Tech*

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open source news articles. Insights from visual analytics projects and a hypothetical scenario show the potential of visual analytics to aid these investigations.

Intelligence analysis has been a key application domain for visual analytics since the US Department of Homeland Security created the National Visualization and Analytics Center in 2004. An initial research roadmap[1] described the challenges and goals of this new domain and identified tasks, data, and analytical scenarios focused on homeland security and terrorism prevention. A passage from the roadmap characterizes the grand challenge for visual analytics systems:

> The analysis of overwhelming amounts of disparate, conflicting, and dynamic information is central to identifying and preventing emerging threats, protecting our borders, and responding in the event of an attack or other disaster. This analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations to both detect the expected and discover the unexpected.

Much of this challenge remains,[2] although visual analytics technologies certainly hold great promise. A few commercial visual analytics tools for intelligence analysis are emerging, such as Analyst's Notebook from IBM i2 (www.ibm.com/software/industry/i2/software), nSpace from Oculus (www.oculusinfo.com/nspace), and Palantir's suite of systems (www.palantir.com). However, continued progress depends heavily on gaining a deeper understanding of intelligence analysis and the role of an analyst, as well as clarifying how visual analytics can help investigators. These insights must inform any design of any visual analytics system that aims to support intelligence analysis.

The "Investigative Scenarios" sidebar describes the kinds of analysis that are prevalent in the intelligence domain. The scale, diversity, and complexity of the information to be explored make such analyses cognitively demanding. Information is often narrative text, not quantitative data, and as such is not as amenable to automated analysis. Particularly challenging are the "pain points" in the intelligence process—the cost of scanning, recognizing (assessing), and selecting items for further attention; a limited attention span for evidence and hypotheses; and the difficulty of generating alternative hypotheses.[3]

To better understand requirements for visual analytics support, we conducted a series of research projects on intelligence analysts. The projects, which spanned 2008 to 2015, included an observational study to better understand the intelligence analysis process and its characteristics. During this time, we also developed Jigsaw (www.cc.gatech.

30 COMPUTER Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

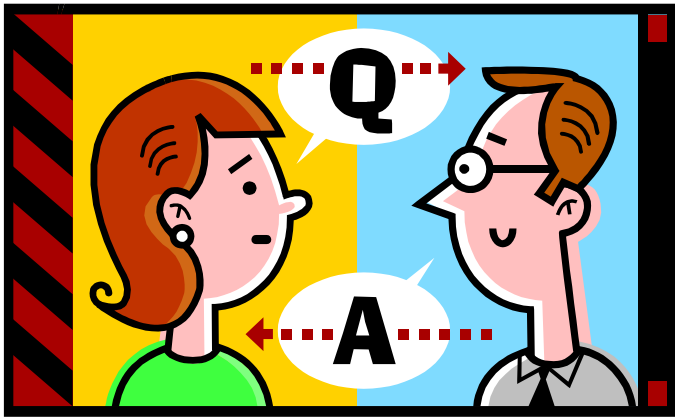
Görg, Kang, Kiu, Stasko
IEEE Computer '13

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

Interaction is Vital

Engage in a dialog with your data



1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect



Toward a Deeper Understanding of the Role of Interaction in Information Visualization

Ji Soo Yi, Youn ah Kang, John T. Stasko, *Member, IEEE*, and Julie A. Jacko

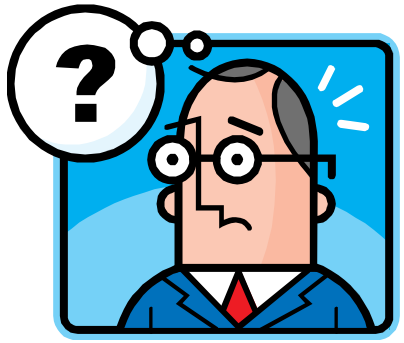
Yi, Kang, Stasko, Jacko
TVCG (InfoVis) ‘07

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

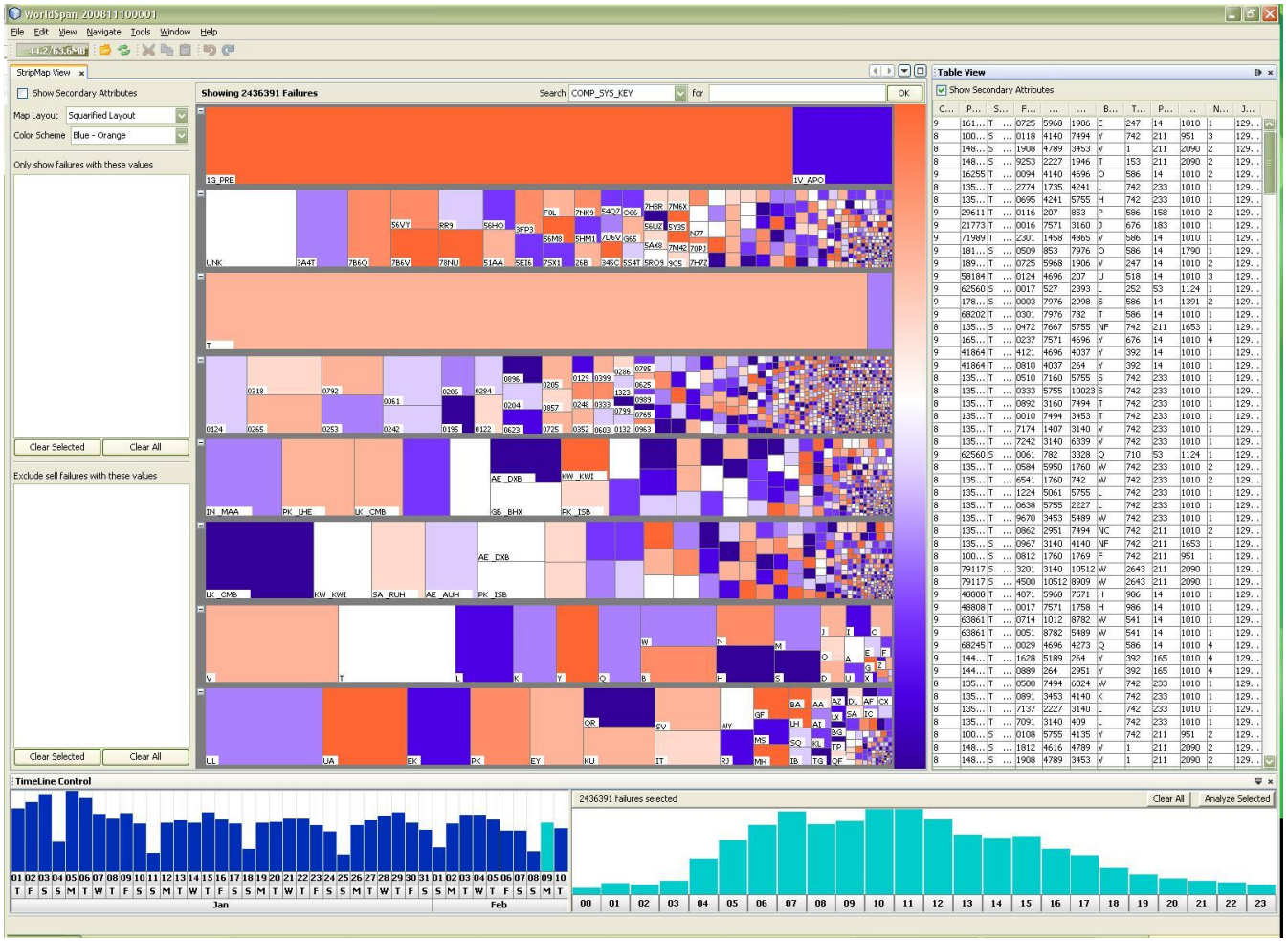
Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Some examples from my group's research

SellTrend



Web-based airline booking requests

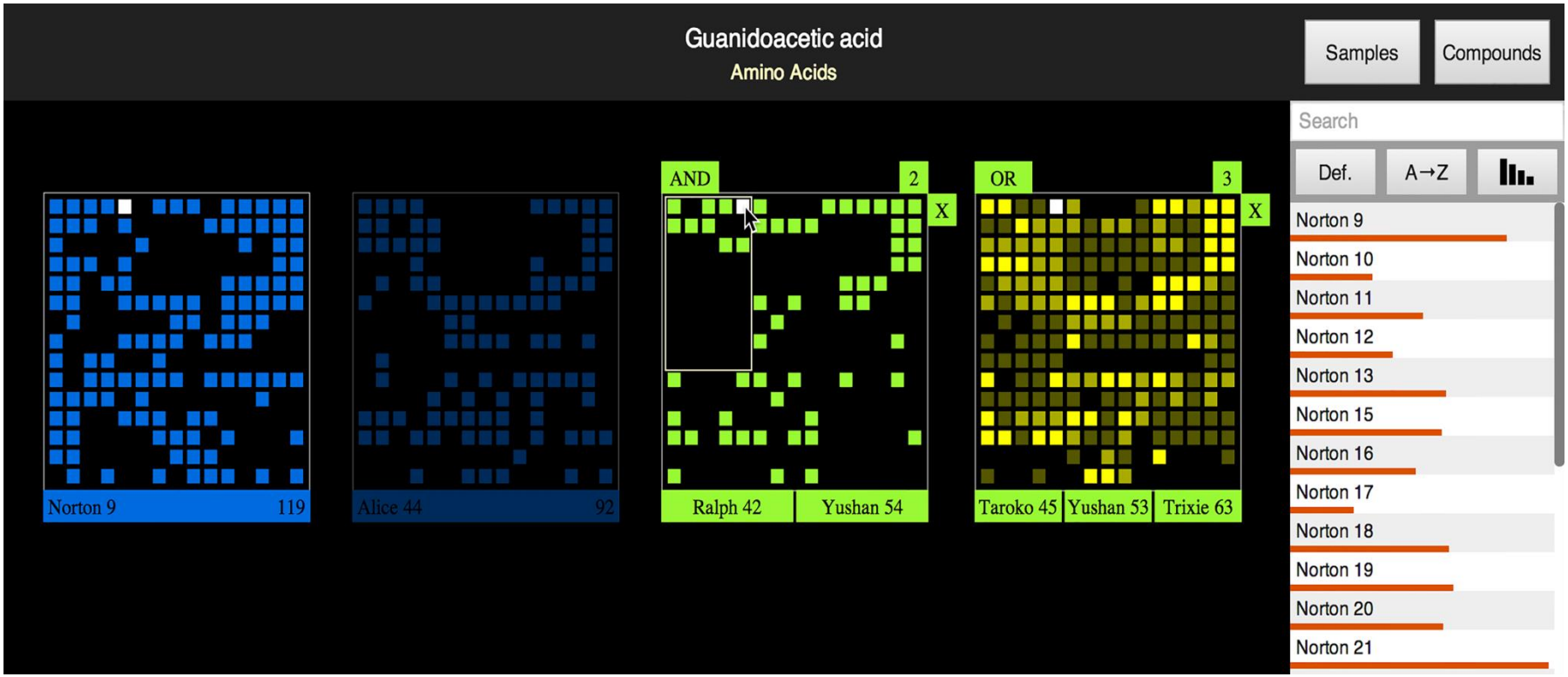
Help recognize irregularities and analyze what the problem is

http://www.cc.gatech.edu/gvu/ii/selltrend/

Liu, Sullivan, Stasko
TVCG (InfoVis) '09

OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/

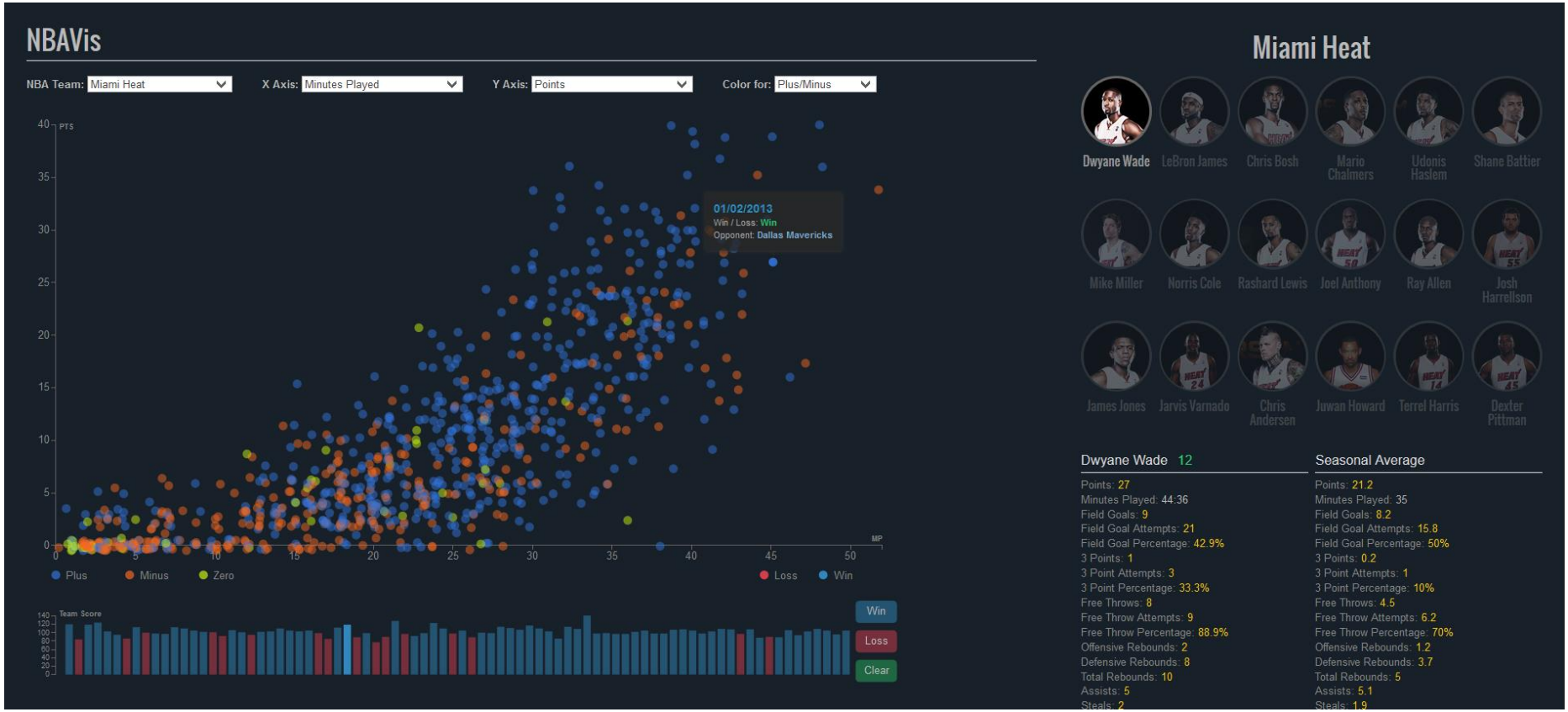


Demo

Sadana, Major, Dove, & Stasko

NBA Player and Team Data

http://www.cc.gatech.edu/gvu/ii/sportvis/nbaVis/

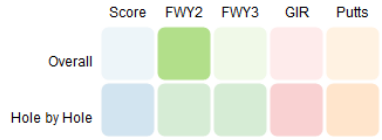


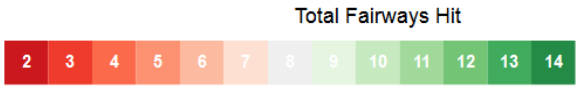
Project by Fengbo Li

Demo

PGA Tour Season Recap

Using ShotLink data





2013 | Phil Mickelson

Sort by Date	Round 1	Round 2	Round 3	Round 4	Position	Total	Strokes
Humana Challenge					T37	-17	271
Farmers Insurance Open					T51	E	288
Phoenix Open					1	-28	256
Pebble Beach Pro-Am					T60	-1	285
Northern Trust Open					T21	-3	281
WGC-Cadillac Champ.					T3	-14	274
Arnold Palmer Invitational					CUT	+8	152
Houston Open					T16	-10	278
The Masters					T54	+9	297
Wells Fargo Champ.					3	-7	281
PLAYERS Champ.					CUT	+1	145
St. Jude Classic					T2	-10	270
U.S. Open					T2	+3	283
The Greenbrier					CUT	+2	142
Open Champ.					1	-3	281
Bridgestone					T21	+1	281
PGA					T72	+12	292
Barclays					T6	-9	275
Deutsche Bank					T41	-8	276
BMW Champ.					T33	-1	283

with Anand Sainath, Rahul Basole

“Putting the pieces together”

Jigsaw

Computational analysis of documents’ text

- Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

- Views are highly coordinated



Help “investigators” explore, analyze and understand large document collections

Academic Papers

Health Forums

Product Reviews

Police Reports



Designing and Implementing an Interac
Scatterplot Visualization for a Tablet Con

Ramik Sadana
Georgia Institute of Technology
ramik@gatech.edu

John Stasi
Georgia Institute of T
stasko@cc.gate

ABSTRACT

Tablet computers now offer screen sizes and computing capabilities that are competitive with traditional desktop PCs. Their popularity has grown tremendously, but we are just beginning to see information visualization applications designed for this platform. One potential reason for this limited development is the challenge of designing and implementing a multi-touch interface for visualizations on mobile, tablet devices. In this work, we identify the primary challenges that touch screen interactions pose for information visualization applications. We explore the design space of multi-touch interactions for visualizations and present a prototype information visualization application using a specific technique, a dynamic scatterplot, for an iPad.

Categories and Subject Descriptors
H.5.0 [Information Interfaces and Presentation]: General

General Terms
Design, Algorithms, Human Factors.

Keywords
Information visualization, multi-touch interaction, scatterplot, tablet computer, gesture.

1. INTRODUCTION

The popularity of tablet computers has grown tremendously in recent years. Devices such as the Apple iPad, Amazon Kindle, and Microsoft Surface constitute a significant portion of computer sales today and a wide variety of applications have been tailored to tablet platform. However, one area with just a few initial applications for tablets is information visualization. Though popular commercial systems such as Tableau and Spotfire have introduced tablet versions in the last two years, these systems still feel much like a port of a desktop application. They do not yet leverage the set of gestural interactions that touch-based interfaces provide and do not offer a rich, multi-touch interface in the style of other tablet applications.

We speculate that the limited development of information visualization applications for tablets has resulted from the challenge of designing and implementing the interface. First, most tablets still provide a smaller screen size, which is a limitation for data visualization. Perhaps more importantly, information visualization applications generally have many small visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the authors must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AVI'14, May 27 - 29 2014, Como, Italy
Copyright is held by the authors. Publication rights licensed to ACM.
ACM 978-1-4503-2775-6/14/05...$15.00.

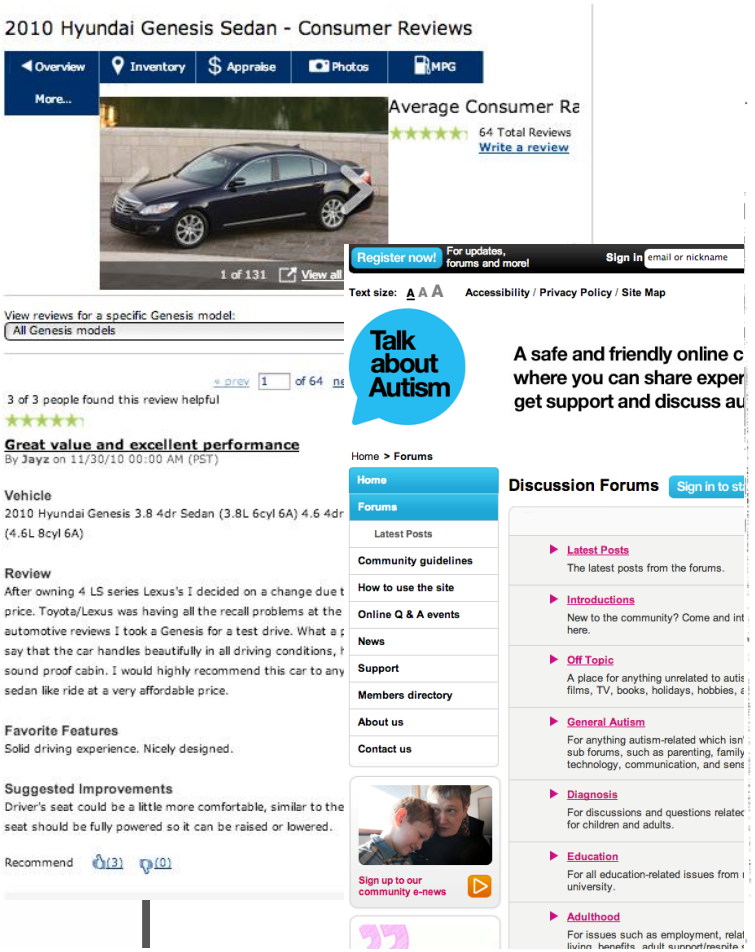
objects to select and manipulate, interactive widgets such as buttons, boxes. Desktop information visualiz make extensive use of a WIMP (w interface. Translating visualization operations to a finger-directed multi keyboard and mouse is simply a very

Some recent research has begun however. One particular avenue development of multi-user, multi applications for tabletop computers [1 explored the design space for s developing for tablet computers is s tabletop computers. Baur et al. [2] rec tablet interface for a StreamGraph-ty This application is much more alig create. Our specific focus, however, used visualization technique. Rzeszot application is a similar technique to driven interactions.

Rather than viewing the movement applications to multi-touch tablets software port, we see this problem challenge. Initial research [25] has interaction design space – for any information visualization syster implementations make sense and Furthermore, we will show that a gestures from tablet applications p applied to data visualizations.

In this work, we further explo implement a working information vi tablet PC. We chose a dynamic sca technique of our focus. Scatterplots a in information visualization; many o suite of provided techniques. Scatterp in data visualization, including the FilmFinder system [1] that led to t Our project starts from "ground zero an existing system. Our application advantage of, and to address all the l a tablet interface using only fingers a

The contributions of our work are sections of this article. By studying commercial implementations of a dyn set of tasks and interactive operation would support (Section 3). Next, we the tasks and describe how those c design (Section 4). We ultimately c multi-touch operations for each task, for each, propose multiple potenti many of the options in a proto application for the iPad (Section 5). I paper may not convey all the operati overview summarizing the applicatio



2010 Hyundai Genesis Sedan - Consumer Reviews

Overview | Inventory | Appraise | Photos | MPG | More...

Average Consumer Ra
64 Total Reviews
Write a review

1 of 131 View all

Register now! For updates, forums and more! Sign in email or nickname

Text size: A A A Accessibility / Privacy Policy / Site Map

View reviews for a specific Genesis model:
All Genesis models

« prev 1 of 64 ne

3 of 3 people found this review helpful

Great value and excellent performance
By Jayz on 11/30/10 00:00 AM (PST)

Vehicle
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 4dr (4.6L 8cyl 6A)

Review
After owning 4 LS series Lexus's I decided on a change due t price. Toyota/Lexus was having all the recall problems at the automotive reviews I took a Genesis for a test drive. What a p say that the car handles beautifully in all driving conditions, h sound proof cabin. I would highly recommend this car to any sedan like ride at a very affordable price.

Favorite Features
Solid driving experience. Nicely designed.

Suggested Improvements
Driver's seat could be a little more comfortable, similar to the seat should be fully powered so it can be raised or lowered.

Recommend (3) (0)

Talk about Autism

A safe and friendly online c where you can share exper get support and discuss au

Home > Forums

Home
Forums
Latest Posts
Community guidelines
How to use the site
Online Q & A events
News
Support
Members directory
About us
Contact us

Discussion Forums Sign in to st

Latest Posts
The latest posts from the forums.

Introductions
New to the community? Come and int here.

Off Topic
A place for anything unrelated to autis films, TV, books, holidays, hobbies,

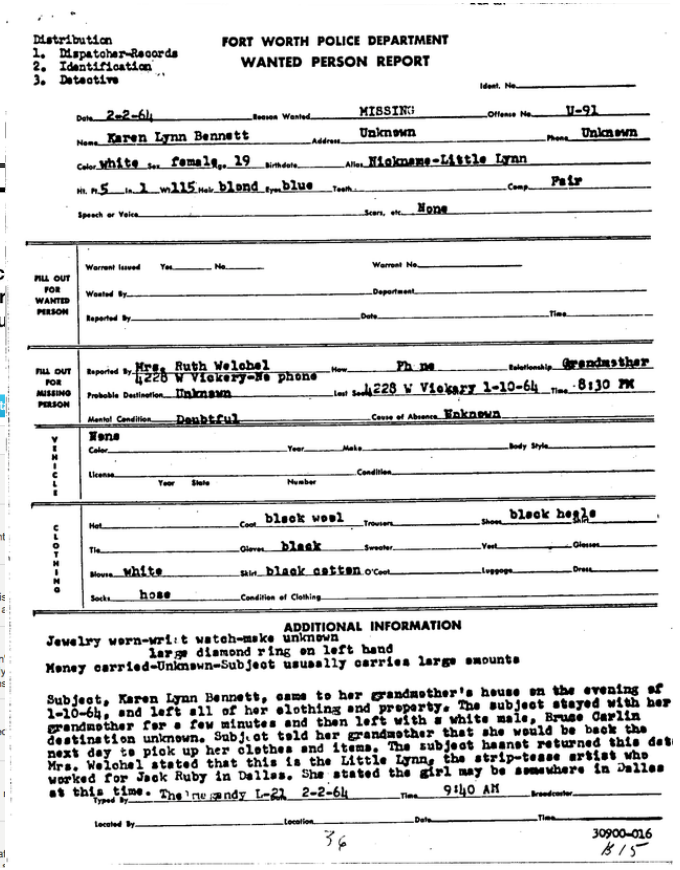
General Autism
For anything autism-related which isn sub forums, such as parenting, family technology, communication, and sens

Diagnosis
For discussions and questions relatec for children and adults.

Education
For all education-related issues from university.

Adulthood
For issues such as employment, relat

Sign up to our community e-news



Distribution
1. Dispatcher-Records
2. Identification
3. Detective

FORT WORTH POLICE DEPARTMENT
WANTED PERSON REPORT

Ident. No.

Date 2-2-64 Reason Wanted MISSING Offense No. U-91
Name Karen Lynn Bennett Address Unknown Phone Unknown
Color White Sex female Age 19 Birthdate Alias Nickname-Little Lynn
Ht. ft. 5 in. 1 Wt. 115 Hair blond Eyes blue Teeth Comp. Fair
Speech or Voice Scars, etc. None

FILL OUT FOR WANTED PERSON
Warrant Issued Yes No Warrant No.
Wanted By Department
Reported By Date Time

FILL OUT FOR MISSING PERSON
Reported By Mrs. Ruth Welchel How Ph me Relationship Grandmother
Probable Destination Unknown Last Seen 4228 W Vickery 1-10-64 Time 8:30 PM
Mental Condition Doubtful Cause of Absence Unknown

VEHICLE
None
Color Year Make Body Style
License Condition
Year State Number

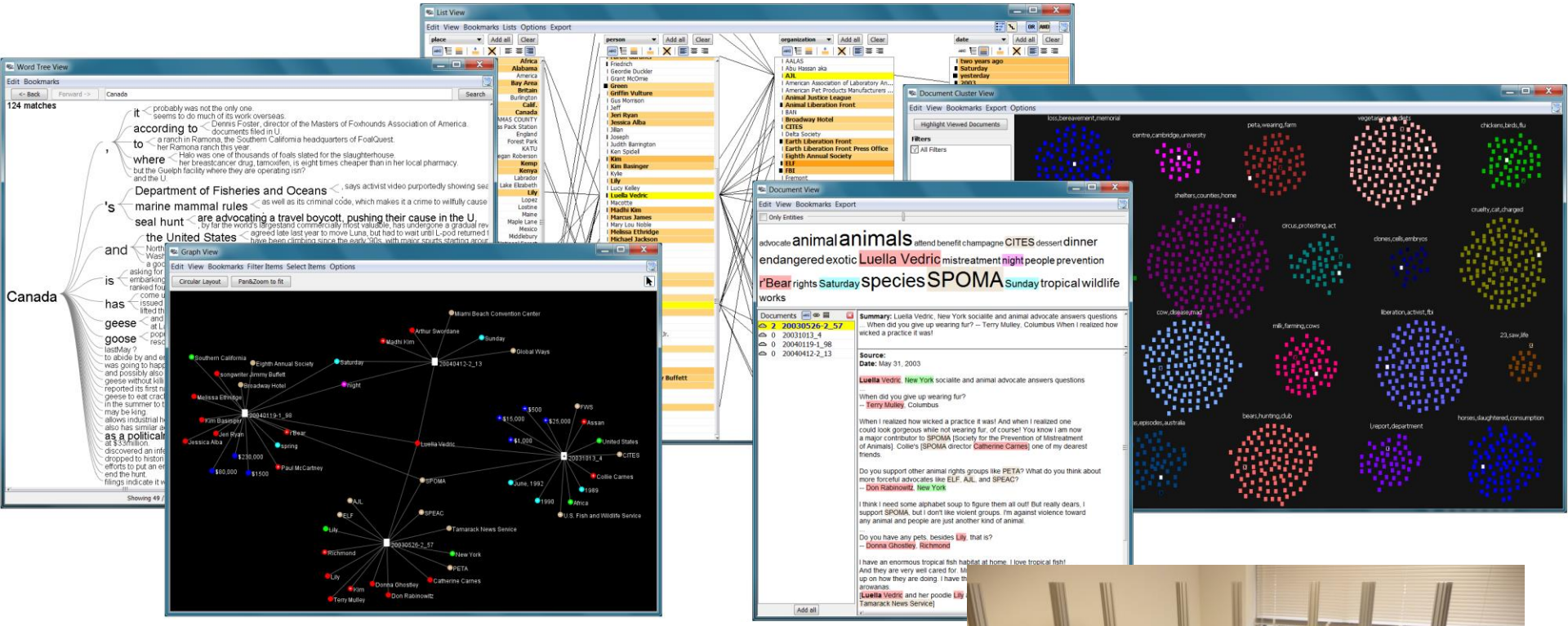
CLOTHING
Hat Coat black wool Trousers Shoes black hegle Belt
Tie Gloves black Sweater Vest Glasses
Blouse white Skirt black cotton Overcoat Luggage Dress
Socks hose Condition of Clothing

ADDITIONAL INFORMATION
Jewelry worn-wrist watch-make unknown
large diamond ring on left hand
Money carried-Unknown-Subject usually carries large amounts

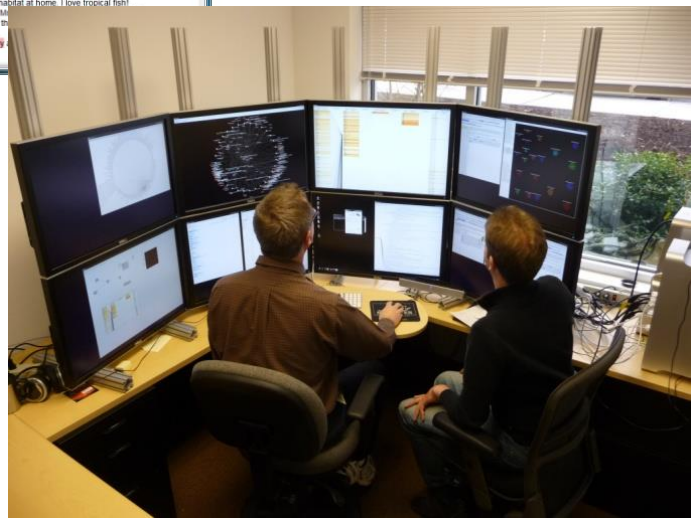
Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothing and property. The subject stayed with her grandmother for a few minutes and then left with a white male, Bruce Carlin destination unknown. Subject told her grandmother that she would be back the next day to pick up her clothes and items. The subject hasnot returned this da Mrs. Welchel stated that this is the Little Lynn, the strip-tease artist who worked for Jack Ruby in Dallas. She stated the girl may be somewhere in Dallas at this time. The investigation L-21 2-2-64 Time 9:40 AM
Typed By
Located By Location Date Time
36
30900-016
8/5

http://www.cc.gatech.edu/gvu/ii/jigsaw

Jigsaw



Visual analytics of large text document collections



Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG '13

MEMBER LOGIN: Username: Password: LOGIN

Forgot My Username/Password

Remember Login Info

eRobertParker.com

The Independent Consumer's Guide to Fine Wines



Rock and Rhône:
Parker Explores the Greatest Unheralded Grapes in the World–Grenache, Mourvèdre, and Syrah

HOME SUBSCRIBE GIFT SUBSCRIPTIONS FAQs VIRTUAL TOUR SUPPORT SITE MAP CONTACT US

What's Inside

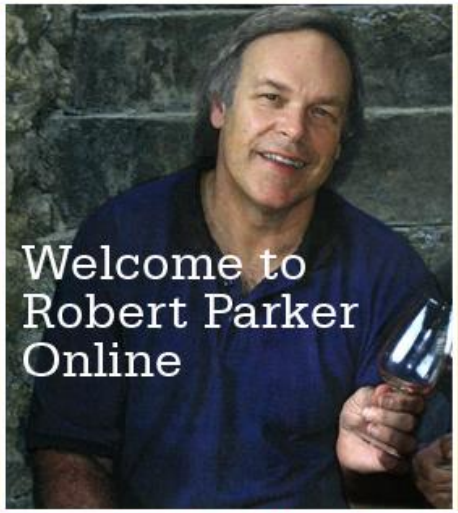
Explore Wines
Over 190,000 searchable notes from *The Wine Advocate*, Robert Parker's books and other sources.

The Hedonist's Gazette
An insider's look at our reviewers' wine and food experiences.

Article Archive
In-depth commentary from *The Wine Advocate* and other sources.

My Wines
Powerful, easy-to-use tools for purchasing/managing a cellar and participating in our wine community.

Robert Parker Mobi
Search the notes from any Internet-enabled mobile device.





Robert Parker's Wine Advocates

The Latest Information from the World's Wine Experts

Robert Parker
Reviews and Articles

Antonio Galloni's
In the Cellar

Neal Martin's
Wine Journal

Jay Miller
Reviews and Articles

Lisa Perrotti-Brown MW
In Asia

David Schildknecht
Reviews and Articles

Mark Squires'
E-zine and Bulletin Board

Parker & Zraly
Wine Certification Program

Information and Products for Non-Subscribers

Wine of the Day:
2009 Ridge Chardonnay
View Tasting Note



3W's of Wine iPhone App
Learn More

The Wine Advocate Store
Shop Now



Parker & Zraly Wine Certification
Learn More

Weekly Wine Buys
Powered by winealert
Learn More

Quick Links	More Information	About Our Reviewers	Links
Username/Password	**The Vintage Chart**	Robert Parker	The Wine Advocate Fund
What's New	Executive Wine Seminars	Antonio Galloni	Educational
Gift Subscriptions	Glossary of Wine Terms	Karen MacNeil	Accessory/Storage Links
Get News Via RSS	Parker's Rating System	Neal Martin	Retailer Links
Shop The Wine Advocate Store	*The Wine Advocate*	Jay Miller	Other
TWA Editorial Calendar	Site Map	Lisa Perrotti-Brown MW	
		David Schildknecht	
		Mark Squires	
		Kevin Zraly	
		Past Reviewers	
		Writer Ethics/Standards	





Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, …

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

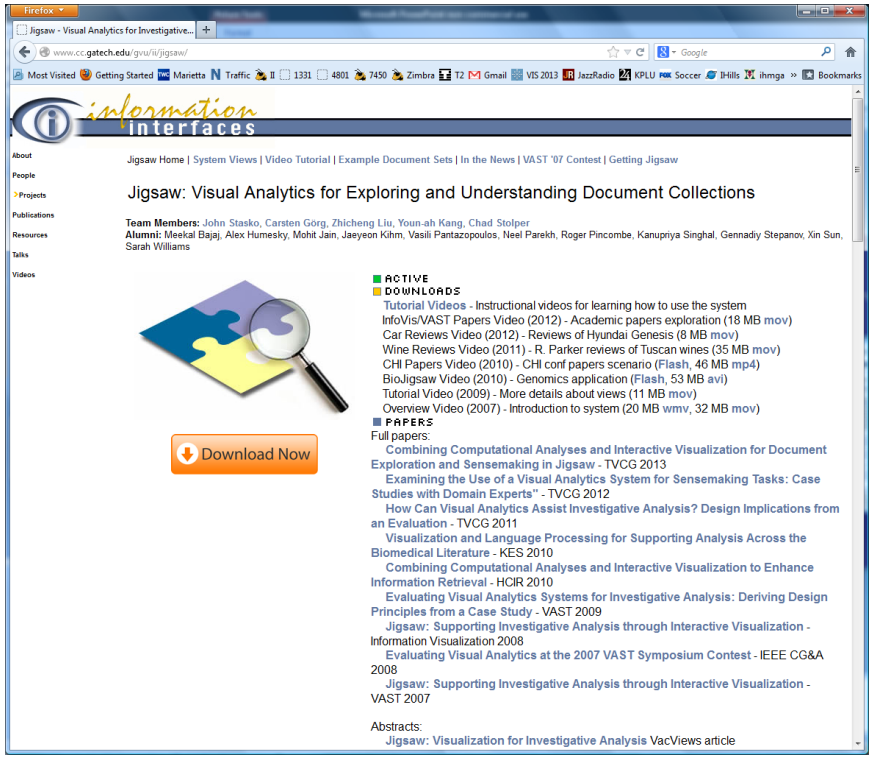
from database of 150,000 reviews





Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw

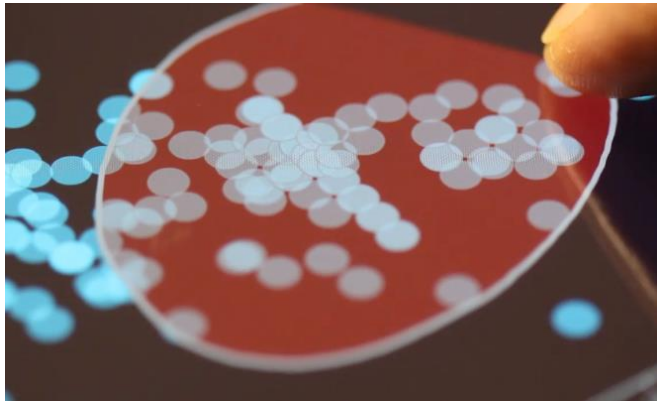




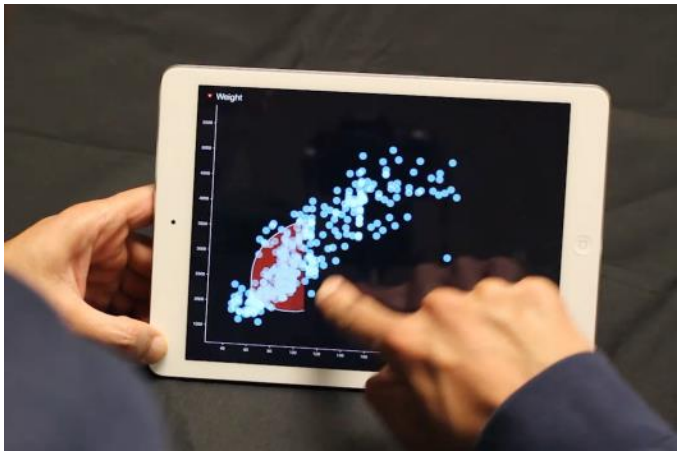
Applications

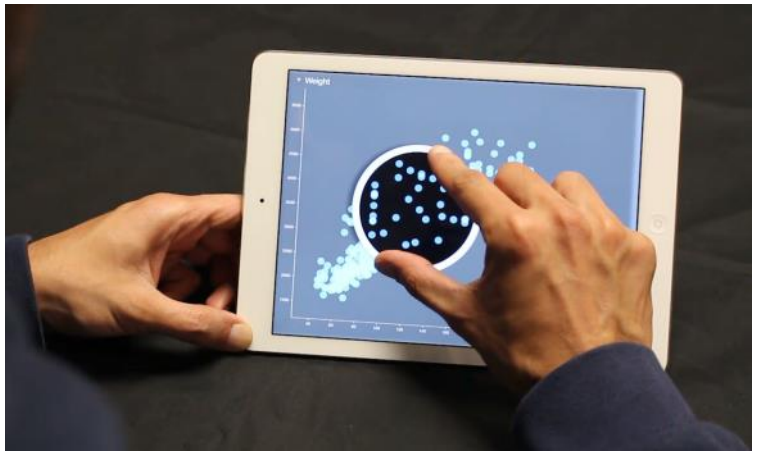
- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

Moving to Tablets



Scatterplot





Video

Sadana & Stasko
AVI ‘14

Different benefits of visualization

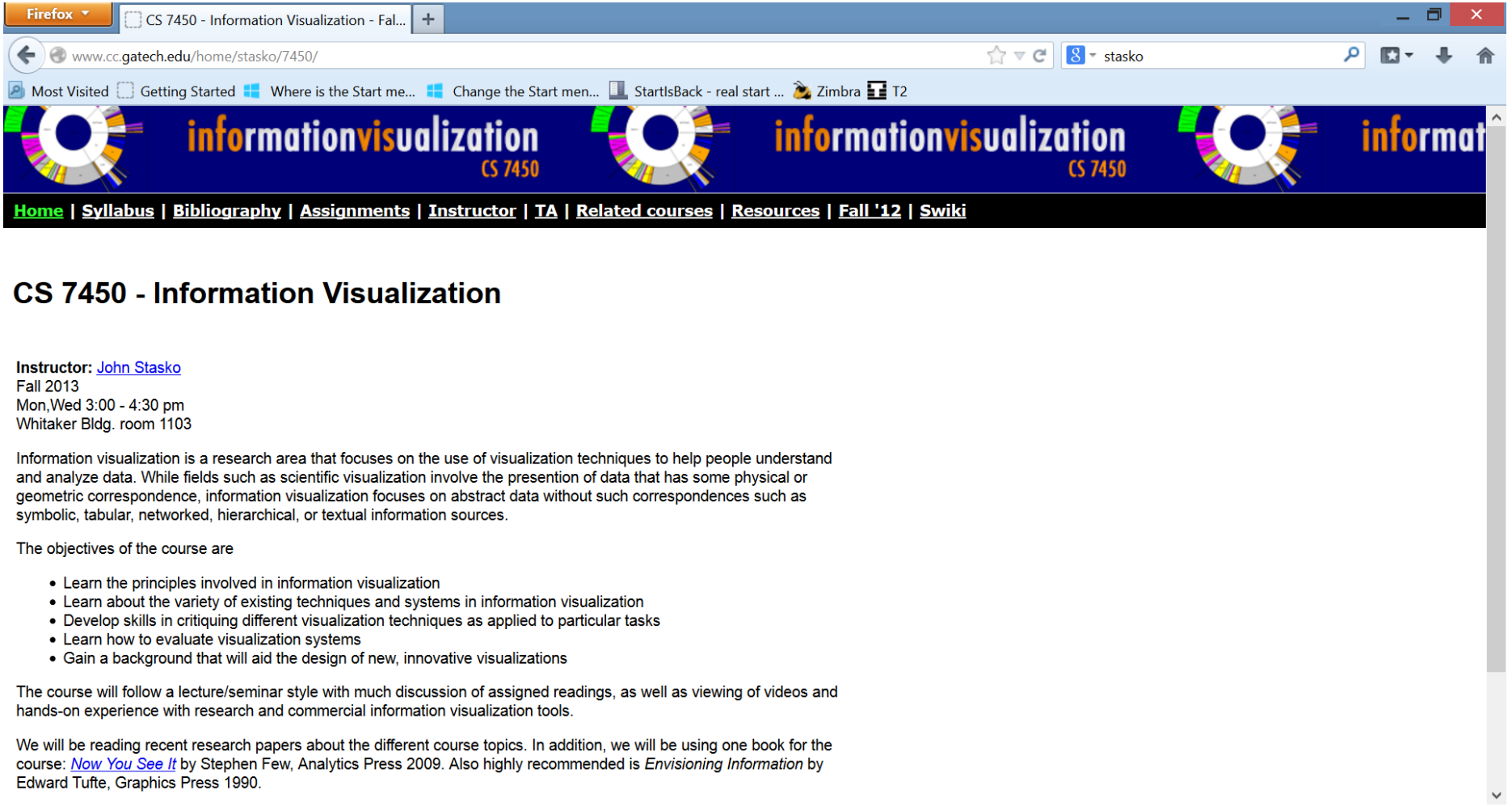
Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

My Class

http://www.cc.gatech.edu/~stasko/7450



My Class

http://www.cc.gatech.edu/~stasko/7450

Syllabus

Link to last year's 2012 syllabus with lecture slides

Overview

Week	Dates	Topic	Topic	HW
1	Aug 19, 21	Introduction	InfoVis overview	HW 1
2	Aug 26, 28	Visual perception	Value/Benefits of visualization	
3	Sep 2, 4	No Class -- Labor Day	Multivariate data & table/graph design	HW 2
4	Sep 9, 11	Few's design guidance	Multivariate visual representations 1	
5	Sep 16, 18	Multivariate visual representations 2	Tasks and analysis	HW 3a HW 3b
6	Sep 23, 25	InfoVis systems & toolkits	Commercial systems demos	HW 4
7	Sep 30, Oct 2	Storytelling	Tufte's design principles	
8	Oct 7, 9	Poster session	Casual InfoVis	HW 5
9	Oct 14, 16	No Class - Fall break	No Class - VIS Conference	
10	Oct 21, 23	Graphs and networks 1	Graphs and networks 2	
11	Oct 28, 30	Hierarchies & trees 1	Hierarchies & trees 2	HW 6
12	Nov 4, 6	Interaction	Overview & detail	
13	Nov 11, 13	Text & documents 1	Text & documents 2	
14	Nov 18, 20	Visual analytics 1	Visual analytics 2	HW 7
15	Nov 25, 27	Time series data	Project work day	
16	Dec 2, 4	Evaluation	Review	

Detail

Below are summaries of all the materials for each class. Lecture slides, references for readings and articles, videos shown, and software demonstrated are listed for each topic. Also available is a more general, alphabetical bibliography.

http://vimeo.com/98986594

For more, my EuroVis ‘14 Capstone Talk





Take Aways

Presentation & analysis

Exploratory data analysis, power from interaction

Developing new questions and insights

Acknowledgments

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)





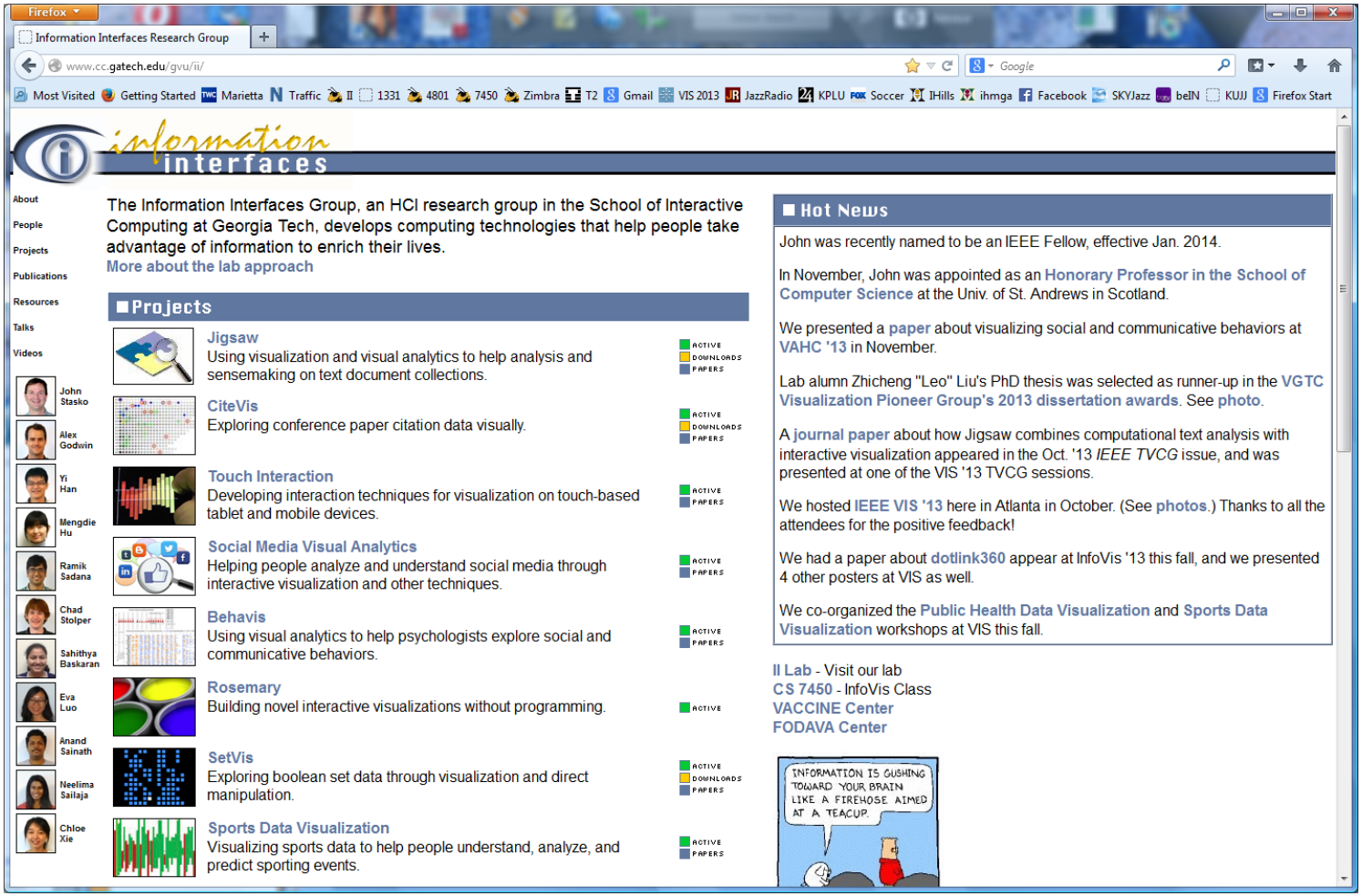
- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA program

My Research

http://www.cc.gatech.edu/gvu/ii



September 24, 2014

The Value of Visualization for Understanding Data and Making Decisions

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu

JISIC 2014





Georgia Tech College of Computing
School of Interactive Computing


Data

Many Data Analysis Approaches

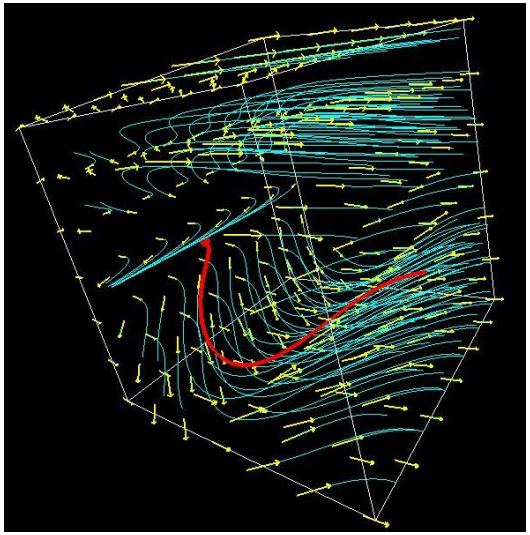
Statistics
Database & information retrieval
Data mining
Machine learning

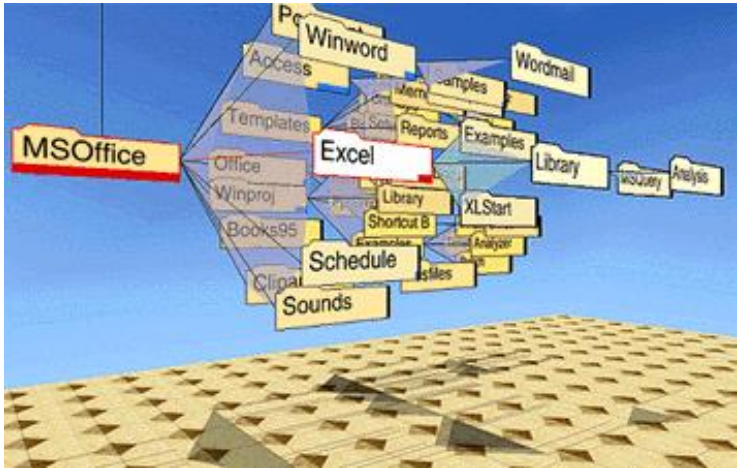


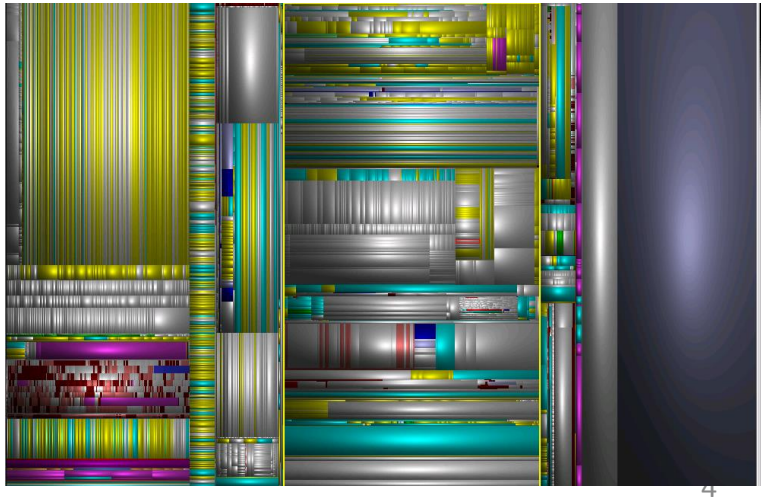


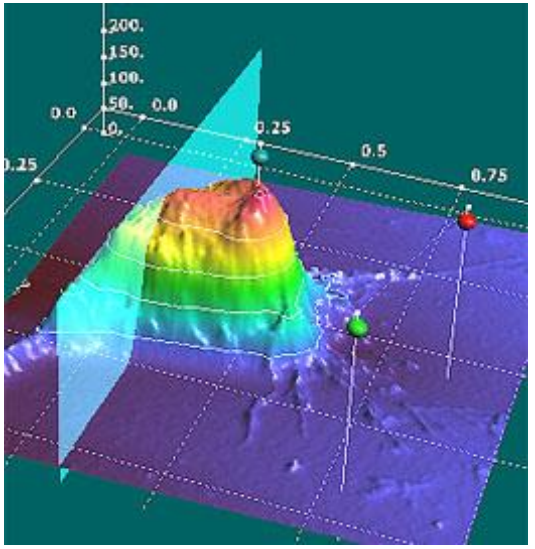
Data Visualization

~~Making pretty pictures?~~









Data Visualization

A cognitive process

Gain an understanding

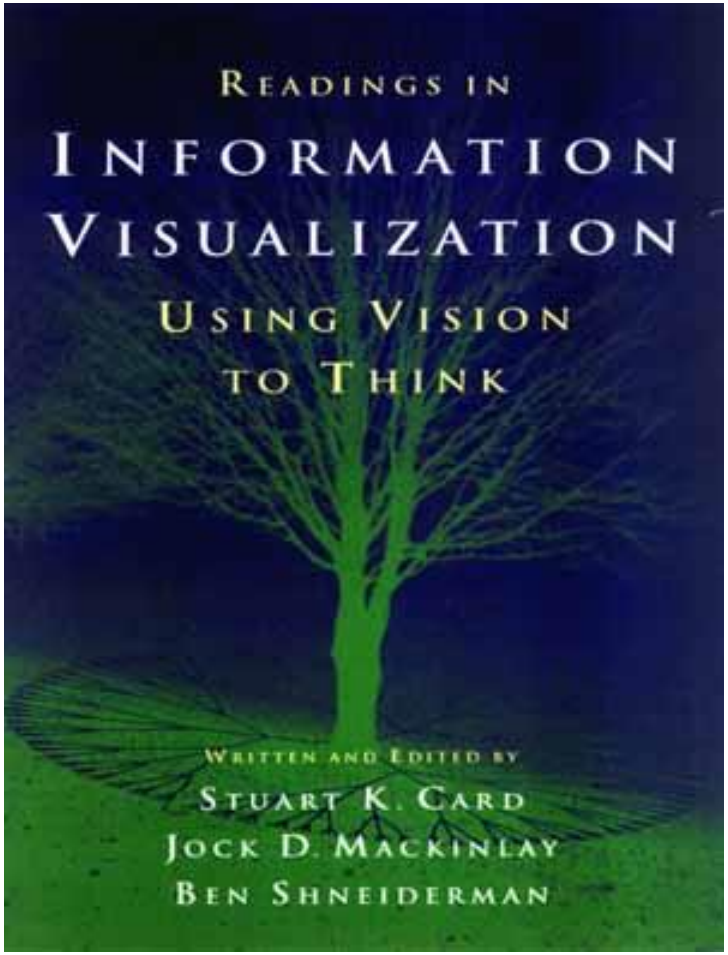


Data Visualization

A cognitive process

Gain an **understanding**





Visualization

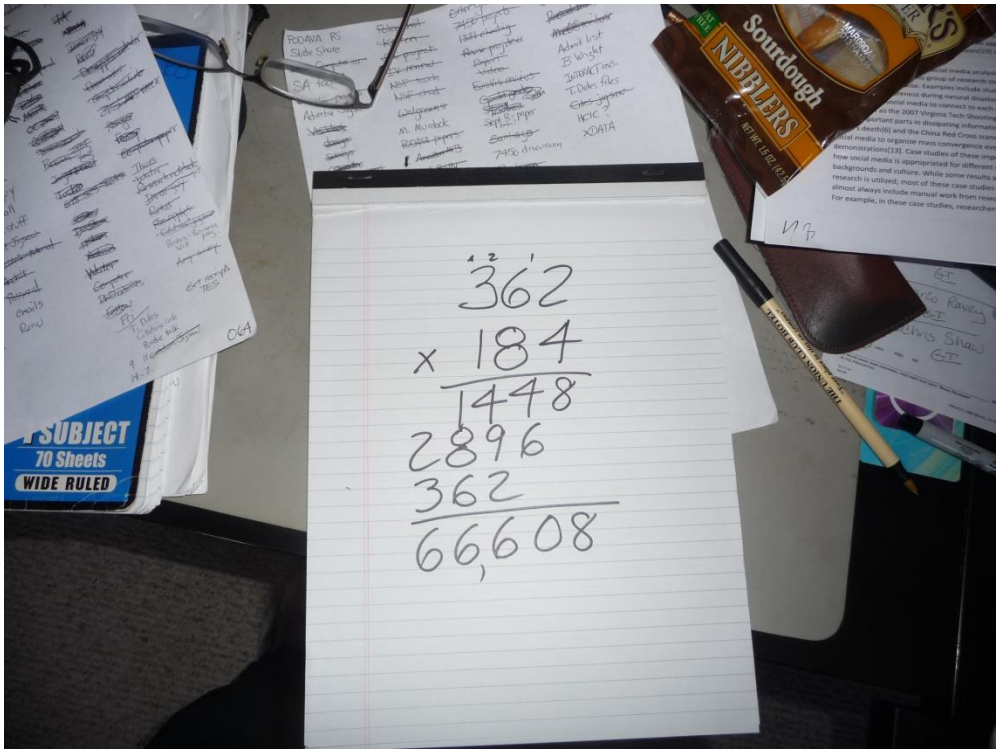
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provide a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Applications of Visualization

Presentation

Analysis

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support

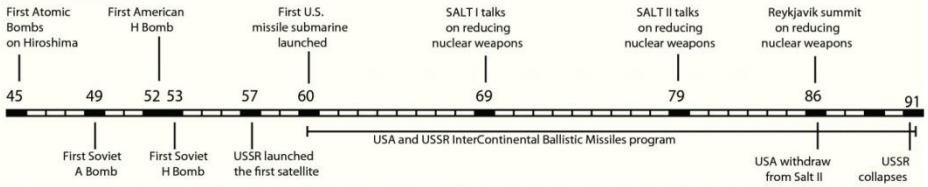


Infographics

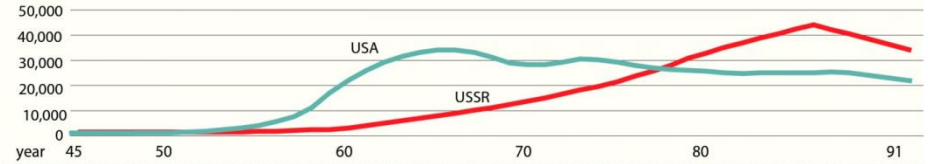
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



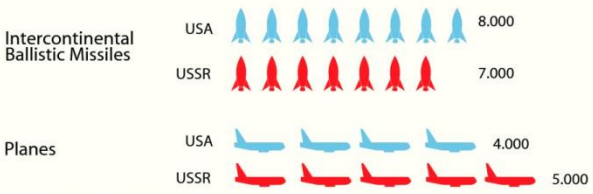
NUCLEAR STOCKPILE



ROCKET MODELS



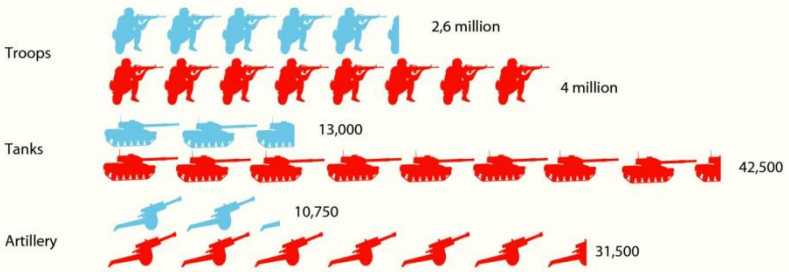
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



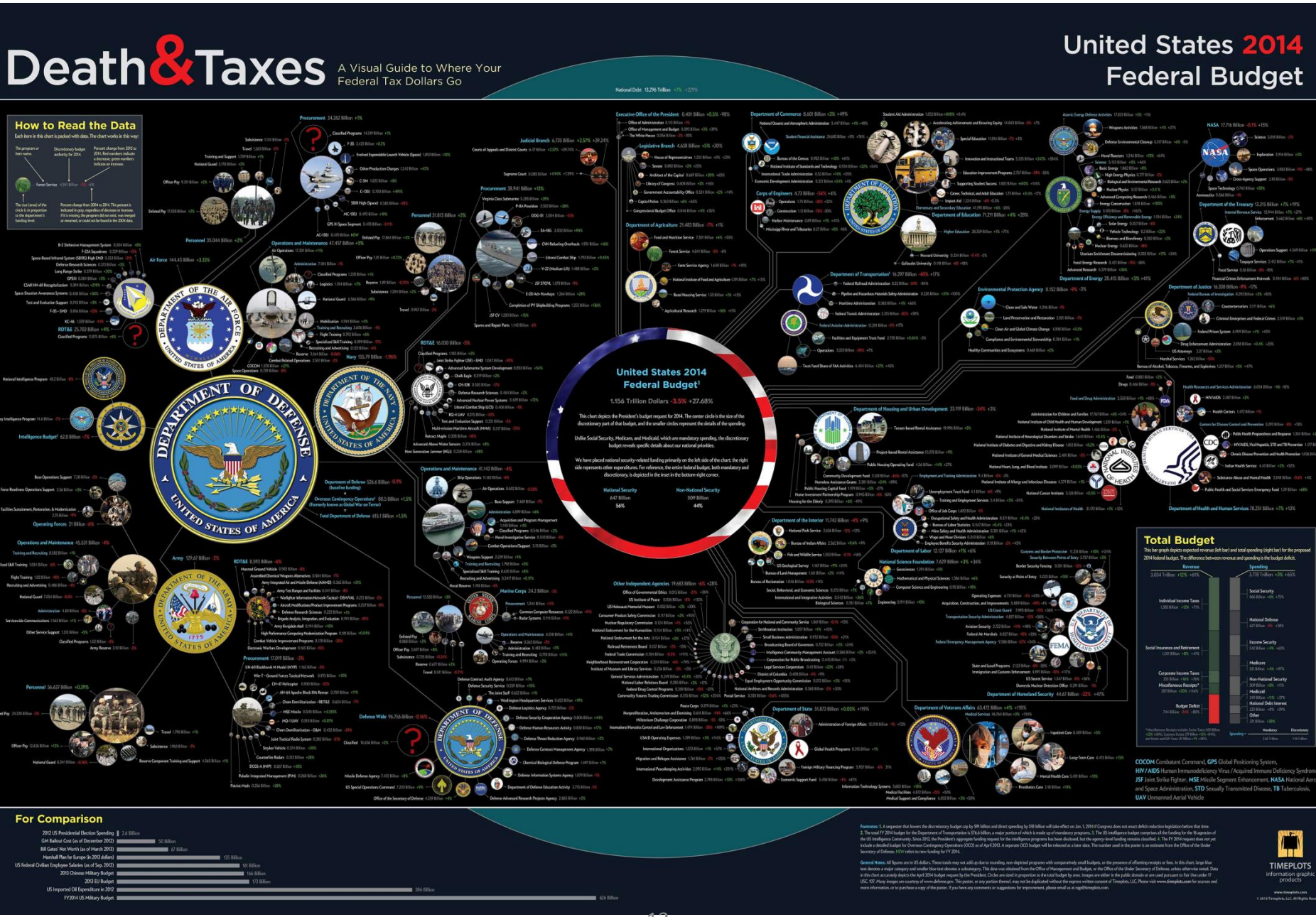


Produced by WatchTheAmericans.com a Fan Website

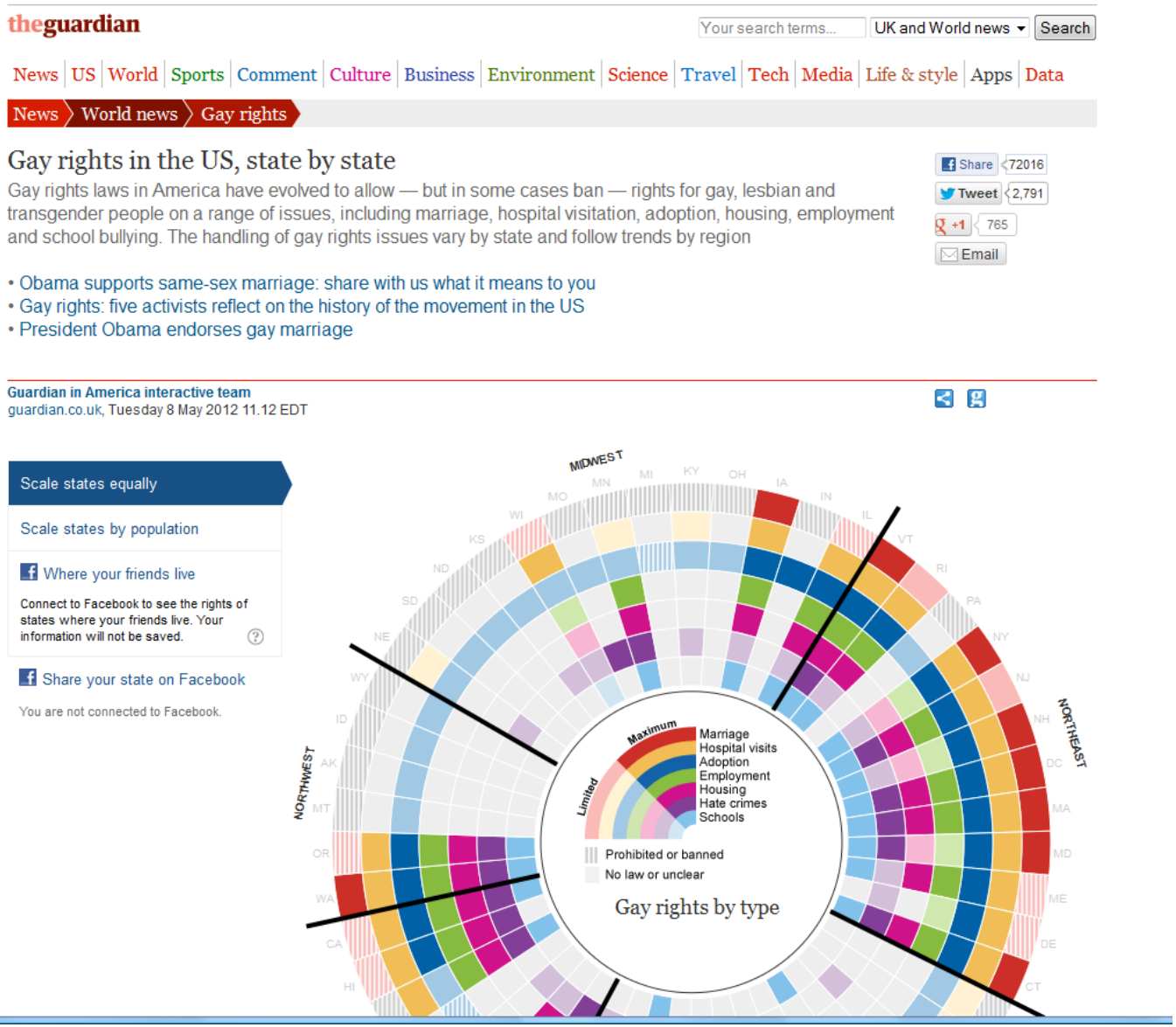
Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget

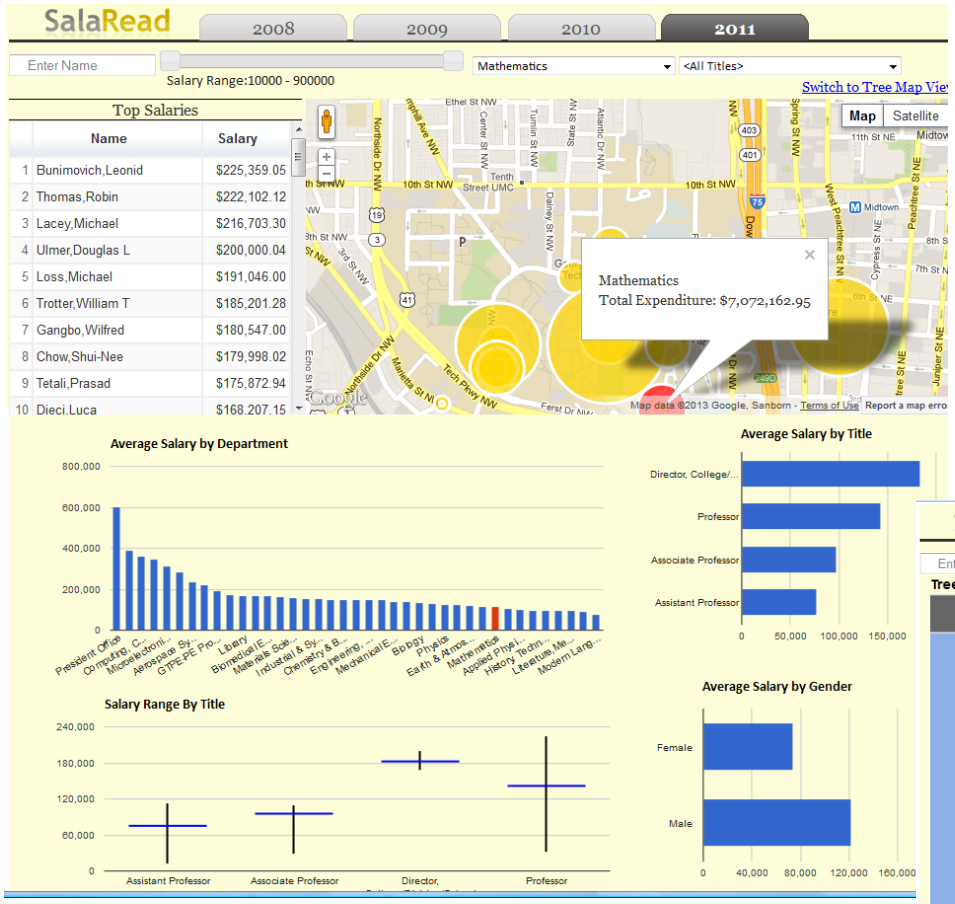


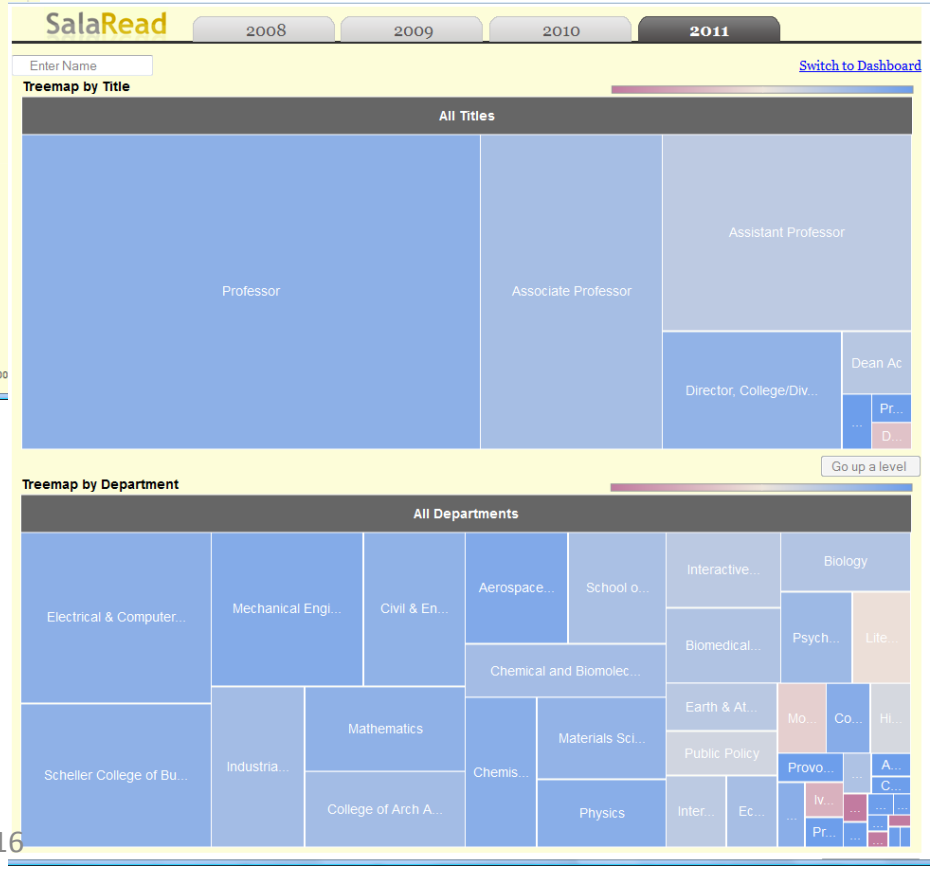
http://www.guardian.co.uk/world/interactive/2012/may/08/gay-rights-united-states?fb=native



Simply presenting data ***visually*** can have a profound impact

CS 7450
Fall ‘12
Design project

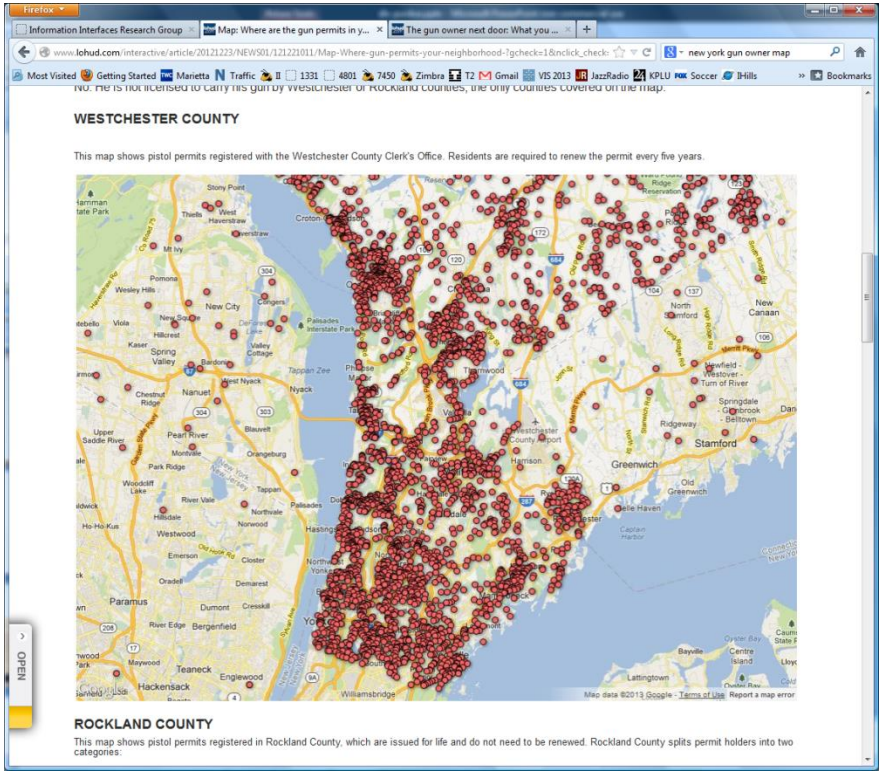




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun ownership in New York counties





http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood-?gcheck=1&nclick_check=1

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

May just show a few variables
and/or a subset of the data cases



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

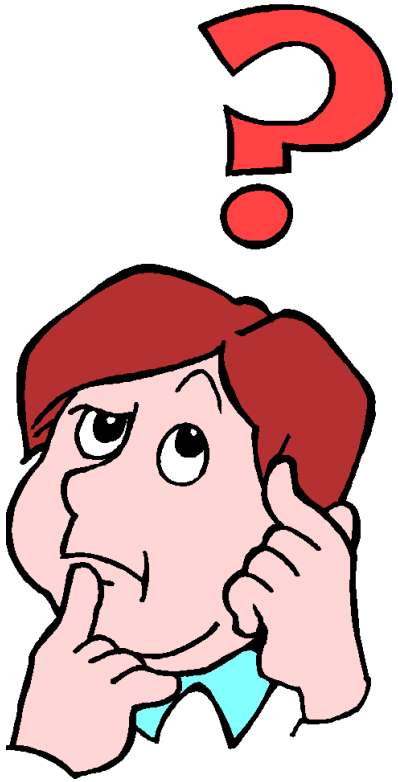
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context



Investigative analysis



COVER FEATURE

Visual Analytics Support for Intelligence Analysis

Carsten Görg, *University of Colorado*
Younah Kang, *Google*
Zhicheng Liu, *Stanford University*
John Stasko, *Georgia Tech*

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open source news articles. Insights from visual analytics projects and a hypothetical scenario show the potential of visual analytics to aid these investigations.

Intelligence analysis has been a key application domain for visual analytics since the US Department of Homeland Security created the National Visualization and Analytics Center in 2004. An initial research roadmap[1] described the challenges and goals of this new domain and identified tasks, data, and analytical scenarios focused on homeland security and terrorism prevention. A passage from the roadmap characterizes the grand challenge for visual analytics systems:

> The analysis of overwhelming amounts of disparate, conflicting, and dynamic information is central to identifying and preventing emerging threats, protecting our borders, and responding in the event of an attack or other disaster. This analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations to both detect the expected and discover the unexpected.

Much of this challenge remains,[2] although visual analytics technologies certainly hold great promise. A few commercial visual analytics tools for intelligence analysis are emerging, such as Analyst's Notebook from IBM i2 (www.ibm.com/software/industry/i2software), nSpace from Oculus (www.oculustinfo.com/nspace), and Palantir's suite of systems (www.palantir.com). However, continued progress depends heavily on gaining a deeper understanding of intelligence analysis and the role of an analyst, as well as clarifying how visual analytics can help investigators. These insights must inform any design of any visual analytics system that aims to support intelligence analysis.

The "Investigative Scenarios" sidebar describes the kinds of analysis that are prevalent in the intelligence domain. The scale, diversity, and complexity of the information to be explored make such analyses cognitively demanding. Information is often narrative text, not quantitative data, and as such is not as amenable to automated analysis. Particularly challenging are the "pain points" in the intelligence process—the cost of scanning, recognizing (assessing), and selecting items for further attention; a limited attention span for evidence and hypotheses; and the difficulty of generating alternative hypotheses.[3]

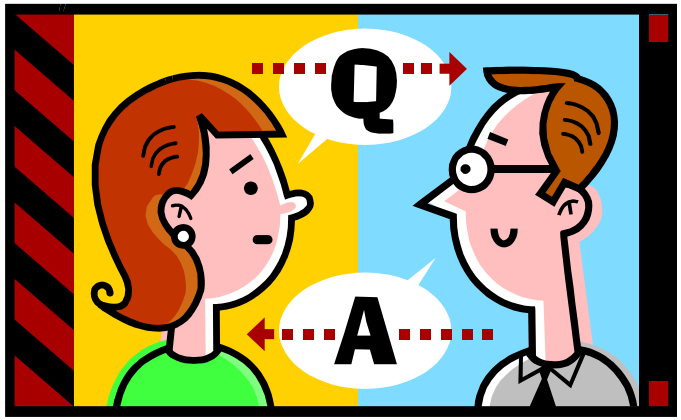
To better understand requirements for visual analytics support, we conducted a series of research projects on intelligence analysis. The projects, which spanned 2008 to 2013, included an observational study to better understand the intelligence analysis process and its characteristics. During this time, we also developed Jigsaw (www.cc.gatech.

 Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

Görg, Kang, Kiu, Stasko
IEEE Computer '13

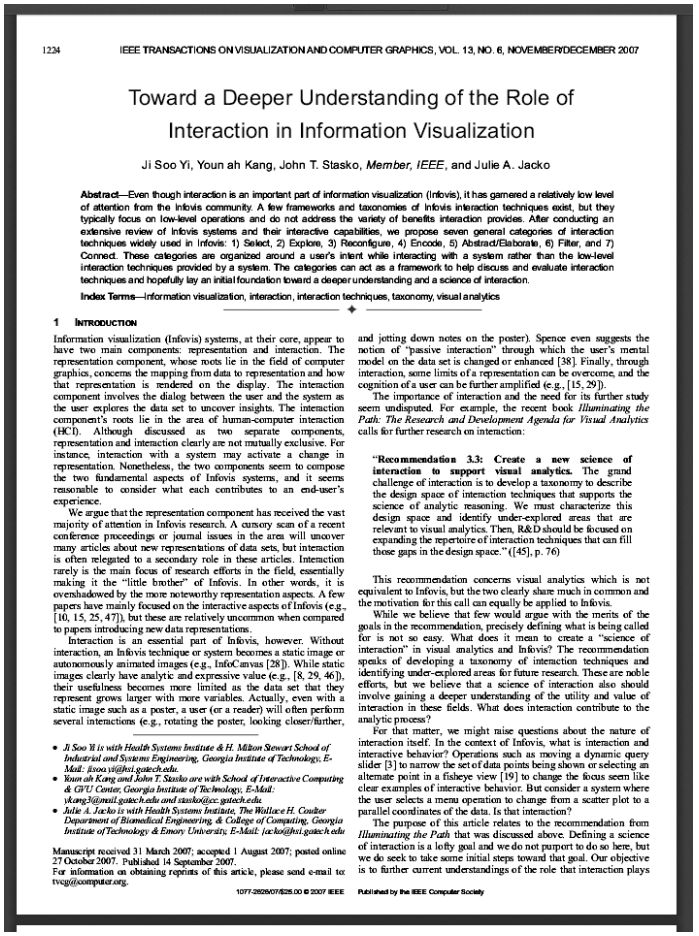
2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction



Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect



1234 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 6, NOVEMBER/DECEMBER 2007

Toward a Deeper Understanding of the Role of Interaction in Information Visualization

Ji Soo Yi, Youn ah Kang, John T. Stasko, *Member, IEEE*, and Julie A. Jacko

Abstract—Even though interaction is an important part of information visualization (Infovis), it has garnered a relatively low level of attention from the Infovis community. A few frameworks and taxonomies of Infovis interaction techniques exist, but they typically focus on low-level operations and do not address the variety of benefits interaction provides. After conducting an extensive review of Infovis systems and their interactive capabilities, we propose seven general categories of interaction techniques widely used in Infovis: 1) Select, 2) Explore, 3) Reconfigure, 4) Encode, 5) Abstract/Elaborate, 6) Filter, and 7) Connect. These categories are organized around a user's intent while interacting with a system rather than the low-level interaction techniques provided by a system. The categories can act as a framework to help discuss and evaluate interaction techniques and hopefully lay an initial foundation toward a deeper understanding and a science of interaction.

Index Terms—Information visualization, interaction, interaction techniques, taxonomy, visual analytics

1 Introduction

Information visualization (Infovis) systems, at their core, appear to have two main components: representation and interaction. The representation component, whose roots lie in the field of computer graphics, concerns the mapping from data to representation and how that representation is rendered on the display. The interaction component involves the dialog between the user and the system as the user explores the data set to uncover insights. The interaction component's roots lie in the area of human-computer interaction (HCI). Although discussed as two separate components, representation and interaction clearly are not mutually exclusive. For instance, interaction with a system may activate a change in representation. Nonetheless, the two components seem to compose the two fundamental aspects of Infovis systems, and it seems reasonable to consider what each contributes to an end-user's experience.

We argue that the representation component has received the vast majority of attention in Infovis research. A cursory scan of a recent conference proceedings or journal issues in the area will uncover many articles about new representations of data sets, but interaction is often relegated to a secondary role in these articles. Interaction rarely is the main focus of research efforts in the field, essentially making it the "little brother" of Infovis. In other words, it is overshadowed by the more noteworthy representation aspects. A few papers have mainly focused on the interactive aspects of Infovis (e.g., [10, 15, 25, 47]), but these are relatively uncommon when compared to papers introducing new data representations.

Interaction is an essential part of Infovis, however. Without interaction, an Infovis technique or system becomes a static image or autonomously animated images (e.g., InfoCanvas [28]). While static images clearly have analytic and expressive value (e.g., [8, 29, 46]), their usefulness becomes more limited as the data set that they represent grows larger with more variables. Actually, even with a static image such as a poster, a user (or a reader) will often perform several interactions (e.g., rotating the poster, looking closer/further, and jotting down notes on the poster). Spence even suggests the notion of "passive interaction" through which the user's mental model on the data set is changed or enhanced [38]. Finally, through interaction, some limits of a representation can be overcome, and the cognition of a user can be further amplified (e.g., [15, 29]).

The importance of interaction and the need for its further study seem undisputed. For example, the recent book *Illuminating the Path: The Research and Development Agenda for Visual Analytics* calls for further research on interaction:

> "**Recommendation 3.3: Create a new science of interaction to support visual analytics.** The grand challenge of interaction is to develop a taxonomy to describe the design space of interaction techniques that supports the science of analytic reasoning. We must characterize this design space and identify under-explored areas that are relevant to visual analytics. Then, R&D should be focused on expanding the repertoire of interaction techniques that can fill those gaps in the design space." ([45], p. 76)

This recommendation concerns visual analytics which is not equivalent to Infovis, but the two clearly share much in common and the motivation for this call can equally be applied to Infovis.

While we believe that few would argue with the merits of the goals in the recommendation, precisely defining what is being called for is not so easy. What does it mean to create a "science of interaction" in visual analytics and Infovis? The recommendation speaks of developing a taxonomy of interaction techniques and identifying under-explored areas for future research. These are noble efforts, but we believe that a science of interaction also should involve gaining a deeper understanding of the utility and value of interaction in these fields. What does interaction contribute to the analytic process?

For that matter, we might raise questions about the nature of interaction itself. In the context of Infovis, what is interaction and interactive behavior? Operations such as moving a dynamic query slider [3] to narrow the set of data points being shown or selecting an alternate point in a fisheye view [19] to change the focus seem like clear examples of interactive behavior. But consider a system where the user selects a menu operation to change from a scatter plot to a parallel coordinates of the data. Is that interaction?

The purpose of this article relates to the recommendation from *Illuminating the Path* that was discussed above. Defining a science of interaction is a lofty goal and we do not purport to do so here, but we do seek to take some initial steps toward that goal. Our objective is to further current understandings of the role that interaction plays

• *Ji Soo Yi is with Health Systems Institute & H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, E-Mail: jisoo.yi@hsi.gatech.edu.*
• *Youn ah Kang and John T. Stasko are with School of Interactive Computing & GVU Center, Georgia Institute of Technology, E-Mail: ykang3@mail.gatech.edu and stasko@cc.gatech.edu.*
• *Julie A. Jacko is with Health Systems Institute, The Wallace H. Coulter Department of Biomedical Engineering, & College of Computing, Georgia Institute of Technology & Emory University, E-Mail: jacko@hsi.gatech.edu*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007. Published 14 September 2007.
For information on obtaining reprints of this article, please send e-mail to tvcg@computer.org.

1077-2626/07/$25.00 © 2007 IEEE Published by the IEEE Computer Society

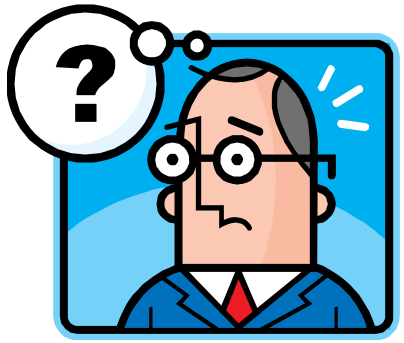
Yi, Kang, Stasko, Jacko
TVCG (InfoVis) '07

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?

Example Tasks and Goals

- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify “hidden” messages or stories in this document collection.
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.

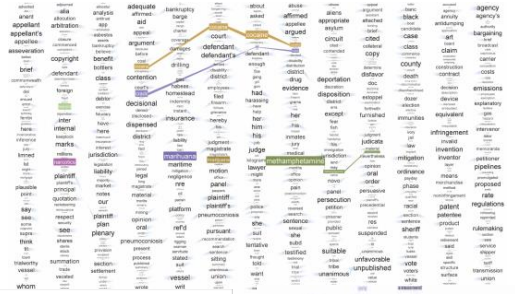
The Challenge

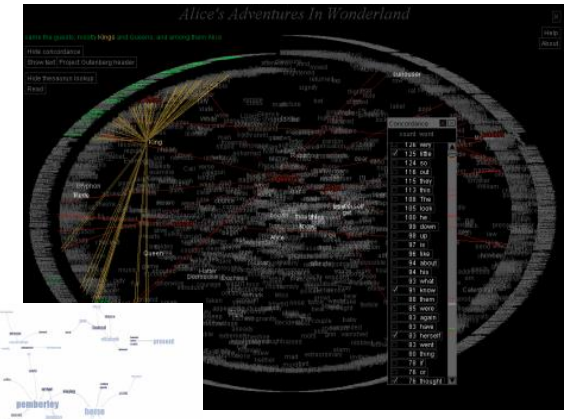
Text is nominal data

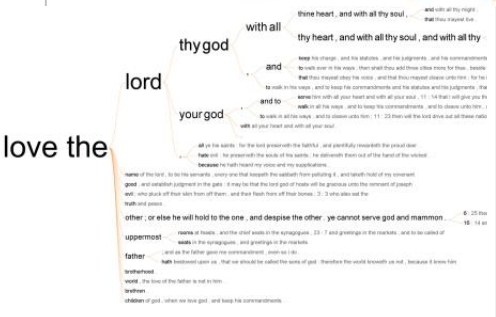
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.









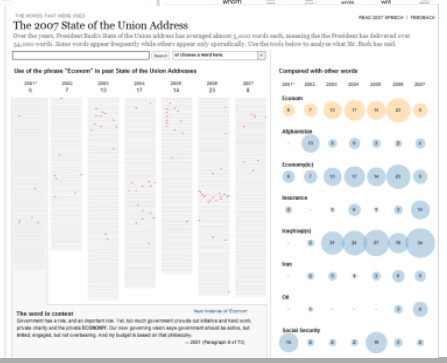
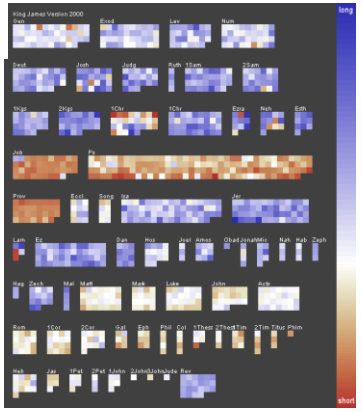
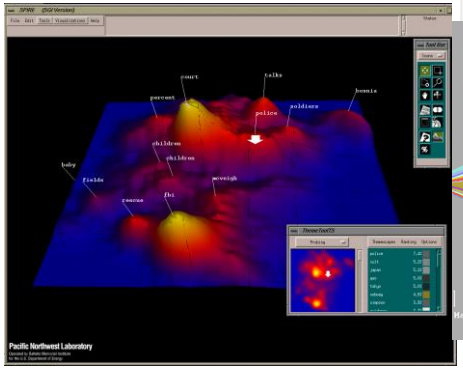
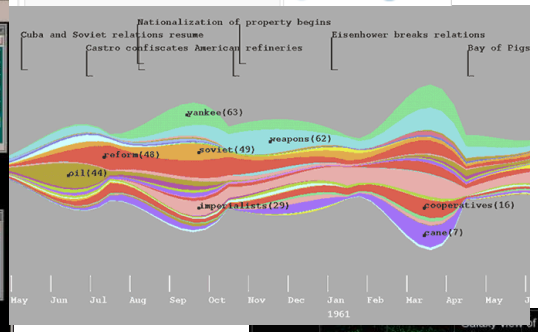


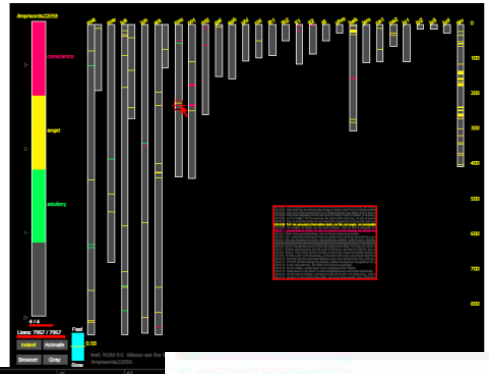


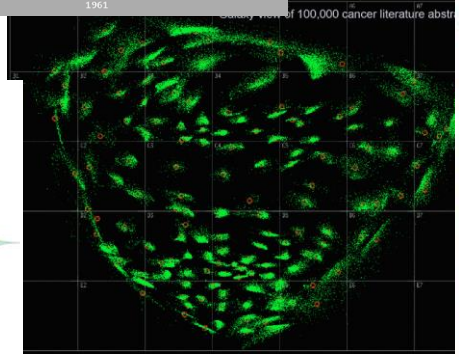
Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with "X and Y" and "X at Y" respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

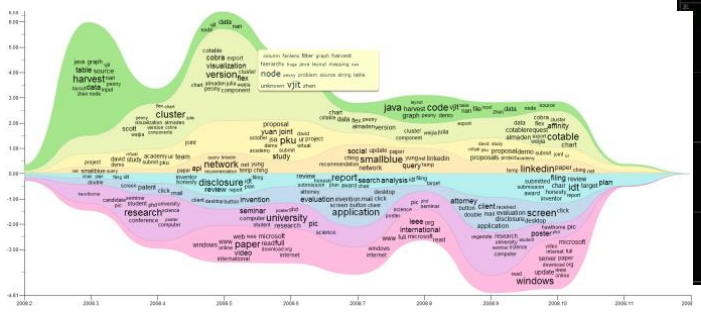




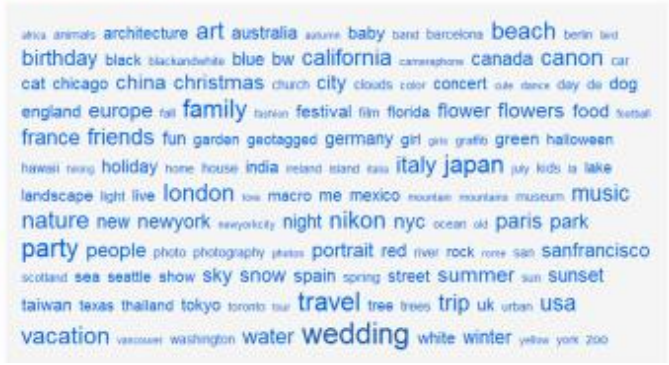








All time most popular tags



“Putting the pieces together”

Jigsaw

Computational analysis of documents’ text

- Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

- Views are highly coordinated



Help “investigators” explore, analyze and understand large document collections

Academic Papers

Health Forums

Product Reviews

Police Reports



Designing and Implementing an Interac
Scatterplot Visualization for a Tablet Con

Ramik Sadana
Georgia Institute of Technology
ramik@gatech.edu

John Stasi
Georgia Institute of T
stasko@cc.gate

ABSTRACT

Tablet computers now offer screen sizes and computing capabilities that are competitive with traditional desktop PCs. Their popularity has grown tremendously, but we are just beginning to see information visualization applications designed for this platform. One potential reason for this limited development is the challenge of designing and implementing a multi-touch interface for visualizations on mobile, tablet devices. In this work, we identify the primary challenges that touch screen interactions pose for information visualization applications. We explore the design space of multi-touch interactions for visualizations and present a prototype information visualization application using a specific technique, a dynamic scatterplot, for an iPad.

Categories and Subject Descriptors
H.5.0 [Information Interfaces and Presentation]: General

General Terms
Design, Algorithms, Human Factors.

Keywords
Information visualization, multi-touch interaction, scatterplot, tablet computer, gesture.

1. INTRODUCTION

The popularity of tablet computers has grown tremendously in recent years. Devices such as the Apple iPad, Amazon Kindle, and Microsoft Surface constitute a significant portion of computer sales today and a wide variety of applications have been tailored to tablet platform. However, one area with just a few initial applications for tablets is information visualization. Though popular commercial systems such as Tableau and Spotfire have introduced tablet versions in the last two years, these systems still feel much like a port of a desktop application. They do not yet leverage the set of gestural interactions that touch-based interfaces provide and do not offer a rich, multi-touch interface in the style of other tablet applications.

We speculate that the limited development of information visualization applications for tablets has resulted from the challenge of designing and implementing the interface. First, most tablets still provide a smaller screen size, which is a limitation for data visualization. Perhaps more importantly, information visualization applications generally have many small visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the authors must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AVI'14, May 27 - 29 2014, Como, Italy
Copyright is held by the authors. Publication rights licensed to ACM.
ACM 978-1-4503-2775-6/14/05...$15.00.

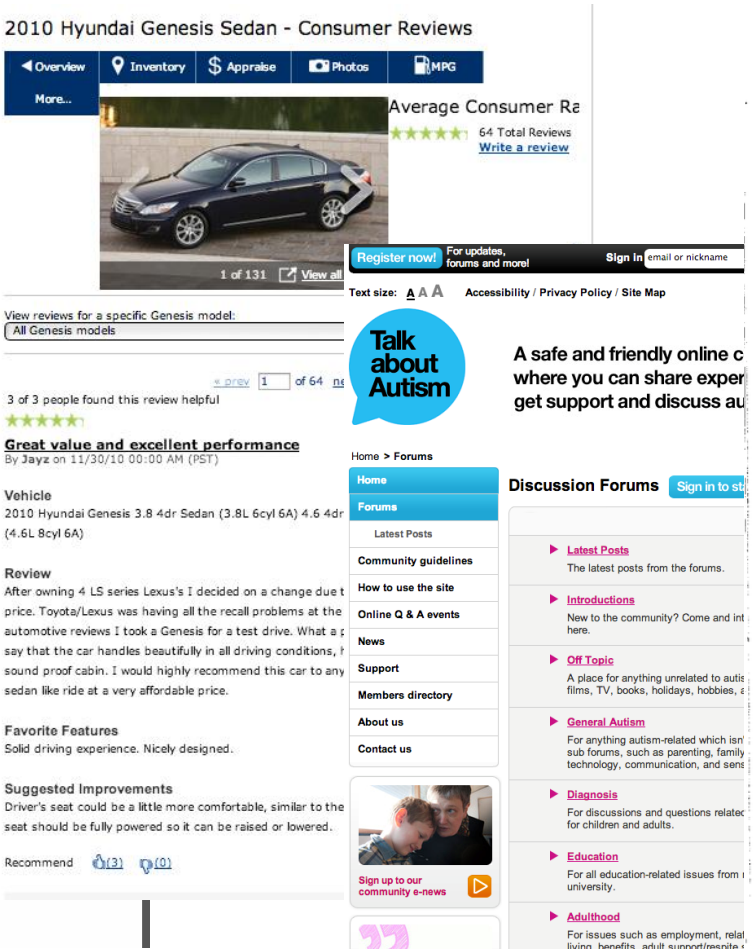
objects to select and manipulate, interactive widgets such as buttons, boxes. Desktop information visualiz make extensive use of a WIMP (w interface. Translating visualization operations to a finger-directed multi keyboard and mouse is simply a very

Some recent research has begun however. One particular avenue development of multi-user, multi applications for tabletop computers [1 explored the design space for s developing for tablet computers is s tabletop computers. Baur et al. [2] rec tablet interface for a StreamGraph-ty This application is much more alig create. Our specific focus, however, used visualization technique. Rzeszot application is a similar technique to driven interactions.

Rather than viewing the movement applications to multi-touch tablets software port, we see this problem challenge. Initial research [25] has interaction design space – for any information visualization syster implementations make sense and Furthermore, we will show that a gestures from tablet applications applied to data visualizations.

In this work, we further explo implement a working information vi tablet PC. We chose a dynamic sca technique of our focus. Scatterplots a in information visualization; many s suite of provided techniques. Scatterp in data visualization, including the FilmFinder system [1] that led to t Our project starts from "ground zero an existing system. Our application advantage of, and to address all the l a tablet interface using only fingers a

The contributions of our work are sections of this article. By studying commercial implementations of a dyn set of tasks and interactive operation would support (Section 3). Next, we the tasks and describe how those c design (Section 4). We ultimately c multi-touch operations for each task, for each, propose multiple potenti many of the options in a proto application for the iPad (Section 5). I paper may not convey all the operati overview summarizing the application



2010 Hyundai Genesis Sedan - Consumer Reviews

Overview | Inventory | Appraise | Photos | MPG | More...

Average Consumer Ra
64 Total Reviews
Write a review

1 of 131 View all

View reviews for a specific Genesis model:
All Genesis models

« prev 1 of 64 ne

3 of 3 people found this review helpful

Great value and excellent performance
By Jayz on 11/30/10 00:00 AM (PST)

Vehicle
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 4dr (4.6L 8cyl 6A)

Review
After owning 4 LS series Lexus's I decided on a change due t price. Toyota/Lexus was having all the recall problems at the automotive reviews I took a Genesis for a test drive. What a p say that the car handles beautifully in all driving conditions, h sound proof cabin. I would highly recommend this car to any sedan like ride at a very affordable price.

Favorite Features
Solid driving experience. Nicely designed.

Suggested Improvements
Driver's seat could be a little more comfortable, similar to the seat should be fully powered so it can be raised or lowered.

Recommend (3) (0)

Register now! For updates, forums and more! Sign in email or nickname

Text size: A A A Accessibility / Privacy Policy / Site Map

Talk about Autism

A safe and friendly online c where you can share exper get support and discuss au

Home > Forums

Home
Forums
Latest Posts
Community guidelines
How to use the site
Online Q & A events
News
Support
Members directory
About us
Contact us

Discussion Forums Sign in to st

Latest Posts
The latest posts from the forums.

Introductions
New to the community? Come and int here.

Off Topic
A place for anything unrelated to autis films, TV, books, holidays, hobbies,

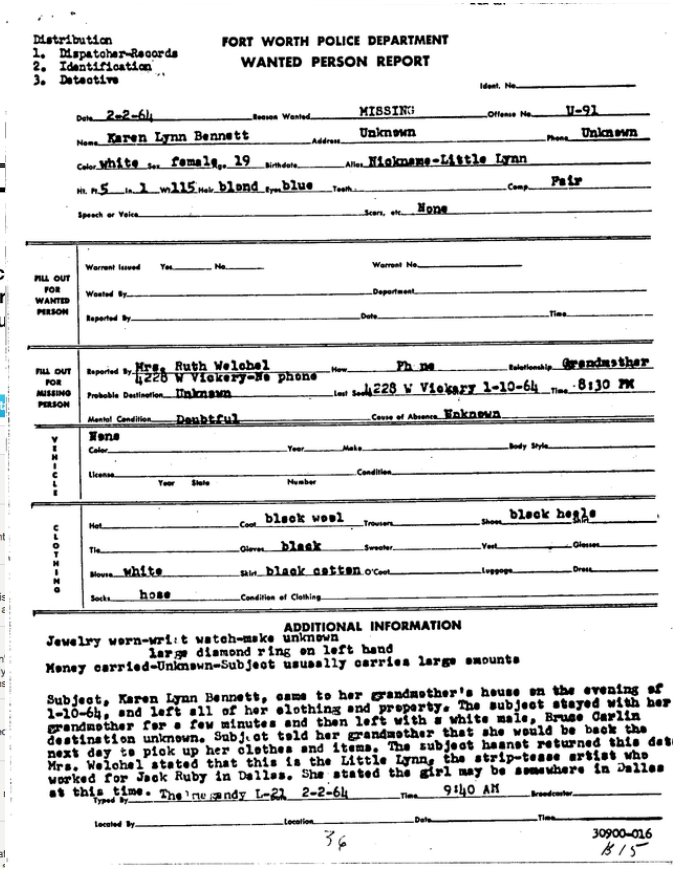
General Autism
For anything autism-related which isn sub forums, such as parenting, family technology, communication, and sens

Diagnosis
For discussions and questions relatec for children and adults.

Education
For all education-related issues from university.

Adulthood
For issues such as employment, relat

Sign up to our community e-news



Distribution
1. Dispatcher-Records
2. Identification
3. Detective

FORT WORTH POLICE DEPARTMENT
WANTED PERSON REPORT

Ident. No.

Date 2-2-64 Reason Wanted MISSING Offense No. U-91
Name Karen Lynn Bennett Address Unknown Phone Unknown
Color White Sex female Age 19 Birthdate Alias Nickname-Little Lynn
Ht. ft. 5 in. 1 Wt. 115 Hair blond Eyes blue Teeth Comp. Fair
Speech or Voice Scars, etc. None

FILL OUT FOR WANTED PERSON
Warrant Issued Yes No Warrant No.
Wanted By Department
Reported By Date Time

FILL OUT FOR MISSING PERSON
Reported By Mrs. Ruth Welchel How Ph me Relationship Grandmother
Probable Destination Unknown Last Seen 4228 W Vickery 1-10-64 Time 8:30 PM
Mental Condition Doubtful Cause of Absence Unknown

VEHICLE
None
Color Year Make Body Style
License Year State Number Condition

CLOTHING
Hat Coat black wool Trousers Shoes black heels
Tie Gloves black Sweater Vest Glasses
Blouse white Skirt black cotton Overcoat Luggage Dress
Socks hose Condition of Clothing

ADDITIONAL INFORMATION
Jewelry worn-wrist watch-make unknown
large diamond ring on left hand
Money carried-Unknown-Subject usually carries large amounts

Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothing and property. The subject stayed with her grandmother for a few minutes and then left with a white male, Bruce Carlin destination unknown. Subject told her grandmother that she would be back the next day to pick up her clothes and items. The subject hasnot returned this dat Mrs. Welchel stated that this is the Little Lynn, the strip-tease artist who worked for Jack Ruby in Dallas. She stated the girl may be somewhere in Dallas at this time. The investigation L-21 2-2-64 Time 9:40 AM

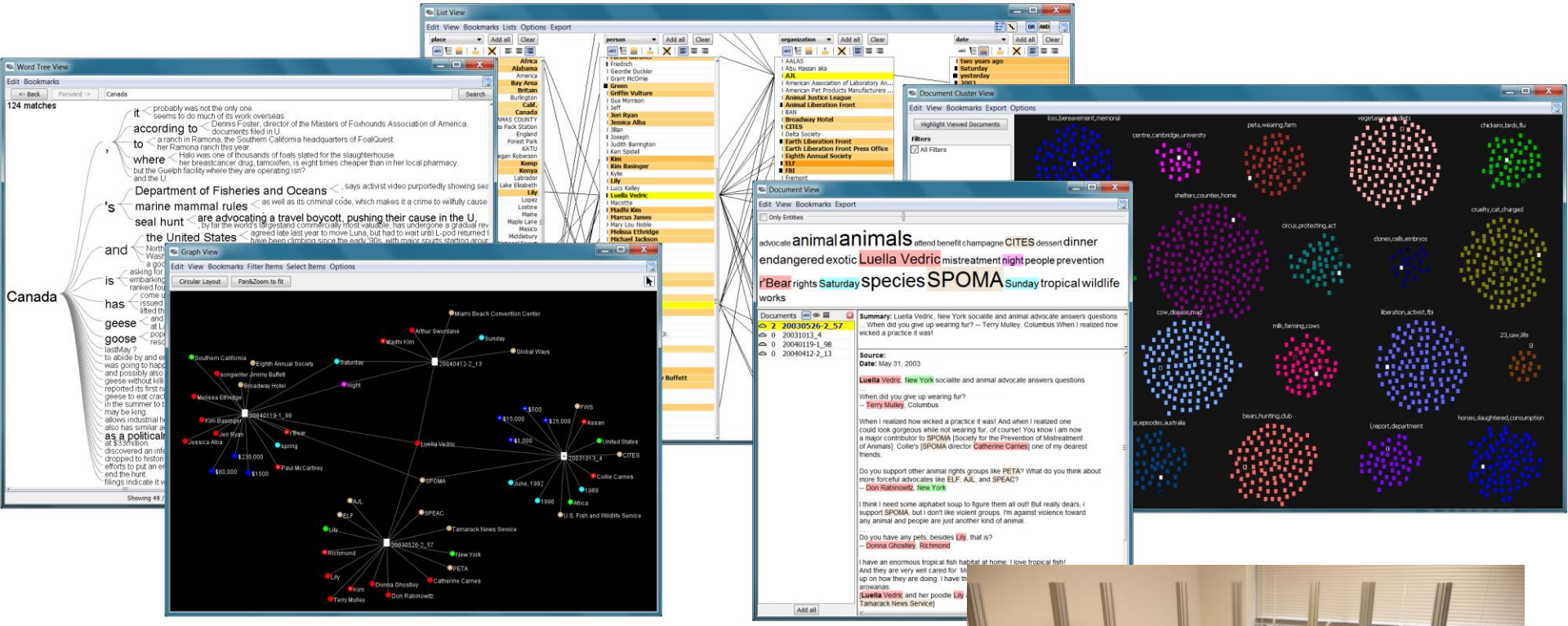
Typed By
Located By Location Date Time

36

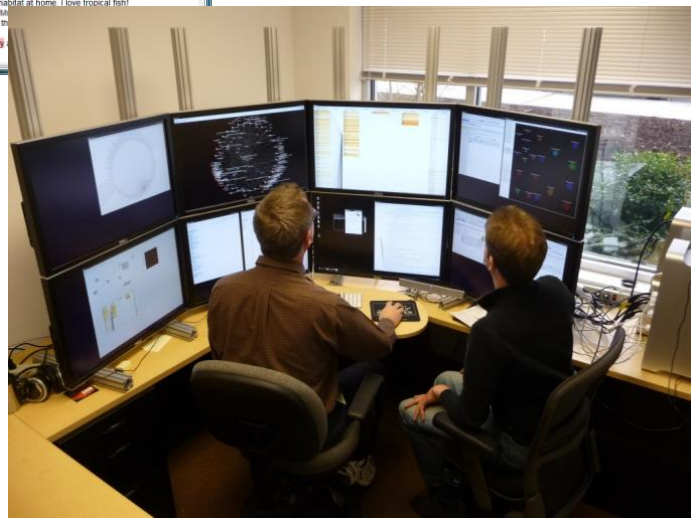
30900-016
8/5

http://www.cc.gatech.edu/gvu/ii/jigsaw

Jigsaw

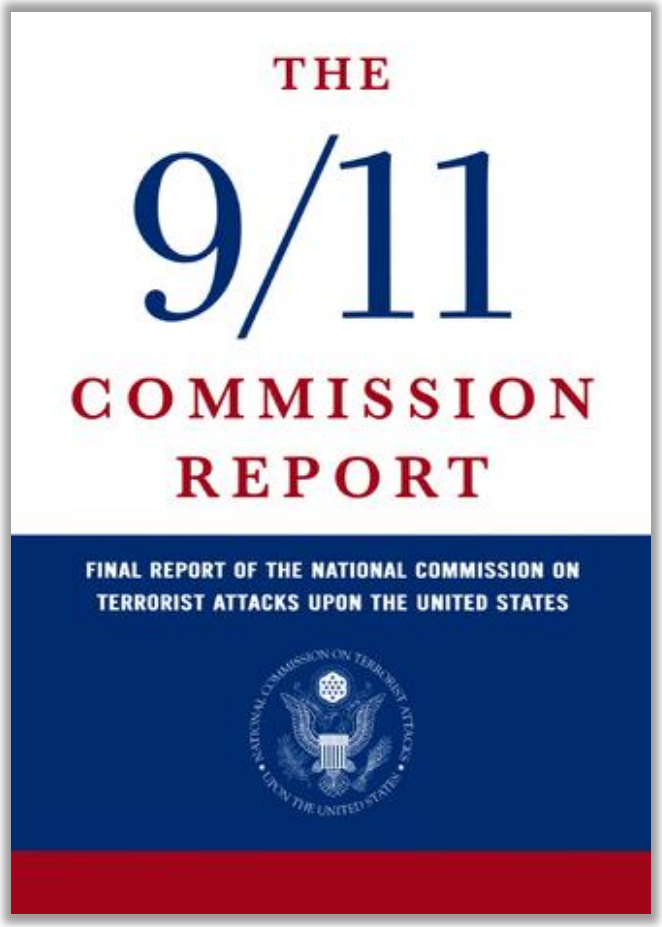


Visual analytics of large text document collections



Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG ‘13

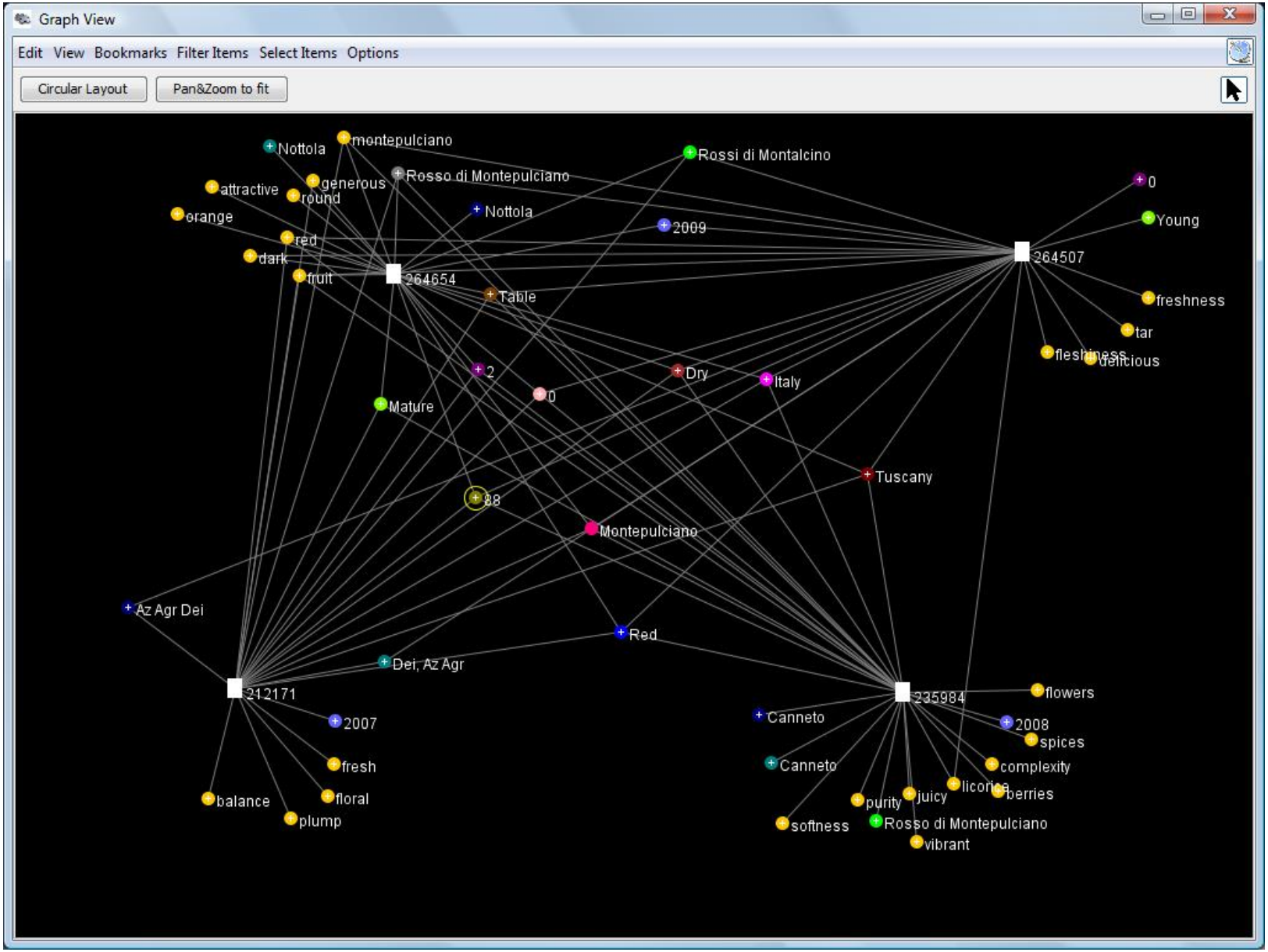
Example



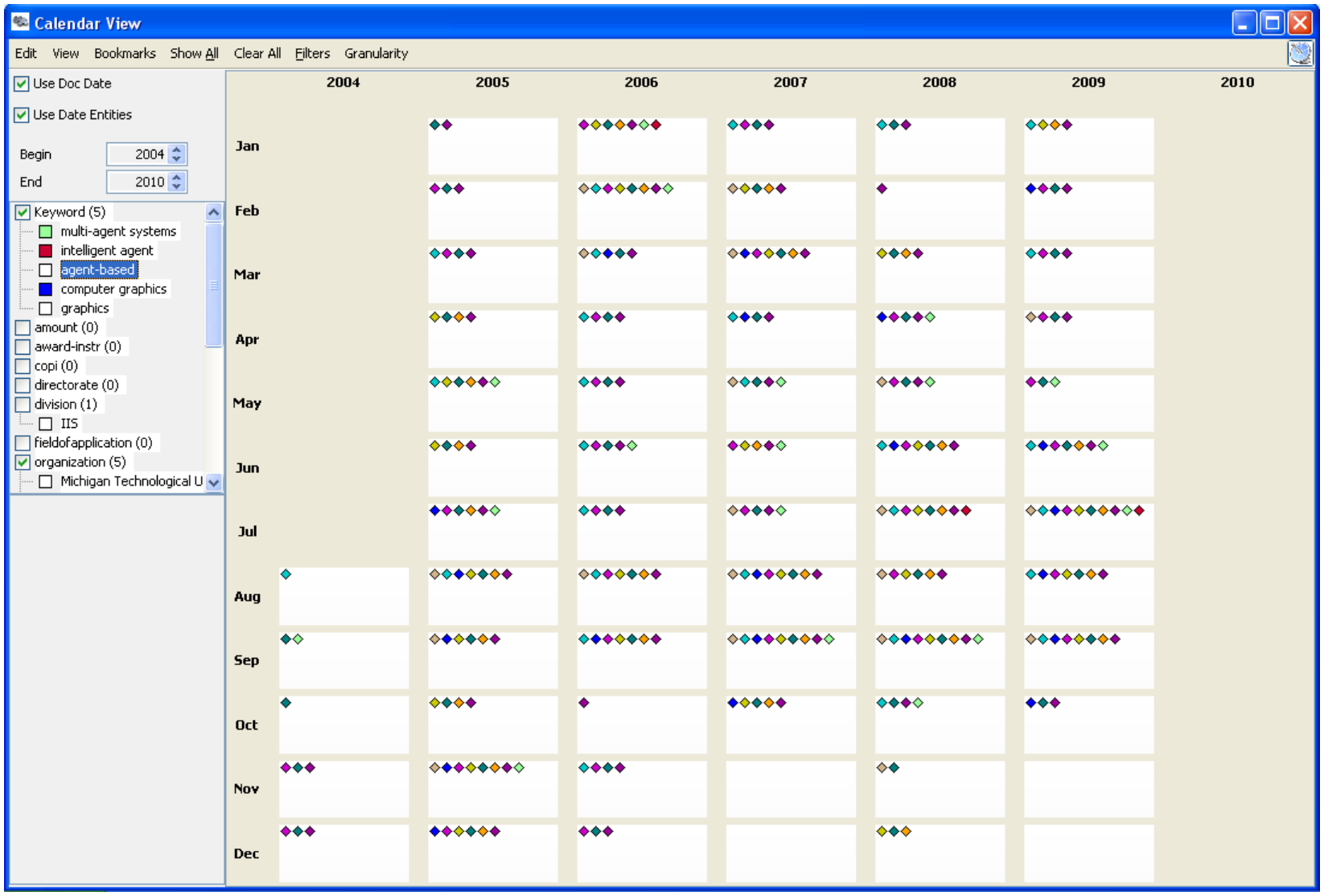
Made each page into a separate “document”
585 in total
Entities: Person, Location, Organization, Date, Money

Demo

Graph View

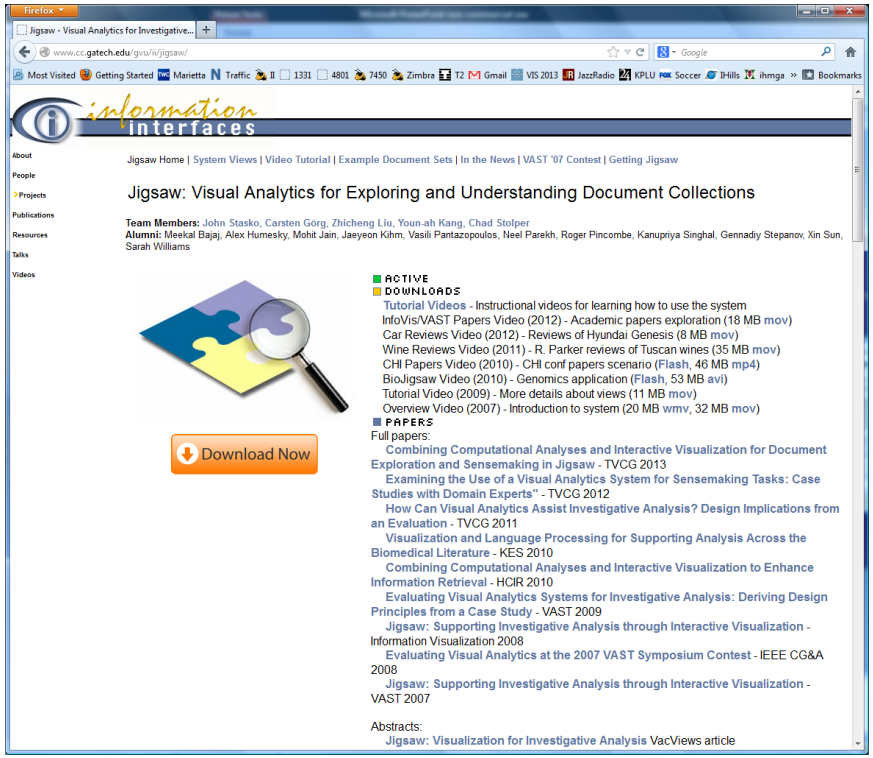

Graph View
Edit View Bookmarks Filter Items Select Items Options
Circular Layout
Pan&Zoom to fit
Nottola
montepulciano
Rossi di Montalcino
Rosso di Montepulciano
generous
attractive
round
orange
Nottola
2009
red
dark
fruit
264654
Table
0
Young
264507
freshness
tar
fleshiness
delicious
2
Dry
Italy
0
Mature
Tuscany
88
Montepulciano
Az Agr Dei
Red
Dei; Az Agr
212171
235984
flowers
2007
Canneto
2008
spices
fresh
Canneto
complexity
licorice
berries
balance
floral
purity
juicy
plump
softness
Rosso di Montepulciano
vibrant

Calendar View



Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw





Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

Further benefits of visualization

Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

EuroVis ‘14 Capstone Talk



http://vimeo.com/98986594



Take Aways

Presentation & analysis

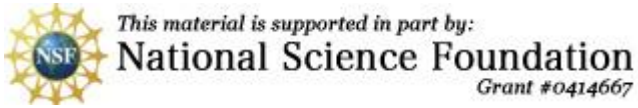
Interaction provides the power

Exploring & developing questions

Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

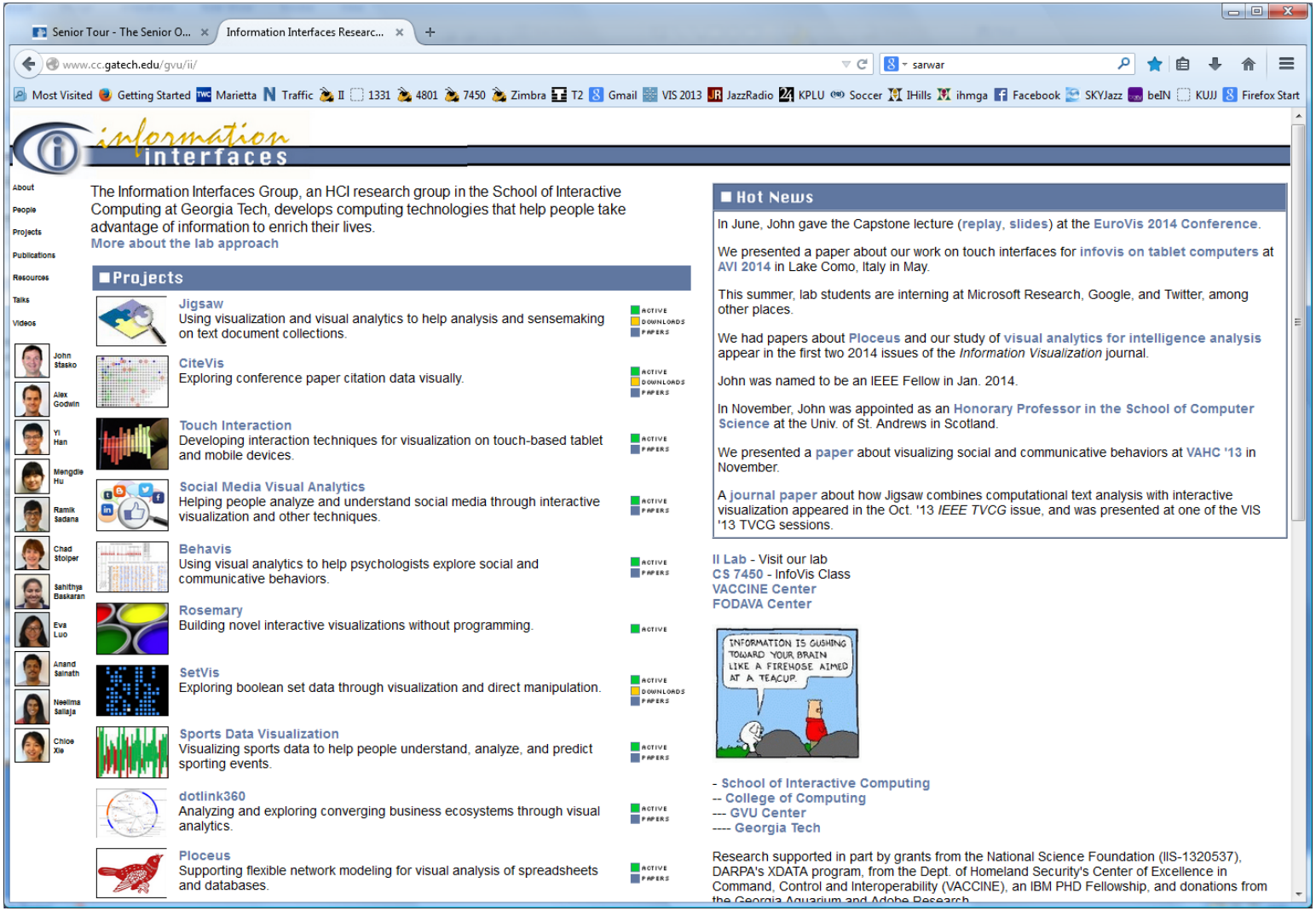
- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA Program

My Research

http://www.cc.gatech.edu/gvu/ii



Value-Driven Evaluation of Visualizations

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu







“No user studies are reported in this paper,
thus it is unclear if the visualization will be useful.”

Really?

- Which cars have the best miles-per-gallon?
- How much does a Ford Taurus cost?
- Is there a correlation between car weight and torque?
- How many different countries manufacture cars?
- What is the range of car’s horsepower?

~~Evaluation~~

Explanation(s) and Illustration(s) of value

Visualization's Value?

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)

Visualization's Value?

$$V_{alue} = T + \mathbf{I} + E + C$$

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to convey an overall **essence** or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)

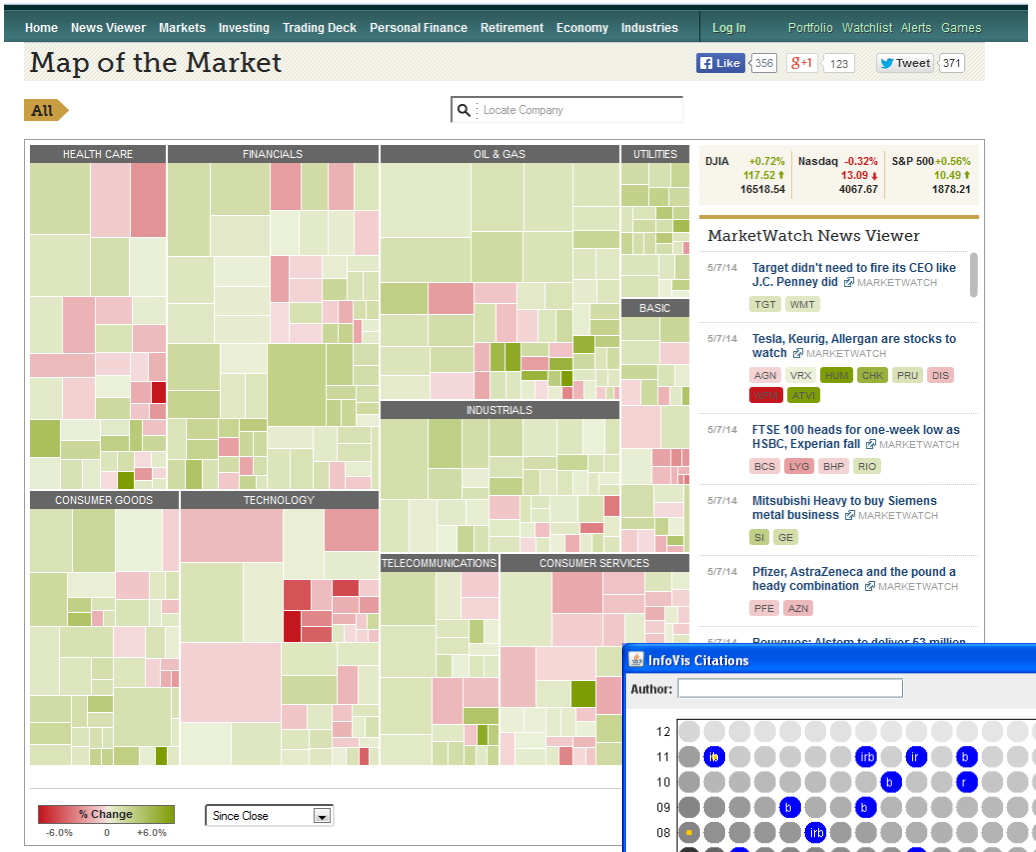
Visualization's Value?

$$Value = T + I + E + C$$

Ability to generate **confidence** and trust about the data, its domain and context

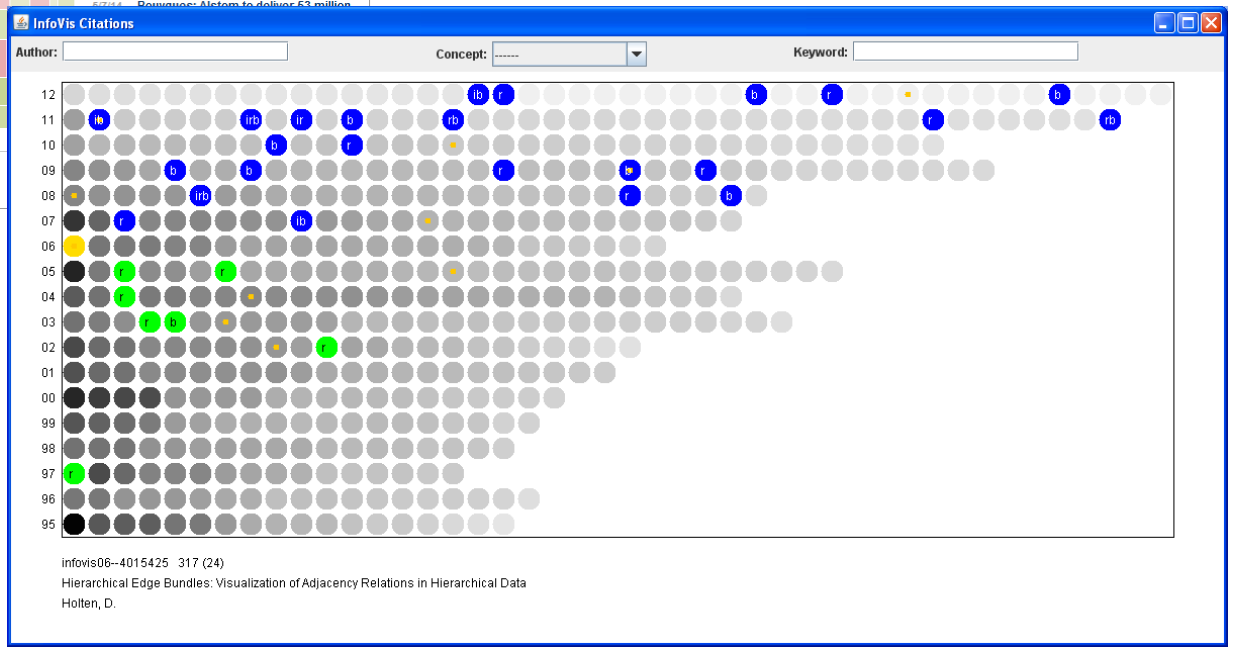
(Beneficial data analysis process side effects)

http://www.marketwatch.com/tools/stockresearch/marketmap



Map of the Market

CiteVis



http://www.cc.gatech.edu/gvu/ii/citevis

Data Visualization 101



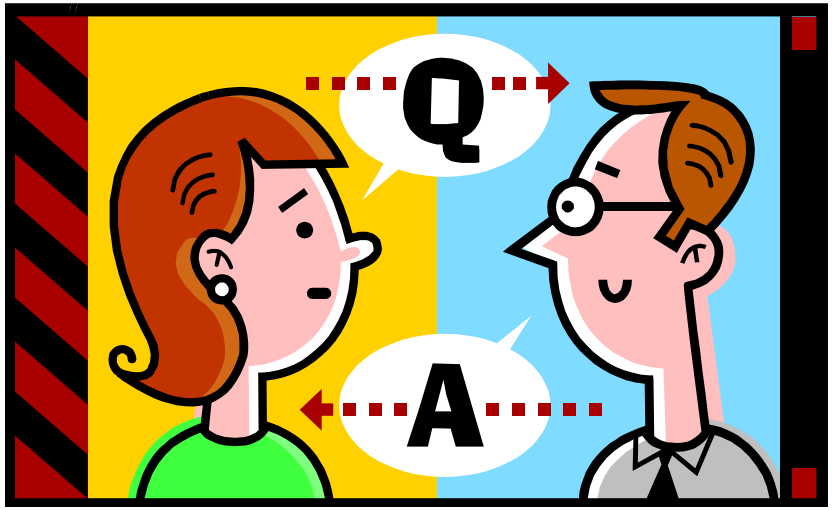
Problem:
You have a lot of data (& attributes) to understand

Do you?

- Pack all the data into one complex representation
- Spread the data into multiple coordinated views
- Use interaction to reveal different subsets of the data

Interaction is Vital

Engage in a dialog with your data



"The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."

S. Few, *Now you see it*

Interaction

Understand the benefits (science of interaction)

Explore new interactive operations

Visualization on other display devices (touch)

http://vimeo.com/98986594

For more, see my EuroVis ‘14 Capstone Talk



Acknowledgments

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)



- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA program

July 28, 2014

Visual Analytics for Investigative Analysis and Exploration of Documents and Data

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu





Georgia Tech College of Computing
School of Interactive Computing

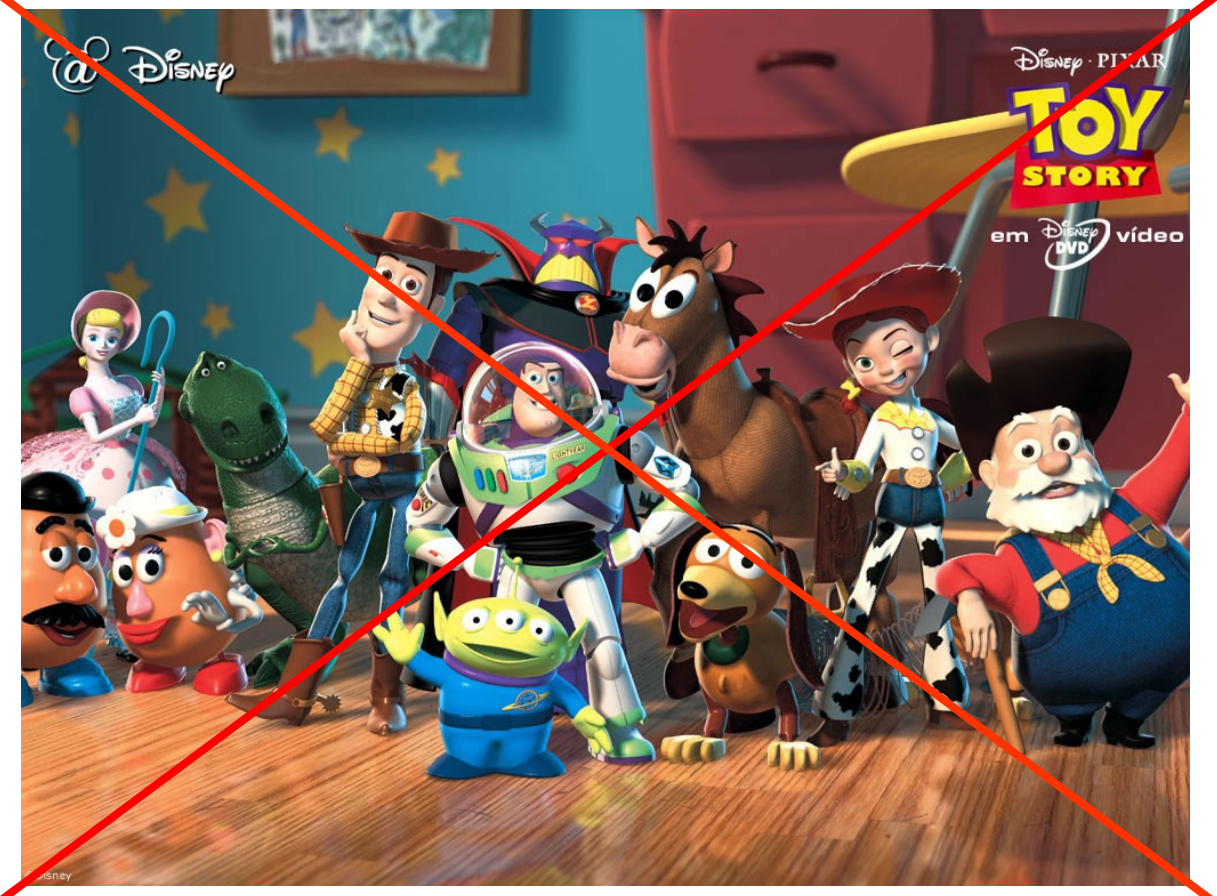
Data Visualization

Making pretty pictures?



Data Visualization

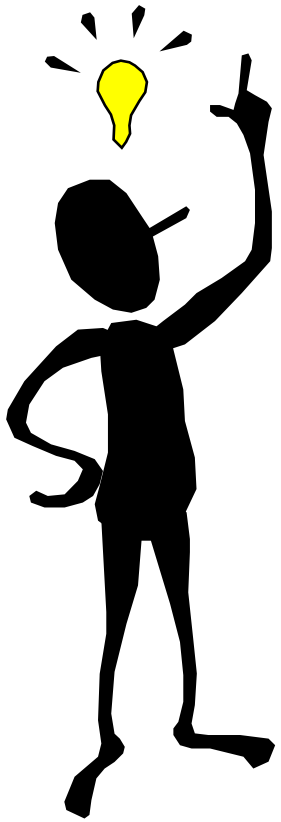
~~Making pretty pictures~~



Data Visualization

A cognitive process

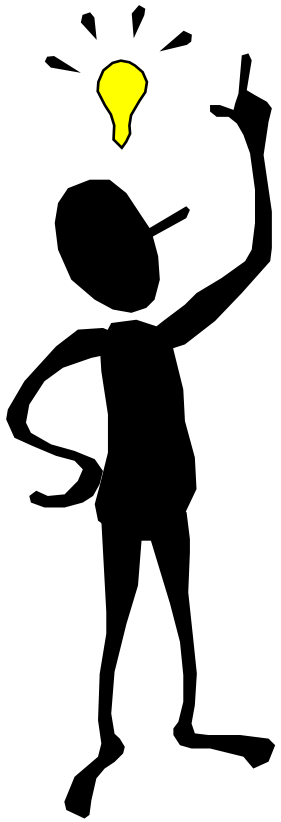
Gain an understanding

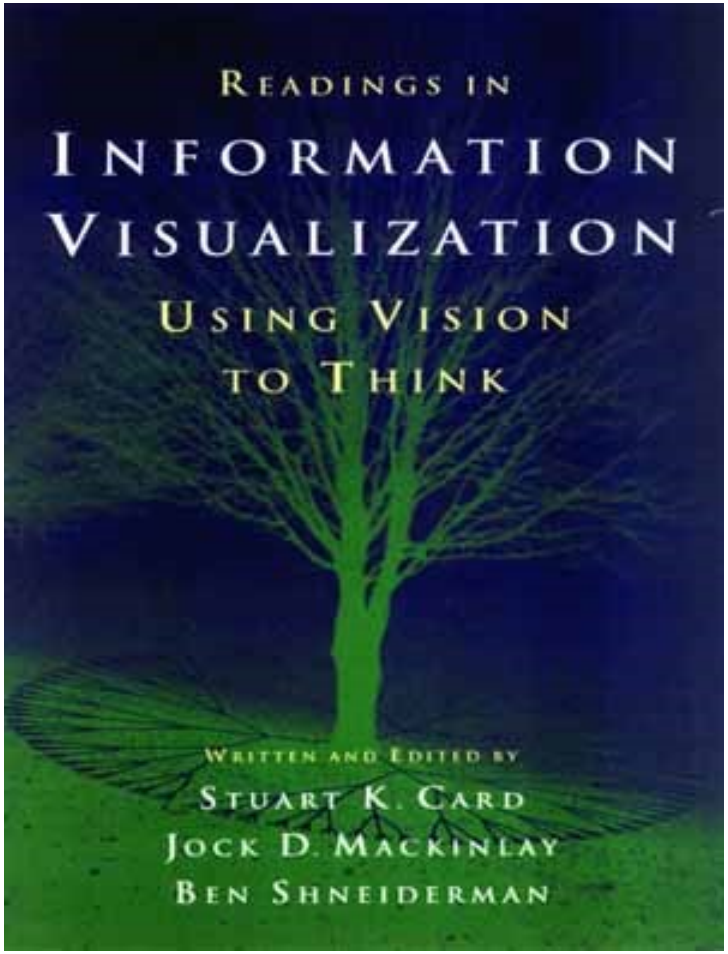


Data Visualization

A cognitive process

Gain an **understanding**





Visualization

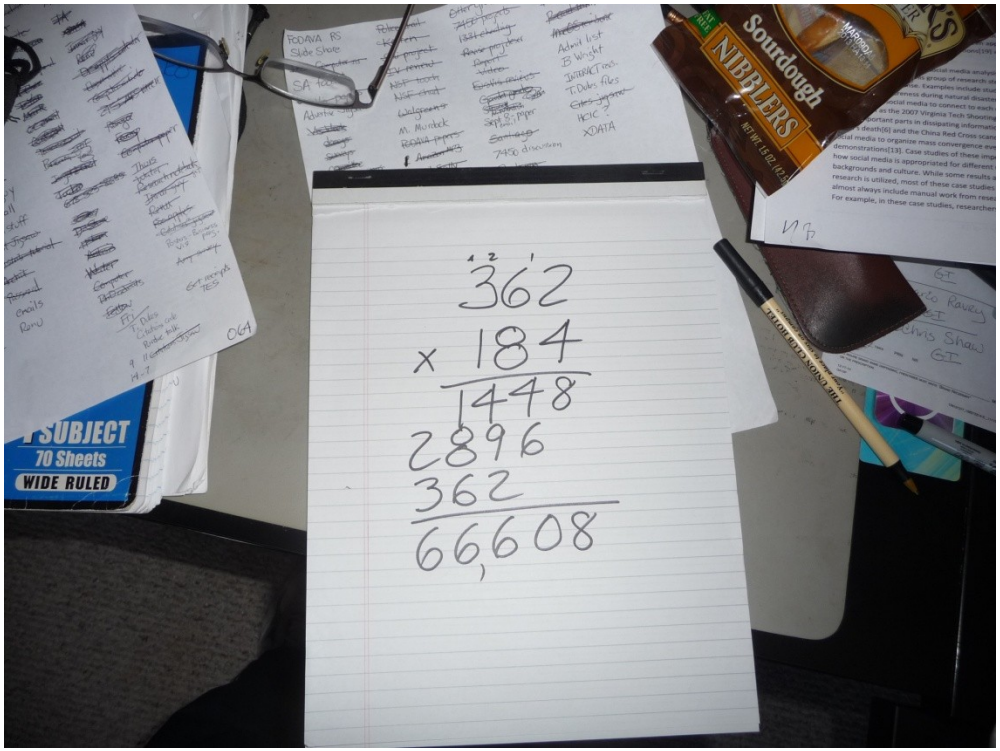
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provide a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Applications of Visualization

Presentation

Analysis

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support

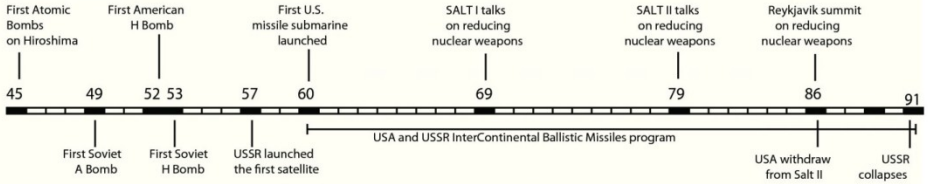


Infographics

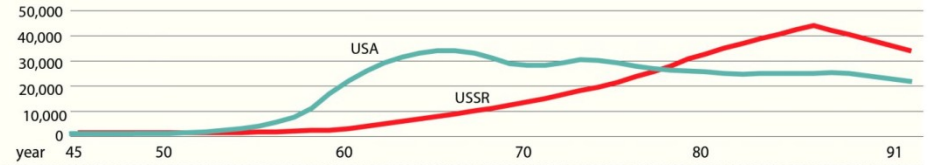
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



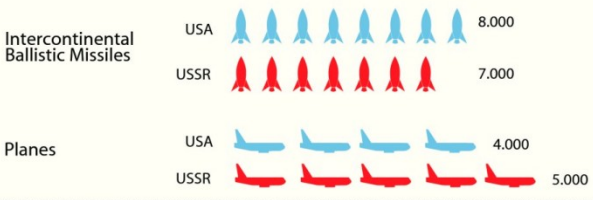
NUCLEAR STOCKPILE



ROCKET MODELS



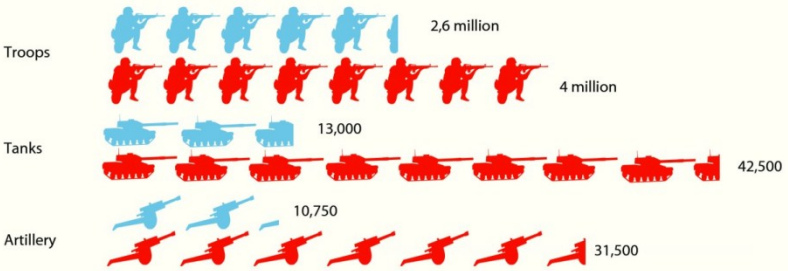
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



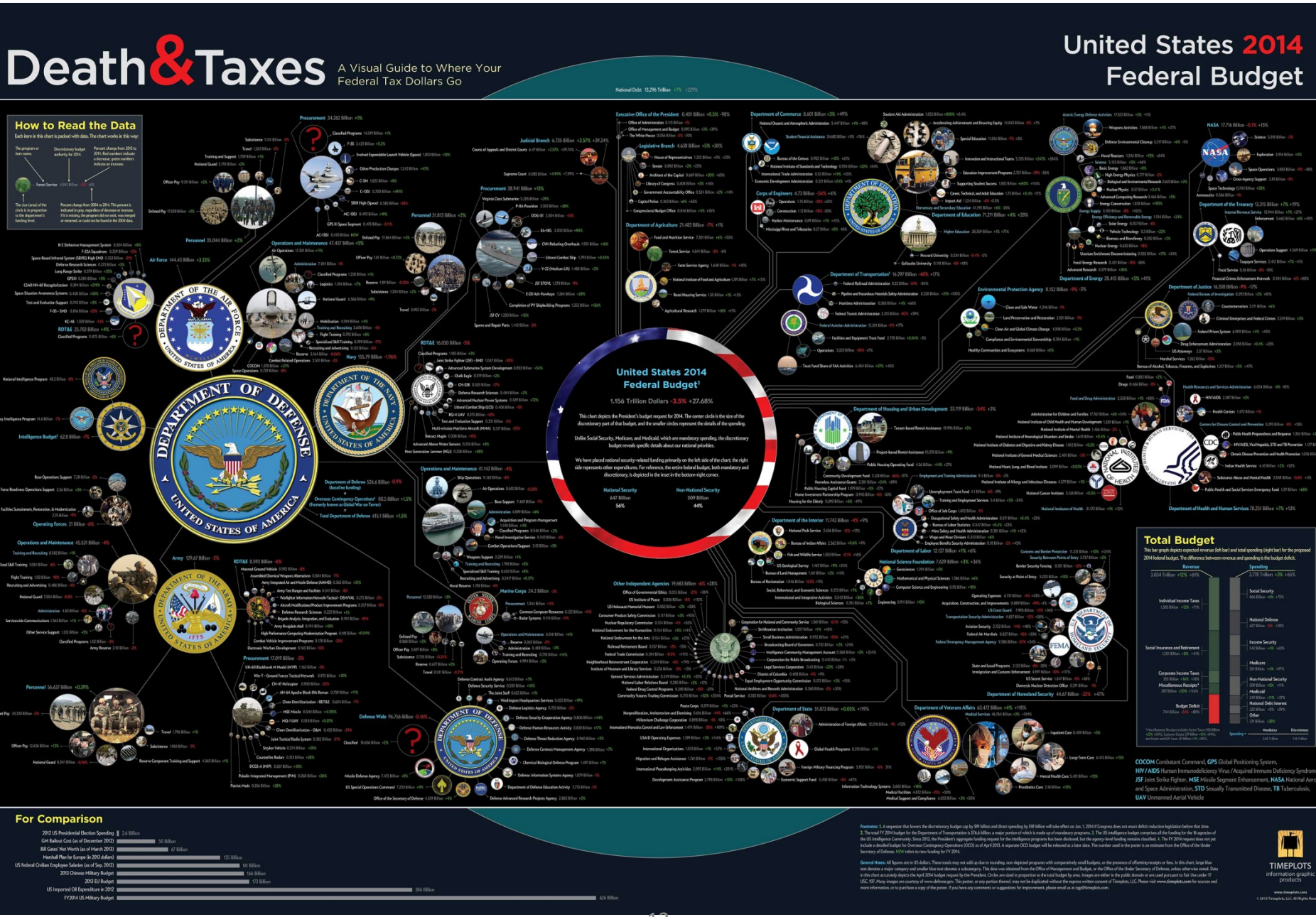


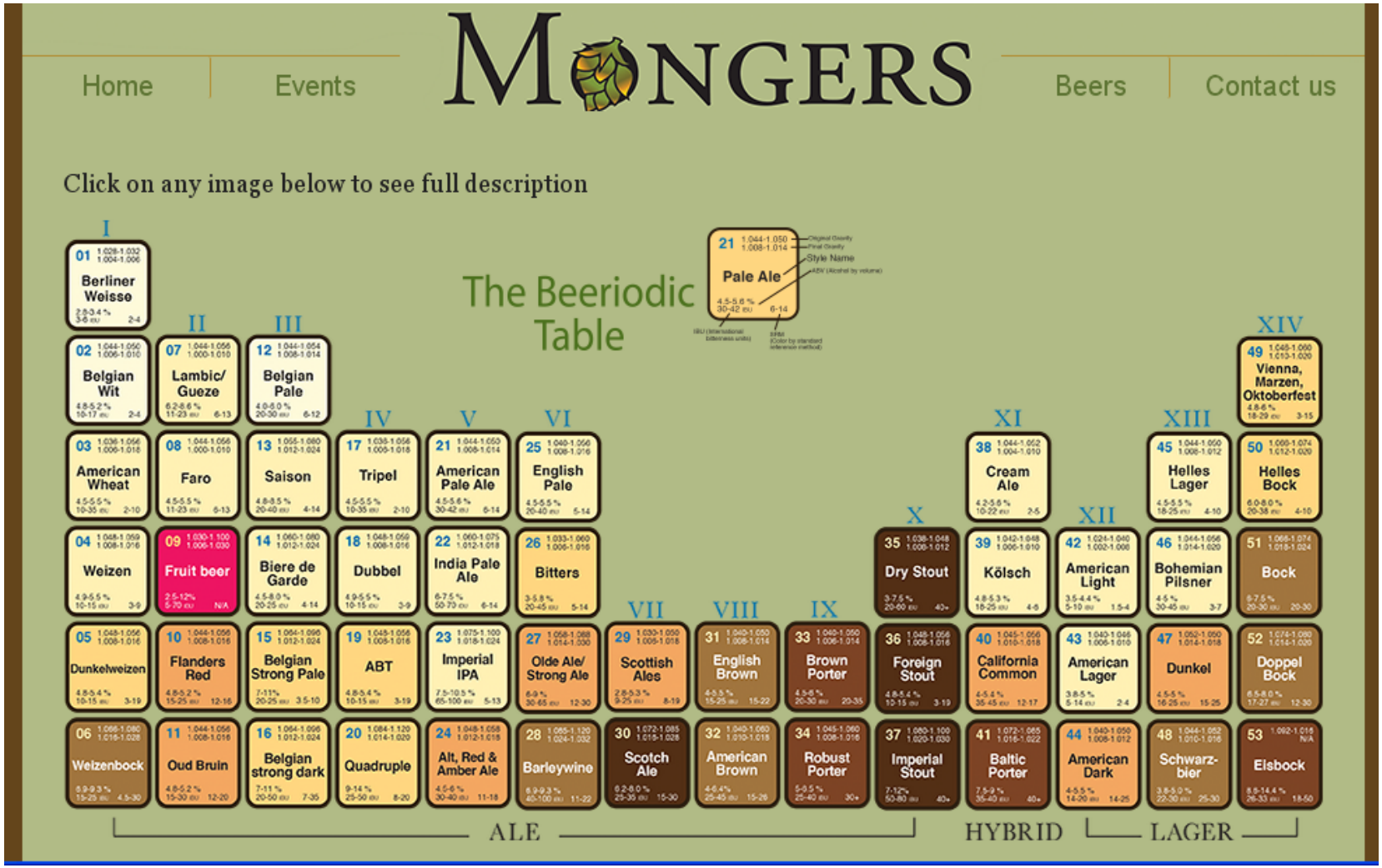
Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget





http://thebeermongers.com/beers/

NY Times 2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons
Bonds was injured last season. He played 14 games and hit 6 homers

600
400
200
0
14th season
5 seasons
10
15

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

Aaron
Actual homers slightly outpace projected homers for five seasons.



Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

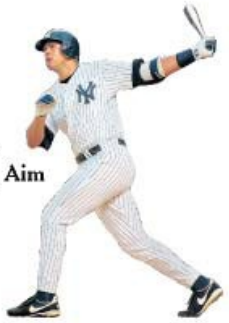


Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

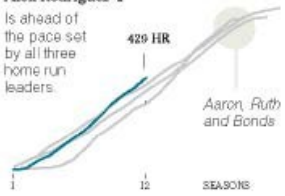


Note: Ages as of July 1 of each season.

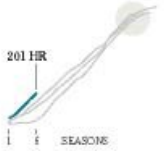
Others Taking Aim



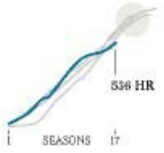
Alex Rodriguez
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

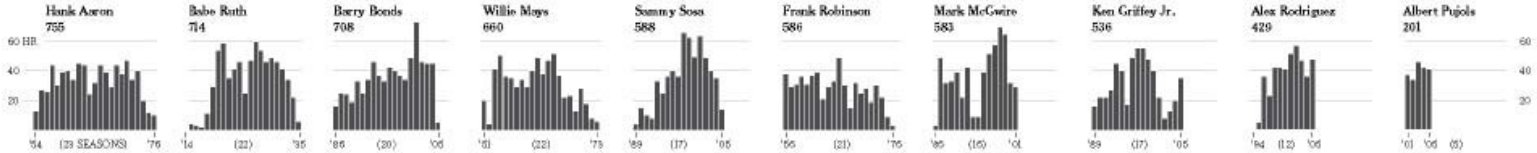


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

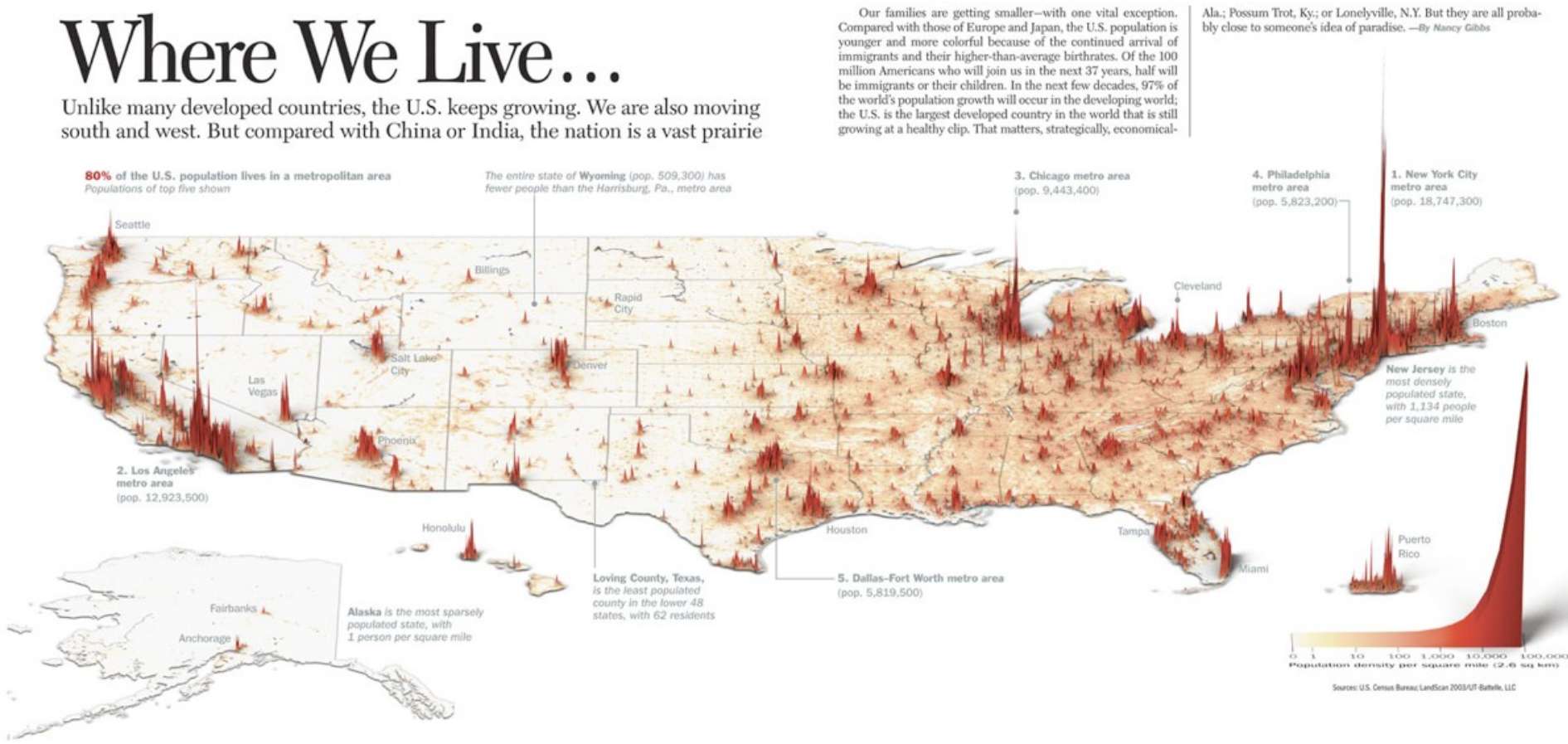


Differing Paths to the Top of the Charts

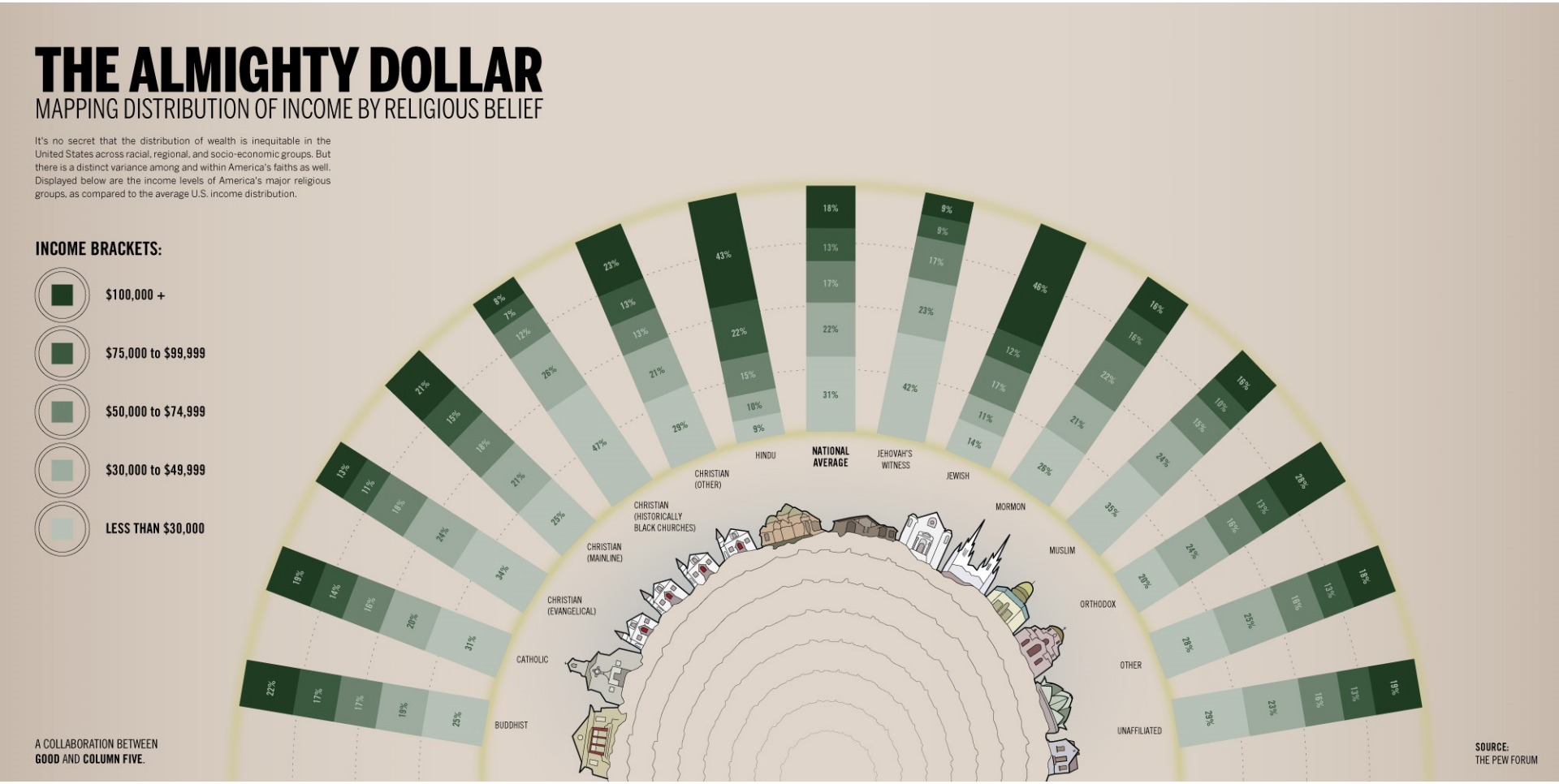
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



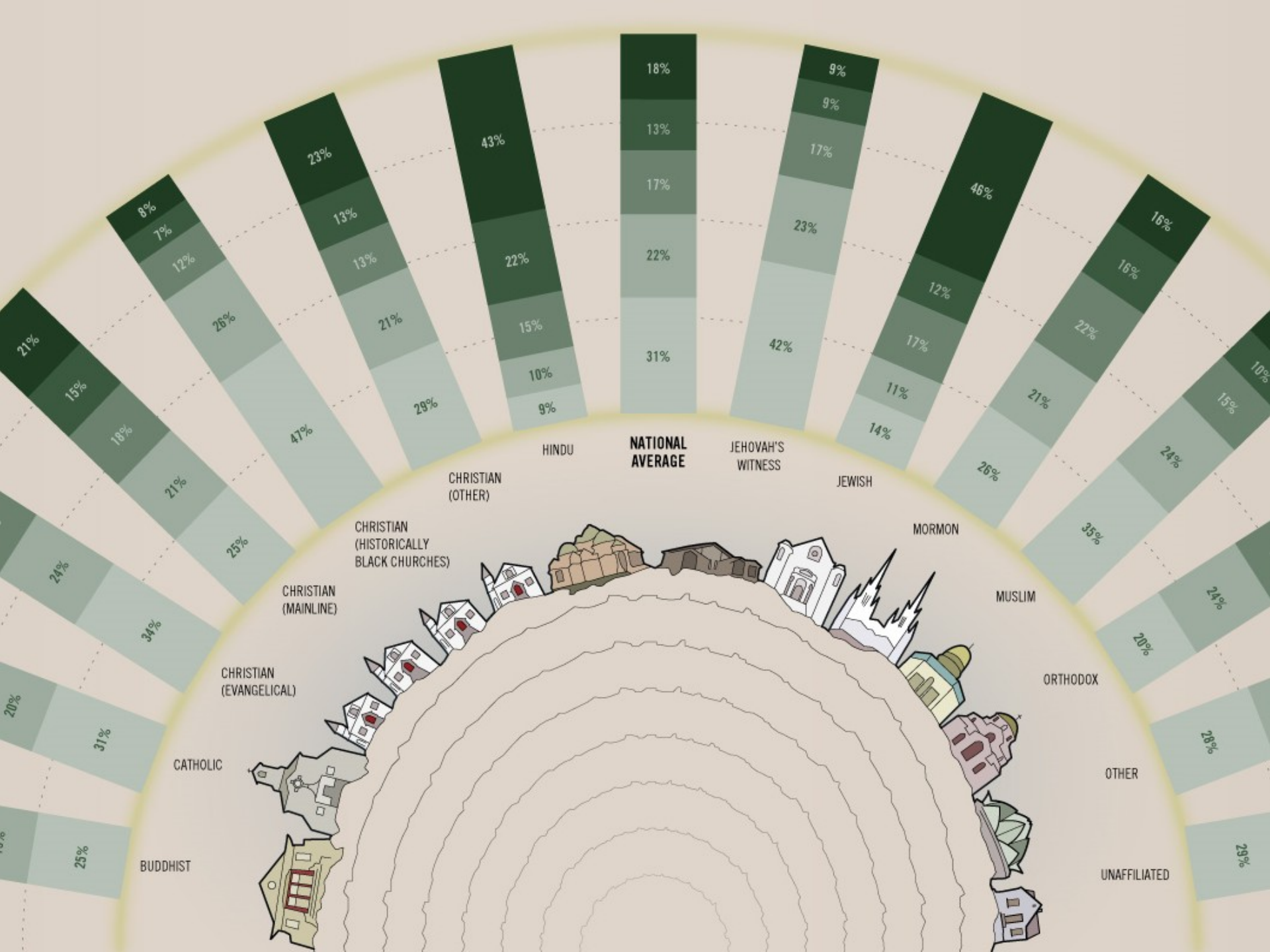
Amanda Cox and Joe Ward/The New York Times



http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html

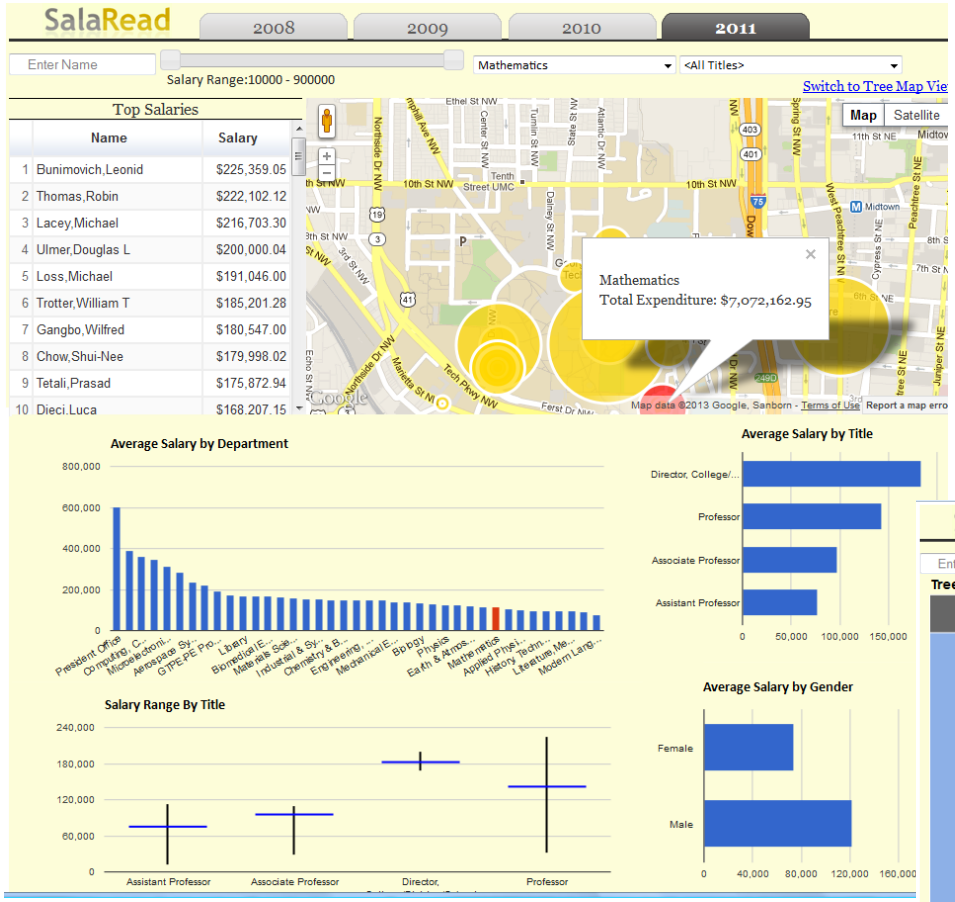


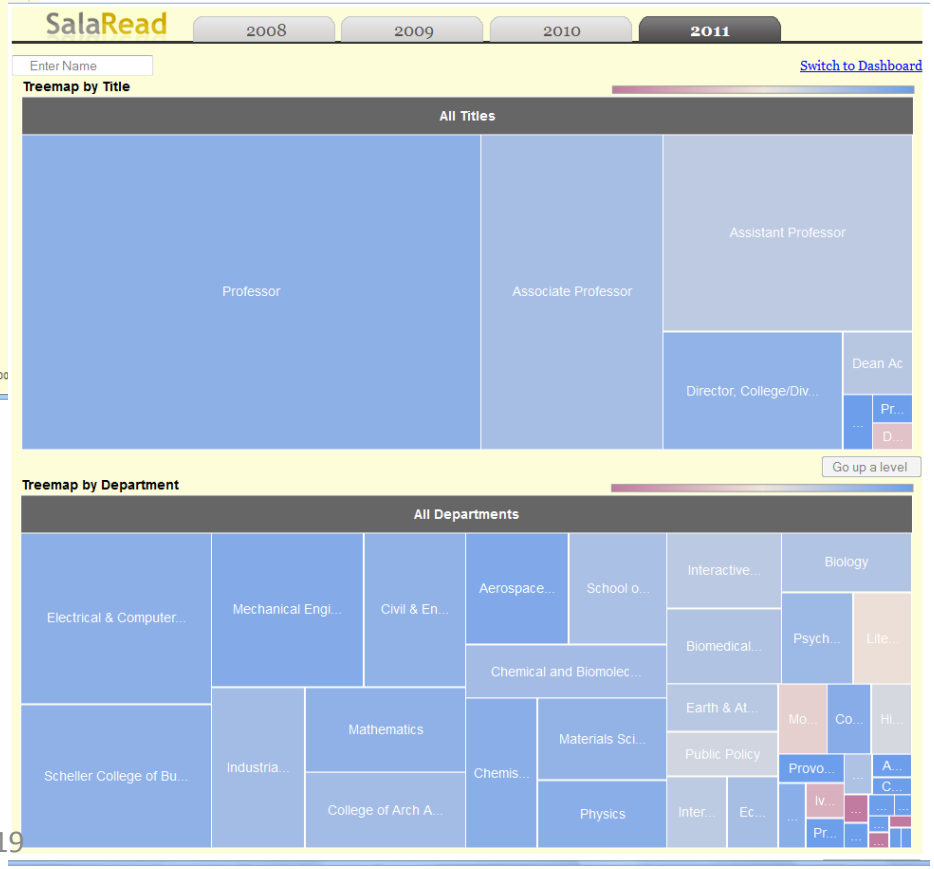
http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg


BUDDHIST
21%
15%
18%
21%
25%
CATHOLIC
20%
31%
24%
34%
CHRISTIAN (EVANGELICAL)
8%
7%
12%
26%
47%
CHRISTIAN (MAINLINE)
23%
13%
13%
21%
29%
CHRISTIAN (HISTORICALLY BLACK CHURCHES)
43%
22%
15%
10%
9%
CHRISTIAN (OTHER)
18%
13%
17%
22%
31%
HINDU
NATIONAL AVERAGE
9%
9%
17%
23%
42%
JEHOVAH'S WITNESS
46%
12%
17%
11%
14%
JEWISH
16%
16%
22%
21%
26%
MORMON
10%
15%
24%
35%
MUSLIM
20%
24%
28%
ORTHODOX
29%
OTHER
UNAFFILIATED

Simply presenting data ***visually*** can have a profound impact

CS 7450
Fall '12
Design project

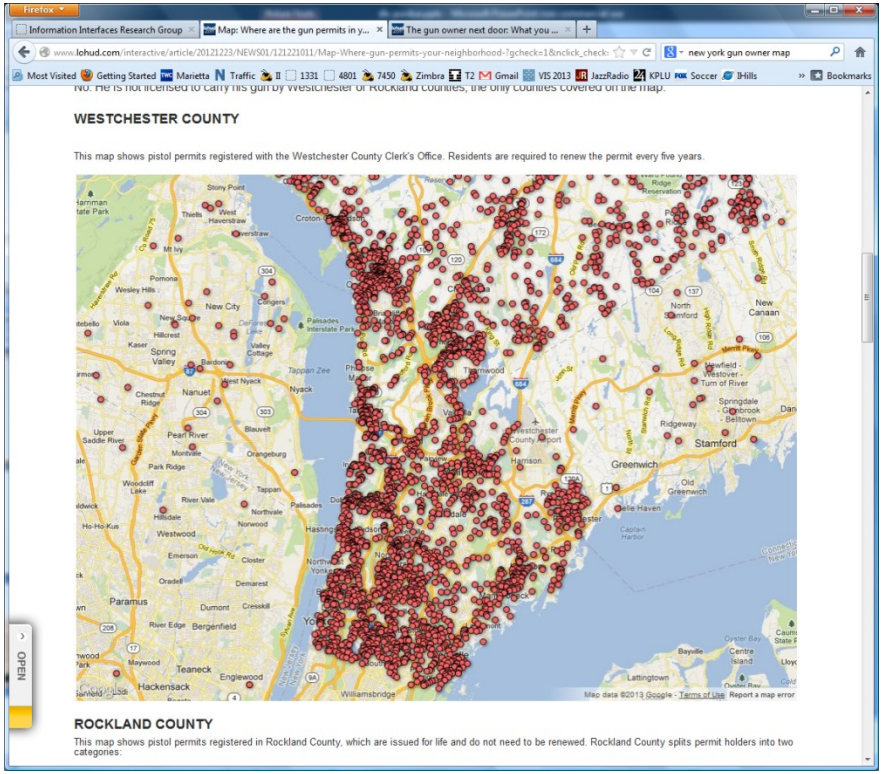




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun ownership in New York counties





http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood-?gcheck=1&nclick_check=1

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

May just show a few variables
and/or a subset of the data cases

2. Analysis

understand, compare, decide, judge, evaluate, assess, determine, …



Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

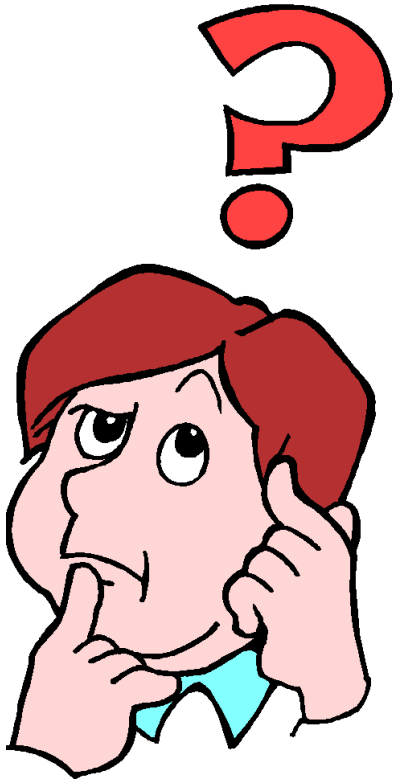
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning

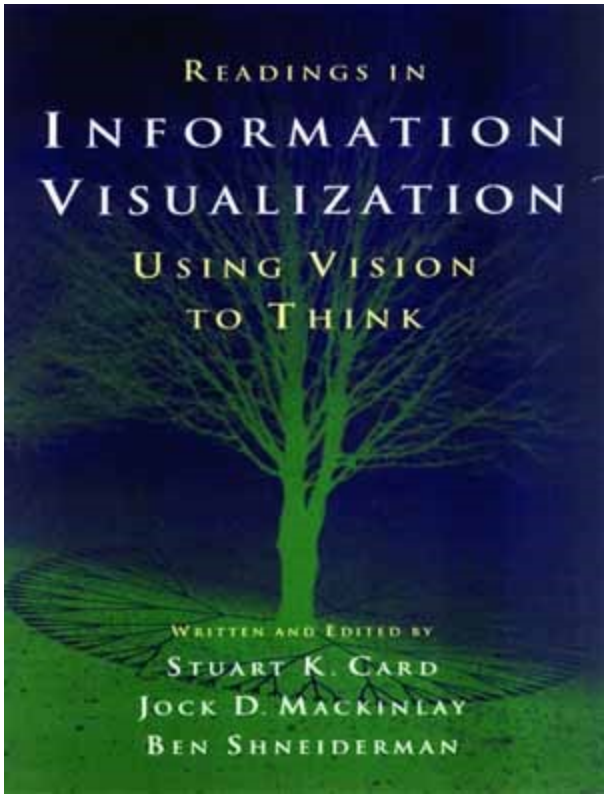




Use them when they get the job done!

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction



Visualization

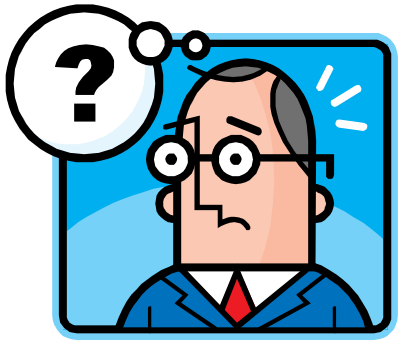
"The use of computer-supported, **interactive** visual representations of data to amplify cognition"

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Text is Everywhere

We use documents as primary information artifacts in our lives

Our access to documents has grown tremendously in recent years due to the networking infrastructure

- WWW
- Digital libraries
- …

Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?

Example Tasks and Goals

- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify “hidden” messages or stories in this document collection.
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.

The Challenge

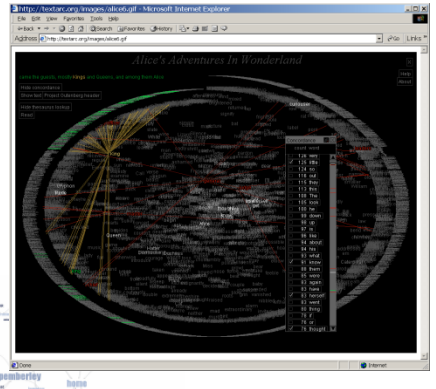
Text is nominal data

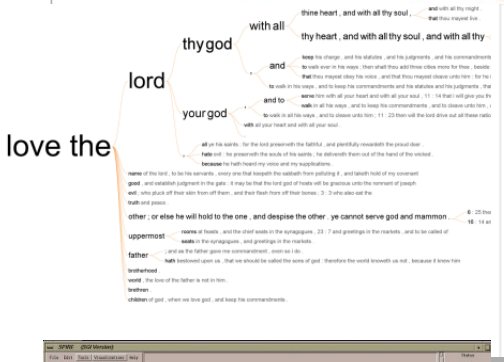
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.









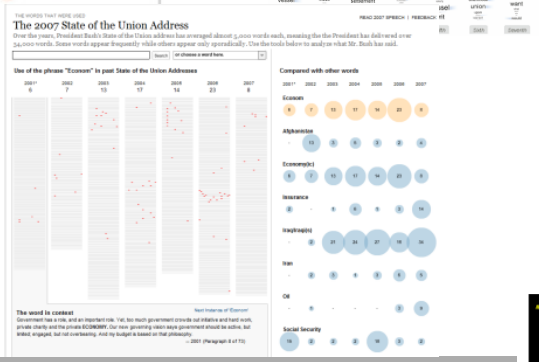
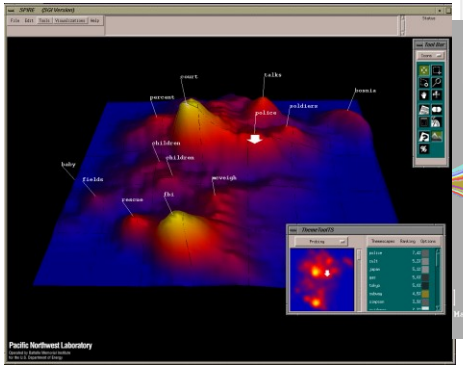
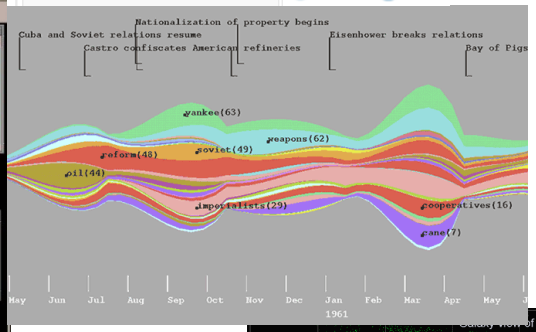
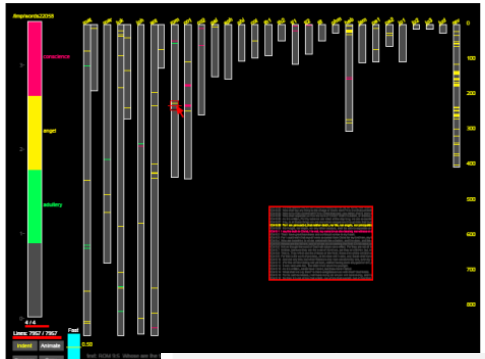


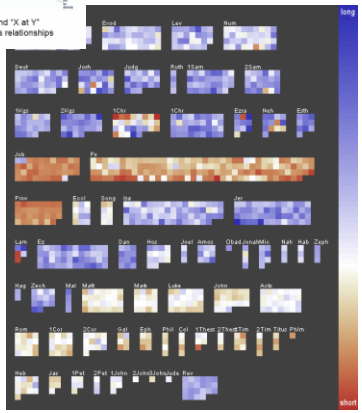


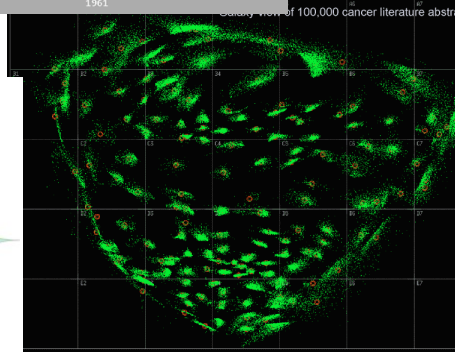
Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with "X and Y" and "X at Y" respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

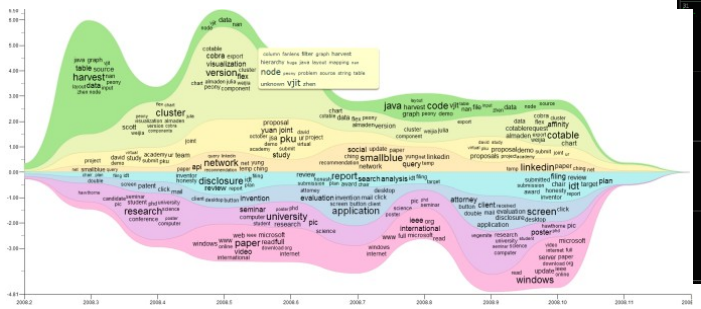












All time most popular tags

"Putting the pieces together"

Jigsaw

Computational analysis of documents' text

Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

Views are highly coordinated



MEMBER LOGIN: Username: Password: LOGIN

Forgot My Username/Password

Remember Login Info

eRobertParker.com

The Independent Consumer's Guide to Fine Wines



Rock and Rhône:
Parker Explores the Greatest Unheralded Grapes in the World–Grenache, Mourvèdre, and Syrah

HOME SUBSCRIBE GIFT SUBSCRIPTIONS FAQs VIRTUAL TOUR SUPPORT SITE MAP CONTACT US

What's Inside

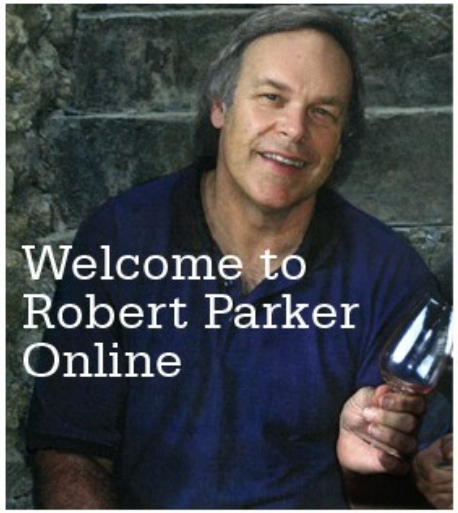
Explore Wines
Over 190,000 searchable notes from *The Wine Advocate*, Robert Parker's books and other sources.

The Hedonist's Gazette
An insider's look at our reviewers' wine and food experiences.

Article Archive
In-depth commentary from *The Wine Advocate* and other sources.

My Wines
Powerful, easy-to-use tools for purchasing/managing a cellar and participating in our wine community.

Robert Parker Mobi
Search the notes from any Internet-enabled mobile device.





Information and Products for Non-Subscribers

Wine of the Day:
2009 Ridge Chardonnay
View Tasting Note



3W's of Wine iPhone App
Learn More

The Wine Advocate Store
Shop Now

Parker & Zraly Wine Certification
Learn More

Weekly Wine Buys
Powered by winealert
Learn More

Quick Links	More Information	About Our Reviewers	Links
Username/Password	**The Vintage Chart**	Robert Parker	The Wine Advocate Fund
What's New	Executive Wine Seminars	Antonio Galloni	Educational
Gift Subscriptions	Glossary of Wine Terms	Karen MacNeil	Accessory/Storage Links
Get News Via RSS	Parker's Rating System	Neal Martin	Retailer Links
Shop The Wine Advocate Store	*The Wine Advocate*	Jay Miller	Other
TWA Editorial Calendar	Site Map	Lisa Perrotti-Brown MW	
		David Schildknecht	
		Mark Squires	
		Kevin Zraly	
		Past Reviewers	
		Writer Ethics/Standards	





Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, …

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

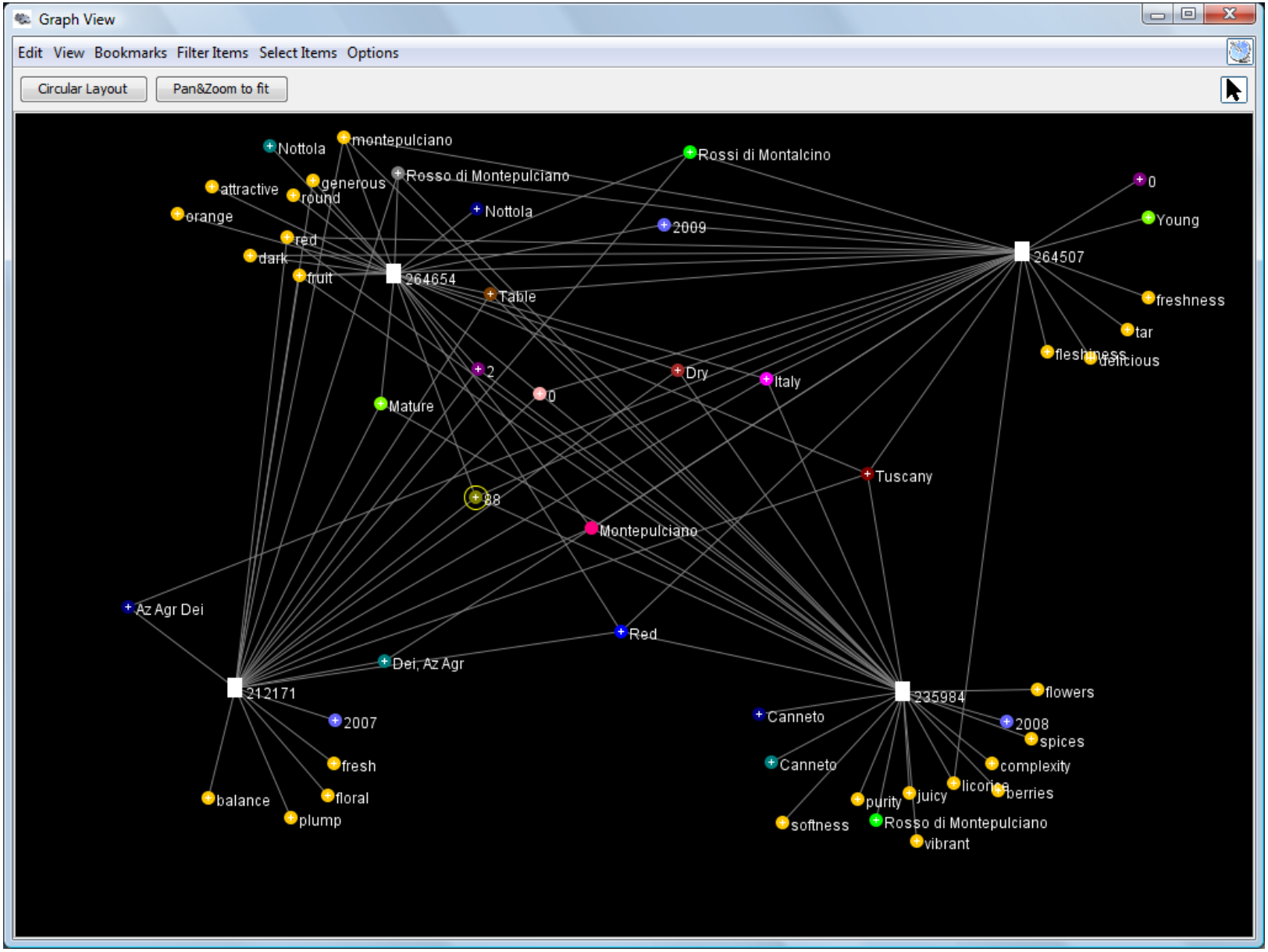
from database of 150,000 reviews



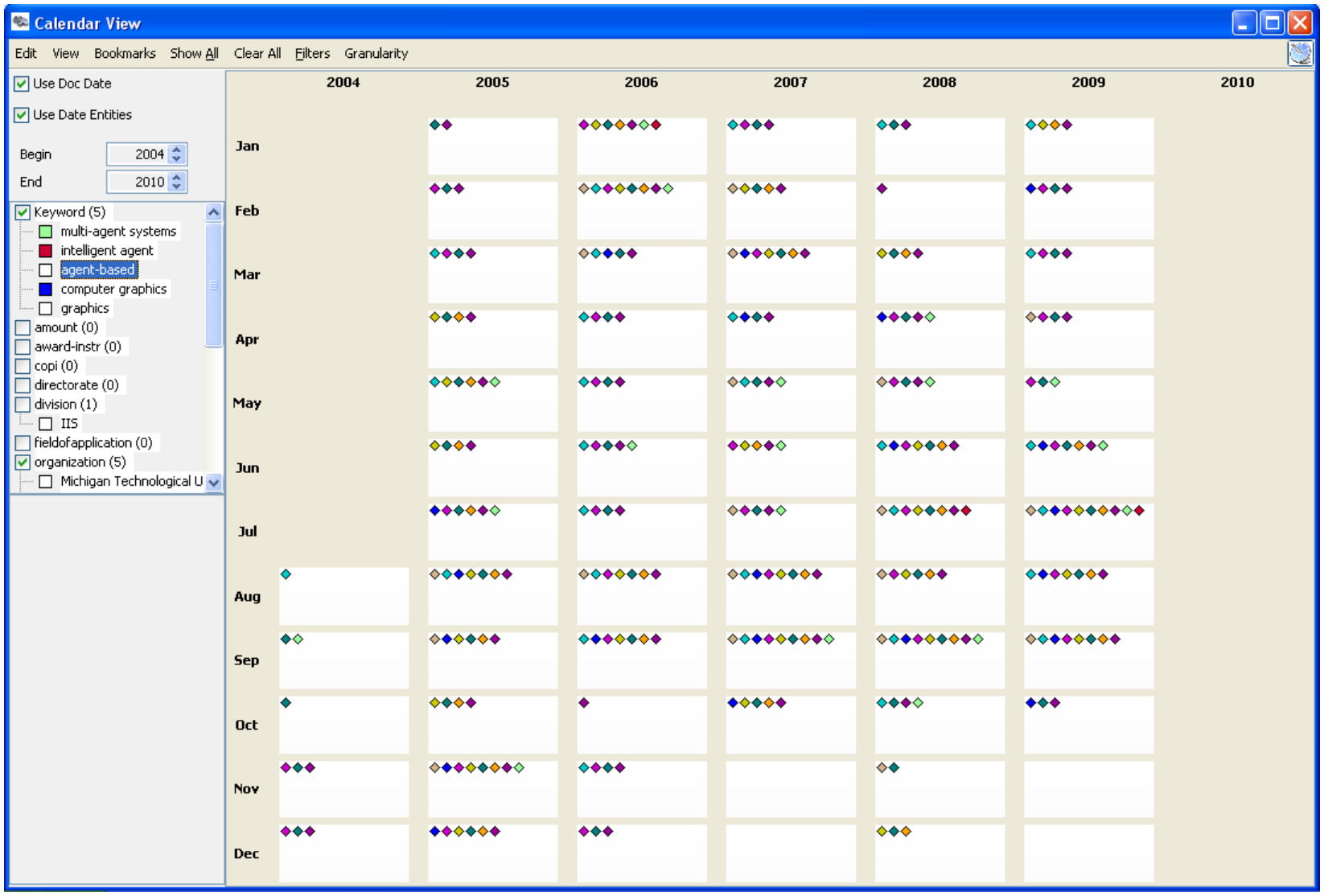
Demo



Graph View

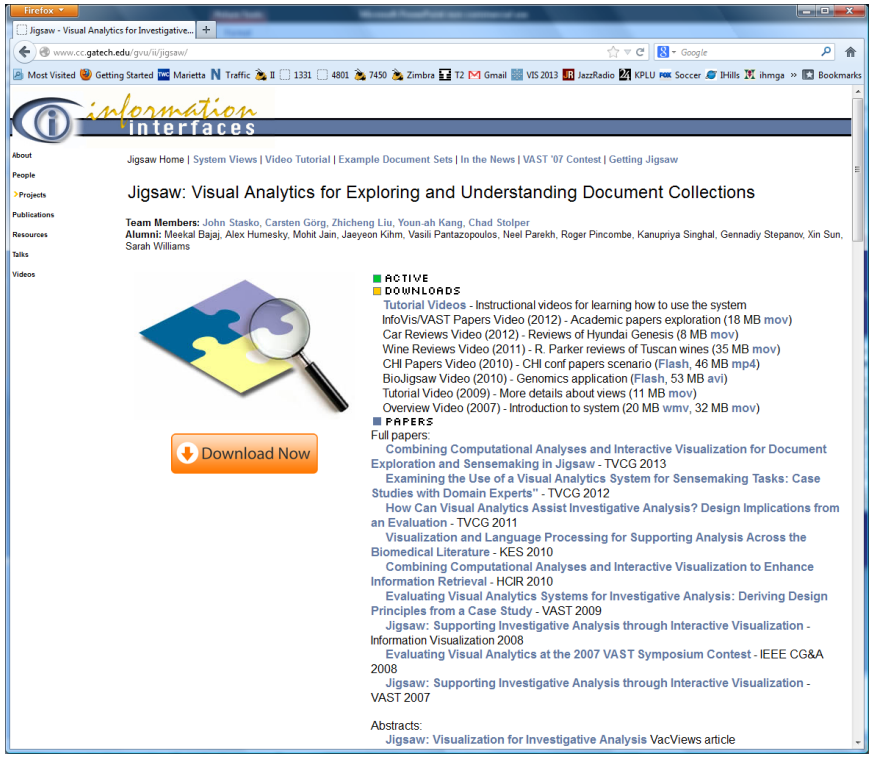

Graph View
Edit View Bookmarks Filter Items Select Items Options
Circular Layout
Pan&Zoom to fit
Nottola
montepulciano
Rossi di Montalcino
Rosso di Montepulciano
generous
attractive
round
orange
Nottola
2009
red
dark
fruit
264654
Table
0
Young
264507
freshness
tar
fleshiness
delicious
2
Dry
Italy
0
Mature
Tuscany
88
Montepulciano
Az Agr Dei
Red
Dei; Az Agr
212171
235984
flowers
2007
Canneto
2008
spices
fresh
Canneto
complexity
balance
floral
purity
juicy
licorice
berries
plump
softness
Rosso di Montepulciano
vibrant

Calendar View



Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw





Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context

Data cleaning example

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

Repeated interactive queries

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Varieties, colors, reviewers, producers, concepts, …

Insights

No great aha! – More learning & knowledge building

Domain knowledge (wines) matters

Further benefits of visualization

Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

EuroVis ‘14 Capstone Talk



http://vimeo.com/98986594



Take Aways

Presentation & analysis

Interaction provides the power

Exploring & developing questions

Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

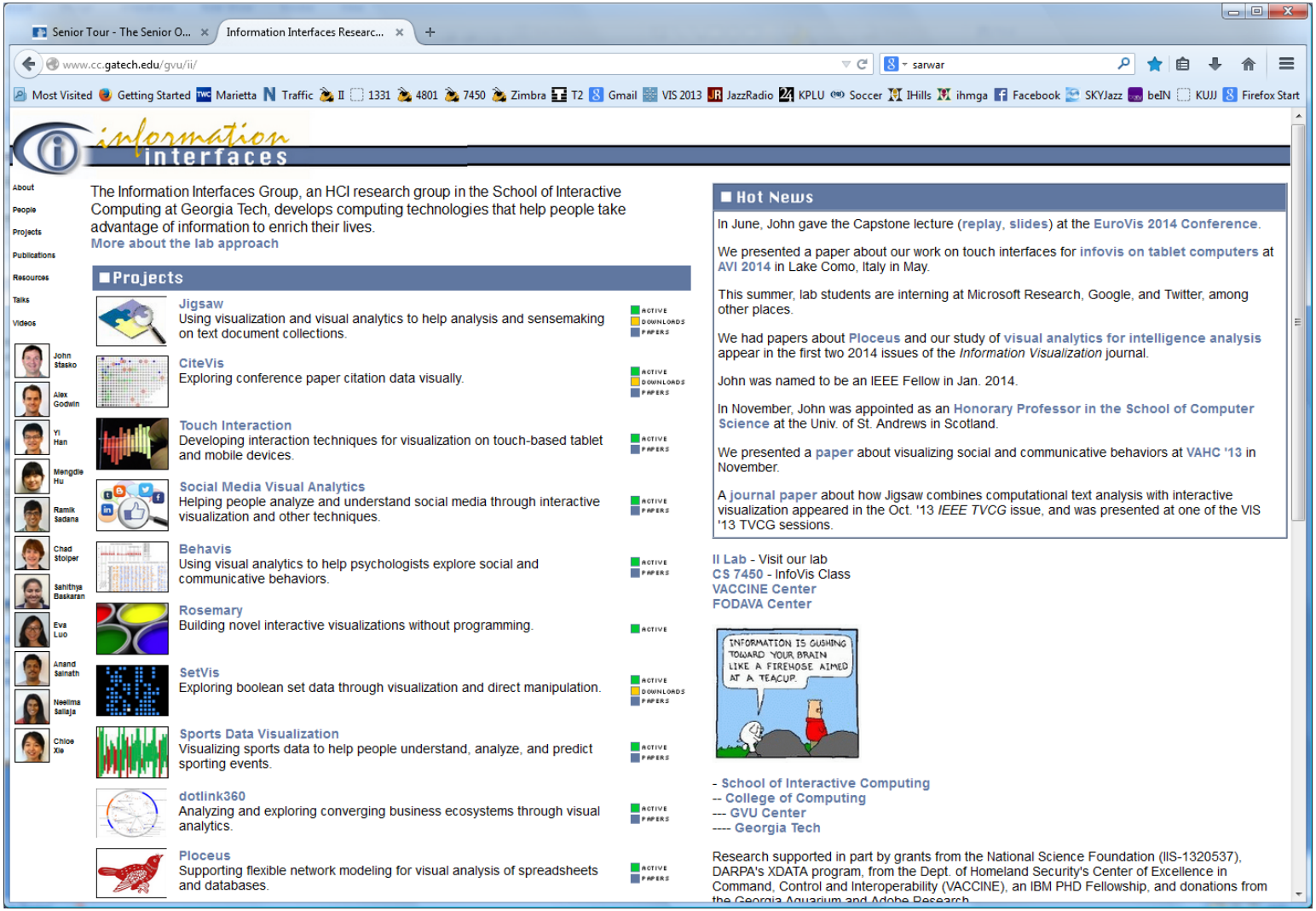
- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

This material is supported in part by:
National Science Foundation
Grant #0414667

- Supported by DARPA's XDATA Program

My Research

http://www.cc.gatech.edu/gvu/ii



VizBi '15 Keynote
March 25, 2015

Data Visualization Principles (Some Things I've Learned)

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu



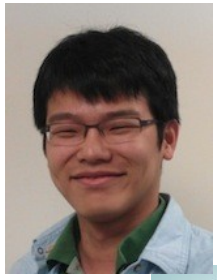


Georgia Tech College of Computing
School of Interactive Computing

Three themes

1. Understanding purpose
2. Embracing interaction
3. Identifying value

Talented students



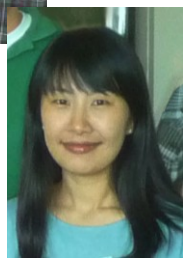
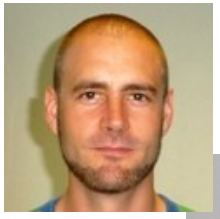
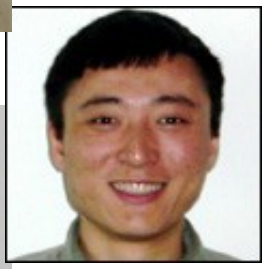




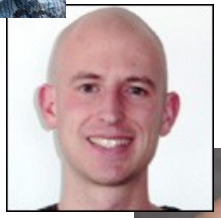


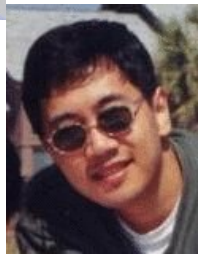


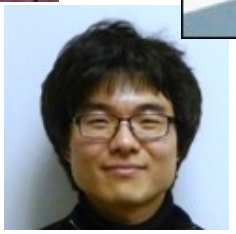
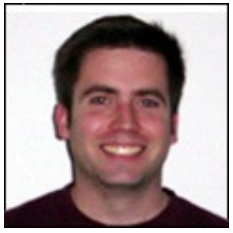







Why visualize?

Applications of Visualization

Presentation	(Explanatory)
Analysis	(Exploratory)

1. Presentation

Communicate data and ideas

Explain and inform

Provide evidence and support

Influence and persuade



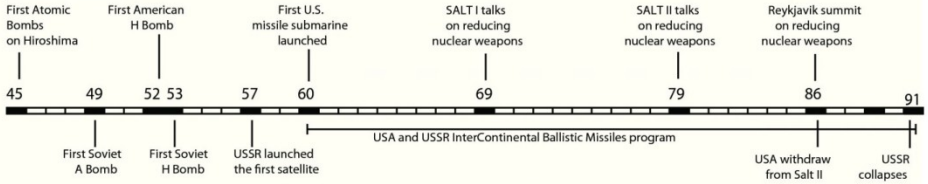
Infographics

All the rage…

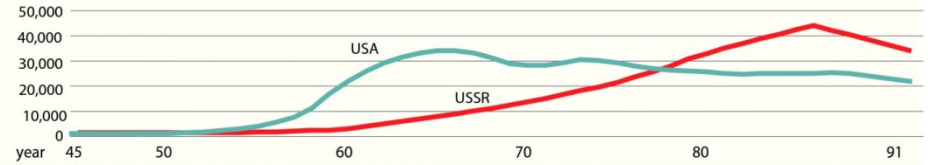
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



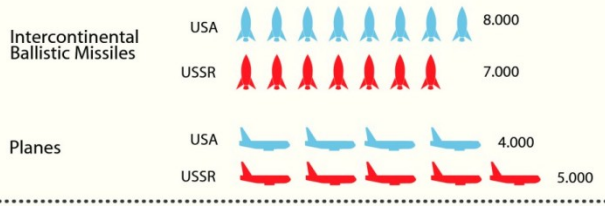
NUCLEAR STOCKPILE



ROCKET MODELS



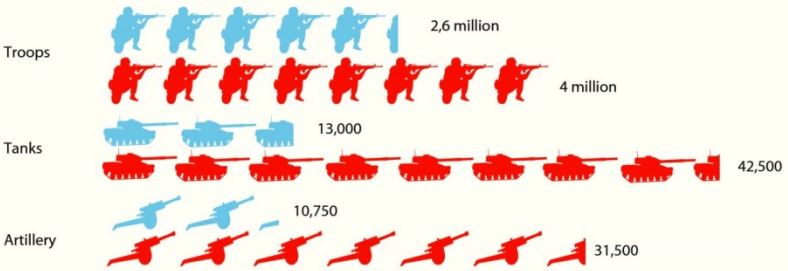
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.





Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

NY Times
2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

Bonds was injured last season. He played 14 games and hit 6 homers.

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

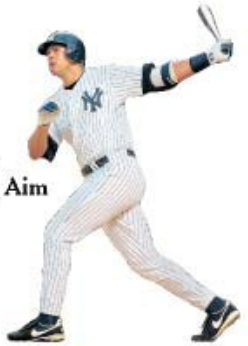
Aaron
Actual homers slightly outpace projected homers for five seasons.

Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

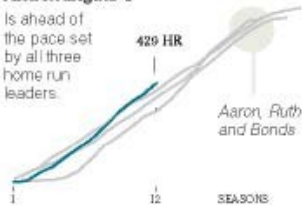
Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

Note: Ages as of July 1 of each season.

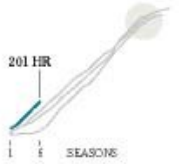
Others Taking Aim



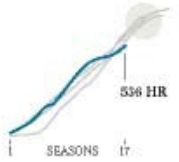
Alex Rodriguez
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

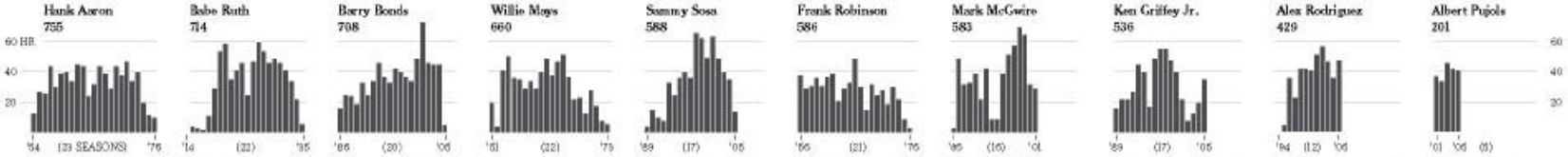


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

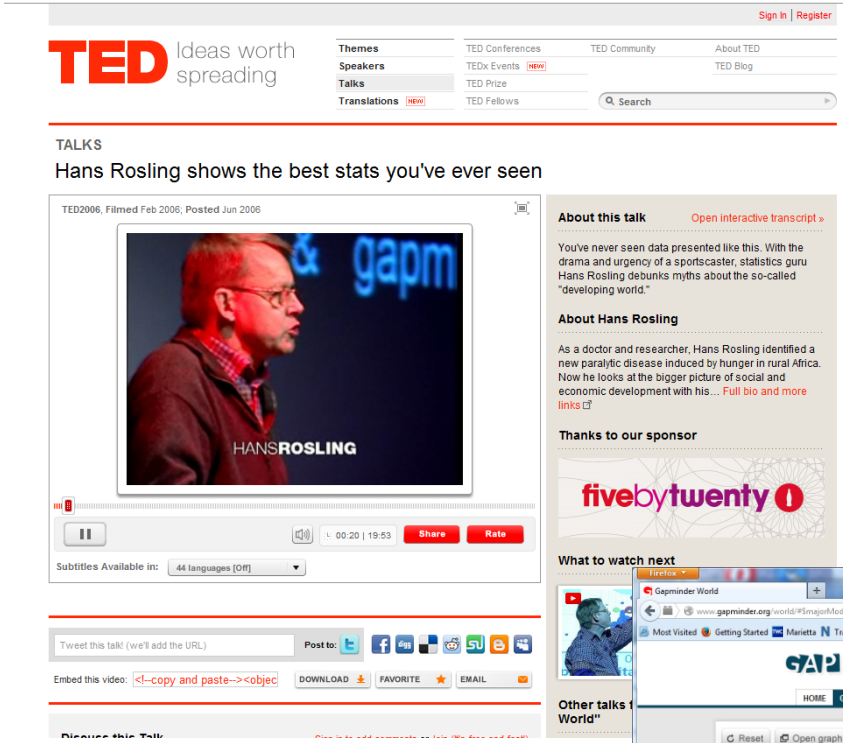


Differing Paths to the Top of the Charts

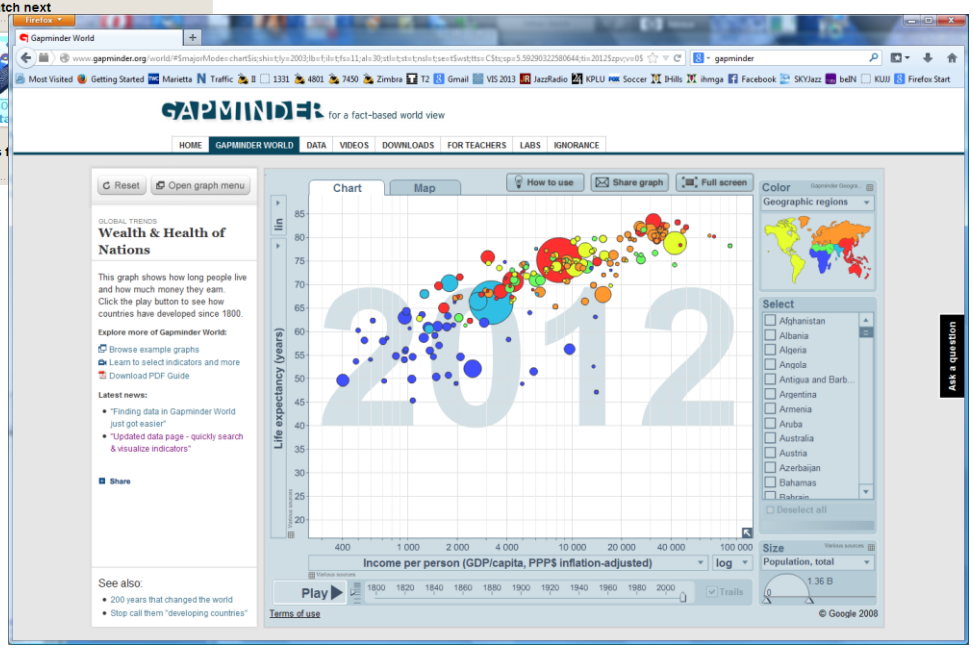
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



Amanda Cox and Joe Ward/The New York Times



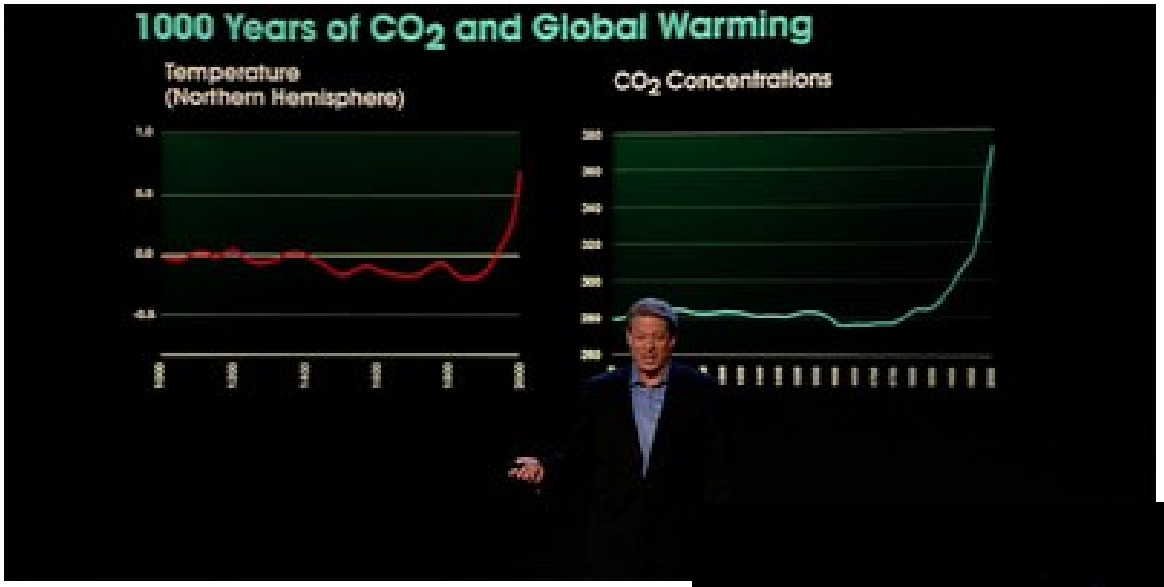
Hans Rosling
Gapminder



2006

http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

An Inconvenient Truth





Gore made extensive use of data graphics

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

Persuade

May just show a few variables
and/or a subset of the data cases



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





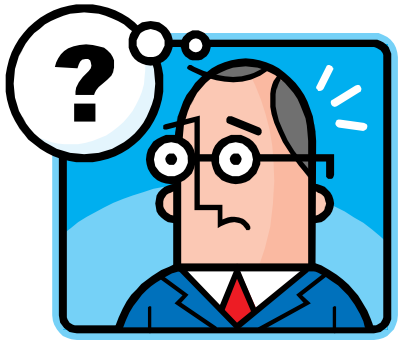
Thought

If you can articulate very precisely what you're seeking, visualization likely isn't your best approach

OK, so what is visualization good for?

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Frequent analysis goals

Show many/all variables

Illustrate overview and detail

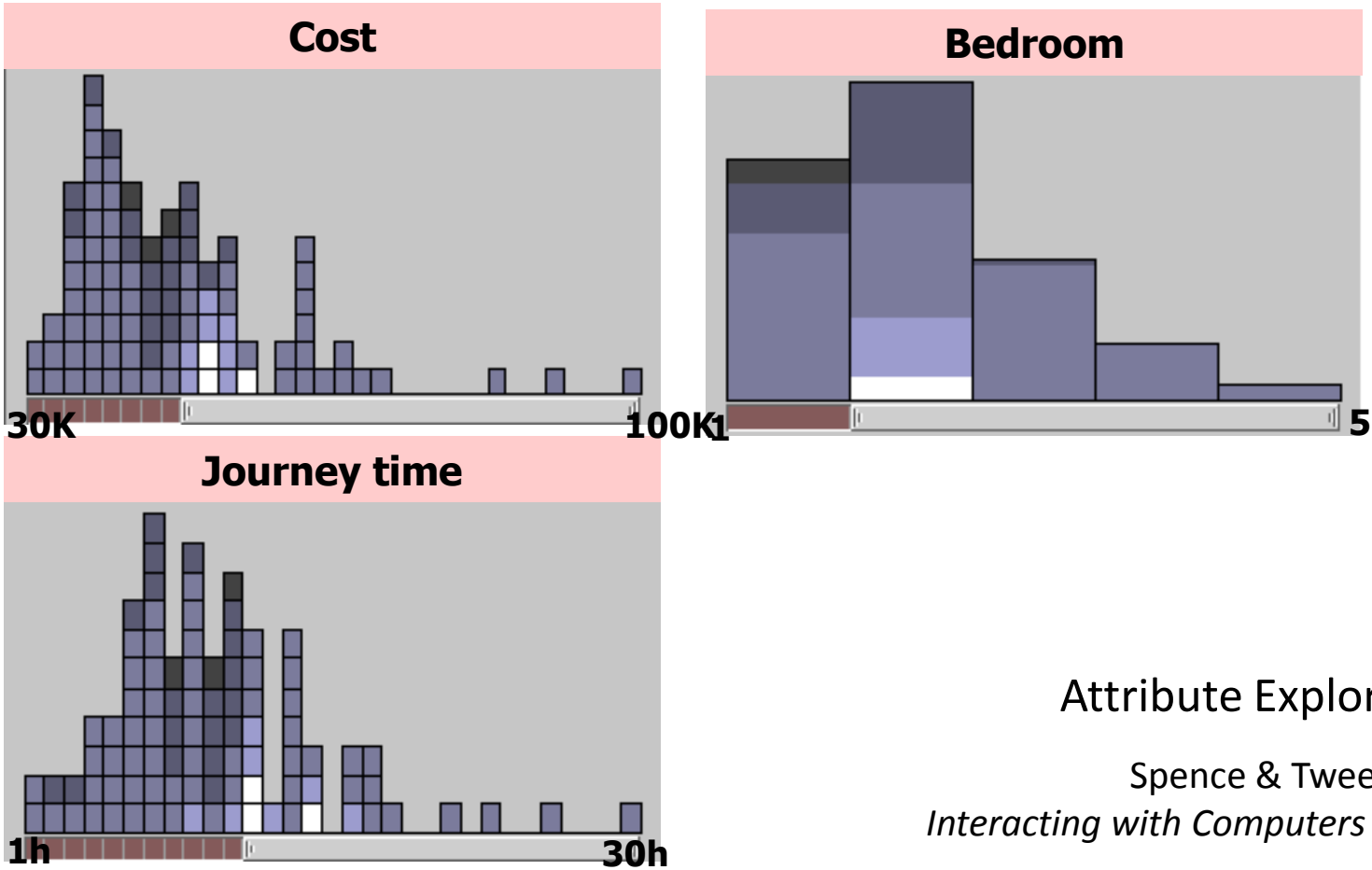
Facilitate comparison

Display may not be easy to interpret at first

At Its Heart: Uncertainty

Nothing is perfect

Humans trade off choices, compromise



Attribute Explorer

Spence & Tweedie

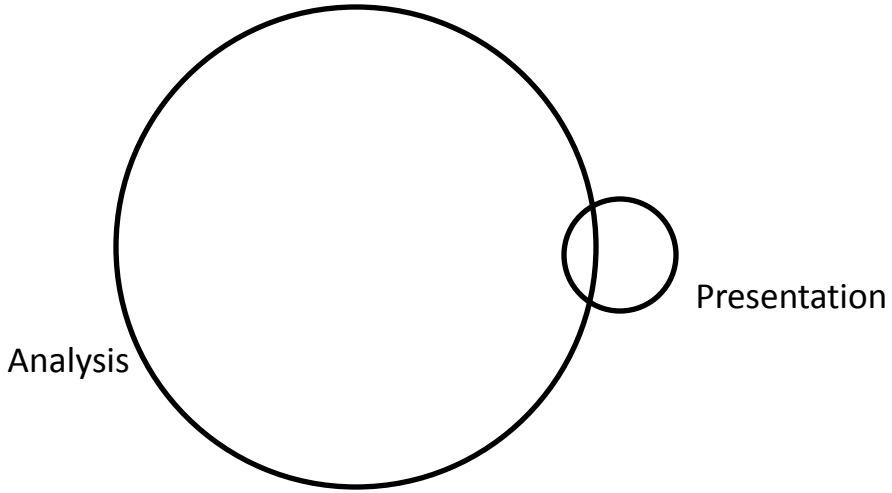
Interacting with Computers '98

Takeaway 1

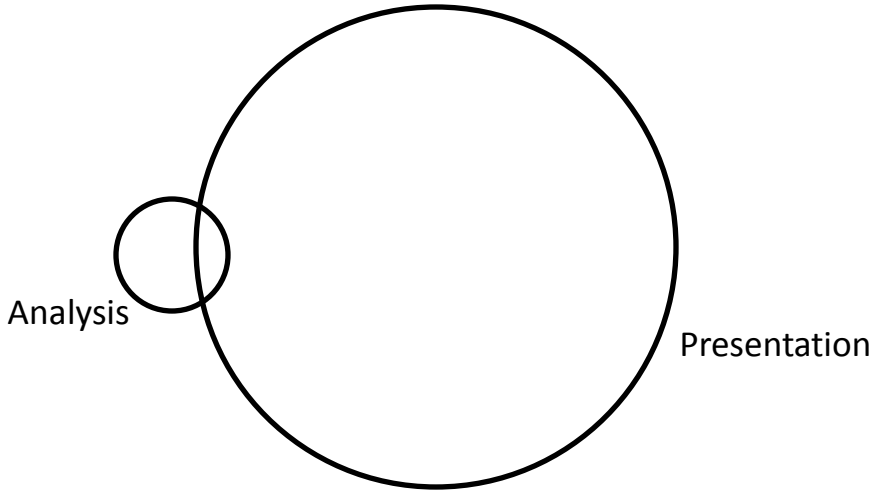
Analysis $\neq$ Presentation

Views of the World

Vis researchers



Everyone else





Narrative Visualization: Telling Stories with Data

Edward Segel and Jeffrey Heer

Abstract—Data visualization is regularly promoted for its ability to reveal stories within data, yet these "data stories" differ in important ways from traditional forms of storytelling. Storytellers, especially online journalists, have increasingly been integrating visualizations into their narratives, in some cases allowing the visualization to function in place of a written story. In this paper, we systematically review the design space of this emerging class of visualizations. Drawing on case studies from news media to visualization research, we identify distinct genres of narrative visualization. We characterize these design differences, together with interactivity and messaging, in terms of the balance between the narrative flow intended by the author (imposed by graphical elements and the interface) and story discovery on the part of the reader (often through interactive exploration). Our framework suggests design strategies for narrative visualization, including promising under-explored approaches to journalistic storytelling and educational media.

Index Terms—Narrative visualization, storytelling, design methods, case study, journalism, social data analysis.

1 Introduction

In recent years, many have commented on the storytelling potential of data visualization. News organizations including the New York Times, Washington Post, and the Guardian regularly incorporate dynamic graphics into their journalism. Politicians, activists, and television reporters use interactive visualizations as a backdrop for stories about global health and economics [10] and election results [9]. A recent feature in The Economist [6] explores the proliferation of digital data and notes that visualization designers are "*melding the skills of computer science, statistics, artistic design and storytelling.*"

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives.

Crafting successful "data stories" requires a diverse set of skills. Gershon and Page [12] note that effective story-telling "*require[s] skills like those familiar to movie directors, beyond a technical expert's knowledge of computer engineering and science.*" While techniques from oration, prose, comic books, video games, and film production are applicable to narrative visualization, we should also expect this emerging medium to possess unique attributes. Data stories differ in important ways from traditional storytelling. Stories in text and film typically present a set of events in a tightly controlled progression. While tours through visualized data similarly can be organized in a linear sequence, they can also be interactive, inviting verification, new questions, and alternative explanations.

Currently, most sophisticated visualization tools focus on data exploration and analysis. Applications such as spreadsheets and visualization tools support an array of analysis routines and visual encodings, but beyond exporting images for presentation typically provide scant support for crafting stories with analysis results. As such, they provide powerful vehicles for discovering "stories", but do little to aid narrative communication of these findings to others. As tools mature and more richly integrate with the web (e.g., Many Eyes [25], Tableau Public [22], GeoTime Stories [8]), they are enabling the publication of dynamic graphics with variably constrained levels of interactivity. It remains an open question how the design of such tools might be evolved to support richer and more diverse forms of storytelling.

In this paper, we investigate the design of narrative visualizations and identify techniques for telling stories with data graphics. We take an empirical approach, analyzing visualizations from online journalism, blogs, instructional videos, and visualization research. After reviewing related work, we share five selected case studies which highlight varied design strategies and illustrate our analytic approach. We then formulate a design space constructed from an analysis of 58 examples. Our analysis identifies salient dimensions of visual storytelling, including how graphical techniques and interactivity can enforce various levels of structure and narrative flow. We describe seven genres of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and video. These genres can be combined with interactivity and messaging to produce varying balances of author-driven and reader-driven experiences. Finally, we discuss the implications of our framework, noting recurring design strategies, promising yet under-utilized approaches to integrating visualization with other media, and the potential for improved user interfaces for crafting data stories. By focusing on the graphical and interactive elements of narrative visualization, our approach gives less attention to the cognitive and emotional experience of the reader. We recognize the importance of these elements, however, and describe directions for future reader-centric research in our conclusion.

2 Related Work

Storytelling and visual expression are integral parts of human culture; storytelling has even been referred to as "*the world's second-oldest profession*" [12]. Without summarizing millennia of achievement, we describe a few of the key concepts informing narrative visualization.

2.1 Narrative Structure

The Oxford English Dictionary defines *narrative* as "an account of a series of events, facts, etc., given in order and with the establishing of connections between them." Central to this definition is the notion of a chain of causally related events. Stories of this form often have a beginning, middle, and end [3, 24]: an introduction to the situation, a series of events often involving tension or conflict, and a resolution.

Since ancient times, people have tried to understand and formalize the elements of storytelling. For example, writers (e.g., [5, 19, 21]) have developed typologies of dramatic situations and identified plot lines common to many narratives, such as the "hero's journey" [5]. This research typically distinguishes between the content of the story and the form in which it is told. While stories often concern interacting characters, they may also present a sequence of facts and observations linked together by a unifying theme or argument.

Storytelling strategies vary among media and genre. For instance, stories told through writing have access to a different set of formal mechanisms and narrative structures (e.g., stream of consciousness) than stories told through film (e.g., split-screen sequences [3]). Blundell [2] describes narrative devices for journalism such as the *anecdo-*

• *The authors are with Stanford University, Stanford, CA 94305. E-mail: {esegel, jheer}@stanford.edu.*

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

1077-2626/10/$26.00 © 2010 IEEE Published by the IEEE Computer Society

Jessica Hullman
Nick Diakopoulos
…

Static ok, Interactive better

Data Visualization 101



Problem:
You have a lot of data (& attributes) to understand

Do you?

- Pack all the data into one complex representation
- Spread the data into multiple coordinated views
- Use interaction to reveal different subsets of the data

Constituents

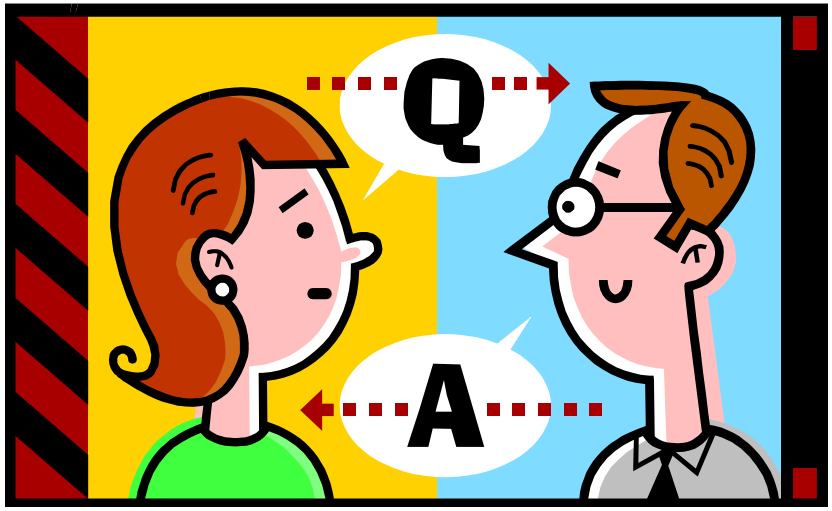
Two key aspects of data visualization

- Representation
- Interaction

> "The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."
>
> S. Few, *Now you see it*

Interaction is Vital

Engage in a dialog with your data



Interaction

Why interact?

Intent

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect

Yi, Kang, Stasko & Jacko
TVCG (InfoVis) '07

Chad Stolper



Interaction

How manifested today?

Tooltips & selection

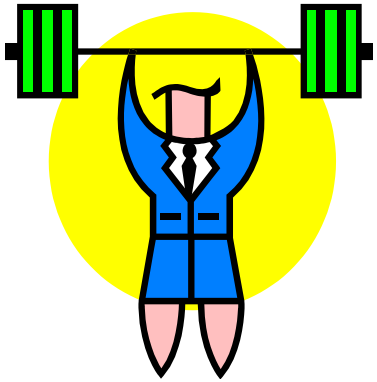
Get details

Navigation

Brushing & linking

Interaction

Can we do more?

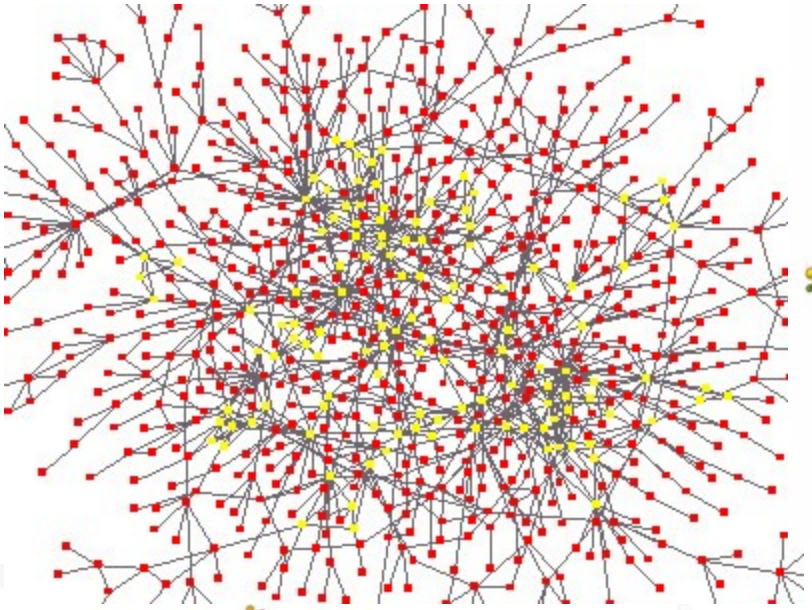


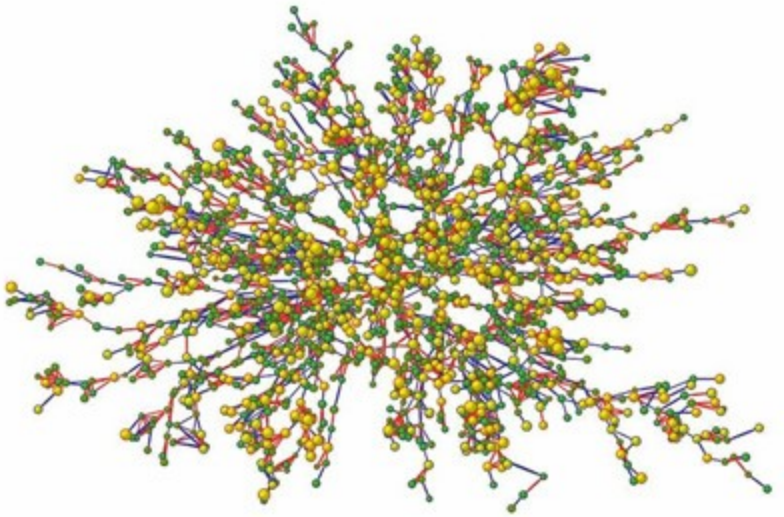
Employ interaction in a more fundamental manner to strengthen the power of visualization

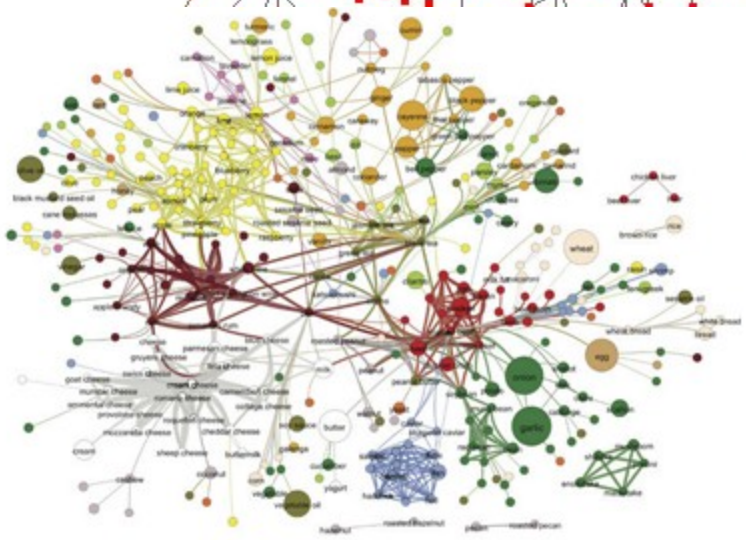
Some examples…

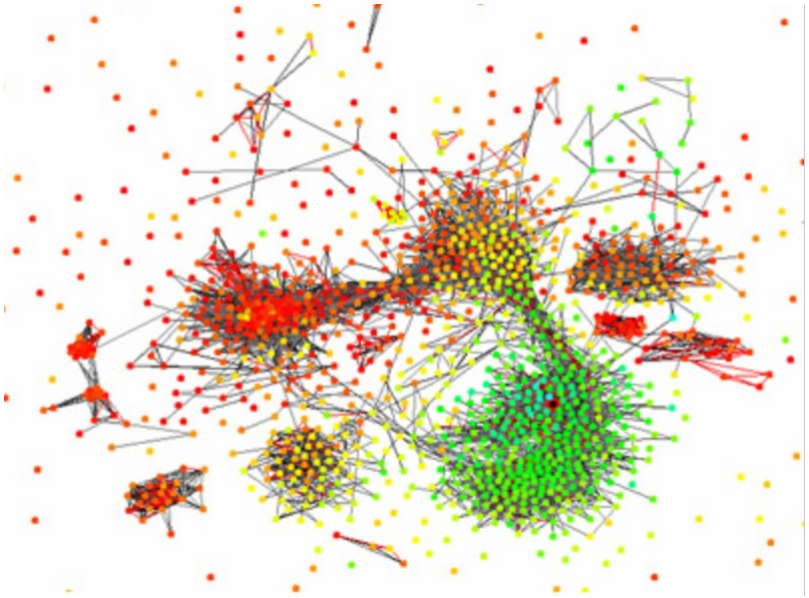
Node-link graphs

http://www.visualcomplexity.com/vc/



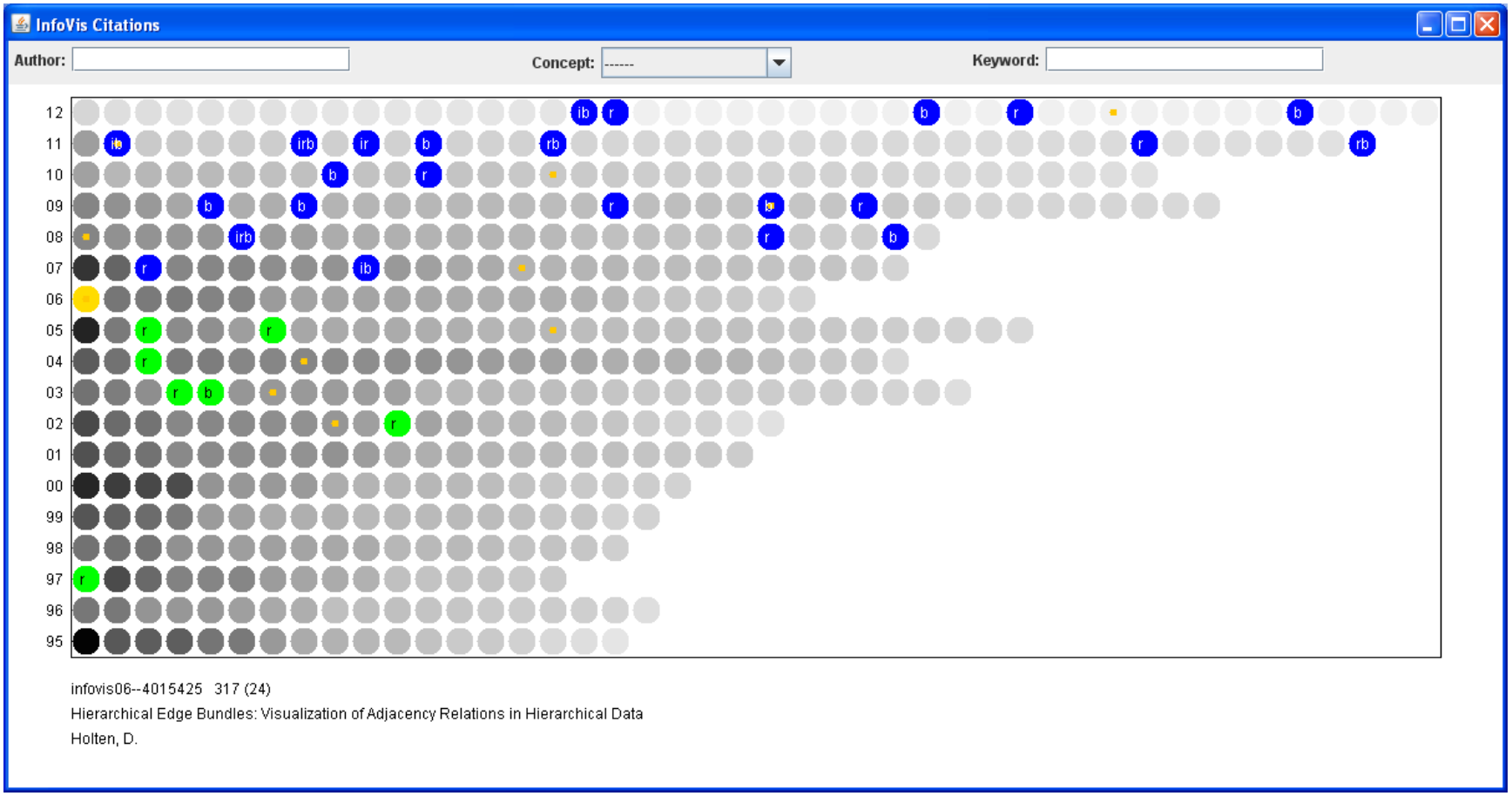






CiteVis

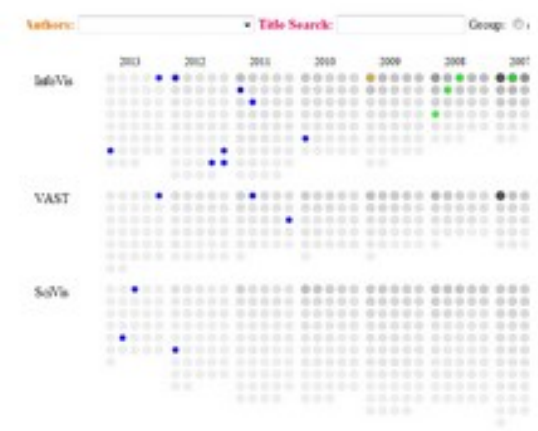
http://www.cc.gatech.edu/gvu/ii/citevis



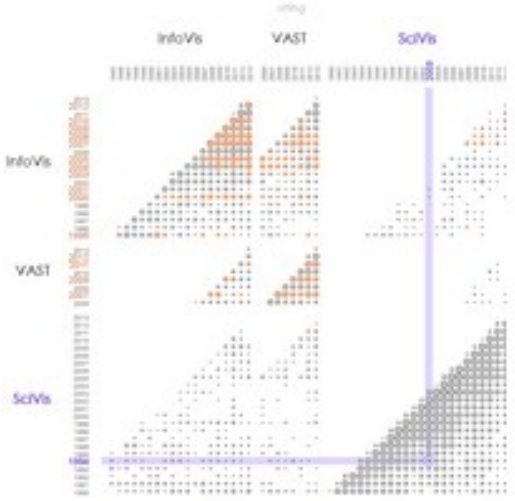
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

VIS25

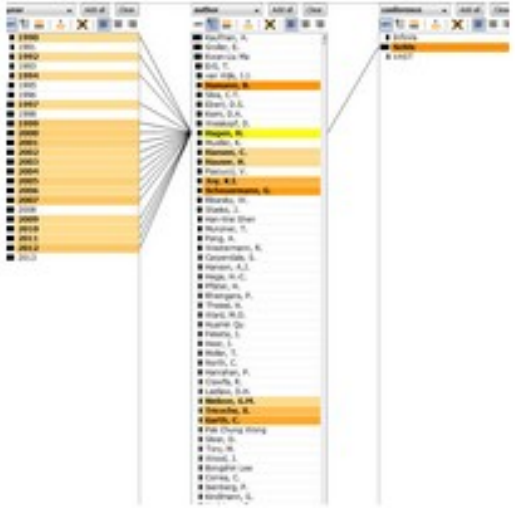
http://www.cc.gatech.edu/gvu/ii/citevis/VIS25/



CiteVis2



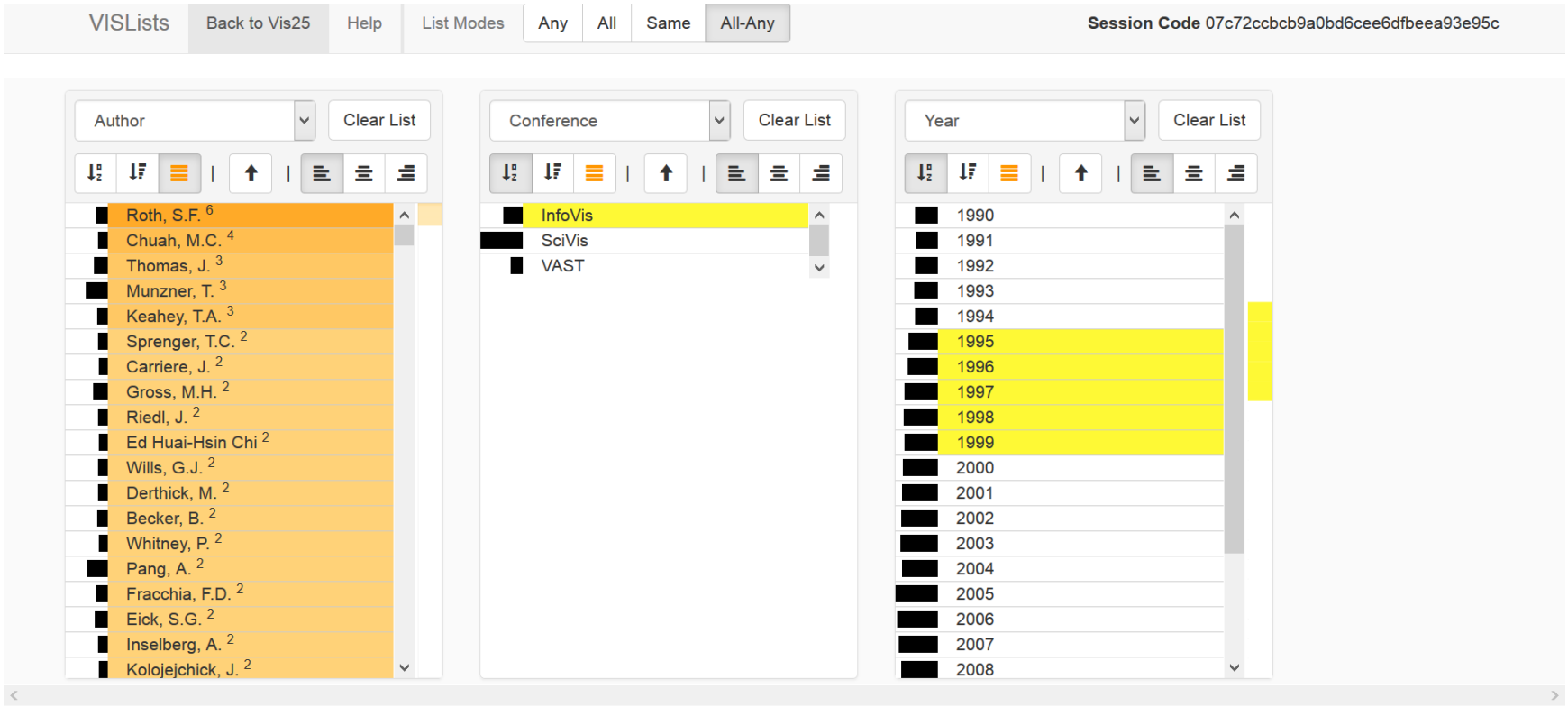
CiteMatrix



VISLists

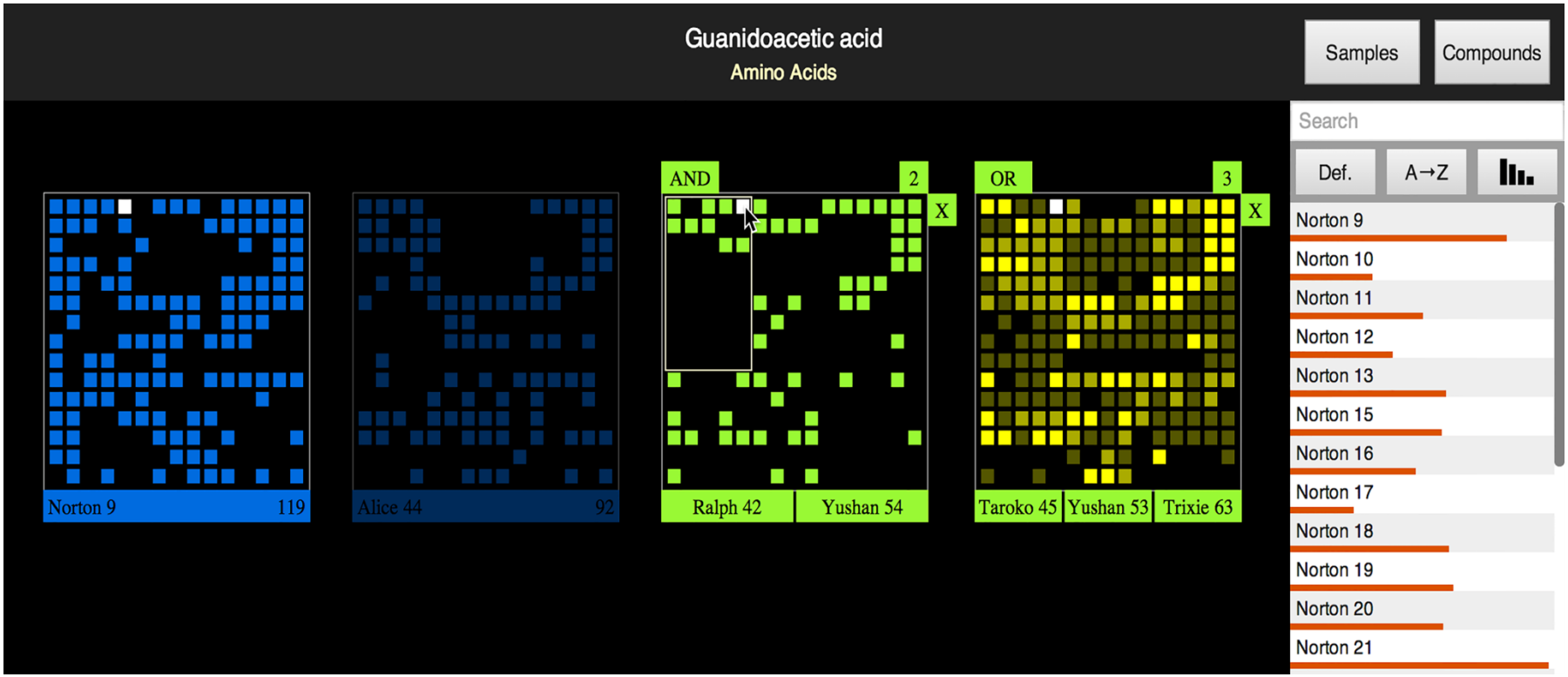
Visualization design: Panpan Xu, Chad Stolper, Anand Sainath

http://vislists.anandsainath.com/list/vis25



OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/



Demo

Sadana, Major, Dove, & Stasko
TVCG (InfoVis) '14

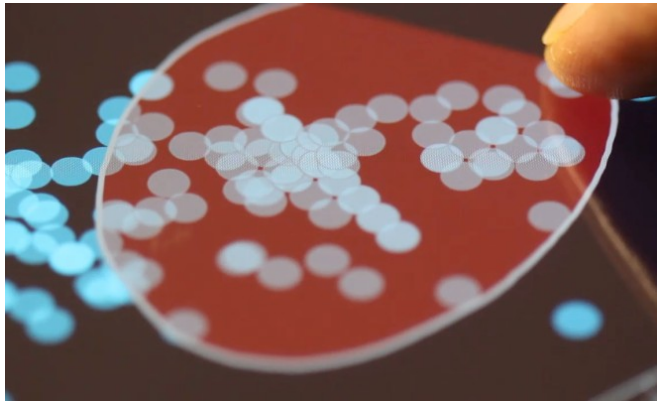
Interaction

What are the tools of interaction?

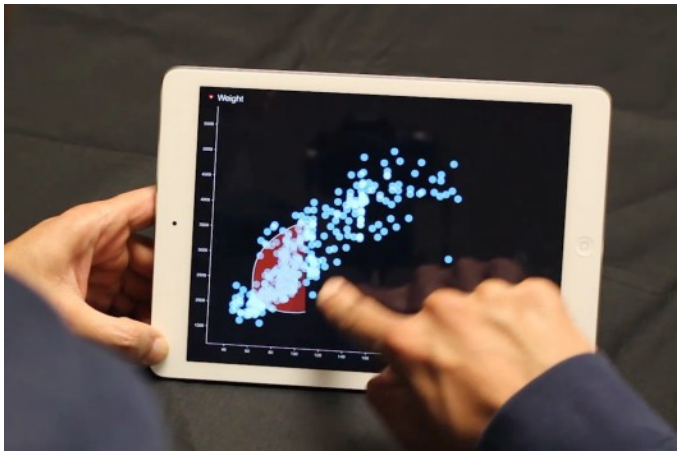
Traditional – Desktop: keyboard, mouse

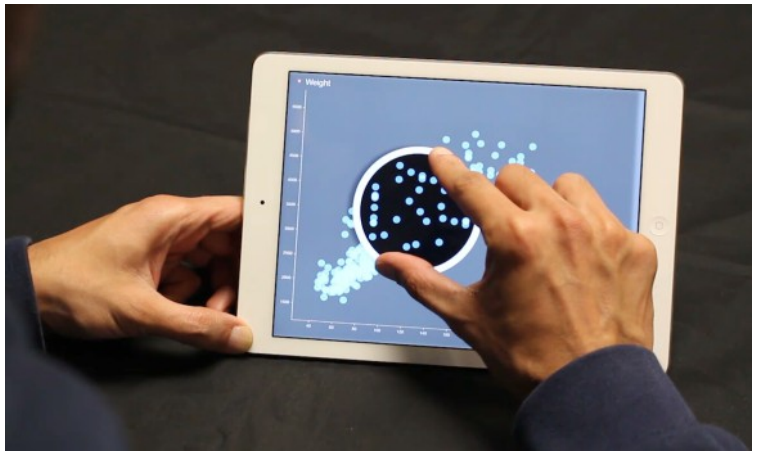
New – Tablet: fingers and multi-touch

Moving to Tablets



Scatterplot





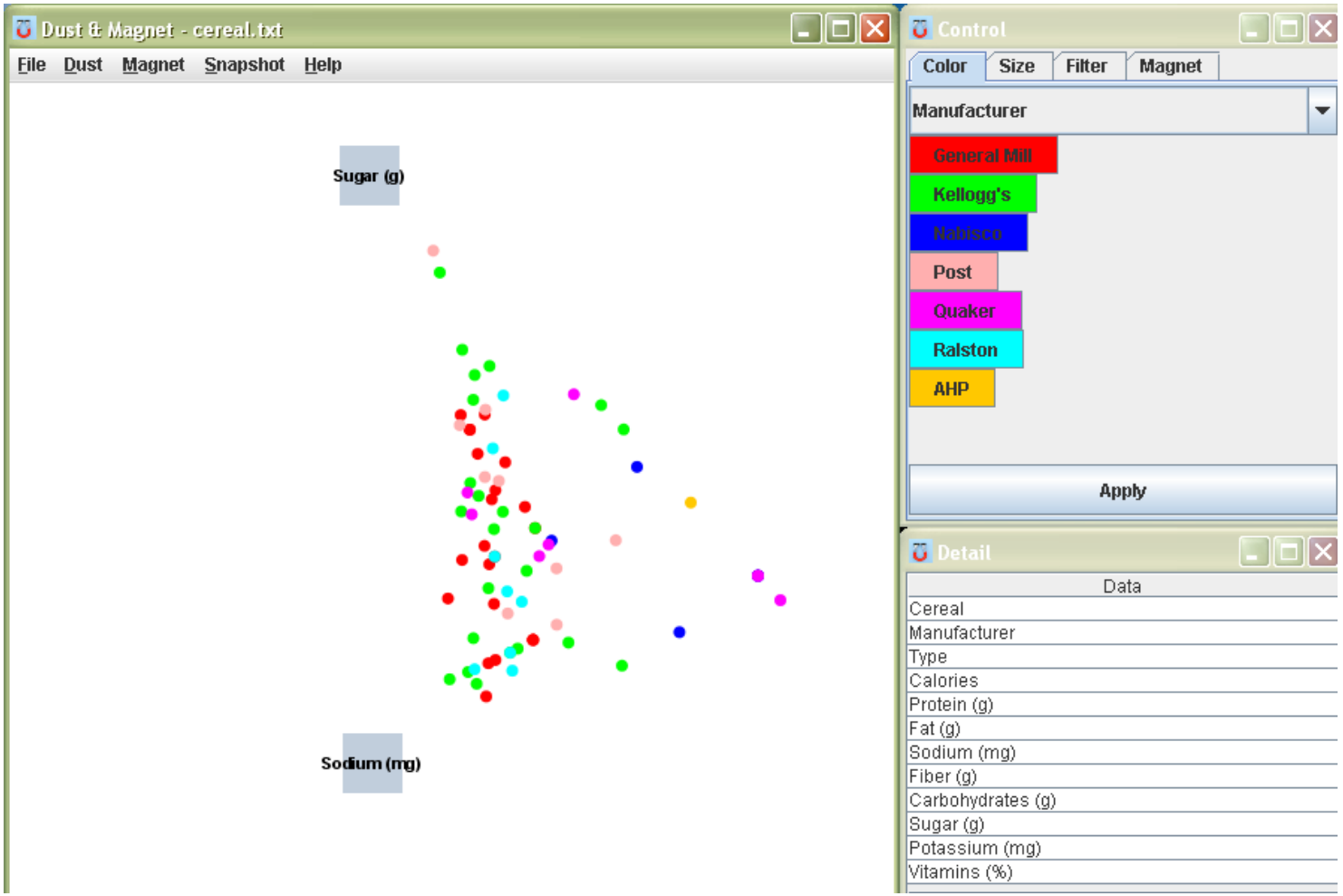


Video

Sadana & Stasko
AVI ‘14

Scatterplot touch video

Dust and Magnet



Yi, Melton, Stasko & Jacko
Information Visualization '05

Pushing it Further

Perceptive Pixel D n M

Takeaway 2

Take advantage of interaction



Visualization's Value

The Value of Information Visualization

Jean-Daniel Fekete[1], Jarke J. van Wijk[2], John T. Stasko[3], and Chris North[4]

[1] Université Paris-Sud, INRIA, Bât 490,
F-91405 Orsay Cedex, France,
Jean-Daniel.Fekete@inria.fr,
http://www.aviz.fr/~fekete/
[2] Department of Mathematics and Computing Science,
Eindhoven University of Technology, P.O. Box 513,
5600 MB EINDHOVEN, The Netherlands,
vanwijk@win.tue.nl,
http://www.win.tue.nl/~vanwijk/
[3] School of Interactive Computing, College of Computing & GVU Center,
Georgia Institute of Technology, 85 5th St., NW,
Atlanta, GA 30332-0760, USA,
stasko@cc.gatech.edu,
http://www.cc.gatech.edu/~john.stasko
[4] Dept of Computer Science, 2202 Kraft Drive,
Virginia Tech, Blacksburg, VA 24061-0106, USA,
north@vt.edu,
http://people.cs.vt.edu/~north/

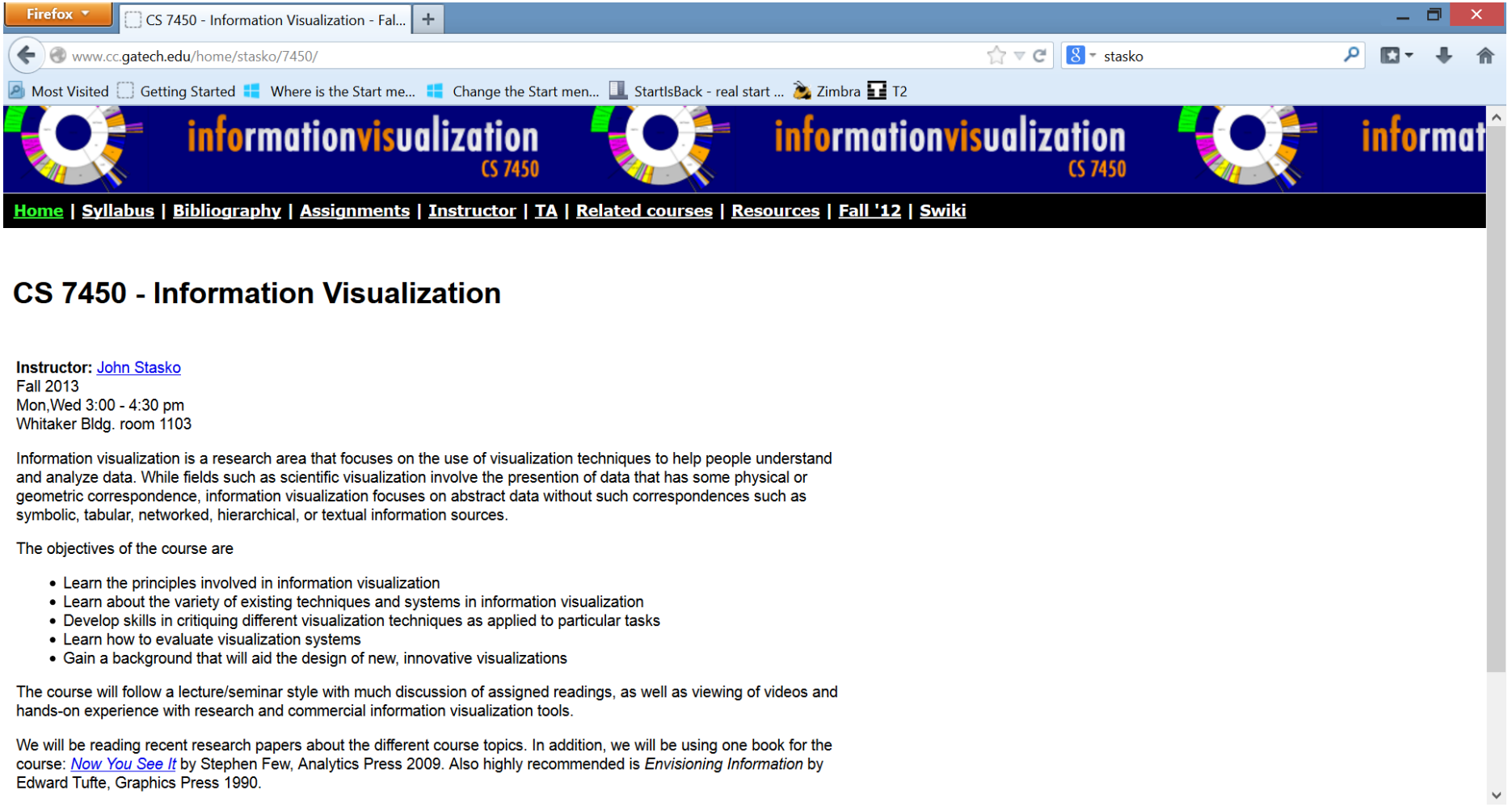
Abstract. Researchers and users of Information Visualization are convinced that it has value. This value can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice. To convince broader audiences, and also, to understand the intrinsic

Information Visualization: Human-Centered Issues and Perspectives
Springer, 2008

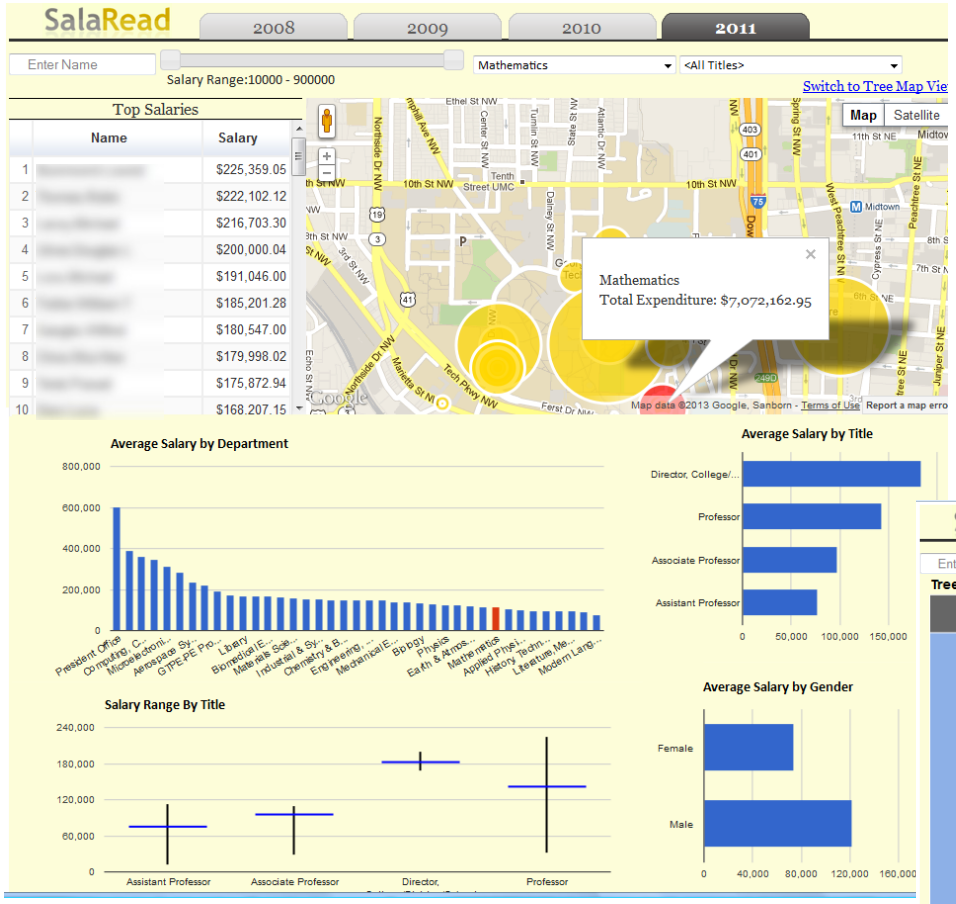
Simply presenting data ***visually*** can have a profound impact

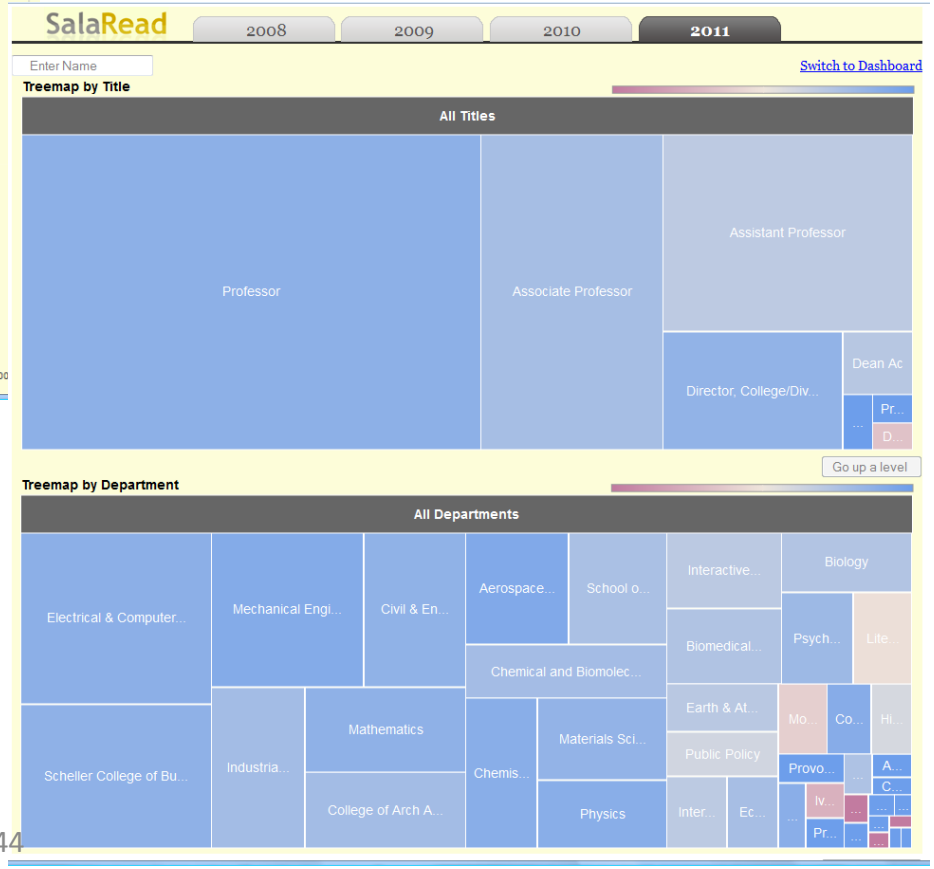
My Class

http://www.cc.gatech.edu/~stasko/7450



CS 7450
Fall ‘12
Design project

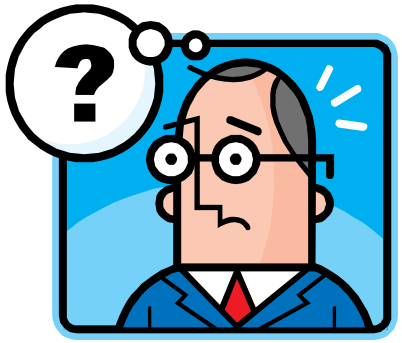




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Visualization's Value?

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Stasko
ACM BELIV '14

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)

What kinds of questions?

"Low-level" tasks

Retrieve value
Filter
Compute derived value
Find extremum
Sort
Determine range
Characterize distribution
Find anomalies
Cluster
Correlate

Amar, Eagan & Stasko
InfoVis '95

Visualization's Value?

$$V_{alue} = T + \mathbf{I} + E + C$$

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)

What is Insight?

An individual observation about the data by the participant, a unit of discovery

Complex
Deep
Qualitative
Relevant
Unexpected

Saraiya, North, & Duca
TVCG '05

North
IEEE CG&A '06

What is Insight?

Sudden grasp of new relationships that are necessary to solve a problem and that were not learned in the past

Bernstein, Penner, Clarke-Stewart & Roy
Psychology, 6th edition

What is Insight?

Is not spontaneous “aha!” moments (eg, in cognitive science)
Is knowledge-building and model-confirmation
    Like a substance that people acquire with the aid of systems

Chang, Ziemkiewicz, Green, & Ribarsky
IEEE CG&A ‘09

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to convey an overall **essence** or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)

Overview and detail

Focus + context

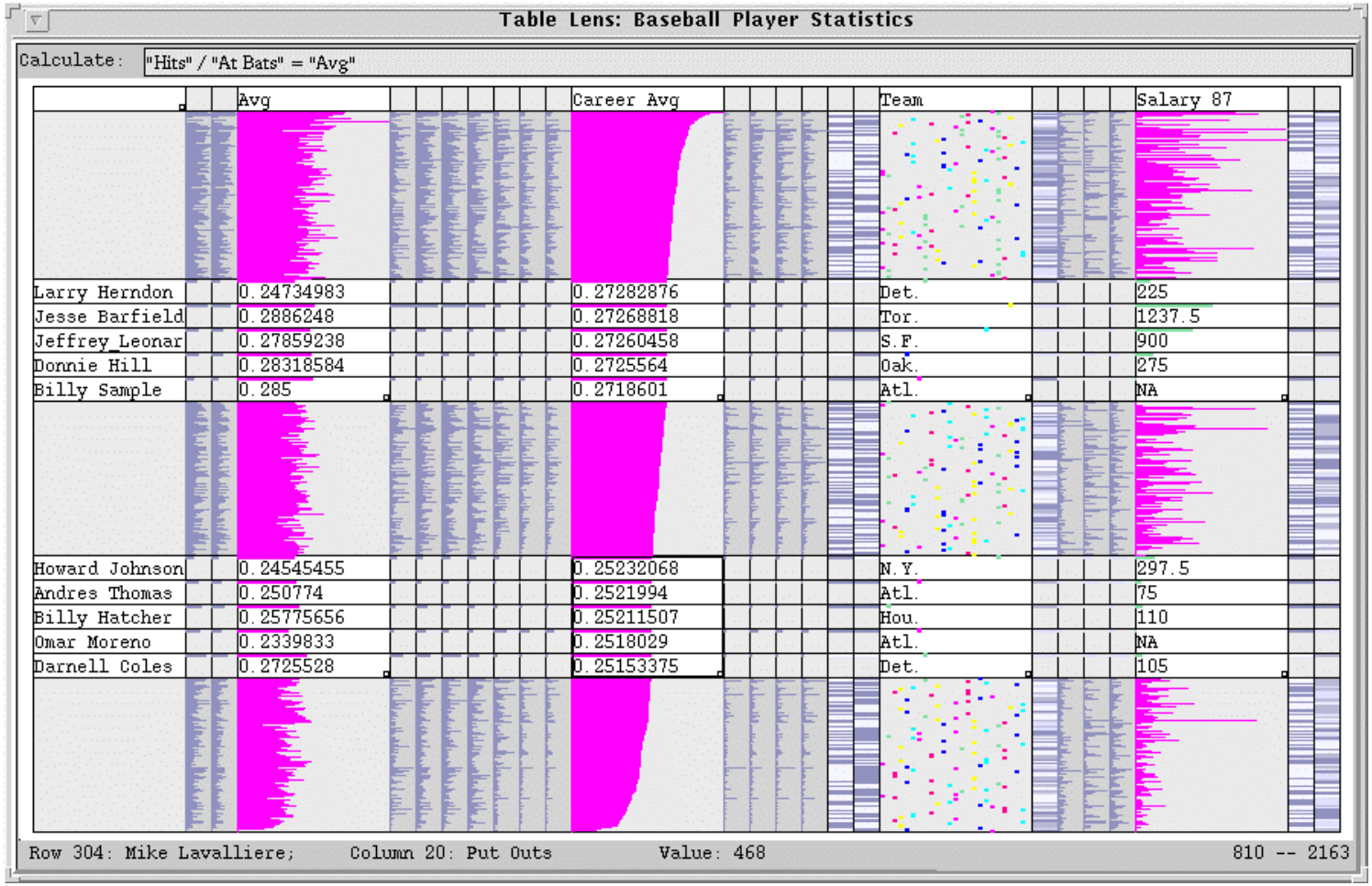


Table Lens

Rao & Card
CHI '94

Visualization's Value?

$$V_{alue} = T + I + E + C$$

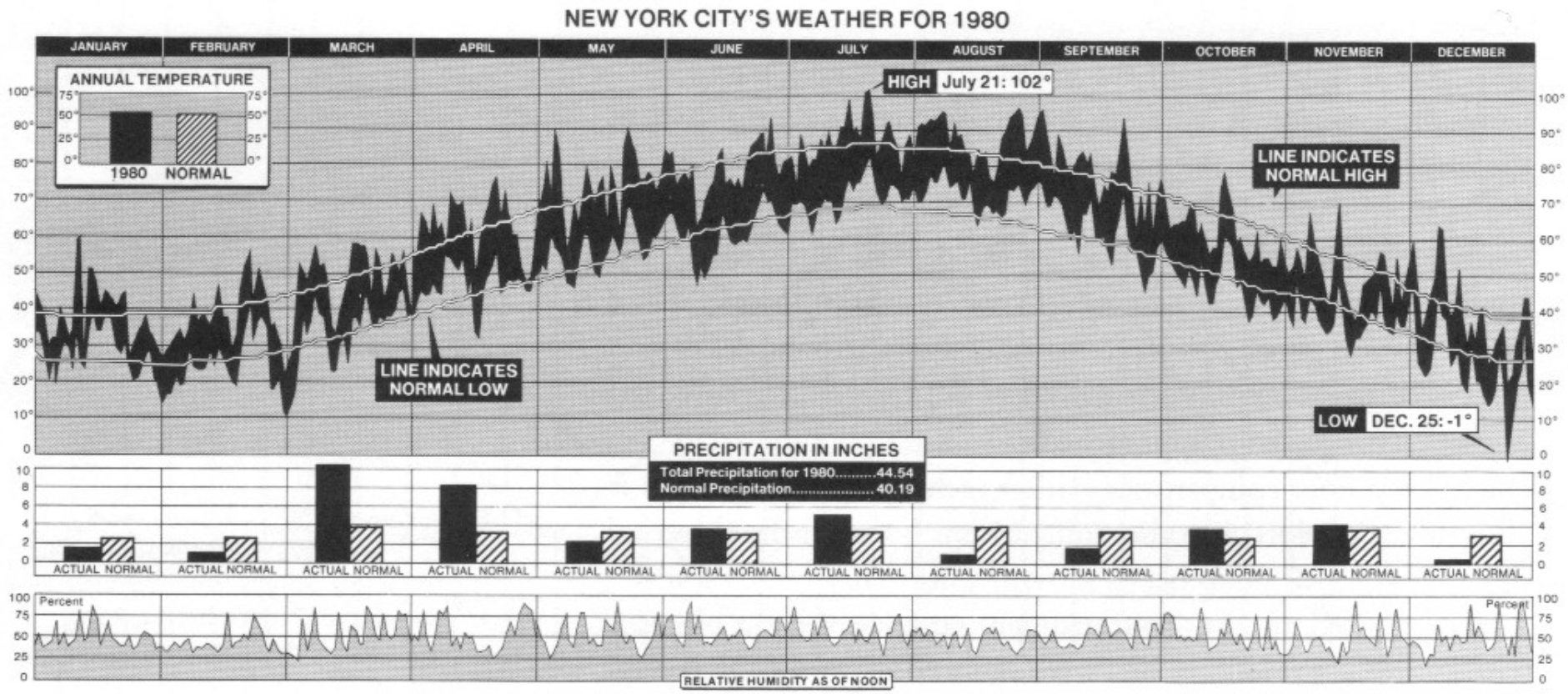
Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)

Some examples

1.

2220 numbers



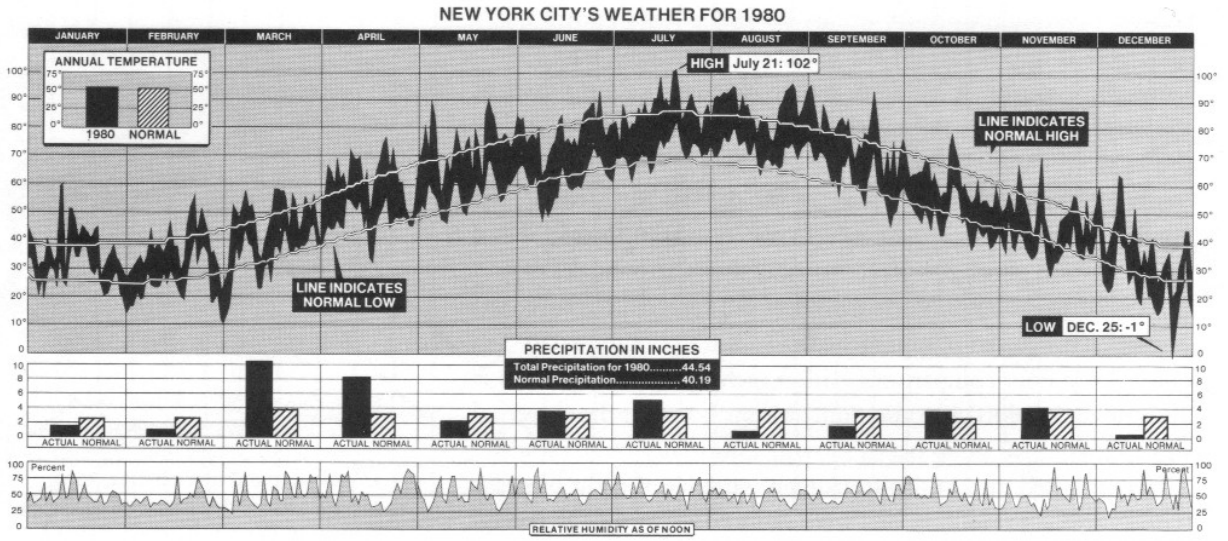
New York Times, January 11, 1981, p. 32.

E. Tufte
The Visual Display of Quantitative Information
1983

Data Values

365	High temp for each day
365	Low temp for each day
365	Avg high temp for each day
365	Avg low temp for each day
365	Precipitation for each day
365	Humidity for each day
12	Precipitation for each month
12	Avg precipitation for each month
1	Precipitation for the year
1	Avg precipitation per year
1	Highest temp (& day) for the year
1	Lowest temp (&day) for the year
1	Avg daily temp for the year
1	Avg daily temp per year

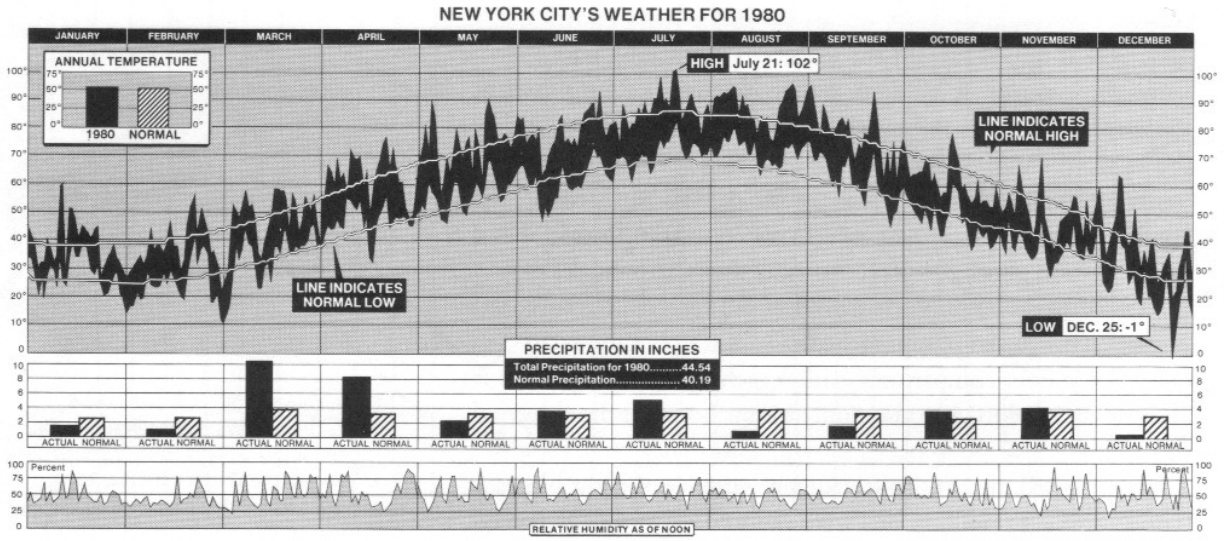
Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

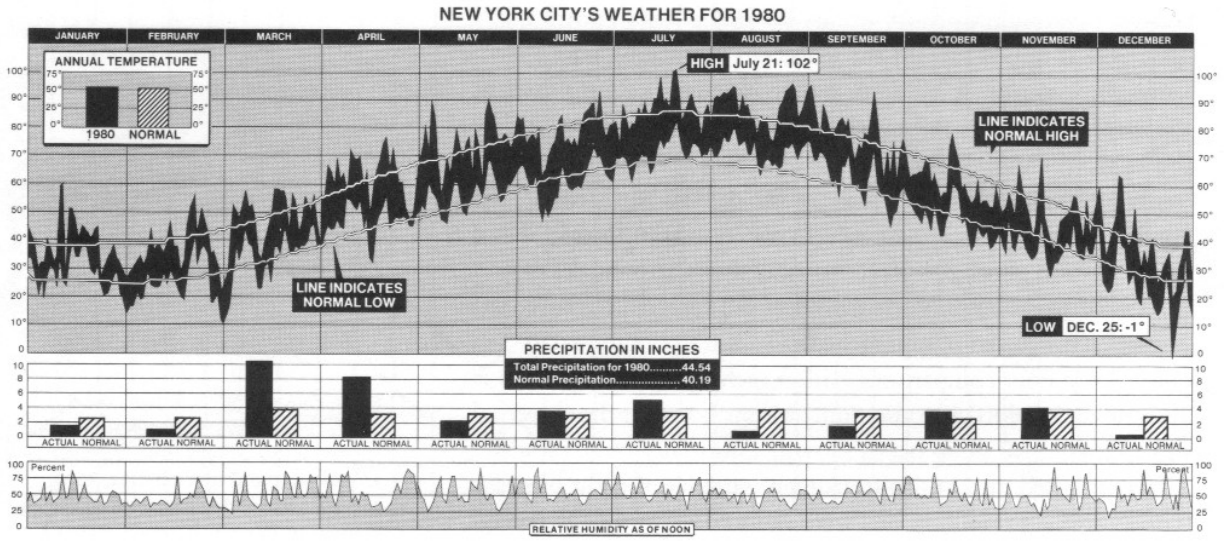
Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

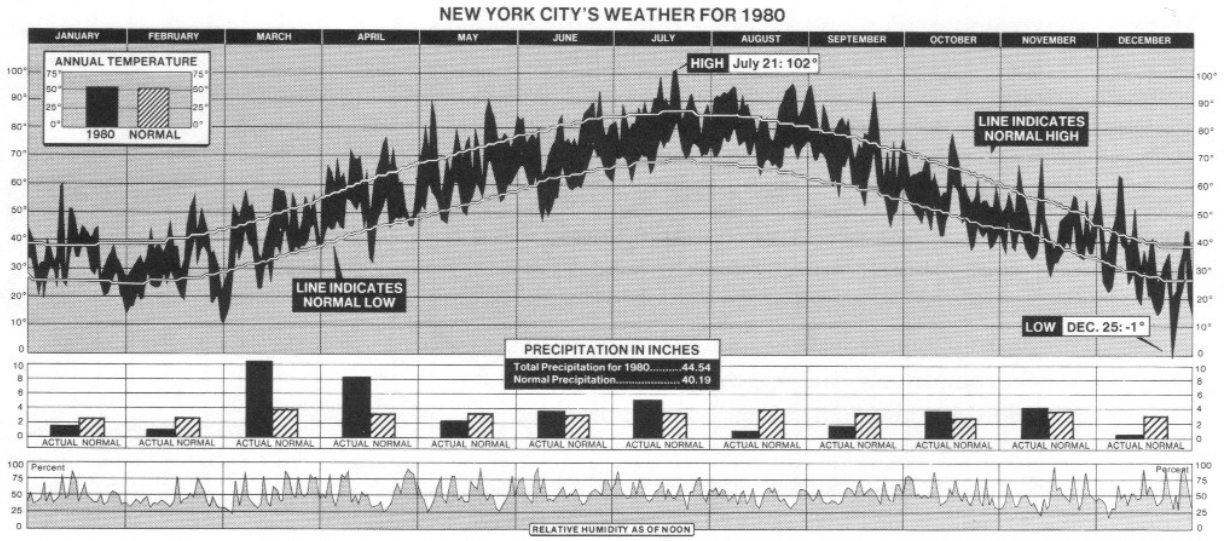
Visualization's Value



New York Times, January 11, 1981, p. 32.

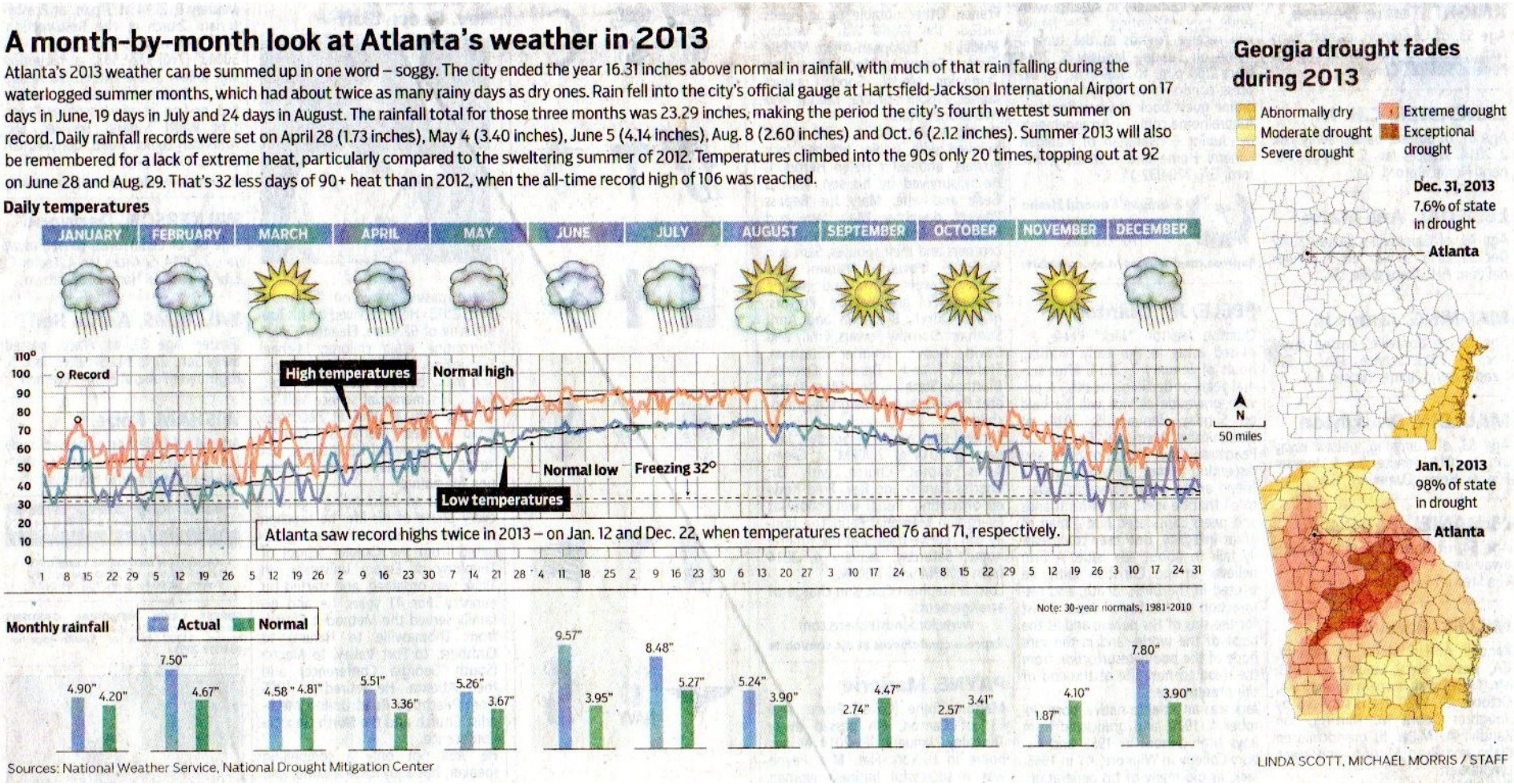
$$V_{alue} = T + I + E + C$$

Visualization's Value



New York Times, January 11, 1981, p. 32.

$$V_{alue} = T + I + E + C$$

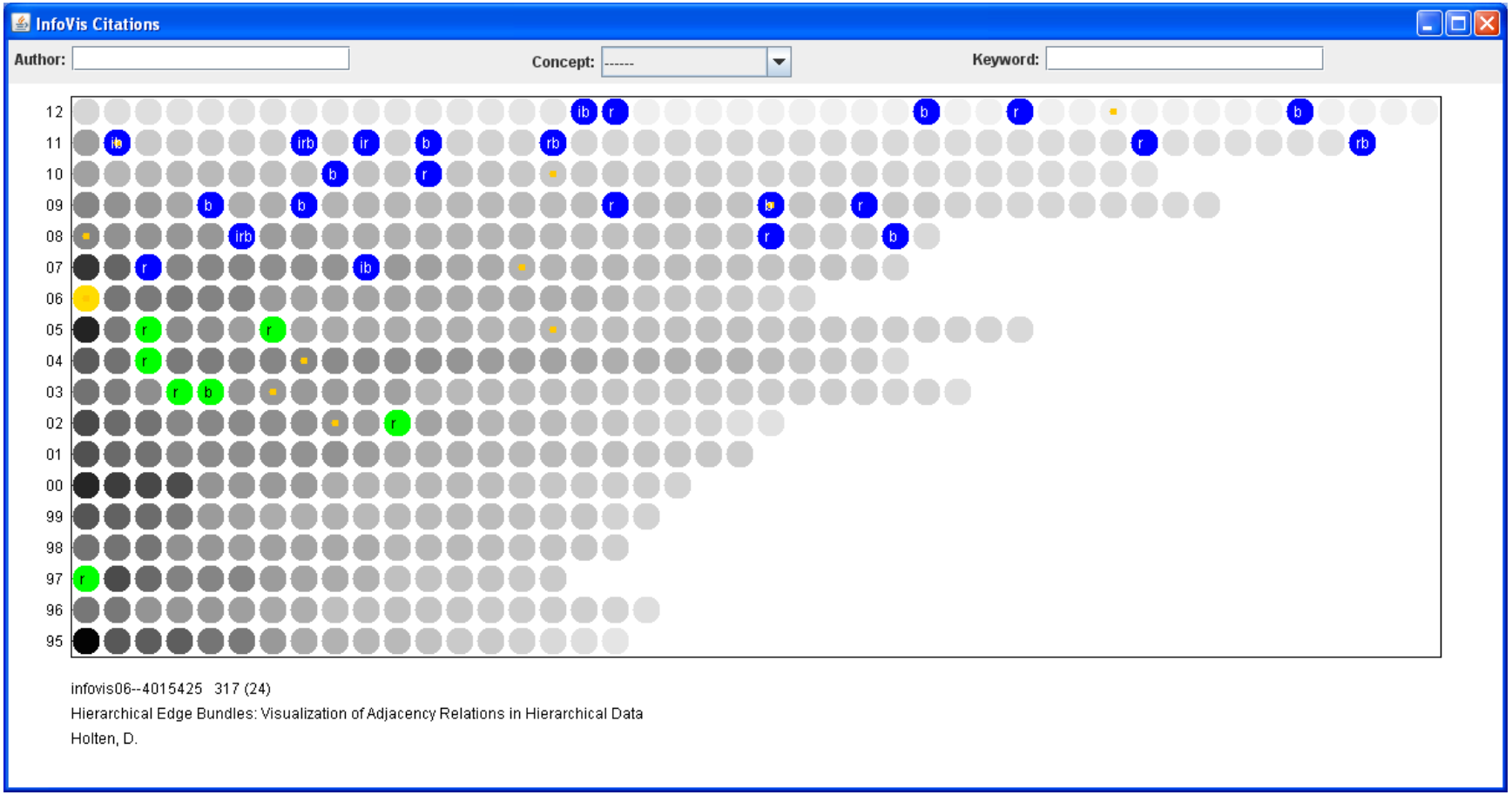


Atlanta Journal Constitution
Jan. 3, 2014

2.

http://www.cc.gatech.edu/gvu/ii/citevis

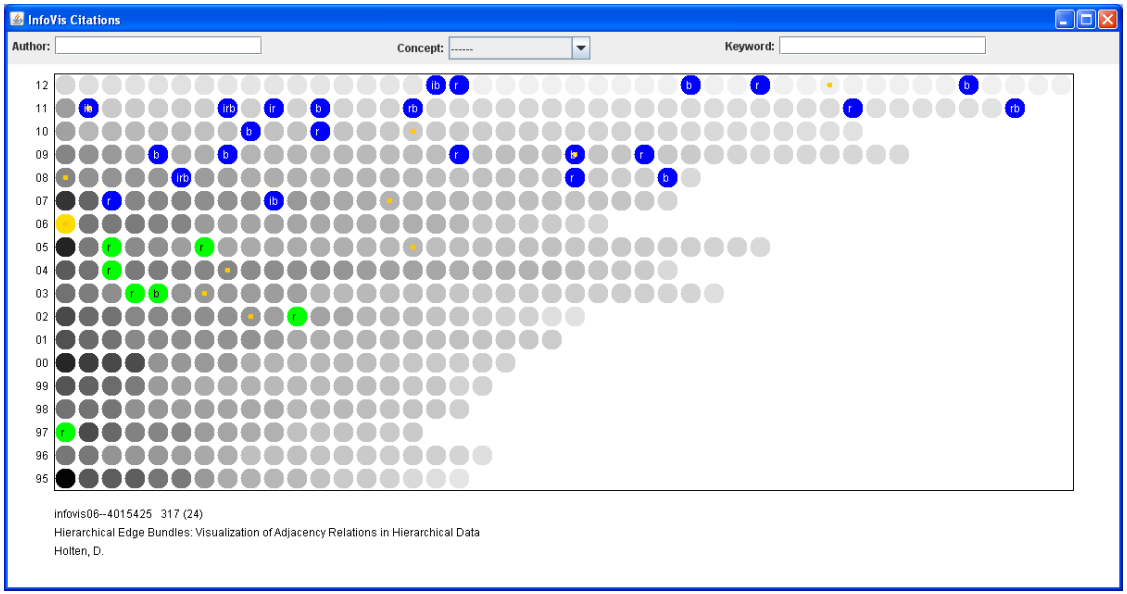
CiteVis



Demo

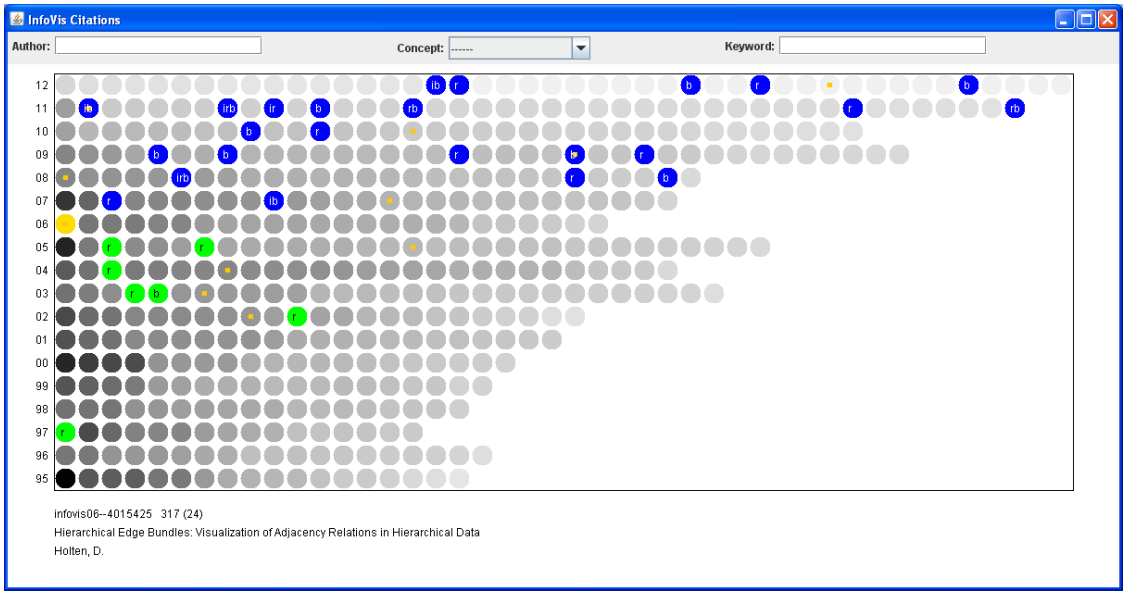
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

Visualization's Value



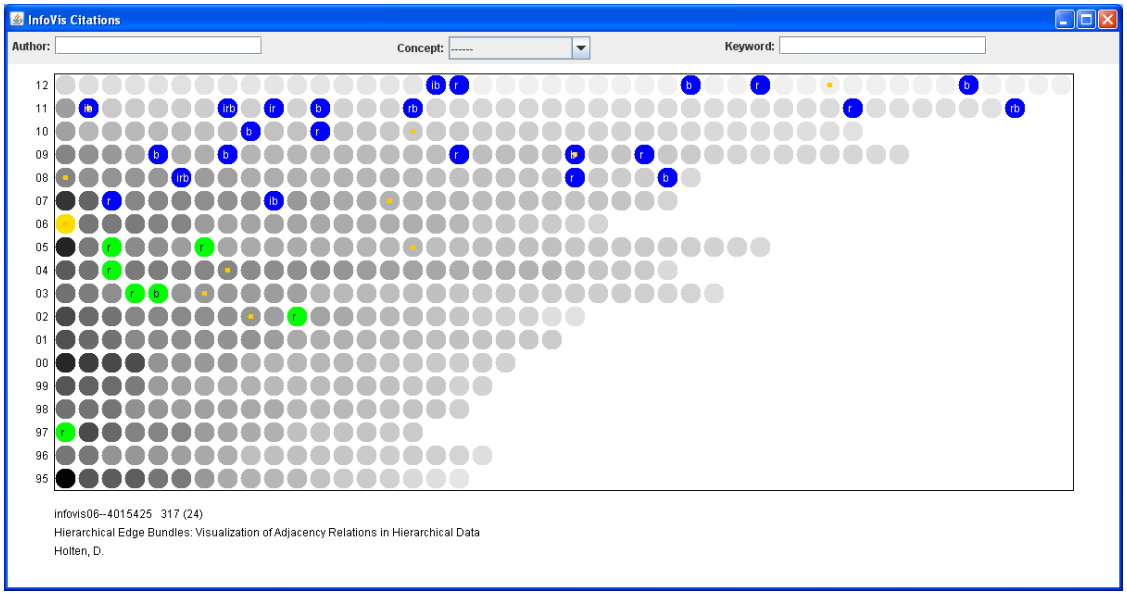
$$V_{alue} = T + I + E + C$$

Visualization's Value



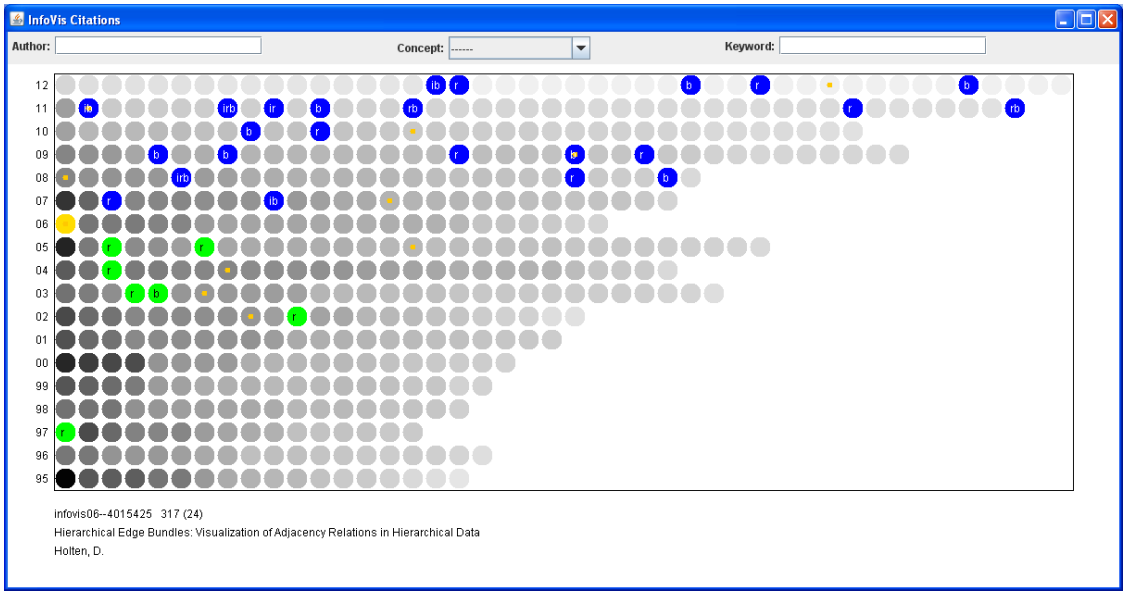
$$V_{alue} = T + I + E + C$$

Visualization's Value



$$V_{alue} = T + I + E + C$$

Visualization's Value



$$V_{alue} = T + I + E + C$$

Value

Not evaluation in traditional sense

User study with 12 grad students not so helpful

Takeaway 3

Understand & communicate the value

Key Open Problems in Visualization

Key Open Problems in Visualization

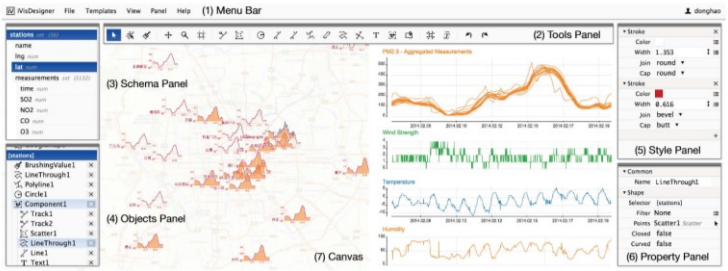
1. Assess and communicate **value**

$$V_{alue} = T + I + E + C$$

Key Open Problems in Visualization

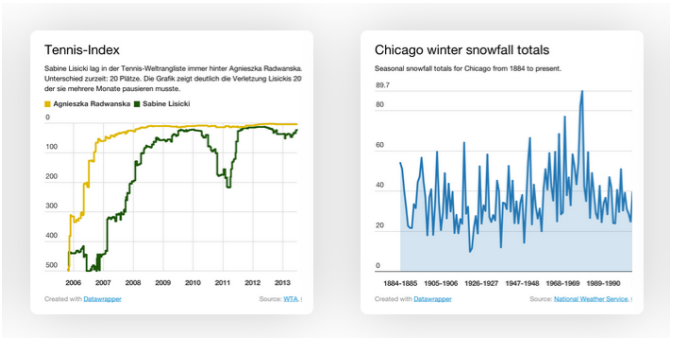
1. Assess and communicate value

2. Make the construction of visualizations easier



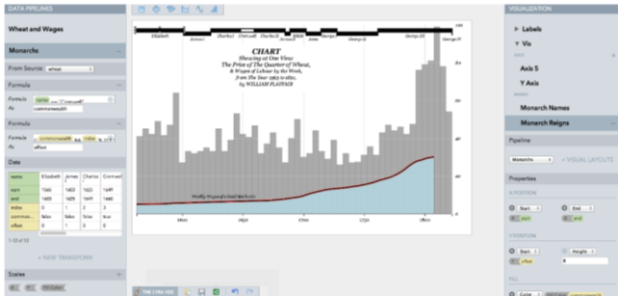
iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) '14

DataWrapper
https://datawrapper.de/





SketchStory
Lee, Kazi & Smith
TVCG (InfoVis) '13



Lyra
Satyanarayan & Heer
EuroVis '14

Key Open Problems in Visualization

1. Assess and communicate value
2. Make the construction of visualizations easier
3. Address real world “big” problems









Take Aways

Understand the purpose

Presentation & analysis, related but different

Interaction provides power, use it

Explain the value (T + I + E + C)

Thanks for your attention

Acknowledgments

- Supported by DARPA's XDATA program

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)





- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



August 28, 2014

The Value of Visualization for Exploring and Understanding Data

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu

Boeing Data Analytics Community of Excellence



information interfaces

Georgia Tech College of Computing
School of Interactive Computing


Data

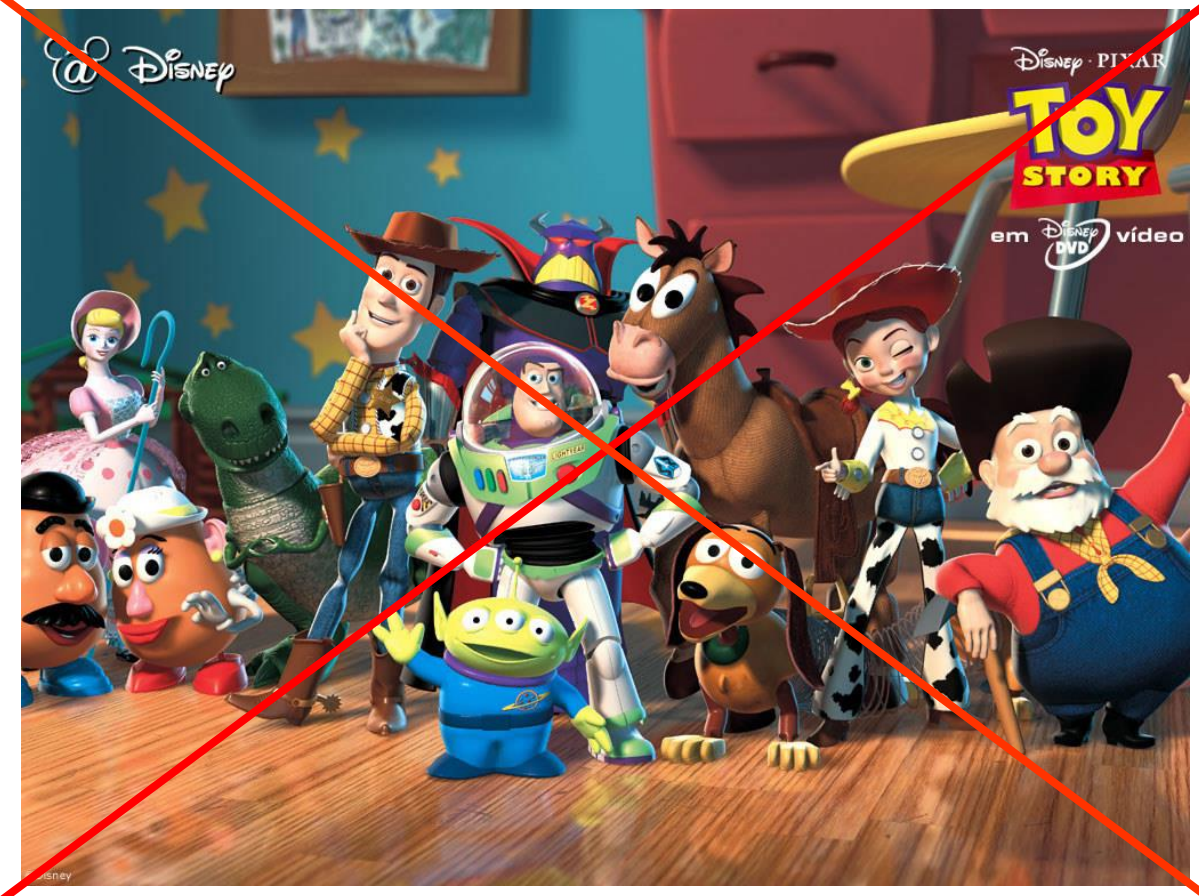
Data Visualization

Making pretty pictures?



Data Visualization

~~Making pretty pictures~~



Data Visualization

A cognitive process

Gain an understanding



Data Visualization

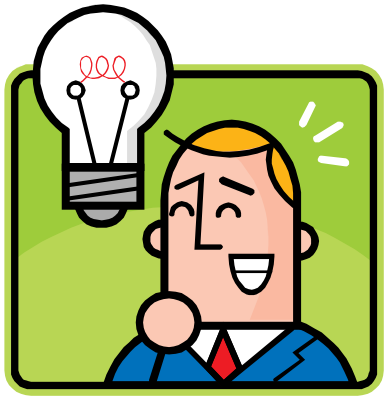
A cognitive process

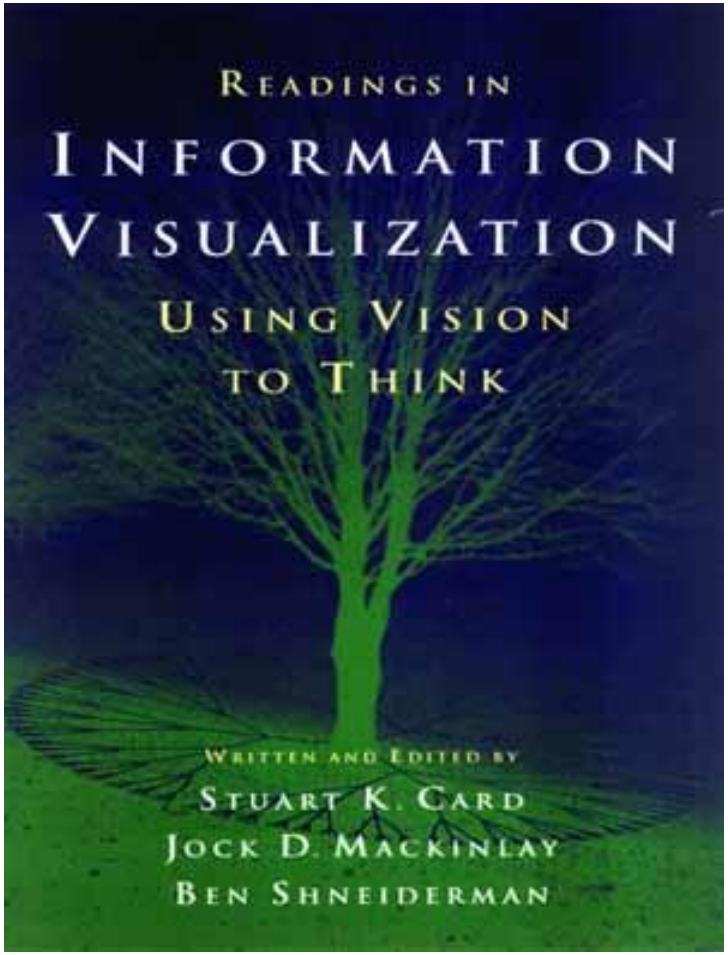
Gain an **understanding**



Challenge

Transform the *data* into *information* (understanding, insight) thus making it useful to people





Visualization

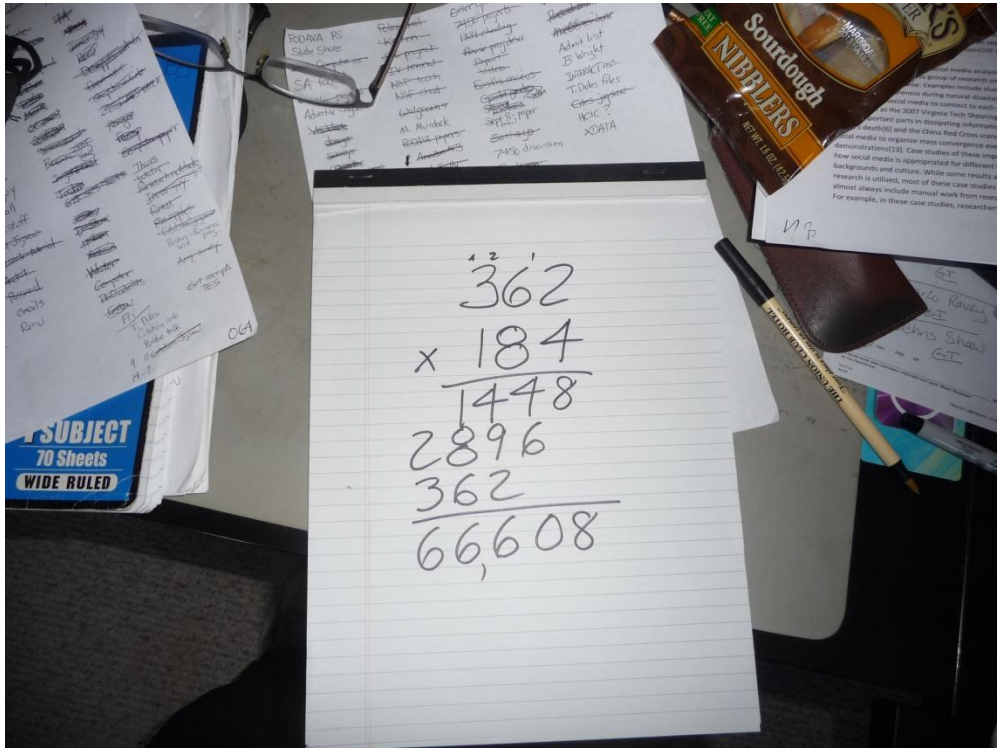
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provides a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Let Me Illustrate

An example…





Questions:

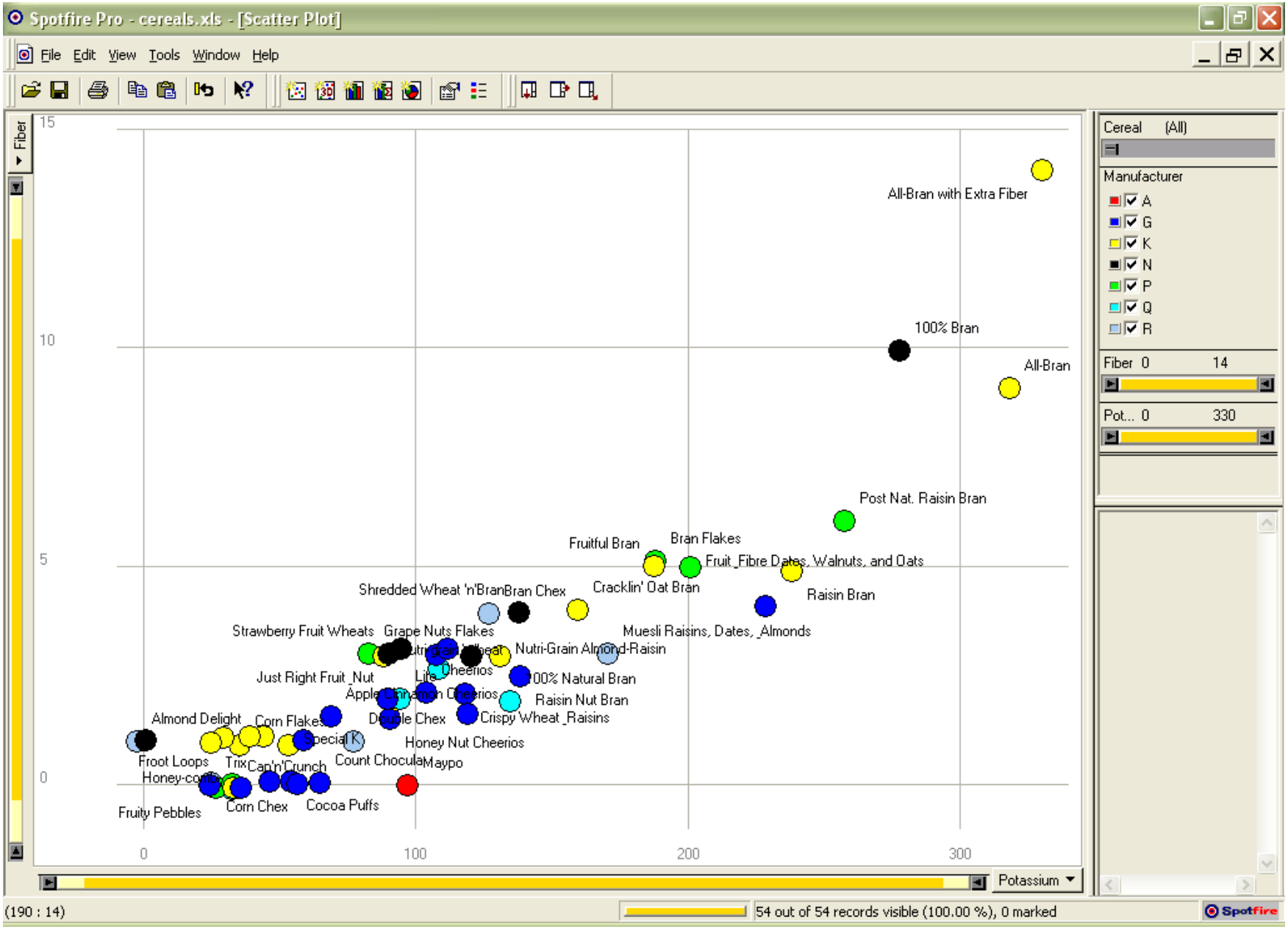
Which cereal has the most/least potassium?

Is there a relationship between potassium and fiber?

If so, are there any outliers?

Which manufacturer makes the healthiest cereals?

	A	B	C	D
1	Cereal	Manufacturer	Fiber	Potassium
2	100% Bran	N	10	280
3	100% Natural Bran	Q	2	135
4	All-Bran	K	9	320
5	All-Bran with Extra Fiber	K	14	330
6	Almond Delight	R	1	0
7	Apple Cinnamon Cheeric	G	1.5	70
8	Bran Chex	R	4	125
9	Bran Flakes	P	5	190
10	Cap'n'Crunch	Q	0	35
11	Cheerios	G	2	105
12	Cocoa Puffs	G	0	55
13	Corn Chex	R	0	25
14	Corn Flakes	K	1	35
15	Count Chocula	G	0	65
16	Cracklin' Oat Bran	K	4	160
17	Cream of Wheat (Quick)	N	1	0
18	Crispy Wheat & Raisins	G	2	120
19	Double Chex	R	1	80
20	Froot Loops	K	1	30
21	Frosted Flakes	K	1	25
22	Fruit & Fibre Dates, Wal	P	5	200
23	Fruitful Bran	K	5	190
24	Fruity Pebbles	P	0	25
25	Golden Grahams	G	0	45
26	Grape Nuts Flakes	P	3	85
27	Honey Nut Cheerios	G	1.5	90
28	Honey-comb	P	0	35
29	Just Right Fruit & Nut	K	2	95
30	Life	Q	2	95
31	Lucky Charms	G	0	55
32	Maypo	A	0	95
33	Muesli Raisins, Dates, &	R	3	170
34	Multi-Grain Cheerios	G	2	90
35	Nutri-Grain Almond-Rais	K	3	130
36	Nutri-grain Wheat	K	3	90
37	Oatmeal Raisin Crisp	G	1.5	120
38	Post Nat. Raisin Bran	P	6	260
39	Product 19	K	1	45
40	Quaker Oatmeal	Q	2.7	110
41	Raisin Bran	K	5	240
42	Raisin Nut Bran	G	2.5	140
43	Rice Krispies	K	0	35
44	Shredded Wheat	N	3	95
45	Shredded Wheat 'n'Bran	N	4	140
46	Shredded Wheat spoon	N	3	120
47	Smacks	K	1	40
48	Special K	K	1	55
49	Strawberry Fruit Wheats	N	3	90
50	Total Corn Flakes	G	0	35
51	Total Raisin Bran	G	4	230
52	Total Whole Grain	G	3	110
53	Trix	G	0	25
54	Wheaties	G	3	110
55	Wheaties Honey Gold	G	1	60

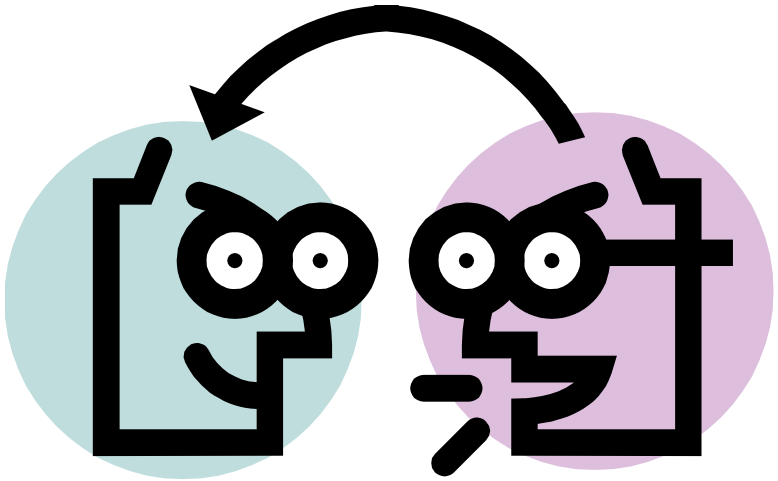


Fiber

Potassium

Even Tougher?

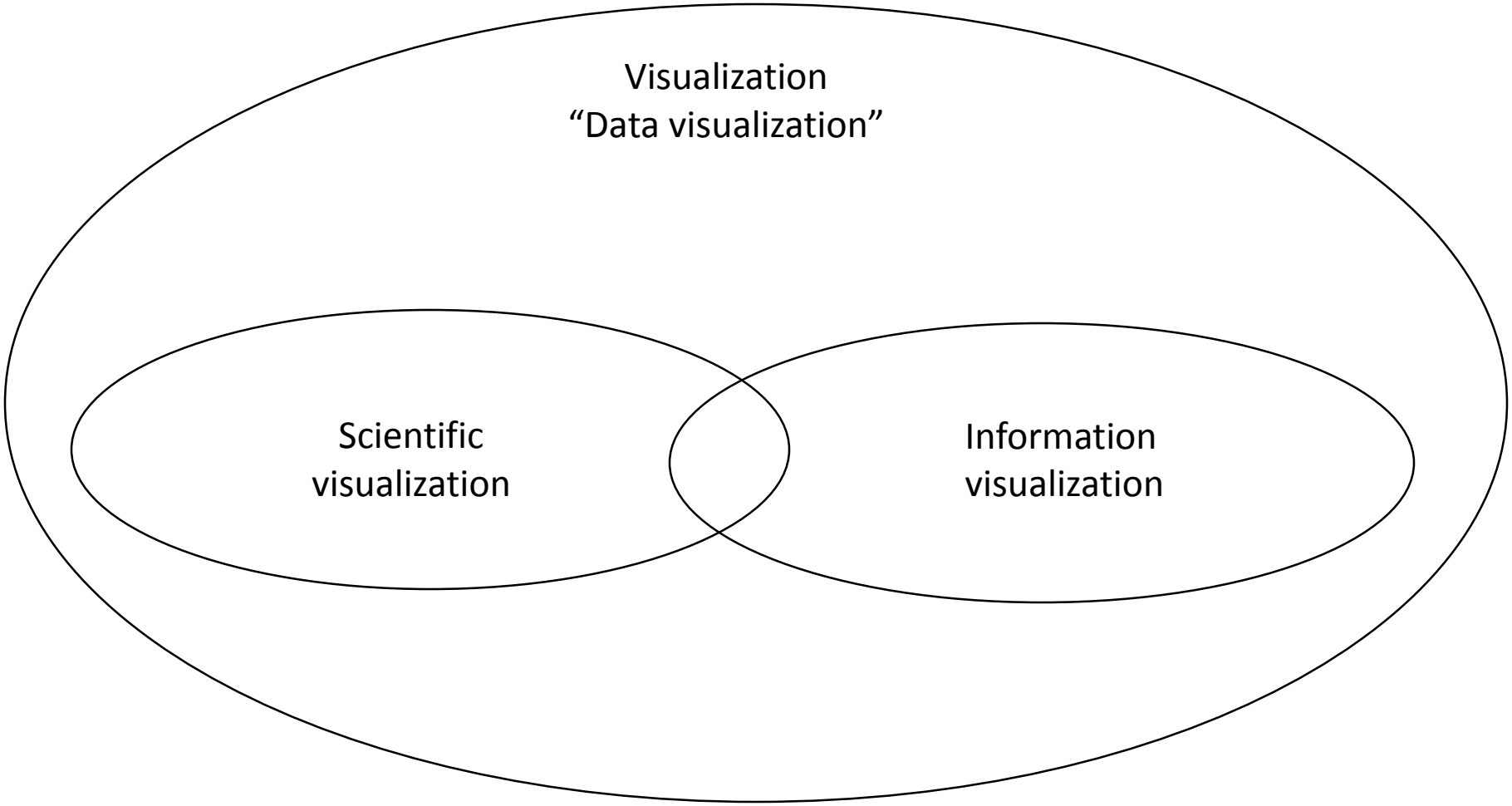
What if you could only see one cereal's data at a time?
(e.g. some websites)
What if I read the data to you?



Part of our Culture

“I see what you’re saying”
“Seeing is believing”
“A picture is worth a thousand words”

Academic Communities



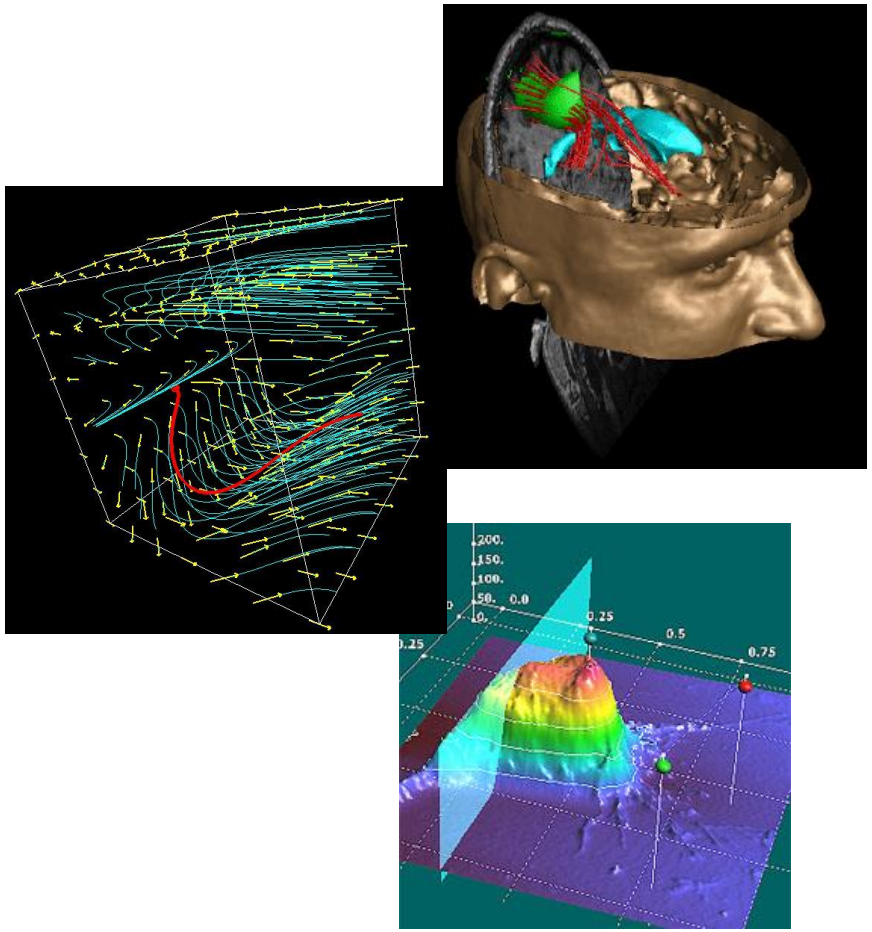
Scientific Visualization (SciVis)

Primarily relates to and represents something spatial, physical or geometric

Often 3-D

Examples

- Air flow over a wing
- Stresses on a girder
- Torrents inside a tornado
- Organs in the human body
- Molecular bonding



Not my research focus

Information Visualization (InfoVis)

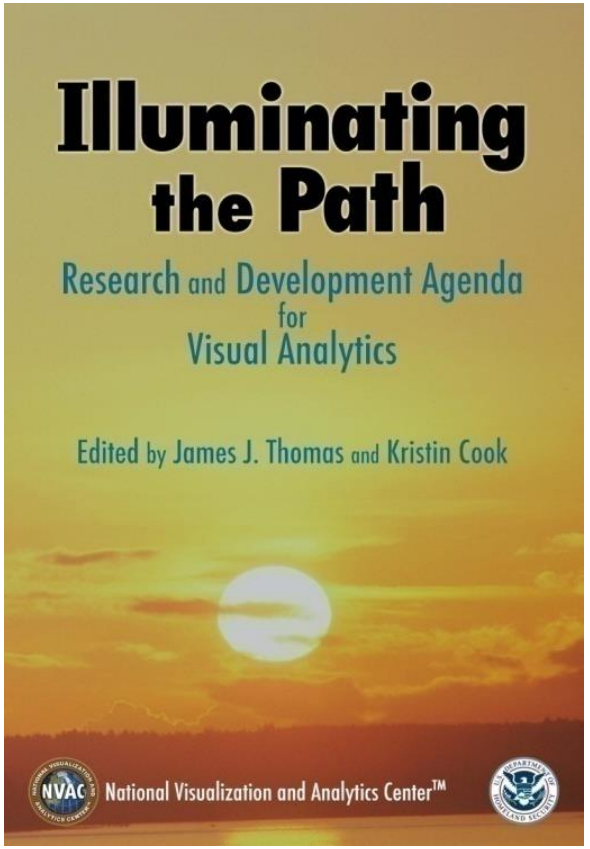
Visualizing *non-spatial* data: Items, entities, things which do not have a direct physical correspondence

- Notion of *abstractness* of the entities is important too
- Examples: baseball statistics, stock trends, connections between criminals, car attributes...

> "A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art."
>
> http://conferences.computer.org/infovis/

Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces



Available at http://nvac.pnl.gov/ in PDF form

Applications of Visualization

Presentation

Analysis

Explanatory

Exploratory

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support



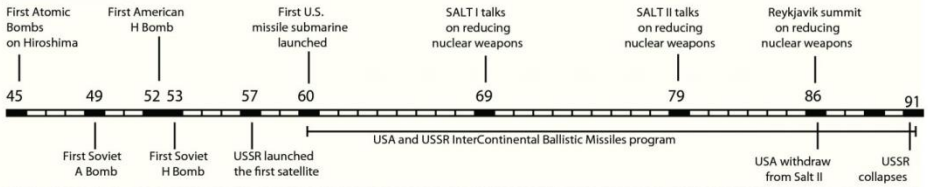
Infographics

All the rage…

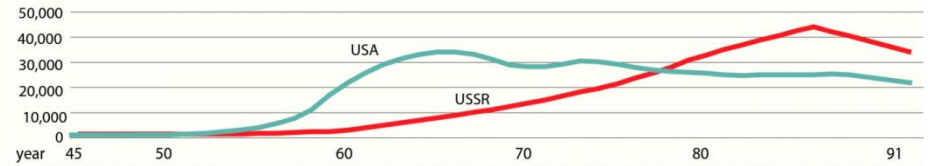
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



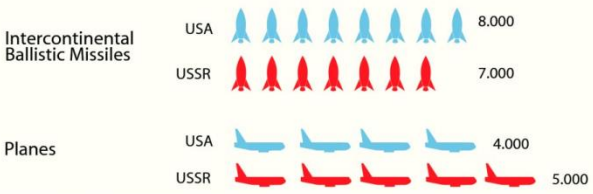
NUCLEAR STOCKPILE



ROCKET MODELS



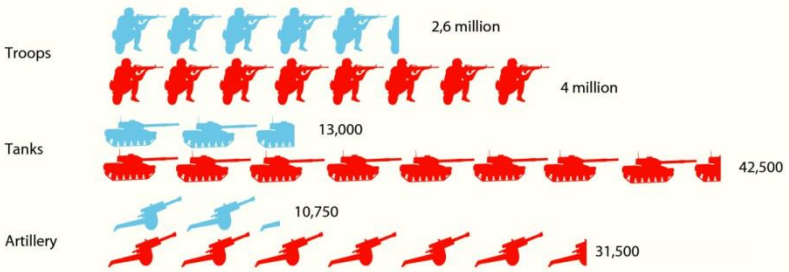
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



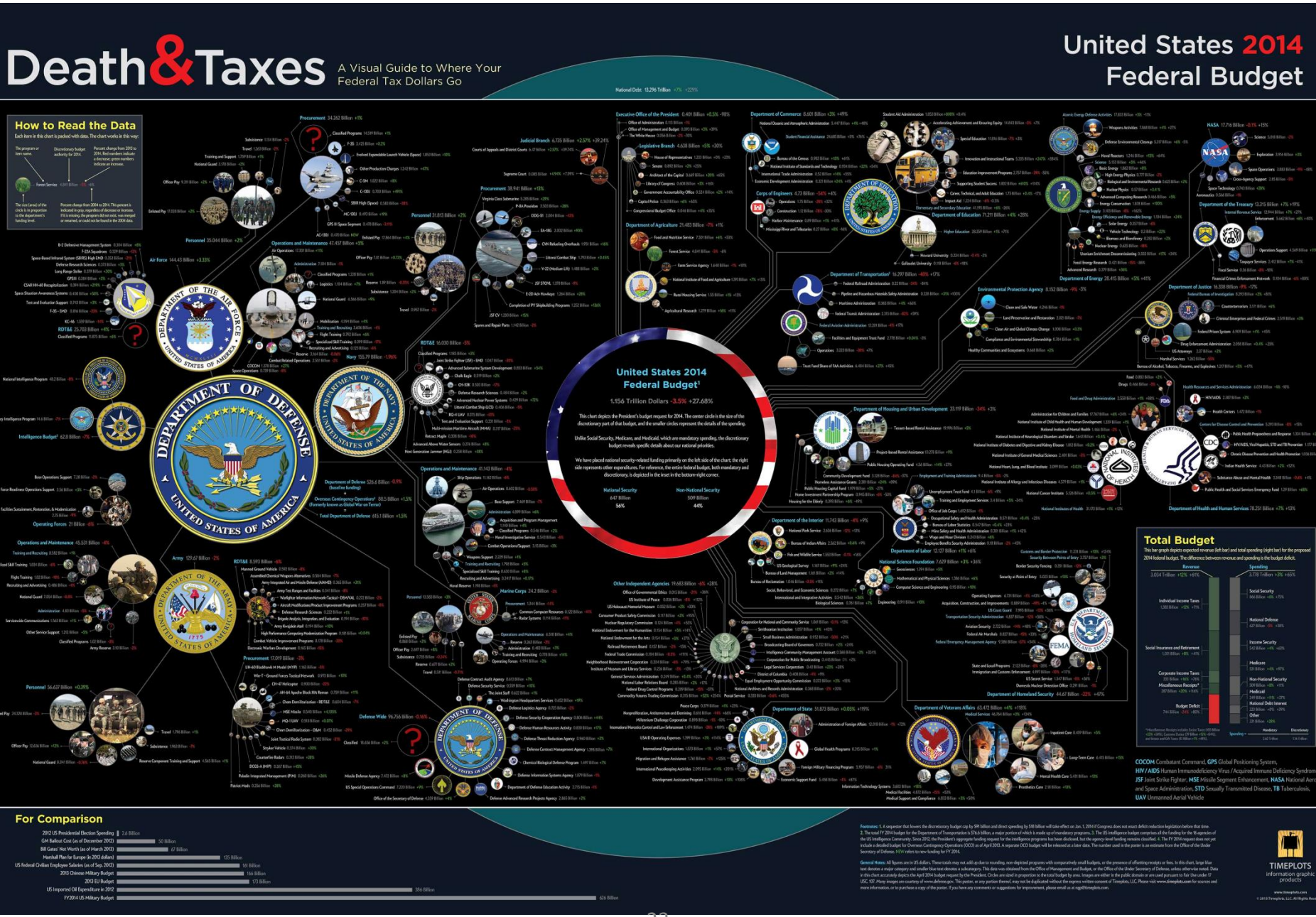


Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget



NY Times 2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons
Bonds was injured last season. He played 14 games and hit 5 homers.

600
400
200
0
14th season
5 seasons
10
15

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

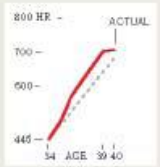
Aaron
Actual homers slightly outpace projected homers for five seasons.



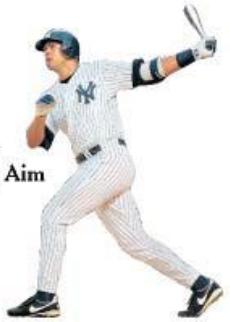
Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.



Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

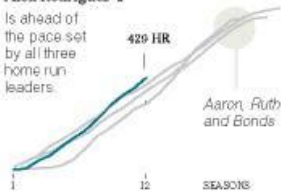


Note: Ages as of July 1 of each season.

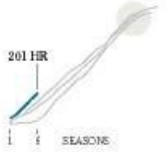


Others Taking Aim

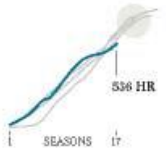
Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

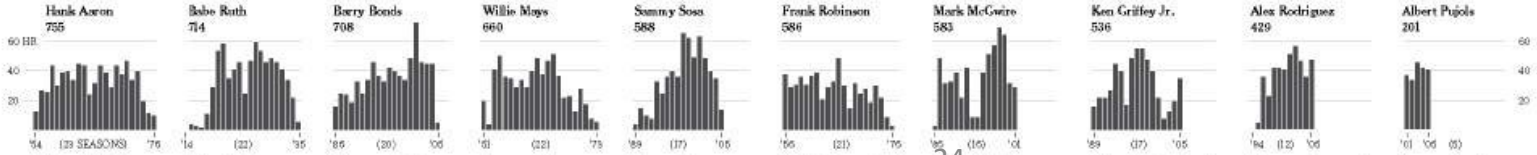


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

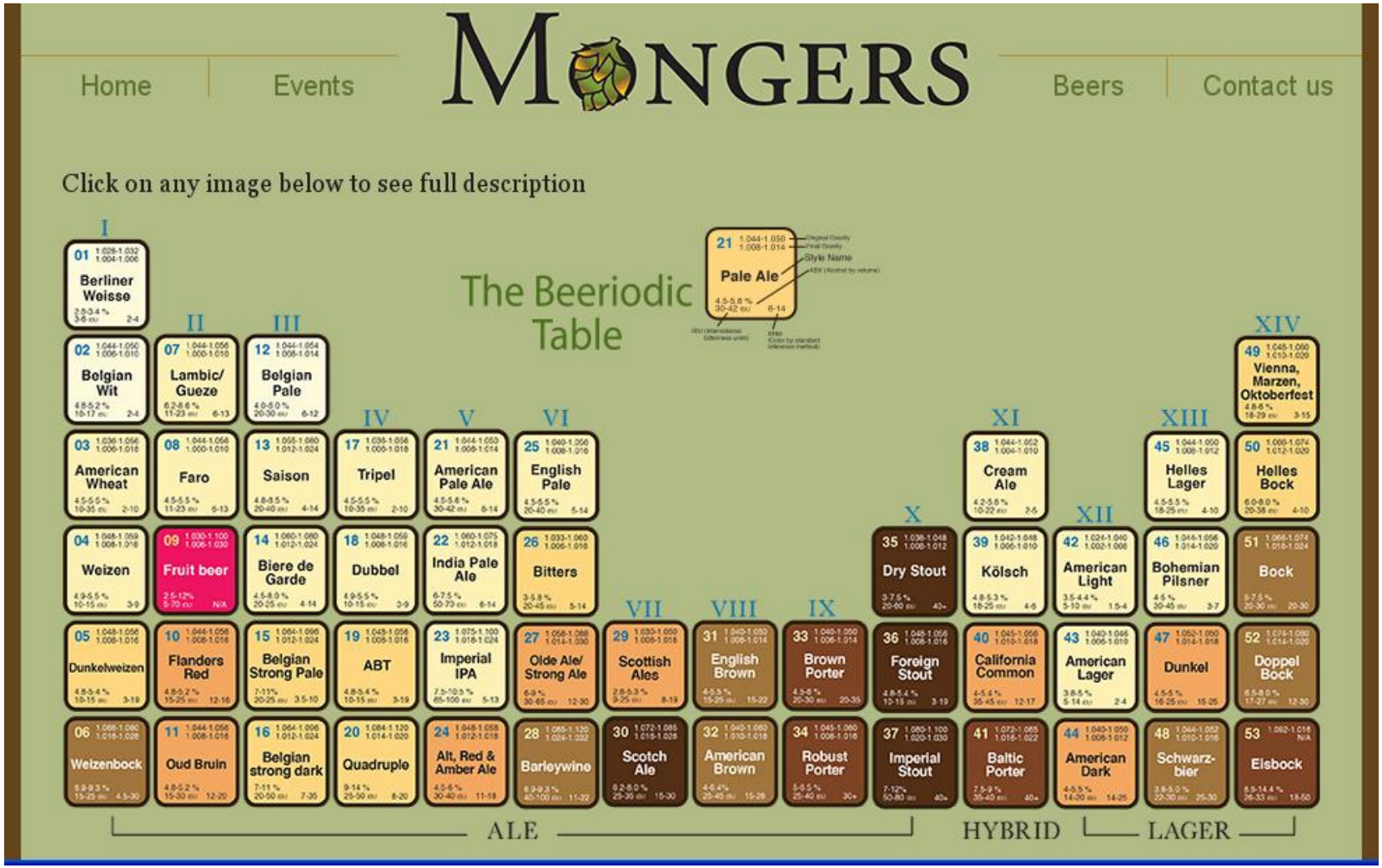


Differing Paths to the Top of the Charts

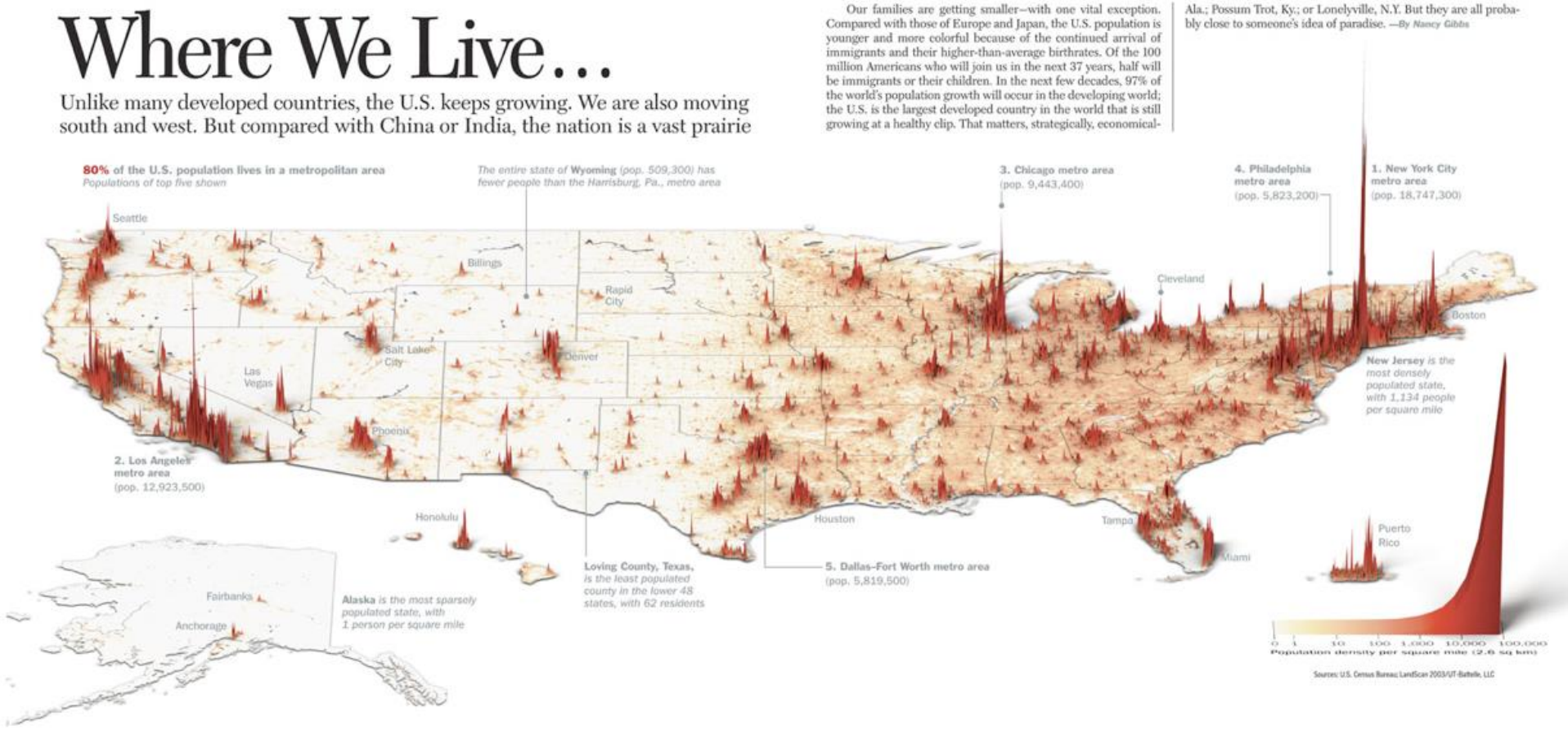
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



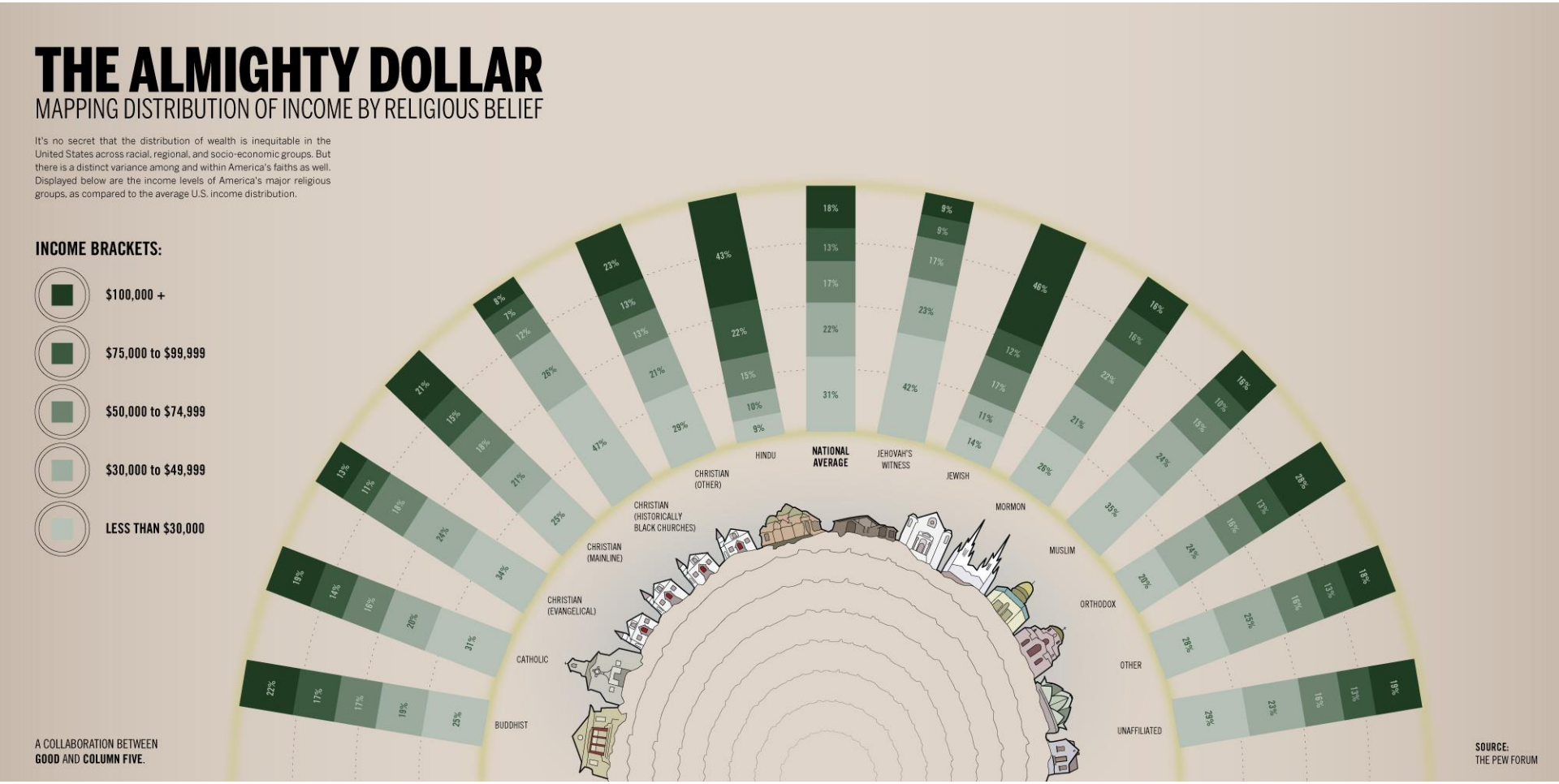
Amanda Cox and Joe Ward/The New York Times



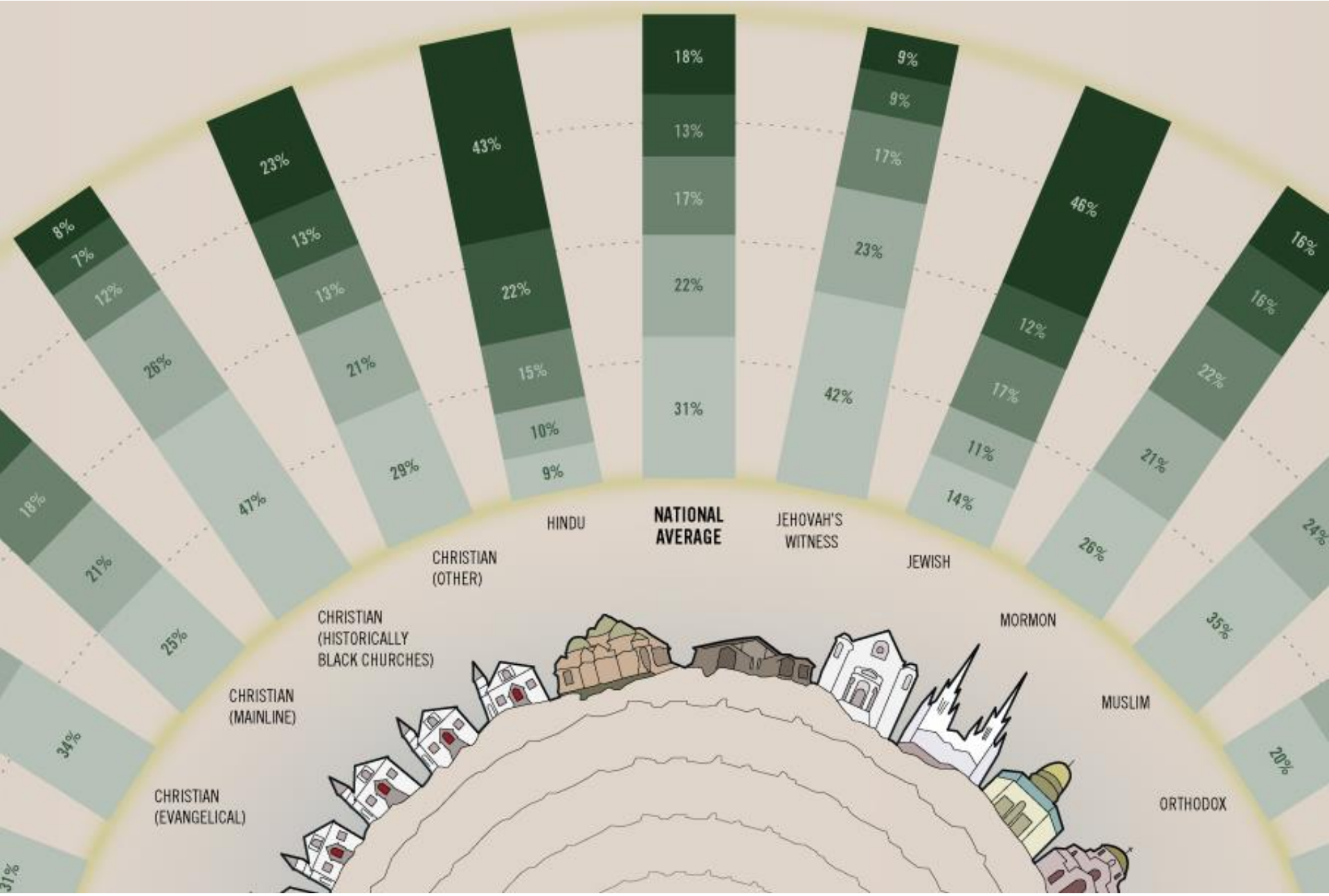
http://thebeermongers.com/beers/



http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html



http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg


8%
7%
12%
26%
41%
CHRISTIAN (EVANGELICAL)
18%
21%
25%
34%
23%
13%
13%
21%
29%
CHRISTIAN (MAINLINE)
43%
22%
15%
10%
9%
CHRISTIAN (HISTORICALLY BLACK CHURCHES)
18%
13%
17%
22%
31%
NATIONAL AVERAGE
9%
9%
17%
23%
42%
HINDU
46%
12%
17%
11%
14%
CHRISTIAN (OTHER)
16%
16%
22%
21%
26%
JEWISH
24%
35%
20%
ORTHODOX
MORMON
MUSLIM
JEHOVAH'S WITNESS

Frequent presentation goals

Clarify

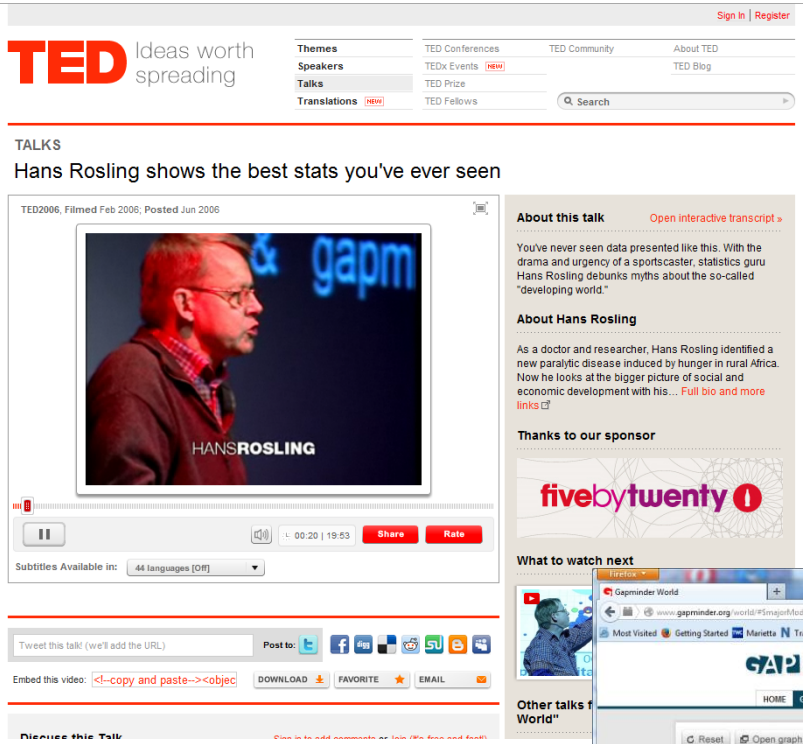
Focus

Highlight

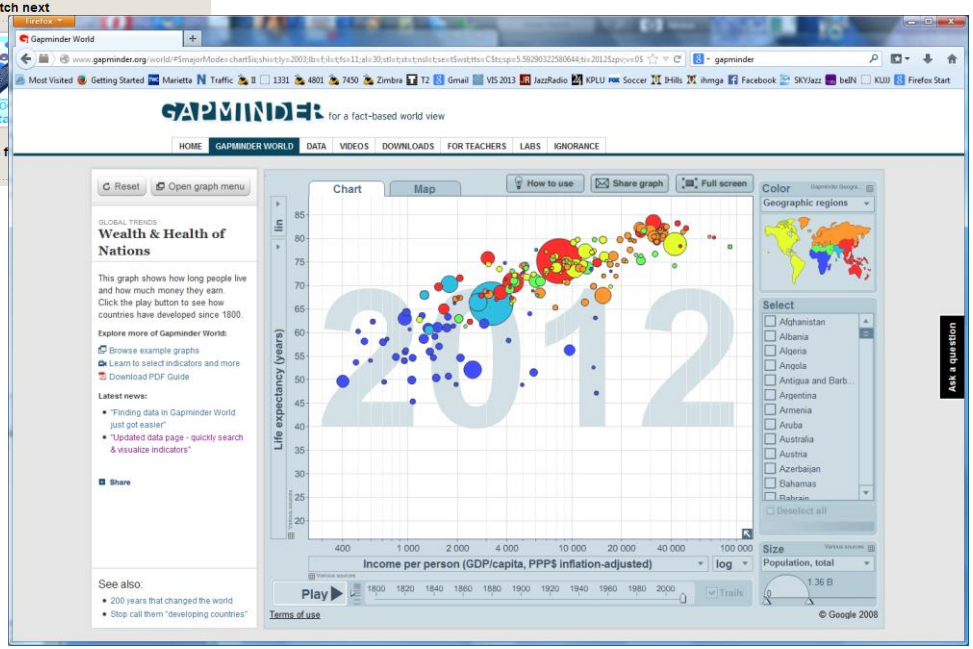
Simplify

Persuade

May just show a few variables
and/or a subset of the data cases



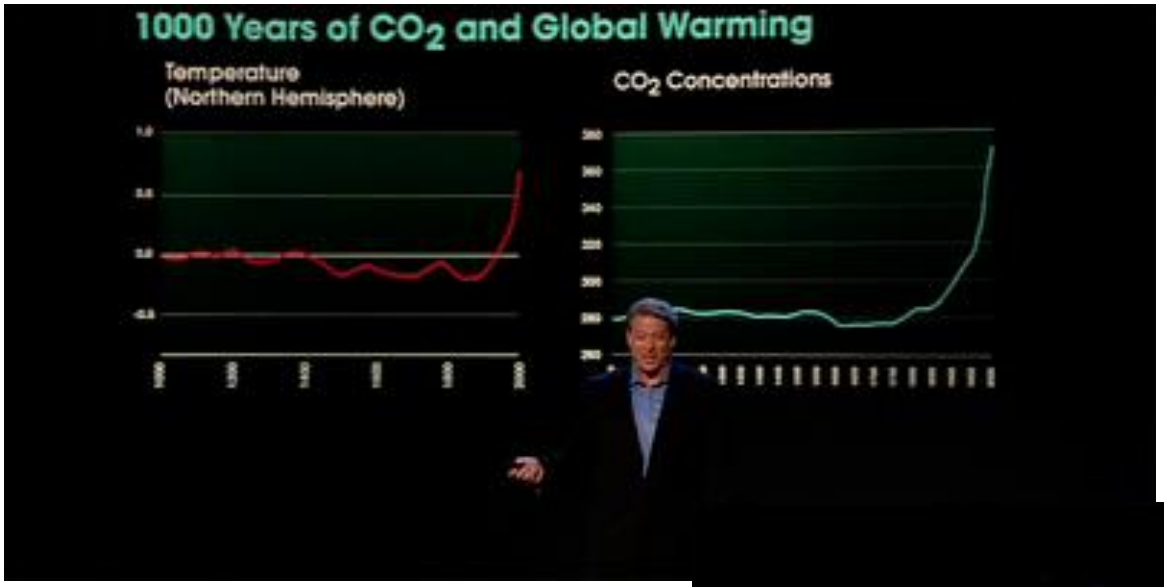
Hans Rosling
Gapminder

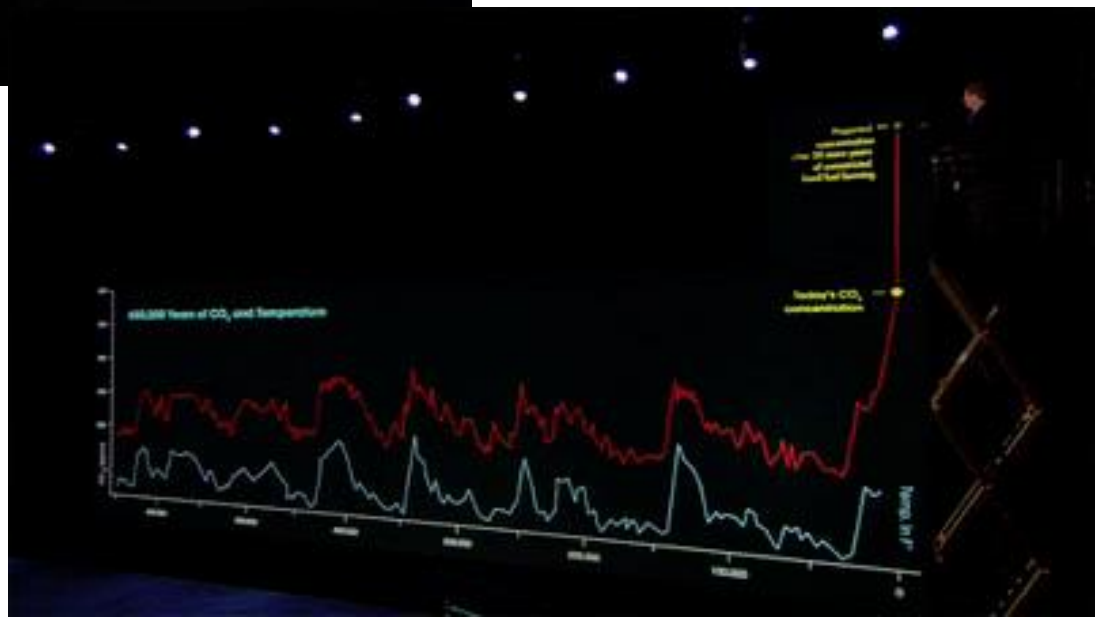


2006

http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

An Inconvenient Truth





Gore made extensive use of data graphics



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

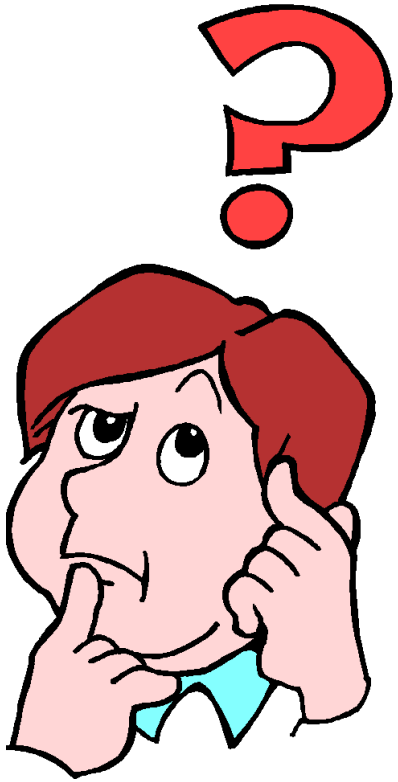
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context



Investigative analysis



COVER FEATURE

Visual Analytics Support for Intelligence Analysis

Carsten Görg, *University of Colorado*
Younah Kang, *Google*
Zhicheng Liu, *Stanford University*
John Stasko, *Georgia Tech*

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open source news articles. Insights from visual analytics projects and a hypothetical scenario show the potential of visual analytics to aid these investigations.

Intelligence analysis has been a key application domain for visual analytics since the US Department of Homeland Security created the National Visualization and Analytics Center in 2004. An initial research roadmap[1] described the challenges and goals of this new domain and identified tasks, data, and analytical scenarios focused on homeland security and terrorism prevention. A passage from the roadmap characterizes the grand challenge for visual analytics systems:

> The analysis of overwhelming amounts of disparate, conflicting, and dynamic information is central to identifying and preventing emerging threats, protecting our borders, and responding in the event of an attack or other disaster. This analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations to both detect the expected and discover the unexpected.

Much of this challenge remains,[2] although visual analytics technologies certainly hold great promise. A few commercial visual analytics tools for intelligence analysis are emerging, such as Analyst's Notebook from IBM i2 (www.ibm.com/software/industry/i2/software), nSpace from Oculus (www.oculusinfo.com/nspace), and Palantir's suite of systems (www.palantir.com). However, continued progress depends heavily on gaining a deeper understanding of intelligence analysis and the role of an analyst, as well as clarifying how visual analytics can help investigators. These insights must inform any design of any visual analytics system that aims to support intelligence analysis.

The "Investigative Scenarios" sidebar describes the kinds of analysis that are prevalent in the intelligence domain. The scale, diversity, and complexity of the information to be explored make such analyses cognitively demanding. Information is often narrative text, not quantitative data, and as such is not as amenable to automated analysis. Particularly challenging are the "pain points" in the intelligence process—the cost of scanning, recognizing (assessing), and selecting items for further attention; a limited attention span for evidence and hypotheses; and the difficulty of generating alternative hypotheses.[3]

To better understand requirements for visual analytics support, we conducted a series of research projects on intelligence analysts. The projects, which spanned 2008 to 2015, included an observational study to better understand the intelligence analysis process and its characteristics. During this time, we also developed Jigsaw (www.cc.gatech.

30 COMPUTER Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

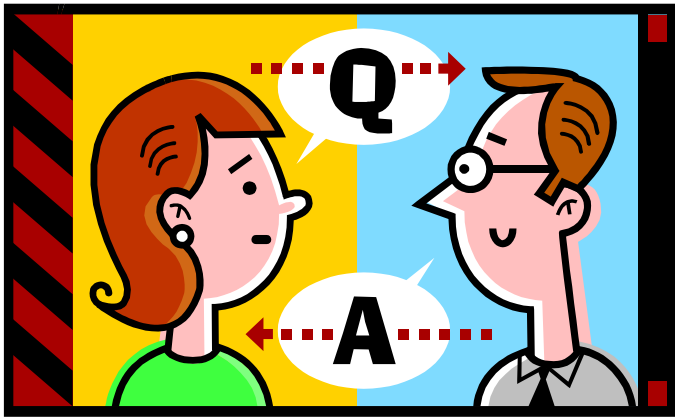
Görg, Kang, Kiu, Stasko
IEEE Computer '13

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

Interaction is Vital

Engage in a dialog with your data



1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect



Toward a Deeper Understanding of the Role of Interaction in Information Visualization

Ji Soo Yi, Youn ah Kang, John T. Stasko, *Member, IEEE*, and Julie A. Jacko

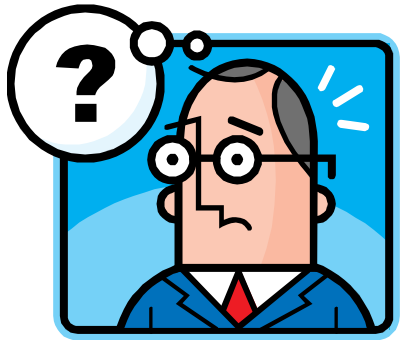
Yi, Kang, Stasko, Jacko
TVCG (InfoVis) ‘07

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

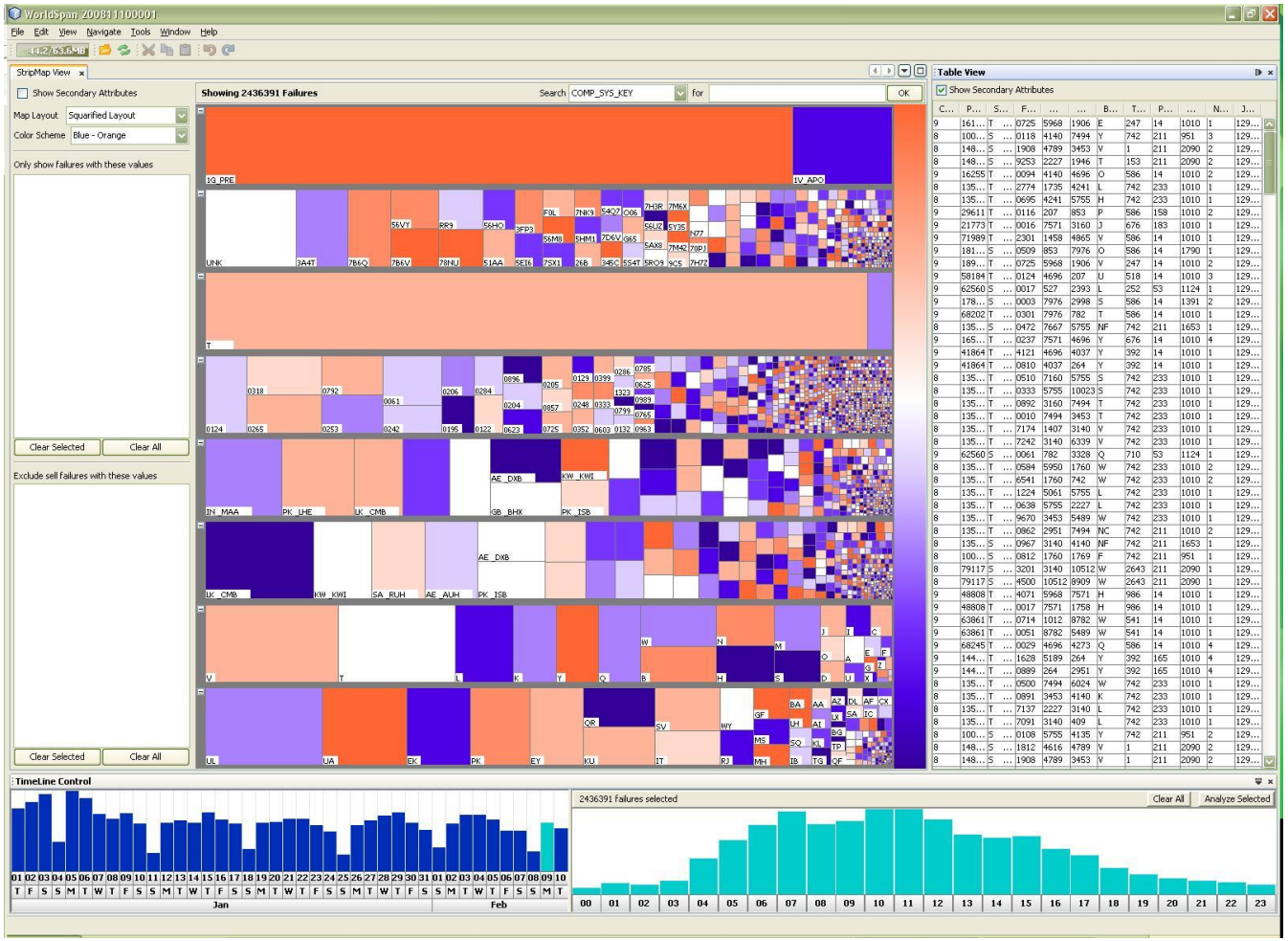
Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Some examples from my group's research

SellTrend



Web-based airline booking requests

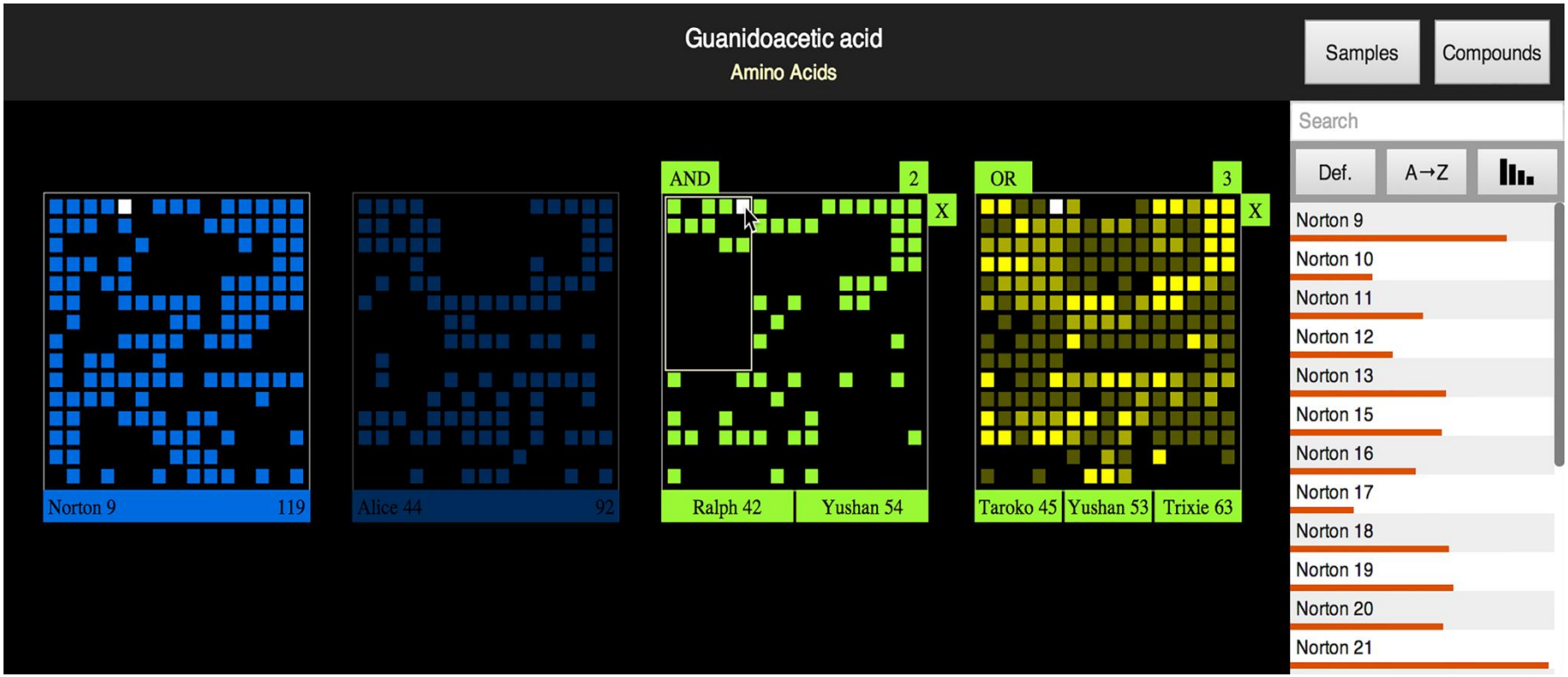
Help recognize irregularities and analyze what the problem is

http://www.cc.gatech.edu/gvu/ii/selltrend/

Liu, Sullivan, Stasko
TVCG (InfoVis) '09

OnSet

http://www.cc.gatech.edu/gvu/ii/setvis/

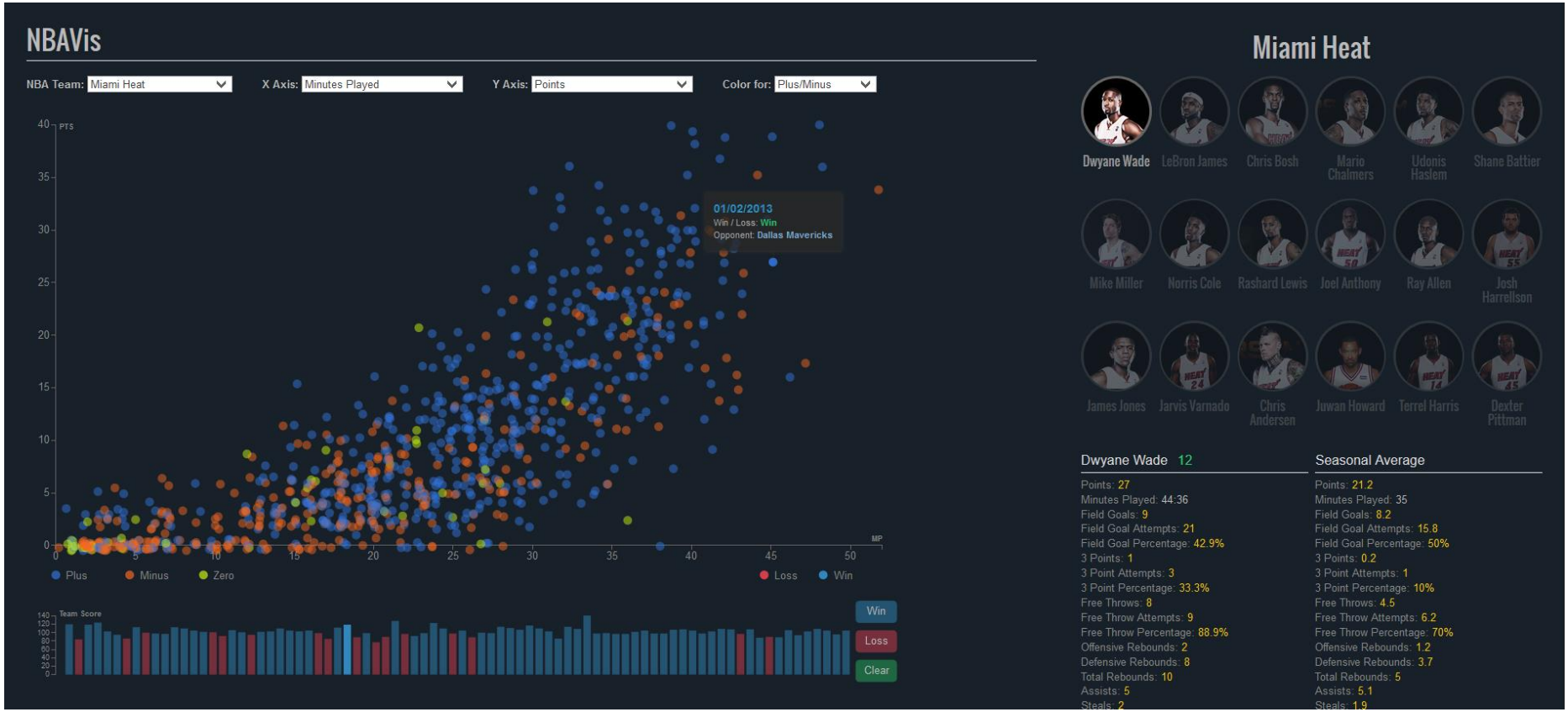


Demo

Sadana, Major, Dove, & Stasko

NBA Player and Team Data

http://www.cc.gatech.edu/gvu/ii/sportvis/nbaVis/

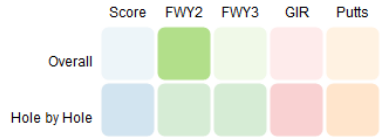


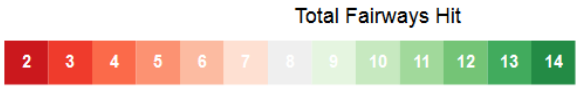
Project by Fengbo Li

Demo

PGA Tour Season Recap

Using ShotLink data





2013 | Phil Mickelson

Sort by Date	Round 1	Round 2	Round 3	Round 4	Position	Total	Strokes
Humana Challenge					T37	-17	271
Farmers Insurance Open					T51	E	288
Phoenix Open					1	-28	256
Pebble Beach Pro-Am					T60	-1	285
Northern Trust Open					T21	-3	281
WGC-Cadillac Champ.					T3	-14	274
Arnold Palmer Invitational					CUT	+8	152
Houston Open					T16	-10	278
The Masters					T54	+9	297
Wells Fargo Champ.					3	-7	281
PLAYERS Champ.					CUT	+1	145
St. Jude Classic					T2	-10	270
U.S. Open					T2	+3	283
The Greenbrier					CUT	+2	142
Open Champ.					1	-3	281
Bridgestone					T21	+1	281
PGA					T72	+12	292
Barclays					T6	-9	275
Deutsche Bank					T41	-8	276
BMW Champ.					T33	-1	283

with Anand Sainath, Rahul Basole

“Putting the pieces together”

Jigsaw

Computational analysis of documents’ text

- Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

- Views are highly coordinated



Help “investigators” explore, analyze and understand large document collections

Academic Papers

Health Forums

Product Reviews

Police Reports



Designing and Implementing an Interac
Scatterplot Visualization for a Tablet Con

Ramik Sadana
Georgia Institute of Technology
ramik@gatech.edu

John Stasi
Georgia Institute of T
stasko@cc.gate

ABSTRACT

Tablet computers now offer screen sizes and computing capabilities that are competitive with traditional desktop PCs. Their popularity has grown tremendously, but we are just beginning to see information visualization applications designed for this platform. One potential reason for this limited development is the challenge of designing and implementing a multi-touch interface for visualizations on mobile, tablet devices. In this work, we identify the primary challenges that touch screen interactions pose for information visualization applications. We explore the design space of multi-touch interactions for visualizations and present a prototype information visualization application using a specific technique, a dynamic scatterplot, for an iPad.

Categories and Subject Descriptors
H.5.0 [Information Interfaces and Presentation]: General

General Terms
Design, Algorithms, Human Factors.

Keywords
Information visualization, multi-touch interaction, scatterplot, tablet computer, gesture.

1. INTRODUCTION

The popularity of tablet computers has grown tremendously in recent years. Devices such as the Apple iPad, Amazon Kindle, and Microsoft Surface constitute a significant portion of computer sales today and a wide variety of applications have been tailored to tablet platform. However, one area with just a few initial applications for tablets is information visualization. Though popular commercial systems such as Tableau and Spotfire have introduced tablet versions in the last two years, these systems still feel much like a port of a desktop application. They do not yet leverage the set of gestural interactions that touch-based interfaces provide and do not offer a rich, multi-touch interface in the style of other tablet applications.

We speculate that the limited development of information visualization applications for tablets has resulted from the challenge of designing and implementing the interface. First, most tablets still provide a smaller screen size, which is a limitation for data visualization. Perhaps more importantly, information visualization applications generally have many small visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the authors must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AVI'14, May 27 - 29 2014, Como, Italy
Copyright is held by the authors. Publication rights licensed to ACM.
ACM 978-1-4503-2775-6/14/05...$15.00.

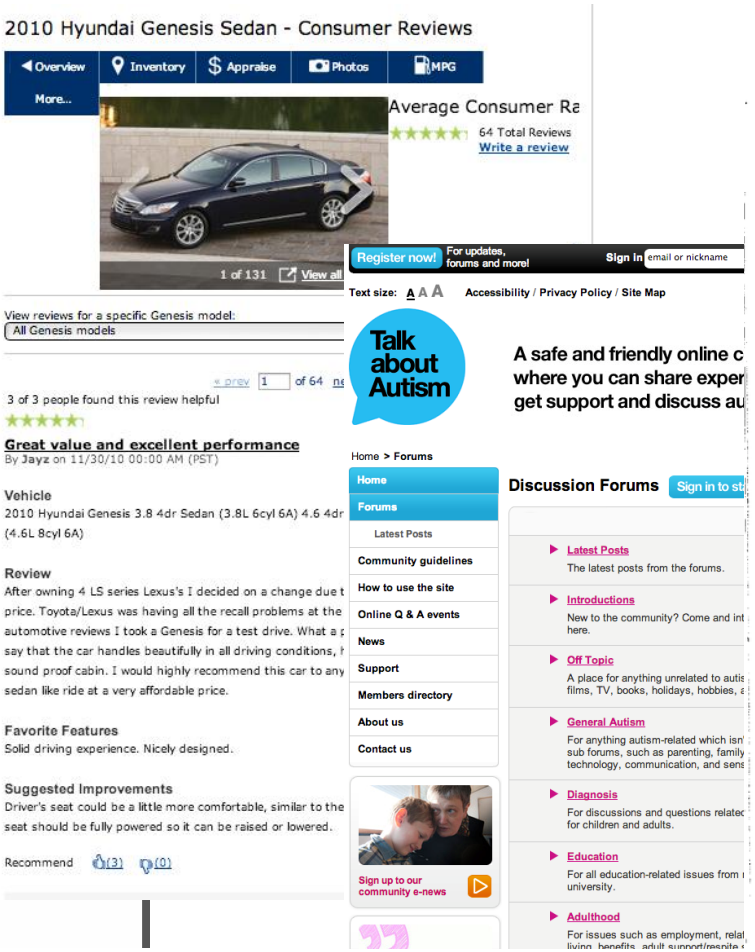
objects to select and manipulate, interactive widgets such as buttons, boxes. Desktop information visualiz make extensive use of a WIMP (w interface. Translating visualization operations to a finger-directed multi keyboard and mouse is simply a very

Some recent research has begun however. One particular avenue development of multi-user, multi applications for tabletop computers [1 explored the design space for s developing for tablet computers is s tabletop computers. Baur et al. [2] rec tablet interface for a StreamGraph-ty This application is much more alig create. Our specific focus, however, used visualization technique. Rzeszot application is a similar technique to driven interactions.

Rather than viewing the movement applications to multi-touch tablets software port, we see this problem challenge. Initial research [25] has interaction design space – for any information visualization syster implementations make sense and Furthermore, we will show that a gestures from tablet applications p applied to data visualizations.

In this work, we further explo implement a working information vi tablet PC. We chose a dynamic sca technique of our focus. Scatterplots a in information visualization; many o suite of provided techniques. Scatterp in data visualization, including the FilmFinder system [1] that led to t Our project starts from "ground zero an existing system. Our application advantage of, and to address all the l a tablet interface using only fingers a

The contributions of our work are sections of this article. By studying commercial implementations of a dyn set of tasks and interactive operation would support (Section 3). Next, we the tasks and describe how those c design (Section 4). We ultimately c multi-touch operations for each task, for each, propose multiple potenti many of the options in a proto application for the iPad (Section 5). I paper may not convey all the operati overview summarizing the applicatio



2010 Hyundai Genesis Sedan - Consumer Reviews

Overview | Inventory | Appraise | Photos | MPG | More...

Average Consumer Ra
64 Total Reviews
Write a review

1 of 131 View all

Register now! For updates, forums and more! Sign in email or nickname

Text size: A A A Accessibility / Privacy Policy / Site Map

View reviews for a specific Genesis model:
All Genesis models

« prev 1 of 64 ne

3 of 3 people found this review helpful

Great value and excellent performance
By Jayz on 11/30/10 00:00 AM (PST)

Vehicle
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 4dr (4.6L 8cyl 6A)

Review
After owning 4 LS series Lexus's I decided on a change due t price. Toyota/Lexus was having all the recall problems at the automotive reviews I took a Genesis for a test drive. What a p say that the car handles beautifully in all driving conditions, h sound proof cabin. I would highly recommend this car to any sedan like ride at a very affordable price.

Favorite Features
Solid driving experience. Nicely designed.

Suggested Improvements
Driver's seat could be a little more comfortable, similar to the seat should be fully powered so it can be raised or lowered.

Recommend (3) (0)

Talk about Autism

A safe and friendly online c where you can share exper get support and discuss au

Home > Forums

Home
Forums
Latest Posts
Community guidelines
How to use the site
Online Q & A events
News
Support
Members directory
About us
Contact us

Discussion Forums Sign in to st

Latest Posts
The latest posts from the forums.

Introductions
New to the community? Come and int here.

Off Topic
A place for anything unrelated to autis films, TV, books, holidays, hobbies,

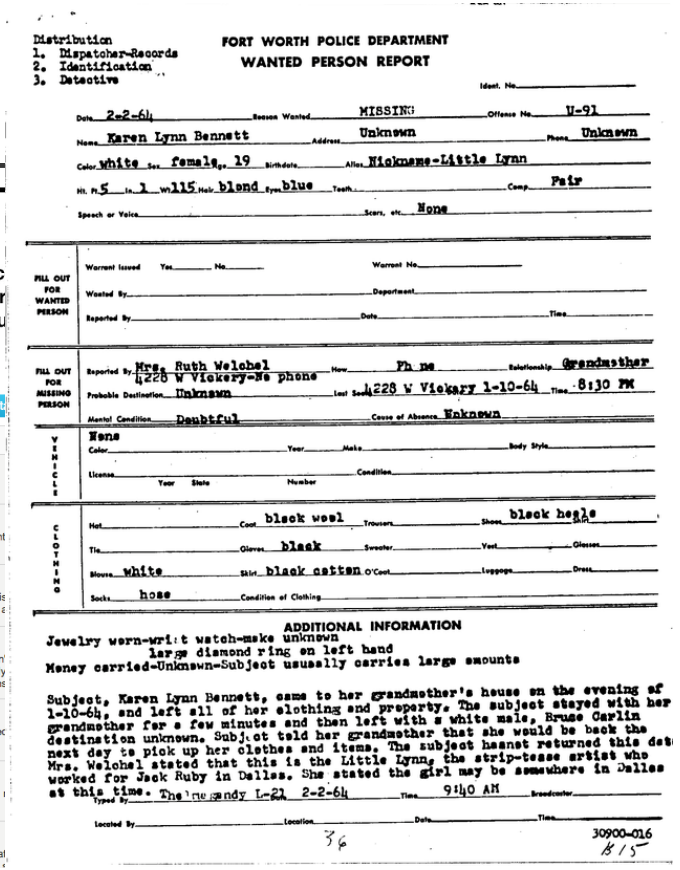
General Autism
For anything autism-related which isn sub forums, such as parenting, family technology, communication, and sens

Diagnosis
For discussions and questions relatec for children and adults.

Education
For all education-related issues from university.

Adulthood
For issues such as employment, relat

Sign up to our community e-news



Distribution
1. Dispatcher-Records
2. Identification
3. Detective

FORT WORTH POLICE DEPARTMENT
WANTED PERSON REPORT

Ident. No.

Date 2-2-64 Reason Wanted MISSING Offense No. U-91
Name Karen Lynn Bennett Address Unknown Phone Unknown
Color White Sex female Age 19 Birthdate Alias Nickname-Little Lynn
Ht. ft. 5 in. 1 Wt. 115 Hair blond Eyes blue Teeth Comp. Fair
Speech or Voice Scars, etc. None

FILL OUT FOR WANTED PERSON
Warrant Issued Yes No Warrant No.
Wanted By Department
Reported By Date Time

FILL OUT FOR MISSING PERSON
Reported By Mrs. Ruth Welchel How Ph me Relationship Grandmother
Probable Destination Unknown Last Seen 4228 W Vickery 1-10-64 Time 8:30 PM
Mental Condition Doubtful Cause of Absence Unknown

VEHICLE
None
Color Year Make Body Style
License Condition
Year State Number

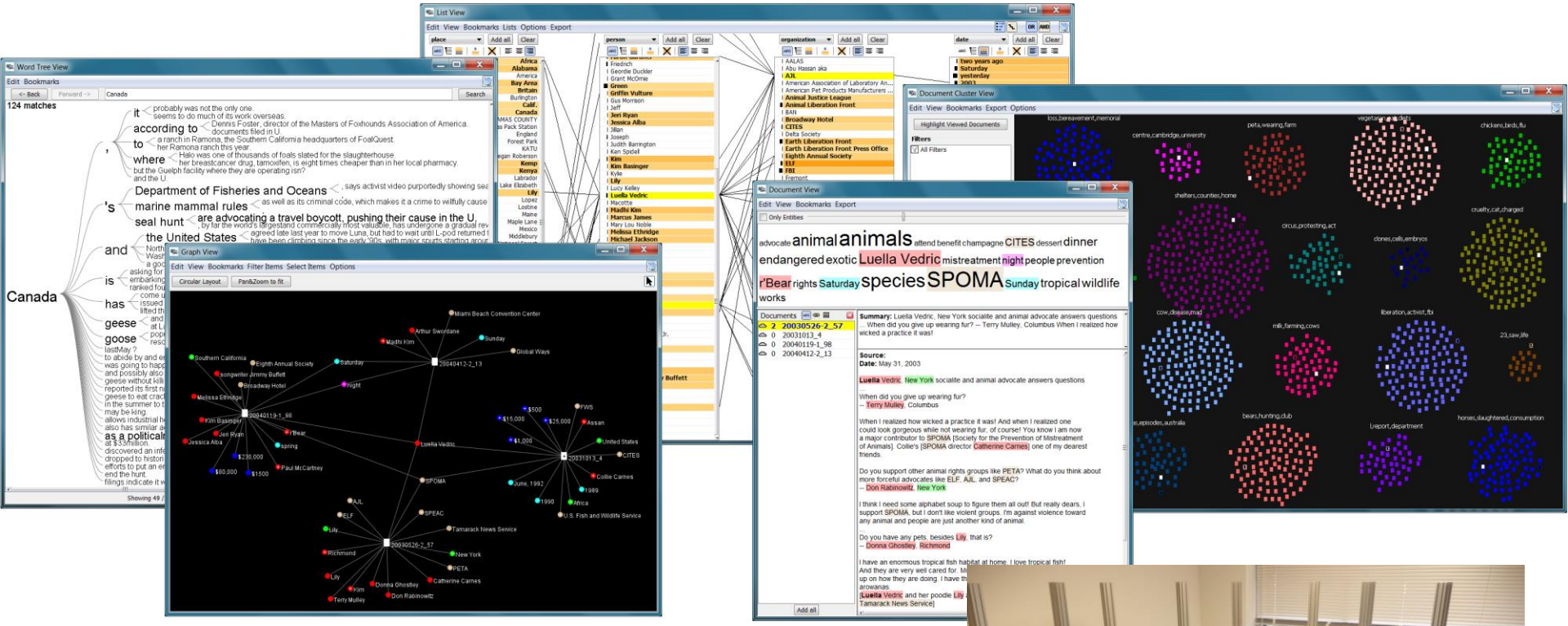
CLOTHING
Hat Coat black wool Trousers Shoes black hegle Belt
Tie Gloves black Sweater Vest Glasses
Blouse white Skirt black cotton Overcoat Luggage Dress
Socks hose Condition of Clothing

ADDITIONAL INFORMATION
Jewelry worn-wrist watch-make unknown
large diamond ring on left hand
Money carried-Unknown-Subject usually carries large amounts

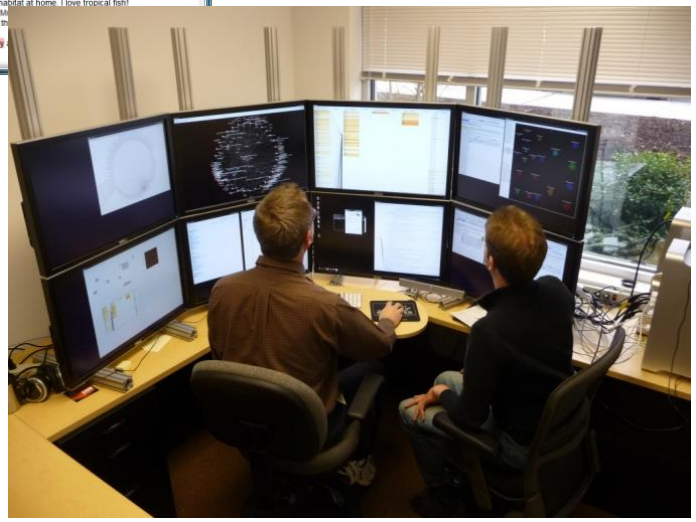
Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothing and property. The subject stayed with her grandmother for a few minutes and then left with a white male, Bruce Carlin destination unknown. Subject told her grandmother that she would be back the next day to pick up her clothes and items. The subject hasnot returned this da Mrs. Welchel stated that this is the Little Lynn, the strip-tease artist who worked for Jack Ruby in Dallas. She stated the girl may be somewhere in Dallas at this time. The investigation L-21 2-2-64 Time 9:40 AM
Typed By
Located By Location Date Time
36
30900-016
8/5

http://www.cc.gatech.edu/gvu/ii/jigsaw

Jigsaw



Visual analytics of large text document collections



Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG '13

MEMBER LOGIN: Username: Password: LOGIN

Forgot My Username/Password

Remember Login Info

eRobertParker.com

The Independent Consumer's Guide to Fine Wines



Rock and Rhône:
Parker Explores the Greatest Unheralded Grapes in the World–Grenache, Mourvèdre, and Syrah

HOME SUBSCRIBE GIFT SUBSCRIPTIONS FAQs VIRTUAL TOUR SUPPORT SITE MAP CONTACT US

What's Inside

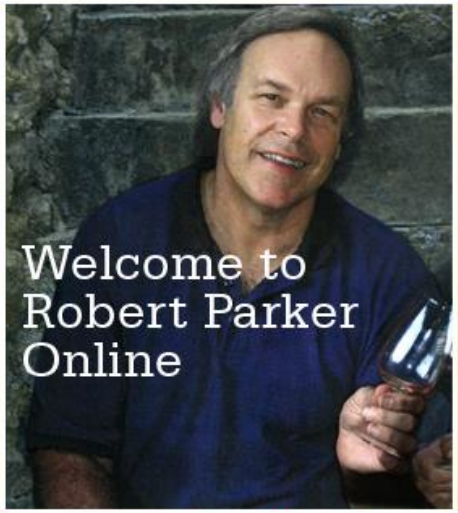
Explore Wines
Over 190,000 searchable notes from *The Wine Advocate*, Robert Parker's books and other sources.

The Hedonist's Gazette
An insider's look at our reviewers' wine and food experiences.

Article Archive
In-depth commentary from *The Wine Advocate* and other sources.

My Wines
Powerful, easy-to-use tools for purchasing/managing a cellar and participating in our wine community.

Robert Parker Mobi
Search the notes from any Internet-enabled mobile device.





Robert Parker's Wine Advocates

The Latest Information from the World's Wine Experts

Robert Parker
Reviews and Articles

Antonio Galloni's
In the Cellar

Neal Martin's
Wine Journal

Jay Miller
Reviews and Articles

Lisa Perrotti-Brown MW
In Asia

David Schildknecht
Reviews and Articles

Mark Squires'
E-zine and Bulletin Board

Parker & Zraly
Wine Certification Program

Information and Products for Non-Subscribers

Wine of the Day:
2009 Ridge Chardonnay
View Tasting Note



3W's of Wine iPhone App
Learn More

The Wine Advocate Store
Shop Now



Parker & Zraly Wine Certification
Learn More

Weekly Wine Buys
Powered by winealert
Learn More

Quick Links	More Information	About Our Reviewers	Links
Username/Password	**The Vintage Chart**	Robert Parker	The Wine Advocate Fund
What's New	Executive Wine Seminars	Antonio Galloni	Educational
Gift Subscriptions	Glossary of Wine Terms	Karen MacNeil	Accessory/Storage Links
Get News Via RSS	Parker's Rating System	Neal Martin	Retailer Links
Shop The Wine Advocate Store	*The Wine Advocate*	Jay Miller	Other
TWA Editorial Calendar	Site Map	Lisa Perrotti-Brown MW	
		David Schildknecht	
		Mark Squires	
		Kevin Zraly	
		Past Reviewers	
		Writer Ethics/Standards	





Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, …

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

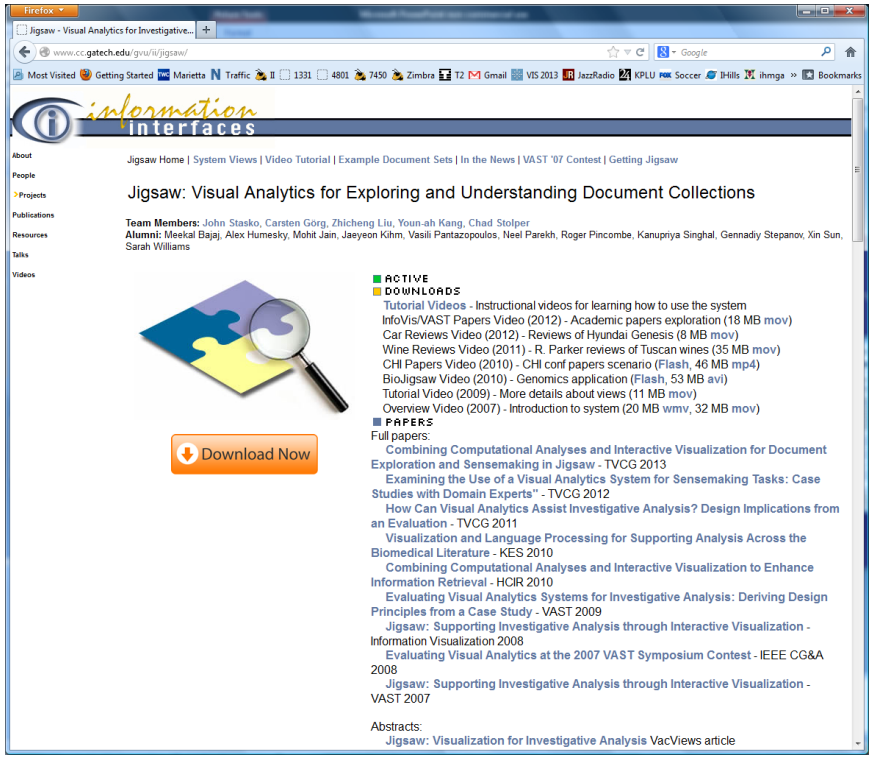
from database of 150,000 reviews





Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw

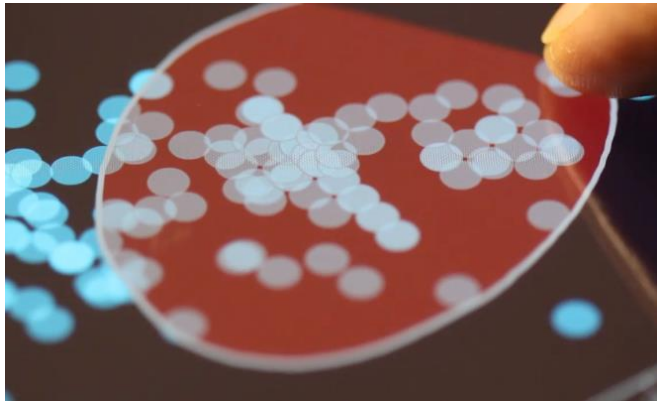




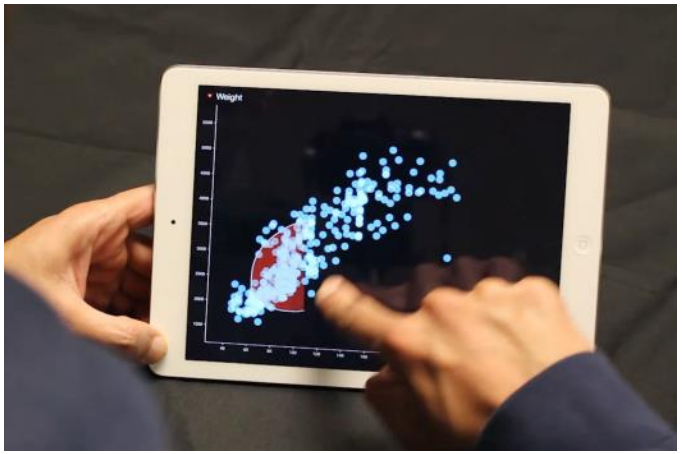
Applications

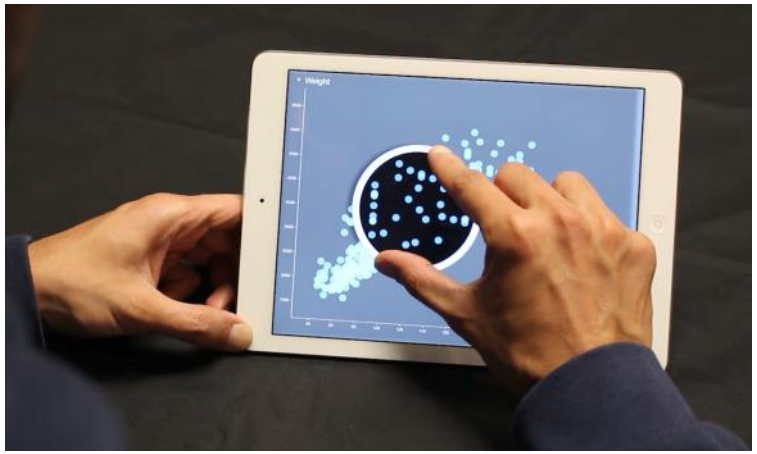
- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

Moving to Tablets



Scatterplot





Video

Sadana & Stasko
AVI ‘14

Different benefits of visualization

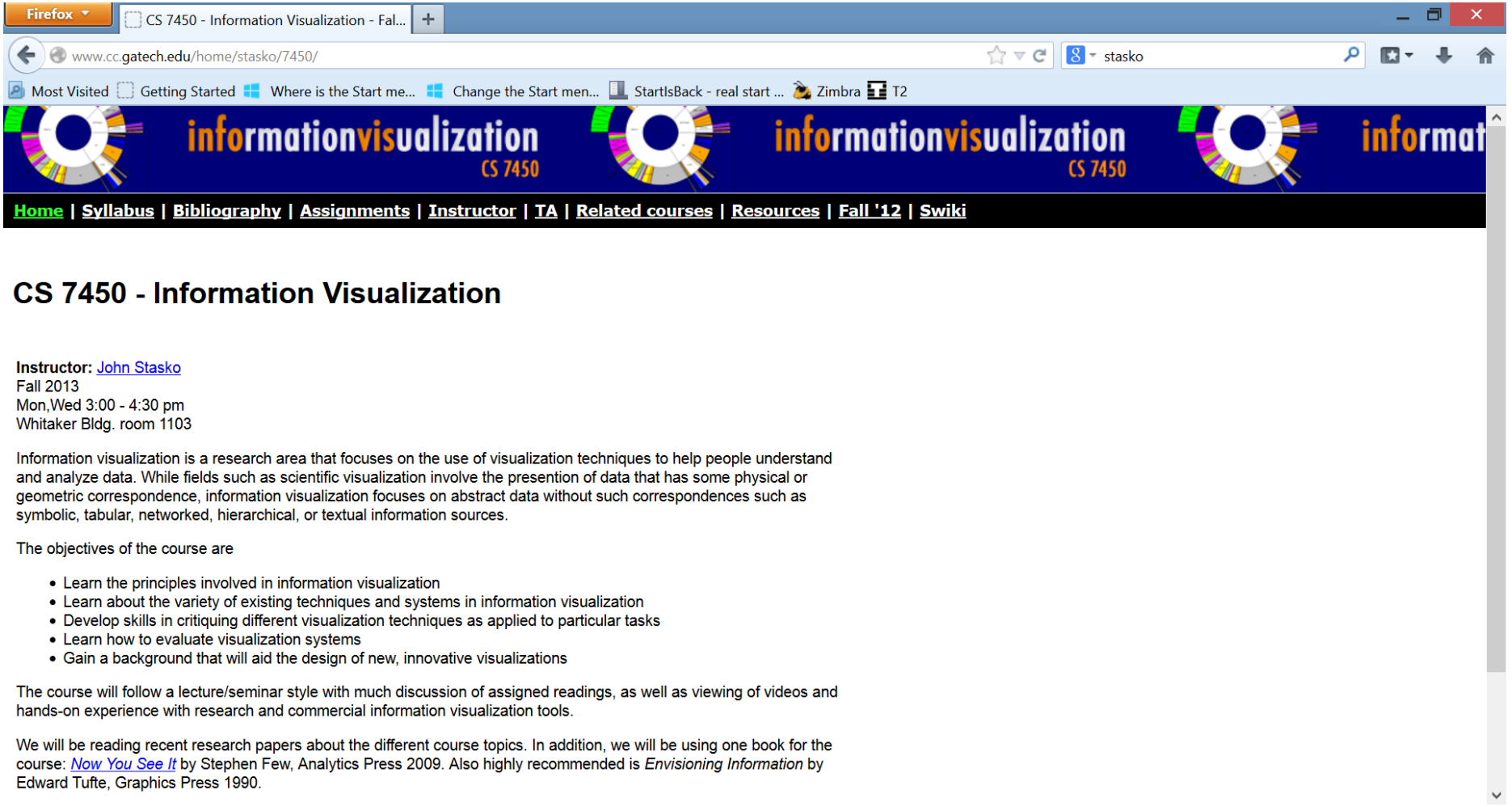
Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

My Class

http://www.cc.gatech.edu/~stasko/7450



My Class

http://www.cc.gatech.edu/~stasko/7450

Syllabus

Link to last year's 2012 syllabus with lecture slides

Overview

Week	Dates	Topic	Topic	HW
1	Aug 19, 21	Introduction	InfoVis overview	HW 1
2	Aug 26, 28	Visual perception	Value/Benefits of visualization	
3	Sep 2, 4	No Class -- Labor Day	Multivariate data & table/graph design	HW 2
4	Sep 9, 11	Few's design guidance	Multivariate visual representations 1	
5	Sep 16, 18	Multivariate visual representations 2	Tasks and analysis	HW 3a HW 3b
6	Sep 23, 25	InfoVis systems & toolkits	Commercial systems demos	HW 4
7	Sep 30, Oct 2	Storytelling	Tufte's design principles	
8	Oct 7, 9	Poster session	Casual InfoVis	HW 5
9	Oct 14, 16	No Class - Fall break	No Class - VIS Conference	
10	Oct 21, 23	Graphs and networks 1	Graphs and networks 2	
11	Oct 28, 30	Hierarchies & trees 1	Hierarchies & trees 2	HW 6
12	Nov 4, 6	Interaction	Overview & detail	
13	Nov 11, 13	Text & documents 1	Text & documents 2	
14	Nov 18, 20	Visual analytics 1	Visual analytics 2	HW 7
15	Nov 25, 27	Time series data	Project work day	
16	Dec 2, 4	Evaluation	Review	

Detail

Below are summaries of all the materials for each class. Lecture slides, references for readings and articles, videos shown, and software demonstrated are listed for each topic. Also available is a more general, alphabetical bibliography.

http://vimeo.com/98986594

For more, my EuroVis ‘14 Capstone Talk





Take Aways

Presentation & analysis

Exploratory data analysis, power from interaction

Developing new questions and insights

Acknowledgments

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)





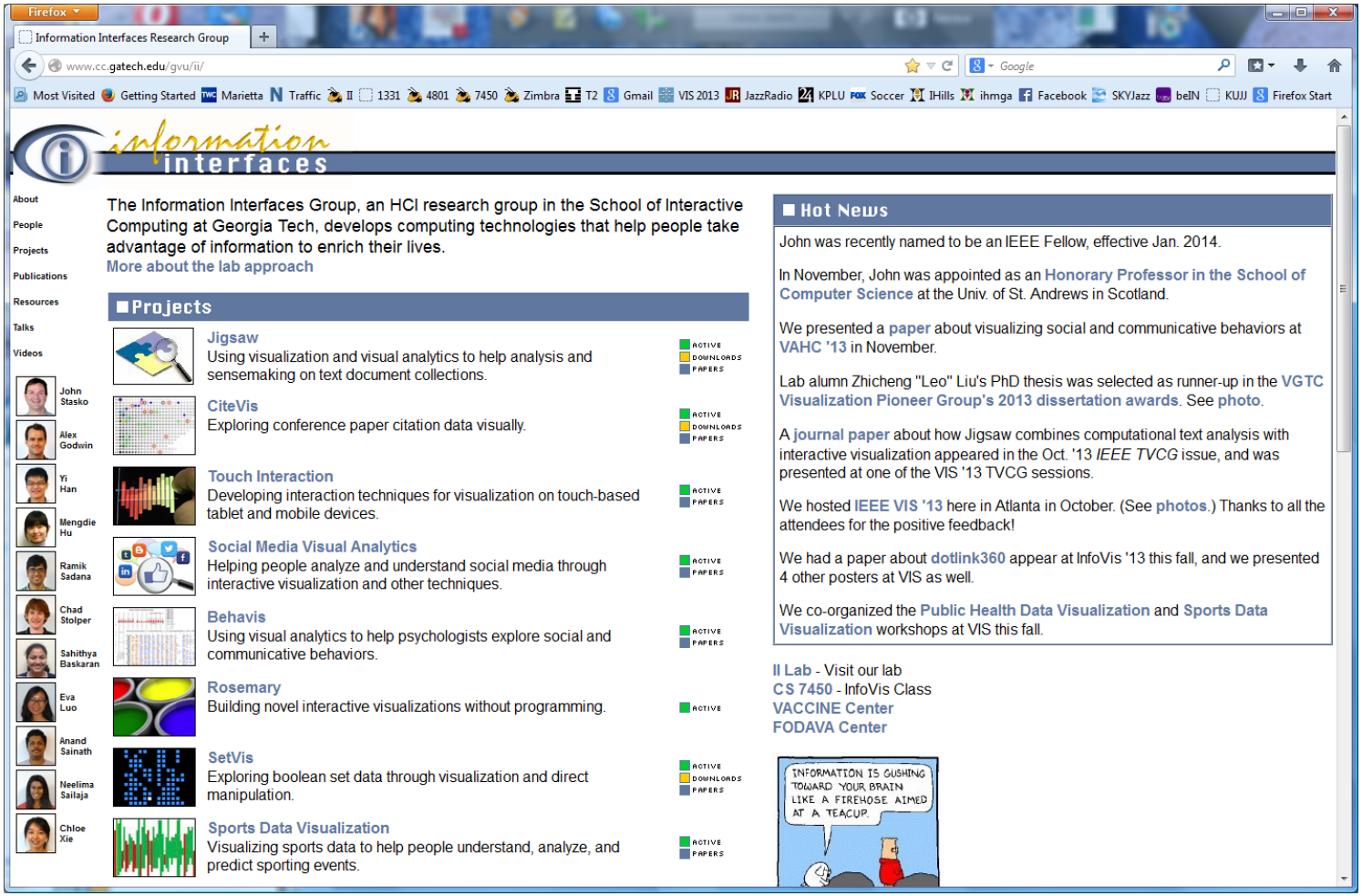
- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA program

My Research

http://www.cc.gatech.edu/gvu/ii



September 24, 2014

The Value of Visualization for Understanding Data and Making Decisions

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu

JISIC 2014





Georgia Tech College of Computing
School of Interactive Computing


Data

Many Data Analysis Approaches

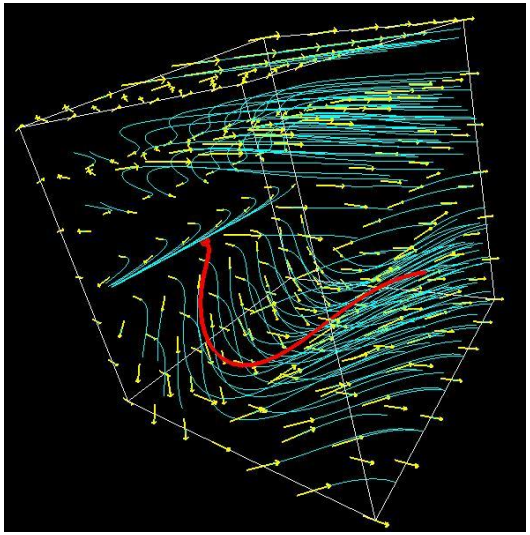
Statistics
Database & information retrieval
Data mining
Machine learning

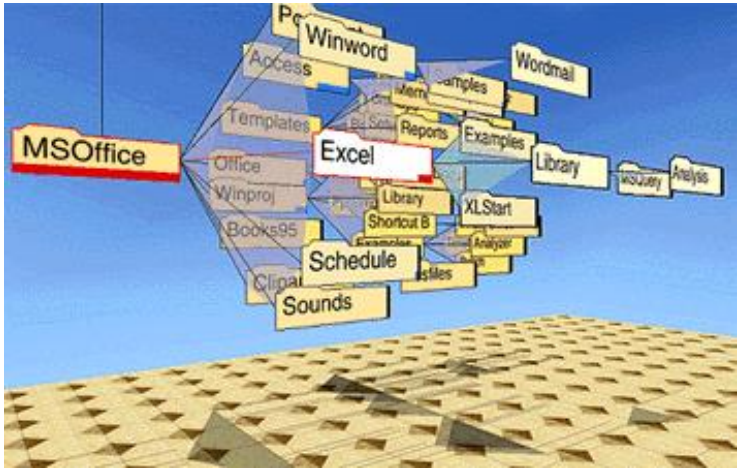


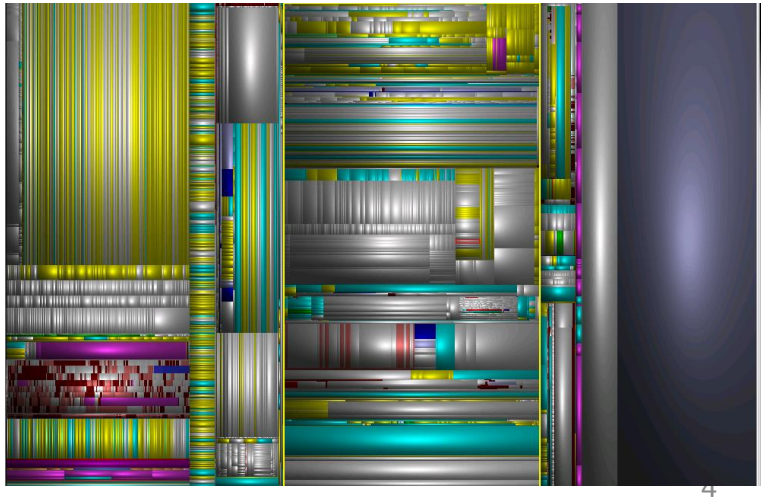


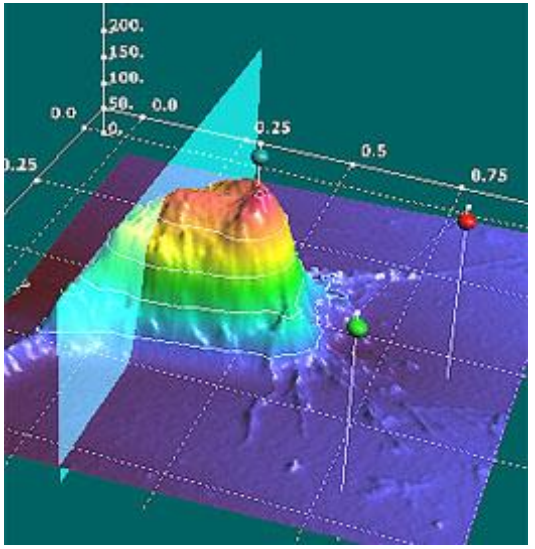
Data Visualization

~~Making pretty pictures?~~









Data Visualization

A cognitive process

Gain an understanding

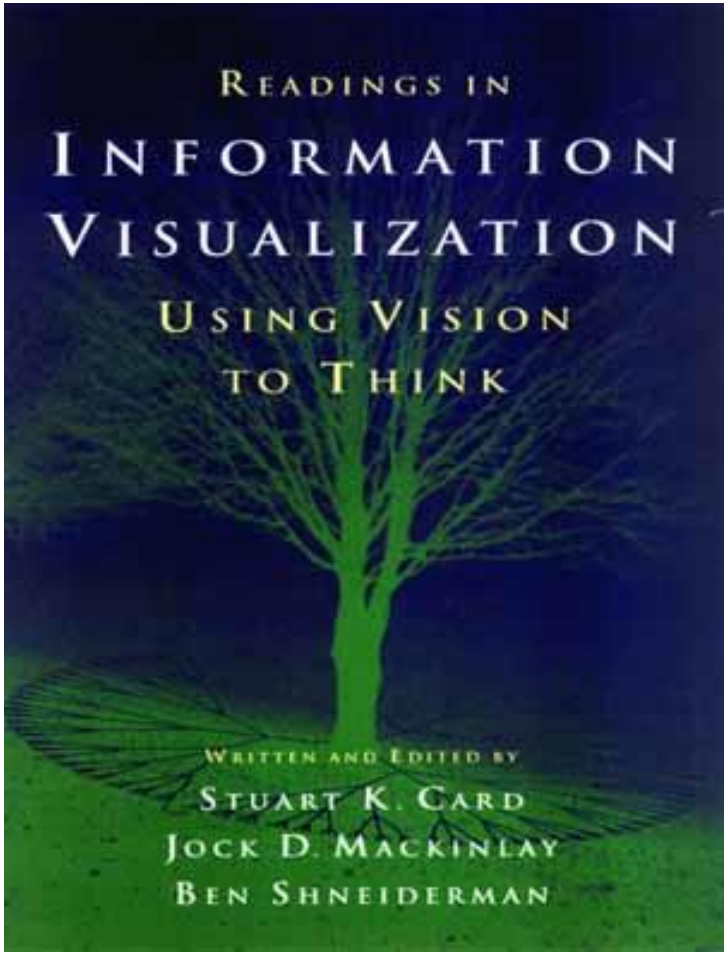


Data Visualization

A cognitive process

Gain an **understanding**





Visualization

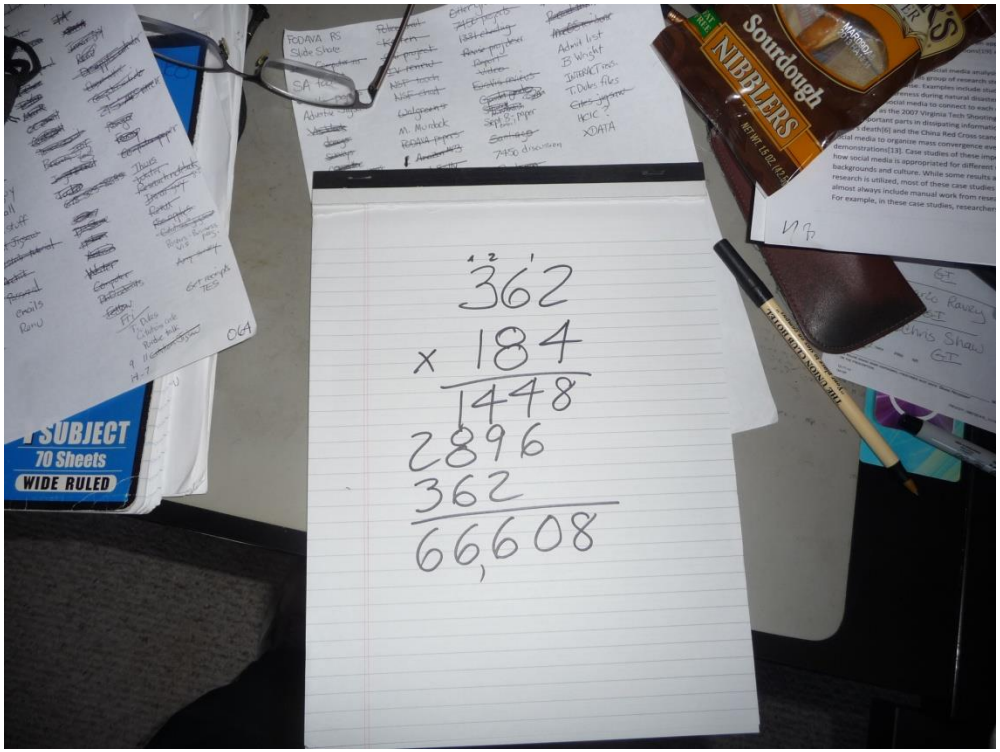
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provide a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Applications of Visualization

Presentation

Analysis

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support

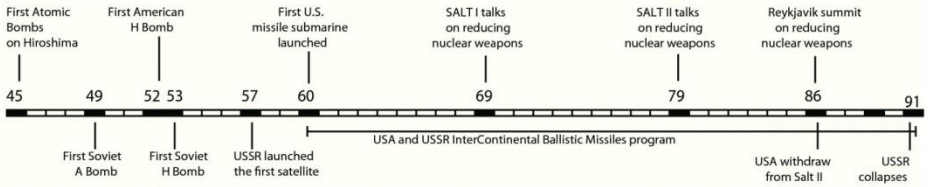


Infographics

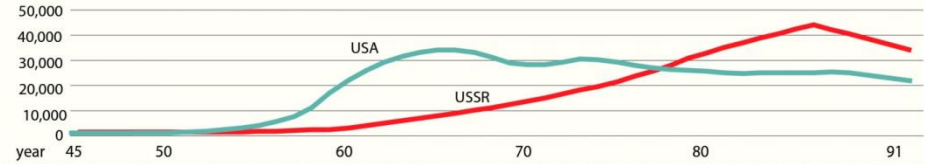
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



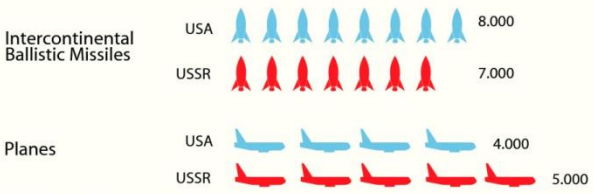
NUCLEAR STOCKPILE



ROCKET MODELS



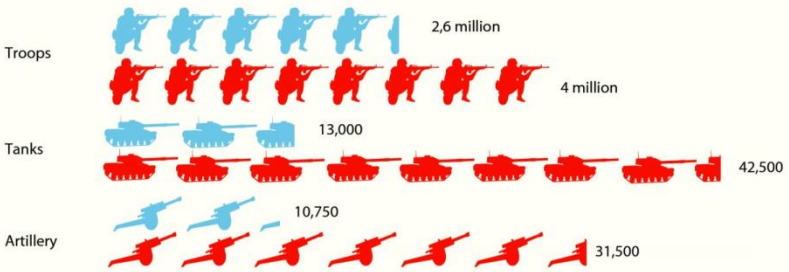
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



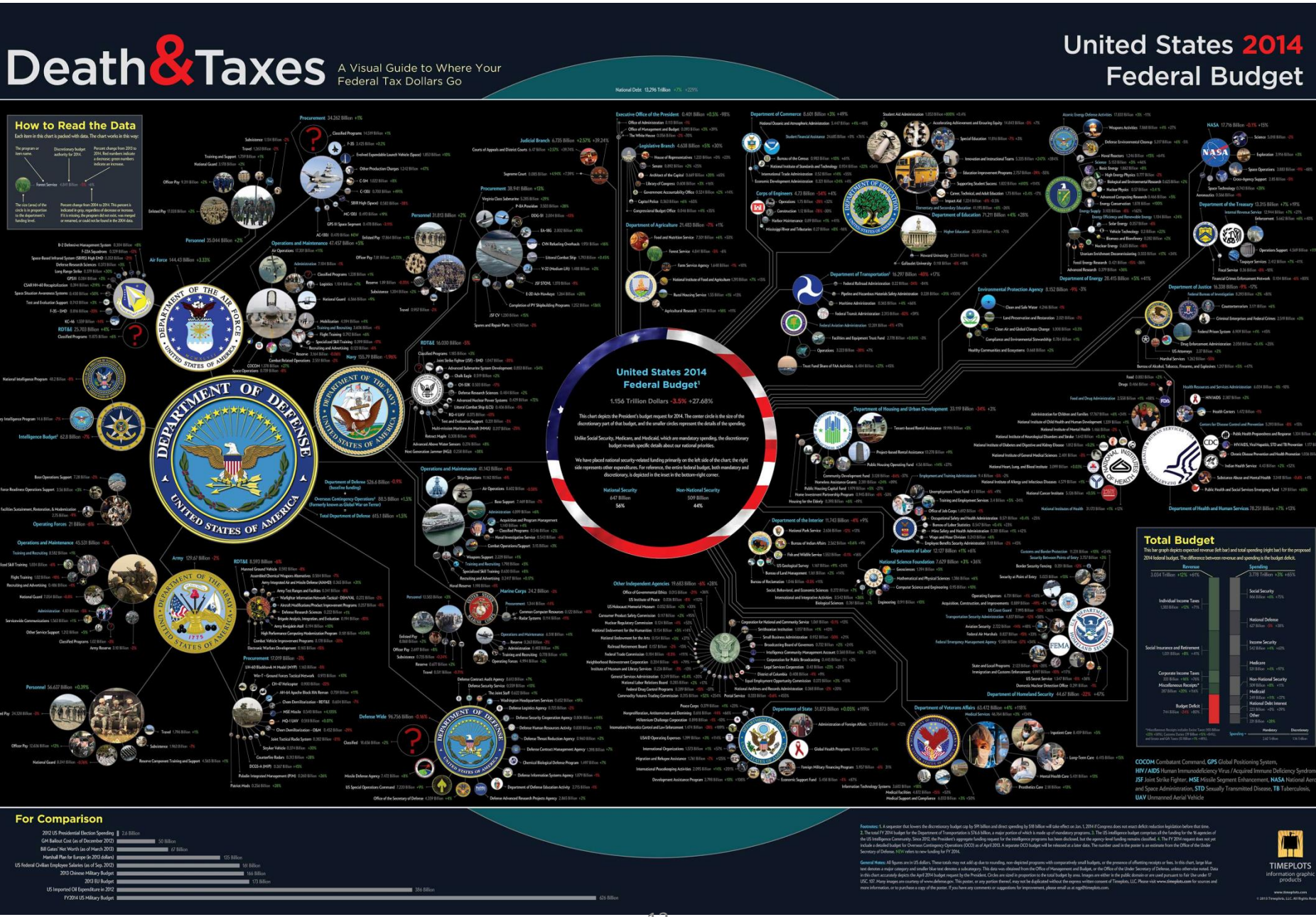


Produced by WatchTheAmericans.com a Fan Website

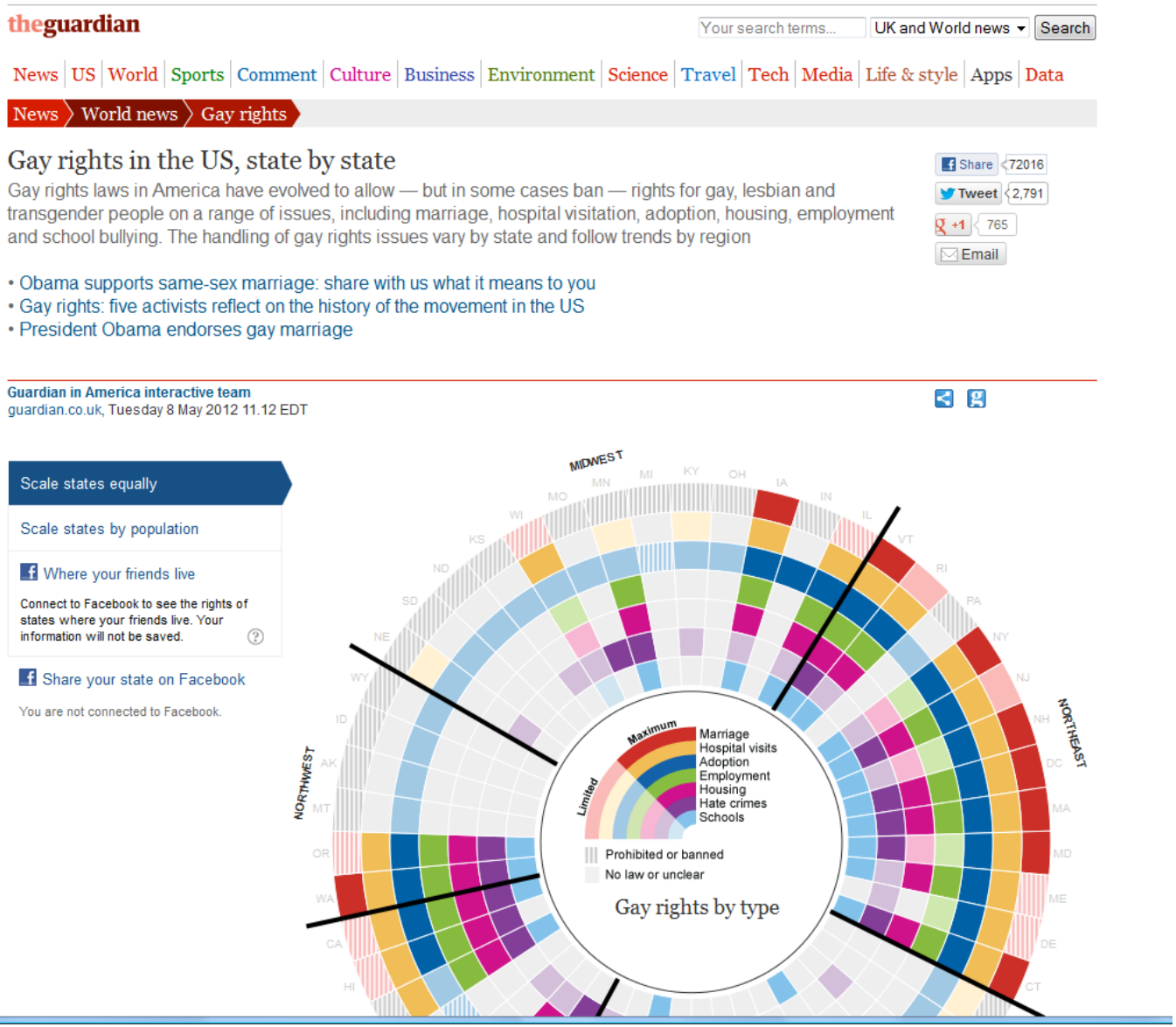
Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget

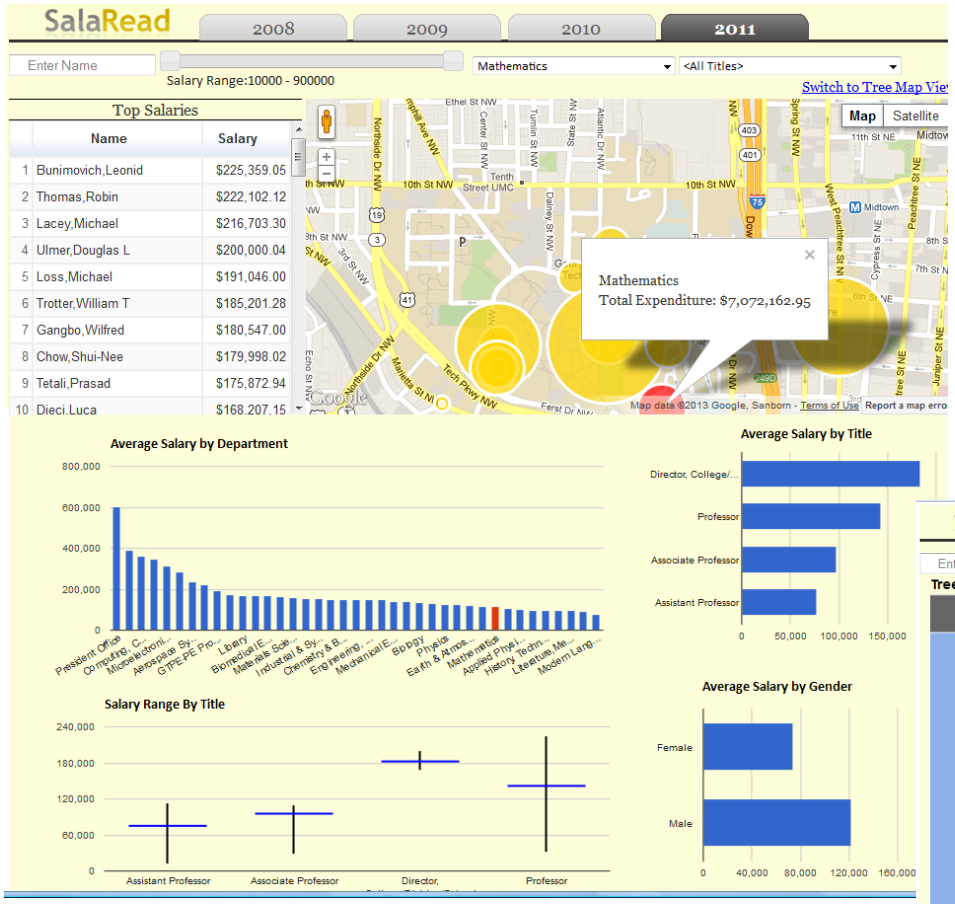


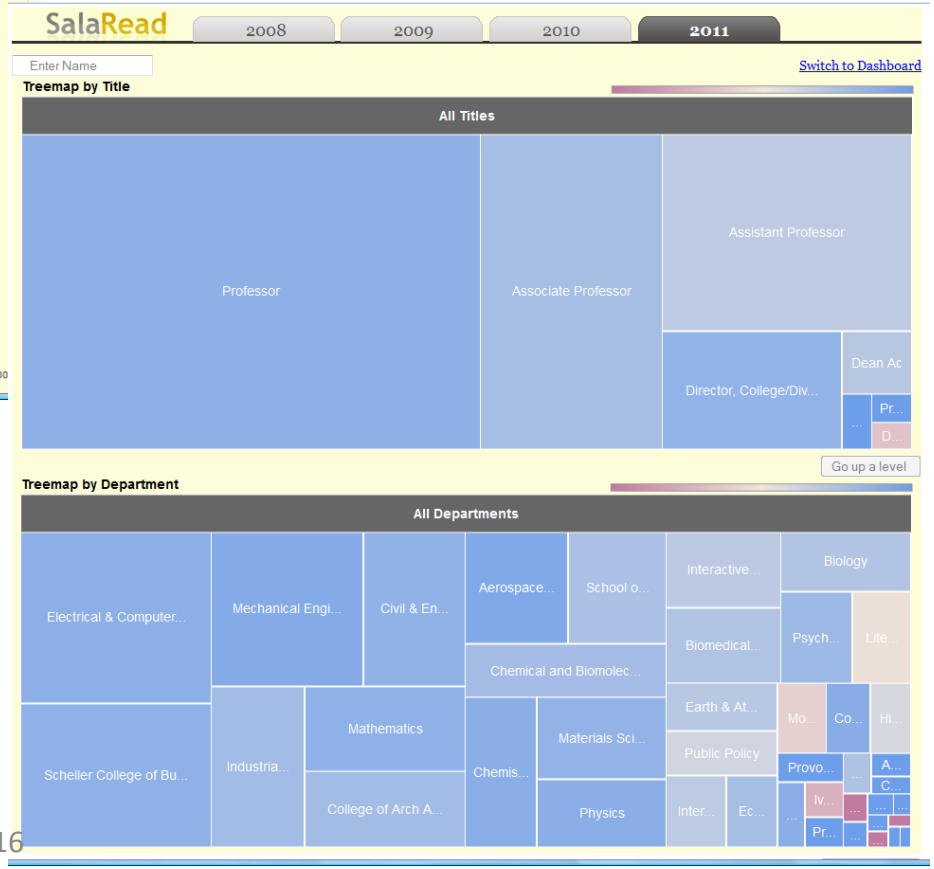
http://www.guardian.co.uk/world/interactive/2012/may/08/gay-rights-united-states?fb=native



Simply presenting data ***visually*** can have a profound impact

CS 7450
Fall ‘12
Design project

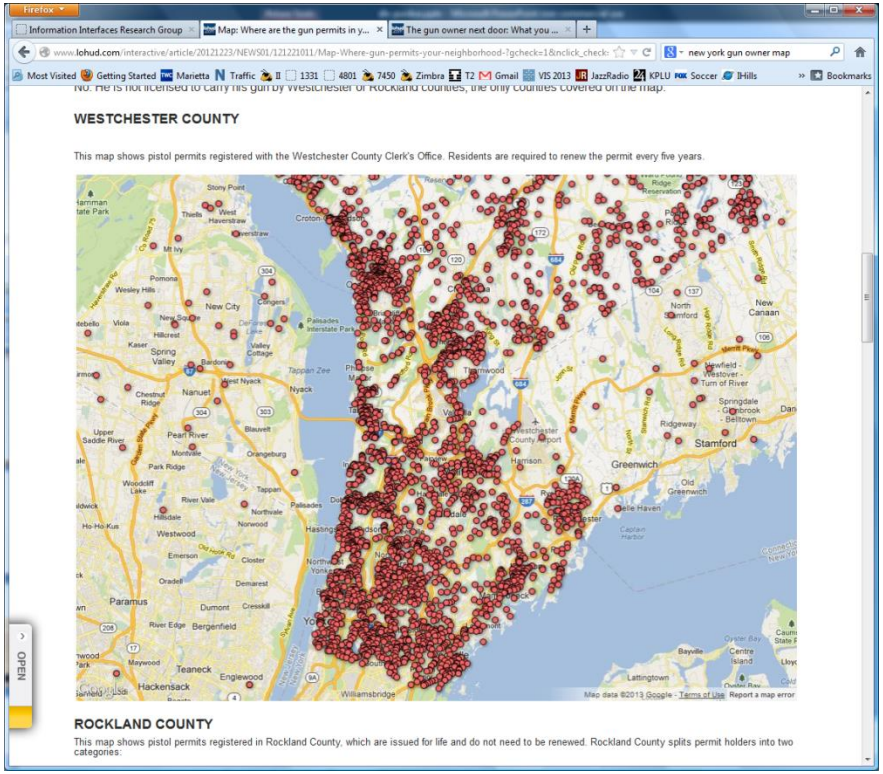




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun ownership in New York counties





http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood-?gcheck=1&nclick_check=1

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

May just show a few variables
and/or a subset of the data cases



2. Analysis

Explore the data

Assess a situation

Determine how to proceed

Decide what to do

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

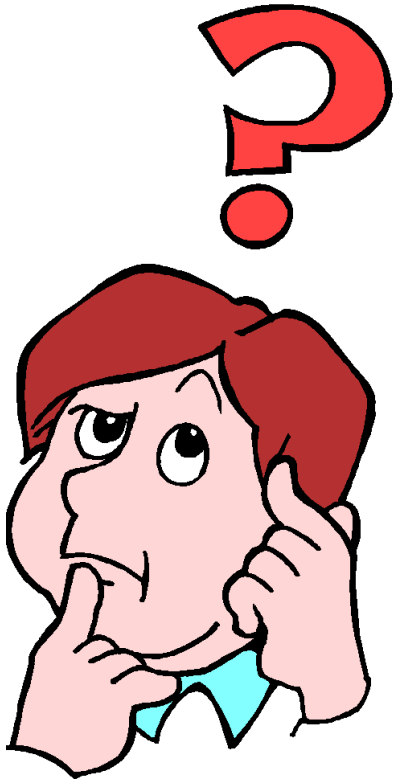
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context



Investigative analysis



COVER FEATURE

Visual Analytics Support for Intelligence Analysis

Carsten Görg, *University of Colorado*
Younah Kang, *Google*
Zhicheng Liu, *Stanford University*
John Stasko, *Georgia Tech*

Intelligence analysts must explore and evaluate volumes of data, from narrative recordings of field agents to open source news articles. Insights from visual analytics projects and a hypothetical scenario show the potential of visual analytics to aid these investigations.

Intelligence analysis has been a key application domain for visual analytics since the US Department of Homeland Security created the National Visualization and Analytics Center in 2004. An initial research roadmap[1] described the challenges and goals of this new domain and identified tasks, data, and analytical scenarios focused on homeland security and terrorism prevention. A passage from the roadmap characterizes the grand challenge for visual analytics systems:

> The analysis of overwhelming amounts of disparate, conflicting, and dynamic information is central to identifying and preventing emerging threats, protecting our borders, and responding in the event of an attack or other disaster. This analysis process requires human judgment to make the best possible evaluation of incomplete, inconsistent, and potentially deceptive information in the face of rapidly changing situations to both detect the expected and discover the unexpected.

Much of this challenge remains,[2] although visual analytics technologies certainly hold great promise. A few commercial visual analytics tools for intelligence analysis are emerging, such as Analyst's Notebook from IBM i2 (www.ibm.com/software/industry/i2software), nSpace from Oculus (www.oculustinfo.com/nspace), and Palantir's suite of systems (www.palantir.com). However, continued progress depends heavily on gaining a deeper understanding of intelligence analysis and the role of an analyst, as well as clarifying how visual analytics can help investigators. These insights must inform any design of any visual analytics system that aims to support intelligence analysis.

The "Investigative Scenarios" sidebar describes the kinds of analysis that are prevalent in the intelligence domain. The scale, diversity, and complexity of the information to be explored make such analyses cognitively demanding. Information is often narrative text, not quantitative data, and as such is not as amenable to automated analysis. Particularly challenging are the "pain points" in the intelligence process—the cost of scanning, recognizing (assessing), and selecting items for further attention; a limited attention span for evidence and hypotheses; and the difficulty of generating alternative hypotheses.[3]

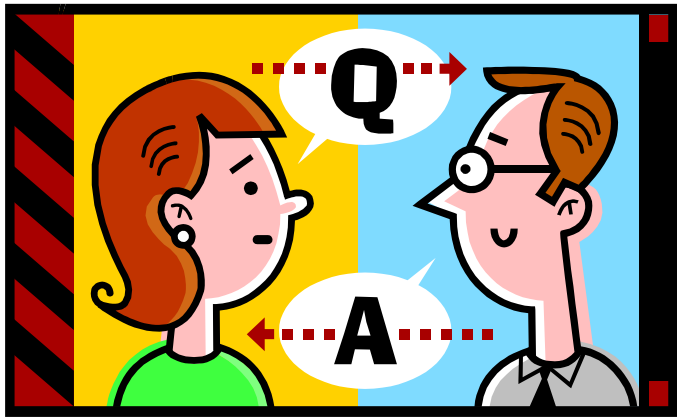
To better understand requirements for visual analytics support, we conducted a series of research projects on intelligence analysis. The projects, which spanned 2008 to 2013, included an observational study to better understand the intelligence analysis process and its characteristics. During this time, we also developed Jigsaw (www.cc.gatech.

 Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE

Görg, Kang, Kiu, Stasko
IEEE Computer '13

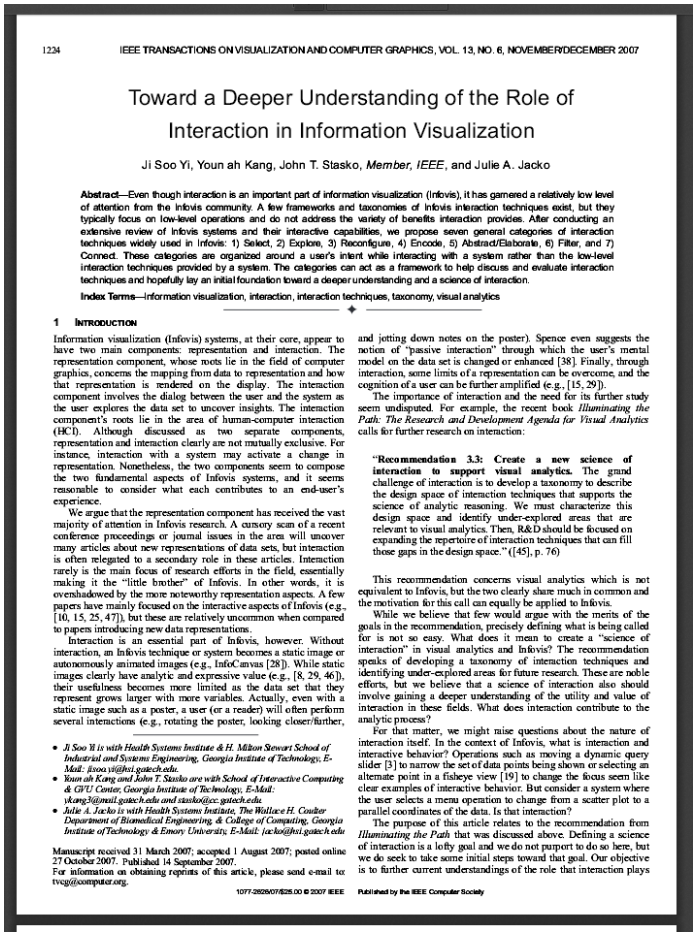
2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction



Engage in a dialog with your data

1. Select
2. Explore
3. Reconfigure
4. Encode
5. Abstract/Elaborate
6. Filter
7. Connect



1234 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 6, NOVEMBER/DECEMBER 2007

Toward a Deeper Understanding of the Role of Interaction in Information Visualization

Ji Soo Yi, Youn ah Kang, John T. Stasko, *Member, IEEE*, and Julie A. Jacko

Abstract—Even though interaction is an important part of information visualization (Infovis), it has garnered a relatively low level of attention from the Infovis community. A few frameworks and taxonomies of Infovis interaction techniques exist, but they typically focus on low-level operations and do not address the variety of benefits interaction provides. After conducting an extensive review of Infovis systems and their interactive capabilities, we propose seven general categories of interaction techniques widely used in Infovis: 1) Select, 2) Explore, 3) Reconfigure, 4) Encode, 5) Abstract/Elaborate, 6) Filter, and 7) Connect. These categories are organized around a user's intent while interacting with a system rather than the low-level interaction techniques provided by a system. The categories can act as a framework to help discuss and evaluate interaction techniques and hopefully lay an initial foundation toward a deeper understanding and a science of interaction.

Index Terms—Information visualization, interaction, interaction techniques, taxonomy, visual analytics

1 Introduction

Information visualization (Infovis) systems, at their core, appear to have two main components: representation and interaction. The representation component, whose roots lie in the field of computer graphics, concerns the mapping from data to representation and how that representation is rendered on the display. The interaction component involves the dialog between the user and the system as the user explores the data set to uncover insights. The interaction component's roots lie in the area of human-computer interaction (HCI). Although discussed as two separate components, representation and interaction clearly are not mutually exclusive. For instance, interaction with a system may activate a change in representation. Nonetheless, the two components seem to compose the two fundamental aspects of Infovis systems, and it seems reasonable to consider what each contributes to an end-user's experience.

We argue that the representation component has received the vast majority of attention in Infovis research. A cursory scan of a recent conference proceedings or journal issues in the area will uncover many articles about new representations of data sets, but interaction is often relegated to a secondary role in these articles. Interaction rarely is the main focus of research efforts in the field, essentially making it the "little brother" of Infovis. In other words, it is overshadowed by the more noteworthy representation aspects. A few papers have mainly focused on the interactive aspects of Infovis (e.g., [10, 15, 25, 47]), but these are relatively uncommon when compared to papers introducing new data representations.

Interaction is an essential part of Infovis, however. Without interaction, an Infovis technique or system becomes a static image or autonomously animated images (e.g., InfoCanvas [28]). While static images clearly have analytic and expressive value (e.g., [8, 29, 46]), their usefulness becomes more limited as the data set that they represent grows larger with more variables. Actually, even with a static image such as a poster, a user (or a reader) will often perform several interactions (e.g., rotating the poster, looking closer/further, and jotting down notes on the poster). Spence even suggests the notion of "passive interaction" through which the user's mental model on the data set is changed or enhanced [38]. Finally, through interaction, some limits of a representation can be overcome, and the cognition of a user can be further amplified (e.g., [15, 29]).

The importance of interaction and the need for its further study seem undisputed. For example, the recent book *Illuminating the Path: The Research and Development Agenda for Visual Analytics* calls for further research on interaction:

> "**Recommendation 3.3: Create a new science of interaction to support visual analytics.** The grand challenge of interaction is to develop a taxonomy to describe the design space of interaction techniques that supports the science of analytic reasoning. We must characterize this design space and identify under-explored areas that are relevant to visual analytics. Then, R&D should be focused on expanding the repertoire of interaction techniques that can fill those gaps in the design space." ([45], p. 76)

This recommendation concerns visual analytics which is not equivalent to Infovis, but the two clearly share much in common and the motivation for this call can equally be applied to Infovis.

While we believe that few would argue with the merits of the goals in the recommendation, precisely defining what is being called for is not so easy. What does it mean to create a "science of interaction" in visual analytics and Infovis? The recommendation speaks of developing a taxonomy of interaction techniques and identifying under-explored areas for future research. These are noble efforts, but we believe that a science of interaction also should involve gaining a deeper understanding of the utility and value of interaction in these fields. What does interaction contribute to the analytic process?

For that matter, we might raise questions about the nature of interaction itself. In the context of Infovis, what is interaction and interactive behavior? Operations such as moving a dynamic query slider [3] to narrow the set of data points being shown or selecting an alternate point in a fisheye view [19] to change the focus seem like clear examples of interactive behavior. But consider a system where the user selects a menu operation to change from a scatter plot to a parallel coordinates of the data. Is that interaction?

The purpose of this article relates to the recommendation from *Illuminating the Path* that was discussed above. Defining a science of interaction is a lofty goal and we do not purport to do so here, but we do seek to take some initial steps toward that goal. Our objective is to further current understandings of the role that interaction plays

• *Ji Soo Yi is with Health Systems Institute & H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, E-Mail: jisoo.yi@hsi.gatech.edu.*
• *Youn ah Kang and John T. Stasko are with School of Interactive Computing & GVU Center, Georgia Institute of Technology, E-Mail: ykang3@mail.gatech.edu and stasko@cc.gatech.edu.*
• *Julie A. Jacko is with Health Systems Institute, The Wallace H. Coulter Department of Biomedical Engineering, & College of Computing, Georgia Institute of Technology & Emory University, E-Mail: jacko@hsi.gatech.edu*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007. Published 14 September 2007.
For information on obtaining reprints of this article, please send e-mail to tvcg@computer.org.

1077-2626/07/$25.00 © 2007 IEEE Published by the IEEE Computer Society

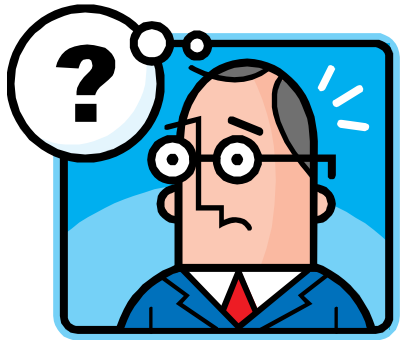
Yi, Kang, Stasko, Jacko
TVCG (InfoVis) '07

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Visualization of different data types

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?

Example Tasks and Goals

- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify “hidden” messages or stories in this document collection.
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.

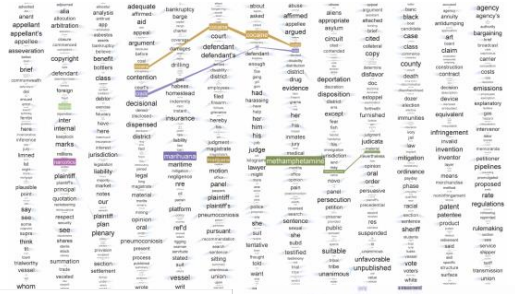
The Challenge

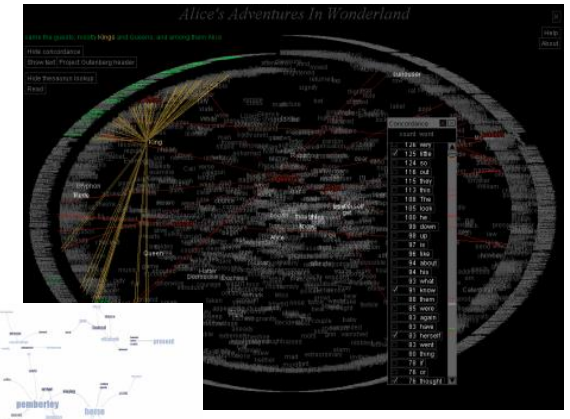
Text is nominal data

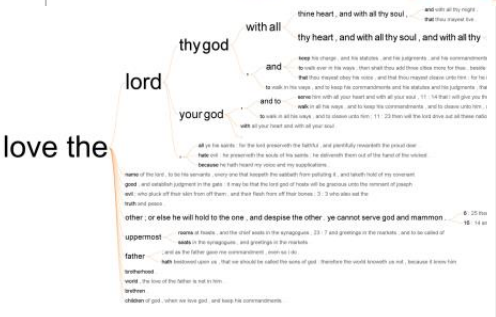
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.









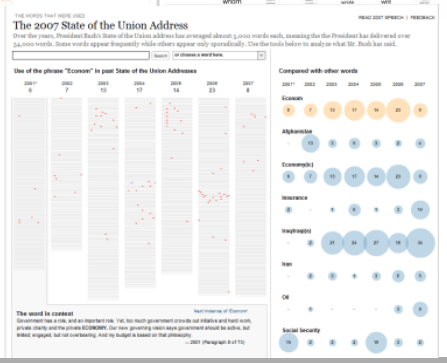
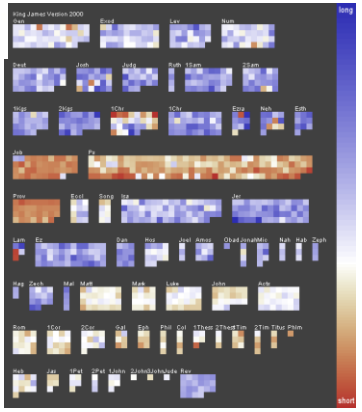
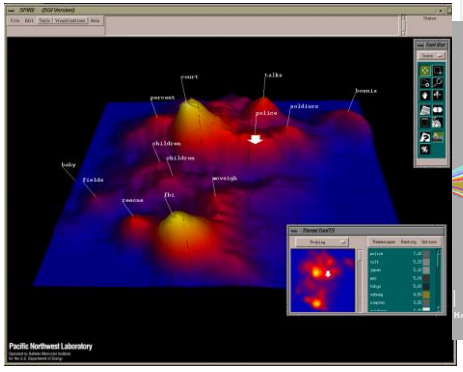
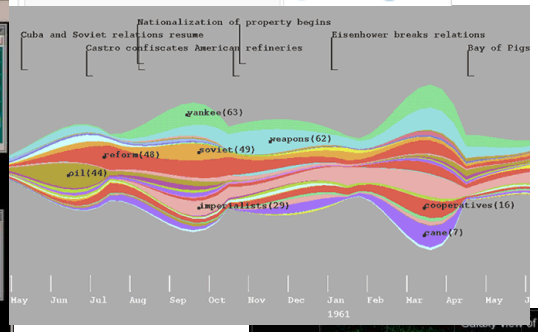


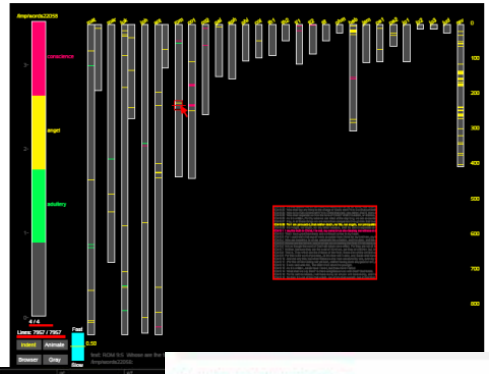


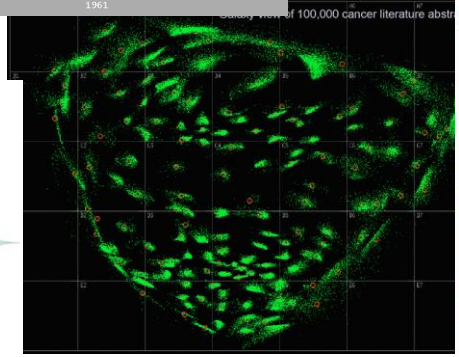
Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with "X and Y" and "X at Y" respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

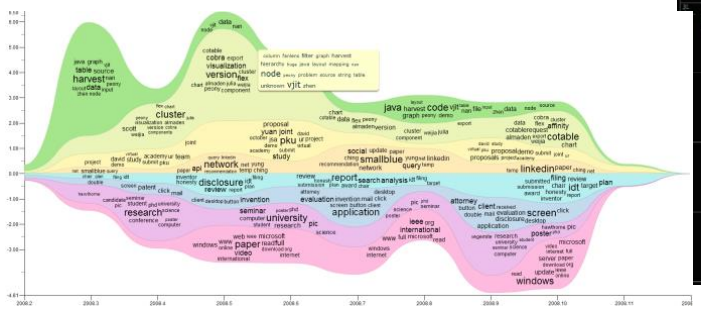




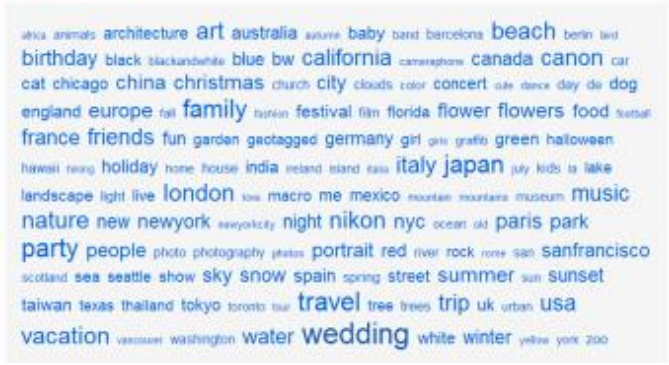








All time most popular tags



“Putting the pieces together”

Jigsaw

Computational analysis of documents’ text

- Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

- Views are highly coordinated



Help “investigators” explore, analyze and understand large document collections

Academic Papers

Health Forums

Product Reviews

Police Reports



Designing and Implementing an Interac
Scatterplot Visualization for a Tablet Con

Ramik Sadana
Georgia Institute of Technology
ramik@gatech.edu

John Stasi
Georgia Institute of T
stasko@cc.gate

ABSTRACT

Tablet computers now offer screen sizes and computing capabilities that are competitive with traditional desktop PCs. Their popularity has grown tremendously, but we are just beginning to see information visualization applications designed for this platform. One potential reason for this limited development is the challenge of designing and implementing a multi-touch interface for visualizations on mobile, tablet devices. In this work, we identify the primary challenges that touch screen interactions pose for information visualization applications. We explore the design space of multi-touch interactions for visualizations and present a prototype information visualization application using a specific technique, a dynamic scatterplot, for an iPad.

Categories and Subject Descriptors
H.5.0 [Information Interfaces and Presentation]: General

General Terms
Design, Algorithms, Human Factors.

Keywords
Information visualization, multi-touch interaction, scatterplot, tablet computer, gesture.

1. INTRODUCTION

The popularity of tablet computers has grown tremendously in recent years. Devices such as the Apple iPad, Amazon Kindle, and Microsoft Surface constitute a significant portion of computer sales today and a wide variety of applications have been tailored to tablet platform. However, one area with just a few initial applications for tablets is information visualization. Though popular commercial systems such as Tableau and Spotfire have introduced tablet versions in the last two years, these systems still feel much like a port of a desktop application. They do not yet leverage the set of gestural interactions that touch-based interfaces provide and do not offer a rich, multi-touch interface in the style of other tablet applications.

We speculate that the limited development of information visualization applications for tablets has resulted from the challenge of designing and implementing the interface. First, most tablets still provide a smaller screen size, which is a limitation for data visualization. Perhaps more importantly, information visualization applications generally have many small visual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the authors must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AVI'14, May 27 - 29 2014, Como, Italy
Copyright is held by the authors. Publication rights licensed to ACM.
ACM 978-1-4503-2775-6/14/05...$15.00.

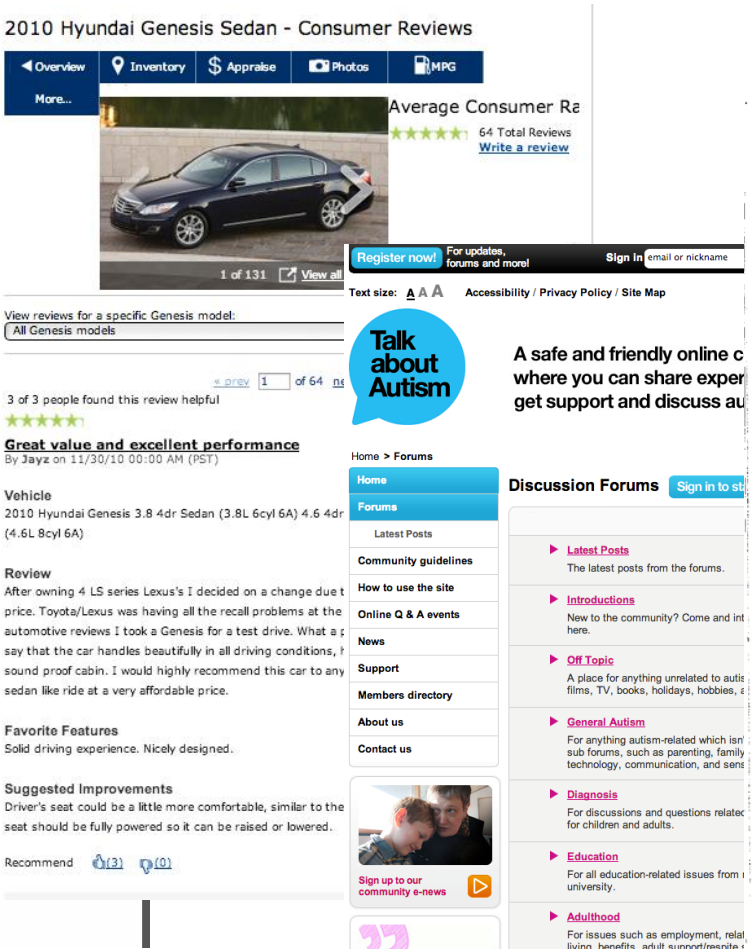
objects to select and manipulate, interactive widgets such as buttons, boxes. Desktop information visualiz make extensive use of a WIMP (w interface. Translating visualization operations to a finger-directed multi keyboard and mouse is simply a very

Some recent research has begun however. One particular avenue development of multi-user, multi applications for tabletop computers [1 explored the design space for s developing for tablet computers is s tabletop computers. Baur et al. [2] rec tablet interface for a StreamGraph-ty This application is much more alig create. Our specific focus, however, used visualization technique. Rzeszot application is a similar technique to driven interactions.

Rather than viewing the movement applications to multi-touch tablets software port, we see this problem challenge. Initial research [25] has interaction design space – for any information visualization syster implementations make sense and Furthermore, we will show that a gestures from tablet applications applied to data visualizations.

In this work, we further explo implement a working information vi tablet PC. We chose a dynamic sca technique of our focus. Scatterplots a in information visualization; many s suite of provided techniques. Scatterp in data visualization, including the FilmFinder system [1] that led to t Our project starts from "ground zero an existing system. Our application advantage of, and to address all the l a tablet interface using only fingers a

The contributions of our work are sections of this article. By studying commercial implementations of a dyn set of tasks and interactive operation would support (Section 3). Next, we the tasks and describe how those c design (Section 4). We ultimately c multi-touch operations for each task, for each, propose multiple potenti many of the options in a proto application for the iPad (Section 5). I paper may not convey all the operati overview summarizing the application



2010 Hyundai Genesis Sedan - Consumer Reviews

Overview | Inventory | Appraise | Photos | MPG | More...

Average Consumer Ra
64 Total Reviews
Write a review

1 of 131 View all

View reviews for a specific Genesis model:
All Genesis models

« prev 1 of 64 ne

3 of 3 people found this review helpful

Great value and excellent performance
By Jayz on 11/30/10 00:00 AM (PST)

Vehicle
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 4dr (4.6L 8cyl 6A)

Review
After owning 4 LS series Lexus's I decided on a change due t price. Toyota/Lexus was having all the recall problems at the automotive reviews I took a Genesis for a test drive. What a p say that the car handles beautifully in all driving conditions, h sound proof cabin. I would highly recommend this car to any sedan like ride at a very affordable price.

Favorite Features
Solid driving experience. Nicely designed.

Suggested Improvements
Driver's seat could be a little more comfortable, similar to the seat should be fully powered so it can be raised or lowered.

Recommend (3) (0)

Register now! For updates, forums and more! Sign in email or nickname

Text size: A A A Accessibility / Privacy Policy / Site Map

Talk about Autism

A safe and friendly online c where you can share exper get support and discuss au

Home > Forums

Home
Forums
Latest Posts
Community guidelines
How to use the site
Online Q & A events
News
Support
Members directory
About us
Contact us

Discussion Forums Sign in to st

Latest Posts
The latest posts from the forums.

Introductions
New to the community? Come and int here.

Off Topic
A place for anything unrelated to autis films, TV, books, holidays, hobbies,

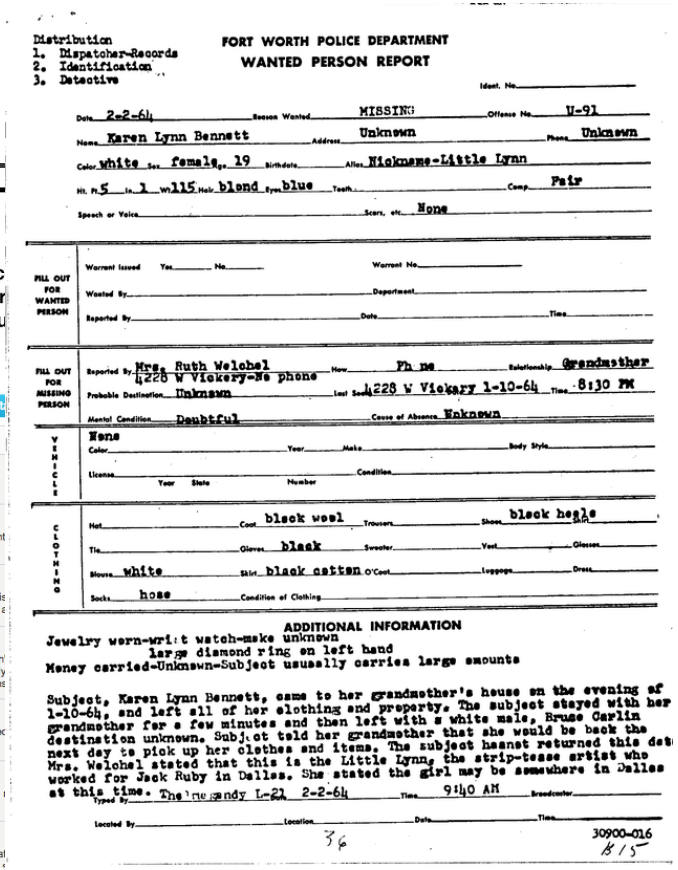
General Autism
For anything autism-related which isn sub forums, such as parenting, family technology, communication, and sens

Diagnosis
For discussions and questions relatec for children and adults.

Education
For all education-related issues from university.

Adulthood
For issues such as employment, relat

Sign up to our community e-news



Distribution
1. Dispatcher-Records
2. Identification
3. Detective

FORT WORTH POLICE DEPARTMENT
WANTED PERSON REPORT

Ident. No.

Date 2-2-64 Reason Wanted MISSING Offense No. U-91
Name Karen Lynn Bennett Address Unknown Phone Unknown
Color White Sex female Age 19 Birthdate Alias Nickname-Little Lynn
Ht. ft. 5 in. 1 Wt. 115 Hair blond Eyes blue Teeth Comp. Fair
Speech or Voice Scars, etc. None

FILL OUT FOR WANTED PERSON
Warrant Issued Yes No Warrant No.
Wanted By Department
Reported By Date Time

FILL OUT FOR MISSING PERSON
Reported By Mrs. Ruth Welchel How Ph me Relationship Grandmother
Probable Destination Unknown Last Seen 4228 W Vickery 1-10-64 Time 8:30 PM
Mental Condition Doubtful Cause of Absence Unknown

VEHICLE
None
Color Year Make Body Style
License Year State Number Condition

CLOTHING
Hat Coat black wool Trousers Shoes black heels
Tie Gloves black Sweater Vest Glasses
Blouse white Skirt black cotton Overcoat Luggage Dress
Socks hose Condition of Clothing

ADDITIONAL INFORMATION
Jewelry worn-wrist watch-make unknown
large diamond ring on left hand
Money carried-Unknown-Subject usually carries large amounts

Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothing and property. The subject stayed with her grandmother for a few minutes and then left with a white male, Bruce Carlin destination unknown. Subject told her grandmother that she would be back the next day to pick up her clothes and items. The subject hasnot returned this dat Mrs. Welchel stated that this is the Little Lynn, the strip-tease artist who worked for Jack Ruby in Dallas. She stated the girl may be somewhere in Dallas at this time. The investigation L-21 2-2-64 Time 9:40 AM

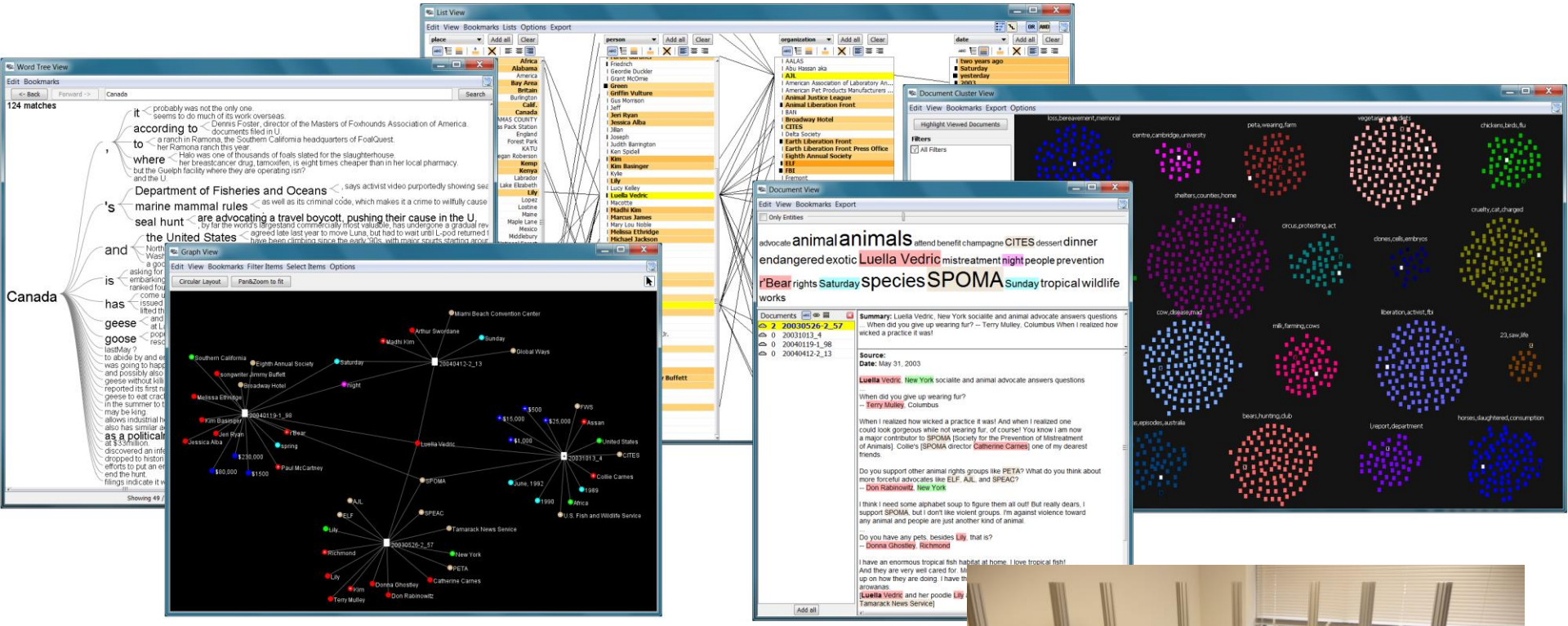
Typed By
Located By Location Date Time

36

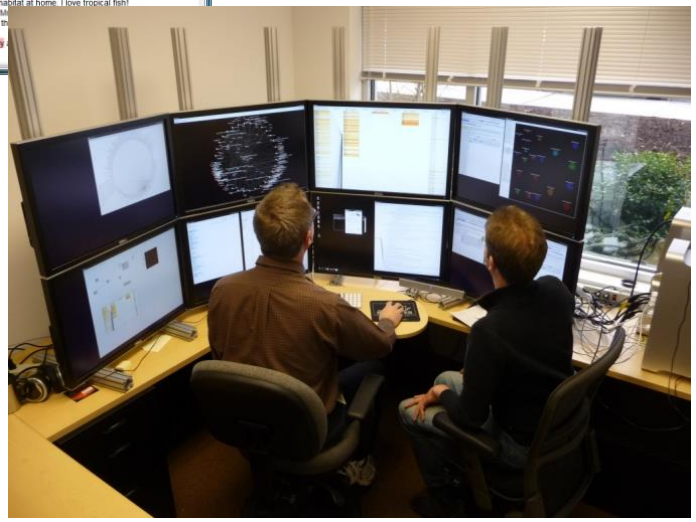
30900-016
8/5

http://www.cc.gatech.edu/gvu/ii/jigsaw

Jigsaw

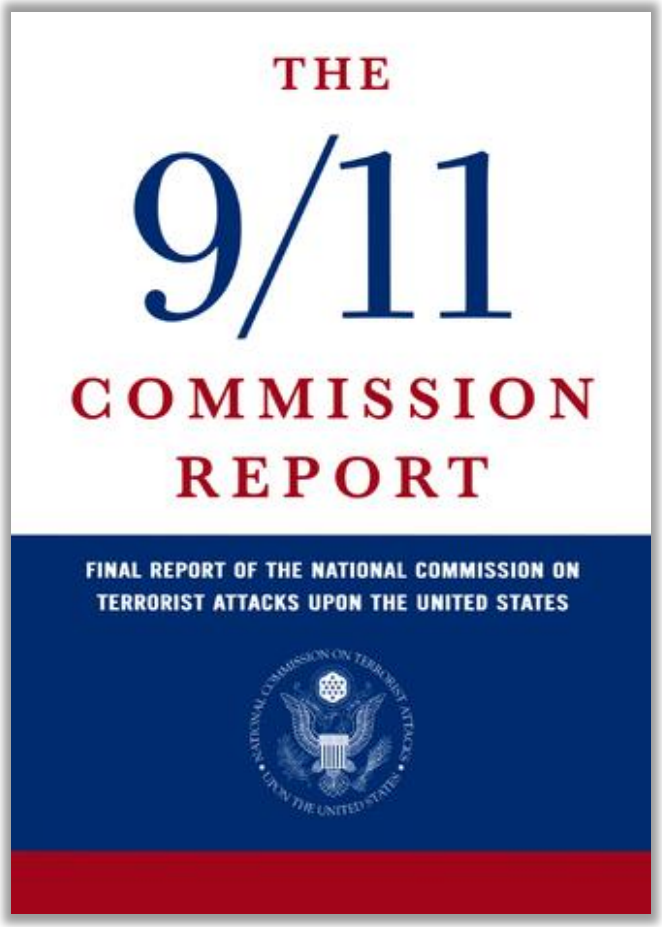


Visual analytics of large text document collections



Görg, Liu, Kihm, Choo, Park, & Stasko
TVCG ‘13

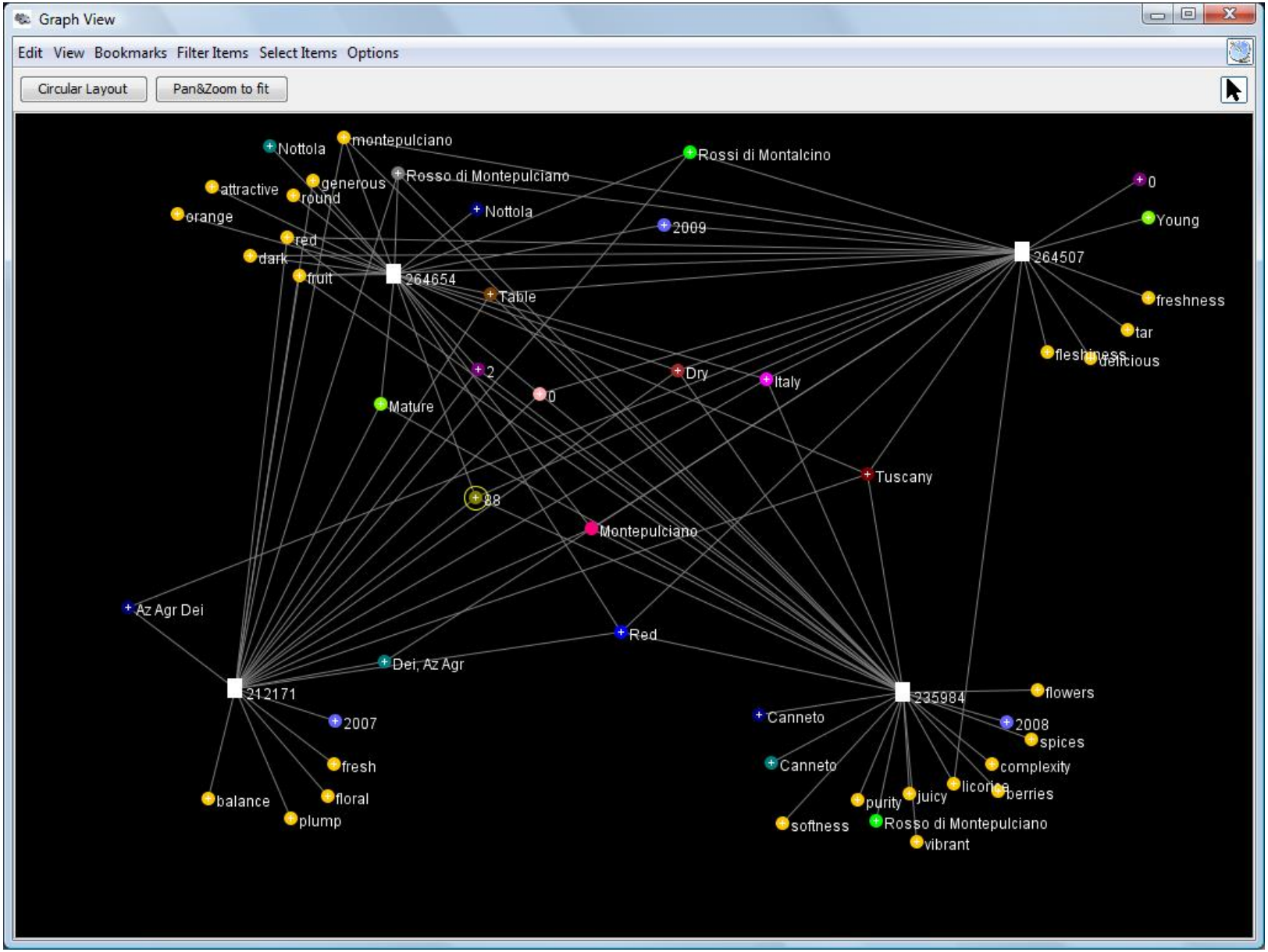
Example



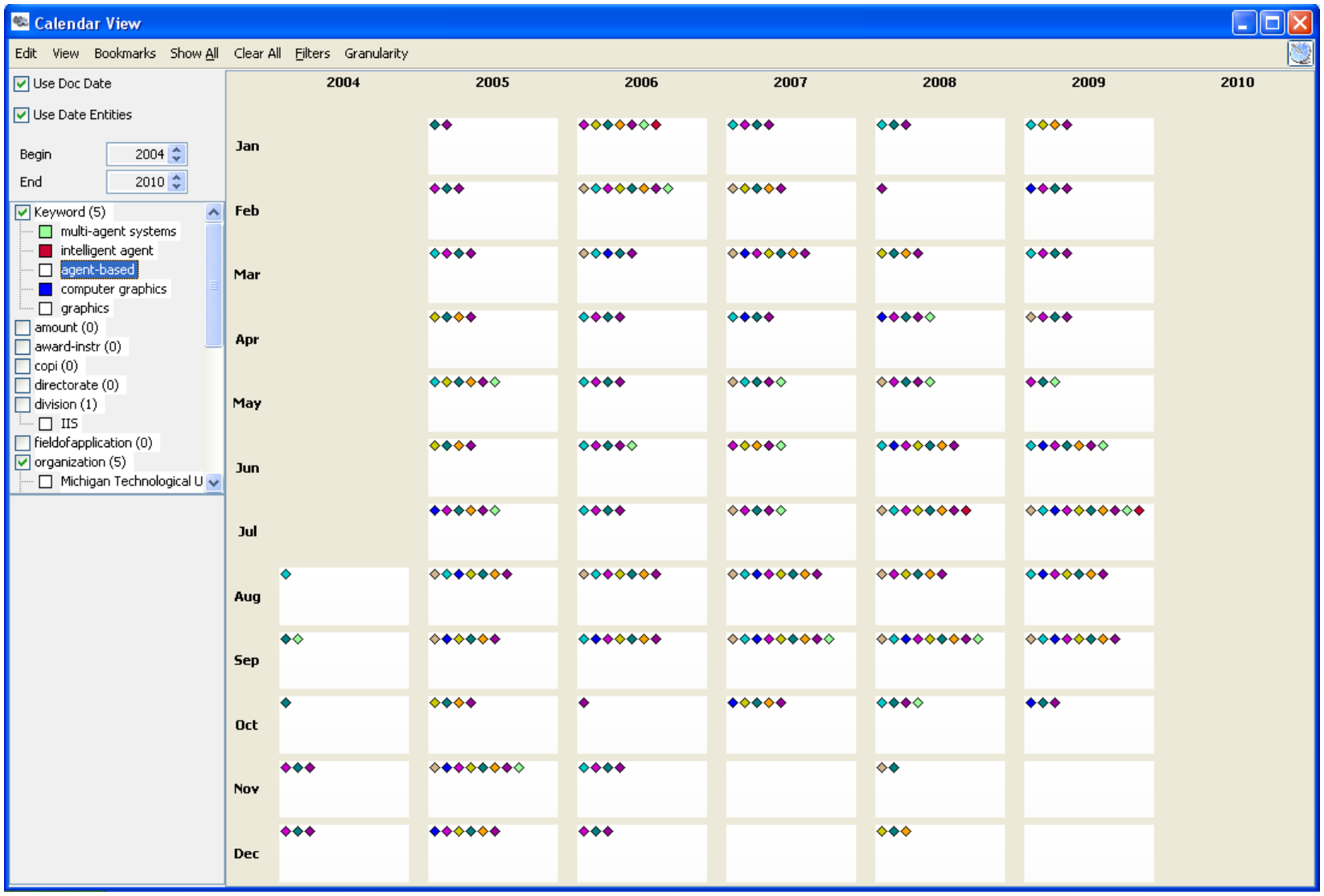
Made each page into a separate “document”
585 in total
Entities: Person, Location, Organization, Date, Money

Demo

Graph View

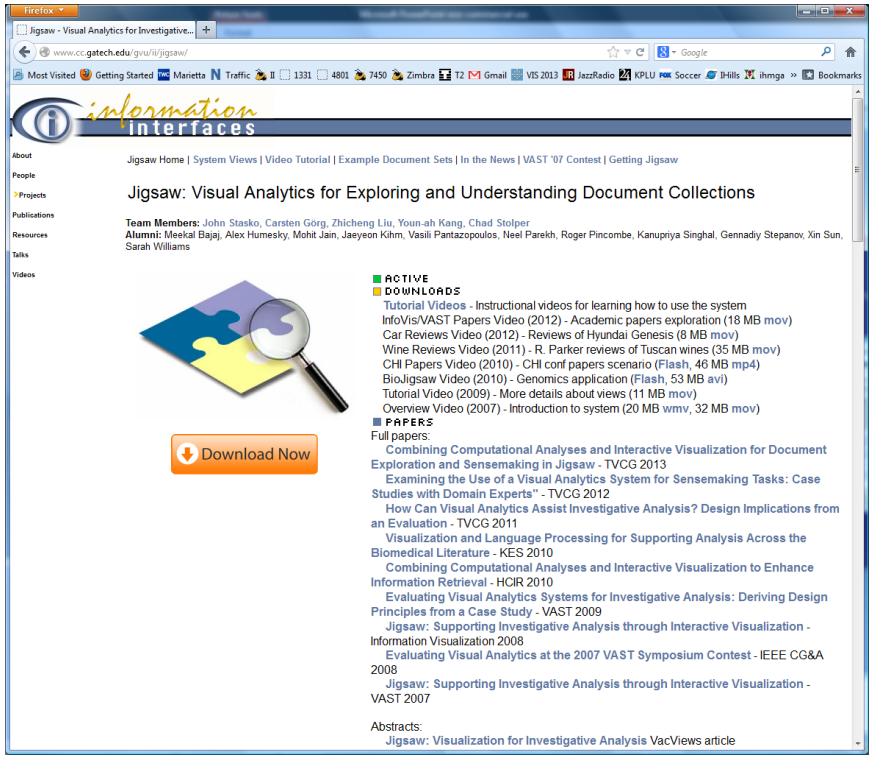

Graph View
Edit View Bookmarks Filter Items Select Items Options
Circular Layout
Pan&Zoom to fit
Nottola
montepulciano
Rossi di Montalcino
Rosso di Montepulciano
generous
attractive
round
orange
Nottola
2009
red
dark
fruit
264654
Table
0
Young
264507
freshness
tar
fleshiness
delicious
2
Dry
Italy
0
Mature
Tuscany
88
Montepulciano
Az Agr Dei
Red
Dei; Az Agr
212171
235984
flowers
2007
Canneto
2008
spices
fresh
Canneto
complexity
licorice
berries
balance
floral
purity
juicy
plump
softness
Rosso di Montepulciano
vibrant

Calendar View



Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw





Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

Further benefits of visualization

Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

EuroVis ‘14 Capstone Talk



http://vimeo.com/98986594



Take Aways

Presentation & analysis

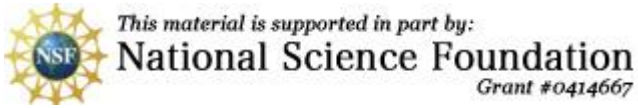
Interaction provides the power

Exploring & developing questions

Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

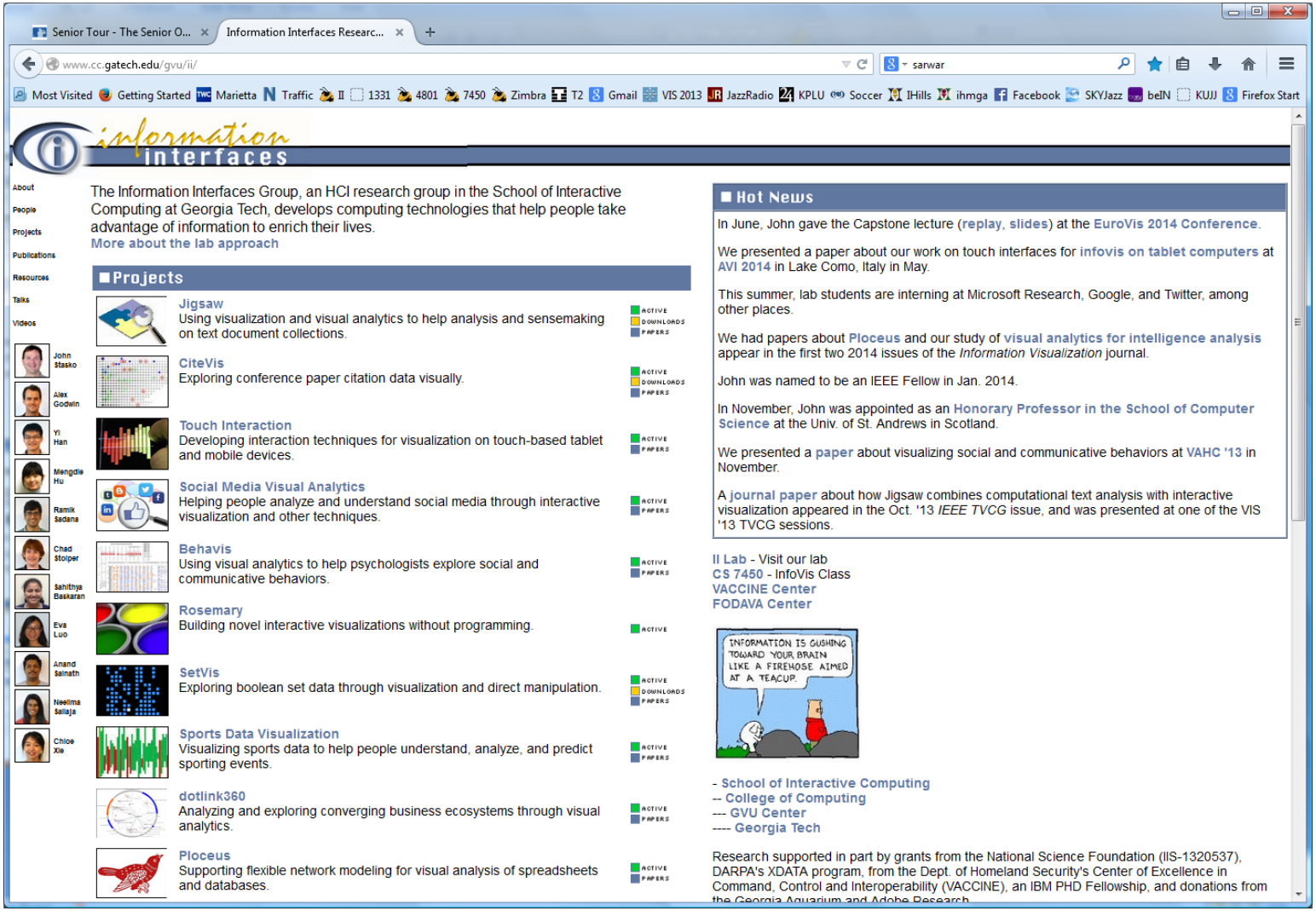
- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA Program

My Research

http://www.cc.gatech.edu/gvu/ii



Value-Driven Evaluation of Visualizations

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu







“No user studies are reported in this paper,
thus it is unclear if the visualization will be useful.”

Really?

- Which cars have the best miles-per-gallon?
- How much does a Ford Taurus cost?
- Is there a correlation between car weight and torque?
- How many different countries manufacture cars?
- What is the range of car’s horsepower?

~~Evaluation~~

Explanation(s) and Illustration(s) of value

Visualization's Value?

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to minimize the total **time** needed to answer a wide variety of questions about the data

(Without formal queries, Interaction really helps)

Visualization's Value?

$$V_{alue} = T + \mathbf{I} + E + C$$

Ability to spur and discover **insights** or insightful questions about the data

(Would be very difficult with only the data)

Visualization's Value?

$$V_{alue} = T + I + E + C$$

Ability to convey an overall **essence** or take-away sense of the data

(The big picture: Whole is greater than the sum of the parts)

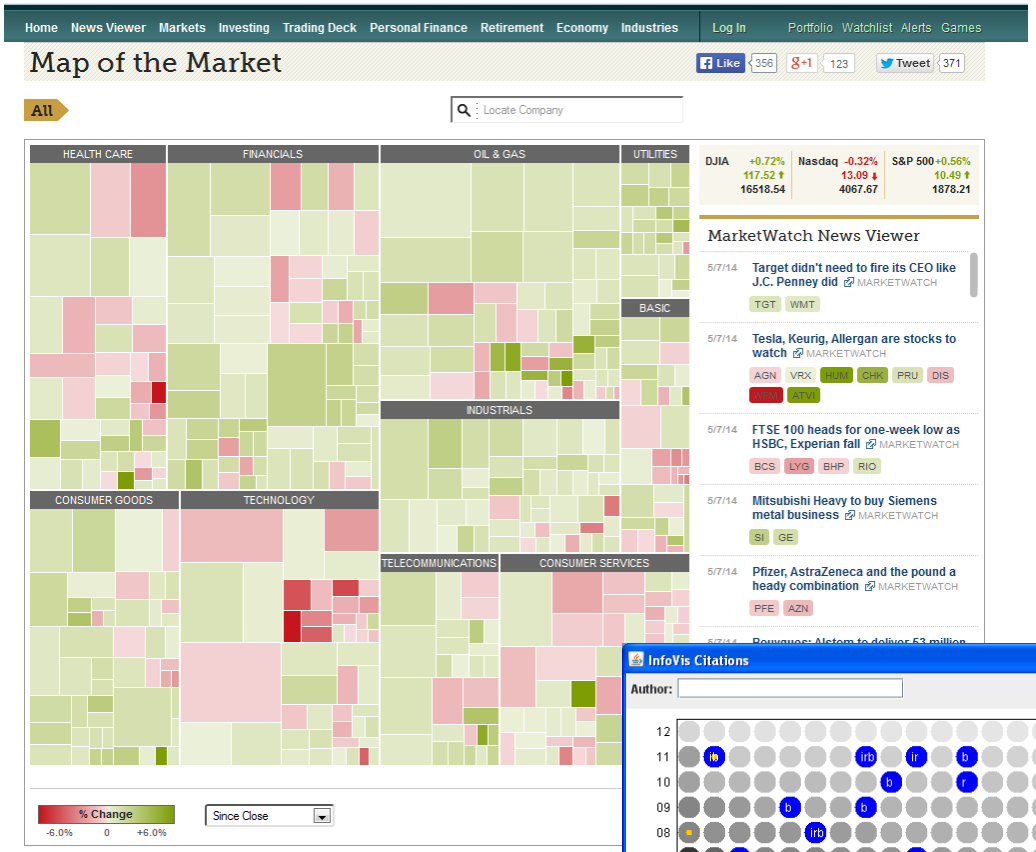
Visualization's Value?

$$Value = T + I + E + C$$

Ability to generate **confidence** and trust about the data, its domain and context

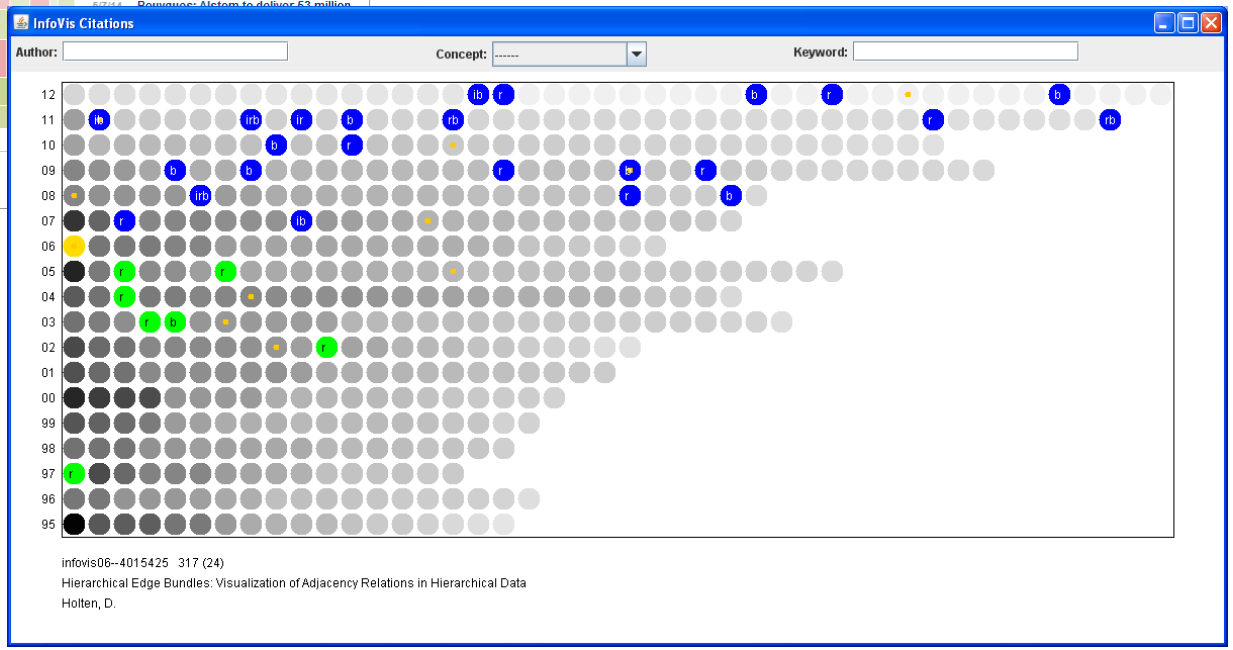
(Beneficial data analysis process side effects)

http://www.marketwatch.com/tools/stockresearch/marketmap



Map of the Market

CiteVis



http://www.cc.gatech.edu/gvu/ii/citevis

Data Visualization 101



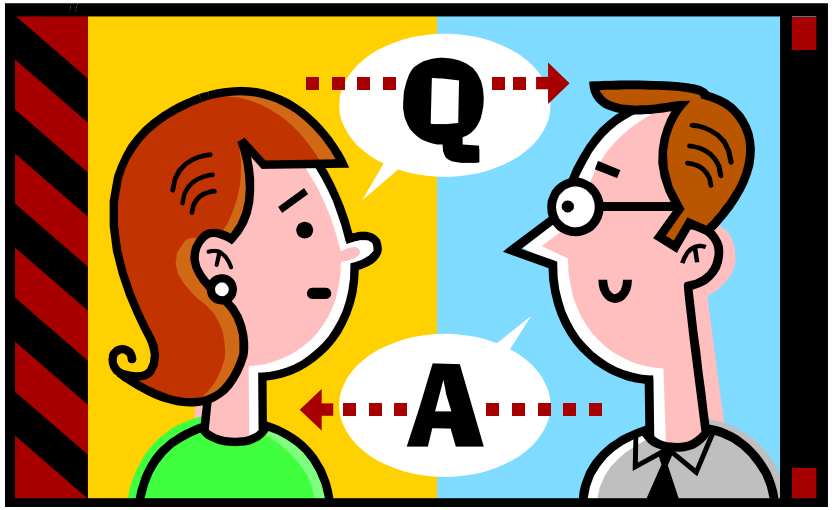
Problem:
You have a lot of data (& attributes) to understand

Do you?

- Pack all the data into one complex representation
- Spread the data into multiple coordinated views
- Use interaction to reveal different subsets of the data

Interaction is Vital

Engage in a dialog with your data



"The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means."

S. Few, *Now you see it*

Interaction

Understand the benefits (science of interaction)

Explore new interactive operations

Visualization on other display devices (touch)

http://vimeo.com/98986594

For more, see my EuroVis ‘14 Capstone Talk



Acknowledgments

- Supported by the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)



- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by DARPA's XDATA program

July 28, 2014

Visual Analytics for Investigative Analysis and Exploration of Documents and Data

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu





Georgia Tech College of Computing
School of Interactive Computing

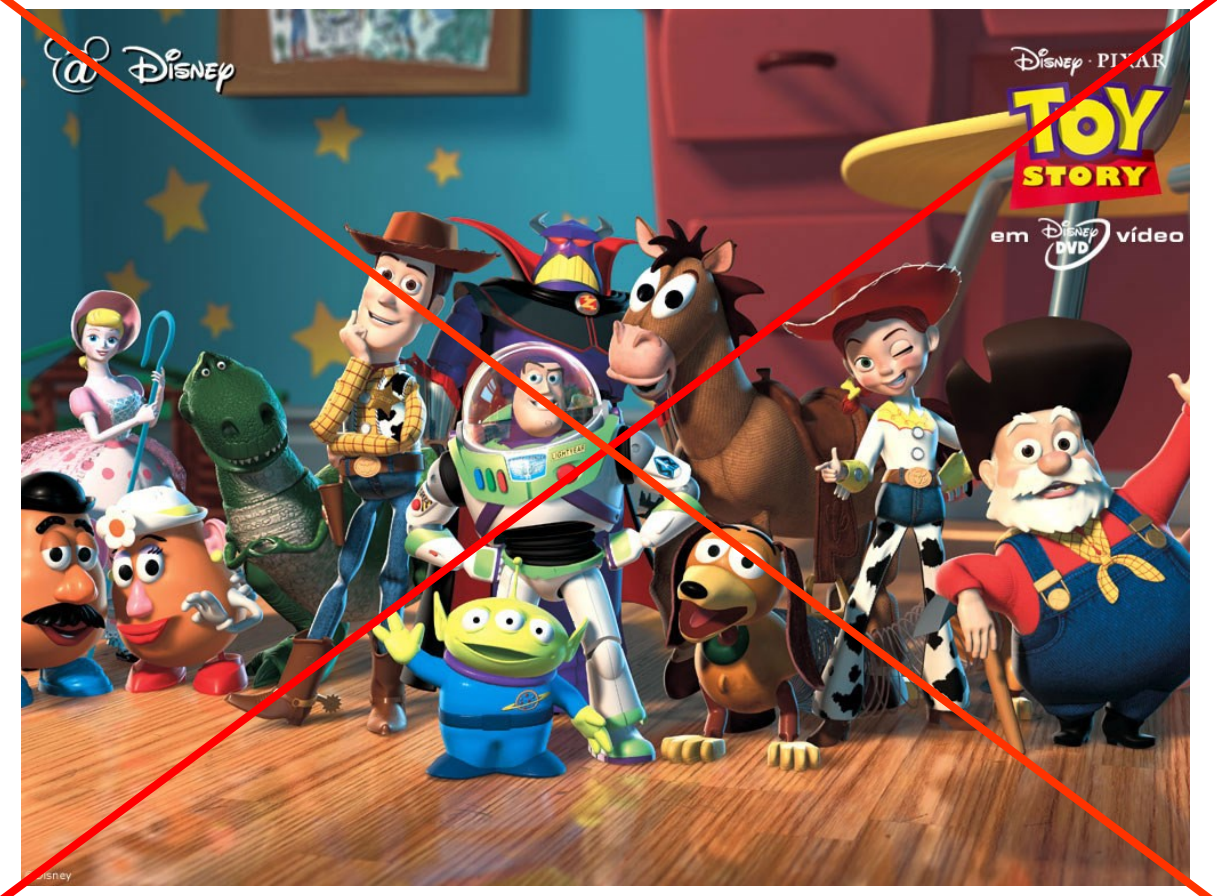
Data Visualization

Making pretty pictures?



Data Visualization

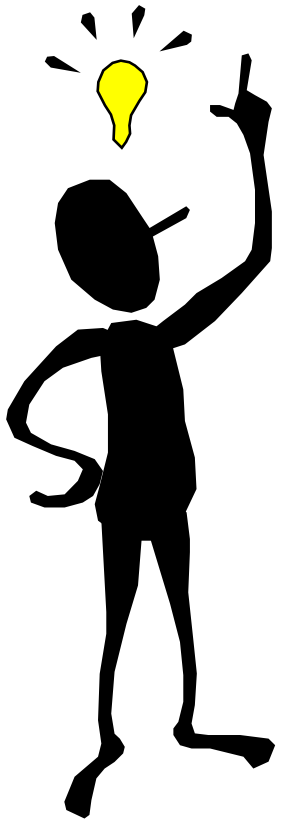
~~Making pretty pictures~~



Data Visualization

A cognitive process

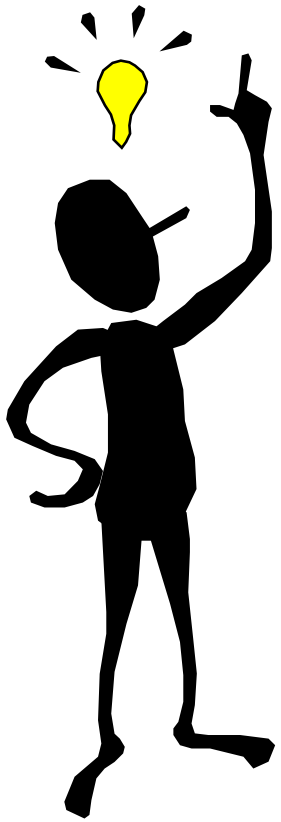
Gain an understanding

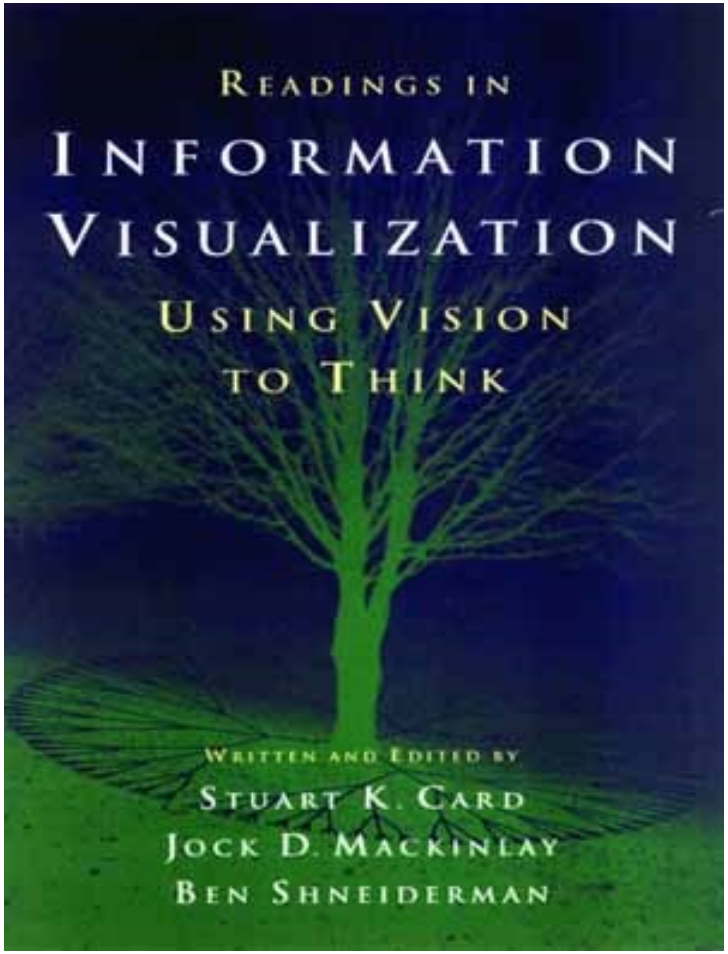


Data Visualization

A cognitive process

Gain an **understanding**





Visualization

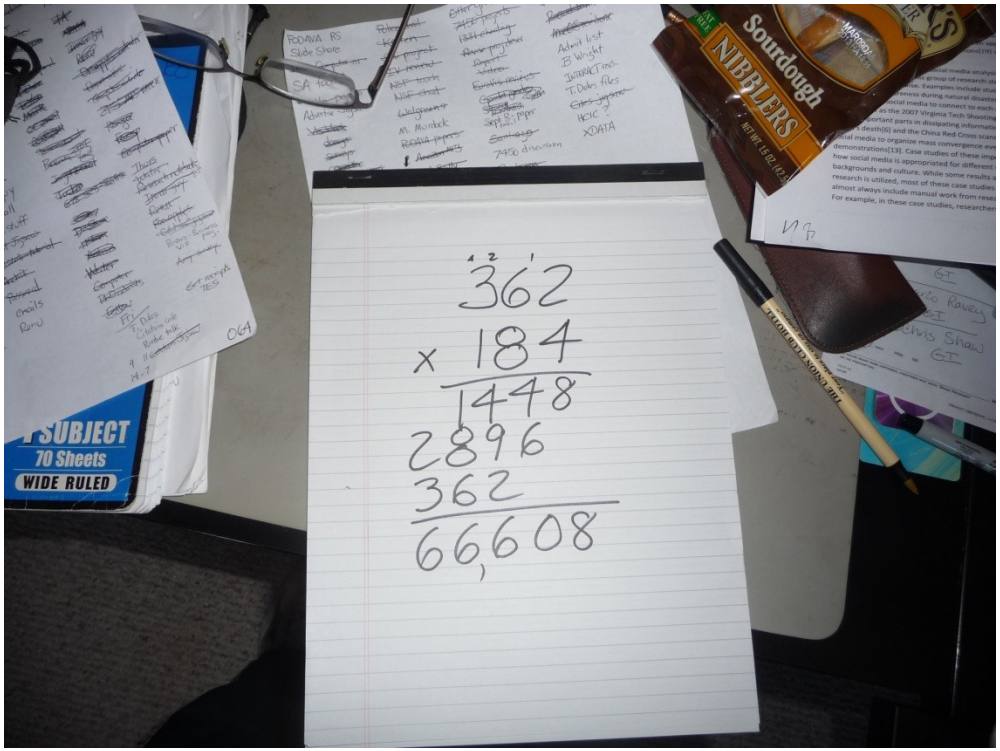
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provide a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

Pattern matching



Applications of Visualization

Presentation

Analysis

1. Presentation

Communicate data and ideas

Explain and inform

Influence and persuade

Provide evidence and support

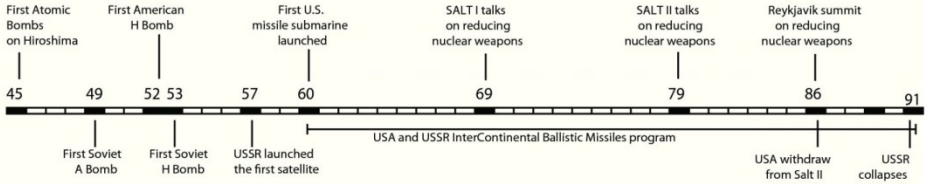


Infographics

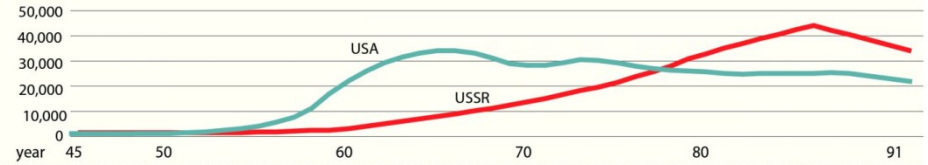
THE NUCLEAR ARMS RACE

It was the main issue in the Cold War when both America and Russia challenging each other to increase their stockpiles of nuclear weapons.

TIMELINE



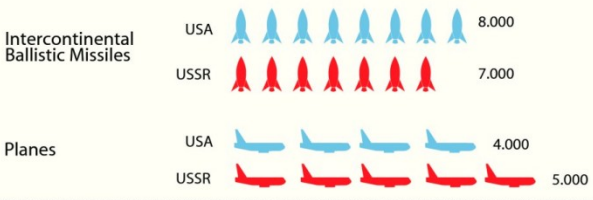
NUCLEAR STOCKPILE



ROCKET MODELS



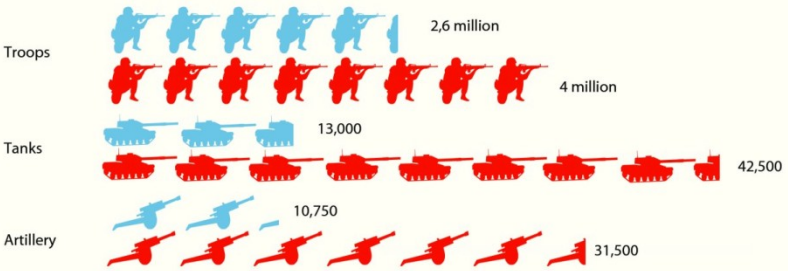
AIR CONTROL



A WORLD BREAK IN TWO

NATO and Warsaw Pact were both mutual defense treaties between states.
The first one started on 1949 till today while the Warsaw Pact lasted from 1955 till 1991.



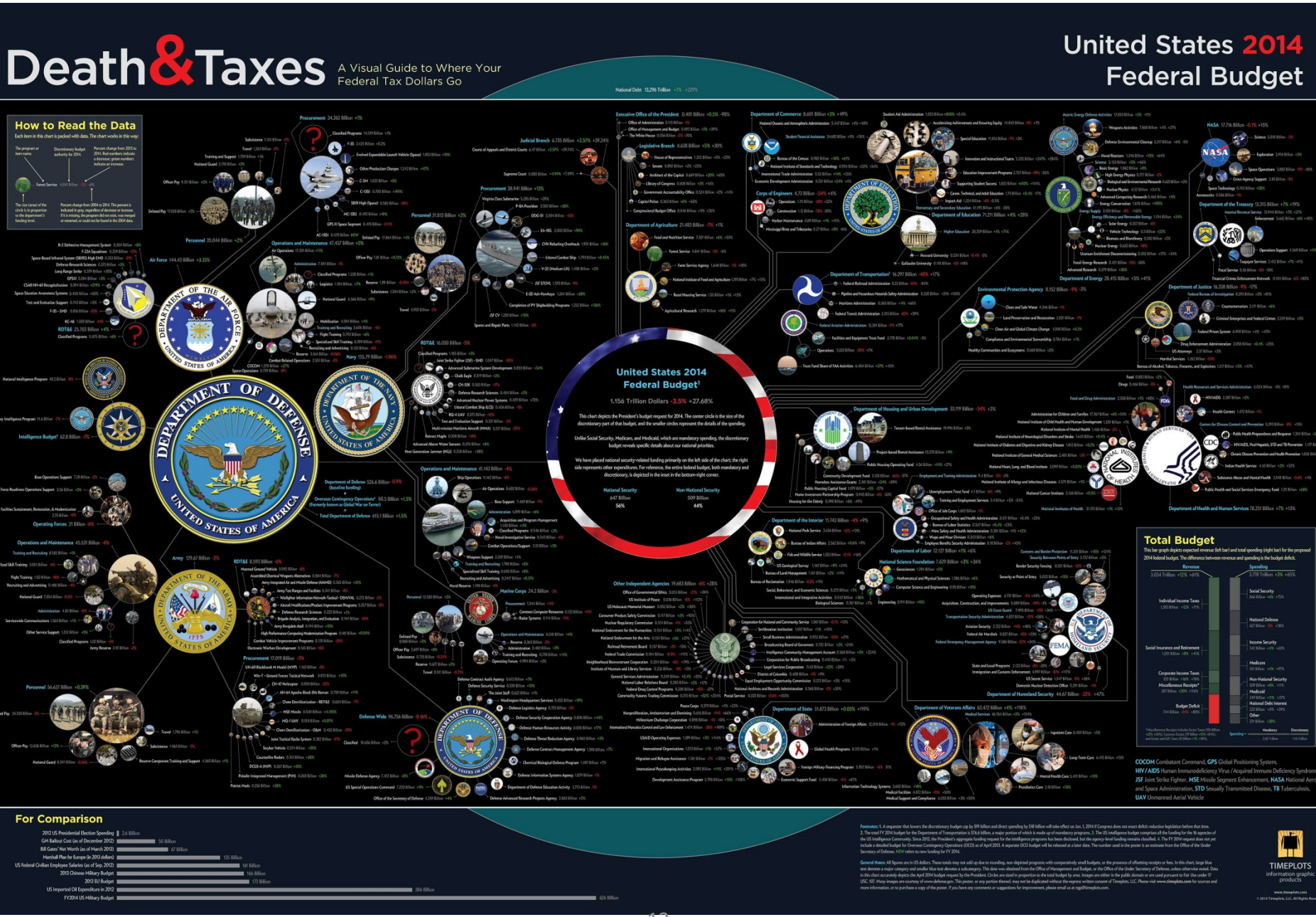


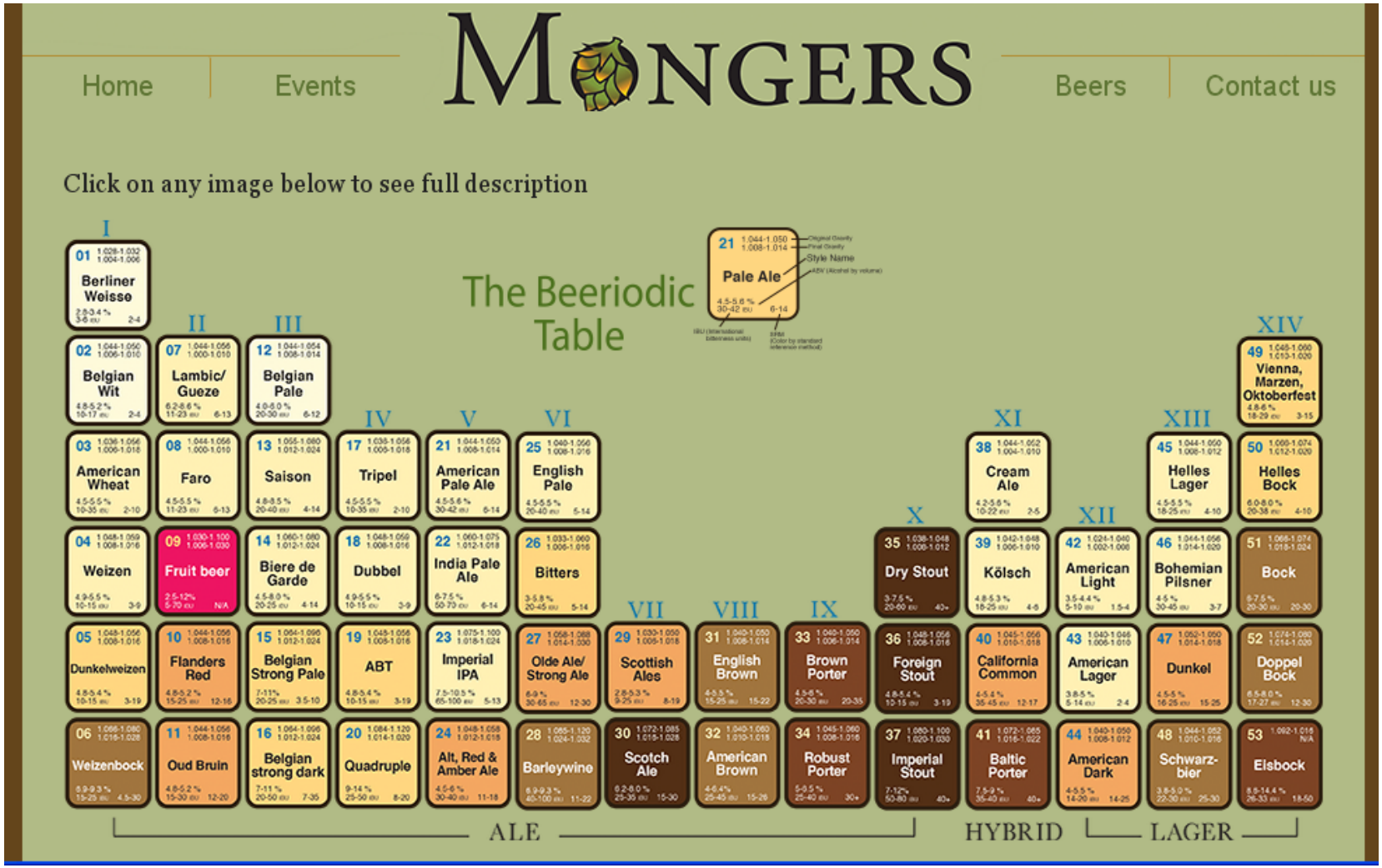
Produced by WatchTheAmericans.com a Fan Website

Designed By Gino Selva

http://visual.ly/nuclear-arms-race

http://visual.ly/death-and-taxes-2014-us-federal-budget





http://thebeermongers.com/beers/

NY Times 2006

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.



Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons



Babe Ruth
714 homers
22 seasons



Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons
Bonds was injured last season. He played 14 games and hit 6 homers

600
400
200
0
14th season
5 seasons
10
15

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

Aaron
Actual homers slightly outpace projected homers for five seasons.



Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

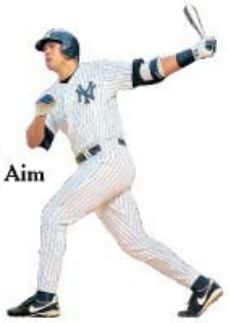


Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

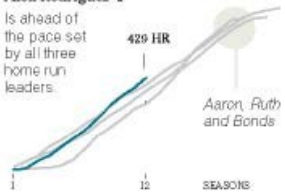


Note: Ages as of July 1 of each season.

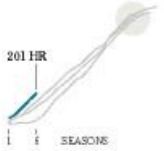
Others Taking Aim



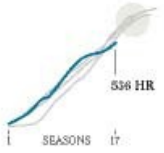
Alex Rodriguez
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

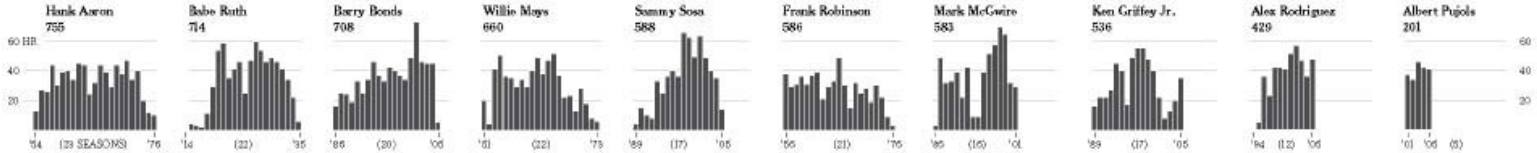


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

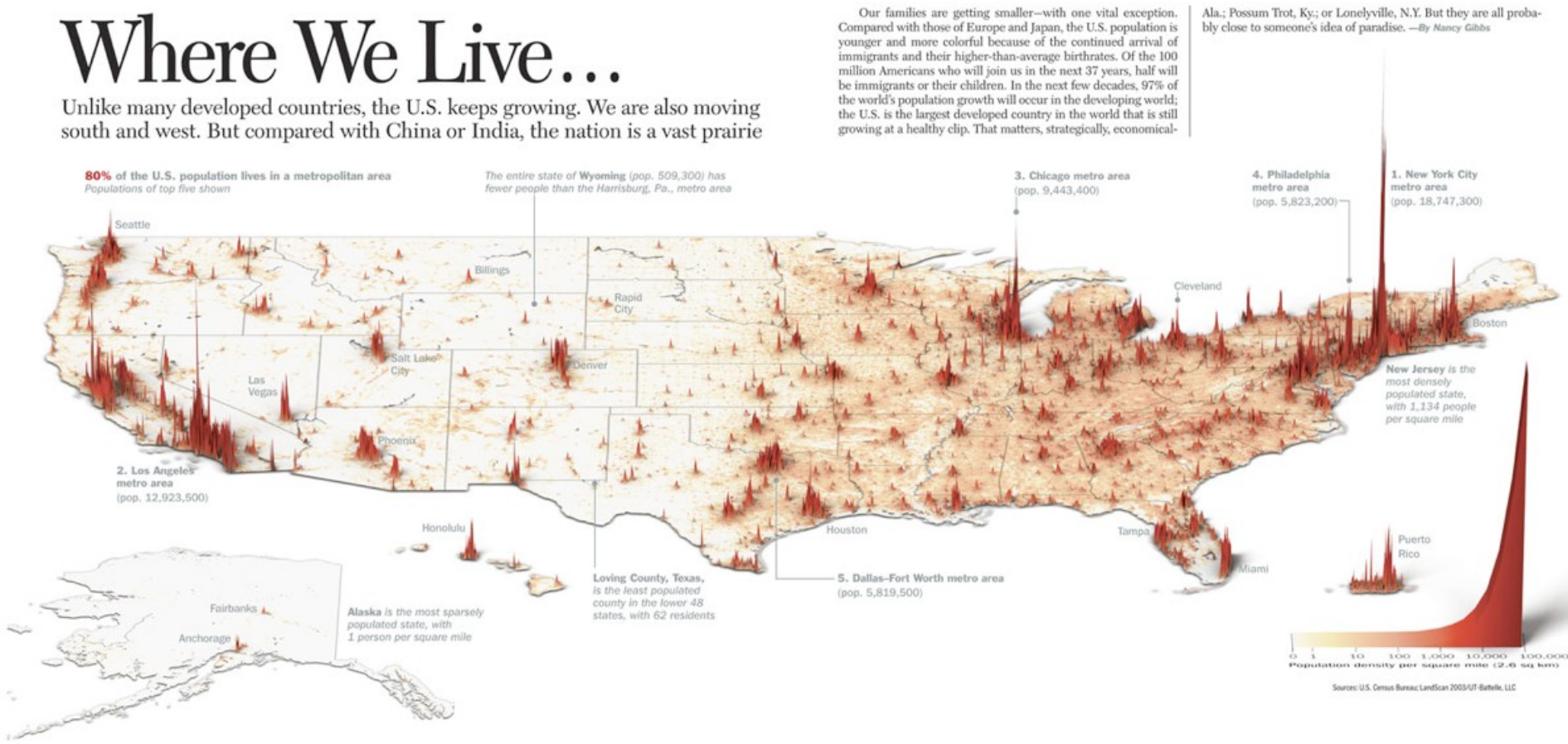


Differing Paths to the Top of the Charts

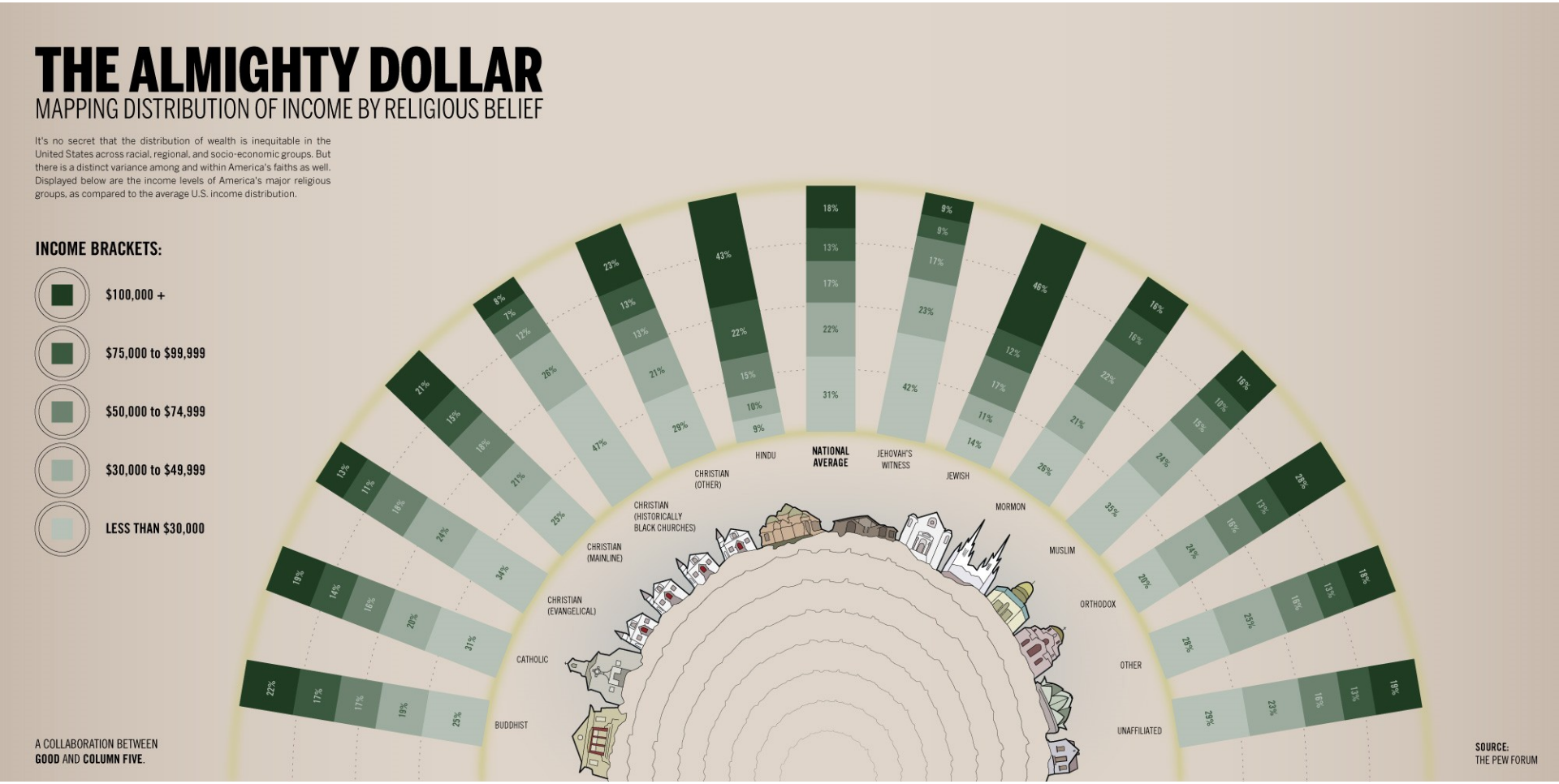
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



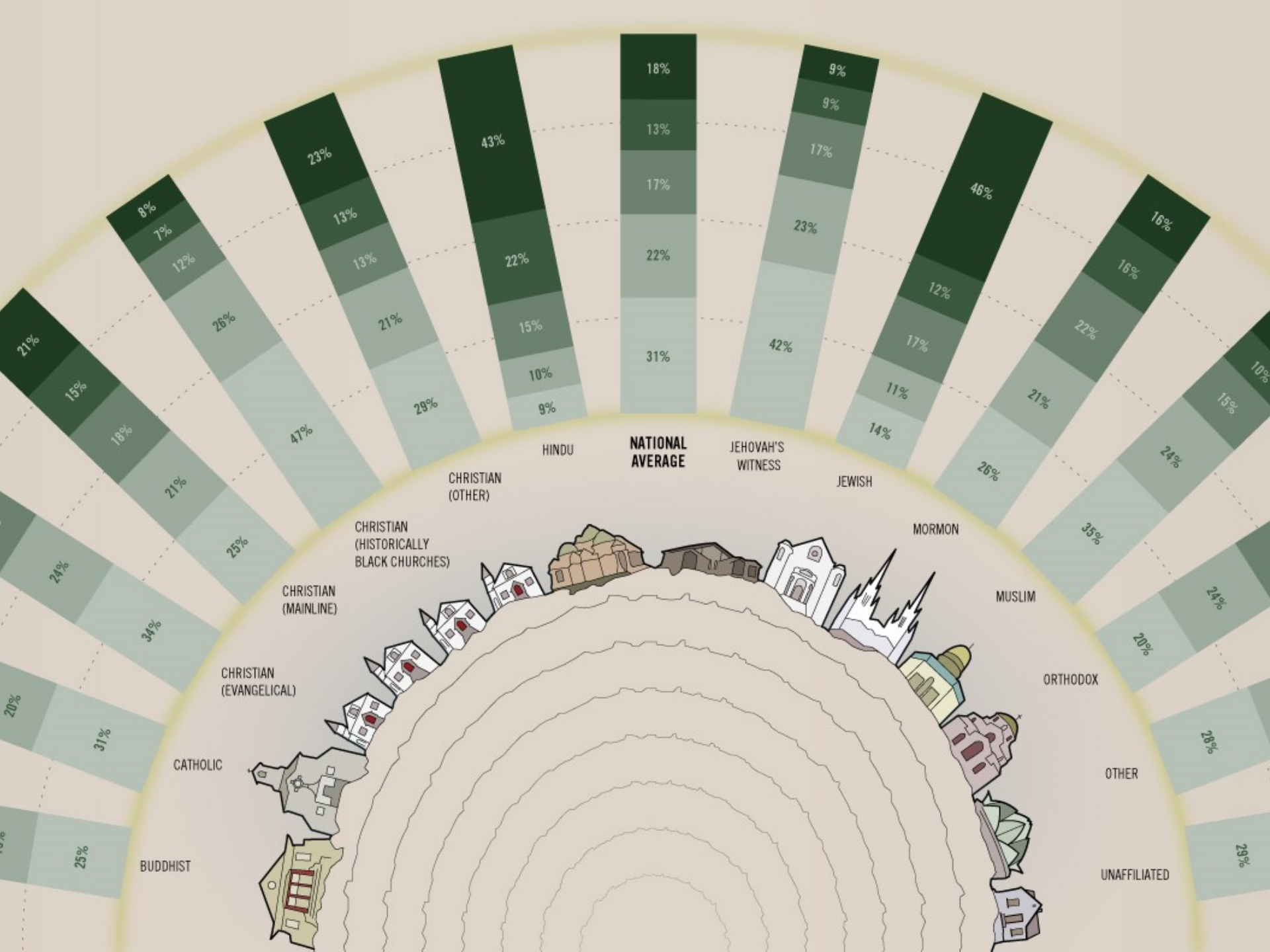
Amanda Cox and Joe Ward/The New York Times



http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html

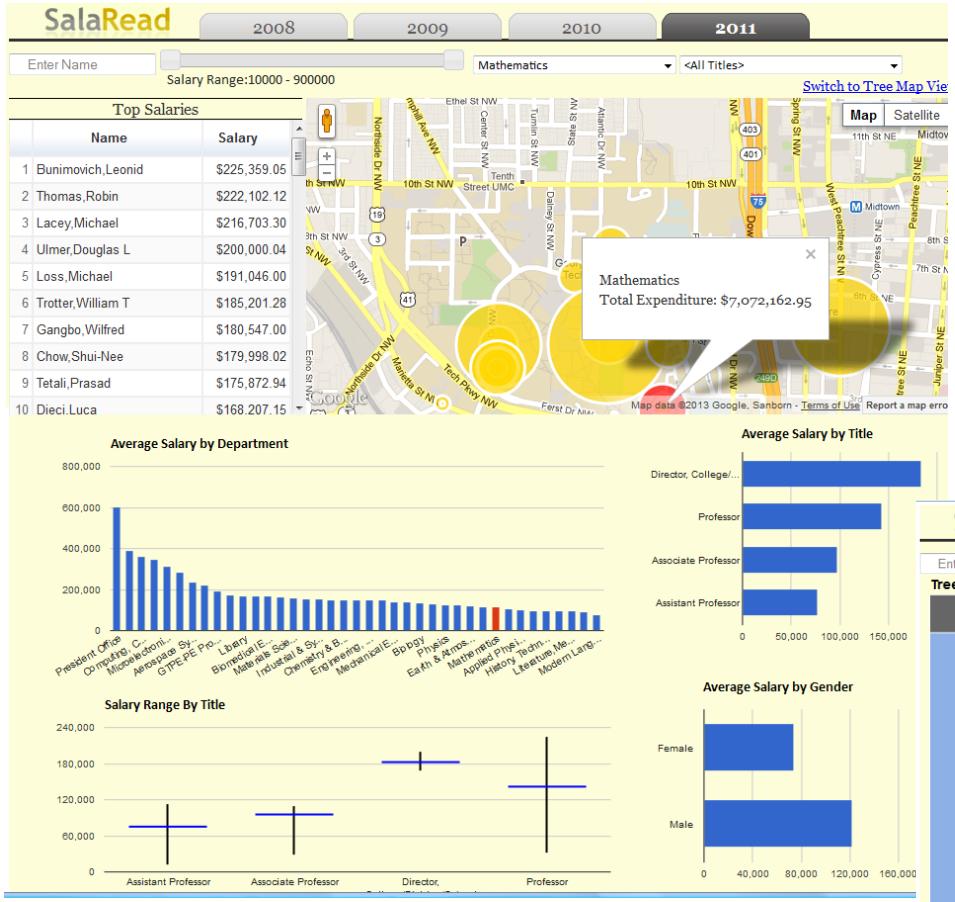


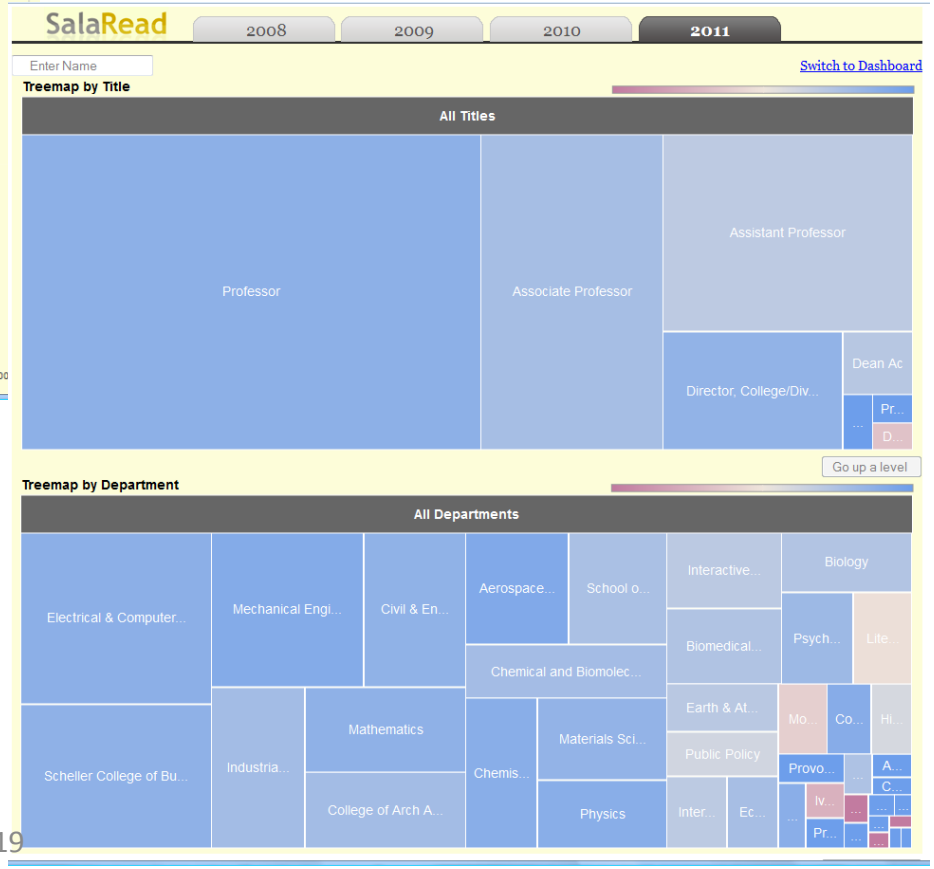
http://awesome.good.is.s3.amazonaws.com/transparency/web/1002/almighty-dollar/transparency.jpg


BUDDHIST
21%
15%
18%
21%
25%
CATHOLIC
20%
31%
24%
34%
CHRISTIAN (EVANGELICAL)
8%
7%
12%
26%
47%
CHRISTIAN (MAINLINE)
23%
13%
13%
21%
29%
CHRISTIAN (HISTORICALLY BLACK CHURCHES)
43%
22%
15%
10%
9%
CHRISTIAN (OTHER)
18%
13%
17%
22%
31%
HINDU
NATIONAL AVERAGE
9%
9%
17%
23%
42%
JEHOVAH'S WITNESS
46%
12%
17%
11%
14%
JEWISH
16%
16%
22%
21%
26%
MORMON
10%
15%
24%
35%
MUSLIM
20%
24%
28%
ORTHODOX
29%
OTHER
UNAFFILIATED

Simply presenting data ***visually*** can have a profound impact

CS 7450
Fall '12
Design project

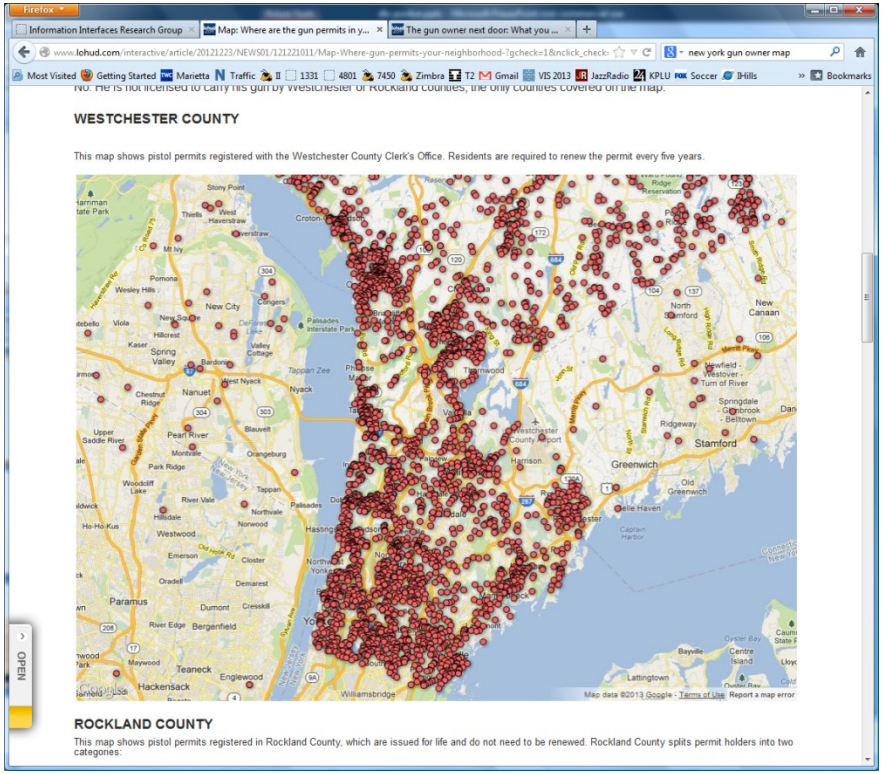




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun ownership in New York counties





http://www.lohud.com/apps/pbcs.dll/article?AID=2012312230056&nclick_check=1

http://www.lohud.com/interactive/article/20121223/NEWS01/121221011/
Map-Where-gun-permits-your-neighborhood-?gcheck=1&nclick_check=1

Frequent presentation goals

Clarify

Focus

Highlight

Simplify

May just show a few variables
and/or a subset of the data cases

2. Analysis

understand, compare, decide, judge, evaluate, assess, determine, …



Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning





"Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data…To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection."

W. Cleveland
The Elements of Graphing Data

Frequent analysis goals

Show many variables

Illustrate overview and detail

Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility

Answering specific questions and accomplishing specific analytic tasks

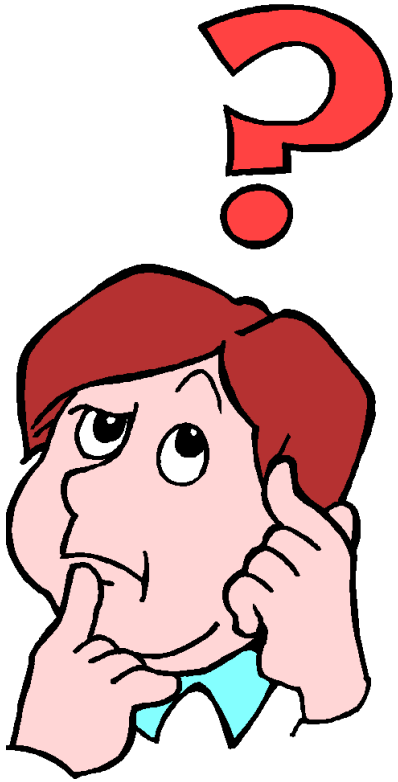
Generating unexpected, serendipitous discoveries and insights

"Finding a needle in a haystack"



Yes, but not what it's best for

So what is visualization most useful for?



1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context

Many Data Analysis Approaches

Statistics

Database & information retrieval

Data mining

Machine learning

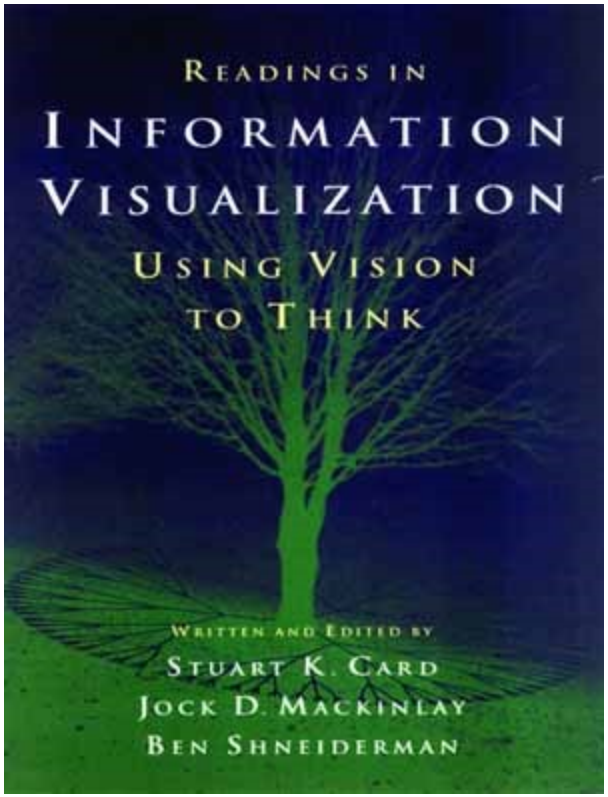




Use them when they get the job done!

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction



Visualization

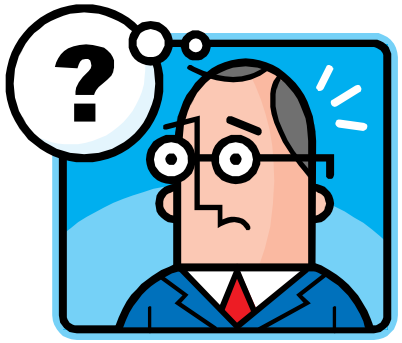
"The use of computer-supported, **interactive** visual representations of data to amplify cognition"

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask



Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Example Domains for Visual Analytics

- Text & documents
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Text is Everywhere

We use documents as primary information artifacts in our lives

Our access to documents has grown tremendously in recent years due to the networking infrastructure

- WWW
- Digital libraries
- …

Why Visualize Text & Documents?

For what purpose(s)?
Accomplish what tasks?
Help with which problems?

Example Tasks and Goals

- Which documents contain text on topic XYZ?
- Which documents are of interest to me?
- Are there other documents that are similar to this one (so they are worthwhile)?
- How are different words used in a document or a document collection?
- What are the main themes and ideas in a document or a collection?
- Which documents have an angry tone?
- How are certain words or themes distributed through a document?
- Identify “hidden” messages or stories in this document collection.
- Quickly gain an understanding of a document or collection in order to subsequently do XYZ.
- Find connections between documents.

The Challenge

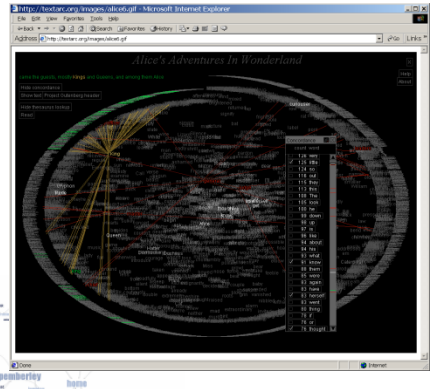
Text is nominal data

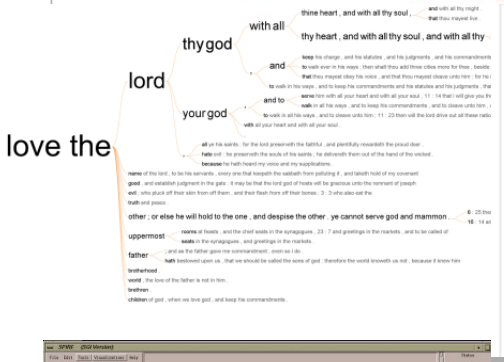
Does not seem to map to geometric/graphical presentations as easily as ordinal and quantitative data

- Bar charts, line charts, scatterplots, etc.









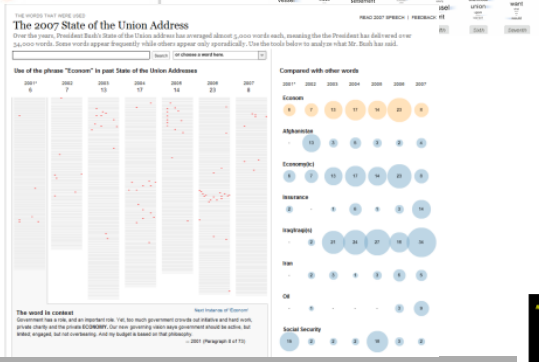
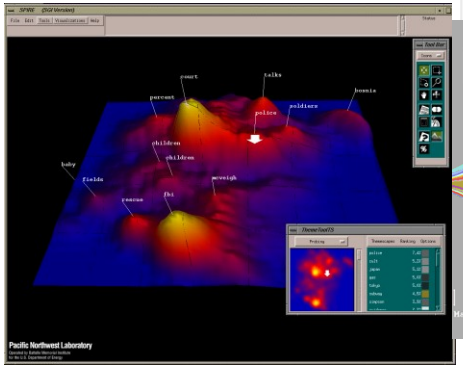
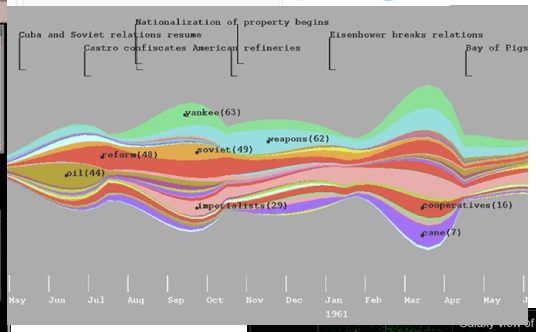
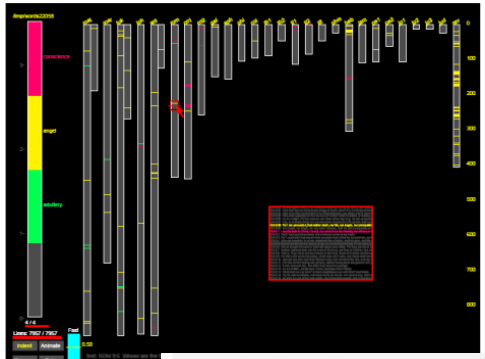


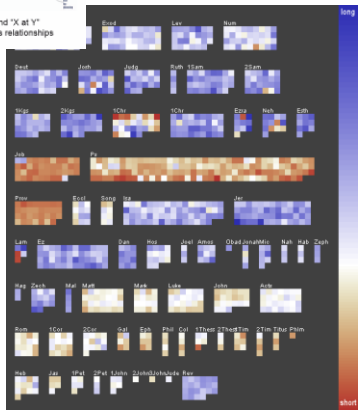


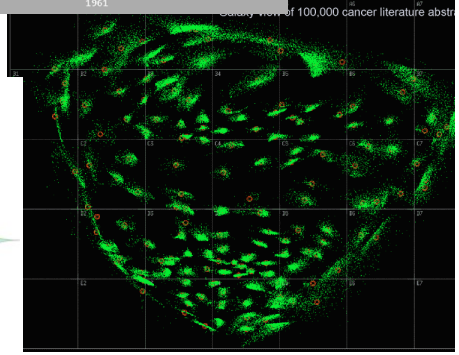
Fig 5. Matching different patterns on the same text. Here we analyzed Jane Austen's *Pride and Prejudice* with "X and Y" and "X at Y" respectively. The left image shows relationships between the main characters amongst others, while the right image shows relationships between locations.

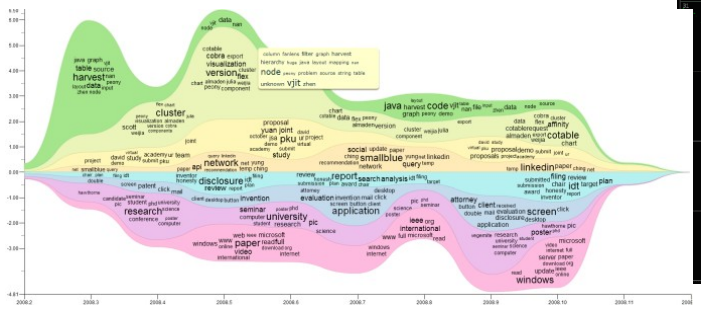












All time most popular tags

"Putting the pieces together"

Jigsaw

Computational analysis of documents' text

Entity identification, document similarity, clustering, summarization, sentiment

Multiple visualizations of documents, analysis results, entities, and their connections

Views are highly coordinated



MEMBER LOGIN: Username: Password: LOGIN

Forgot My Username/Password

Remember Login Info

eRobertParker.com

The Independent Consumer's Guide to Fine Wines



Rock and Rhône:
Parker Explores the Greatest Unheralded Grapes in the World–Grenache, Mourvèdre, and Syrah

HOME SUBSCRIBE GIFT SUBSCRIPTIONS FAQs VIRTUAL TOUR SUPPORT SITE MAP CONTACT US

What's Inside

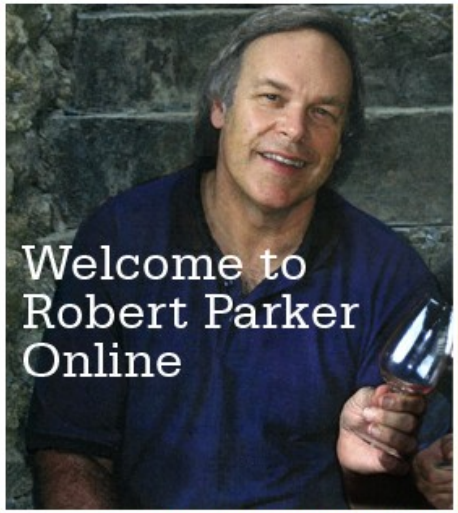
Explore Wines
Over 190,000 searchable notes from *The Wine Advocate*, Robert Parker's books and other sources.

The Hedonist's Gazette
An insider's look at our reviewers' wine and food experiences.

Article Archive
In-depth commentary from *The Wine Advocate* and other sources.

My Wines
Powerful, easy-to-use tools for purchasing/managing a cellar and participating in our wine community.

Robert Parker Mobi
Search the notes from any Internet-enabled mobile device.





Information and Products for Non-Subscribers

Wine of the Day:
2009 Ridge Chardonnay
View Tasting Note



3W's of Wine iPhone App
Learn More

The Wine Advocate Store
Shop Now

Parker & Zraly Wine Certification
Learn More

Weekly Wine Buys
Powered by winealert
Learn More

Quick Links	More Information	About Our Reviewers	Links
Username/Password	**The Vintage Chart**	Robert Parker	The Wine Advocate Fund
What's New	Executive Wine Seminars	Antonio Galloni	Educational
Gift Subscriptions	Glossary of Wine Terms	Karen MacNeil	Accessory/Storage Links
Get News Via RSS	Parker's Rating System	Neal Martin	Retailer Links
Shop The Wine Advocate Store	*The Wine Advocate*	Jay Miller	Other
TWA Editorial Calendar	Site Map	Lisa Perrotti-Brown MW	
		David Schildknecht	
		Mark Squires	
		Kevin Zraly	
		Past Reviewers	
		Writer Ethics/Standards	





Example

Reviews of wines from Washington state from ‘07-’10

Text: review narrative

Entities: variety, producer, rating, vintage, color, location, producer, “descriptor”, …

Descriptor terms (~9000), eg: abrasive, oaky, cherry, mocha, textured

1323 reviews

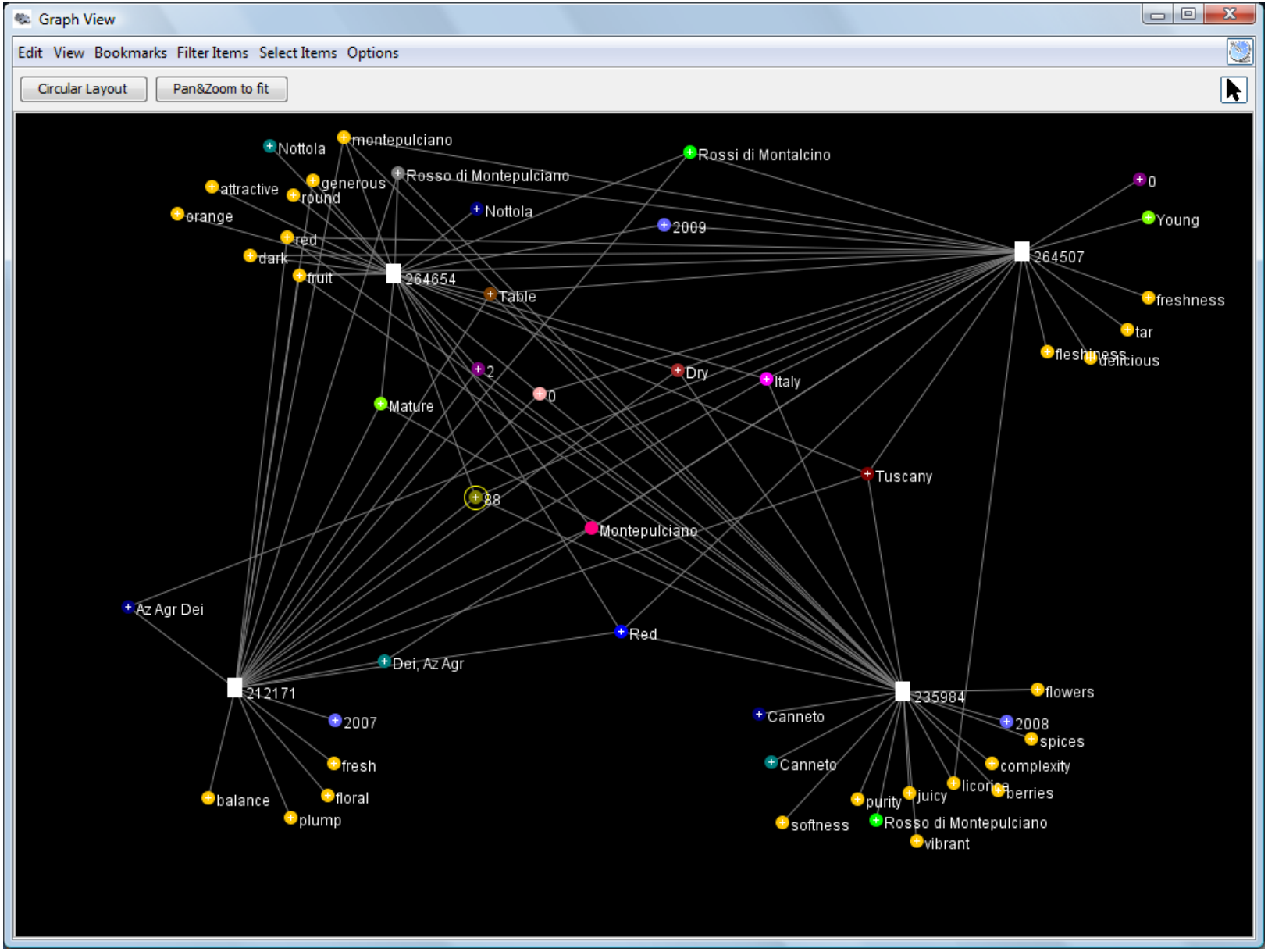
from database of 150,000 reviews



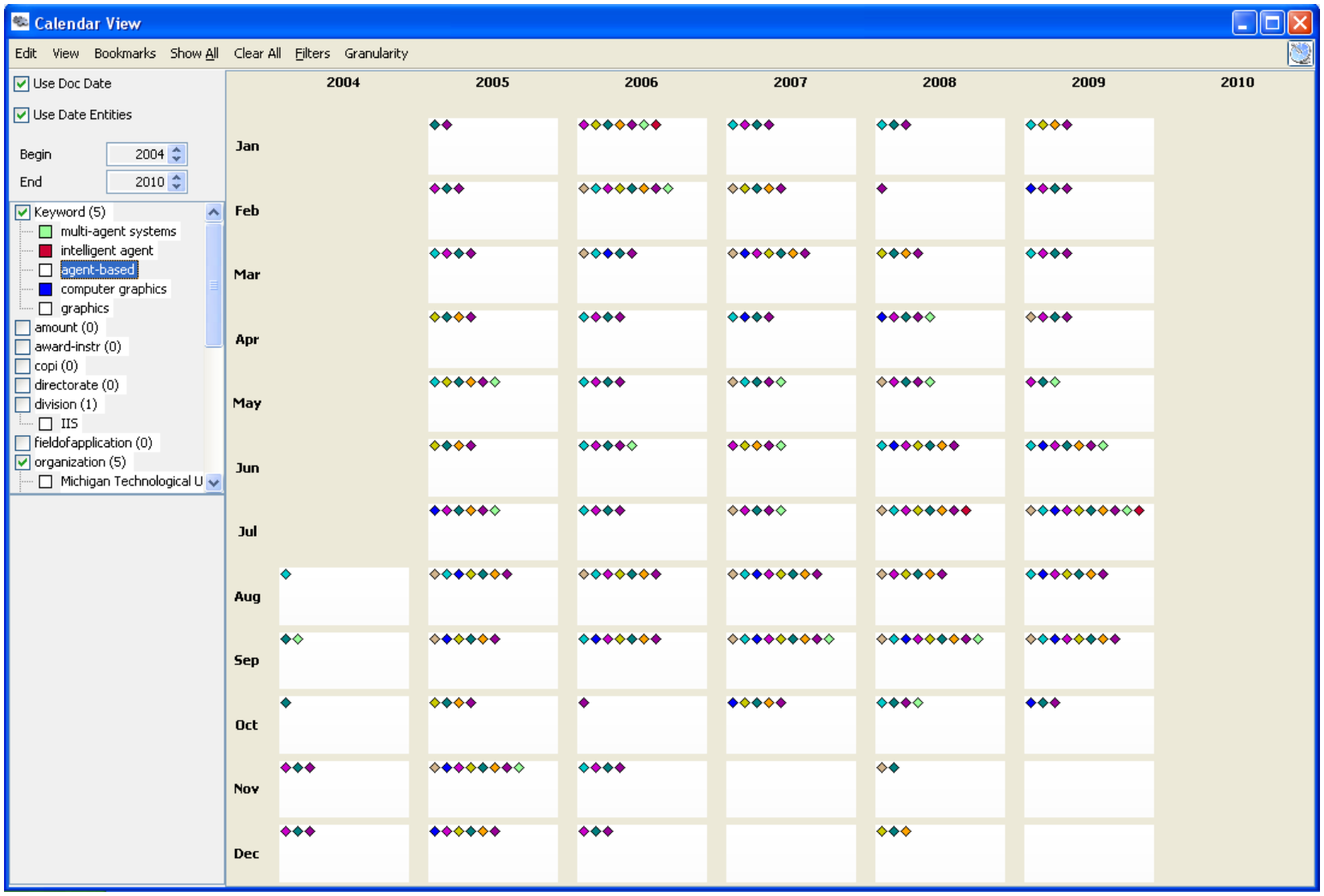
Demo



Graph View

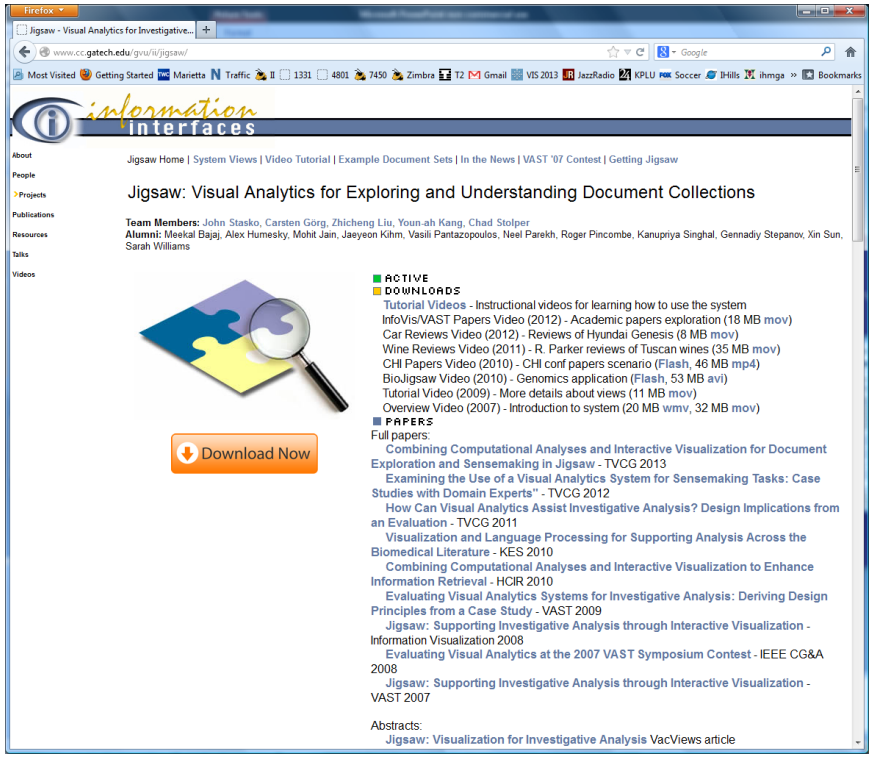

Graph View
Edit View Bookmarks Filter Items Select Items Options
Circular Layout
Pan&Zoom to fit
Nottola
montepulciano
Rossi di Montalcino
Rosso di Montepulciano
generous
attractive
round
orange
Nottola
2009
red
dark
fruit
264654
Table
0
Young
264507
freshness
tar
fleshiness
delicious
2
Dry
Italy
0
Mature
Tuscany
88
Montepulciano
Az Agr Dei
Red
Dei; Az Agr
212171
235984
flowers
2007
Canneto
2008
spices
fresh
Canneto
complexity
balance
floral
purity
juicy
licorice
berries
plump
softness
Rosso di Montepulciano
vibrant

Calendar View



Download for free

http://www.cc.gatech.edu/gvu/ii/jigsaw





Applications

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES '10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000
- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, wine reviews
  - Görg et al, HCIR '10
- Business Intelligence
  - Patents, press releases, corporate agreements, …
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis '10

1. Visualization is more than just answering specific questions about data (as is often the case for automated analysis methods); it also facilitates the investigative analysis process, which supports analysts in developing awareness of, learning about, and generating trust in their data, its domain, and its context.

Learning, awareness, trust, context

Data cleaning example

2. Visualization, primarily through its interactive capabilities, promotes a dialog of inquiry between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

Repeated interactive queries

3. Visualization rapidly and efficiently facilitates flexible exploration of data to foster both a broad and deep understanding of the information contained therein.

Broad and deep understanding quickly

Varieties, colors, reviewers, producers, concepts, …

Insights

No great aha! – More learning & knowledge building

Domain knowledge (wines) matters

Further benefits of visualization

Ease of specifying queries

Opportunistic discovery of relevant data

Spurs the generation of new questions

EuroVis ‘14 Capstone Talk



http://vimeo.com/98986594



Take Aways

Presentation & analysis

Interaction provides the power

Exploring & developing questions

Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)

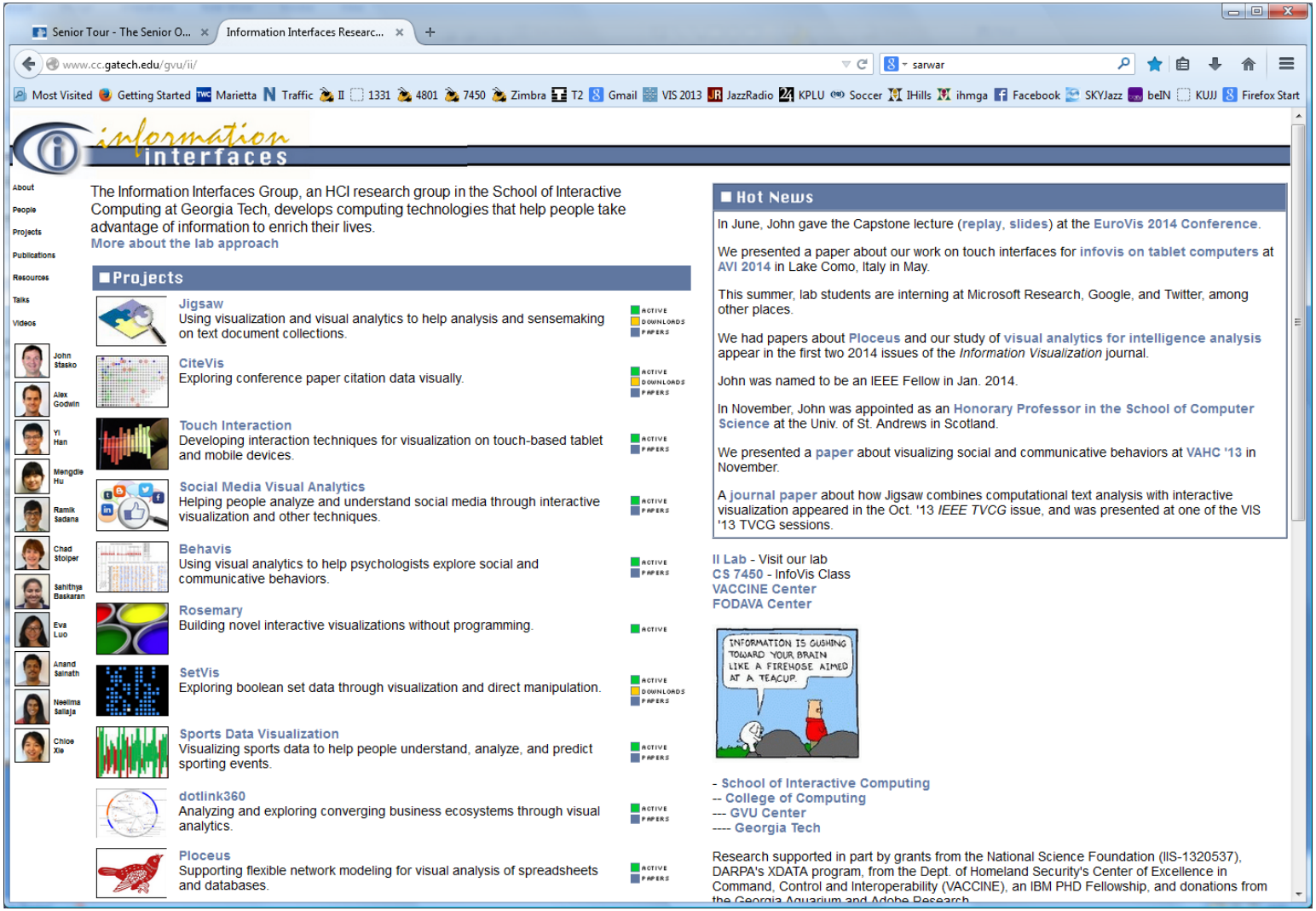
- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

This material is supported in part by:
National Science Foundation
Grant #0414667

- Supported by DARPA's XDATA Program

My Research

http://www.cc.gatech.edu/gvu/ii



iLeaps

Android Native App Development for iLaw Enforcement Application Program for Students

Kofi Nyarko | Benjamin Hall | Alfred Shoetan | Leonardo Ferreira | Roberta Virgil

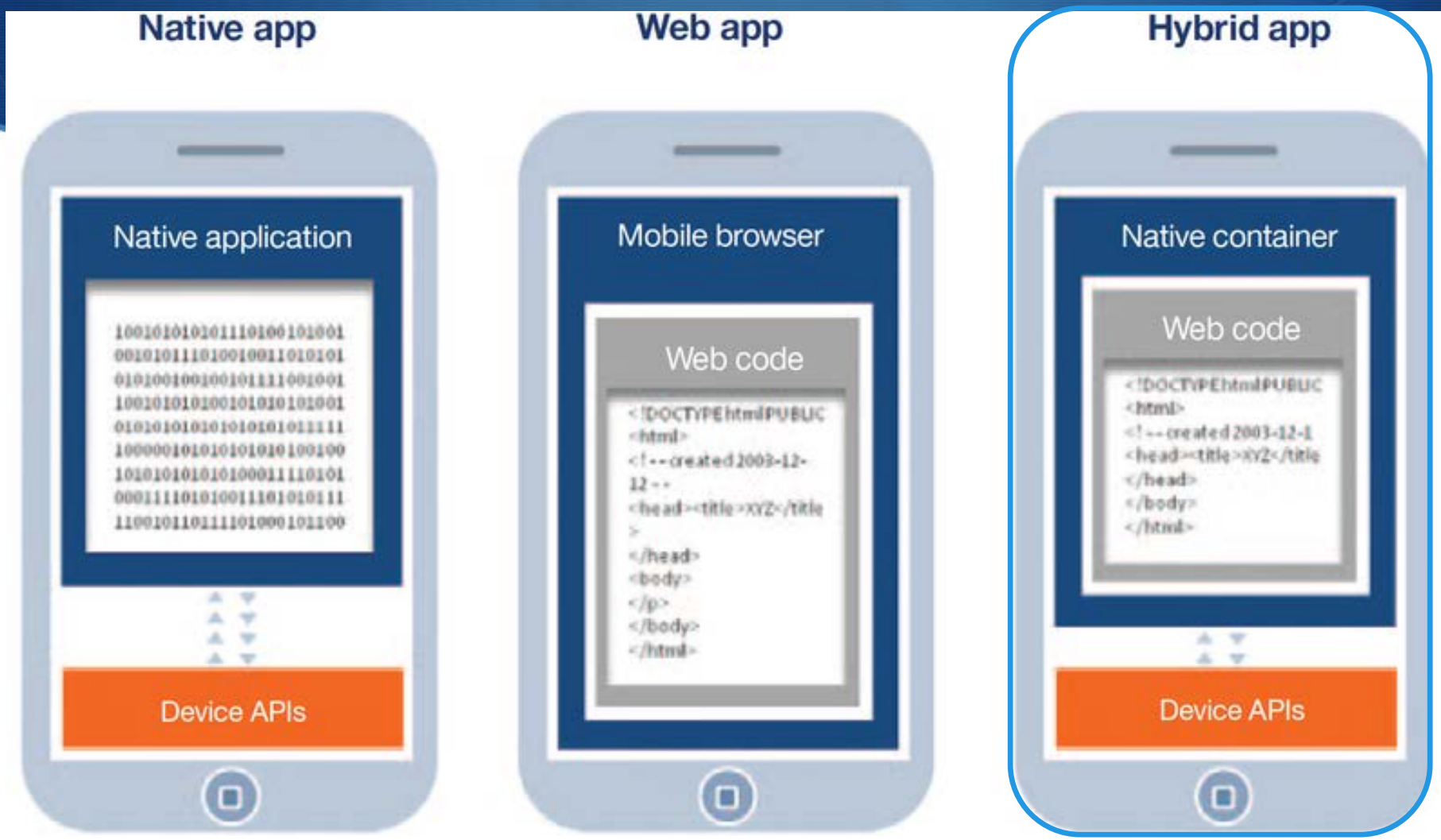
June 30, 2015

Background

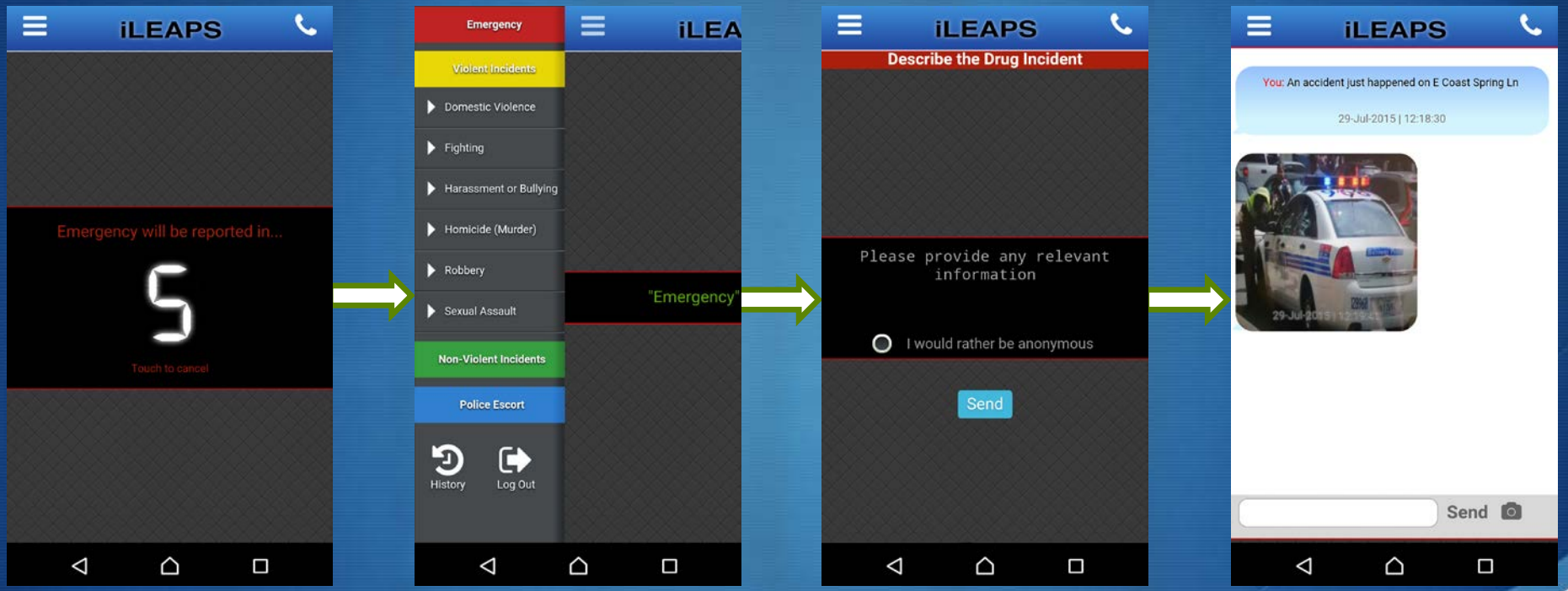
- Incidents on college campuses have made safety a central issue
- Need a platform that can improve response time for first-responders while simultaneously providing situational awareness
- Platform enables individuals to report incidents that can be geo-referenced and cataloged for forensic analysis as well as predictive analytics

- Decrease the response time of emergency responders to an incident

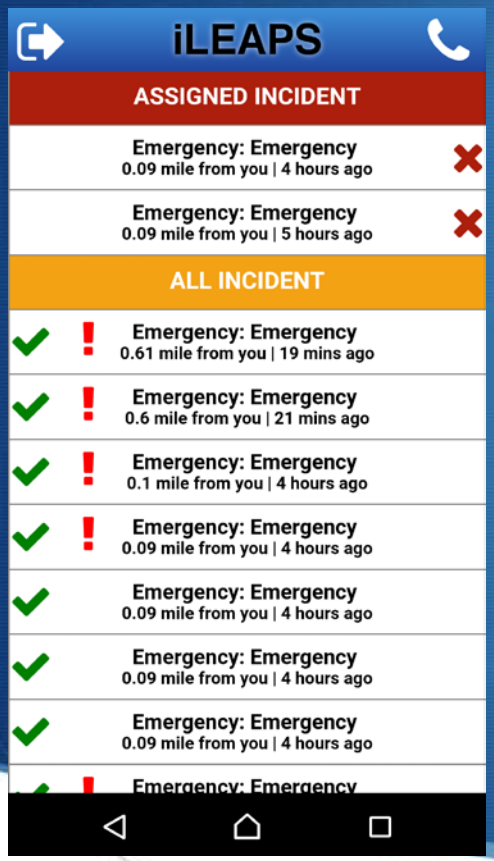


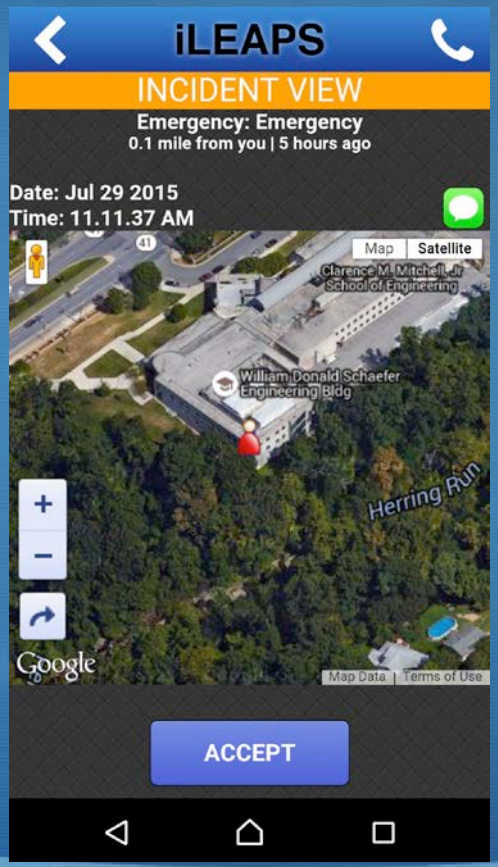

Native app
Native application
Device APIs
Web app
Mobile browser
Web code
Hybrid app
Native container
Web code
Device APIs

User


iLEAPS
Emergency will be reported in...
5
Touch to cancel
Emergency
Violent Incidents
Domestic Violence
Fighting
Harassment or Bullying
Homicide (Murder)
Robbery
Sexual Assault
Non-Violent Incidents
Police Escort
History
Log Out
iLEA
"Emergency"
iLEAPS
Describe the Drug Incident
Please provide any relevant information
I would rather be anonymous
Send
iLEAPS
You: An accident just happened on E Coast Spring Ln
29-Jul-2015 | 12:18:30
Send

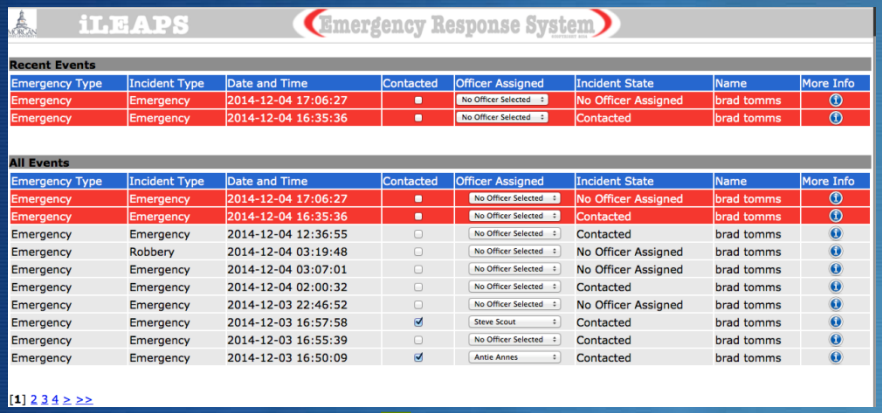
Officer

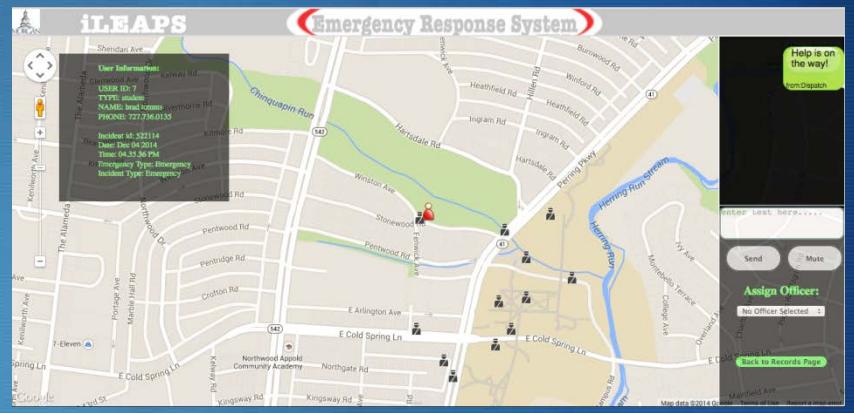

iLEAPS
ASSIGNED INCIDENT
Emergency: Emergency
0.09 mile from you | 4 hours ago
Emergency: Emergency
0.09 mile from you | 5 hours ago
ALL INCIDENT
Emergency: Emergency
0.61 mile from you | 19 mins ago
Emergency: Emergency
0.6 mile from you | 21 mins ago
Emergency: Emergency
0.1 mile from you | 4 hours ago
Emergency: Emergency
0.09 mile from you | 4 hours ago
Emergency: Emergency
0.09 mile from you | 4 hours ago
Emergency: Emergency
0.09 mile from you | 4 hours ago
Emergency: Emergency
0.09 mile from you | 4 hours ago

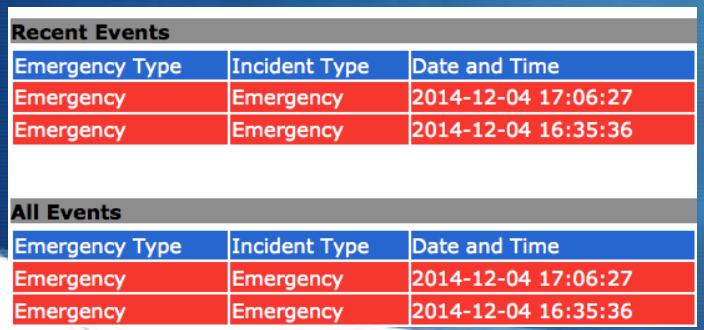

iLEAPS
INCIDENT VIEW
Emergency: Emergency
0.1 mile from you | 5 hours ago
Date: Jul 29 2015
Time: 11.11.37 AM
Map
Satellite
Clarence M. Mitchell, Jr. School of Engineering
William Donald Schaefer Engineering Bldg
Herring Run
Google
Map Data
Terms of Use
ACCEPT

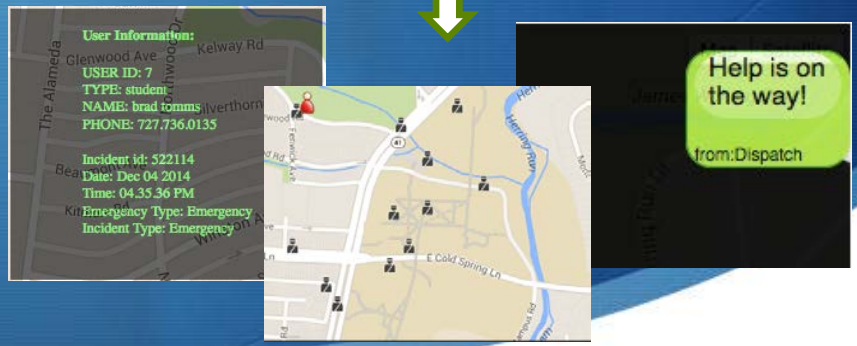

iLEAPS
PR: An accident just happened on E Coast Spring Ln
29-Jul-2015 | 15:18:30
29-Jul-2015 | 12:19:41
you: I can't make it dispatch
29-Jul-2015 | 15:39:09
Enter Message
Send

Dispatch


iLEAPS
Emergency Response System
Recent Events
Emergency Type | Incident Type | Date and Time | Contacted | Officer Assigned | Incident State | Name | More Info
Emergency | Emergency | 2014-12-04 17:06:27 | No Officer Selected | No Officer Assigned | brad tomms
Emergency | Emergency | 2014-12-04 16:35:36 | No Officer Selected | Contacted | brad tomms
All Events
Emergency | Emergency | 2014-12-04 17:06:27 | No Officer Selected | No Officer Assigned | brad tomms
Emergency | Emergency | 2014-12-04 16:35:36 | No Officer Selected | Contacted | brad tomms
Emergency | Emergency | 2014-12-04 12:36:55 | No Officer Selected | Contacted | brad tomms
Emergency | Robbery | 2014-12-04 03:19:48 | No Officer Selected | No Officer Assigned | brad tomms
Emergency | Emergency | 2014-12-04 03:07:01 | No Officer Selected | No Officer Assigned | brad tomms
Emergency | Emergency | 2014-12-04 02:00:32 | No Officer Selected | Contacted | brad tomms
Emergency | Emergency | 2014-12-03 22:46:52 | No Officer Selected | No Officer Assigned | brad tomms
Emergency | Emergency | 2014-12-03 16:57:58 | Steve Scout | Contacted | brad tomms
Emergency | Emergency | 2014-12-03 16:55:39 | No Officer Selected | Contacted | brad tomms
Emergency | Emergency | 2014-12-03 16:50:09 | Antie Annes | Contacted | brad tomms
[1] 2 3 4 > >>


iLEAPS
Emergency Response System
User Information:
USER ID: 7
TYPE: student
NAME: brad tomms
PHONE: 727.736.0135
Incident id: 522114
Date: Dec 04 2014
Time: 04.35.36 PM
Emergency Type: Emergency
Incident Type: Emergency
Help is on the way!
from:Dispatch
Assign Officer:
Back to Records Page

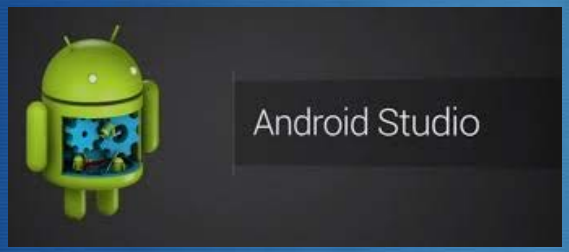

Recent Events
Emergency Type | Incident Type | Date and Time
Emergency | Emergency | 2014-12-04 17:06:27
Emergency | Emergency | 2014-12-04 16:35:36
All Events
Emergency Type | Incident Type | Date and Time
Emergency | Emergency | 2014-12-04 17:06:27
Emergency | Emergency | 2014-12-04 16:35:36


User Information:
USER ID: 7
TYPE: student
NAME: brad tomms
PHONE: 727.736.0135
Incident id: 522114
Date: Dec 04 2014
Time: 04.35.36 PM
Emergency Type: Emergency
Incident Type: Emergency
Help is on the way!
from:Dispatch


Google Play services


AJAX
Asynchronous Javascript And XML


Google cloud messaging


Android Studio


HTML
JS
CSS


jQuery
write less, do more.


php
MySQL

- **Officer and User apps**
  - Android push notifications with Google Cloud Messaging
  - Geo-tagging
  - Taking pictures of incidents

- **Android push notifications with Google Cloud Messaging**
  - Send data from your server to your users' devices, and receive messages from devices on the same connection
  - The GCM service handles all aspects of queueing of messages and delivery to client applications running on target devices
  - Distribute messages to single devices, to groups of devices, or to devices subscribed to topics.
  - Battery-efficient connection and free

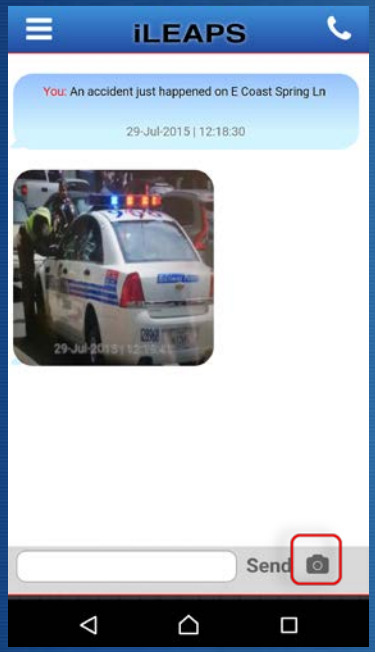
- **Android push notifications with Google Cloud Messaging. Example:**

```
https://gcm-http.googleapis.com/gcm/send
Content-Type:application/json
Authorization:key=AIzaSyZ-1u...0GBYzPu7Udno5aA
{
  "to": "/topics/foo-bar",
  "data": {
    "message": "This is a GCM Topic Message!",
   }
}
```

```
@Override
public void onMessageReceived(String from, Bundle data) {
   String message = data.getString("message");
   Log.d(TAG, "From: " + from);
   Log.d(TAG, "Message: " + message);
   // Handle received message here.
}
```

- **Geo-tagging**
  - Google Location Services API, part of Google Play Services
  - Provides a more powerful, high-level framework that automatically handles location providers, user movement, and location accuracy.
  - It also handles location update scheduling based on power consumption parameters you provide.
  - In most cases, you'll get better battery performance, as well as more appropriate accuracy, by using the Location Services API.

- **Taking pictures of incidents**



Handles the event

Calls a Android Method

Invokes the Activity for taking photos



Upload photo

Store photo on server

GCM

Send a GCM message

- Core functionality on all three interfaces almost complete
- The team is also testing and fixing issues
- Future work:
  - to perform predictive analytics on forensic data (discovery of trends and patterns in collected data)
  - expand to other organizations

Questions ?

Geographic Information Retrieval workshop - GIR'14

Construction and First Analysis of a Corpus for the Evaluation and Training of Microblog/Twitter Geoparsers

JAN OLIVER WALLGRÜN, FRANK HARDISTY, ALAN M. MACEACHREN, **MORTEZA KARIMZADEH (PRESENTER)**, YITING JU, SCOTT PEZANOWSKI

Manually annotated Corpus

Goal: Building a gold standard corpus of microblog posts with manually annotated place references.

- Evaluation, comparison, training of geoparsers.
- Cognitive studies. What is a place name? Analysis of disagreement between lay persons' annotations.

Microblog posts (Tweets)

- Suitable for monitoring social activities, elections and society's sentiment, situational awareness in crisis times, etc.
- ~%1-2 geotagged.
- ~%10 contain references to places.

Corpus Building Framework

- Crowdsourcing
- GeoVisual Analytics

Corpus Status:

- 6000 tweets
- %30 with at least one place name, annotated with GeoNames ID and Geographic Coordinates.

General Corpus Building Framework

"Where are the best #steakhouses in New York City? Find out: http://t.co/KJLKJL #nyc"

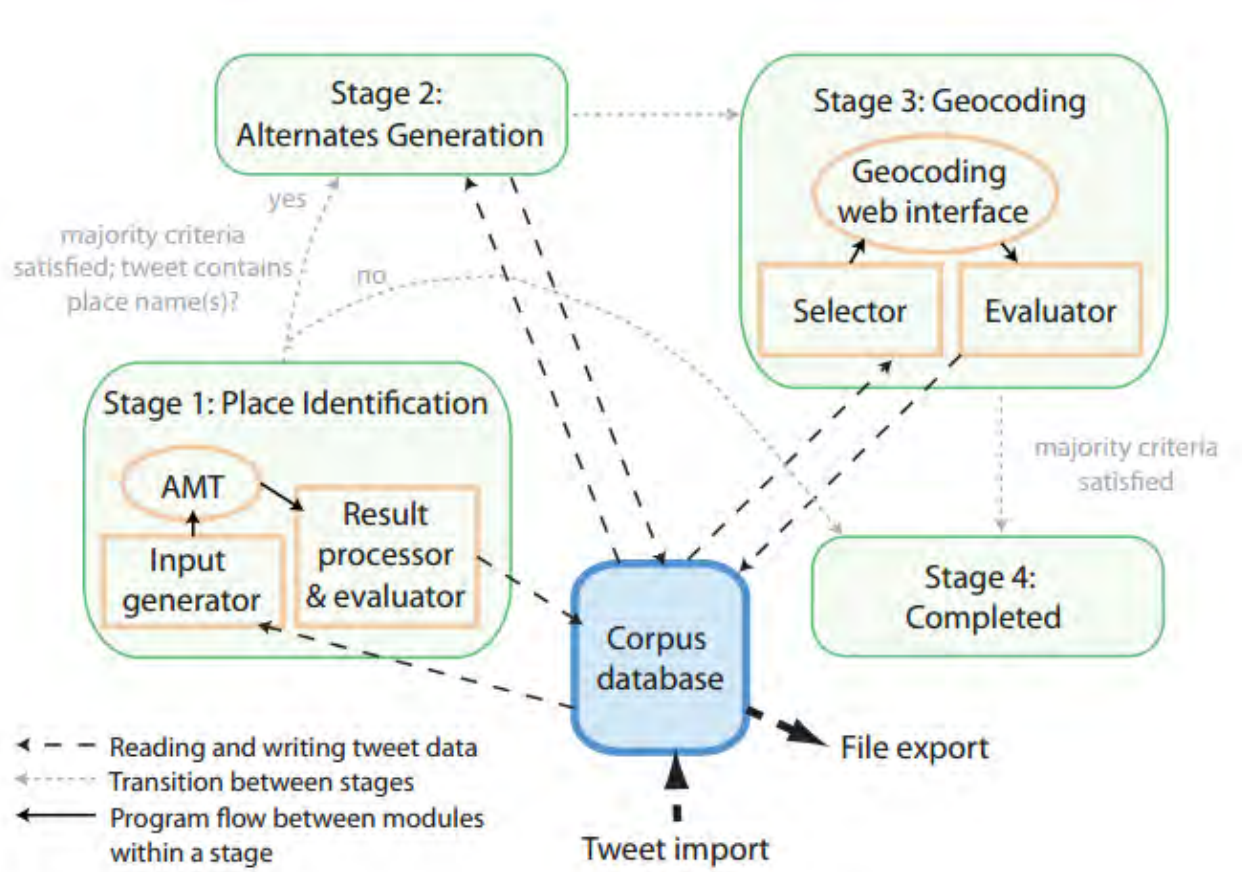
Each occurrence would be annotated with the information:

(1) Toponym: New York City;

(2) Geonames ID: 5128581;

(3) Latitude: 40.71427;

(4) Longitude: -74.00597.



Tweet Collection and Sampling

Twitter Streaming API

- ~1 billion Tweets collected, from Jan 2014 to Sept 2014.
- Keywords based on current (crisis related) events.
- Random Sample of 6000 tweets (500 for each keyword).
  - dengue, malaria, earthquake, measles, fire, protest, flood, rebels, flu, riot, gun, tornado.

More place-focused tweets that talk about

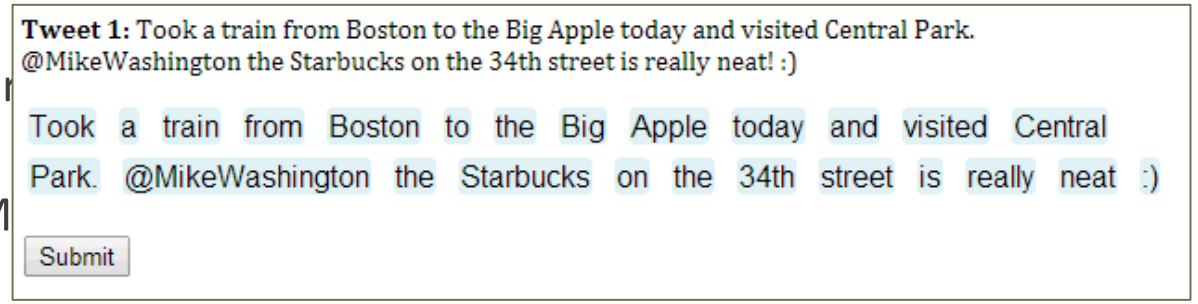
- Natural Disasters.
- Infections disease.
- Human threats/violence.

- Filter for “the”.
  - The most common word in the English language.

Crowdsourcing Place Identification

- Amazon Mechanical Turk (AMT)
  - Distributing small tasks: Human Intelligence Tasks (HITs).
  - Recruitment and payment handled by AMT.

- Web Interface for AMT workers to identify place names
  - Detailed instructions for workers.
  - Only allowing for workers with a HIT acceptance rate of 97% or higher from pr
  - 2 tweets per HIT
  - Each Tweet tagged by at least 5 different AM
  - 4 cents per assignment.
  - Overall costs of $600 to collect 5 AMT results for each of the 6000 tweets.



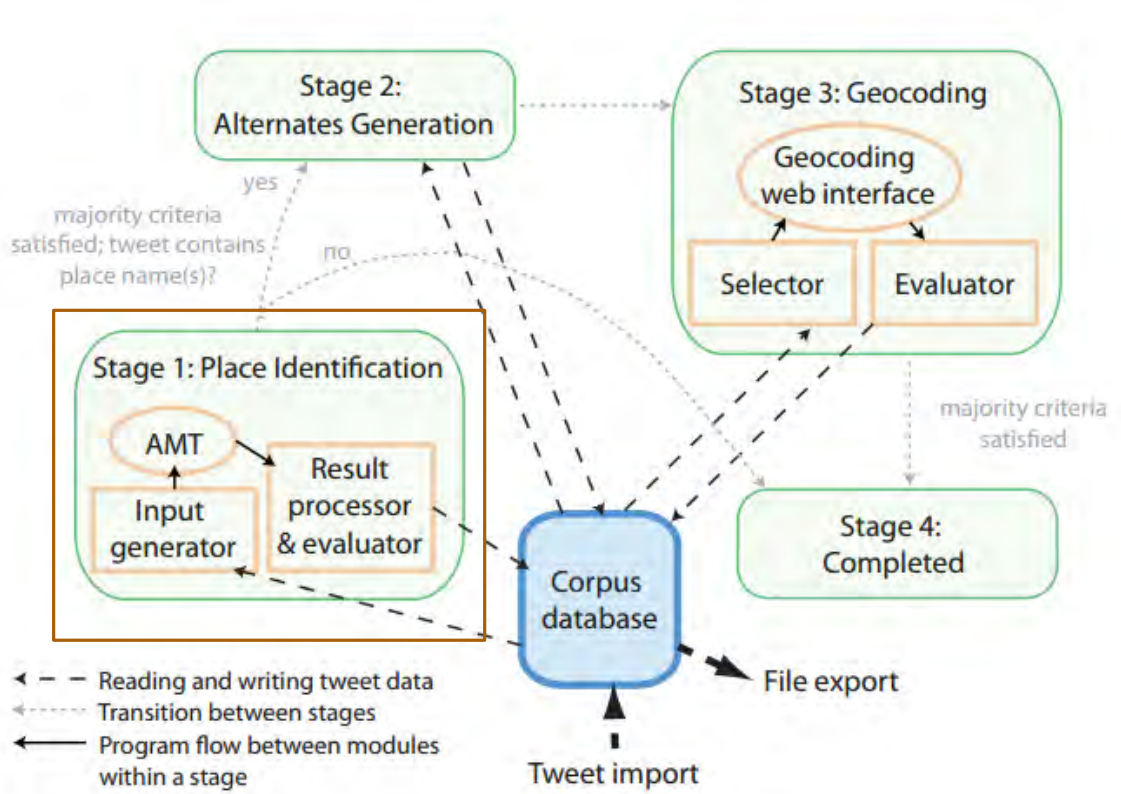
Place Identification Instructions

- Annotate as a place reference:
  1. Any named town, city, county, state, or country (e.g., Los Angeles, Jefferson County, NY, Italy)
  2. Named buildings (e.g., Eiffel Tower, Dodgers Stadium, Alcatraz, James J. Ferris High School);
  3. Named areas (e.g., Grand Canyon, Pacific Ocean, Washington Mall, Hyde Park);
  4. Street and highway names (Atherton Street, 1st Ave, Highway 1, I-70);
  5. Place names inside hashtags (#newyork).

- Do NOT annotate as a place reference:
  1. Businesses (Starbucks, Microsoft, Texas Steak House, Baltimore Ravens)
  2. Organizations (Lutheran Church, Red Cross, United Nations, Grand Canyon Historical Society);
  3. Place names used as descriptors (U.S. dollar, Philadelphia cheesesteak).

Agreement Criteria

1. ≥ 5 results from different workers exist for this tweet.
2. For each word in the tweet, there is at least a 70% majority agreement on whether that word is part of a place name and if so together with which other words.

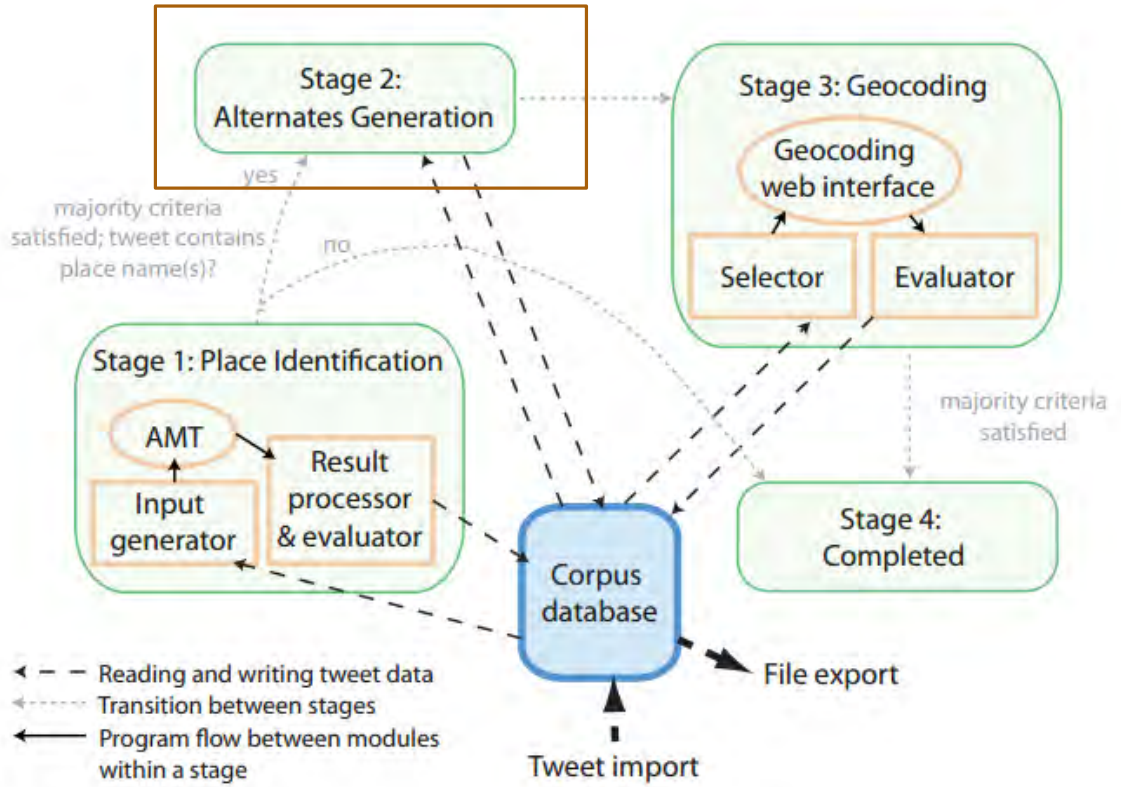


M. Potthast. Crowdsourcing a Wikipedia vandalism corpus. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 789–790, 2010

Alternate Generation

Generate the alternates using relevance ranking and geocoding component from our own geoparser, GeoTxt*.

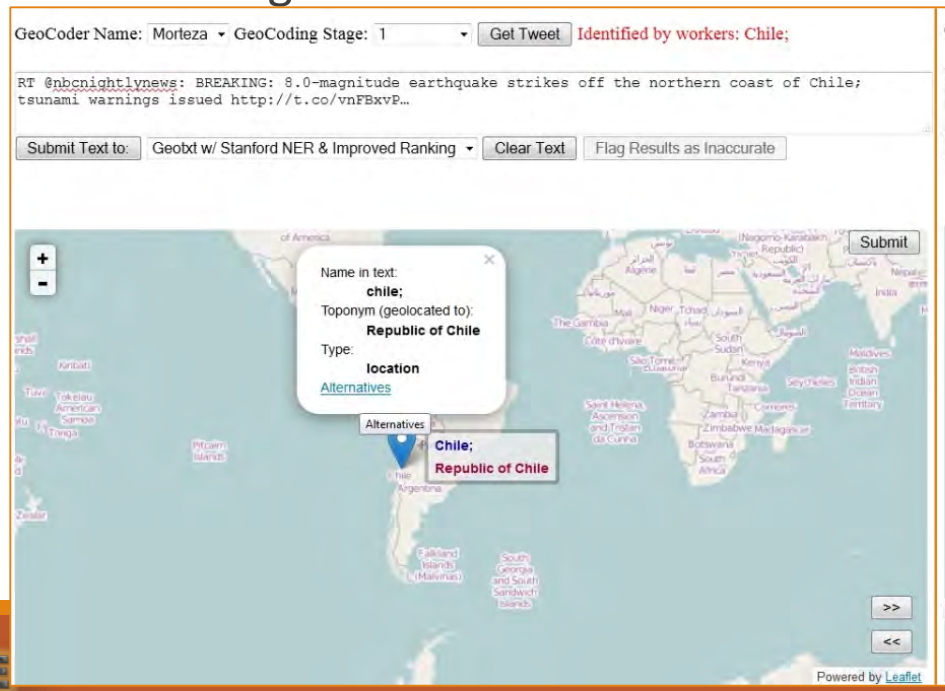
- Name similarity
- Population
- Geographic prominence

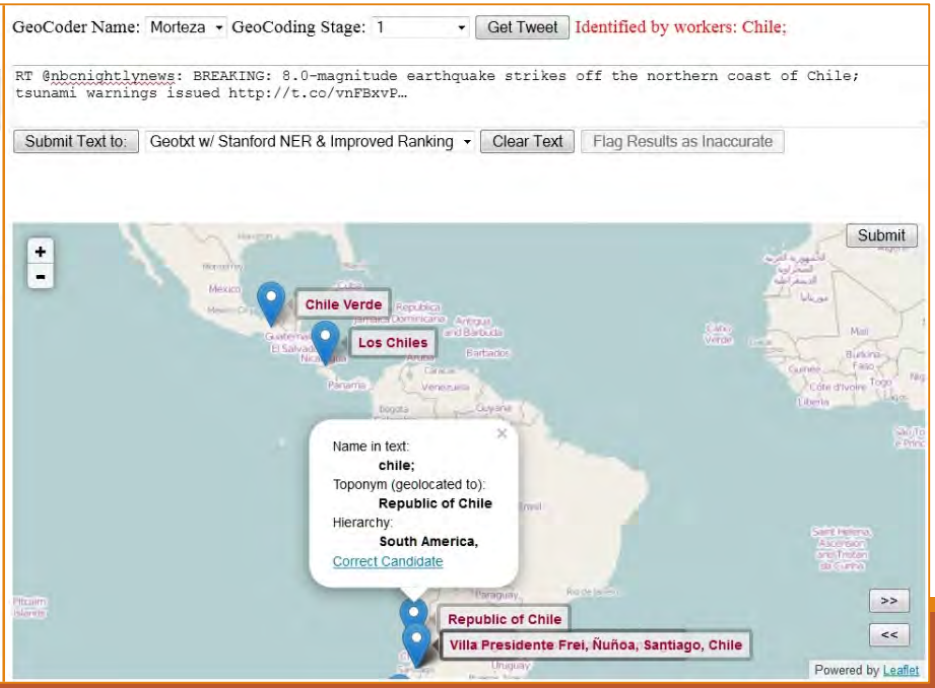


* Karimzadeh M, Huang W, Banerjee S, et al. (2013) GeoTxt: a web API to leverage place references in text. Proceedings of the 7th Workshop on Geographic Information Retrieval. Orlando, Florida: ACM, 72-73.

Manual Geocoding

- “Expert” Geocoding: Graduate Students and faculty members in Geography.
  1. Only tweets with place names (%30).
  2. Either having local knowledge or general knowledge and willingness to research.





Corpus Analysis

- Of the event-based sample of tweets making up the corpus
  - 70.0% contain no place names;
  - 22.3% contain one place name;
  - 5.7% contain two place names;
  - 1.3% more than two place names.

- Reached agreement for 91.6% of the 6000 tweets in the corpus.

- Disagreement analysis useful for
  - Place conceptualization.
  - Improving corpus building and NER.

Disagreement Analysis

- Resolution of disagreements:
  1. Manual resolution by experts that decide what are correct submissions.
  2. Automatic resolution via algorithms that can correct or discard incorrect submissions.
  3. Introducing new rules to instruct workers more clearly and collect more results via AMT.

- Disagreement Categories:
  1. **1% - Tweeter errors (TWE)** → Manual resolution, new instructions, auto-correction.
     - *"Its been 20 years since the north ridge earthquake? Wow time flies"*
  2. **Worker errors**
     - workers make mistakes that clearly were against the instructions given to them.
  3. **Place Definition**
     - What constitutes a place name? Reasonable people may disagree.
     - The most common of the sources of disagreement.

Worker Errors

- 32 %- Familiarity (FAM) –> Manual resolution.
  - *"... help us find missing white large dog Leechan http://t.co/pxZY5wmpov Sendai #Japan. one of ..."*
- 5% - Misidentified Tokens (MTN) –> Manual resolution.
  - *"@lakeline @HockeenightsCT great. Now you've done it CT. Youve ..."*
- 20% - Non-merging Errors (NME) → Automatic resolution.
  - *"... Earthquake Shakes Mexicos Pacific Coast: A strong earthquake has shaken the southern Pacific... "*
- 6% - Skipped repeats (REP) → Automatic resolution.
  - *"... Text: No measles outbreak in Cavite TRECE MARTIRES CITY, Cavite There is still no ..."*

Place Definition

- 23% - Adjectival Usage (ADJ) → Improve instructions, manual resolution.
  - *"S'Sudan rebels reinforce captured city defences: South Sudan's rebels are strengthening ..."*
- 16% - Organization with a place in name (ORG) → Improve instructions, manual resolution.
  - *"#Patriots How the* Indianapolis *Colts Defense Can Keep Tom Brady in Check http://t.co/TxBtIhHlSS"*
- 11% - Place Hierarchy Related Issue (PHR) → Improve Instructions, automatic resolution.
  - *RT @NewEarthquake: 5.6 earthquake, 297km ENE of Grytviken, South Georgia and the ..."*
- 5% - Non-proper name (NPN) → Improve instructions.
  - *"... Flash flood watch has now been trimmed back to just include the south shore through 9am."*
- 10% - Vaguely qualified place name (VQP) → Improve instructions, potentially special annotation.
  - *"... wildfires in south #California: At least two structures burned to the ground an..."*
  - *"... Theres an outbreak of measles in the Bronx and upper Manhattan, officials warn: ..."*
- 6% - Inclusion or exclusion of kind of place (KND) → Improve instructions, mandatory for physical features, flexibility in comparison in the benchmarking framework, automatic resolution.
  - *"... How Adelia, 9, and the island of Sabang, #Indonesia, triumphed over #malaria ..."*
- 16% - Hashtag Related Issues (HAS) → Improve the annotation platform, improve instructions, automatic resolution.
  - *"... everyone please pray for the people there. #prayforindonesia"*

Outlook

- Find appropriate licensing/terms for corpus release.
- Implementing our three strategies to deal with disagreements in annotations.
- Use the same framework for other kinds of text.
- Evaluate and improve our own Geoparsing service.

- Summary of disagreem

TWE	FAM	MTN	NME	REP	ADJ
1%	32%	5%	20%	6%	23%
ORG	PHR	NPN	VQP	KND	HAS
16%	11%	5%	10%	6%	16%

Suggestions? ☺

?

Karimzadeh@psu.edu

An Intelligent Crowdsourcing System for Forensic Analysis of Surveillance Video

Khalid Tahboub, Neeraj J. Gadgil, Javier Ribera, Blanca Delgado and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana USA

Outline

- **Motivation**
- **System Requirements**
- **Proposed Approach**
- **Experimental Results**
- **Conclusion and Future Work**

Motivation

- **Hundreds of search and rescue missions are carried out in North America and many more worldwide**
- **Search and rescue missions and "after-the-event" investigations require careful analysis of videos of tens or hundreds of hours of duration**
- **Law enforcement authorities have used some forms of crowdsourcing to reach out to volunteers**

Motivation

- **Crowdsourcing is often considered as an effective solution to problems that involve cognitive tasks**
- **Use the collective intelligence of crowds to do tasks that machines find very difficult**
- **Goal: Build a system to enhance crowdsourcing by incorporating machine learning methods and distinguish the ability of crowd members**

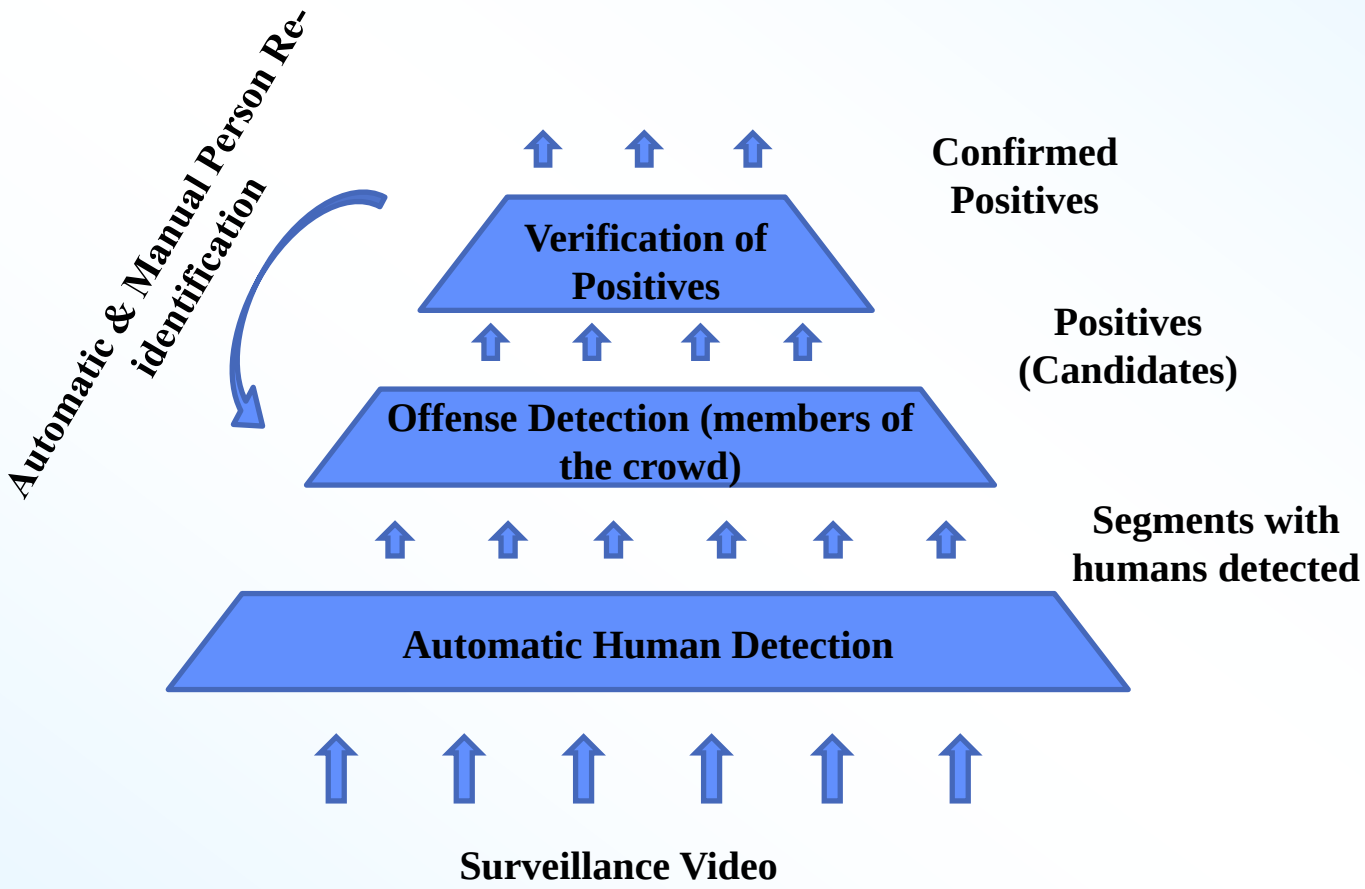
System Requirements

- **Web-based system**
- **Enable training and teaching**
- **Upload a set of videos as one investigation task**
- **Accelerate the process using machine learning methods**
- **No intervention required – autonomous from the perspective of law enforcement**
- **Produce the final result in the form of a set of video segments specifying the events of interest as one storyline**

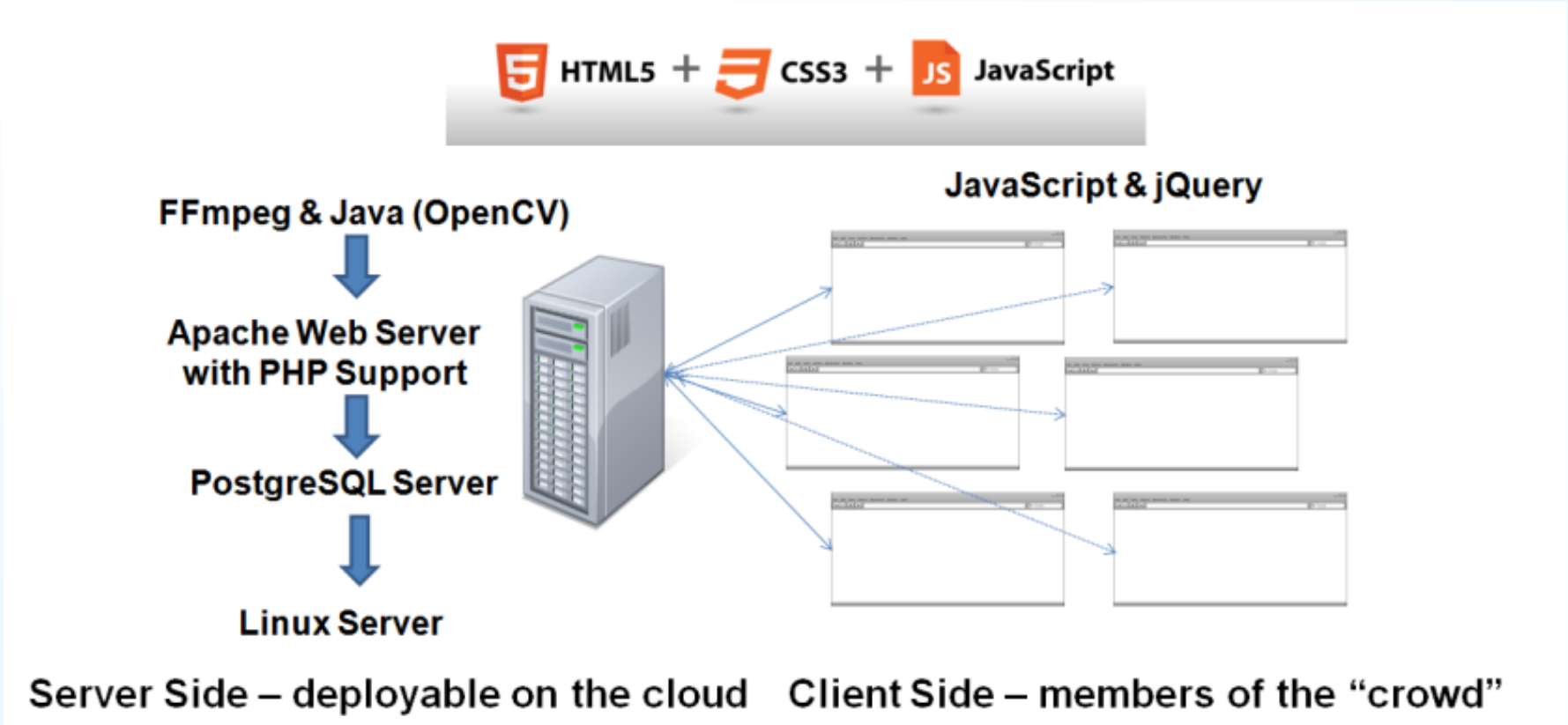
Proposed Approach

- **Use a hierarchal pyramid model**
- **Incorporate human detection and re-identification**
- **Enhance the annotation platform using object tracking**

Proposed Approach

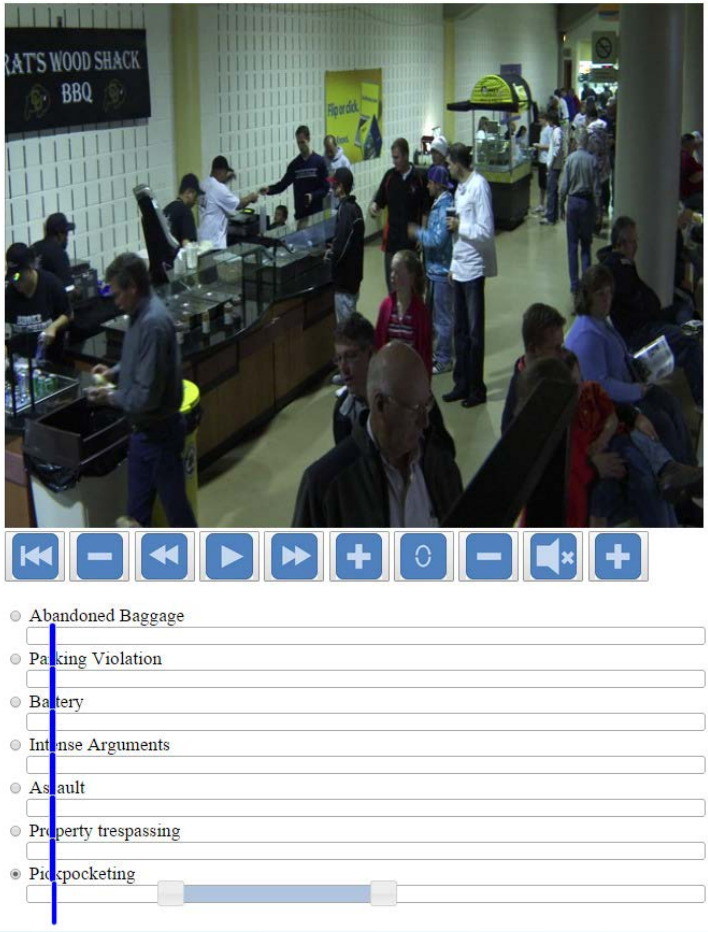


Proposed Approach



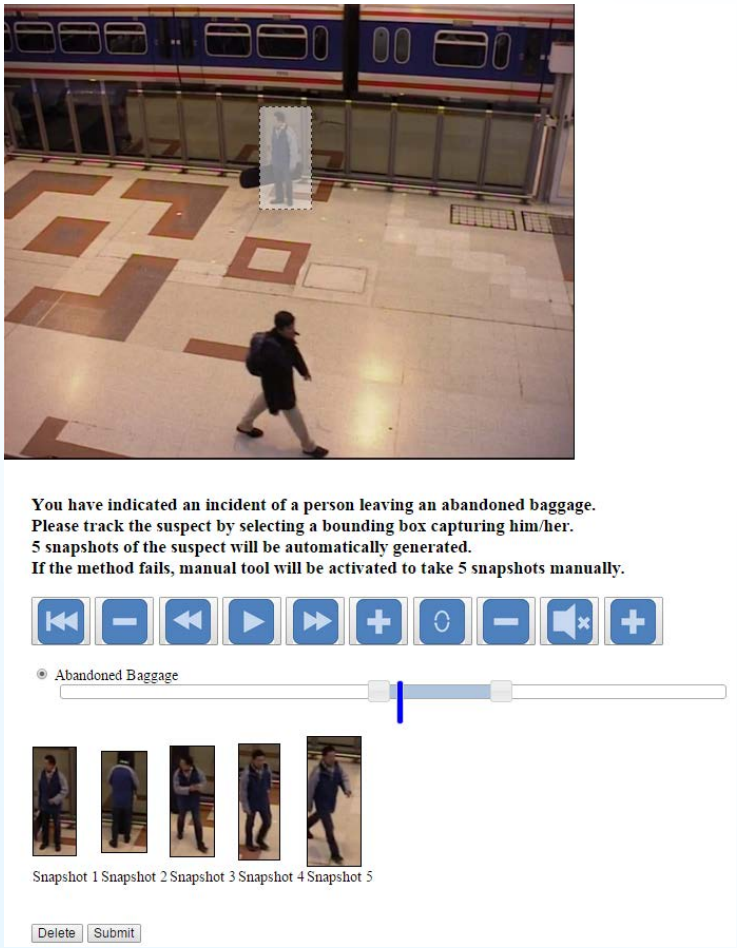
http://lorenz.ecn.purdue.edu/~vat/

Proposed Approach: Enhanced Annotation Platform



- **Standard video player functionalities**
- **Implemented using JavaScript and HTML5**
- **User-friendly interface**
- **Temporal and spatial annotations**

Proposed Approach: Enhanced Annotation Platform



- **Corners are detected according to the features from accelerated segment test (FAST)**
- **Corner point description and matching are based on binary robust independent elementary features (BRIEF)**

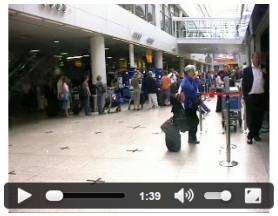
Proposed Approach: Training Management

- Add a label with training modules
- Assign training to workers
- Assess training performance (Pass/Fail)

Gold Answer

Performance Assessment

Training Video Information



Label Information

Name: Suspicious Bag Activity
Description: Abandoned or swap
Status: Ready
Is passing required? Yes

Level: Medium
Passing criterion: Same number of annotations

Neeraj Gadgil	Admin Task	Date Assigned: 2014-01-02	Time Assigned: 02:40:13	Completed.	Date Completed: 2014-01-02	Time Completed: 02:41:46
Dahjung Chung	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:50:53	Assigned, not started.	Date Completed: 2001-01-01	Time Completed: 00:00:00
He Li	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:50:37	Pass	Date Completed: 2014-01-02	Time Completed: 12:57:35
Joonsoo Kim	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:50:48	Pass	Date Completed: 2014-01-02	Time Completed: 17:28:46
Soonam Lee	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:50:42	Pass	Date Completed: 2014-01-02	Time Completed: 15:23:27
Albert Parra Pozo	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:50:29	Fail	Date Completed: 2014-01-02	Time Completed: 12:56:38
Khalid Tahboub	Turker Training Task	Date Assigned: 2014-01-02	Time Assigned: 02:51:00	Fail	Date Completed: 2014-01-02	Time Completed: 03:14:12

Assign Training to turkers

Proposed Approach: Human Detection

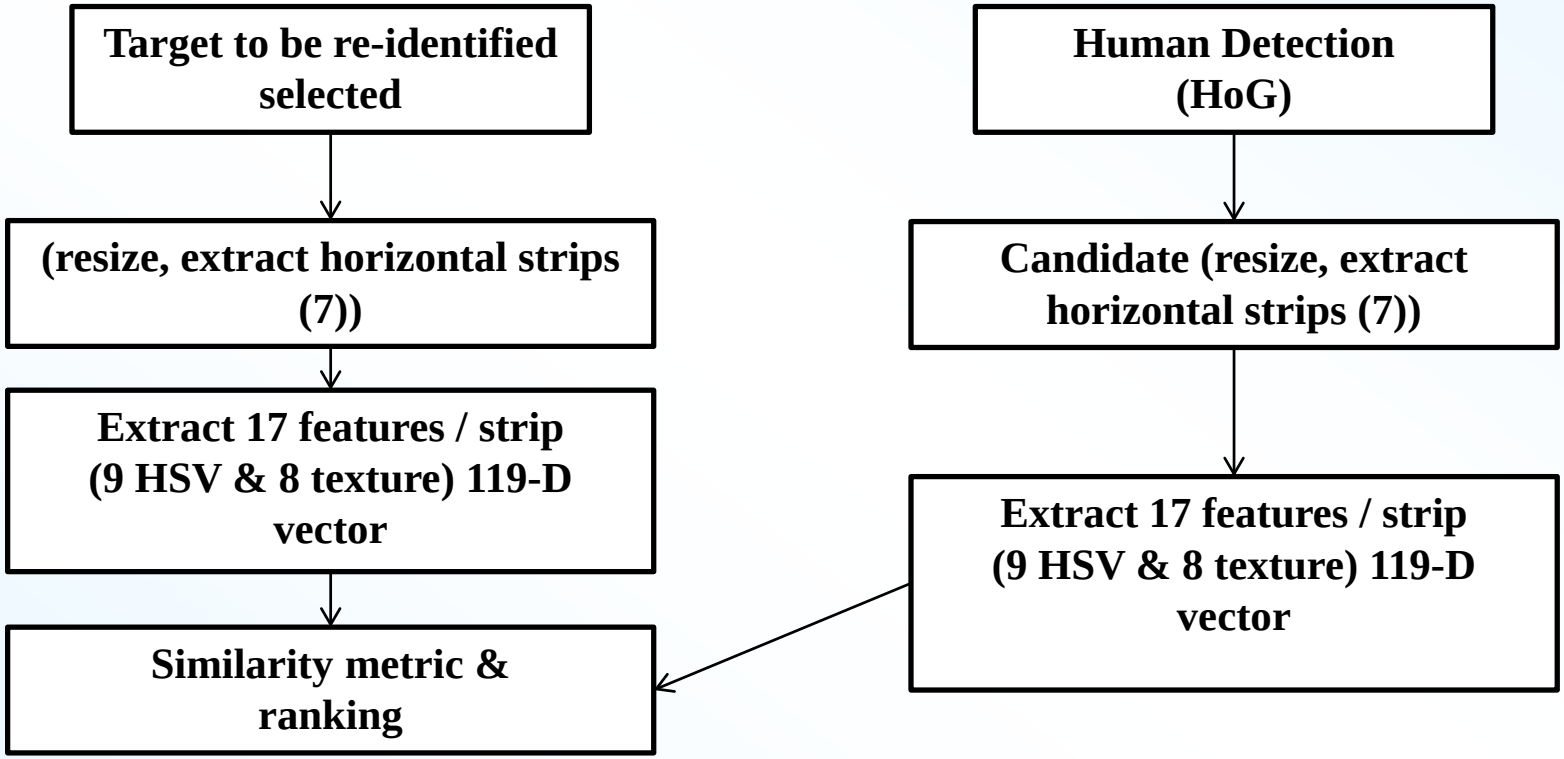
- **An OpenCV implementation of the histograms of oriented gradients (HOG)**
- **Linear support vector machine (SVM) classifier**
- **The classifier is trained with the INRIA person dataset**

Proposed Approach: Person Re-identification

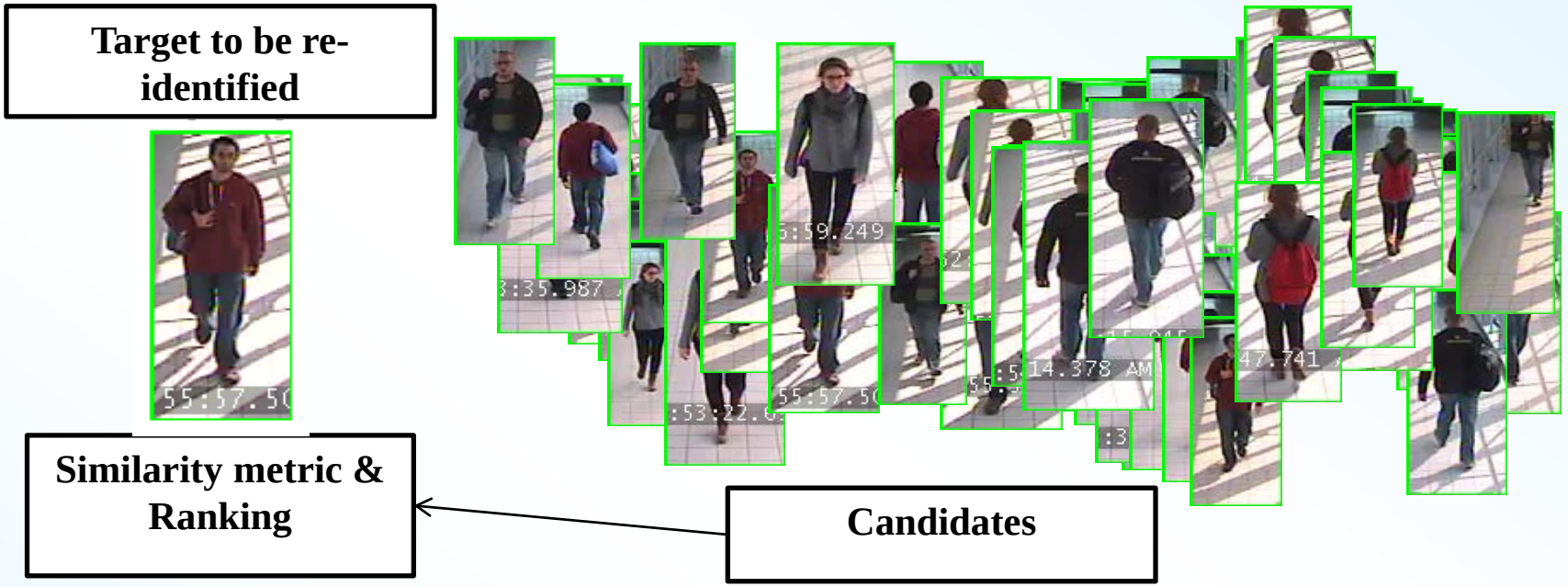
- **Aims to find all the reappearances of a confirmed suspect in the surveillance video**

- **Based on color and texture features**

Proposed Approach: Person Re-identification



Proposed Approach: Person Re-identification



Experimental Results

- **Used public datasets: BEHAVE, Lunds University traffic, PETS 2006, PETS 2007 series, i-Lids-AVSS 2007 and videos from public safety communication research (PSCR) laboratory**
- **Performance of trained versus non-trained members of the crowd**
- **Proposed approach performance under a staged investigation**

Experimental Results: Training

- **Two labels for training**

(Suspicious bag activity, Traffic violation)

- **For trained crowd, tasks were only assigned after successful completion of training modules**
- **Total 8 workers participated**

Event type	Trained group			Untrained group		
	No. of Annotators	Precision	Recall	No. of Annotators	Precision	Recall
Suspicious bag activity	4	0.89	1	4	1	0.5
Traffic violation	3	0.5	0.67	4	0.2	0.25

- **Increase in the number of true event detection**

Experimental Results: Staged Investigation

- **Video content includes crowded scenes in stadiums, hallways, parking lots, stores and building entrances**

Number of surveillance videos	113
Total duration	115 minutes
Offense type	Abandoned baggage
Number of videos showing offense	1
Number of suspect reappearances	3

Experimental Results: Staged Investigation

- **Human detection:**

Number of true positives	83
Number of false positives	13
Number of true negatives	15
Number of false negatives	2

- **Person Re-id:**

Number of true positives	1
Number of false positives	5
Number of true negatives	89
Number of false negatives	1

Experimental Results: Staged Investigation

- **Manual Person Re-id:**

Number of true positives	3
Number of false positives	2
Number of true negatives	91
Number of false negatives	0

- **All false negatives were rejected by experienced crowd members**
- **Final output is the correct identification of the offense and the three reappearances**
- **Total execution time is 17 minutes**

Conclusion and Future work

- **Proposed a crowdsourcing system for forensic analysis of surveillance video**
- **Used a hierarchal pyramid model to distinguish the crowd members based on their ability, experience and performance record**
- **Enhanced crowdsourcing by incorporating human detection, re-identification and tracking methods**
- **In future, we plan to incorporate more machine learning methods with our model**

An Intelligent Crowdsourcing System for Forensic Analysis of Surveillance Video

Khalid Tahboub, Neeraj J. Gadgil, Javier Ribera, Blanca Delgado and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana USA

Automatic Detection of Abnormal Human Events on Train Platforms

Blanca Delgado[1] , Khalid Tahboub[2] and Edward J. Delp[2]

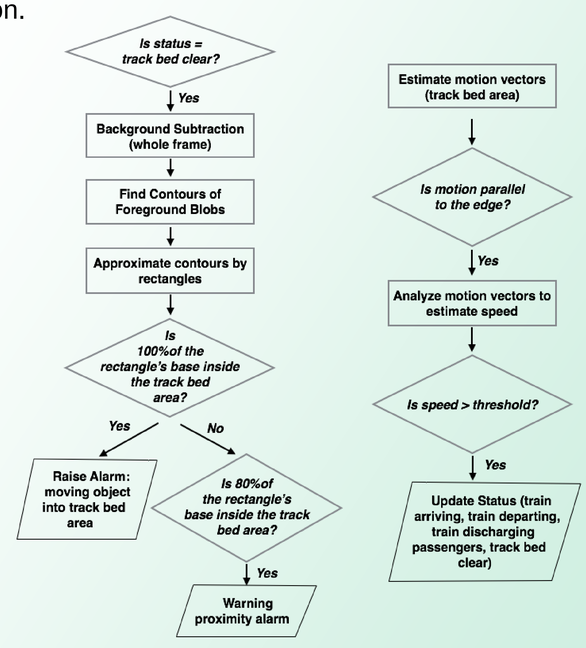
[2] **Universitat Politècnica de Catalunya, Barcelona, Spain**
[1] **Purdue University, West Lafayette, IN, U.S.A.**

Motivation

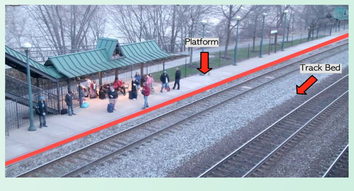
- Video surveillance systems that contain a large number of cameras makes the continuous monitoring of the video feeds nearly impossible task.
- Many video analytic techniques have been developed to provide automatic analysis of the video data.
- Many cities witnessed fatal incidents where people were hit and killed by trains.
- Automatic detection of abnormal human events on train platforms can save lives.

Proposed Method

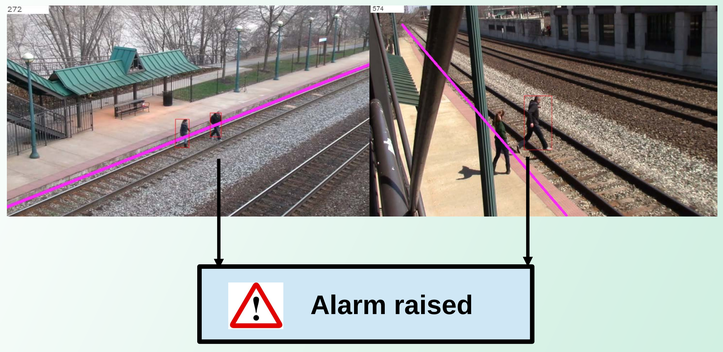
- We detect moving objects and analyze their location in relative to the edge of the track bed. Train presence is investigated by analyzing motion vectors in the track bed region.



- A line separating the track bed and the platform is determined manually.



- Moving object detection is performed using background subtraction.
- Every pixel of the background is modeled as a random variable. The mean or average represents the actual background and its variation with respect to the mean is due to the noise introduced by the camera sensor and other sources.
- To account for variations introduced by dynamic environmental conditions, a Gaussian Mixture Model is used as a statistical model.
- Morphological operations are used to remove noise and unify the blobs of the foreground mask.
- For each blob, the contours are estimated and approximated by a rectangle.
- When the base of the rectangular contour (assumed to be the feet of a person) completely crosses the line marking the edge between the platform and the track bed, an alarm is raised indicating the detection of a person into the track bed.



Train Presence

- To detect the presence of the train, its motion information is used, taking into account that the motion of the train is parallel to the edge of the platform.
- To estimate motion vectors, we find the most prominent corners of each frame constrained to the track bed region.
- We use a point-based tracker to track the prominent corners which is based on Lucas Kanade Feature Tracker.
- Displacement vectors serve as an indicator of the train speed. When the train is moving at a high speed, the displacement distance is large.

Examples of prominent corners:



Results

- The experimental evaluation was conducted using our new dataset: videos have been recorded from three different views and cameras were placed at a height similar to typical surveillance cameras.



- The dataset has been divided in short clips containing the following events:
  - No events or trains
  - Person crossing the track bed
  - Train arriving
  - Train waiting or discharging passengers
  - Train departing
- Precision was found to be 90%, while recall was 100%.
- Train presence was correctly identified at all times.

Conclusion and Future Work

- An experimental evaluation was conducted using a video dataset that was recorded at a local train station, precision and recall were found to be 90% and 100% respectively.
- Train presence was correctly identified at all times.
- For future work, we want to investigate:
  - Multi-camera scenarios.
  - Study the impact of objects moving in the immediate vicinity to the camera.
  - Estimate the depth of moving objects using motion information.

Publications

B. Delgado, K. Tahboub and E. J. Delp, “Automatic Detection of Abnormal Human Events on Train Platforms”, To appear, Proceedings of the IEEE National Aerospace & Electronics Conference, Dayton, Ohio, June 2014.

Acknowledgements

This work is partially funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001.

Automatic Recognition and Interpretation of Gang Graffiti (GARI)

Edward J. Delp

Video and Image Processing Laboratory

School of Electrical and Computer Engineering

Purdue University

West Lafayette, Indiana USA

ace@ecn.purdue.edu

Acknowledgements

- **This work is funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001**
- **Indiana Gang Intelligence Network (INGangNetwork)**
- **Cook County Sheriff Department**

GARI System Goals

- System that allows first responders to:
  - Collect gang graffiti and gang tattoos images
  - Analyze and interpret the images
  - Image retrieval
  - Interact with a database of images
  - Browse database using an interactive map
  - SQL database
- Platforms
  - Android and iPhone application
  - Web-based interface
  - CGAP (Civilian Gang Application Program)

GARI System Overview



Original image

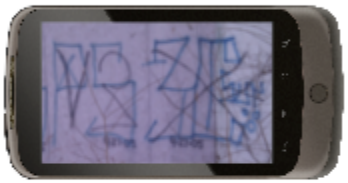
1

Offline automatic analysis and labeling
- Geoposition
- Date and time
- Extracted Features

3



Original Database



2

4

Filtered results
Info + thumbnails



Server



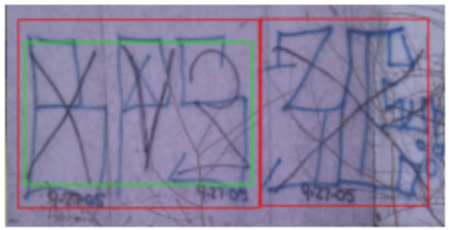
Filtered Database

5

Offline manual filtering

6

Manual labeling
- Additional Features



Labeled image

7

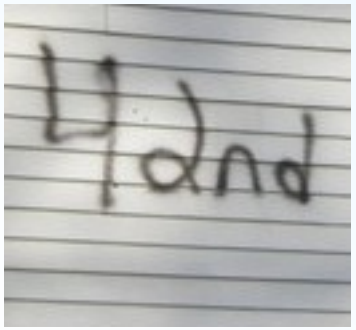
Addition to Database

Gang Graffiti Interpretation



SHAPE

Simple, Straightforward



NUMBERS

42nd street gang



SYMBOLS

6-point star, pitchforks



COLOR

Goon Squad: Red/Black

Gang Graffiti Interpretation



LETTERS

East Side Gang



TIME

Black: 18 ST (18th Street Gang)

Red: 13 SUR (Sureños 13)



POSITION/ALIGNMENT

Letters at star points

Numbers in the middle

Letters at the bottom

Pitchforks upright

Image Analysis – Color Recognition





MEXICANOS
MALDITOS SUREÑOS 13





18 STREET GANG



SUREÑOS 13

Image Analysis – Scene Analysis

GANG RIVALRY TRACKING



- **Date:** 08/19/2010
- **Time:** 3.25 PM
- **Geo:** 41.387917, 2.169919



- **Date:** 01/03/2011
- **Time:** 5.11 PM
- **Geo:** 41.387917, 2.169919

Gang Tattoos

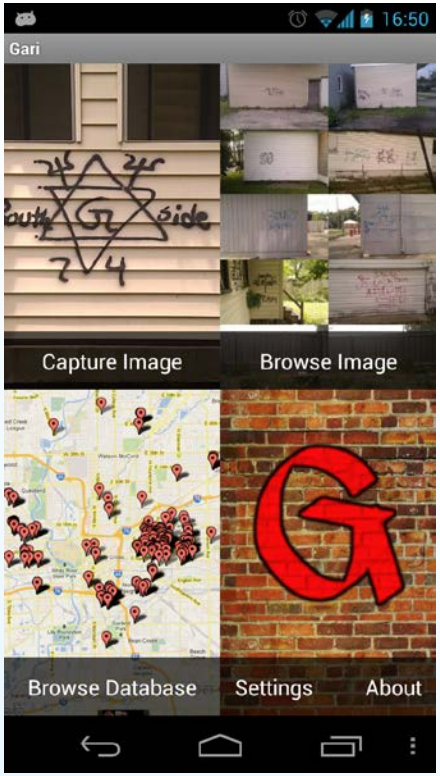




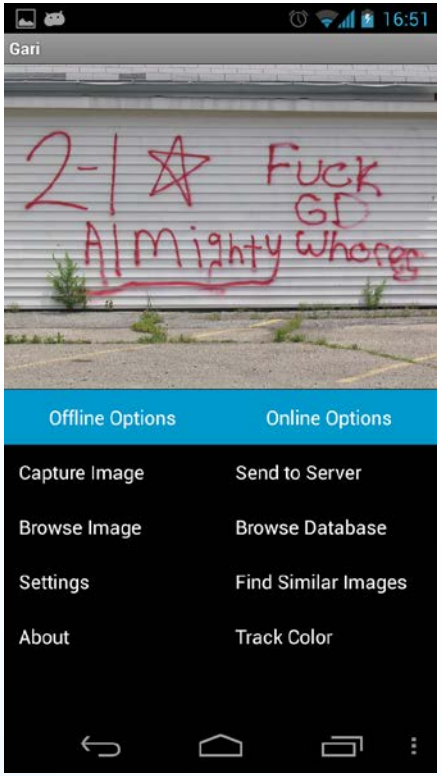




User Interface



Main menu



User options

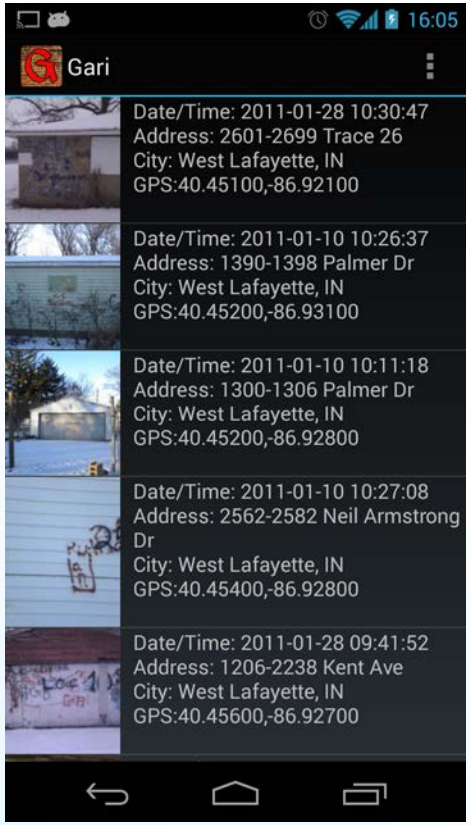


Analysis results

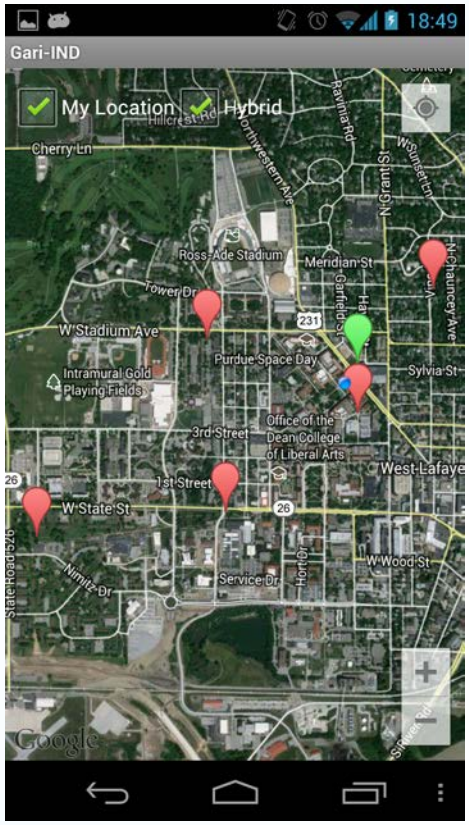


Image matching

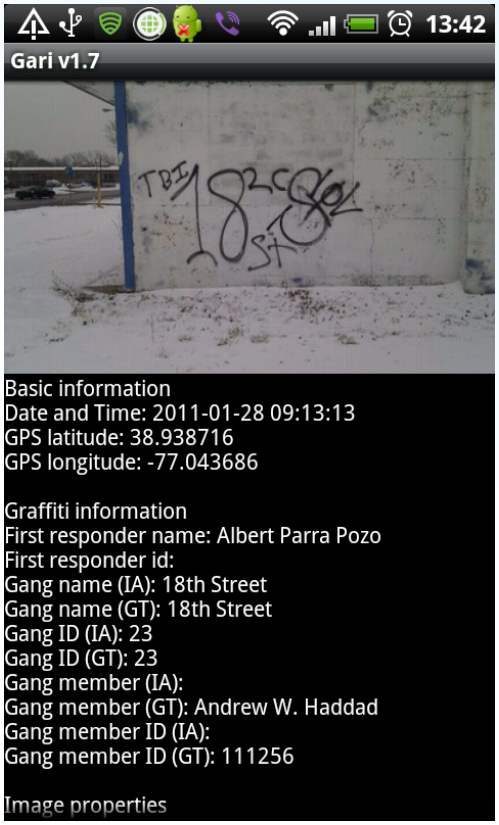
Browse Database



Browse general results



Show graffiti on map



Inspect specific graffiti

Augmented Reality Map



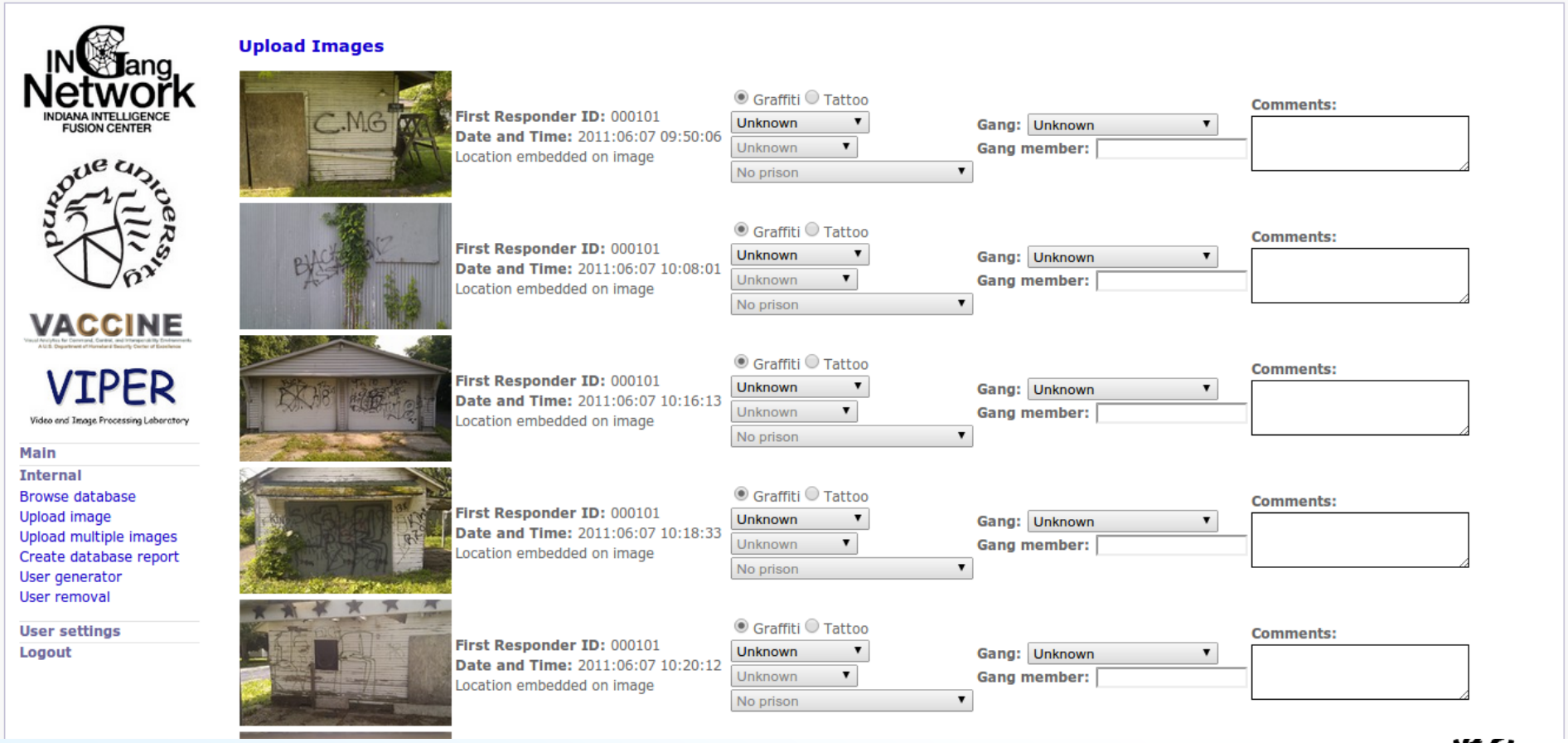






Multiple Image Upload

Desktop Version



Browse Database

Desktop Version

- **Browse by**
  - **Radius**
  - **Date**
  - **Address**
  - **Graffiti/Tattoo**



Database browsing screenshots



Browse Database

Desktop Version

VIPER

Video and Image Processing Laboratory

VACCINE

Main

Archive

Browse Database

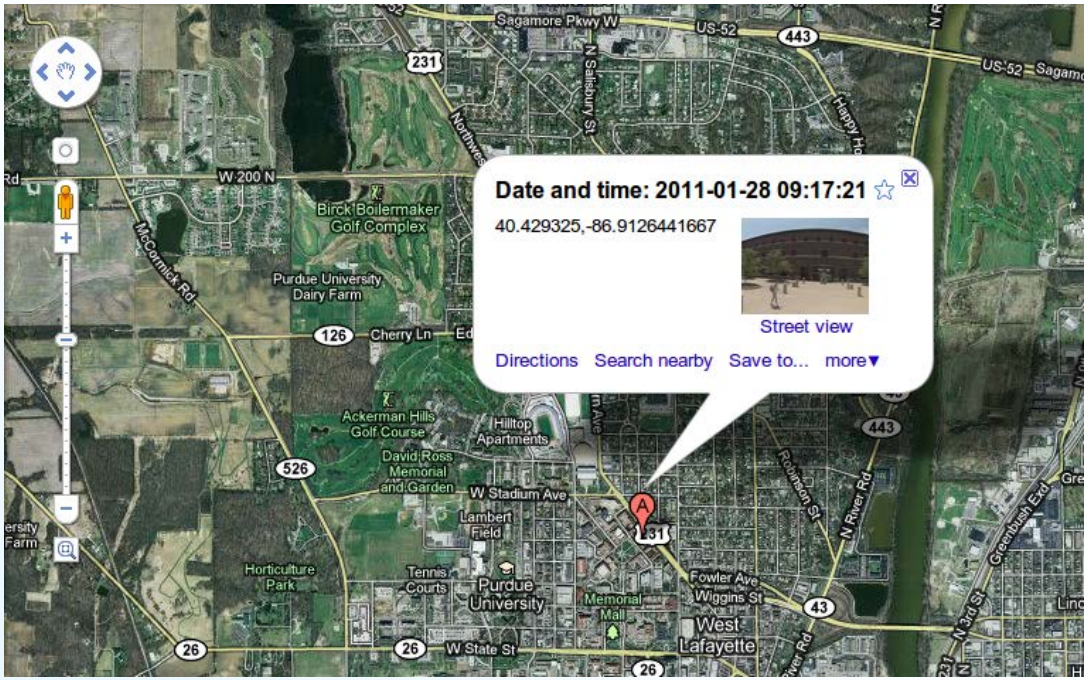


Date and Time: 2011-01-28 09:17:21
GPS latitude: 40.429325
GPS longitude: -86.9126441667
File Size: 664501 bytes
Height: 1552 px
Width: 2592 px
Focal length: 4.31 mm
Camera make: HTC
Camera model: HTC Desire
Image id: 1111
Show in map

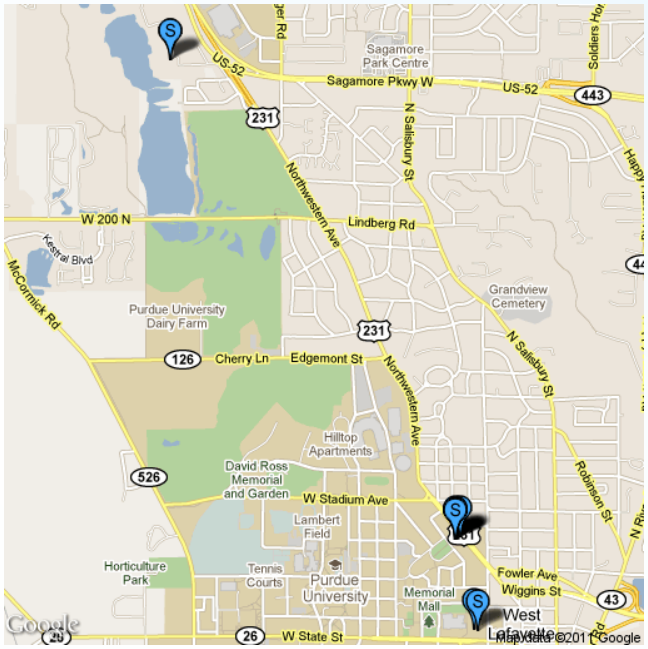
Browse Database

Desktop Version

Google maps



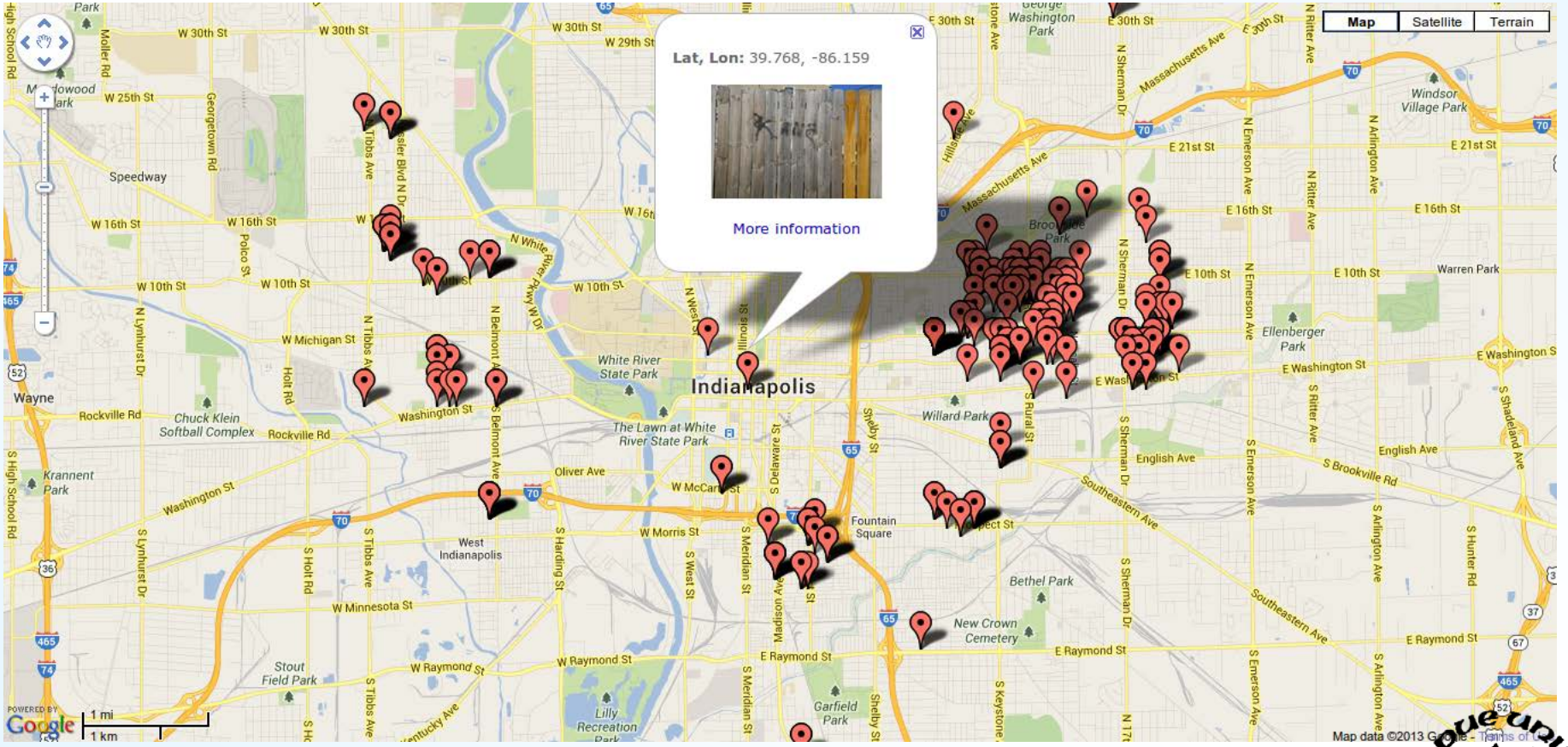
Specific image



All images in radius

Browse Database

Desktop Version



Gang Graffiti – Image Analysis





Light Sensor Value

Blur Prevention and Detection

Input Image

1

Color Correction

2

Touchscreen Tracing

3

Gaussian Color Segmentation

4

Segmentation Enhancement

5

Background Stripe Removal

6

Graffiti Component Reconnection

7

Graffiti Component Extraction

8

Content-Based Image Retrieval

9

Gang Graffiti

Gang Graffiti – Blur Prevention

- **Camera automatically takes image after 4 second countdown**
- **Using smartphone's accelerometer to detect shaking**
  - **If shaking, counter restarts**
- **Forcing camera to autofocus before taking image**



Gang Graffiti – Color Correction

- **Allows color correction without fiducial marker**
- **Non-intrusive**



- **Automatically assigns color correction matrix to light sensor value**



Before





After

Gang Graffiti – Color Segmentation

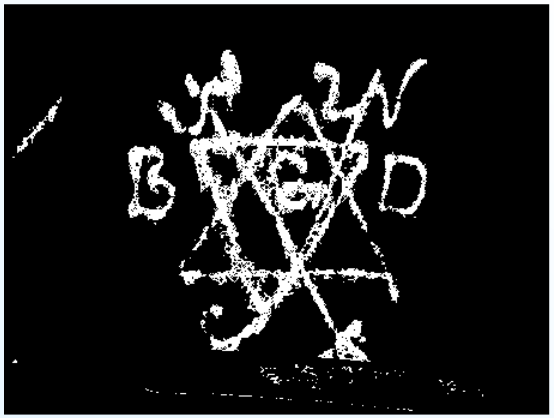




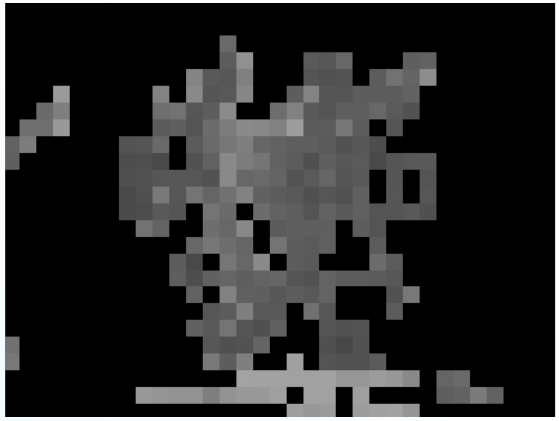
Gang Graffiti – Segmentation Enhancement



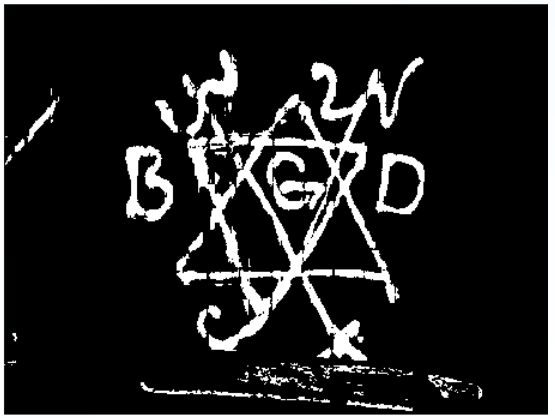
Original Image



Initial Segmentation



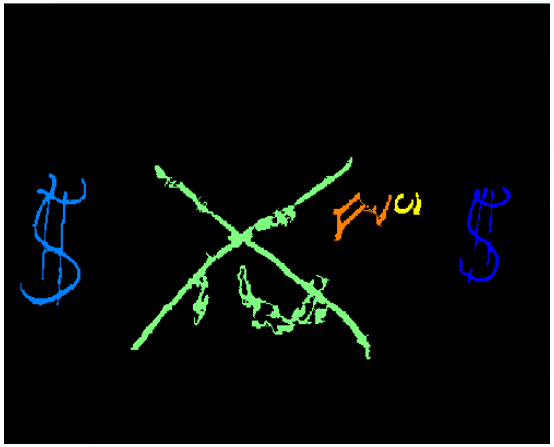
Block-Wise Luma

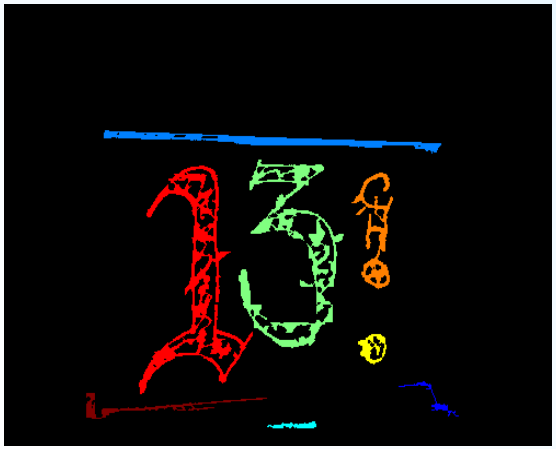


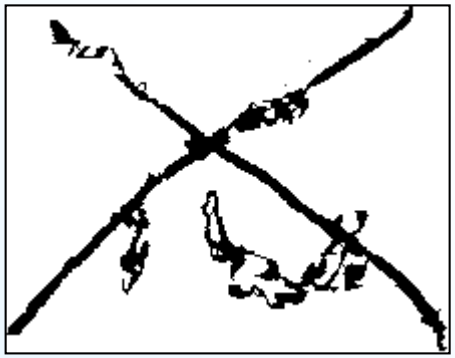
Enhanced

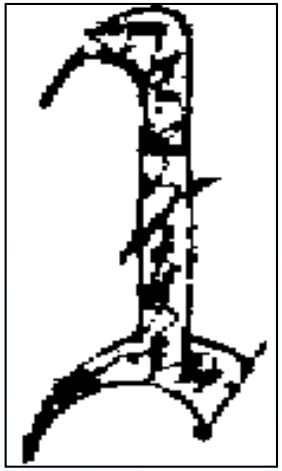
Gang Graffiti – Component Extraction









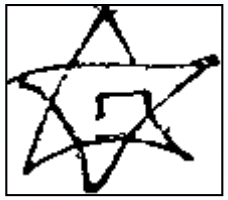




Gang Graffiti – Component Classifcation







Sybmol: 6-Point Star

Confidence: 84%



Symbol: Crown

Confidence: 80%



Symbol: Pitchfork

Confidence: 75%



Symbol: Unknown



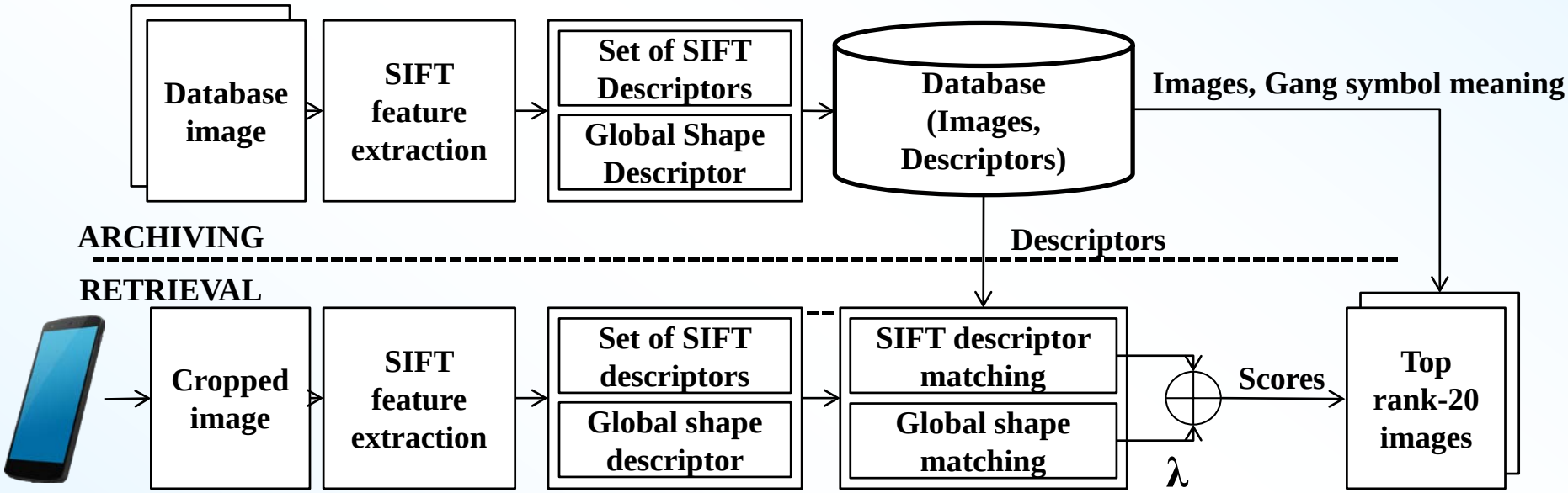
Symbol: Unknown

Gang Graffiti – Find Similar Images



Retrieval: Accuracy: 99.10%, Time: 70 ms

Gang Tattoo – Image Analysis



Gang Tattoo – Image Analysis



Original Image



Cropped Image

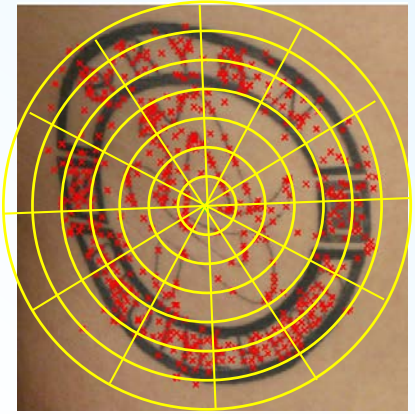


Image Analysis

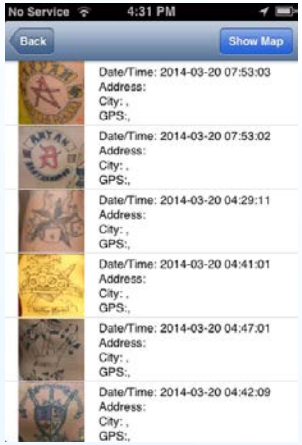
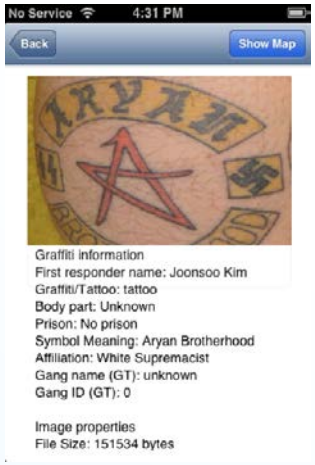


Image Matching



Analysis Results

Graffiti information
First responder name: Joonsoo Kim
Graffiti/Tattoo: tattoo
Body part: Unknown
Prison: No prison
Symbol Meaning: Aryan Brotherhood
Affiliation: White Supremacist
Gang name (GT): unknown
Gang ID (GT): 0

Gang Tattoo – Find Similar Images



Retrieval Accuracy: 81.37% Top-10 (57.84% Top-1)

GARI Availability

- **The GARI system is available free to any law enforcement agency**

- **INGang Members can receive GARI now (all images are shared)**
- **Not in Indiana**
  - **MOU with Purdue University**
  - **Agency must agree to host a GARI server or cooperate with an agency currently using GARI**
  - **Software and server setup (and support) provided by VACCINE/Purdue University**

For more information: gari@ecn.purdue.edu

Characterizing the Uncertainty of Classification Methods and Its Impact on the Performance of Crowdsourcing

Javier Ribera, Khalid Tahboub, Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana



Outline

1. **Motivation**
2. **Uncertainty characterizations**
3. **Crowdsourcing**
4. **Crowd flow estimation**
5. **Results**
6. **Conclusions**

Motivation

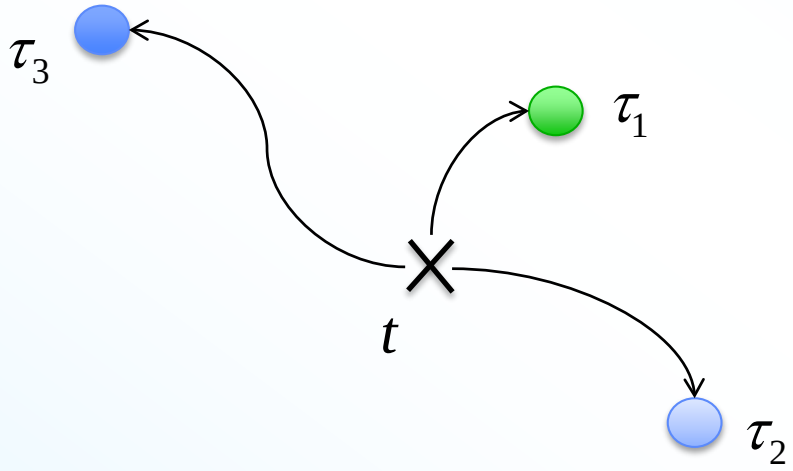
- **Video surveillance applications demand robust video analytics**
- **Changing environments, occlusions, shadows, video distortions**
- **Uncertainty in classification methods is quantified**

Motivation

- **Incorporate crowdsourcing**
- **Online active learning**
- **How to aggregate the contributions of the crowd?**

Uncertainty

- **Classification is crucial to many methods**
- **New data points need to be labeled using the given model**



Uncertainty Characterizations

- **When is the unlabeled data point "close enough" to the nearest labeled data point?**

- **Uncertainty quantification of the classification: μ**

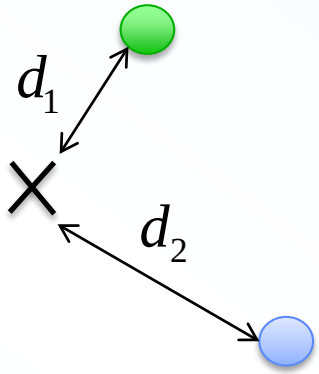
- **Quantify and threshold (maximum uncertainty allowed)**

$$\mu < \alpha$$

Uncertainty Characterizations

– 1st uncertainty characterization

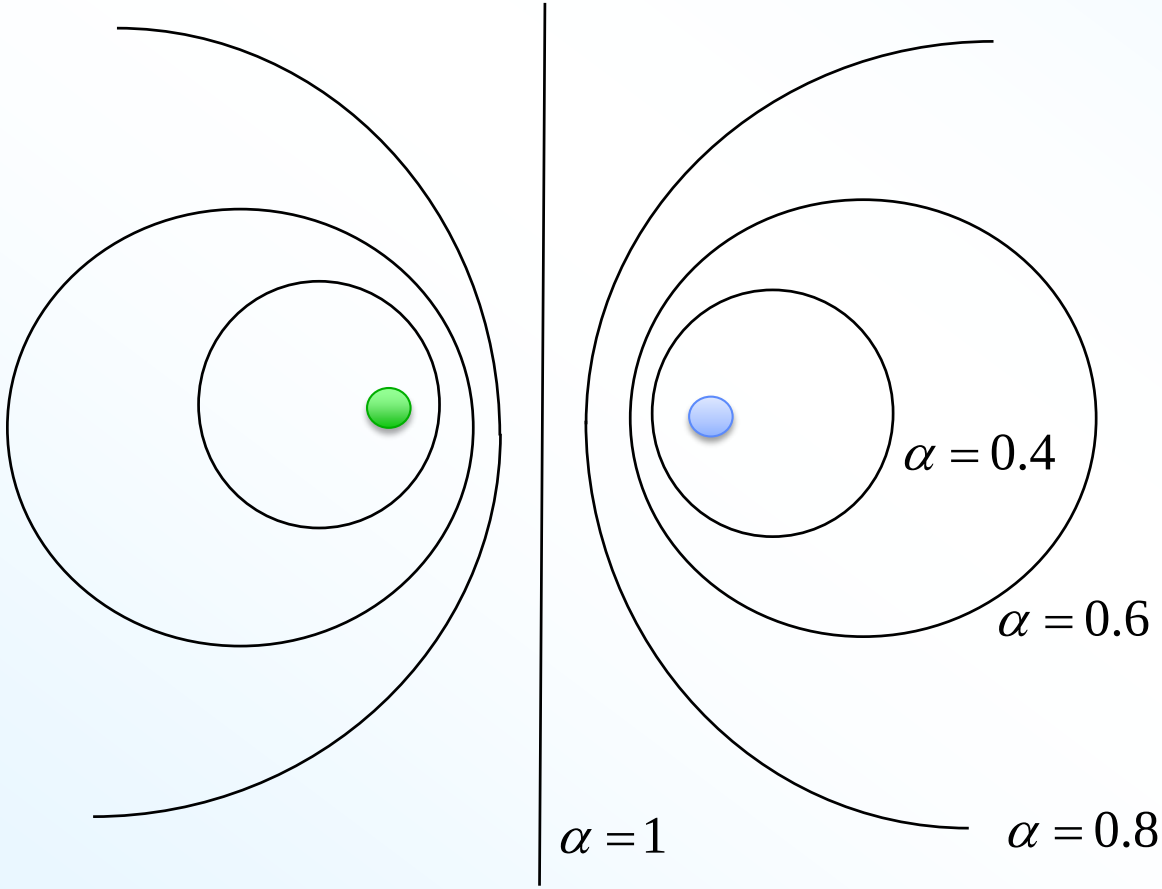
$$\mu_1 = \frac{d_1}{d_2} < \alpha$$



Uncertainty Characterizations

– **1st uncertainty characterization**

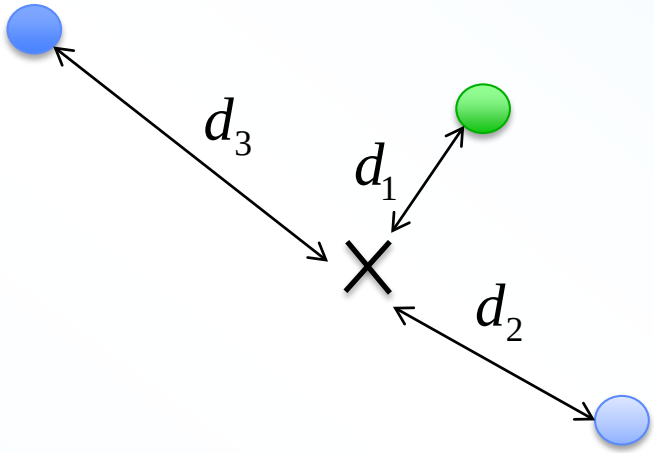
$$\mu_1 = \frac{d_1}{d_2} < \alpha$$



Uncertainty Characterizations

– 2nd uncertainty characterization

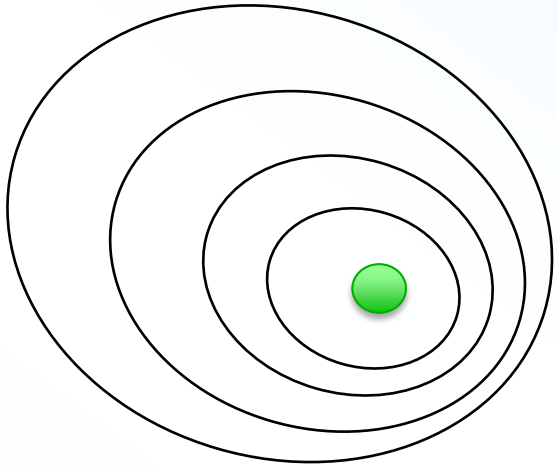
$$\mu_2 = \sum_{i=2}^{L} \frac{d_1}{d_i} < \alpha$$



Uncertainty Characterizations

– 2nd uncertainty characterization

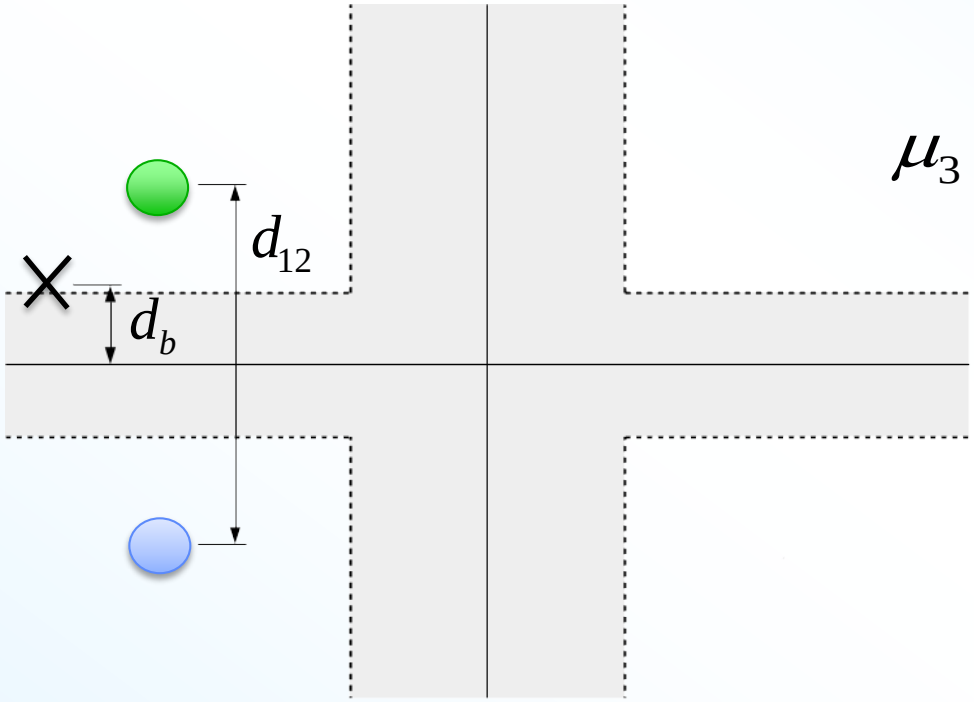
$$\mu_2 = \sum_{i=2}^{L} \frac{d_1}{d_i} < \alpha$$



Uncertainty Characterizations

– 3rd uncertainty characterization

$$\mu_3' = \frac{d_{12}}{d_b}$$

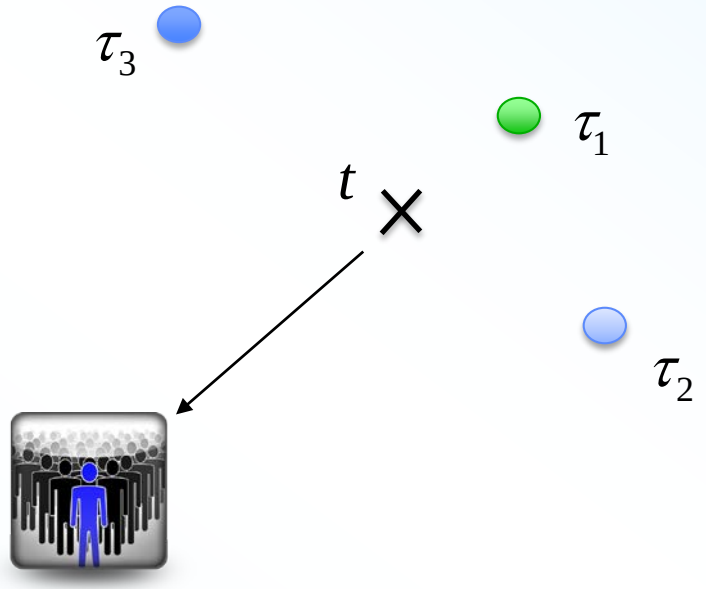


$$\mu_3 = \frac{1}{1 + \frac{1}{\mu_3'}}$$

Crowdsourcing

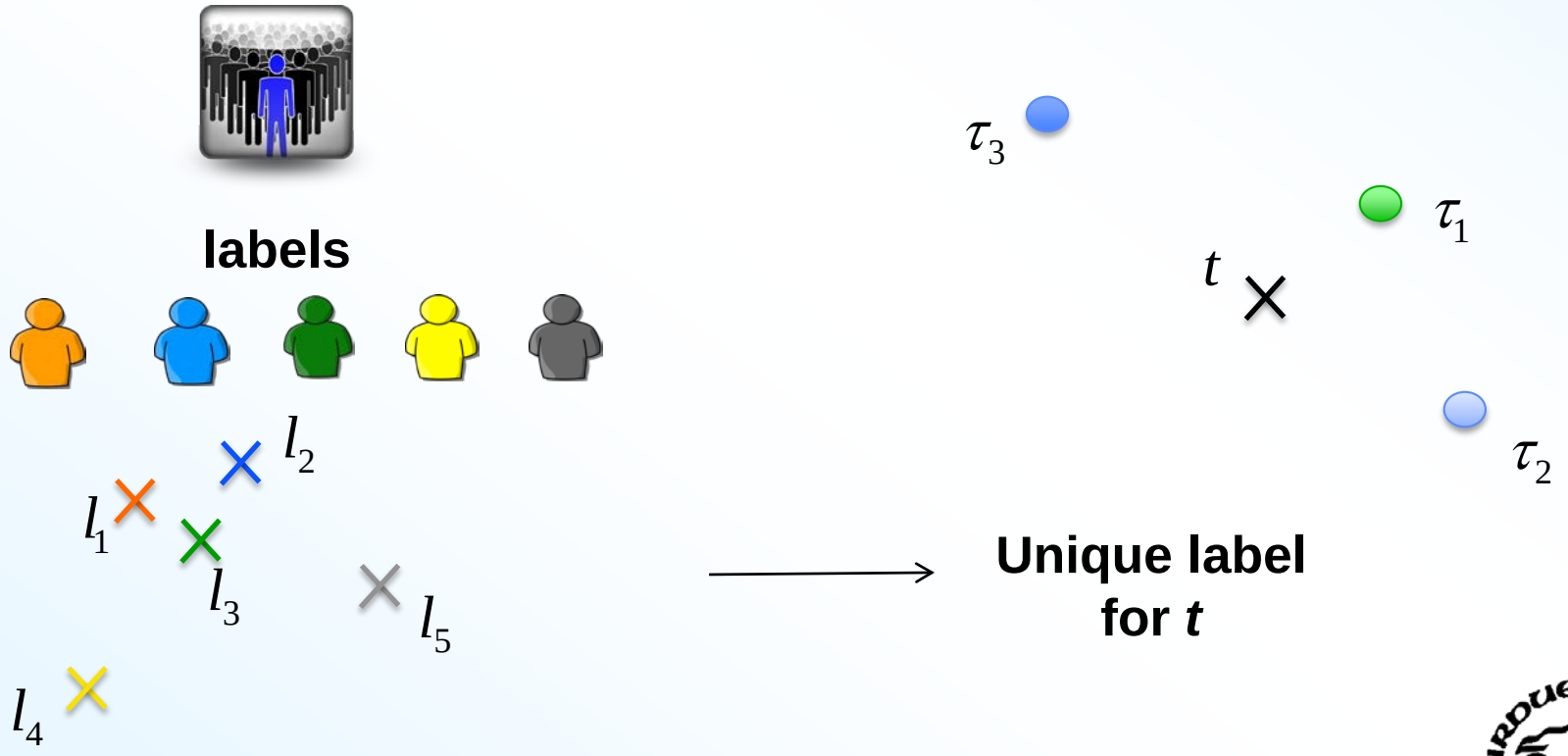
- **Crowdsourcing is incorporated when the classification uncertainty is estimated to be high**

$$\mu > \alpha \Rightarrow \text{crowdsourcing}$$



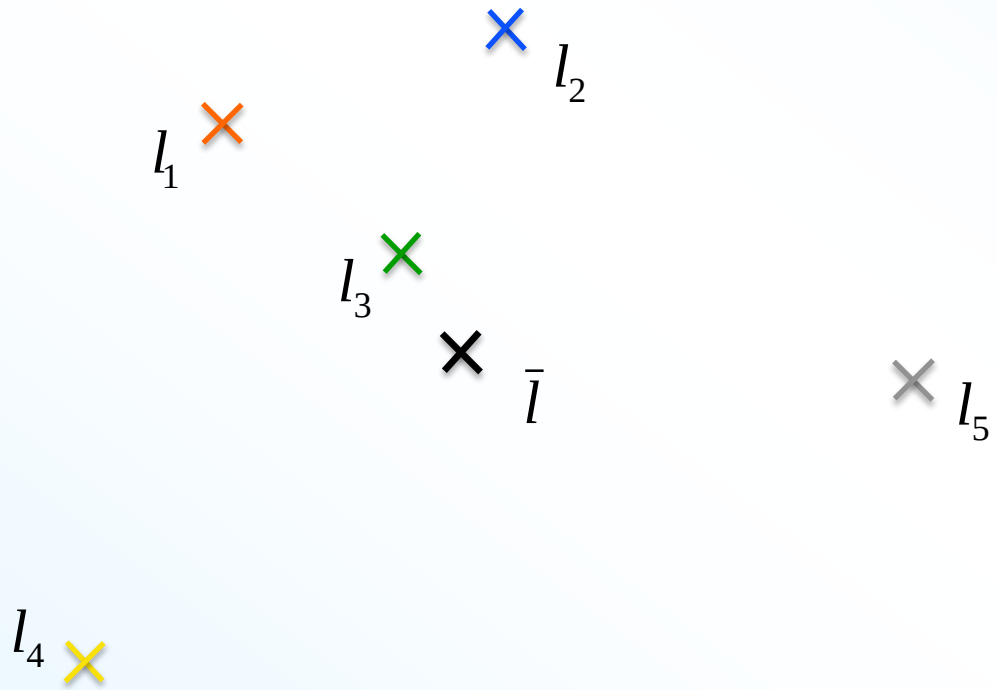
Crowdsourcing

- **How to aggregate the labels of the crowd?**



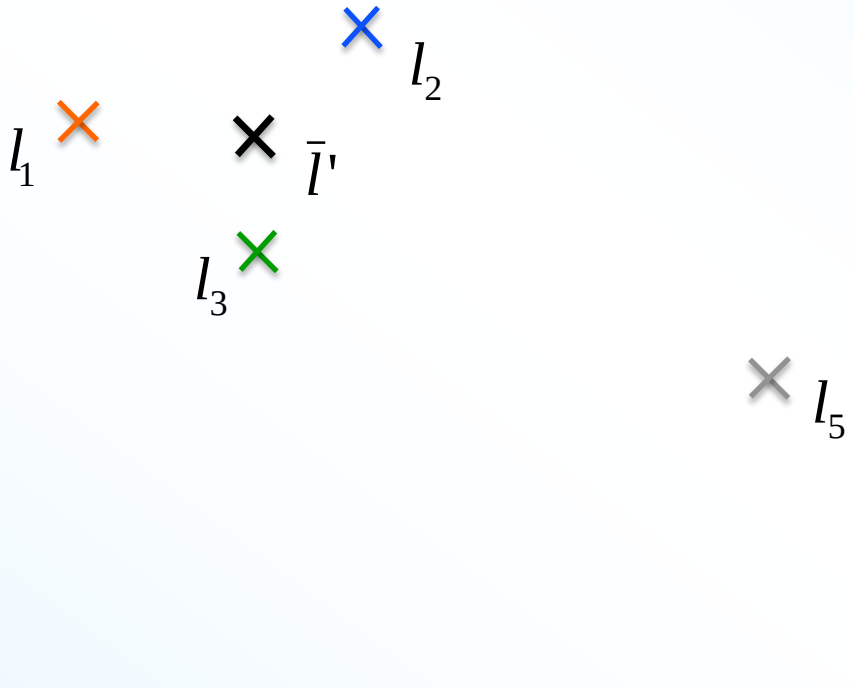
Crowdsourcing

1. **Average crowd labels for the same data point**



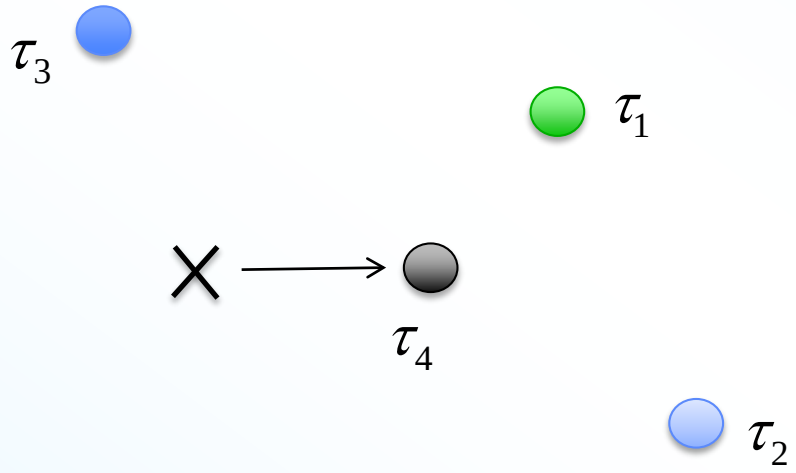
Crowdsourcing

2. **Use the subset with lowest variance (discard the outliers)**



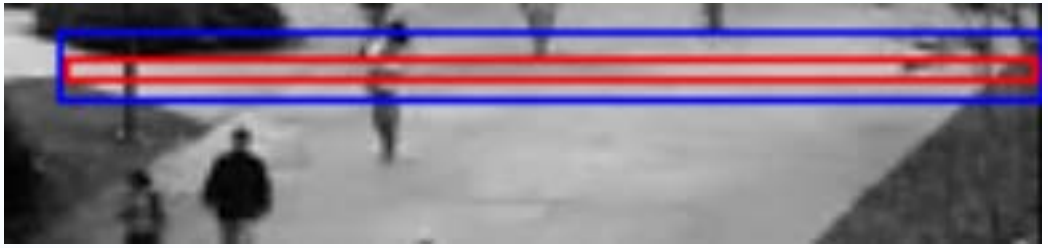
Active Learning

- **New labeled data point is incorporated into the model to enhance future classifications**



Crowd Flow Estimation

- **Application: estimate the number of people that have crossed a specific region**



Tripwire (red)
Region of Interest (blue)

Crowd Flow Estimation

- **Estimation: proportional to the number of foreground pixels in the Tripwire**

$$\upsilon = \frac{S_N}{C}$$

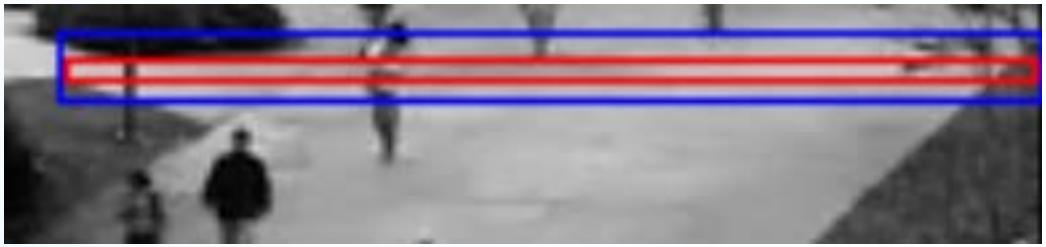


Tripwire region

- ***C* is the number of pixels seen per person: depends on the level of crowdedness (occlusions)**

Crowd Flow Estimation

1. **Compute foreground segmentation in the Tripwire and accumulate foreground pixels.**
2. **Estimate level of crowdedness in the Region of Interest by means of texture features.**

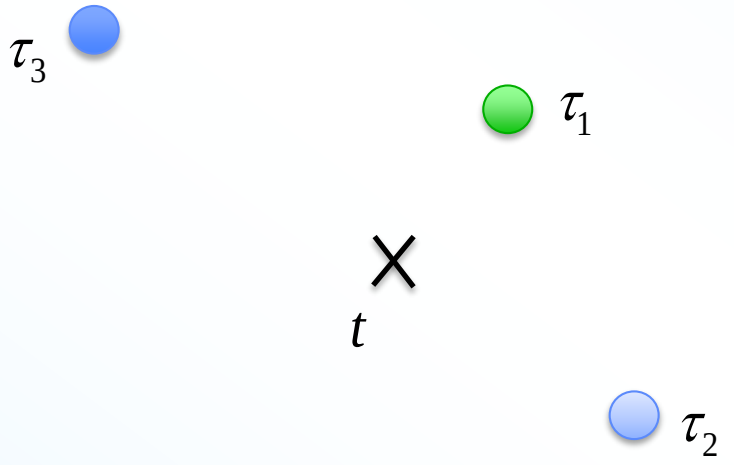


Tripwire (red)
ROI (blue)

- **Feature vector: 16-D texture feature vector built using the Gray Level Co-occurrence Matrix.**

Crowd Flow Estimation

- **The texture feature vector *t* is the testing data point for our model in a 16-D texture feature space**
- **The "reference" feature vectors τ_i have been manually labeled in the training stage with their corresponding *C*, based on their level of crowdedness**



Experiments

- **UCSD pedestrian dataset**
- **Experiments run with degraded quality**
- **Which uncertainty characterization leads to better results?**
- **How crowd labels aggregation methods impact the performance?**

Results

- **With no crowdsourcing (α=1):**

	Average error rate
Original quality	8.1%
Degraded quality	31.7%

- **Degradation consisted of boosting contrast and compressing the video file to 40%**

Results

- **Comparison of uncertainty characterizations:**

	Uncertainty characterization	α	Average error rate
Degraded quality	# 1	0.6	15.2 %
Degraded quality	# 2	0.5	12.5 %

- **Degradation consisted in boosting contrast and compressing the video file to a 40%**

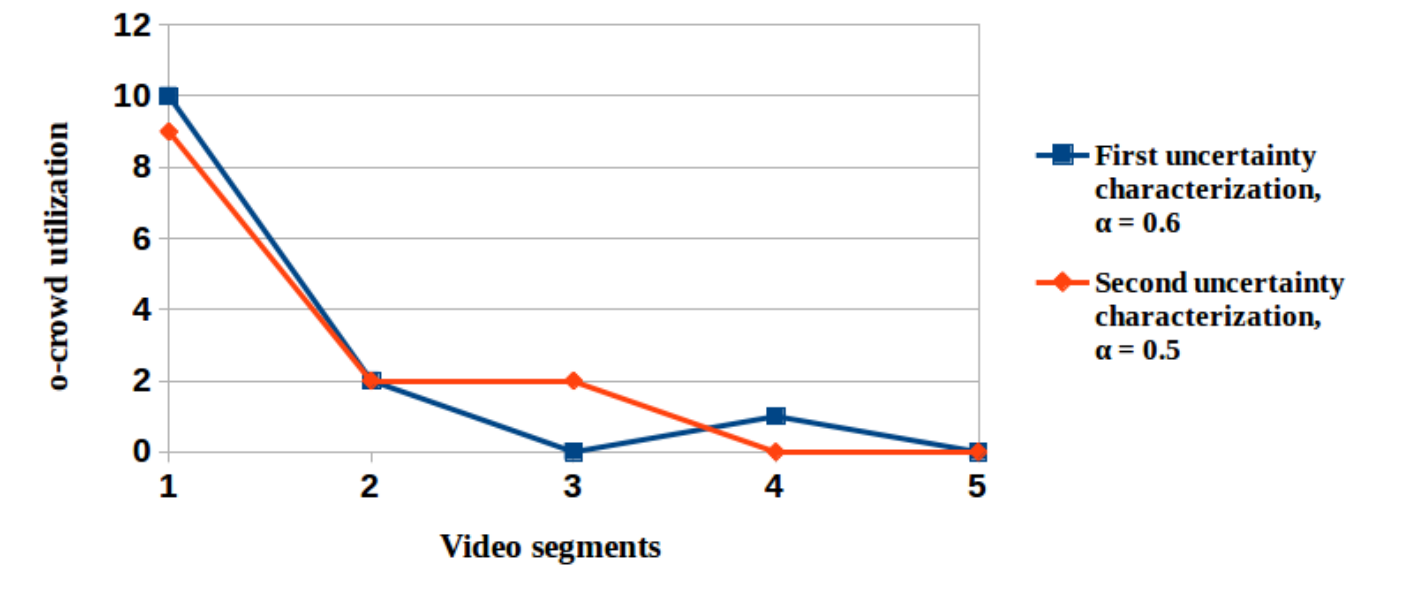
Results

- **Comparison of crowd aggregations:**

Aggregation of K of the crowd members	Average error rate
K = 5	26 %
K = 3	19.5 %

Results

- **Active learning**
- **The method learns from crowdsourcing and the model adapts to the new conditions.**



Conclusions

- **Video analytics can be enhanced by characterizing uncertainties and use crowdsourcing to address it**
- **Different characterizations of the uncertainty of a classification are proposed and evaluated**
- **Aggregating results from the labelers with smallest variance yields better results**
- **Future work includes adding the history of the performance of the crowd members into the model**
- **Further performance analysis**

Characterizing the Uncertainty of Classification Methods and Its Impact on the Performance of Crowdsourcing

Javier Ribera, Khalid Tahboub, Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana



Efficient Graph-Cut Tattoo Segmentation

Joonsoo Kim, Albert Parra, He Li, and Edward J. Delp
Video and Image Processing Laboratory (*VIPER)*
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu

Outline

- **Motivation**
- **Overview of Existing Methods**
- **Proposed Method**
- **Experimental Results**
- **Conclusions and Discussion**

Motivation

- **Law enforcement is interested in exploiting tattoos as an information source to identify, track and prevent gang-related crimes**

- **The use of tattoos for person identification and tattoo symbol interpretation is an interesting problem for law enforcement**









Motivation

- **In a retrieval system tattoo segmentation is an important step for retrieval accuracy since segmentation removes background information in a tattoo image**

- **Many existing methods for tattoo segmentation have problems**
  - **Some existing methods use pre-cropped images whose background is skin only**
  - **Existing methods using uncropped image are not robust when background has color similar to skin tones, there is an illumination change, and an image has diverse skin tones**

Overview of Existing Methods

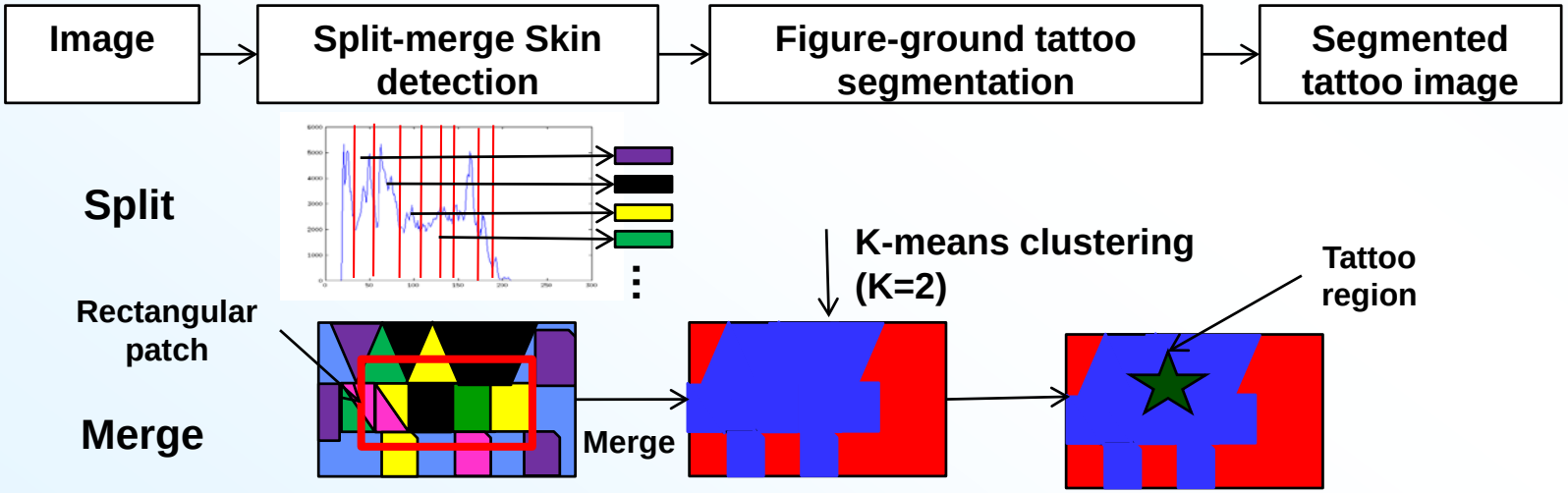
P. Duangphasuk and W. Kurutach, "Tattoo skin detection and segmentation using image negative method," Proceedings of the International Symposium on Communications and Information Technologies, pp. 354-359, September 2013, SuratThani, Thailand

- **The HSV color space is used to find skin regions**
- **Suggested a fixed color range to detect skin pixels - restricted to Asian people**
- **Can be affected by illumination, background, and camera characteristics**

Overview of Existing Methods

J. Allen, N. Zhao, J. Yuan, and X. Liu, "Unsupervised tattoo segmentation combining bottom-up and top-down cues," Proceedings of the SPIE Mobile Multimedia/Image Processing, Security, and Applications, vol. 8063, pp. 1–9, April 2011, Orlando, FL

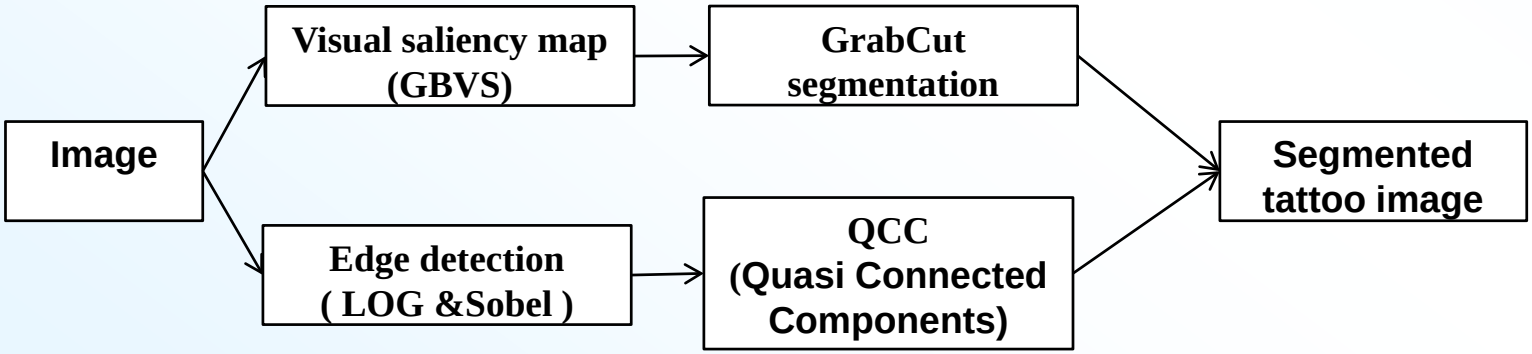
- Used histogram on image intensity(gray-scale) and multiple clusters
- Merge most of the clusters
- K-means clustering is used in RGB space to distinguish a tattoo/skin region



Overview of Existing Methods

B. Heflin, W. Scheirer, and T.E. Boult, "Detecting and classifying scars, marks, and tattoos found in the wild," Proceedings of the IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 31–38, September 2012

- **Region near the tattoo is detected based on a visual saliency model**
- **GrabCut segmentation is used in the saliency region**
- **Edge map with QCC(Quasi Connected Components) is combined with the Grabcut segmentation result**
- **The accuracy of this method highly depends on the accuracy of a visual saliency map**



Proposed Method

- **Two important assumptions**
  - **Tattoos are surrounded by skin (closed loop)**
  - **Edges between tattoo and skin**
- **Tattoo detection becomes detecting skin pixels around tattoo**
  - **Skin model**
  - **Regions near image edges are only considered for "possible segmentation regions"**
  - **Segmentation errors outside of the possible segmentation regions can be avoided**
  - **Computationally efficient**

Approach



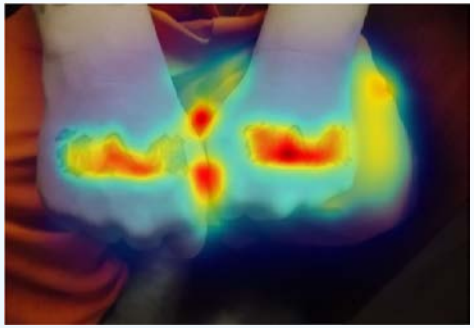
(a) Original image



(b) Possible Segmentation Regions



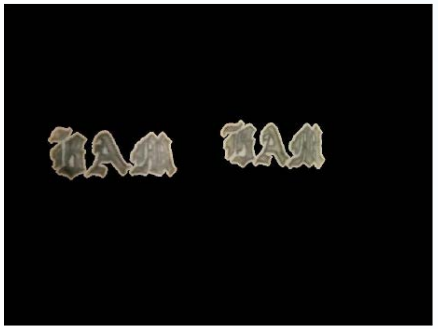
(c) Skin Probability map



(d) Visual Saliency map

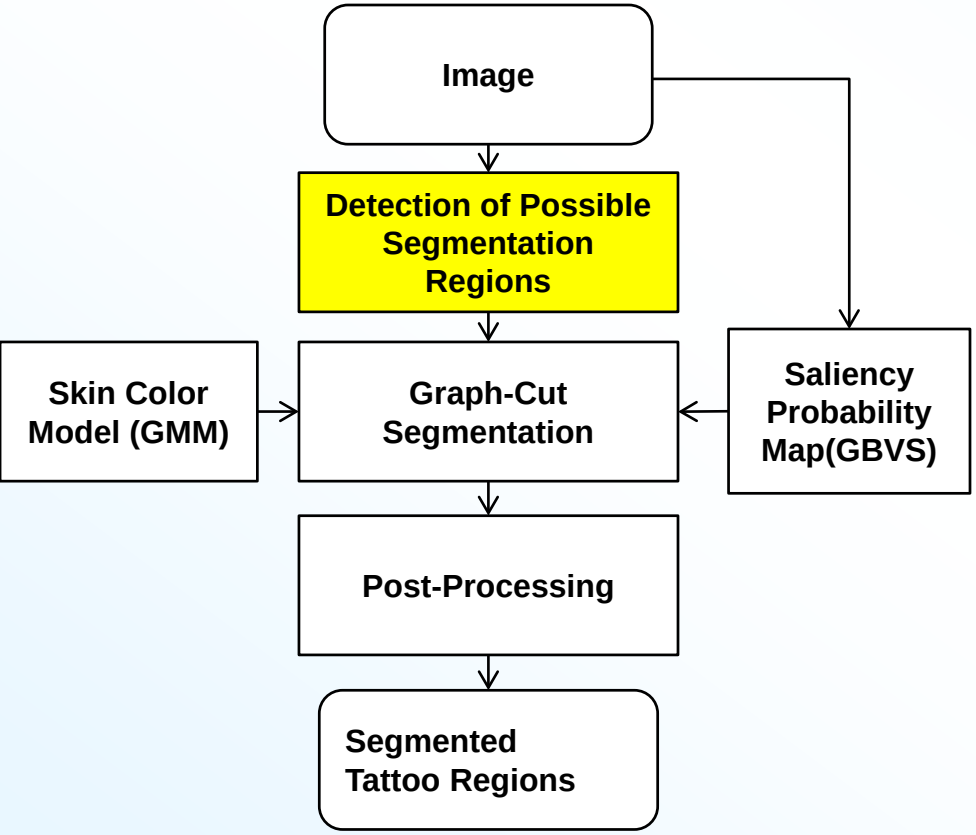


(e) Graph-Cut Segmentation



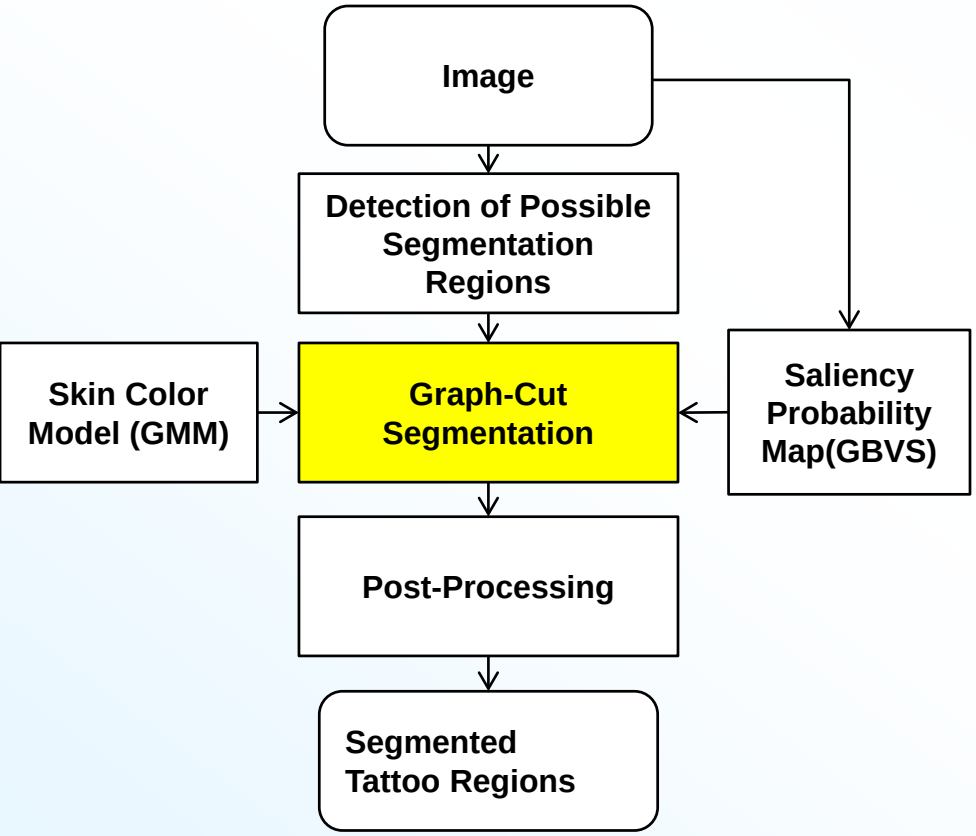
(f) Tattoo segmentation

Our Method



- Canny edge detector is used to find edges for each RGB color channel
- The edge regions from each color channel are combined
- Morphological dilation is used to the combined edge regions for finding the possible segmentation regions

Segmentation



- **Graph–cut segmentation – minimize the Gibbs energy**

$$E(x) = \sum_{i \in I} D(x_i) + \lambda \sum_{i \in I, j \in N_i} V(x_i, x_j),$$

$D(x_i)$: the data term

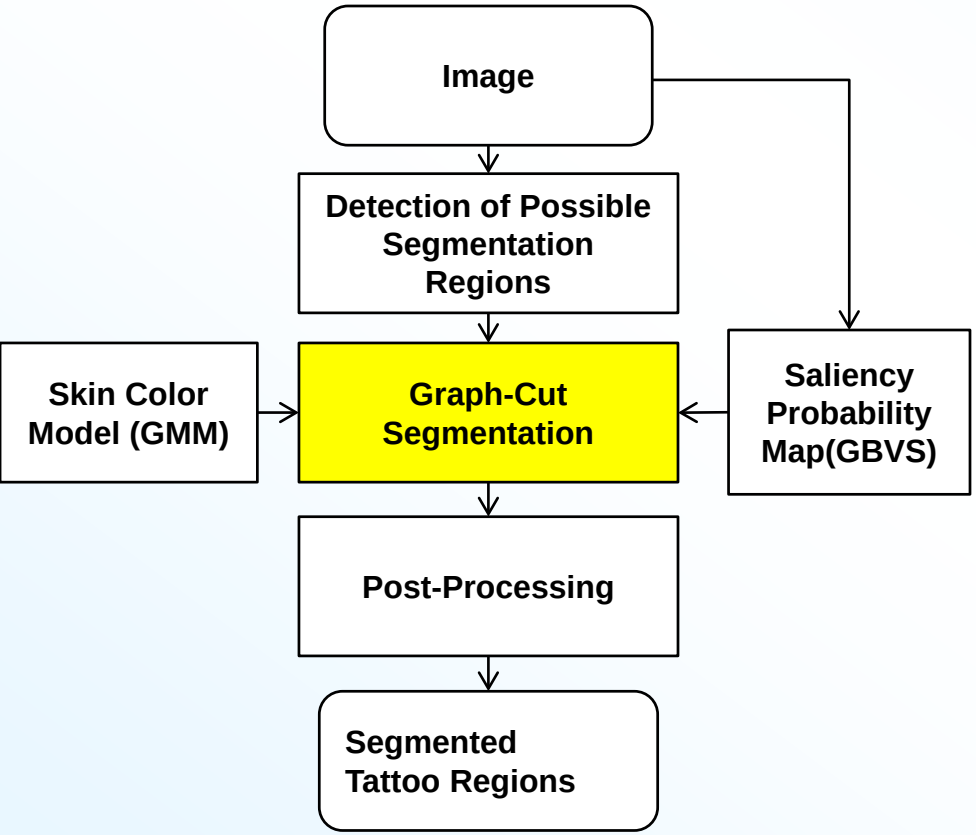
$V(x_i,x_j)$: the smoothness term

I : image

N_i : neighborhood pixels of pixel I

$x_i \in$ { 0(background), 1(foreground) }

Segmentation



- **$D(x_i)$: the data term**

$$D(x_i) = w_1 D_1(x_i) + w_2 D_2(x_i),$$

$D_1(x_i)$: energy for skin or non-skin model based on GMM

$D_2(x_i)$: energy for normalized saliency map

$$V(x_i, x_j) = |x_i - x_j| * f(C_{ij})$$

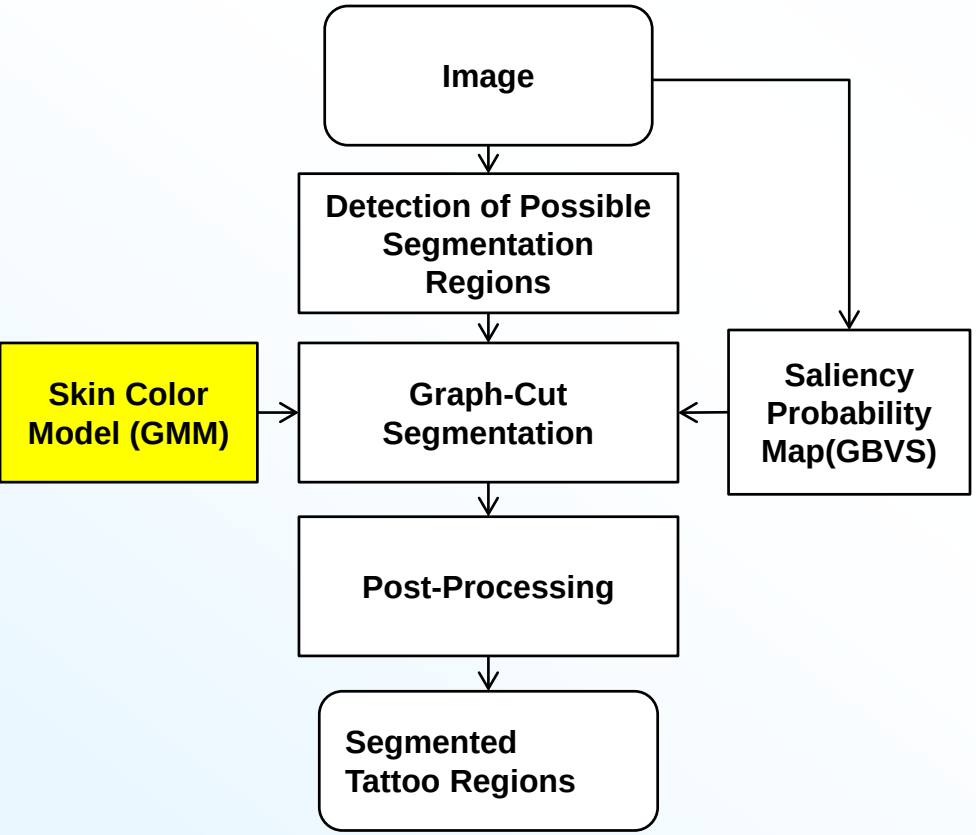
- **$V(x_i,x_j)$: the smoothness term**

$f(\varepsilon) = 1/(1+\varepsilon)$,

$C_{ij} = \|C_i - C_j\|^2$

C_i : YCbCr color of pixel i

Skin Model



- **GMM skin model - train on diverse skin tones**

$$D(x_i) = w_1 D_1(x_i) + w_2 D_2(x_i),$$

$D_1(x_i)$: energy for skin/non-skin

$$D_1(x_i = 1) = -log(p_1(x_i = 1))$$

$$p(x_i = 1) = \sum_{j=1}^{M} \pi_j * g(C_i | u_j, \Sigma_j)$$

C_i : YCbCr color of i^{th} pixel

π_j : j^{th} mixture weight

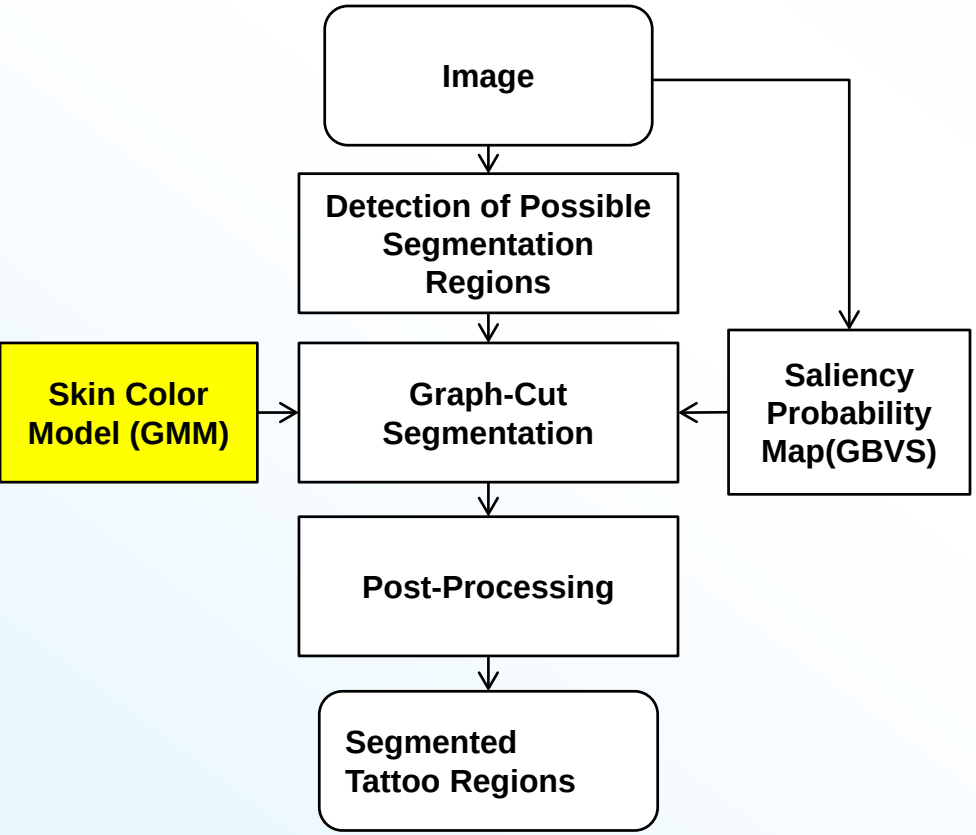
M : # of mixture components

g() : Gaussian density function

u_j, Σ_j : Gaussian parameters

- **π_j u_j, Σ_j are estimated using EM**

Skin Model

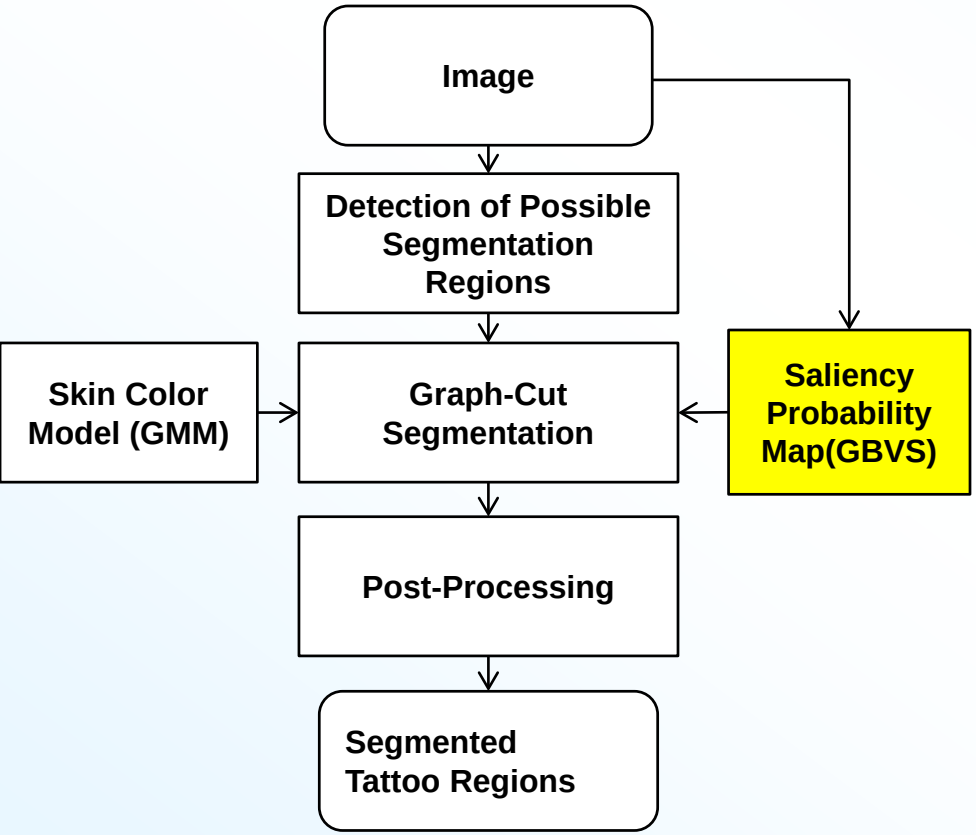


- **Cannot estimate all non-skin colors - use adaptive threshold for non-skin model**

$$D_1(x_i = 0) = -log(wt_1 * \frac{1}{n_I} \sum_{i \in I} p_1(x_i = 1))$$

- **Adaptive threshold helps segment skin when the skin colors in an image do not fit the skin model**

Saliency



- **Case for color similar to skin in background, use visual saliency map to focus on skin near tattoo**

$$D(x_i) = w_1 D_1(x_i) + w_2 D_2(x_i),$$

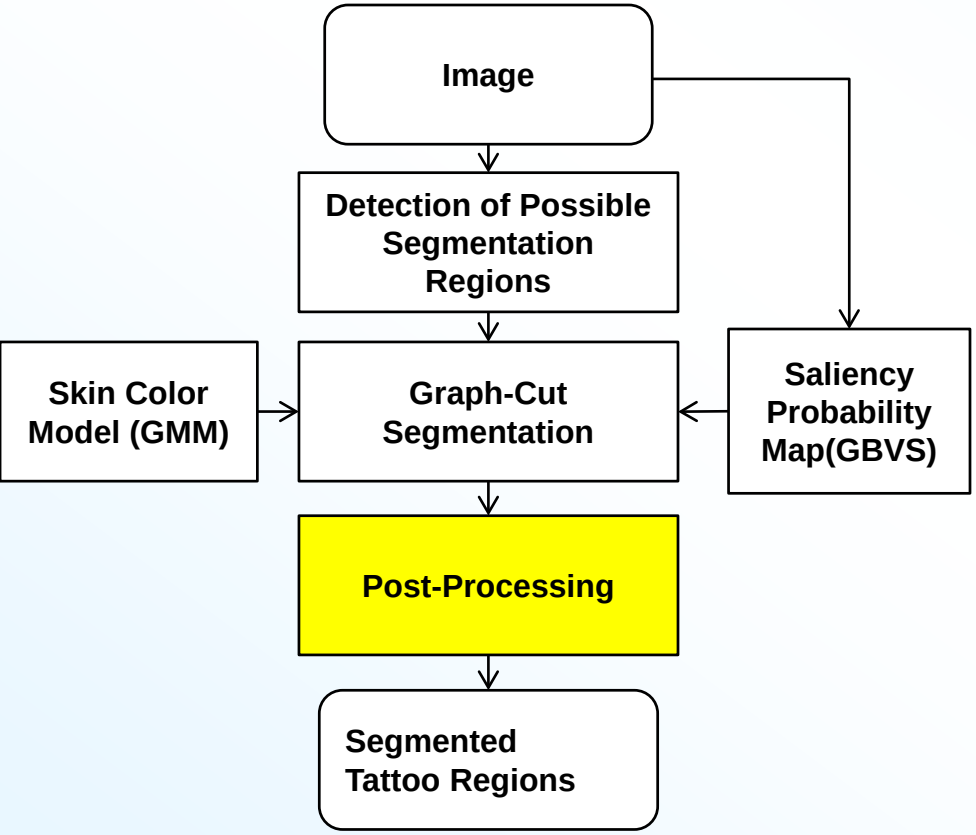
- **$D_2(x_i)$: energy for normalized saliency map**
- **Normalized GBVS saliency map is used for $p_2(x_i=1)$**

$$D_2(x_i = 1) = -log(p_2(x_i = 1))$$

- **Adaptive threshold is used for p2(xi=0)**

$$D_2(x_i = 0) = -log(wt_2 * \frac{1}{n_I} \sum_{i \in I} p_2(x_i = 1))$$

Post Processing



- **Detect both skin pixels near boundaries of the body and skin pixels near tattoos**
- **For skin pixels near tattoos, examine which set of skin pixels connected with each other forms a closed contour with holes inside**
- **Graph-cut segmentation errors - small holes(false contours)**
- **We check**

$$\frac{n_c + n_h}{n_b} >= t_f$$

n_c : the number of pixels- closed contour

n_h : the number of pixels for holes inside the closed contour

n_b : the number of pixels in the minimum bounding rectangular box including the closed contour

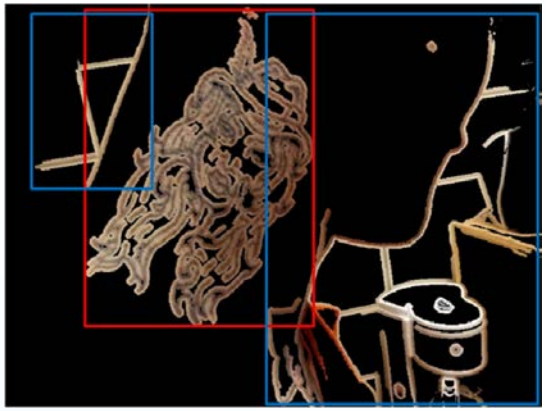
Post Processing

- **Example of Post processing**

: minimum bounding box including false contour

: minimum bounding box including a tattoo







(a) Original image **(b) Graph-Cut Segmentation** **(c) Tattoo segmentation after post processing**

Experimental Results

- **Tattoo images - Indiana Gang Network (INGangNetwork)**
- **TDSD dataset used to train skin model - includes diverse skin tones**
- **GMM - M=5**
- **Image retrieval system - do not segment the tattoo inside the closed contour found after post processing**
- **Compared with existing methods: Allen's method and Heflin's method**

Experimental Results



(a) Original image (b) Allen's method (c) Heflin's method (d) Proposed method

Experimental Results



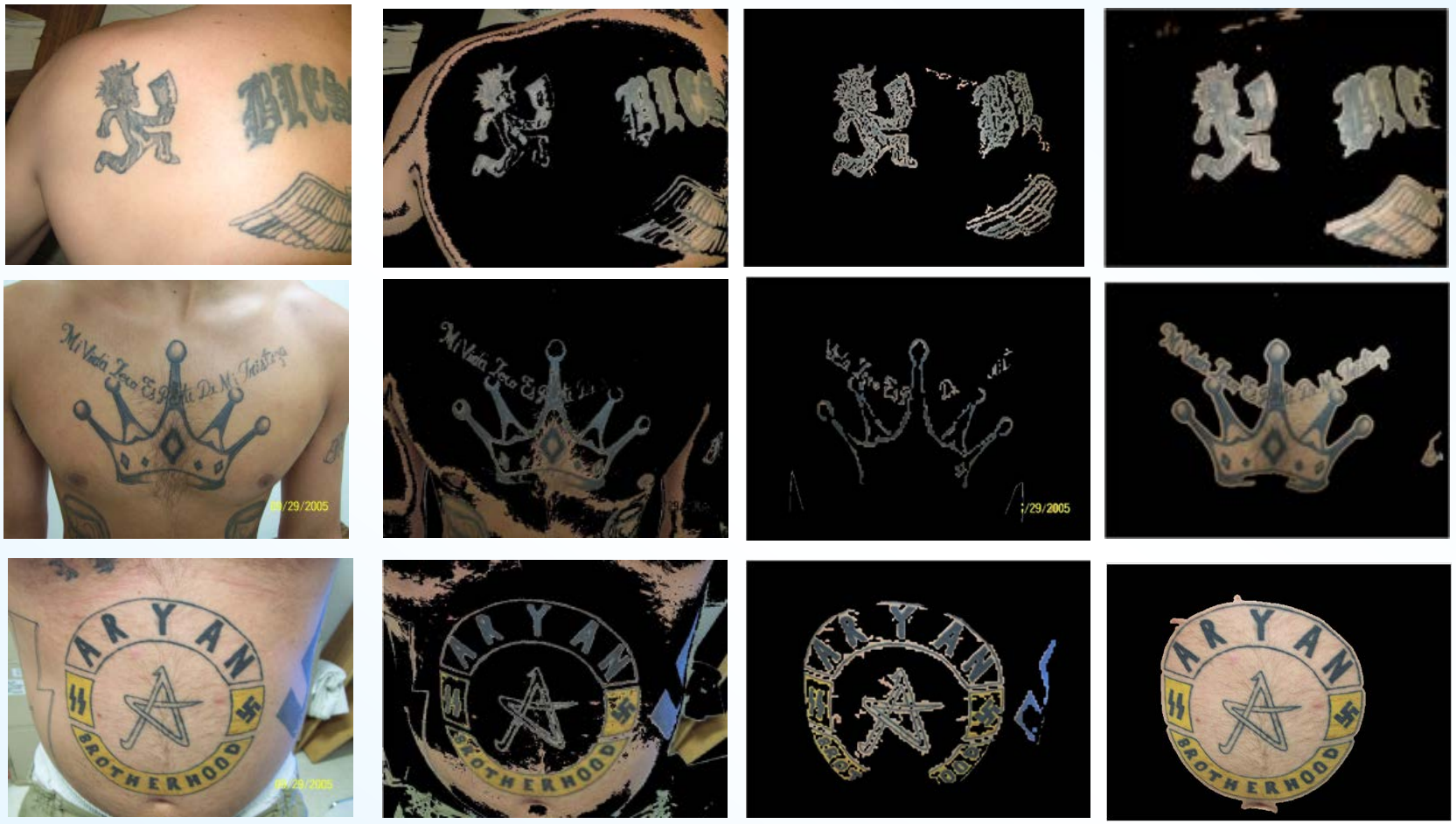
(a) Original image (b) Allen's method (c) Heflin's method (d) Proposed method

Experimental Results



(a) Original image (b) Allen's method (c) Heflin's method (d) Proposed method

Experimental Results



(a) Original image (b) Allen's method (c) Heflin's method (d) Proposed method

Experimental Results



(a) Original image **(b) Allen's method** **(c) Heflin's method** **(d) Proposed method**

Experimental Results

- **Our method segmented tattoo regions incorrectly when tattoo regions are very close to boundaries of the body and there are hair regions with strong edges inside a body**



(a) Original image (b) Allen's method (c) Heflin's method (d) Proposed method

Conclusions and Discussion

- **Using skin pixels near image edges allows one to segment tattoo regions more efficiently**
- **Our method sometimes fails to extract tattoo regions correctly when there are tattoo regions close to boundaries of the body and/or near “hair regions”**
- **In the future we will investigate techniques to detect these regions**

Efficient Graph-Cut Tattoo Segmentation

Joonsoo Kim, Albert Parra, He Li, and Edward J. Delp
Video and Image Processing Laboratory (*VIPER)*
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu

GARI
A Gang Tattoo and Graffiti System

He Li, Joonsoo Kim, Jiaju Yue, Edward J. Delp
(ace@ecn.purdue.edu)

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

Acknowledgements

- **This work is funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001**

- **Cook County Sheriff Department**

- **Indiana Gang Intelligence Network (INGangNetwork)**

GARI System Goals

- System that allows first responders to:
  - Collect gang graffiti and gang tattoos images
  - Analyze and interpret the images
  - Image retrieval
  - Interact with a database of images
  - Browse database using an interactive map
  - SQL database
- Platforms
  - Android and iPhone application
  - Web-based interface (report generation, bulk image upload, same features as mobile app)
  - CGAP (Civilian Gang Application Program)

GARI System Overview



Original image

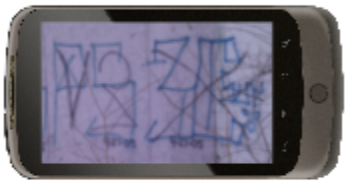
1

Offline automatic analysis and labeling
- Geoposition
- Date and time
- Extracted Features

3



Original Database



2

4

Filtered results
Info + thumbnails



Server



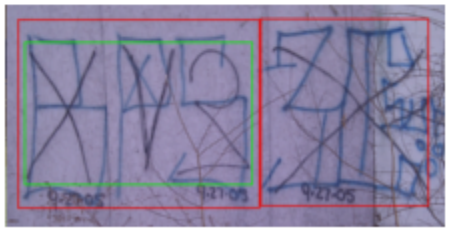
Filtered Database

5

Offline manual filtering

6

Manual labeling
- Additional Features



Labeled image

7

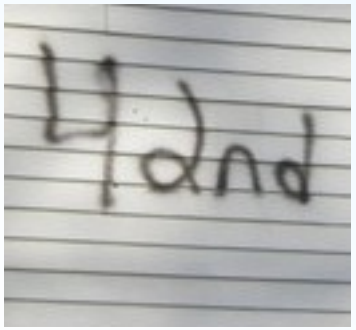
Addition to Database

Gang Graffiti Interpretation



SHAPE

Simple, Straightforward



NUMBERS

42nd street gang



SYMBOLS

6-point star, pitchforks



COLOR

Goon Squad: Red/Black

Gang Graffiti Interpretation



LETTERS

East Side Gang



TIME

Black: 18 ST (18th Street Gang)

Red: 13 SUR (Sureños 13)



POSITION/ALIGNMENT

Letters at star points

Numbers in the middle

Letters at the bottom

Pitchforks upright

Image Analysis – Color Recognition





MEXICANOS
MALDITOS SUREÑOS 13





18 STREET GANG



SUREÑOS 13

Image Analysis – Scene Analysis

GANG RIVALRY TRACKING



- **Date:** 08/19/2010
- **Time:** 3.25 PM
- **Geo:** 41.387917, 2.169919



- **Date:** 01/03/2011
- **Time:** 5.11 PM
- **Geo:** 41.387917, 2.169919

Gang Tattoos

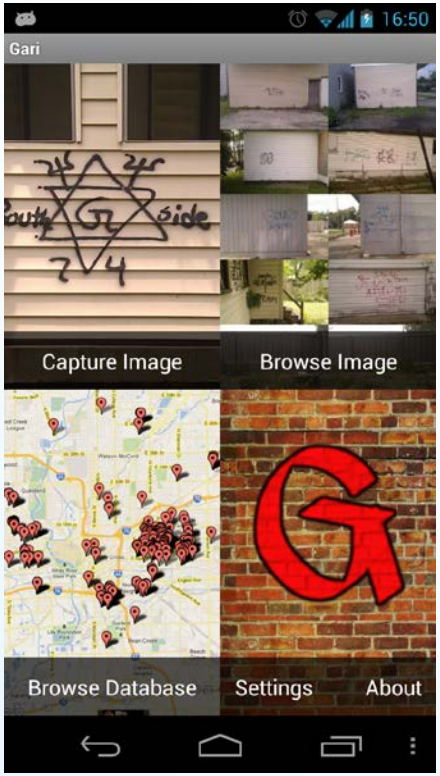




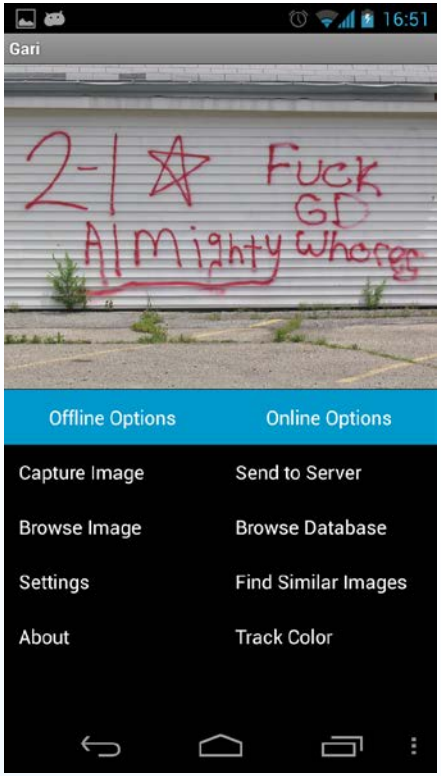




Mobile App User Interface



Main menu



User options

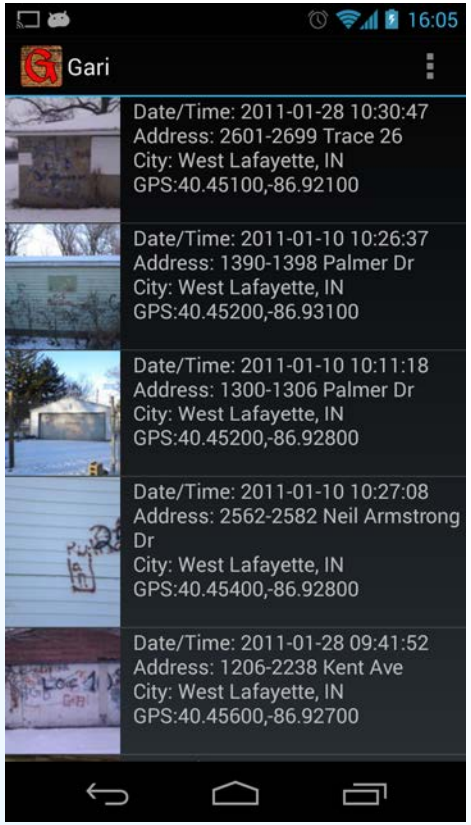


Analysis results

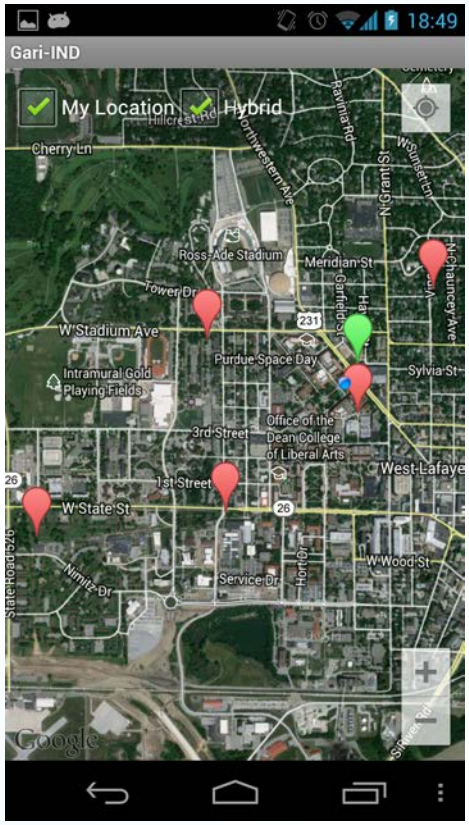


Image matching

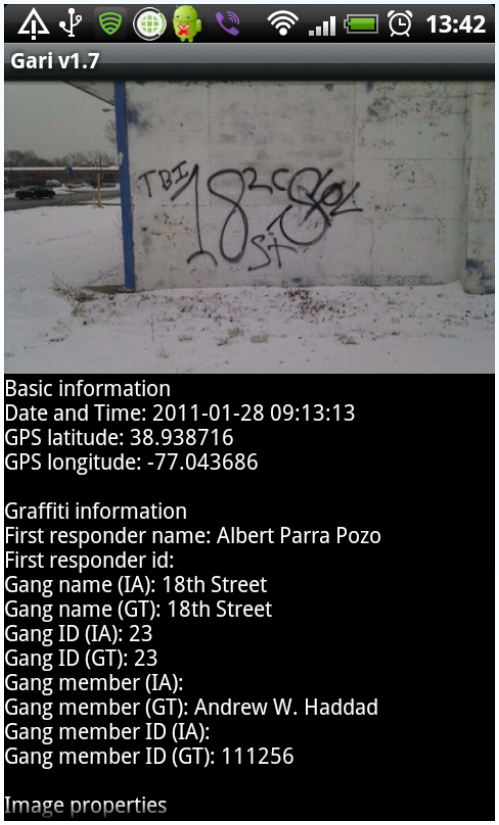
Browse Database



Browse general results



Show graffiti on map



Inspect specific graffiti

Augmented Reality Map









Browse Database

Desktop Version

- **Browse by**
  - **Radius**
  - **Date**
  - **Address**
  - **Zipcode**
  - **Firstresponder ID**



Database browsing screenshots



Browse Database

Desktop Version

VIPER

Video and Image Processing Laboratory

VACCINE

Main

Archive

Browse Database

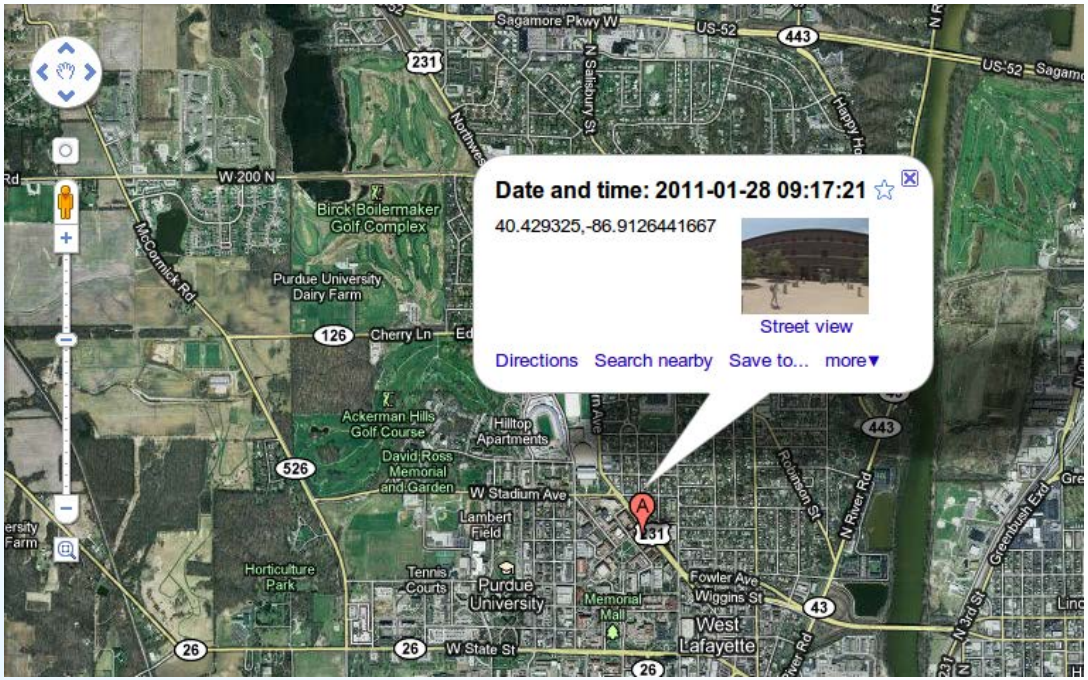


Date and Time: 2011-01-28 09:17:21
GPS latitude: 40.429325
GPS longitude: -86.9126441667
File Size: 664501 bytes
Height: 1552 px
Width: 2592 px
Focal length: 4.31 mm
Camera make: HTC
Camera model: HTC Desire
Image id: 1111
Show in map

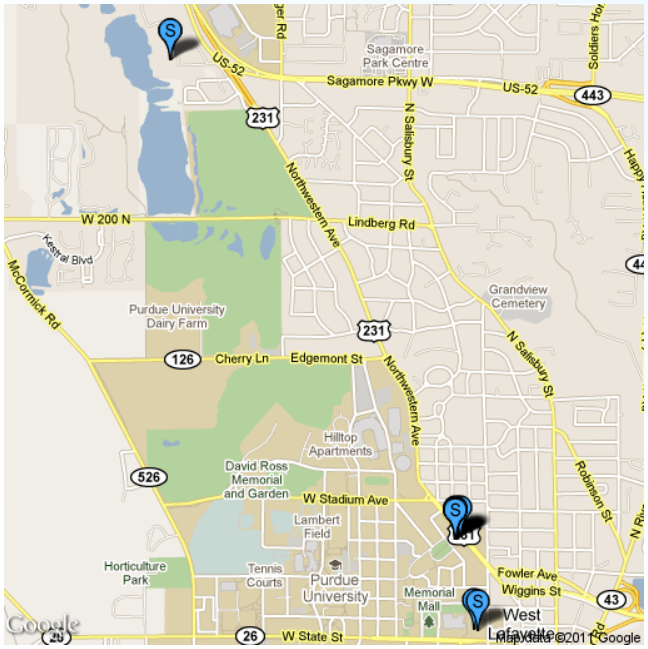
Browse Database

Desktop Version

Google maps



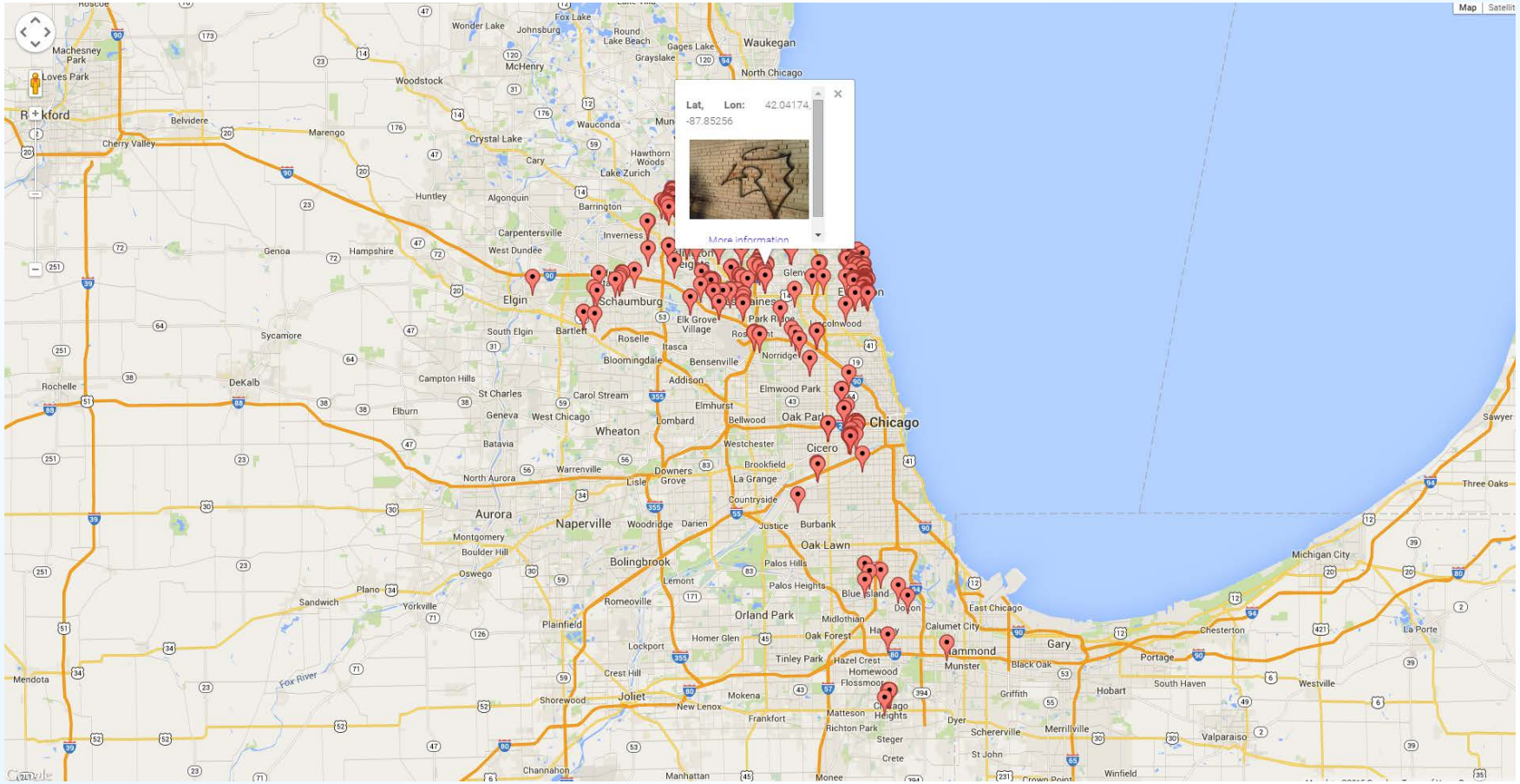
Specific image



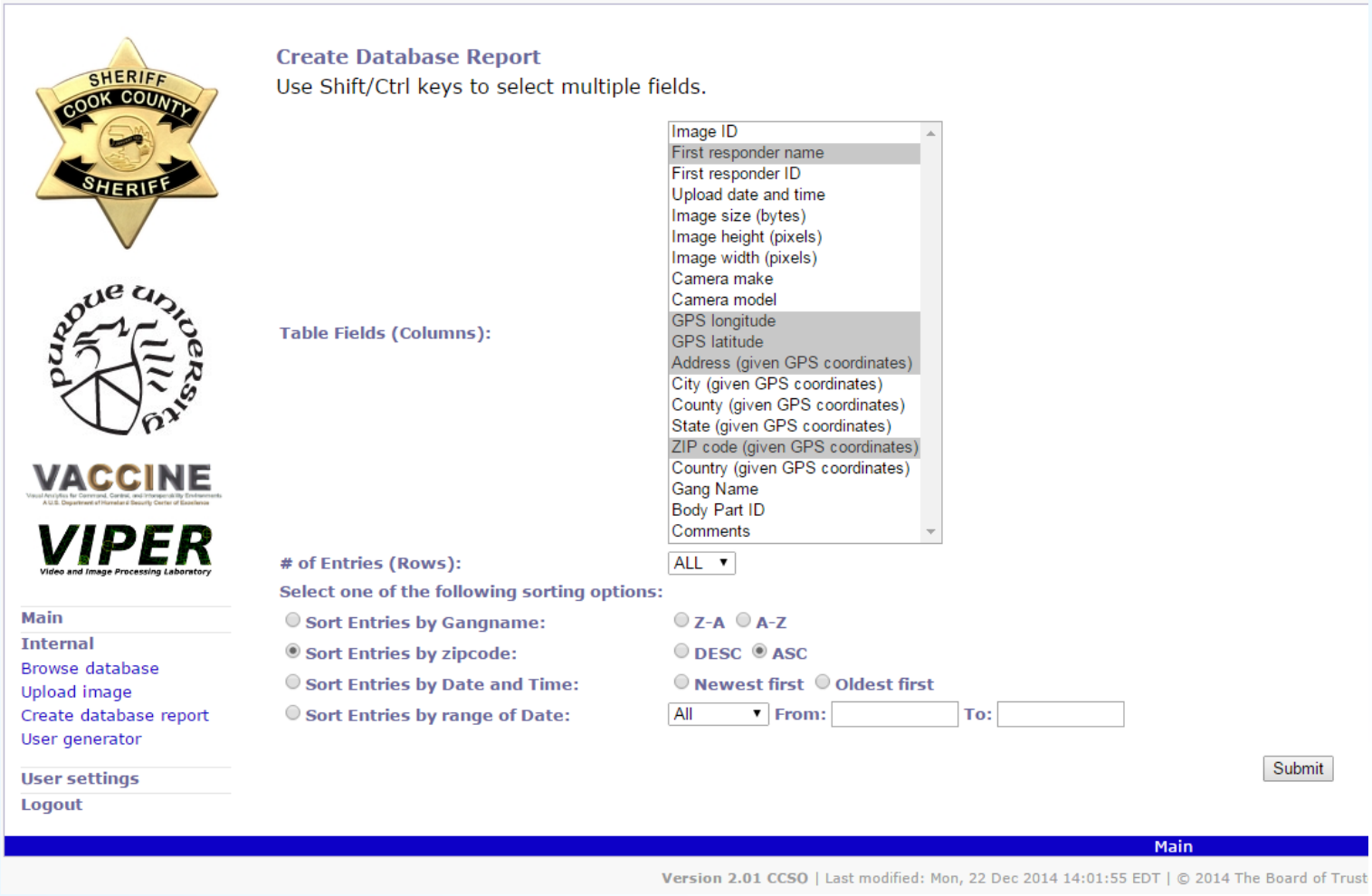
All images in radius

Browse Database

Desktop Version



Create Database Report



Create Database Report

zip firstrespo	filedatetimeuplo	gpslongitu	gpslatitude	address	zip
1101	5/16/2011 22:09	-77.034	38.904	1424 L St NW	20005
1101	5/16/2011 22:04	-86.885	40.43	1018 Roberts St	47904
1101	5/16/2011 22:04	-86.875	40.416	26 Grant St	47904
1101	5/16/2011 22:04	-86.865	40.41	2810 Kossuth St	47904
1101	5/16/2011 22:04	-122.41	37.786	3319 Longlois Dr	47904
1101	5/16/2011 22:04	-86.885	40.407	905 S 10th St	47905
1101	5/16/2011 22:04	-86.872	40.405	1141 S 22nd St	47905
1101	5/16/2011 22:04	-86.885	40.402	1000 Potomac Ave	47905
1101	5/16/2011 22:09	-86.846	40.422	3607 Driftwood Dr S	47905
1101	5/16/2011 22:04	-86.921	40.451	2601-2699 Trace 26	47906
1101	5/16/2011 22:04	-86.912	40.428	Memorial Mall	47906
1101	5/16/2011 22:04	-86.92	40.431	1165 W Stadium Ave	47906
1101	5/16/2011 22:04	-86.929	40.423	136-174 Nimitz Dr	47906
1101	5/16/2011 22:04	-86.936	40.418	2-98 Aviation Drive.	47906
1101	5/16/2011 22:04	-86.927	40.417	148 Arnold Dr	47906
1101	5/16/2011 22:09	-86.931	40.452	1390-1398 Palmer Dr	47906
1101	5/16/2011 22:09	-86.928	40.452	1300-1306 Palmer Dr	47906
1101	5/16/2011 22:09	-86.928	40.454	2562-2582 Neil Armst	47906
1101	5/16/2011 22:04	-86.908	40.433	812 Vine St	47906
1101	5/16/2011 22:04	-77.08	38.939	465 Northwestern Av	47906
1101	5/16/2011 22:09	-77.076	38.936	465 Northwestern Av	47906
1101	5/16/2011 22:04	-122.405	37.785	510 Hayes St	47906
1101	5/16/2011 22:04	-122.399	37.786	610 Purdue Mall	47906
1101	5/16/2011 22:04	-122.404	37.781	465 Northwestern Av	47906
1101	5/16/2011 22:09	-122.394	37.782	465 Northwestern Av	47906
1101	5/16/2011 22:04	0	0	422 Vine St	47906
1101	5/16/2011 22:04	0	0	773-799 Northwester	47906
1101	5/16/2011 22:04	0	0	1203 W State St	47906
1101	5/16/2011 22:09	0	0	465 Northwestern Av	47906
1101	5/16/2011 22:04	-86.919	40.424	S Russell Dr	47907
1112	12/12/2014 13:34	-86.9125	40.42934	501 Purdue Mall	47907
1112	12/12/2014 15:10	-86.9125	40.42933	501 Purdue Mall	47907
200069	8/7/2014 8:37	-87.9708	42.07957	213 South Dryden Pla	60004
200069	11/20/2014 9:53	-88.0135	42.15278	2400-2498 Crystal Lan	60004

Gang Graffiti – Image Analysis



Light Sensor Value

Blur Prevention and Detection

Input Image

1

Color Correction

2

Touchscreen Tracing

3

Gaussian Color Segmentation

4

Segmentation Enhancement

5

Background Stripe Removal

6

Graffiti Component Reconnection

7

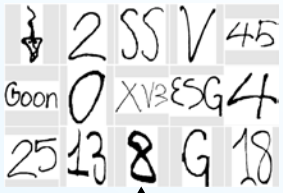
Graffiti Component Extraction

8

Content-Based Image Retrieval

9

Gang Graffiti



Gang Graffiti – Color Correction

- **Allows color correction without fiducial marker**
- **Non-intrusive**



- **Automatically assigns color correction matrix to light sensor value**



Before





After

Gang Graffiti – Color Segmentation

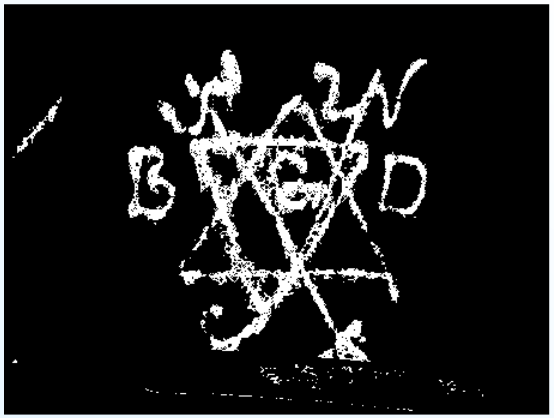




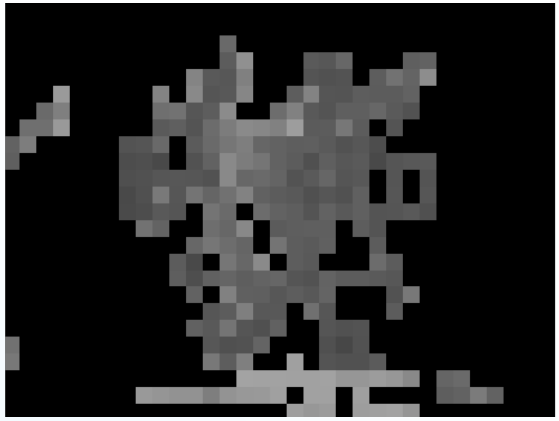
Gang Graffiti – Segmentation Enhancement



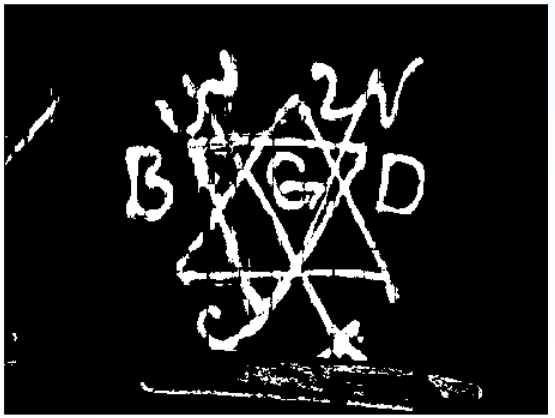
Original Image



Initial Segmentation

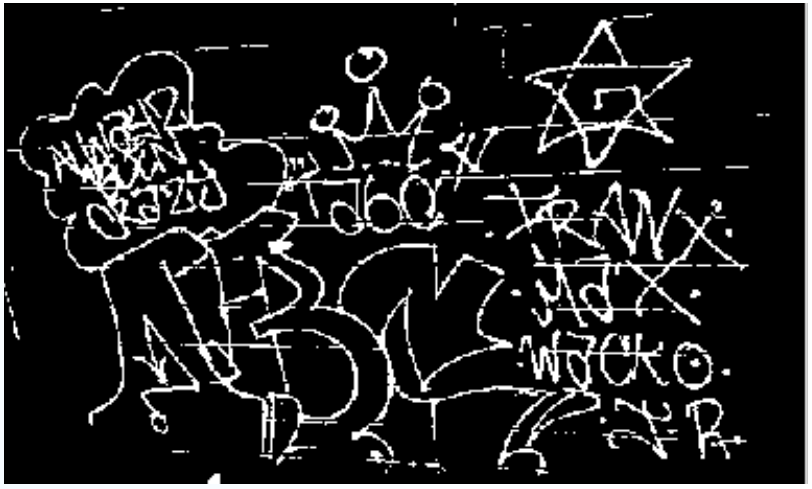


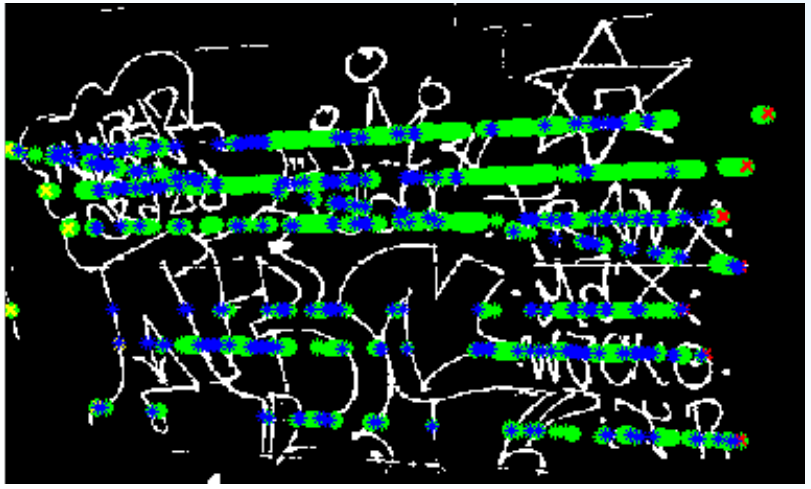
Block-Wise Luma

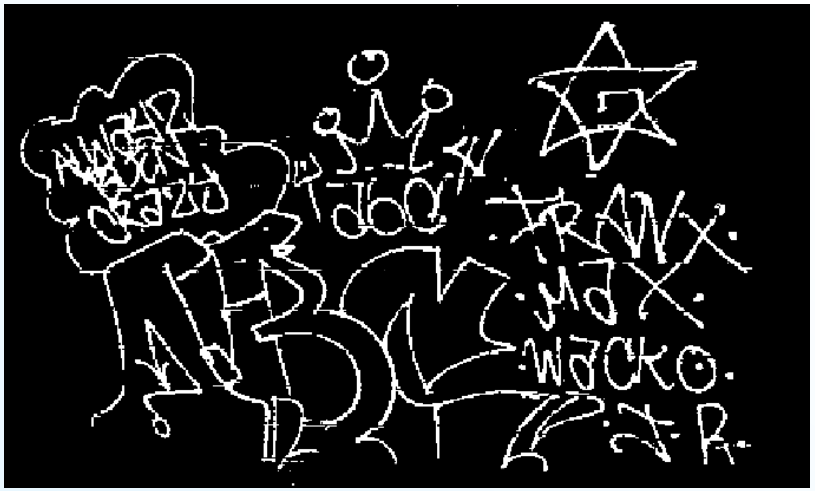


Enhanced

Gang Graffiti – Stripe Removal

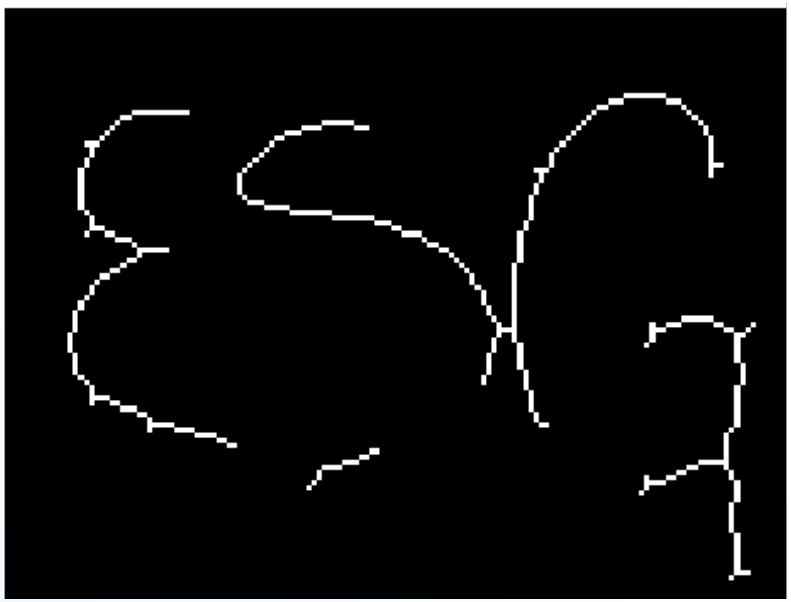




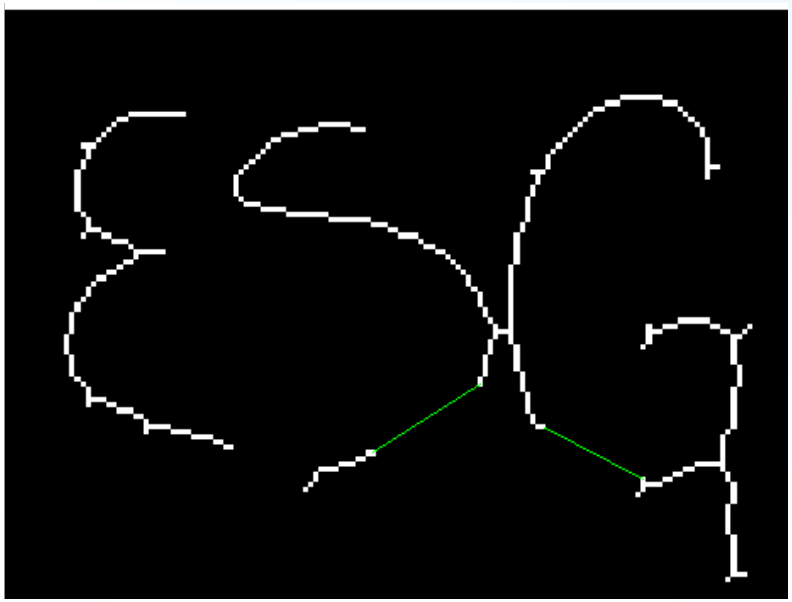


Segments to be removed (green) and ignored (blue)

Gang Graffiti – Component Reconnection



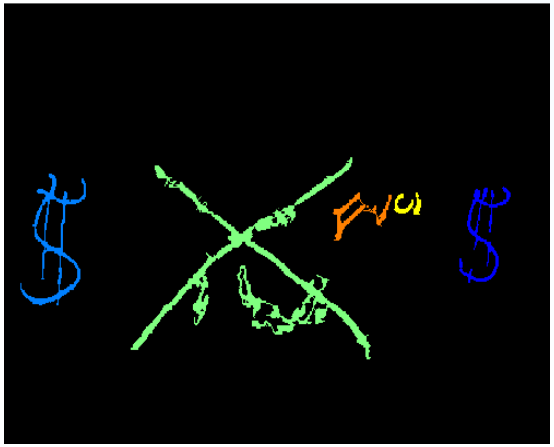
Before

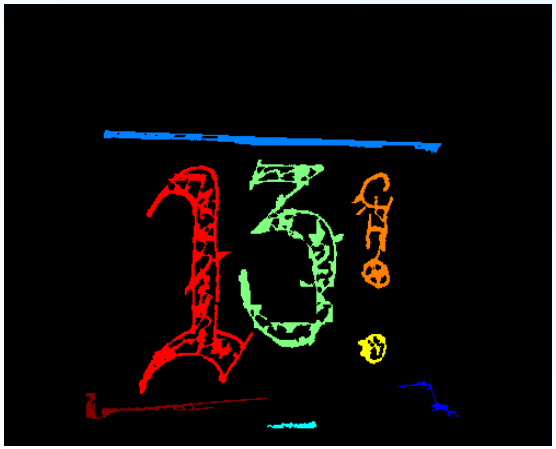


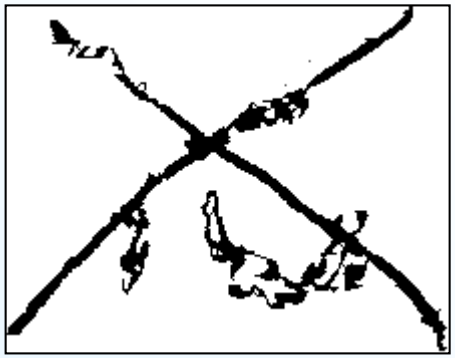
After

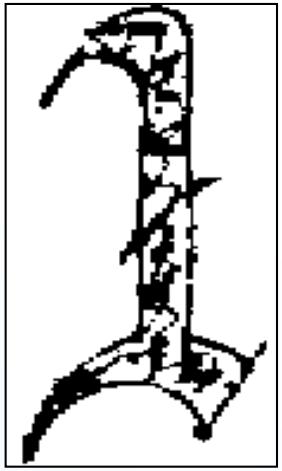
Gang Graffiti – Component Extraction









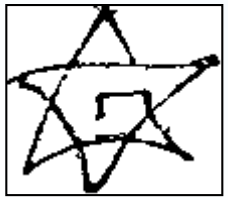




Gang Graffiti – Component Classifcation







Sybmol: 6-Point Star

Confidence: 84%



Symbol: Crown

Confidence: 80%



Symbol: Pitchfork

Confidence: 75%



Symbol: Unknown



Symbol: Unknown

Gang Graffiti – Find Similar Images



Retrieval: Accuracy: 99.10%, Time: 70 ms

Gang Tattoo – Image Analysis



Original Image



Cropped Image



Image Analysis

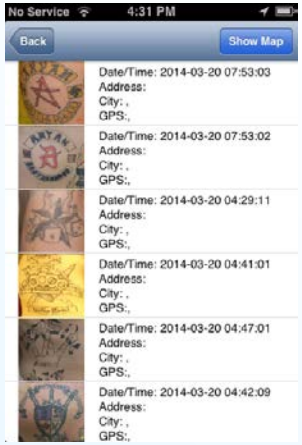
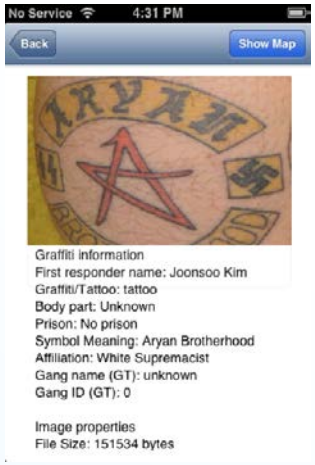


Image Matching



Analysis Results

Graffiti information
First responder name: Joonsoo Kim
Graffiti/Tattoo: tattoo
Body part: Unknown
Prison: No prison
Symbol Meaning: Aryan Brotherhood
Affiliation: White Supremacist
Gang name (GT): unknown
Gang ID (GT): 0

Gang Tattoo – Find Similar Images



Tattoo Recognition Technology – Challenge (Tatt-C)

Tattoo Identification And Region of Interest

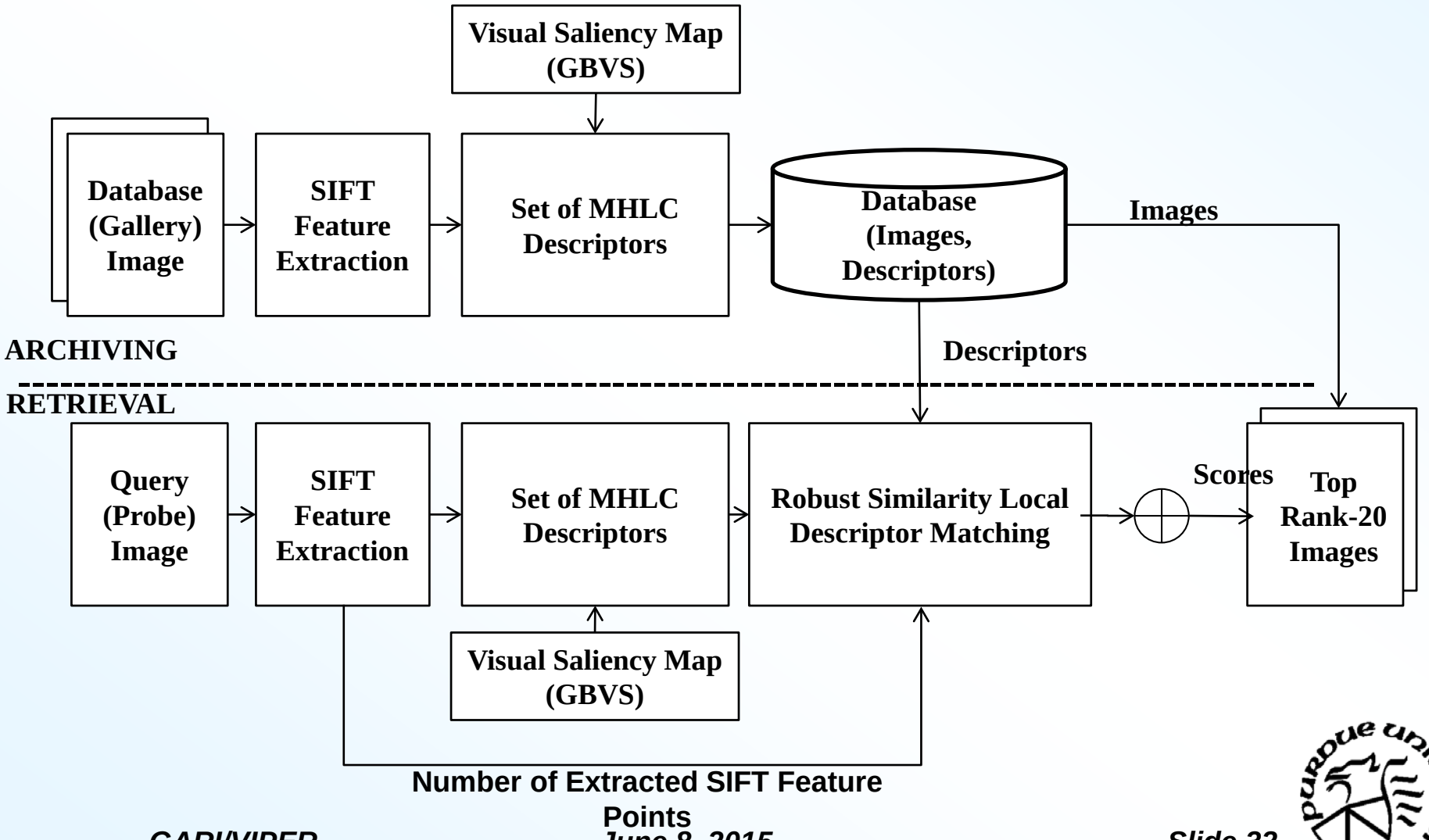
Introduction for Recognition Technology-Challenge

- **The Tattoo Recognition Technology-Challenge (Tatt-C) : a competition to test image-based tattoo matching technology. This challenge will assess the capability of tattoo recognition methods to detect and retrieve tattoos**

- **We have participated in**
  - ***Tattoo Identification* : matching different instances of the same tattoo images from the same subject over time**
  - ***Region of Interest* : matching a small region of interest that is contained in a larger image**

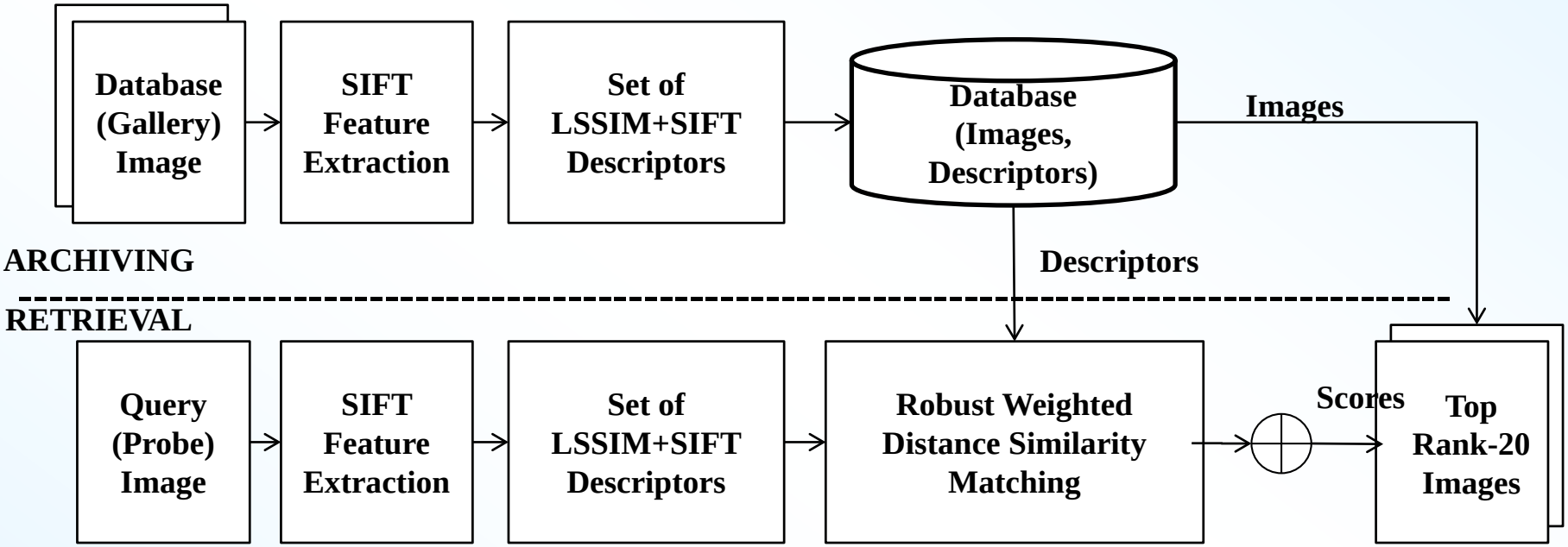
Tattoo Identification

- **Block Diagram**



Region of Interest

- **Block Diagram**



LSSIM : Local Self-SIMilarities

Experimental Results

- **Definition of probe, gallery, and background image**
  - **Probe image: A query image used to search against the dataset – ideally find the true matches for a probe image**
  - **Gallery: A set of images that will be searched against to find the matches for the probe image**
  - **Background image: Irrelevant images that can be added to the gallery to test the robustness of the method – "distractor"**

Experimental Results

Dataset Information

	Tattoo Identification	Region of Interest
Number of gallery images	215	157
Number of probe images	157	297
Number of background images (“Distractor”)	4332	4332

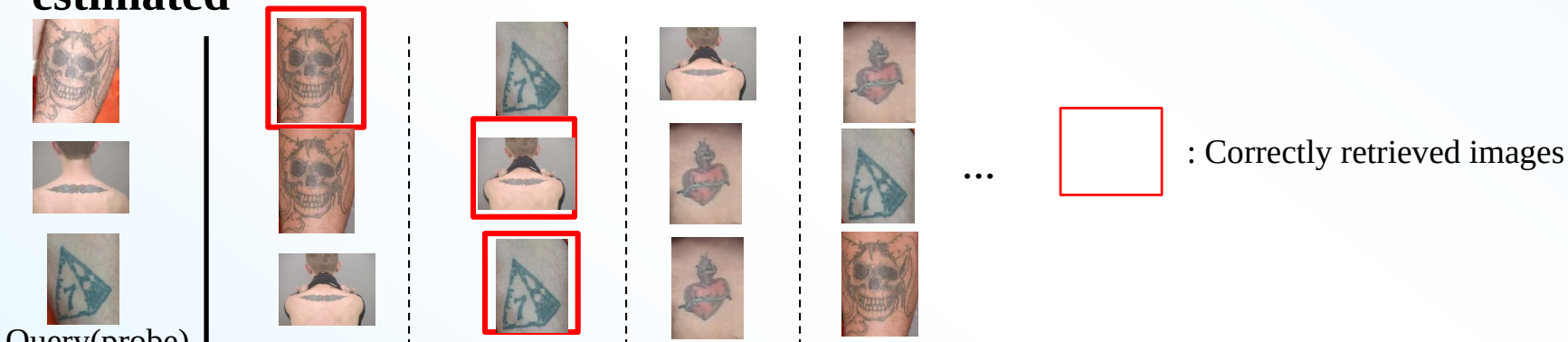
In some experiments the gallery and background are merged

Experimental Results

- **Performance Metrics**
  - **Cumulative Match Characteristic (CMC)**
    - **CMC = The probability that one or more correctly matched images for a probe is observed within the top K ranks**

Experimental Results

- **Cumulative Match Characteristic (CMC):**
  - **CMC(k) = sum of P(r) where r <=k**
- **Rank Probability Mass Function P(r) = at rank r, the number of correctly matched probe images in the candidate list divided by total number of probe images**
- **Given a database and set of queries (probes), this probability can be estimated**



P(r) :	1/ 3	2/ 3	0	0
CMC(k) :	1/ 3	1/3 + 2/ 3 = 1	1	1

Experimental Results

5-fold Cross Validation

- **We conducted 5 separate experiment using leave-one-out-cross validation**
- **In each experiment, 4 of subset were combined to form a training set, with the 5^{th} subset used for testing**
- **Using this scheme, we determined the weight between SIFT descriptor and MHLC descriptor in tattoo identification and the weight between SIFT descriptor and self similarity descriptor in region of interest**

Experimental Results (Tattoo Identification)

- **Example of Matched Results (Without Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched

Probe Image probe_004.jpg





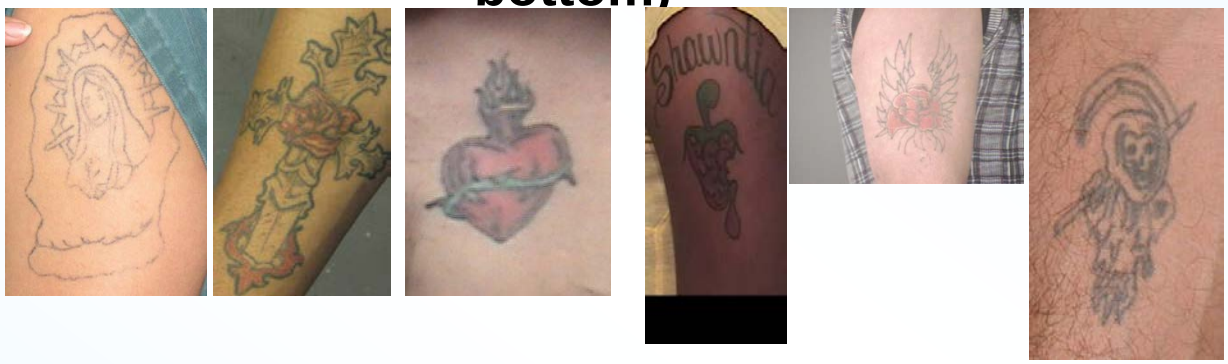
Experimental Results (Tattoo Identification)

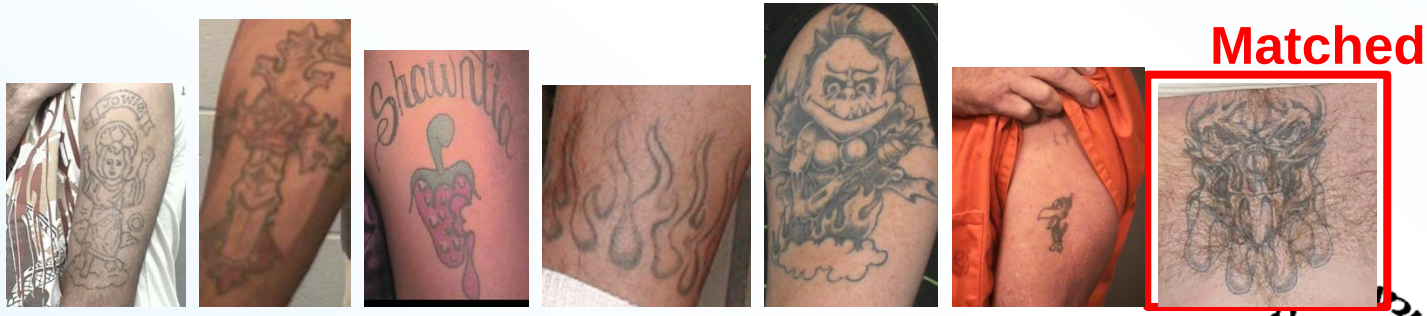
- **Example of Matched Results (Without Background Images)**

Probe Image probe_053.jpg



Top 13 ranks (left to right, top to bottom)





Matched

Experimental Results (Tattoo Identification)

- **Example of Matched Results (Without Background Images)**

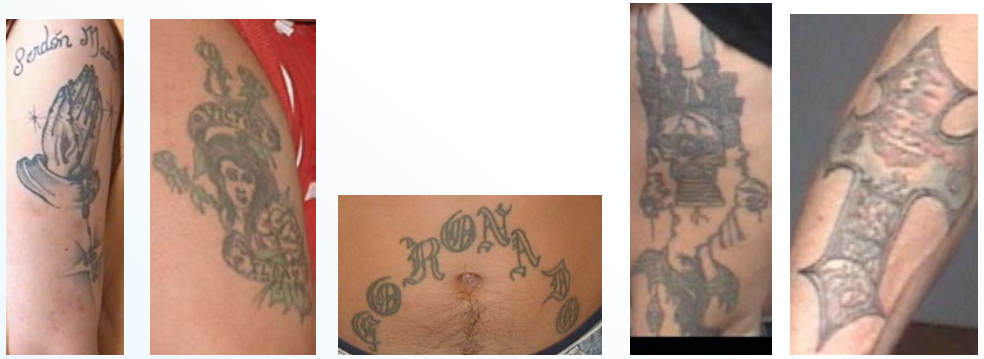
Probe Image probe_056.jpg



Top 10 ranks (left to right, top to bottom)

Matched





Experimental Results (Tattoo Identification)

- **Example of Matched Results (With Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched

Probe Image probe_004.jpg

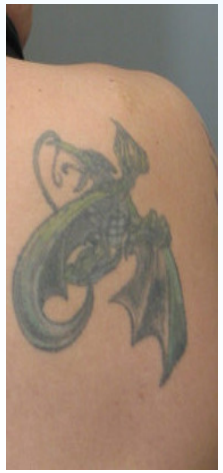


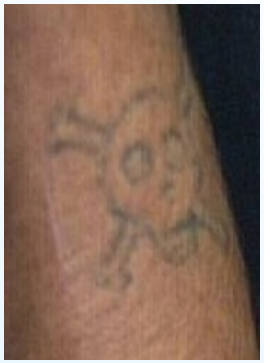




















Experimental Results (Tattoo Identification)

- **Example of Matched Results (With Background Images)**

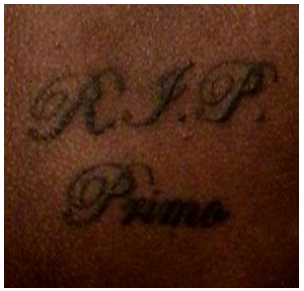
Top 10 ranks (left to right, top to bottom)

No Match

Probe Image probe_053.jpg











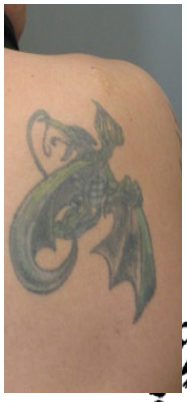












Experimental Results (Tattoo Identification)

- **Example of Matched Results (With Background Images)**

Top 10 ranks (left to right, top to bottor

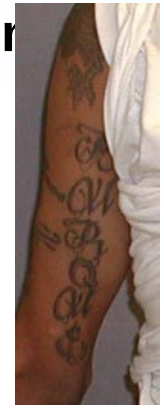
Probe Image probe_056.jpg

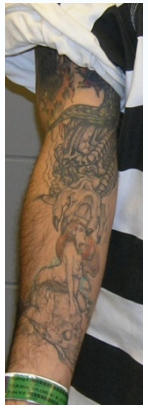


Matched



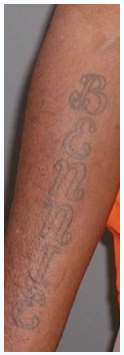


















Experimental Results (Tattoo Identification)

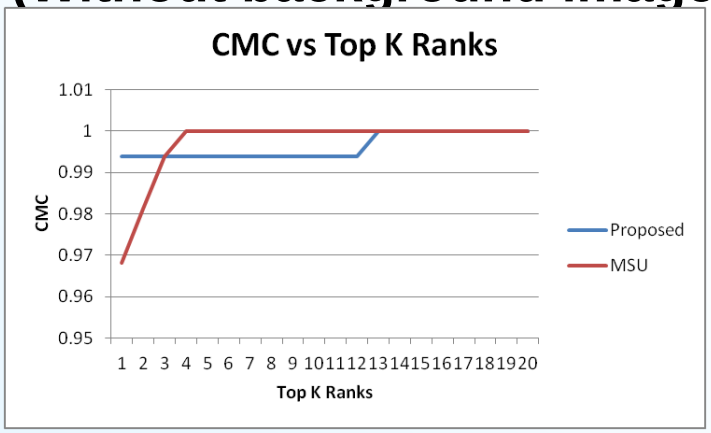
- Comparison Between Different Methods

Top rank	Proposed	MSU
1	0.99376	0.96814
3	0.99376	0.99376
5	0.99376	1
10	0.99376	1
15	1	1
20	1	1

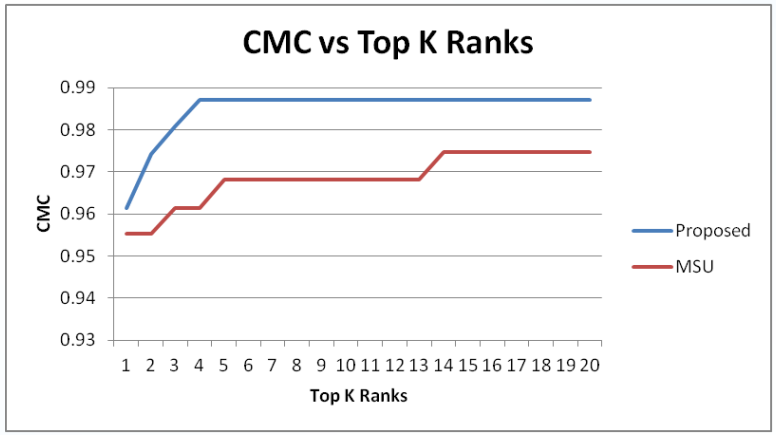
CMC vs Top K rank Table (K=1,3,5,10,15,20) (Without background images)

Top rank	Proposed	MSU
1	0.96146	0.95522
3	0.98086	0.96146
5	0.9871	0.96814
10	0.9871	0.96814
15	0.9871	0.9748
20	0.9871	0.9748

CMC vs Top K rank Table (K=1,3,5,10,15,20) (With background images)



CMC vs Top K rank Plot (Without background images)

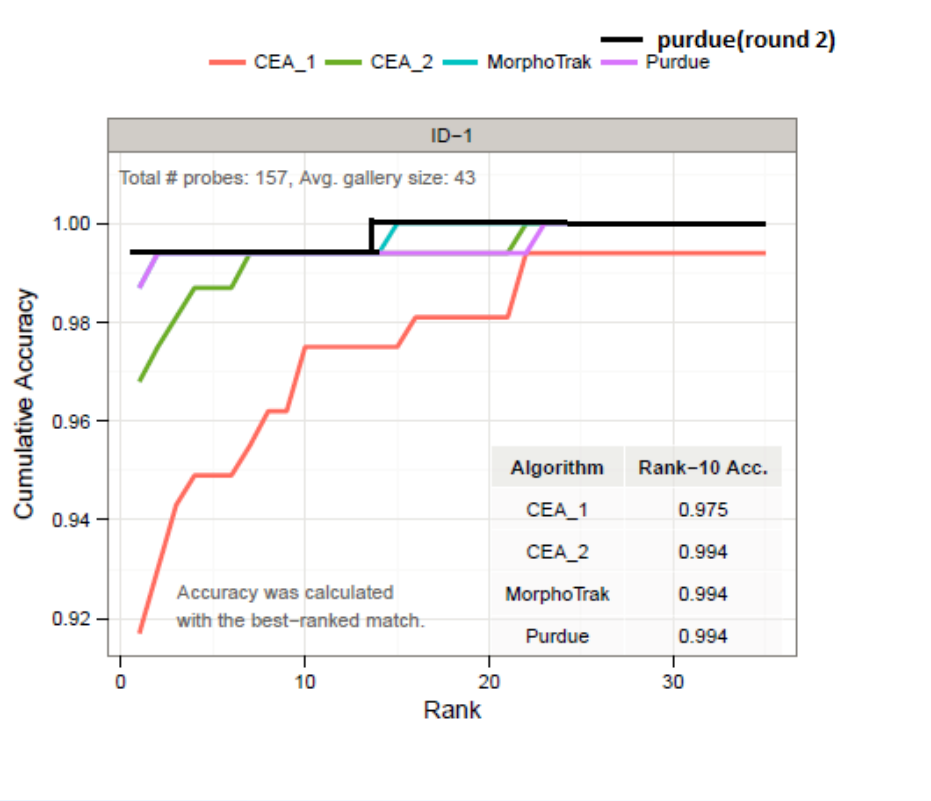


CMC vs Top K rank Plot (With background images)

Experimental Results (Tattoo Identification)

- **Comparison with Phase 1 results**

(Without Background Images)

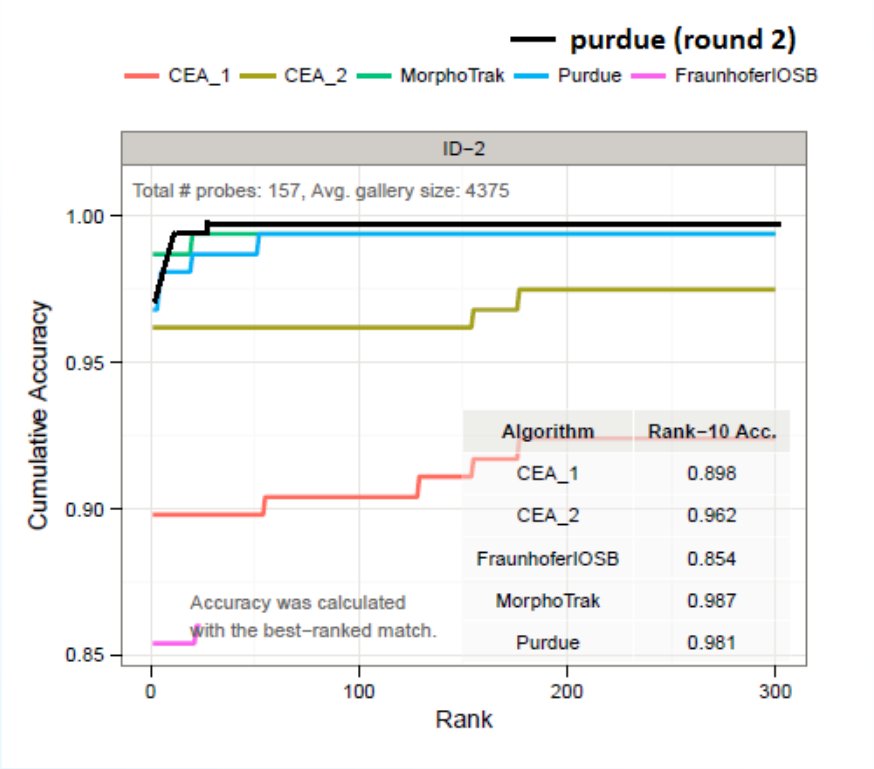


Purdue : 0.994

Experimental Results (Tattoo Identification)

- **Comparison with Phase 1 results**

(With Background Images)



Purdue : 0.987

Experimental Results (Region of Interest)

- **Example of Matched Results (Without Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_001.jpg





Experimental Results (Region of Interest)

- **Example of Matched Results (Without Background Images)**

Top 10 ranks (left to right, top to bottom)

Probe Image
probe_002.jpg





Experimental Results (Region of Interest)

- **Example of Matched Results (Without Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_0013.jpg





Experimental Results (Region of Interest)

- **Example of Matched Results (With Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_001.jpg





Experimental Results (Region of Interest)

- **Example of Matched Results (With Background Images)**

Top 10 ranks (left to right, top to bottom)

No Matching…

Probe Image
probe_002.jpg







Experimental Results (Region of Interest)

- **Example of Matched Results (With Background Images)**

Top 10 ranks (left to right, top to bottom)

Matched!!

Probe Image
probe_013.jpg





Experimental Results (Region of Interest)

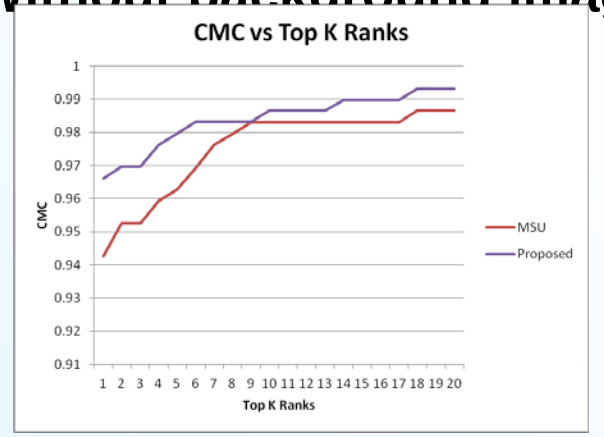
- Comparison Between Different Methods

Top rank	Proposed	MSU
1	0.96614	0.94262
3	0.96964	0.95262
5	0.97962	0.96262
10	0.98648	0.98296
15	0.98982	0.98296
20	0.99316	0.98648

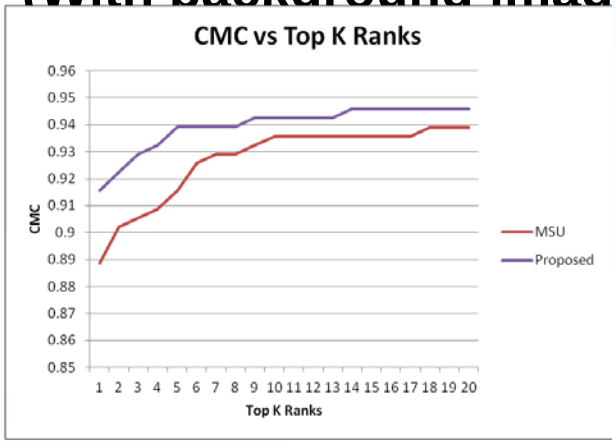
CMC vs Top K rank Table (K=1,3,5,6,10,15,20) (Without background images)

Top rank	Proposed	MSU
1	0.9156	0.88858
3	0.92912	0.90542
5	0.93914	0.9156
10	0.94246	0.9358
15	0.94596	0.9358
20	0.94596	0.93912

CMC vs Top K rank Table (K=1,3,5,6,10,15,20) (With background images)



CMC vs Top K rank Plot (Without background images)

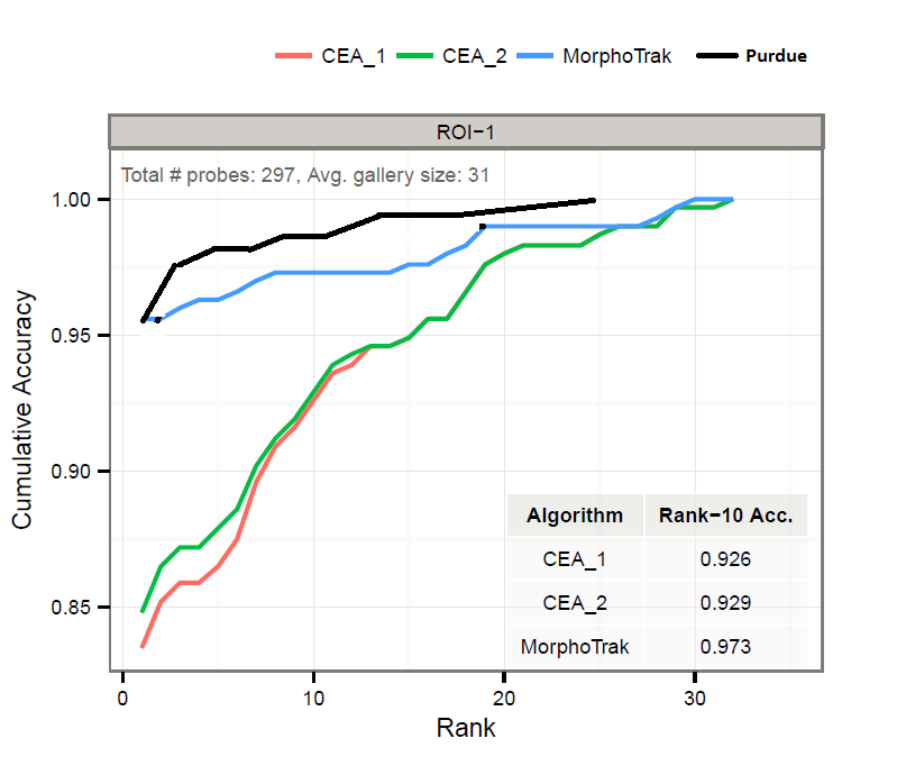


CMC vs Top K rank Plot (With background images)

Experimental Results (Region of Interest)

- **Comparison with Phase 1 results**

(Without Background Images)

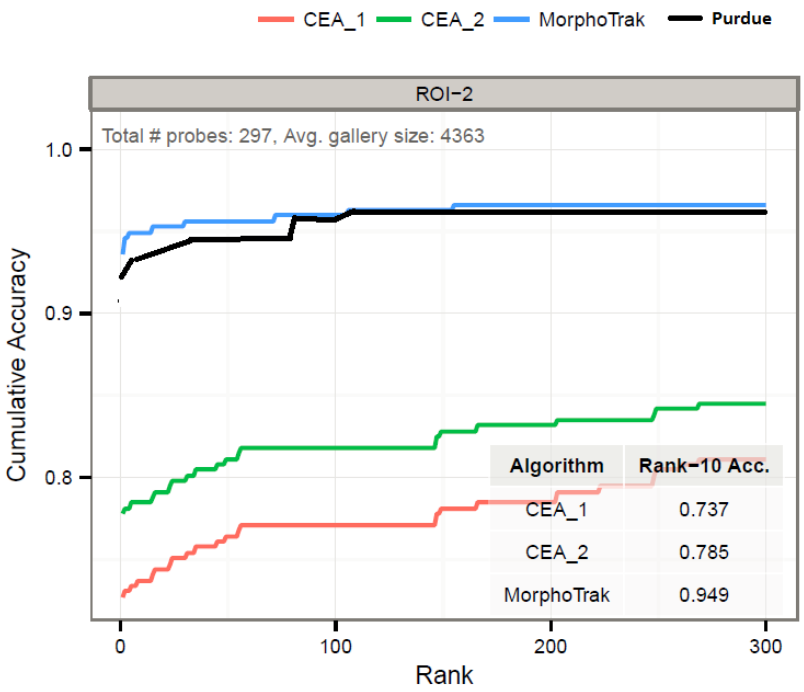


Purdue : 0.983

Experimental Results (Region of Interest)

- **Comparison with Phase 1 results**

(Without Background Images)



Purdue : 0.942

Issues and Problems

- **How do the methods scale?**
- **How much domain specific/location specific information is needed – it not not just an image analysis problem**
- **User interface!**
- **How does an graffiti or tattoo image analysis system fit into existing law enforcement work flow**

GARI Availability

- **The GARI system is available free to any law enforcement agency**
  - **– MOU required**
  - **– Hosting of the GARI server**

For more information: gari@ecn.purdue.edu

Video-Based Re-Identification of Subjects for The Greater Cleveland Regional Transit Authority

eBOLO

These slides supersede the slides dated March 6, 2015













We acknowledge support of eBOLO from the U.S. Department of Homeland Security and the DHS Centers of Excellence

eBOLO Goals

- Demonstrate a proof of concept of a surveillance system that would allow GCTRA to identify persons of interest from a video and determine if and when they re-appear at a later date

- Add a subject to the eBOLO list by manually highlighting them; use computer vision and image processing techniques to generate the “eBOLO fingerprint”

eBOLO Goals

- Use eBOLO fingerprint to continuously monitor the video surveillance system to determine when that subject reappears in the video => flag or alarm GCRTA that the subject has reappeared and provide location details

- Provide recommendations to GCTRA with respect to infrastructure improvements (e.g., camera types, camera placement, background networking and data storage) so that an eBOLO system would be more effective

eBOLO Research Team

Octavia I. Camps

- **Mengran Gou, Tom Hebble, Sadjad Asgari Esfeden**

Edward J. Delp

- **Khalid Tahboub, Blanca Delgado**

Patrick J. Flynn

- **Joel Brogan, Nikita Amelchenko, Amy Li**

Yun Fu

- **Ming Shao**

Richard J. Radke

- **Srikrishna Karanam, Yang Li, Gyanendra Sharma**

Body Matching







Octavia I. Camps

- **Mengran Gou, Tom Hebble, Sadjad Asgari Esfeden**

Edward J. Delp

- **Khalid Tahboub, Blanca Delgado**

Richard J. Radke

- **Srikrishna Karanam, Yang Li, Gyanendra Sharma**









Face Recognition





Patrick J. Flynn, Kevin Bowyer

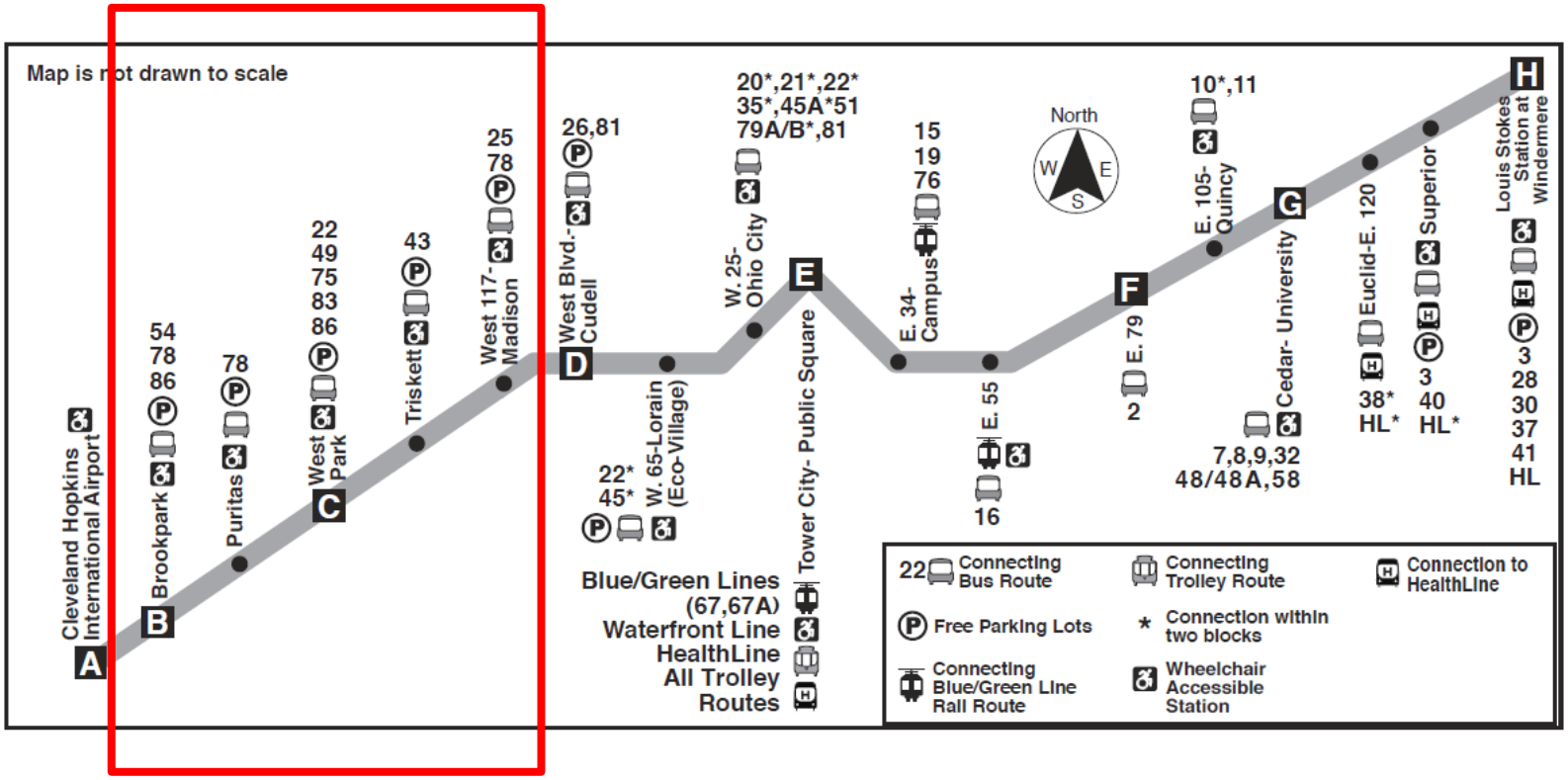
- **Joel Brogan, Amy Li, Nikita Amelchenko**

Yun Fu

- **Ming Shao**

Dataset Collection

- In October 2014, the team conducted experiments and collected extensive data from 5 stations: Brookpark, Puritas, West Park, Triskett, and West 117
- A total of **266** hours of footage from **9** cameras was collected



Camera Fields of View

BrookPark
Bus Wait



BrookPark
Tunnel 1



BrookPark
Tunnel 2



Puritas
Tunnel



Triskett
Bridge



Triskett
Main entrance door



West 117th
Lobby west



West 117th
Tunnel



West Park
Tunnel



Resolutions and Frame Rate

Station	Resolution	Max Resolution	Frame Rate (fps)	Max Frame Rate (fps)
Brookpark	352 x 240	720 x 480	3.75	30
Puritas	640 x 480	N/A	4	30
West Park	352 x 240	N/A	17	30
Triskett	352 x 240	720 x 480	3.75	30
W117	352 x 240	720 x 480	3	30

- Due to system constraints, the maximum resolution and frame rate for each camera is rarely achievable

Ground Truth Generation











- Students manually outlined re-appearances of subjects from the data collection- this created what is known as “ground truth” data – this data will be used for training and validation of the analysis methods
- The sizes and appearances of the subjects vary considerably
- Several “distractor” pedestrian bounding boxes were also generated



Camera Locations For Ground Truth Data

- Two views were selected for ground truthing:
  - Triskett Bridge
  - West117 Tunnel

- These are camera locations with significant illumination changes and a clear view of people as they enter/leave the camera field of view

- The selected segments are for different times of the day over two separate days

Camera Locations For Ground Truth Data

- Summary of the videos ground truthed for the two cameras:

West117 Tunnel

Video File Name	Date	Start Time	End time
117_tunnel_10a-11a	16-Oct-14	10:00 AM	11:00 AM
117_tunnel_5p-6p	16-Oct-14	5:00 PM	6:00 PM
117_tunnel_oct17_1145a-1245p	17-Oct-14	11:45 AM	12:45 PM

Triskett Bridge

Video File Name	Date	Start Time	End time
tsktt_bridge_oct17_1140a-1240p	17-Oct-14	11:40 AM	12:20 PM
tskt_bridge_1030a-1130a	16-Oct-14	5:30 PM	6:30 PM
tskt_bridge_1031a-1131a	16-Oct-14	10:30 AM	11:30 AM

Ground Truth Data

- The ground truth data was organized in terms of appearances - an appearance is a video segment outlining one person as he or she enters the camera field of view and walks/loiters until he/she leaves the camera field of view
- Multi-shot images outlining the person during the entire time he/she is within the camera field of view are generated and stored as part of the ground truth data
- An example of Ground Truth data- single subject appearance at West177-tunnel:



Ground Truth Data Organization

- Ground truth data was divided into training and testing sets
- Testing data is used for method evaluation
- Training data is used to tune method parameters
- Summary of testing data:

Triskett-Bridge

Student 1	16 appearances
Student 2	28 appearances
Student 3	27 appearances
Student 4	18 appearances
Student 5	16 appearances
Student 6	16 appearances
Non-students	60 appearances

West177-Tunnel

Student 1	13 appearances
Student 2	19 appearances
Student 3	20 appearances
Student 4	9 appearances
Student 5	11 appearances
Student 6	12 appearances
Non-students	36 appearances

- Summary of training data:

Triskett-Bridge

Student 7	28 appearances
Non-students	7 appearances

West177-Tunnel

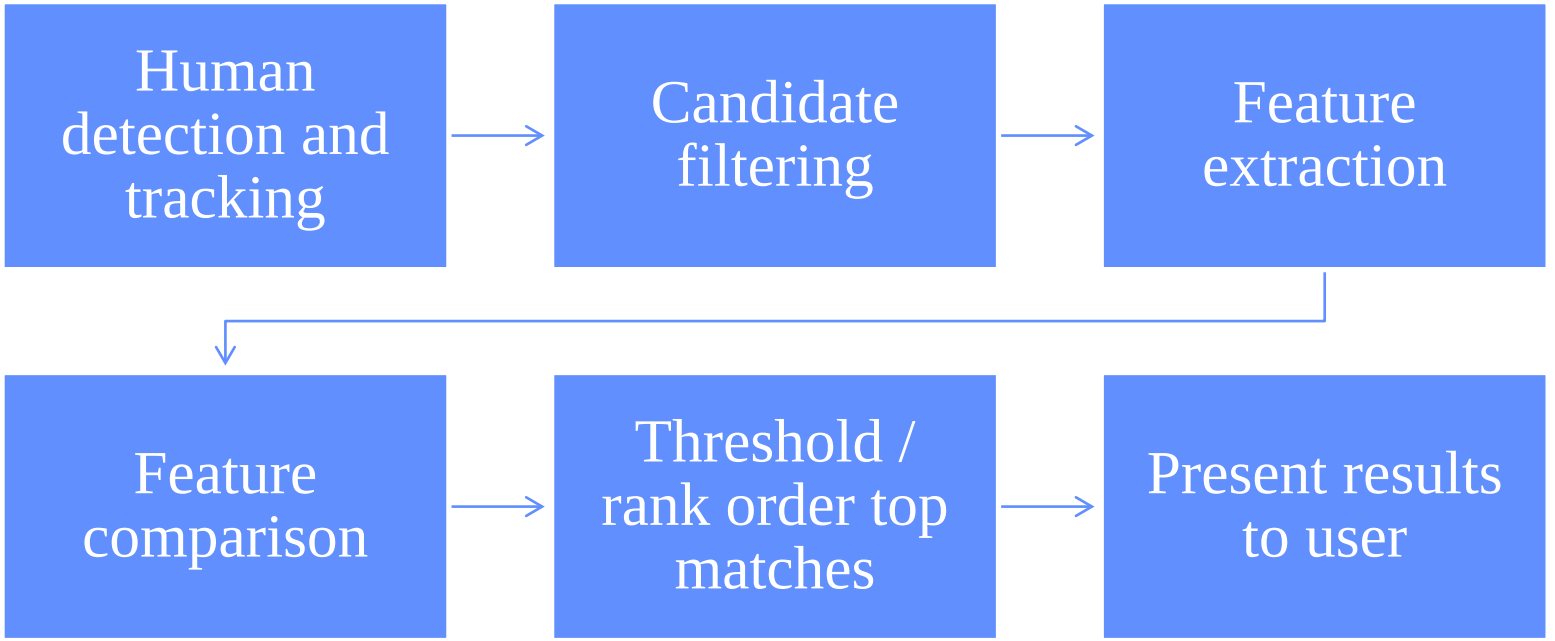
Student 7	19 appearances
Non-students	11 appearances

Body Matching Re-Identification Pipeline

Target specification:



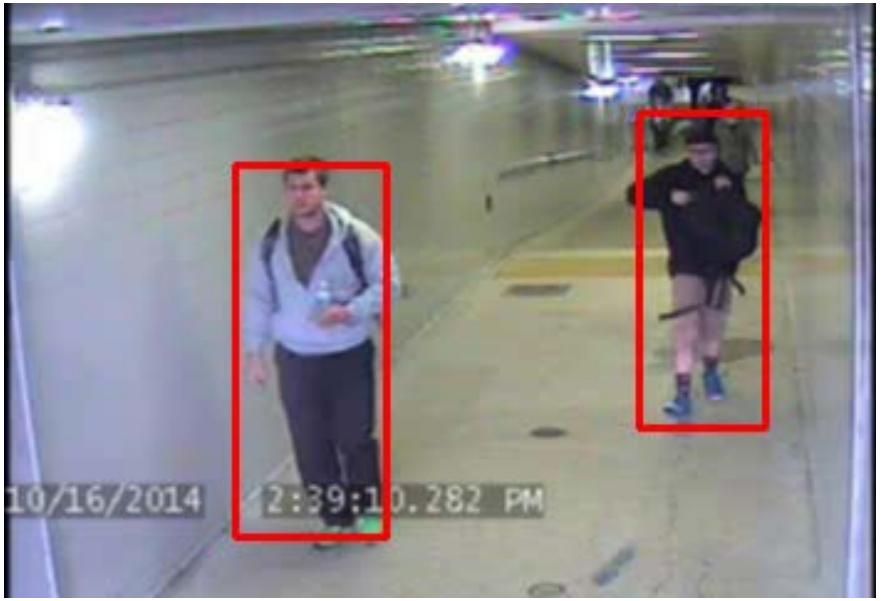
Target re-identification:



Human Detection and Tracking

- Foreground detection to get regions of interest; mixture of Gaussians
- Person detection based on improvements to aggregated channel features (Dollar et al. 2014) in each region of interest
- Algorithm is based on sliding windows, multi-scale feature pyramids, and boosted decision trees





Human Detection

- MOG Background Subtraction allows for reduced search space
- HOG and HAAR Cascades used to detect bodies within search space



Triskett Bridge busy scene (right) and MOG areas of foreground (left, in white)

Color Features

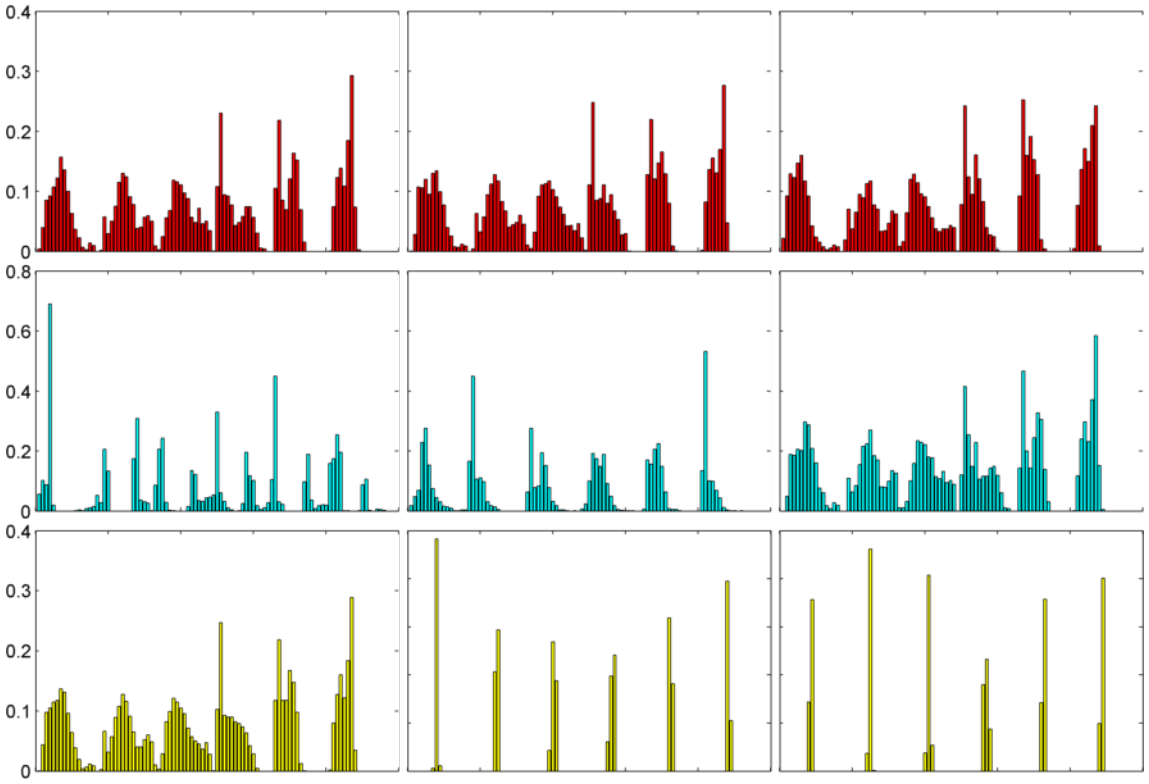
Each target and candidate is divided into several horizontal strips. Features are extracted from each strip based on the distribution of colors



RGB

HSV

YCbCr

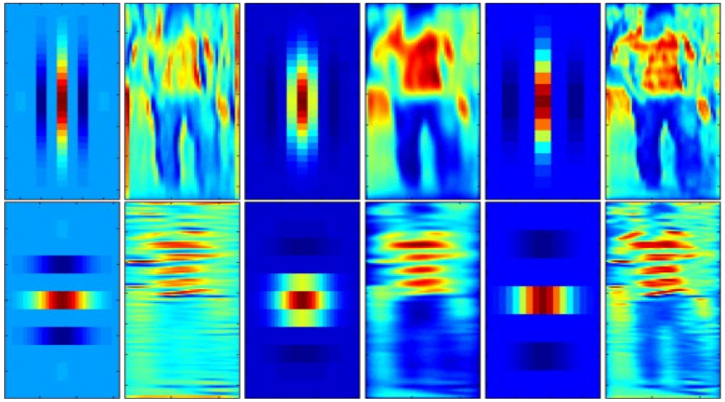


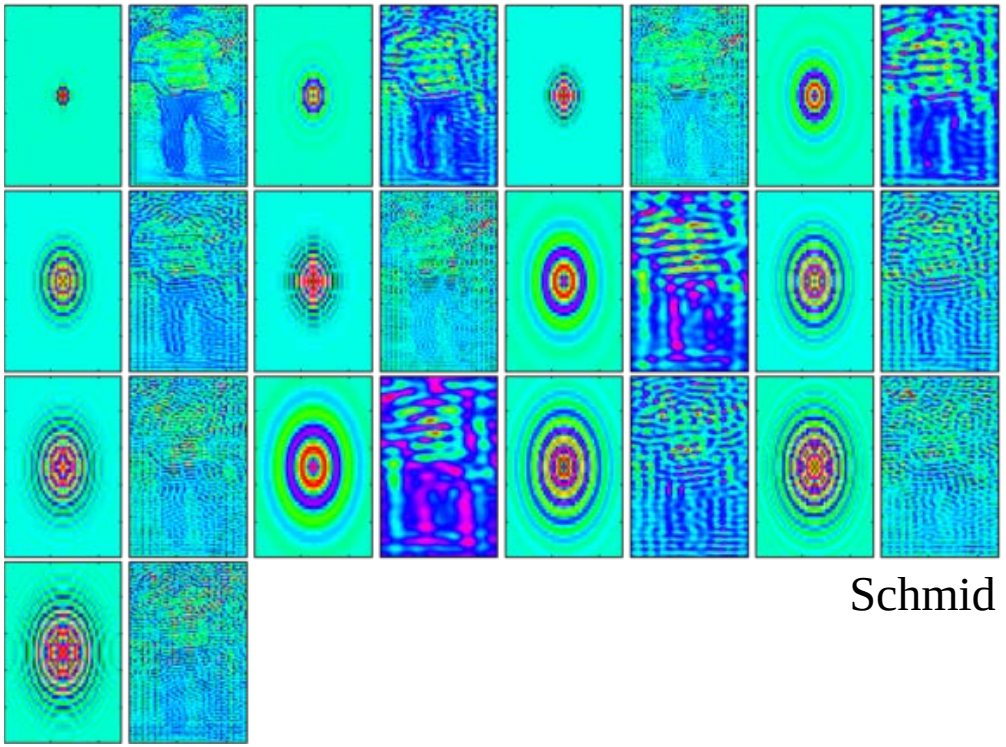
Texture Features

Texture features based on the patterns of intensity in each strip are also calculated



Gabor



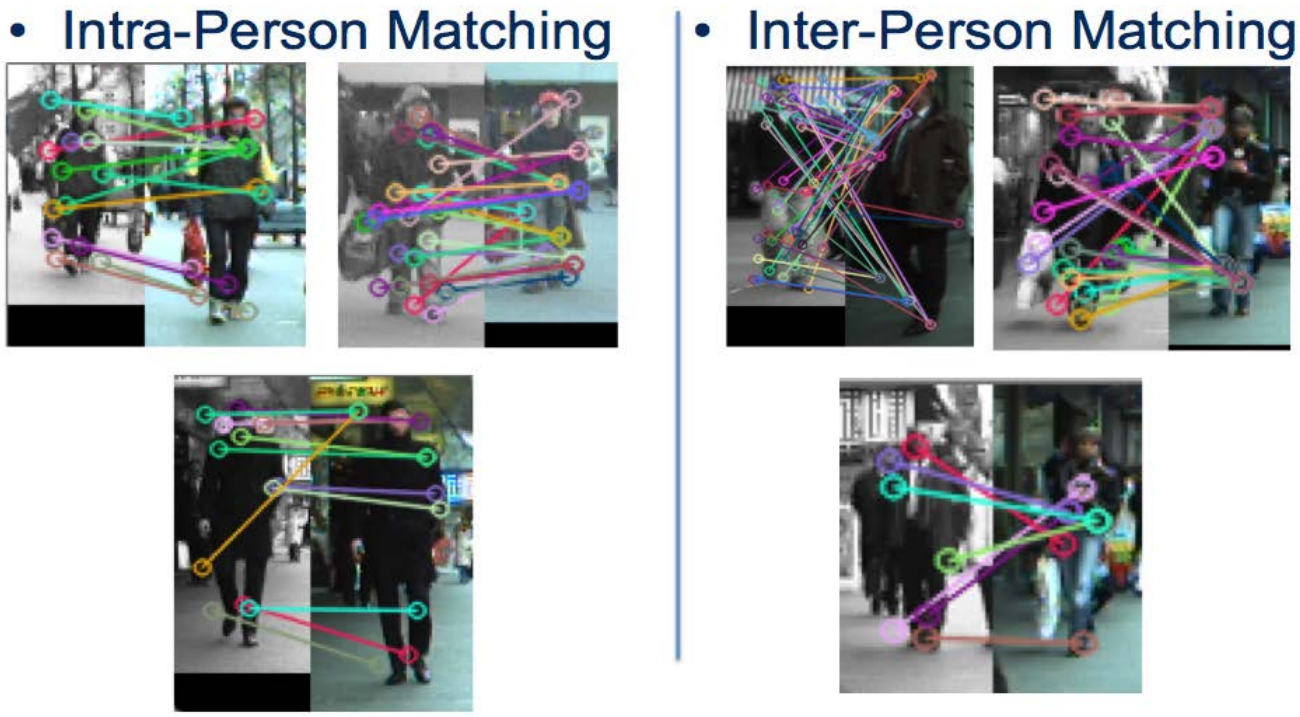


Schmid

We are currently investigating the combinations of color and texture feature that yield the best performance

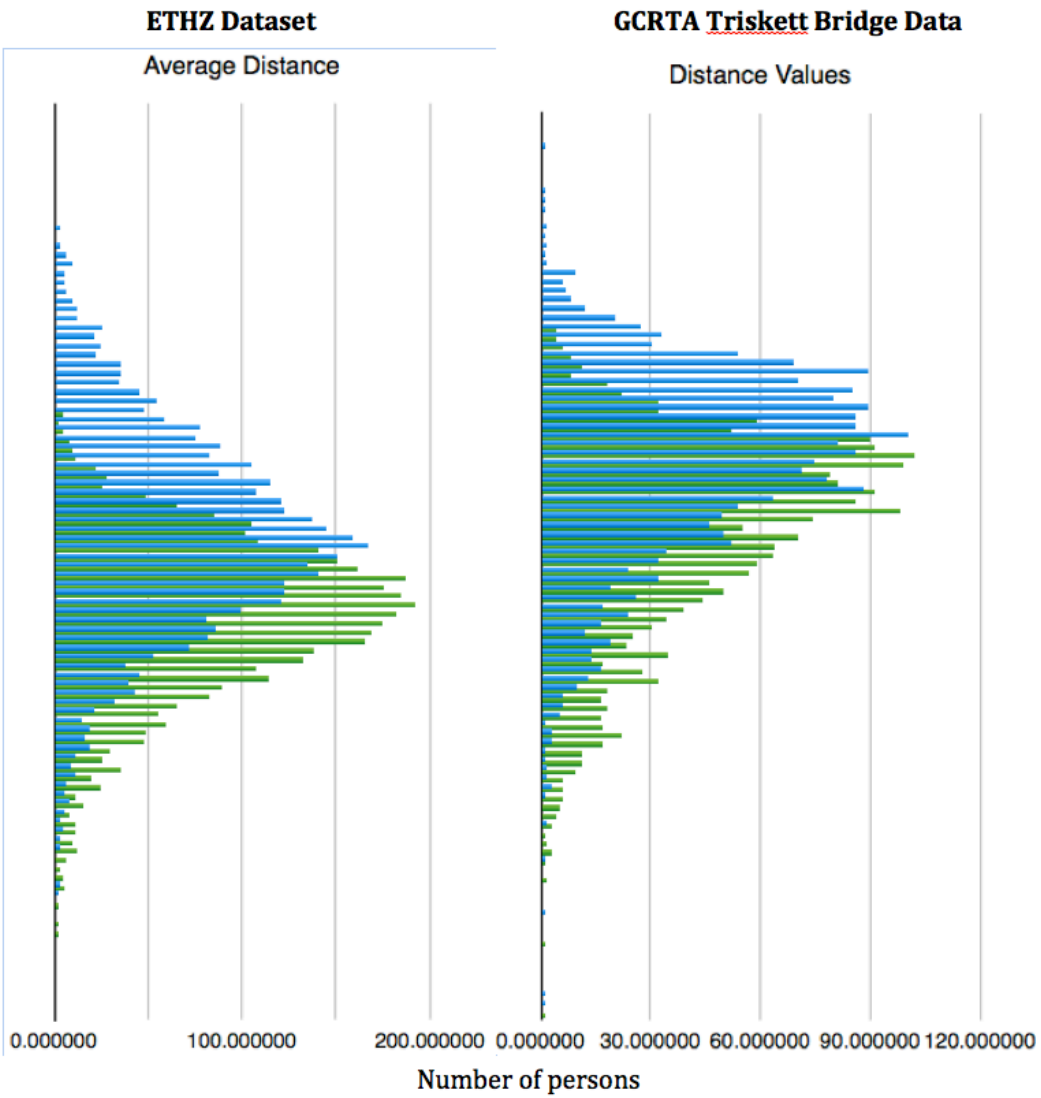
SURF Features

- Spectral analysis performed using SURF features on ETHZ dataset and annotated version of Triskett Bridge video (matching descriptor Cartesian distances ("Global Distance")



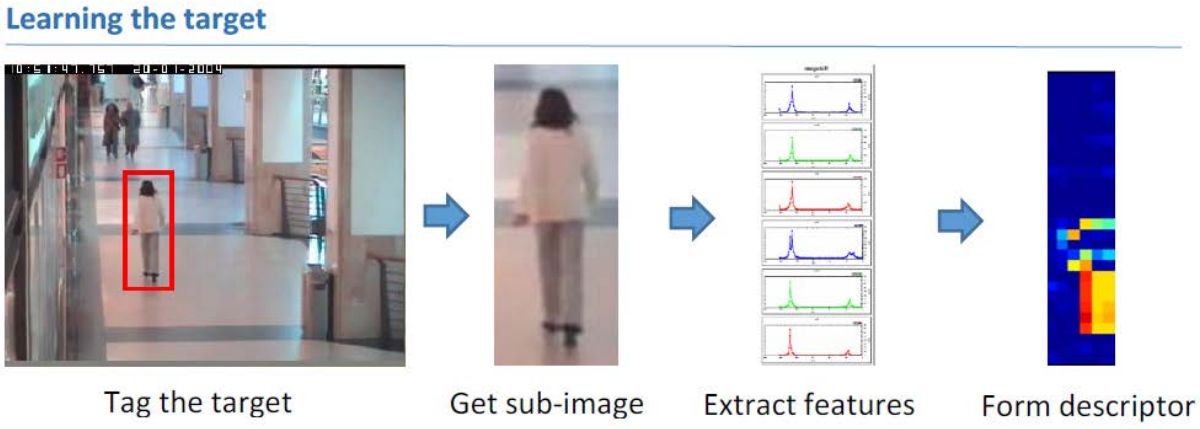
Examples of Intra-Person (true) matching, and Inter-Person (false matching) on pedestrians

SURF Feature - Body Analysis



Histograms of intra-person match distances (green) vs. inter-person match distances (blue).

Re-Identification Method



- We are investigating re-id methods including RankSVM (Prosser et al. 2010) and LFDA (Pedagadi 2013)
- In each case, offline training is slow, but online comparison is instantaneous

Metric Learning Re-Identification



Image pairs from
the same person

Classifier

Pairwise
Constraints



Image pairs from
different people

Discriminative Content Using Adaptive Boosting

- Find localized patches of pixels that contain discriminative content for a specific target
- Color features are extracted within these confined patches of pixels
- Build the classifier to look for these features

‘Target’



‘Non-Targets’



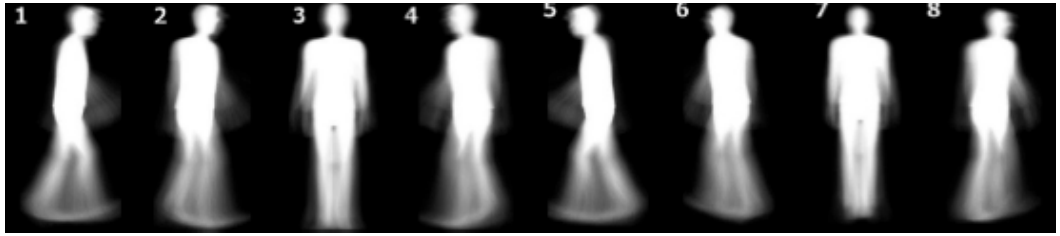
Recognition Of Person-Based Attributes

- Identify the target based on its attributes - backpacks, hats, eyewear, ...

- The main attributes we are investigating:
  - carried objects such as backpacks and other types of luggage
  - irregularities in the silhouette of a person

Carried Objects

- We use prior information relative to the expected locations of carried objects
- Create templates for various poses



- Candidates are matched against pose templates
- Unmatched regions are expected to correspond to carried objects

Preliminary Results: RPI group

Station 1: Triskett Bridge

Input Video: tskt_bridge_1030a-1130a

Target Video: tskt_bridge_530p-630p

Station 2: West 117

Input Video: 117_tunnel_10a-11a

Target Video: 117_tunnel_5p-6p

Probe Images



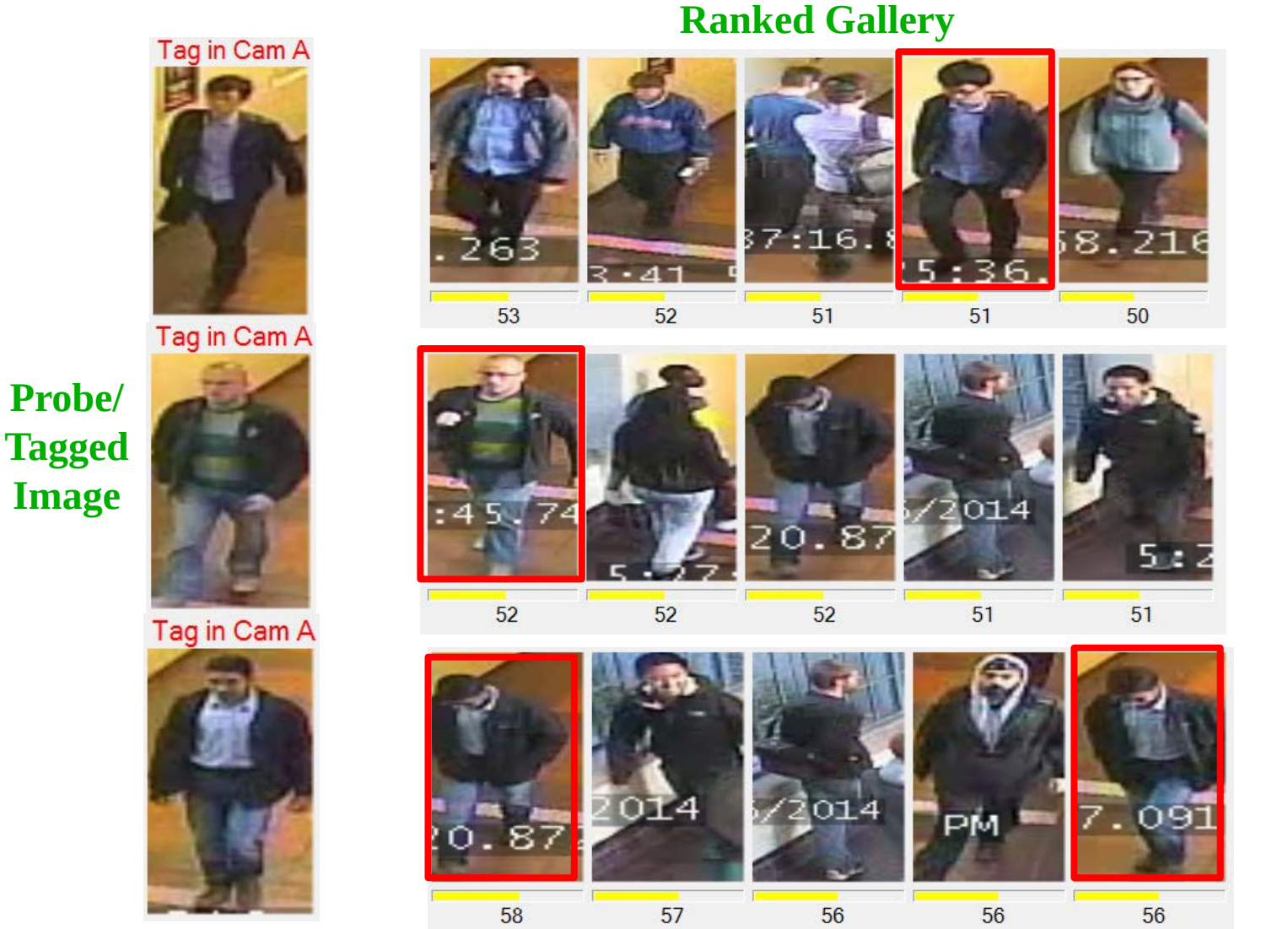
Gallery Images



Preliminary Results: RPI group



Preliminary Results: RPI group



Summary of RPI preliminary results

Input Video	Target Video	Person	Rank
tskt_bridge_1030a-1130a	tskt_bridge_530p-630p	Subject1	>5
		Subject2	5
		Subject3	1
		Subject4	1
		Subject5	1
		Subject6	3
% in rank 1			**50%**
% in rank 5			**83%**
117_tunnel_10a-11a	117_tunnel_5p-6p	Subject1	1
		Subject2	3
		Subject3	2
		Subject4	1
		Subject5	1
		Subject6	2
% in rank 1			**50%**
% in rank 5			**100%**

Preliminary results: NEU group

Basic information:
Station: West Park (highest FPS, ~17)
Video: w_pk_11a-12p; w_pk_530p-630p

Data:
Each person is represented by a video
Probe set: basic outfits
Gallery set: complement outfits + 25 randomly picked distractors

Features:
15 frame length dense trajectories + HOG3D ; Fisher Vector Encoded

Metric learning:
Kernel Linear Fisher Discriminative Analysis (kLFDA)

Training data:
SAIVT Soft-bio dataset



Actors



Sample Distractors

Re-identification via dynamics





Key idea: use *dynamics* of subject's motion for re-identification instead of their appearance

(Click images to play videos)

Preliminary results: NEU group

(Click images to play videos)

Probe/ Tagged Image

Robust to appearance change!



Preliminary results: NEU group

Ranked Gallery

(Click images to play videos)



Probe/ Tagged Image

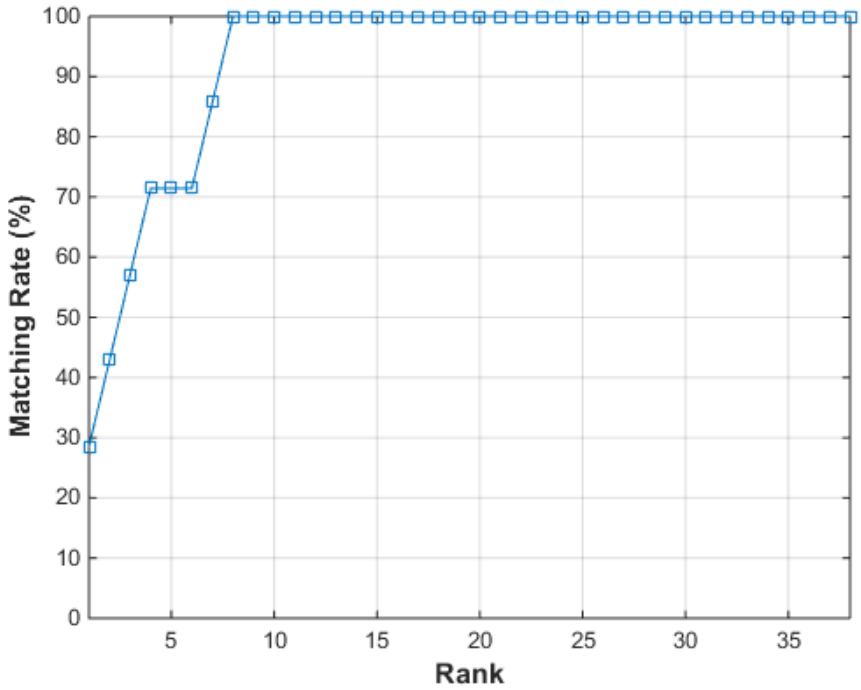


Robust to appearance change!



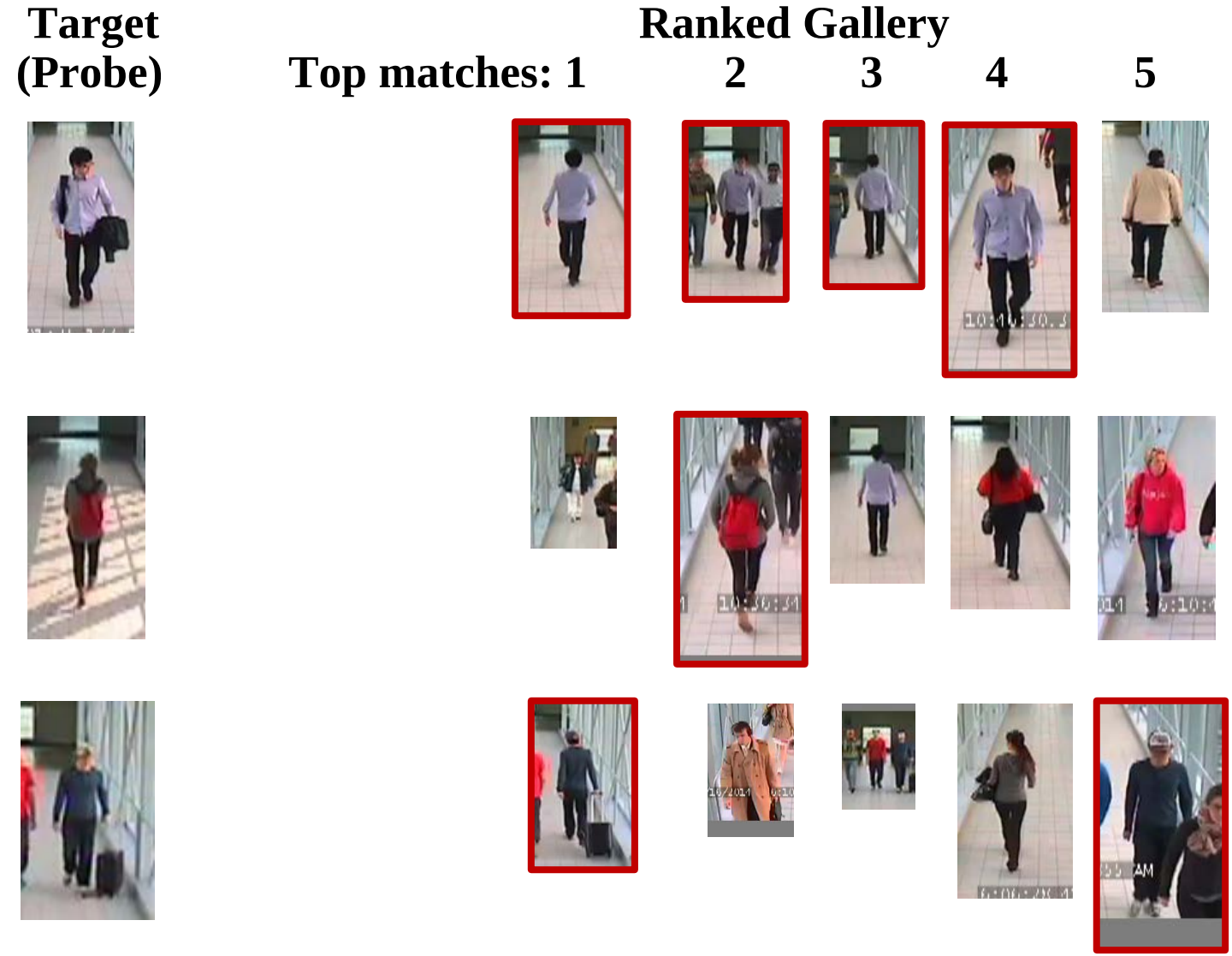
Preliminary results: NEU group

Gallery size: 38

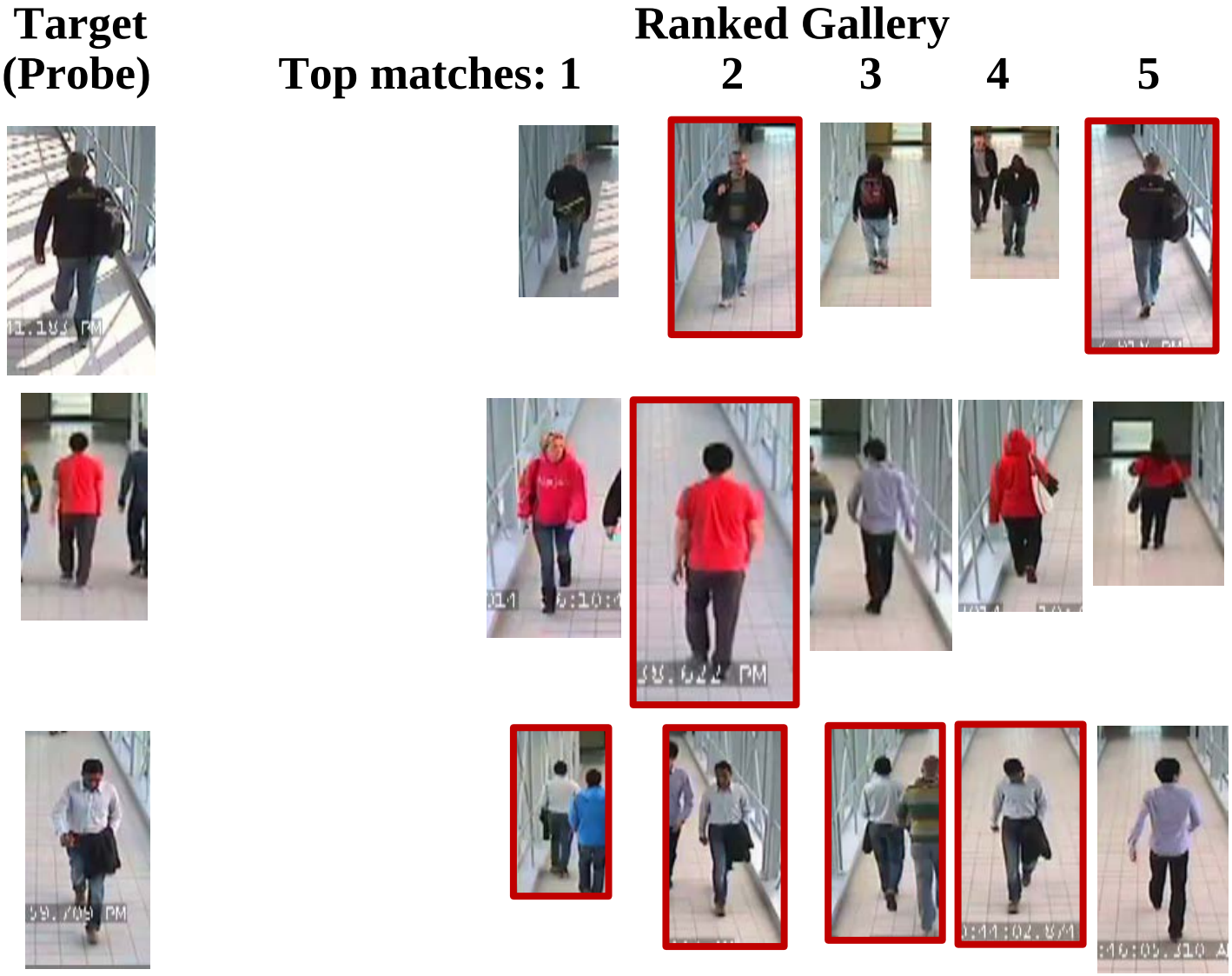


Rank	1	2	3	5	10
Matching Rate (%)	28.6	42.9	57.1	71.4	100

Preliminary Results: Purdue Group



Preliminary Results: Purdue Group



Preliminary Results: Purdue Group

Triskett Bridge

Rank	1	2	3	4	5	6
Matching Accuracy	53%	63%	77%	87%	93%	100%

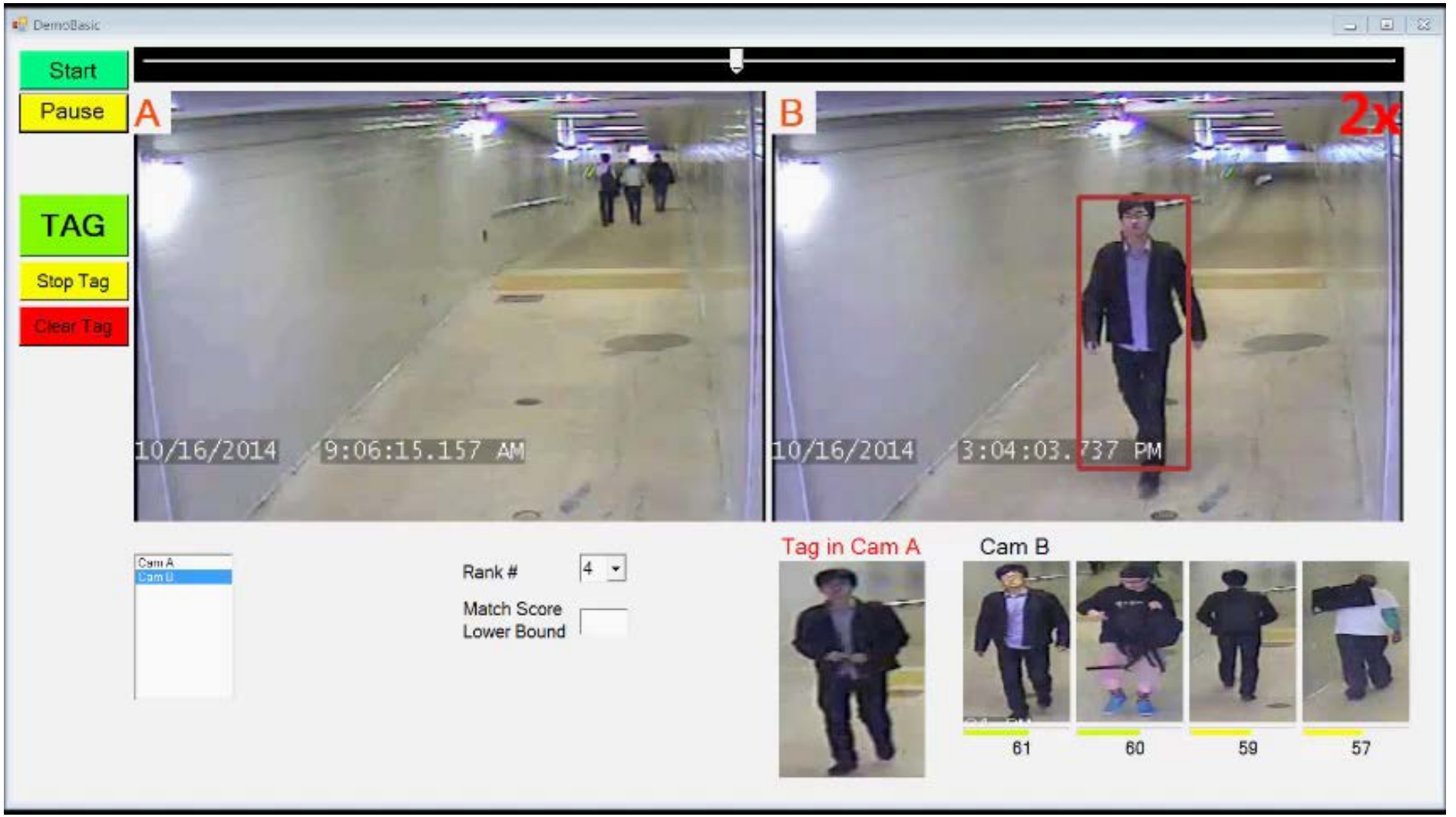


West 177 - Tunnel

Rank	1	2	3	4	5	6
Matching Accuracy	53%	77%	90%	100%	100%	100%



Prototype Graphical User Interface



This slide is a playable video

Current and Future Work (Body Re-id)

- Hybrid generative/discriminative approach for fast, robust human detection and tracking
- Further investigation of effective features (e.g., visual saliency and subjective attributes)
- Per-camera tuning of feature extraction and re-id metric
- Boosting approach for offline training
- Subject-discriminative features and weights learned on-the-fly
- Intelligent selection of distinctive target and candidate snapshots for multi-shot re-identification

Face Recognition

Patrick J. Flynn, Kevin Bowyer

- **Joel Brogan, Amy Li, Nikita Amelchenko**

Yun Fu

- **Ming Shao**

Motivation

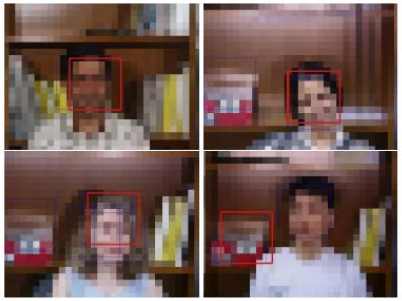
- Face detection in low-resolution images is challenging
- Detection becomes harder for both machine and human from resolutions of 24×24 to 6×6
- Results below are produced by combination of “Viola-Jones detector” + color heuristics



Resolution of faces
24 × 24



16 × 16



8 × 8



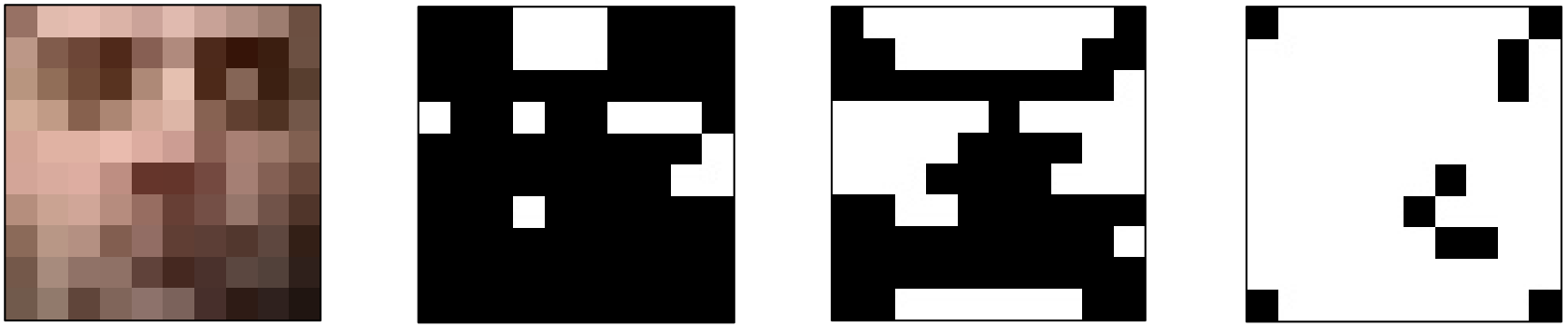
6 × 6

Motivation

- Poor lighting and low frame rate will fool detectors relying on skin color and face tracking
- In eBOLO dataset, lighting and frame rate are poor
- Commercial software is likely to fail
- To address these problems, we propose to:
  - Use body bounding box, and detect faces only on upper body
  - Empirically enlarge images (by image interpolation) to achieve better results

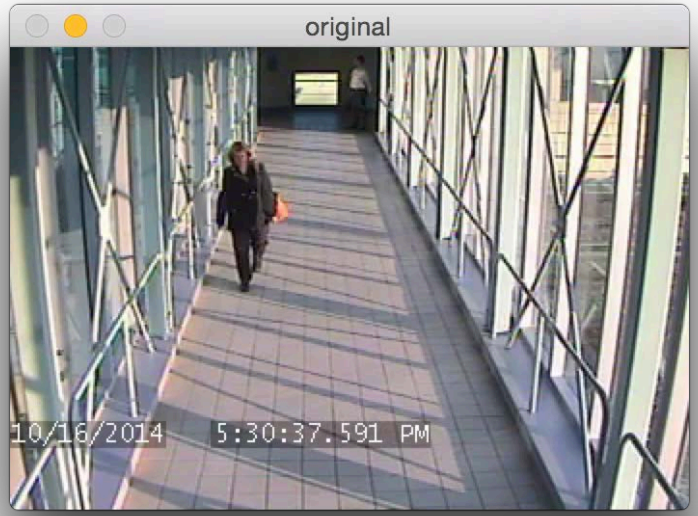
Face Detection: Notre Dame

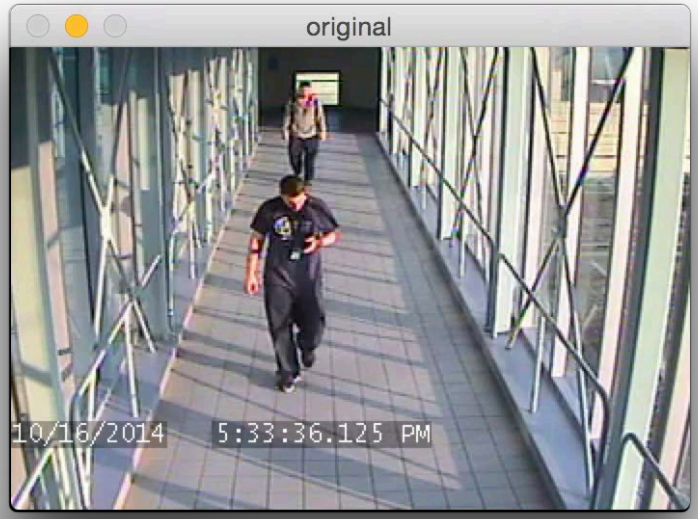
- Custom face detector implemented based on literature
- 3 stage detector, robust to illumination changes
- To prevent false negatives, we allow false positives to contribute to the final face detection
- A density filter of face detection points allows us to robustly filter out the actual facial area from all TP and FP detections from a given pedestrian



Left: 10x10 Face. Right: 3 cascading classifier stages. White areas are pixels used for classification of a frame determined by a modified adaboost method

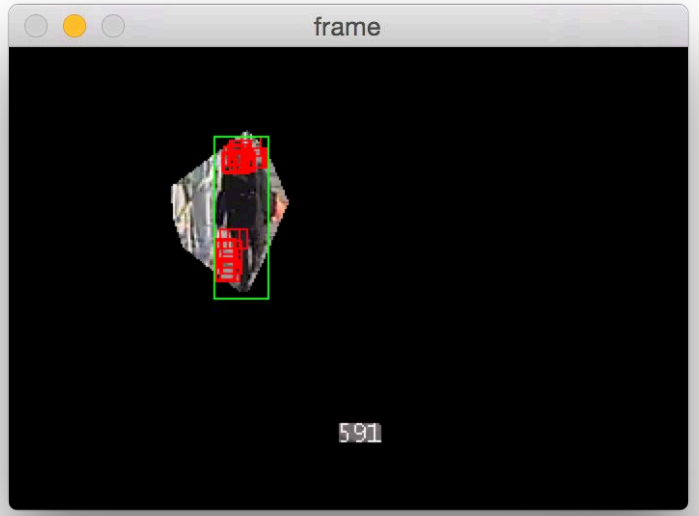
Preliminary Results: Notre Dame

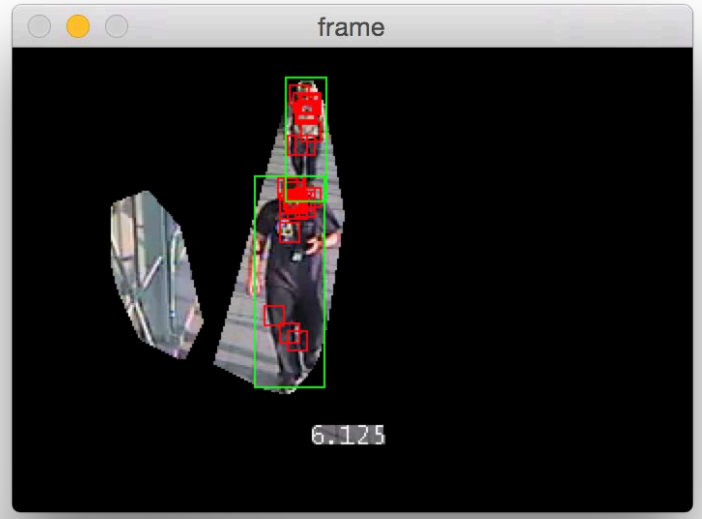






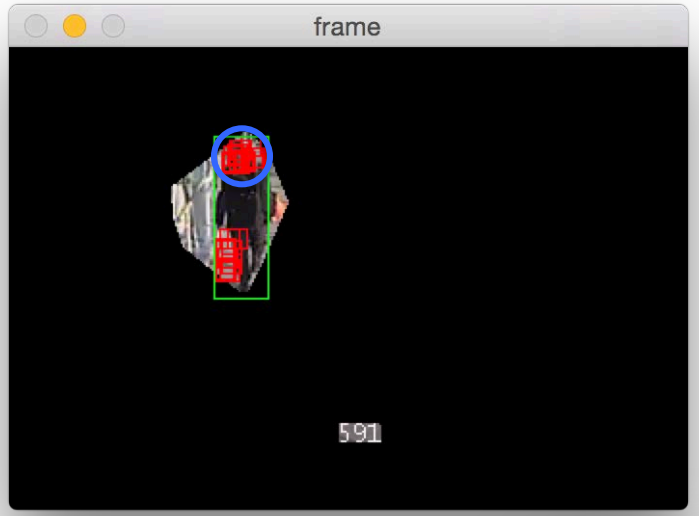
Preliminary Results: Notre Dame

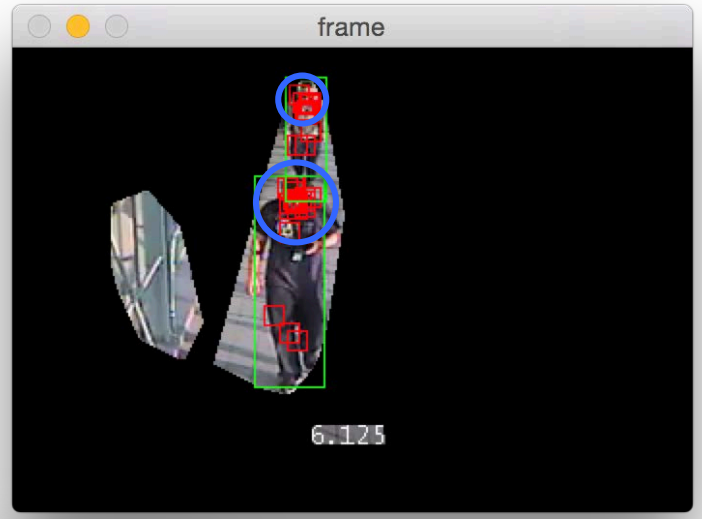






Preliminary Results: Notre Dame

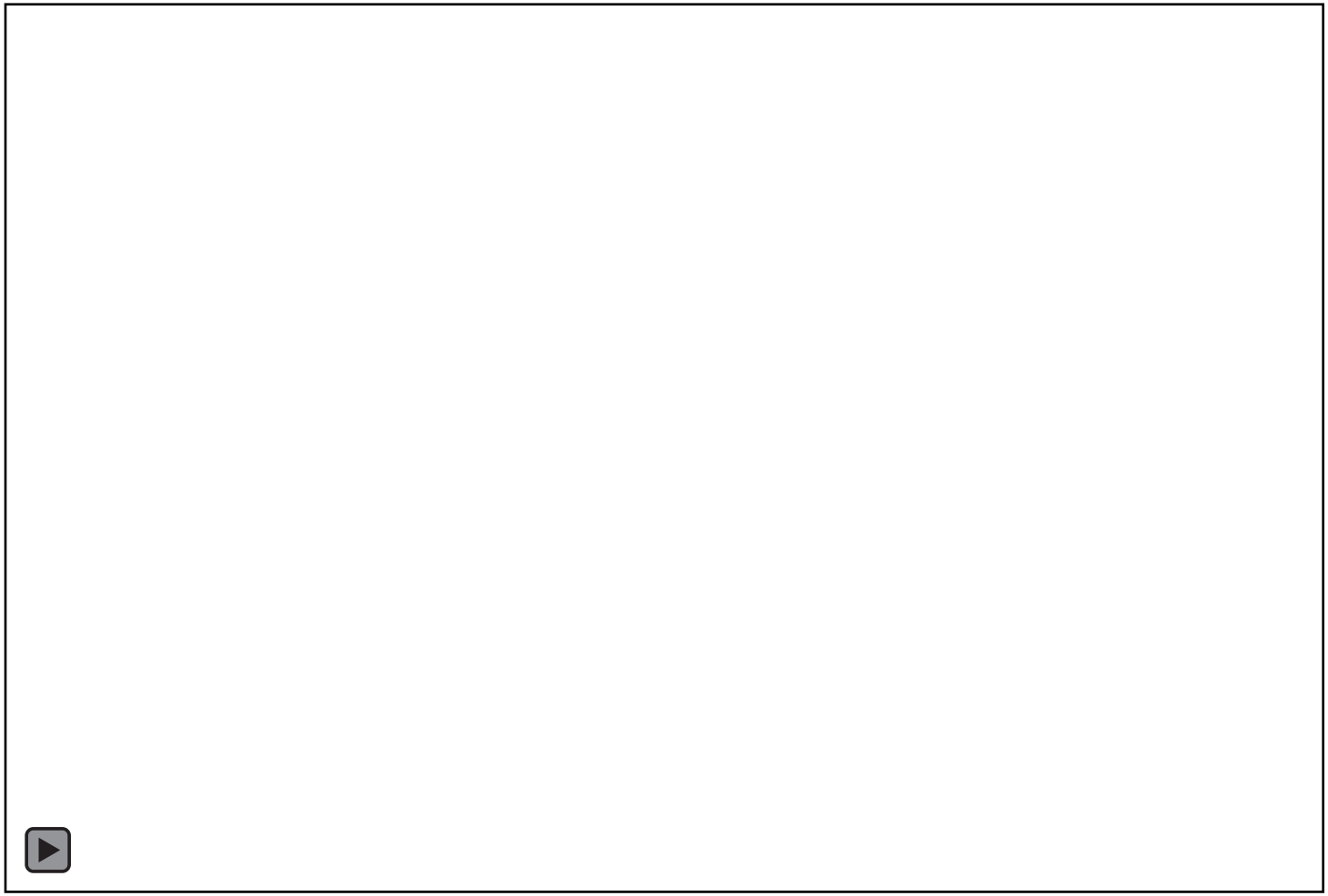






Preliminary Results: Notre Dame

Video Demo



Face Detection: NEU Group

- Viola-Jones face detector
  - Haar feature + weak classifiers + cascade structure
- Implementation: OpenCV Library
  - Version 2.4
  - Requirement on minimum resolution: 20 × 20
- Randomly sample 100 images from
  - West 117 Tunnel 5p-6p
  - Triskett Bridge 1030a-1130a





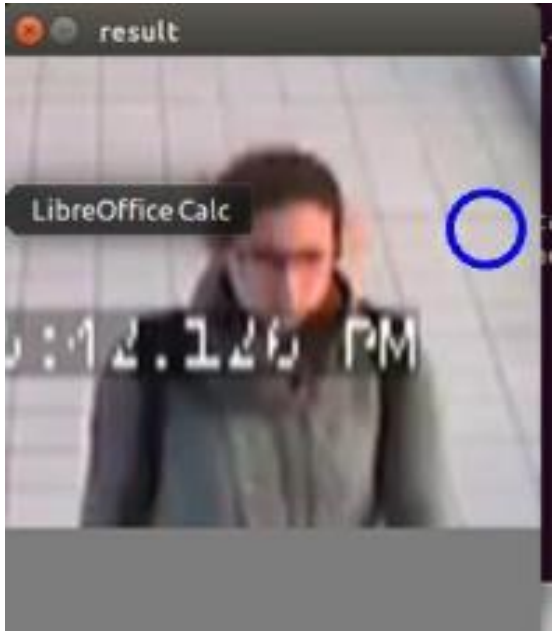
Preliminary Results: NEU Group

- Correct detections in low-resolution images (after image interpolation)



Preliminary Results: NEU Group

- Incorrect detections in low-resolution images, but can be corrected by heuristics (face is on top of upper body)





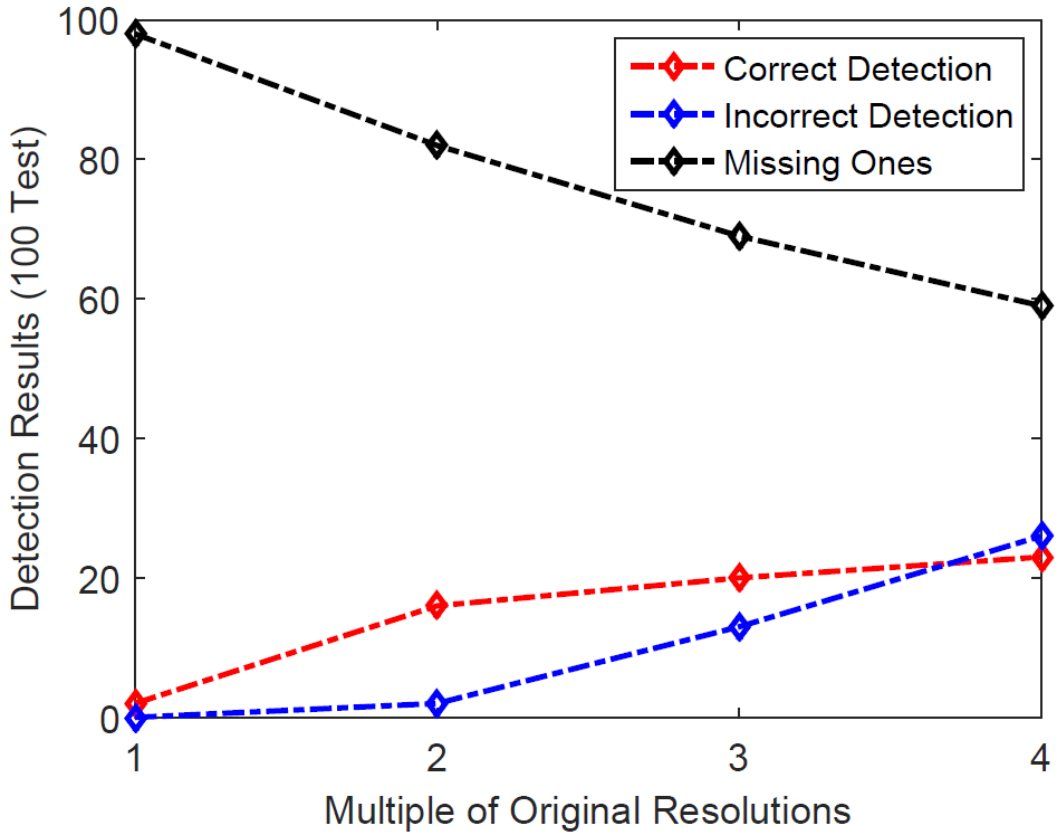






Preliminary Results: NEU Group

- Detection results of correct/incorrect and missed detections
- Using interpolation to increase the image resolution up to 4 times the original



Face Recognition: Notre Dame

- Transforming a simple face space using Modified Census Transform with local colors allows for significant illumination/color invariance
- Due to Very Low Resolution (VLR) conditions, localized feature extraction becomes less practical
- Using training data, the face-space relationships between cameras can be calculated as homographies
  - Face data still being annotated

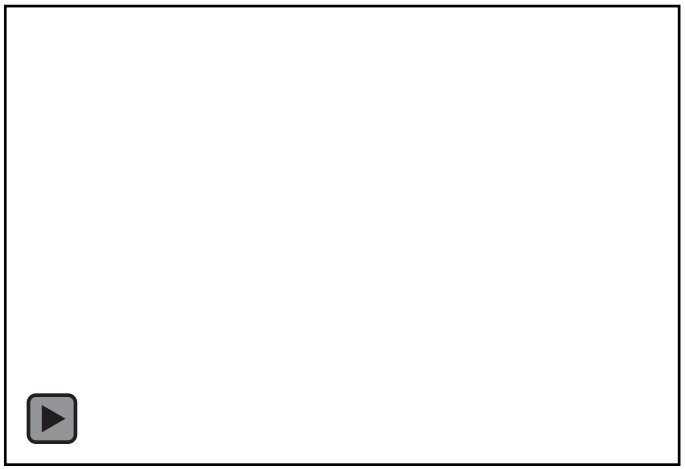
Face Recognition

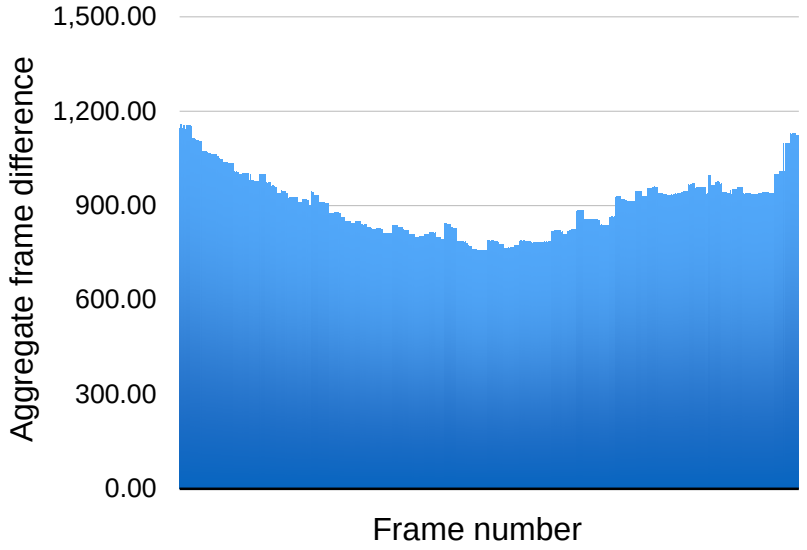
- Face Hallucination
  - Constructing higher-resolution faces for better feature extraction and matching
  - Currently determining viability for the Very Low Resolution case

Video Summarization: Notre Dame

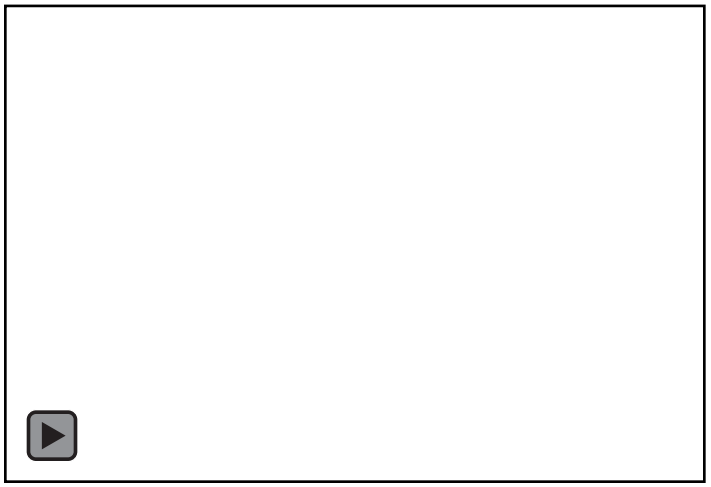
- Personnel: Nikita Amelchenko (CS undergraduate)
- Concept: remove video where nothing is happening
- Challenge: something is always happening
  - Noise
  - Timestamp/annotation
  - Other nonstationarities (sunlight)
- Approach
  - Baseline noise distribution
  - Masking
  - Thresholding difference

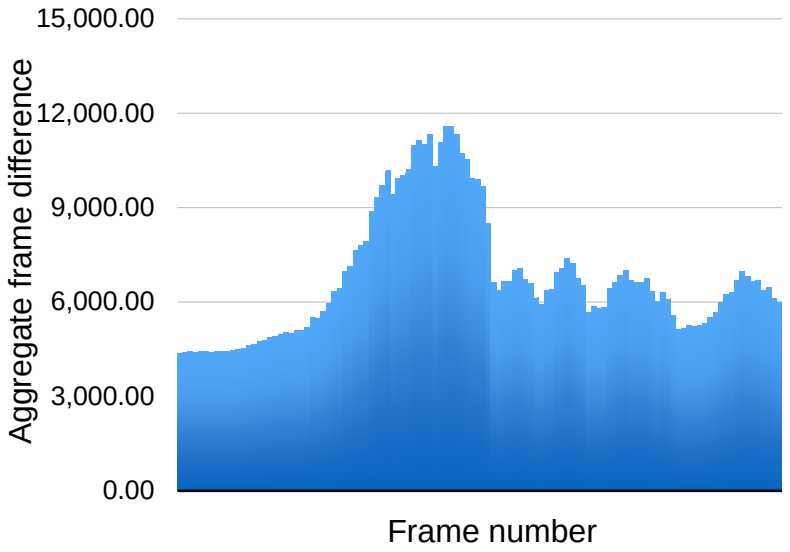
Baseline



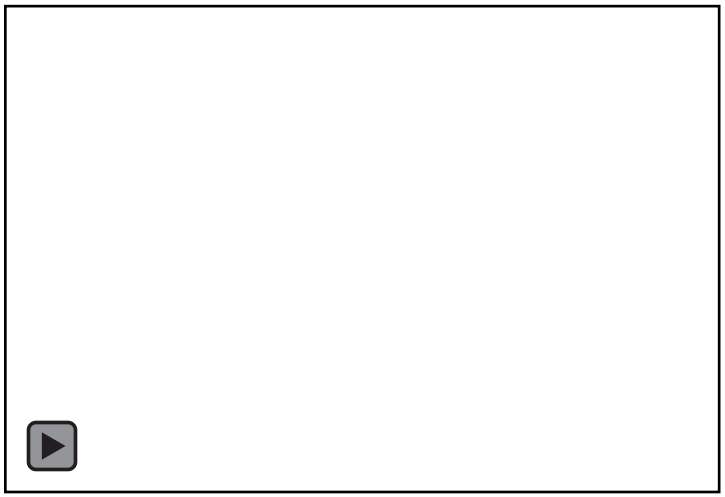

1,500.00
1,200.00
900.00
600.00
300.00
0.00
Aggregate frame difference
Frame number

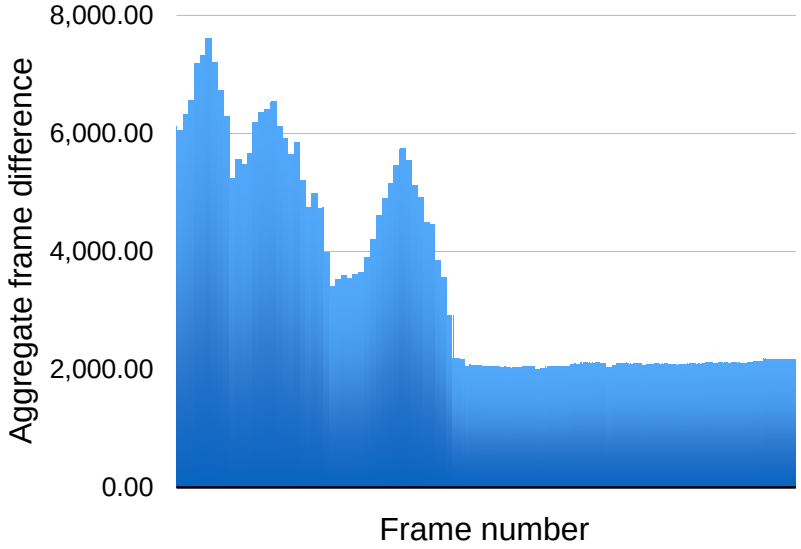
Motion




15,000.00
12,000.00
9,000.00
6,000.00
3,000.00
0.00
Aggregate frame difference
Frame number

Strong motion





Annotation Tools: NEU Group

FDViewer: A Face Detector Interface

- **Single program** to detect and analyze faces.
- **Visualize** images with labels.
- **Well organized** using Corpus structures.
  - Handle image collections in a fluent fashion.
- **Quickly** browse results and record metrics.
  - Count FP, FN, Missed Detections w/out ground truth
  - Note that with ground truth manual counting is minimal
  - Annotations are stored as CSV files:

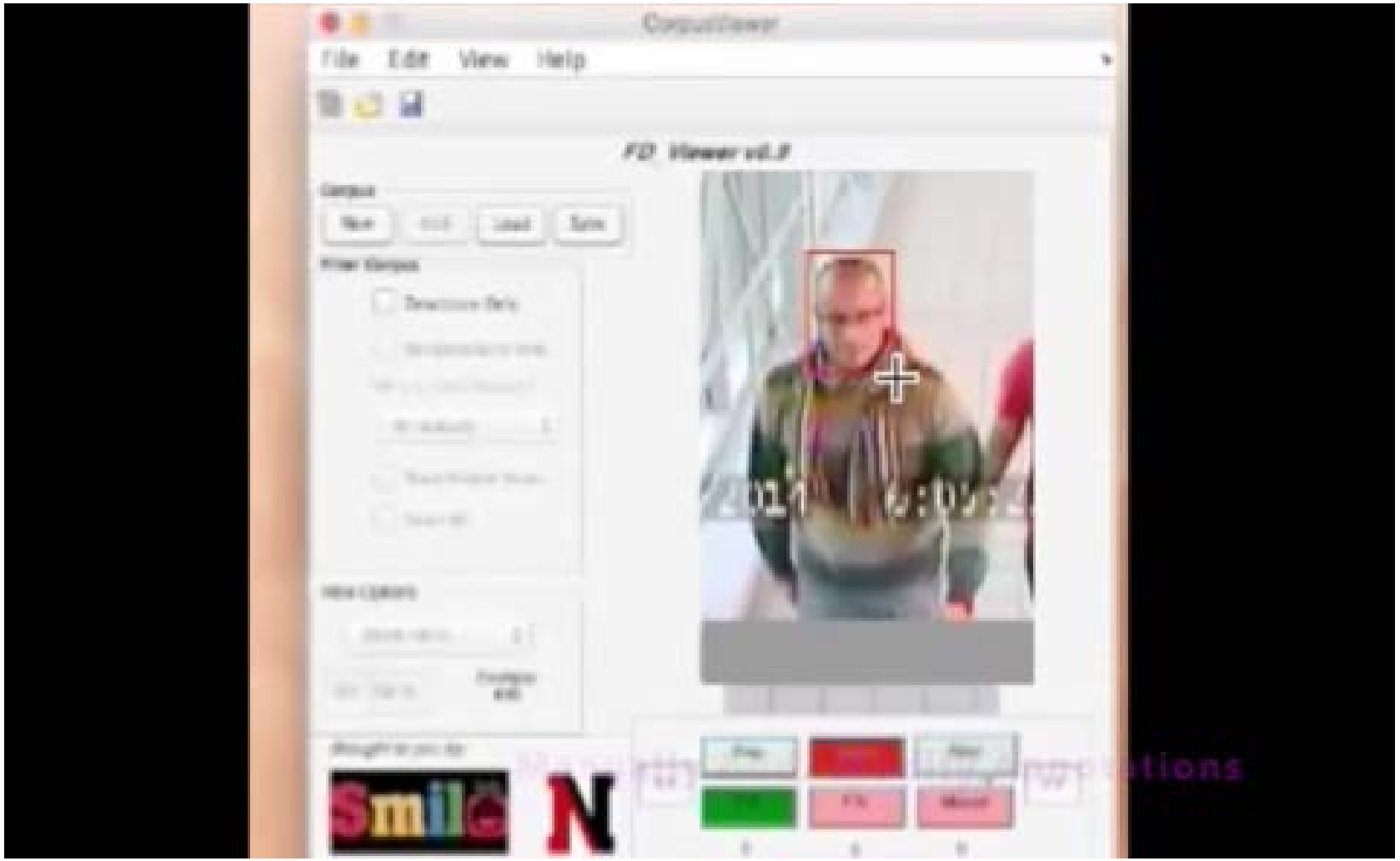
    <image_name, numFN, numFN, numMissed>
- **Easy** to create, save and load Corpus.





Annotation Tools

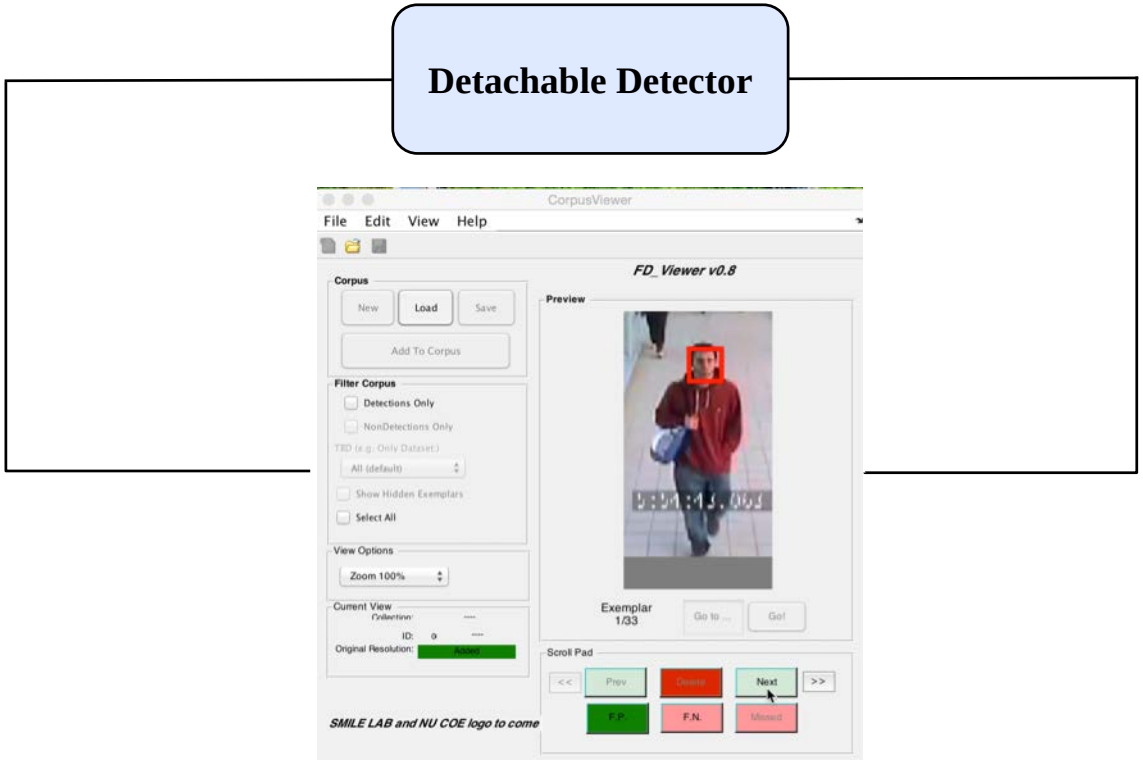
FDViewer: Video Demo



Annotation Tools

FDViewer: An Extendable (Modular) Framework

- **Easily attach** any object detector of choice.
  - **Change** detectors via dropdown menu.
  - **Analyze** and **Compare** different detectors.



Annotation Tools

Future Work

1. Enable Corpus merging.
   - Add user bookmarks.
2. Provide views (exemplar) filter.
   - View type specific results.
3. Allow images to be rescaled.
   - Reprocess scaled images.
4. Add ‘Delete’ to drop unwanted samples.



Future Work (face re-id)

- Fine tune Viola-Jones detector with appropriate image pre-processing
- Try different open source/commercial detectors (e.g., face plus plus)
- Compute quantitative results for detection, e.g., ROC, based on our face annotation tools (FDViewer)
- Investigate applicability of face recognition and video summarization to eBOLO application
- Integrate face algorithms with body-based algorithms to improve ranking of correct candidates

AUTOMATED CROWD FLOW ESTIMATION ENHANCED BY CROWDSOURCING

Javier Ribera[1] , Khalid Tahboub[2] and Edward J. Delp[2]

[1] **Universitat Politècnica de Catalunya, Barcelona, Spain**
[2] **Purdue University, West Lafayette, IN, U.S.A.**

Motivation

- Crowd flow is the process of estimating the number of people moving through a region in a given period of time.
- How robust is the automatic analysis of crowd flow?
- Can we identify uncertainties and utilize human intelligence to help machines perform better?
- How effective is crowdsourcing in enhancing the accuracy of crowd flow estimation?

Crowd Flow Estimation

- The number of people crossing the desired region of the image is proportional to the foreground pixels in the region.

$$v_N = \frac{\tilde{S}_N}{C}$$

$\tilde{S}_N$ is the weighted accumulation of foreground pixels related to moving objects inside the region and during the period of N frames.

v_N is the number of people that have crossed the region during a period of N frames.

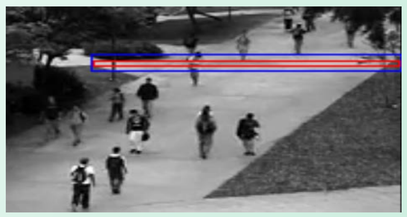
C is a scaling factor representing the number of foreground pixels per person

Estimation of Crowdedness

- Large crowd density results in a large number of occlusions which means a reduced number of foreground pixels per person.
- The level of crowdedness should be estimated in order to scale the foreground pixel count properly.
- Crowded scene presents a fine texture, while a sparse scene has a coarse texture.
- To estimate the level of crowdedness, we use a 16-D texture feature based on the Gray Level Co-occurrence Matrix (GLCM).
- Four GLCM matrices are created, each matrix represents a different direction for the adjacent pixels: right, top-right, top and top-left.
- For each matrix we extract four scalar features: energy, entropy, homogeneity and contrast to get a 16-D vector.

A region through which people passing are counted (Tripwire) and a Region of Interest surrounding it (ROI) are both selected.

ROI
Tripwire



Training process generates feature vectors to represent the texture of each level of crowdedness. It also determines the scaling factor for each level (number of foreground pixels per moving person)

Step 1: Extract the texture feature vector and classify it into one of the levels of crowdedness. Classification is based on the nearest reference vector.

Step 2: Find the count of foreground pixels related to moving objects and scale it properly using the scaling factor for this specific level of crowdedness.

Enhancement Using Crowdsourcing

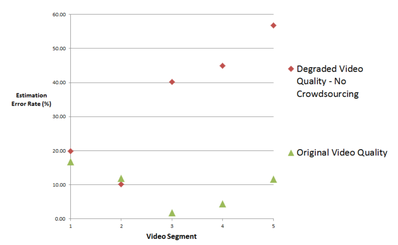
- To enhance the classification performance, we use crowdsourcing to reduce uncertainty.
- The distance between a texture feature vector and the nearest reference vector, d_1, might be very comparable to the distance to the second nearest neighbor, d_2.
- We define "uncertainty" as the ratio d_1/d_2 to represent the uncertainty of the classification decision.
- We use crowdsourcing and ask members of the crowd to classify frames whenever "uncertainty" is greater than a predefined threshold:

$$\mu = \frac{d_1}{d_2} > \alpha$$

- The new feature vectors are incorporated into the classifier to reduce future uncertainty

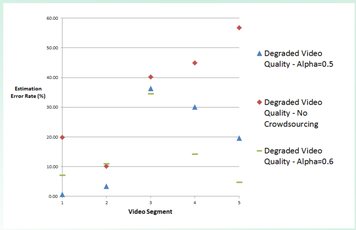
Robustness against Compression and Contrast Change

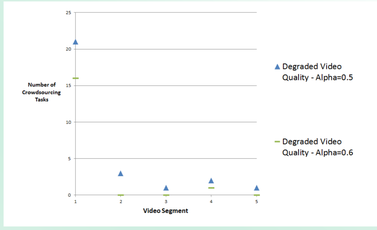
Degradation of video quality can significantly impact the automatic analysis.



Degradation of video quality is a result of compression and contrast change. H.264 compression was implemented using FFmpeg Constant Rate Factor (CRF), CRF was set to 33. Contrast was boosted to 1:68 using FFmpeg eq2 filter.

Results – Crowdsourcing Enhancement





This demonstrates that our proposed method manages to identify uncertainties and use the crowd in an effective way to lower the error rates.

Conclusion and Future Work

- We present enhancements to our previously developed crowd flow estimation method.
- Experimental evaluation demonstrates that our proposed method identifies uncertainties and use crowdsourcing in an effective way.
- Future work includes building a pyramid model to differentiate the accuracy of various crowd members and incorporating it in our method.

Publications

J. Ribera, K. Tahboub and E. J. Delp, "Automated crowd flow estimation enhanced by crowdsourcing", To appear, Proceedings of the IEEE National Aerospace & Electronics Conference, Dayton, Ohio, June 2014.

Acknowledgements

- This work is partially funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001.
- We used the following publically available surveillance dataset for our experiments:
  – UCSD pedestrian dataset : A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 5, pp. 909–926, May 2008. doi: [10.1109/TPAMI.2007.70738]

Visual Saliency Models Based on Spectrum Processing

Bin Zhao and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana, USA

IEEE Winter Conference on Applications of Computer Vision

SUMMARY

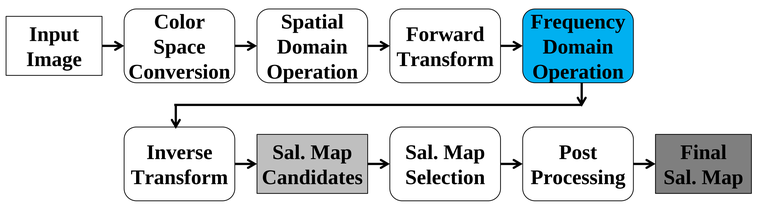
Visual saliency is the perceptual process that makes attractive objects "stand out" from their surroundings in the low-level human visual system. A map of the "salient objects" or a saliency map is generated to indicate the locations of salient regions in a scene.

We describe four frequency domain visual saliency models based on new spectrum processing methods, i.e. the Gamma Corrected Spectrum (GCS) model, the Gamma Corrected Log Spectrum (GCLS) model, the Gaussian Filtered Spectrum (GFS) model, and the Gaussian Filtered Log Spectrum (GFLS) model.

A set of saliency map candidates are generated by inverse transform of a set of modified spectrums. An output saliency map is selected by minimizing the entropy among the set of saliency map candidates. Experimental results show that four extensions of our GCS, GCLS, GFS, and GFLS models are more accurate and efficient than some state-of-the-art saliency models in predicting eye fixation on standard image datasets.

Frequency Domain Visual Saliency Models

- **Generalize existing frequency domain models**



The Generalization of Frequency Domain Visual Saliency Models

- **Frequency Domain Representation**

(1) Transformed Image

$$\mathcal{F}(u,v) = T[f(x,y)]$$

(2) Amplitude Spectrum

$$\mathcal{A}(u,v) = |\mathcal{F}(u,v)|$$

(3) Phase Spectrum

$$\mathcal{P}(u,v) = angle(\mathcal{F}(u,v))$$

(4) Log Amplitude Spectrum

$$\mathcal{L}(u,v) = log_e(\mathcal{A}(u,v))$$

- **Entropy Based Saliency Map Selection**
  - Select a "good" saliency map by minimizing the entropy among the set of saliency map candidates

- **Post-Processing**
  - Each element of saliency map is squared individually
  - Convolute saliency map with a Gaussian burring kernel

Proposed Visual Saliency Models

- **Gamma Corrected Spectrum (GCS) Model**
  - Gamma correction is a nonlinear operation
  - A set of gamma corrections with different gamma values are utilized to modify the amplitude spectrum

$$\mathcal{A}_{GCS}(u,v,k) = (\mathcal{A}(u,v))^{\gamma_k}$$

$$S(x,y,k) = T^{-1}[(\mathcal{A}(u,v))^{\gamma_k} \cdot exp(i \cdot \mathcal{P}(u,v))]$$

- **Gamma Corrected Log Spectrum (GCLS) Model**
  - A set of gamma corrections with different gamma values are utilized to modify the log amplitude spectrum

$$\mathcal{L}_{GCLS}(u,v,k) = (\mathcal{L}(u,v))^{\gamma_k}$$

$$S(x,y,k) = T^{-1}[exp((\mathcal{L}(u,v))^{\gamma_k} + i \cdot \mathcal{P}(u,v))]$$

- **Gaussian Filtered Spectrum (GFS) Model**
  - A set of Gaussian filters with various standard deviations are used to filter the amplitude spectrum

$$\mathcal{A}_{GFS}(u,v,k) = \mathcal{A}(u,v) \star GF(u,v,k)$$

$$S(x,y,k) = T^{-1}[\mathcal{A}_{GFS}(u,v,k) \cdot exp(i \cdot \mathcal{P}(u,v))]$$

- **Gaussian Filtered Log Spectrum (GFLS) Model**
  - A set of Gaussian filters with various standard deviations are used to filter the log amplitude spectrum

$$\mathcal{L}_{GFLS}(u,v,k) = \mathcal{L}(u,v) \star GF(u,v,k)$$

$$S(x,y,k) = T^{-1}[exp(\mathcal{L}_{GFLS}(u,v,k) + i \cdot \mathcal{P}(u,v))]$$

- **Naming Convention of The Extended Models (A-B-C)**
  - **A** represents a proposed model's abbreviation or an existing model's abbreviation
  - **B** represents a specific transform used for the model
  - **C** represents a particular color space used for the model

Eye Fixation Image Datasets

- **Saliency models provide accurate prediction of human eye fixation on natural images in free viewing**

- **Two standard image datasets**

(1) The Bruce and Tsotsos (BT) dataset [Bruce2009]
- 120 color images with resolution of 681x511 pixels from indoor and outdoor scenes and the eye fixation data is based on 20 subjects

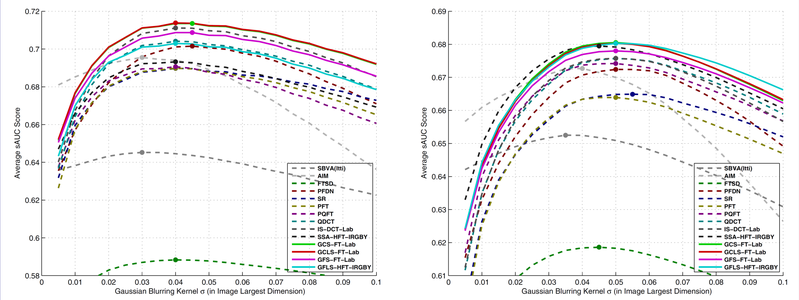
(2) The Li *et al*.'s (Li) dataset [Li2013]
- 235 color images with resolution of 640x480 pixels in six categories based on 19 subjects

EXPERIMENTAL RESULTS

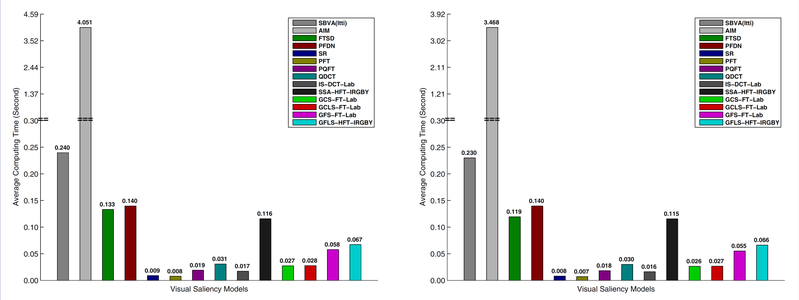
- **Comprehensive Evaluation:** Comparing our 4 best extended saliency models with 10 existing saliency models using two standard image datasets
  - Our 4 best extended models:**GCS-FT-Lab**, **GCLS-FT-Lab**, **GFS-FT-Lab**, and **GFLS-HFT-IRGBY**
  - 10 existing models:**SBVA** [Itti1998], **AIM** [Bruce2009], **FTSD** [Achanta2009], **PFDN** [Bian2010], **SR** [Hou2007], **PFT** [Guo2008], **PQFT** [Guo2010], **QDCT** [Schauerte2012], **IS-DCT-Lab** [Hou2012], **SSA-HFT-IRGBY** [Li2013]
- **Evaluation Measure:** Adopted the shuffled Area Under Curve (sAUC) score as the evaluation measure and developed an enhanced evaluation tool using the original MATLAB implementation of the existing models
  - Our 4 best extended models are better than some state-of-the-art models in predicting eye fixation

Image Dataset	SBVA [20]	AIM [6]	FTSD [1]	PFDN [3]	SR [18]	PFT [13]	PQFT [14]	QDCT [25]	IS-DCT [17]	SSA-HFT [22]	GCS-FT -Lab	GCLS-FT -Lab	GFS-FT -Lab	GFLS-HFT -IRGBY
BT [6] Rank	13	8	14	7	11	11	10	5	**3**	9	**2**	**1**	**4**	6
Max sAUC	0.6453	0.6956	0.5883	0.7015	0.6898	0.6898	0.6906	0.7041	0.7111	0.6932	0.7135	0.7138	0.7087	0.7030
σ_{opt}	0.030	0.030	0.040	0.045	0.040	0.040	0.040	0.040	0.040	0.040	0.045	0.040	0.045	0.040
Li [22] Rank	13	9	14	10	11	12	8	7	6	**4**	**1**	**2**	5	**2**
Max sAUC	0.6525	0.6727	0.6186	0.6724	0.6649	0.6639	0.6742	0.6757	0.6758	0.6795	0.6805	0.6803	0.6780	0.6803
σ_{opt}	0.035	0.040	0.045	0.050	0.055	0.050	0.050	0.050	0.050	0.045	0.050	0.050	0.050	0.050

The rank, maximum sAUC score, and associated Gaussian σ_{opt} of each model



The sAUC score of each model for BT (left) and Li (right) datasets



The average execution time of each model for BT (left) and Li (right) datasets

CONCLUSIONS

- We investigated low-complexity visual saliency models using spectral analysis approaches and proposed four visual saliency models based on spectrum processing methods
- **Acknowledgement:** This work was supported by the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) Center of the U.S. Department of Homeland Security (DHS) under Award Number 2009-ST-061-CI000

Visual Saliency Models Based on Spectrum Processing

Bin Zhao and Edward J. Delp

Video and Image Processing Laboratory (VIPER)

School of Electrical and Computer Engineering

Purdue University

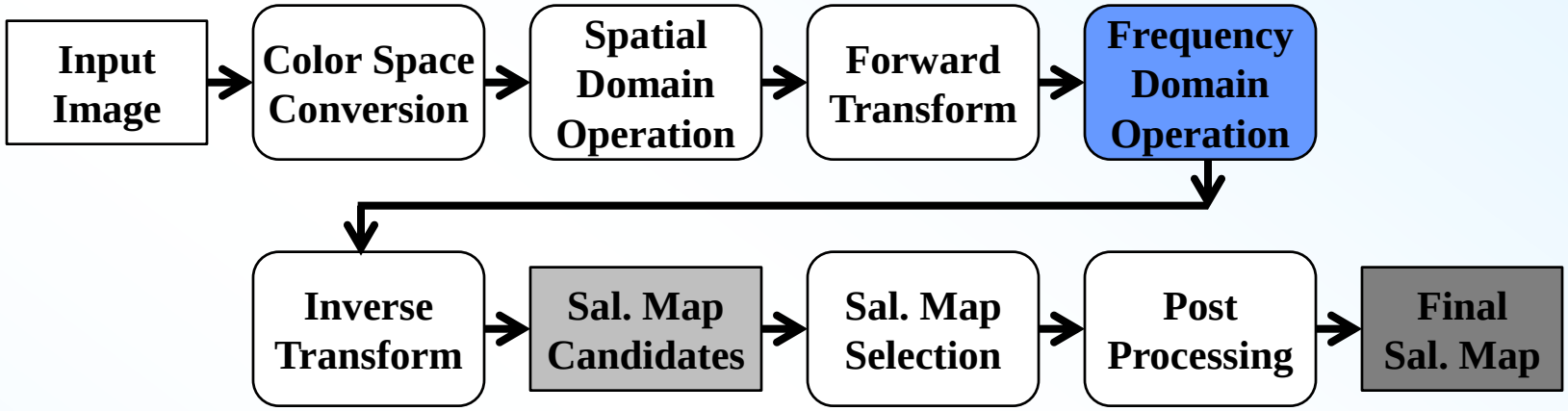
West Lafayette, Indiana, USA

Summary

- **Investigated low-complexity bottom-up visual saliency using spectral analysis approaches**
- **Proposed 4 visual saliency models based on new spectrum processing methods**
- **A saliency map is selected by minimizing the entropy among a set of saliency map candidates**
- **Developed an evaluation tool for benchmarking visual saliency models**
- **Experimental results show that four extensions of our models are more accurate and efficient than previous saliency models in predicting eye fixation**

PURDUE UNIVERSITY

Proposed Visual Saliency Models



- **Propose new visual saliency models based on different spectrum processing methods in the frequency domain**
  - **Gamma Correction (non-linear operation)**
    - **Gamma Corrected Spectrum (GCS) Model**
    - **Gamma Corrected Log Spectrum (GCLS) Model**
  - **Gaussian Filtering (linear operation)**
    - **Gaussian Filtered Spectrum (GFS) Model**
    - **Gaussian Filtered Log Spectrum (GFLS) Model**

Experiments

- **Visual saliency models should be capable of providing accurate prediction of human eye fixation on natural images in free viewing**
- **Use shuffled Area Under Curve (sAUC) as the fair evaluation measure to eliminate the center-bias effect in eye fixation data**
- **Test the proposed and existing visual saliency models to predict eye fixation on two standard image datasets**

  (1) The Bruce and Tsotsos dataset (BT)

  – 120 color images with resolution of 681x511 pixels

  (2) The Li *et al.*'s dataset (Li)

  – 235 color images with resolution of 640x480 pixels

Comparison with Other Models

- **A comprehensive evaluation of the 4 best extended saliency models by comparing with 10 other saliency models using two standard image datasets**
  - **SBVA(Itti) [Itti1998], AIM [Bruce2009], FTSD [Achanta2009], PFDN [Bian2010], SR [Hou2007], PFT [Guo2008], PQFT [Guo2010], QDCT [Schauerte2012], IS-DCT-Lab [Hou2012], SSA-HFT-IRGBY [Li2013]**
  - **SBVA(Itti), AIM, FTSD are spatial domain models**
  - **PFDN, SR, PFT, PQFT, IS-DCT-Lab, SSA-HFT-IRGBY and the 4 best extended models (GCS-FT-Lab, GCLS-FT-Lab, GFS-FT-Lab, and GFLS-HFT-IRGBY) are all frequency domain models**

PURDUE UNIVERSITY

Comparison with Other Models

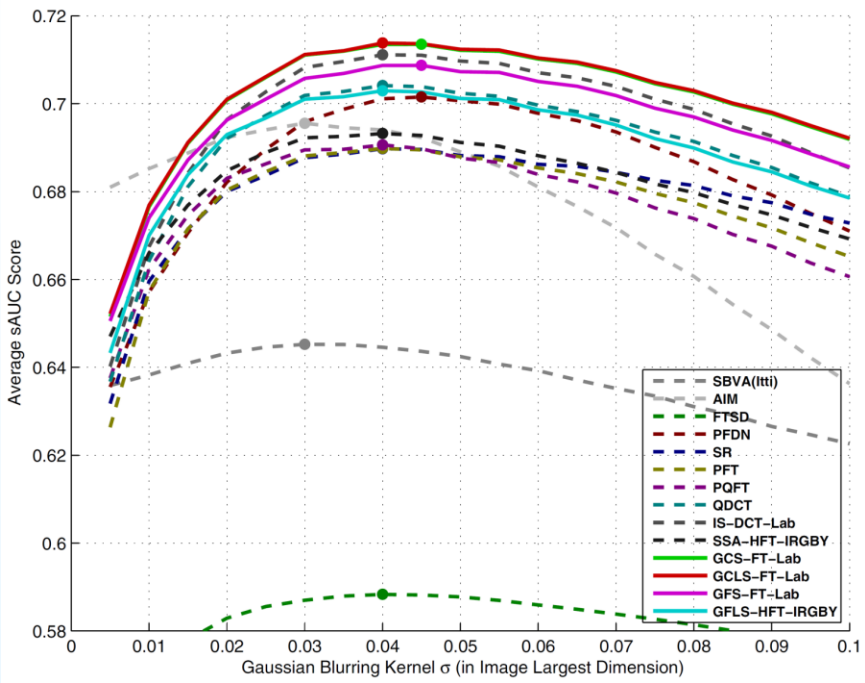
- **The summary of each model**
  - **The rank in the same image dataset, the maximum sAUC score, and the associated Gaussian σ_{opt}**

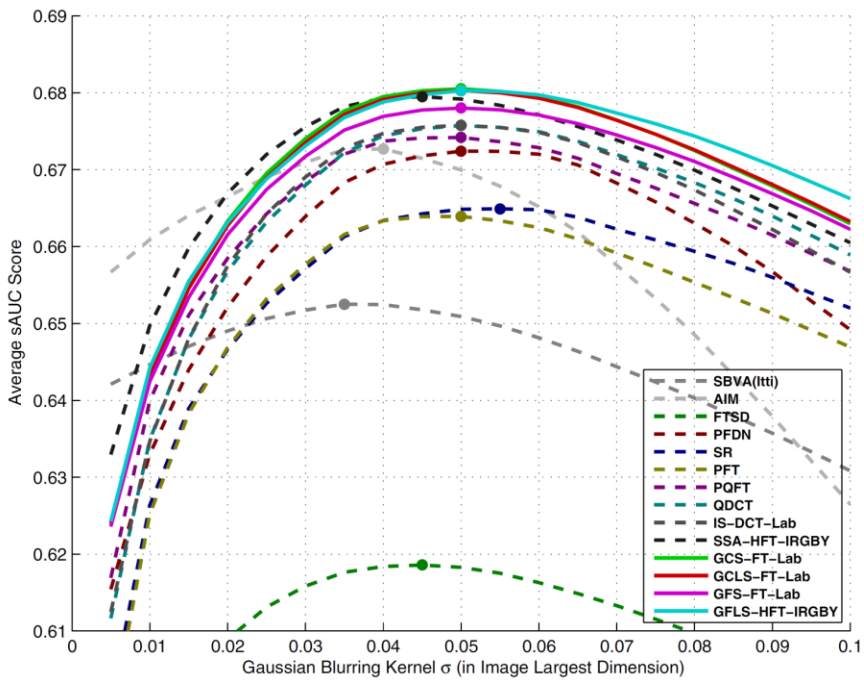
Image Dataset	SBVA [20]	AIM [6]	FTSD [1]	PFDN [3]	SR [18]	PFT [13]	PQFT [14]	QDCT [25]	IS-DCT [17]	SSA-HFT [22]	GCS-FT -Lab	GCLS-FT -Lab	GFS-FT -Lab	GFLS-HFT -IRGBY
BT [6] Rank	13	8	14	7	11	11	10	5	**3**	9	**2**	**1**	**4**	6
Max sAUC	0.6453	0.6956	0.5883	0.7015	0.6898	0.6898	0.6906	0.7041	0.7111	0.6932	0.7135	0.7138	0.7087	0.7030
σ_{opt}	0.030	0.030	0.040	0.045	0.040	0.040	0.040	0.040	0.040	0.040	0.045	0.040	0.045	0.040
Li [22] Rank	13	9	14	10	11	12	8	7	6	**4**	**1**	**2**	5	**2**
Max sAUC	0.6525	0.6727	0.6186	0.6724	0.6649	0.6639	0.6742	0.6757	0.6758	0.6795	0.6805	0.6803	0.6780	0.6803
σ_{opt}	0.035	0.040	0.045	0.050	0.055	0.050	0.050	0.050	0.050	0.045	0.050	0.050	0.050	0.050

- **The 4 best extended models are generally better than previous models**
  - **The top 6 out of all the 14 models for the BT dataset**
  - **The top 5 out of all the 14 models for the Li dataset**
- **The best model: GCS-FT-Lab**

PURDUE UNIVERSITY

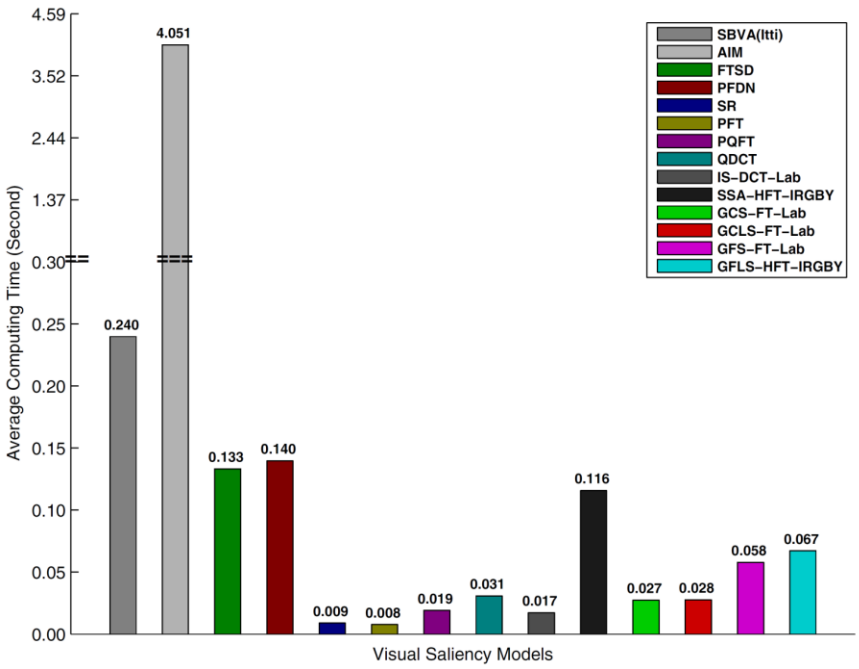
Comparison with Other Models

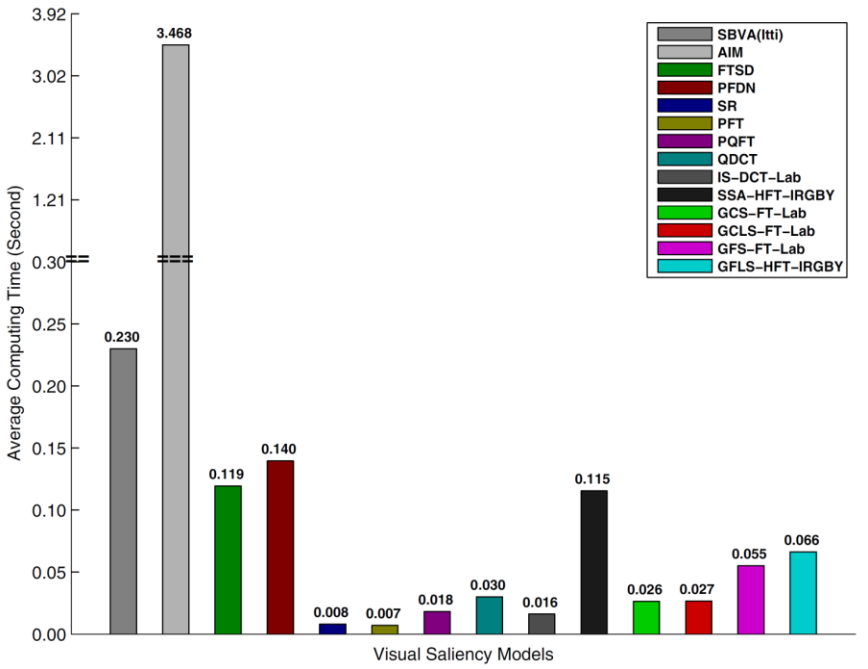




The sAUC score of each model for BT (left) and Li (right) datasets

Comparison with Other Models





The average execution time of each model for BT (left) and Li (right) datasets

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Beyond Big Data Analytics: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

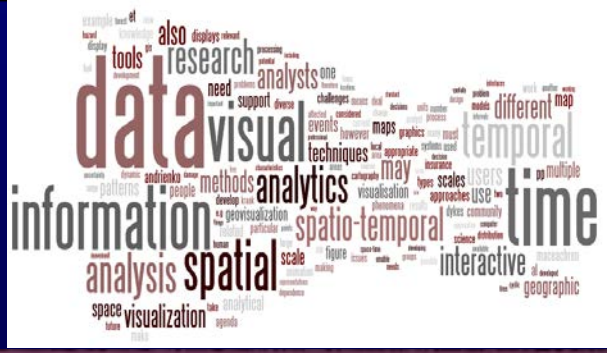
CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

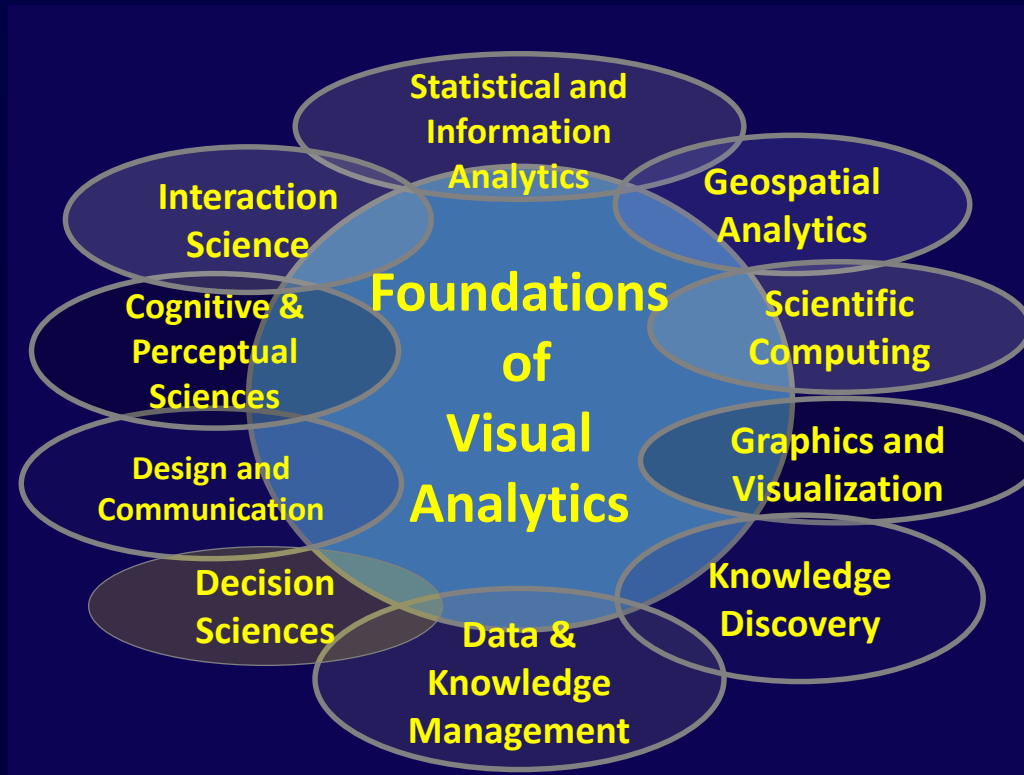
February 2015

What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

Big Data



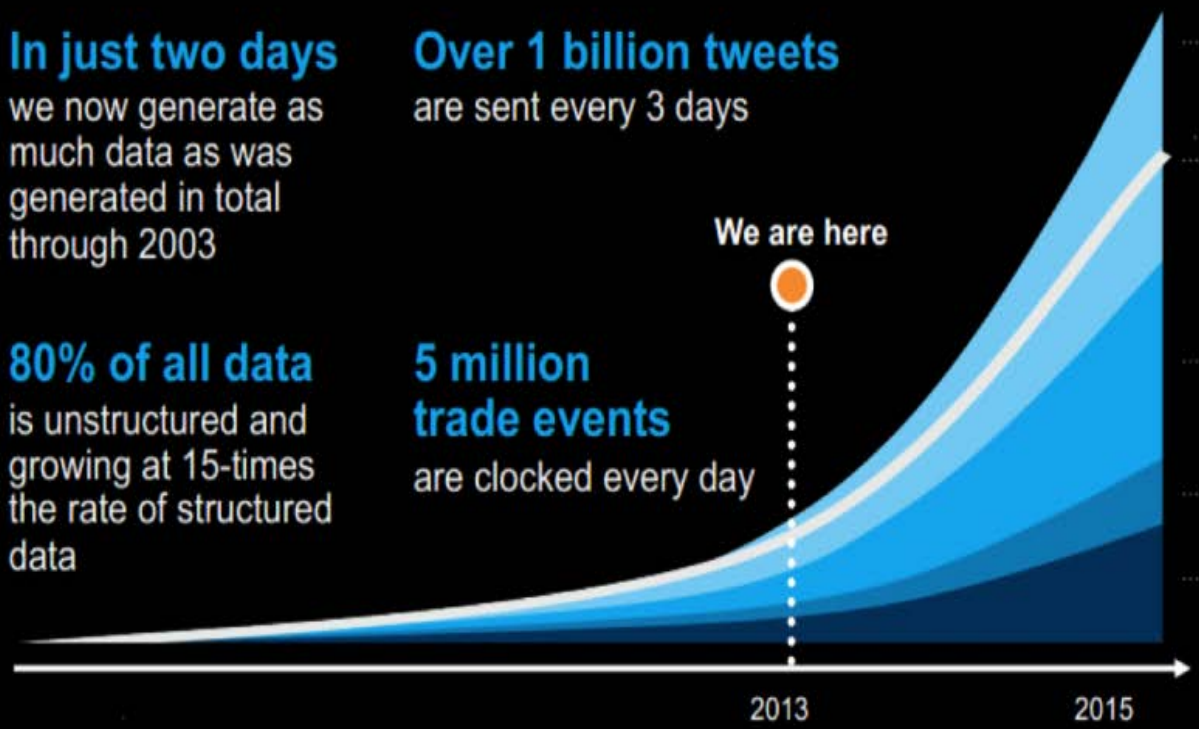
These forces fuel an explosion of data – a new economic asset forming the basis of opportunity

In just two days
we now generate as much data as was generated in total through 2003

Over 1 billion tweets
are sent every 3 days

80% of all data
is unstructured and growing at 15-times the rate of structured data

5 million trade events
are clocked every day



Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

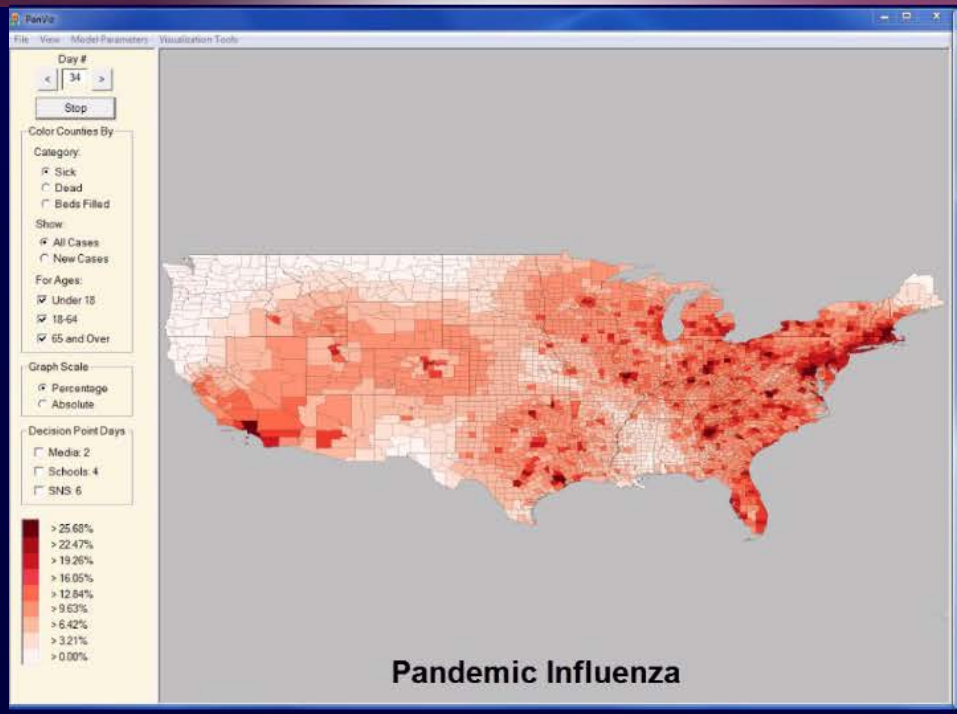
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

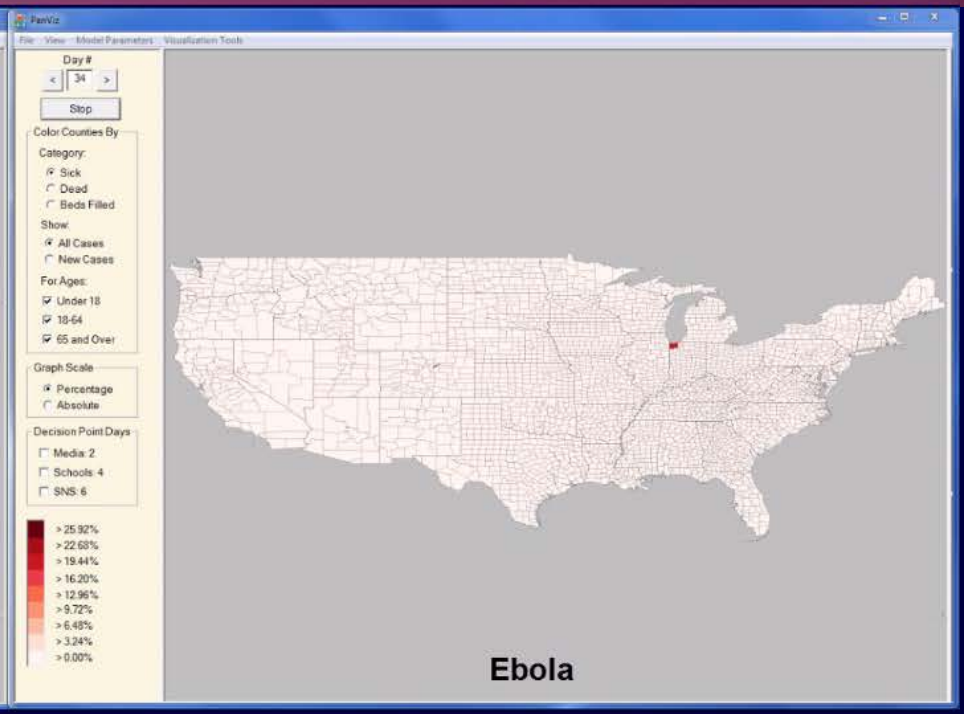
VADM Robert Parker, 2012 MRS Keynote Address



Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL





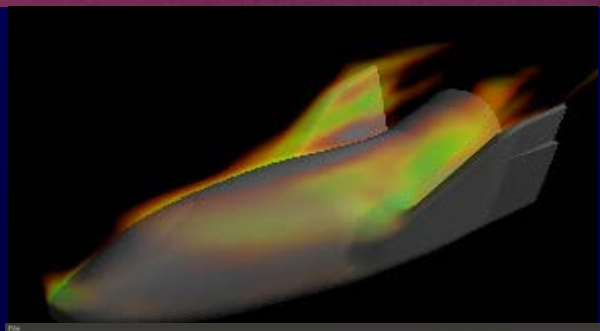
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

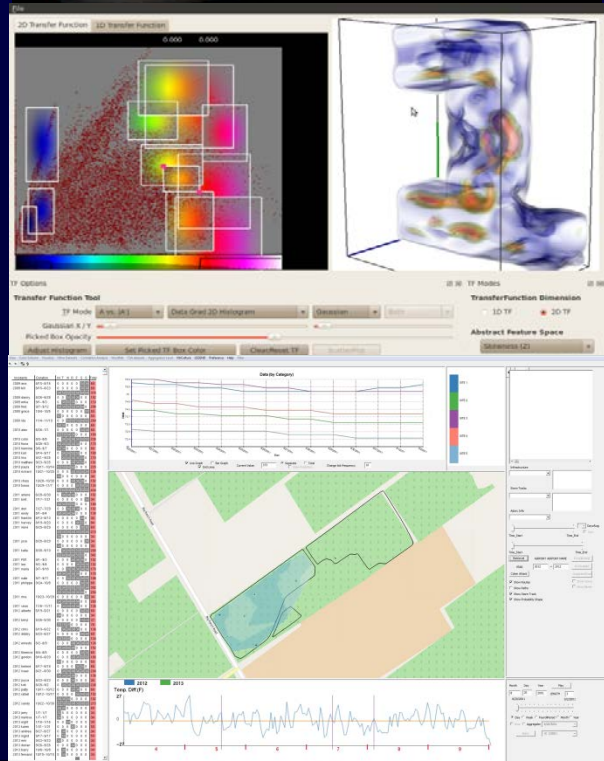
VACCINE

Scientific Discovery

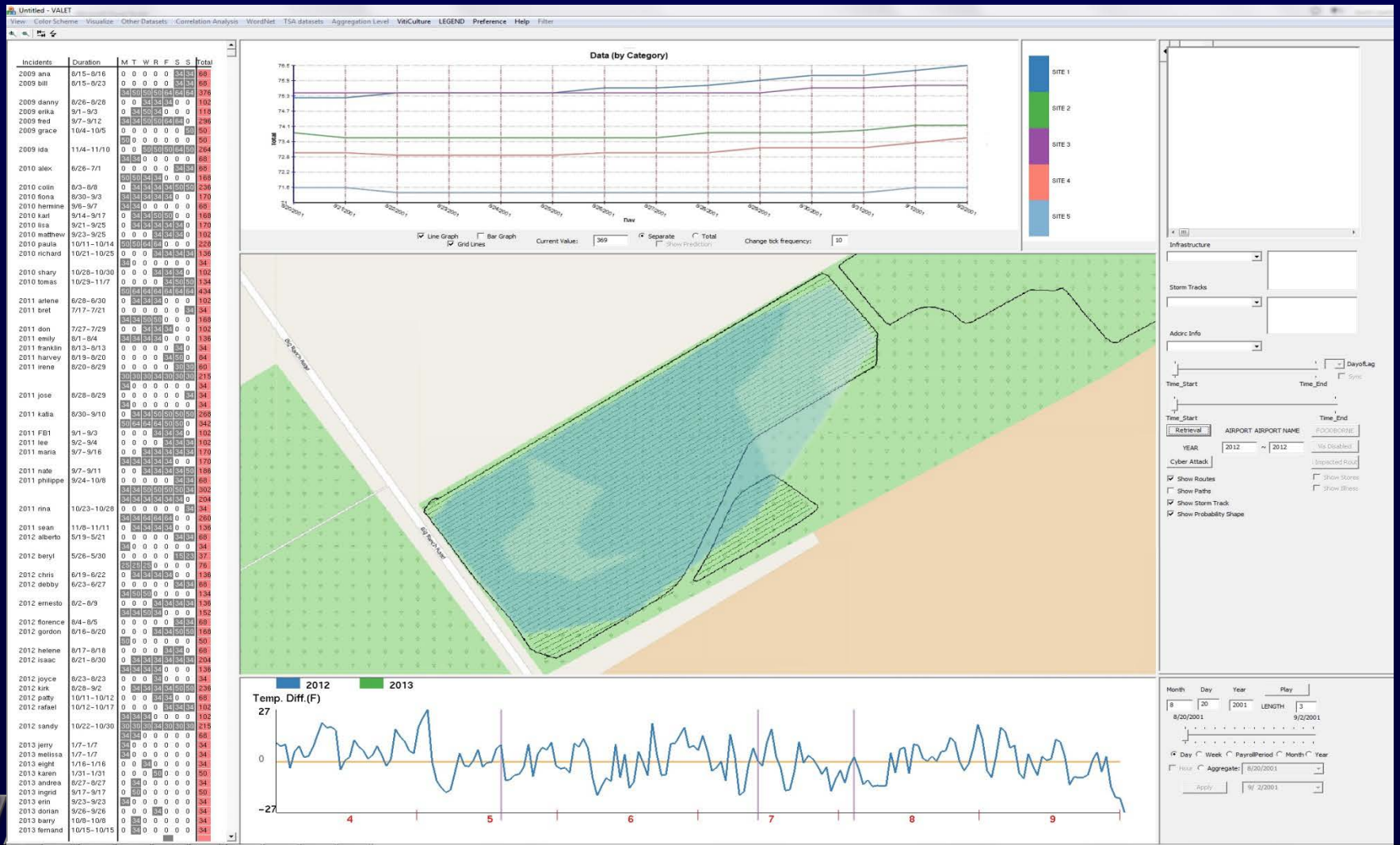


- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality



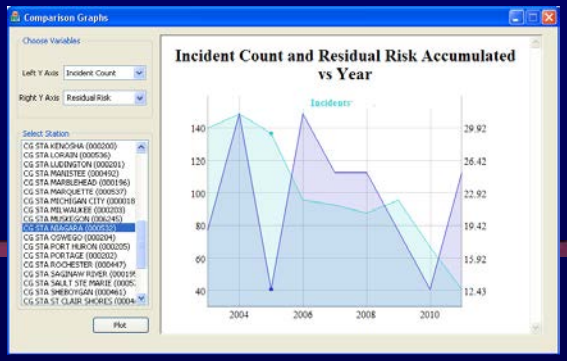


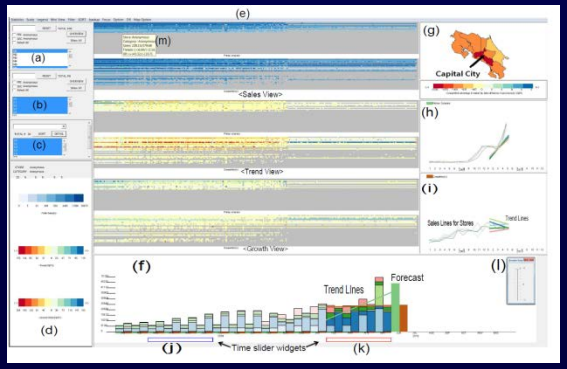
Precision Agriculture VA



Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation



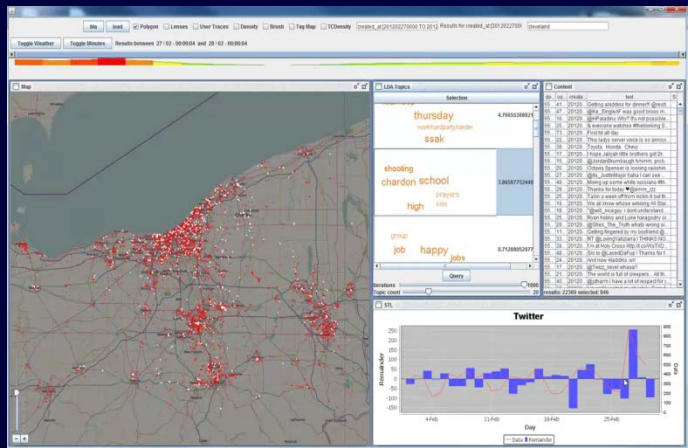




Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation

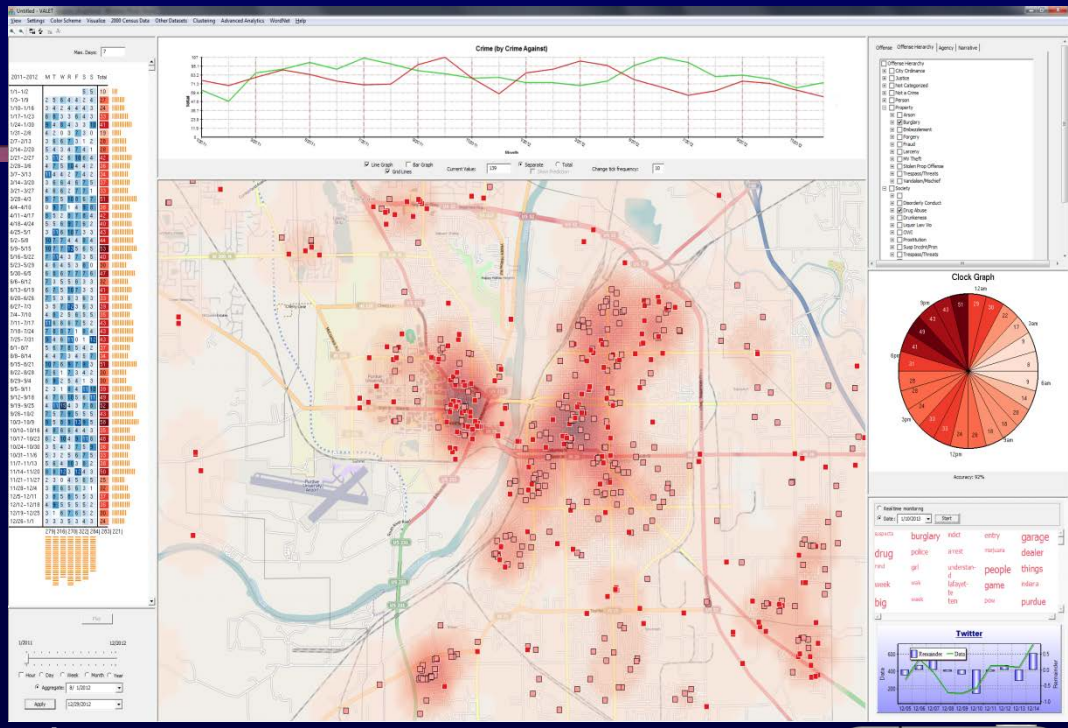


Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

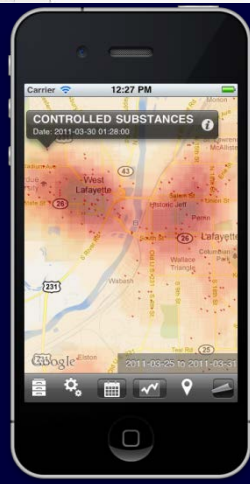
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

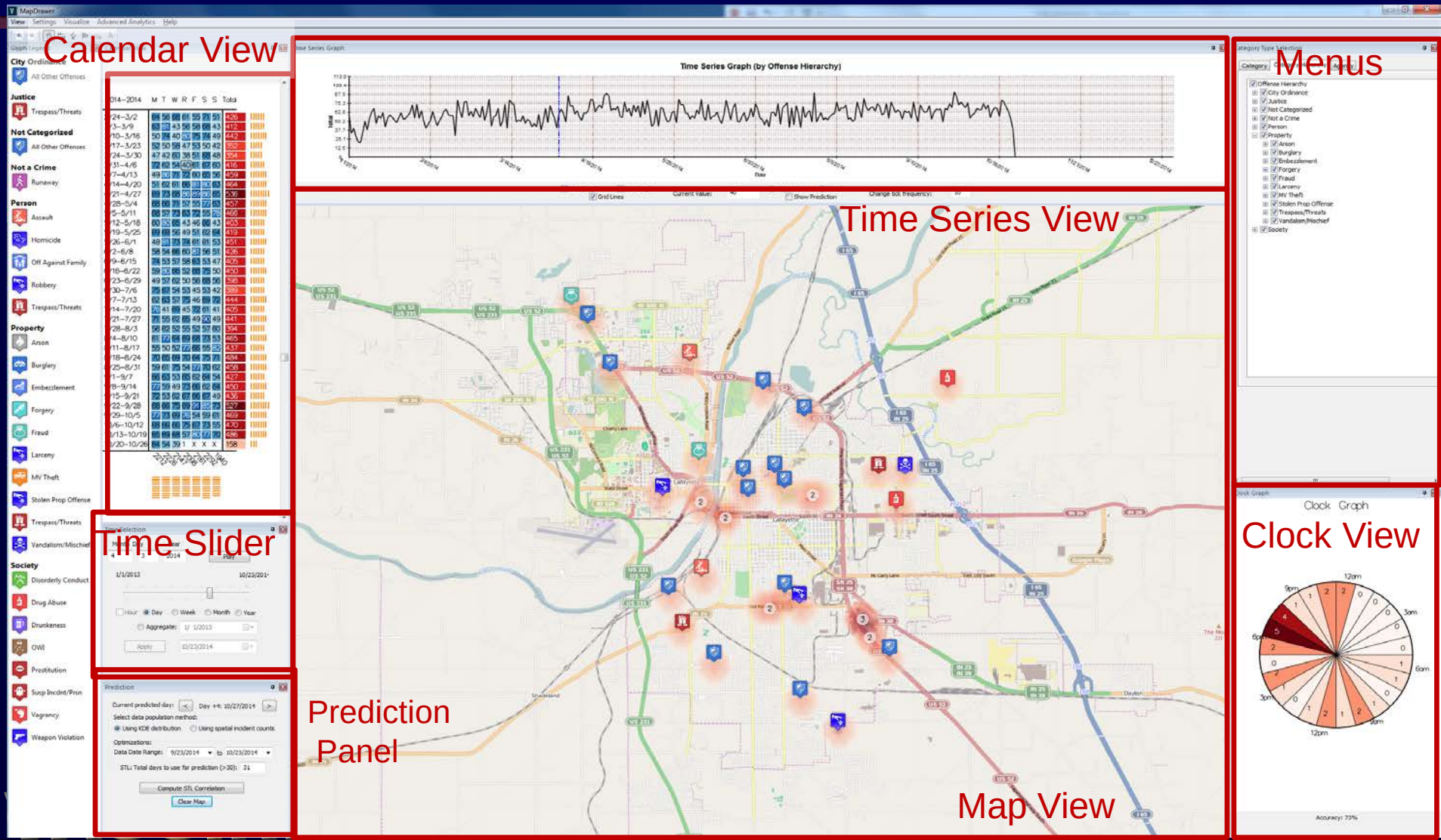
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

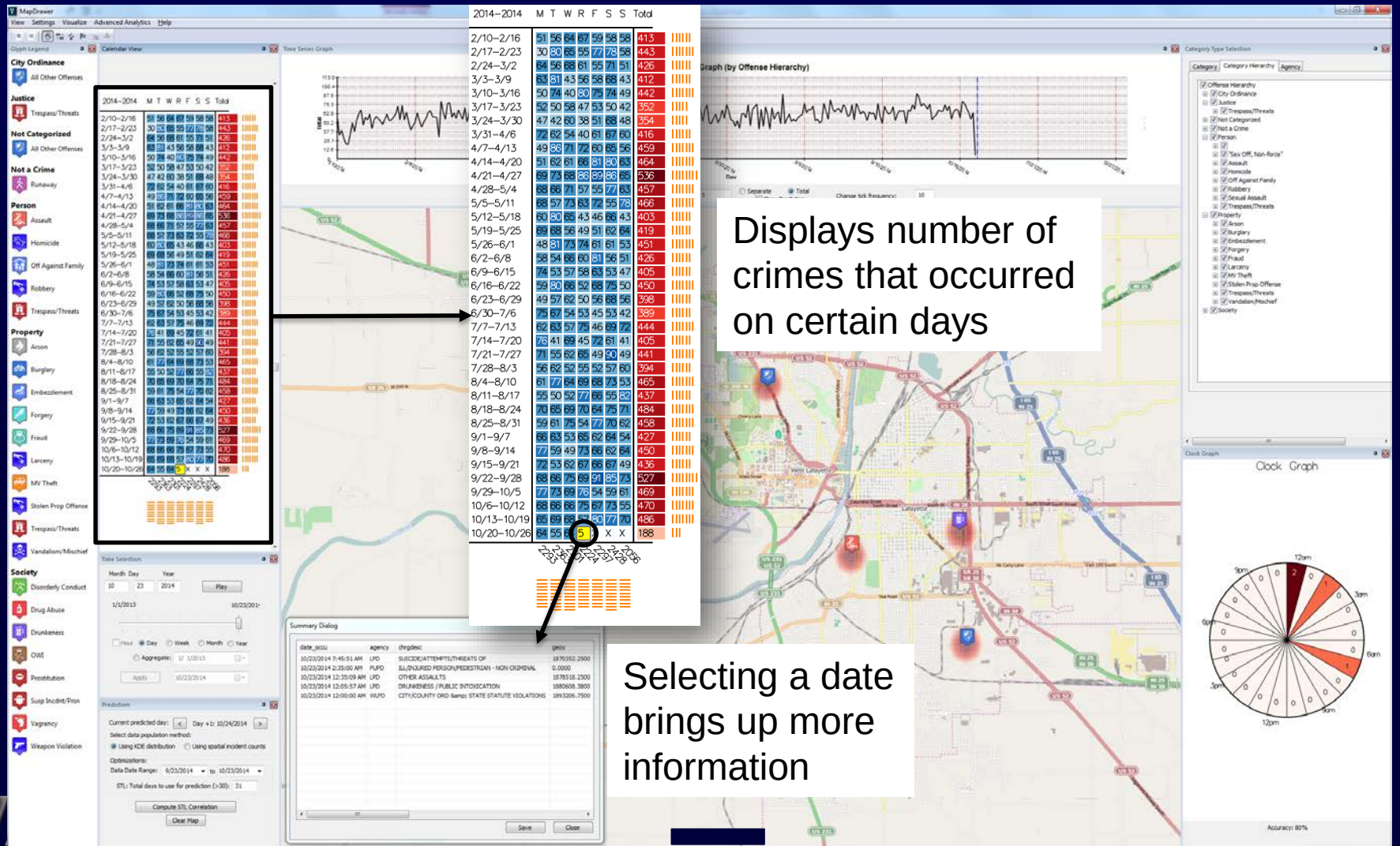


VACCINE

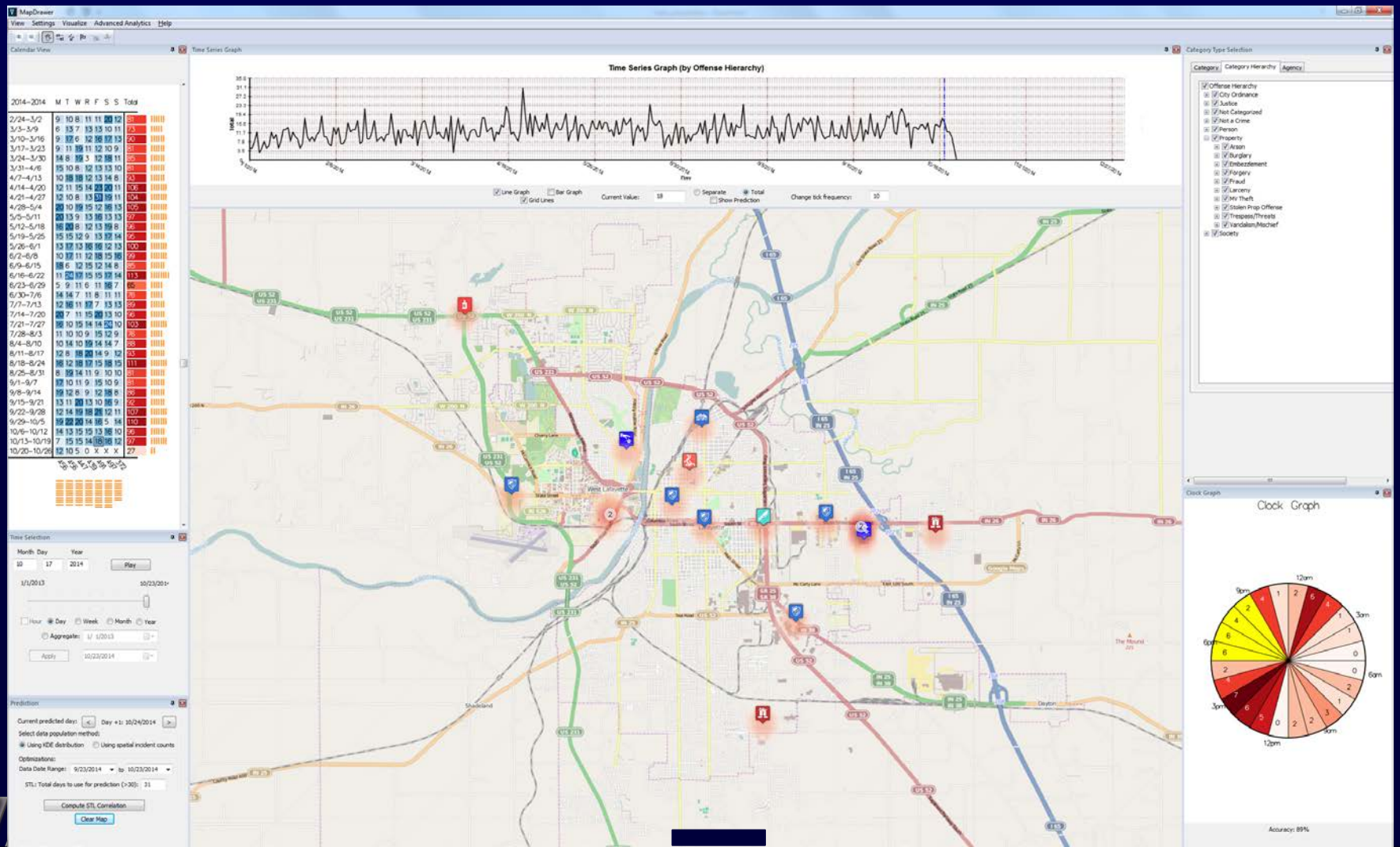
VALET Layout



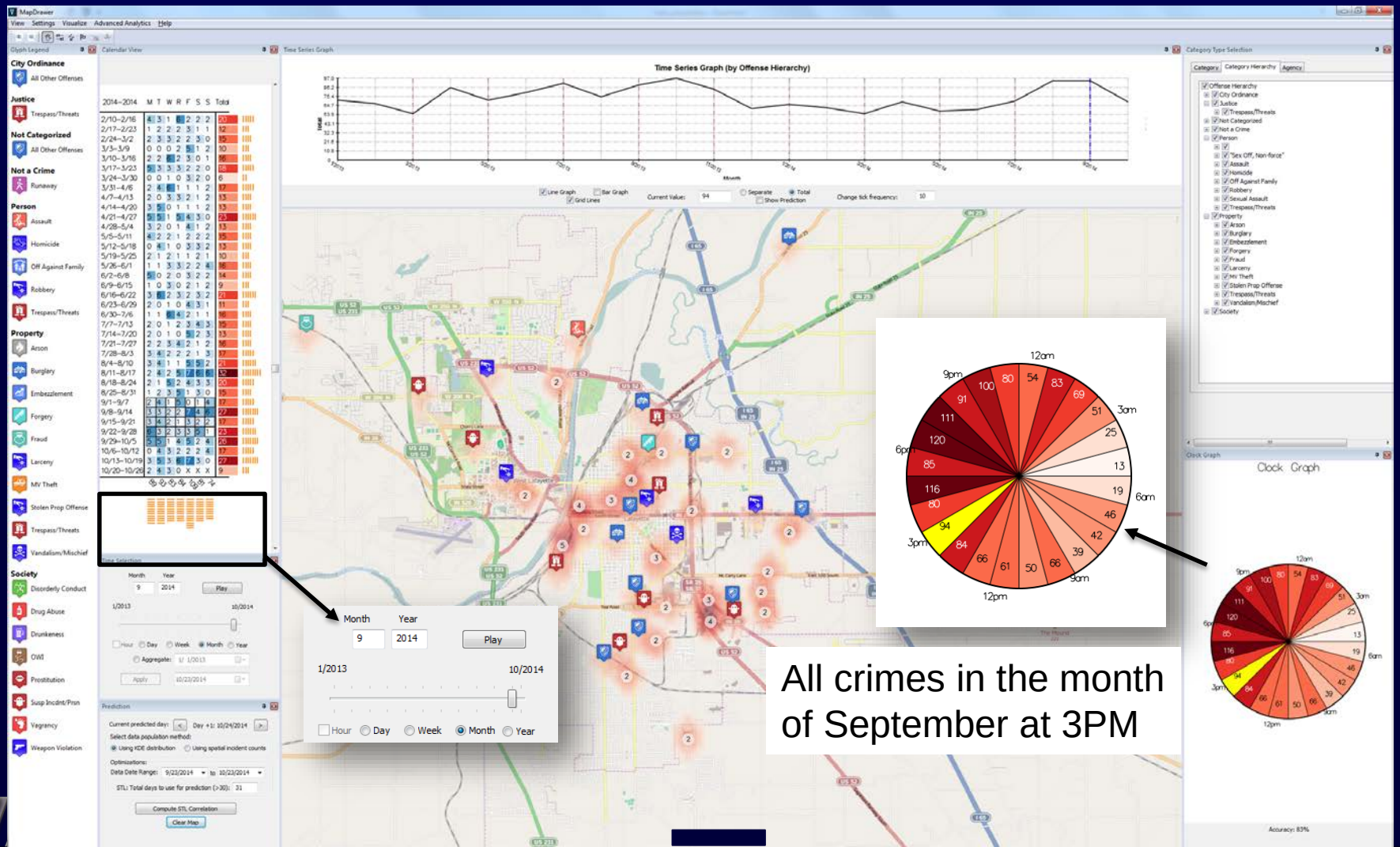
Calendar Display



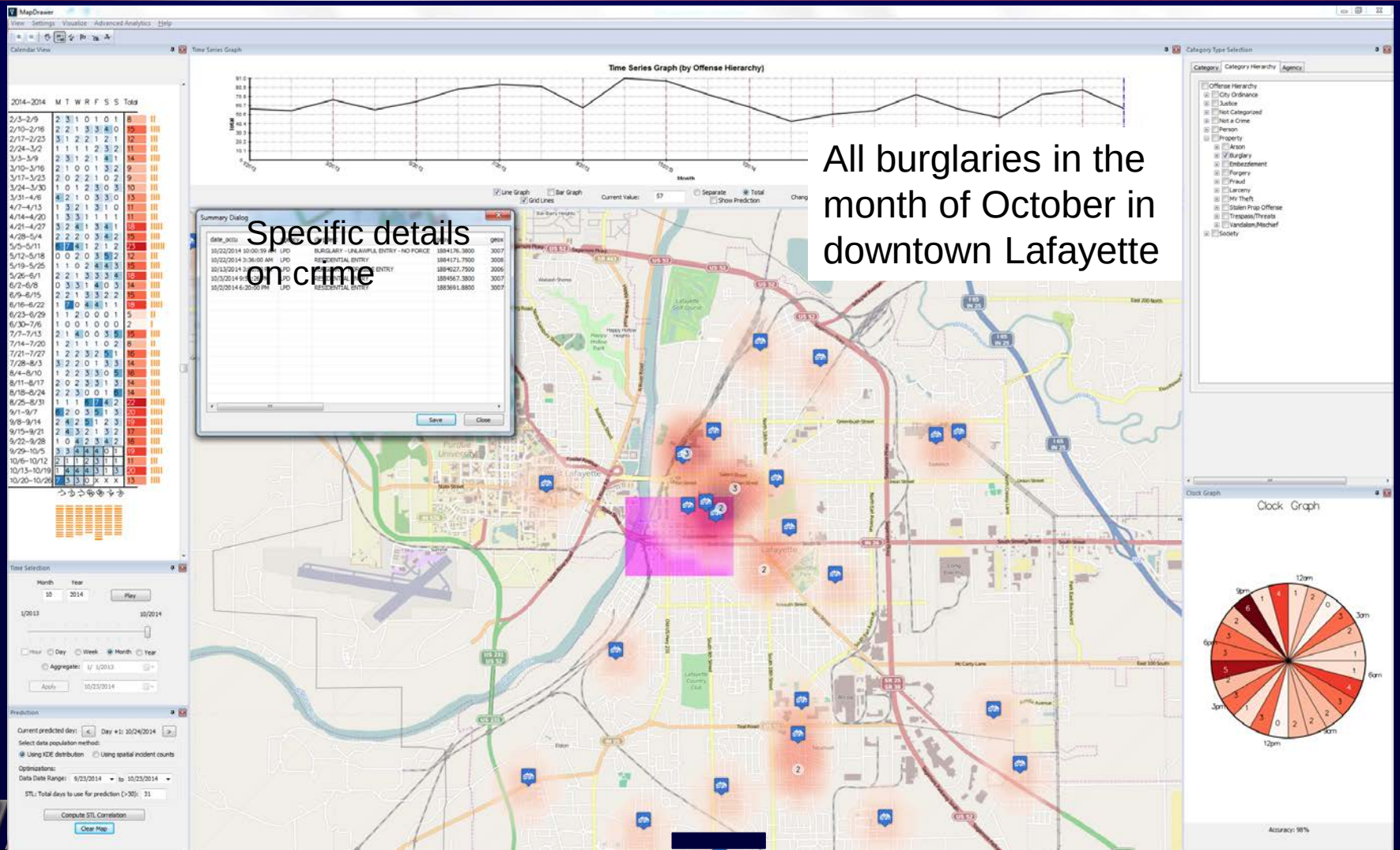
Crimes In A Single Day From 6PM – 9PM



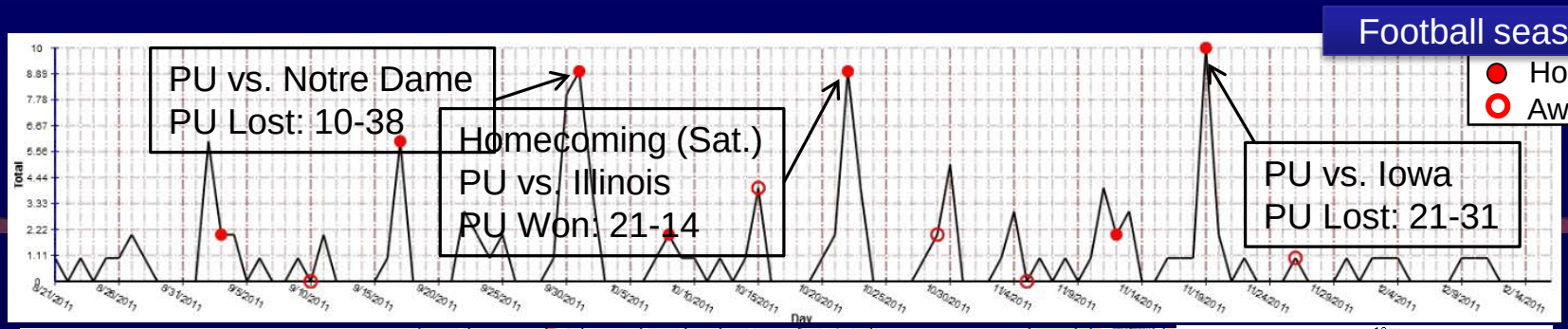
Crimes During Certain Hours In A Month

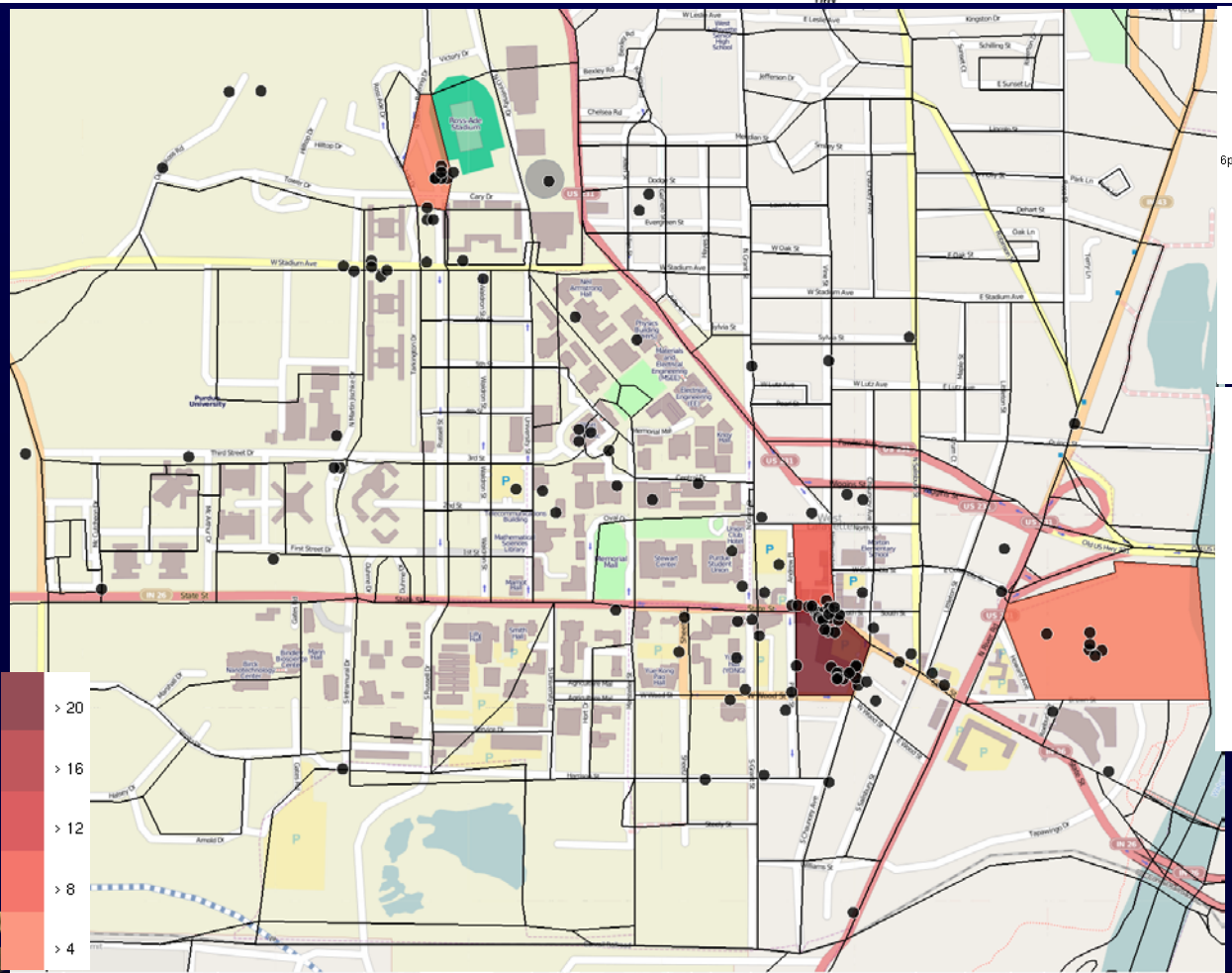


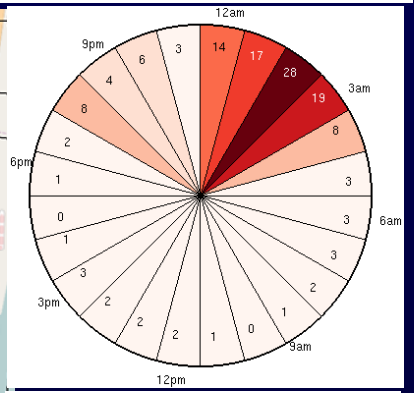
Select Specific Region To Display Crimes

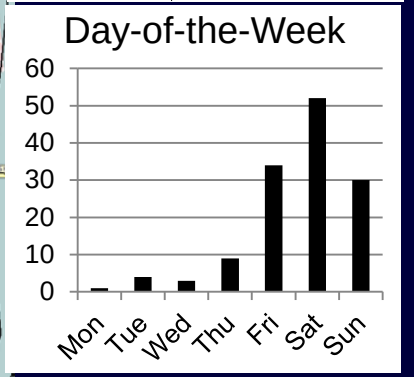


Example: More Data Sources – Intoxication

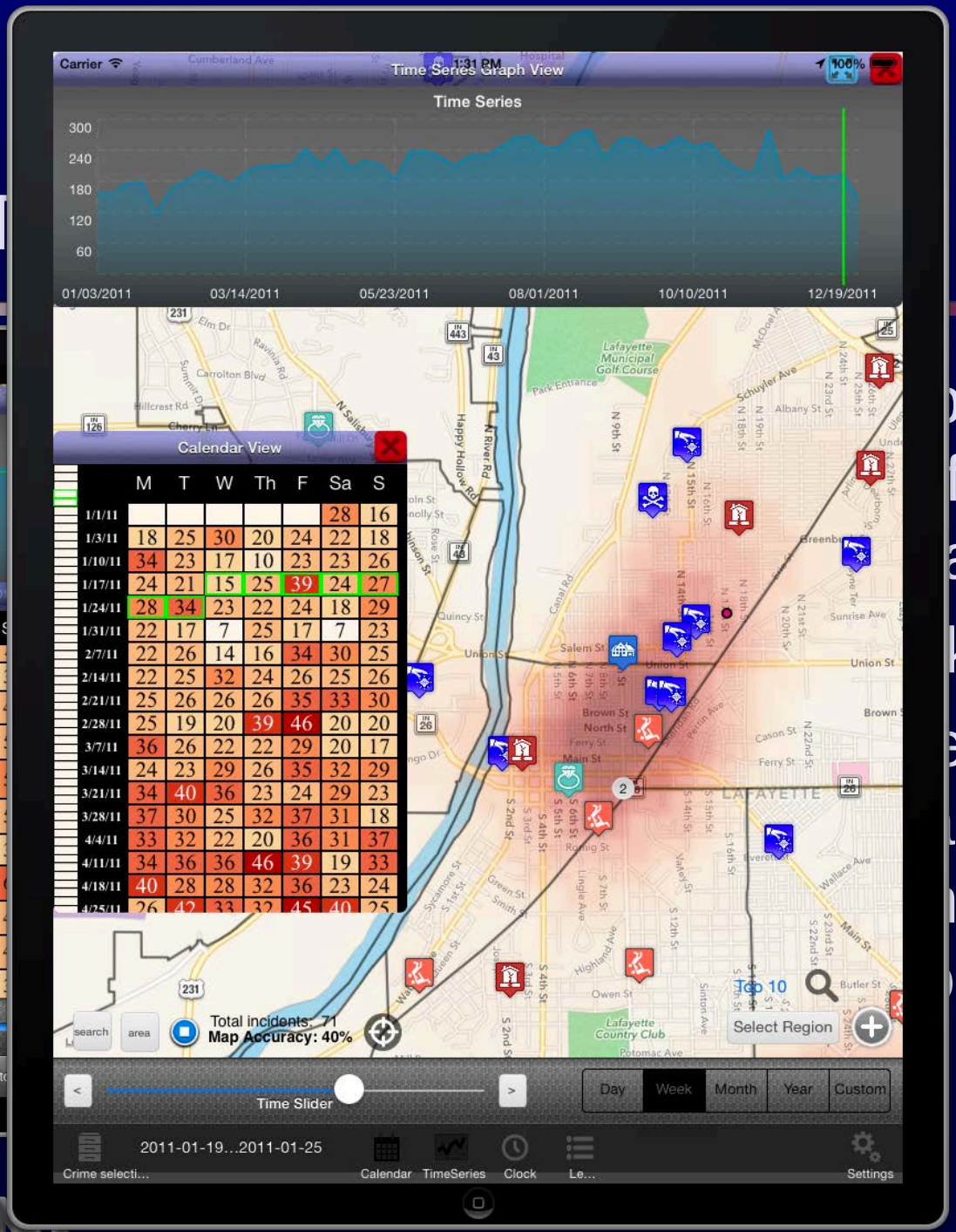




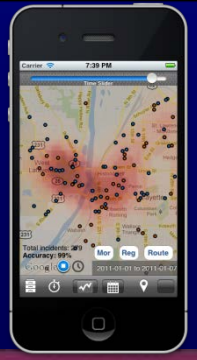


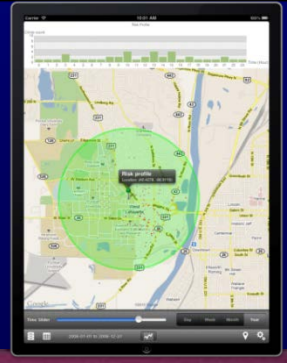


iVALET









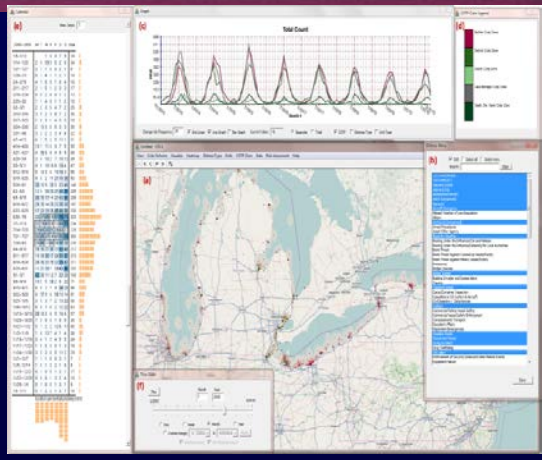


lore criminal,
fic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis

VACCINE

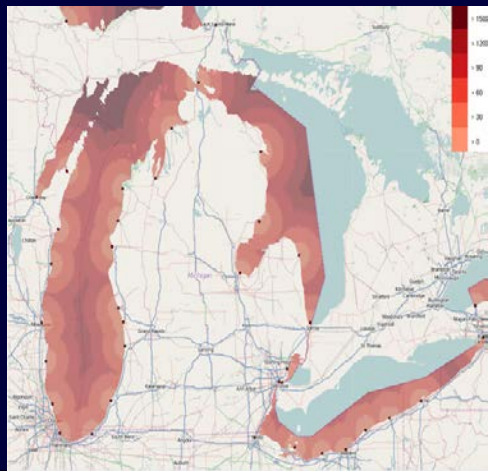
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



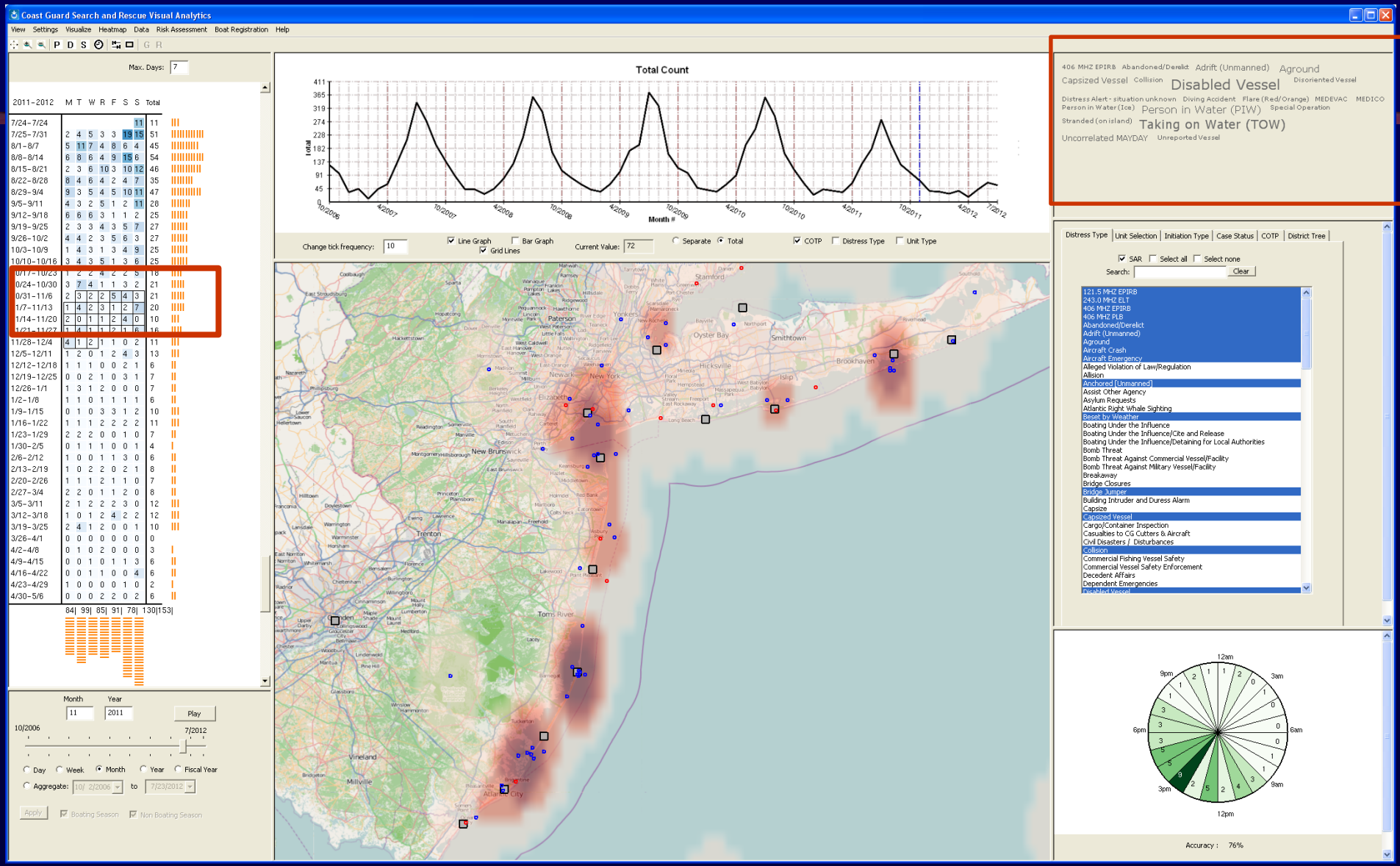
IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

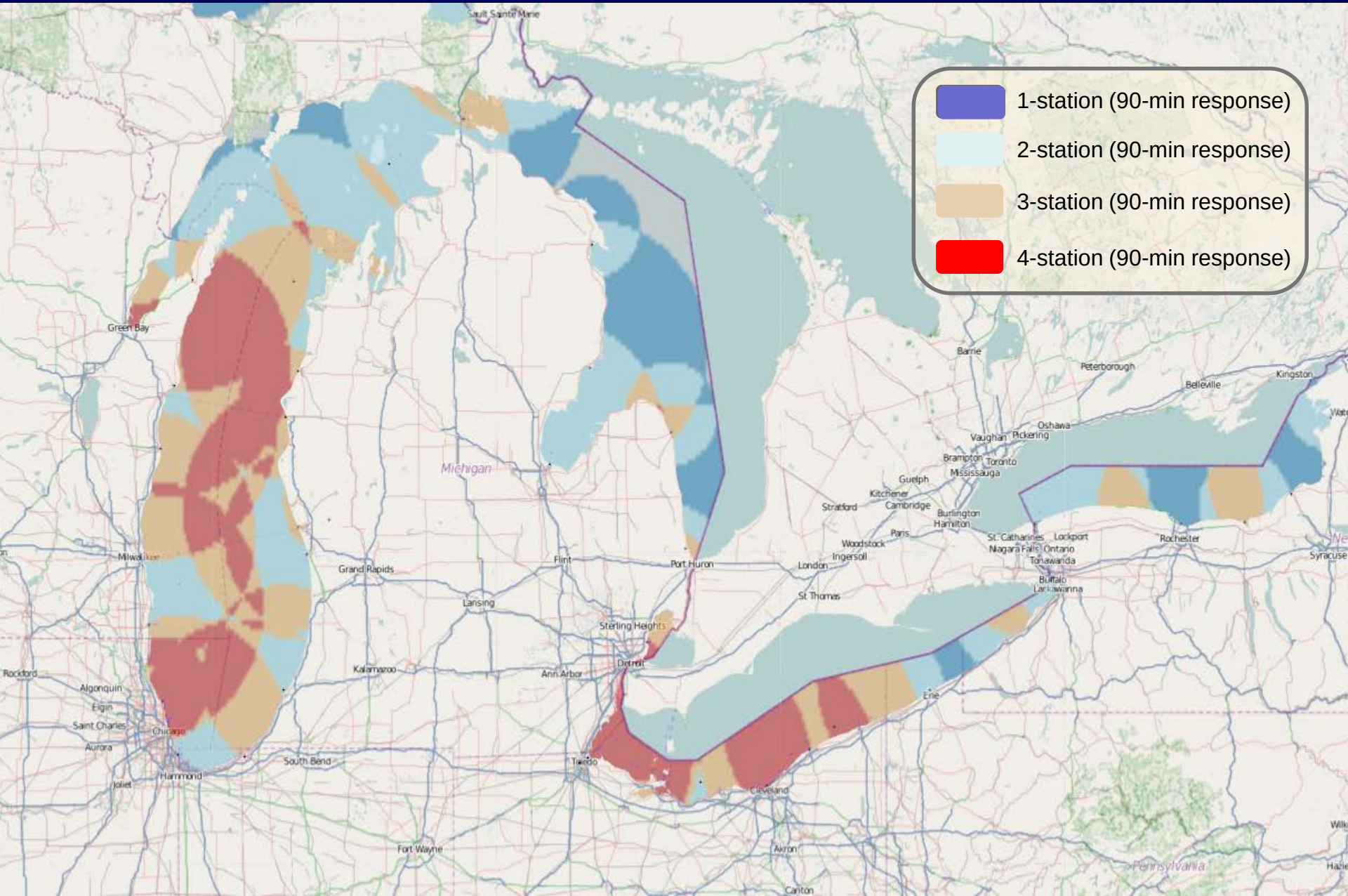


Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets

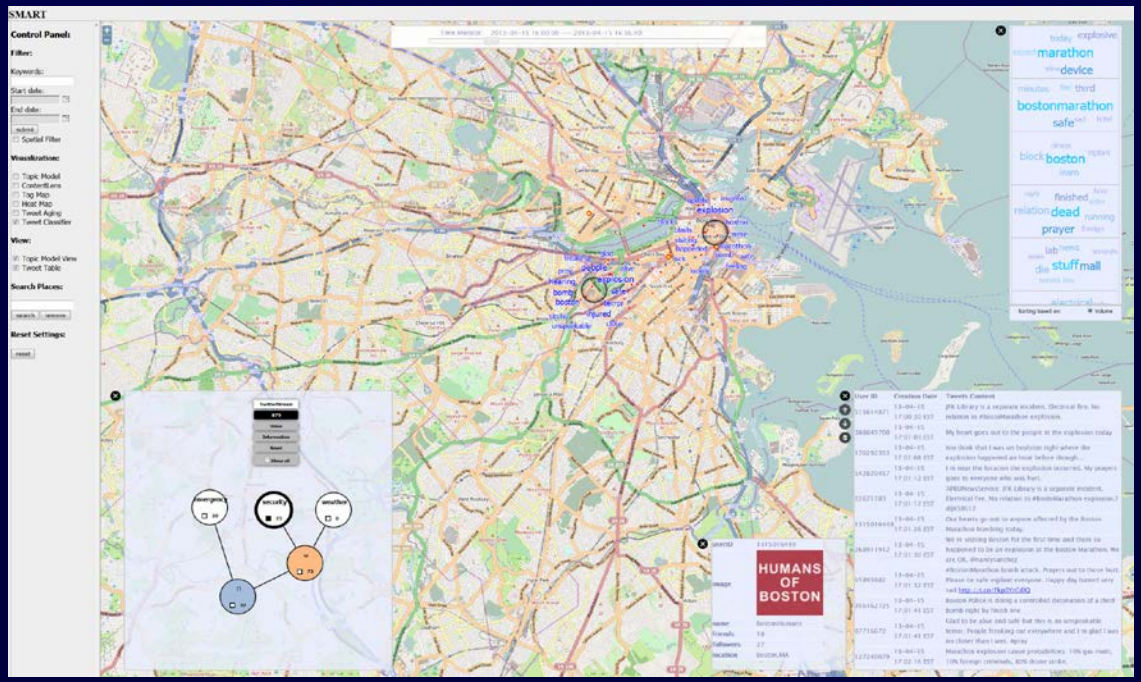


VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

February 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

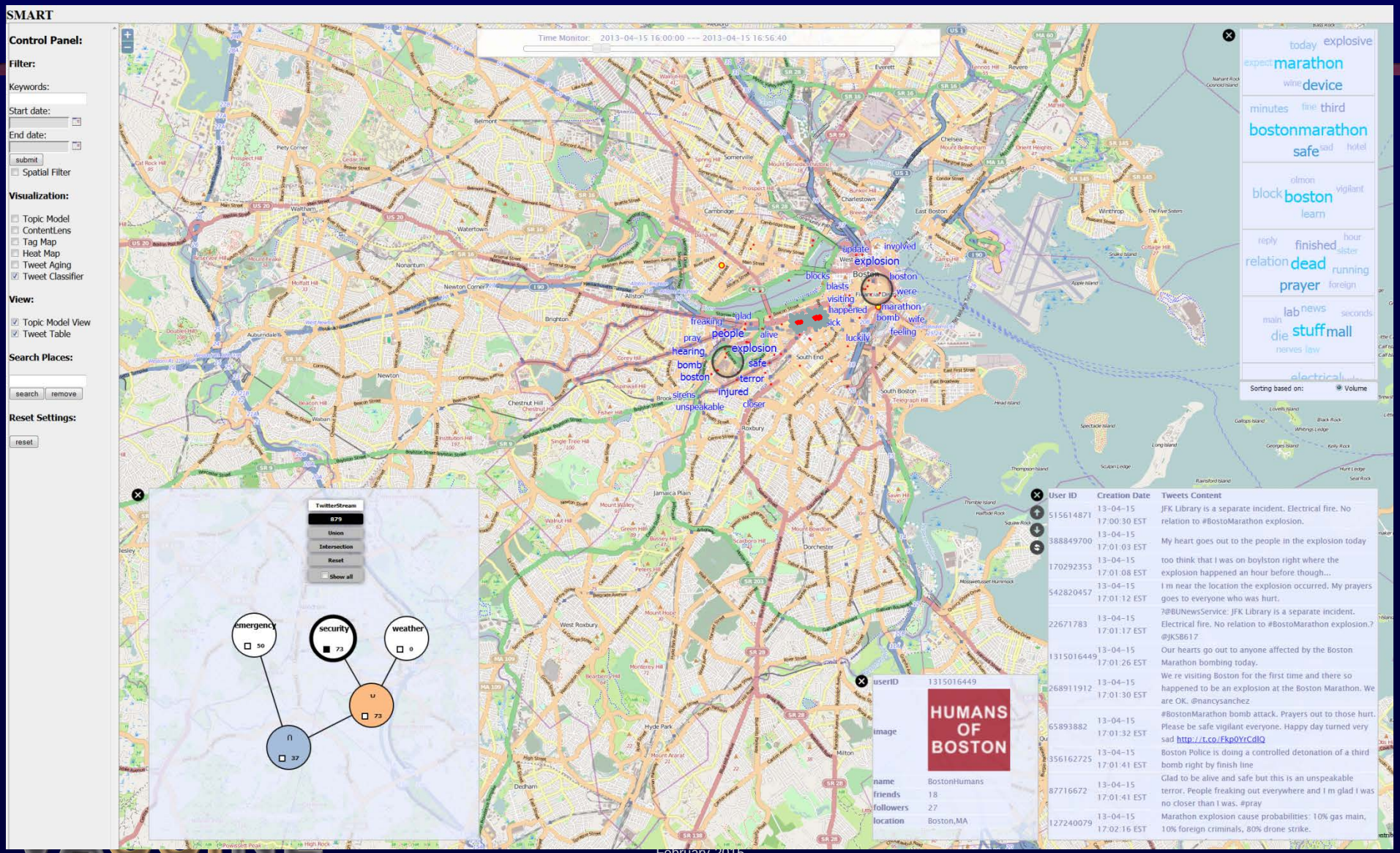
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

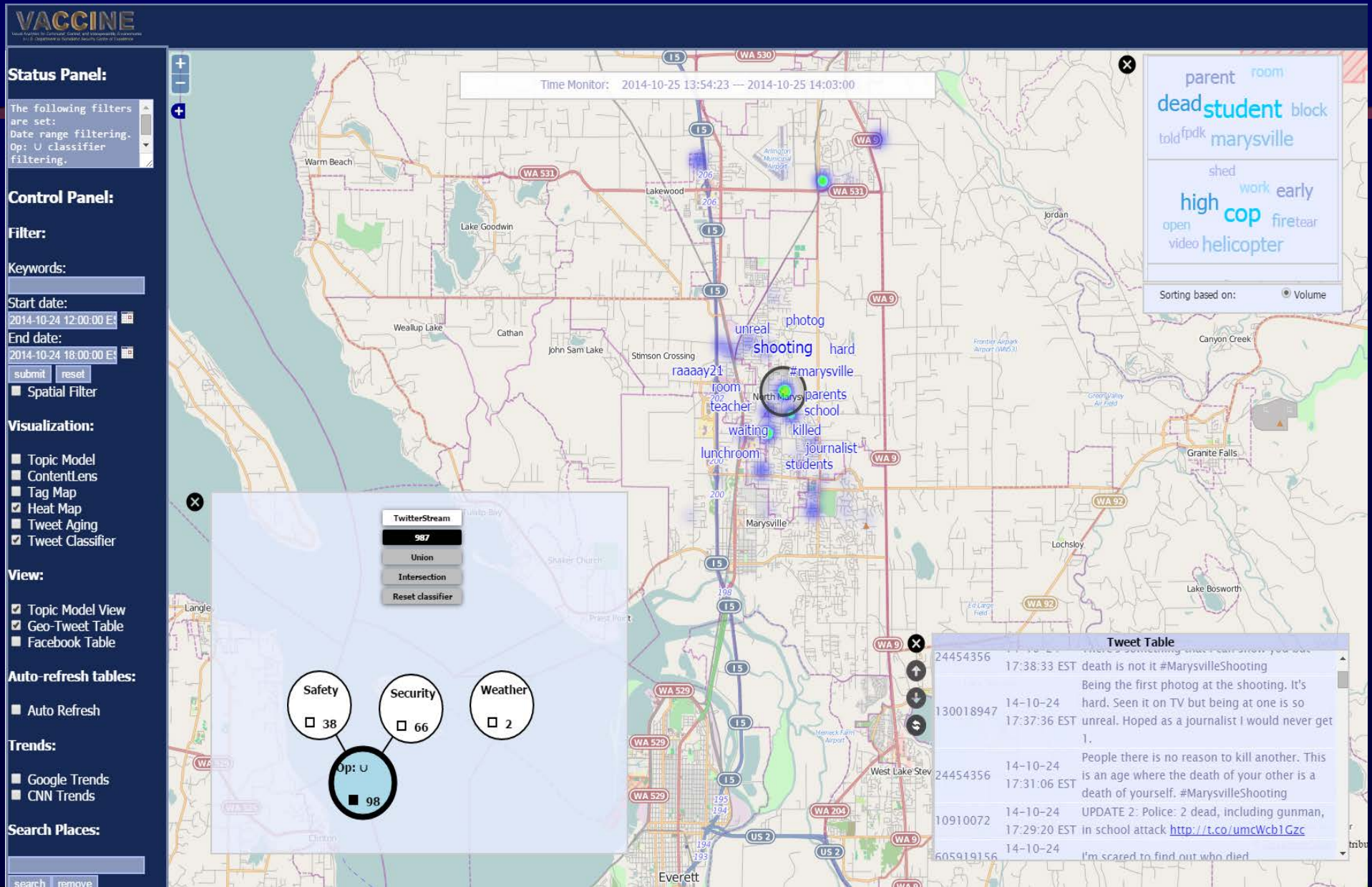


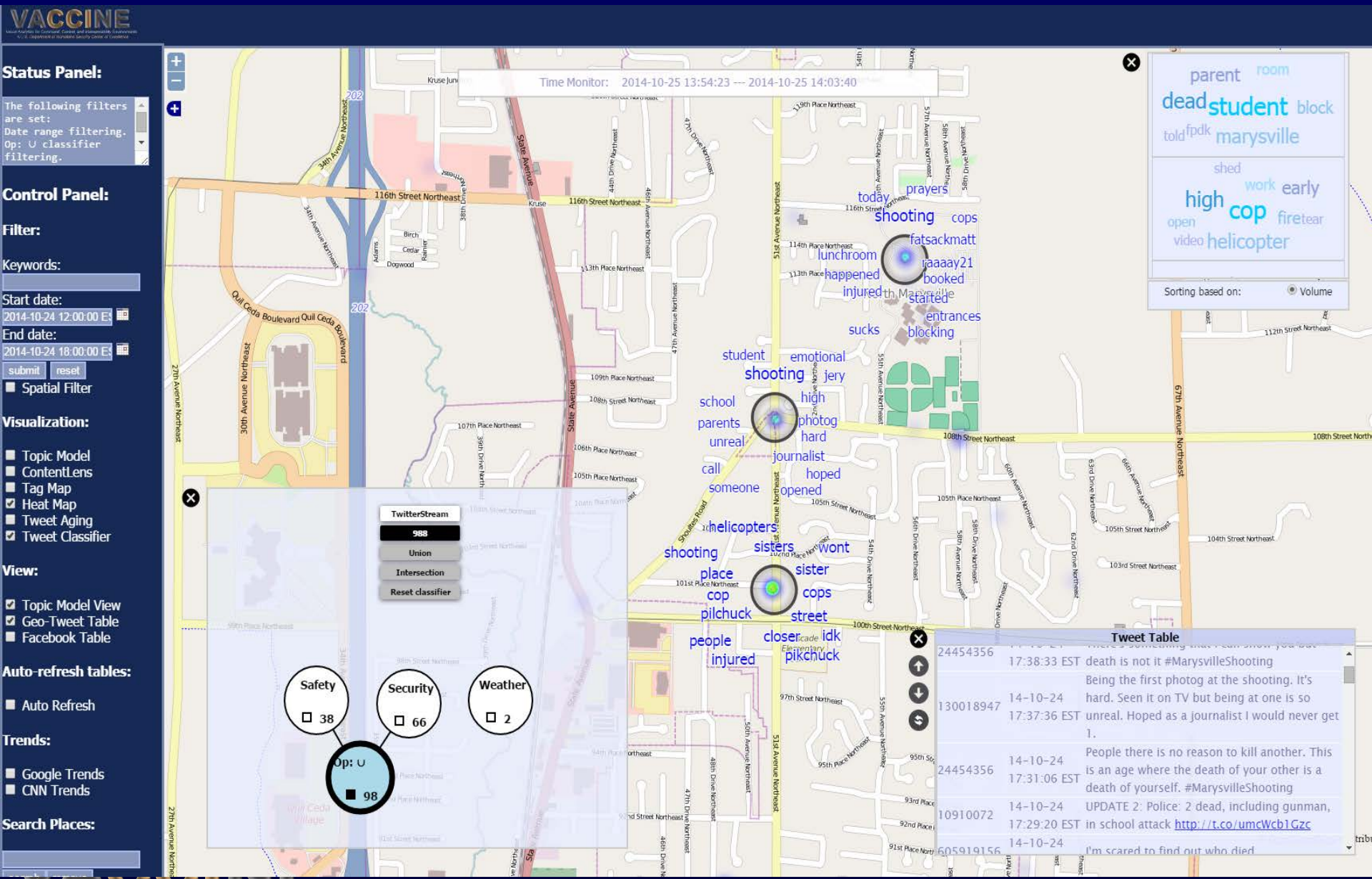
VACCINE

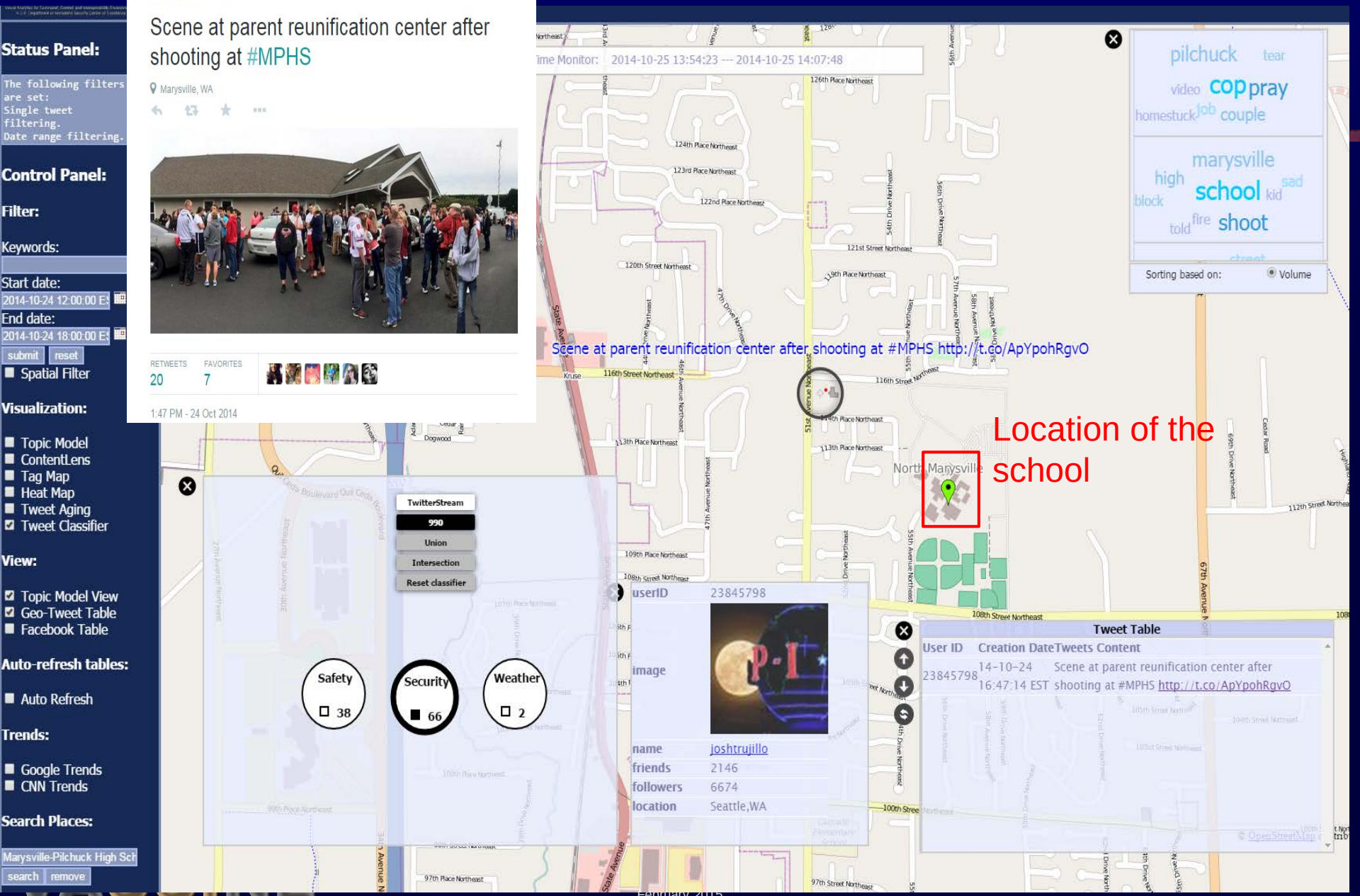
Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24 17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24 17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24 17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died

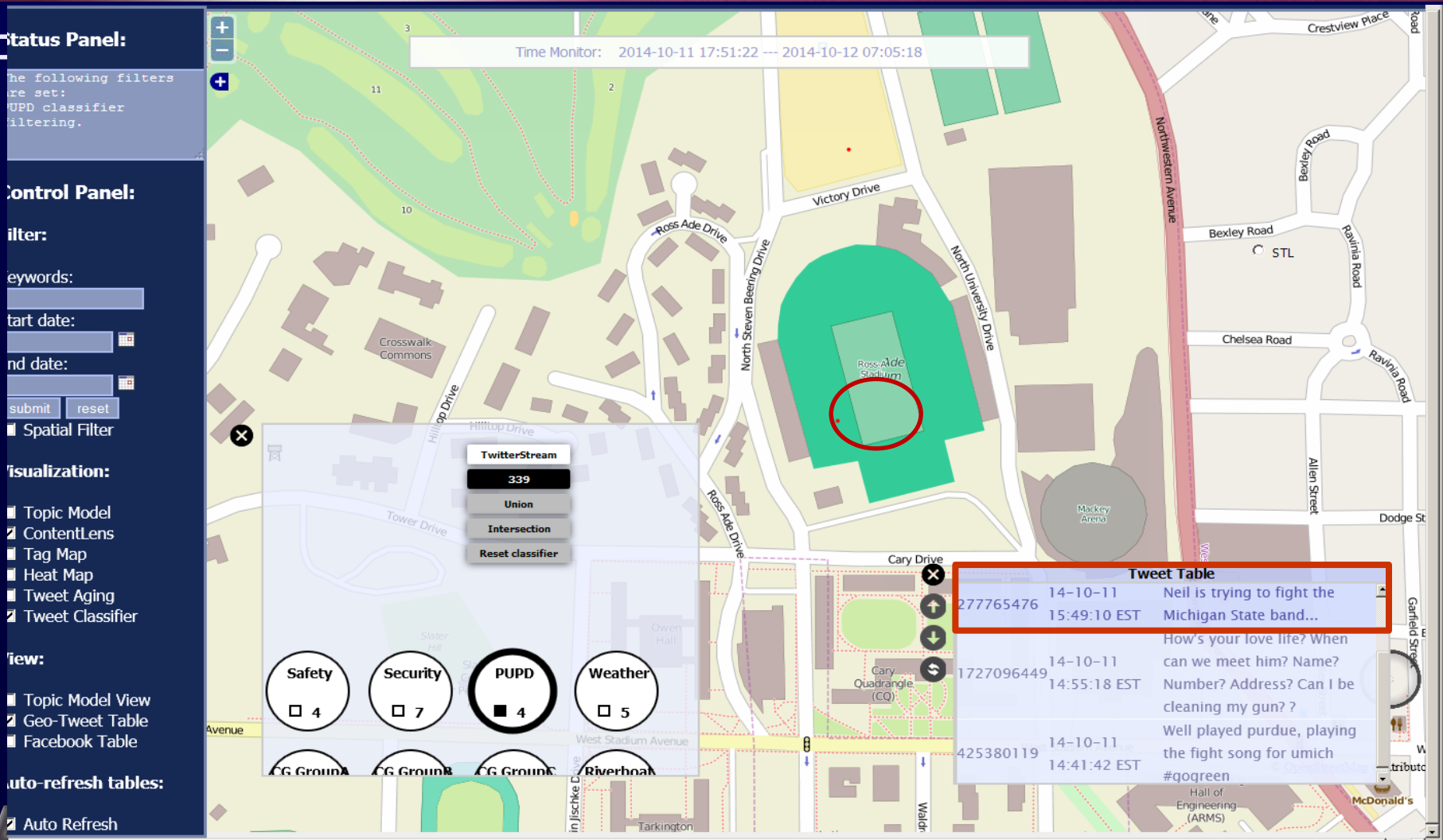

Joshua Trujillo
@joshtrujillo
Follow
Scene at parent reunification center after shooting at #MPHS
Marysville, WA
RETWEETS 20
FAVORITES 7
1:47 PM - 24 Oct 2014
Status Panel:
The following filters are set:
Single tweet filtering.
Date range filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 ES
End date:
2014-10-24 18:00:00 ES
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Marysville-Pilchuck High Sch
search
remove
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48
pilchuck
tear
video
cop
pray
homestuck
job
couple
marysville
high
school
kid
sad
block
told
fire
shoot
Sorting based on:
Volume
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO
Location of the school
North Marysville
TwitterStream
990
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
userID 23845798
image
name joshtrujillo
friends 2146
followers 6674
location Seattle,WA
Tweet Table
User ID
Creation Date
Tweets Content
23845798
14-10-24 16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

Where Can It Be Helpful?

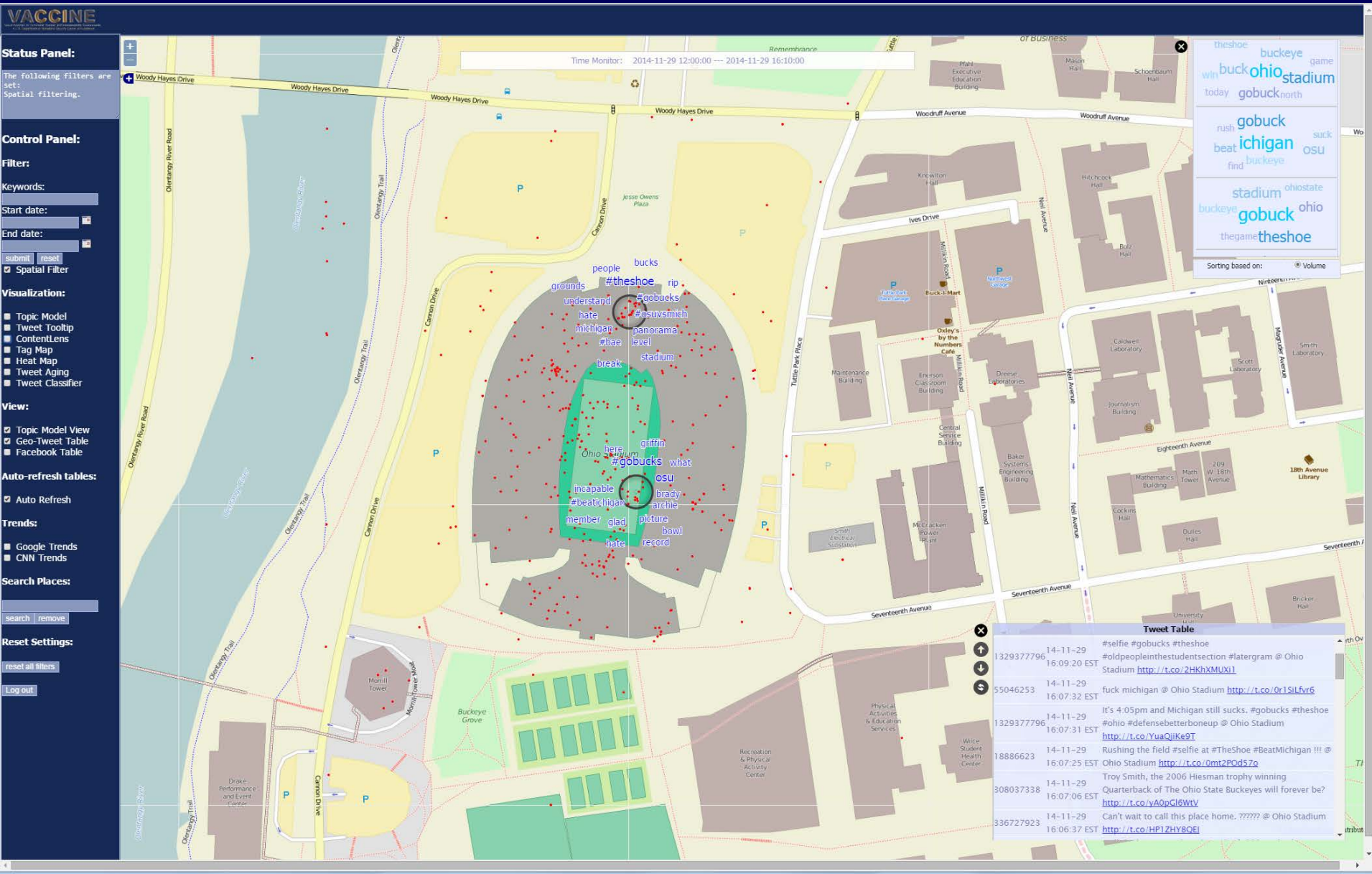
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



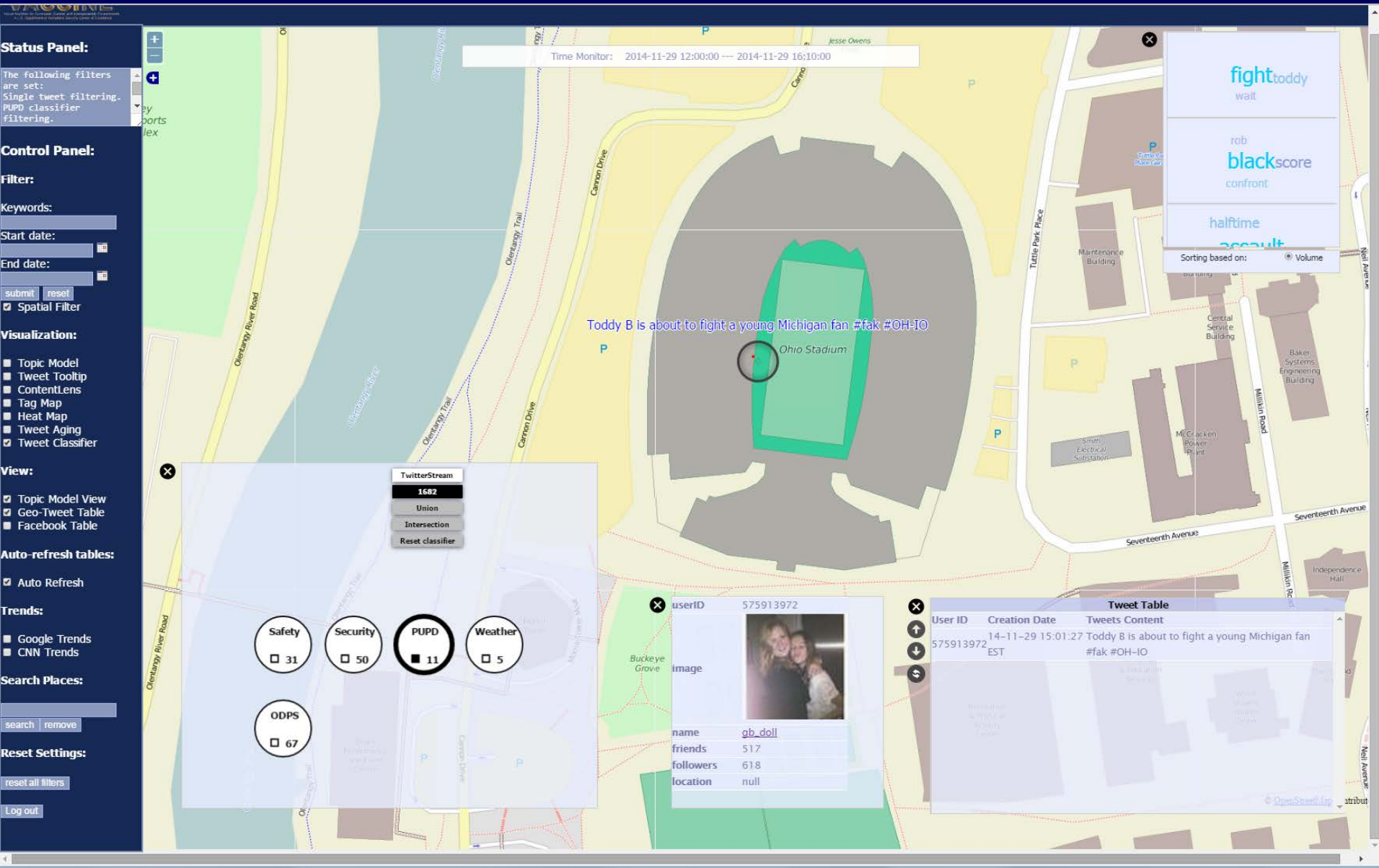
Example From Ohio State Football 11/29/14



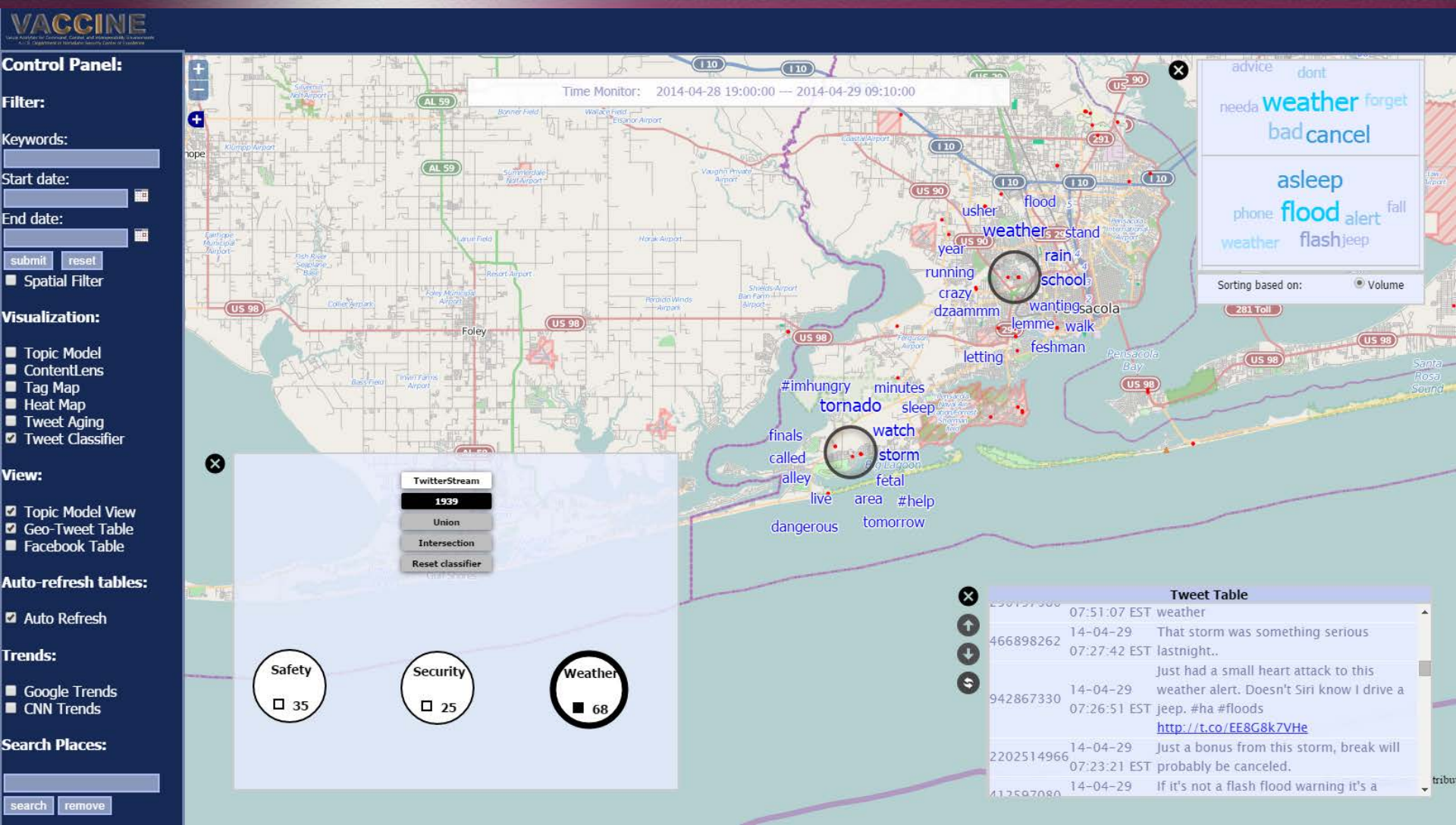
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



Where Can It Be Helpful?

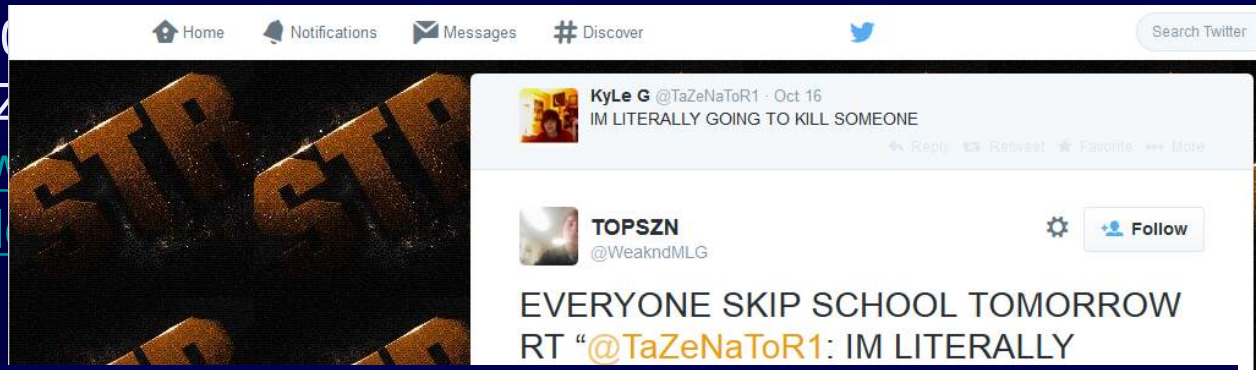


Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of California San Diego**
- **University Washington**
- **Arizona State University**
- **Bethune-Cookman U.**
- **Simon Fraser University**
- **University of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **Oak Ridge National Laboratories**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

Some of Our Engaged End-Users

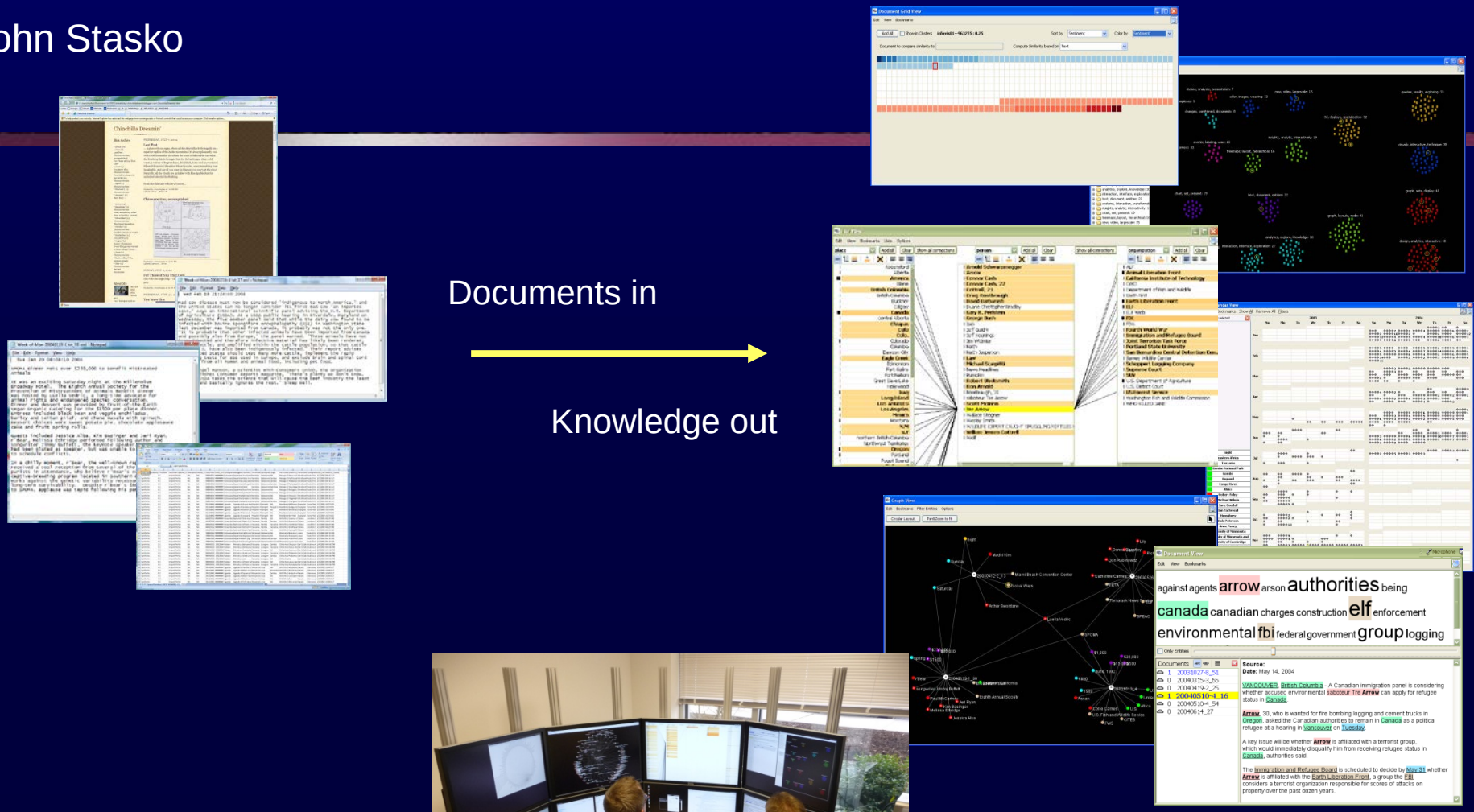
- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America



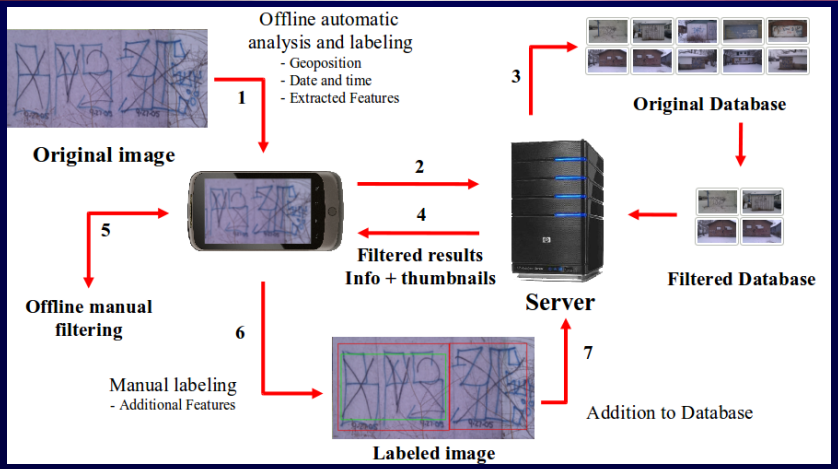


Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

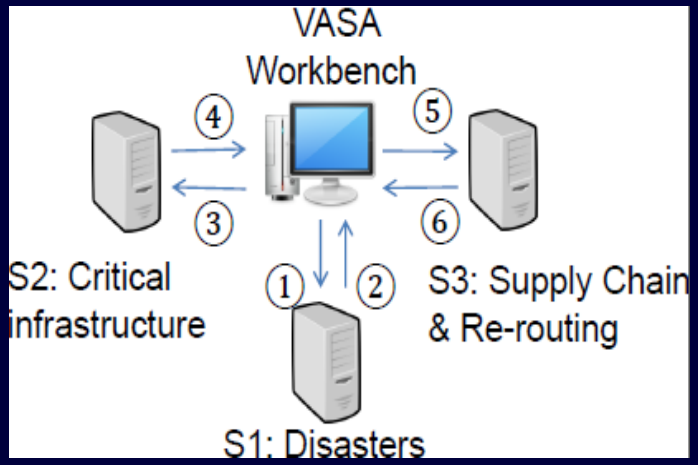
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

TACC

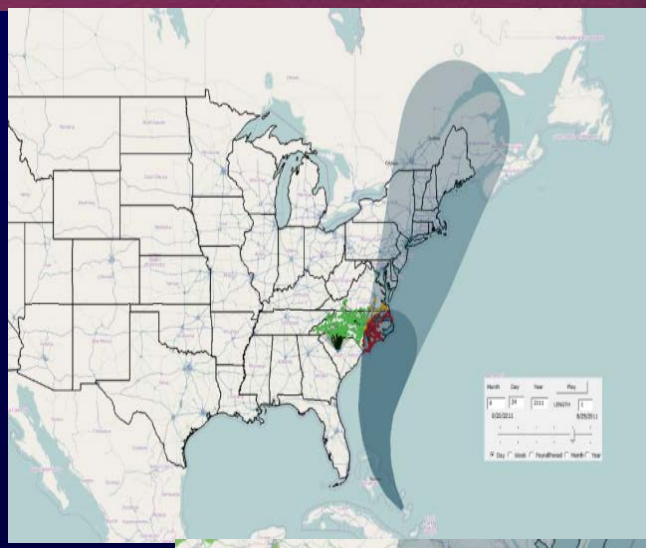
VACCINE

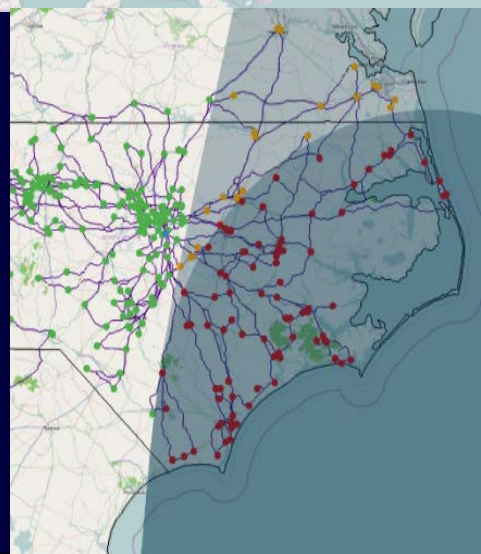
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





For Further Information

www.VisualAnalytics-CCI.org



Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Beyond Big Data Analytics: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

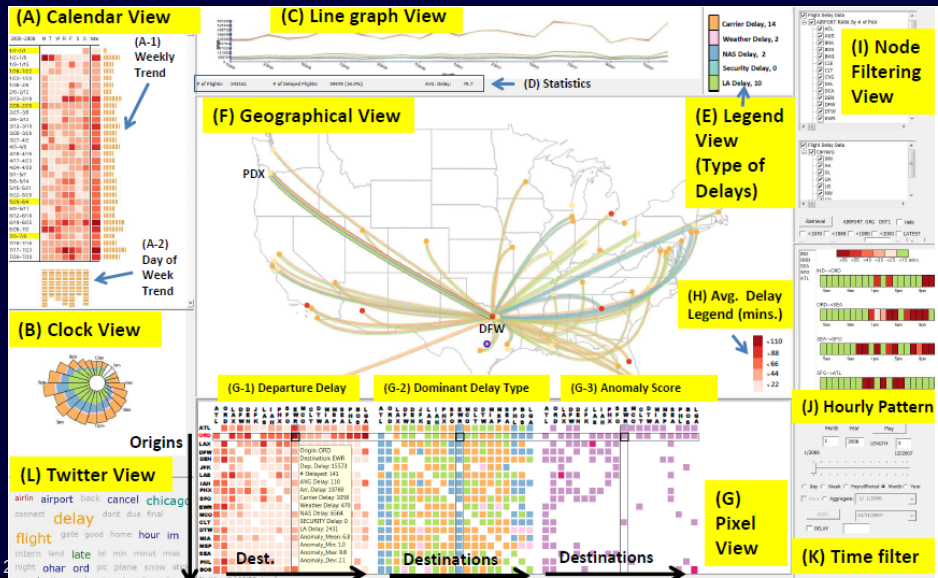
Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

December 2014

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Research approach
- Example projects
- Path forward



VACCINE

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

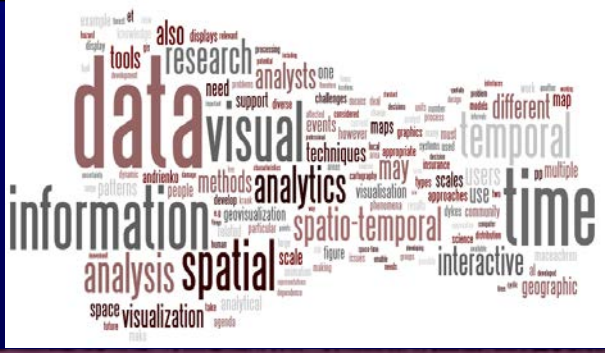
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America



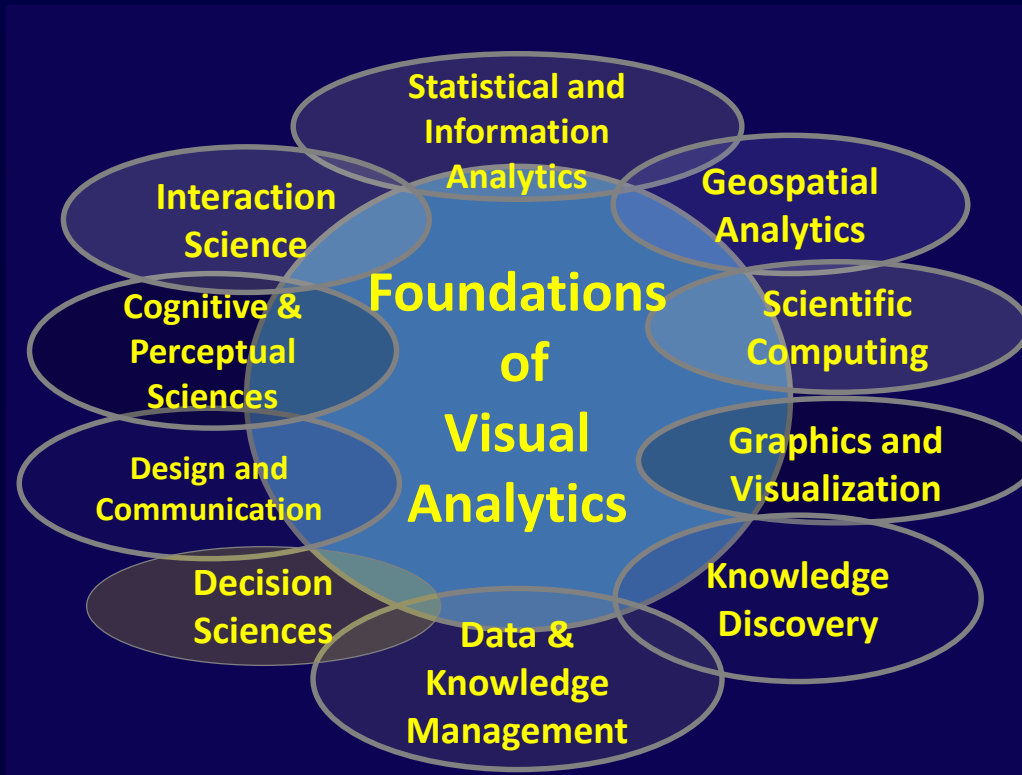


What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

Big Data



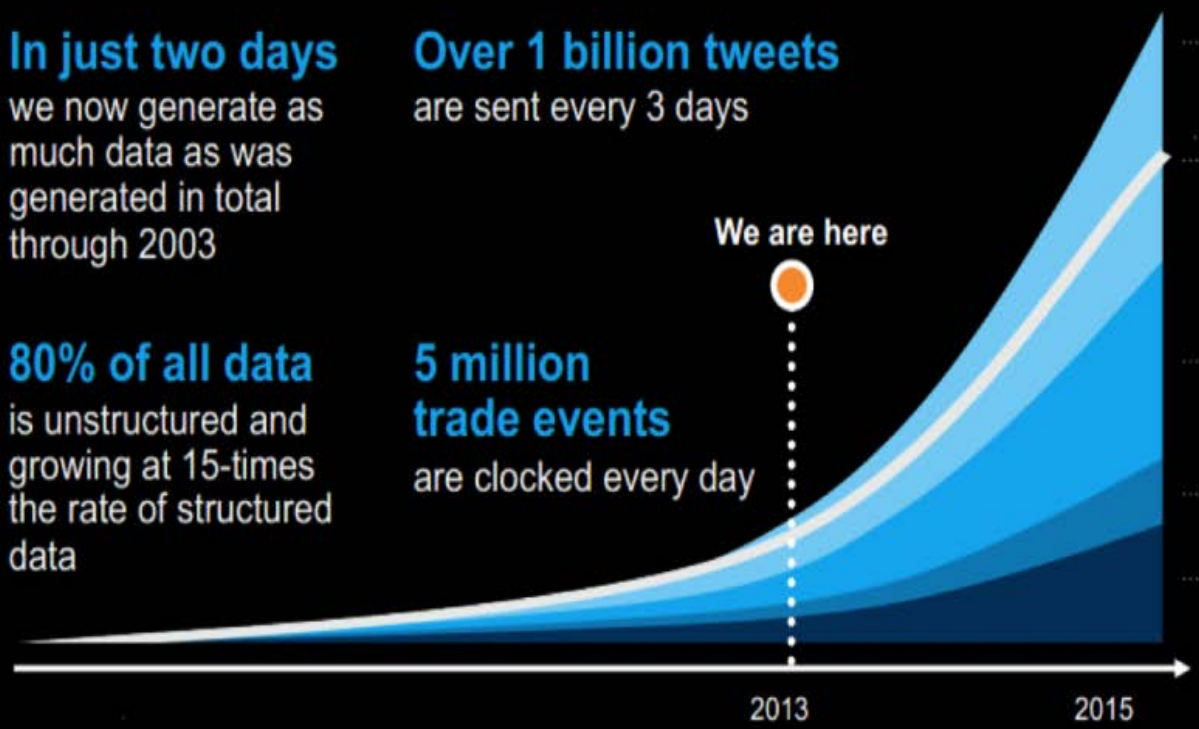
These forces fuel an explosion of data – a new economic asset forming the basis of opportunity

In just two days
we now generate as much data as was generated in total through 2003

Over 1 billion tweets
are sent every 3 days

80% of all data
is unstructured and growing at 15-times the rate of structured data

5 million trade events
are clocked every day



Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

Our Visual Analytics Process

(extended and adapted from Bertin, Normal, Tversky)

Process:

- Define problem / question
- Determine:
  - Workflow and interaction
  - Relevant data
  - Data characteristics (e.g., types, quality)
  - Abstraction/representation level
  - Natural scales
- Map to appropriate visual representation for task

- Utilize iterative refinement with user feedback from the start

Guiding Principles:

- Appropriateness principle
  - Neither more nor less
- Naturalness principle
  - Match with cognitive models
- Matching principle
  - Match the task, offer action affordances
- Principle of apprehension
  - Content should be accurately and easily perceived

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

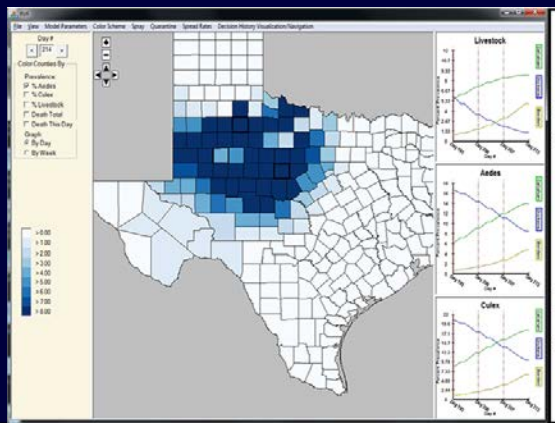
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

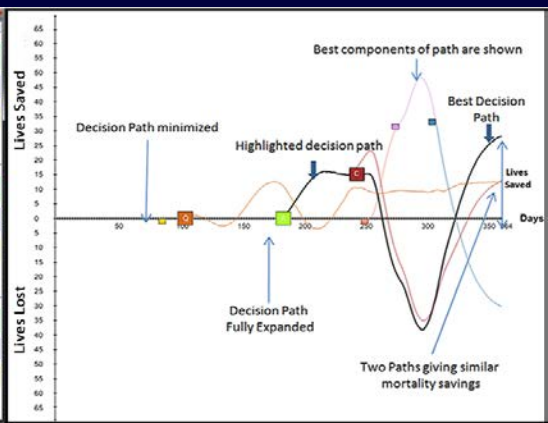
VADM Robert Parker, 2012 MRS Keynote Address



Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Research topics
- **Example application areas and projects**
- Path forward

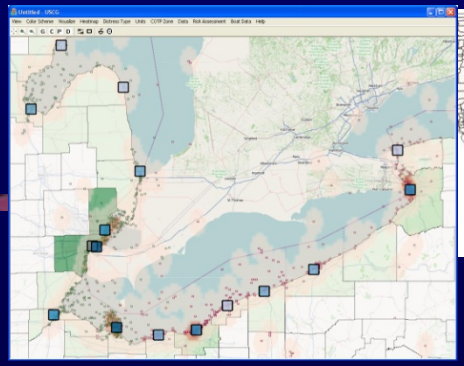


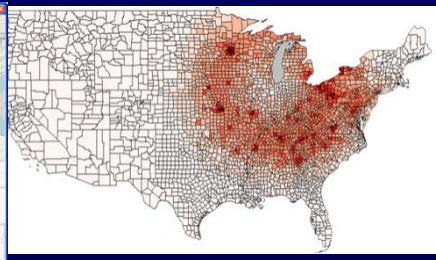


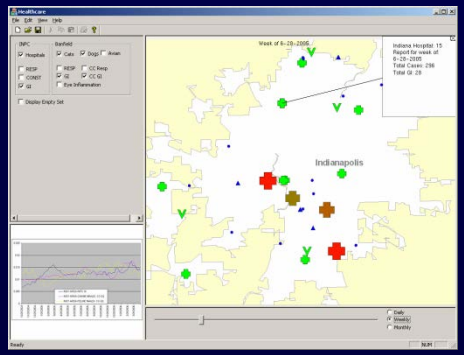
VACCINE

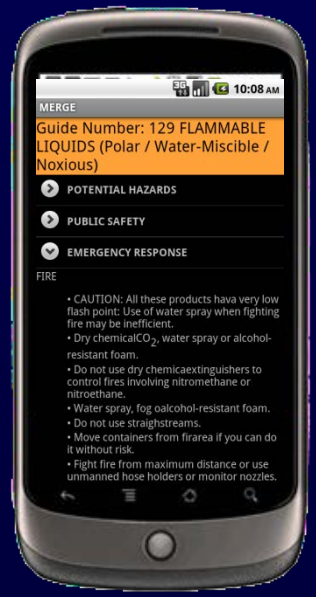
Public Safety & Health

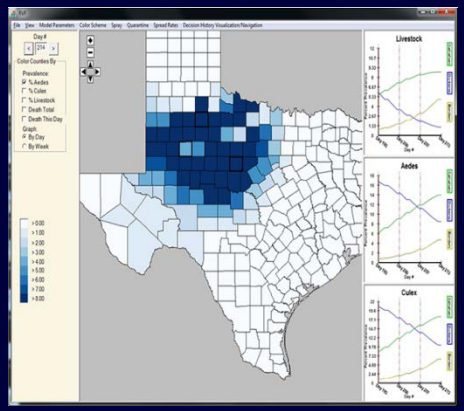
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering







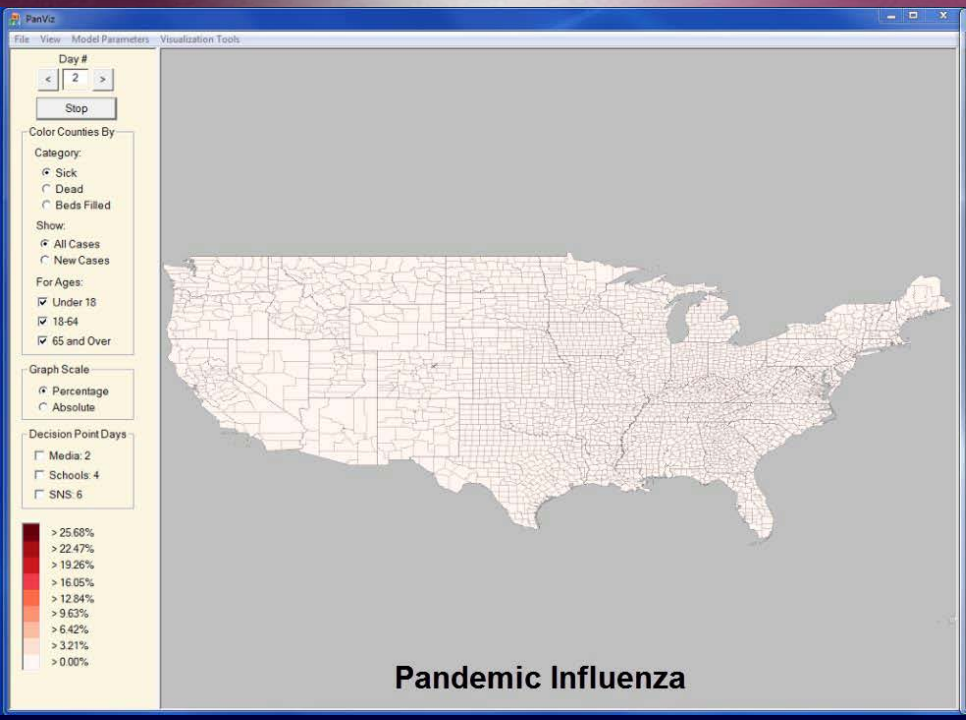


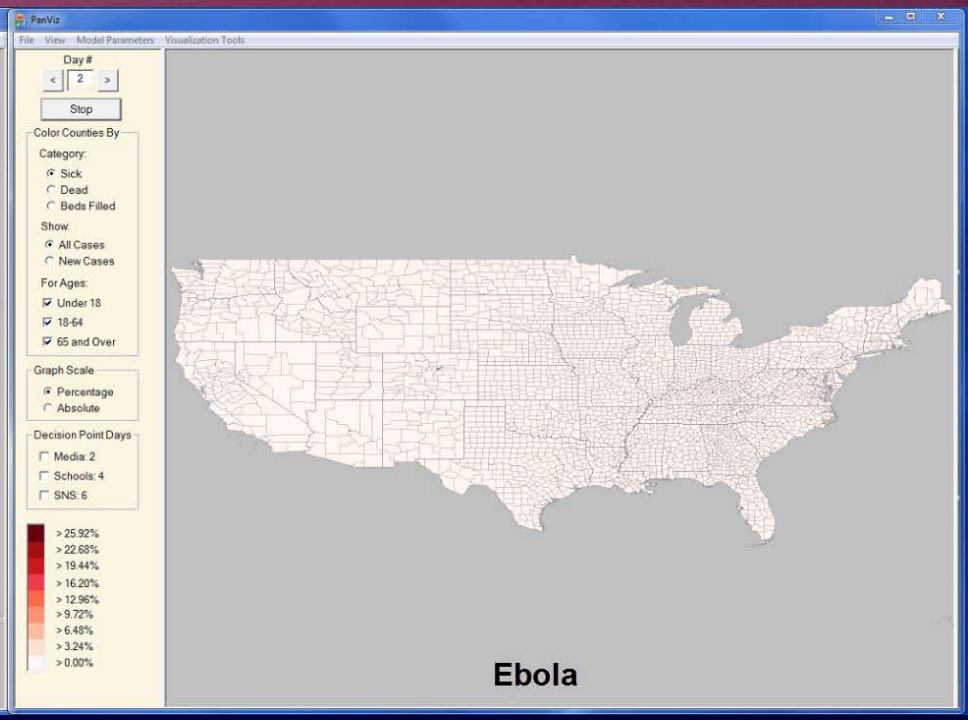




VACCINE

Ebola Vs. Pandemic Influenza (H1N1)



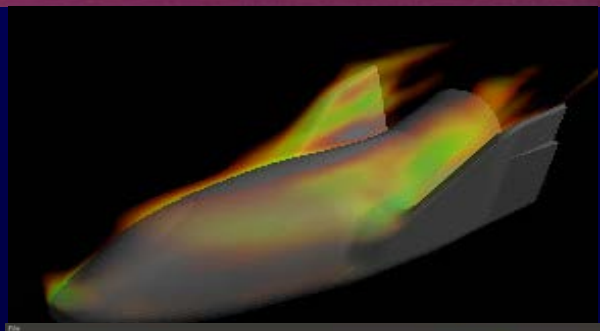


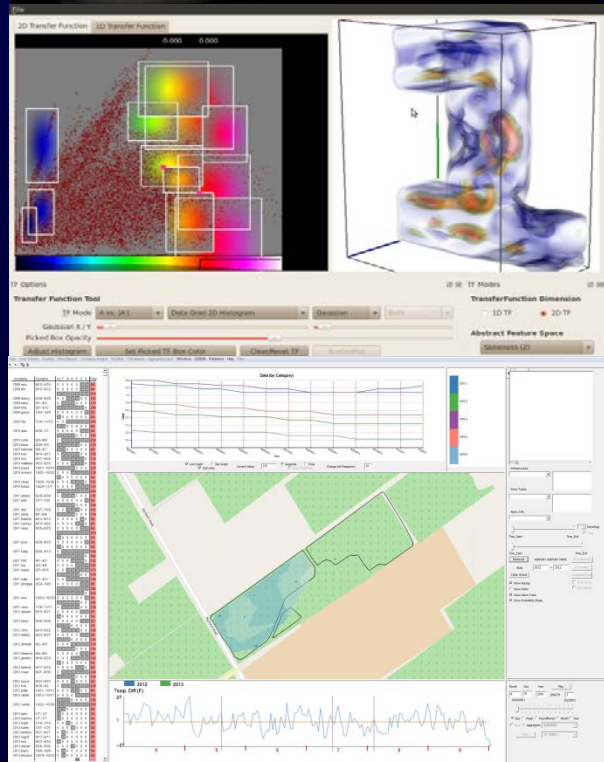
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

Scientific Discovery

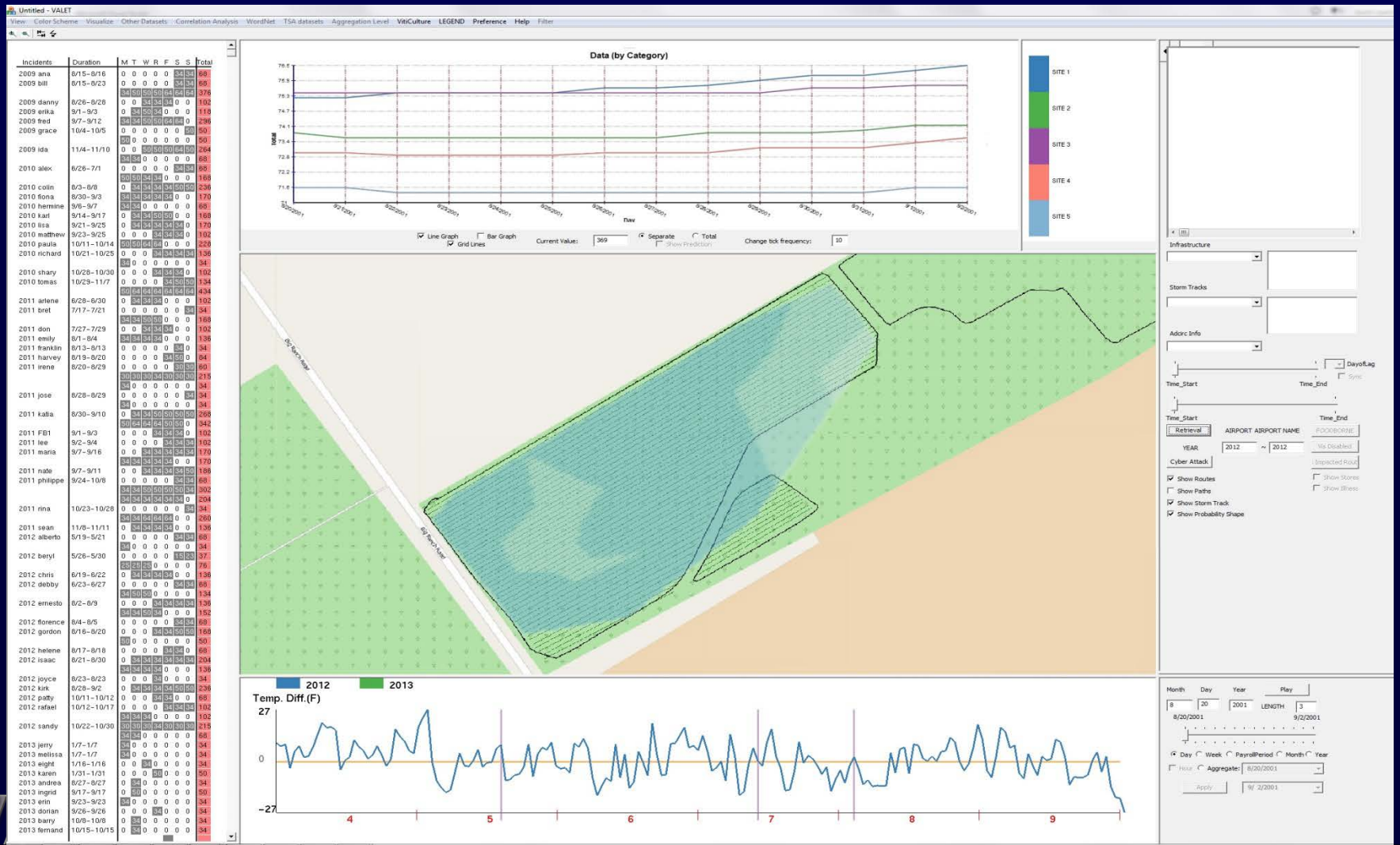


- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality



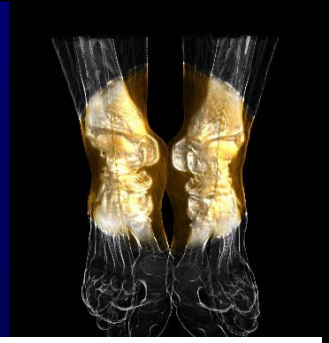


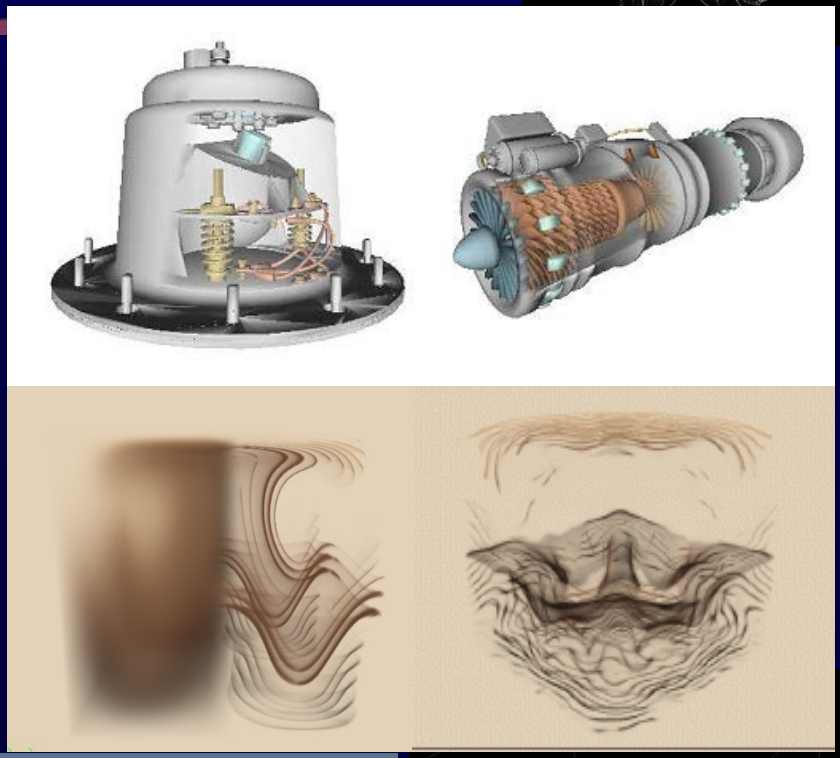
Precision Agriculture VA

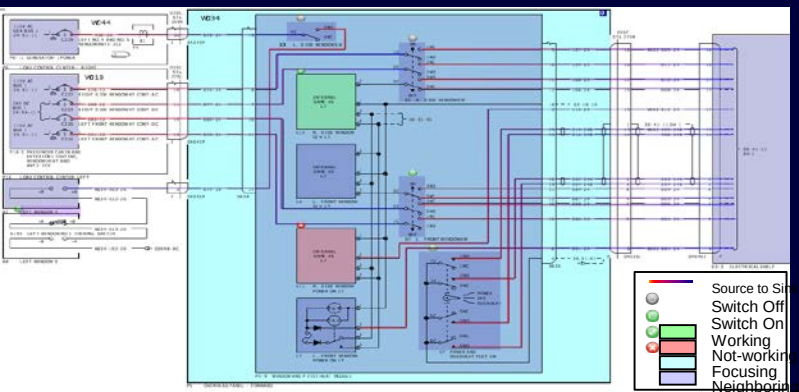


Scientific Discovery

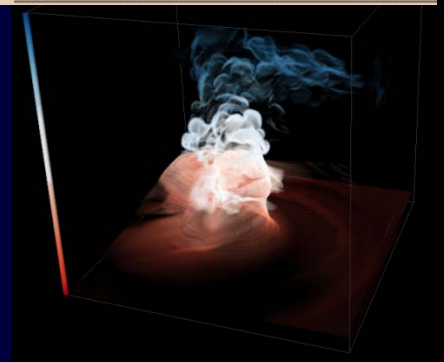
- Illustrative Visualization
  - Mechanical assemblies
  - Schematics and wiring diagrams
  - Illustrative flow visualization
  - Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics





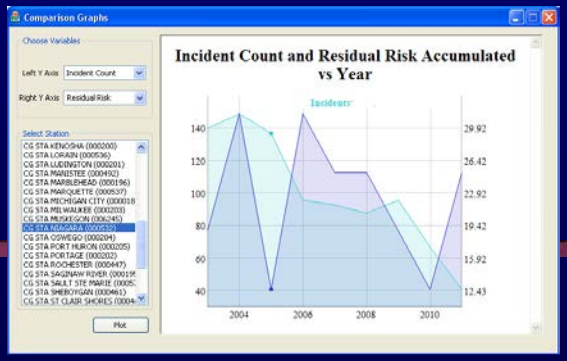


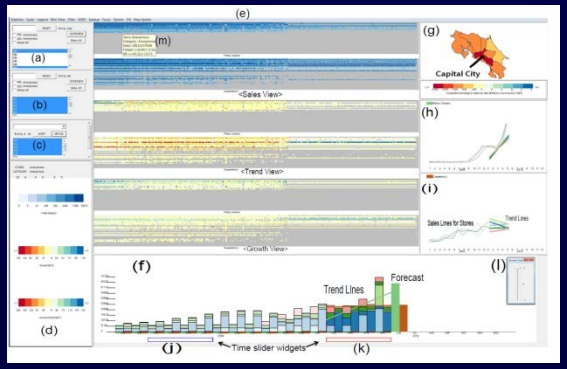




Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation



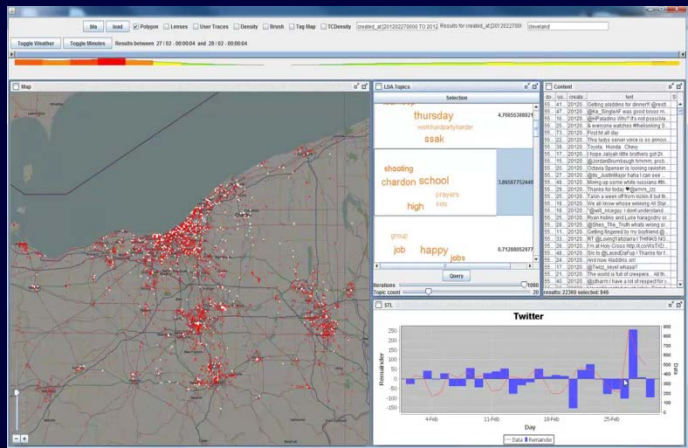




Visual Analytics Uses for Public Safety



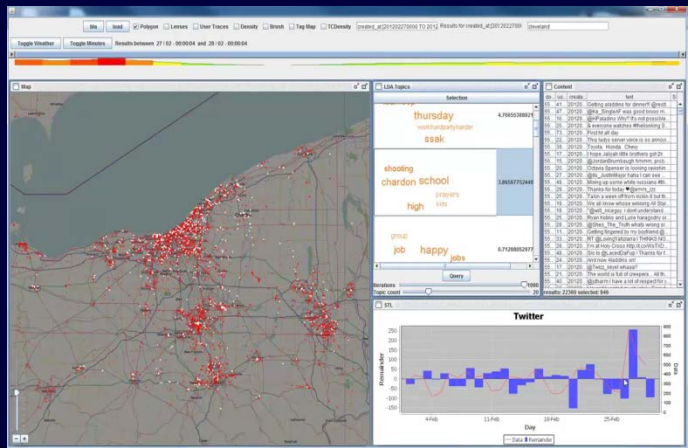
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Our Spatiotemporal Visual Analytics Work Contributions

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

The Scale Problem

- Data scale:
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues
- User / problem scale:
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally

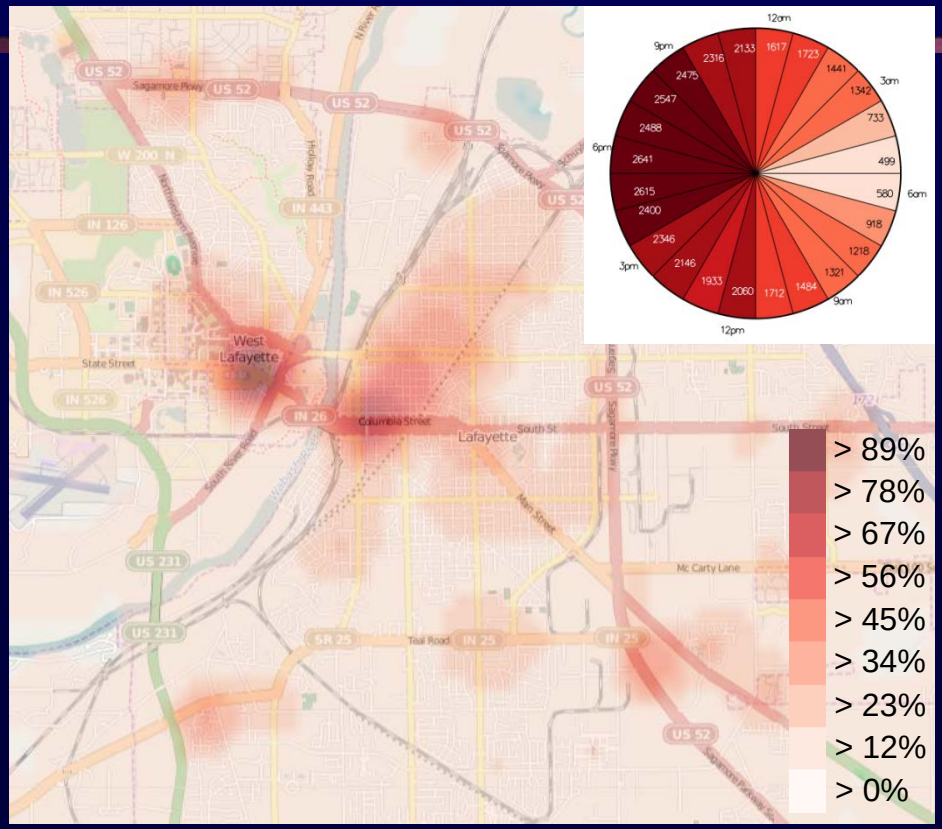
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness[1]
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: **Empower the casual expert to be more effective**

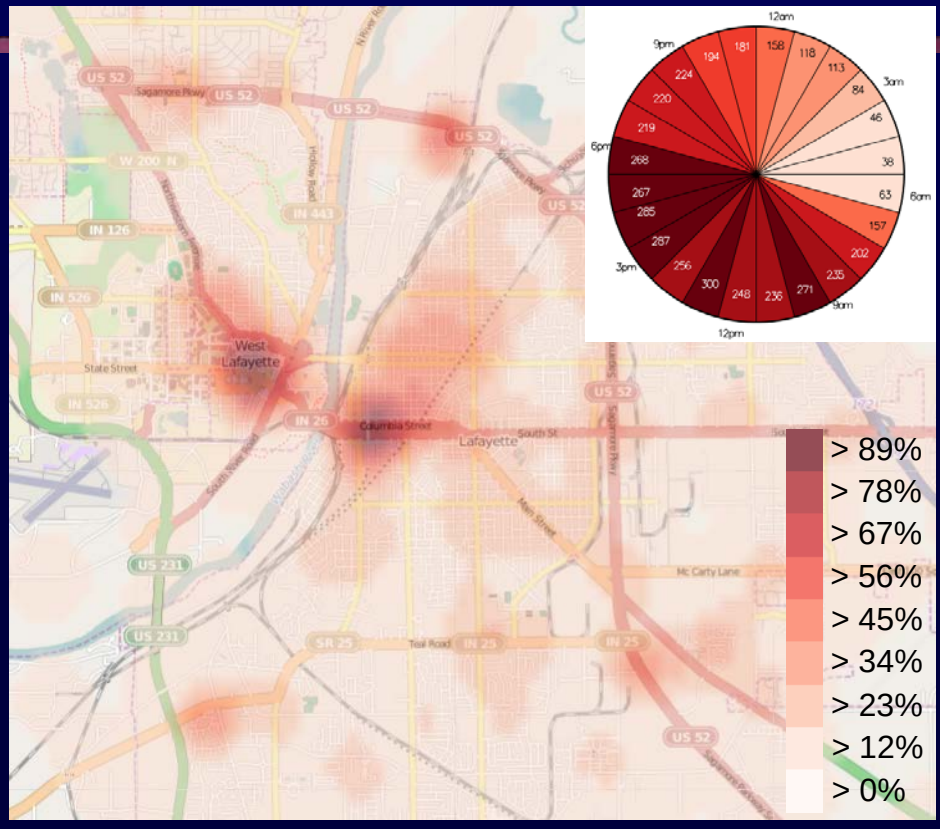
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

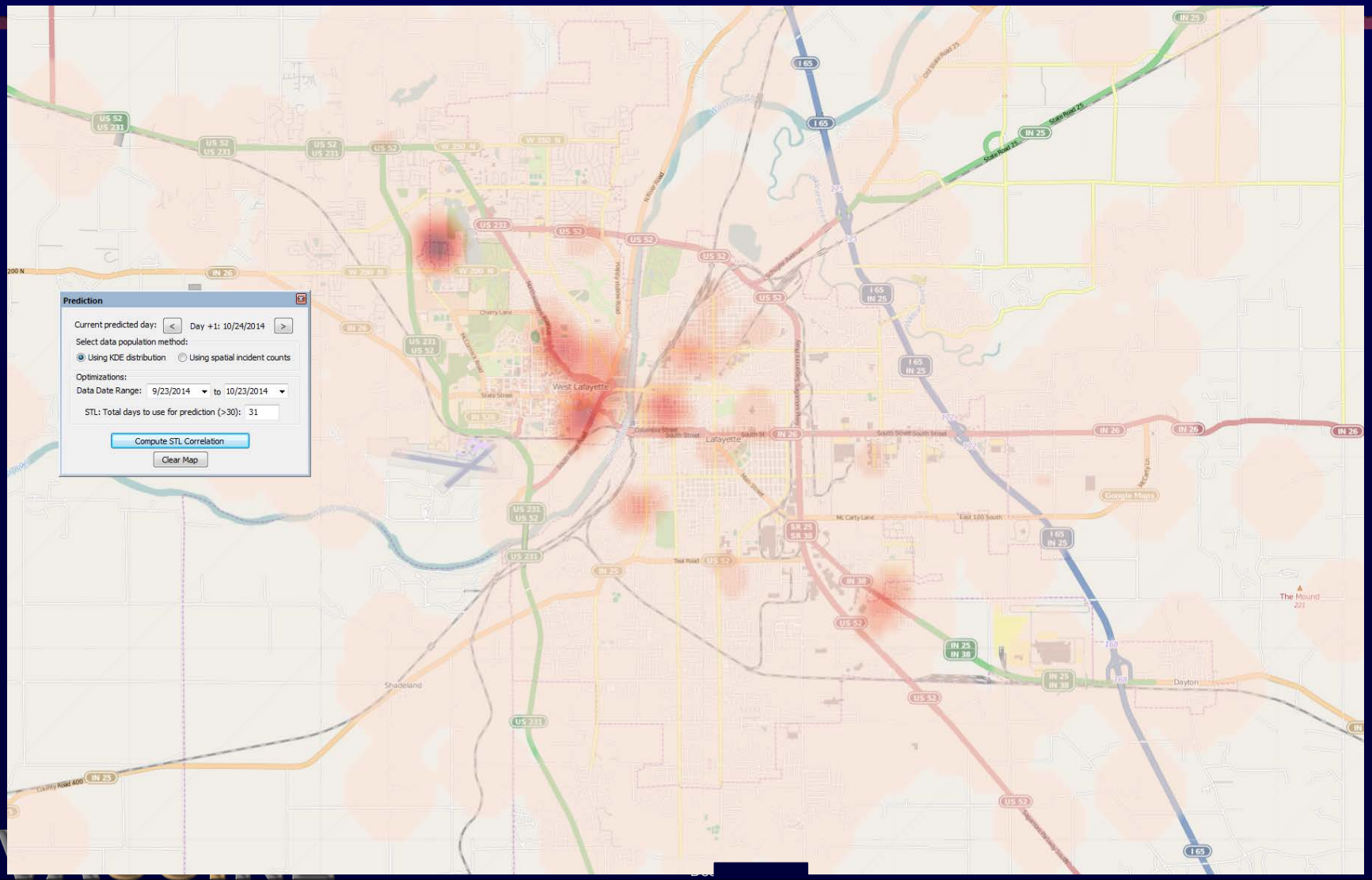


Past 2 Years
3/11/2012 – 3/10/2014

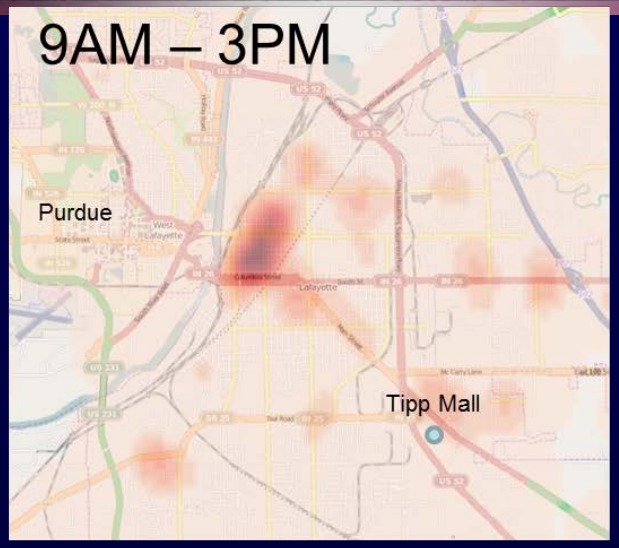


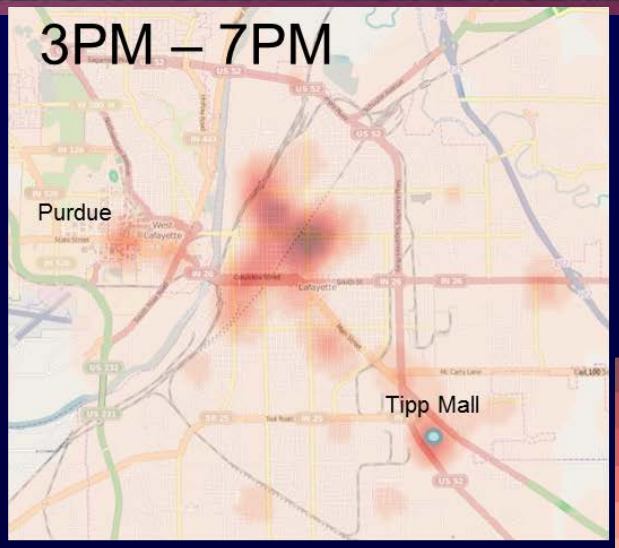
March + Tuesday Filter
(Past 10 Years)

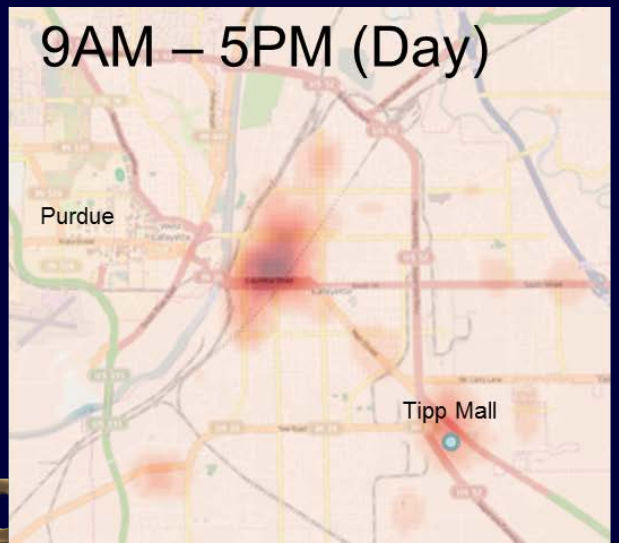
Prediction Using KDE Distribution

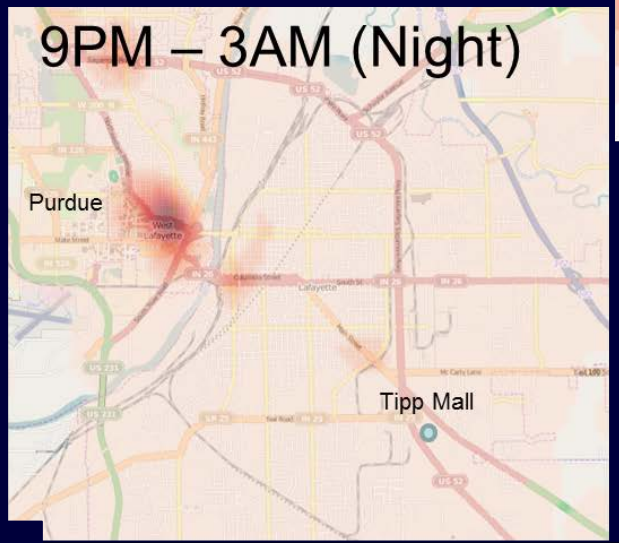


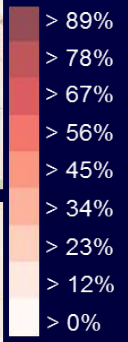
Prediction During Certain Hours In The Day





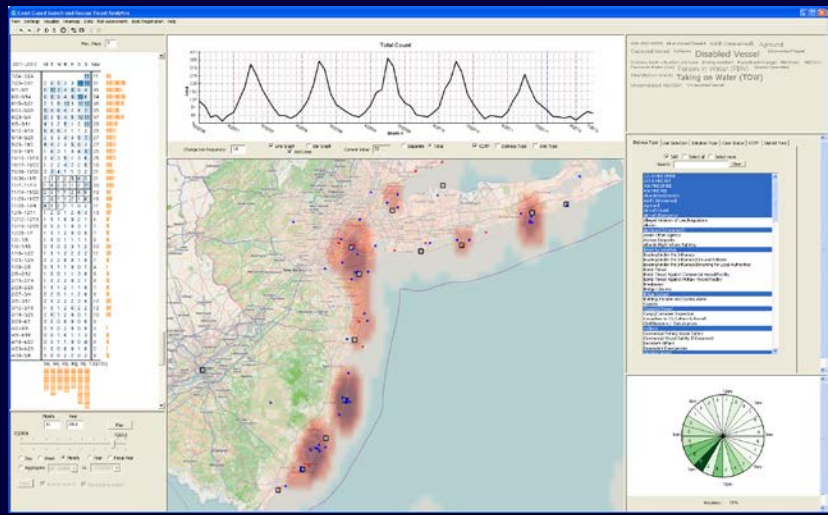




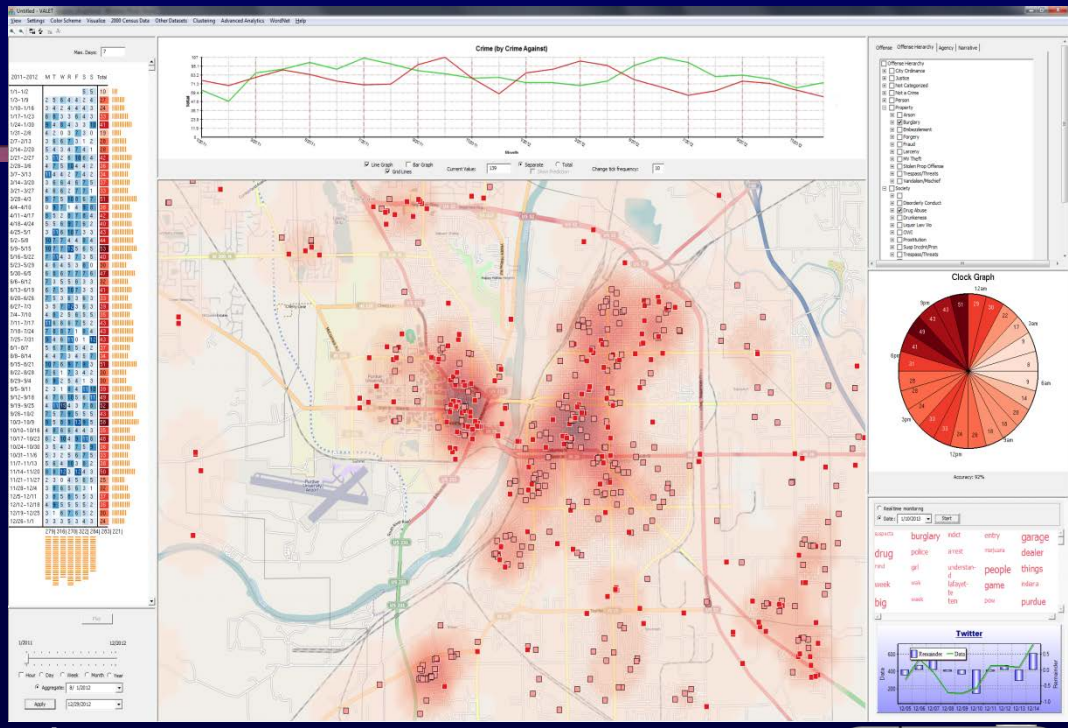


Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system result
  - US Coast Guard
  - Social Media
- Path forward



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

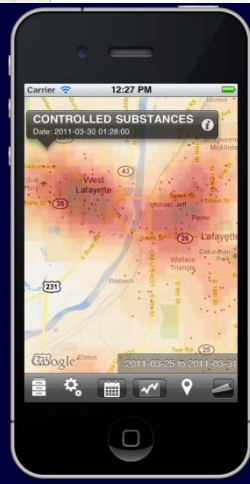
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

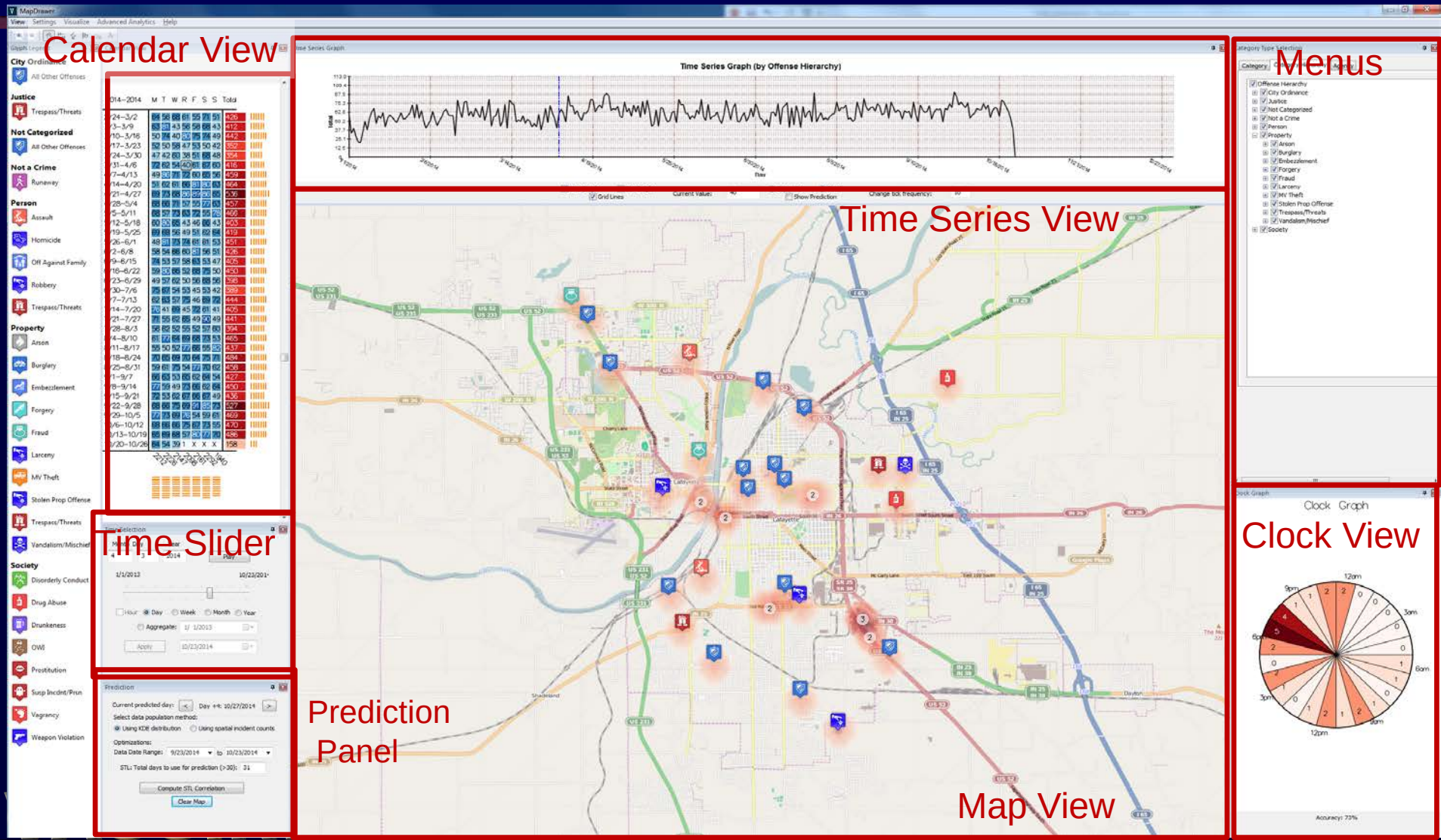
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

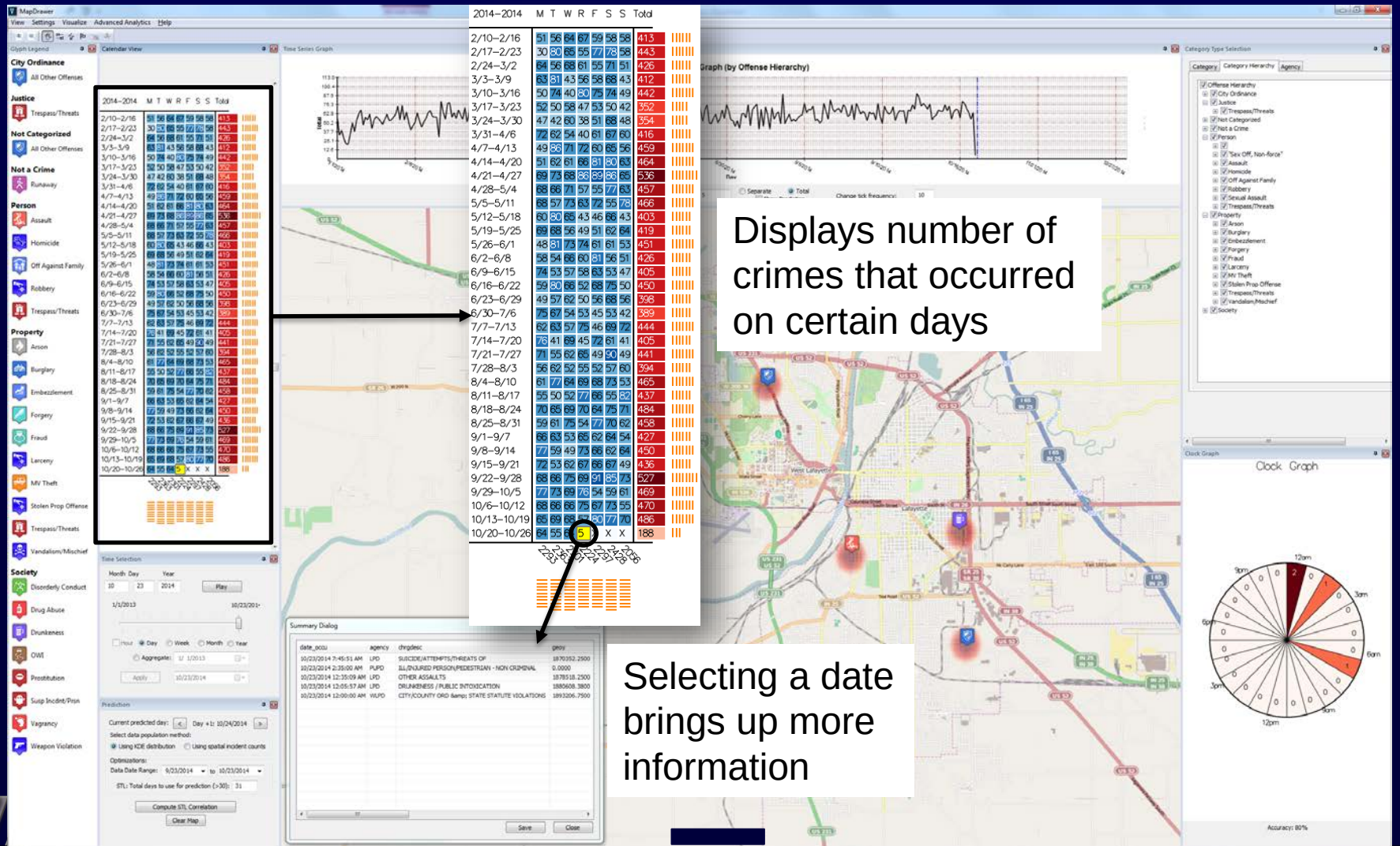


VACCINE

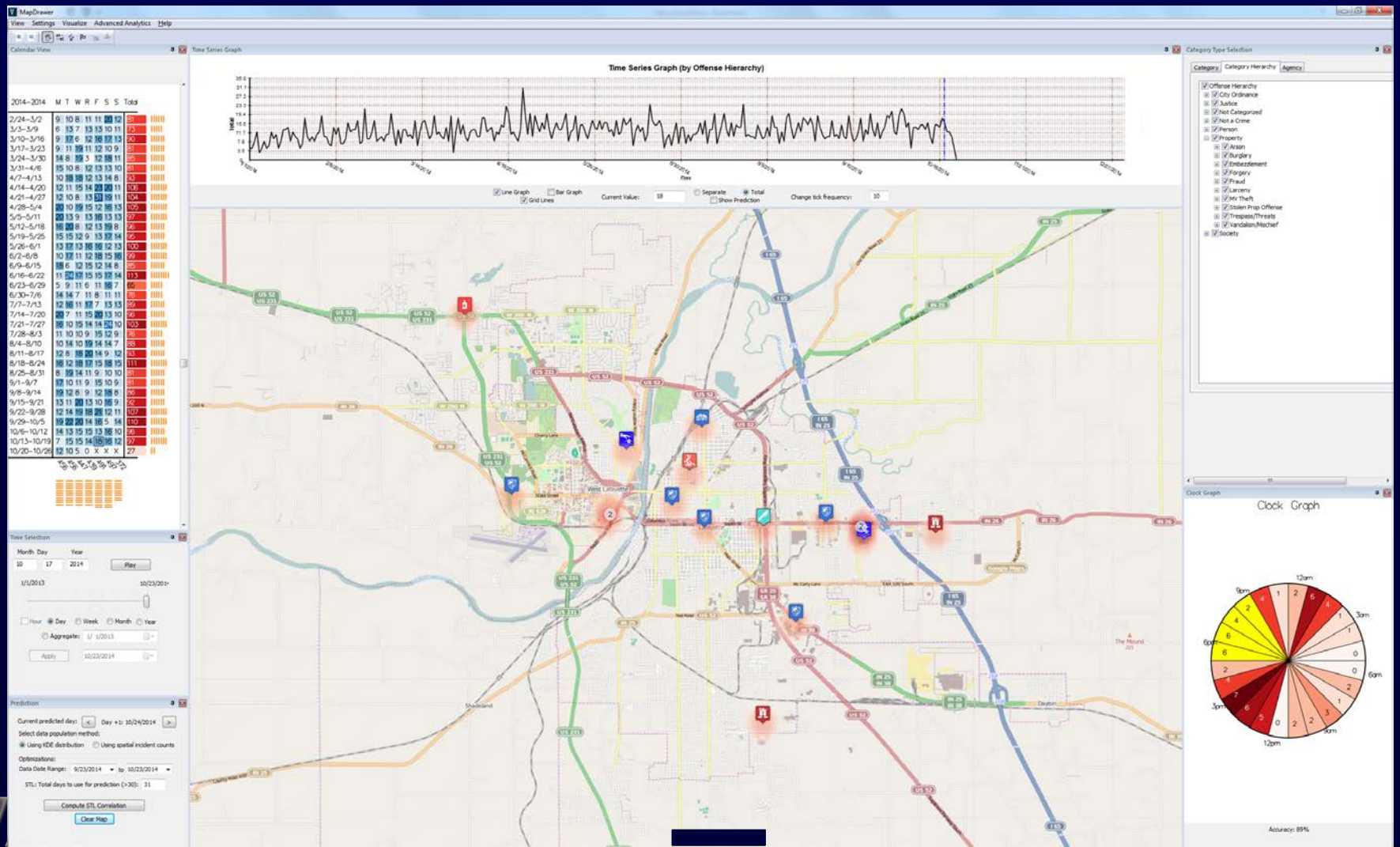
VALET Layout



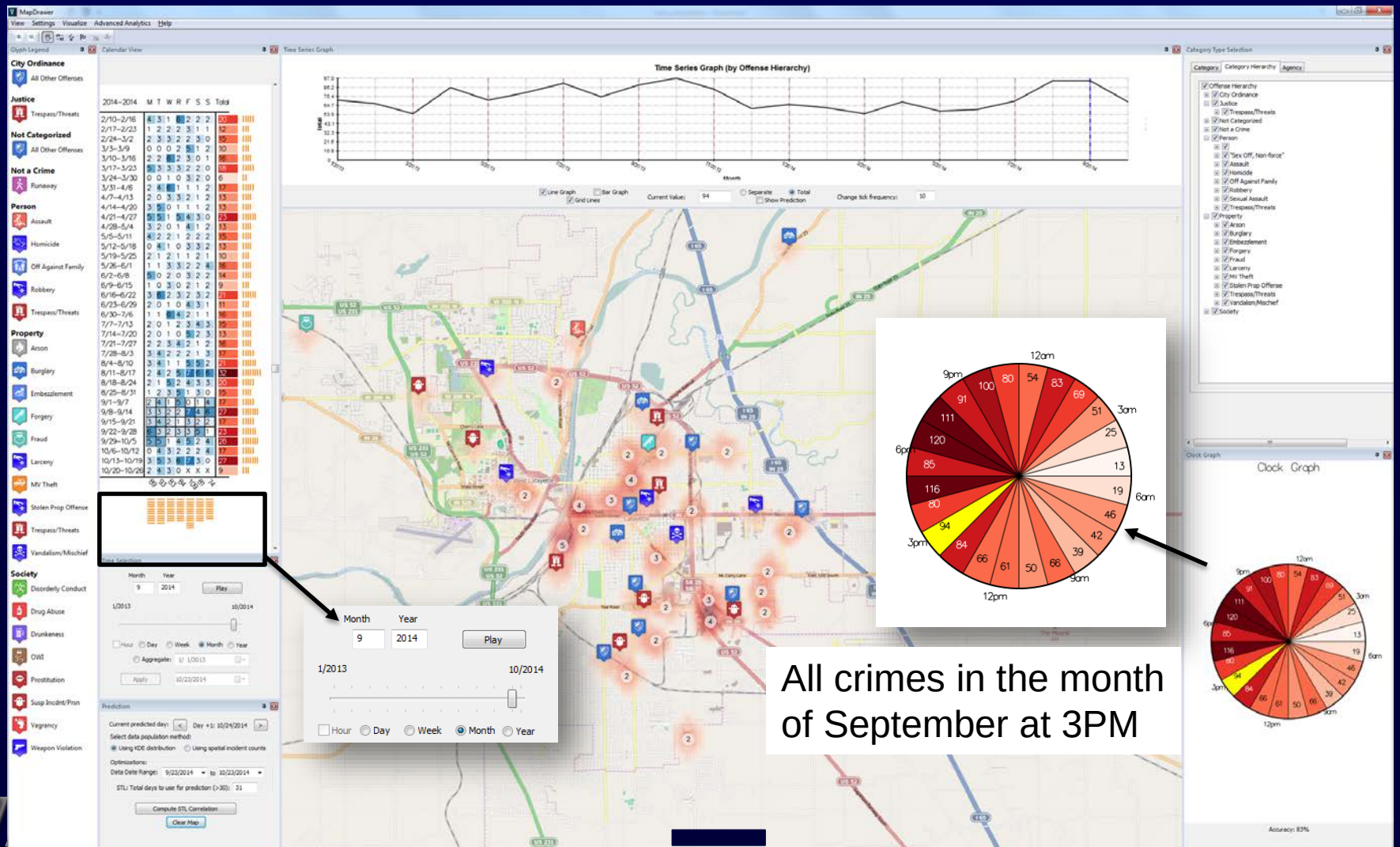
Calendar Display



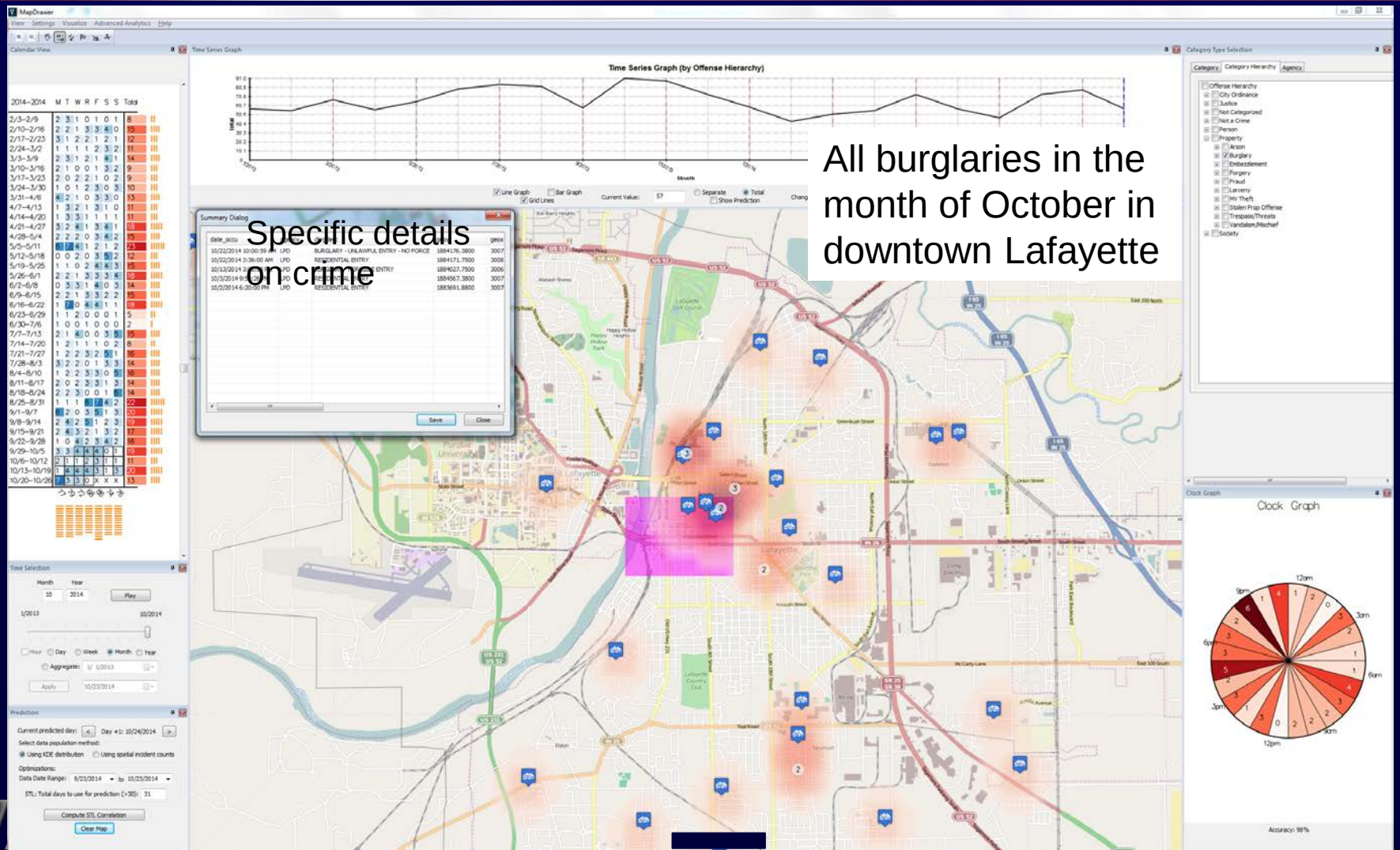
Crimes In A Single Day From 6PM – 9PM



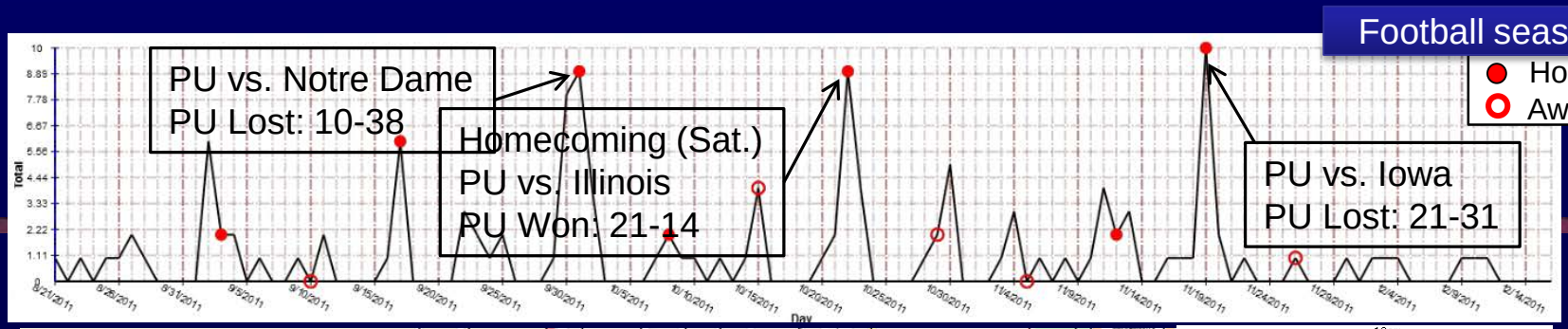
Crimes During Certain Hours In A Month

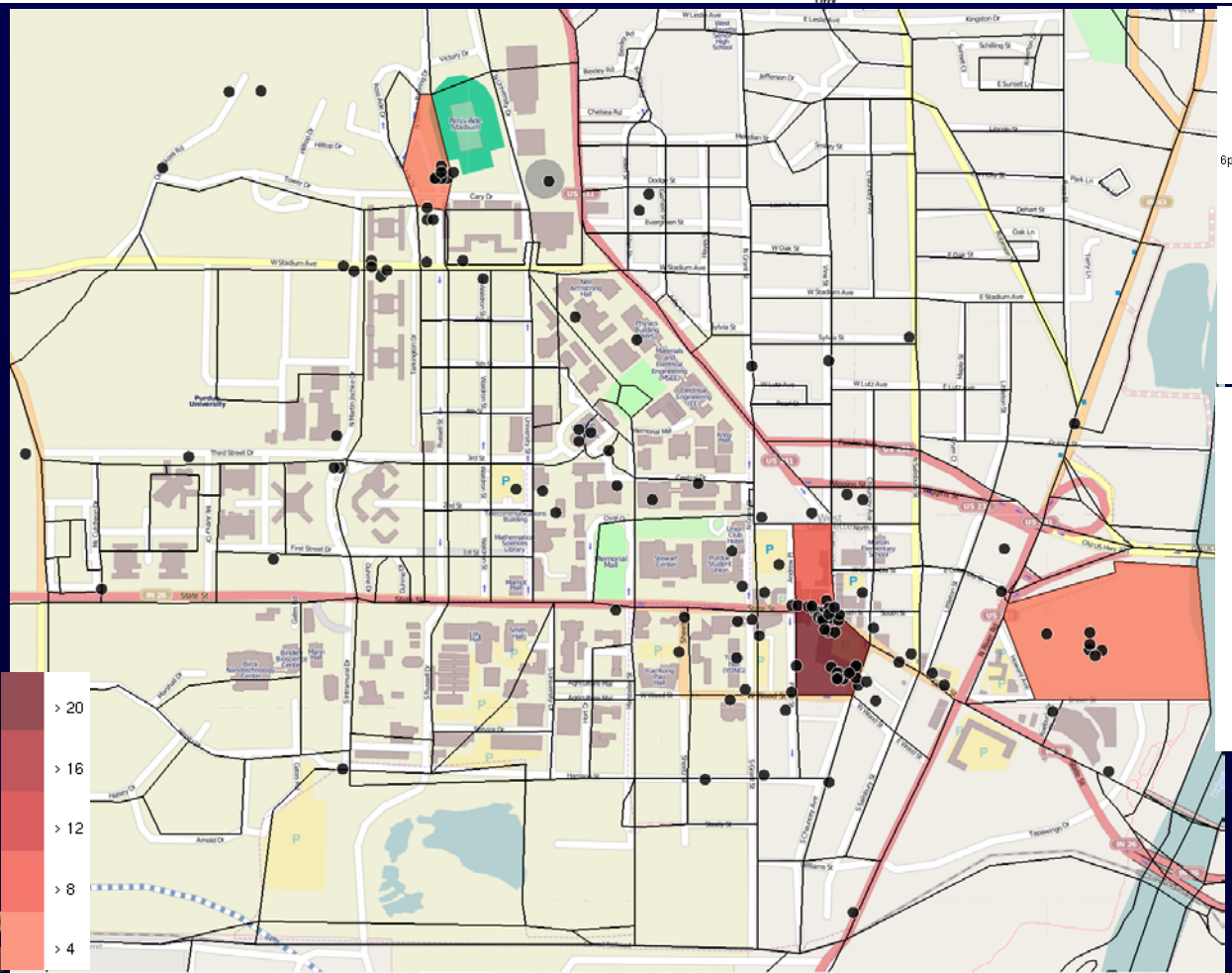


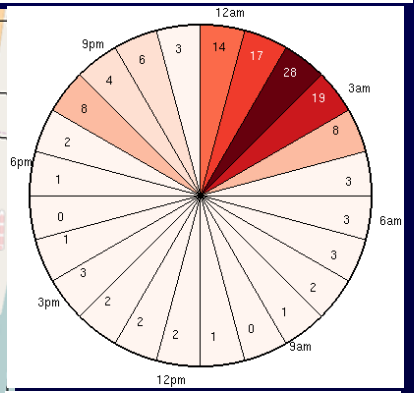
Select Specific Region To Display Crimes

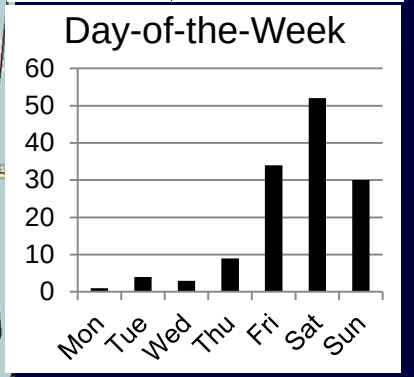


Example: More Data Sources – Intoxication

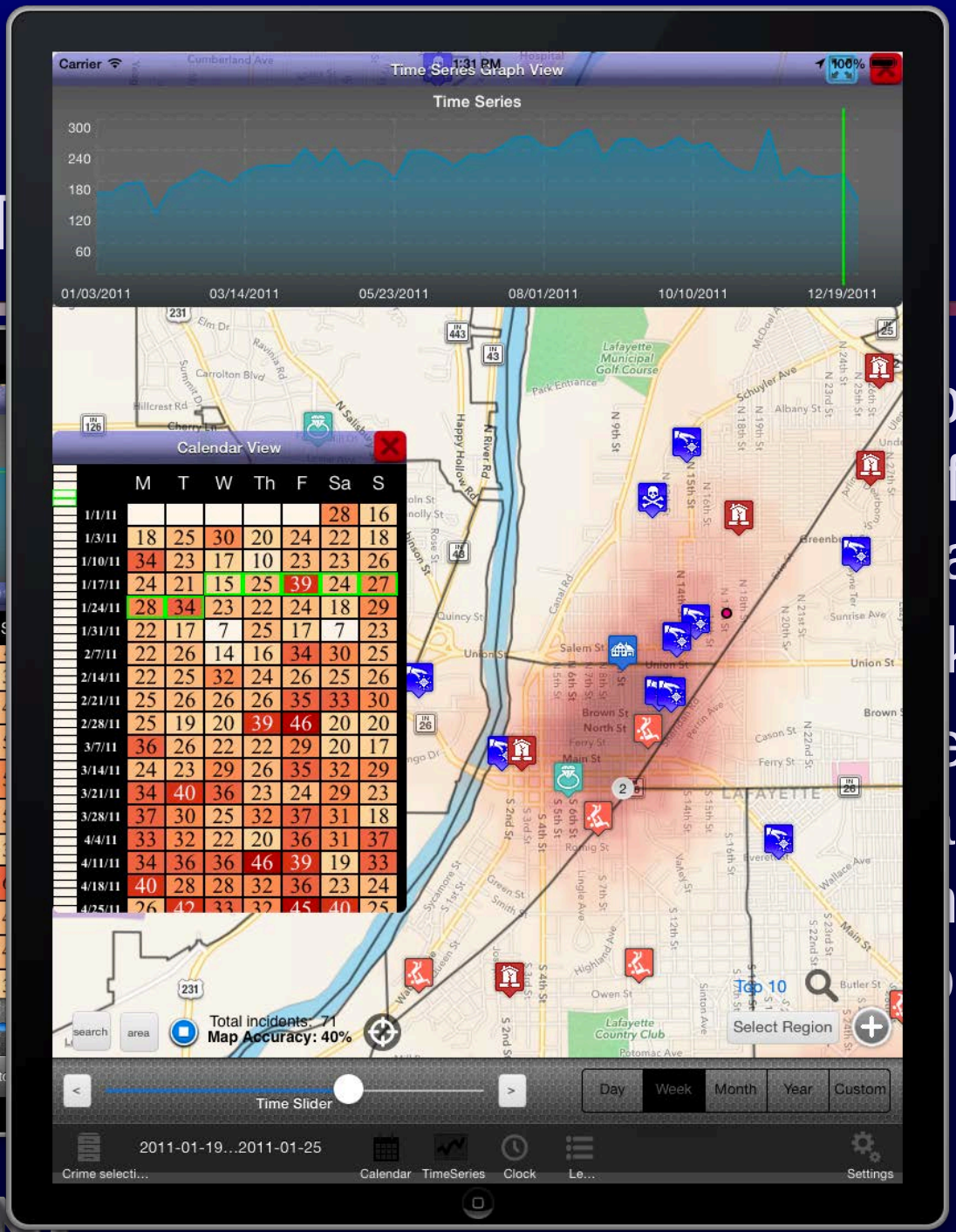




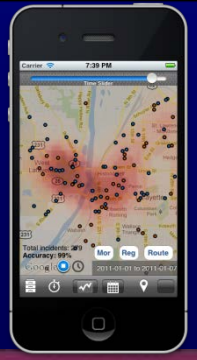


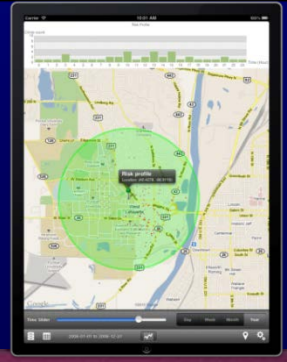


iVALET









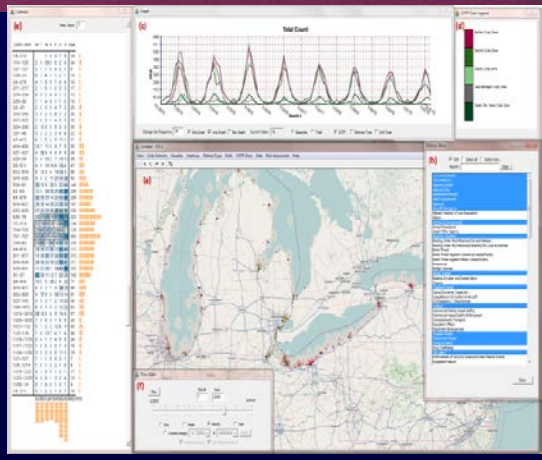


lore criminal,
fic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis

VACCINE

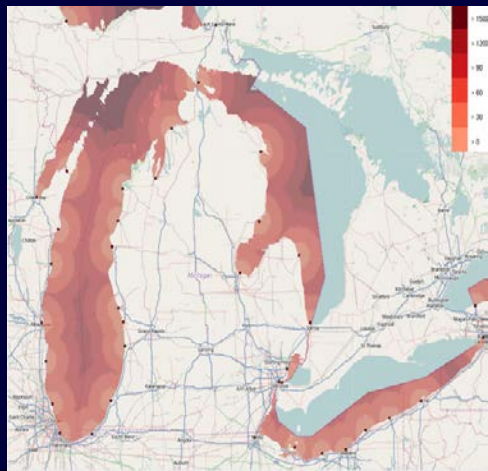
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

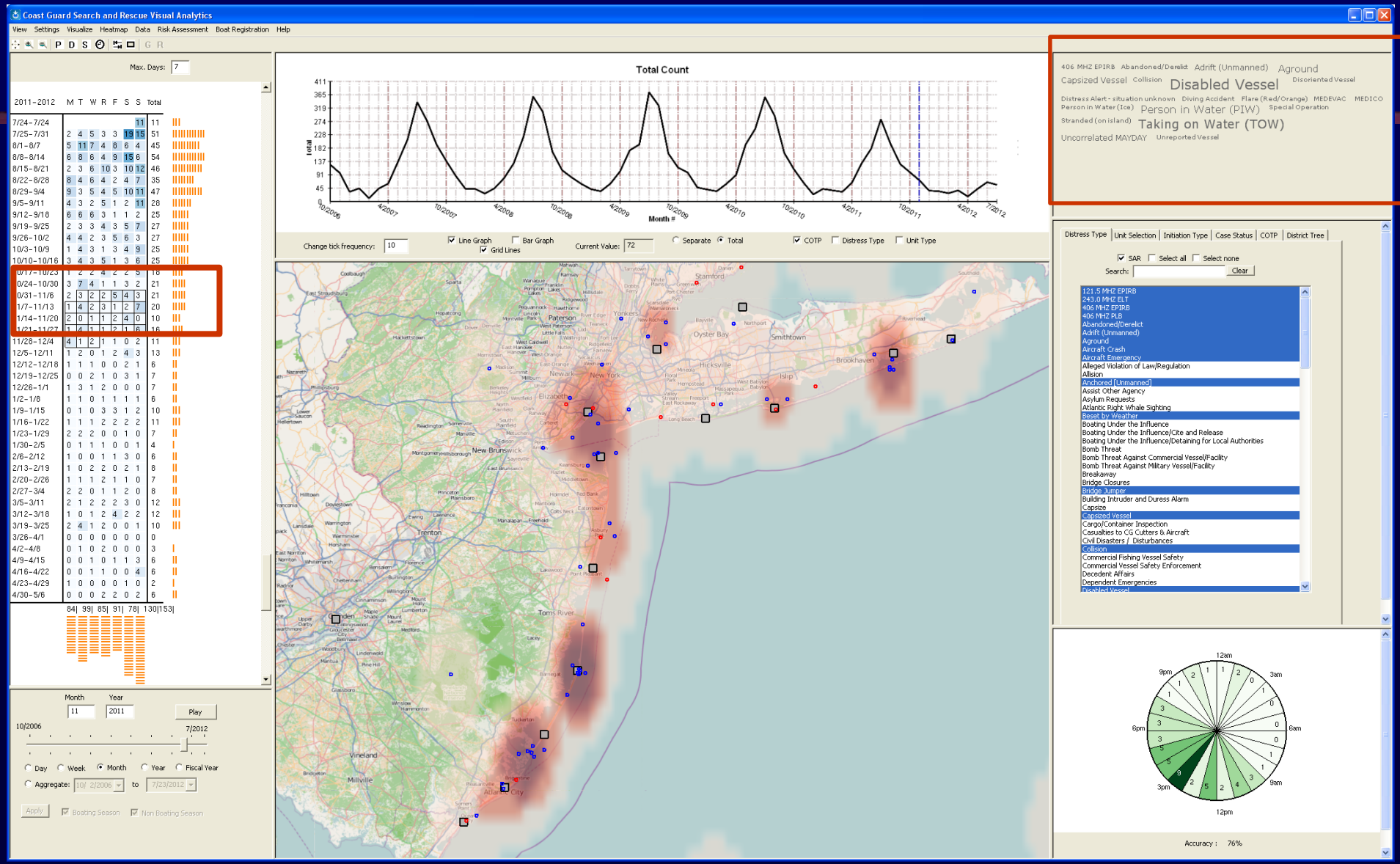
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



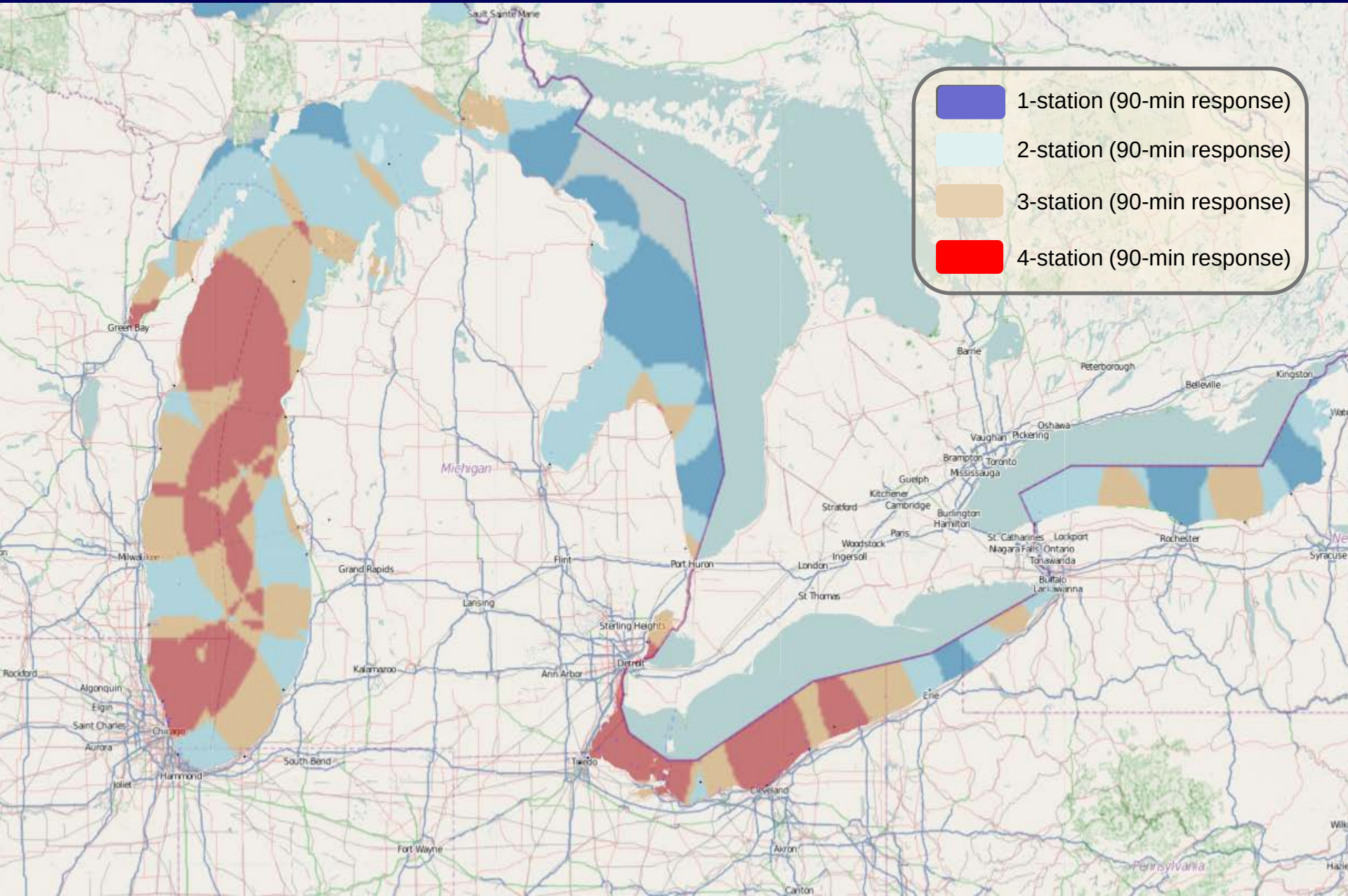
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

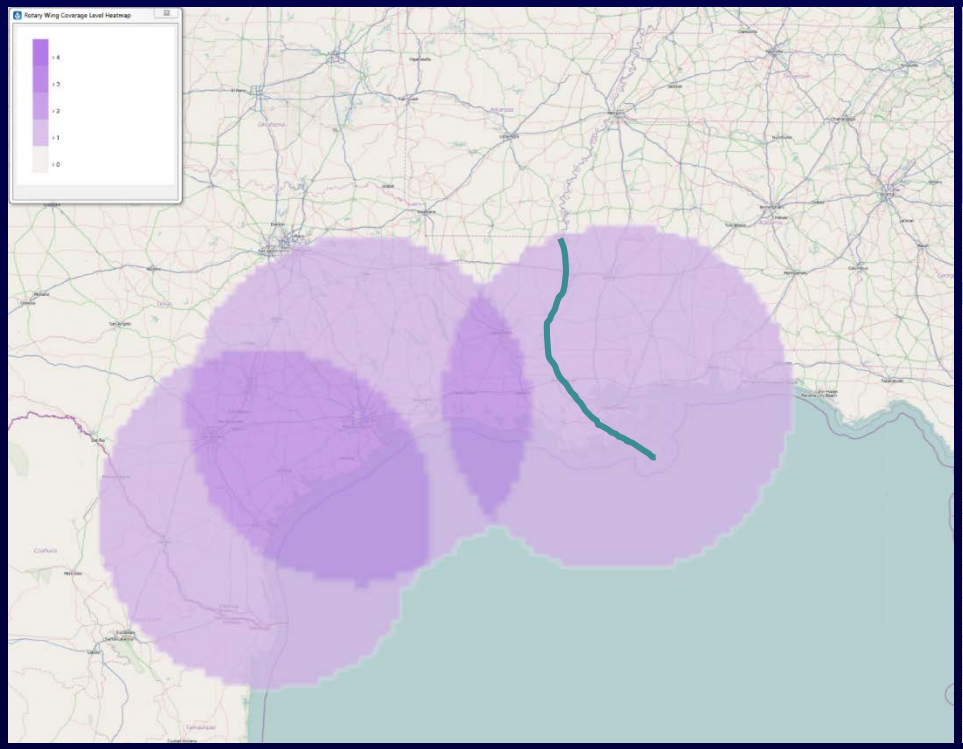


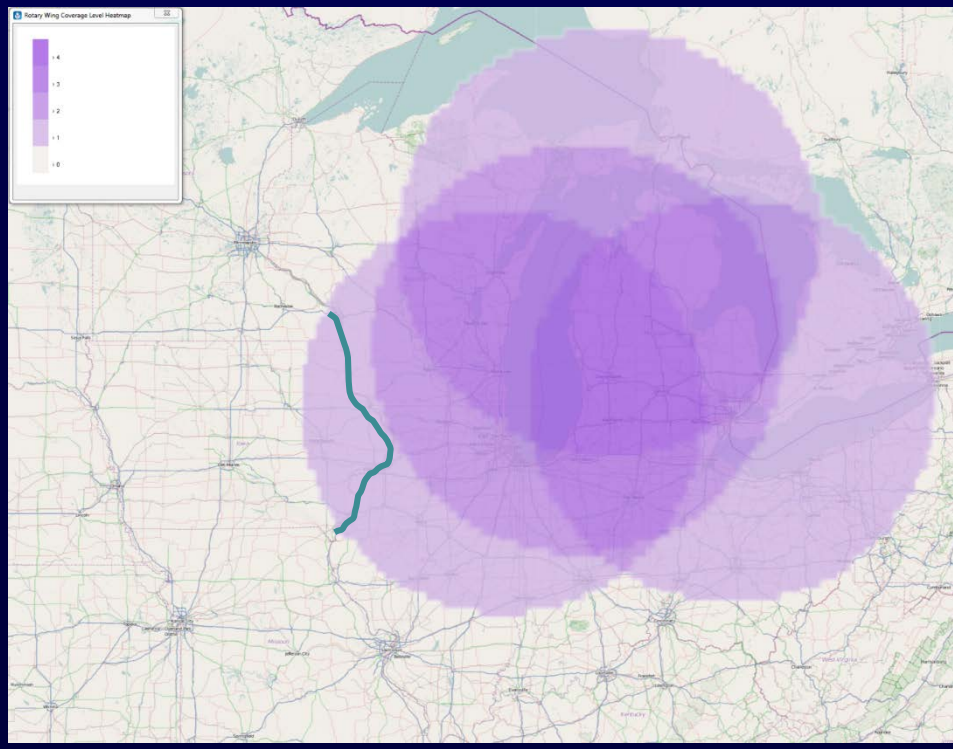
Response Efficiency – Potential Future Assets



Airstation Utility Visualization Comparison

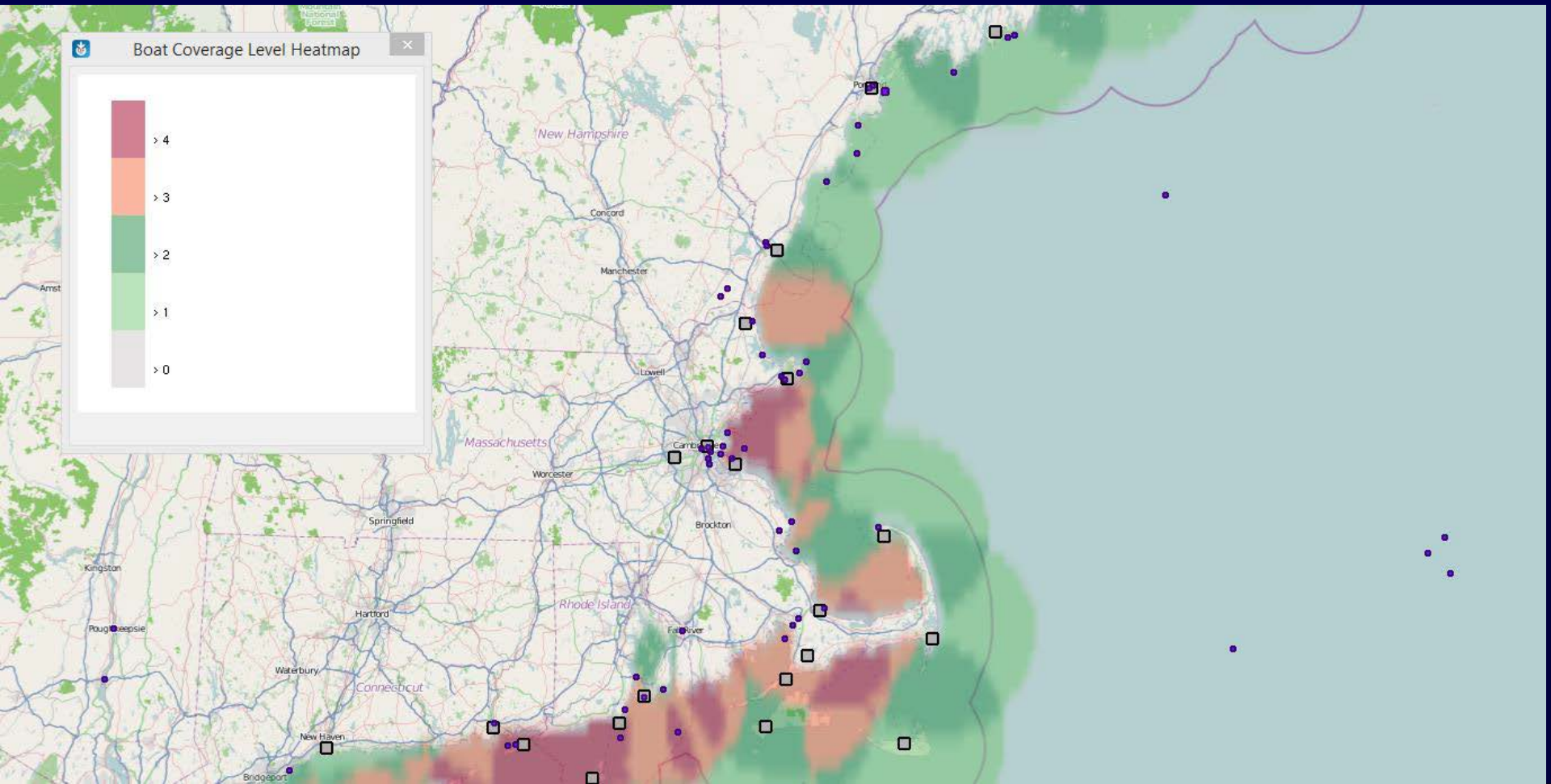
- Rotary Wing Coverage Map by AirSTA





Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



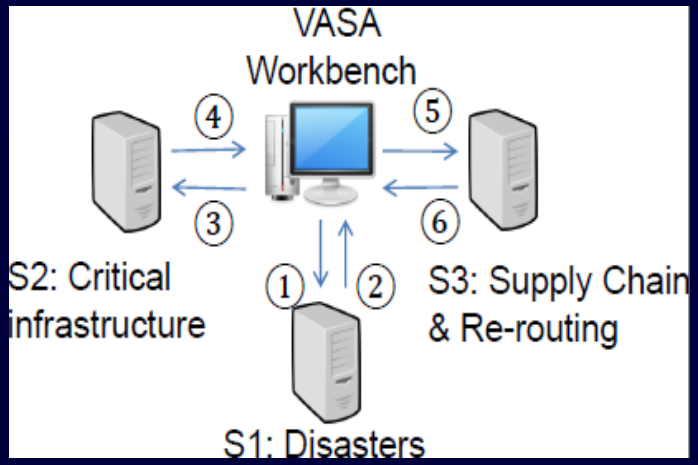
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

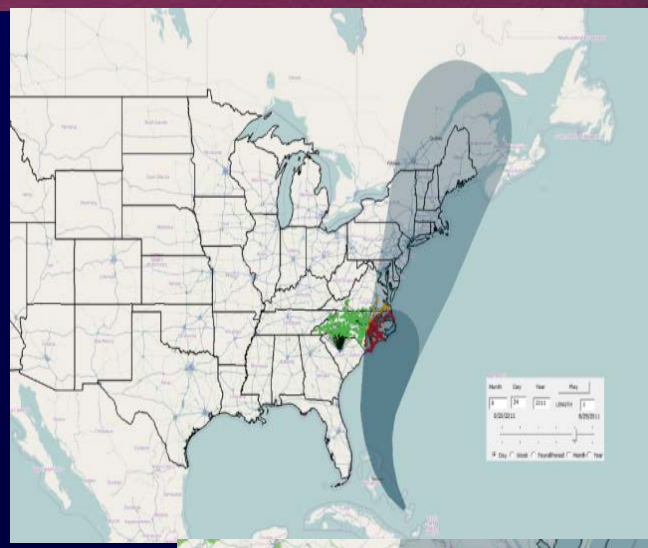
TACC

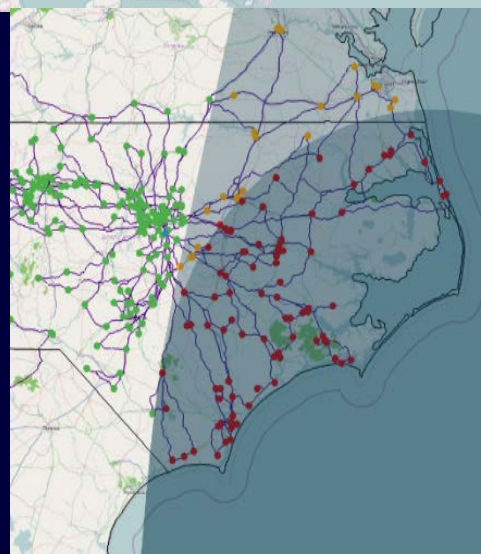
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models



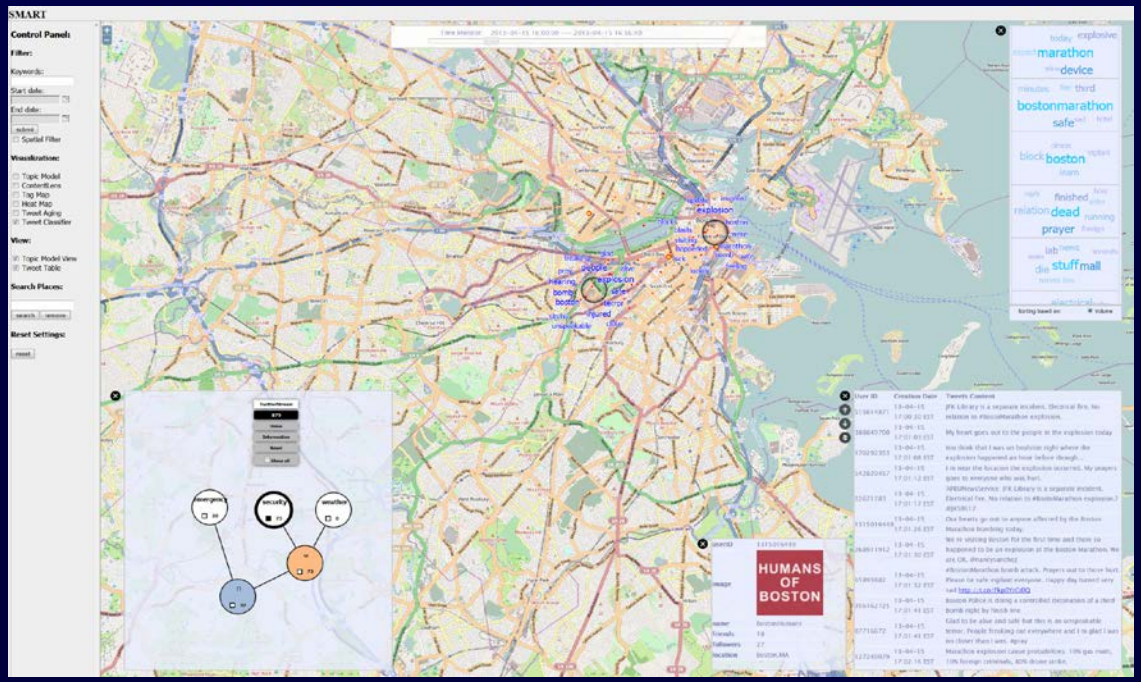


VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

December 2014

Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting

Time: around 10:40 am PST
(01:40 pm EST), 10/24/2014

Earliest Tweets About The Event

601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances
461749842	14-10-24 13:55:45 EST	Cops are fucking EVERYWHERE right now
461749842	14-10-24 13:55:21 EST	Cops are fucking EVWRYWHERE right now
707981743	14-10-24 13:55:12 EST	There is a shooting at mp?
818397241	14-10-24 13:54:06 EST	@Miguel_Rocha34 do you have any idea who the person with the gun was
818397241	14-10-24 13:52:12 EST	you got some serious issues that you need to seek help for if you shoot up a fucking high school
601648236	14-10-24 13:49:11 EST	@Ray_Raaaay21 one of the kids in the lunchroom started shooting and everyone booked it
601648236	14-10-24 13:46:17 EST	shooting at mp

707981743	14-10-24 14:17:43 EST	2 people are dead...
335778696	14-10-24 14:17:26 EST	Wtf is going on?? Shooting at mp?
70255303	14-10-24 14:16:37 EST	Shooting at Marysville Pilchuck. Praying for the kids and the families.
1853421522	14-10-24 14:14:39 EST	@Lake_stonerrr a student shot several students, possible 2 students dead 6 injured
1099441382	14-10-24 14:10:38 EST	Shooting at pikchuck???
599963823	14-10-24 14:08:26 EST	Something's fucking wrong !!!!! Or my mom wouldn't be crying about the MP shooting !!!
130018947	14-10-24 14:07:51 EST	About 5 mins away from the scene. Sad. Another school shooting. #MPHS
601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances

Some online news feed:
http://www.foxnews.com/us/2014/10/24/police-respond-to-shooting-at-washington-state-high-school/
http://www.cnn.com/2014/10/24/us/washington-school-shooting/index.html

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

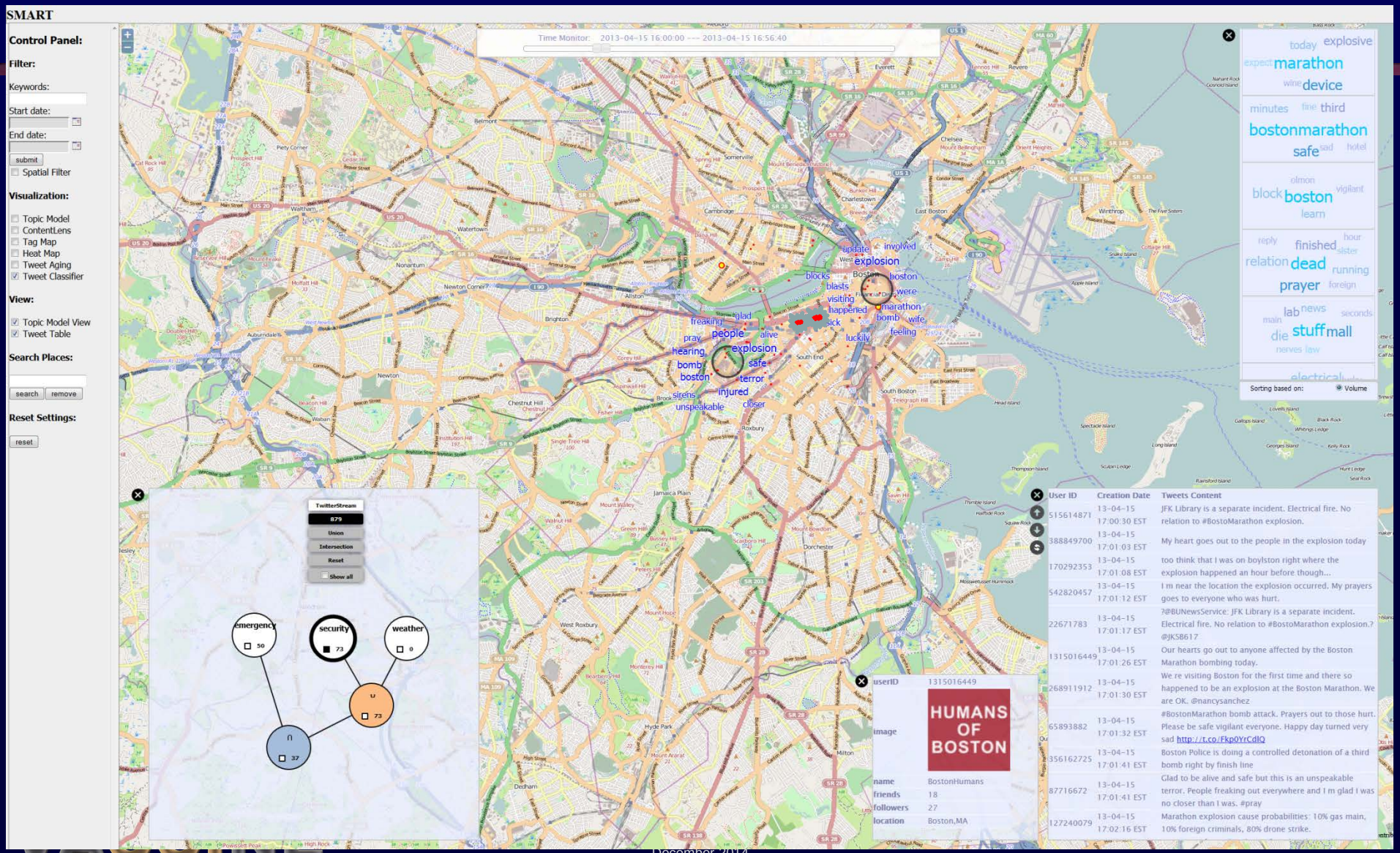
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

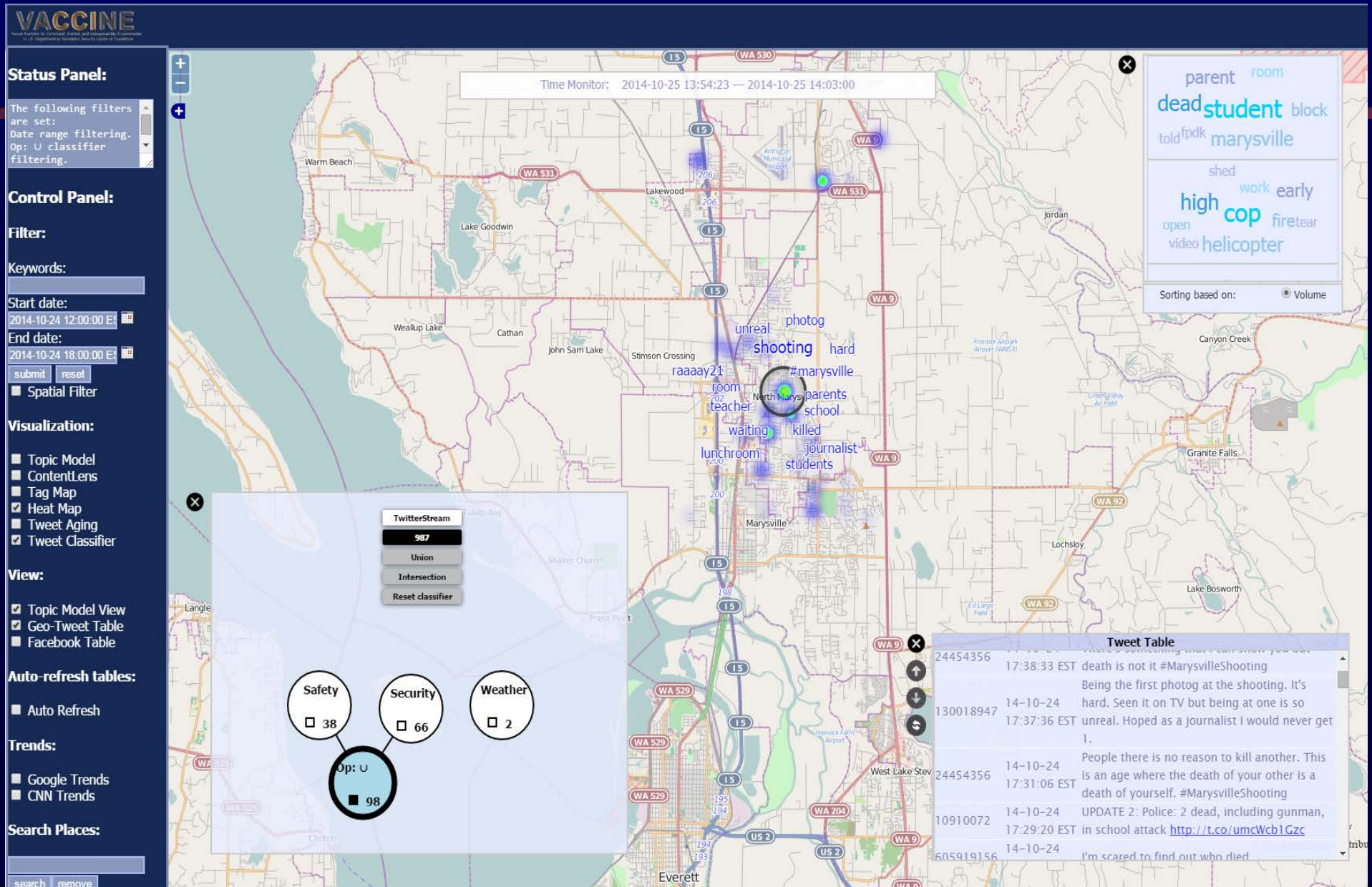


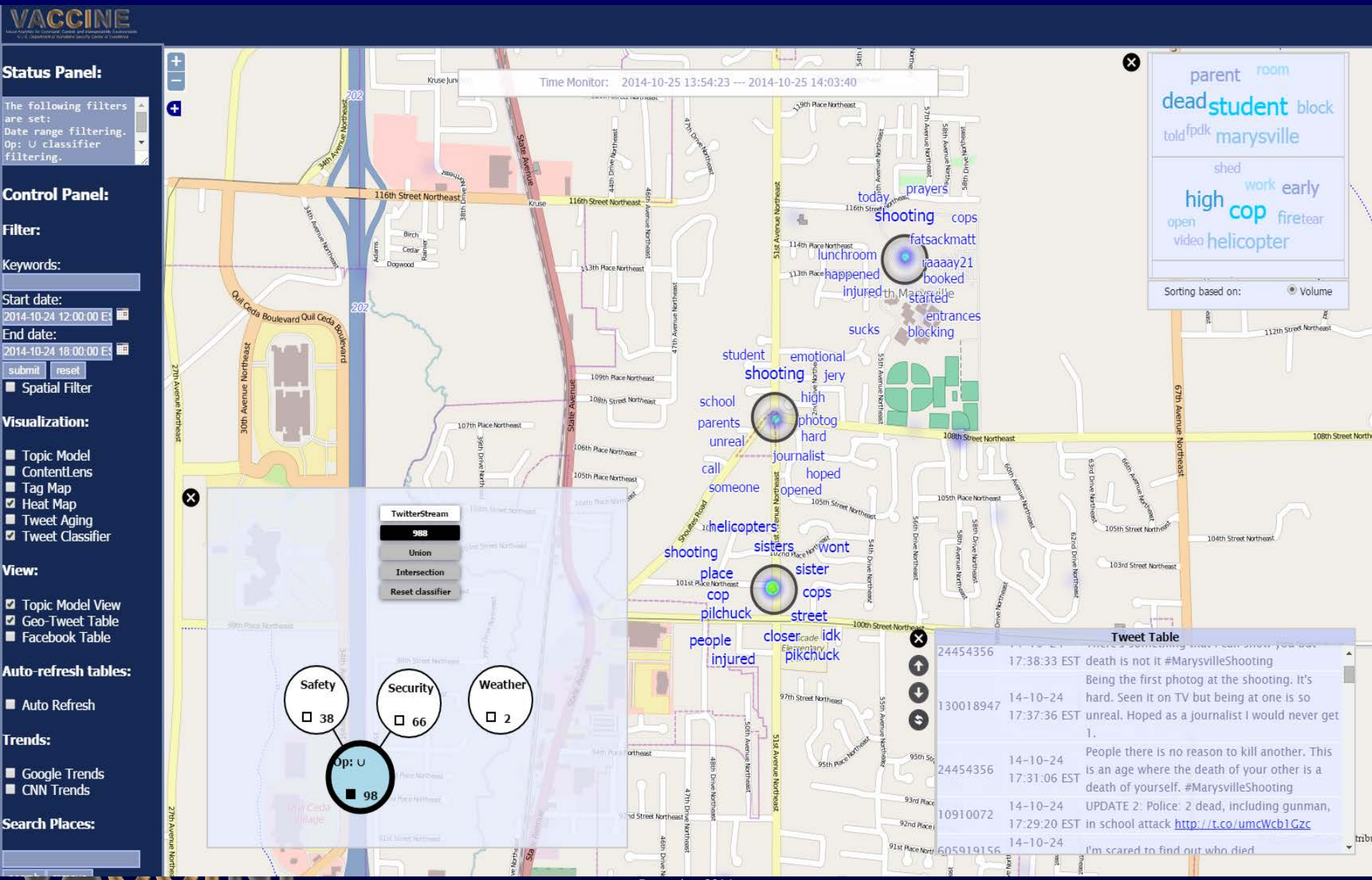
VACCINE

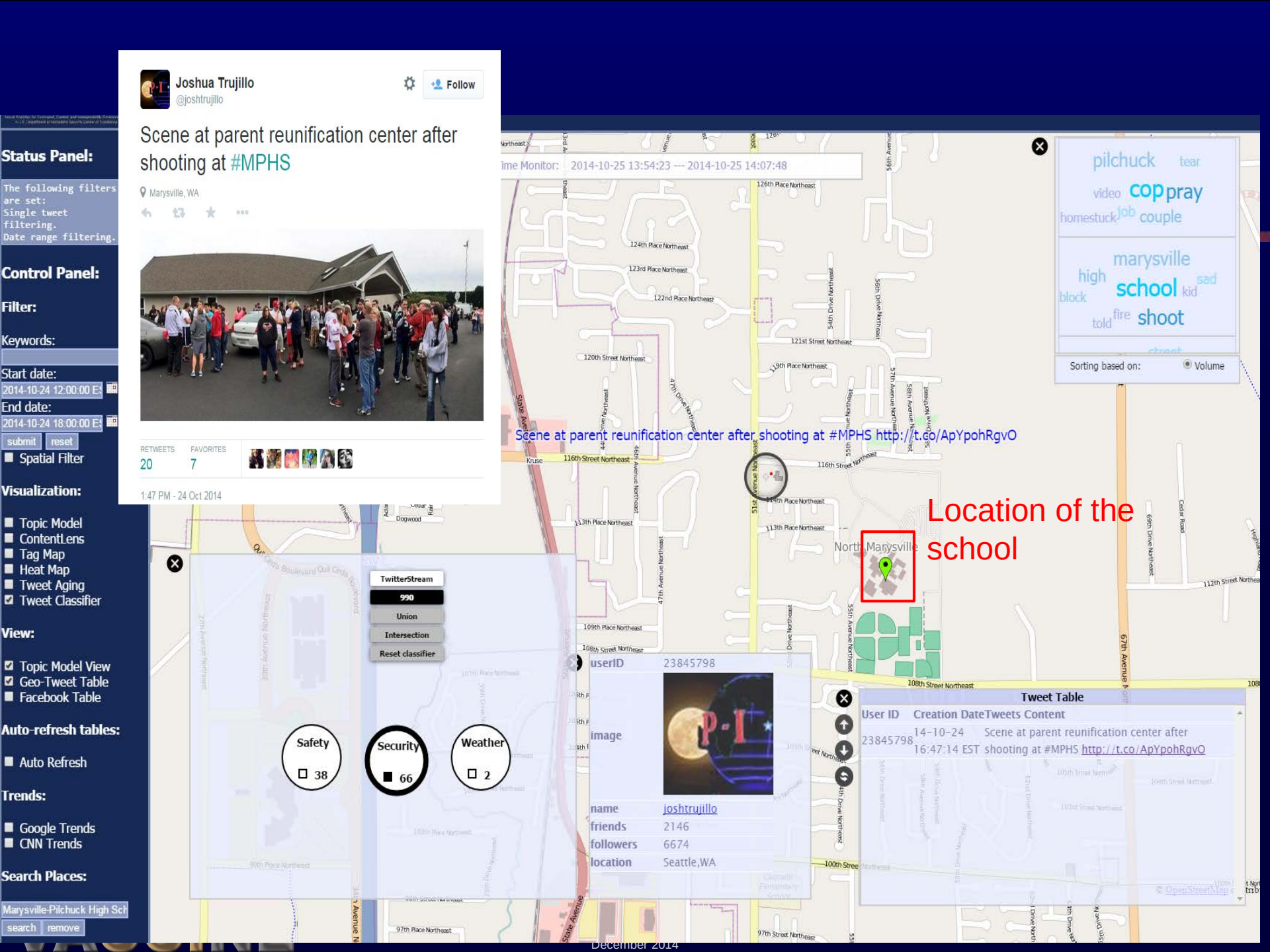
Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24
17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24
17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24
17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died


Joshua Trujillo
@joshtrujillo
Follow
Scene at parent reunification center after shooting at #MPHS
Marysville, WA
RETWEETS
20
FAVORITES
7
1:47 PM - 24 Oct 2014
Status Panel:
The following filters are set:
Single tweet filtering.
Date range filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 ES
End date:
2014-10-24 18:00:00 ES
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Marysville-Pilchuck High Sch
search
remove
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48
pilchuck
tear
video
cop
pray
homestuck
job
couple
marysville
high
school
kid
sad
block
told
fire
shoot
Sorting based on:
Volume
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO
Location of the school
North Marysville
TwitterStream
990
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
userID
23845798
image
name
joshtrujillo
friends
2146
followers
6674
location
Seattle,WA
Tweet Table
User ID
Creation Date
Tweets Content
23845798
14-10-24 16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

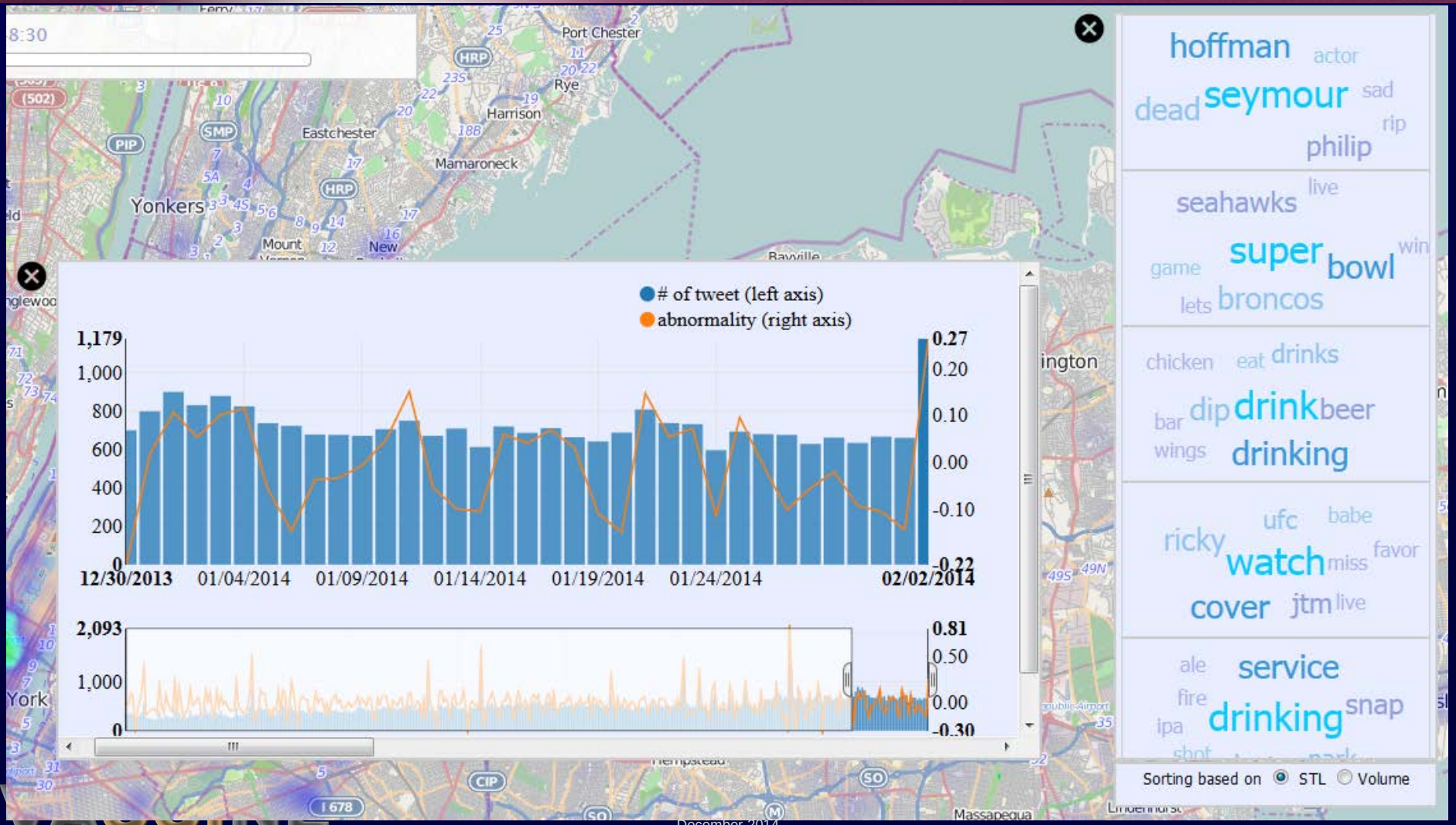
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring event
  - Purdue football games – the p
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – L



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

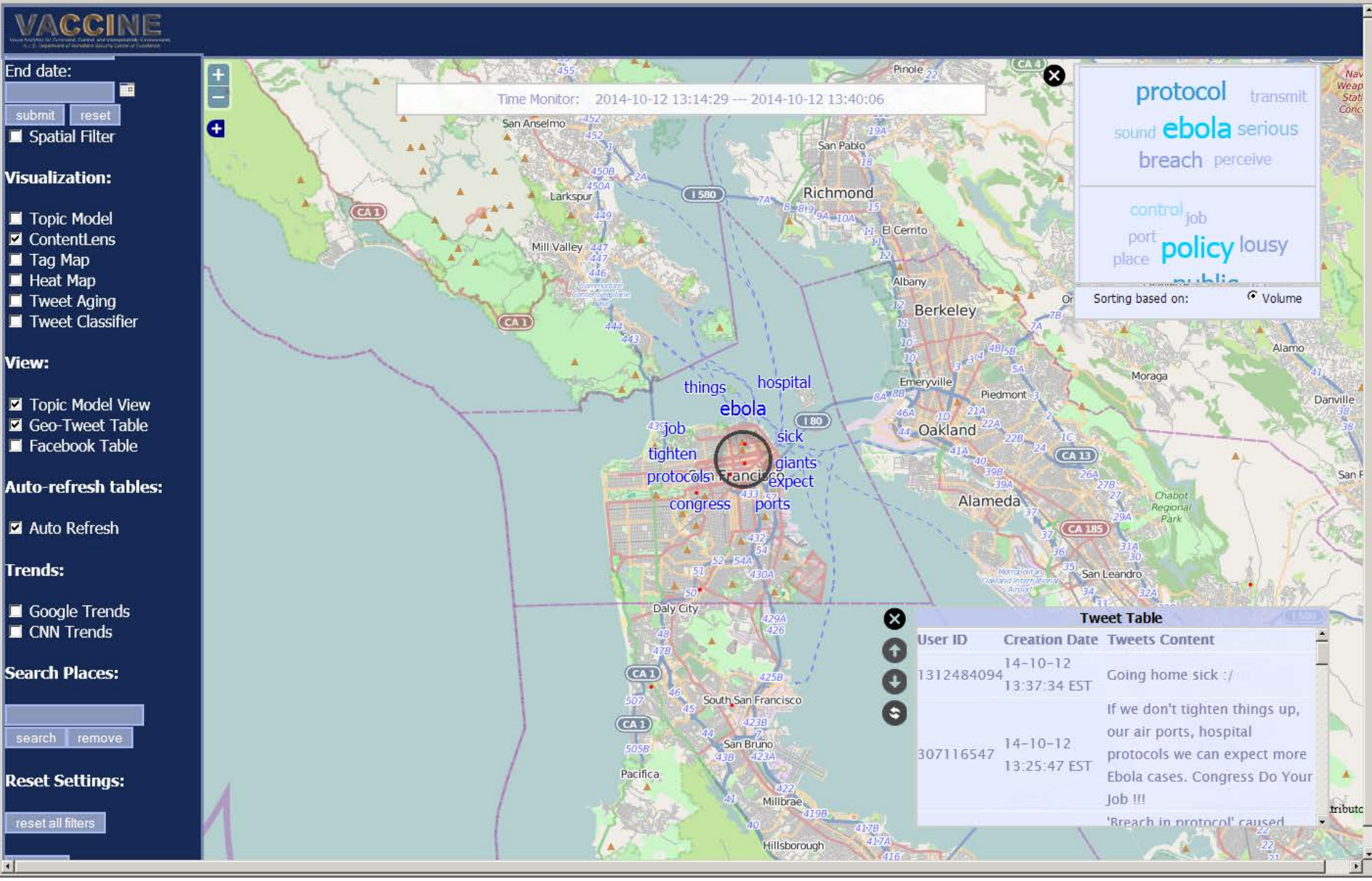


RETWEETS 6

FAVORITES 4

12:56 PM - 12 Oct 2014

Ebola tweets during Fleet Week 10/12/14

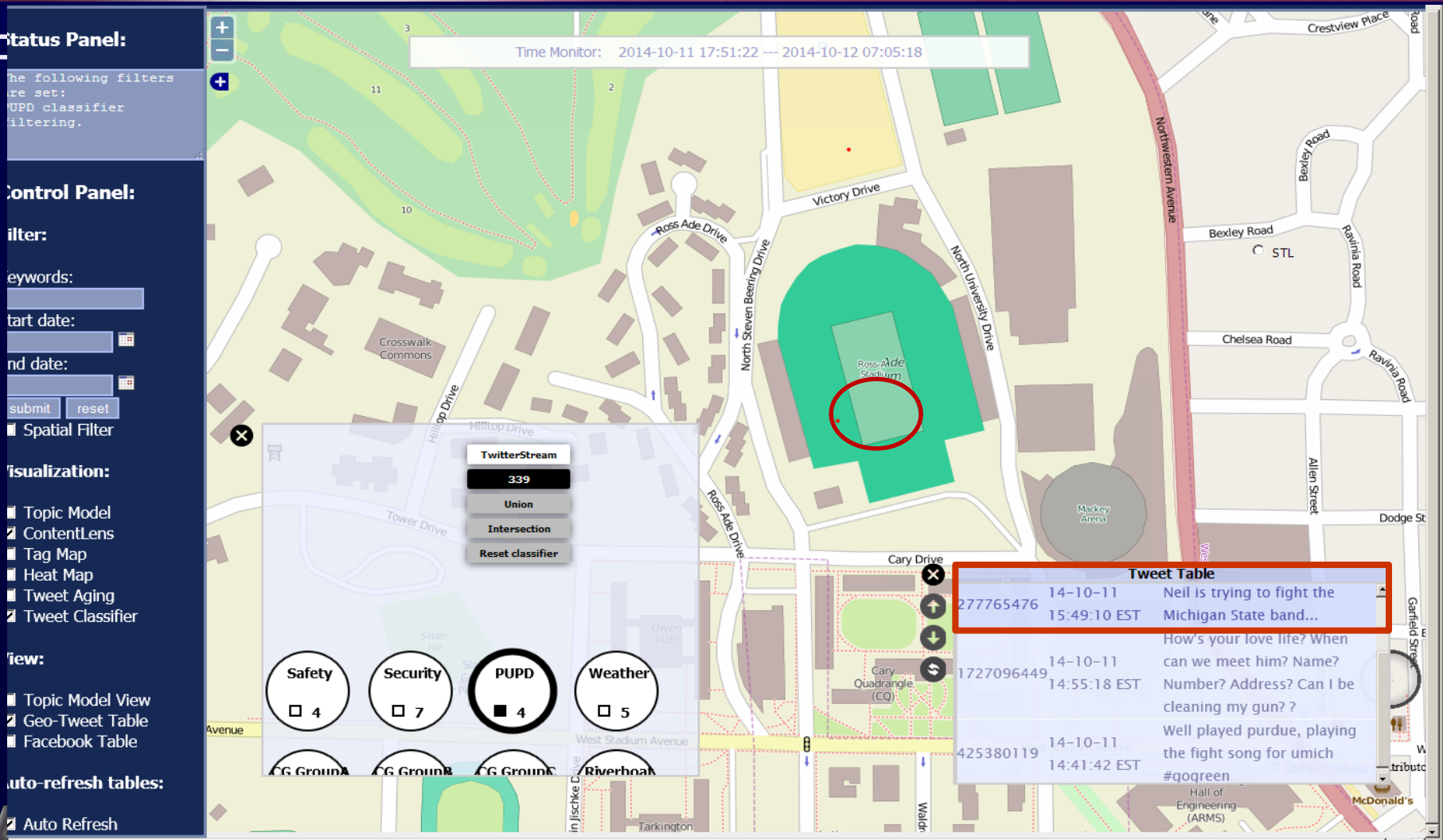


Where Can It Be Helpful?

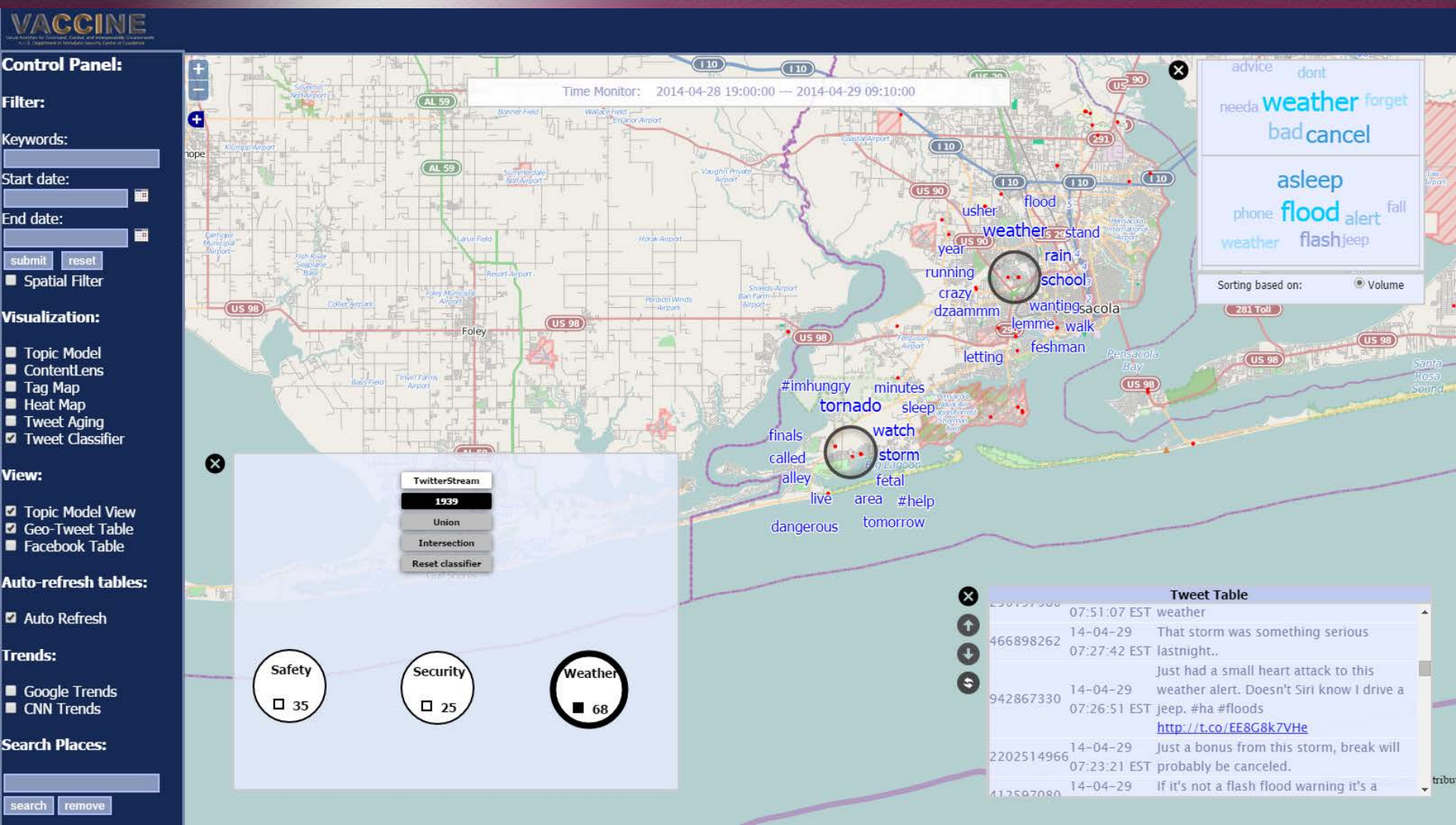
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



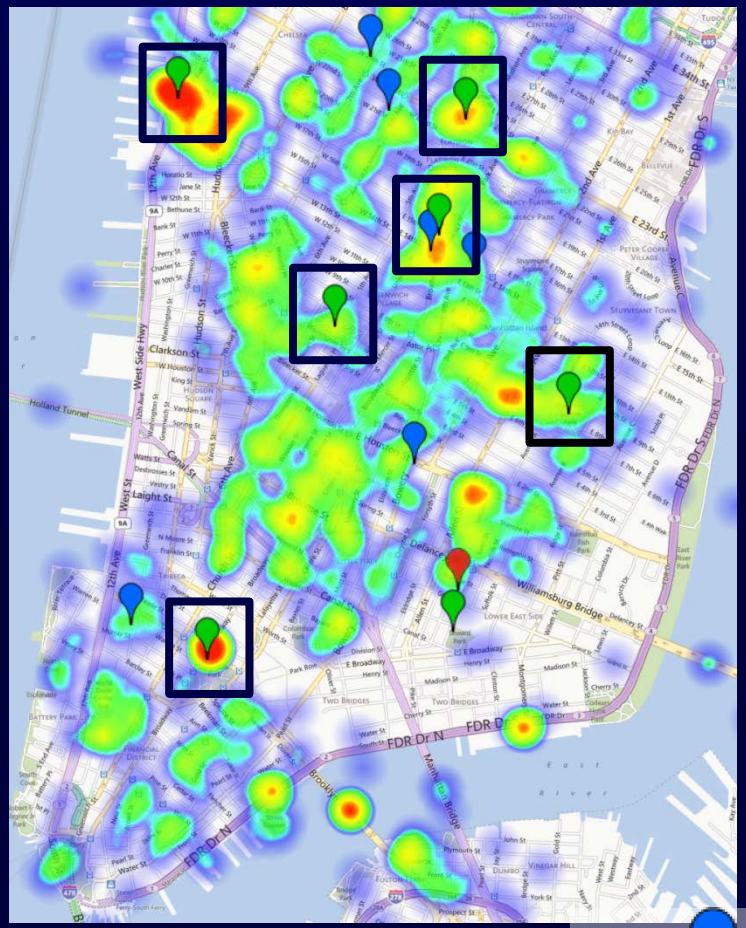
Where Can It Be Helpful?

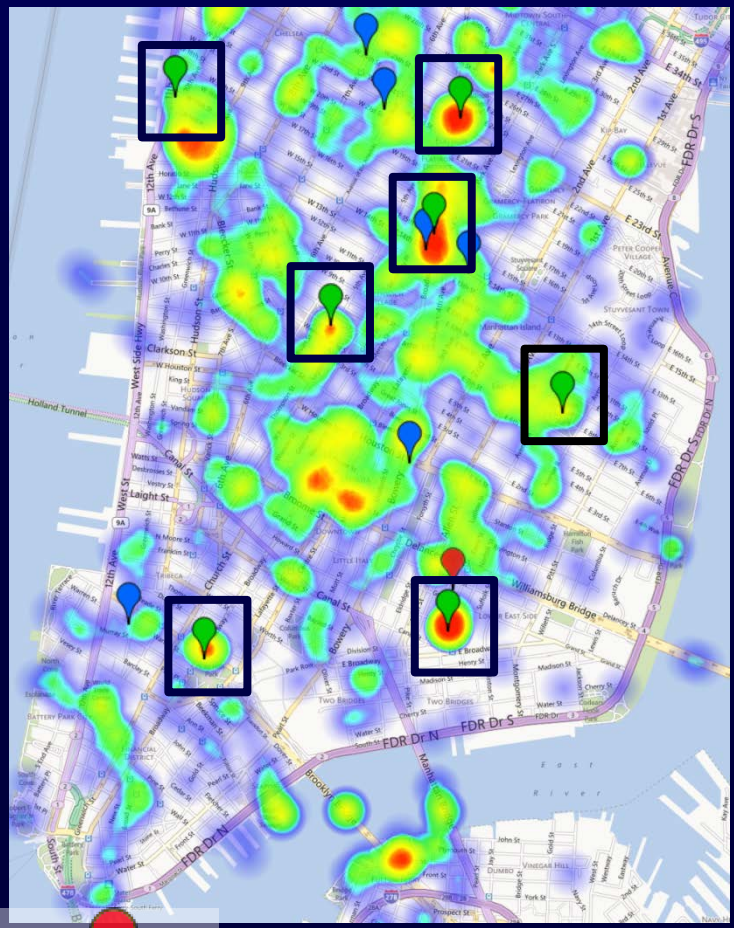


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

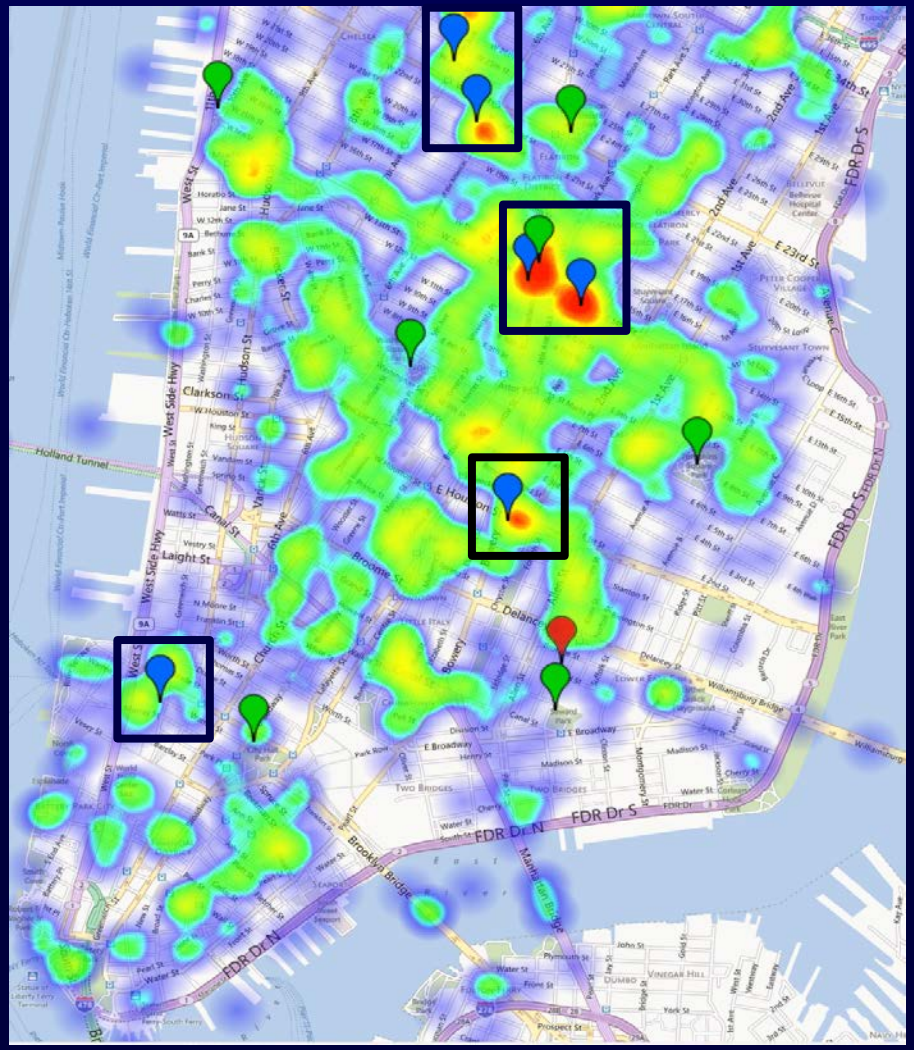
December 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

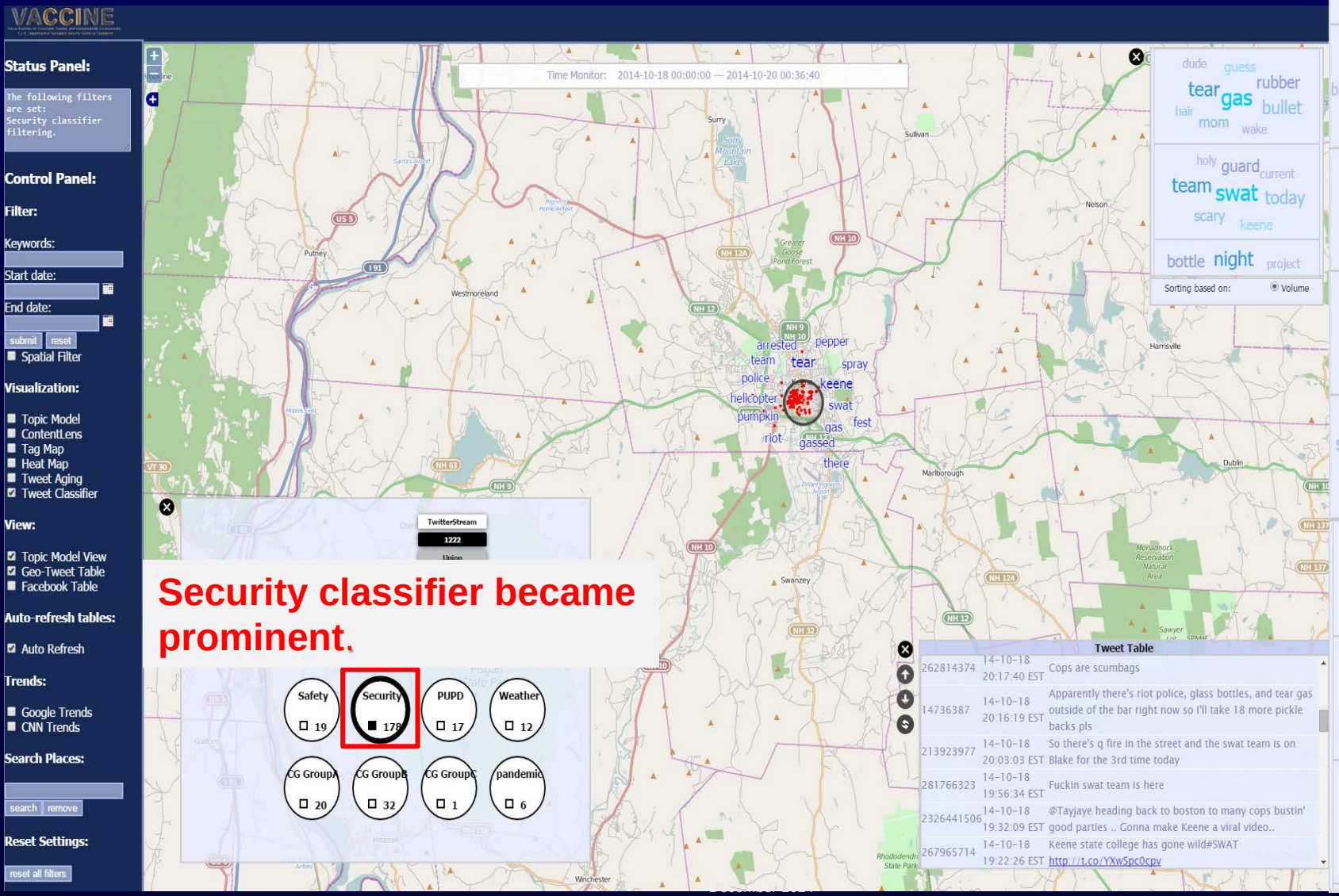
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



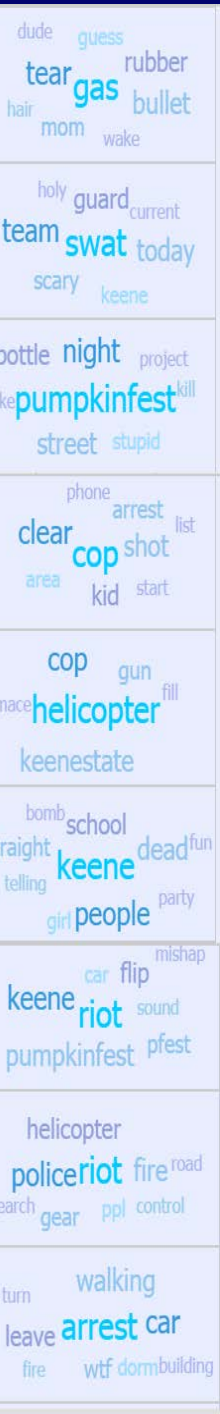
VACCINE

Keene, NH Pumpkin festival 10/18/14 – 10/19/14 Overview

Major topics:

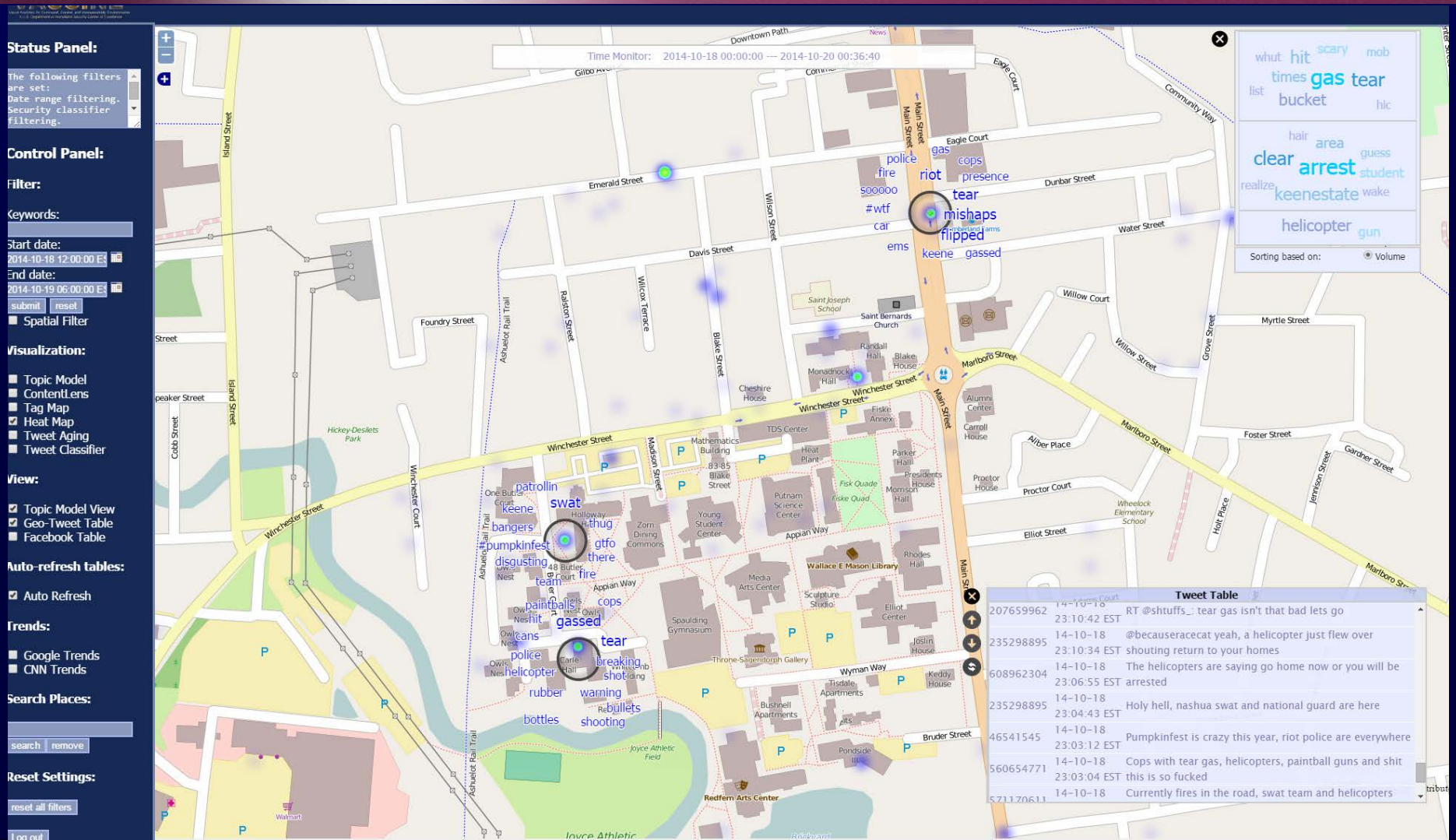


Security classifier became prominent.



Keene, NH

Time range:10/18 12:00pm – 10/19 06:00am, Security classifier)

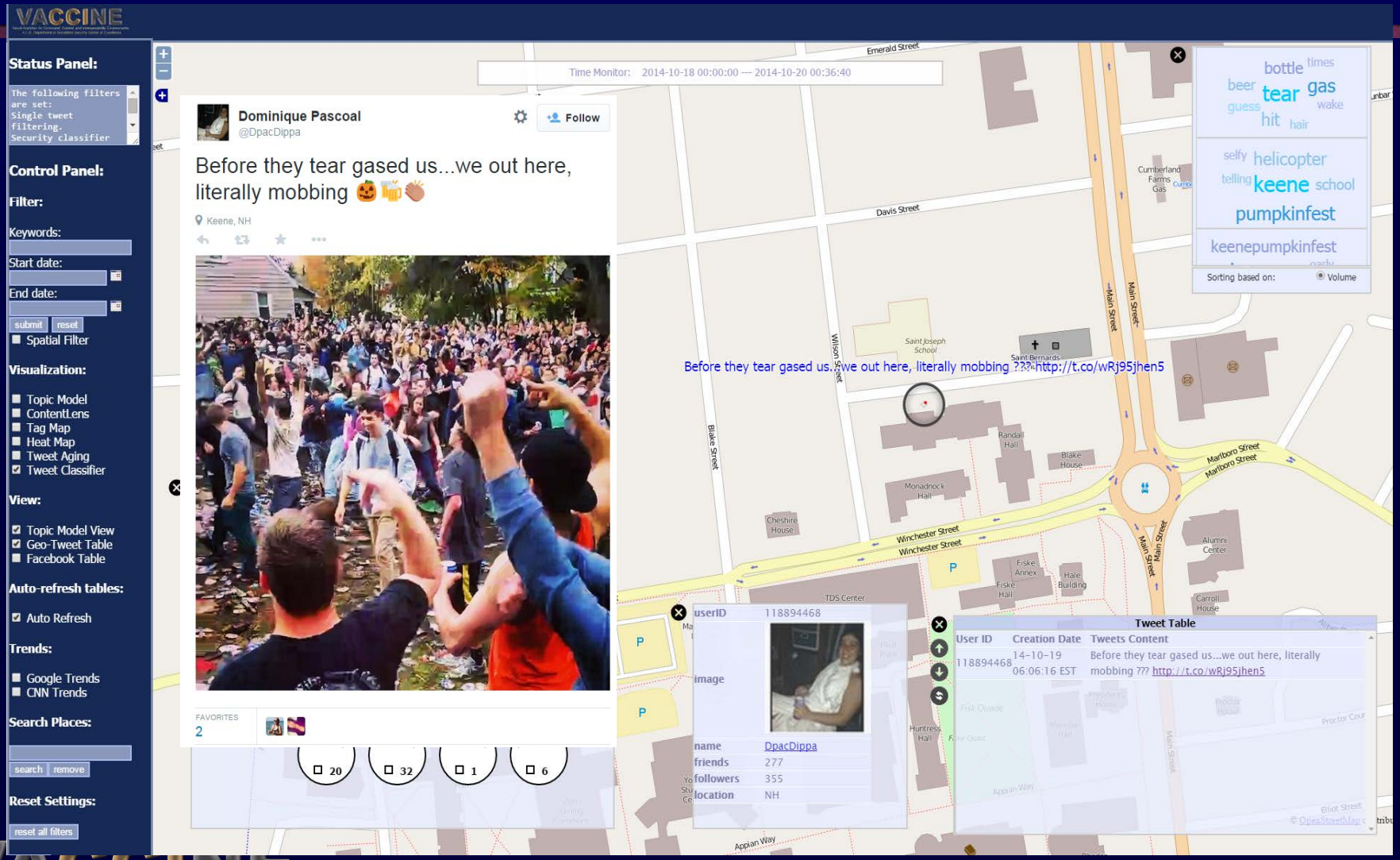


Keene, NH



The picture was later used by ABC news(link).

Keene, NH

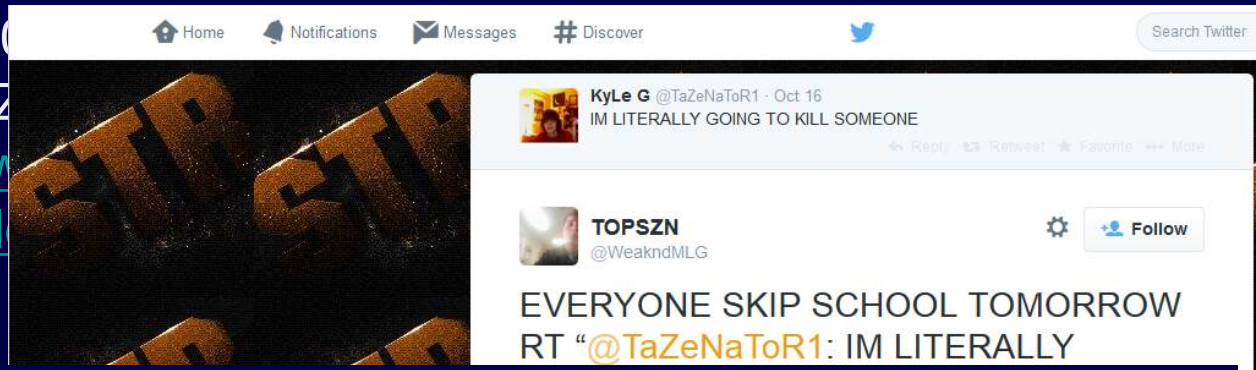

VACCINE
Status Panel:
The following filters are set:
Single tweet filtering.
Security classifier
Control Panel:
Filter:
Keywords:
Start date:
End date:
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
search
remove
Reset Settings:
reset all filters
Time Monitor: 2014-10-18 00:00:00 --- 2014-10-20 00:36:40
Dominique Pascoal
@DpacDippa
Follow
Before they tear gased us...we out here, literally mobbing
Keene, NH
FAVORITES
2
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
bottle
times
beer
tear
gas
guess
wake
hit
hair
selfy
helicopter
telling
keene
school
pumpkinfest
keenepumpkinfest
Sorting based on:
Volume
userID 118894468
image
name DpacDippa
friends 277
followers 355
location NH
Tweet Table
User ID
Creation Date
Tweets Content
118894468
14-10-19 06:06:16 EST
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
Emerald Street
Davis Street
Main Street
Wilson Street
Blake Street
Winchester Street
Marlboro Street
Saint Joseph School
Saint Bernards
Randall Hall
Blake House
Monadnock Hall
Cheshire House
Fiske Annex
Fiske Hall
Hale Building
TDS Center
Alumni Center
Carroll House
Huntress Hall
Appian Way
Cumberland Farms Gas

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



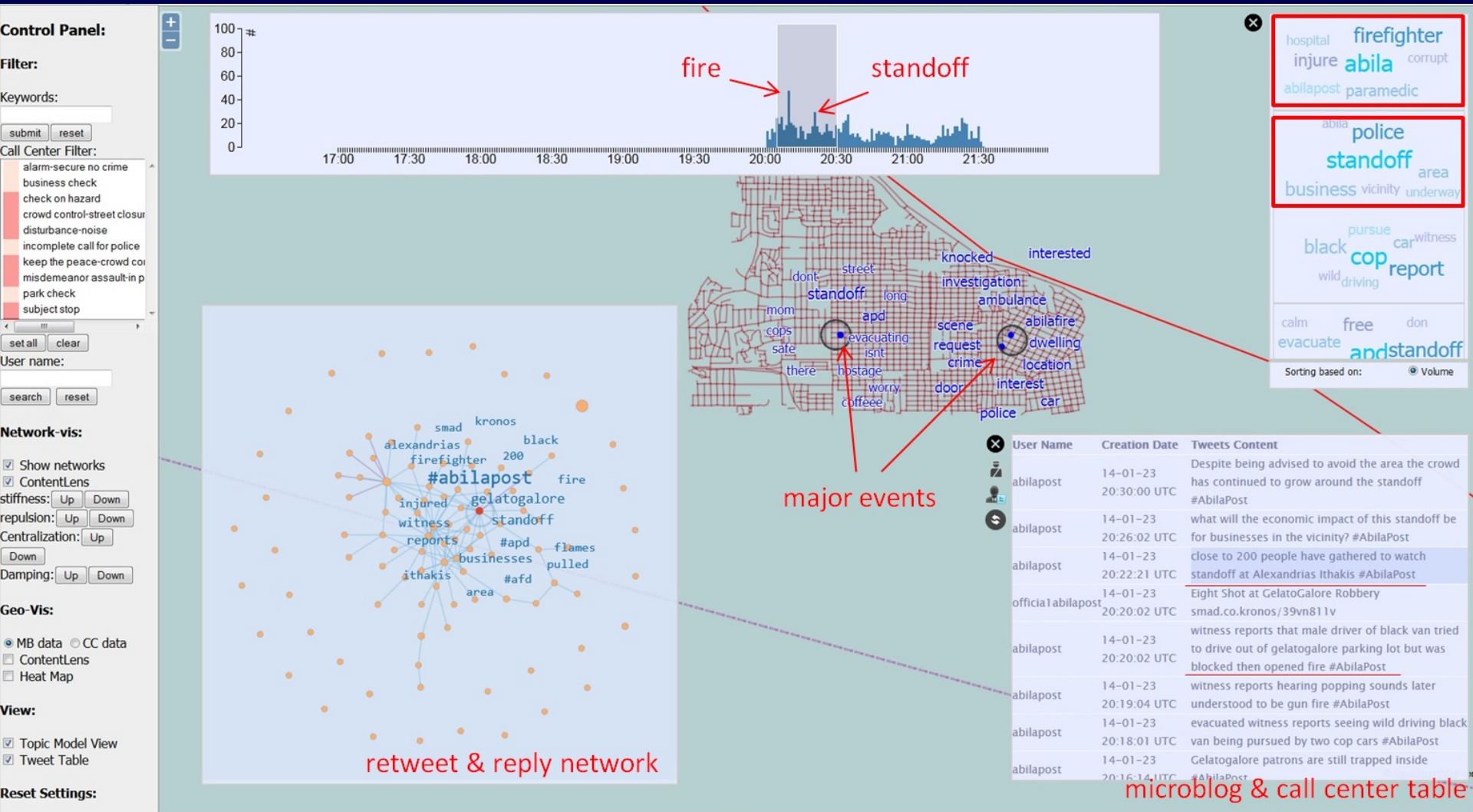
Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

SMART: Network Linkage



Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Directions Forward, Keys to Success & Challenges

Seven Challenges for Proactive & Predictive Visual Analytics

1. Creating computer-human visual cognition environments
2. Integrating interactive simulation and analytics
3. Solving specific scale issues and cross-scale issues
4. Managing uncertainty and time
5. Enabling risk-based decision making environments
6. Developing the Science of Interaction for Visual Analytics
7. Defining appropriate evaluation methodologies for visual analytics

Visual Analytics: Remember…

- We need to be cognizant of parameters for visual representations
- Appropriate analysis can guide users to interesting features in the data
- Refined analysis through user interaction and their domain knowledge can help discover hidden problems
- There is no single catch-all visual representation or analysis

Keys for Success

- User and problem driven
- Balance human cognition and automated analysis and modeling
  - Often applied on-the-fly for specific components identified by the user
- Interactivity and easy interaction
  - Utilizing HPC and novel analysis approaches
- Understandability of why predicted value is what it is
- Intuitive visual cognition
- Not overloaded with features

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Big Data and Visual Analytics: Decision-Makers of the Digital Age

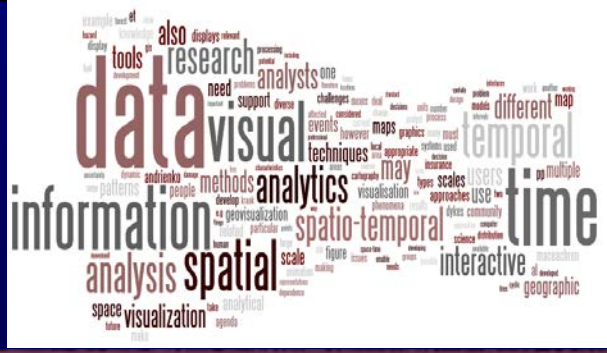
David S. Ebert

School of ECE

Purdue University

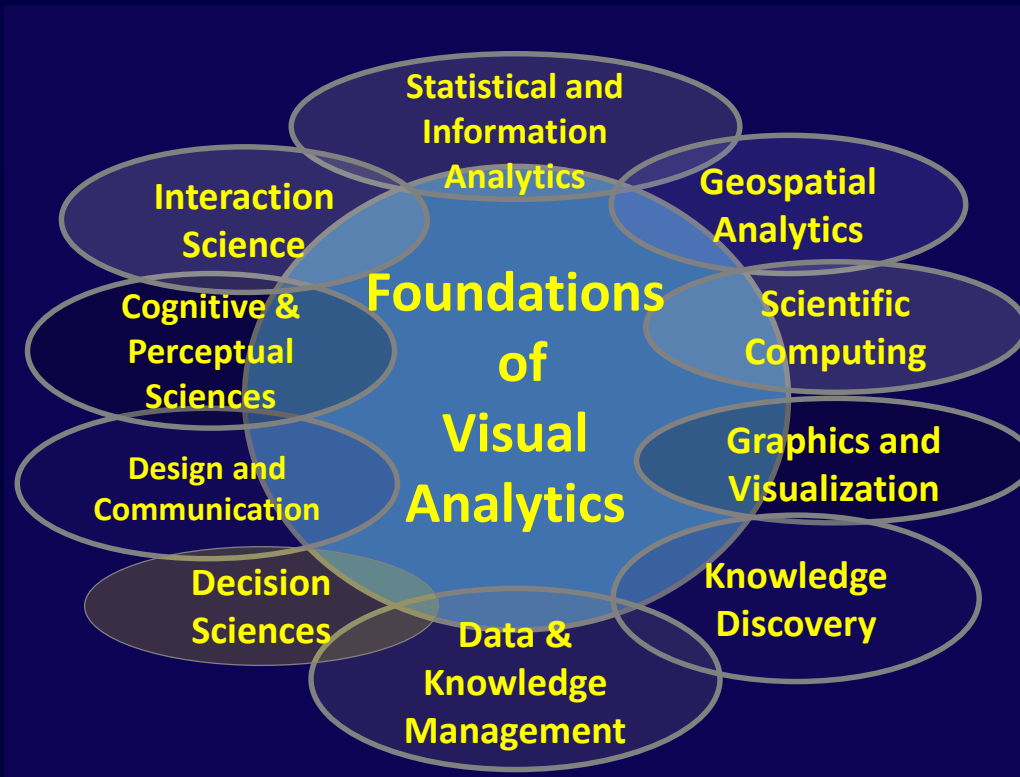
ebertd@purdue.edu

Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

Big Data Challenges



Big Data



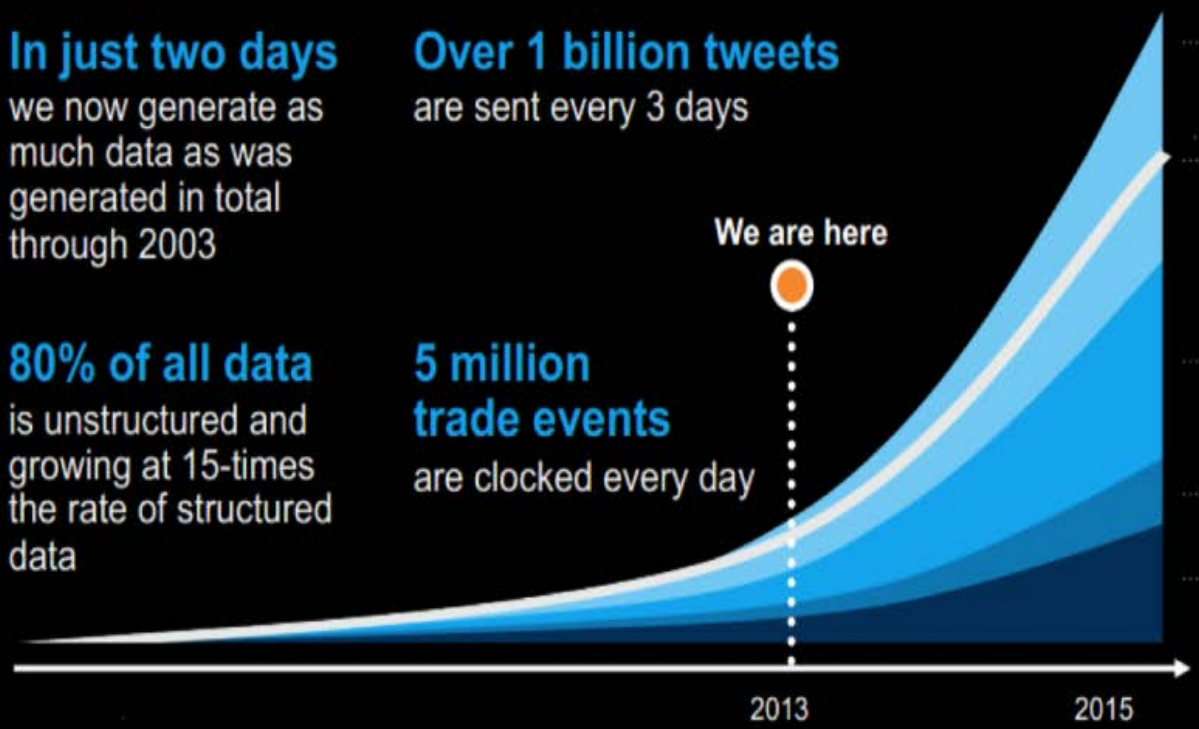
These forces fuel an explosion of data – a new economic asset forming the basis of opportunity

In just two days
we now generate as much data as was generated in total through 2003

Over 1 billion tweets
are sent every 3 days

80% of all data
is unstructured and growing at 15-times the rate of structured data

5 million trade events
are clocked every day



Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

VA Approaches to Big Data

- Don’t display all the data!
- Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space
- Drive process with User interaction and expertise

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

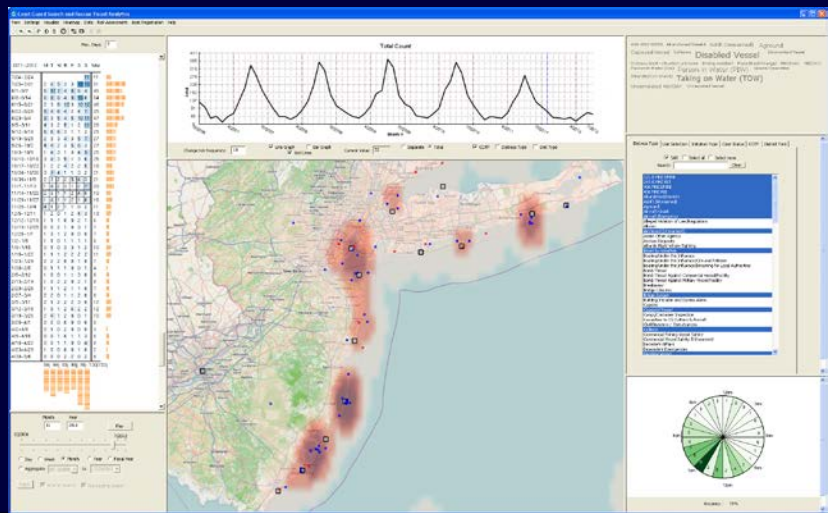
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address



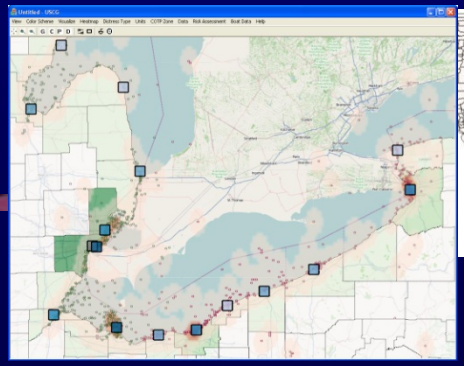
Research Area Overview and Example Projects

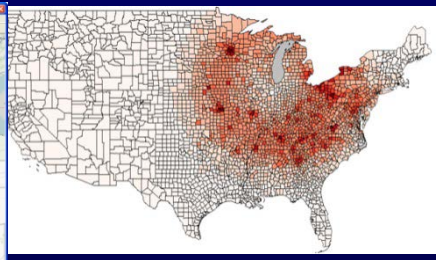
- Research applications areas
- VACCINE
- Example projects
  - Public safety and law enforcement
  - US Coast Guard
  - Social media visual analytics

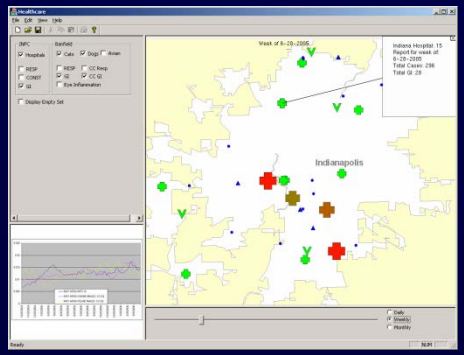


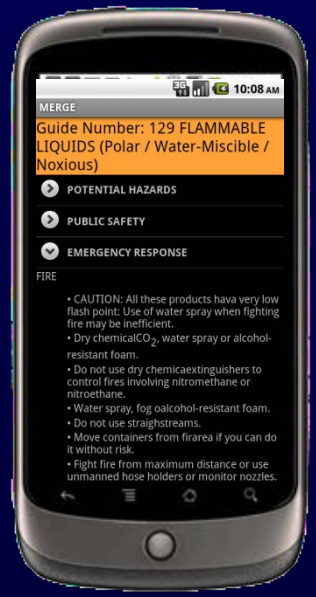
Public Safety & Health

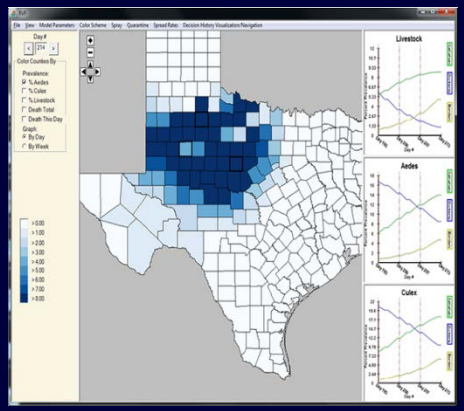
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering

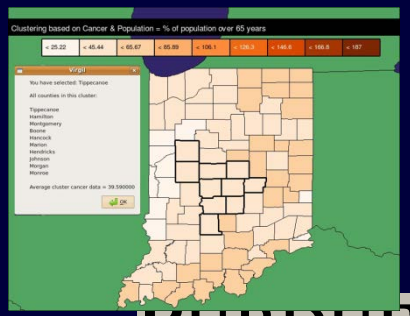








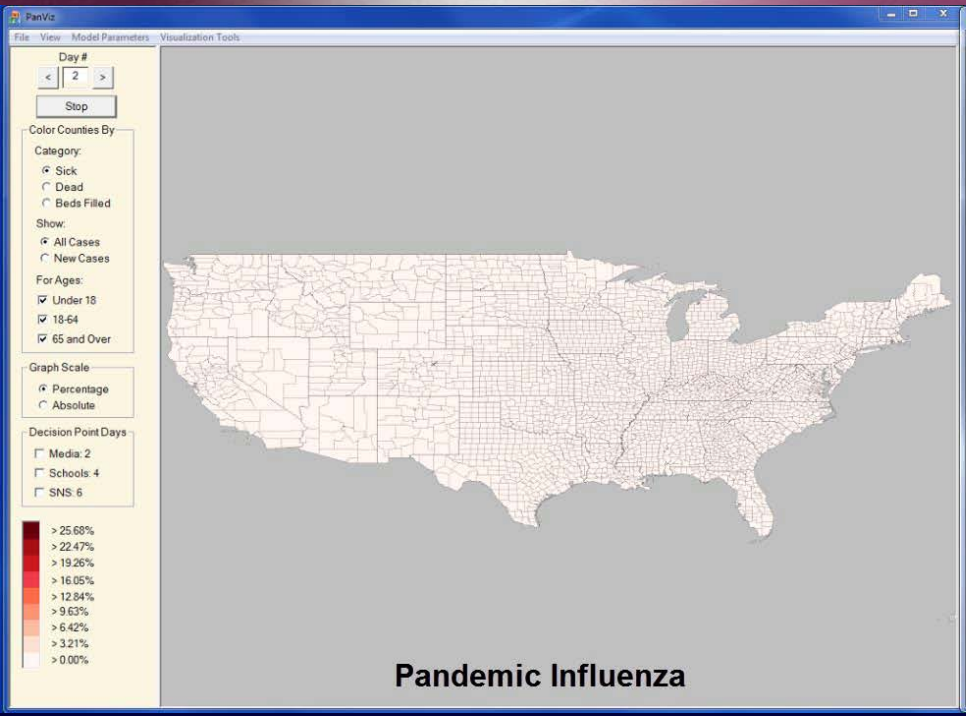


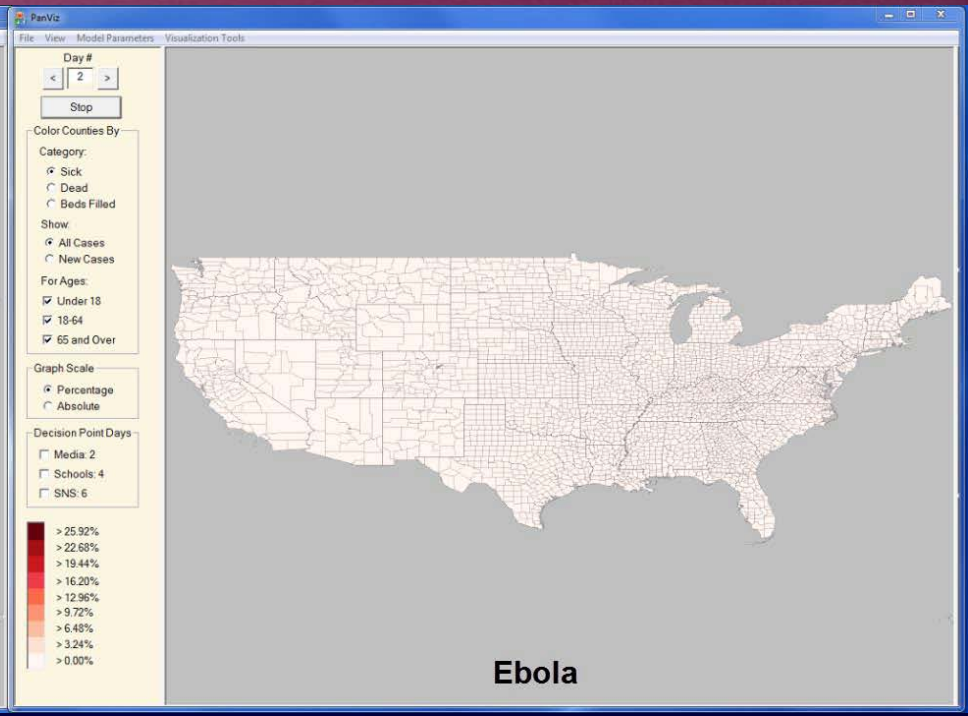


VACCINE

PURDUE UNIVERSITY

Ebola Vs. Pandemic Influenza (H1N1)

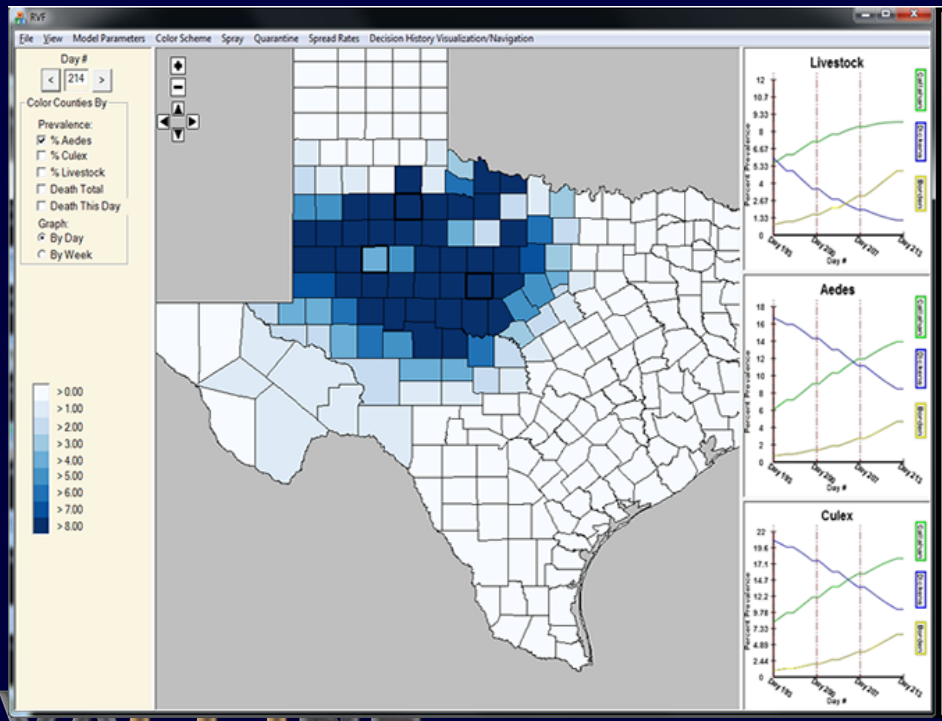


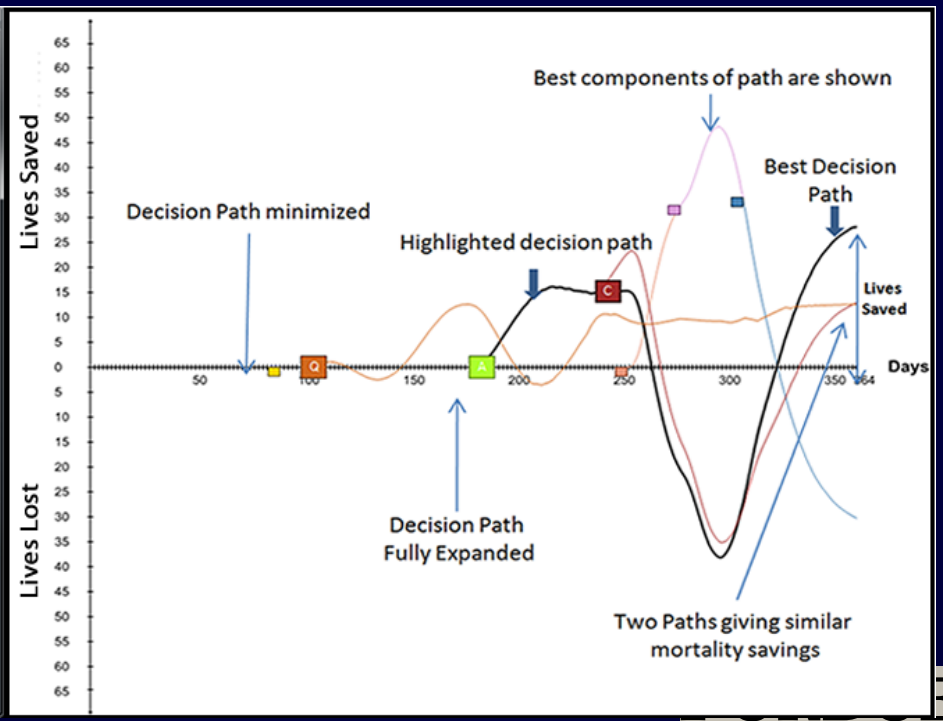


Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)

- Integrated simulation models
- Explore alternative courses of action in decision space and real space

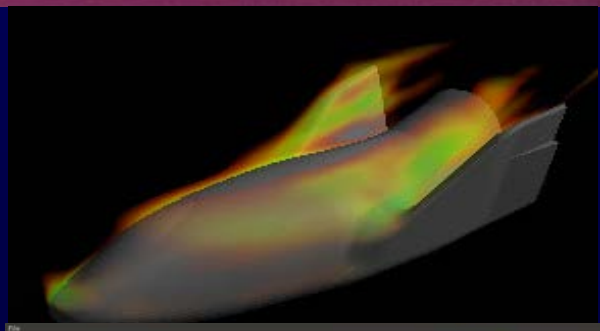


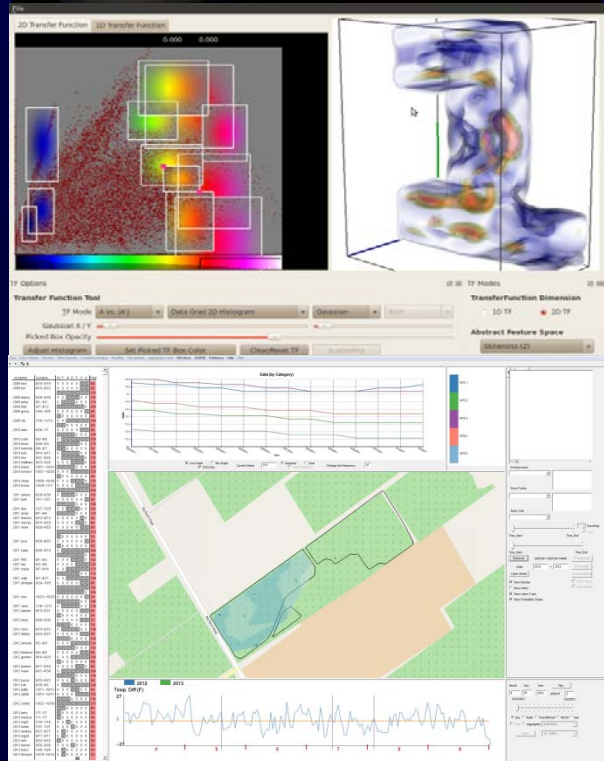


Scientific Discovery

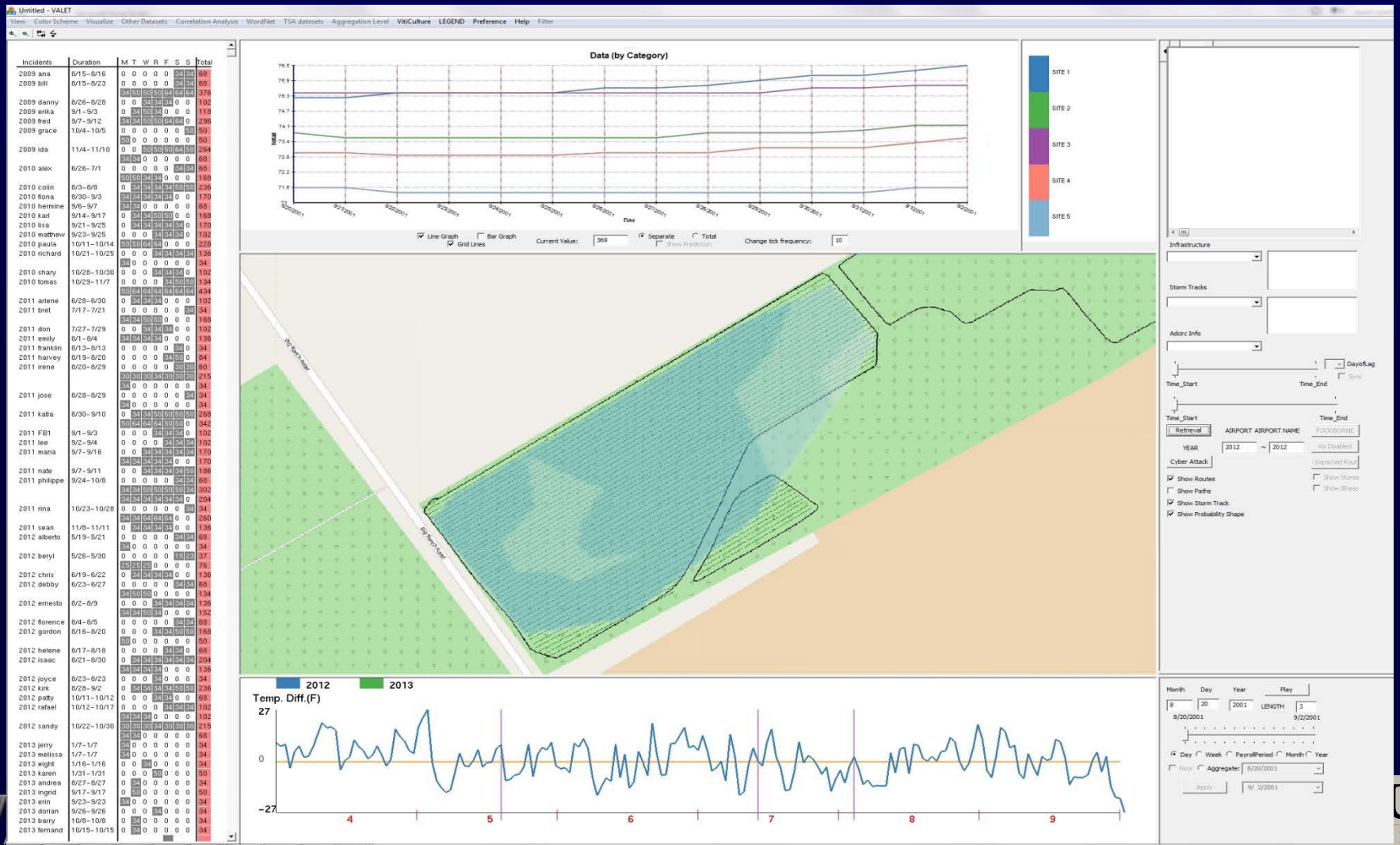


- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality

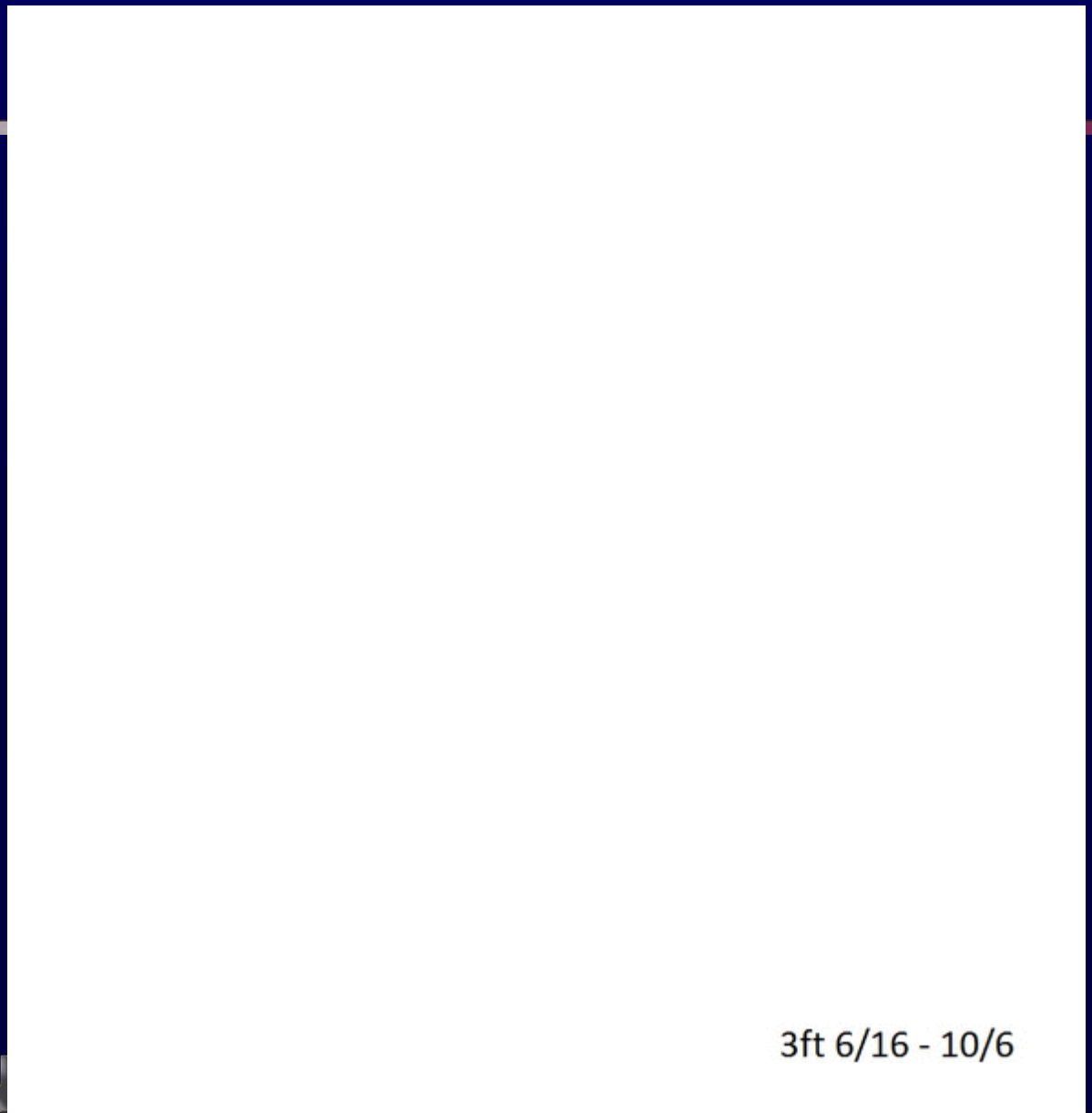




Precision Agriculture VA

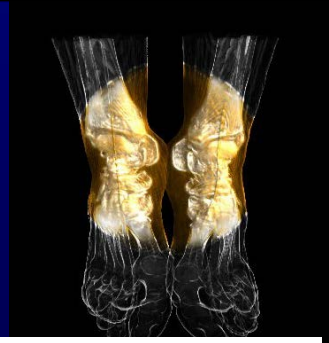


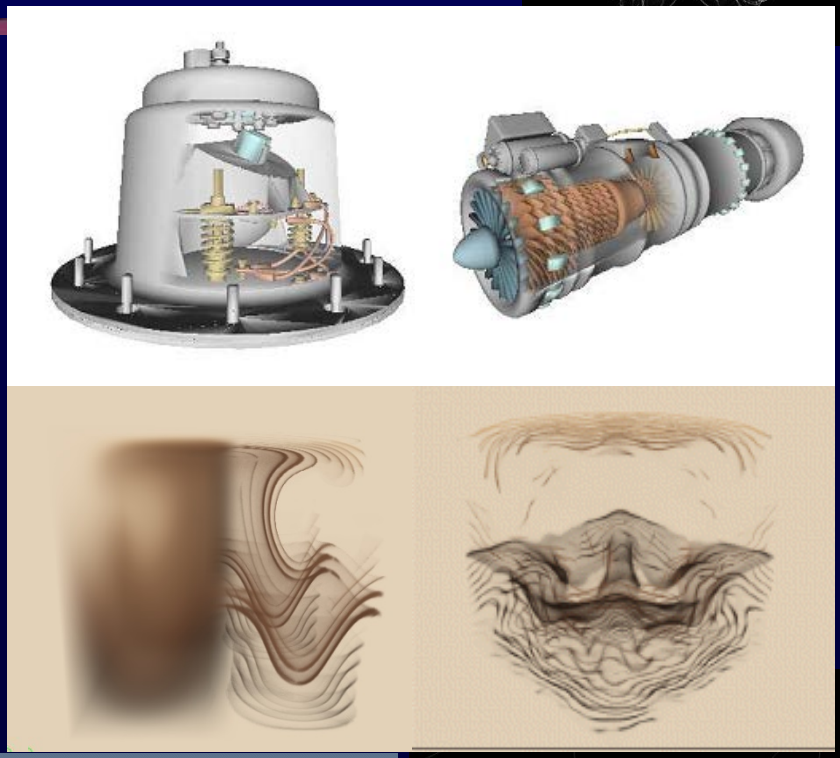
Precision Agriculture VA

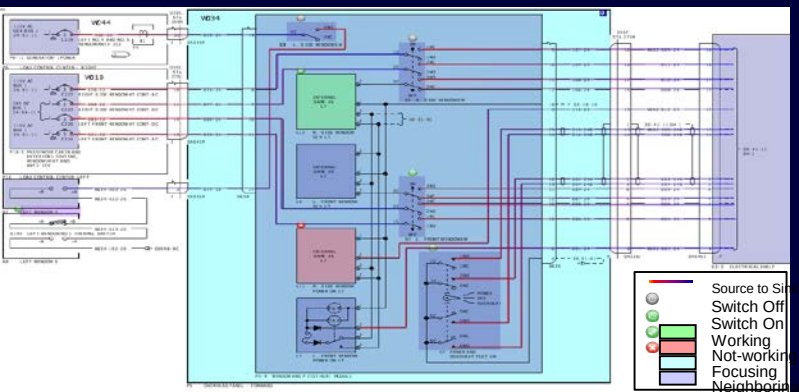


Scientific Discovery

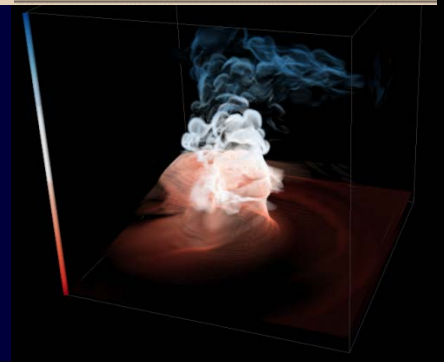
- Illustrative Visualization
  - Mechanical assemblies
  - Schematics and wiring diagrams
  - Illustrative flow visualization
  - Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics





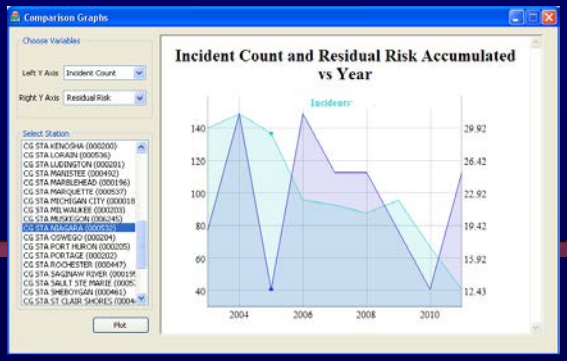


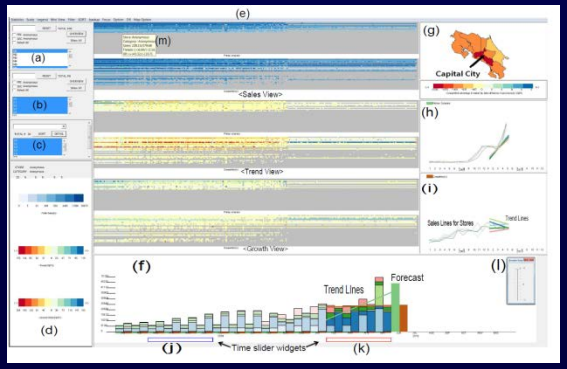




Business Visual Analytics

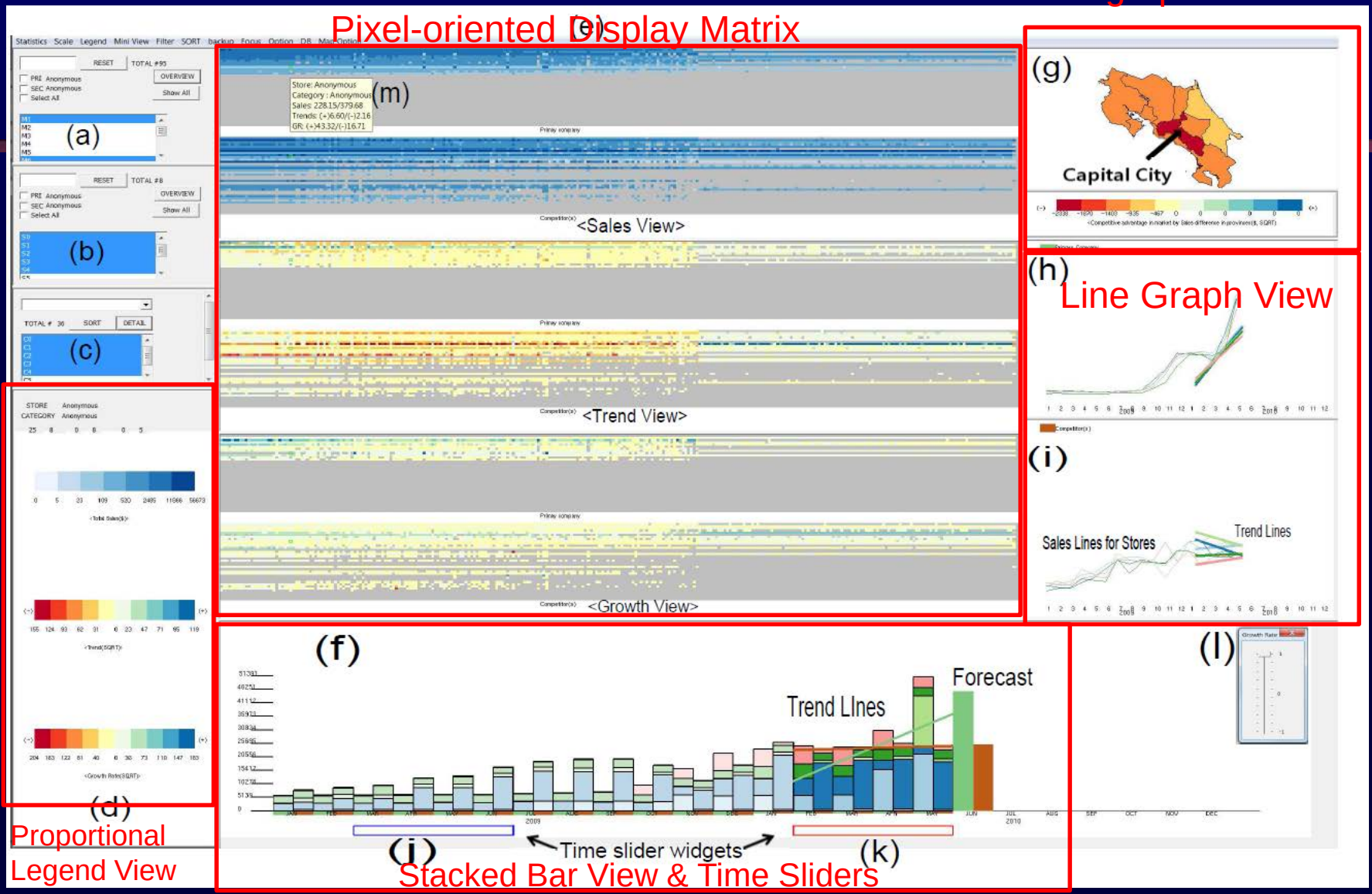
- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







Market Analyzer

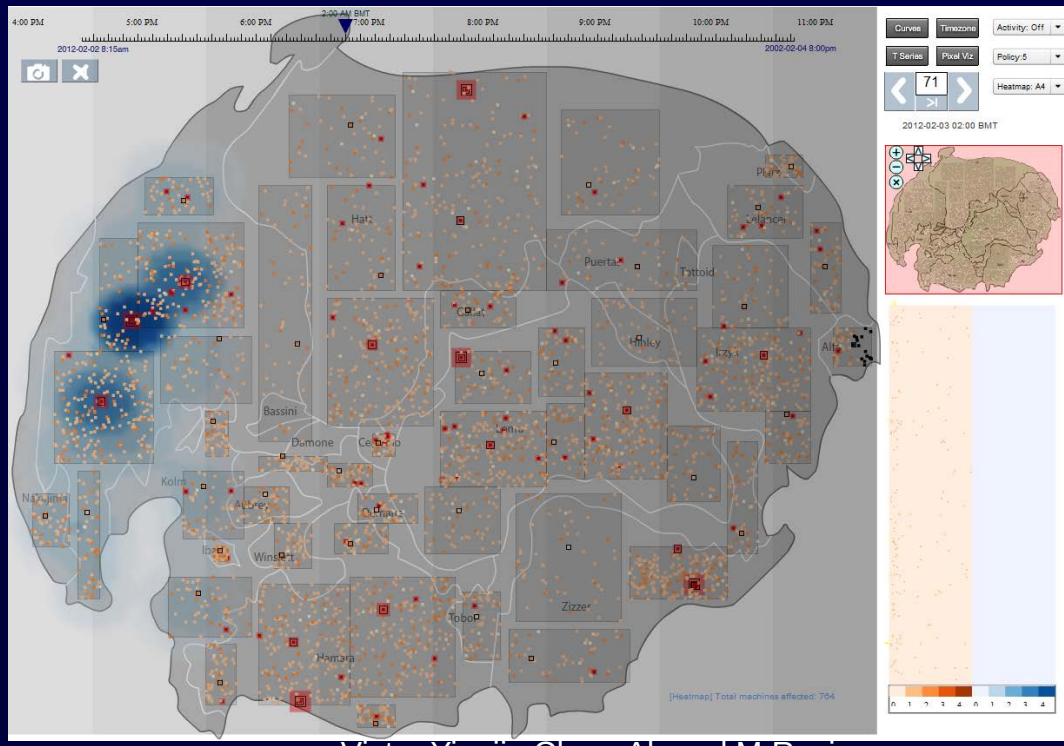


SemanticPrism: A Multi-aspect View of Large High-dimensional Data (Purdue University)

- **VAST 2012 Mini Challenge 1 Award: Outstanding Integrated Analysis and Visualization**

- **Geo-Temporal**
- **Time-serial**
- **Pixel-based**

- **Semantic Zoom**



Victor Yingjie Chen, Ahmad M Razip, Sungahn Ko, Cheryl Zhenyu Qian, David S.Ebert

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

October 2014

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Justice Institute of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America

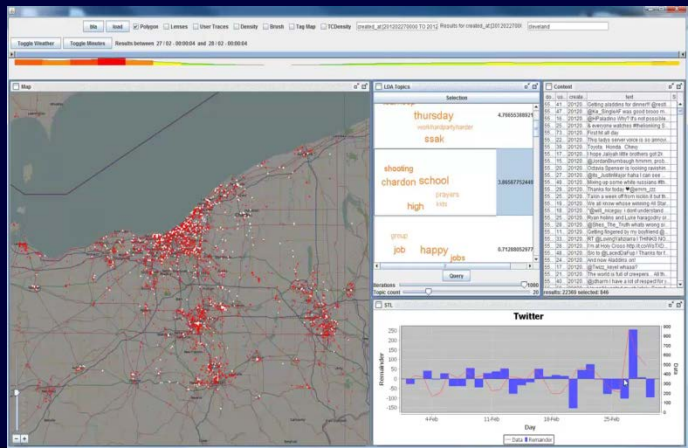




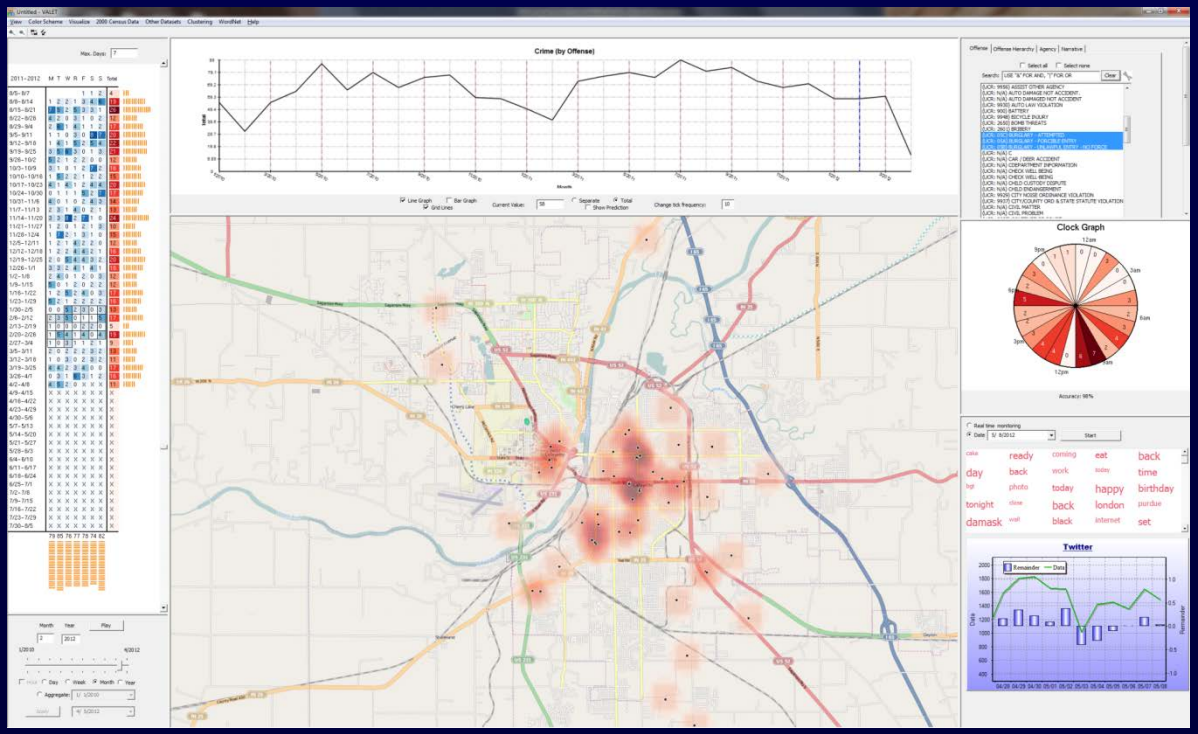
Visual Analytics Uses for Public Safety

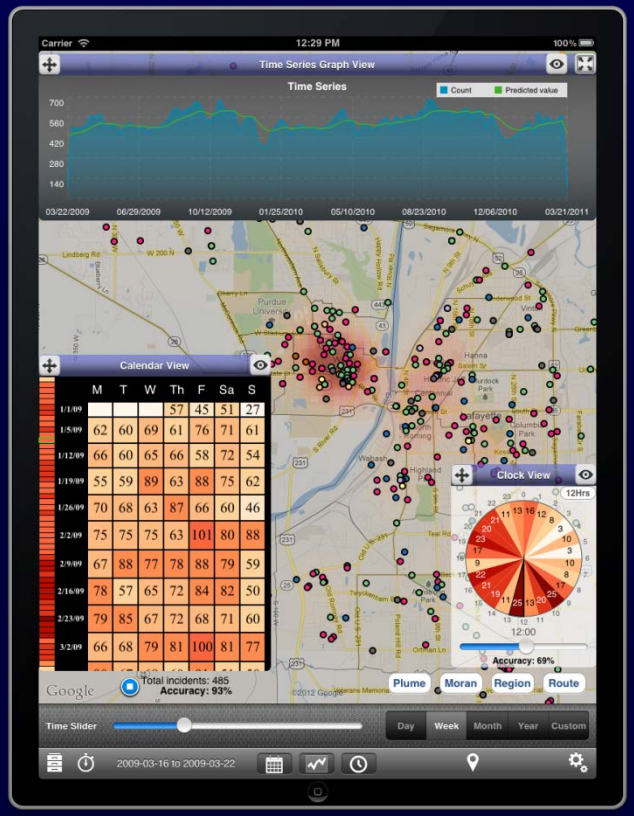


- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



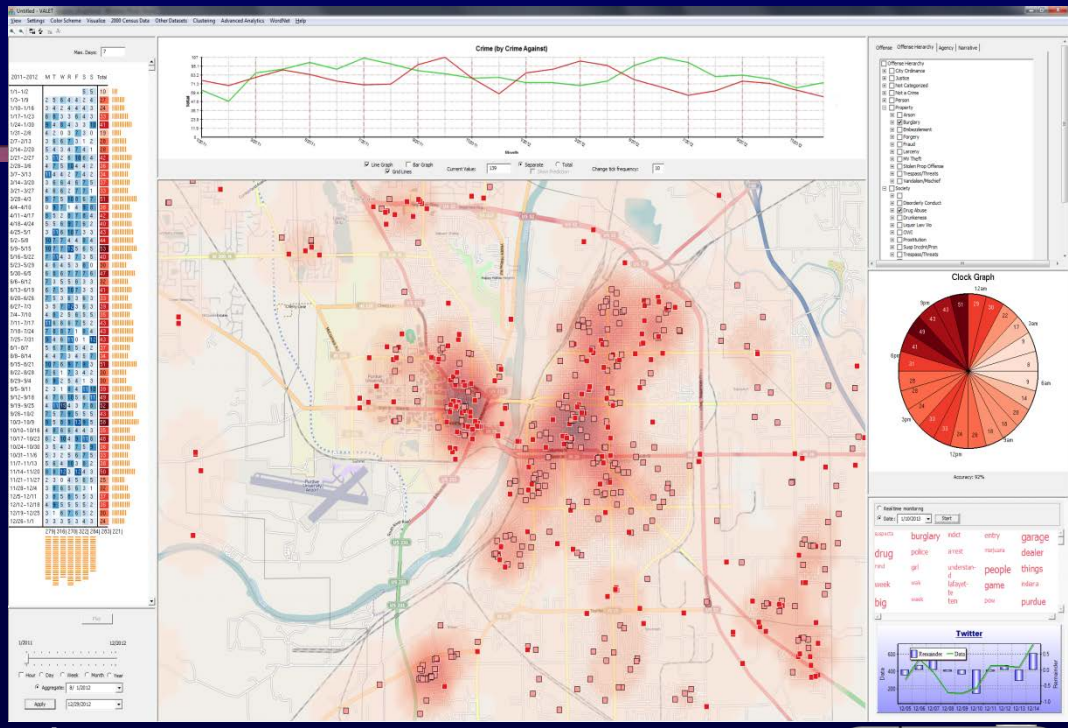
Visual Analytics Law Enforcement Toolkit (i)VALET





Example Projects: Public Safety and Law Enforcement Applications

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

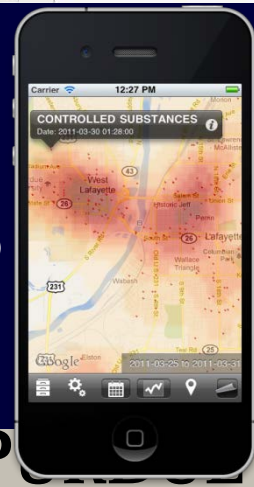
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

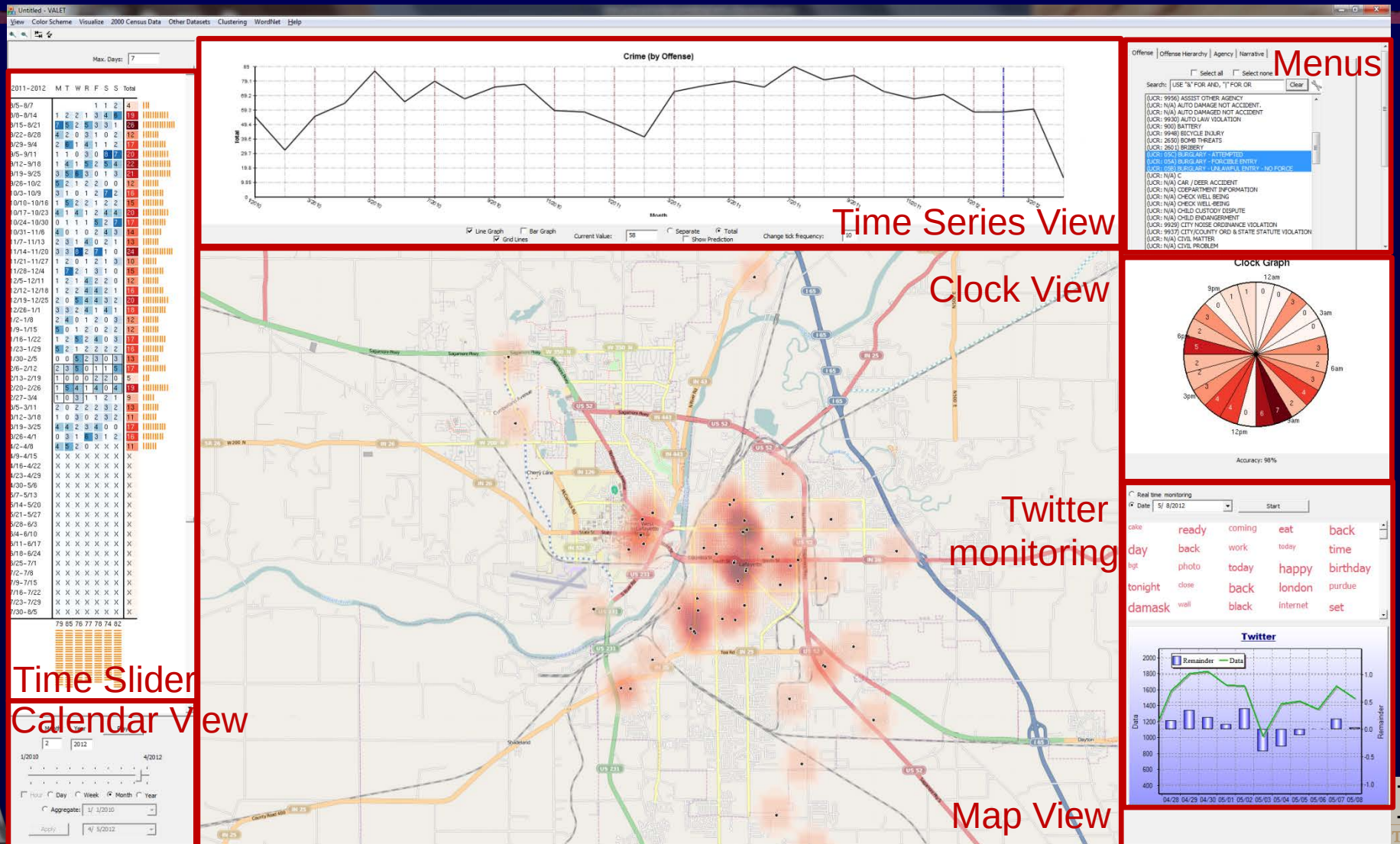
- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:

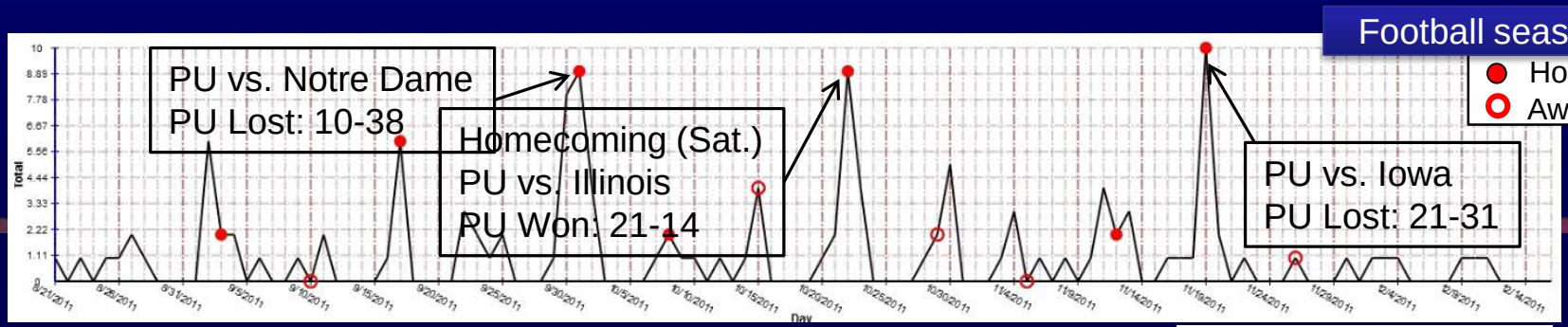
- October 2011: Purdue, WL Police

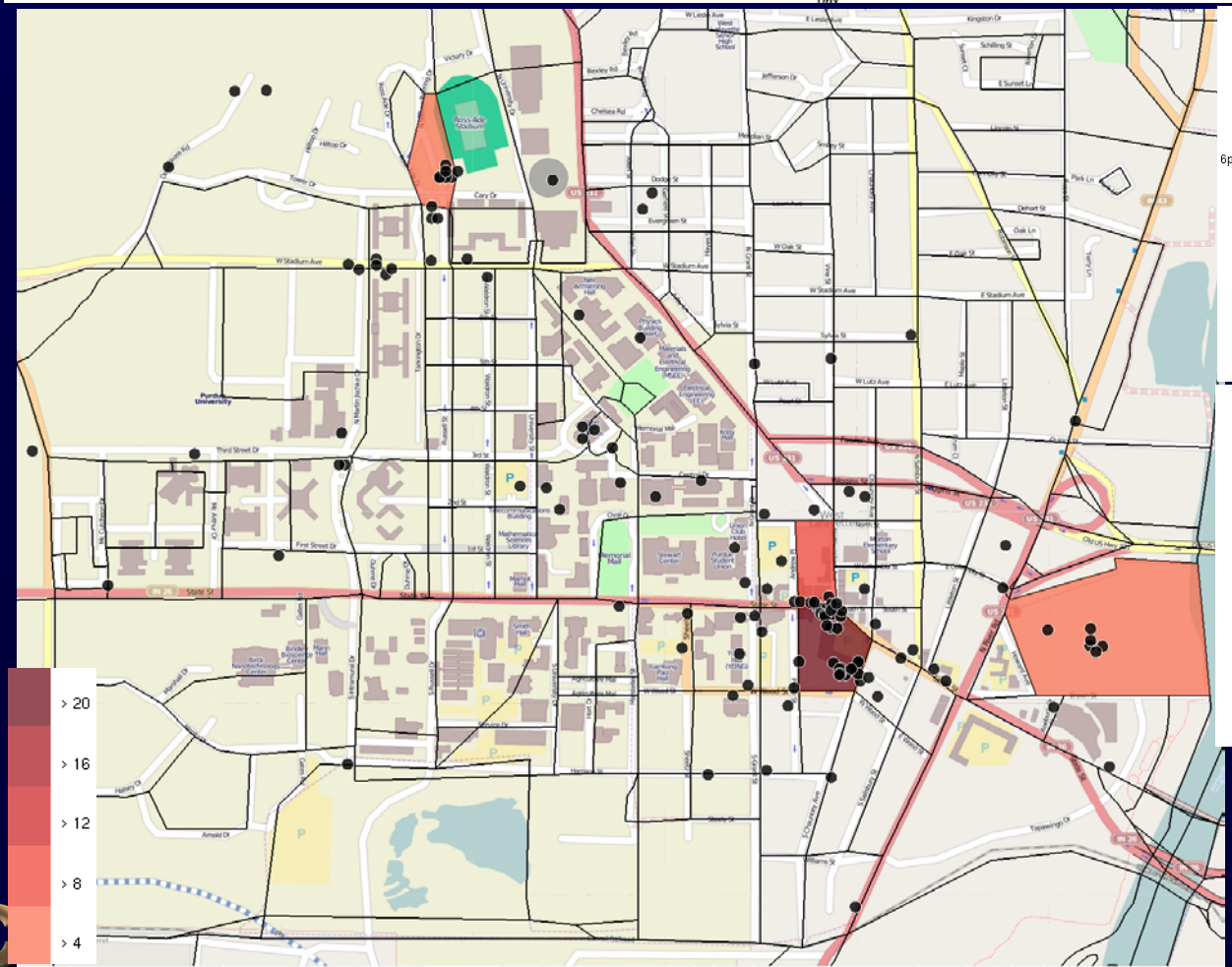


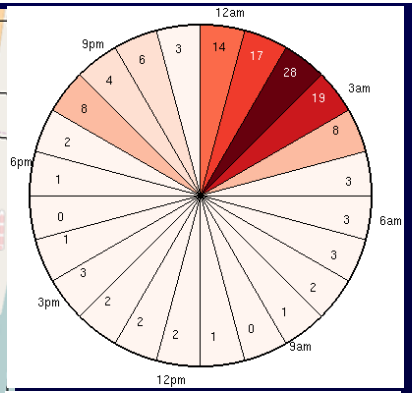
VALET Overview

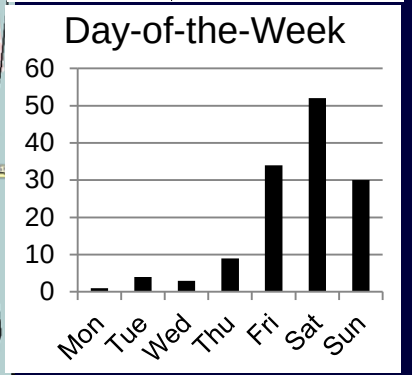


Example: Drunkenness / Public Intoxication



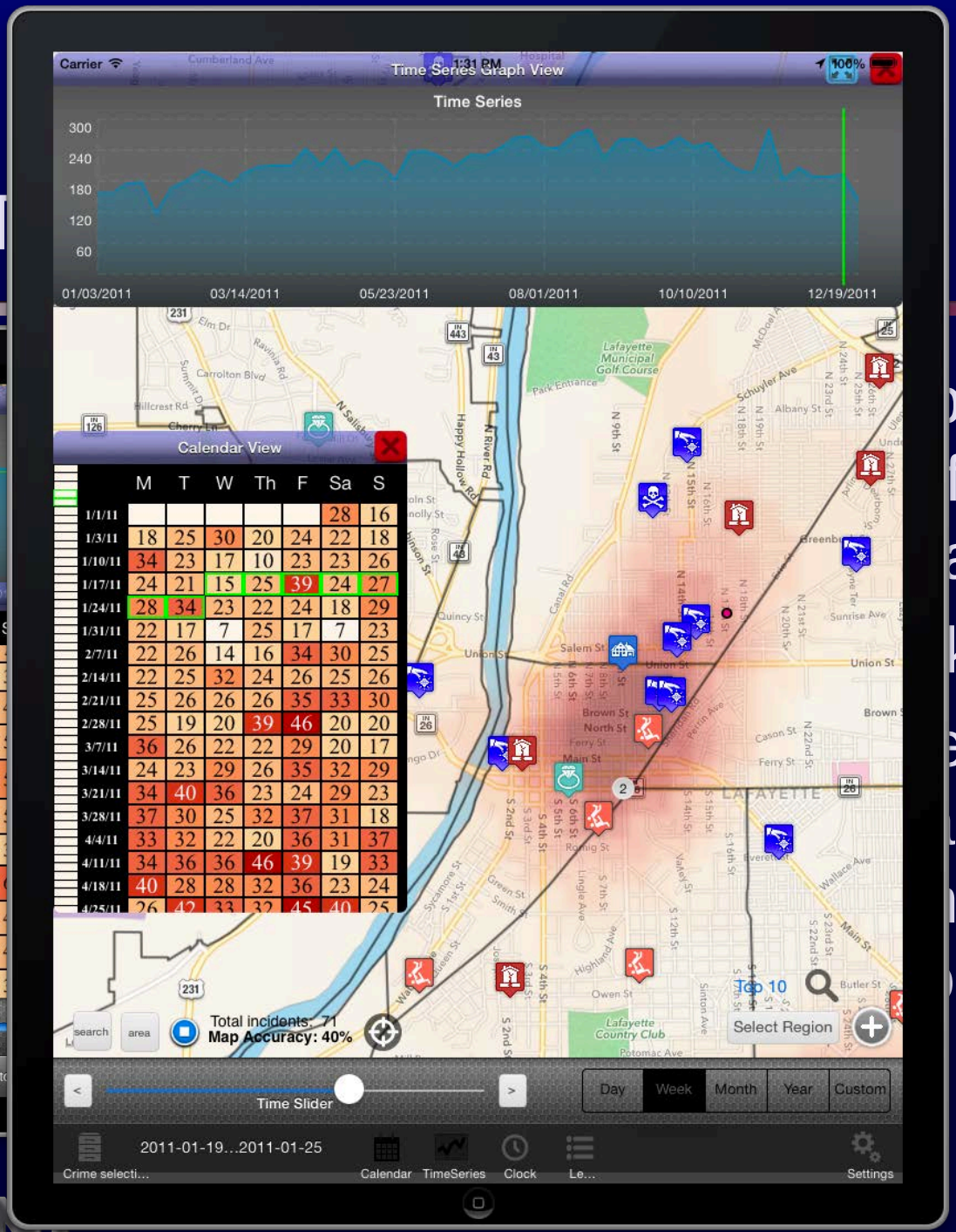




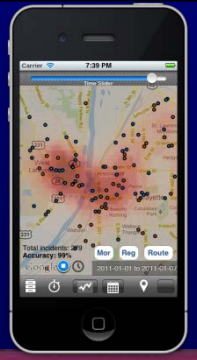


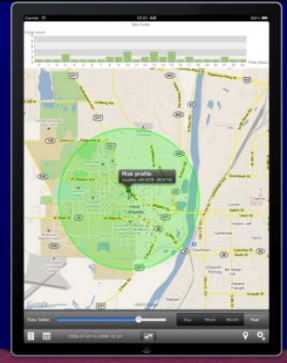
VALET Demo

iVALET











lore criminal,
fic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis

Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)

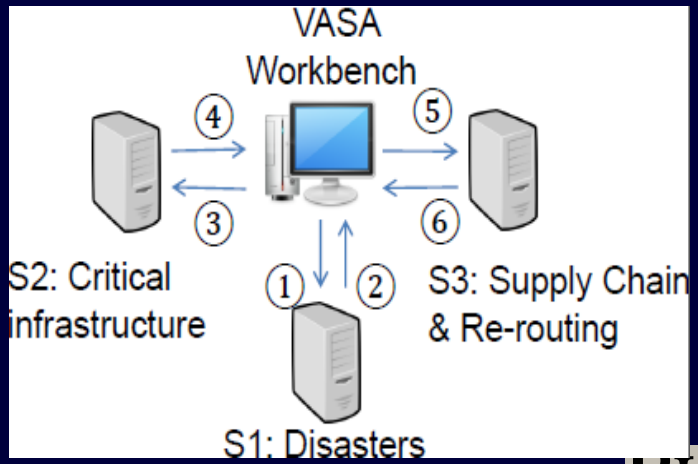


- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC

U Minn



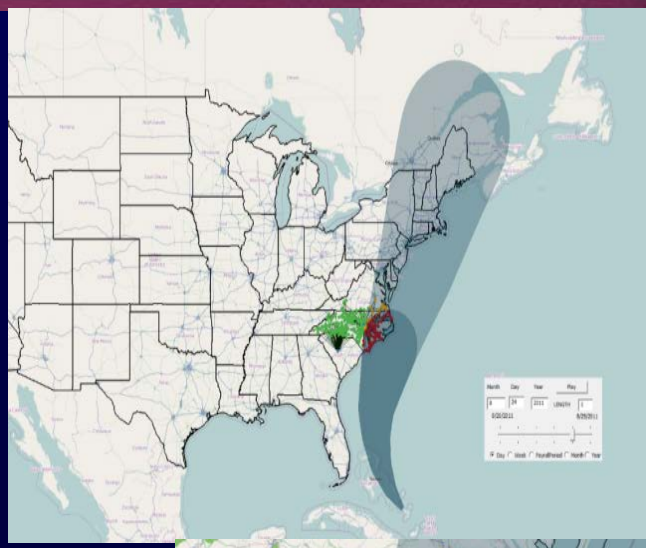
TACC

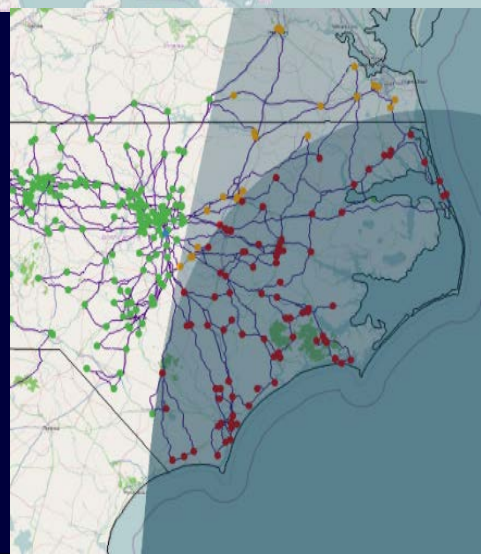
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





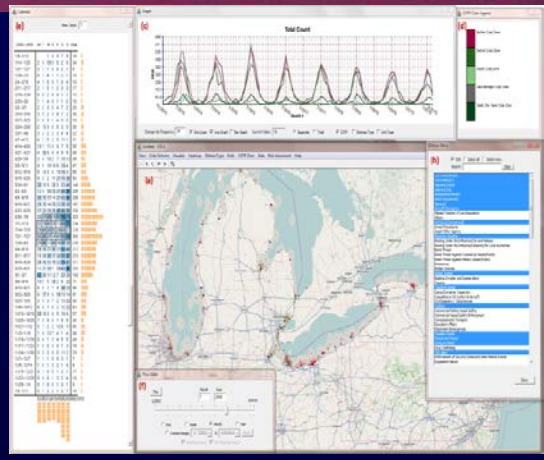
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



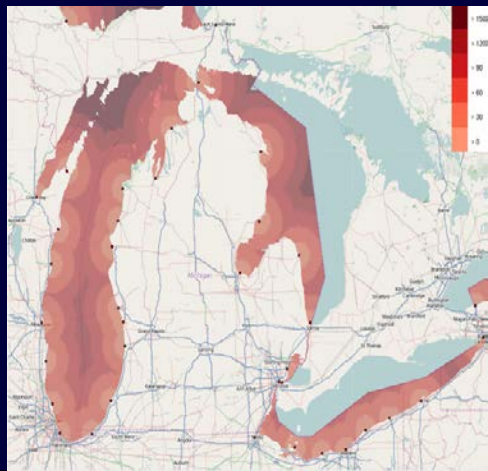
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771



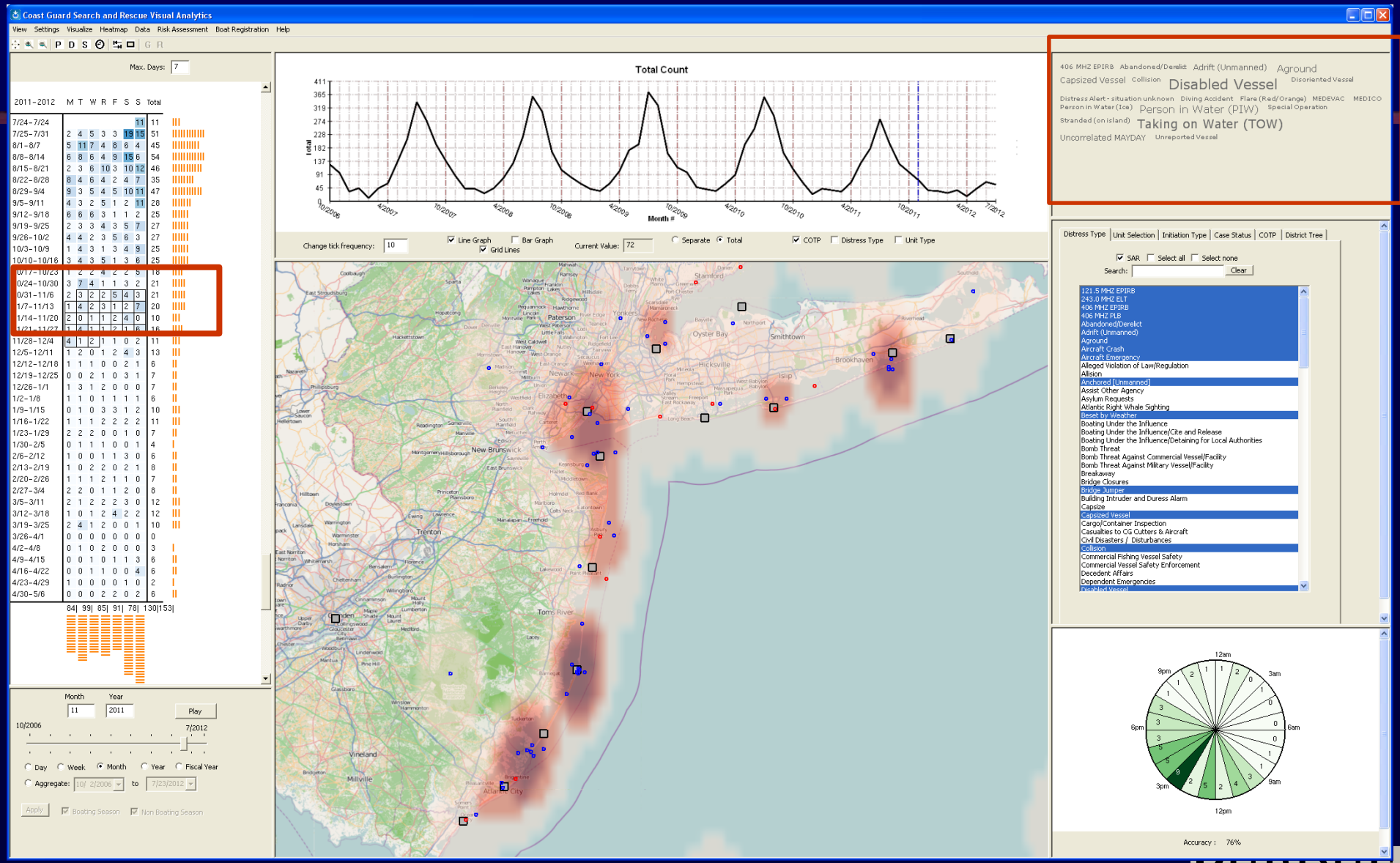
IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

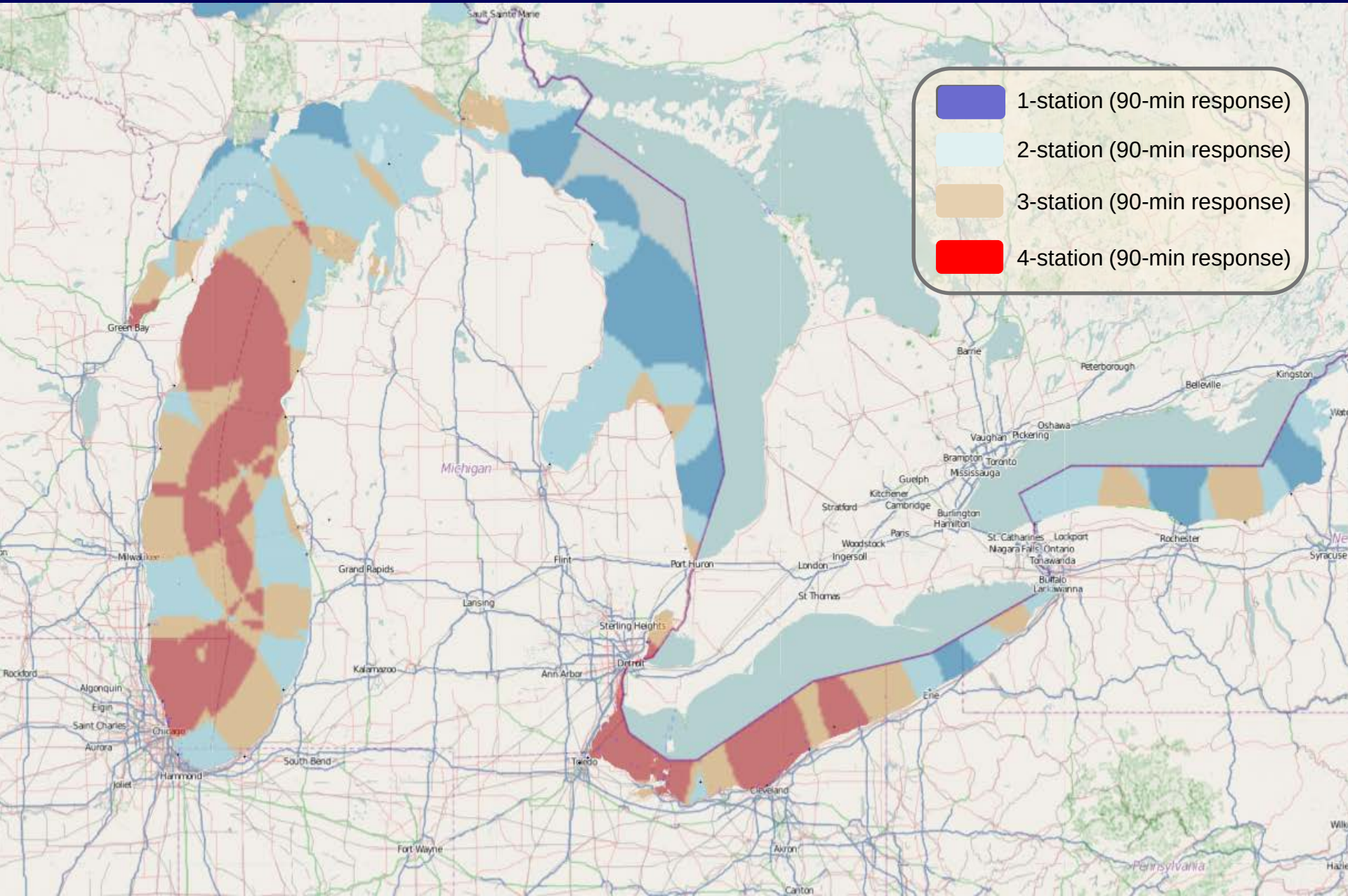


Example: Risks and Consequences From Sandy:

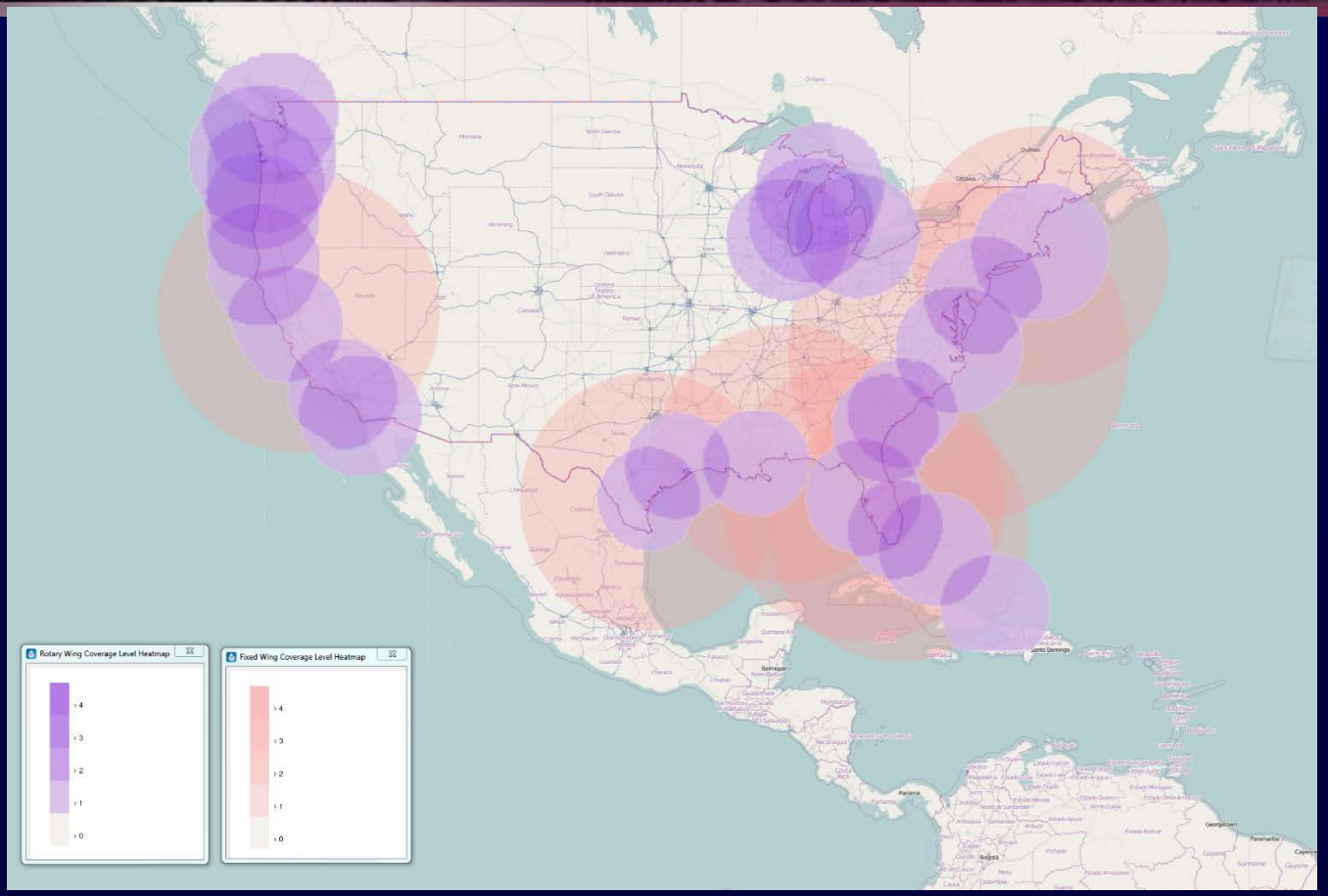
SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



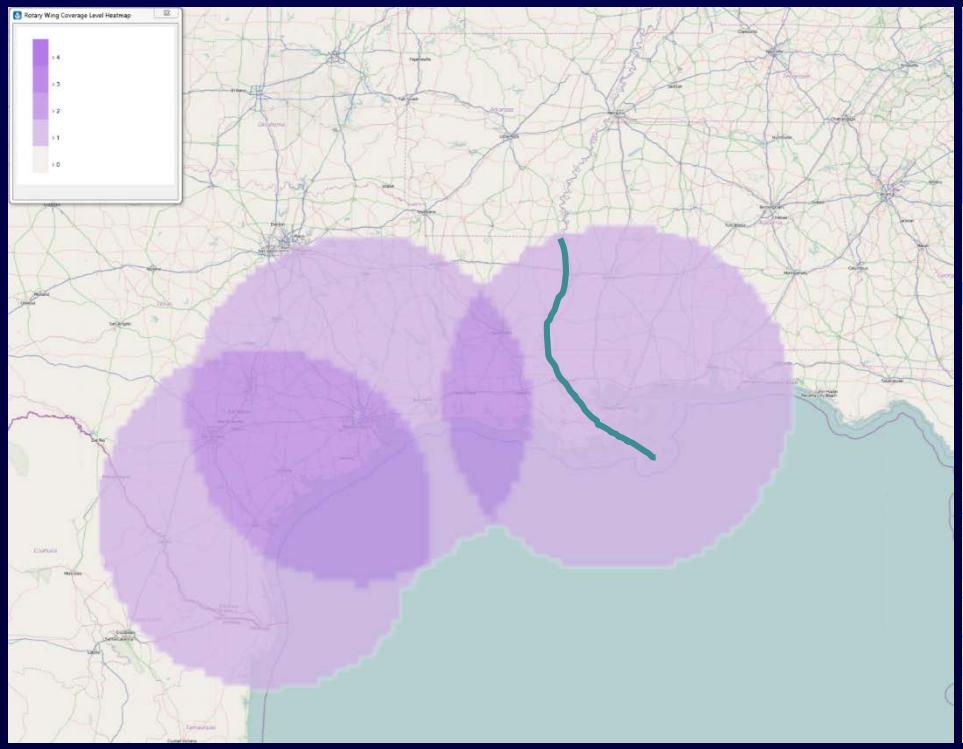
cgSARVA Results – Aircraft coverage

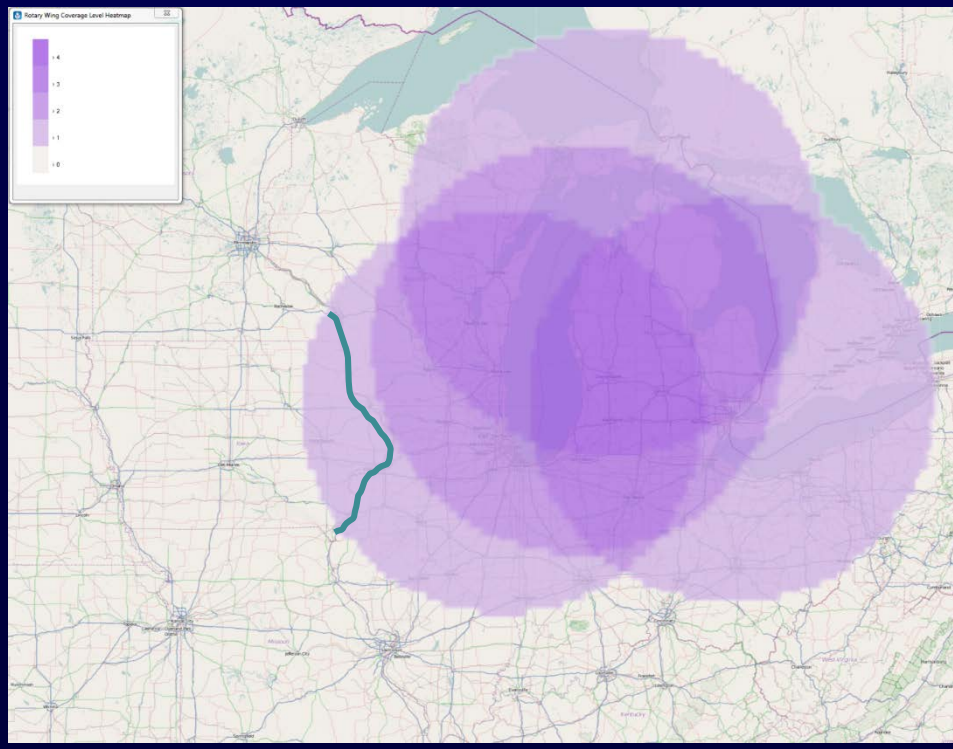


Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.

Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA





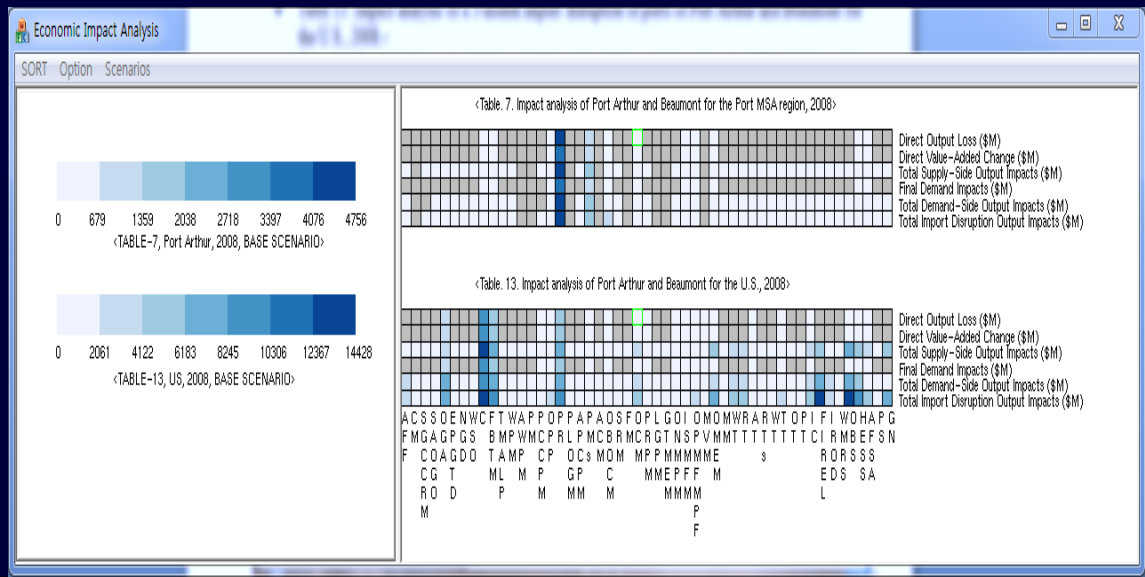
Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



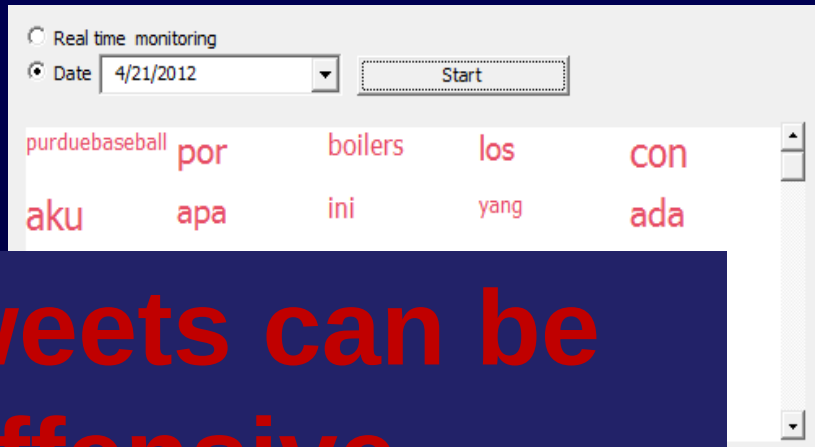
VACCINE

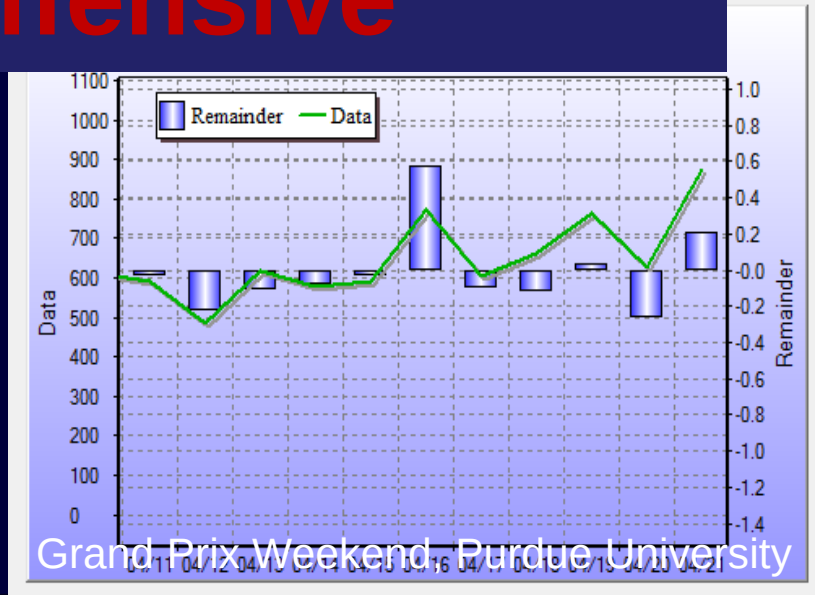
Social Media: Real-time Visual Analytics

(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- T
  - d
- S
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC

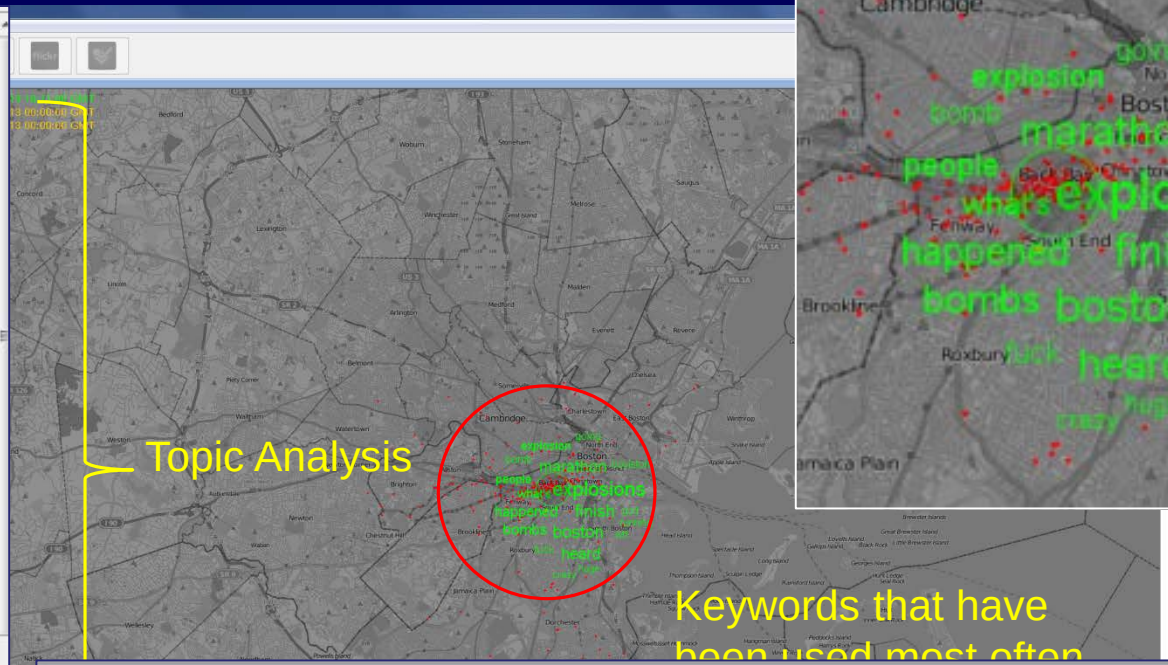
Warning: Many tweets can be explicit and offensive





Explosion Area in Boston





Keywords that have been used most often

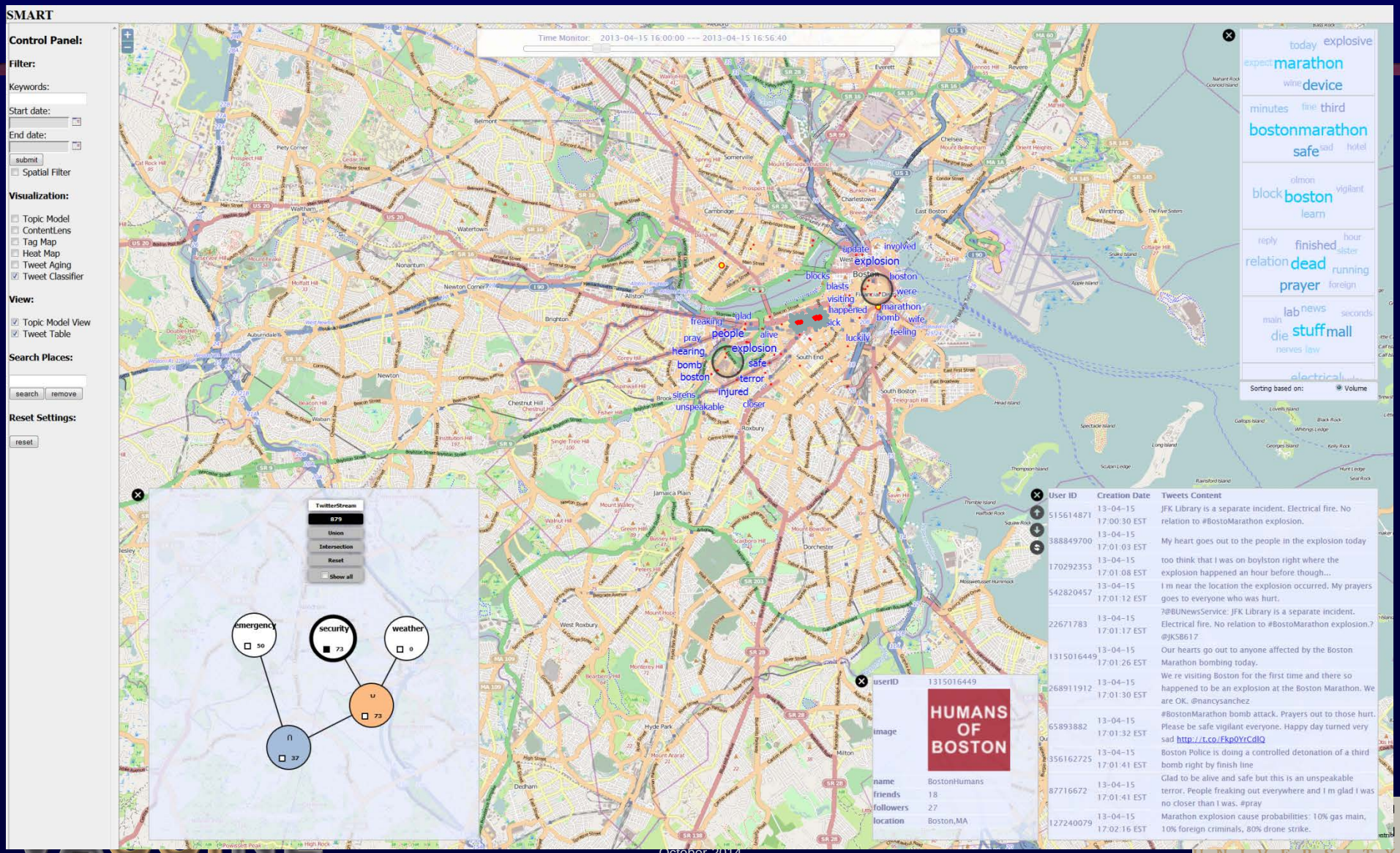
Tweets

Text
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped.
two bombs just went off on boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finsh line this afternoon.
What the FUCK was that
Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalciO I figured 3 people would get the joke
Two explosions just rocked the finish line of the Boston Marathon. Sirens galore. People running in fear. Wonder what happened.
This is crazy i seen that blow up #bostonmarathon
can someone tell me what that explosion was!? #boston #bostonmarathon

First Response (Tweet & Picture)



Social Media Analysis and Reporting Toolkit (SMART)



Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)

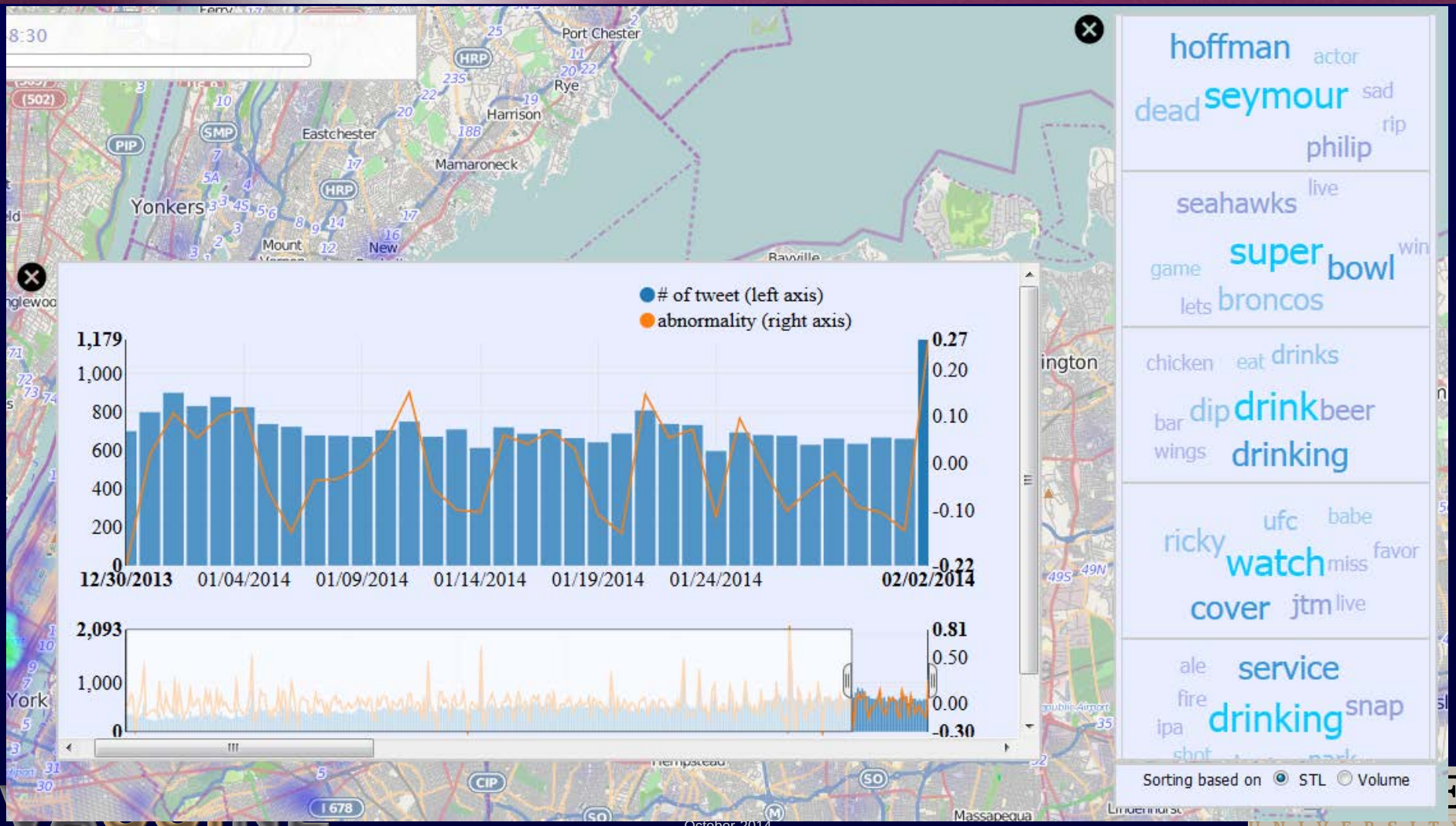


VACCINE

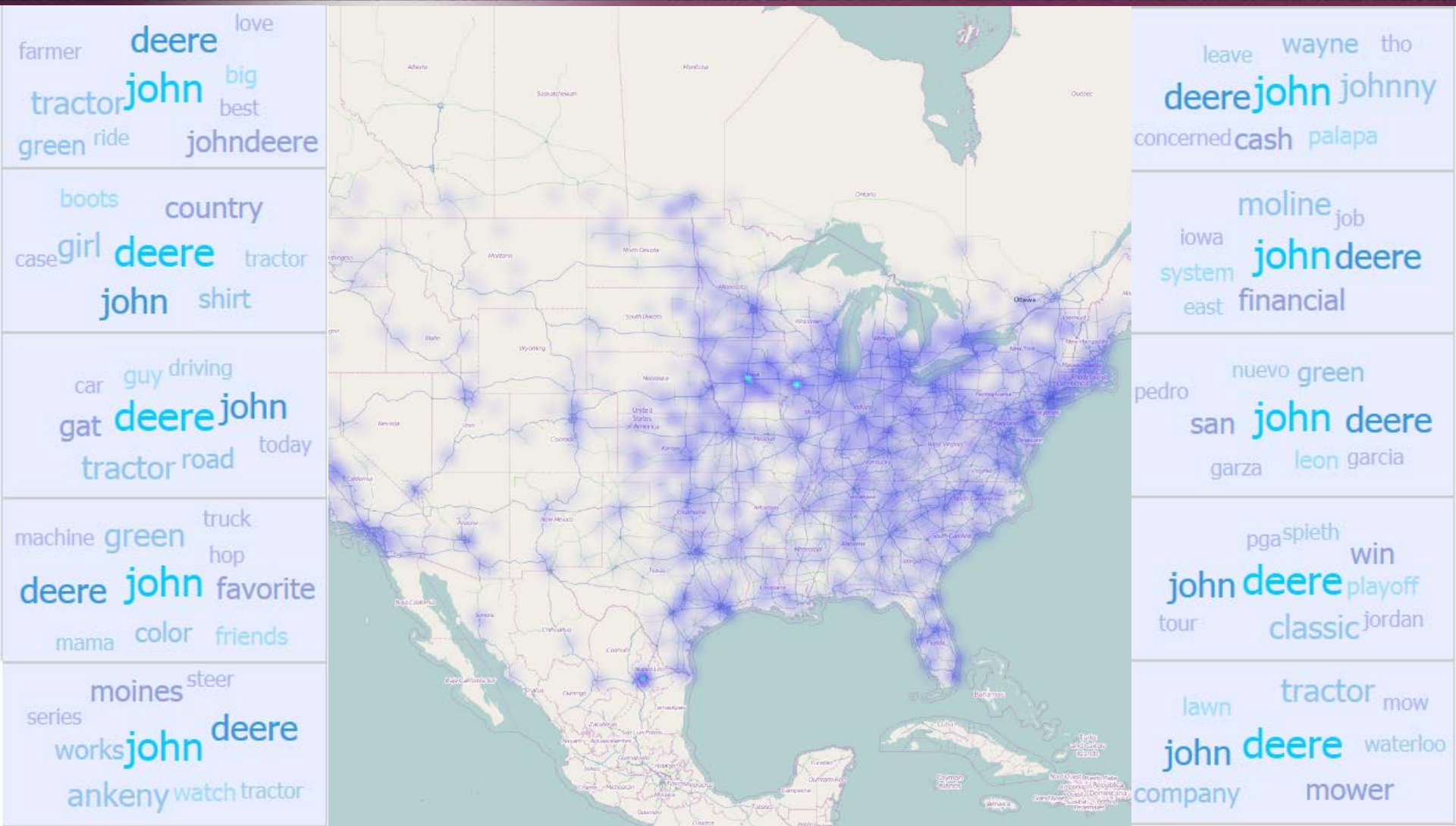
Social Media Analysis and Reporting Toolkit (SMART)

Hurricane Sandy

Anomaly Detection

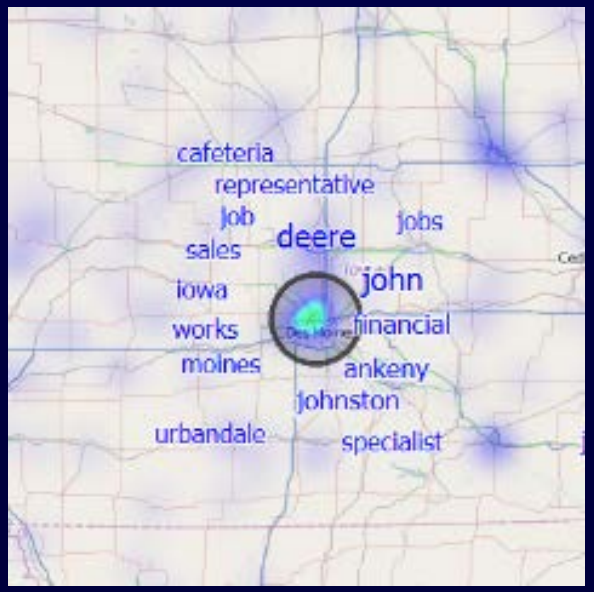


Tweets about John Deere

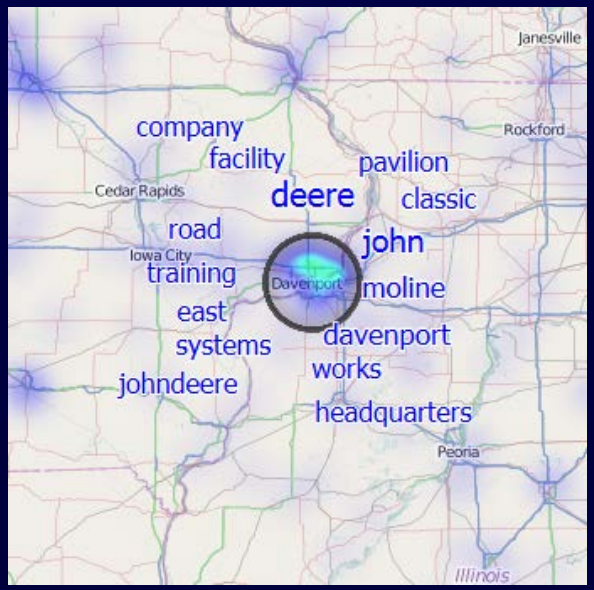


Keywords(Contentlens) in the 3 Hotspots

Des Moines, Iowa



Davenport, Iowa

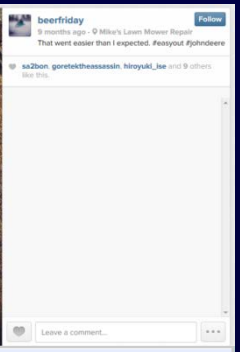


Monterrey, Mexico



Tweets related to ‘break down, shut down, stop working’ and ‘maintenance, repair’





User ID	Creation Date	Tweets Content
630487309	13-06-14 15:55:47 EST	1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of
35458727	13-10-18 22:59:17 EST	@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking

User ID	Creation Date	Tweets Content
933628327	13-06-09 14:27:17 EST	And, for your viewing pleasure, a nicely maintained 1930s John Deere. http://t.co/Zavel5QnDI
50371494	13-06-07 10:13:50 EST	The Class of 2013 presents its class gift at commencement – a John Deere Gator for the Maintenance Department. http://t.co/ju1rchEuLm
494325743	13-08-16 15:07:21 EST	Rondorf yard maintenance! #chevy #johndeere http://t.co/HyhJy5pTNR
404120195	13-09-03 16:13:01 EST	Just saw a clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #Townielife
1038577016	13-07-28 22:58:57 EST	#labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenan... http://t.co/azEBmrPMt5
502127999	13-08-07 18:47:22 EST	That went easier than I expected. #easyout #johndeere @ Mike s Lawn Mower Repair http://t.co/8pgiu0Uv7g

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

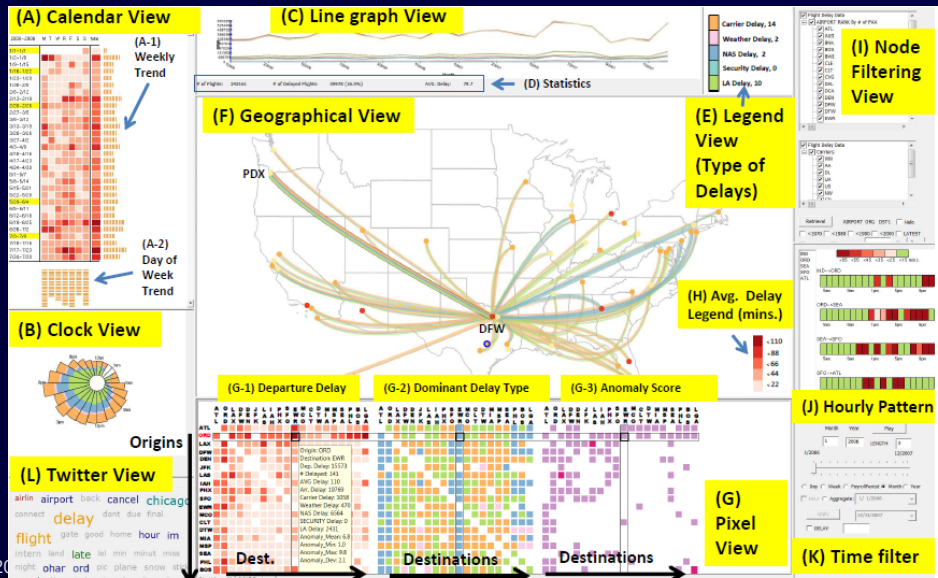
Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

January 2015

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects



VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of California, San Diego**
- **University Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **Oak Ridge National Laboratories**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

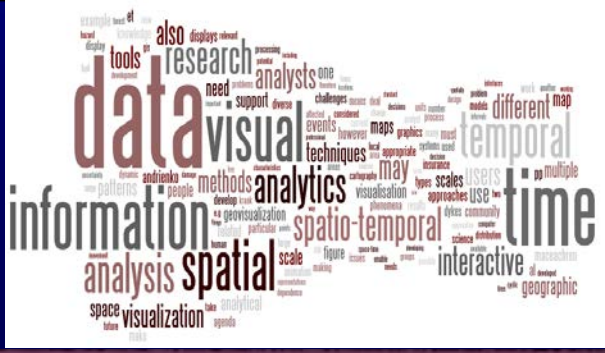
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America



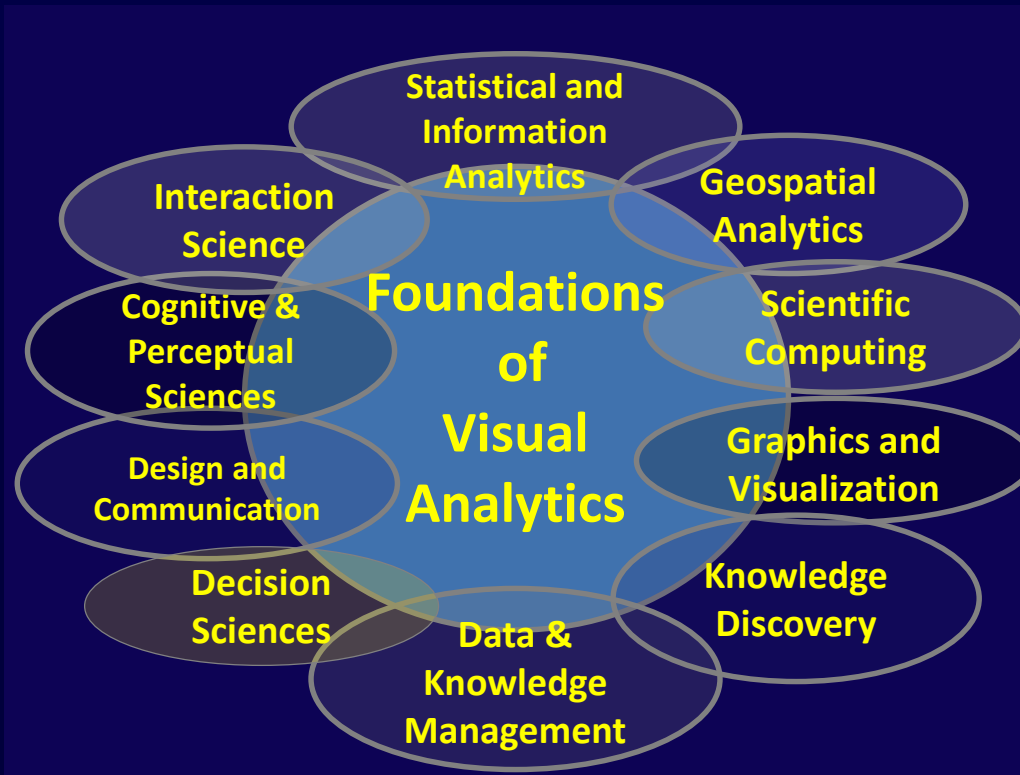


What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

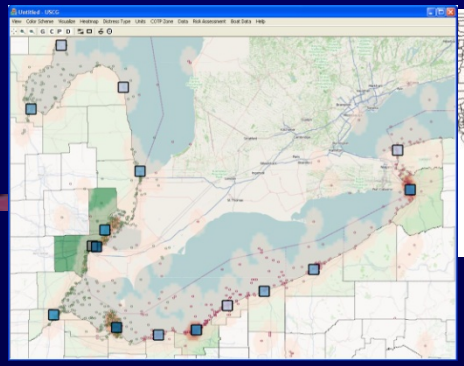
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

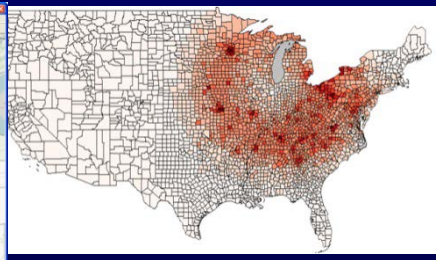
VADM Robert Parker, 2012 MRS Keynote Address

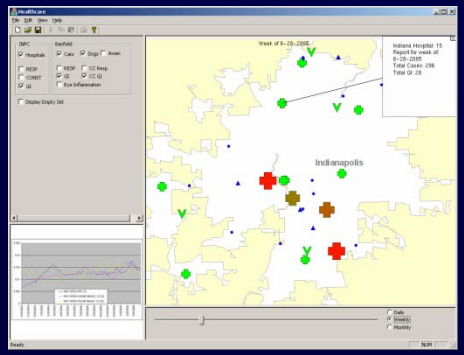


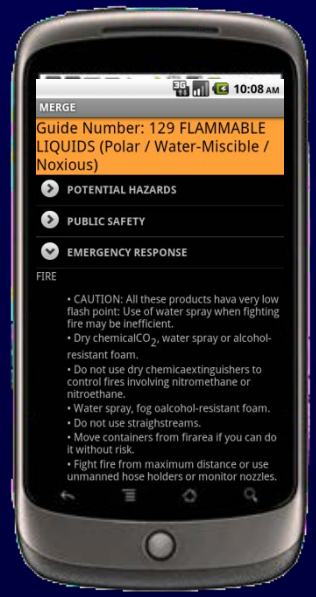
Public Safety & Health

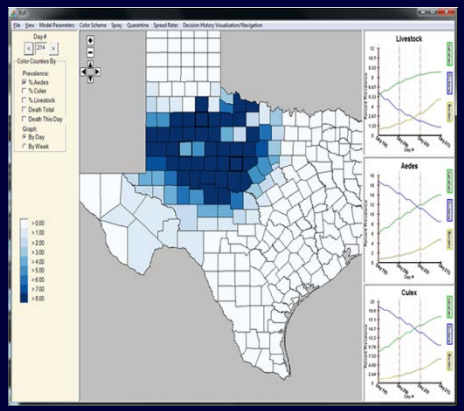
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering









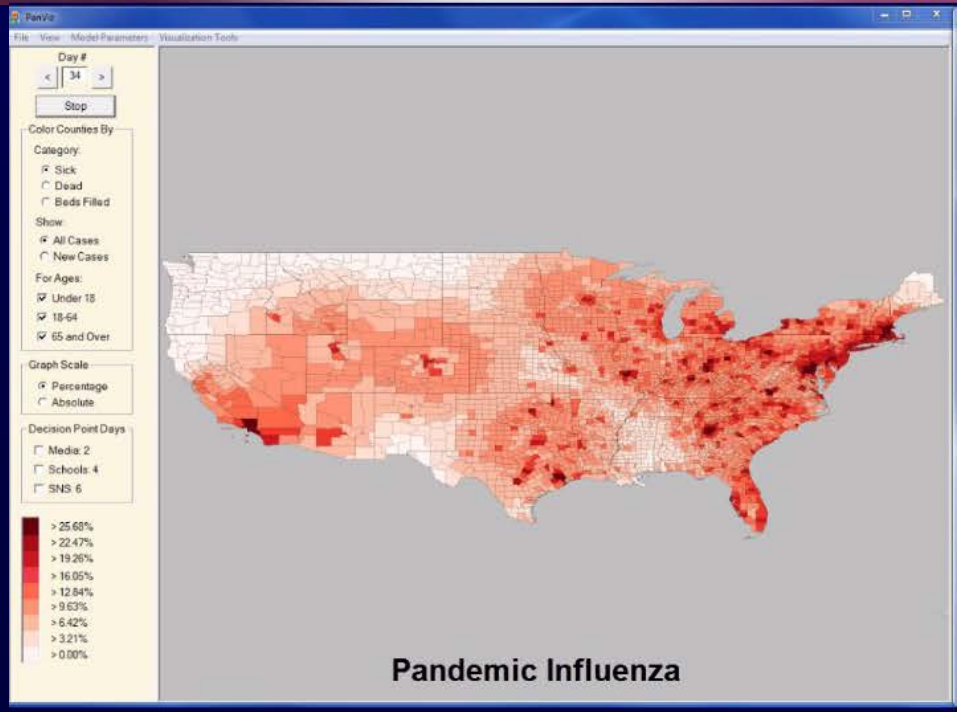


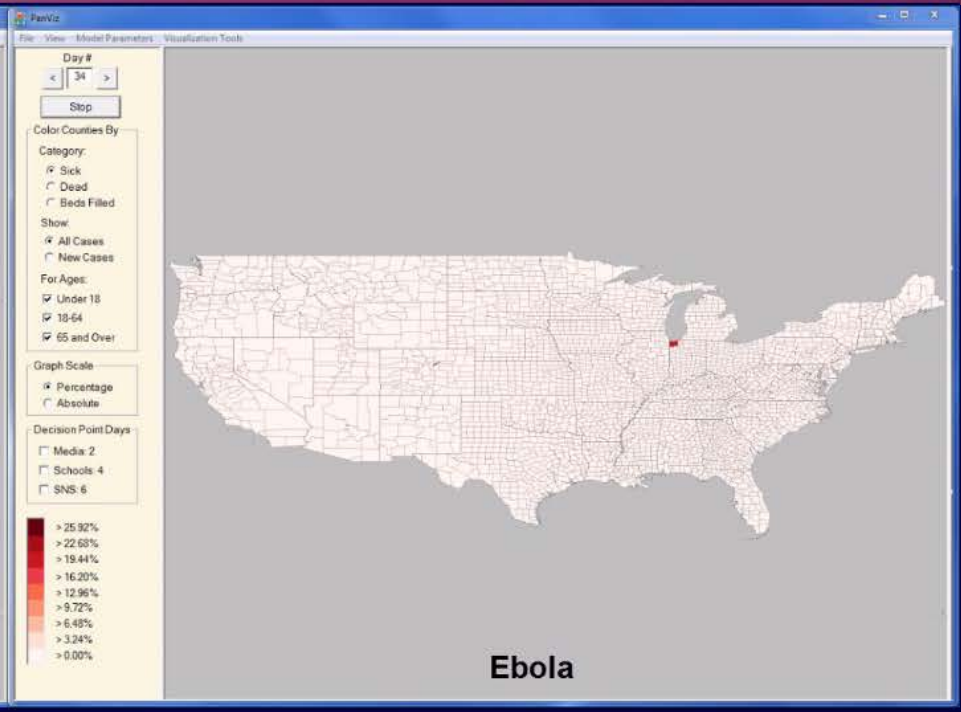


VACCINE

Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL



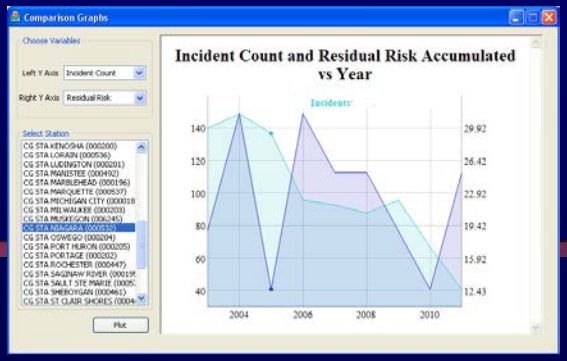


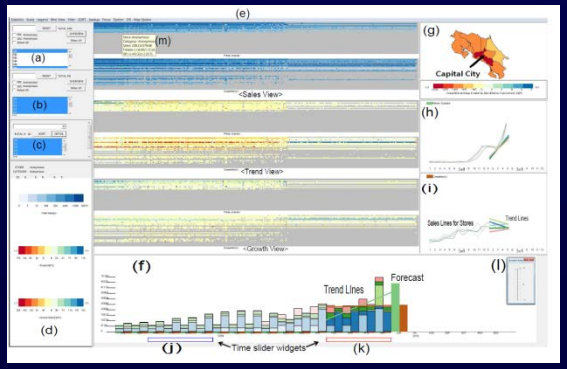
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

VACCINE

Business Visual Analytics

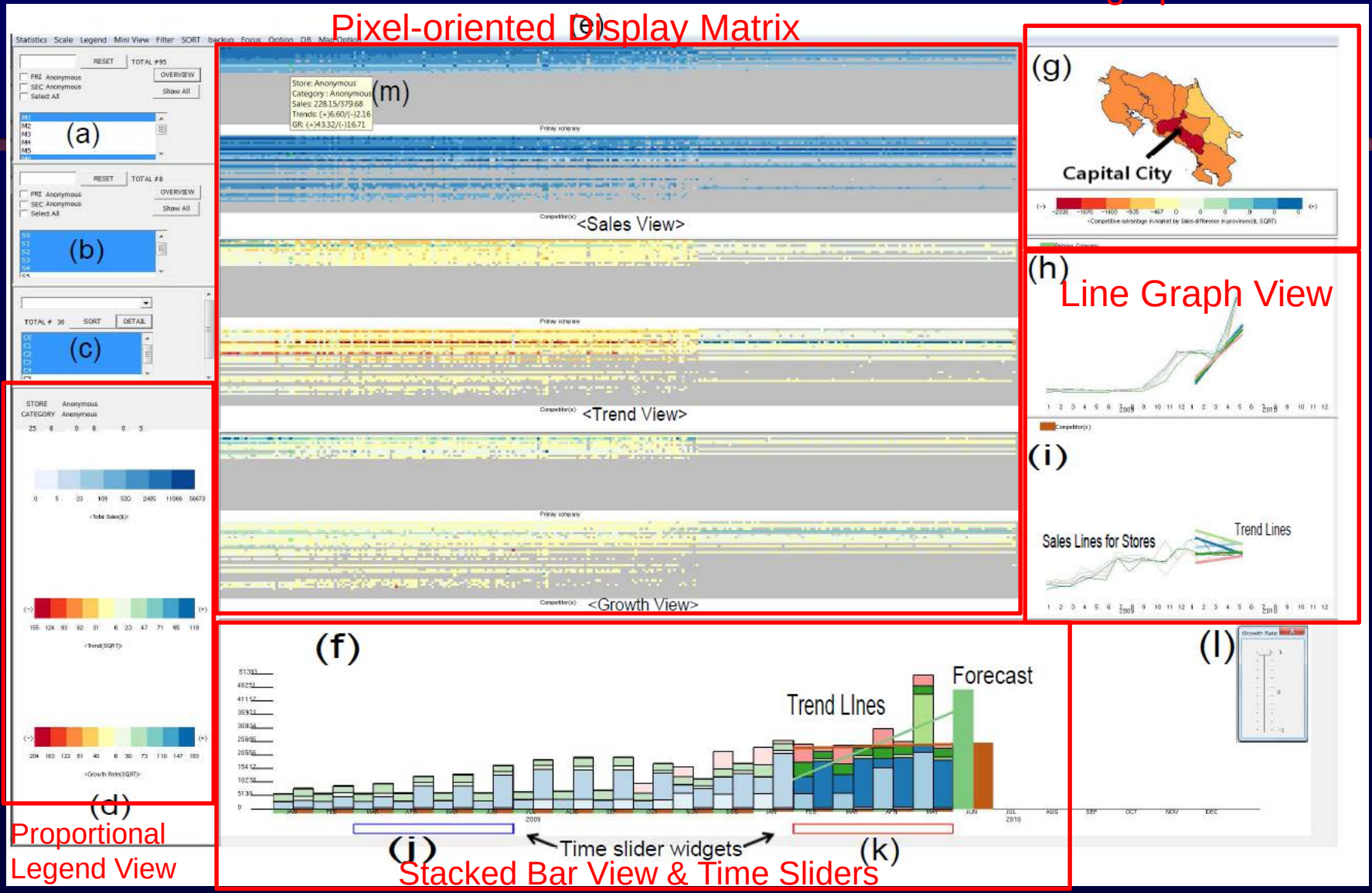
- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







Market Analyzer

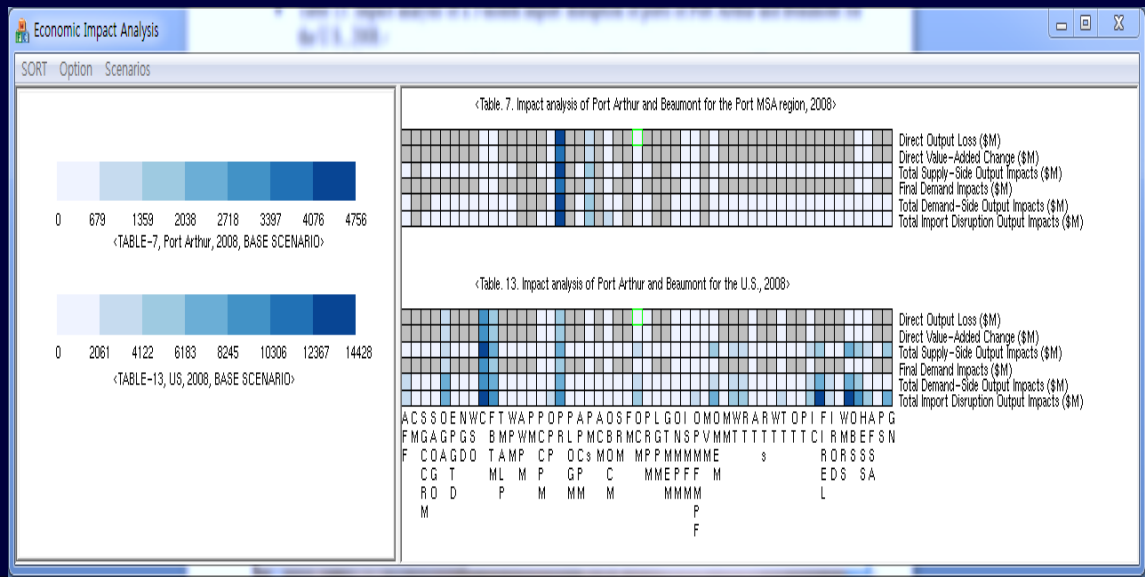


USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies

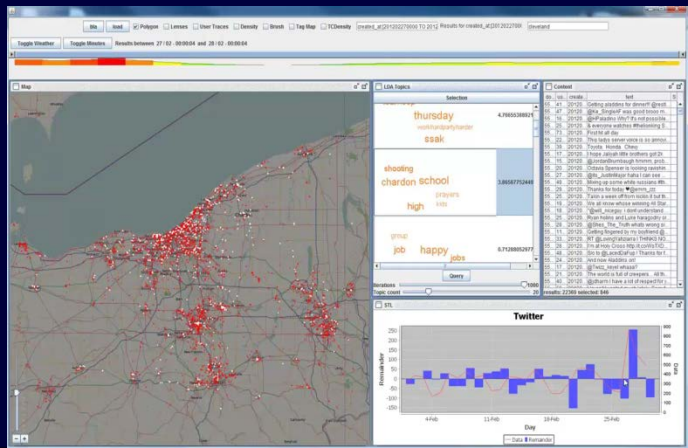


VACCINE

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Our Spatiotemporal Visual Analytics Work Contributions

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

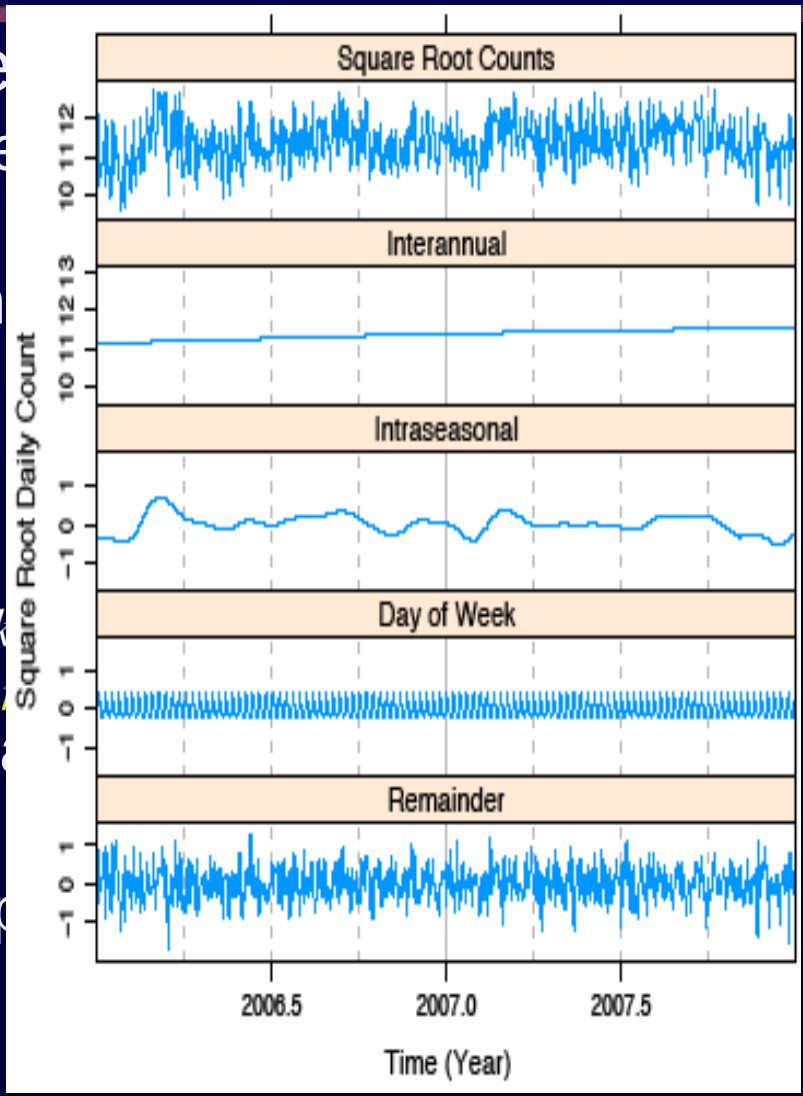
Event/Anomaly Detection - STL

- Temporal trend analysis and pre
  - Seasonal Trend decomposition base
    weighted regression) (STL)
  - Time series data viewed as the sum

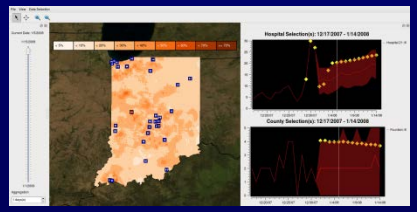
$$Y = T$$

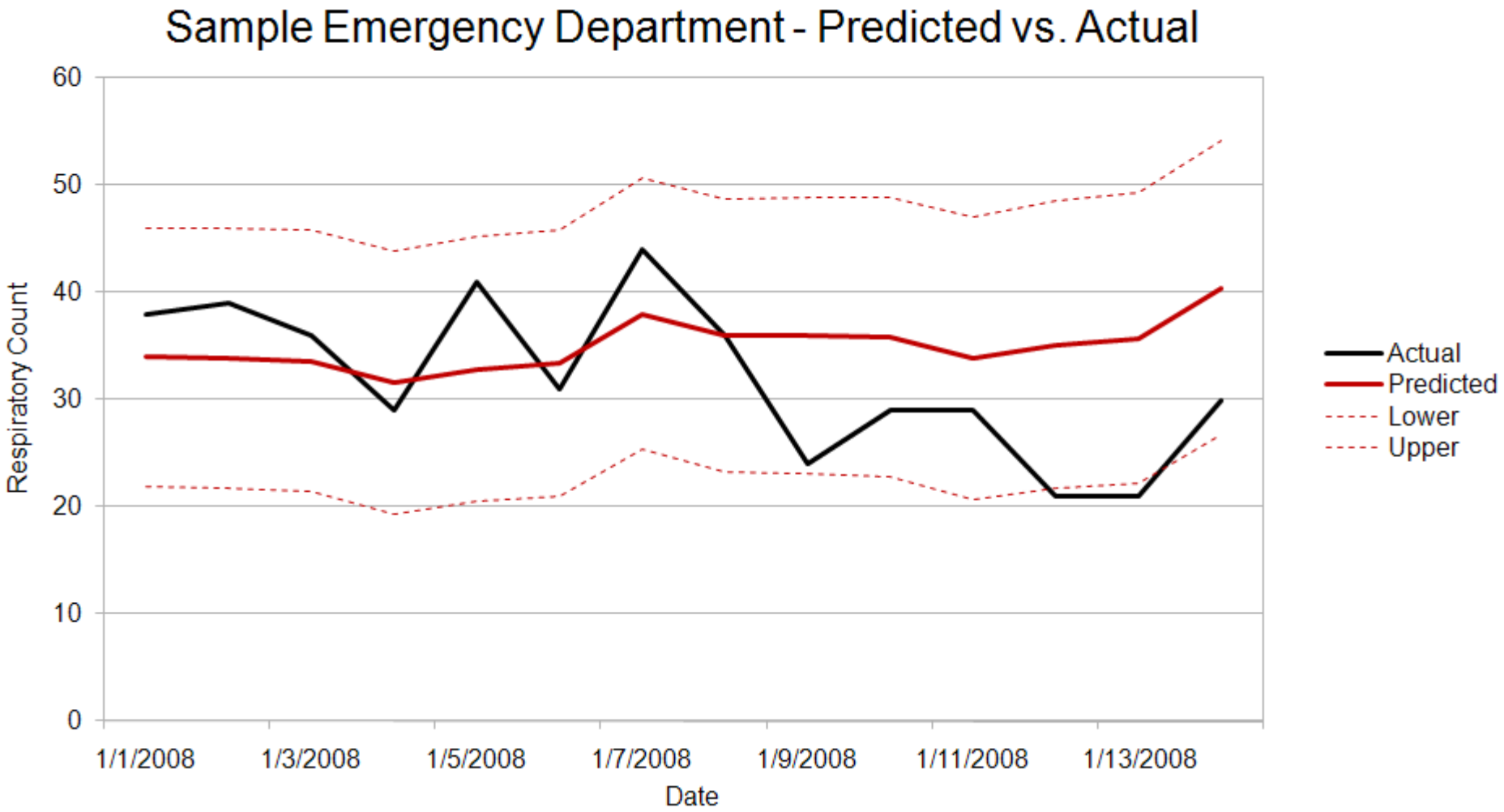
Y : *original time series*
T : *trend component*
S : *seasonal (daily/week*
R : *remainder compone*

  - Large value of *R* indicates substanti
  - Filter out the noise component
  - If value varies by 2 SD from STL exp



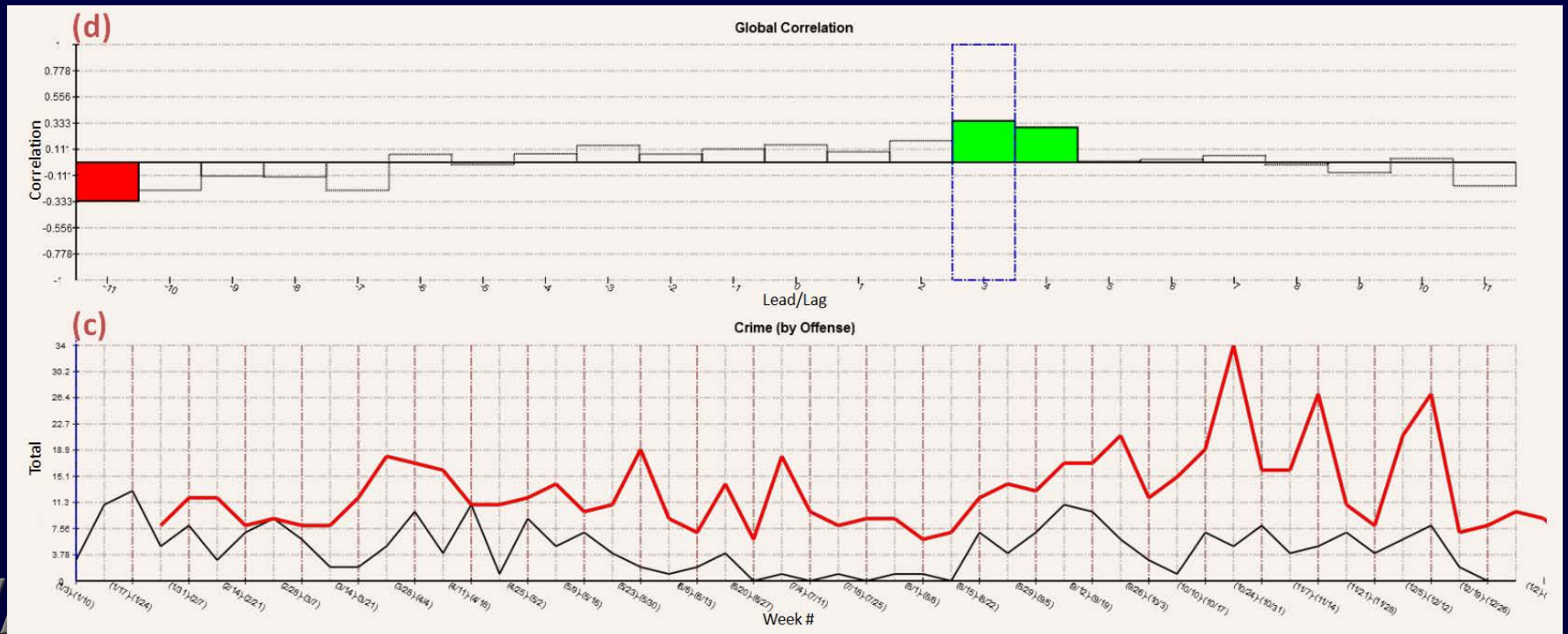
Predictive Visual Analytics



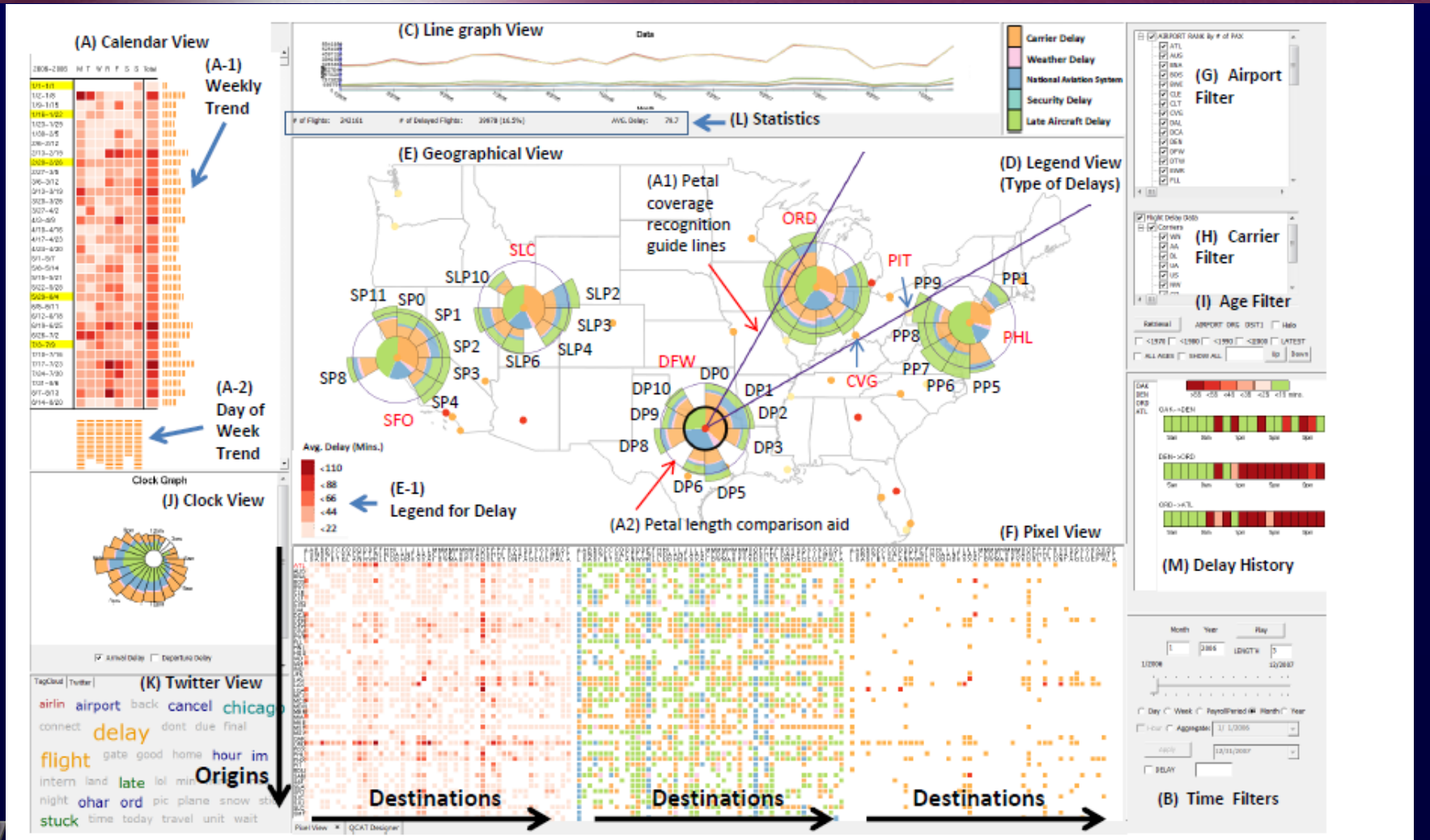


Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space



Multivariate Network Analytics with Information Theory Anomaly Detection



The Scale Problem

- Data scale:
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues
- User / problem scale:
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally

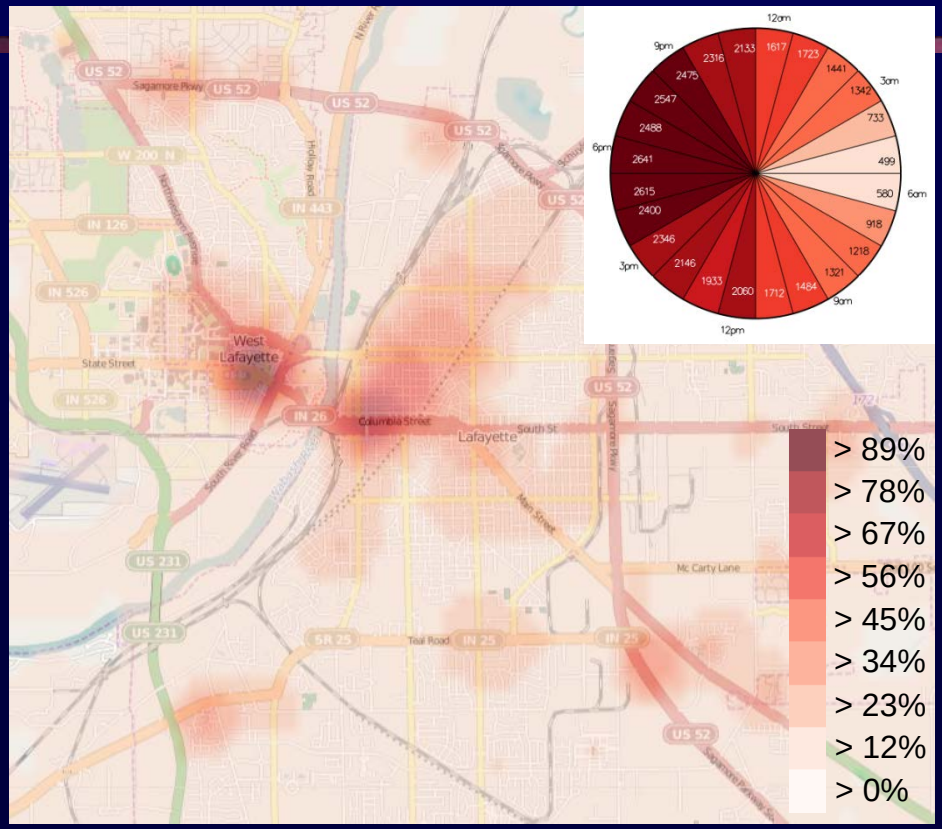
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness[1]
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: **Empower the casual expert to be more effective**

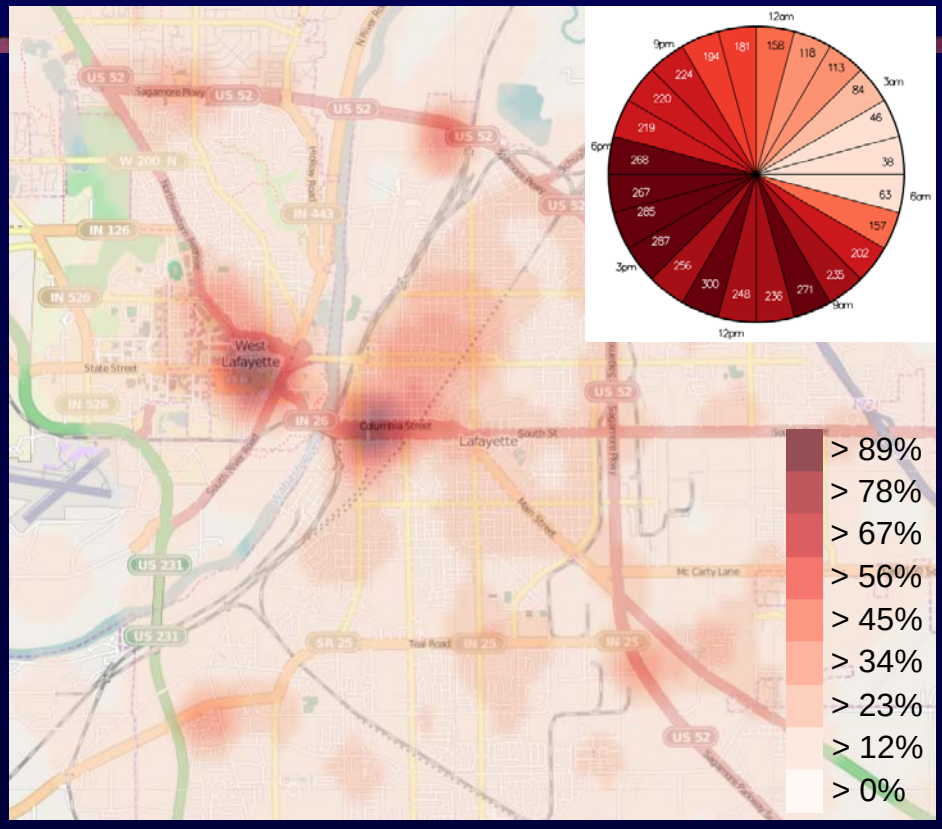
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

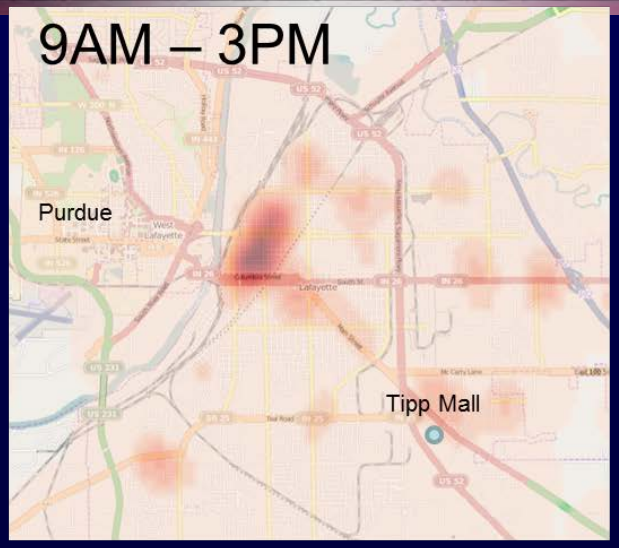


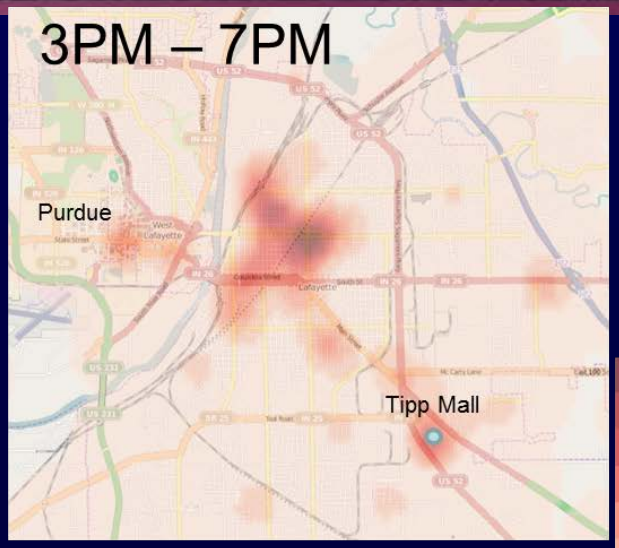
Past 2 Years
3/11/2012 – 3/10/2014

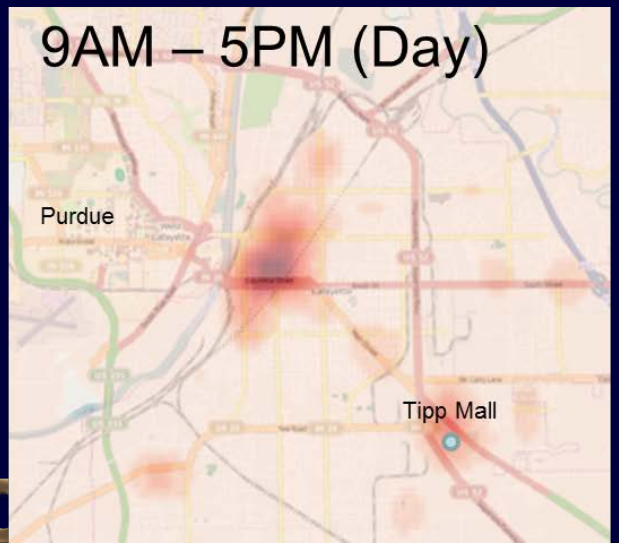


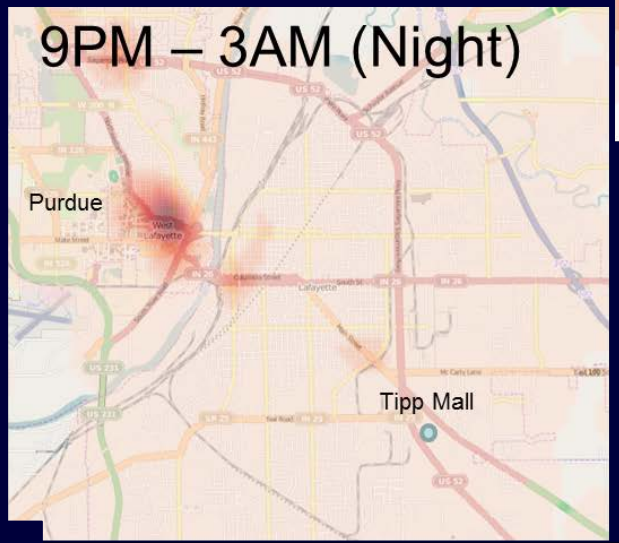
March + Tuesday Filter
(Past 10 Years)

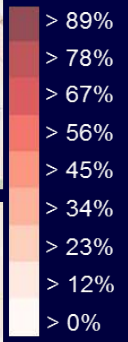
Prediction During Certain Hours In The Day





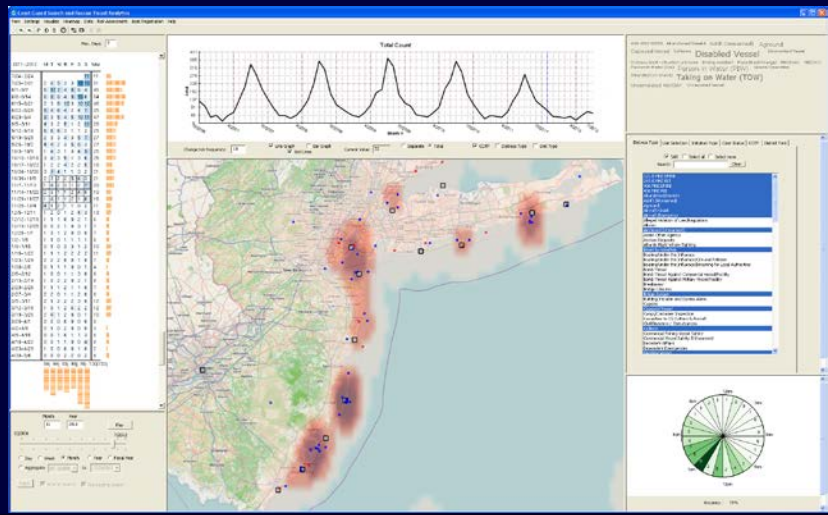




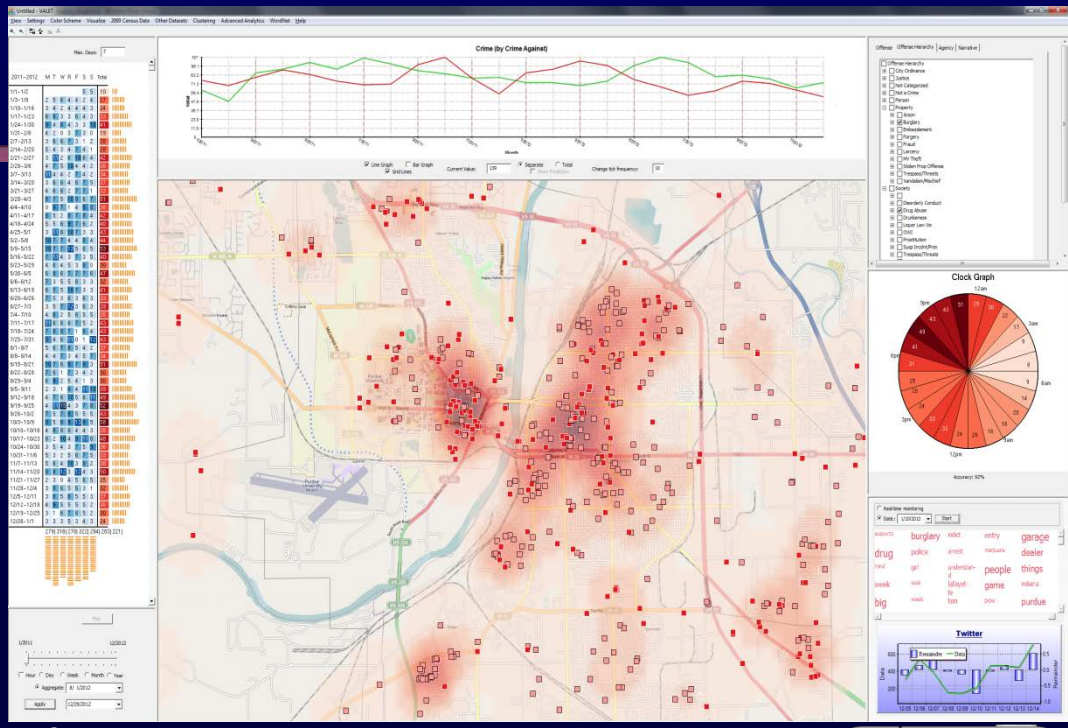


Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

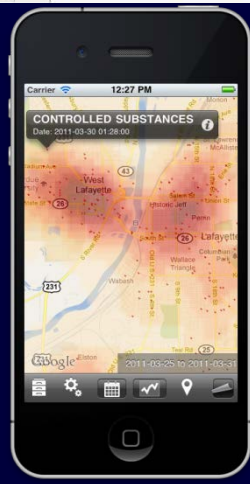
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

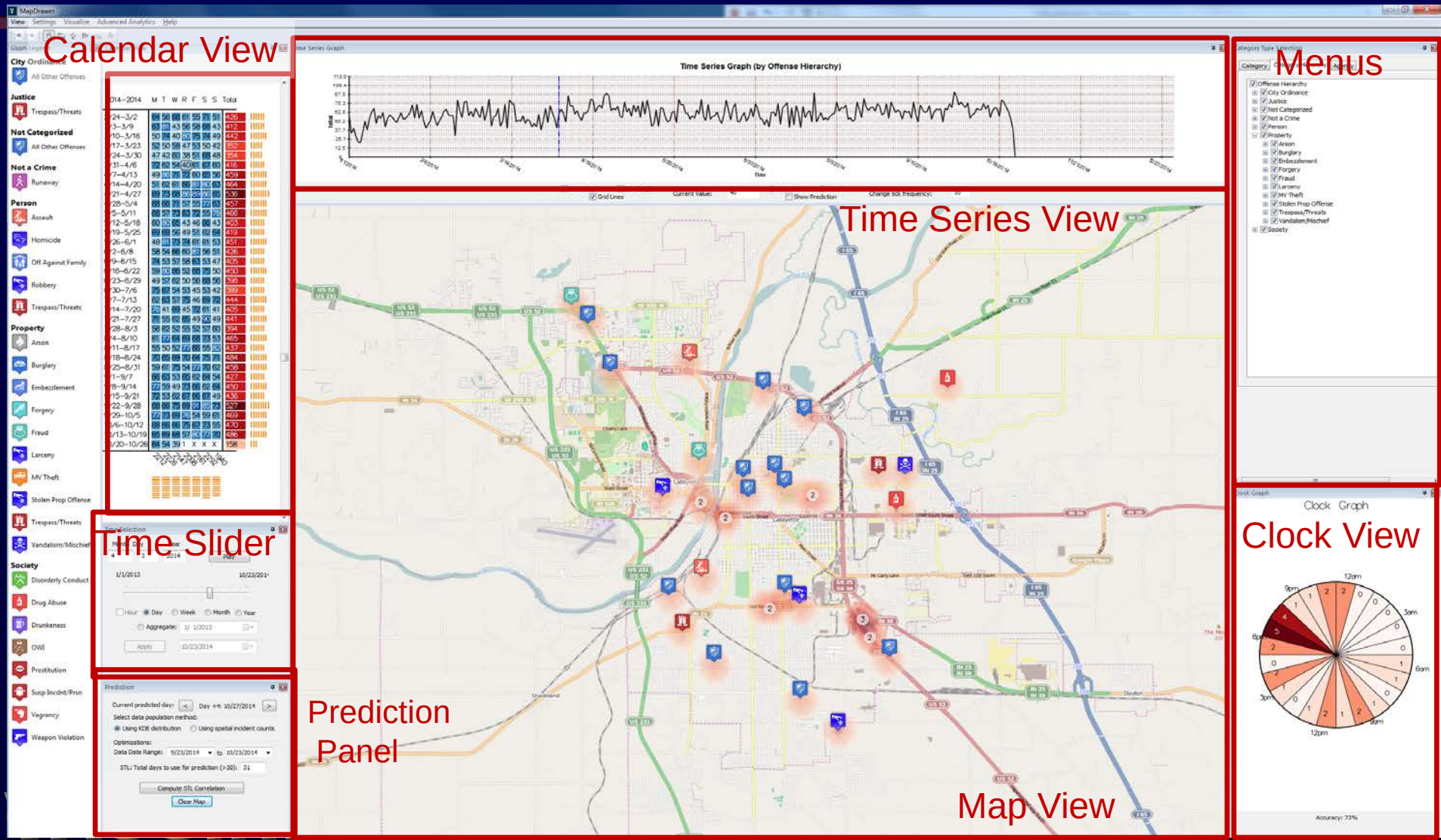
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

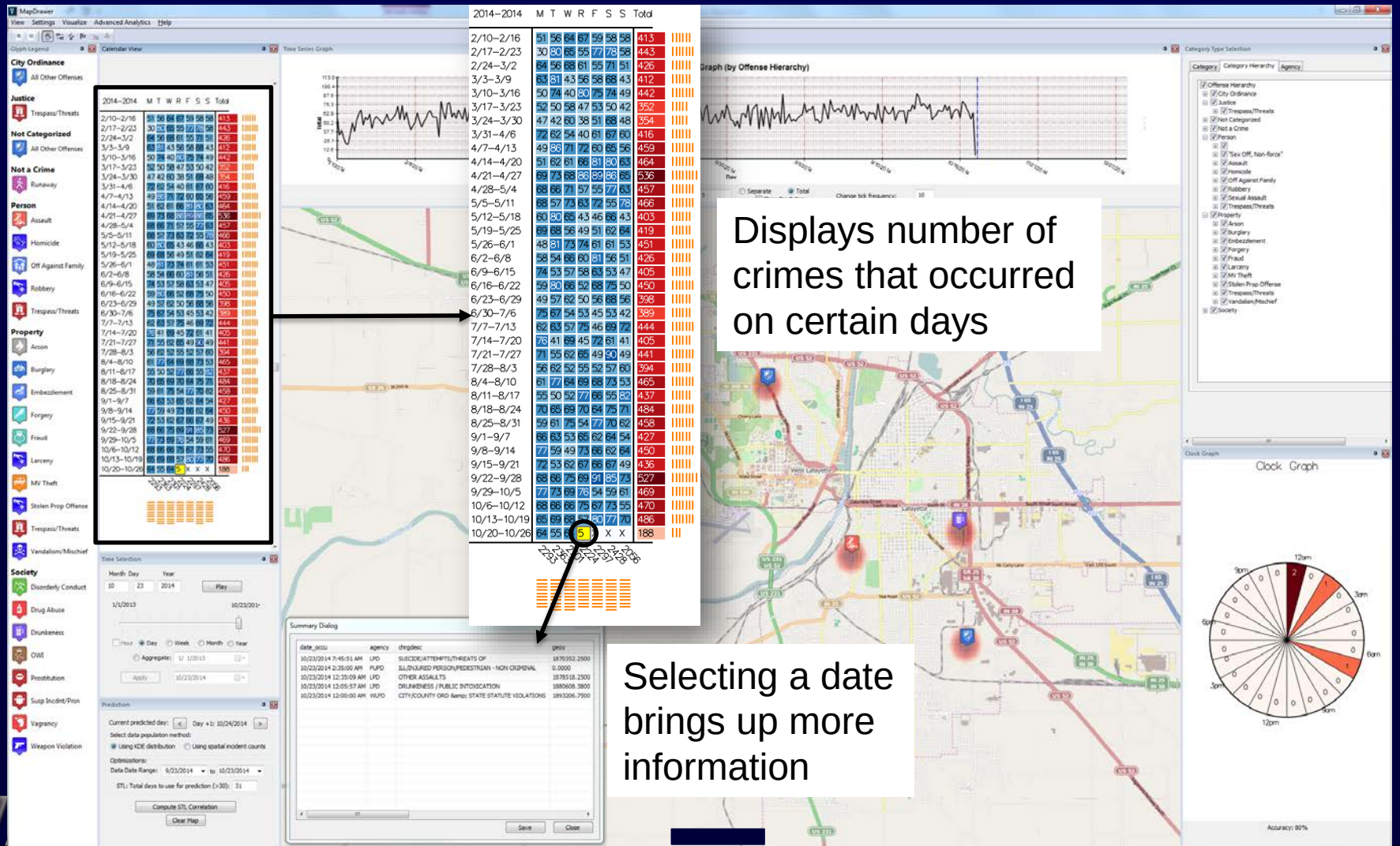


VACCINE

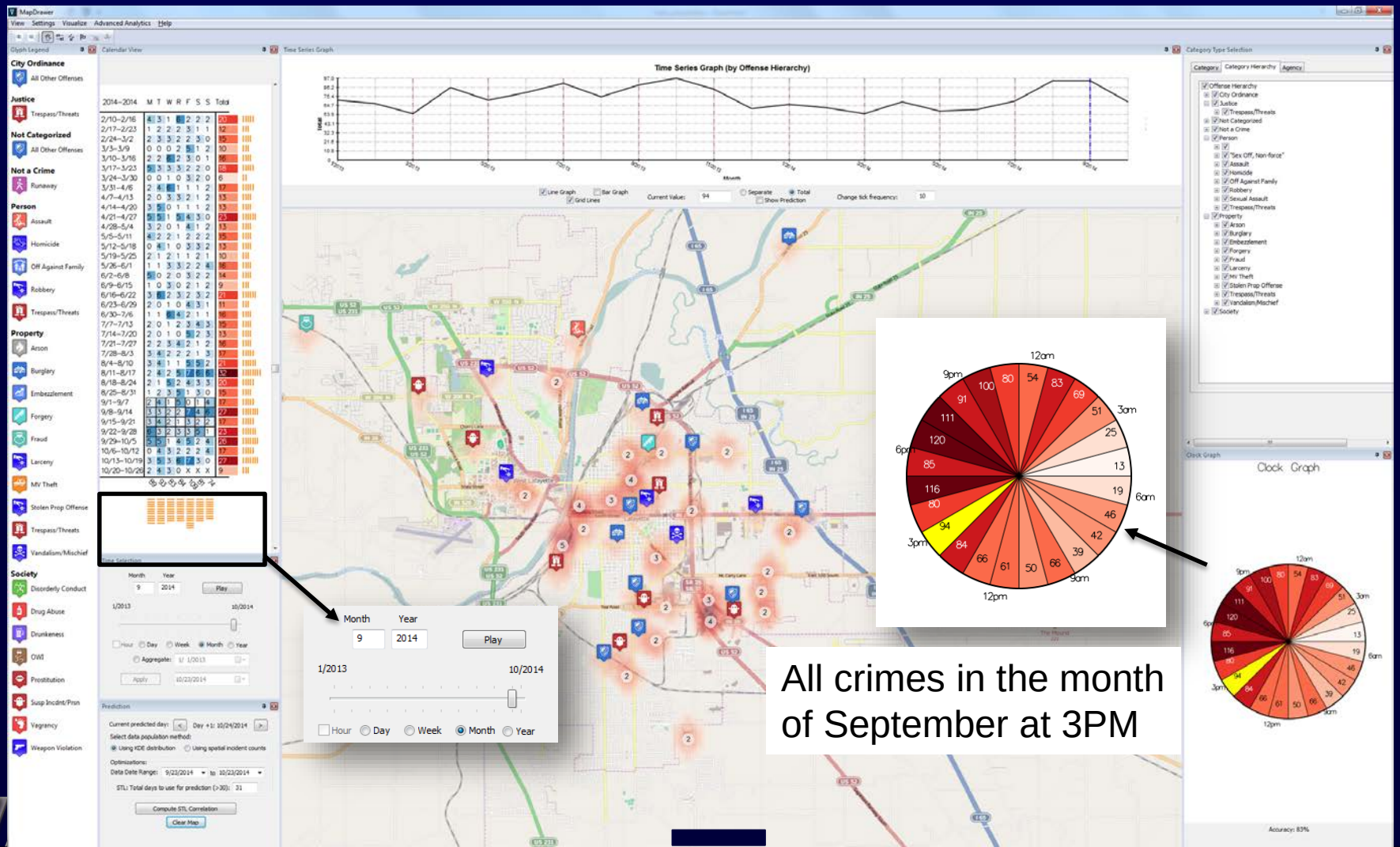
VALET Layout



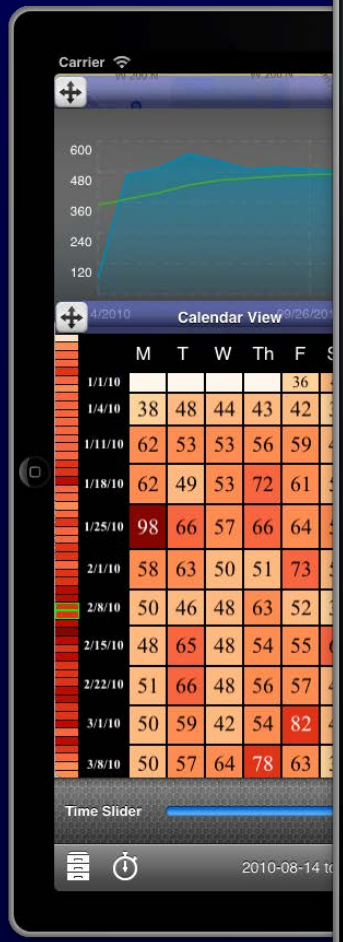
Calendar Display

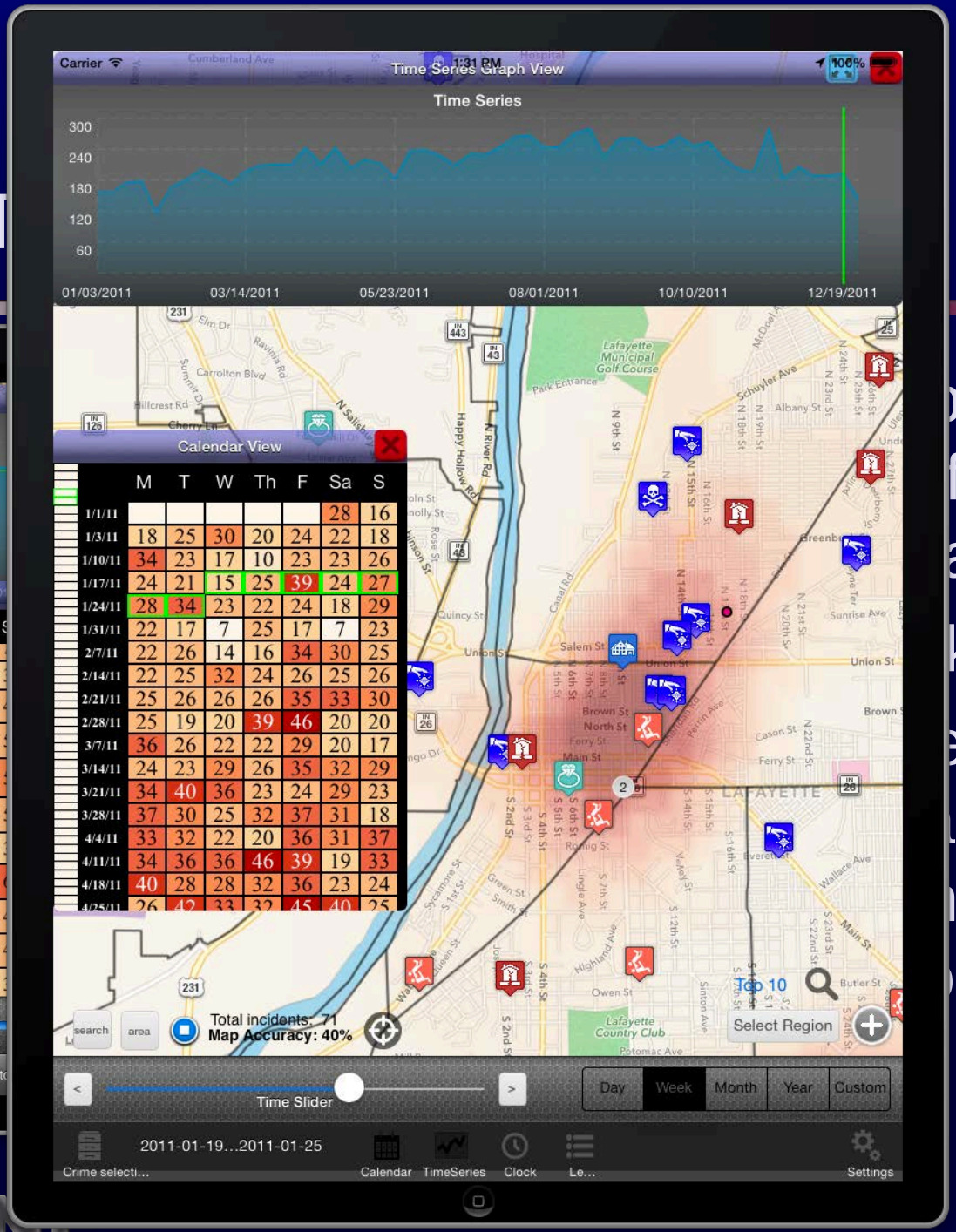


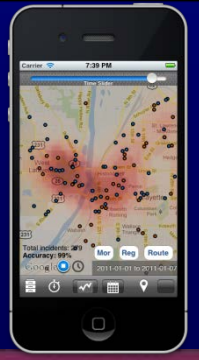
Crimes During Certain Hours In A Month

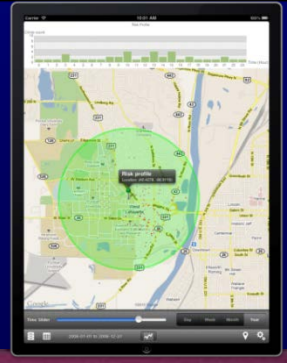


iVALET









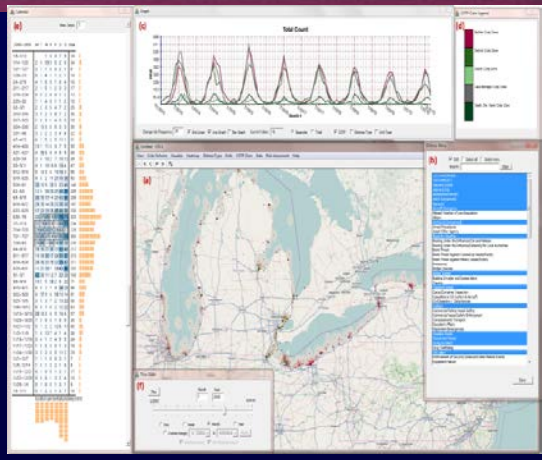
olore criminal,
ffic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis



VACCINE

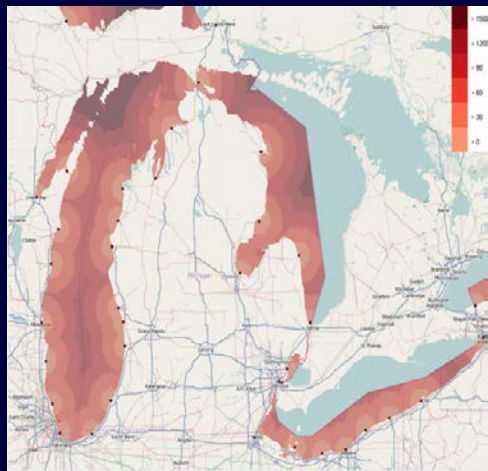
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

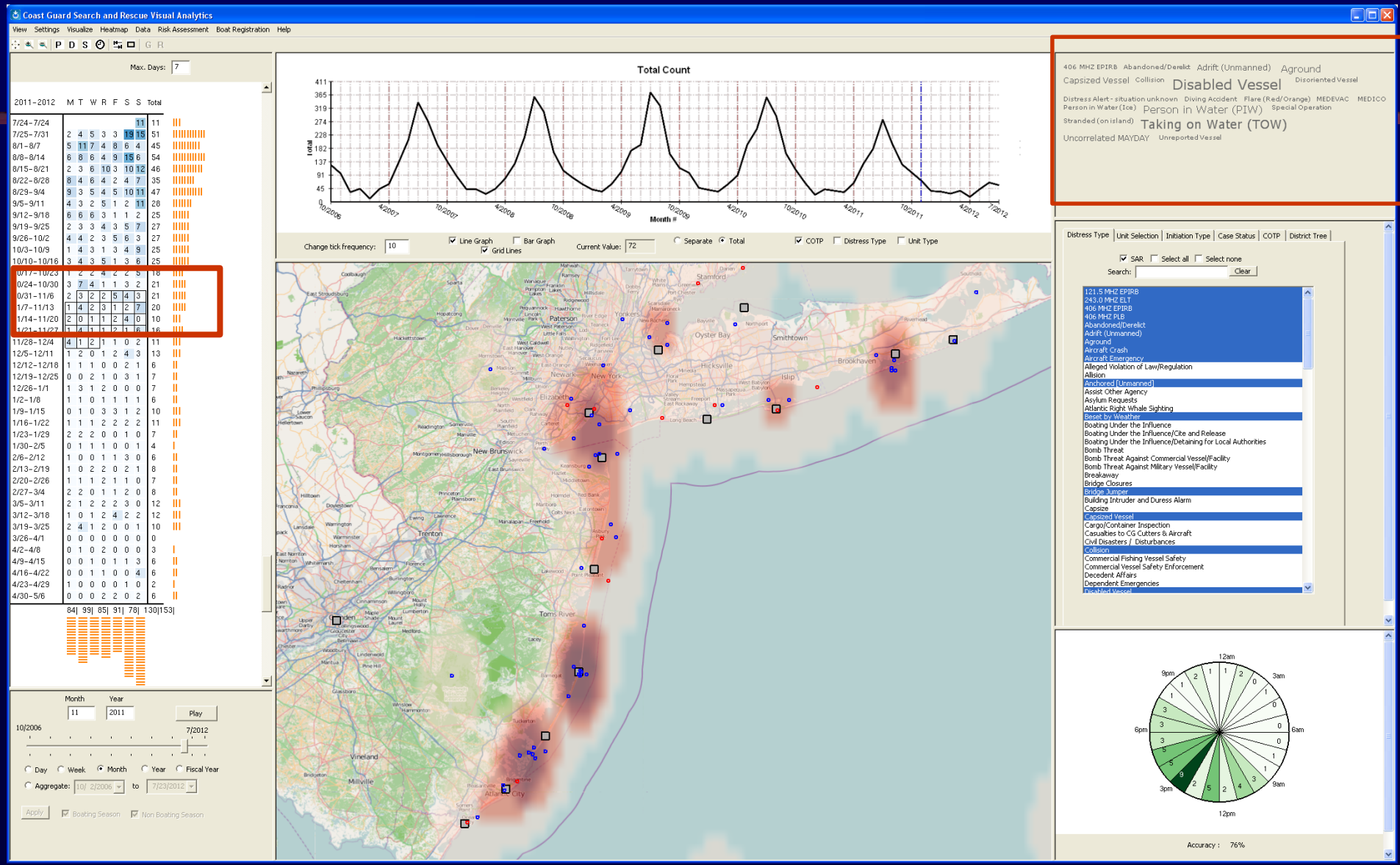
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



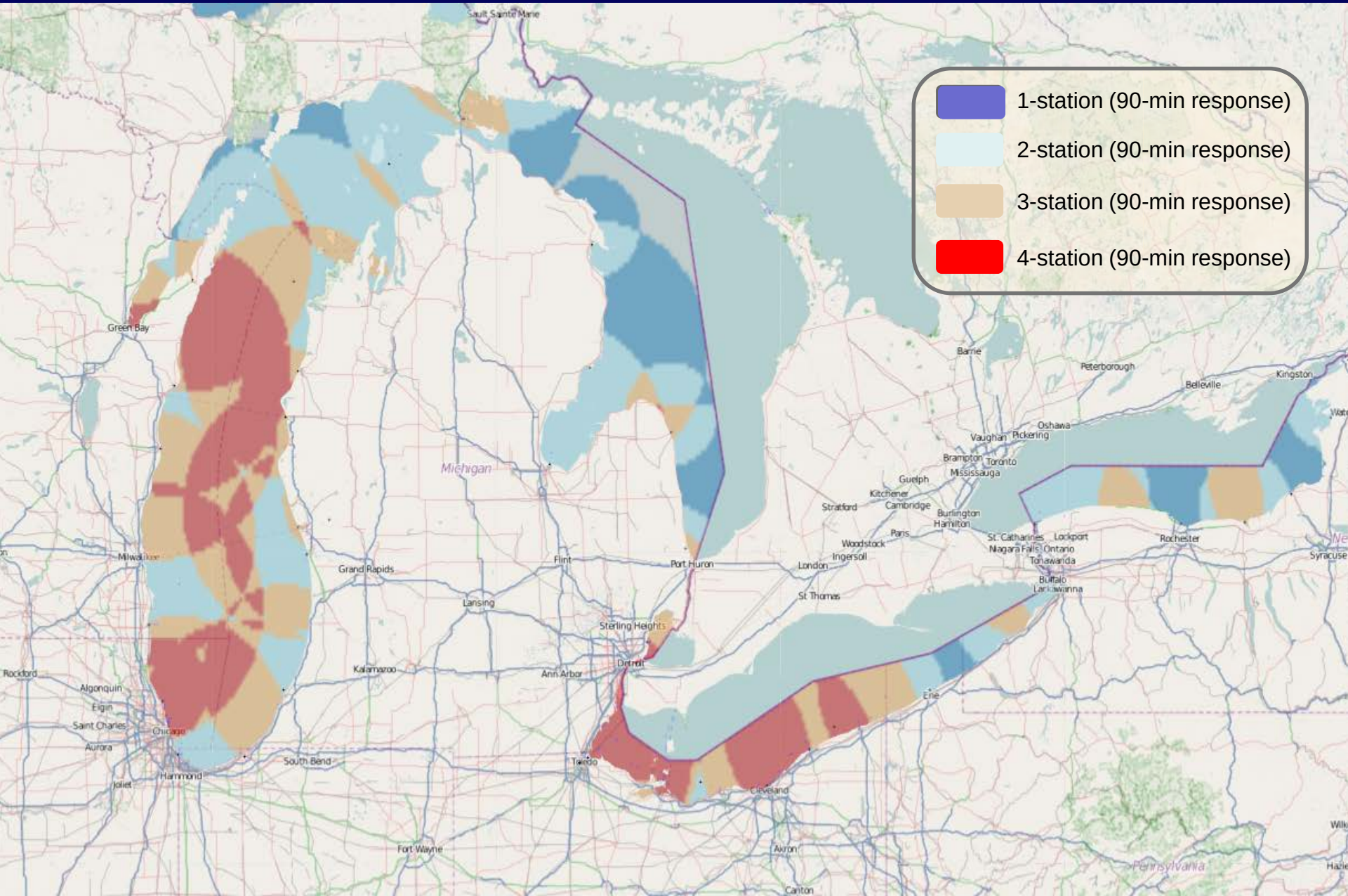
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

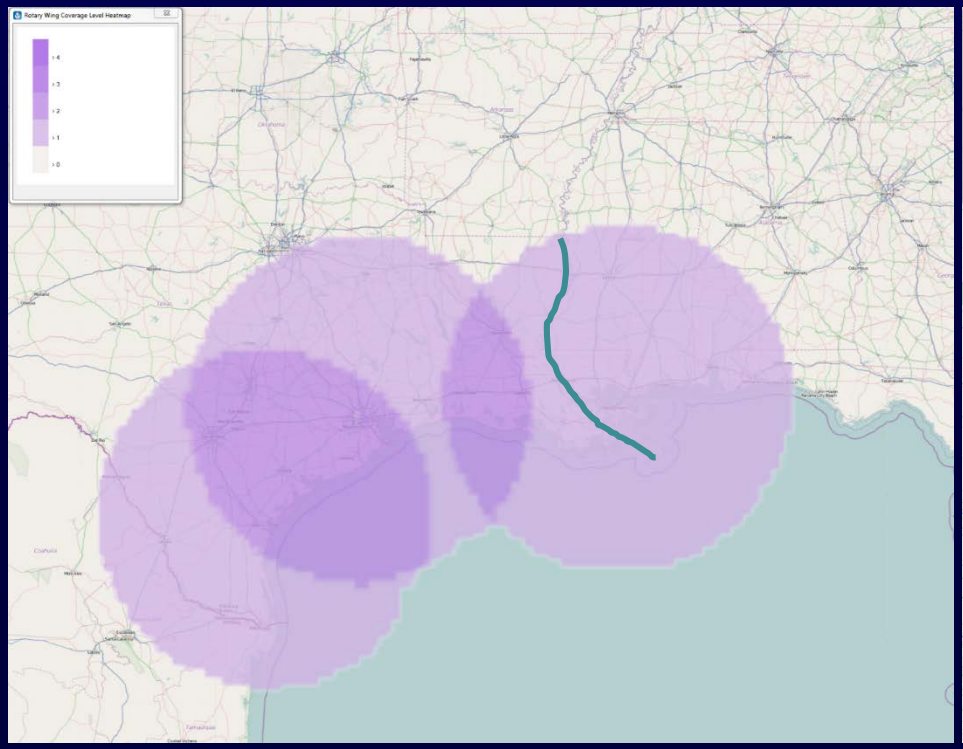


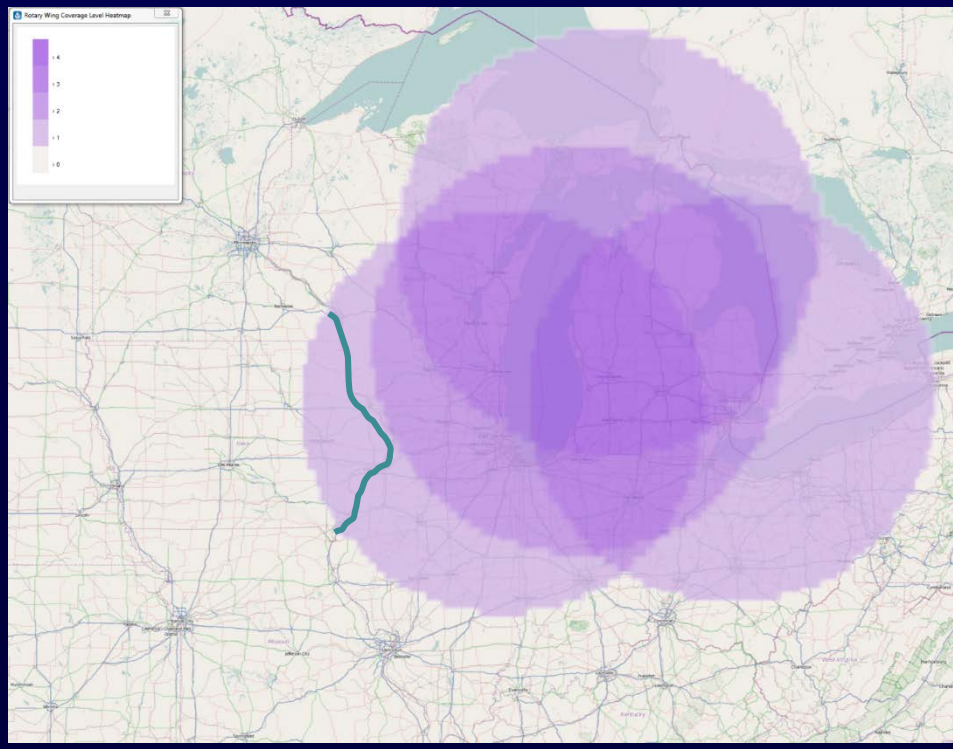
Response Efficiency – Potential Future Assets



Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

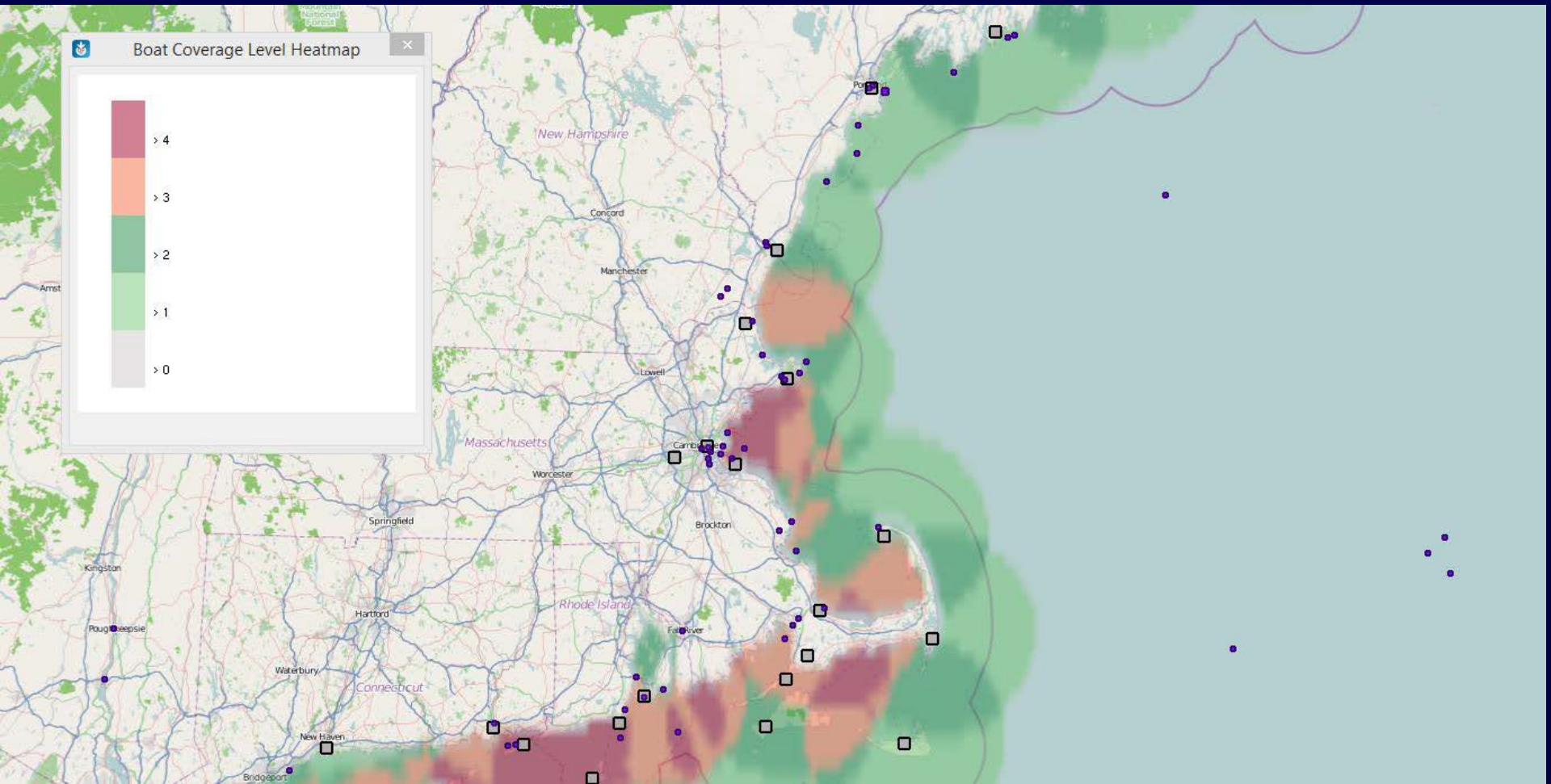




Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

VACCINE

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



Interactive Operational Performance Assessment Report (iOPAR)

- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels

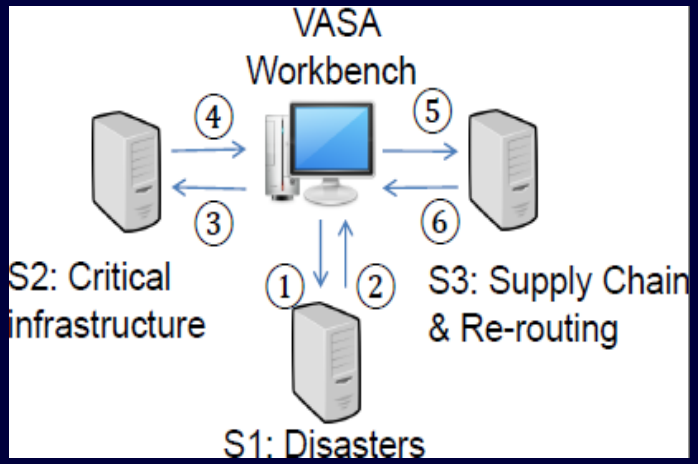
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC

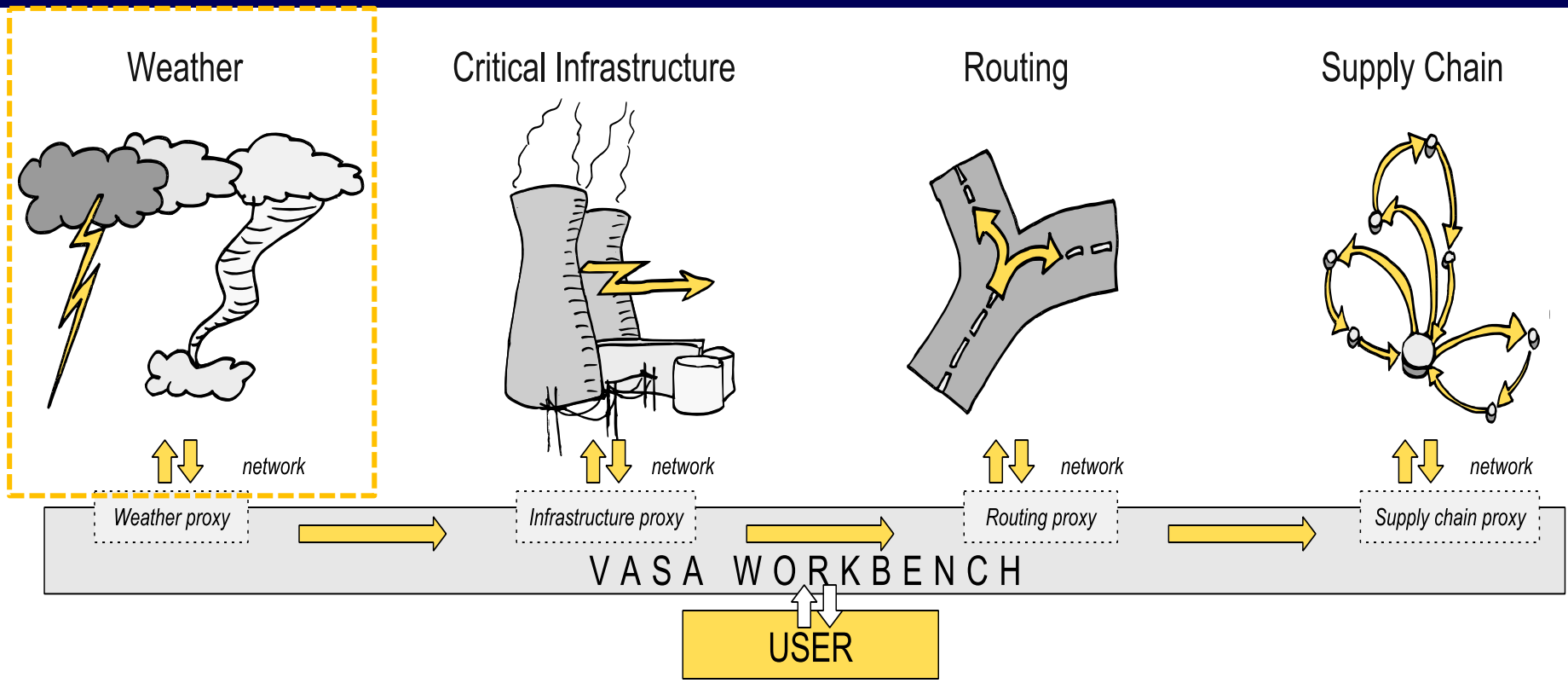


U Minn

TACC

VACCINE

VASA COMPONENT:
CRITICAL INFRASTRUCTURE

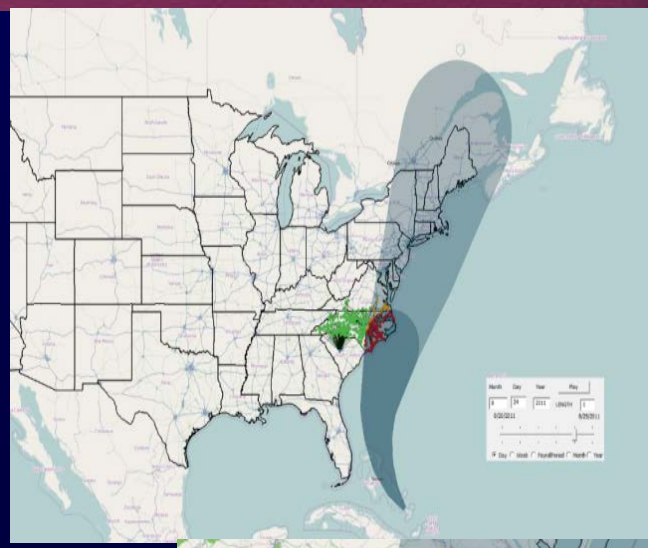


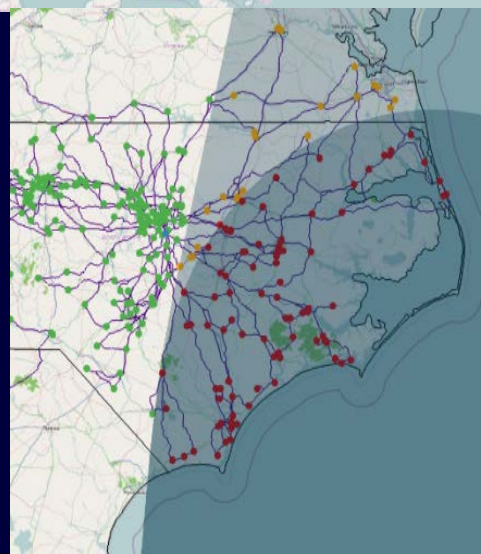
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models

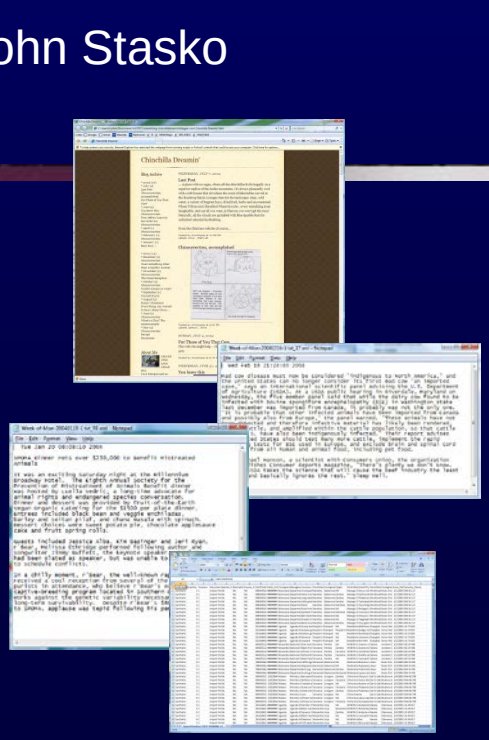




VACCINE

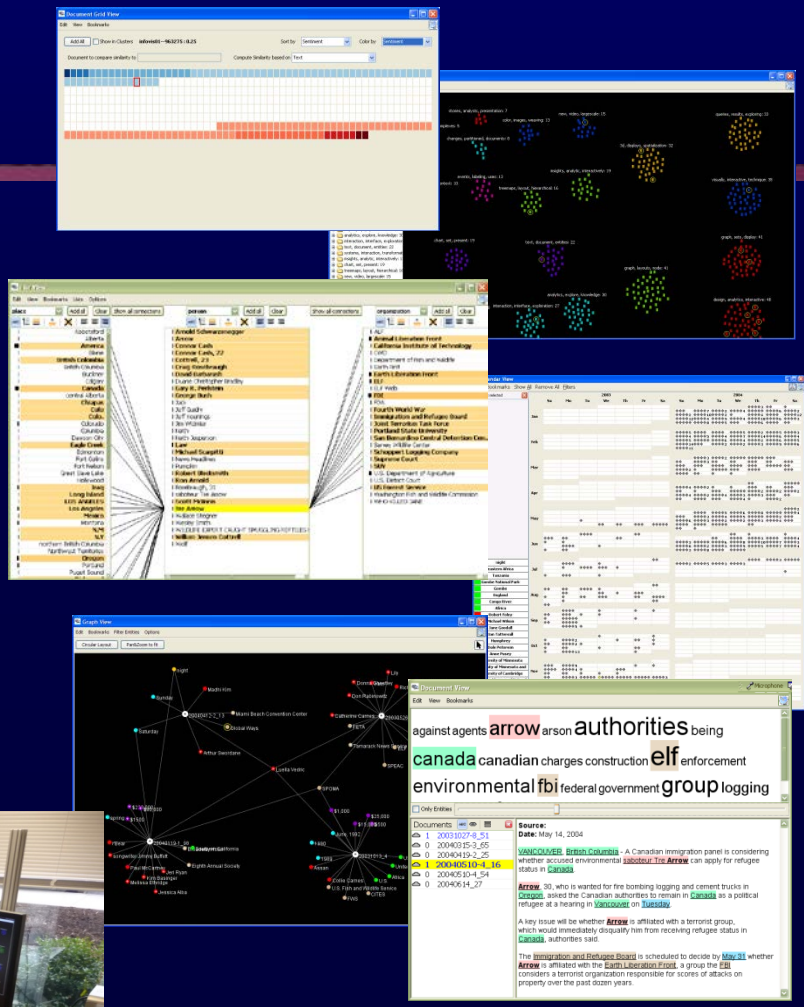
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



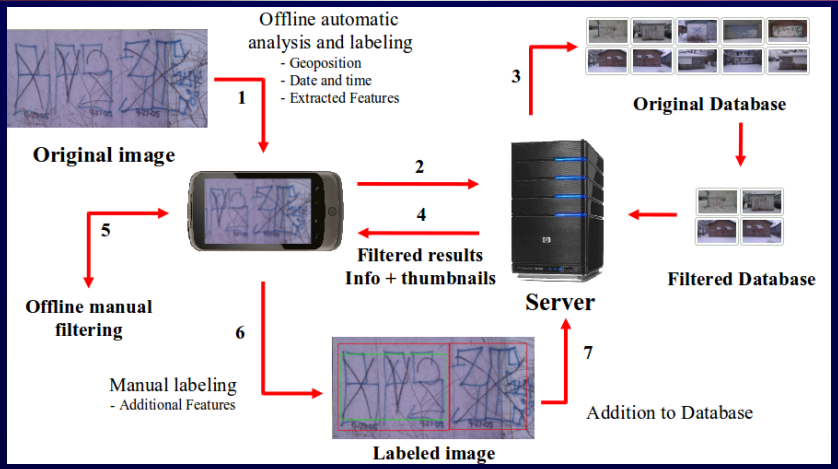
Documents in

Knowledge out





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

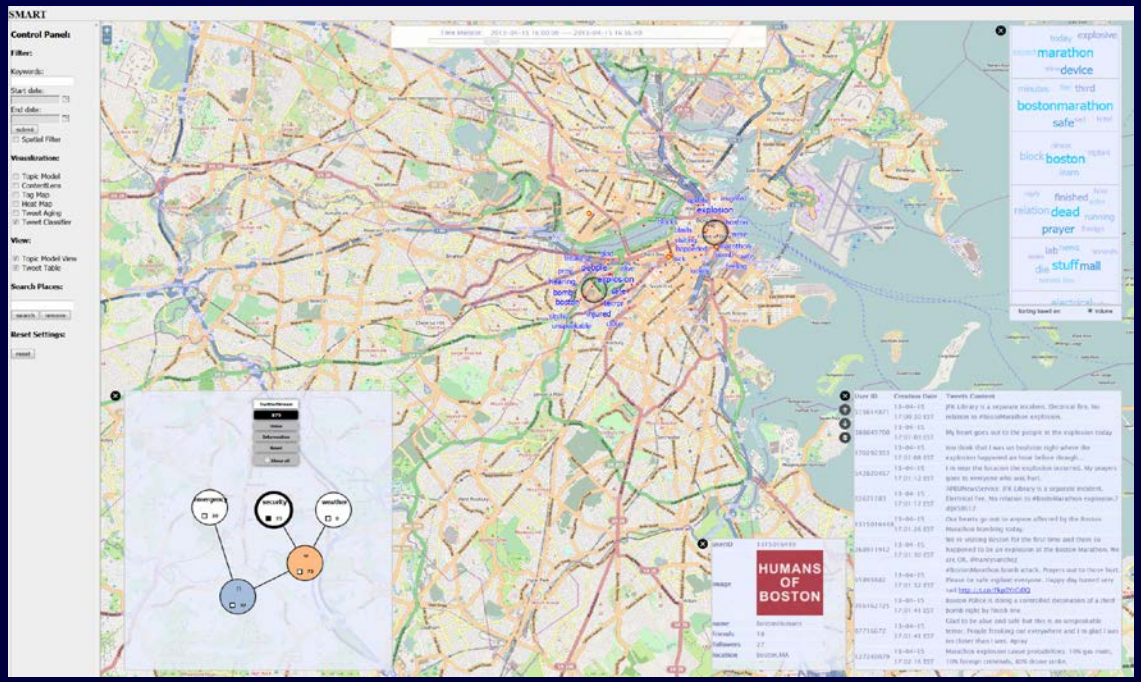
VACCINE

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

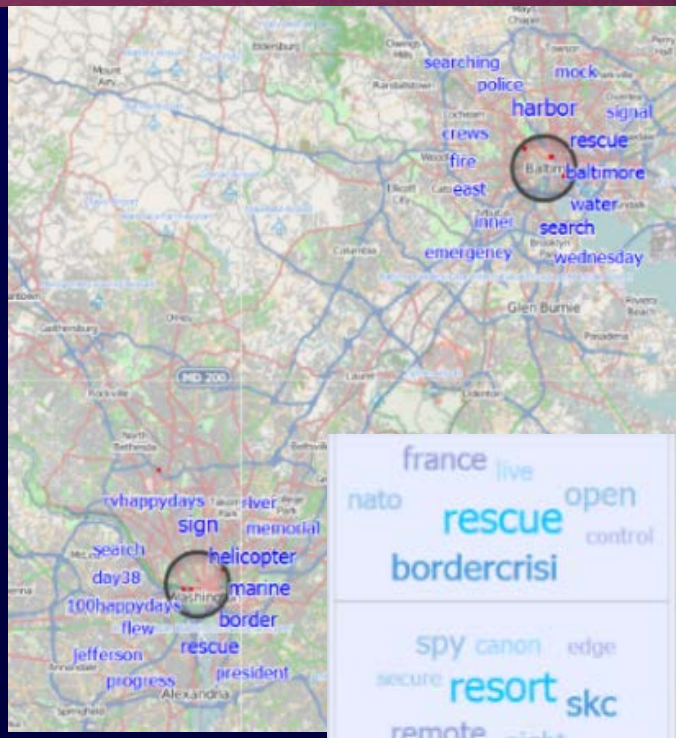
January 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

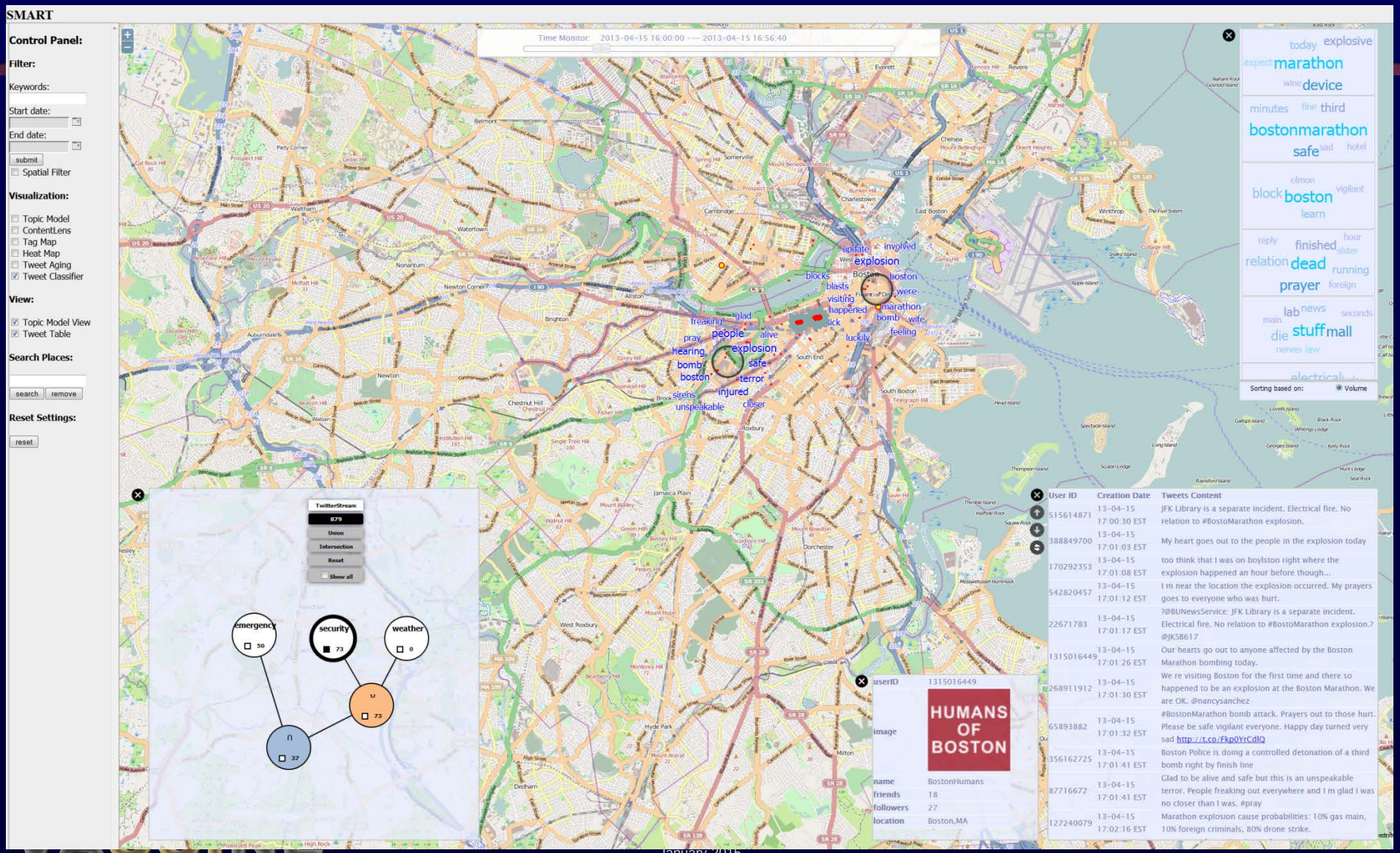
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools



VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



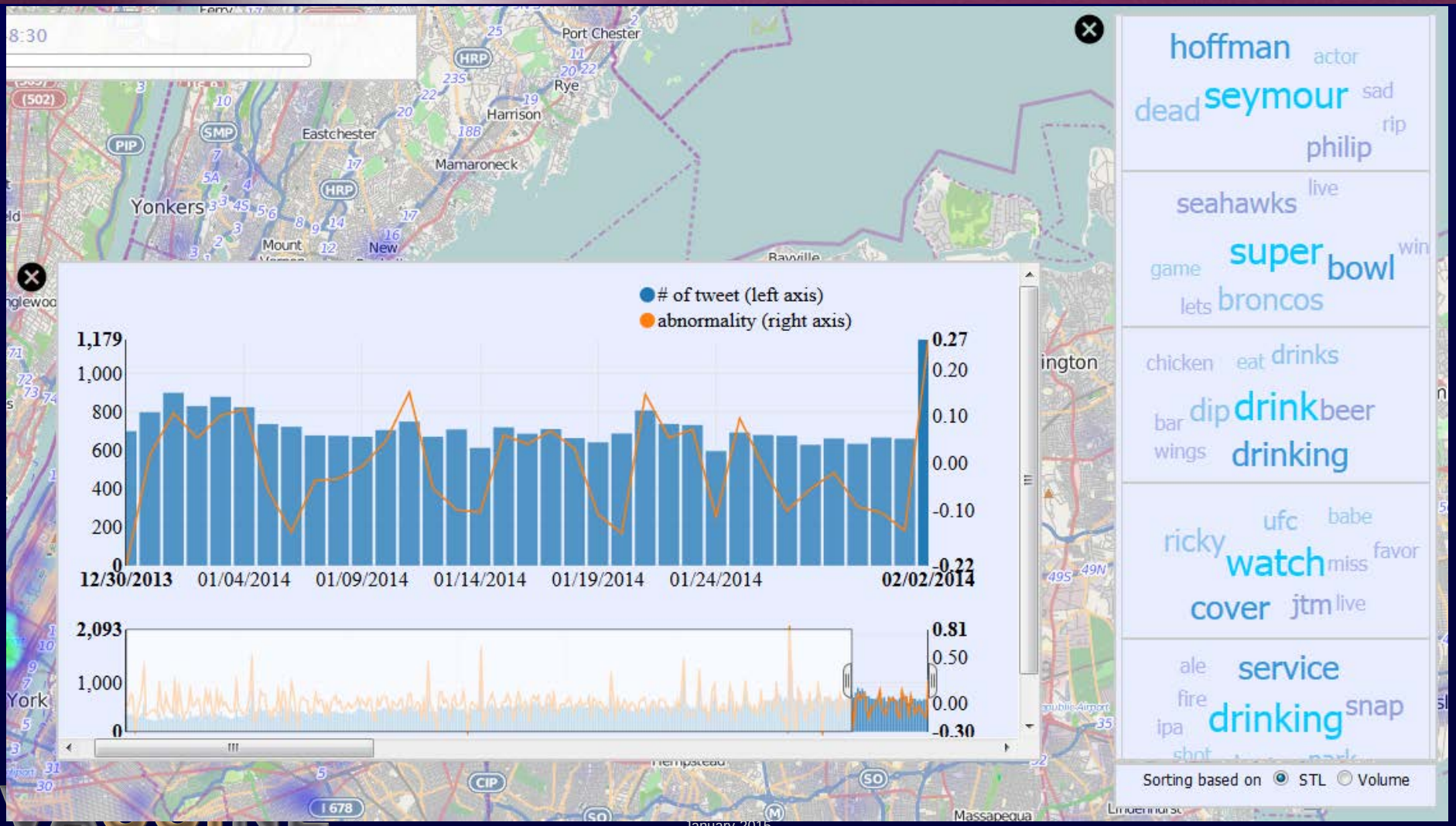
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring event
  - Purdue football games – the p
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – L



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More



RETWEETS 6

FAVORITES 4

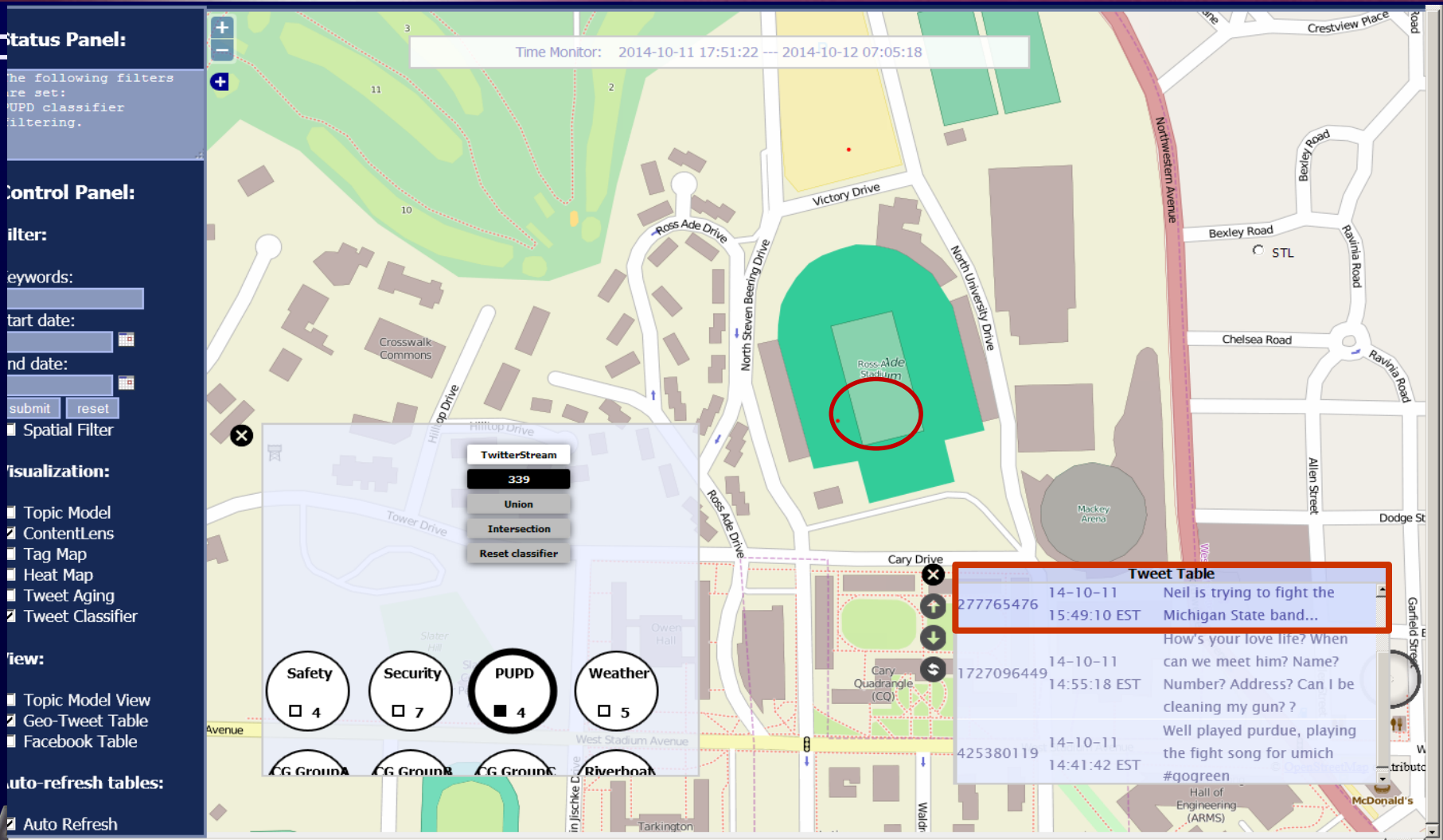
12:56 PM - 12 Oct 2014

Where Can It Be Helpful?

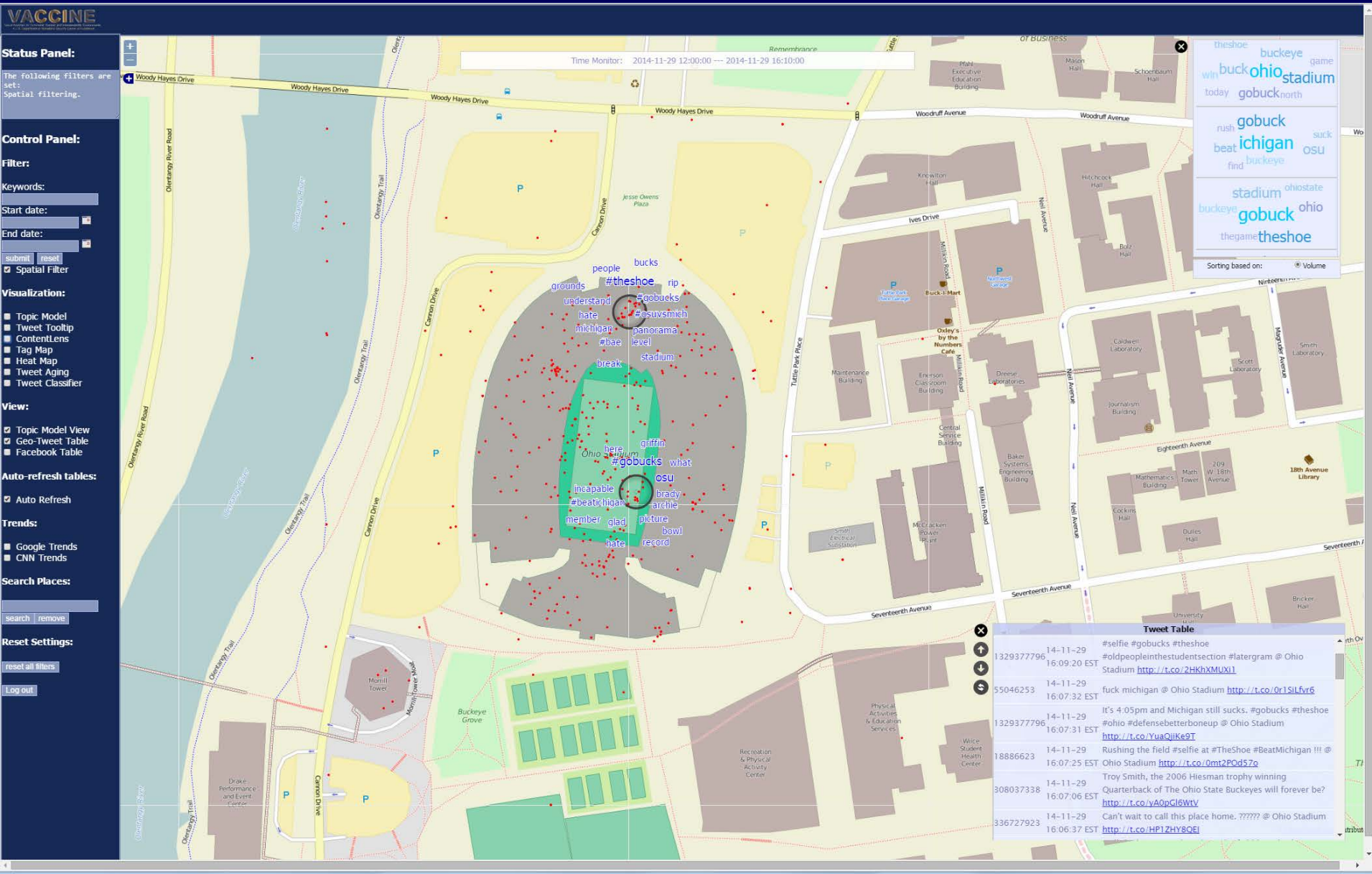
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



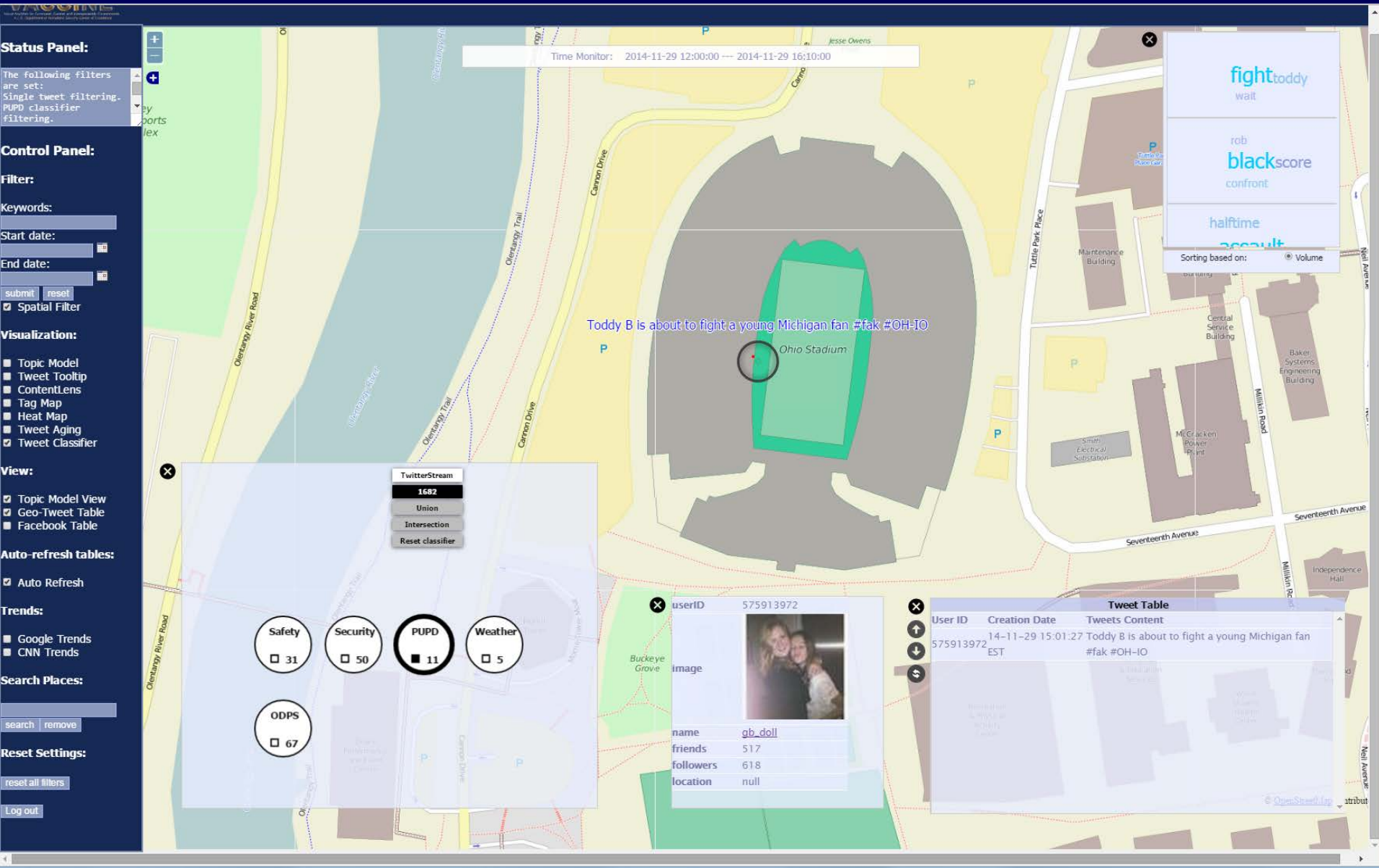
Example From Ohio State Football 11/29/14



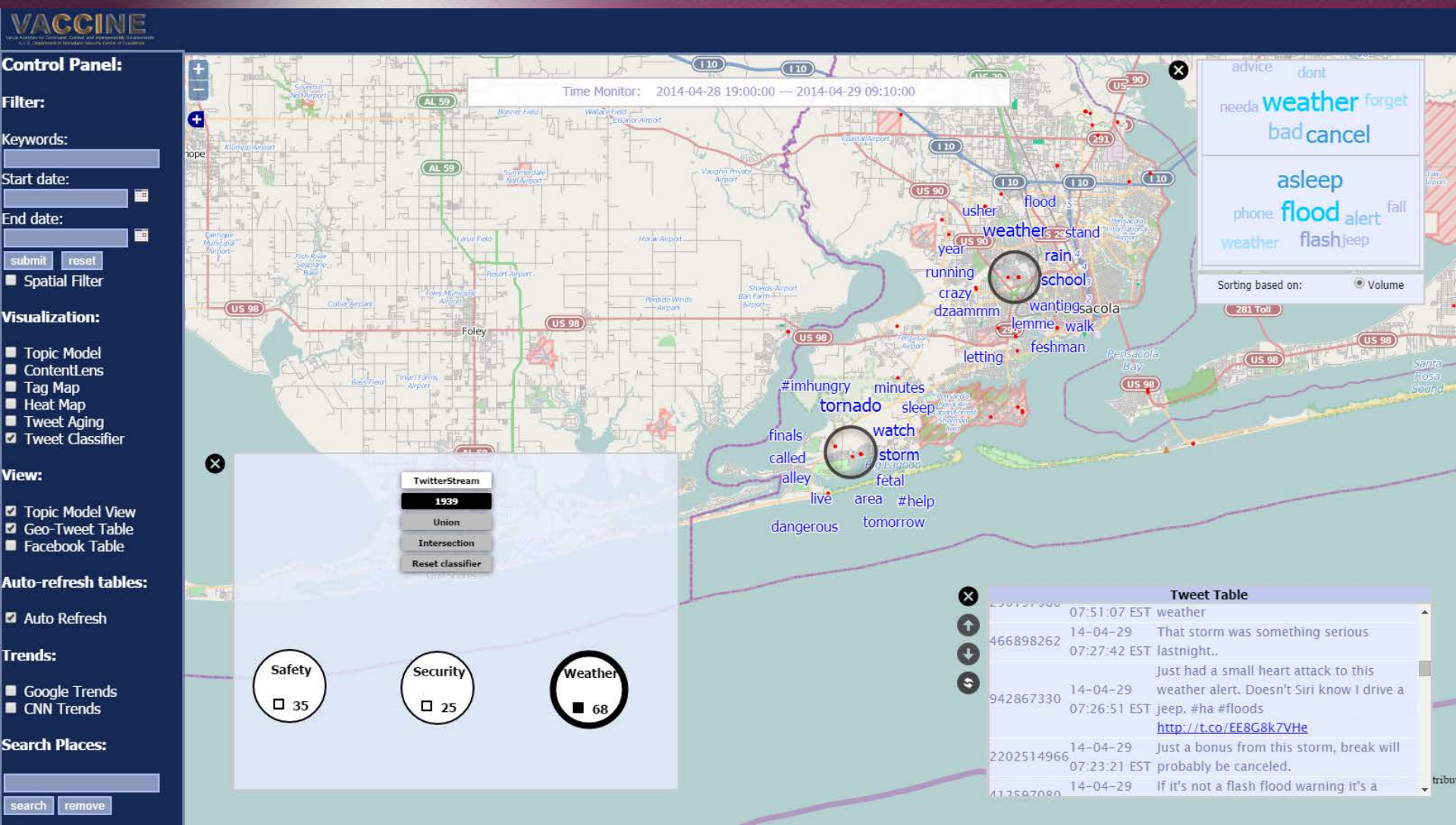
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



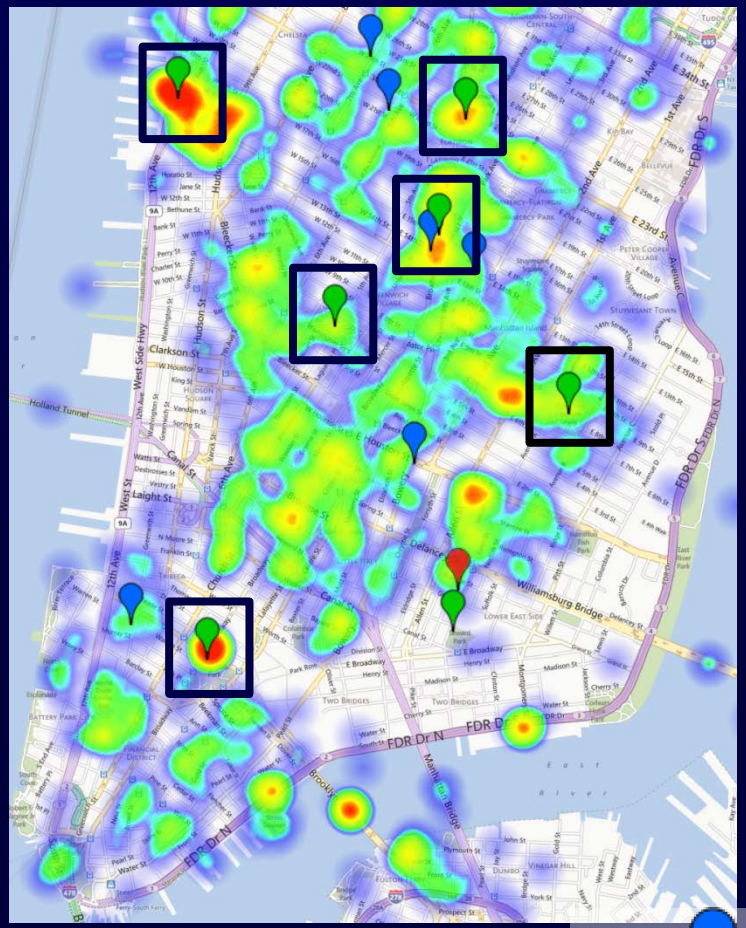
Where Can It Be Helpful?

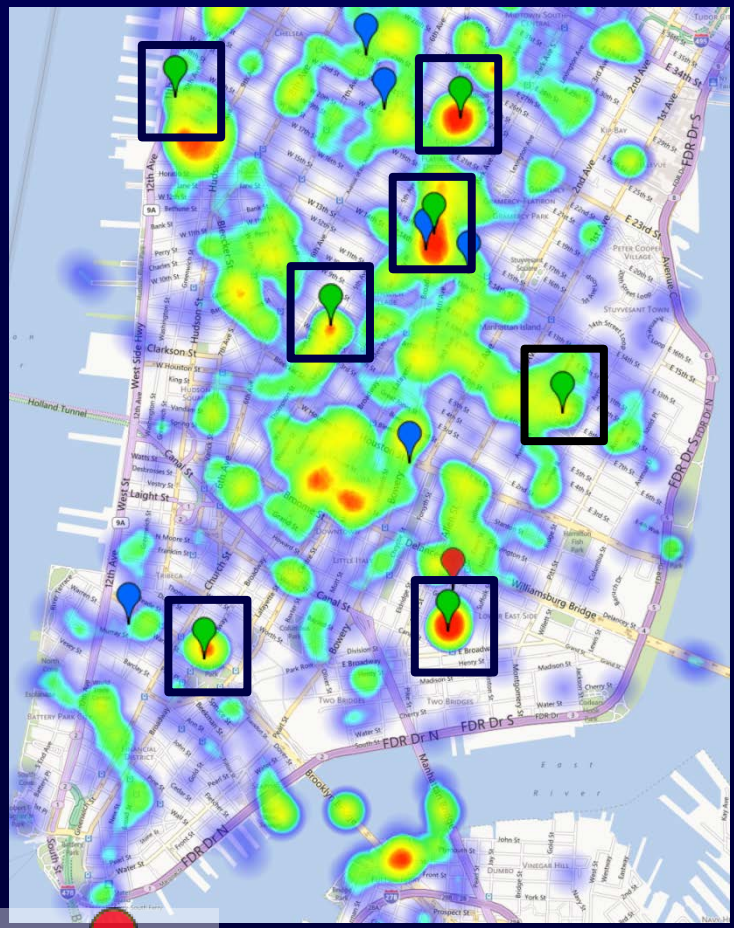


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

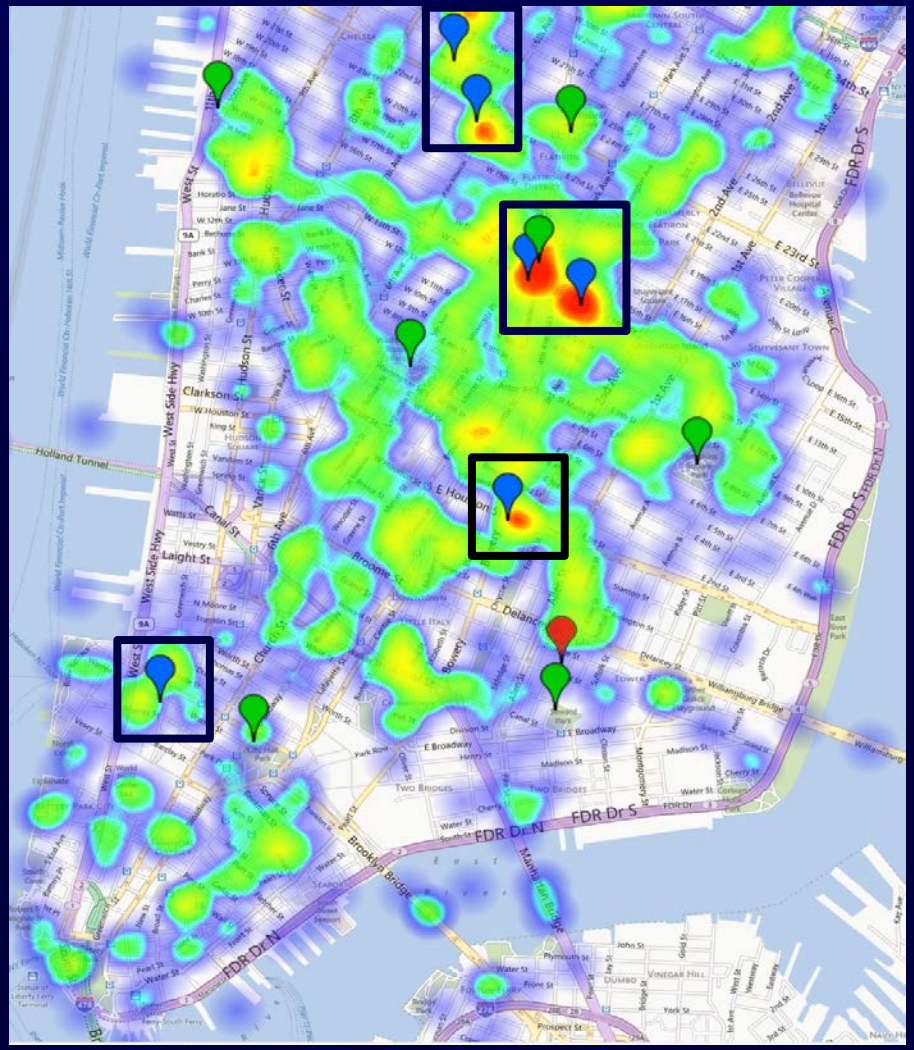
VACCINE

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



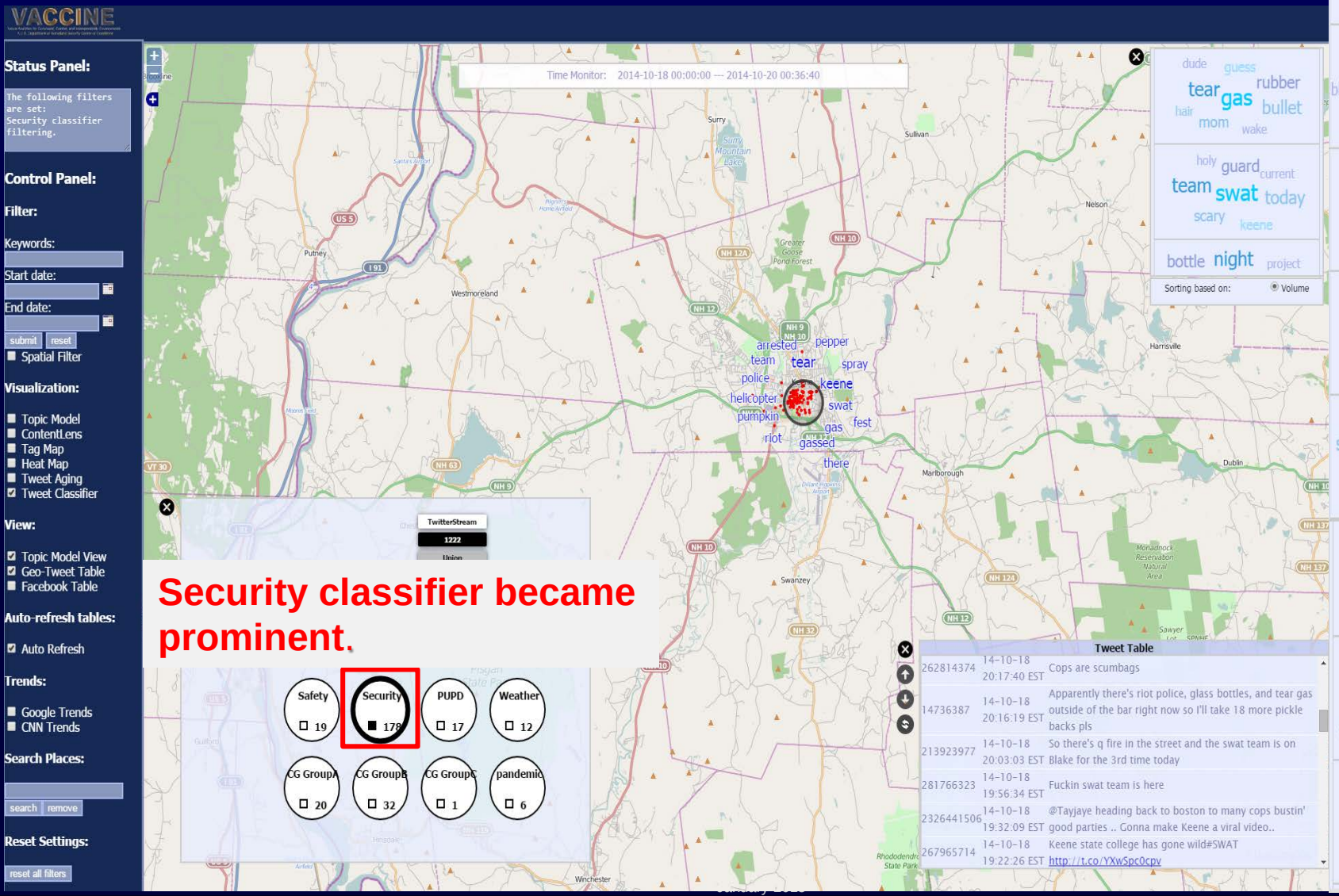
Supermarket

Park

Shelter

Keene, NH Pumpkin festival 10/18/14 – 10/19/14 Overview

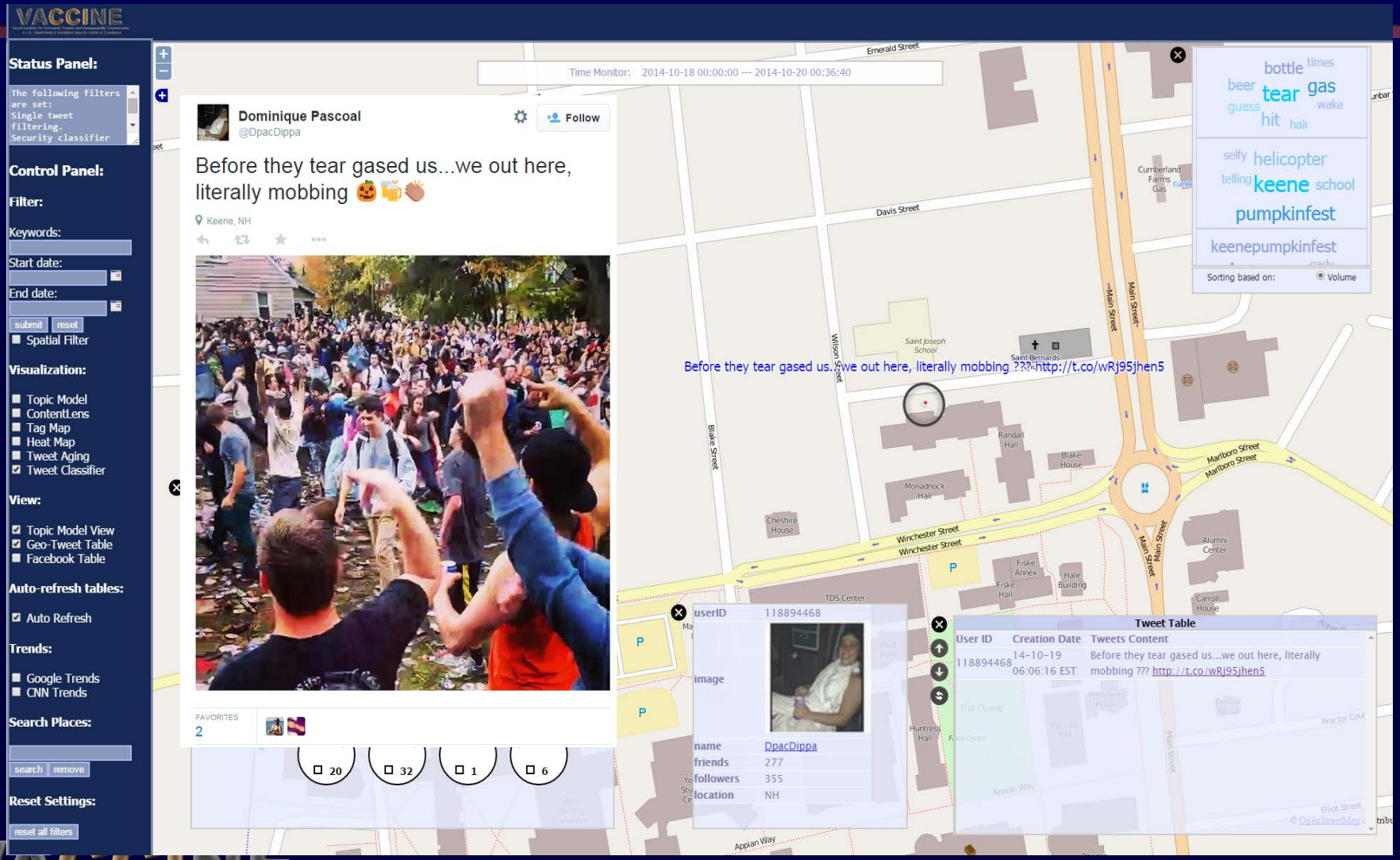
Major topics:



Security classifier became prominent.



Keene, NH

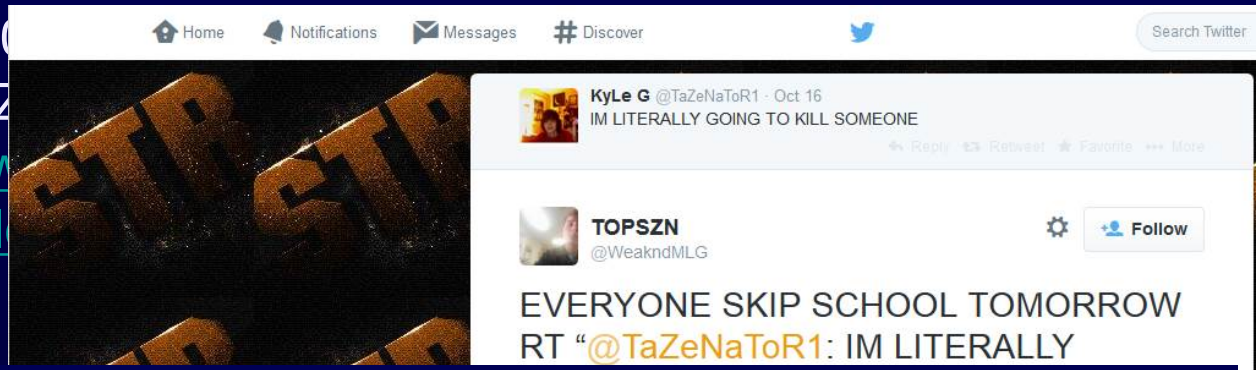

VACCINE
Status Panel:
The following filters are set:
Single tweet filtering.
Security classifier
Control Panel:
Filter:
Keywords:
Start date:
End date:
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
search
remove
Reset Settings:
reset all filters
Time Monitor: 2014-10-18 00:00:00 --- 2014-10-20 00:36:40
Dominique Pascoal
@DpacDippa
Follow
Before they tear gased us...we out here, literally mobbing
Keene, NH
FAVORITES
2
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
bottle
times
beer
tear
gas
guess
wake
hit
hair
selfy
helicopter
telling
keene
school
pumpkinfest
keenepumpkinfest
Sorting based on:
Volume
userID 118894468
image
name DpacDippa
friends 277
followers 355
location NH
Tweet Table
User ID
Creation Date
Tweets Content
118894468
14-10-19 06:06:16 EST
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
Emerald Street
Davis Street
Main Street
Wilson Street
Blake Street
Winchester Street
Marlboro Street
Appian Way
Saint Joseph School
Saint Bernards
Randall Hall
Blake House
Monadnock Hall
Cheshire House
Fiske Annex
Fiske Hall
Hale Building
TDS Center
Alumni Center
Carroll House
Huntress Hall
Proctor Court
Elliot Street
© OpenStreetMap

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



ORAM Visualization

Introduction

Risk-based Decision Making

Process that collects and organizes information about possible outcomes in an ordered structure that helps analysts make informed choices.[B. Macesker]

Performing risk analysis is a complex and challenging analytical task.

- Complex and large multivariate datasets.
- Clear understanding of the risk at hand.
- Cognitive limitations and information overload.

Related Work

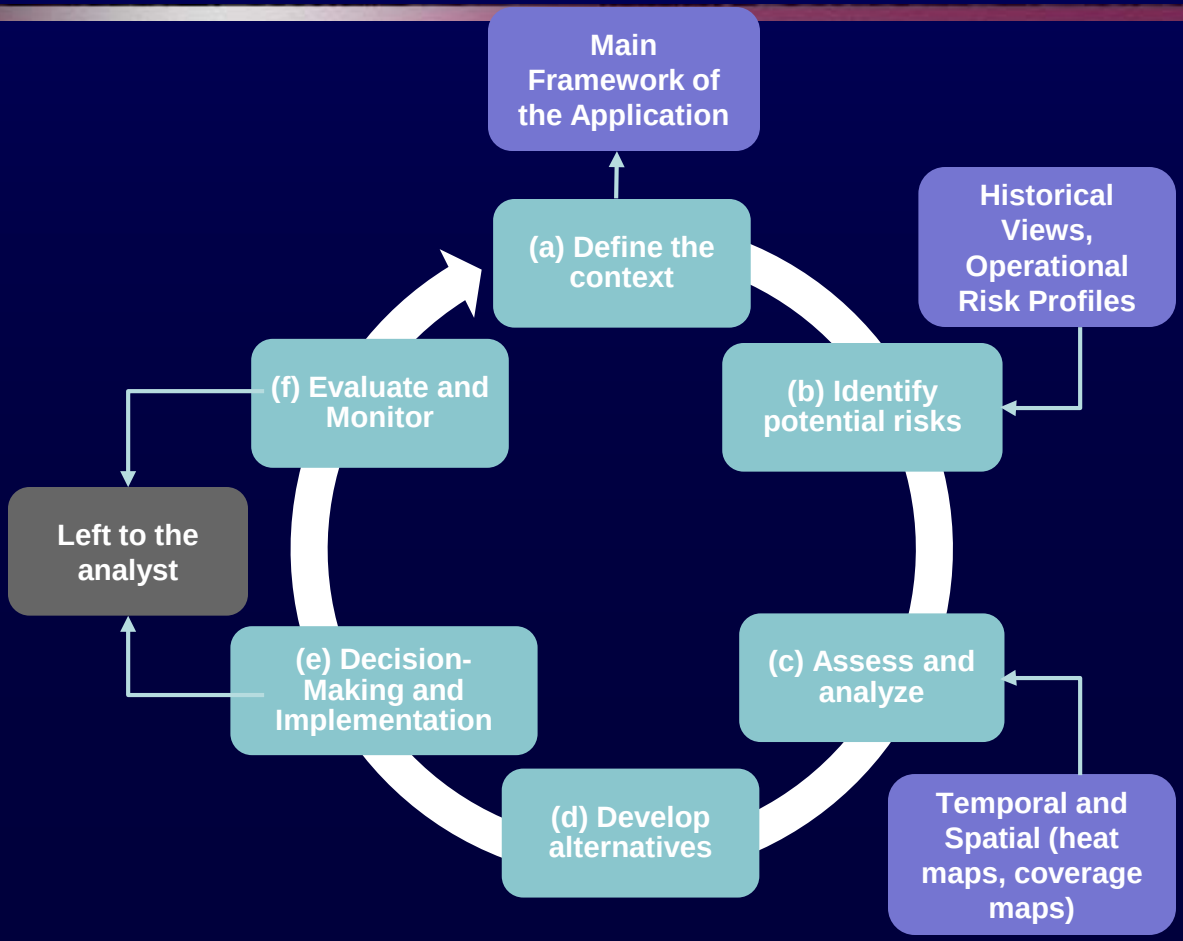
Communication of Risk

- Static risk displays [Lipkus and Holland]
- Portfolio risk management [Gershon and Eick]

Risk Analysis and Modeling

- Operational Risk Visualization [Bonafede and Marmo]
- Visualization Techniques to Present Risks and Options for Risk Mitigation[Feather et al.]
- Visual exploration and classification models for risk assessment [Migut and Worring]

Problem Characterization: Risk Management Process



Risk Terminology

- ORAM – Operational Risk Assessment Model
- 11 Missions in the Coast Guard – 10 Missions with risk values
- RIN – Risk index numbers.
- Total Risk – Mitigated Risk = Residual Risk.

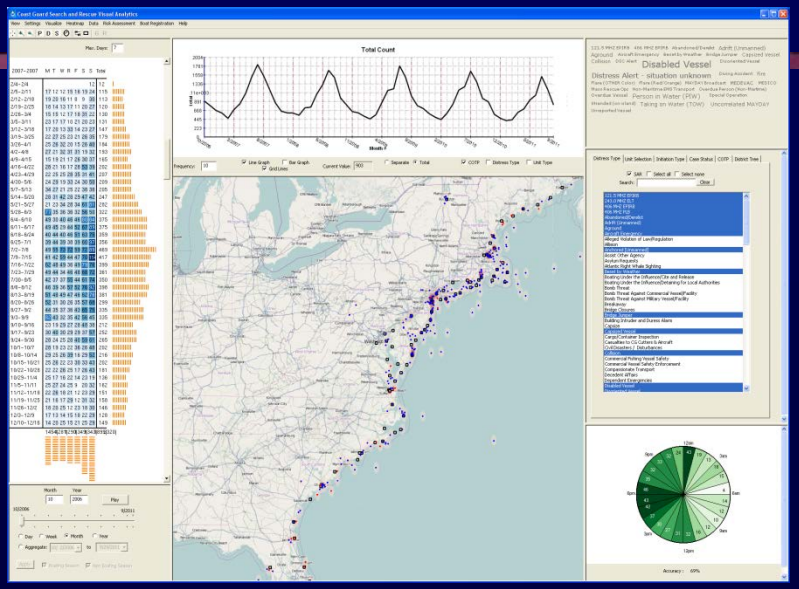
Visual Analytics Environment

Design Requirements:

Interaction with the system.

Integration with previously used tool.

Creation of customizable scenarios.





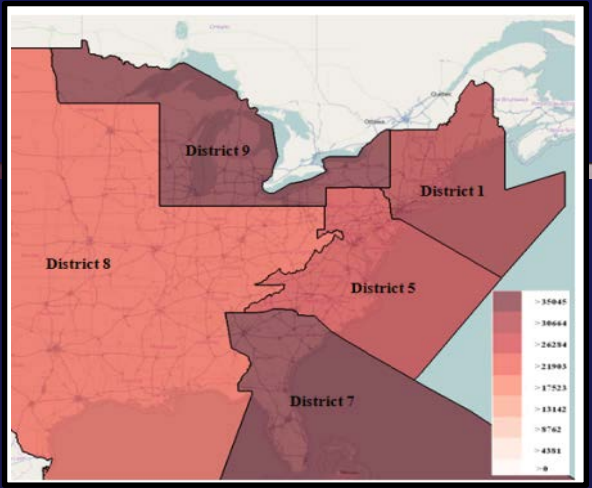
VACCINE

Scenario

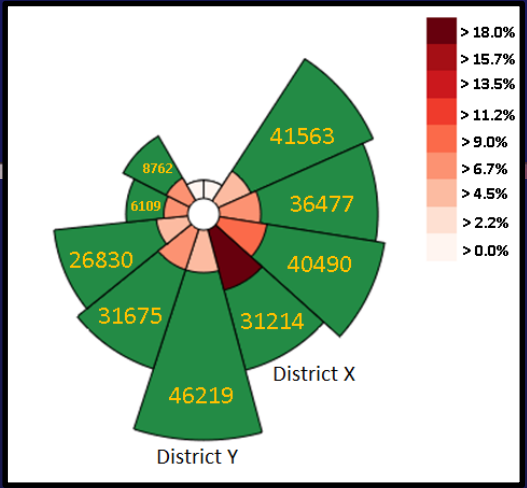
Questions that drive the planning strategy:

- What risks exist in the region and where they are distributed?
- Where are our resources allocated?
- What constraints exist in the system that will require a prioritization of resource use?

Identify and Analyze Potential Risk



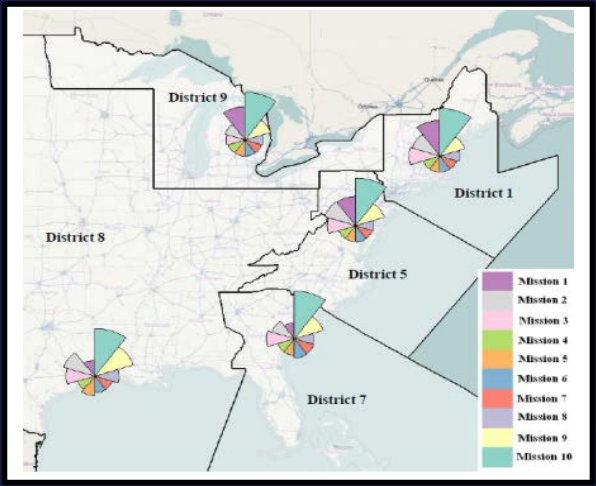
(A) District Choropleth Map for Total Risk across all missions



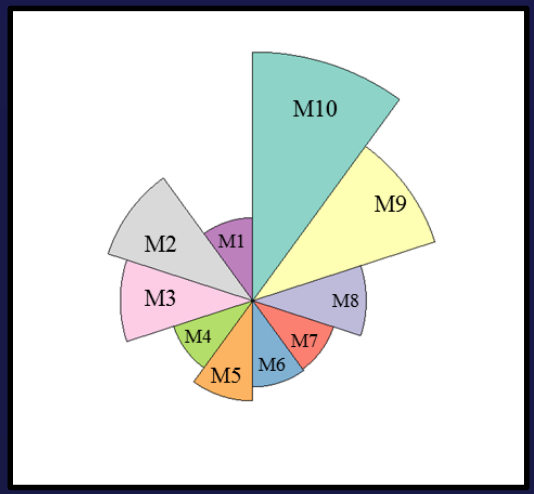
(C) Risk profiles across all districts and mission areas



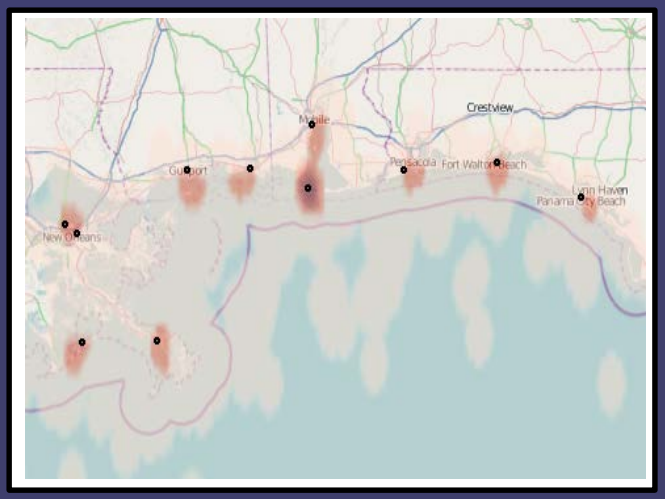
(E) Total M10 Risk Values distributed among District X's station AOR



(B) Total Risk distribution by mission for any given district.



(D) District X Risk Profiles across all mission area



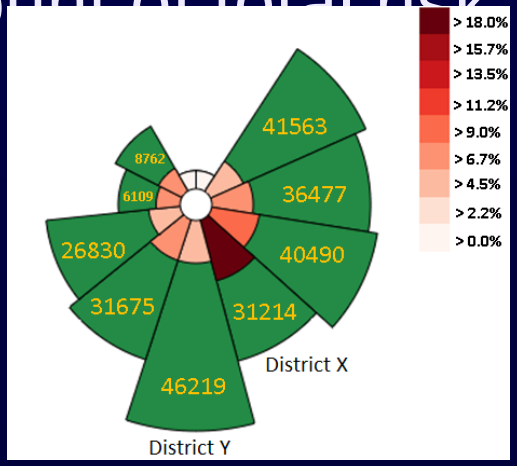
(F) KDE heatmap showing Total M10 Risk values

General understanding on how the risk is distributed between the different areas and missions.

Specific areas can be analyzed by time and the analyst can obtain details by demand.

Scenario

- 1. In a resource constrained environment, we want to use resources in the mission area that provides the greatest return on investment (large amount of total risk but very little residual risk).



- District Y has the largest total risk values, it mitigated most of the risk effectively.
- District X shows less total risk, but the amount of residual risk as well as residual to total risk ratio is the highest as encoded by the darkest red shade.
- District X can be seen as more problematic than District Y

Scenario

2. Determine the key drivers of risk within a district. What mission drives the risk?

3. Are there several big events that drive the risk, or are there many small events with smaller consequences accumulated to affect the operation?

4. Evalu sou





Feedback

- The prototype components went through an iterative design refinement process with the collaboration of four Coast Guard personnel: an operation research analyst, a former Coast Guard officer, one in-field officer, and a high level officer.

- Overall they emphasized the system allowed them to answer questions derived from developing the planning strategy in a fast manner. It also allows them to observe the underlying factors that drive the risk profiles.

Summary

- The system presented assists in the cognitive processes of quantifying evidence and comparing lines of evidence, as well as adapting to new information.
- Case Study providing useful visual reference that can communicate recommendations based on risk management.
- The risk-based decision making process described serves as a blueprint for future systems dealing with risk values and resource planning.

Acknowledgments

• This work was partially funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0003. Jang's work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A1011170).
• Thank you to Ben Maule and Joe DiRenzo from the US Coast Guard for their help.

Backup Slides

Definition of Risk

- “Potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences”
- R. S. Committee. DHS Risk Lexicon. Homeland Security, 2010

Coast Guard Missions

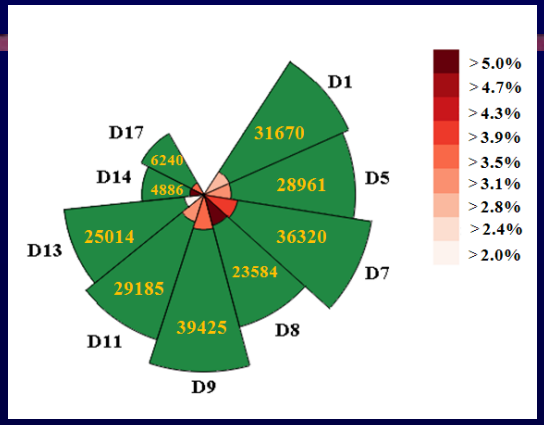
Coast Guard Mission	CG Mission (Short Name)
Ports, Waterways, and Coastal Security	PWCS
Drug Interdiction	CD
Migrant Interdiction	AMIO
Living Marine Resources	LMR
Other Law Enforcement	OLE
Marine Safety	MARSAFE
Search and Rescue	SAR
Marine Environmental Protection	MEP
Defense Readiness	DR
Ice Operations	DOMICE
Aids to Navigation	ATON

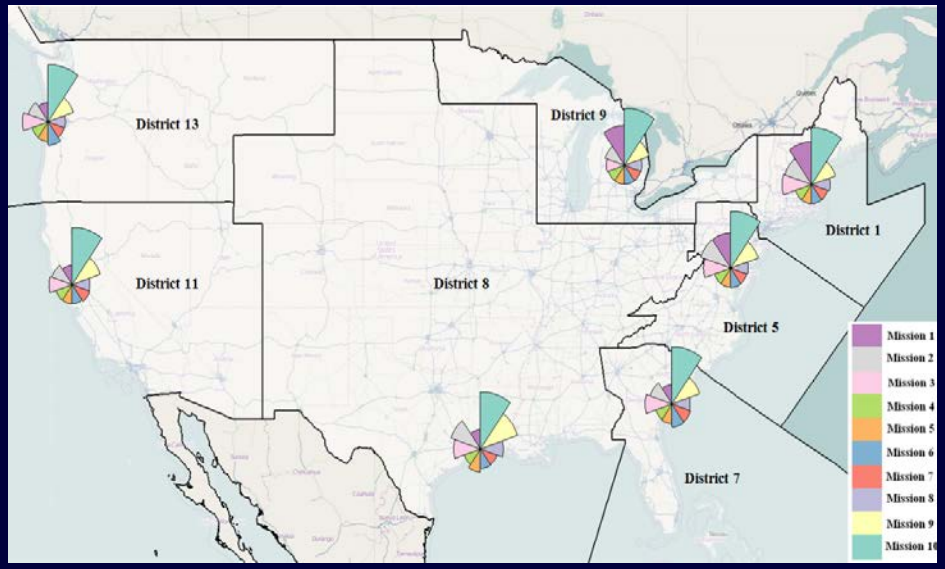
Operational Risk Profiles

Compare risk values against districts.

Compare risk values against missions

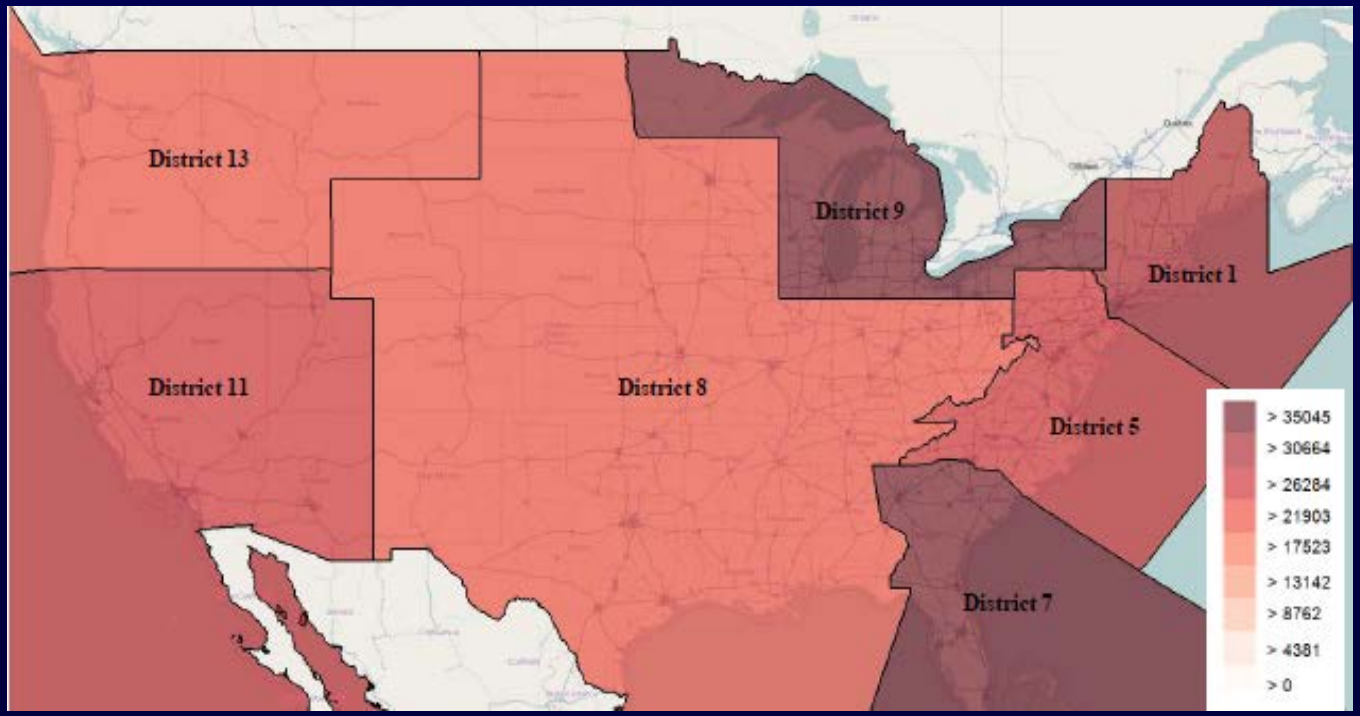
Iterative process with feedback from the Coast Guard





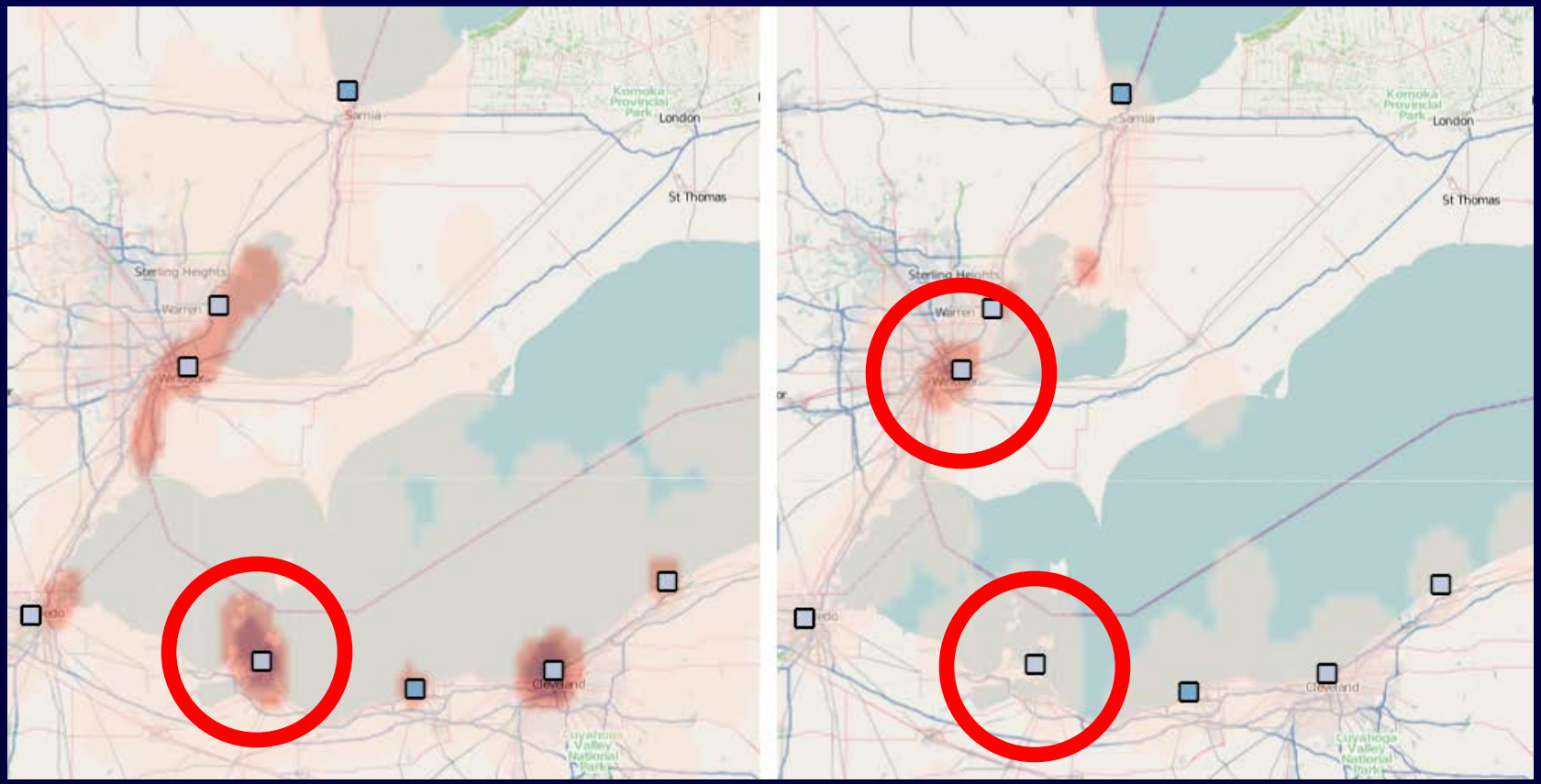
VACCINE

Choropleth Maps



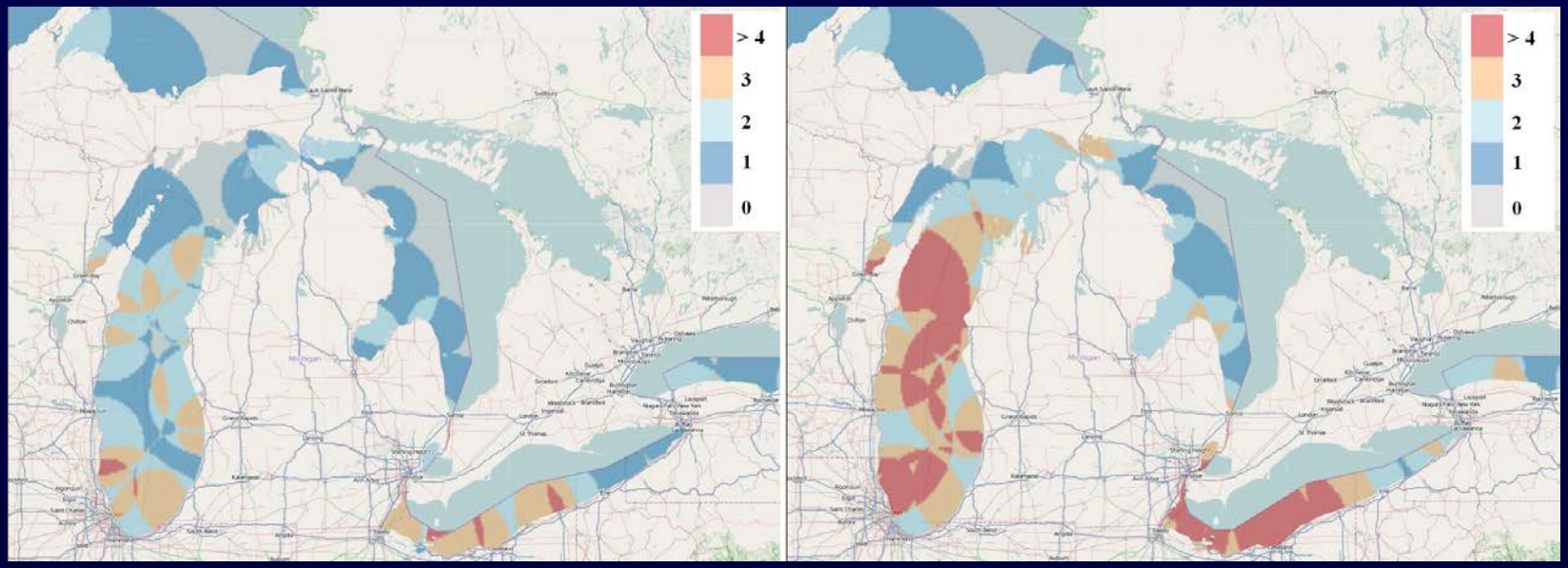
Heatmap for risk values

- Total vs Residual comparison.

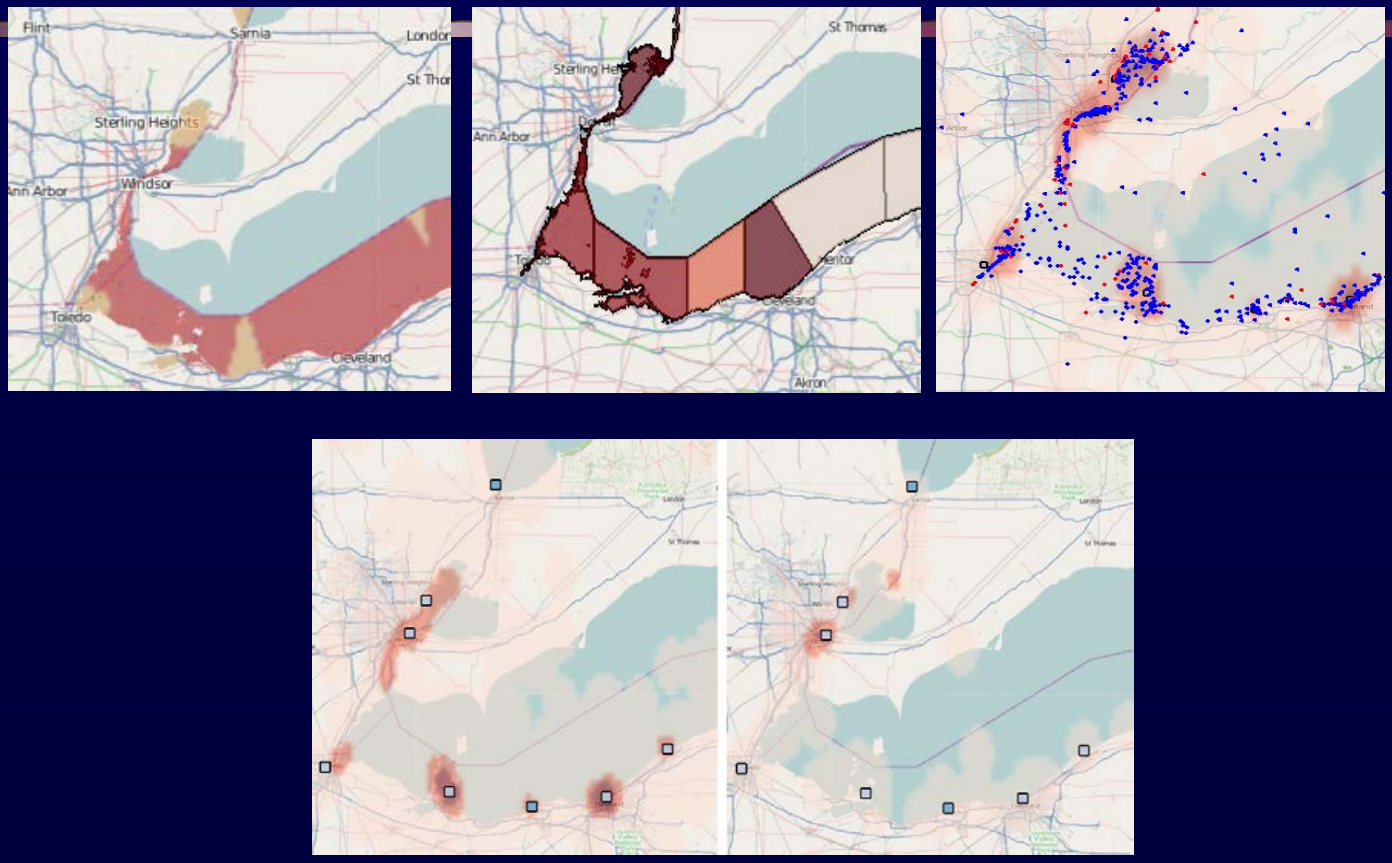


VACCINE

Coverage Maps for Performance Evaluation

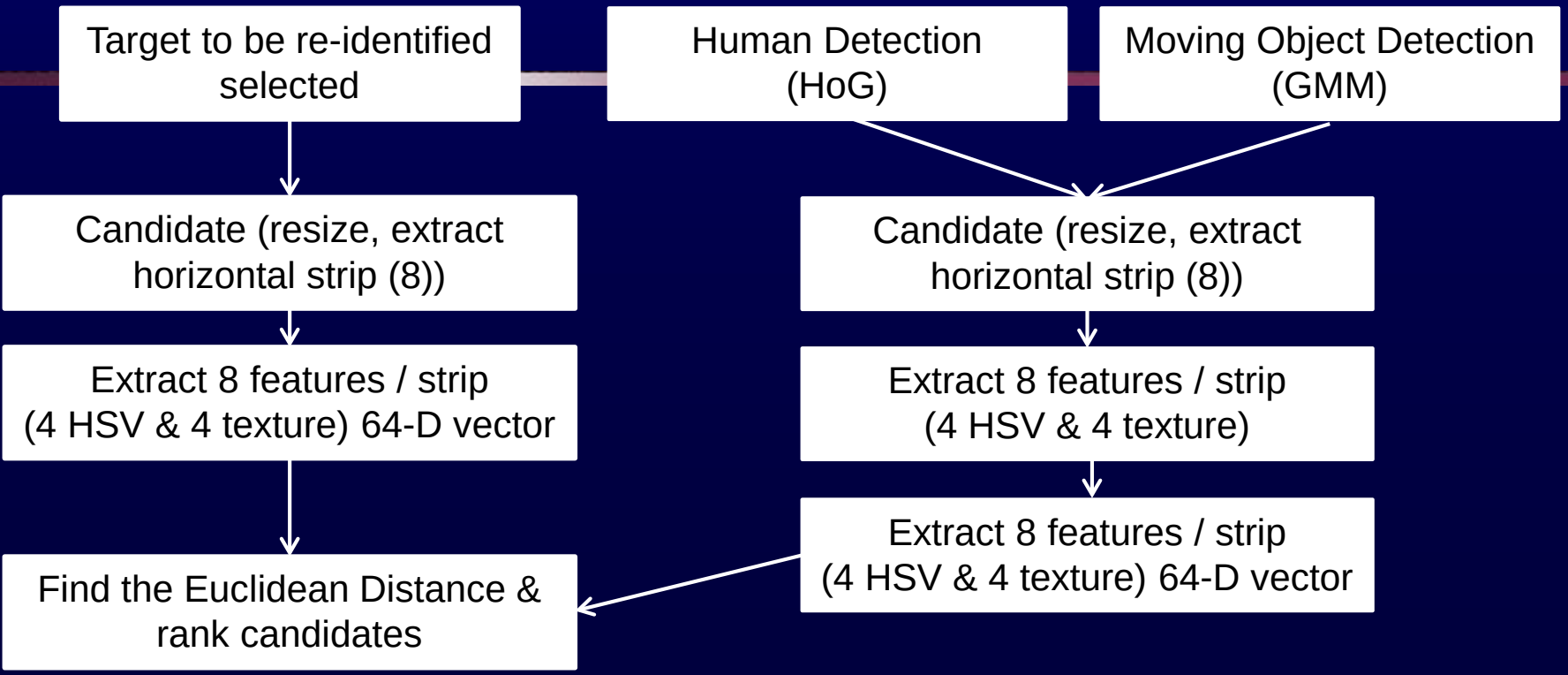


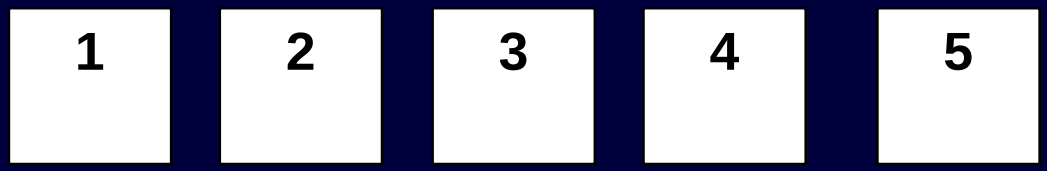
Analyst's Evaluation



VACCINE

EBOLO





The top 5 identified targets

EBOLO Analysis



Human Detection (HoG)

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

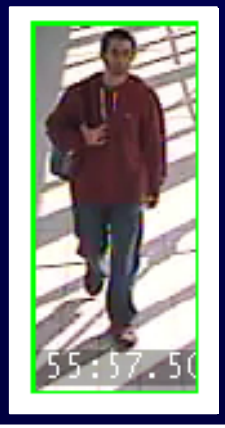
Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

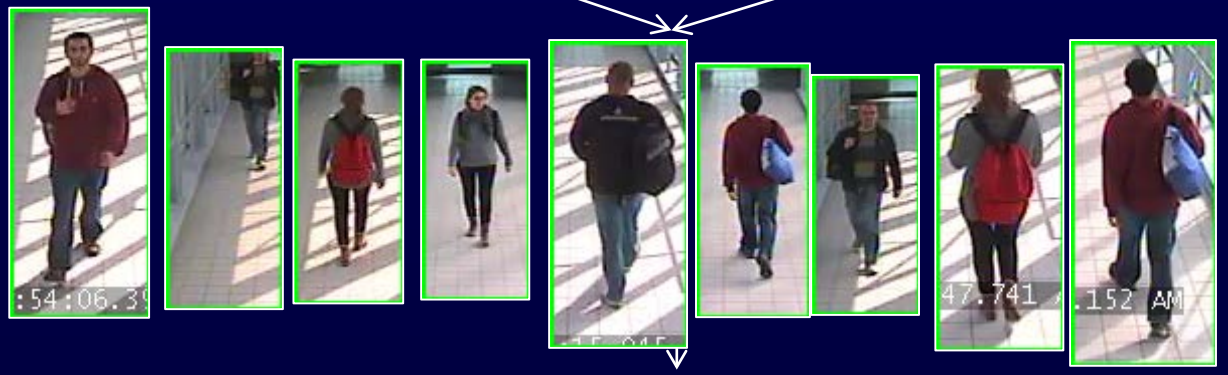
Experimental Results

Target to be re-identified selected

Human Detection (HoG)

Moving Object Detection (GMM)





Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

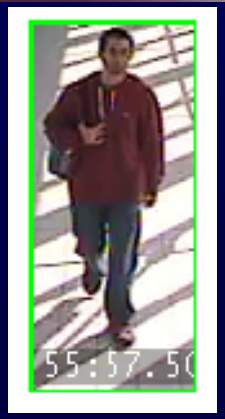
Find the Euclidean Distance & rank candidates

1	2	3	4	5

The top 5 identified targets

Experimental Results

Target to be re-identified selected



Find the Euclidean Distance & rank candidates





The top 5 identified targets

VIPER Lab

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

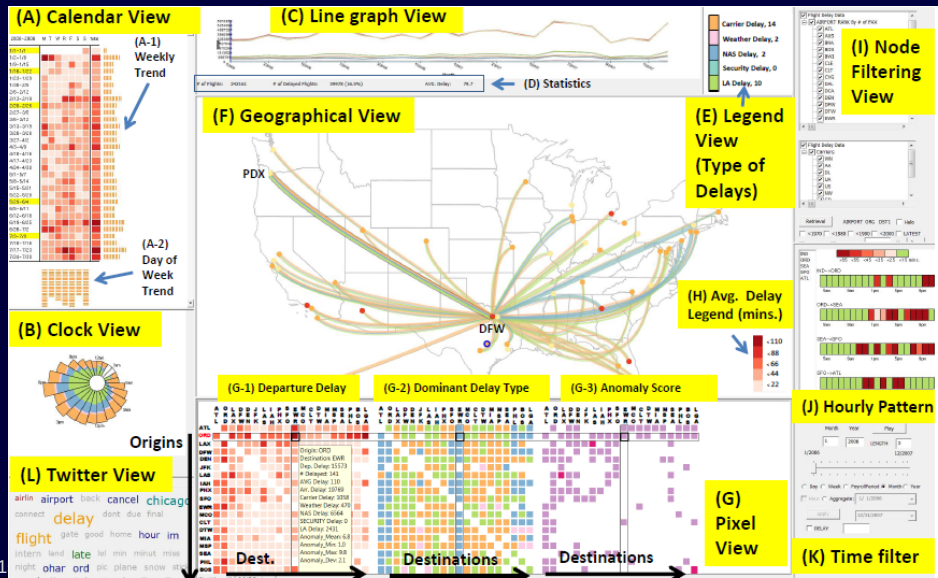
Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

July 2015

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects



VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

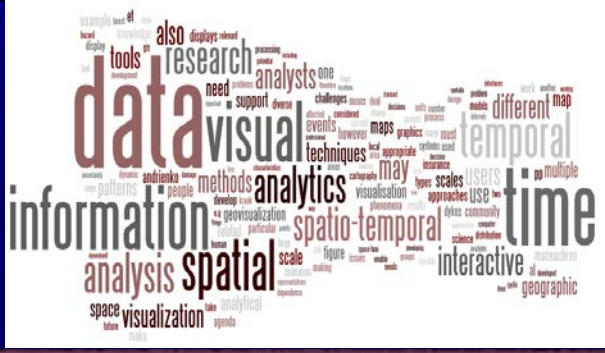
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America



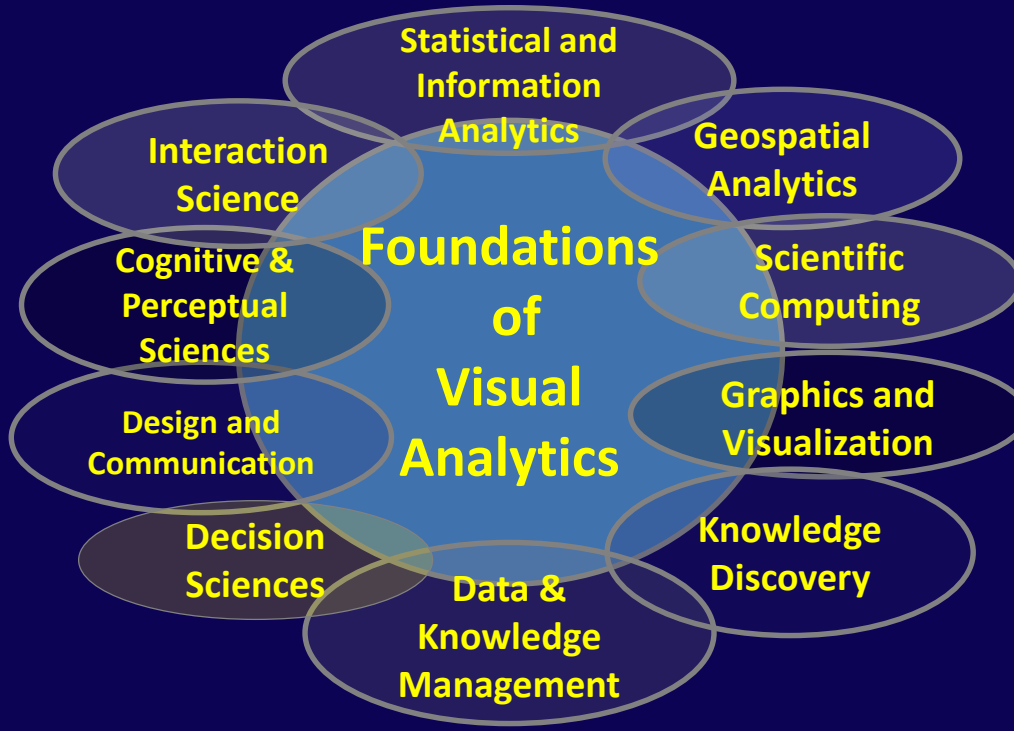


What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

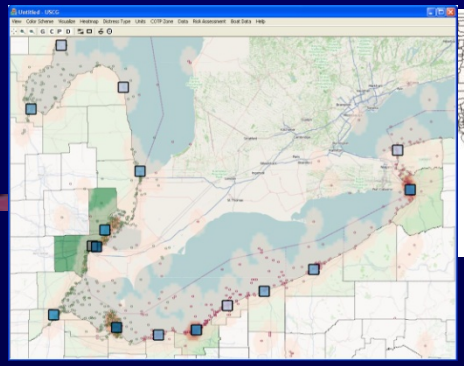
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

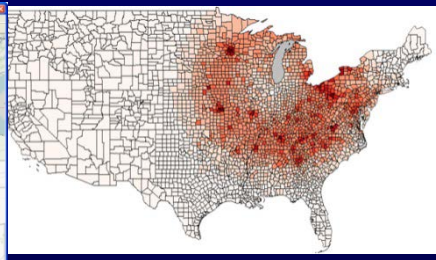
VADM Robert Parker, 2012 MRS Keynote Address

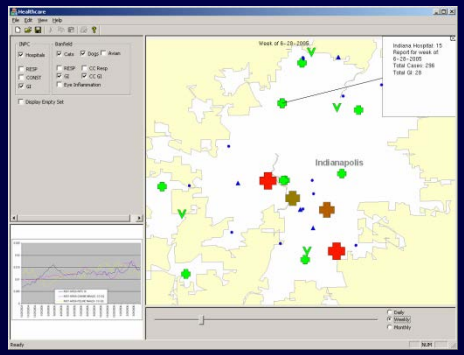


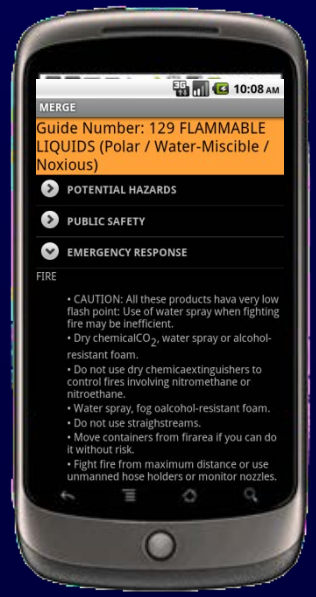
Public Safety & Health

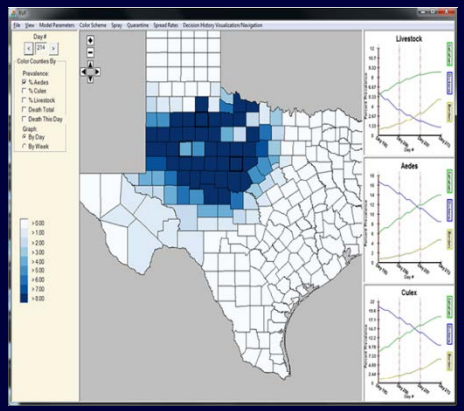
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering









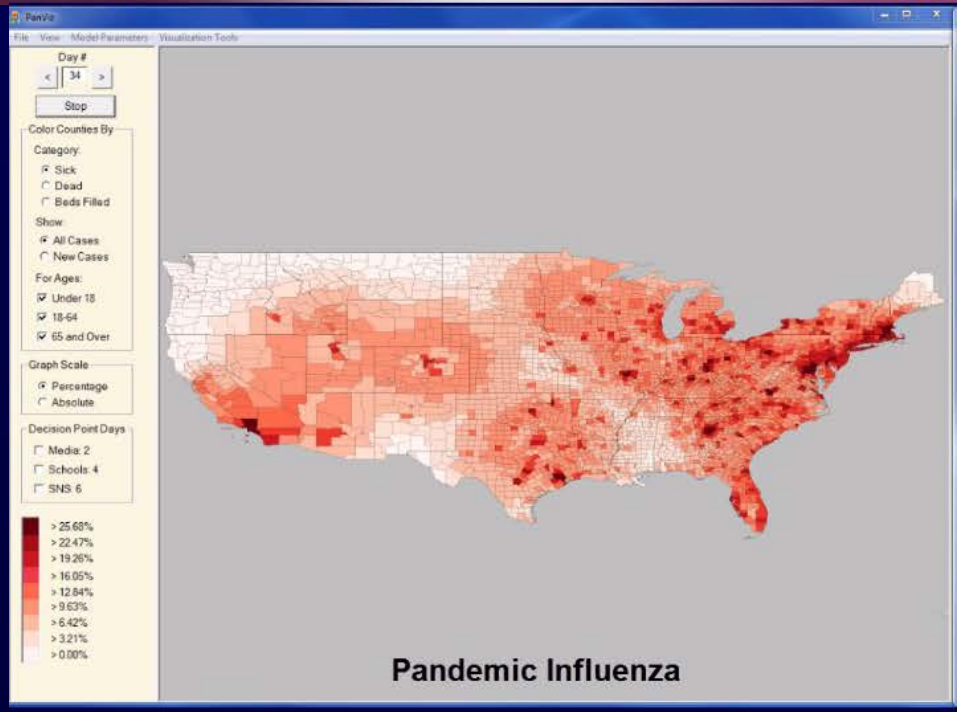


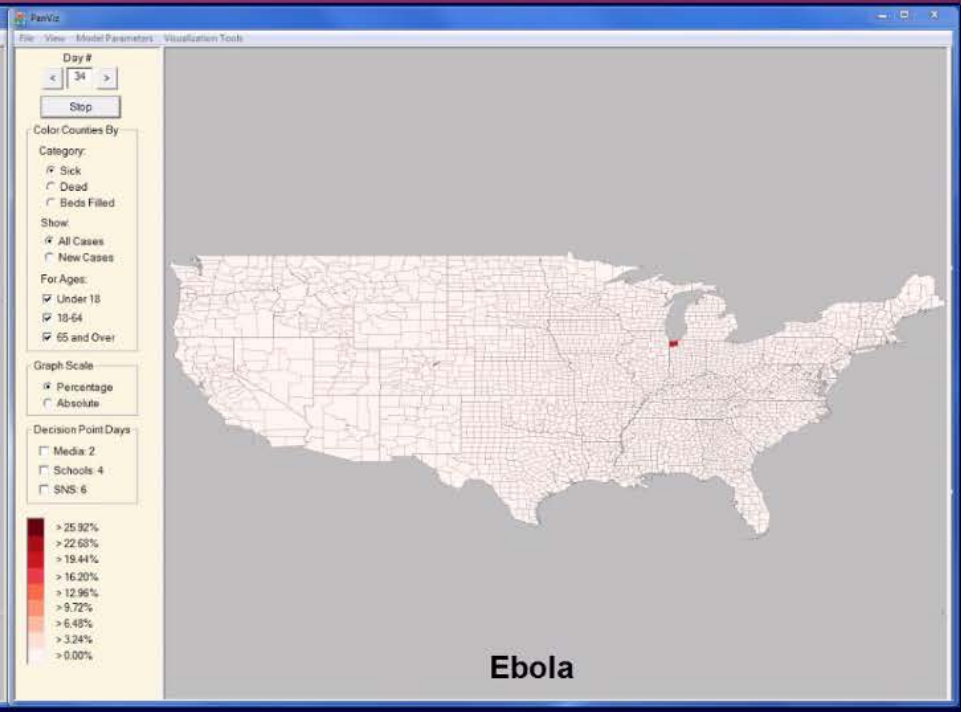


VACCINE

Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL



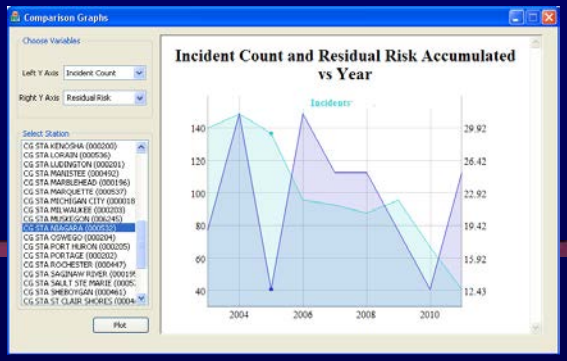


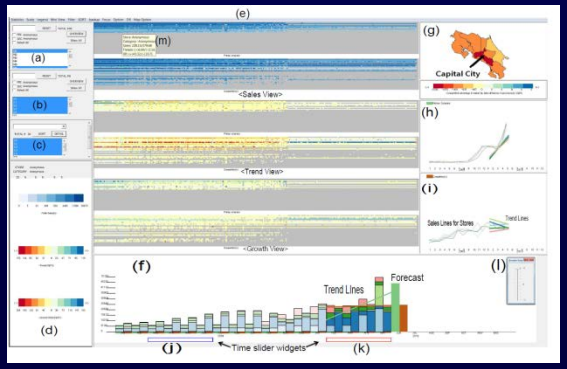
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

VACCINE

Business Visual Analytics

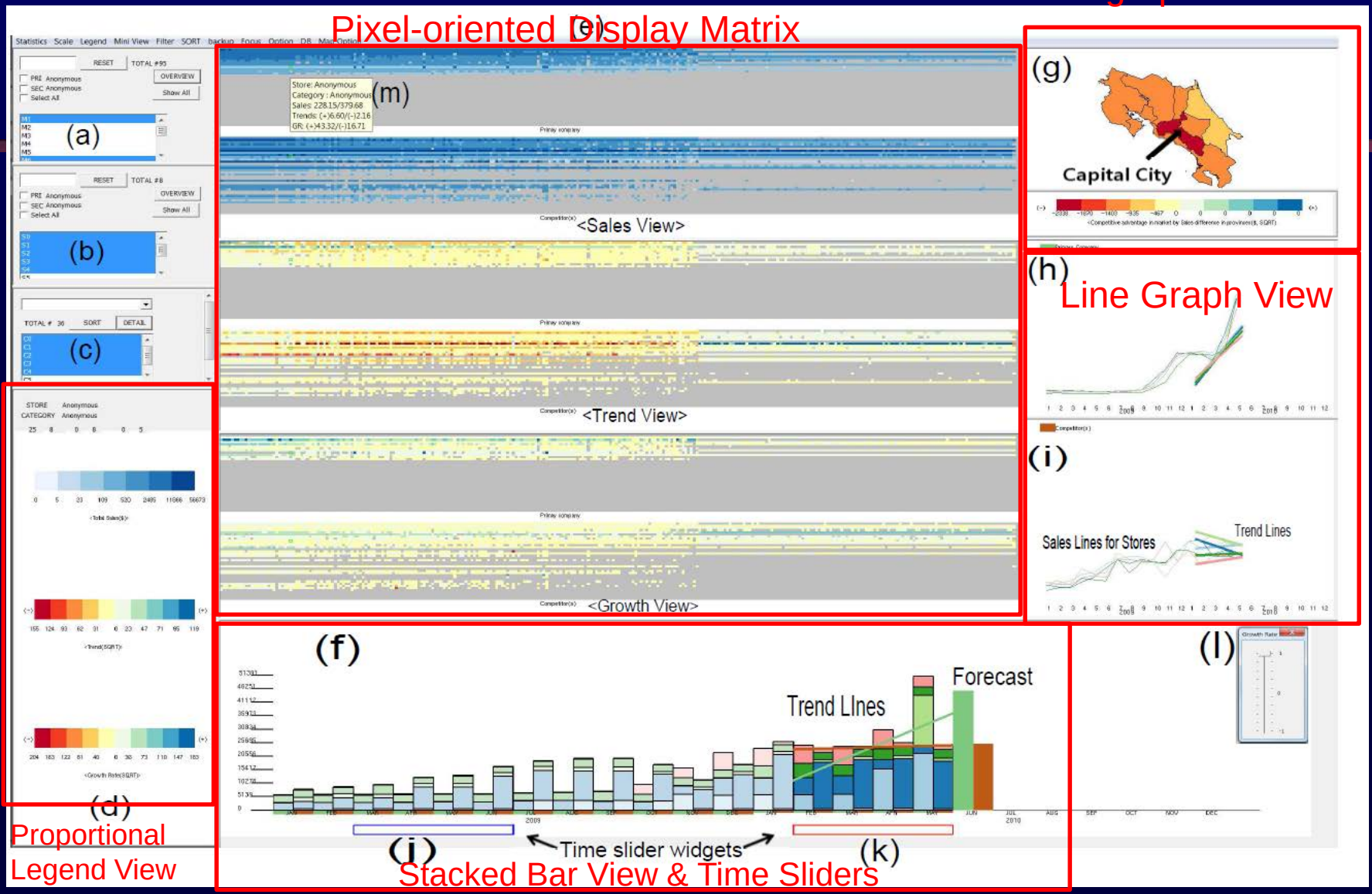
- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







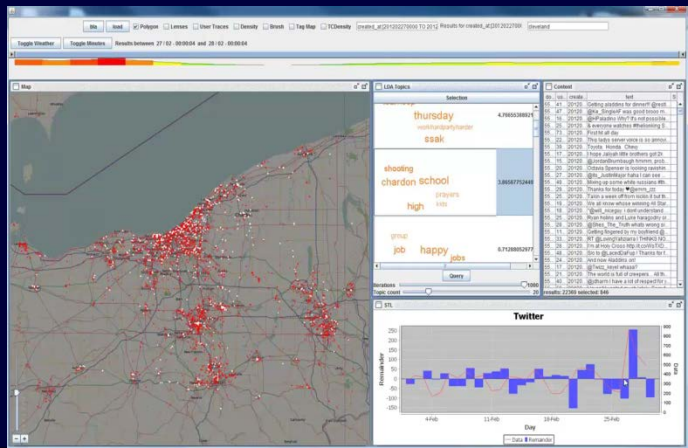
Market Analyzer



Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Our Spatiotemporal Visual Analytics Work Contributions

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

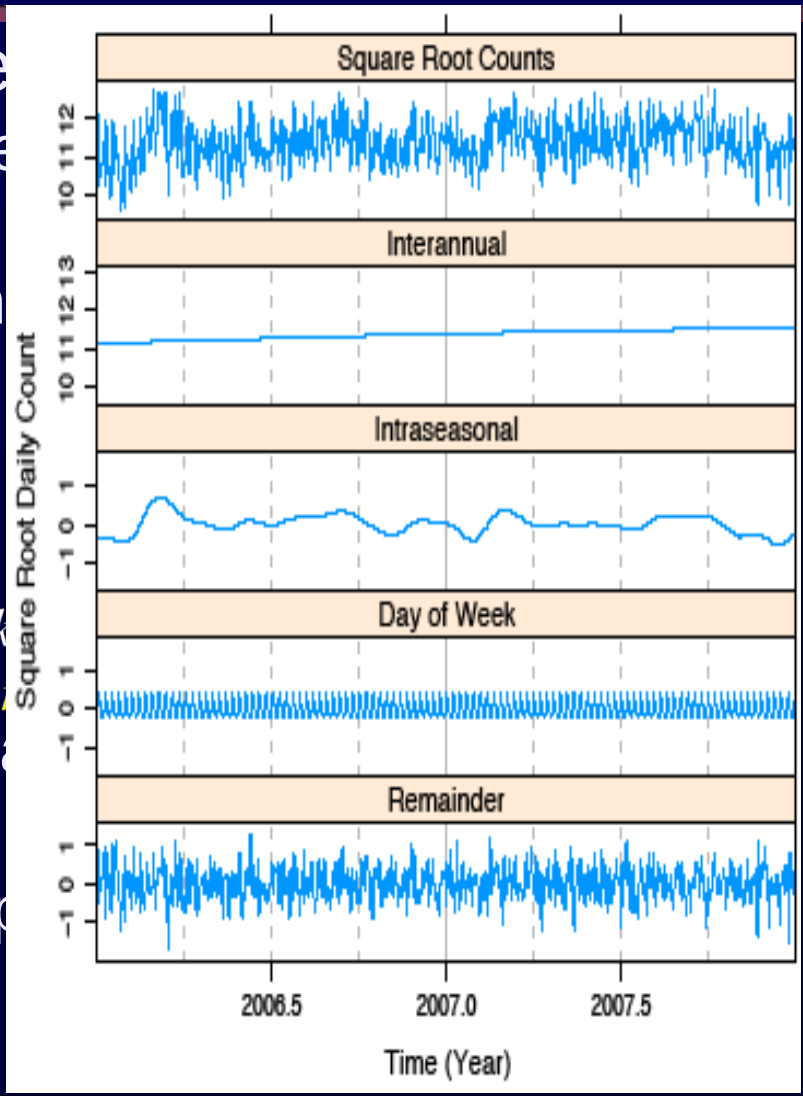
Event/Anomaly Detection - STL

- Temporal trend analysis and pre
  - Seasonal Trend decomposition base
    weighted regression) (STL)
  - Time series data viewed as the sum

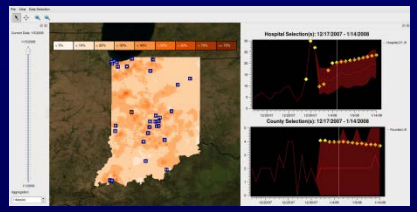
$$Y = T$$

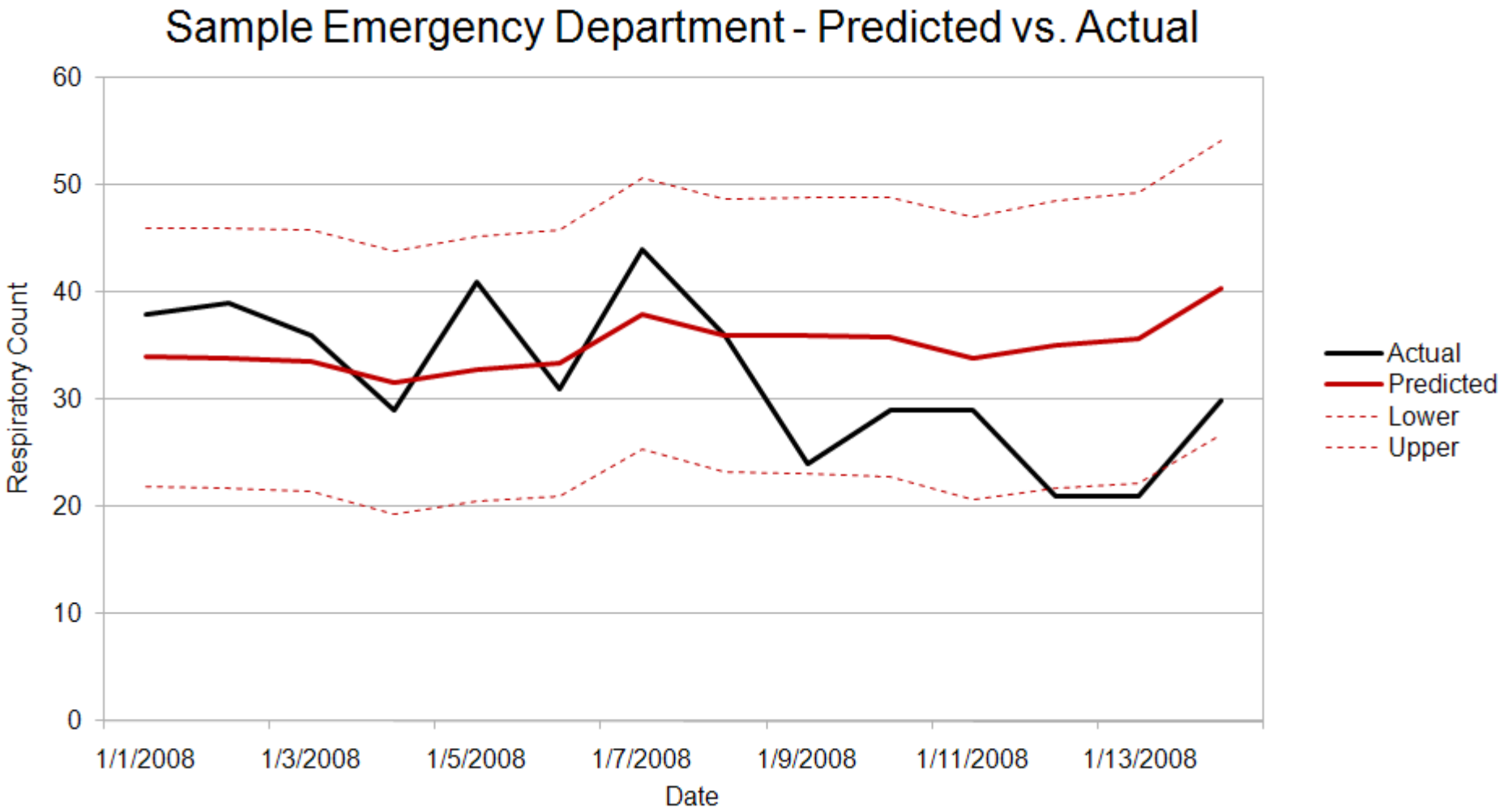
Y : *original time series*
T : *trend component*
S : *seasonal (daily/week*
R : *remainder compone*

  - Large value of *R* indicates substanti
  - Filter out the noise component
  - If value varies by 2 SD from STL exp



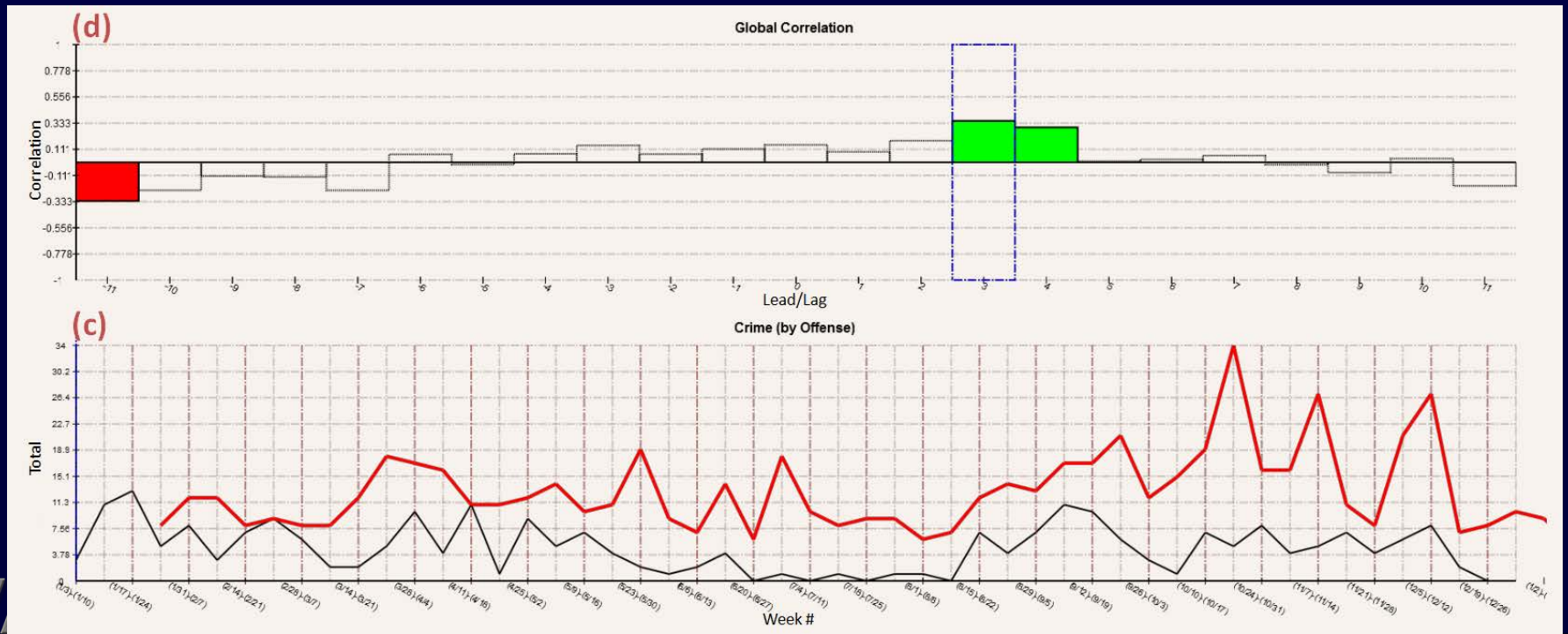
Predictive Visual Analytics



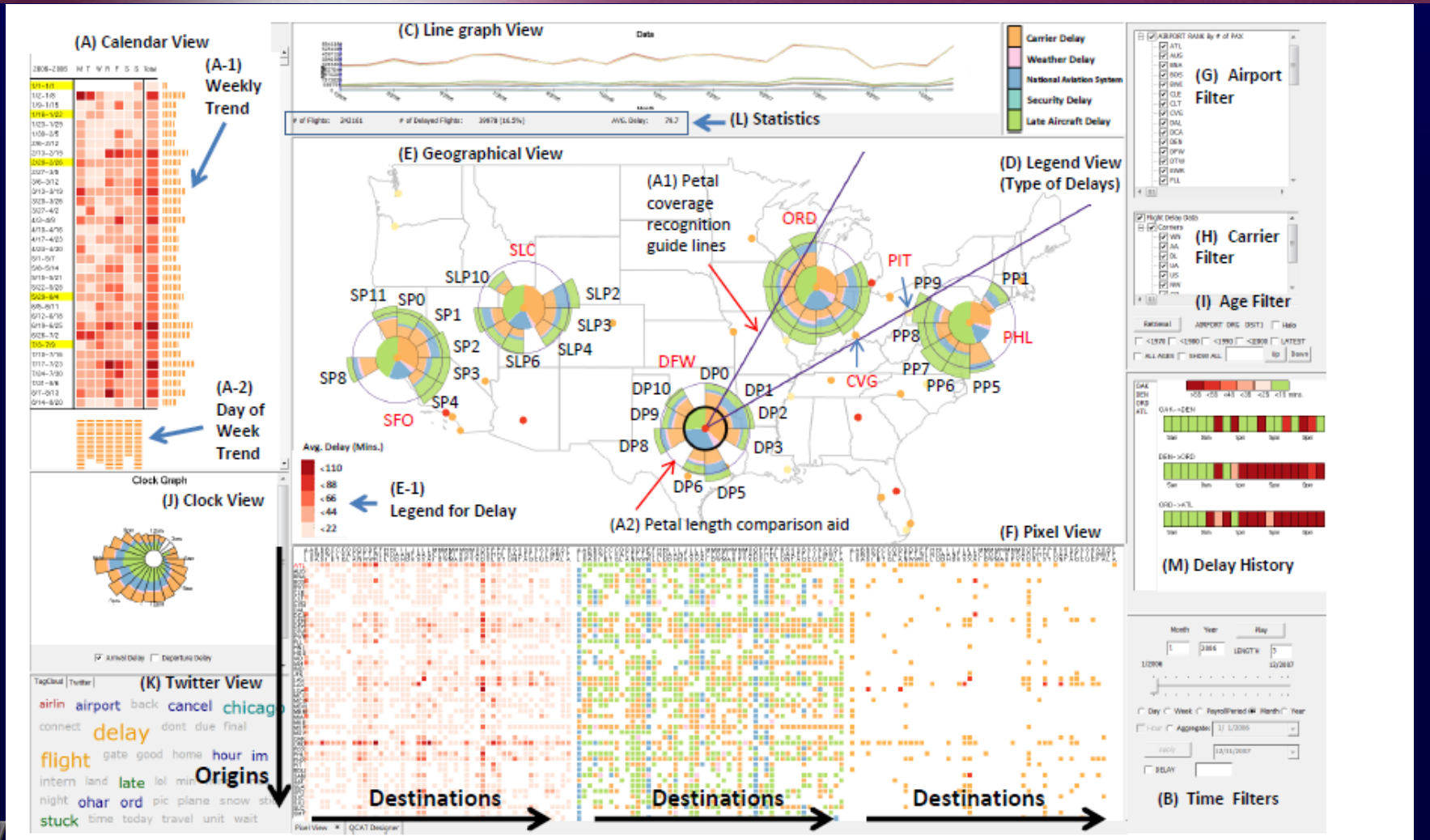


Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space



Multivariate Network Analytics with Information Theory Anomaly Detection



The Scale Problem

- Data scale:
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues
- User / problem scale:
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally

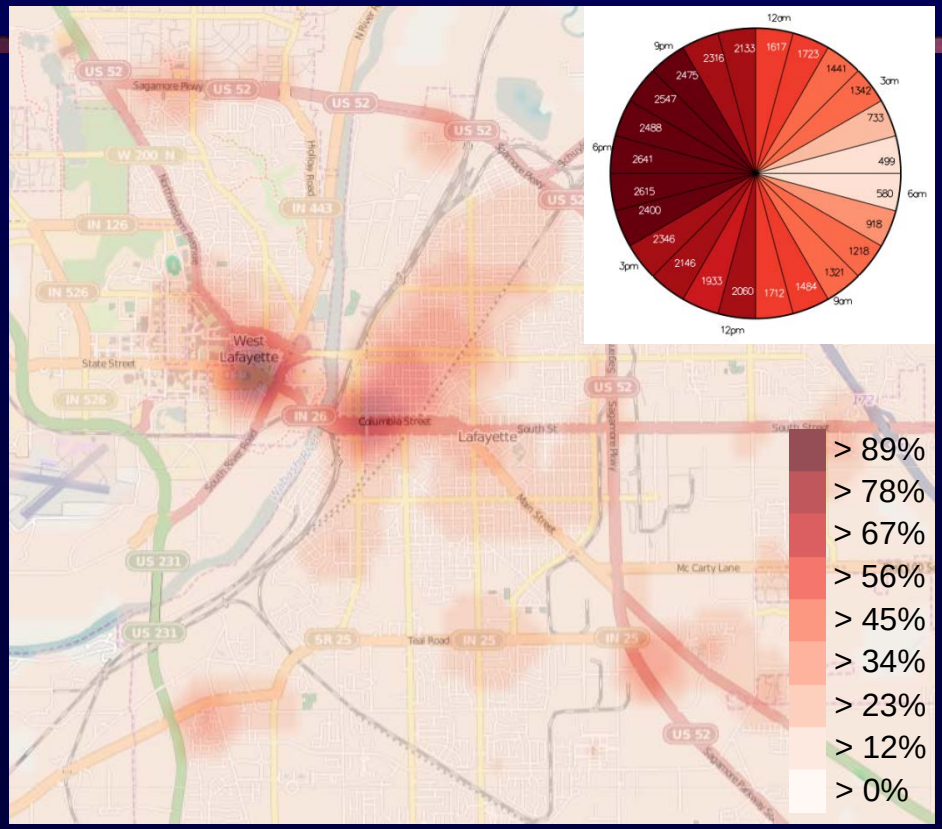
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness[1]
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: **Empower the casual expert to be more effective**

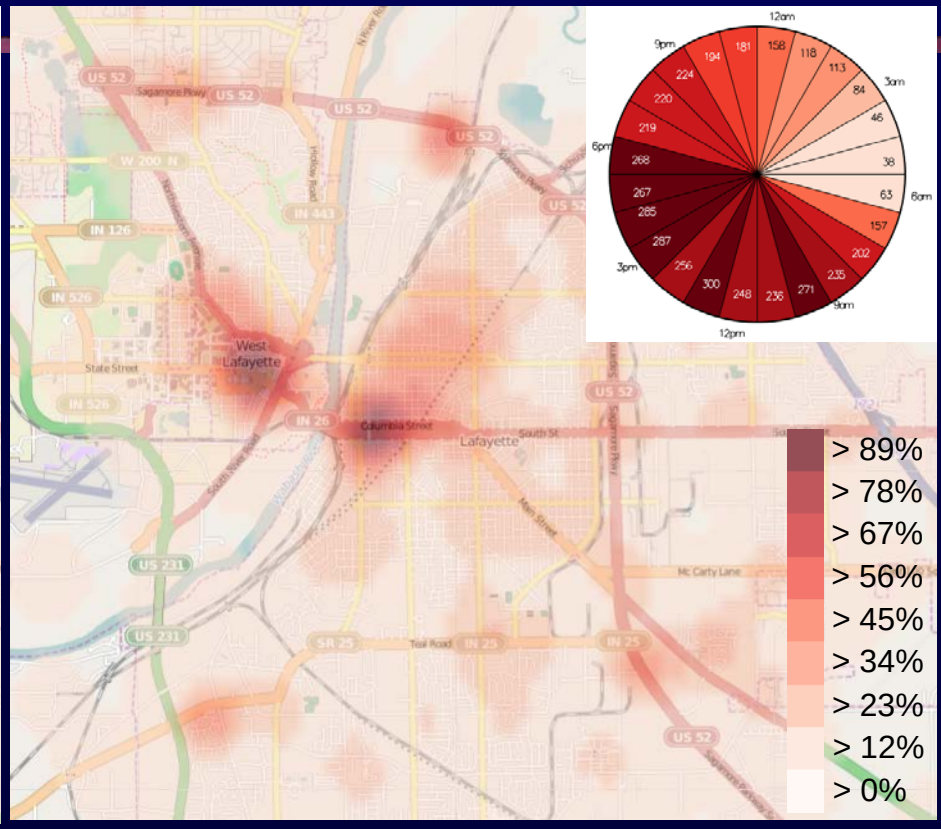
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data



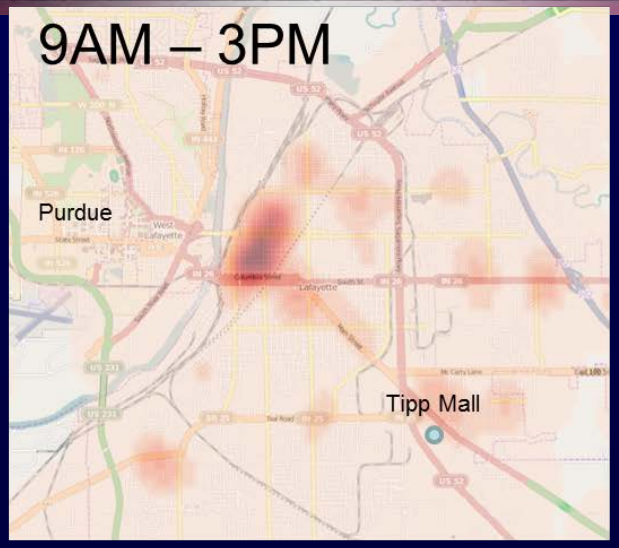
Past 2 Years
3/11/2012 – 3/10/2014

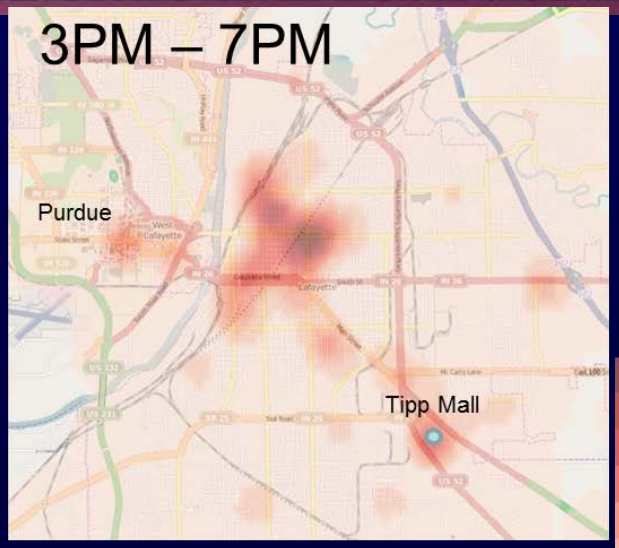


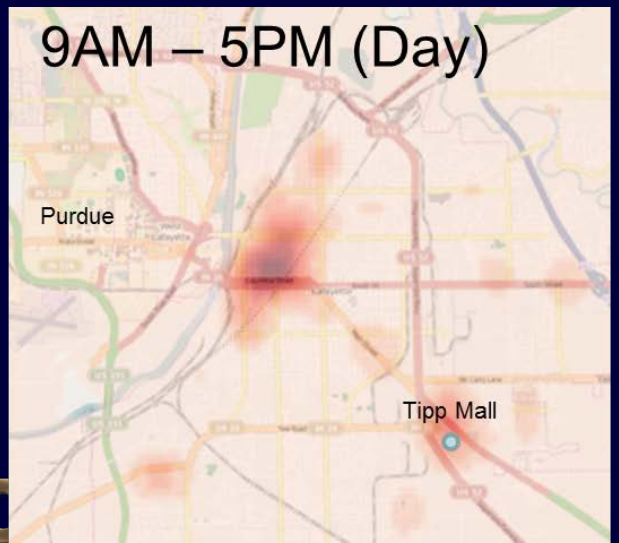
March + Tuesday Filter
(Past 10 Years)

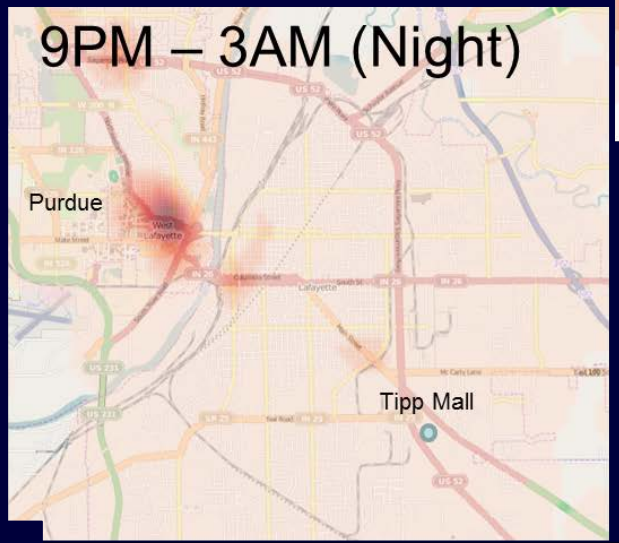
VACCINE

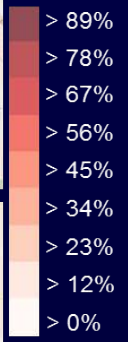
Prediction During Certain Hours In The Day





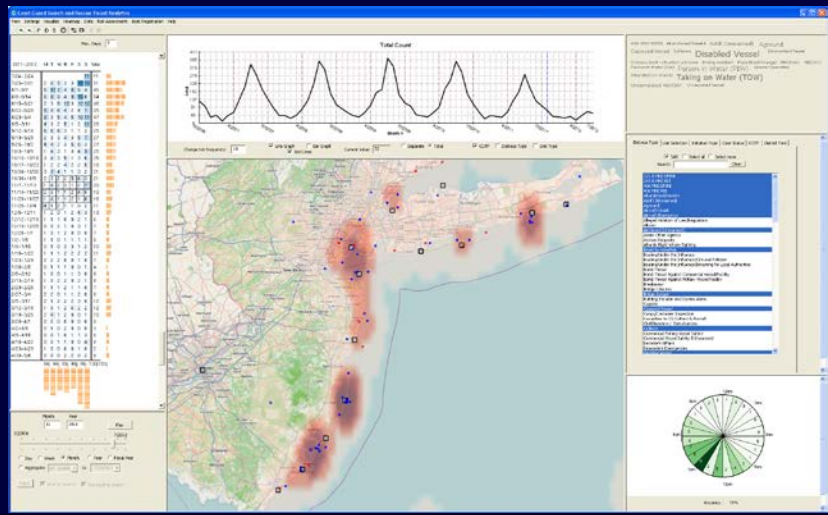




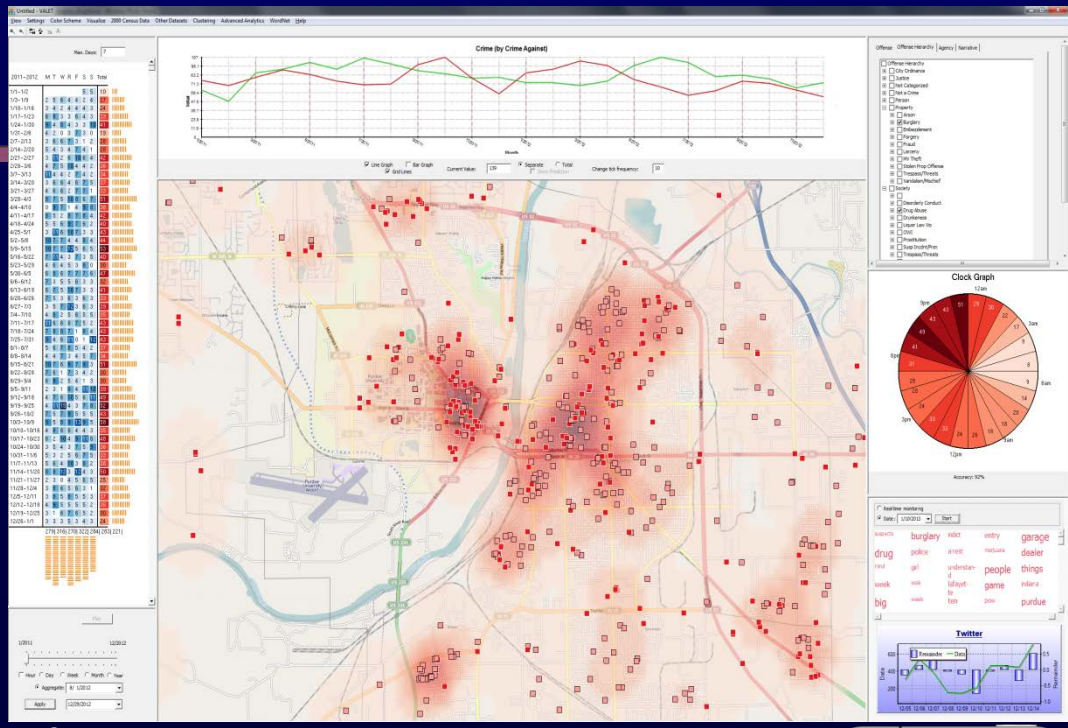


Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

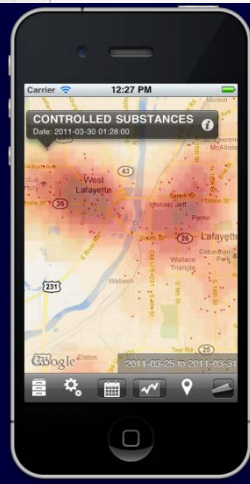
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

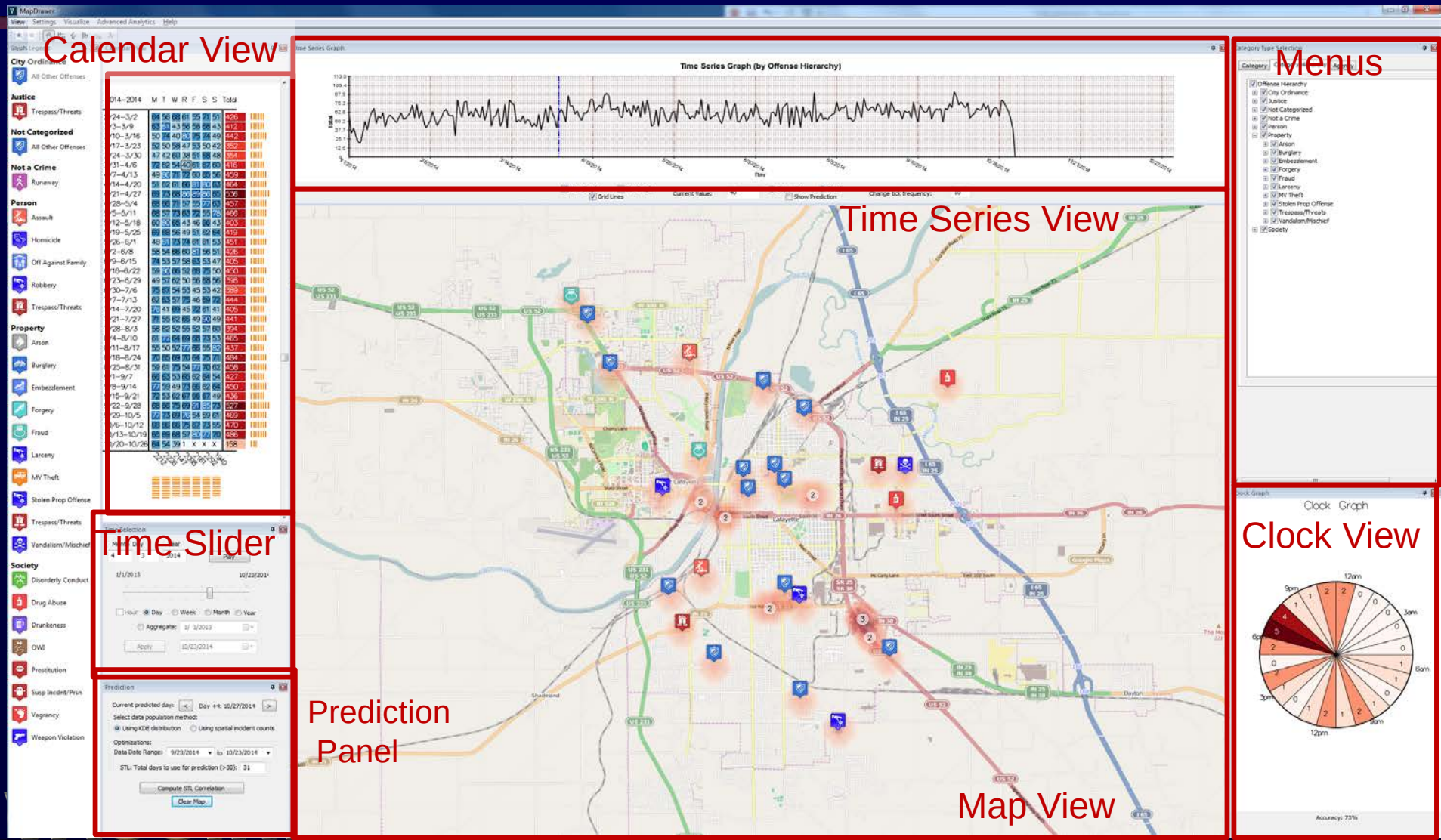
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

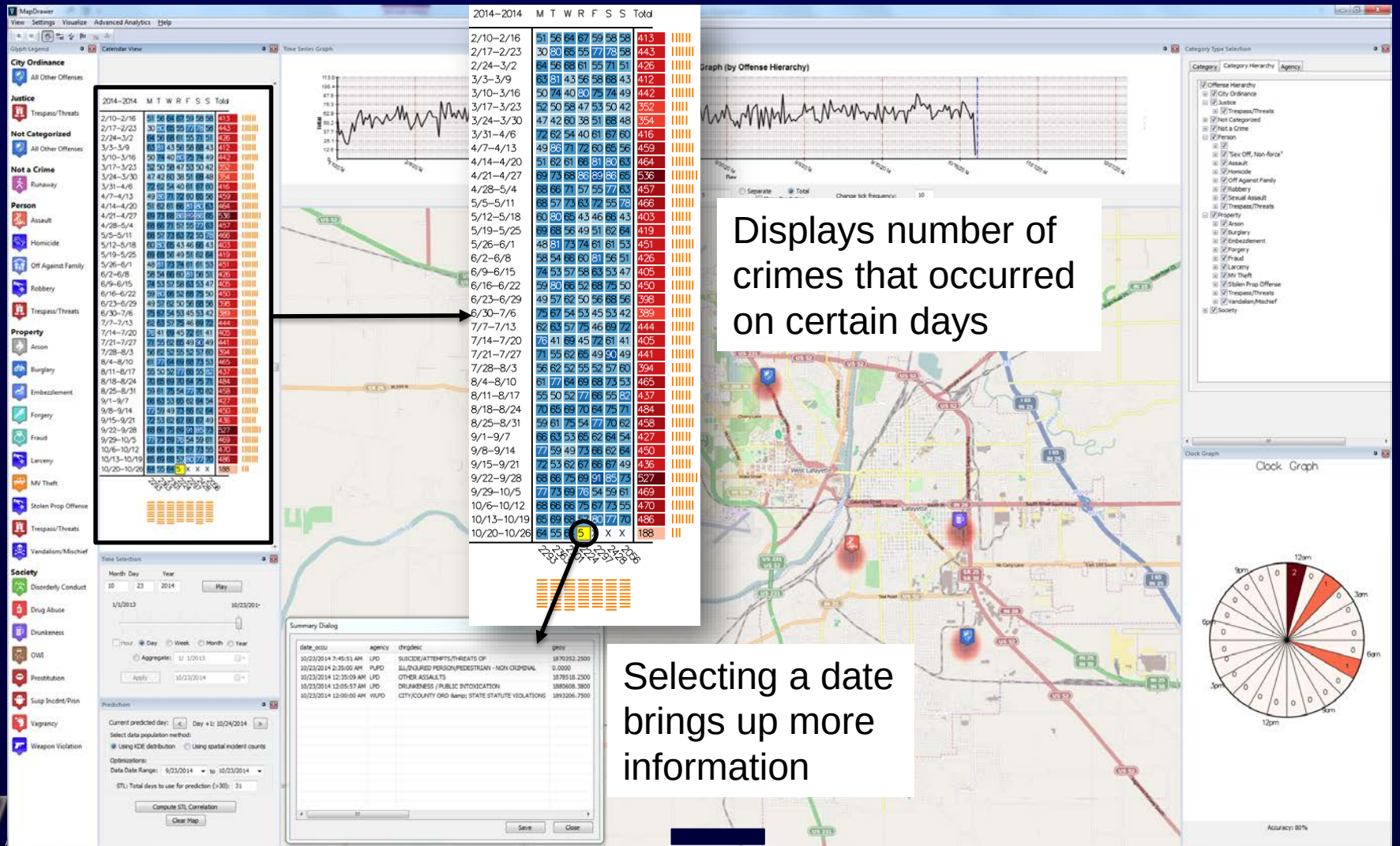


VACCINE

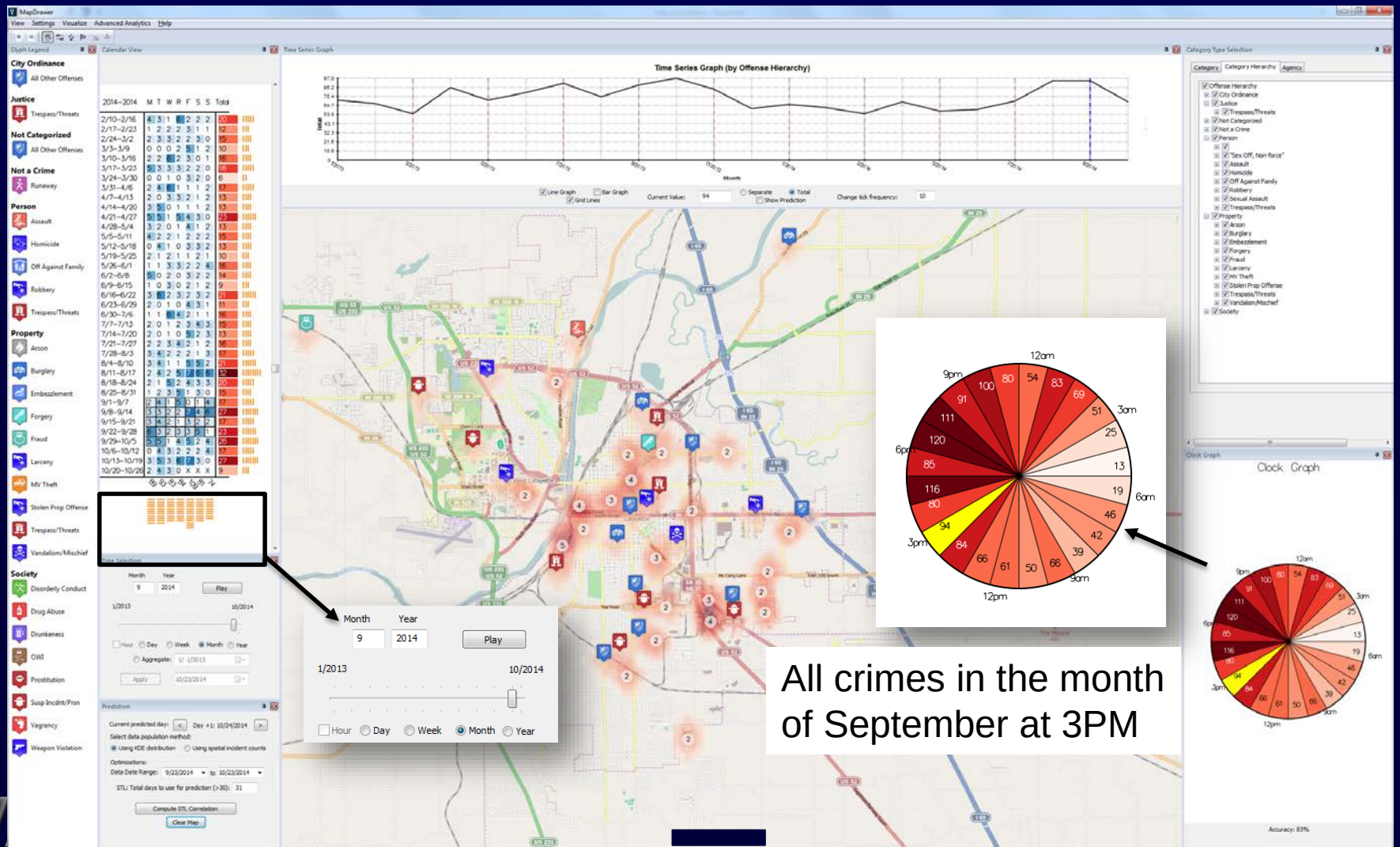
VALET Layout



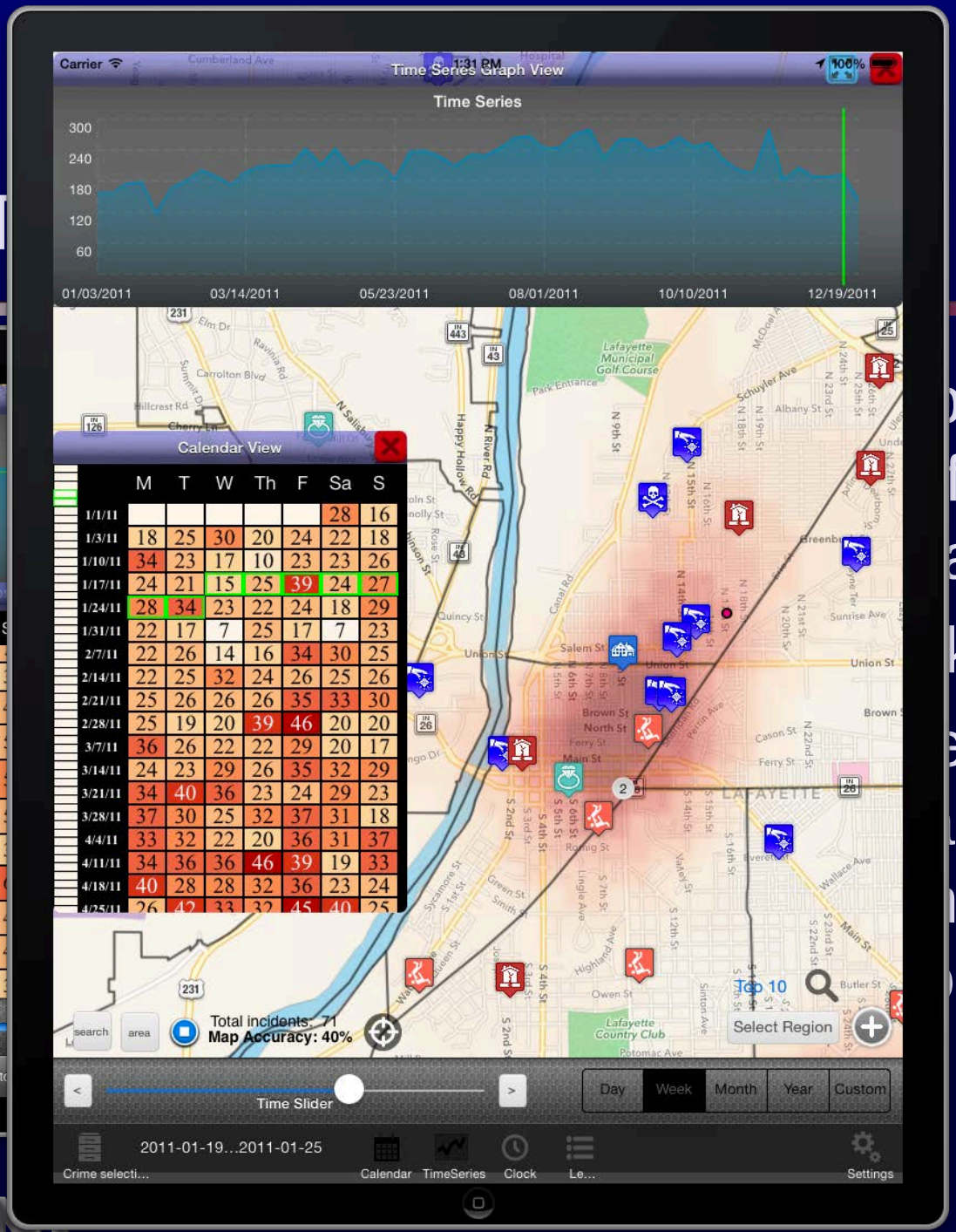
Calendar Display



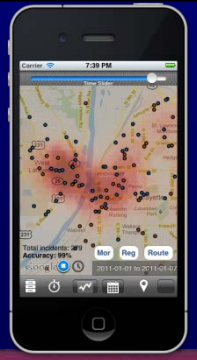
Crimes During Certain Hours In A Month

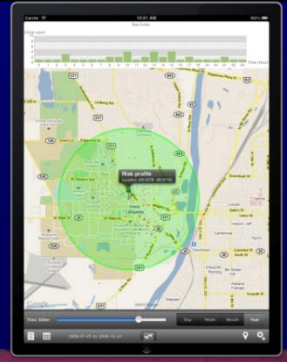


iVALET









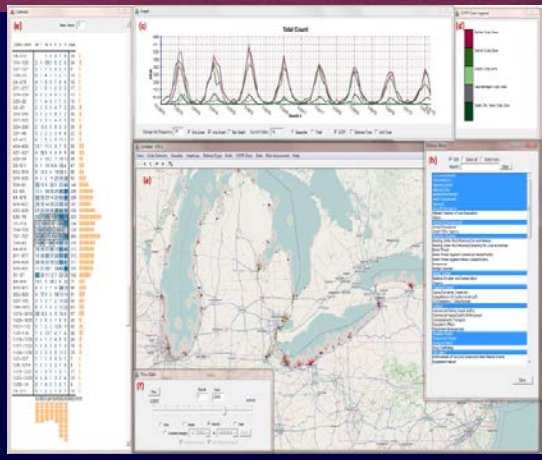


...lore criminal,
...ffic and civil
...a on-the-go
...k assessment
...e current
...tial +
...poral context
... analysis

VACCINE

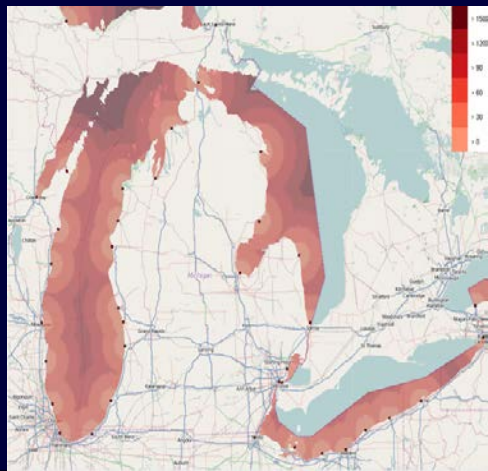
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

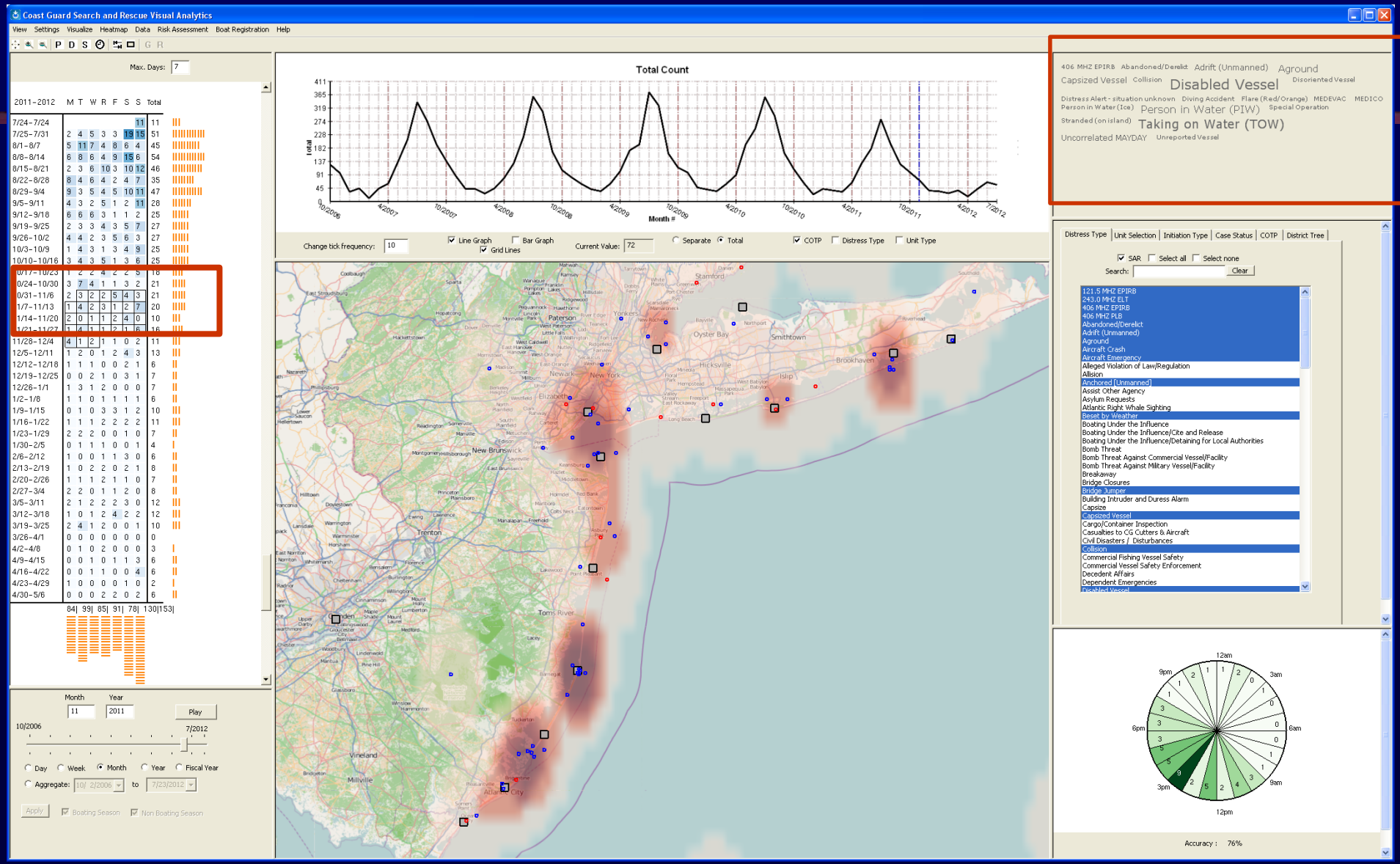
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



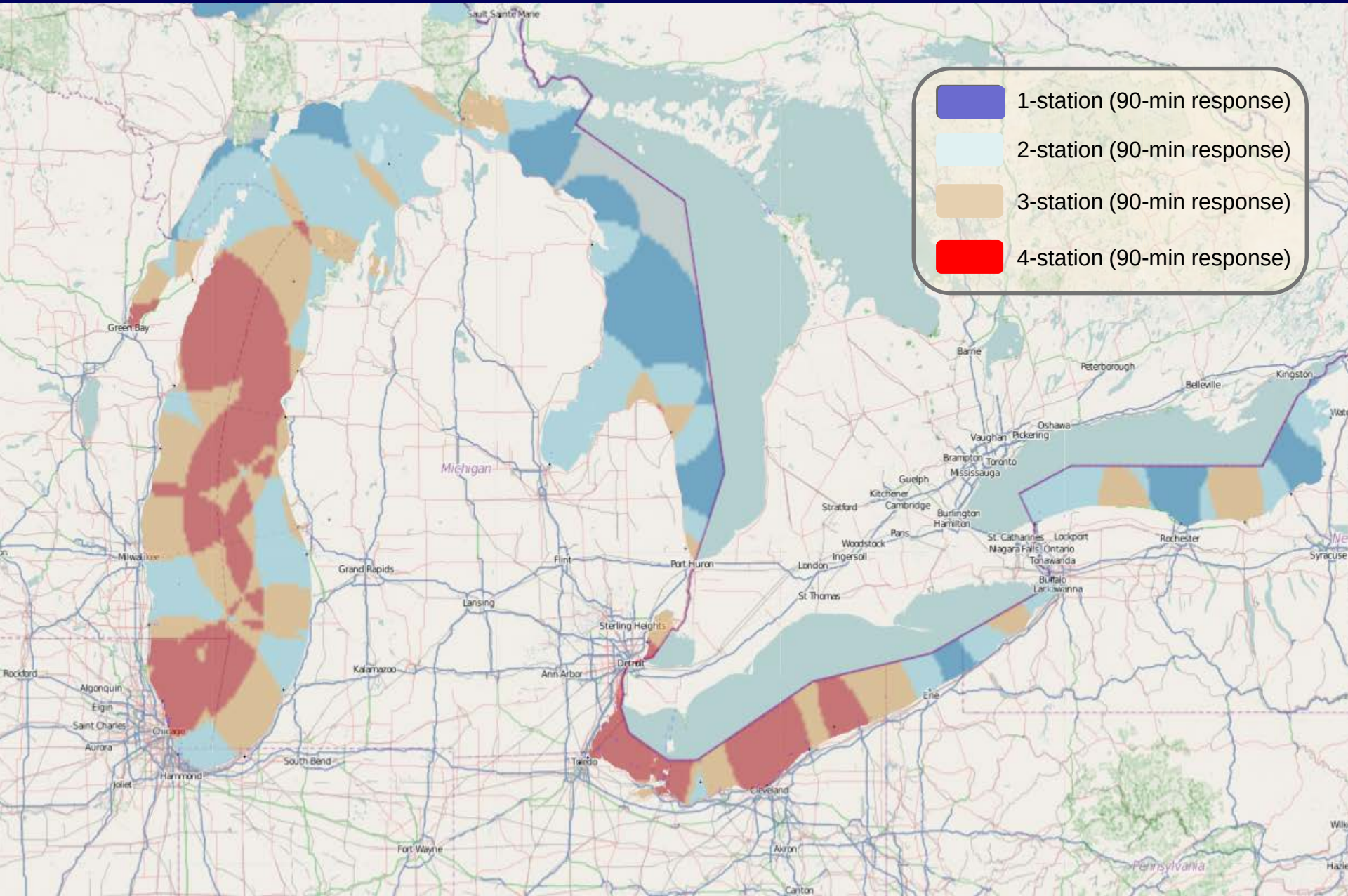
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

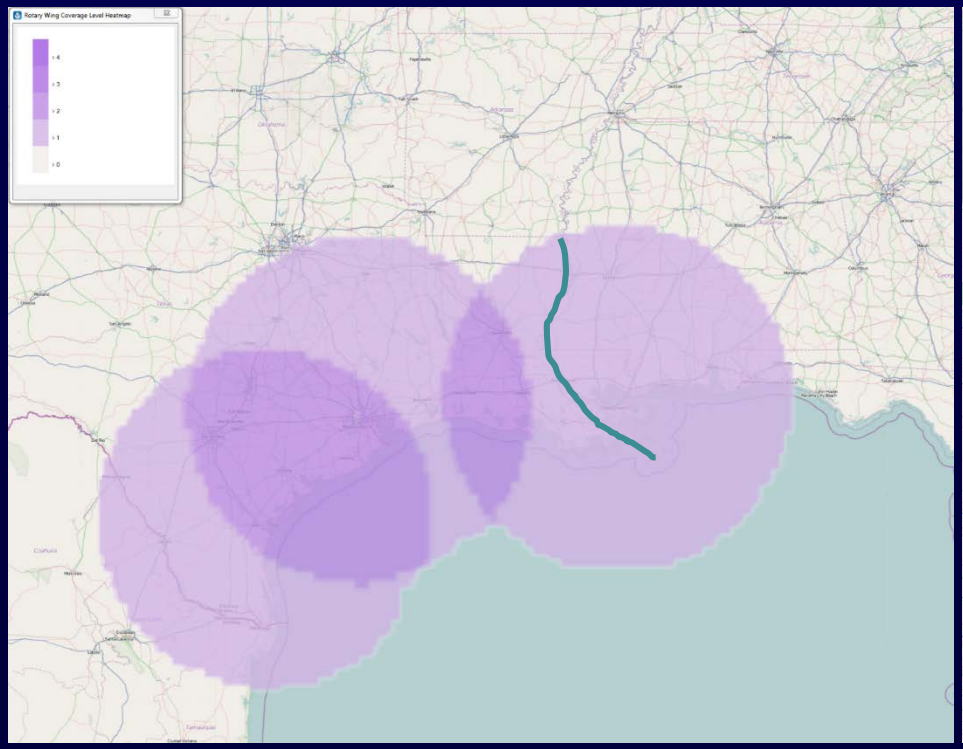


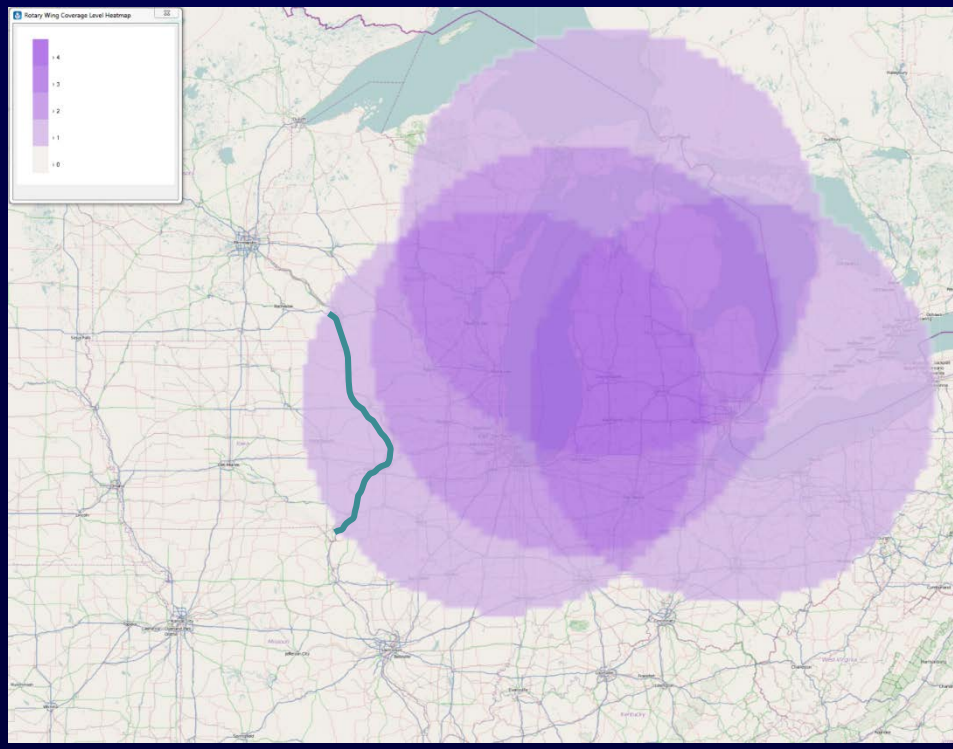
Response Efficiency – Potential Future Assets



Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

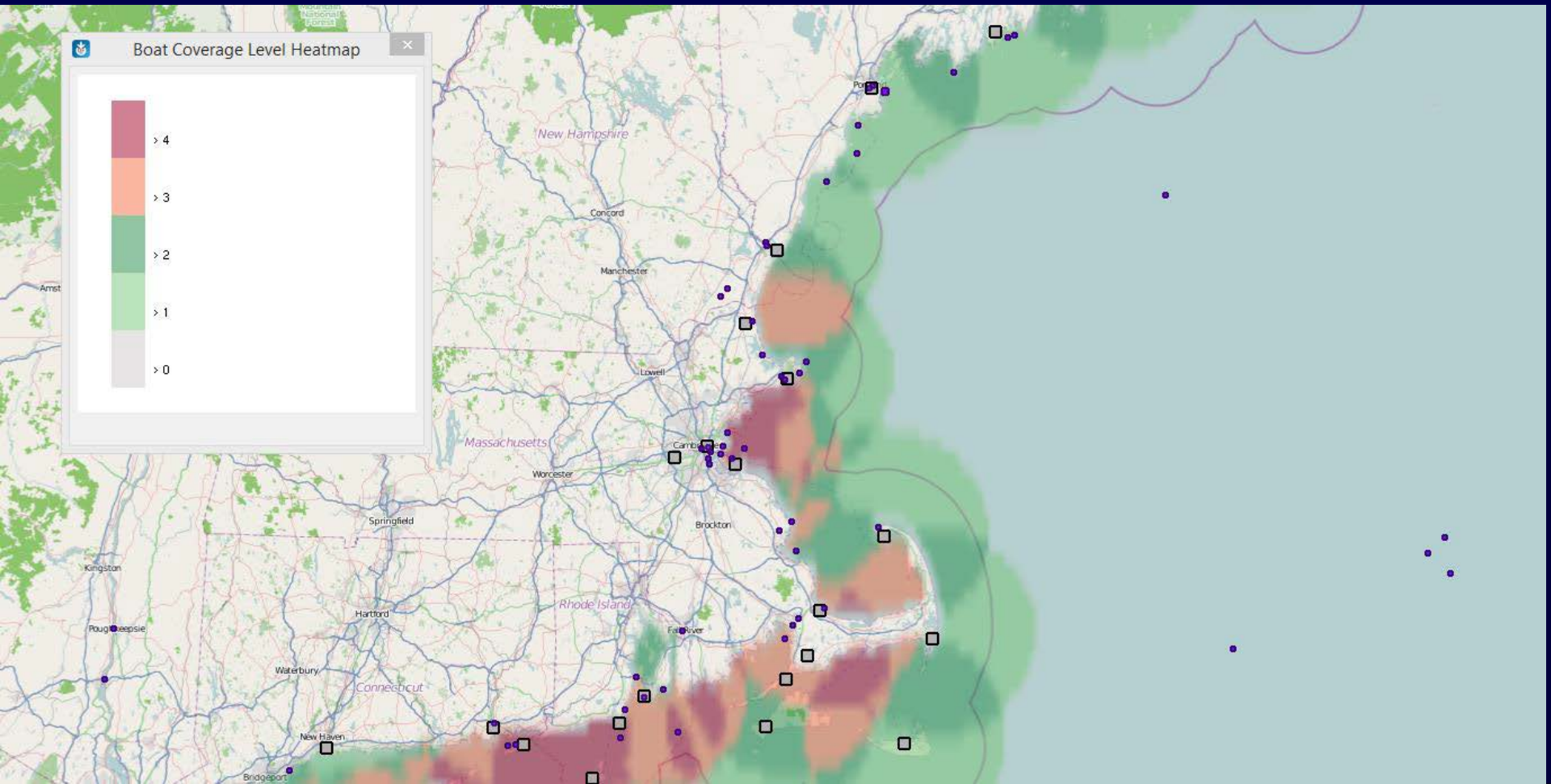




Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

VACCINE

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007

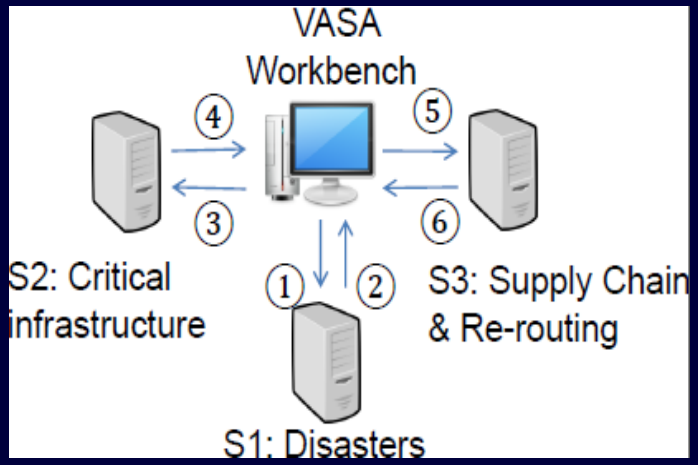


Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

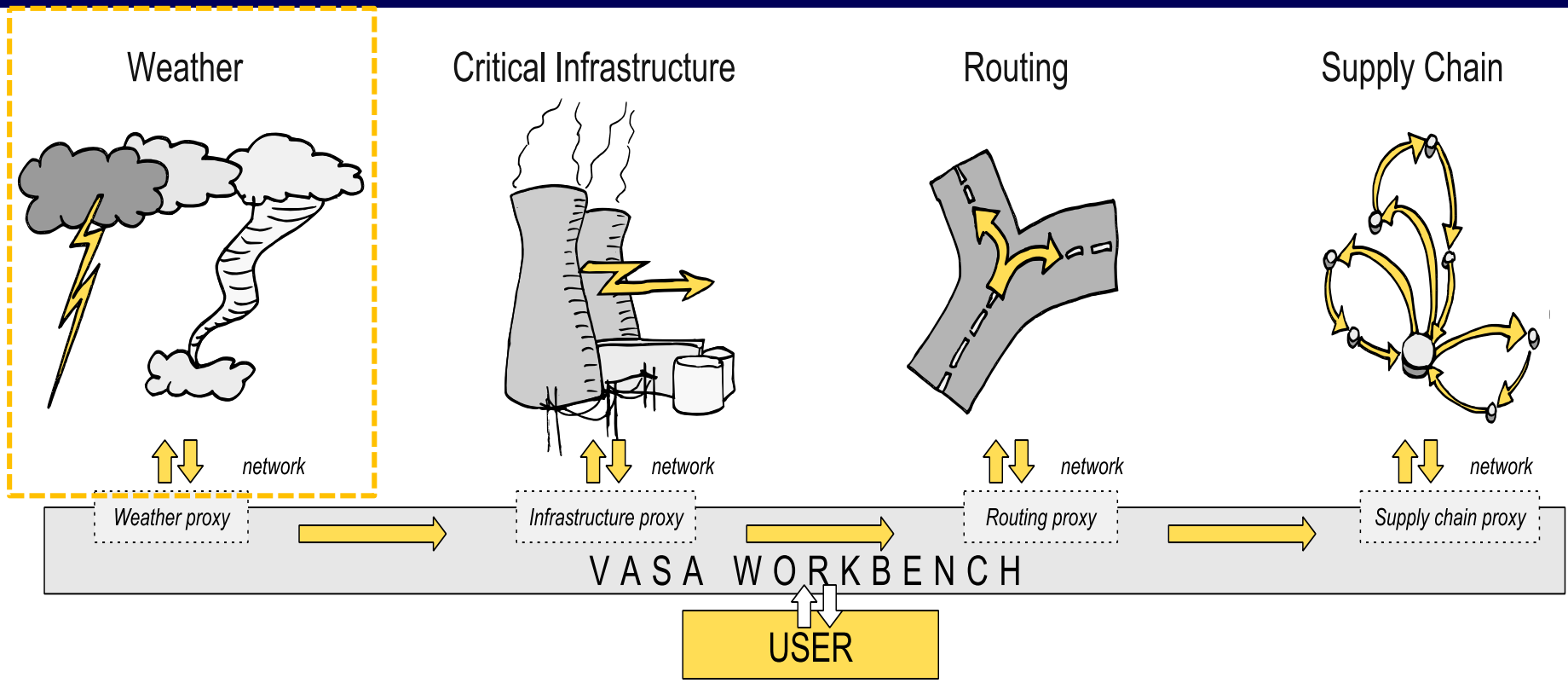


UNCC

U Minn

TACC

VASA COMPONENT:
CRITICAL INFRASTRUCTURE

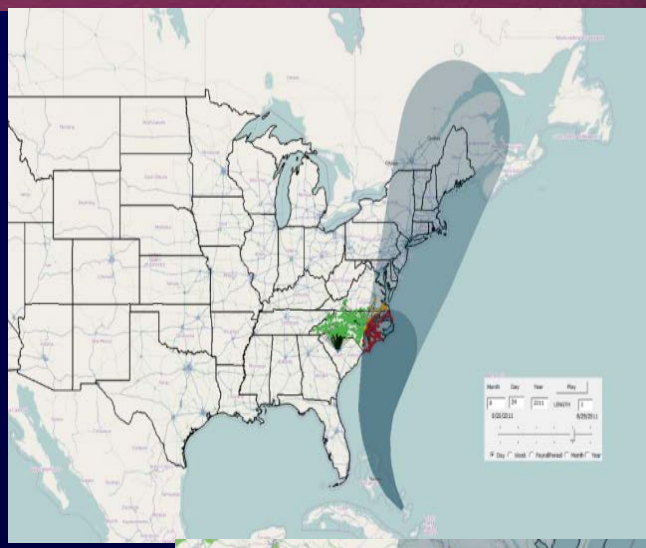


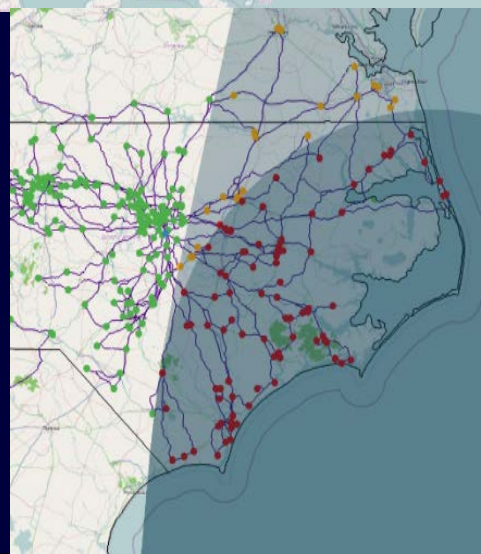
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models

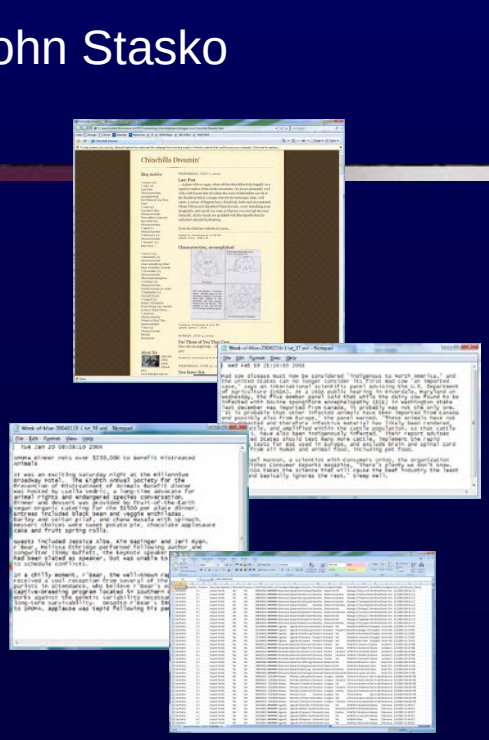




VACCINE

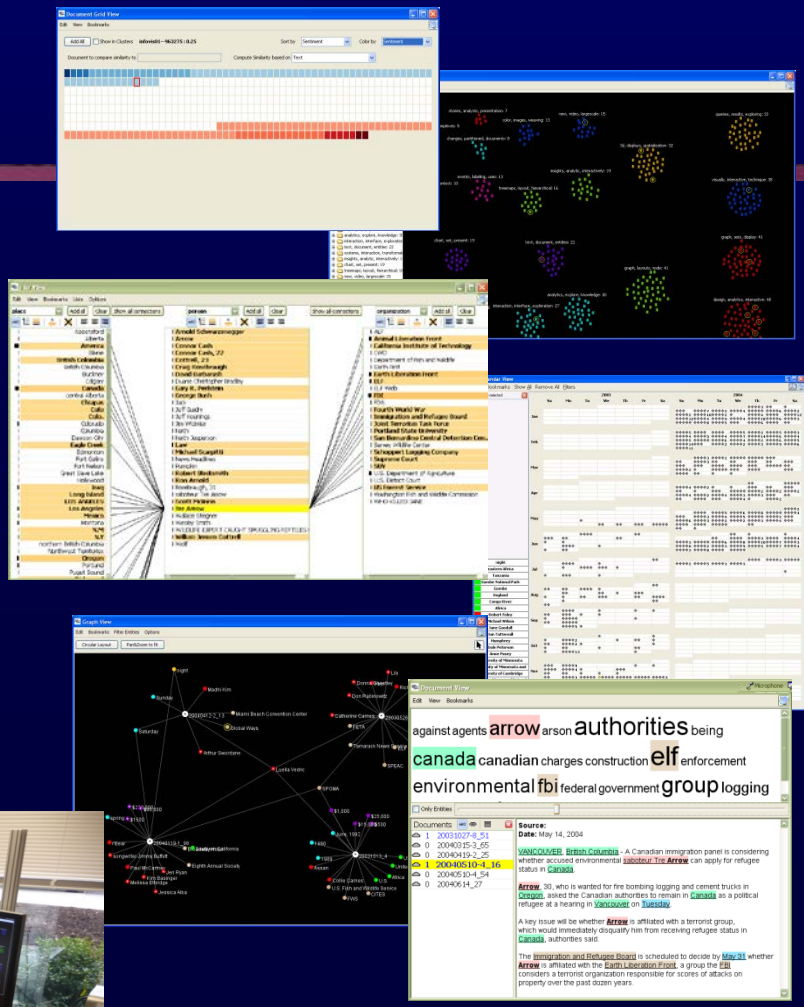
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



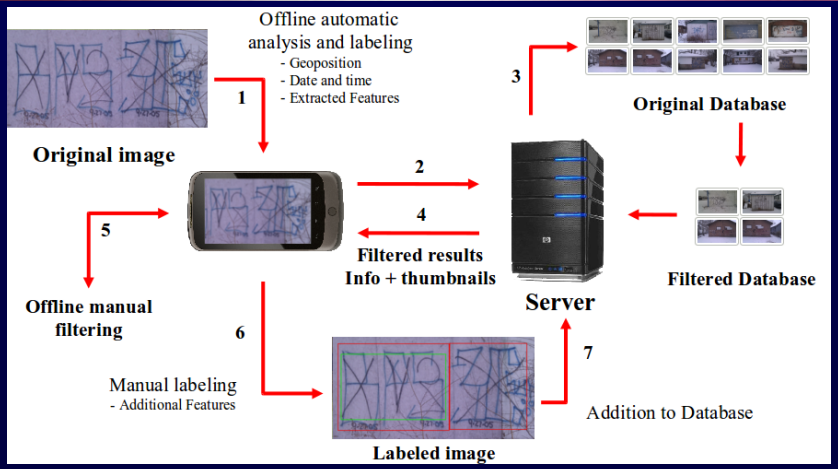
Documents in

Knowledge out





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

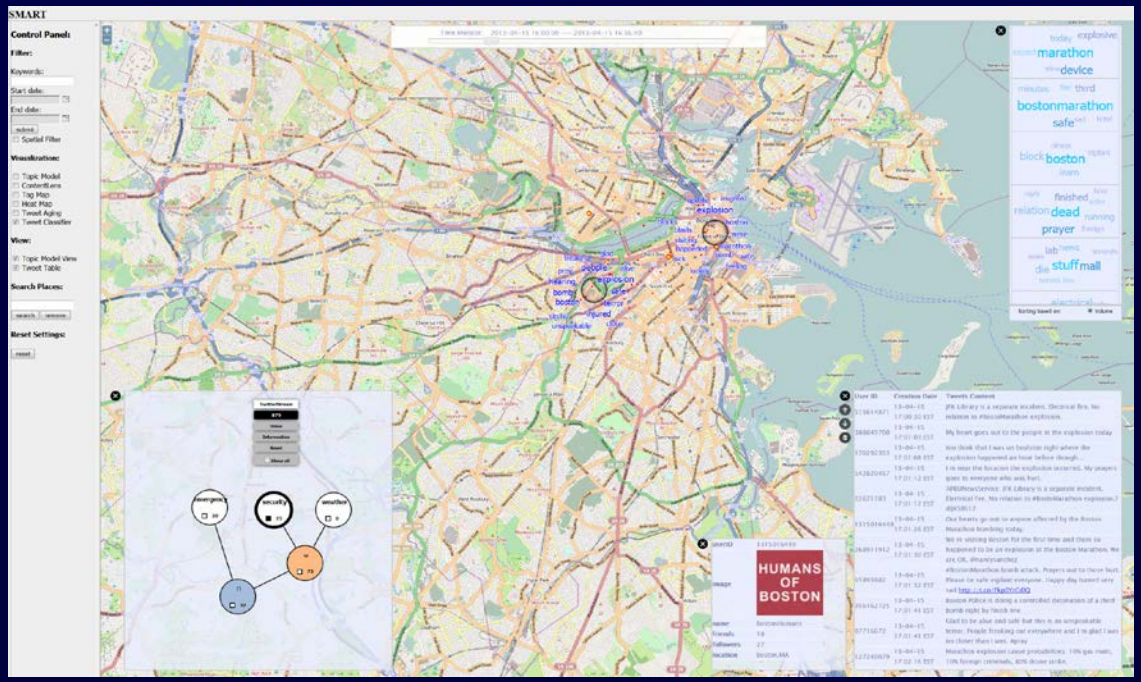
VACCINE

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

July 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

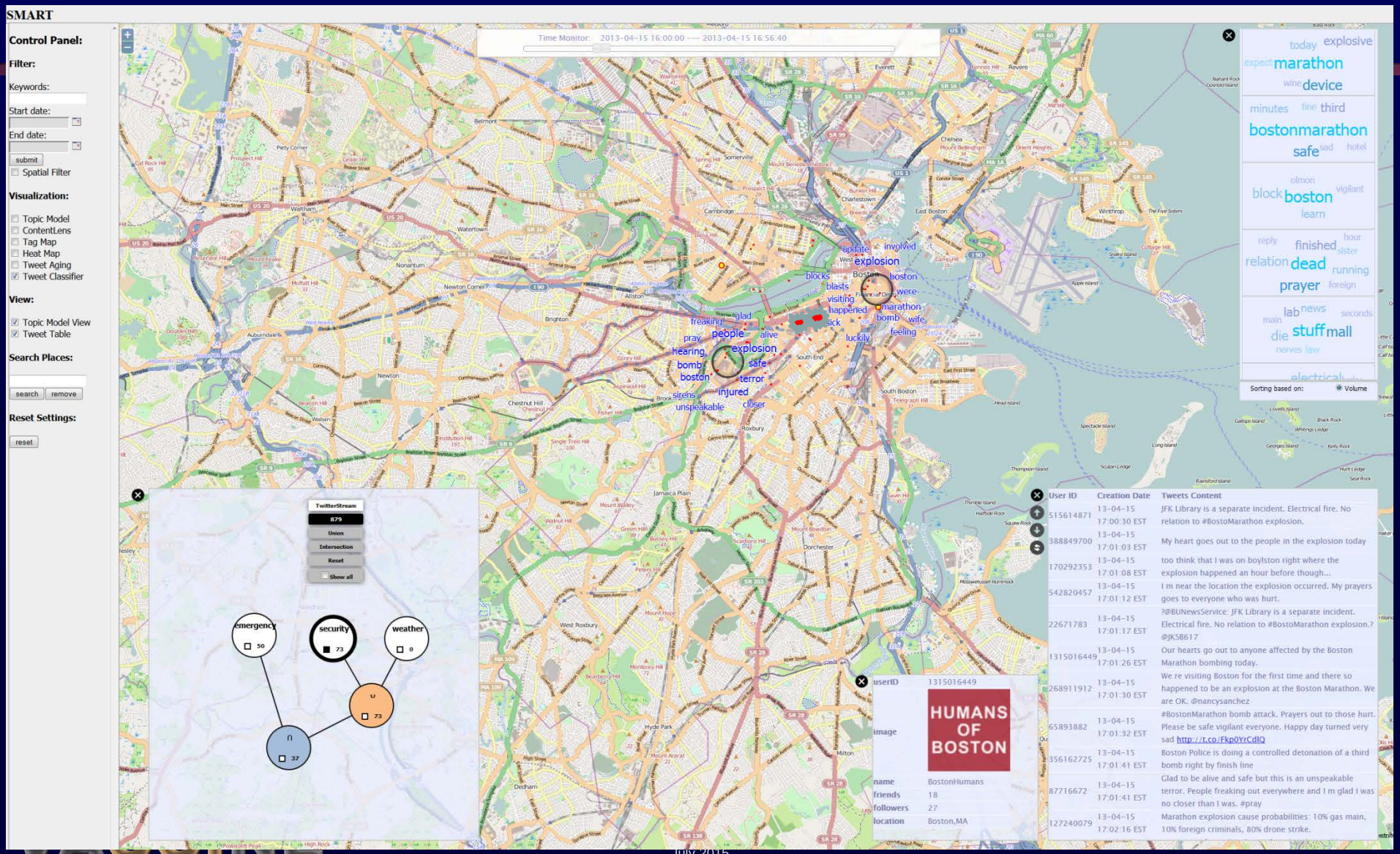
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

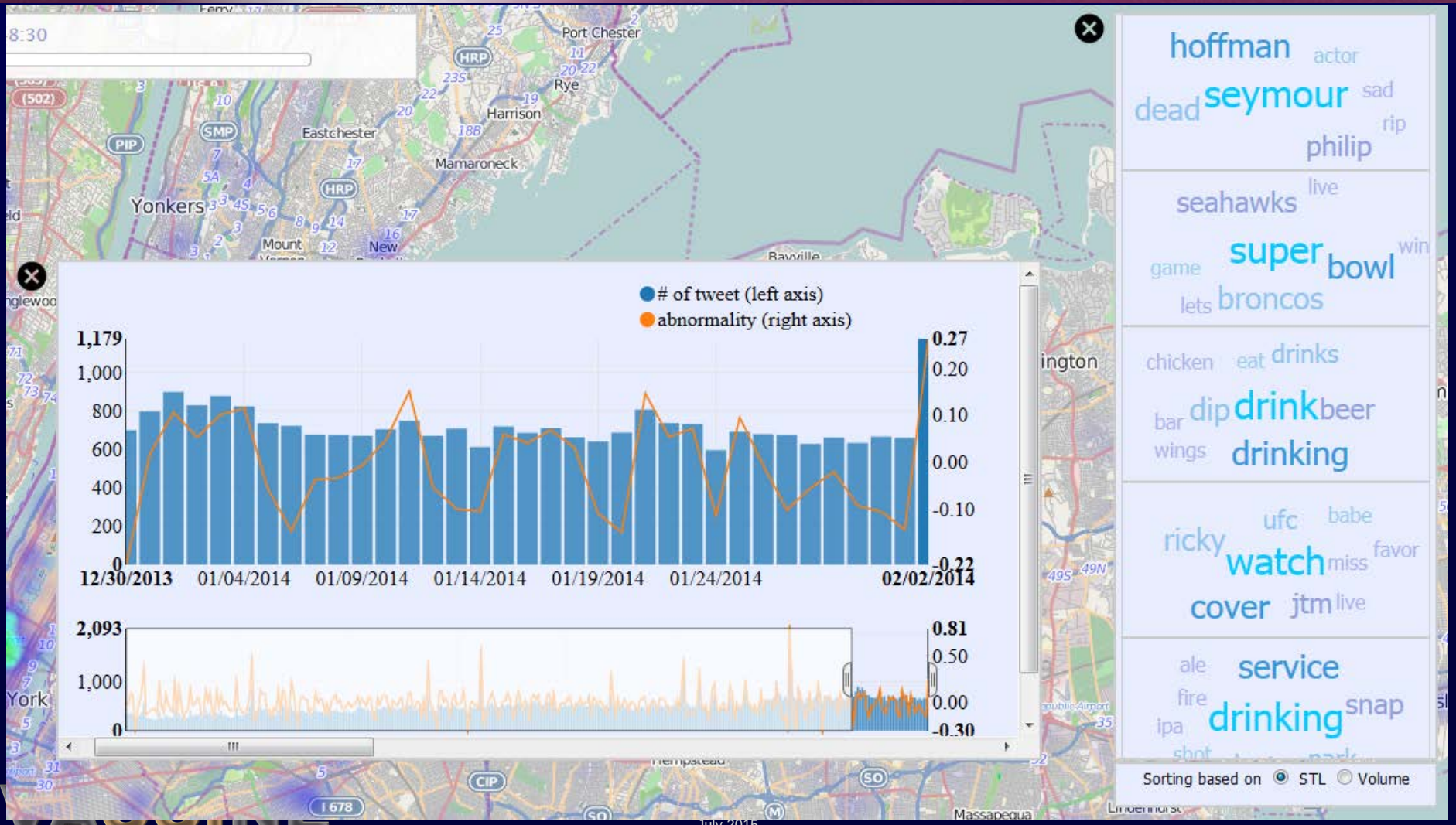


VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More



RETWEETS 6

FAVORITES 4

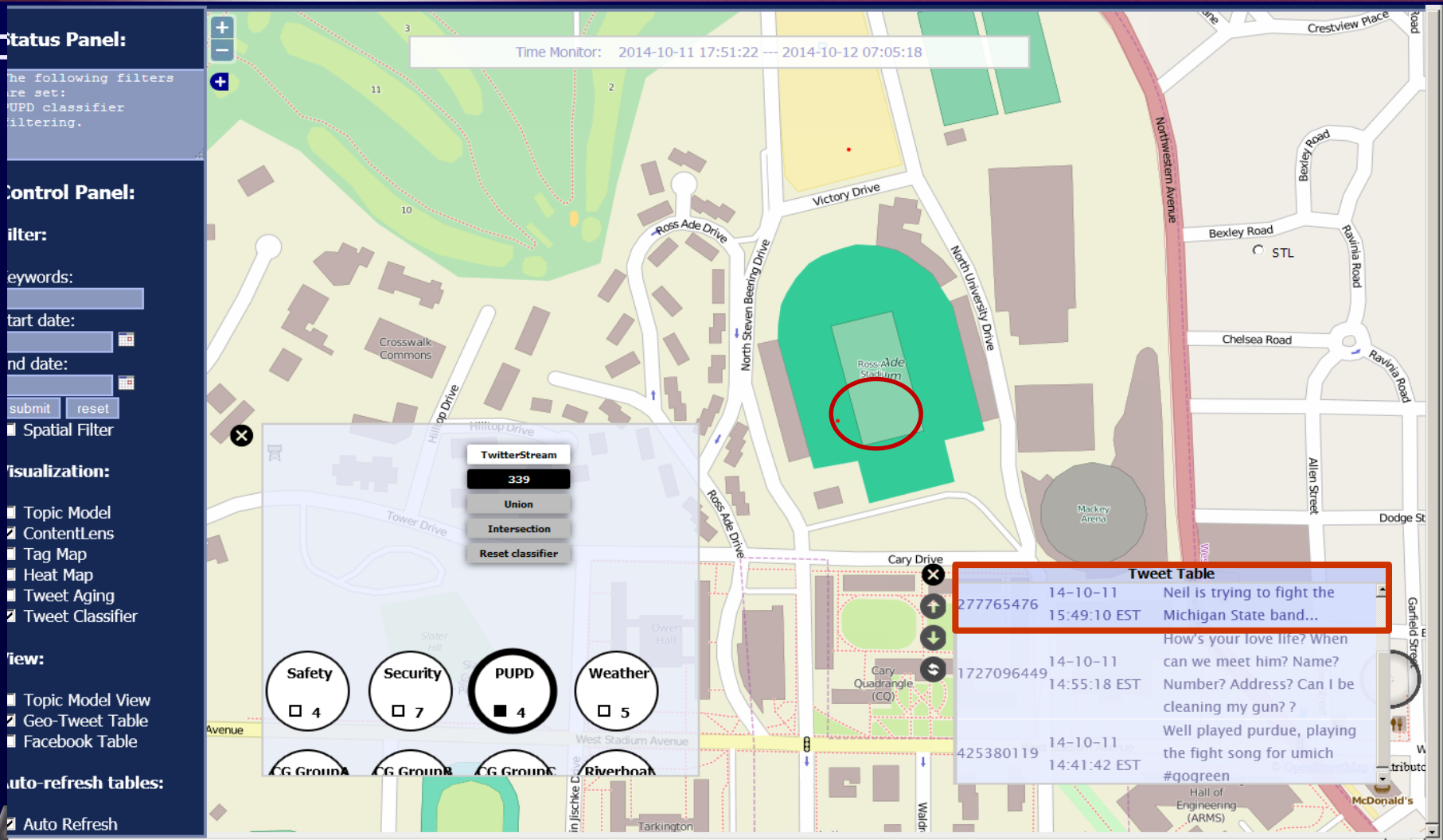
12:56 PM - 12 Oct 2014

Where Can It Be Helpful?

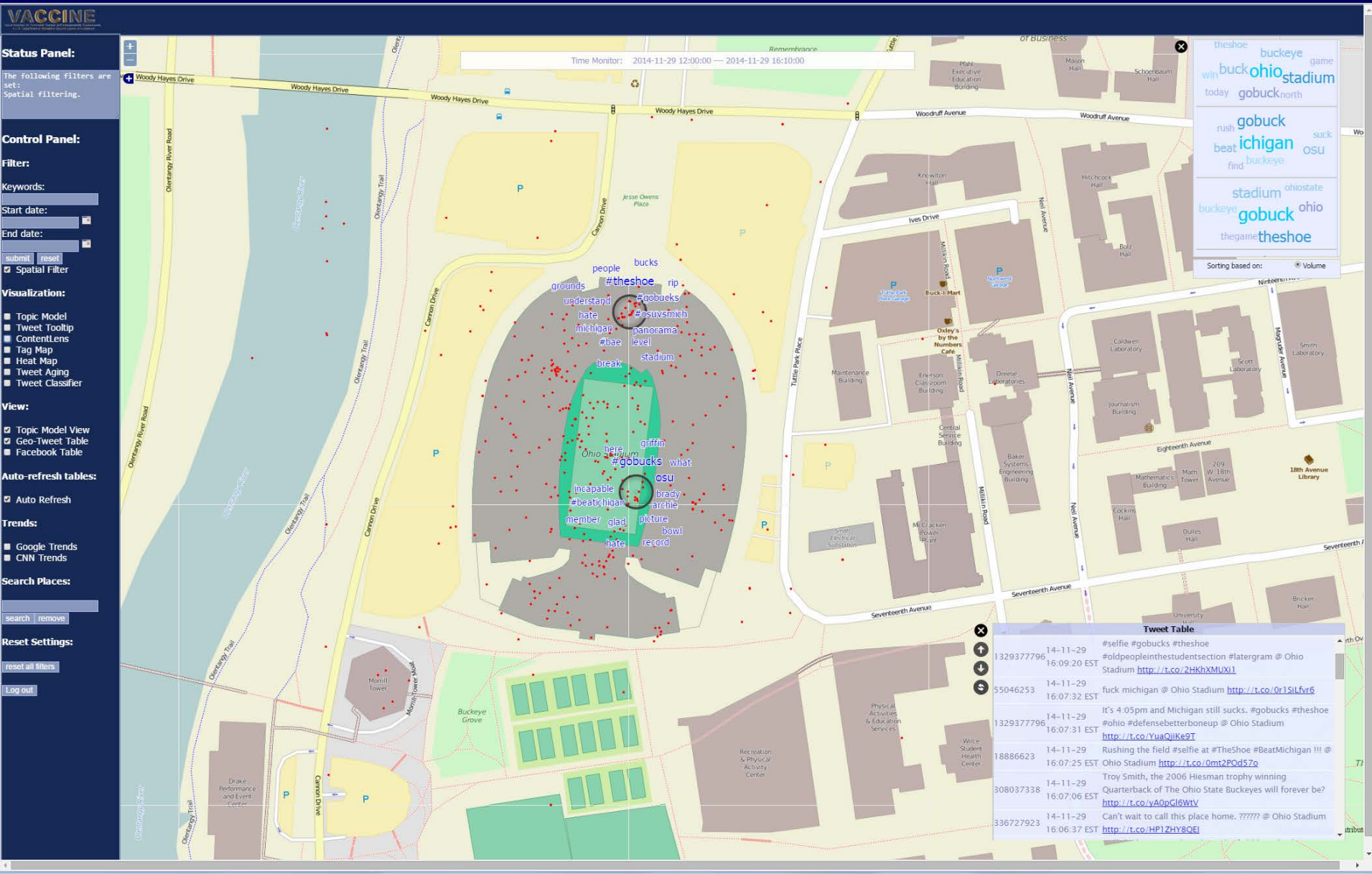
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



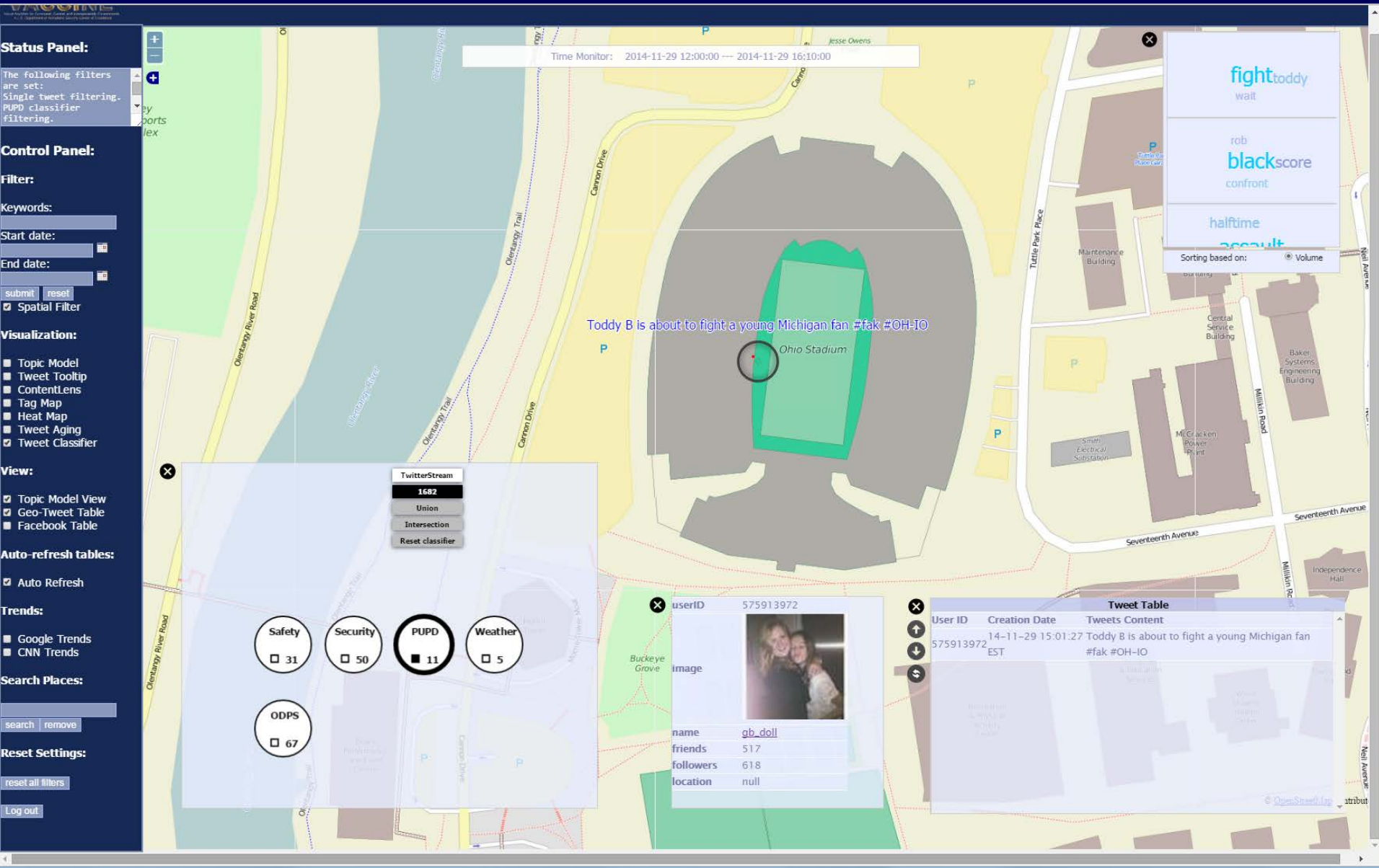
Example From Ohio State Football 11/29/14



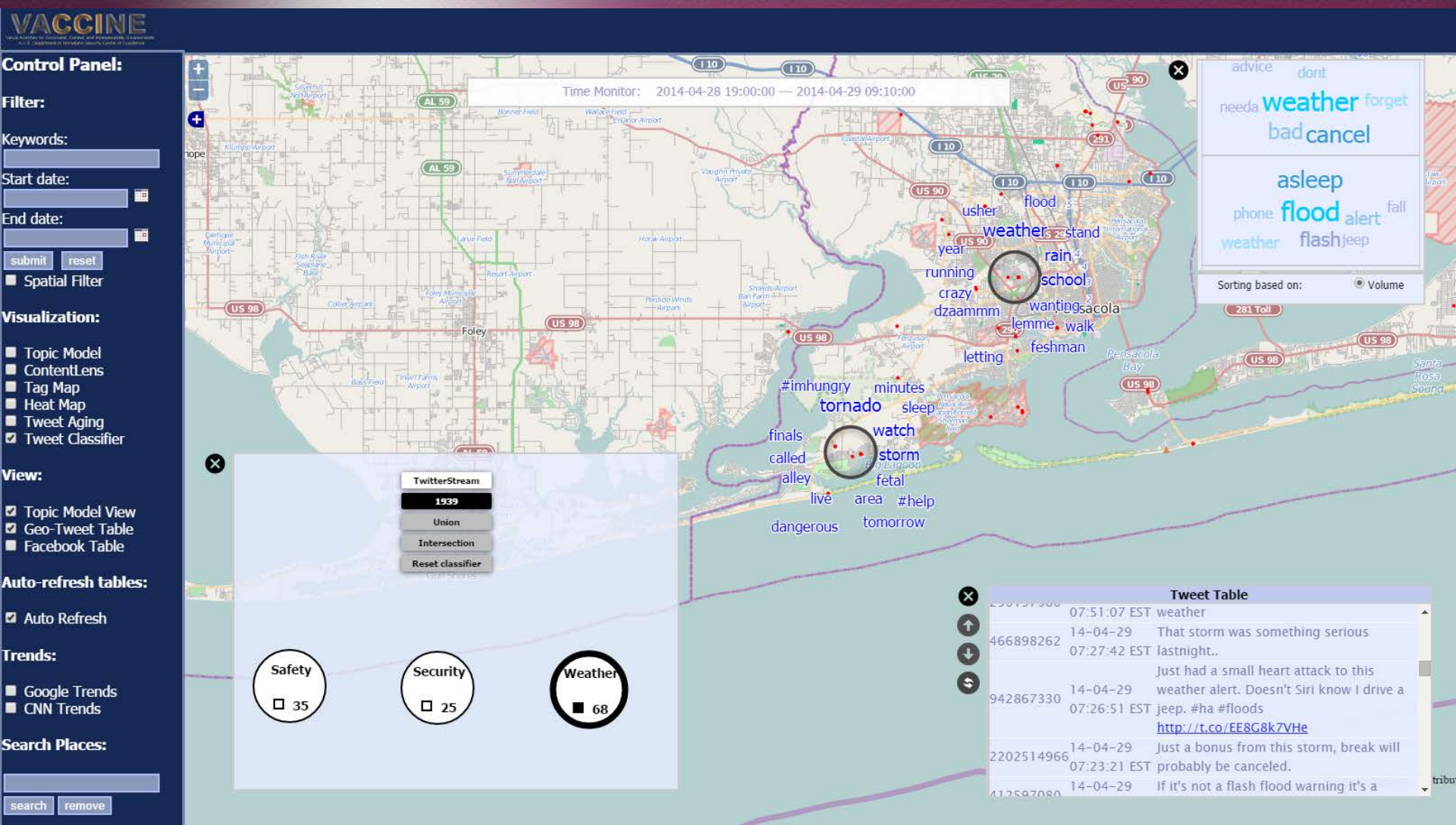
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



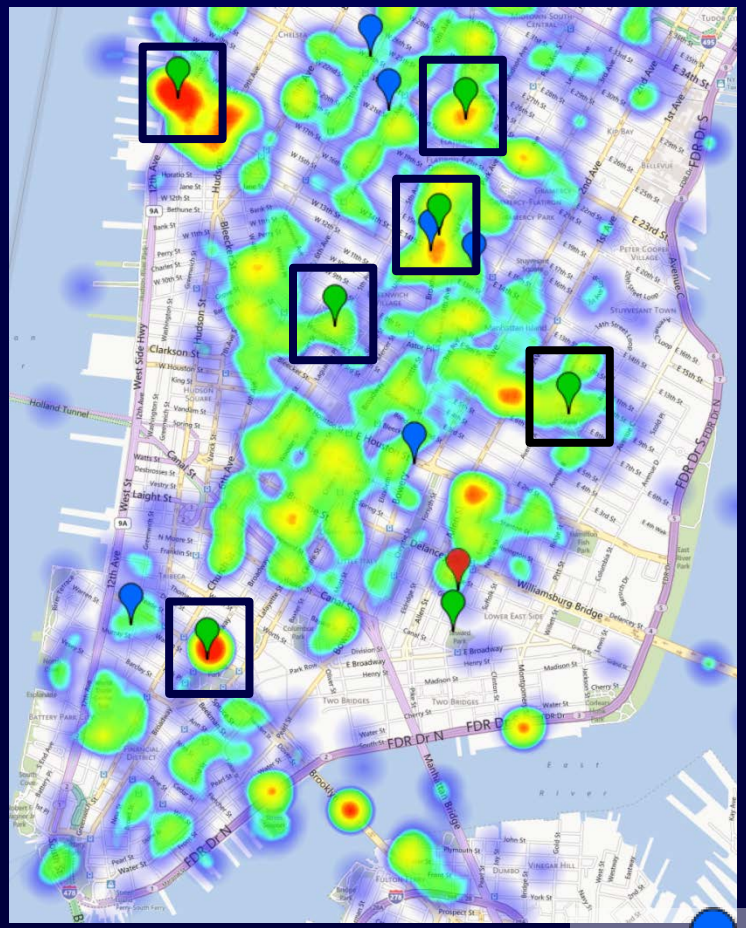
Where Can It Be Helpful?

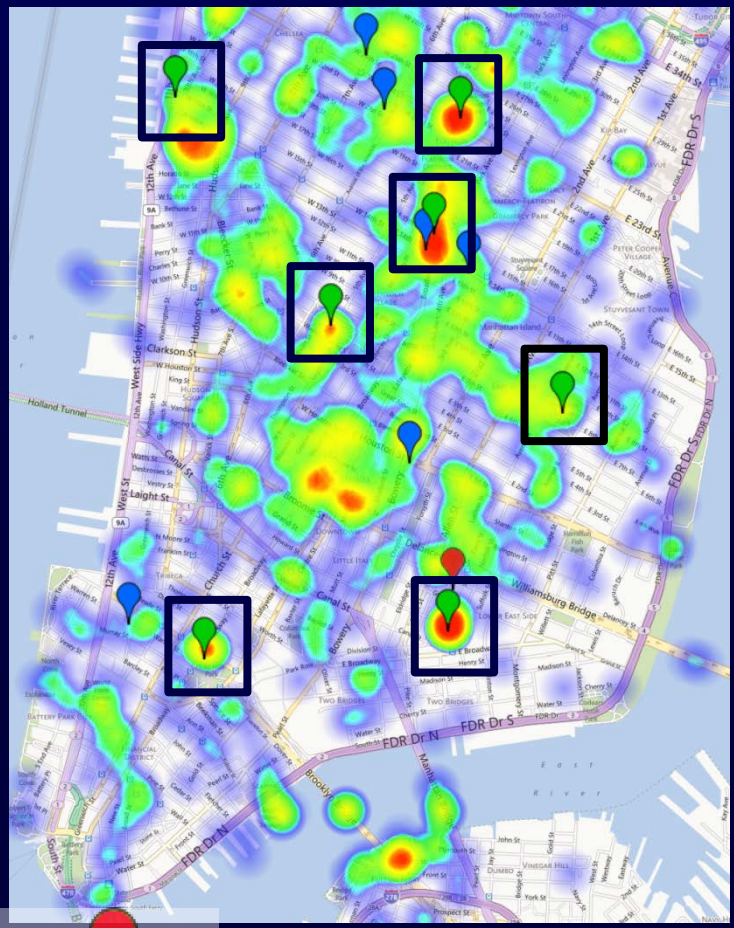


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

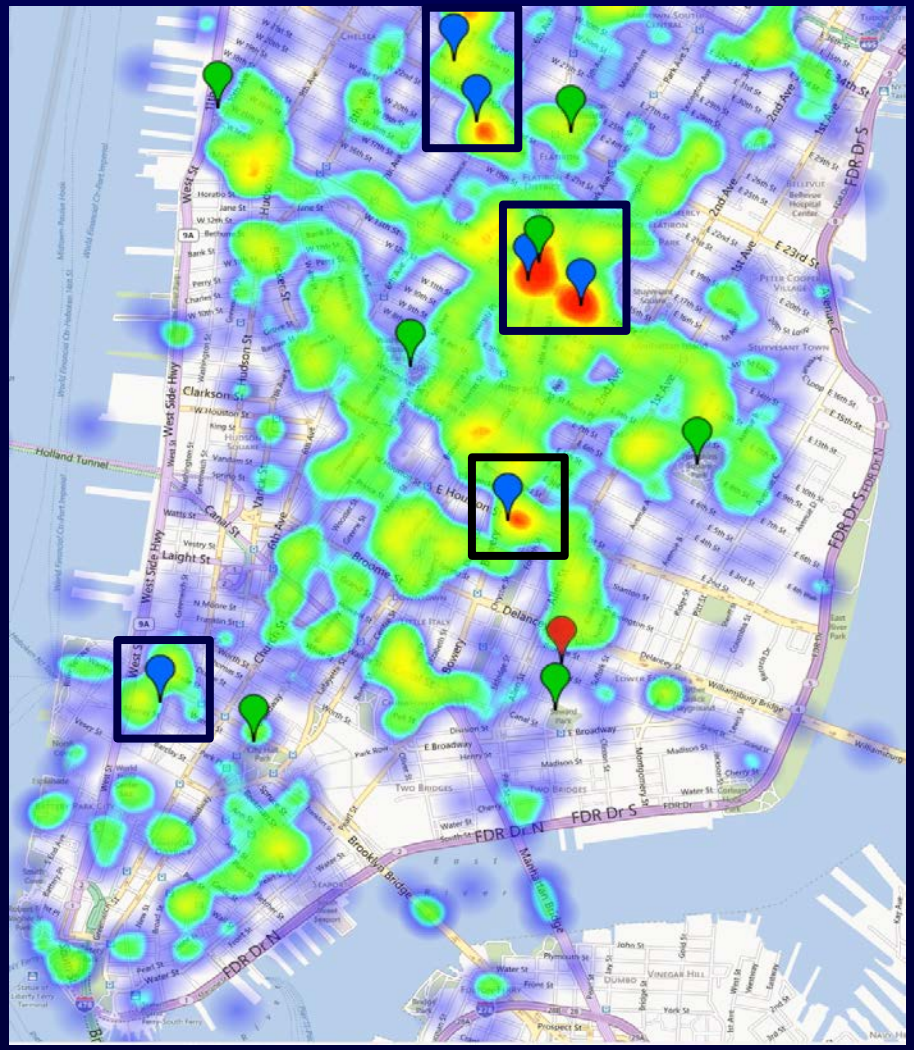
July 2015

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM

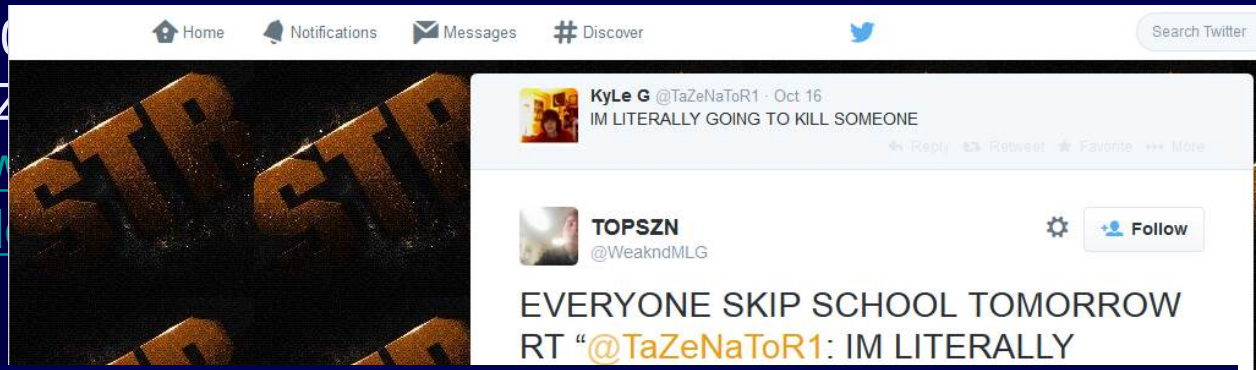


Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

For Further Information

www.VisualAnalytics-CCI.org



ORAM Visualization

Introduction

Risk-based Decision Making

Process that collects and organizes information about possible outcomes in an ordered structure that helps analysts make informed choices.[B. Macesker]

Performing risk analysis is a complex and challenging analytical task.

- Complex and large multivariate datasets.
- Clear understanding of the risk at hand.
- Cognitive limitations and information overload.

Related Work

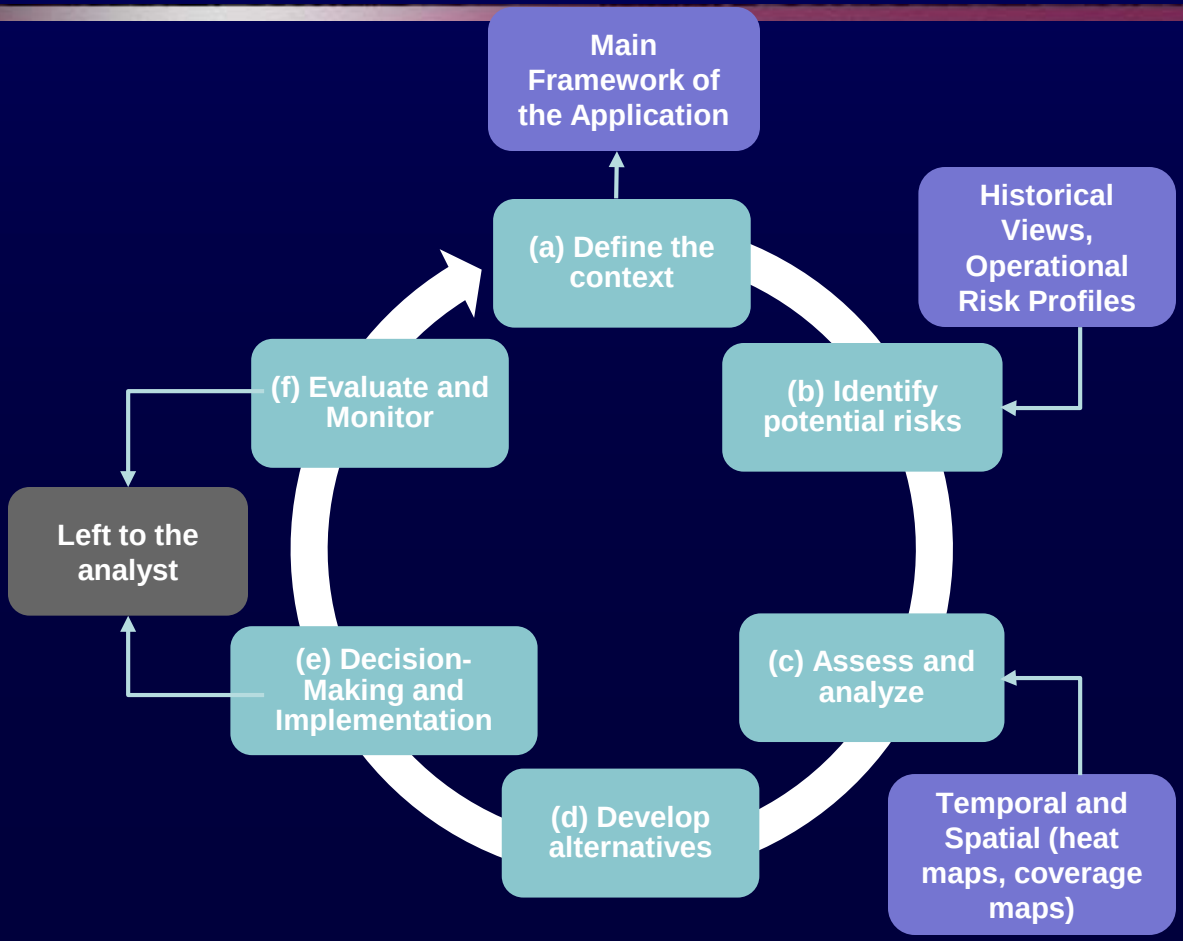
Communication of Risk

- Static risk displays [Lipkus and Holland]
- Portfolio risk management [Gershon and Eick]

Risk Analysis and Modeling

- Operational Risk Visualization [Bonafede and Marmo]
- Visualization Techniques to Present Risks and Options for Risk Mitigation[Feather et al.]
- Visual exploration and classification models for risk assessment [Migut and Worring]

Problem Characterization: Risk Management Process



Risk Terminology

- ORAM – Operational Risk Assessment Model
- 11 Missions in the Coast Guard – 10 Missions with risk values
- RIN – Risk index numbers.
- Total Risk – Mitigated Risk = Residual Risk.

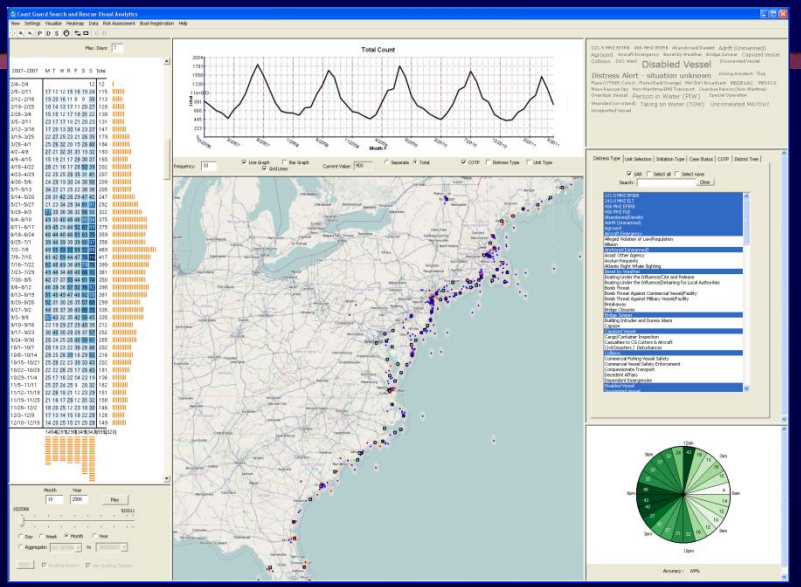
Visual Analytics Environment

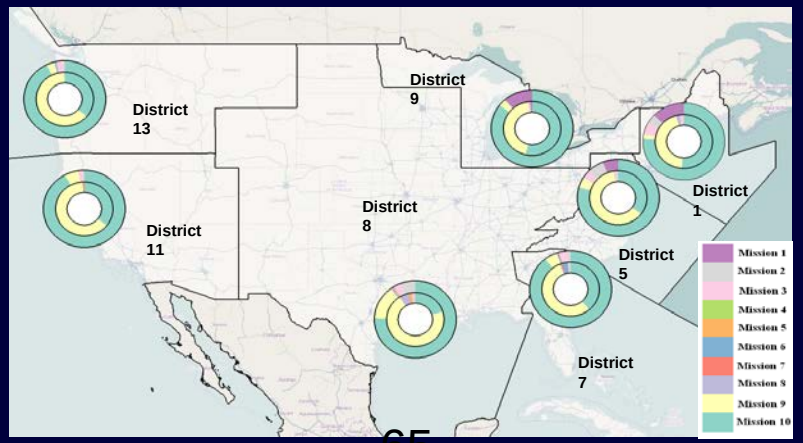
Design Requirements:

Interaction with the system.

Integration with previously used tool.

Creation of customizable scenarios.





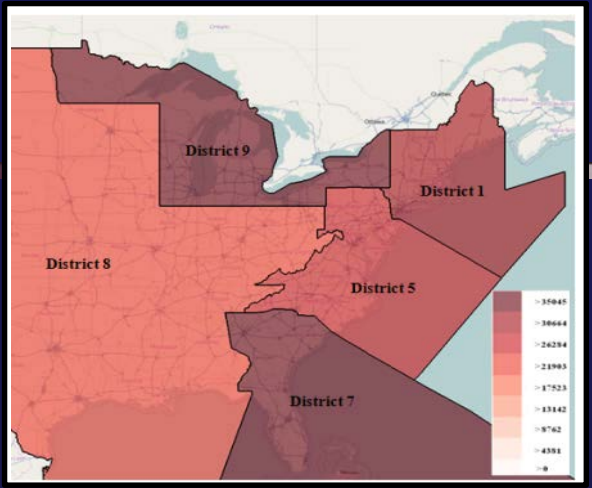
VACCINE

Scenario

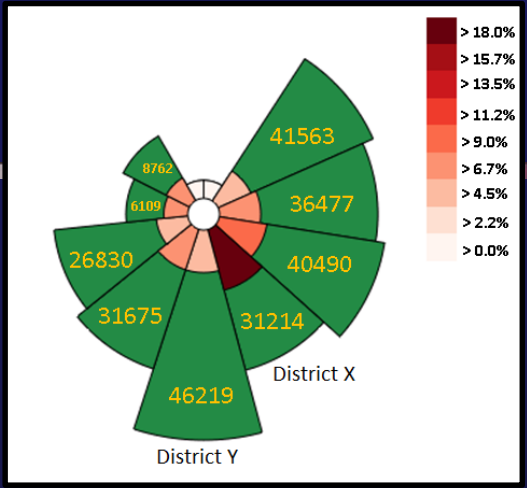
Questions that drive the planning strategy:

- What risks exist in the region and where they are distributed?
- Where are our resources allocated?
- What constraints exist in the system that will require a prioritization of resource use?

Identify and Analyze Potential Risk



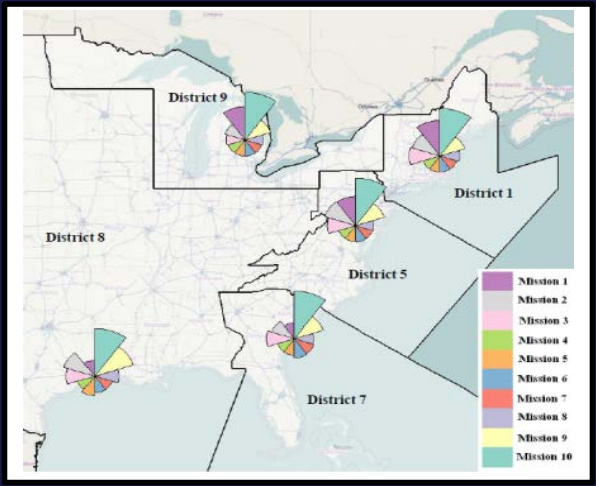
(A) District Choropleth Map for Total Risk across all missions



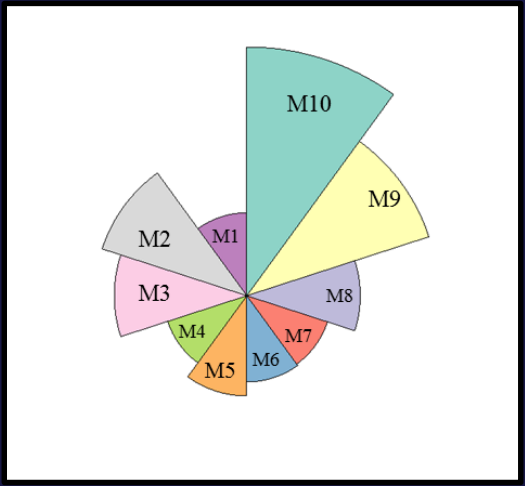
(C) Risk profiles across all districts and mission areas



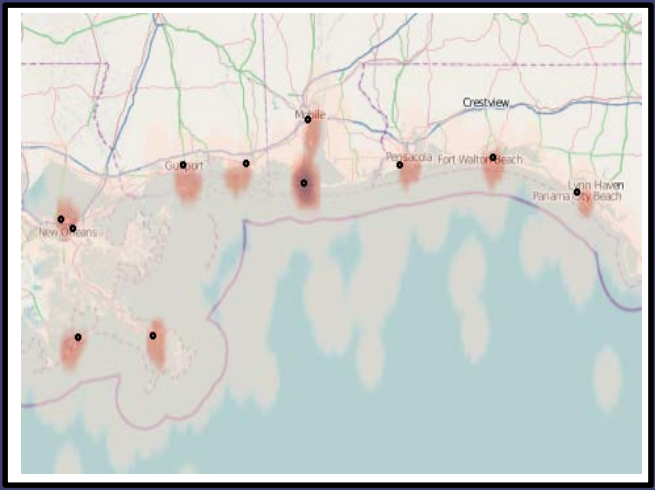
(E) Total M10 Risk Values distributed among District X's station AOR



(B) Total Risk distribution by mission for any given district.



(D) District X Risk Profiles across all mission area



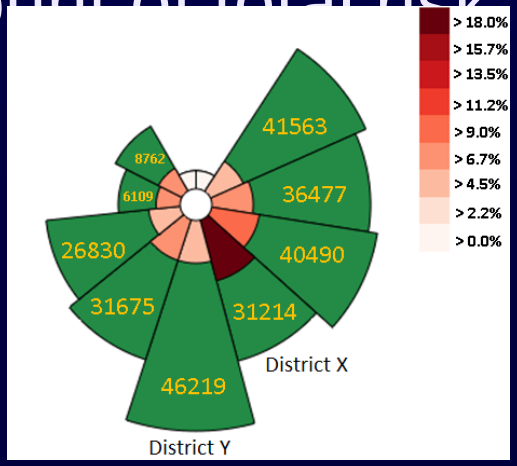
(F) KDE heatmap showing Total M10 Risk values

General understanding on how the risk is distributed between the different areas and missions.

Specific areas can be analyzed by time and the analyst can obtain details by demand.

Scenario

- 1. In a resource constrained environment, we want to use resources in the mission area that provides the greatest return on investment (large amount of total risk but very little residual risk).



- District Y has the largest total risk values, it mitigated most of the risk effectively.
- District X shows less total risk, but the amount of residual risk as well as residual to total risk ratio is the highest as encoded by the darkest red shade.
- District X can be seen as more problematic than District Y

Scenario

2. Determine the key drivers of risk within a district. What mission drives the risk?

3. Are there several big events that drive the risk, or are there many small events with smaller consequences accumulated to affect the operation?

4. Evalu sou





Feedback

- The prototype components went through an iterative design refinement process with the collaboration of four Coast Guard personnel: an operation research analyst, a former Coast Guard officer, one in-field officer, and a high level officer.

- Overall they emphasized the system allowed them to answer questions derived from developing the planning strategy in a fast manner. It also allows them to observe the underlying factors that drive the risk profiles.

Summary

- The system presented assists in the cognitive processes of quantifying evidence and comparing lines of evidence, as well as adapting to new information.
- Case Study providing useful visual reference that can communicate recommendations based on risk management.
- The risk-based decision making process described serves as a blueprint for future systems dealing with risk values and resource planning.

Acknowledgments

• This work was partially funded by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0003. Jang's work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A1011170).
• Thank you to Ben Maule and Joe DiRenzo from the US Coast Guard for their help.

Backup Slides

Definition of Risk

- “Potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences”
- R. S. Committee. DHS Risk Lexicon. Homeland Security, 2010

Coast Guard Missions

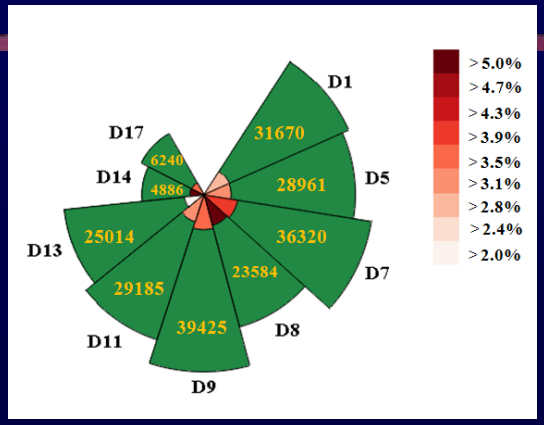
Coast Guard Mission	CG Mission (Short Name)
Ports, Waterways, and Coastal Security	PWCS
Drug Interdiction	CD
Migrant Interdiction	AMIO
Living Marine Resources	LMR
Other Law Enforcement	OLE
Marine Safety	MARSAFE
Search and Rescue	SAR
Marine Environmental Protection	MEP
Defense Readiness	DR
Ice Operations	DOMICE
Aids to Navigation	ATON

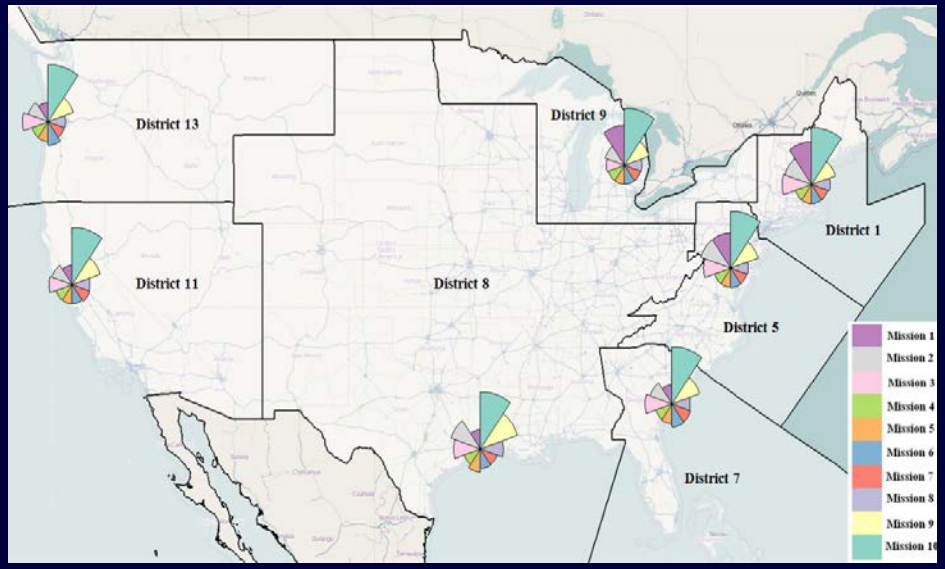
Operational Risk Profiles

Compare risk values against districts.

Compare risk values against missions

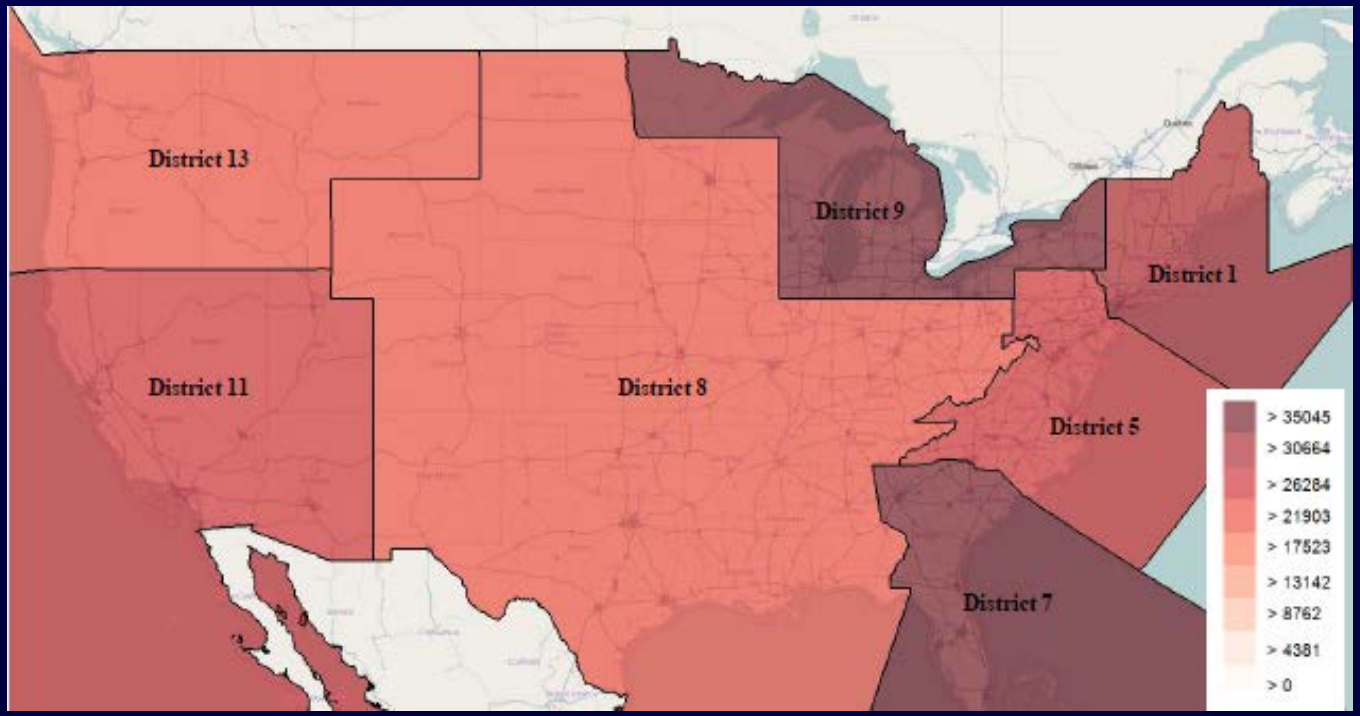
Iterative process with feedback from the Coast Guard





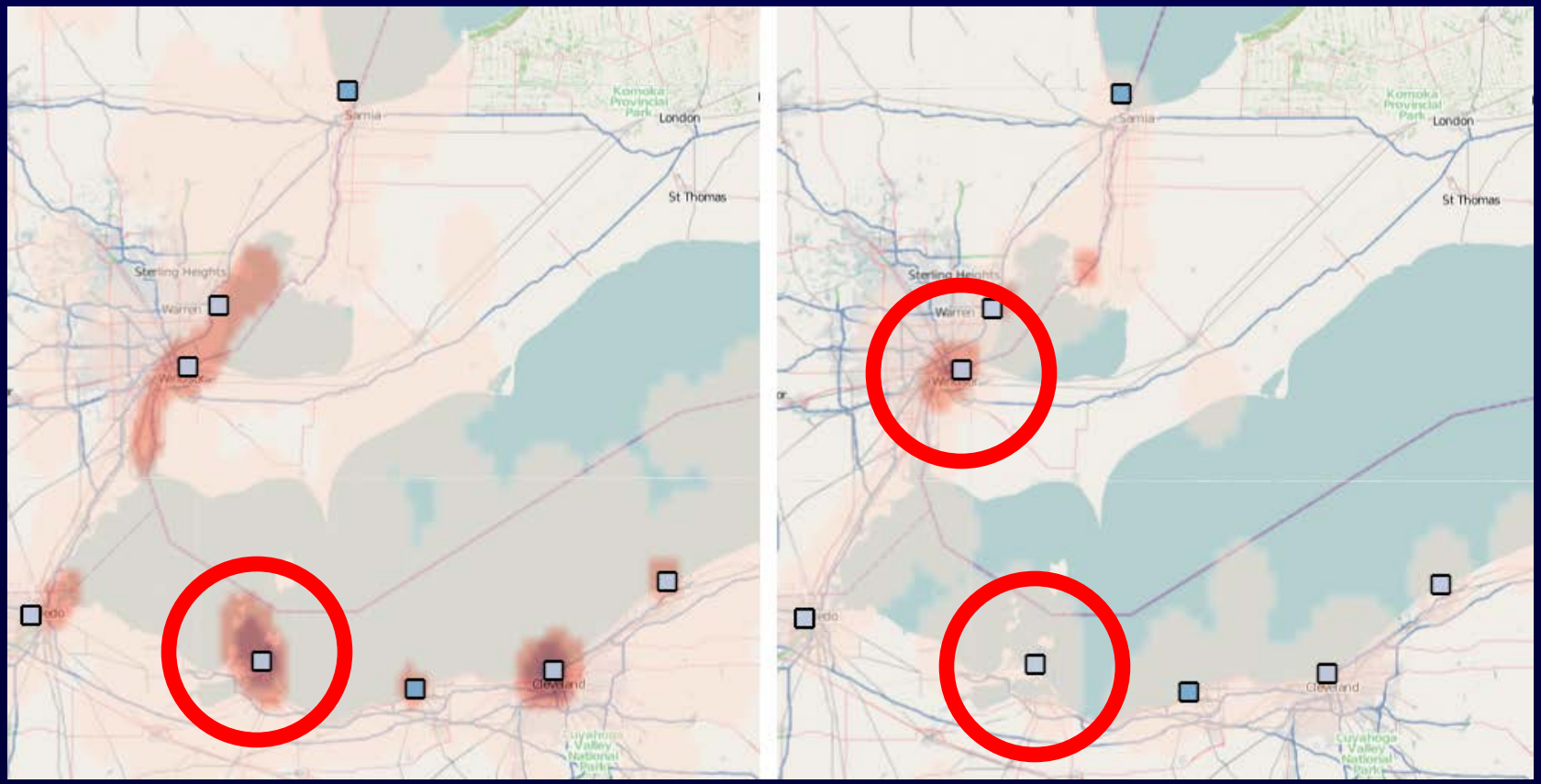
VACCINE

Choropleth Maps



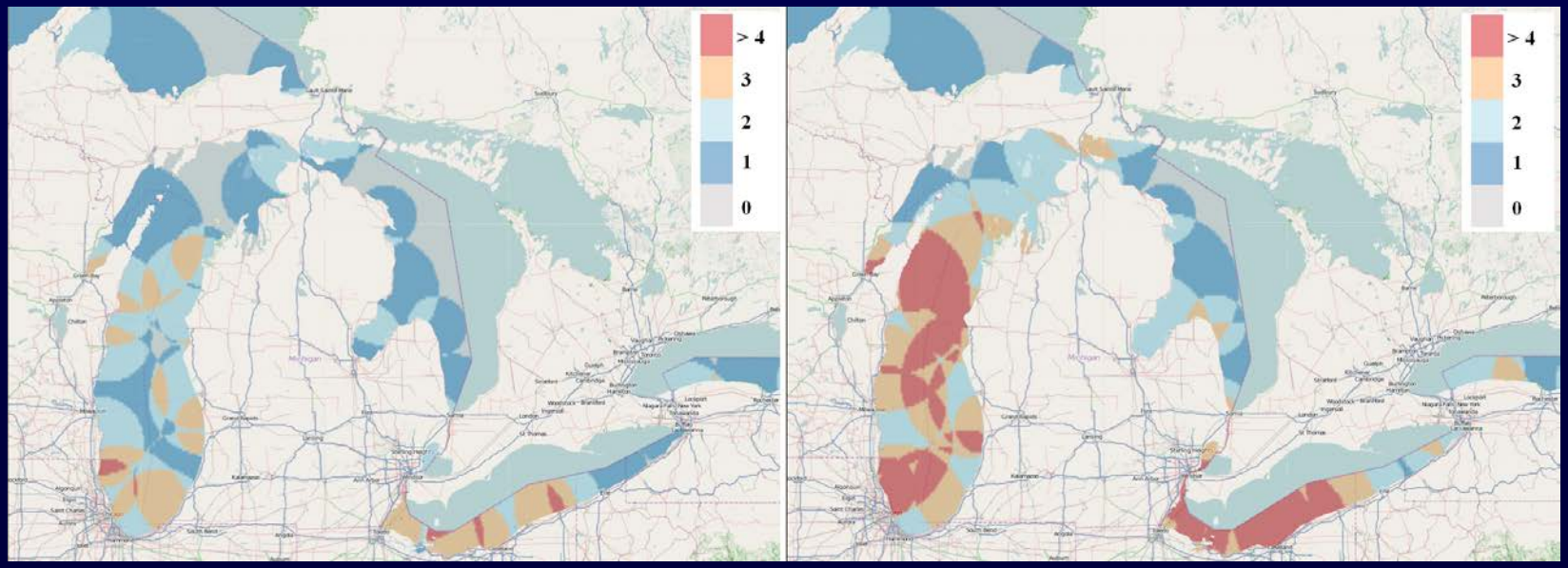
Heatmap for risk values

- Total vs Residual comparison.

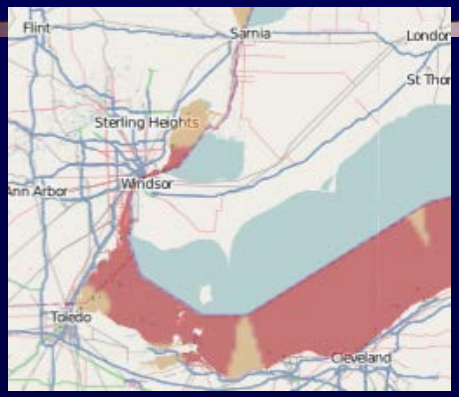


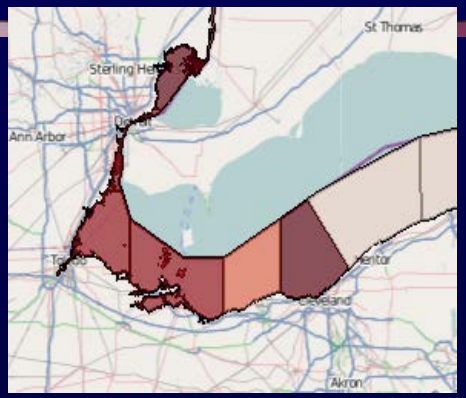
VACCINE

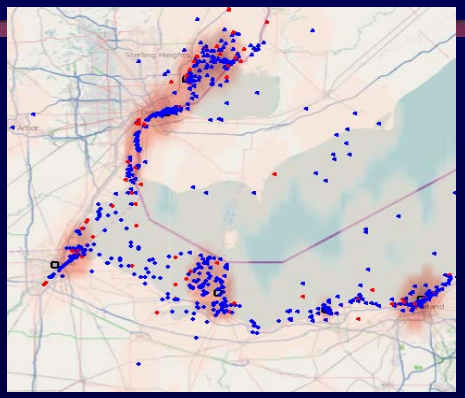
Coverage Maps for Performance Evaluation



Analyst's Evaluation





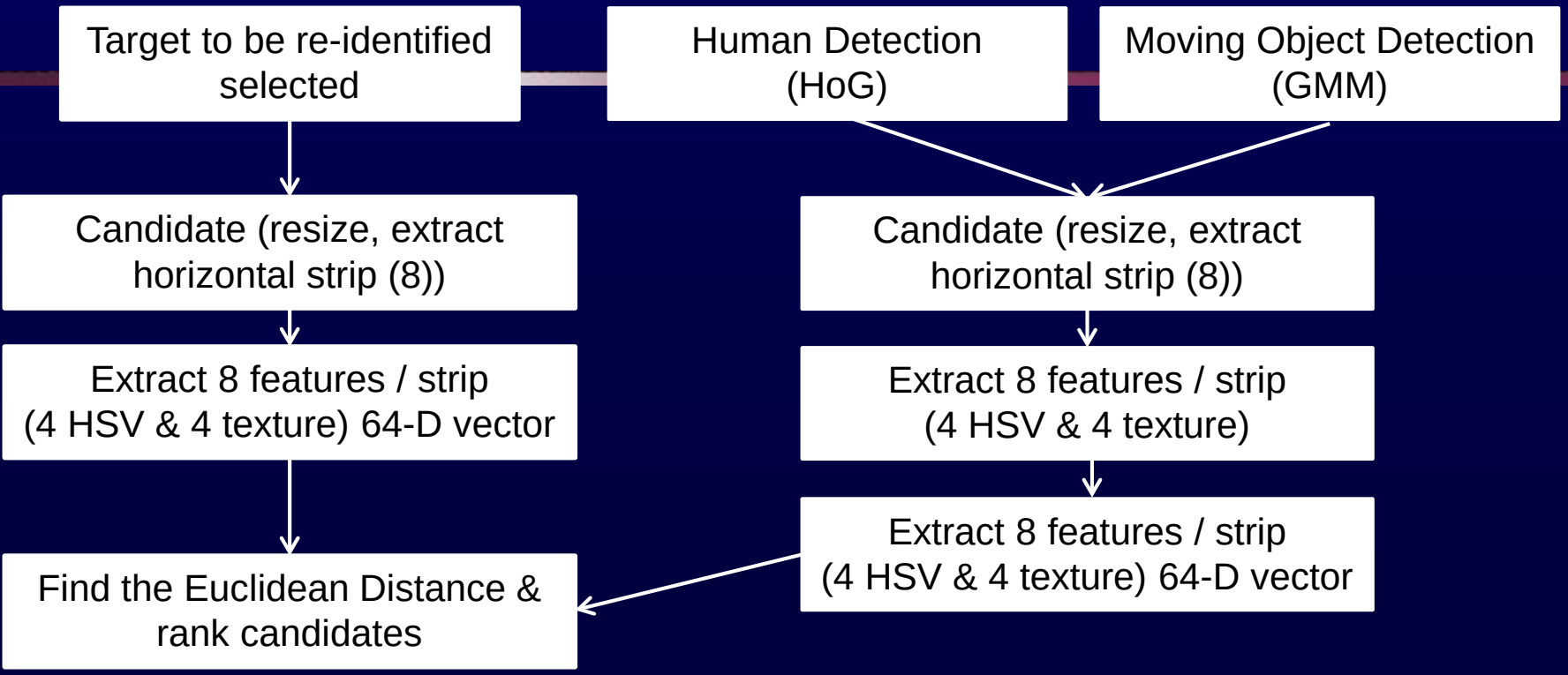


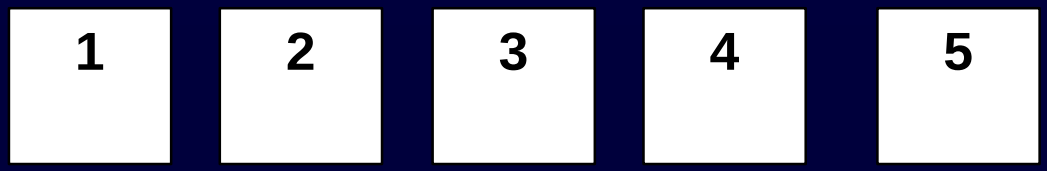


VACCINE

VACCINE

EBOLO





The top 5 identified targets

EBOLO Analysis



Human Detection (HoG)

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

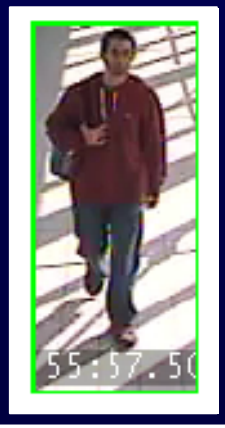
Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

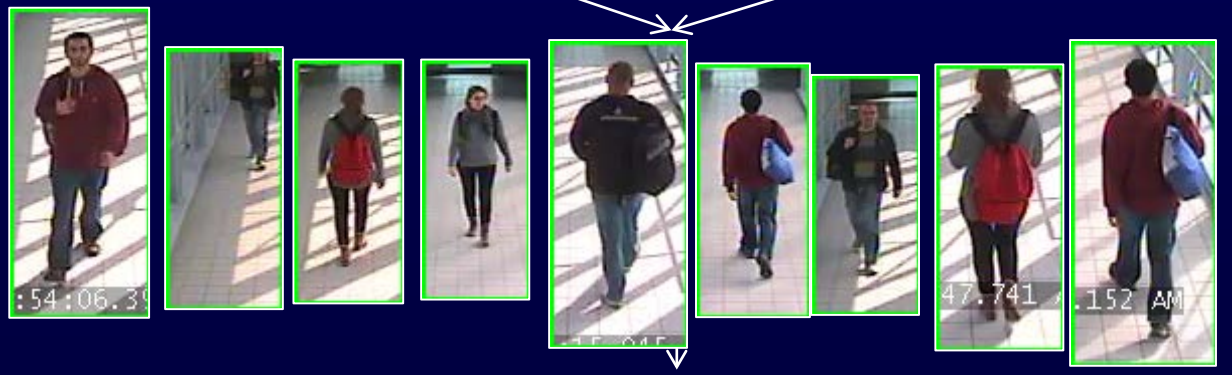
Experimental Results

Target to be re-identified selected

Human Detection (HoG)

Moving Object Detection (GMM)





Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

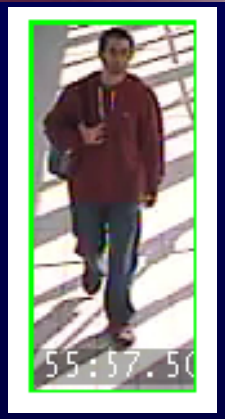
Find the Euclidean Distance & rank candidates

1	2	3	4	5

The top 5 identified targets

Experimental Results

Target to be re-identified selected



Find the Euclidean Distance & rank candidates



The top 5 identified targets

VIPER Lab

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

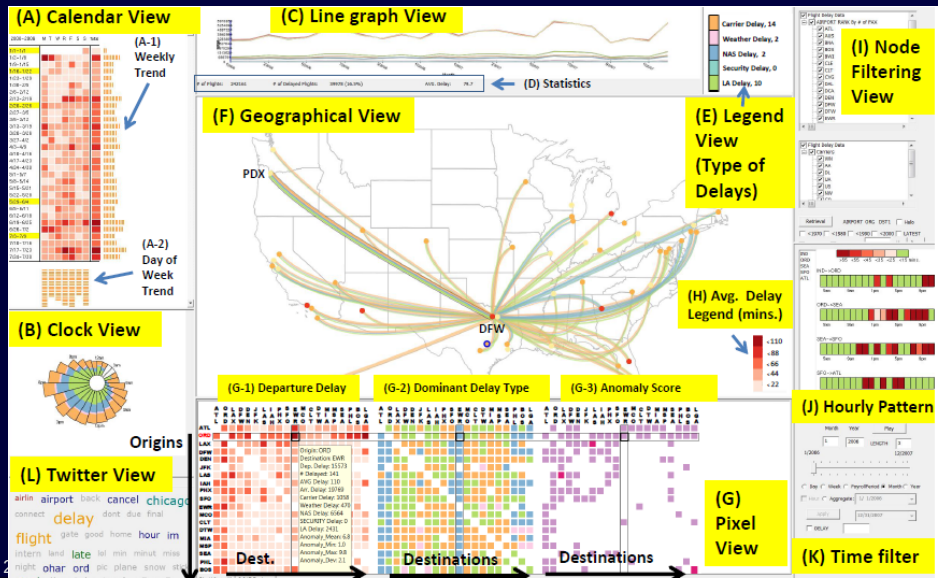
Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

December 2014

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects



VACCINE

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

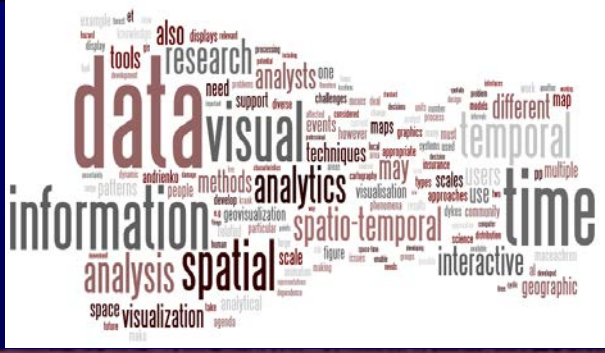
Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America



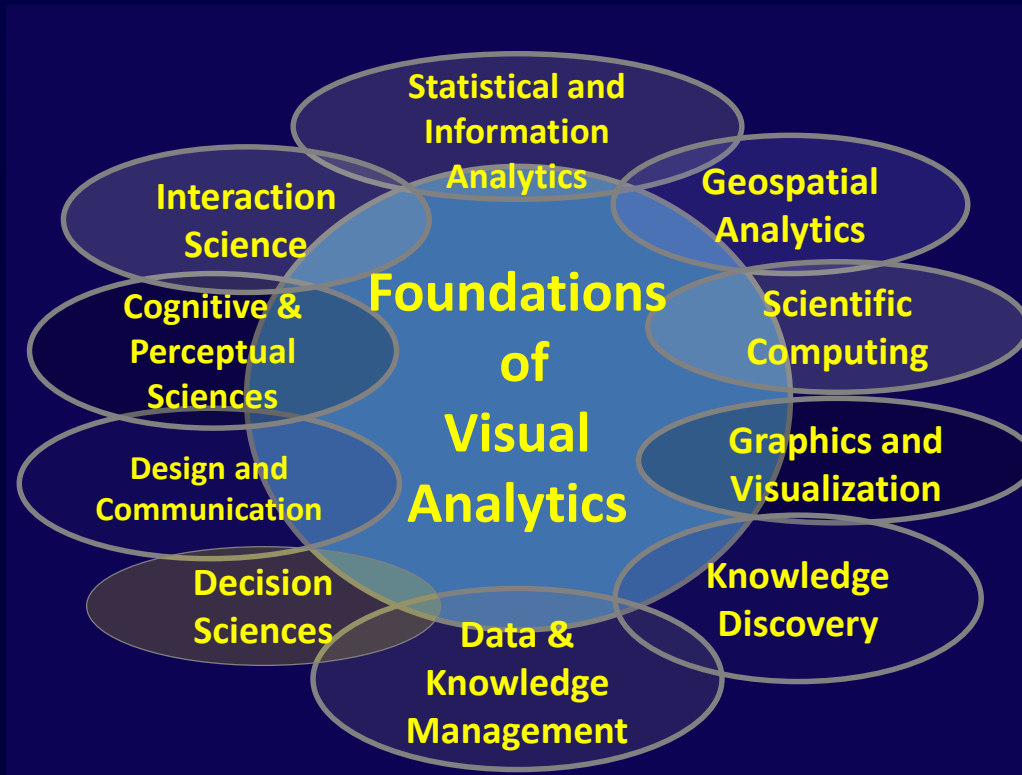


What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

Big Data



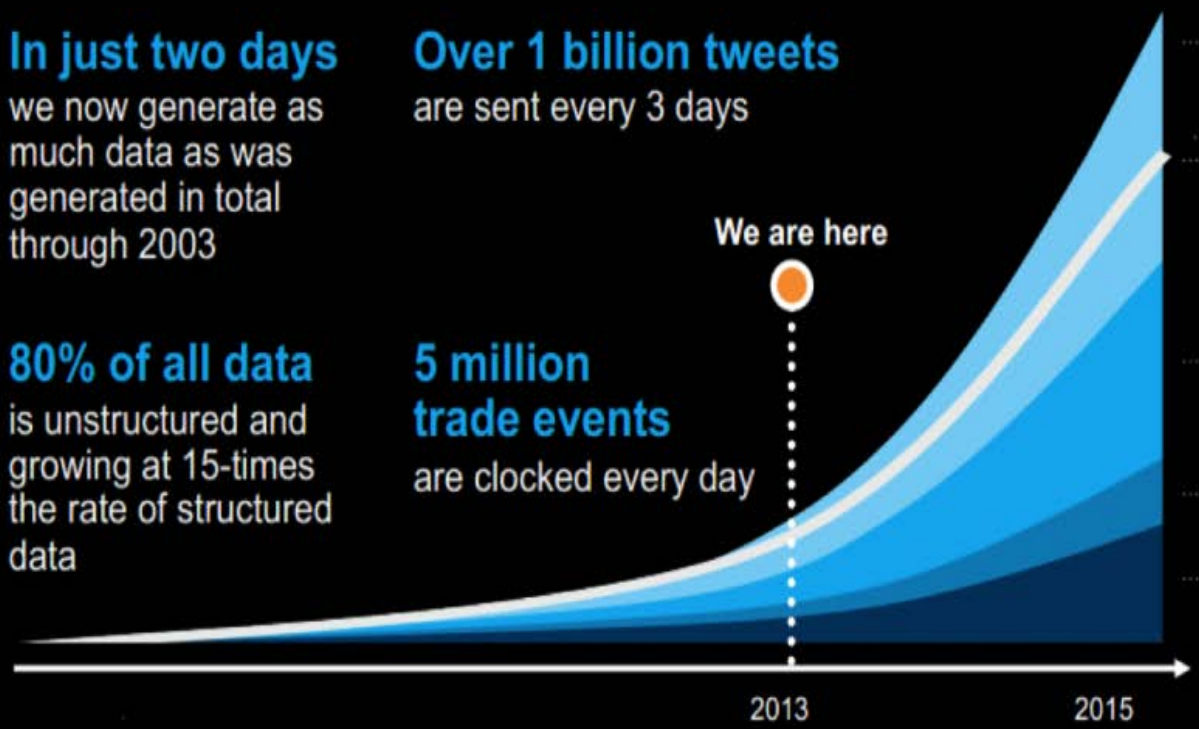
These forces fuel an explosion of data – a new economic asset forming the basis of opportunity

In just two days
we now generate as much data as was generated in total through 2003

Over 1 billion tweets
are sent every 3 days

80% of all data
is unstructured and growing at 15-times the rate of structured data

5 million trade events
are clocked every day



Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

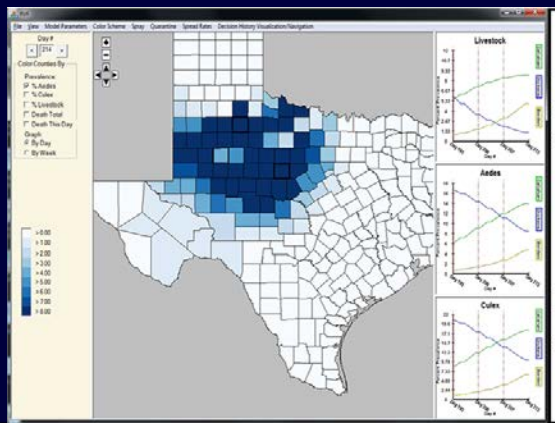
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

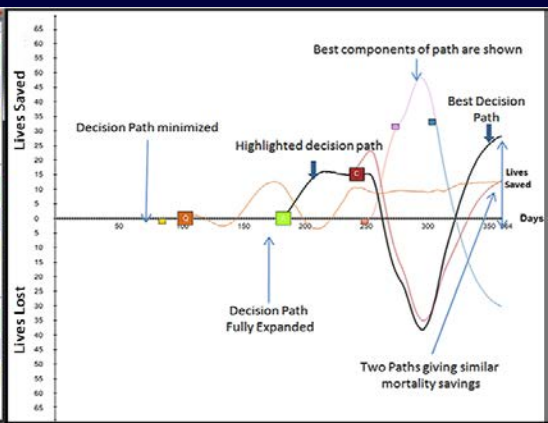
VADM Robert Parker, 2012 MRS Keynote Address



Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Research topics
- **Example application areas and projects**

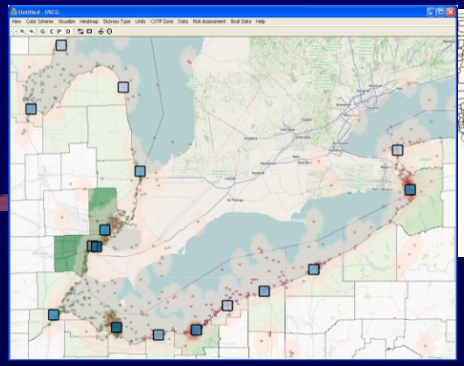


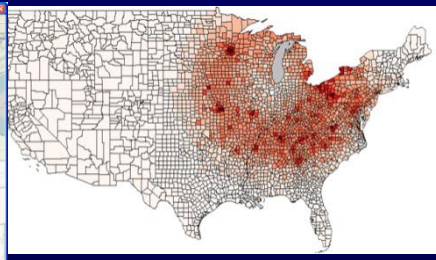


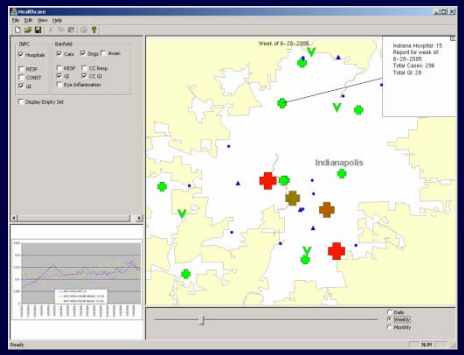
VACCINE

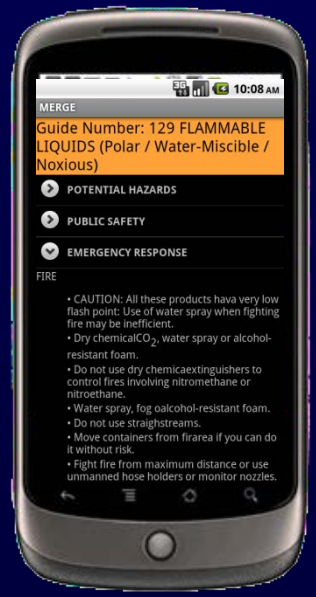
Public Safety & Health

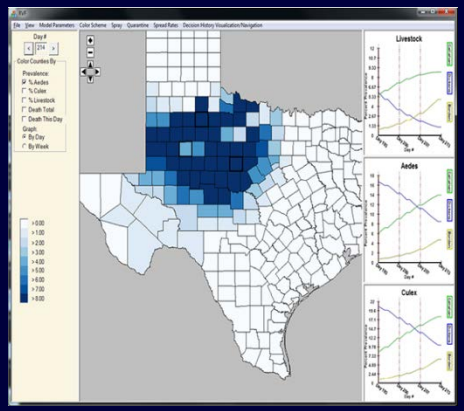
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering







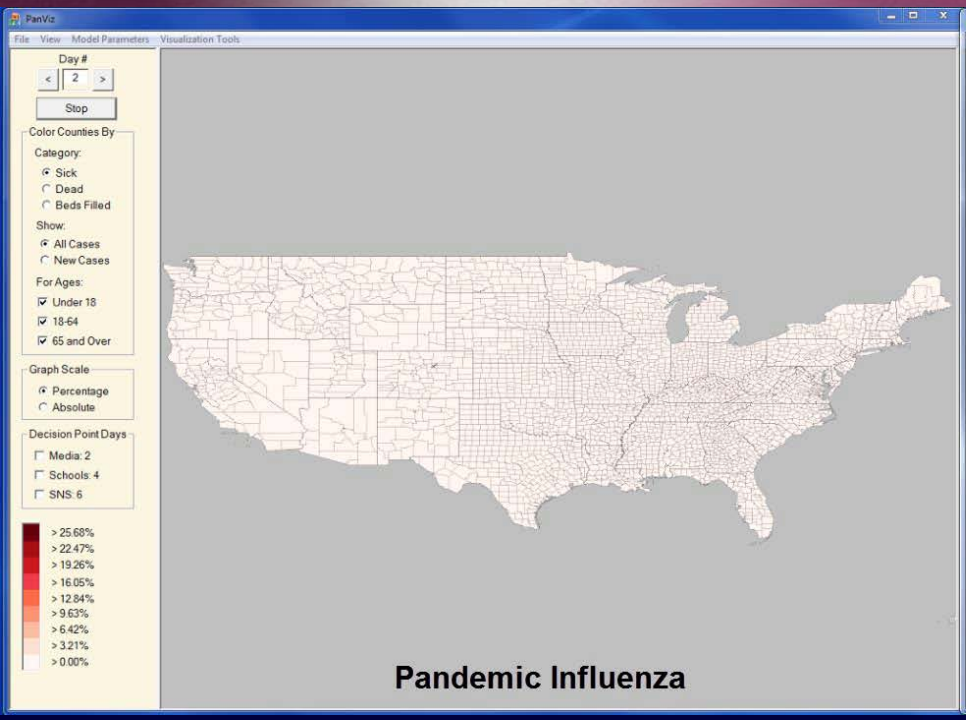


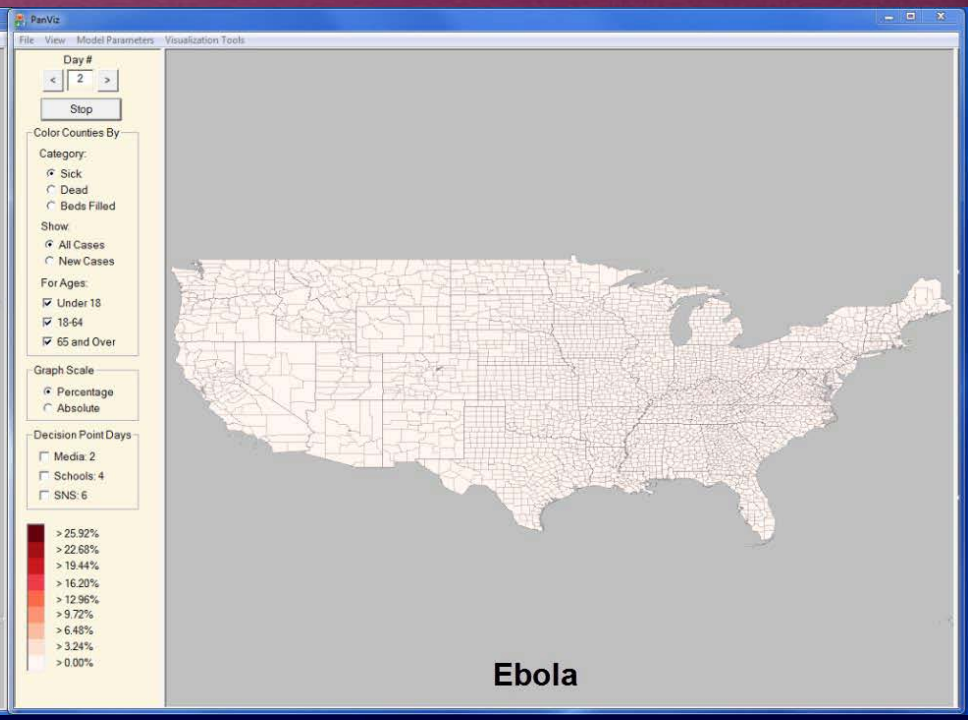




VACCINE

Ebola Vs. Pandemic Influenza (H1N1)

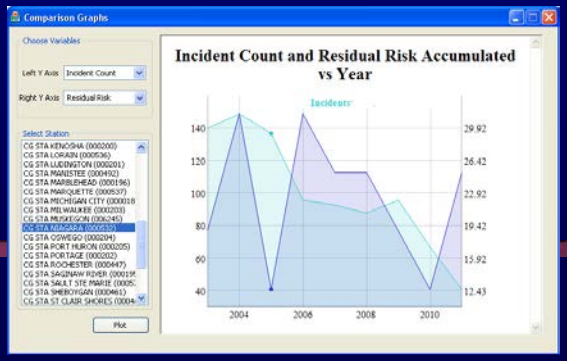


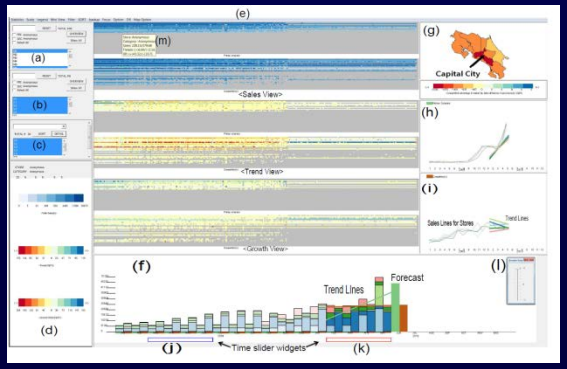


Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

Business Visual Analytics

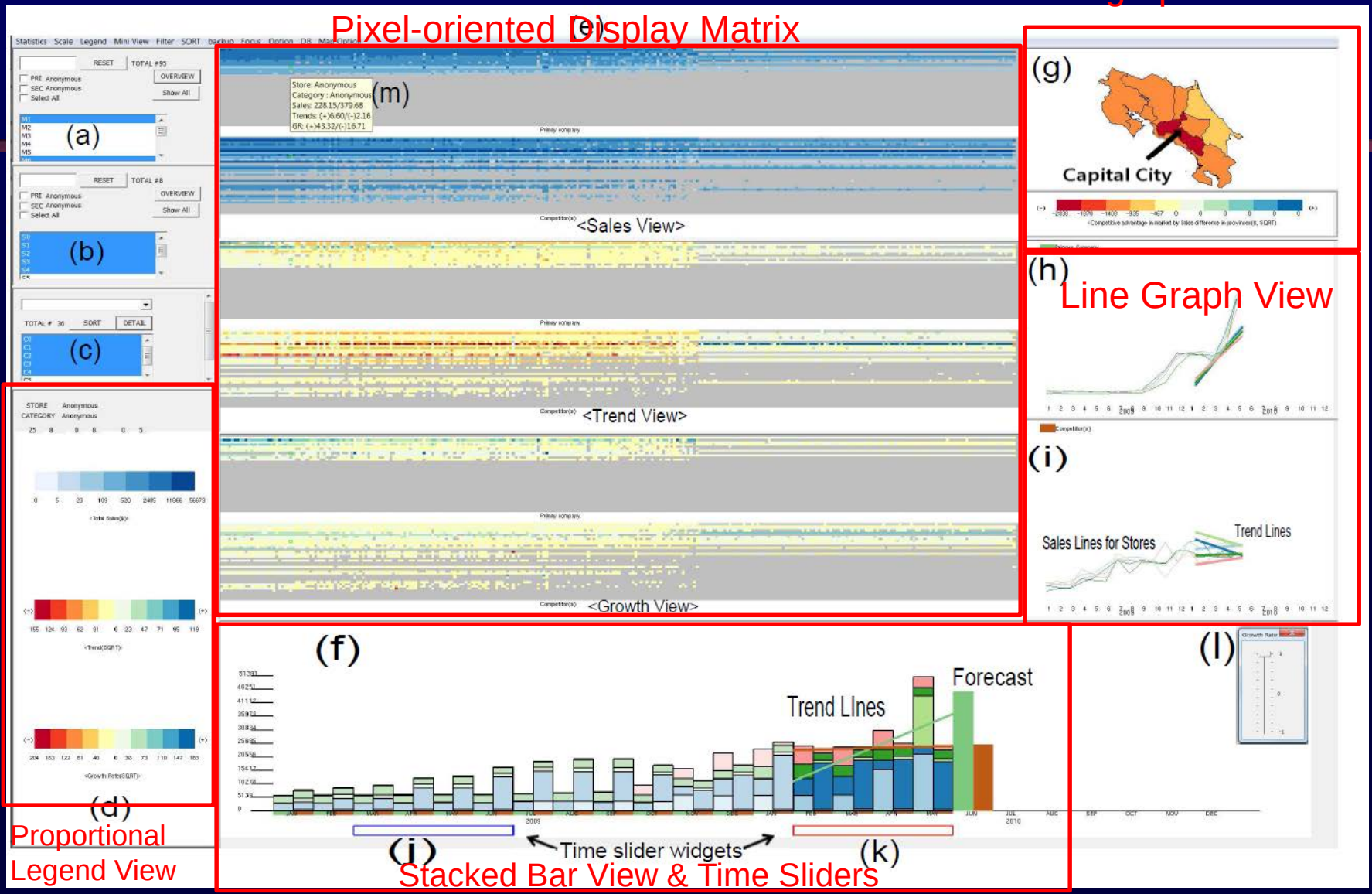
- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







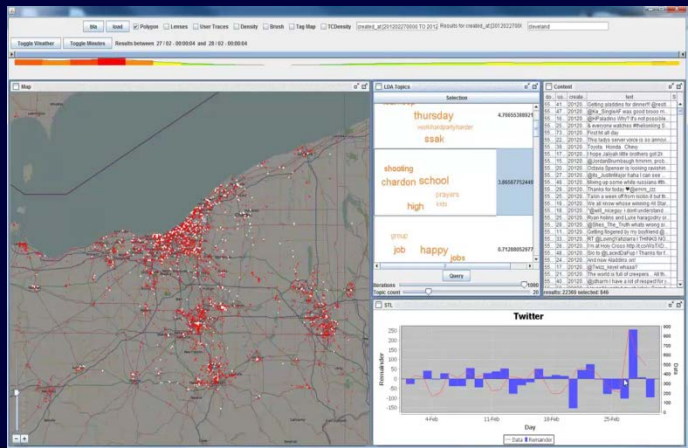
Market Analyzer



Visual Analytics Uses for Public Safety



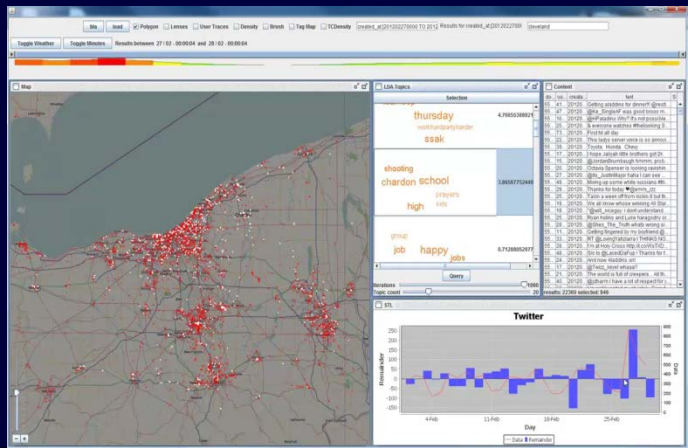
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Our Spatiotemporal Visual Analytics Work Contributions

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

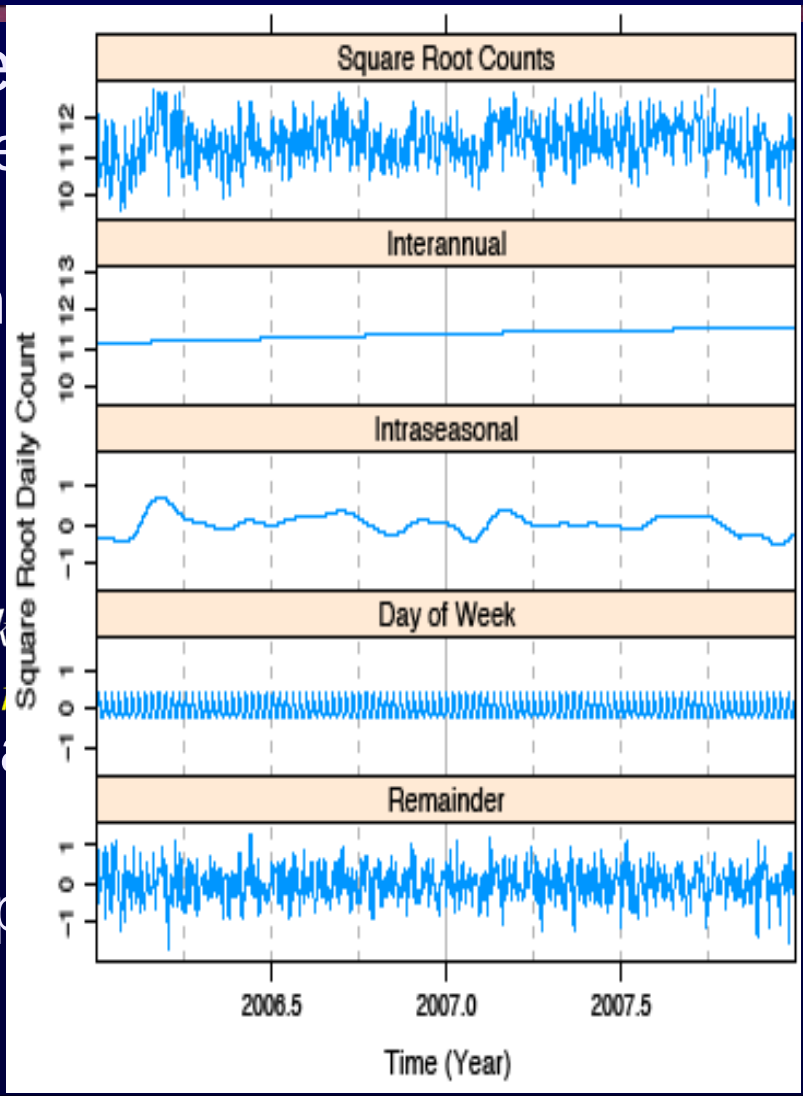
Event/Anomaly Detection - STL

- Temporal trend analysis and pre
  - Seasonal Trend decomposition base
    weighted regression) (STL)
  - Time series data viewed as the sum

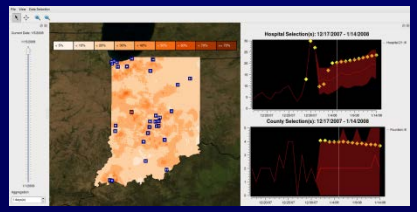
$$Y = T$$

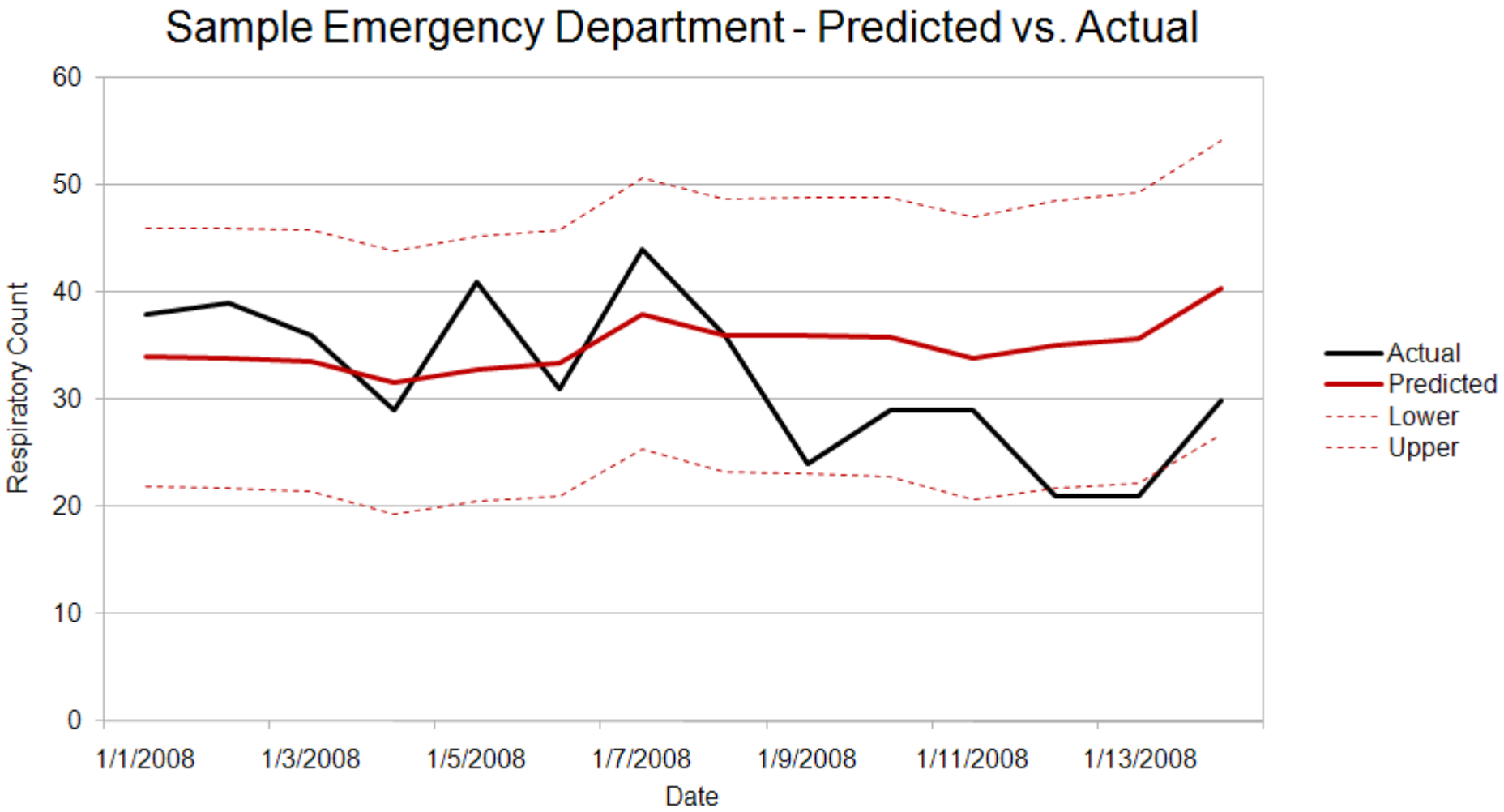
Y : original time series
T : trend component
S : seasonal (daily/week
R : remainder compone

  - Large value of *R* indicates substanti
  - Filter out the noise component
  - If value varies by 2 SD from STL exp



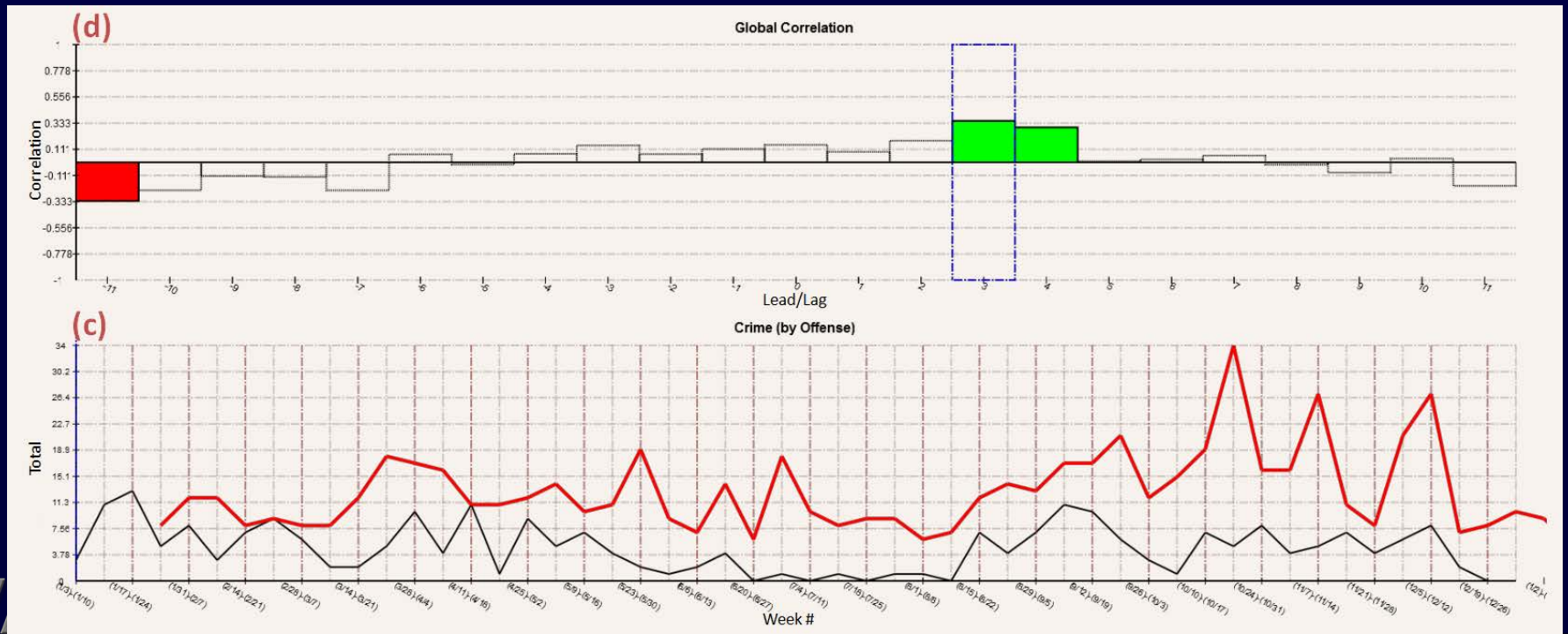
Predictive Visual Analytics



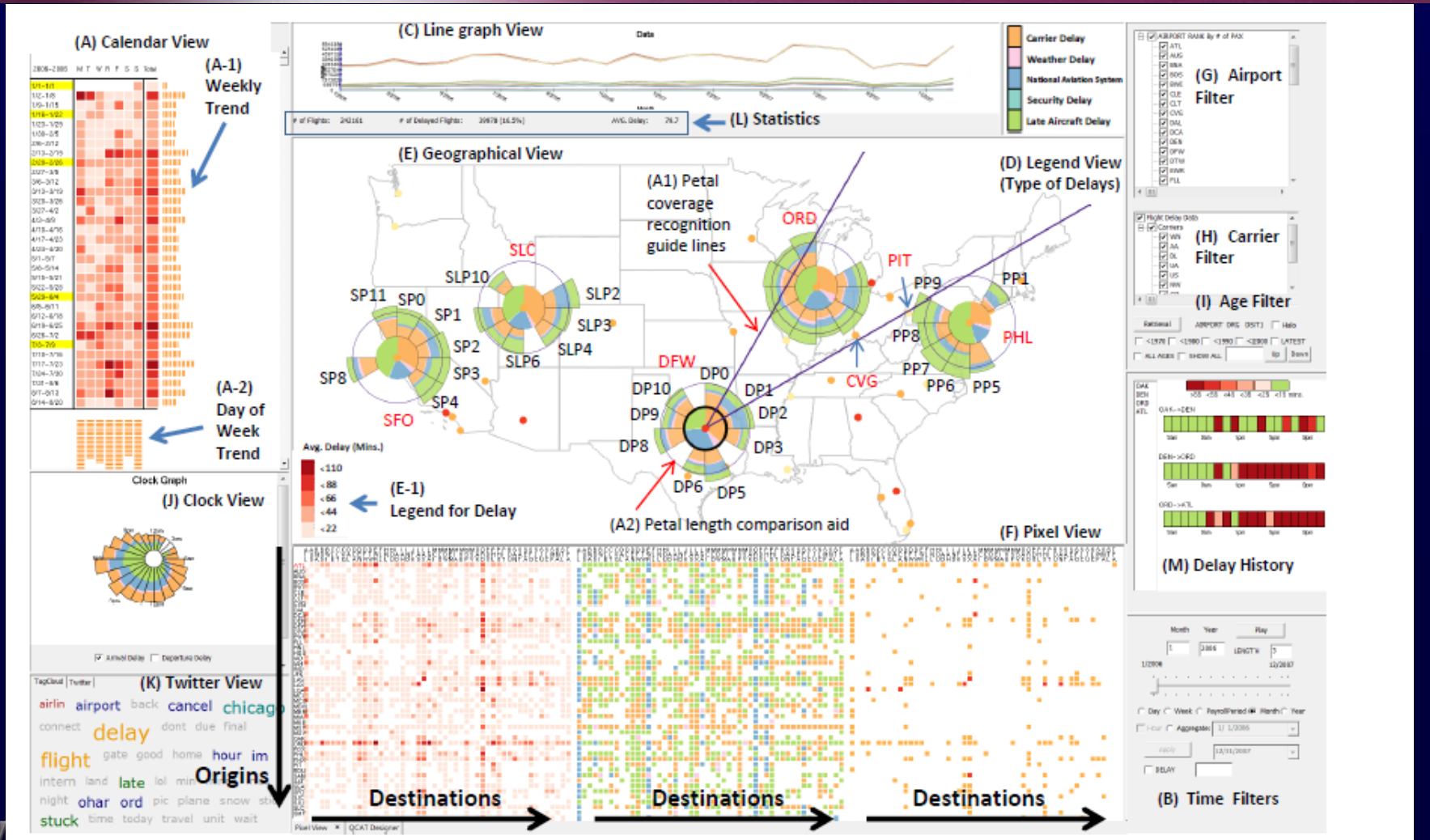


Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space



Multivariate Network Analytics with Information Theory Anomaly Detection



The Scale Problem

- Data scale:
  - Coarse scale analysis can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis may limit applicability because of fine resolution and can have data sparseness issues
- User / problem scale:
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally

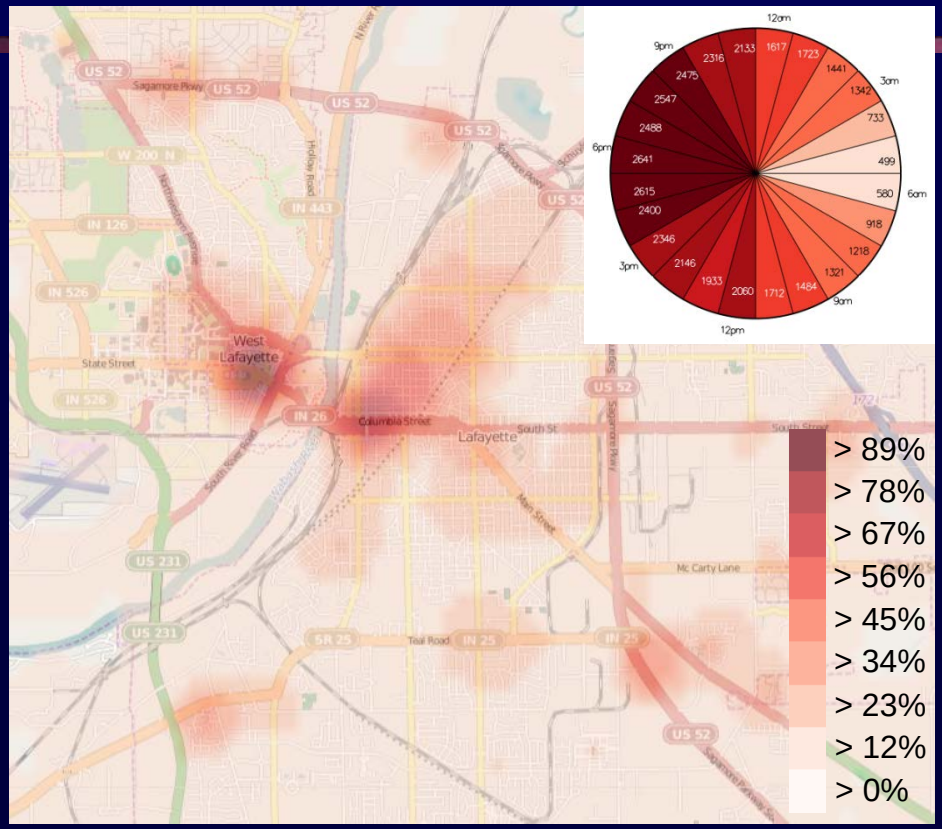
Our Approach: Natural Scale Templates

- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness[1]
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Natural Scales: Mappings that facilitate human perception to comprehend the data/task/problem
  - Can facilitate discovery from data
  - Goal: **Empower the casual expert to be more effective**

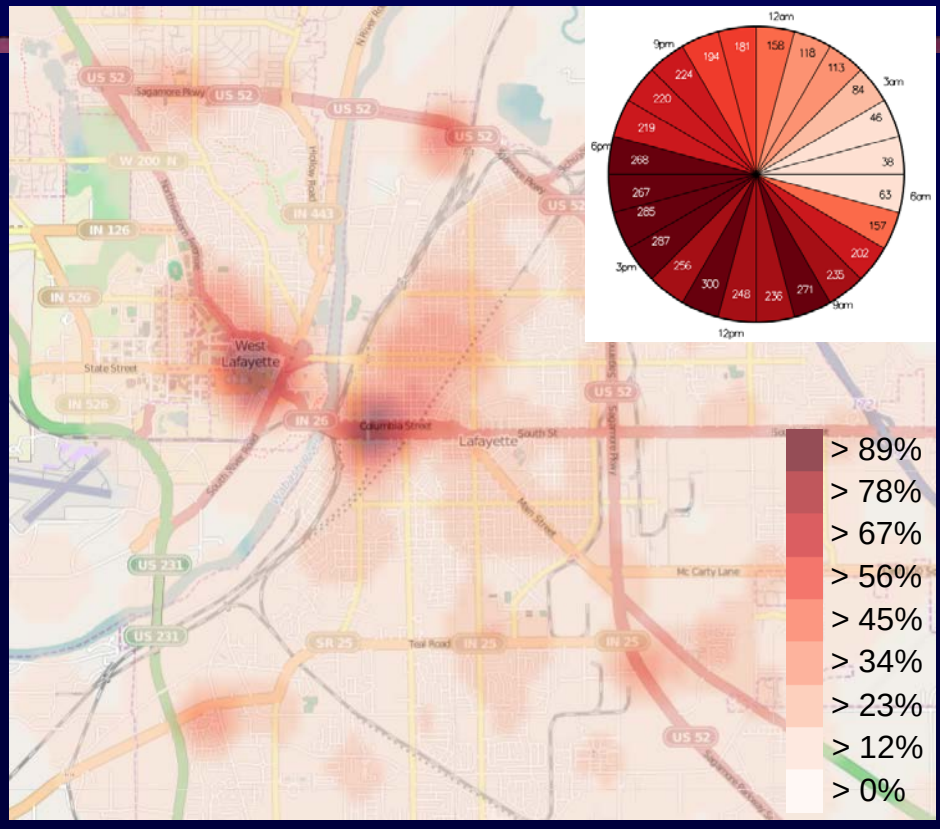
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

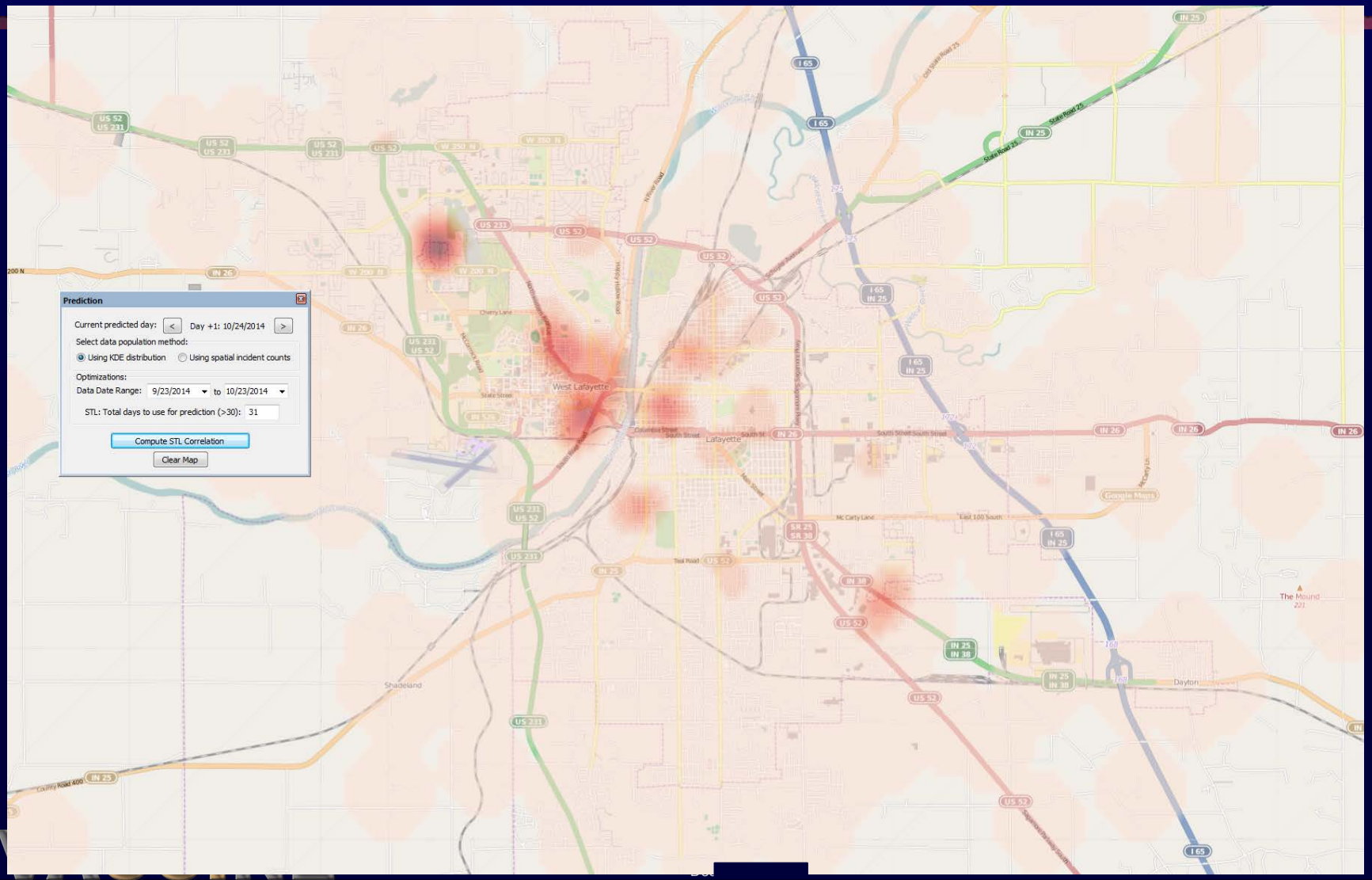


Past 2 Years
3/11/2012 – 3/10/2014

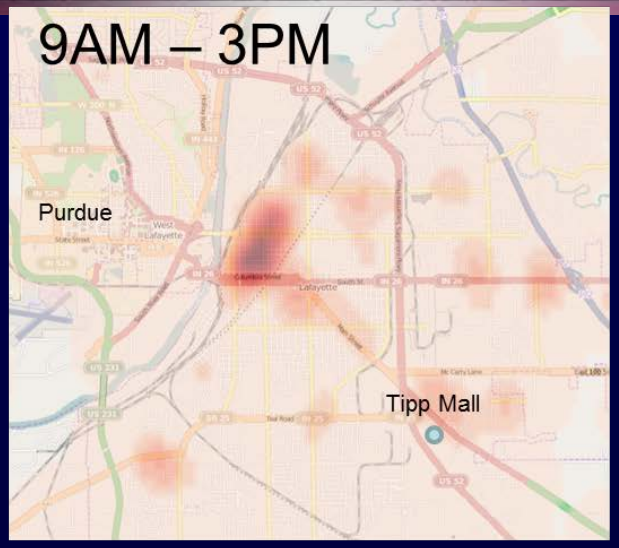


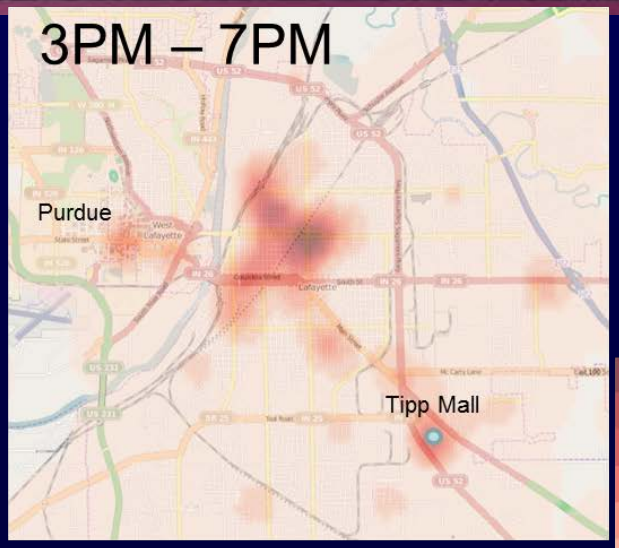
March + Tuesday Filter
(Past 10 Years)

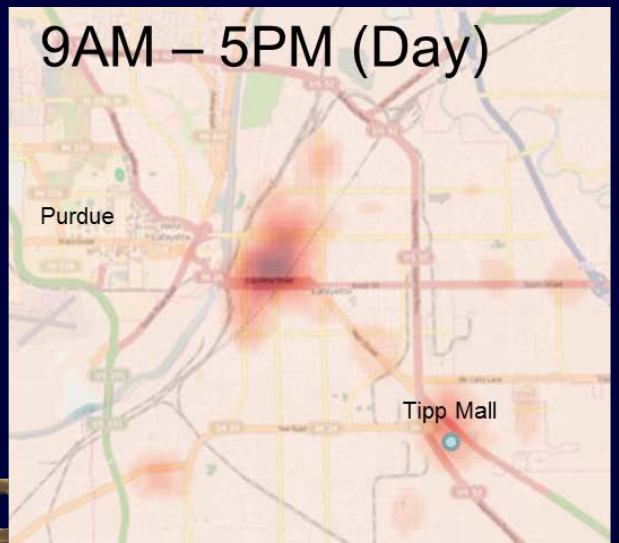
Prediction Using KDE Distribution

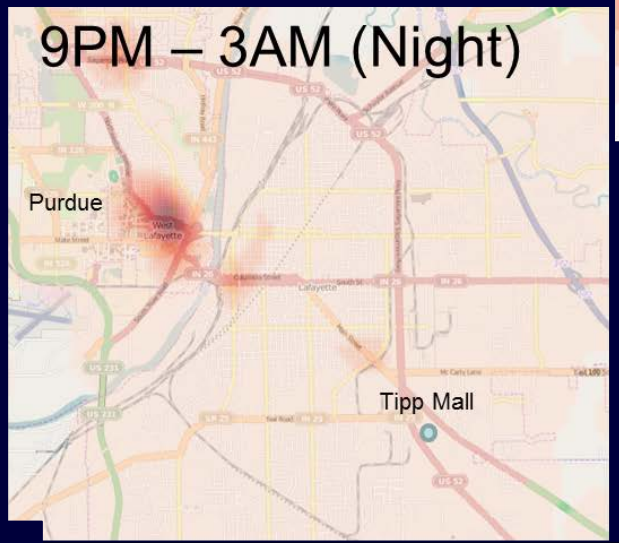


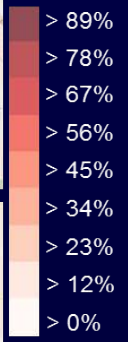
Prediction During Certain Hours In The Day





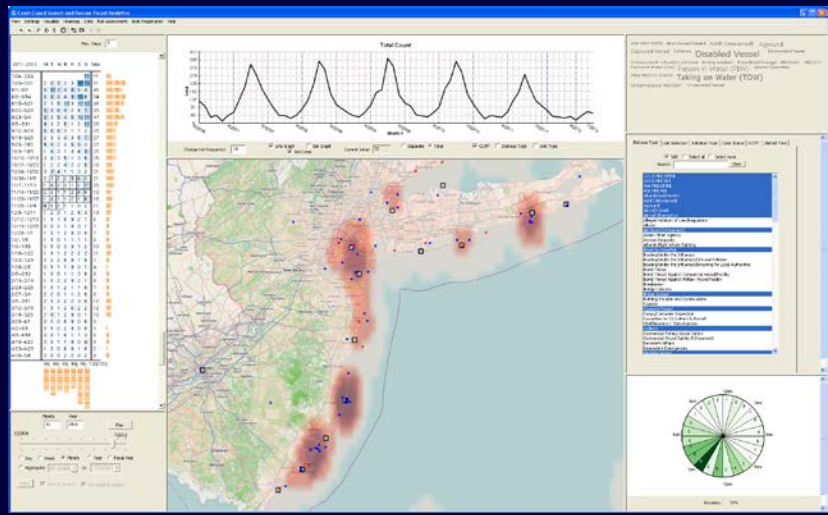




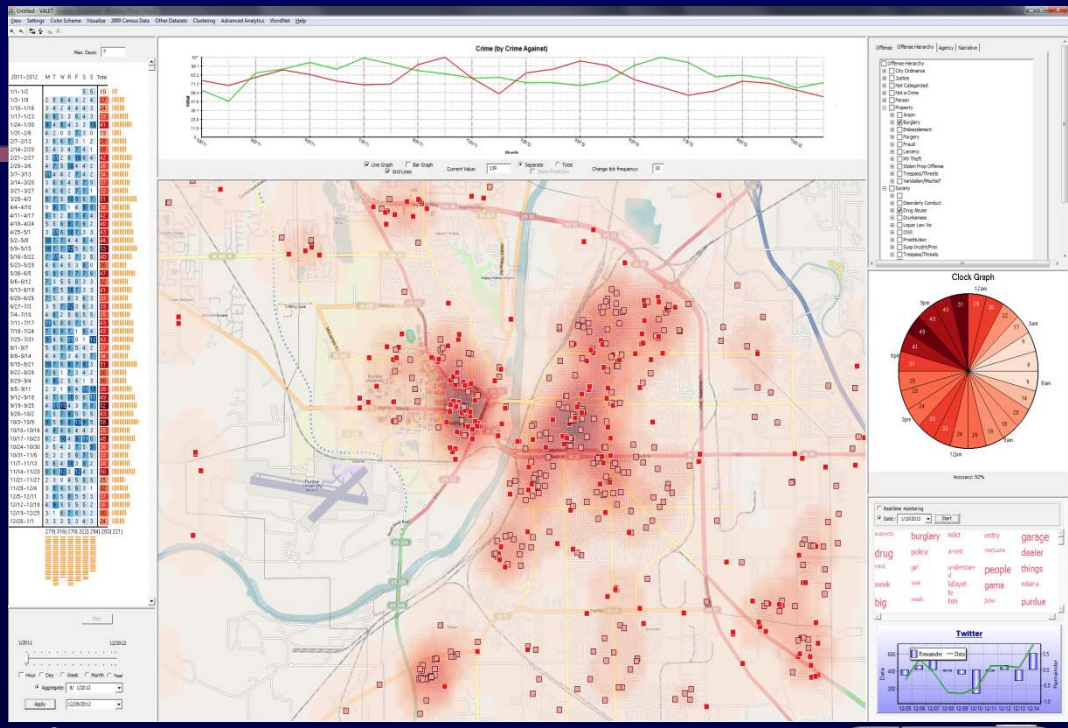


Overview

- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

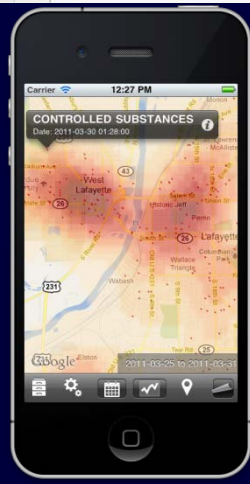
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

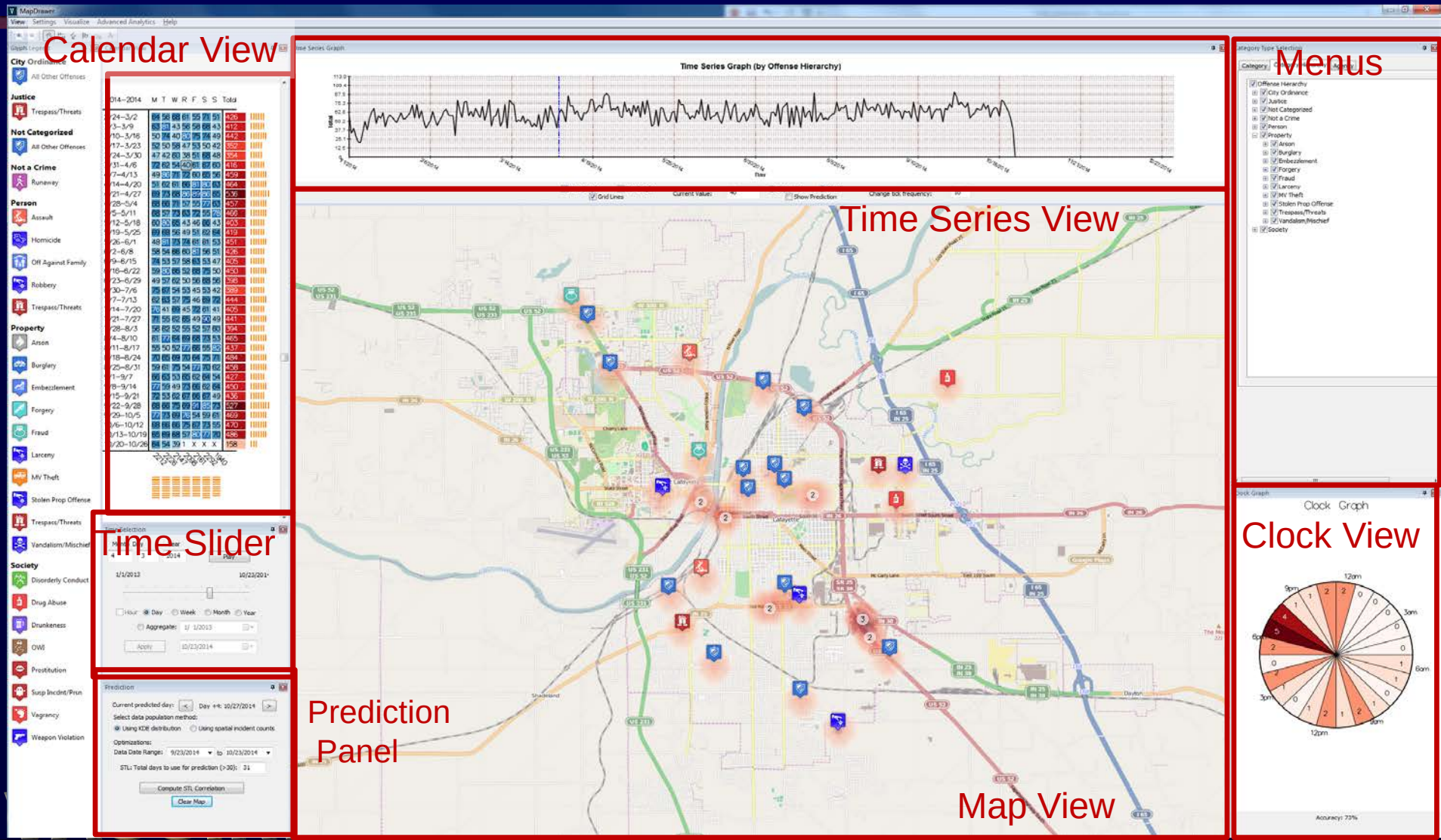
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

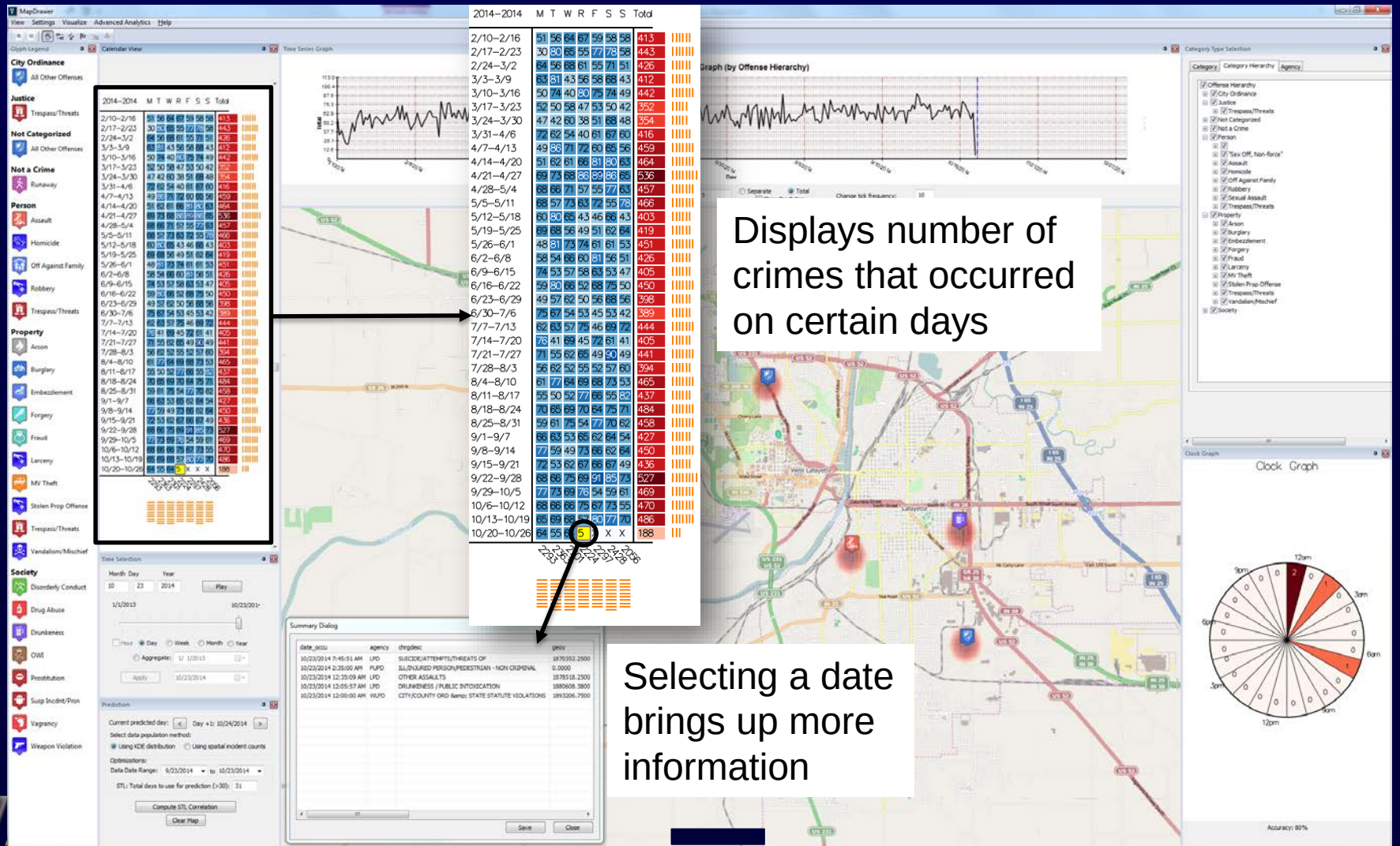


VACCINE

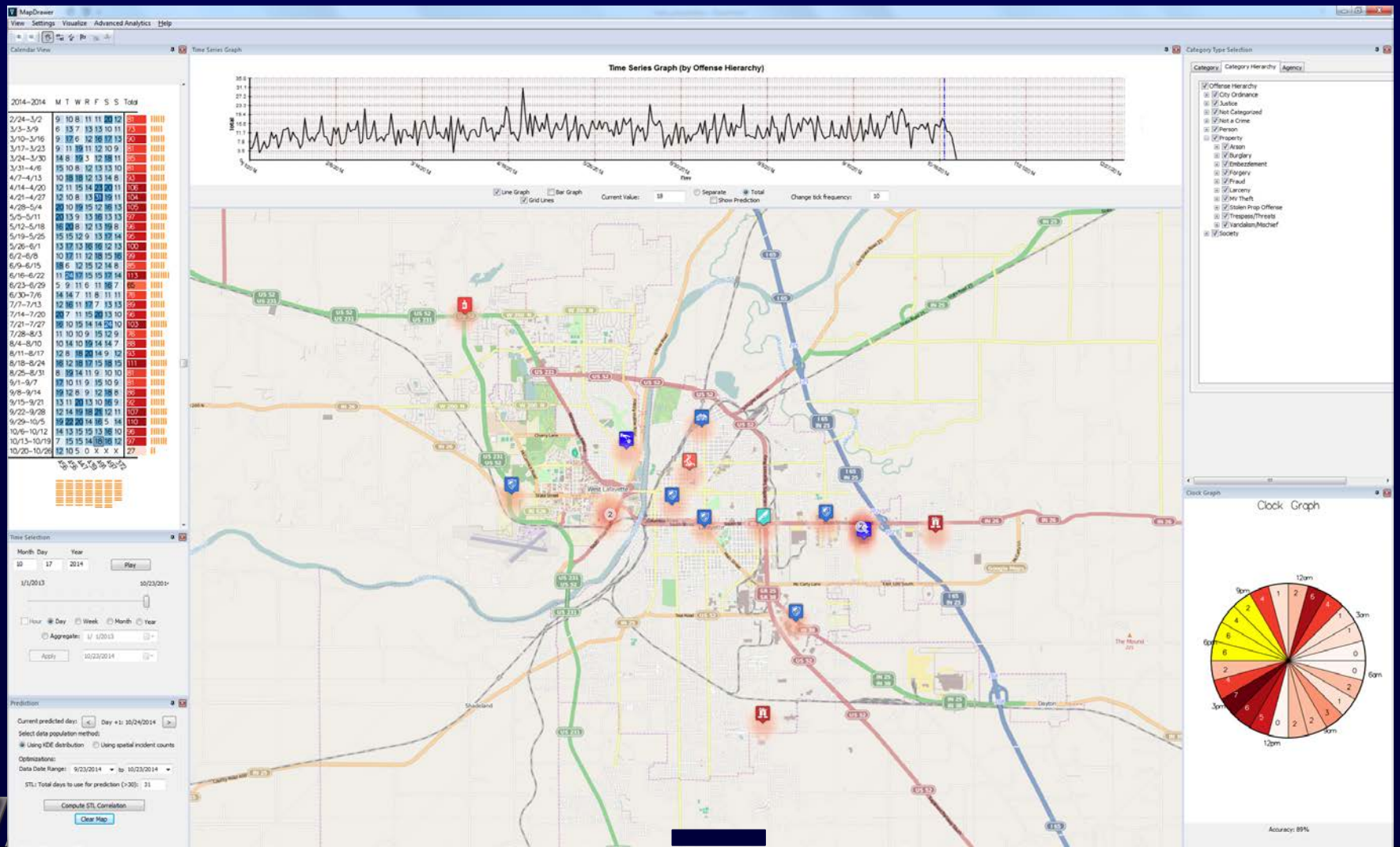
VALET Layout



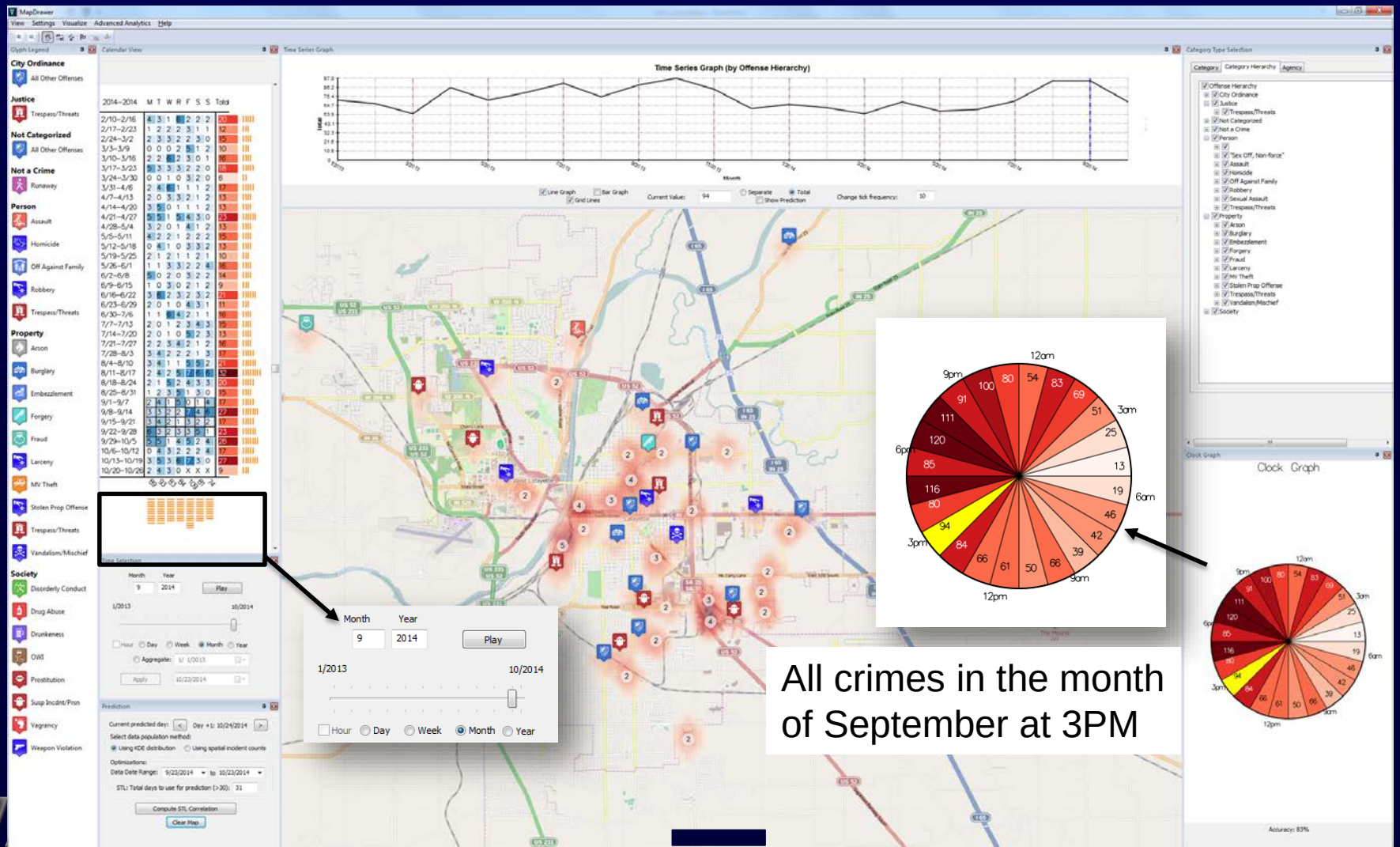
Calendar Display



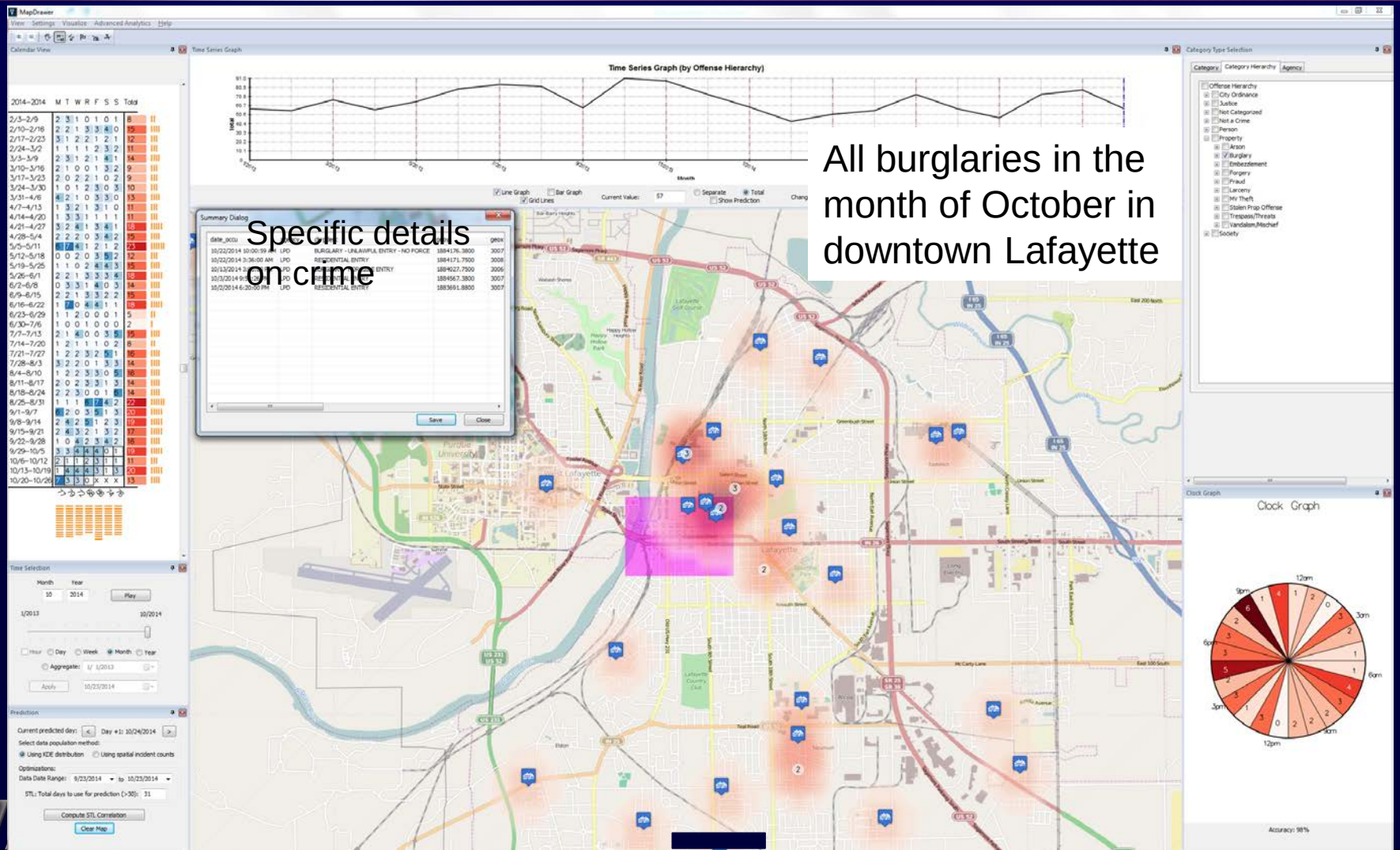
Crimes In A Single Day From 6PM – 9PM



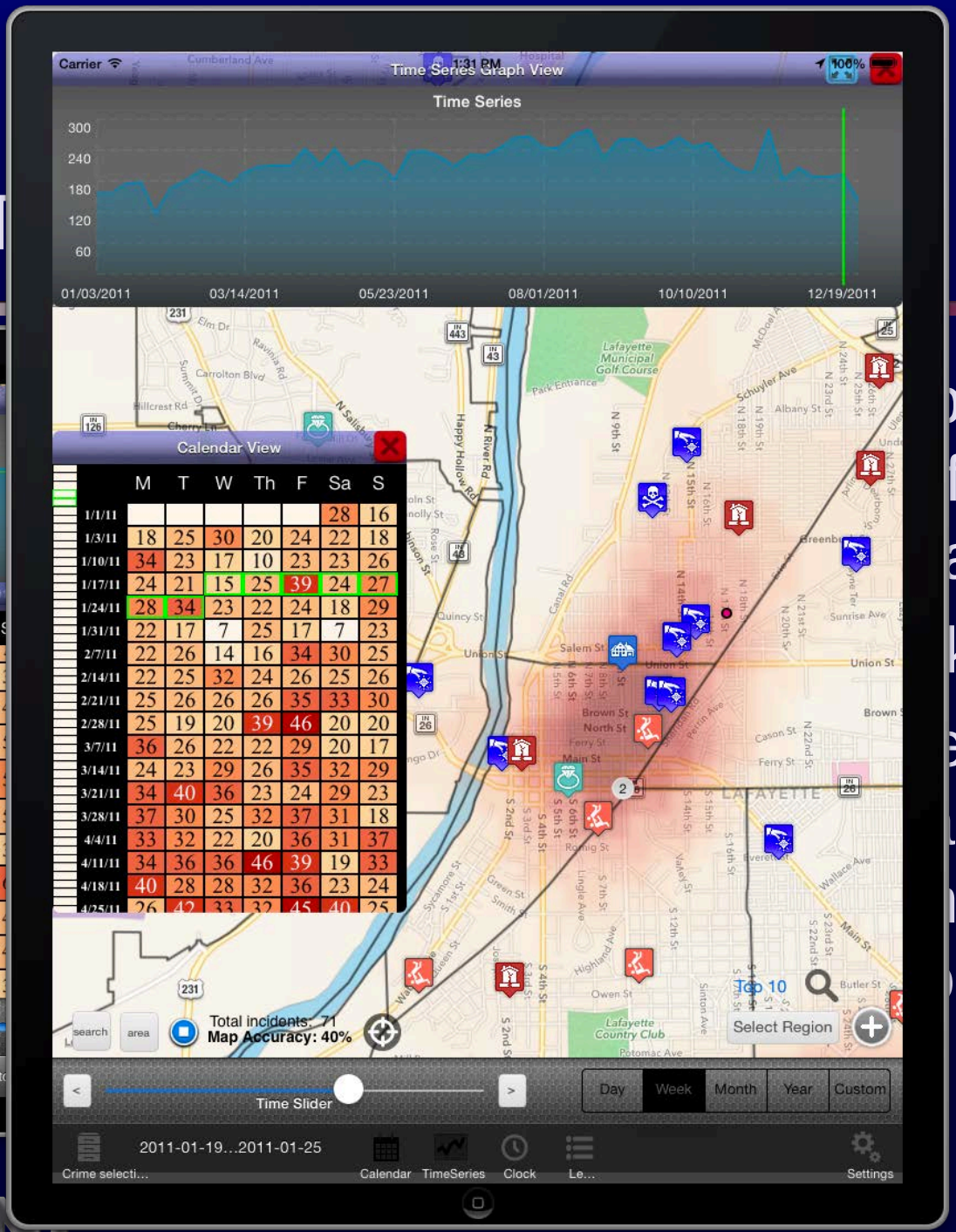
Crimes During Certain Hours In A Month



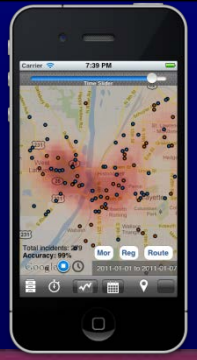
Select Specific Region To Display Crimes

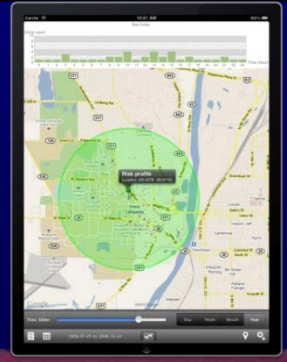


iVALET









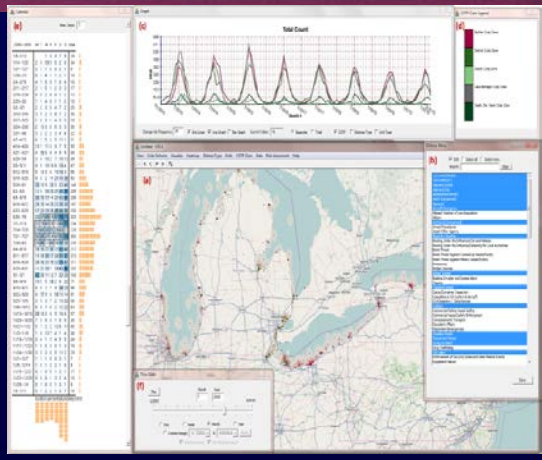


lore criminal,
fic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis

VACCINE

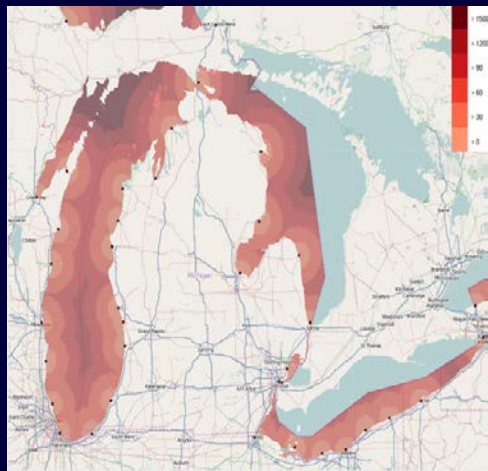
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

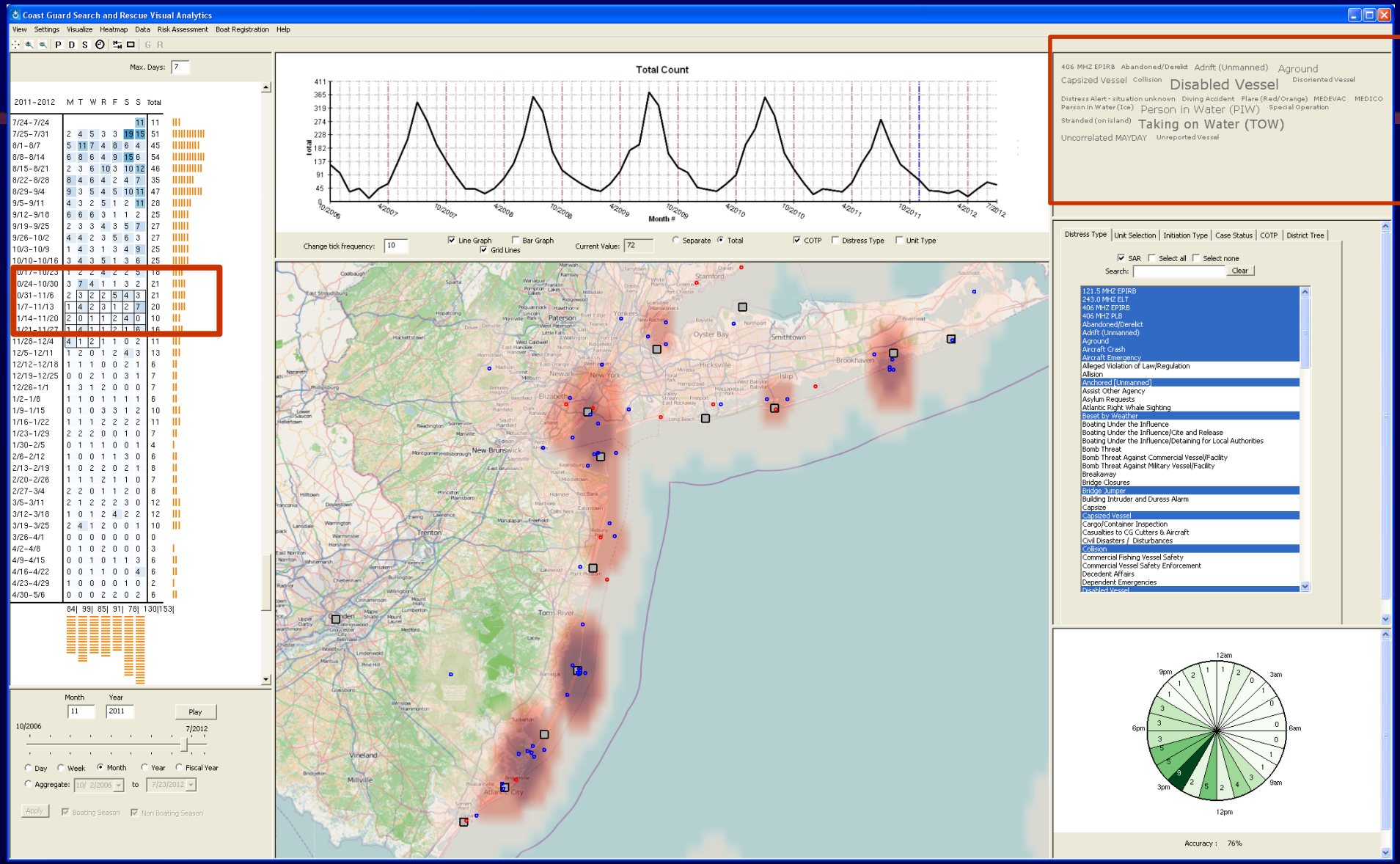
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



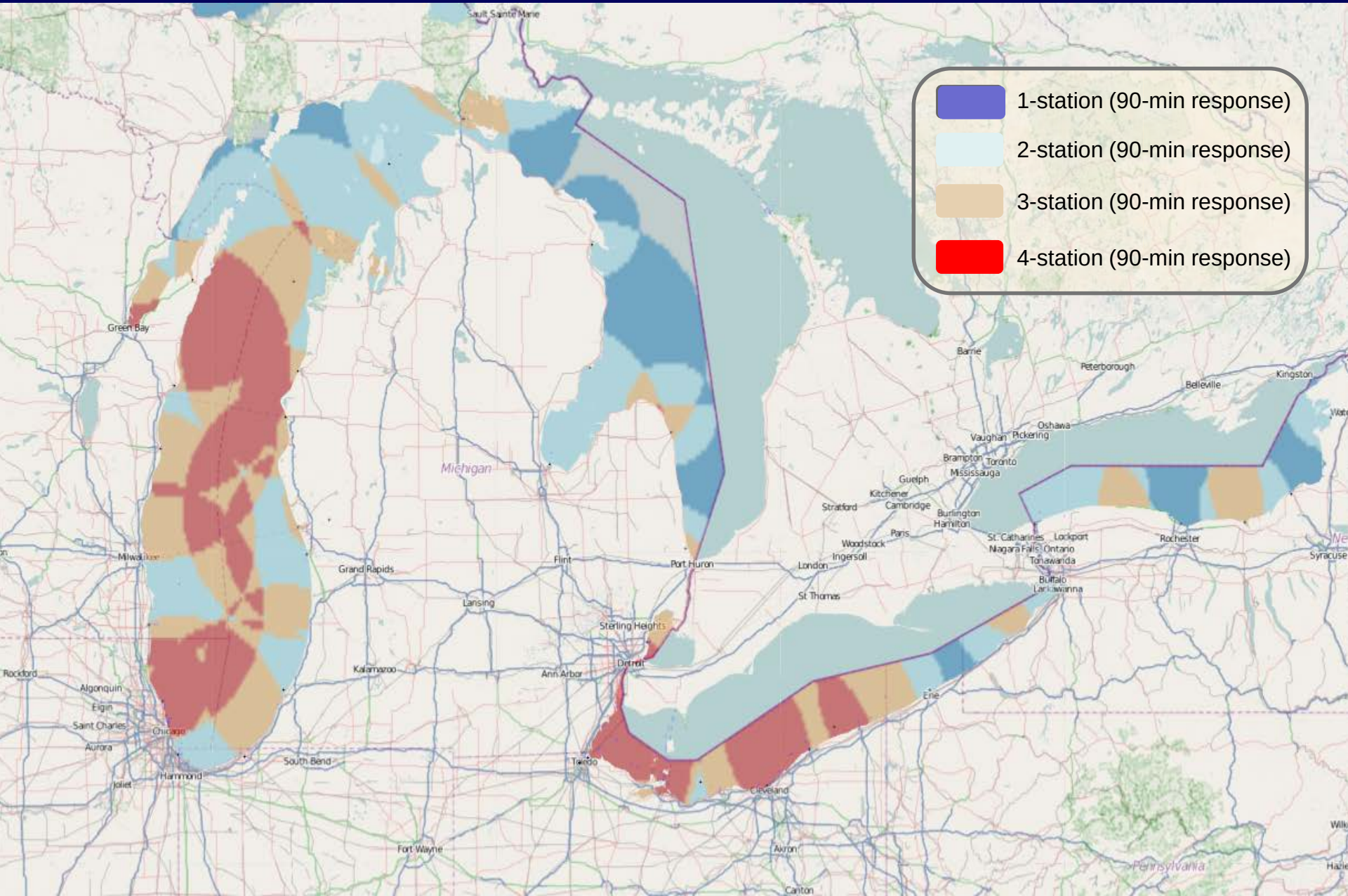
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

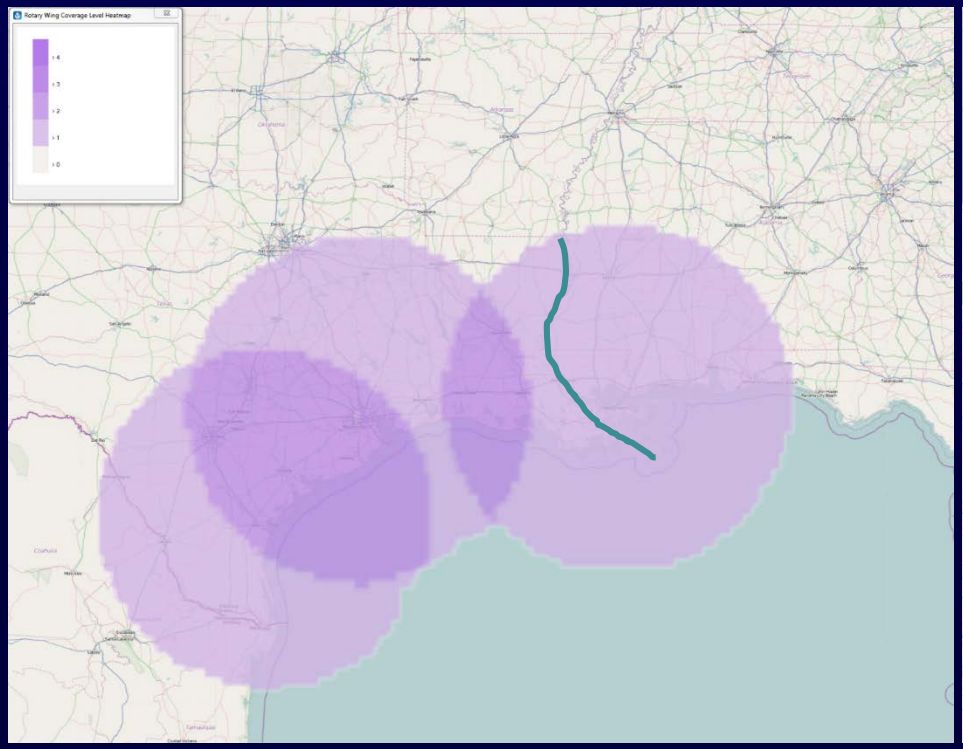


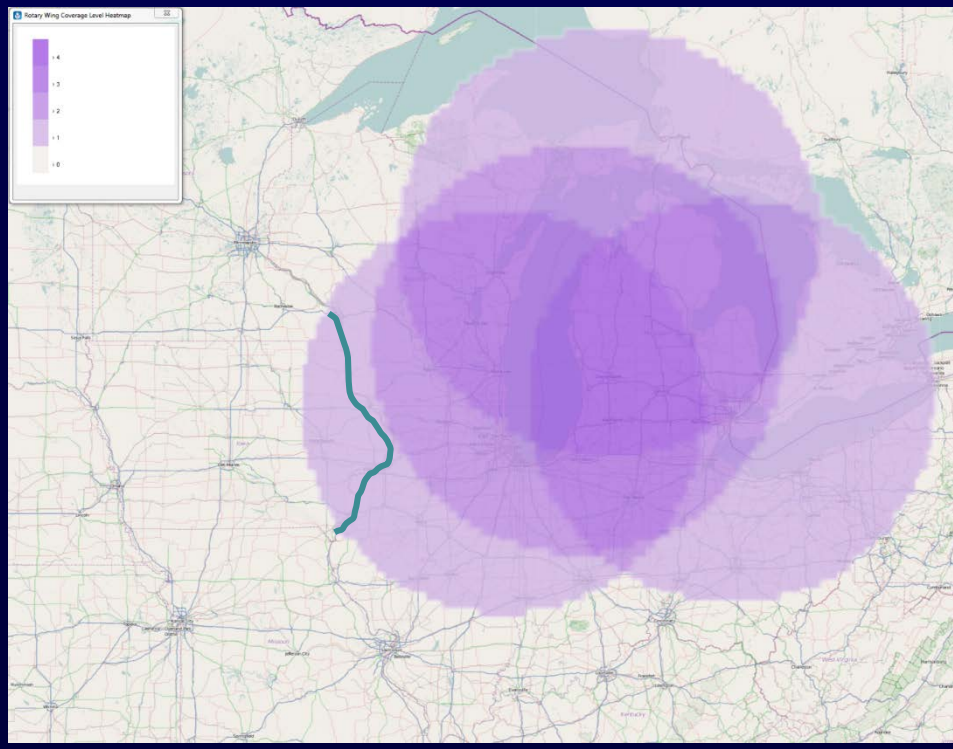
Response Efficiency – Potential Future Assets



Airstation Utility Visualization Comparison

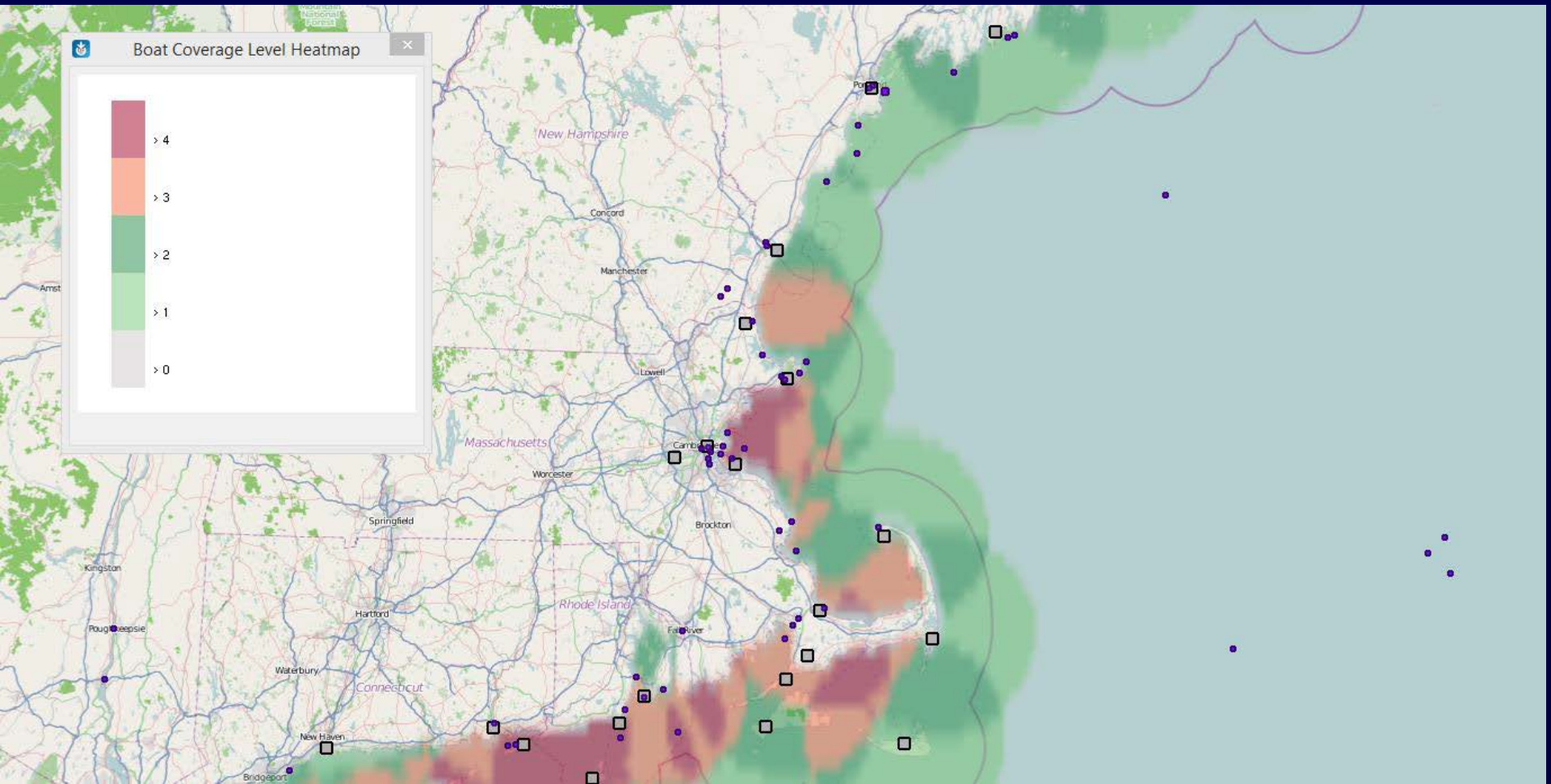
- Rotary Wing Coverage Map by AirSTA





Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007

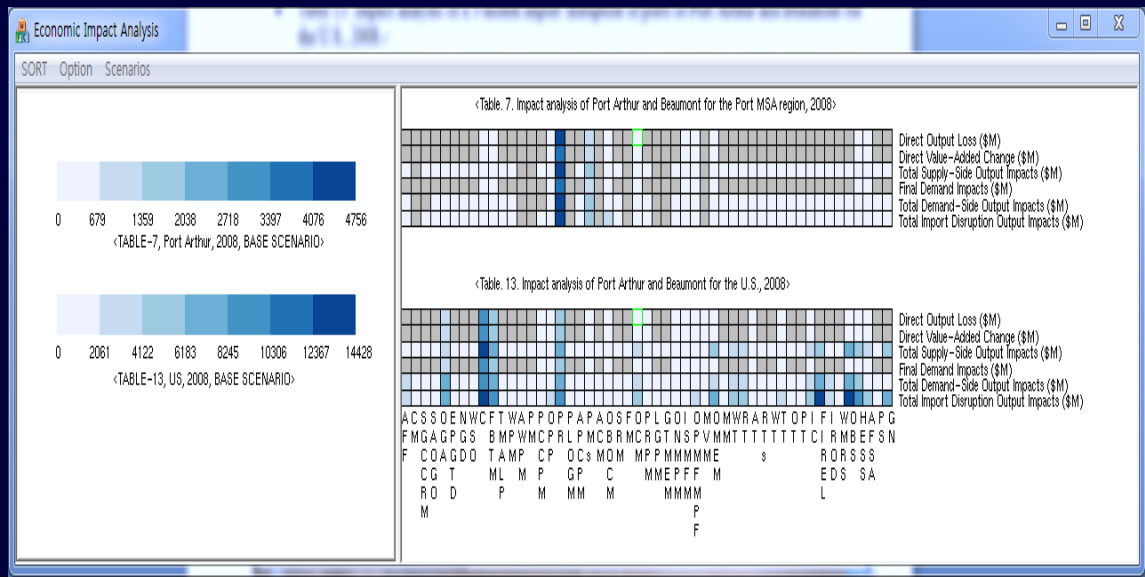


USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



VACCINE

Interactive Operational Performance Assessment Report (iOPAR)

- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels

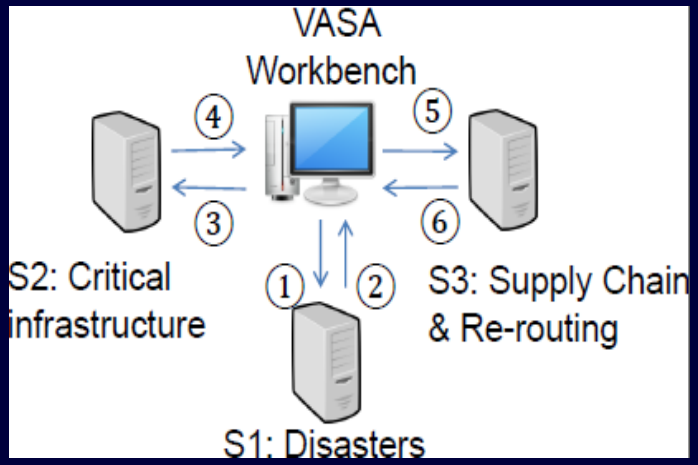
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

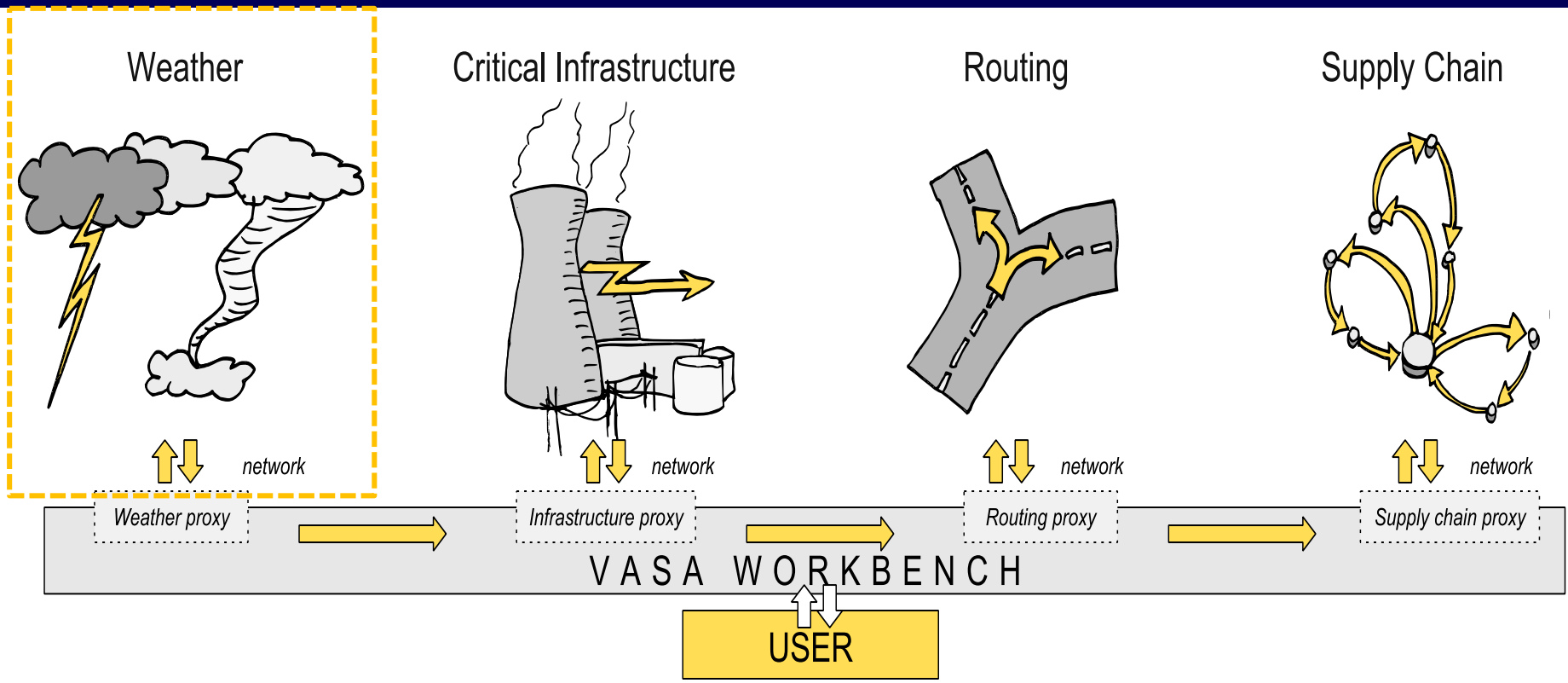
UNCC



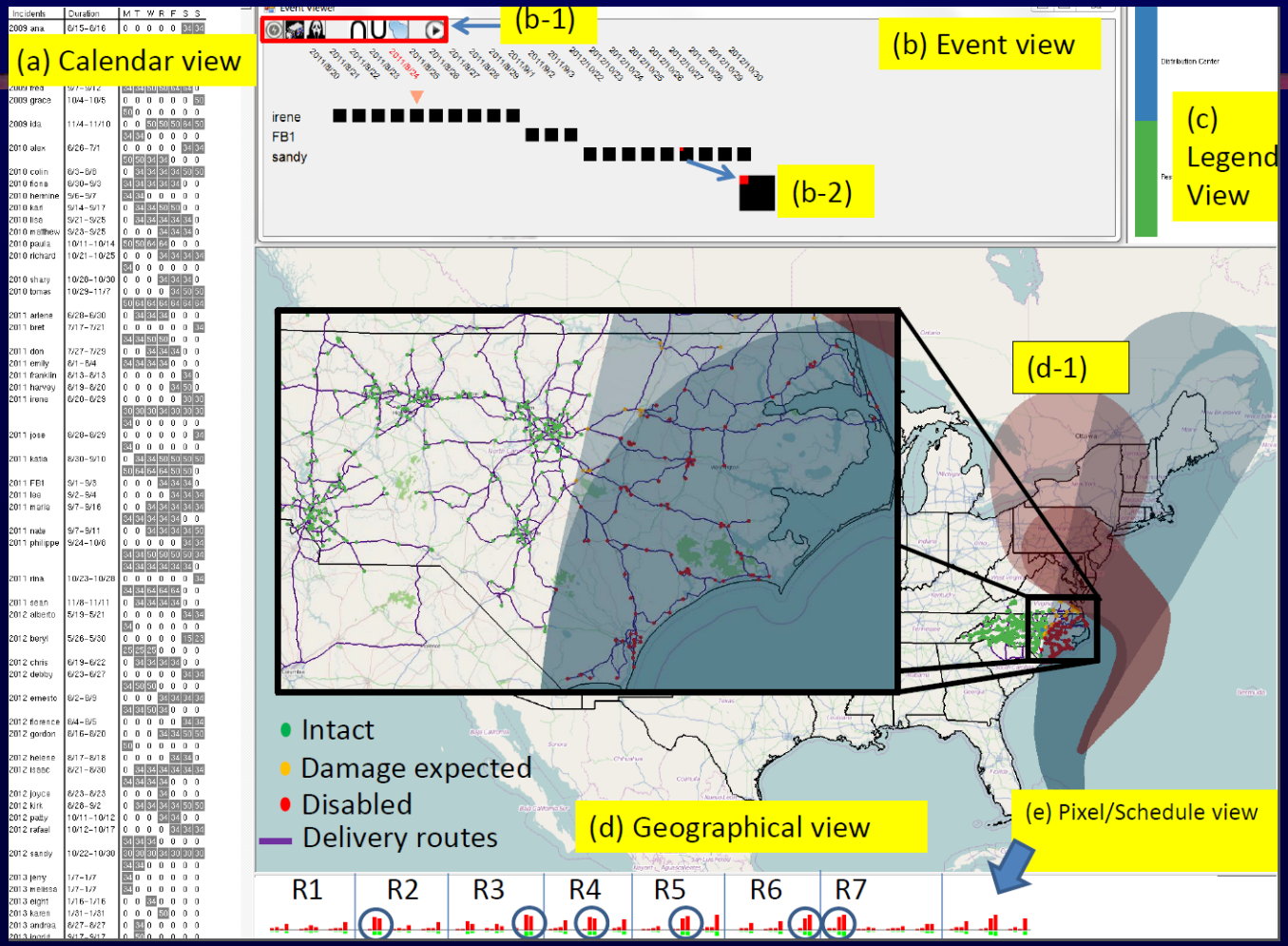
U Minn

TACC

VASA COMPONENT:
CRITICAL INFRASTRUCTURE



System Interface

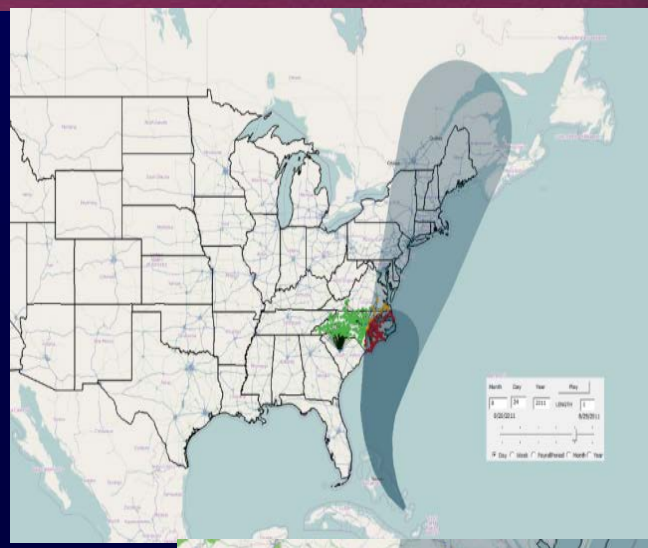


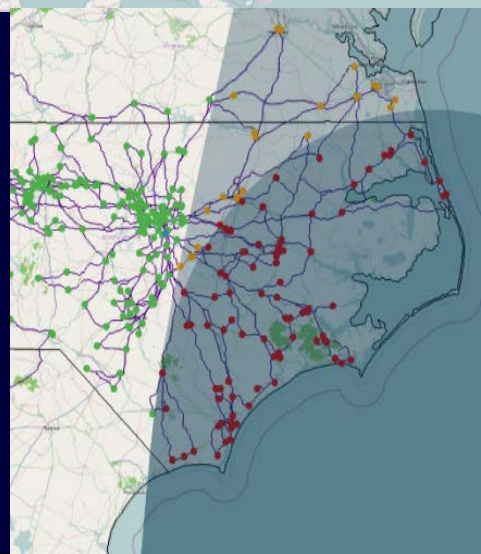
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

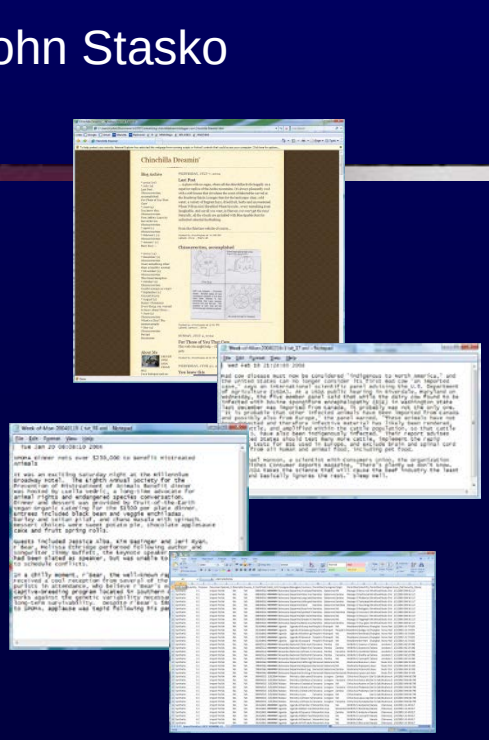
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





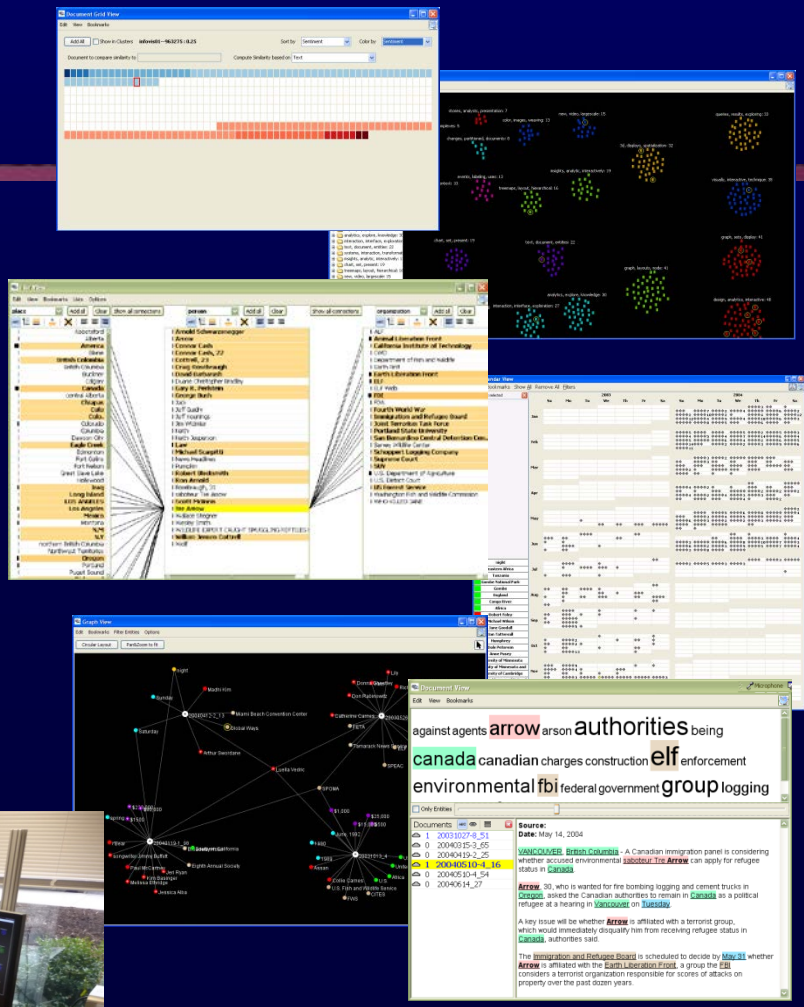
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



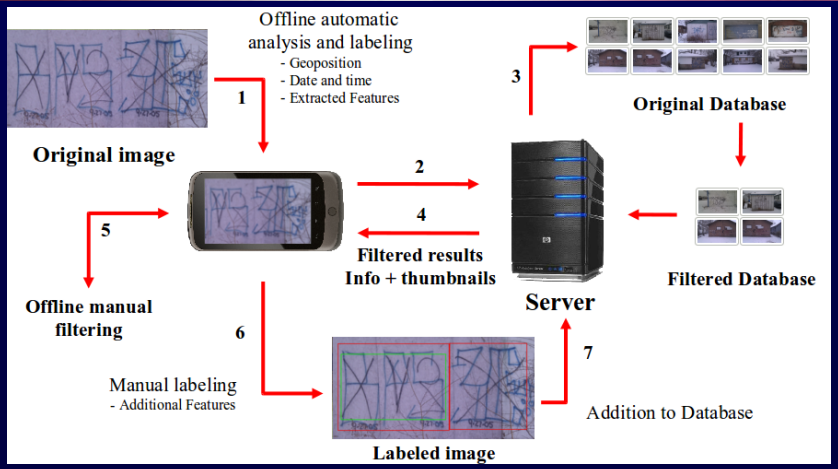
Documents in

Knowledge out





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

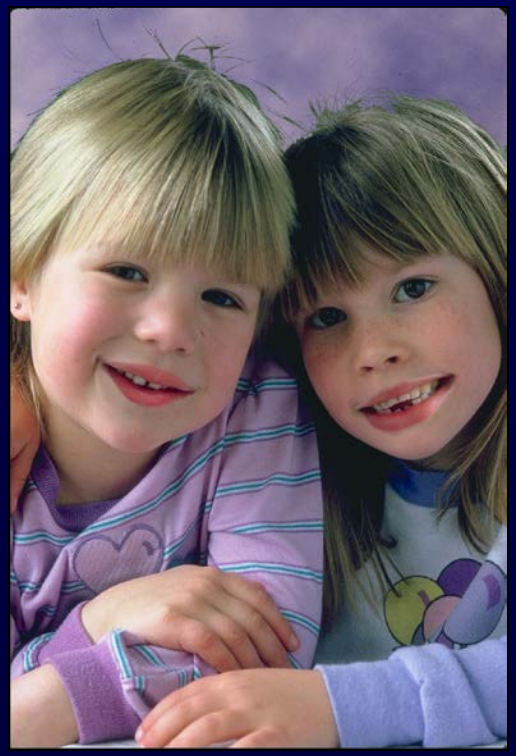
VACCINE

Camera/Image Forensics



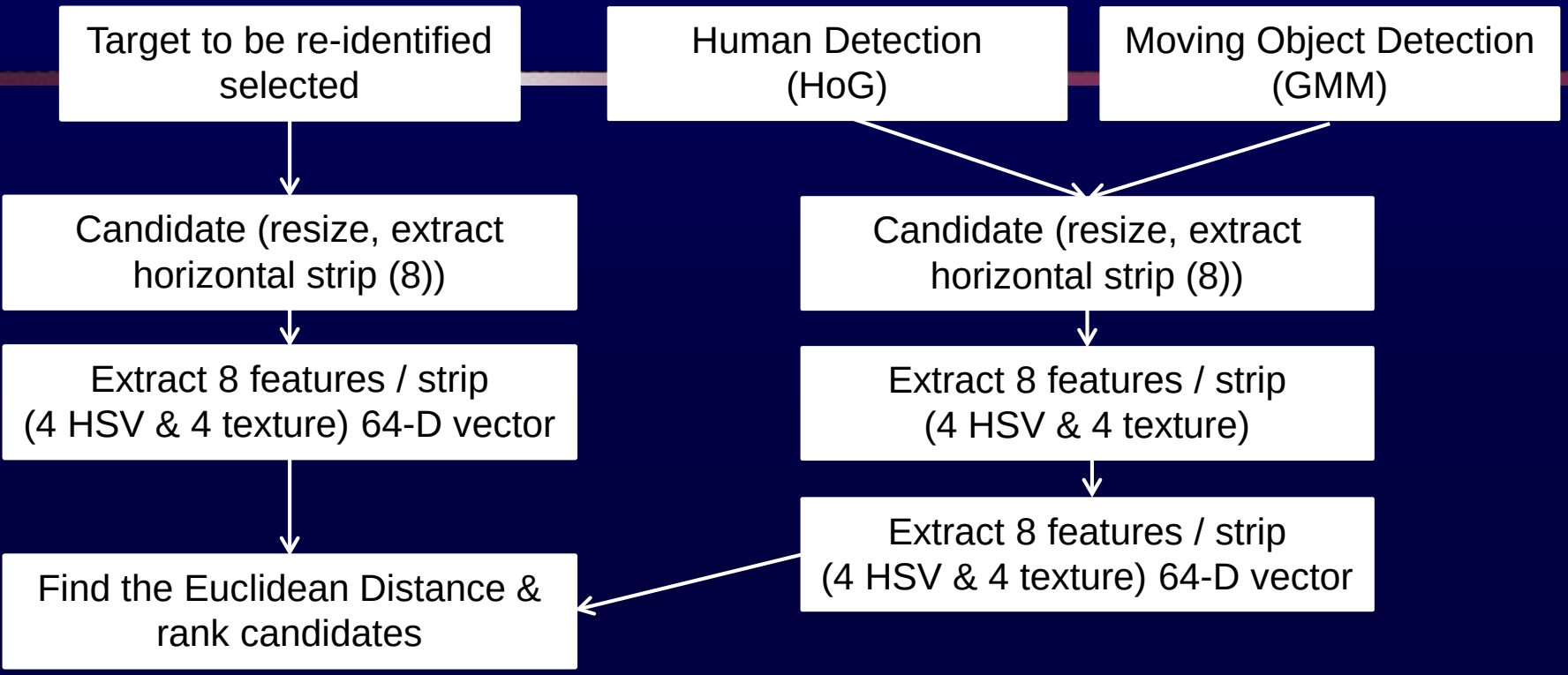
← Original “Girls”

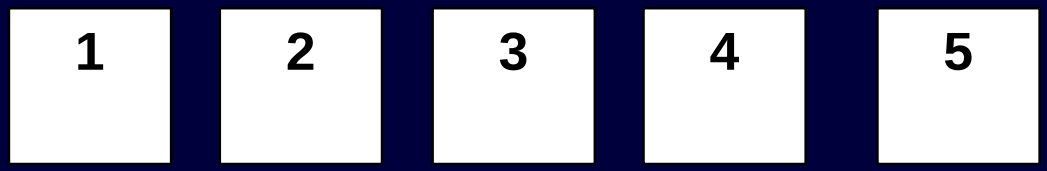
Altered “Girls” →





EBOLO





The top 5 identified targets

EBOLO Analysis



Human Detection (HoG)

Moving Object Detection (GMM)

Candidate (resize, extract horizontal strip (8))

Extract 8 features / strip (4 HSV & 4 texture)

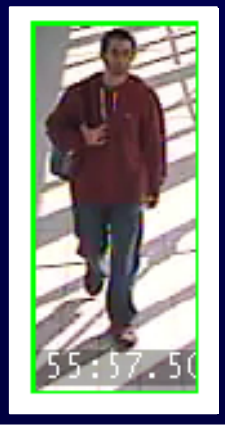
Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

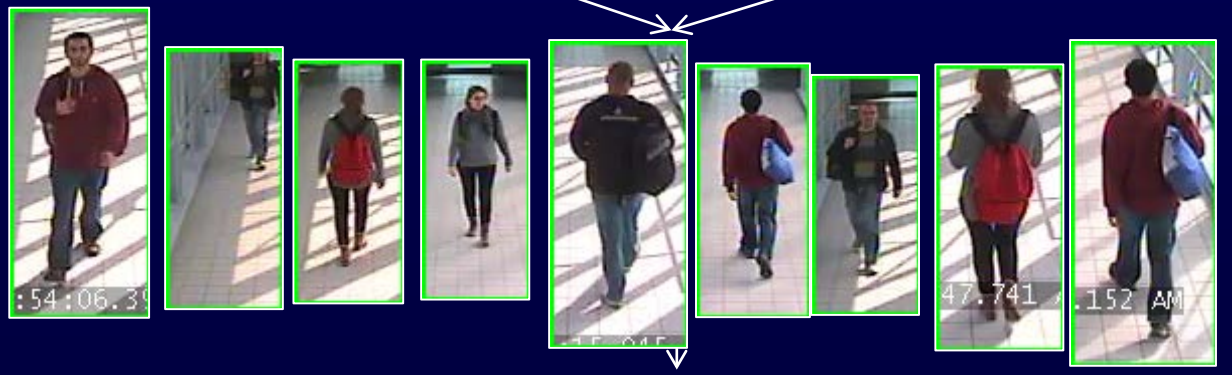
Experimental Results

Target to be re-identified selected

Human Detection (HoG)

Moving Object Detection (GMM)





Extract 8 features / strip (4 HSV & 4 texture) 64-D vector

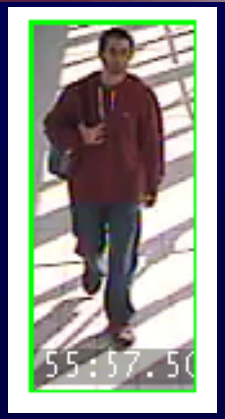
Find the Euclidean Distance & rank candidates

1	2	3	4	5

The top 5 identified targets

Experimental Results

Target to be re-identified selected



Find the Euclidean Distance & rank candidates





The top 5 identified targets

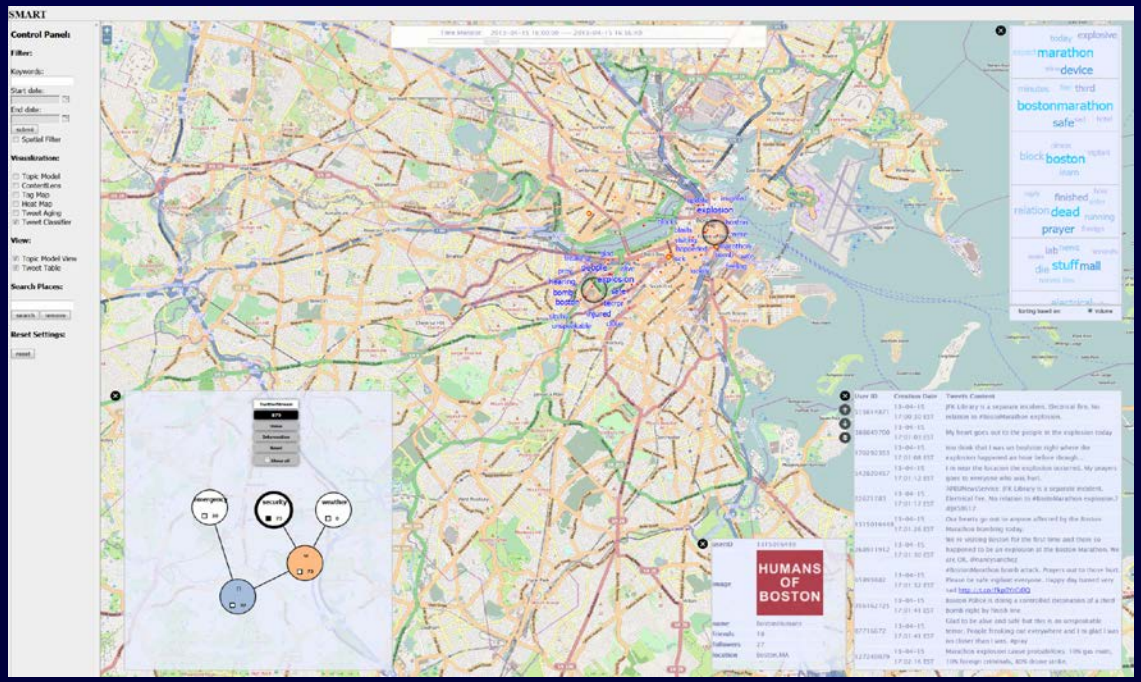
VIPER Lab

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

December 2014

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

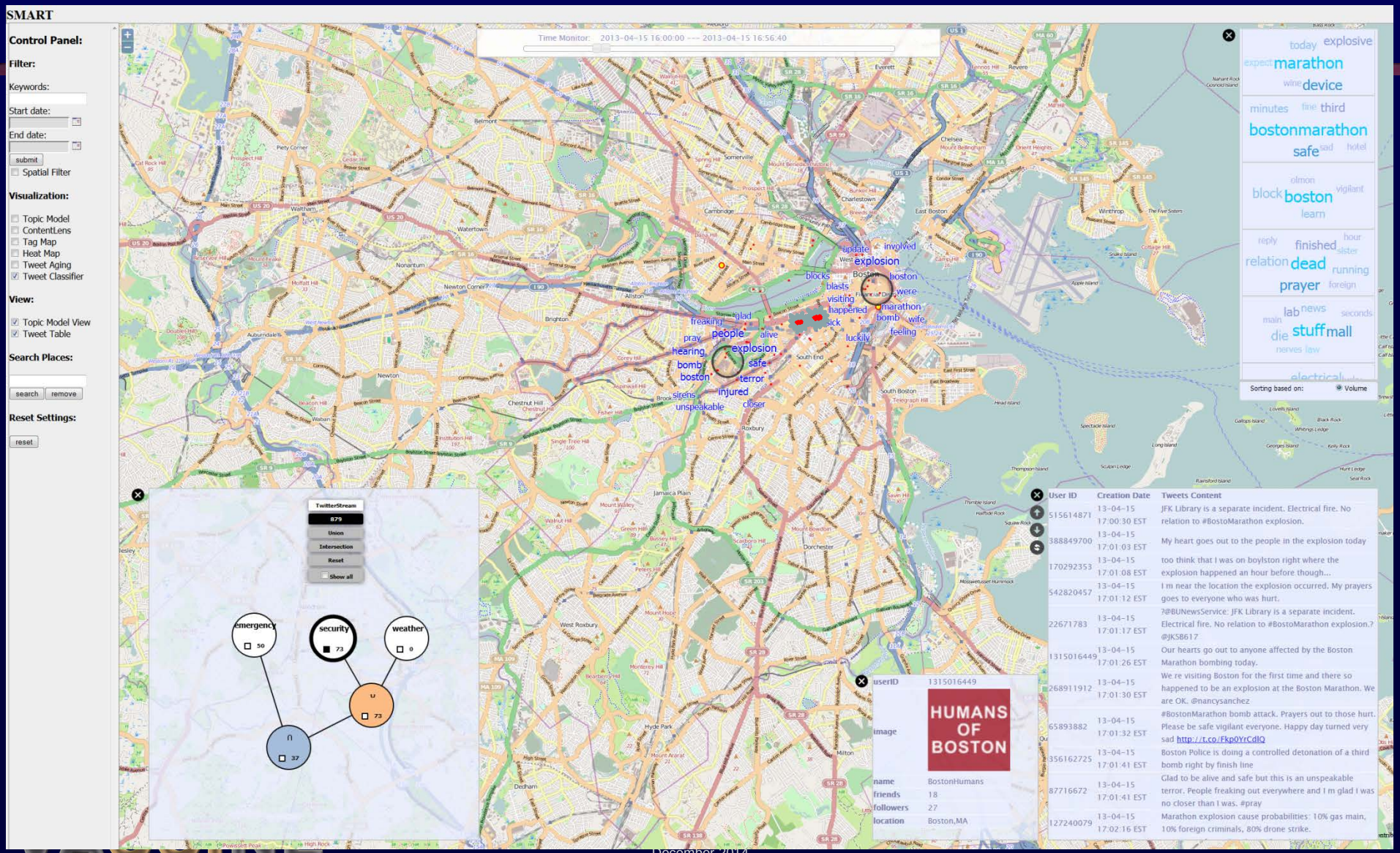
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

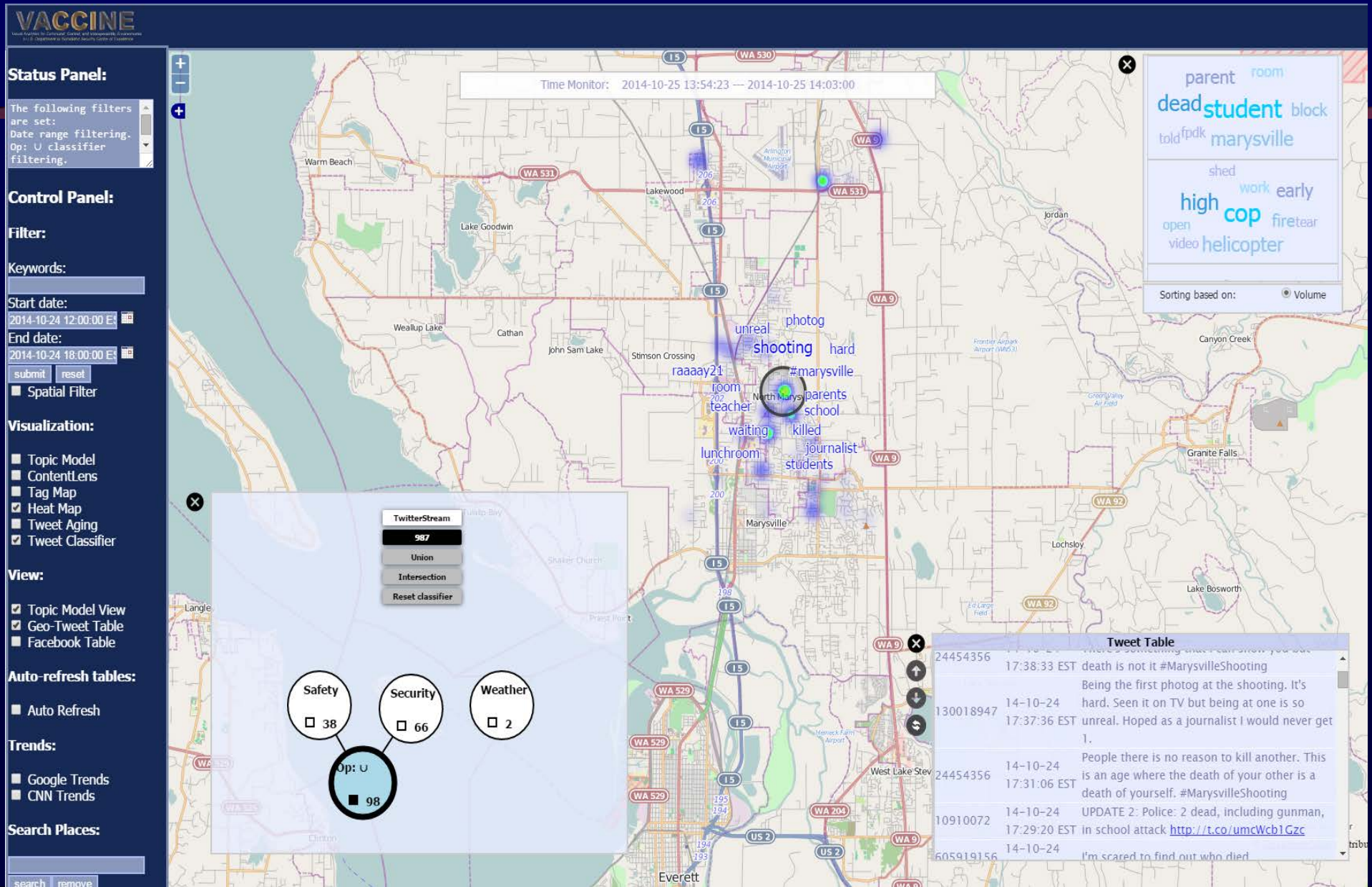


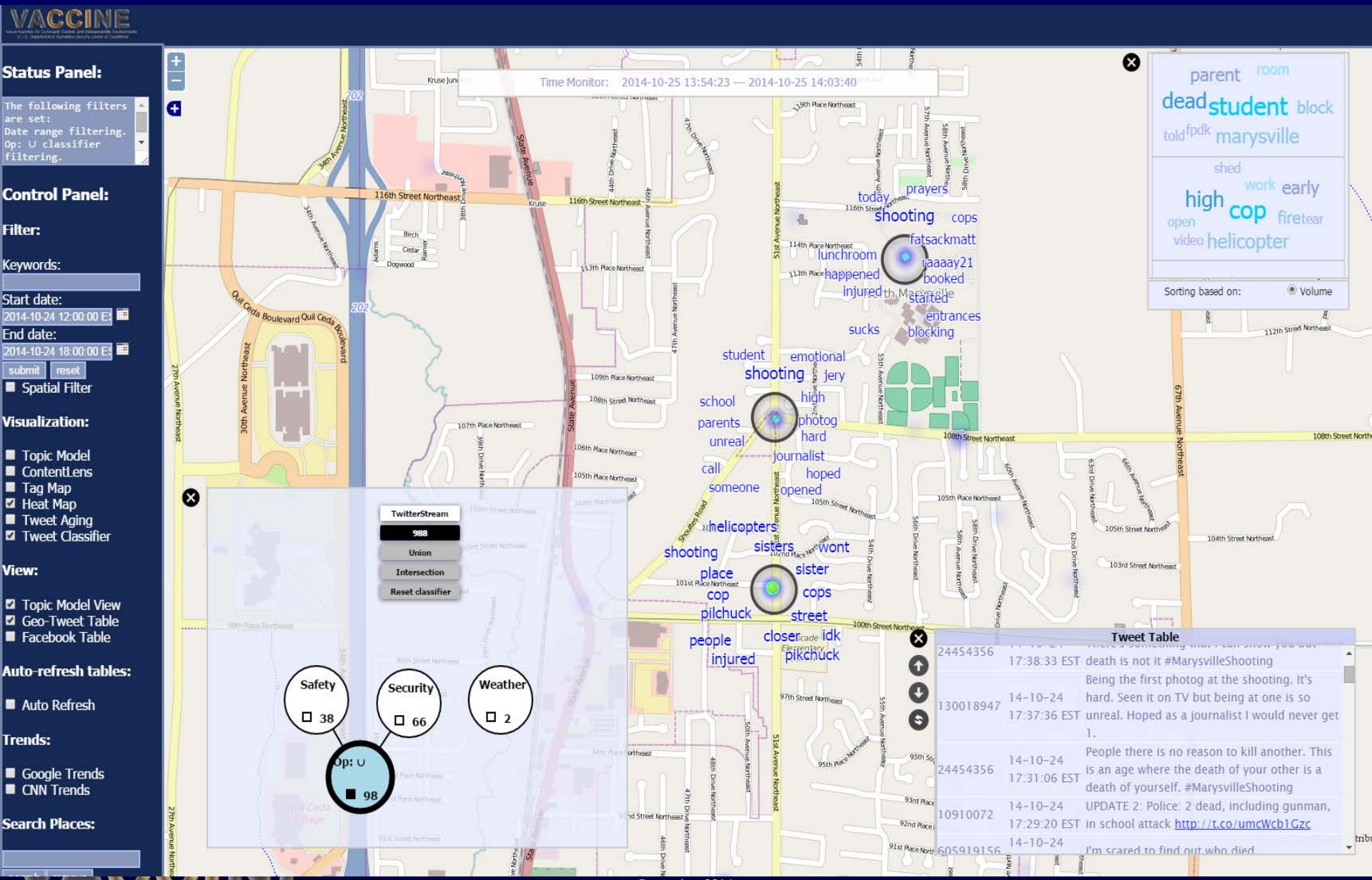
VACCINE

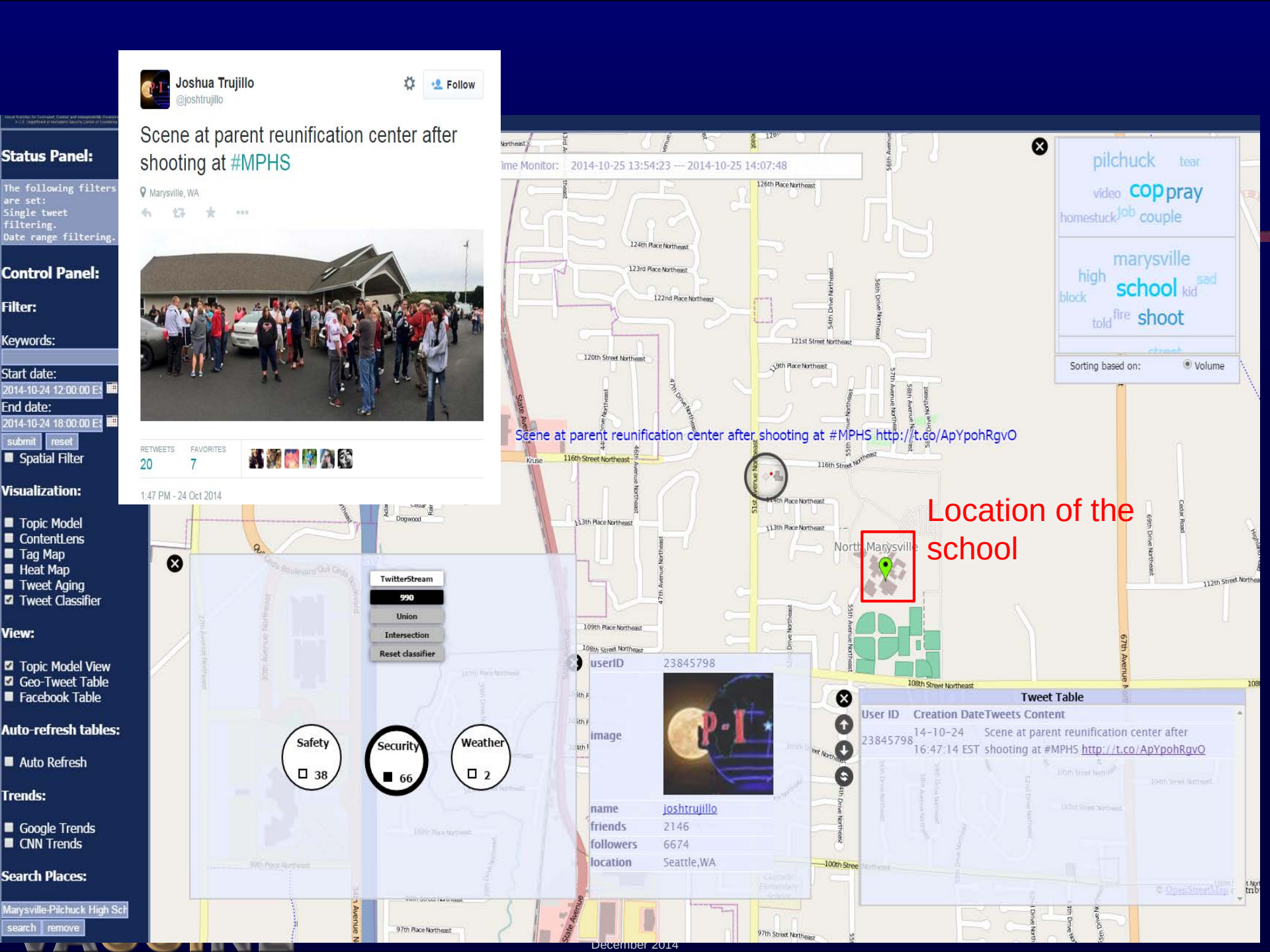
Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24
17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24
17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24
17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died


Joshua Trujillo
@joshtrujillo
Follow
Scene at parent reunification center after shooting at #MPHS
Marysville, WA
RETWEETS
20
FAVORITES
7
1:47 PM - 24 Oct 2014
Status Panel:
The following filters are set:
Single tweet filtering.
Date range filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 ES
End date:
2014-10-24 18:00:00 ES
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Marysville-Pilchuck High Sch
search
remove
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48
pilchuck
tear
video
cop
pray
homestuck
job
couple
marysville
high
school
kid
sad
block
told
fire
shoot
Sorting based on:
Volume
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO
Location of the school
North Marysville
TwitterStream
990
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
userID
23845798
image
name
joshtrujillo
friends
2146
followers
6674
location
Seattle,WA
Tweet Table
User ID
Creation Date
Tweets Content
23845798
14-10-24 16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

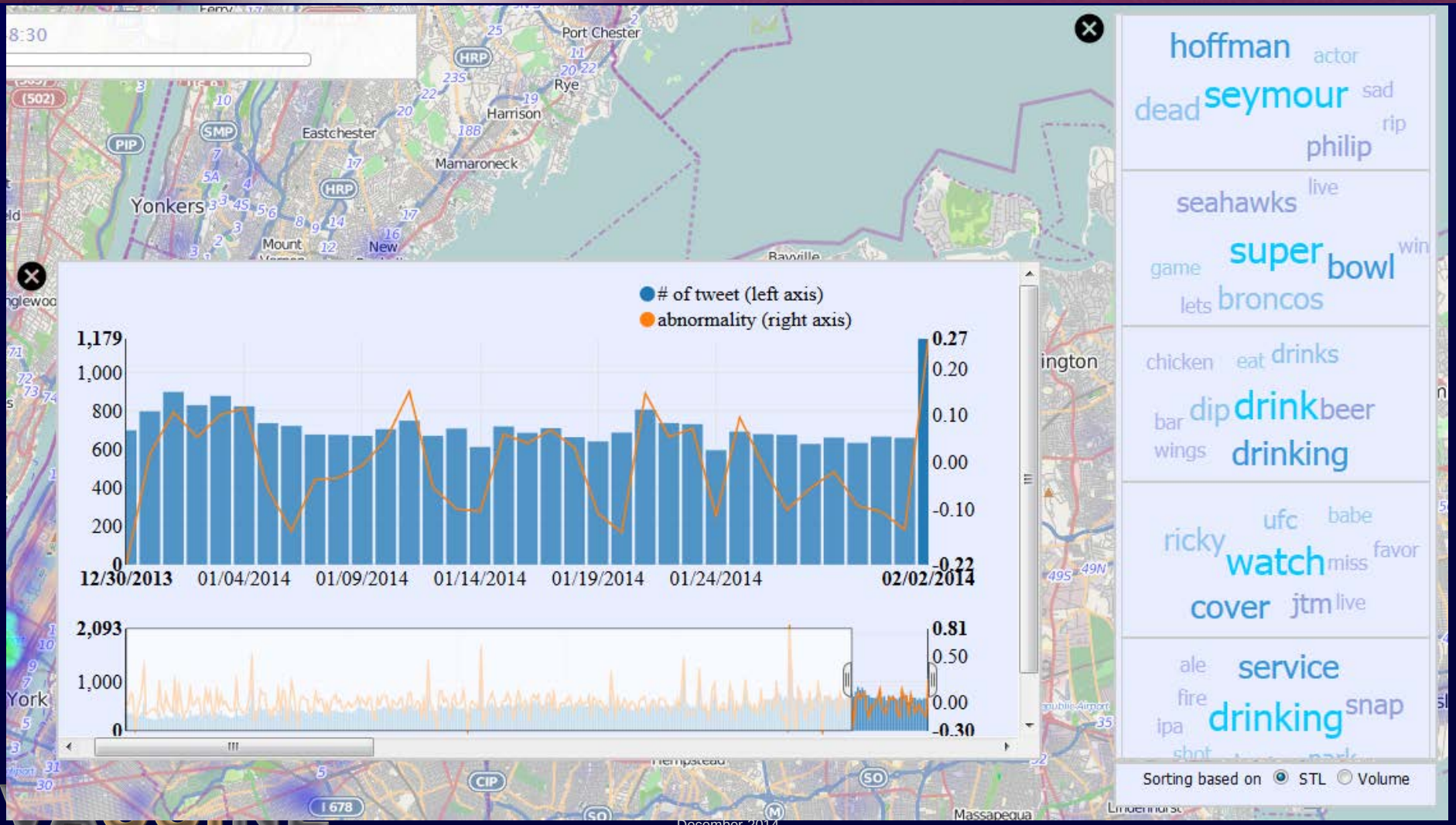
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring event
  - Purdue football games – the p
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – L



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More



RETWEETS 6

FAVORITES 4

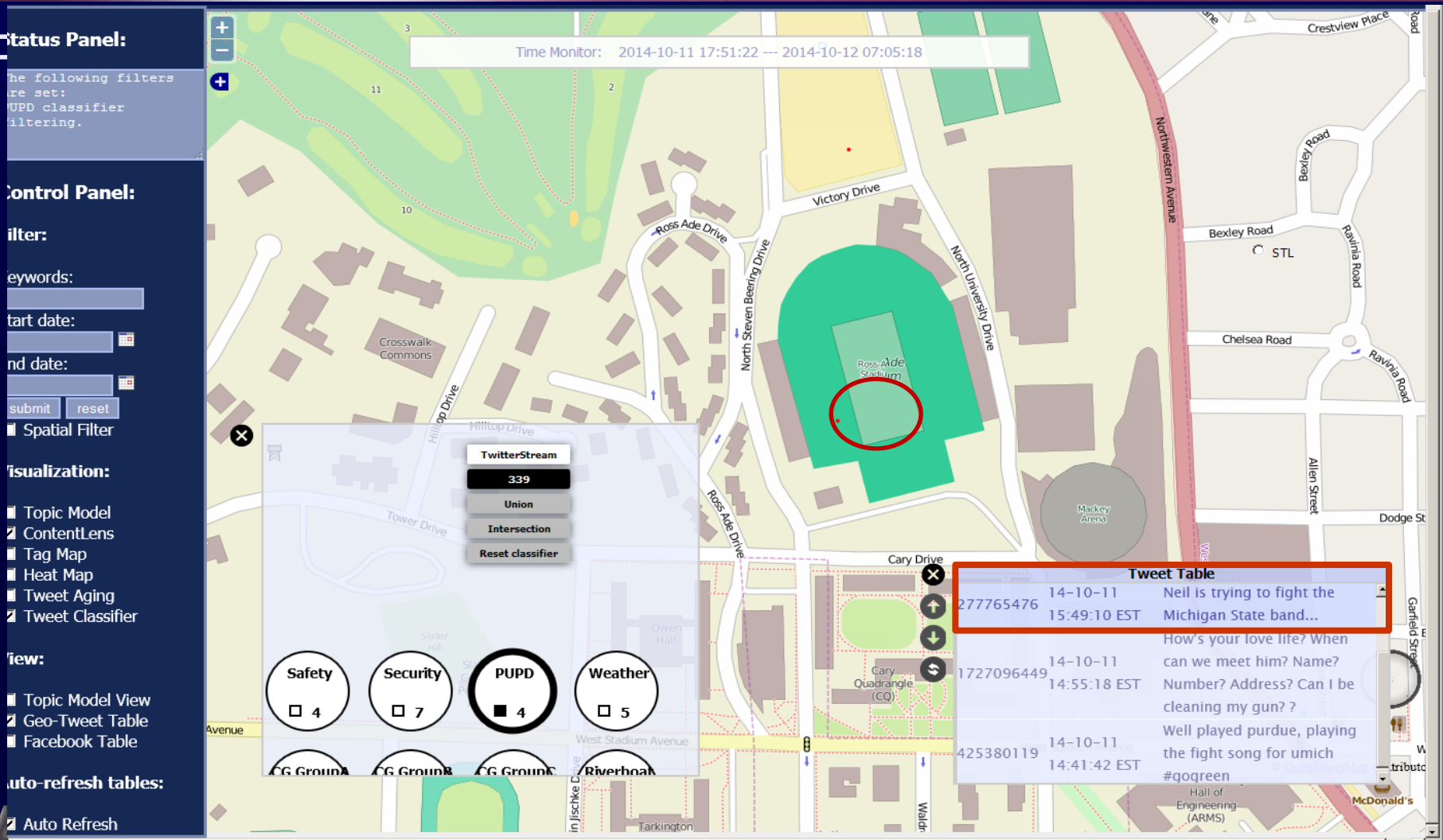
12:56 PM - 12 Oct 2014

Where Can It Be Helpful?

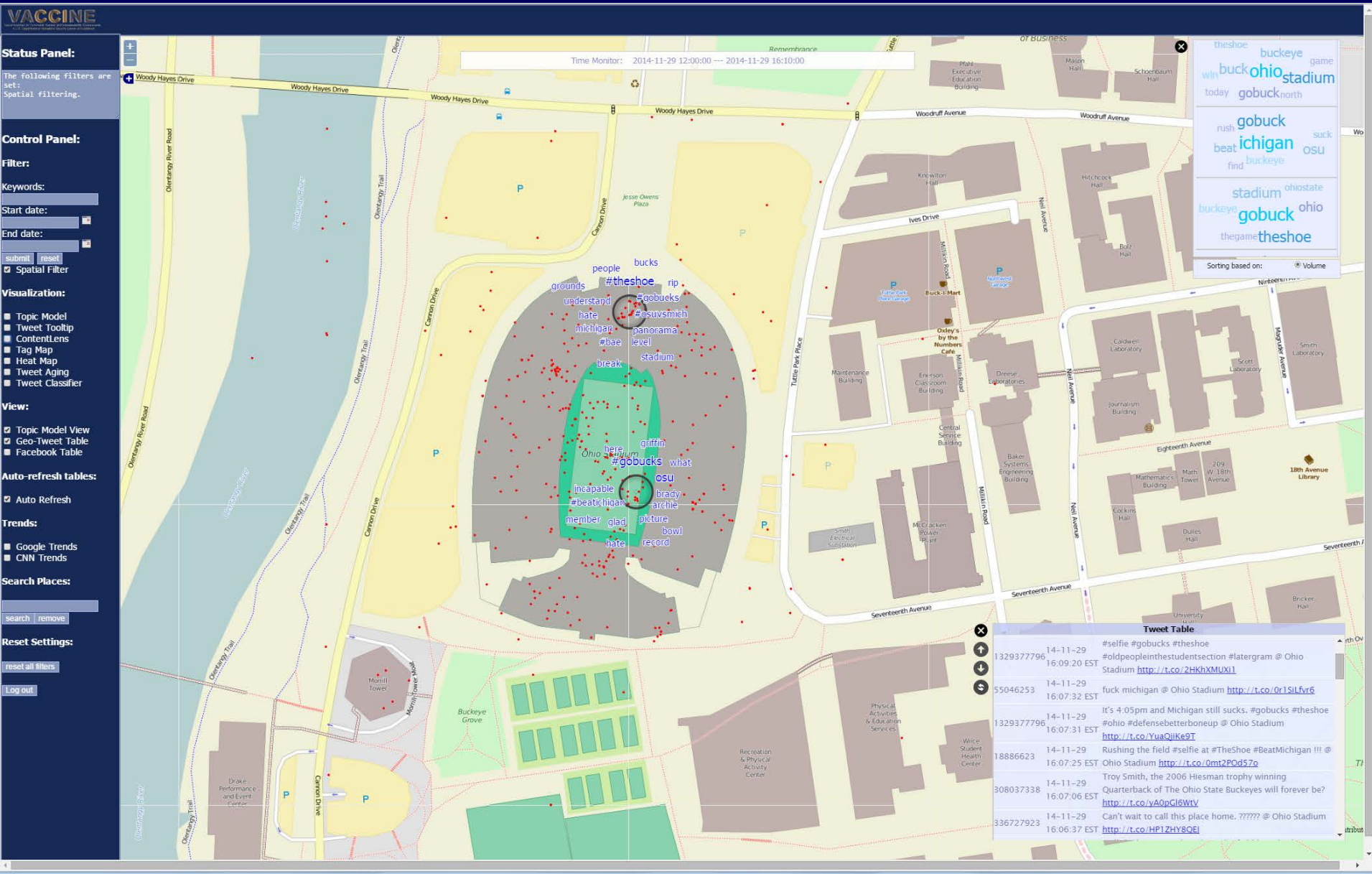
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



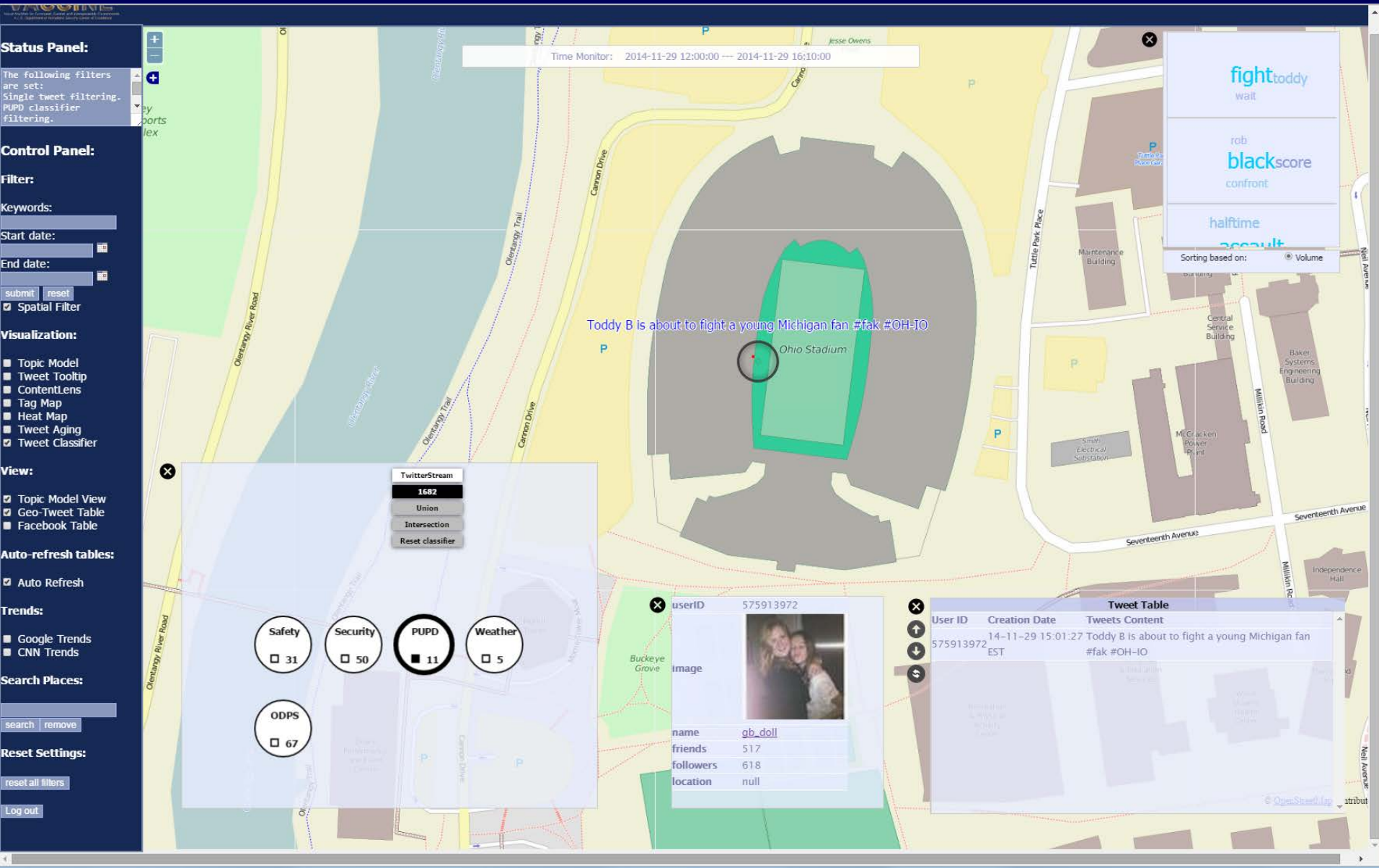
Example From Ohio State Football 11/29/14



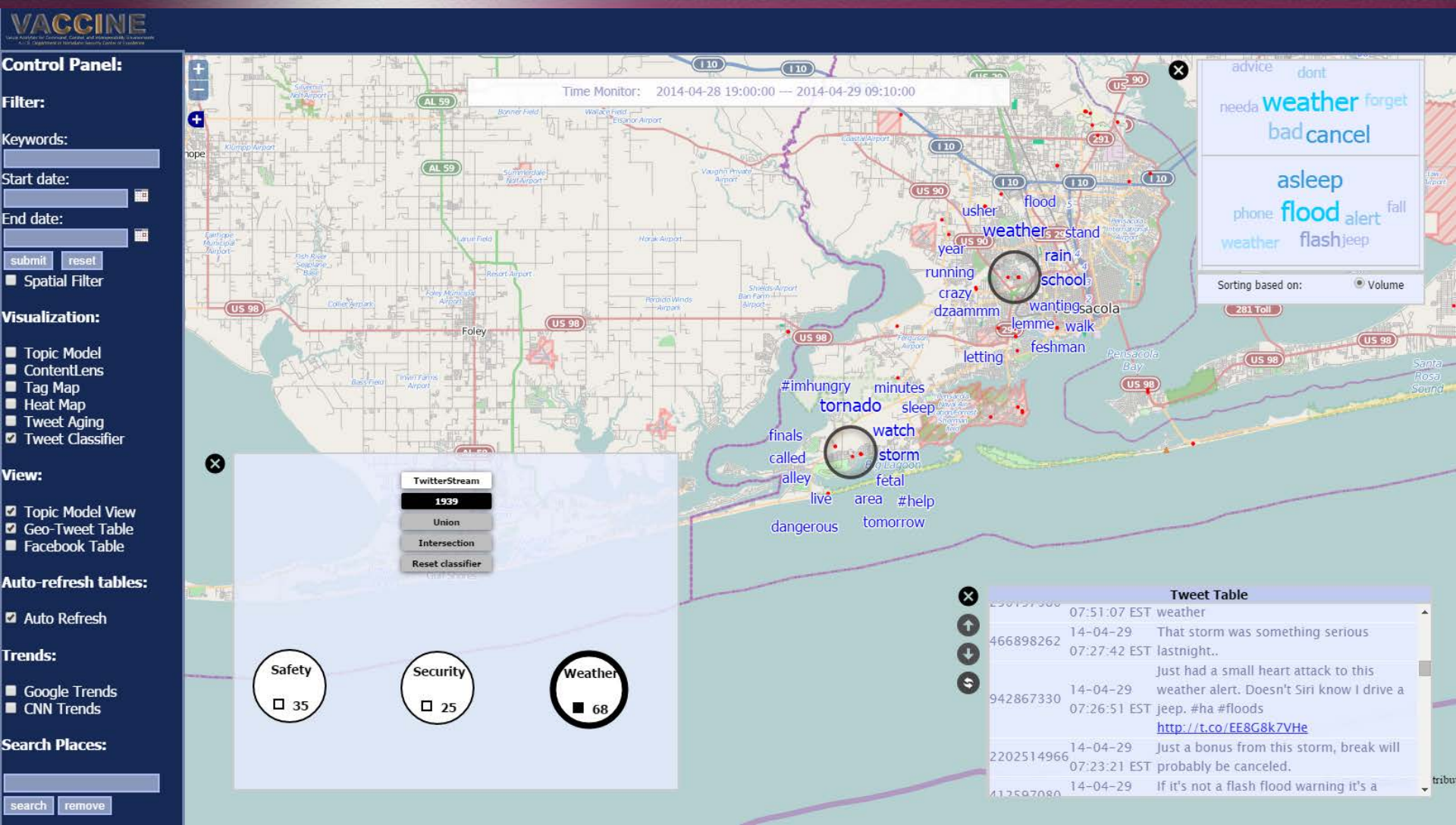
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



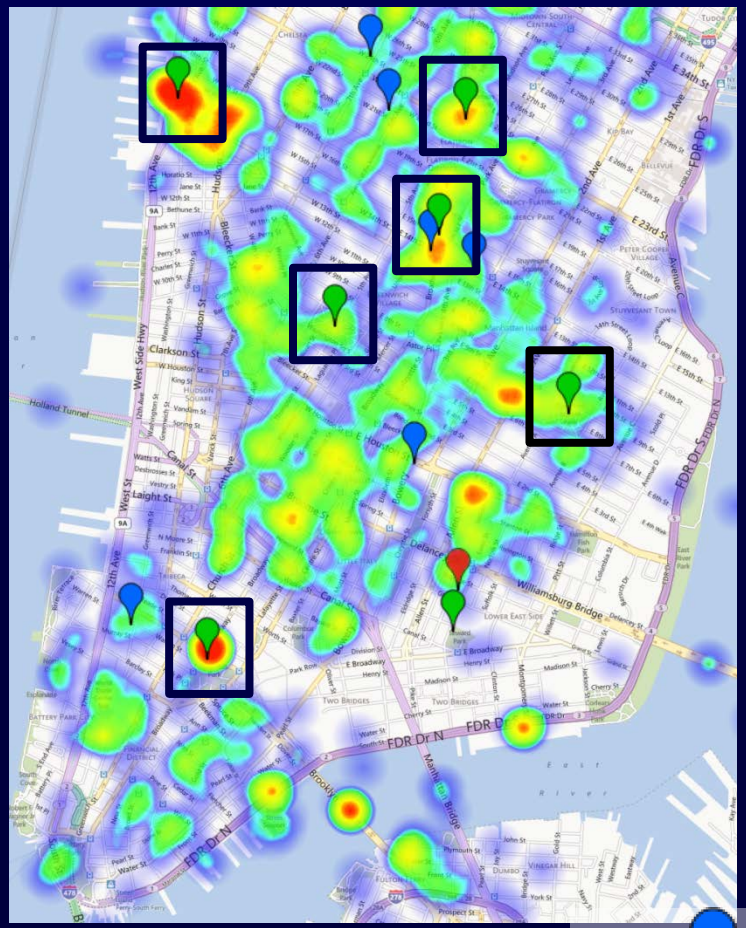
Where Can It Be Helpful?

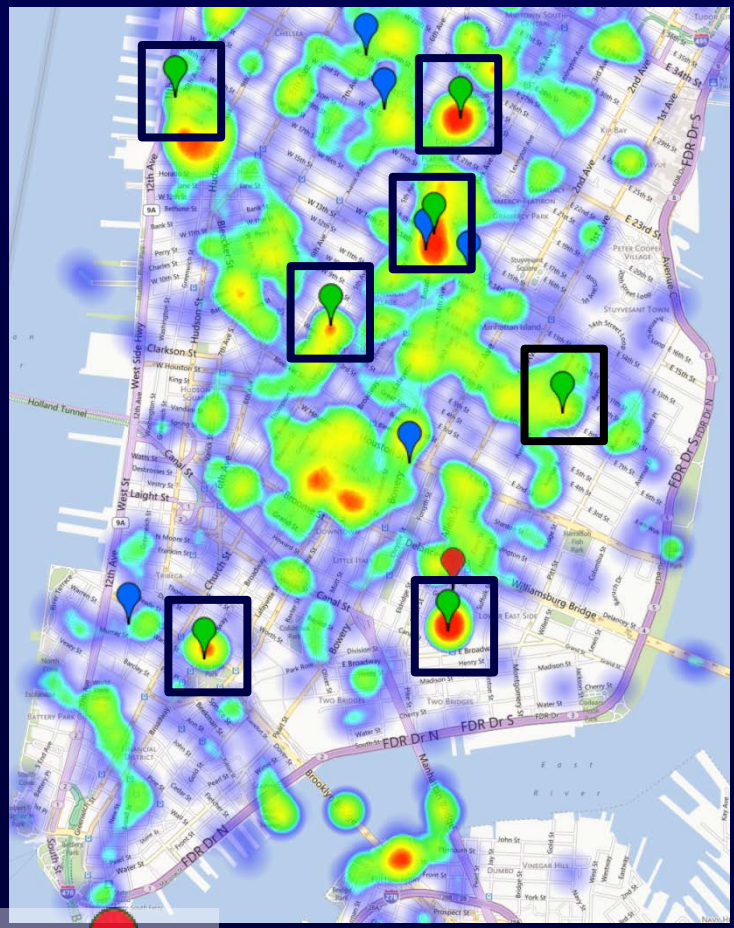


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

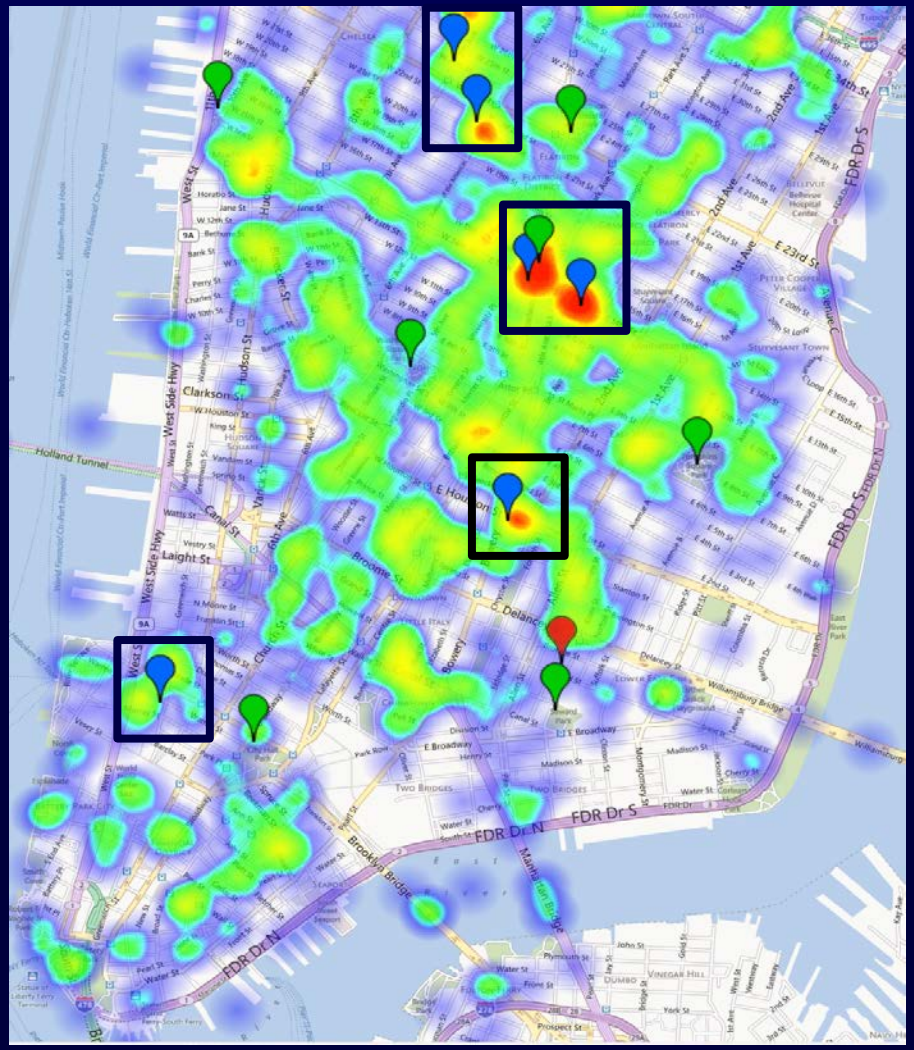
December 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

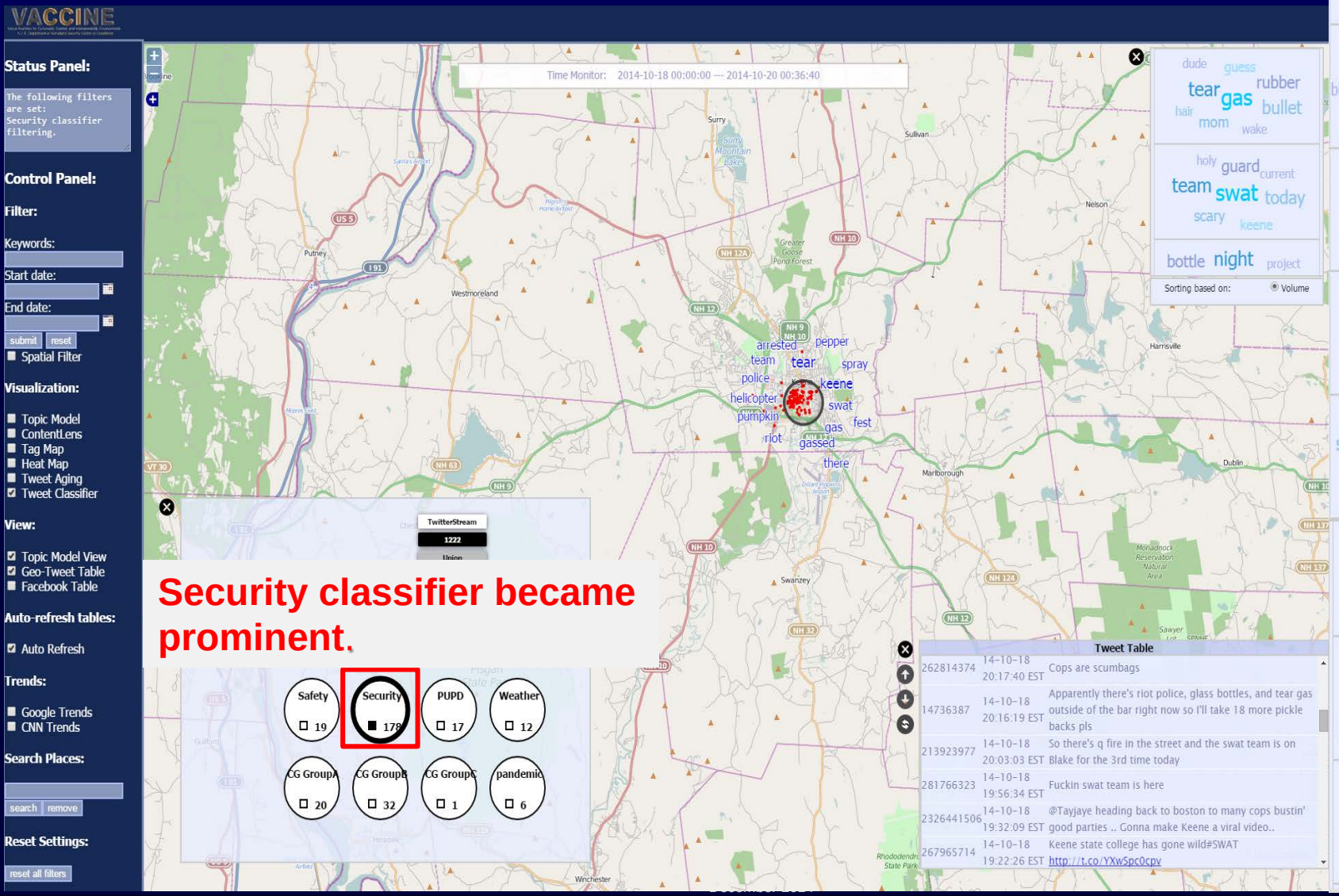
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



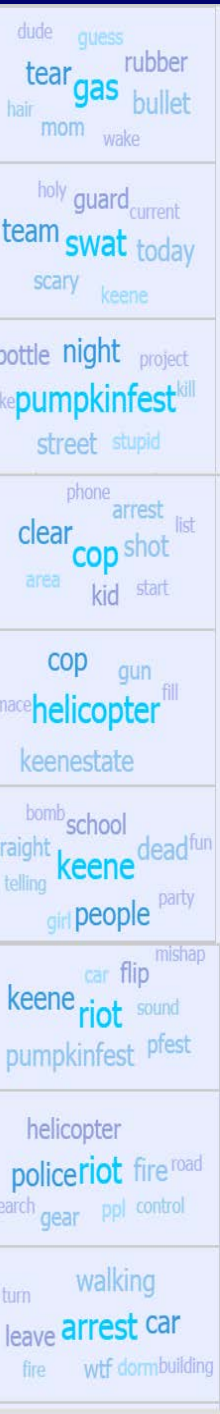
VACCINE

Keene, NH Pumpkin festival 10/18/14 – 10/19/14 Overview

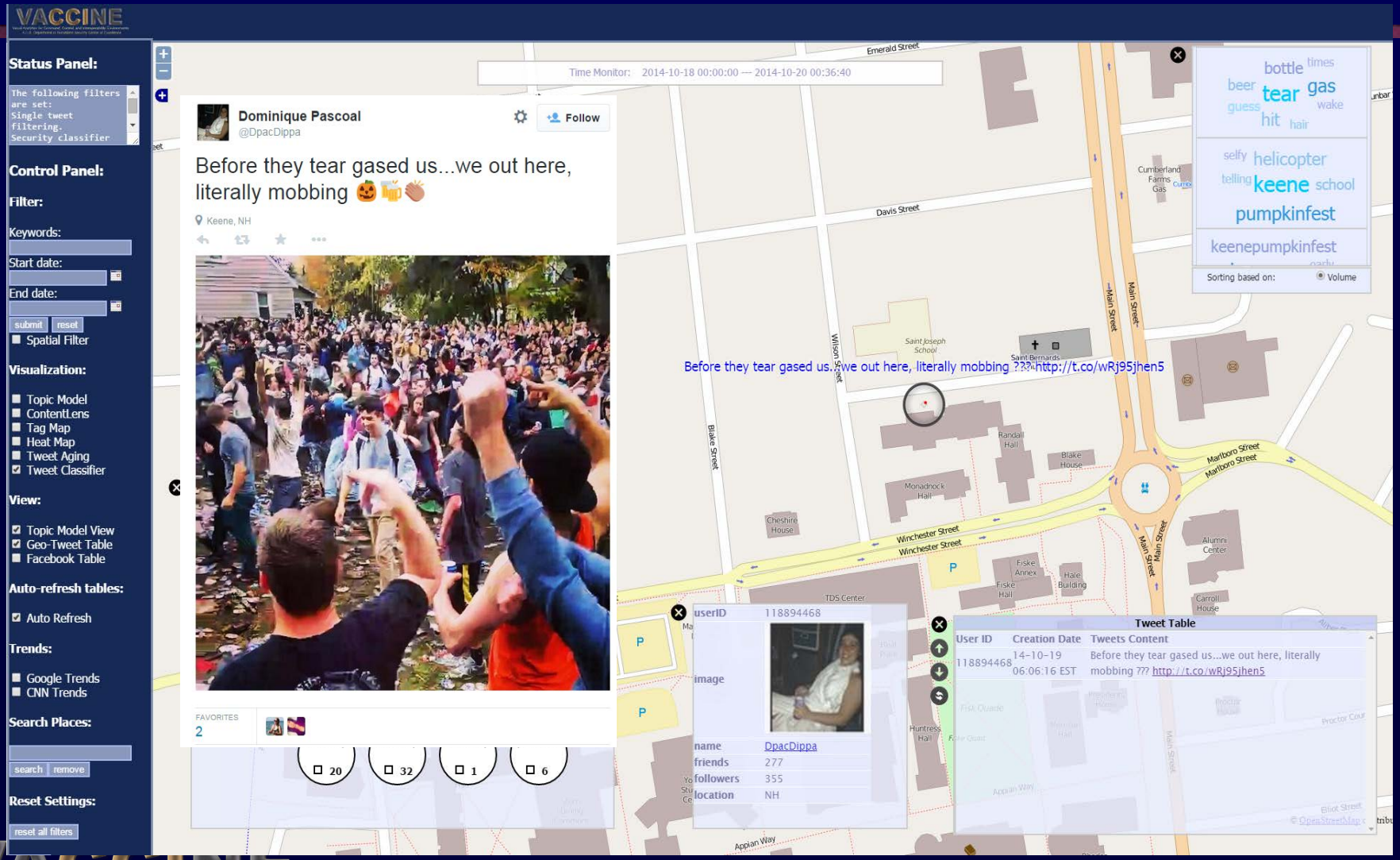
Major topics:



Security classifier became prominent.



Keene, NH

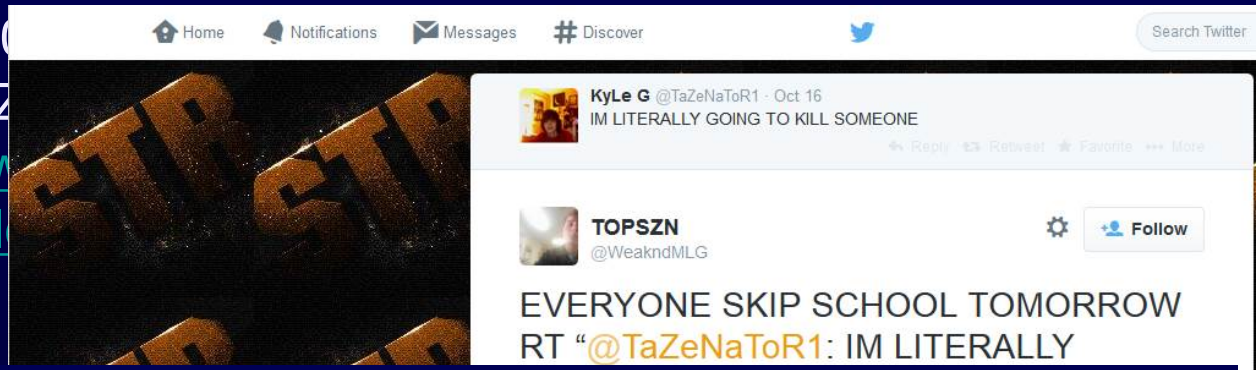

VACCINE
Status Panel:
The following filters are set:
Single tweet filtering.
Security classifier
Control Panel:
Filter:
Keywords:
Start date:
End date:
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
search
remove
Reset Settings:
reset all filters
Time Monitor: 2014-10-18 00:00:00 --- 2014-10-20 00:36:40
Dominique Pascoal
@DpacDippa
Follow
Before they tear gased us...we out here, literally mobbing
Keene, NH
FAVORITES
2
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
bottle
times
beer
tear
gas
guess
wake
hit
hair
selfy
helicopter
telling
keene
school
pumpkinfest
keenepumpkinfest
Sorting based on:
Volume
userID 118894468
image
name DpacDippa
friends 277
followers 355
location NH
Tweet Table
User ID
Creation Date
Tweets Content
118894468
14-10-19 06:06:16 EST
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
Emerald Street
Davis Street
Main Street
Wilson Street
Blake Street
Winchester Street
Marlboro Street
Saint Joseph School
Saint Bernards
Randall Hall
Blake House
Monadnock Hall
Cheshire House
Fiske Annex
Fiske Hall
Hale Building
TDS Center
Alumni Center
Carroll House
Huntress Hall
Appian Way
Cumberland Farms Gas

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



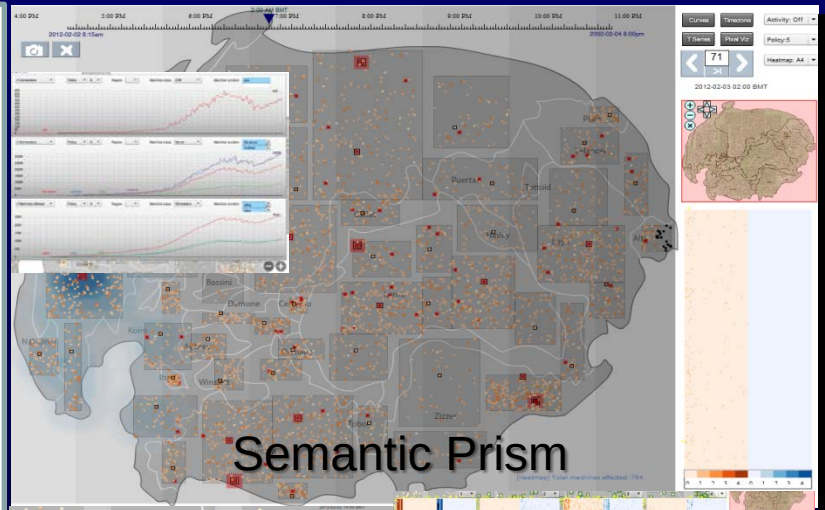
VACCINE

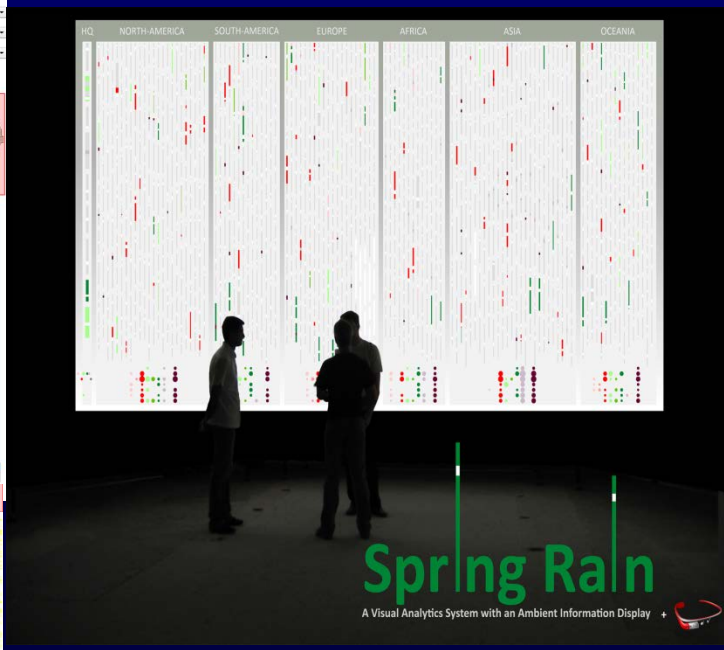
December 2014

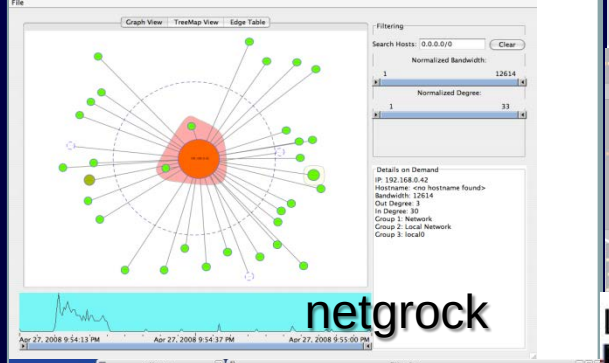
Intrusion &Packet/Traffic Vis

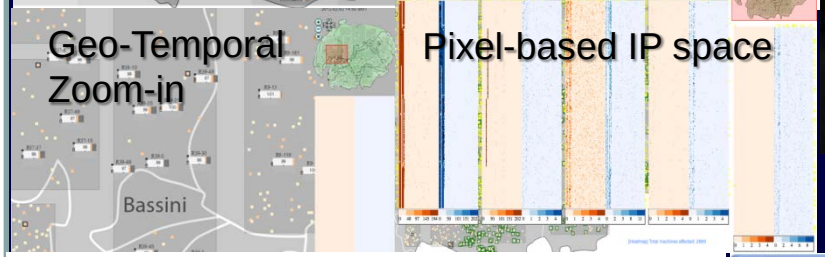
System-Wide Multi-Scale Vis



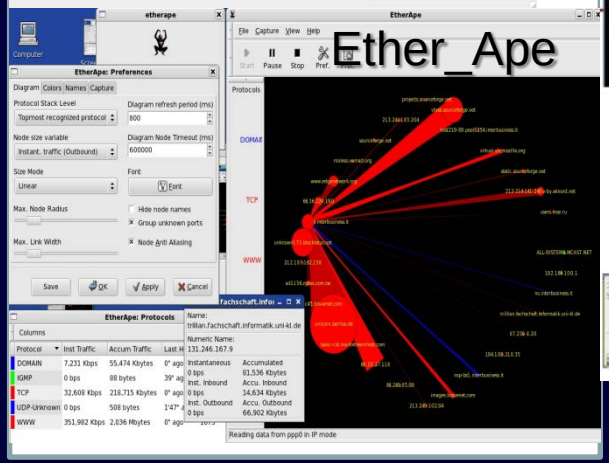


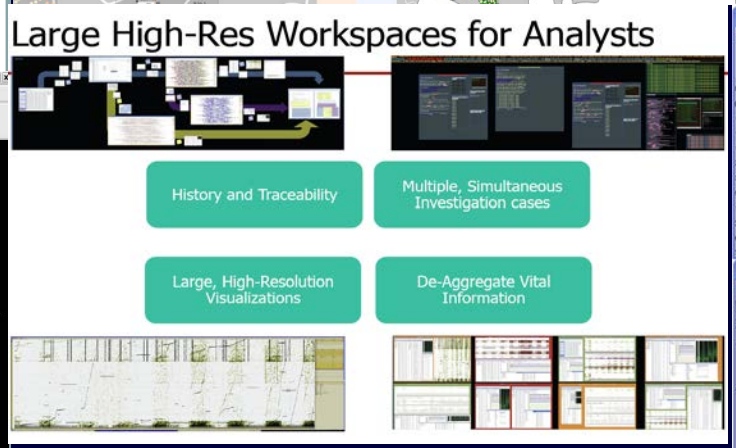


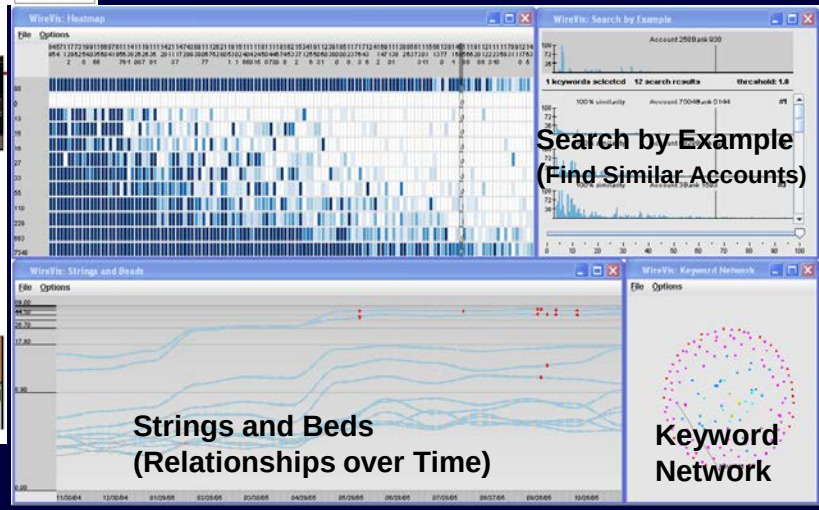




WireVis:
Wire Transaction Fraud Vis







VACCINE

CVADA

Center for Visual Analytics and Data Analytics
A U.S. Department of Homeland Security Center of Excellence

VACCINE: Advanced Interactive Visual Analytic Environments

David Ebert
Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments
A U.S. Department of Homeland Security Center of Excellence

January 2015

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of California, San Diego**
- **University Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **Oak Ridge National Laboratories**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 50 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America





Big Data: Solution to Global Challenges?



- Big Data is often defined along three dimensions:
  - **Volume** – size
  - **Velocity** – rate of input, update, change
  - **Variety** – different types, sources, variables
- Need:
  - Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information (TechAmerica Foundation)
  - **Enable effective, efficient analysis, decision making, planning, and action**

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

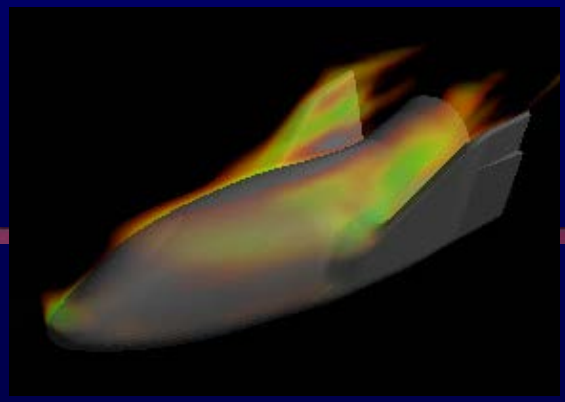
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

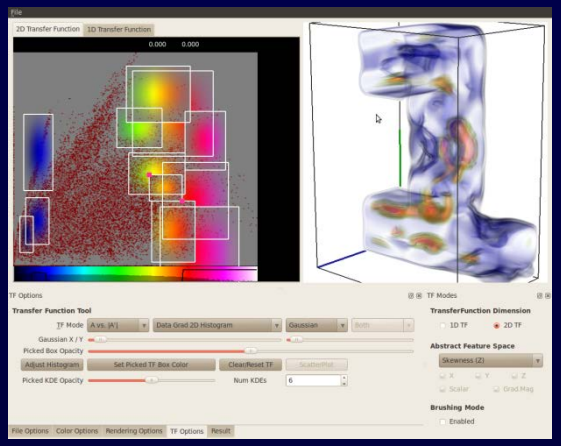
VADM Robert Parker, 2012 MRS Keynote Address

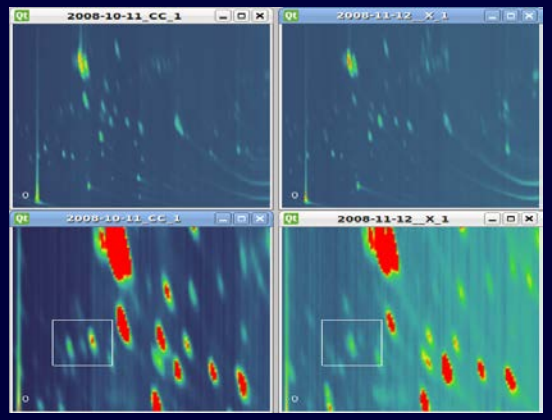


Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Cancer biomarker
  - Visual analysis suite for exploring multiple samples of data from cancer care engineering



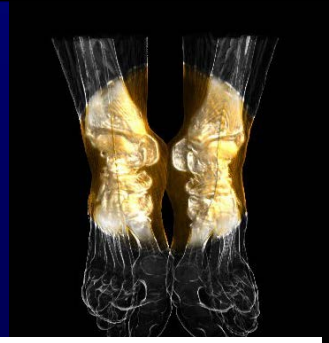


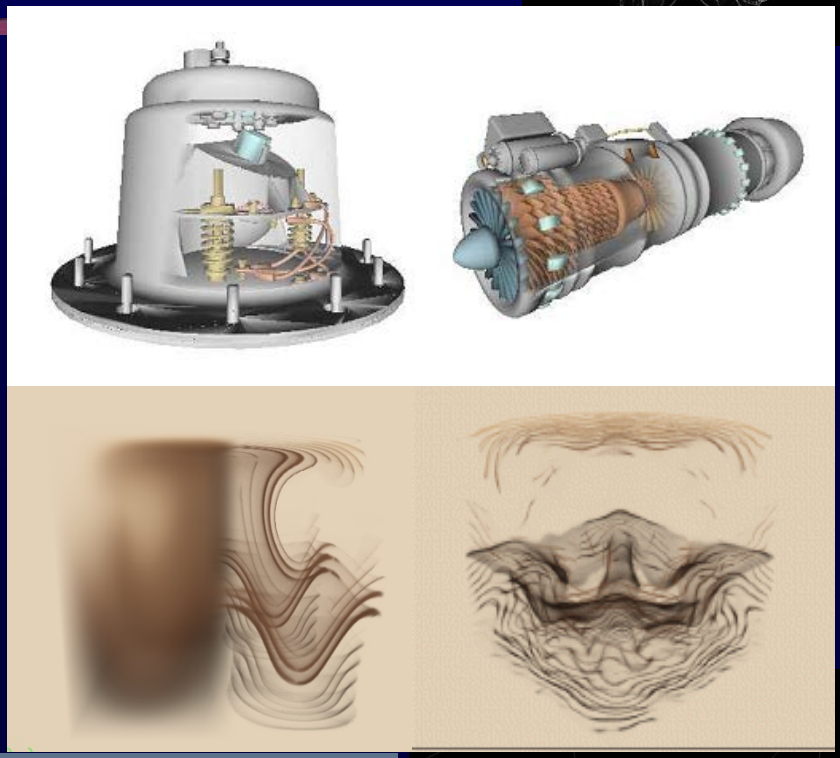


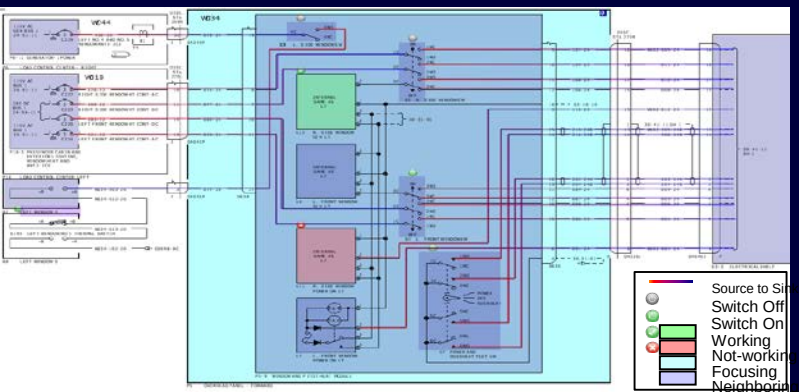
VACCINE

Scientific Discovery

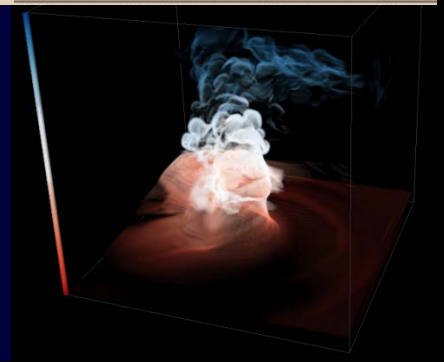
- Illustrative Visualization
  - Mechanical assemblies
  - Schematics and wiring diagrams
  - Illustrative flow visualization
  - Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics







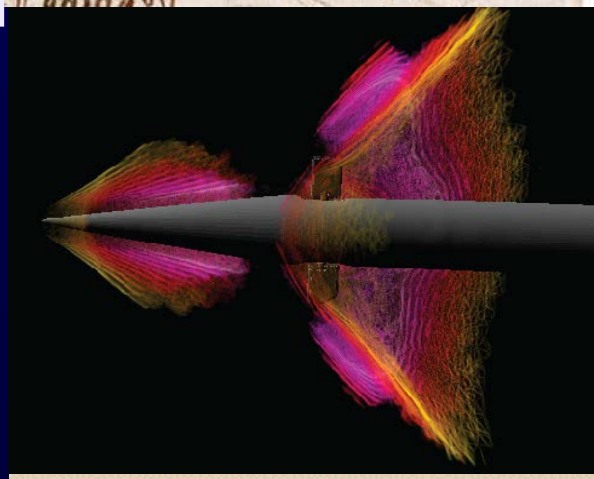


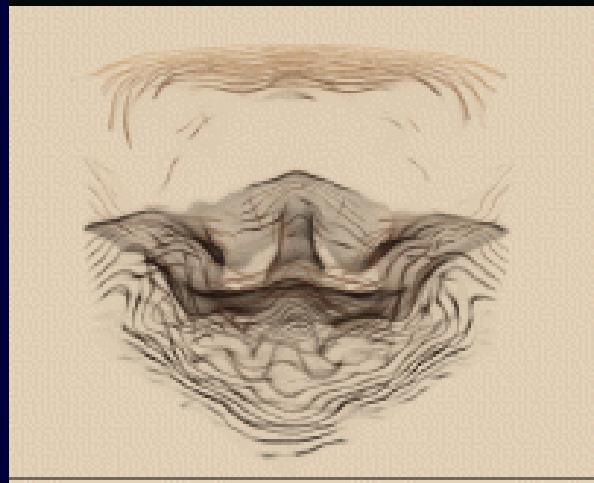


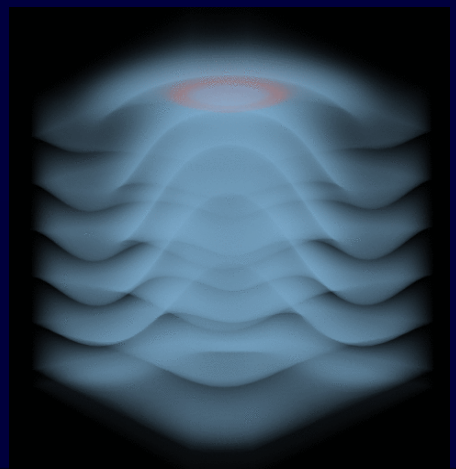
Flow Visual Analytics

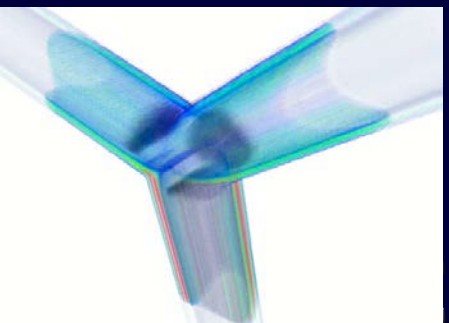
- Inspired by work of Leonardo da Vinci
- Allow understanding of flow processes an effects
- Enable interactive analysis of massive unstructured data through functional encoding, analysis and rendering
- Enable Interactive correlative feature analysis







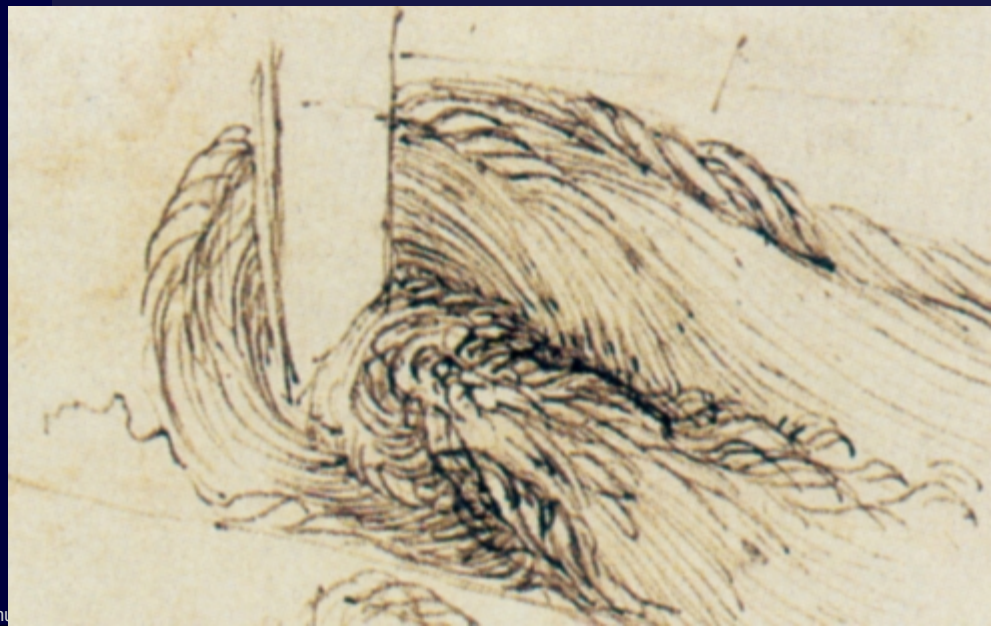




Flow Analytics

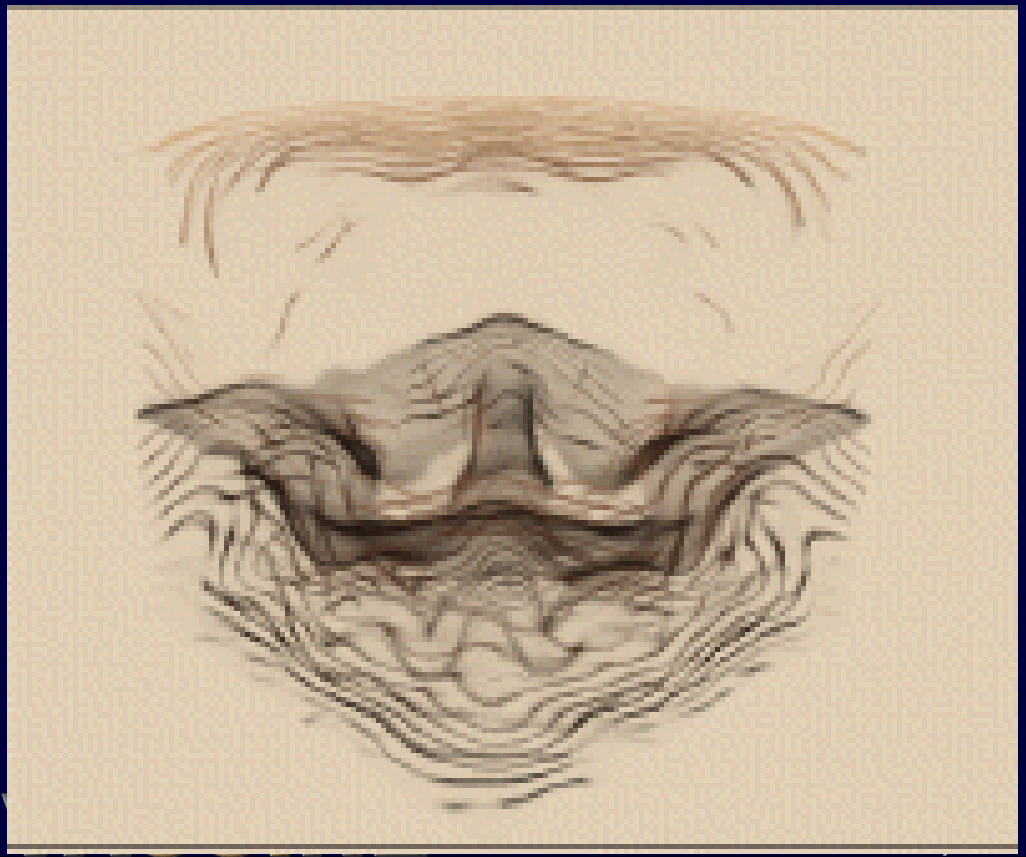
- Inspired by work of Leonardo da Vinci
- Main feature conveyed with a few simple strokes
- Easy to understand 3D turbulent flow

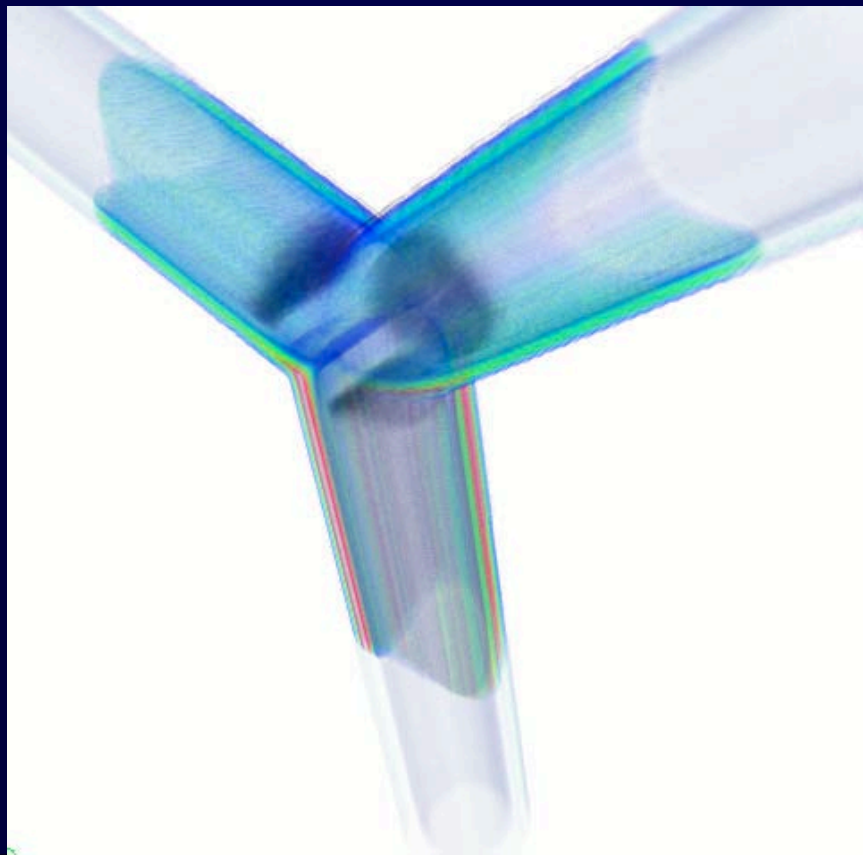




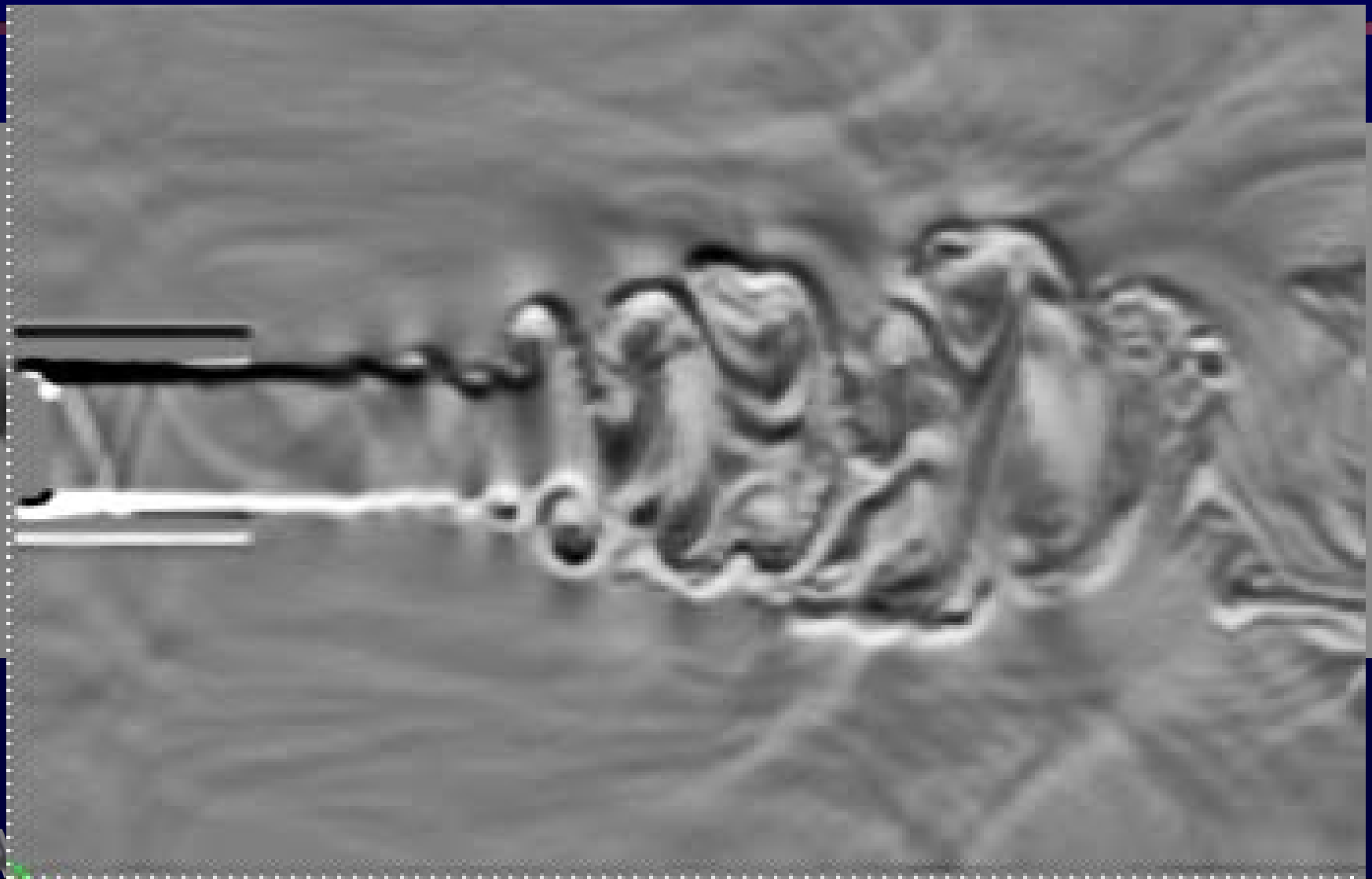
Illustrative Flow Analytics

- Two examples

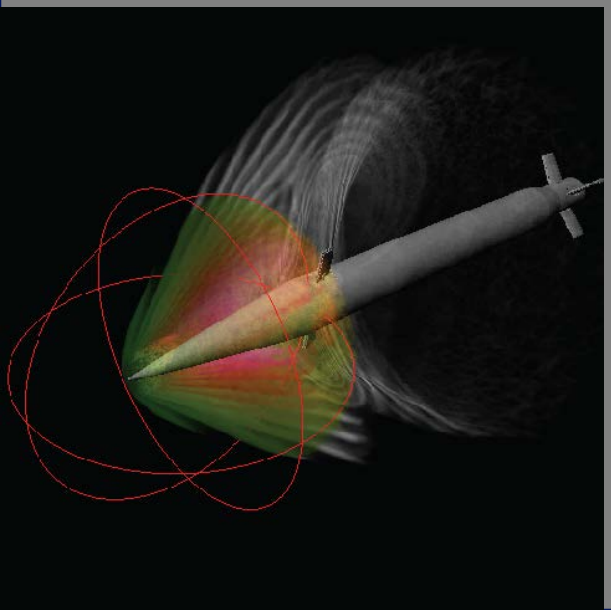


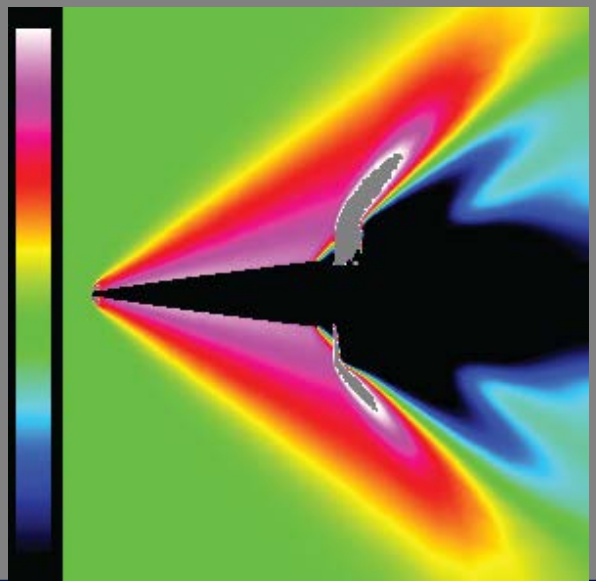


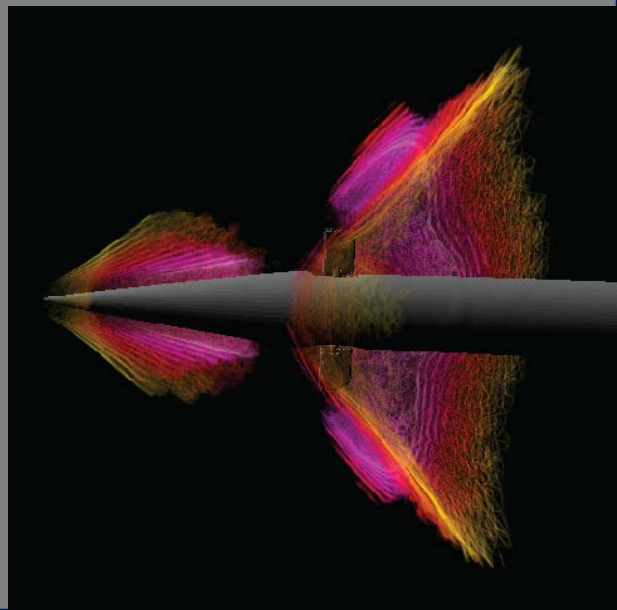
Flow - Photographic Inspired Techniques



Flow Analytic Environment

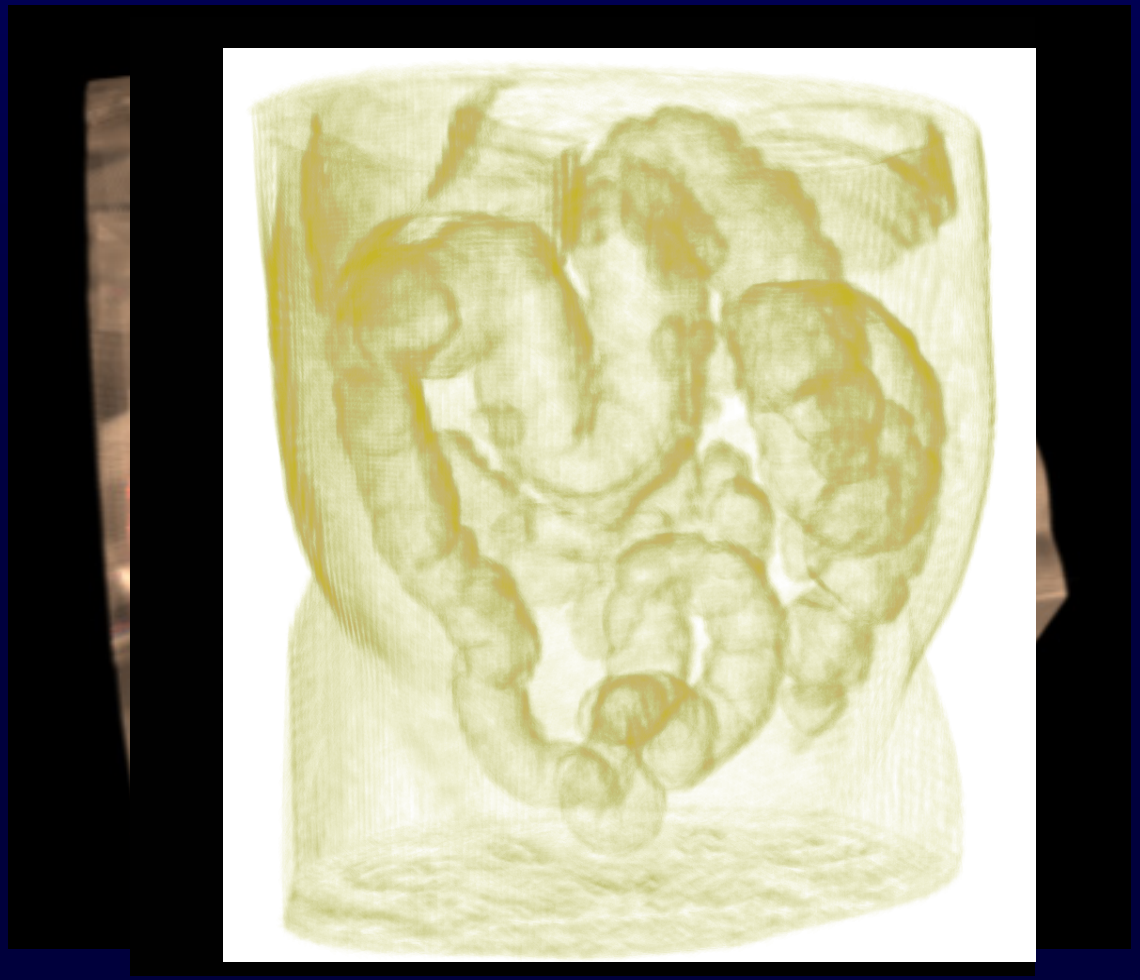






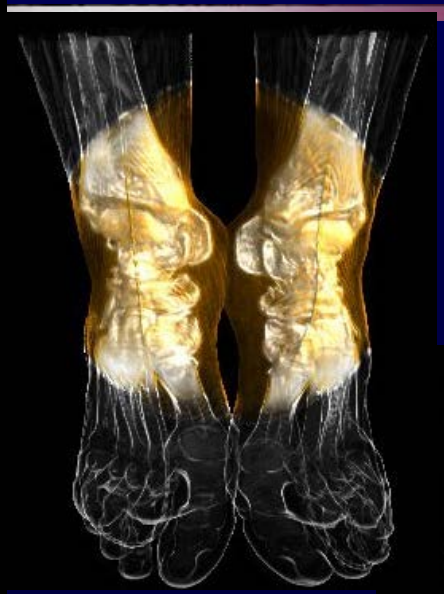
Volume Illustration: Overview

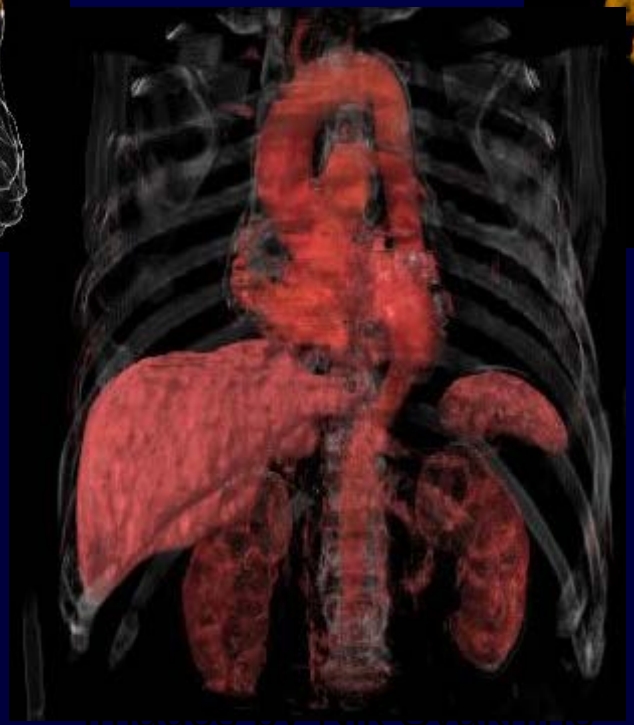
- Utilize illustration principles and techn

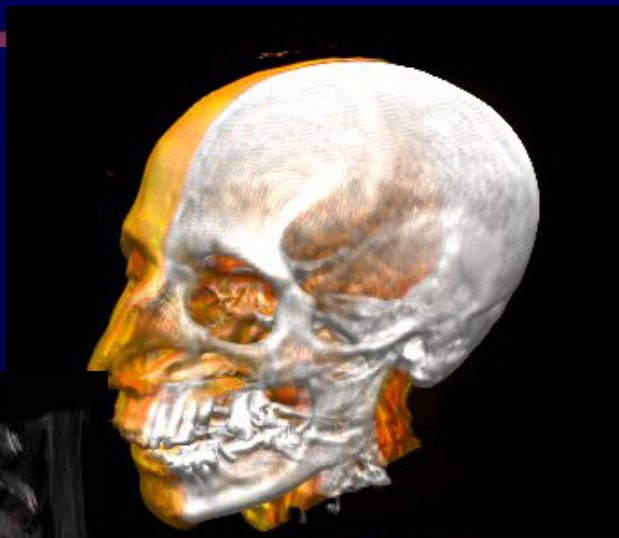


Interactive Volume Illustration Motifs: Anatomical Illustration





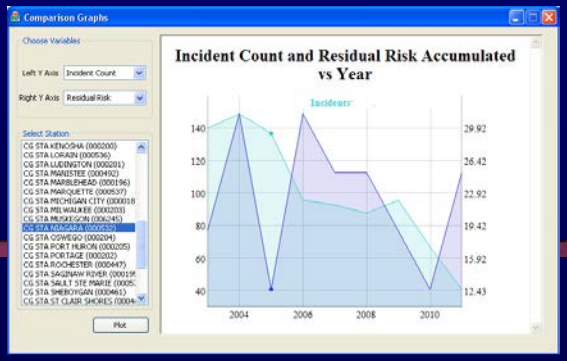


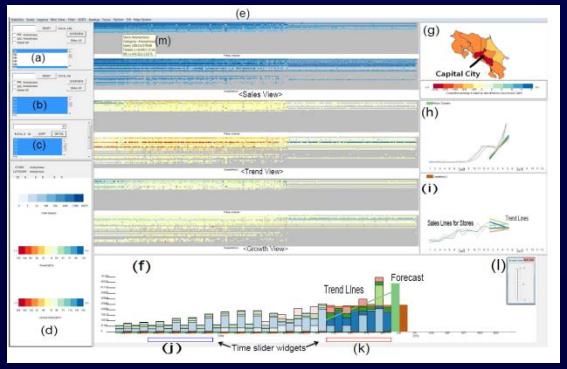




Business Visual Analytics

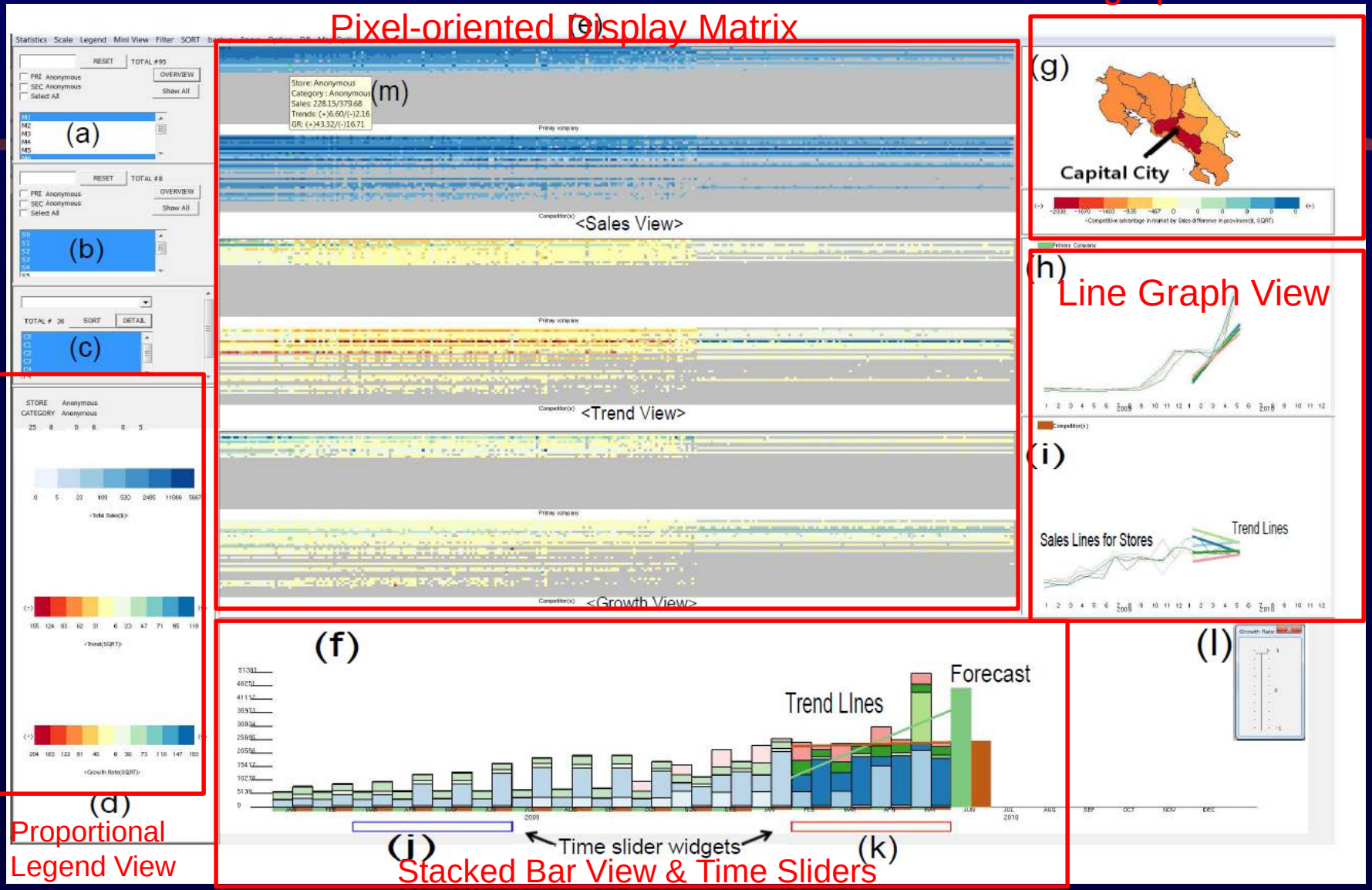
- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation





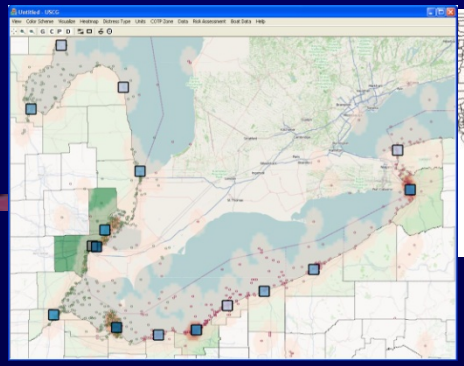


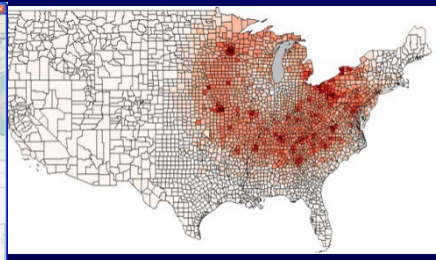
Market Analyzer

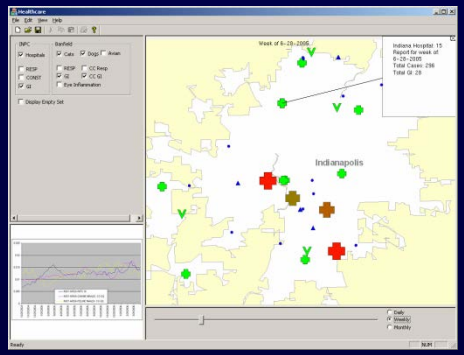


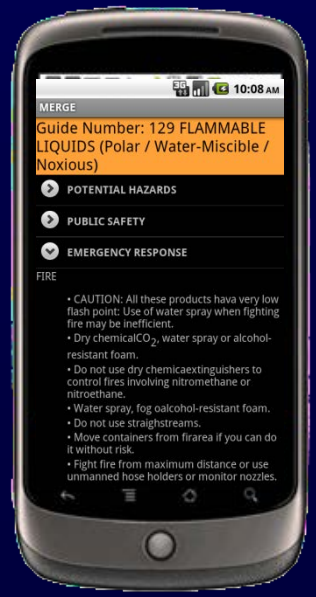
Public Safety & Health

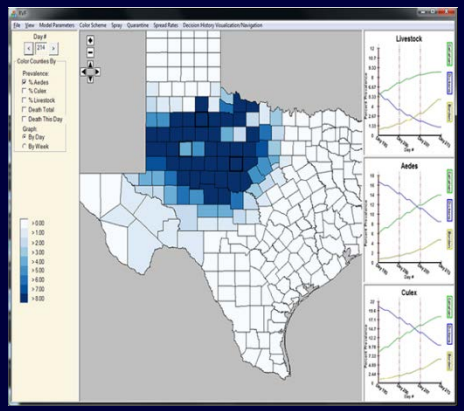
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering













VACCINE

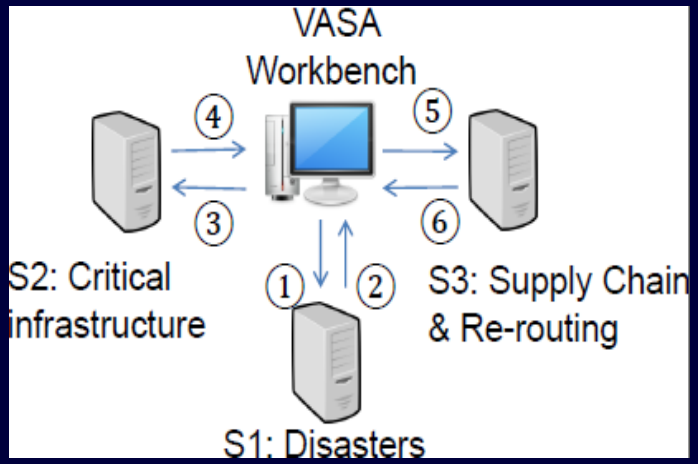
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC

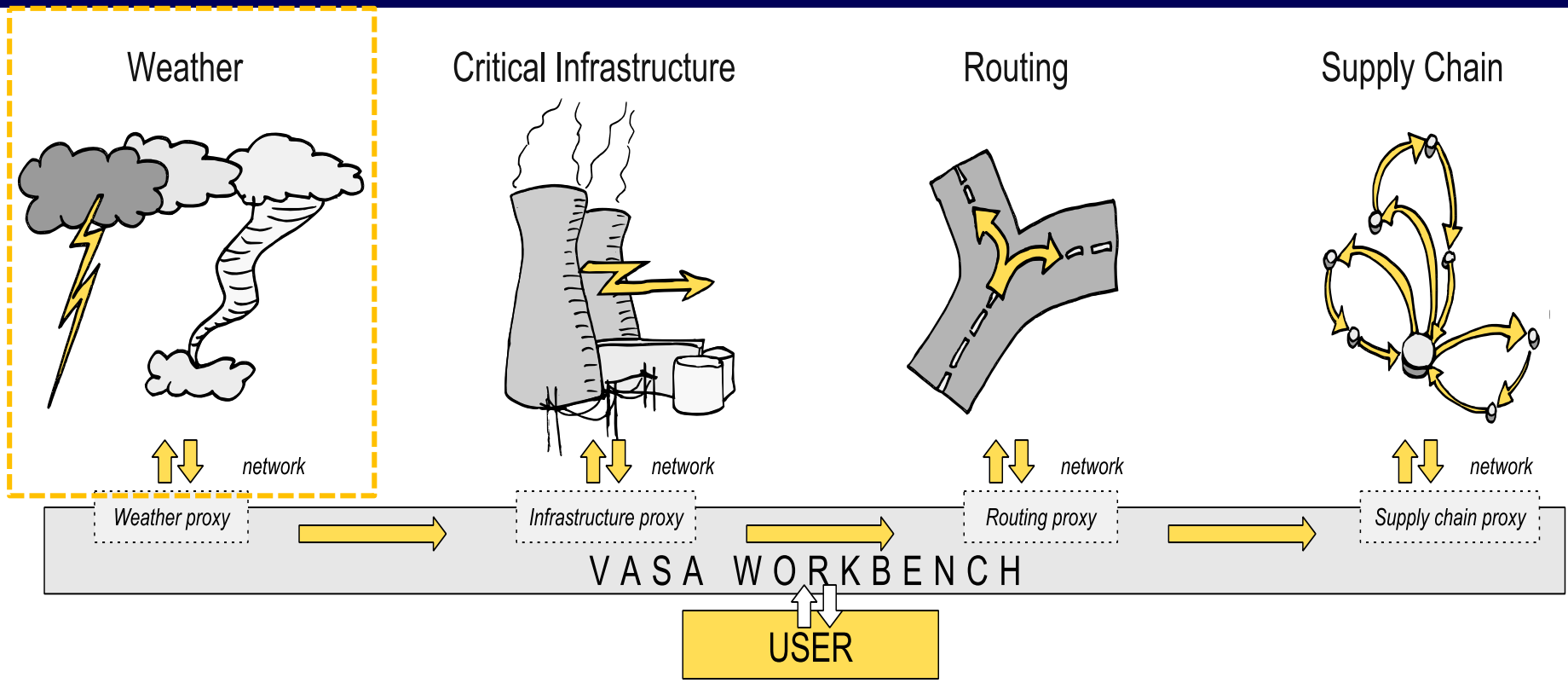


U Minn

TACC

VACCINE

VASA COMPONENT:
CRITICAL INFRASTRUCTURE

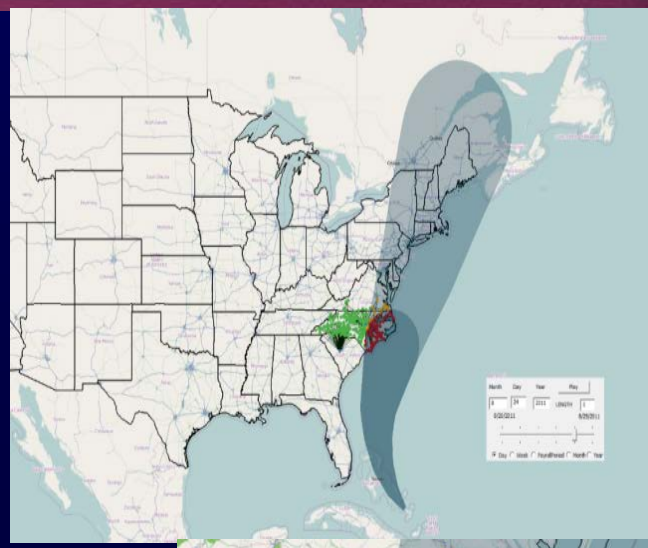


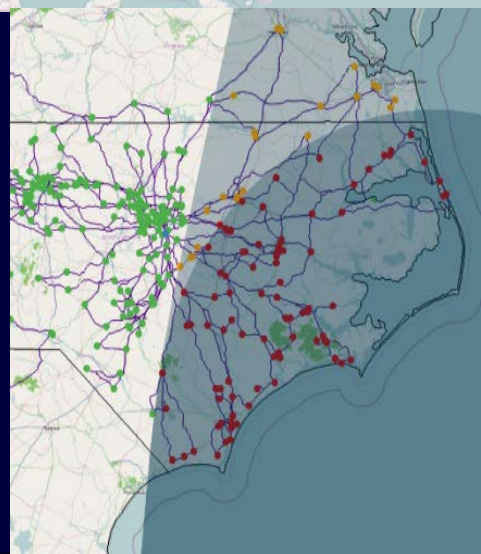
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models



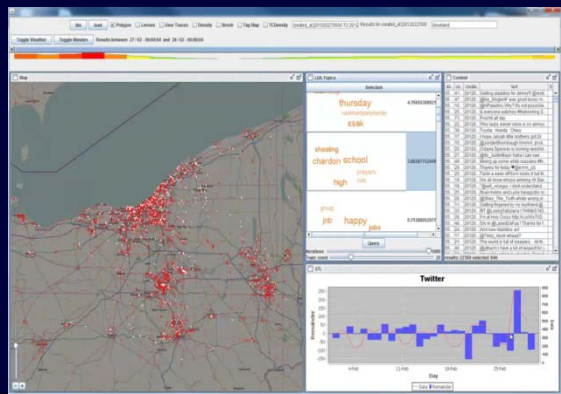


VACCINE

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Our Spatiotemporal Visual Analytics Work Contributions

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

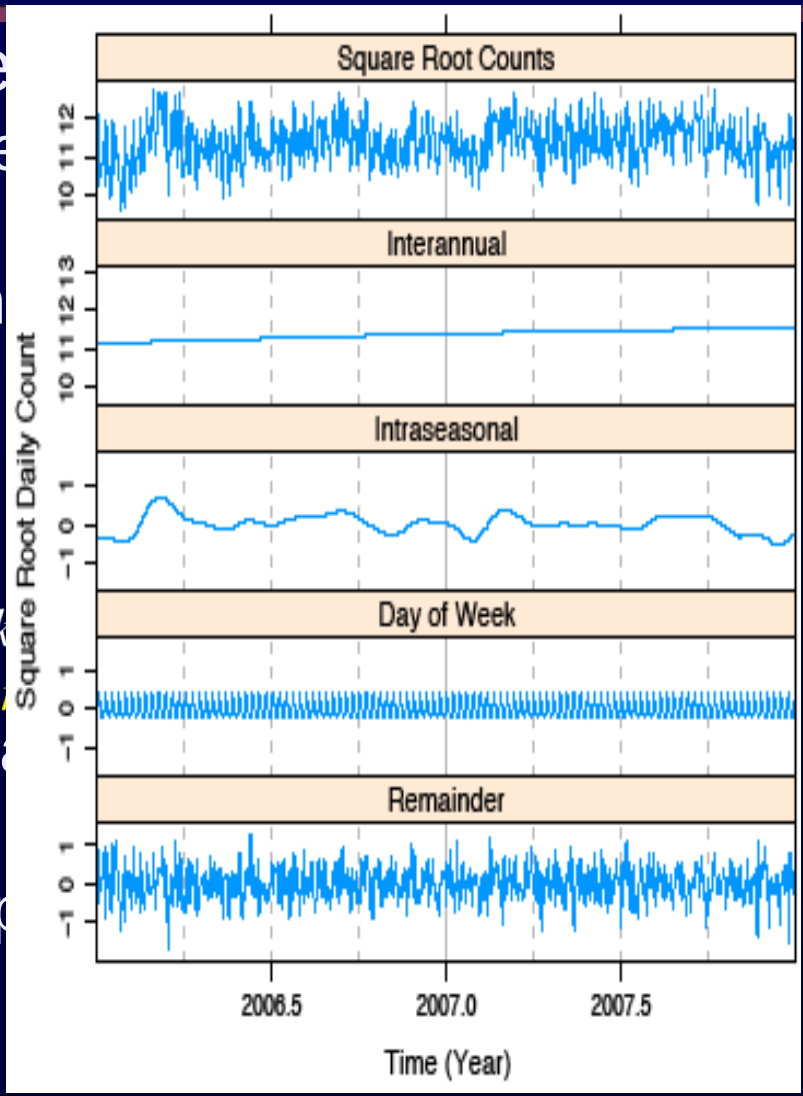
Event/Anomaly Detection - STL

- Temporal trend analysis and pre
  - Seasonal Trend decomposition base
    weighted regression) (STL)
  - Time series data viewed as the sum

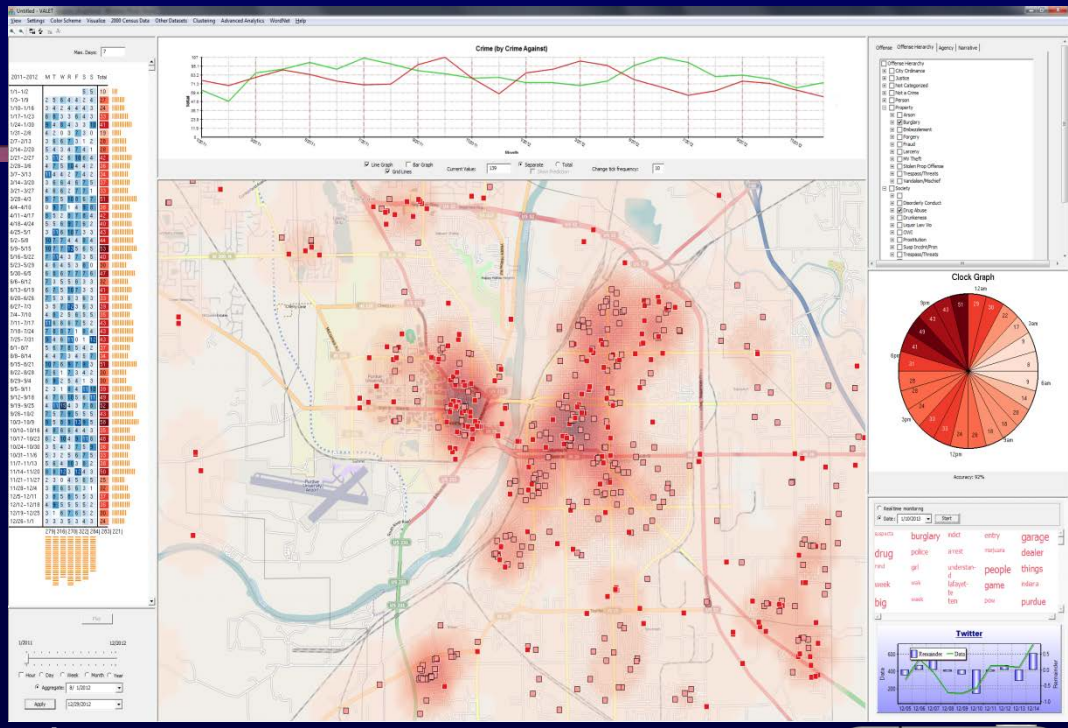
$$Y = T$$

Y : *original time series*
T : *trend component*
S : *seasonal (daily/week*
R : *remainder compone*

  - Large value of *R* indicates substanti
  - Filter out the noise component
  - If value varies by 2 SD from STL exp



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

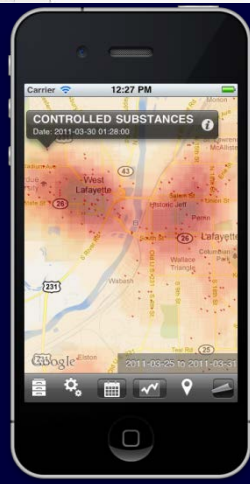
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

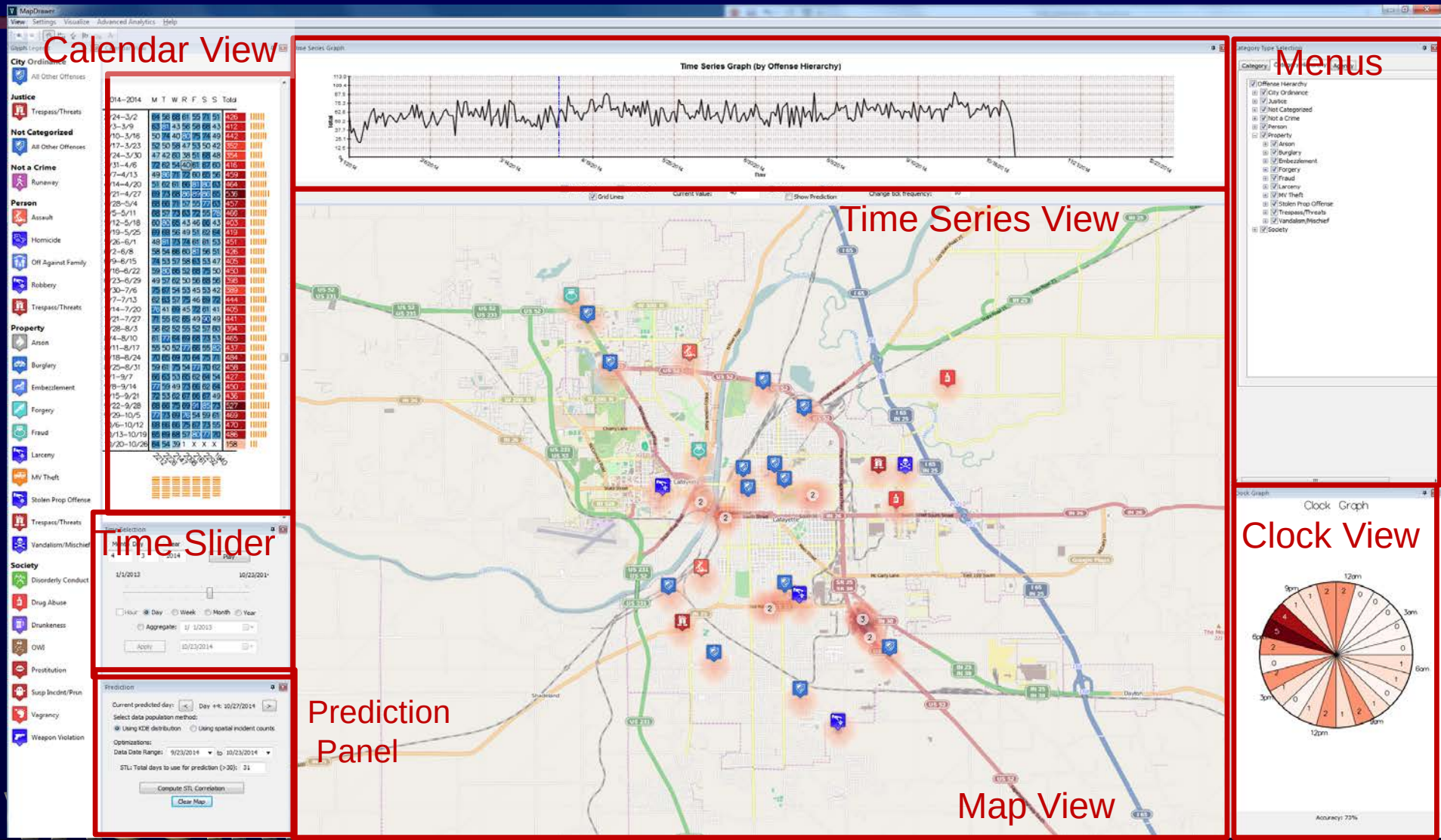
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

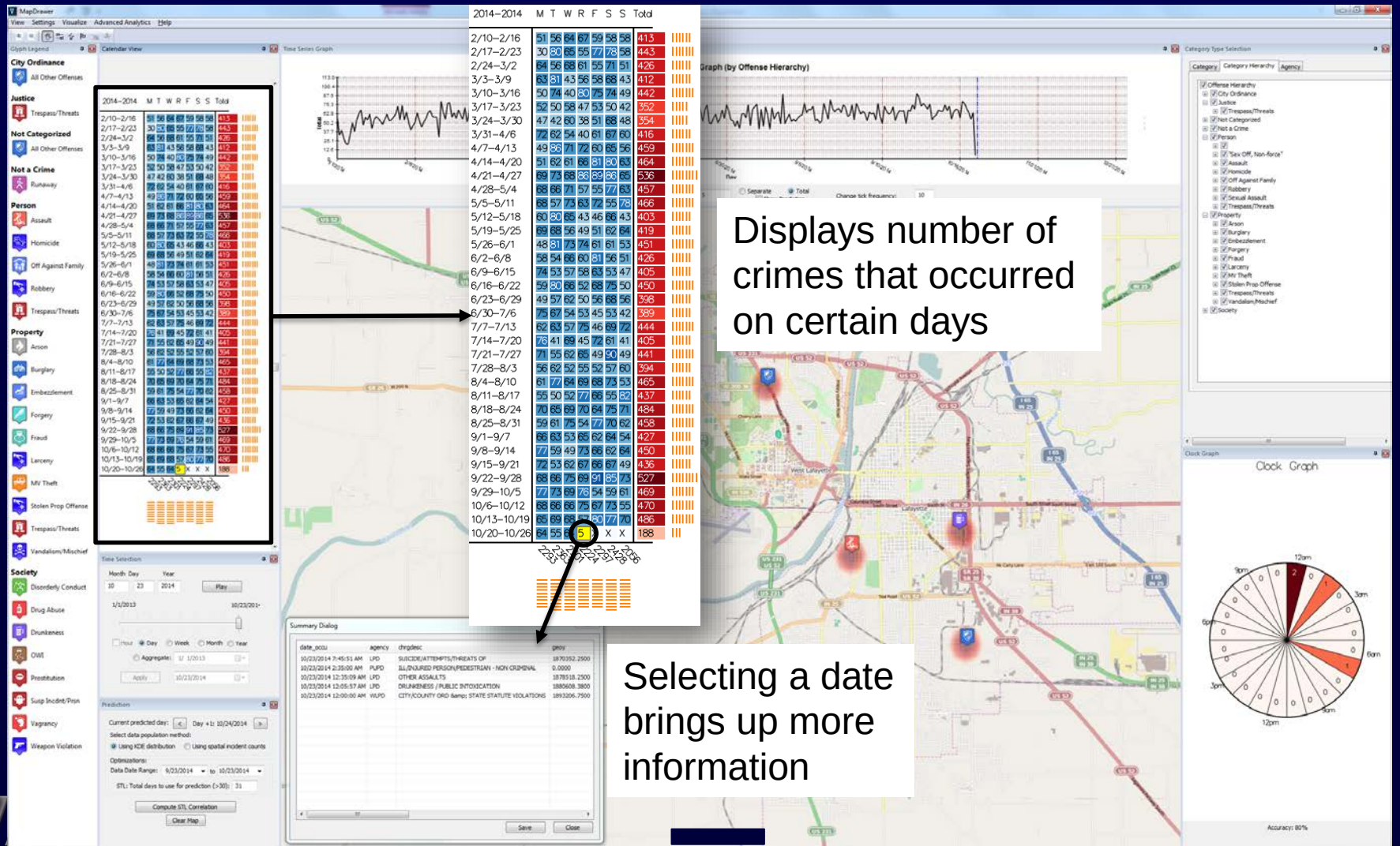


VACCINE

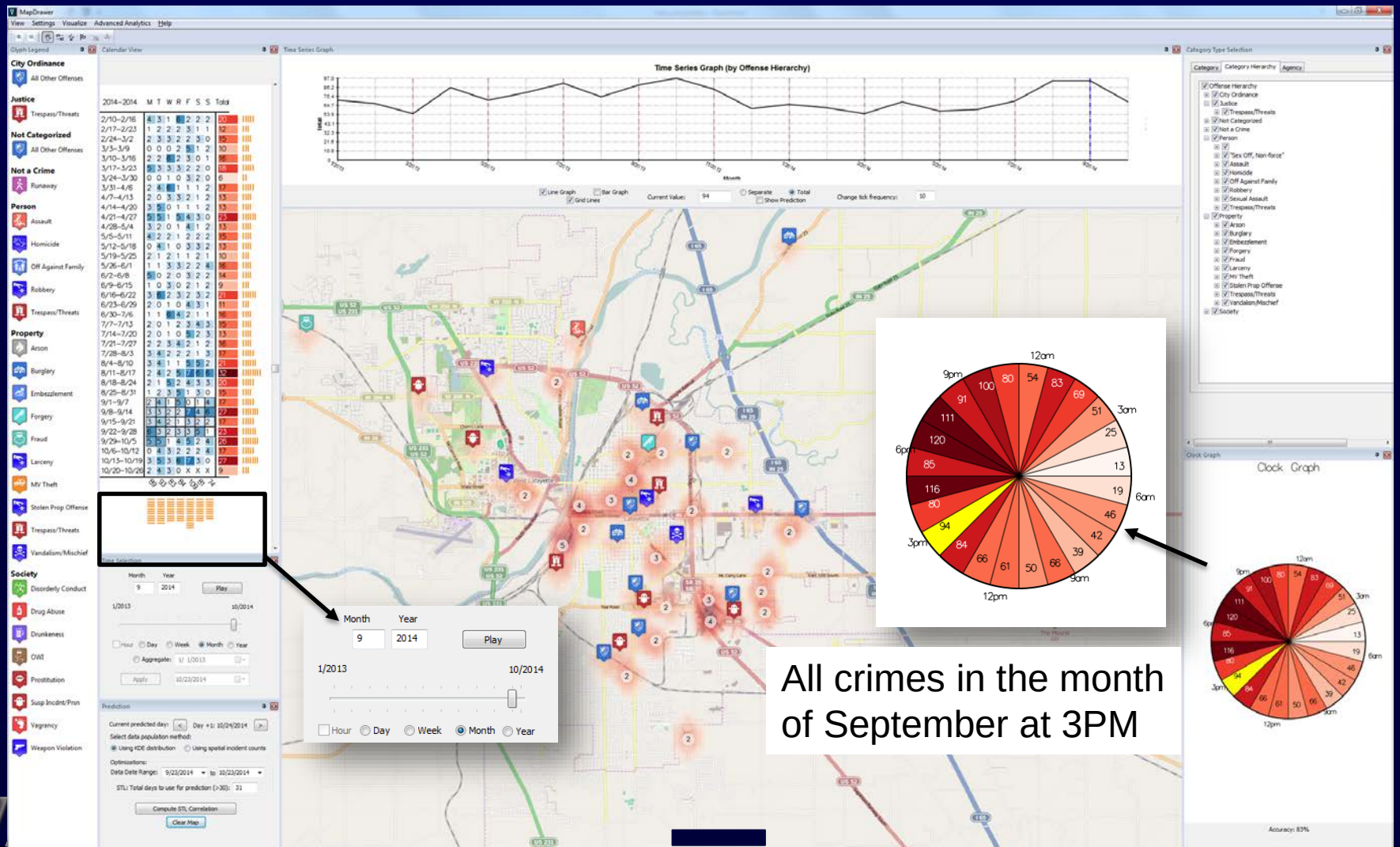
VALET Layout



Calendar Display



Crimes During Certain Hours In A Month

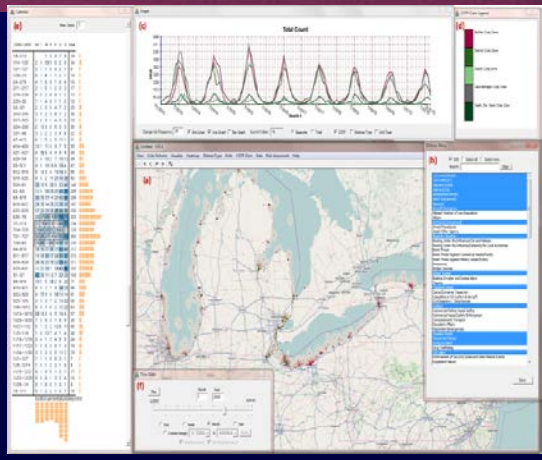


Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

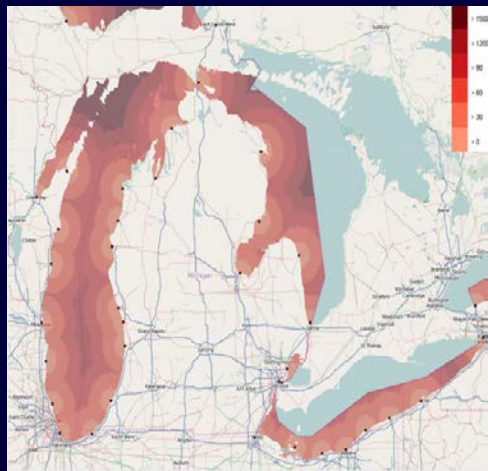
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771



IMPACTS:

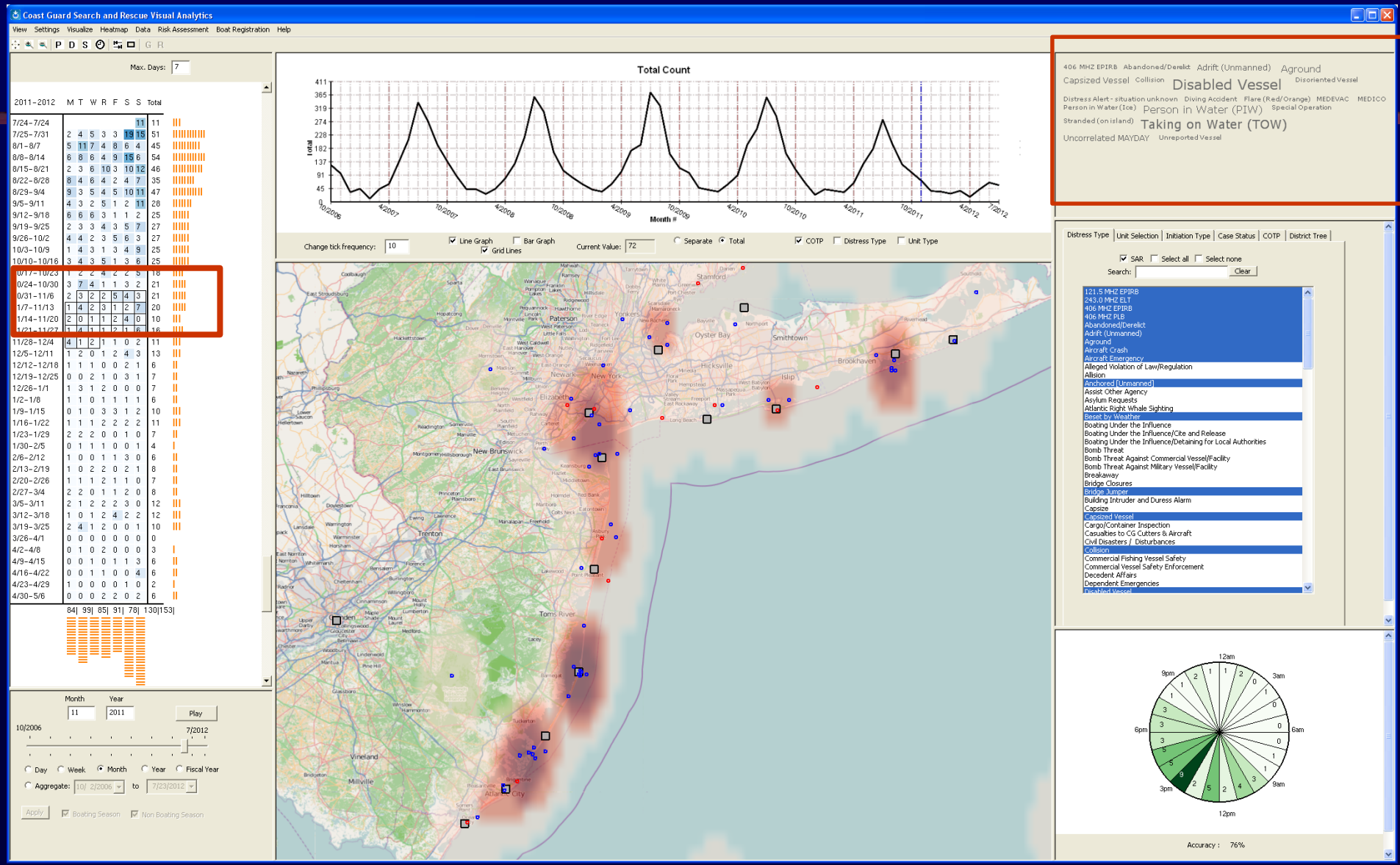
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

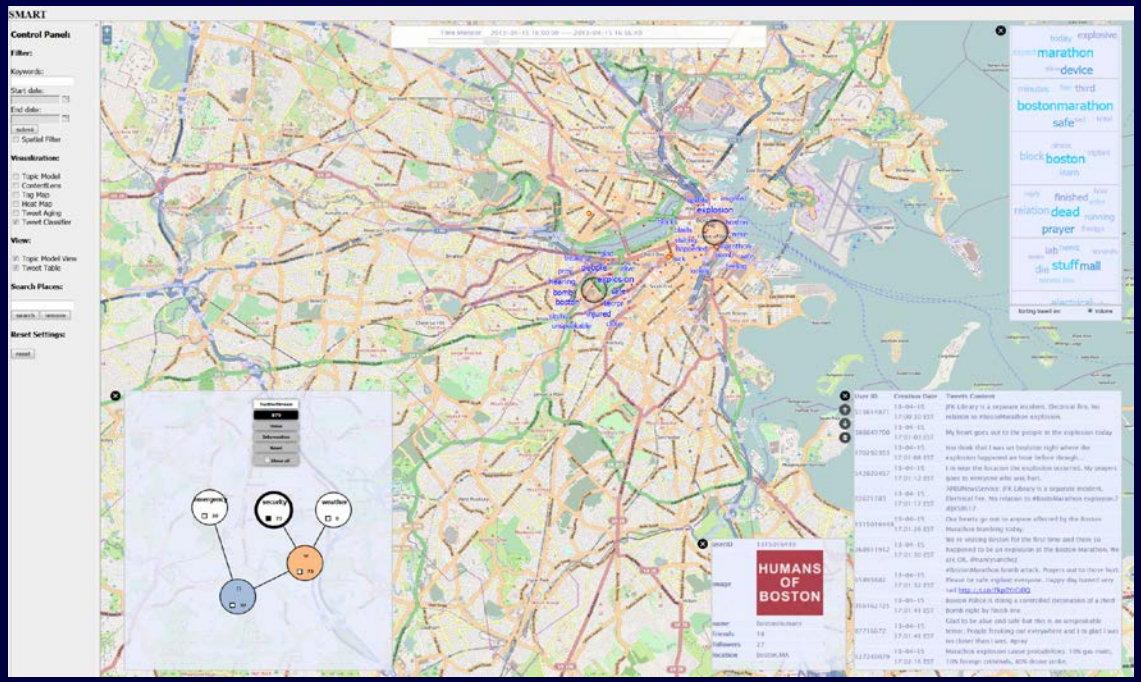


VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

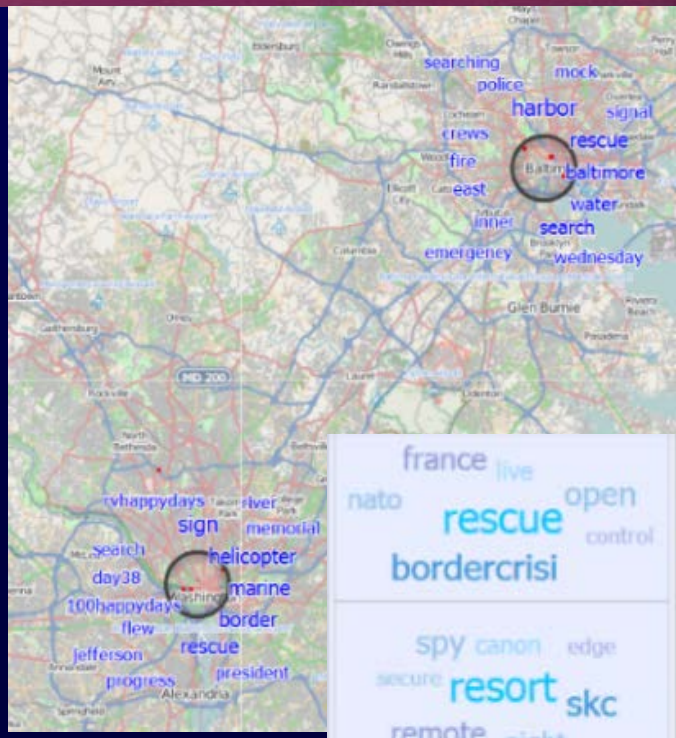
January 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

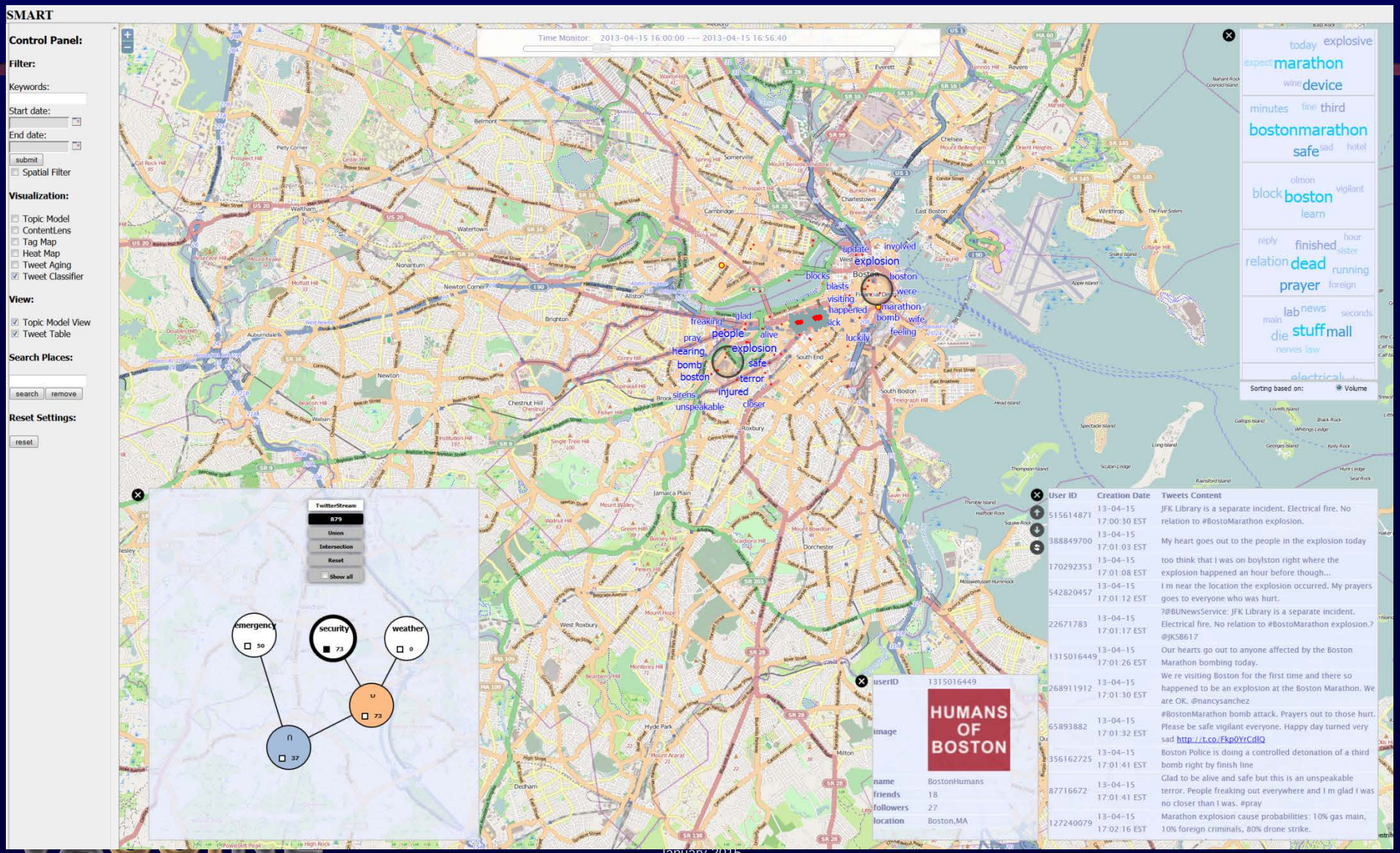
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools



VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



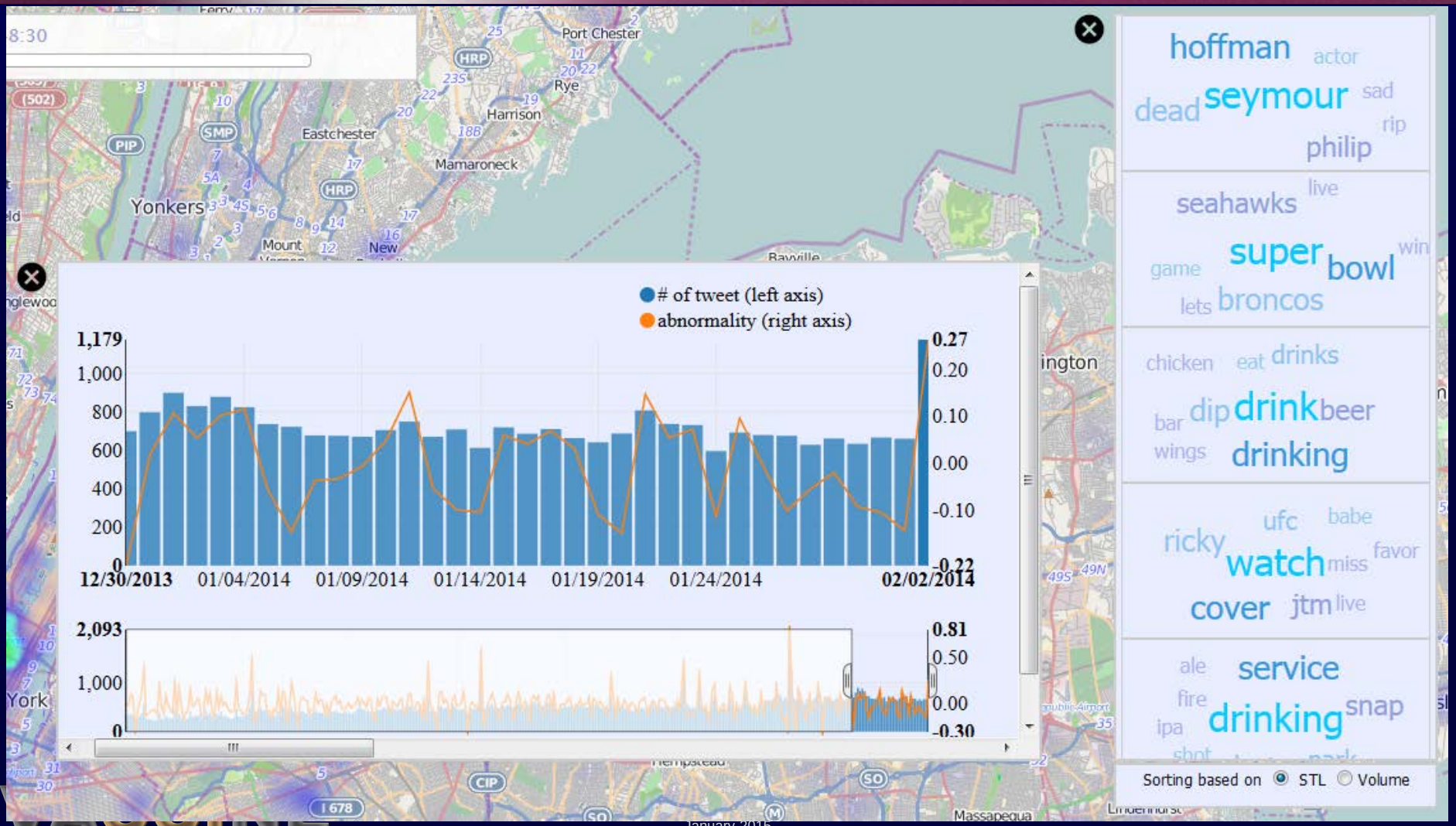
Where Can It Be Helpful?

- Understanding Customers
- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Fran
- During recurring event
  - Purdue football games – the poo
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Loui



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More



RETWEETS 6

FAVORITES 4

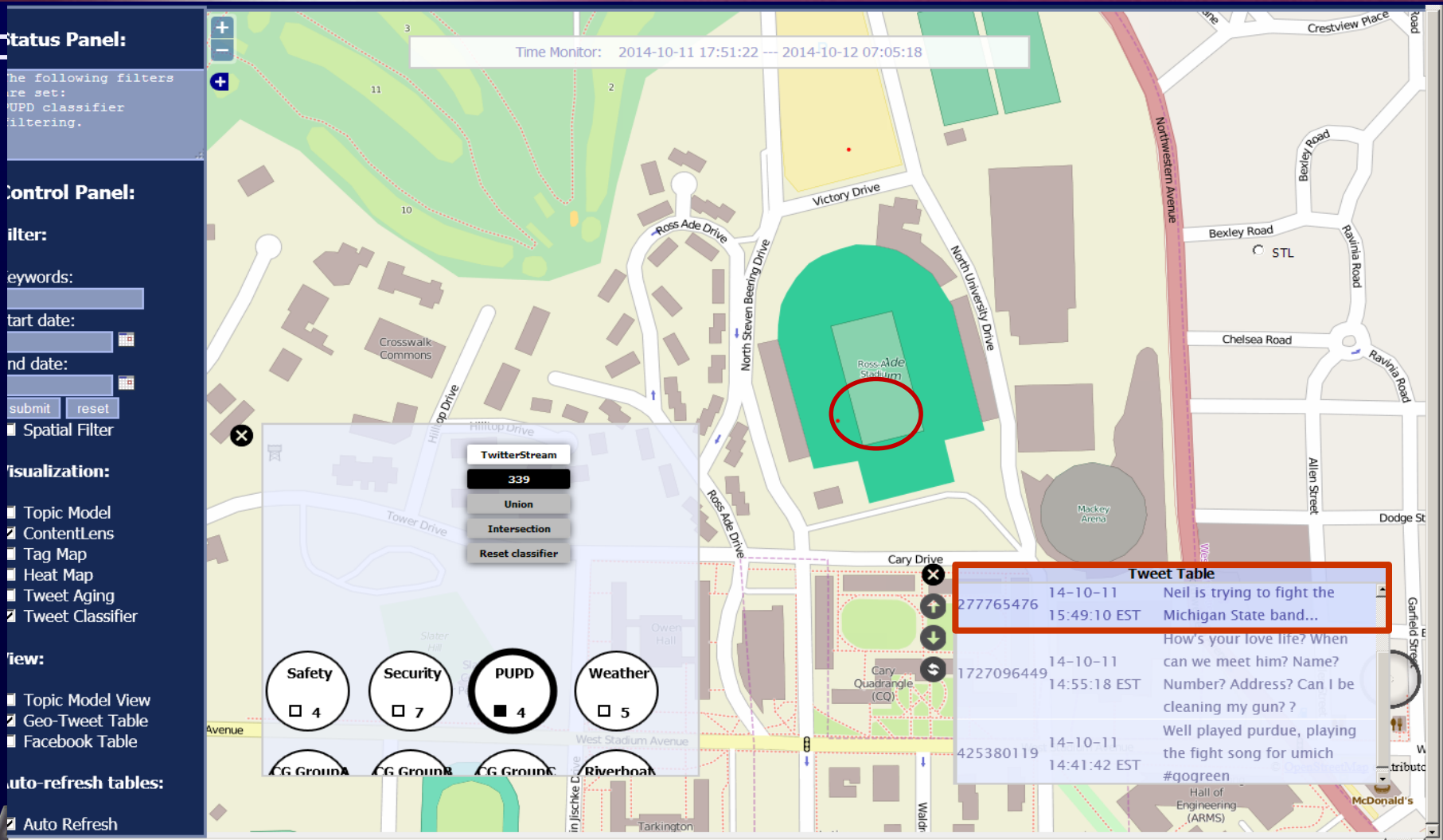
12:56 PM - 12 Oct 2014

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F

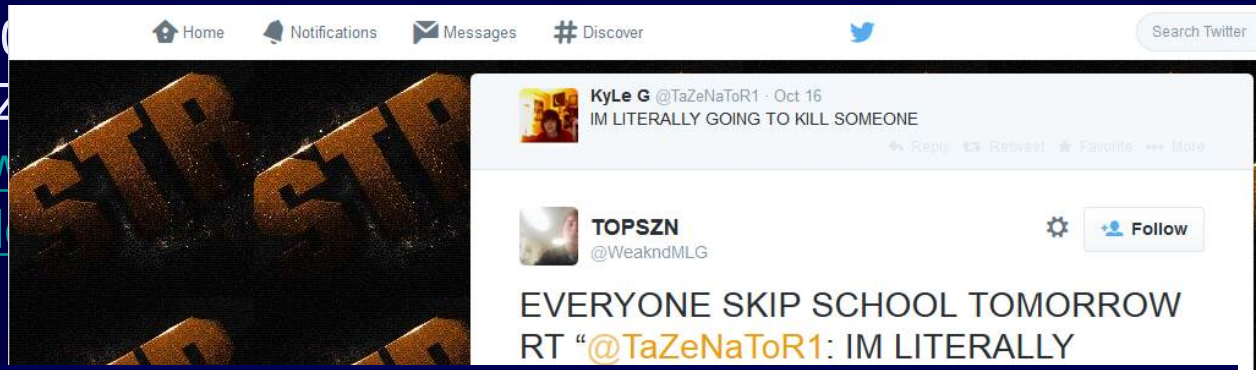


Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

VACCINE: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

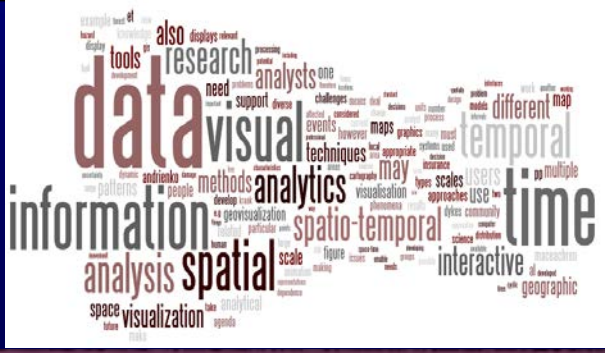
May 2015

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 29 Institutions



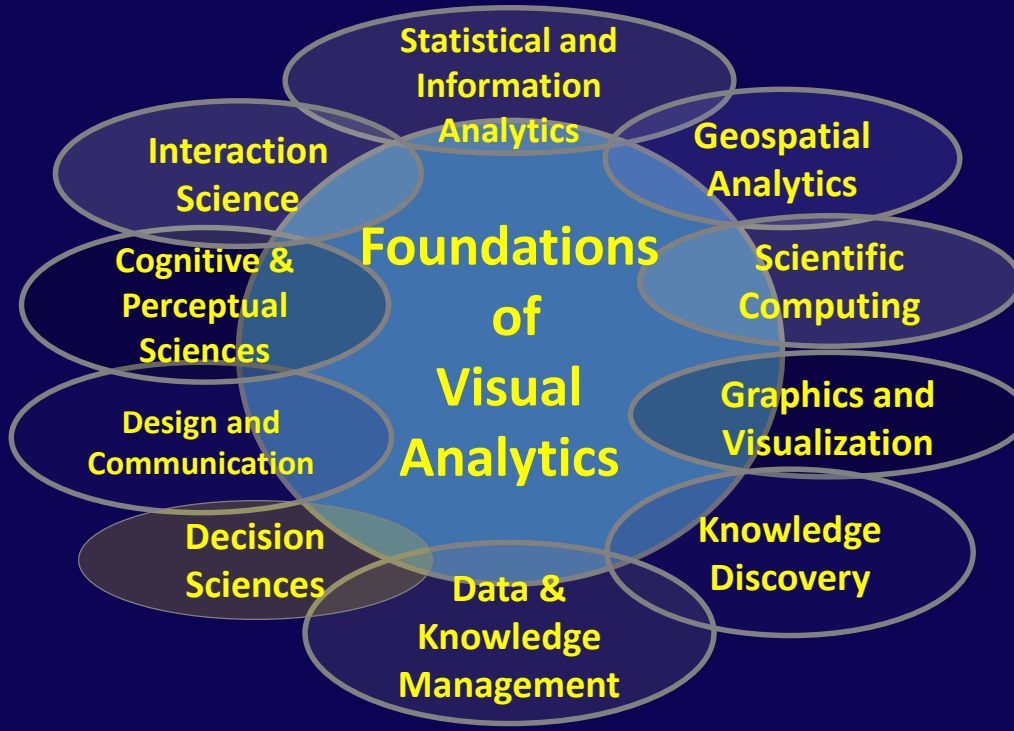
- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California San Diego
- University of Maryland
- University Washington
- Arizona State University
- Bethune-Cookman U.
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

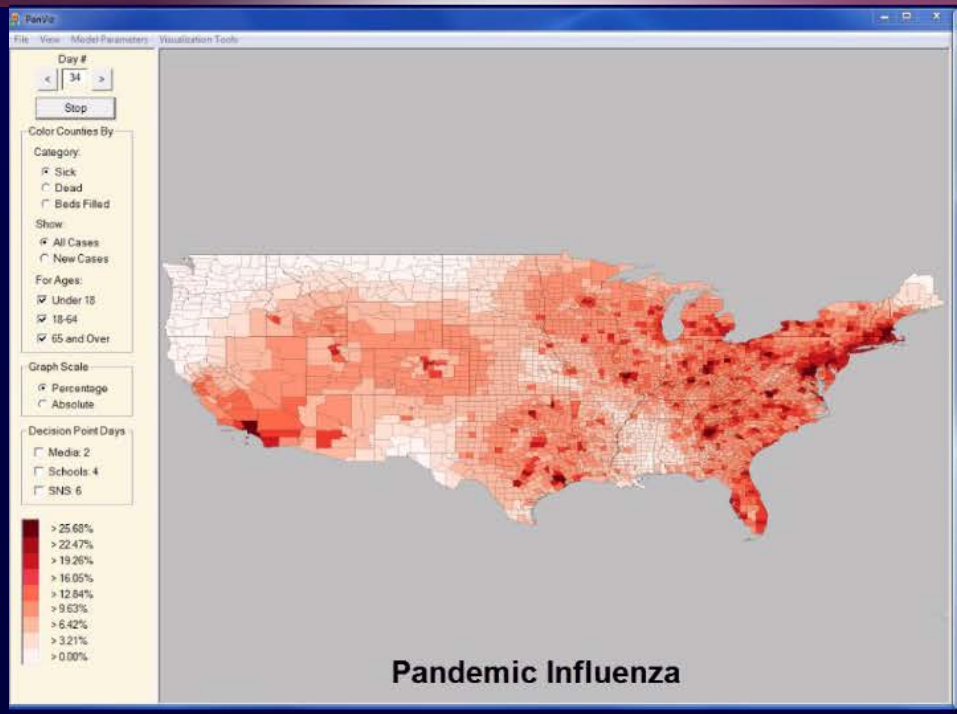
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

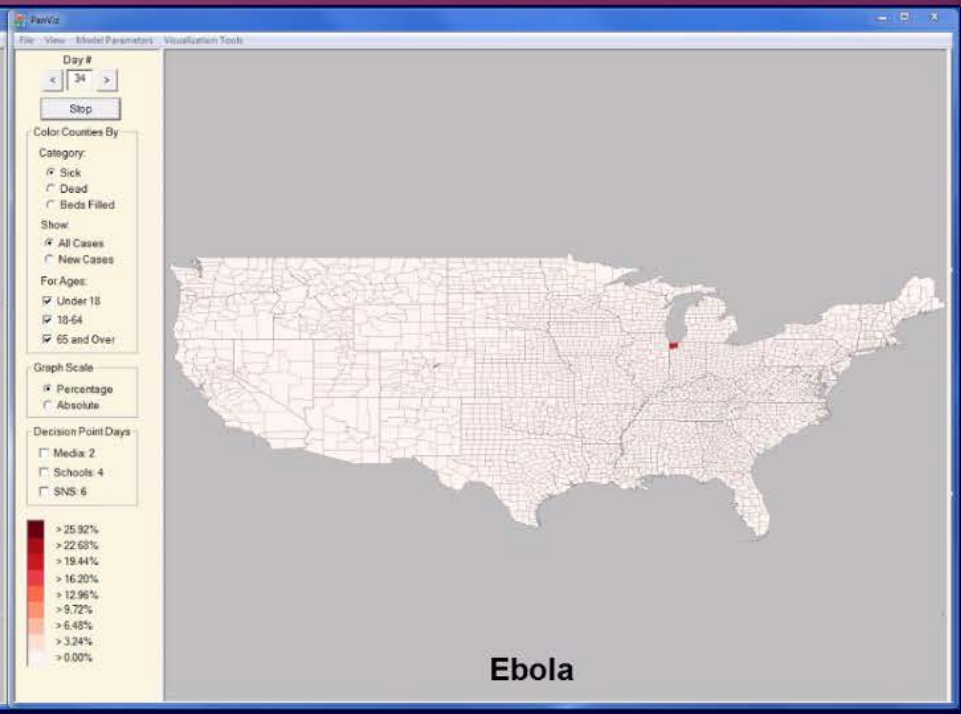
VADM Robert Parker, 2012 MRS Keynote Address



Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL





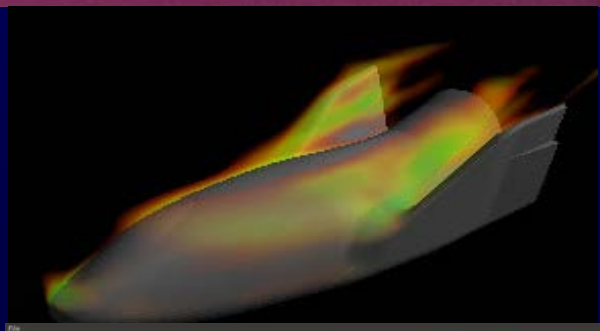
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

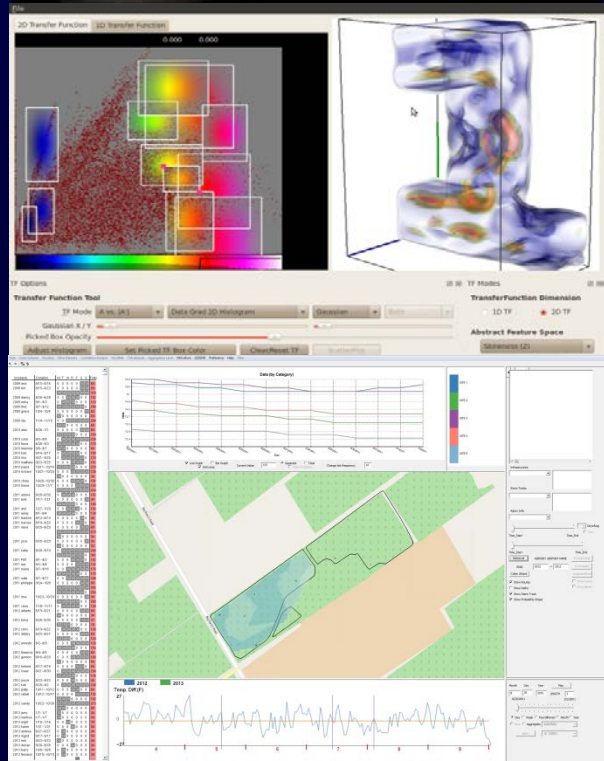
VACCINE

Scientific Discovery



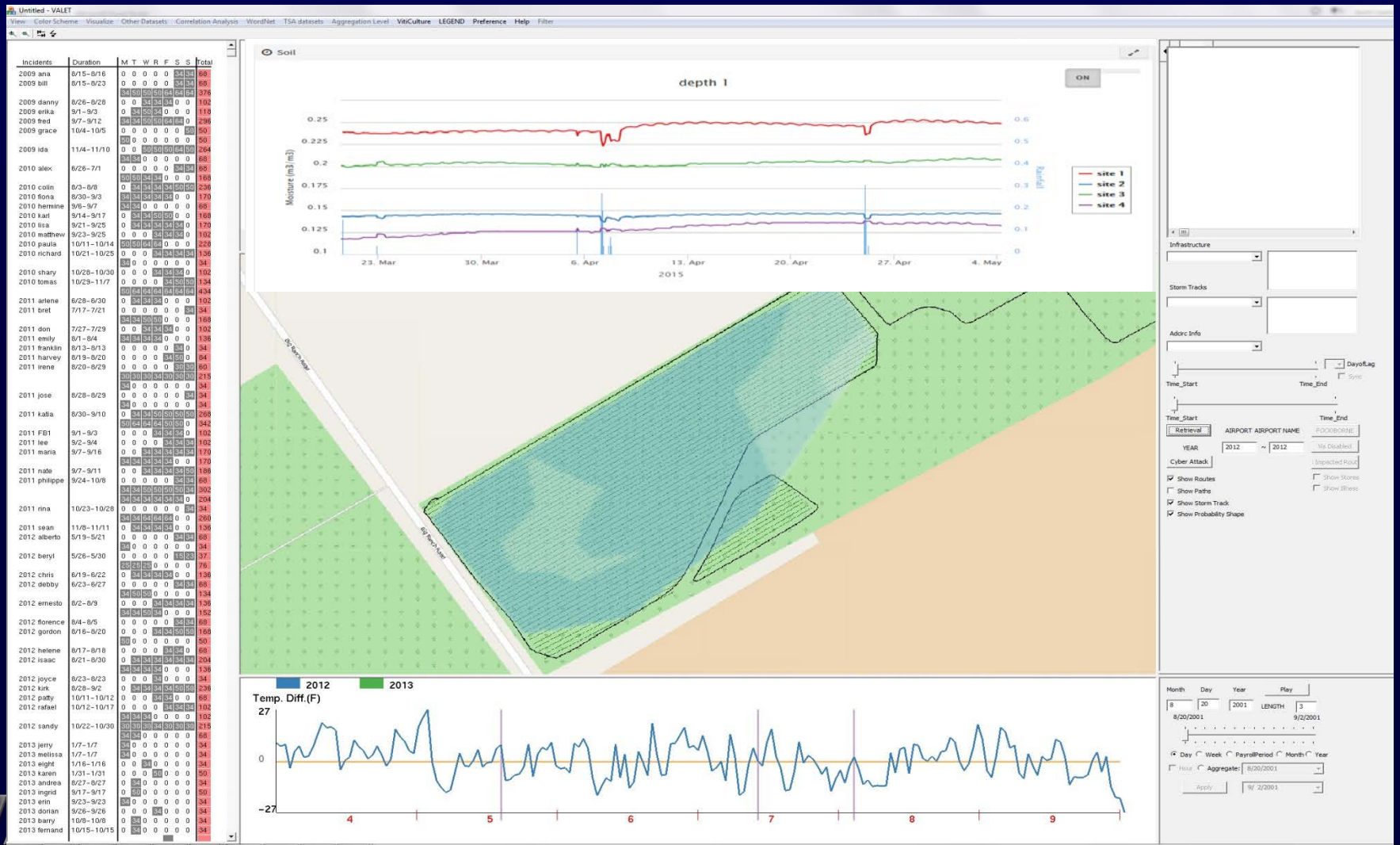
- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality





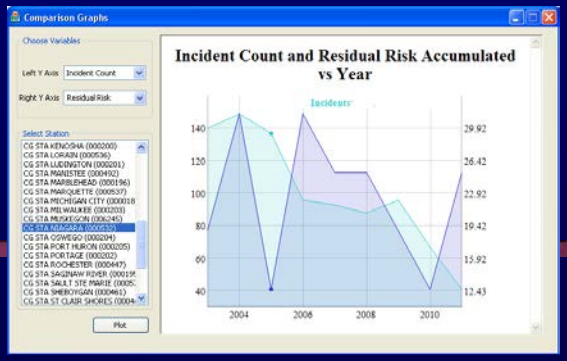
VACCINE

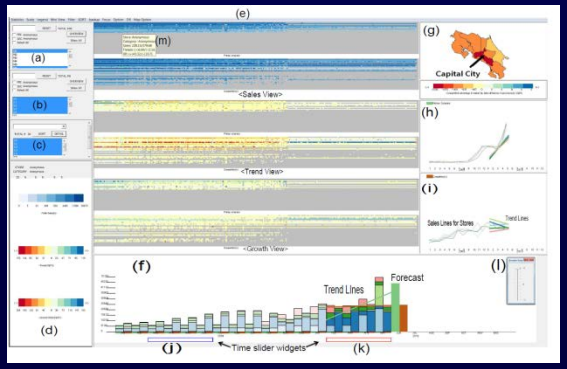
Precision Agriculture VA



Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

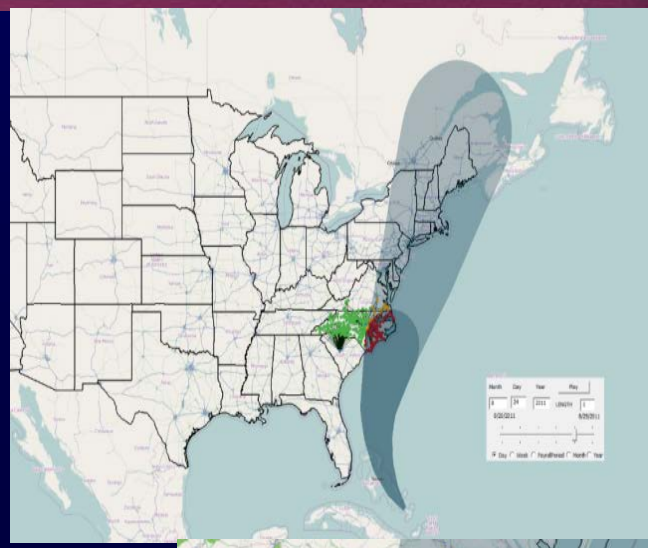


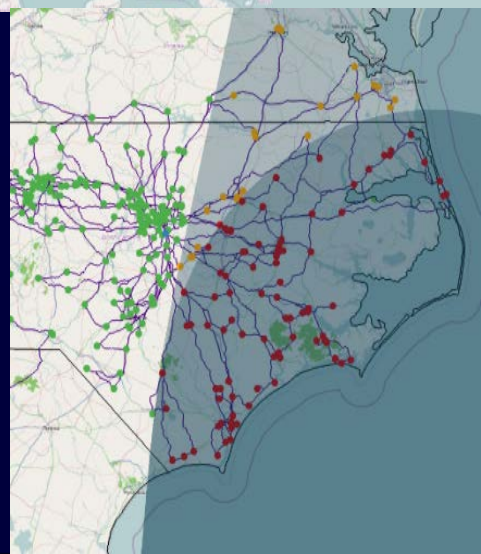
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

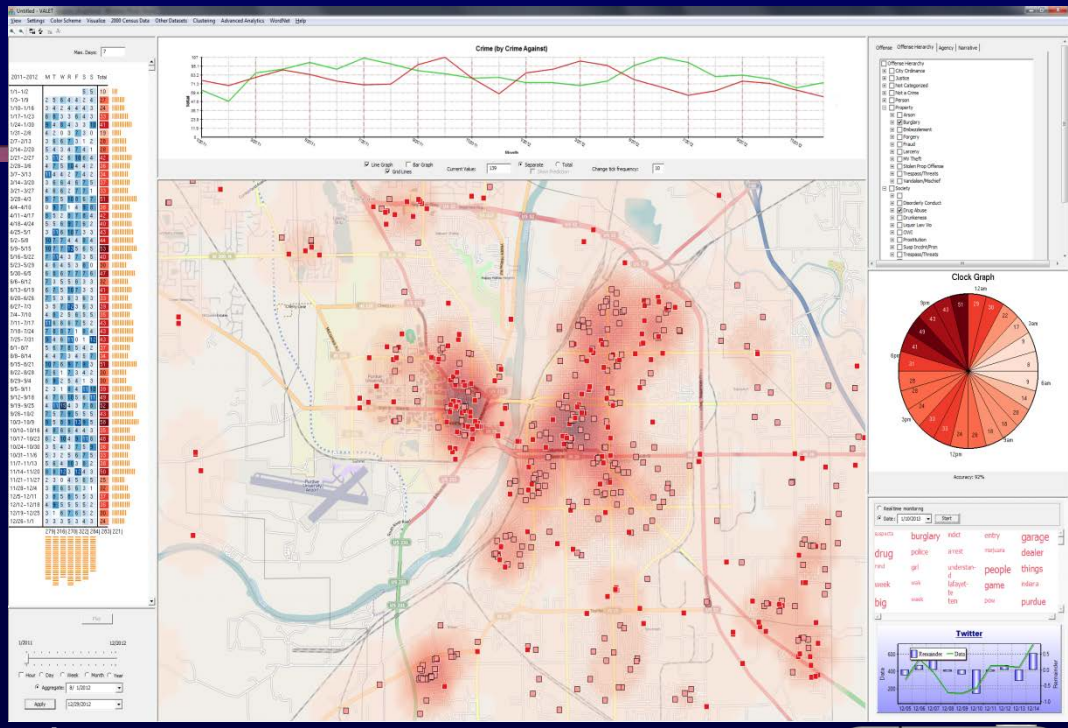
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





VACCINE

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

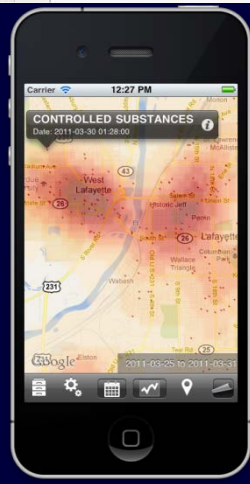
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD



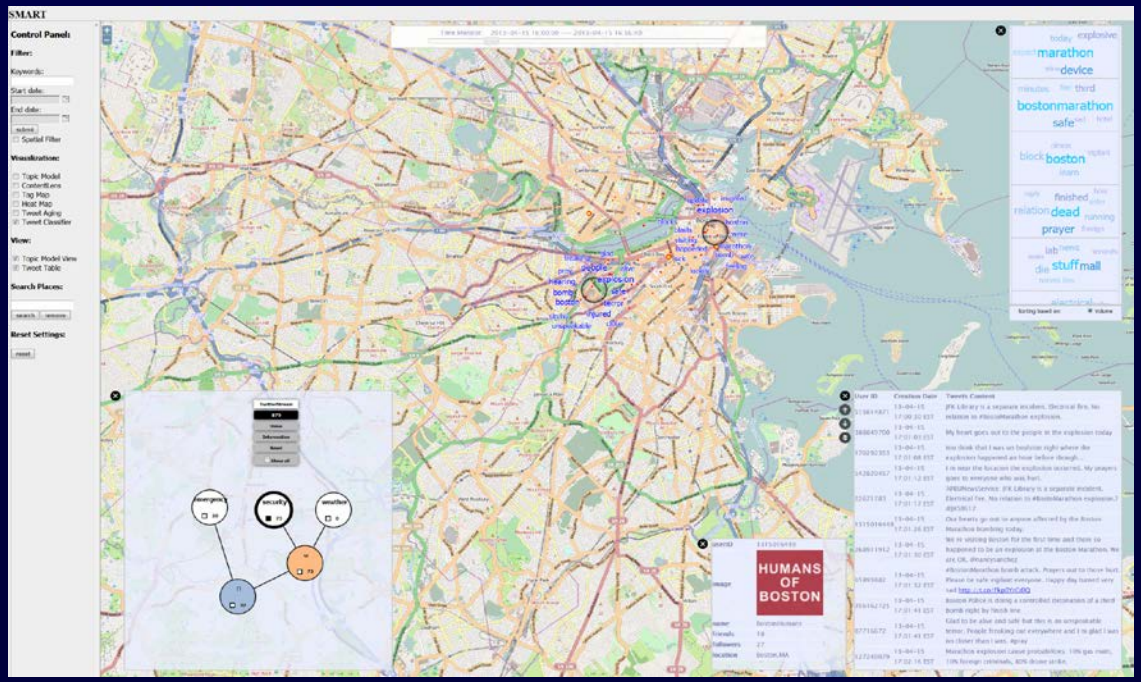
VACCINE

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

May 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

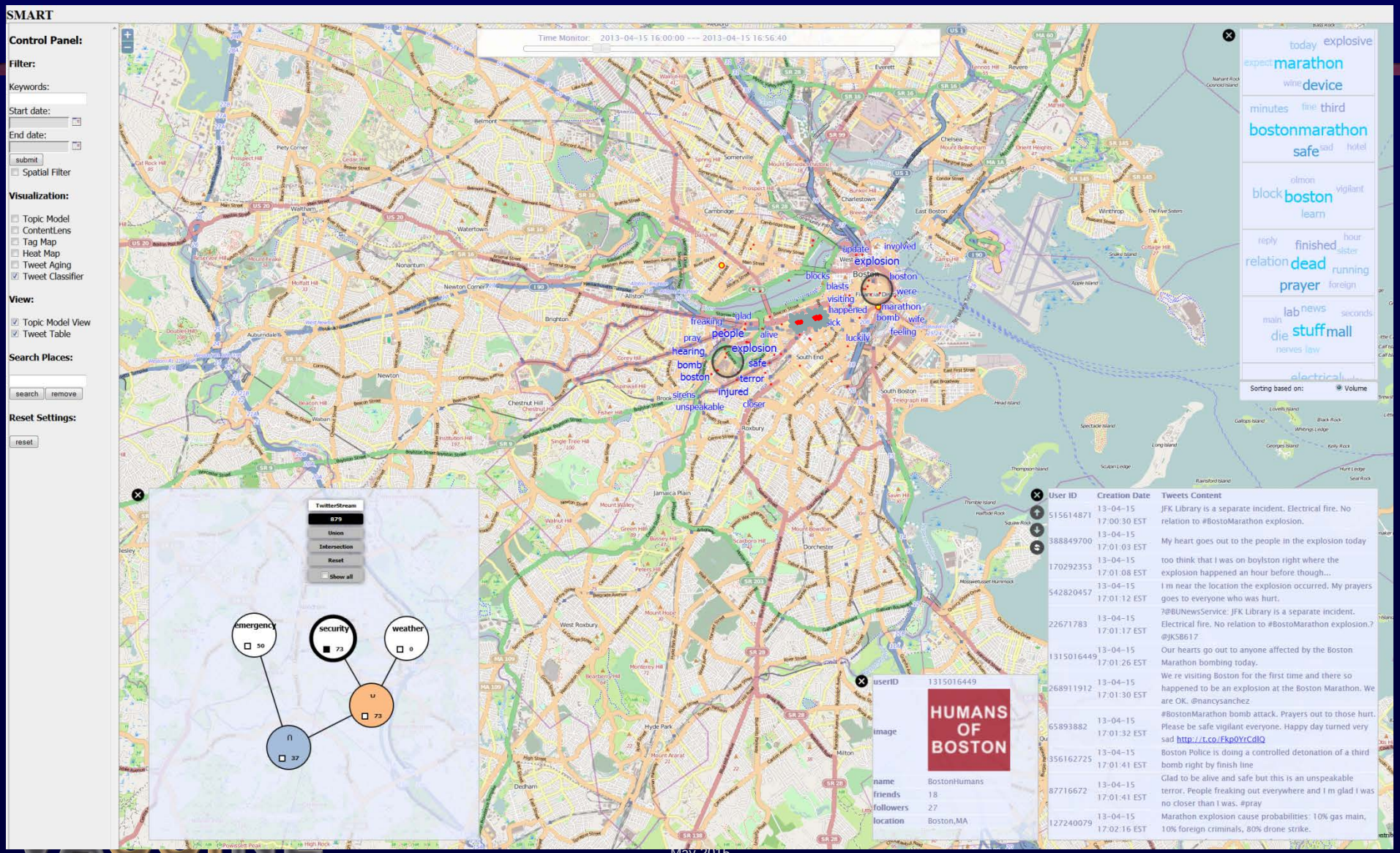
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

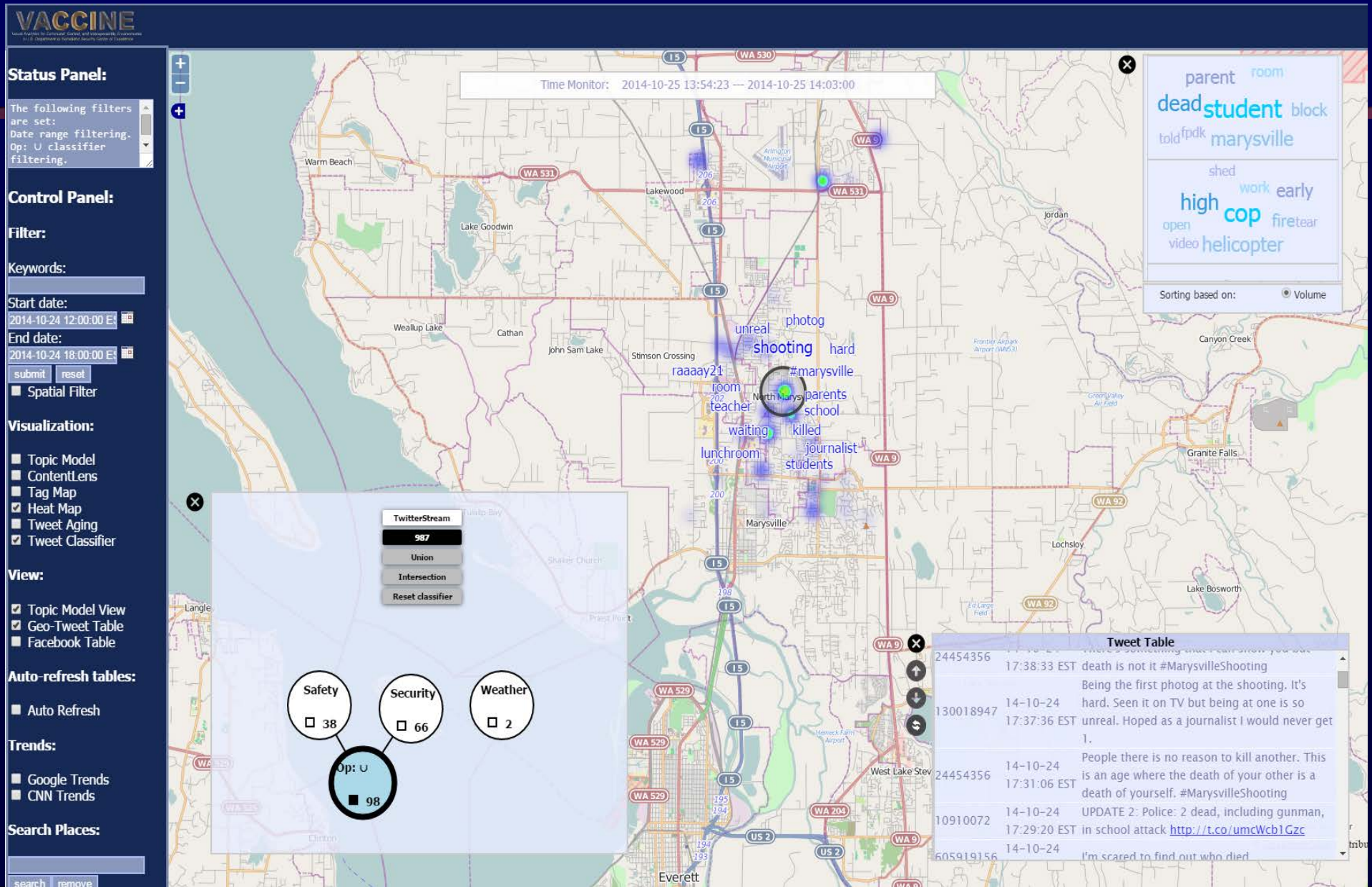


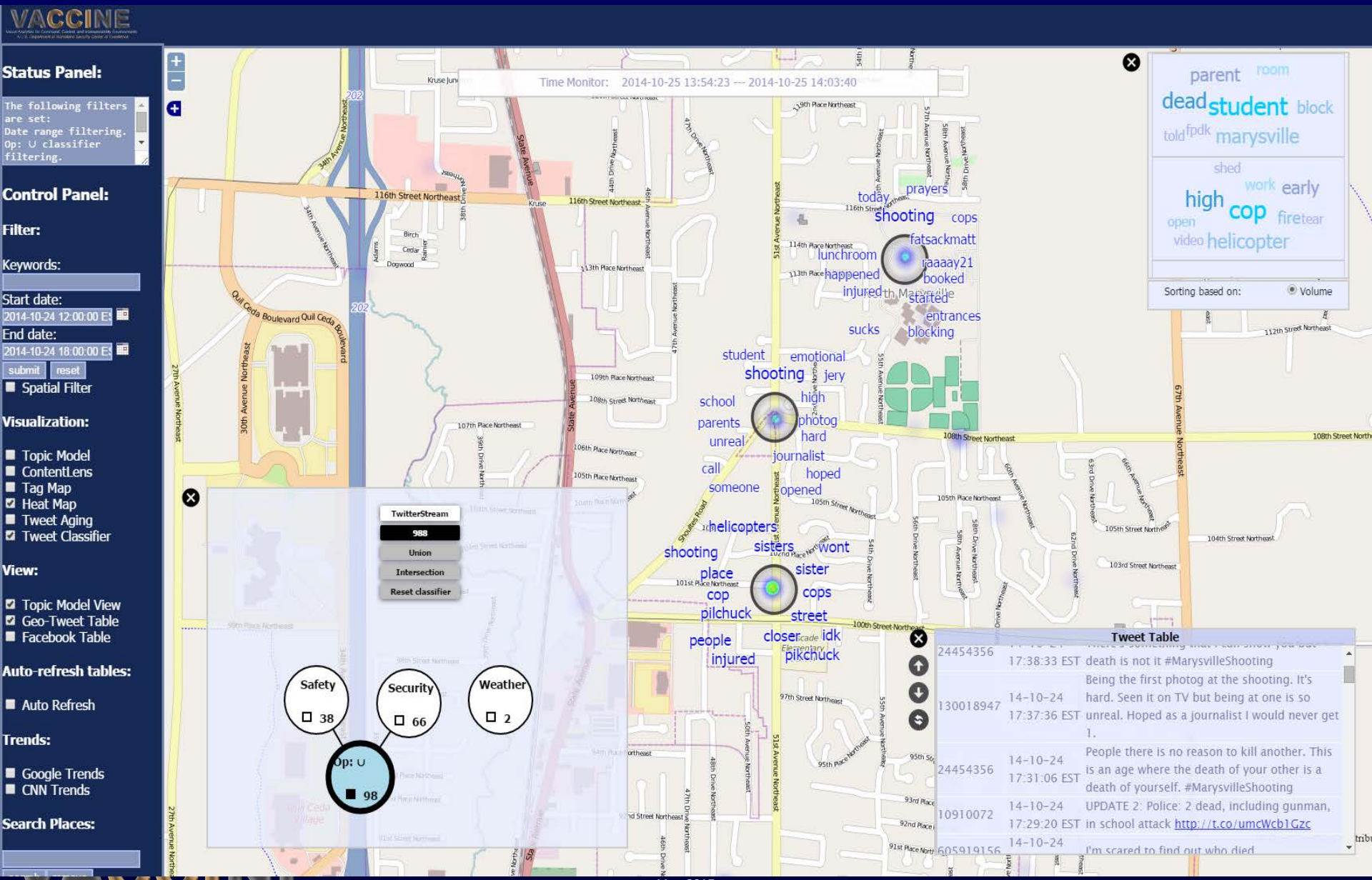
VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24 17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24 17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24 17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died



Status Panel:

The following filters are set:
Single tweet filtering.
Date range filtering.

Control Panel:

Filter:

Keywords:

Start date:
2014-10-24 12:00:00 ES

End date:
2014-10-24 18:00:00 ES

submit reset

- Spatial Filter

Visualization:

- Topic Model
- ContentLens
- Tag Map
- Heat Map
- Tweet Aging
- Tweet Classifier

View:

- Topic Model View
- Geo-Tweet Table
- Facebook Table

Auto-refresh tables:

- Auto Refresh

Trends:

- Google Trends
- CNN Trends

Search Places:

Marysville-Pilchuck High Sch

search remove

Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48

pilchuck tear video cop pray homestuck job couple marysville high school sad block kid told fire shoot

Sorting based on: Volume

Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

Location of the school

North Marysville

TwitterStream
990
Union
Intersection
Reset classifier

Safety 38
Security 66
Weather 2

userID 23845798

image

name joshtrujillo
friends 2146
followers 6674
location Seattle,WA

Tweet Table

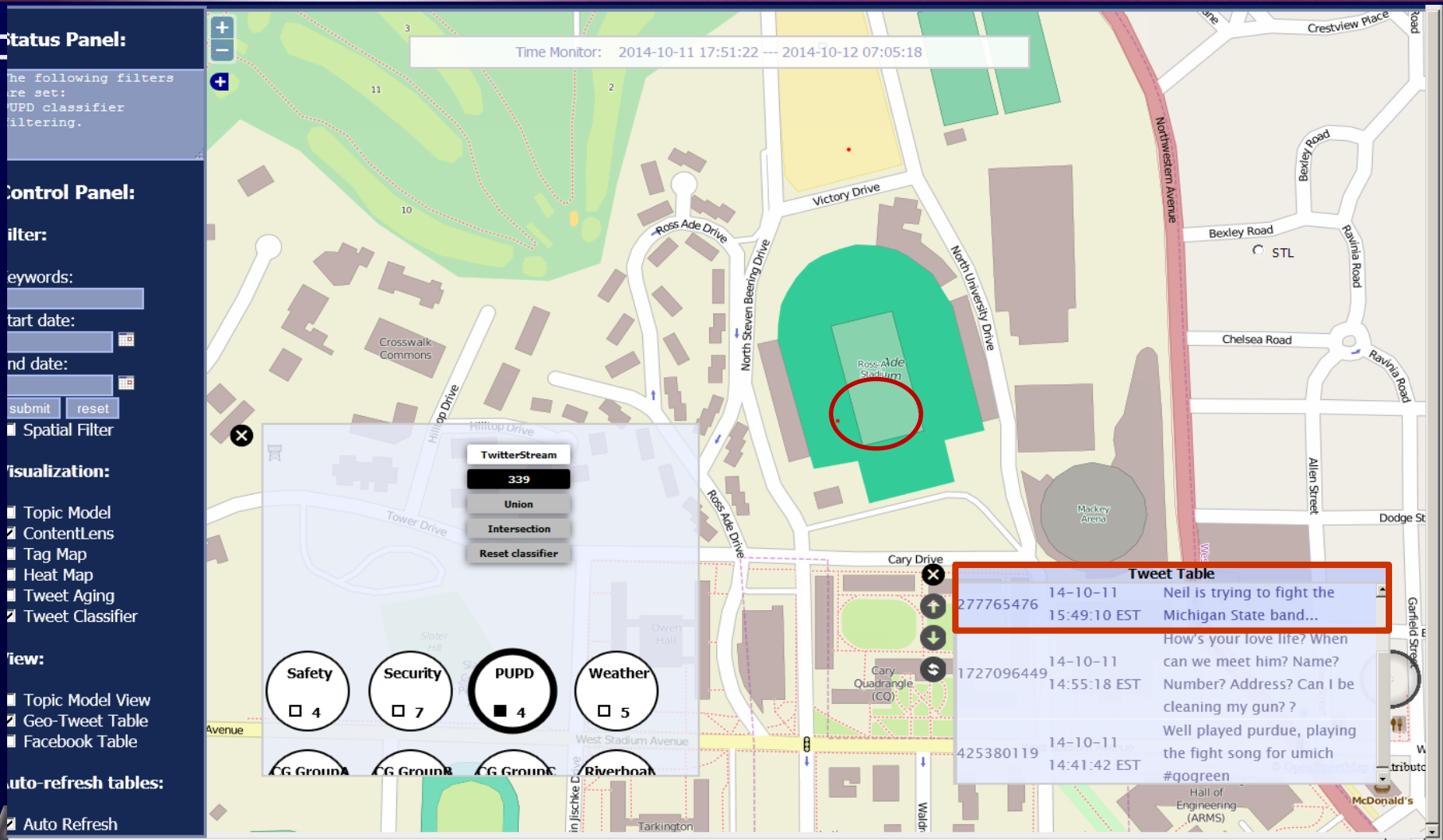
User ID	Creation Date	Tweets Content
23845798	14-10-24 16:47:14 EST	Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

Where Can It Be Helpful?

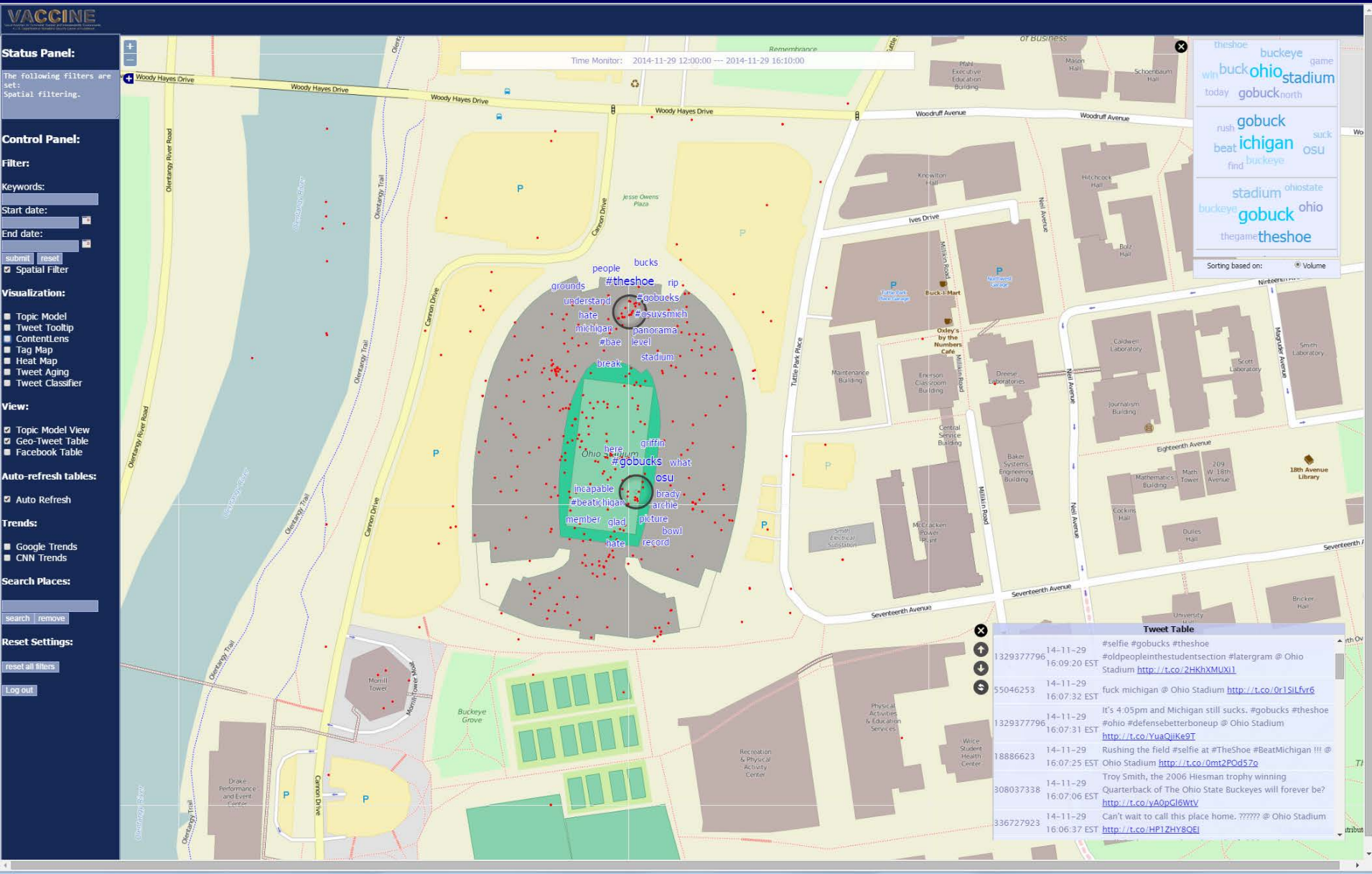
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



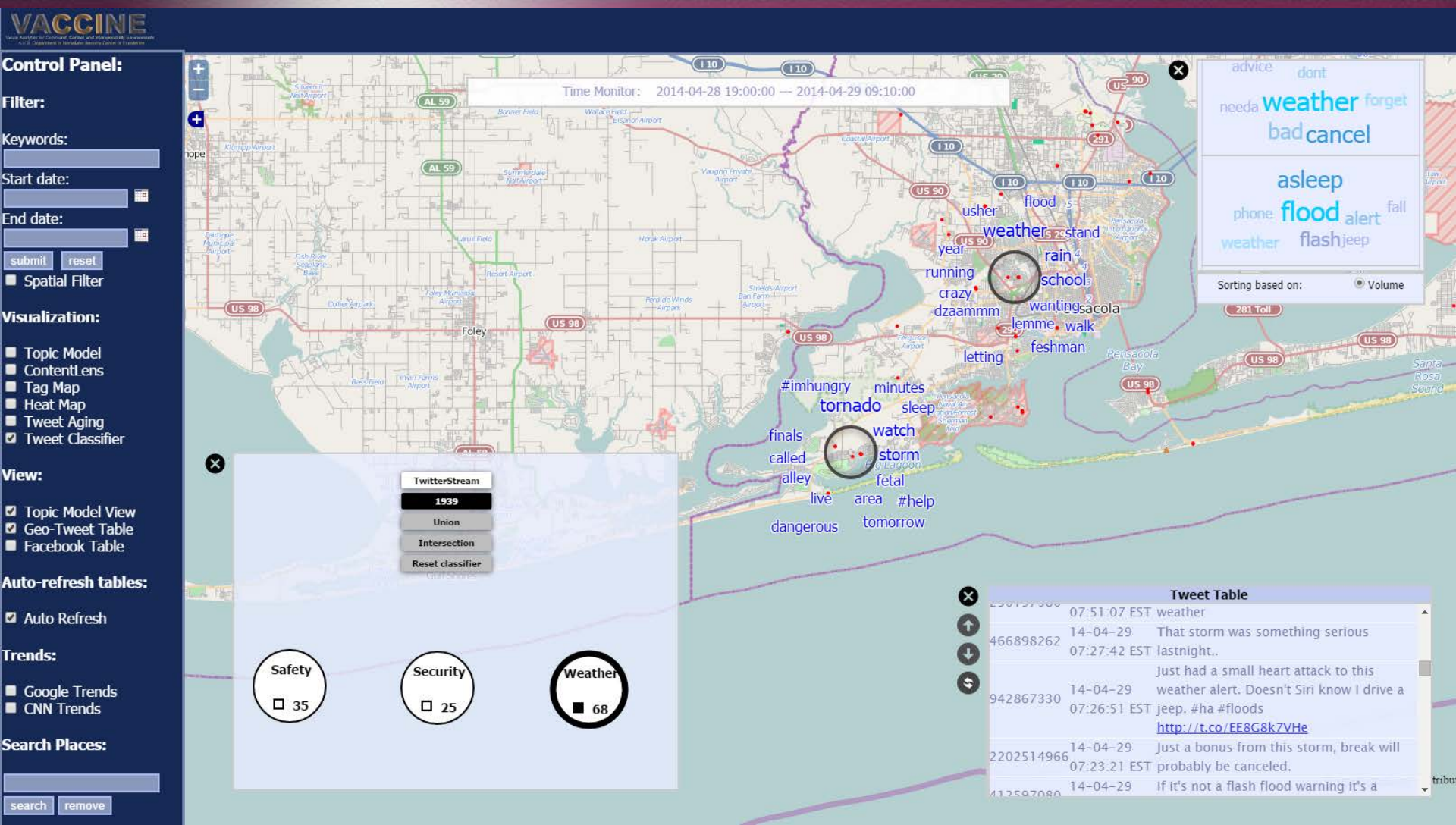
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



Where Can It Be Helpful?

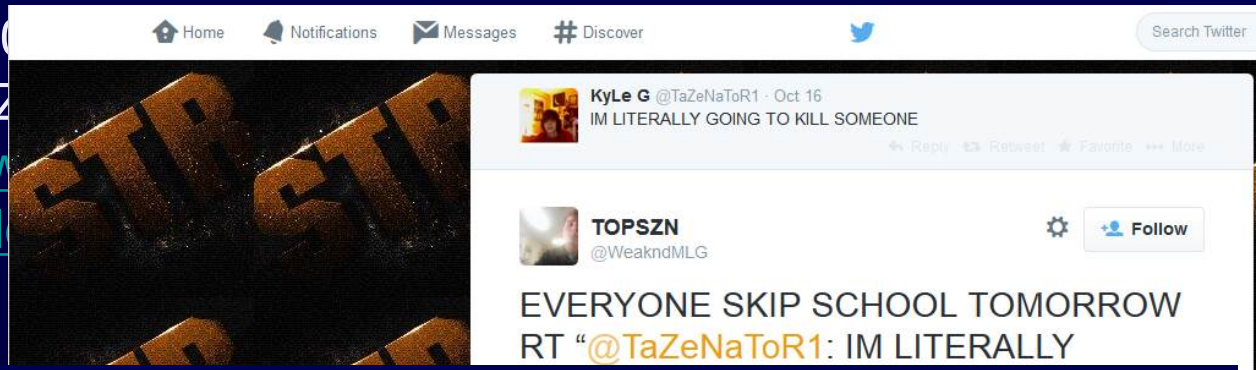


Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE



Wayfast Whitey
@WayFastWhitey44

Follow

pnbarrie

Follow

1 month ago · ARIA Resort & Casino

Nice production at the John Deere event Thursday night for Conexpo. #conexpo #johndeere #rdoequipment #lovemyjob

shepnelson, ottoreynolds, mycranelife and connshawn like this.

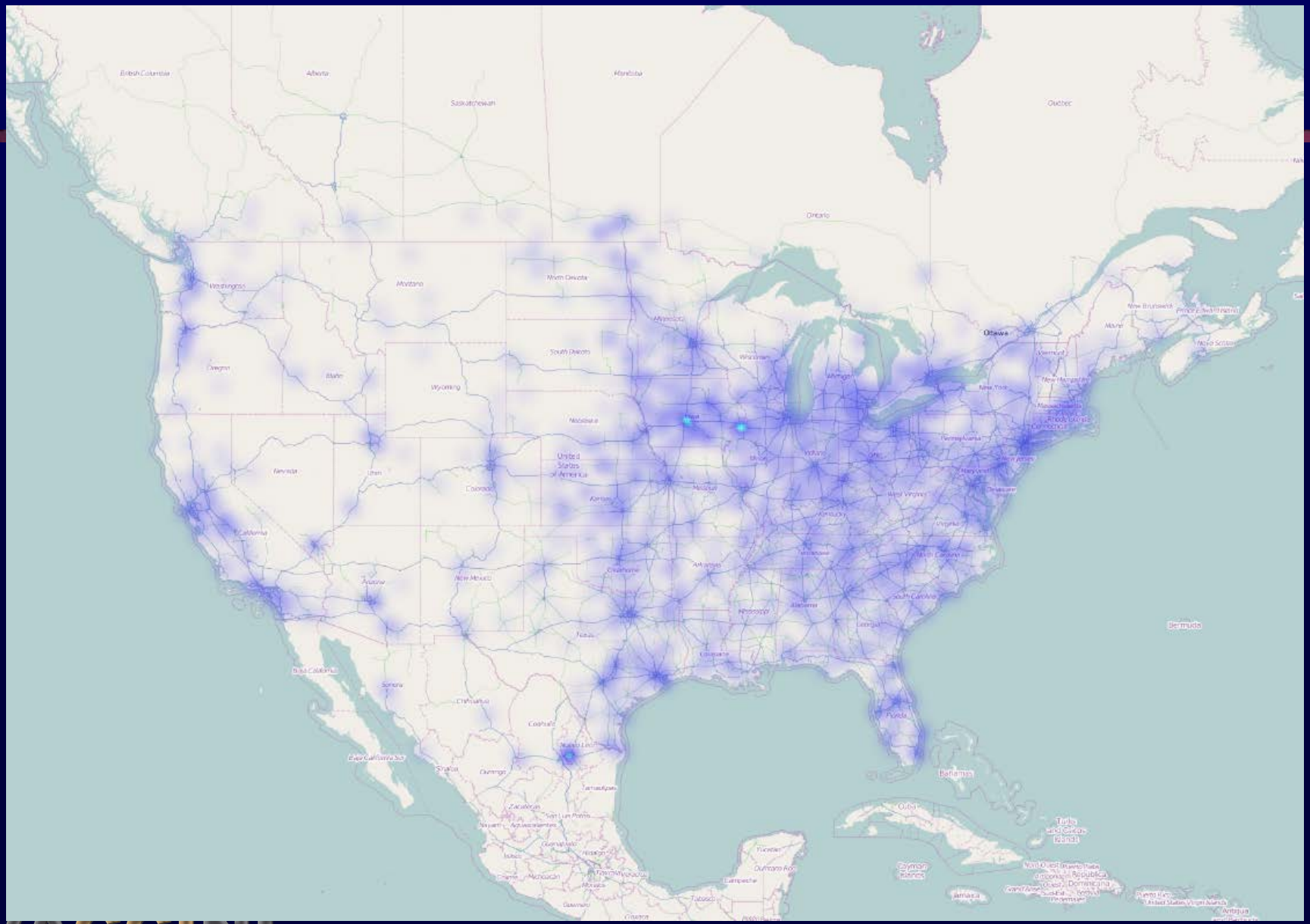
Leave a comment...

VACCINE

May 2

Reply to @WayFastWhitey44

John Deere


British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
Ottawa
Montana
North Dakota
South Dakota
Wyoming
Nebraska
United States of America
Colorado
Kansas
Oklahoma
Texas
New Mexico
Arizona
Nevada
California
Oregon
Idaho
Utah
Minnesota
Wisconsin
Michigan
Illinois
Indiana
Ohio
Kentucky
Tennessee
Mississippi
Alabama
Georgia
Florida
South Carolina
North Carolina
Virginia
West Virginia
Pennsylvania
New York
Vermont
New Hampshire
Maine
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
Louisiana
Arkansas
Missouri
Iowa
Bermuda
Bahamas
Cuba
Jamaica
Cayman Islands
Haiti
Dominican Republic
Puerto Rico
Baja California
Sonora
Chihuahua
Coahuila
Nuevo Leon
Tamaulipas
Durango
Zacatecas
San Luis Potosi
Aguascalientes
Jalisco
Michoacan
Guanajuato
Queretaro
Hidalgo
Veracruz
Tabasco
Campeche
Yucatan
Quintana Roo

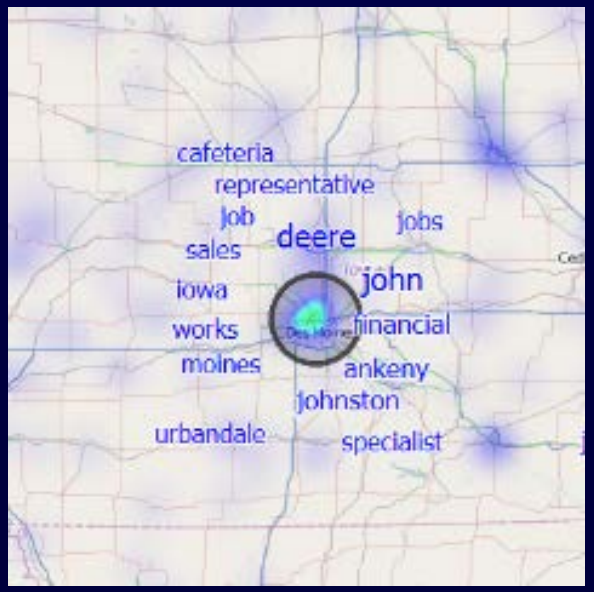
Ten Dominant Topics Related to John Deere



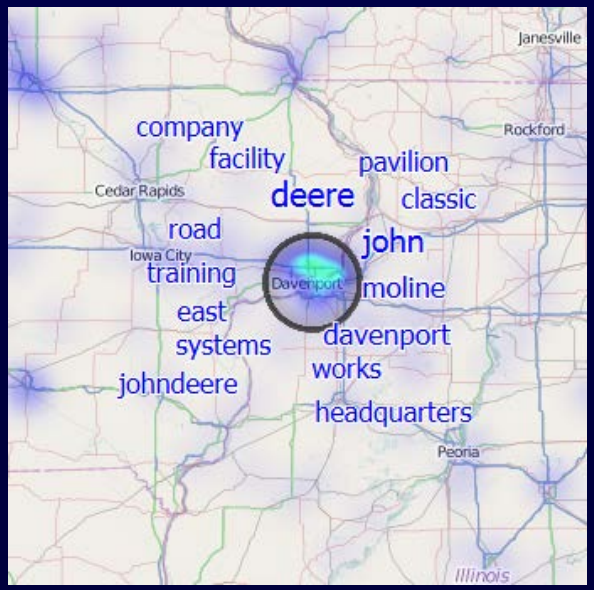


Keywords(Contentlens) in the 3 Hotspots

Des Moines, Iowa



Davenport, Iowa



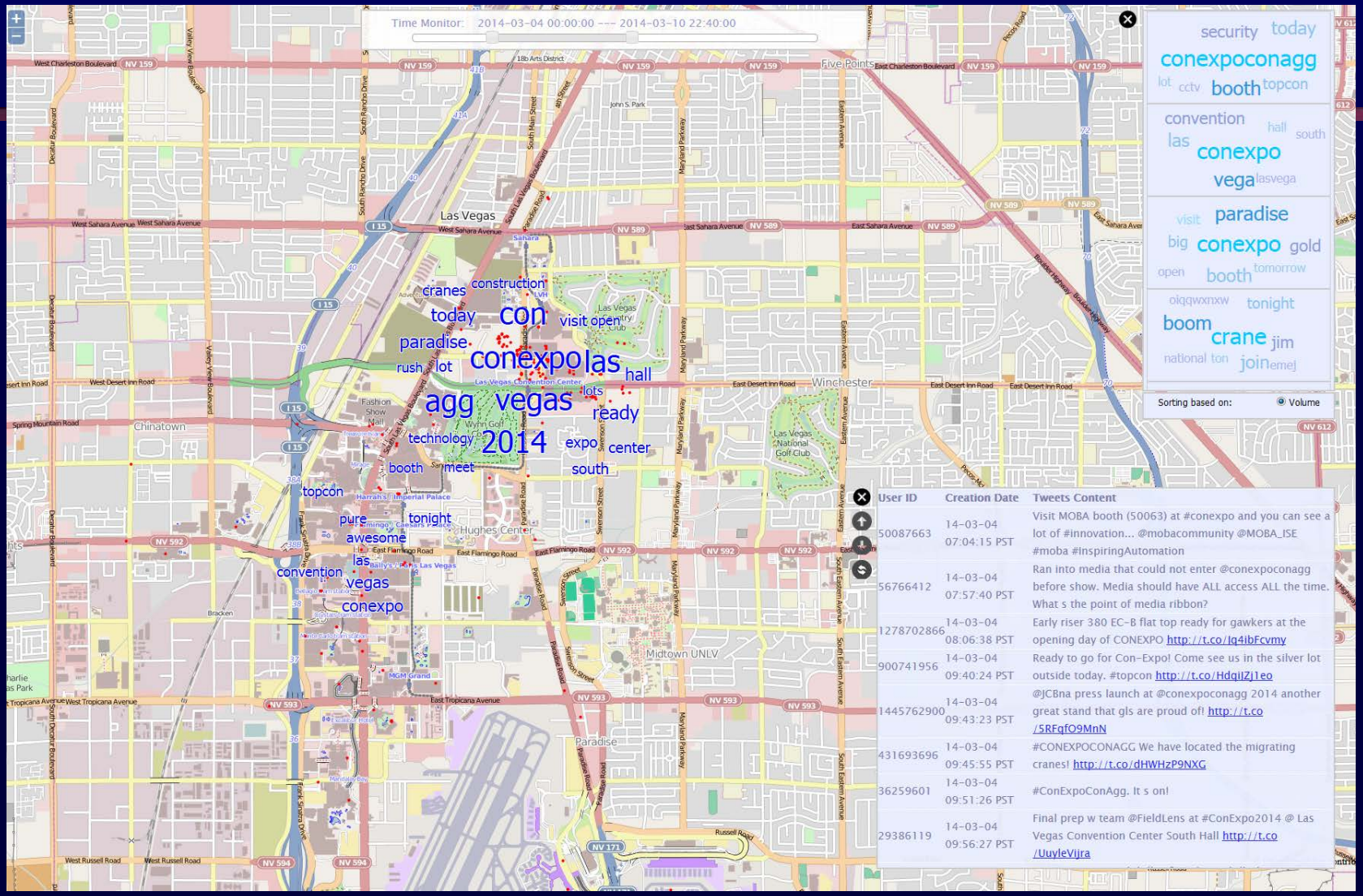
Monterrey, Mexico



Hashtags Related to John Deere

- #johndeere, #nothingrunslikeadeere
- #tractor, #tractortuesday
- #farm, #farming, #farmlife
- #jobs, #work, #harvest
- #green, #country, #summer
- #diesel, # dieselgenerator, #jasonaldean
- #genset, #wayouthere

Con Expo


Time Monitor: 2014-03-04 00:00:00 --- 2014-03-10 22:40:00
cranes
construction
today
con
visit
open
paradise
conexpo
las
rush
lot
hall
lots
agg
vegas
ready
technology
2014
expo
center
booth
meet
south
topcon
pure
tonight
awesome
las
convention
vegas
conexpo
security
today
conexpoconagg
lot
cctv
booth
topcon
convention
hall
south
las
conexpo
vega
lasvega
visit
paradise
big
conexpo
gold
open
booth
tomorrow
oiqqwxnxw
tonight
boom
crane
jim
national
ton
join
emej
Sorting based on:
Volume
User ID
Creation Date
Tweets Content
50087663
14-03-04 07:04:15 PST
Visit MOBA booth (50063) at #conexpo and you can see a lot of #innovation... @mobacommunity @MOBA_ISE #moba #InspiringAutomation
56766412
14-03-04 07:57:40 PST
Ran into media that could not enter @conexpoconagg before show. Media should have ALL access ALL the time. What s the point of media ribbon?
1278702866
14-03-04 08:06:38 PST
Early riser 380 EC-B flat top ready for gawkers at the opening day of CONEXPO http://t.co/Iq4ibFcvmy
900741956
14-03-04 09:40:24 PST
Ready to go for Con-Expo! Come see us in the silver lot outside today. #topcon http://t.co/HdqiIZj1eo
1445762900
14-03-04 09:43:23 PST
@JCBna press launch at @conexpoconagg 2014 another great stand that gls are proud of! http://t.co/5RFqfO9MnN
431693696
14-03-04 09:45:55 PST
#CONEXPOCONAGG We have located the migrating cranes! http://t.co/dHWHzP9NXG
36259601
14-03-04 09:51:26 PST
#ConExpoConAgg. It s on!
29386119
14-03-04 09:56:27 PST
Final prep w team @FieldLens at #ConExpo2014 @ Las Vegas Convention Center South Hall http://t.co/UuyleVijra

Tweets related to John Deere skidder

wave3news @wave3news · Apr 4
RT to help ISP: Reward offered for stolen **John Deere** log **skidder**, bit.ly/QMzLgr #wave3news pic.twitter.com/XqxOZ7lggn



Buster Beans @drifhfbfbfj · 22h
Just got a new **John Deere skidder** today pic.twitter.com/c8WiMnXVfV

Hide photo Reply Retweet Favorite More



FAVORITE 1

7:46 PM - 26 Apr 2014 · De

Retweeted 53 times
John Deere @JohnDeere · Apr 24
If a lion is the king of the jungle, a **John Deere Skidder** is the king of the forest. pic.twitter.com/afZ76bXvVT



Expand Reply Retweet

Service Manuals @EbookManuals · Apr 4
#download **John Deere** 848G **Skidder** Technical Manual Download TM1 dlvr.it/5KChgF @EbookManuals

Expand Reply Retweet Favorite

J Lowe @jennyfaithlowe · Apr 22
Firestone 16 ply **Skidder** Tires on **john deere** wheels 7/8ths holes #ebaymobile bit.ly/1nn9l24

Expand Reply Retweet Favorite More

Want Ad Digest @Wantaddigest · Apr 16
John Deere 540 B Cable **Skidder** LOWVILLE NY: 23x/x26 tire ring chains diamond chains, runs very good, needs l... bit.ly/1ePVzwM

Expand Reply Retweet Favorite

Burge Mcmanus @manualdownloads · Apr 13
#download **John Deere** 640D **Skidder**, 648D Operation and Tests Manual dlvr.it/5NyshS

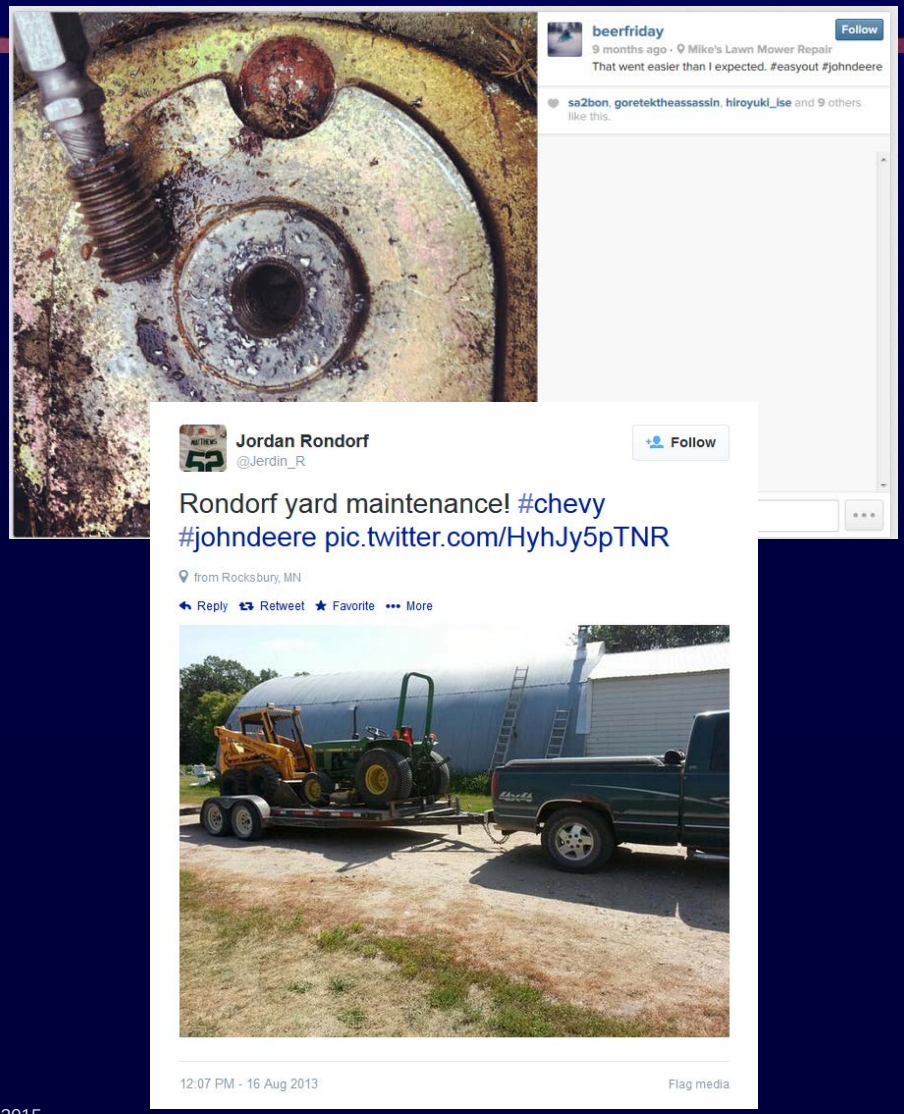
Collapse Reply Retweet Favorite More

9:15 PM - 13 Apr 2014 · Details

VACCINE

Tweets related to ‘maintain, maintenance, repair’

User ID	Creation Date	Tweets Content
933628327	13-06-09 14:27:17 EST	And, for your viewing pleasure, a nicely maintained 1930s John Deere. http://t.co/Zavel5QnDI
50371494	13-06-07 10:13:50 EST	The Class of 2013 presents its class gift at commencement – a John Deere Gator for the Maintenance Department. http://t.co/ju1rcbEuLm
494325743	13-08-16 15:07:21 EST	Rondorf yard maintenance! #chevy #johndeere http://t.co/HyhJy5pTNR
404120195	13-09-03 16:13:01 EST	Just saw a clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #Townielife
1038577016	13-07-28 22:58:57 EST	#labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenan... http://t.co/azERmrPMt5
502127999	13-08-07 18:47:22 EST	That went easier than I expected. #easyout #johndeere @ Mike s Lawn Mower Repair http://t.co/8pgiu0Uv7g

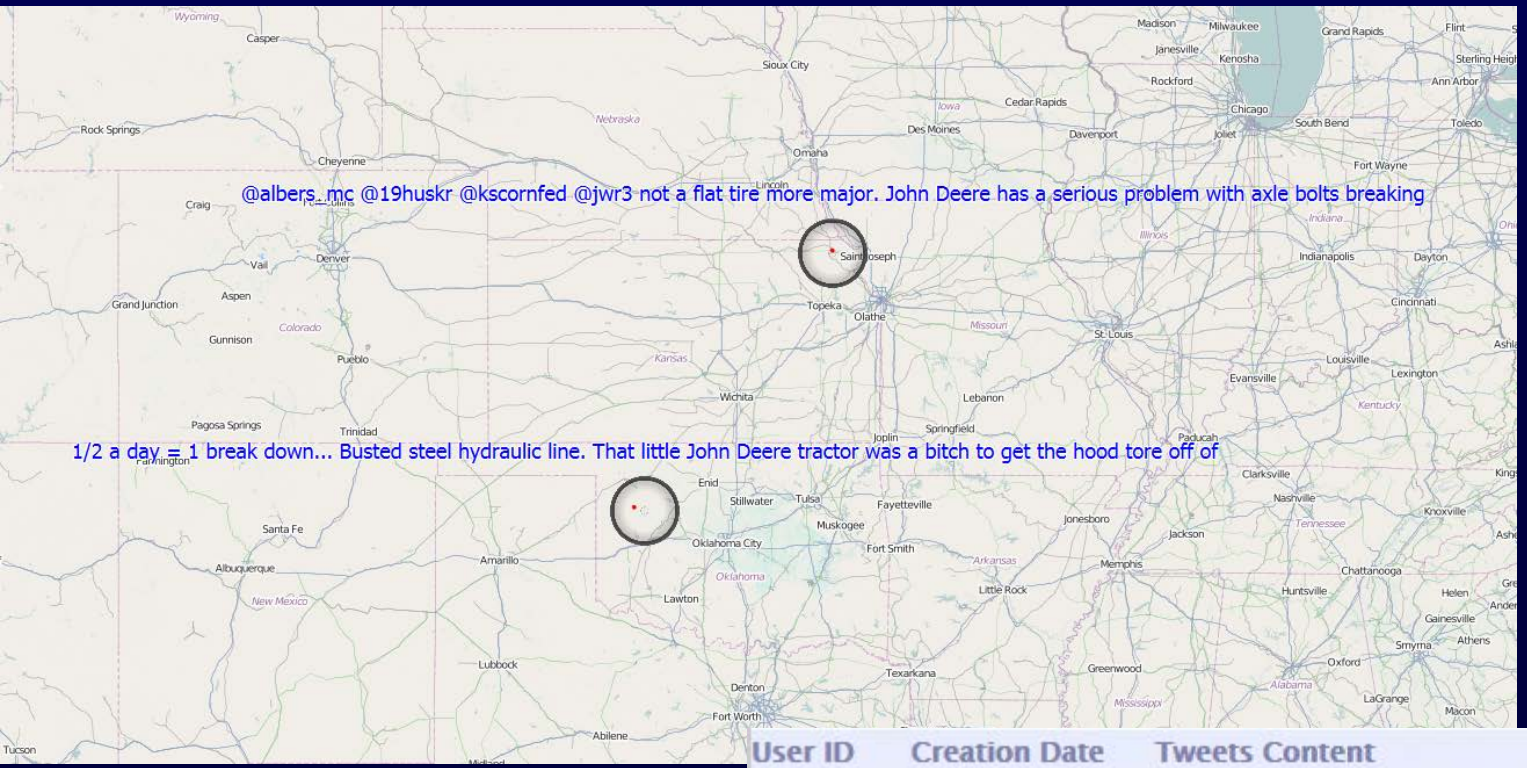


‘Maintain”


Carl
@pzldcharlatan
Follow
And, for your viewing pleasure, a nicely maintained 1930s John Deere.
pic.twitter.com/Zavel5QnDl
from Damascus, PA
Reply Retweet Favorite More


Shady Side Academy
@shady_side
Follow
The Class of 2013 presents its class gift at commencement - a John Deere Gator for the Maintenance Department.
pic.twitter.com/ju1rchEuLm
from Fox Chapel, PA
Reply Retweet Favorite More
RETWEETS
3
FAVORITES
2

Tweets related to ‘break down, shut down, stop working’



User ID	Creation Date	Tweets Content
630487309	13-06-14 15:55:47 EST	1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of
35458727	13-10-18 22:59:17 EST	@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking

VACCINE

For Further Information

www.VisualAnalytics-CCI.org



VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

VACCINE: Empowering Effective Decision Making Through Visual Analytics

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

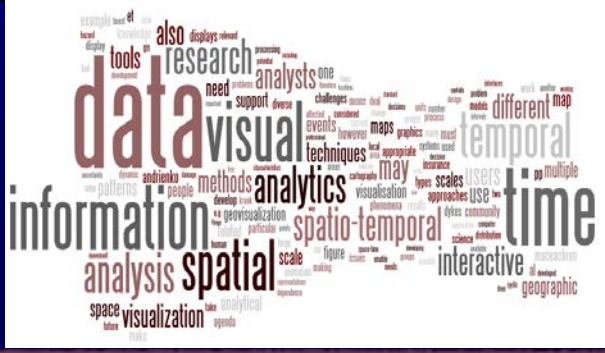
May 2015

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 29 Institutions



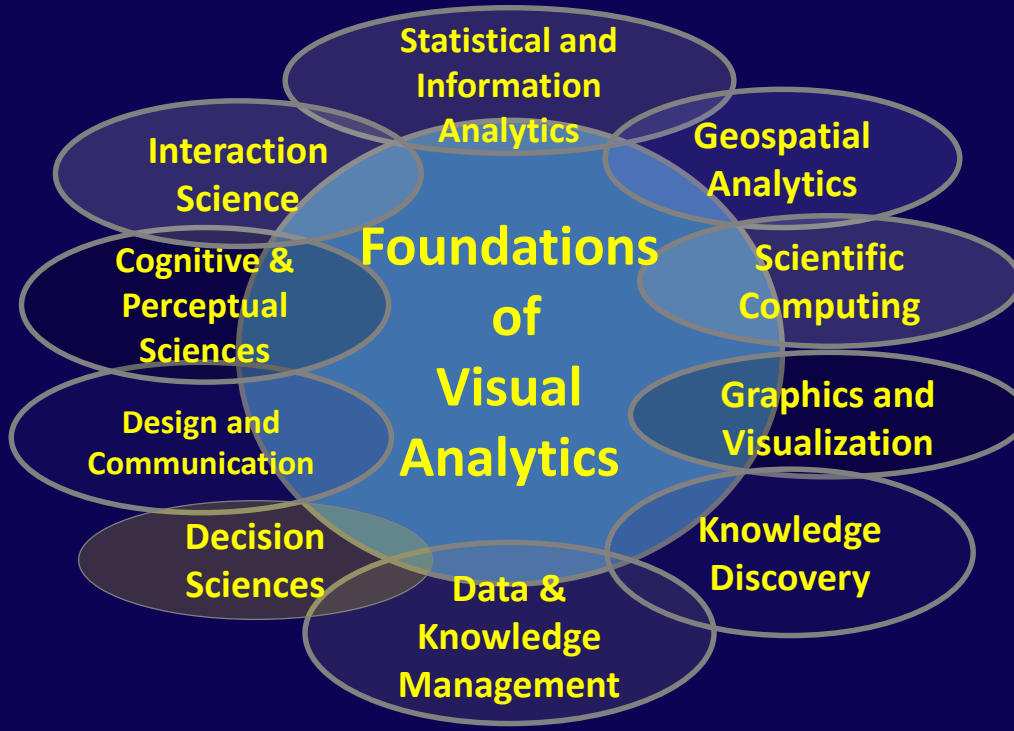
- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California San Diego
- University of Maryland
- University Washington
- Arizona State University
- Bethune-Cookman U.
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

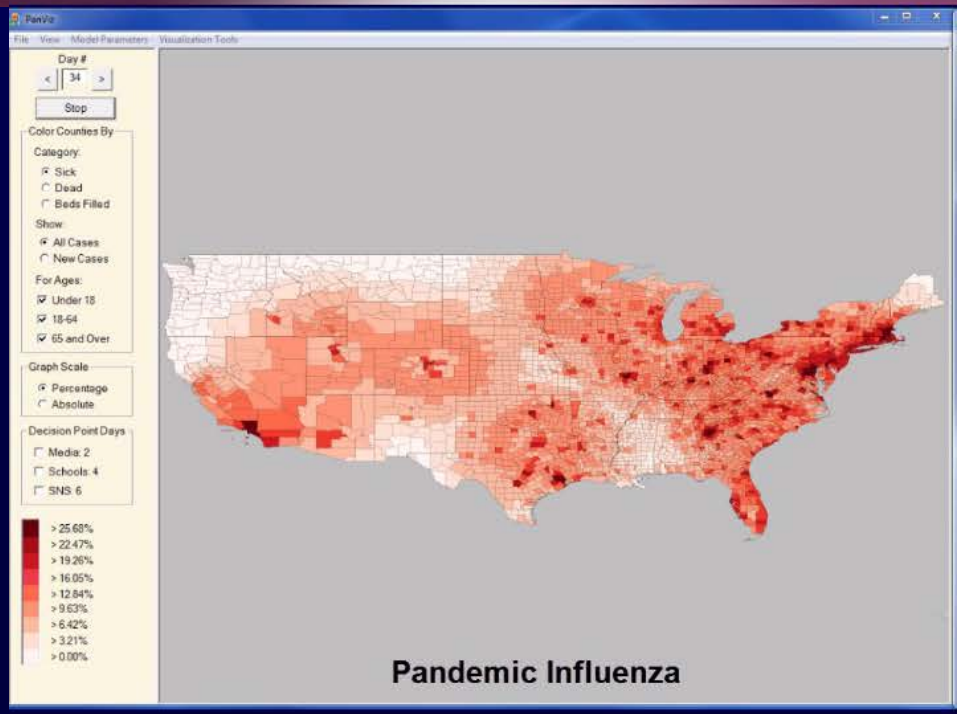
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

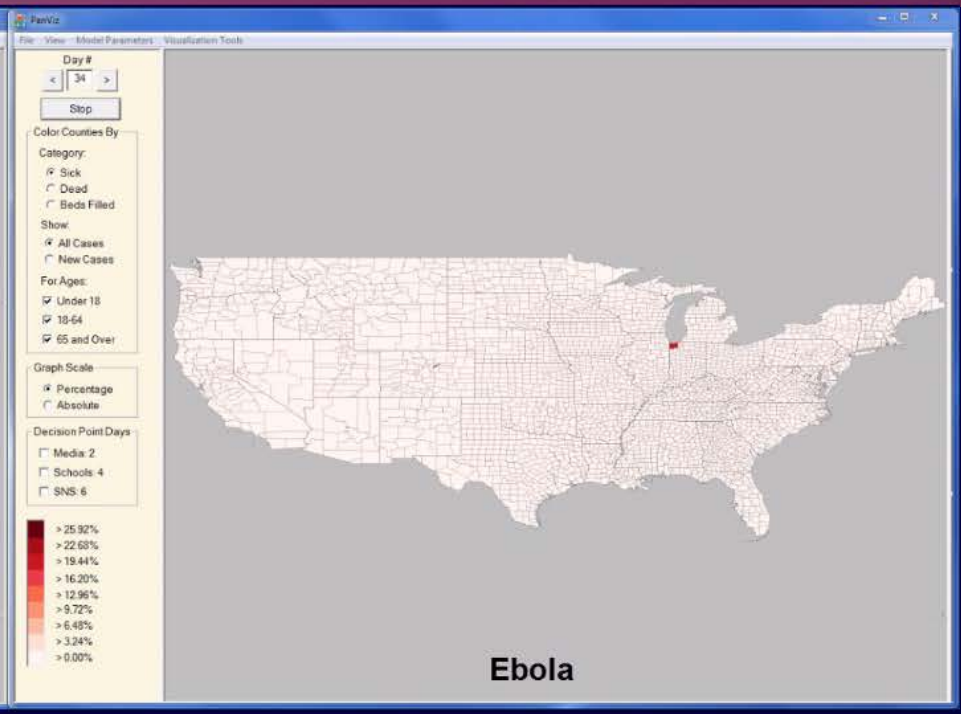
VADM Robert Parker, 2012 MRS Keynote Address



Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL





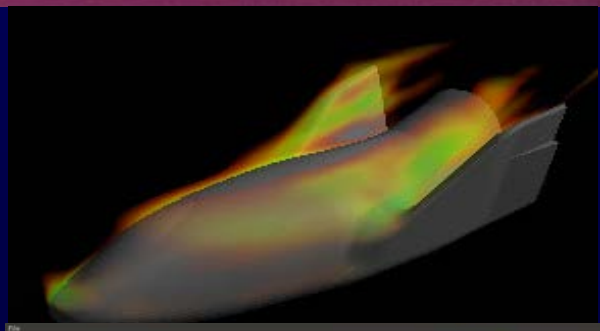
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

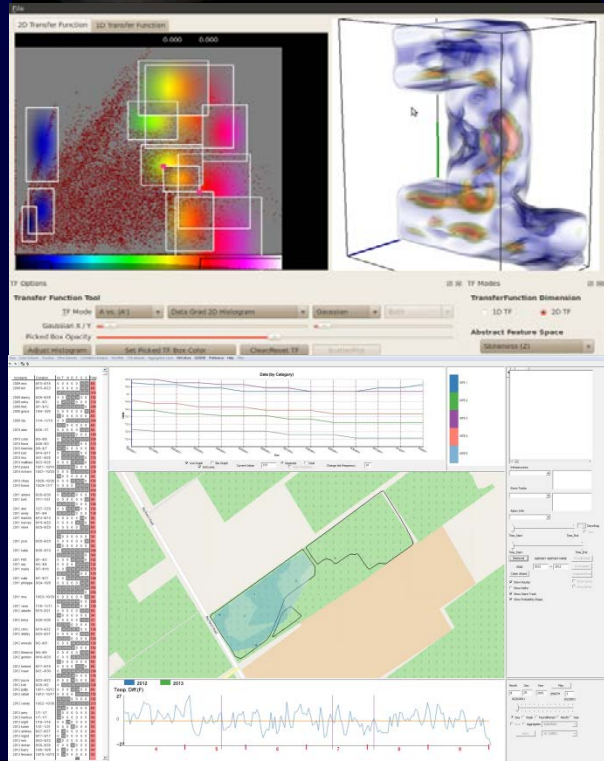
VACCINE

Scientific Discovery



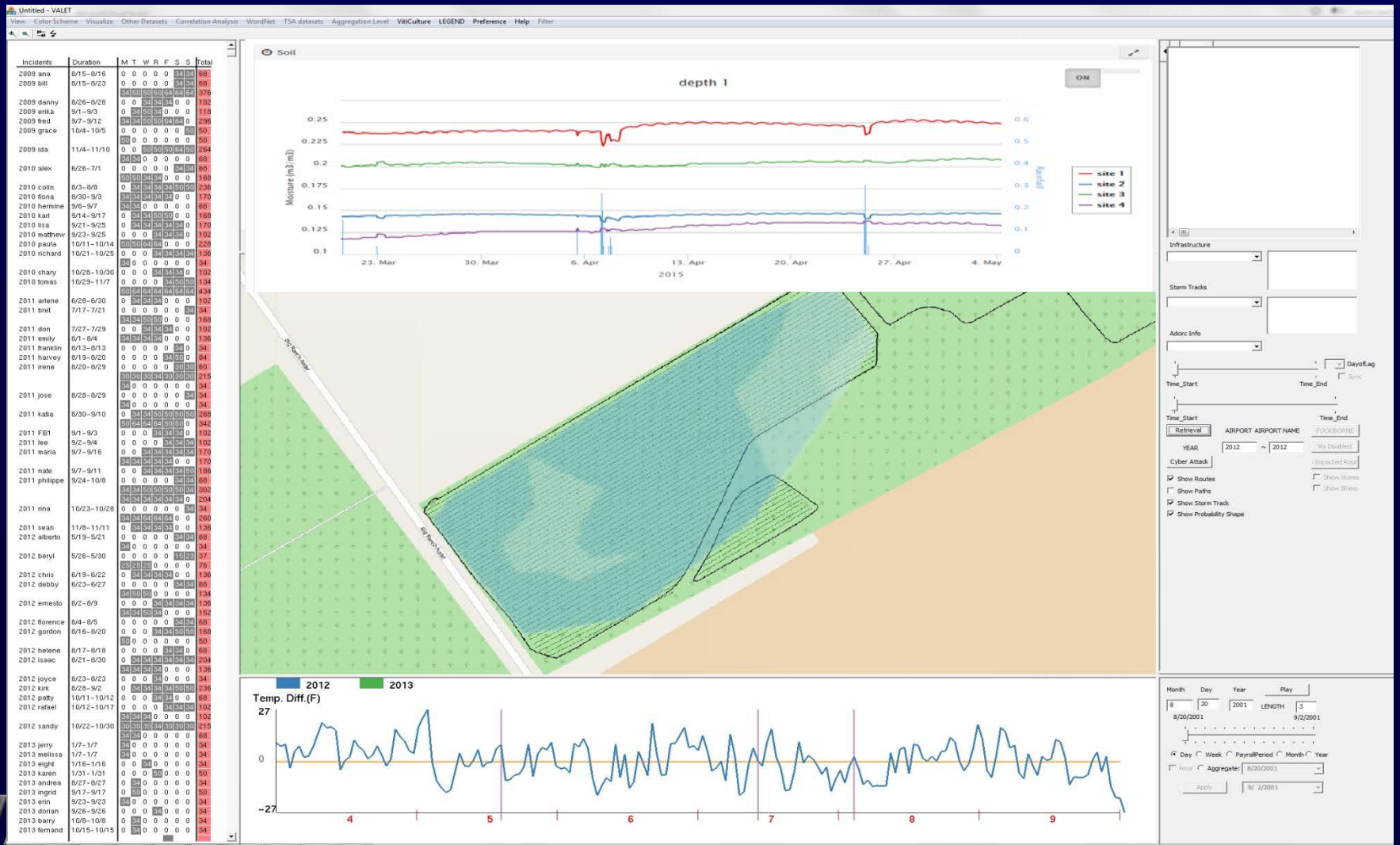
- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality





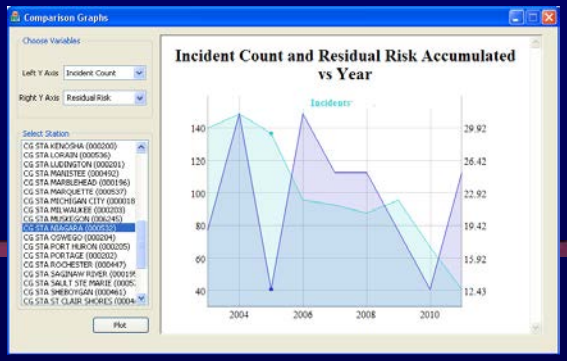
VACCINE

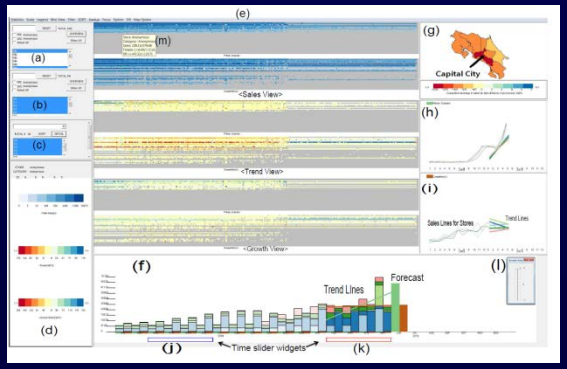
Precision Agriculture VA



Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation







Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

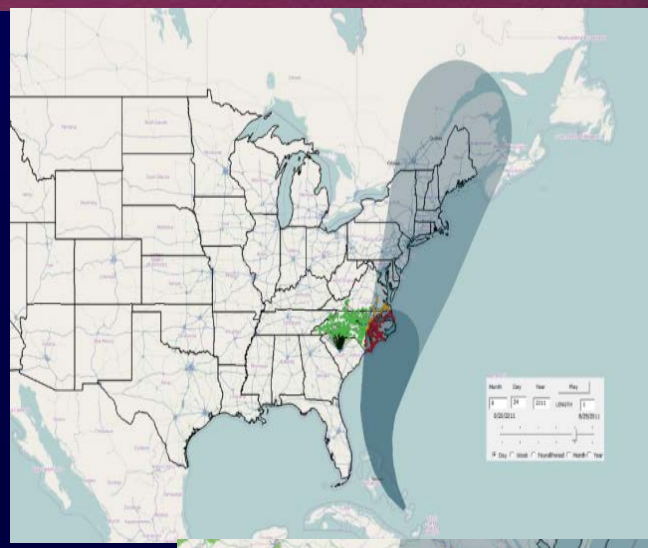


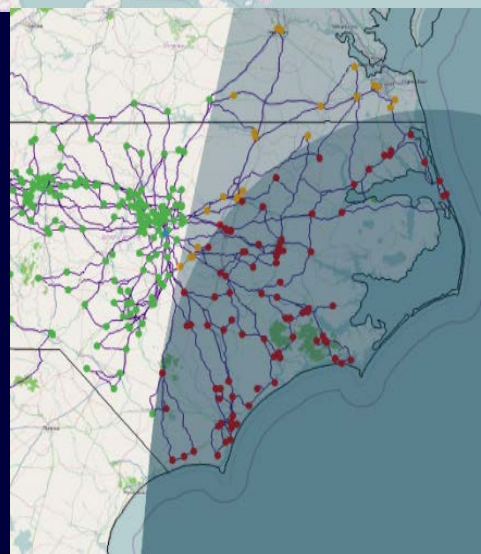
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

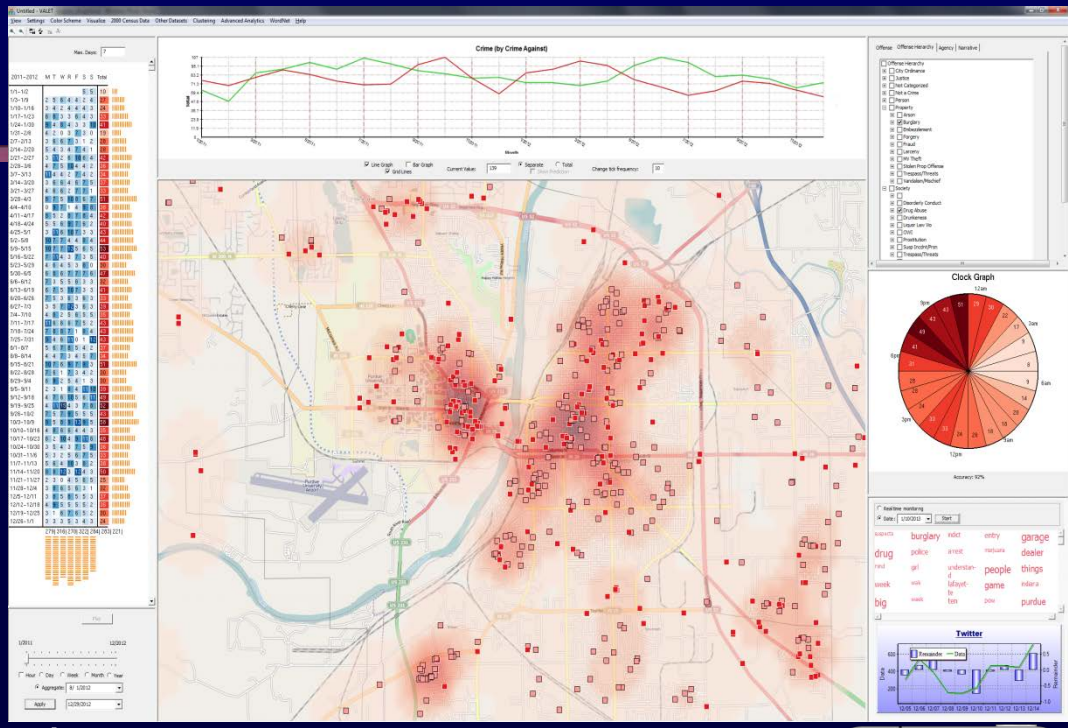
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





VACCINE

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

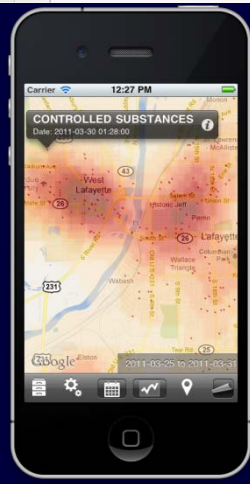
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD



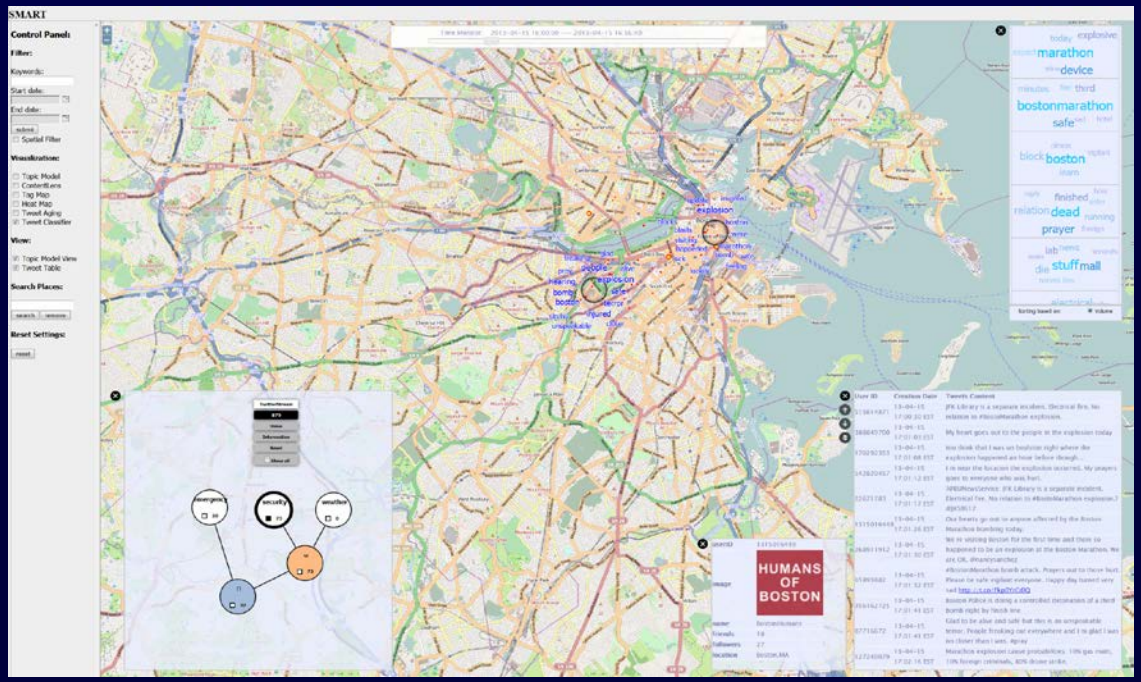
VACCINE

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

May 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

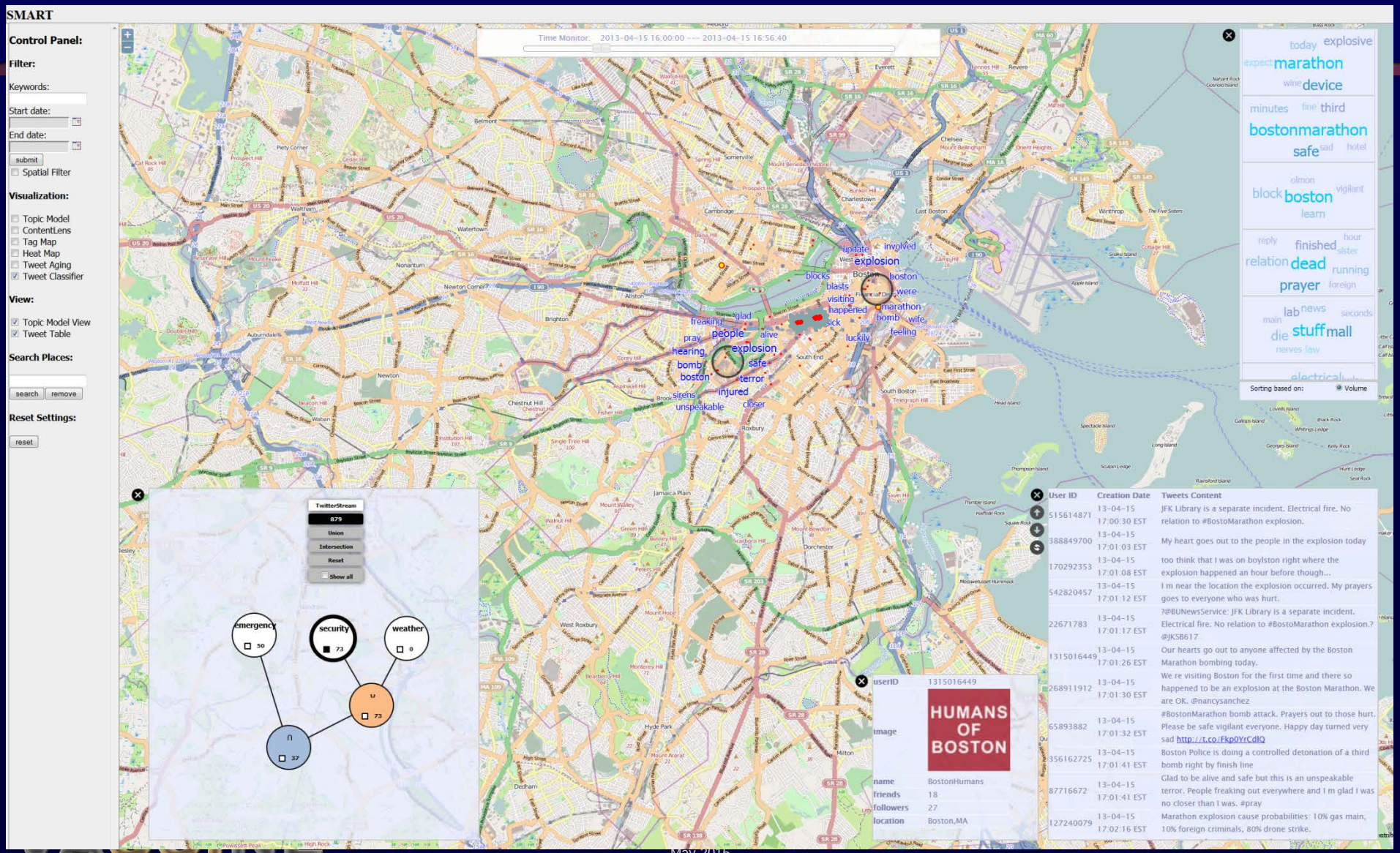
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

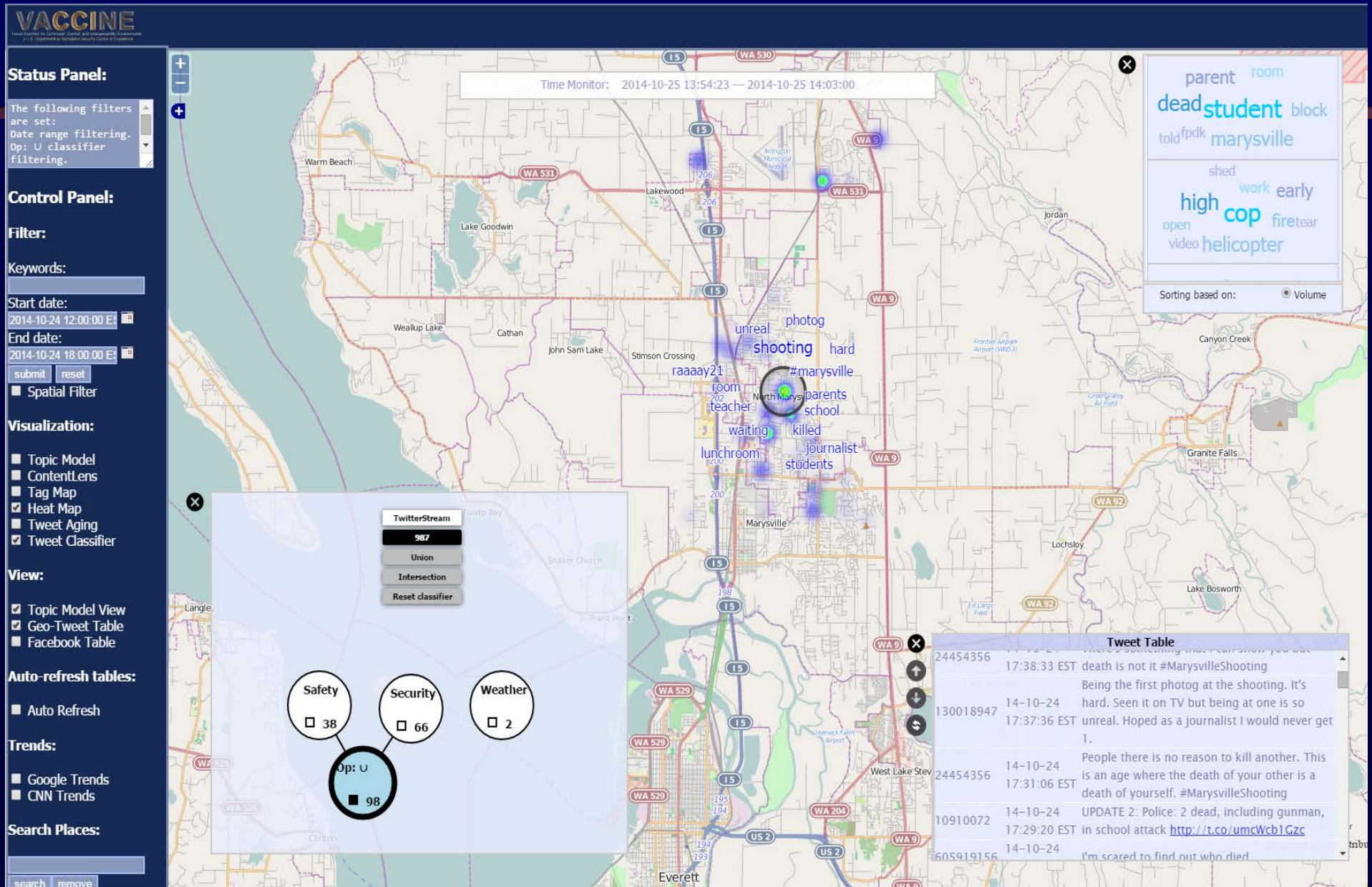


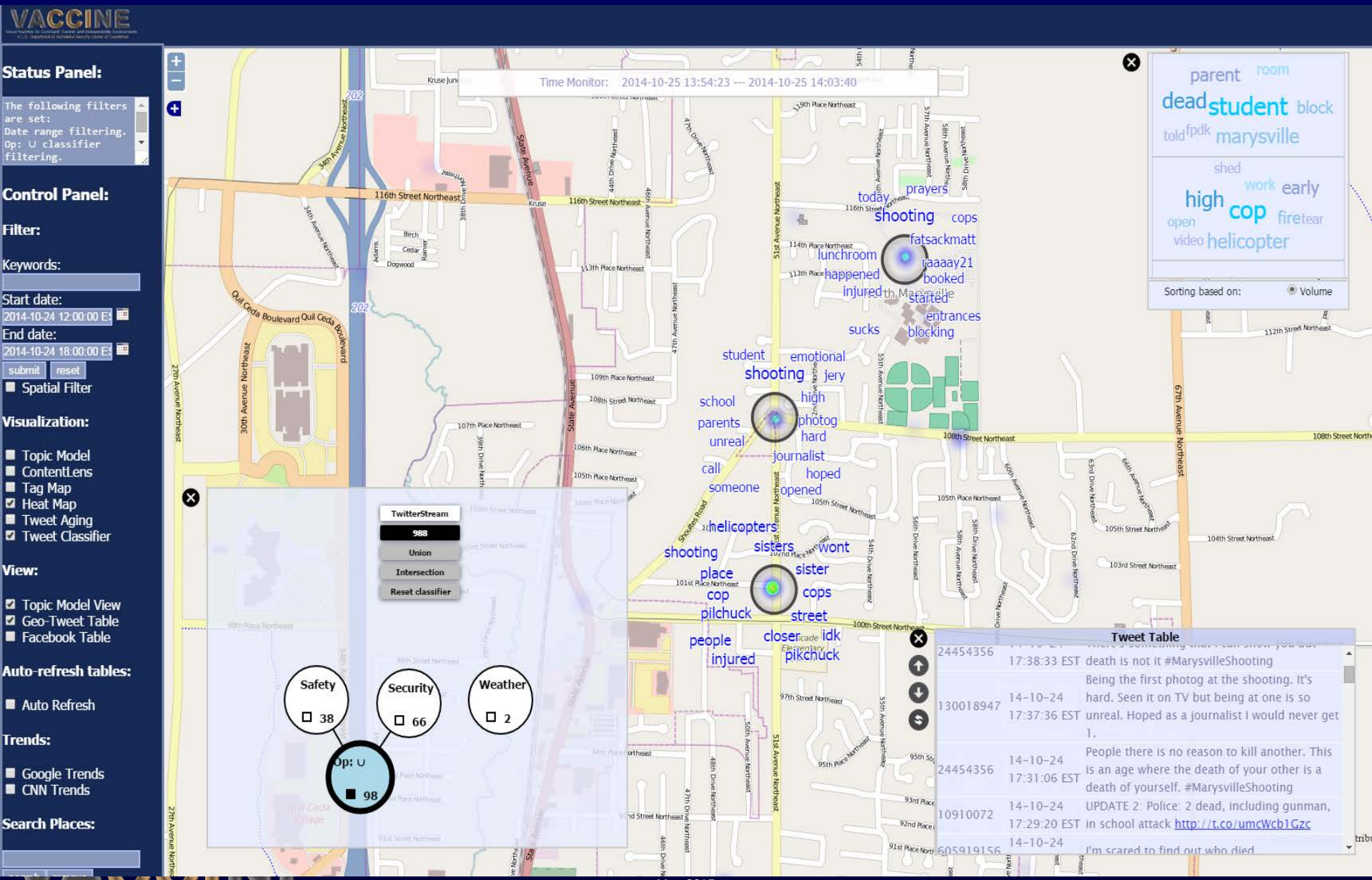
VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24 17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24 17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24 17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died



Status Panel:

The following filters are set:
Single tweet filtering.
Date range filtering.

Control Panel:

Filter:

Keywords:

Start date:
2014-10-24 12:00:00 ES

End date:
2014-10-24 18:00:00 ES

submit reset

- Spatial Filter

Visualization:

- Topic Model
- ContentLens
- Tag Map
- Heat Map
- Tweet Aging
- Tweet Classifier

View:

- Topic Model View
- Geo-Tweet Table
- Facebook Table

Auto-refresh tables:

- Auto Refresh

Trends:

- Google Trends
- CNN Trends

Search Places:

Marysville-Pilchuck High Sch

search remove

Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48

pilchuck tear video cop pray homestuck job couple marysville high school sad block kid told fire shoot

Sorting based on: Volume

Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

Location of the school

North Marysville

TwitterStream
990
Union
Intersection
Reset classifier

Safety 38
Security 66
Weather 2

userID 23845798

image

name joshtrujillo
friends 2146
followers 6674
location Seattle,WA

Tweet Table

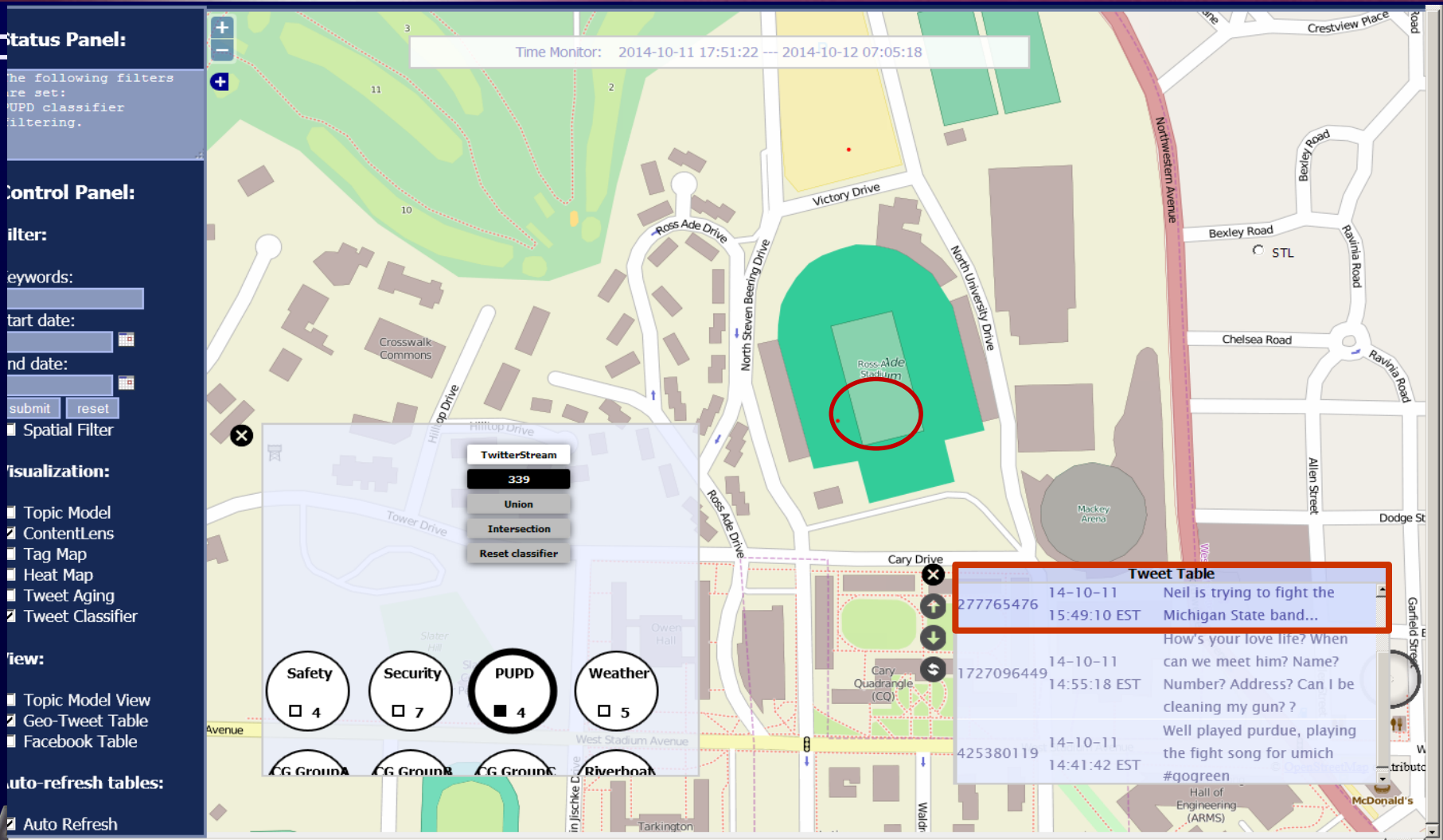
User ID	Creation Date	Tweets Content
23845798	14-10-24 16:47:14 EST	Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

Where Can It Be Helpful?

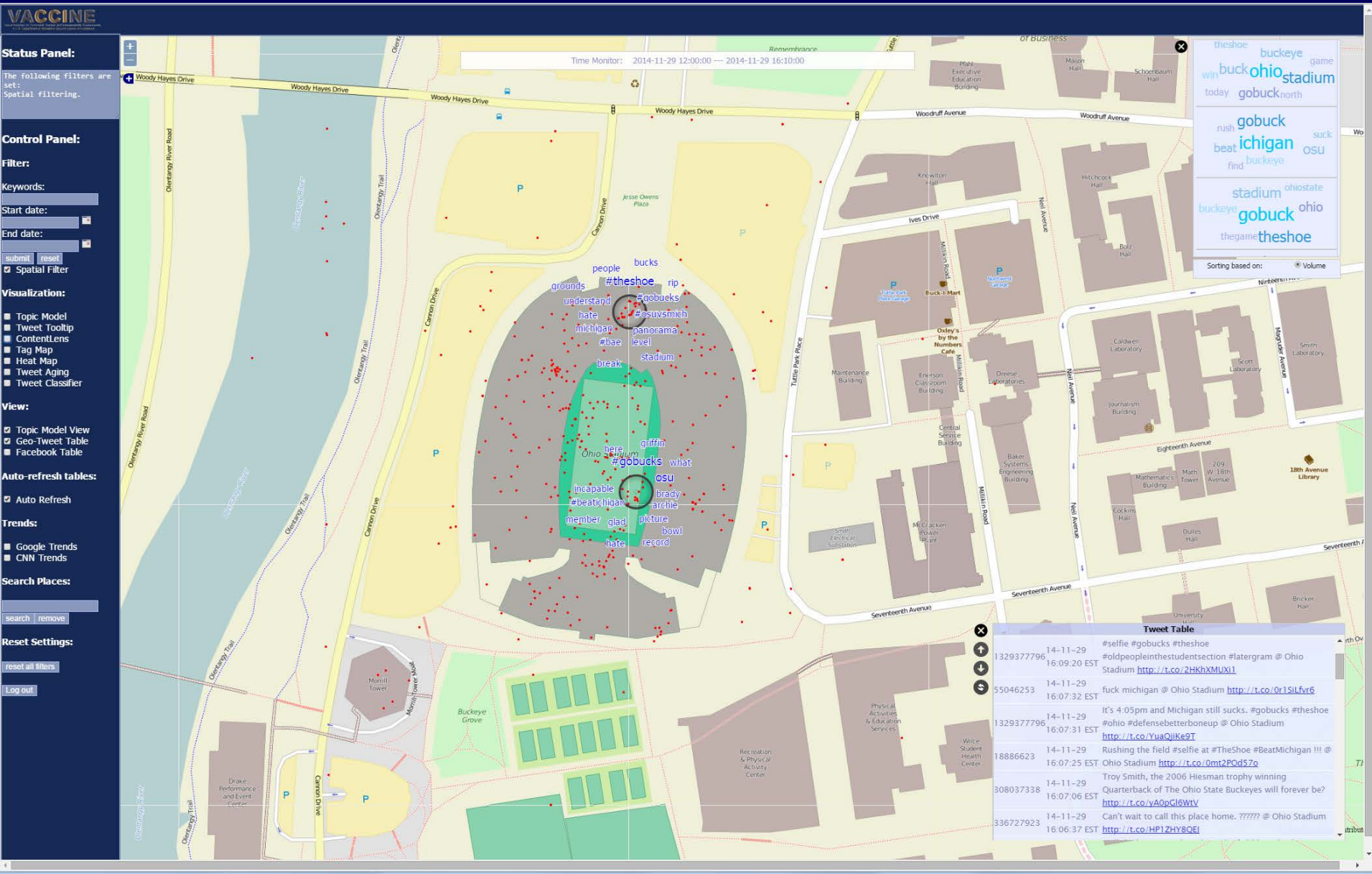
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



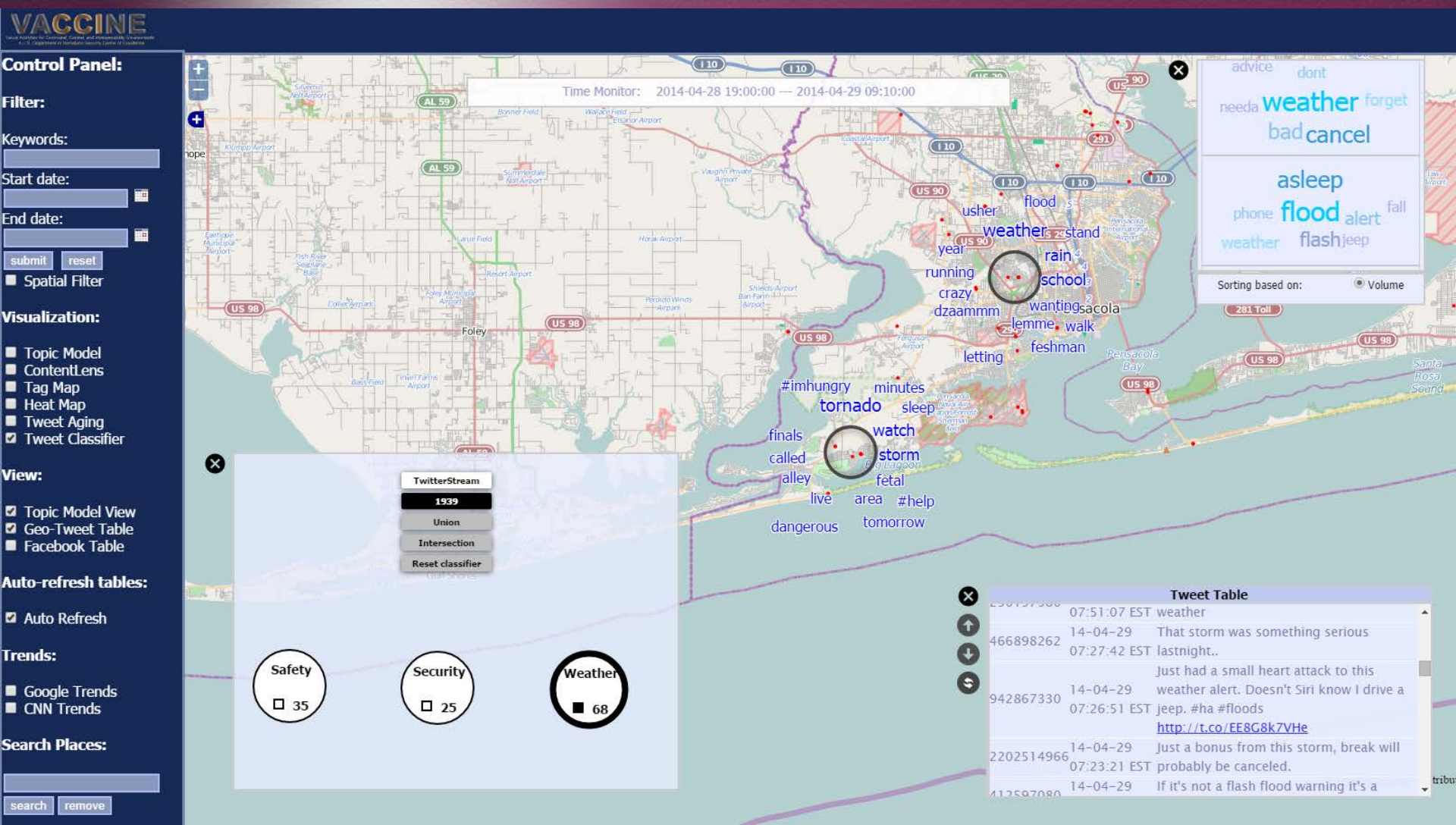
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



Where Can It Be Helpful?

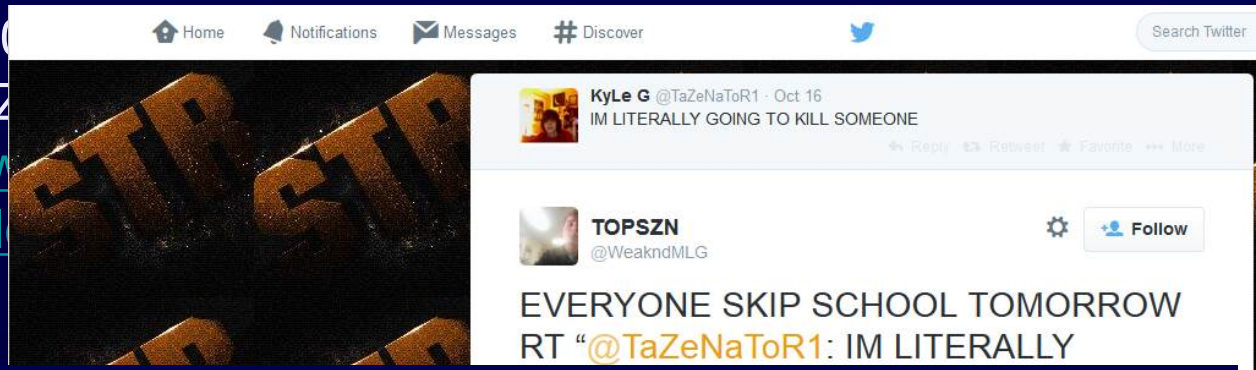


Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE



Wayfast Whitey
@WayFastWhitey44

Follow

pnbarrie

Follow

1 month ago · ARIA Resort & Casino

Nice production at the John Deere event Thursday night for Conexpo. #conexpo #johndeere #rdoequipment #lovemyjob

shepnelson, ottoreynolds, mycranelife and connshawn like this.

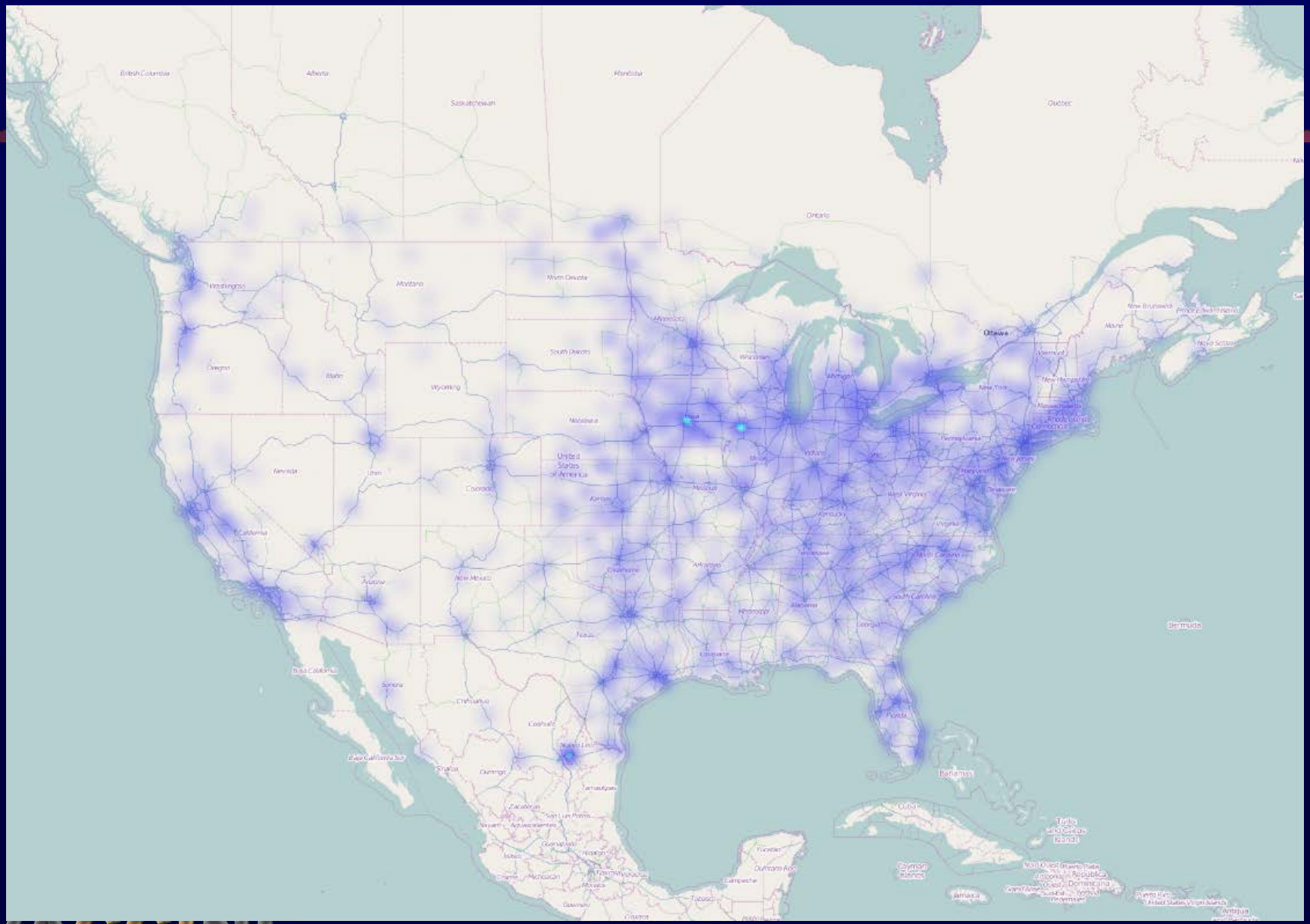
Leave a comment...

VACCINE

May 2

Reply to @WayFastWhitey44

John Deere


British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
Ottawa
Montana
North Dakota
South Dakota
Wyoming
Nebraska
United States of America
Colorado
Kansas
Oklahoma
Texas
New Mexico
Arizona
Nevada
California
Oregon
Idaho
Utah
Minnesota
Wisconsin
Michigan
Illinois
Indiana
Ohio
Kentucky
Tennessee
Mississippi
Alabama
Georgia
Florida
South Carolina
North Carolina
Virginia
West Virginia
Pennsylvania
New York
Vermont
New Hampshire
Maine
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
Louisiana
Arkansas
Missouri
Iowa
Bermuda
Bahamas
Cuba
Jamaica
Cayman Islands
Haiti
Dominican Republic
Puerto Rico
Baja California
Sonora
Chihuahua
Coahuila
Nuevo Leon
Tamaulipas
Durango
Zacatecas
San Luis Potosi
Aguascalientes
Jalisco
Michoacan
Guanajuato
Queretaro
Hidalgo
Veracruz
Tabasco
Campeche
Yucatan
Quintana Roo

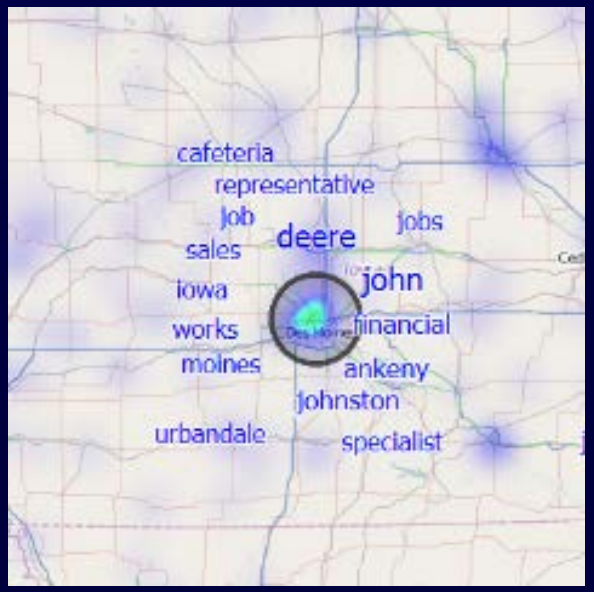
Ten Dominant Topics Related to John Deere



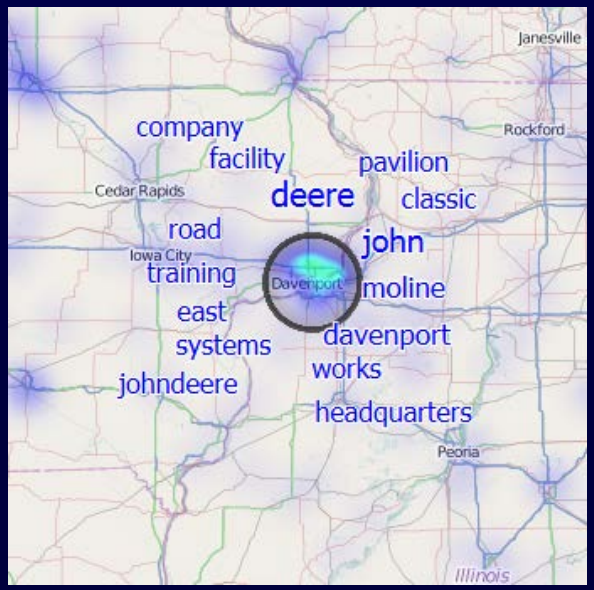


Keywords(Contentlens) in the 3 Hotspots

Des Moines, Iowa



Davenport, Iowa



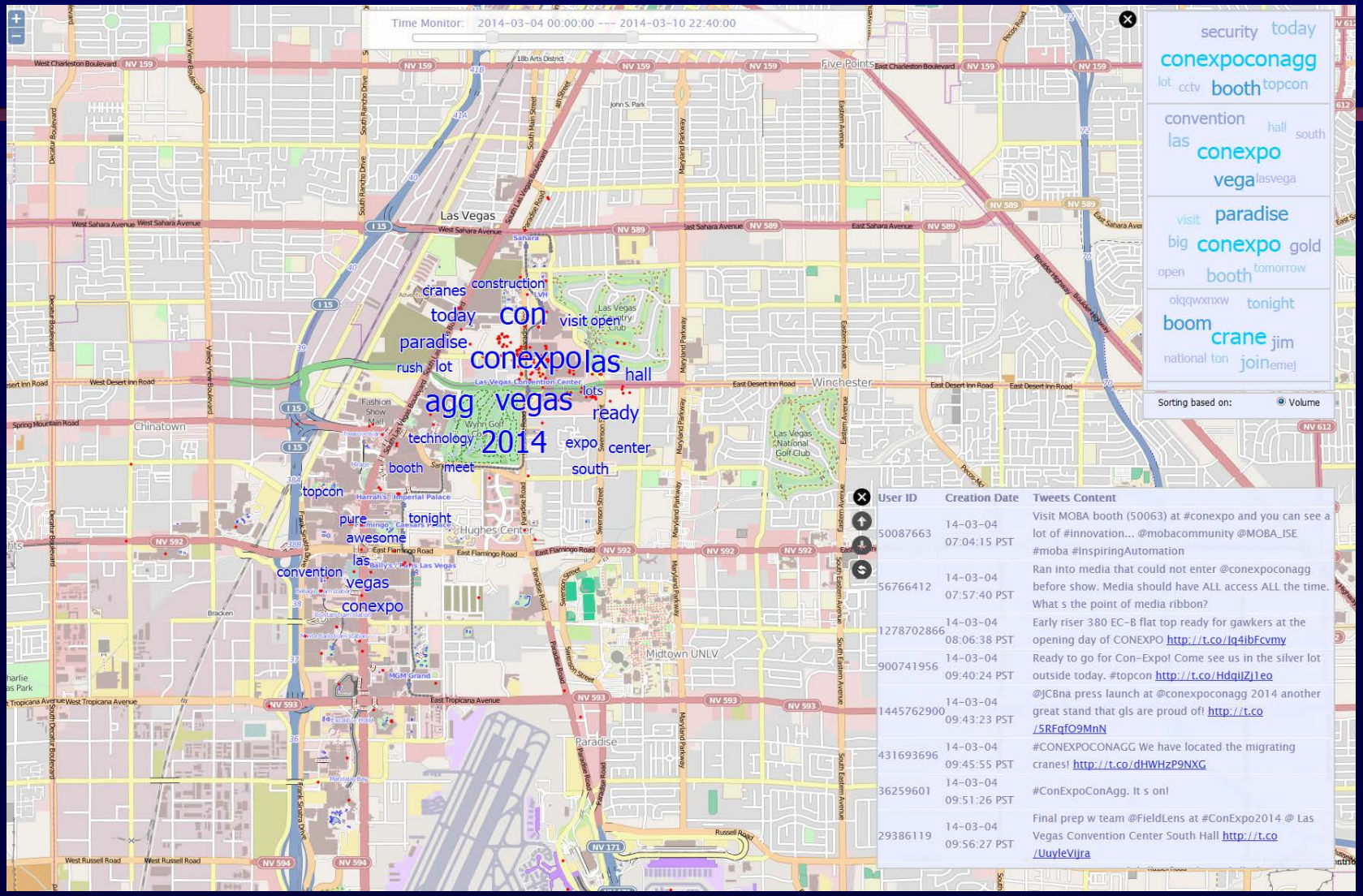
Monterrey, Mexico



Hashtags Related to John Deere

- #johndeere, #nothingrunslikeadeere
- #tractor, #tractortuesday
- #farm, #farming, #farmlife
- #jobs, #work, #harvest
- #green, #country, #summer
- #diesel, # dieselgenerator, #jasonaldean
- #genset, #wayouthere

Con Expo


Time Monitor: 2014-03-04 00:00:00 --- 2014-03-10 22:40:00
cranes
construction
today
con
visit
open
paradise
conexpo
las
rush
lot
hall
lots
agg
vegas
ready
technology
2014
expo
center
booth
meet
south
topcon
pure
tonight
awesome
las
convention
vegas
conexpo
security
today
conexpoconagg
lot
cctv
booth
topcon
convention
hall
south
las
conexpo
vega
lasvega
visit
paradise
big
conexpo
gold
open
booth
tomorrow
oiqqwxnxw
tonight
boom
crane
jim
national
ton
join
emej
Sorting based on:
Volume
User ID
Creation Date
Tweets Content
50087663
14-03-04 07:04:15 PST
Visit MOBA booth (50063) at #conexpo and you can see a lot of #innovation... @mobacommunity @MOBA_ISE #moba #InspiringAutomation
56766412
14-03-04 07:57:40 PST
Ran into media that could not enter @conexpoconagg before show. Media should have ALL access ALL the time. What s the point of media ribbon?
1278702866
14-03-04 08:06:38 PST
Early riser 380 EC-B flat top ready for gawkers at the opening day of CONEXPO http://t.co/Iq4ibFcvmy
900741956
14-03-04 09:40:24 PST
Ready to go for Con-Expo! Come see us in the silver lot outside today. #topcon http://t.co/HdqiIZj1eo
1445762900
14-03-04 09:43:23 PST
@JCBna press launch at @conexpoconagg 2014 another great stand that gls are proud of! http://t.co/5RFqfO9MnN
431693696
14-03-04 09:45:55 PST
#CONEXPOCONAGG We have located the migrating cranes! http://t.co/dHWHzP9NXG
36259601
14-03-04 09:51:26 PST
#ConExpoConAgg. It s on!
29386119
14-03-04 09:56:27 PST
Final prep w team @FieldLens at #ConExpo2014 @ Las Vegas Convention Center South Hall http://t.co/UuyleVijra

Tweets related to John Deere skidder

wave3news @wave3news · Apr 4
RT to help ISP: Reward offered for stolen **John Deere** log **skidder**, bit.ly/QMzLgr #wave3news pic.twitter.com/XqxOZ7lggn



Buster Beans @drifhfbfbfj · 22h
Just got a new **John Deere skidder** today pic.twitter.com/c8WiMnXVfV

Hide photo Reply Retweet Favorite More



FAVORITE 1

7:46 PM - 26 Apr 2014 · De

Retweeted 53 times
John Deere @JohnDeere · Apr 24
If a lion is the king of the jungle, a **John Deere Skidder** is the king of the forest. pic.twitter.com/afZ76bXvVT



Expand Reply Retweet

Service Manuals @EbookManuals · Apr 4
#download **John Deere** 848G **Skidder** Technical Manual Download TM1 dlvr.it/5KChgF @EbookManuals

Expand Reply Retweet Favorite

J Lowe @jennyfaithlowe · Apr 22
Firestone 16 ply **Skidder** Tires on **john deere** wheels 7/8ths holes #ebaymobile bit.ly/1nn9l24

Expand Reply Retweet Favorite More

Want Ad Digest @Wantaddigest · Apr 16
John Deere 540 B Cable **Skidder** LOWVILLE NY: 23x/x26 tire ring chains diamond chains, runs very good, needs l... bit.ly/1ePVzwM

Expand Reply Retweet Favorite

Burge Mcmanus @manualdownloads · Apr 13
#download **John Deere** 640D **Skidder**, 648D Operation and Tests Manual dlvr.it/5NyshS

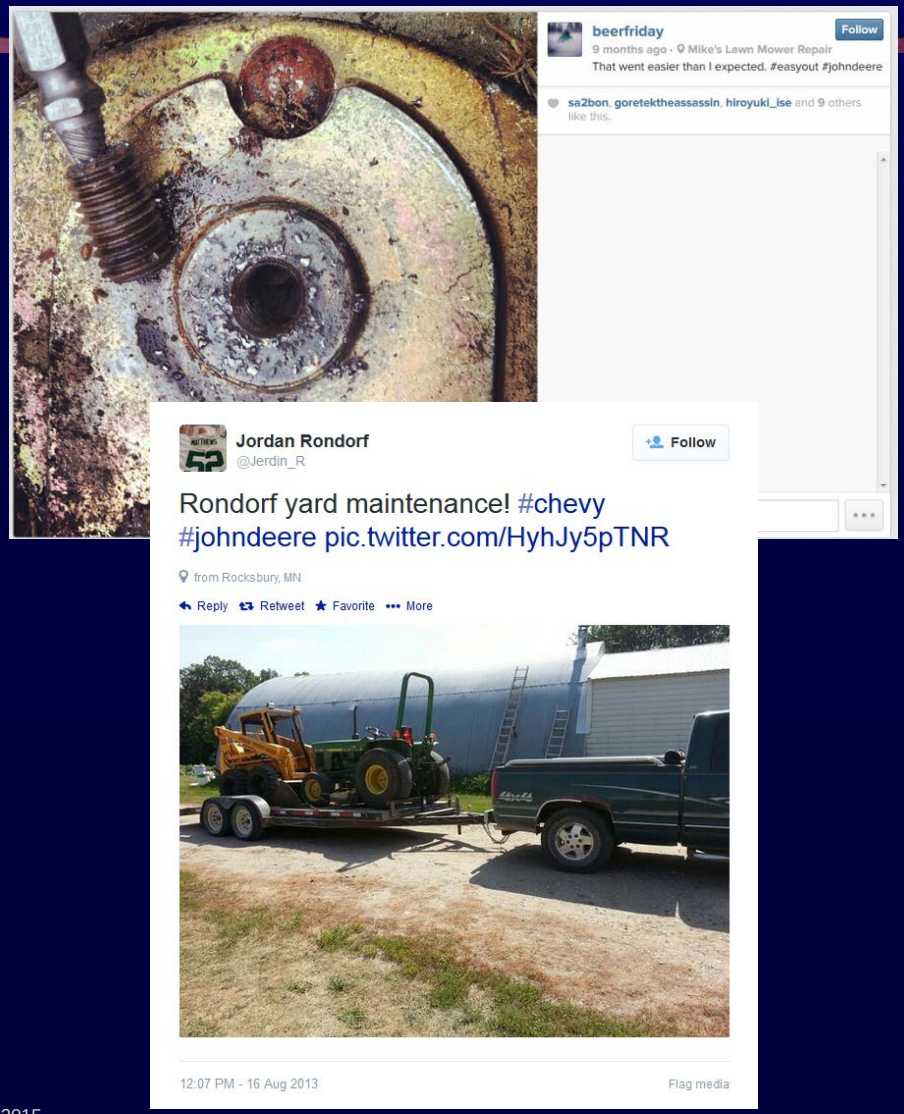
Collapse Reply Retweet Favorite More

9:15 PM - 13 Apr 2014 · Details

VACCINE

Tweets related to ‘maintain, maintenance, repair’

User ID	Creation Date	Tweets Content
933628327	13-06-09 14:27:17 EST	And, for your viewing pleasure, a nicely maintained 1930s John Deere. http://t.co/Zavel5QnDI
50371494	13-06-07 10:13:50 EST	The Class of 2013 presents its class gift at commencement – a John Deere Gator for the Maintenance Department. http://t.co/ju1rcbEuLm
494325743	13-08-16 15:07:21 EST	Rondorf yard maintenance! #chevy #johndeere http://t.co/HyhJy5pTNR
404120195	13-09-03 16:13:01 EST	Just saw a clarkson maintenance worker ride a John Deere lawn tractor to the gas station. No need to bother with a gas can. #Townielife
1038577016	13-07-28 22:58:57 EST	#labor #jobs Industrial Maintenance Technician: position with John Deere Water: Industrial Maintenan... http://t.co/azERmrPMt5
502127999	13-08-07 18:47:22 EST	That went easier than I expected. #easyout #johndeere @ Mike s Lawn Mower Repair http://t.co/8pgiu0Uv7g

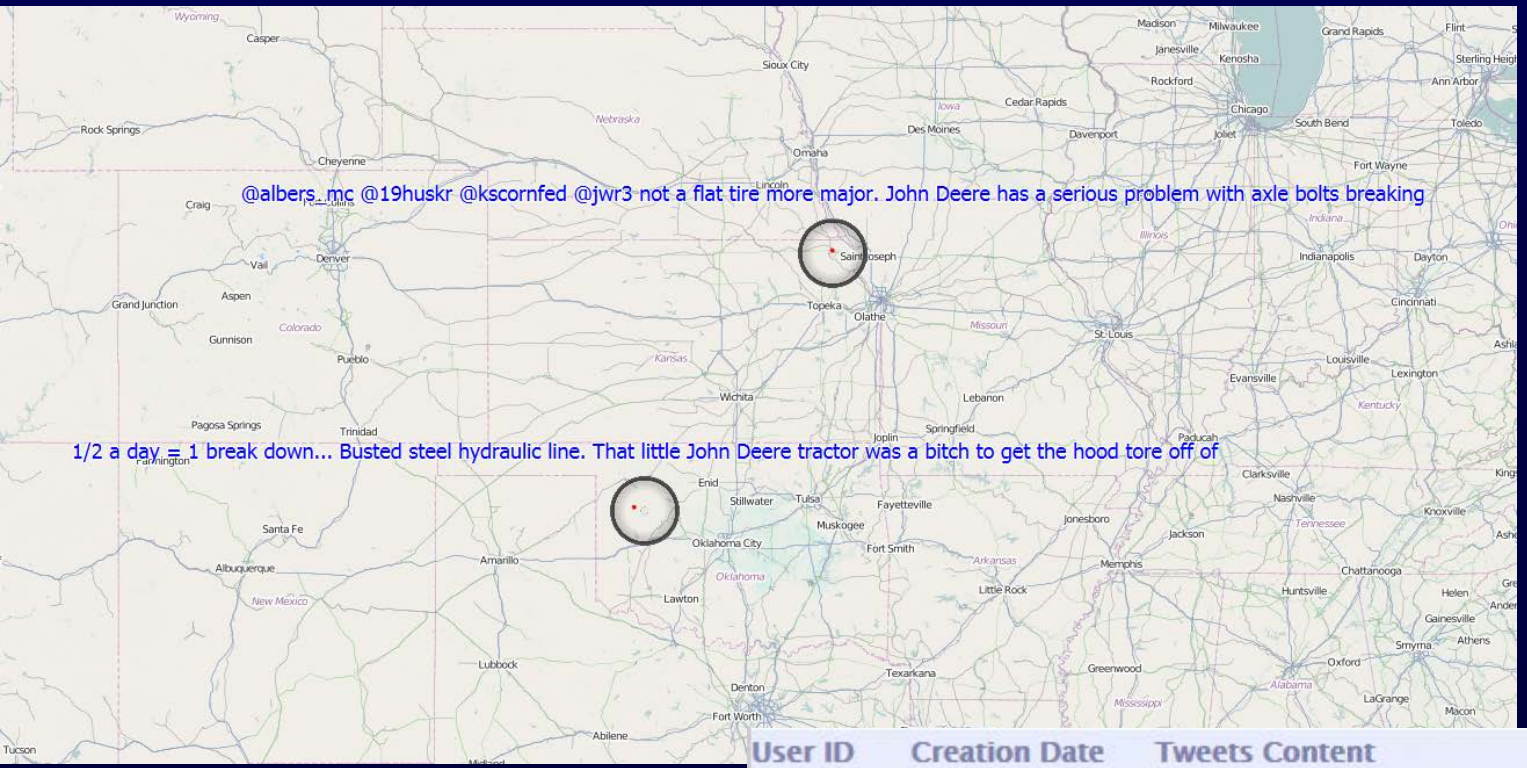


‘Maintain”


Carl
@pzldcharlatan
Follow
And, for your viewing pleasure, a nicely maintained 1930s John Deere.
pic.twitter.com/Zavel5QnDl
from Damascus, PA
Reply Retweet Favorite More


Shady Side Academy
@shady_side
Follow
The Class of 2013 presents its class gift at commencement - a John Deere Gator for the Maintenance Department.
pic.twitter.com/ju1rchEuLm
from Fox Chapel, PA
Reply Retweet Favorite More
RETWEETS
3
FAVORITES
2

Tweets related to ‘break down, shut down, stop working’



User ID	Creation Date	Tweets Content
630487309	13-06-14 15:55:47 EST	1/2 a day = 1 break down... Busted steel hydraulic line. That little John Deere tractor was a bitch to get the hood tore off of
35458727	13-10-18 22:59:17 EST	@albers_mc @19huskr @kscornfed @jwr3 not a flat tire more major. John Deere has a serious problem with axle bolts breaking

VACCINE

For Further Information

www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

VACCINE and Visual Analytics Solutions

David Ebert, Kaethe Beck

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

August 2014

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Justice Institute of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students

"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

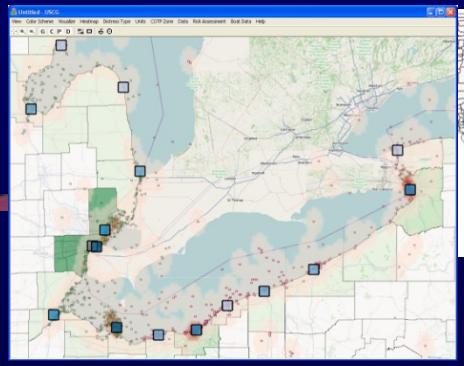
VADM Robert Parker, 2012 MRS Keynote Address

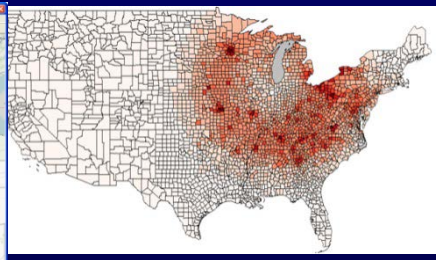
VADM Robert Parker with VACCINE student researchers (cgSARVA, COAST, iOPAR)

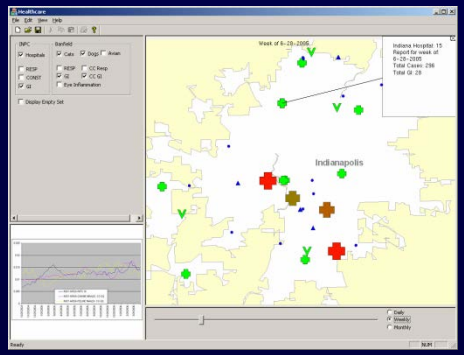


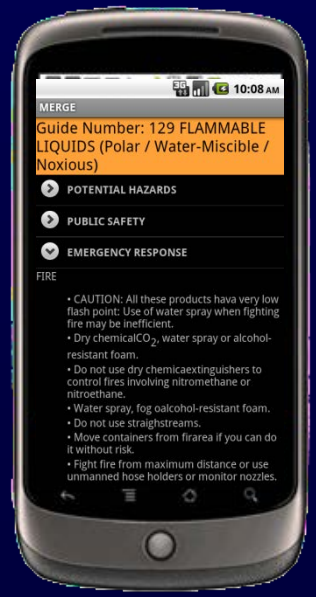
Public Safety & Health

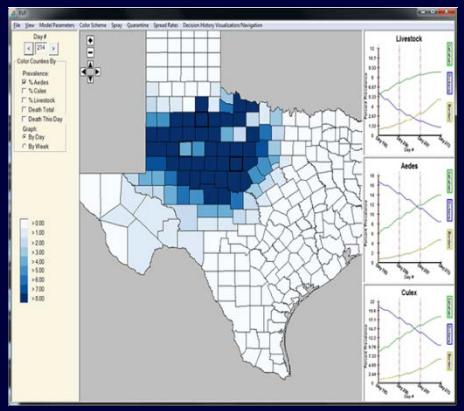
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering









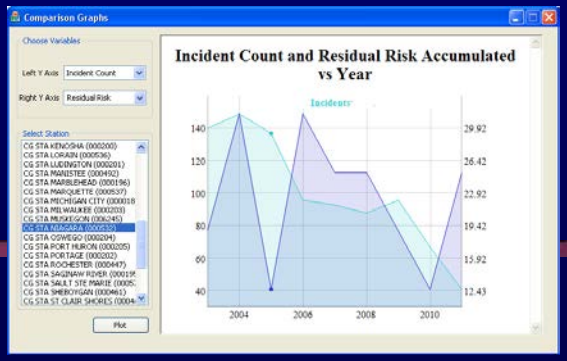


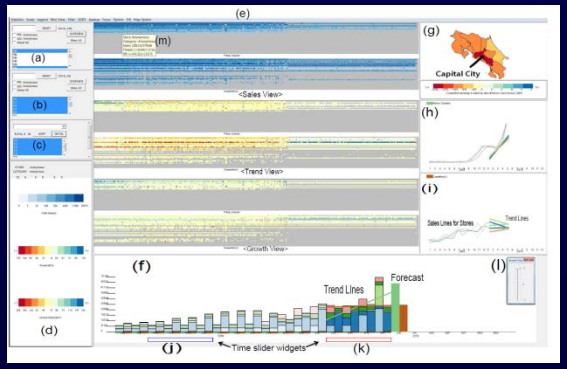


VACCINE

Business Visual Analytics

- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation



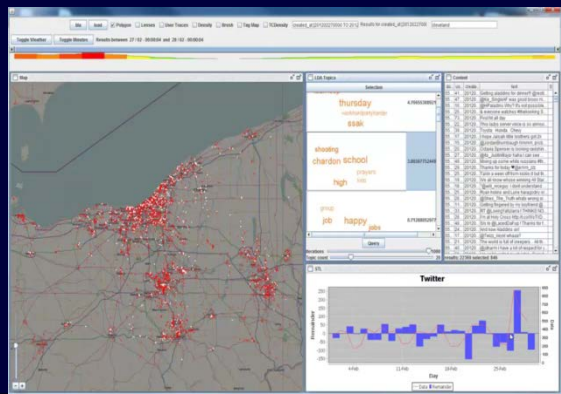




Visual Analytics Uses for Public Safety

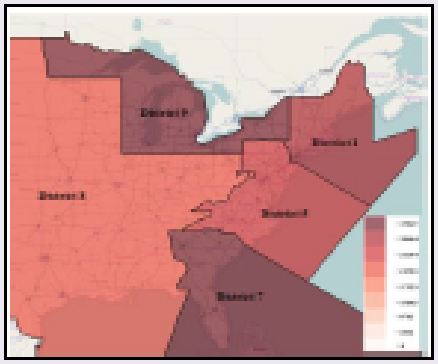


- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation

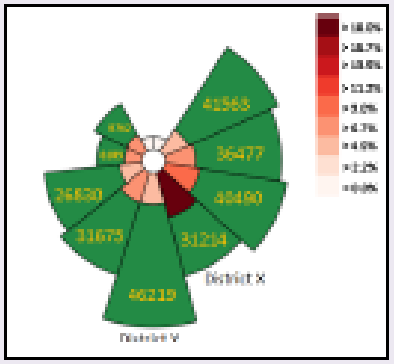


Multi-Visualization, Multivariate VA for Risk-Based Decision Making

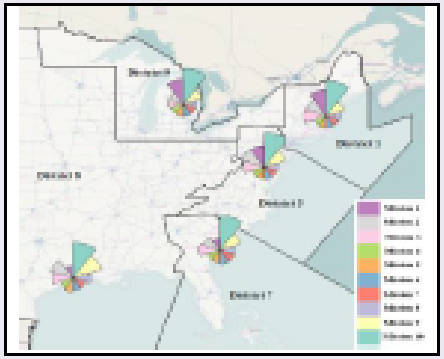
Identify and Analyze Potential Risk



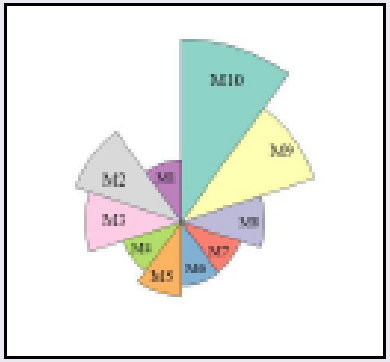
(A) District Choropleth Map for Total Risk across all missions



(C) Risk profiles across all districts and mission areas



(B) Total Risk distribution by mission for any given district.

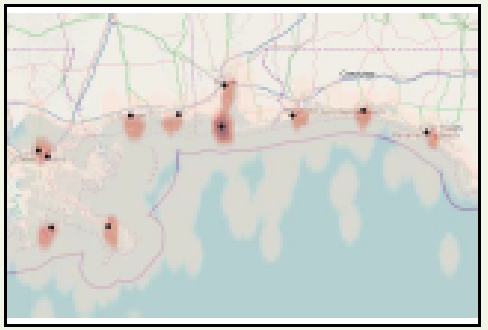


(D) District X Risk Profiles across all mission area

- Define the Context
- Identify potential risks



(E) Total M10 Risk Values distributed among District X's station ADR



(F) KDE heatmap showing Total M10 Risk values

- Assess and Analyze

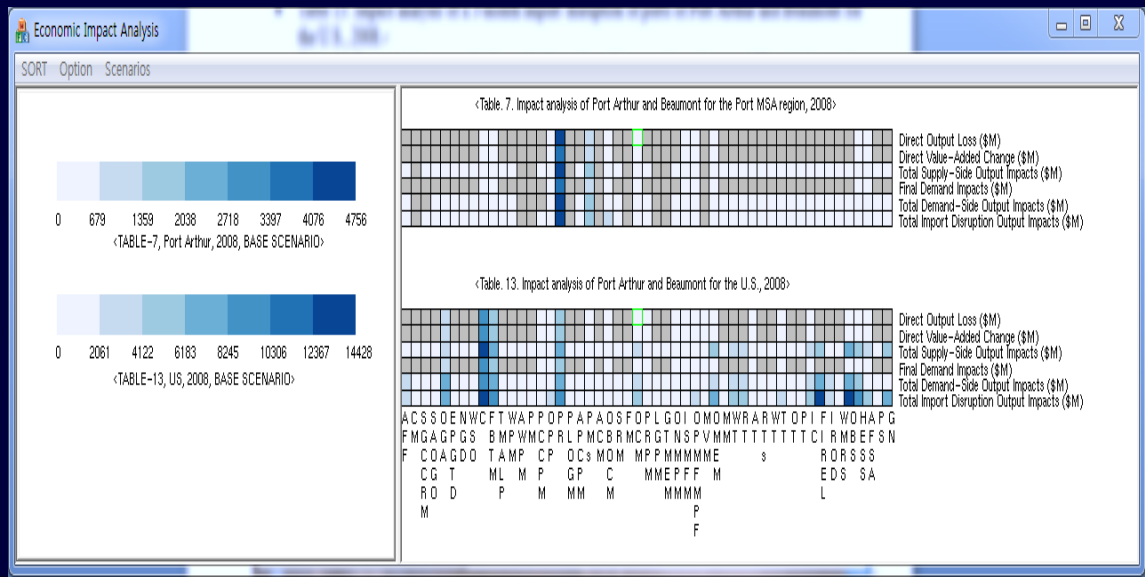
1. Perform spatio-temporal analysis of the risk values.
2. Identify risk priority areas and how they evolve over time and identify the mitigated risk.

USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



VACCINE

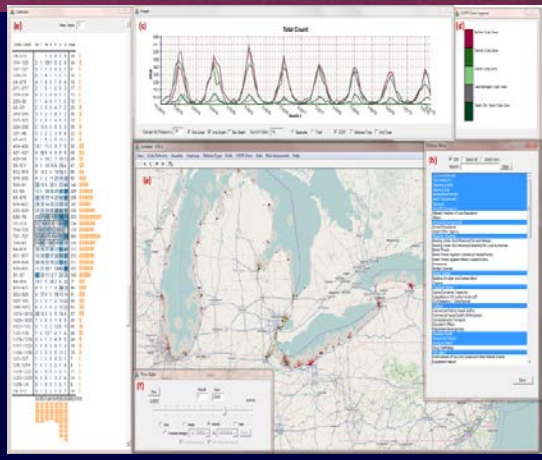
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



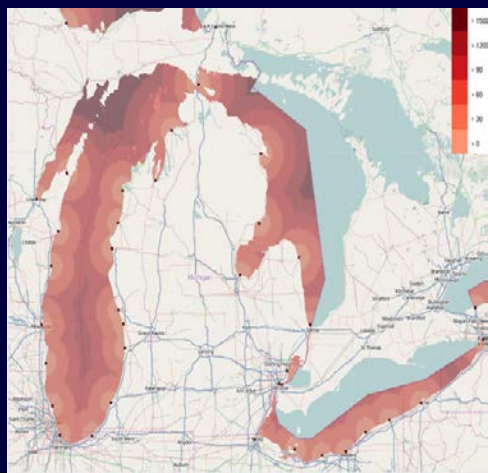
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771



IMPACTS:

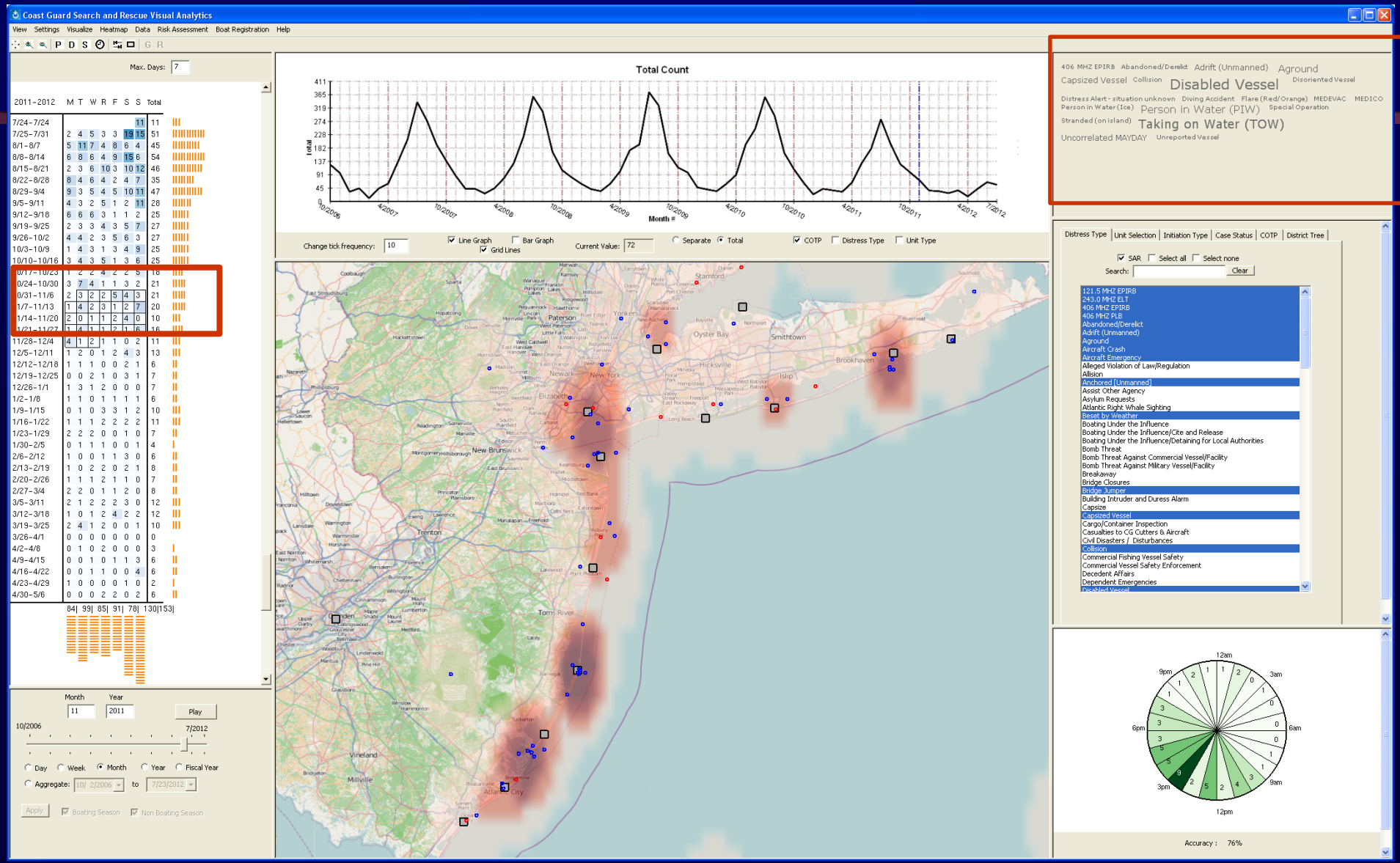
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



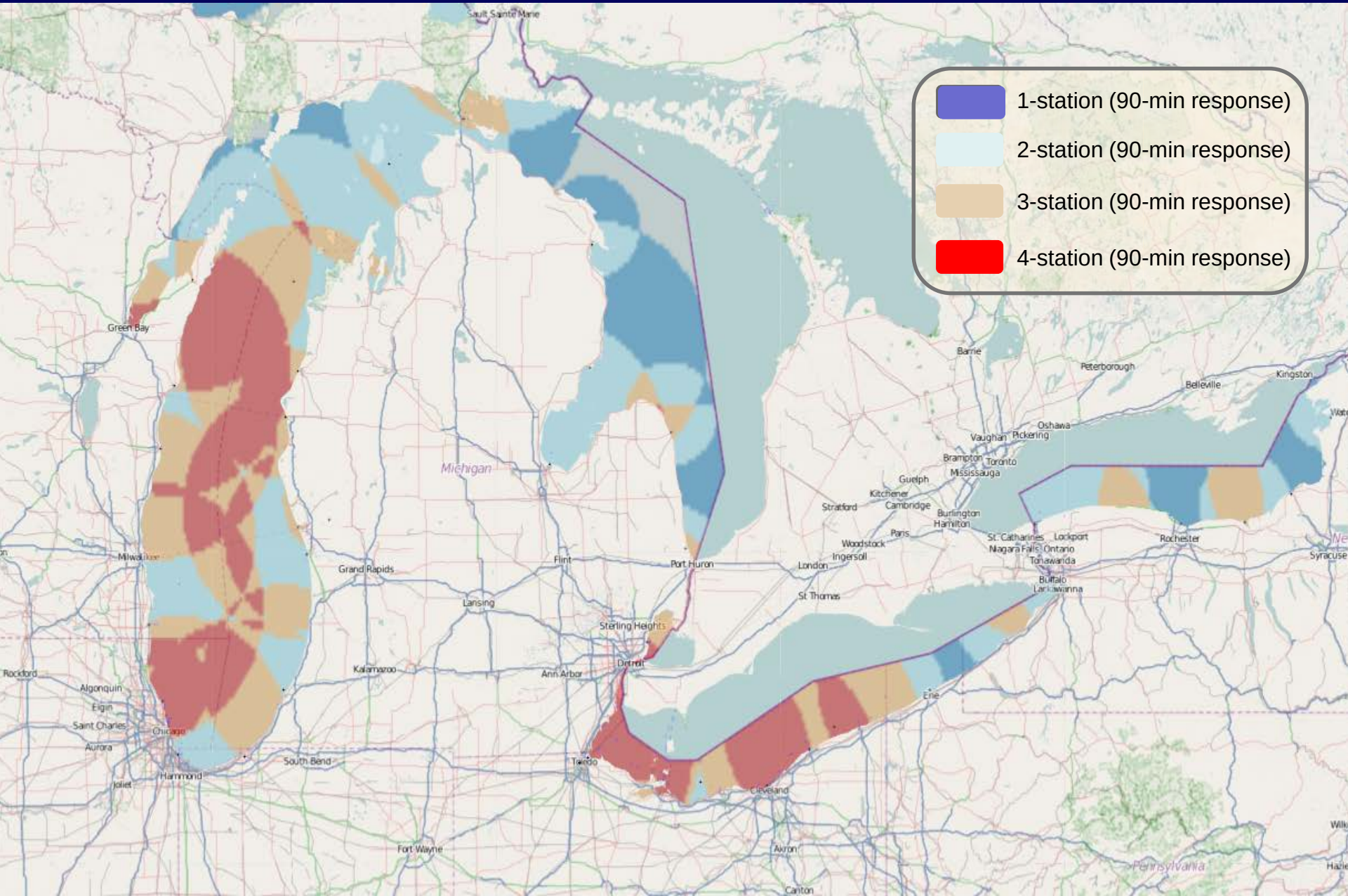
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



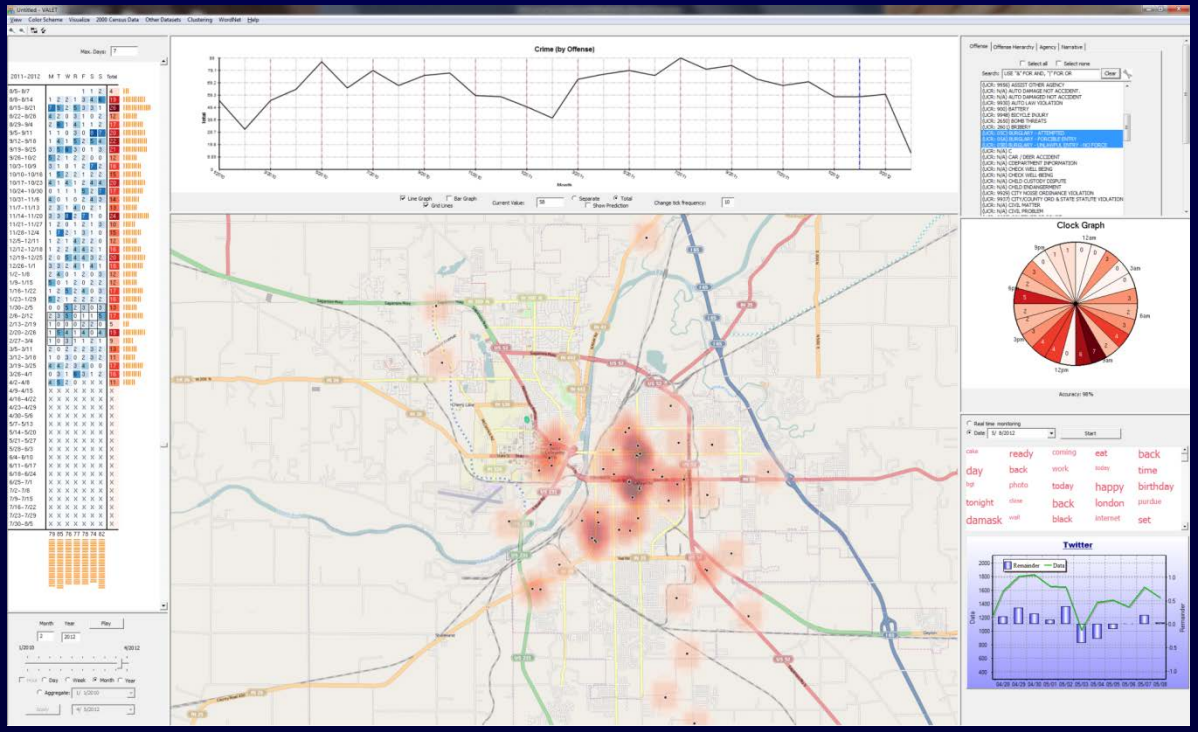
Interactive Operational Performance Assessment Report (iOPAR)

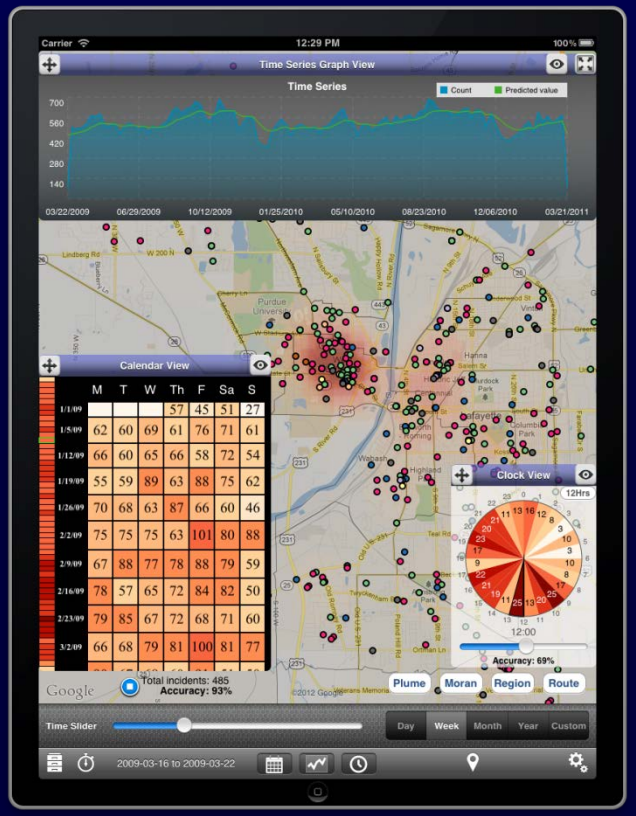
- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels

VACCINE Spatiotemporal VA

- Coast Guard visual analytics
  - Search and rescue
  - Risk-based resource allocation
  - Risk-based operational performance
- Anomaly detection
- Situational awareness
- Predictive policing

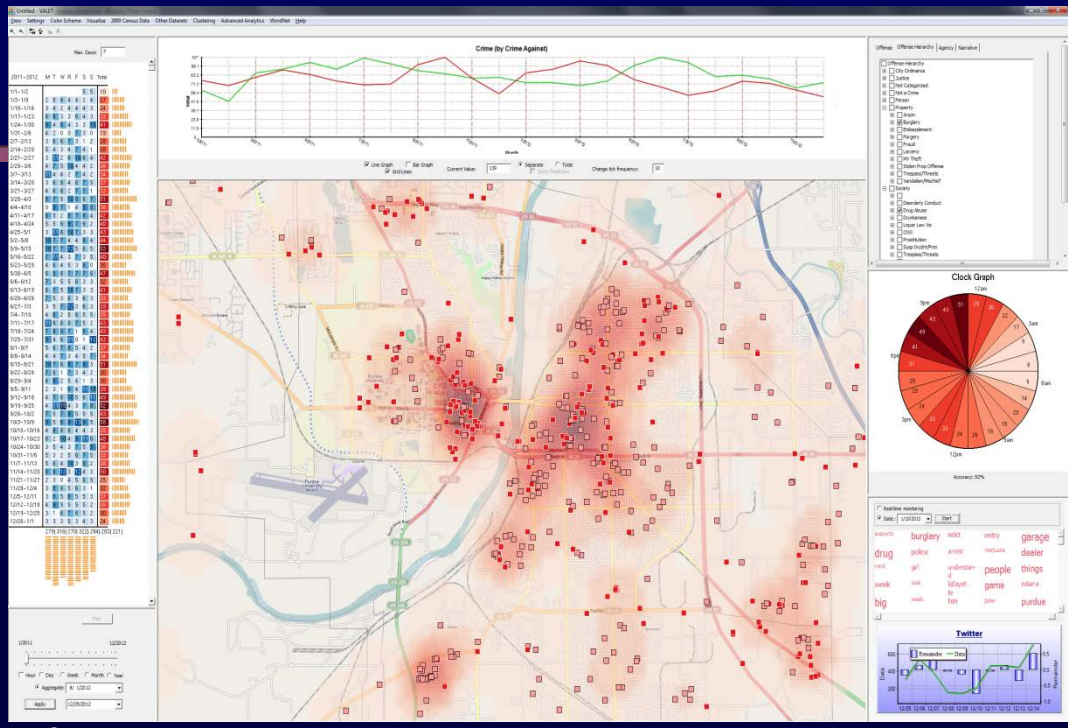
Visual Analytics Law Enforcement Toolkit (i)VALET





VACCINE

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

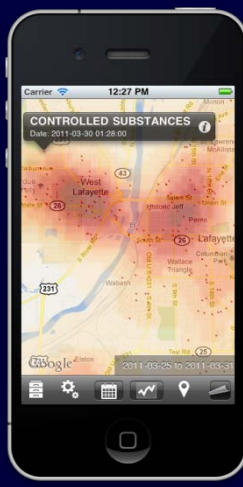
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police

iVALET delivered:

- October 2011: Purdue, WL Police



VACCINE

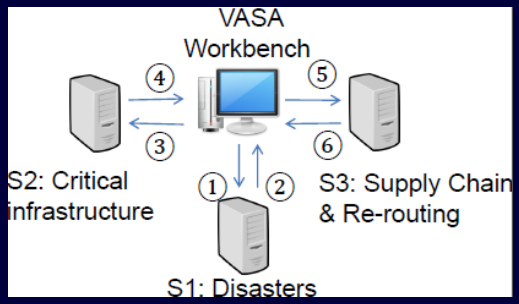
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s): Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

Purdue

UNCC



U Minn

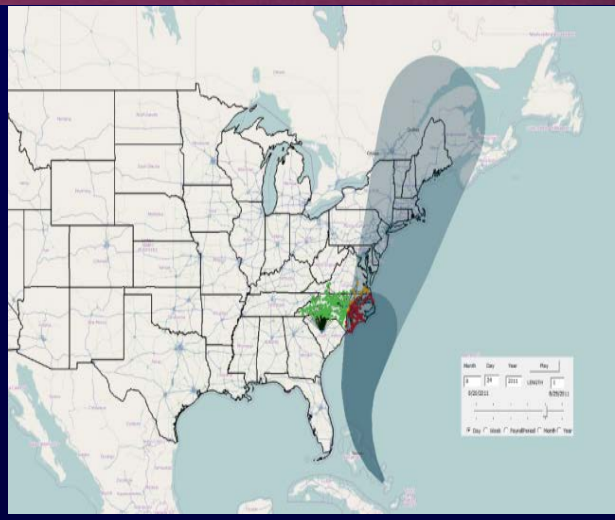
TACC

VACCINE

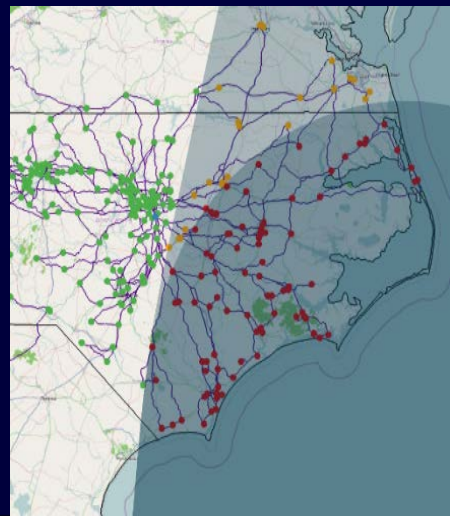
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

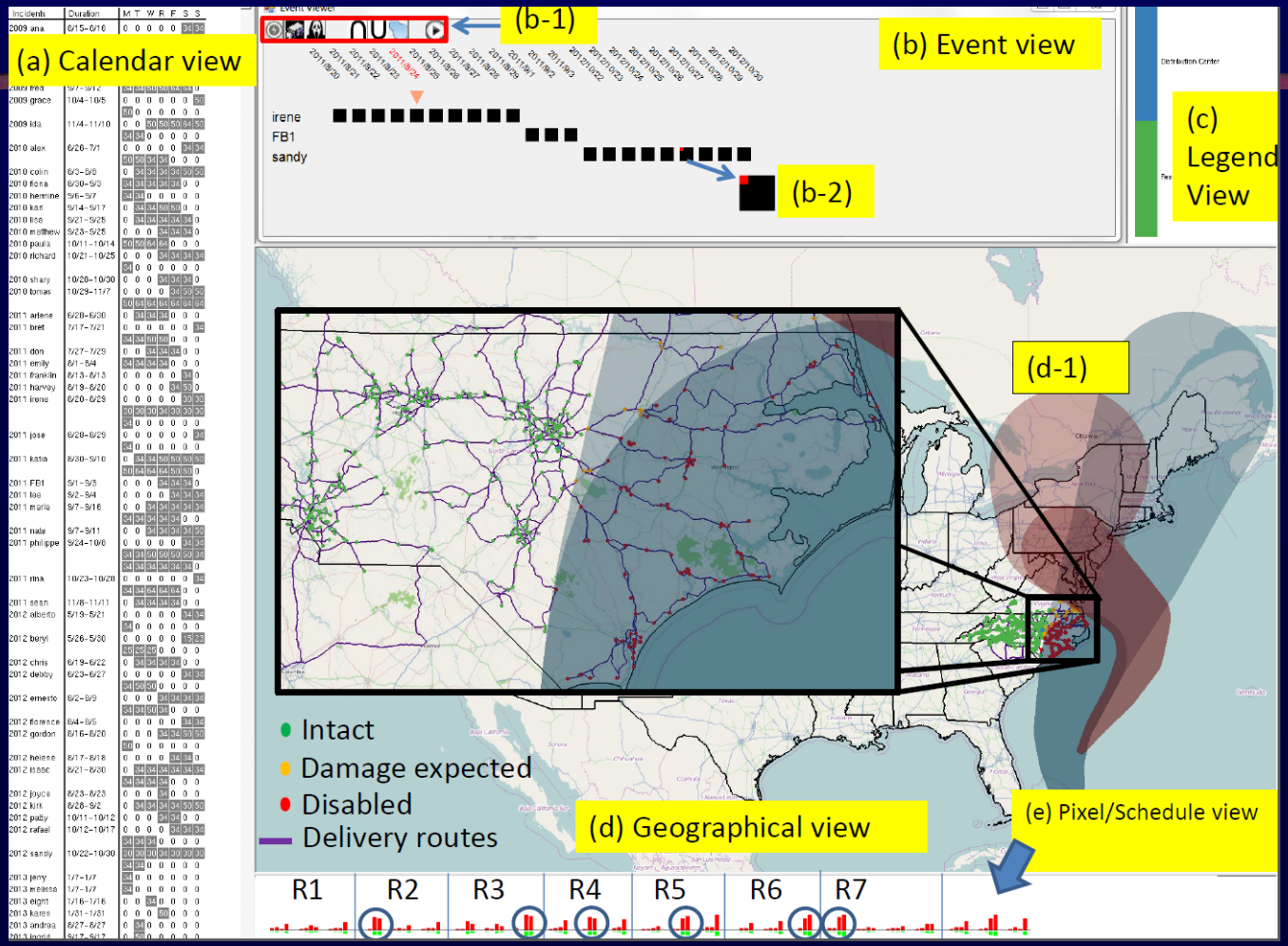
Impacts and Accomplishments:



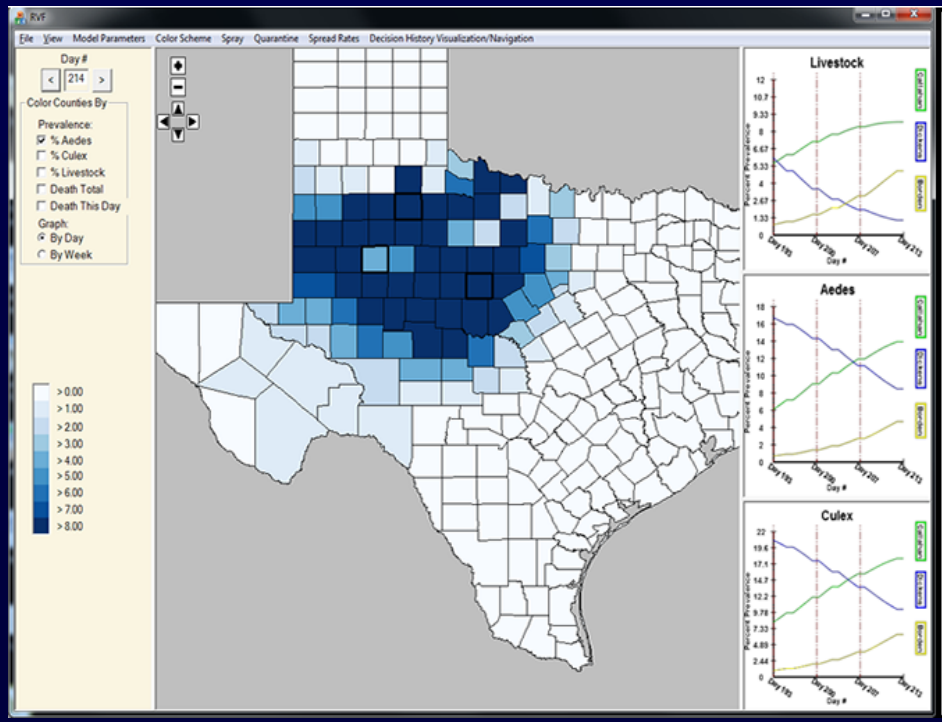
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  combine interactive VA with complex simulation models

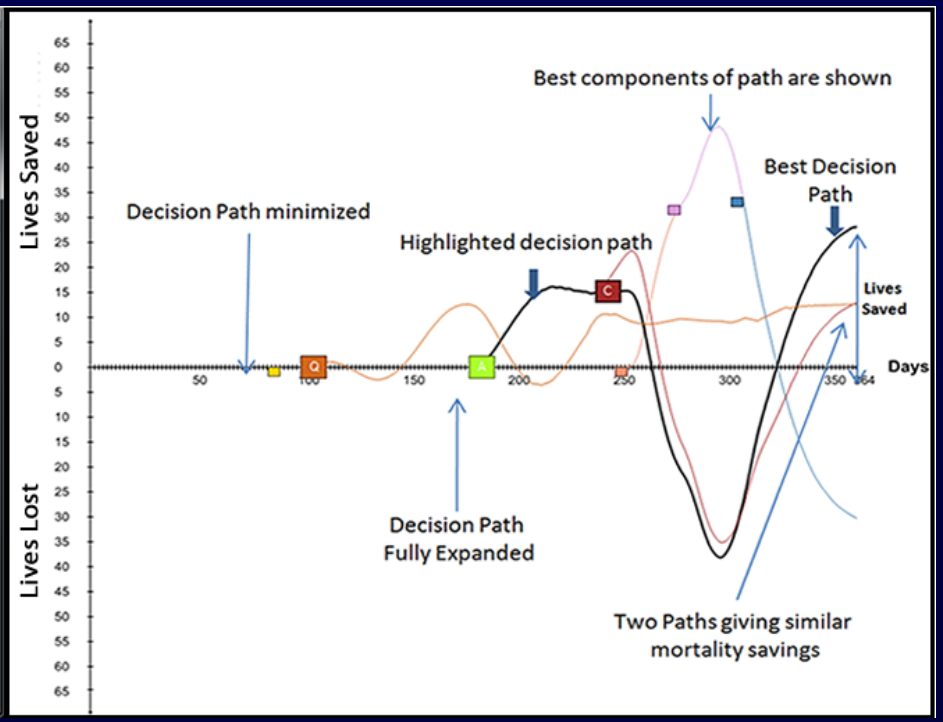


System Interface



Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)





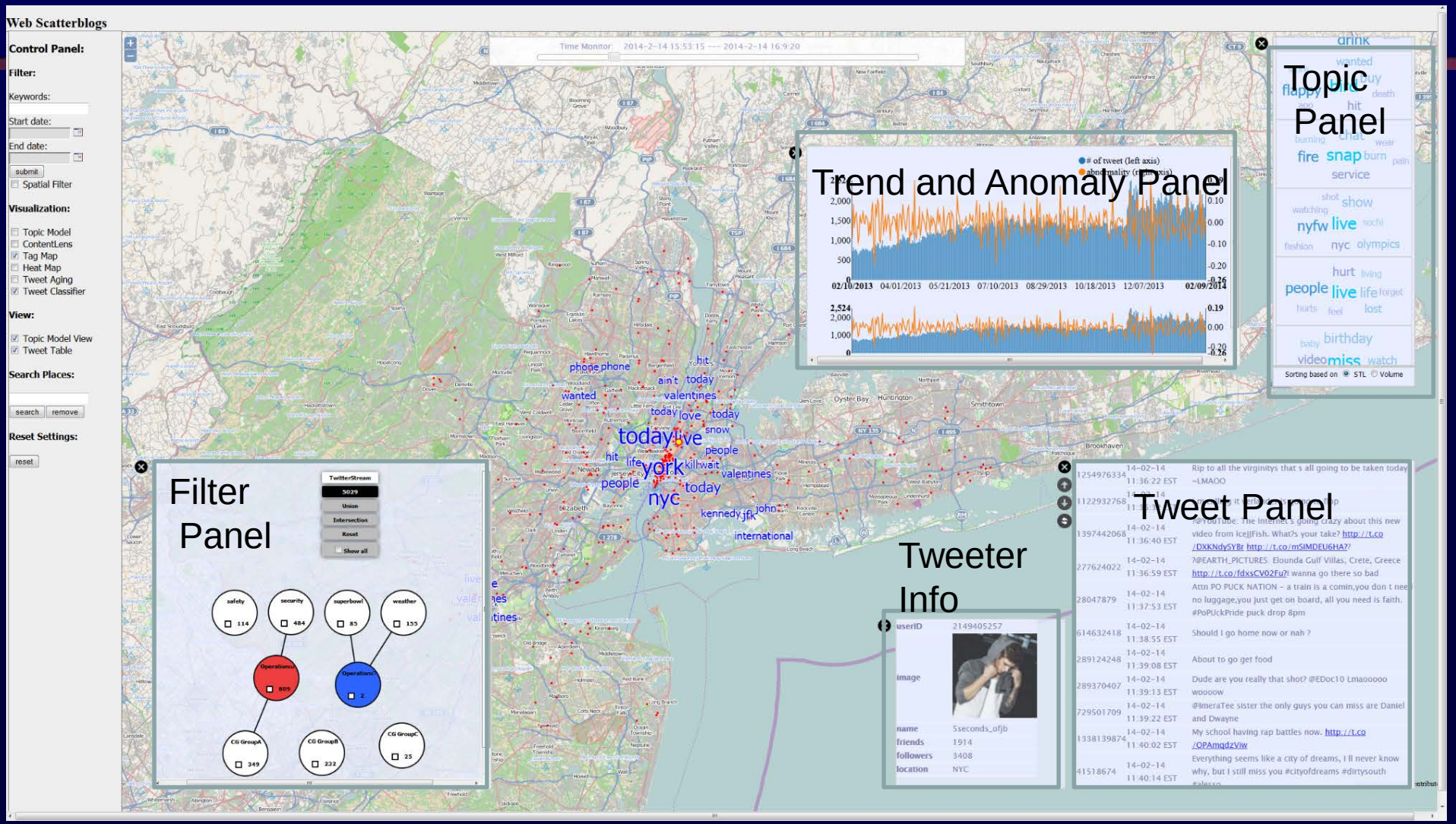
Social Media: Real-time Twitter Visual Analytics

(Purdue, Stuttgart, Penn St.)

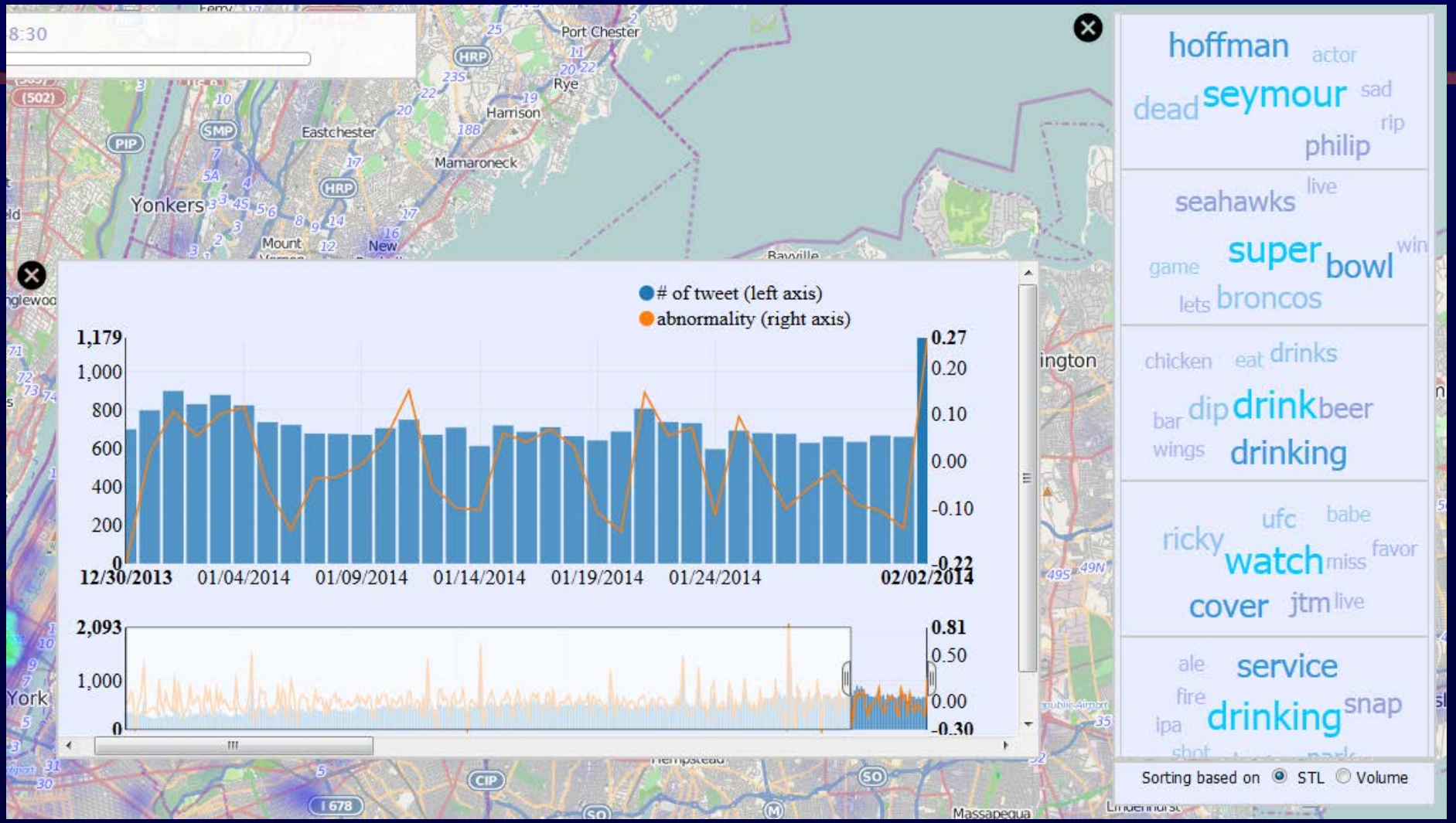
- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud



Social Media Visual Analytics: SMART

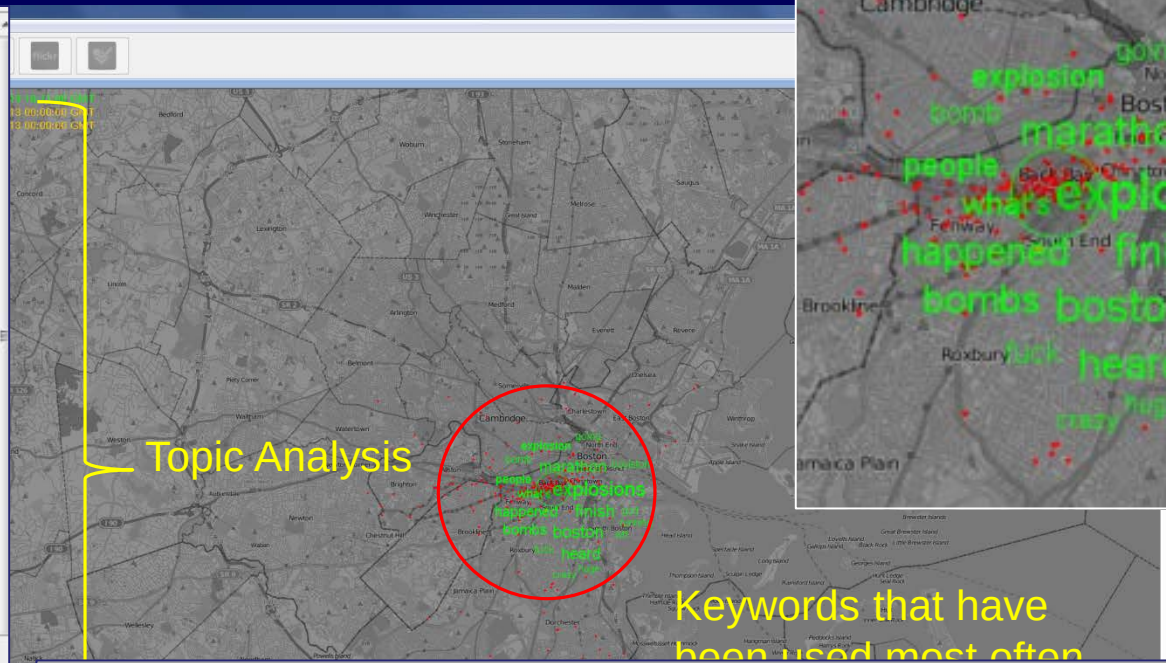


Anomaly Detection



Explosion Area in Boston





Keywords that have been used most often

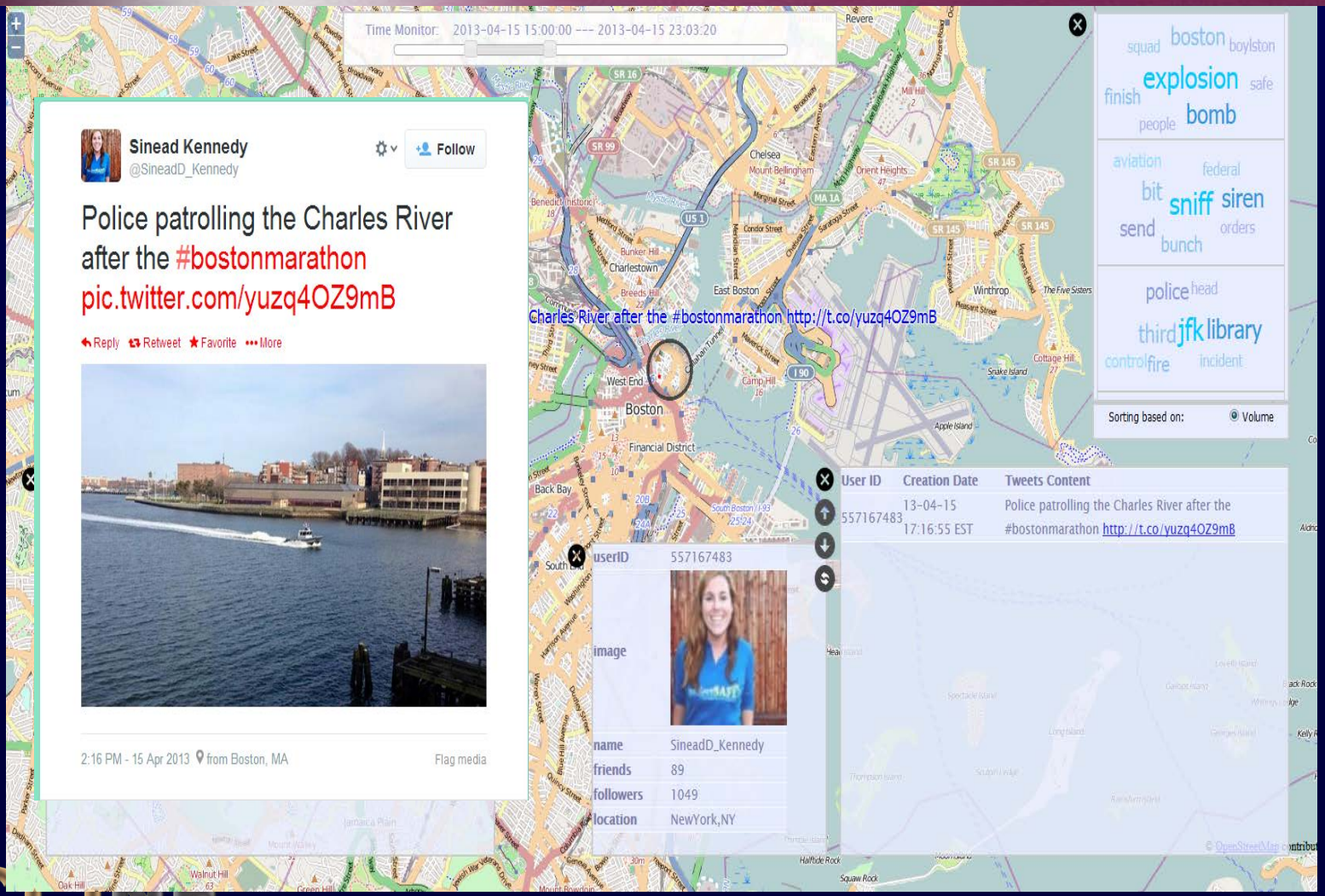
Tweets

Text
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped.
two bombs just went off on boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finsh line this afternoon.
What the FUCK was that
Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalciO I figured 3 people would get the joke
Two explosions just rocked the finish line of the Boston Marathon. Sirens galore. People running in fear. Wonder what happened.
This is crazy i seen that blow up #bostonmarathon
can someone tell me what that explosion was!? #boston #bostonmarathon

First Response (Tweet & Picture)



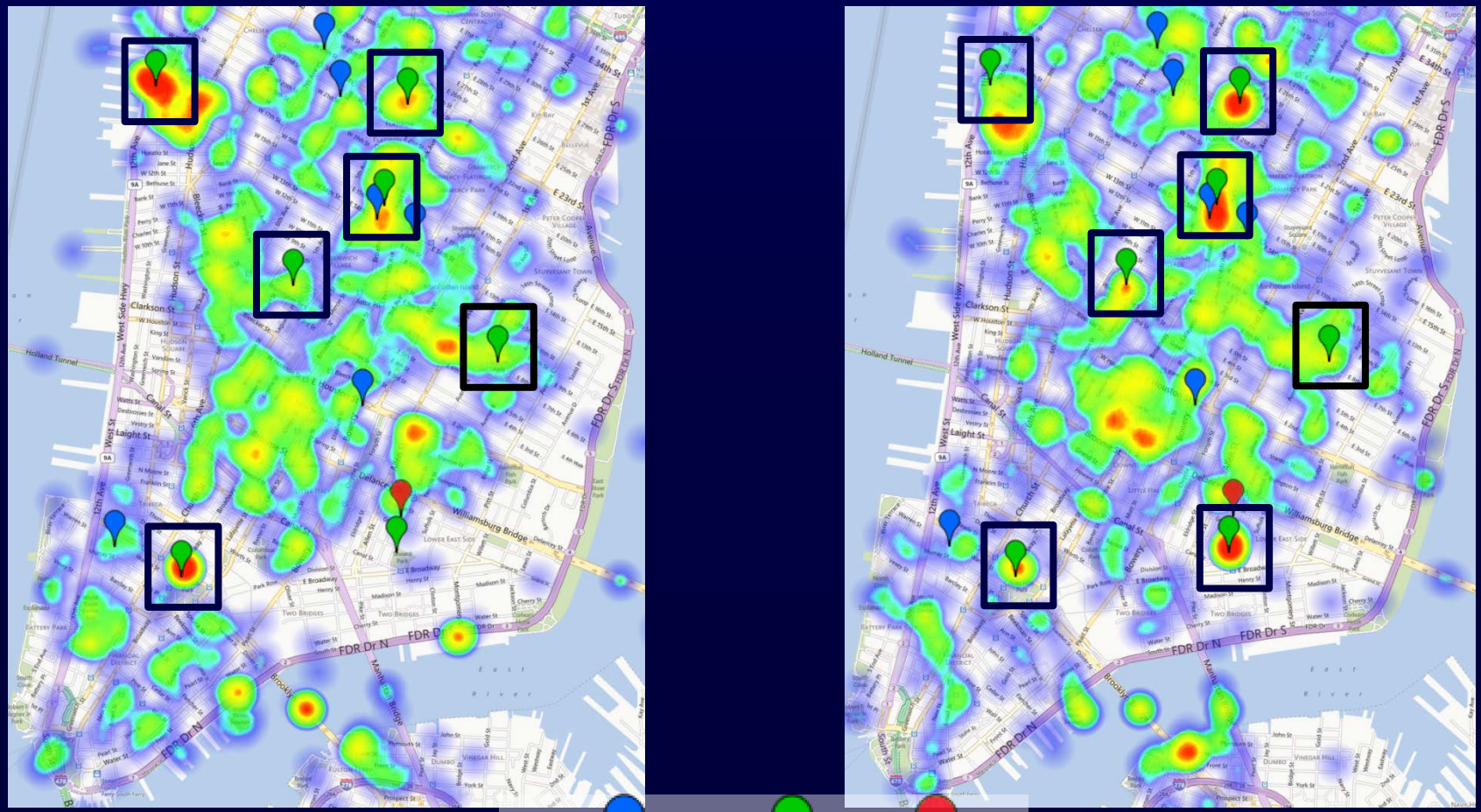
Police Patrolling The River



Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

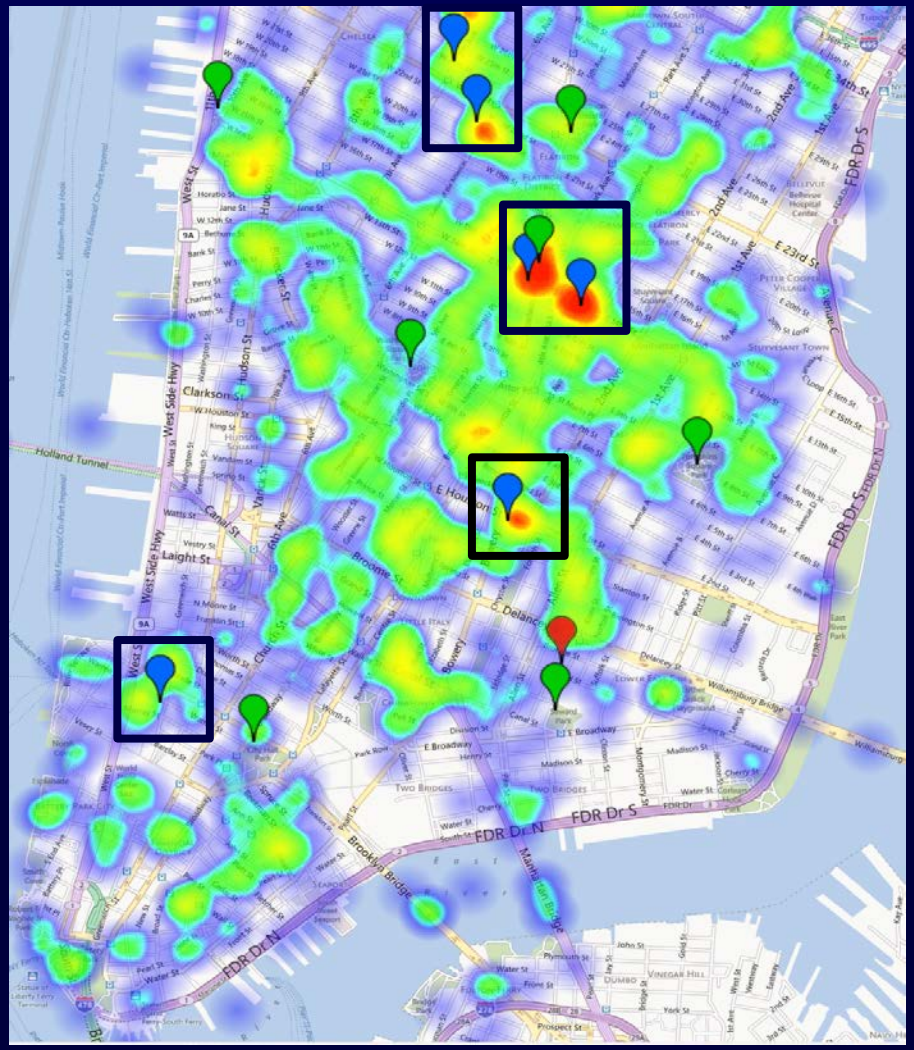


Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM


West St
Clarkson St
Holland Tunnel
West Side Hwy
Laight St
Canal St
Broome St
E Houston St
E Delancey St
Bowery
Chrystie St
Forsyth St
Allen St
Essex St
Pitt St
Avenue A
1st Ave
2nd Ave
3rd Ave
4th Ave
E 23rd St
E 14th St
E 13th St
E 10th St
E 9th St
E 8th St
E 7th St
E 6th St
E 5th St
E 4th St
E 3rd St
E 2nd St
E 1st St
Stanton St
Rivington St
Houston St
Prince St
Spring St
Vestry St
Franklin St
Thomas St
Worth St
Park Row
Chambers St
Warren St
Murray St
Barclay St
Vesey St
Cortlandt St
Fulton St
Dey St
Liberty St
Cedar St
Pine St
Wall St
Beaver St
Whitehall St
Broadway
E Broadway
Madison St
Cherry St
Water St
Division St
Two Bridges
Henry St
Market St
Pearl St
Frankfort St
Gold St
John St
Fulton St
Ann St
Nassau St
Spruce St
Park Pl
Barclay St
Vesey St
Murray St
Warren St
Chambers St
Reade St
Duane St
Thomas St
Worth St
Park Row
Broadway
Lafayette St
Elizabeth St
Mott St
Mulberry St
Baxter St
Centre St
Columbus Park
Foley Square
Tribeca
Chelsea
Hudson St
Greenwich St
Washington St
Varick St
Sixth Ave
Seventh Ave
Eighth Ave
Ninth Ave
Tenth Ave
Eleventh Ave
Twelfth Ave
Bleecker St
Houston St
Canal St
Grand St
Delancey St
Rivington St
Stanton St
Avenue B
Avenue C
Avenue D
FDR Dr S
FDR Dr N
Brooklyn Bridge
Manhattan Bridge
Williamsburg Bridge
Queensboro Bridge
Two Bridges
Chinatown
Little Italy
Soho
Nolita
East Village
Lower East Side
Gramercy
Flatiron
Union Square
Greenwich Village
Meatpacking District
Stuyvesant Town
Peter Cooper Village
Bellevue Hospital Center
Kips Bay
Murray Hill
Midtown South
Midtown Tunnel
Lincoln Tunnel
Holland Tunnel
Battery Park
Financial District
Seaport
Brooklyn
Dumbo
Vinegar Hill
Navy Yard
Prospect St
York St
Plymouth St
Jay St
Gold St
Front St
Water St
Pearl St
Frankfort St
Park Pl
Barclay St
Vesey St
Murray St
Warren St
Chambers St
Reade St
Duane St
Thomas St
Worth St
Park Row
Broadway
Lafayette St
Elizabeth St
Mott St
Mulberry St
Baxter St
Centre St
Columbus Park
Foley Square
Tribeca
Chelsea
Hudson St
Greenwich St
Washington St
Varick St
Sixth Ave
Seventh Ave
Eighth Ave
Ninth Ave
Tenth Ave
Eleventh Ave
Twelfth Ave
Bleecker St
Houston St
Canal St
Grand St
Delancey St
Rivington St
Stanton St
Avenue A
1st Ave
2nd Ave
3rd Ave
4th Ave
E 23rd St
E 14th St
E 13th St
E 10th St
E 9th St
E 8th St
E 7th St
E 6th St
E 5th St
E 4th St
E 3rd St
E 2nd St
E 1st St
Stanton St
Rivington St
Houston St
Prince St
Spring St
Vestry St
Franklin St
Thomas St
Worth St
Park Row
Chambers St
Warren St
Murray St
Barclay St
Vesey St
Cortlandt St
Fulton St
Dey St
Liberty St
Cedar St
Pine St
Wall St
Beaver St
Whitehall St
Broadway
E Broadway
Madison St
Cherry St
Water St
Division St
Two Bridges
Henry St
Market St
Pearl St
Frankfort St
Gold St
John St
Fulton St
Ann St
Nassau St
Spruce

For Further Information

www.VisualAnalytics-CCI.org



VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics For Effective Decision Making

David S. Ebert

Purdue University

ebertd@purdue.edu

CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

December 2014

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America





What Our Visual Analytic Solutions Offer

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

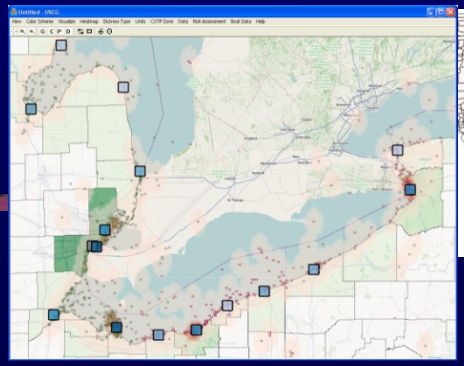
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

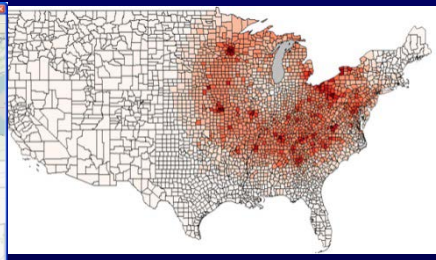
VADM Robert Parker, 2012 MRS Keynote Address

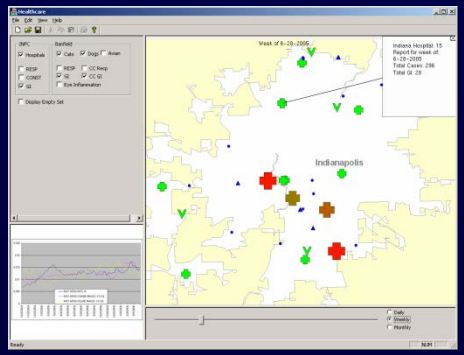


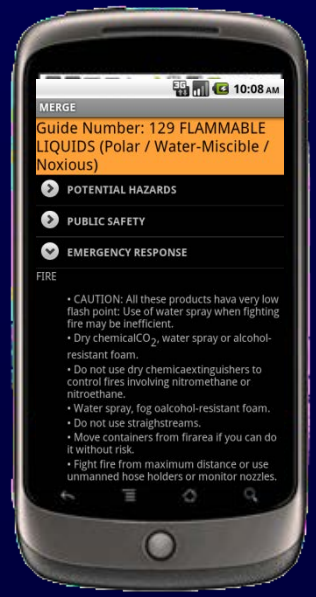
Public Safety & Health

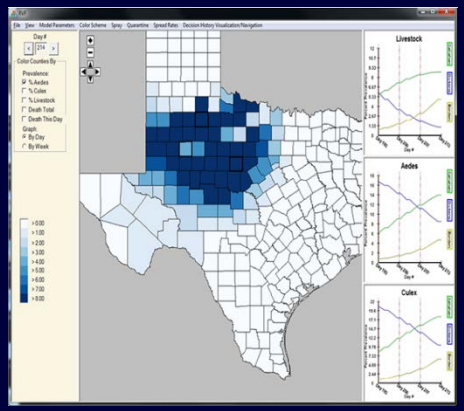
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering







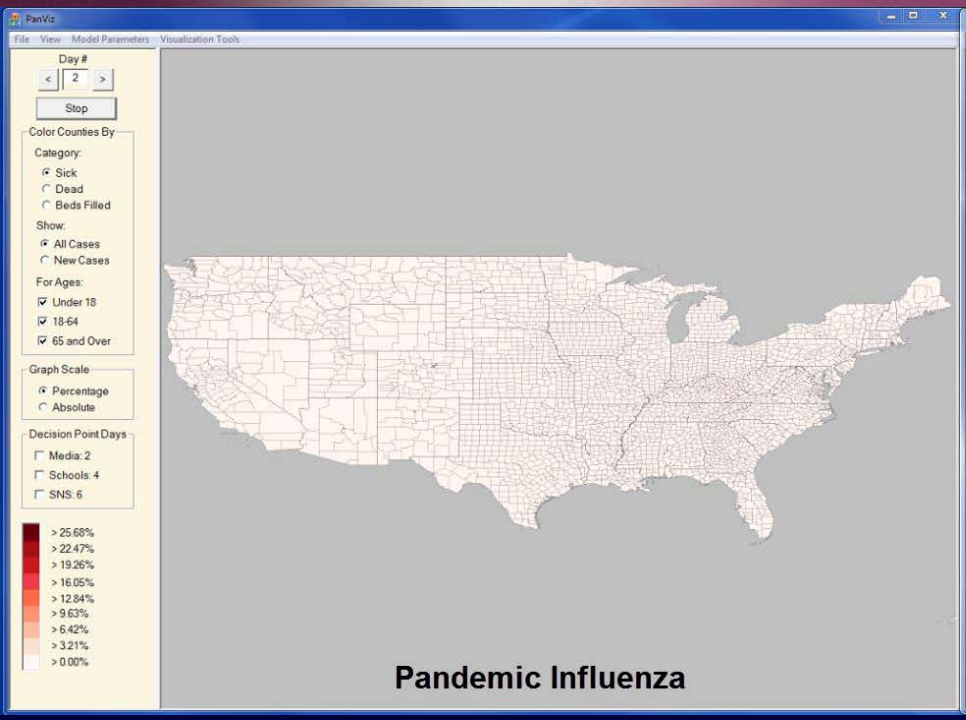


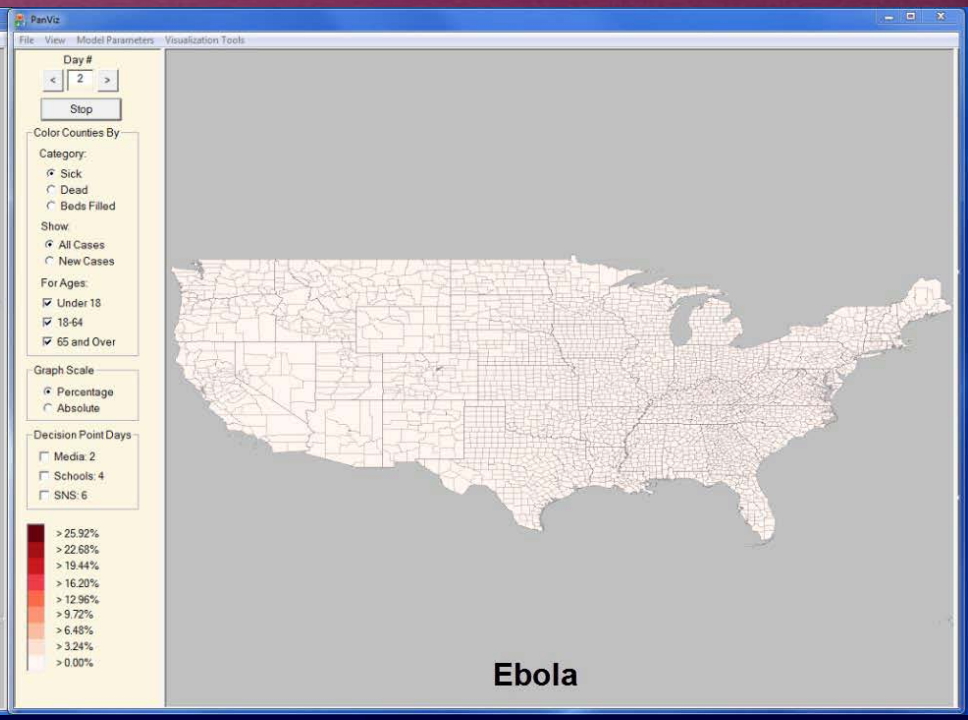




VACCINE

Ebola Vs. Pandemic Influenza (H1N1)



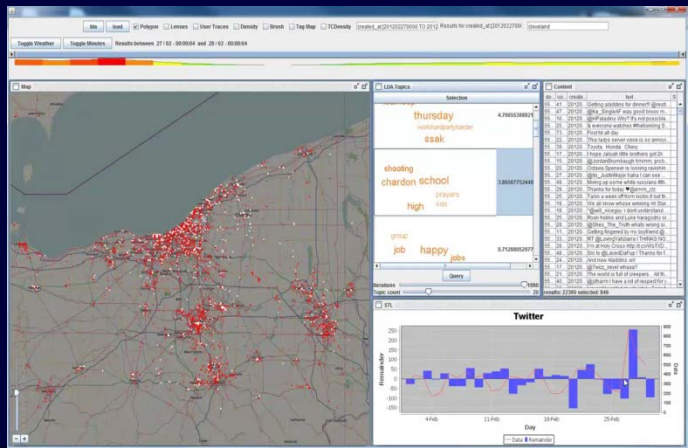


Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

Visual Analytics Uses for Public Safety



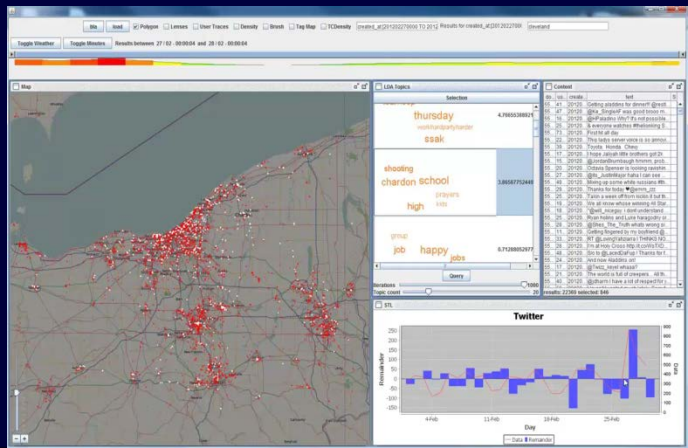
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



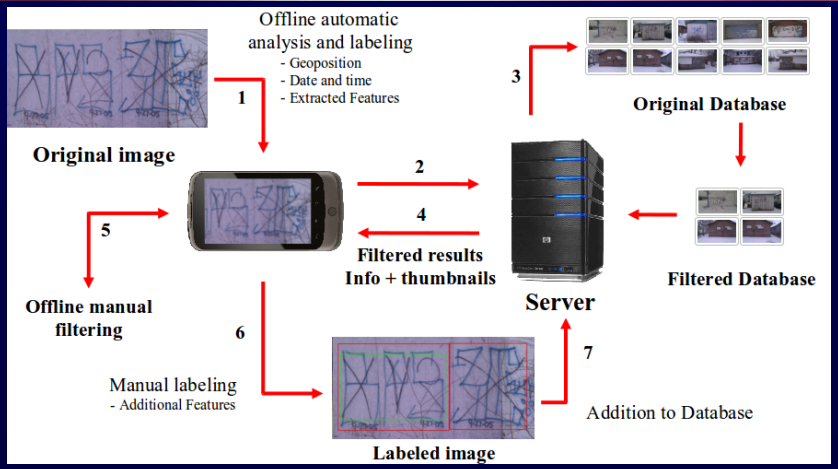
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level



Example Tools

Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 2000 graffiti and tatoo images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
- Spring 2014
  - Public reporting version

VACCINE

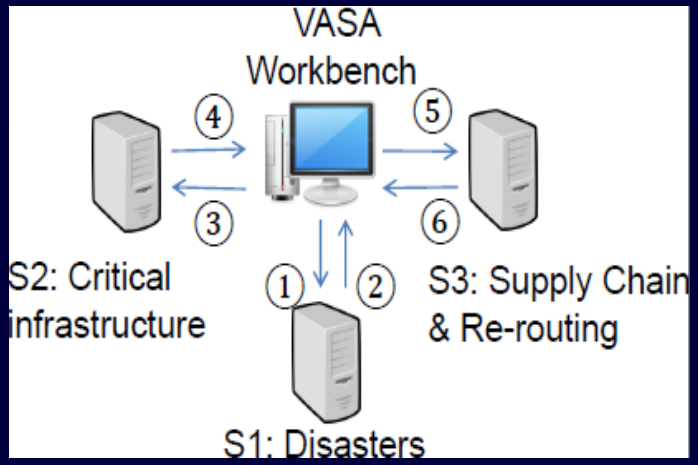
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

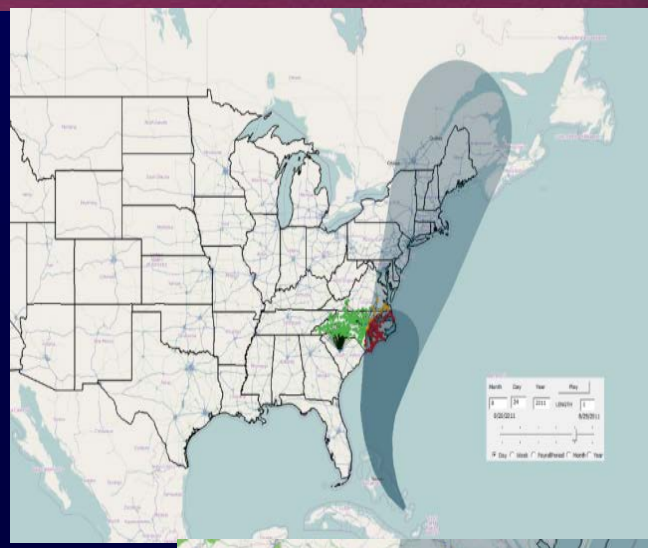
TACC

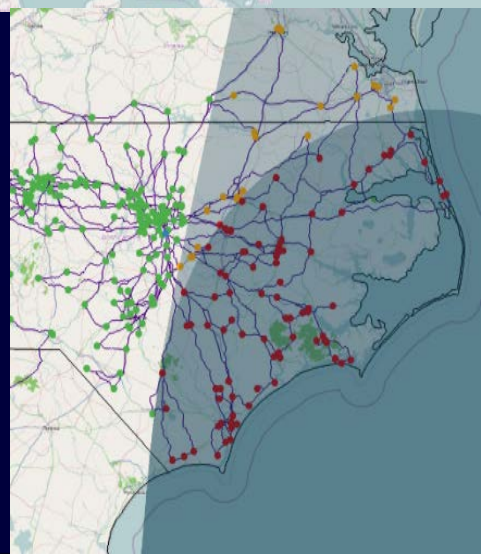
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

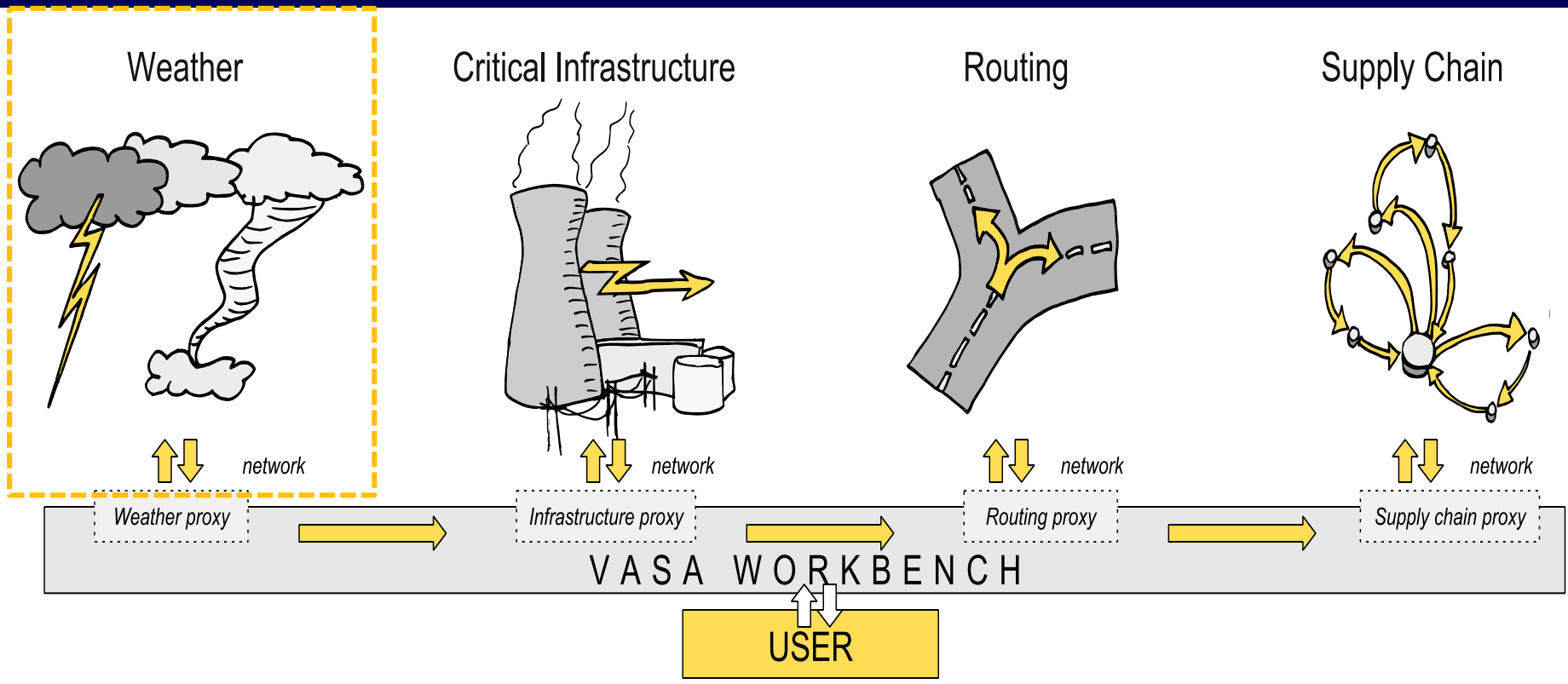
Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models



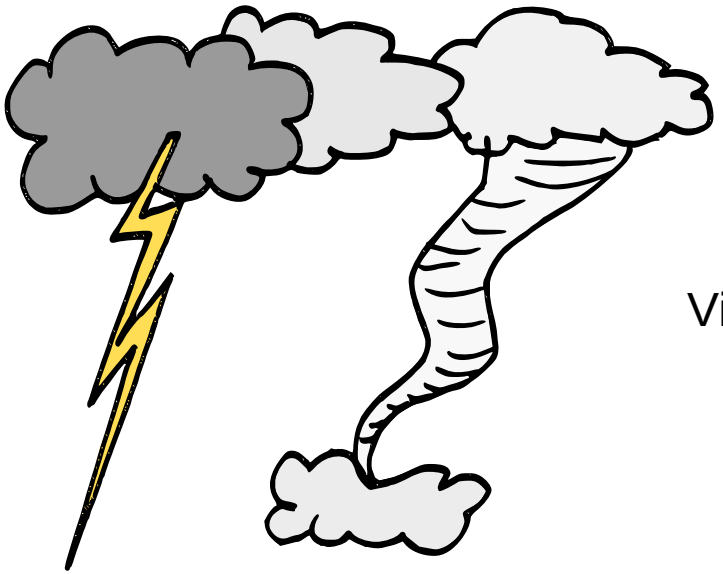


VASA COMPONENT: **WEATHER**



VASA COMPONENT: **WEATHER**

Historic data
(Irene, Sandy, etc)



Prepares **event datasets** from server

ADCIRC: Advanced Circulation

Visualizes **hurricane cone** over time using slider

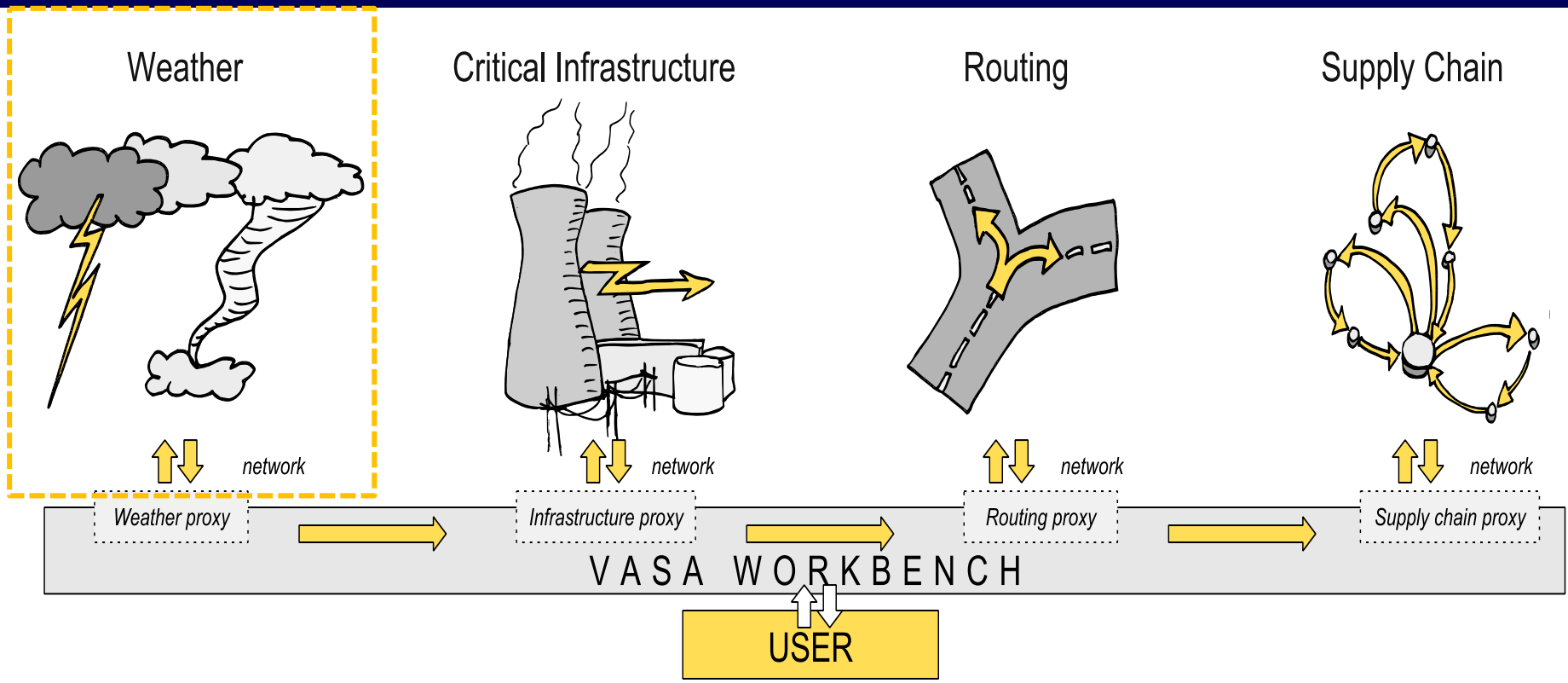
NOAA: National Oceanic and Athmospheric Administration

Generates **inputs** to downstream components

MODEL

PROXY

VASA COMPONENT:
CRITICAL INFRASTRUCTURE



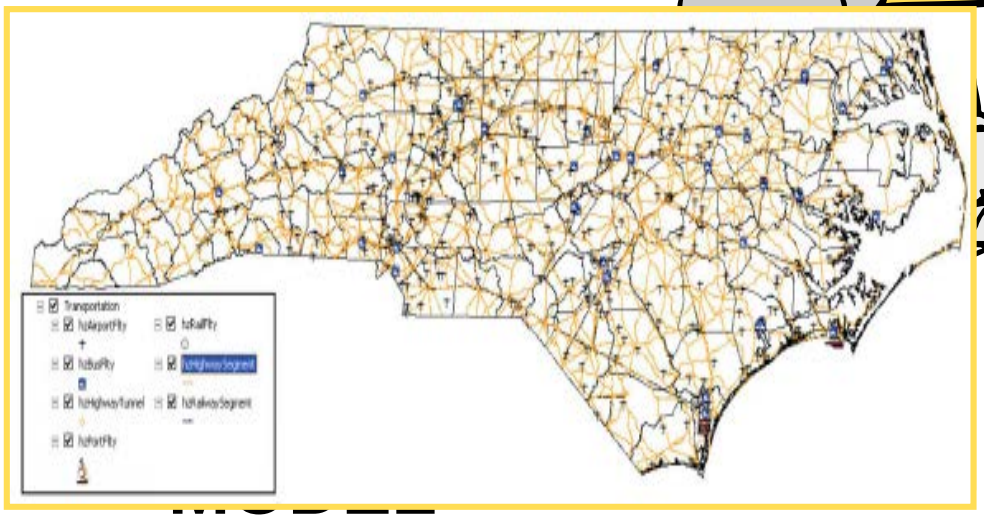
VASA COMPONENT:
CRITICAL INFRASTRUCTURE

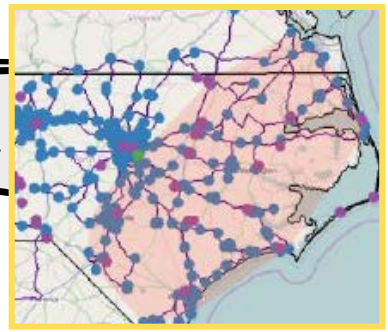
Vu environment with submodels

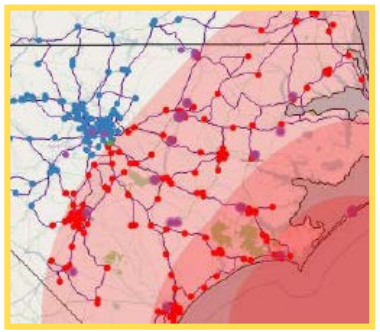
Simplified **connectivity graph** of important structures

E.g. Electric grids, telecom networks, gas distribution

Example: show impact of hurricanes on restaurants

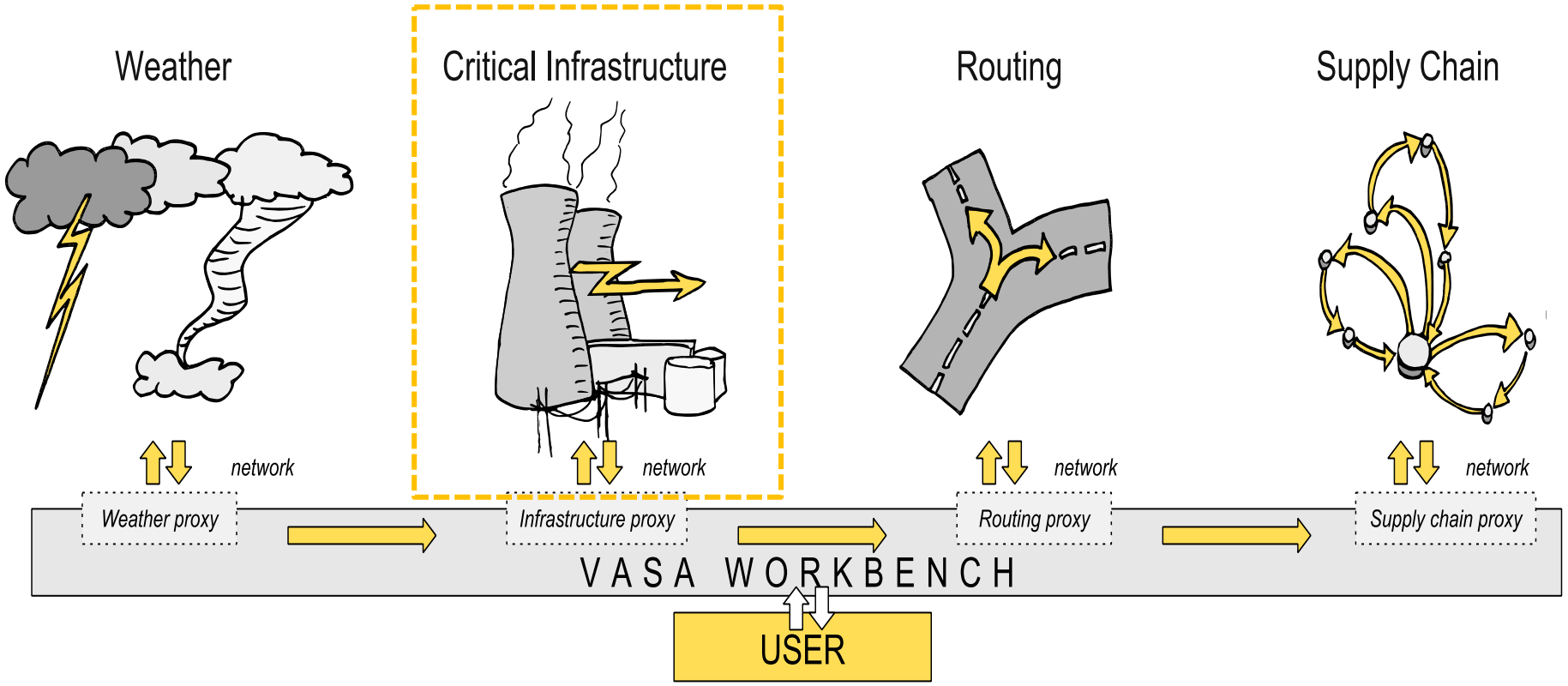






PROXY

VASA COMPONENT: **ROUTING**

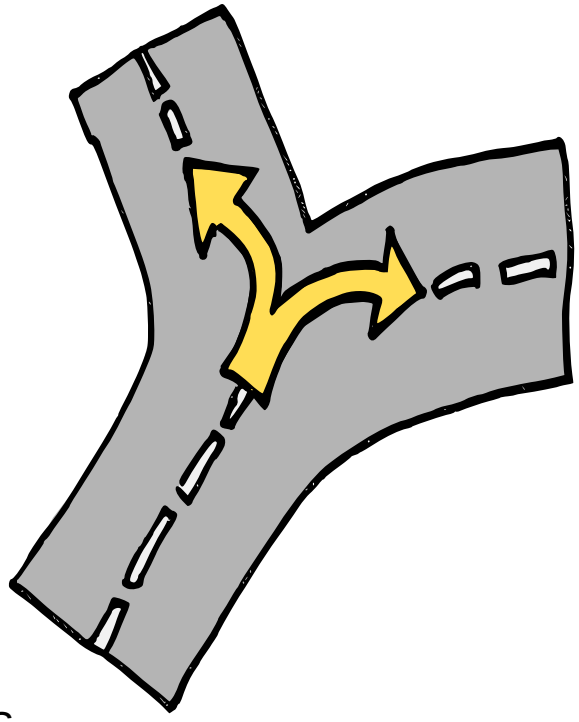


VASA COMPONENT:
ROUTING

Maintains **road network** for critical infrastructure

Simulation engine:
ArcGIS Server

Inputs: barriers and closures
Outputs: new transport routes



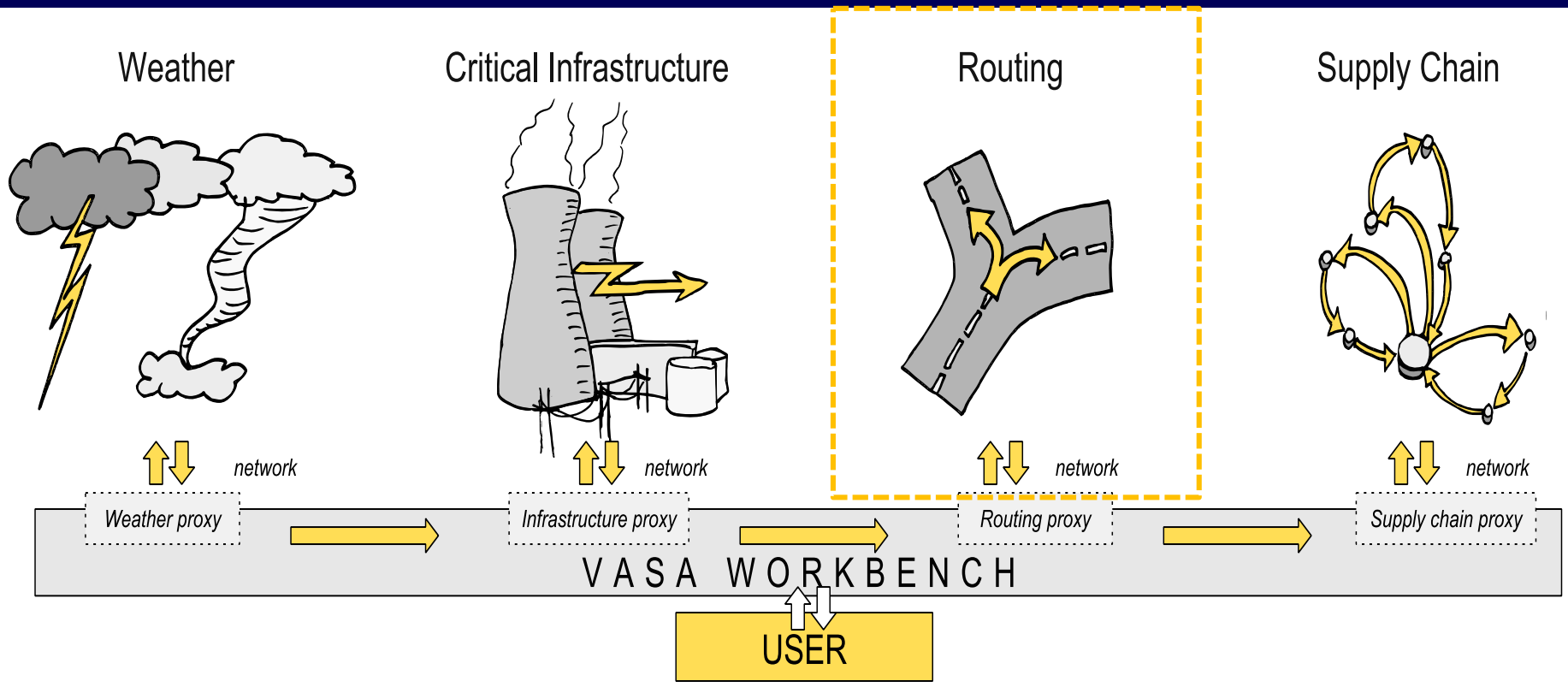
Approximates disabled routes and facilities

Visualizes disabled routes and facilities

MODEL

PROXY

VASA COMPONENT:
SUPPLY CHAIN

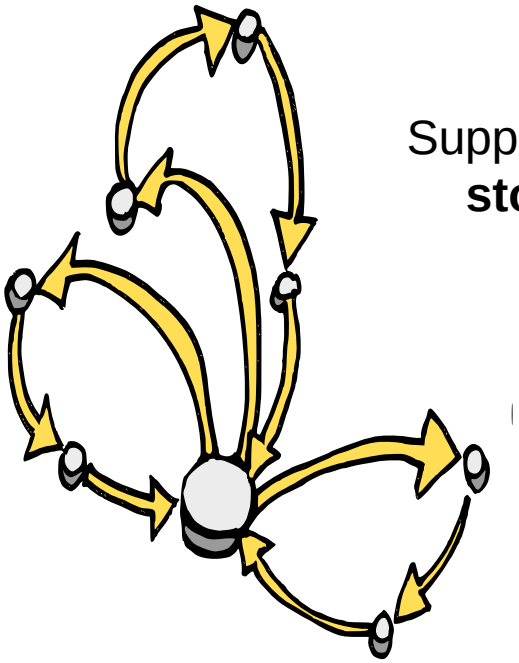


VASA COMPONENT: **SUPPLY CHAIN**

Supply chains depend on **business** and **goods**

Our models: poultry firm + fast food = farm to restaurant

Discrete event simulation on chain in geolocated facilities



Accepts external inputs (**weather** and **roads**)

Supports **road closures, powerless stores** and **flooding** (polygons)

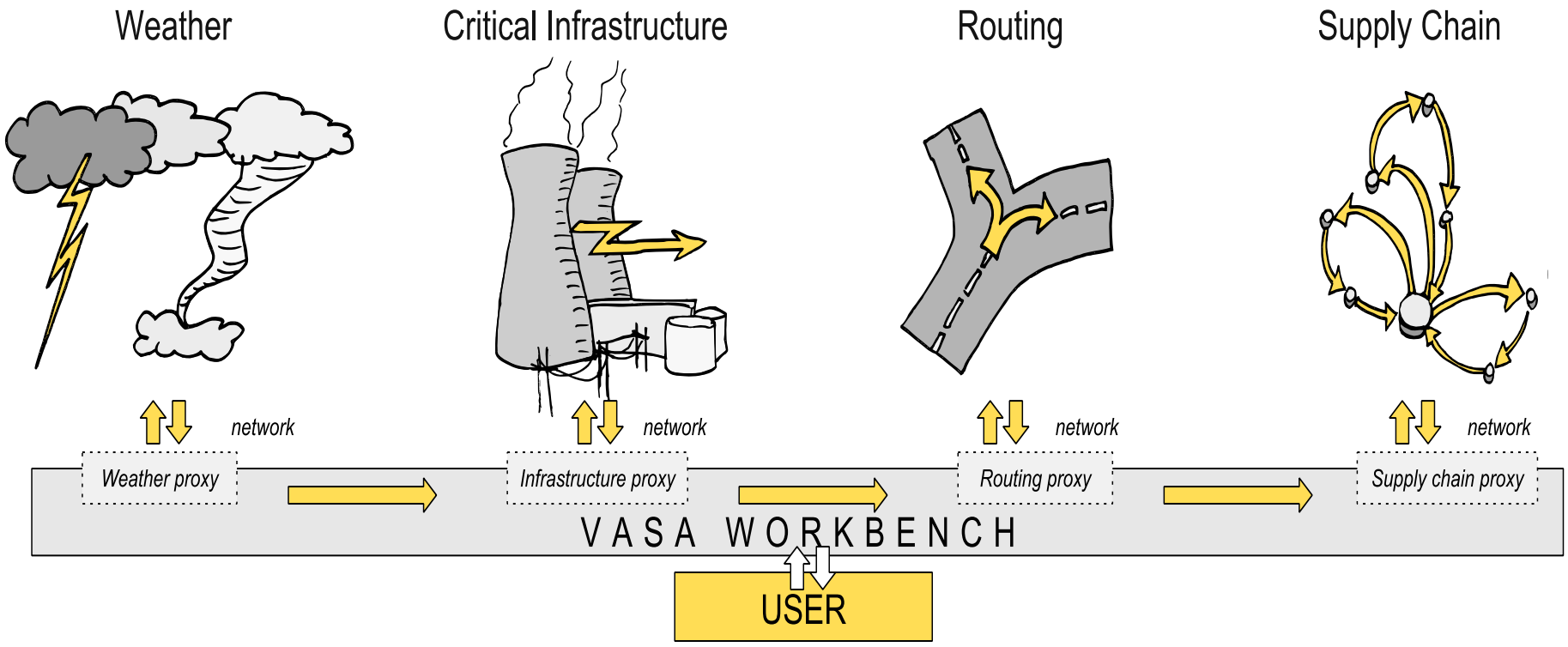
Optimizes distribution and even redistributes products

Food contamination also modeled and visualized

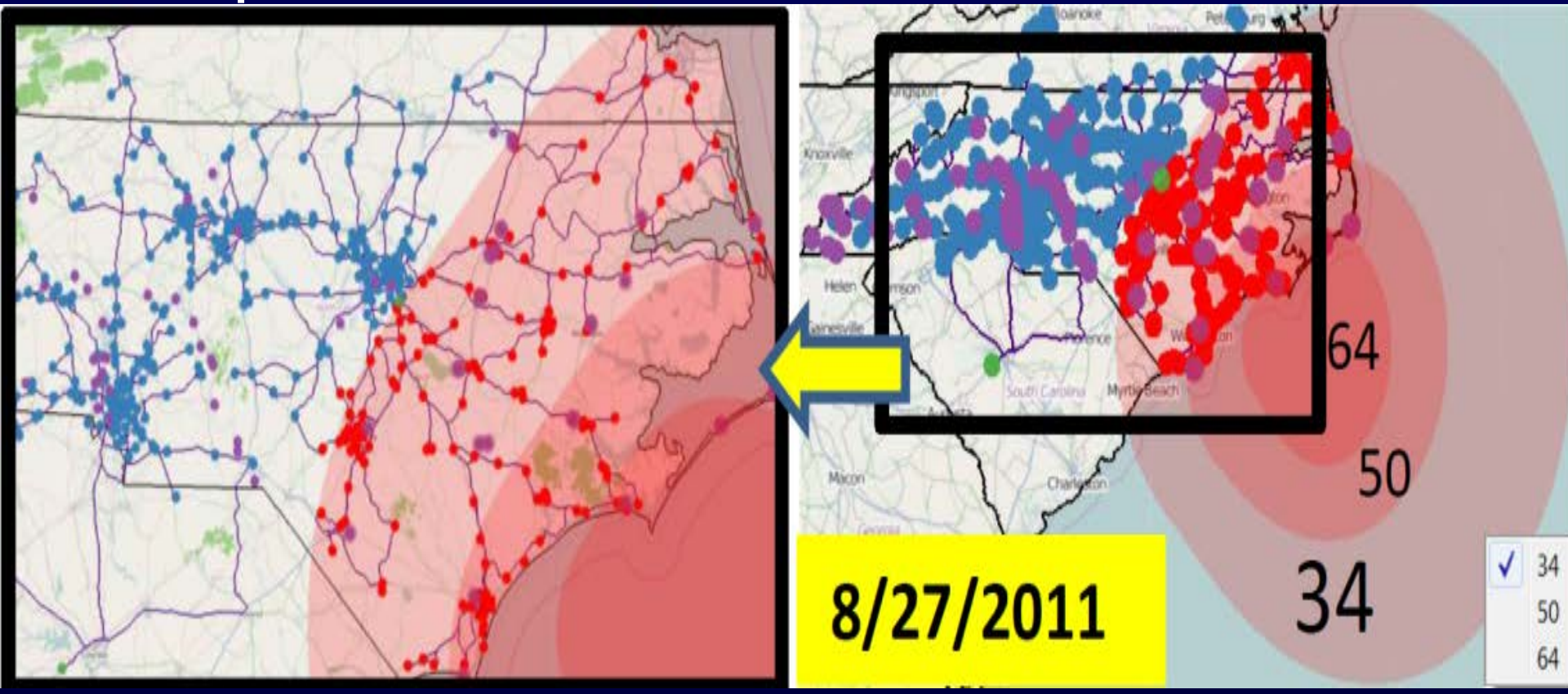
MODEL

PROXY

VASA


Weather
Critical Infrastructure
Routing
Supply Chain
network
network
network
network
Weather proxy
Infrastructure proxy
Routing proxy
Supply chain proxy
VASA WORKBENCH
USER

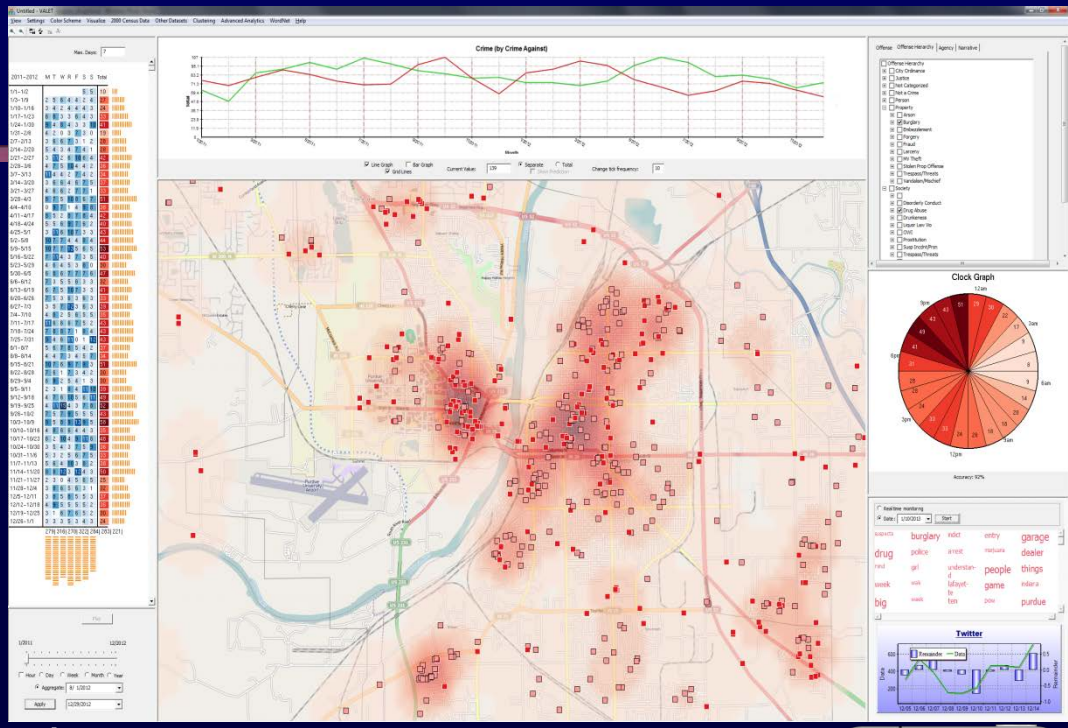
Example: U.S. Hurricane Season



34-knot winds batter the coast; critical infrastructure proxy estimates **impacted restaurant**

Hurricane Irene hits North Carolina on August 27, 2011

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

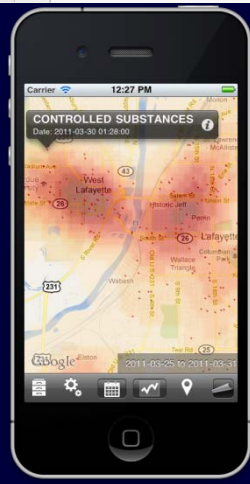
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

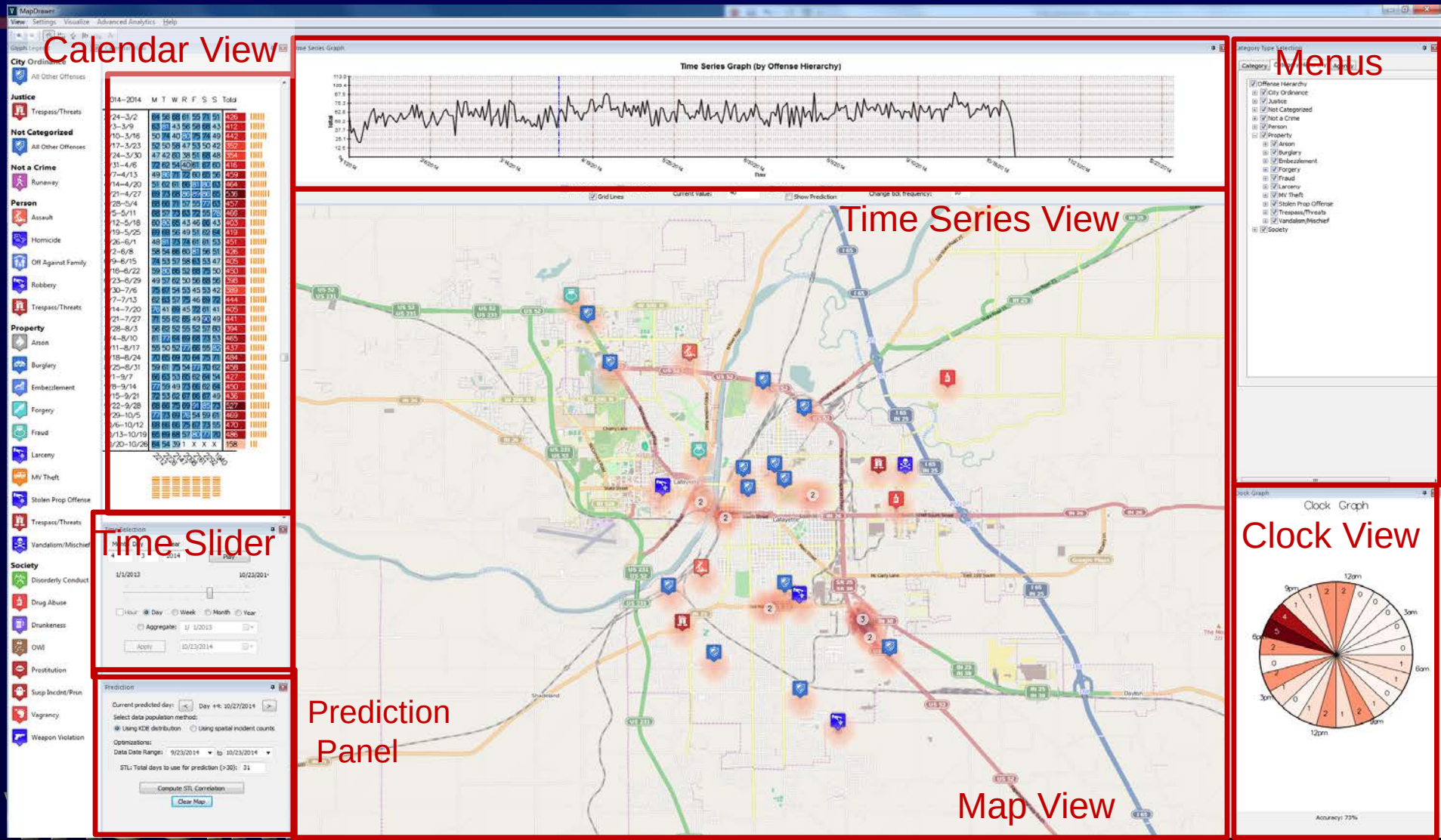
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

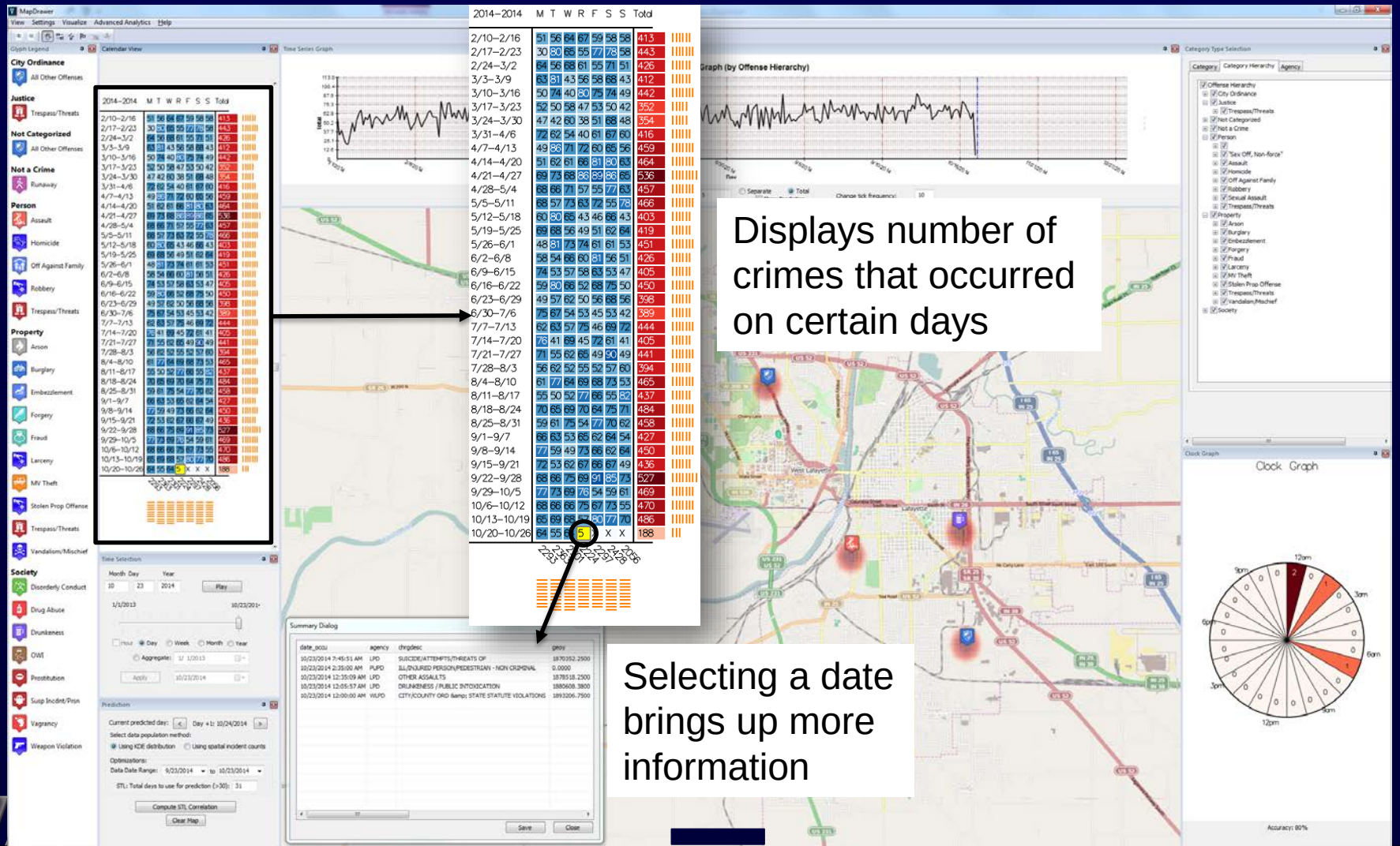


VACCINE

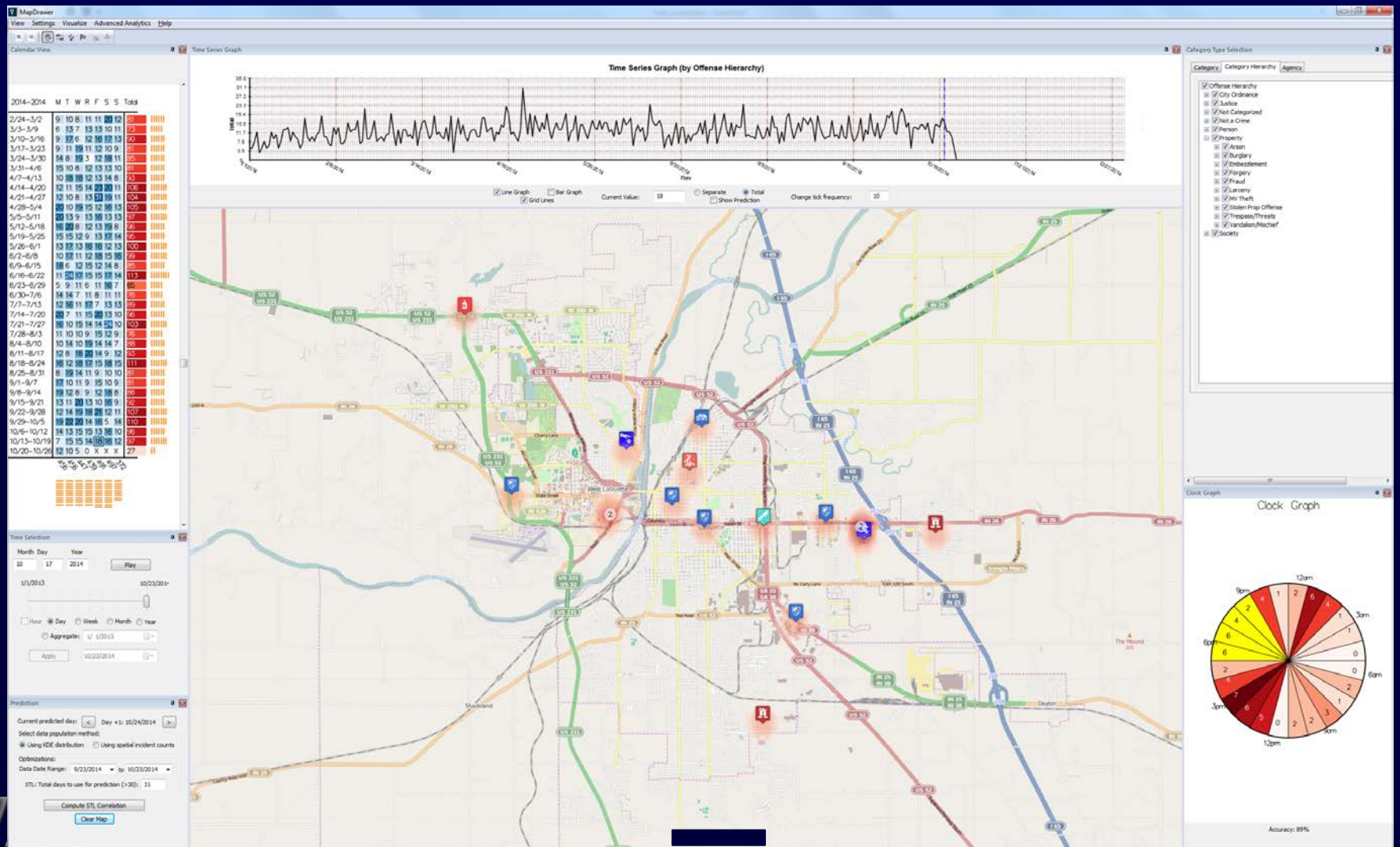
VALET Layout



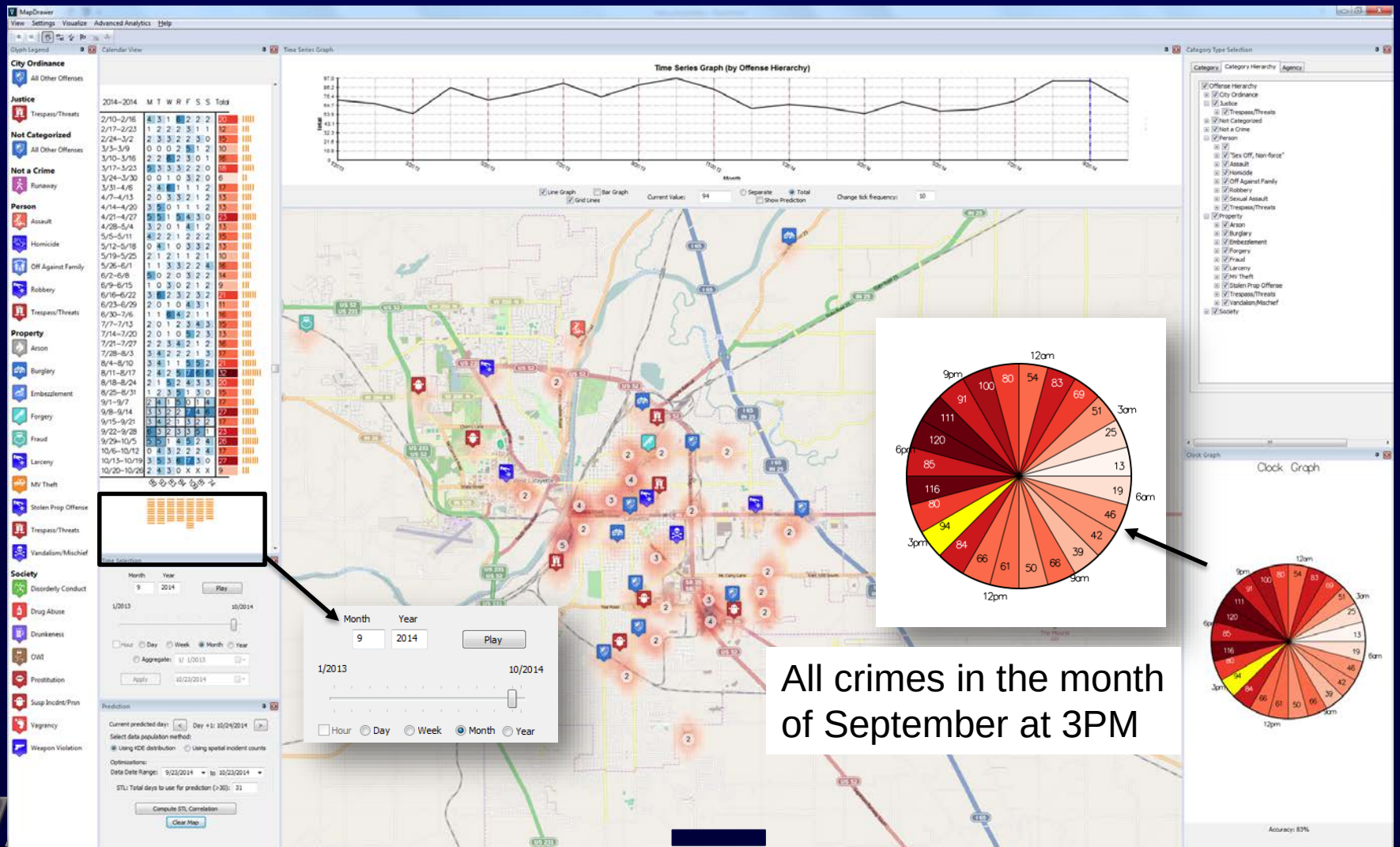
Calendar Display



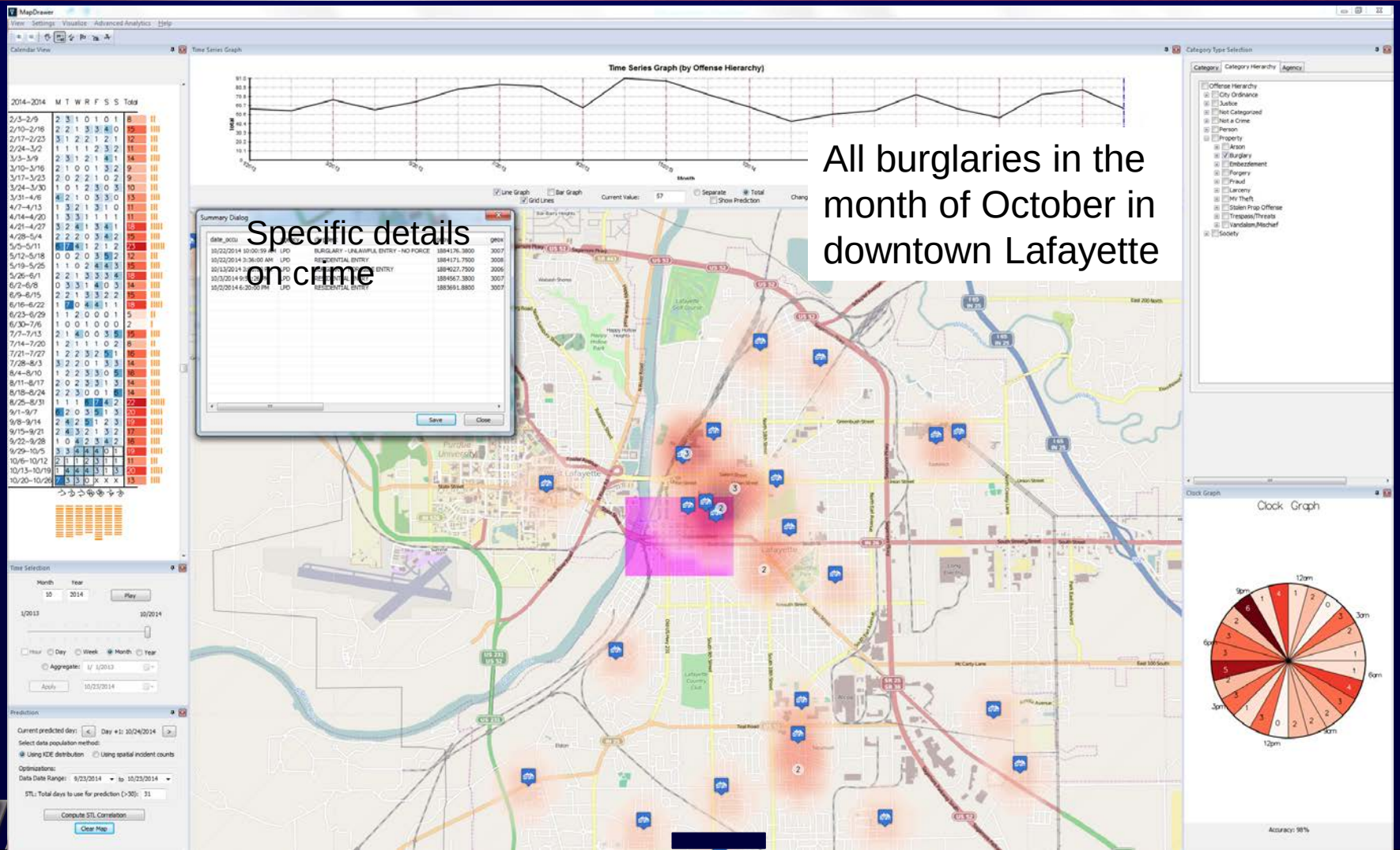
Crimes In A Single Day From 6PM – 9PM



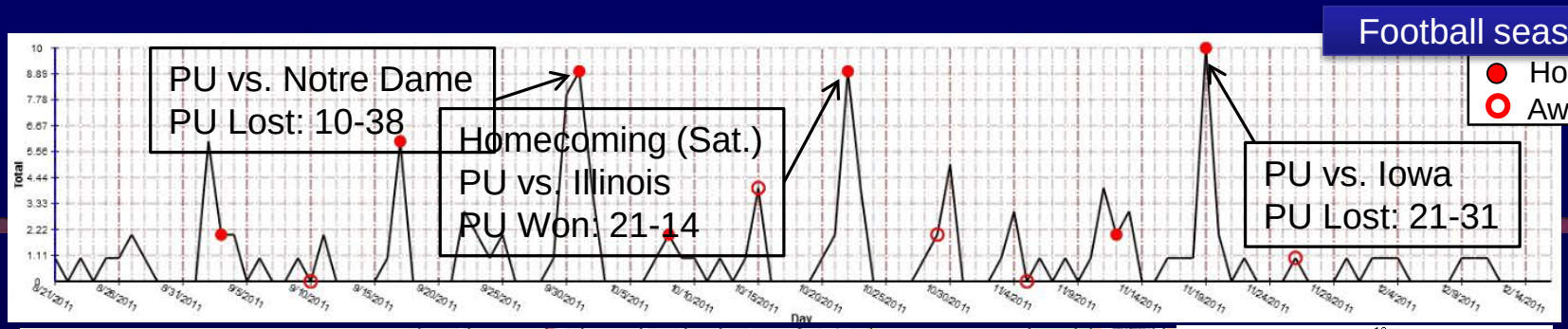
Crimes During Certain Hours In A Month

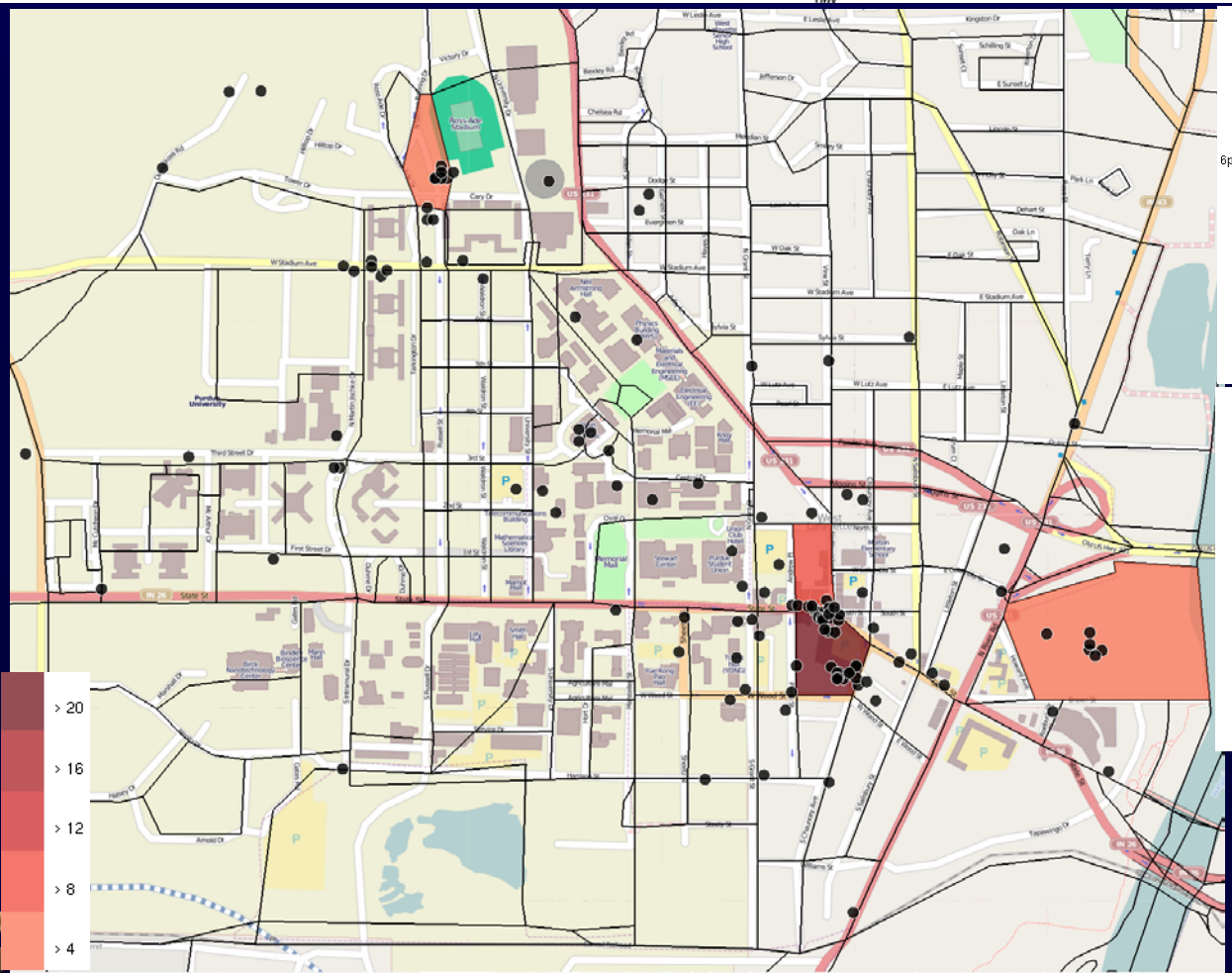


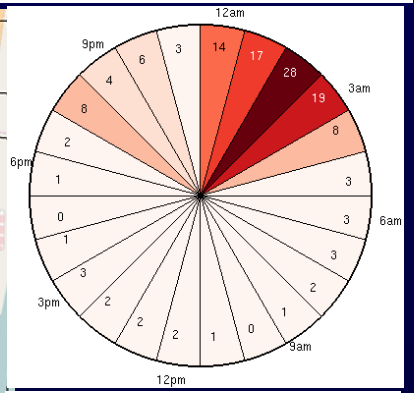
Select Specific Region To Display Crimes

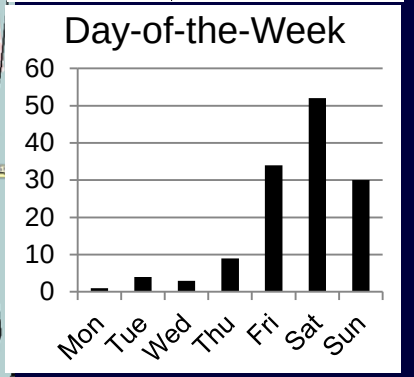


Example: More Data Sources – Intoxication

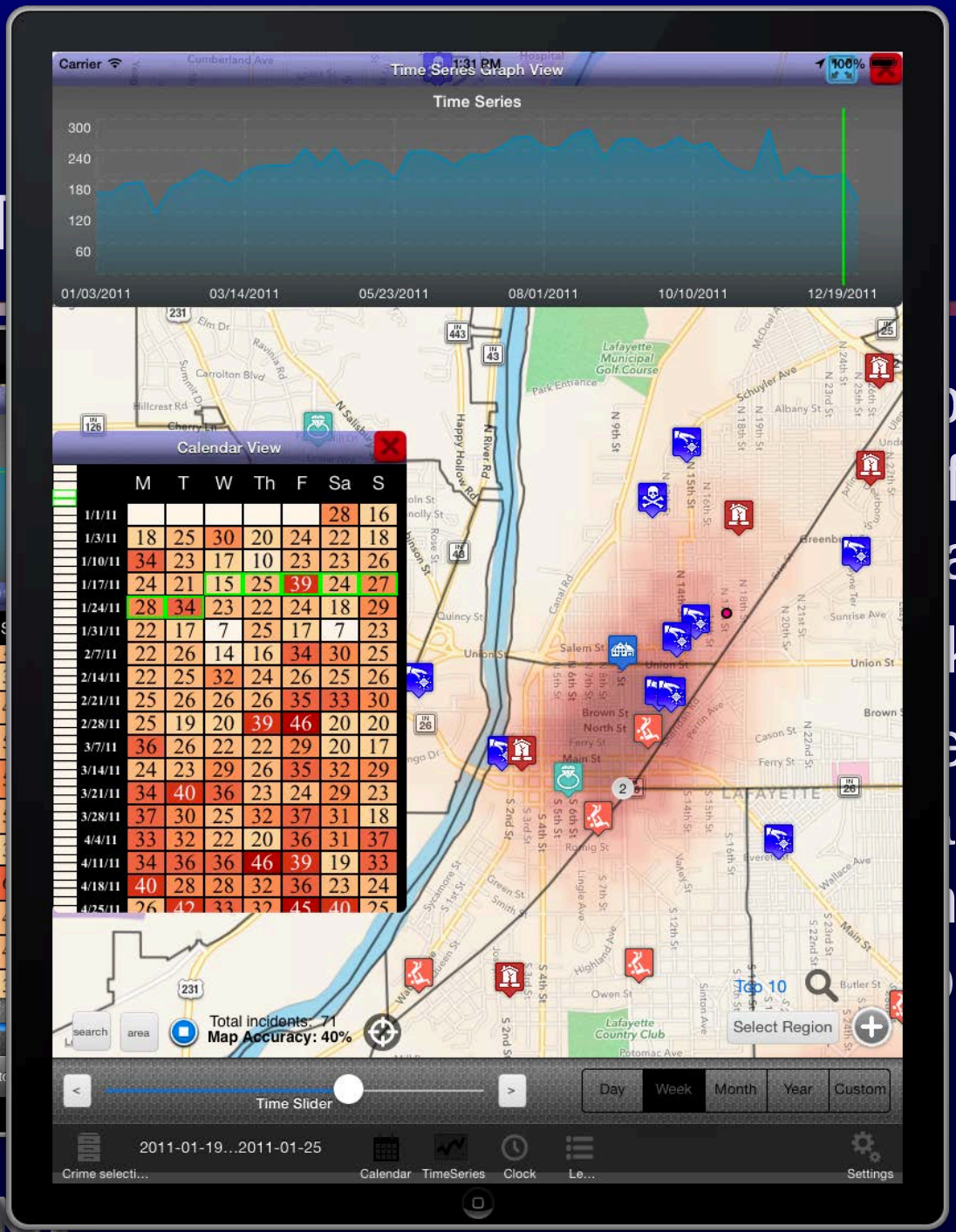




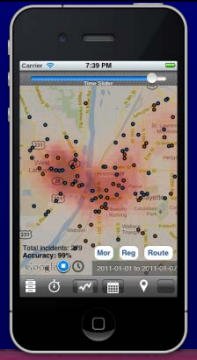


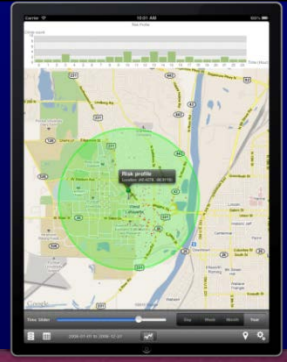


iVALET











lore criminal,
fic and civil
a on-the-go
k assessment
e current
tial +
poral context
analysis

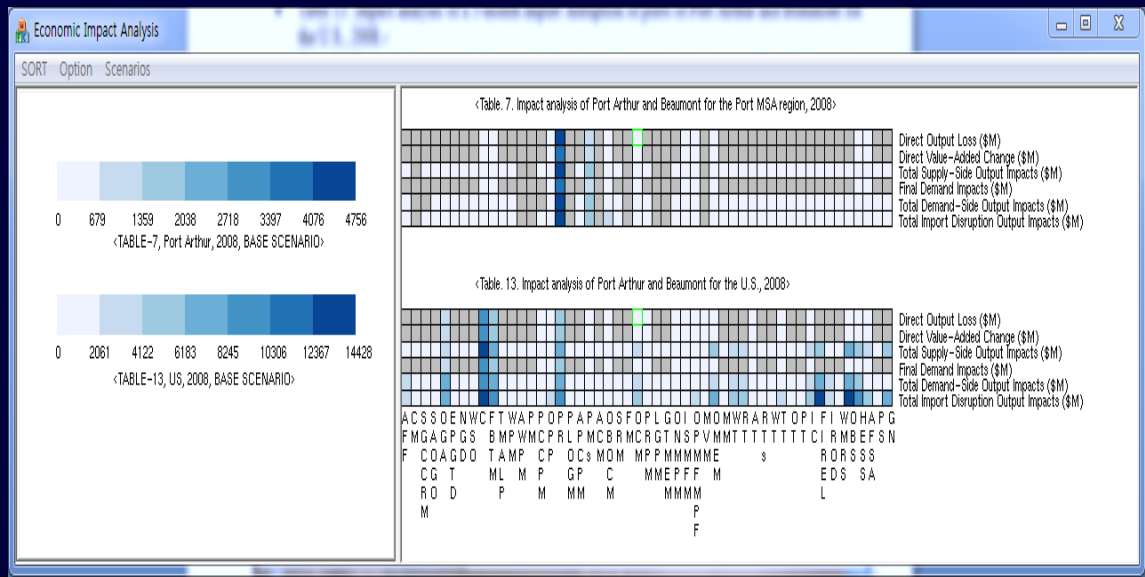
VACCINE

USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

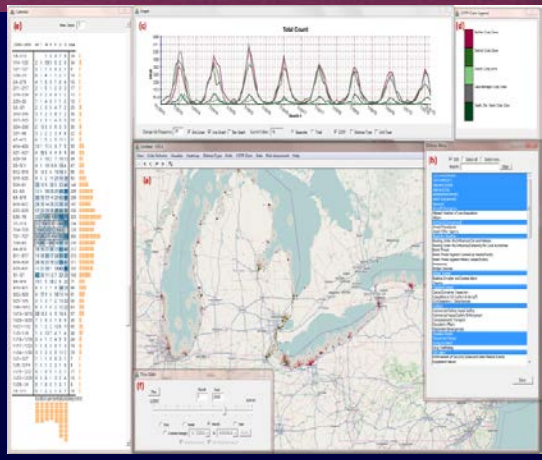
- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



VACCINE

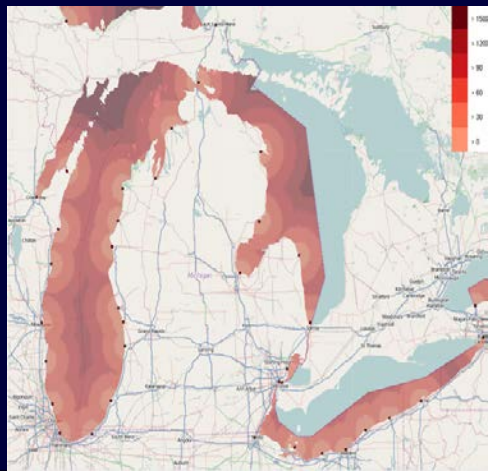
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

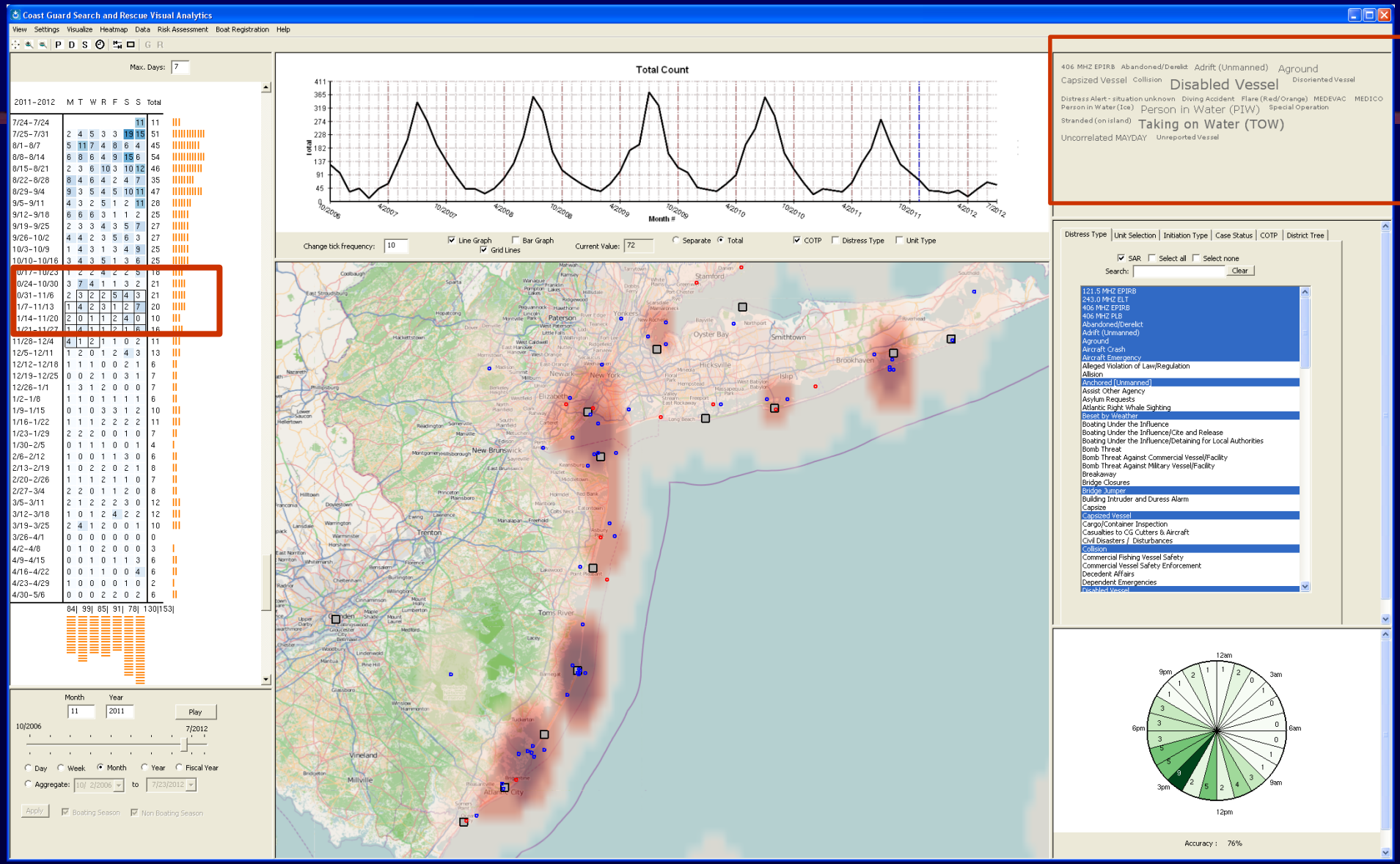


VACCINE

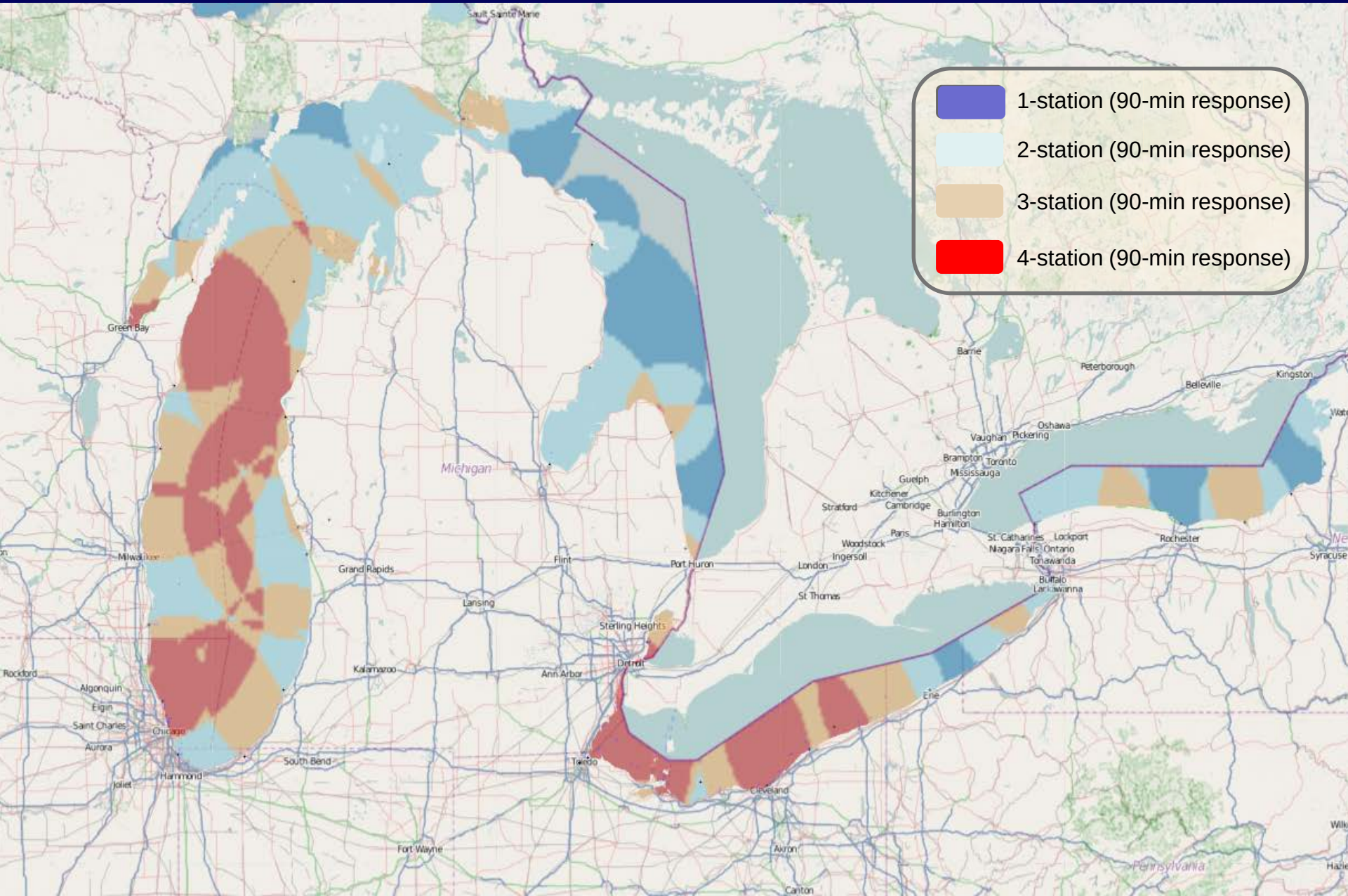
Video

Example: Risks and Consequences From Sandy:

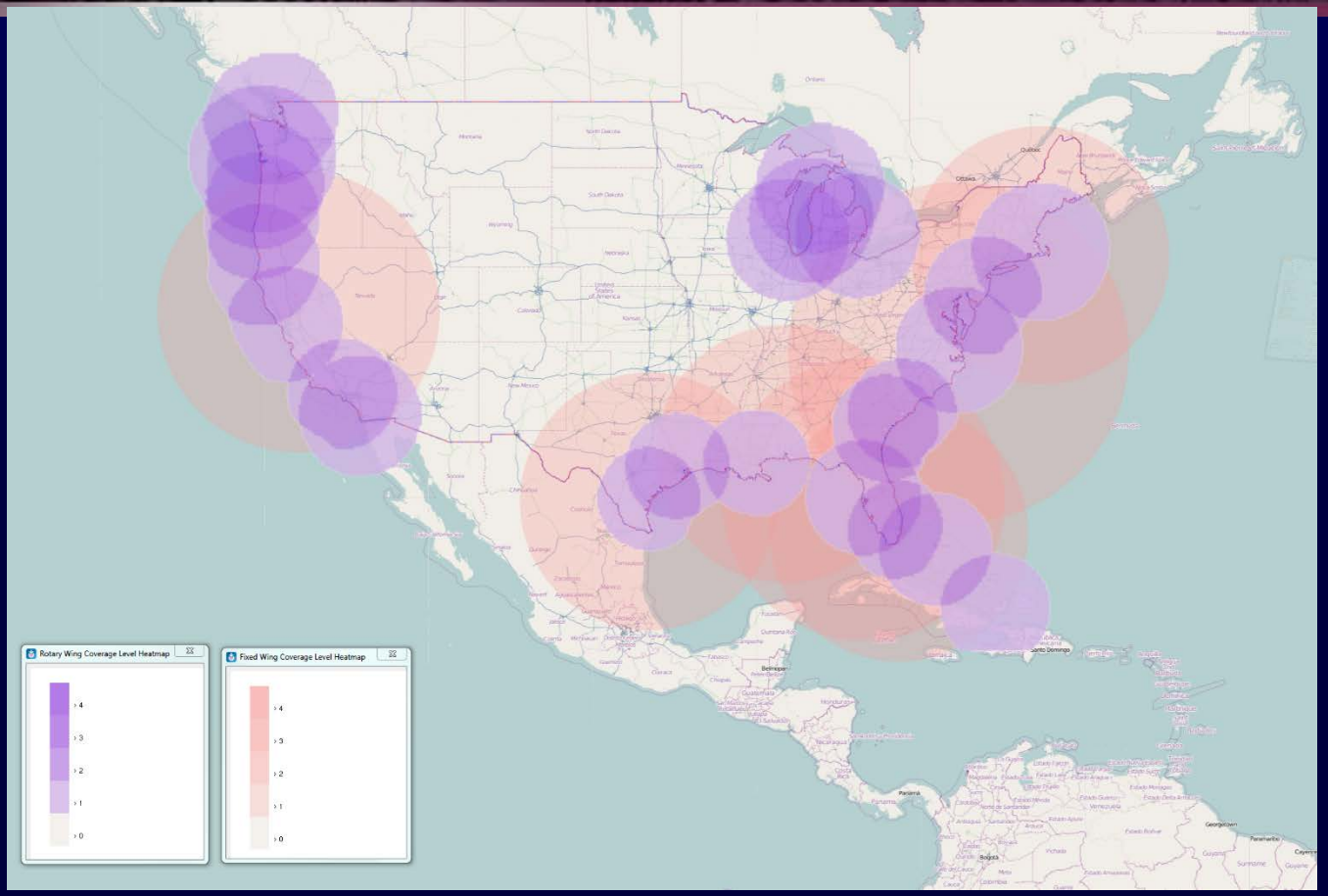
SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



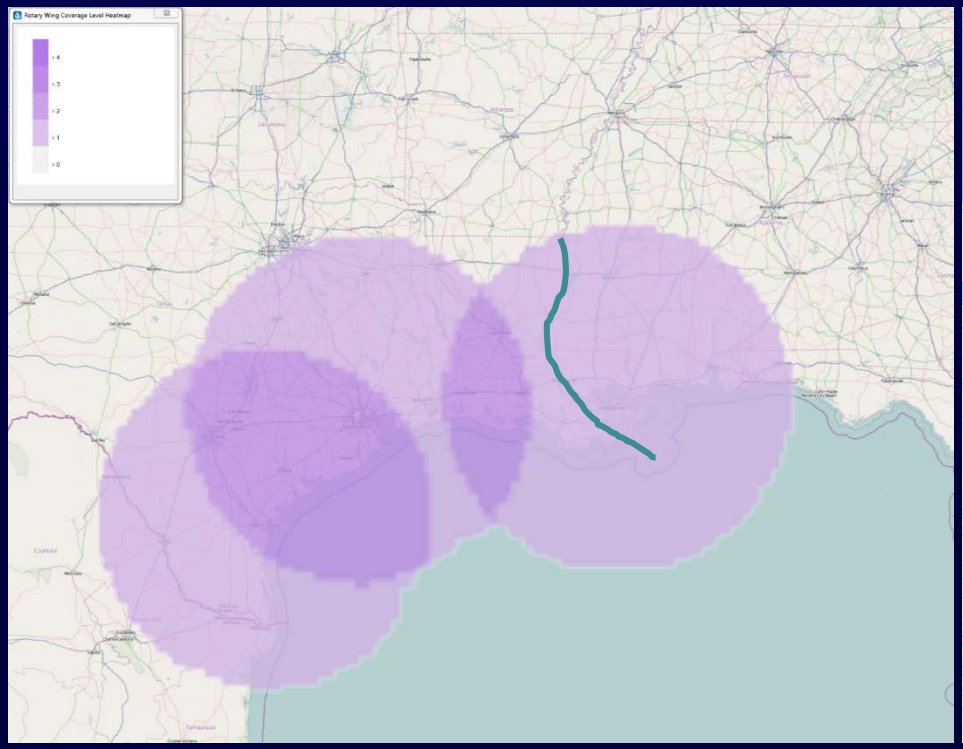
cgSARVA Results – Aircraft coverage

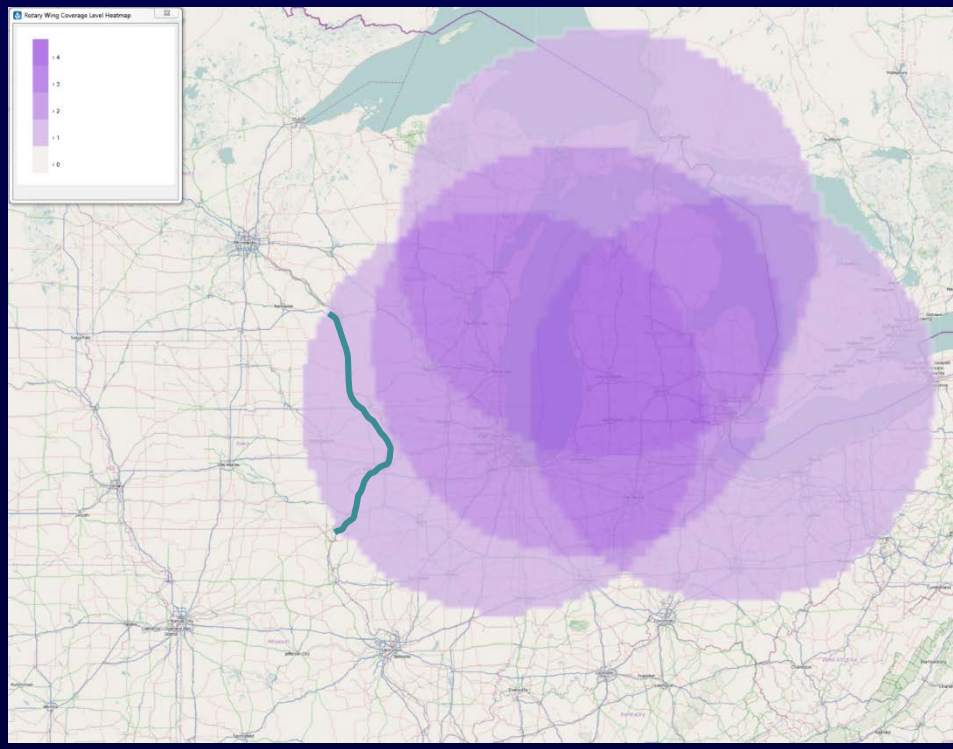


Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.

Airstation Utility Visualization Comparison

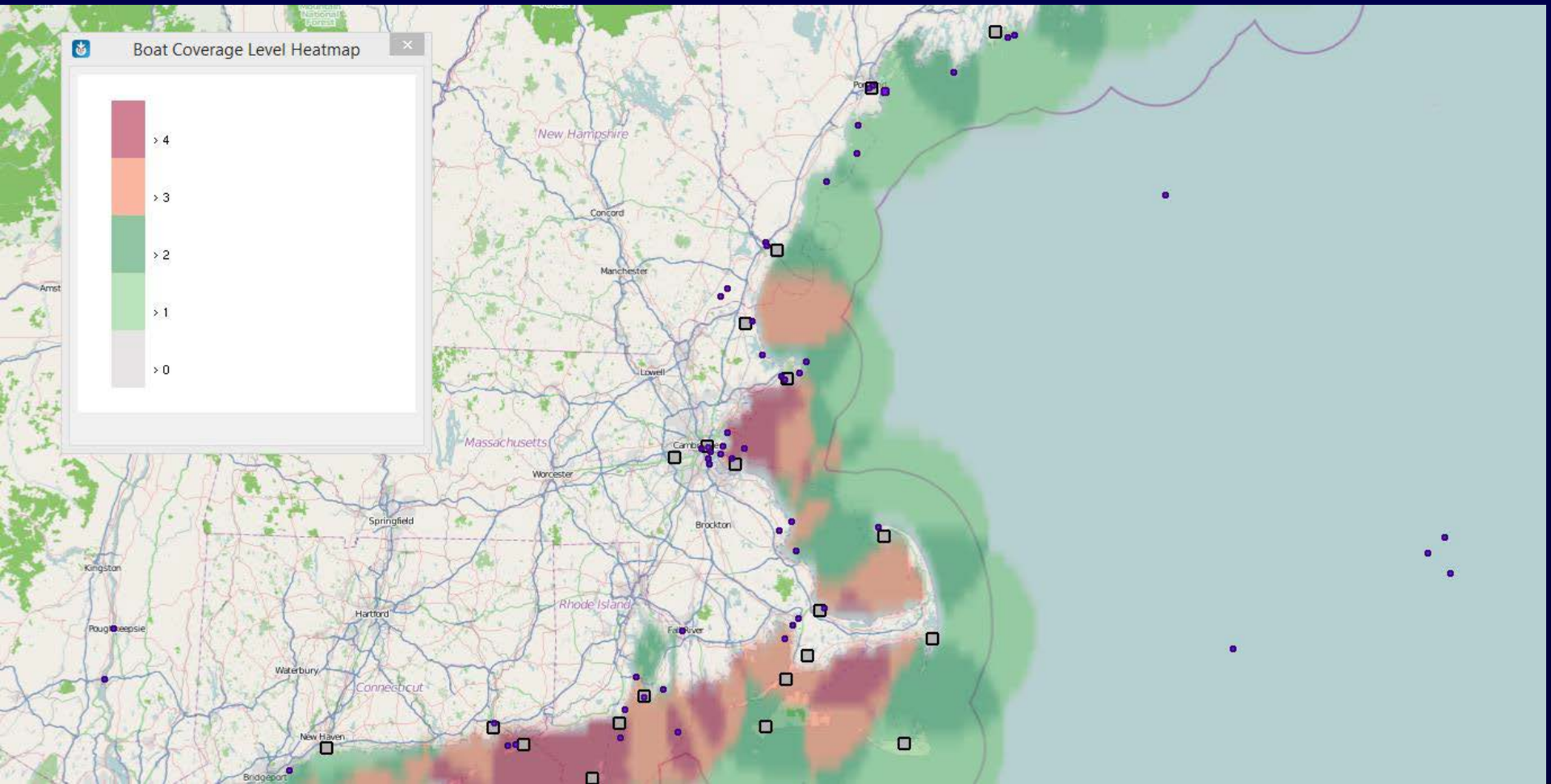
- Rotary Wing Coverage Map by AirSTA

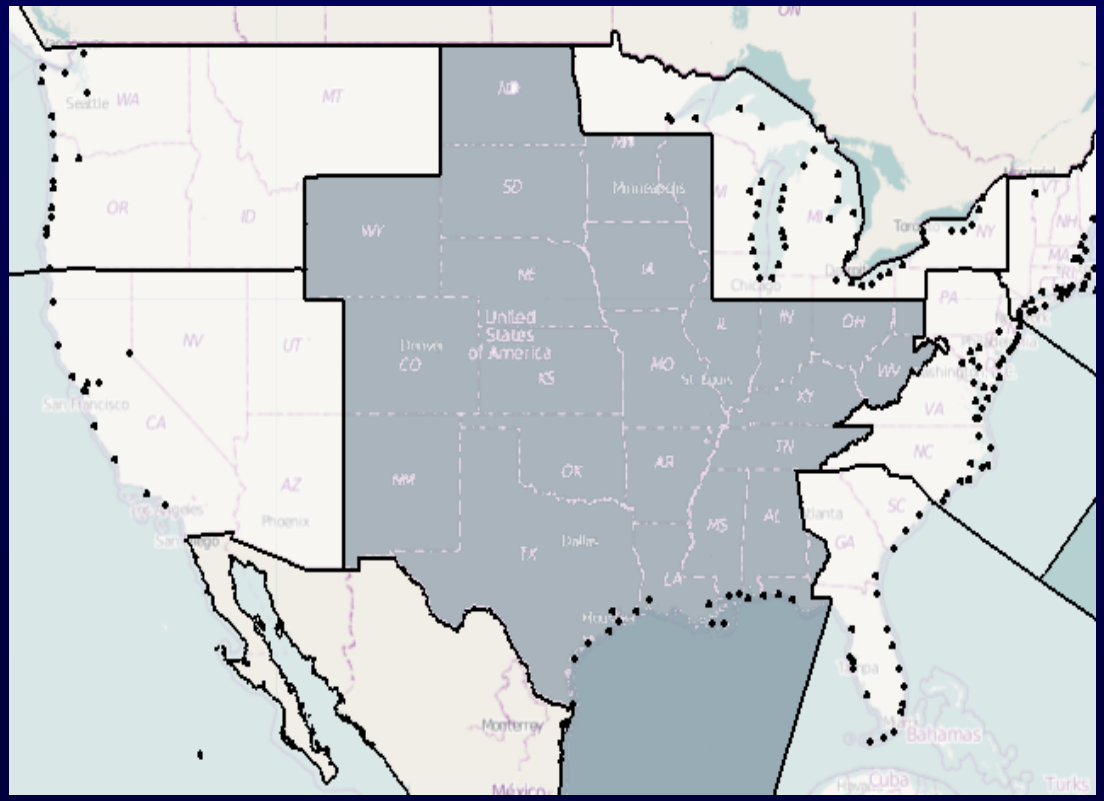




Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



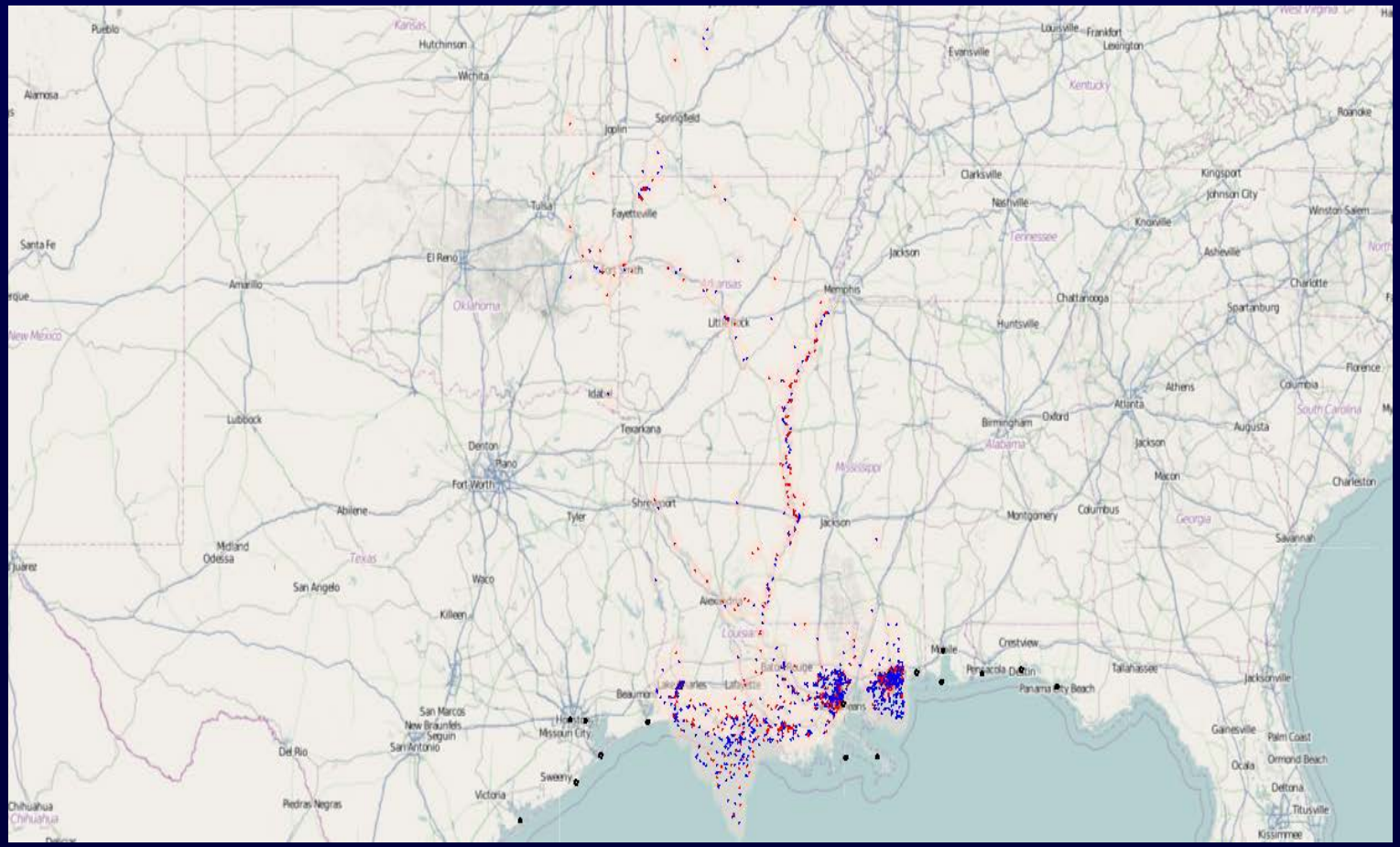


cgSARVA District 8

Mississipi River

D8 SAR Cases:

Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport



Aggregate of SAR Cases 2006-2013

D8 Hazard To Navigation Cases:

Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport



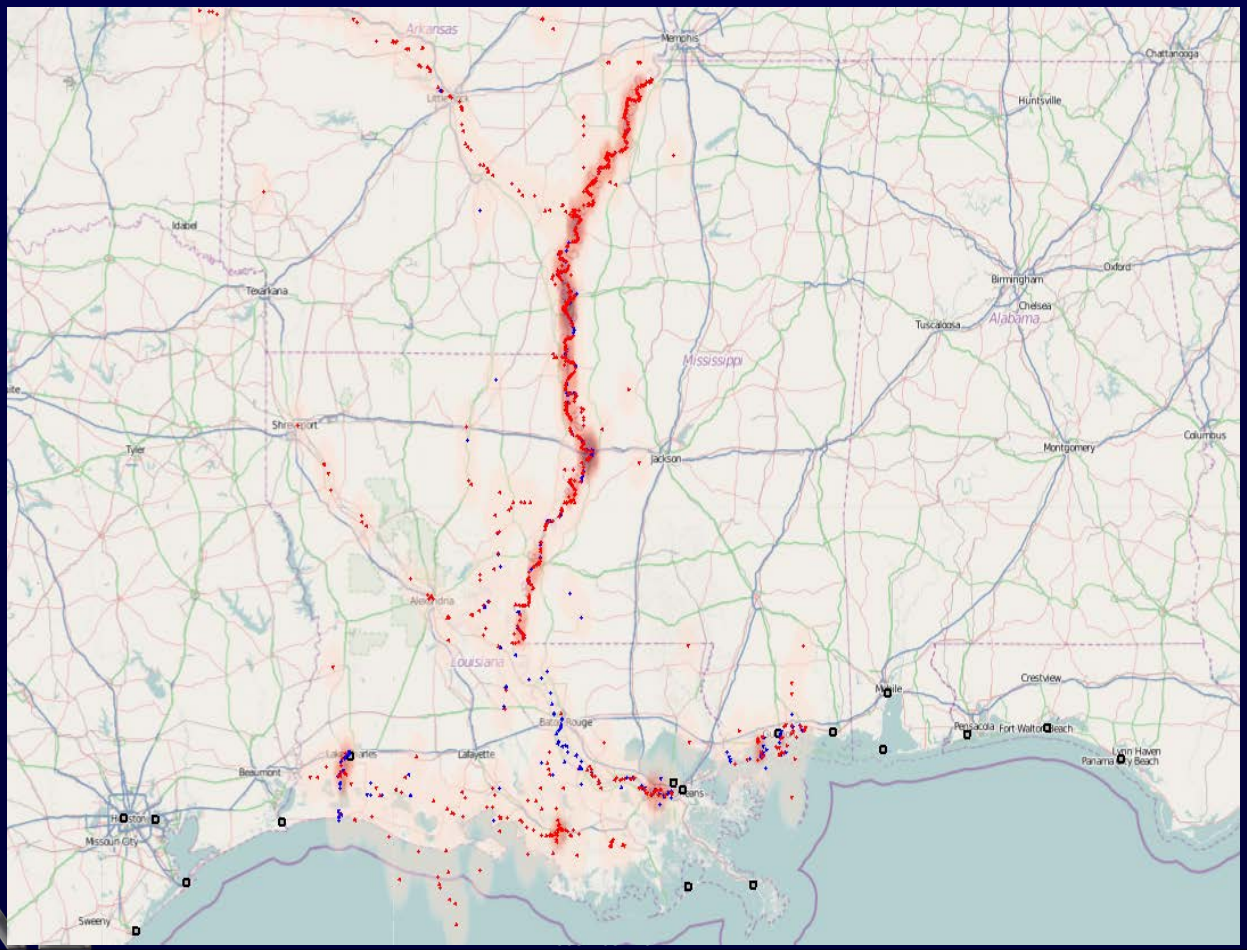
Aggregate of cases 2006-2013

VACCINE

D8 MS Cases:

Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

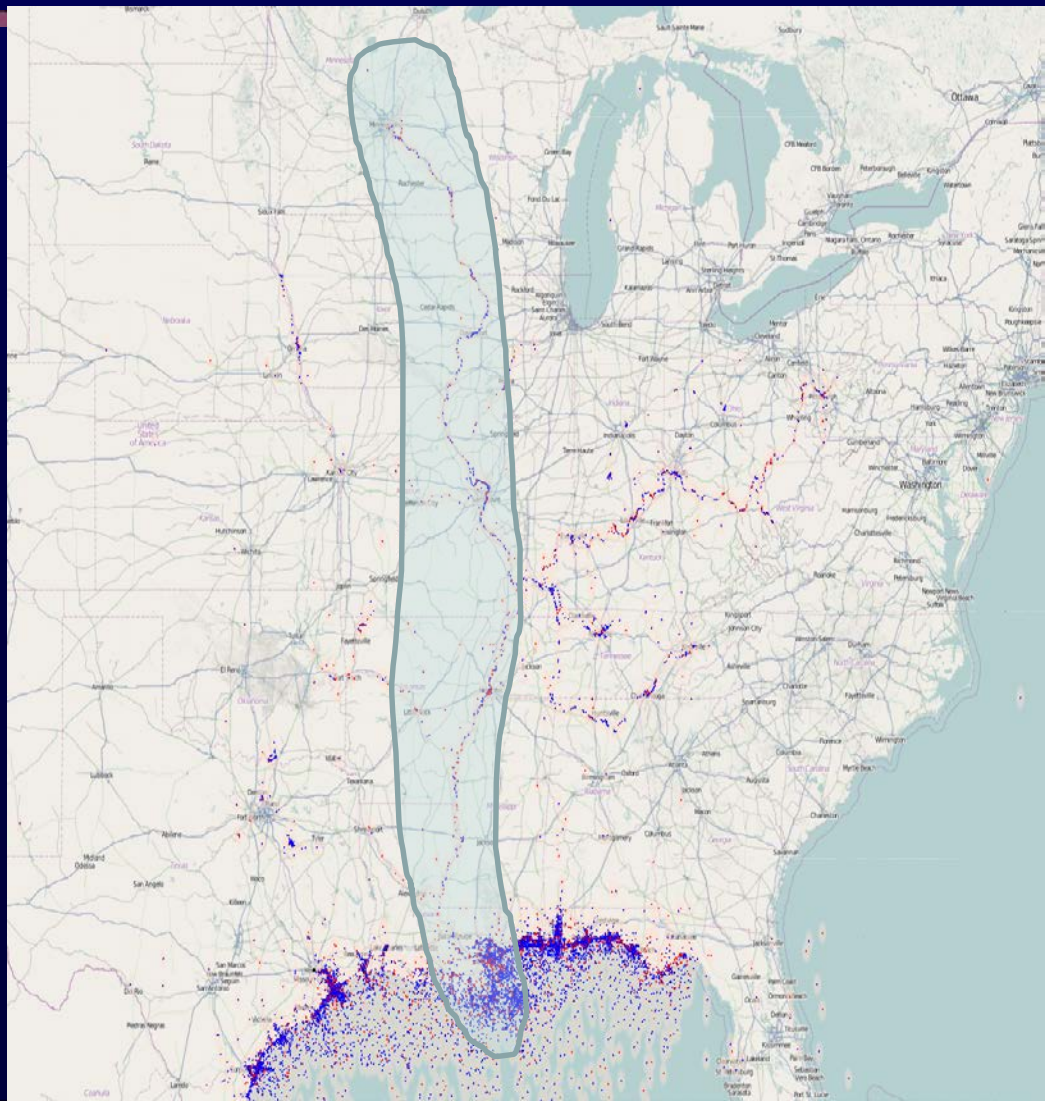
Aggregate of MS Cases 2006-2013. Red indicates cases which status is “open”



VACCIN

SAR Cases -Overview

- All SAR cases near the Mississippi River in District 8.
- 10/2/2006-9/29/2013



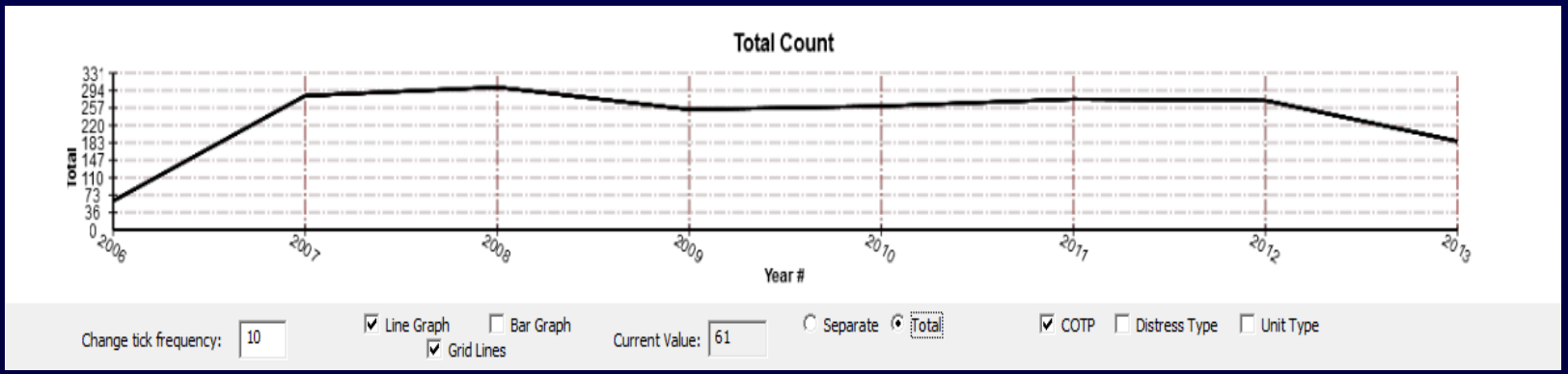
Aggregated Cases nearthe Mississippi River

VACCINE

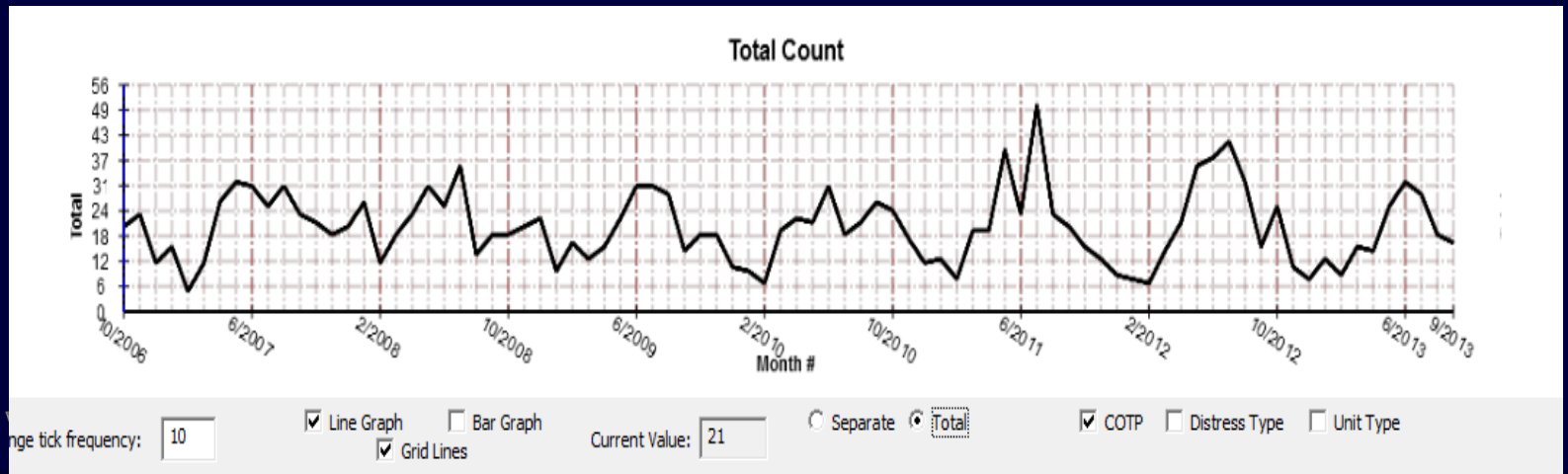
SAR Case Distribution in D8

Abandoned/Derelict Adrift (Unmanned) Aground Aircraft Emergency
Anchored [Unmanned] Beset by Weather Bridge Jumper Capsized Vessel Collision
Disabled Vessel Disoriented Vessel
Distress Alert - situation unknown Diving Accident Fire Hazard to Navigation
MAYDAY Broadcast MEDEVAC MEDICO Non-Maritime EMS Transport
Overdue Person (Non-Martime) Overdue Vessel Person in Water (Ice)
Person in Water (PIW) Special Operation-SAR Stranded (on island)
Taking on Water (TOW) Uncorrelated MAYDAY Unreported Vessel

Aggregate of SAR Cases 2006-2013 and respective word cloud:



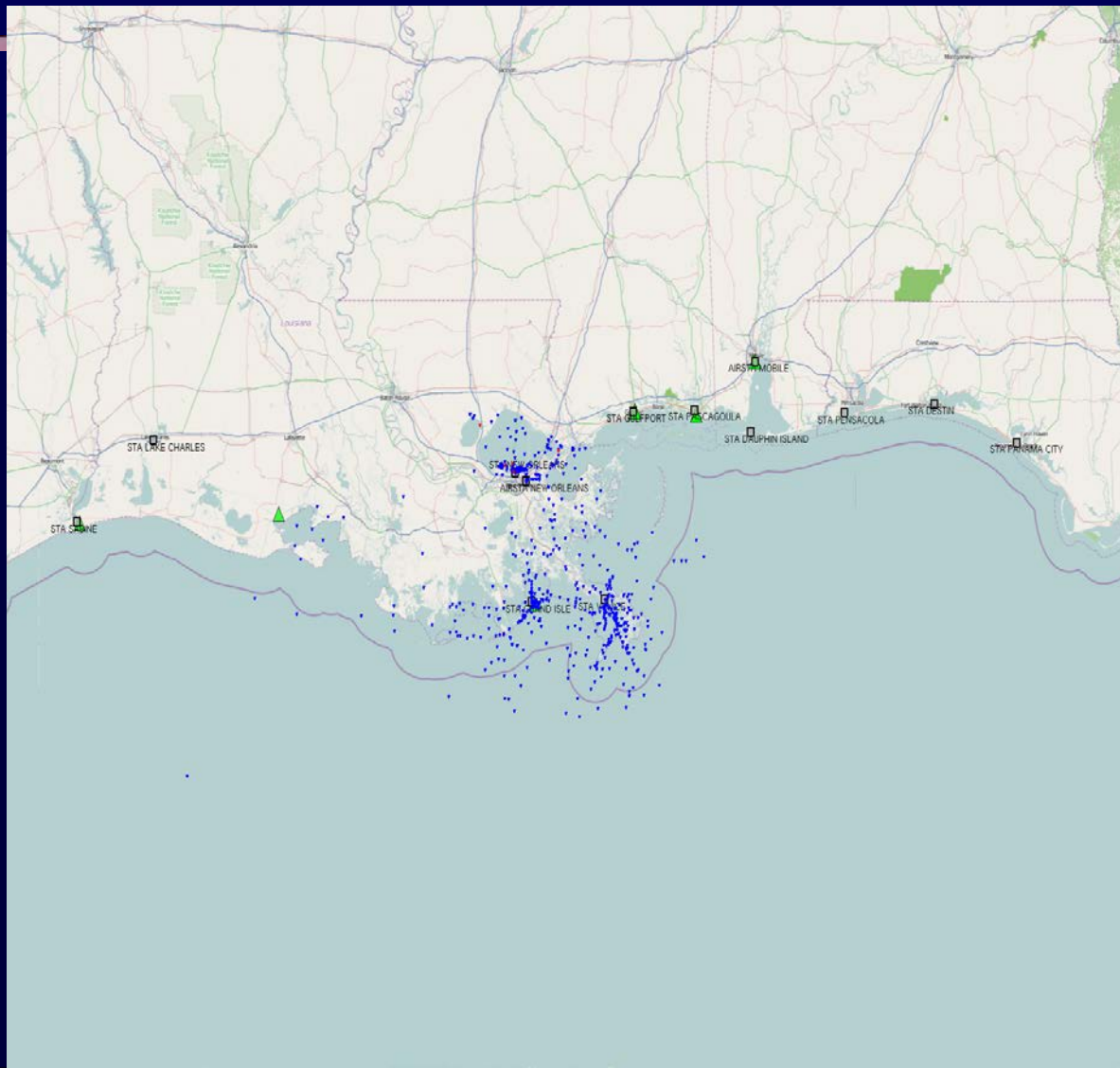
SAR Cases don't follow direct seasonal pattern:



SAR Data Around The Southeast Portion Of MS River

- Three D8 stations
  - STA GRAND ISLE
  - STA NEW ORLEANS
  - STA VENICE
- 10/2/2006--9/29/2013

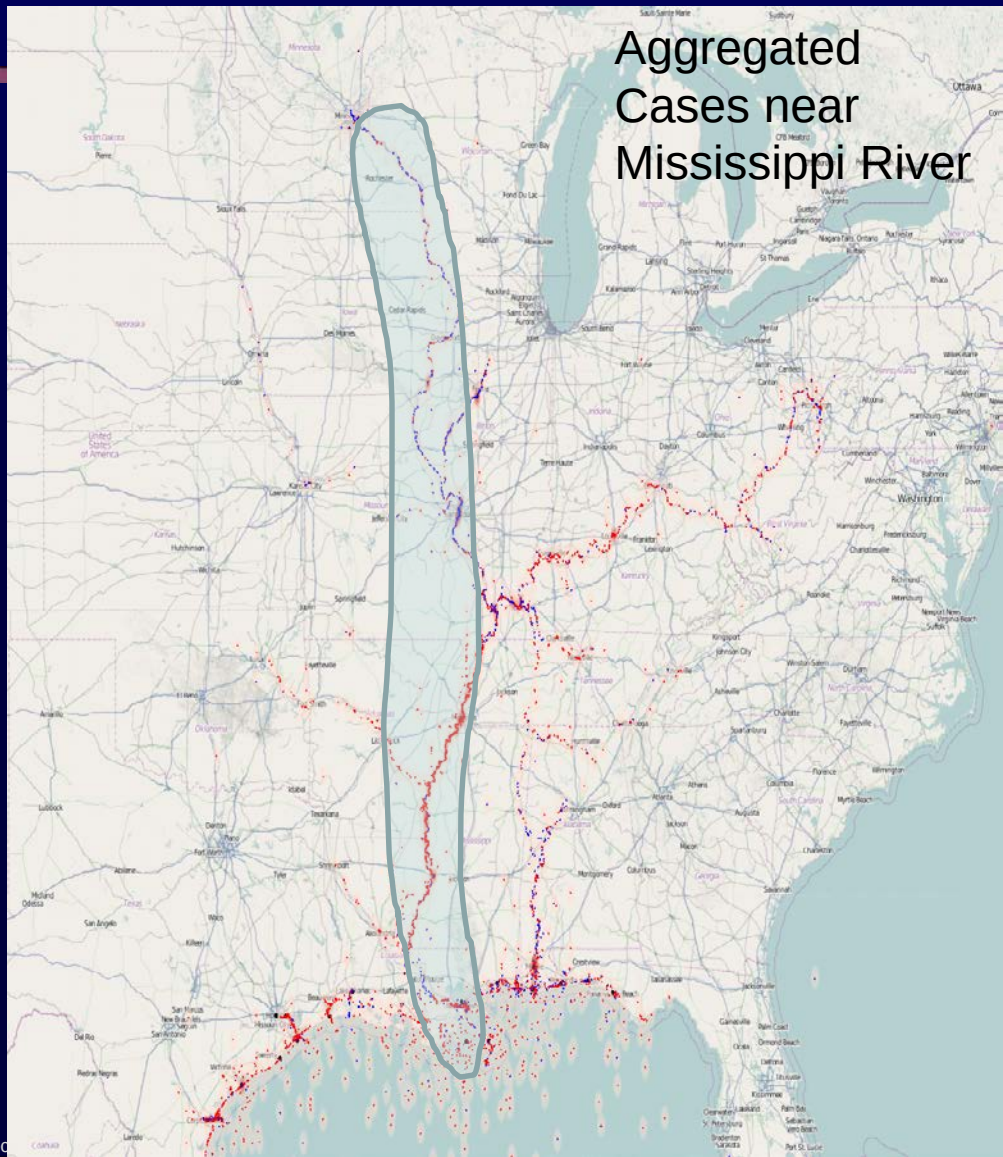
Abandoned/Derelict Adrift (Unmanned) Aground
Beset by Weather Bridge Jumper Capsized Vessel Collision
Disabled Vessel
Disoriented Vessel
Distress Alert - situation unknown Diving Accident Fire
MEDEVAC MEDICO Overdue Vessel Person in Water (PIW)
Stranded (on island) Taking on Water (TOW)
Uncorrelated MAYDAY Unreported Vessel



MS Cases-Overview

- Marine Safety
- MS taken from MISLE column: “Case Distress Class”
- 10/2/2006-9/29/2013
- MS Cases include the following:

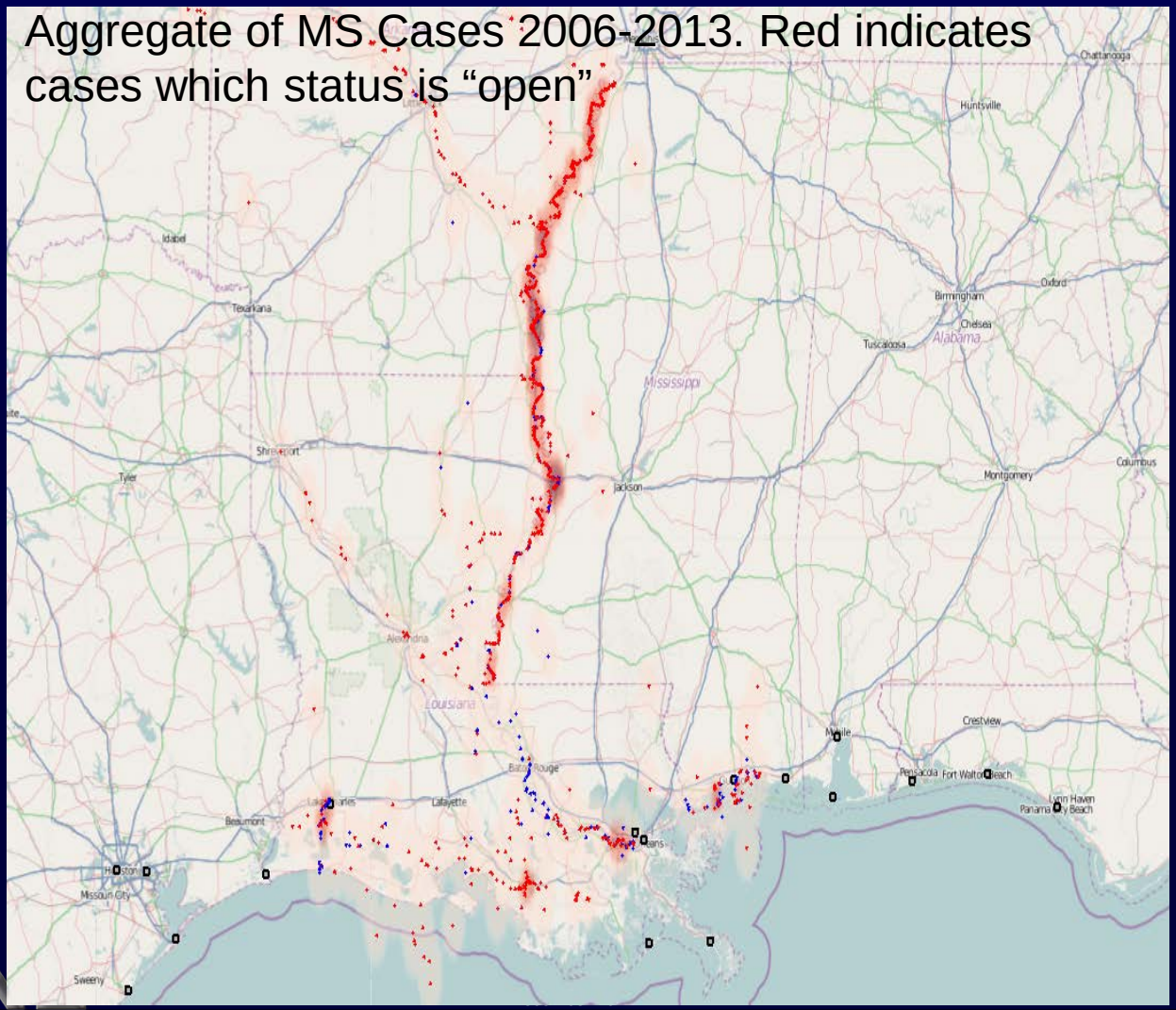
  - Alleged Violation of Law/Regulation
  - Loss of Life/Injury
  - Equipment Failure
  - Fire
  - Allision
  - Explosion
  - Grounding
  - Hazardous Vessel Operation
  - Capsize
  - Hazard to Navigation
  - Collision
  - Force Majeure Claim
  - Breakaway
  - Maritime Explosive Ordnance Incidents
  - Irregular Navigation Incident
  - Waterway Restriction/Closure
  - Near-miss Situation
  - Special Operation
  - Heavy Weather Damage
  - Bridge Closures
  - Vsl Inspection/PSC
  - Cargo/Container Inspection
  - Facility Inspection
  - Towing Vsl Exam
  - Cargo Transfer Monitor



VACCINE

D8 MS Cases:

Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport

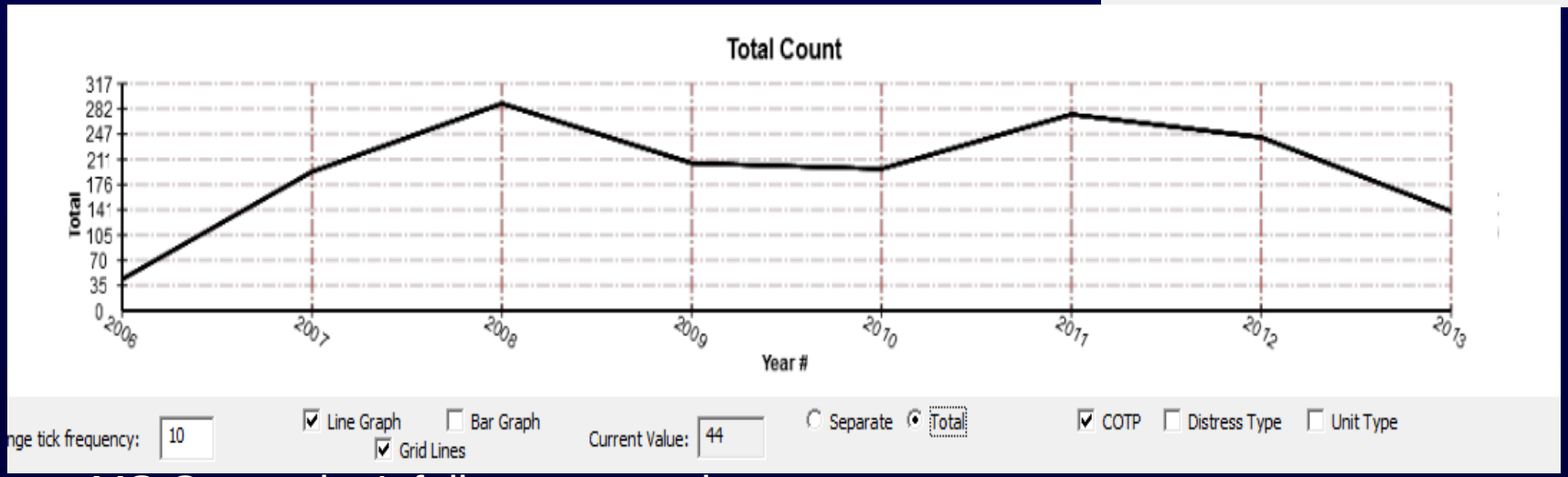


VACCIN

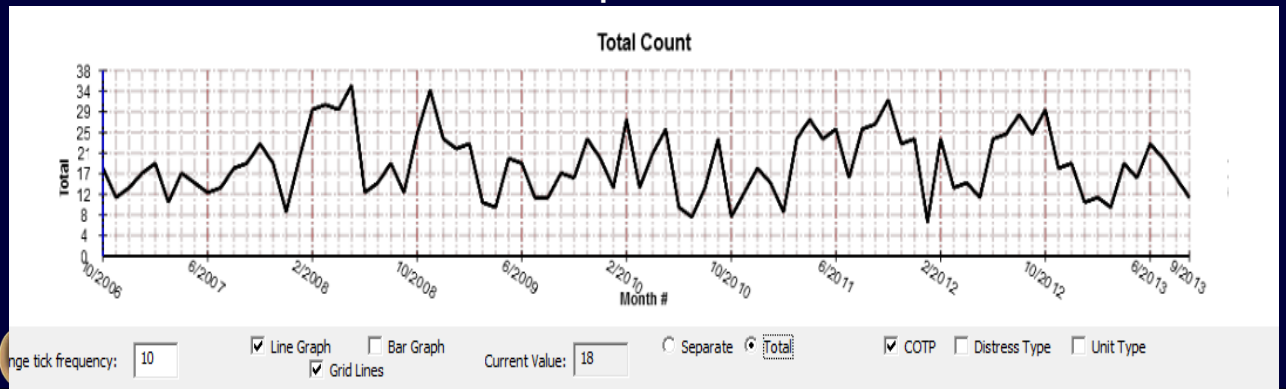
MS Case Distribution (in D8)



Aggregate of MS Cases 2006-2013 and respective word cloud:



MS Cases don't follow seasonal patterns:



D8 Hazard To Navigation Cases:

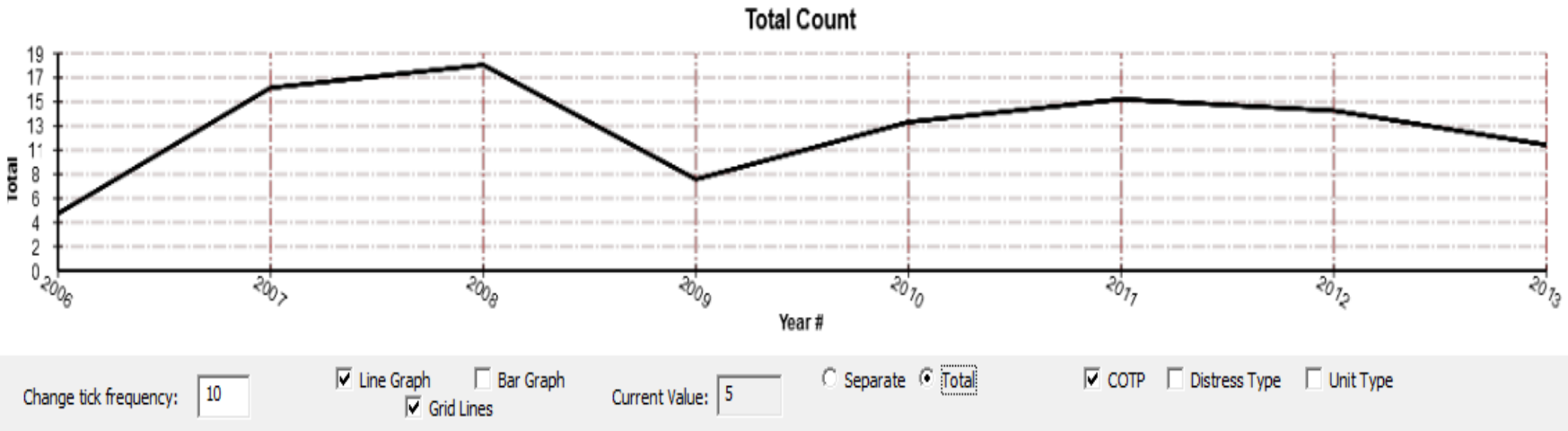
Filtered To Only Show Region Between Sta Lake Charles And Sta Gulfport



Aggregate of cases 2006-2013

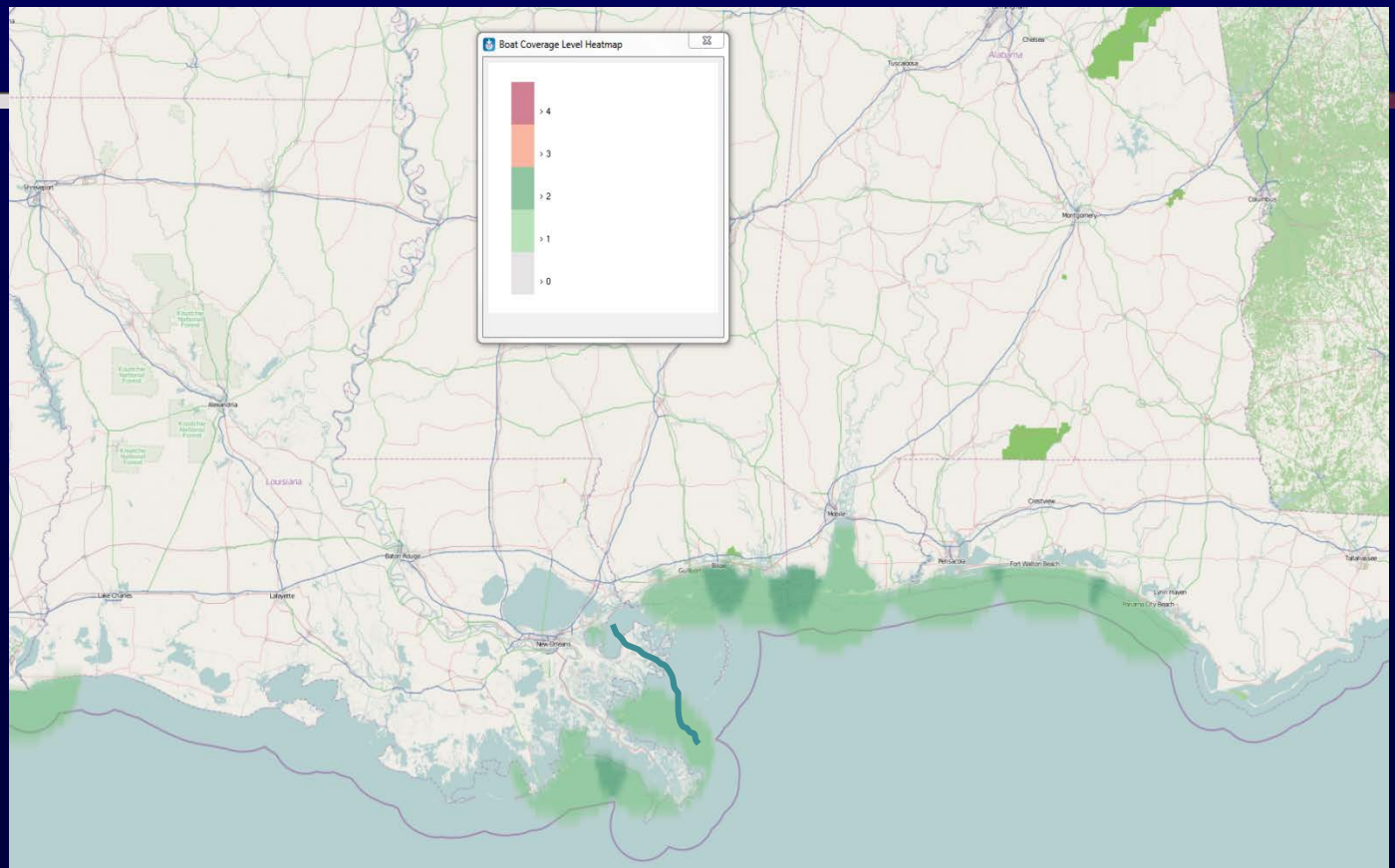
VACCINE

Hazard to Navigation Case Distribution



Utility Visualization

- Boat Coverage Map
  - STA GRAND ISLE
  - STA VENICE

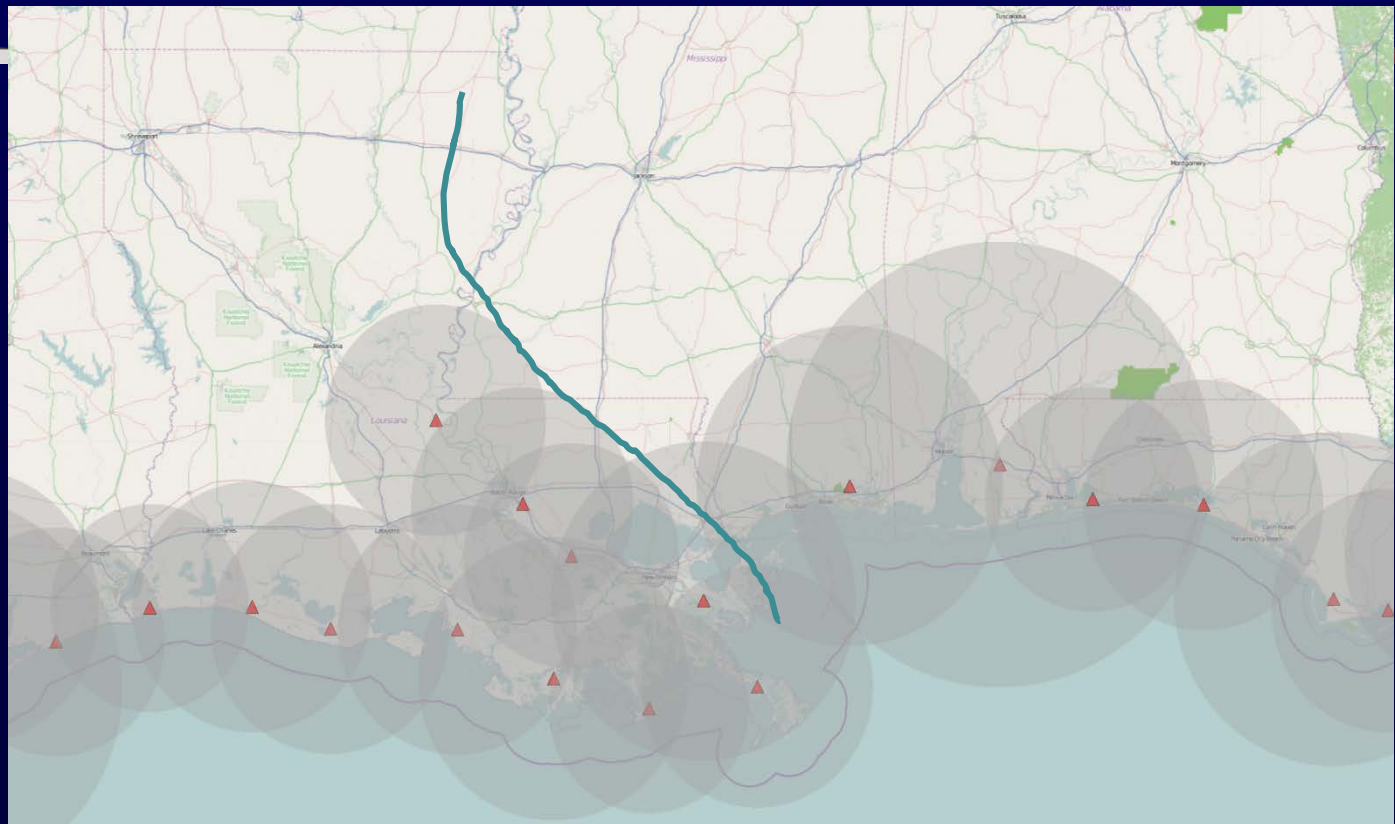


Covered Portion of Mississippi River

VACCINE

Utility Visualization

- Rescue21 Towers in District 8

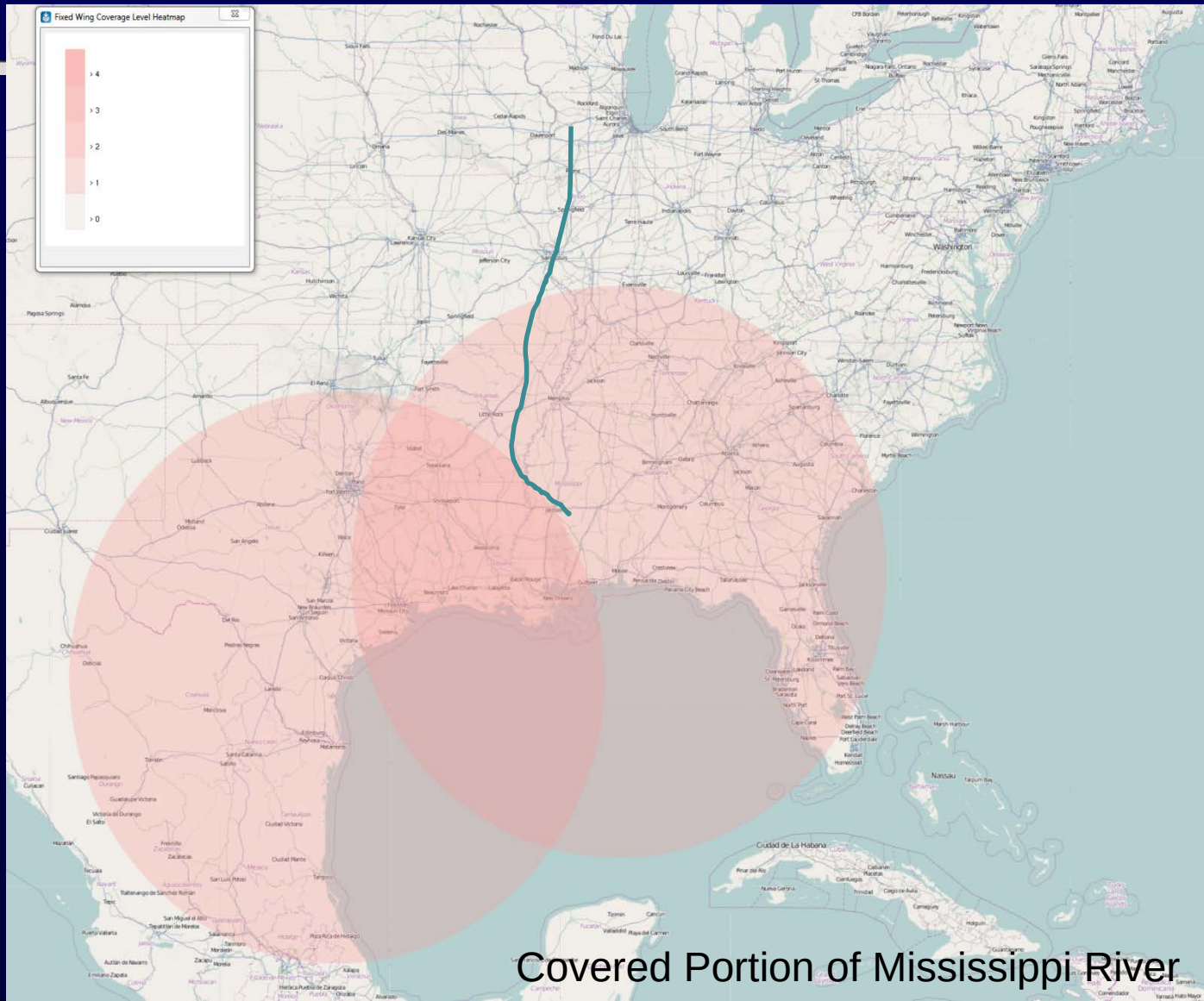


Covered Portionof Mississippi River

VACCINE

Utility Visualization

- FixedWing Coverage Map
  - AIRSTA MOBILE
  - AIRSTA CORPUS CHRISTI

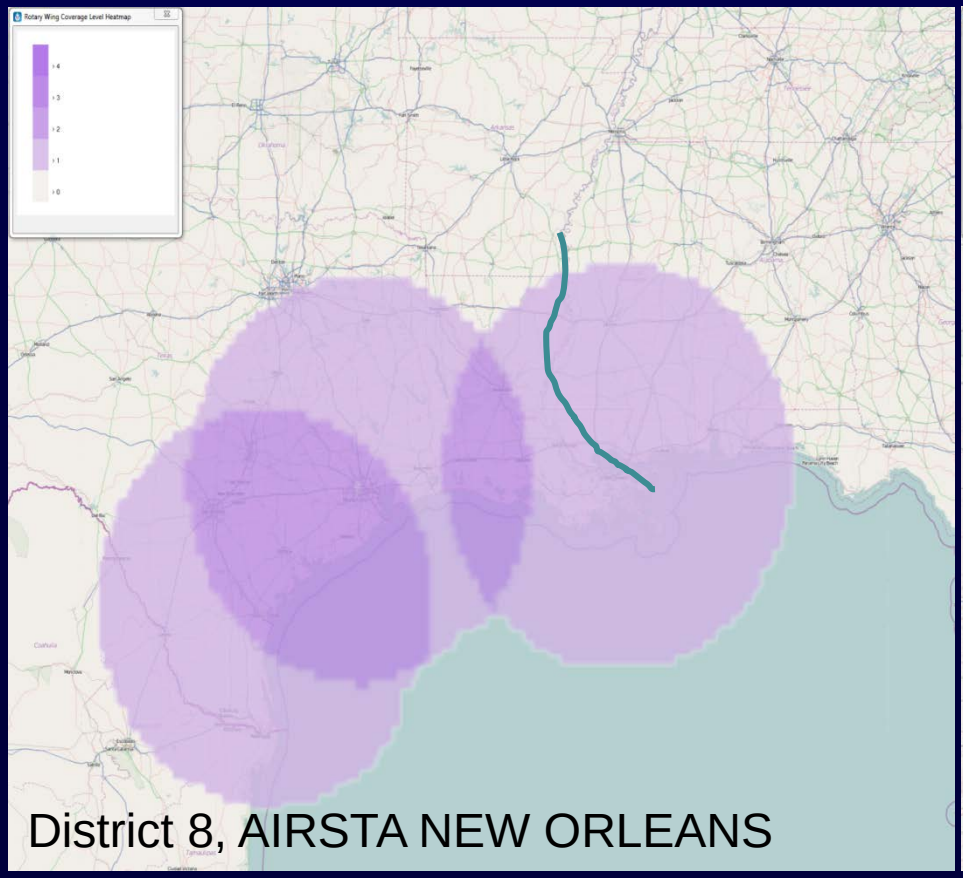


Covered Portion of Mississippi River

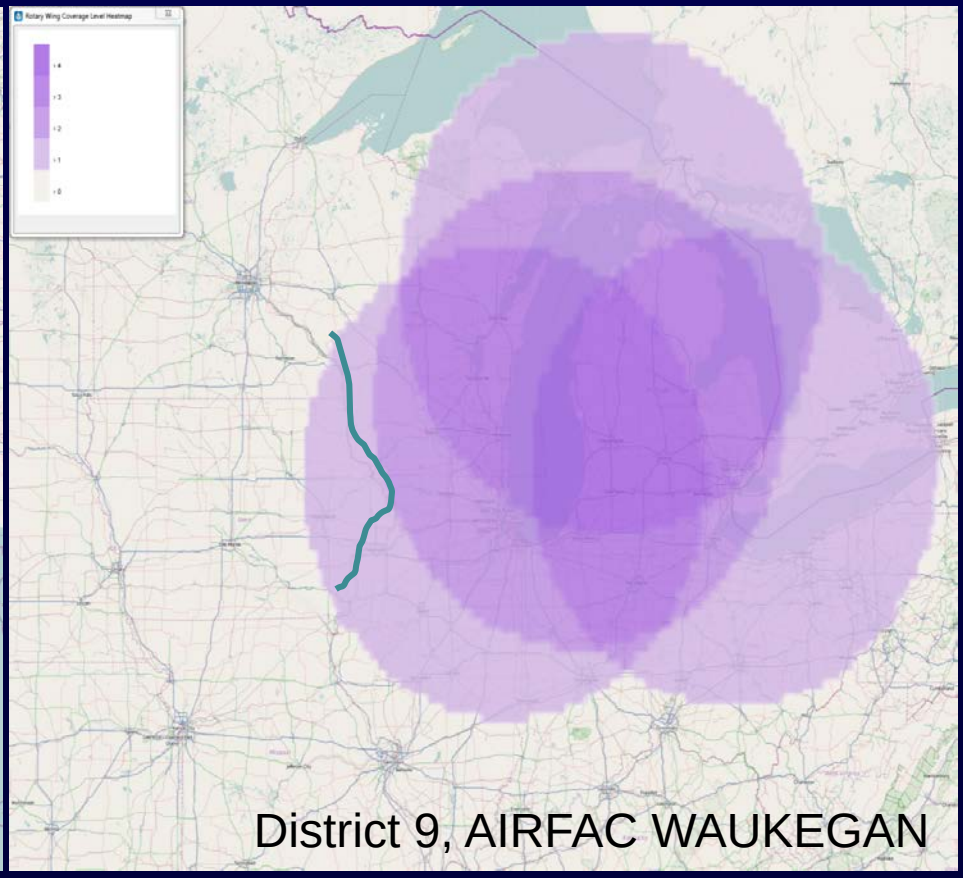
VACCINE

Utility Visualization

- Rotary Wing Coverage Map by AirSTA



District 8, AIRSTA NEW ORLEANS



District 9, AIRFAC WAUKEGAN

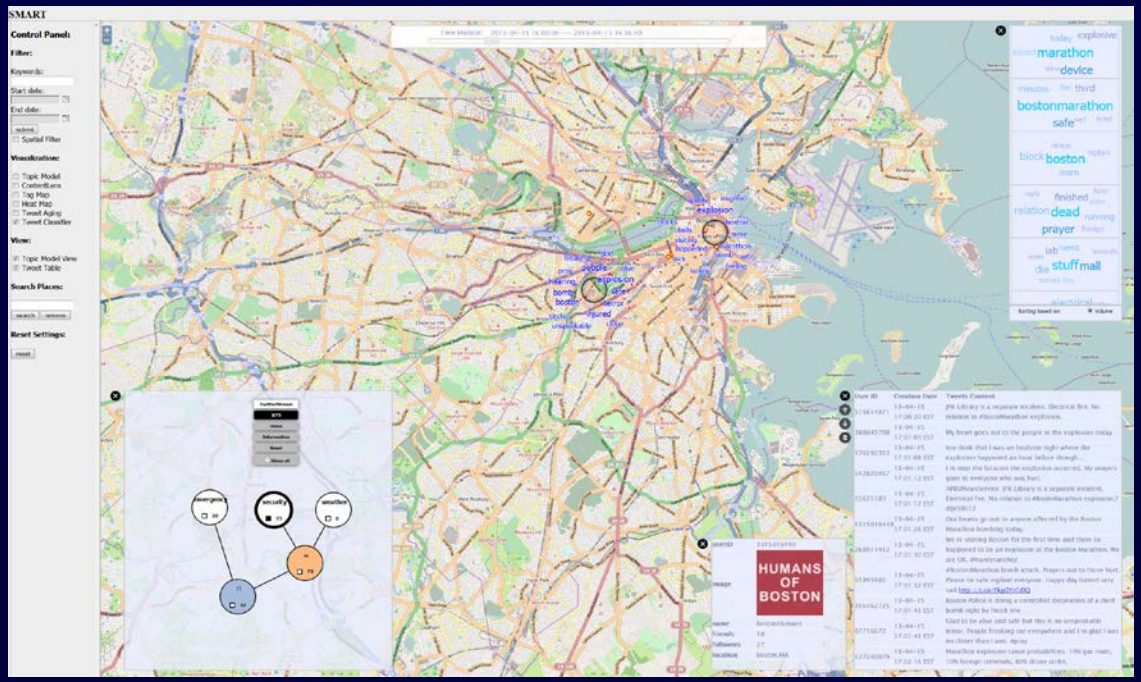
Covered Portion of Mississippi River

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

December 2014

Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting

Time: around 10:40 am PST
(01:40 pm EST), 10/24/2014

Earliest Tweets About The Event

601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances
461749842	14-10-24 13:55:45 EST	Cops are fucking EVERYWHERE right now
461749842	14-10-24 13:55:21 EST	Cops are fucking EVWRYWHERE right now
707981743	14-10-24 13:55:12 EST	There is a shooting at mp?
818397241	14-10-24 13:54:06 EST	@Miguel_Rocha34 do you have any idea who the person with the gun was
818397241	14-10-24 13:52:12 EST	you got some serious issues that you need to seek help for if you shoot up a fucking high school
601648236	14-10-24 13:49:11 EST	@Ray_Raaaay21 one of the kids in the lunchroom started shooting and everyone booked it
601648236	14-10-24 13:46:17 EST	shooting at mp

707981743	14-10-24 14:17:43 EST	2 people are dead...
335778696	14-10-24 14:17:26 EST	Wtf is going on?? Shooting at mp?
70255303	14-10-24 14:16:37 EST	Shooting at Marysville Pilchuck. Praying for the kids and the families.
1853421522	14-10-24 14:14:39 EST	@Lake_stonerrr a student shot several students, possible 2 students dead 6 injured
1099441382	14-10-24 14:10:38 EST	Shooting at pikchuck???
599963823	14-10-24 14:08:26 EST	Something's fucking wrong !!!!! Or my mom wouldn't be crying about the MP shooting !!!
130018947	14-10-24 14:07:51 EST	About 5 mins away from the scene. Sad. Another school shooting. #MPHS
601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances

Some online news feed:
http://www.foxnews.com/us/2014/10/24/police-respond-to-shooting-at-washington-state-high-school/
http://www.cnn.com/2014/10/24/us/washington-school-shooting/index.html

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

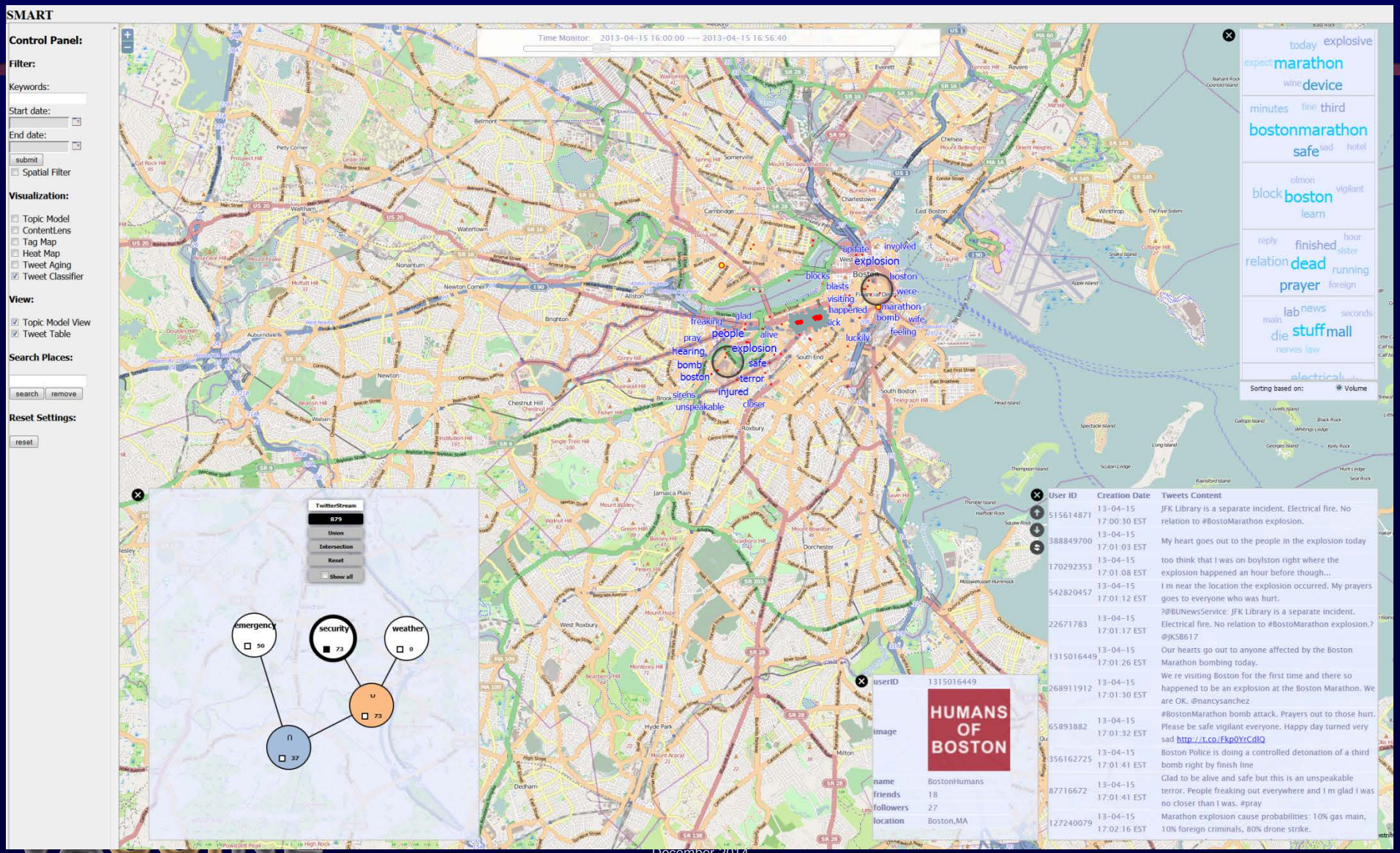
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)
- Ohio State Football – more successful than commercial tools

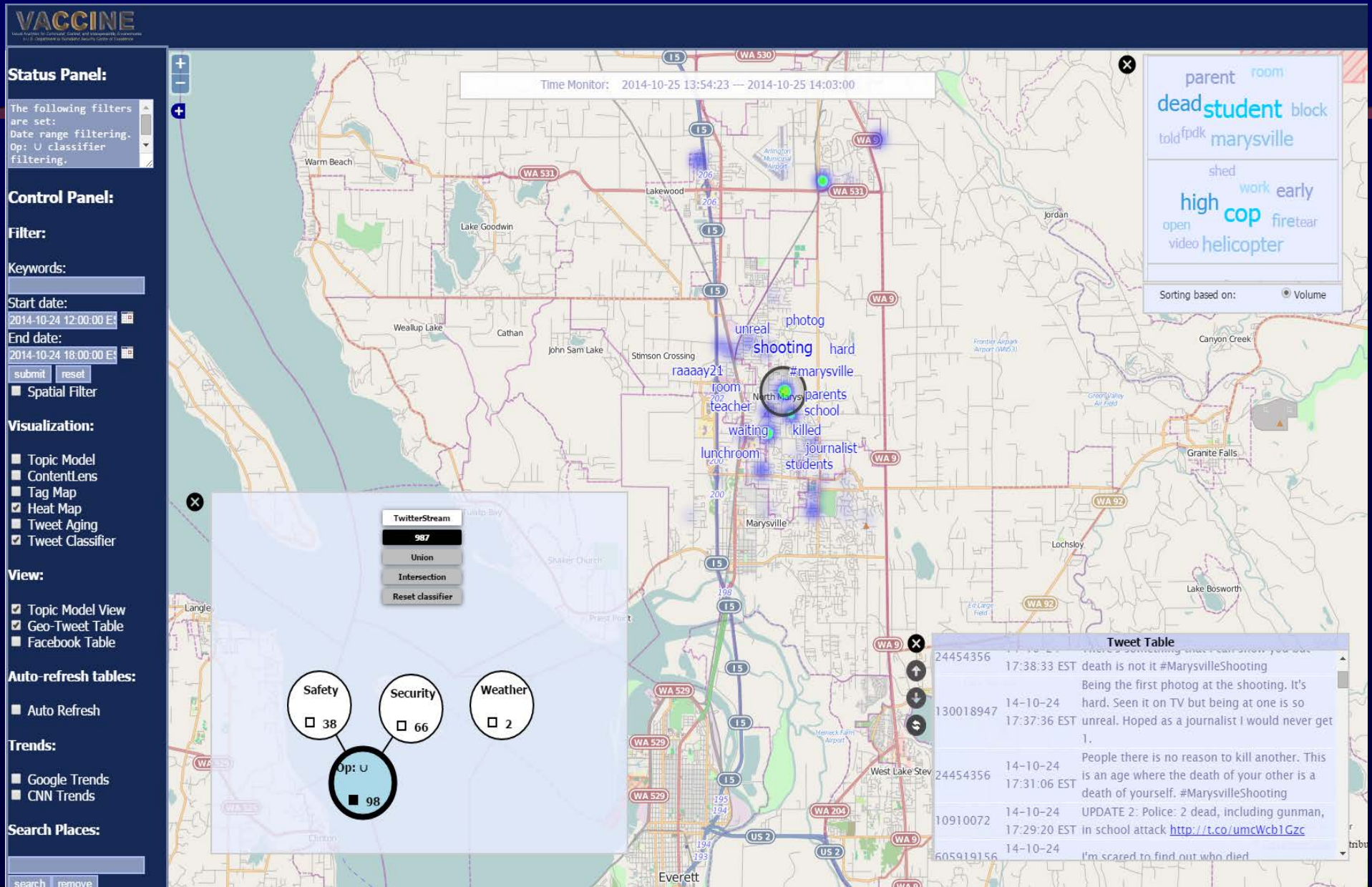


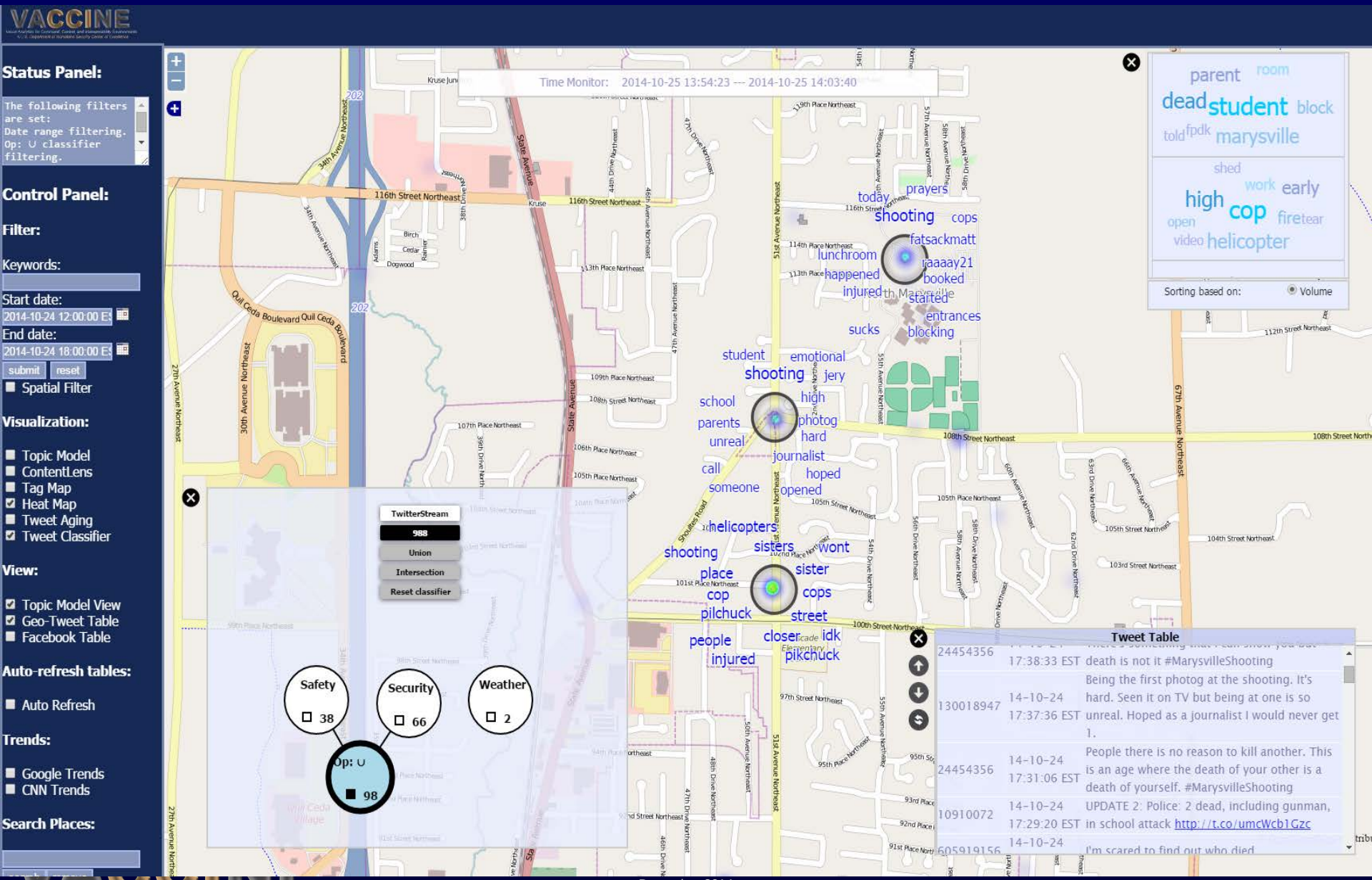
VACCINE

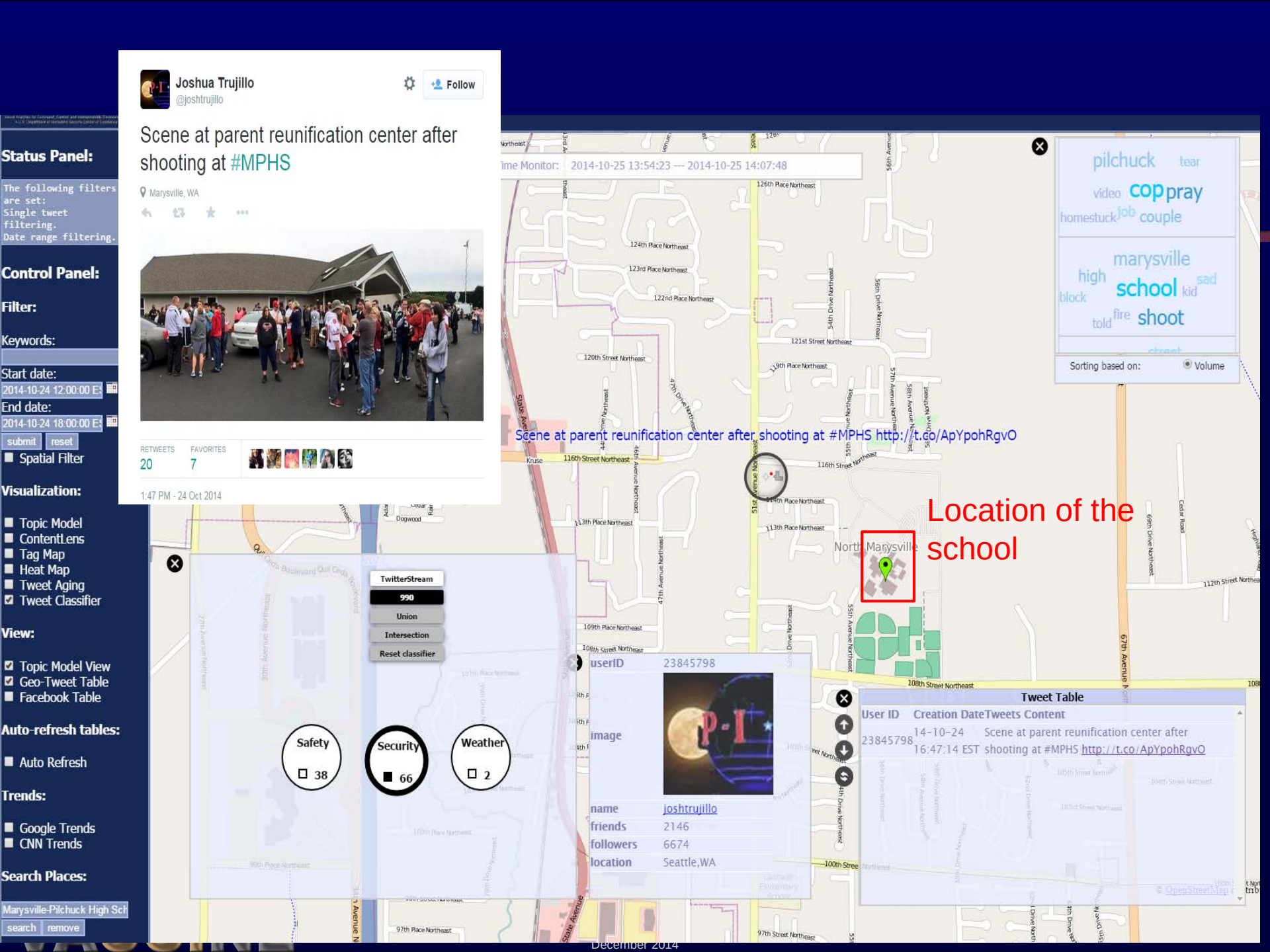
Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
jaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24
17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24
17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24
17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died


Joshua Trujillo
@joshtrujillo
Follow
Scene at parent reunification center after shooting at #MPHS
Marysville, WA
RETWEETS
20
FAVORITES
7
1:47 PM - 24 Oct 2014
Status Panel:
The following filters are set:
Single tweet filtering.
Date range filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 ES
End date:
2014-10-24 18:00:00 ES
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Marysville-Pilchuck High Sch
search
remove
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48
pilchuck
tear
video
cop
pray
homestuck
job
couple
marysville
high
school
kid
sad
block
told
fire
shoot
Sorting based on:
Volume
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO
Location of the school
North Marysville
TwitterStream
990
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
userID
23845798
image
name
joshtrujillo
friends
2146
followers
6674
location
Seattle,WA
Tweet Table
User ID
Creation Date
Tweets Content
23845798
14-10-24 16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

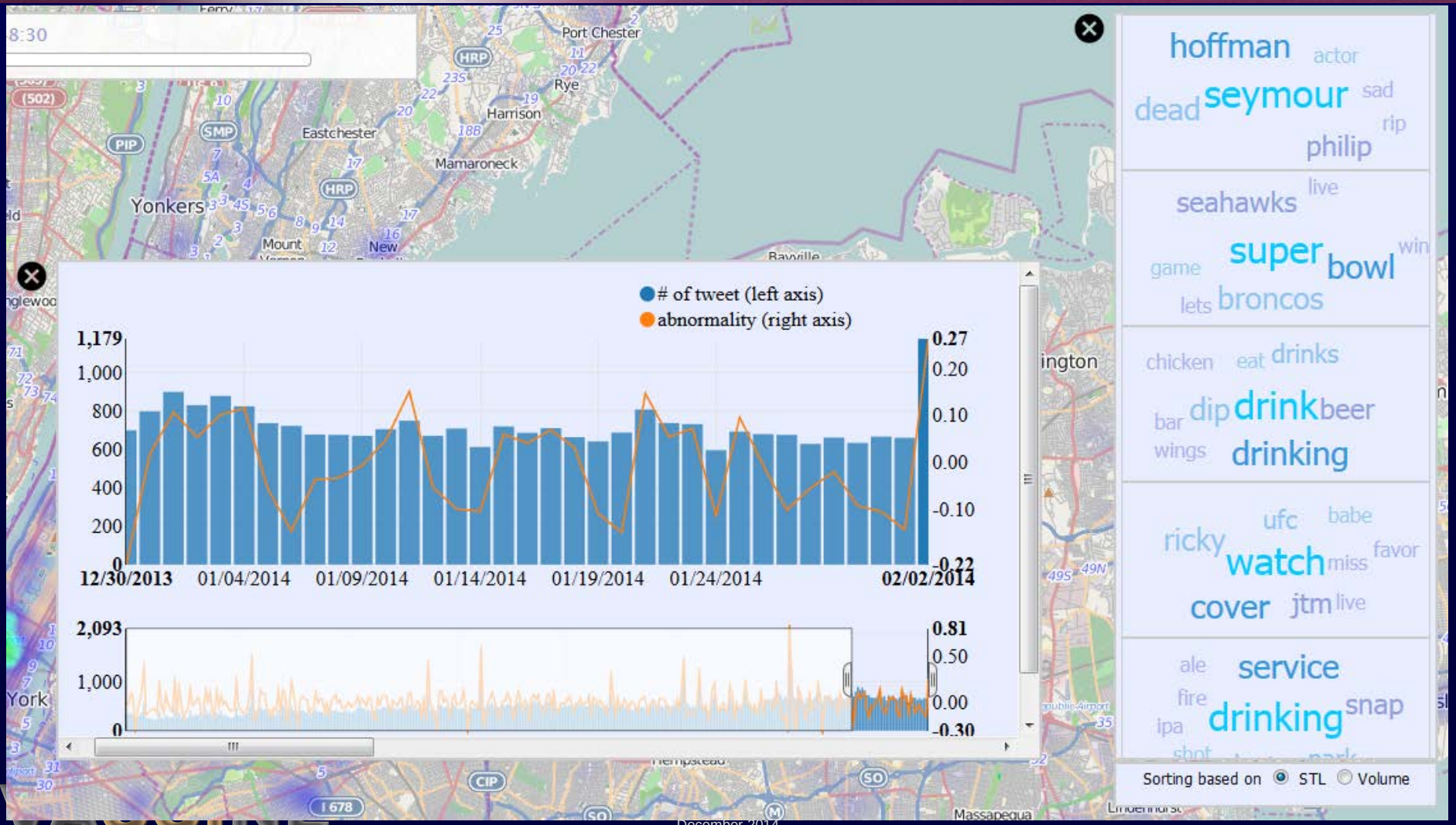
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring event
  - Purdue football games – the p
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – L



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

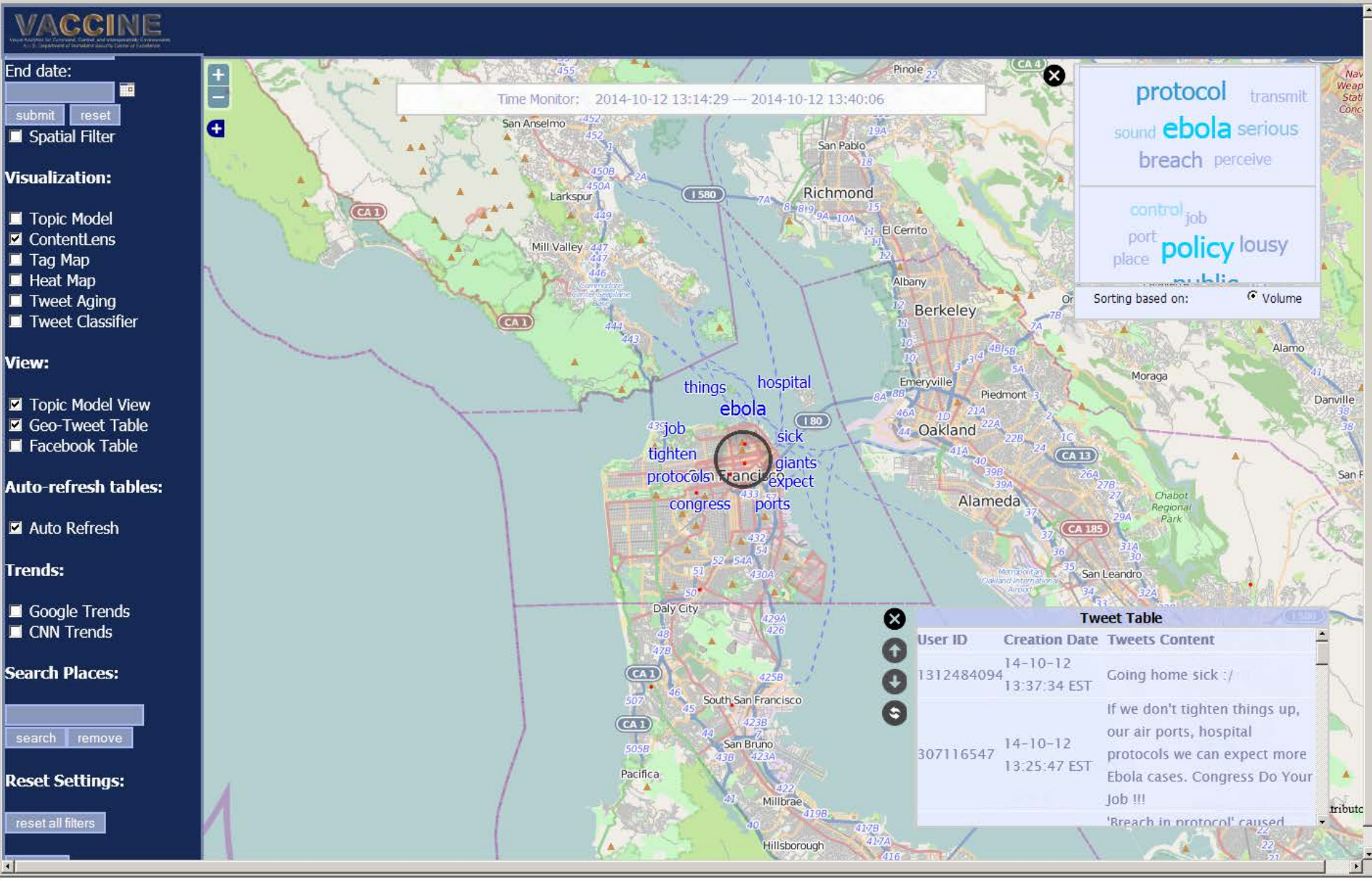


RETWEETS 6

FAVORITES 4

12:56 PM - 12 Oct 2014

Ebola tweets during Fleet Week 10/12/14

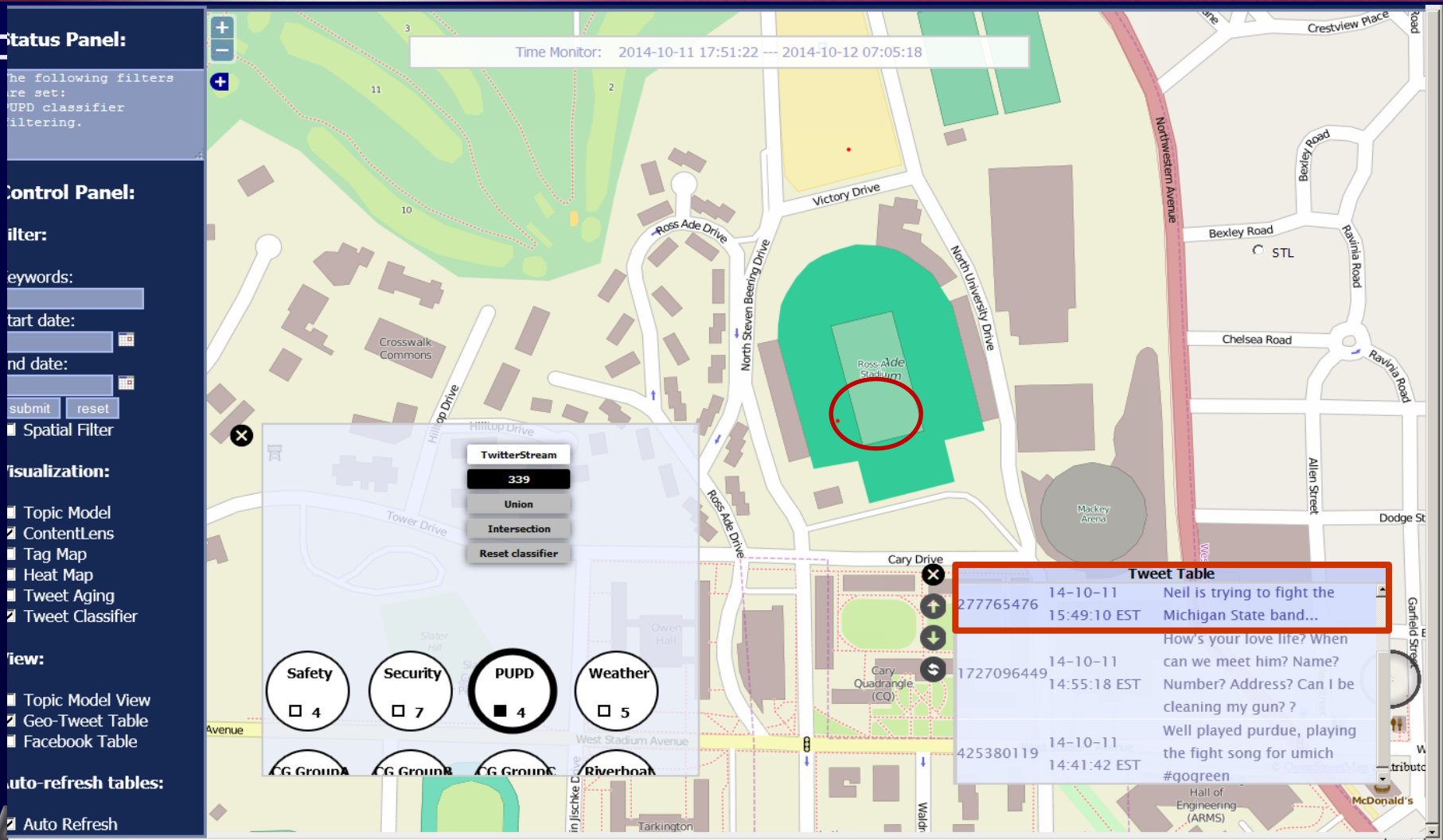


Where Can It Be Helpful?

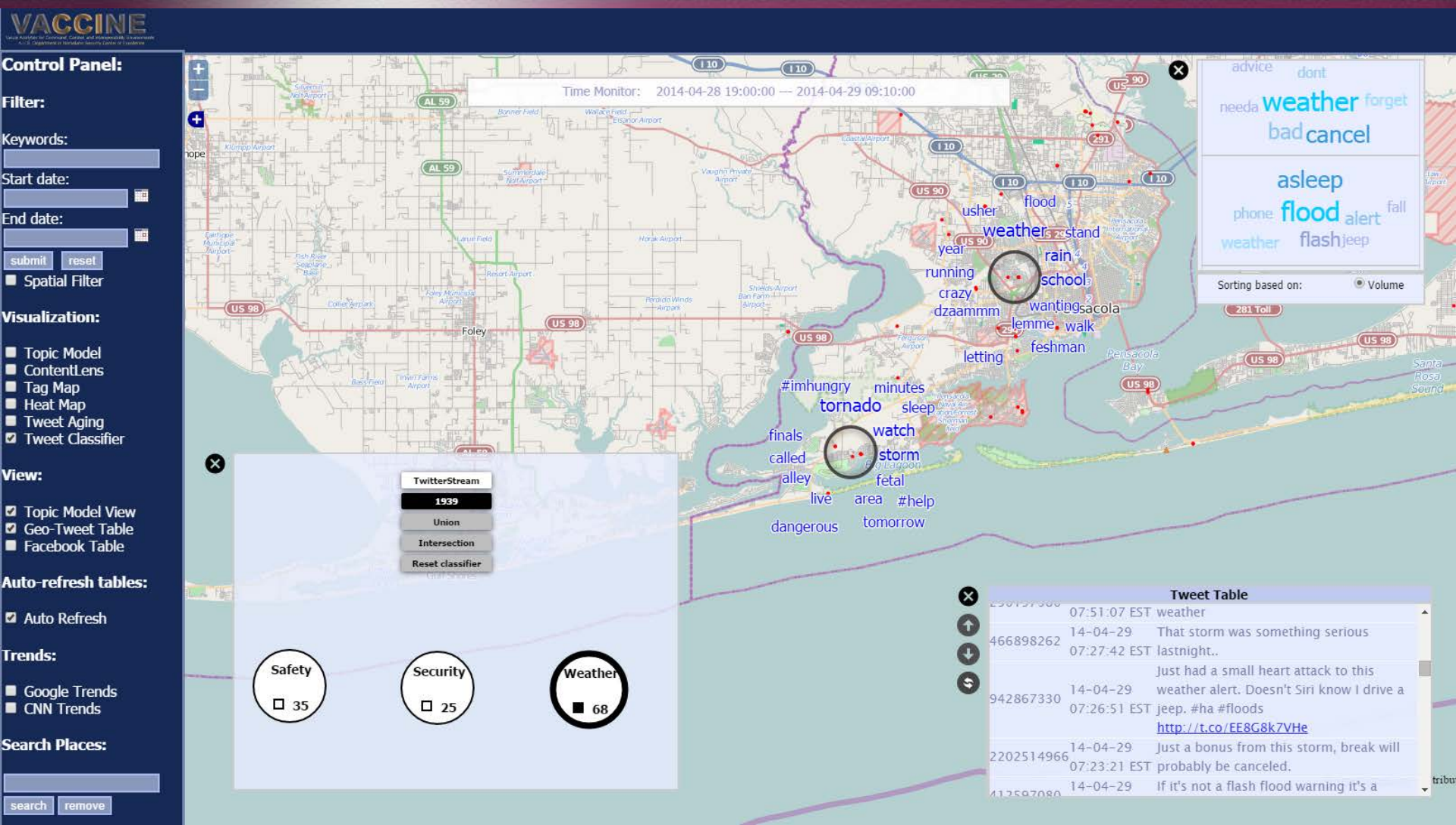
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



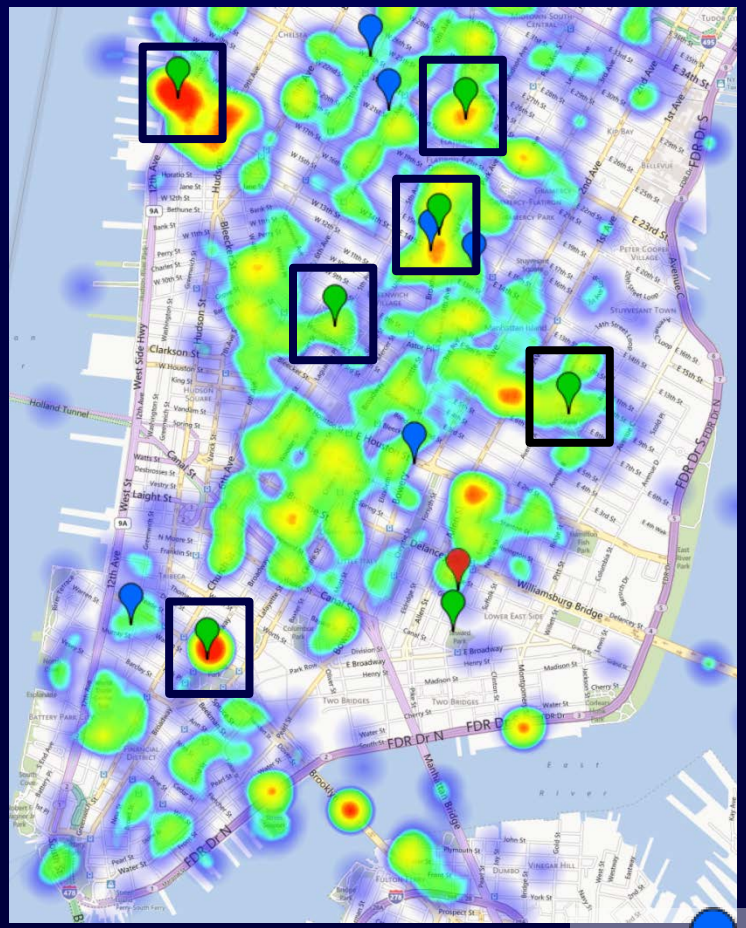
Where Can It Be Helpful?

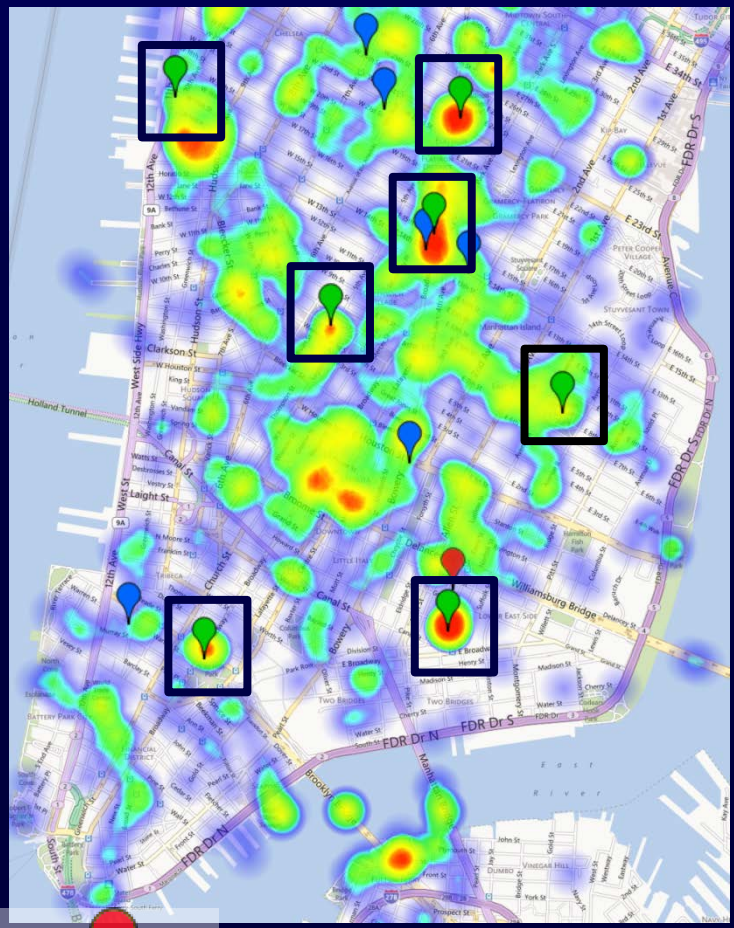


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

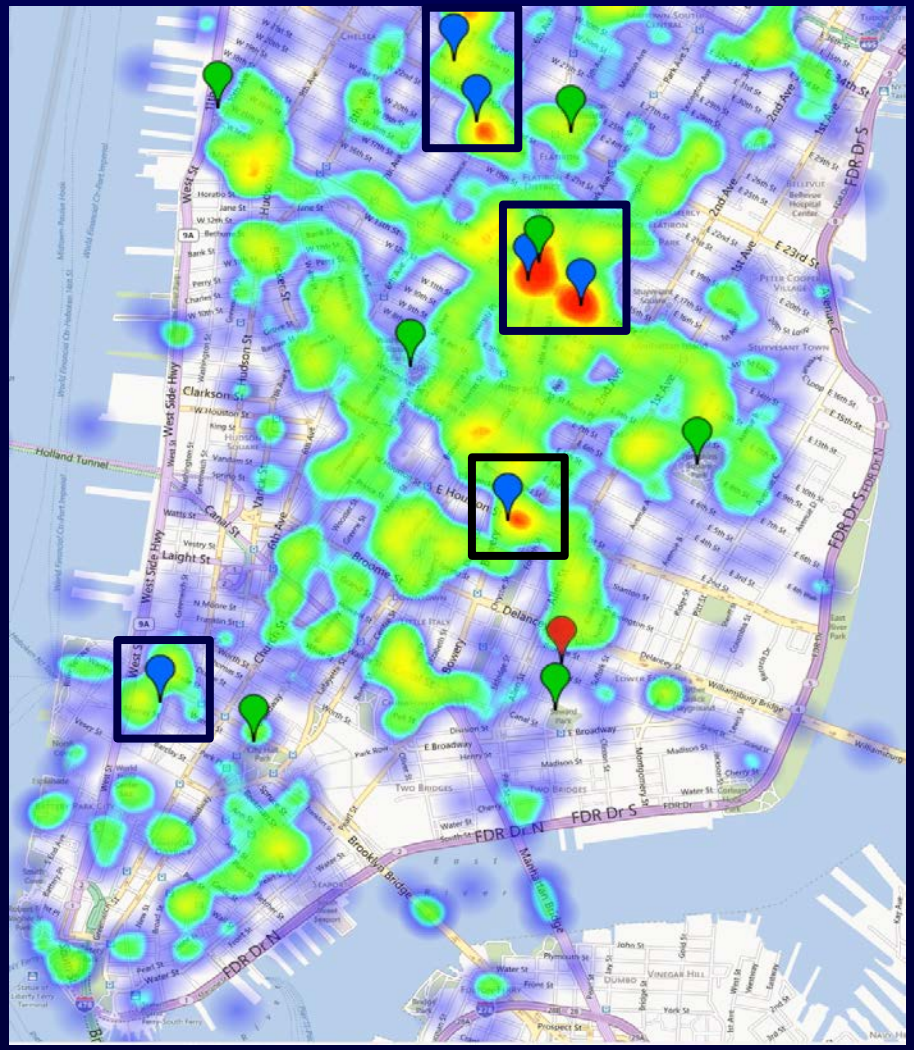
December 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

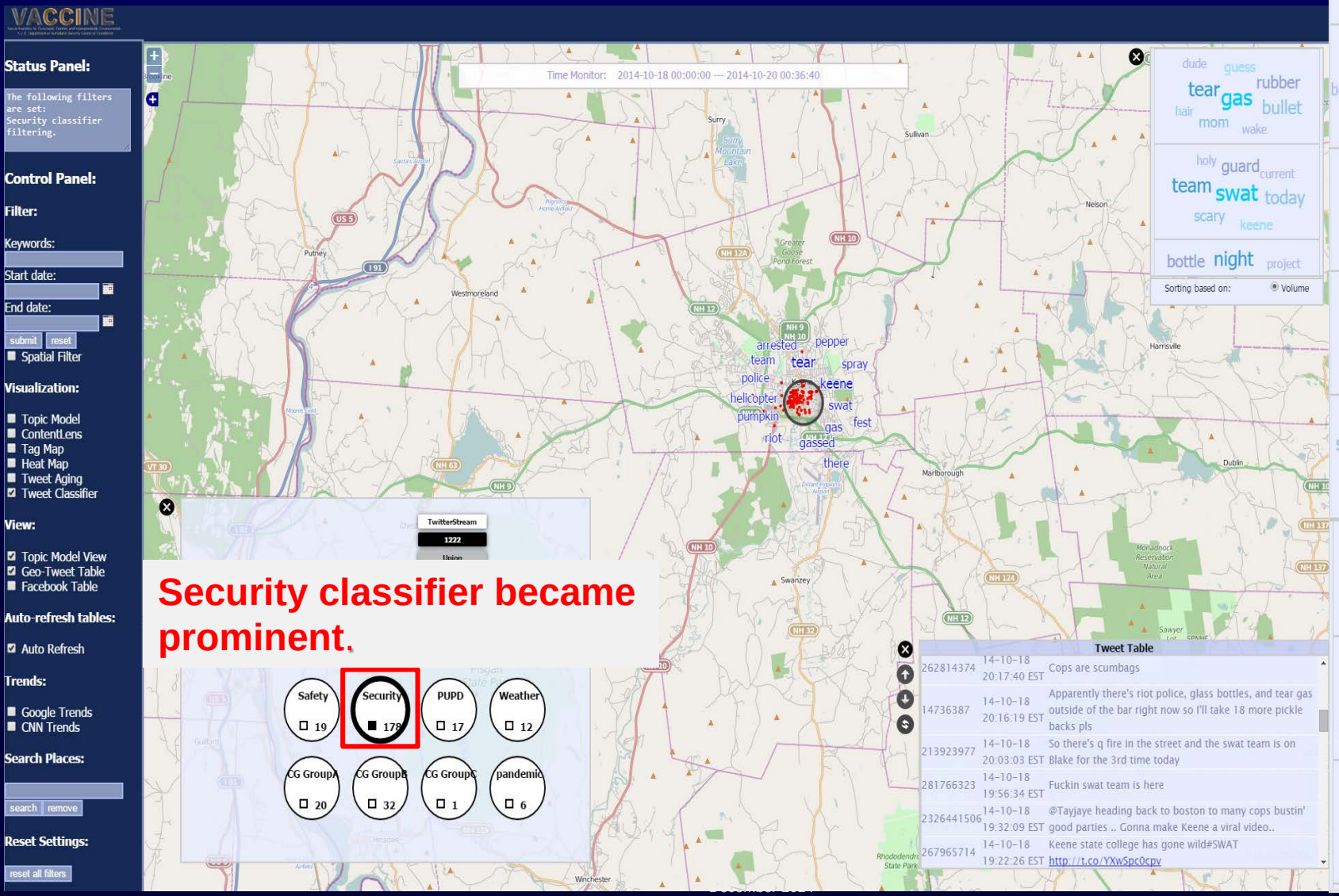
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



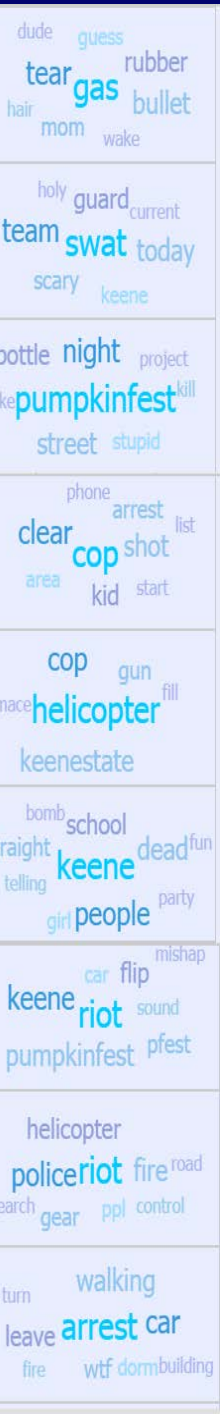
VACCINE

Keene, NH Pumpkin festival 10/18/14 – 10/19/14 Overview

Major topics:

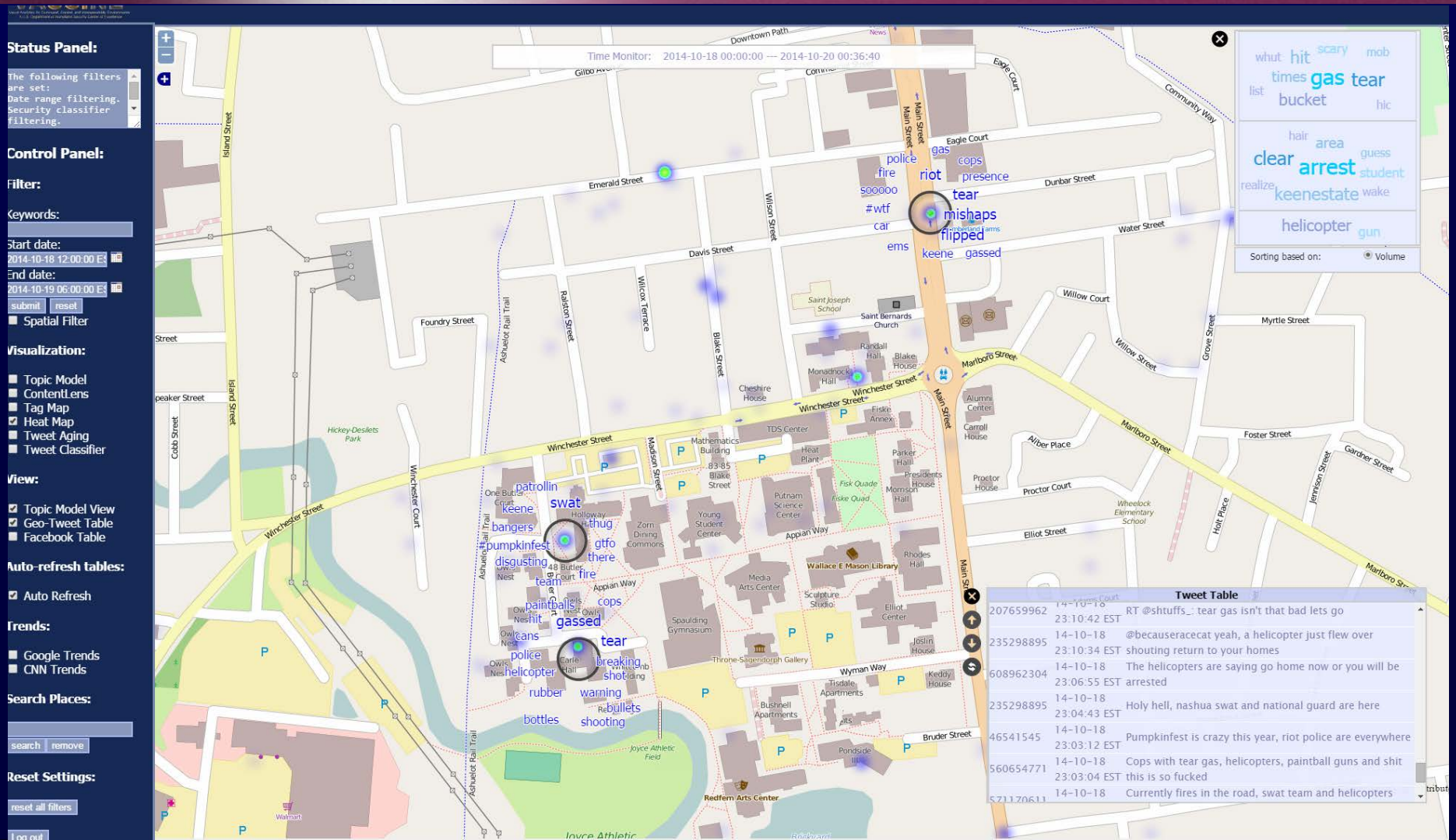


Security classifier became prominent.



Keene, NH

Time range:10/18 12:00pm – 10/19 06:00am, Security classifier)

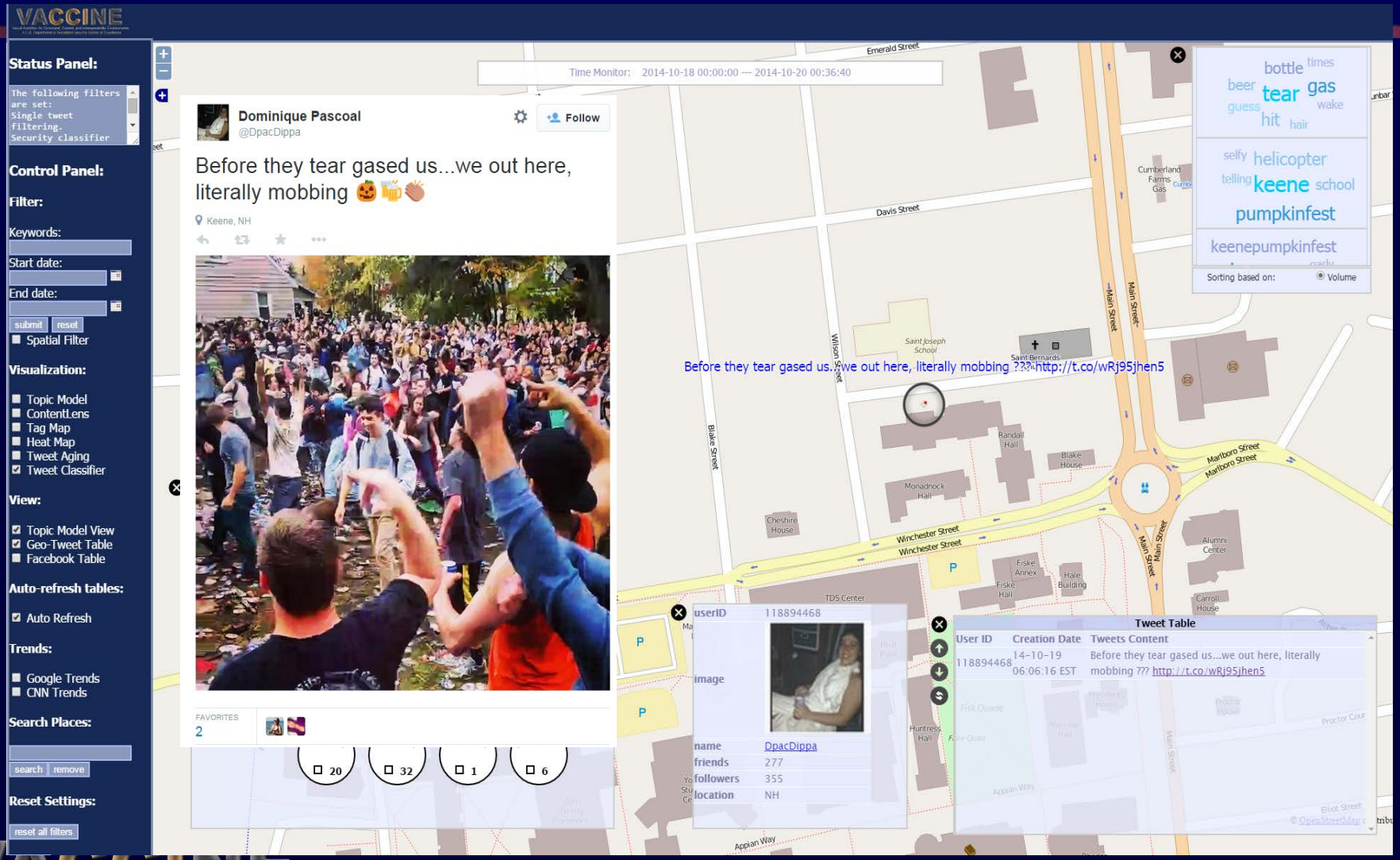


Keene, NH



The picture was later used by ABC news(link).

Keene, NH

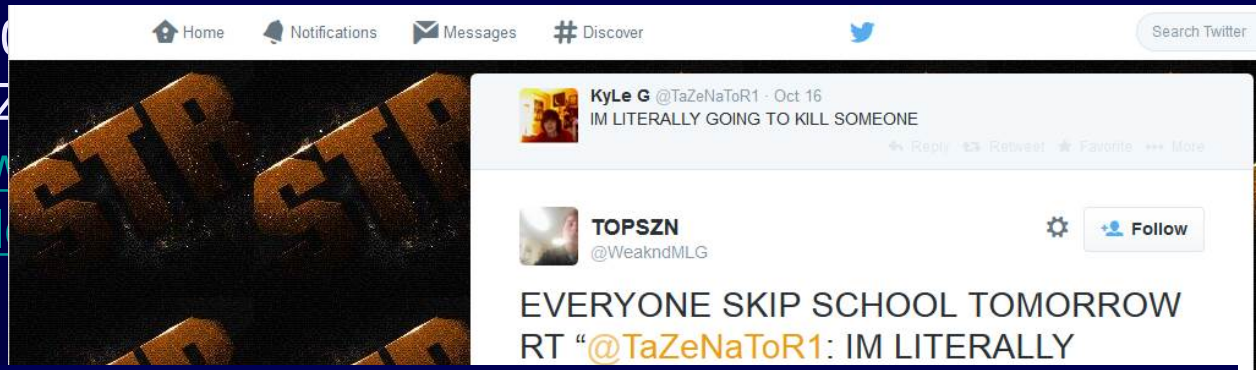

VACCINE
Status Panel:
The following filters are set:
Single tweet filtering.
Security classifier
Control Panel:
Filter:
Keywords:
Start date:
End date:
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
search
remove
Reset Settings:
reset all filters
Time Monitor: 2014-10-18 00:00:00 --- 2014-10-20 00:36:40
Dominique Pascoal
@DpacDippa
Follow
Before they tear gased us...we out here, literally mobbing
Keene, NH
FAVORITES
2
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
bottle
times
beer
tear
gas
guess
wake
hit
hair
selfy
helicopter
telling
keene
school
pumpkinfest
keenepumpkinfest
Sorting based on:
Volume
userID 118894468
image
name DpacDippa
friends 277
followers 355
location NH
Tweet Table
User ID
Creation Date
Tweets Content
118894468
14-10-19 06:06:16 EST
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
Emerald Street
Davis Street
Main Street
Wilson Street
Blake Street
Winchester Street
Marlboro Street
Saint Joseph School
Saint Bernards
Randall Hall
Blake House
Monadnock Hall
Cheshire House
Fiske Annex
Fiske Hall
Hale Building
TDS Center
Alumni Center
Carroll House
Huntress Hall
Appian Way
Cumberland Farms Gas

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



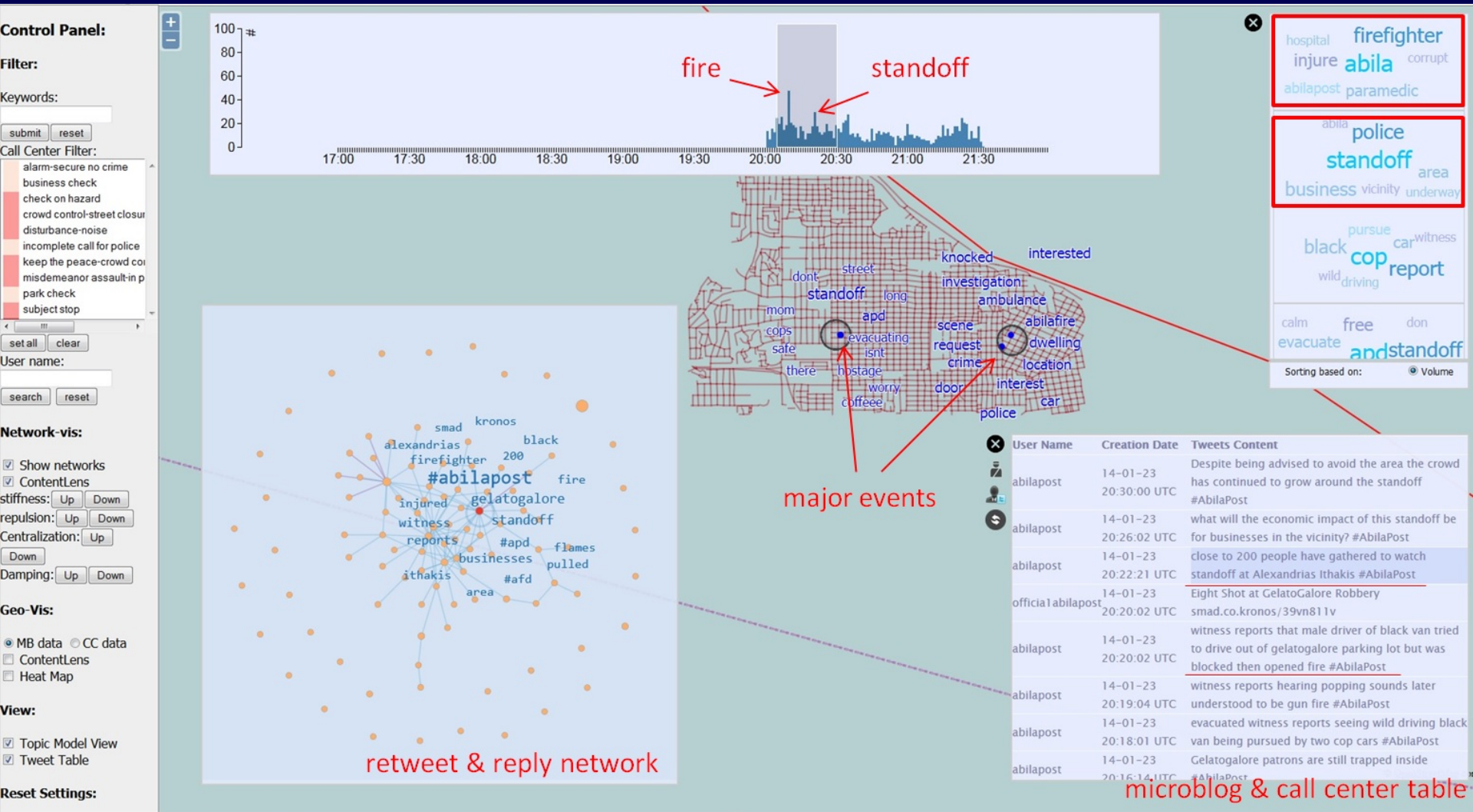
Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

SMART: Network Linkage



Summary

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Conclusions

- Visual Analytics is a powerful tool for harnessing big data
- Empowering the user, not replacing them

- *If you have a problem…*
  we can develop a solution!

For Further Information

www.VisualAnalytics-CCI.org



CYBER SECURITY DIVISION

2014 R&D SHOWCASE AND TECHNICAL WORKSHOP

Visual Analytics for Security Applications

VACCINE, Purdue University

David S. Ebert

December 18, 2014

Homeland Security

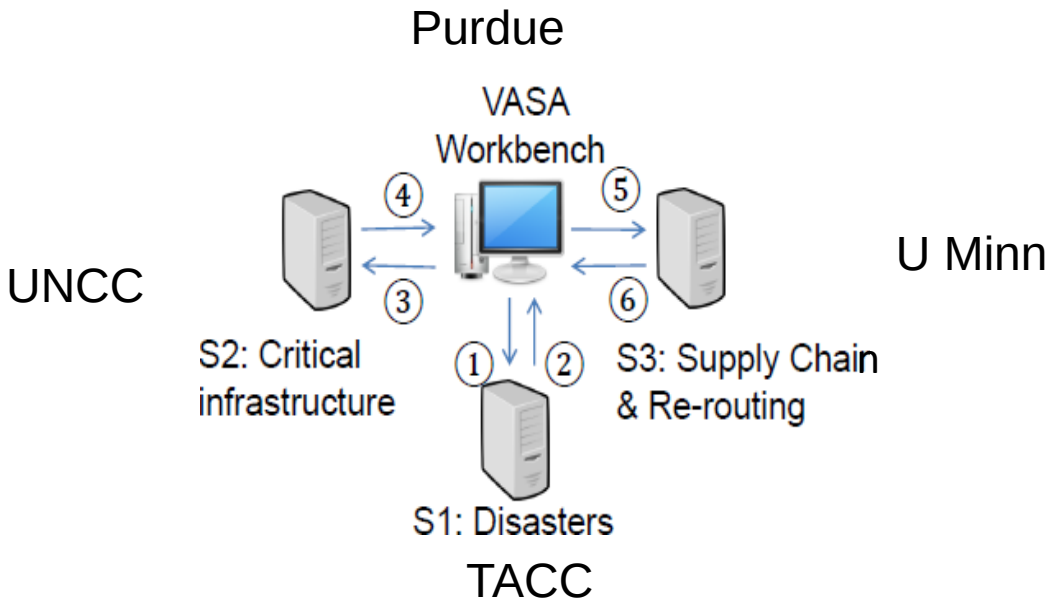
Science and Technology

VACCINE

Team Profile

- **Purdue – VASA workbench and visualization**
- **UNC Charlotte – critical infrastructure**
- **U. Minn. (NCFPD) - supply chain**
- **U. Texas (TACC) – disaster simulation**

- U. Konstanz
- U. Stuttgart
- Fraunhofer IGD
- Siemens
- German utilities



Society is under threat from many sources...



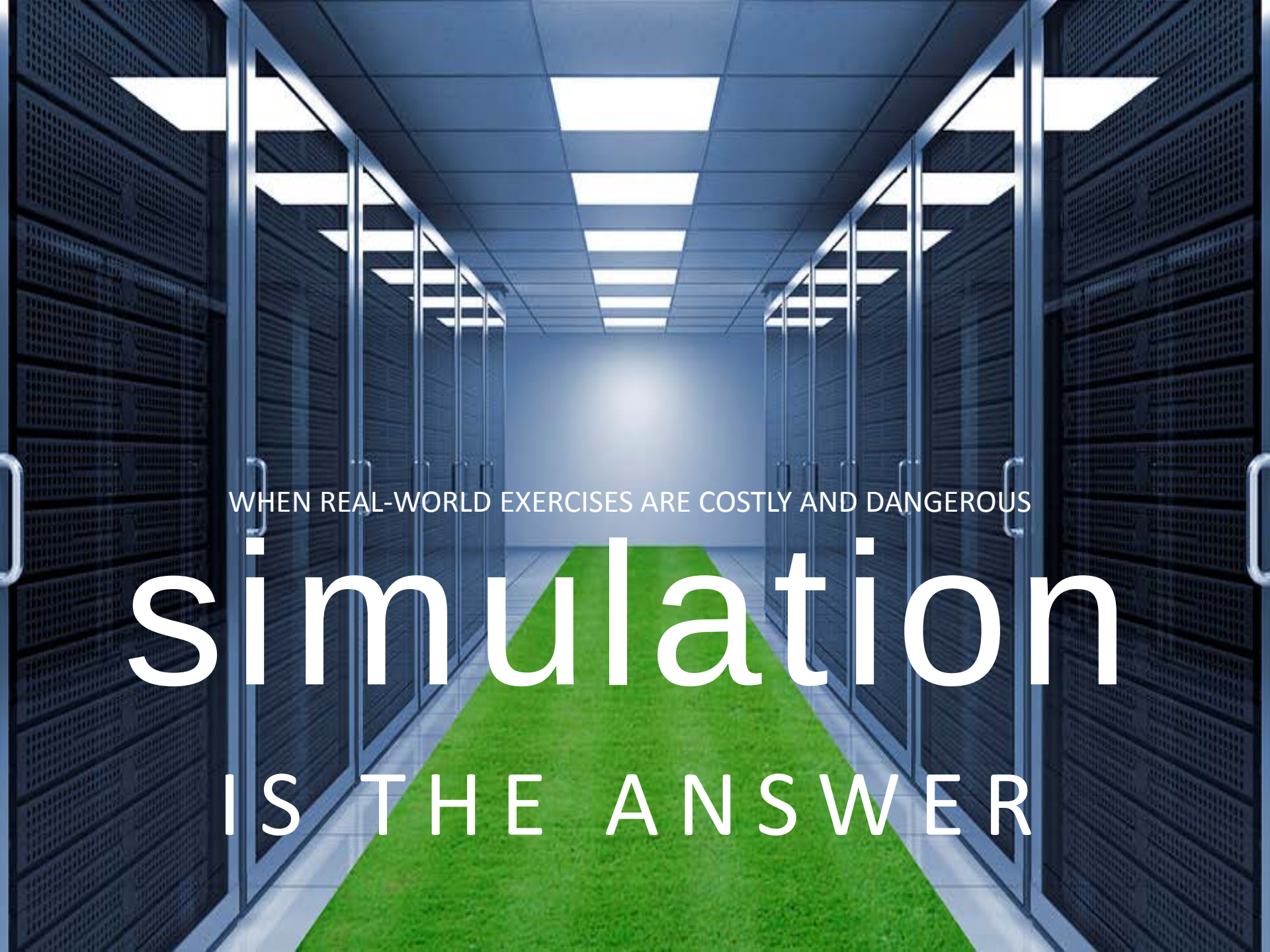

Power grid


Power plants


Highways


Supply chains


When DISASTER
strikes…
HOW CAN WE PREPARE?


WHEN REAL-WORLD EXERCISES ARE COSTLY AND DANGEROUS
simulation
IS THE ANSWER



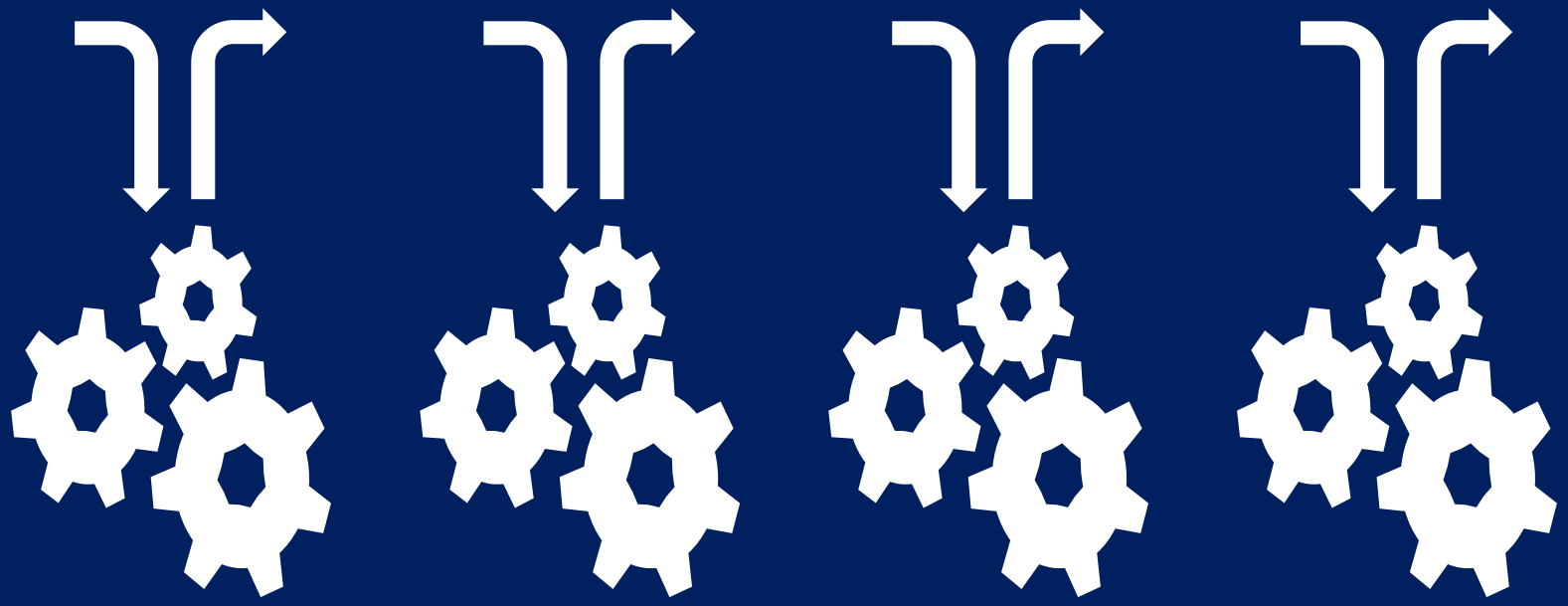
VISUAL ANALYTICS FOR SIMULATION-BASED ACTION

Distributed component-based framework for **computational steering** of systems-of-systems **simulations** for societal infrastructure based on a **visual analytics** approach

VASA Main Idea:

Pipeline Of Asynchronous Simulations

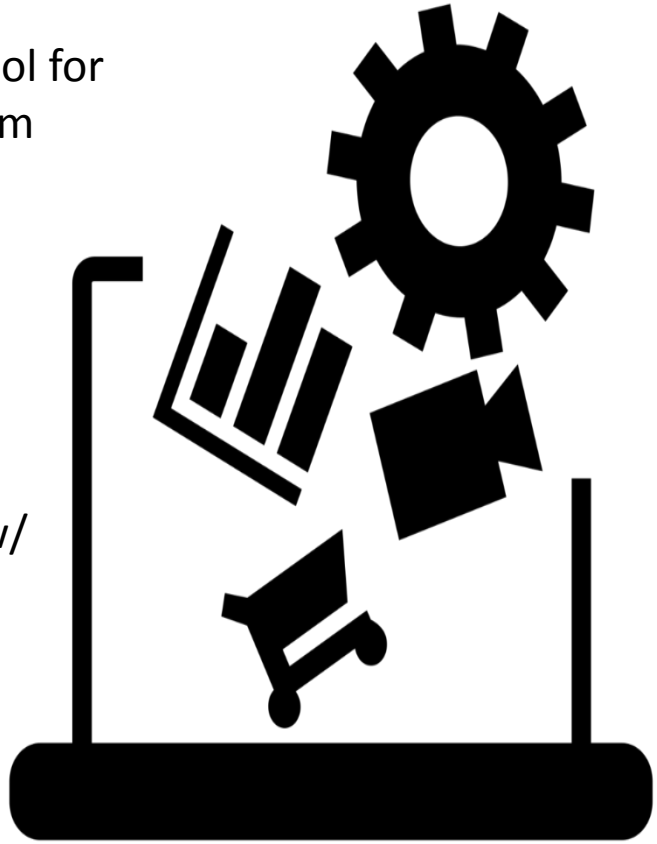
(System of System)



VASA Workbench

Interactive **desktop** tool for a distributed system

Control **distributed simulations** using REST API



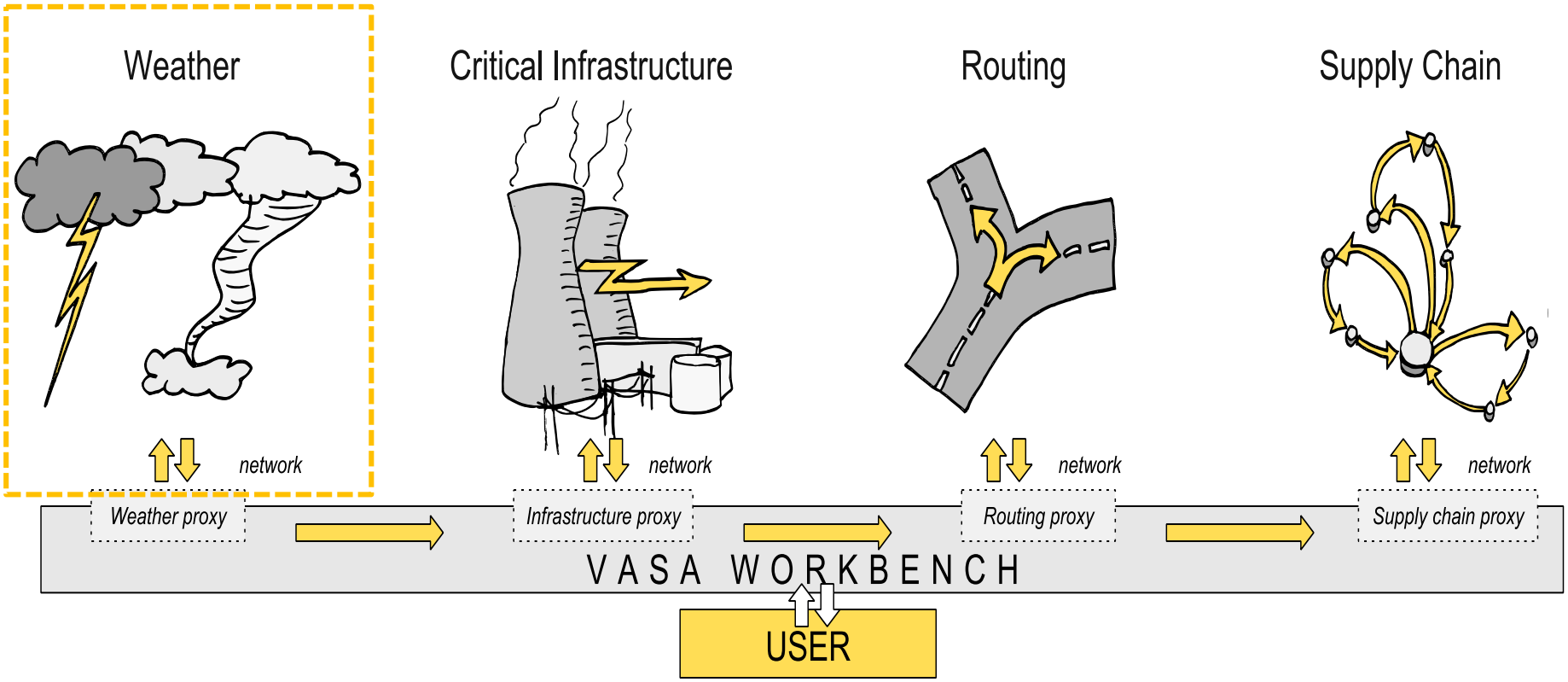
Visual analytics **dashboard** w/ multiple coordinated views

Simulation proxy provides real-time response

Configure + steer + explore (simulation models)

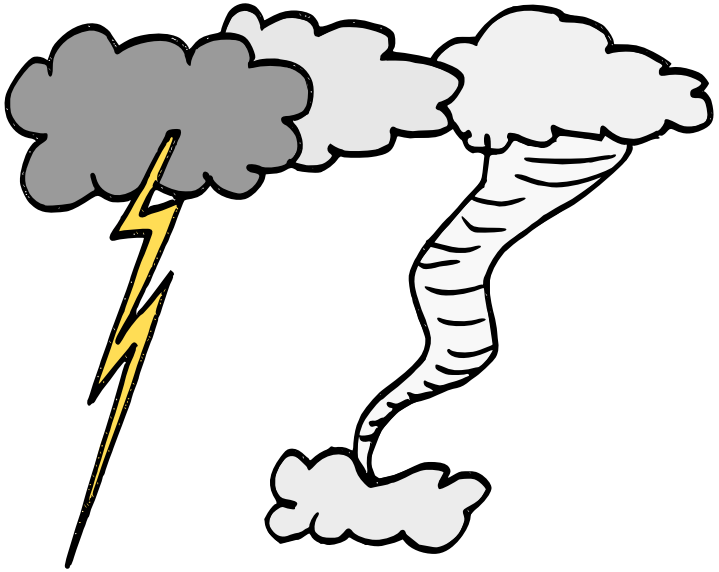
VACCINE

VASA COMPONENT: **WEATHER**



VASA COMPONENT: **WEATHER**

Historic data (Irene, Sandy, etc)



Prepares **event datasets** from server

ADCIRC: Advanced Circulation

Visualizes **hurricane cone** over time using slider

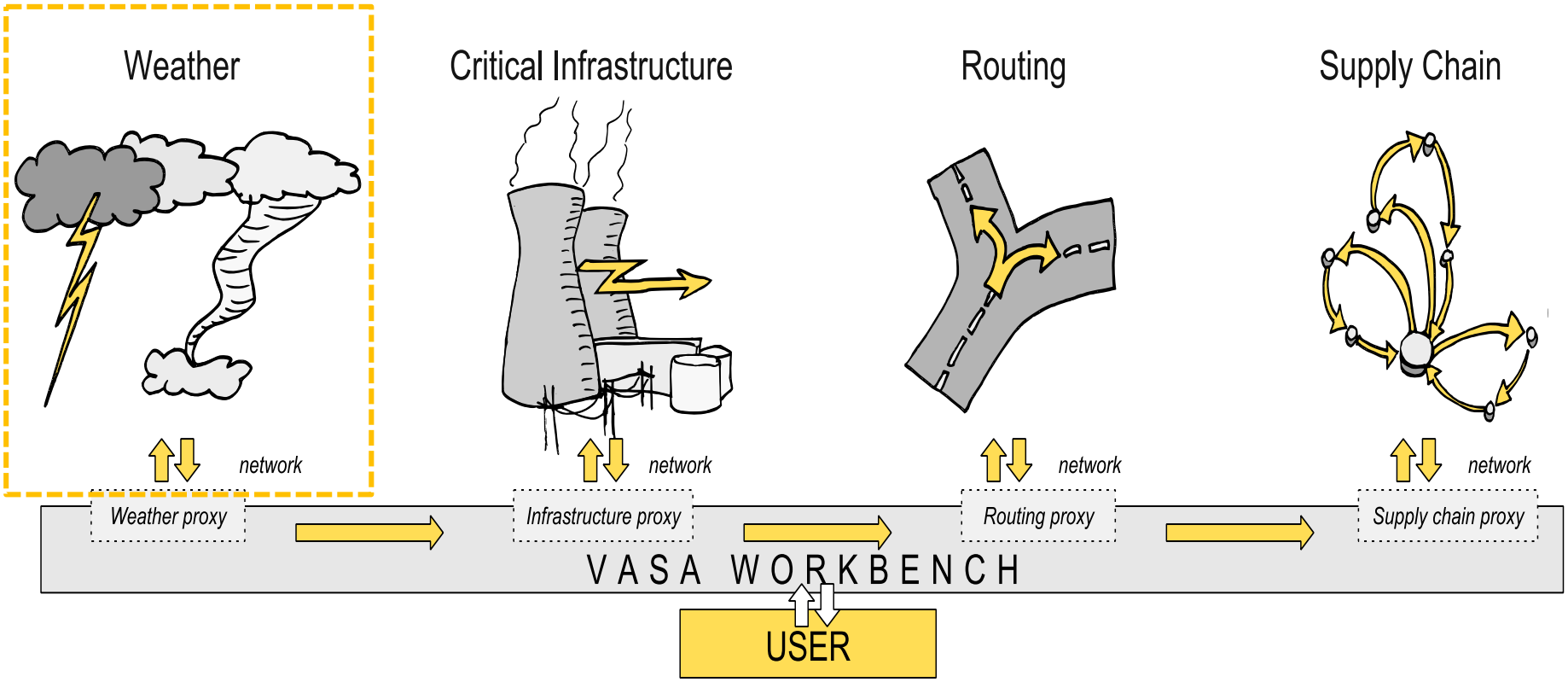
NOAA: National Oceanic and Athmospheric Administration

Generates **inputs** to downstream components

MODEL

PROXY

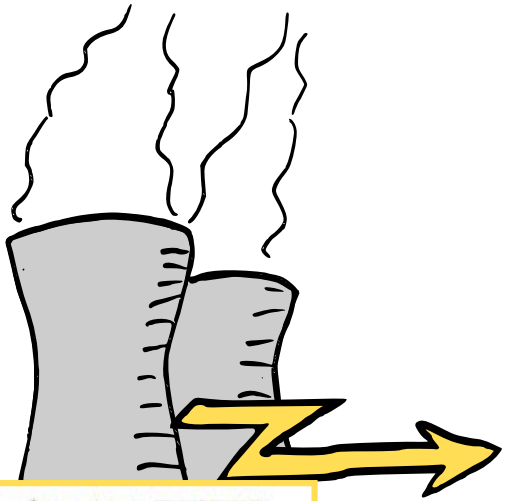
VASA COMPONENT:
CRITICAL INFRASTRUCTURE



VASA COMPONENT:
CRITICAL INFRASTRUCTURE

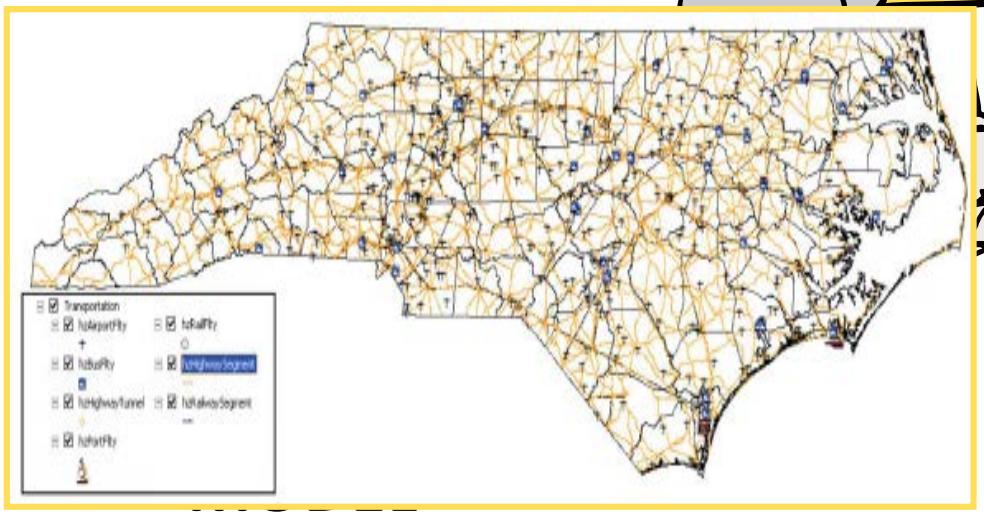
Vu environment with submodels

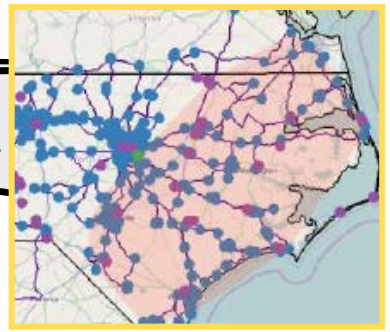
Simplified **connectivity graph** of important structures

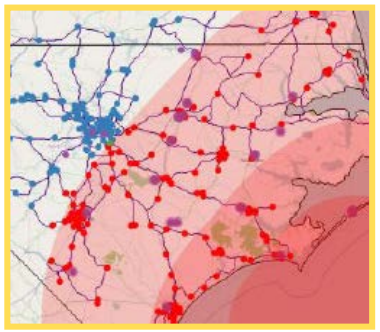


E.g. Electric grids, telecom networks, gas distribution

Example: show impact of hurricanes on restaurants

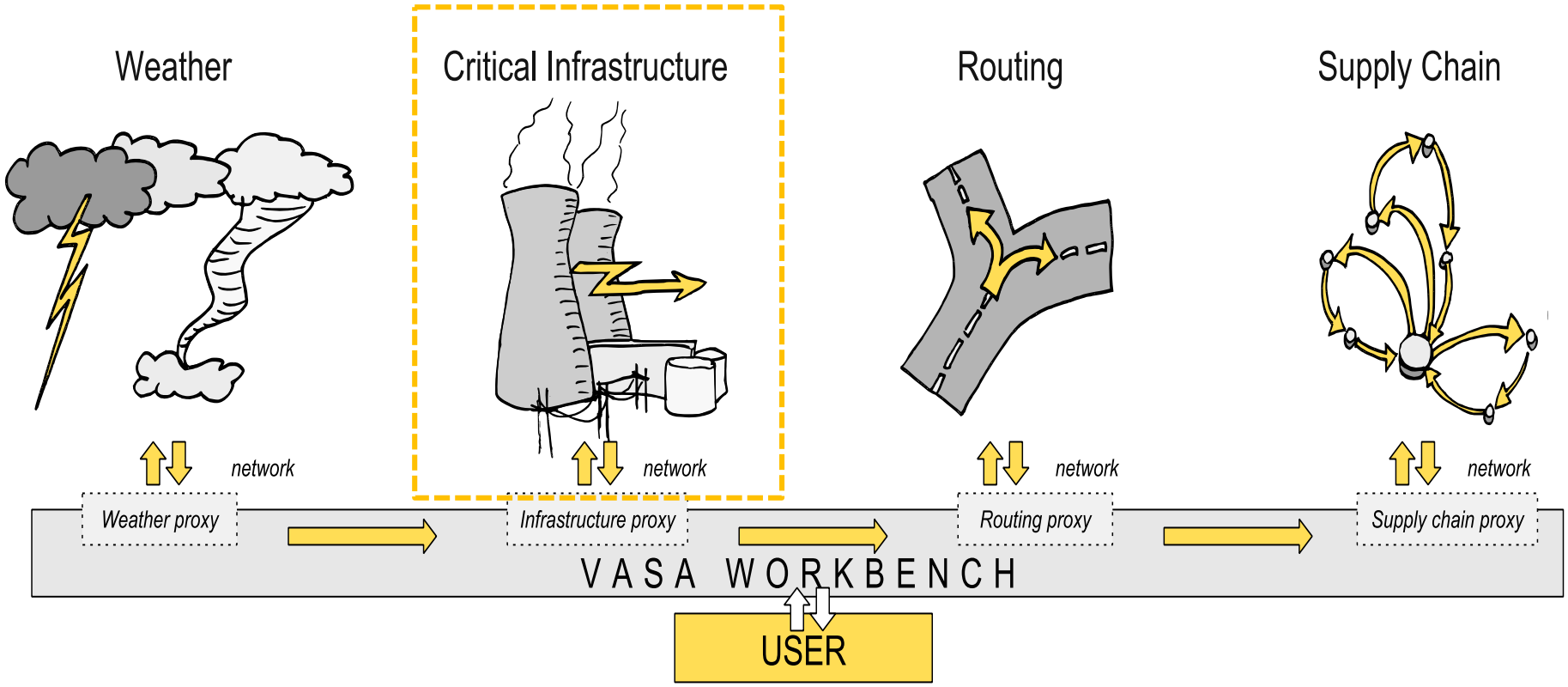






PROXY

VASA COMPONENT: **ROUTING**

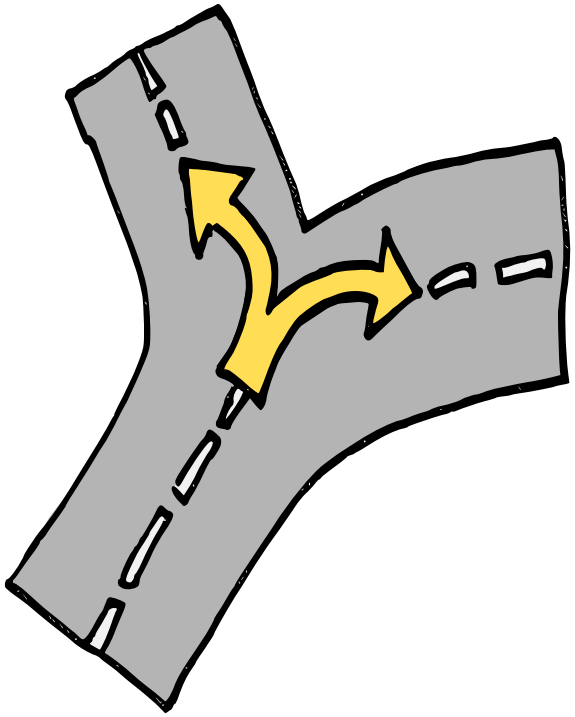


VASA COMPONENT: **ROUTING**

Maintains **road network** for critical infrastructure

Simulation engine: **ArcGIS Server**

Inputs: barriers and closures
Outputs: new transport routes



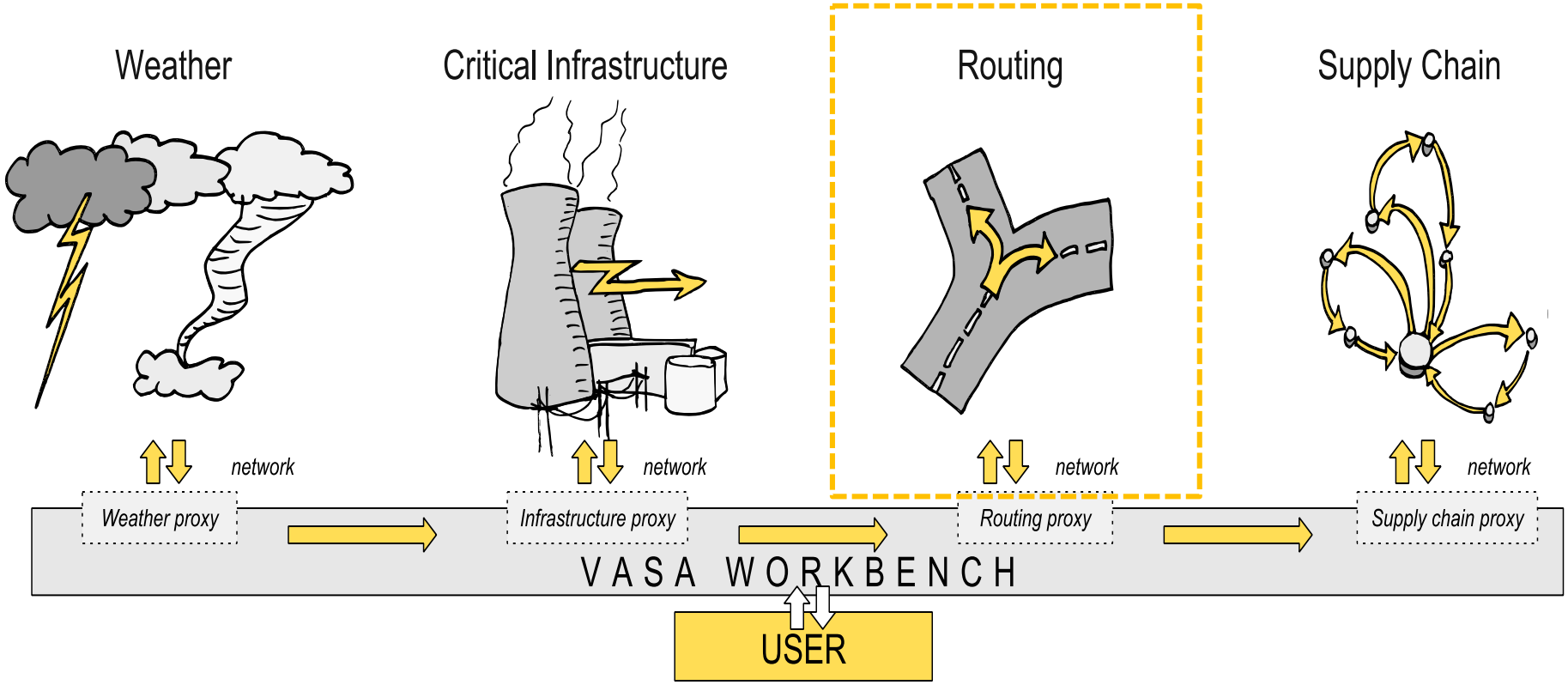
Approximates disabled routes and facilities

Visualizes disabled routes and facilities

MODEL

PROXY

VASA COMPONENT:
SUPPLY CHAIN

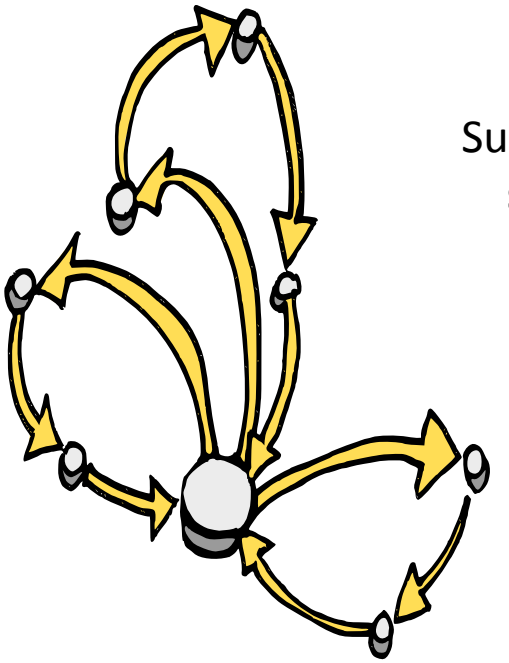


VASA COMPONENT:
SUPPLY CHAIN

Supply chains depend on **business** and **goods**

Our models: poultry firm + fast food = farm to restaurant

Discrete event simulation on chain in geolocated facilities



Accepts external inputs (**weather** and **roads**)

Supports **road closures, powerless stores** and **flooding** (polygons)

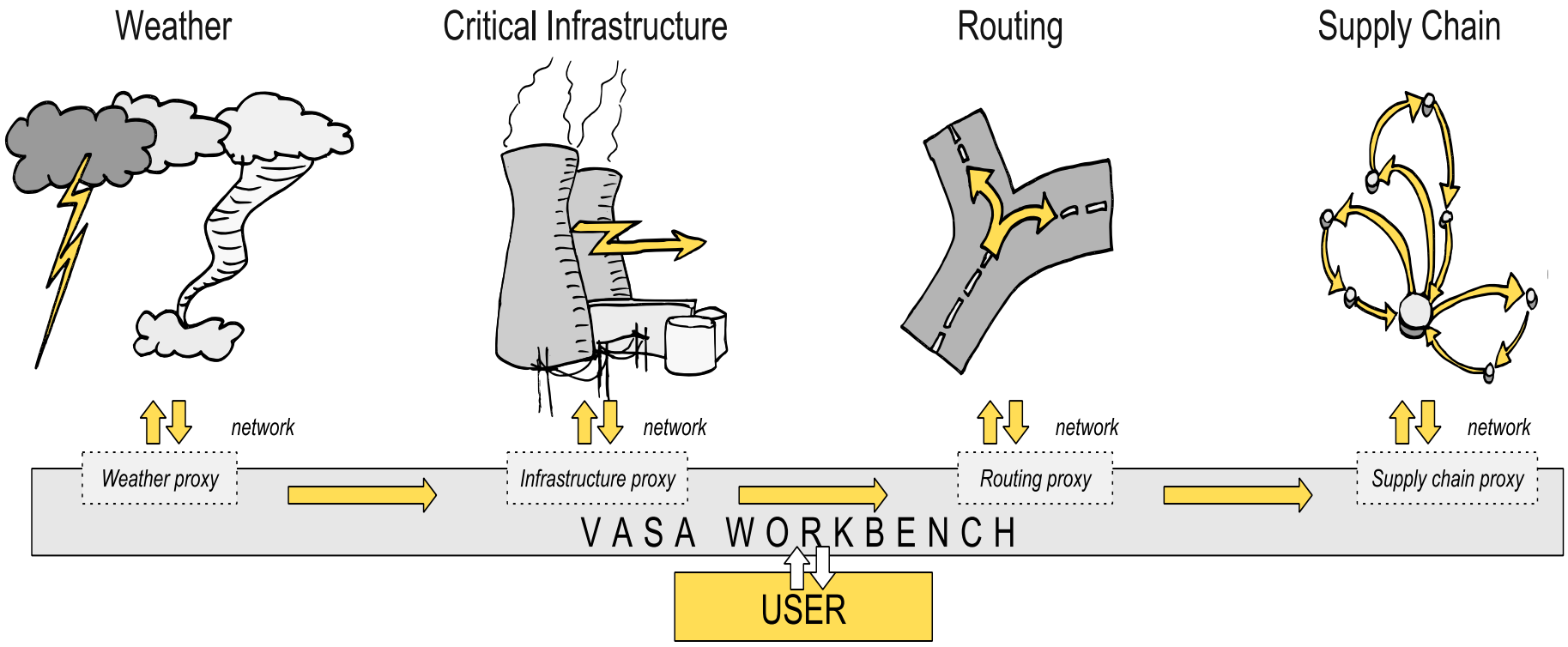
Optimizes distribution and even redistributes products

Food contamination also modeled and visualized

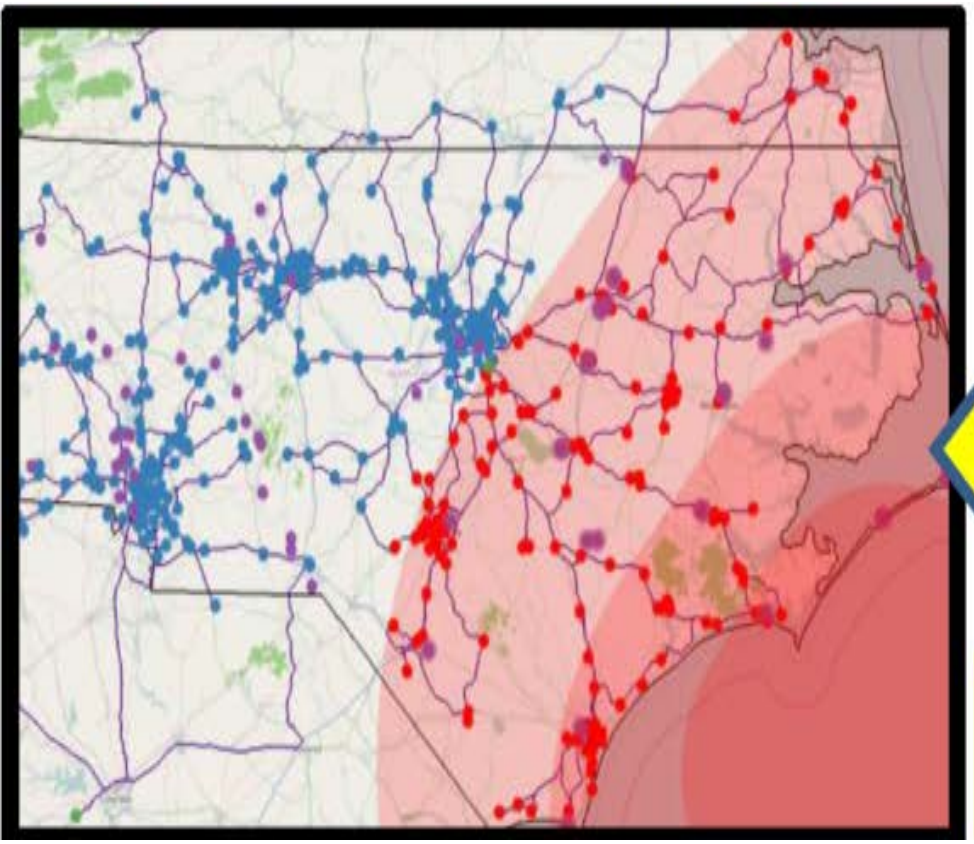
MODEL

PROXY

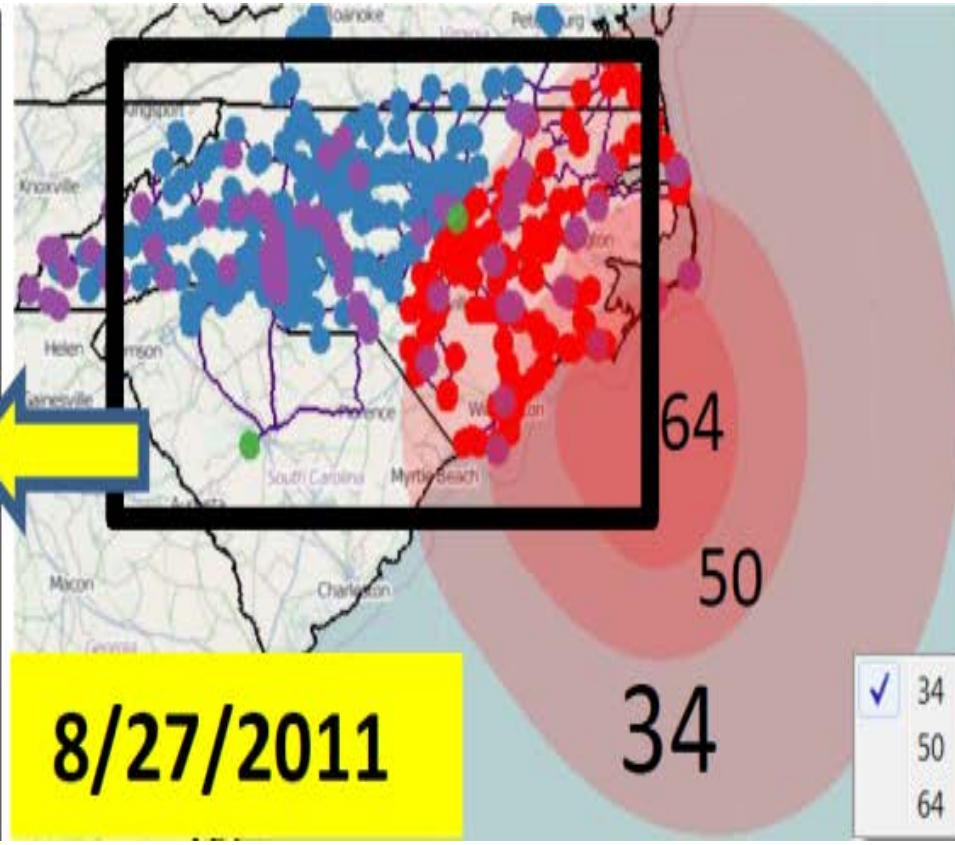
VASA


Weather
Critical Infrastructure
Routing
Supply Chain
network
network
network
network
Weather proxy
Infrastructure proxy
Routing proxy
Supply chain proxy
VASA WORKBENCH
USER

Example: U.S. Hurricane Season



34-knot winds batter the coast; critical infrastructure proxy estimates **impacted restaurant**



Hurricane Irene hits North Carolina on August 27, 2011

VACCINE

Transition: Initial Partners

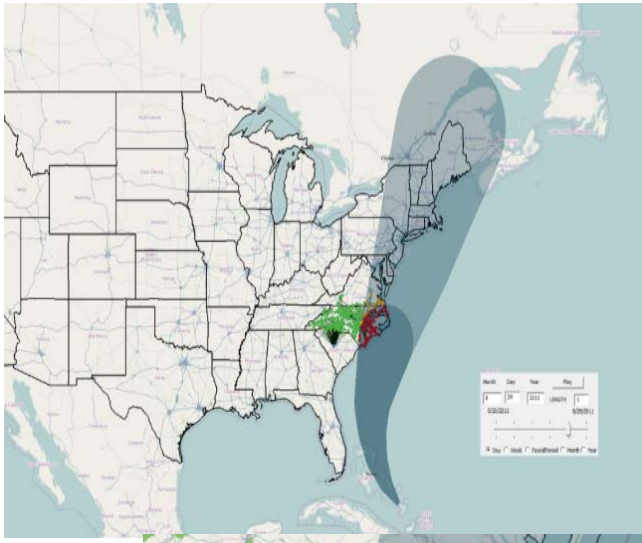
- FEMA – regional FEMA feedback:
  - Unprecedented work
  - Whole community approach
- U.S. Coast Guard
  - Emergency preparedness interest
  - Dire need with no apparent solution available

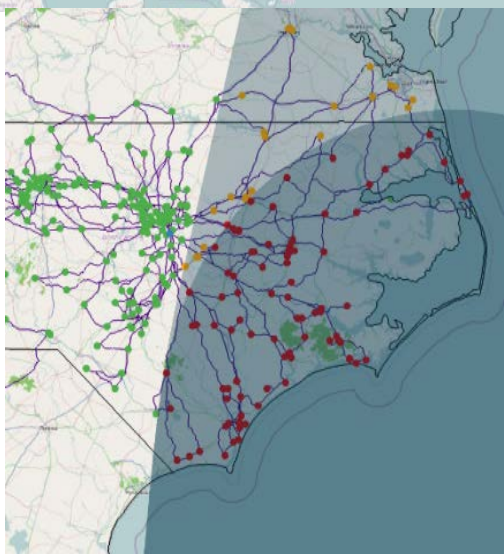
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

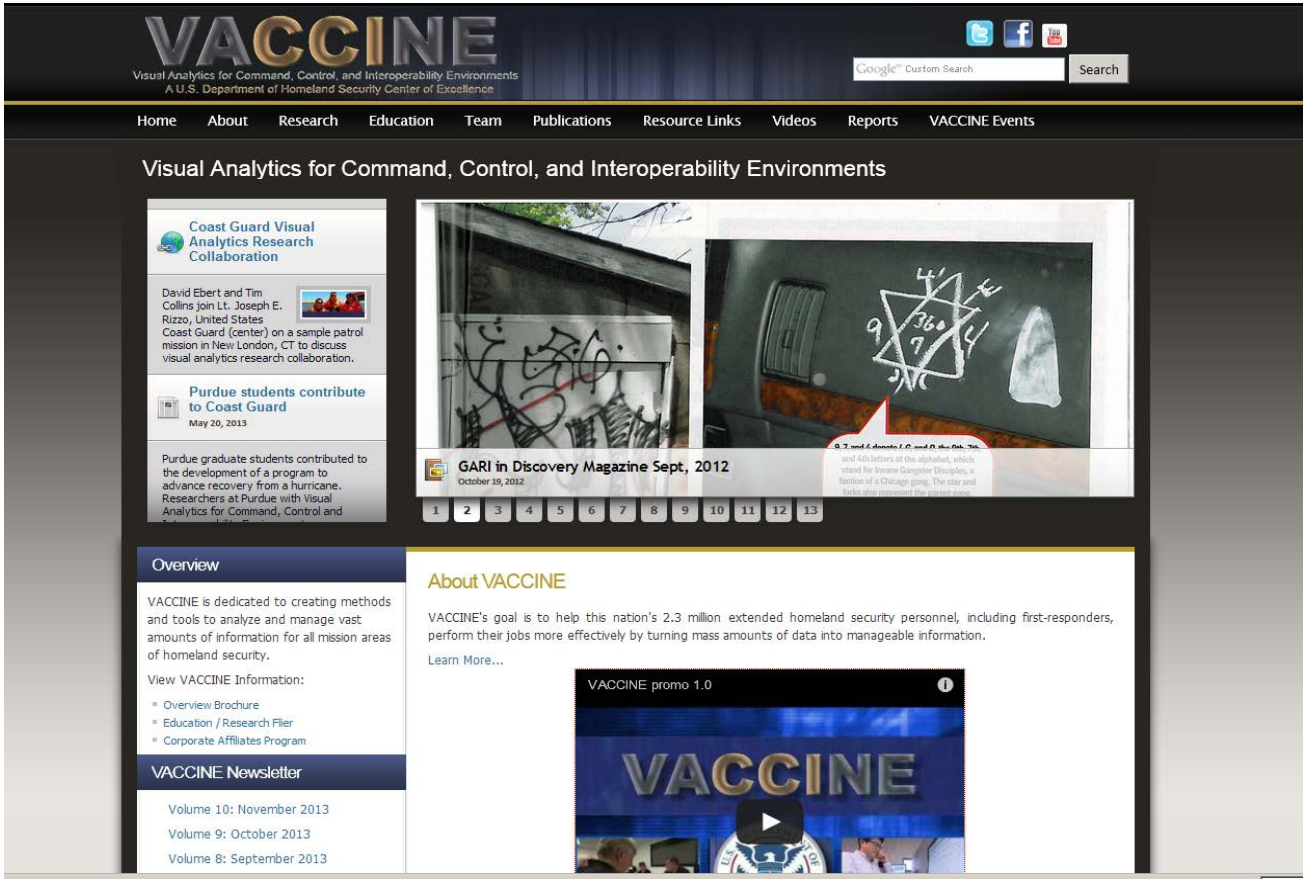
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





Contact

- ebertd@purdue.edu
- www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics
A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Data-Driven Decision Making

David Ebert
Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments
A U.S. Department of Homeland Security Center of Excellence

September 2014

Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina Charlotte
- University of Washington
- Arizona State University
- Oak Ridge National Lab
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

September 2014

VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

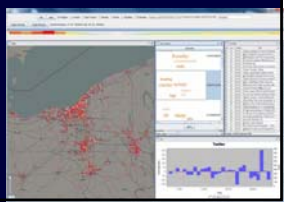
VACCINE

September 2014

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



VACCINE

September 2014

Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

VACCINE

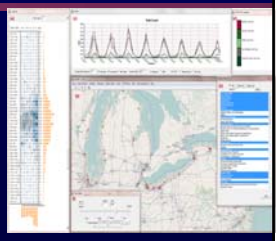
September 2014

U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



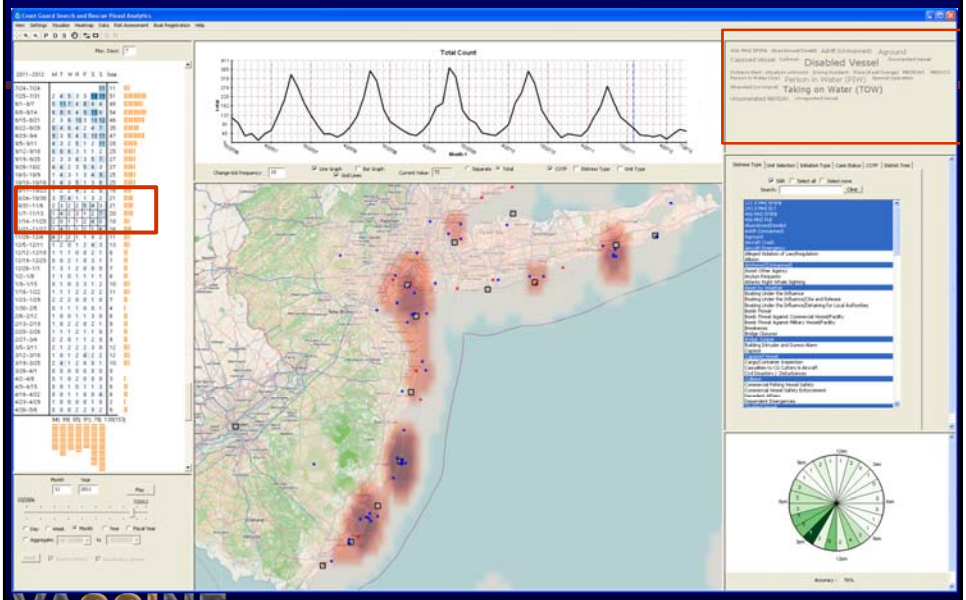


VACCINE

September 2014

Example: Risks and Consequences From Sandy:

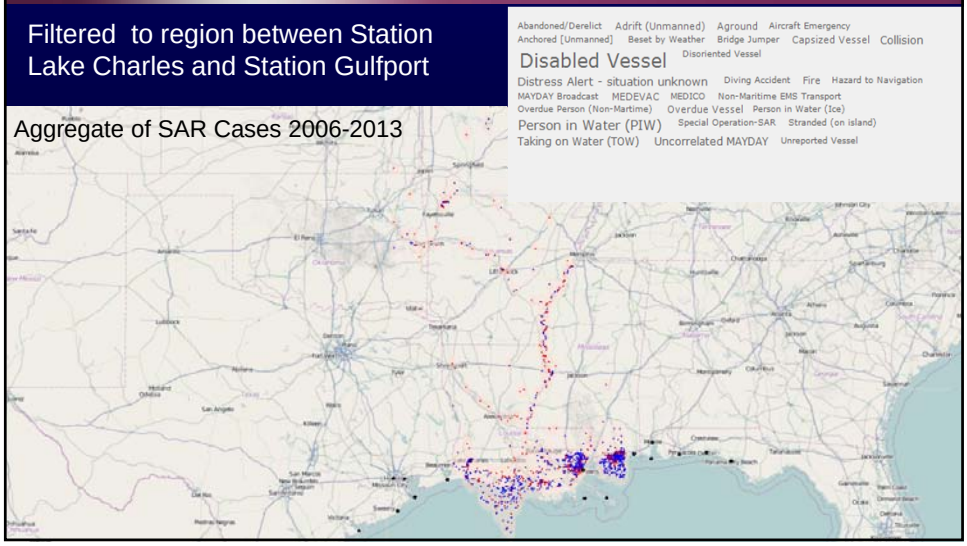
SAR Cases November 2011 NJ/NYC Area



Example: D8 Western Rivers and Coastal Operations - SAR Cases

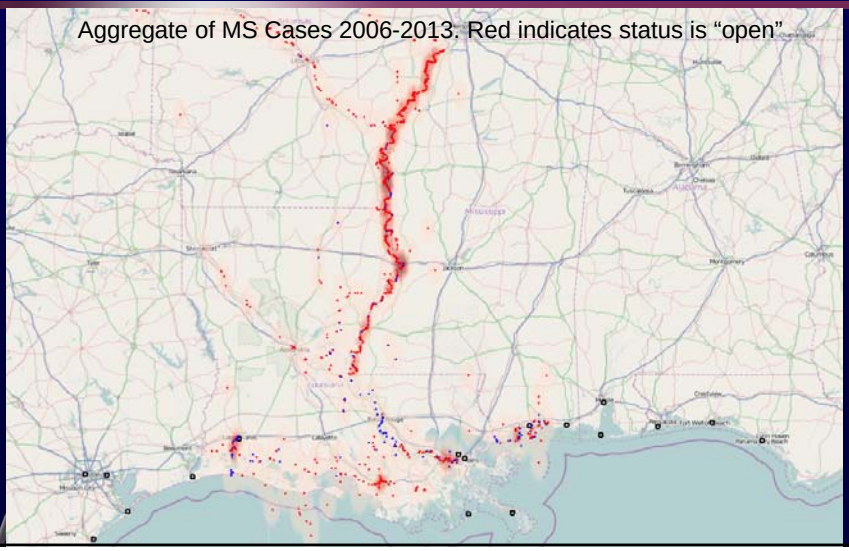
Filtered to region between Station Lake Charles and Station Gulfport

Aggregate of SAR Cases 2006-2013



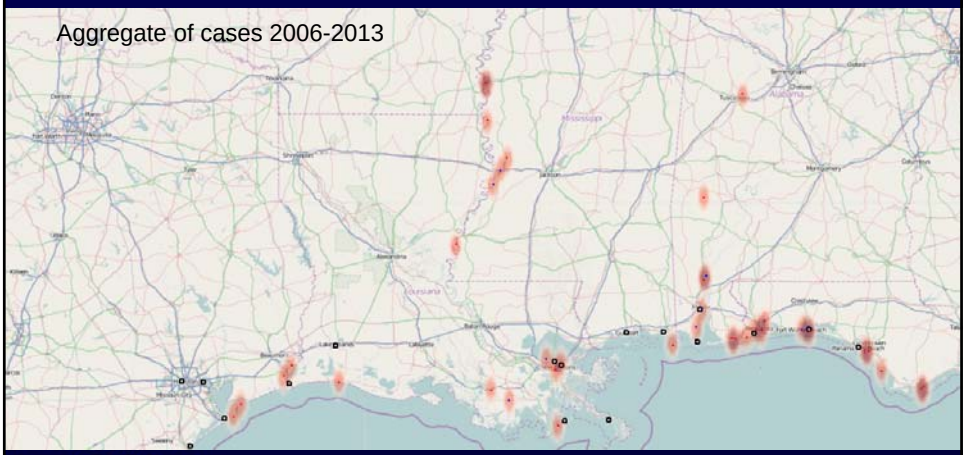
Example: D8: Maritime Safety Cases

Filtered to region between Station Lake Charles and Station Gulfport



Example: D8 Hazards to Navigation

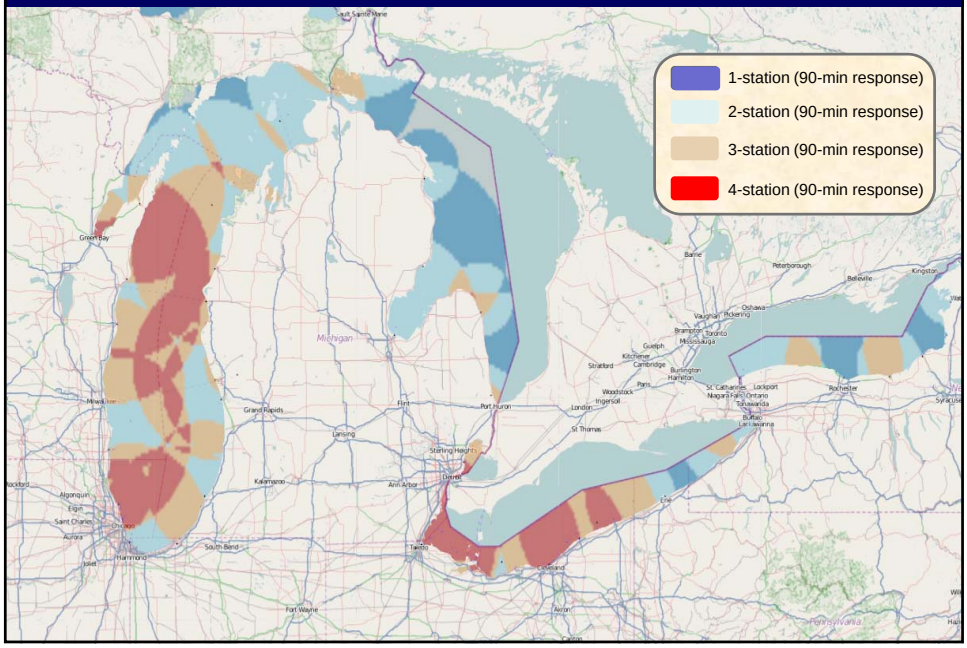
Filtered to region between Station Lake Charles and Station Gulfport



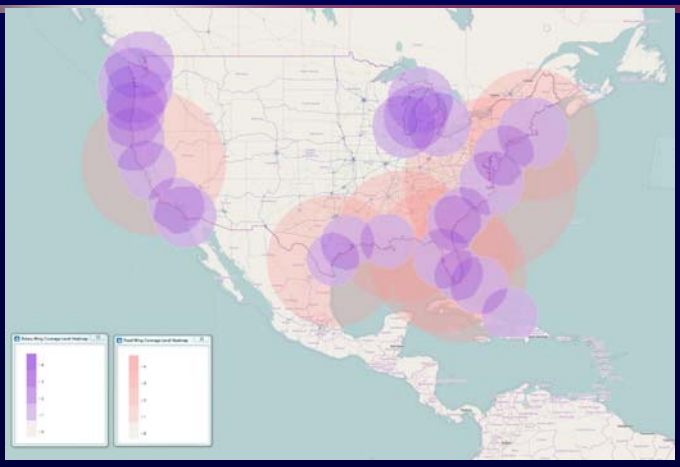
VACCINE

September 2014

Response Efficiency – Potential Future Assets



cgSARVA Results – Aircraft coverage



Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.

VACCINE

September 2014

Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

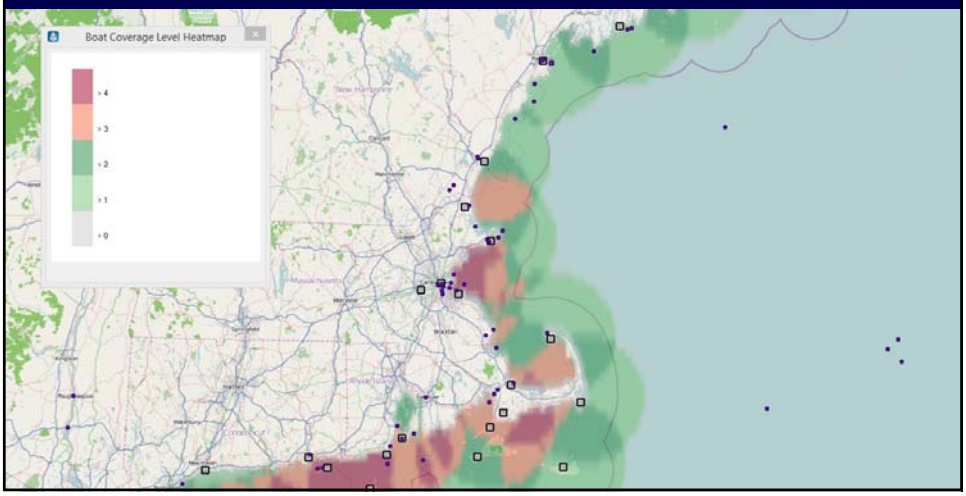


Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

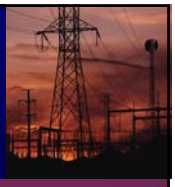
VACCINE

September 2014

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



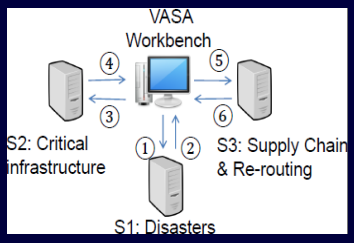
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

TACC

VACCINE

September 2014

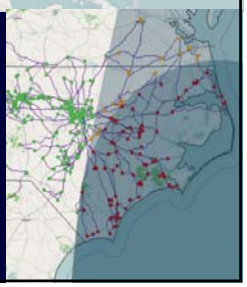
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow "what-if" exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





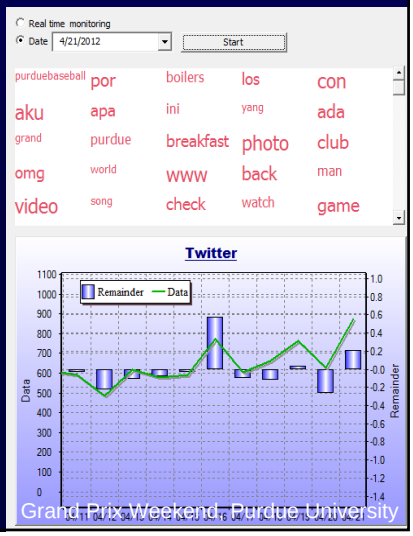
VACCINE

September 2014

Social Media: Real-time Visual Analytics

(Purdue, Stuttgart, Penn St.)

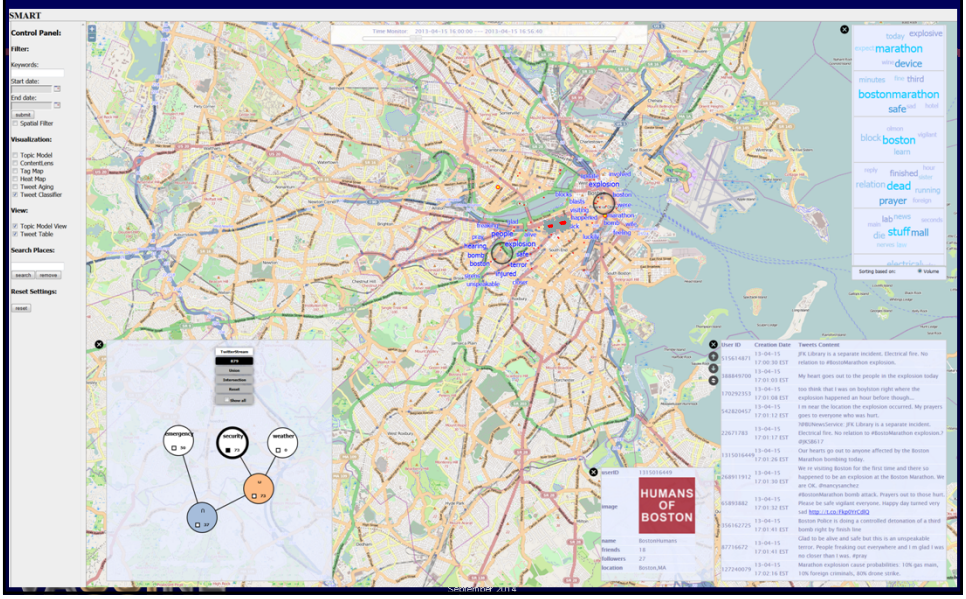
- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- Topics extracted shown as a dynamic word cloud
- Sources: Twitter, facebook, flickr
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC



VACCINE

September 2014

Social Media Analysis and Reporting Toolkit (SMART)



Social Media Analysis and Reporting Toolkit (SMART)

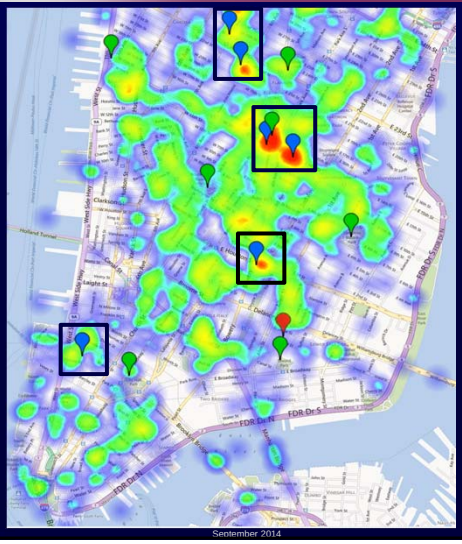
Hurricane Sandy

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM





VACCINE

Coast Guard False Distress Call Analysis Using SMART

1. Overview of the problem

Case	Time	Location
1	1:22 AM on 5 January 2014	Buxton, NC
2	5-6:00 PM on 11 August 2014	Annapolis, MD

2. Data

Data Type	Time Range	Data Source	Geo Bound	After filtering / total size
Geo Twitter	01/2013 -- present	VACCINE Twitter DB	US	9,335 / 1,484,449,638
Non-geo Twitter	01/2014 -- present	VACCINE Twitter DB plus Twitter.com	None	7,422 / 959,291,538
Public Facebook	08/2014 -- present	VACCINE DB	None	238 / 3,183,363

3. Process

We monitored social media traffic, filtered based on carefully defined keyword lists, restricted to detailed area and time ranges, created classifiers to categorize messages of different topics and narrowed to suspicious messages and users.

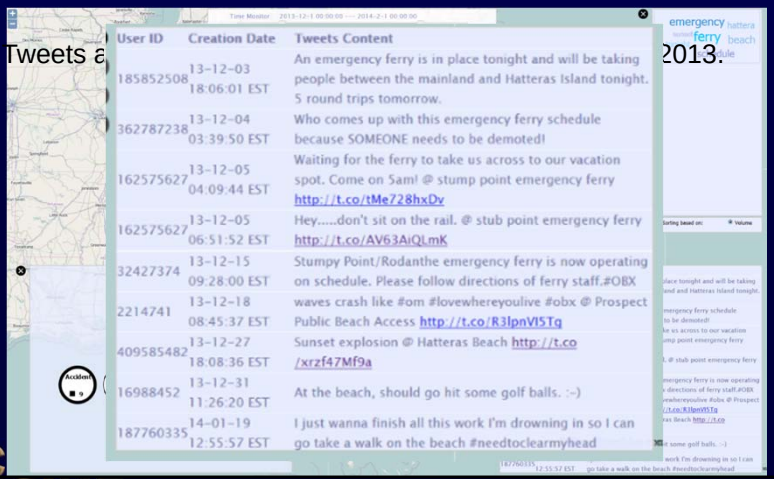
VACCINE

September 2014

Results – Buxton Case (Jan. 5th, 2014)

Searched tweets posted 12/2013-1/2014 within a radius of 60 miles, covering LOB (Line of Bearing) provided – **Only tweets related to emergency ferries found**

Tweets a 2013.

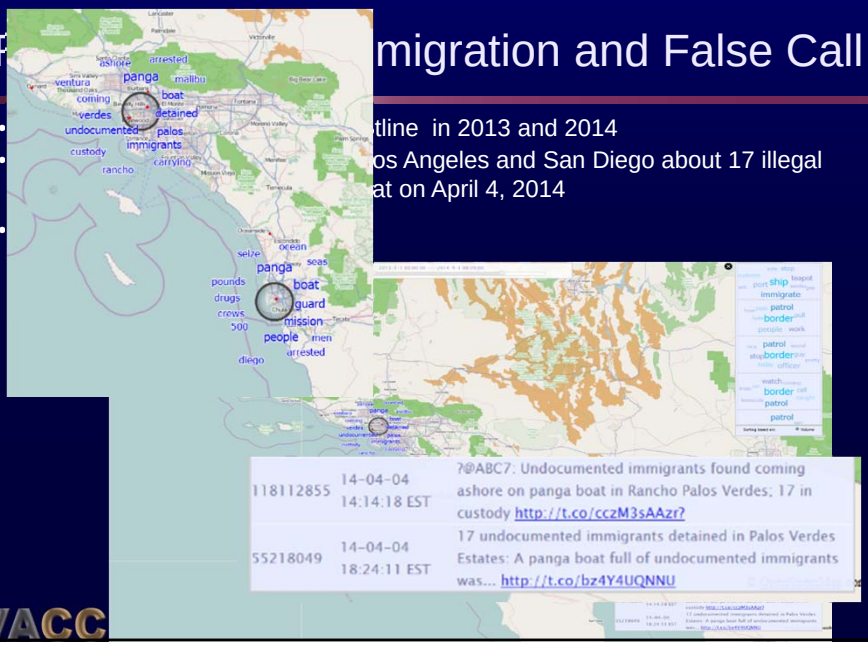


User ID	Creation Date	Tweets Content
185852508	13-12-03 18:06:01 EST	An emergency ferry is in place tonight and will be taking people between the mainland and Hatteras Island tonight. 5 round trips tomorrow.
362787238	13-12-04 03:39:50 EST	Who comes up with this emergency ferry schedule because SOMEONE needs to be demoted!
162575627	13-12-05 04:09:44 EST	Waiting for the ferry to take us across to our vacation spot. Come on 5am! @ stump point emergency ferry http://t.co/tMe728hxDv
162575627	13-12-05 06:51:52 EST	Hey.....don't sit on the rail. @ stub point emergency ferry http://t.co/AV63AiQLmK
32427374	13-12-15 09:28:00 EST	Stumpy Point/Rodanthe emergency ferry is now operating on schedule. Please follow directions of ferry staff.#OBX
2214741	13-12-18 08:45:37 EST	waves crash like #om #lovewhereyoulive #obx @ Prospect Public Beach Access http://t.co/R3IpnVI5Tq
409585482	13-12-27 18:08:36 EST	Sunset explosion @ Hatteras Beach http://t.co/xrzf47Mf9a
16988452	13-12-31 11:26:20 EST	At the beach, should go hit some golf balls. :-)
187760335	14-01-19 12:55:57 EST	I just wanna finish all this work I'm drowning in so I can go take a walk on the beach #needtoclearmyhead

VAC

P migration and False Call

- tline in 2013 and 2014
- os Angeles and San Diego about 17 illegal
- at on April 4, 2014



VACC

For Further Information

www.VisualAnalytics-CCI.org



VAC

CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Data-Driven Decision Making

David Ebert

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

July 2014

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America





VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

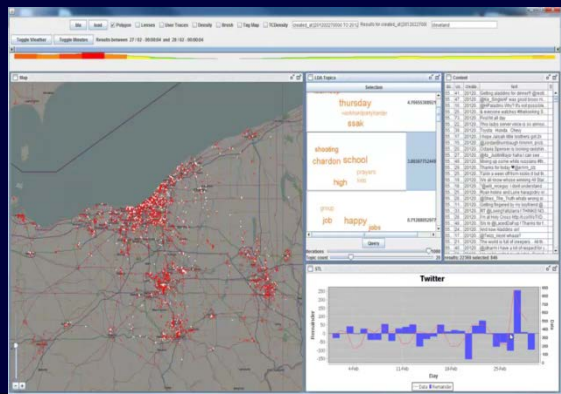
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

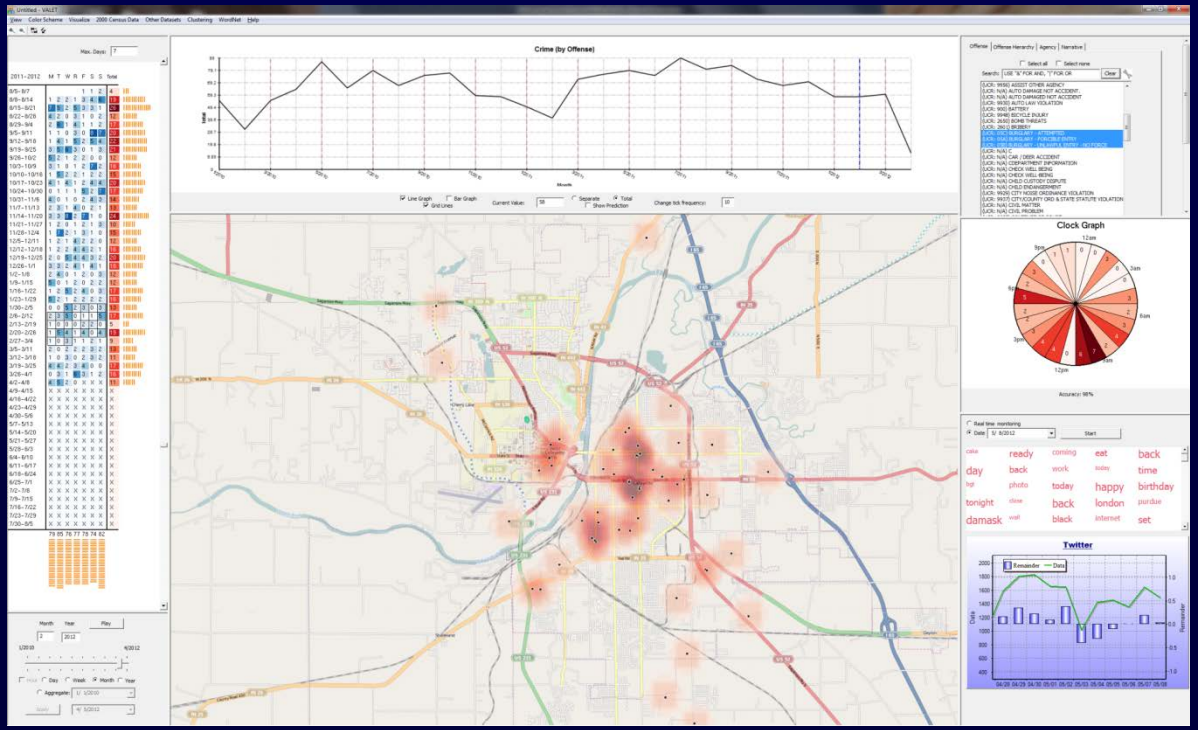
Visual Analytics Uses for Public Safety

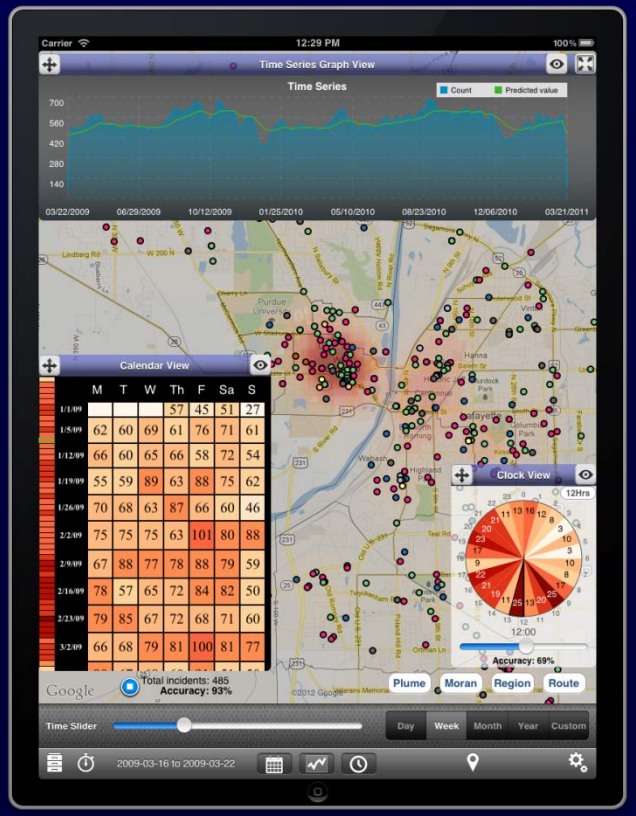


- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation

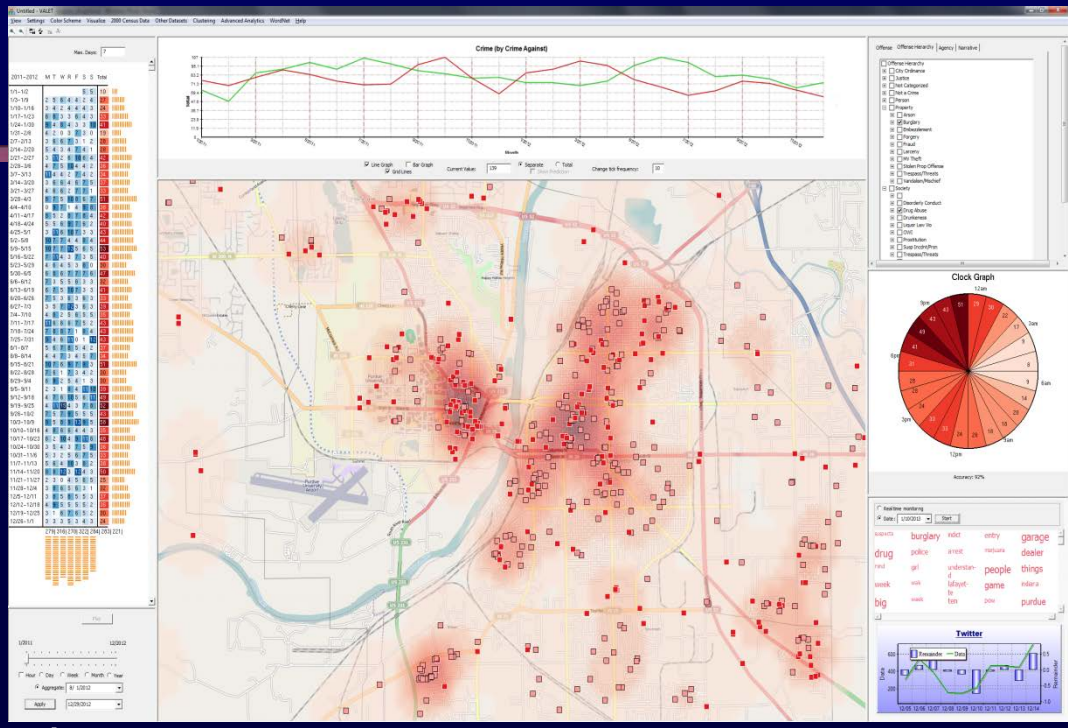


Visual Analytics Law Enforcement Toolkit (i)VALET





Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

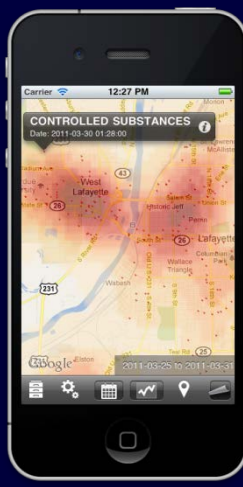
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police

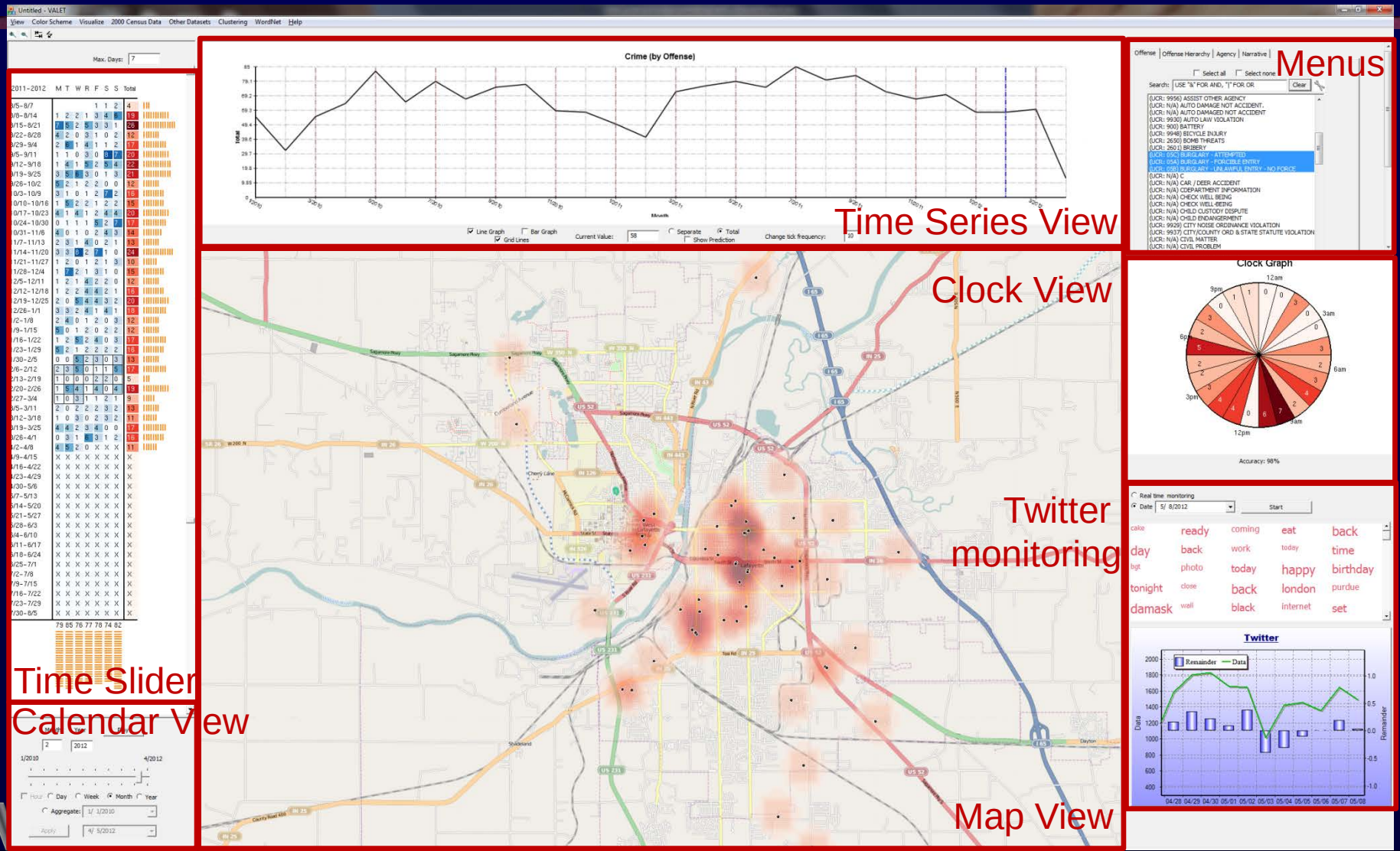
iVALET delivered:

- October 2011: Purdue, WL Police

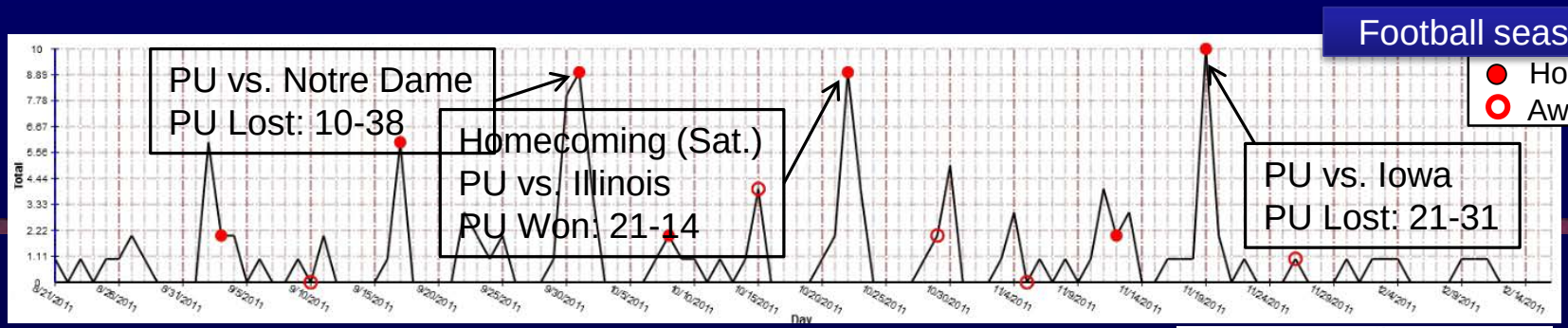


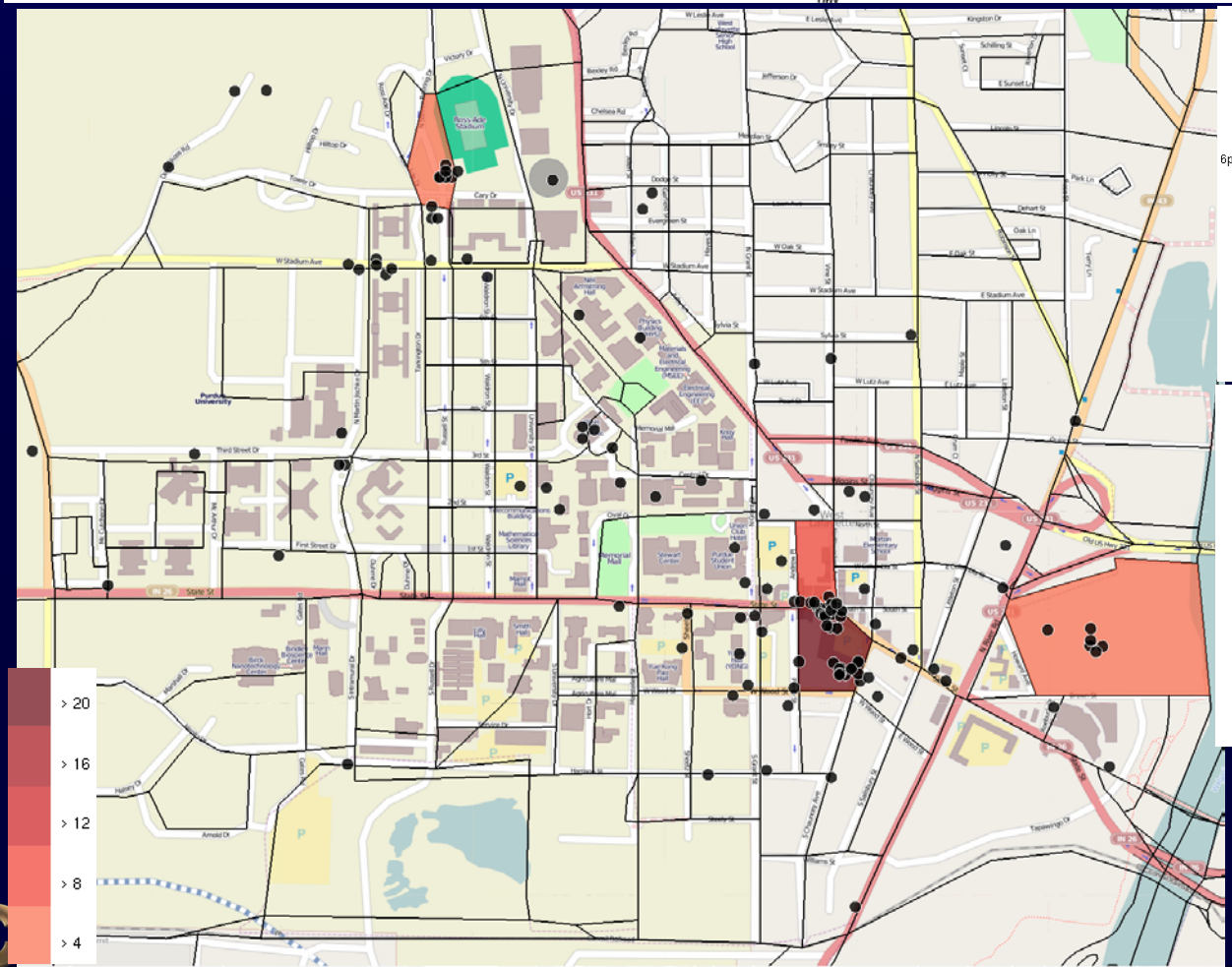
VACCINE

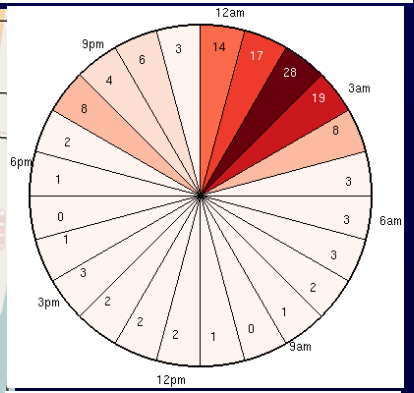
Example for Crime Analytics: VALET

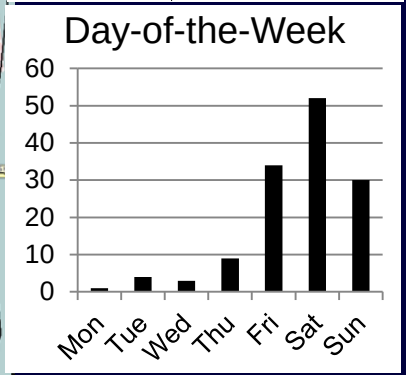


Example: Drunkenness / Public Intoxication









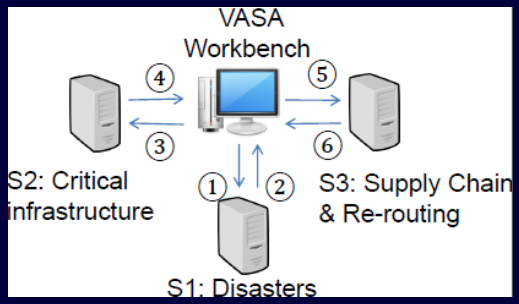
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s): Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

Purdue

UNCC



U Minn

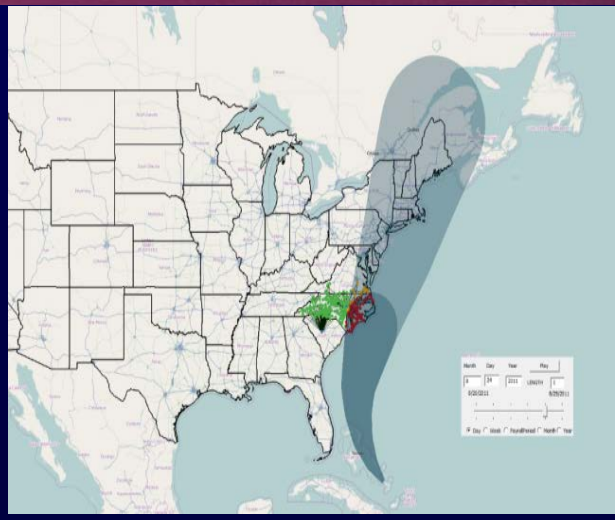
TACC

VACCINE

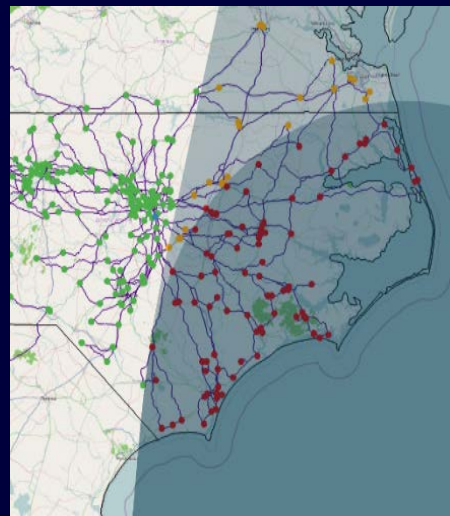
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

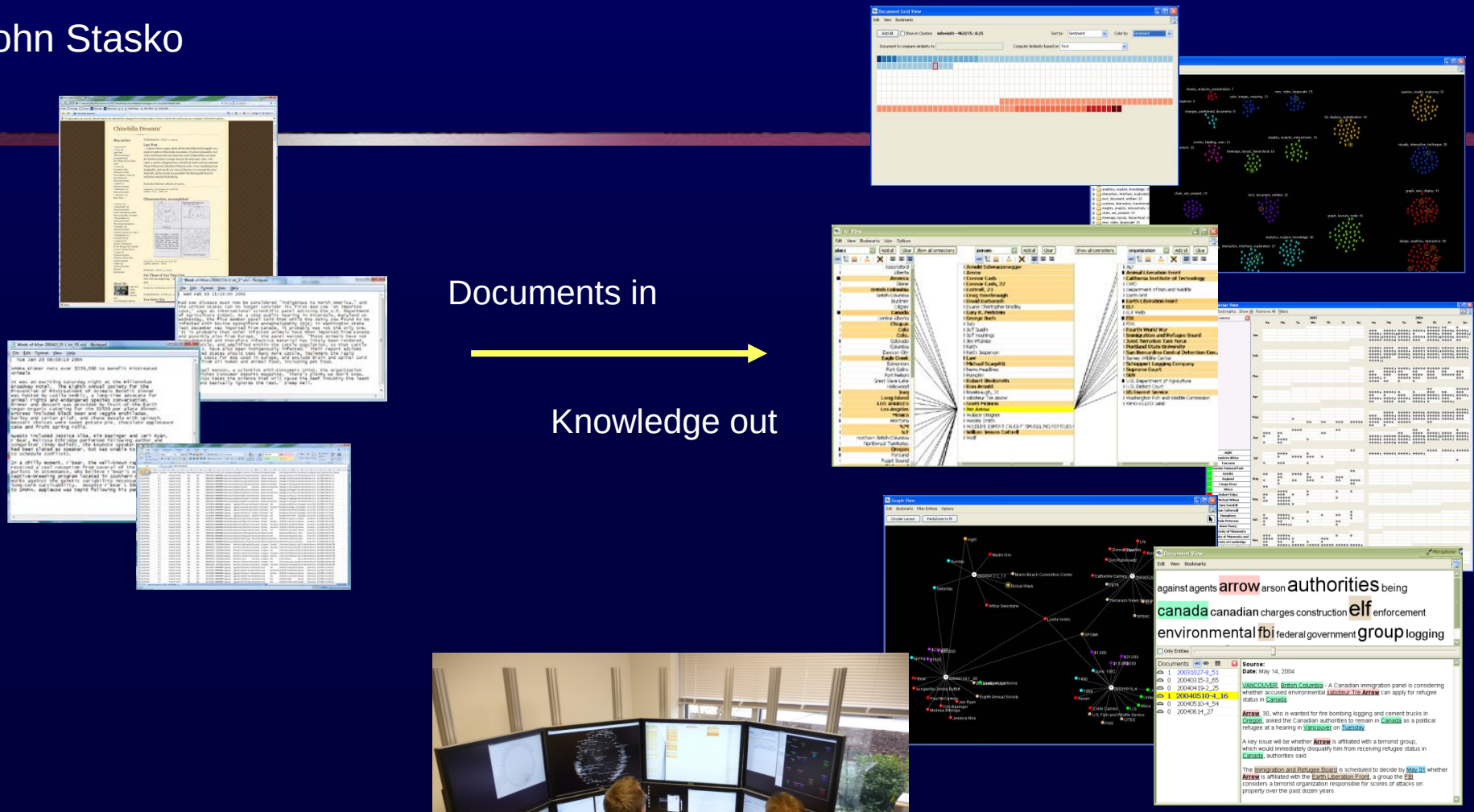


- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  combine interactive VA with complex simulation models

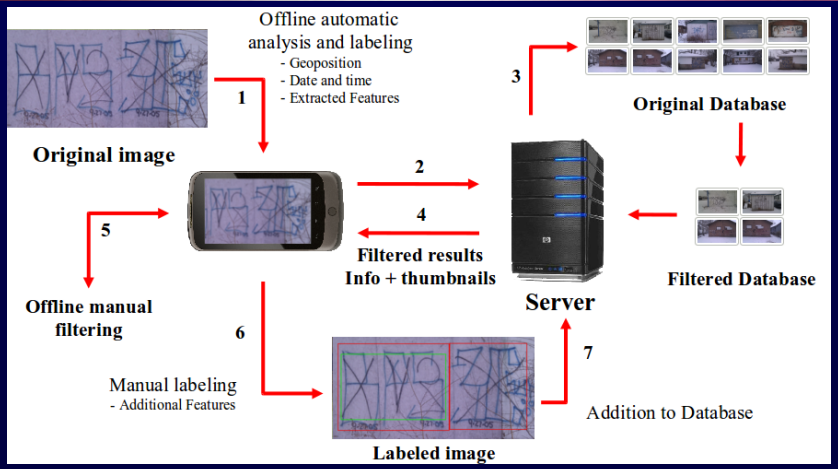


Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

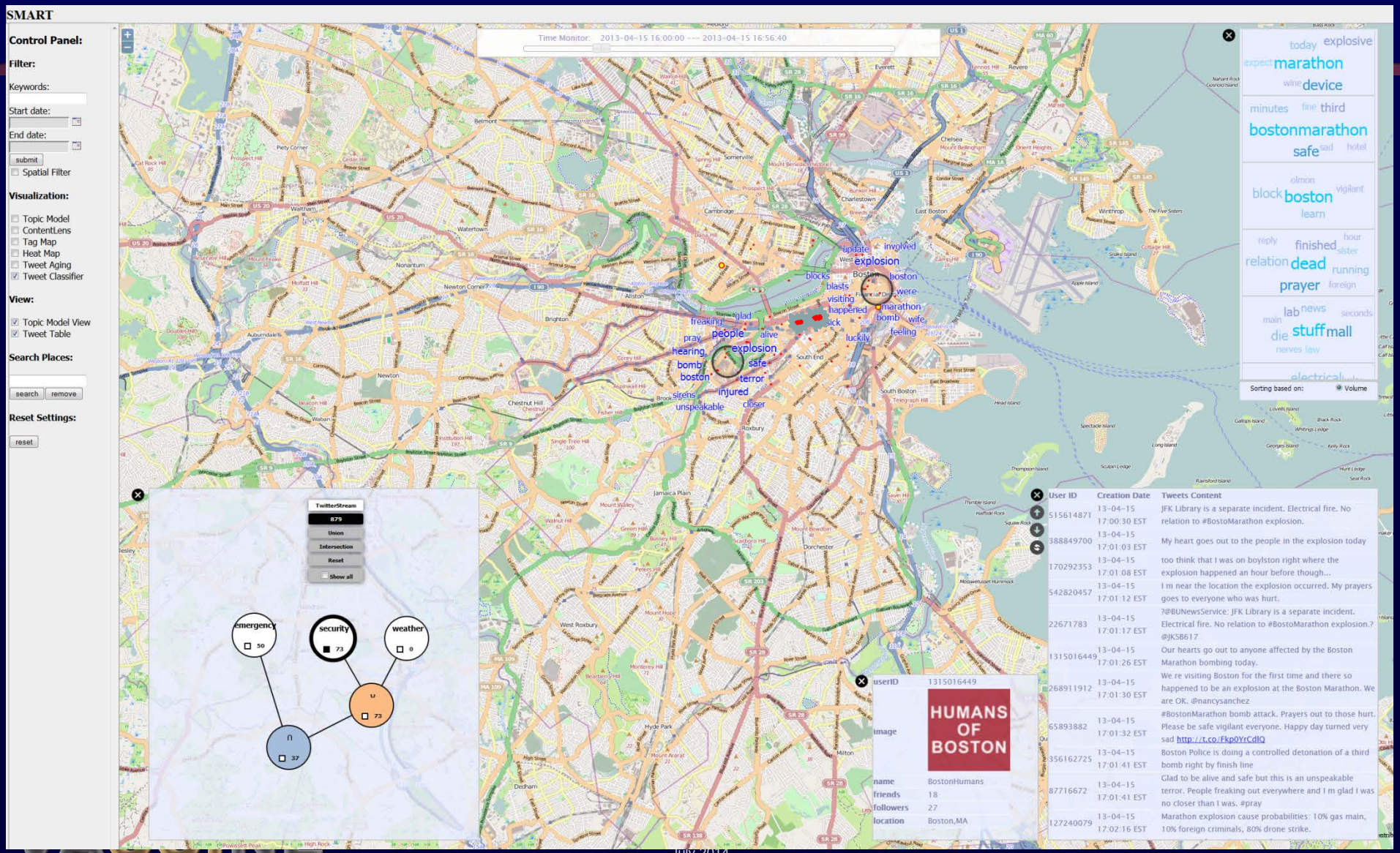
Social Media: Real-time Twitter Visual Analytics

(Purdue, Stuttgart, Penn St.)

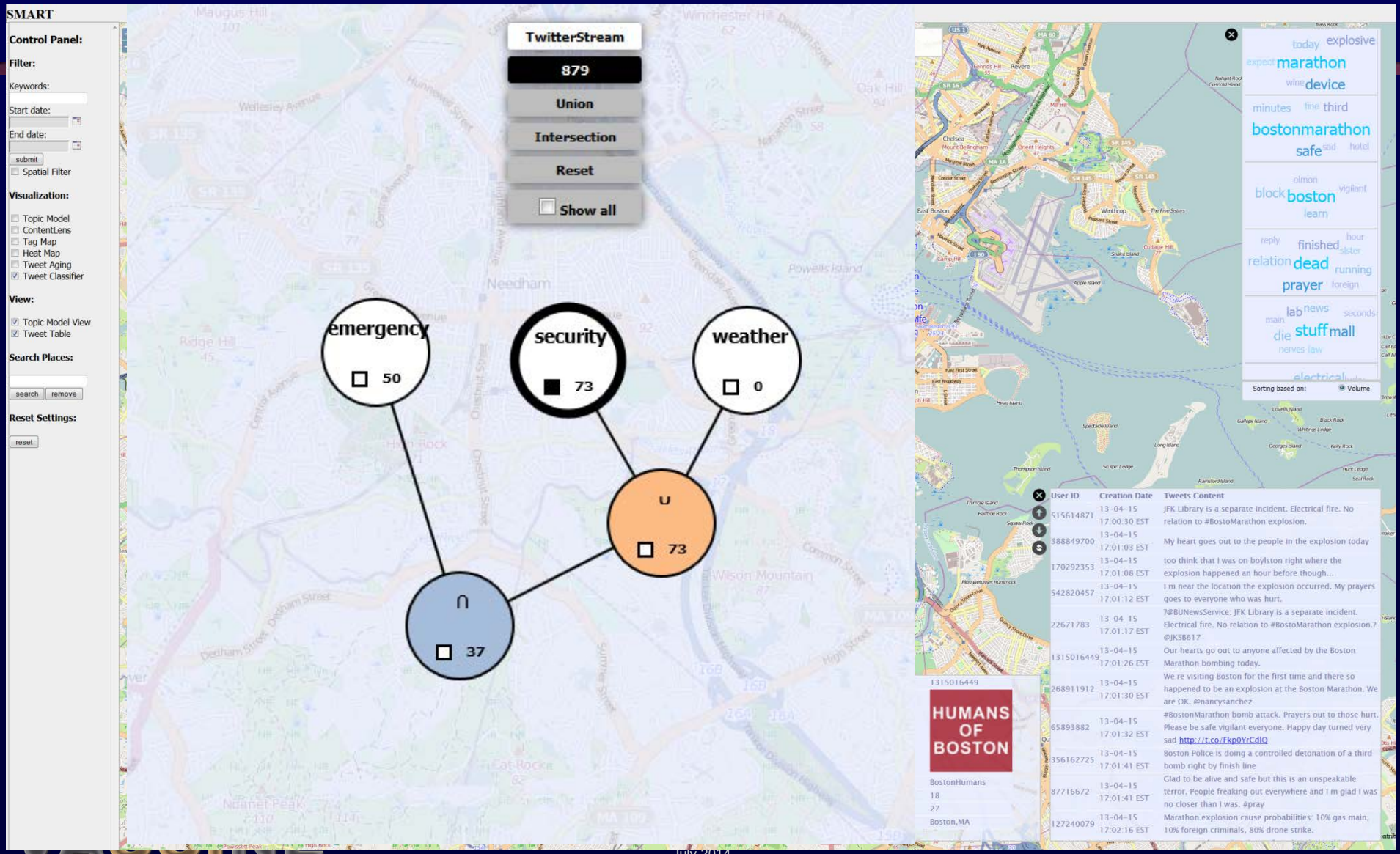
- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud



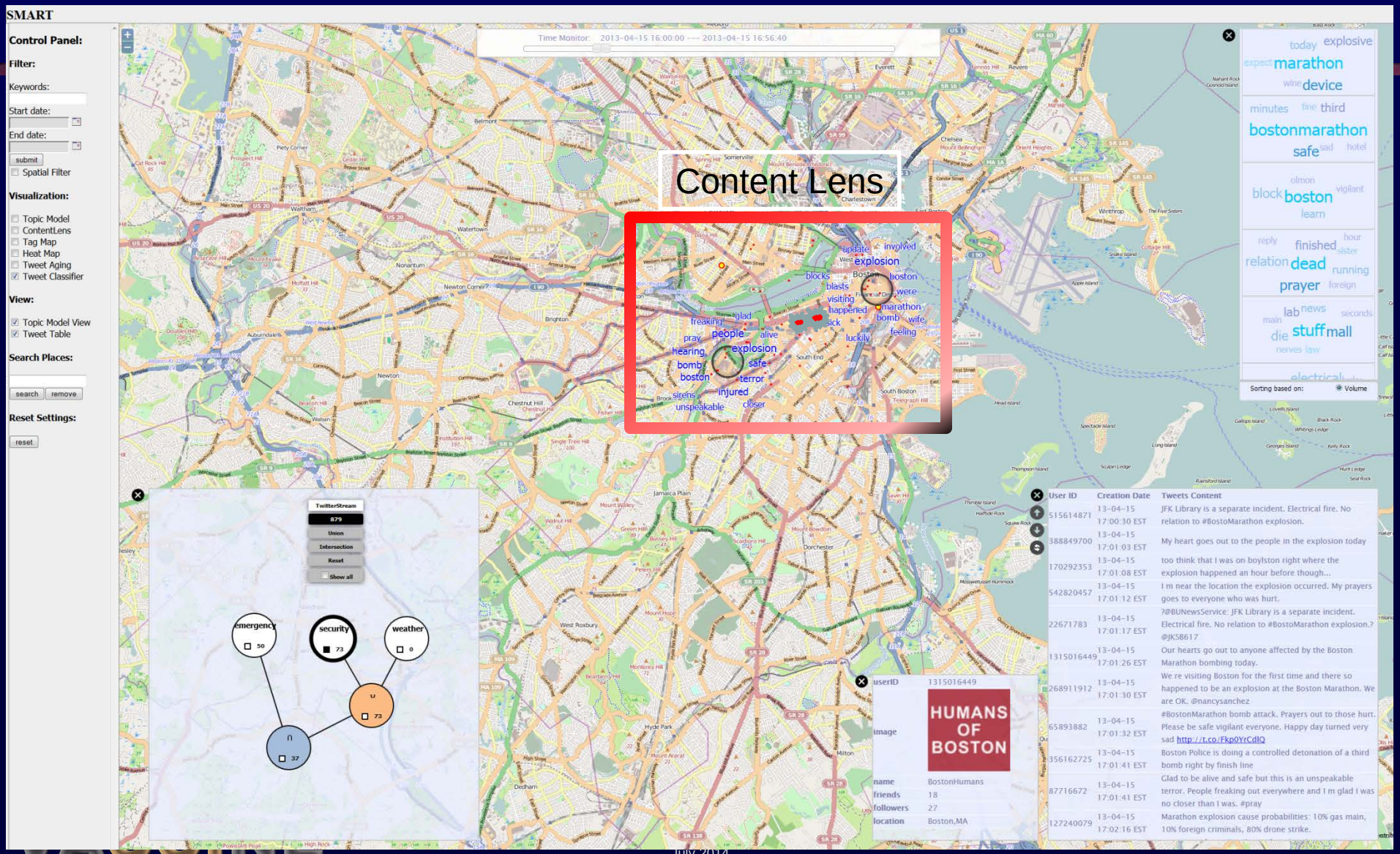
Social Media Analysis and Reporting Toolkit (SMART)



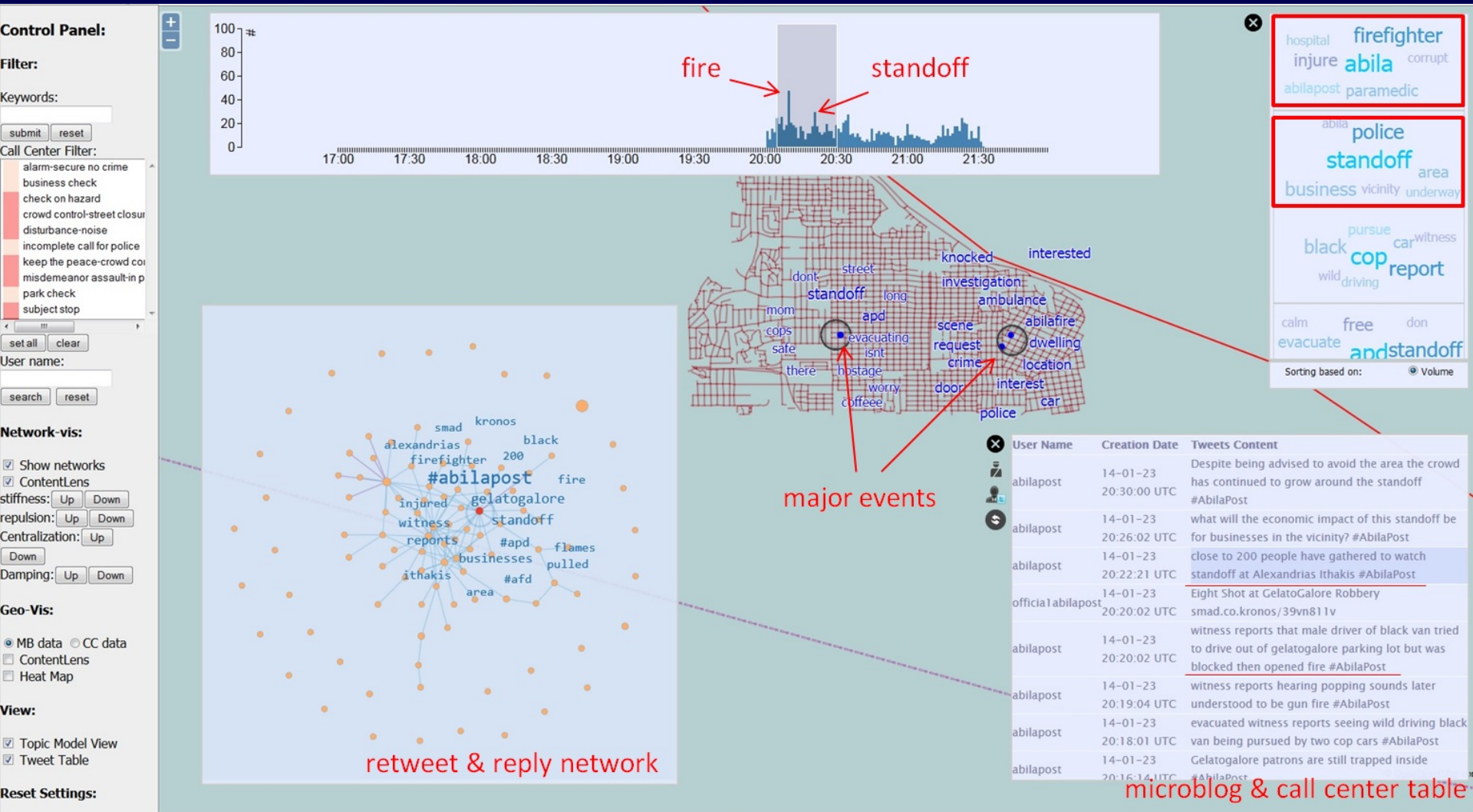
Social Media Analysis and Reporting Toolkit (SMART)



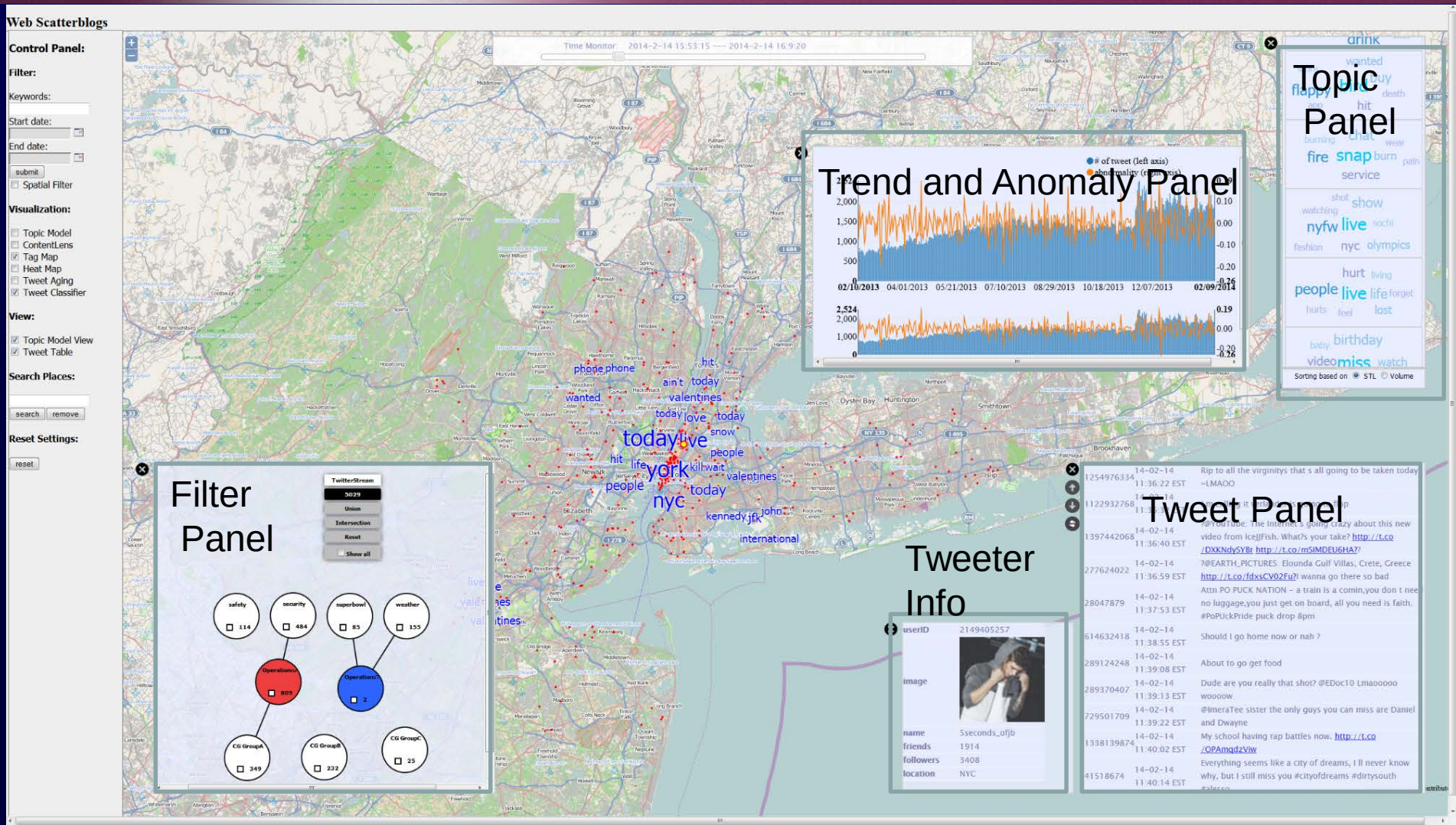
Social Media Analysis and Reporting Toolkit (SMART)



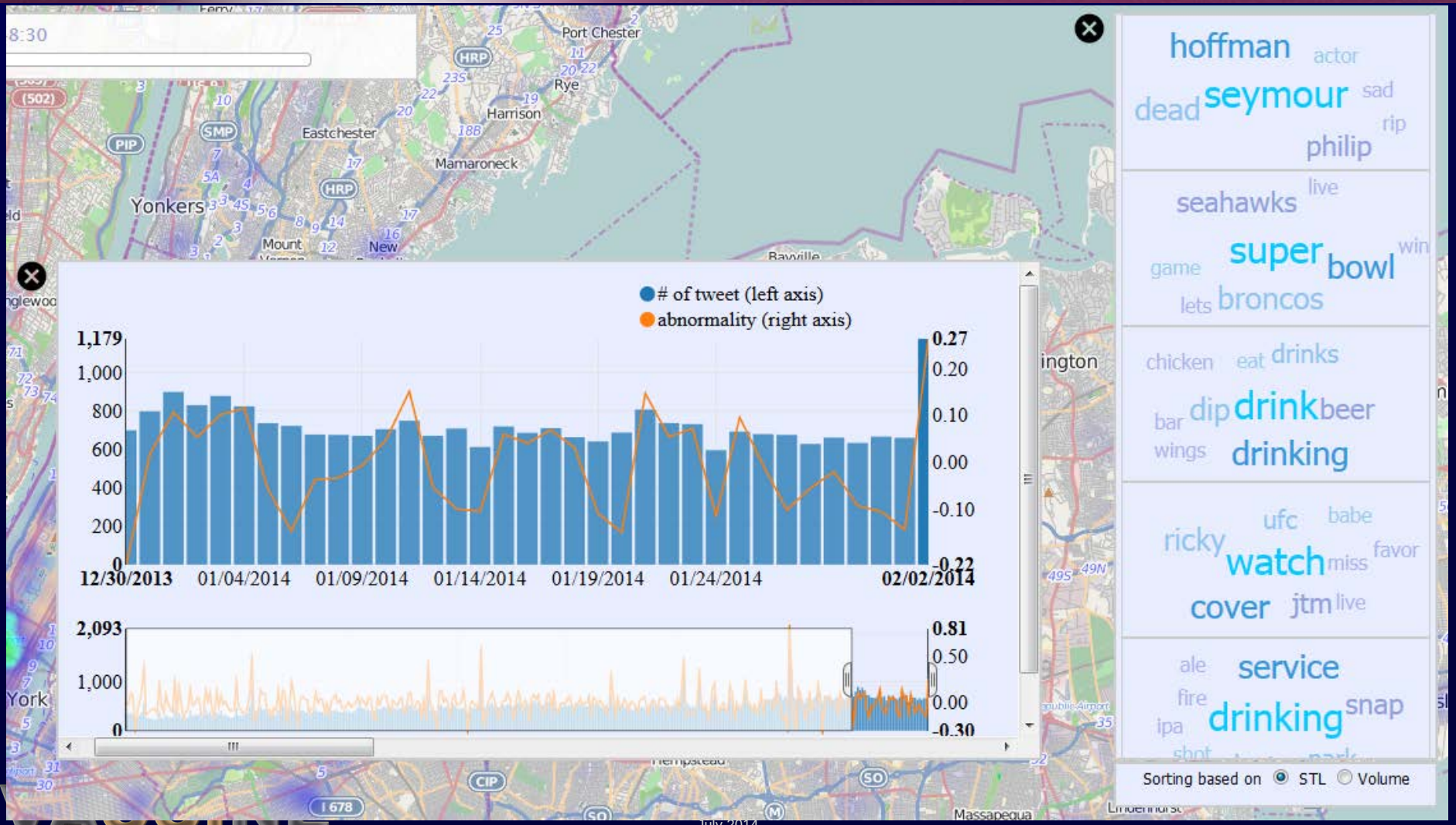
SMART: Network Linkage



SMART Anomaly Detection

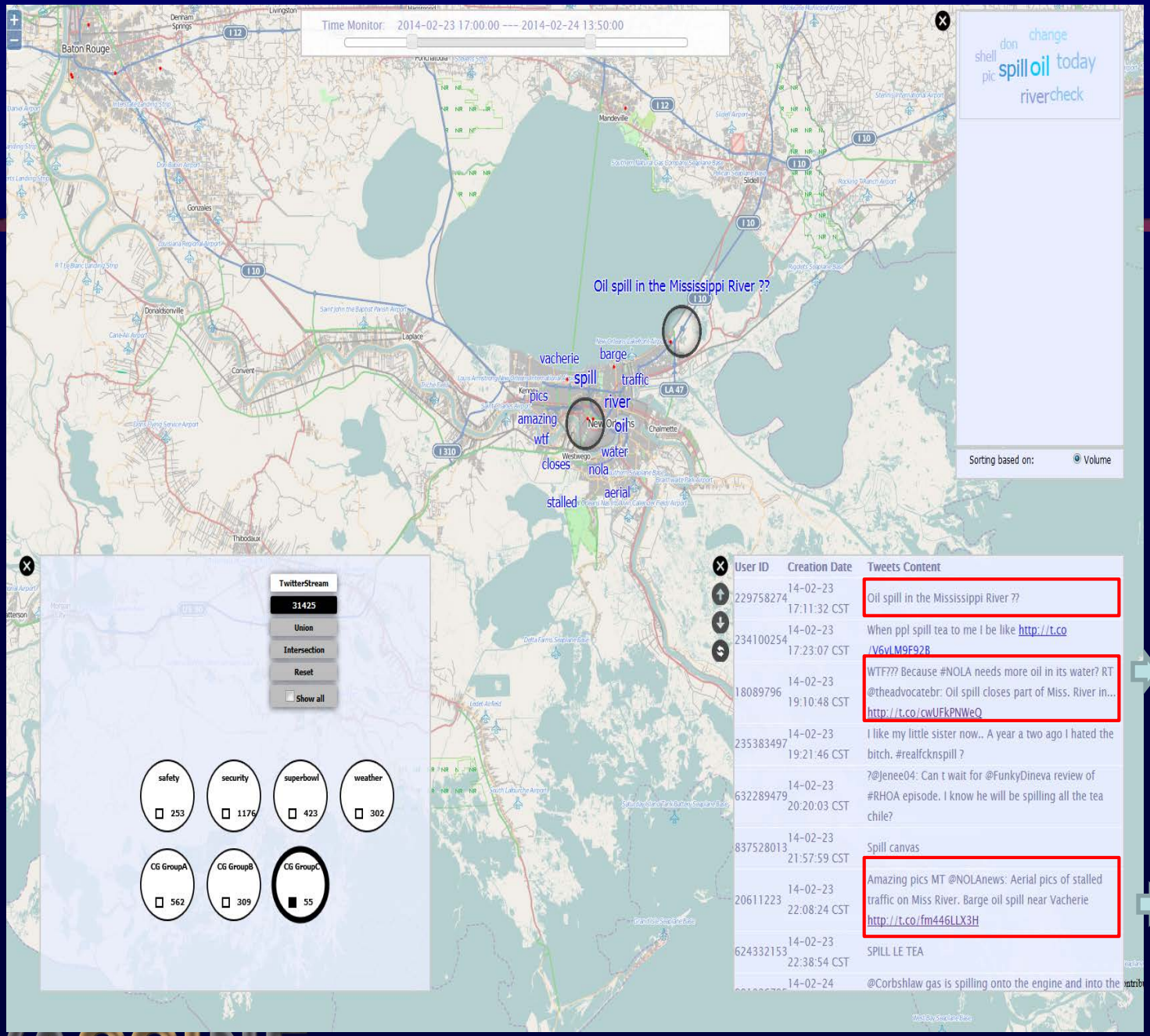


Anomaly Detection

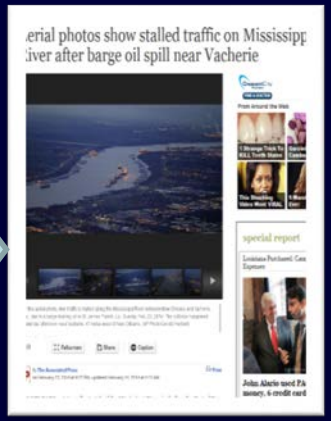


Oil Spill in Mississippi River

2014/02/23


Time Monitor: 2014-02-23 17:00:00 --- 2014-02-24 13:50:00
change
don
shell
pic
spill
oil
today
rivercheck
Oil spill in the Mississippi River ??
vacherie
barge
spill
traffic
pics
river
amazing
oil
wtf
water
closes
nola
aerial
stalled
Sorting based on:
Volume
TwitterStream
31425
Union
Intersection
Reset
Show all
safety
253
security
1176
superbowl
423
weather
302
CG GroupA
562
CG GroupB
309
CG GroupC
55
User ID
Creation Date
Tweets Content
229758274
14-02-23 17:11:32 CST
Oil spill in the Mississippi River ??
234100254
14-02-23 17:23:07 CST
When ppl spill tea to me I be like http://t.co/V6vLM9F92B
18089796
14-02-23 19:10:48 CST
WTF?? Because #NOLA needs more oil in its water? RT @theadvocatebr: Oil spill closes part of Miss. River in... http://t.co/cwUFkPNWeQ
235383497
14-02-23 19:21:46 CST
I like my little sister now.. A year a two ago I hated the bitch. #realfcknspill ?
632289479
14-02-23 20:20:03 CST
?@Jenee04: Can t wait for @FunkyDineva review of #RHOA episode. I know he will be spilling all the tea chile?
837528013
14-02-23 21:57:59 CST
Spill canvas
20611223
14-02-23 22:08:24 CST
Amazing pics MT @NOLAnews: Aerial pics of stalled traffic on Miss River. Barge oil spill near Vacherie http://t.co/fm446LLX3H
624332153
14-02-23 22:38:54 CST
SPILL LE TEA
14-02-24
@Corbshlaw gas is spilling onto the engine and into the


Oil spill closes part of Mississippi River in La.


Aerial photos show stalled traffic on Mississipp
River after barge oil spill near Vacherie

VACCINE

Non-geo Tweets On Twitter About Oil Spill Event

ProPublica @ProPublica · 3h
An **oil spill** just shut down 65 miles of the Miss. River. As production surges, how do we ship oil safely? propub.ca/1duUl9g @Bloomberg
View summary · Reply · Retweet · Favorite · More

Top news story
New Orleans LA @New_OrleansLA · 22h
Mississippi **Oil Spill** Highlights Risk of US Oil Boom - Bloomberg ift.tt/1ds9cRO #NewOrleans
Expand · Reply · Retweet · Favorite · More

Bloomberg News
Mississippi Oil Spill Highlights Risk of U.S. Oil Boom
A barge crash that spilled enough oil to temporarily shut a stretch of the Mississippi River highlights the transportation risks of the U.S. energy boom just as regulators respond to several rail...
View on web

Retweeted 19 times
Guardian Environment @guardianeco · Feb 24
Mississippi river closed at New Orleans after barge crash prompts **oil spill** gu.com/p/3n2q9/tf
View summary · Reply · Retweet · Favorite · More

Motherboard @motherboard · 4h
The Mississippi already a massive toxic cesspool every other day of the year: bit.ly/1flVUpJ
View summary · Reply · Retweet · Favorite · More

Marcy McGowan @marcylauren · 5h
Oil Spill Shuts Down Sixty-Five Mile Stretch of Mississippi River | Americans Against the Tea Party aattp.org/oil-spill-shut...
View summary · Reply · Retweet · Favorite · More

EDF Louisiana @EDF_Louisiana · 8h
#**oilspill** update: Lower Mississippi River reopens after **oil spill** near New Orleans. Watch the video bit.ly/MZB7ST
View summary · Reply · Retweet · Favorite · More

Sierra Club @sierraclub · 9h
The Mississippi River Was a Toxic Mess Long Before this Weekend's **Oil Spill**: sc.org/MSriverspill (via @motherboard)
Expand · Reply · Retweet · Favorite · More

Motherboard
The Mississippi River Was a Toxic Mess Long Before the Oil Spill
Everyone is outraged about a barge spilling oil into the Mississippi—perhaps everyone forgot that it is a massive toxic cesspool every other day of the year, too.
View on web



National Wildlife @NWF · 12h
A barge has crashed into a tugboat, spilling 31,500 gallons of oil into the Mississippi River. Read more: bit.ly/Mrrz2d
Expand · Reply · Retweet · Favorite · More

Margarita Noriega™ @margafret · 13h
31,500 gallons of light crude oil spilled into the Mississippi River weather.com/news/science/e...
View summary · Reply · Retweet · Favorite · More

National Wildlife @NWF · 14h
New **Oil Spill** Fouls Mississippi River: bit.ly/1gzqS01
Expand · Reply · Retweet · Favorite · More

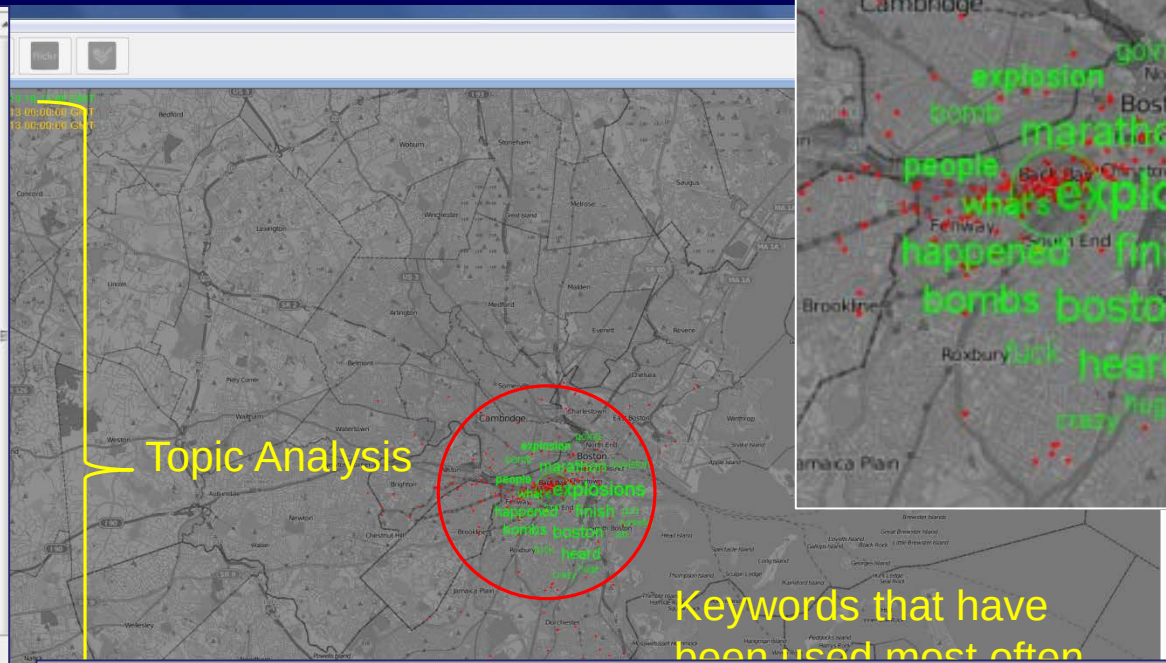
inhabitat @inhabitat · Feb 24
It is unclear when the ships stuck on the Mississippi River will be able to move as crews clean up a new **oil spill** bit.ly/1dpdo4V
Expand · Reply · Retweet · Favorite · More

inhabitat
Louisiana Oil Spill Shuts Down New Orleans Port and Parts of the...
The Port of New Orleans and surrounding areas were closed Sunday after a barge collided with a tow boat, causing oil to spill into the Mississippi River. The boats were located between Baton Rouge...
View on web



Explosion Area in Boston





Keywords that have been used most often

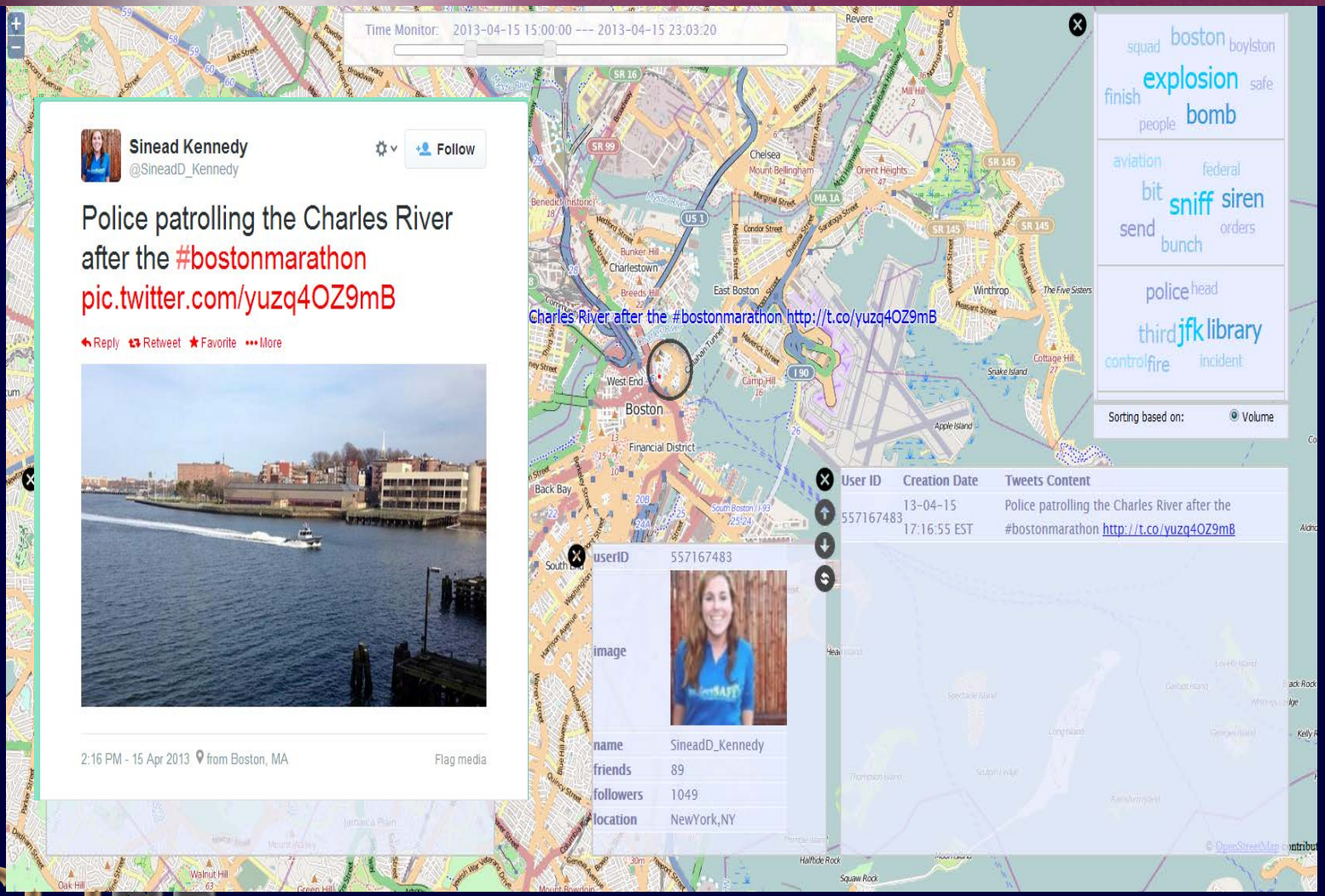
Text
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped.
two bombs just went off on boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finsh line this afternoon.
What the FUCK was that
Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalciO I figured 3 people would get the joke
Two explosions just rocked the finish line of the Boston Marathon. Sirens galore. People running in fear. Wonder what happened.
This is crazy i seen that blow up #bostonmarathon
can someone tell me what that explosion was!? #boston #bostonmarathon

Tweets

First Response (Tweet & Picture)



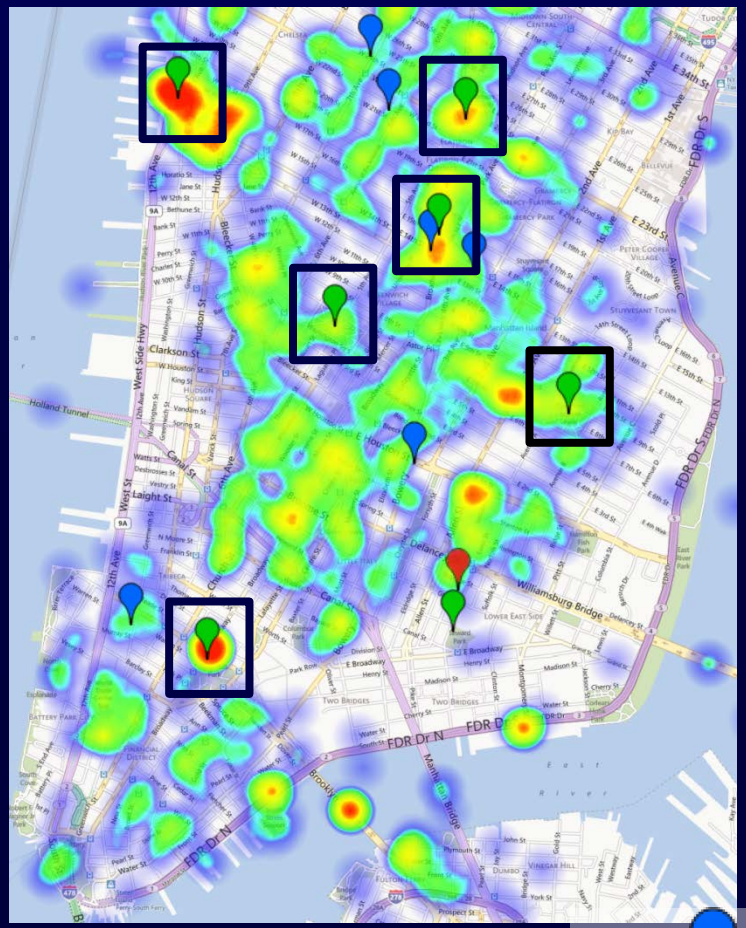
Police Patrolling The River

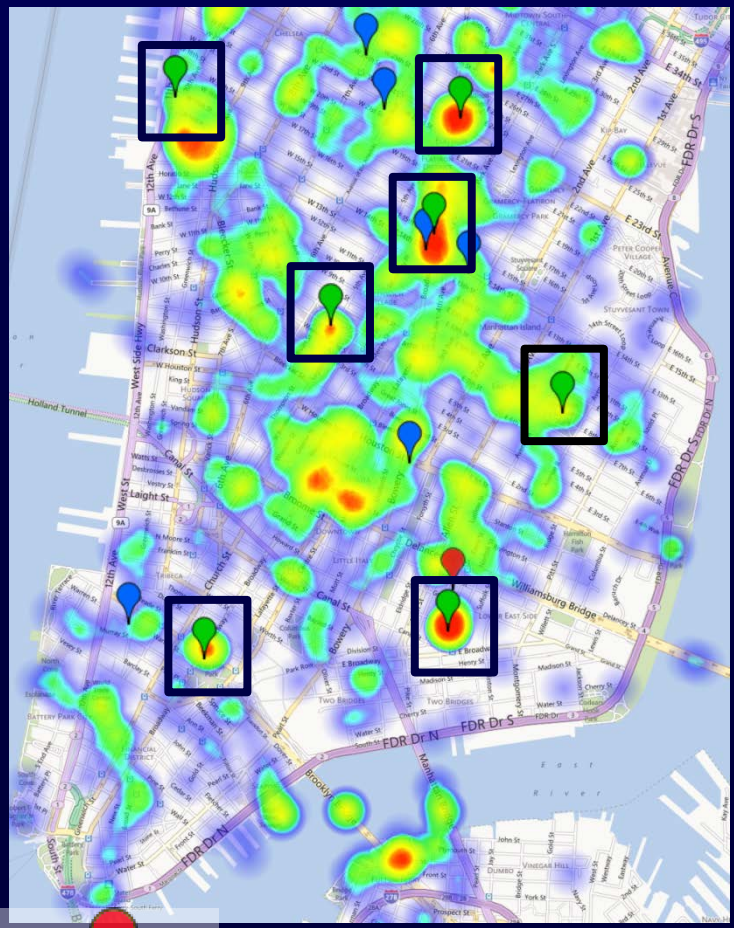


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

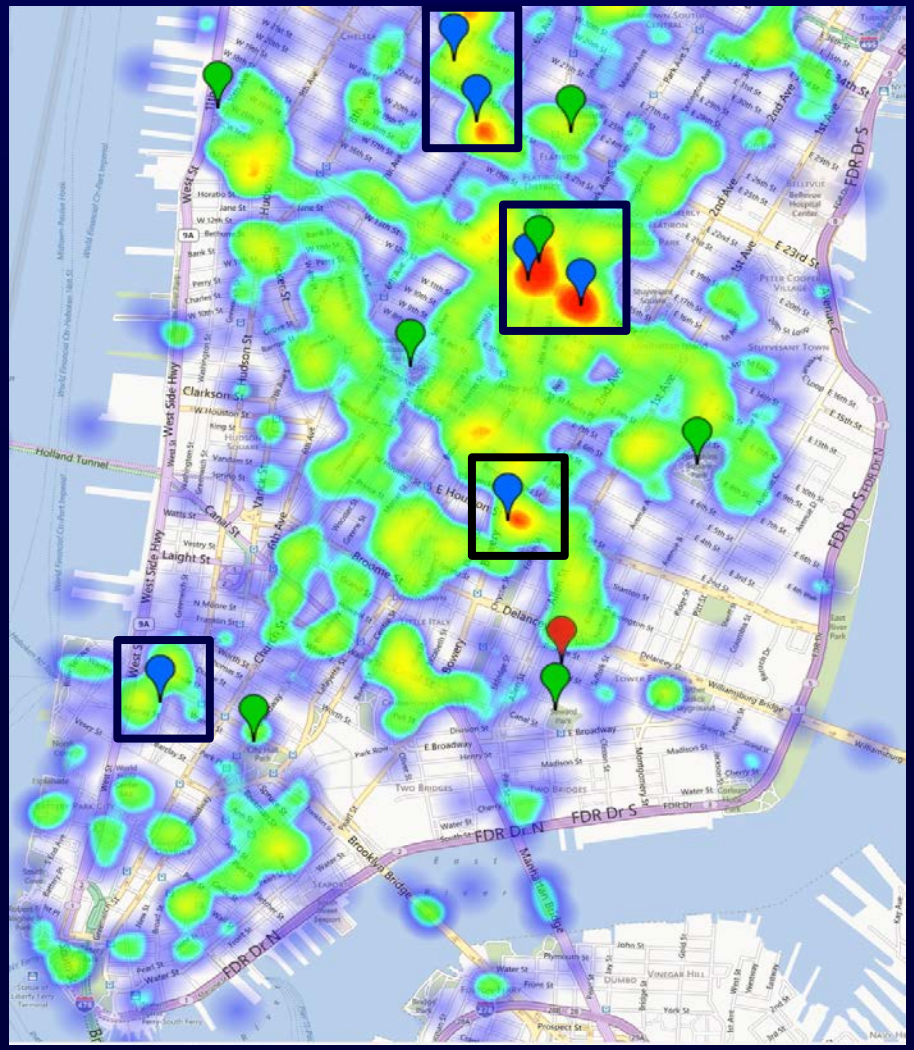
July 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

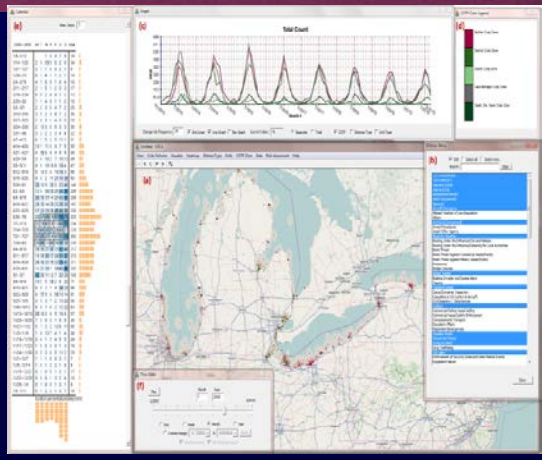
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



VACCINE

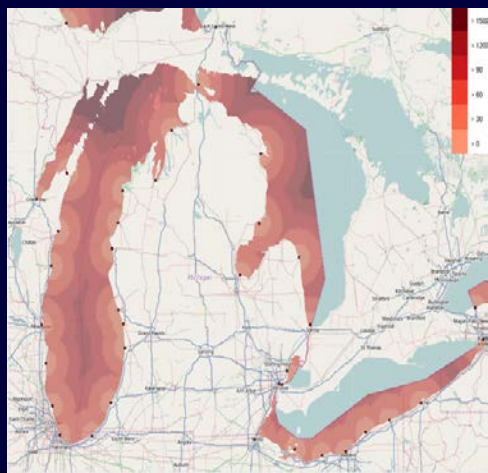
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771



IMPACTS:

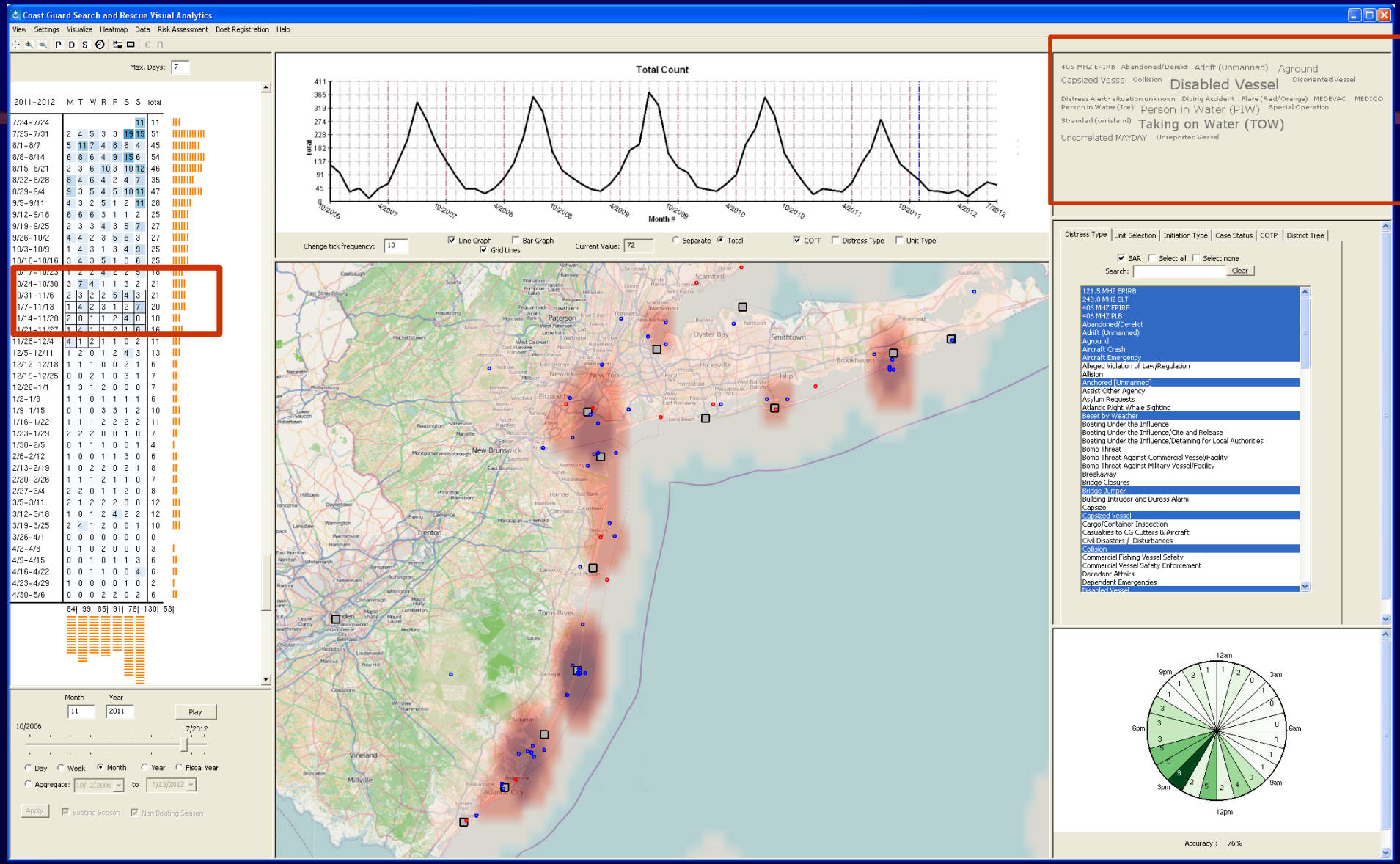
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



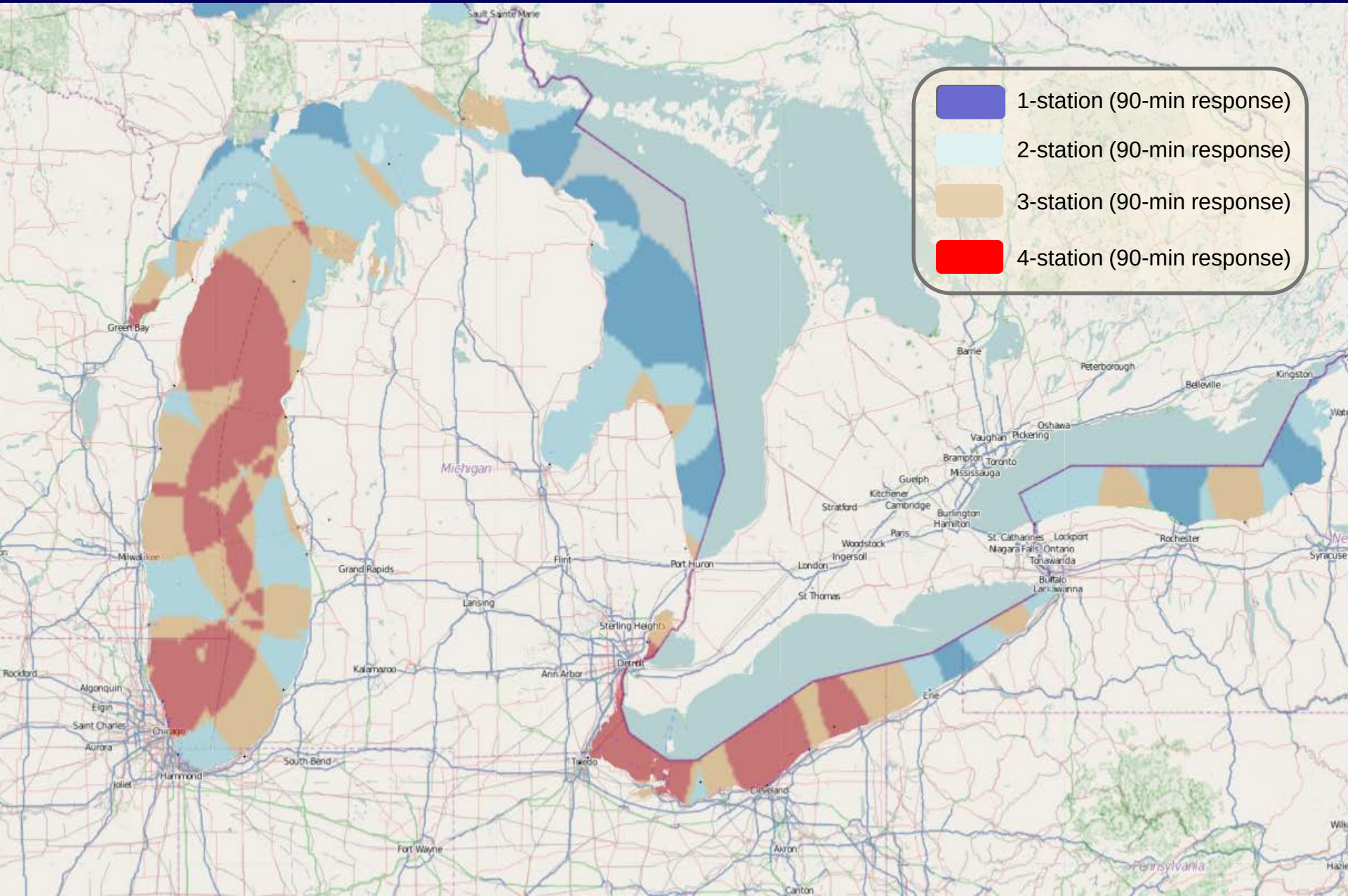
VACCINE

Example: Risks and Consequences From Sandy:

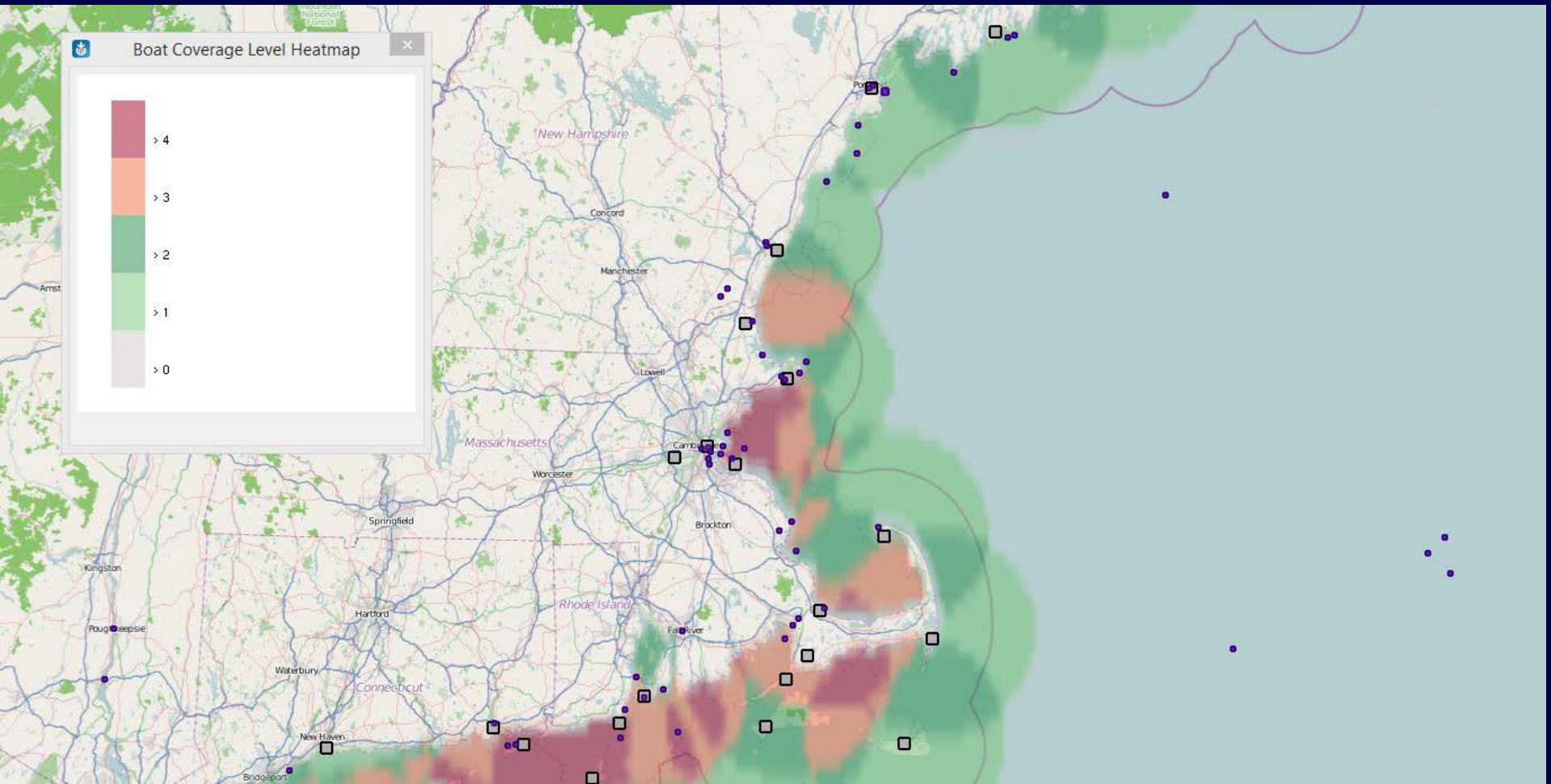
SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



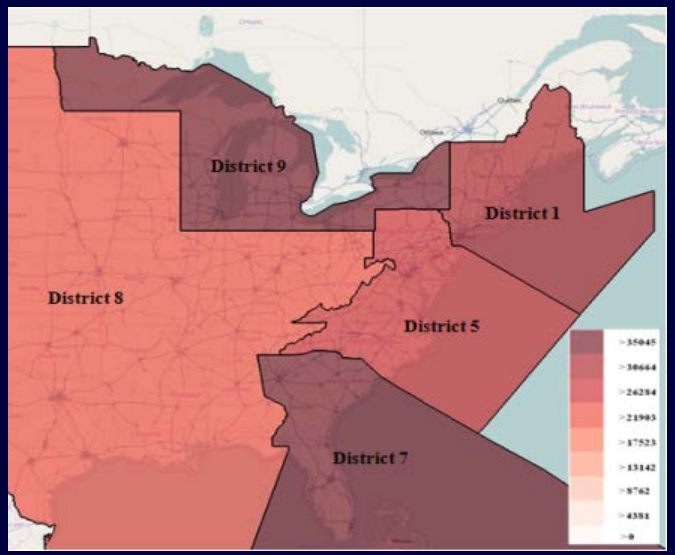
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



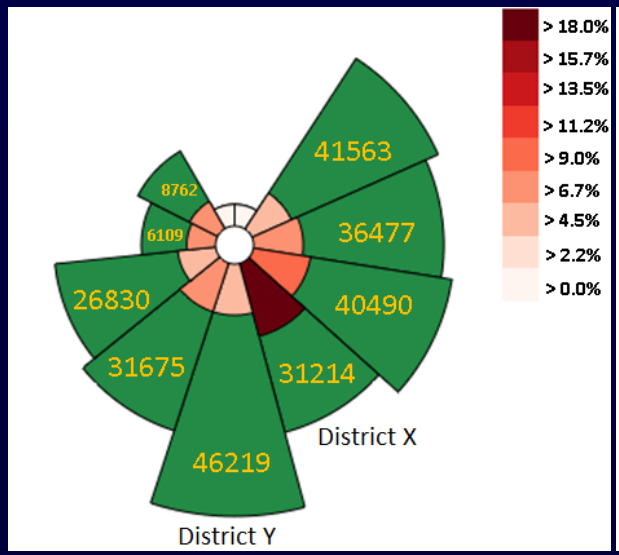
ORAM Visualization

Component Overview: Identification of Potential Risks

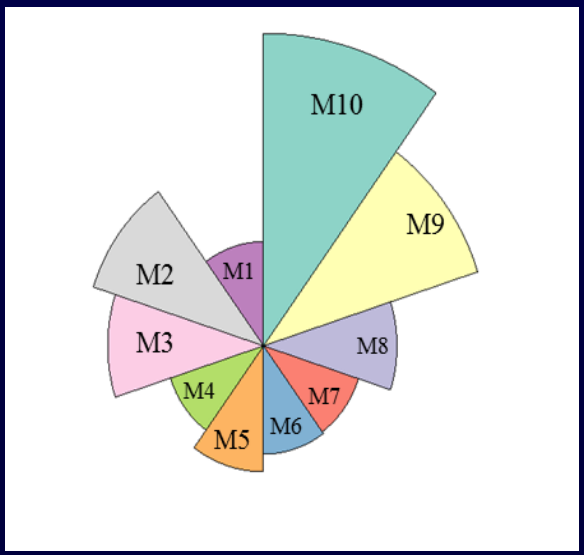
General understanding on how the risk is distributed between the different areas and missions.



(A) District Choropleth Map for Total Risk across all missions



(B) Risk profiles across all districts and mission areas



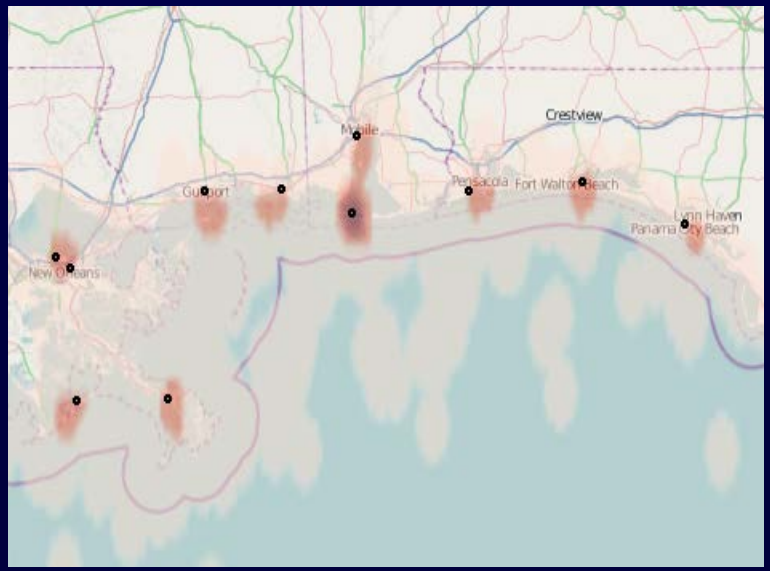
(C) District X Risk Profiles across all mission area

Component Overview: Analysis and Assessment

Specific areas can be analyzed by time and details on demand.



(D) Total M10 Risk Values distributed among District X's station AOR



(E) KDE heatmap showing Total M10 Risk values

Interactive Operational Performance Assessment Report (iOPAR)

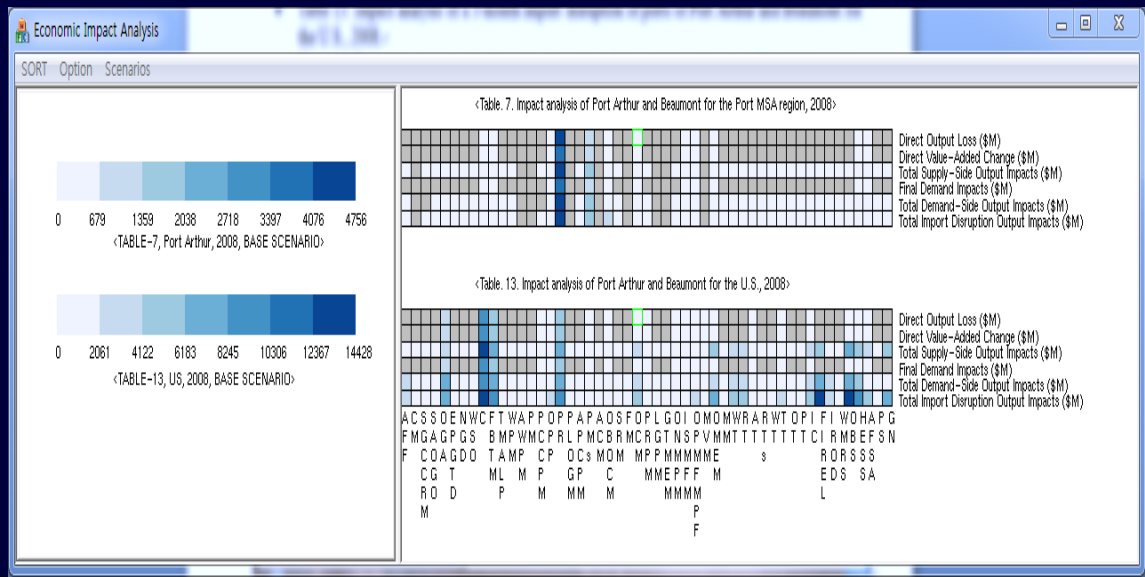
- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels

USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



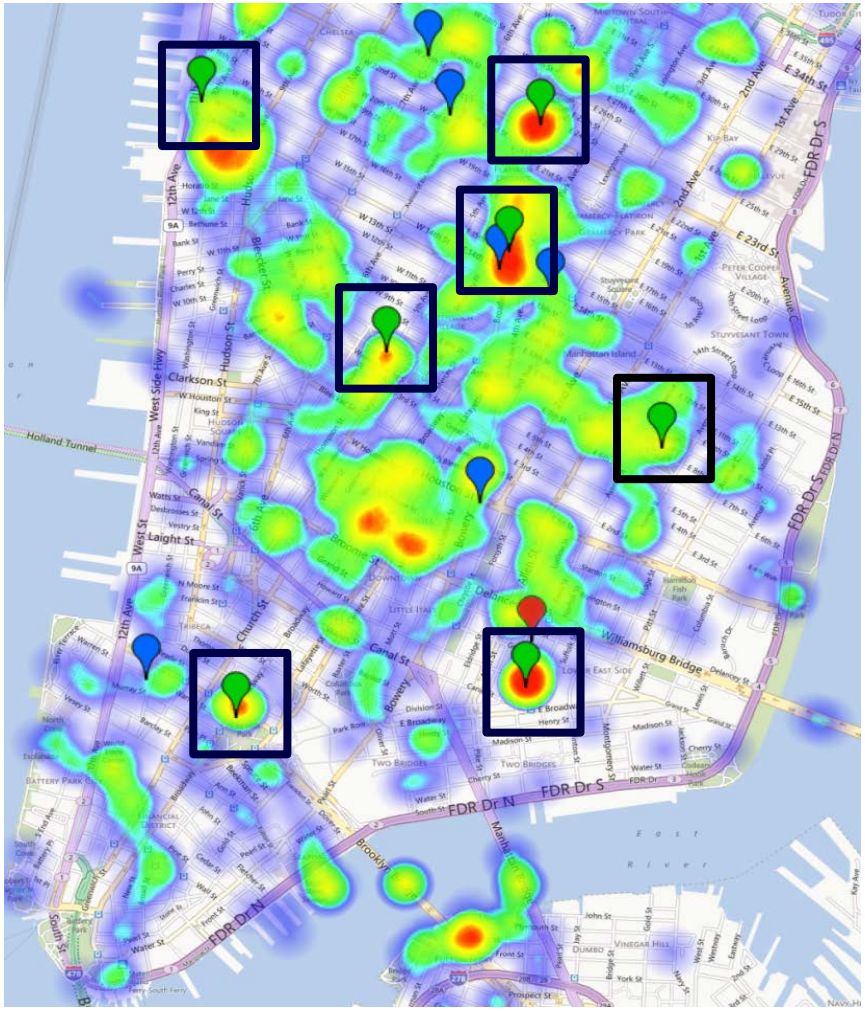
VACCINE

For Further Information

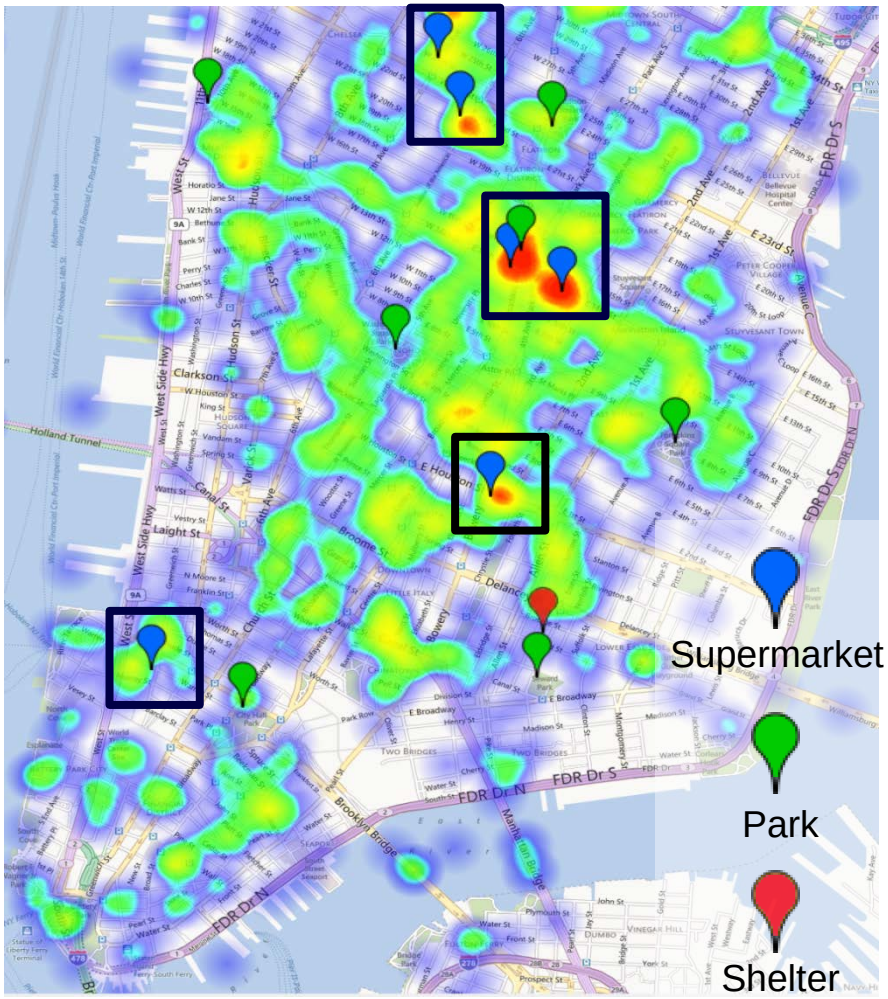
www.VisualAnalytics-CCI.org



One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00



After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Data-Driven Decision Making

David Ebert

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

August 2014

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Justice Institute of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (GA, IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America



VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**
  - Interdisciplinary world leading team of researchers and students

"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address

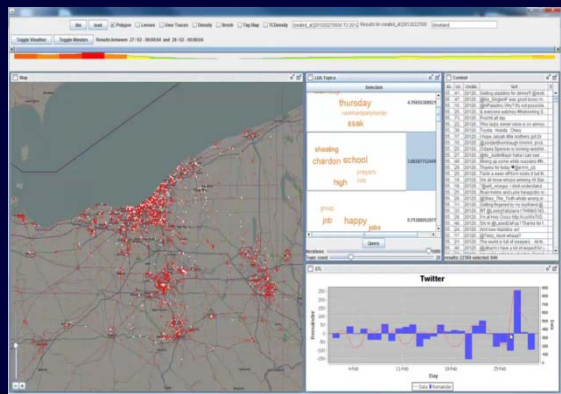


VADM Robert Parker with VACCINE student researchers (cgSARVA, COAST, iOPAR)

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation

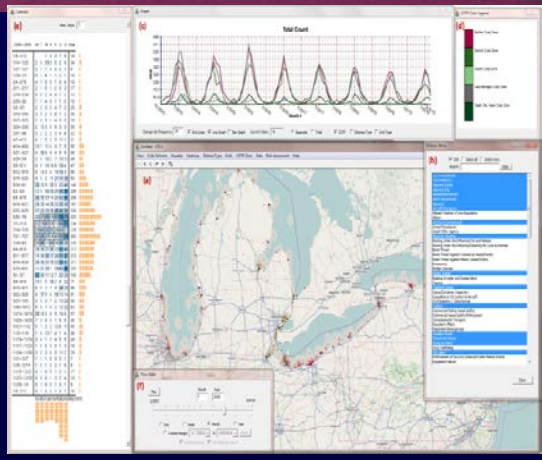


Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

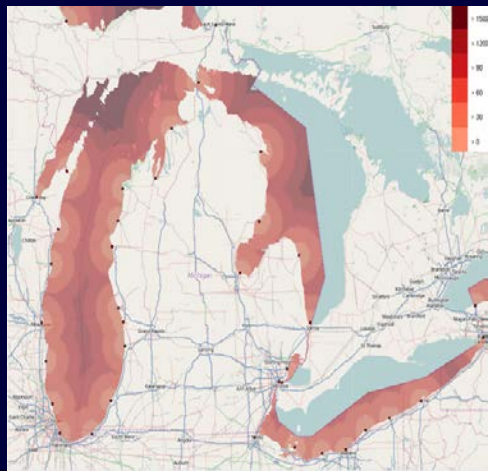
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

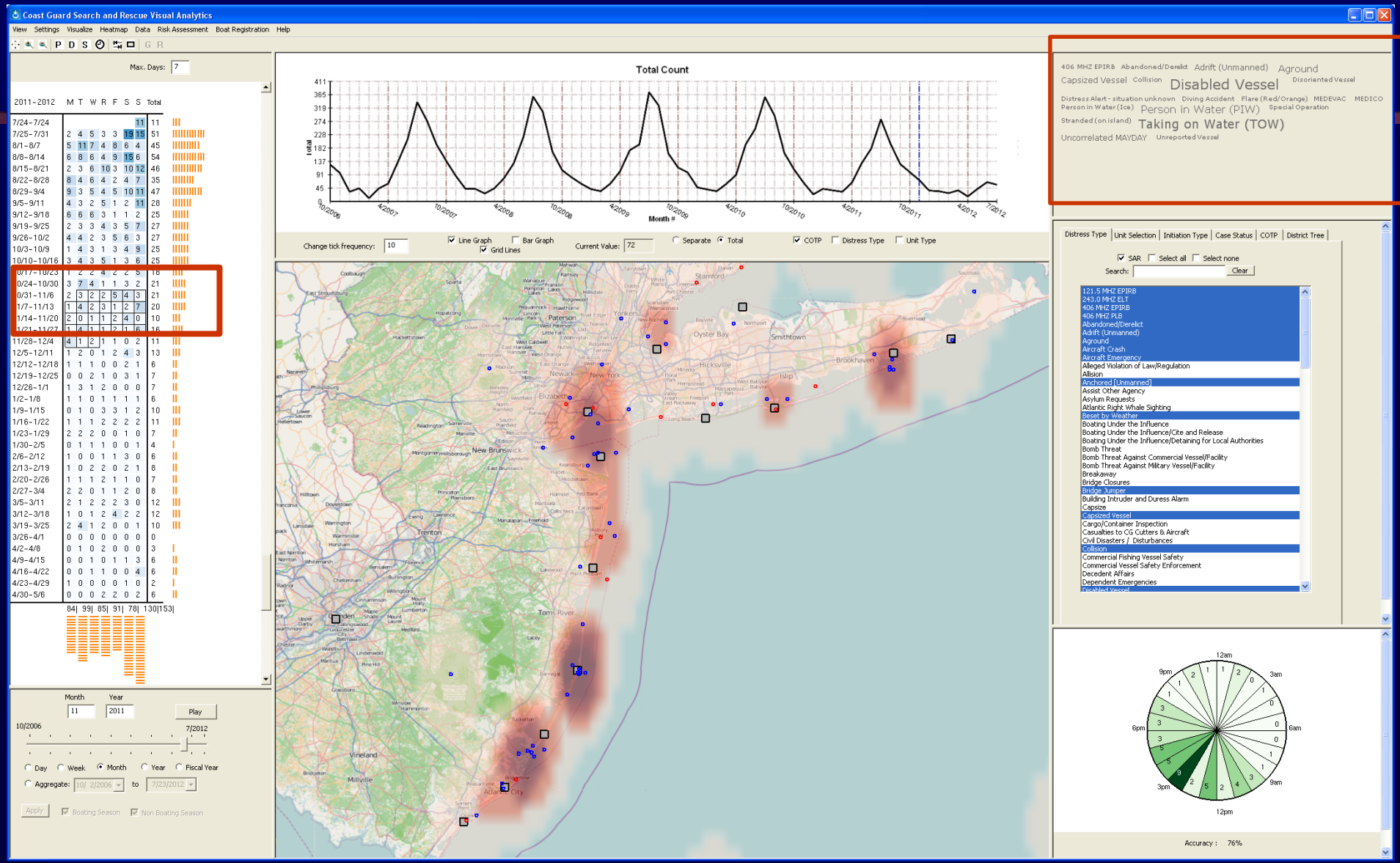
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



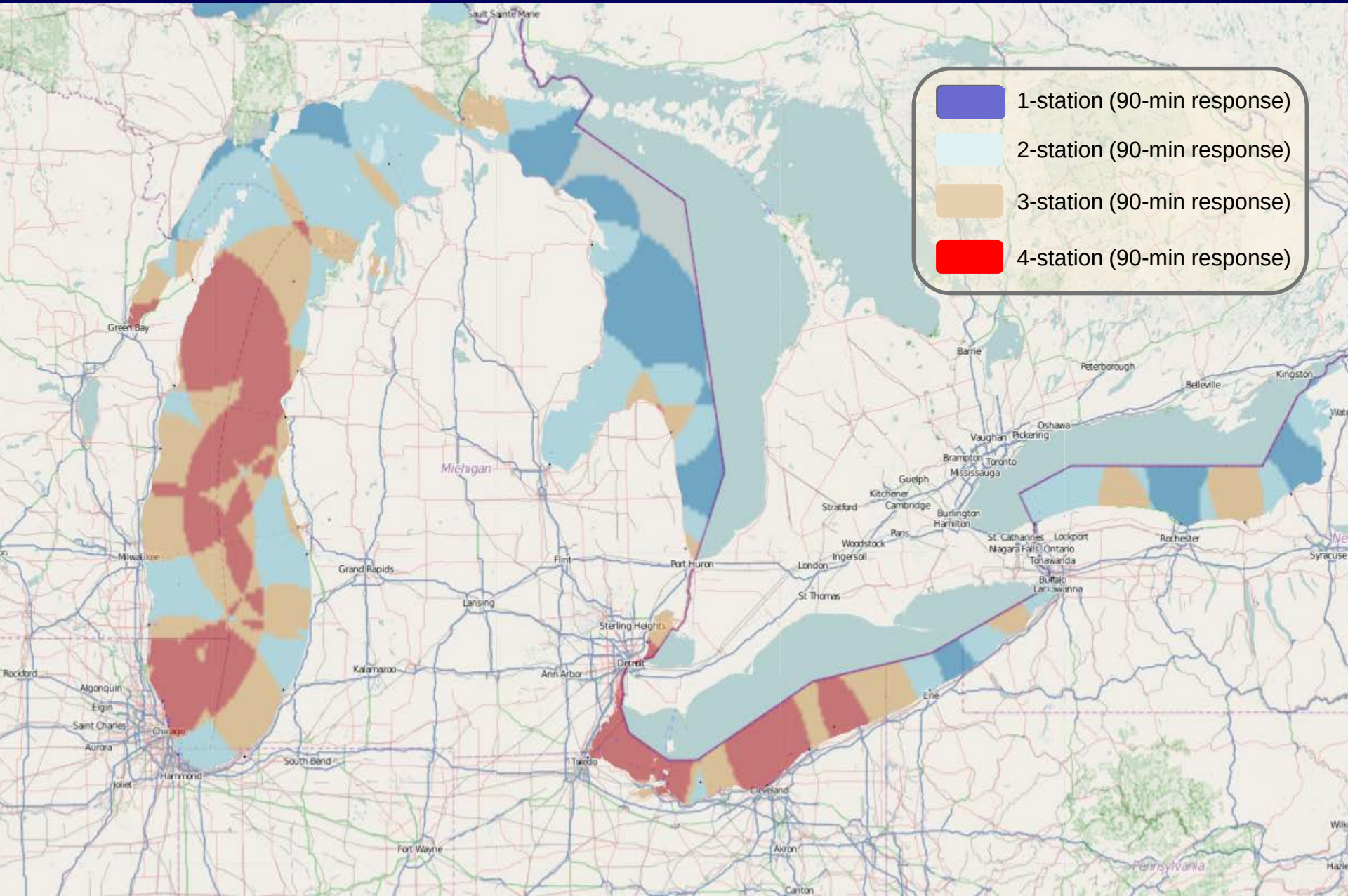
VACCINE

Example: Risks and Consequences From Sandy:

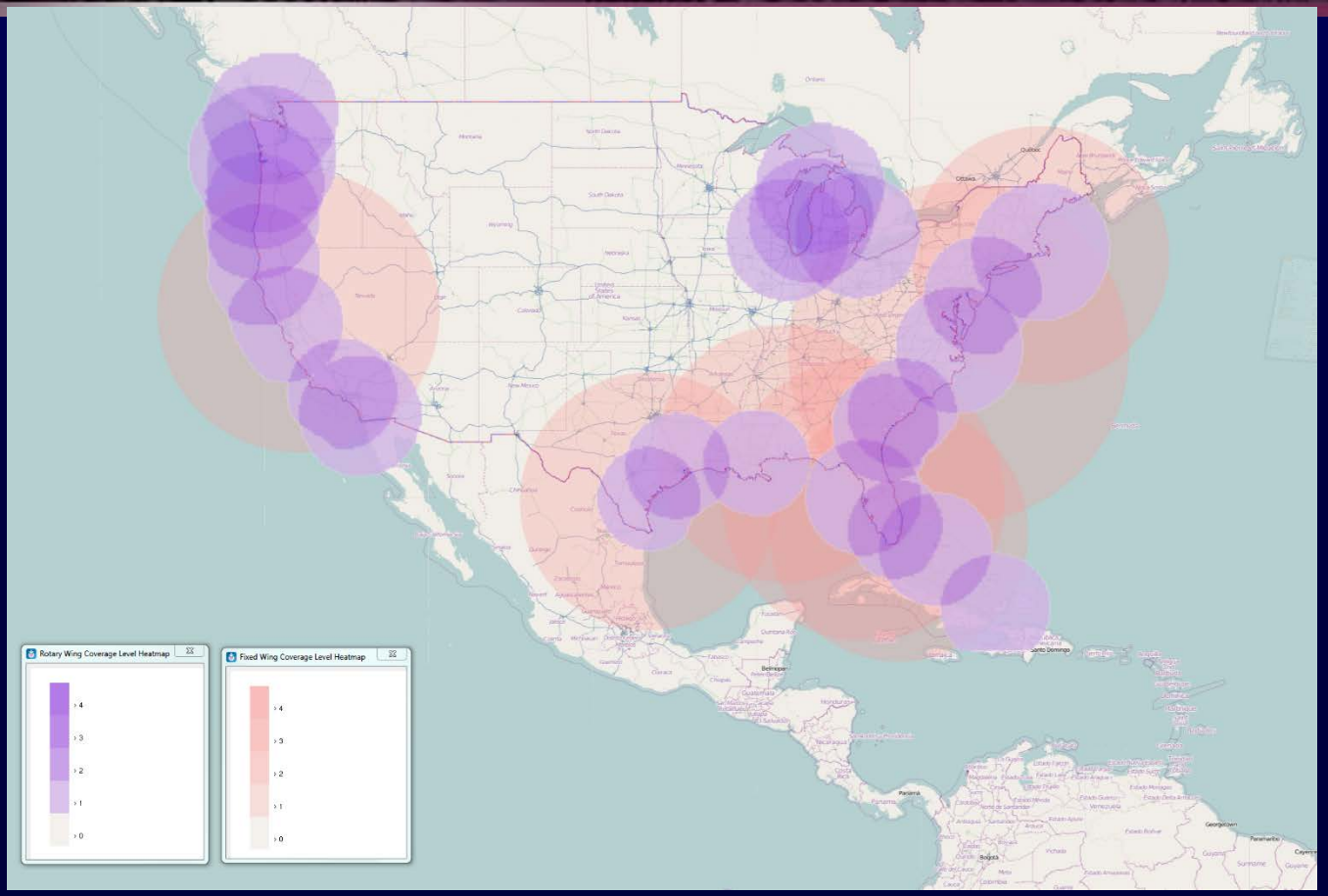
SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



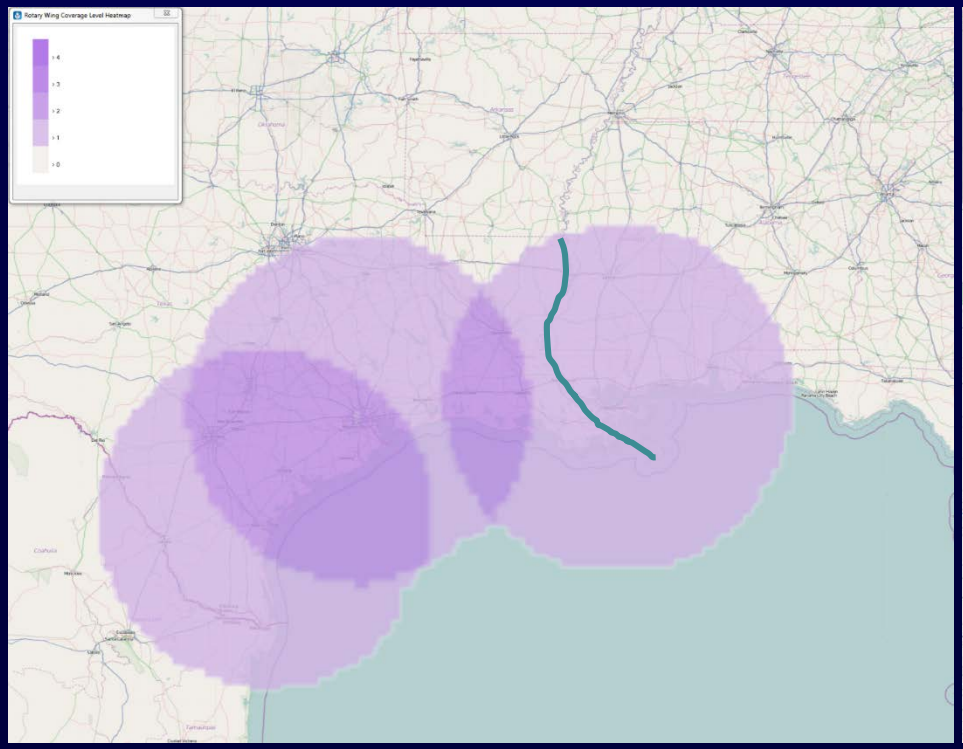
cgSARVA Results – Aircraft coverage

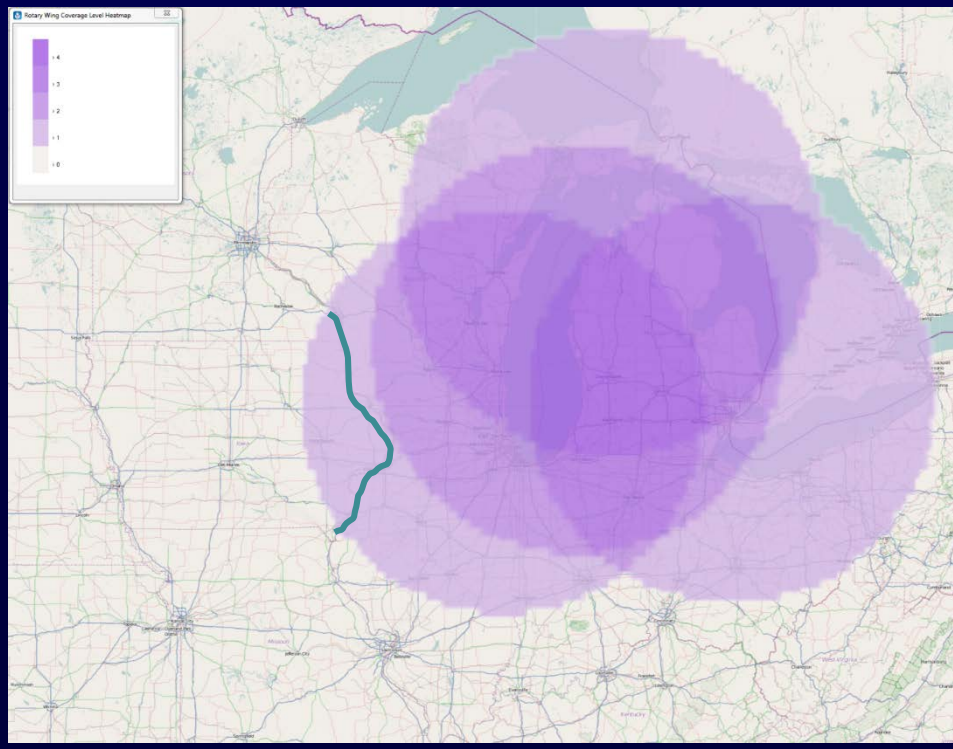


Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.

Airstation Utility Visualization Comparison

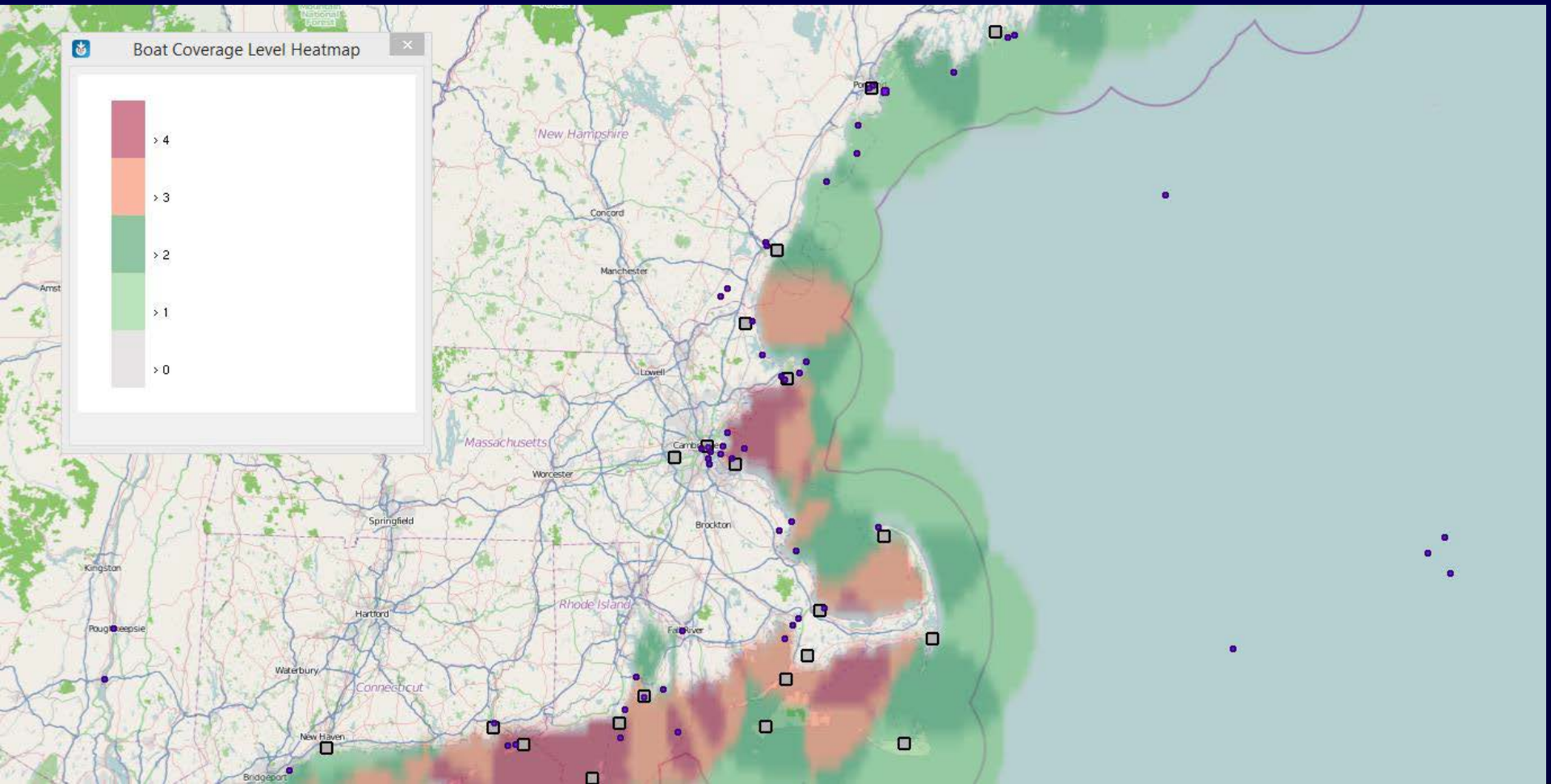
- Rotary Wing Coverage Map by AirSTA





Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



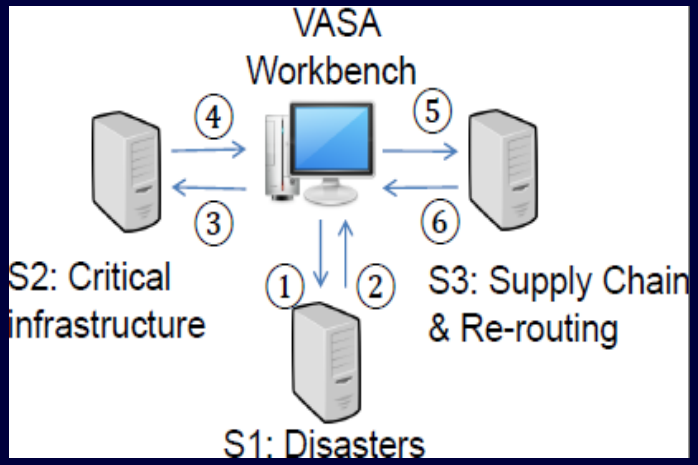
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

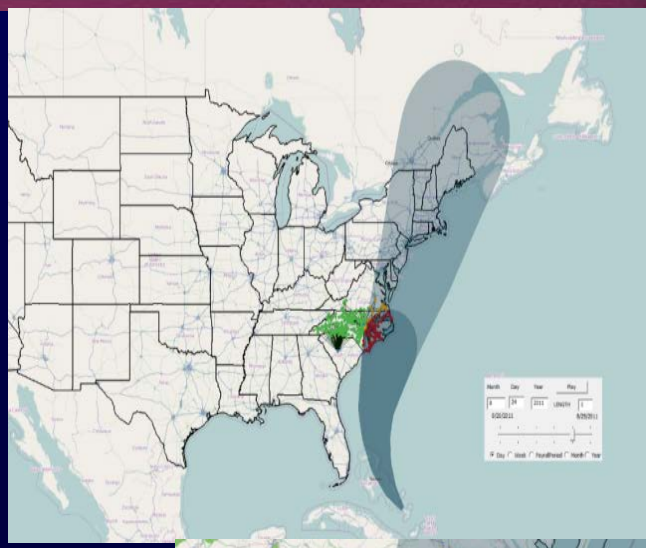
TACC

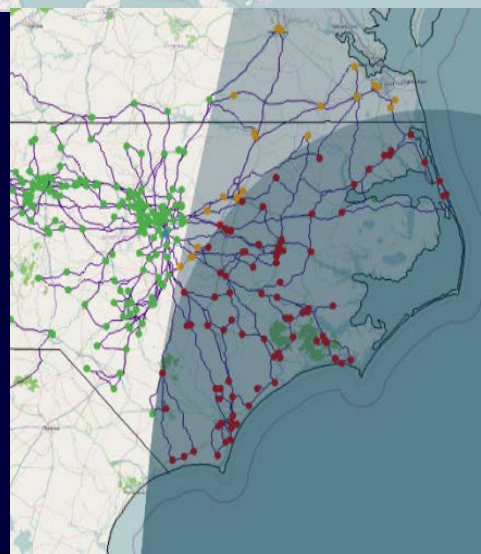
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





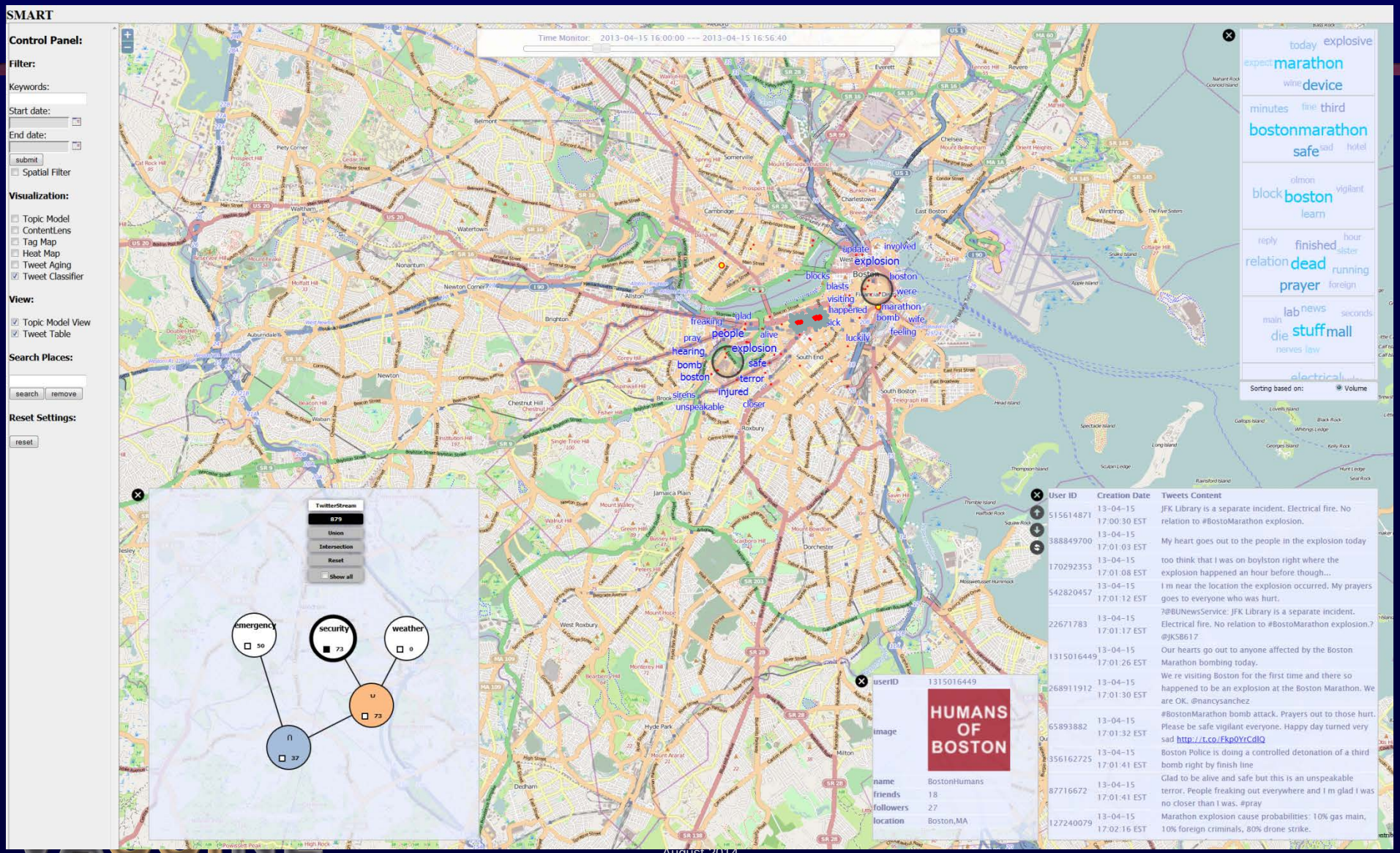
Social Media: Real-time Visual Analytics

(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- Topics extracted shown as a dynamic word cloud
- Sources: Twitter, facebook, flickr
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC

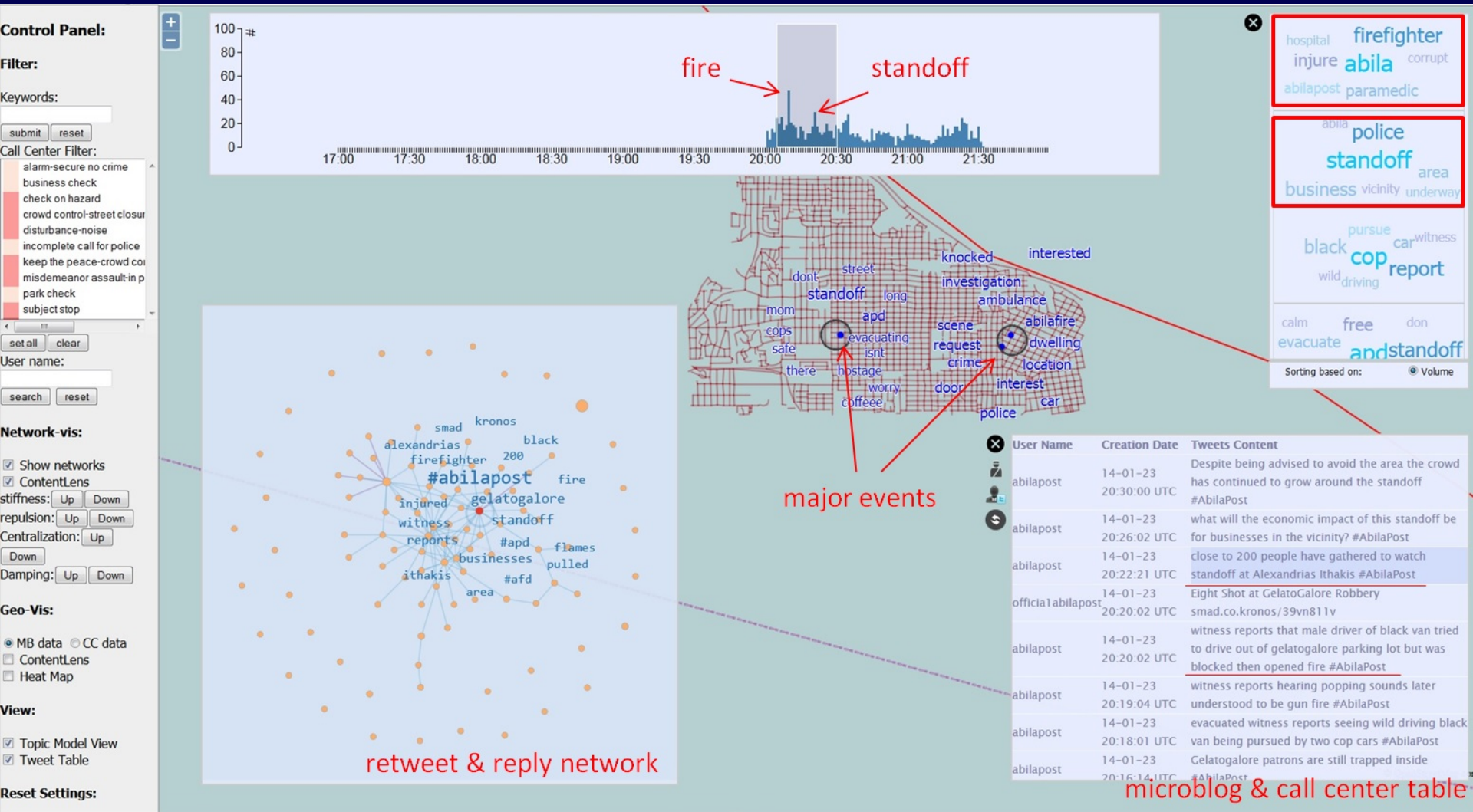


Social Media Analysis and Reporting Toolkit (SMART)



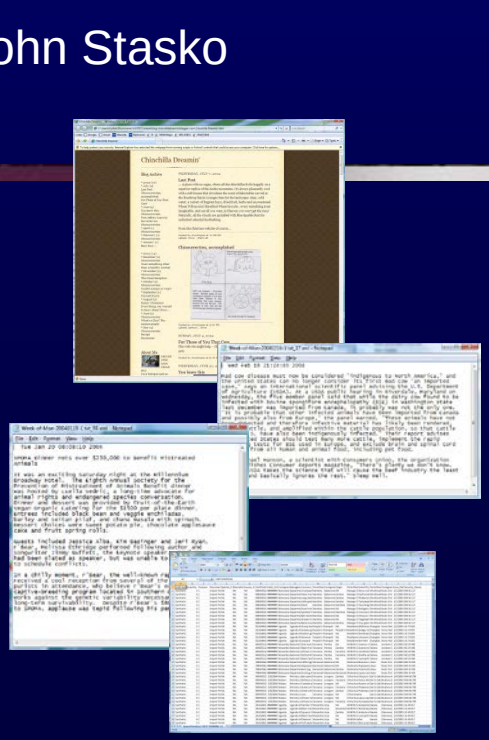
Social Media Analysis and Reporting Toolkit (SMART)

SMART: Network Linkage



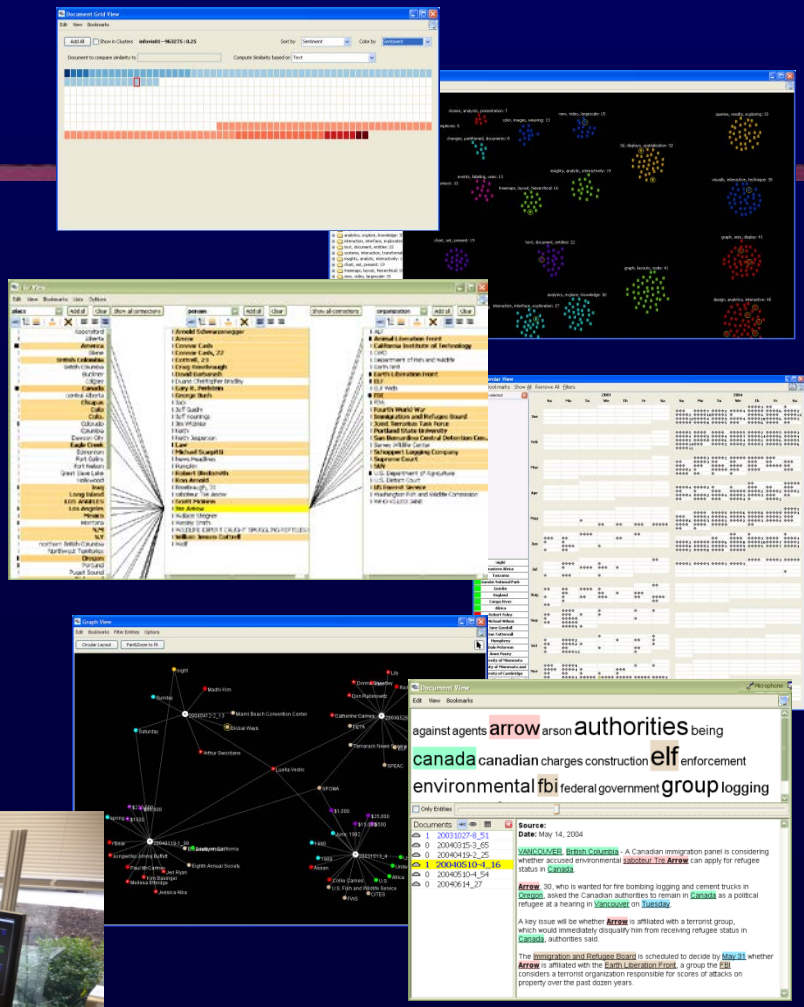
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



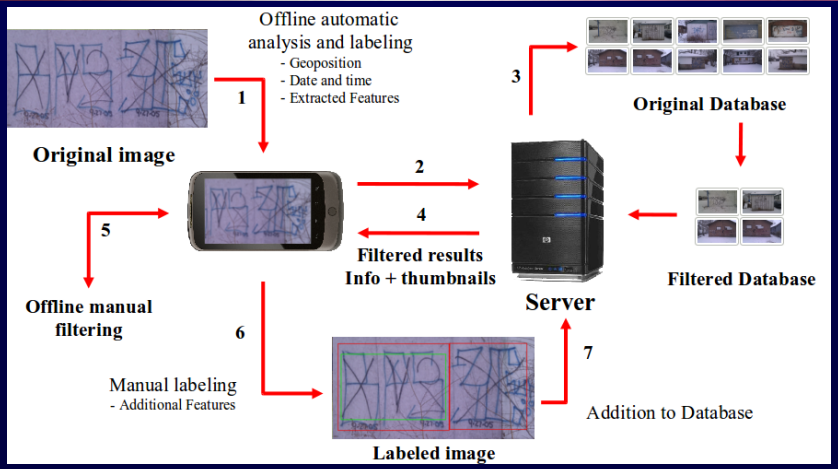
Documents in

Knowledge out





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

For Further Information

www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Data-Driven Decision Making

David Ebert

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

July 2014

VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

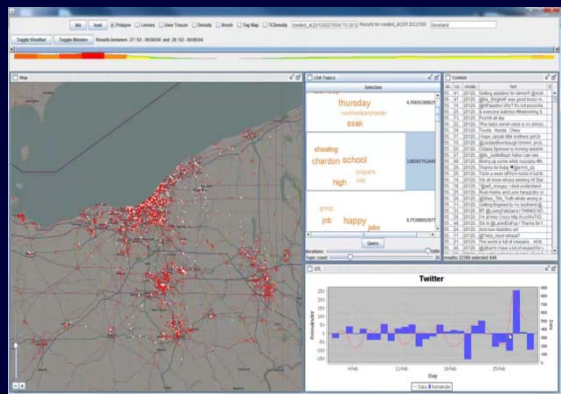
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information
    Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

Visual Analytics Uses for Public Safety

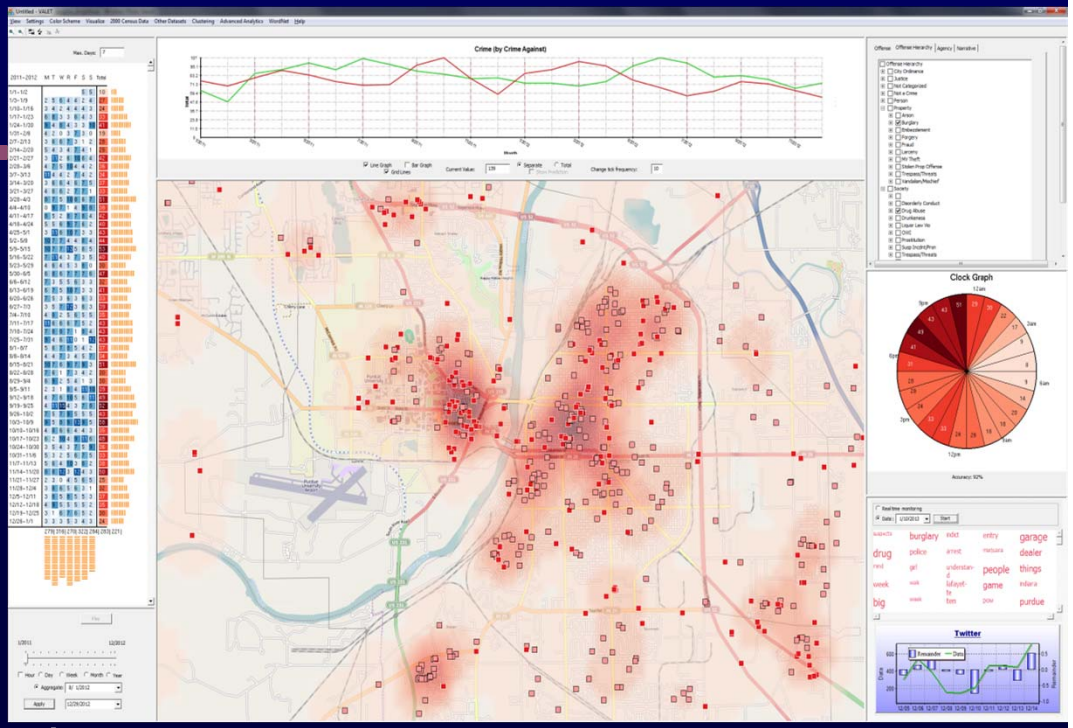


- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



VACCINE

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

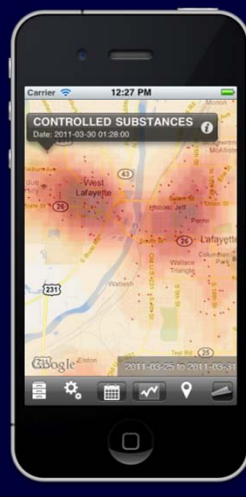
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police

iVALET delivered:

- October 2011: Purdue, WL Police



VACCINE

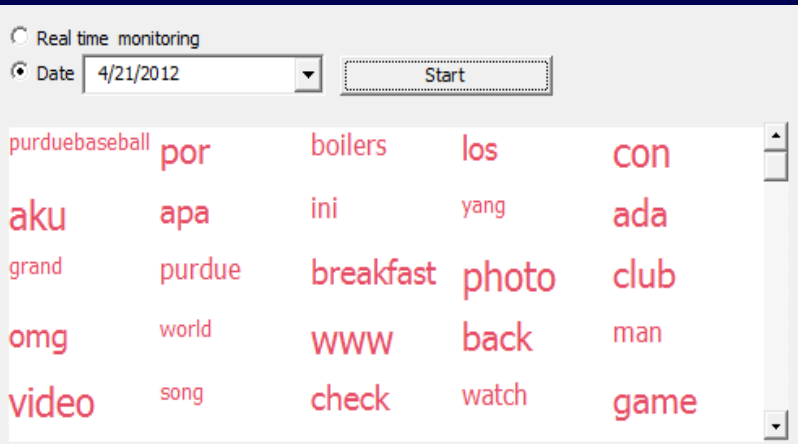
Effective Risk-based Decision Making and Resource Allocation Visual Analytics

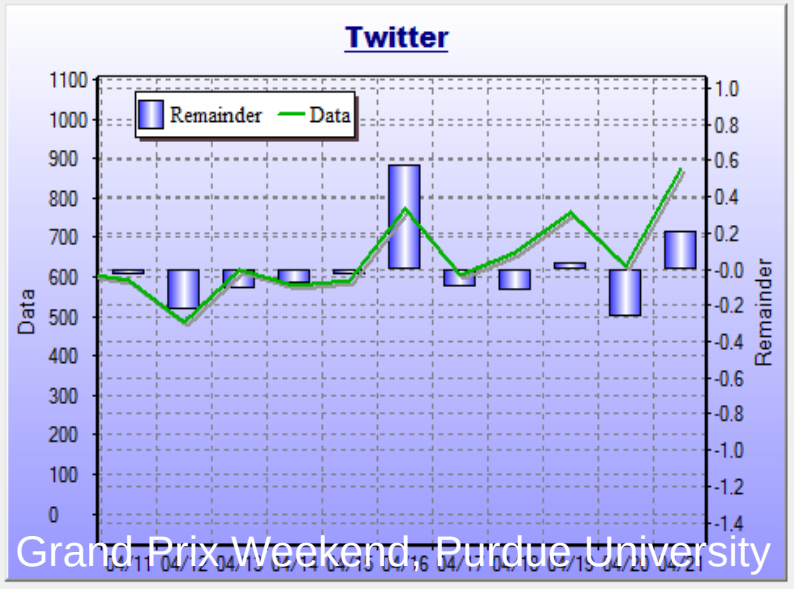
- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

Social Media: Real-time Twitter Visual Analytics

(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud

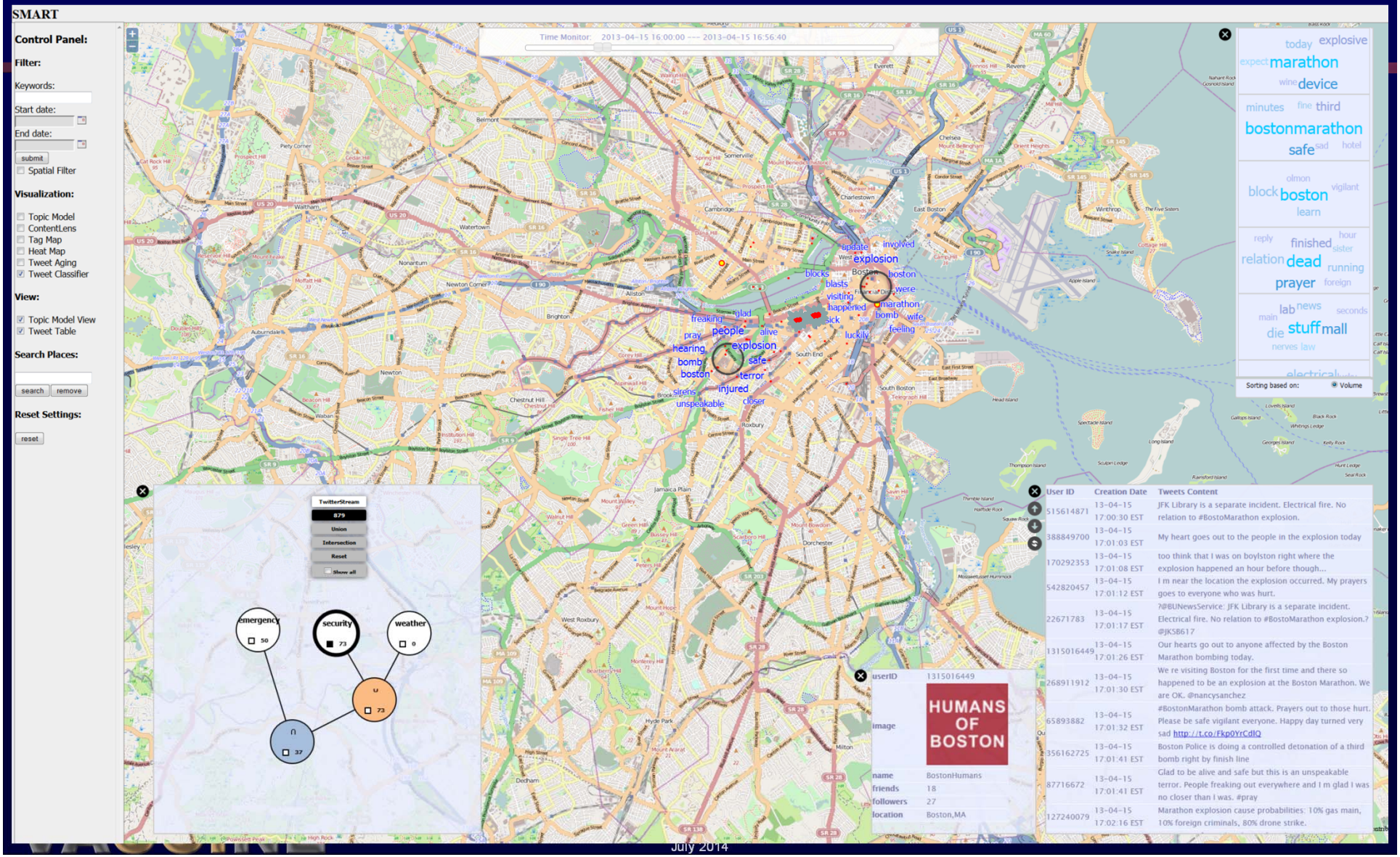




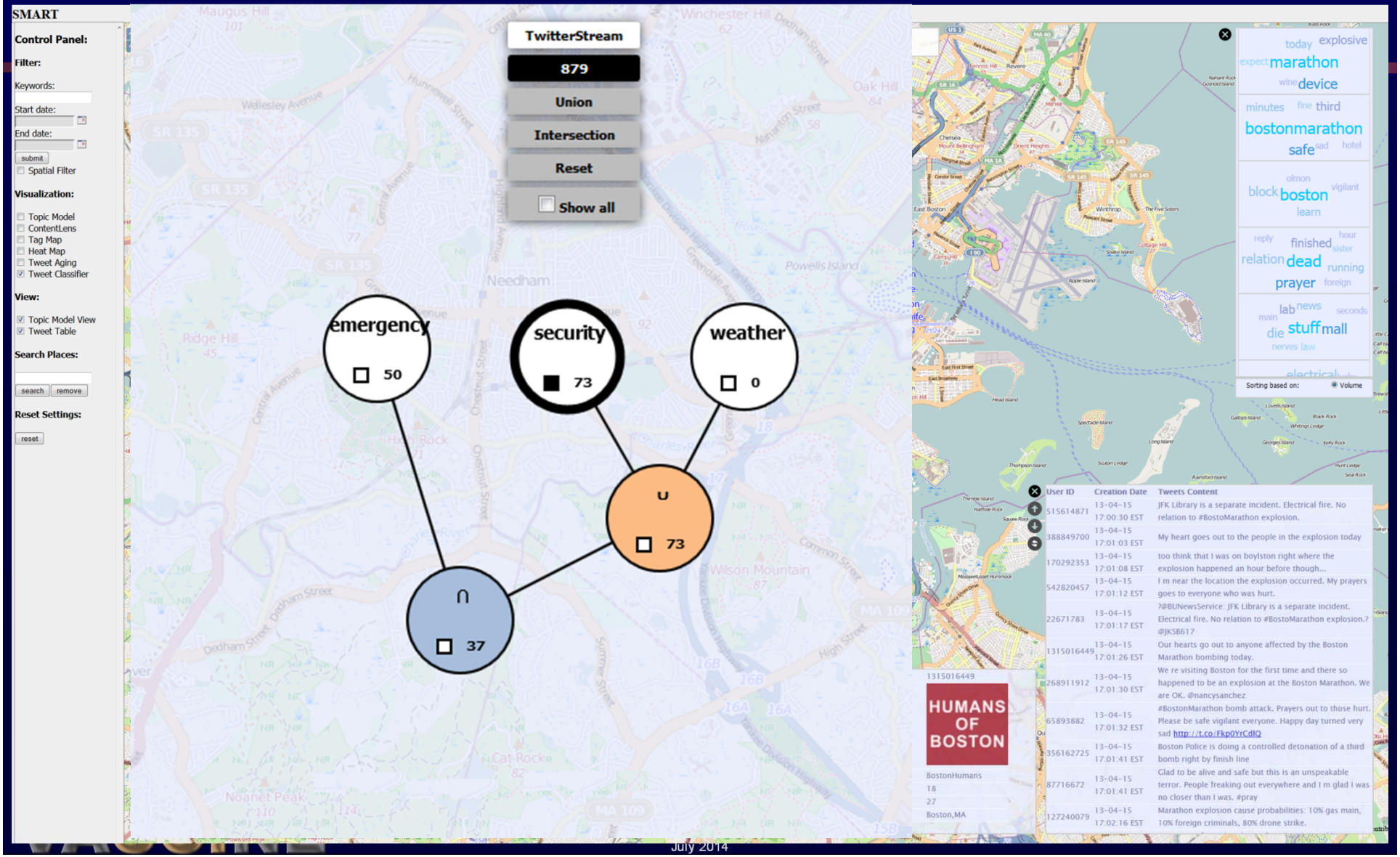
Grand Prix Weekend, Purdue University

VACCINE

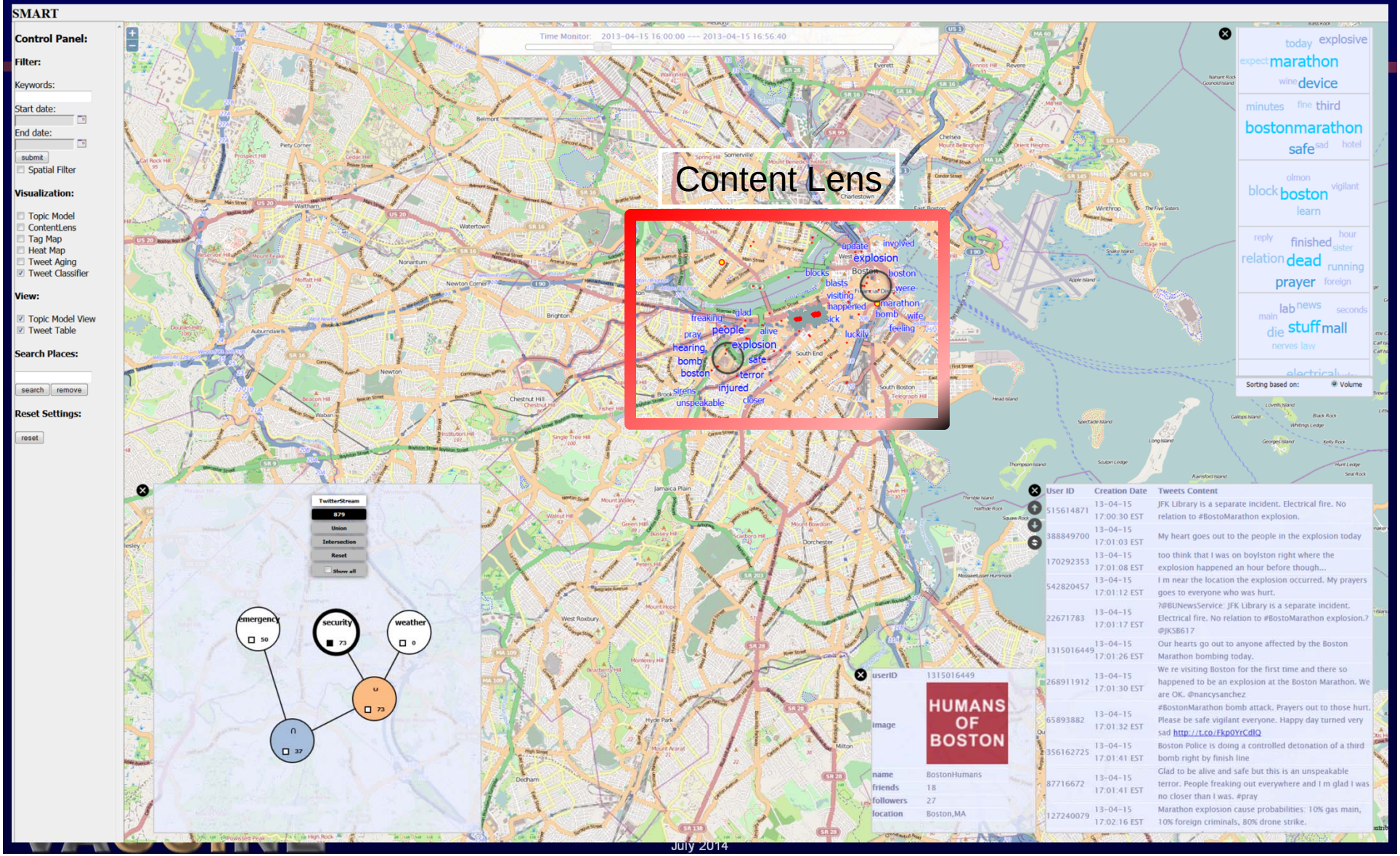
Social Media Analysis and Reporting Toolkit (SMART)



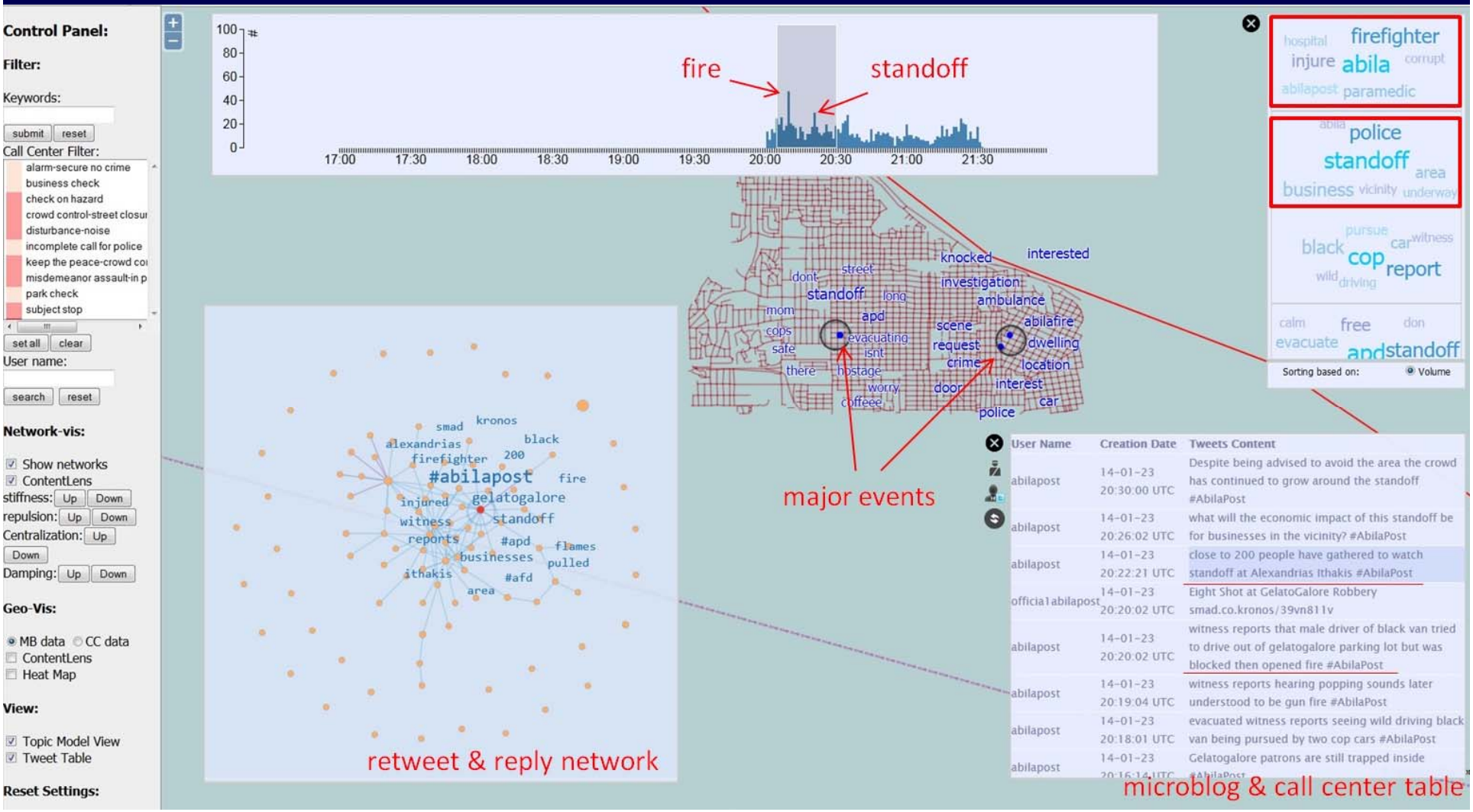
Social Media Analysis and Reporting Toolkit (SMART)



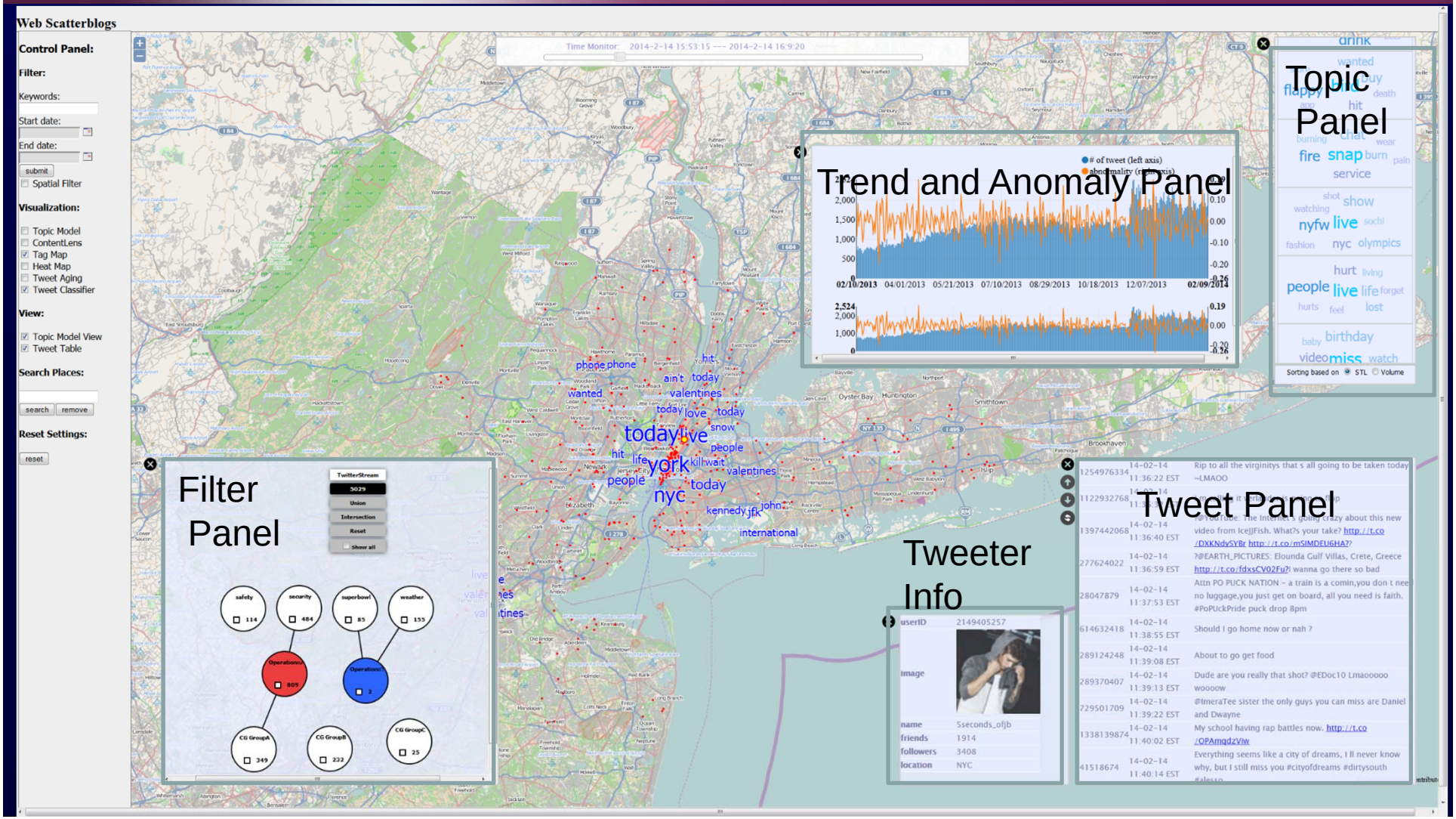
Social Media Analysis and Reporting Toolkit (SMART)



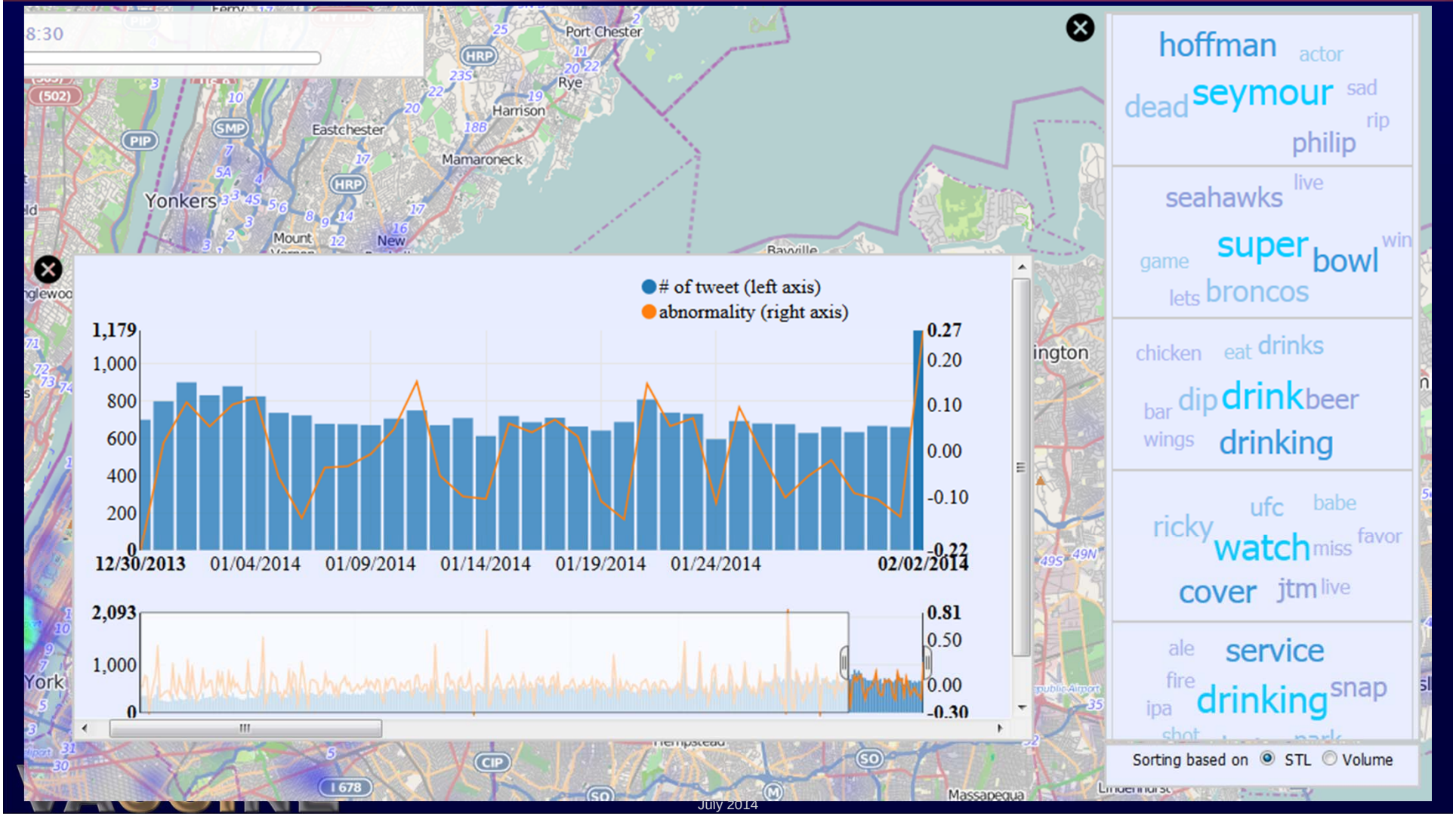
SMART: Network Linkage



SMART Anomaly Detection

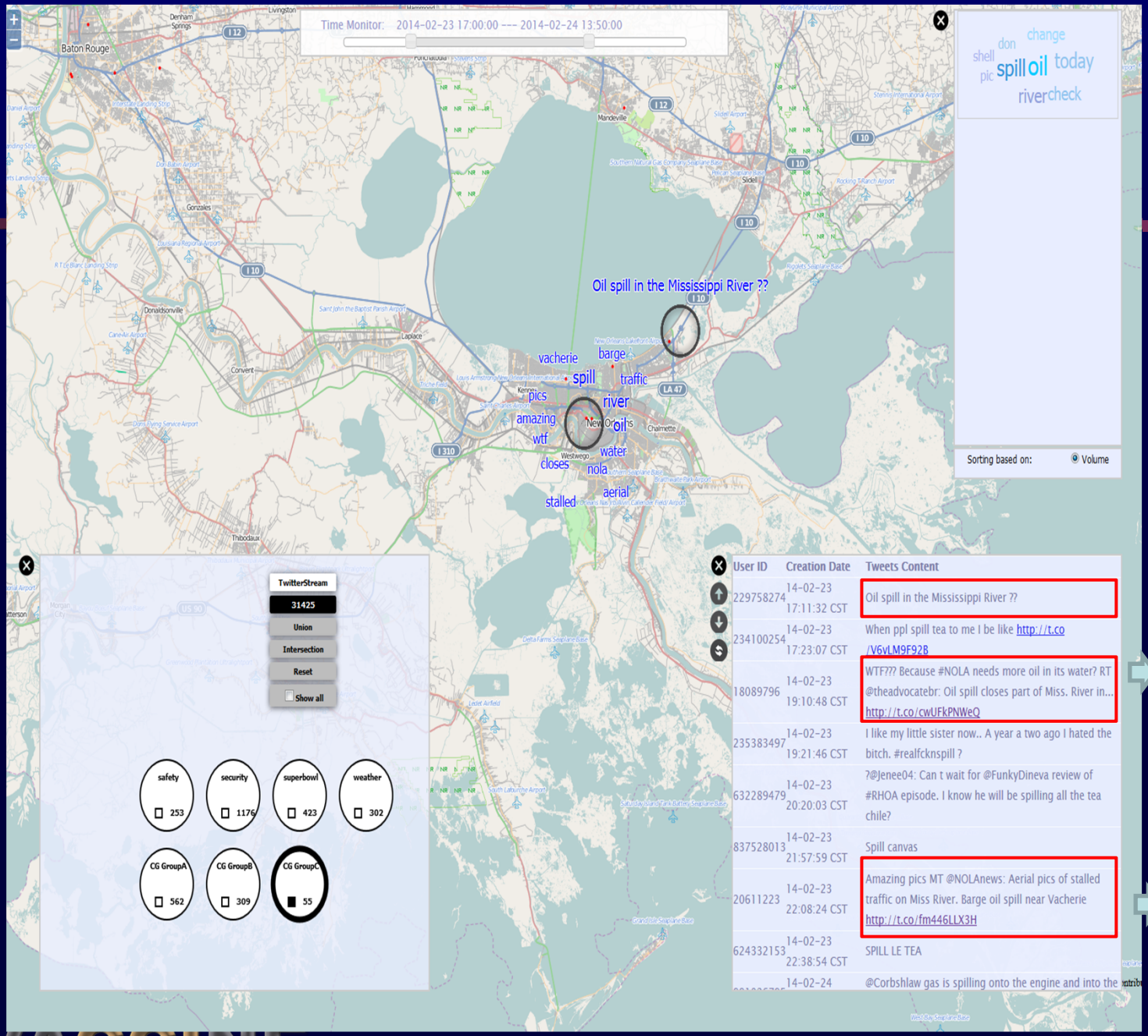


Anomaly Detection

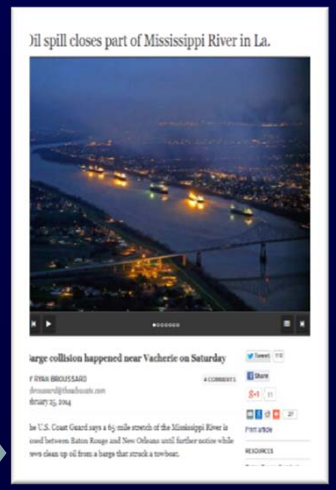


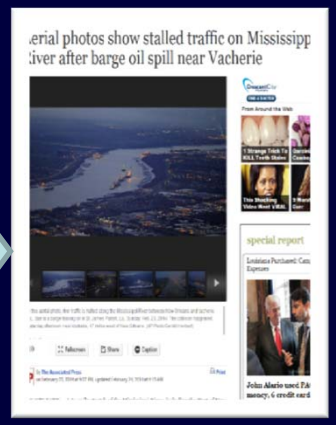
Oil Spill in Mississippi River

2014/02/23



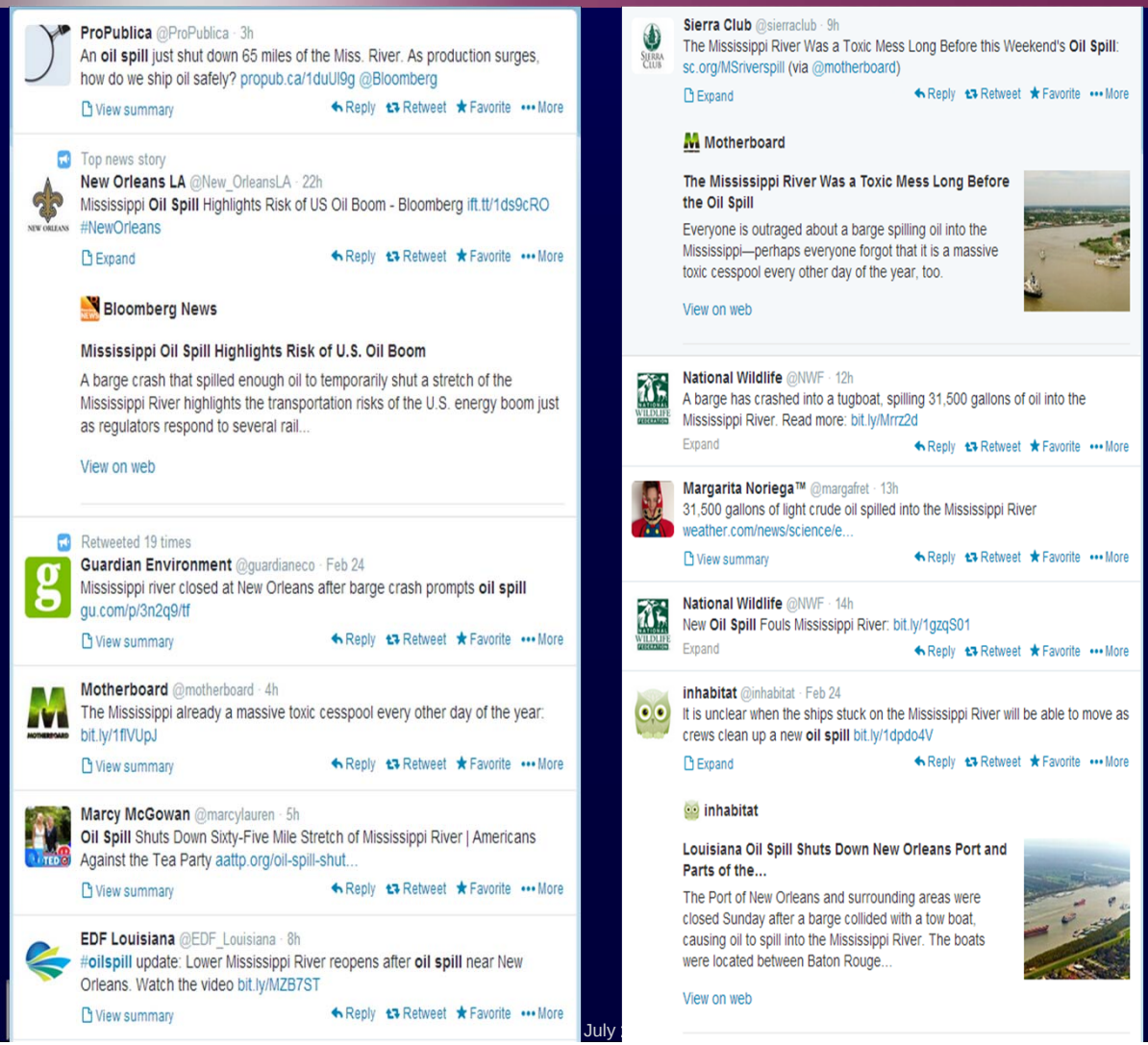
User ID	Creation Date	Tweets Content
229758274	14-02-23 17:11:32 CST	Oil spill in the Mississippi River ??
234100254	14-02-23 17:23:07 CST	When ppl spill tea to me I be like http://t.co/V6vLM9F92B
18089796	14-02-23 19:10:48 CST	WTF?? Because #NOLA needs more oil in its water? RT @theadvocatebr: Oil spill closes part of Miss. River in... http://t.co/cwUFkPNWeQ
235383497	14-02-23 19:21:46 CST	I like my little sister now.. A year a two ago I hated the bitch. #realfcknspill ?
632289479	14-02-23 20:20:03 CST	?@Jenee04: Can t wait for @FunkyDineva review of #RHOA episode. I know he will be spilling all the tea chile?
837528013	14-02-23 21:57:59 CST	Spill canvas
20611223	14-02-23 22:08:24 CST	Amazing pics MT @NOLAnews: Aerial pics of stalled traffic on Miss River. Barge oil spill near Vacherie http://t.co/fm446LLX3H
624332153	14-02-23 22:38:54 CST	SPILL LE TEA
	14-02-24	@Corbshlaw gas is spilling onto the engine and into the





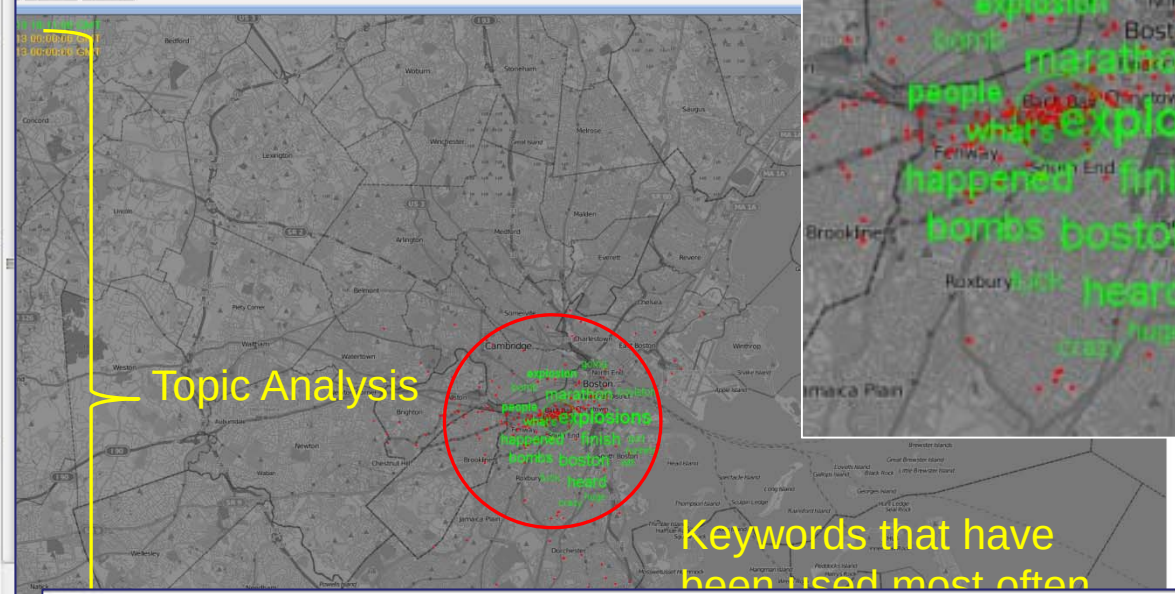
VACCINE

Non-geo Tweets On Twitter About Oil Spill Event



Explosion Area in Boston







Keywords that have been used most often

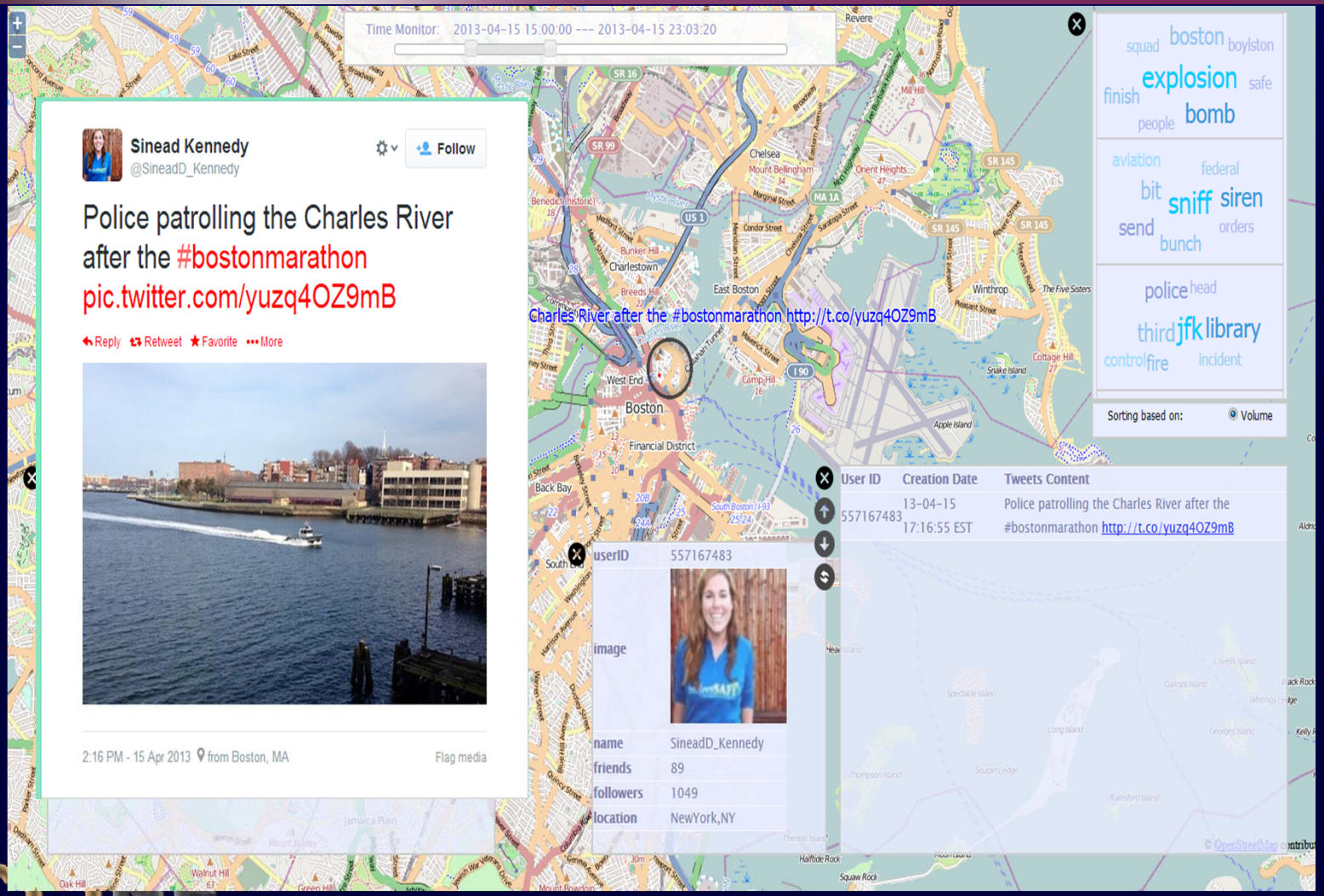
Tweets

Text
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped.
two bombs just went off on boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finsh line this afternoon.
What the FUCK was that
Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalcIO I figured 3 people would get the joke
Two explosions just rocked the finish line of the Boston Marathon. Sirens galore. People running in fear. Wonder what happened.
This is crazy i seen that blow up #bostonmarathon
can someone tell me what that explosion was!? #boston #bostonmarathon

First Response (Tweet & Picture)



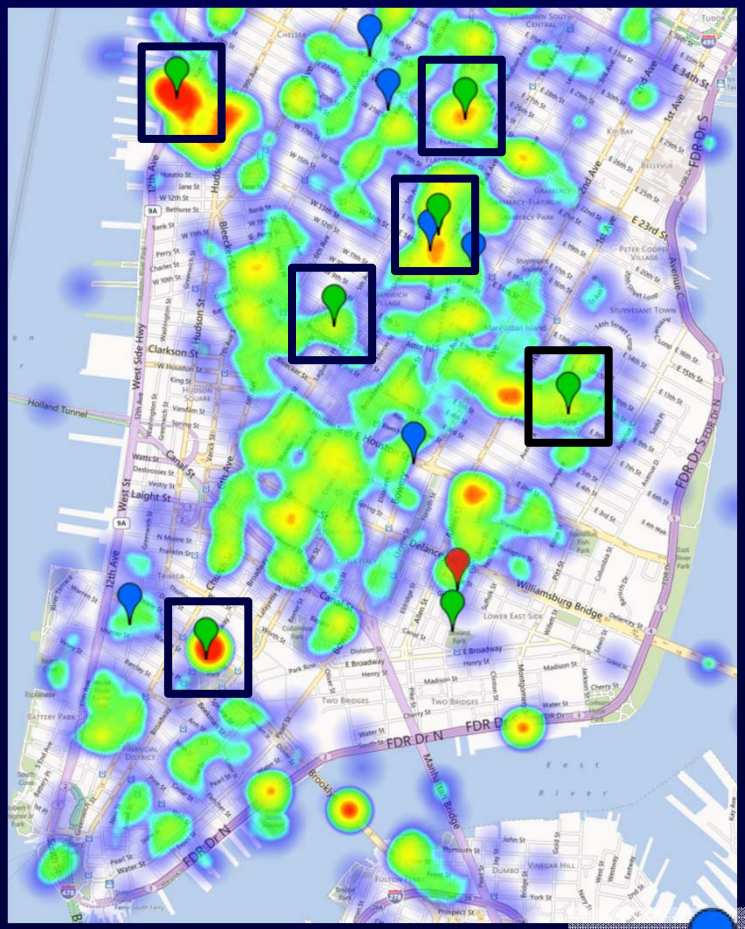
Police Patrolling The River

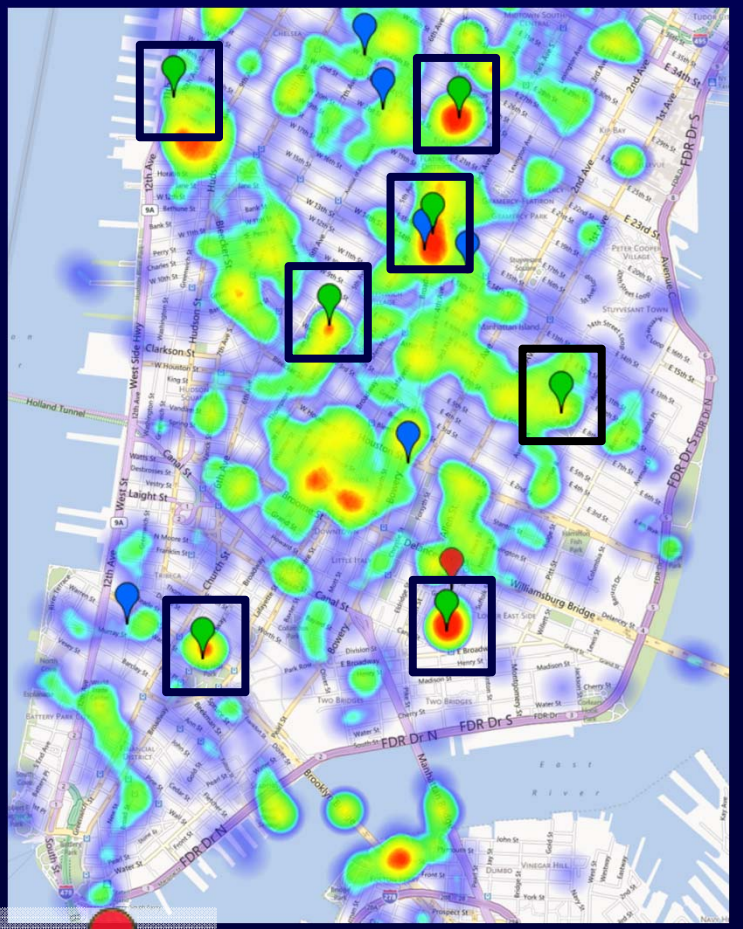


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

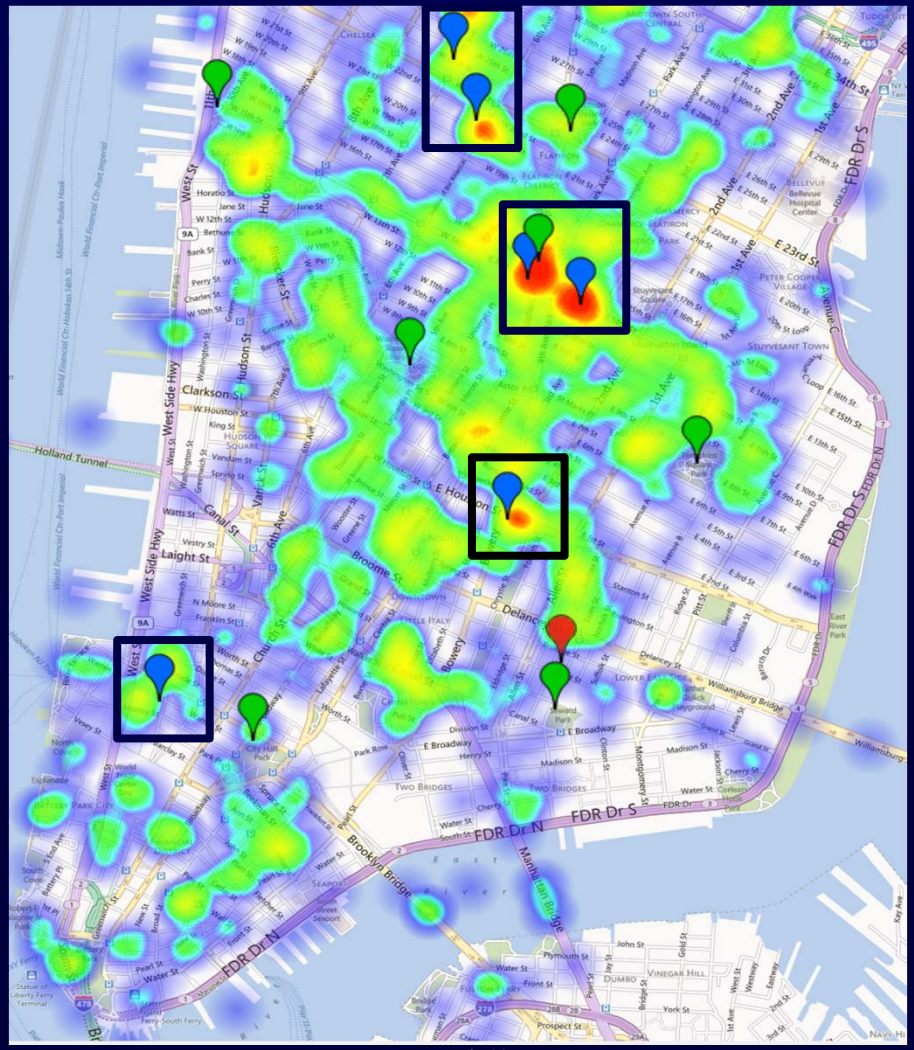
July 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



VACCINE

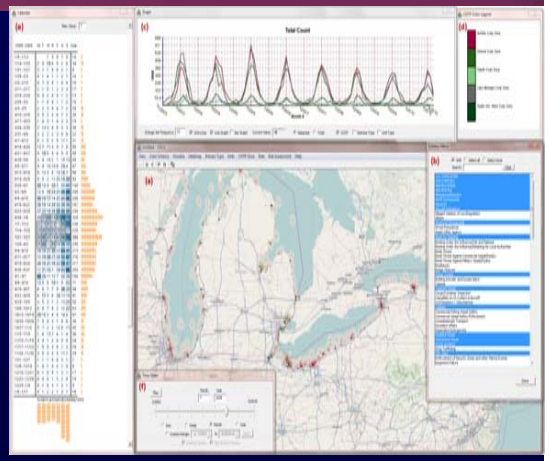
July 2014

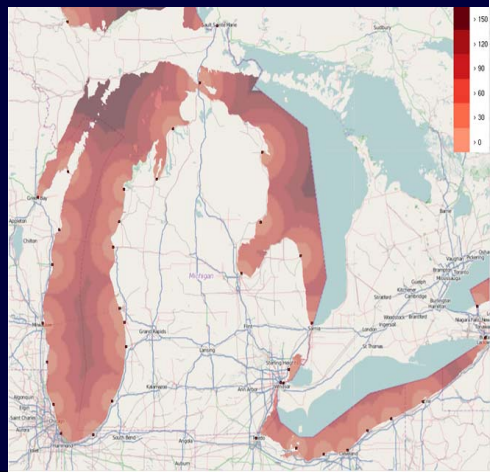
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

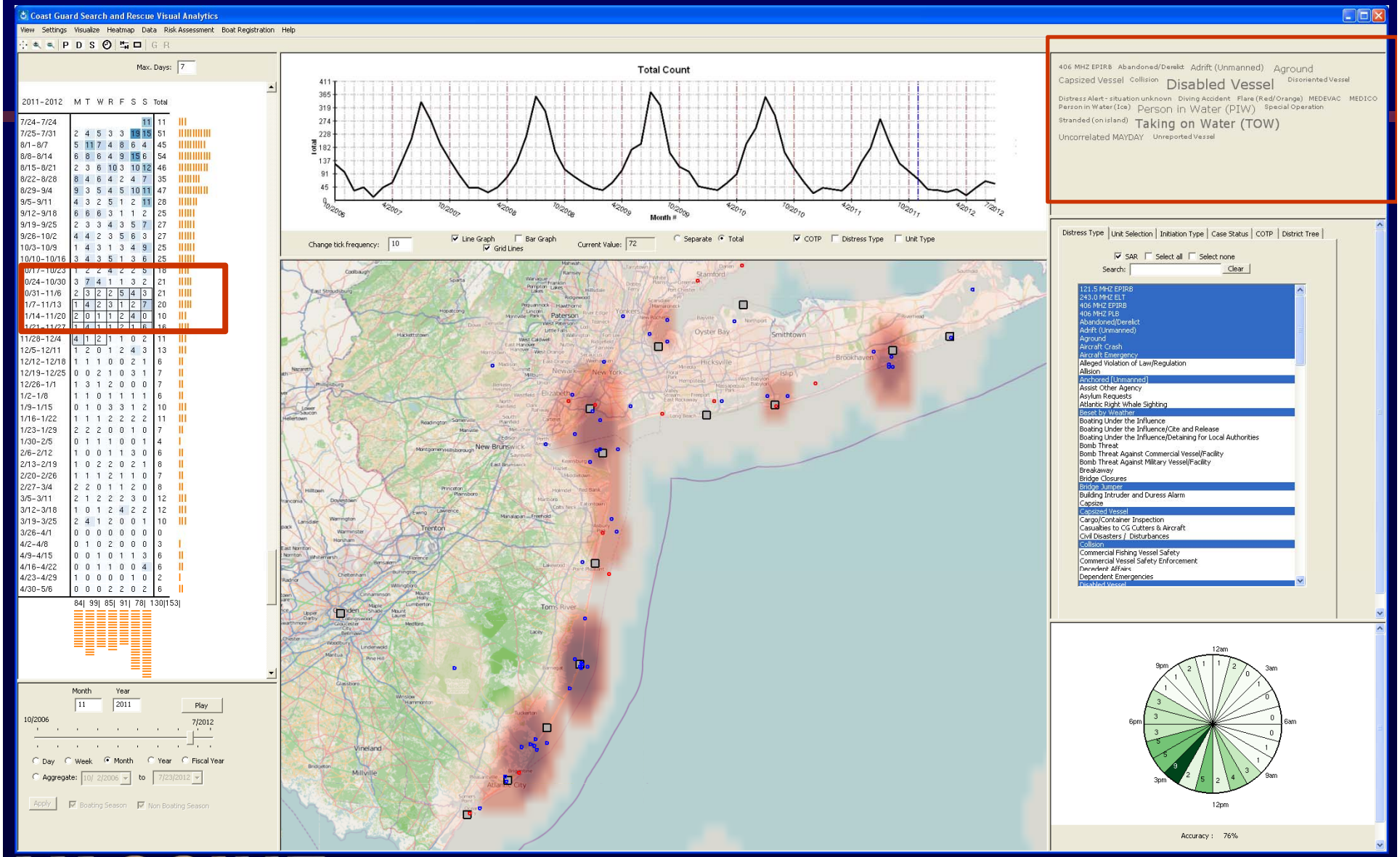




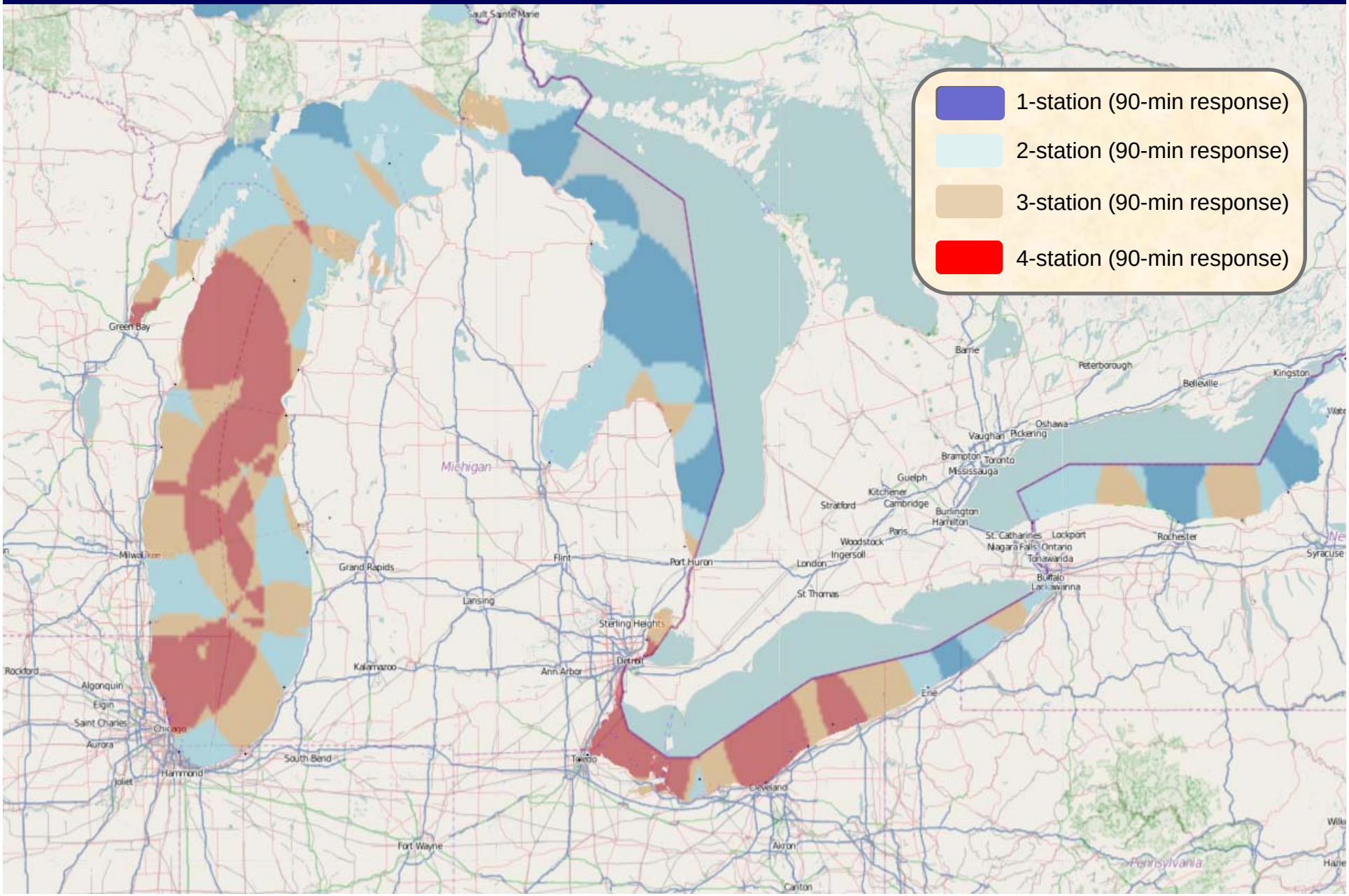
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



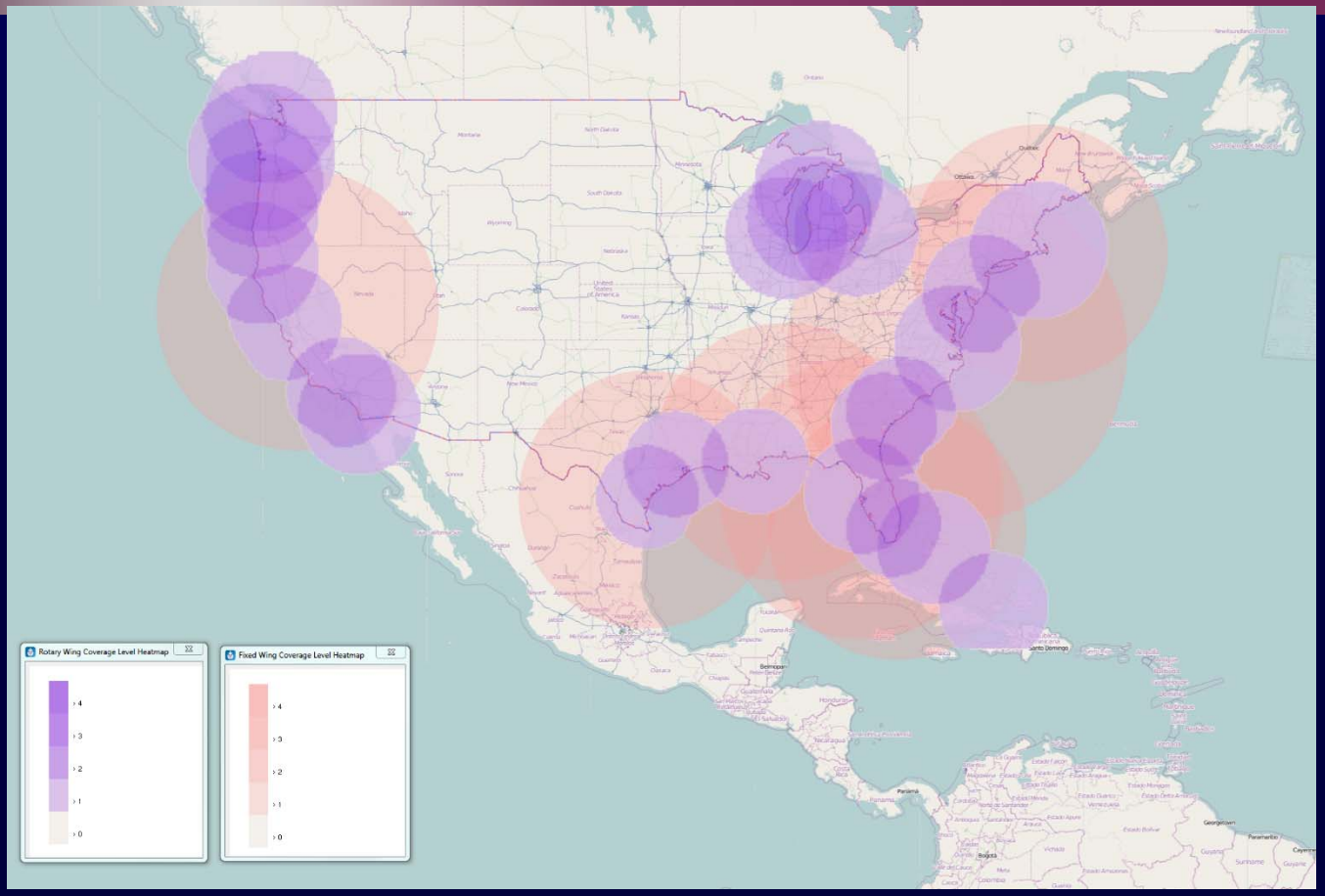
cgSARVA 2.0 New Features (1/2)

- Rotary-Wing Coverage
  - The new shapes depict the maximum range from land that can be achieved if the aircraft returns to the nearest point of land possible
- Offshore and Inland Filtering
- Siting Plans
  - Shows the current inventory of boats at the station
- Reliability Coverage
  - Indicates the probability of the boats a the given station being able to respond to that point on the water

cgSARVA 2.0 New Features (2/2)

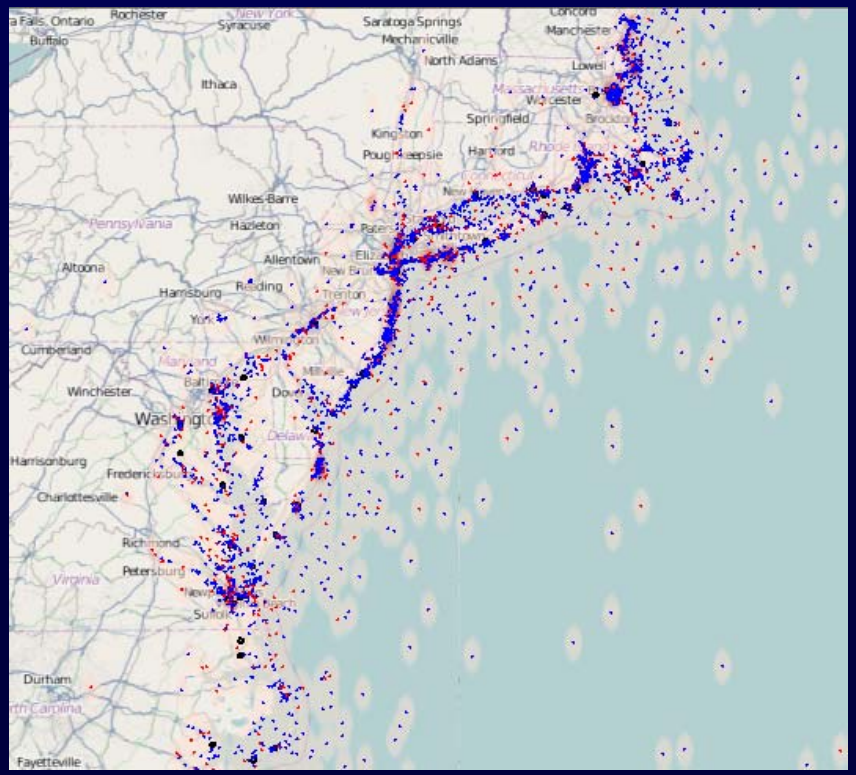
- Rescue 21
  - Visualize the VHF Receiving Range of the R-21 Towers and allows the user to turn on/off each tower
- Patrol Boats
  - Shows the home port for the different cutters, and displays the level of effort that is defined by the percentage of presence the group of patrol boats provide to a Sector relative to a number of hours parameter

cgSARVA Results – Aircraft coverage

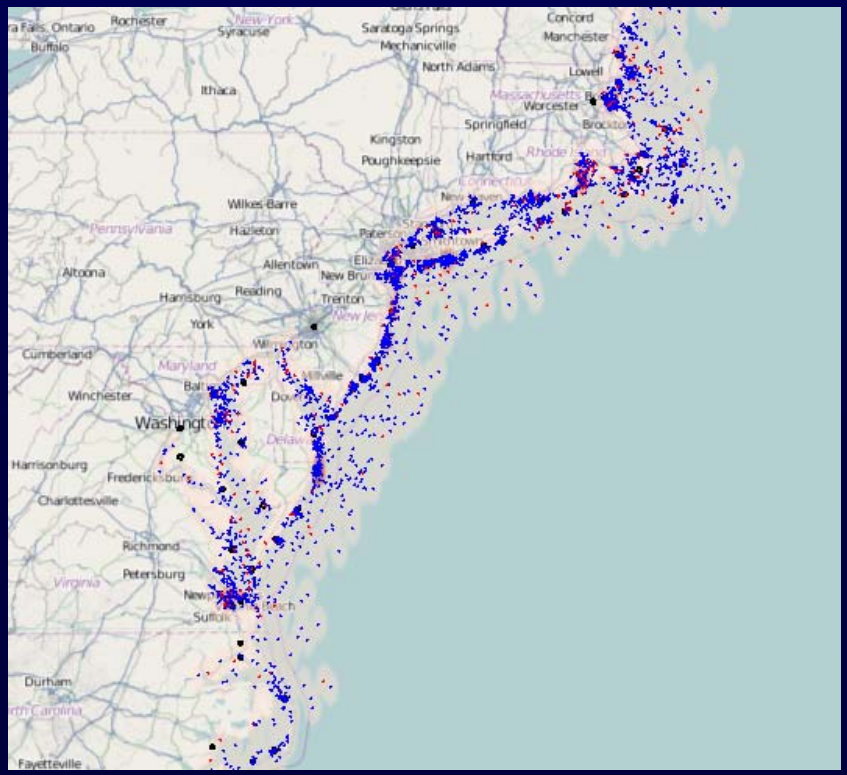


Rotary wing coverage shown in purple. Fixed wing coverage shown in orange.

cgSARVA Results – Offshore and Inland Filtering



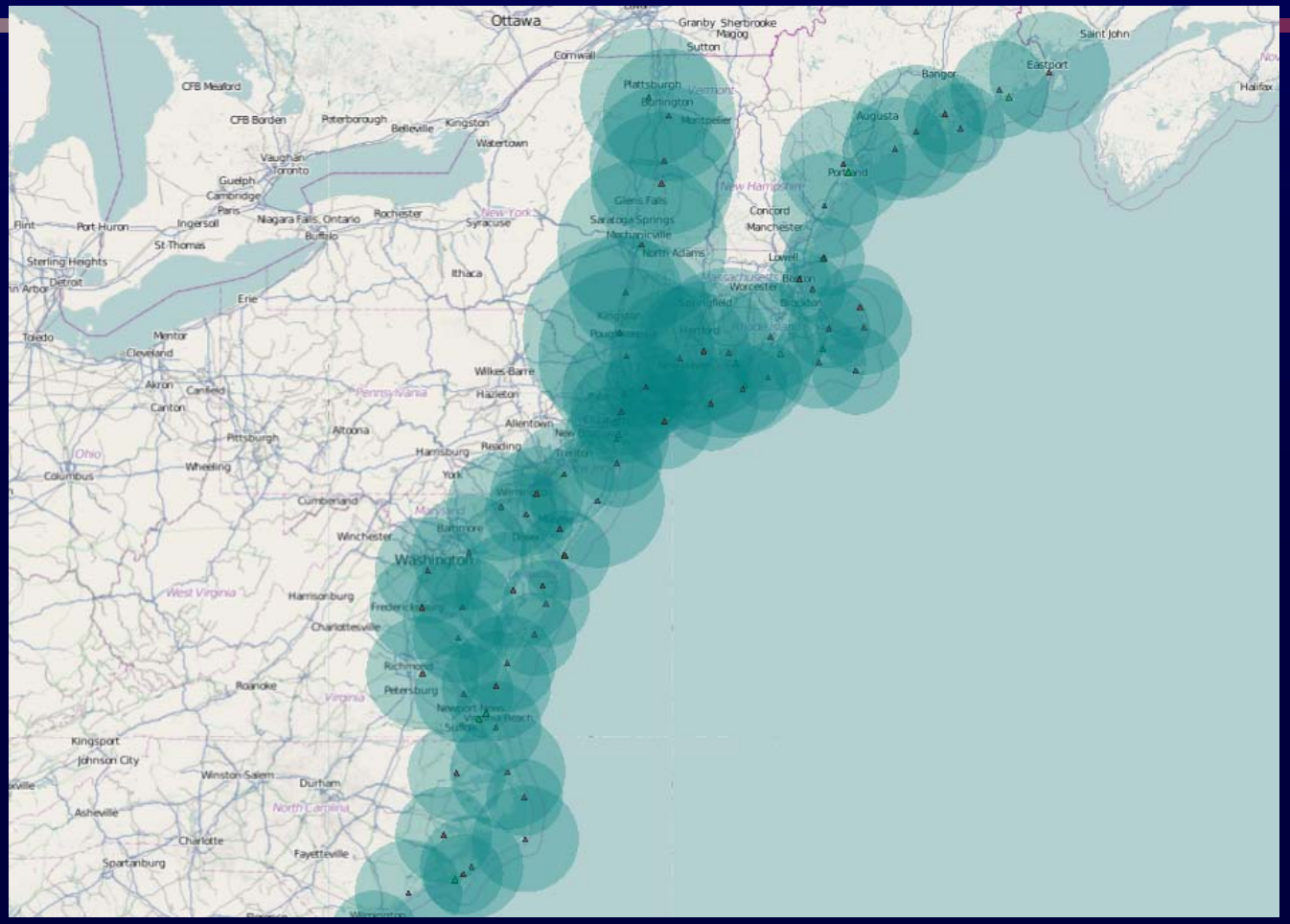
Before filtering. Year 2010



Filtered cases using 5 NM inland and 50 NM offshore. Year 2010

VACCINE

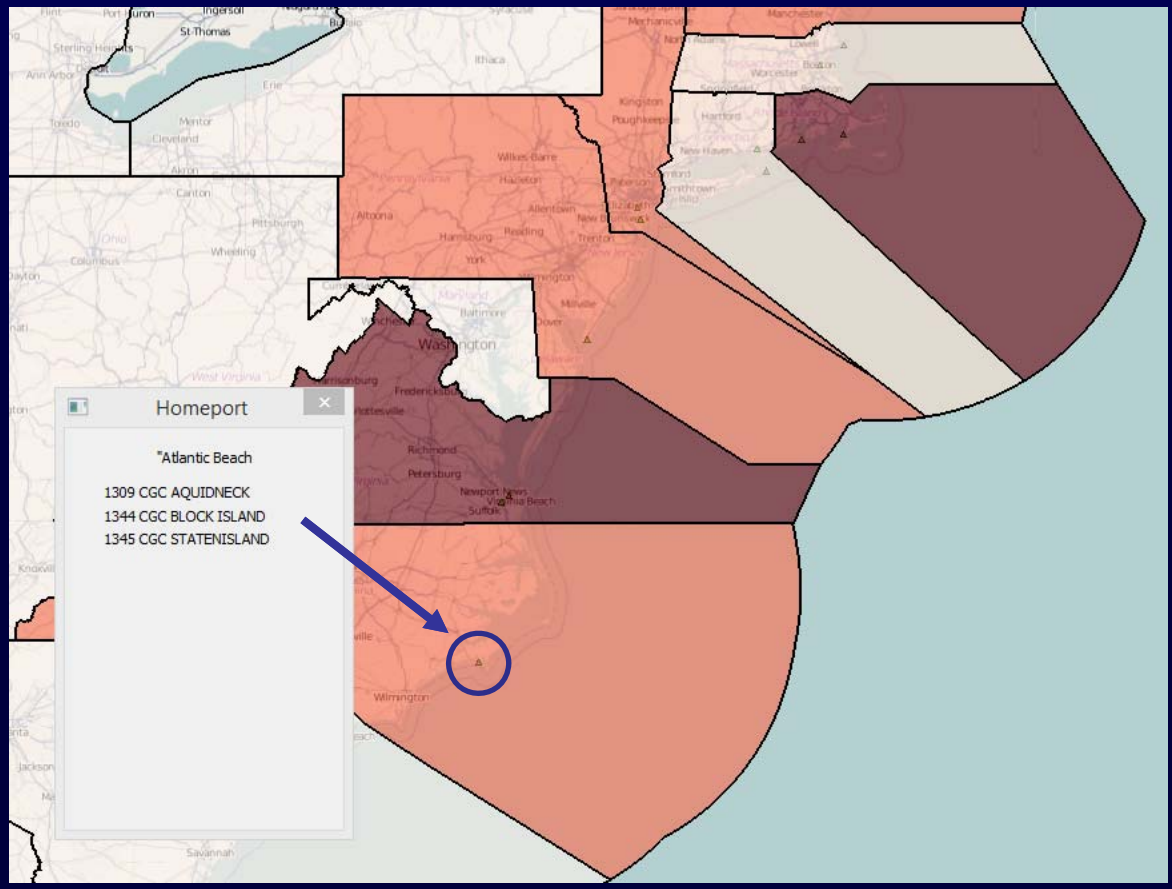
cgSARVA Results – Rescue 21



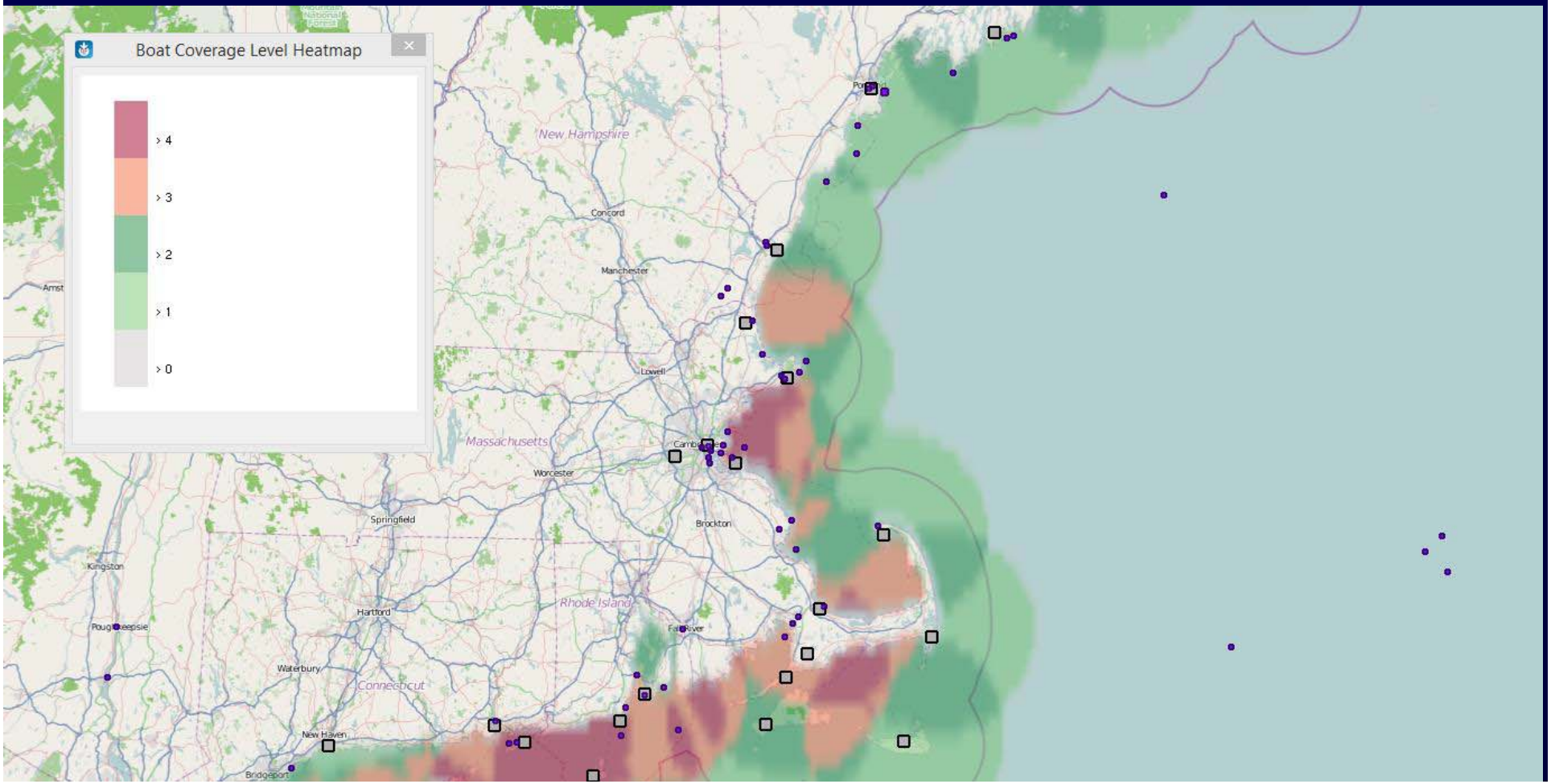
cgSARVA Results – Patrol Boat

Showing homeport of Atlantic Beach and its 3 cutters.

The choropleth map represents the aggregate number of hours for all the boats in the Sector.



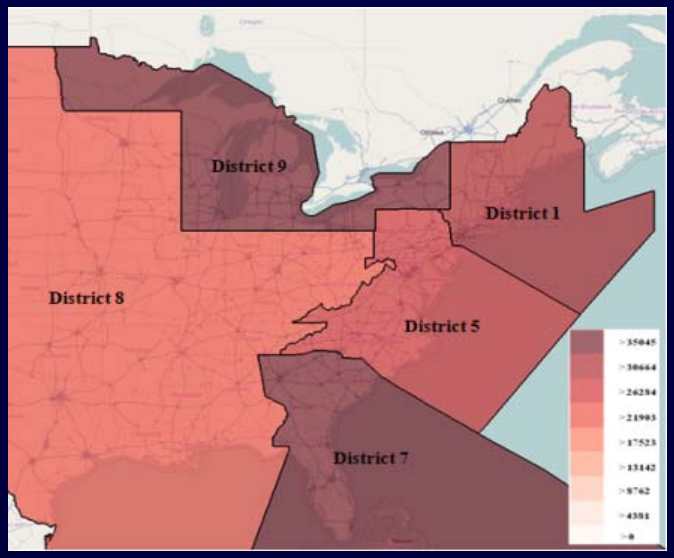
Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



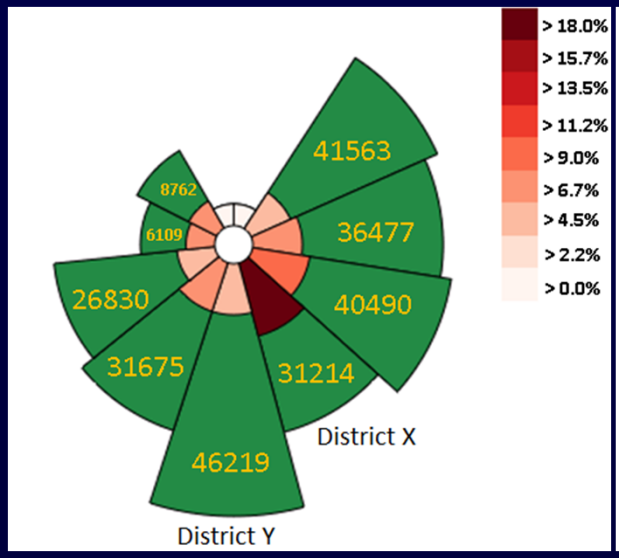
ORAM Visualization

Component Overview: Identification of Potential Risks

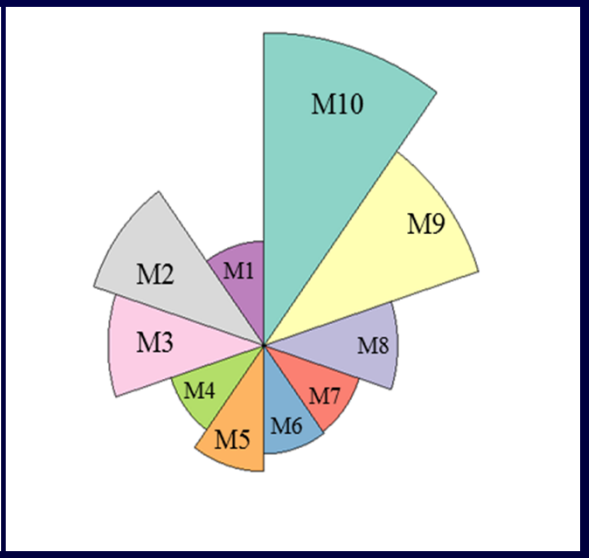
General understanding on how the risk is distributed between the different areas and missions.



(A) District Choropleth Map for Total Risk across all missions



(B) Risk profiles across all districts and mission areas



(C) District X Risk Profiles across all mission area

Component Overview: Analysis and Assessment

Specific areas can be analyzed by time and details on demand.



(D) Total M10 Risk Values distributed among District X's station AOR



(E) KDE heatmap showing Total M10 Risk values

Interactive Operational Performance Assessment Report (iOPAR)

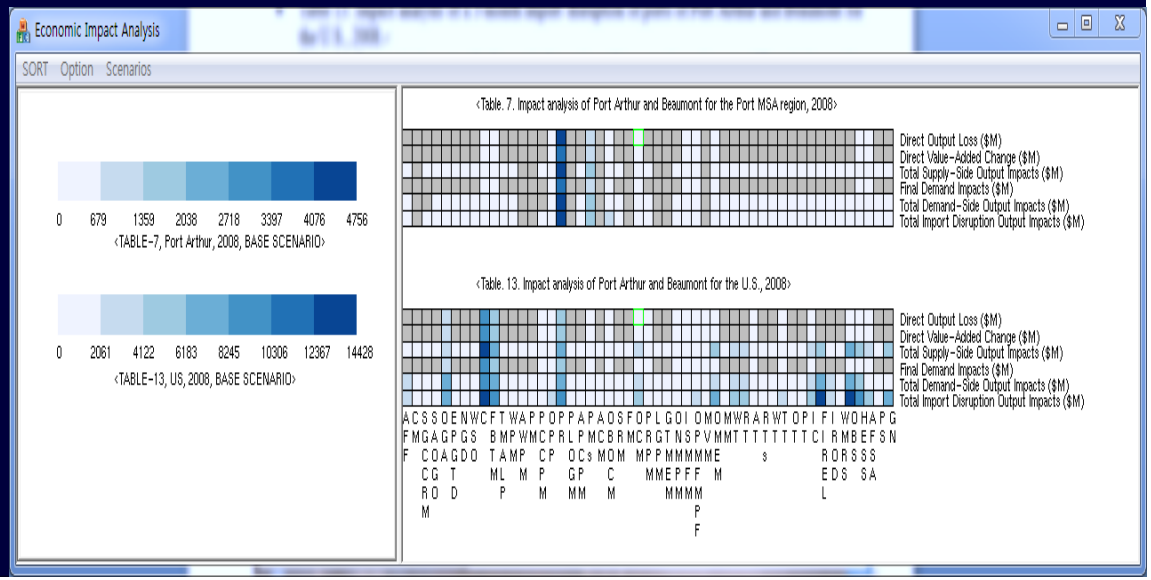
- OPAR
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Anaylze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels

USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

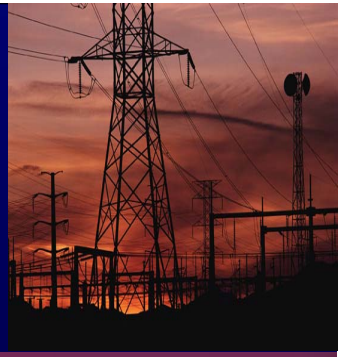
IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



VACCINE

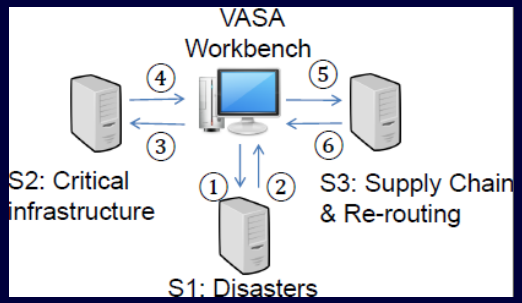
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s):
  Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s): Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

Purdue

UNCC



U Minn

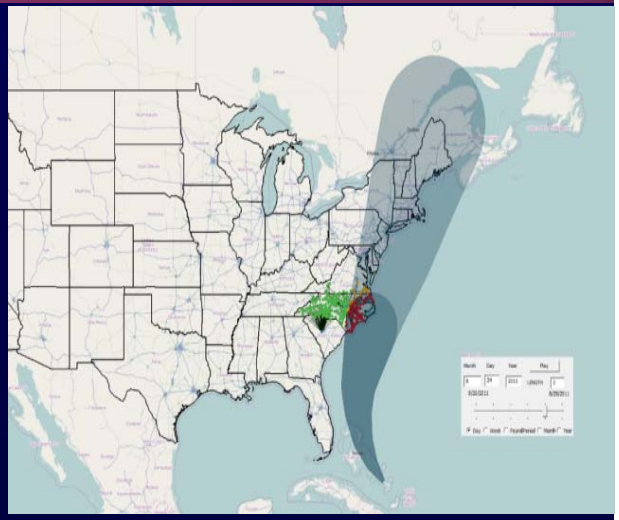
TACC

VACCINE

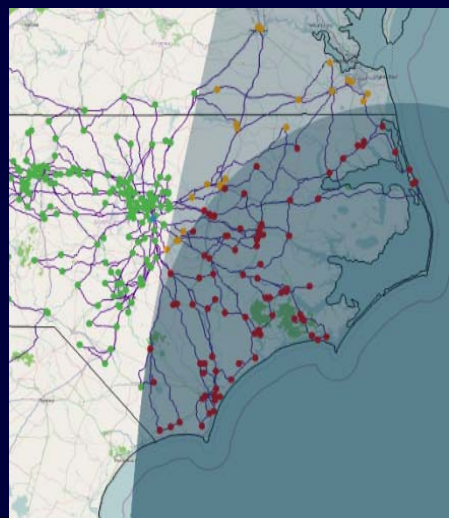
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:



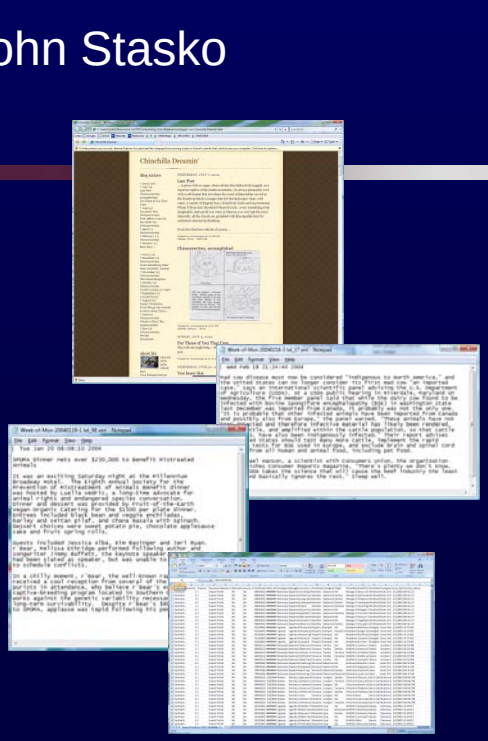
- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  combine interactive VA with complex simulation models



VACCINE

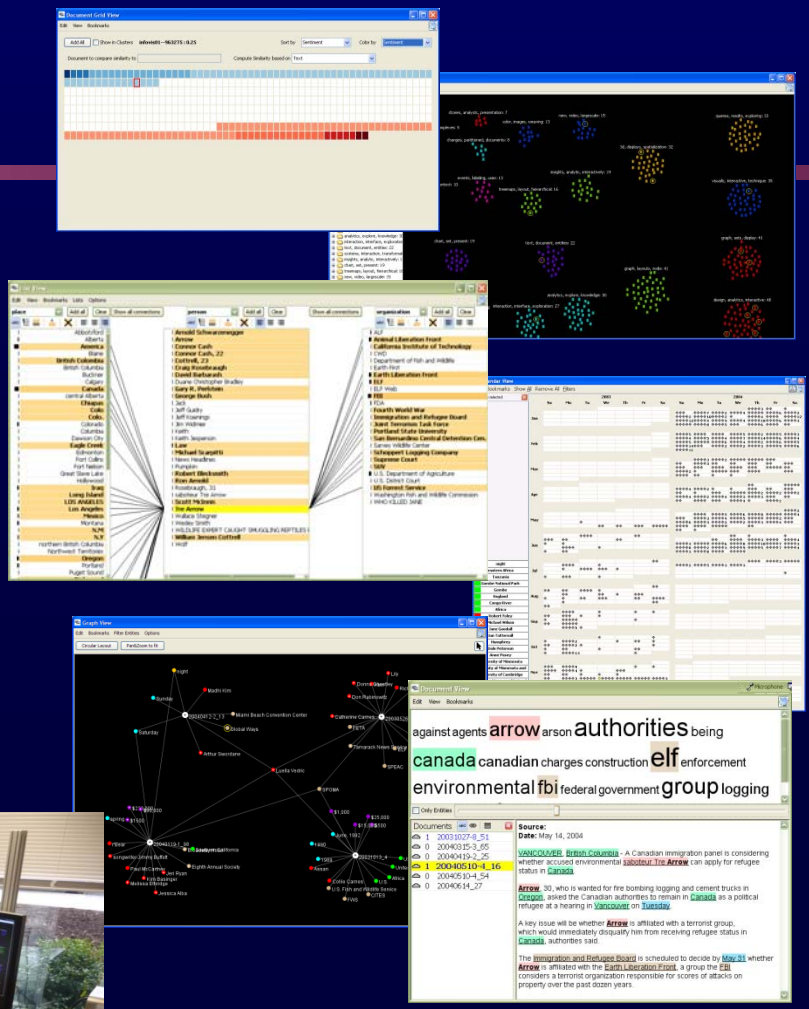
Jigsaw: Visual Analytics for Investigative Analysis

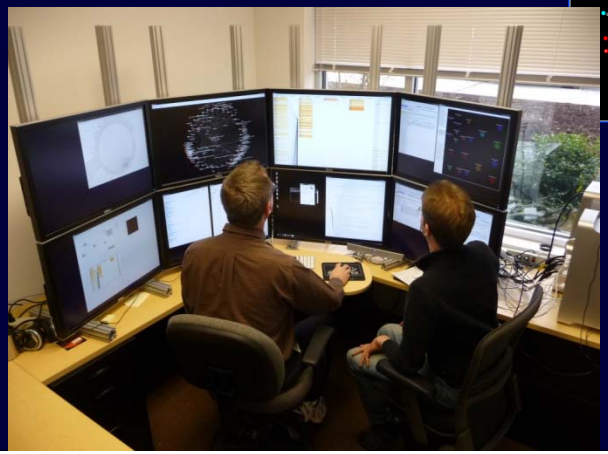
John Stasko



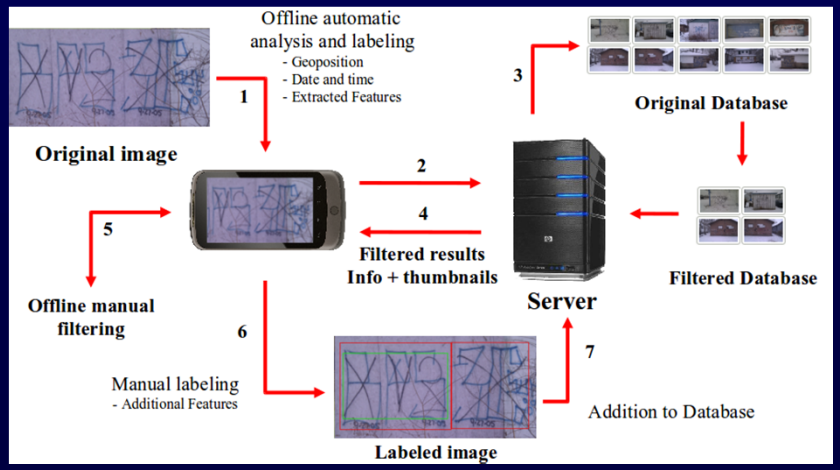
Documents in

Knowledge out





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

For Further Information

www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Data-Driven Decision Making

David Ebert

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

October 2014

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- **Purdue University**
- **Georgia Institute of Technology**
- **Pennsylvania State University**
- **Stanford University**
- **University of North Carolina at Charlotte**
- **University of Washington**
- **Arizona State University**
- **Simon Fraser University**
- **University of British Columbia**
- **Justice Institute of British Columbia**
- **Ontario Institute of Technology**
- **Dalhousie University**
- **University of Houston, Downtown**
- **Virginia Tech**
- **Indiana University**
- **Florida International University**
- **University of Texas at Austin**
- **Morgan State University**
- **Navajo Technical College**
- **University of Stuttgart**
- **University of Swansea**
- **Oxford University**
- **University of Calgary**
- **University of Manitoba**
- **Carleton University**
- **University of Victoria**

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (GA, IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America



VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**
  - Interdisciplinary world leading team of researchers and students

"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address

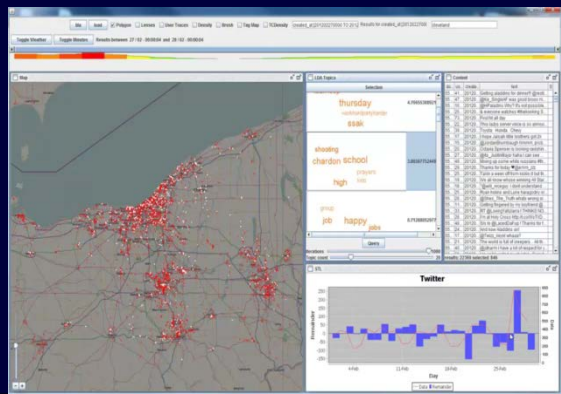


VADM Robert Parker with VACCINE student researchers (cgSARVA, COAST, iOPAR)

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

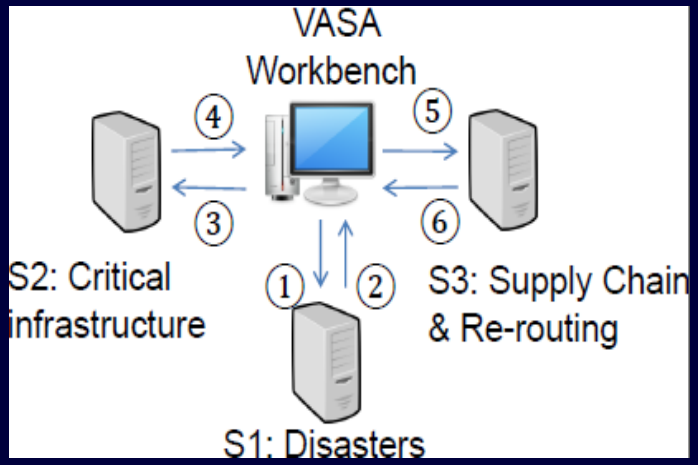
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

TACC

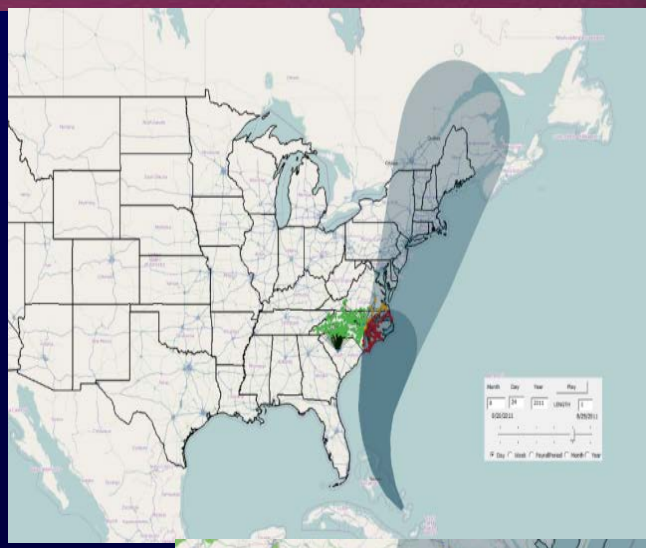
VACCINE

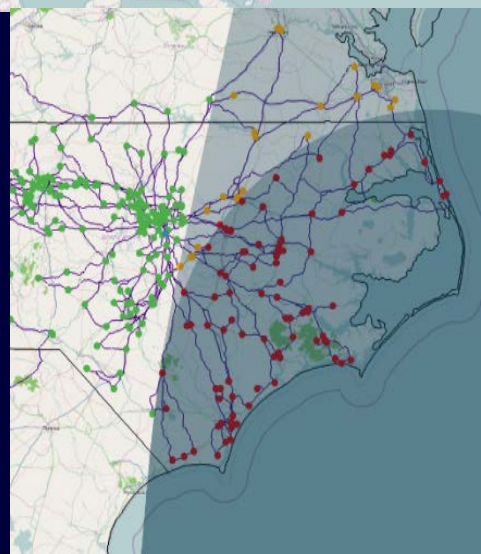
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

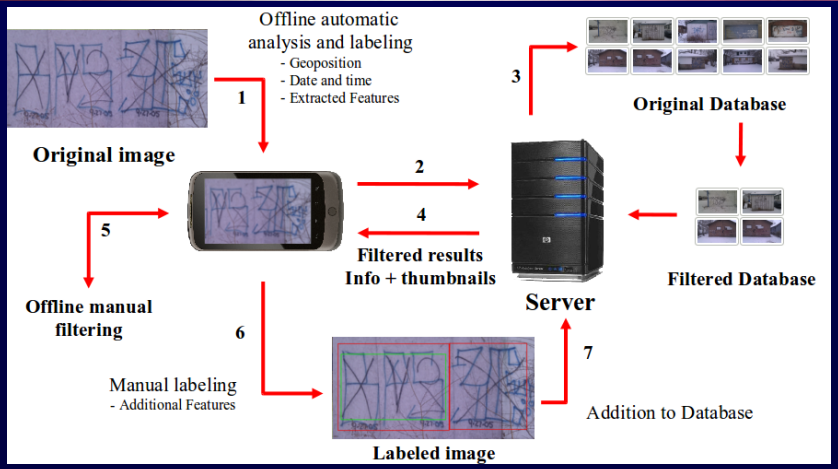
Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models





Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 2000 graffiti and tatoo images acquired

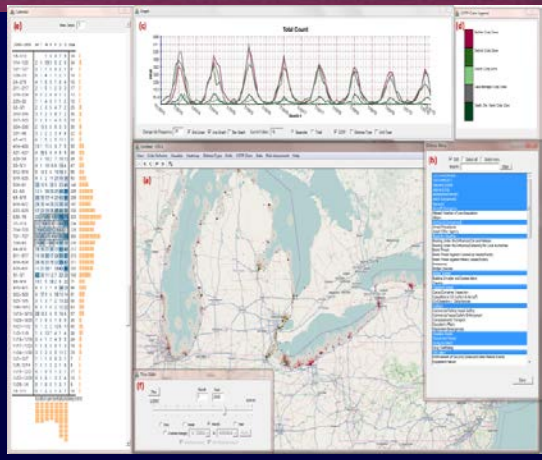
GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added
- Spring 2014
  - Public reporting version

VACCINE

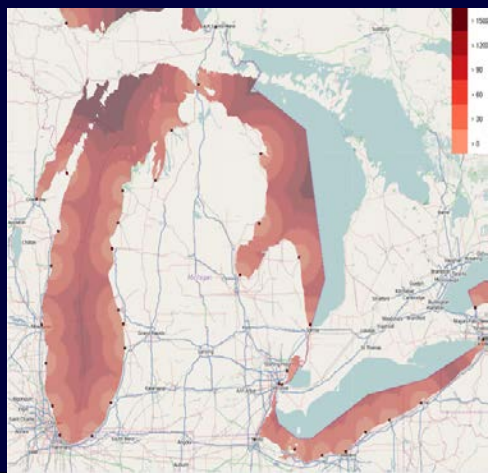
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

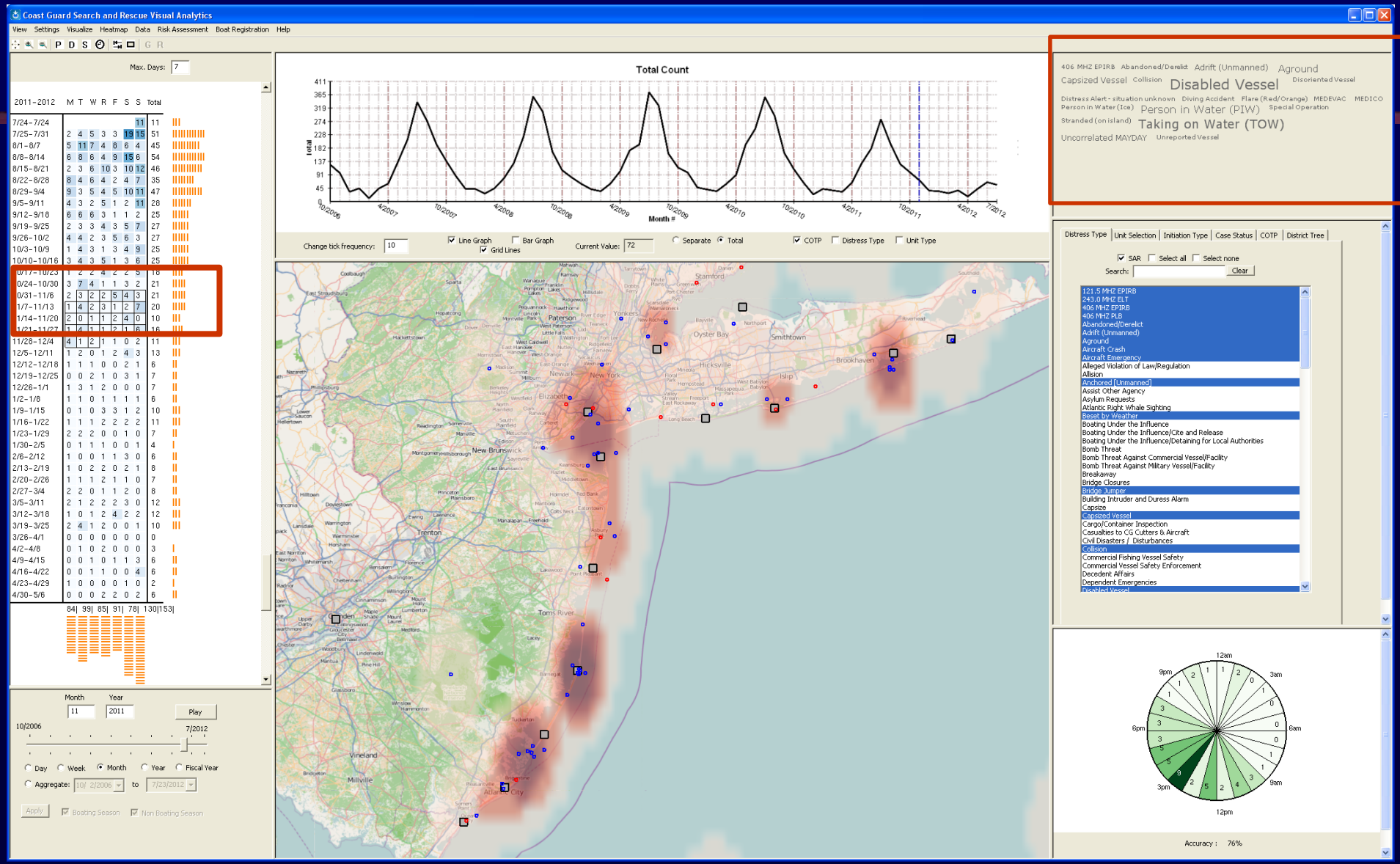
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



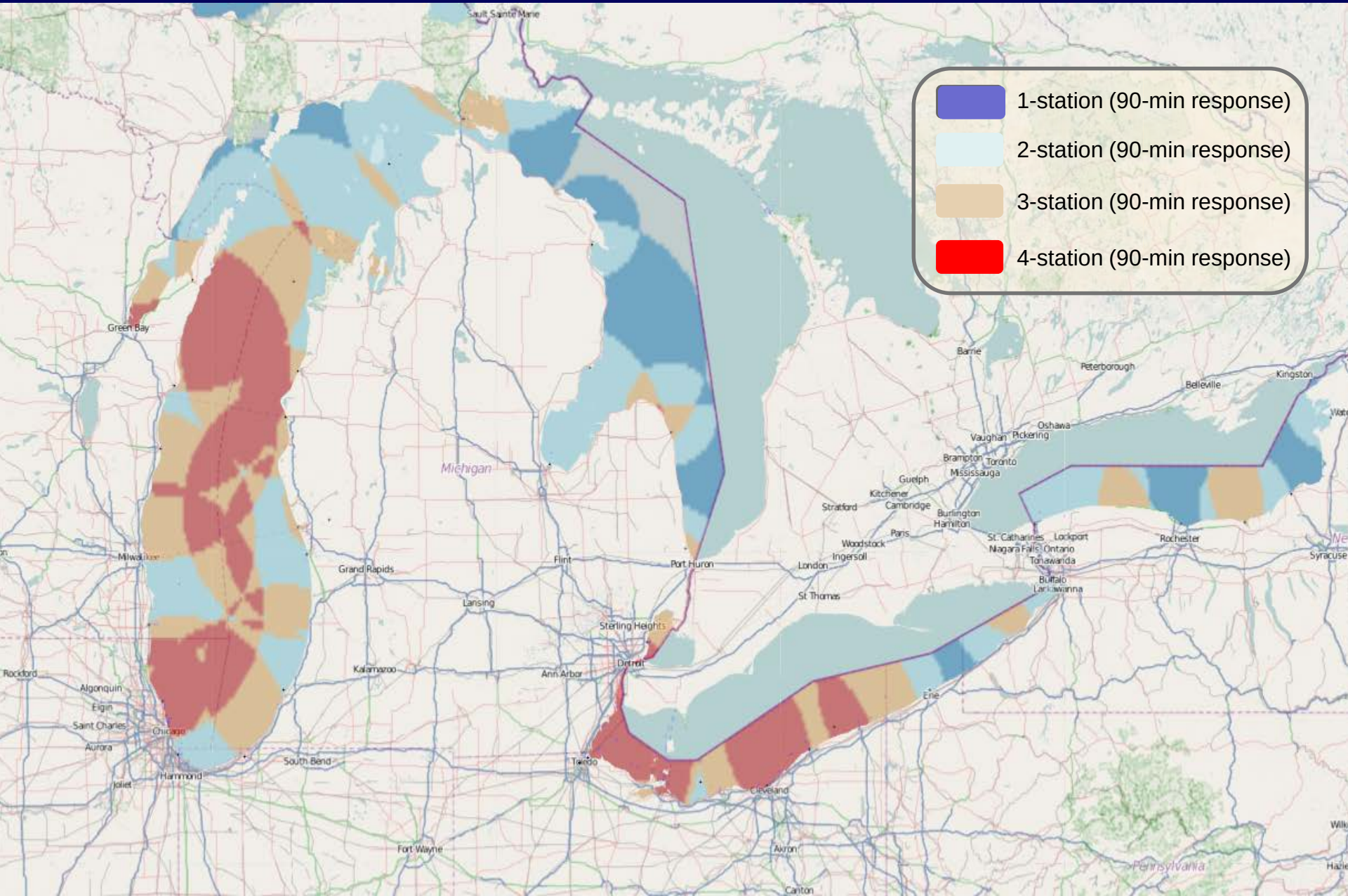
VACCINE

Example: Risks and Consequences From Sandy:

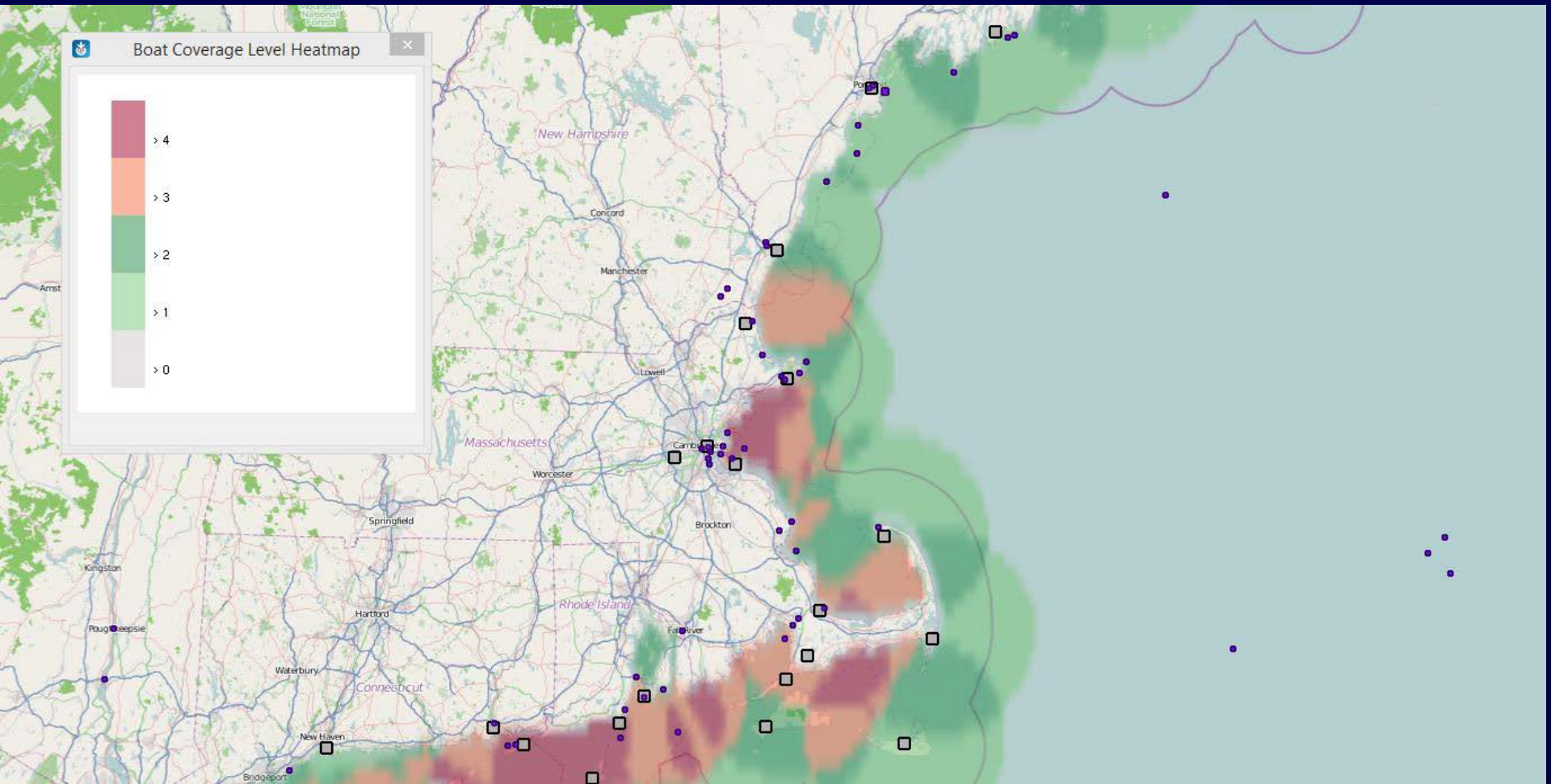
SAR Cases November 2011 NJ/NYC Area



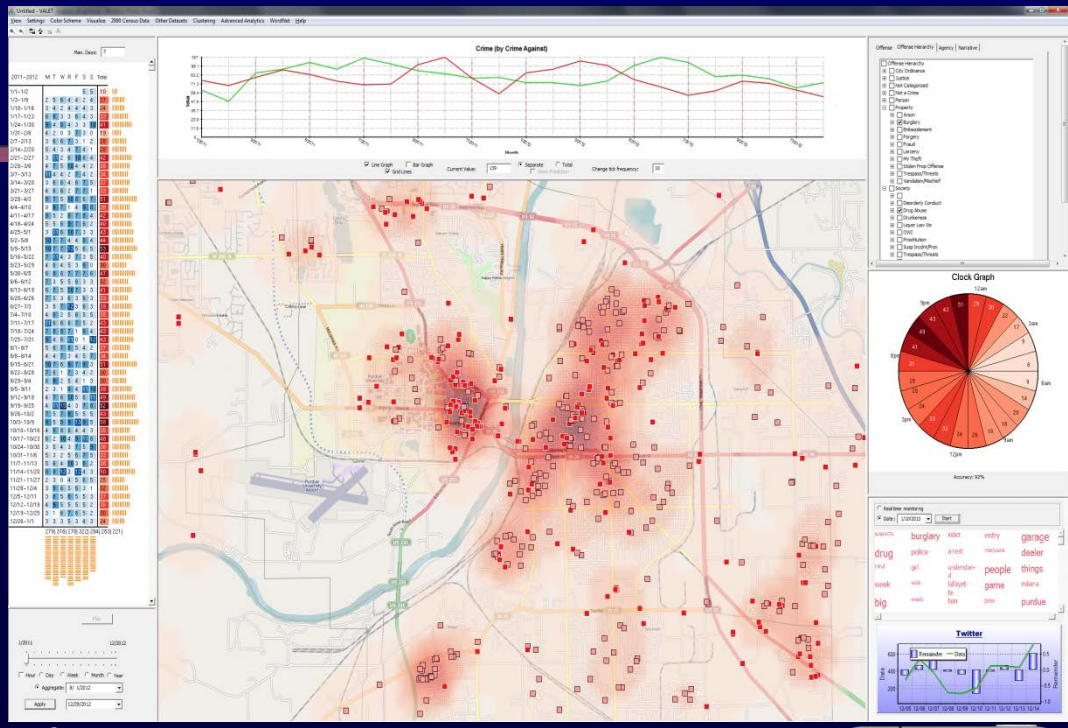
Response Efficiency – Potential Future Assets



Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

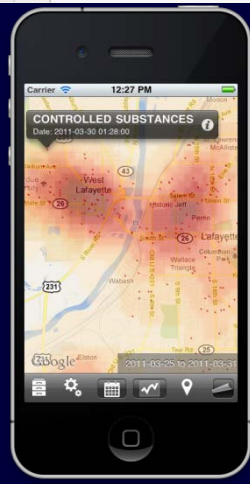
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

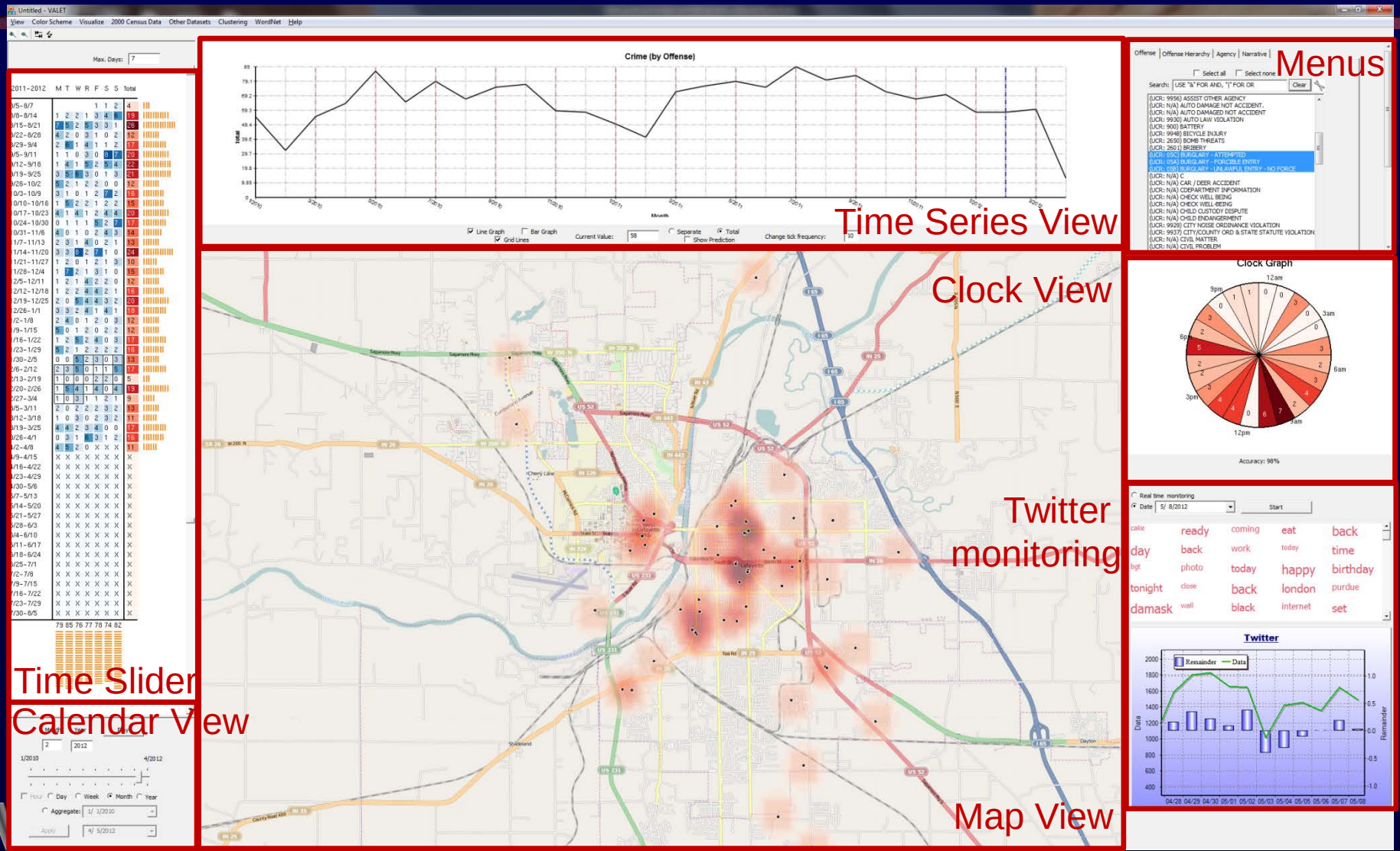
iVALET delivered:

- October 2011: Purdue, WL Police

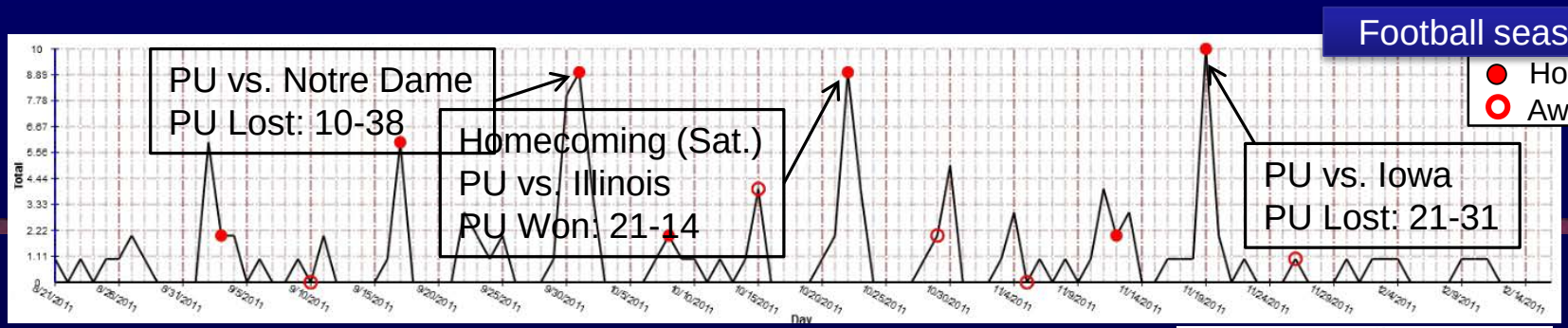


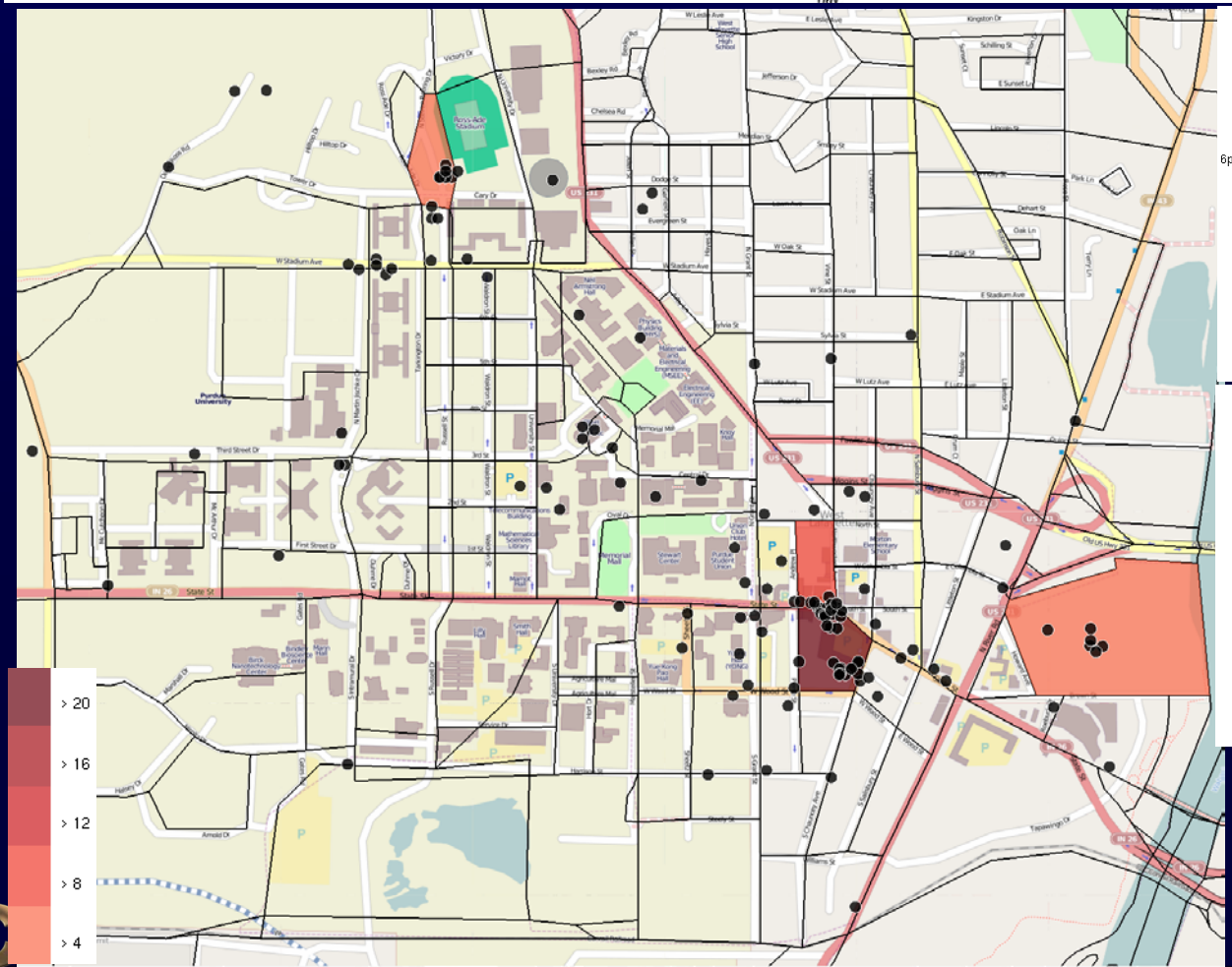
VACCINE

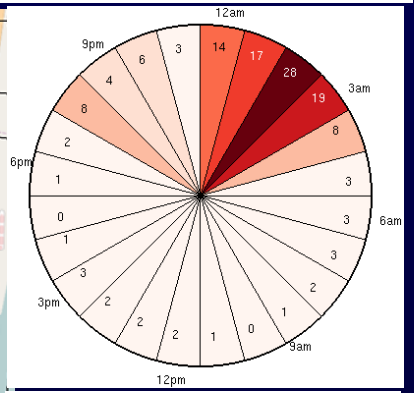
Example for Crime Analytics: VALET

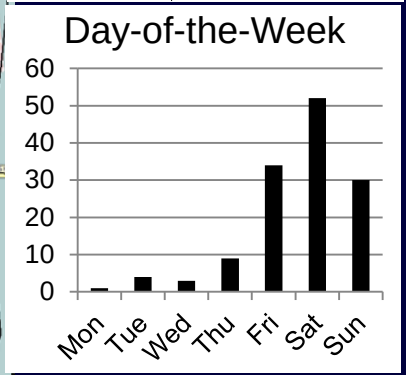


Example: Drunkenness / Public Intoxication









Social Media: Real-time Visual Analytics

(Purdue, Stuttgart, Penn St.)

- Anomalous topic extraction (LDA, novel STL based estimation)
- Dynamically linked views
- Topics extracted shown as a dynamic word cloud
- Sources: Twitter, facebook, flickr
- Uses:
  - USCG Sector Ohio Valley
  - Indianapolis PD, Purdue PD
  - Fleet week Nov. 2014: USCG PAC



Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

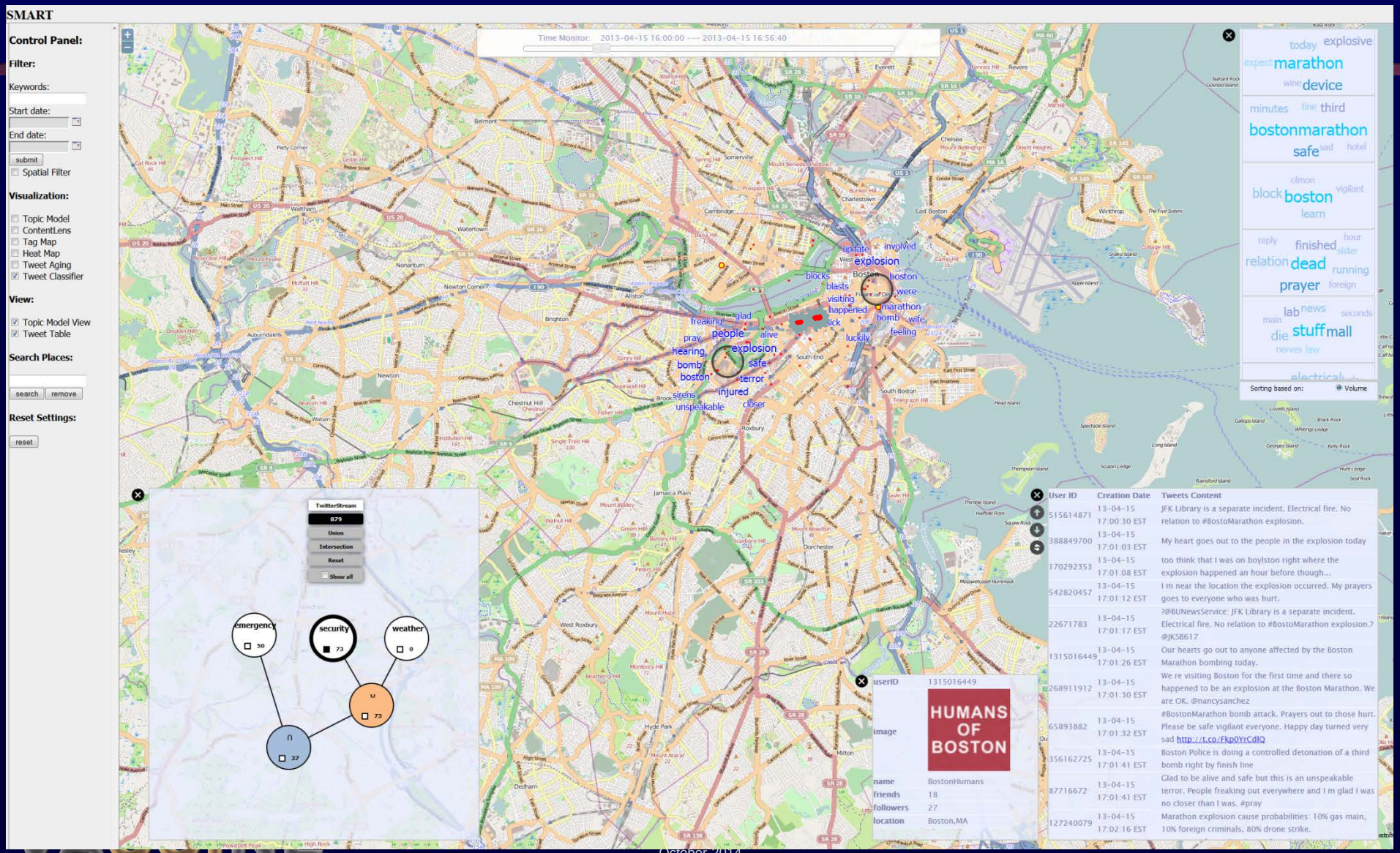
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)



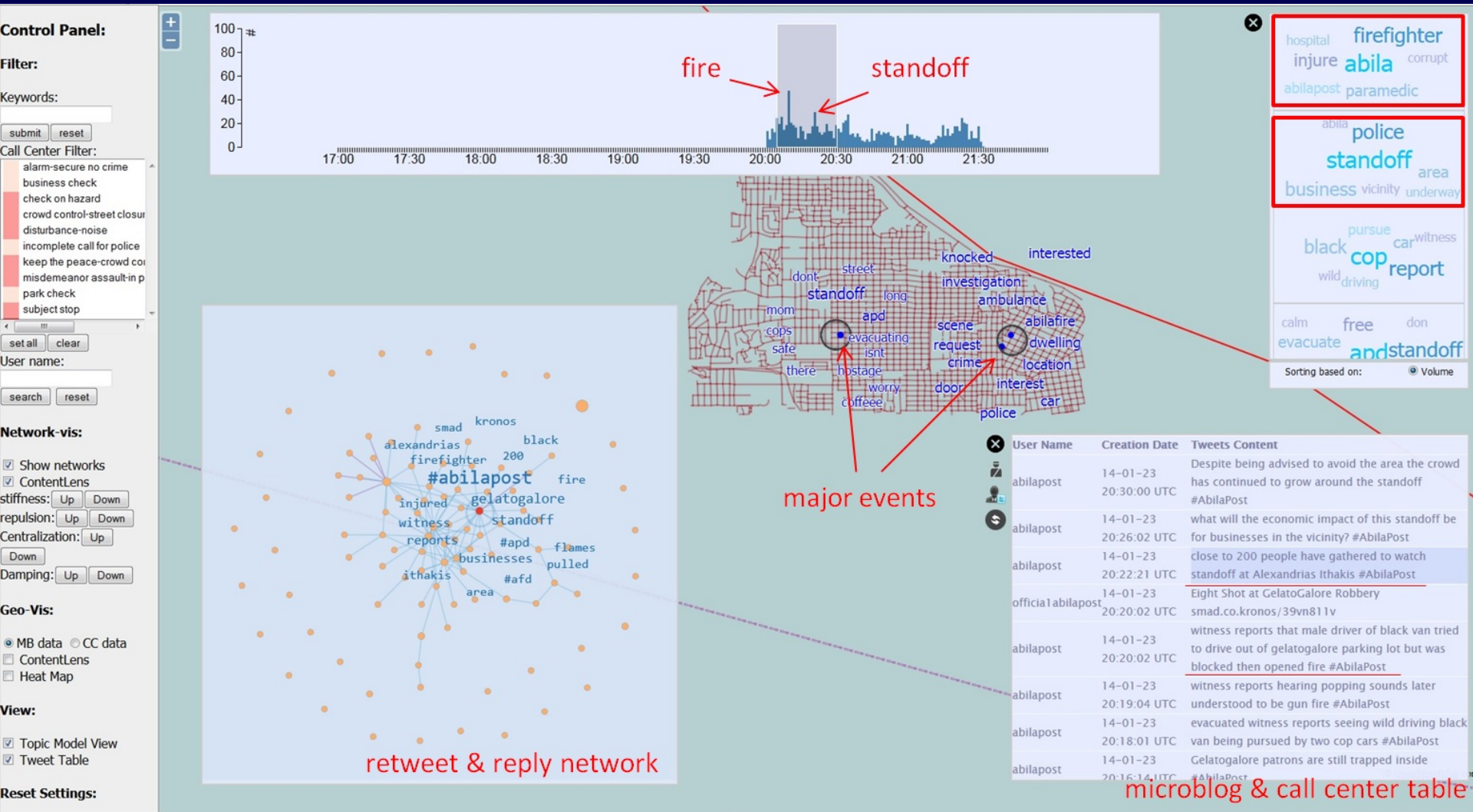
VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



Social Media Analysis and Reporting Toolkit (SMART)

SMART: Network Linkage



For Further Information

www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for Spatially Data-Driven Decision Making

David Ebert

Purdue University

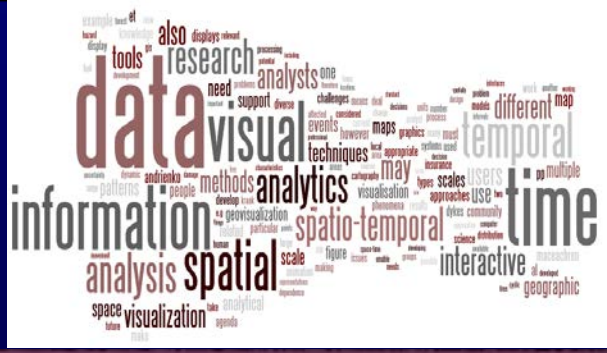
VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

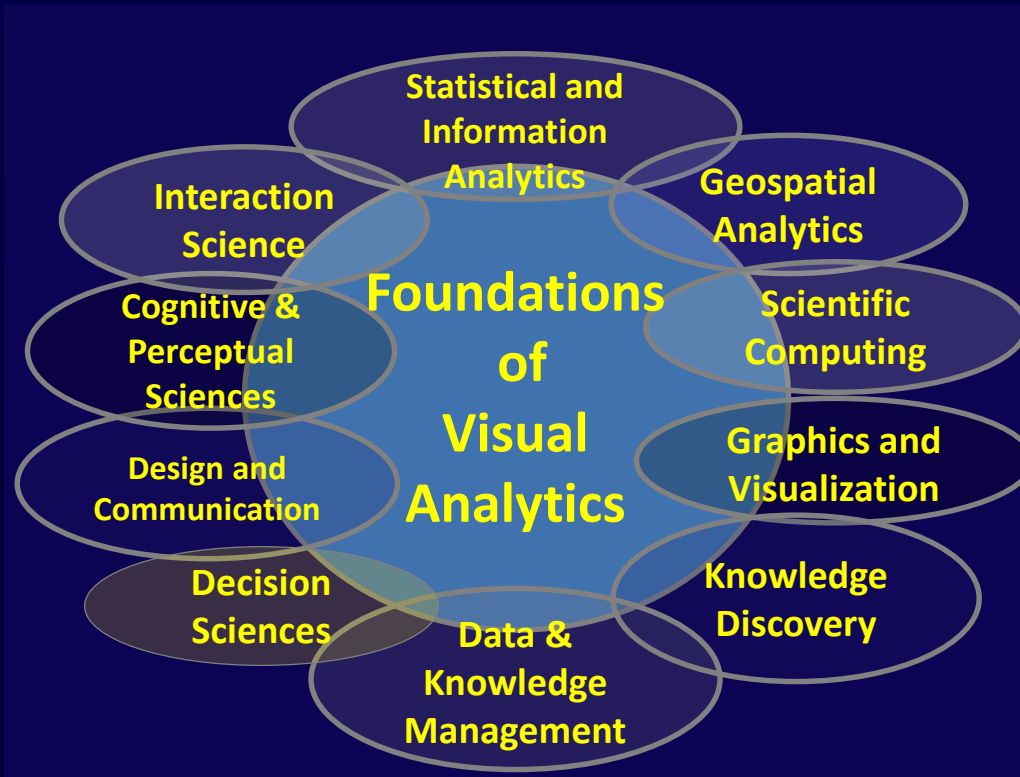
October 2014

Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

VACCINE - Who We Are:
International Team of Experts
75+ Faculty, 27 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of California, San Diego
- University Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- Oak Ridge National Laboratories
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Duke Energy, Bank of America





VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**
  - Interdisciplinary world leading team of researchers and students

"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address



VADM Robert Parker with VACCINE student researchers (cgSARVA, COAST, iOPAR)

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

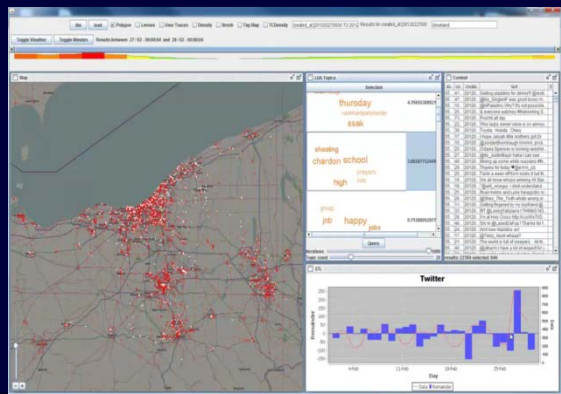
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

Visual Analytics Uses for Public Safety



- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

Our Spatiotemporal Visual Analytics Work

- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale analytics
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

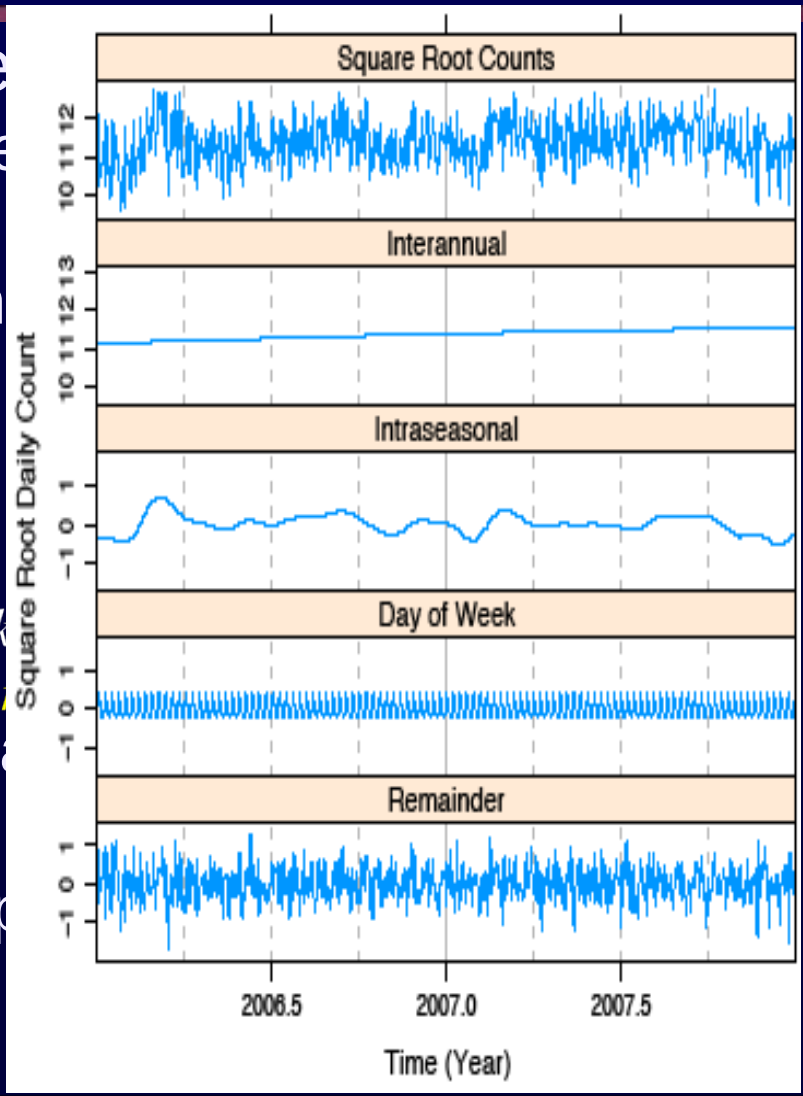
Event/Anomaly Detection - STL

- Temporal trend analysis and pre
  - Seasonal Trend decomposition base
    weighted regression) (STL)
  - Time series data viewed as the sum

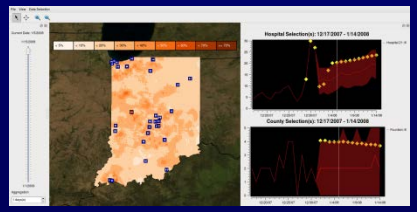
$$Y = T$$

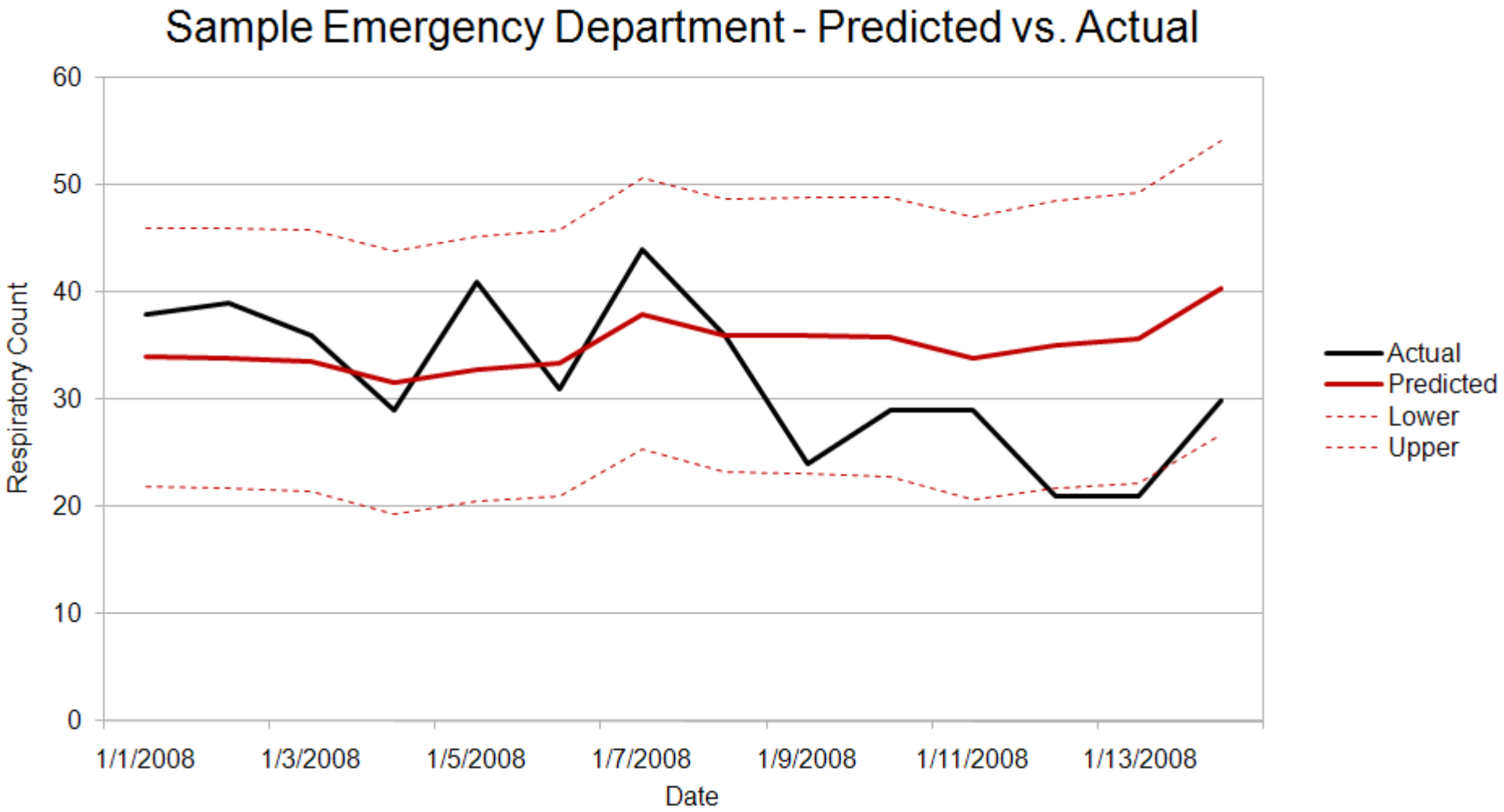
Y : *original time series*
T : *trend component*
S : *seasonal (daily/week*
R : *remainder compone*

  - Large value of *R* indicates substanti
  - Filter out the noise component
  - If value varies by 2 SD from STL exp



Predictive Visual Analytics



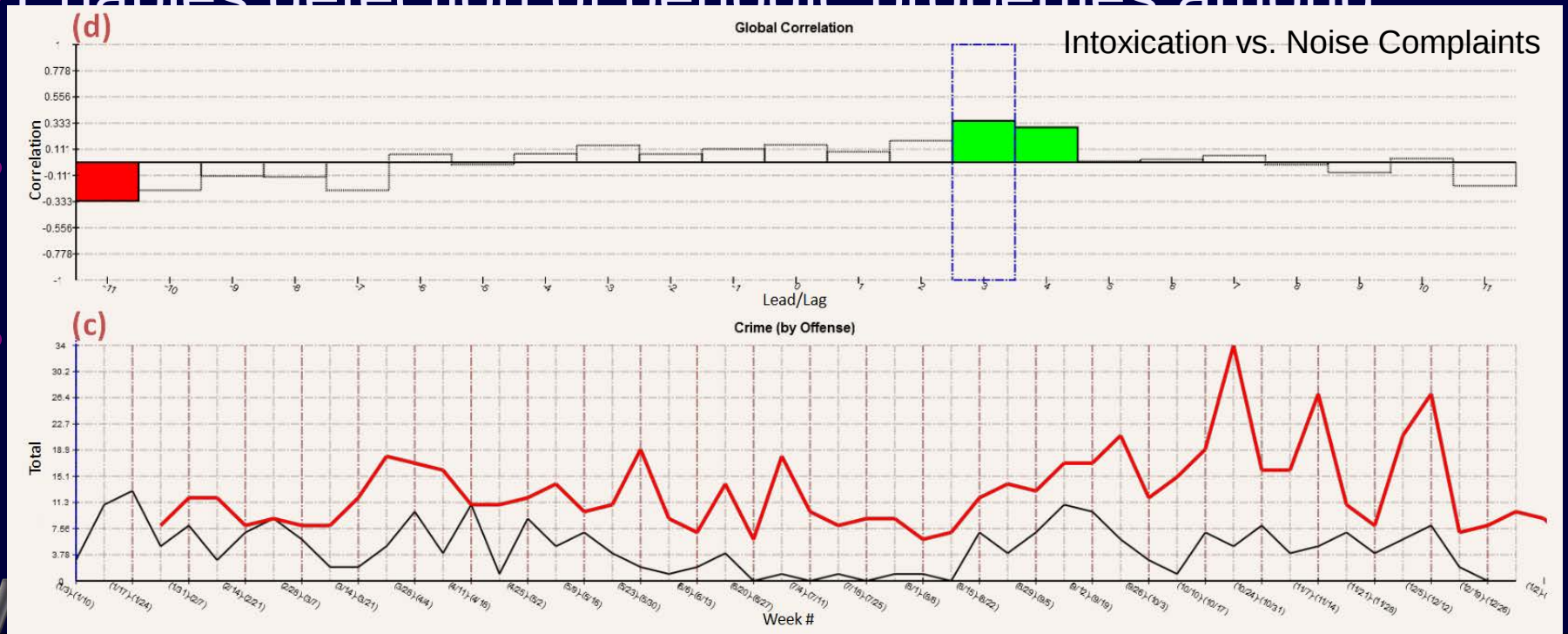


Visual Correlative Analytics: Value of Visual Correlative Exploration?

- Understanding the relationships between spatial and temporal trends
  - Acts as precursor to creating predictive models from data
- Provides insights into the workings of real world environments
- Different forces that determine nature of relationships
  - E.g., one event precursor to another
  - Requires exploration of leads/lags between datasets
  - Determine if one dataset is related to another at a different temporal location
- Aids analysts and decision makers in hypothesis generation

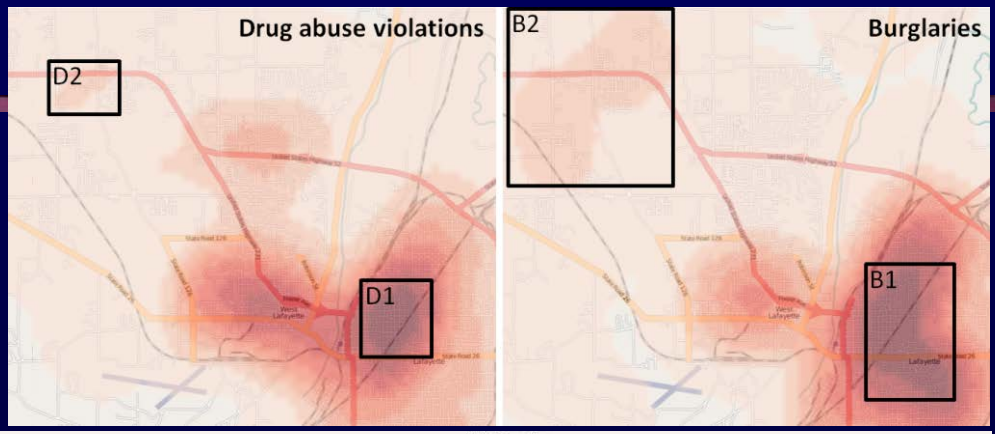
Multivariate Correlative Predictive Analytics

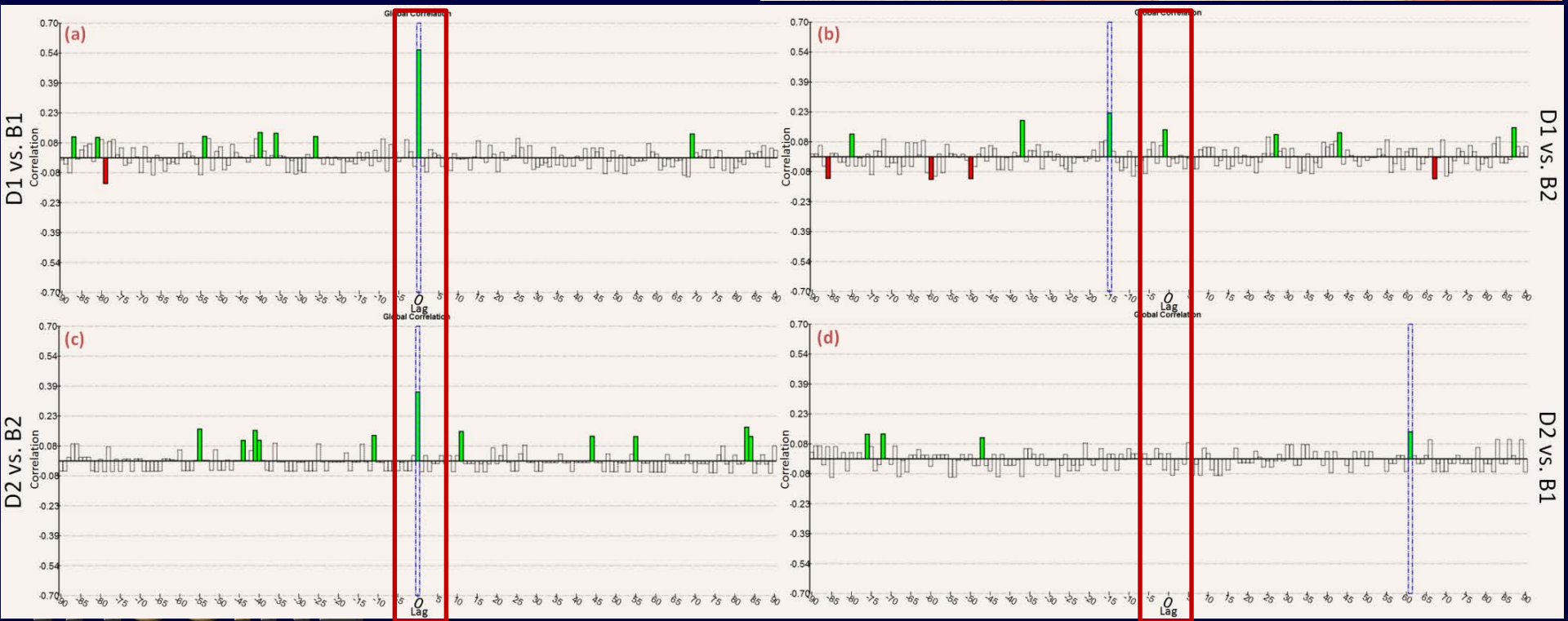
- Two Key Features:
  - Automatic correlation computation against lead/lags
  - Temporal and spatial windowing
- Enables detection of periodic properties among



Spatial Windowing: Drug Abuse Violations vs. Burglaries

- Over geospatial hotspots





The Scale Problem

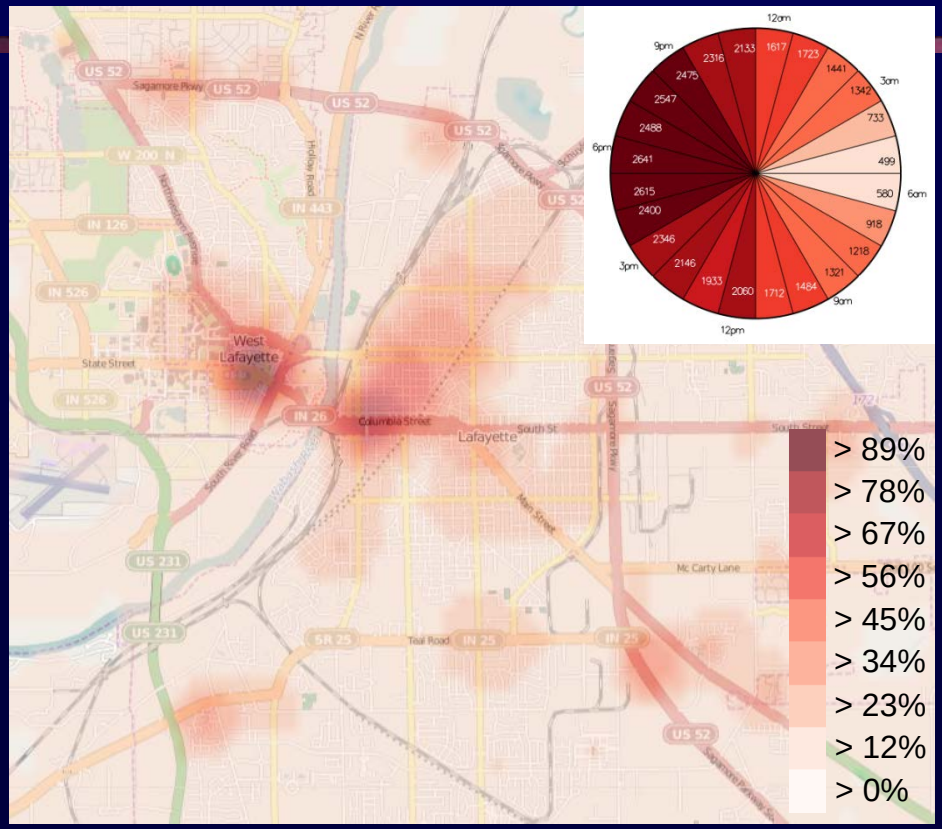
- Data scale:
  - Coarse scale analysis is quick but can have too much variance or contain combinations of multiple signals reducing effectiveness
  - Fine scale analysis is fast but may limit applicability because of fine resolution and can have data sparseness issues
- User/problem scale
  - Users perform analysis at scales that are natural to them and the task/problem
  - Characteristics might not be well articulated or available digitally

Natural Scale Templates

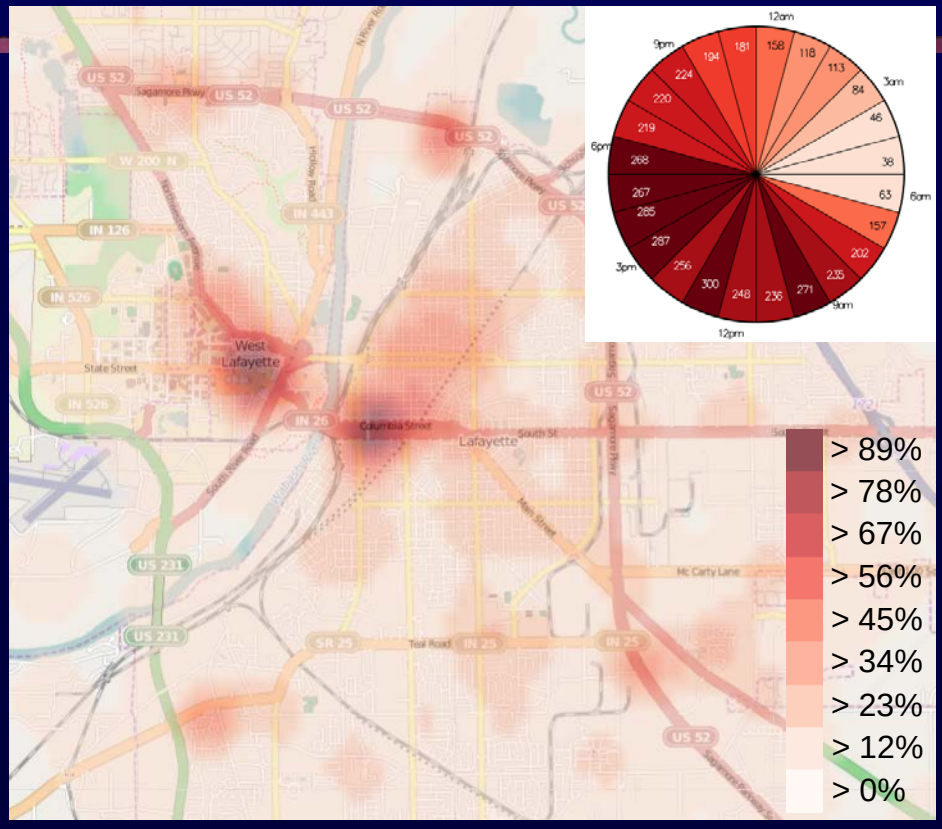
- Spatial and temporal visualization and analysis templates that balance data scale issues and human decision maker scale issues
- Some important questions:
  - How do we facilitate appropriate analysis for the casual expert?
  - How can casual expert improve performance of analytical results?
  - How do we effectively communicate the result, accuracy, and certainty?

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data

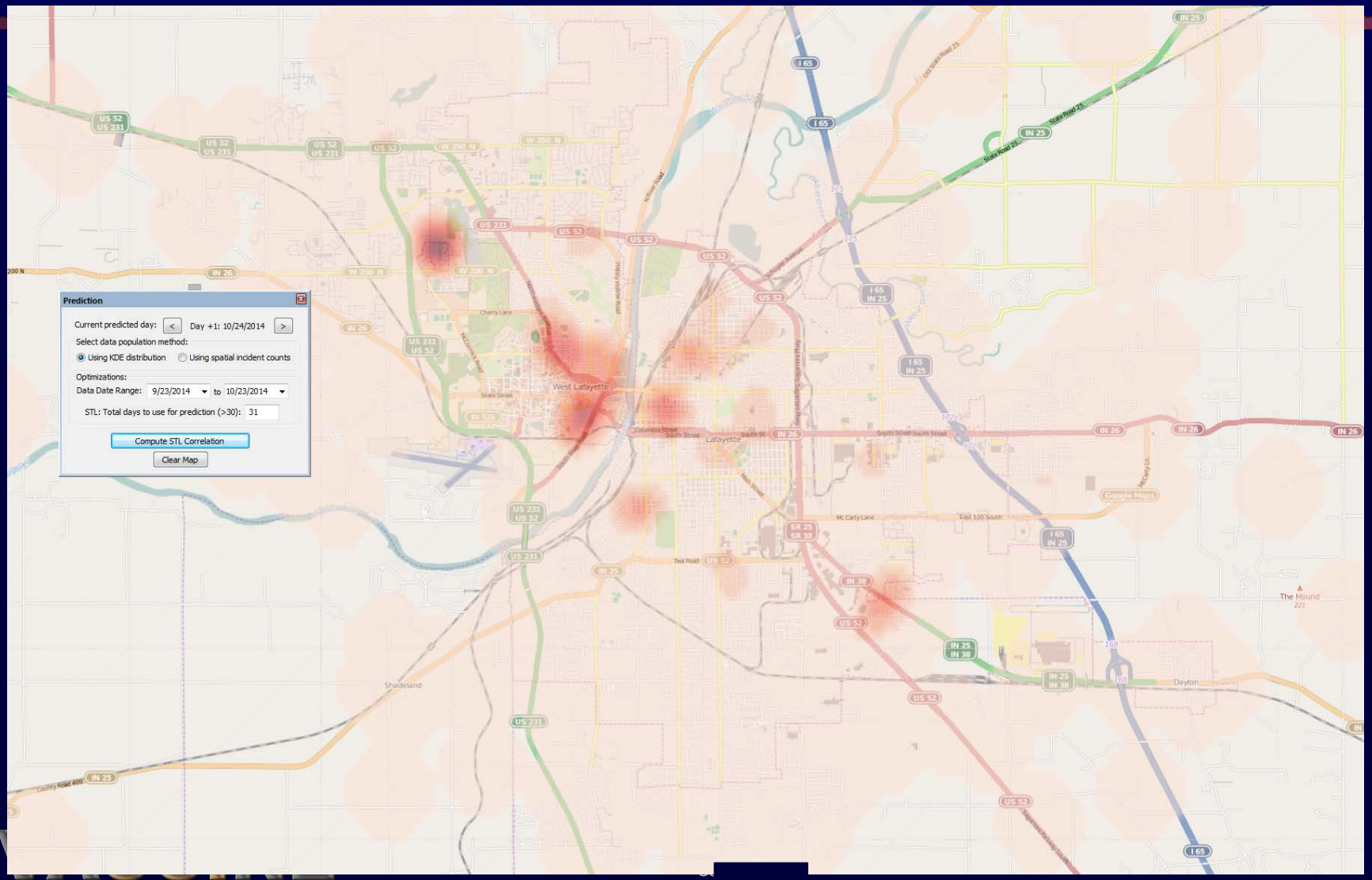


Past 2 Years
3/11/2012 – 3/10/2014

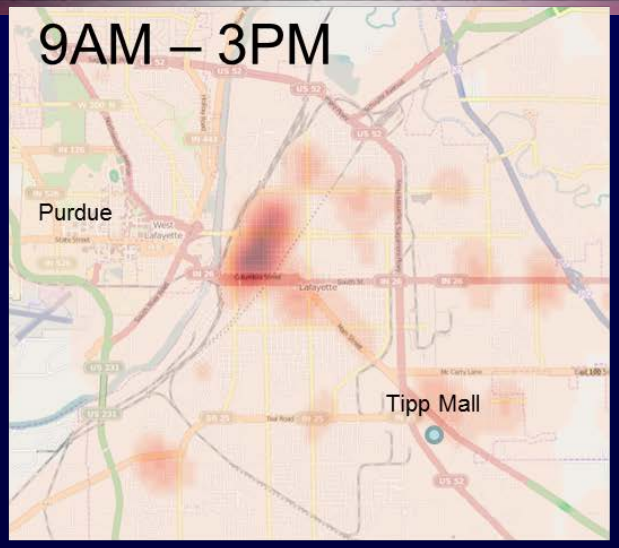


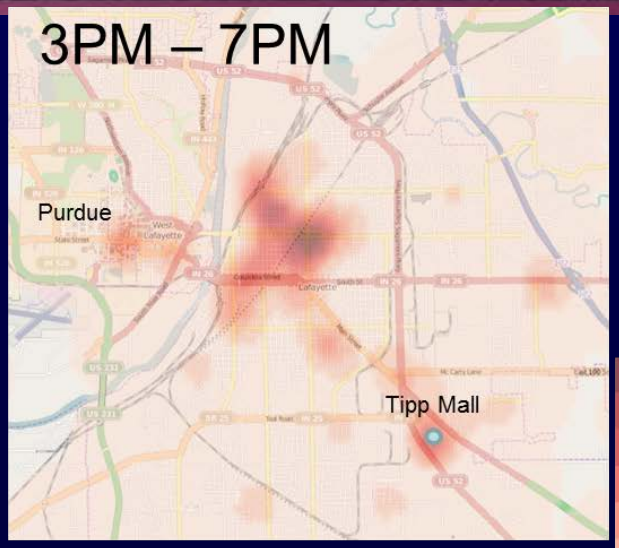
March + Tuesday Filter
(Past 10 Years)

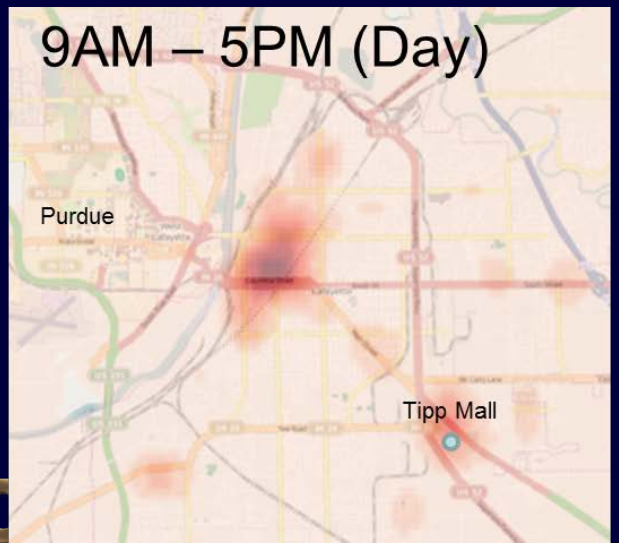
Prediction Using KDE Distribution

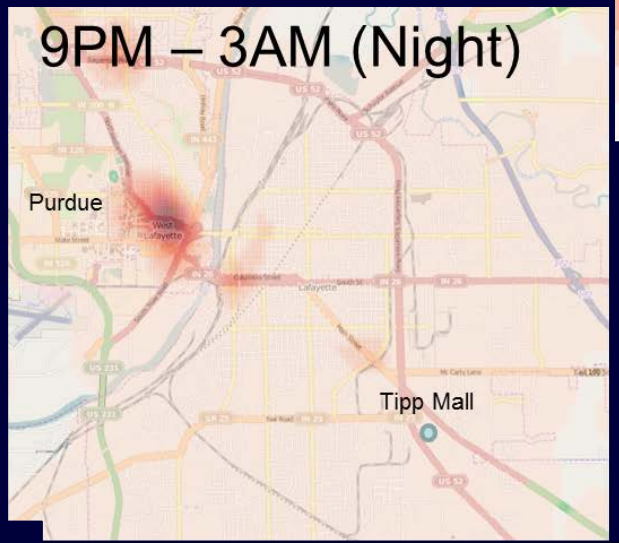


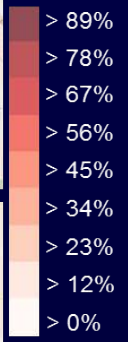
Prediction During Certain Hours In The Day







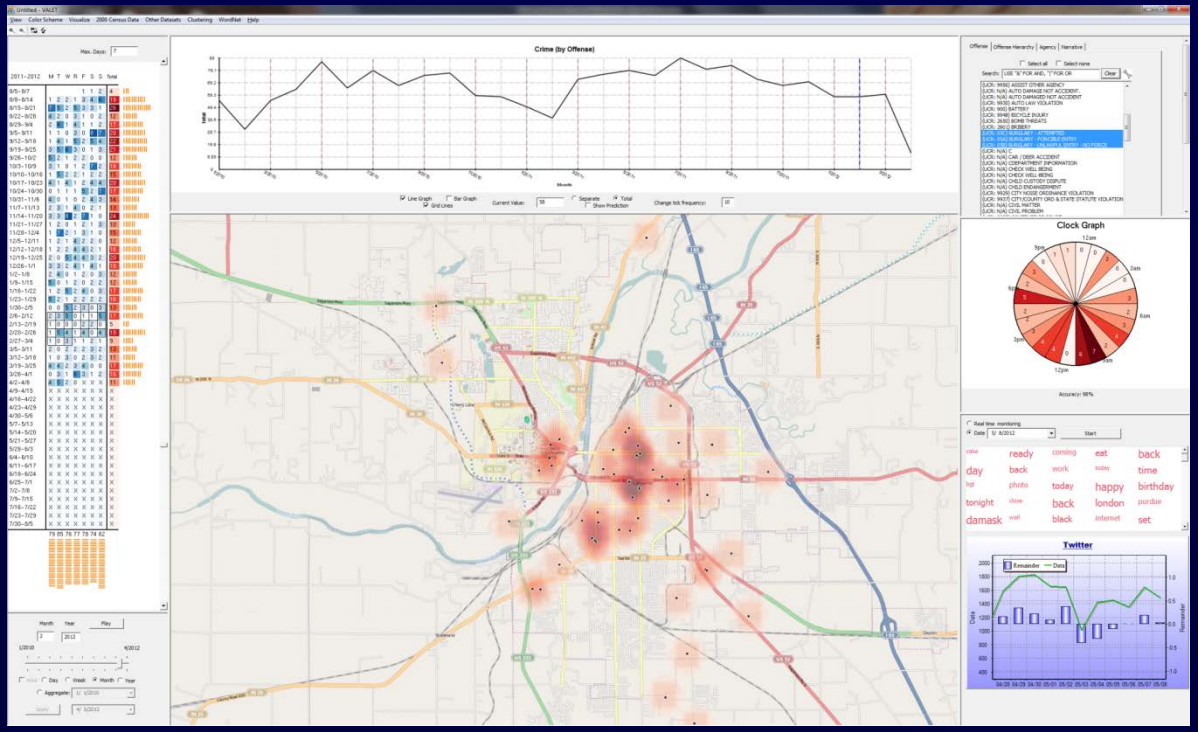


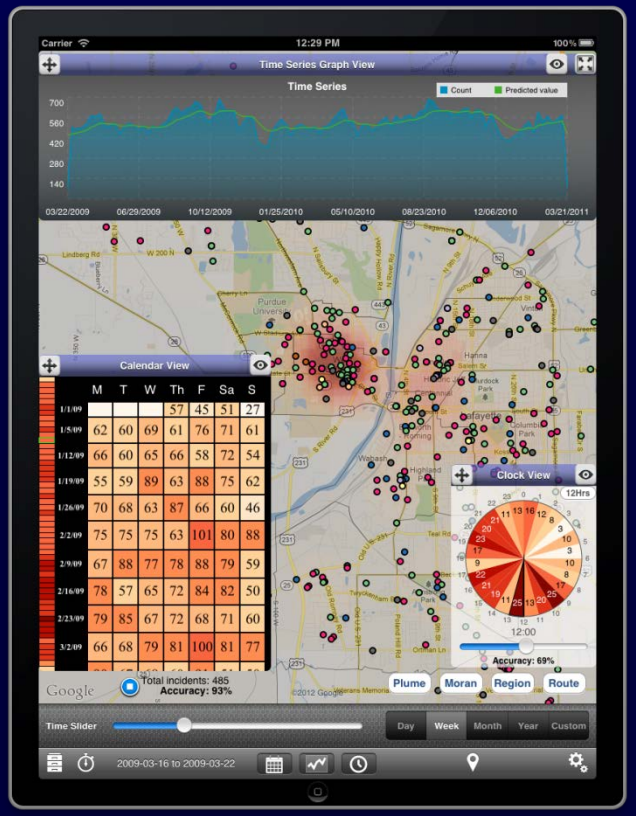


Summary of Prediction Features

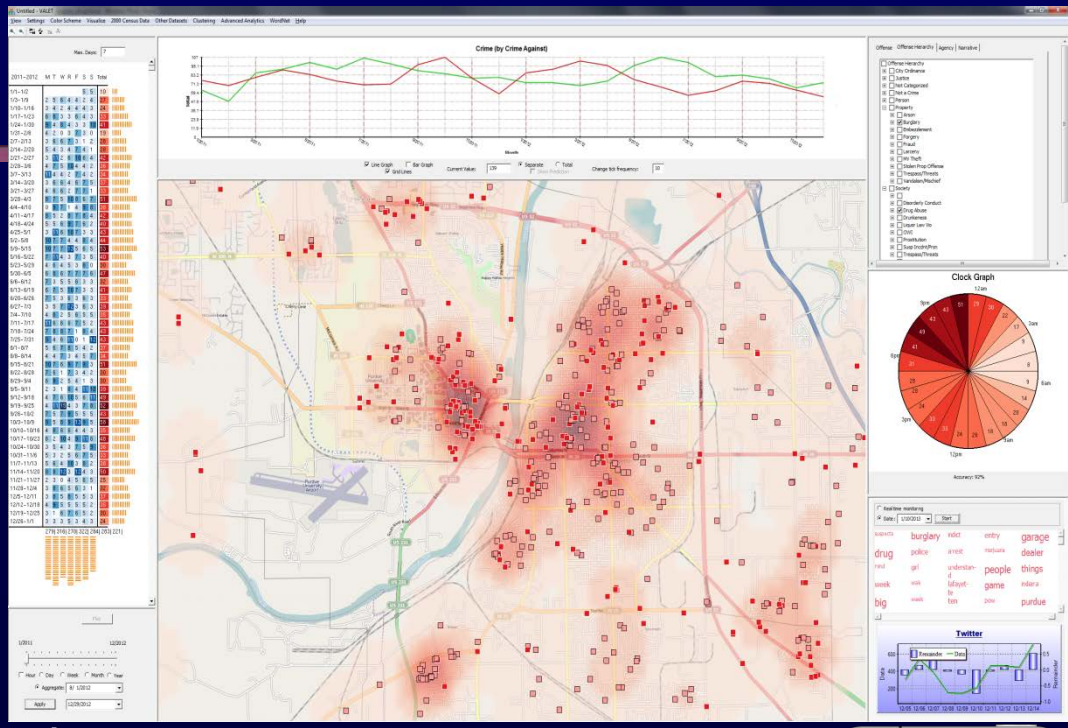
- Novel prediction techniques based on seasonal, day of week, and time of day patterns
- Determination of right spatial and temporal level for reliable predictions
- Correlation of multiple types of crime with lag/lead indicators included
- Location aware updating of relevant information

Visual Analytics Law Enforcement Toolkit (i)VALET





Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

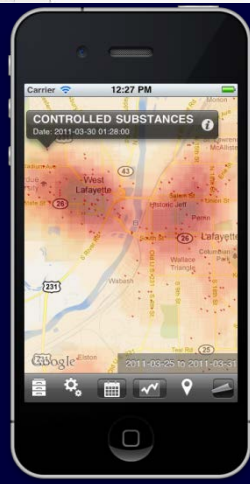
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Mobile version being released to public (Spring 2015) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police
- Fall 2013: Ohio State Highway Patrol
- Spring 2014: NYPD
- Fall 2014: Evansville PD, New Albany PD

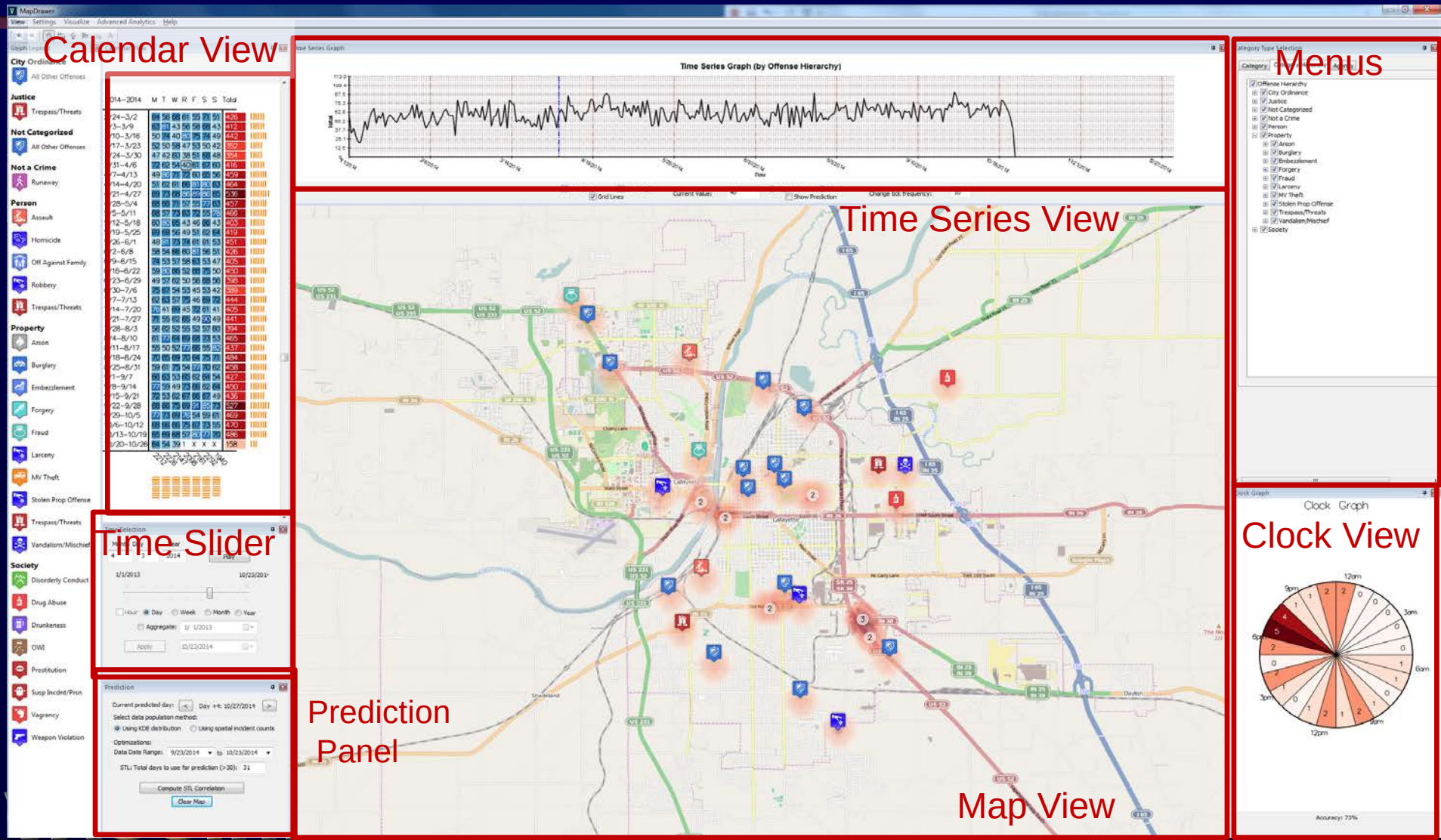
iVALET delivered:

- October 2011: Purdue, WL Police, Lafayette PD

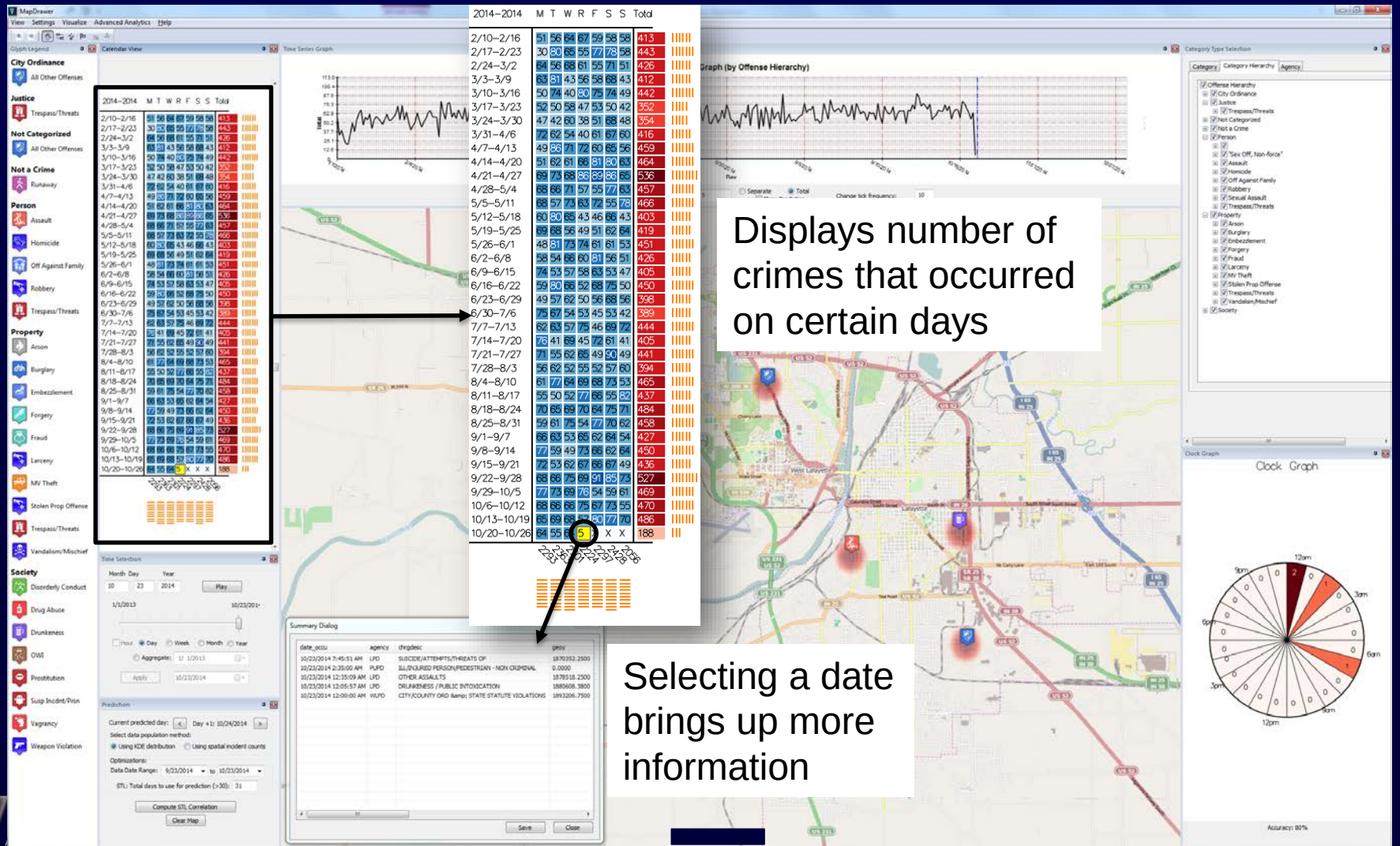


VACCINE

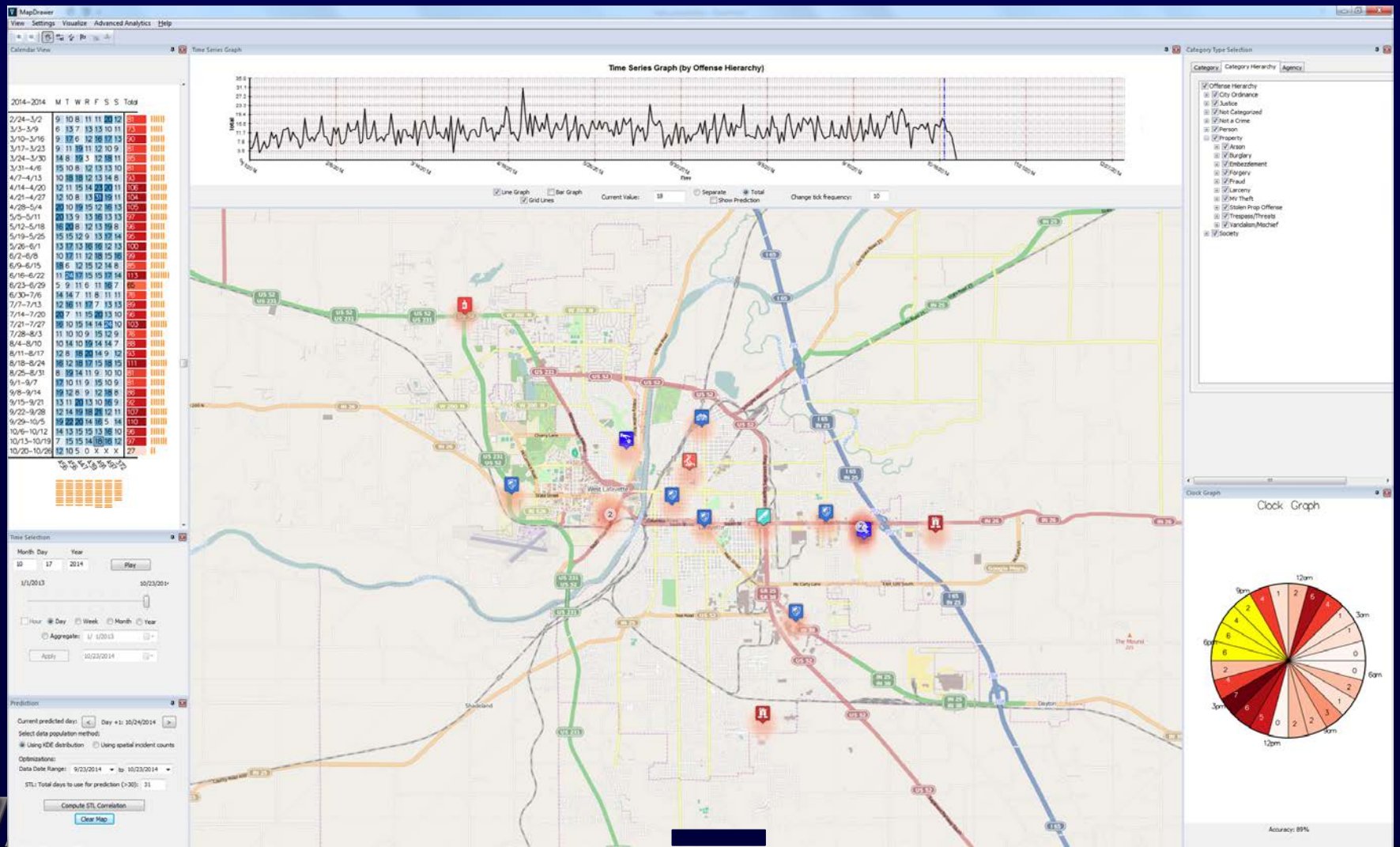
VALET Layout



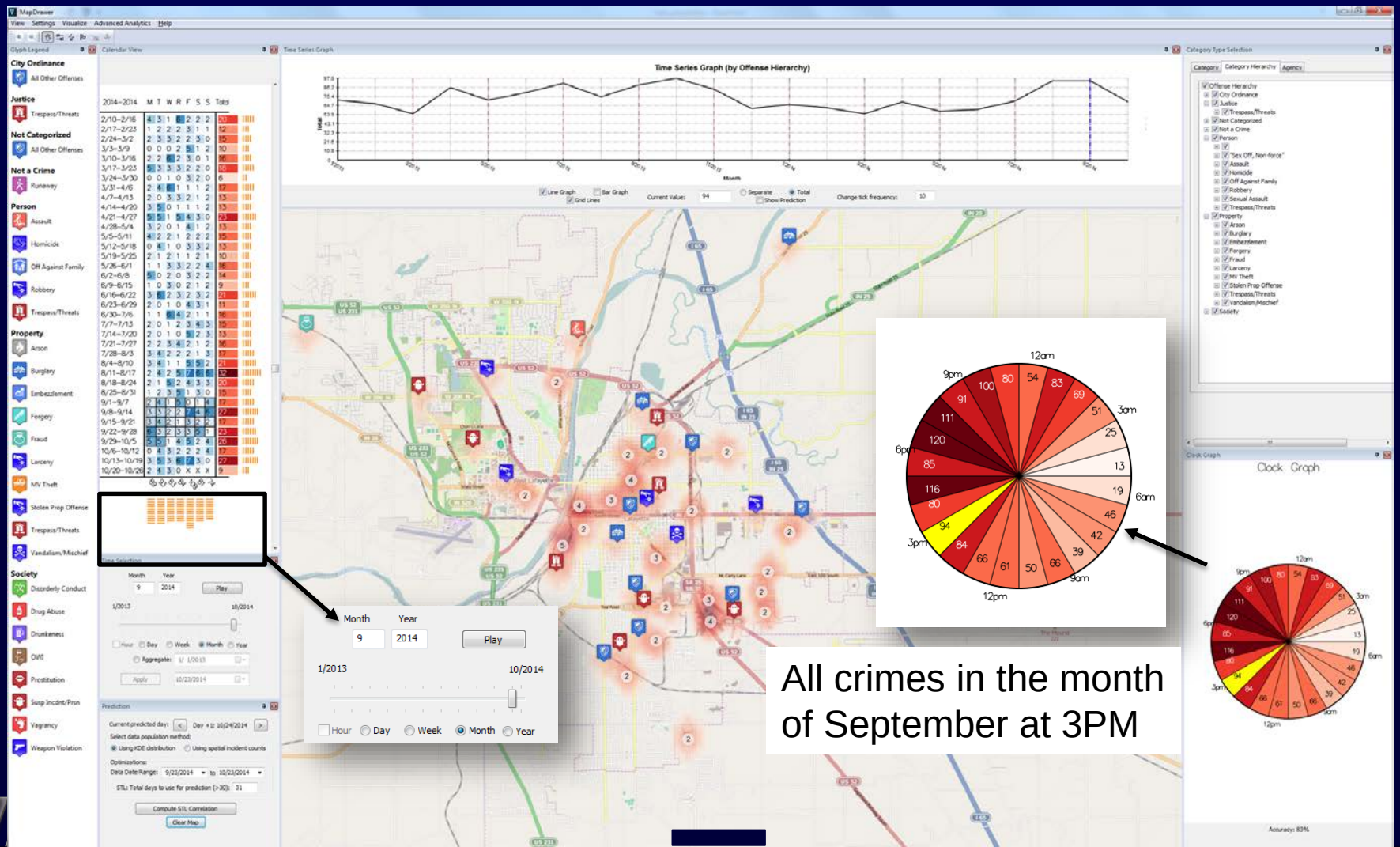
Calendar Display



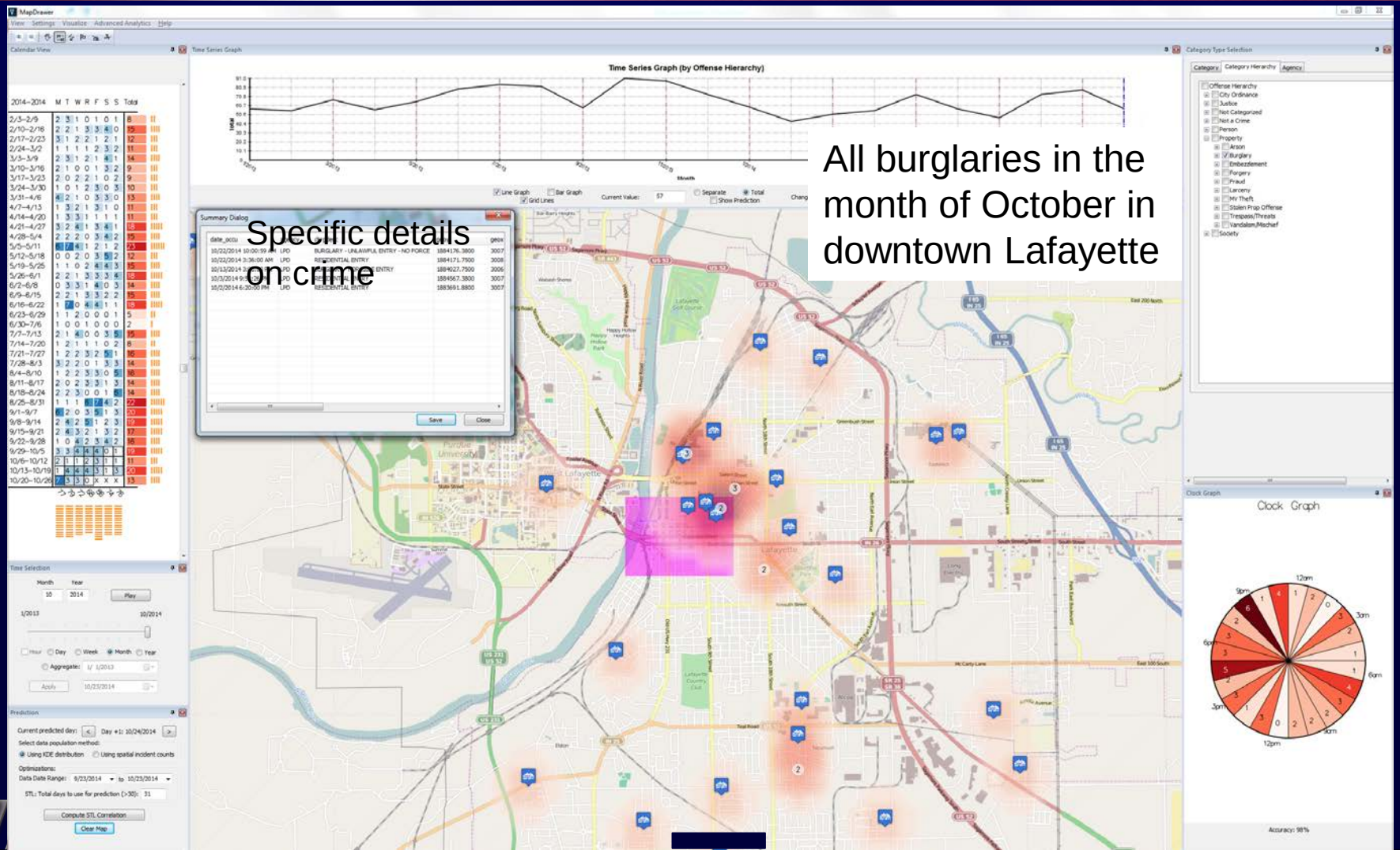
Crimes In A Single Day From 6PM – 9PM



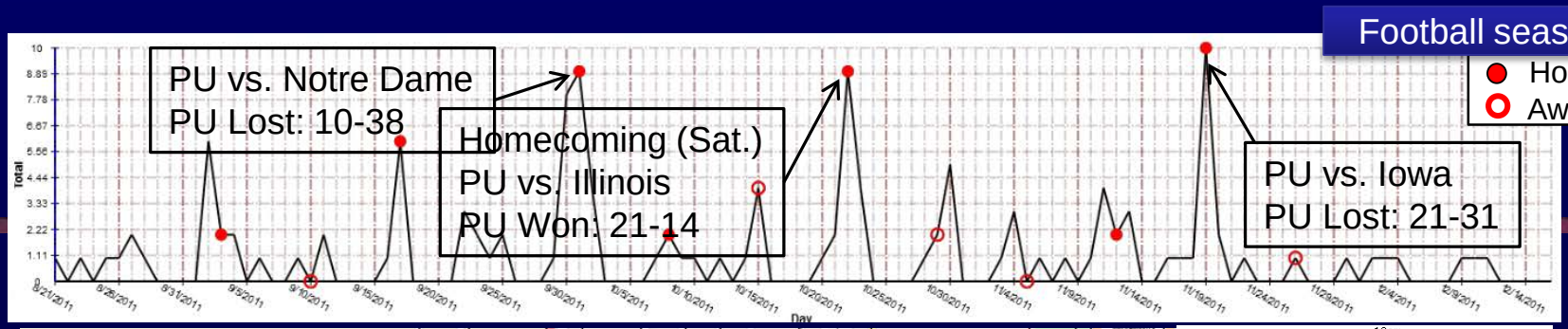
Crimes During Certain Hours In A Month

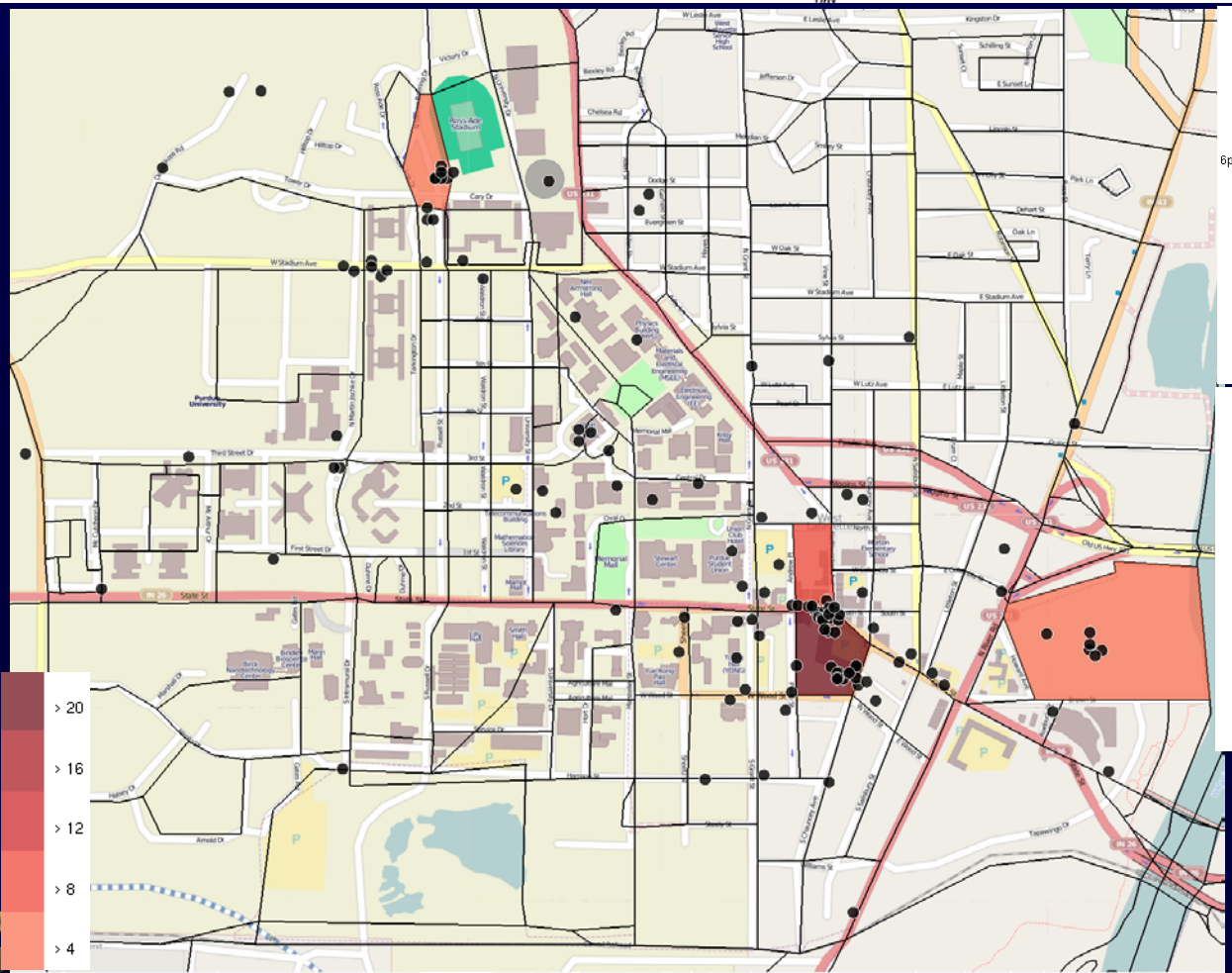


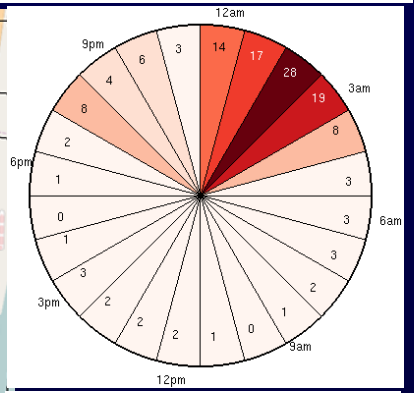
Select Specific Region To Display Crimes

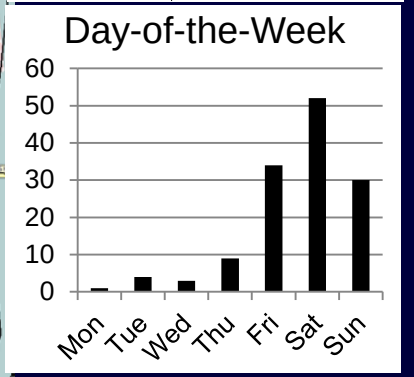


Example: More Data Sources – Intoxication

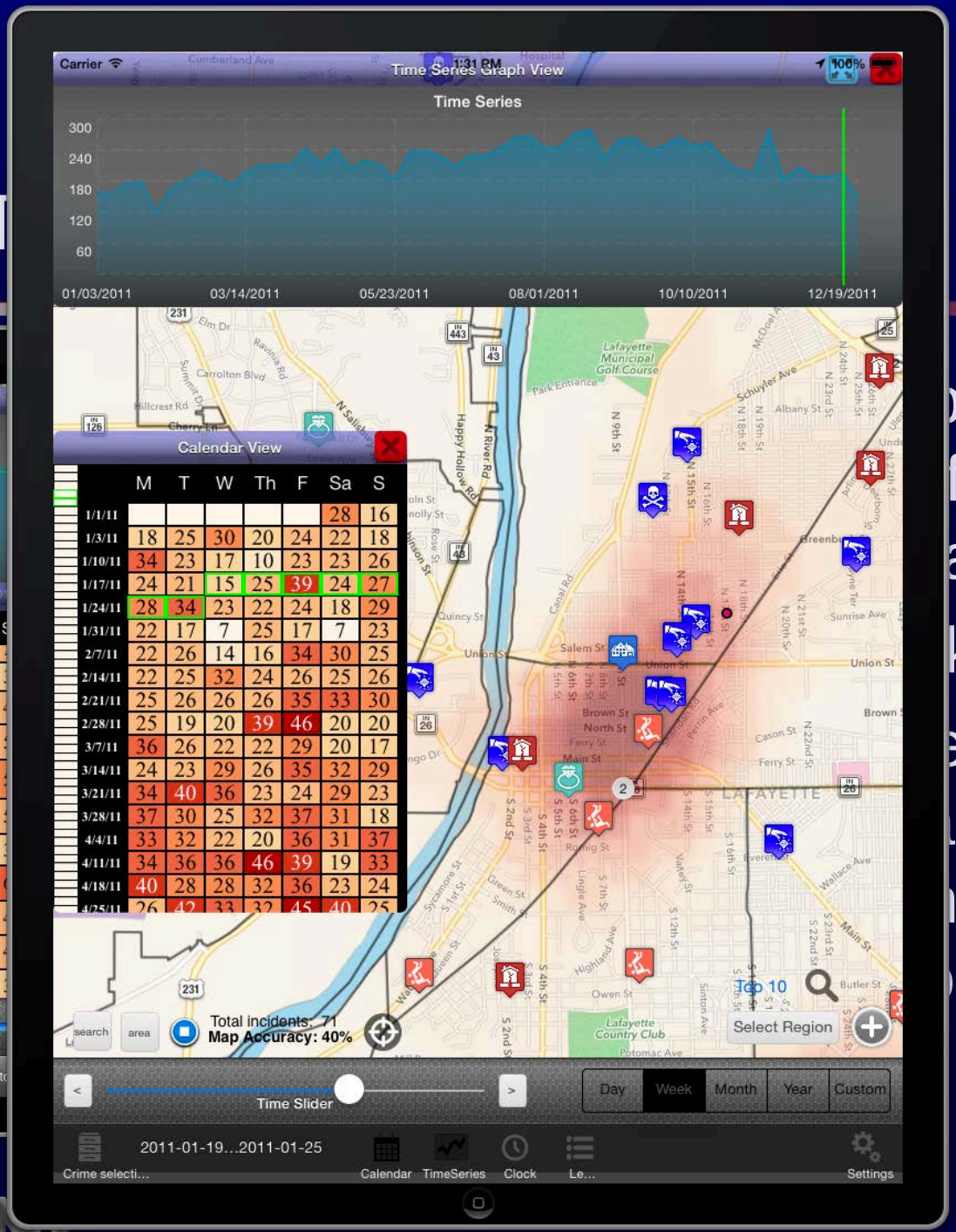




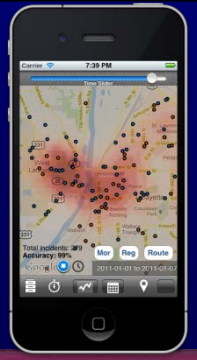


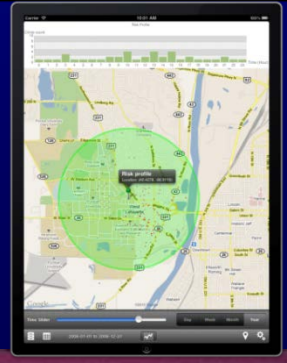


iVALET









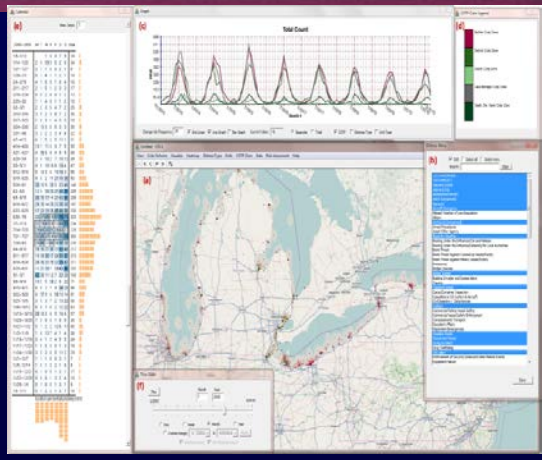


...lore criminal,
...ffic and civil
...a on-the-go
...k assessment
...e current
...tial +
...poral context
... analysis

VACCINE

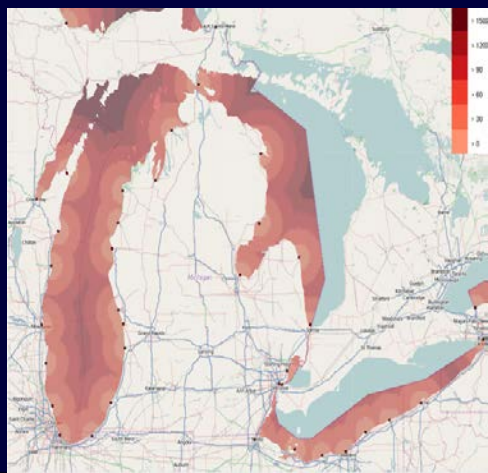
U.S. Coast Guard COAST/ SARVA (cgSARVA)

Partners: USCG LANT 7, USCG HQ 771 , USCG D9, USCG D5, RDC



IMPACTS:

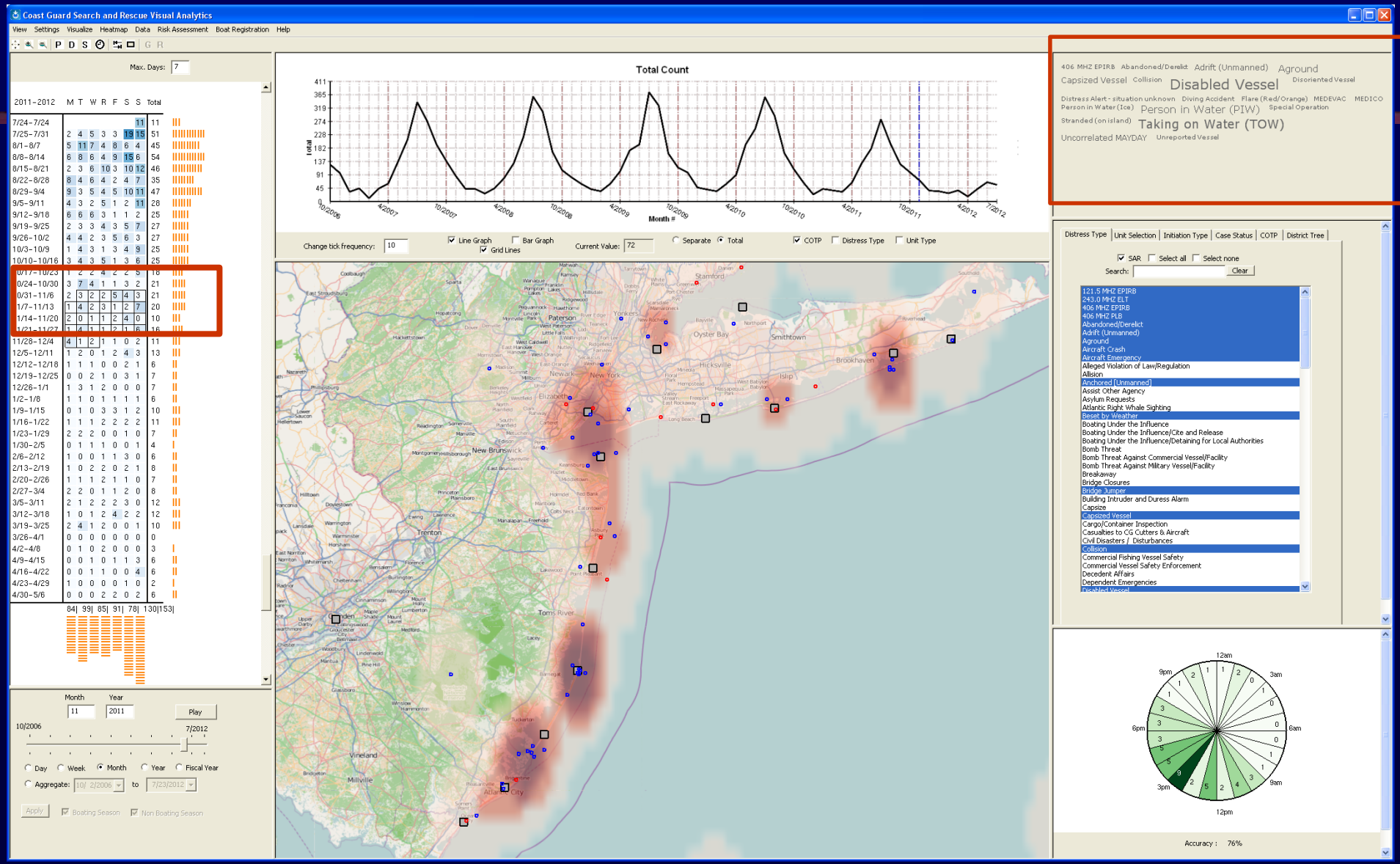
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



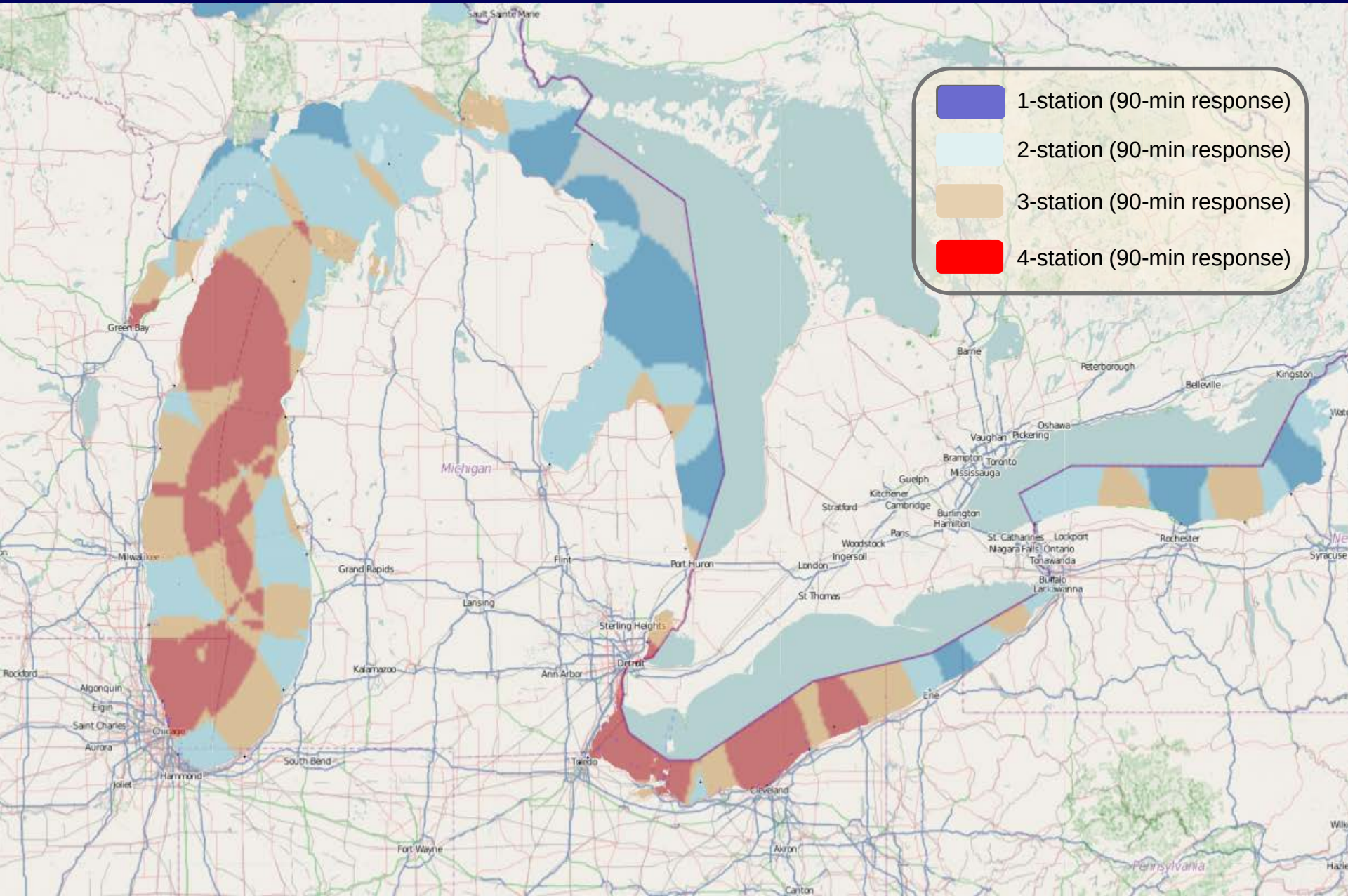
VACCINE

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

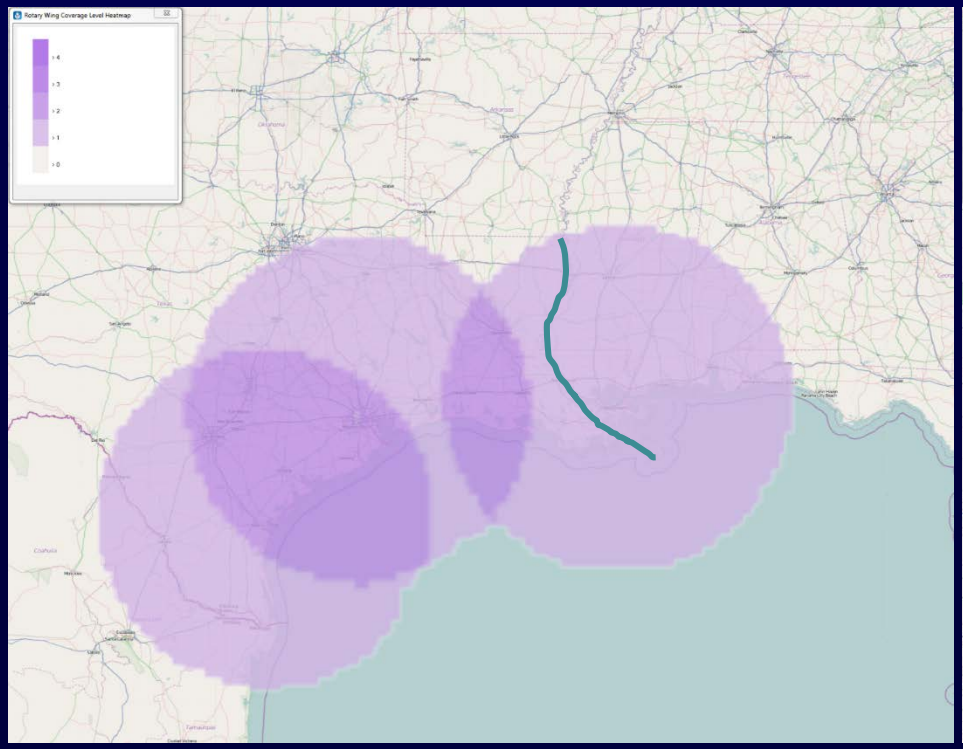


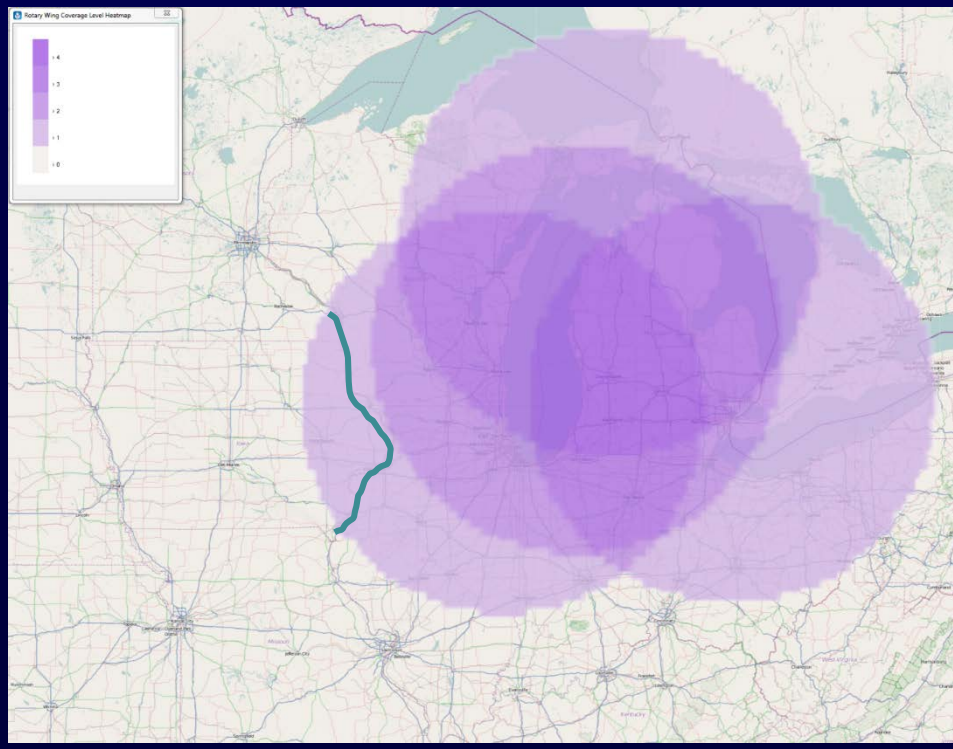
Response Efficiency – Potential Future Assets



Airstation Utility Visualization Comparison

- Rotary Wing Coverage Map by AirSTA

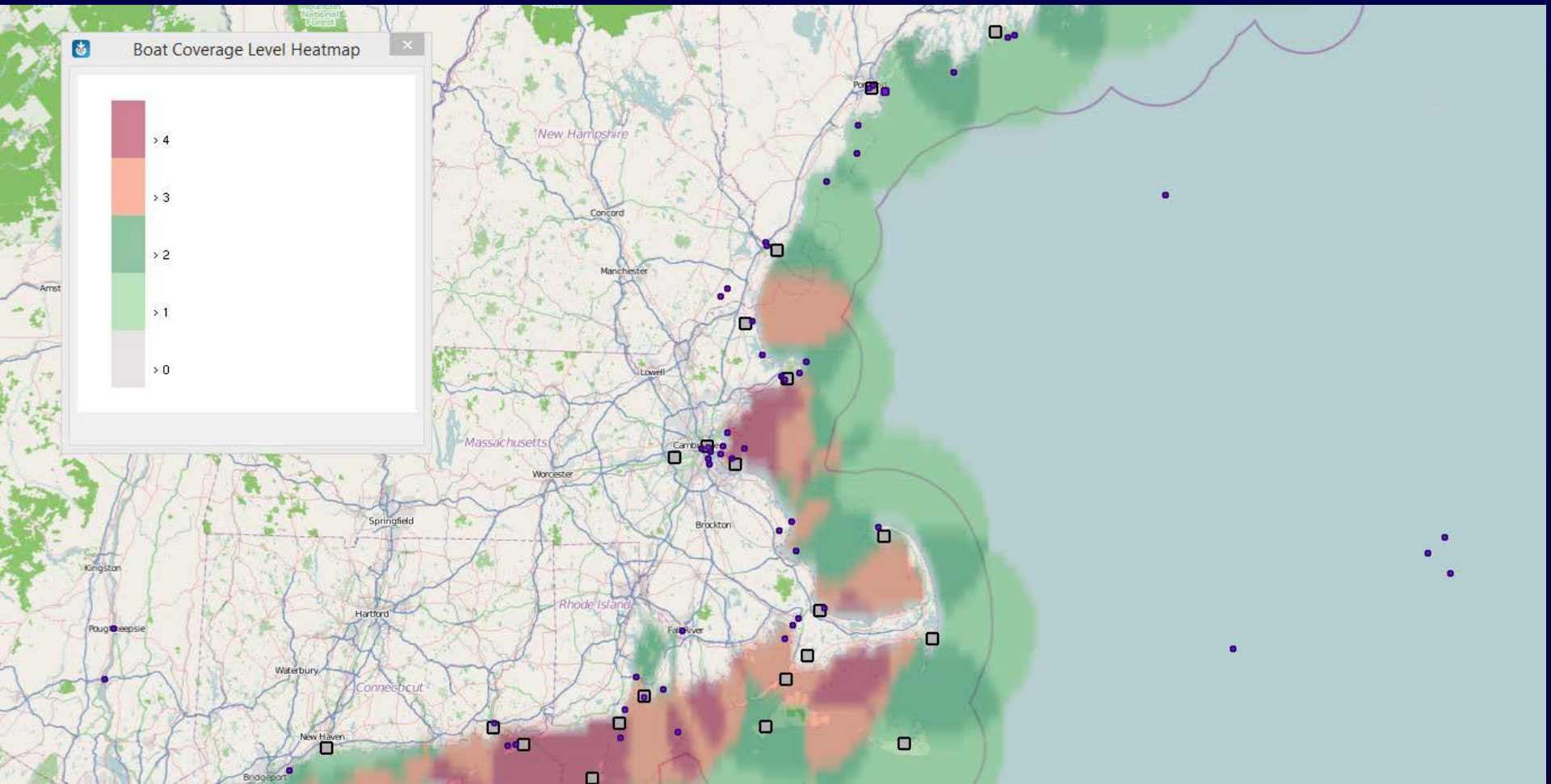




Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

VACCINE

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



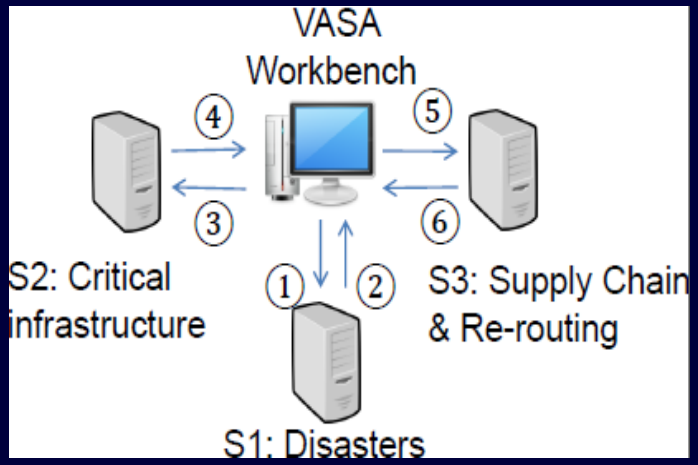
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s):
  - Power Suppliers (e.g., Duke Energy)
  - Cyber Community (e.g., Cisco)
  - Quick Service Restaurants and suppliers,
  - Logistics transportation
  - Food supply monitors

Purdue

UNCC



U Minn

TACC

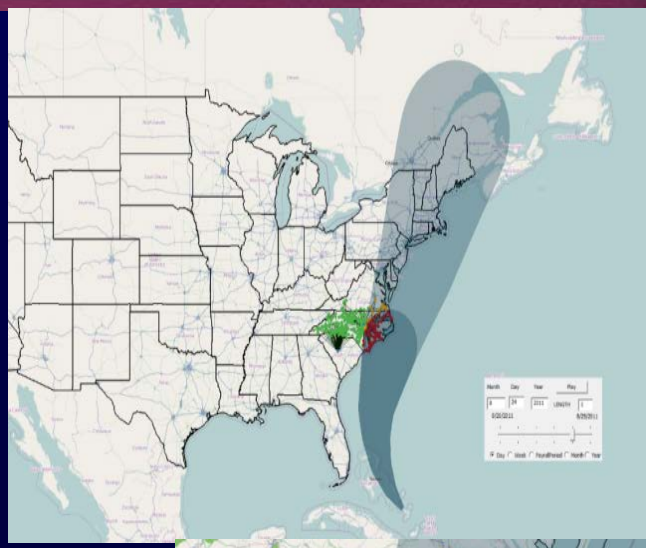
VACCINE

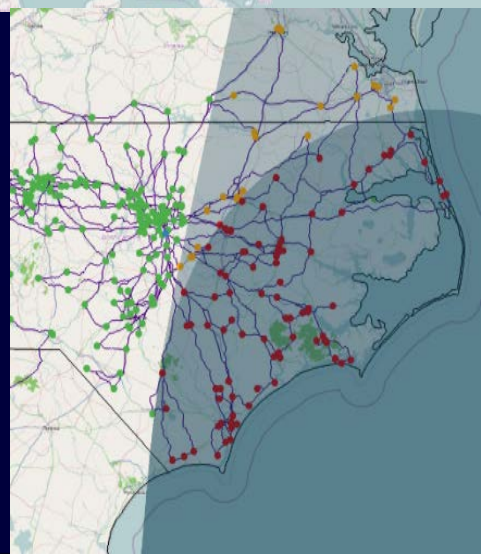
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  Combine interactive VA with complex simulation models



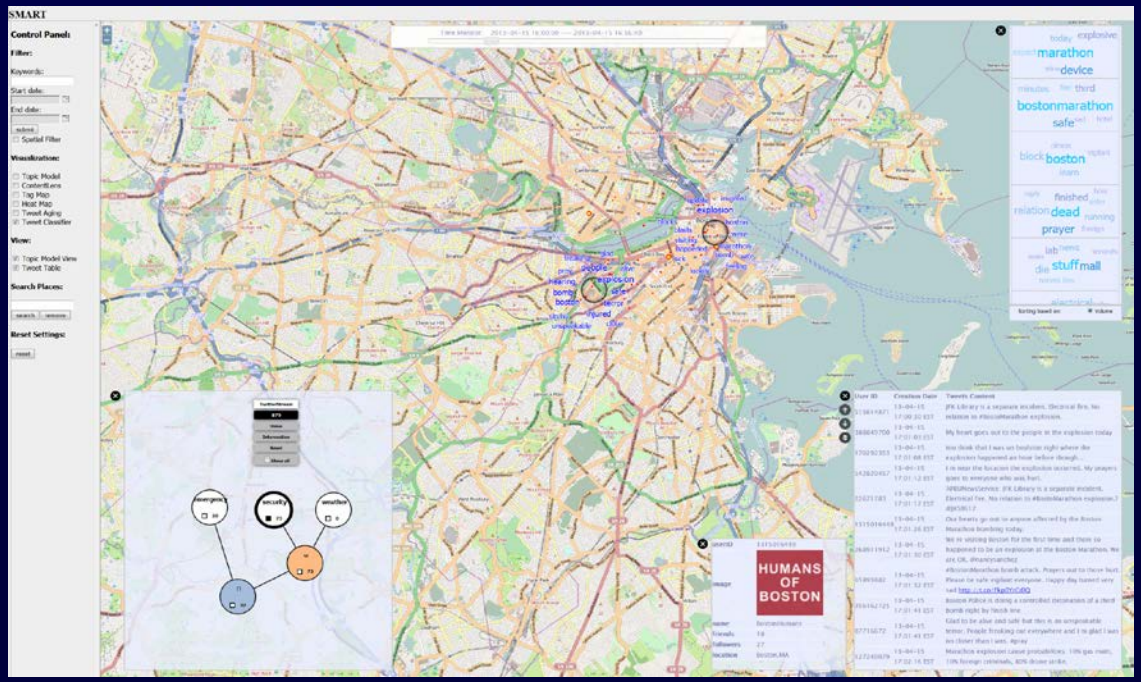


VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media Analytics and Reporting Tool



David S. Ebert, Purdue University

ebertd@purdue.edu

October 2014

Why Use Social Media?

Marysville-Pilchuck High School (WA) Shooting

Time: around 10:40 am PST
(01:40 pm EST), 10/24/2014

Earliest Tweets About The Event

601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances
461749842	14-10-24 13:55:45 EST	Cops are fucking EVERYWHERE right now
461749842	14-10-24 13:55:21 EST	Cops are fucking EVWRYWHERE right now
707981743	14-10-24 13:55:12 EST	There is a shooting at mp?
818397241	14-10-24 13:54:06 EST	@Miguel_Rocha34 do you have any idea who the person with the gun was
818397241	14-10-24 13:52:12 EST	you got some serious issues that you need to seek help for if you shoot up a fucking high school
601648236	14-10-24 13:49:11 EST	@Ray_Raaaay21 one of the kids in the lunchroom started shooting and everyone booked it
601648236	14-10-24 13:46:17 EST	shooting at mp

707981743	14-10-24 14:17:43 EST	2 people are dead...
335778696	14-10-24 14:17:26 EST	Wtf is going on?? Shooting at mp?
70255303	14-10-24 14:16:37 EST	Shooting at Marysville Pilchuck. Praying for the kids and the families.
1853421522	14-10-24 14:14:39 EST	@Lake_stonerrr a student shot several students, possible 2 students dead 6 injured
1099441382	14-10-24 14:10:38 EST	Shooting at pikchuck???
599963823	14-10-24 14:08:26 EST	Something's fucking wrong !!!!! Or my mom wouldn't be crying about the MP shooting !!!
130018947	14-10-24 14:07:51 EST	About 5 mins away from the scene. Sad. Another school shooting. #MPHS
601648236	14-10-24 14:01:49 EST	@_FatSackMatt cops are blocking all the entrances

Some online news feed:
http://www.foxnews.com/us/2014/10/24/police-respond-to-shooting-at-washington-state-high-school/
http://www.cnn.com/2014/10/24/us/washington-school-shooting/index.html

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

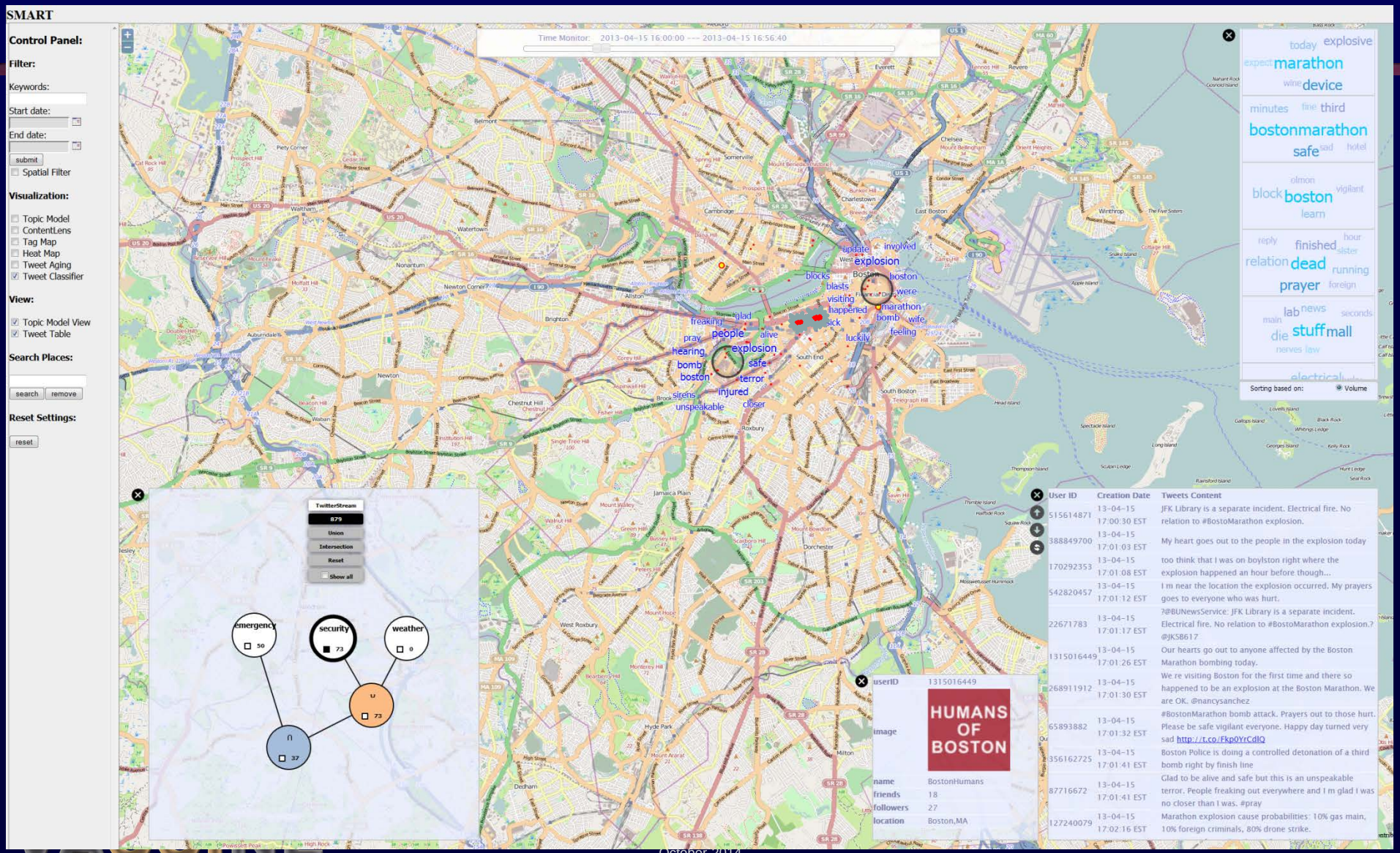
IMPACTS:

- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games Fall 2014
- Indianapolis Metropolitan Police for special events
- US CBP for investigations Fall 2014 (Boston, AMOC)

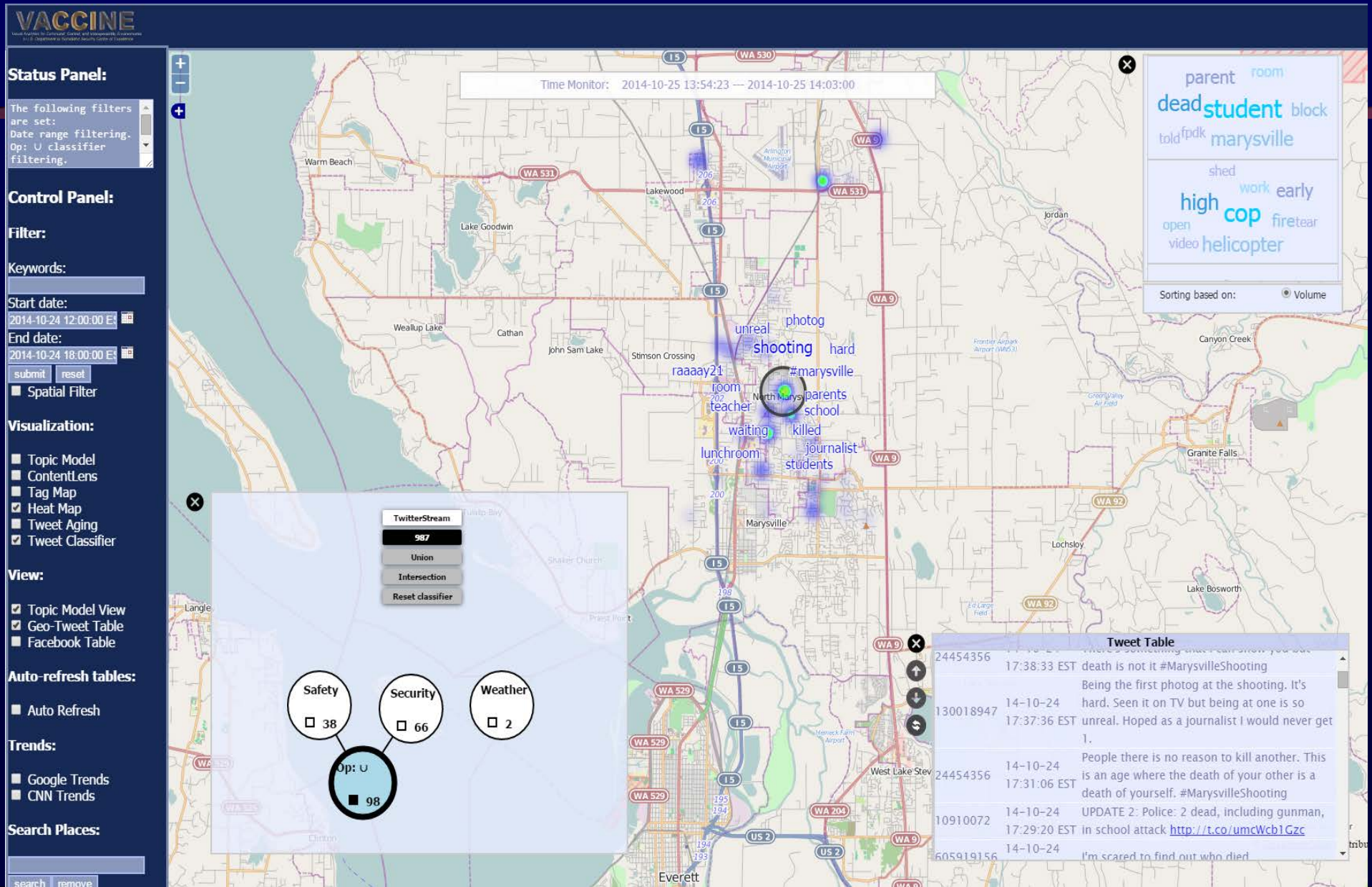


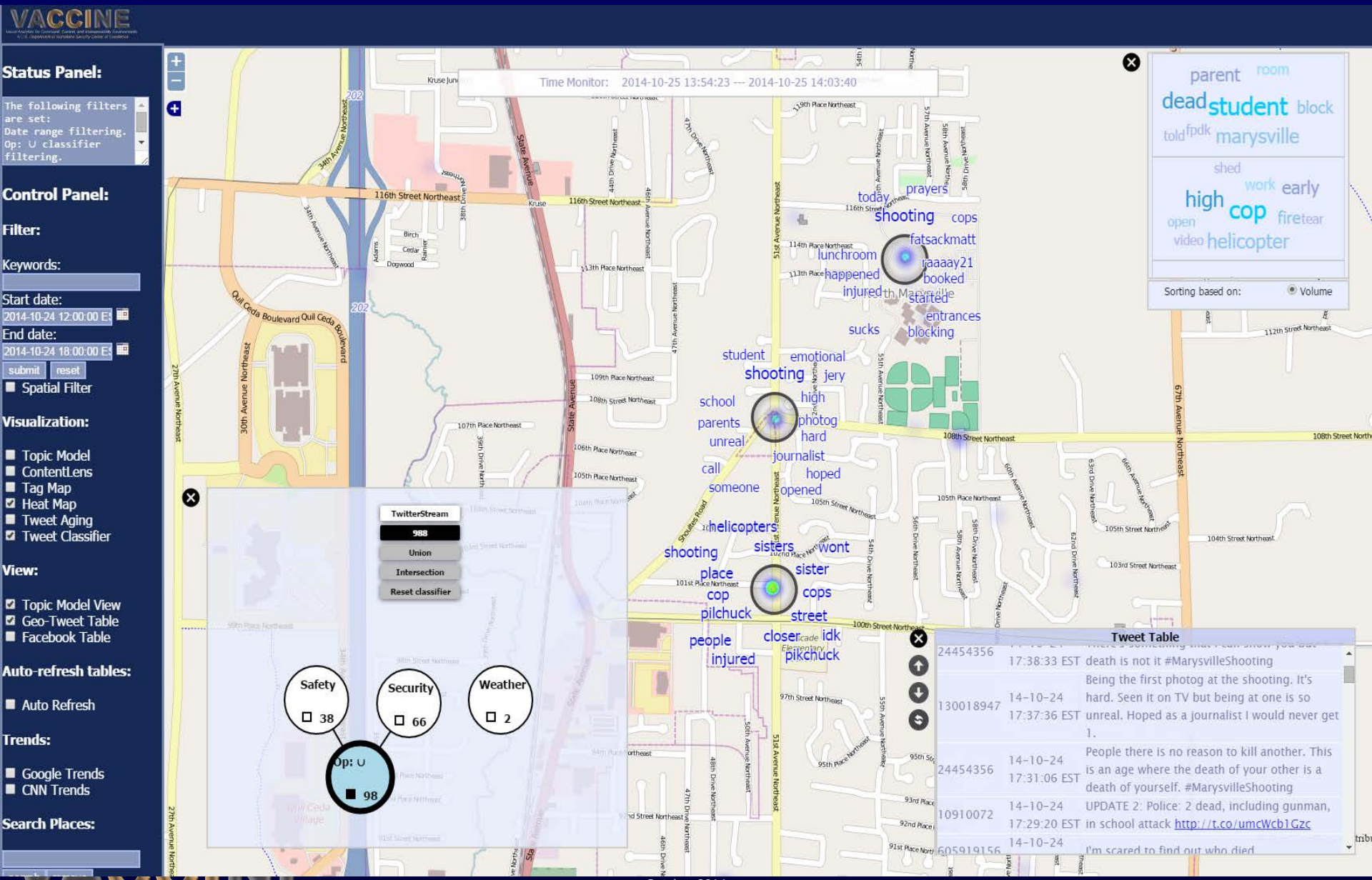
VACCINE

Social Media Analysis and Reporting Toolkit (SMART)



Marysville-Pilchuck High School (WA) Shooting




VACCINE
Status Panel:
The following filters are set:
Date range filtering.
Op: ∪ classifier filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 E
End date:
2014-10-24 18:00:00 E
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:03:40
parent
room
dead
student
block
told
fpdk
marysville
shed
work
early
high
cop
firetear
open
video
helicopter
Sorting based on:
Volume
today
prayers
shooting
cops
fatsackmatt
lunchroom
iaaaay21
happened
booked
injured
started
entrances
sucks
blocking
student
emotional
shooting
jery
school
high
parents
photog
unreal
hard
journalist
call
hoped
someone
opened
helicopters
shooting
sisters
wont
place
sister
cop
cops
pilchuck
street
people
closer
idk
injured
pikchuck
TwitterStream
988
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
Op: ∪
98
Tweet Table
24454356
17:38:33 EST
death is not it #MarysvilleShooting
130018947
14-10-24
17:37:36 EST
Being the first photog at the shooting. It's hard. Seen it on TV but being at one is so unreal. Hoped as a journalist I would never get 1.
24454356
14-10-24
17:31:06 EST
People there is no reason to kill another. This is an age where the death of your other is a death of yourself. #MarysvilleShooting
10910072
14-10-24
17:29:20 EST
UPDATE 2: Police: 2 dead, including gunman, in school attack http://t.co/umcWcb1Gzc
605919156
14-10-24
I'm scared to find out who died


Joshua Trujillo
@joshtrujillo
Follow
Scene at parent reunification center after shooting at #MPHS
Marysville, WA
RETWEETS
20
FAVORITES
7
1:47 PM - 24 Oct 2014

Status Panel:
The following filters are set:
Single tweet filtering.
Date range filtering.
Control Panel:
Filter:
Keywords:
Start date:
2014-10-24 12:00:00 ES
End date:
2014-10-24 18:00:00 ES
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
Marysville-Pilchuck High Sch
search
remove
Time Monitor: 2014-10-25 13:54:23 --- 2014-10-25 14:07:48
pilchuck
tear
video
cop
pray
homestuck
job
couple
marysville
high
school
sad
block
kid
told
fire
shoot
Sorting based on:
Volume
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO
Location of the school
North Marysville
TwitterStream
990
Union
Intersection
Reset classifier
Safety
38
Security
66
Weather
2
userID
23845798
image
name
joshtrujillo
friends
2146
followers
6674
location
Seattle,WA
Tweet Table
User ID
Creation Date
Tweets Content
23845798
14-10-24 16:47:14 EST
Scene at parent reunification center after shooting at #MPHS http://t.co/ApYpohRgvO

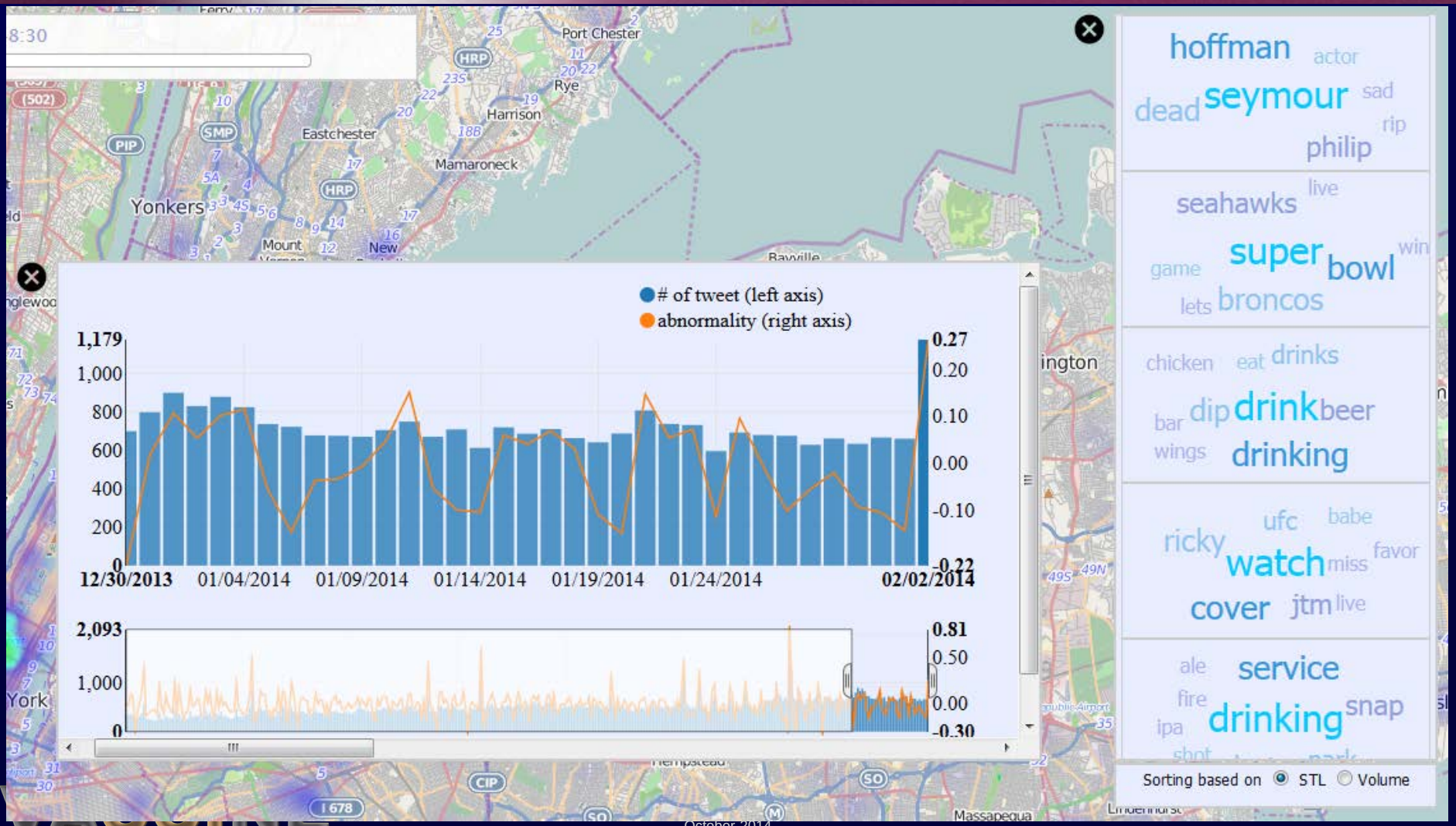
Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Boston Marathon
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring event
  - Purdue football games – the p
- During unexpected eve
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – L



VACCINE

Anomaly Detection Superbowl Weekend



Examples From

Fleet Wee

VACCINE

End date:

submit reset

Spatial Filter

Visualization:

Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier

View:

Topic Model View
Geo-Tweet Table
Facebook Table

Auto-refresh tables:

Auto Refresh

Trends:

Google Trends
CNN Trends

Search Places:

search remove

Reset Settings:







Tiffany Wilson
@TWilsonTV

Follow

#FleetWeekSF crowds have Embarcadero at a crawl. I think pedestrians are moving faster than cars

San Francisco, CA

Reply Retweet Favorite More

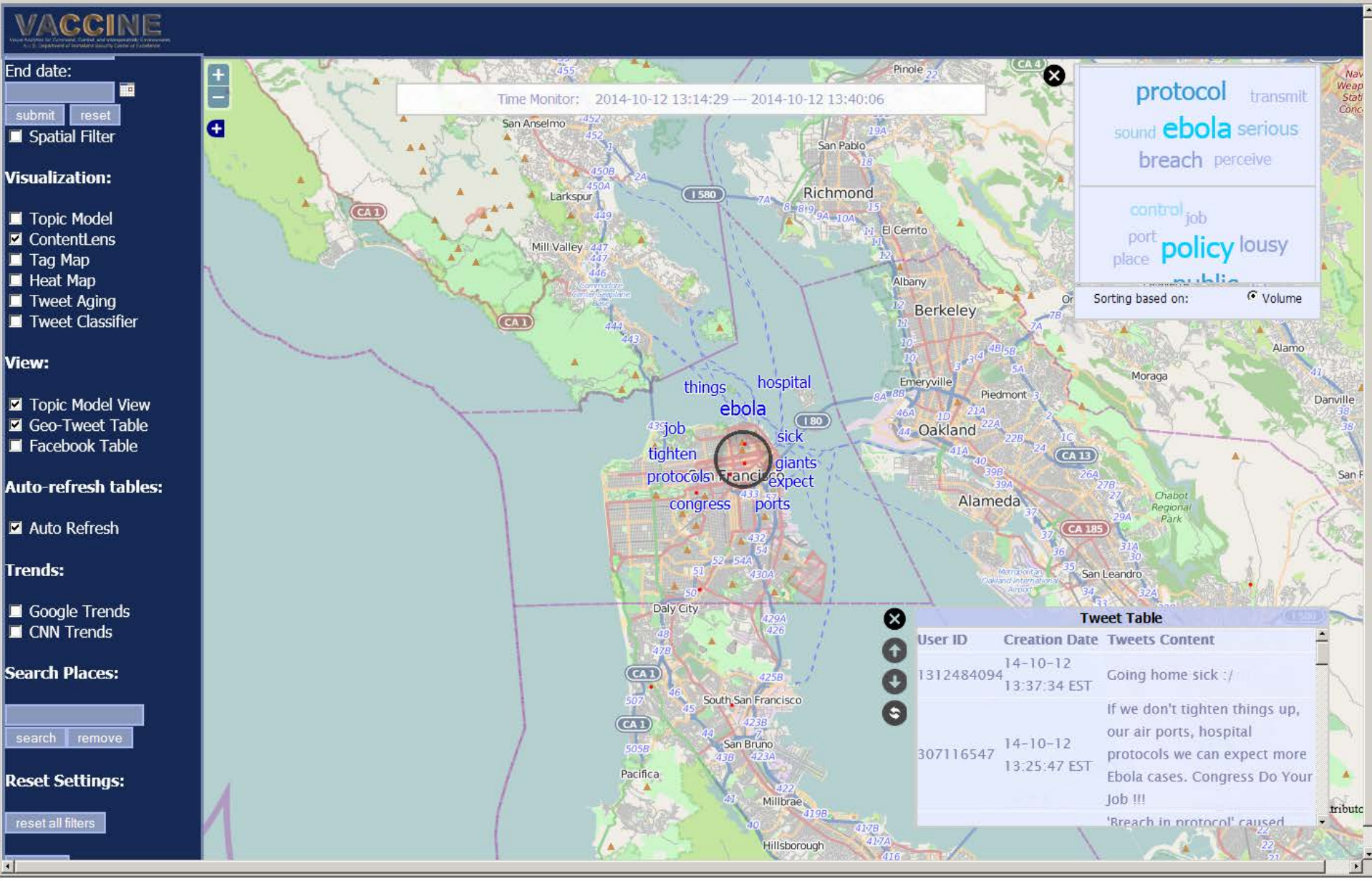


RETWEETS 6

FAVORITES 4

12:56 PM - 12 Oct 2014

Ebola tweets during Fleet Week 10/12/14

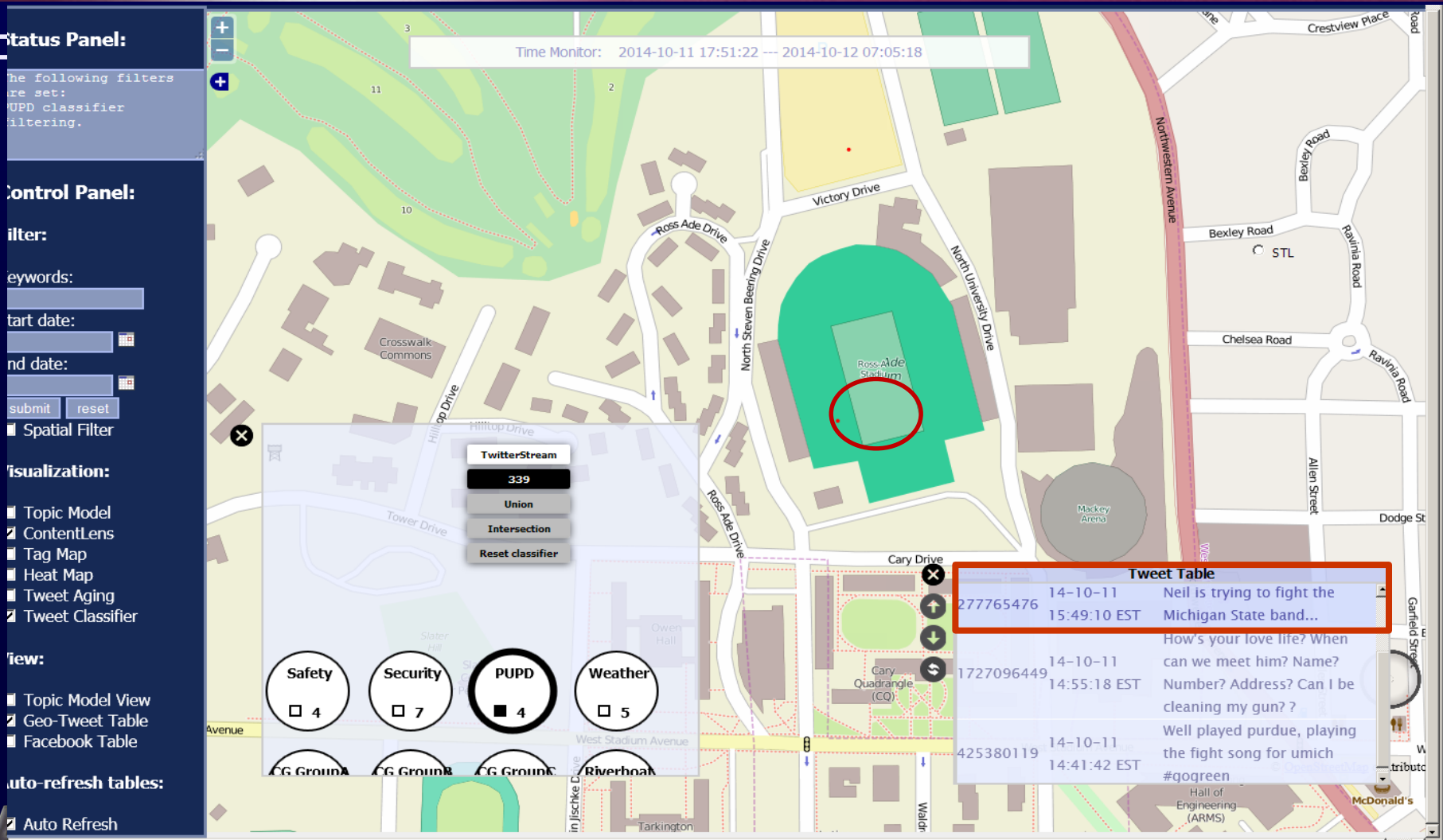


Where Can It Be Helpful?

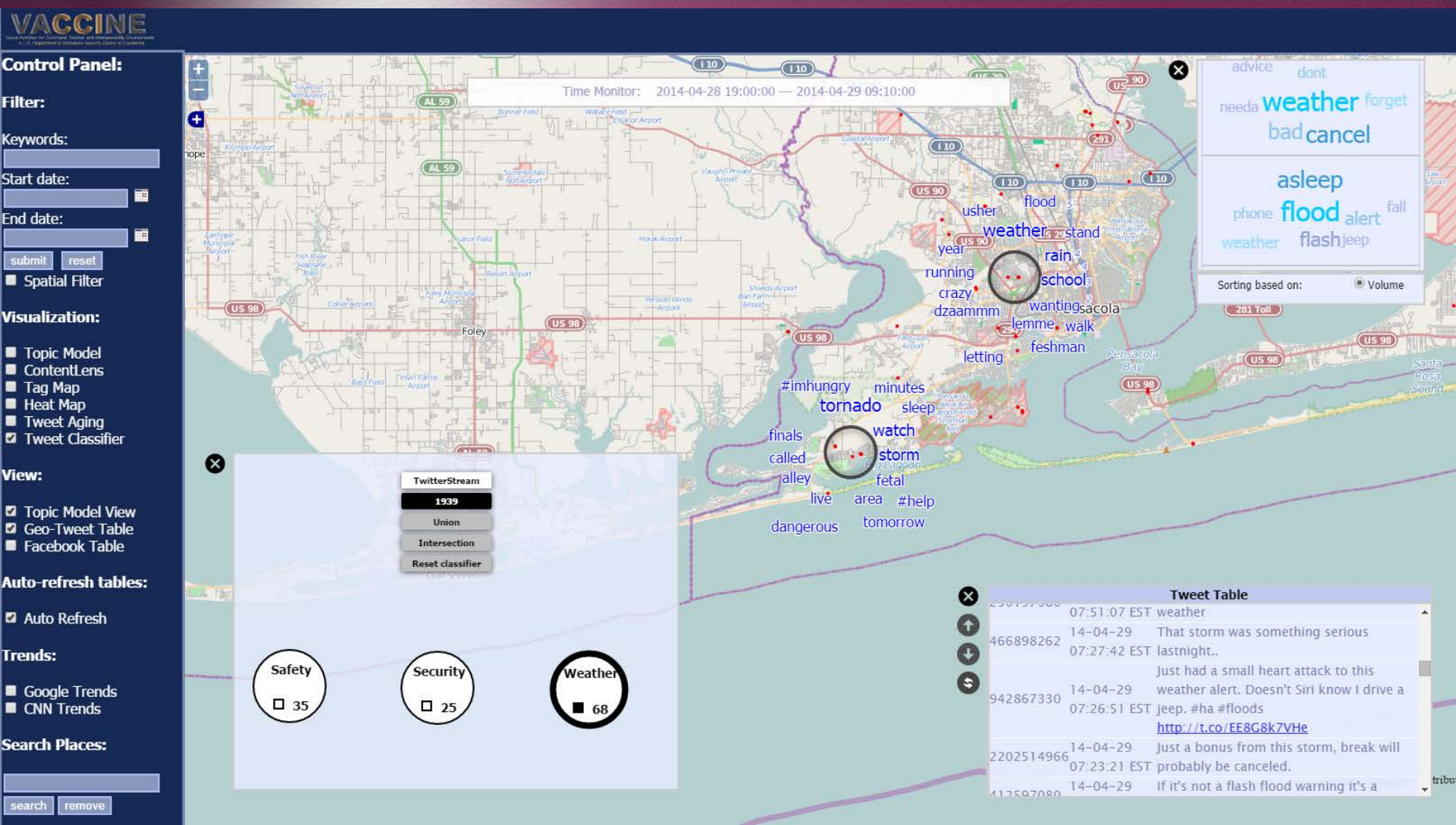
- During large-scale planned events, e.g.,
  - Boston Marathon bombing
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



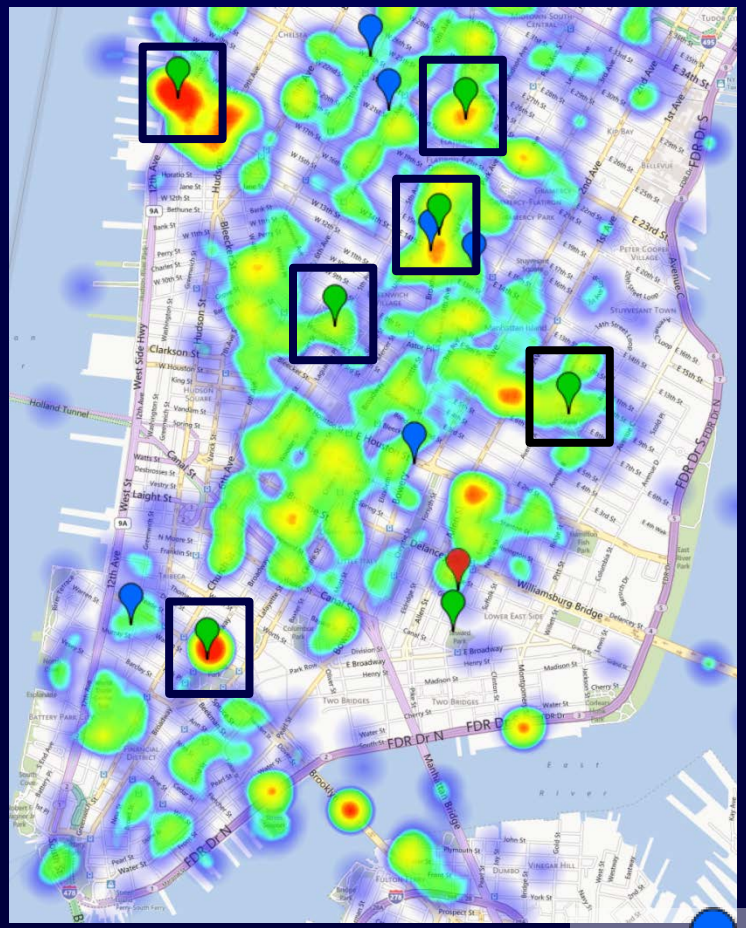
Where Can It Be Helpful?

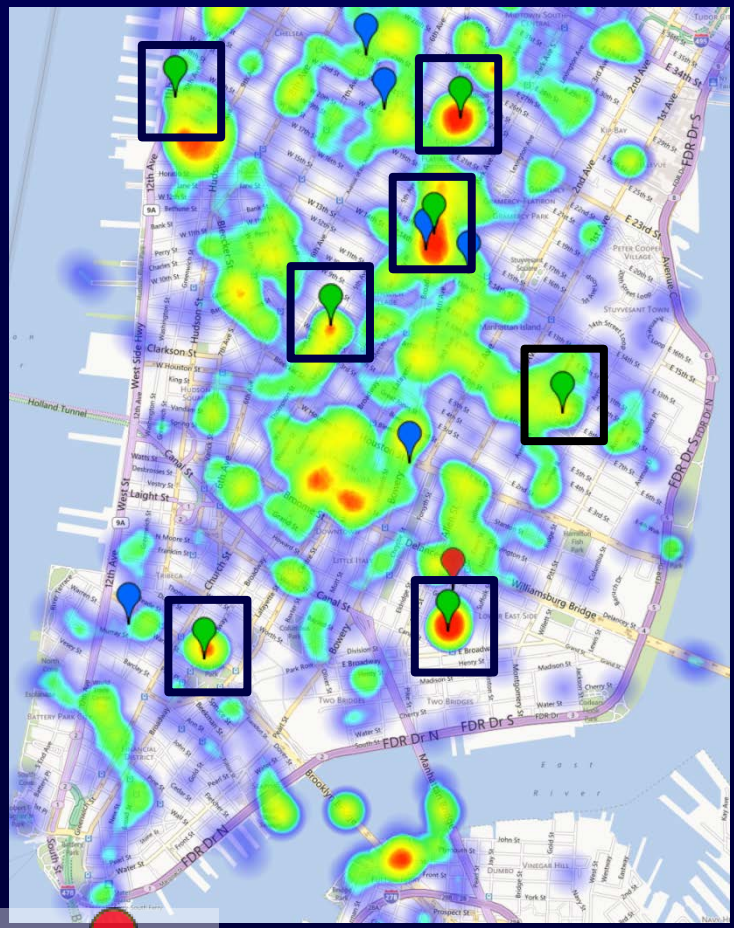


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

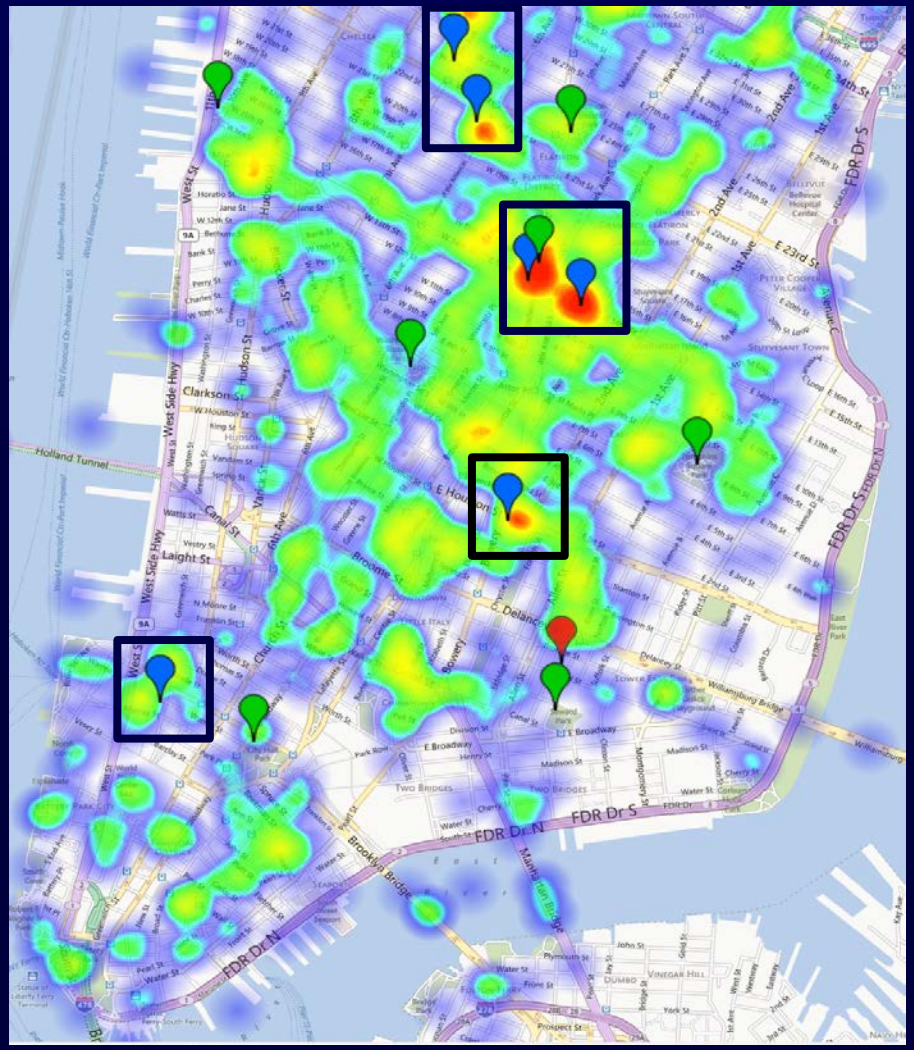
VACCINE

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

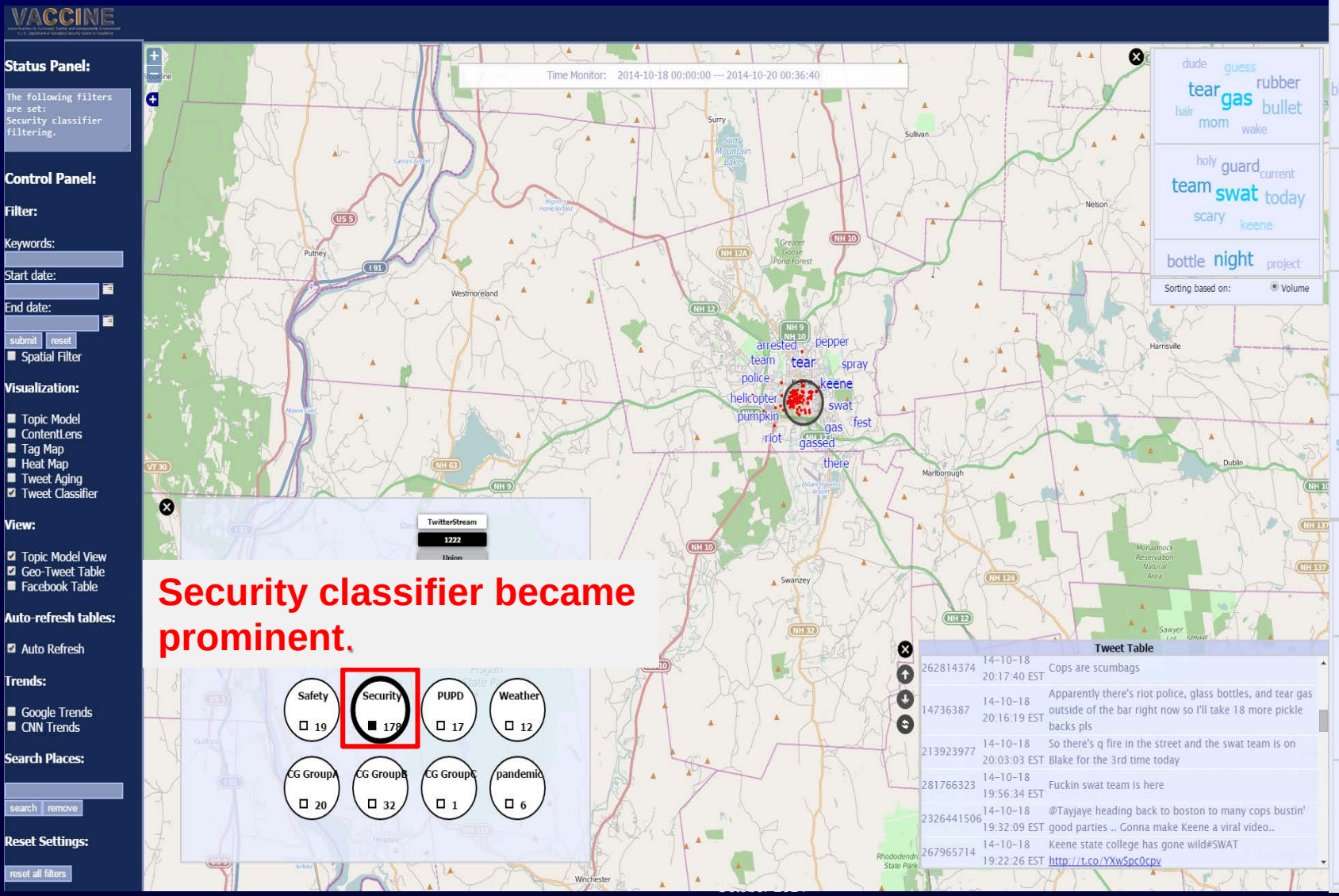
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



VACCINE

Keene, NH Pumpkin festival 10/18/14 – 10/19/14 Overview

Major topics:

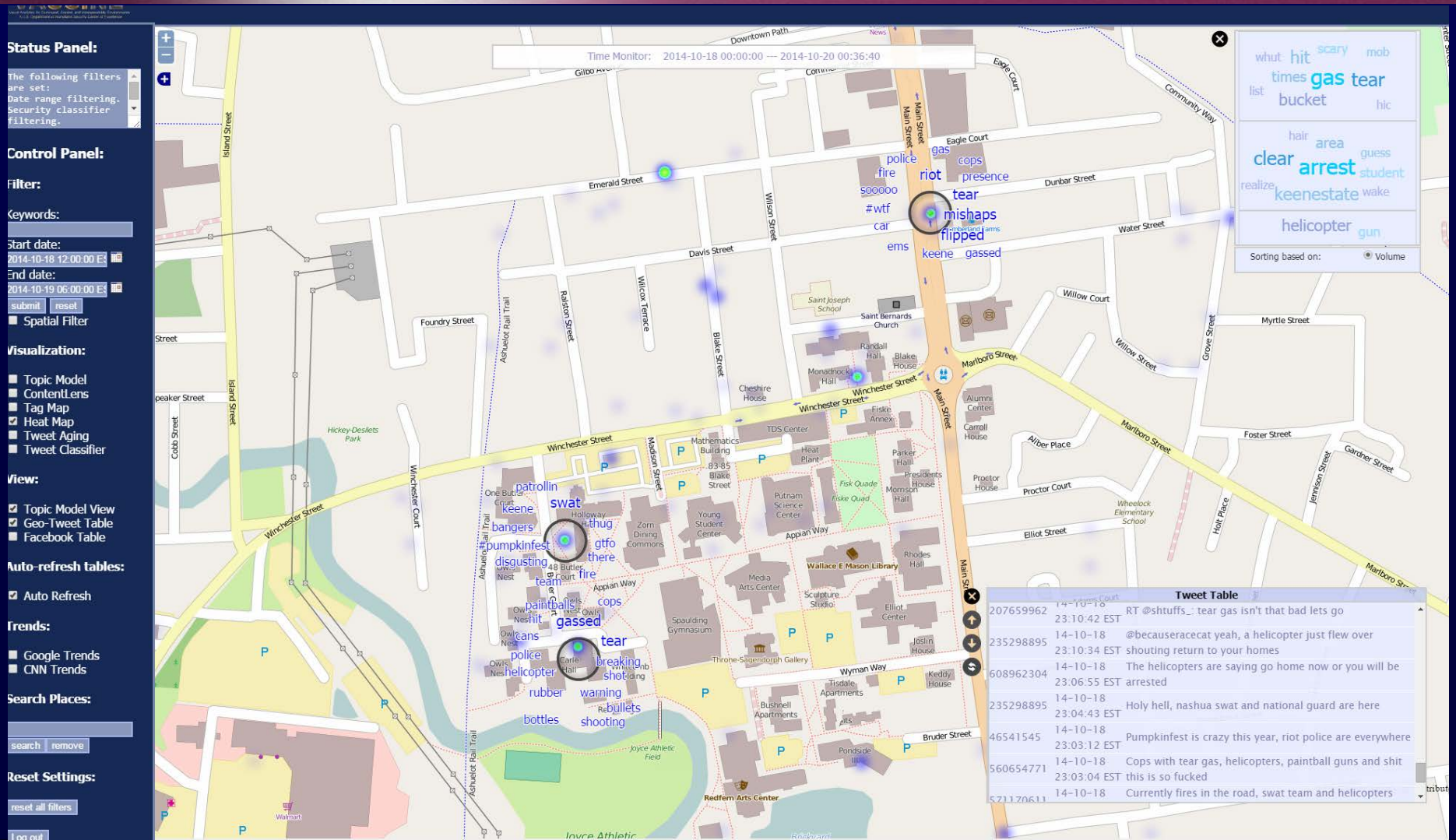


Security classifier became prominent.



Keene, NH

Time range:10/18 12:00pm – 10/19 06:00am, Security classifier)

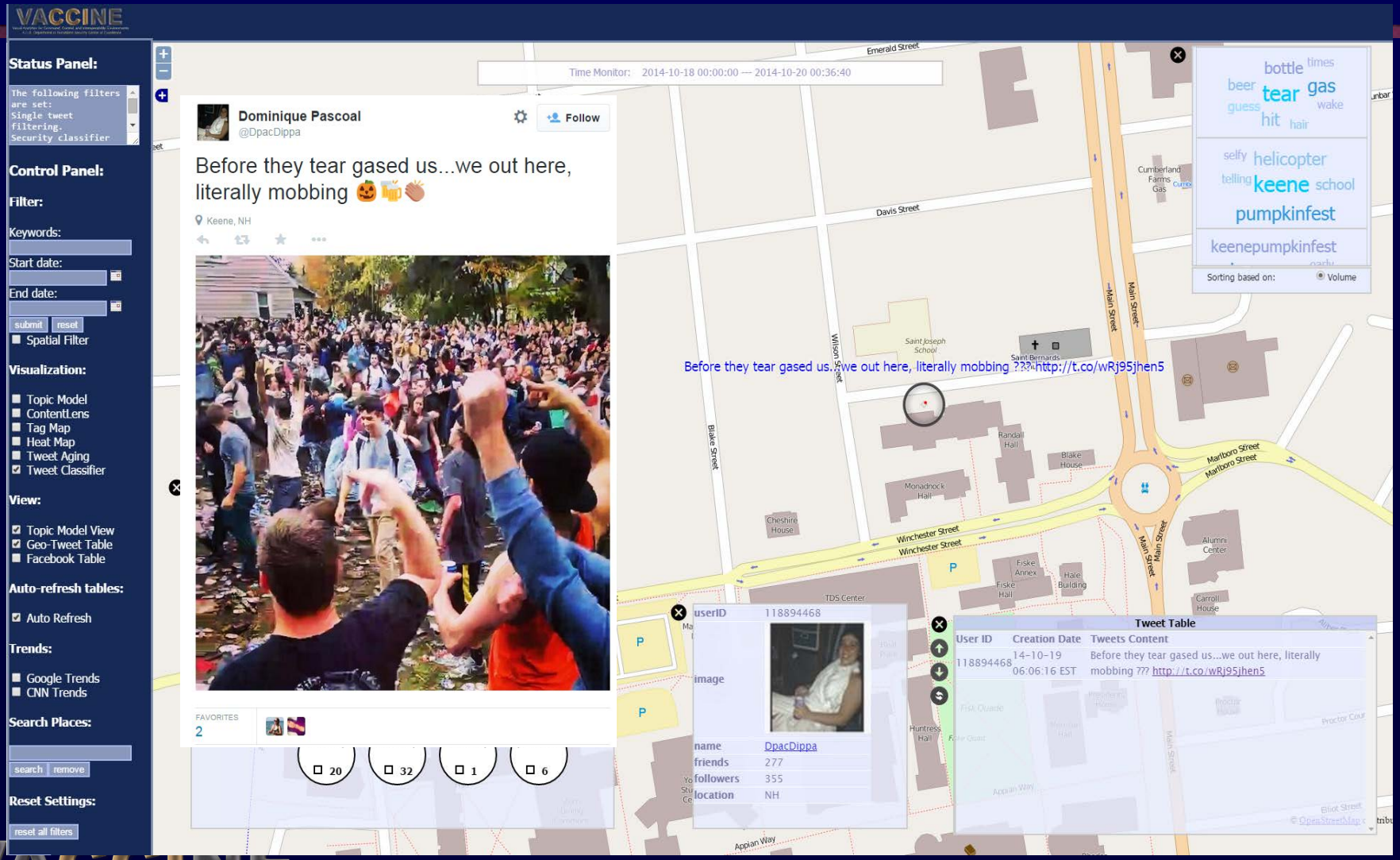


Keene, NH



The picture was later used by ABC news(link).

Keene, NH

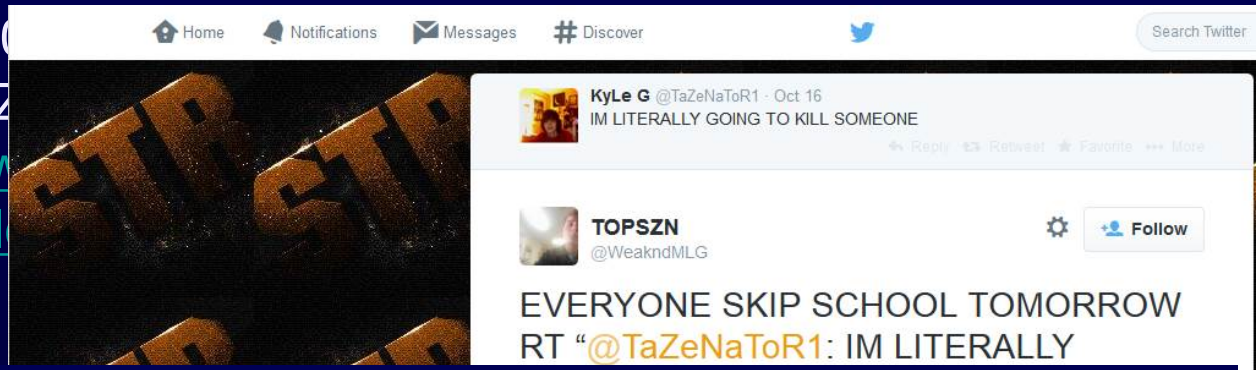

VACCINE
Status Panel:
The following filters are set:
Single tweet filtering.
Security classifier
Control Panel:
Filter:
Keywords:
Start date:
End date:
submit
reset
Spatial Filter
Visualization:
Topic Model
ContentLens
Tag Map
Heat Map
Tweet Aging
Tweet Classifier
View:
Topic Model View
Geo-Tweet Table
Facebook Table
Auto-refresh tables:
Auto Refresh
Trends:
Google Trends
CNN Trends
Search Places:
search
remove
Reset Settings:
reset all filters
Time Monitor: 2014-10-18 00:00:00 --- 2014-10-20 00:36:40
Dominique Pascoal
@DpacDippa
Follow
Before they tear gased us...we out here, literally mobbing
Keene, NH
FAVORITES
2
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
bottle
times
beer
tear
gas
guess
wake
hit
hair
selfy
helicopter
telling
keene
school
pumpkinfest
keenepumpkinfest
Sorting based on:
Volume
userID 118894468
image
name DpacDippa
friends 277
followers 355
location NH
Tweet Table
User ID
Creation Date
Tweets Content
118894468
14-10-19 06:06:16 EST
Before they tear gased us...we out here, literally mobbing ??? http://t.co/wRj95jhen5
Emerald Street
Davis Street
Main Street
Wilson Street
Blake Street
Winchester Street
Marlboro Street
Saint Joseph School
Saint Bernards
Randall Hall
Blake House
Monadnock Hall
Cheshire House
Fiske Annex
Fiske Hall
Hale Building
Alumni Center
Carroll House
TDS Center
Huntress Hall
Cumberland Farms Gas
Appian Way
20
32
1
6

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Louisville, KY Oct 16 – School Shooting Threat

- WeakndMLG 10-16 23:(TOMORROW RT "@TaZ SOMEONE" http://www https://www.googl



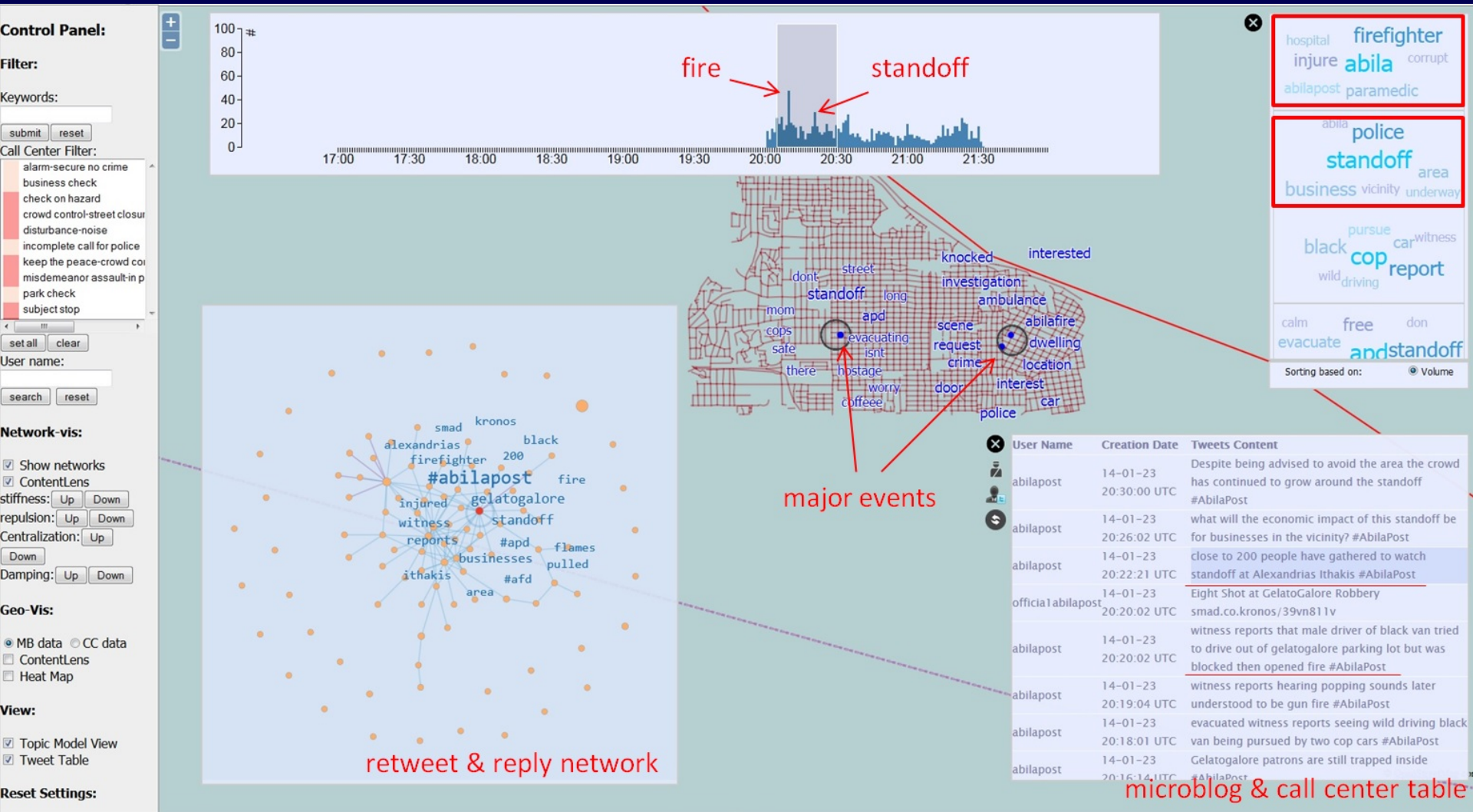
Follow Up Tweets

- WeakndMLG 10-17 16:46:20 EDT Now the coast guard will be monitoring my tweets? http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125221,-85.817108
- WeakndMLG 10-17 17:08:07 EDT @TommyTheGhoster they said the coast guard intelligence unit picked up on my tweets or something they had a Manila folder with print outs http://www.twitter.com/WeakndMLG https://www.google.com/maps/place/38.125206,-85.817072



VACCINE

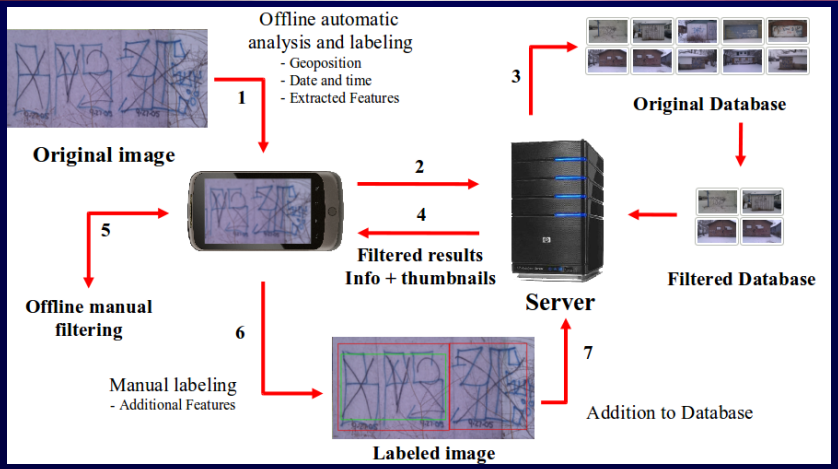
SMART: Network Linkage



Conclusions

- Powerful tool for daily use and during special events
- Realtime monitoring and email alerts
- More features than many tools on the market
- How to get SMART?
- Visit: www.vaccinetools.org
- Contact us: vaccine@purdue.edu

Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

For Further Information

www.VisualAnalytics-CCI.org



CVADA

Center for Visual Analytics and Data Analytics

A U.S. Department of Homeland Security Center of Excellence

Visual Analytics Solutions for The Real World

David Ebert

Purdue University

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

July 2014

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions



- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE

Some of Our Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Citizenship and Immigration Service
  - US Federal Emergency Management Agency
  - US Customs and Border Protection
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies (IN, IL, OH, PA, NC, NY)
- State Fusion Centers
  - Ohio (SAIC), Indiana (IIFC)
- Companies
  - Motorola, Kimberly Clark, Banfield, Bank of America





VACCINE: What We Do – Visual Analytic Solutions

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Why VA For Big Data?

(Or What Big Data Analytics Can't Do)

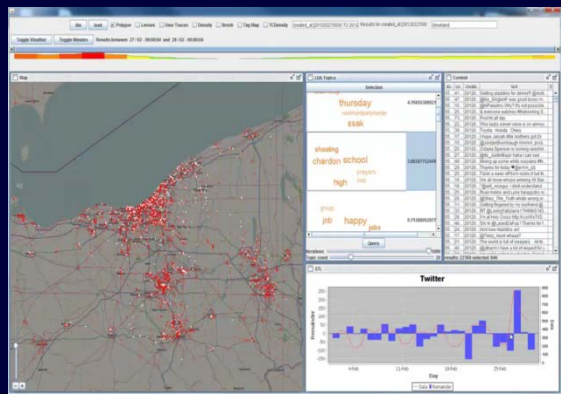
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

- Qualitative, fuzzy, and social data
  - Preferences, significance of one relationship over another; Trust
- Context
  - Data is rarely complete, nor incorporates all relevant information Humans have extensive information and experience that never make it into the collected data
- Spurious vs. Significant
  - Big data means more statistically significant events and correlations, but they may not have any relevance
  - Increases noise to signal ratio
- Big problems
  - Complex, multifaceted, multiparameter big challenges with unquantified dependencies

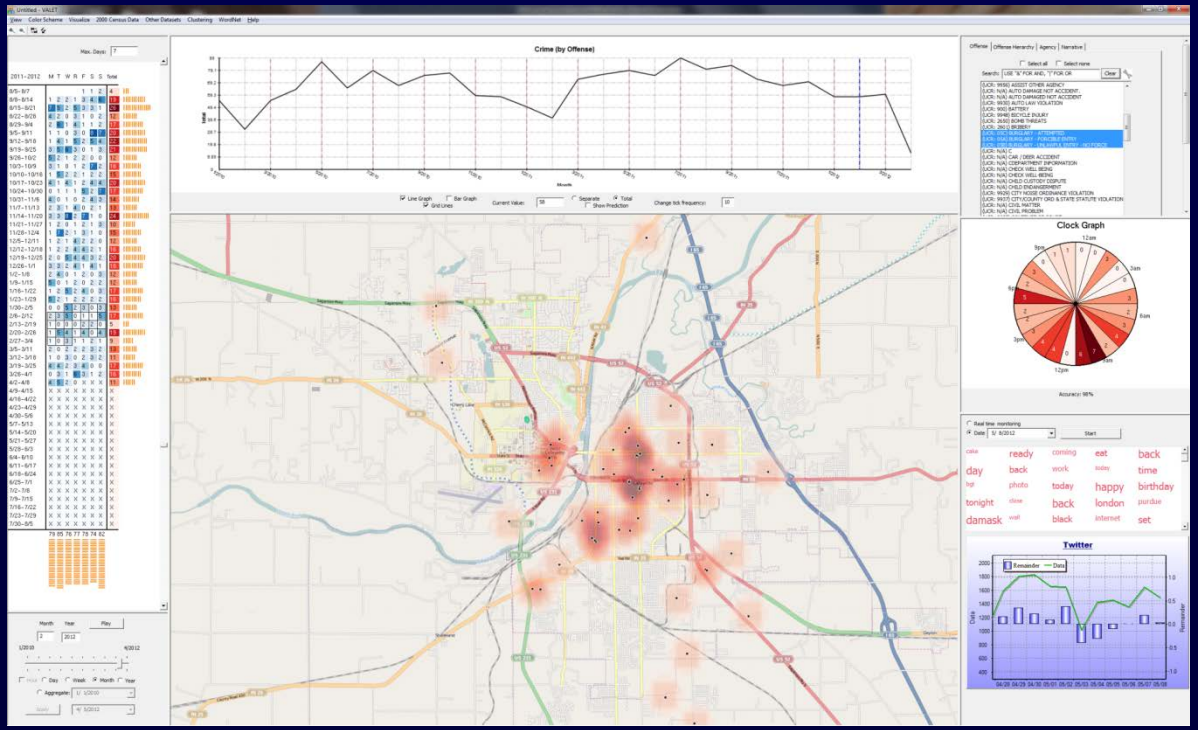
Visual Analytics Uses for Public Safety

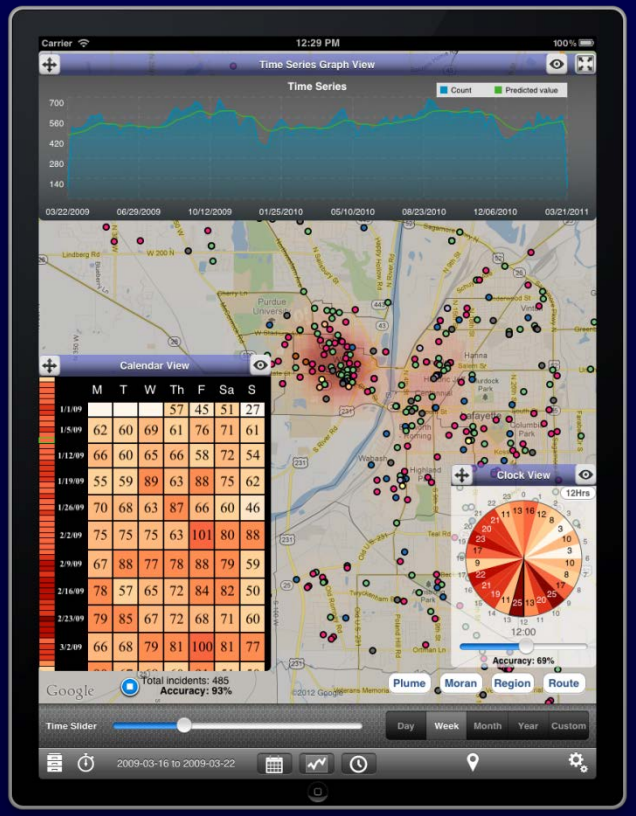


- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Data-driven policy analysis
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation

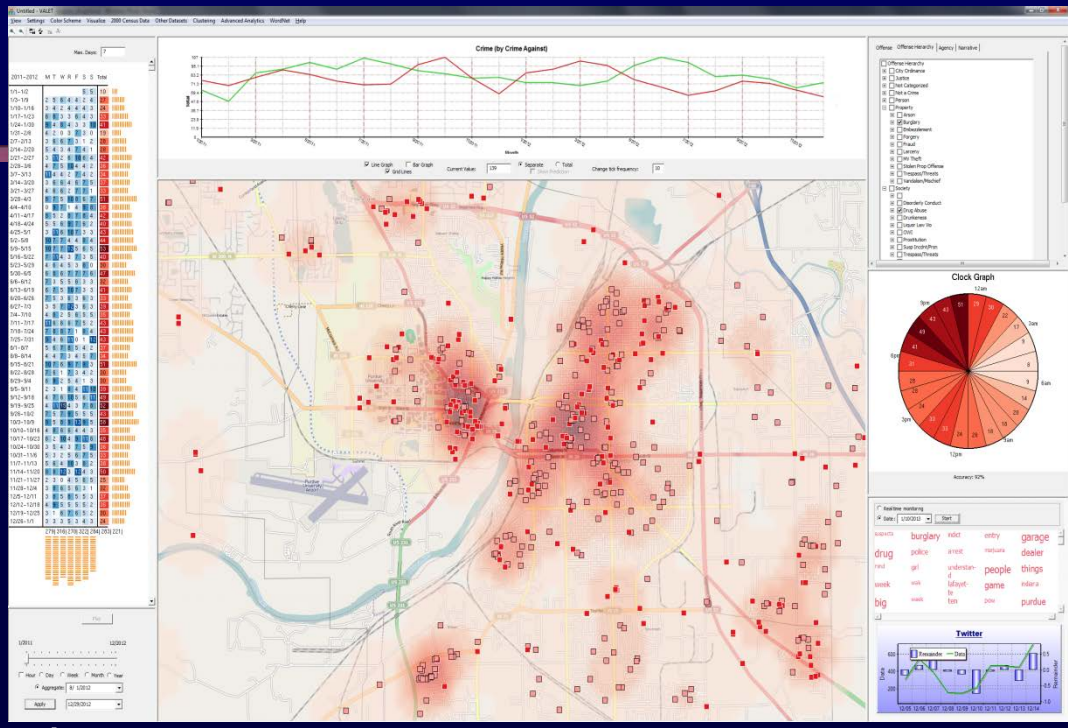


Visual Analytics Law Enforcement Toolkit (i)VALET





Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

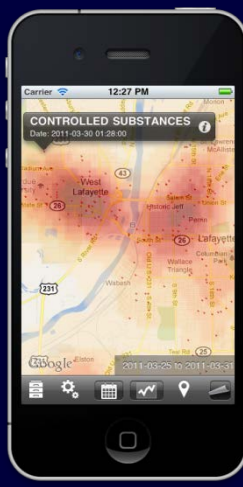
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:

- Spring 2011: WL, Lafayette Police

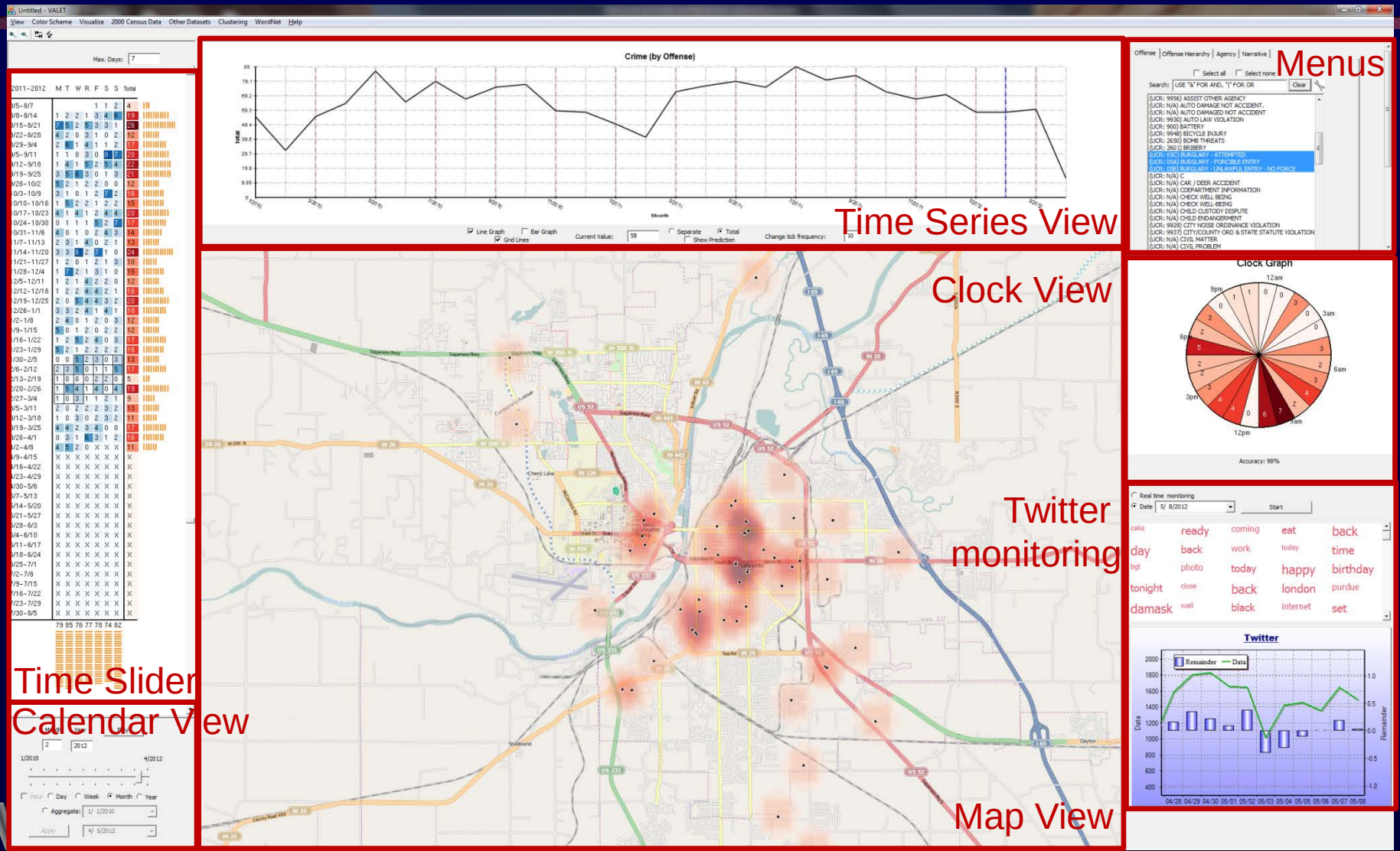
iVALET delivered:

- October 2011: Purdue, WL Police

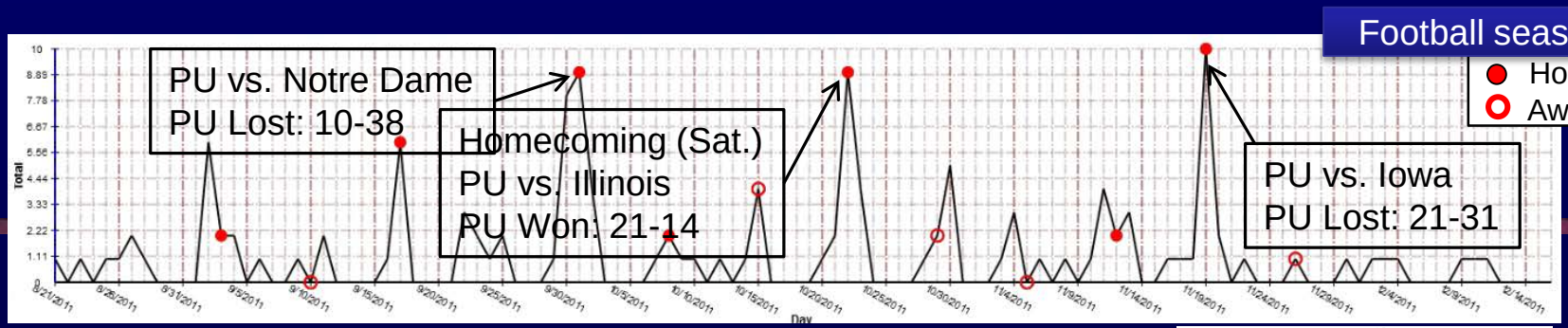


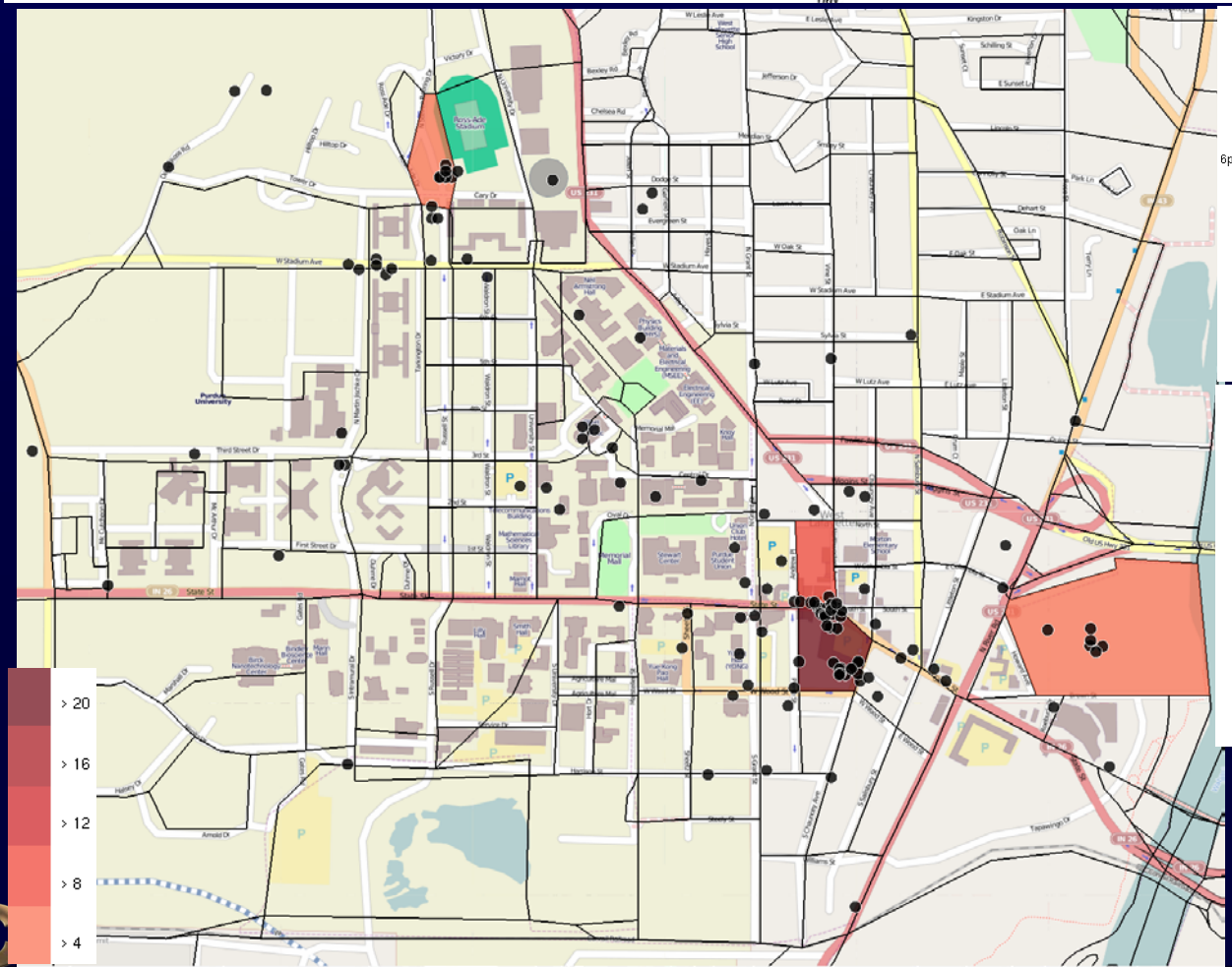
VACCINE

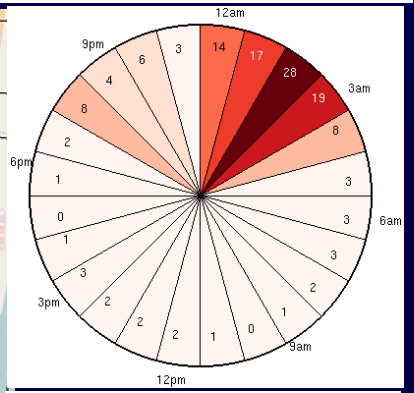
Example for Crime Analytics: VALET

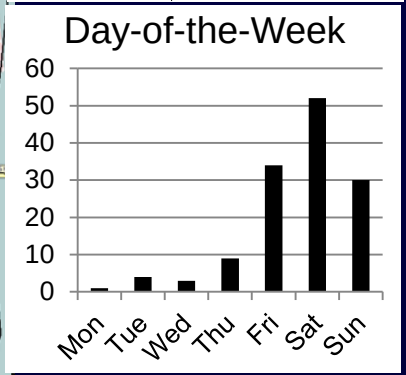


Example: Drunkenness / Public Intoxication









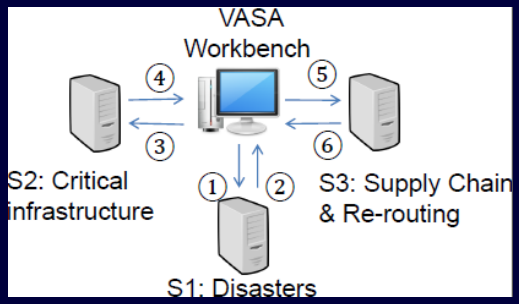
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Texas (TACC), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s): Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

Purdue

UNCC



U Minn

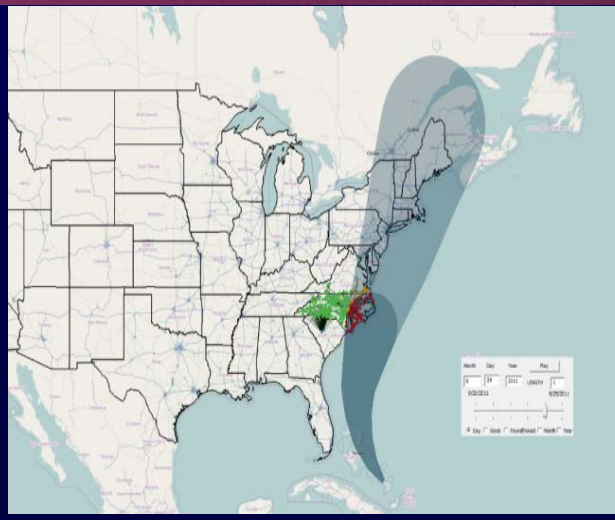
TACC

VACCINE

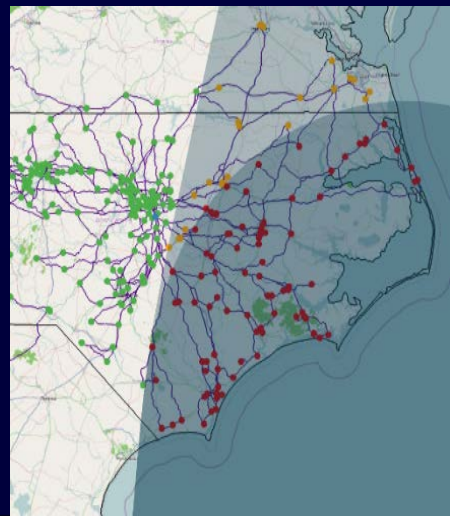
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

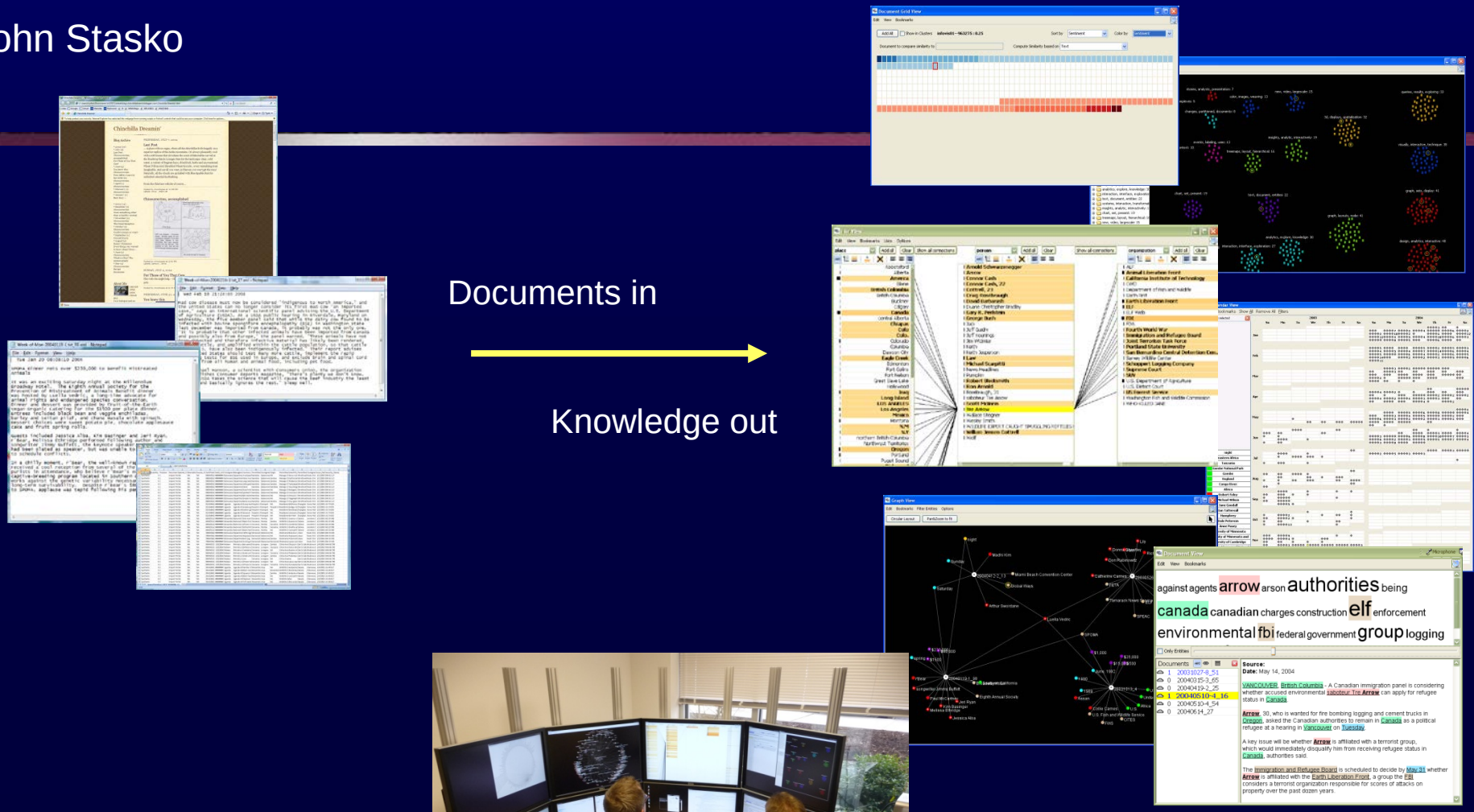


- Support decision-making for extreme weather and natural disaster scenarios
  - Combine real and simulation data
  - Allow “what-if” exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**:
  combine interactive VA with complex simulation models

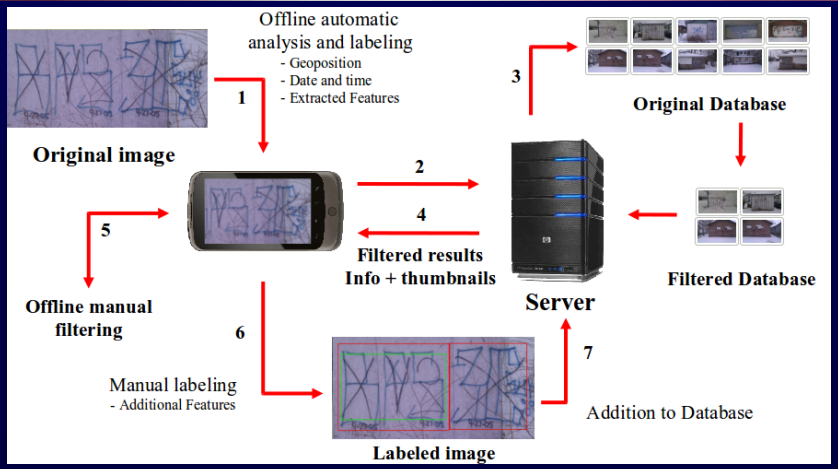


Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



Gang Graffiti Analysis and Recognition Using a Mobile Telephone (GARI)





IMPACT:

- Allows police to catalog and analyze gang graffiti images into a database system to better track and determine gang activity throughout a region
- Will allow the graffiti images to be "interpreted"
- More than 140 users and 1300 graffiti images acquired

GARI delivered:

- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

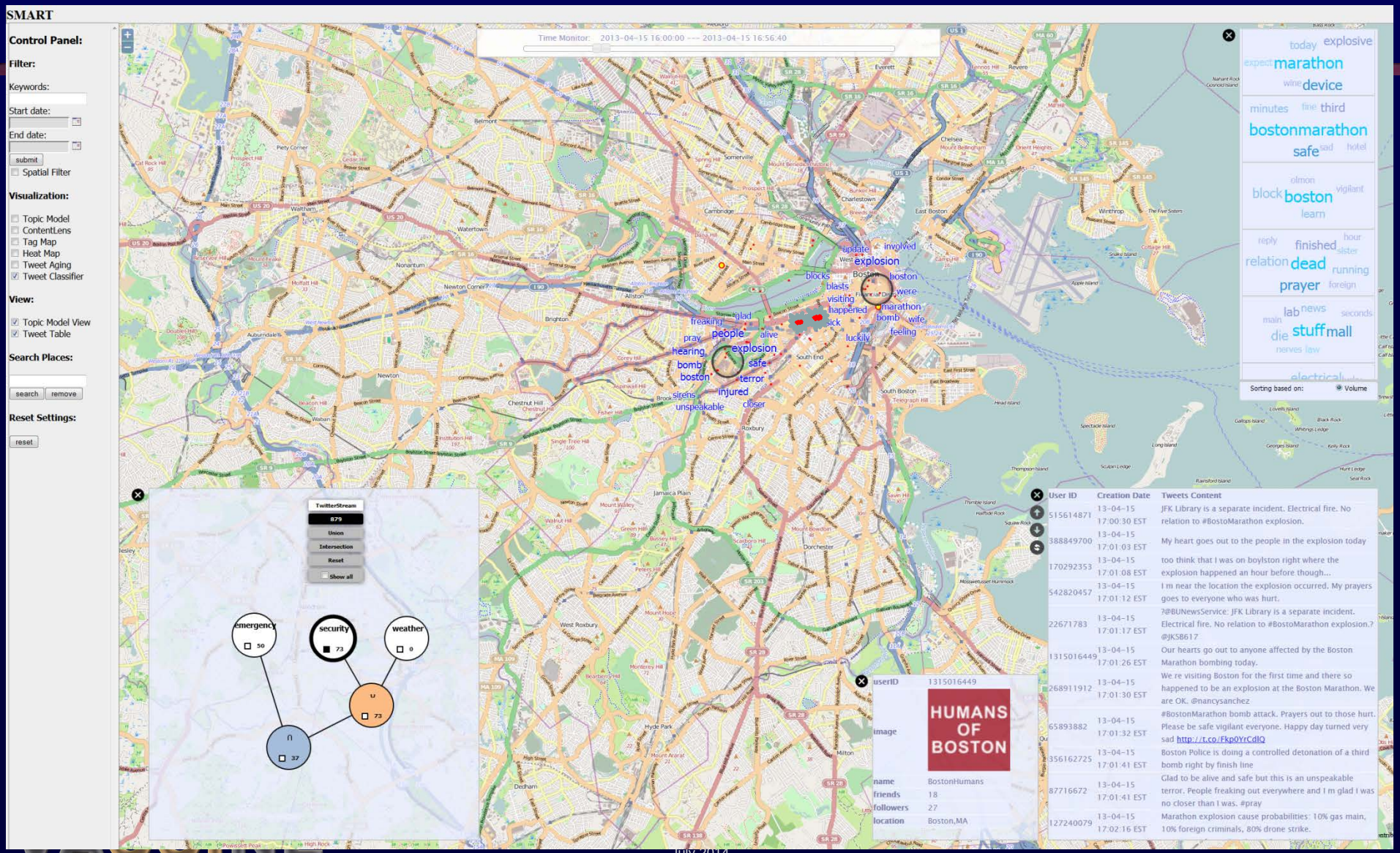
Social Media: Real-time Twitter Visual Analytics

(Purdue, Stuttgart, Penn St.)

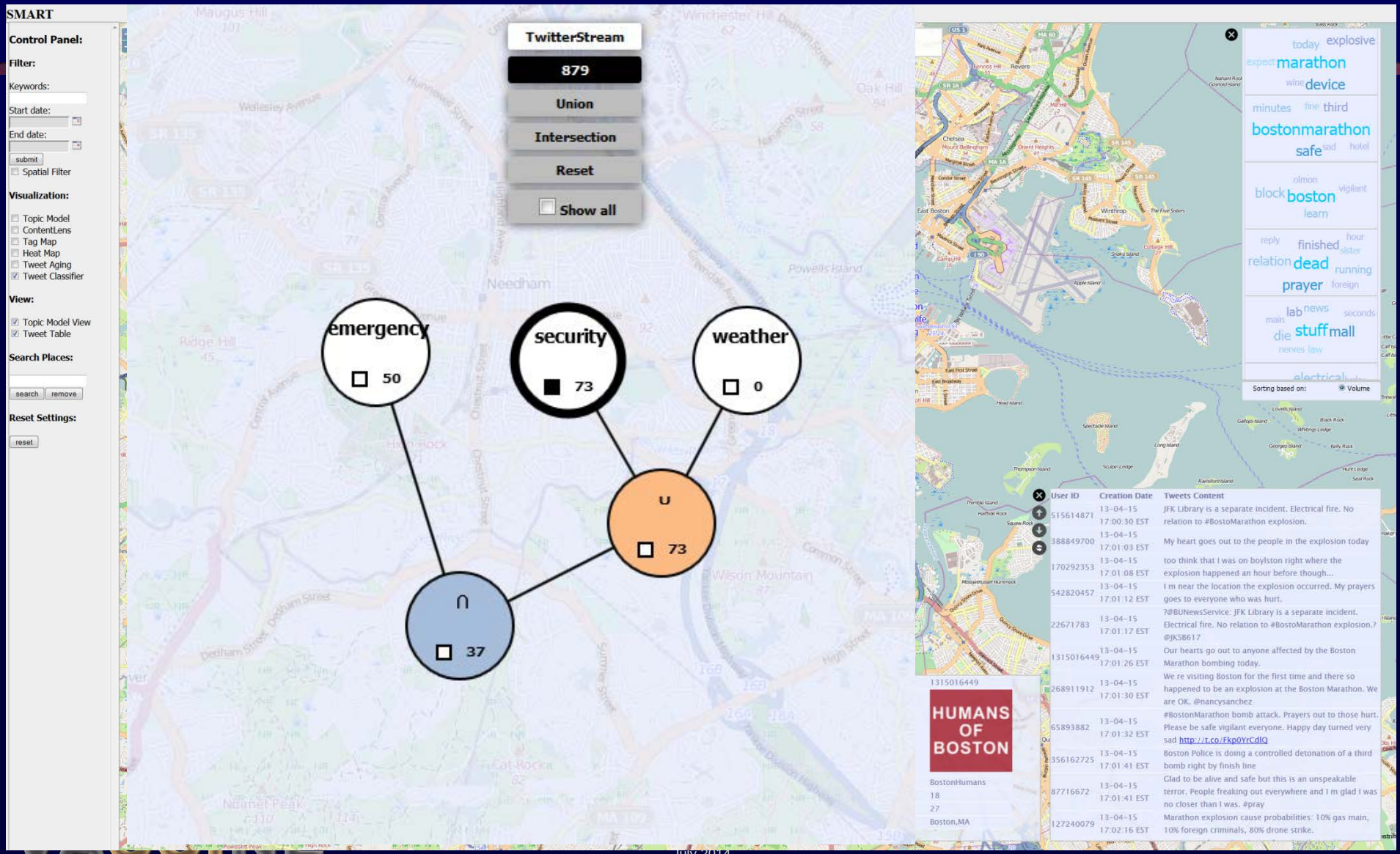
- Anomalous topic extraction using LDA and novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud



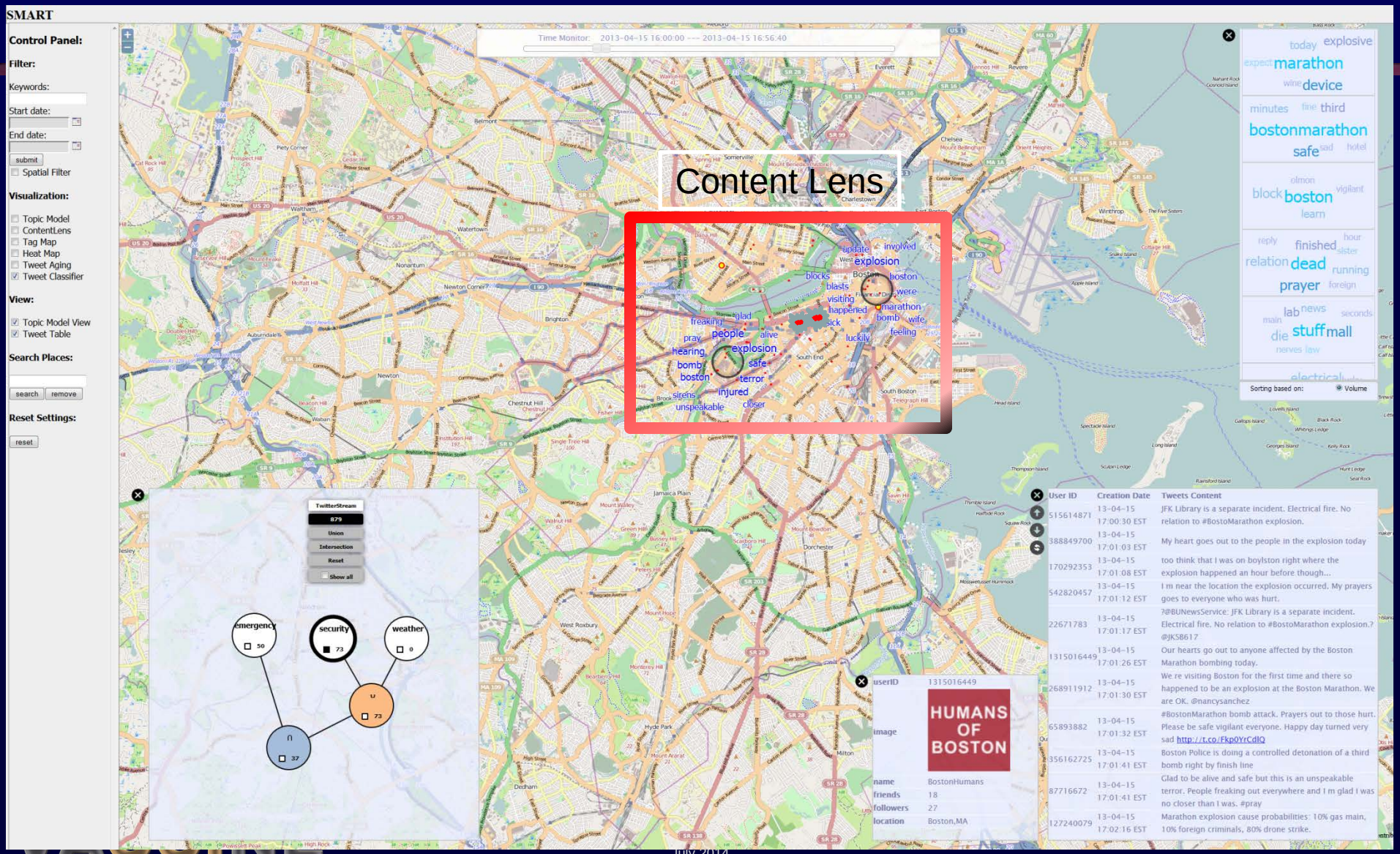
Social Media Analysis and Reporting Toolkit (SMART)



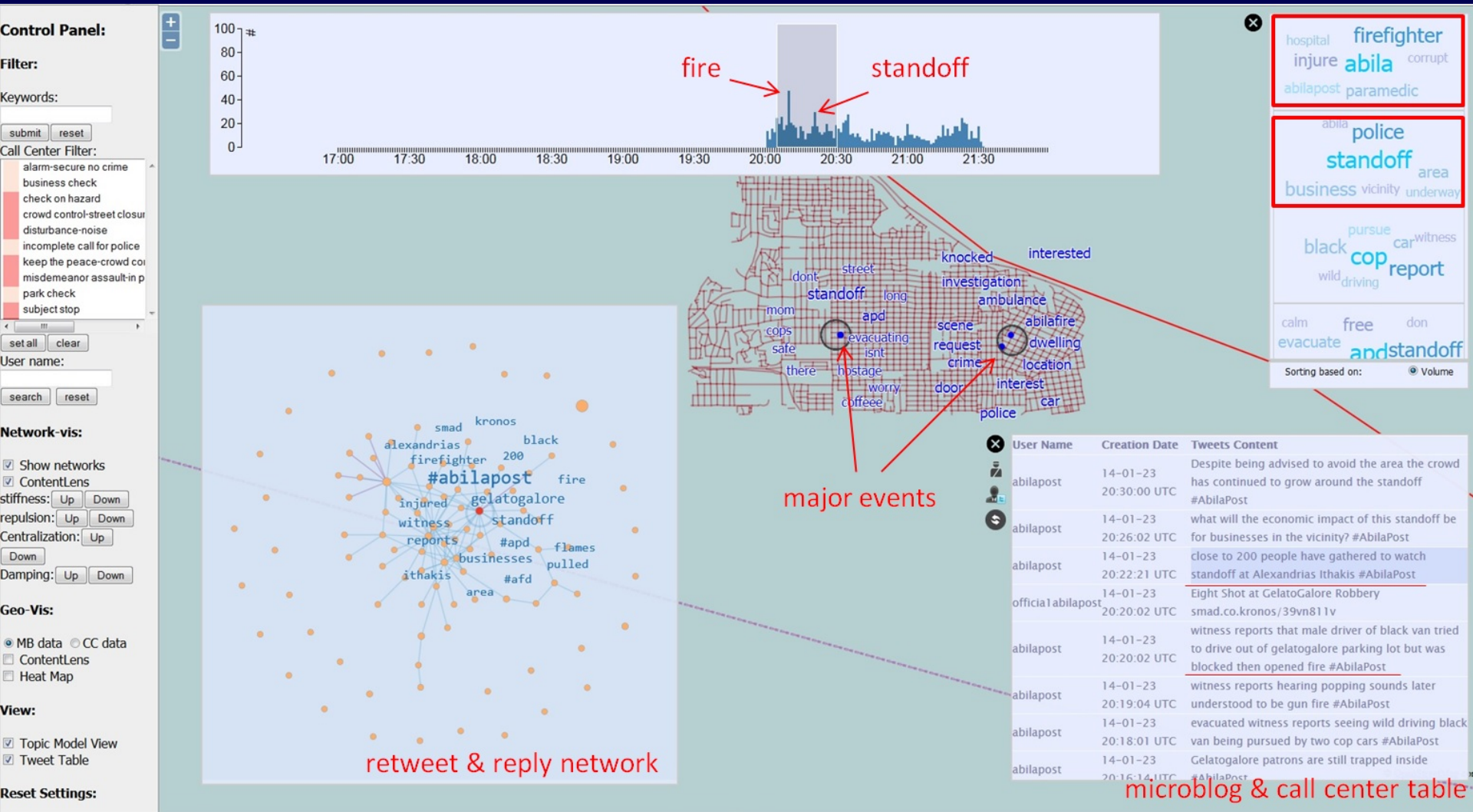
Social Media Analysis and Reporting Toolkit (SMART)



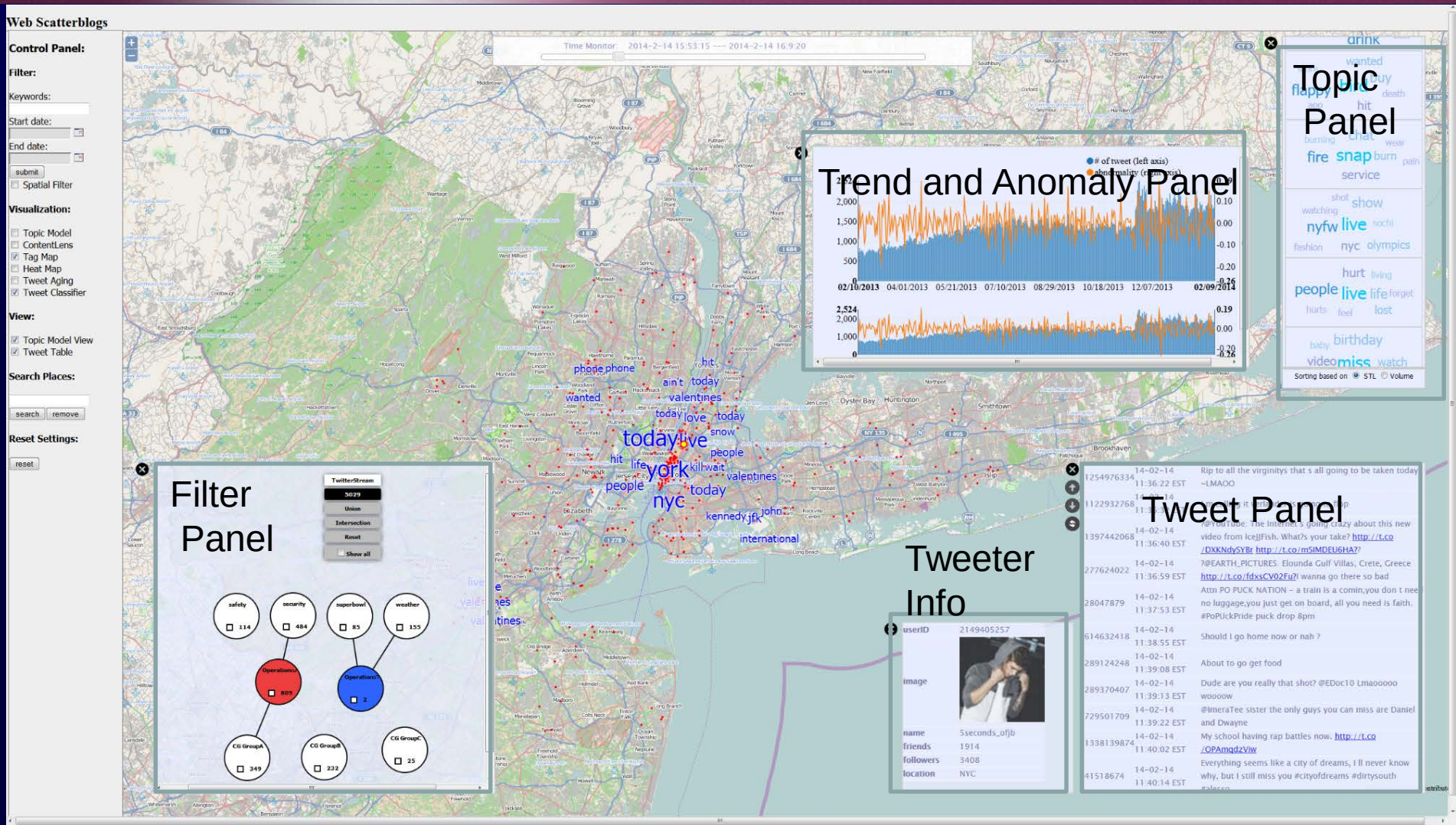
Social Media Analysis and Reporting Toolkit (SMART)



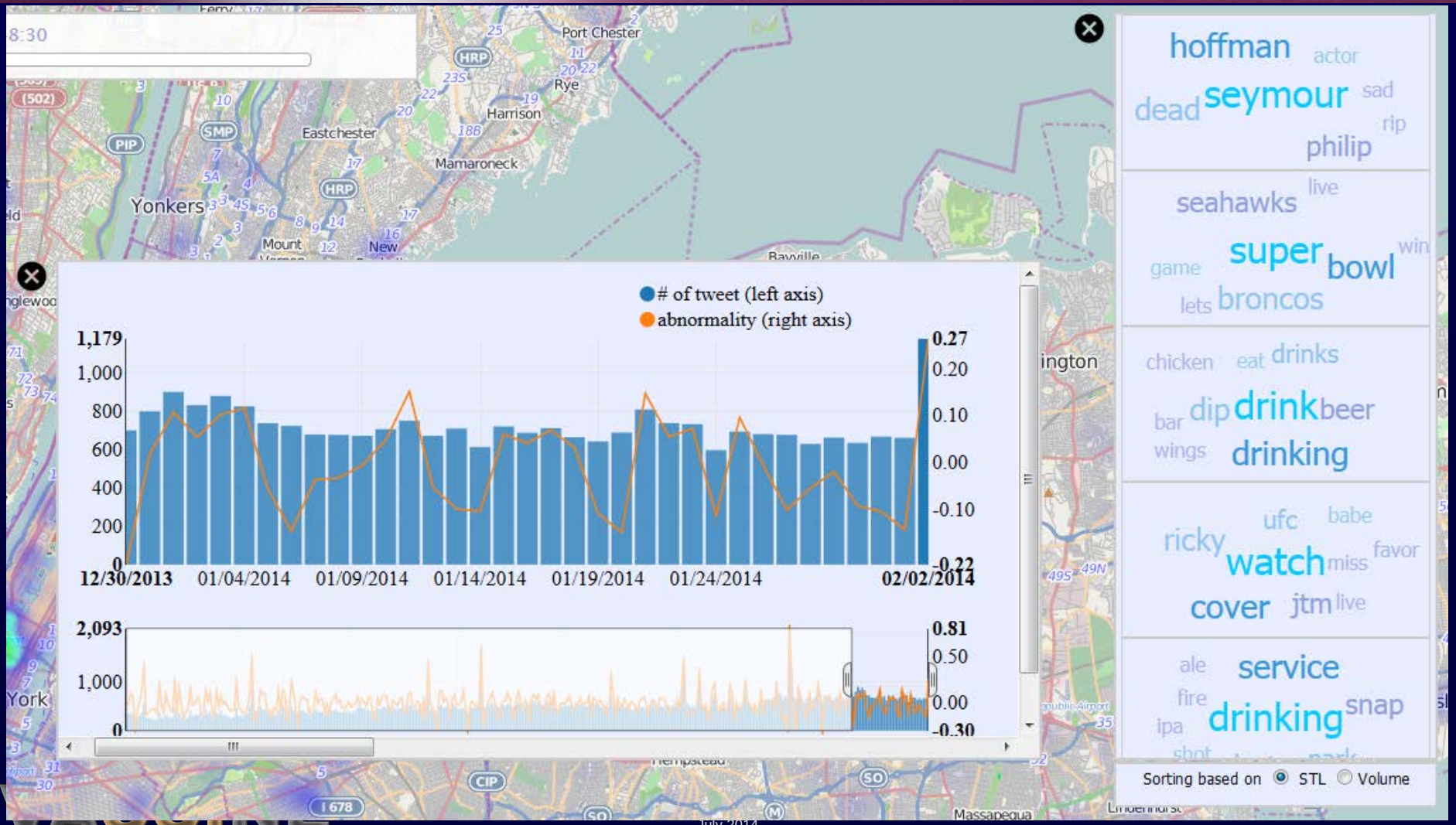
SMART: Network Linkage



SMART Anomaly Detection

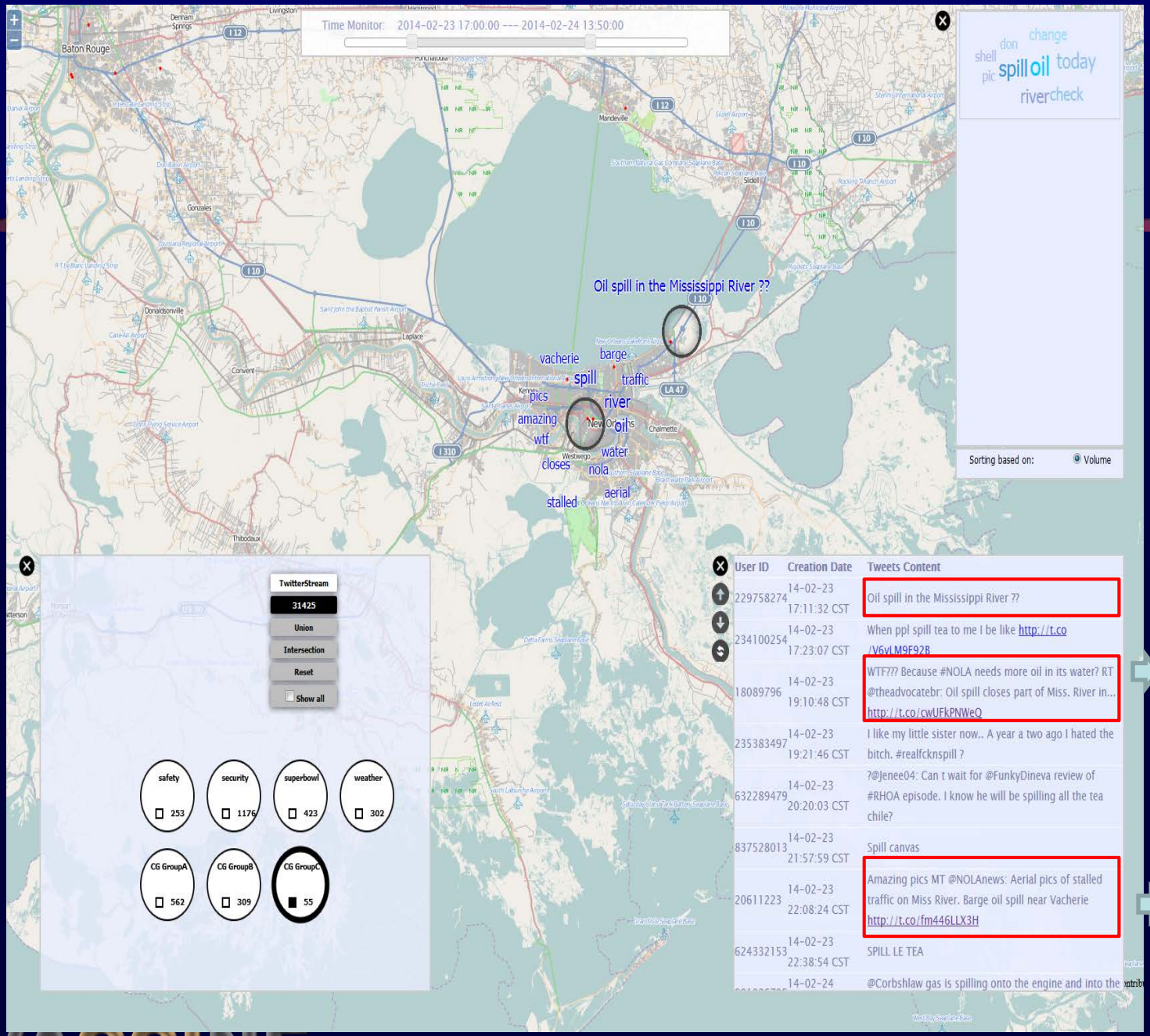


Anomaly Detection

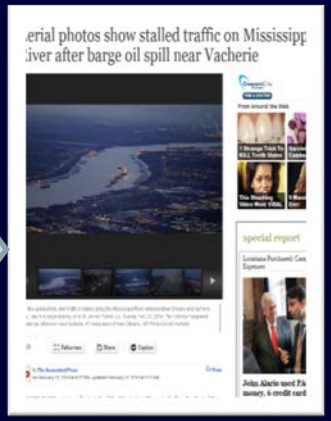


Oil Spill in Mississippi River

2014/02/23


Time Monitor: 2014-02-23 17:00:00 --- 2014-02-24 13:50:00
change
don
shell
pic
spill
oil
today
rivercheck
Oil spill in the Mississippi River ??
vacherie
barge
spill
traffic
pics
river
amazing
oil
wtf
water
closes
nola
aerial
stalled
Sorting based on:
Volume
TwitterStream
31425
Union
Intersection
Reset
Show all
safety
253
security
1176
superbowl
423
weather
302
CG GroupA
562
CG GroupB
309
CG GroupC
55
User ID
Creation Date
Tweets Content
229758274
14-02-23 17:11:32 CST
Oil spill in the Mississippi River ??
234100254
14-02-23 17:23:07 CST
When ppl spill tea to me I be like http://t.co/V6vLM9F92B
18089796
14-02-23 19:10:48 CST
WTF?? Because #NOLA needs more oil in its water? RT @theadvocatebr: Oil spill closes part of Miss. River in... http://t.co/cwUFkPNWeQ
235383497
14-02-23 19:21:46 CST
I like my little sister now.. A year a two ago I hated the bitch. #realfcknspill ?
632289479
14-02-23 20:20:03 CST
?@Jenee04: Can t wait for @FunkyDineva review of #RHOA episode. I know he will be spilling all the tea chile?
837528013
14-02-23 21:57:59 CST
Spill canvas
20611223
14-02-23 22:08:24 CST
Amazing pics MT @NOLAnews: Aerial pics of stalled traffic on Miss River. Barge oil spill near Vacherie http://t.co/fm446LLX3H
624332153
14-02-23 22:38:54 CST
SPILL LE TEA
14-02-24
@Corbshlaw gas is spilling onto the engine and into the


Oil spill closes part of Mississippi River in La.


Aerial photos show stalled traffic on Mississipp
River after barge oil spill near Vacherie

VACCINE

Non-geo Tweets On Twitter About Oil Spill Event

ProPublica @ProPublica · 3h
An **oil spill** just shut down 65 miles of the Miss. River. As production surges, how do we ship oil safely? propub.ca/1duUl9g @Bloomberg
View summary · Reply · Retweet · Favorite · More

Top news story
New Orleans LA @New_OrleansLA · 22h
Mississippi **Oil Spill** Highlights Risk of US Oil Boom - Bloomberg ift.tt/1ds9cRO #NewOrleans
Expand · Reply · Retweet · Favorite · More

Bloomberg News
Mississippi Oil Spill Highlights Risk of U.S. Oil Boom
A barge crash that spilled enough oil to temporarily shut a stretch of the Mississippi River highlights the transportation risks of the U.S. energy boom just as regulators respond to several rail...
View on web

Retweeted 19 times
Guardian Environment @guardianeco · Feb 24
Mississippi river closed at New Orleans after barge crash prompts **oil spill** gu.com/p/3n2q9/tf
View summary · Reply · Retweet · Favorite · More

Motherboard @motherboard · 4h
The Mississippi already a massive toxic cesspool every other day of the year: bit.ly/1flVUpJ
View summary · Reply · Retweet · Favorite · More

Marcy McGowan @marcylauren · 5h
Oil Spill Shuts Down Sixty-Five Mile Stretch of Mississippi River | Americans Against the Tea Party aattp.org/oil-spill-shut...
View summary · Reply · Retweet · Favorite · More

EDF Louisiana @EDF_Louisiana · 8h
#**oilspill** update: Lower Mississippi River reopens after **oil spill** near New Orleans. Watch the video bit.ly/MZB7ST
View summary · Reply · Retweet · Favorite · More

Sierra Club @sierraclub · 9h
The Mississippi River Was a Toxic Mess Long Before this Weekend's **Oil Spill**: sc.org/MSriverspill (via @motherboard)
Expand · Reply · Retweet · Favorite · More

Motherboard
The Mississippi River Was a Toxic Mess Long Before the Oil Spill
Everyone is outraged about a barge spilling oil into the Mississippi—perhaps everyone forgot that it is a massive toxic cesspool every other day of the year, too.
View on web



National Wildlife @NWF · 12h
A barge has crashed into a tugboat, spilling 31,500 gallons of oil into the Mississippi River. Read more: bit.ly/Mrrz2d
Expand · Reply · Retweet · Favorite · More

Margarita Noriega™ @margafret · 13h
31,500 gallons of light crude oil spilled into the Mississippi River weather.com/news/science/e...
View summary · Reply · Retweet · Favorite · More

National Wildlife @NWF · 14h
New **Oil Spill** Fouls Mississippi River: bit.ly/1gzqS01
Expand · Reply · Retweet · Favorite · More

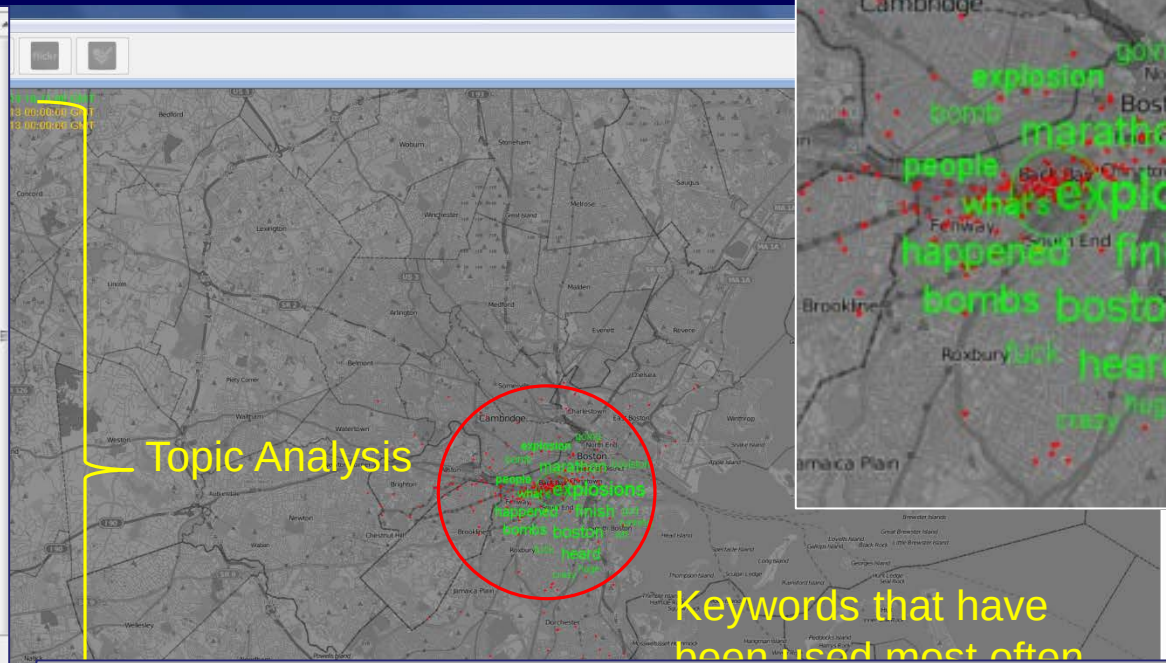
inhabitat @inhabitat · Feb 24
It is unclear when the ships stuck on the Mississippi River will be able to move as crews clean up a new **oil spill** bit.ly/1dpdo4V
Expand · Reply · Retweet · Favorite · More

inhabitat
Louisiana Oil Spill Shuts Down New Orleans Port and Parts of the...
The Port of New Orleans and surrounding areas were closed Sunday after a barge collided with a tow boat, causing oil to spill into the Mississippi River. The boats were located between Baton Rouge...
View on web



Explosion Area in Boston





Keywords that have been used most often

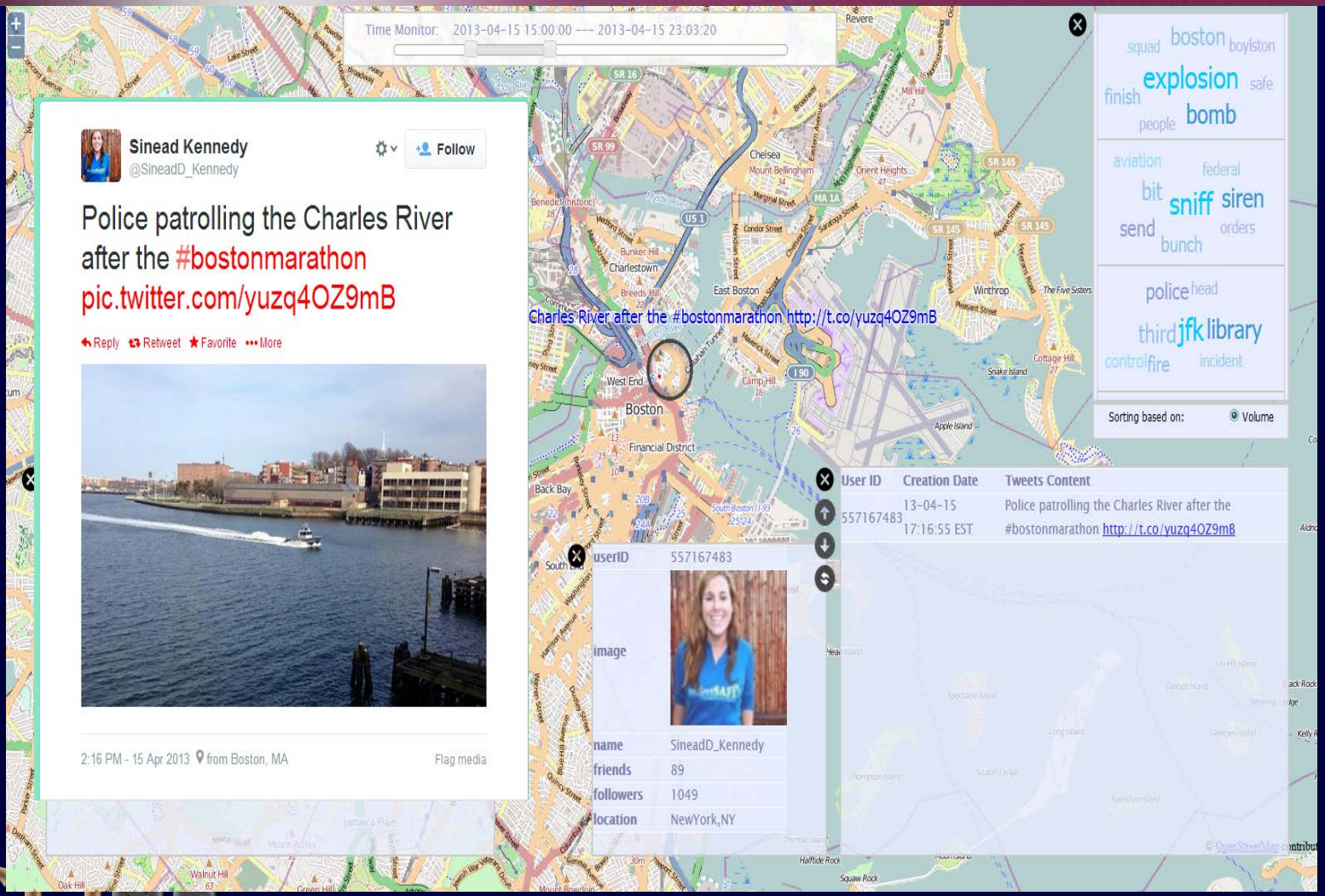
Text
Oh my god what just happened
Something happened at the end of the Boston Marathon. Something bad and there is a lot of chatter on Twitter. What's going on?
Multiple people are injured near the Boston Marathon finish line after two explosions. The #BostonMarathon has been stopped.
two bombs just went off on boylston
BREAKING NEWS: Two powerful explosions detonated in quick succession right next to the Boston Marathon finsh line this afternoon.
What the FUCK was that
Back in Sept, @croon1 solicited me for $2000. He now has a music video with William Shatner. If you watch it (god forbid) keep that on mind.
Literally what the fuck get me out of here
@DTenenbaum my office right next to it
@FRANCESCalciO I figured 3 people would get the joke
Two explosions just rocked the finish line of the Boston Marathon. Sirens galore. People running in fear. Wonder what happened.
This is crazy i seen that blow up #bostonmarathon
can someone tell me what that explosion was!? #boston #bostonmarathon

Tweets

First Response (Tweet & Picture)



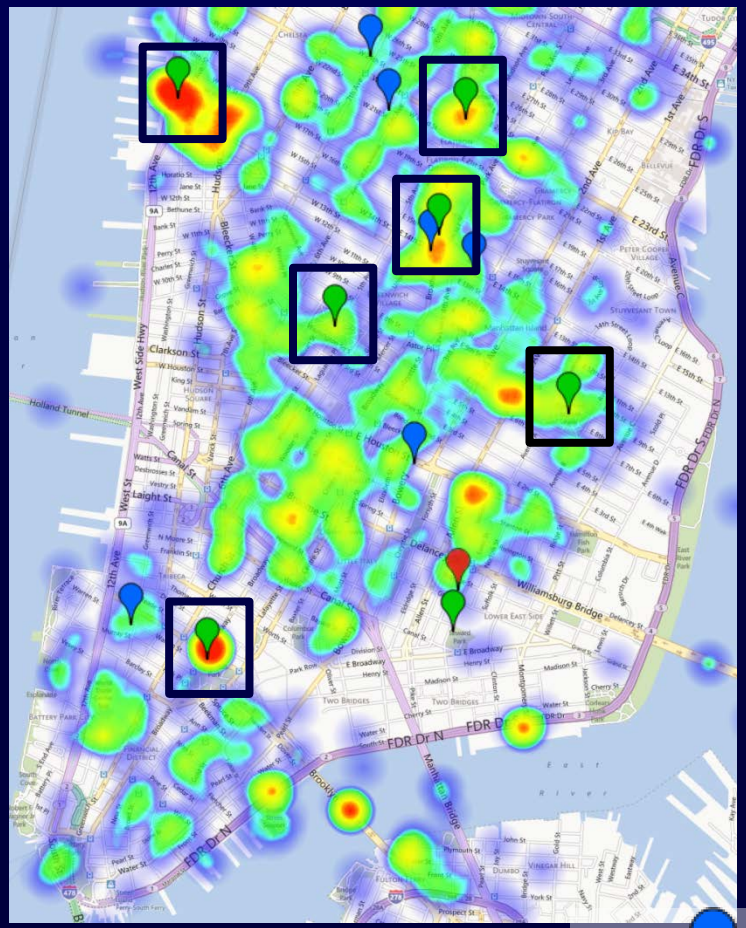
Police Patrolling The River

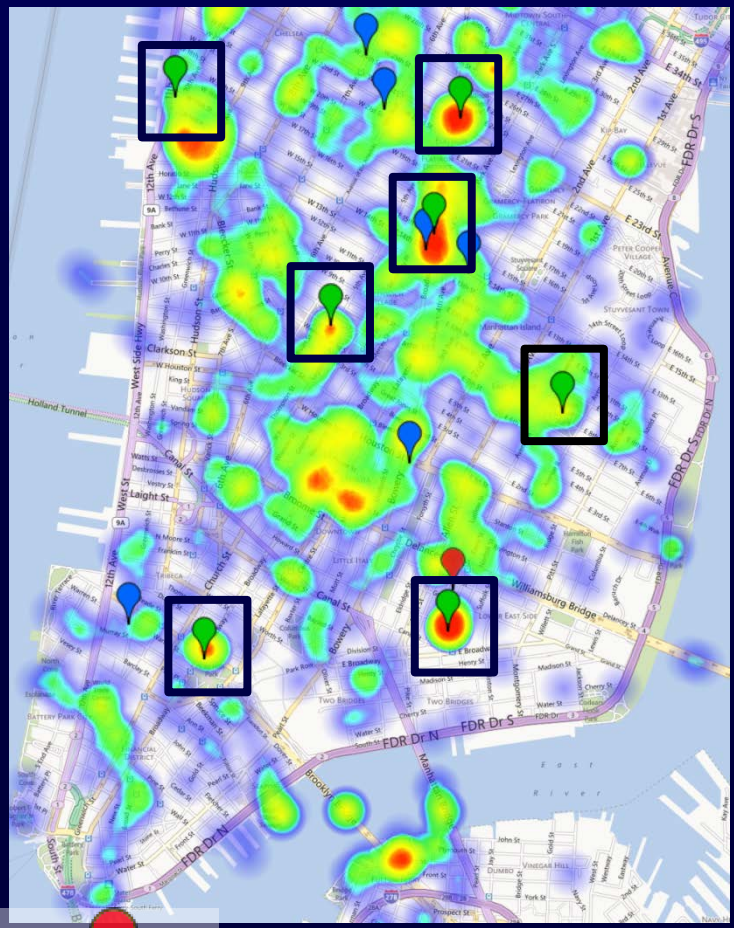


Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00





Supermarket Park Shelter

VACCINE

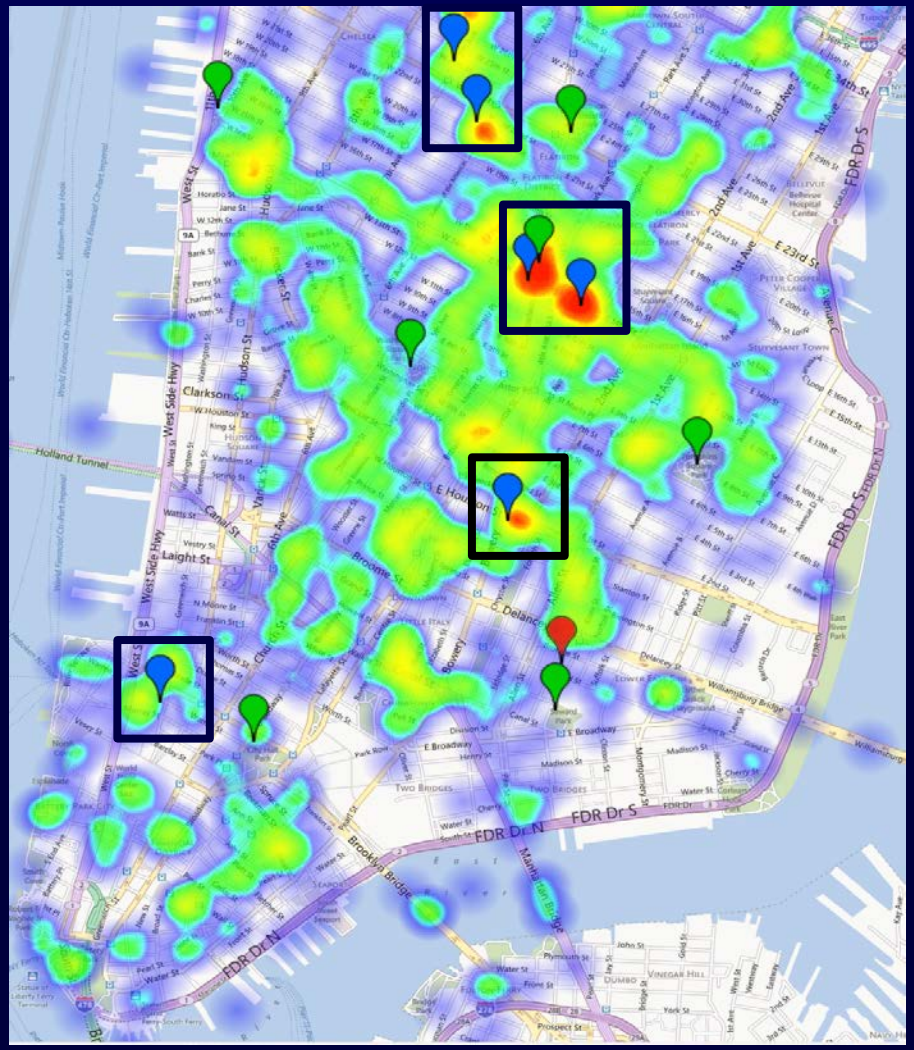
July 2014

Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

Evacuation order:
10/28, 10:30 AM

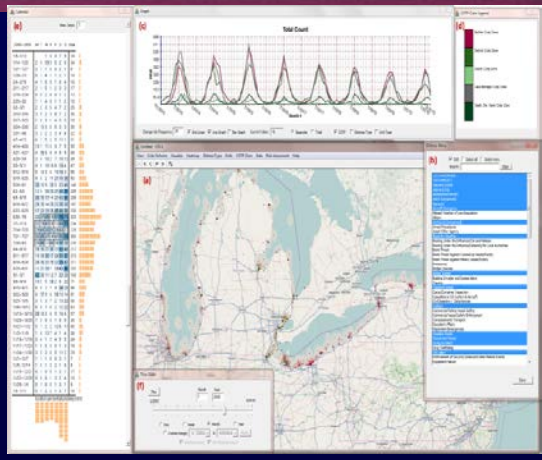
Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



VACCINE

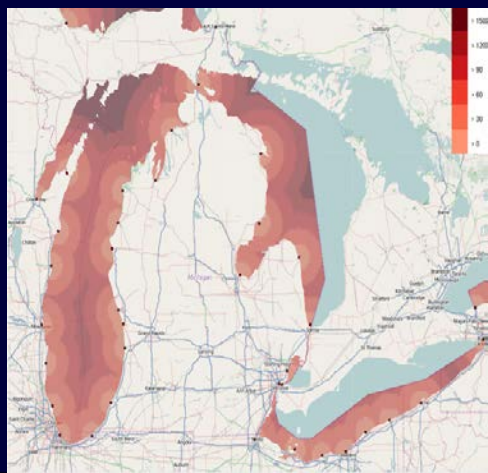
U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771



IMPACTS:

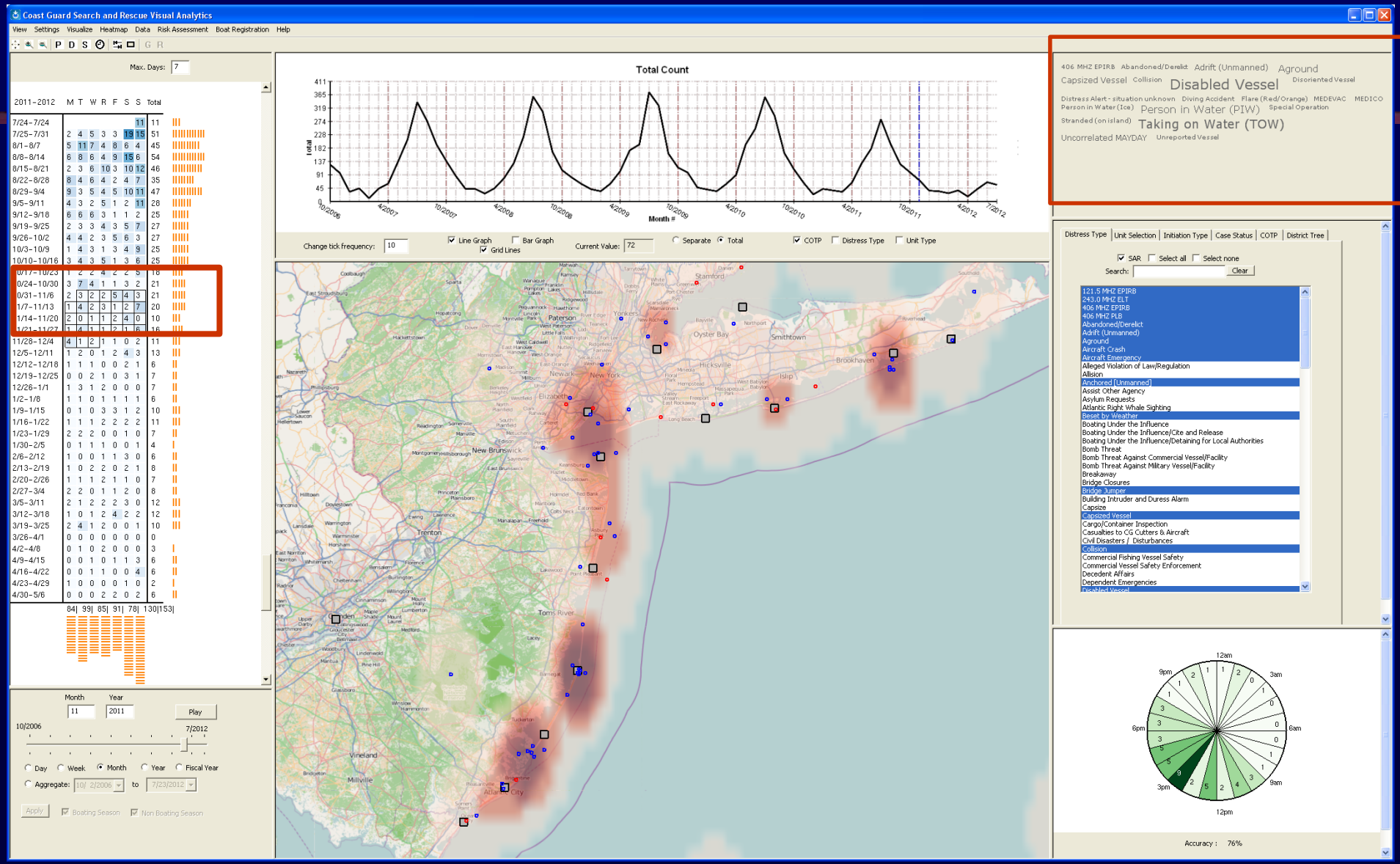
- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ



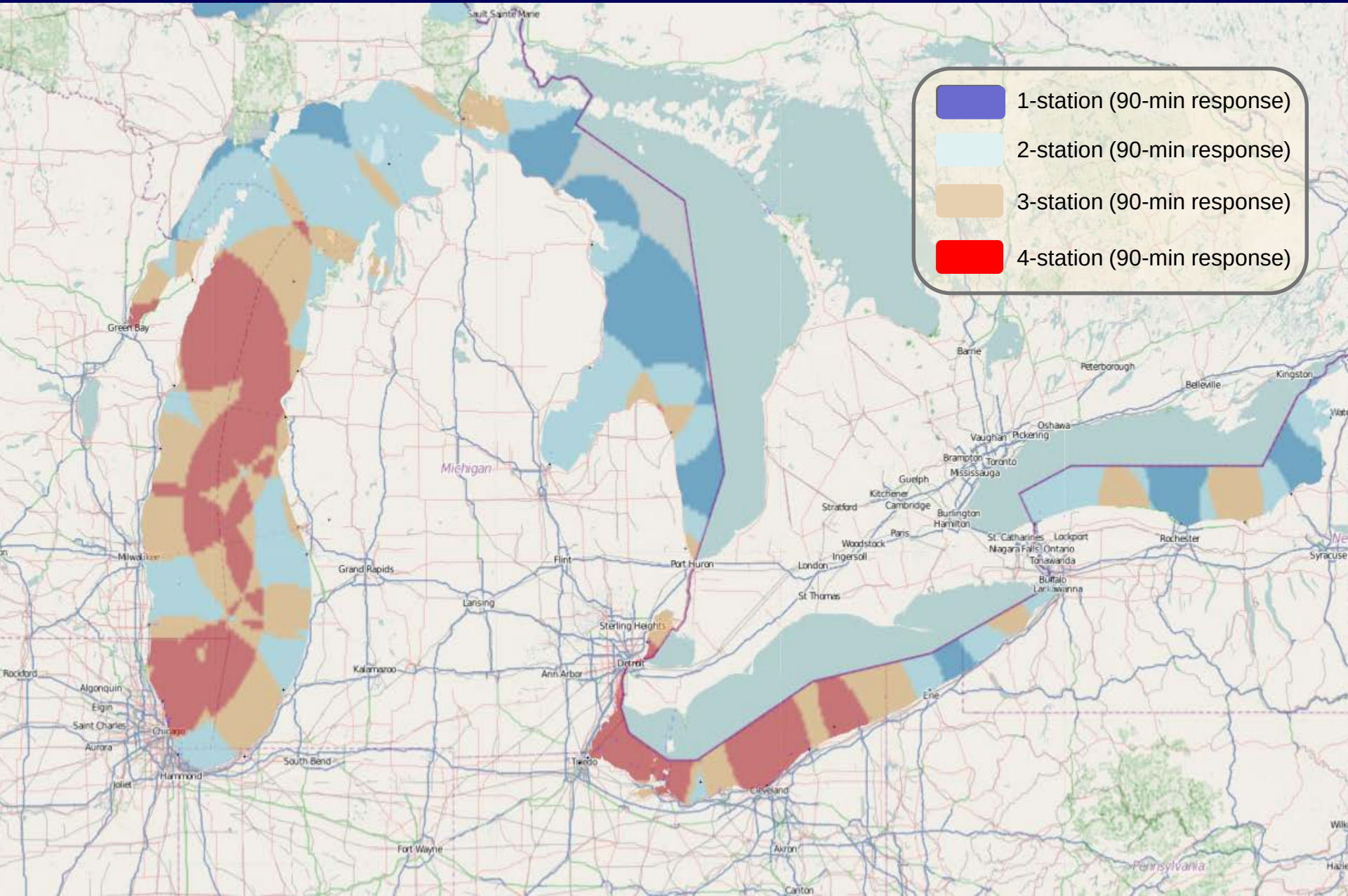
VACCINE

Example: Risks and Consequences From Sandy:

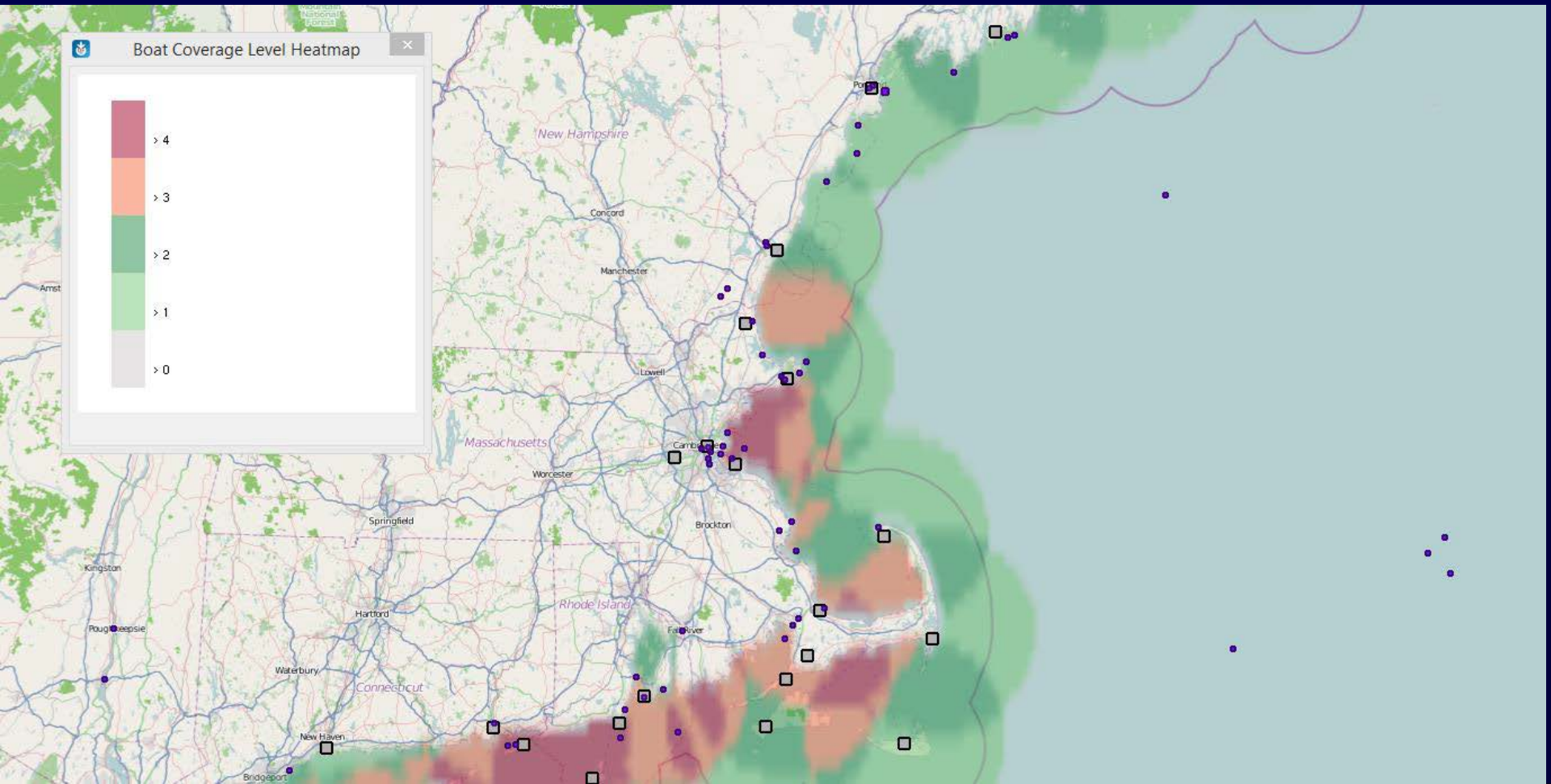
SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



Coverage vs Lives Lost Aggregated Since Fiscal Year 2007

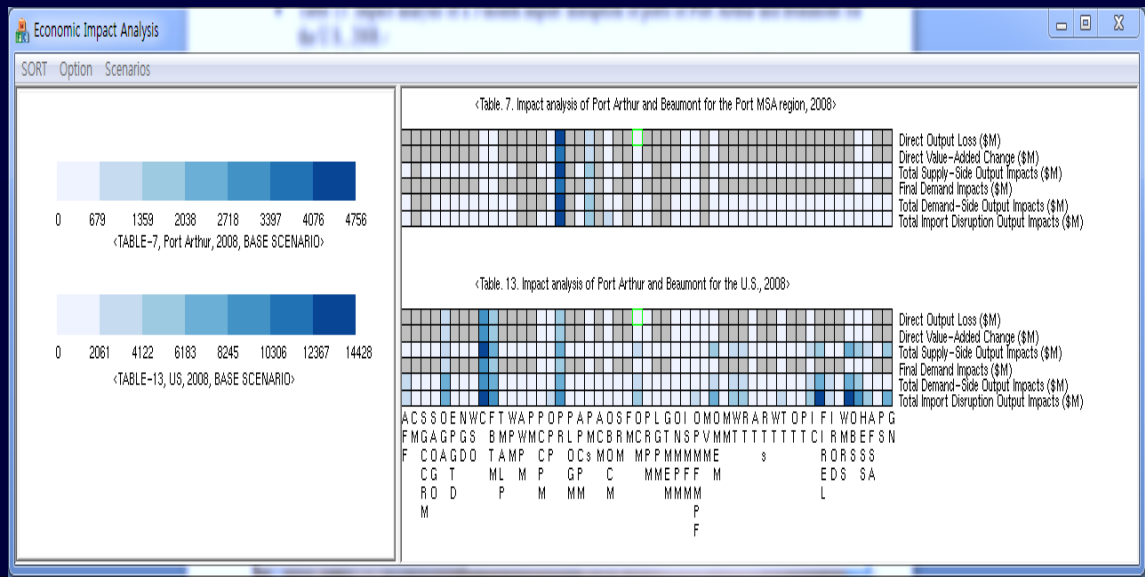


USCG Port Closure Economic Impact VA

Partners: USC CREATE, USCG RDC, USCG D7, USCG LANT

IMPACT:

- Provided tool for use analysis and planning for impact of port closure in Port Arthur, Tx
- Economic sector impact, local and national impact
- Impact and effectiveness of alternative mitigation strategies



VACCINE

For Further Information

www.VisualAnalytics-CCI.org

