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 7

Addendum A - Presentations

JULY 1, 2015 – JUNE 30, 2016

Cooperative Agreement No. 2009-ST-061-CI0001

PURDUE UNIVERSITY

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

Arizona State University







esri

ng, Wei Luo, Elizabeth Mack

Ross Maciejewski

& Data Exploration Research Lab

Problem

- Multivariate clustering has been widely applied in various domains over various geographical areas
  - Regional ecosystems[1]
  - Demographic maps[2]
  - At risk neighborhood[3]

1. MILLS R. T., HOFFMAN F. M., KUMAR J., HARGROVE W. W.: Cluster analysis-based approaches for geospatiotemporal data mining of massive data sets for identification of forest threats. Procedia Computer Science 4 (2011), 1612–1621.
2. CHESHIRE J., MATEOS P., LONGLEY P. A.: Delineating Europe's cultural regions: Population structure and surname clustering. Human Biology 83, 5 (2011), 573–598
3. ASHBY D. I., LONGLEY P. A.: Geocomputation, geodemographics and resource allocation for local policing. Transactions in GIS 9, 1 (2005), 53–72.

Problem and Goal

- Local variations in the data and choices of clustering parameters can greatly impact the resultant visualization.
  - These methods (kmeans) are usually, but not always[1] applied at a level that is agnostic to the spatial relationships between the data

- The goal is to develop an understanding of global and local multivariate clustering for geographical visualization

1. MASON G., JACOBSON R.: Fuzzy geographically weighted clustering. In Proceedings of the 9th international conference on geocomputation, Maynooth, Eire, Ireland (2007), pp. 3–5.

Clustering evaluation and comparison

- Related Work
  - Rand Index[1] (Rand, 1971)

$$R = \frac{a+b}{a+b+c+d}$$

  - Normalized Mutual Information[2] (Manning et al., 2008)

$$NMI(X,Y) = \frac{H(X)+H(Y)}{H(X,Y)}$$

  - Variation of Information[3] (Meila, 2005)

$$VI(X,Y) = H(X) + H(Y) - I(X,Y)$$

  - Clustering gain[4] (Jung et al, 2003)

$$\delta = \sum_{j=1}^{k} (n_j - 1)||p_0 - p_0^{(j)}||_2^2$$

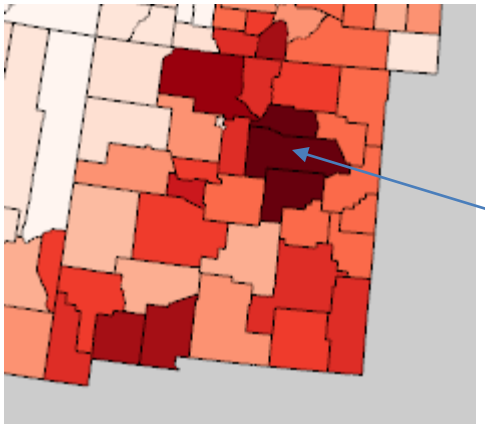
They often remain at the arithmetic level or lack the spatial context

1. RAND W. M.: Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association,* 66, 336 (1971)
2. MANNING C. D., RAGHAVAN P., SCHÜTZE H.: *Introduction to information retrieval*, vol. 1. Cambridge university press, 2008
3. MEILA M.: Comparing clusterings: An axiomatic view. *In Proceedings of the 22nd international conference on Machine learning* (2005)
4. JUNG Y., PARK H., DU D.-Z., DRAKE B. L.: A decision criterion for the optimal number of clusters in hierarchical clustering. *Journal of Global Optimization* 25, 1 (2003).

Geographical Variation

- *Spatial dependency:* similarity in attribute values of nearby spatial units

Tobler's first law of geography[1]:
"Everything is related to everything else, but near things are more related than distant things."

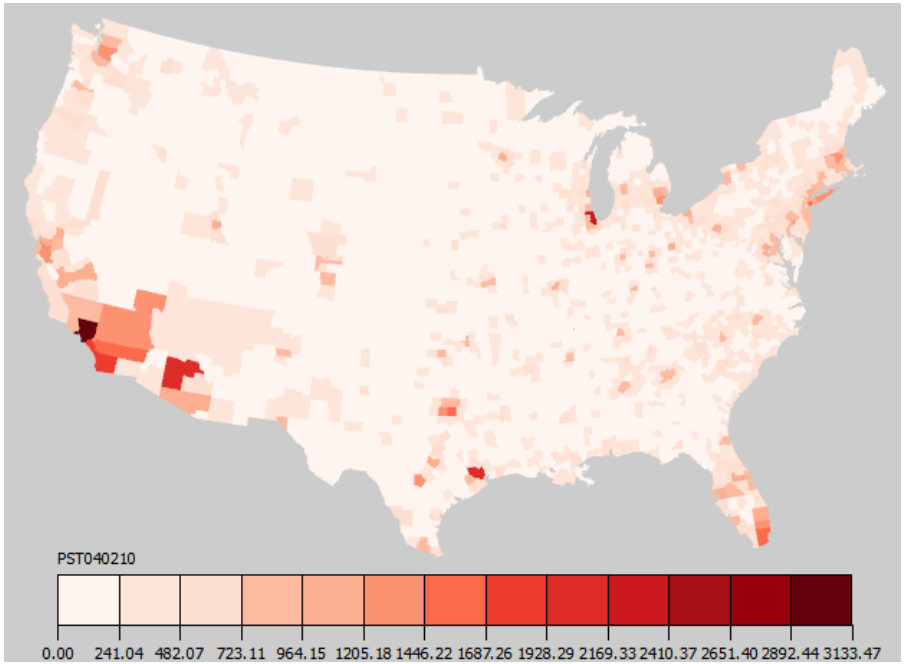


Percent of Hispanic or Latino around San Miguel County in New Mexico

1. TOBLER W. R.: A computer movie simulating urban growth in the detroit region. Economic geography (1970), 234–240.

Geographical Variation

- *Spatial heterogeneity*[1]*:* refers to the uneven distribution of a trait, event, or relationship across a region.



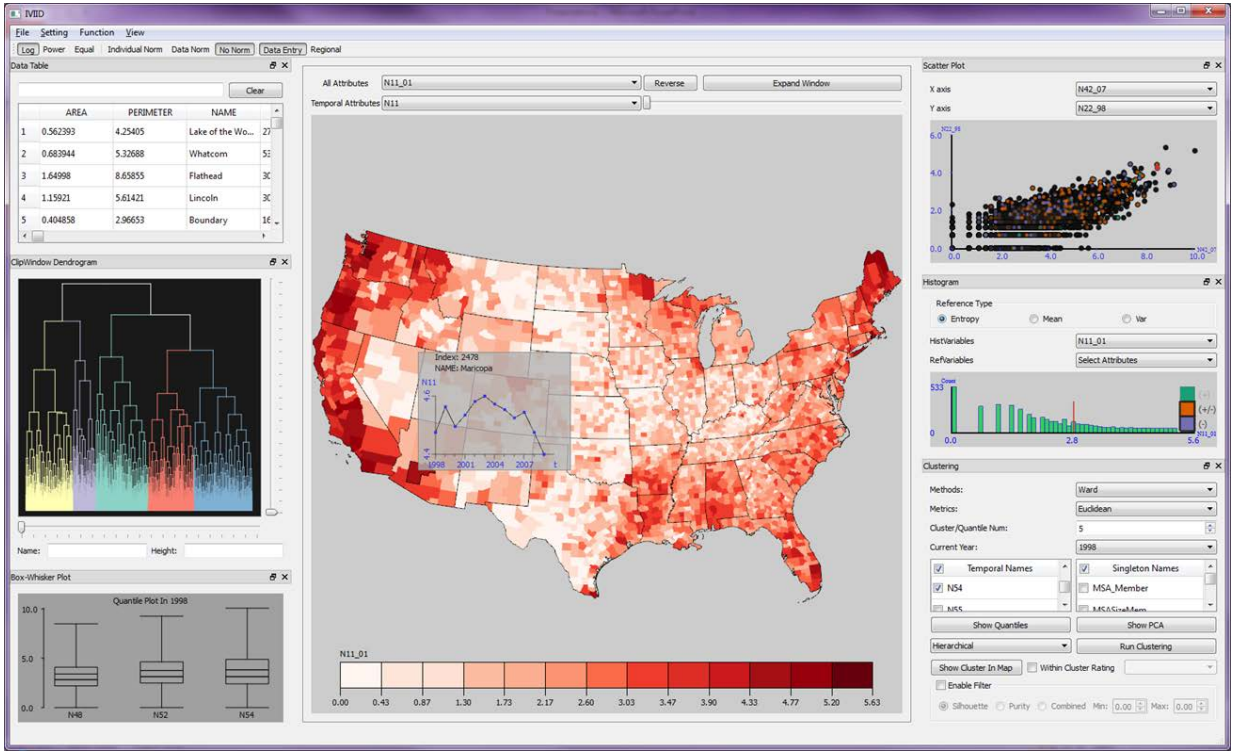
variation rather than similarity in values for a particular measures across all spatial units

Distribution of estimated population in 2010. It should not be surprising that there are clusters of high density in some places and areas that have very low density in between

1. Anselin, Luc. Thirty years of spatial econometrics. *Papers in Regional Science* 89(1) (2010) : 3–25

Interactive Visual Analytics Framework

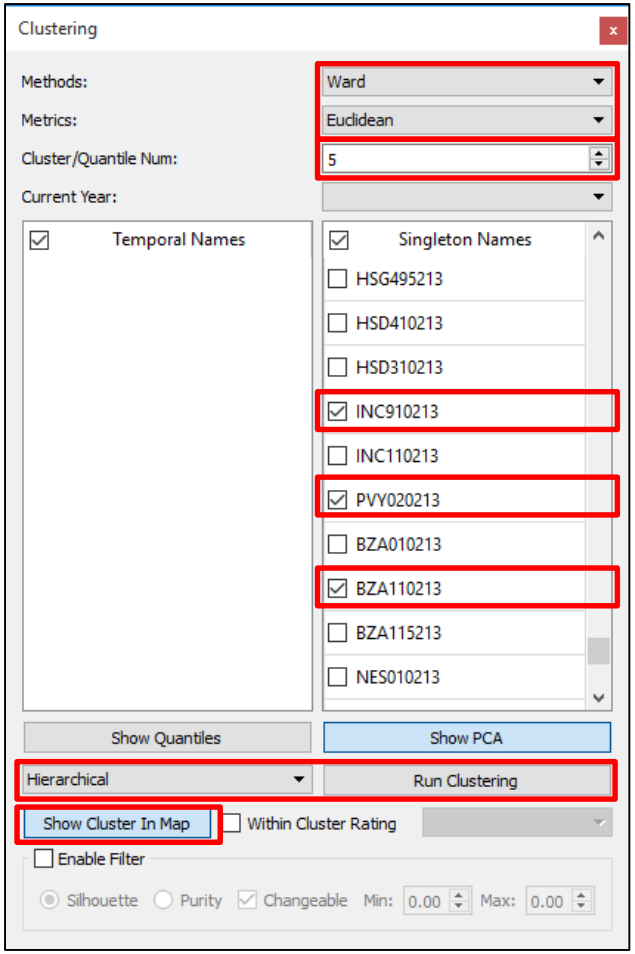
A framework[1,2] that integrates various visualization techniques (linked brushing, dynamic tick-marks, focus+context highlighting).

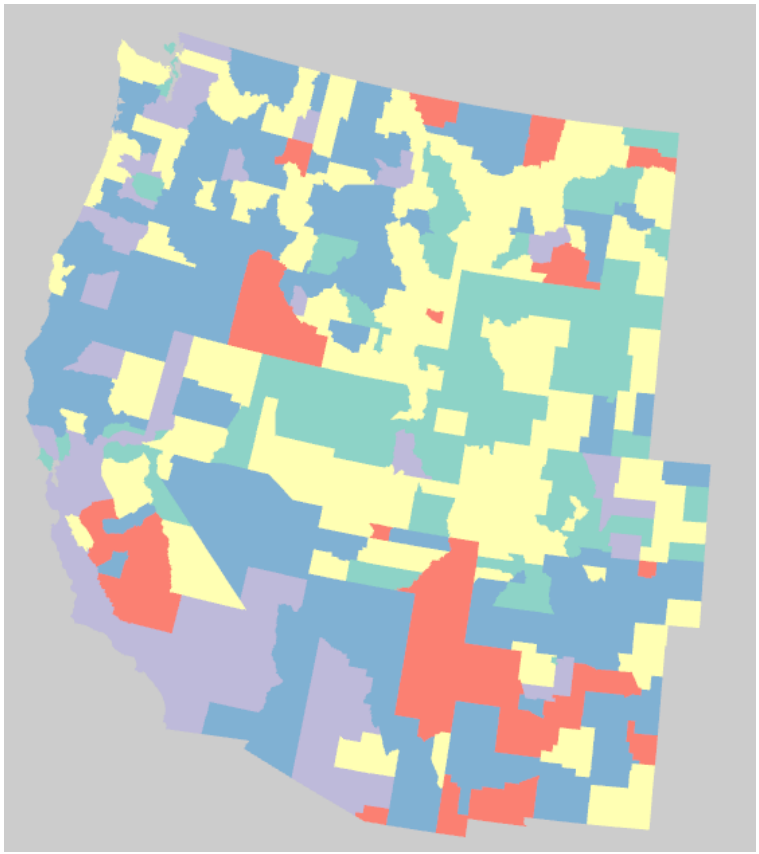


1. Mack, E. A., Zhang, Y., Rey, S., & Maciejewski, R. (2014). Spatio-temporal analysis of industrial composition with IVIID: an interactive visual analytics interface for industrial diversity. *Journal of Geographical Systems*, 16(2), 183-209.
2. Landis, S., Rezaeedaryakenari, B. , Zhang, Y., Thies, C.G., Hubbard, K., Maciejewski, R., & Luo, W. (2016). Fording Differences? Analyzing Climate Instability, Political Violence, and Social Unrest in the Niger River Basin. Revise and Resubmit

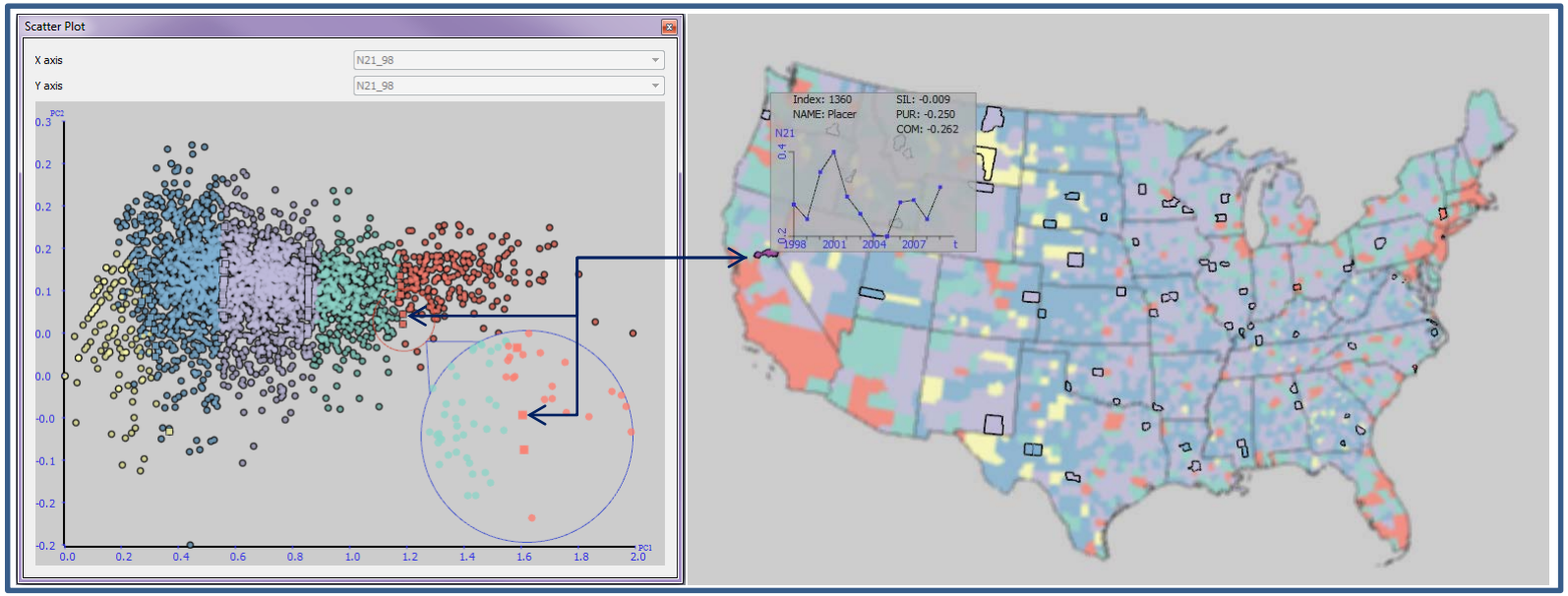
Interactive Visual Analytics Framework

To use the framework on multivariate map classification





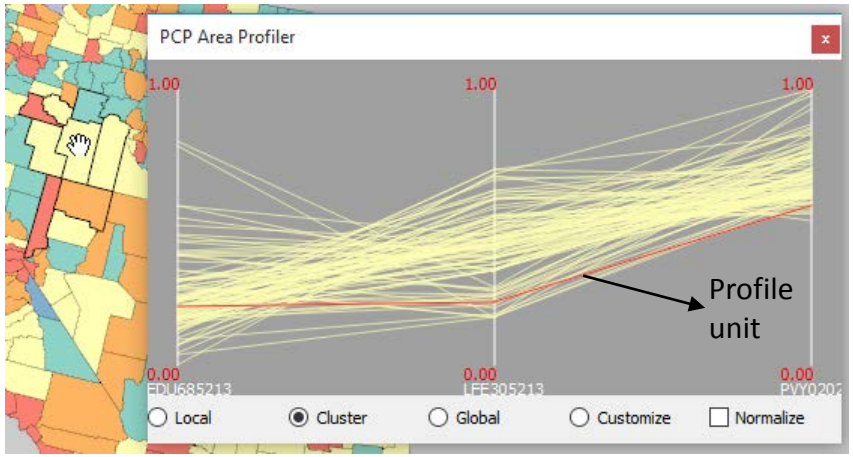
Clustering Exploration

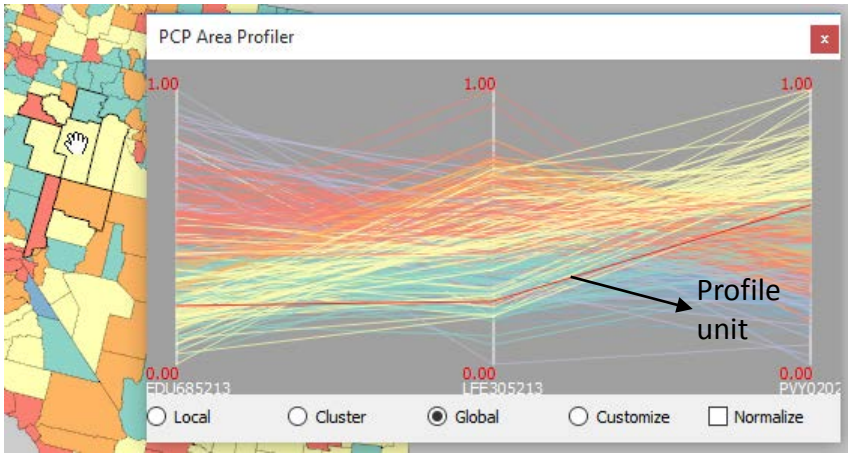


- PCA (Principle Component Analysis) scatterplot: generalized an overview of how clusters are distributed

Clustering Exploration



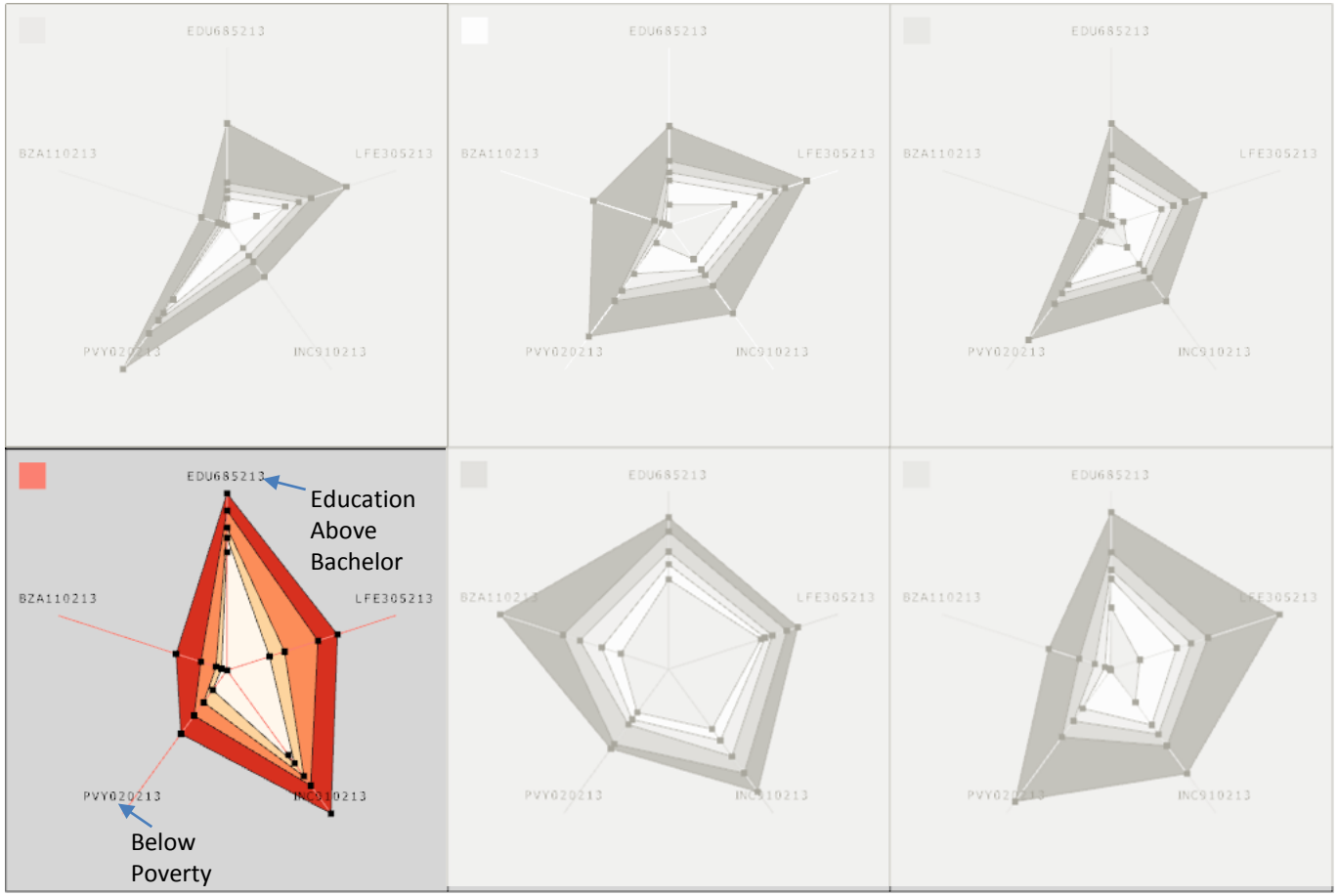






PCP Area Profiler: visualize multivariate relations of different area profiles

Clustering Exploration



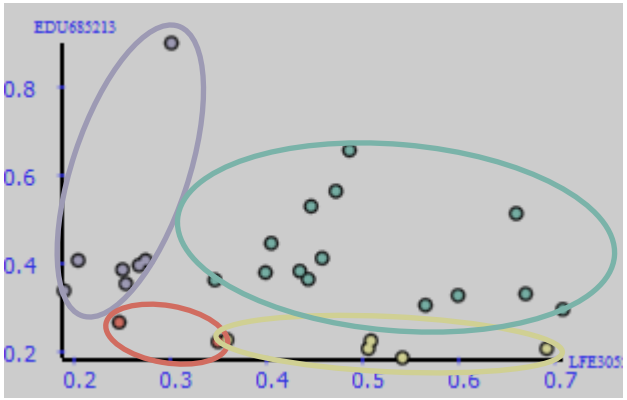
Rose Plot: statistical characteristics (Min, Max, 3 Quartiles) of each variable for each cluster

Clustering Exploration

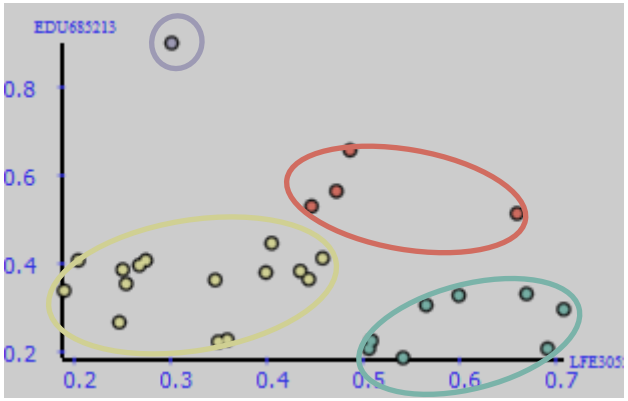
- Silhouette coefficient[1]: as a means of assessing the goodness of a clustering result

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}},$$

$$-1 \leq S(i) \leq 1$$



0.245



0.549

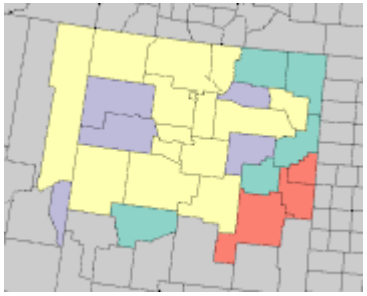
1. ROUSSEEUW P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*. 1987

Clustering Exploration

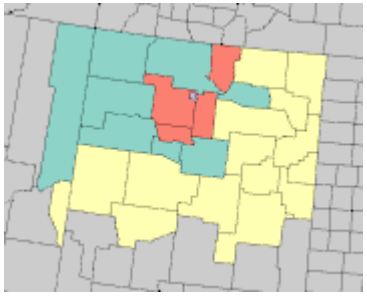
- Purity Indicator: Gini-index[1] like indicator that based on number of units that belong to the same class

$$P(i) = (\frac{n_{C_i}}{N(i)})^2 - \sum_{C_i \neq C_j} (\frac{n_{C_j}}{N(i)})^2,$$

$$-1 \leq P(i) \leq 1$$



-0.027

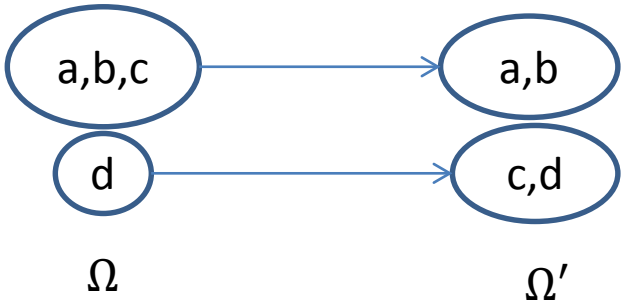


0.301

1. GINI C.: *Variabilità e mutabilità*. 1912.

Clustering Comparison

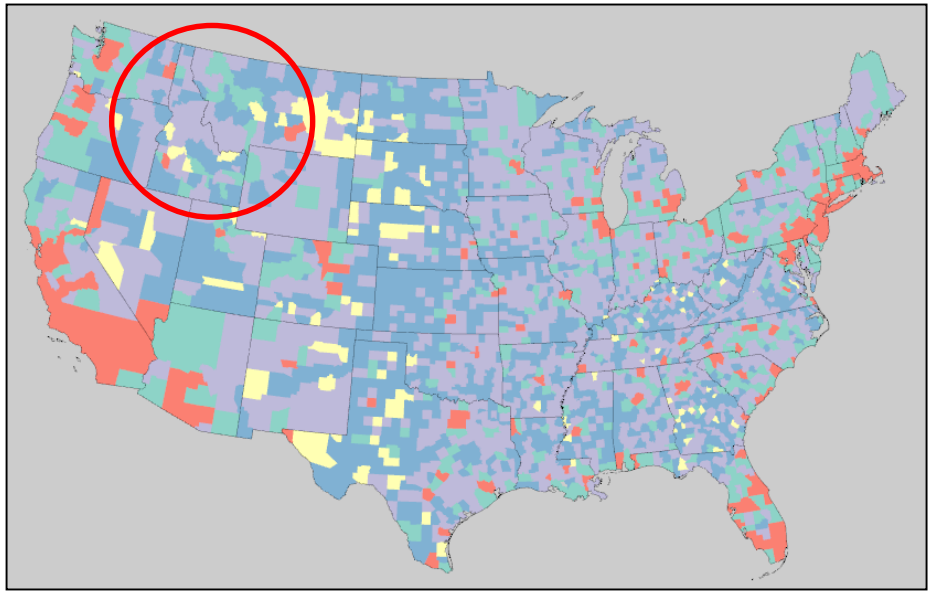
- Cluster coloring[1]
  - Intuitive way of comparing clustering results by matching the color
  - Attempt to keep labels of spatial units consistent
    - Kuhn-Munkres (KM) algorithm[2] which treats the clusters between different clusterings as a weighted bipartite graph and solves it as an optimal assignment problem

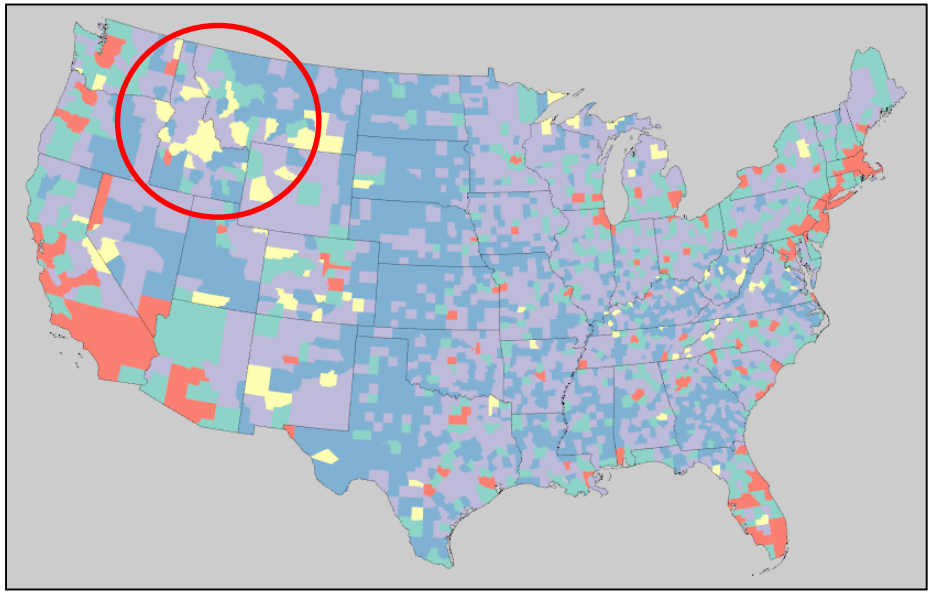


minimize the number of spatial units that need to be relabeled

1. Hu, Yifan, Stephen G. Kobourov, and Sankar Veeramoni. "*Embedding, clustering and coloring for dynamic maps*." PacificVis. IEEE, 2012.
2. J. Munkres, "Algorithms for the Assignment and Transportation Problems", *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957

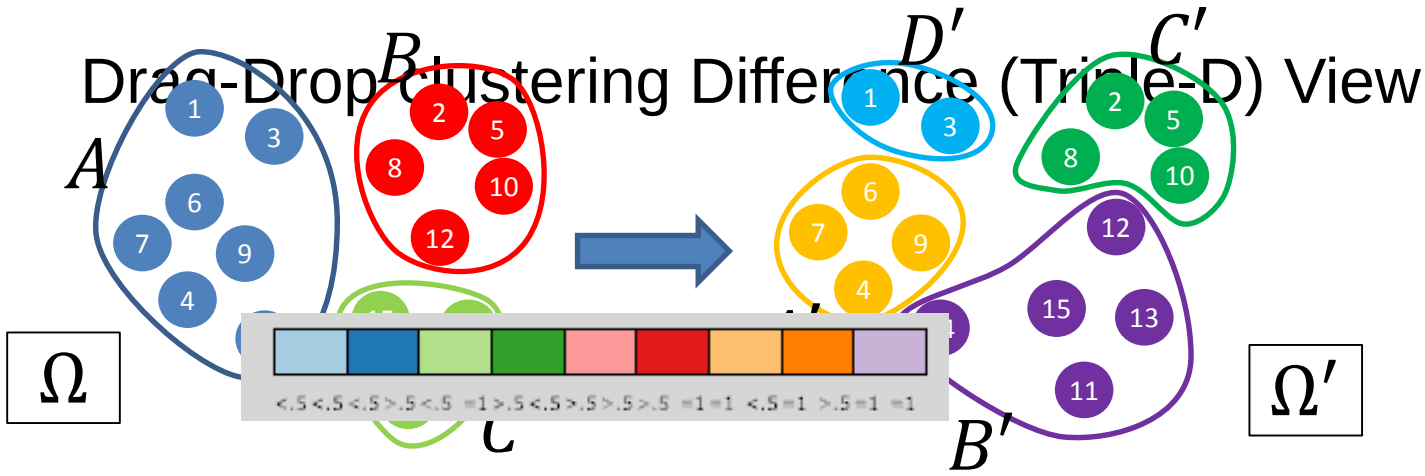
Clustering Comparison

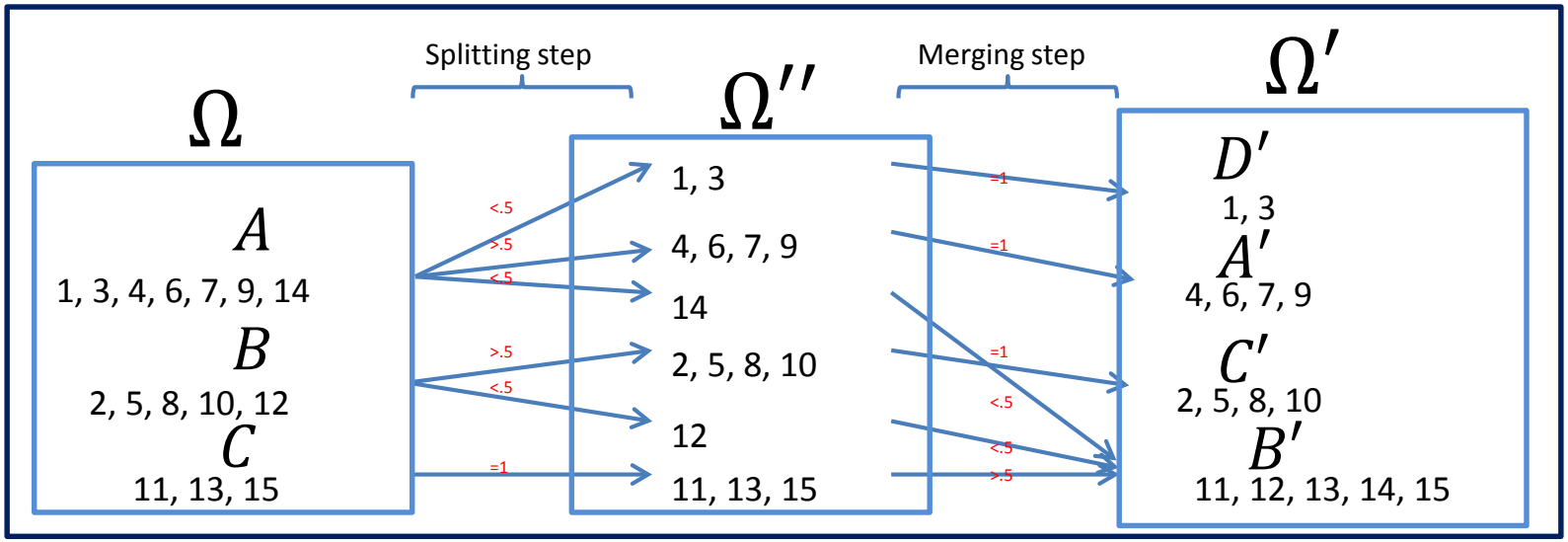




1. It lacks a detail comparison capability
2. It can only compare clustering with the exact same cluster numbers and units

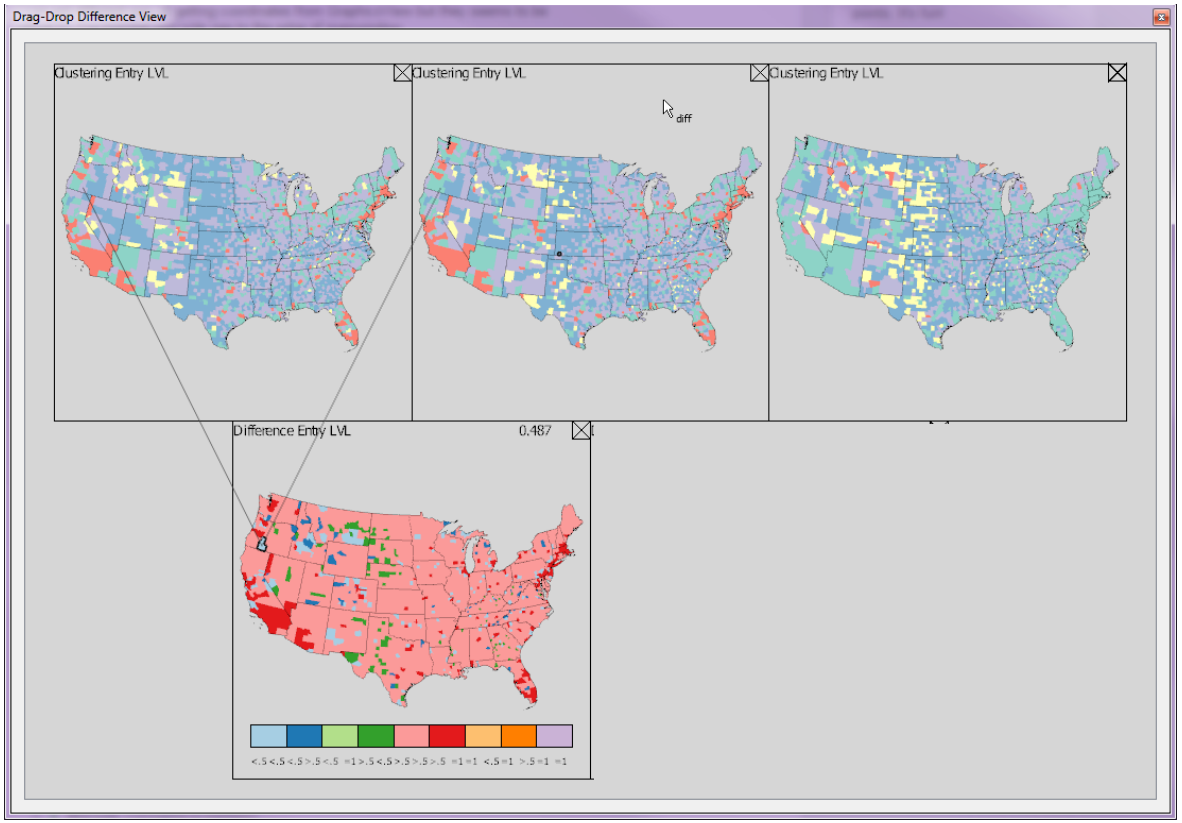
Clustering Comparison





Clustering Comparison

- Drag-Drop clustering Difference View
  - Dragging and dropping between two clustering results, the Triple-D view will map the changes (i.e., the intermediate sub-clusters) under the two clustering being compared



Clustering Comparison

- split-merge metric:

$$SM(\Omega, \Omega'') = -\sum_i \sum_j \frac{|C_i \cap C'_j|}{N} log \frac{|C_i \cap C'_j|^2}{|C_i||C'_j|},$$

Larger metric values represent more dissimilarity between clusterings.

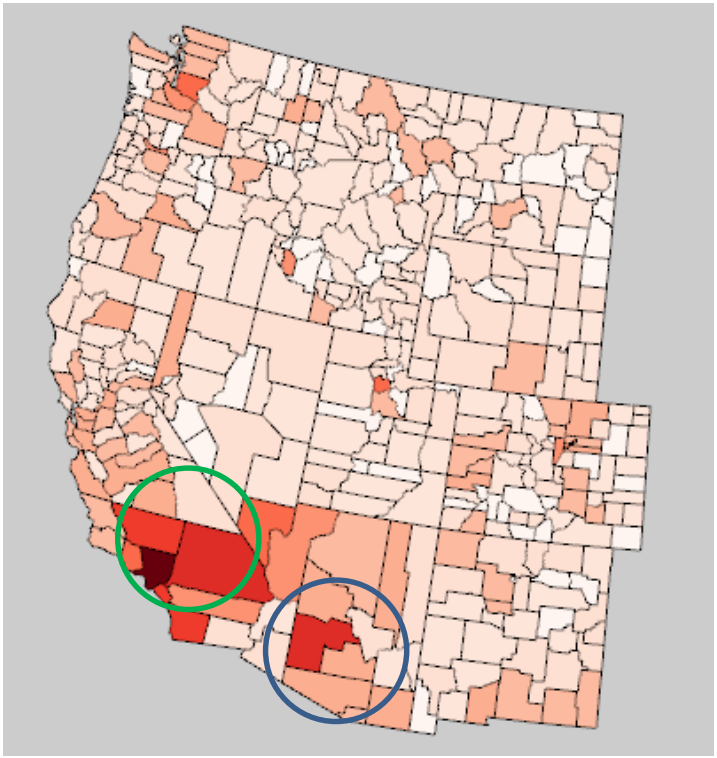
Characterize Space

1. Discrete spatial extent
2. Discrete geographical features
3. User-defined (Continuous) spatial extent
4. Continuous geographical resolution

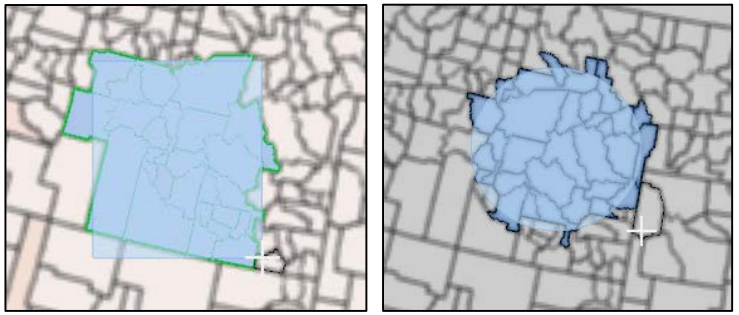
1. Zhang YF, Luo W, Mack AE, and Maciejewski R. (accepted) Visualizing the Impact of Geographical Variations on Multivariate Clustering. *Eurographics Conference on Visualization*. 2016.

Characterize Space

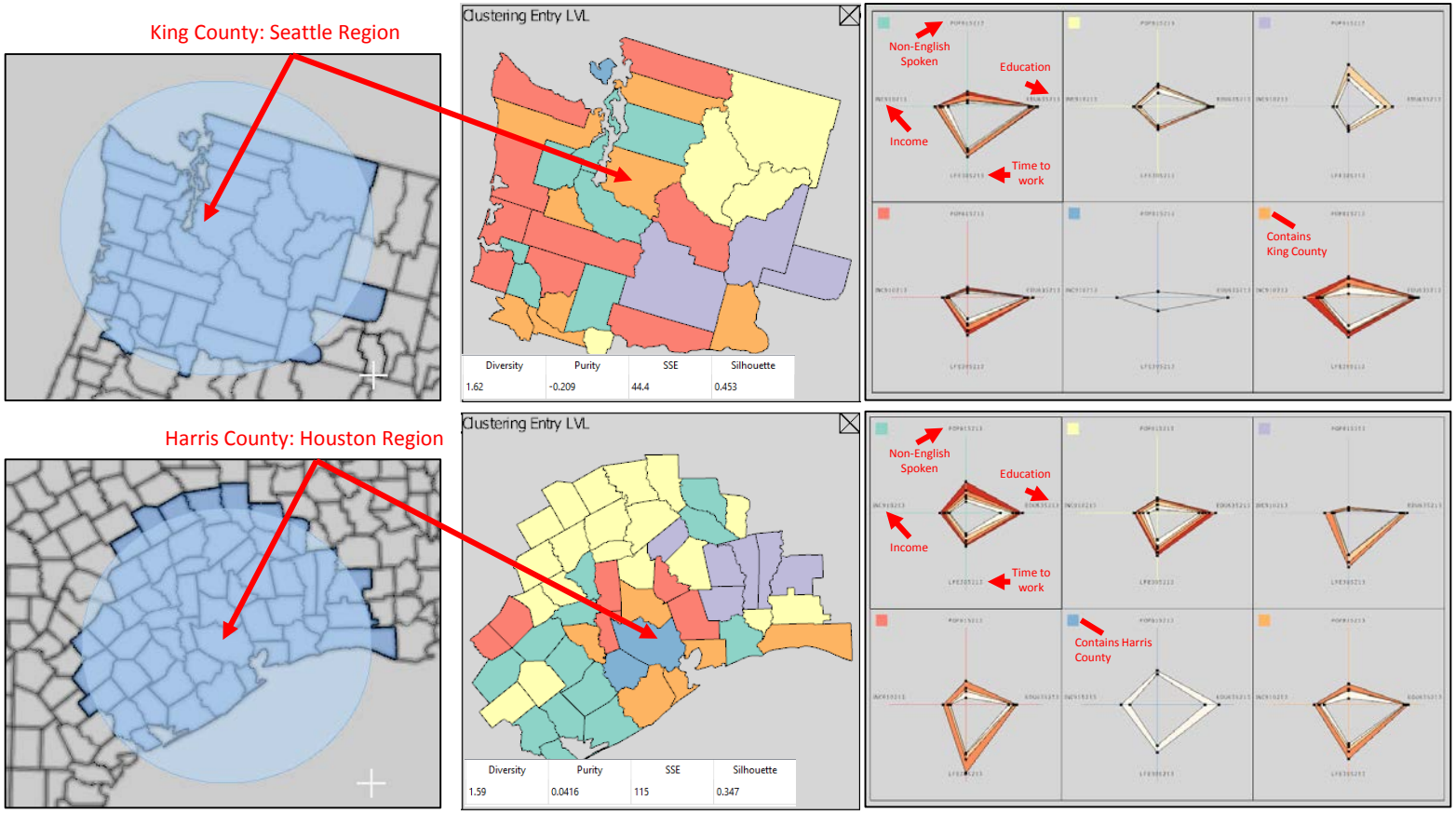
1. Discrete spatial extent
   – geographical selections at a constant spatial extent and apply clustering to the selected spatial extent



Rubber band selection in geographic space



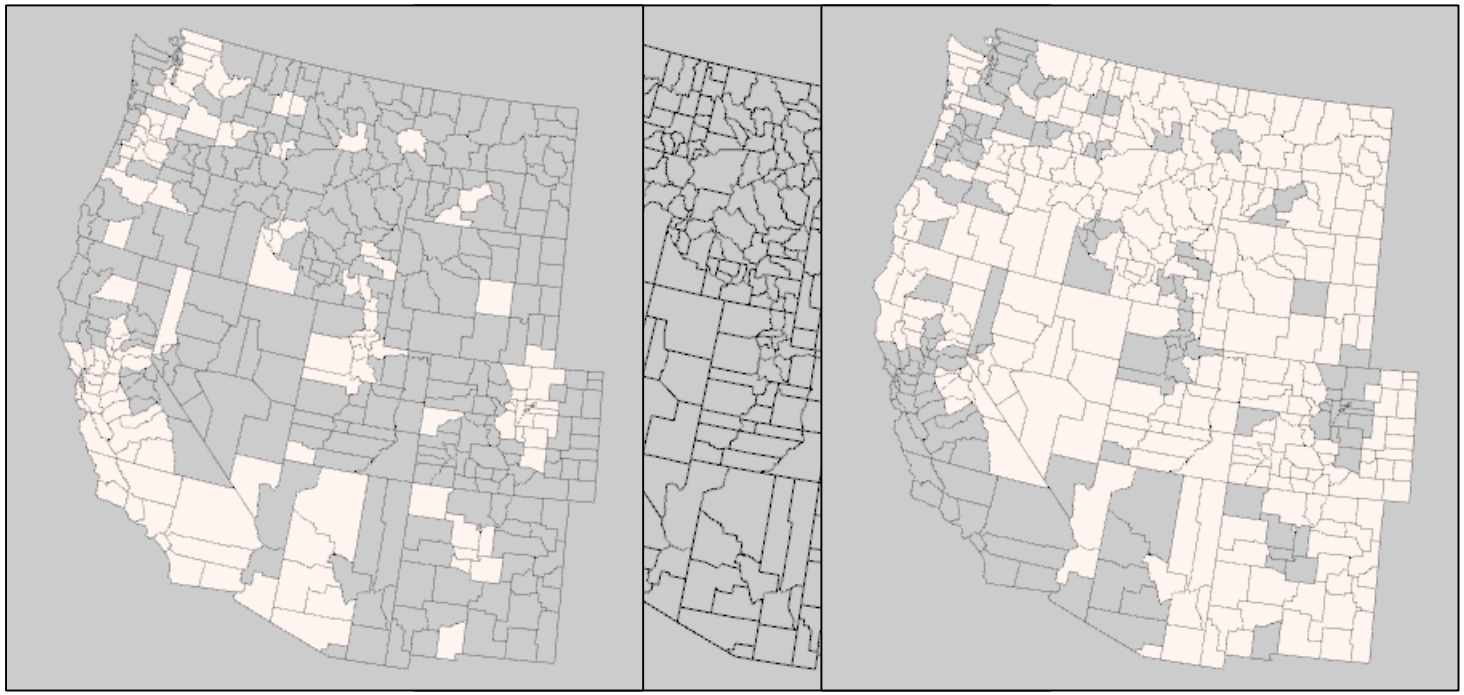
Validation Examples



Discrete spatial extent (King County and Harris County).
Demographics data for US counties: Non English spoken, Education above high school, Mean travel time to work, and Income level.

Characterize Space

2. Discrete geographical features
   – distinguish geographical features (e.g., urban vs. rural) and apply clustering to the geographical features of interest

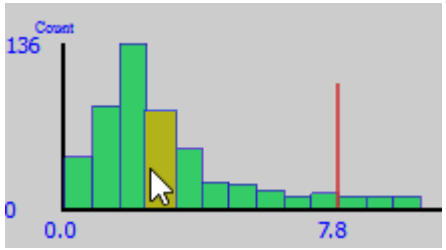


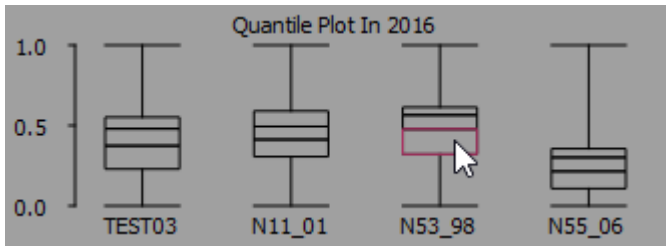
Metropolitan statistical area **Non metropolitan statistical area**

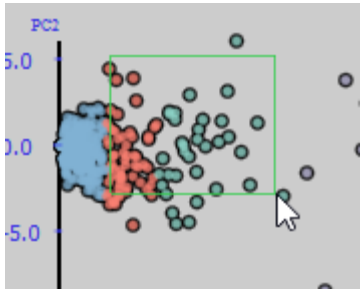
Characterize Space

2. Discrete geographical features
   – distinguish geographical features (e.g., urban vs. rural) and apply clustering to the geographical features of interest

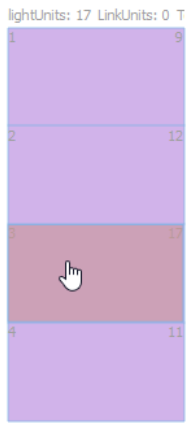
Selections from multivariate space



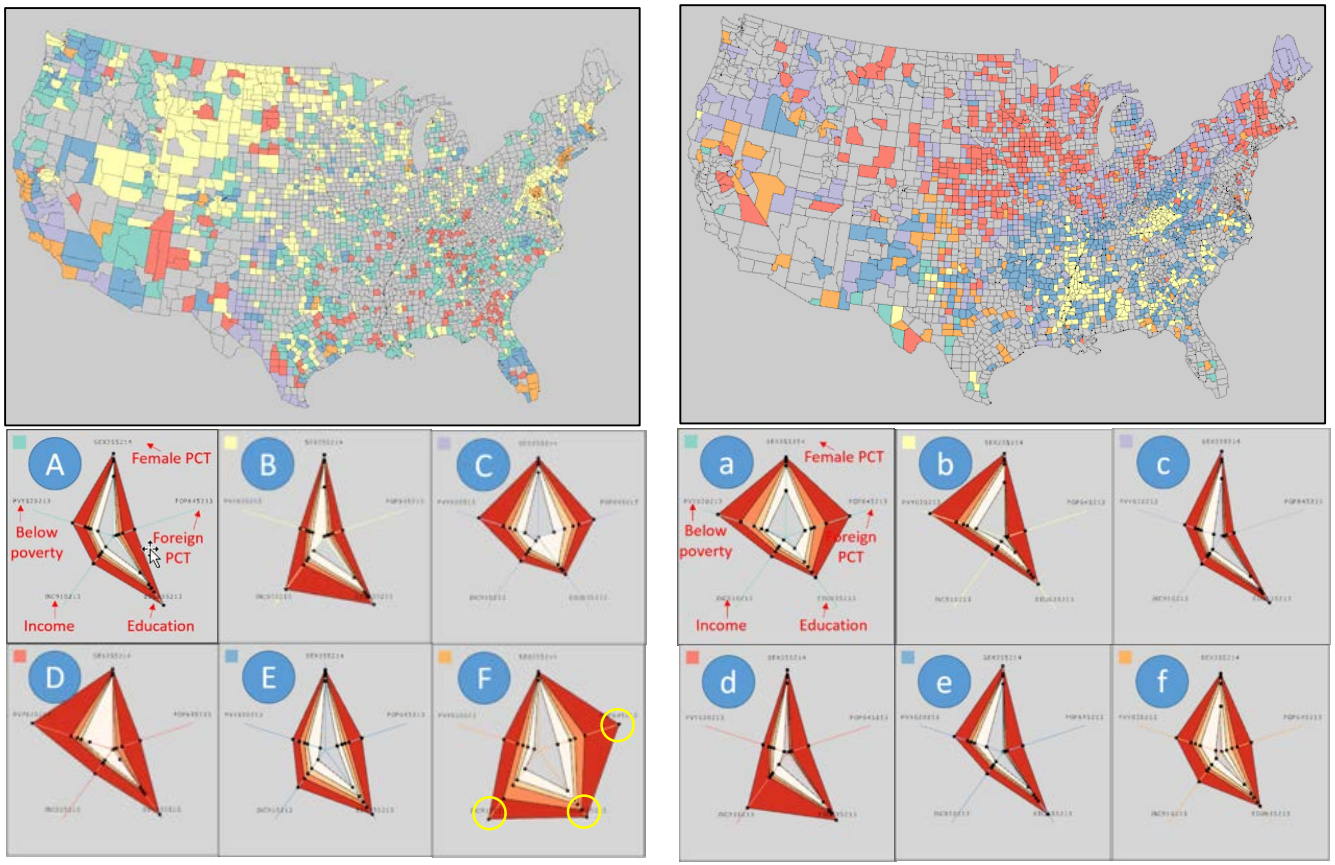




Variable Name	Type	Cat. Num
AREA	Ordinal	1
PERIMETER	Ordinal	1
ST99_D00_	Ordinal	1
ST99_D00_I	Ordinal	1
STATE	Ordinal	1
NAME	Nominal	1
LSAD	Ordinal	1
REGION	Nominal	4
DIVISION	Ordinal	1
LSAD_TRANS	Nominal	1



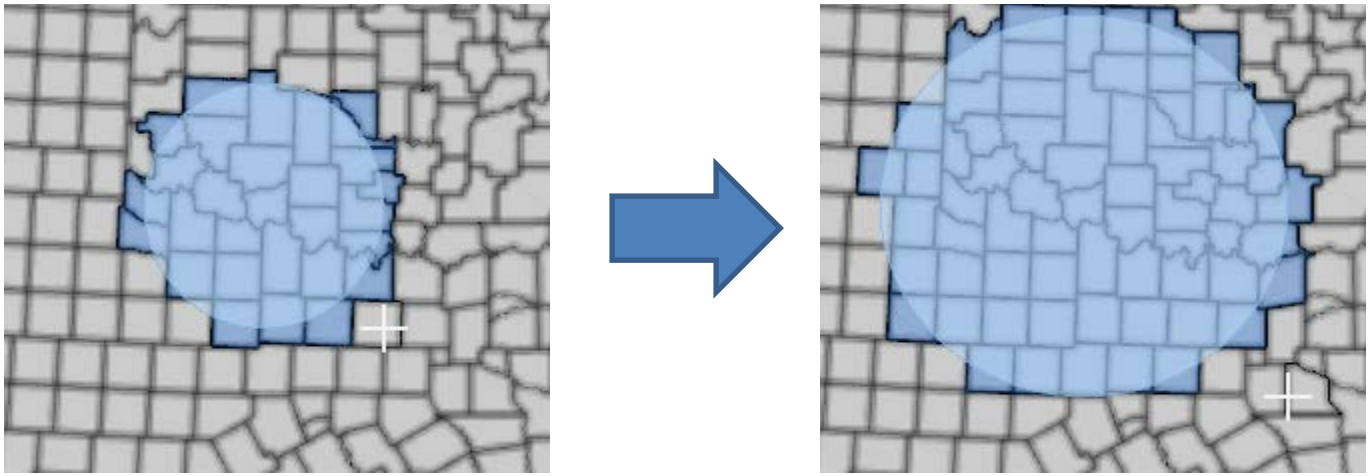
Validation Examples



Discrete geographical features (Positive VS. Negative population change). Demographics data for US counties: Female, Foreign born people, Education above high school, Income level, and below poverty.

Characterize Space

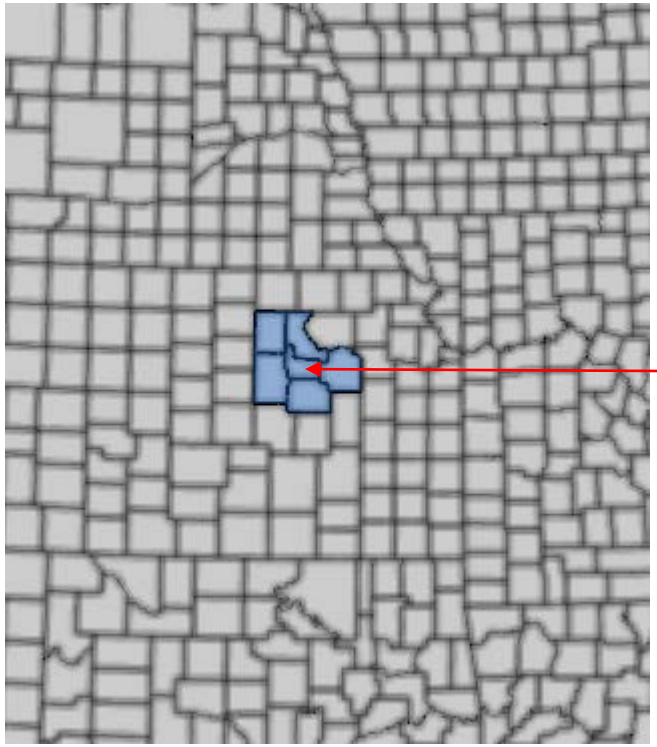
3. User-defined (Continuous) spatial extent
   – adjust spatial scales around a fixed location to understand the impact of the spatial dependence



Characterize Space

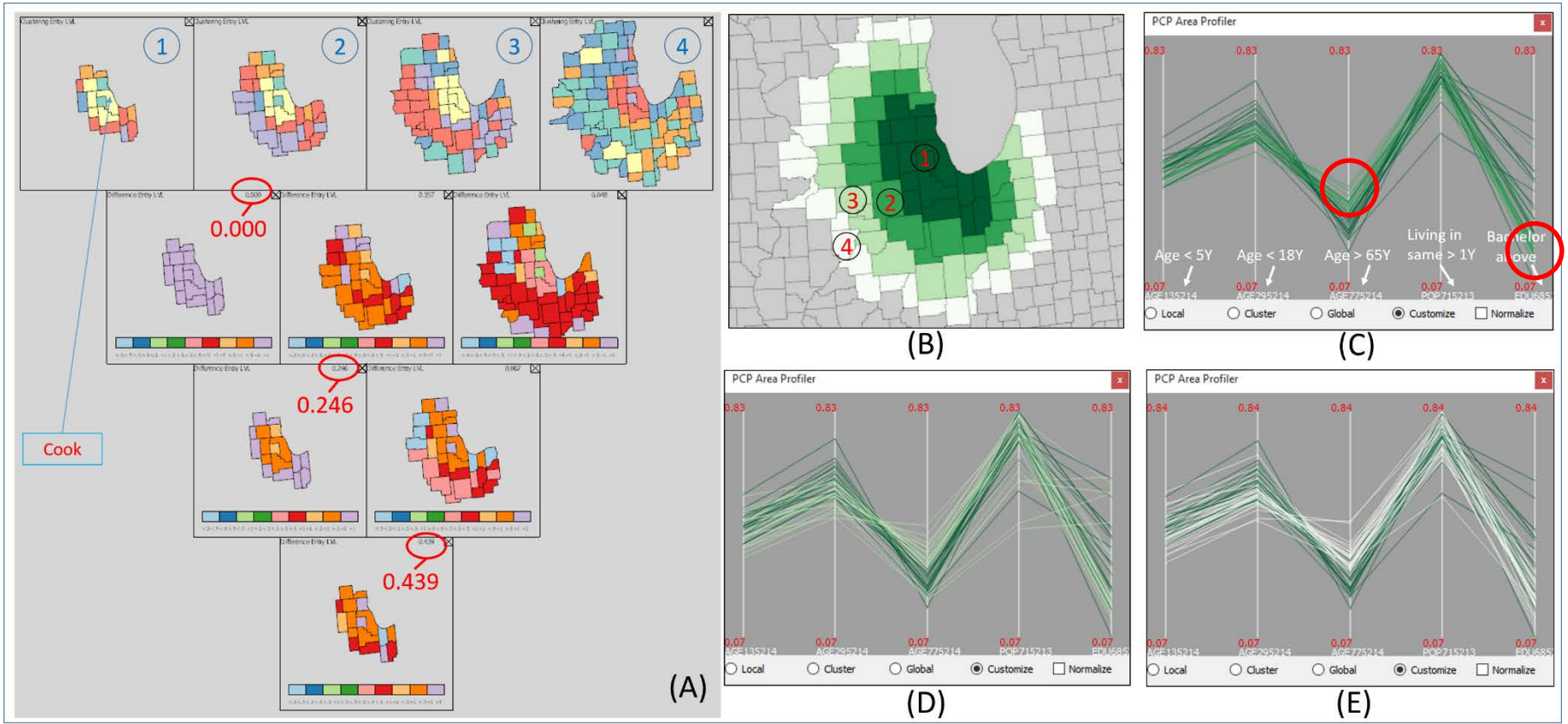
3. User-defined (Continuous) spatial extent

Automated geographical selections



Selecting the unit and set the growing rule: first neighbor

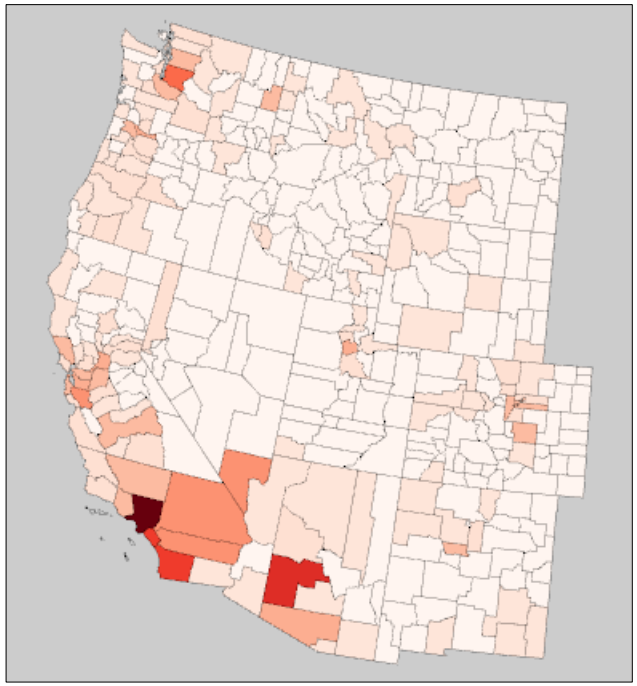
Validation Examples

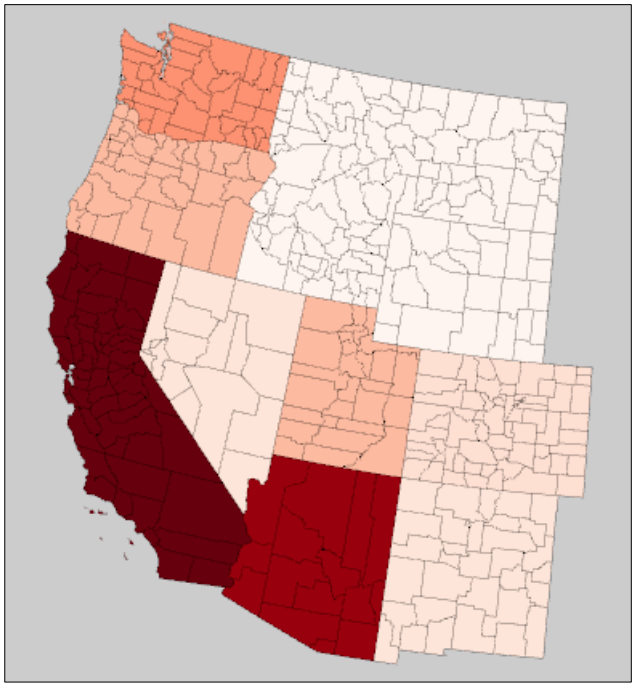


Continuous spatial extent: Demographics data for US counties: variables of age (percent of person under 5, under 18, and above 65 respectively), house living(percent of living in same house more than 1 year), and education.

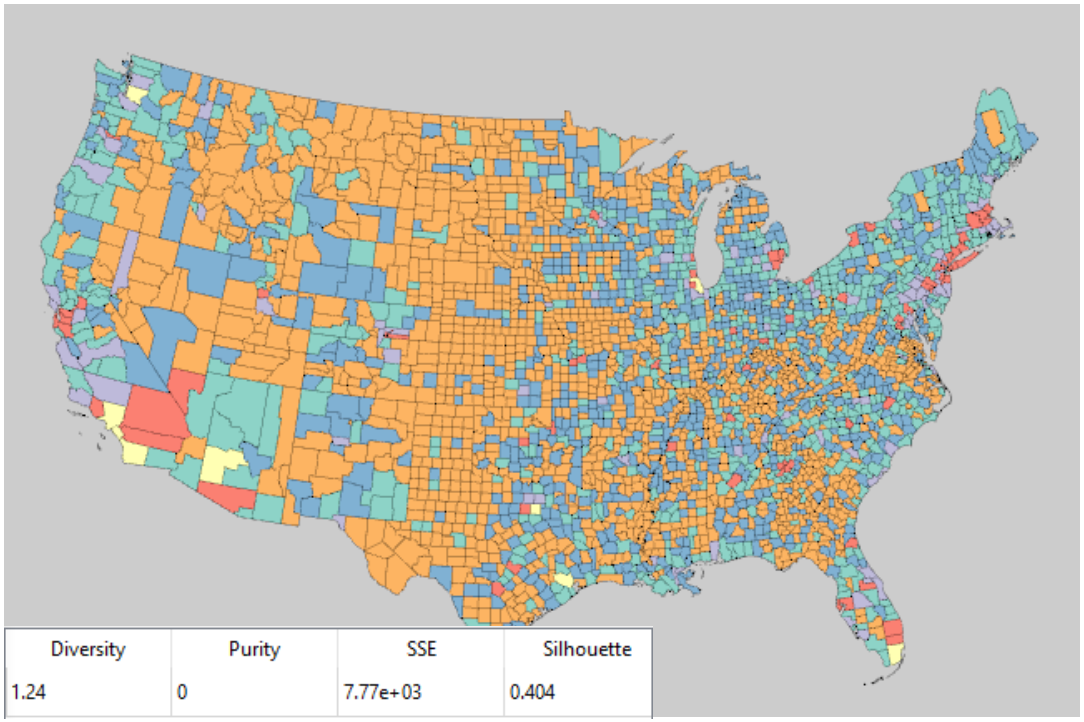
Characterize Space

4. Continuous geographic resolution
   – aggregate multivariate attributes at different spatial resolutions (e.g., city, county, state)

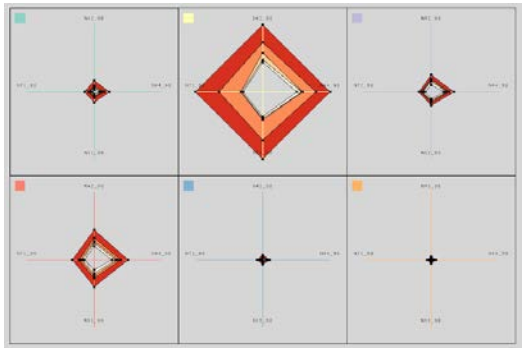


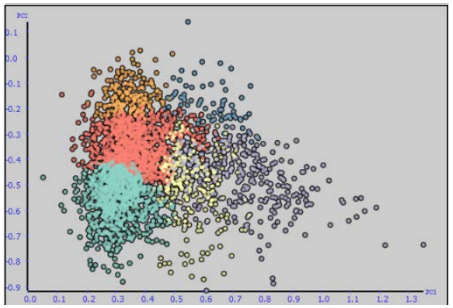


Validation Examples



Diversity	Purity	SSE	Silhouette
1.24	0	7.77e+03	0.404





Continuous geographical resolution (County to State).
Demographics data for US counties: Education above bachelor's degree, Veterans, Mean travel time to work, and Private nonfarm employment.

1. SUN M., WONG D. W.: Spatial aggregation as a means to improve data quality. *In Proceedings of the 13th International Conference on GeoComputation* (2015).

Conclusion & Future Work

- Better understanding of the multivariate map classification in different geographical space

  By allowing users to explore the impact of geographical variations across locations and scales for multivariate map classification.

- Highlight and identify statistically significant similarities and differences between clusters
- Extensions to the spatiotemporal domain
- Automated exploration on the impact of scale
  - by defining levels of aggregation and present a summary comparison to end users to suggest appropriate scales of analysis for the data.

Acknowledgments

- NSF under Grant No. 1350573
- U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-0610-CI0001
- Anonymous reviewers whose comments helped improve the manuscript.
- Thanks everyone for attending

- For details visit :*vader.lab.asu.edu*



VIS 2015

VAST * INFOVIS * SCIVIS

VAST Challenge 2015: Grand Challenge - Team VADER/VIS

Award for Outstanding Comprehensive Submission

Michael Steptoe[1], Robert Krueger[2], Yifan Zhang[1], Xing Liang[1], **Rolando Garcia[1]**, Sagarika Kadambi[1], Wei Luo[1], Thomas Ertl[2], Ross Maciejewski[1]

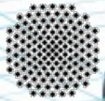


Visual Analytics & Data Exploration Research

ASU Arizona State University



Institute for Visualization and Interactive Systems



Universität Stuttgart

Challenges

- Raw trajectory data + POI data + communication data
- How to align and explore?

- Similar to last year's challenge but:
  - Larger data volumes (~11.000 visitors)

- **Semi-automatic approaches** become more important!
  - Exploration, **group/community & outlier detection**, hypothesis forming & validation

VIS 2015
VAST • INFOVIS • SCIVIS

Preprocessing

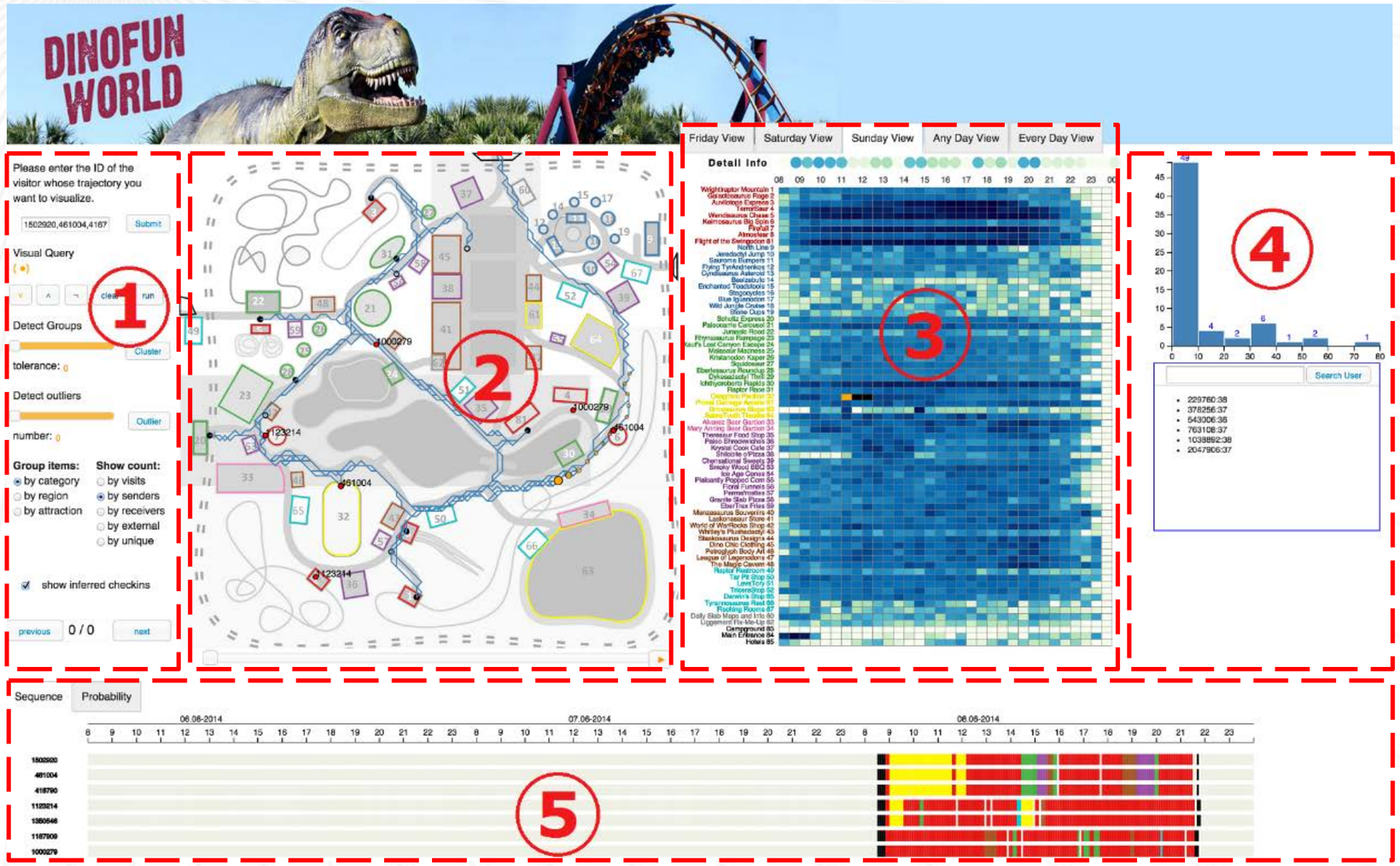
- Align trajectories with attractions
  - For rides: attraction visits are known
  - For other attractions: align movements with attraction positions (inferred)
  - Raw trajectories -> Semantic trajectories -> Attraction visit sequences



- Align attraction visits with communication
  - For each communication: retrieve the last movement for both sender and receiver before that communication timestamp. Then assign the attraction position if the movement is within a certain area of that attraction.

VIS 2015
VAST • INFOVIS • SCIVIS

System Components



1. Analytics Interface
2. Map View
3. Calendar View
4. Distribution View
5. Sequence View

VIS 2015
VAST • INFOVIS • SCIVIS

Analytics Interface

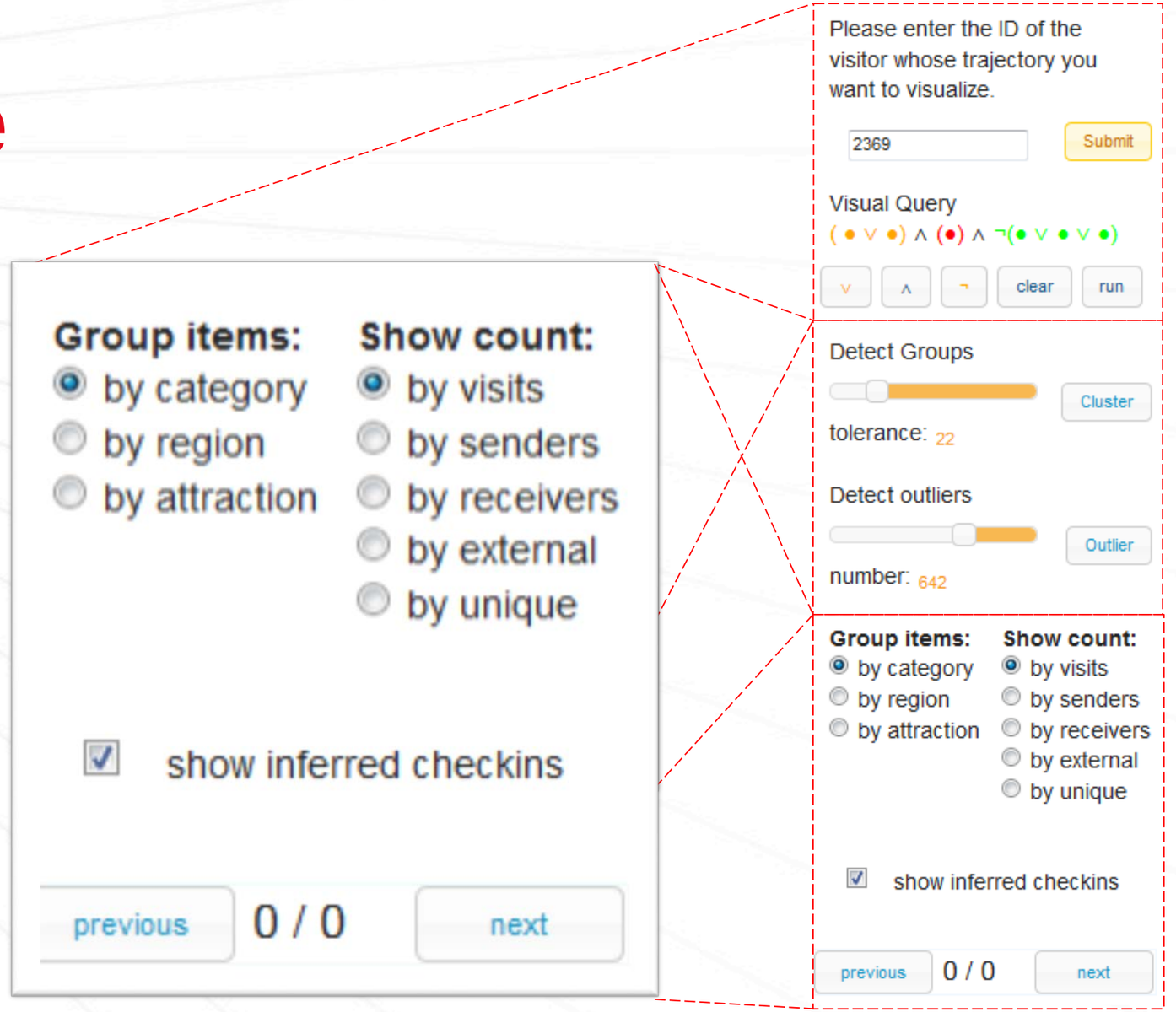
A. ID Selection: inspect a single visitor

B. Visual Query: create queries about visited attractions to create ,validate and refine hypotheses *(Boolean operations)*

C. Cluster: find visitor groups travelling together *(clustering with Levenshtein distance)*

D. Outliers: find visitors with outlying attraction visit behavior *(rank rows in distance matrix)*

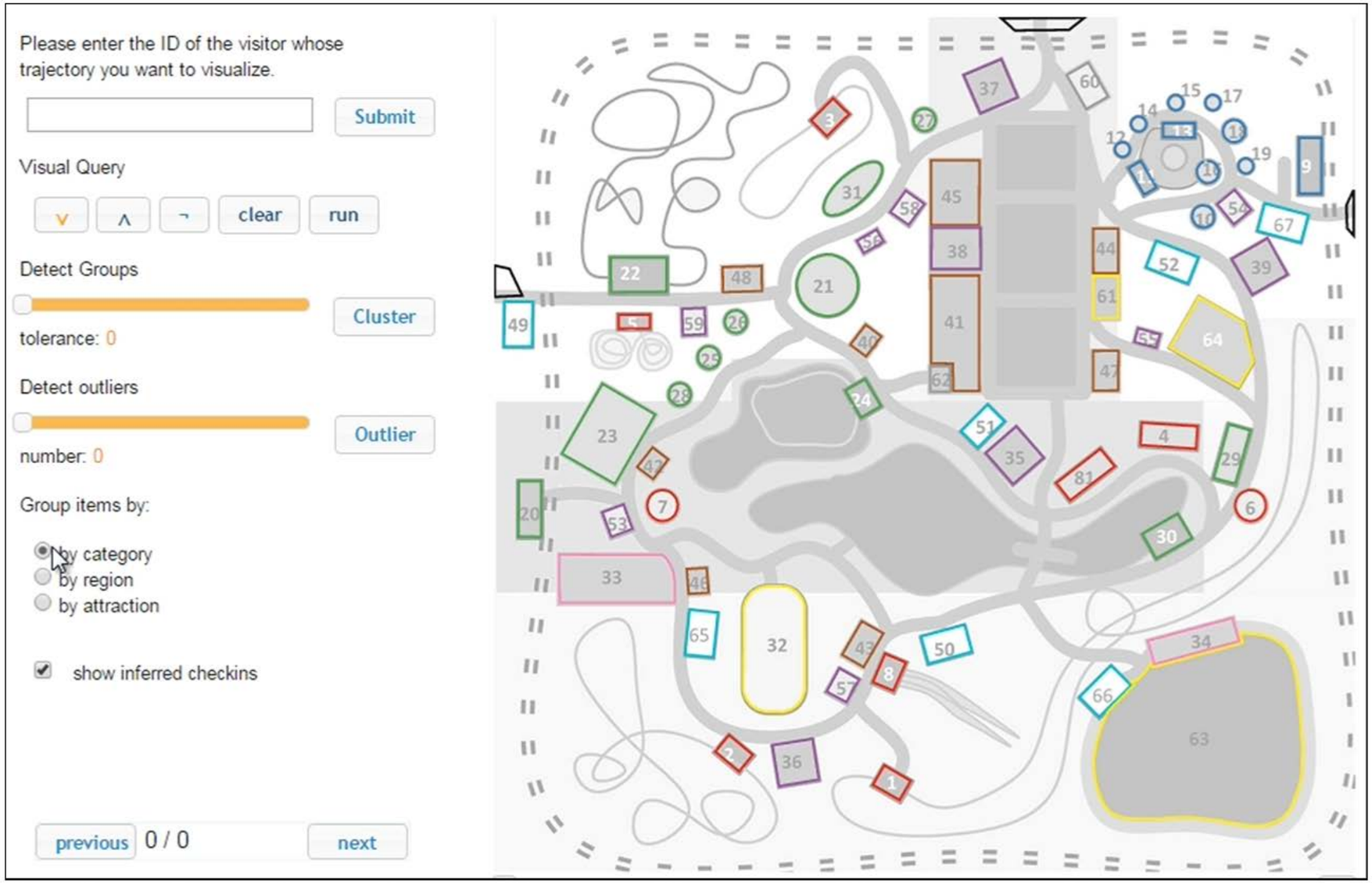
E. Grouping: order and color coding, select statistics to show (visits/communication)



VIS2015
VAST • INFOVIS • SCIVIS

Map View

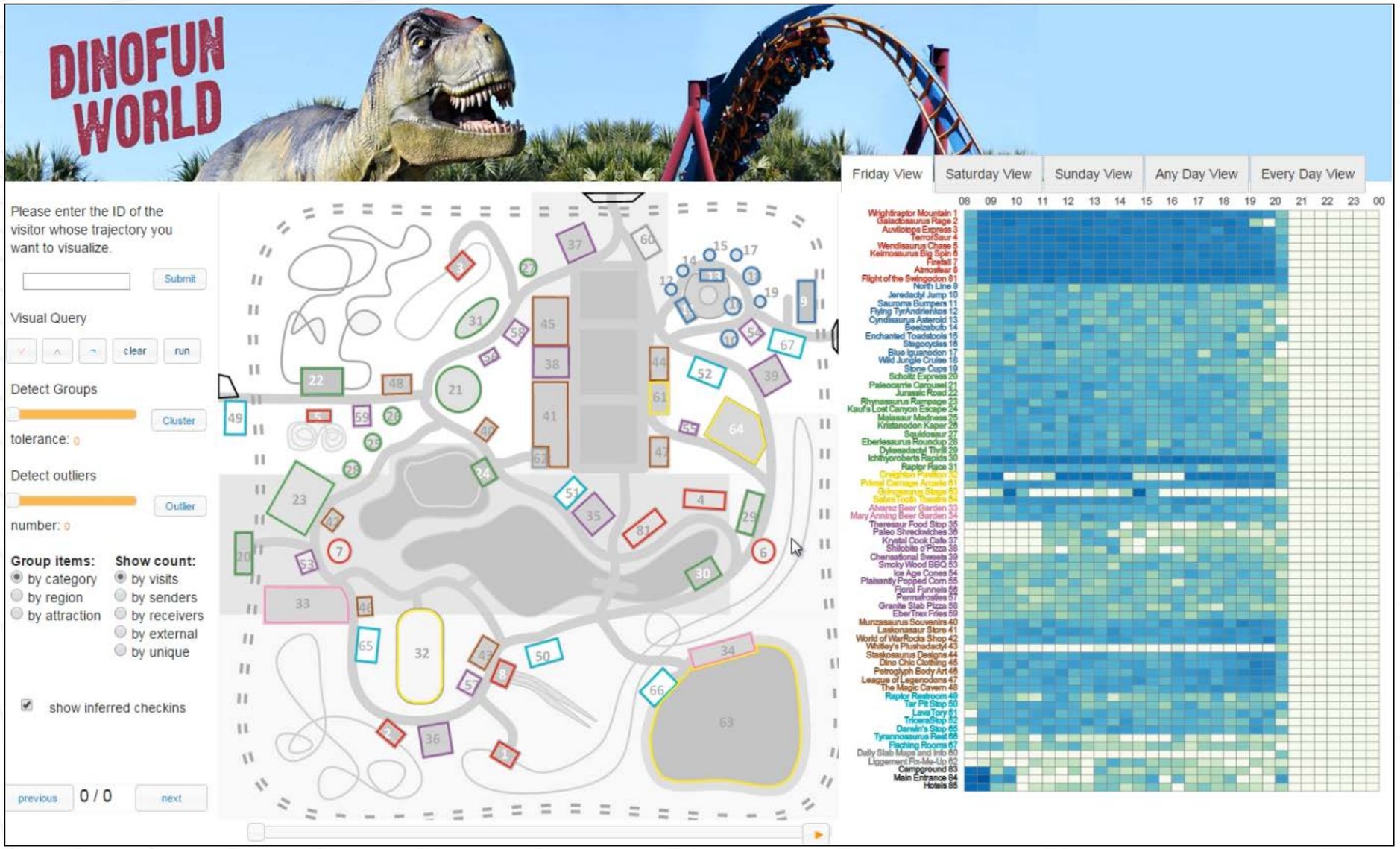
- Trajectories & movement animation
- Bivariate heatmap
- Footprints



VIS 2015
VAST • INFOVIS • SCIVIS

Calendar View

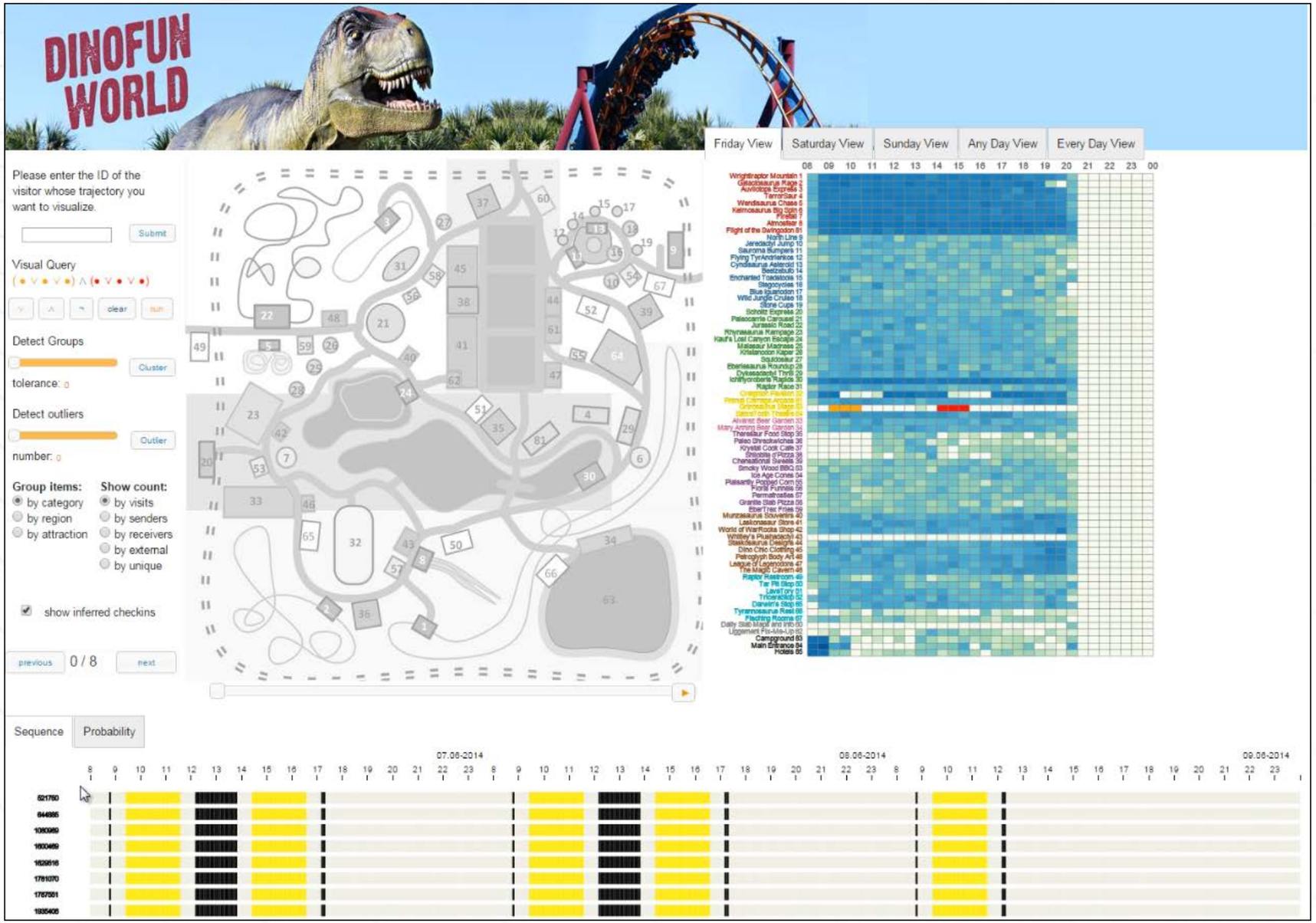
- Temporal attraction visit & communication overview
- Visual query



VIS 2015
VAST • INFOVIS • SCIVIS

Sequence View

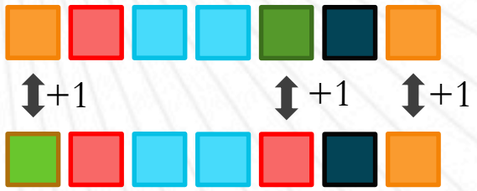
- Temporal attraction visit sequences
- Groups of similar visit behavior
- Outlying visitor behavior



VIS 2015
VAST • INFOVIS • SCIVIS

Sequence View - *Clustering and Outlier Detection*

- Levenshtein distance

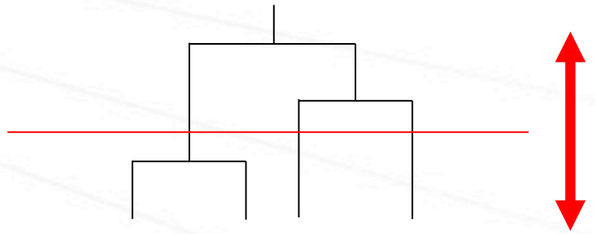


1. Replace
2. Replace
3. Insert

- Distance matrix *~11.000 x ~11.000*

	0	1	2	3	4	5	…
0							
1							
2							
3							
4							
5							
…							

- Agglomerative clustering



- Outlier detection
  - Rank by sum of distances

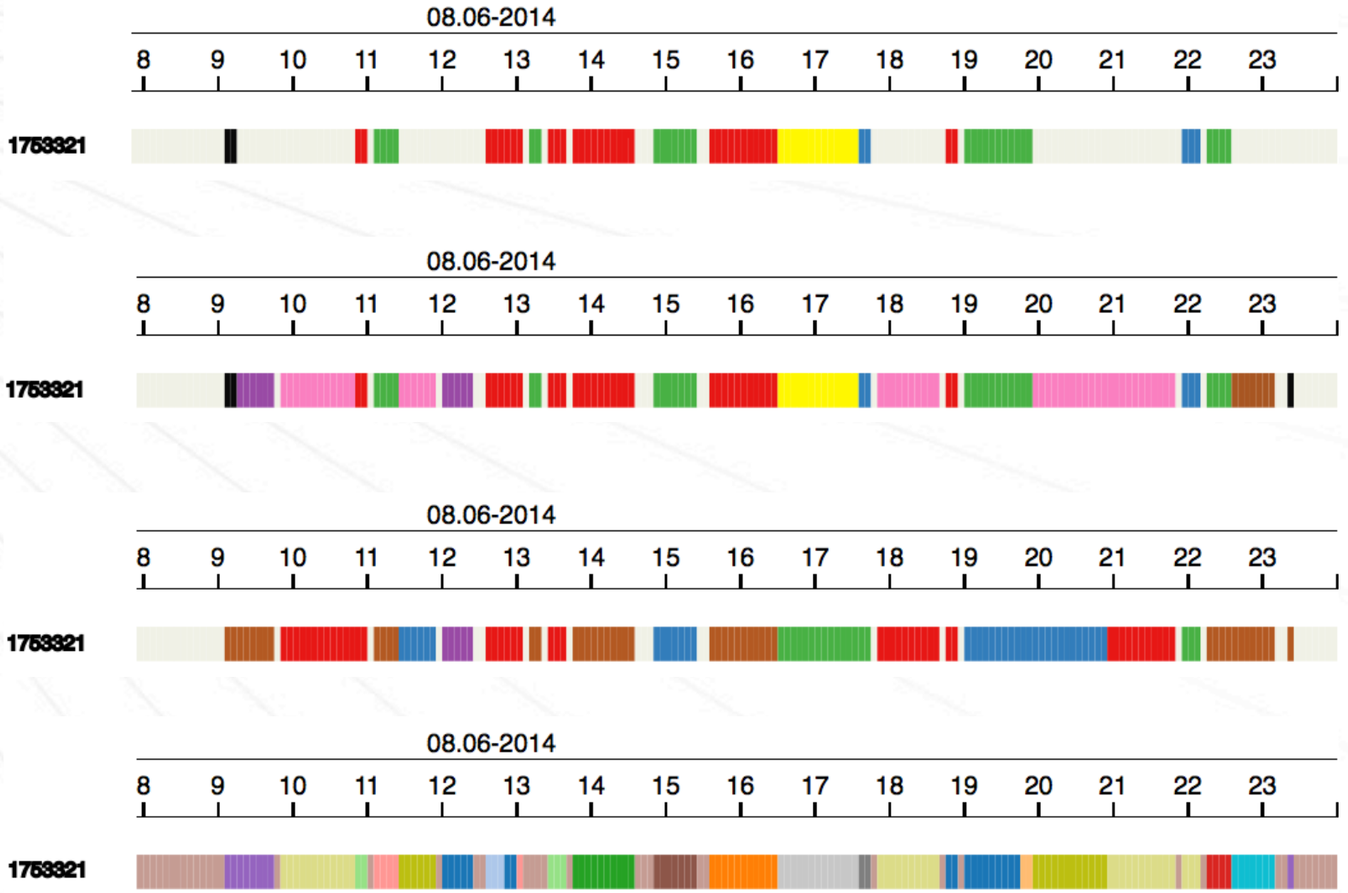
X	22	10	3	16	17

= 68

VIS 2015
VAST • INFOVIS • SCIVIS

Sequence View

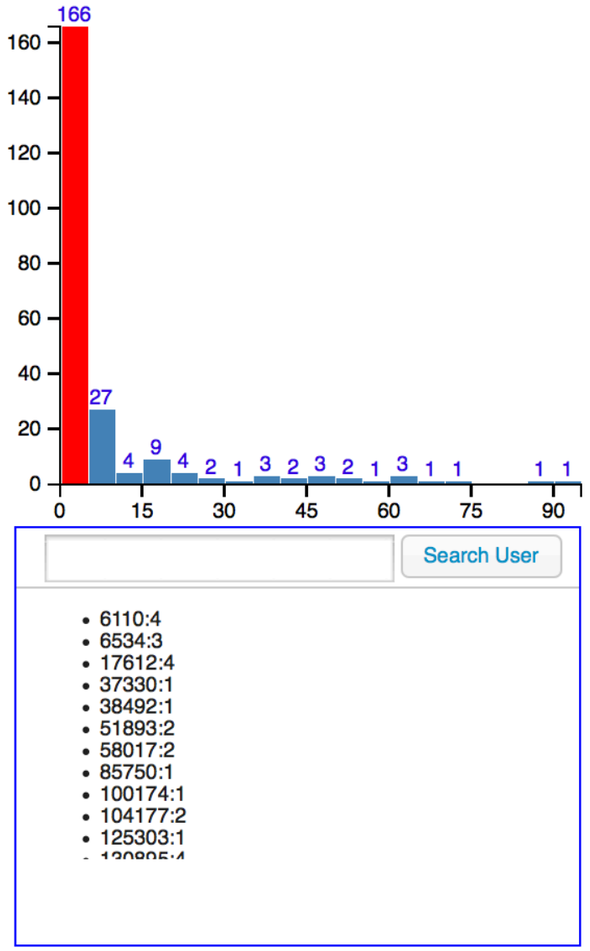
- Category without inferred check-ins
- Category with inferred check-ins
- Region with inferred check-ins
- Attraction with inferred check-ins



VIS 2015
VAST • INFOVIS • SCIVIS

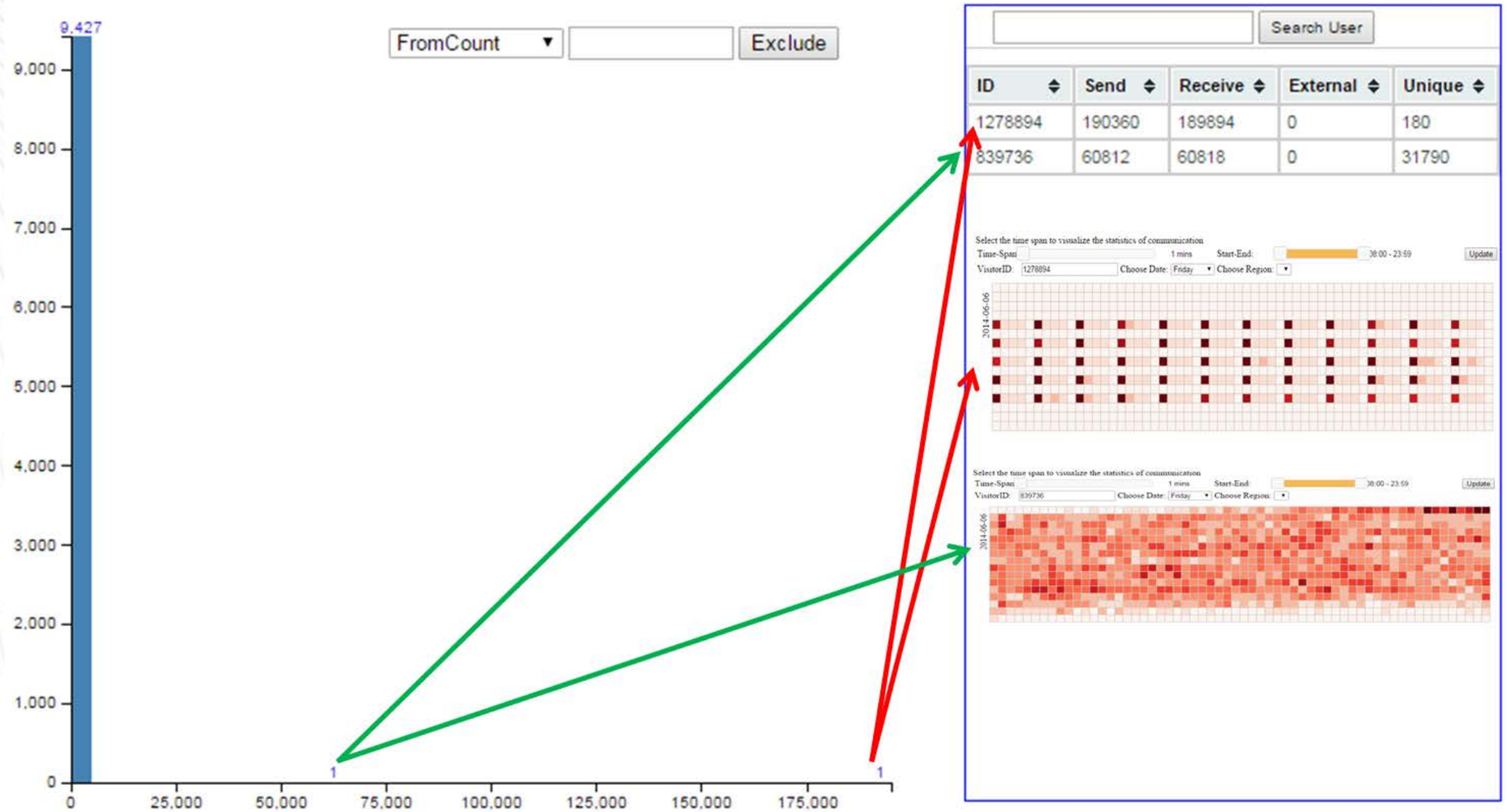
Distribution View

- Call distribution for selected attraction and time.
  - X-axis: Number of communications
  - Y-axis: Number of IDs
- Interactive bins
  - Shows all IDs within the selected bin and the amount they communicated.



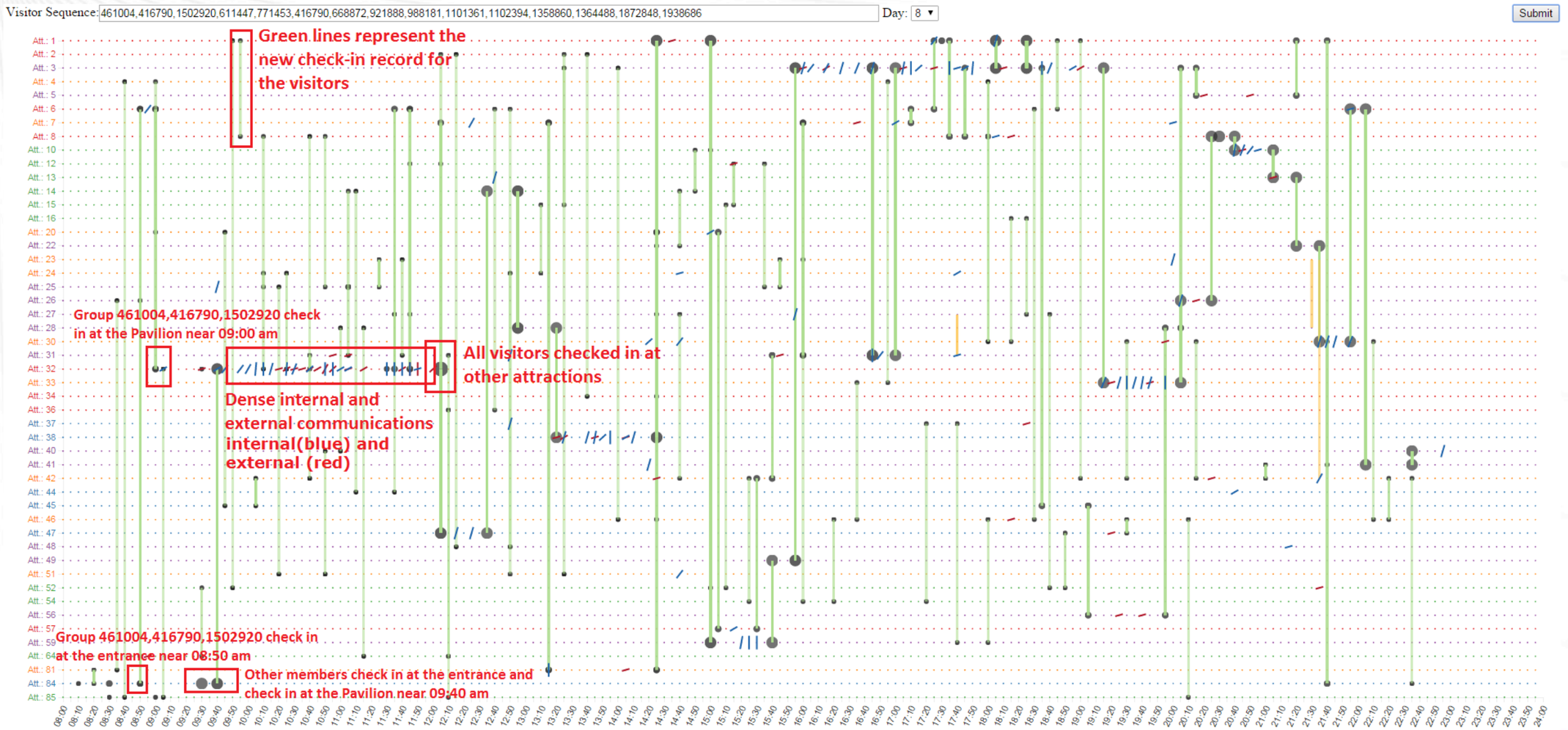
VIS2015
VAST • INFOVIS • SCIVIS

Communication Visualization



VIS 2015
VAST • INFOVIS • SCIVIS

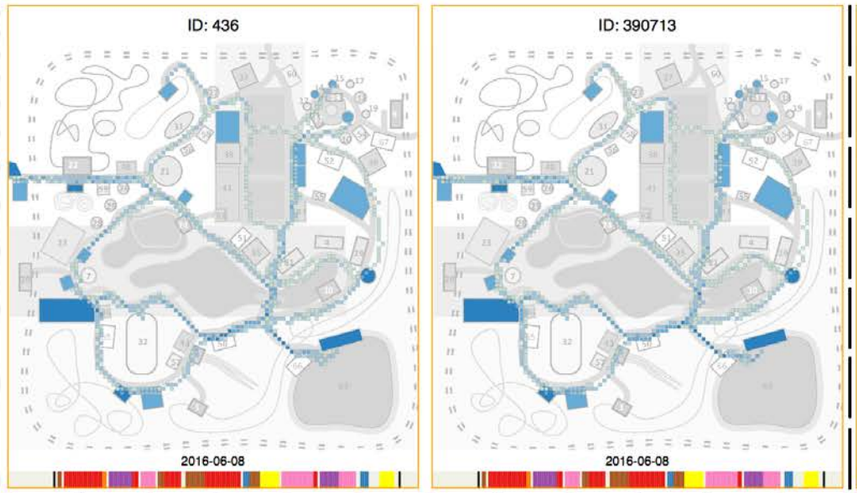
Network Analysis Visualization

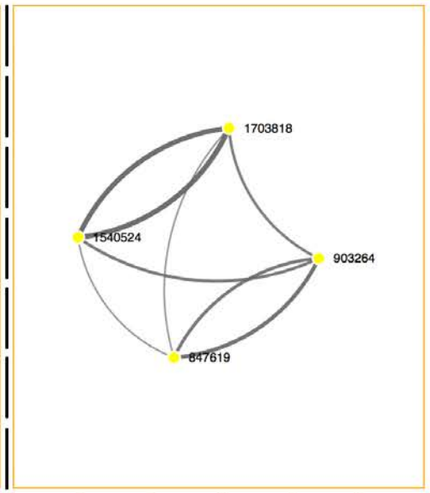


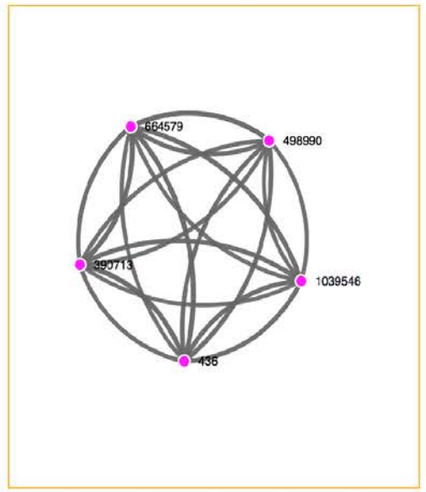
VIS 2015
VAST • INFOVIS • SCIVIS

Reporting Visualization

Movement View

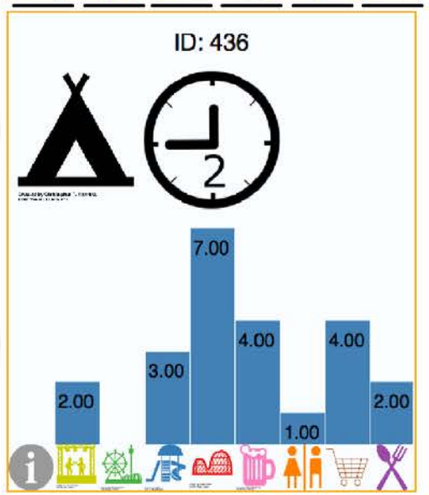


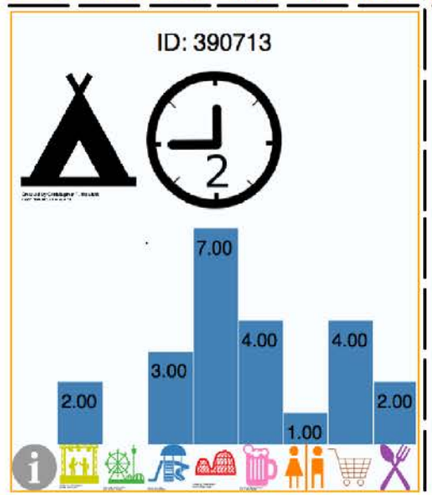


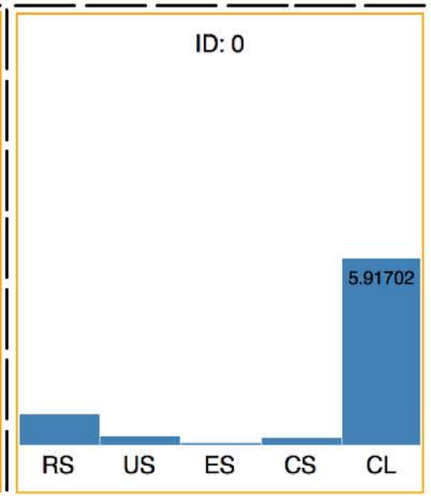


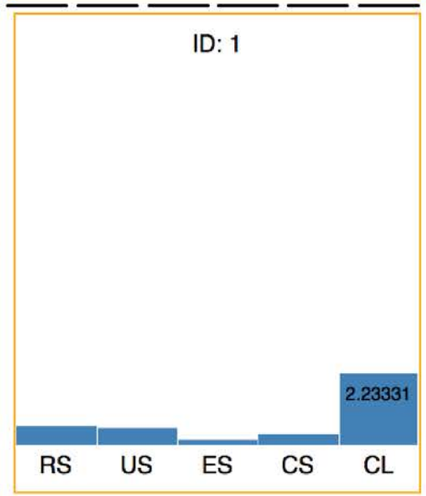
Network View

Feature View









Communication View

VIS2015
VAST • INFOVIS • SCIVIS

Movement View

Please select the view type.

Visitor | Group | Group Type

Please enter the ID(s) of the visitor you want to visualize.

941 ×

Submit

Legend

μ

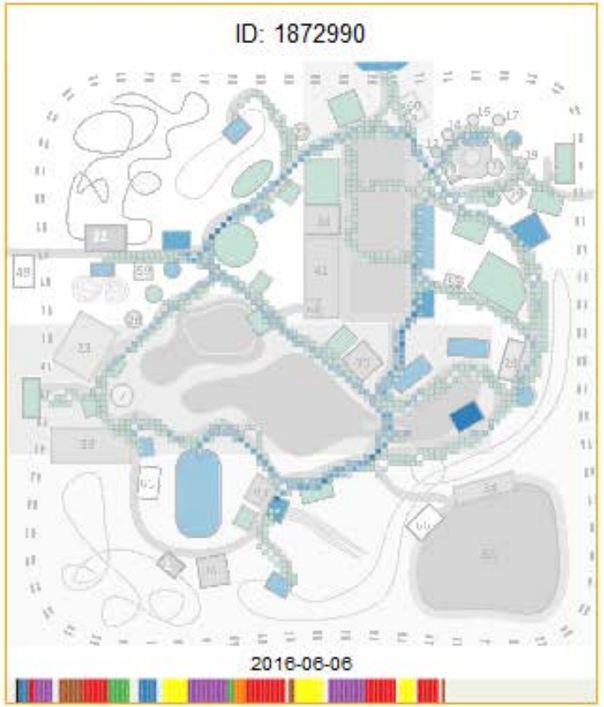
σ

Color sequence by:

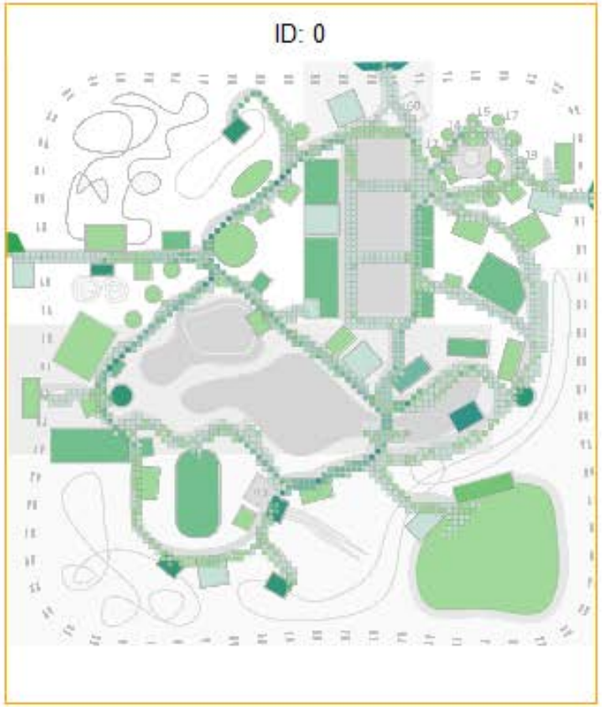
Category | Region | Attraction

Please select the sequence date to be displayed.

Friday | Saturday | Sunday | All







VIS 2015
VAST • INFOVIS • SCIVIS

Feature View

Please select the view type.

Visitor | Group | Group Type

Please enter the ID(s) of the visitor you want to visualize.

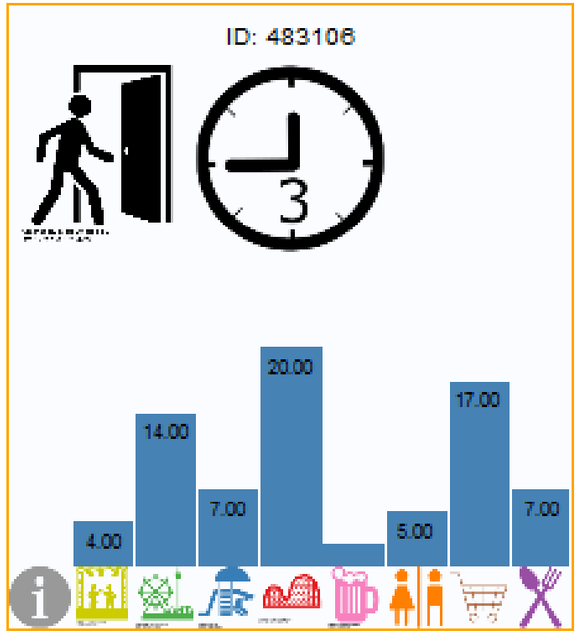
483106 ×

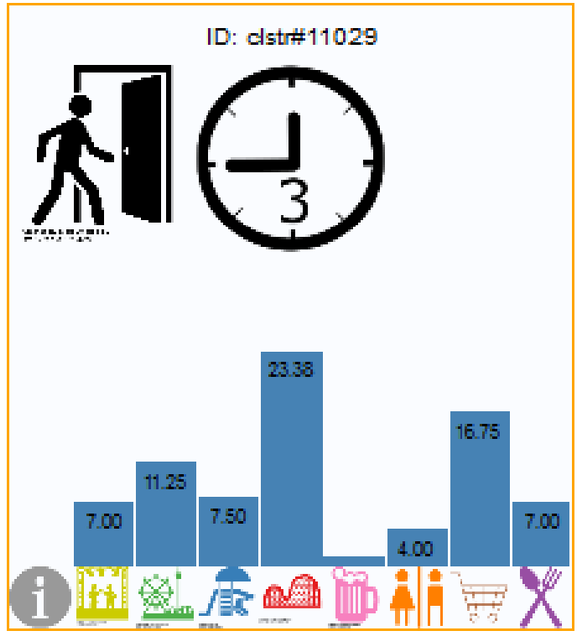
Submit

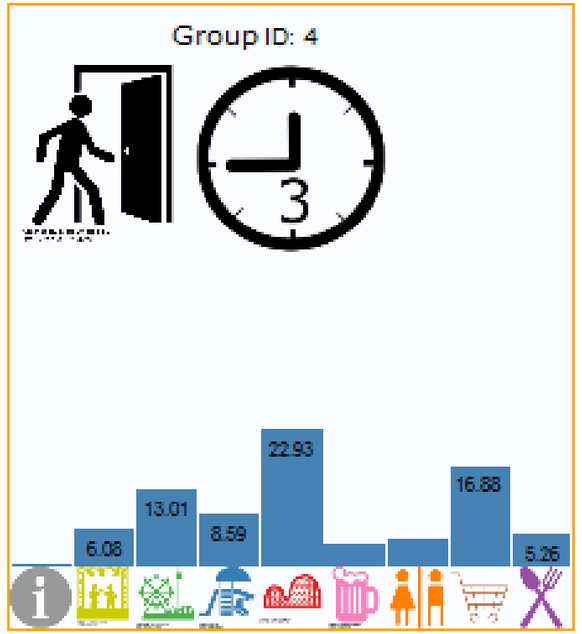
Load All IDs

Filter by features:

Submit







VIS 2015
VAST • INFOVIS • SCIVIS

Network View

Please enter the ID(s) of the group you want to visualize.

clstr#3878 × clstr#5519 ×

Submit

Please enter the ID(s) of the visitor you want to visualize.

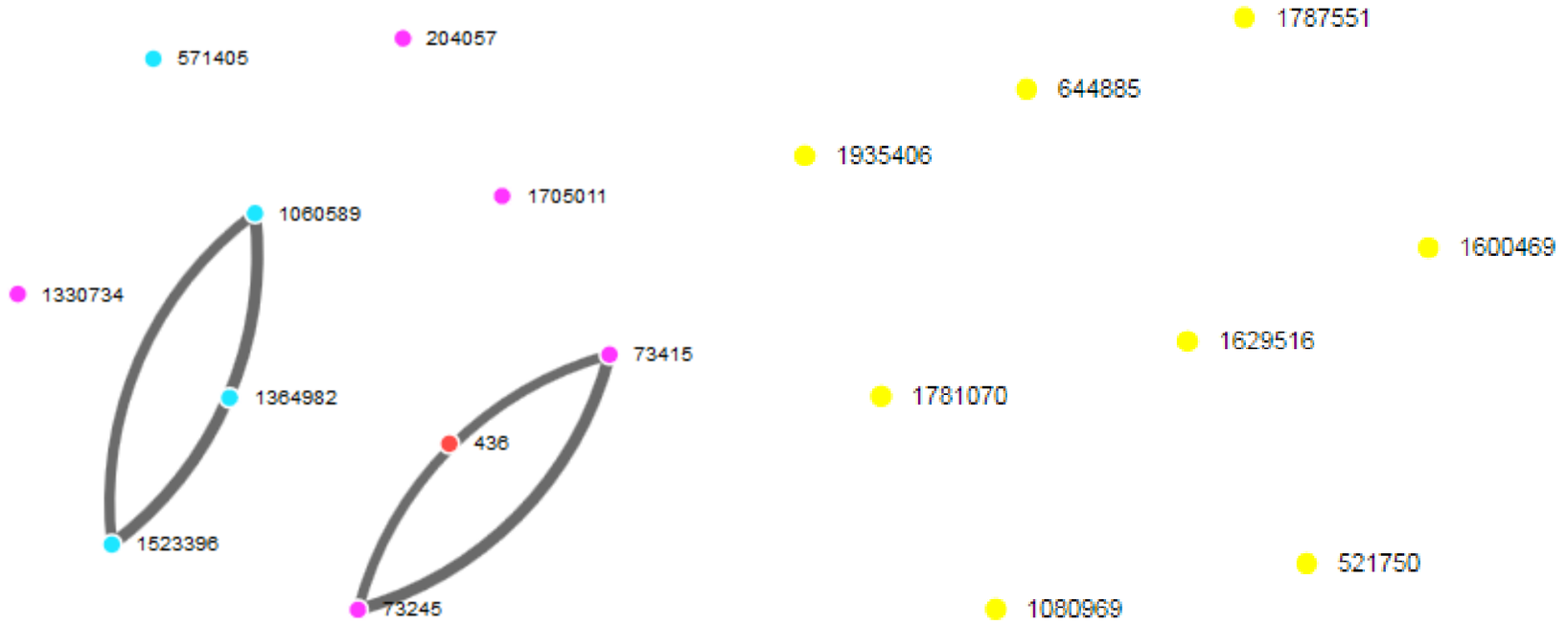
436 ×

Add

Please enter the ID(s) of the group you want to add to the visualize.

Add

Clear All



VIS 2015
VAST • INFOVIS • SCIVIS

Communication View

Number of group types:

10

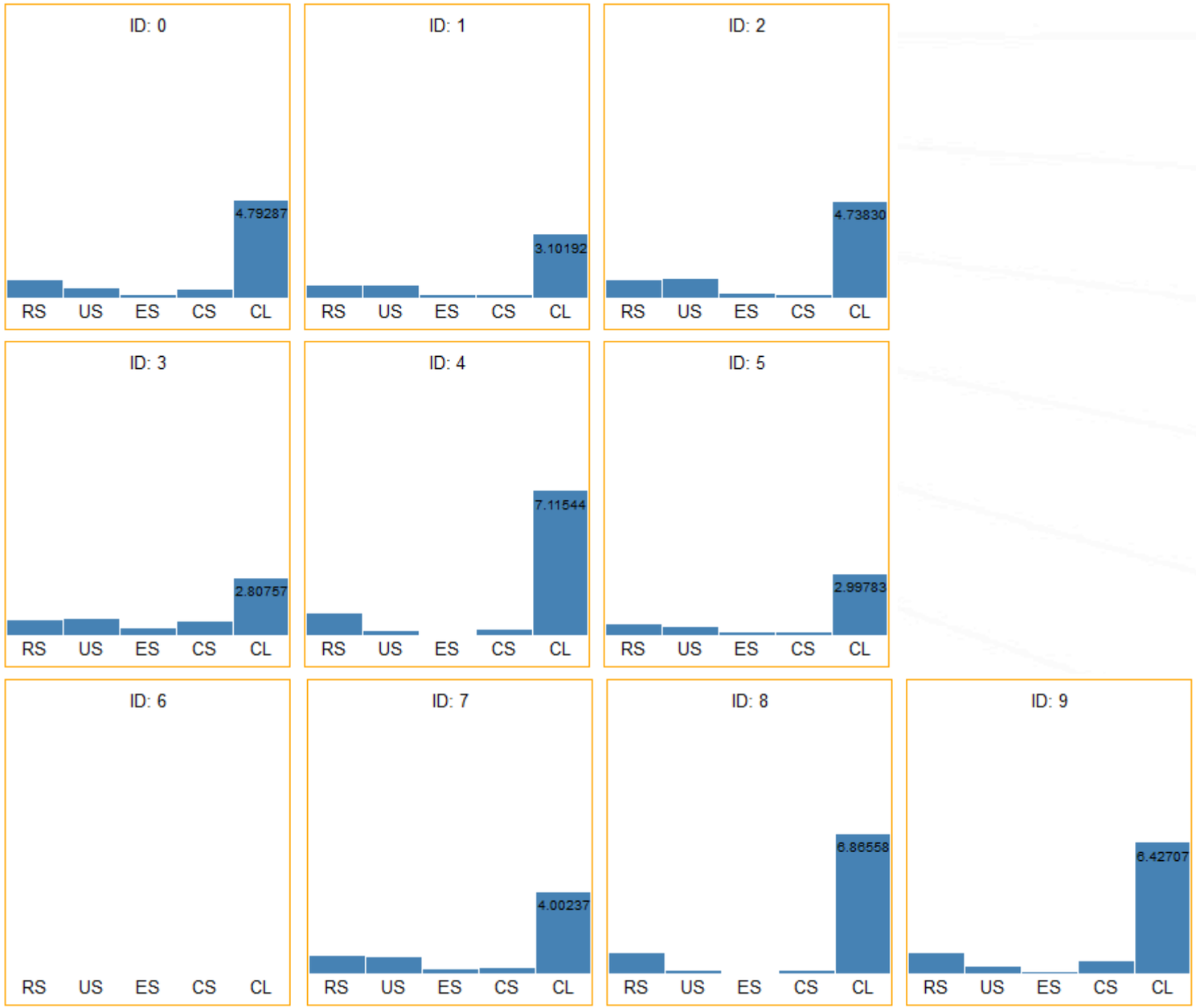
Cluster by features:

RS US
ES CS
CL

Submit

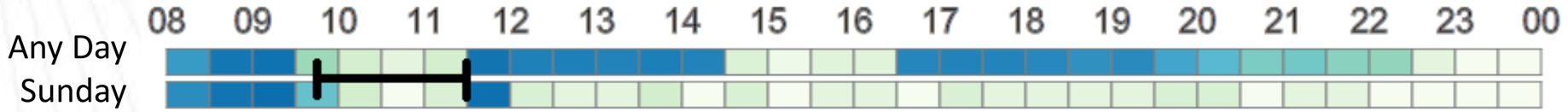
Return

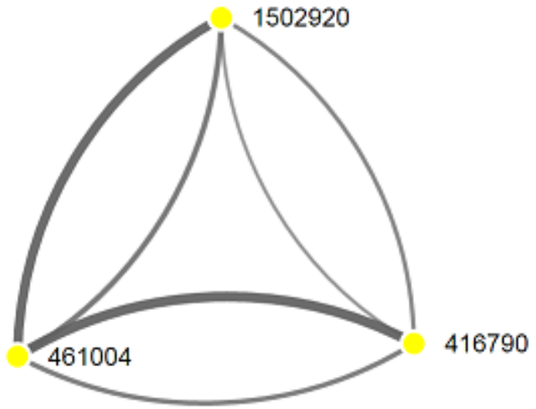
Clear All



Time of Crime

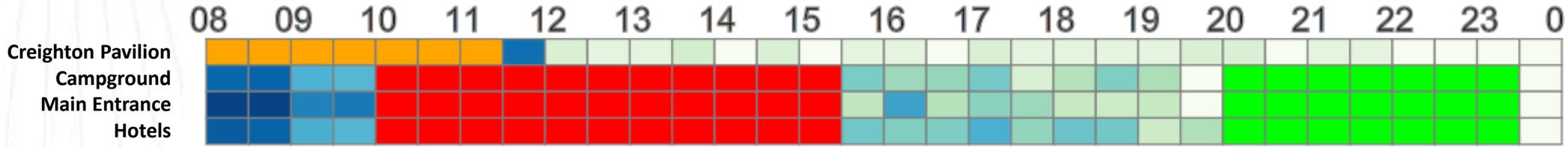
- Sunday between 9:45 AM and 11:30 AM
- Staff-members in the pavilion start external communications at 10:20 AM
- Public returns to the pavilion after Scott's performance at 11:30 AM





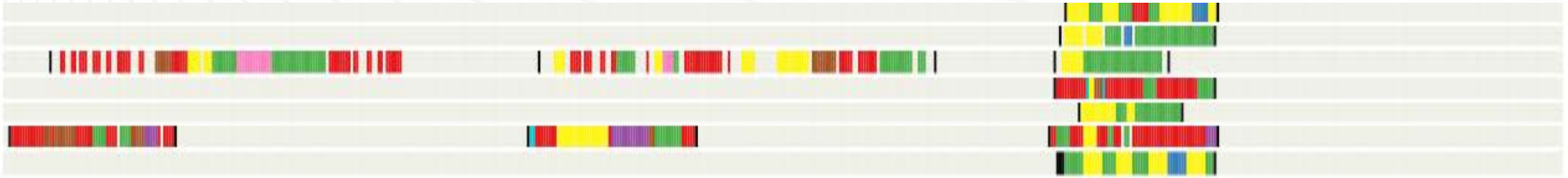
Filtering Suspects

- People who were at the Pavilion before 11:30 AM
- Exited the park between 10:00 AM and 3:00 PM
- Did not return to the park



Investigating Suspects

- 39 visitors fit the criteria
- Of the 39, only these 7 visitors moved alone:

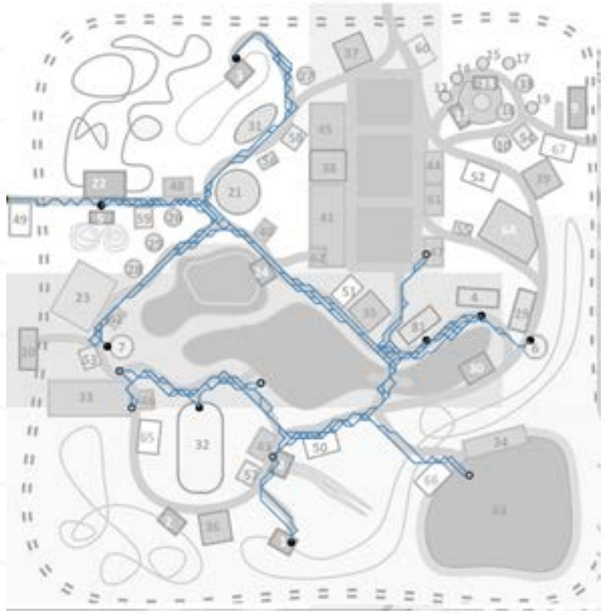


- We used our tool to explore the trajectories of each of these 7 visitors
- We selected the trajectory below as the culprit because of its behavior

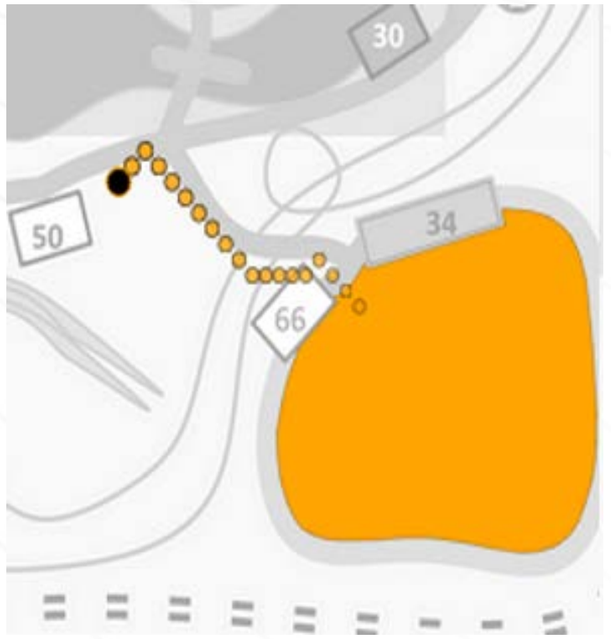


Culprit's Behavior Friday

- Thrill Rides
- Shopping
- Creighton Pavilion
- Grinosaurus Stage
- Ichthyoroberts Rapids
- Alvarez Beer Garden





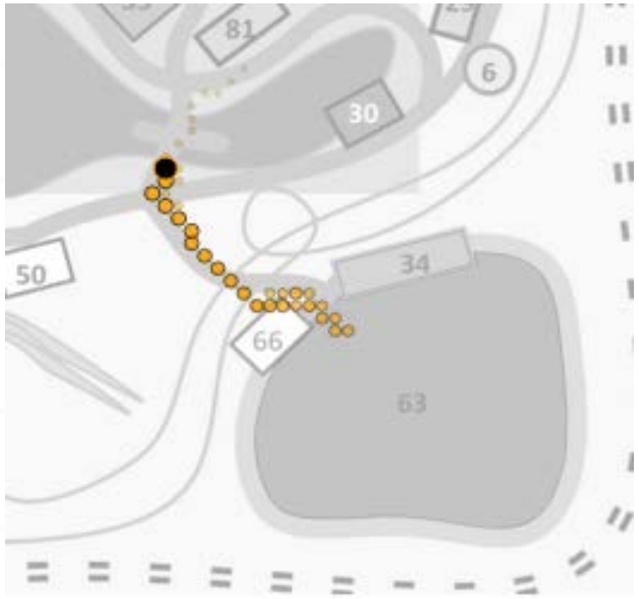




Culprit's Behavior Saturday

- Creighton Pavilion
- Thrill Rides & Grinosaurus Stage
- Scholtz Express
- Creighton Pavilion
- Alvarez Beer Garden
- Grinosaurus Stage
- Creighton Pavilion
- Shopping
- Rhynasaurus Rampage
- Dykesadactyl Thrill









Culprit's Behavior Saturday

- Creighton Pavilion
- Thrill Rides & Grinosaurus Stage
- Scholtz Express
- Creighton Pavilion
- Alvarez Beer Garden
- Grinosaurus Stage
- Creighton Pavilion
- Shopping
- Rhynasaurus Rampage
- Dykesadactyl Thrill



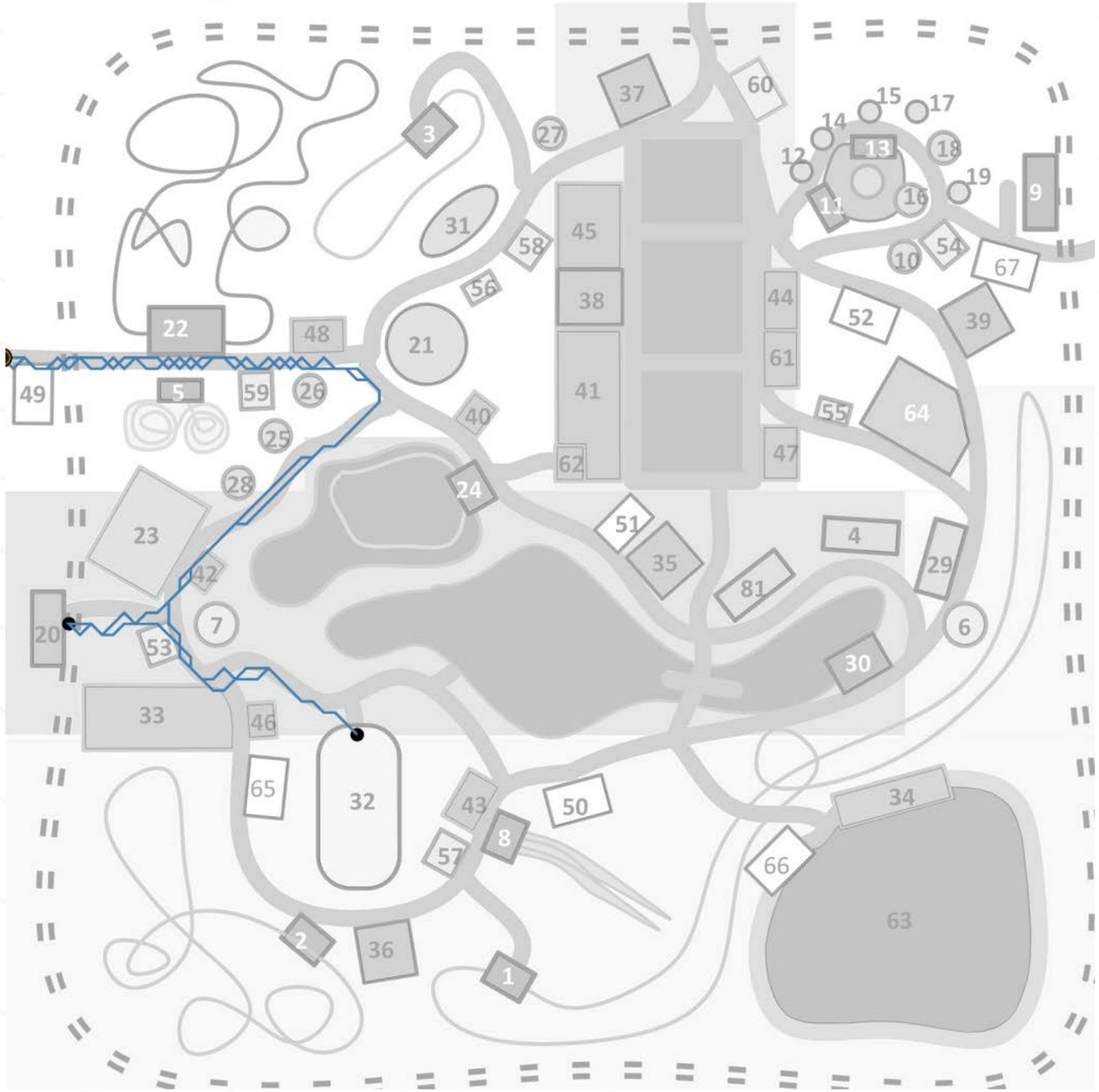






Culprit's Behavior Sunday

- Creighton Pavilion
- Scholtz Express



Interpretation

- 4 friends — 3 college scholarships
- Neglect friends and family to pursue goals
- Poor, misguided and disgruntled figure from Scott's past

- Resentful: Cannot bear to witness Scott's performance
- Obsessed: Cannot stop visiting stage and pavilion
- Depressed: After showing evidence of resentful and obsessive behavior, visits beer garden.

Georgia Institute of Technology



Introduction to Visual Analytics:

Coupling Cognition and Computation through Interactive Visual Interfaces

VA.lab

Georgia Tech Visualization Lab

Alex Endert
Assistant Professor
School of Interactive Computing
Georgia Tech
endert@gatech.edu

goals of this talk

- give you basic knowledge of a collection of topics that ground VA
- *[insert coffee here]*
- show you some examples of tools
- give you the chance to ask some questions and discuss

Overview first

- **Foundations** of VA
  - What is VA? Why show data visually?
  - Why does visualization work so well? Grounded in areas including: Perception, Cognition, Visualization, Computation, **User Interaction**
  - What **tasks** do people use VA for?
  - How can we integrate data mining, statistics, machine learning? Data models, dimensionality
- **Applications** of VA
  - demos, videos
- Open areas and exciting **new directions** for VA
- Q&A

and now 184 slides of details! :)

Why do we show data visually?

it all starts with data…

- Everyday life in the digital age creates data
  - even my cat is a data creator!
- Data creates fields of study (data science), and changes the way that we perform science in many well-established fields
  - astrophysics, biology, …
- data is a quantitative record of sensed, real-world phenomena — **Exciting**!

Data present opportunity

- Challenge: how do we let people gain insight from data?
  - we want people to make decisions based on data
  - form (and sometimes change) opinions about important issues
  - learn
  - discover new phenomena

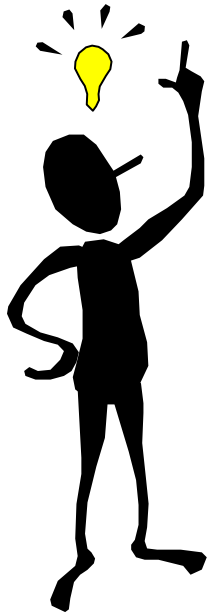
"The purpose of computing is insight, not numbers."

Richard Hamming, 1962

The Challenge

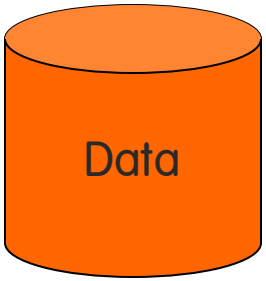
- Transform the data into information (understanding, insight) thus making it useful to people

- How do we let people understand phenomena, given the data?



The data transfer problem

Web,
Books,
Papers,
Game scores,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News



Data Transfer

How?



Vision: 100 MB/s
Ears: <100 b/s
Telepathy
Haptic/tactile
Smell
Taste

next few slides courtesy of Chris North

Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity

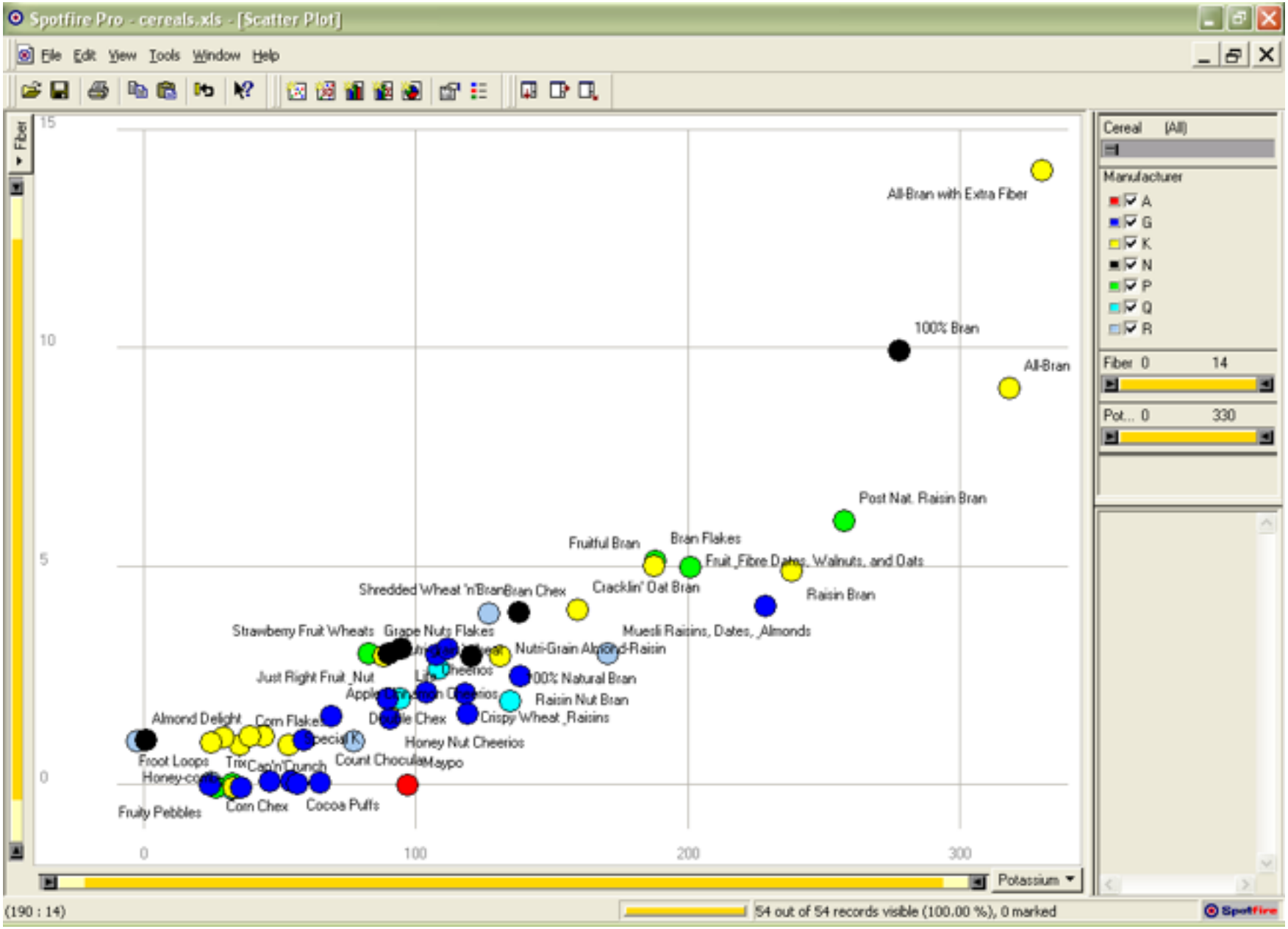
  (Multiplication test)
- People think visually

Impressive. Lets use it!

An Example

- Why visualization helps…

	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



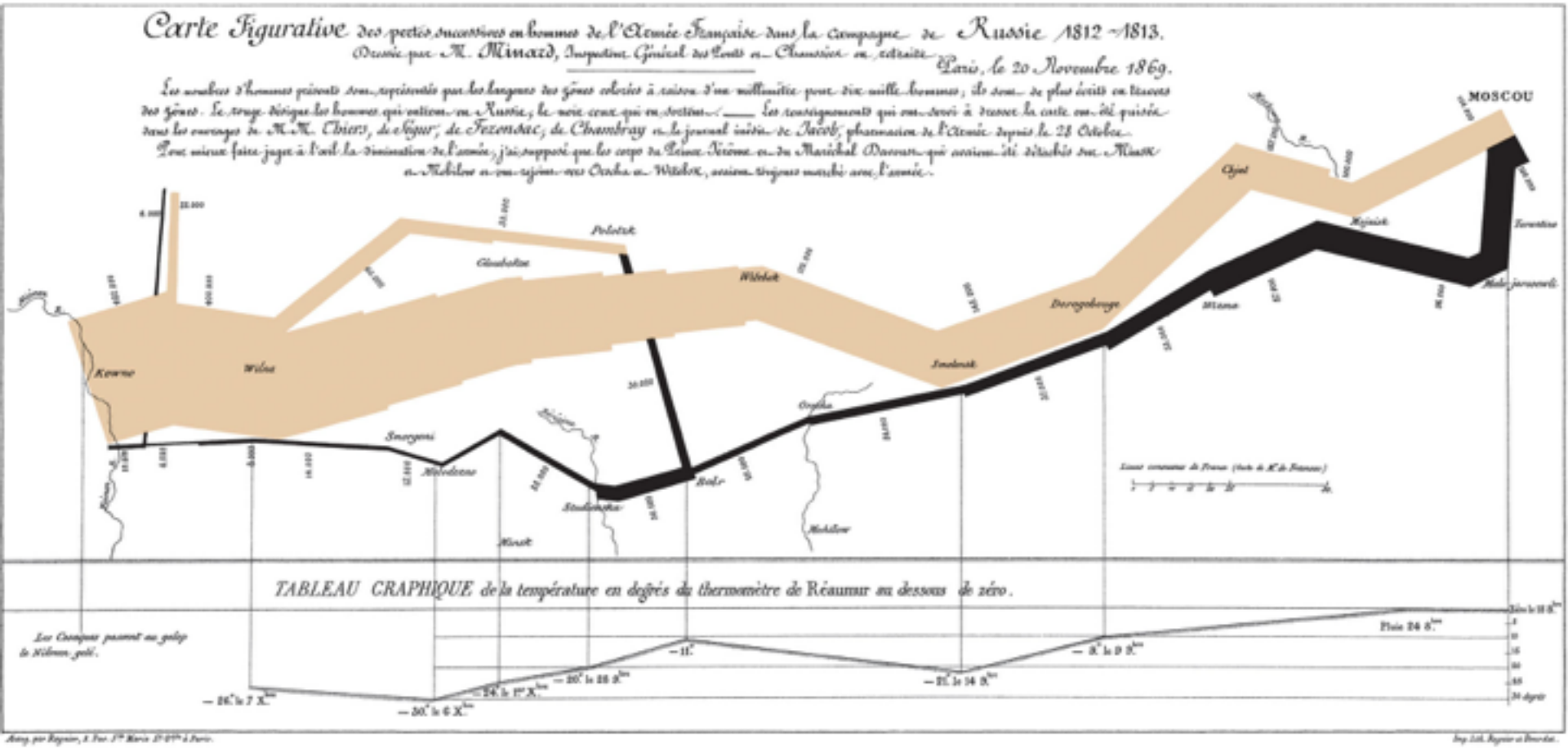
Fiber

Potassium

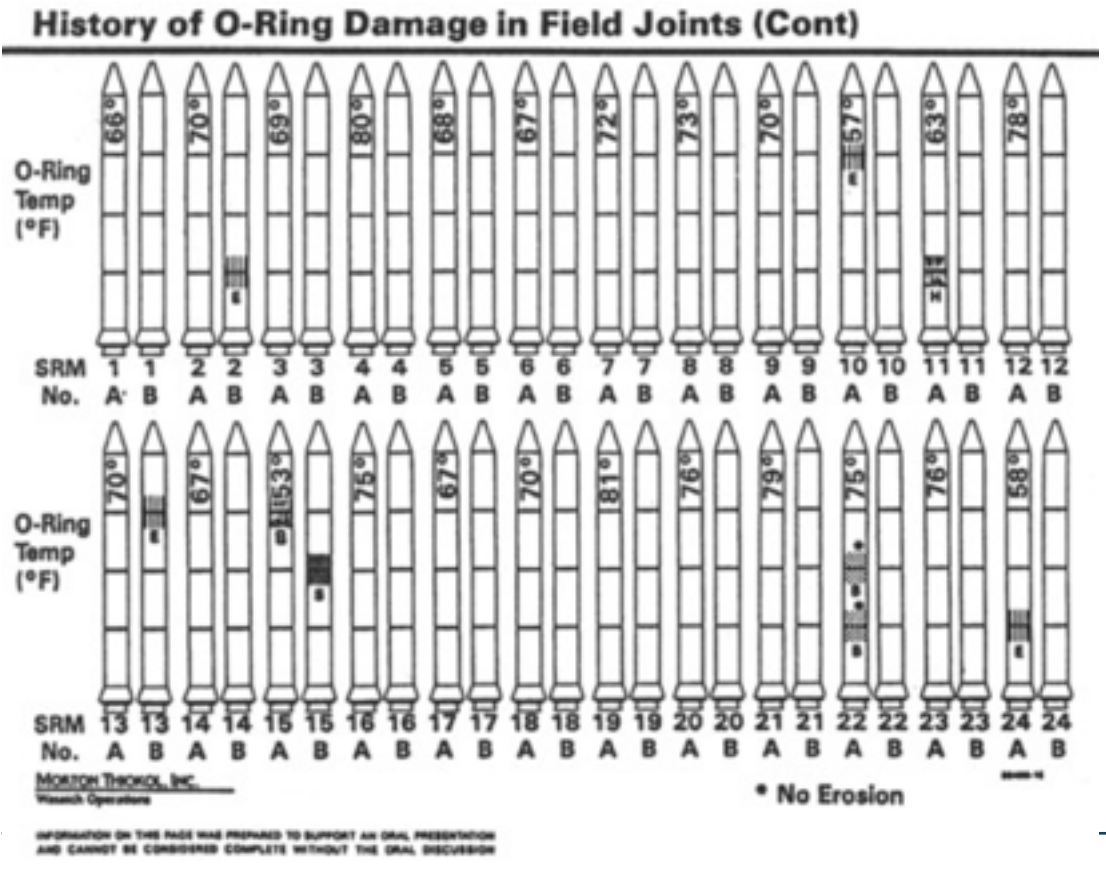
Even Tougher?

- 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?

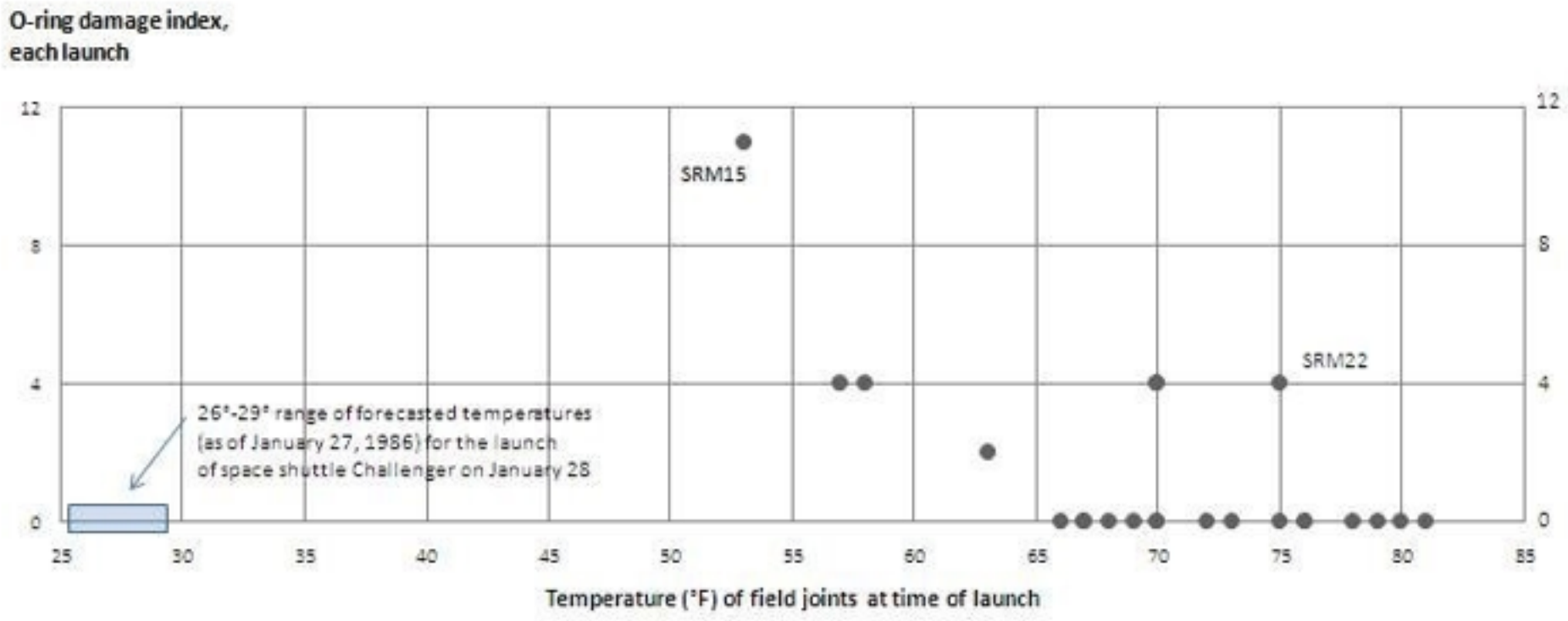
many other examples exist



many other examples exist



many other examples exist



Tufte, Edward R., and P. R. Graves-Morris. The visual display of quantitative information. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

Visualization

- “The use of computer-supported, interactive visual representations of data to amplify cognition.”

  From [Card, Mackinlay Shneiderman ’98]

Visualization

- Often thought of as process of making a graphic or an image
  - "show me the data!"
  - "I've got all this data, visualize it!"
- Really is about fostering **cognitive processes**
  - Form a mental model of phenomena
  - Internalize understanding
  - think long division
- "The purpose of visualization is insight, not pictures"
  - Insight: discovery, decision making, explanation

Main Idea

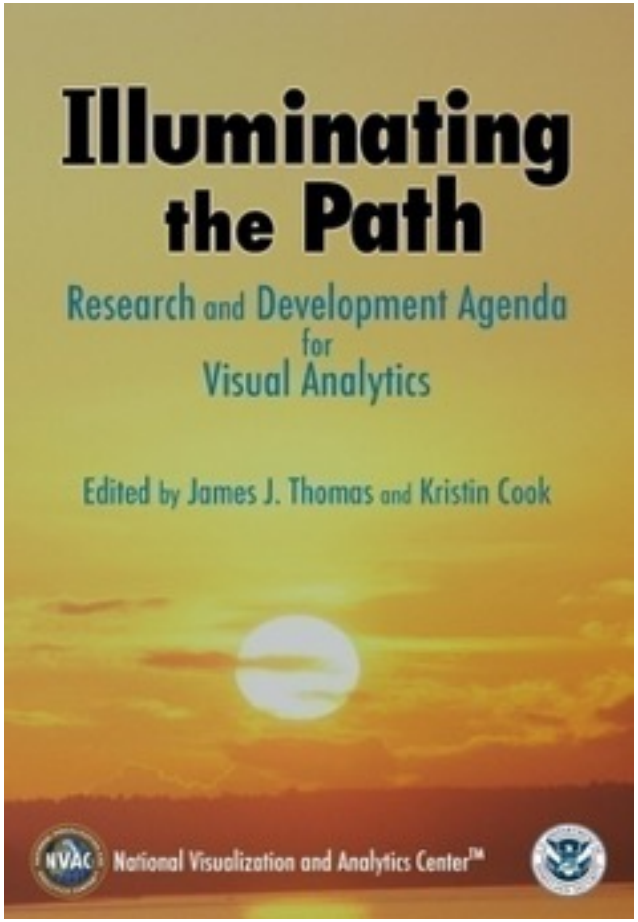
- Visuals help us think
  - Provide a frame of reference, a temporary storage area
- Cognition → Perception
- Pattern matching
- External cognition aid
  - Role of external world in thinking and reason
  - We use it everyday!

Larkin & Simon '87
Card, Mackinlay, Shneiderman '98

scoping and defining
visual analytics

Visual Analytics

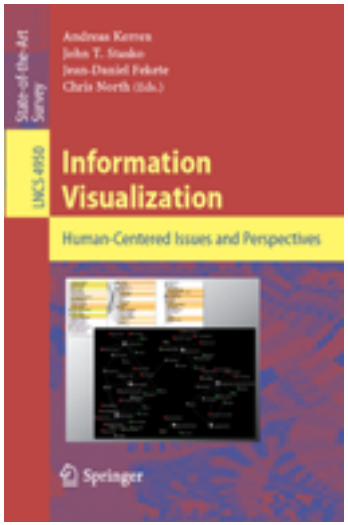
Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces



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

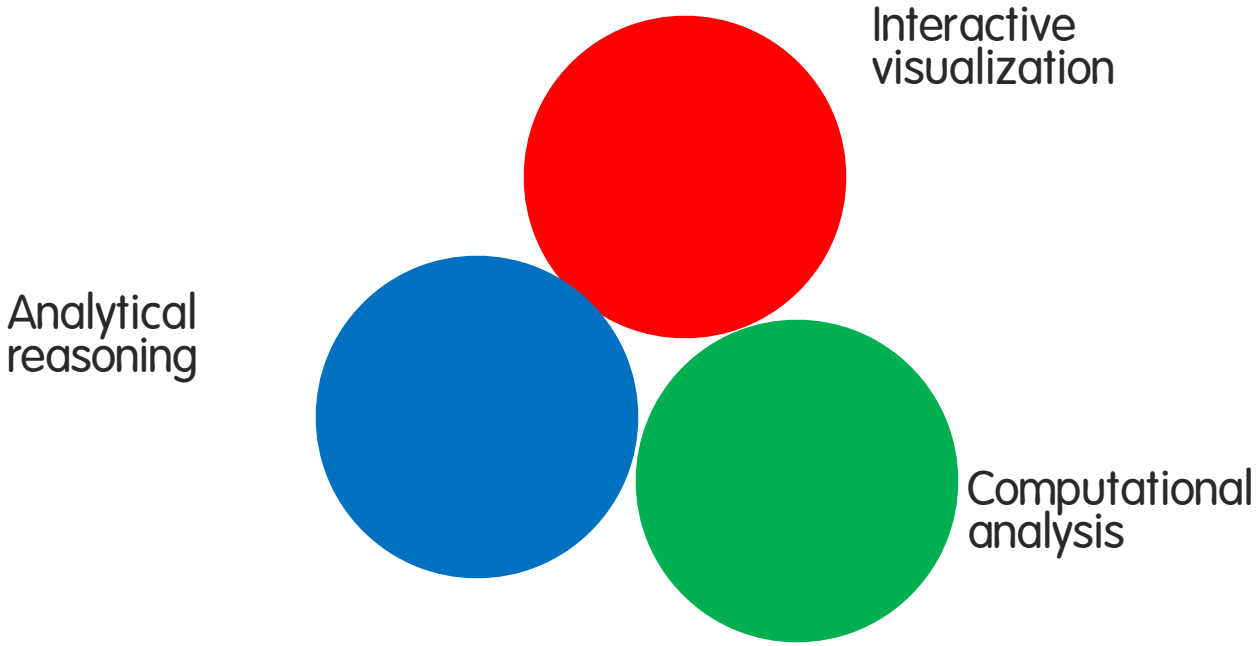
Alternate Definition

Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets

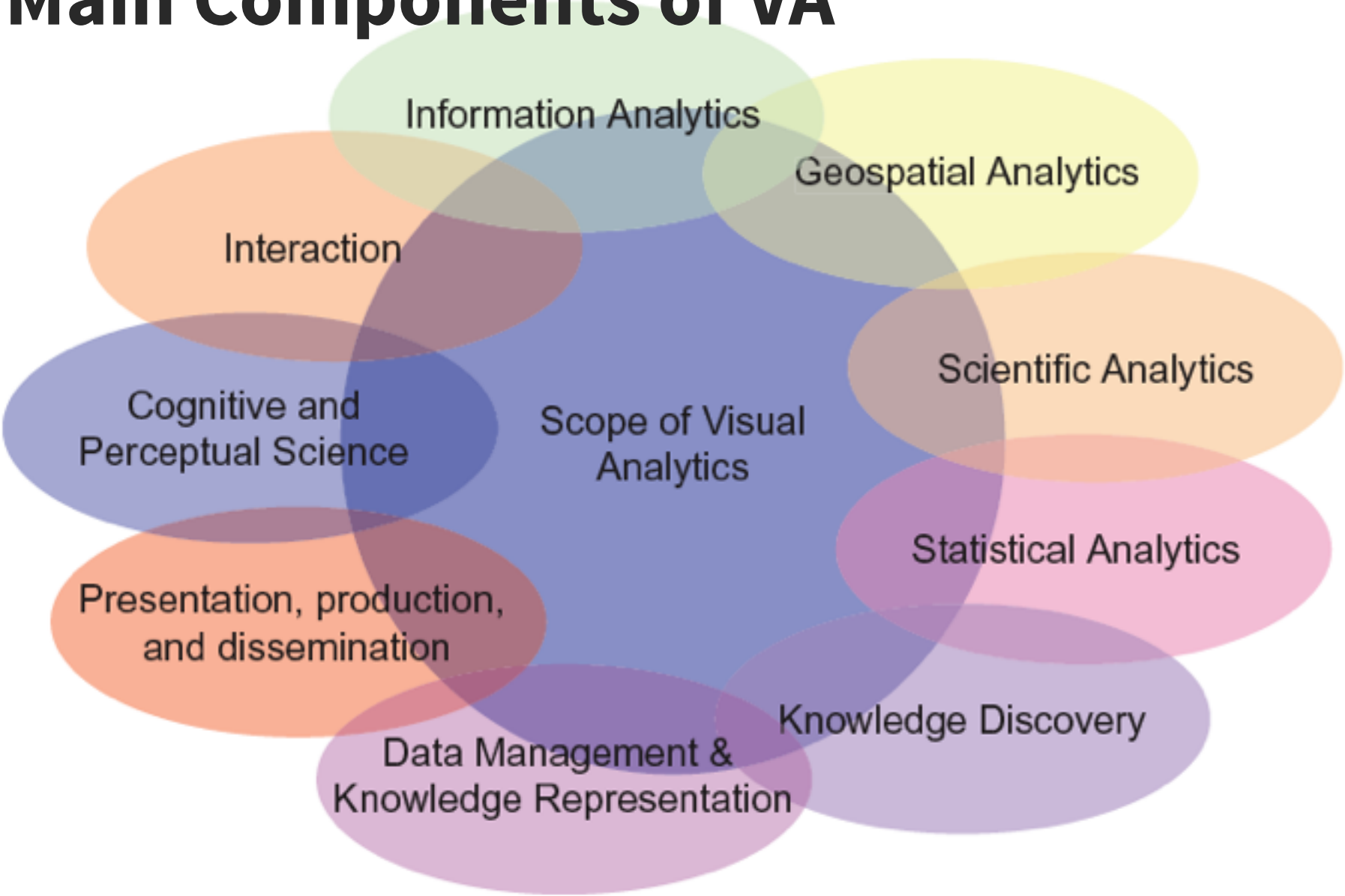


Keim et al
chapter in ***Information Visualization: Human-Centered Issues and Perspectives***, 2008

Main Components of VA



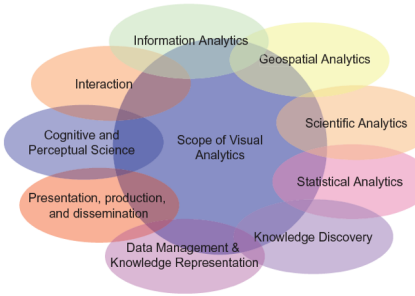
Main Components of VA



Main Components of VA

I won't cover all of these today, but much of what I'm going to talk about has ties to many.

First, let's start with some fundamentals



why does vis work so well for gaining insight from data?

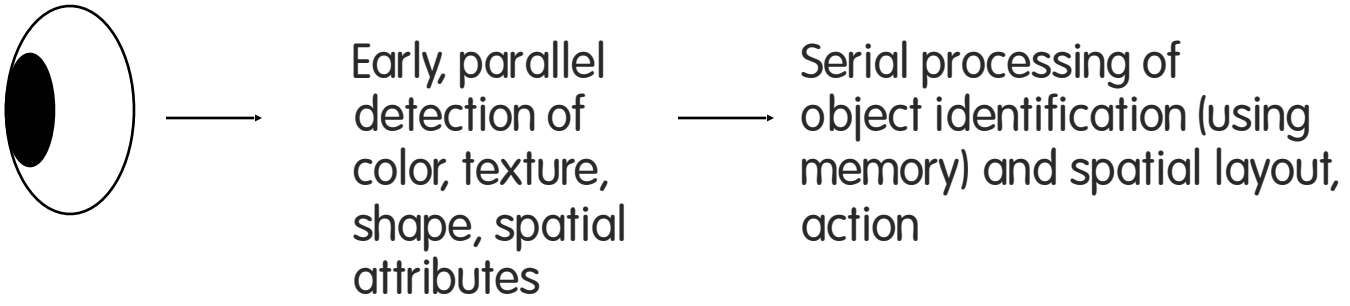
perception **cognition**

Perceptual Processing

- Seeks to better understand visual perception and visual information processing
  - Multiple theories or models exist
  - Need to understand physiology and cognitive psychology
  - This is how we see, therefore the “inputs” to the brain for visualization

One (simple) Model

- Two stage process
  - Parallel extraction of low-level properties of scene
  - Sequential goal-directed processing



Ware 2000

Preattentive Processing

- How does human visual system analyze images?
  - Some things seem to be done preattentively, without the need for focused attention
  - Generally less than 200-250 msecs (eye movements take 200 msecs)
  - Seems to be done in parallel by low-level vision system

Drawn from
C. Healey web article

What Kinds of Tasks is it good for?

- Target detection
  - Is something there?
- Boundary detection
  - Can the elements be grouped?
- Counting
  - How many elements of a certain type are present?

How Many 3's?

128176875613897654698450698560498282676298098584582245098564589450984509809435859091030209905959595772564675050678904567884578980982167765487636490856091294968 6

How Many 3's?

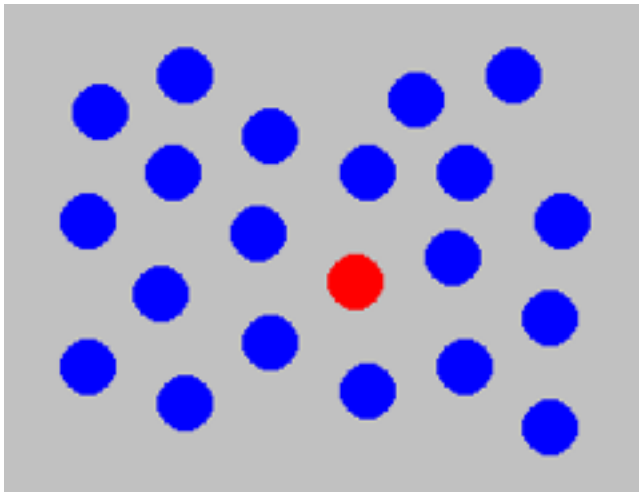
128176875613897654698450698560498282676 2
980985845822450985645894509845098094 3585
909103020990595959577256467505067890456 7
884578983982167765487636490856091294968 6

Example

- Determine if a red circle is present

- raise your hand when you see it

Hue



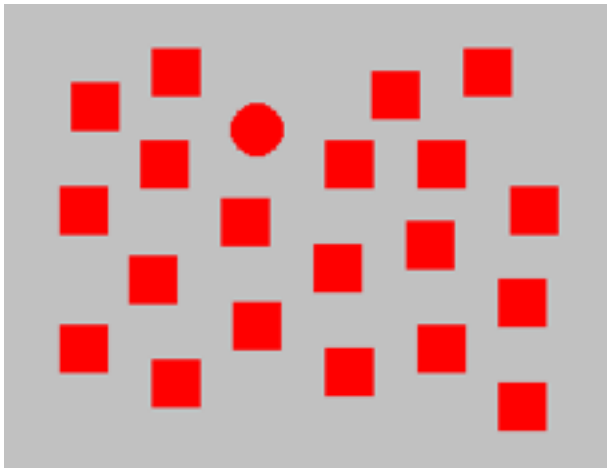
Can be done rapidly (preattentively) by people
Surrounding objects called “distractors”

Example

- Determine if a red circle is present

- raise your hand when you see it

Shape

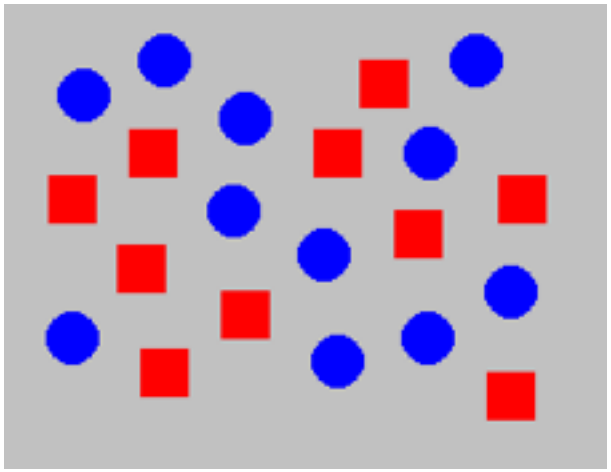


Can be done preattentively by people

Example

- Determine if a red circle is present

Hue and Shape

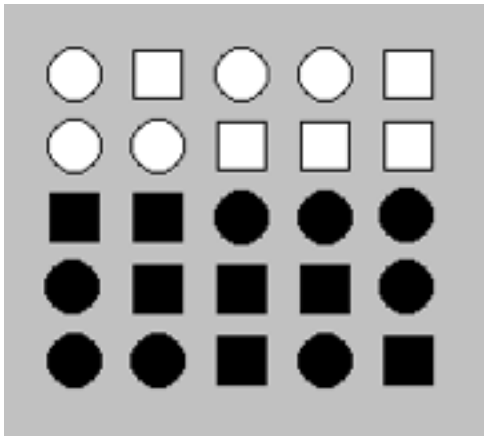


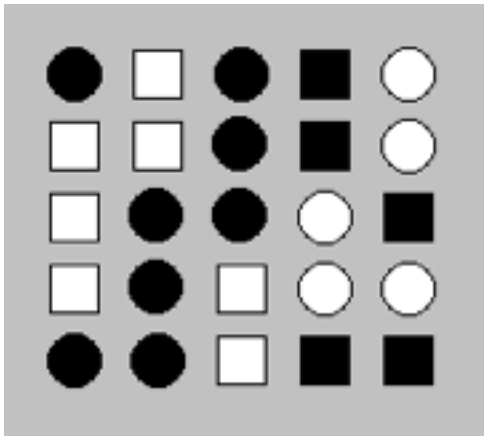
- Cannot be done preattentively
- Must perform a sequential search
- Conjuction of features (shape and hue) causes it

Example

- Is there a boundary in the display?

Fill and Shape



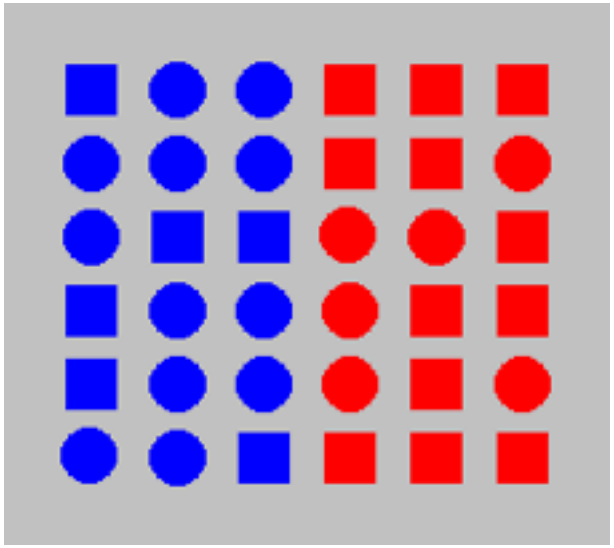


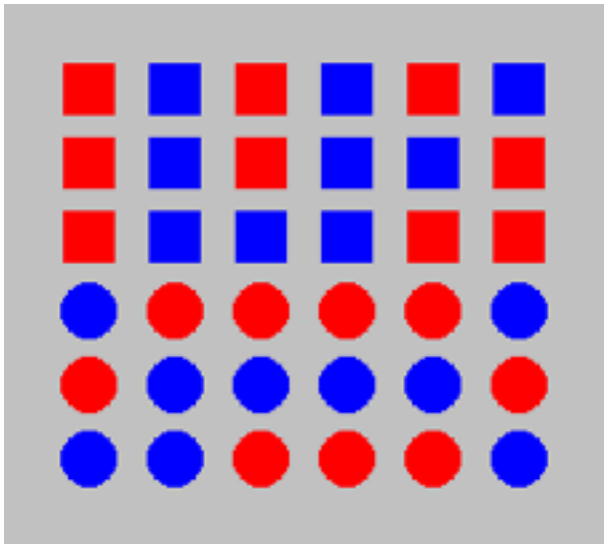
- Left can be done preattentively since each group contains one unique feature
- Right cannot (there is a boundary!) since the two features are mixed (fill and shape)

Example

- Is there a boundary in the display?

Hue versus Shape





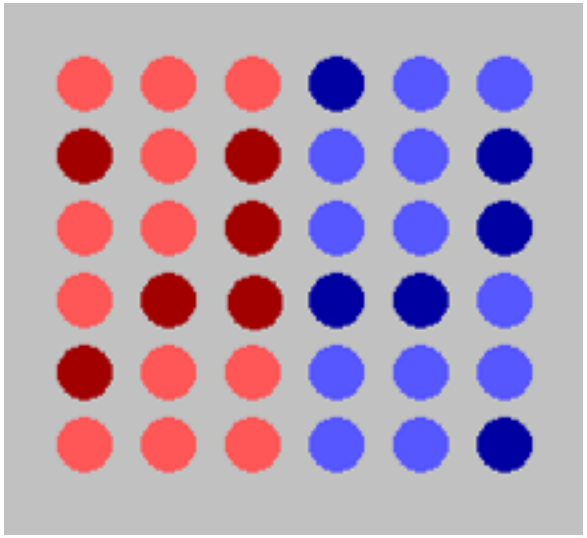
Left: Boundary detected preattentively based on hue regardless of shape

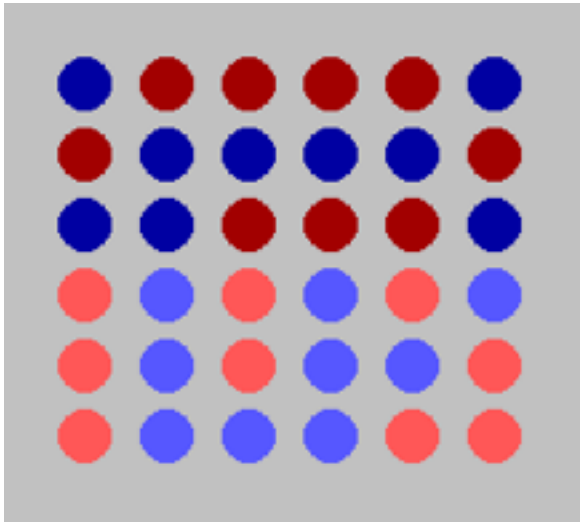
Right: Cannot do mixed color shapes preattentively

Example

- Is there a boundary?

Hue versus brightness





Left: Varying brightness seems to interfere
Right: Boundary based on brightness can be done preattentively

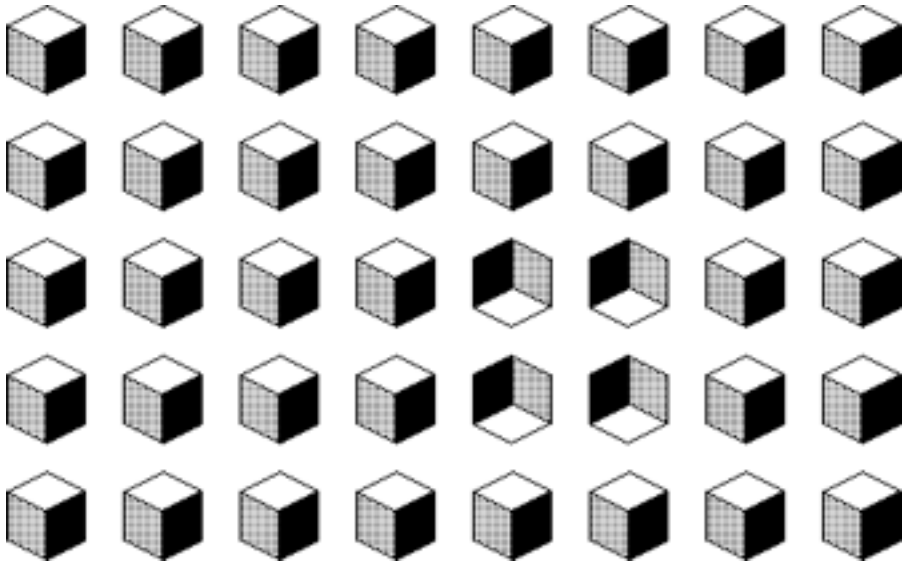
Example Applet

- Nice on-line tutorial and example applet
  - http://www.csc.ncsu.edu/faculty/healey/PP/index.html
  - Chris Healey, NC State
  - Prior pictures taken from site

Preattentive Features

- Certain visual forms lend themselves to preattentive processing
- Variety of forms seem to work

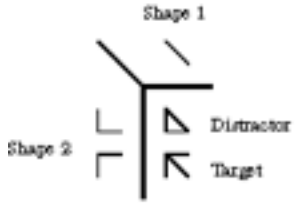
3-D Figures



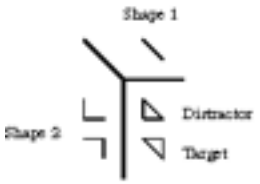


3-D visual reality has an influence

Emergent Features









Gestalt Laws

- Background
  - German psychologists, early 1900's
  - Attempt to understand pattern perception
  - Founded Gestalt school of psychology
  - Provided clear descriptions of many basic perceptual phenomena

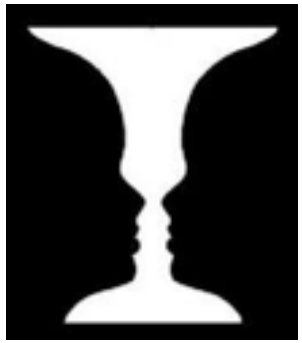
    → Gestalt Laws of Pattern Perception

Gestalt Laws

- Proximity
  - Things close together are perceptually grouped together
- Similarity
  - Similar elements get grouped together
- Connectedness
  - Connecting different objects by lines unifies them
- Continuity
  - More likely to construct visual entities out of smooth, continuous visual elements

Gestalt Laws

- Symmetry
  - Symmetrical patterns are perceived more as a whole
- Closure
  - A closed contour is seen as an object
- Relative Size
  - Smaller components of a pattern as perceived as objects
- Figure & Ground
  - Figure is foreground, ground is behind



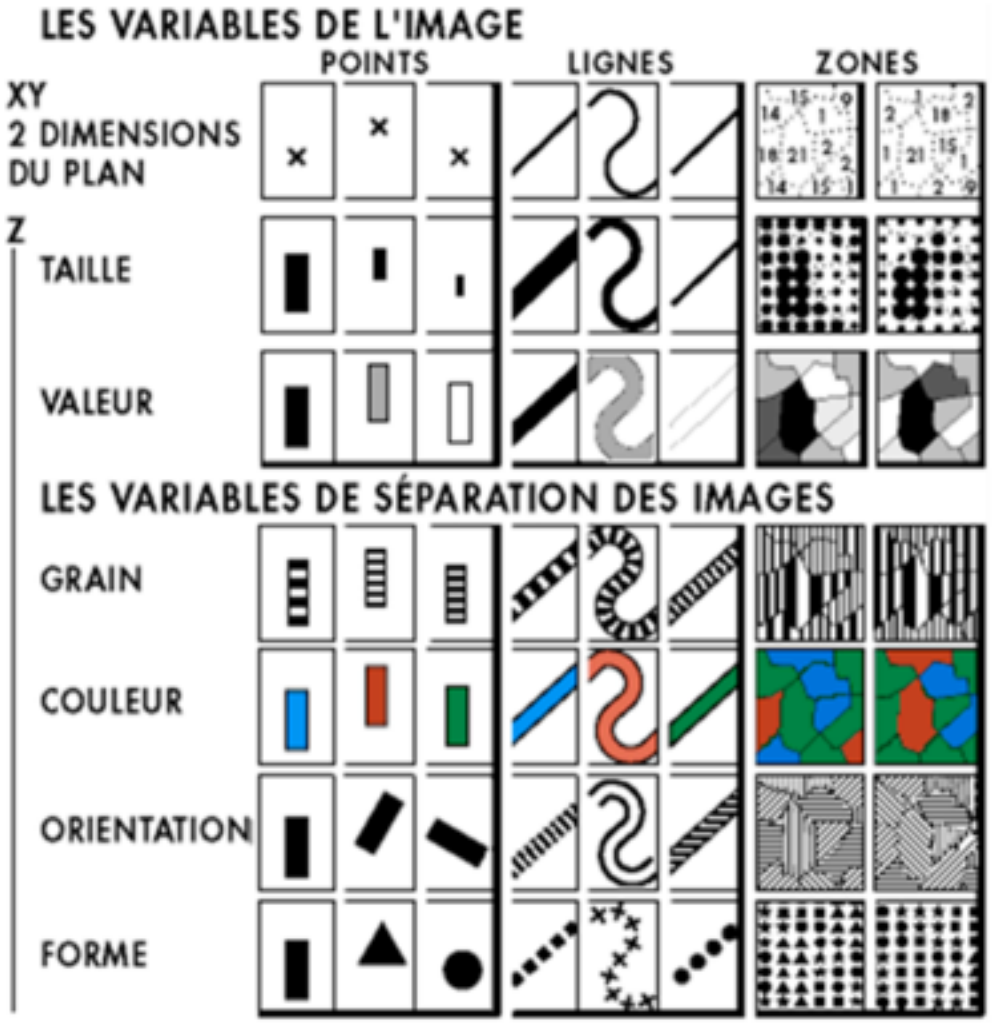
Basics of Visual Representation

- binding data values to visual glyphs
- emphasize salient features of data for perceptual system
- augments cognitive processes of people to reason about data

Visual Variables

Position

Size
Value
Color
Texture
Orientation
Shape



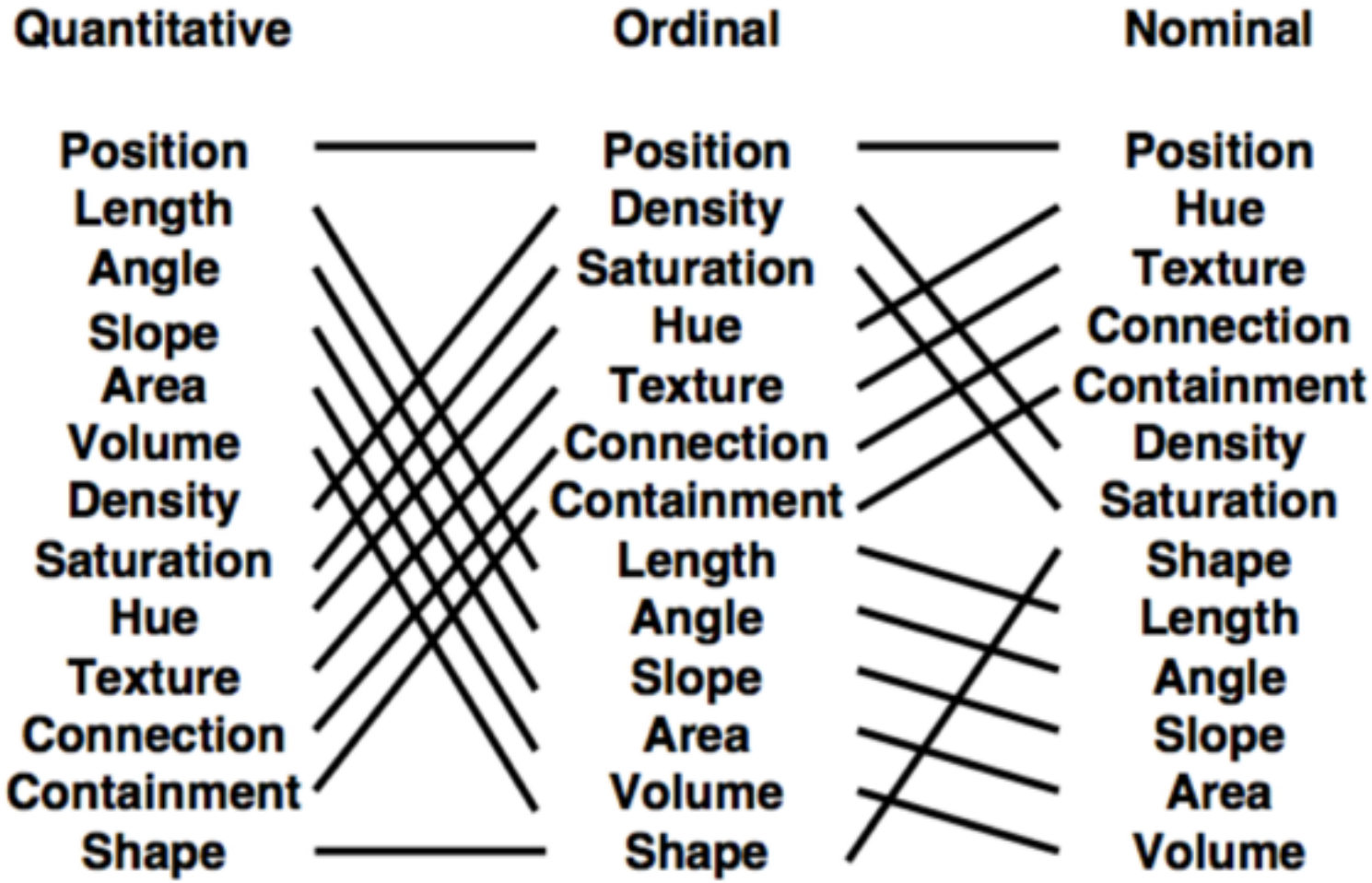
Bertin, Jacques. "Semiology of graphics: diagrams, networks, maps." (1983).

How many to use?

- Can only show limited amount of features
  - map each feature to a visual variable

- That means for lots of complex “big” data, we need methods to aggregate or combine data features
  - even then, we may still need more help
  - e.g., dimension reduction - will cover this later

which visual feature is best?

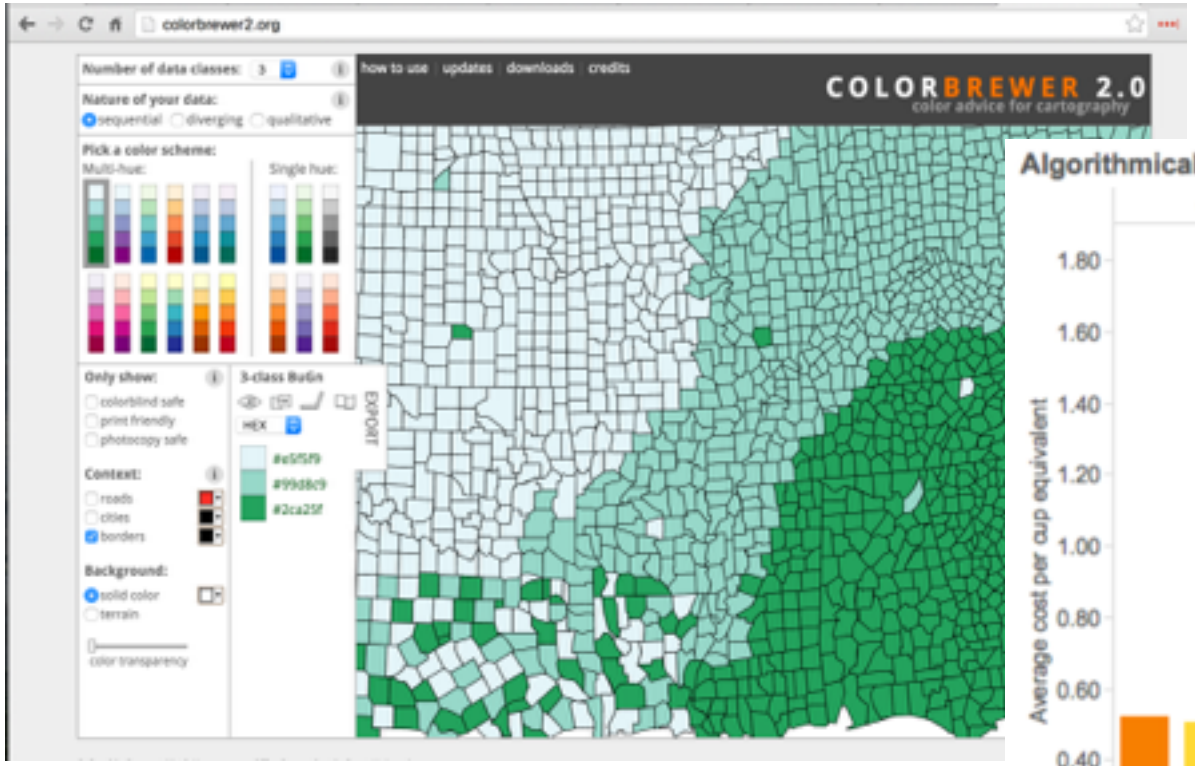


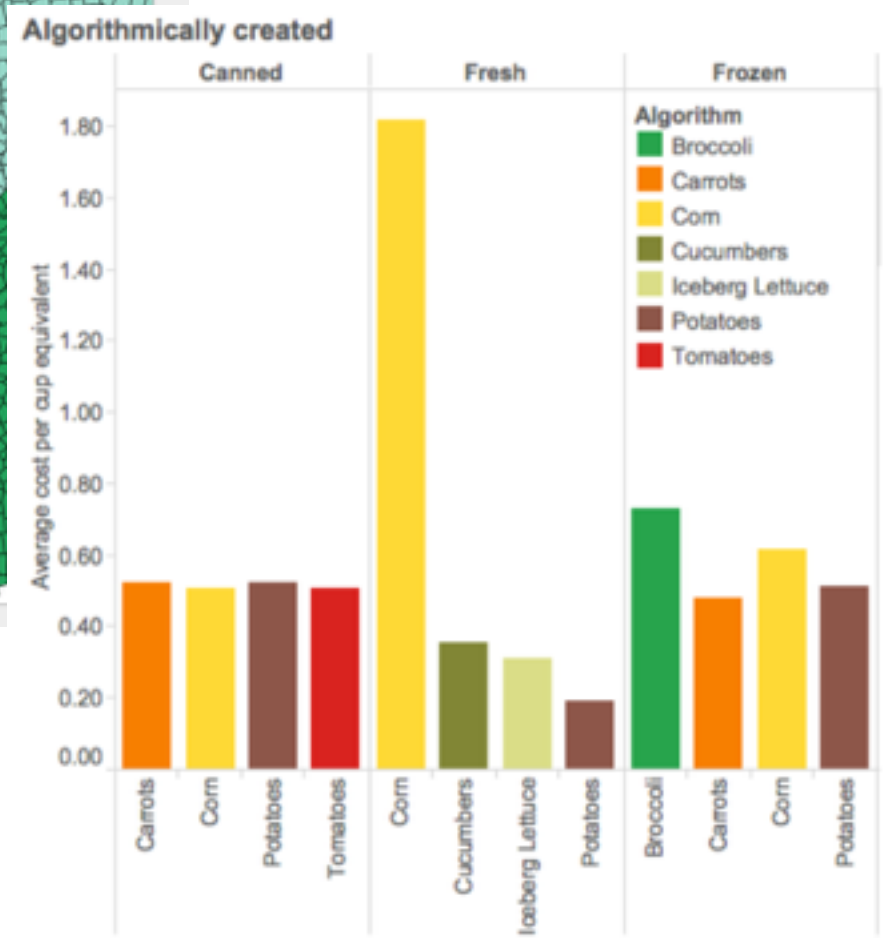
Mackinlay, Jock. "Automating the design of graphical presentations of relational information." Acm Transactions On Graphics (Tog) 5.2 (1986): 110-141.

color is tricky

- lots of research on color ramps, color theory, etc.
- I won't go into details, but will say this:
  - chose wisely
  - ordering of color should be used when data has order
  - separation in color ramp should indicate separation in data
  - there are limits to the number of items/colors people can distinguish between (~7-9)

picking a color scheme





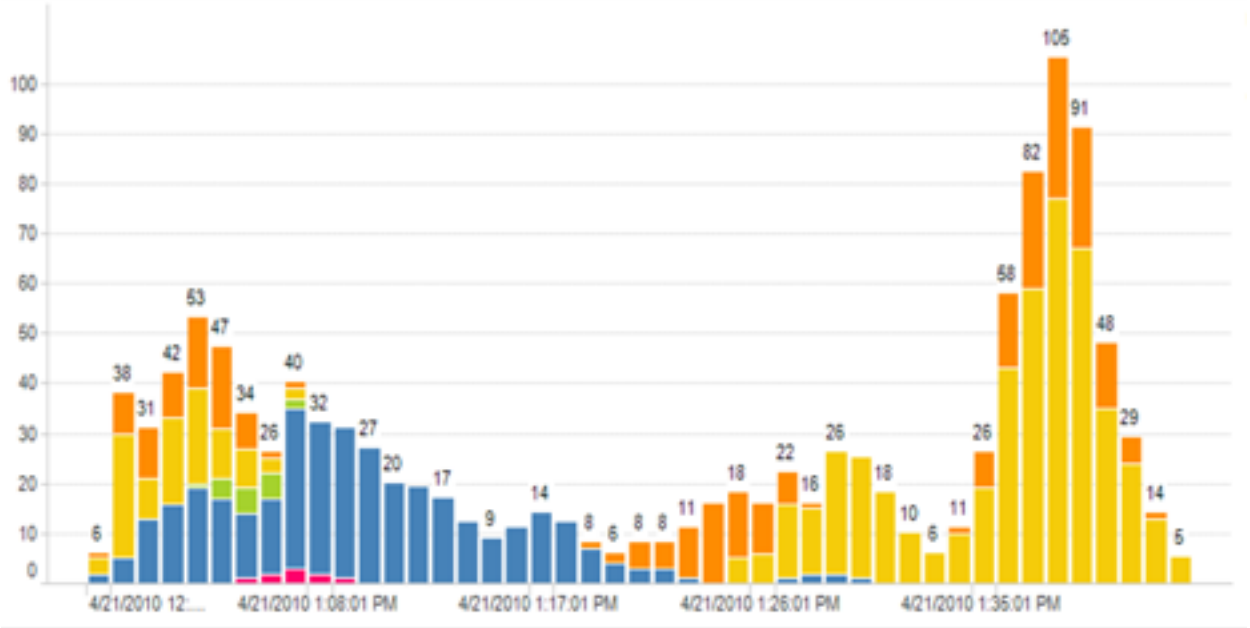
Setlur, Vidya, and Maureen C. Stone. "A Linguistic Approach to Categorical Color Assignment for Data Visualization." Visualization and Computer Graphics, IEEE Transactions on 22.1 (2016): 698-707.

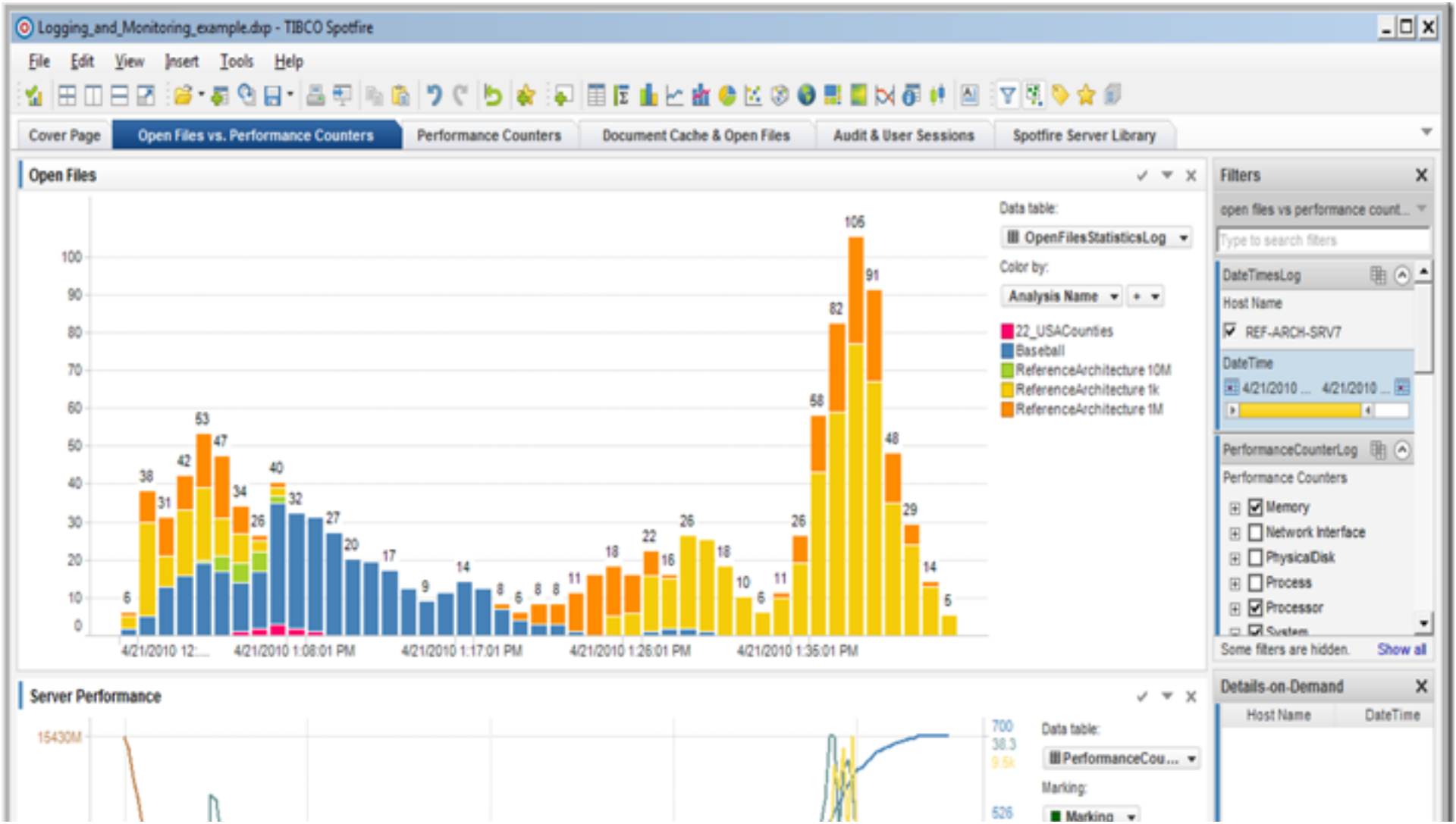
how do we involve the person in the analysis?

user interaction

we've talked a lot about visual representation

- What about interaction?
- How do we let people become active participants, not passive viewers?


100
90
80
70
60
50
40
30
20
10
0
6
38
31
42
53
47
34
26
40
32
27
20
17
9
14
8
6
8
8
11
18
22
16
26
18
10
6
11
26
58
82
106
91
48
29
14
5
4/21/2010 12:...
4/21/2010 1:08:01 PM
4/21/2010 1:17:01 PM
4/21/2010 1:26:01 PM
4/21/2010 1:35:01 PM


Logging_and_Monitoring_example.dxp - TIBCO Spotfire
File Edit View Insert Tools Help
Cover Page
Open Files vs. Performance Counters
Performance Counters
Document Cache & Open Files
Audit & User Sessions
Spotfire Server Library
Open Files
Data table:
OpenFilesStatisticsLog
Color by:
Analysis Name
22_USACounties
Baseball
ReferenceArchitecture 10M
ReferenceArchitecture 1k
ReferenceArchitecture 1M
4/21/2010 12:...
4/21/2010 1:08:01 PM
4/21/2010 1:17:01 PM
4/21/2010 1:26:01 PM
4/21/2010 1:35:01 PM
Filters
open files vs performance count...
Type to search filters
DateTimesLog
Host Name
REF-ARCH-SRV7
DateTime
4/21/2010 ... 4/21/2010 ...
PerformanceCounterLog
Performance Counters
Memory
Network Interface
PhysicalDisk
Process
Processor
Some filters are hidden.
Show all
Server Performance
15430M
Data table:
PerformanceCou...
Marking:
Marking
Details-on-Demand
Host Name
DateTime

The InfoVis Pipeline

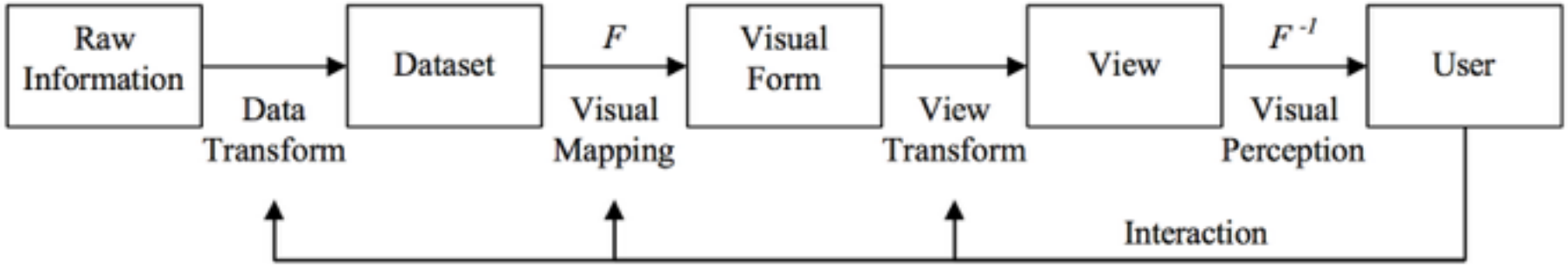


Figure 2: The visualization pipeline, converting information into interactive visual representations (adapted from [Card et al., 1999]).

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 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

IEEE TVCG 13(6), '07

Study Methodology

- Survey
  - 59 papers

    Papers introducing new interaction systems

    Well-known papers in subareas of Infovis
  - 51 systems

    Commercial Infovis Systems (SeeIT, Spotfire, TableLens, InfoZoom, etc.)
  - Collected 311 individual interaction techniques
- Affinity Diagram Method

Main Idea

- Don’t focus so much on particular interactive operations and how they work

- Interaction is ultimately being done by a person for a purpose
  - Seeking more information, solving a problem
  - Fundamental aspect of exploratory, analytic discourse

Results

7 categories

Select

Explore

Reconfigure

Encode

Abstract/Elaborate

Filter

Connect

1. Select

“Mark something as interesting”

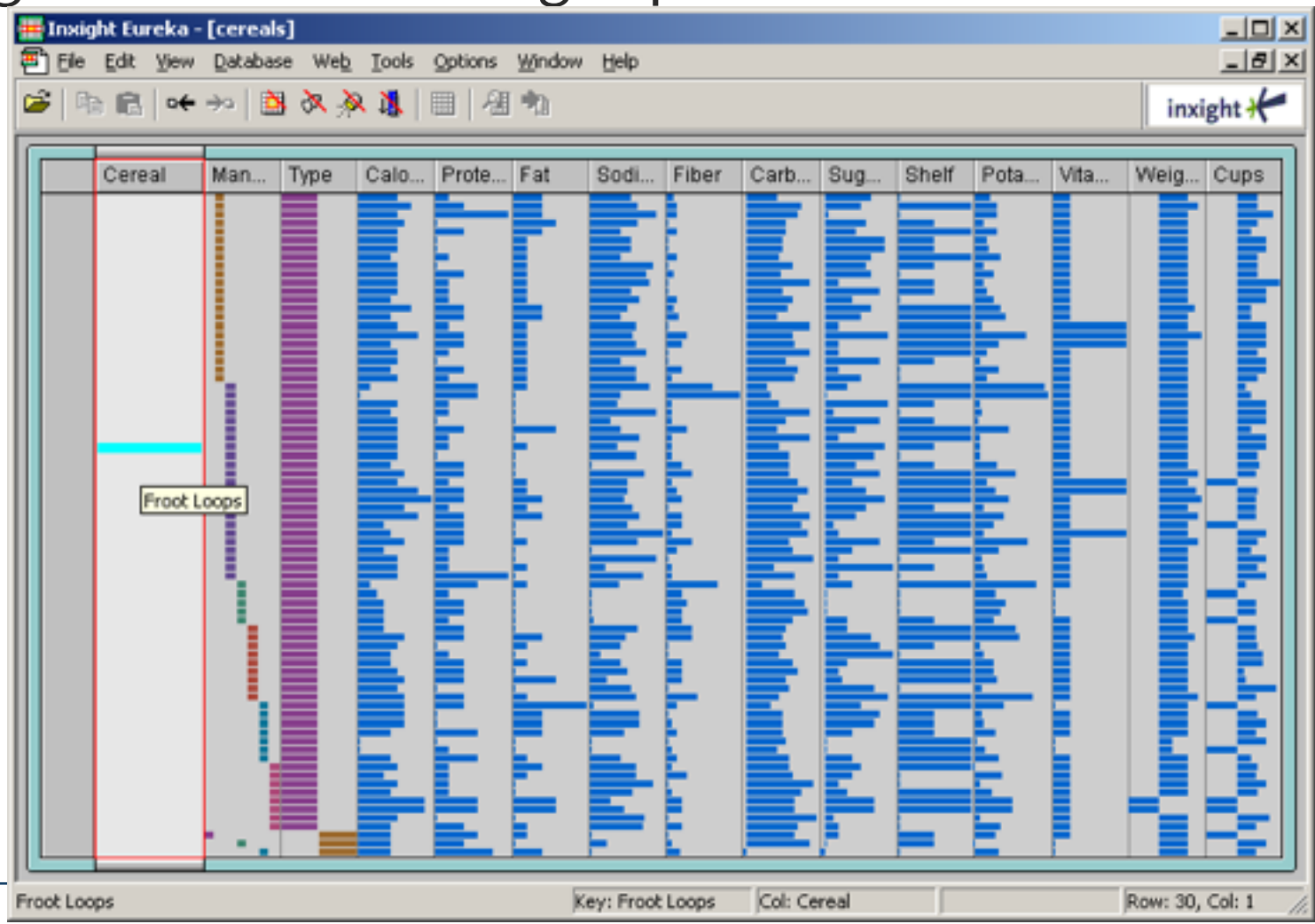
- Mark items of interest to keep track
- Seems to often work as a preceding action to subsequent operations.

e.g.,

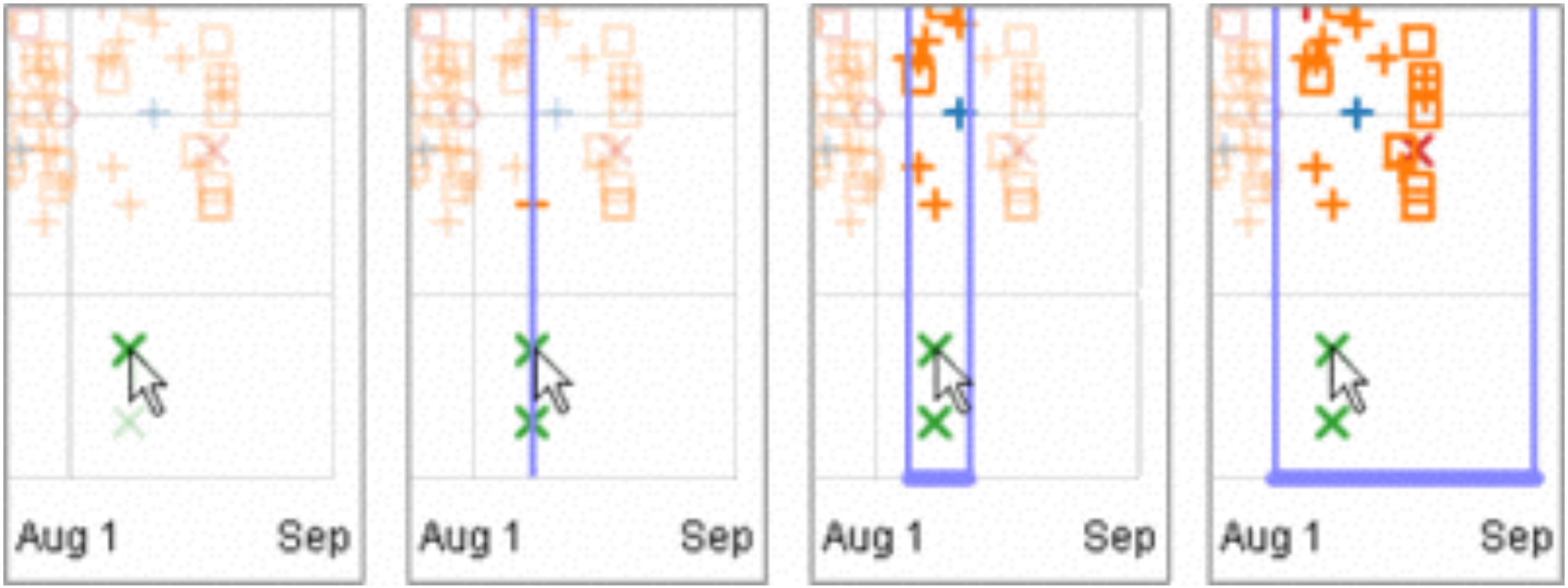
- Selecting a placemark in Google Map
- The Focus feature in TableLens

Pop-up tooltips

- Hovering mouse cursor brings up details of item



Query Relaxation



As you dwell on your mouse click, the selection criteria broaden and you can choose sets of items

Heer, Agrawala, Willett
CHI '08

2. Explore

“Show me something different”

- Enable users to examine a different subset of data
- Overcome the limitation of display size

e.g.,

- Panning in Google Earth
- expanding a hierarchical category

3. Reconfigure

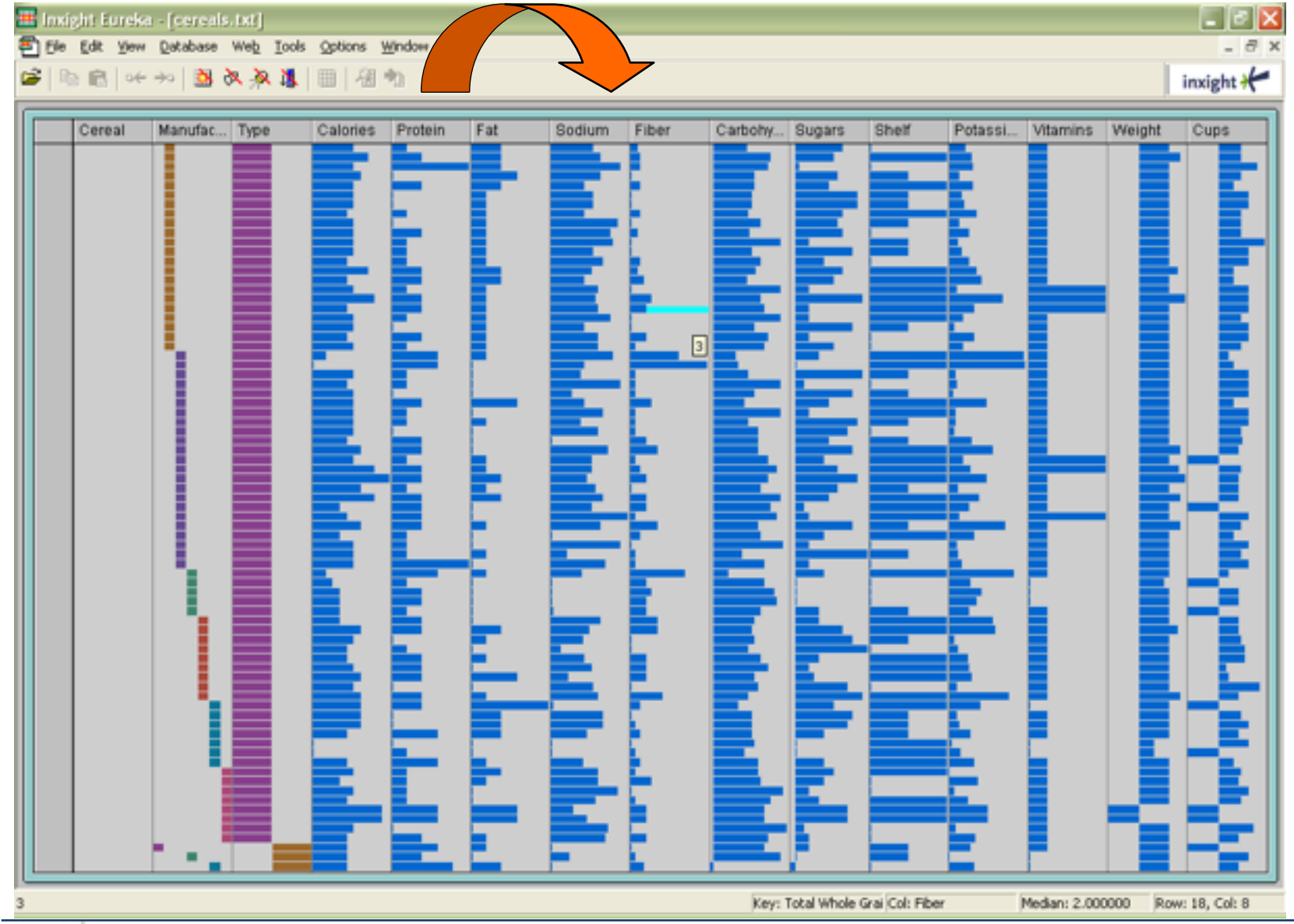
"Show me a different arrangement"

- Provide different perspectives by changing the spatial arrangement of representation

e.g.,

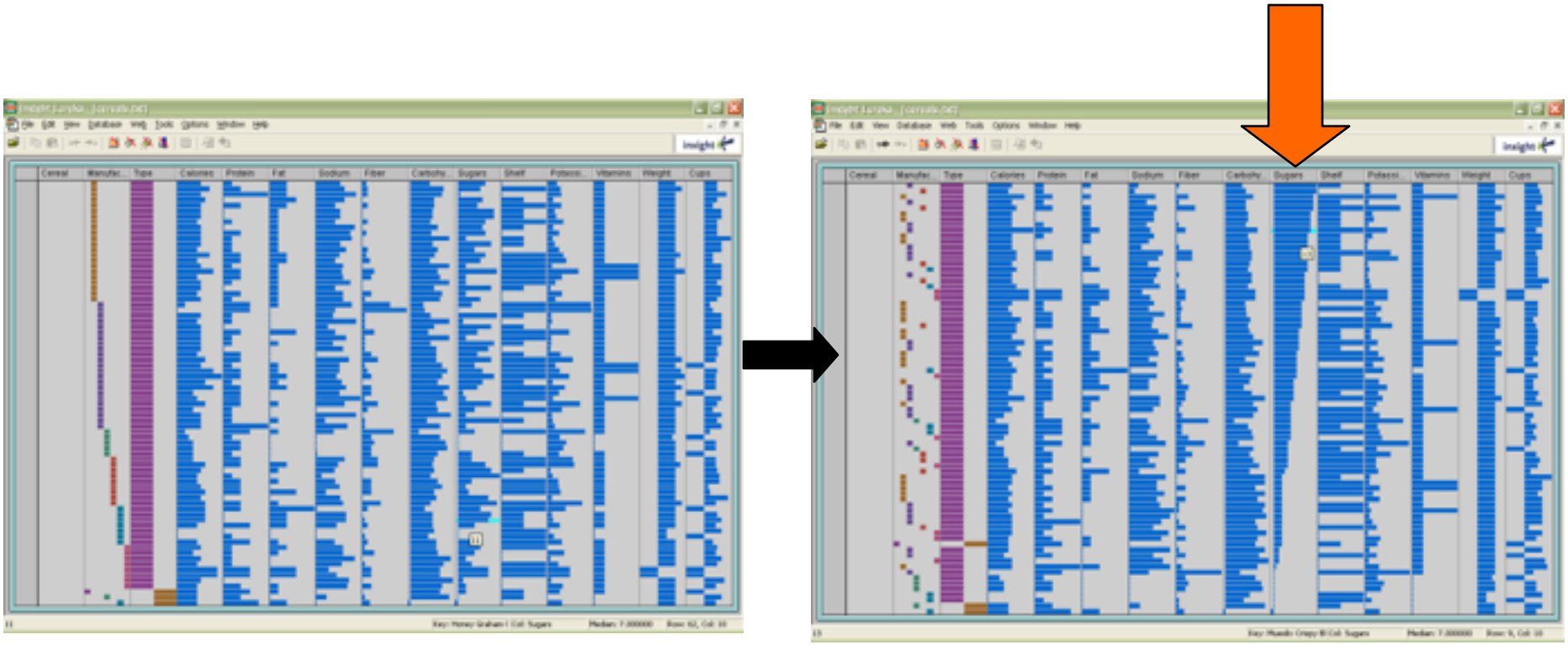
- Sorting and rearranging columns in TableLens
- Changing the attributes in a scatter plot
- The baseline adjustment feature in Stacked Histogram
- The "Spread Dust" feature in Dust & Magnet

Rearrange


Inxight Eureka - [cereals.txt]
File Edit View Database Web Tools Options Window
inxight
Cereal
Manufac...
Type
Calories
Protein
Fat
Sodium
Fiber
Carbohy...
Sugars
Shelf
Potassi...
Vitamins
Weight
Cups
3
Key: Total Whole Grai
Col: Fiber
Median: 2.000000
Row: 18, Col: 8

Sorting

Can sort data with respect to a particular attribute in Table Lens



4. Encode

“Show me a different representation”

- Change visual appearances

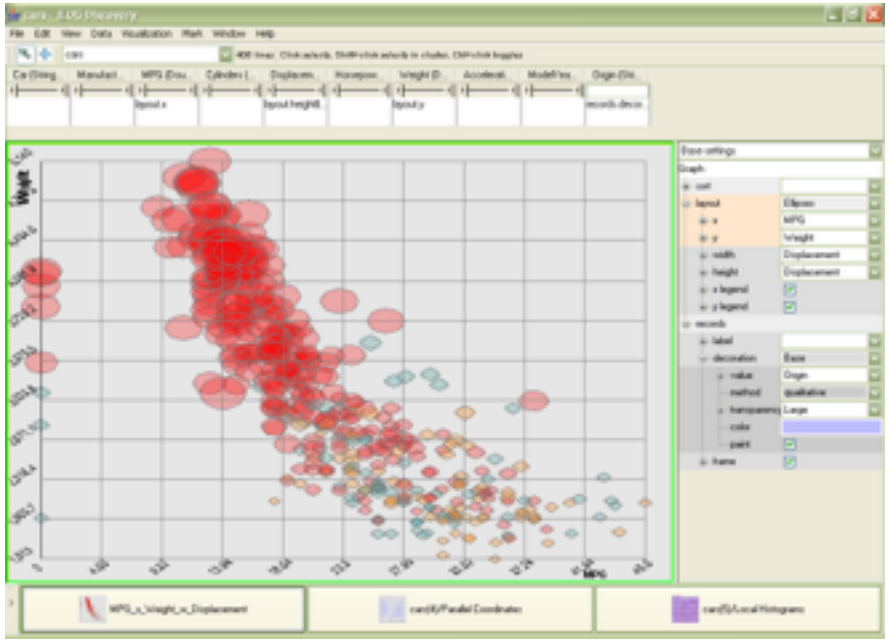
e.g.,

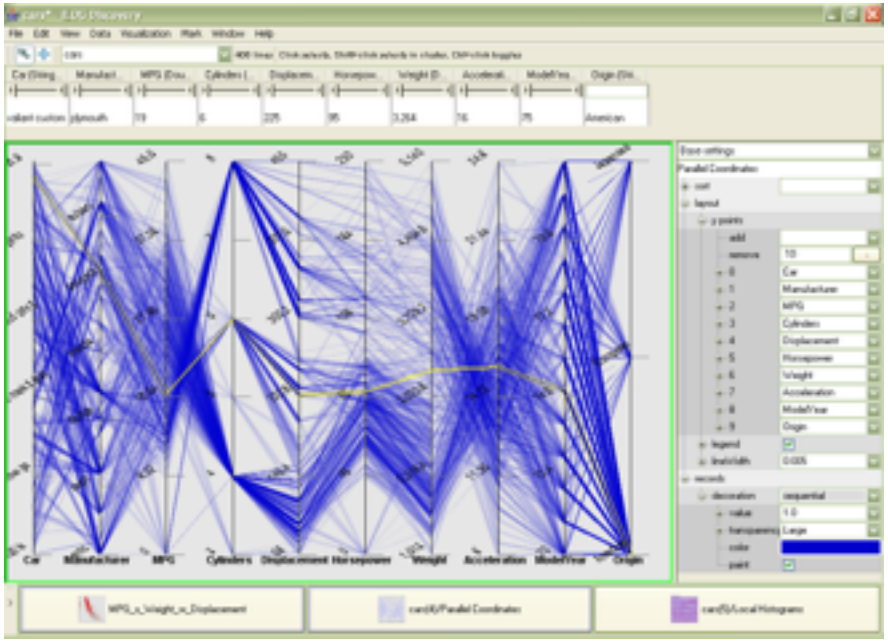
- Changing color encoding
- Changing size
- Changing orientation
- Changing font
- Changing shape

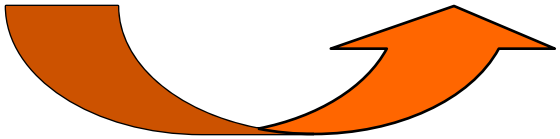
Changing Representation

- May interactively change entire data representation
  - Looking for new perspective
  - Limited real estate may force change

Example







Selecting different representation from options at bottom

5. Abstract/Elaborate

"Show me more or less detail"

- Adjust the level of abstraction (overview and details)

e.g.,

- Unfolding sub-categories in an interactive pie chart
- Drill-down in Treemap
- Details-on-demand in Sunburst
- The tool-tip operation in SeeIT
- Zooming (geometric zooming)

Details-on-Demand

- Term used in infovis when providing viewer with more information/details about data case or cases
- May just be more info about a case
- May be moving from aggregated view to individual view
  - May not be showing all the data due to scale problem
  - May be showing some abstraction of groups of elements
  - Expand set of data to show more details, perhaps individual cases

6. Filter

"Show me something conditionally"

- Change the set of data items being presented based on some specific conditions.

e.g.,

- Dynamic query
- Attribute Explorer
- Keystoke based filtering in NameVoyager
- QuerySketch

Dynamic Query

- Probably best-known and one of most useful infovis techniques
- Let’s explore more details...

Dynamic Query Constituents

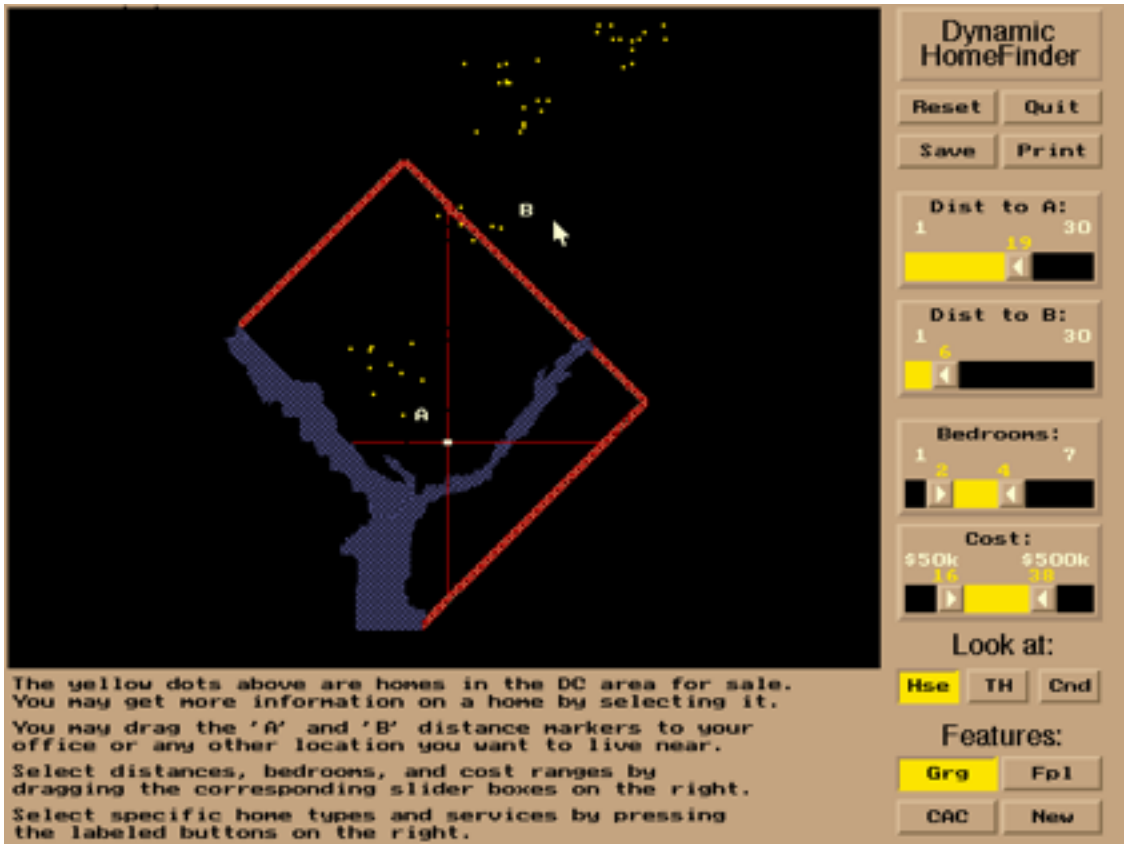
- Visual representation of world of action including both the objects and actions
- Rapid, incremental and reversible actions
- Selection by pointing (not typing)
- Immediate and continuous display of results

Shneiderman
IEEE Software '94

Ahlberg & Shneiderman
CHI '94

DQ Examples

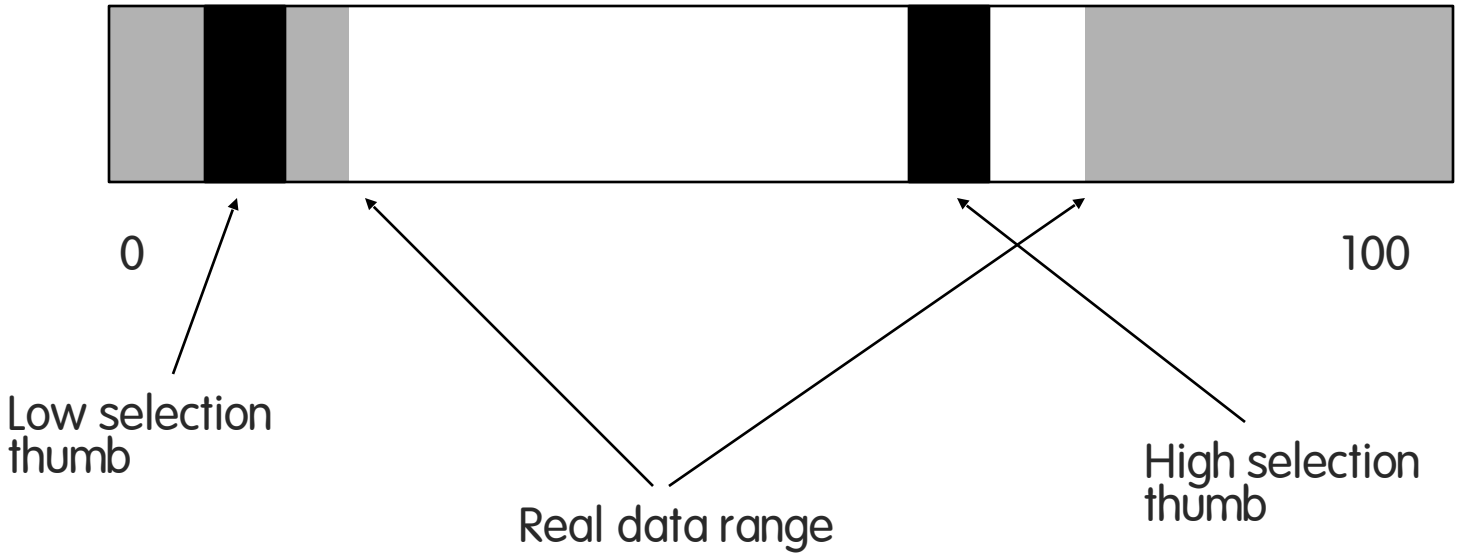
- HomeFinder - Univ. of Maryland



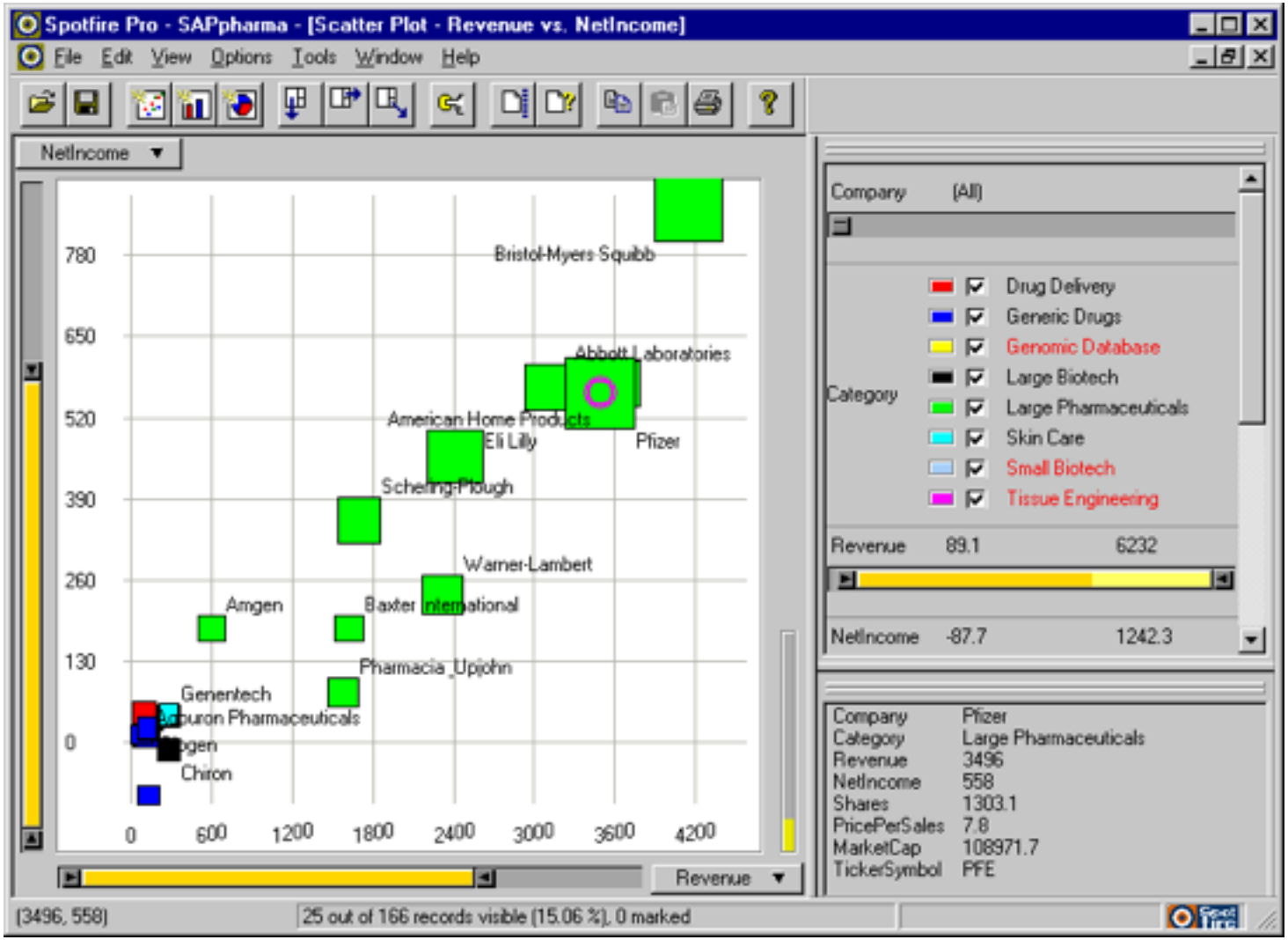
Query Controls

- Variable types
  - Binary nominal - Buttons
  - Nominal with low cardinality - Radio buttons
  - Ordinal, quantitative - sliders

Rangeslider

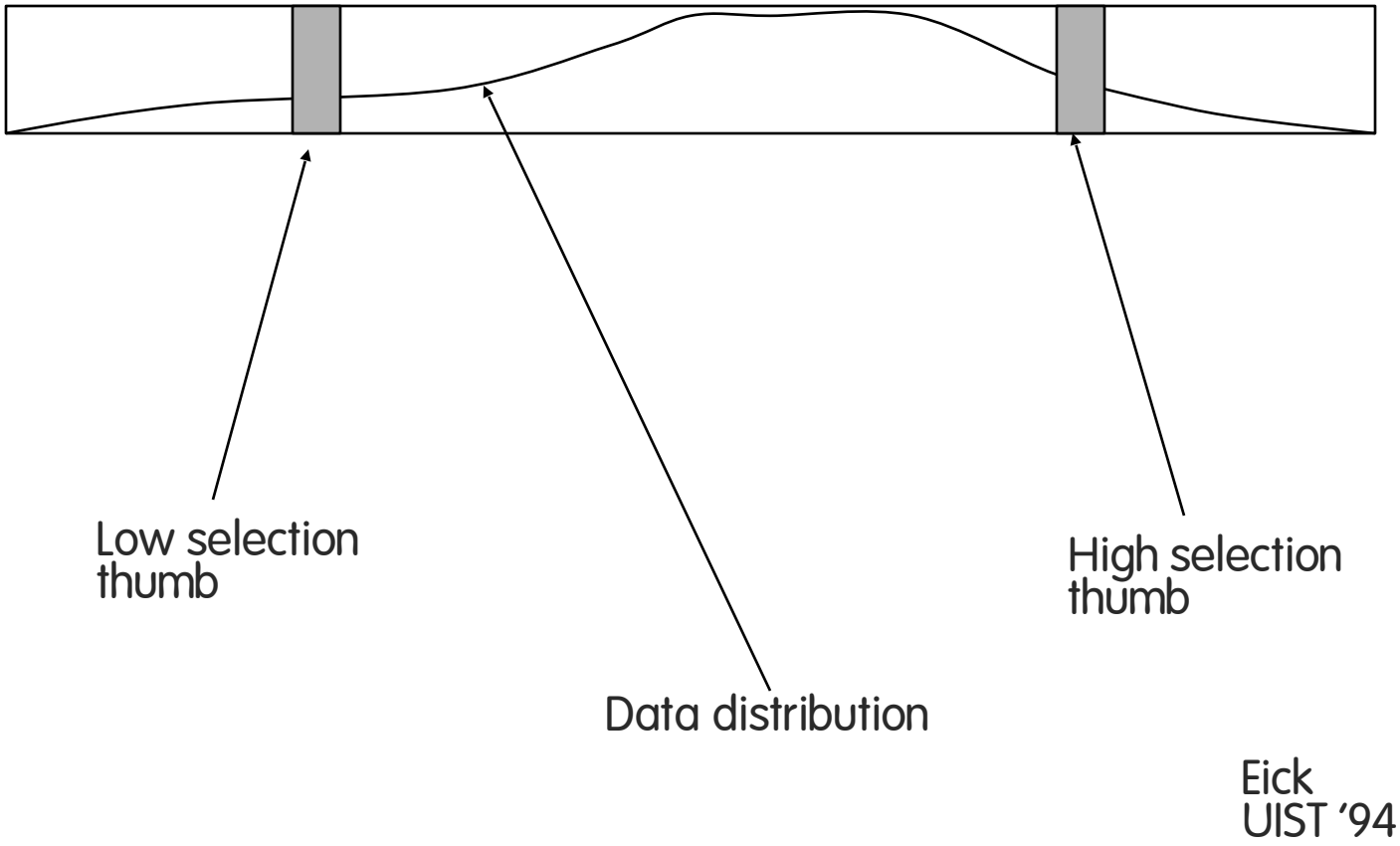


Spotfire


Spotfire Pro - SAPpharma - [Scatter Plot - Revenue vs. NetIncome]
File Edit View Options Tools Window Help
NetIncome
780
650
520
390
260
130
0
0 600 1200 1800 2400 3000 3600 4200
Bristol-Myers Squibb
Abbott Laboratories
American Home Products
Eli Lilly
Pfizer
Schering-Plough
Warner-Lambert
Amgen
Baxter International
Pharmacia_Upjohn
Genentech
Chiron
Revenue
Company (All)
Category
Drug Delivery
Generic Drugs
Genomic Database
Large Biotech
Large Pharmaceuticals
Skin Care
Small Biotech
Tissue Engineering
Revenue 89.1 6232
NetIncome -87.7 1242.3
Company Pfizer
Category Large Pharmaceuticals
Revenue 3496
NetIncome 558
Shares 1303.1
PricePerSales 7.8
MarketCap 108971.7
TickerSymbol PFE
(3496, 558)
25 out of 166 records visible (15.06 %), 0 marked

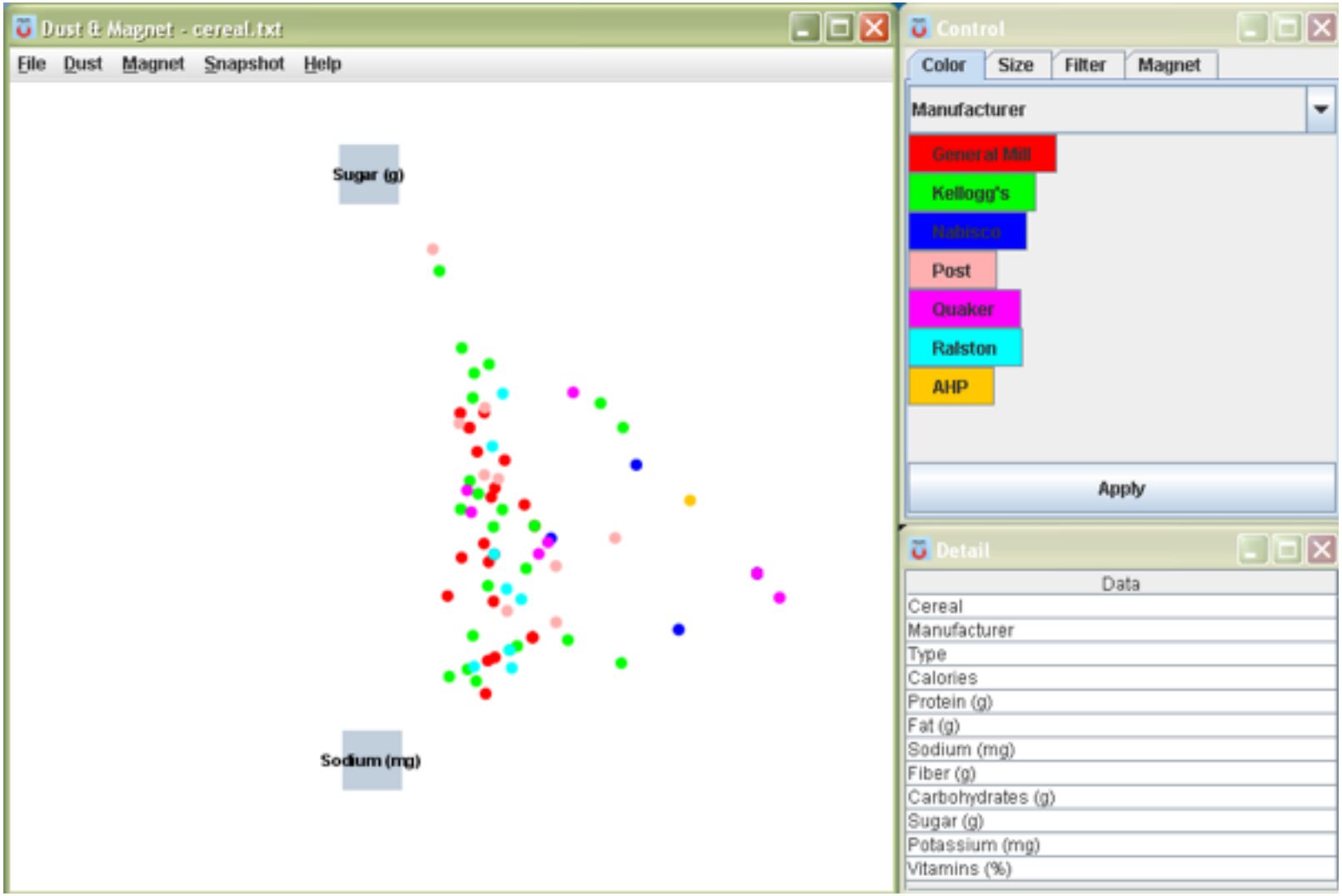
Data Visualization Sliders

- Vis in the controls…



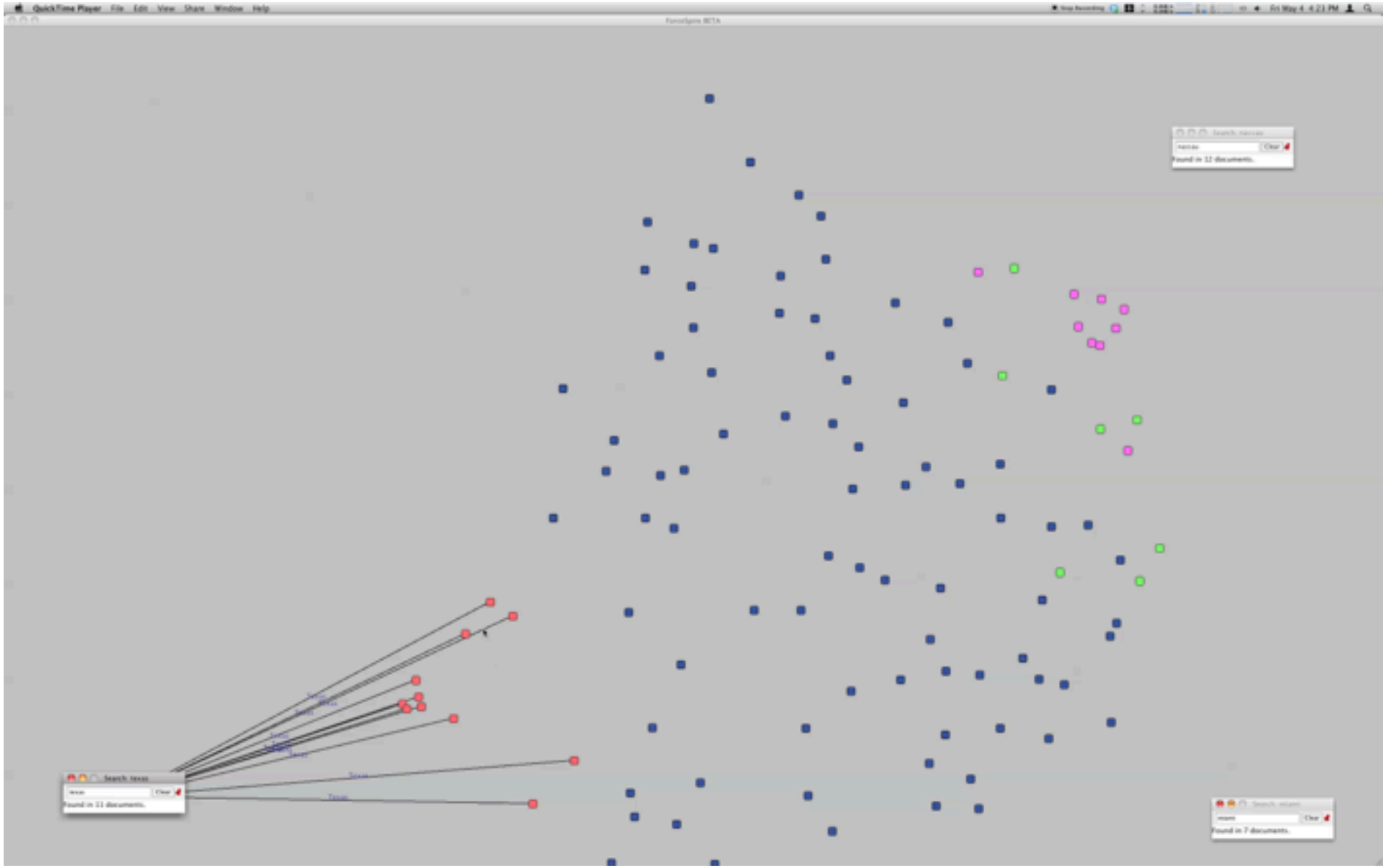
Dust & Magnet

- Controls in the visualization…



Yi et al Information Visualization '05

Semantic Interaction



7. Connect

“Show me related items”

- Highlight associations and relationships
- Show hidden data items that are relevant to a specified item

e.g.,

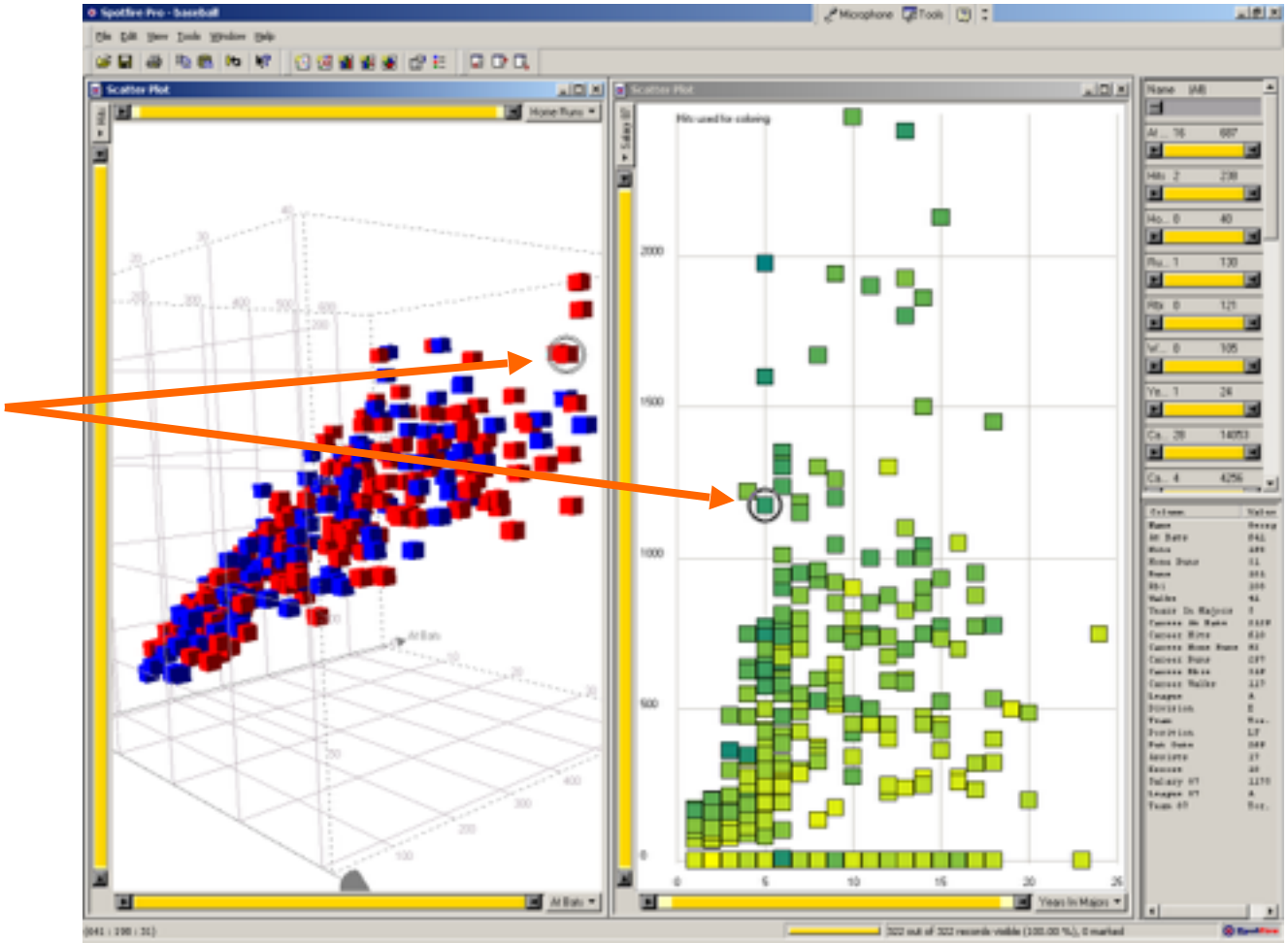
- Highlighting directly connected nodes
- Brushing across multiple views

Brushing

- Applies when you have multiple views of the same data
- Selecting or highlighting a case in one view generates highlighting the case in the other views
- Very common technique in InfoVis

Brushing

Same item



tasks

what do people do with visualizations?

Purpose of InfoVis

- Two main uses of InfoVis
  - **Analysis** – Understand your data better and act upon that understanding
  - **Presentation** – Communicate and inform others more effectively

1. Analysis

- Given all the data, then
  - understand, compare, decide, judge, evaluate, assess, determine, …
- Ultimately, it’s about solving problems

When (not?) to use vis

- Many other techniques for data analysis
  - Statistics, DB, data mining, machine learning
  - These give you explicit answers/results
  - E.g., the average temperature is 74.82 degrees

- Good when you know what questions to ask

When to use vis

- 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

- Should be open to let yourself discover, but also explain and answer specific questions

“A graphic display has many purposes but it achieves its highest value when it forces us to see what we were not expecting.”
H. Wainer

EDA Example 1

- Business
  - Why has Apple sold so many iPhones?
  - How much does their marketing campaign have to do with it?
  - Have their phones improved in terms of quality?
  - How do they compare to the alternatives?

EDA Example 2

- House Hunting
  - Why do you “like” some homes, and not others? # of bedrooms, baths, price /// smell? feel of area?
  - What will happen to the market? Is the home a “good deal”?

[interesting thought]

Could a data mining algorithm explicitly tell me what house to buy? What is the optimization function? Would you trust it?

Tasks for InfoVis

- Search (OK)
  - Finding a specific piece of information
  - “How far is it from LHR to Middlesex?”

- Browsing (Better)
  - Look over or inspect something in a more casual manner, seek interesting information
  - “What are the tradeoffs between different routes and transportation options to get from LHR to Middlesex?”
  - “What are some interesting articles in the news today?”

Tasks for InfoVis

- Analysis
  - Comparison-Difference
  - Outliers, Extremes
  - Patterns
- Monitoring - what has changed? what changes matter?
- Awareness

2. Presentation

- Use visualization to communicate ideas, influence, explain, persuade
- Visuals can serve as evidence or support
- They can accompany other medium for presentation (text, video, etc.)

Key Benefits of Visualization

- Facilitating awareness and understanding
- Helping to raise new questions and supply answers
- Generating insights
- Telling a story and making a point

Key Challenge

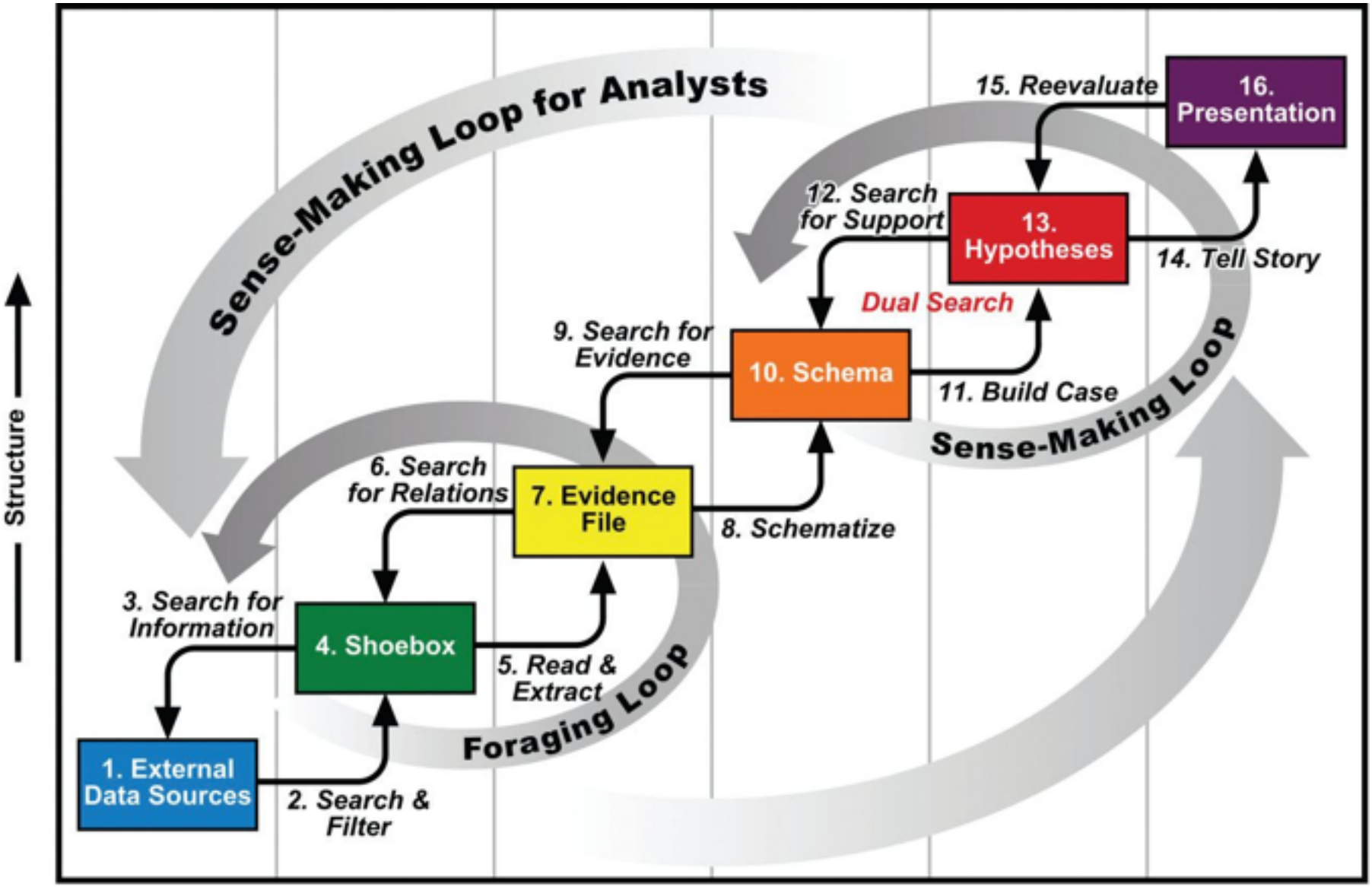
- How to measure and prove?
  - All those benefits are not easily quantifiable and measured
- Evaluation is an open research challenge for visualization
- For (2) Presentation, user interaction can be a challenge

tasks, at a higher-level

- we have models the describe the higher-level cognitive tasks
- formation of mental models
- will only go into this briefly, y'all will hear more from William Wong about this later

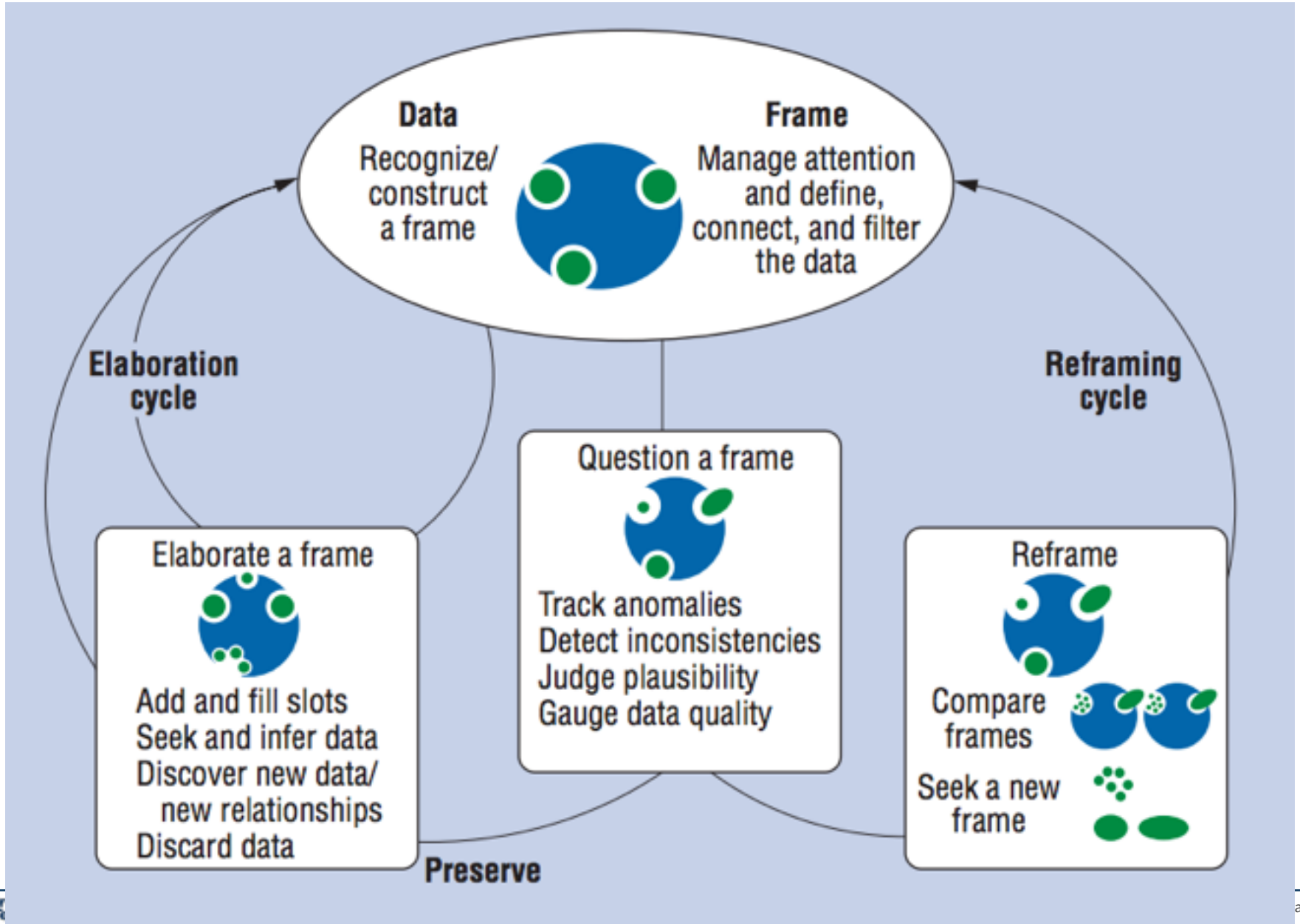
Sensemaking Loop

Pirolli & Card, 2001



Data-Frame Model

Klein et al., 2006



so is that it?

- Traditional visualization has several advantages
  - focuses include novel representations, bindings between data and visual glyphs

- remember, visual analytics advances the science further by creating interactive visual interfaces of data that **integrate computation**

recap

- VA leverages well-established principles in perception and cognition for data transfer
  - similar to InfoVis
- **interaction** transforms viewers into “users”
- these people perform **tasks**, both low-level and high-level
  - low: zoom, filter, select, …
  - high: sensemaking, analysis, …

after the break

- how do we incorporate data analytics, statistics, machine learning?
  - make interactive VA even more powerful
- a few examples and demos
- during the break, think about some questions to ask at the end

putting it all together (after coffee):

visual analytics:

coupling computation with interactive visualization

The InfoVis Pipeline

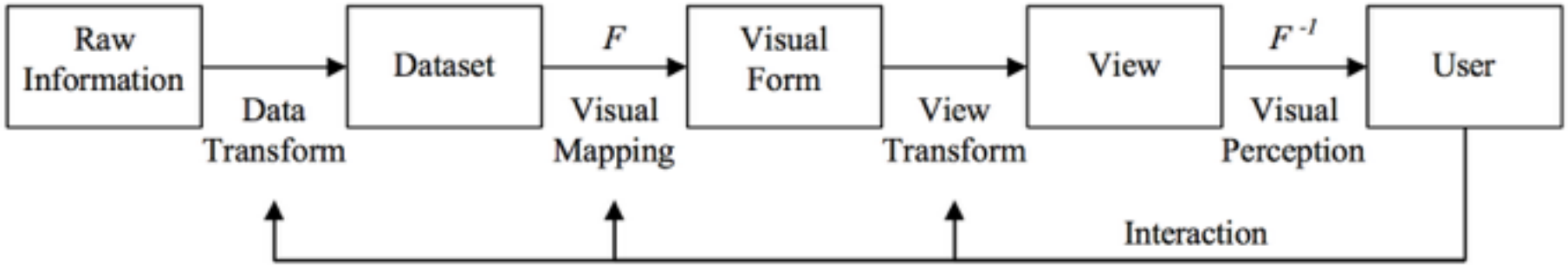
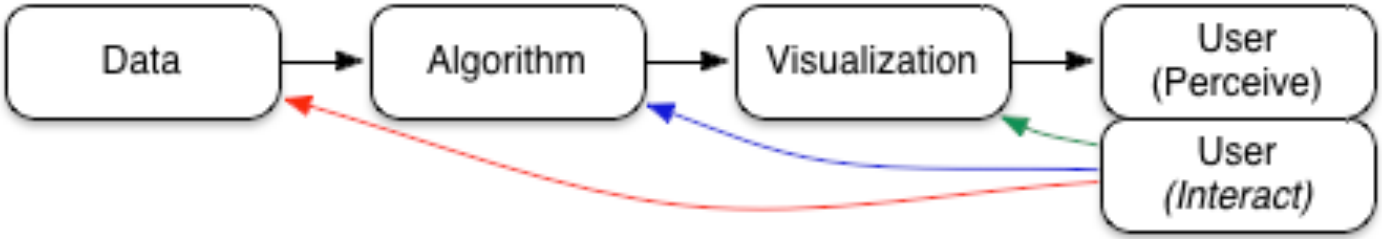


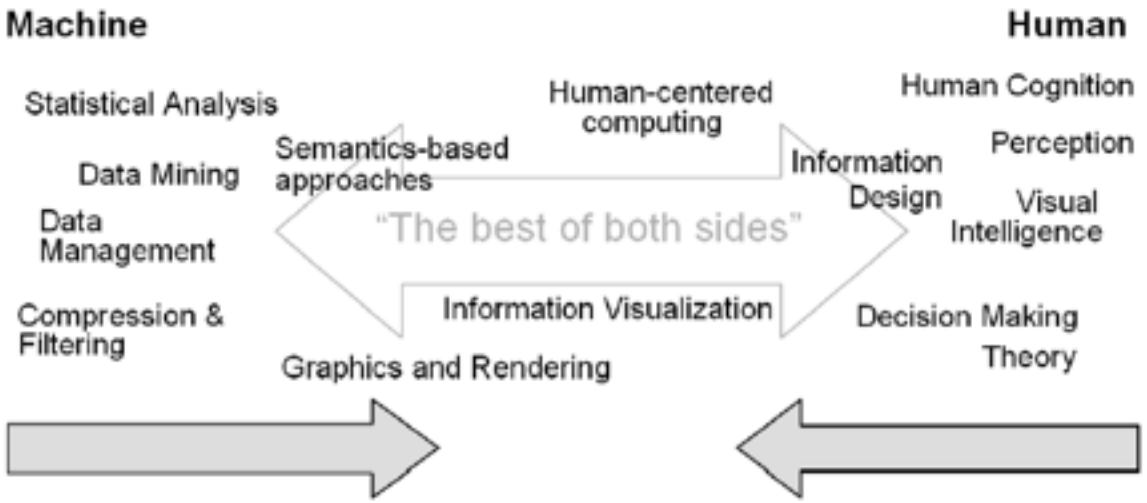
Figure 2: The visualization pipeline, converting information into interactive visual representations (adapted from [Card et al., 1999]).

Visual Analytics Pipeline



Synergy

- Combine strengths of both human and electronic data processing
  - Gives a semi-automated analytical process
  - Use strengths from each



From Keim

Integrating Data Mining and ML

- Dimension Reduction
  - PCA, MDS, …
- Feature Selection
  - Entity extraction, …
- Similarity functions
  - “find more like this”, …
- Prediction
  - will the stock price go up or down?

First, let's talk data

- **Dimensionality** of data is an important deciding factor for which vis technique to use
  - a.k.a. features, attributes, variables
- Dimensionality can lead to complexity of data
- Recall, increased dimensionality becomes more difficult to visualize

Data

- Data is taken from and/or representing some phenomena from the world
  - important distinction: what data *is* vs what data *approximates*
- Data describes something of interest to us
- Can be sensed, manually created, machine-produced, …

Example

- Cars
  - make
  - model
  - year
  - miles per gallon
  - cost
  - number of cylinders
  - weights
  - ...

Data Models

- Often characterize data through three components
  - **Items**

    Items of interest

    (students, courses, terms, …)
  - **Attributes**

    Characteristics or properties of data

    (name, age, GPA, number, date, …)
  - **Relations**

    How two or more objects relate

    (student takes course, 2 ppl friends on Facebook, …)

Data Tables

- We take raw data and transform it into a model/form that is more workable
- Main idea:
  - Individual items are called cases or objects
  - Cases have variables (attributes)
  - Relational: Relations between cases (not our main focus today)

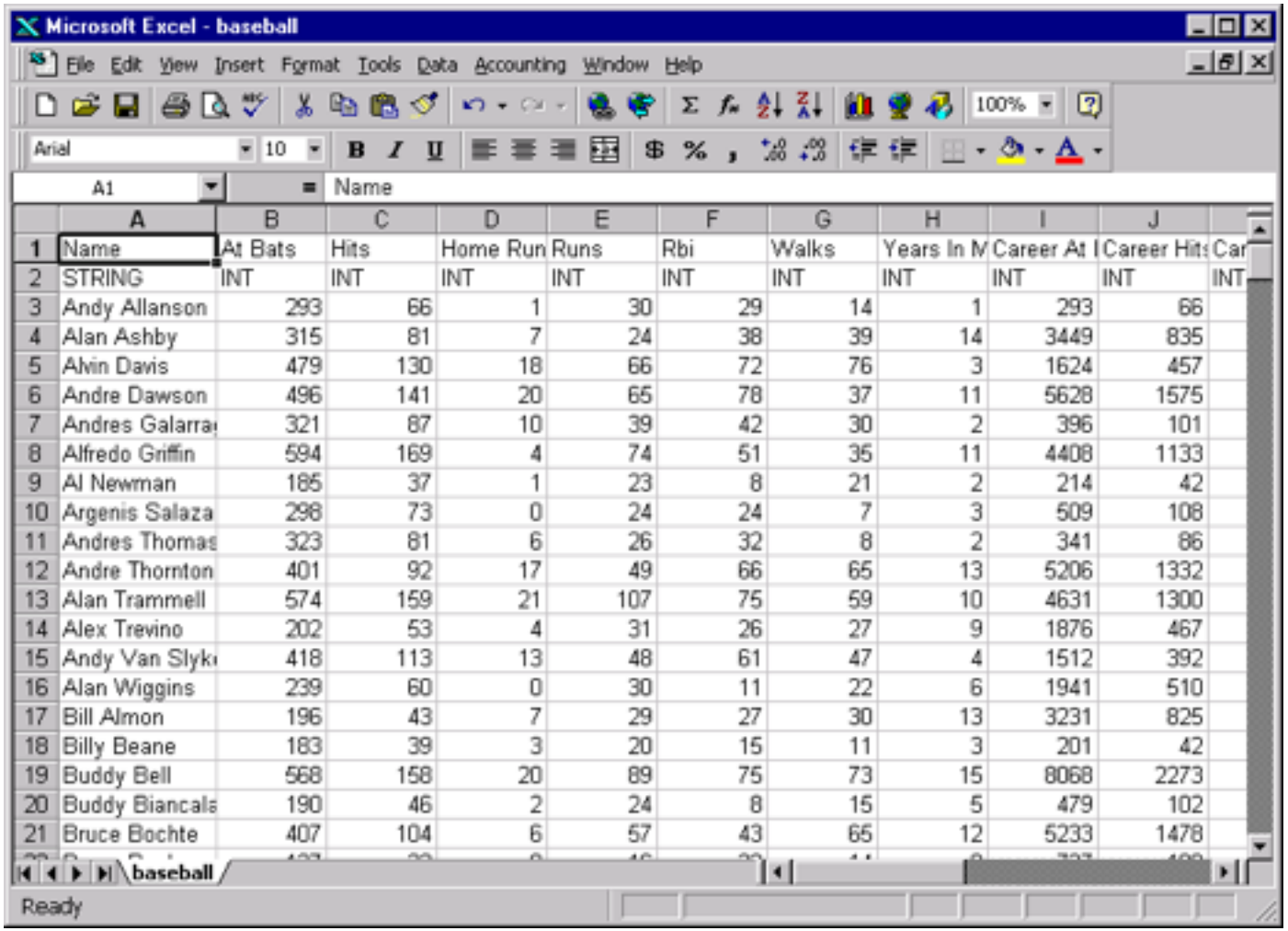
Data Table Format

	Case1	Case2	Case3	...
Variable1	Value11	Value21	Value31	
Variable2	Value12	Value22	Value32	
Variable3	Value13	Value23	Value33	
...				

Example

	Mary	Jim	Sally	Mitch	...
SSN	145	294	563	823	
Age	23	17	47	29	
Hair	brown	black	blonde	red	
GPA	2.9	3.7	3.4	2.1	
...					

E.g., Spreadsheets and Tables



Microsoft Excel - baseball

File Edit View Insert Format Tools Data Accounting Window Help

Arial 10 B I U

A1 = Name

	A	B	C	D	E	F	G	H	I	J	
1	Name	At Bats	Hits	Home Run	Runs	Rbi	Walks	Years In M	Career At I	Career Hits	Car
2	STRING	INT	INT	INT	INT	INT	INT	INT	INT	INT	INT
3	Andy Allanson	293	66	1	30	29	14	1	293	66	
4	Alan Ashby	315	81	7	24	38	39	14	3449	835	
5	Alvin Davis	479	130	18	66	72	76	3	1624	457	
6	Andre Dawson	496	141	20	65	78	37	11	5628	1575	
7	Andres Galarra	321	87	10	39	42	30	2	396	101	
8	Alfredo Griffin	594	169	4	74	51	35	11	4408	1133	
9	Al Newman	185	37	1	23	8	21	2	214	42	
10	Argenis Salaza	298	73	0	24	24	7	3	509	108	
11	Andres Thomas	323	81	6	26	32	8	2	341	86	
12	Andre Thornton	401	92	17	49	66	65	13	5206	1332	
13	Alan Trammell	574	159	21	107	75	59	10	4631	1300	
14	Alex Trevino	202	53	4	31	26	27	9	1876	467	
15	Andy Van Slyk	418	113	13	48	61	47	4	1512	392	
16	Alan Wiggins	239	60	0	30	11	22	6	1941	510	
17	Bill Almon	196	43	7	29	27	30	13	3231	825	
18	Billy Beane	183	39	3	20	15	11	3	201	42	
19	Buddy Bell	568	158	20	89	75	73	15	8068	2273	
20	Buddy Biancala	190	46	2	24	8	15	5	479	102	
21	Bruce Bochte	407	104	6	57	43	65	12	5233	1478	

baseball

Ready

Attribute Types

- Three main types of variables
  - N-**Nominal** (equal or not equal to other values)

    Example: gender
  - O-**Ordinal** (obeys < relation, ordered set)

    Example: fr,so,jr,sr / grande < venti / small, medium, large shirts
  - Q-**Quantitative** (can do math on them)

    Example: age, weight

How Many Dimensions?

- Data sets of dimensions 1, 2, 3 are common
- Number of variables per class
  - 1 - Univariate data - bar chart?
  - 2 - Bivariate data - scatterplot?
  - 3 - Trivariate data - 3d scatterplot? color+scatterplot?
  - >3 - Hypervariate data — also see a lot of these

- For hypervariate, VA often needs help from computation

Hypervariate Data

- How about 4 to 20 or more attributes (for instance)?
  - Many data sets fall into this category

- Many datasets consist of **thousands of attributes**

- **what do we do?**

should I use a table or a visual?

- Tables are good when:
  - direct textual representation
  - important to know individual values
  - precise values are required
  - comparison across precise, individual values are needed
- Graphs are good when:
  - symbolic
  - when the shape of is important
  - when you need to look at relationships between data items

graphs

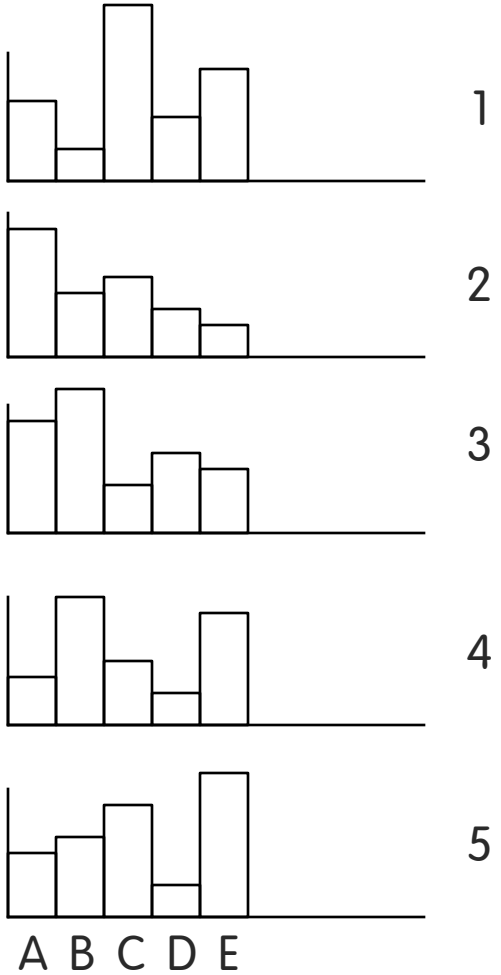
- visual display that illustrates relationships among entities
- shorthand notation

- considerations:
  - why are we trying to show a graph?
  - what are the questions we want people to be able to answer?
  - who is the audience?

another approach, **Multiple Views**

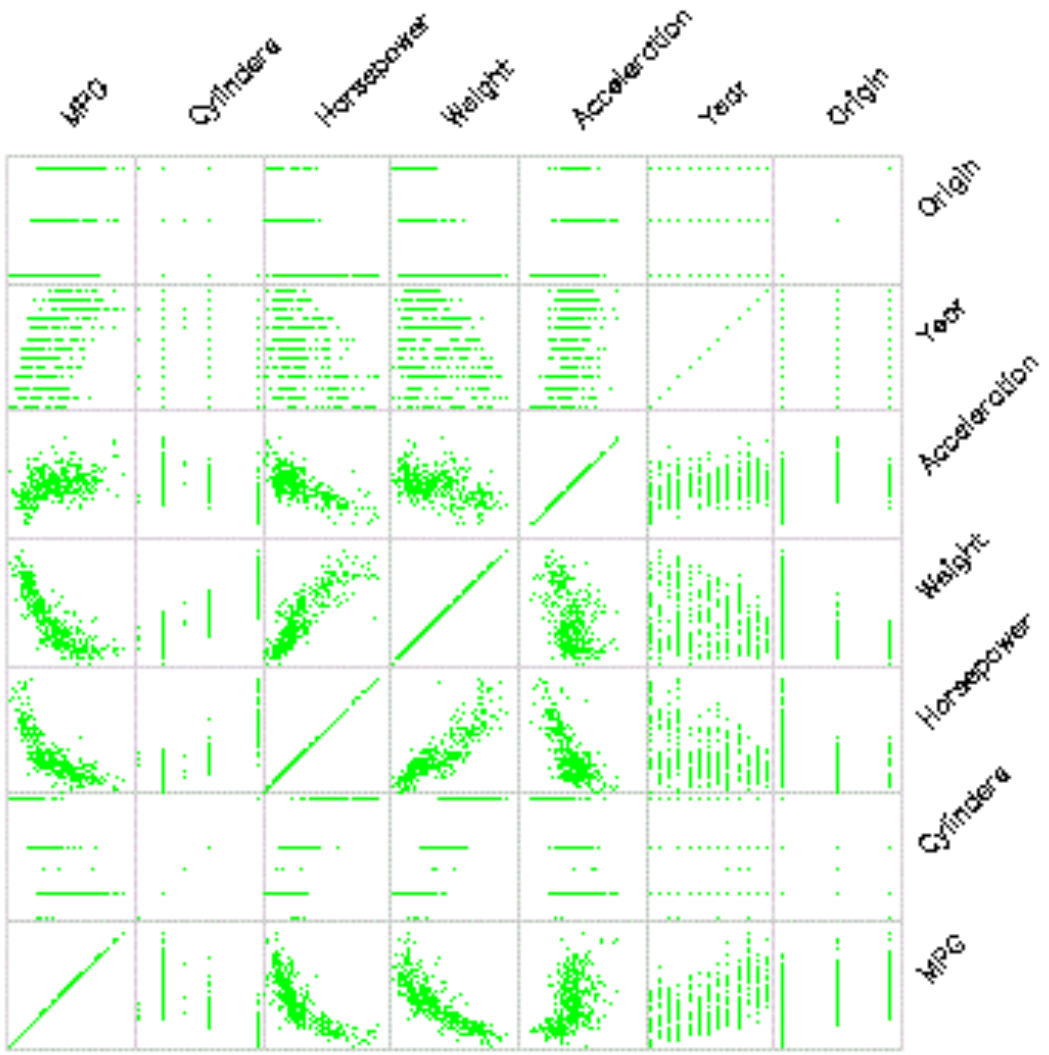
Give each variable its own display

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5
5	3	4	5	1	7

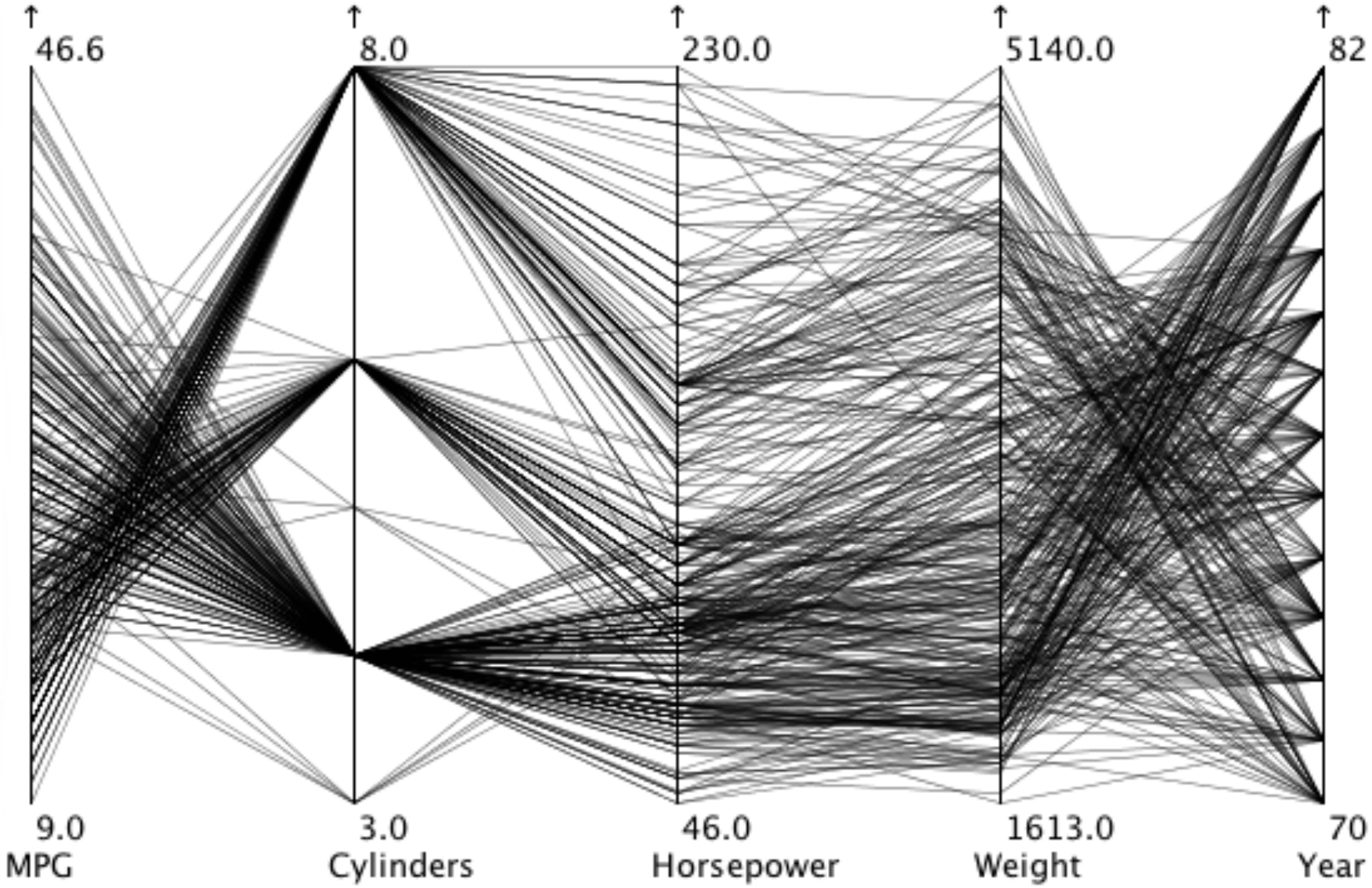


Scatterplot Matrix

Represent each possible pair of variables in their own 2-D scatterplot



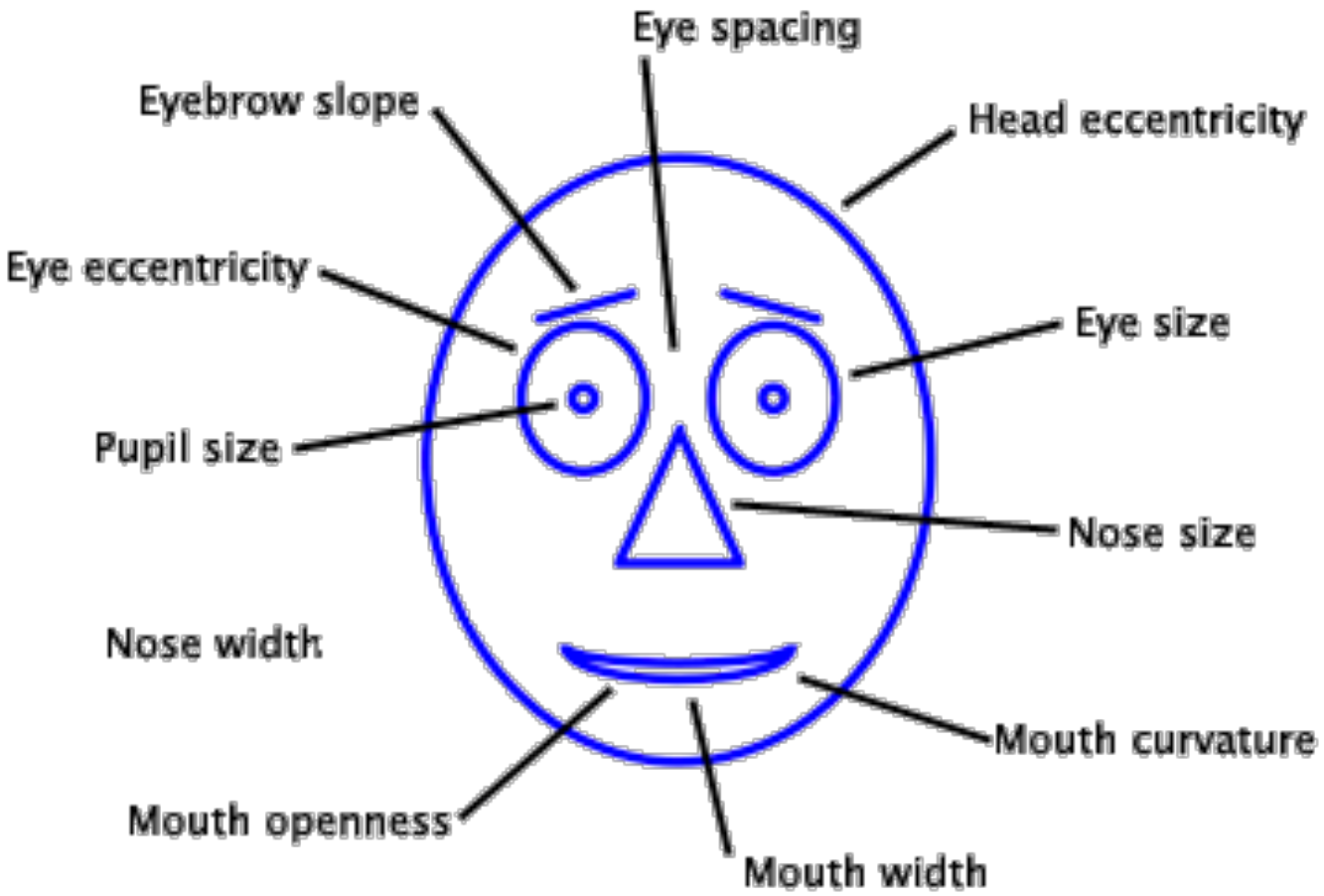
Parallel Coordinates



Parallel Coordinates

- Issues exist
  - Overplotting
  - Hard to trace data point (line) across all dimensions

a fun one - Chernov faces



or, use **Dimension Reduction!**

- computational techniques to solve for similar attributes so that the final number or attributes used to describe the dataset are less than the original
  - this is lossy

- for VA, we often reduce from n, to 2
  - we can show 2 dimension in a scatterplot
  - which attributes do we discard? we can't show them all!
  - need to find most salient data attributes and show those

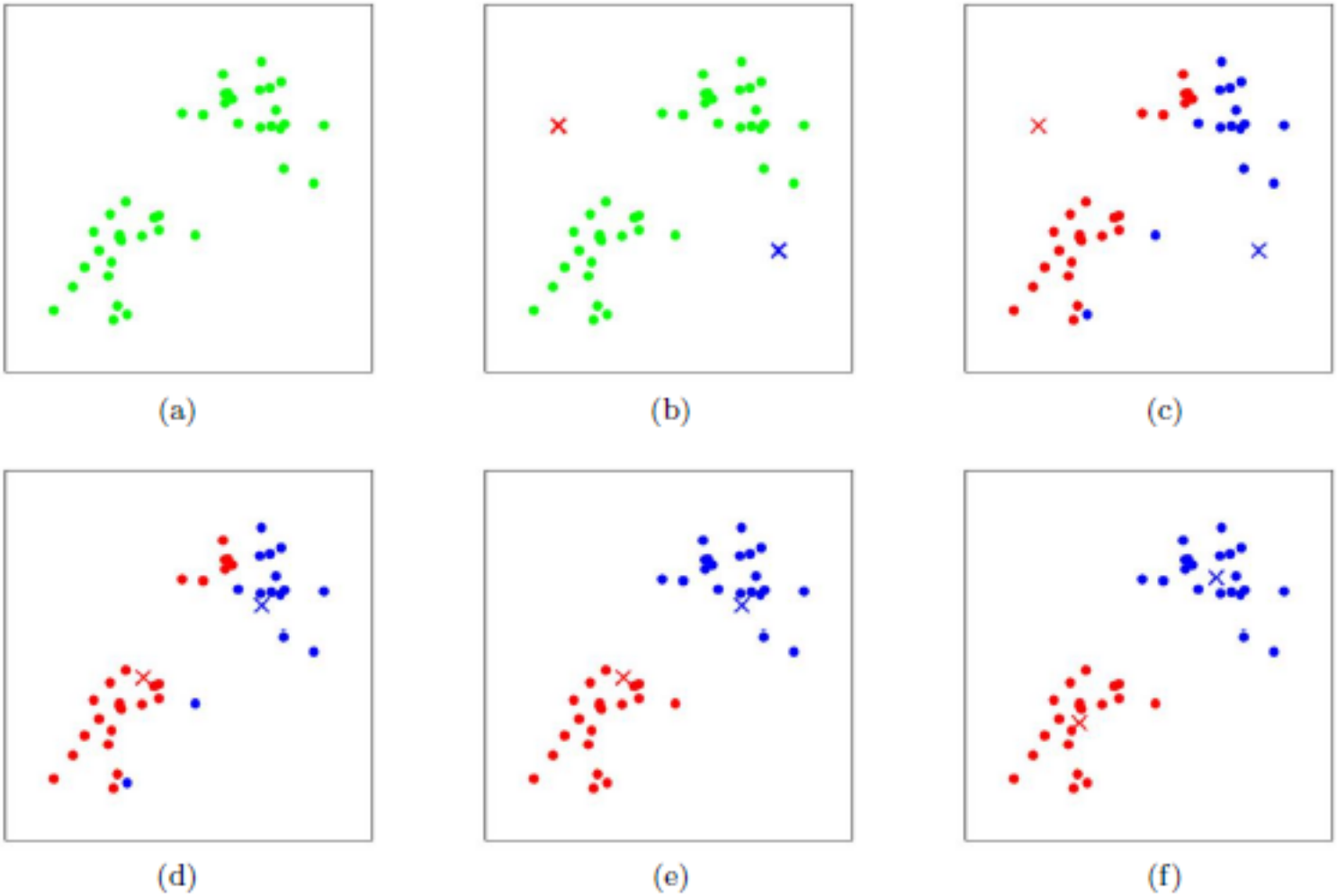
Principal Component Analysis (PCA)

- Frequently used deterministic technique
- transform data table with lots of attributes to principal components, which can be mapped in pairs to x and y axes of scatterplots
  - computed PCs will be <= original attributes of data
- solve for maximum variance

potential PCA limitations

- no guarantee that resulting PCs are “meaningful”, or most important
  - importance is a human, subjective metric
  - max variance may not be most meaningful
- sensitive to outliers
- linear function, what if data/phenomena is non-linear?

k-means clustering



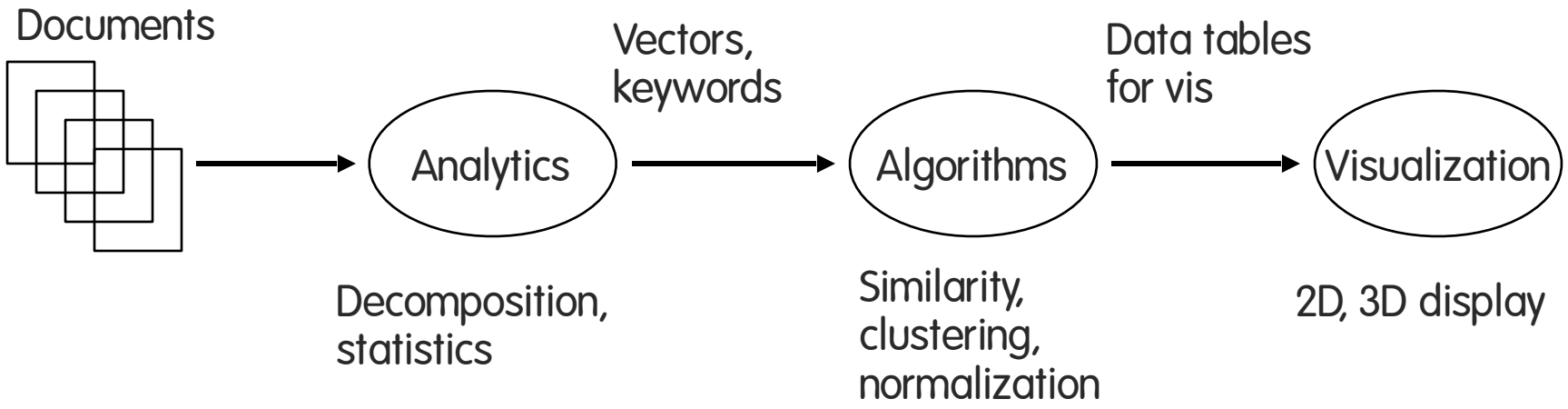
given k, average and move centroid until stable

k-means clustering

- goals is to partition our data points into k clusters
- potential drawback/tradeoff
  - have to define k (how many clusters do you want?)
  - uses Euclidean distance for computation
- PCA and k-means are good, fundamental computations to start with
  - lots and lots of others, and derivatives of those, and combinations of those, …

let's walk through an example

- Text data can be represented as hypervariate data



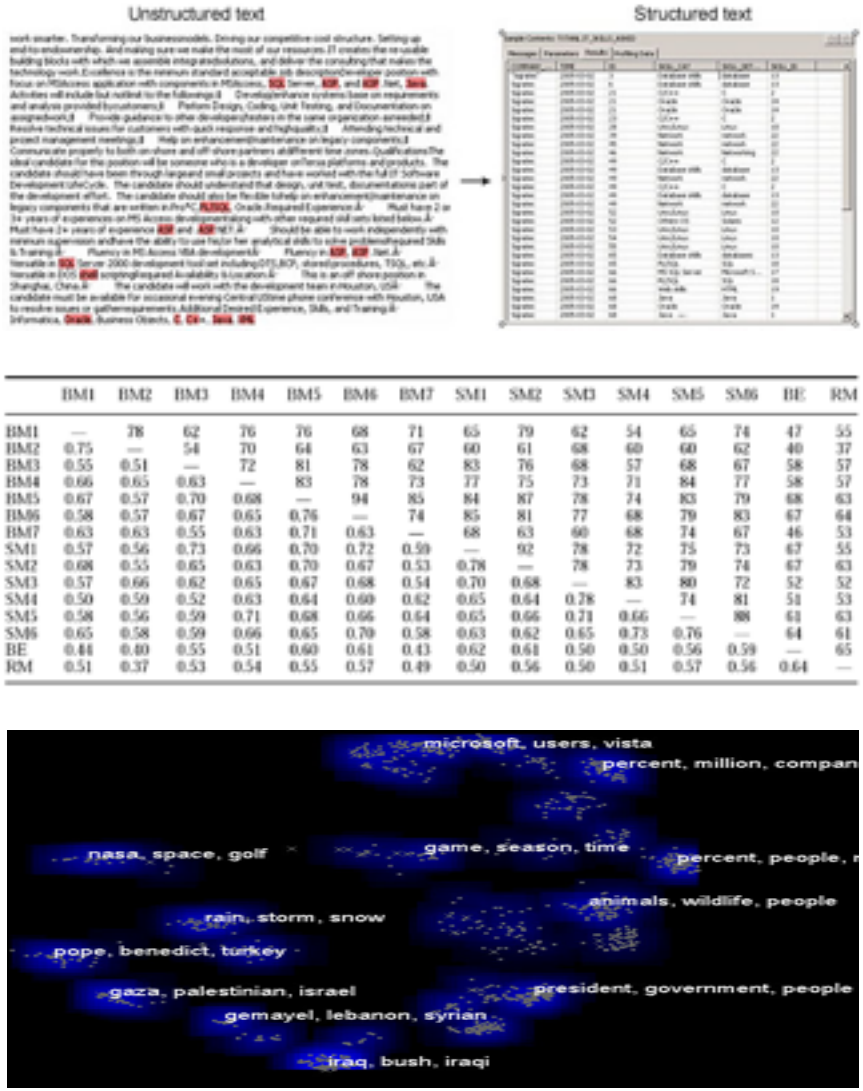
Unstructured Text

Extract Entities

High Dimensional Data

Dimension Reduction Model

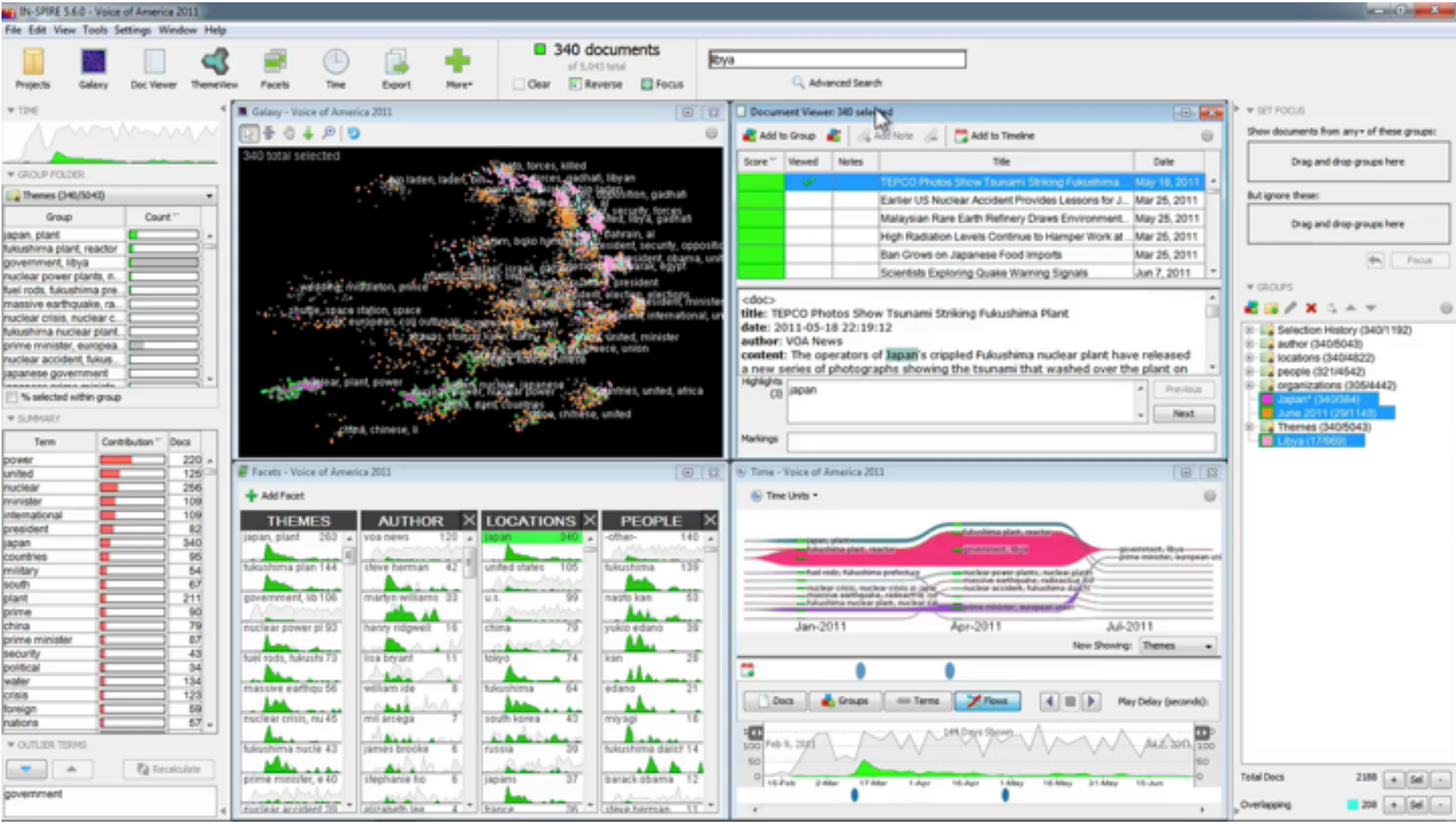
Choice of Visual Encodings/ Metaphor



tf-idf for entity selection/extraction

- term frequency–inverse document frequency
  - TF(t) = (num times term t appears in a document) / (num of terms in the document)
  - IDF(t) = log(Total num of documents / num of documents with term t in it)
  - Then, TF x IDF
- commonly used way to apply structure and weighting to unstructured text
  - with structure, we can compute and visualize

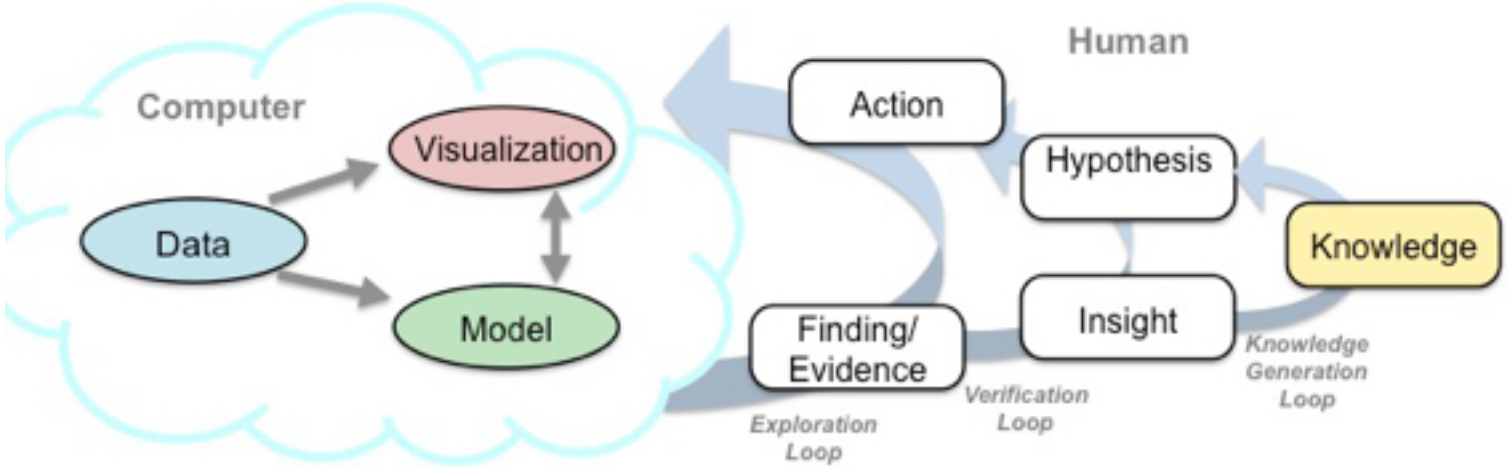
IN-SPIRE



Wise, James A., et al. "Visualizing the non-visual: spatial analysis and interaction with information from text documents." Information Visualization, 1995. Proceedings.. IEEE, 1995.

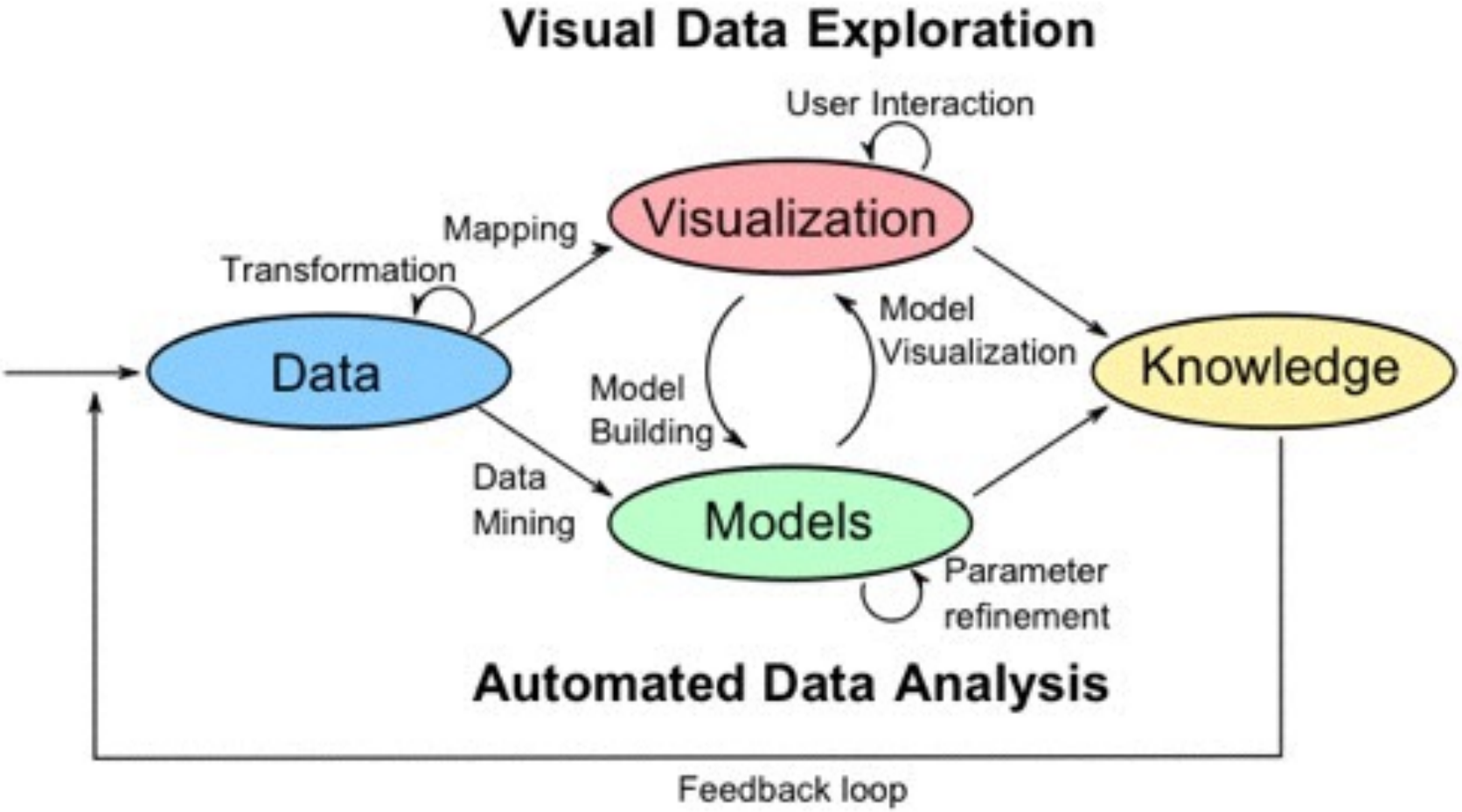
incorporating user feedback

- unsupervised computation is rarely the “final” solution
- “human-in-the-loop” design advocates for including user feedback in the holistic system
- people have domain expertise, can balance subjective tradeoffs, etc.
- let’s look at a few examples of user interaction in VA



Sacha, Dominik, et al. "Knowledge generation model for visual analytics." Visualization and Computer Graphics, IEEE Transactions on 20.12 (2014): 1604-1613.

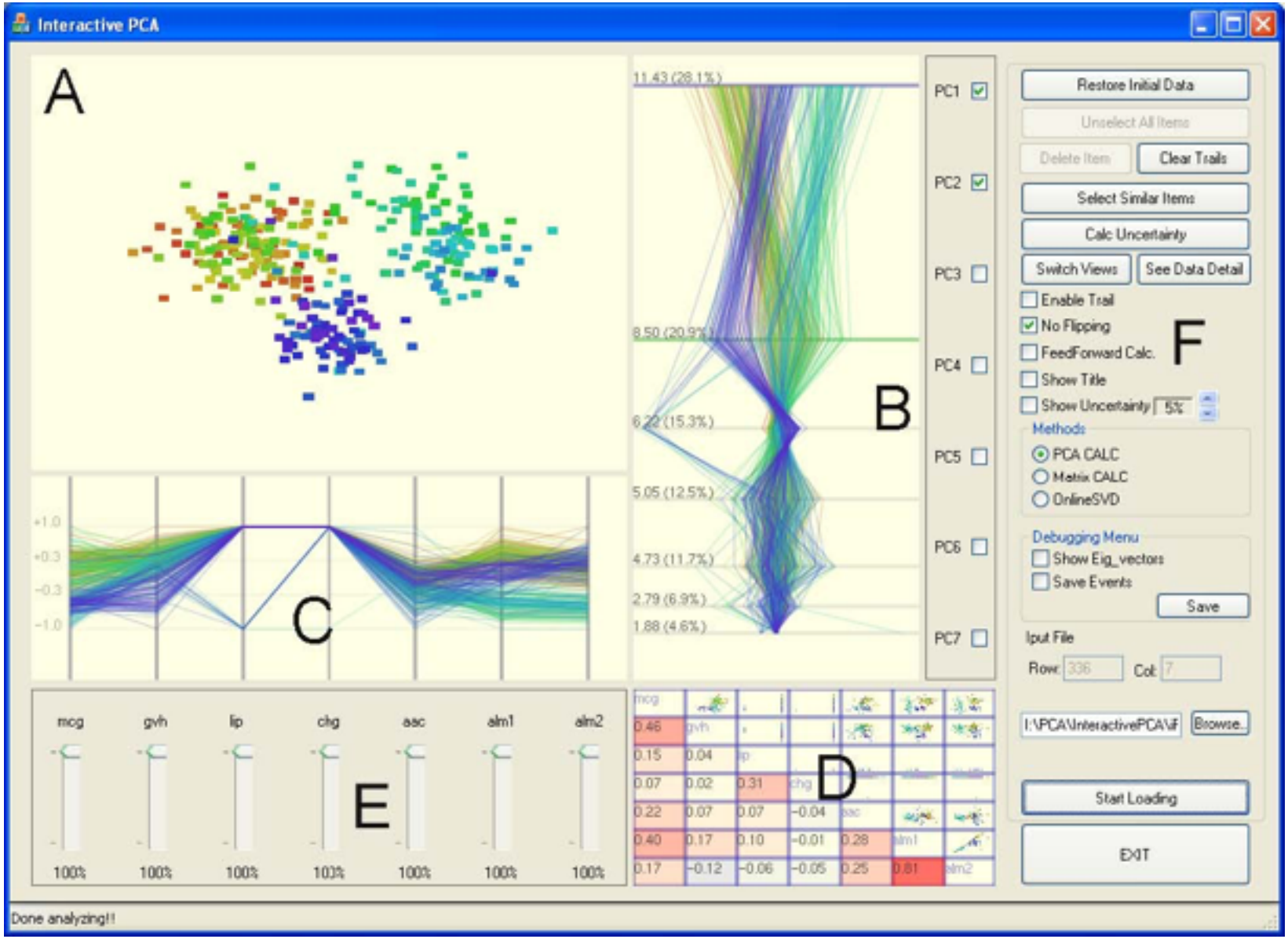
incorporating user feedback



"Mastering the Information Age Solving Problems with Visual Analytics",edited by Daniel Keim, Jörn Kohlhammer,Geo rey Ellis and Florian Mansmann

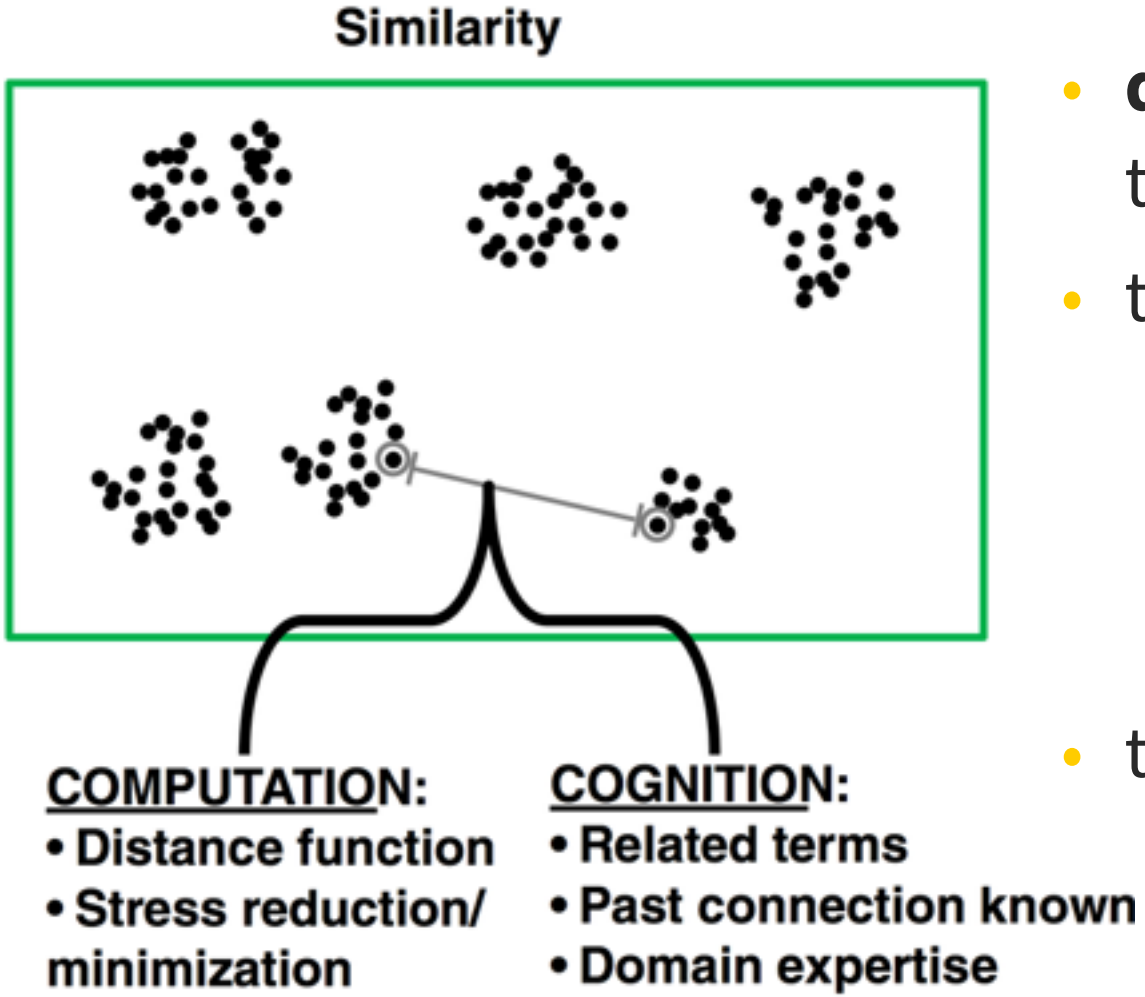
iPCA

https://youtu.be/oUx-7Hca_i4?t=1m30s



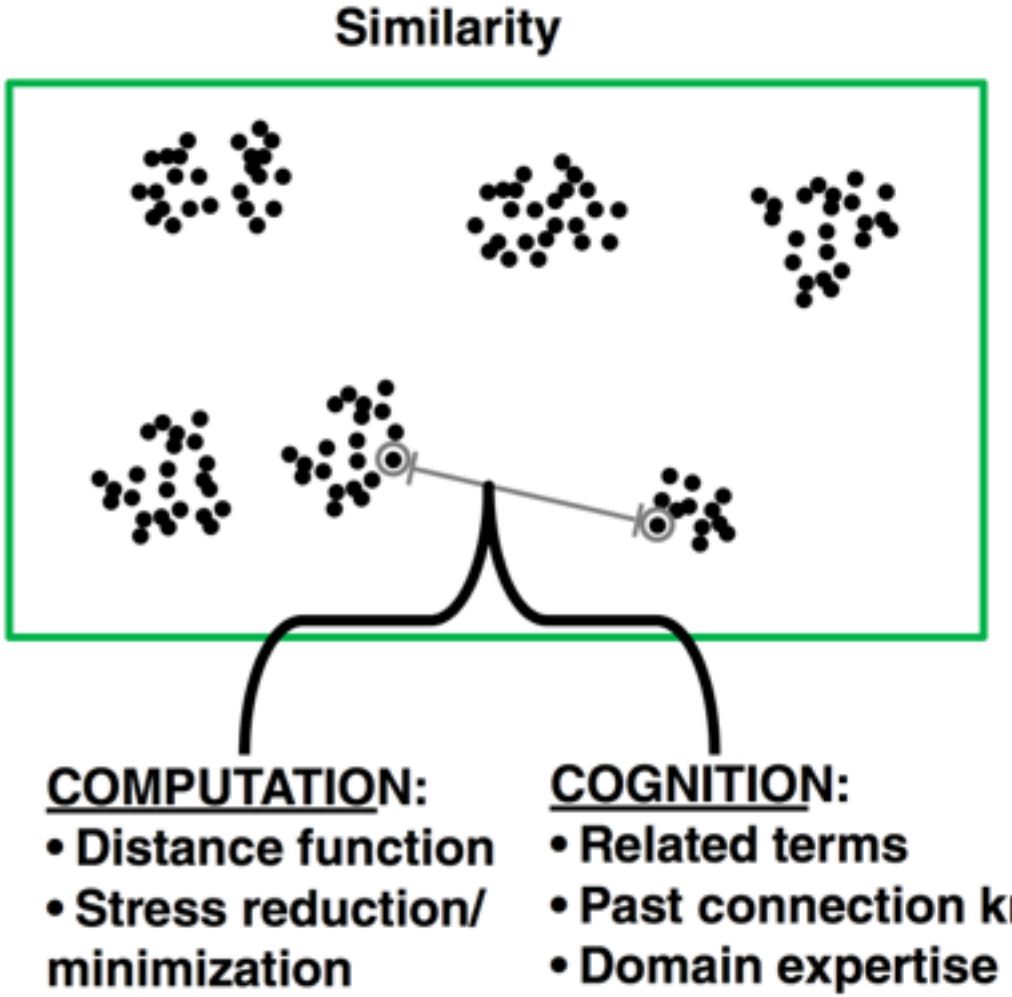
Jeong et al.
CG&A 2009

coupling mental models & data models



- **distance** has meaning to the system and the user
- to the **system**:
  - uses distance functions and similarity metrics
  - preserves distance in high-dimensional feature space in the low-dimensional scatterplot
- to the **user:**
  - proximity approximates similarity
  - we put things nearby if they have some conceptual similarity

coupling mental models & data models



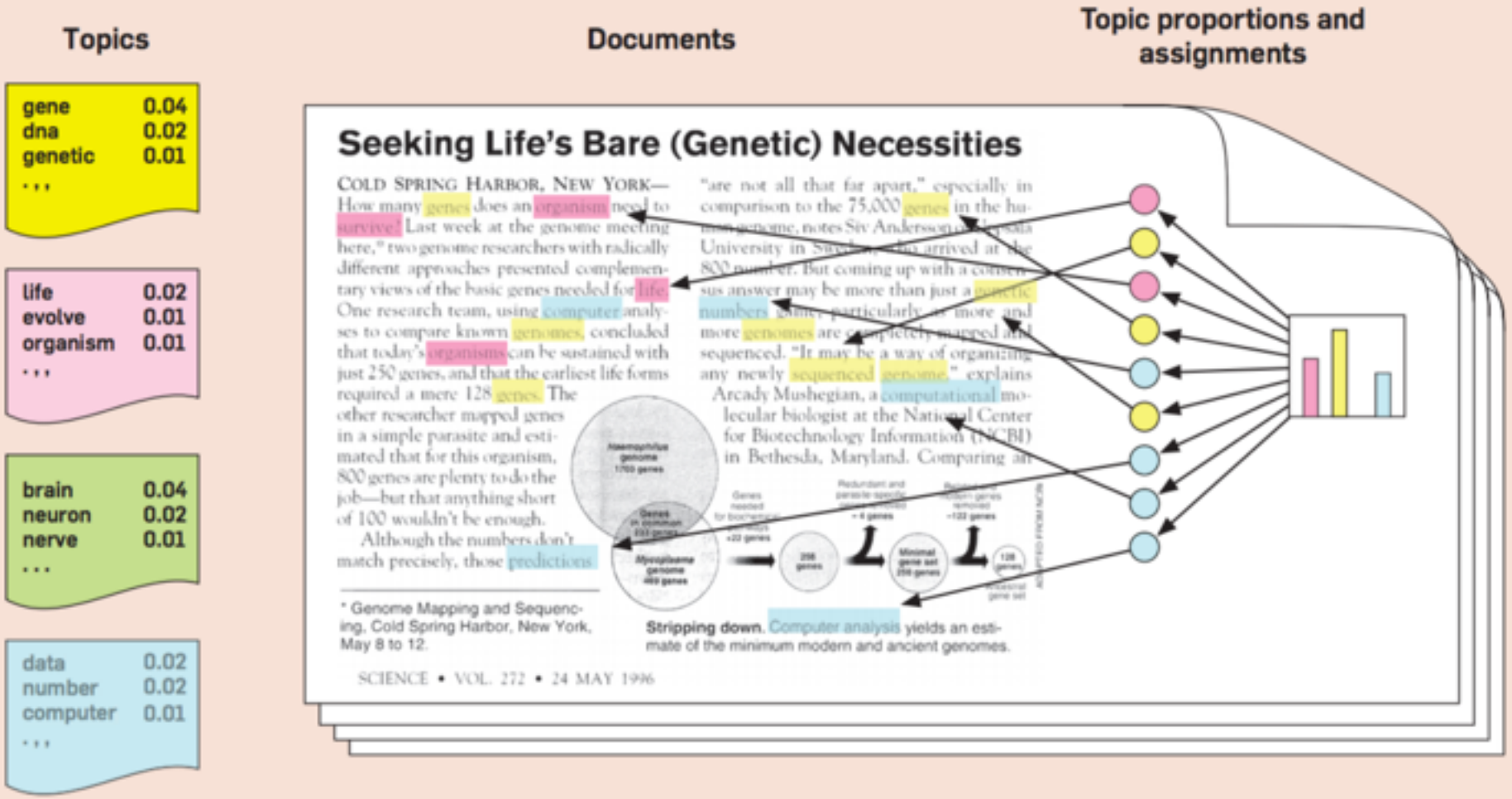
- opens an interesting avenue of research
  - how do we take advantage of this connection?
- we'll get back to this later if we have time!

Space to Think



Andrews, Christopher, Alex Endert, and Chris North. "Space to think: large high-resolution displays for sensemaking." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010.

topic modeling



Blei, David M. "Probabilistic topic models." Communications of the ACM 55.4 (2012): 77-84.

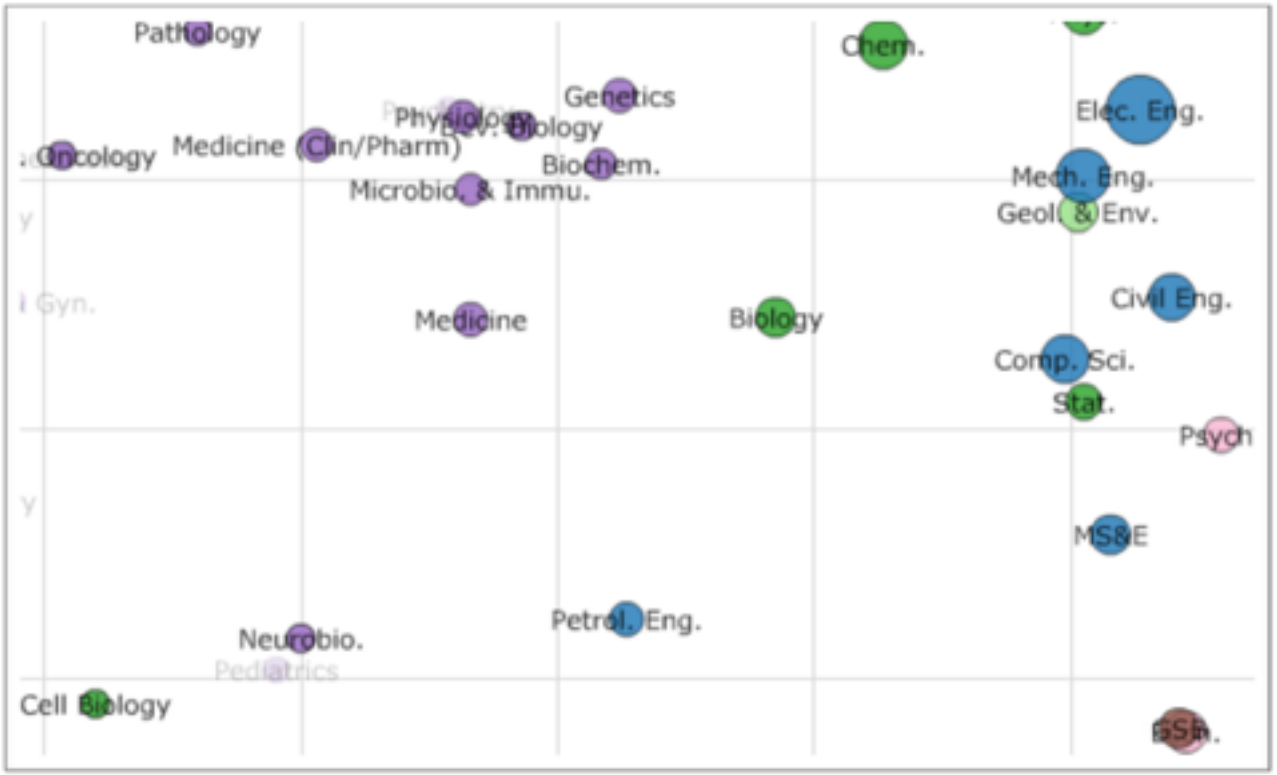
Serendip



Alexander, Eric, et al. "Serendip: Topic model-driven visual exploration of text corpora." Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on. IEEE, 2014.

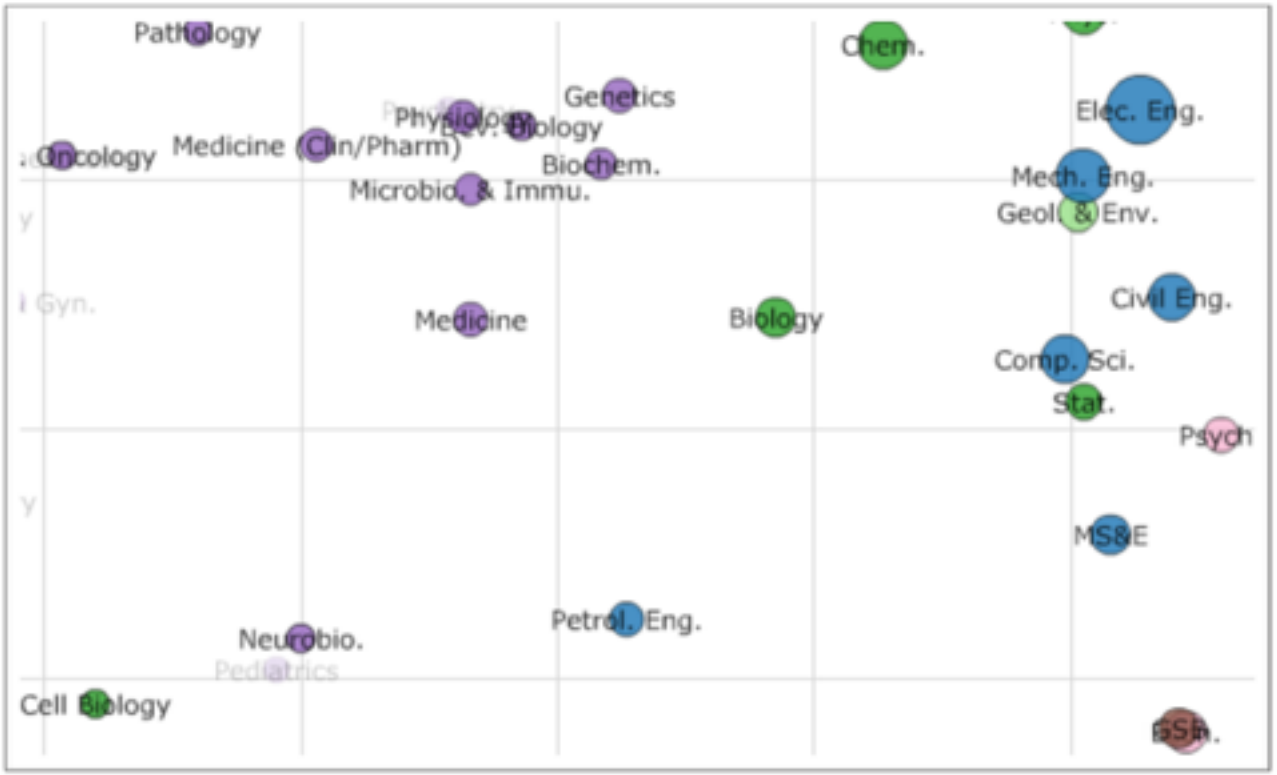
potential challenges

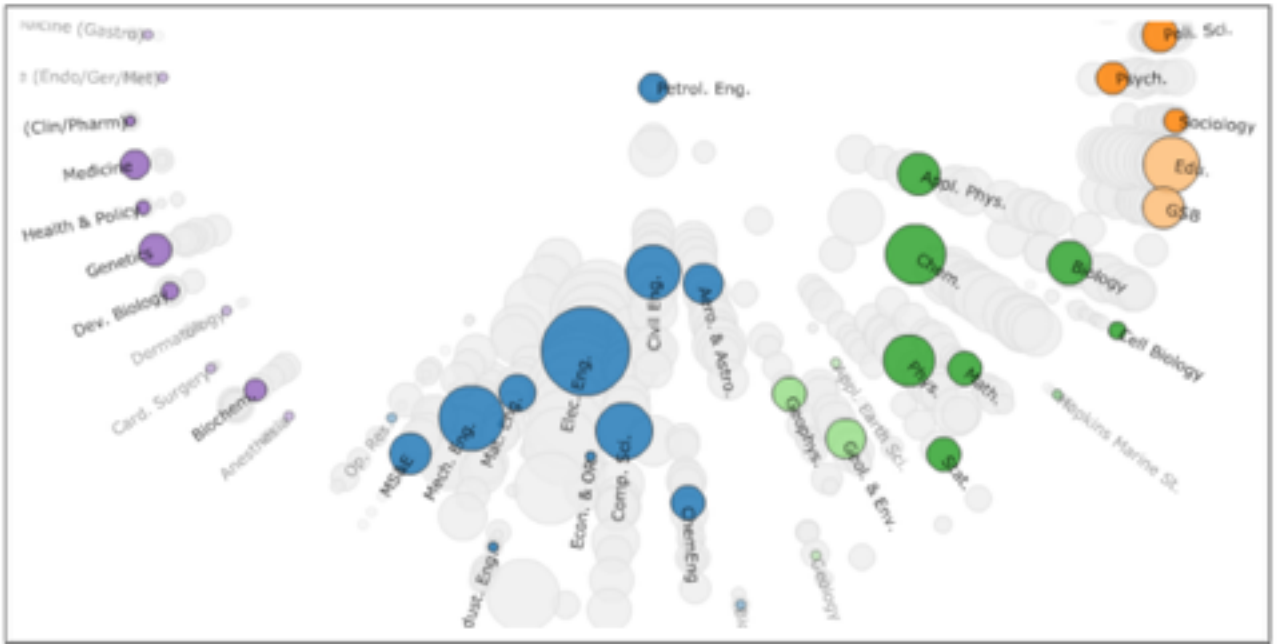
- **interpretability** and **trust**
- computation **approximates** data characteristics and phenomena, but sometimes get is wrong
- good case study by Chuang et al.



DR-produced view has distortions that can create visual artifacts

Petrol Eng. is similar to Neurobiology?





DR-produced view has distortions that can create visual artifacts

Petrol Eng. is similar to Neurobiology?

transformation of view based on user selection can remove these

including ML seems to work well

- helps offload computationally-attractive tasks to machines - nice!
- frees up user to focus cognitive energy on other things - also good

- but we can do better!

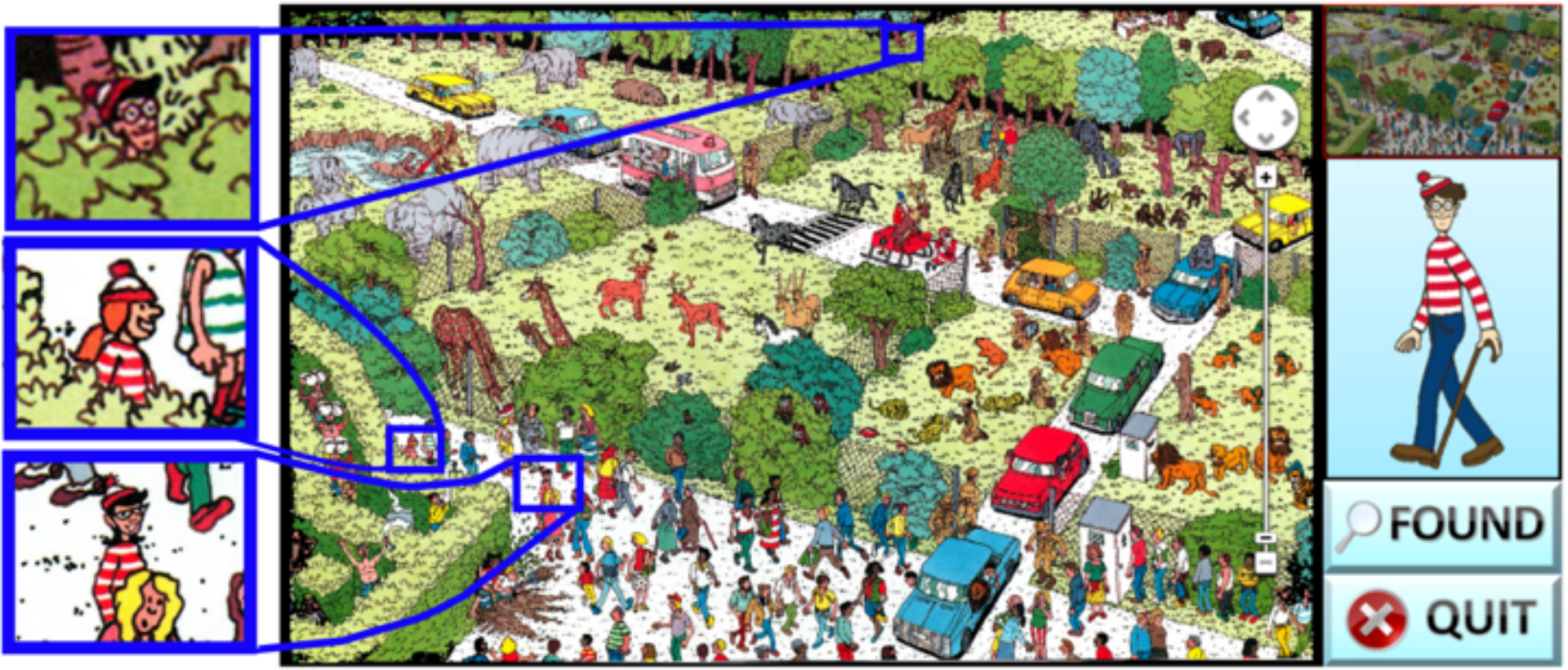
warning! \begin{alex_rant}

- do not treat user interaction as an easy way out
  - “oh, we’ll just give the user a menu for that!”
- do not make your problems (as developers and designers) the user’s problems
- interaction should be meant to explore data, ask questions, gain insight

user interaction is valuable data

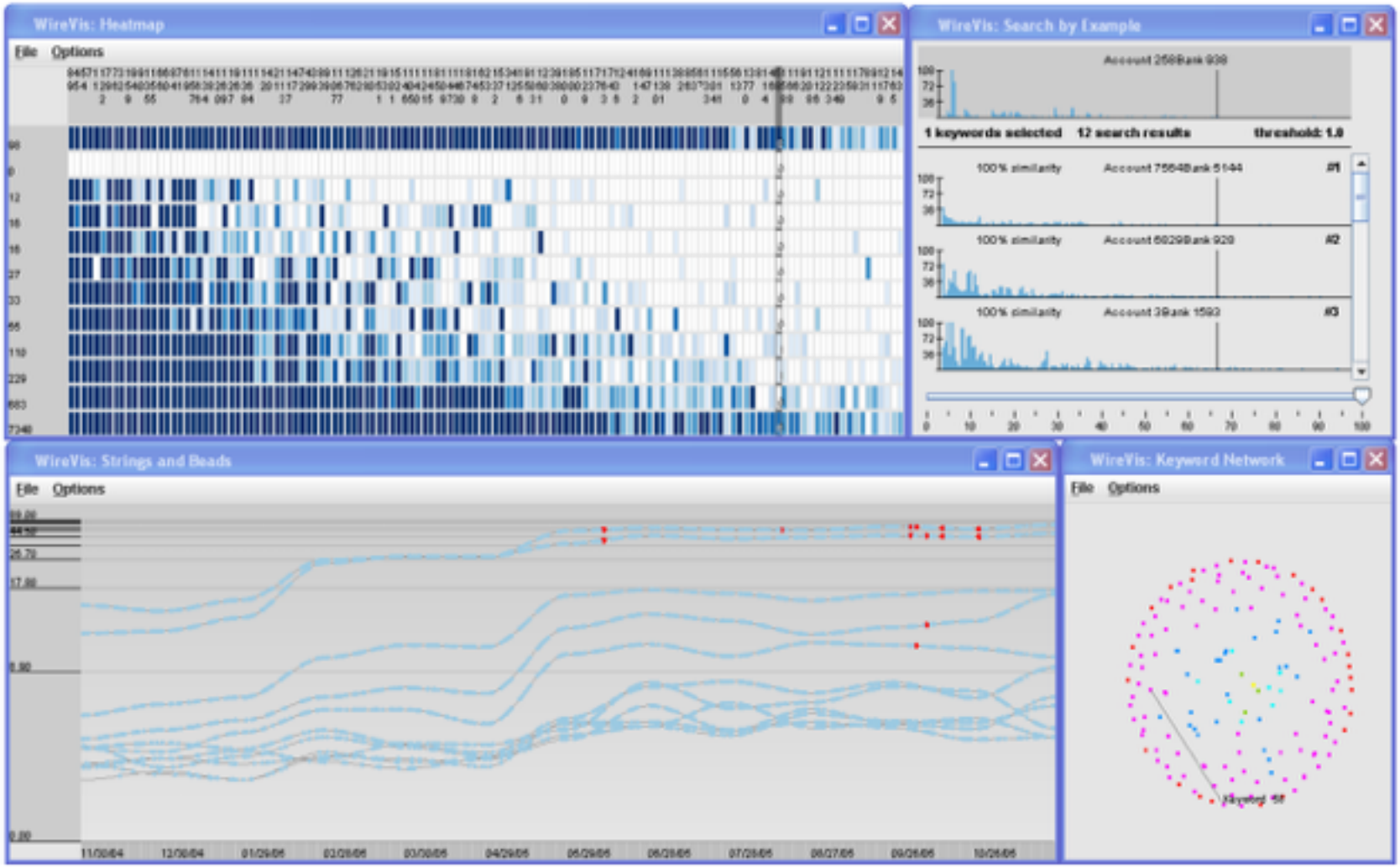
- quantitative record of analysis and exploration process
- what can we learn from this data?
  - data models? sensemaking stages?
- “interaction is the insight”
- let’s look at a few examples

are you good at Where's Waldo?



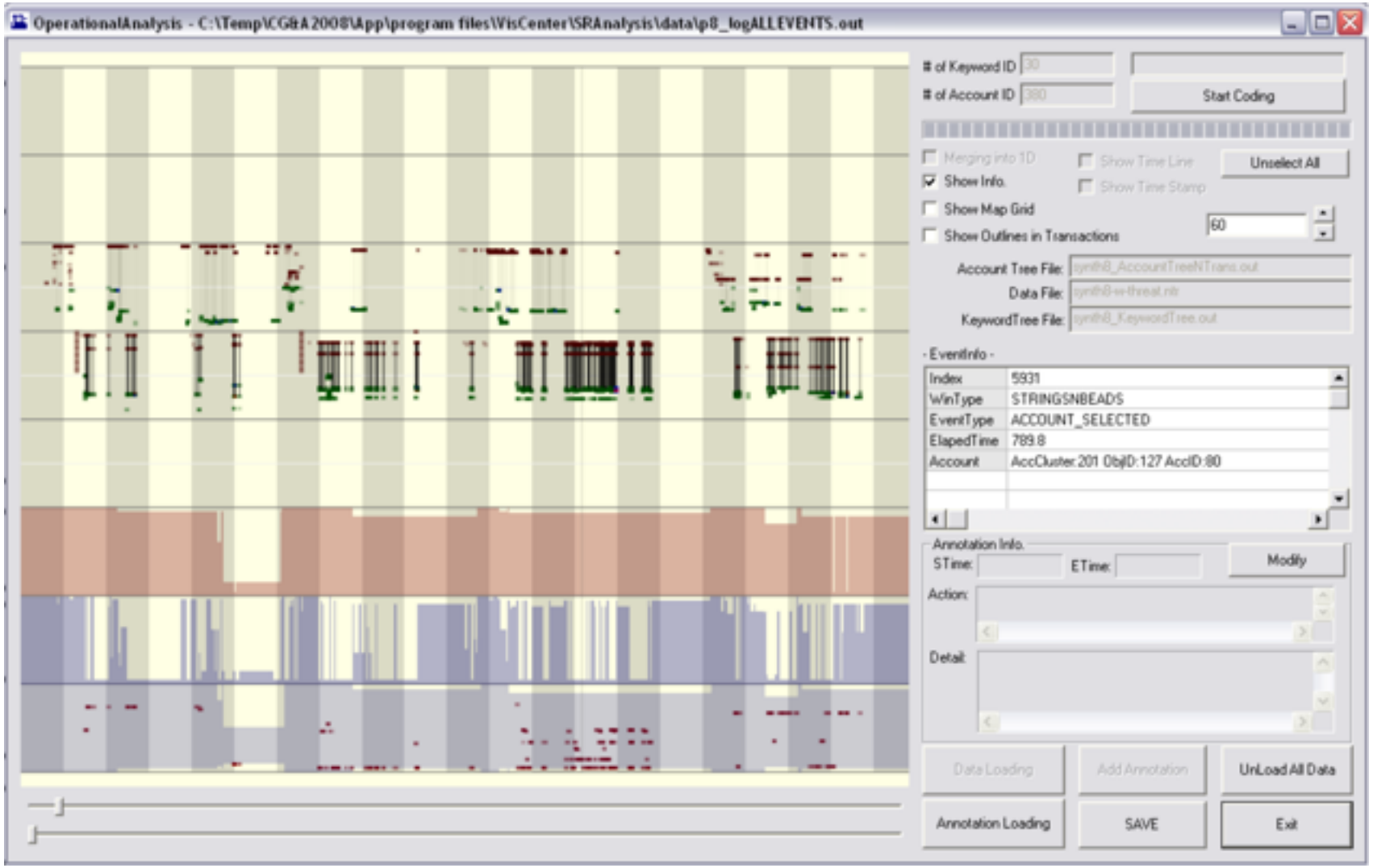
Brown, Eli T., et al. "Finding waldo: Learning about users from their interactions." Visualization and Computer Graphics, IEEE Transactions on 20.12 (2014): 1663-1672.

recover reasoning process: capture



Dou, Wenwen, et al. "Recovering reasoning processes from user interactions." IEEE Computer Graphics and Applications 3 (2009): 52-61.

recover reasoning process: analyze



recover reasoning process: results

- analyzing user interactions after the analysis, researchers can recover:
  - 60% of strategies
  - 60% of methods
  - 79% of findings

- Impressive!
- Can we systematically analyze the user interaction in the system at runtime to steer the models?

another way to look at this

- instead of asking users to directly tune and parameterize models, how can systems learn from their exploratory user interactions?

Mixed-Initiative Systems for VA

Mixed-initiative sys principles

- develop value-added automation
- consider uncertainty about user goals
- consider status of user attention w/r/t timing
- infer ideal action
- employ dialog to resolve uncertainty
- (system) continue learning by observing
- maintain working memory of recent interactions
- …

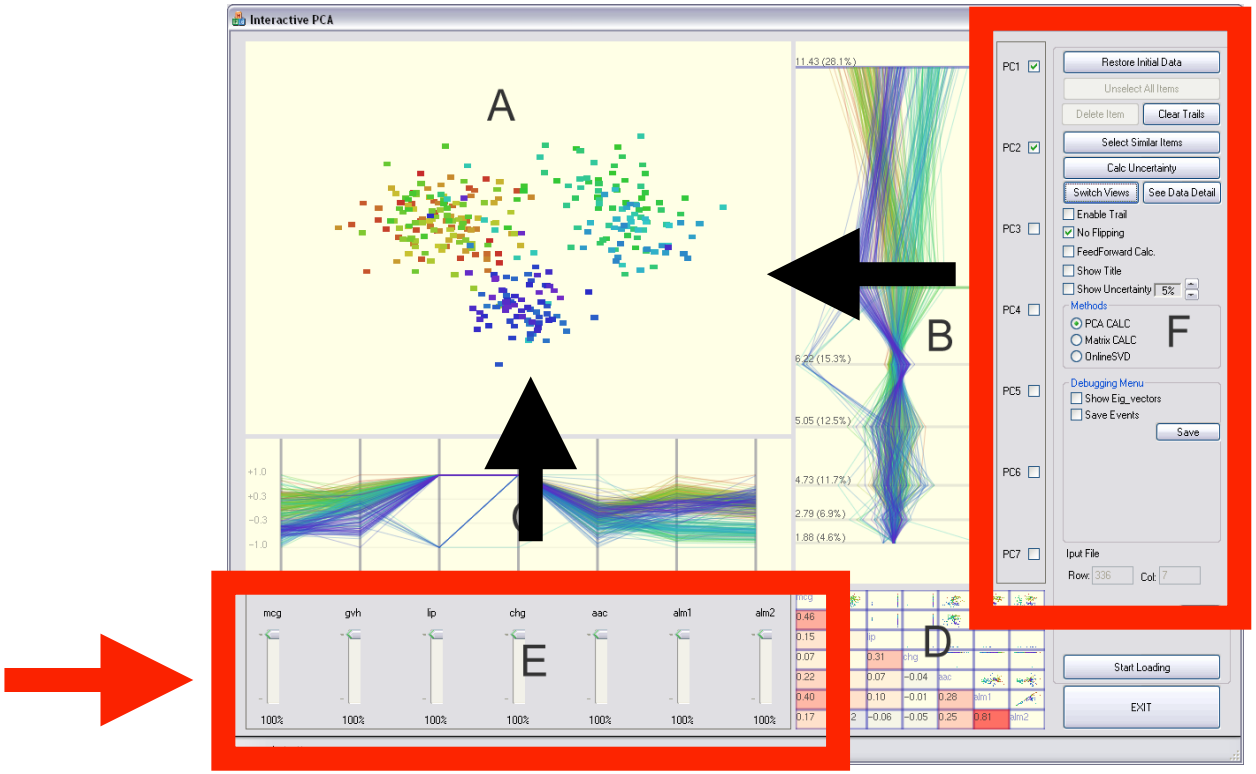
Horvitz, Eric. "Principles of mixed-initiative user interfaces." Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, 1999.

Voyager

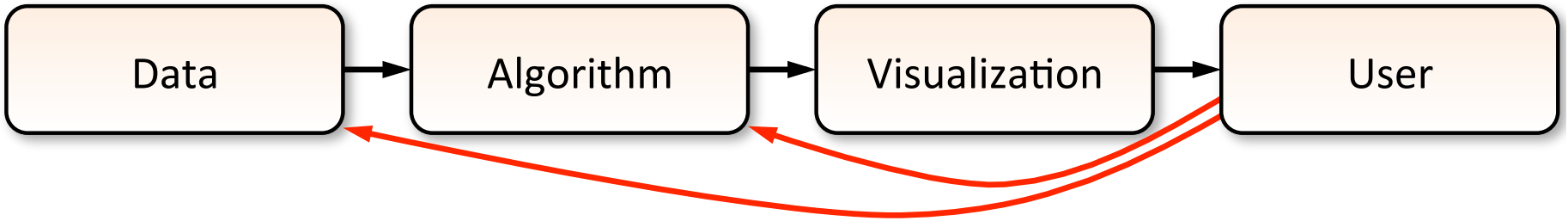
https://idl.cs.washington.edu/papers/voyager/

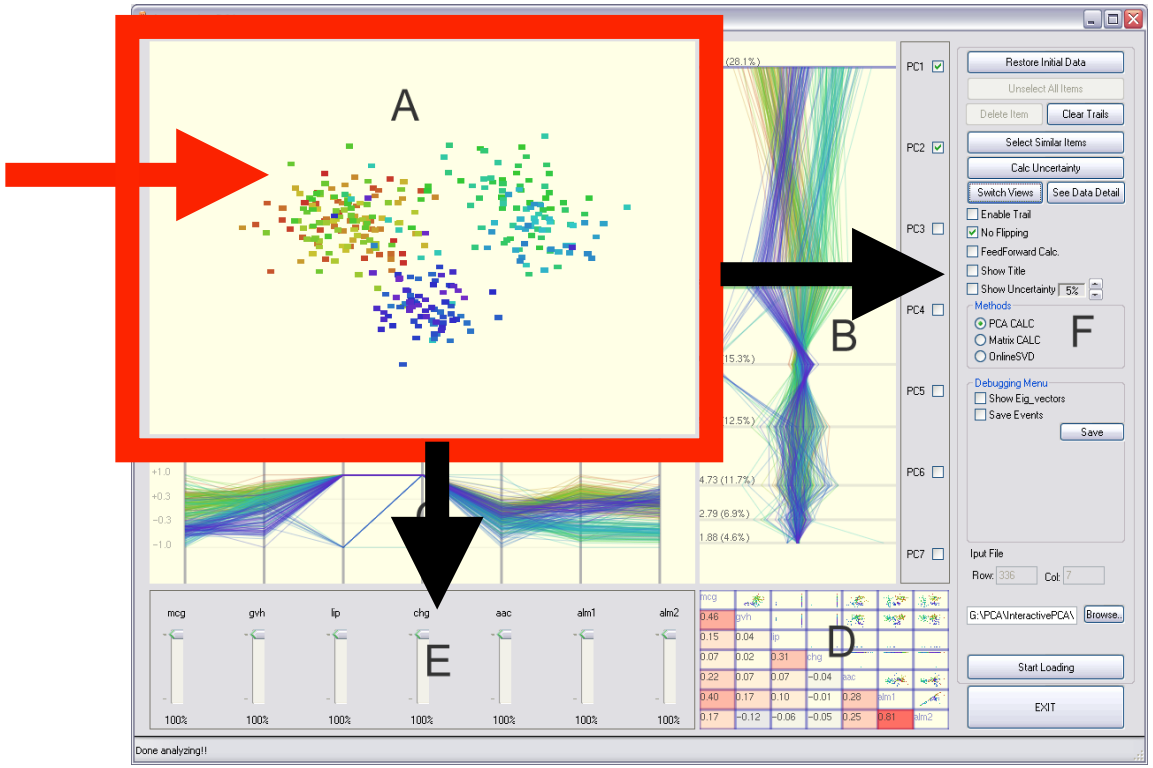


Wongsuphasawat, Kanit, et al. "Voyager: Exploratory analysis via faceted browsing of visualization recommendations." Visualization and Computer Graphics, IEEE Transactions on 22.1 (2016): 649-658.

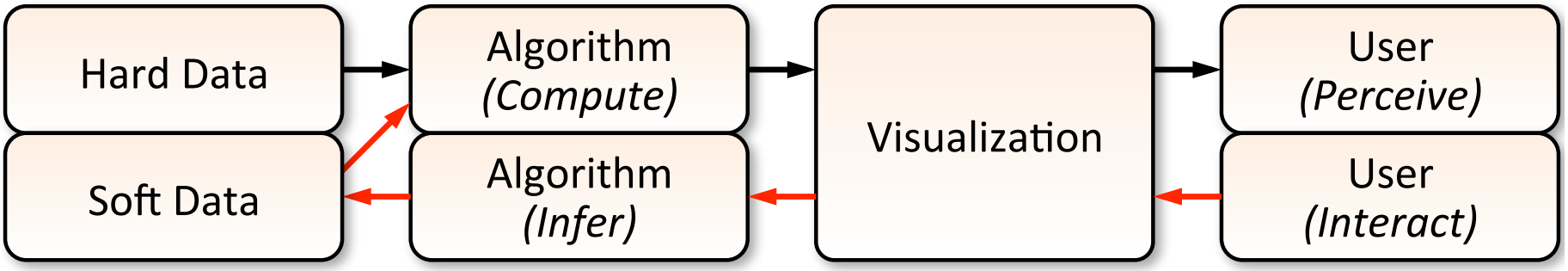


High Formality needed for exploration



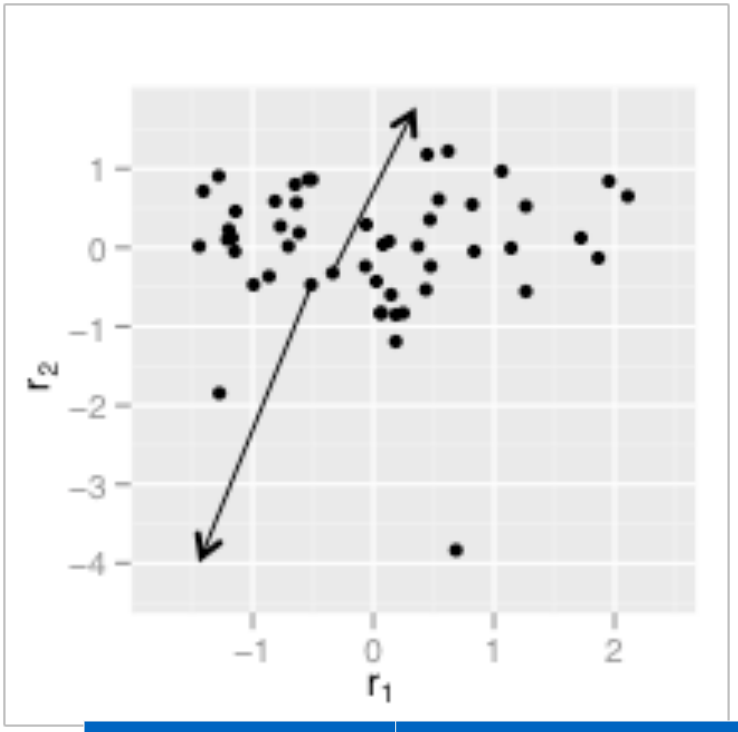


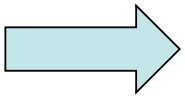
Formality incrementally quantified

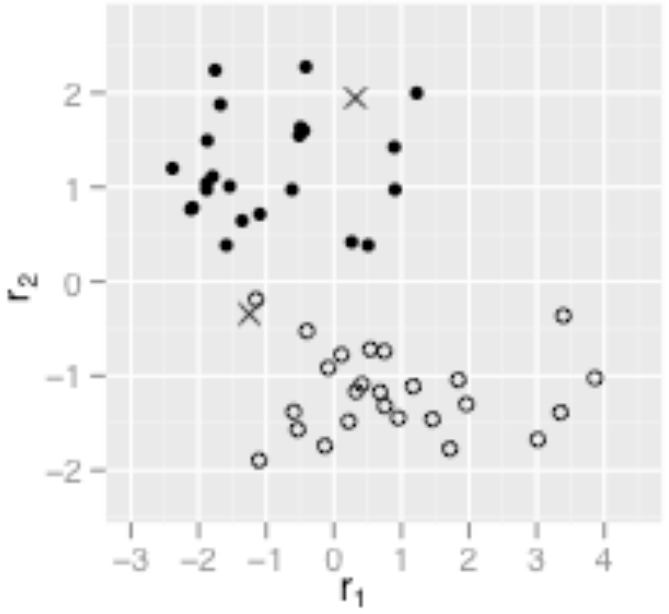


Semantic Interaction Principles

- **Support** analysts' spatial cognition
- **Use** semantic interactions within visual metaphor
- **Interpret** semantic interactions for underlying model
- **Shield** users from complexity of model
- Models **learn** incrementally
- **Provide** visual feedback
- **Reuse** learned model parameters







Convey	Interaction	Solve for
Similarity	Move two data points closer	Parallel residual eigenvector
Dissimilarity	Move two data points apart	Orthogonal residual eigenvector

Endert, A., Han, C., Maiti, D., House, L., Leman, S., & North, C. Observation-level interaction with statistical models for visual analytics. IEEE Visual Analytics Science and Technology (VAST), 2011.

e.g., **Dis-Function**

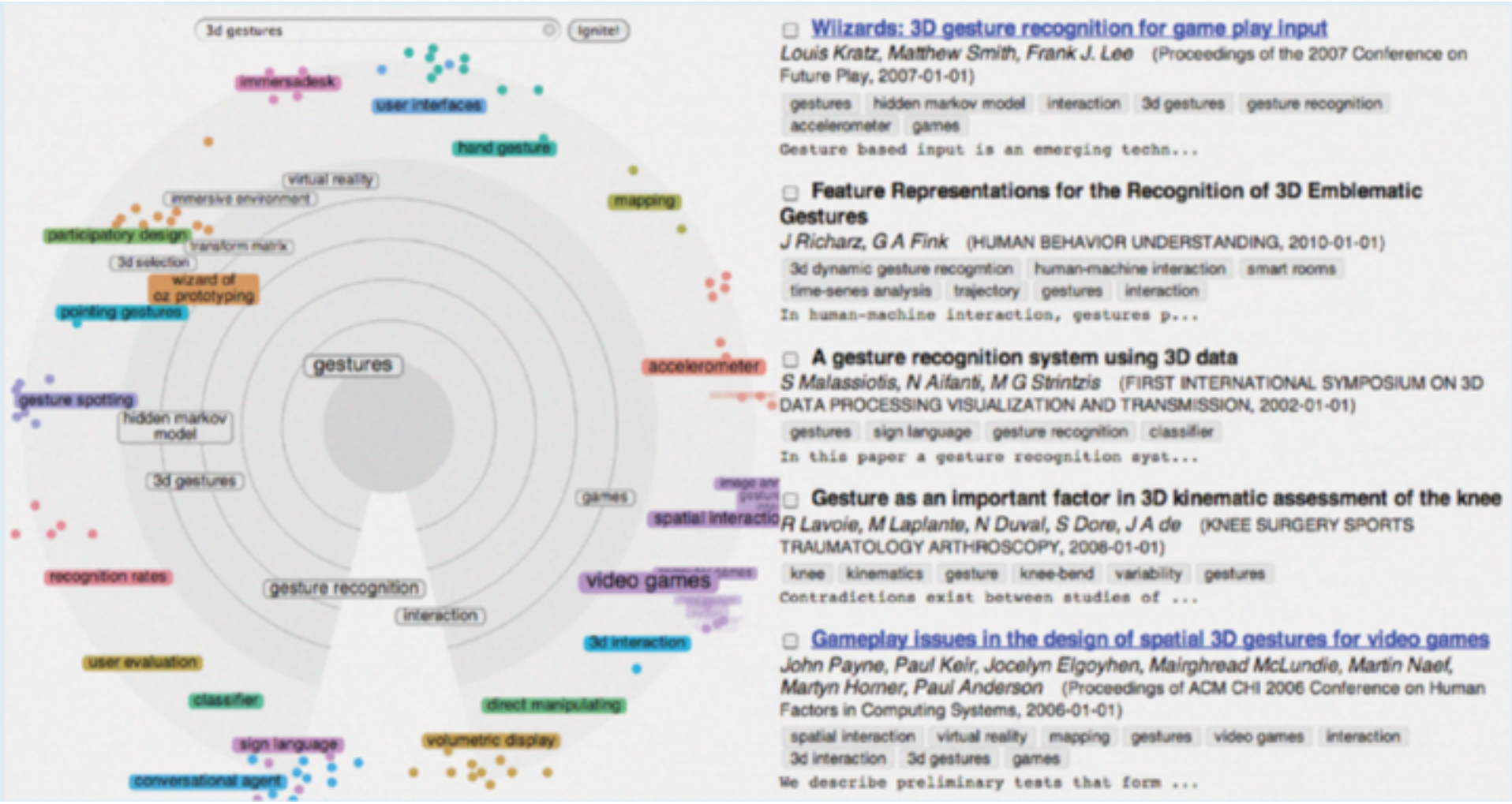


Brown, Eli T., et al. "Dis-function: Learning distance functions interactively." Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on. IEEE, 2012.

can also do this for other models

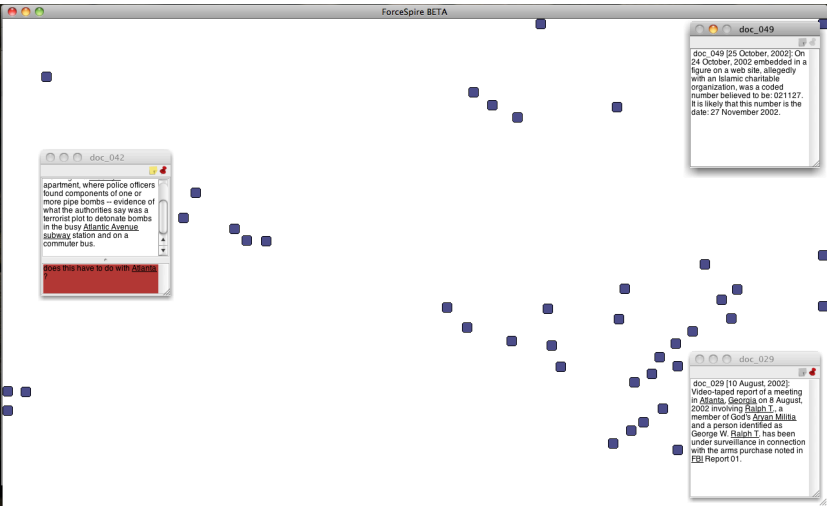
- remember, the general framework for VA encompasses other computational models also
- In addition to DR,
  - entity extraction (for text)
  - information retrieval
- let's take a look at some of those examples

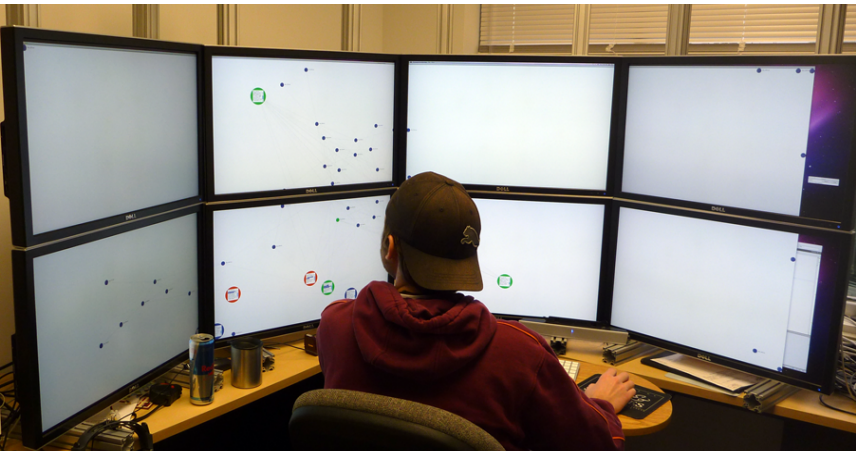
Interactive Intent Modeling

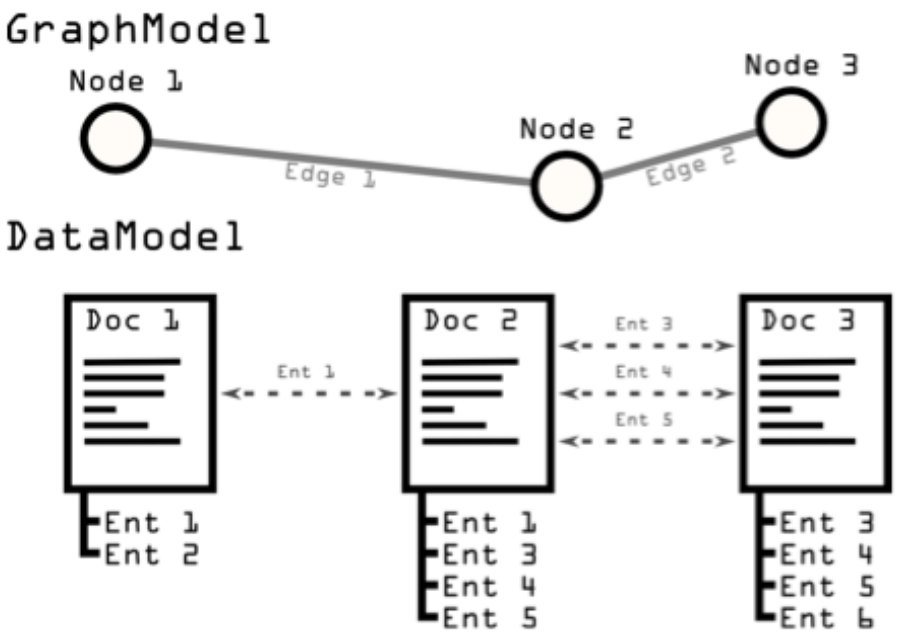


Ruotsalo, Tuukka, et al. "Interactive intent modeling: Information discovery beyond search." Communications of the ACM 58.1 (2015): 86-92.

ForceSPIRE


ForceSpire BETA
doc_049
doc_049 [25 October, 2002]: On 24 October, 2002 embedded in a figure on a web site, allegedly with an Islamic charitable organization, was a coded number believed to be: 021127. It is likely that this number is the date: 27 November 2002.
doc_042
apartment, where police officers found components of one or more pipe bombs -- evidence of what the authorities say was a terrorist plot to detonate bombs in the busy Atlantic Avenue subway station and on a commuter bus.
does this have to do with Atlanta
doc_029
doc_029 [10 August, 2002]: Video-taped report of a meeting in Atlanta, Georgia on 8 August, 2002 involving Ralph T., a member of God's Aryan Militia and a person identified as George W. Ralph T. has been under surveillance in connection with the arms purchase noted in FBI Report 01.




GraphModel
Node 1
Node 2
Node 3
Edge 1
Edge 2
DataModel
Doc 1
Doc 2
Doc 3
Ent 1
Ent 3
Ent 4
Ent 5
Ent 1
Ent 2
Ent 1
Ent 3
Ent 4
Ent 5
Ent 3
Ent 4
Ent 5
Ent 6

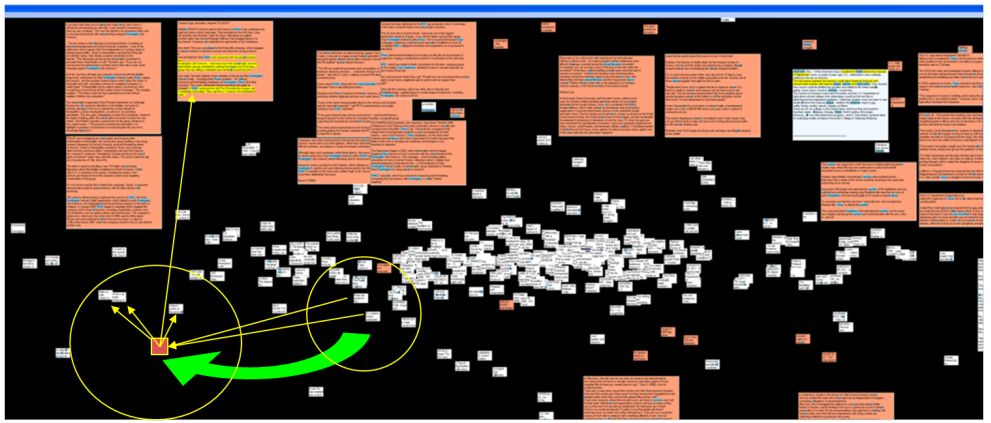
ForceSPIRE

Semantic Interaction	Model Learning
Expressive Movement	Dimension Weighting
Exploratory Movement	Spatial landmarks/constraints
Highlight	Guide entity extraction
Search	Spatial landmarks, dimension weighting
Annotation	Add domain information to the model

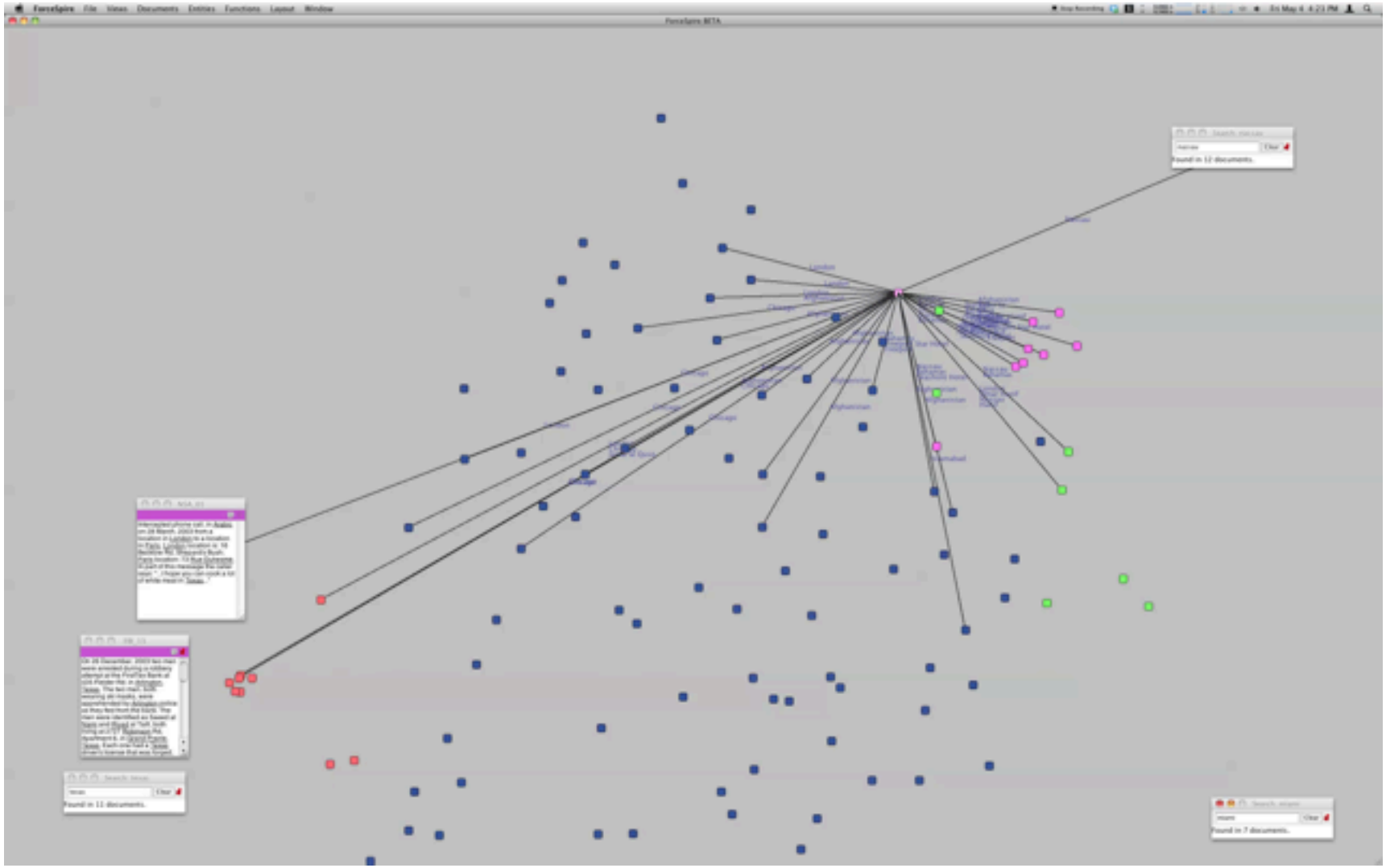
Document Movement

Create Spatial Landmarks
"Exploratory"



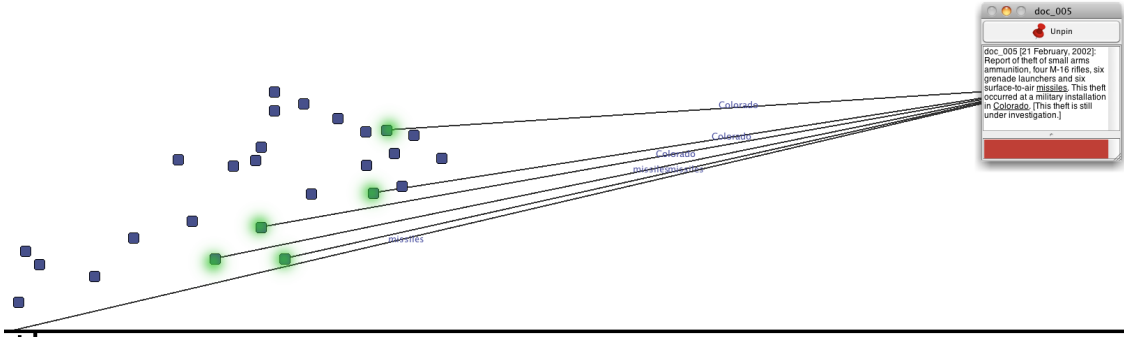


Observation-Level Interaction
MDS, GTM, PPCA

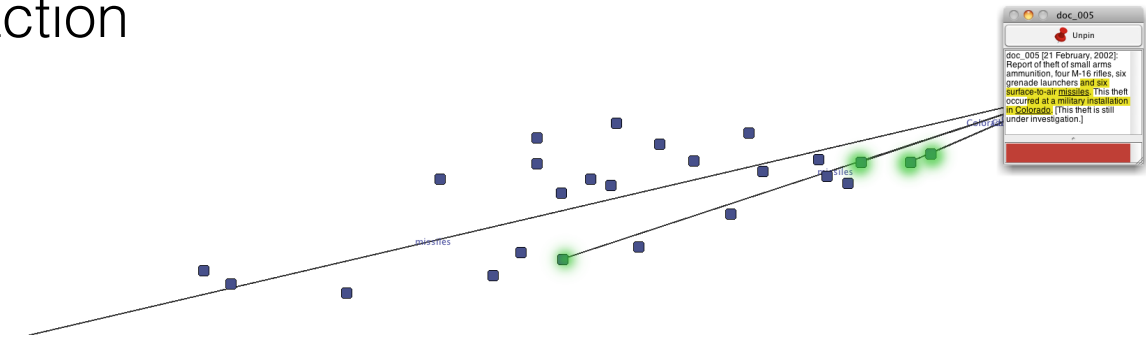


Highlighting

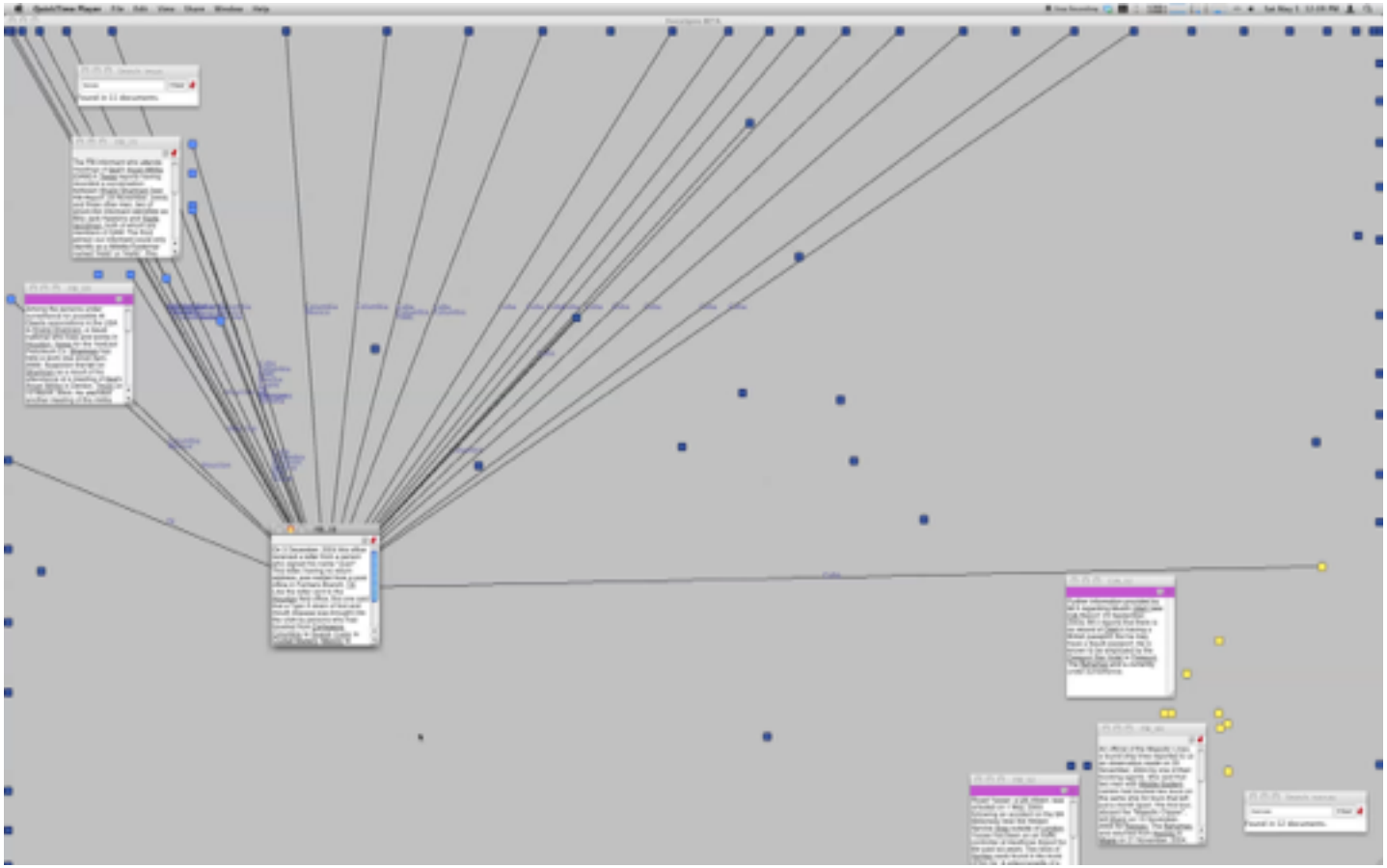
Emphasize phrase
more precise guidance



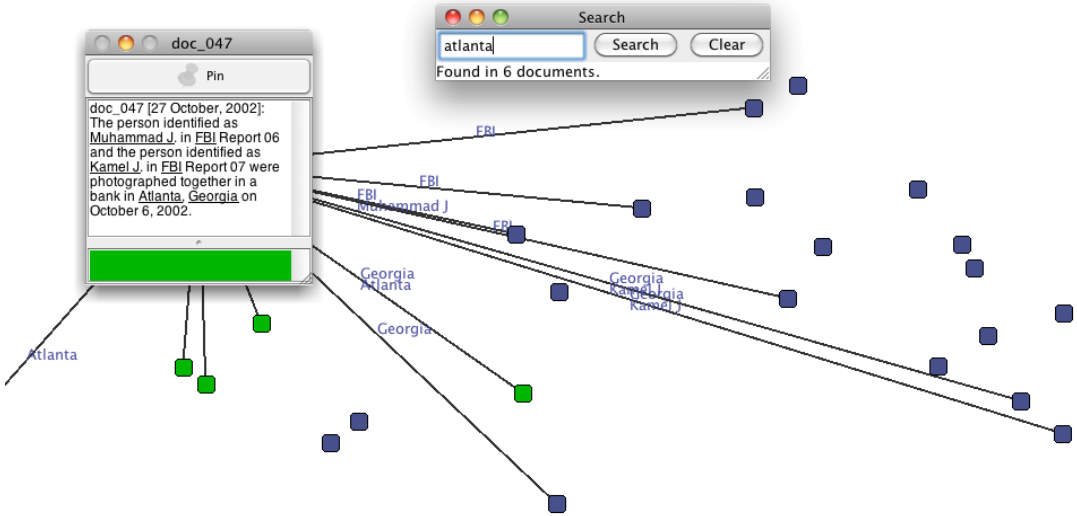
More aggressive entity extraction



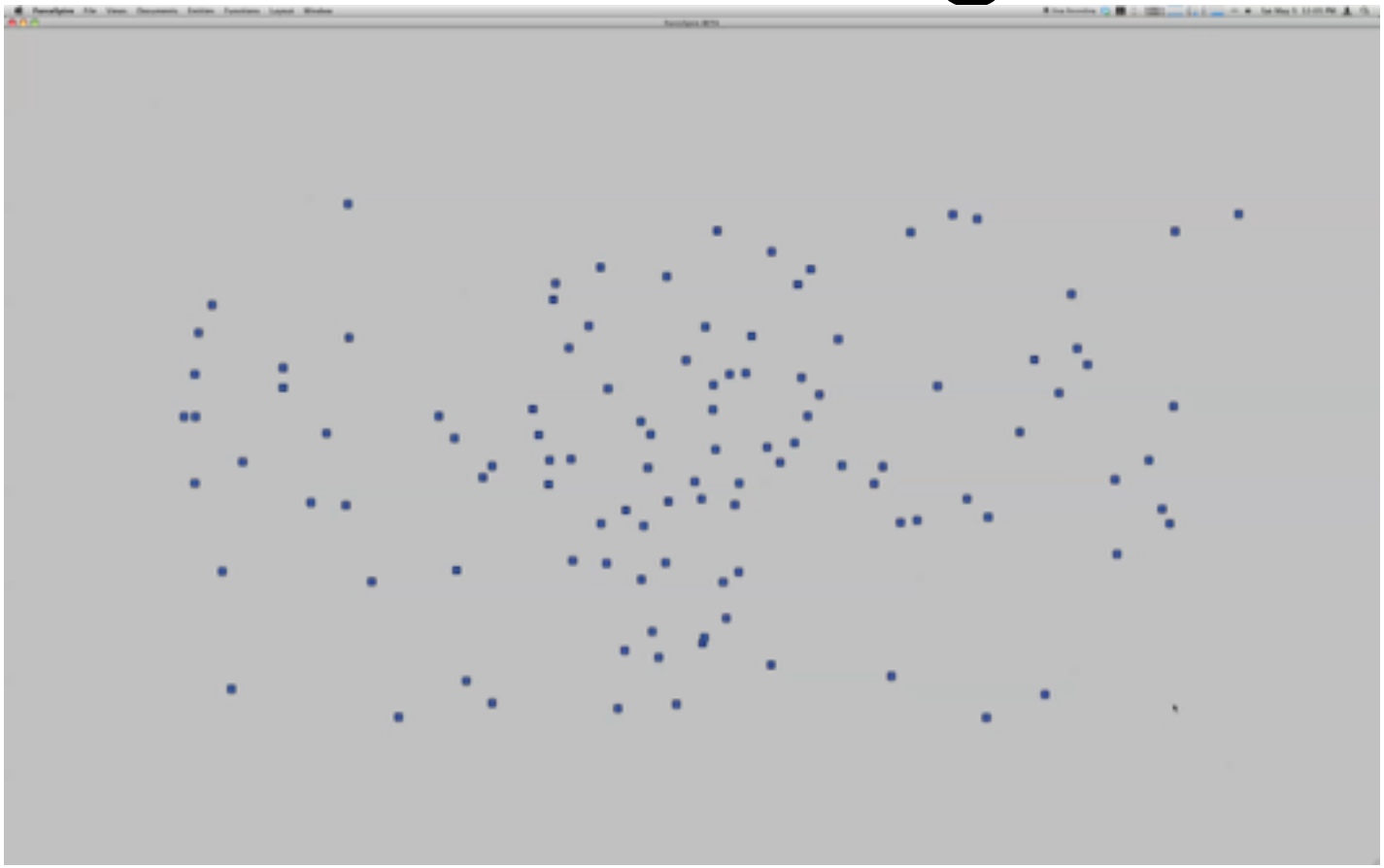
Highlighting



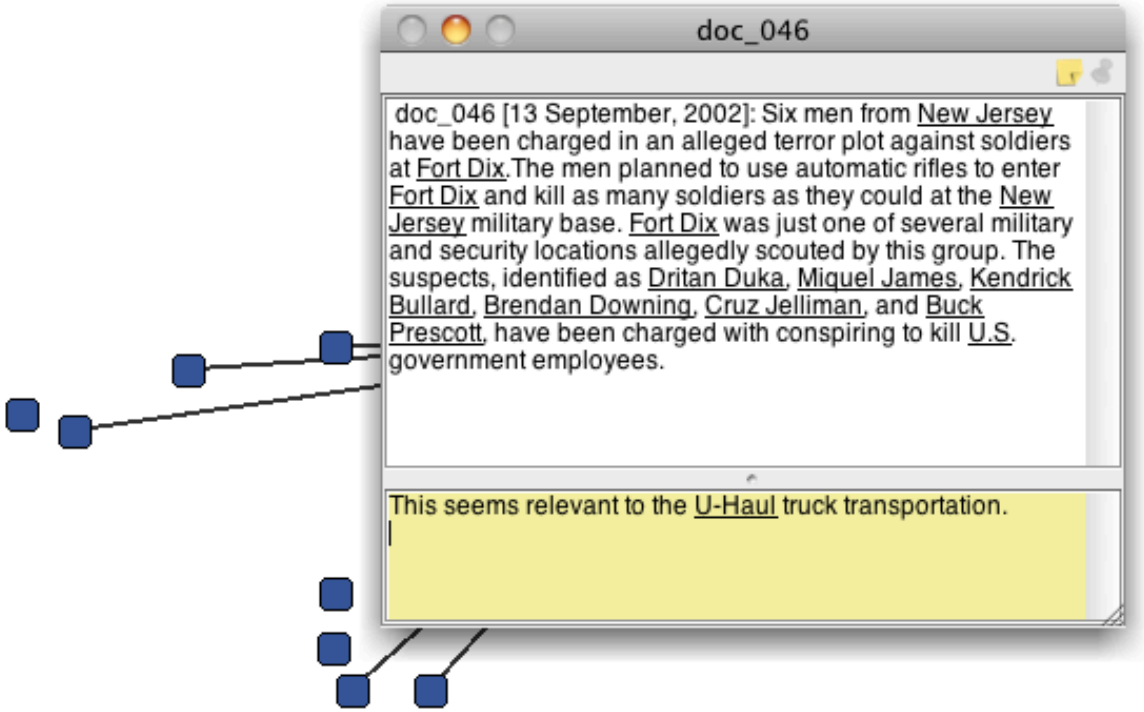
Searching


doc_047
Pin
doc_047 [27 October, 2002]:
The person identified as
Muhammad J. in FBI Report 06
and the person identified as
Kamel J. in FBI Report 07 were
photographed together in a
bank in Atlanta, Georgia on
October 6, 2002.
Search
atlanta
Search
Clear
Found in 6 documents.
FBI
FBI
FBI
Georgia
Atlanta
Georgia
Atlanta

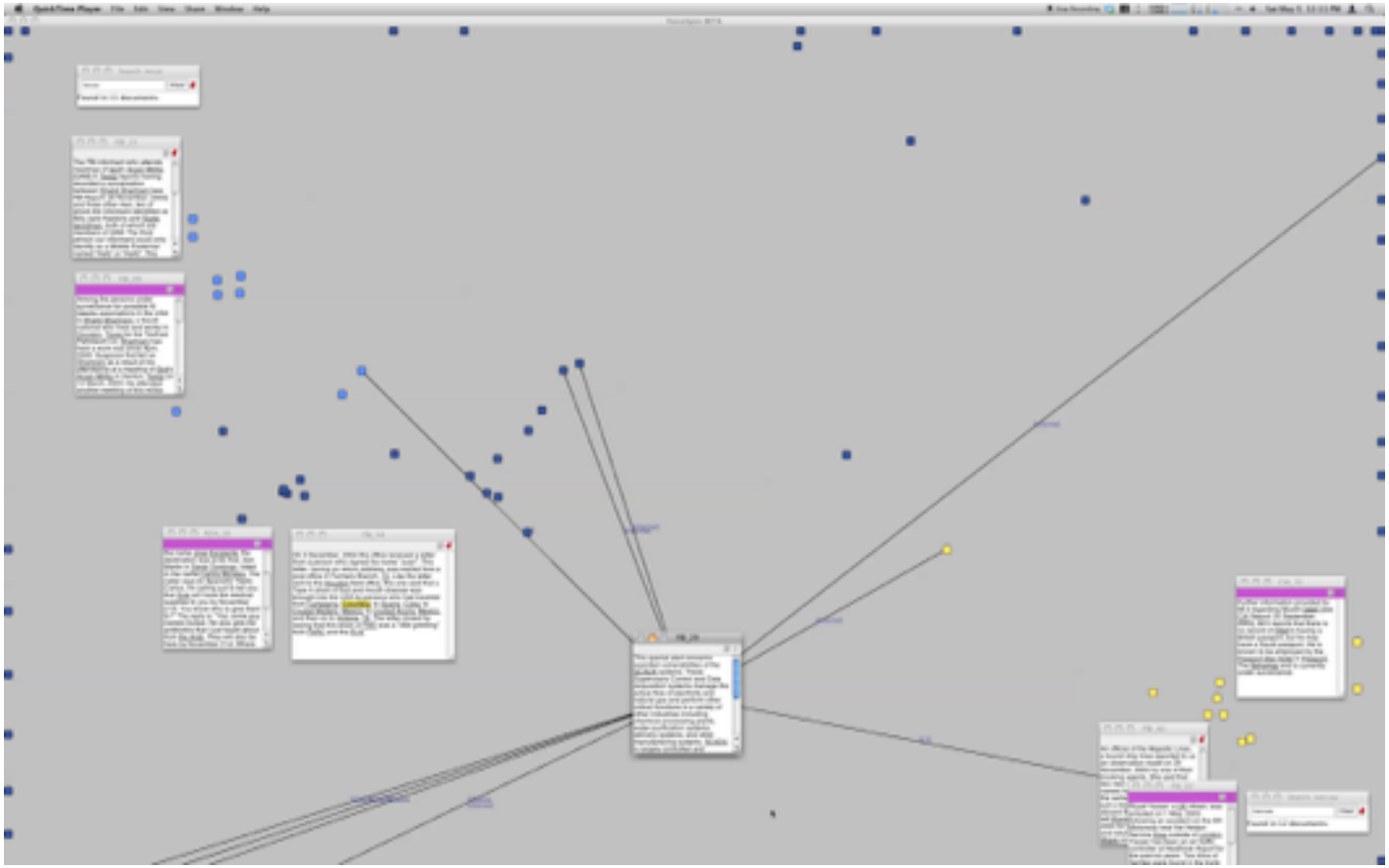
Searching



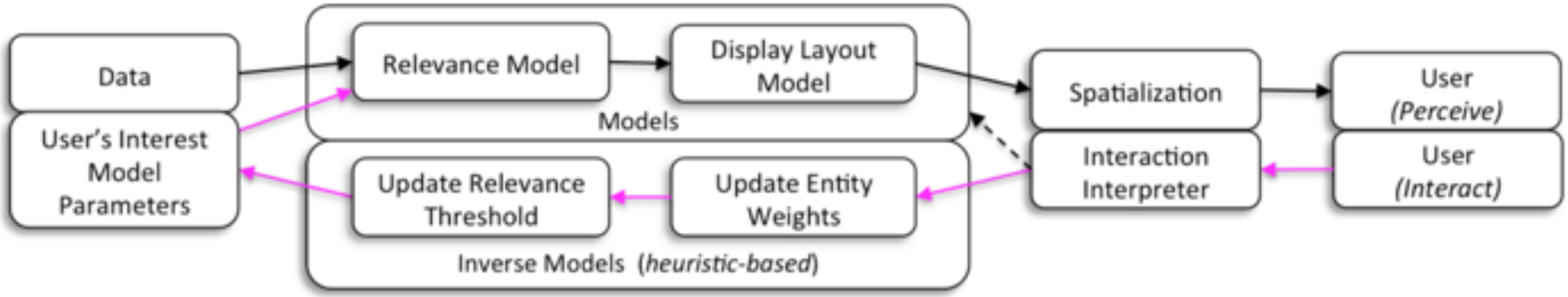
Annotating


doc_046
doc_046 [13 September, 2002]: Six men from New Jersey have been charged in an alleged terror plot against soldiers at Fort Dix.The men planned to use automatic rifles to enter Fort Dix and kill as many soldiers as they could at the New Jersey military base. Fort Dix was just one of several military and security locations allegedly scouted by this group. The suspects, identified as Dritan Duka, Miquel James, Kendrick Bullard, Brendan Downing, Cruz Jelliman, and Buck Prescott, have been charged with conspiring to kill U.S. government employees.
This seems relevant to the U-Haul truck transportation.

Annotating



Multi-Model Semantic Interaction



Information Retrieval

Bradel, Lauren, Chris North, and Leanna House. "Multi-model semantic interaction for text analytics." Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on. IEEE, 2014.

Model Steering for non-text

- InterAxis [demo, if internet]
  - http://va.gatech.edu/projects/interaxis/

Active Data Environment



Mixed-Initiative Visual Analytics Using Task-Driven Recommendations

Kristin Cook (PNNL), Nick Cramer (PNNL), David Israel (SRI), Michael Wolverton (SRI), Joe Bruce (PNNL), Russ Burtner (PNNL), and Alex Endert (GT)

[Contact: kris.cook@pnnl.gov]

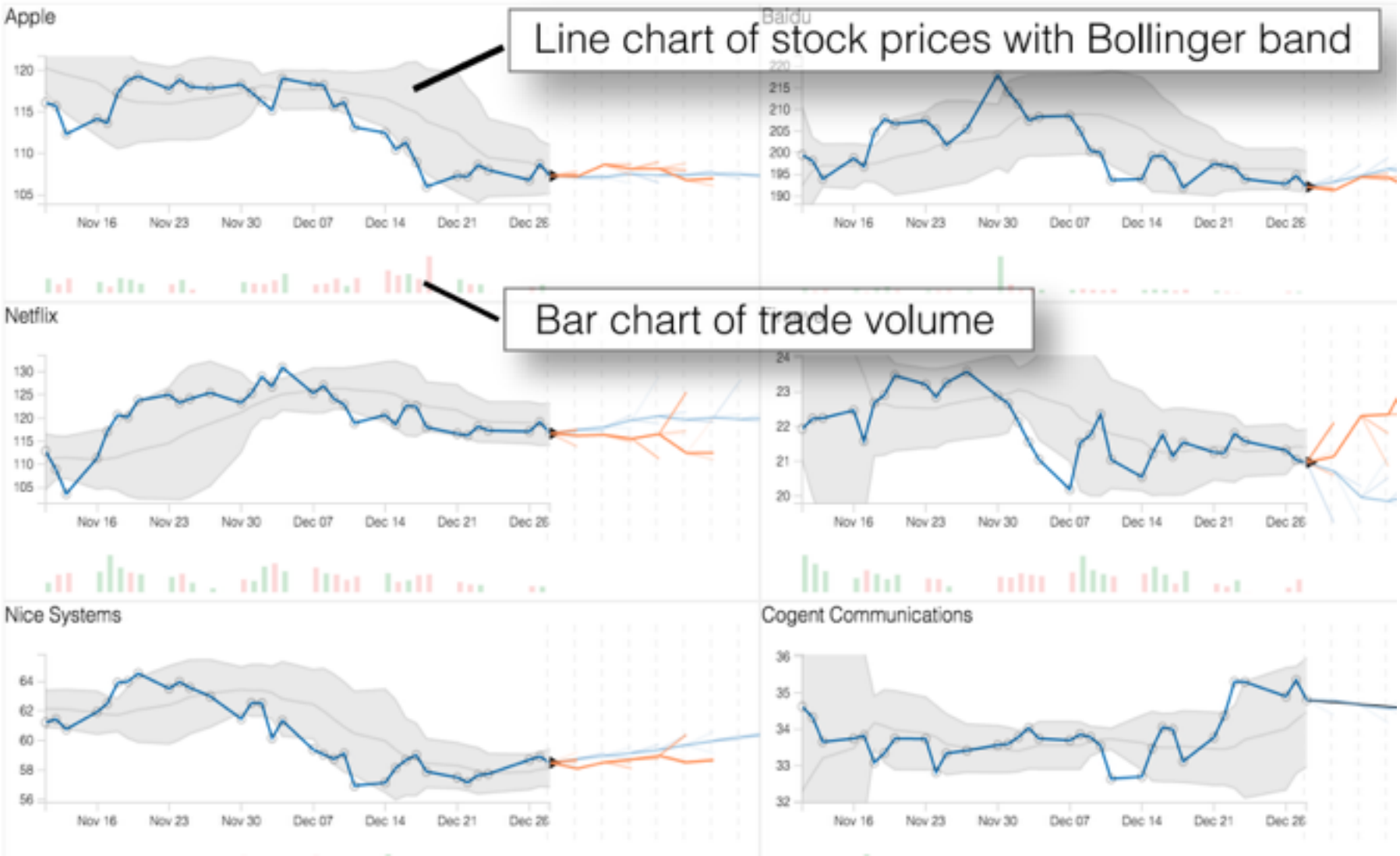
Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

SRI International

Cook, Kristin, Nick Cramer, David Israel, Michael Wolverton, Joe Bruce, Russ Burtner, and Alex Endert. "Mixed-initiative visual analytics using task-driven recommendations." In Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on, pp. 9-16. IEEE, 2015.

TimeFork: Prediction Models



Badam, Sriram Karthik, et al. "TimeFork: Mixed-initiative time-series prediction." Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on. IEEE, 2014.

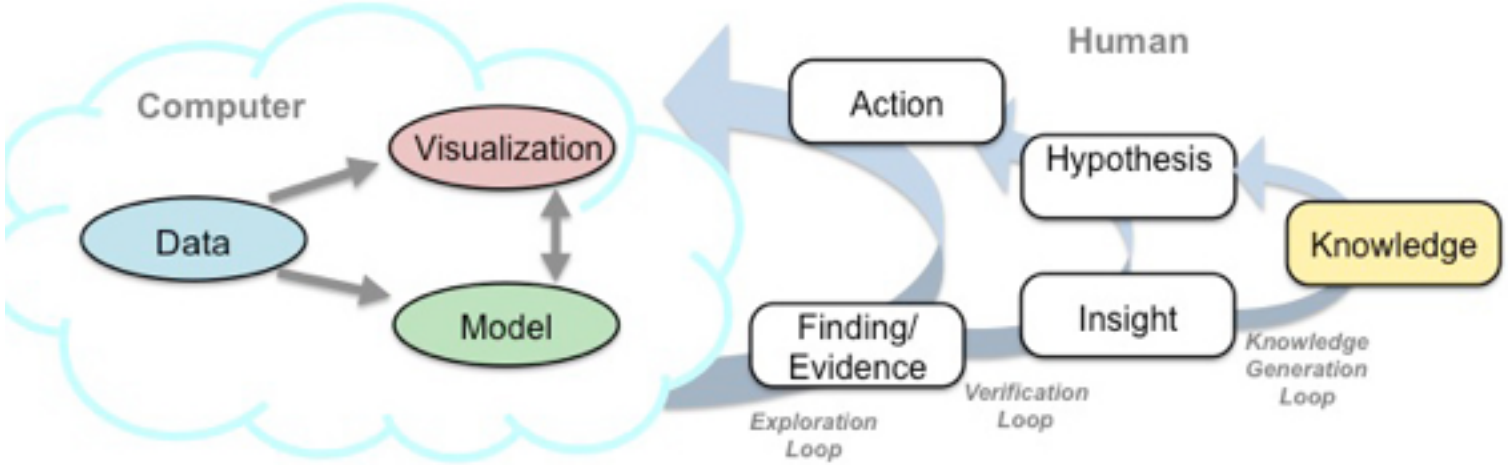
Wrangler

http://vis.stanford.edu/wrangler/app/

- a lot of techniques I've discussed assume that data is clean, organized, etc.
- wrangler helps getting your data "ready to vis"
- enables data transformation expressions via user demonstration in data table

Recap

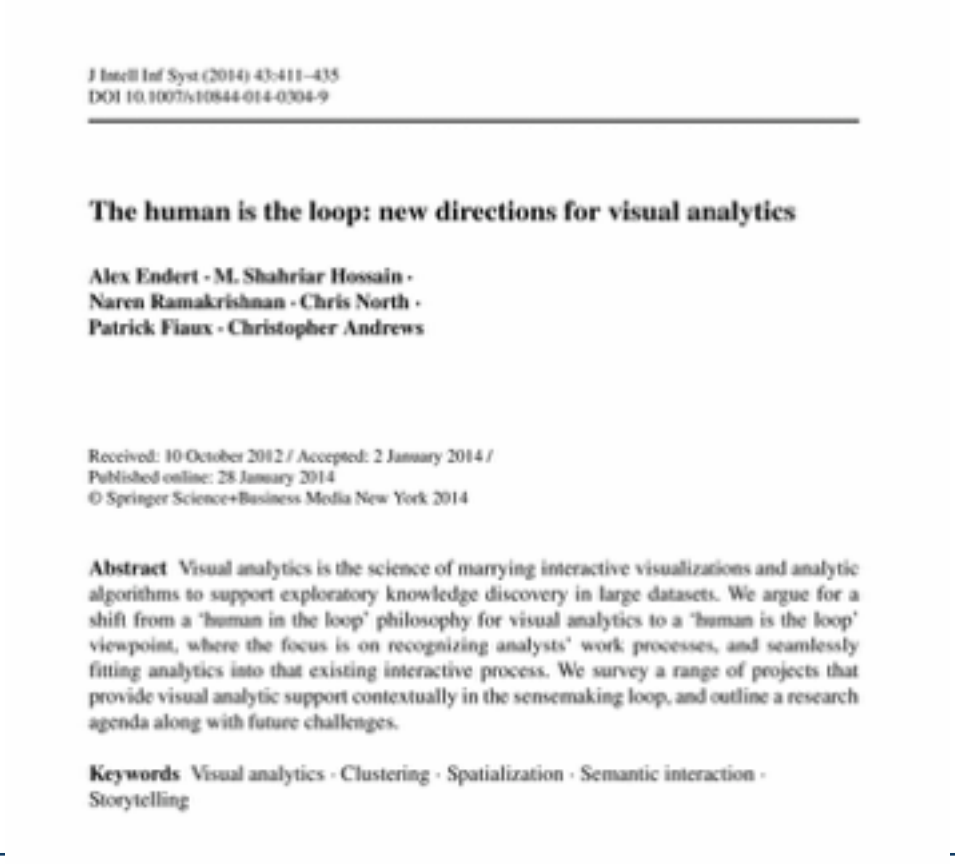
- lots about human-in-the-loop



- there are other views on this

human-*is*-the-loop?

- how do we tie computation to the various tasks that users perform during sensemaking?



J Intell Inf Syst (2014) 43:411–435
DOI 10.1007/s10844-014-0304-9

The human is the loop: new directions for visual analytics

Alex Endert · M. Shahriar Hossain · Naren Ramakrishnan · Chris North · Patrick Fiaux · Christopher Andrews

Received: 10 October 2012 / Accepted: 2 January 2014 /
Published online: 28 January 2014
© Springer Science+Business Media New York 2014

Abstract Visual analytics is the science of marrying interactive visualizations and analytic algorithms to support exploratory knowledge discovery in large datasets. We argue for a shift from a 'human in the loop' philosophy for visual analytics to a 'human is the loop' viewpoint, where the focus is on recognizing analysts' work processes, and seamlessly fitting analytics into that existing interactive process. We survey a range of projects that provide visual analytic support contextually in the sensemaking loop, and outline a research agenda along with future challenges.

Keywords Visual analytics · Clustering · Spatialization · Semantic interaction · Storytelling

human-*on*-the-loop?



humans should be involved

- regardless of where in the loop, many decisions today require users to be included in the computation and holistic decision-making process.

there's much, much more

- What I did not cover:
  - cyber security
  - geospatial
  - storytelling
  - streaming analytics
  - image and video analysis
  - analytic provenance
  - VA on other interaction and display modalities
  - many more

let’s recap some key points

visualization helps us think

- not only helps us think about data, but also helps us use our visual system to help us reason in general
- visualization is high-bandwidth, good for data transfer to humans
- lots of perceptual and cognition literature tells us many guidelines for visualization

Visual Analytics

- VA leverages many InfoVis techniques and makes them more powerful with computation
- need to think about user task, domain expertise, data characteristics when choosing which analytic model to use
- data models, analytic models, and mental models are different.
  - leverage them all, but respect their differences

acknowledgements

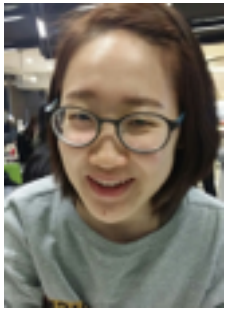
- many graduate students made the work I’ve showed possible

- many slides were developed by John Stasko (Georgia Tech) and Chris North (Virginia Tech)

- your time



Bahador Saket
PhD, CS



Hannah Kim
Phd, CSE



Emily Wall
Phd, CS



Fred Hohman
Phd, CSE



Subhajit Das
Phd, CS



Arjun Srinivasan
Phd, CS

Georgia Tech School of Interactive Computing

GVU Georgia Tech

VACCINE

VA.lab
va.gatech.edu

KEYSIGHT TECHNOLOGIES

Microsoft

Pacific Northwest NATIONAL LABORATORY



Sandia National Lab
May 26, 2016

New Approaches for Information Visualization: Rethinking Existing Notions

John Stasko

School of Interactive Computing
Georgia Institute of Technology

stasko@cc.gatech.edu



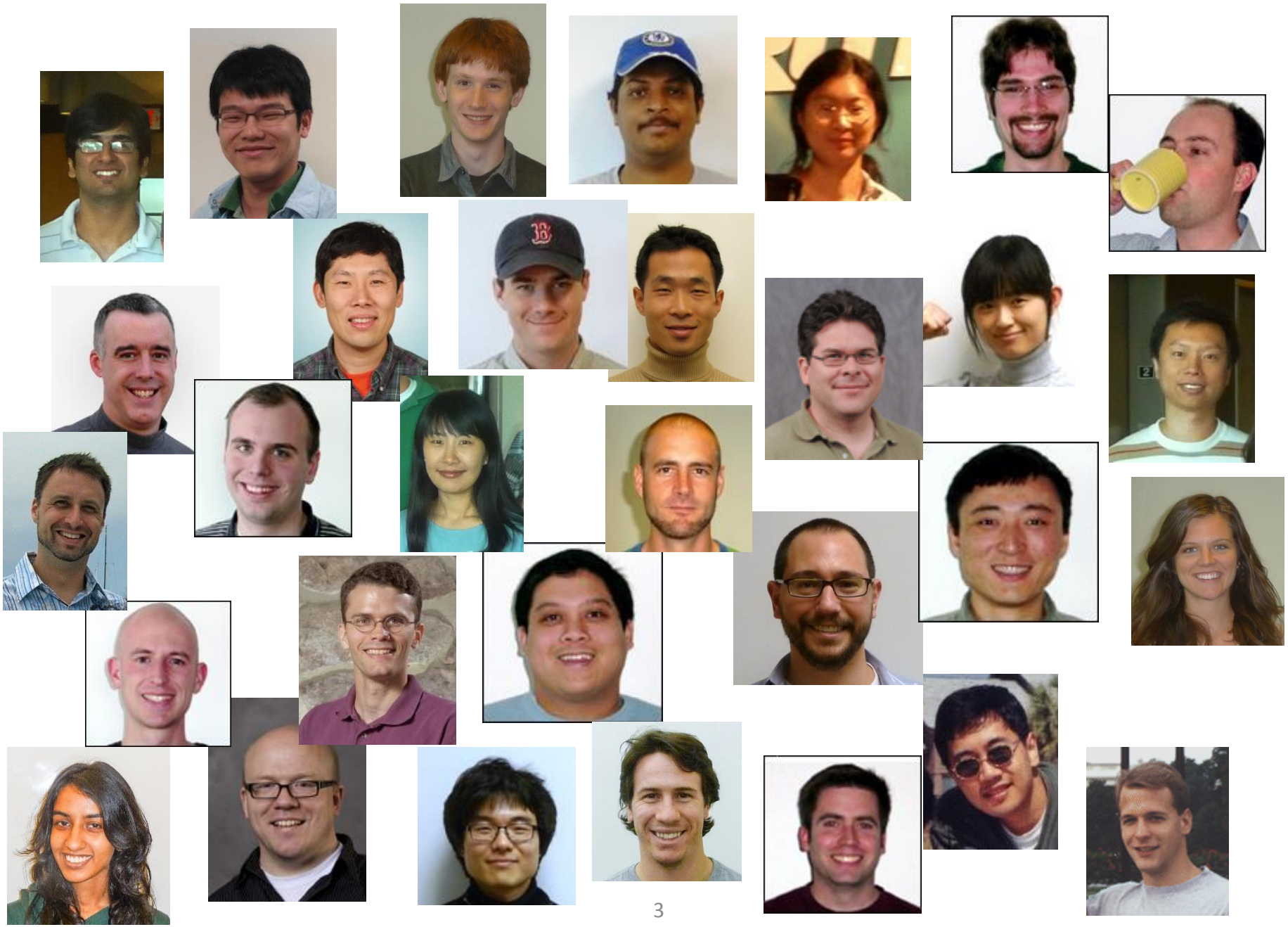


Georgia Tech College of Computing
School of Interactive Computing

Three existing notions

1. Networks as force-directed node-link diagrams
2. Representation is first and foremost
3. Evaluation through user studies and experiments

Talented students

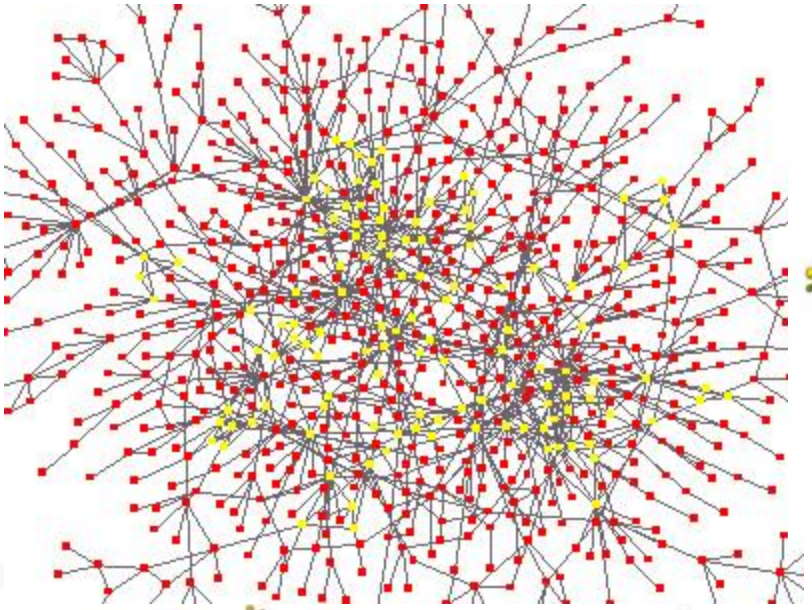


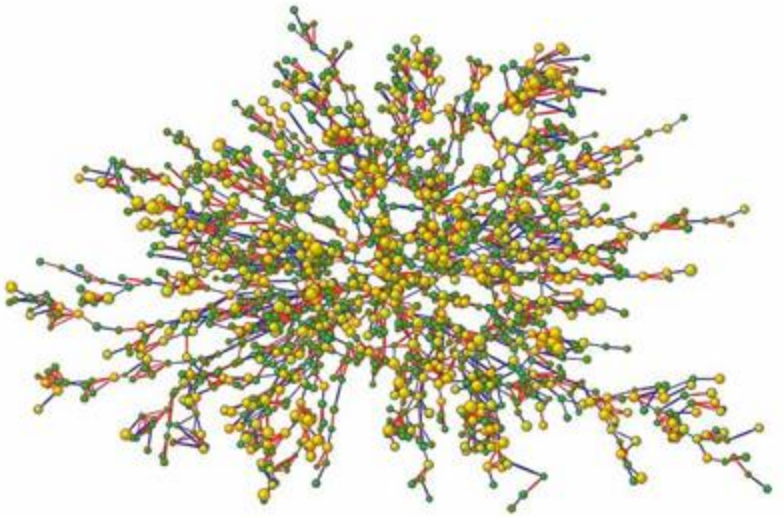
1. Network Vis as force-directed node link

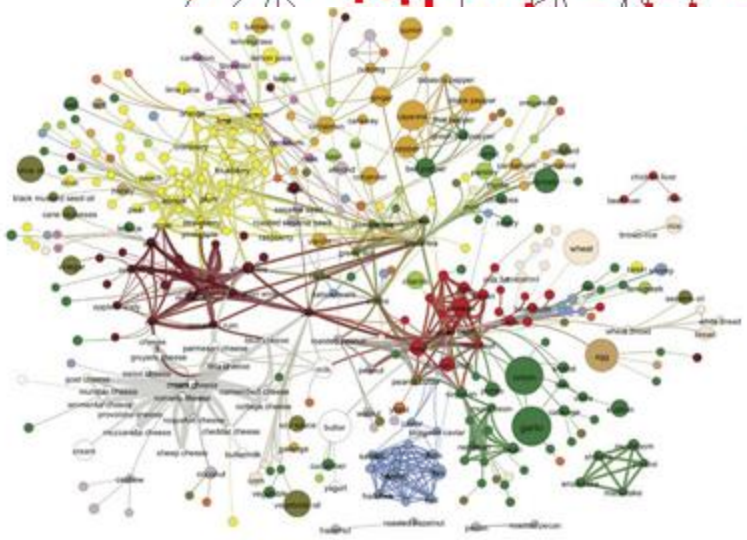
Look impressive, but interpretable?

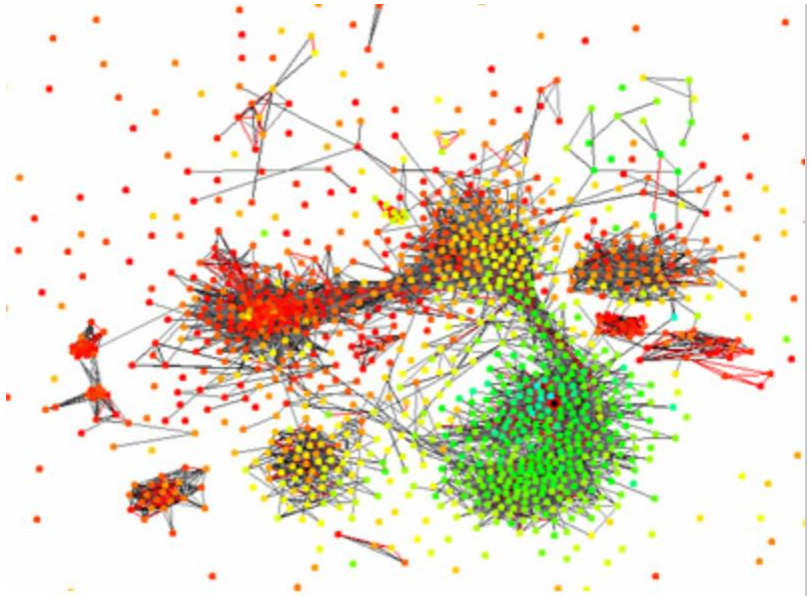
Node-link graphs

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





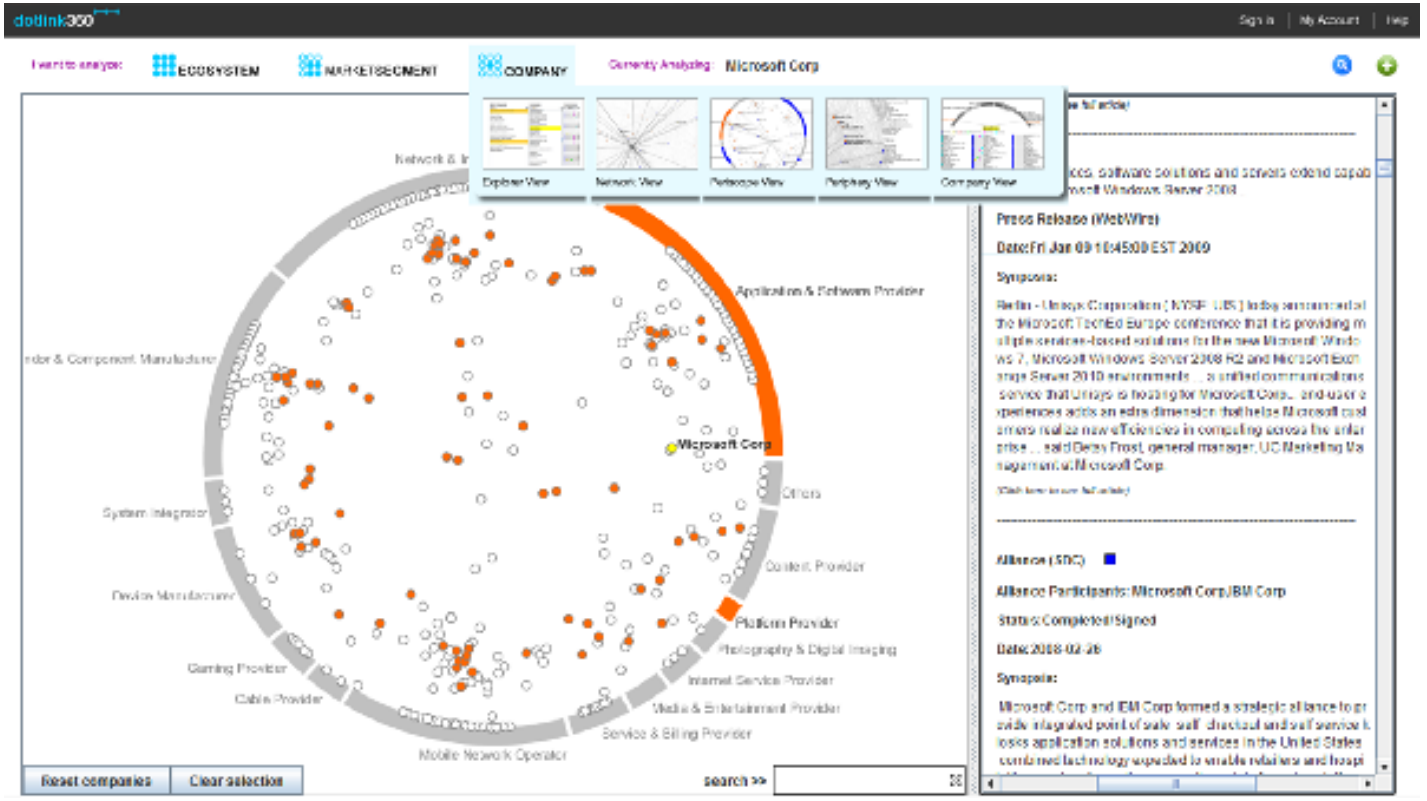




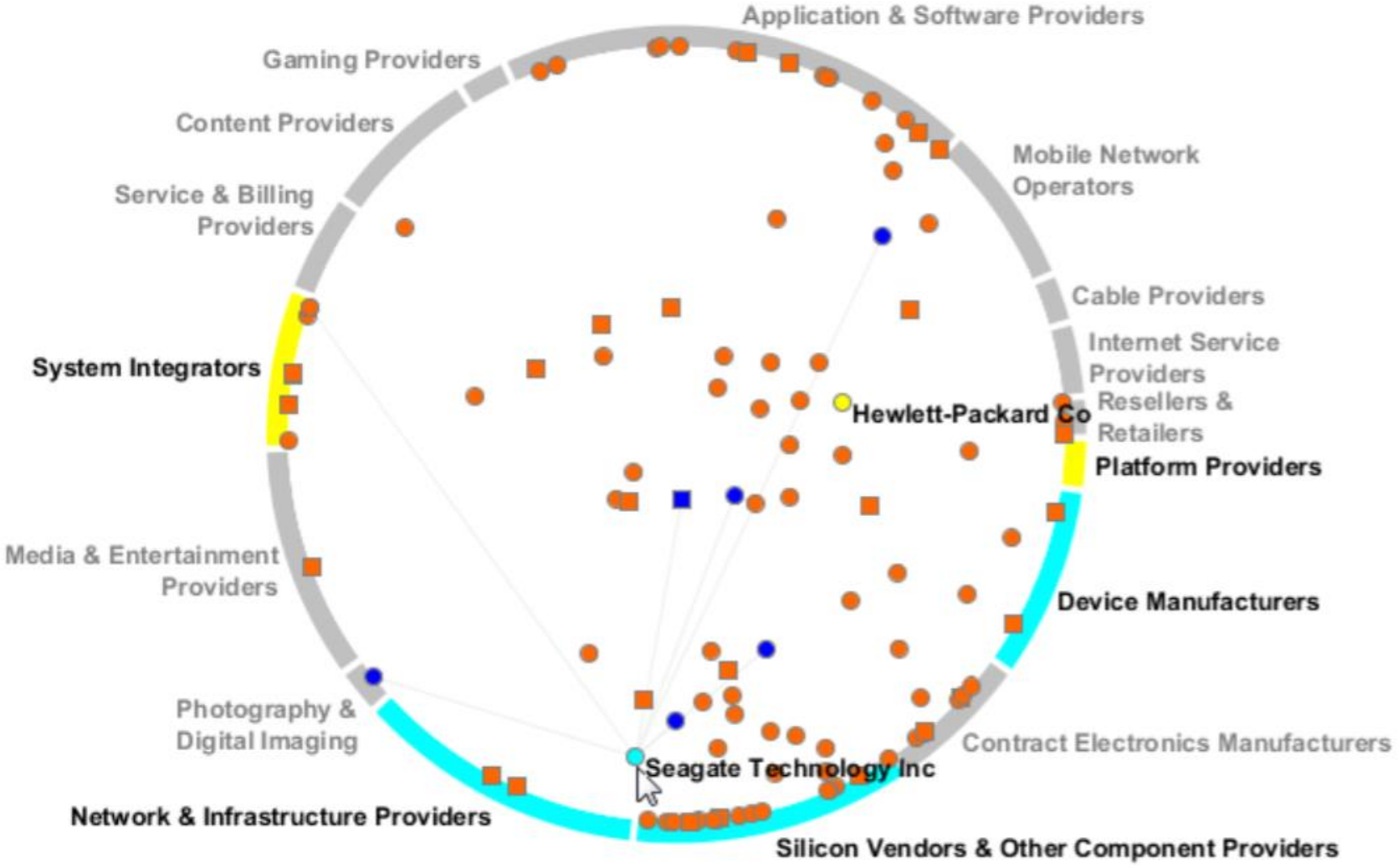
Attribute-based Network Layout

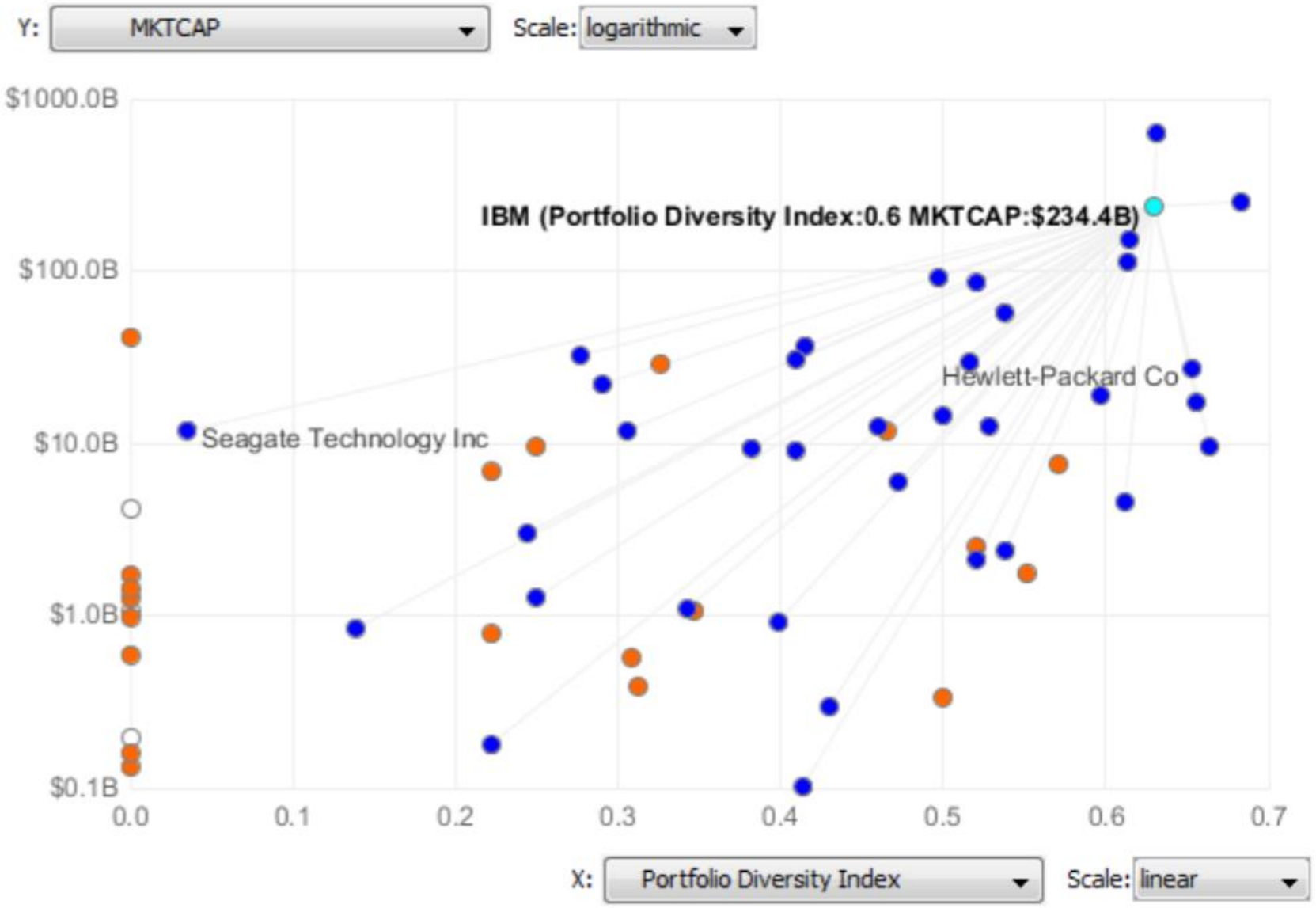
dotlink360

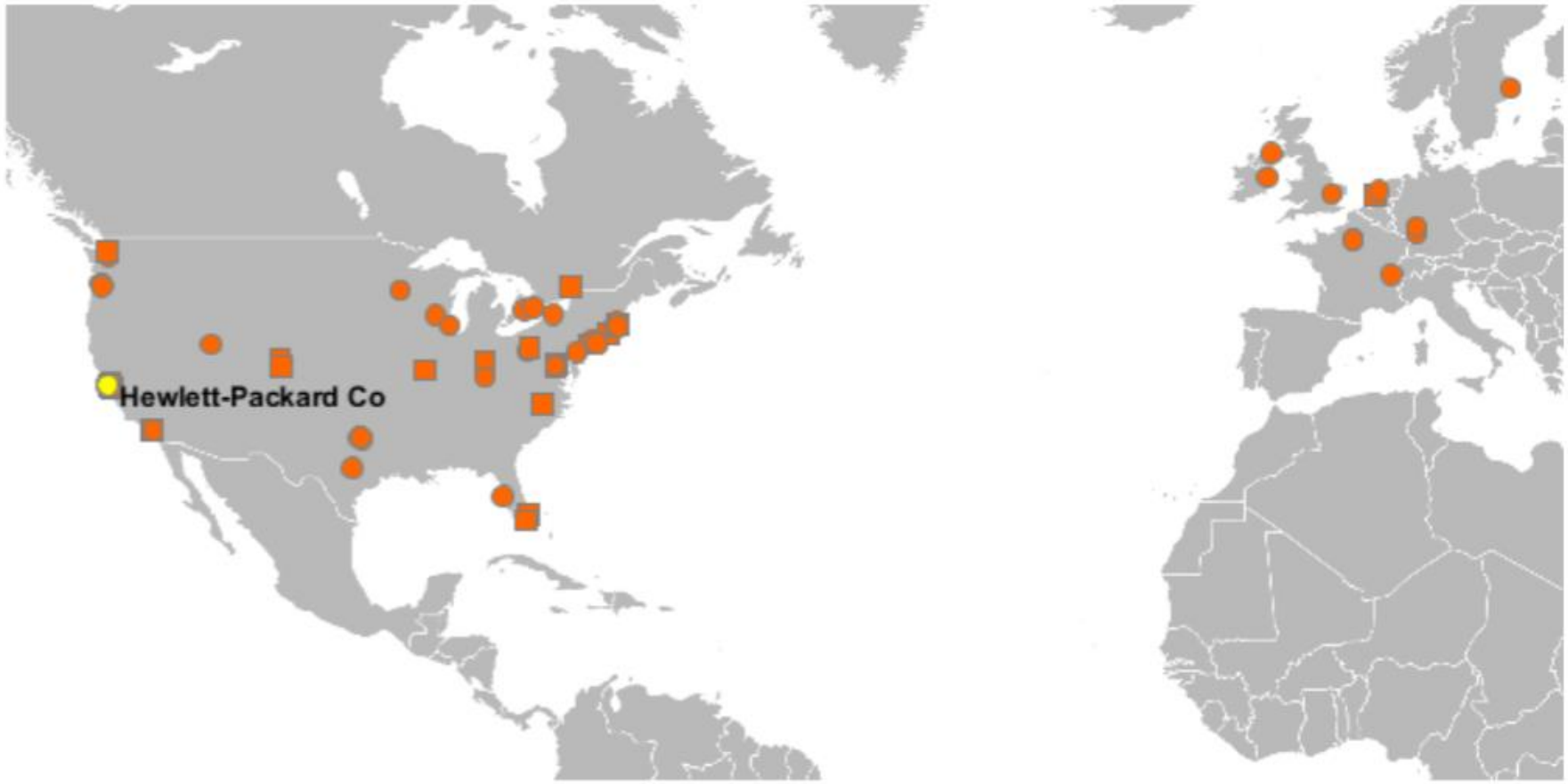
Design study:
Corporate alliances



Basole, et al
TVCG (InfoVis) ‘13


Application & Software Providers
Gaming Providers
Content Providers
Service & Billing Providers
Mobile Network Operators
Cable Providers
Internet Service Providers
Resellers & Retailers
Platform Providers
Device Manufacturers
Contract Electronics Manufacturers
Silicon Vendors & Other Component Providers
Network & Infrastructure Providers
Photography & Digital Imaging
Media & Entertainment Providers
System Integrators
Hewlett-Packard Co
Seagate Technology Inc

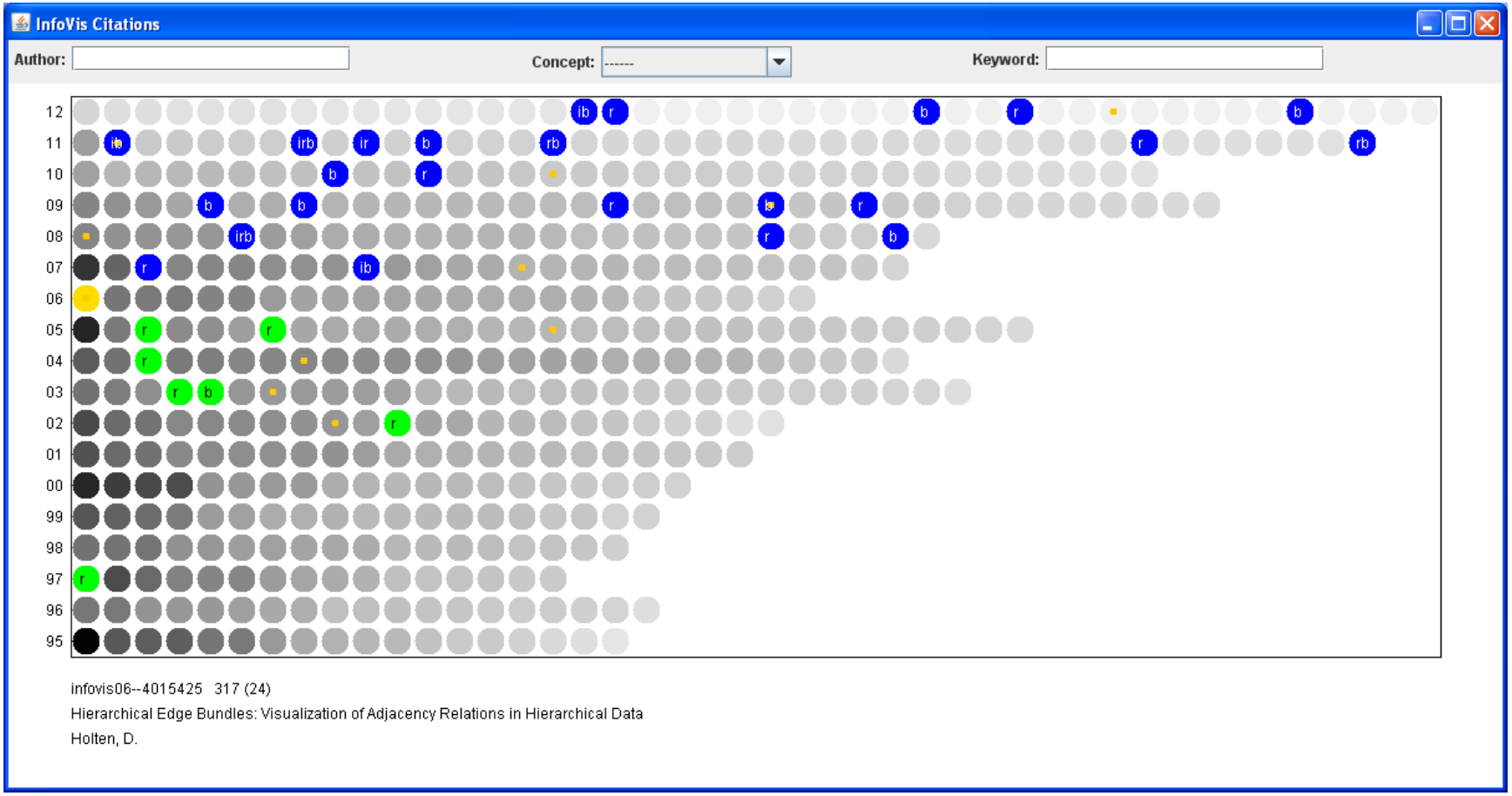

Y:
MKTCAP
Scale:
logarithmic
$1000.0B
$100.0B
$10.0B
$1.0B
$0.1B
IBM (Portfolio Diversity Index:0.6 MKTCAP:$234.4B)
Hewlett-Packard Co
Seagate Technology Inc
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
X:
Portfolio Diversity Index
Scale:
linear


Hewlett-Packard Co

Paper citations

CiteVis

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

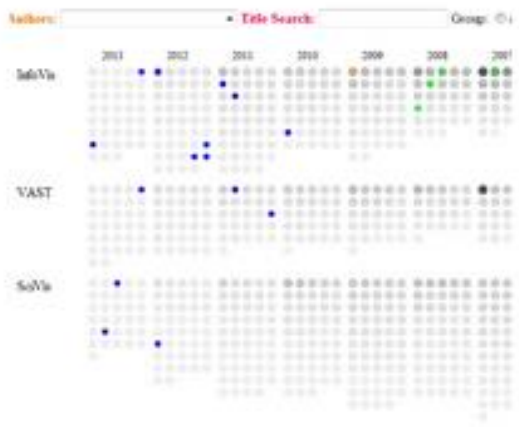


Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

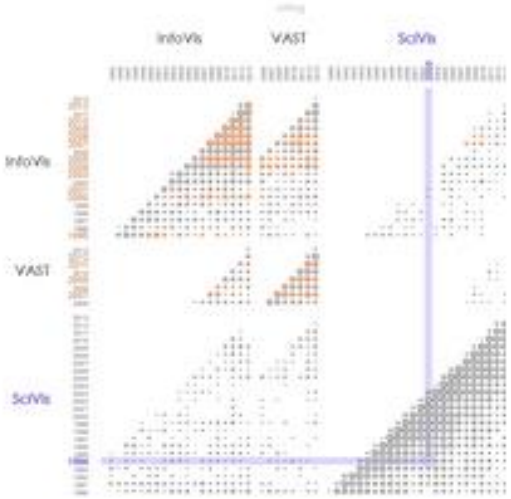
Demo

VIS25

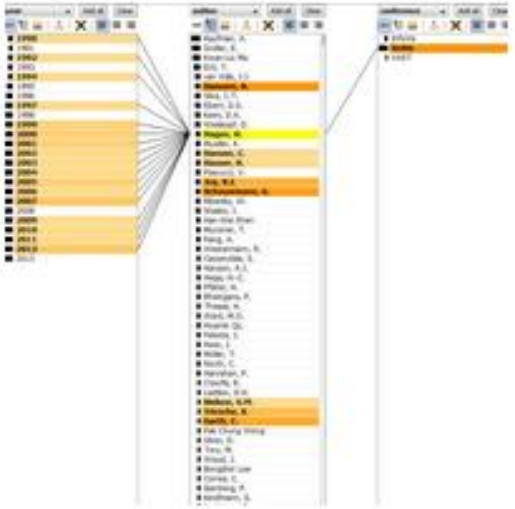
http://www.cc.gatech.edu/gvu/ii/citevis/VIS25/



CiteVis2



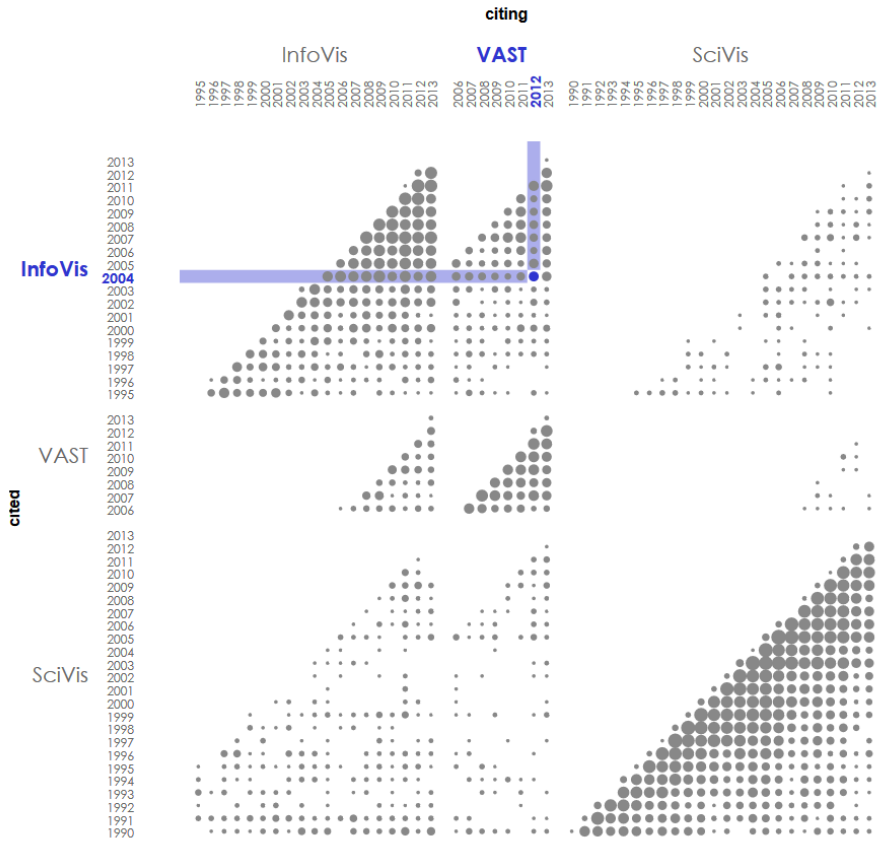
CiteMatrix



VISLists

Visualization design: Panpan Xu, Chad Stolper, Anand Sainath

CiteMatrix



Help | Back to VIS 25 home

Author: | Clear All

Cited Papers

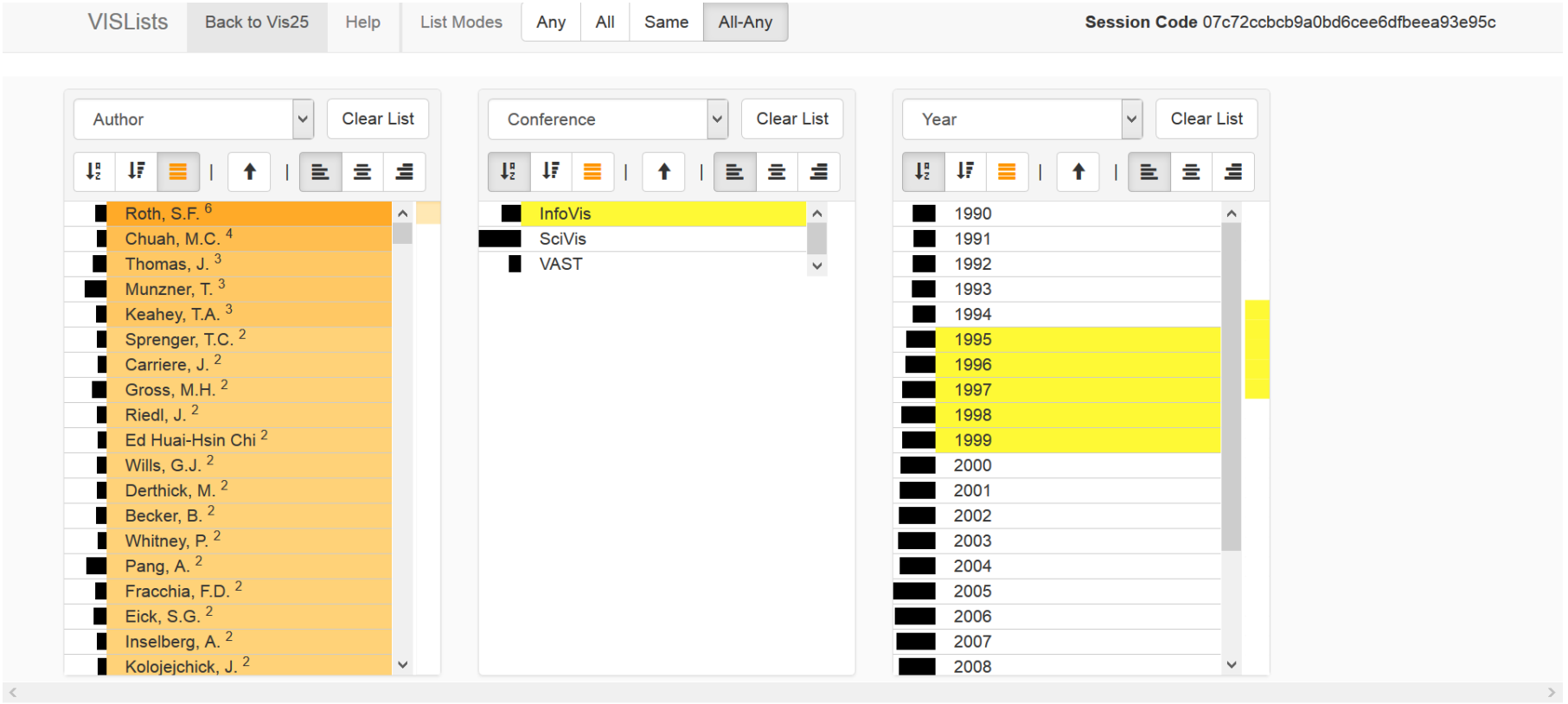
3--GeoTime Information Visualization
2--A Rank-by-Feature Framework for Unsupervised Multidimensional Data Exploration Using Low Dimensional Projections
2--Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering
2--Building Highly-Coordinated Visualizations in Improvise
1--An Evaluation of Microarray Visualization Tools for Biological Insight
1--A History Mechanism for Visual Data Mining
1--Value and Relation Display for Interactive Exploration of High Dimensional Datasets
1--Time-Varying Data Visualization Using Information Flocking Boids
1--The InfoVis Toolkit

P. Xu



Demo

VisLists

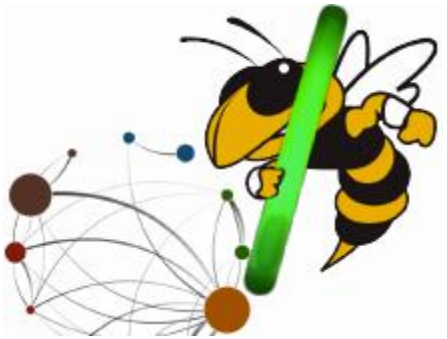


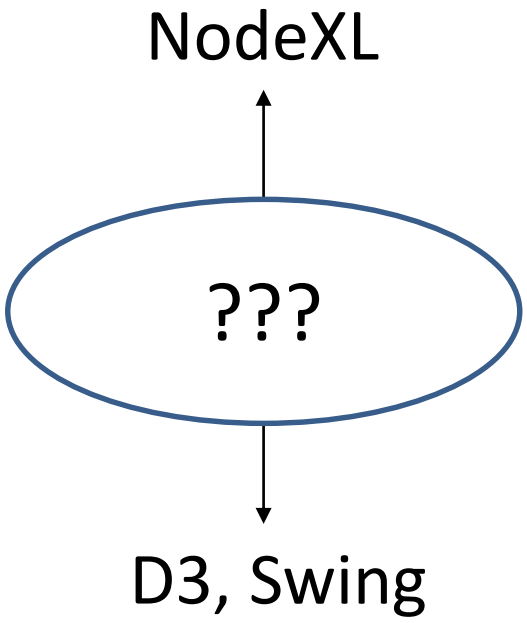
Demo

A. Sainath



Graph-Level Operations (GLOs)





Stolper, et al
TVCG (InfoVis) ‘14

GLO-STIX: Graph-Level Operations for Specifying Techniques and Interactive eXploration

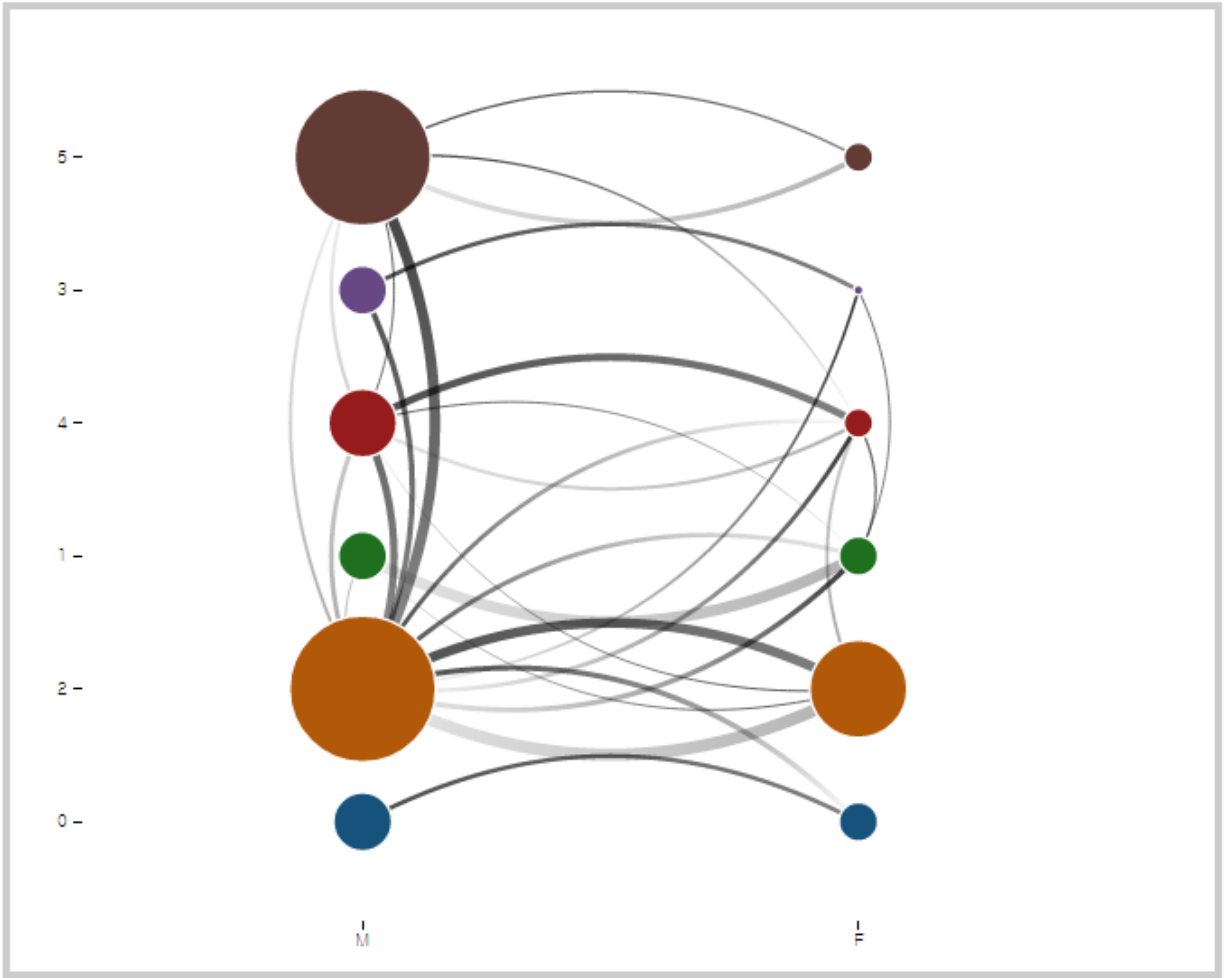
Select GLOs

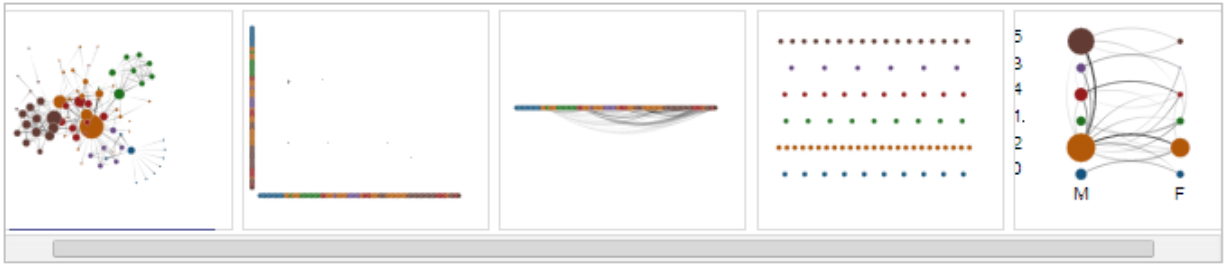
- Positioning Nodes
  - Align Left
  - Align Center
  - Align Right
  - Align Top
  - Align Middle
  - Align Bottom
  - Evenly Distribute on X
  - Evenly Distribute on Y
  - Evenly Distribute on X by Category
  - Evenly Distribute on Y by Category
  - Substrate on X by Gender
  - Substrate on Y by Category
  - Evenly Distribute on x Within Substrates
  - Position on X Relatively by Betweenness Cent.
- Modifying Element Properties
- Cloning
- Aggregation
- Modifying Display Properties

Applied GLOs

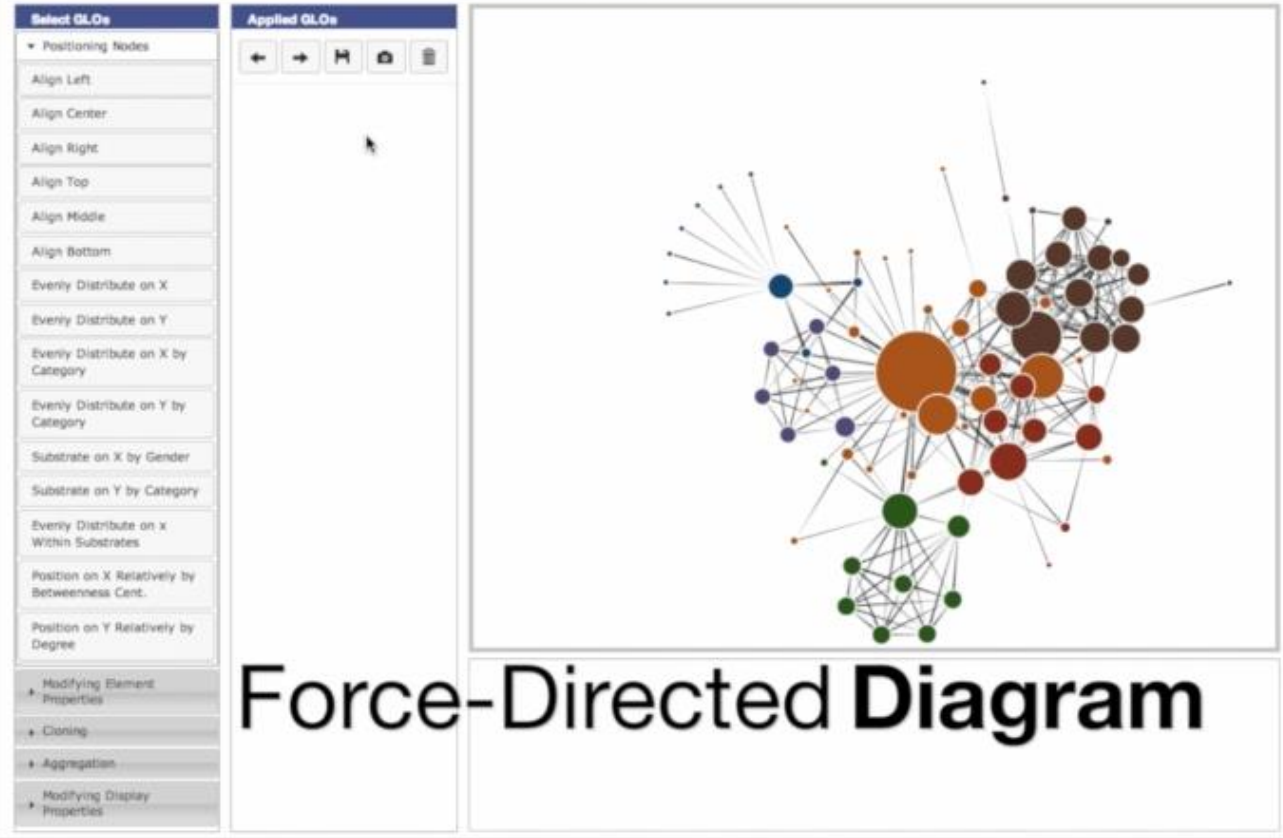


- Evenly Distribute on Y
- Align Left
- Set Source Generation 1
- Links to Circles
- Links to Curved
- Set Source Generation 0
- Remove Generation 1
- Align Middle
- Substrate on Y by Category
- Evenly Distribute on x Within Substrates
- Show Select Links
- Show All Links
- Substrate on X by Gender
- Show X Axis
- Show Y Axis
- Aggregate by Gender and Category





GLO-STIX: Graph-Level Operations for Specifying Techniques and Interactive eXploration



Rethought Notion 1

Consider alternative network representations

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

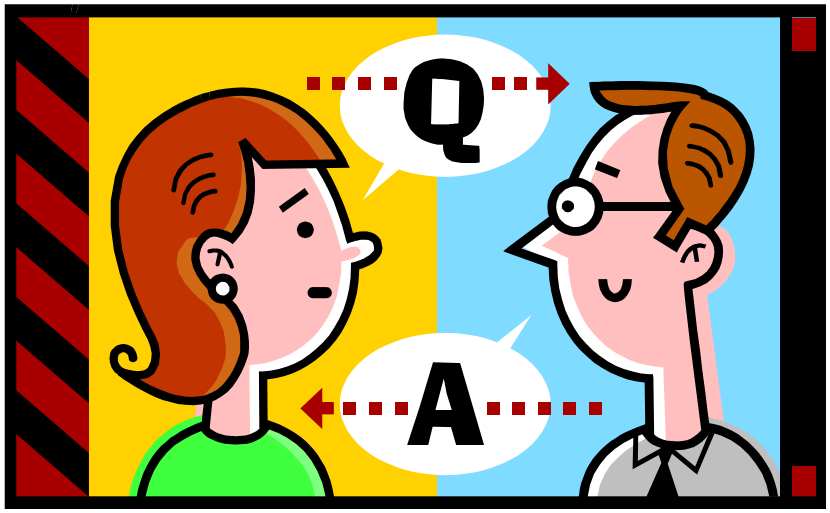
Two key aspects of data visualization

- 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

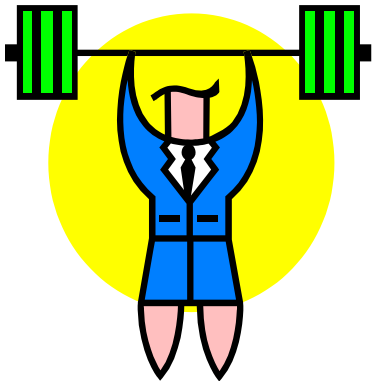
Get details

Navigation

Brushing & linking

Interaction

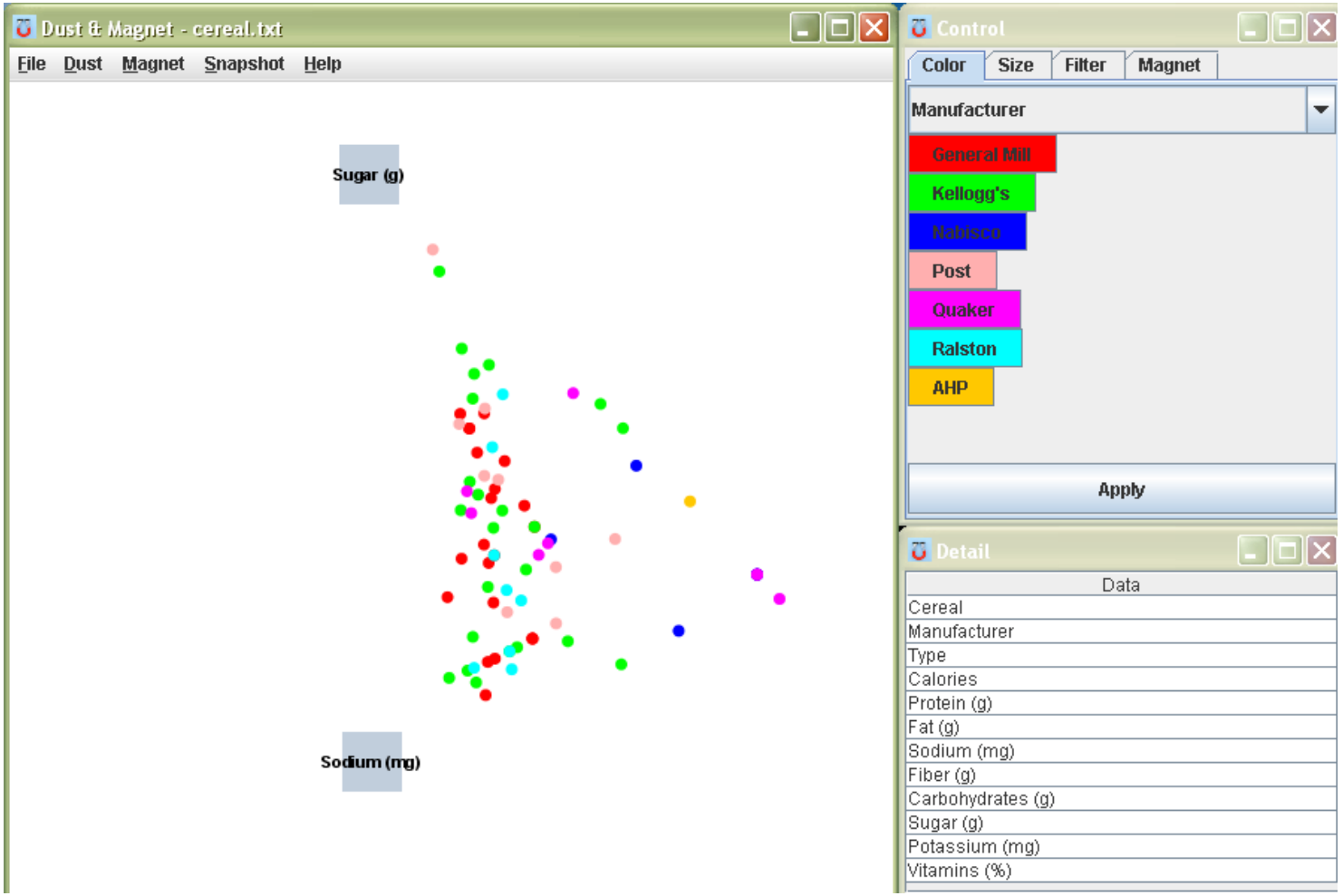
Can we do more?



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

Some examples…

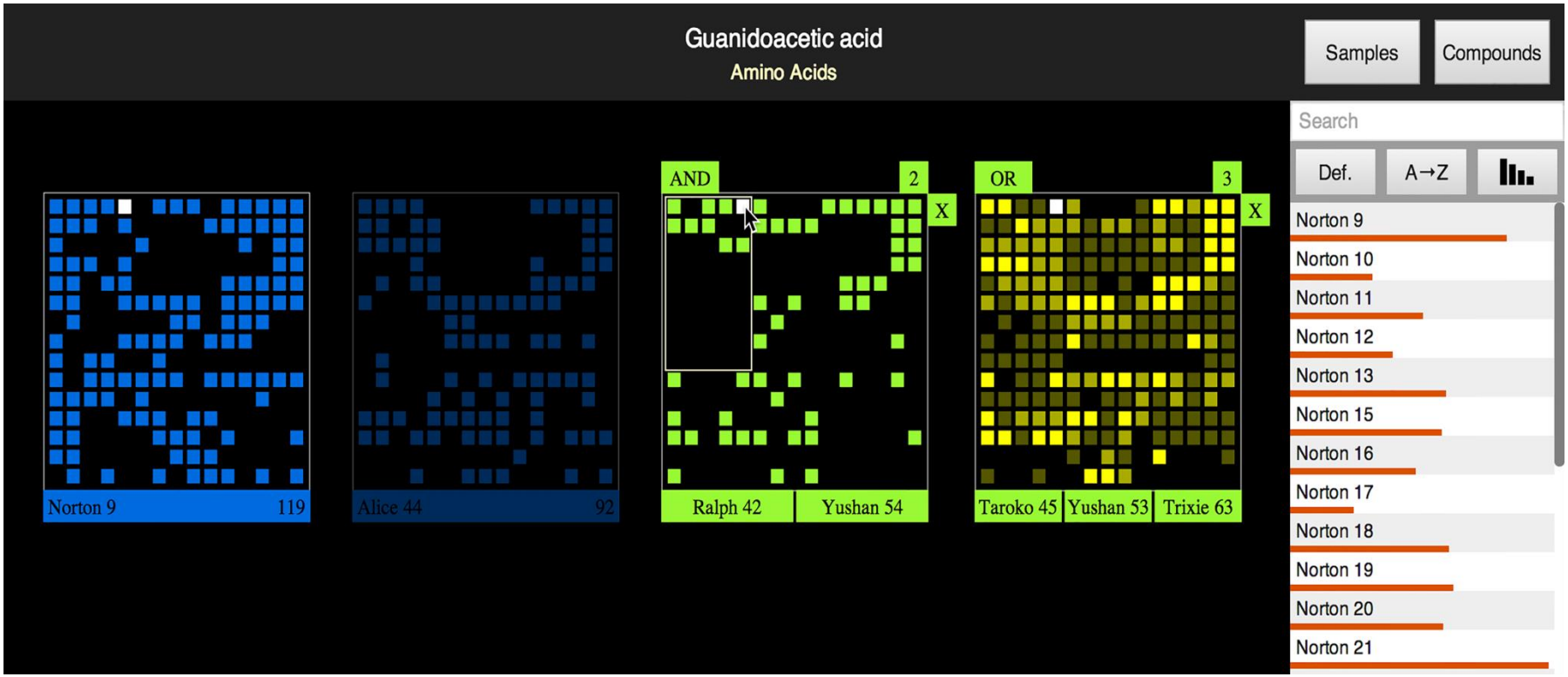
Dust and Magnet



Yi, Melton, Stasko & Jacko
Information Visualization '05

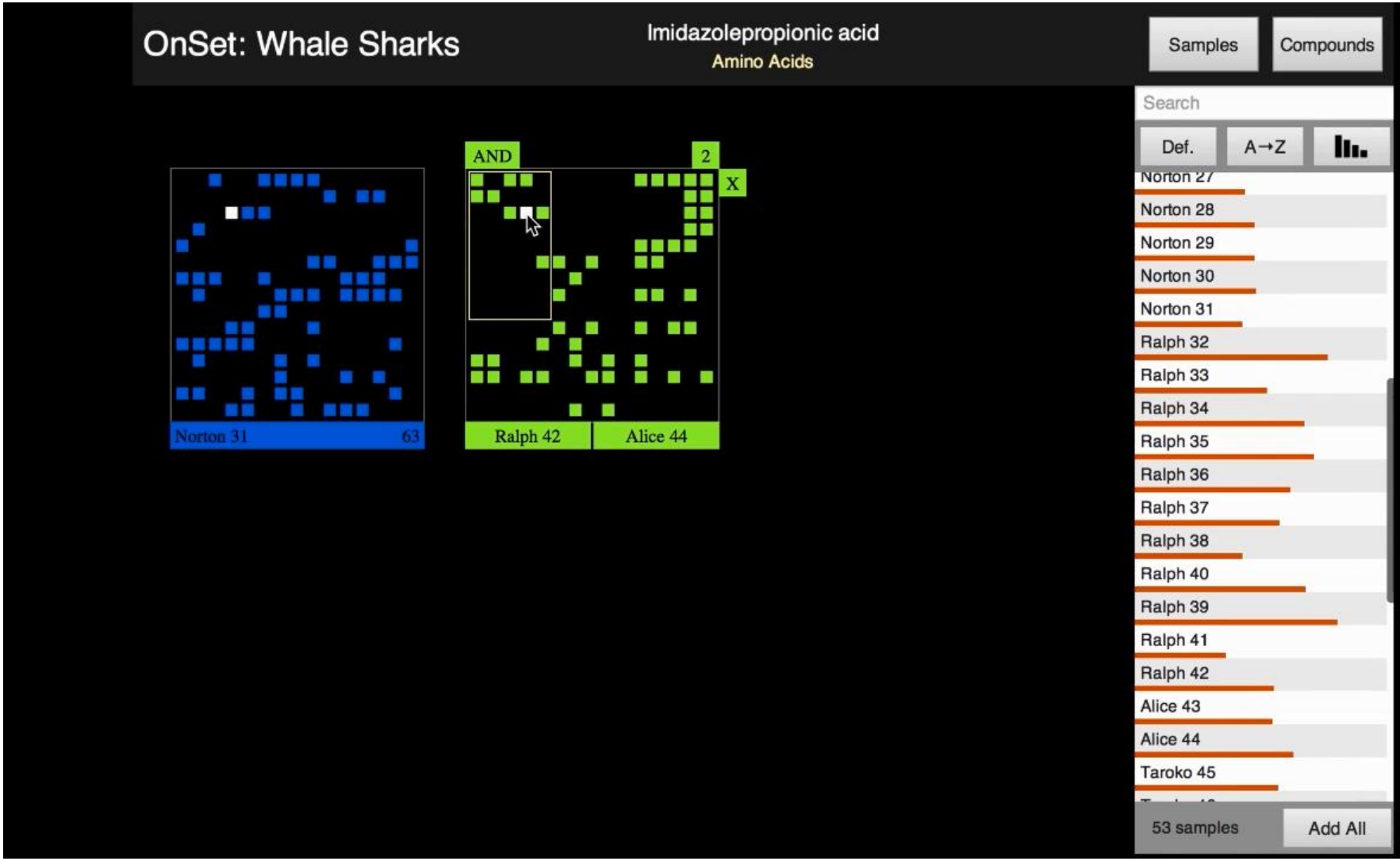
OnSet

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



Demo

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


OnSet: Whale Sharks
Imidazolepropionic acid
Amino Acids
Samples
Compounds
Search
Def.
A→Z
Norton 27
Norton 28
Norton 29
Norton 30
Norton 31
Ralph 32
Ralph 33
Ralph 34
Ralph 35
Ralph 36
Ralph 37
Ralph 38
Ralph 40
Ralph 39
Ralph 41
Ralph 42
Alice 43
Alice 44
Taroko 45
53 samples
Add All
AND
2
X
Norton 31
63
Ralph 42
Alice 44

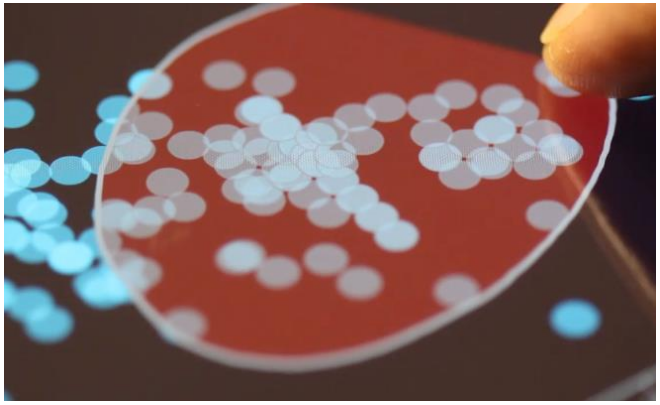
Interaction

What are the tools of interaction?

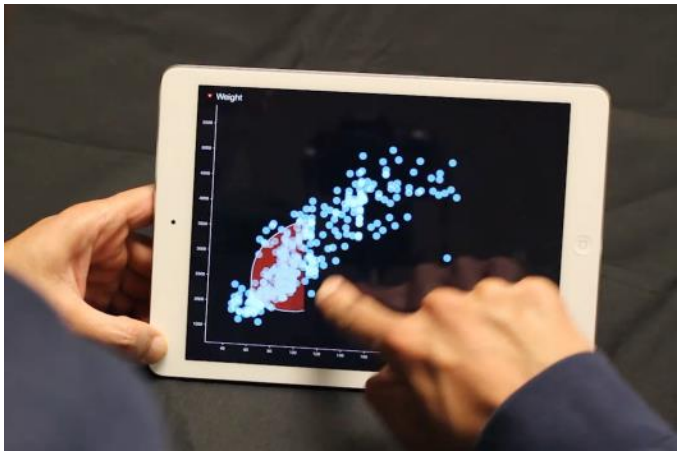
Traditional – Desktop: keyboard, mouse

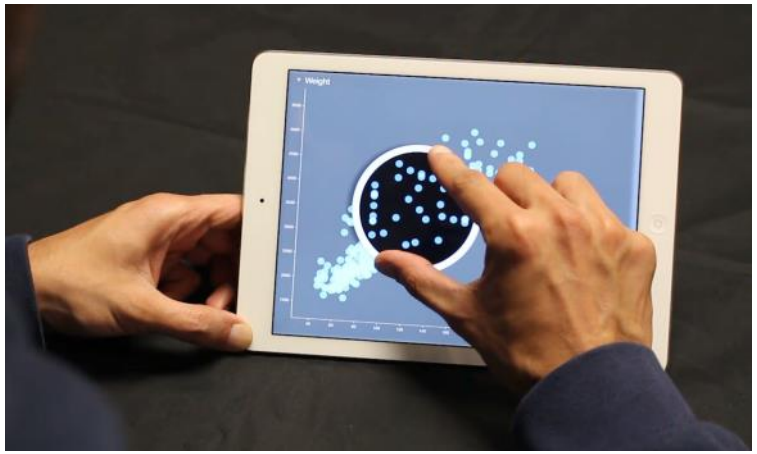
New – Tablet: fingers and multi-touch

Moving to Tablets



Scatterplot



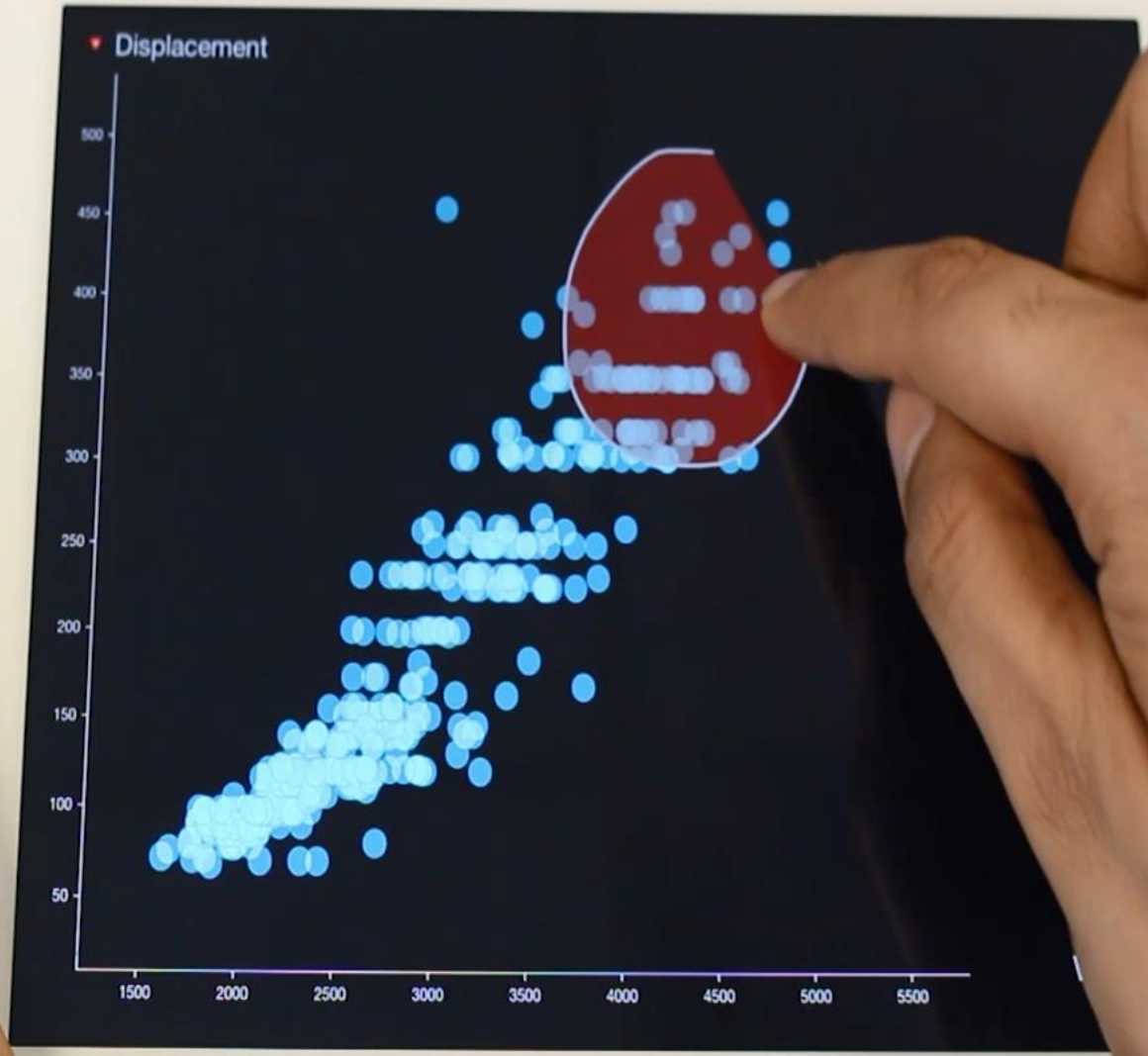




Video

Sadana & Stasko
AVI ‘14

Lasso Select



Dust and Magnet - Pushing it Further



Dai et al
InfoVis ‘15 poster

“Putting the pieces together”

Jigsaw

Computational analysis of documents’ text

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

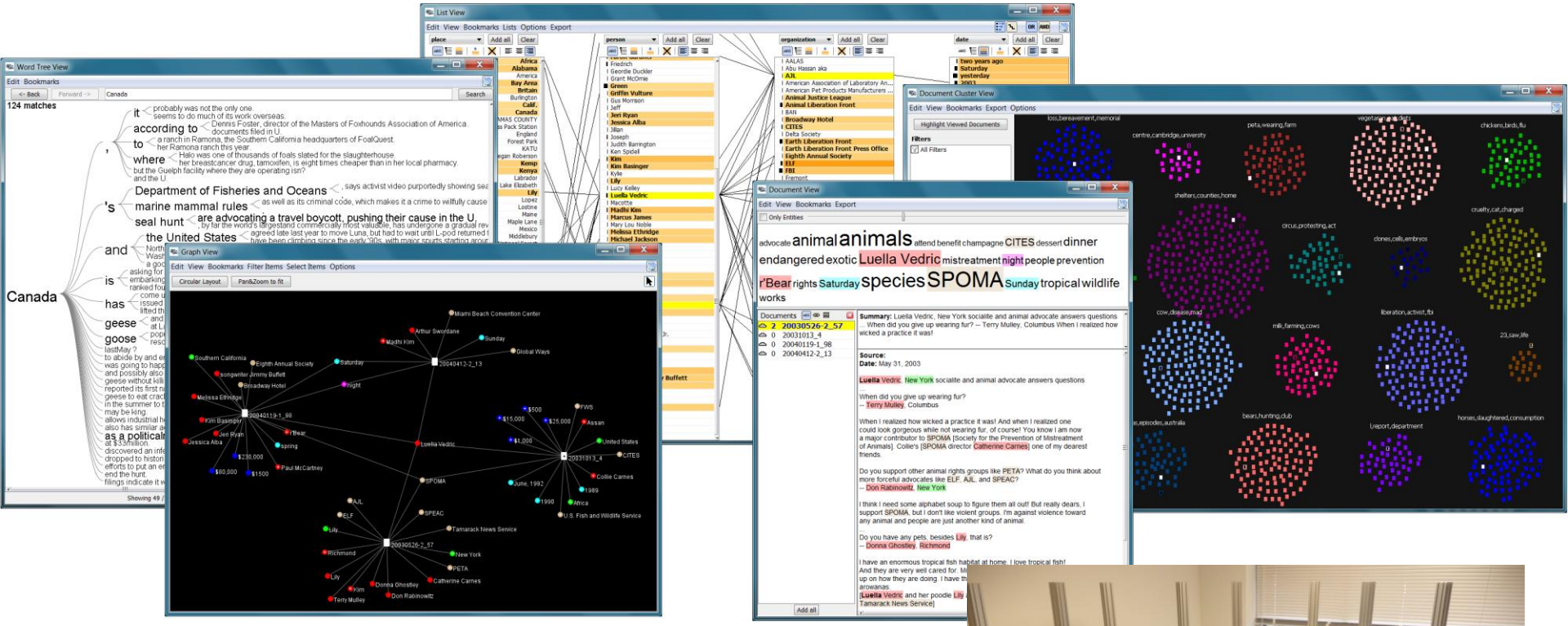
Multiple visualizations of documents, analysis results, entities, and their connections

- Views are highly coordinated

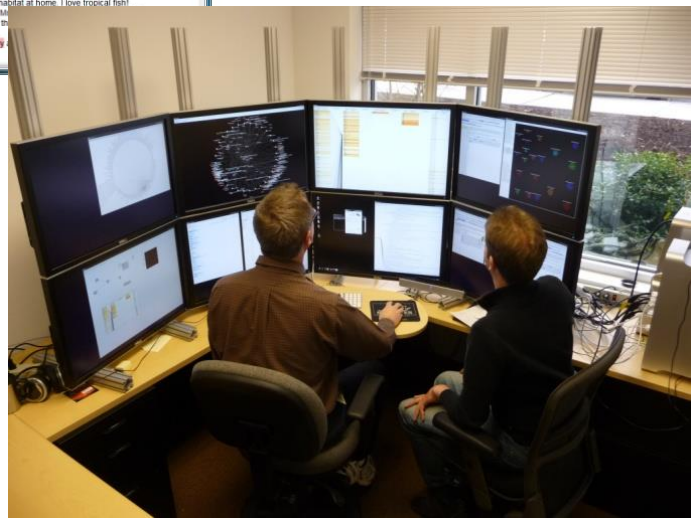


http://www.cc.gatech.edu/gvu/ii/jigsaw

Jigsaw



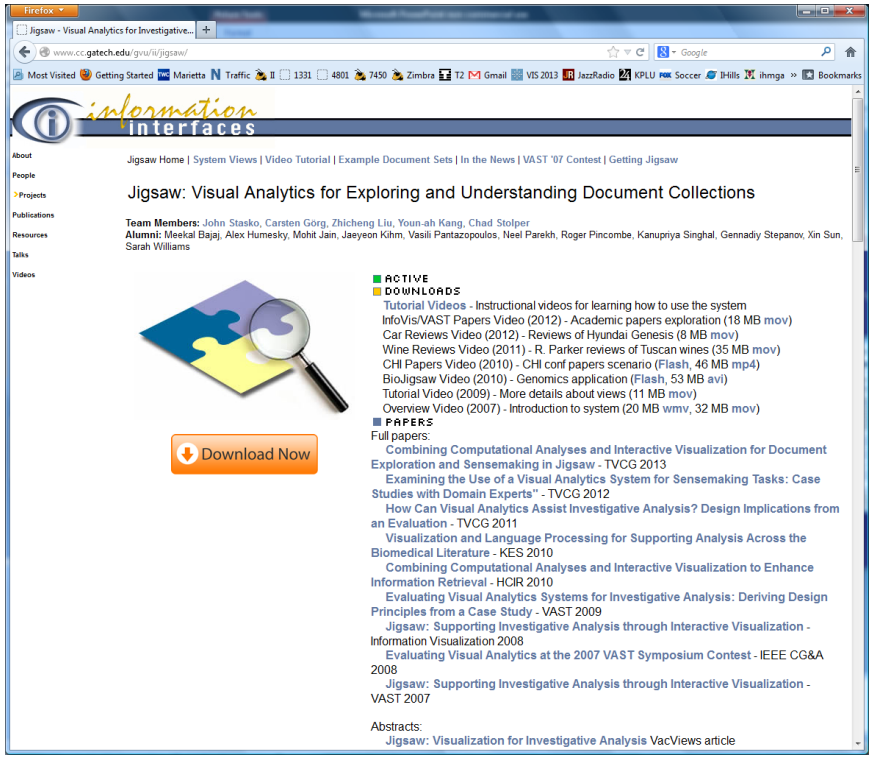
Visual analytics of large text document collections



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

Download for free

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





Rethought Notion 2

Take better advantage of interaction

How do we determine if a visualization is “good”?

User Studies (typically)

Short, fact-based questions & tasks

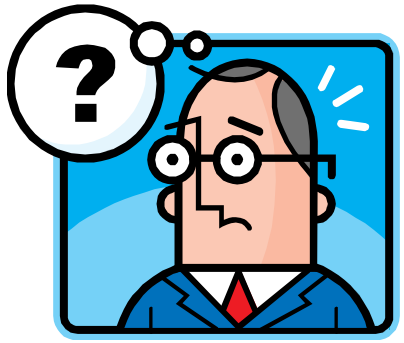
Example: Car dataset

- 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?

Important for usability, but can get those answers in other ways

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



Rethink evaluation as identifying a visualization's value

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

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

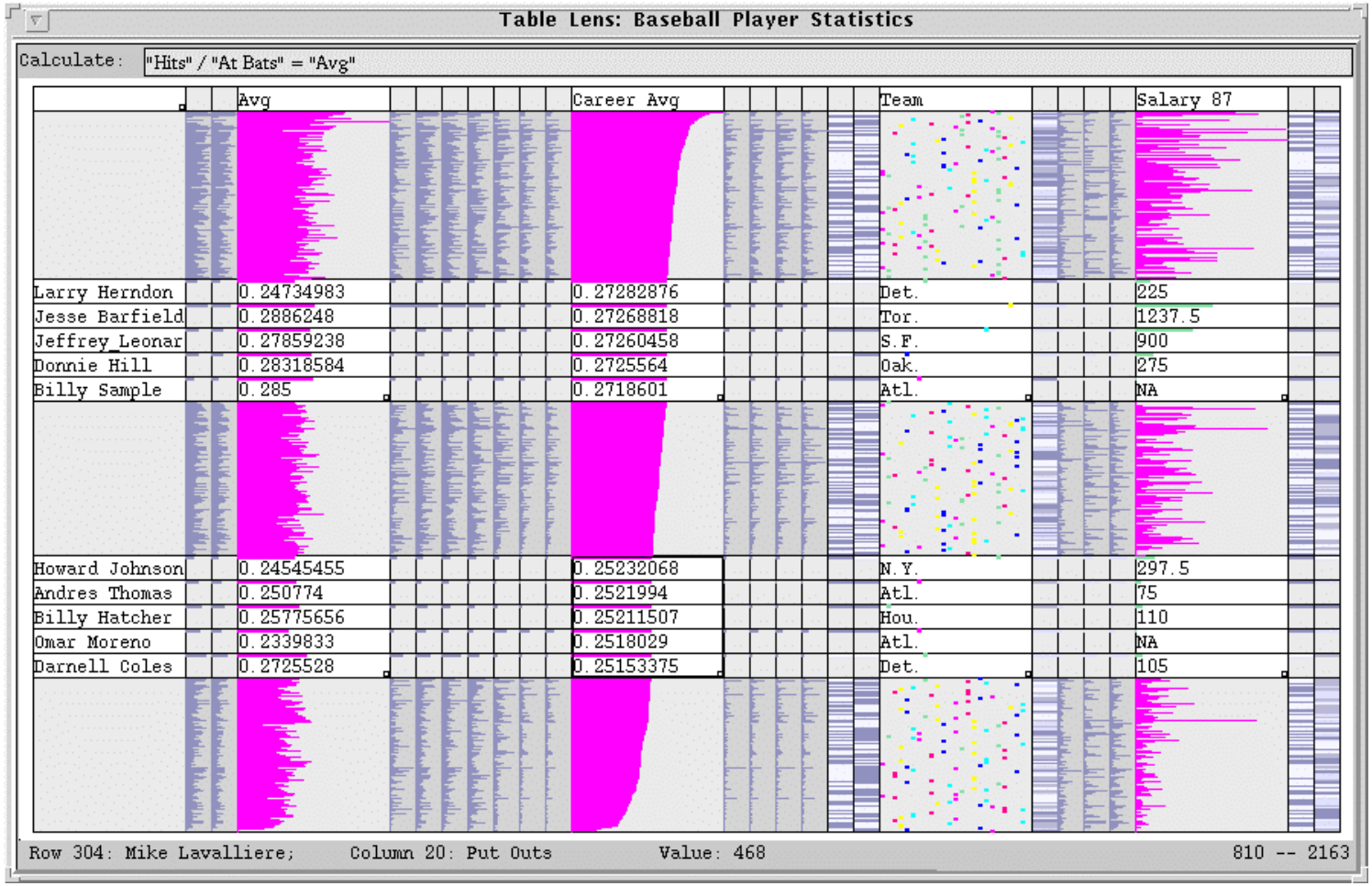


Table Lens

Rao & Card
CHI ‘94

Visualization's Value?

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

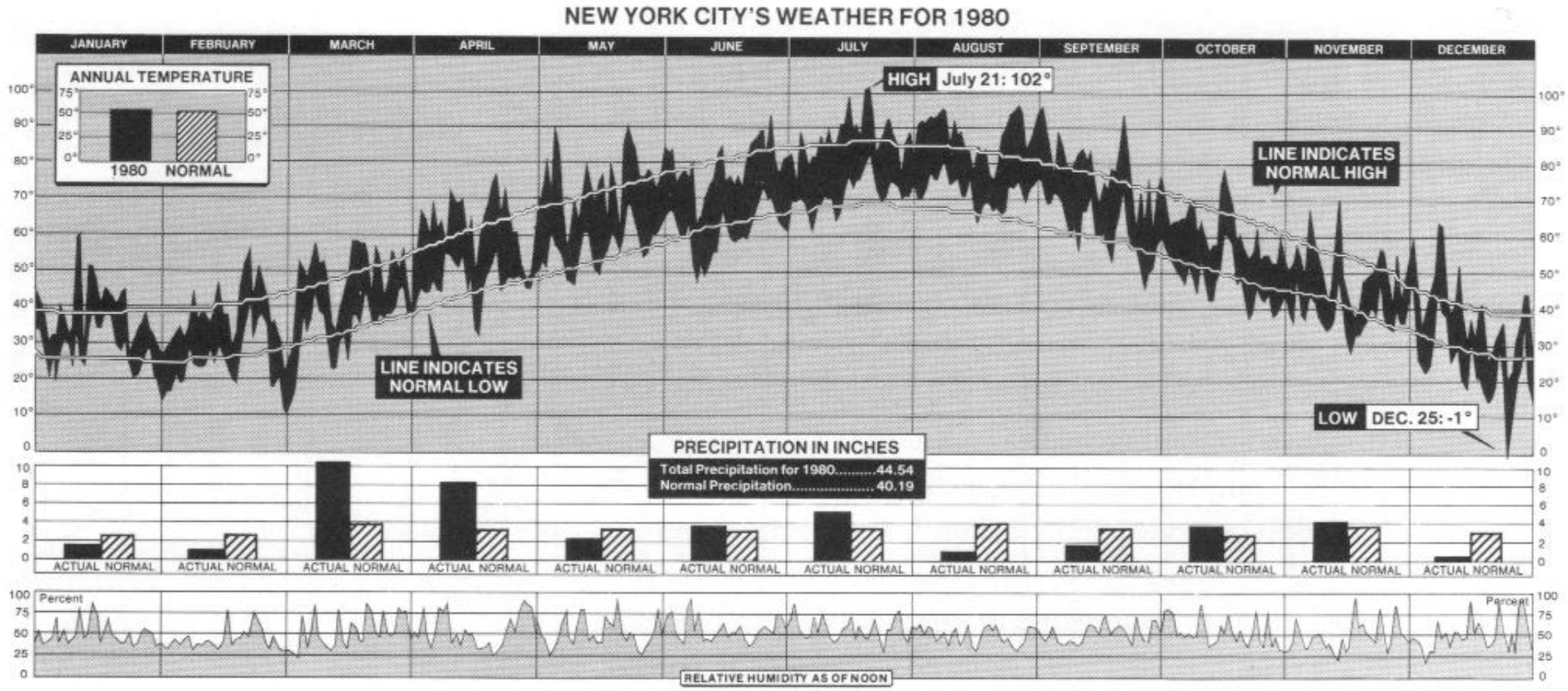
Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)

Some examples

1.

2220 numbers



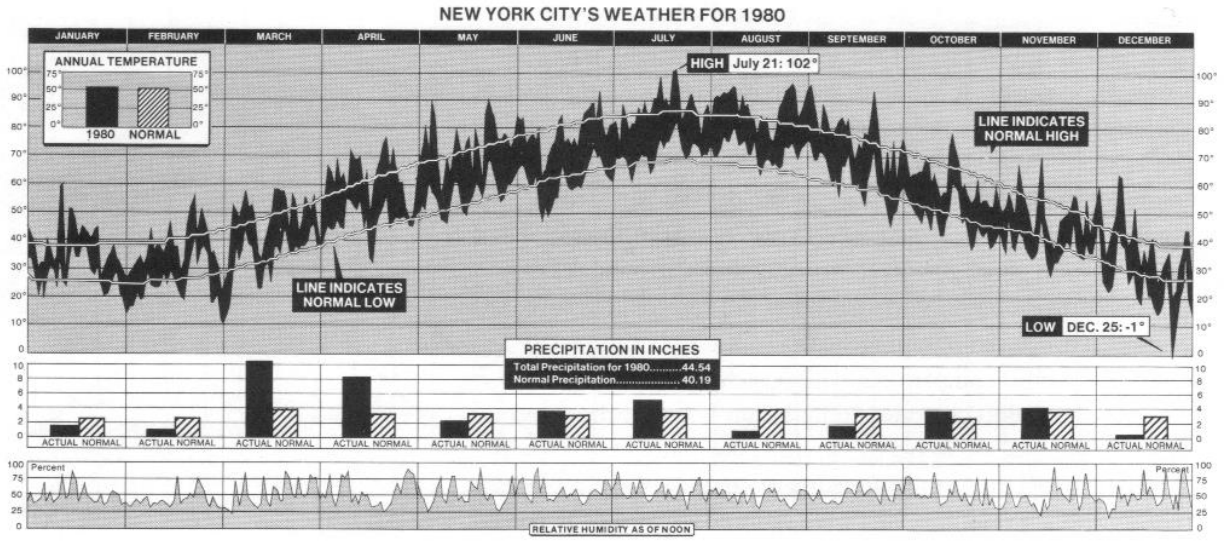
New York Times, January 11, 1981, p. 32.

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

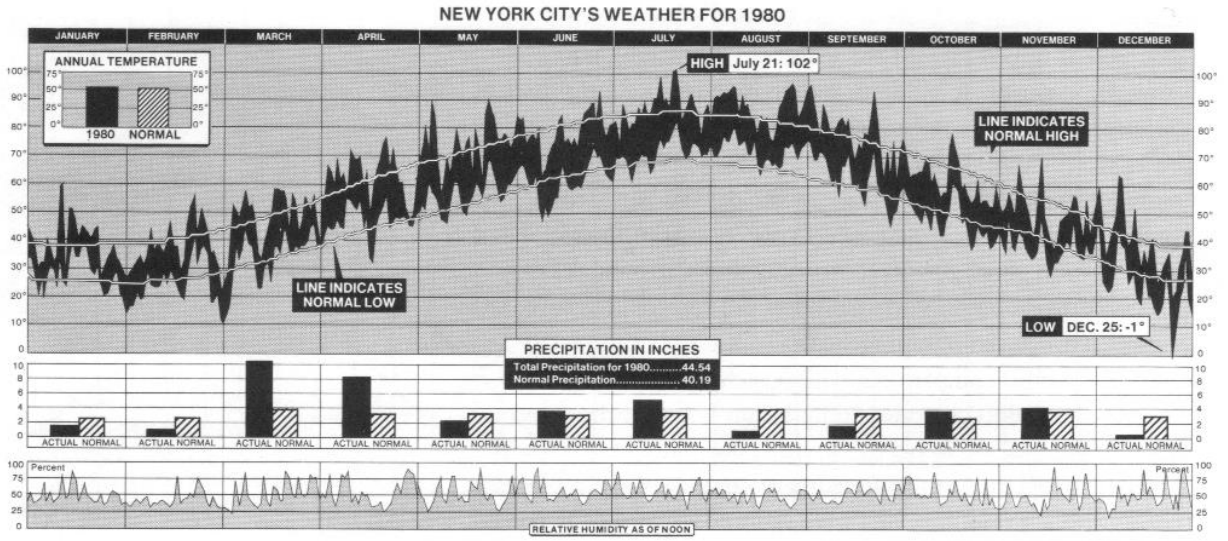
Visualization's Value



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

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

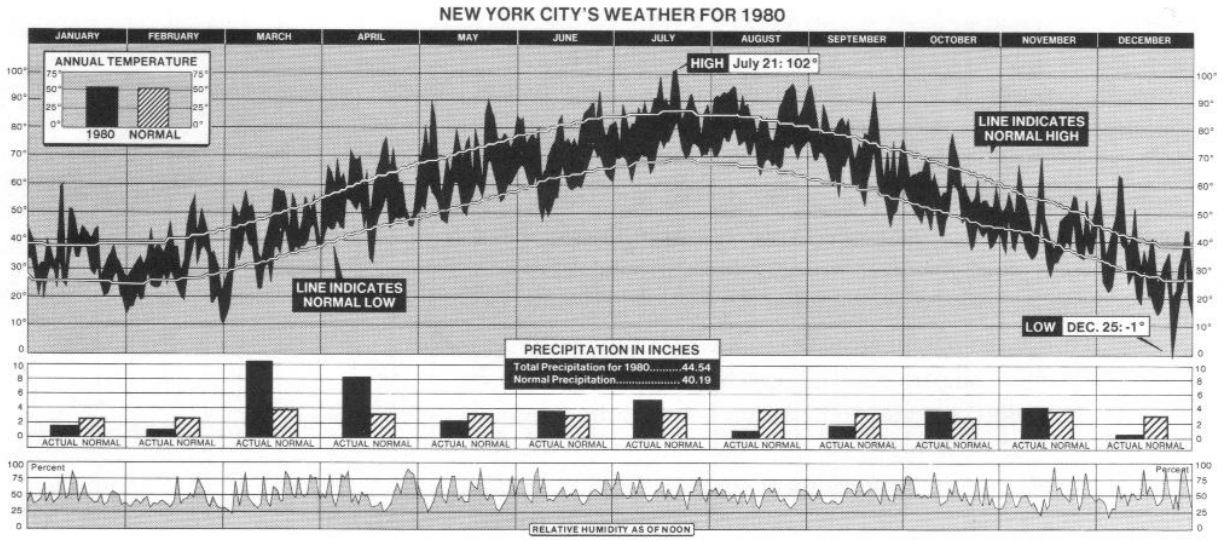
Visualization's Value



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

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

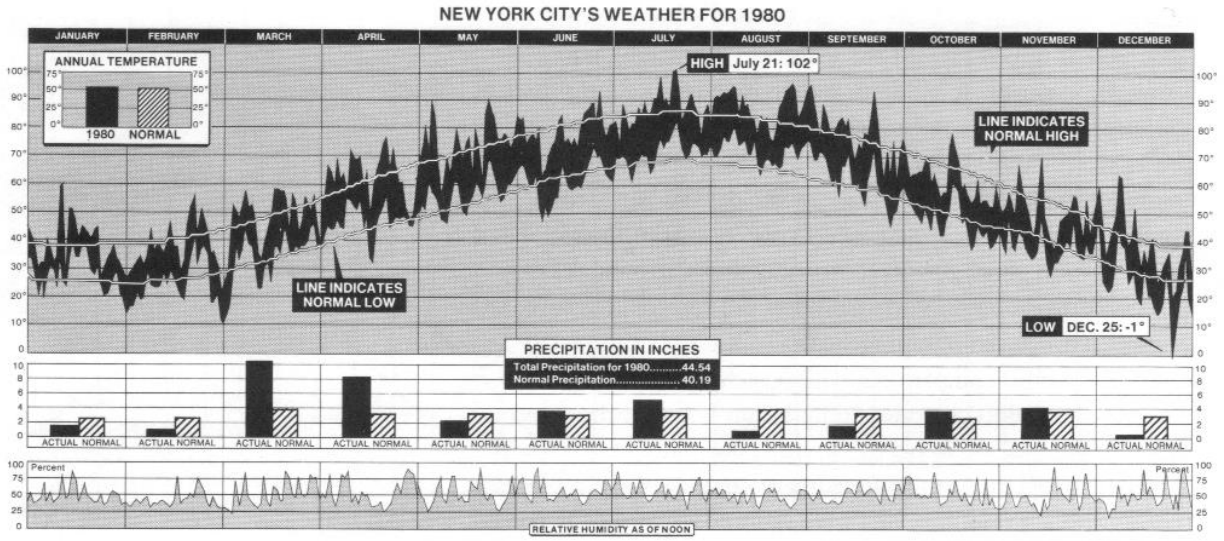
Visualization's Value



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

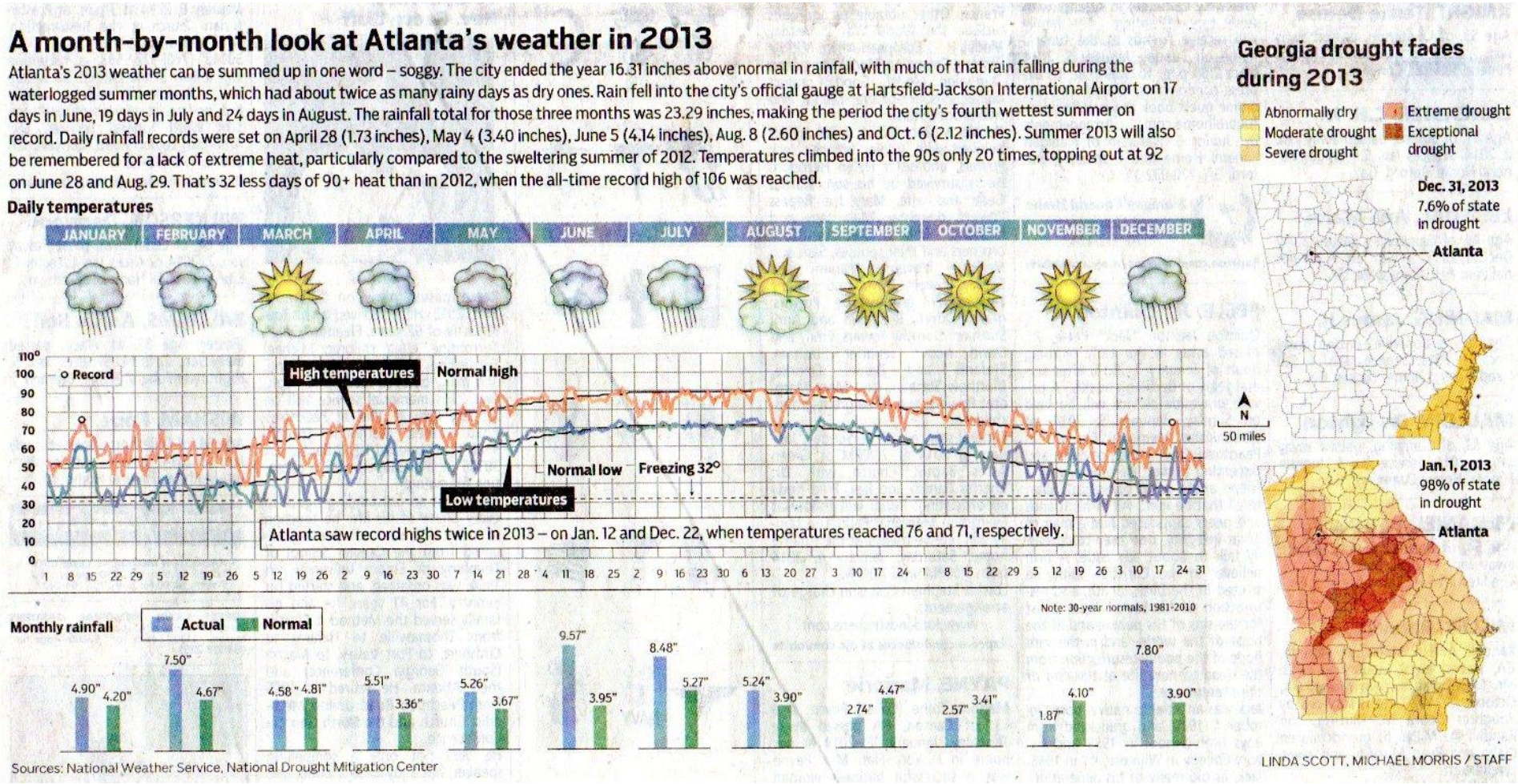
$$V_{alue} = T + I + E + C$$

Visualization's Value



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

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

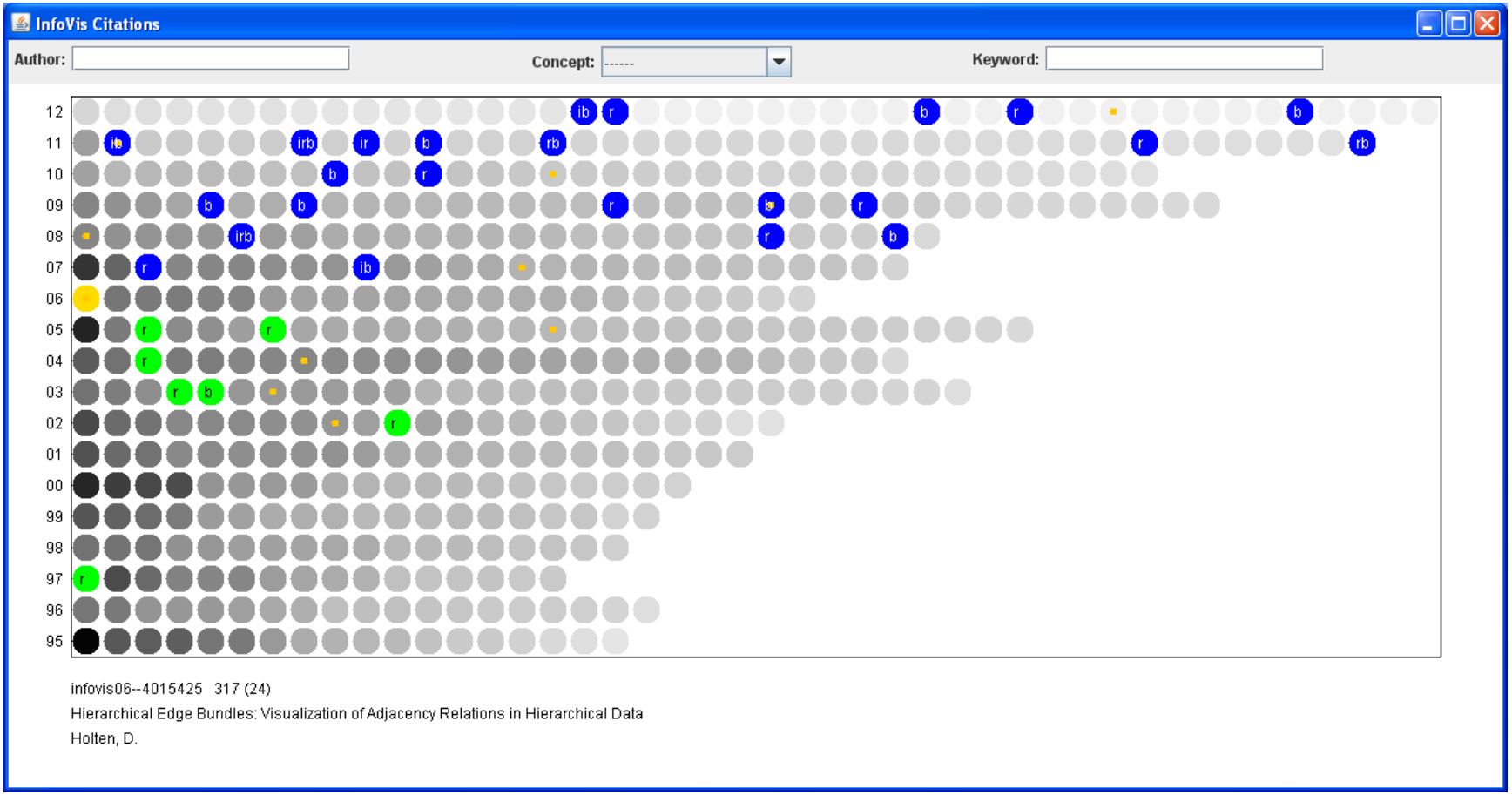


Atlanta Journal Constitution
Jan. 3, 2014

2.

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

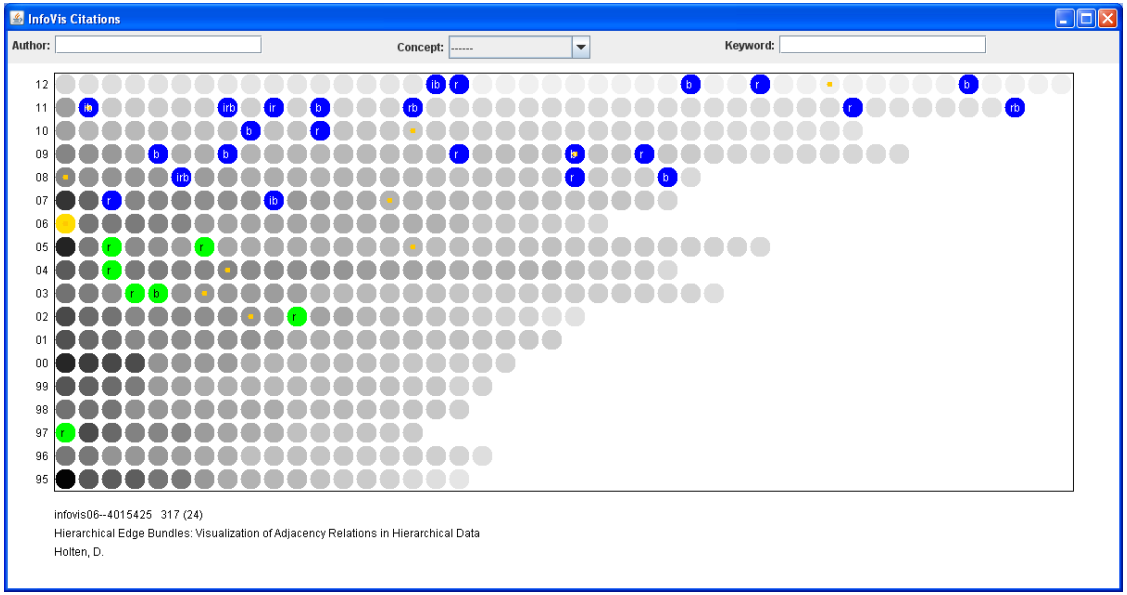
CiteVis



Earlier demo

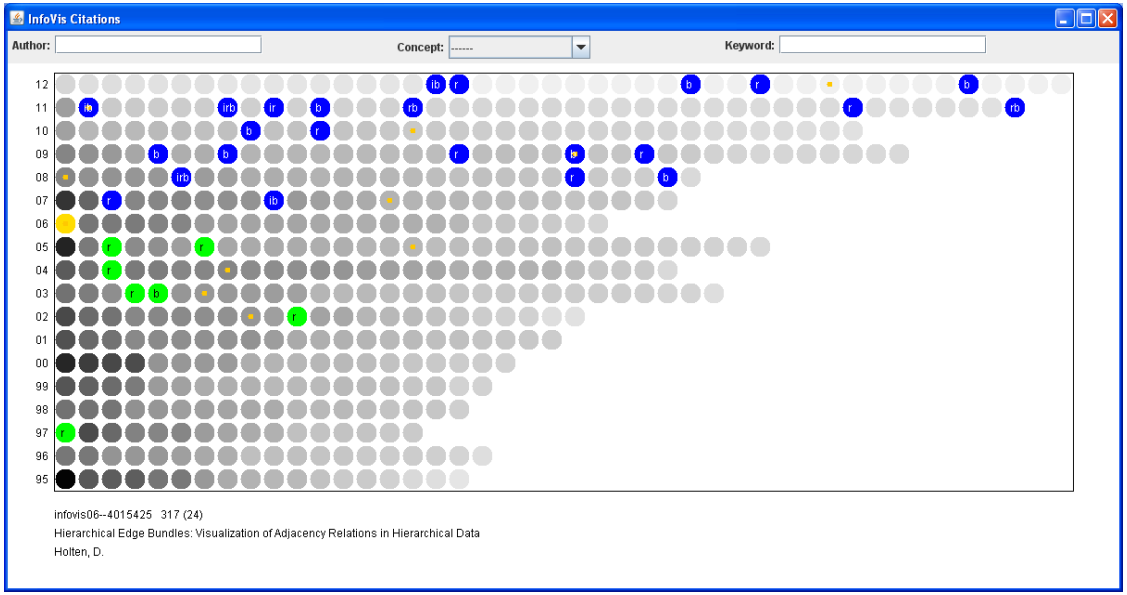
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

Visualization's Value



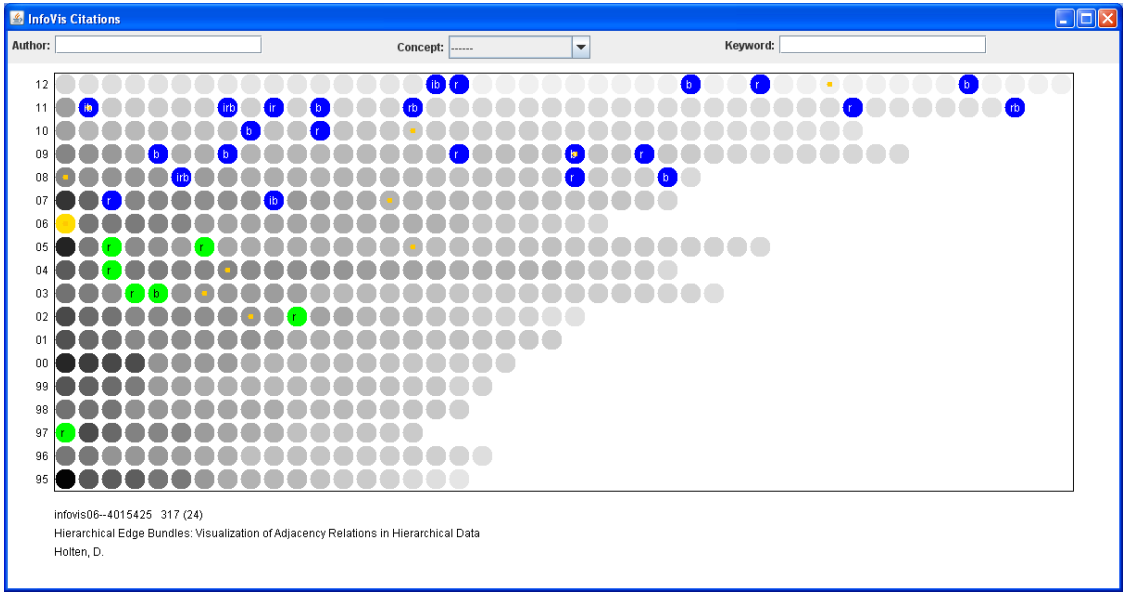
$$V_{alue} = T + I + E + C$$

Visualization's Value



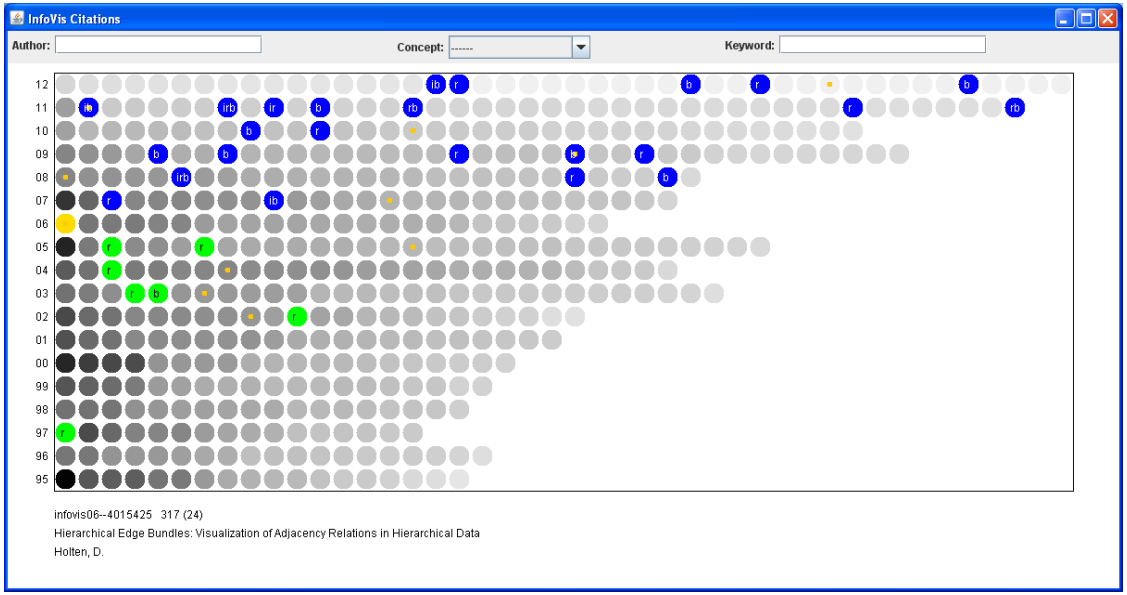
$$V_{alue} = T + I + E + C$$

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

Rethought Notion 3

Evaluation through value identification & communication

Key Open Problems in Visualization

Key Open Problems in Visualization

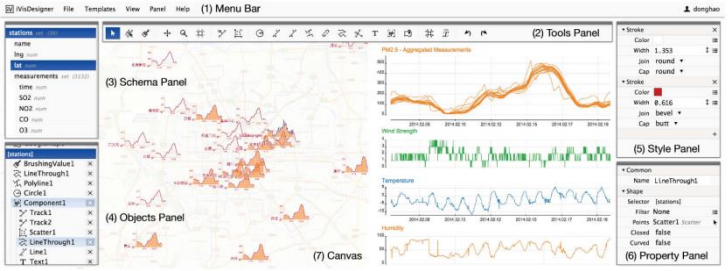
1. Assess and communicate **value**

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

Key Open Problems in Visualization

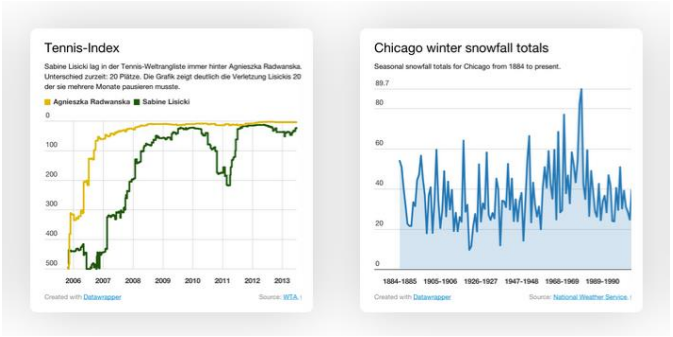
1. Assess and communicate value

2. Make the construction of visualizations easier



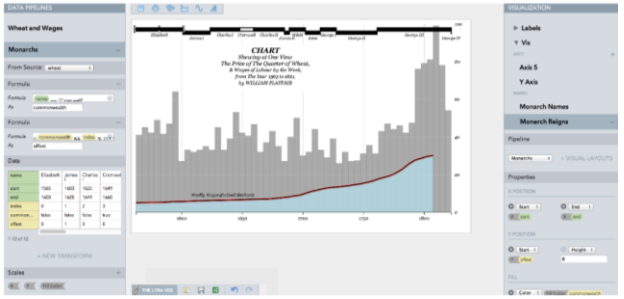
iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) '14

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

Reconsider network visualization

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)



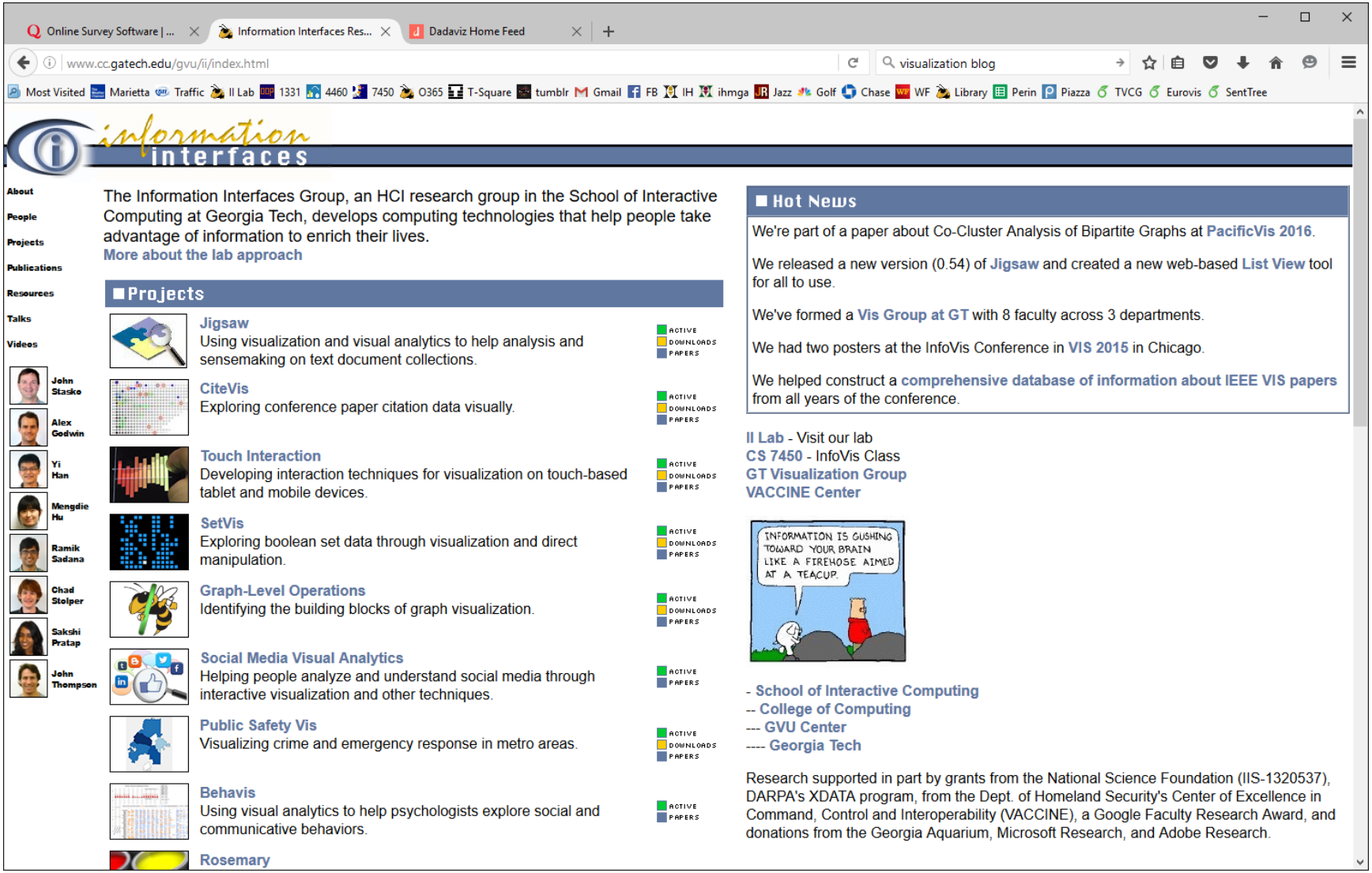


- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537



- Supported by gifts from Microsoft, Google, Adobe, and Nokia

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



Putting User Interaction First: **Interactive Visual Affordances** for Visual Analytics

Alex Endert
Assistant Professor
School of Interactive Computing, Georgia Tech
endert@gatech.edu

VA.lab

goals for this talk

- **share an idea** that's interesting to my students and I
- motivate the idea with some **prior and current work**
- stimulate **discussion** and **collaboration**

a bit about me

2008-2012	Virginia Tech
2012 - 2014	PNNL
(current)	Georgia Tech



Interactive Computing



Rahul C. Basole
Associate Professor



Alex Endert
Assistant Professor



Jim Foley
Professor



John Stasko
Professor

Computational Science & Engineering



Polo Chau
Assistant Professor
Associate Director,
MS Analytics



Jimeng Sun
Associate Professor

Literature, Media & Communications



Nassim JafariNaimi
Assistant Professor



Lauren Klein
Assistant Professor



Yanni Loukissas
Assistant Professorr

vis.gatech.edu

Georgia Tech Visualization Lab

Interactive **Affordances**

James J. Gibson



- Affordances
  - objects in our physical environment afford **opportunities for action**
  - based on prior knowledge, context and situation, perception/vision, and situated actions and goals

"The Theory of Affordances", 1977

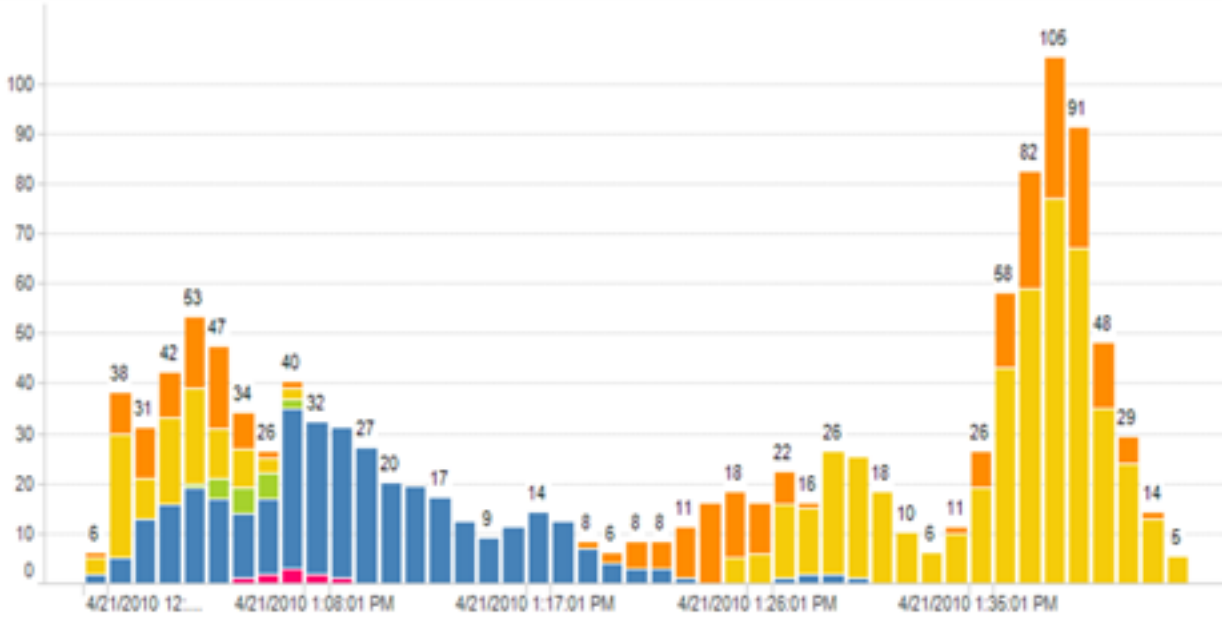




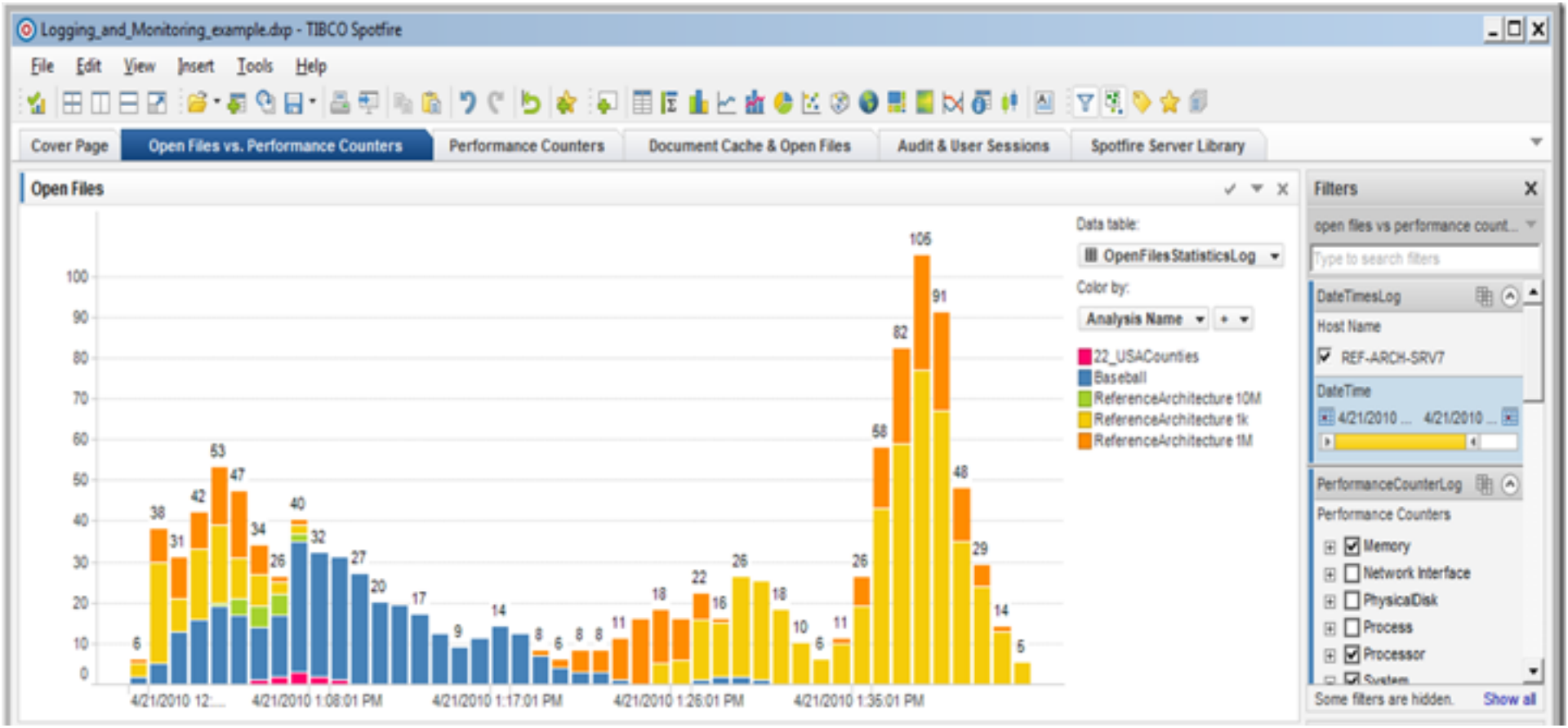
Information Visualization

- **Binding data values to graphical encodings** to produce **interactive** visual representations of datasets

- foster **analytic discourse** with the data through user interaction
  - discovery
  - sensemaking



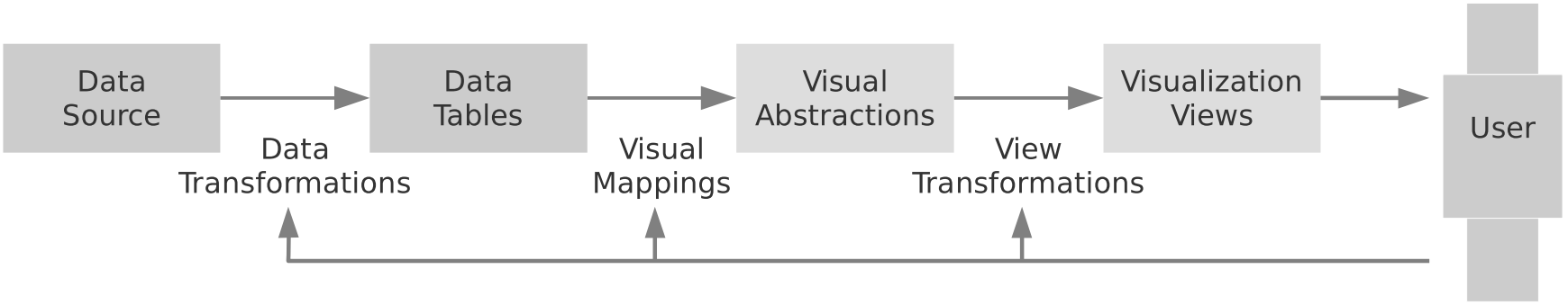
information visualization



interactive information visualization

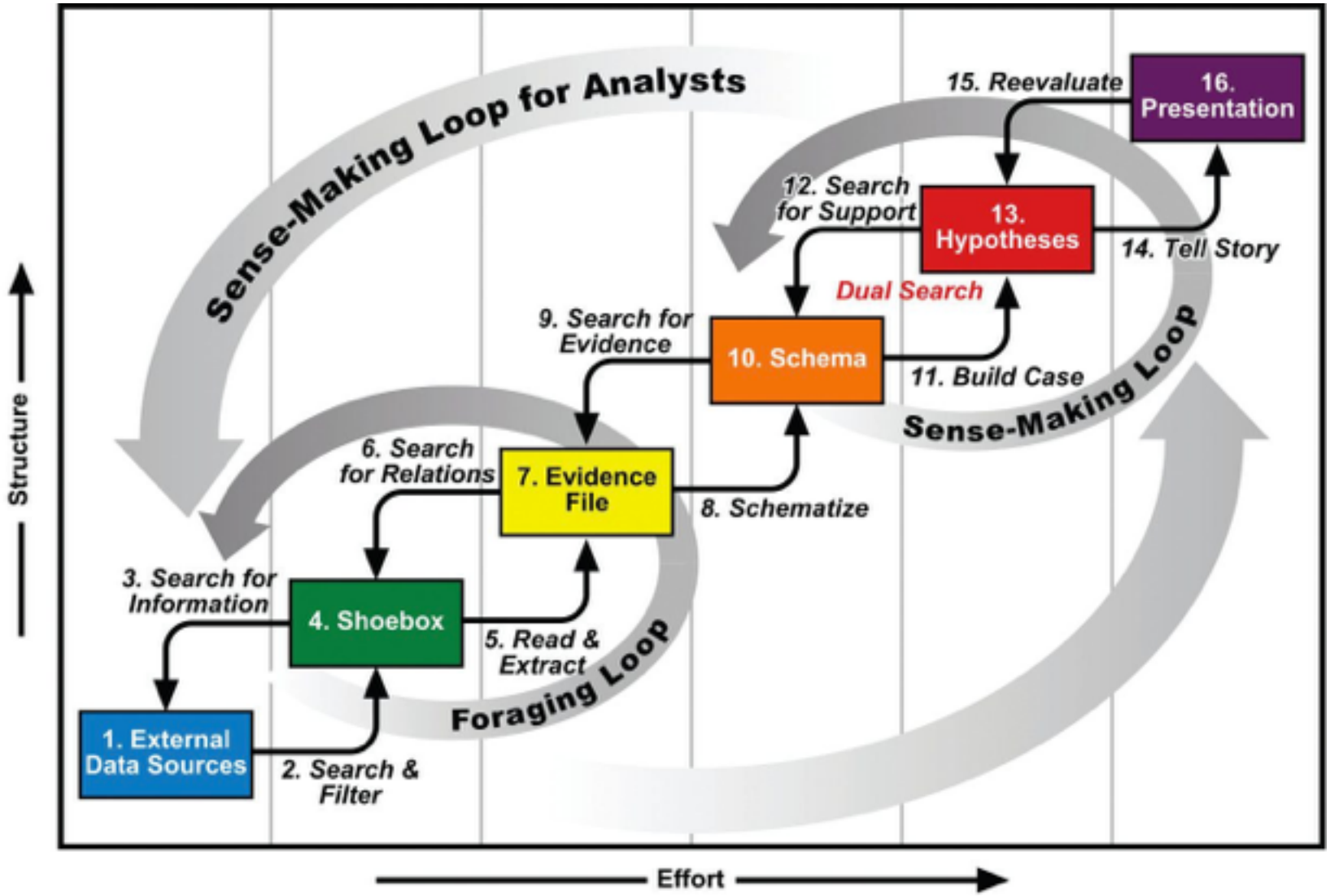
the importance of user interaction

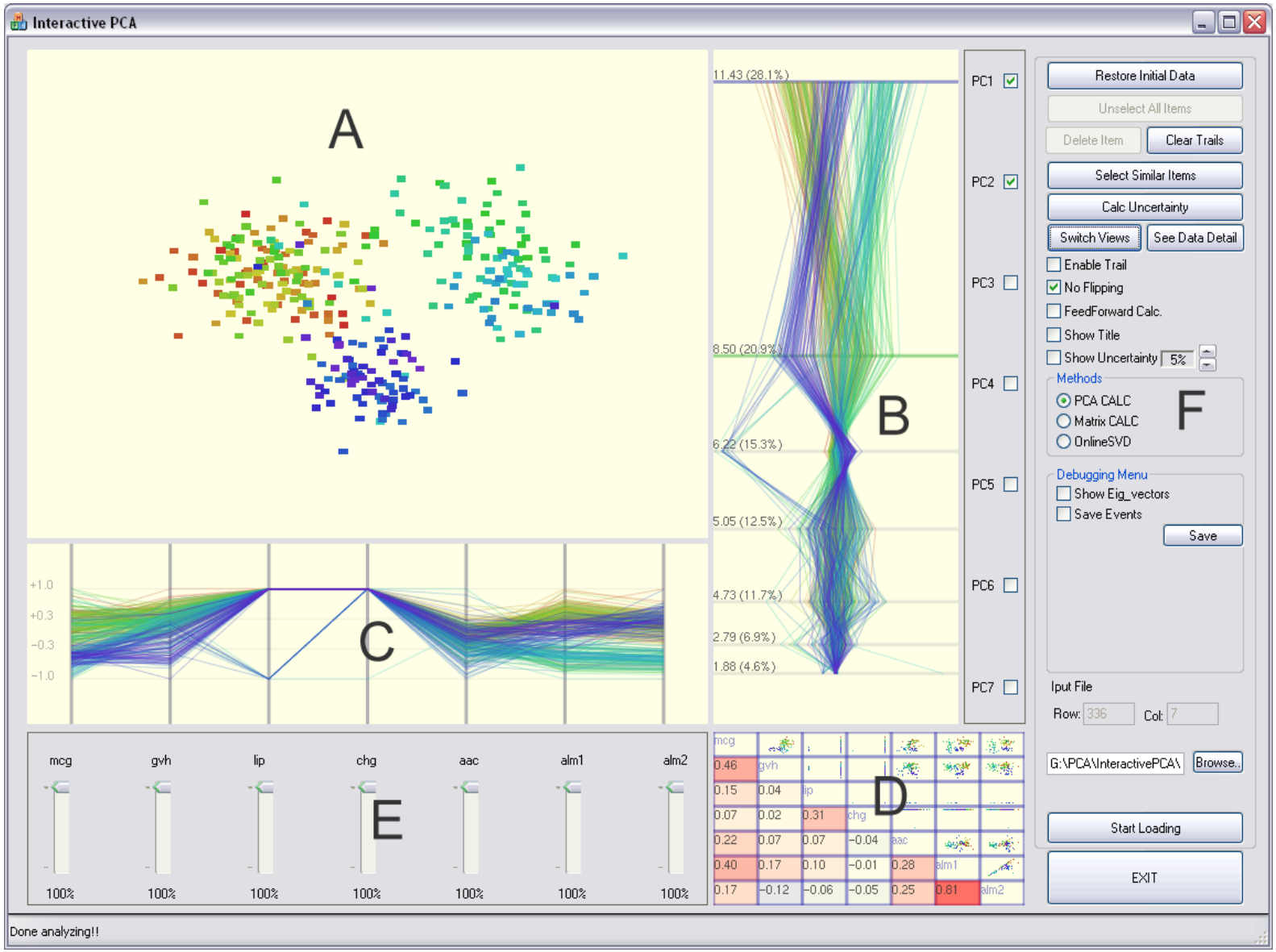
- direct manipulation (Shneiderman, 1985)
  - direct, **explicit control** over visualization and data parameters



Card et al., 1999

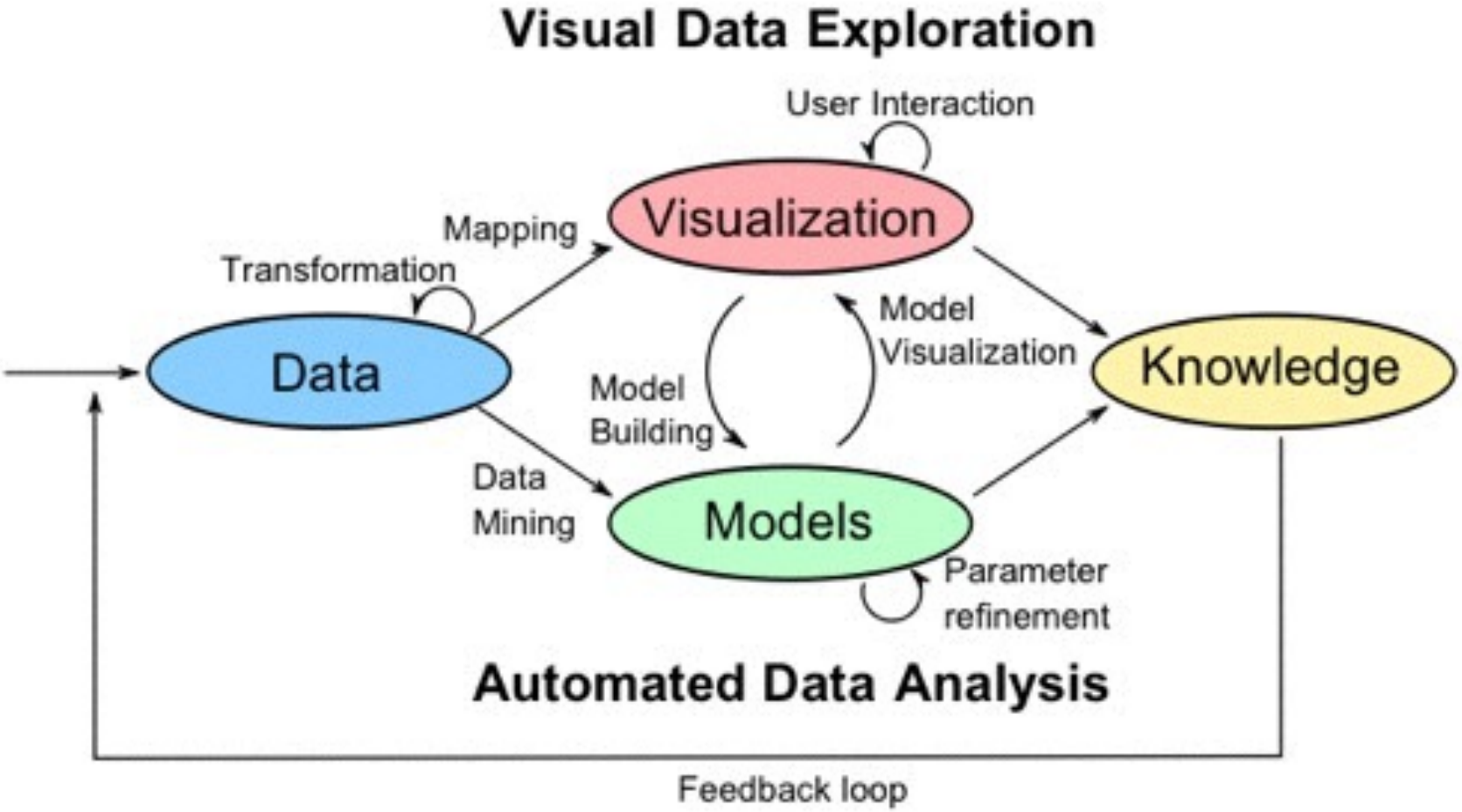
Sensemaking Loop (Pirolli and Card)





interactive visual analytics

incorporating user feedback



"Mastering the Information Age Solving Problems with Visual Analytics",edited by Daniel Keim, Jörn Kohlhammer,Geo rey Ellis and Florian Mansmann

opportunity

new paradigm for user interaction
in visual analytics

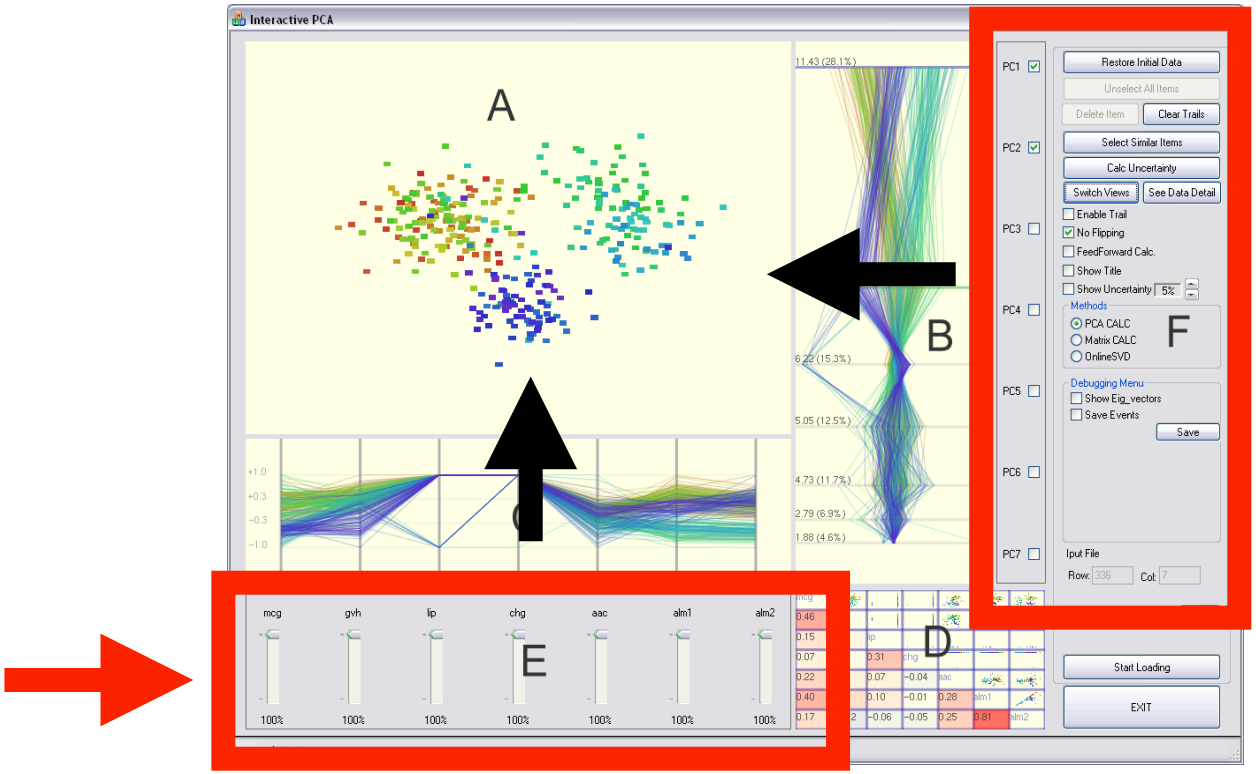
approach:

how do we enable people to **act on graphical encodings** directly?

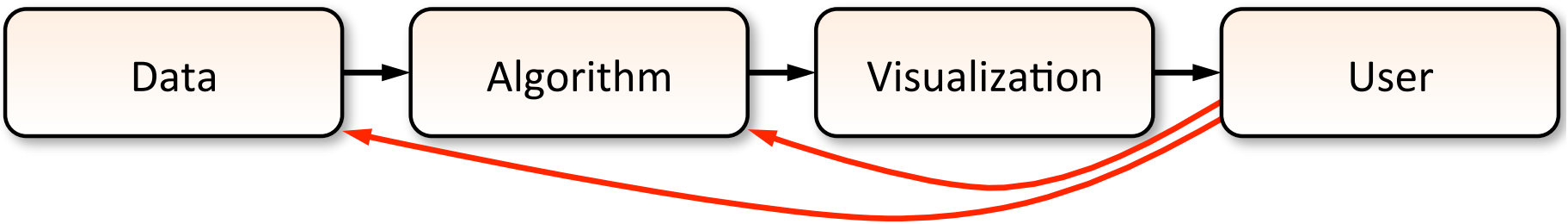
how do we enable users to **act on graphical encodings** directly?

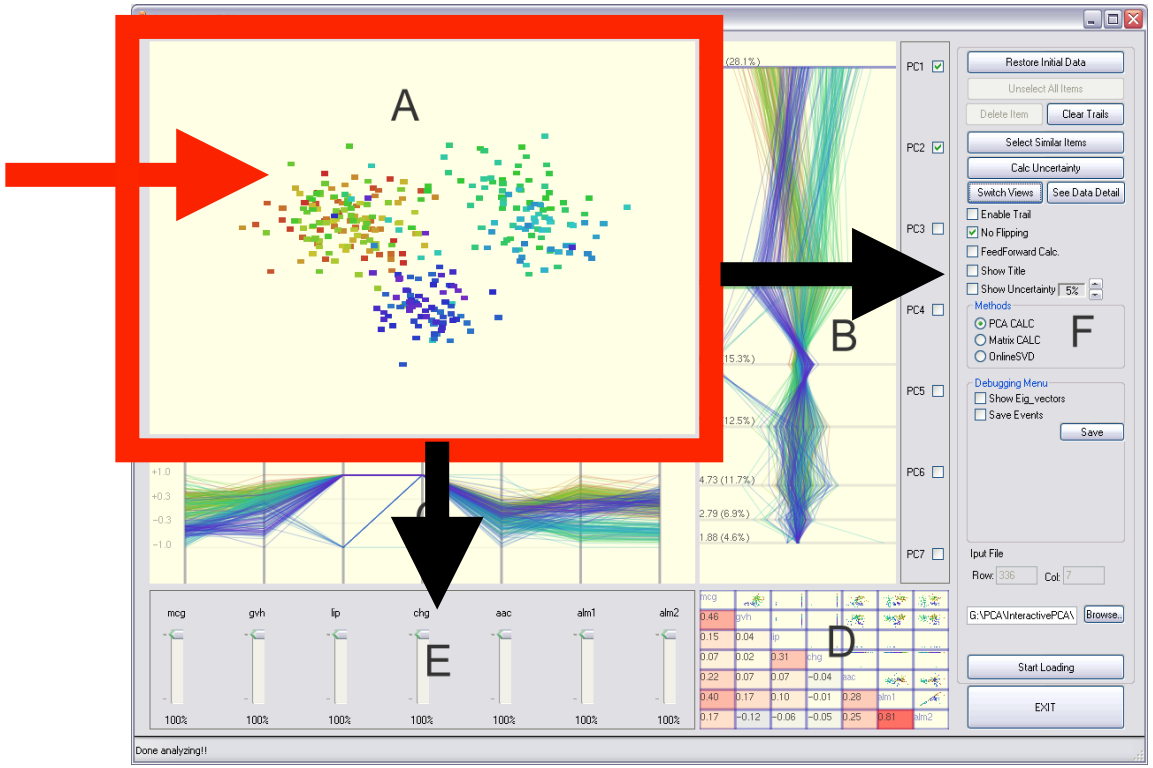
interactive visual affordances

the ability for humans to **plan and act** on hypotheses and questions about data through directly manipulating **interactive visual primitives**

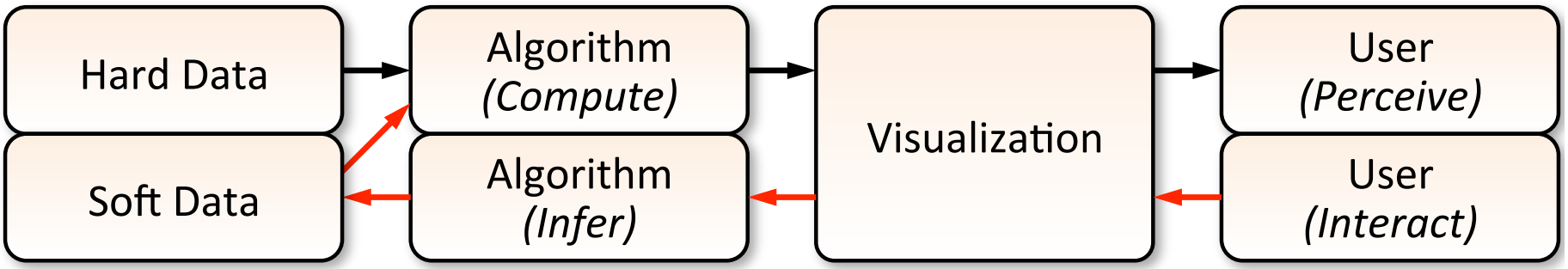


Affordances designed explicitly as **graphical controls**



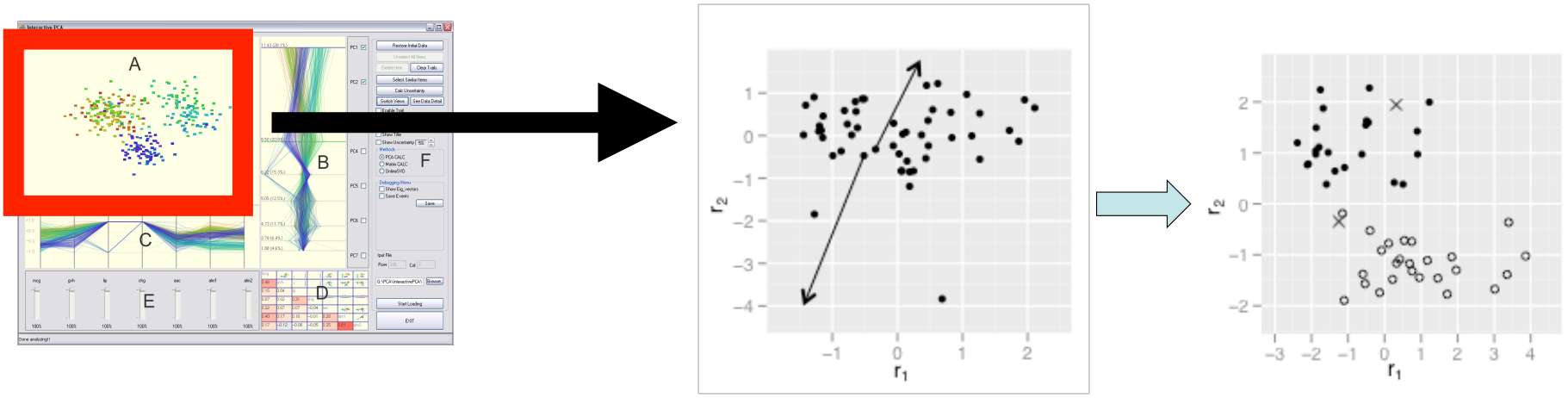


Affordances placed **into the visual metaphor**



interactive visual affordance
examples

Observation-Level Interaction:
Inverting Dimension Reduction Models



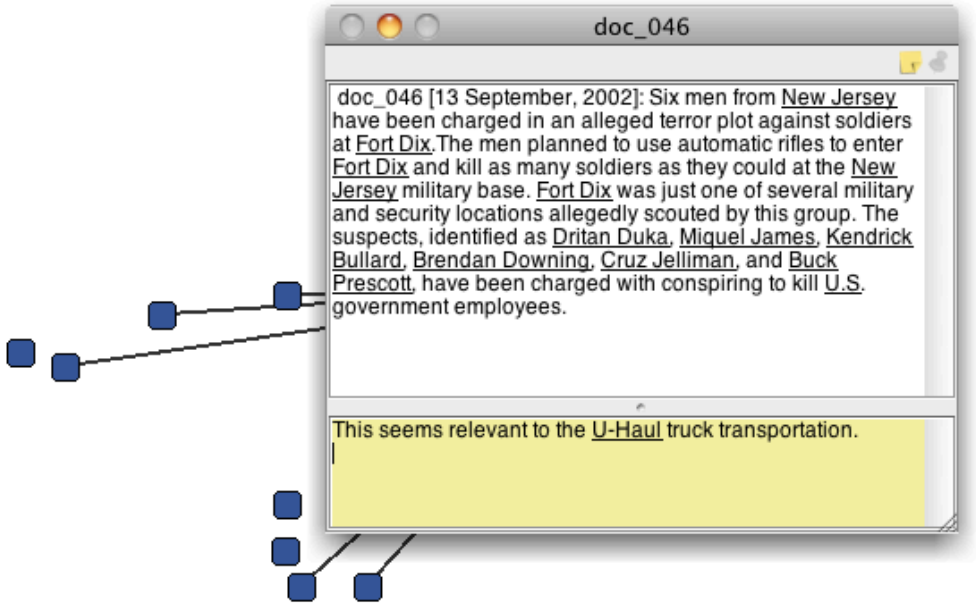
Convey	Interaction	Solve for
Similarity	Move two data points closer	Parallel residual eigenvector
Dissimilarity	Move two data points apart	Orthogonal residual eigenvector

Endert, A., Han, C., Maiti, D., House, L., Leman, S., & North, C. Observation-level interaction with statistical models for visual analytics. IEEE Visual Analytics Science and Technology (VAST), 2011.

ForceSPIRE: visual text analysis

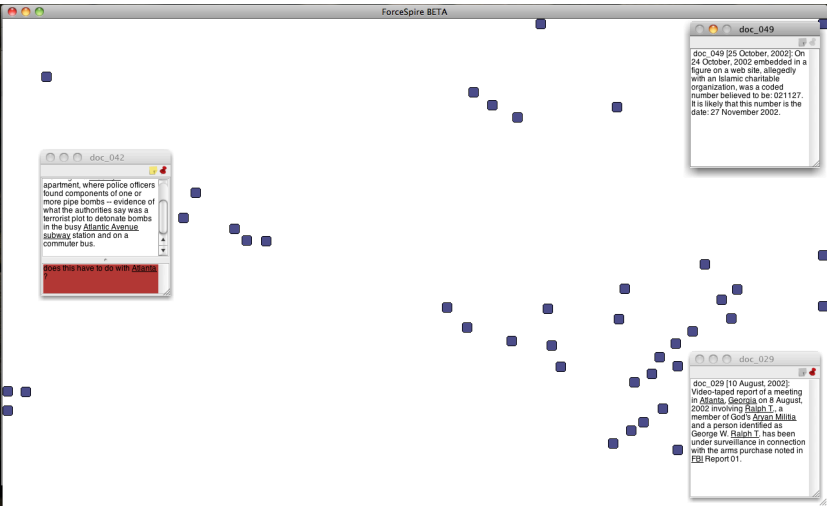
- uses *semantic interaction*
  - every interaction carries some aspect of user intent
  - system learn from these user interactions
  - steering underlying analytic models (layout, entity extraction, entity weighting)

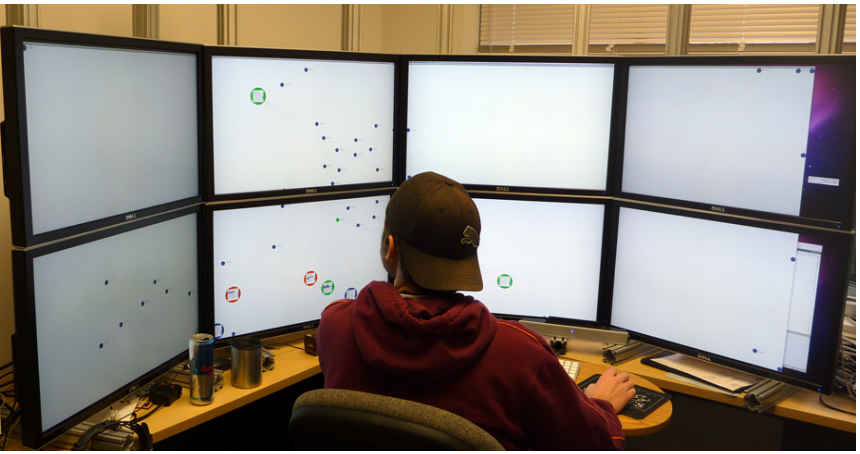
- [video]

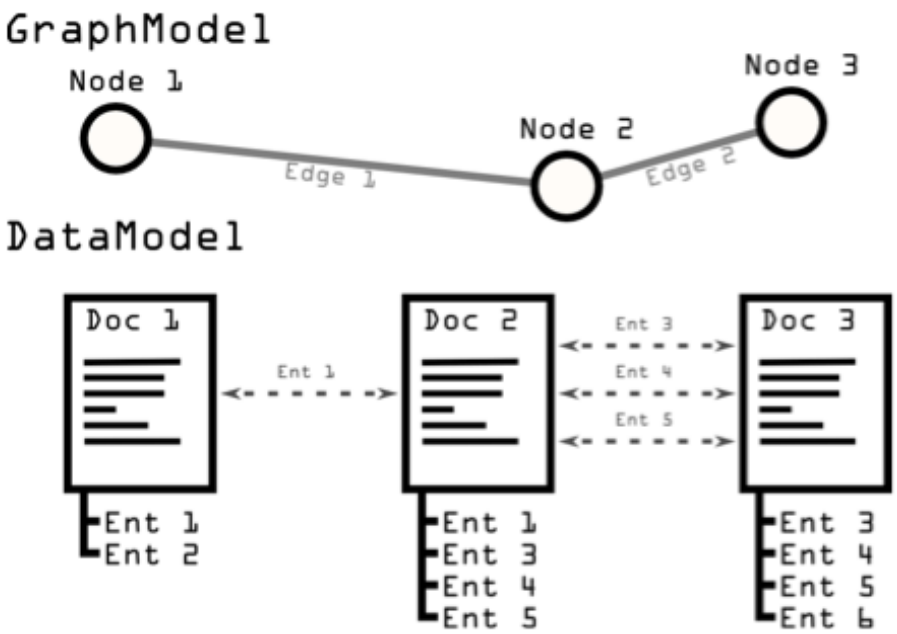


Endert, A., Fiaux, P., North, C., "Semantic Interaction for Visual Text Analytics", ACM Human Factors in Computing Systems (CHI), 2012.

ForceSPIRE


ForceSpire BETA
doc_049
doc_049 [25 October, 2002]: On 24 October, 2002 embedded in a figure on a web site, allegedly with an Islamic charitable organization, was a coded number believed to be: 021127. It is likely that this number is the date: 27 November 2002.
doc_042
apartment, where police officers found components of one or more pipe bombs -- evidence of what the authorities say was a terrorist plot to detonate bombs in the busy Atlantic Avenue subway station and on a commuter bus.
does this have to do with Atlanta ?
doc_029
doc_029 [10 August, 2002]: Video-taped report of a meeting in Atlanta, Georgia on 8 August, 2002 involving Ralph T., a member of God's Aryan Militia and a person identified as George W. Ralph T. has been under surveillance in connection with the arms purchase noted in FBI Report 01.




GraphModel
Node 1
Node 2
Node 3
Edge 1
Edge 2
DataModel
Doc 1
Doc 2
Doc 3
Ent 1
Ent 3
Ent 4
Ent 5
Ent 1
Ent 2
Ent 1
Ent 3
Ent 4
Ent 5
Ent 3
Ent 4
Ent 5
Ent 6

ForceSPIRE Semantic Interactions

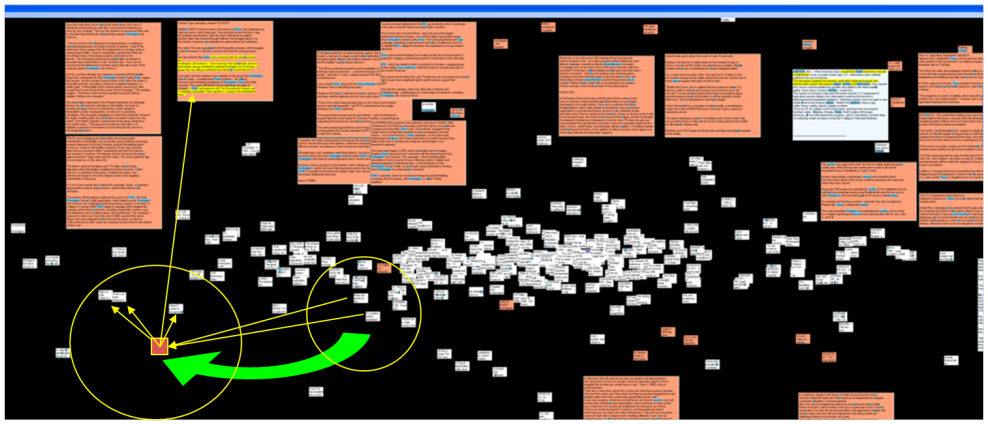
Semantic Interaction	Model Learning
Expressive Movement	Dimension Weighting
Exploratory Movement	Spatial landmarks/constraints
Highlight	Guide entity extraction
Search	Spatial landmarks, dimension weighting
Annotation	Add domain information to the model

Document Movement

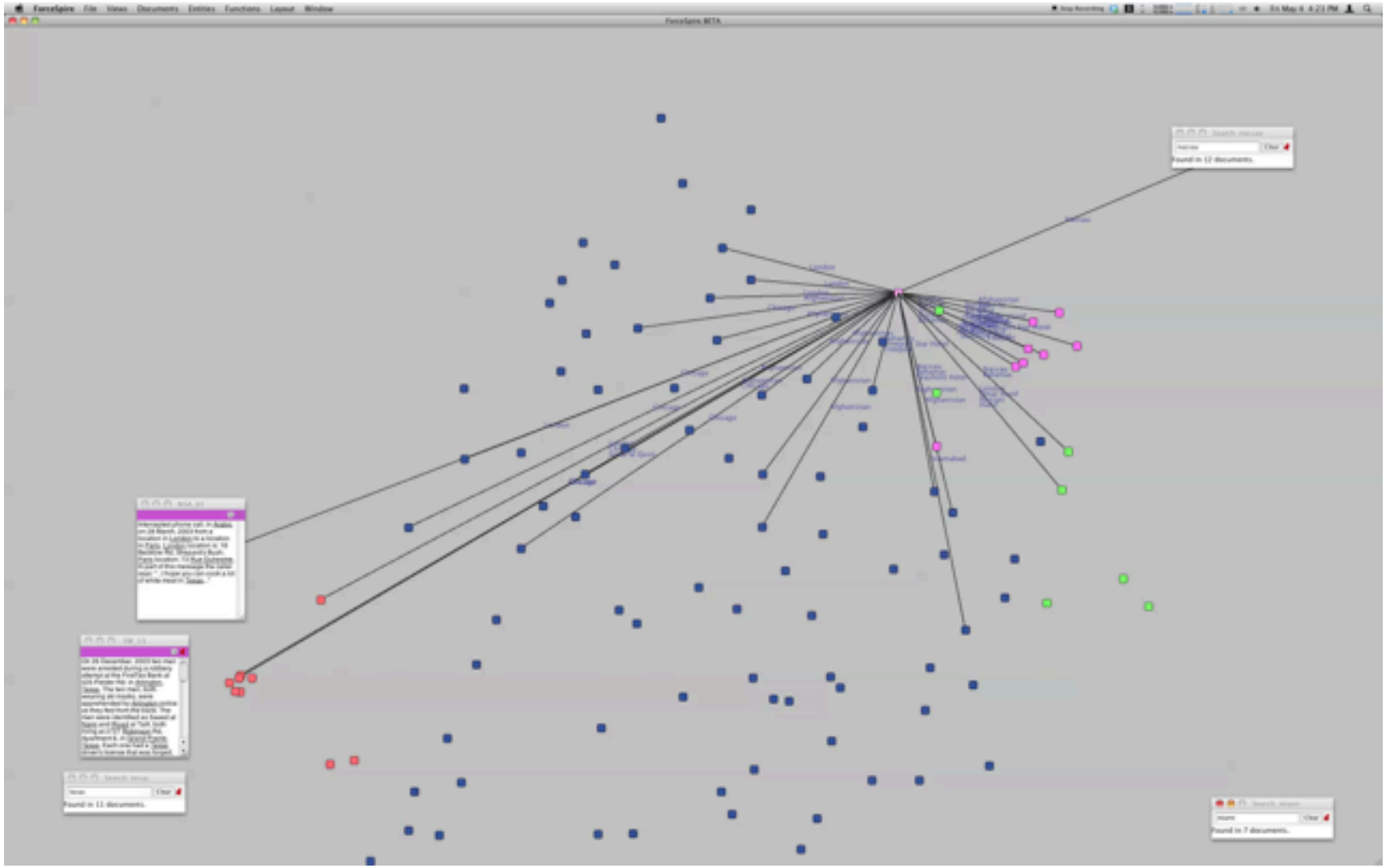
Create Spatial Landmarks
"Exploratory"



Emphasize Features
"Expressive"

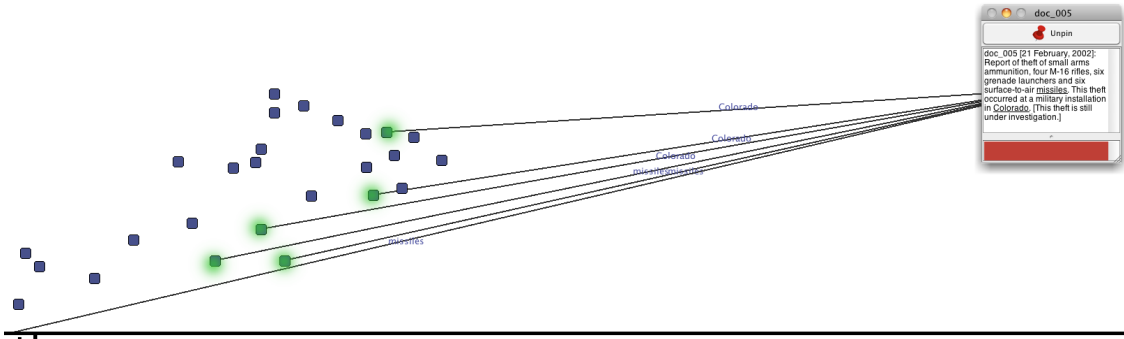


Observation-Level Interaction
MDS, GTM, PPCA

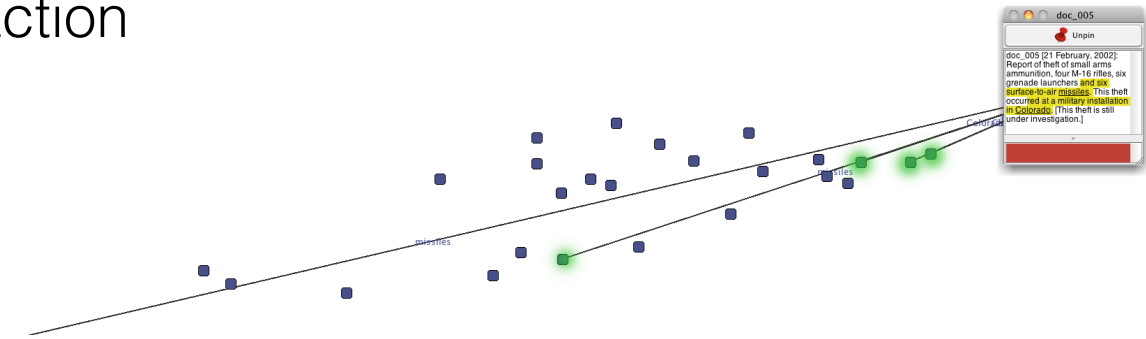

ForceSpire File Views Documents Entities Functions Layout Window
ForceSpire BETA
Found in 12 documents.
Found in 11 documents.
Found in 7 documents.

Highlighting

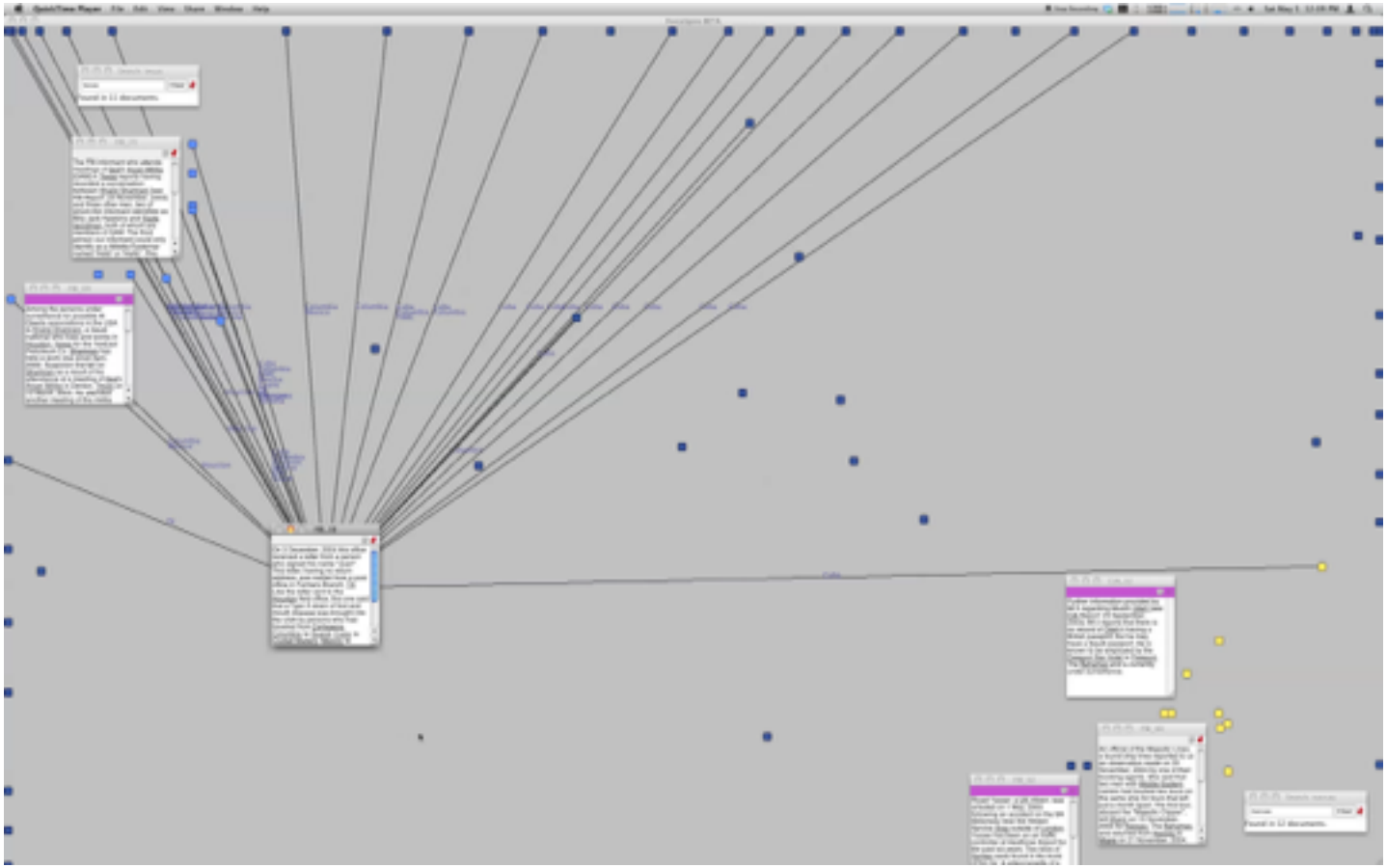
Emphasize phrase
more precise guidance



More aggressive entity extraction



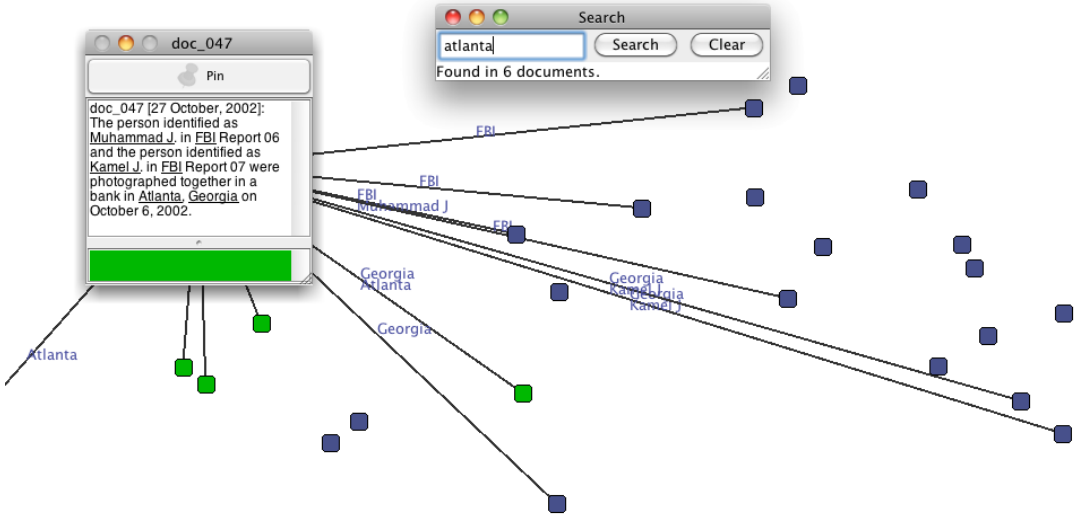
Highlighting



Searching

Create spatial landmark
define axes

Add constraint to model



Evaluating Semantic Interaction

Ability for Semantic Interaction to couple spatialization, mathematical model, and user insights through inferring reasoning embedded in user interactions.

express features of interest for model steering

enable analysis within spatialization

enable spatialization co-creation

What did they find?

Top Keywords from User Findings and Model Weighting

Bold

entity extracted from semantic interaction

Underlined

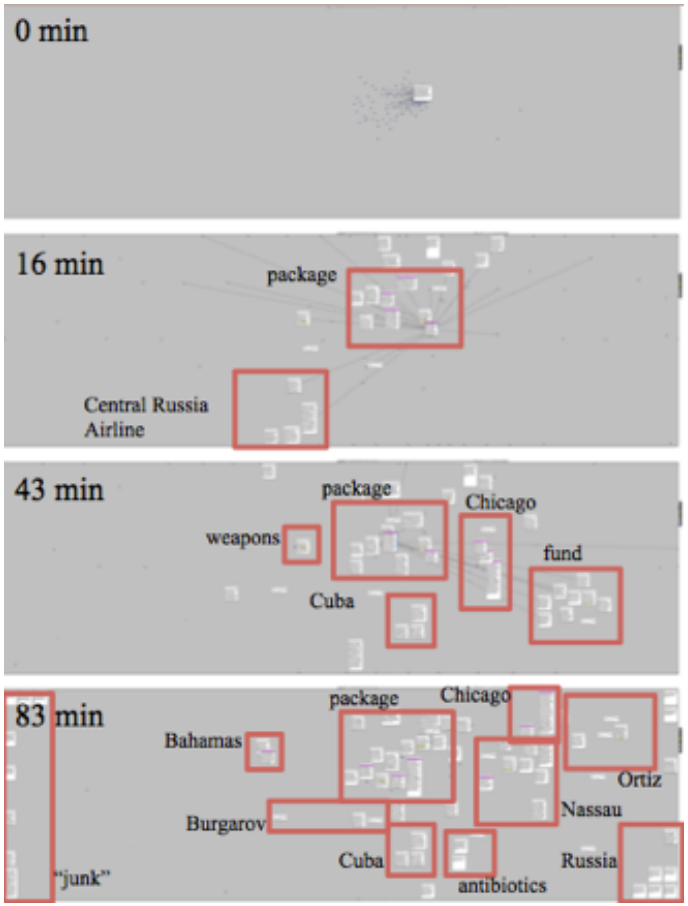
match between user and model

1 (*user*)	**1** (*model*)	**2** (*user*)	**2** (*model*)	**3** (*user*)	**3** (*model*)	**4** (*user*)	**4** (*model*)	**5** (*user*)	**5** (*model*)	**6** (*user*)	**6** (*model*)
diamonds	**diamonds**	package	Nassau	explosives	Nassau	diamonds	package	Al Queda	Nassau	diamonds	**weapons**
scholarship	**weapons**	Hanif	Hanif	weapons	**students**	Nassau	Chicago	Caribbean	Miami	antibiotics	**diamonds**
jihad	**graduate**	antibiotics	Freeport	Nassau	**weapons**	Burgarov	Russia	Russia	Freeport	Ortiz	Nassau
weapons	Jamal	diamonds	Miami	students	**scholarship**	antibiotics	Nassau	Hanif	Apple St.	Bahamas	**graduate**
freeport	Nassau	Nassau	**package**	scholarship	Jamal	package	Burgarov	Odeh	**weapons**	Hijazi	Hijazi

Spatialization co-creation

User defined regions throughout study

Spatialization placement per user-defined and model-defined document locations



Spatialization co-creation

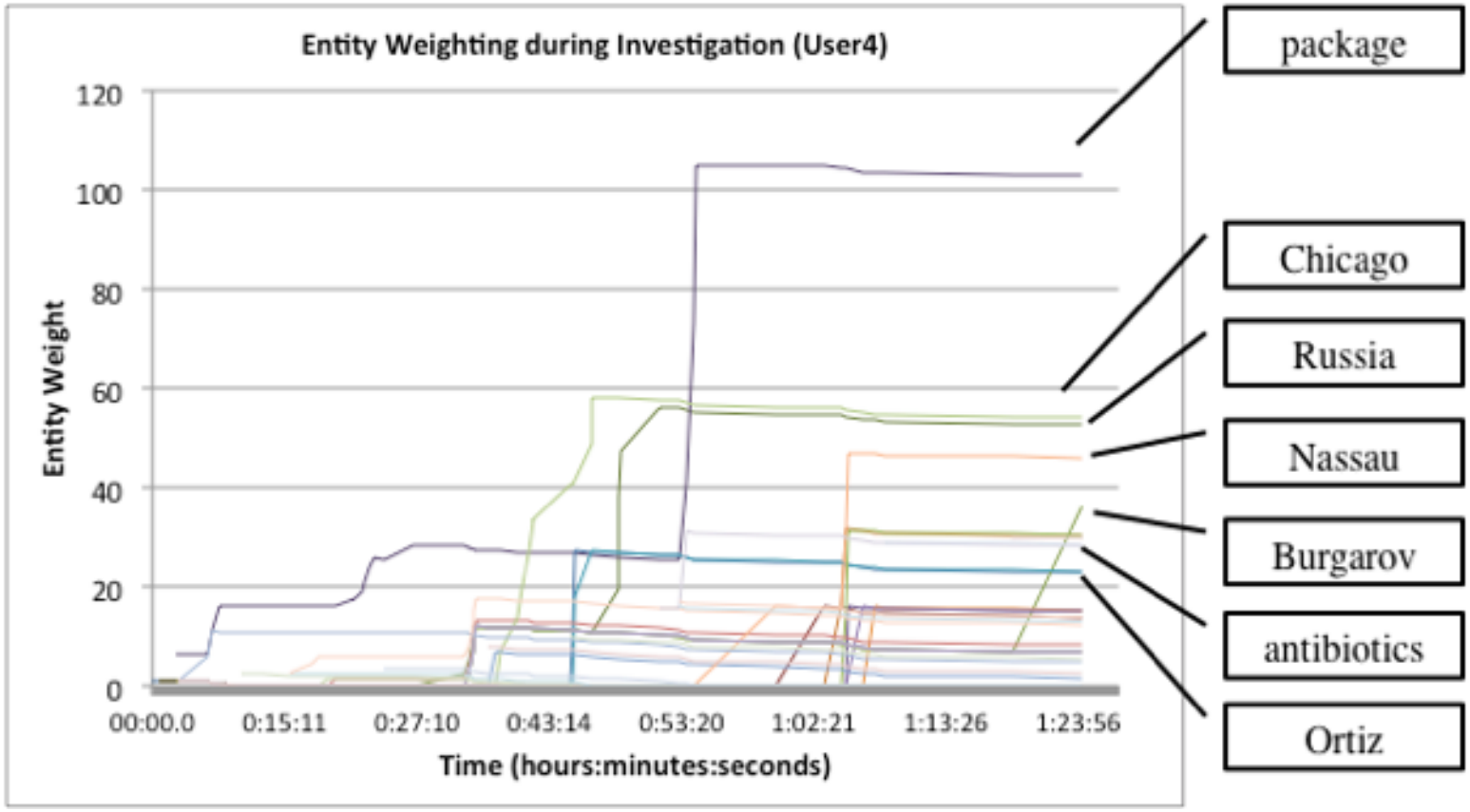
	User-Defined	Model-Defined
All users	32%	68%
User4	40%	60%



Spatialization co-creation



4 (*user*)	4 (*model*)
diamonds	package
Nassau	Chicago
Burgarov	Russia
antibiotics	Nassau
package	Burgarov

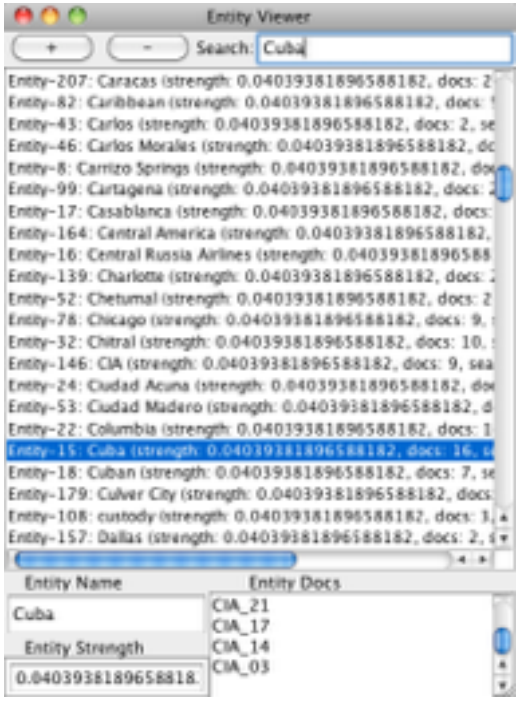


Participant Feedback

“[expressive movements] tell the system something”

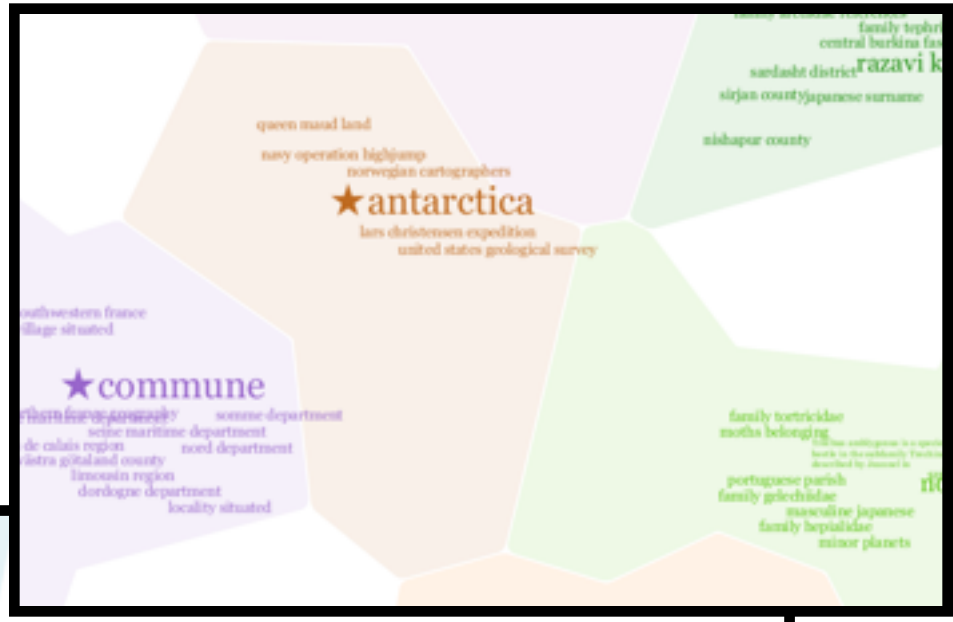
“Ooops, oh well. [ForceSPIRE] will fix that later”

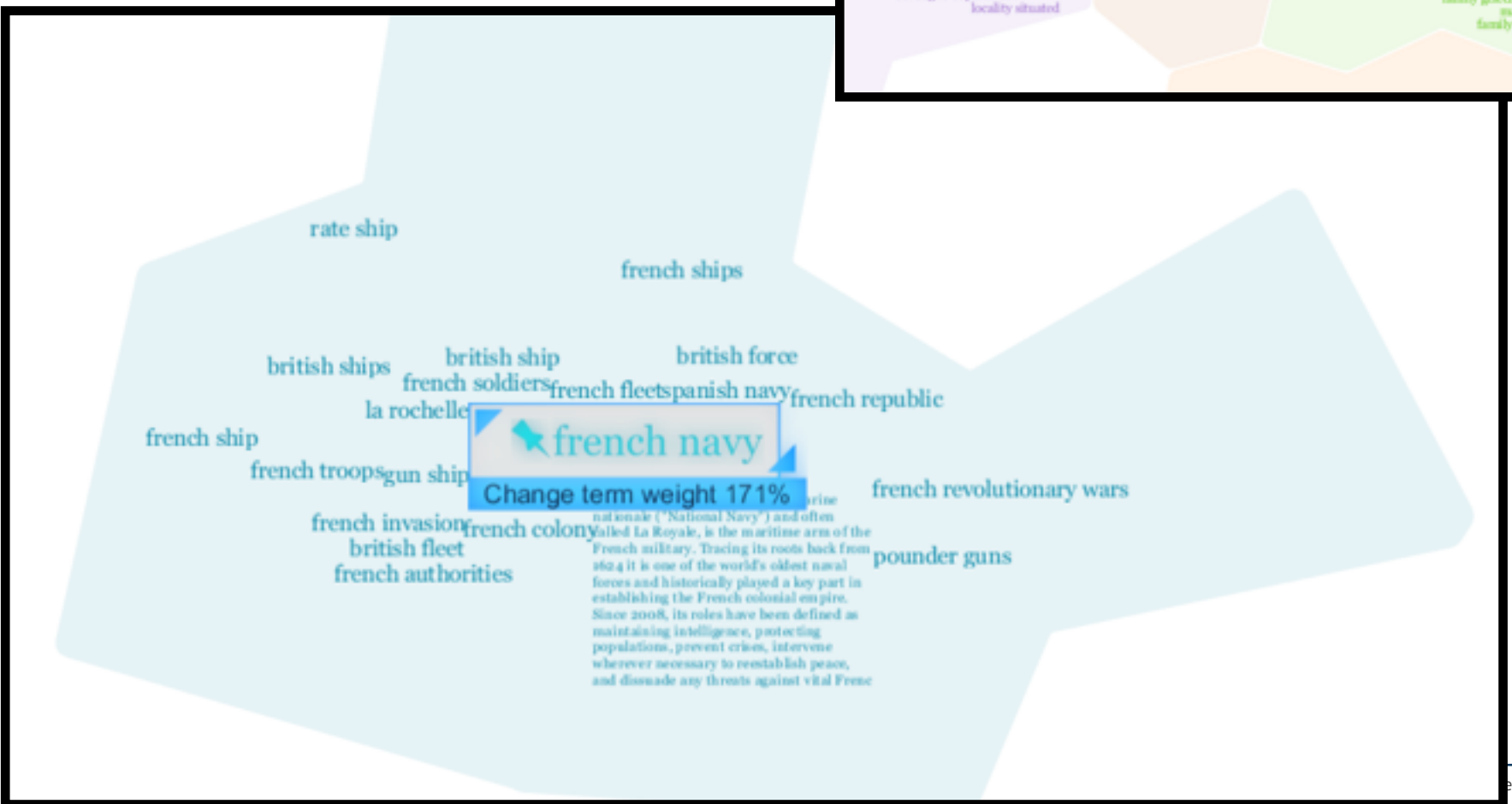
“Look! [ForceSPIRE]’s moving other junk into my junk pile for me!”



Direct Model Control?

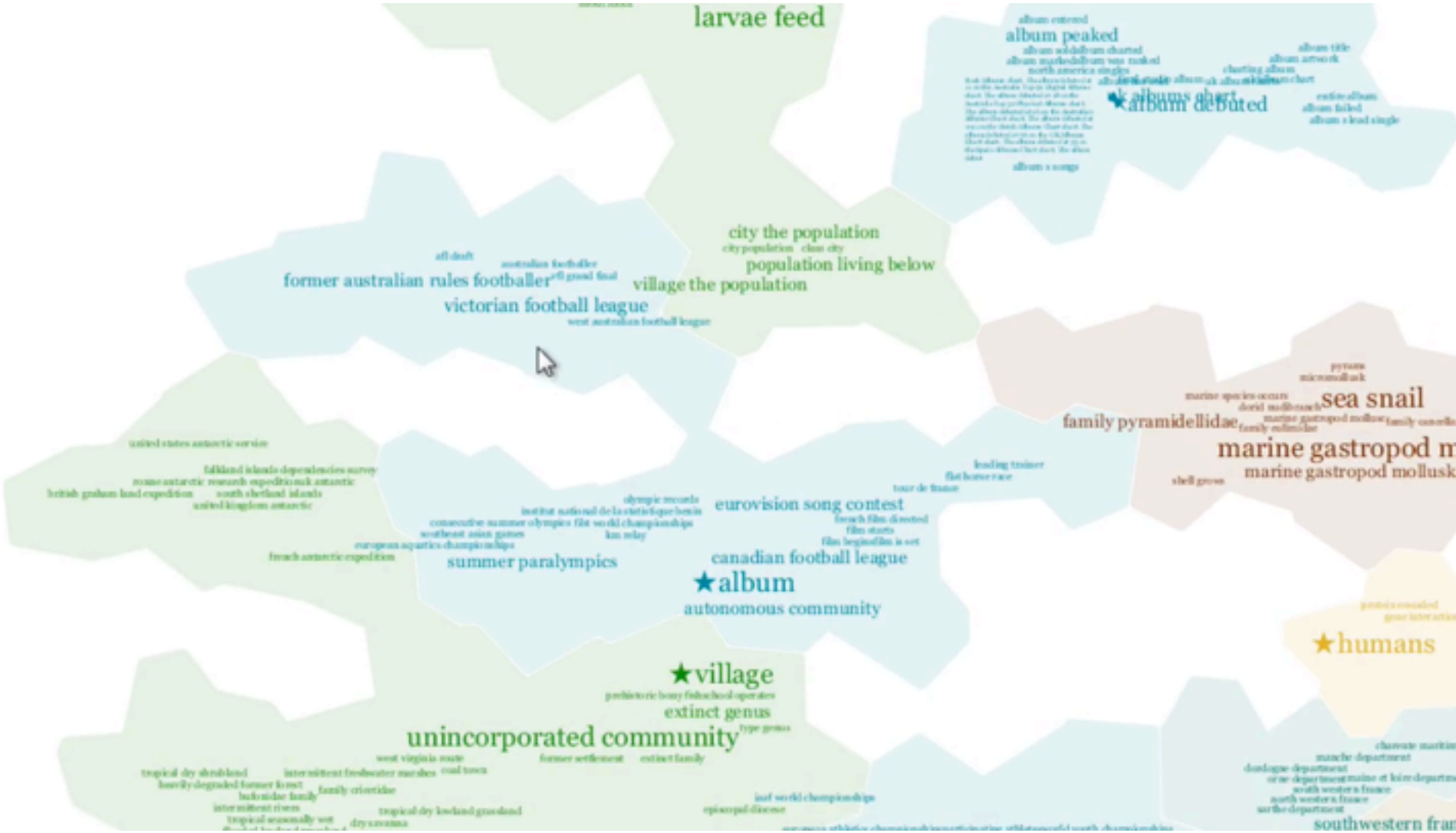
TexTonic



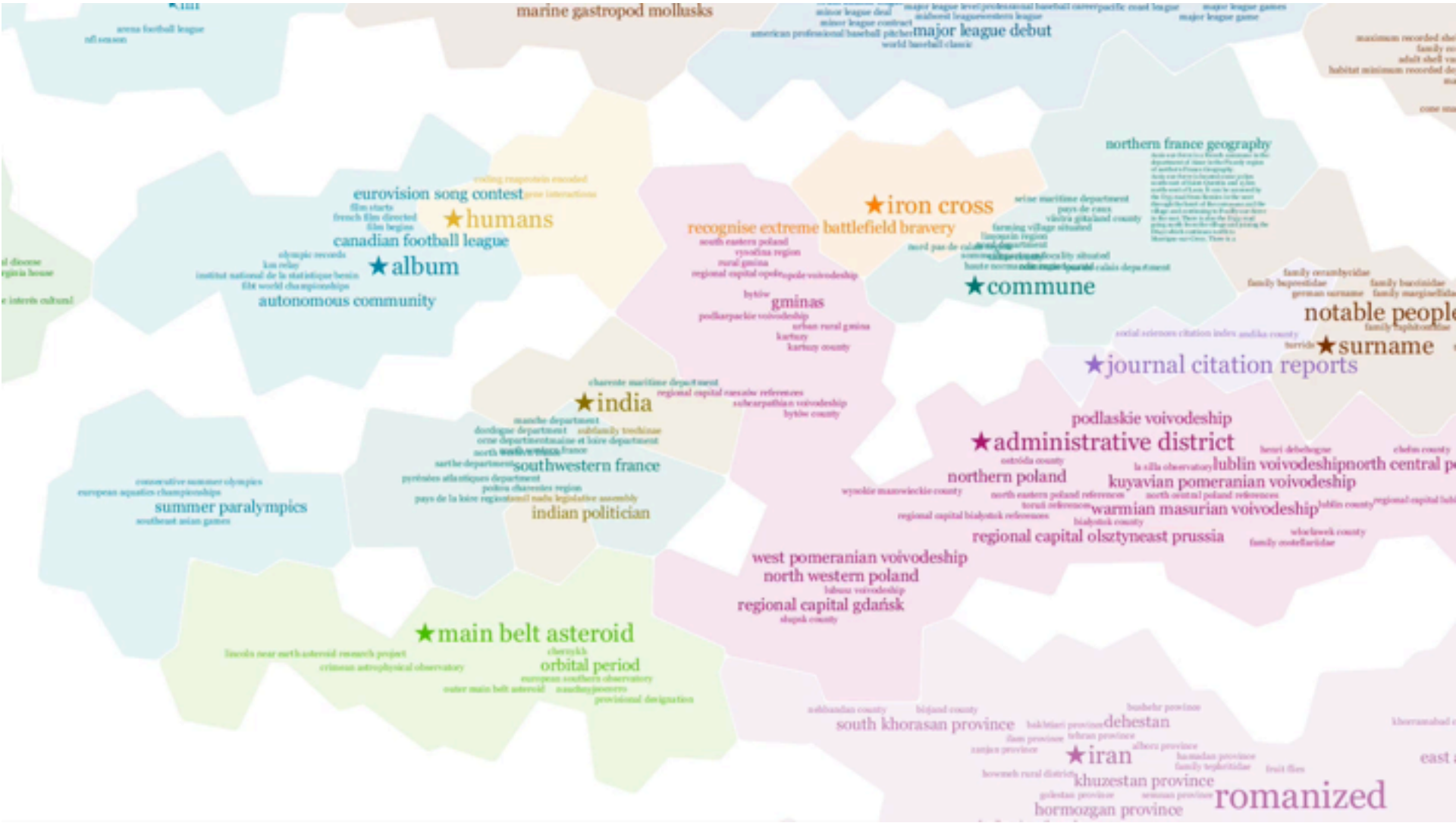


ert@gatech.edu

TexTonic

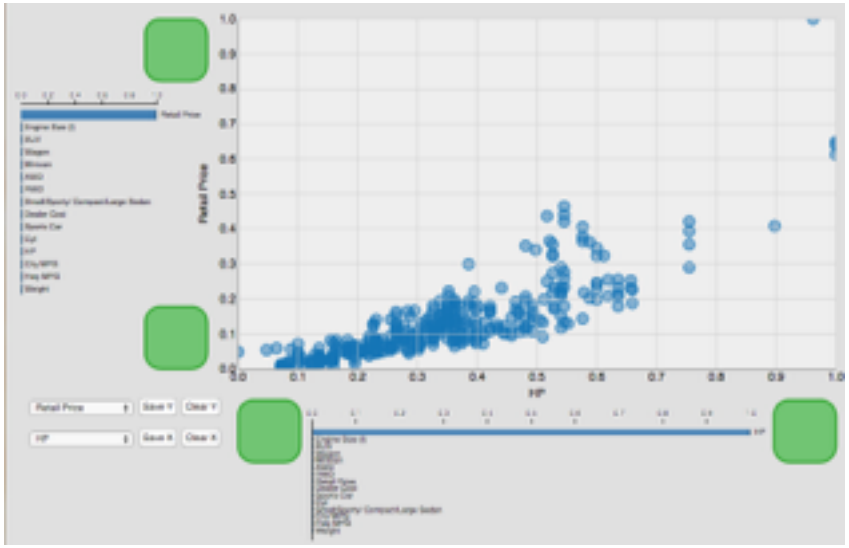

larvae feed
album peaked
album debuted
city the population
population living below
village the population
former australian rules footballer
victorian football league
sea snail
family pyramidellidae
marine gastropod m
marine gastropod mollusk
eurovision song contest
canadian football league
summer paralympics
★album
autonomous community
★humans
★village
extinct genus
unincorporated community
southwestern fran

TexTonic


marine gastropod mollusks
major league debut
northern france geography
eurovision song contest
humans
canadian football league
album
autonomous community
iron cross
recognise extreme battlefield bravery
commune
gminas
notable people
surname
journal citation reports
india
podlaskie voivodeship
administrative district
southwestern france
northern poland
kuyavian pomeranian voivodeship
summer paralympics
indian politician
warmian masurian voivodeship
regional capital olsztyneast prussia
west pomeranian voivodeship
north western poland
regional capital gdańsk
main belt asteroid
orbital period
south khorasan province
dehestan
iran
khuzestan province
romanized
hormozgan province

InterAxis

- Enable users to directly specify and manipulate **scatterplot axes**
- Fosters **understanding data attributes** that quantify subjective preferences
- http://www.cc.gatech.edu/~hkim708/InterAxis/cars.html

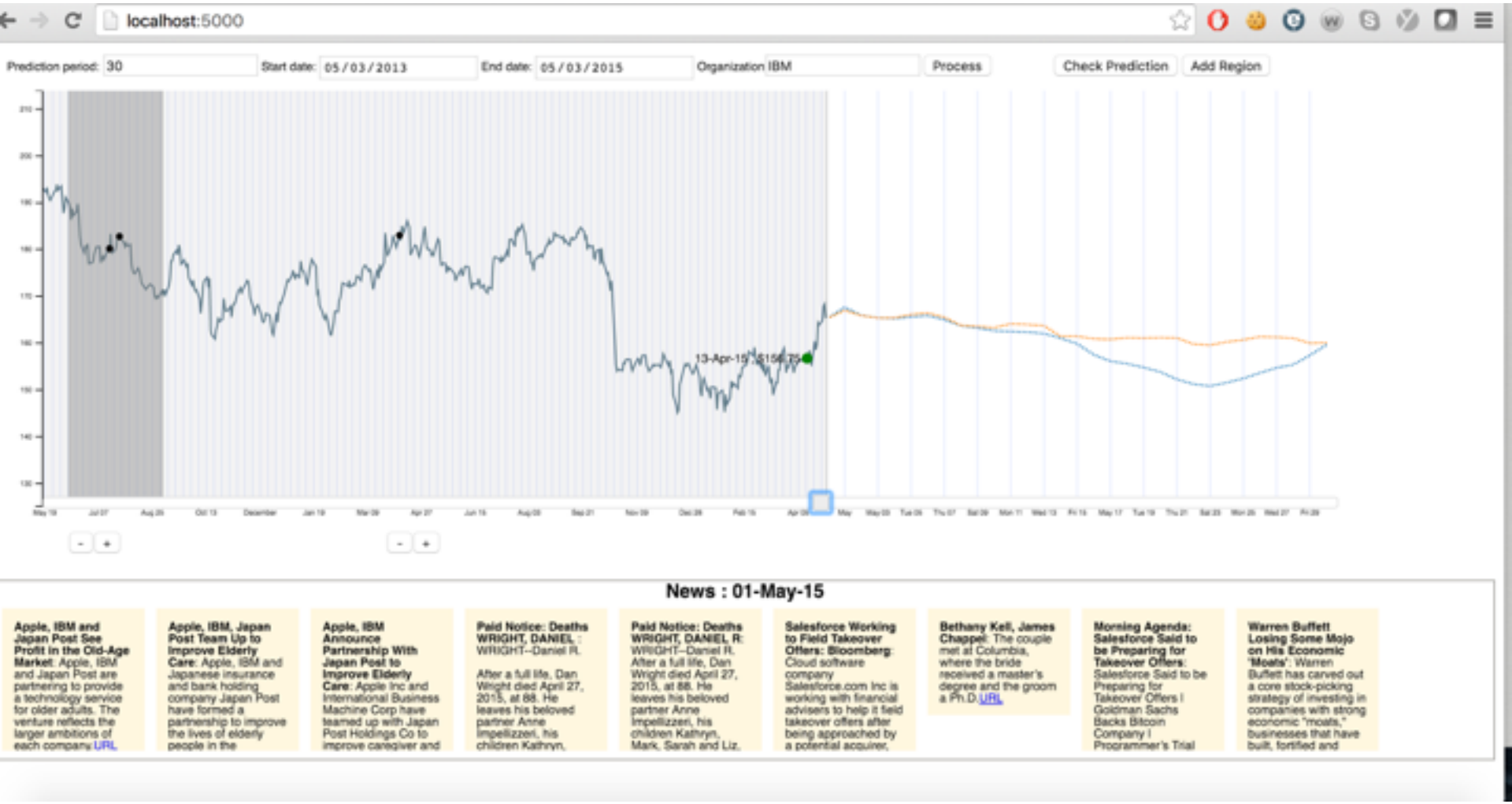


Kim, H., Choo, J., Park, H., Endert, A., InterAxis: Steering Scatterplot Axes via Observation-Level Interaction. IEEE VAST, 2015.

InterAxis

http://va.gatech.edu/projects/interaxis

User-Driven Predictive Analytics

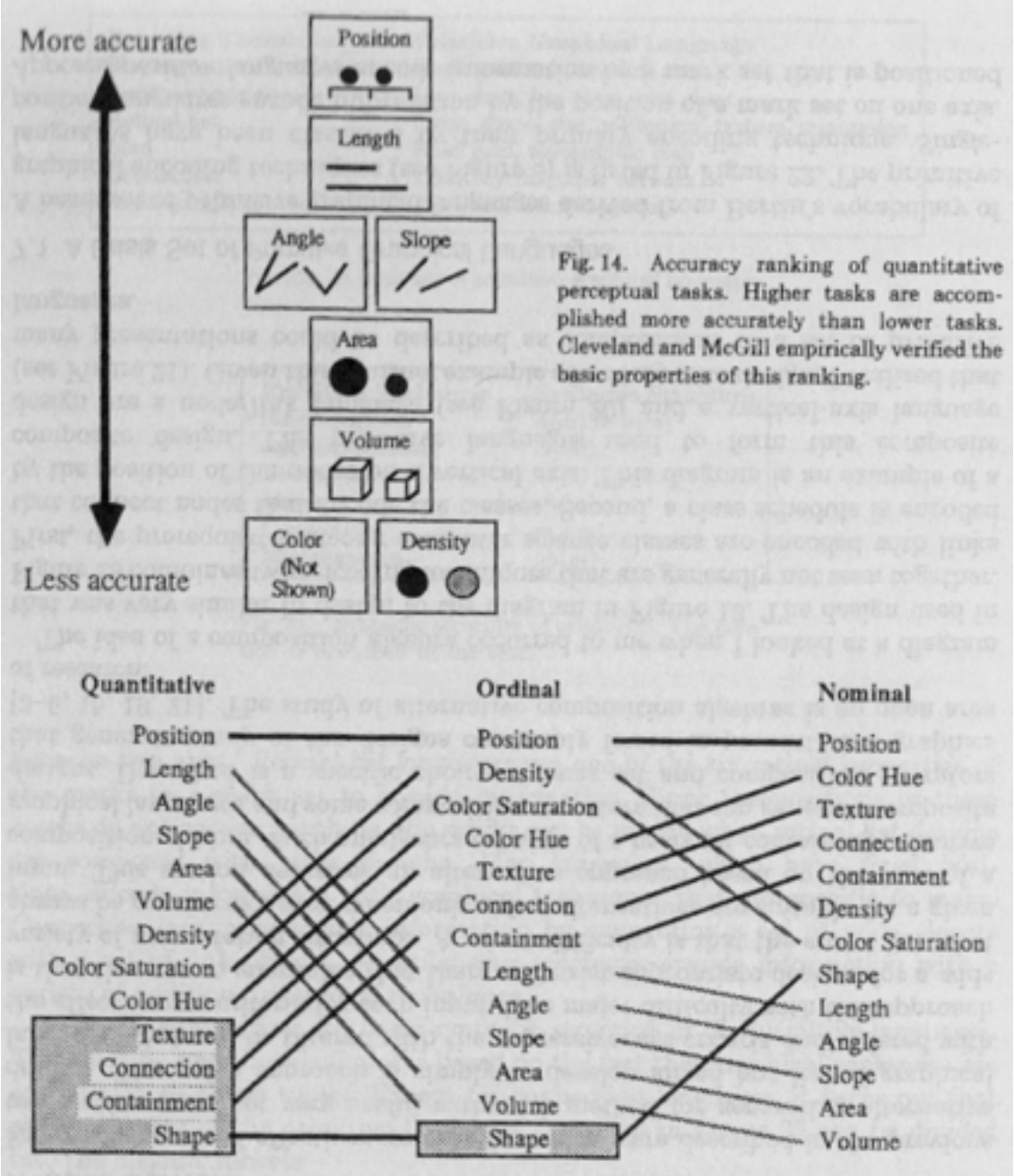


looking ahead

evaluating
interactive visual primitives

graphical encodings

Accuracy ranking of quantitative perceptual tasks

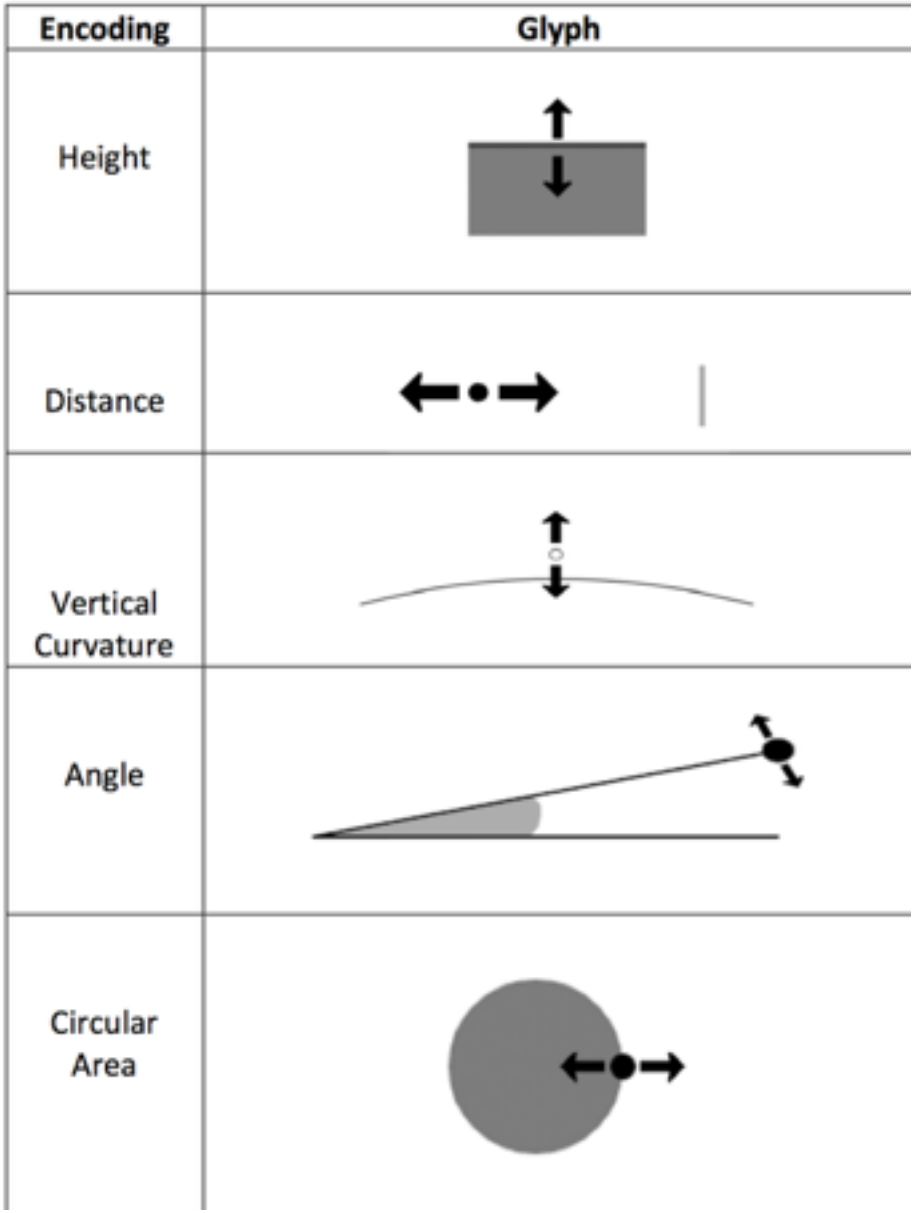


From Jock Mackinlay's 1986 thesis

how effective are fundamental graphical encodings for **interaction**?

what can we learn from the user interaction to inform us about the **connection between perception and action?**

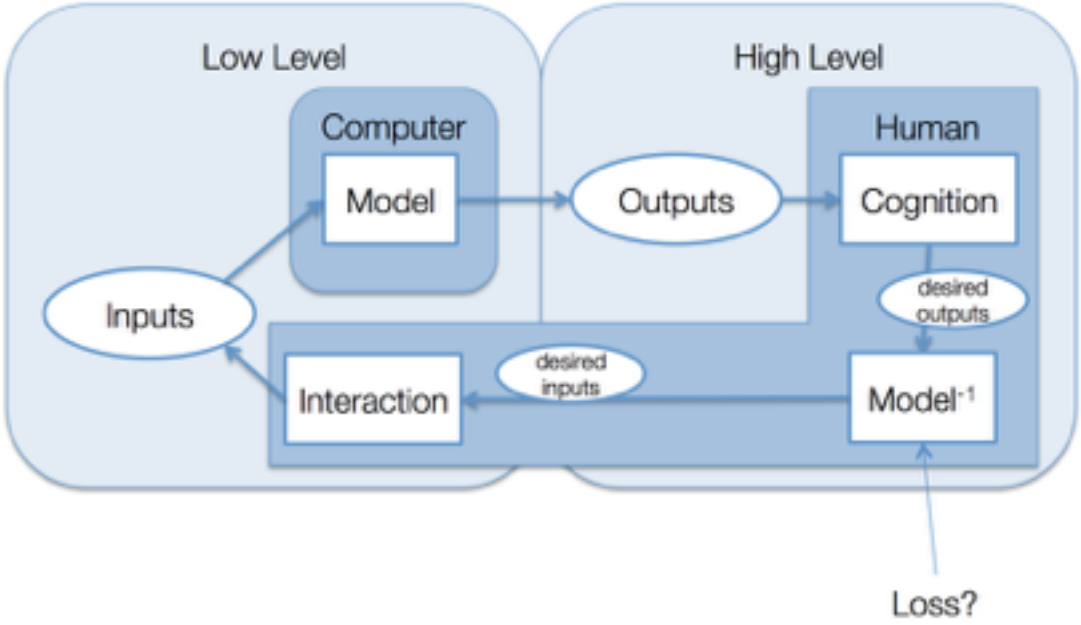
stay tuned. results are coming.

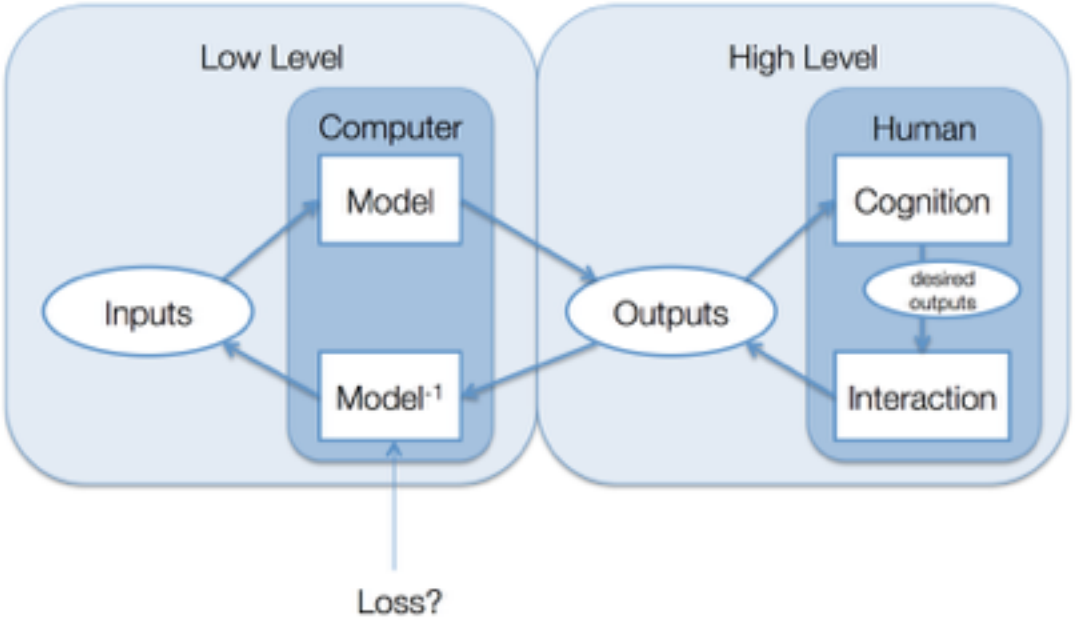


Human-*is*-the-loop?

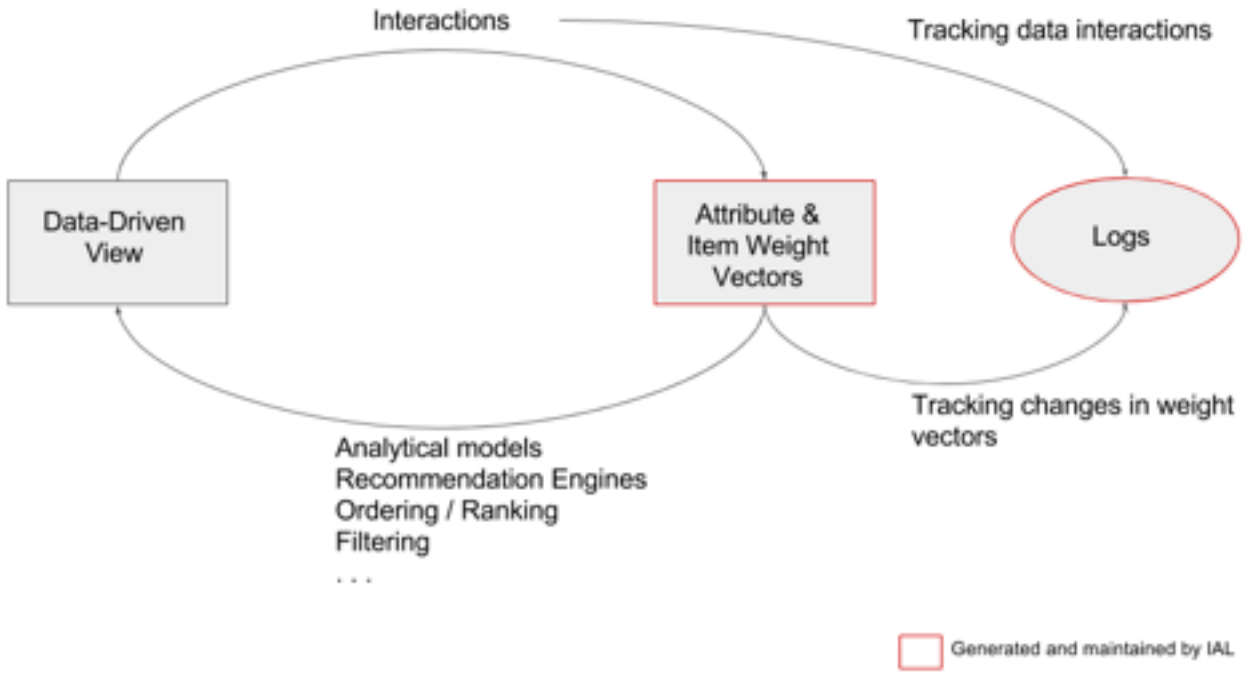
- Should people understand analytics,

or analytics understand humans?





Interactive Analytics Library



- Incorporate interaction into visualizations created with **d3**
- Perform **passive model steering** based on the user interaction
- enable development of visualizations that are **inherently interactive**

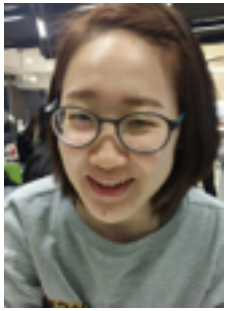
http://gtvalab.github.io/ial/

warning! \begin{alex_rant}

- do not treat user interaction as an easy way out
  - “oh, we’ll just give the user a menu for that!”
- do not make your problems (as developers and designers) the user’s problems
- interaction should be meant to explore data, ask questions, gain insight



Bahador Saket
PhD, CS



Hannah Kim
Phd, CSE



Emily Wall
Phd, CS



Fred Hohman
Phd, CSE



Subhajit Das
Phd, CS



Arjun Srinivasan
Phd, CS

Georgia Tech School of Interactive Computing

GVU Georgia Tech

VACCINE

VA.lab
va.gatech.edu

KEYSIGHT TECHNOLOGIES

Microsoft

Pacific Northwest NATIONAL LABORATORY



April 12, 2016
Bentley Univ.



The Value of Visualization for Exploring and Understanding Data

John Stasko

School of Interactive Computing
Georgia Institute of Technology

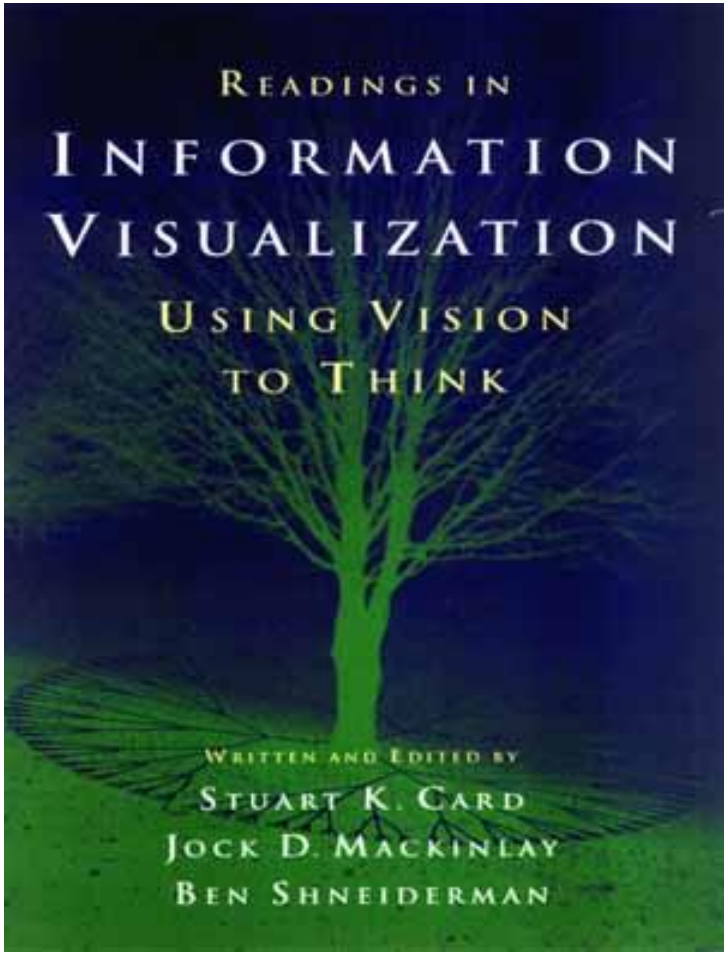
stasko@cc.gatech.edu








Data



Visualization

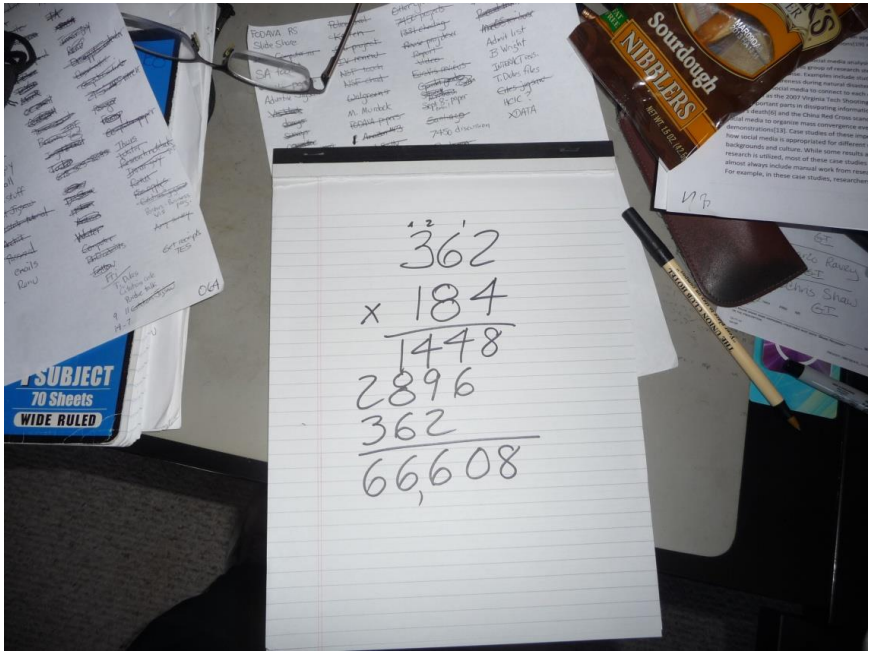
"The use of computer-supported, interactive visual representations of data to amplify cognition"

Visuals help us think

Provide a frame of reference, temporary storage area

Cognition $\rightarrow$ Perception

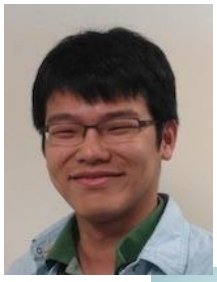
Pattern matching



Three keys to success with visualization

1. Understand purpose
2. Embrace interaction
3. Identify value

Talented students





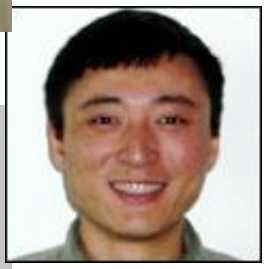


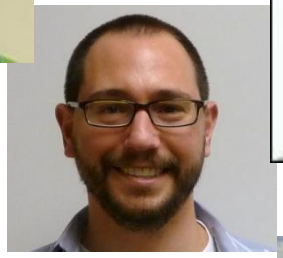






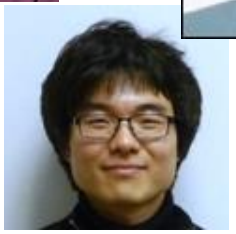
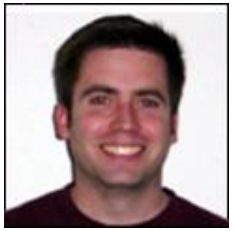











Why visualize?

Applications of Visualization

Analysis (Exploratory)

Presentation (Explanatory)



1. 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





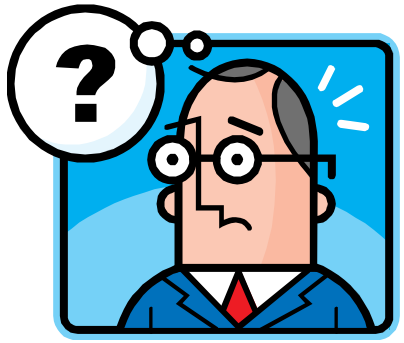
Thought

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



Will return to this idea later…

Frequent analysis goals

Show many/all variables

Illustrate overview and detail

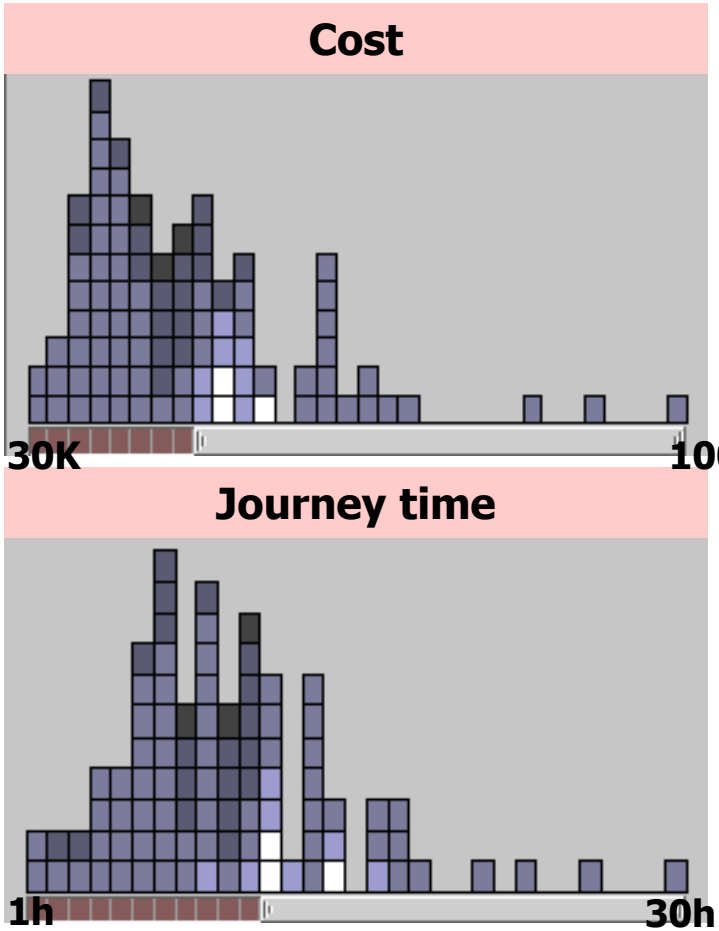
Facilitate comparison

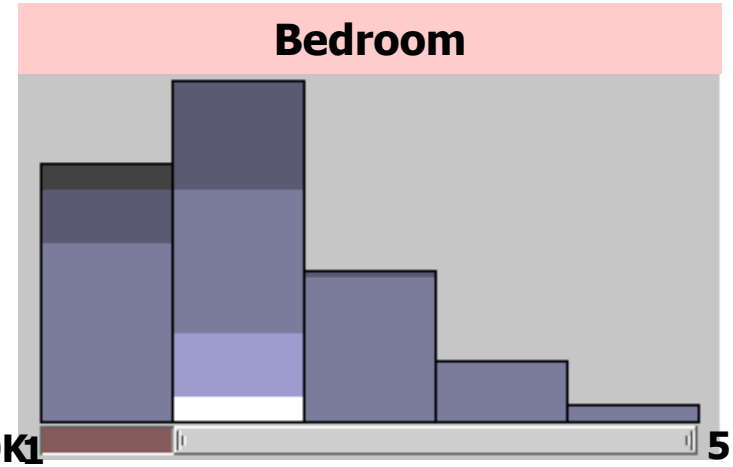
Display may not be easy to interpret at first

At Its Heart: Uncertainty

Nothing is perfect

Humans trade off choices, compromise

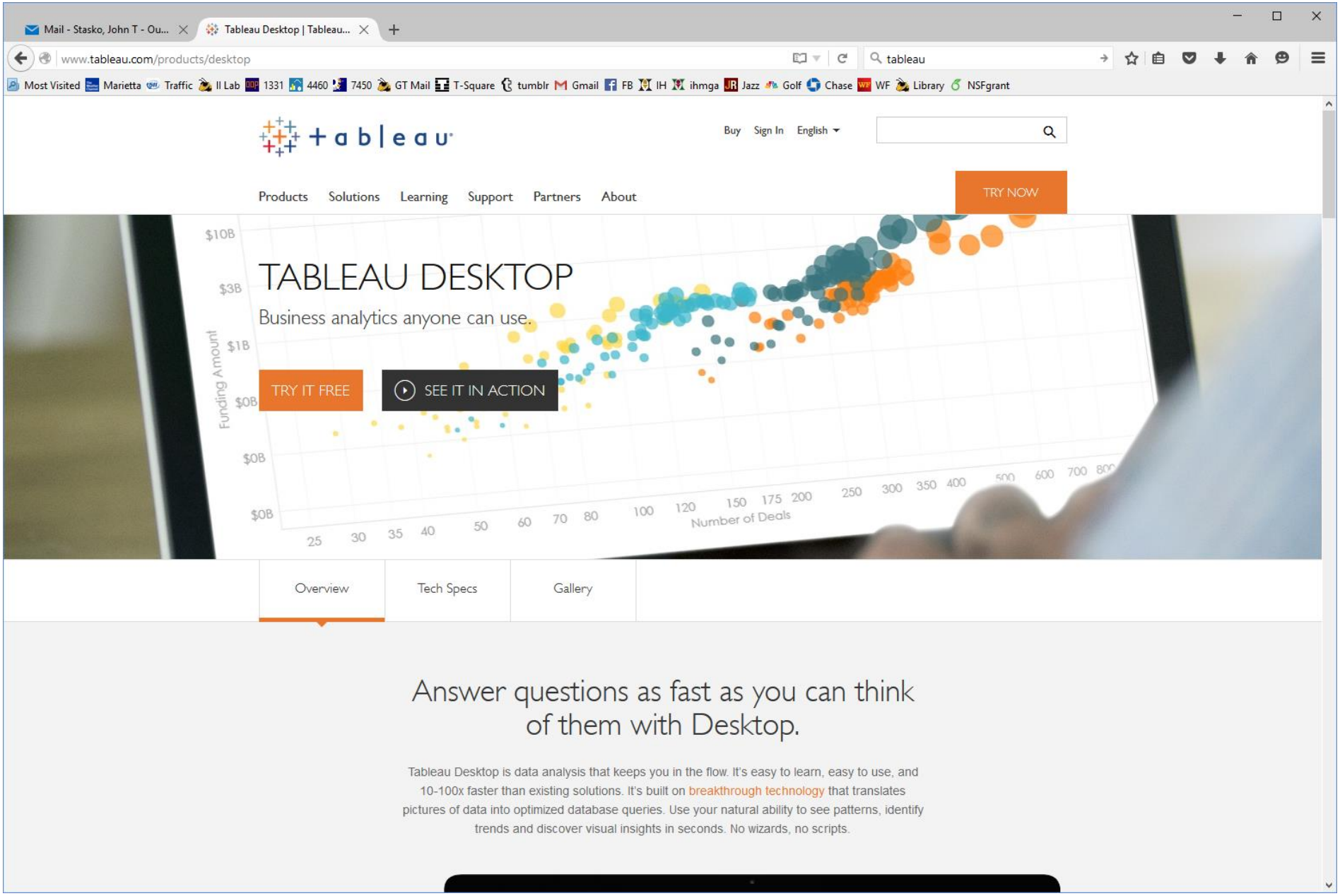




Attribute Explorer

Spence & Tweedie

Interacting with Computers '98


Mail - Stasko, John T - Ou...
Tableau Desktop | Tableau...
www.tableau.com/products/desktop
tableau
Most Visited Marietta Traffic Il Lab 1331 4460 7450 GT Mail T-Square tumblr Gmail FB IH ihmga Jazz Golf Chase WF Library NSFgrant
+ a b l e a u
Buy Sign In English
Products Solutions Learning Support Partners About
TRY NOW
TABLEAU DESKTOP
Business analytics anyone can use.
TRY IT FREE
SEE IT IN ACTION
Funding Amount
$10B
$3B
$1B
$0B
$0B
$0B
25 30 35 40 50 60 70 80 100 120 150 175 200 250 300 350 400 500 600 700
Number of Deals
Overview
Tech Specs
Gallery
Answer questions as fast as you can think of them with Desktop.
Tableau Desktop is data analysis that keeps you in the flow. It's easy to learn, easy to use, and 10-100x faster than existing solutions. It's built on breakthrough technology that translates pictures of data into optimized database queries. Use your natural ability to see patterns, identify trends and discover visual insights in seconds. No wizards, no scripts.

2. Presentation

Communicate data and ideas

Explain and inform

Provide evidence and support

Influence and persuade



Infographics

All the rage…

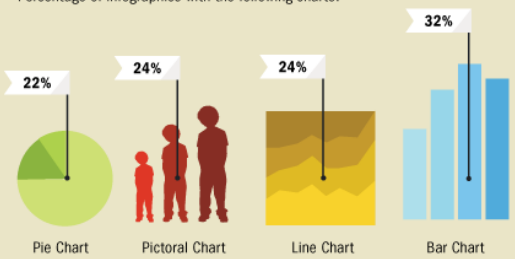
INFOGRAPHIC OF INFOGRAPHICS

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

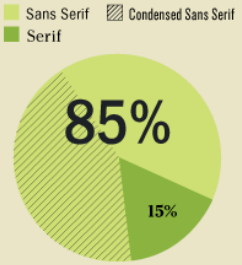
DESIGN

CHART STYLE

Percentage of infographics with the following charts:

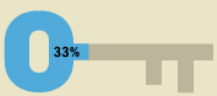


FONT



KEY INFO

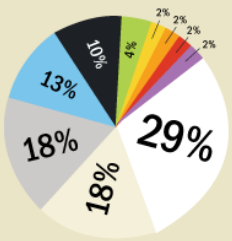
Percentage of infographics with key:



Average number of symbols per key: **5.1**

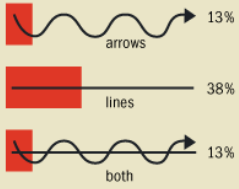


BASE COLOR



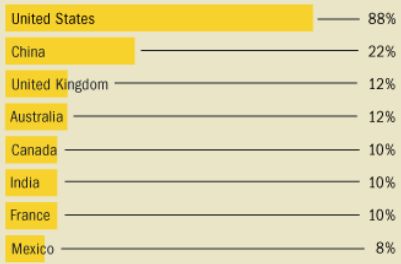
NAVIGATIONAL ICONOGRAPHY

Frequency of arrows & connecting lines in infographics:



CONTENT

COUNTRIES FEATURED



THEME

Relative popularity of different infographic themes:



SECTIONS



CREDITED SOURCES

Average number of sources per infographic: **2.29**



TITLE

Average number of words per infographic title: **4.36**

"RICHEST AND POOREST AMERICAN NEIGH

CONCEPT & DESIGN Ivan Cash **SOURCE** 49 infographics collected at random from www.good.is/infographics

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

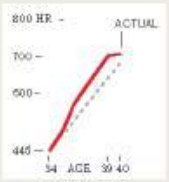
Aaron
Actual homers slightly outpace projected homers for five seasons.



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

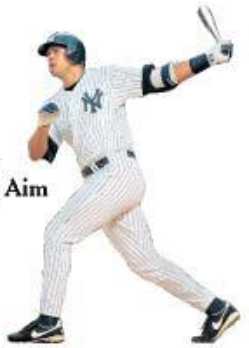


Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

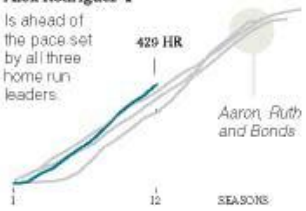


Note: Ages as of July 1 of each season.

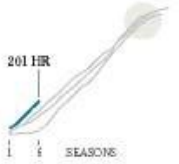
Others Taking Aim



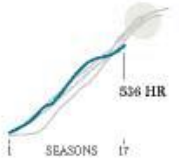
Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.



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

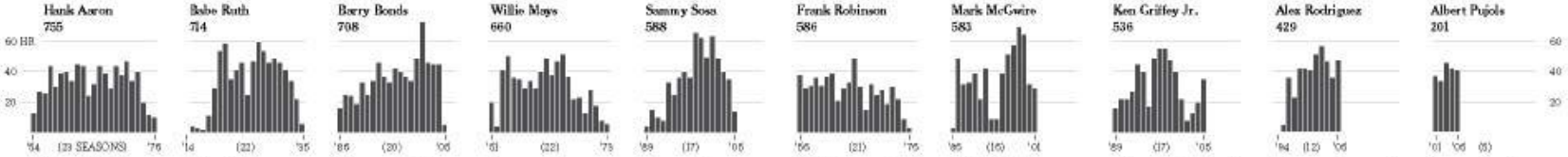


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

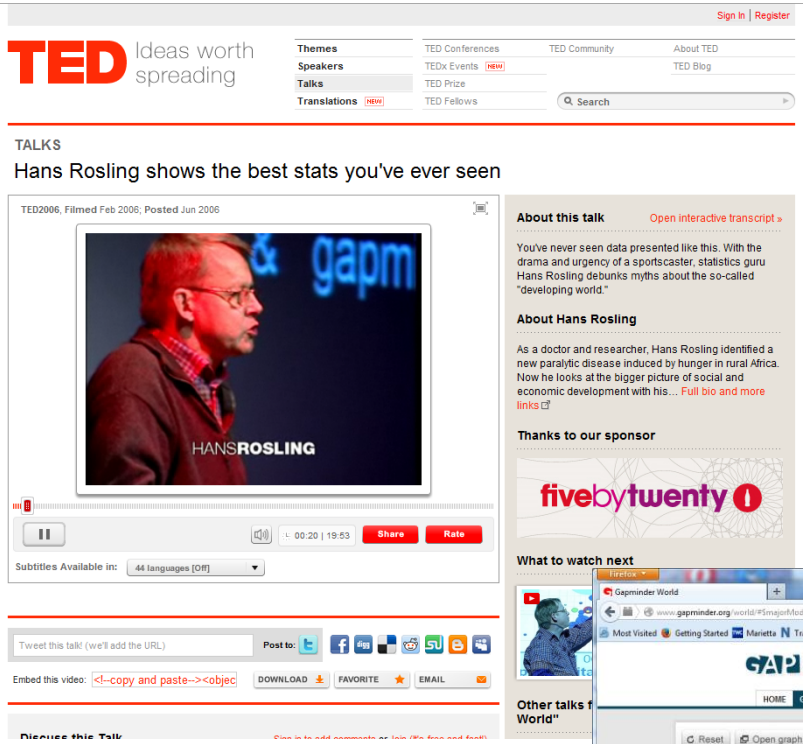


Differing Paths to the Top of the Charts

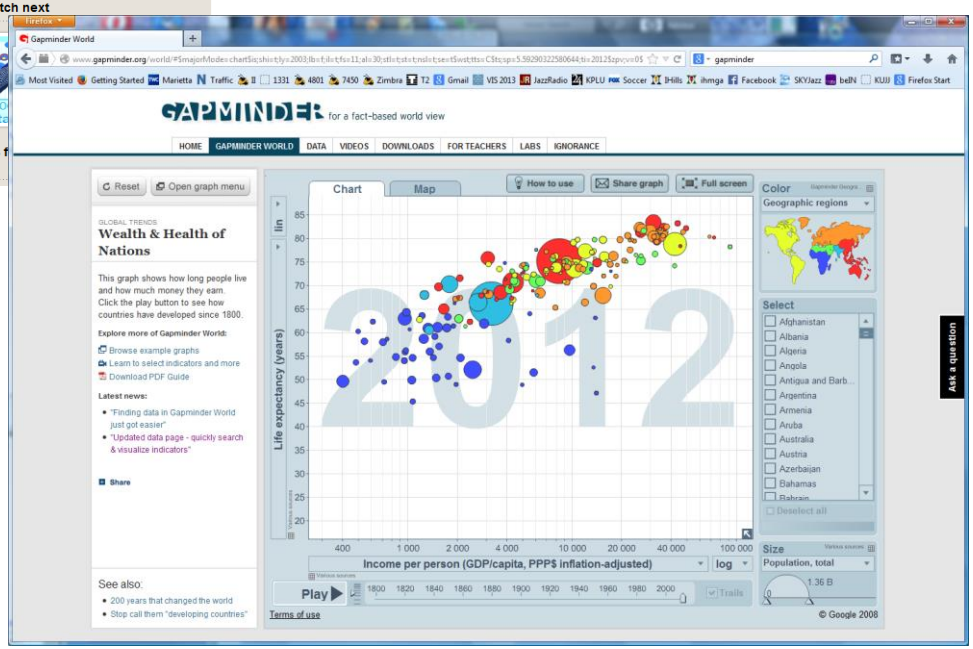
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).



Amanda Cox and Joe Ward/The New York Times



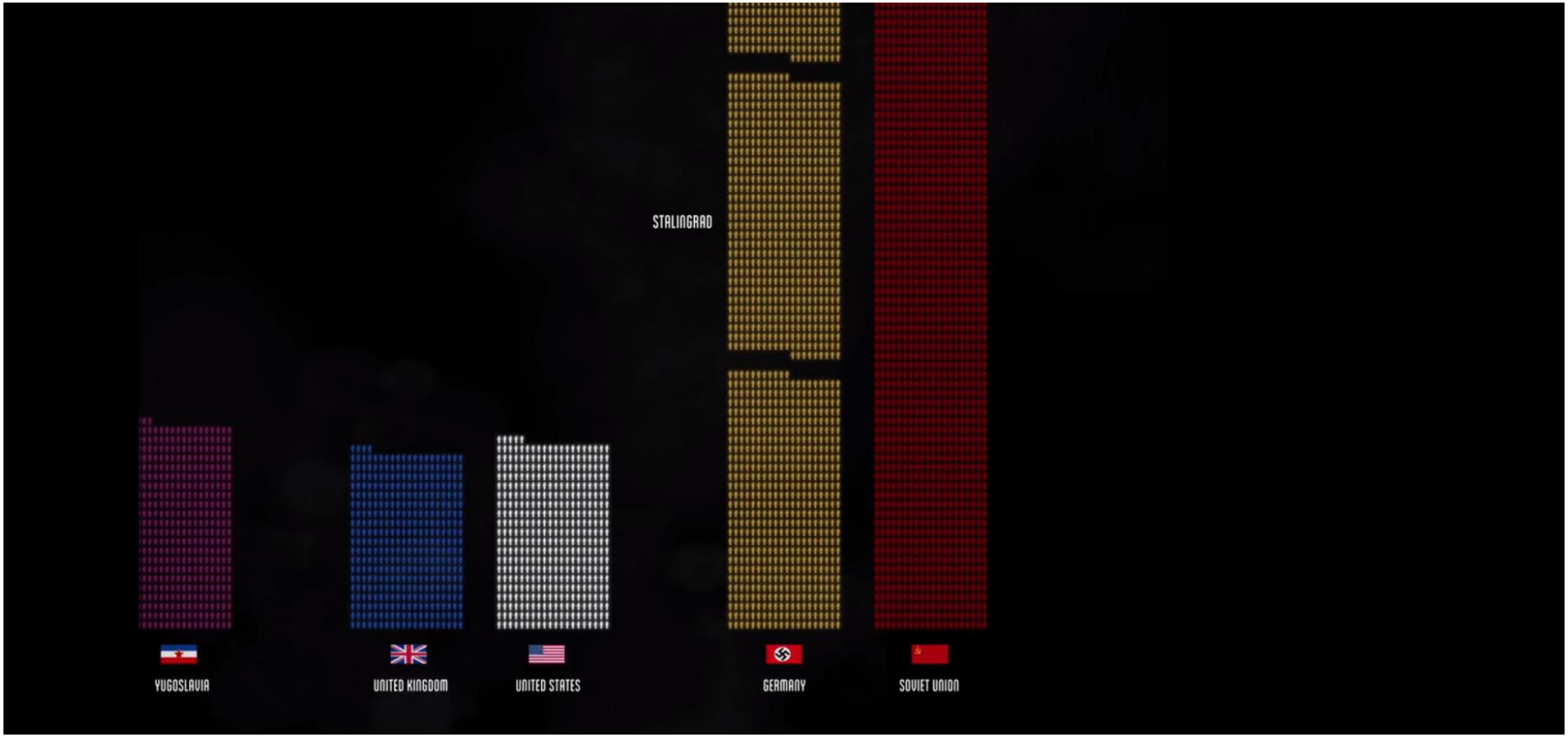
Hans Rosling
Gapminder



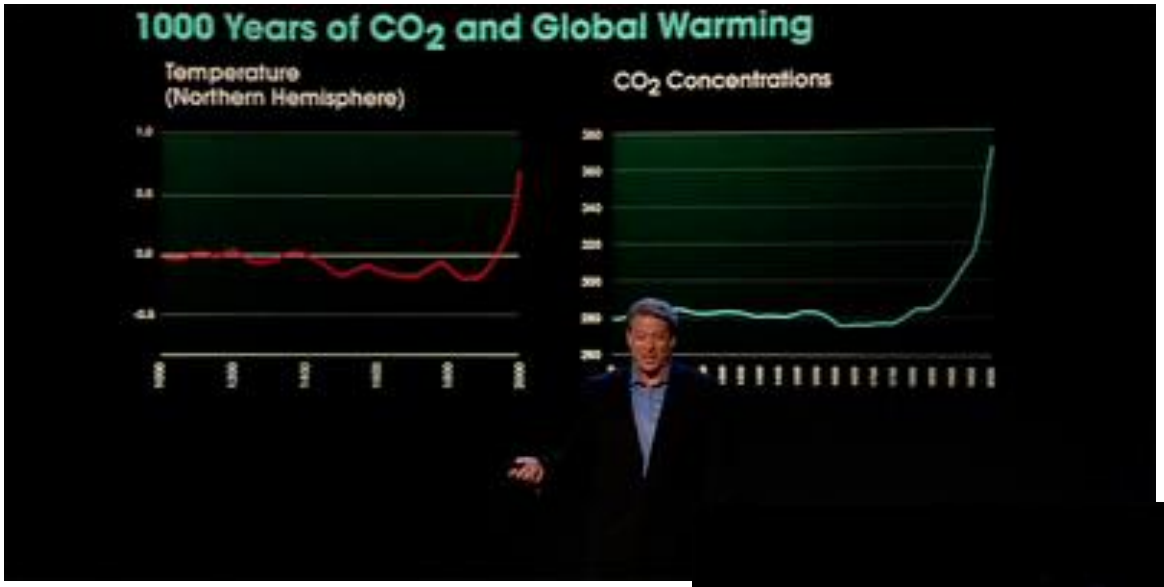
2006

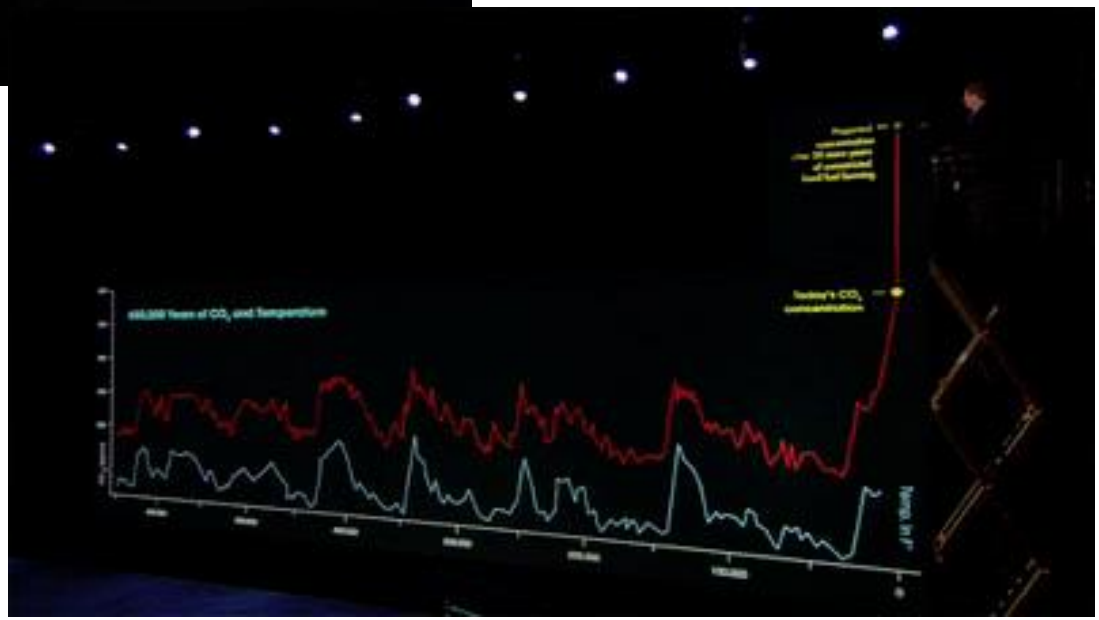
http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

http://www.fallen.io/ww2/



An Inconvenient Truth





Gore made extensive use of data graphics

Frequent presentation goals

Clarify

Focus

Highlight

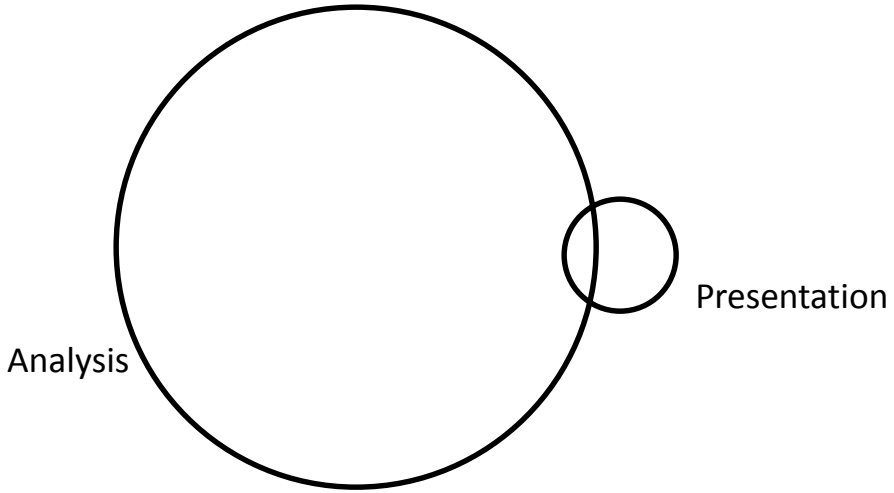
Simplify

Persuade

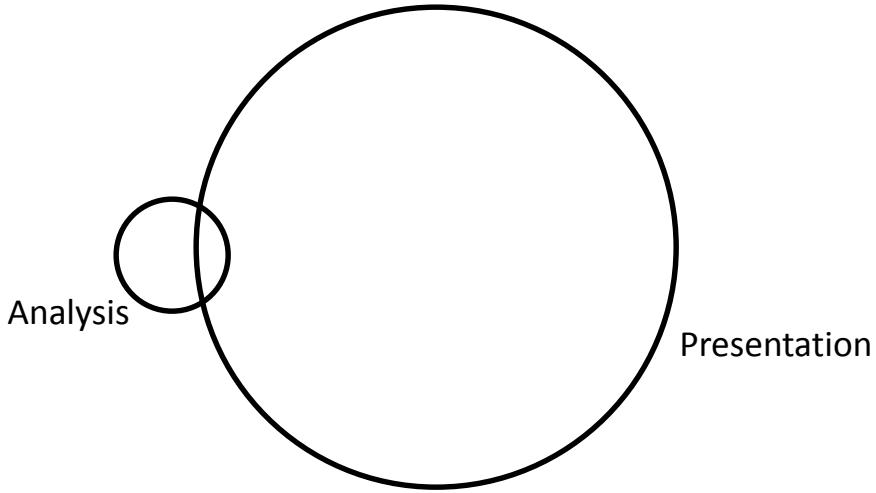
May just show a few variables
and/or a subset of the data cases

Views of the World

Vis researchers



Everyone else



Takeaway 1

Analysis $\neq$ Presentation

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

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

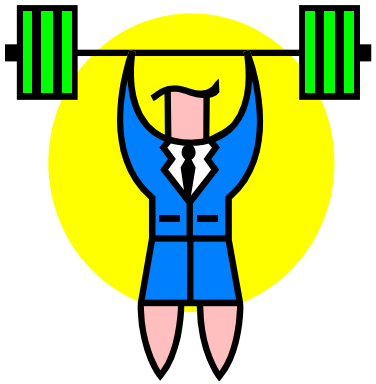
Get details

Navigation

Brushing & linking

Interaction

Can we do more?

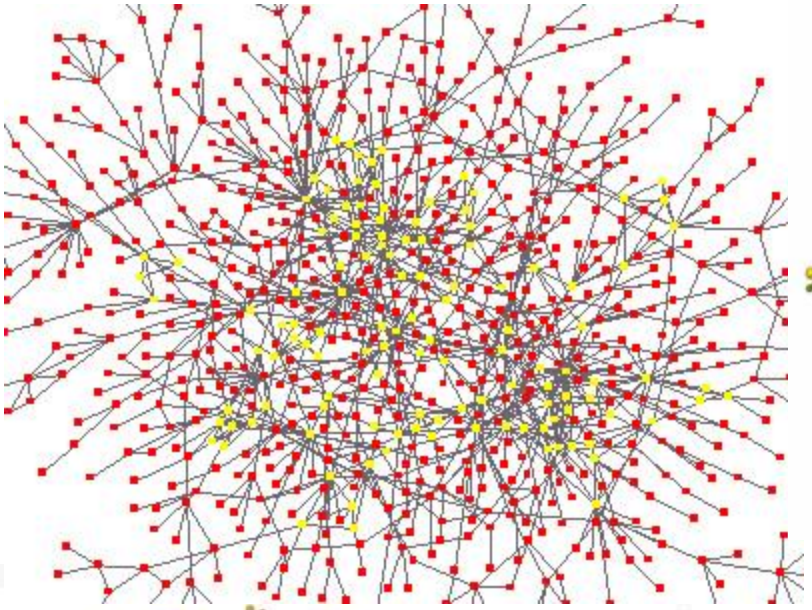


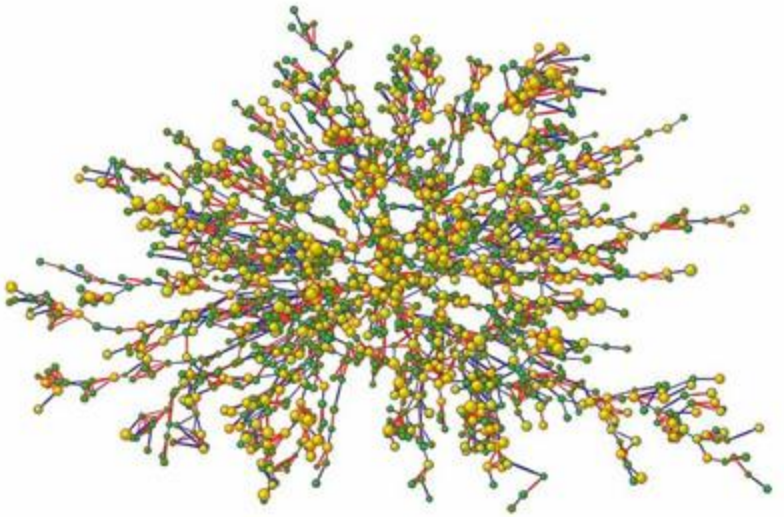
Employ interaction in a more fundamental manner to strengthen the power of visualization

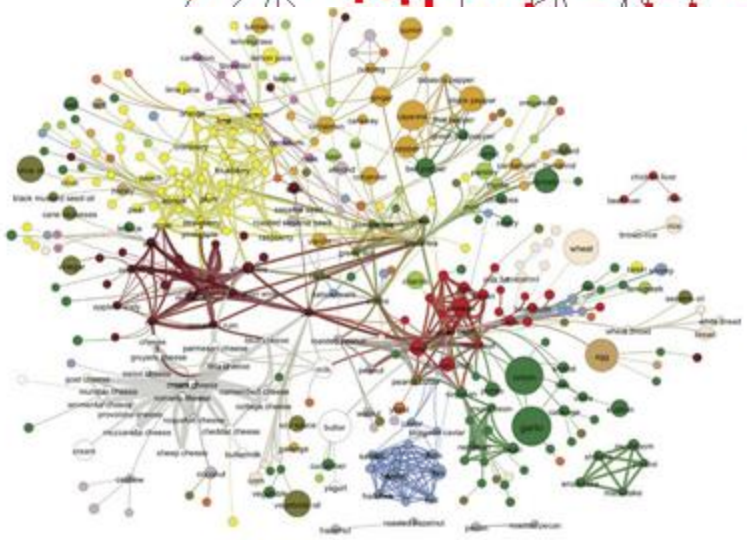
Some examples…

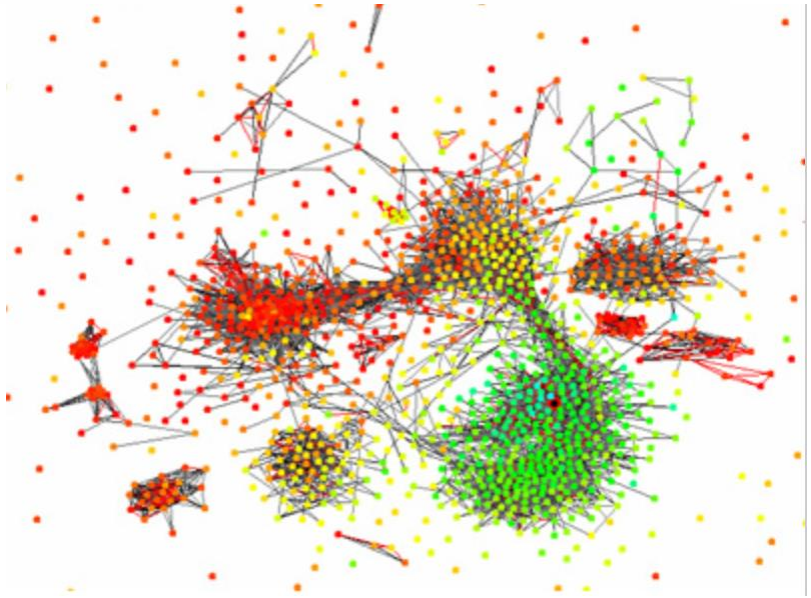
Node-link graphs

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



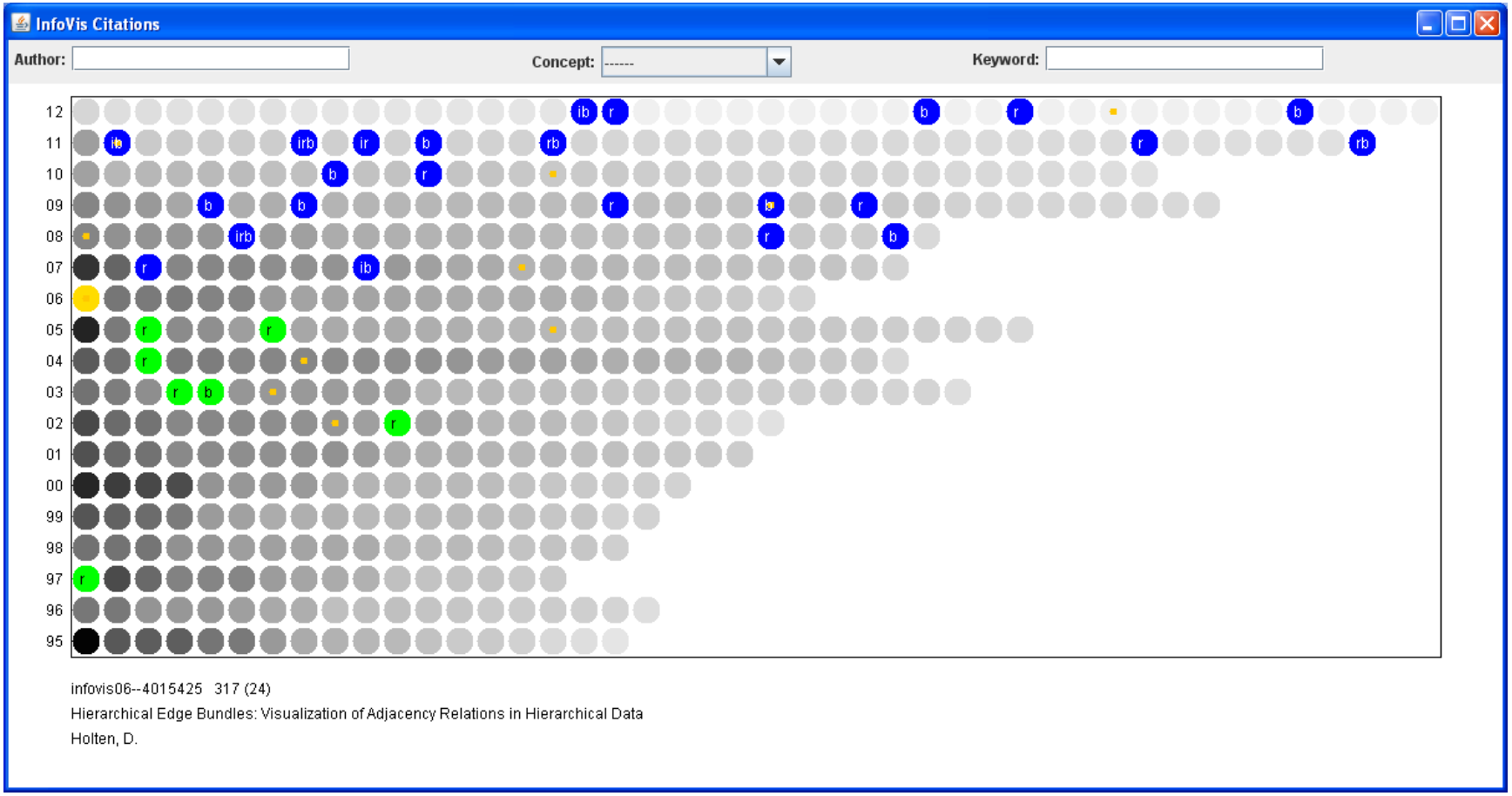






CiteVis

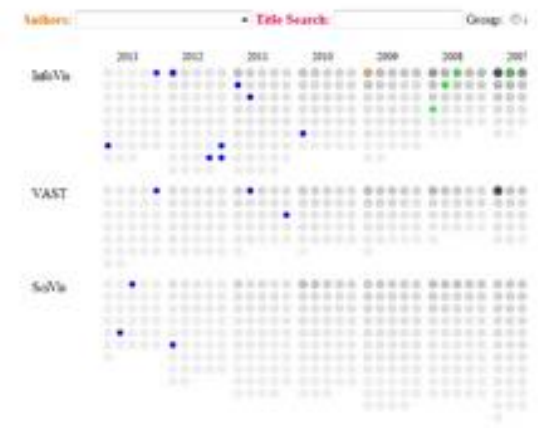
http://www.cc.gatech.edu/gvu/ii/citevis



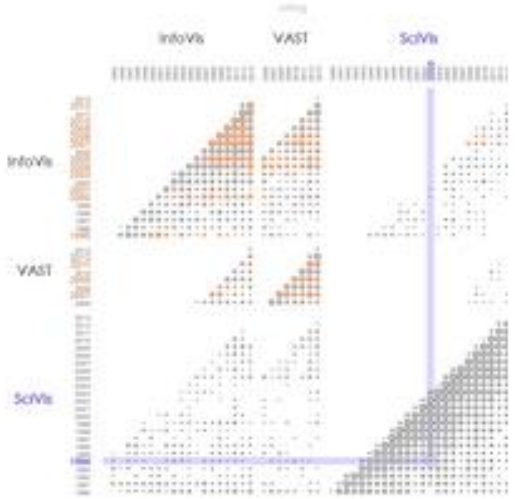
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

VIS25

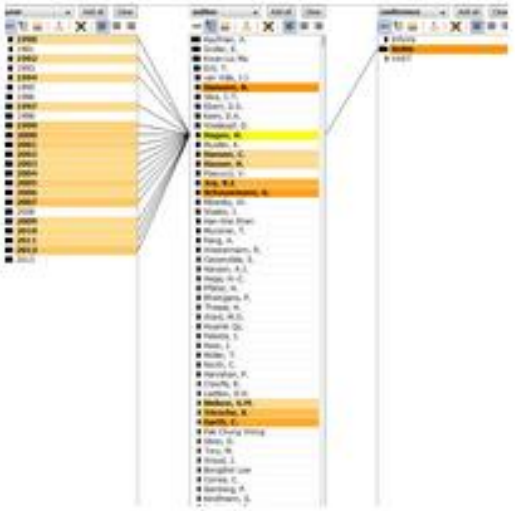
http://www.cc.gatech.edu/gvu/ii/citevis/VIS25/



CiteVis2



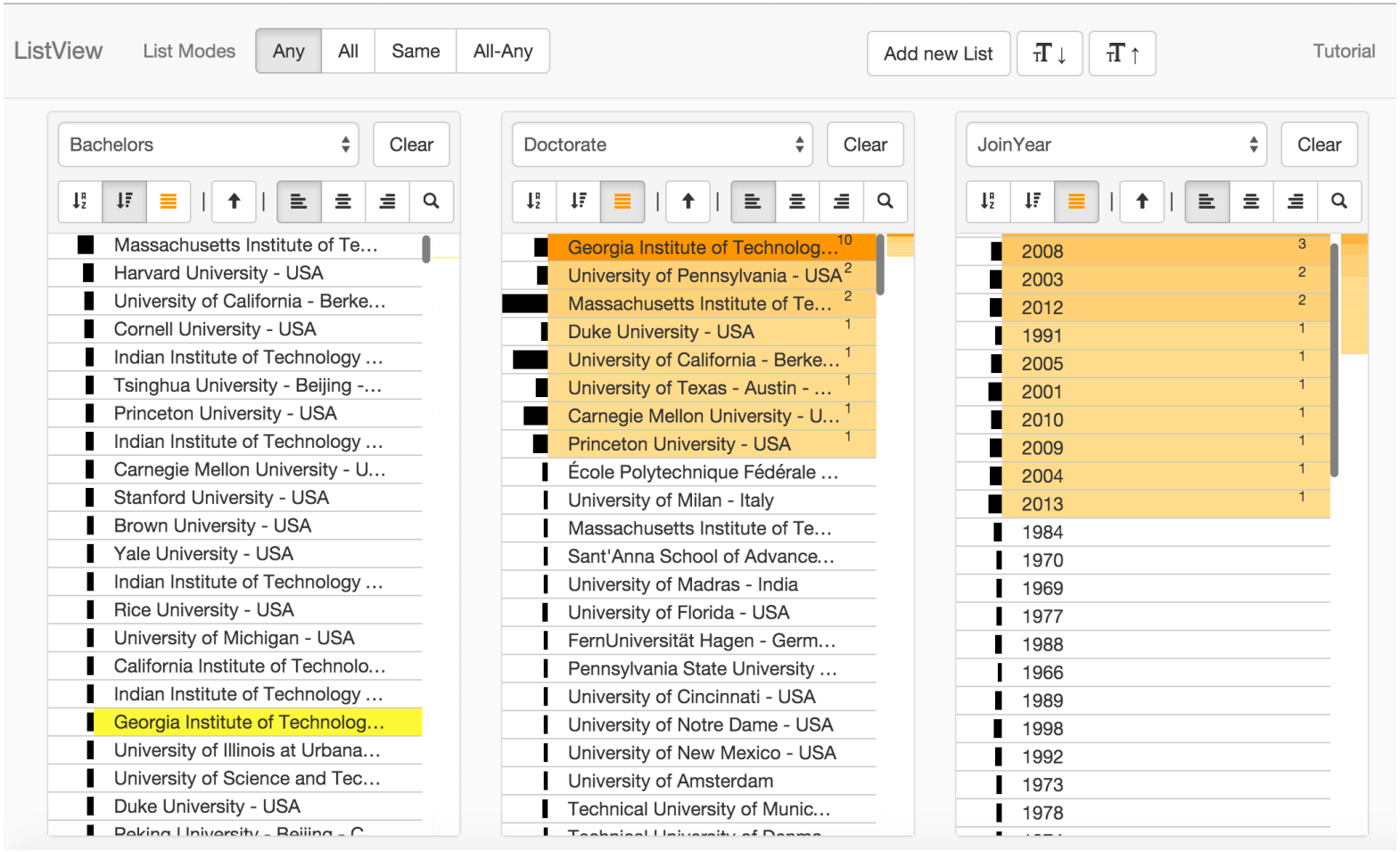
CiteMatrix



VISLists

Visualization design: Panpan Xu, Chad Stolper, Anand Sainath

ListView



S. Pratap A. Sainath

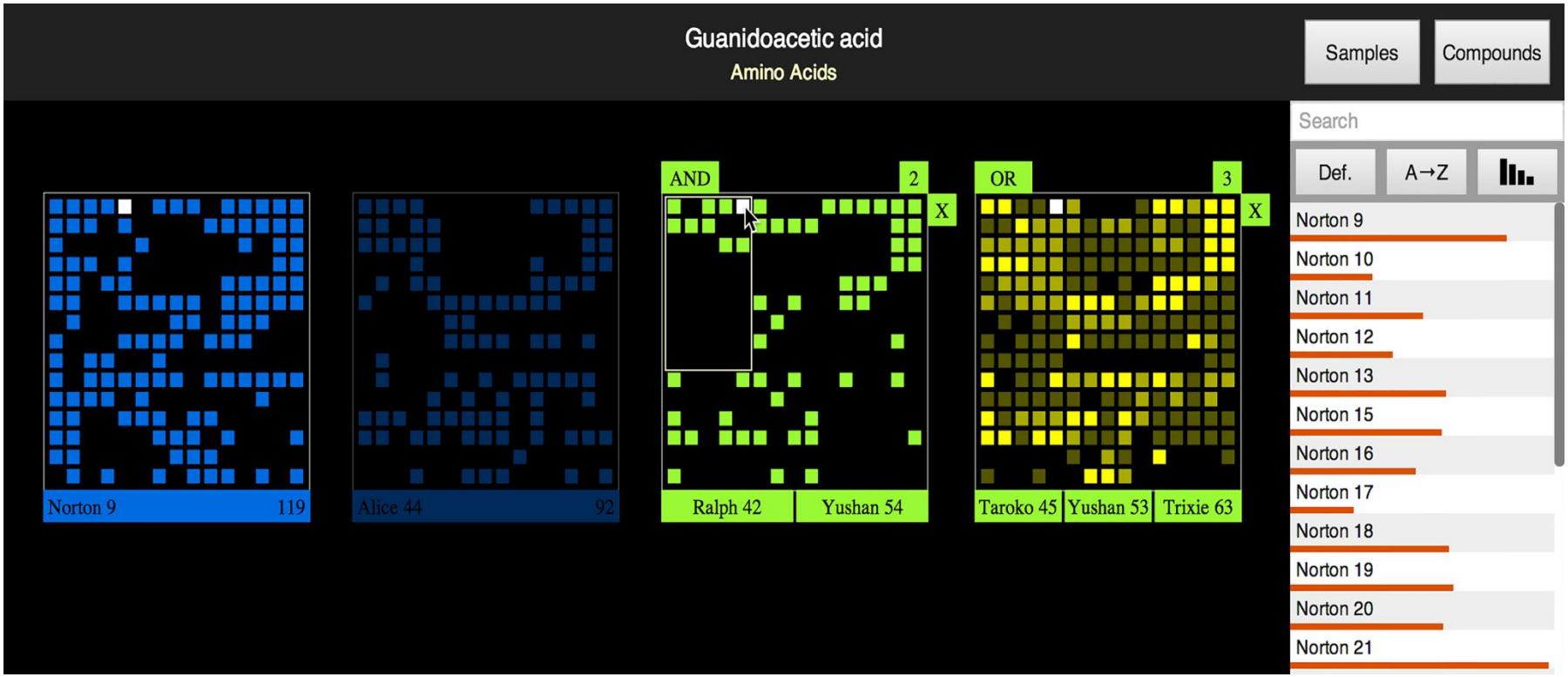




Demo

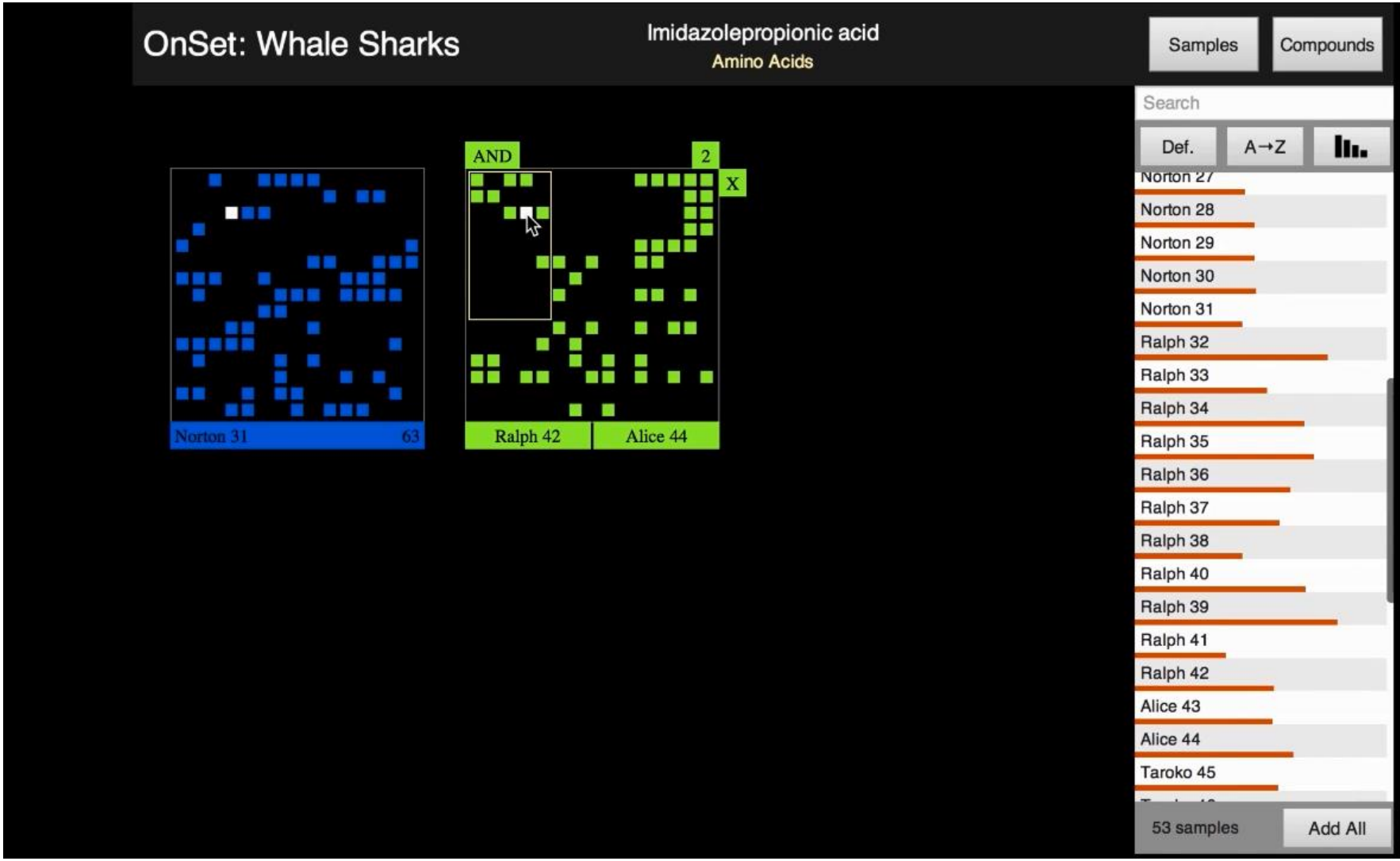
OnSet

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



Demo

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


OnSet: Whale Sharks
Imidazolepropionic acid
Amino Acids
Samples
Compounds
Search
Def.
A→Z
Norton 27
Norton 28
Norton 29
Norton 30
Norton 31
Ralph 32
Ralph 33
Ralph 34
Ralph 35
Ralph 36
Ralph 37
Ralph 38
Ralph 40
Ralph 39
Ralph 41
Ralph 42
Alice 43
Alice 44
Taroko 45
53 samples
Add All
AND
2
X
Norton 31
63
Ralph 42
Alice 44

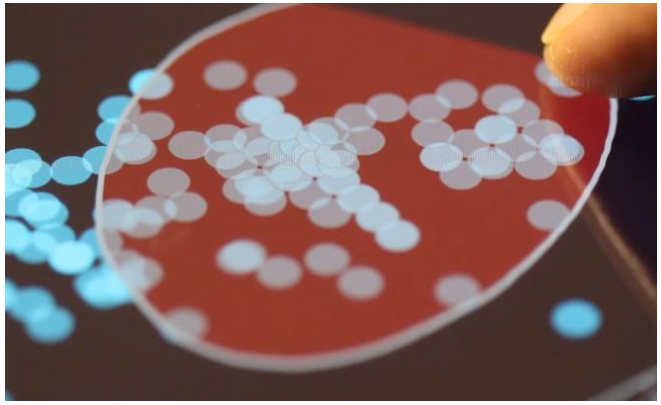
Interaction

What are the tools of interaction?

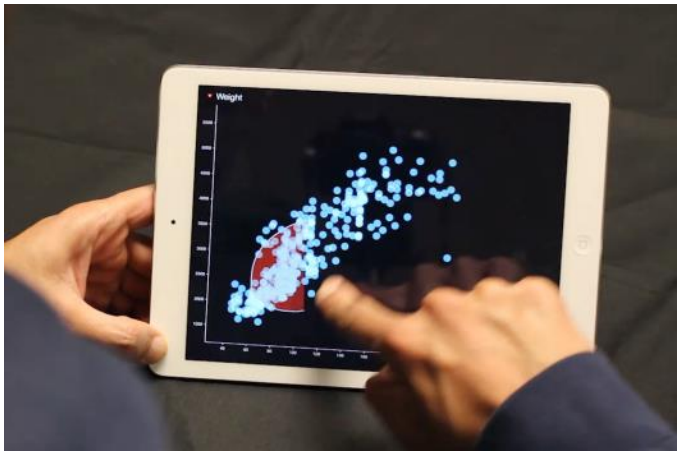
Traditional – Desktop: keyboard, mouse

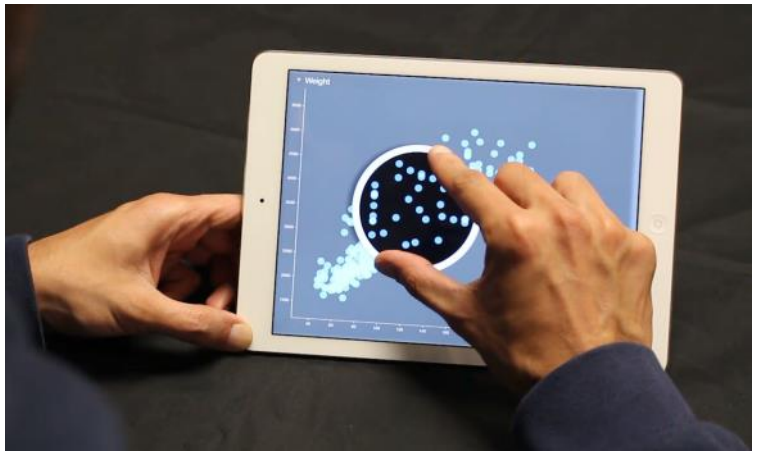
New – Tablet: fingers and multi-touch

Moving to Tablets



Scatterplot



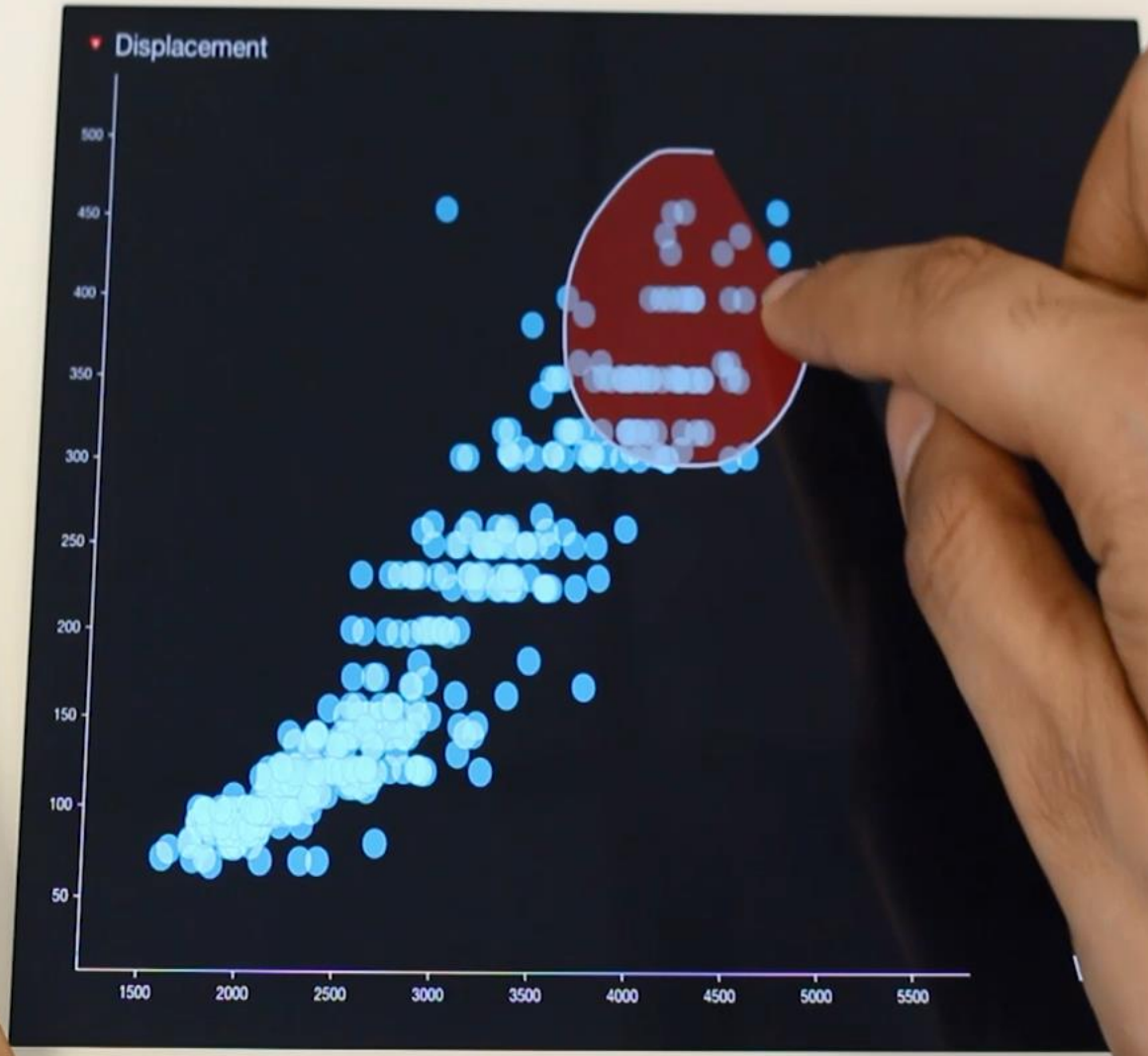




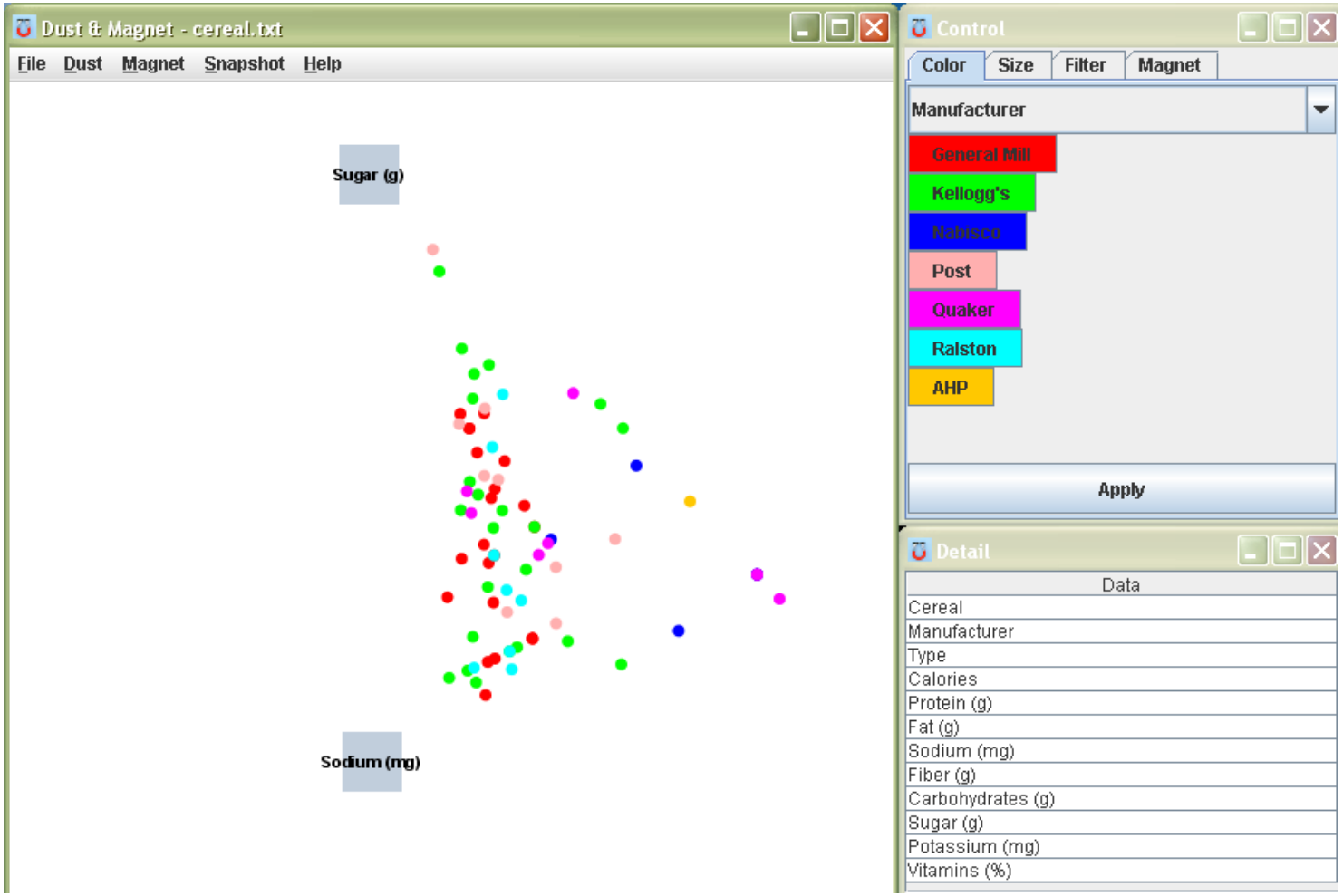
Video

Sadana & Stasko
AVI ‘14

Lasso Select



Dust and Magnet



Yi, Melton, Stasko & Jacko
Information Visualization '05

Dust and Magnet - Pushing it Further



Dai et al
InfoVis ‘15 poster

Takeaway 2

Take advantage of interaction

3. Identify Value



Visualization's Value

User Studies (typically)

Short, fact-based questions & tasks

Example: Car dataset

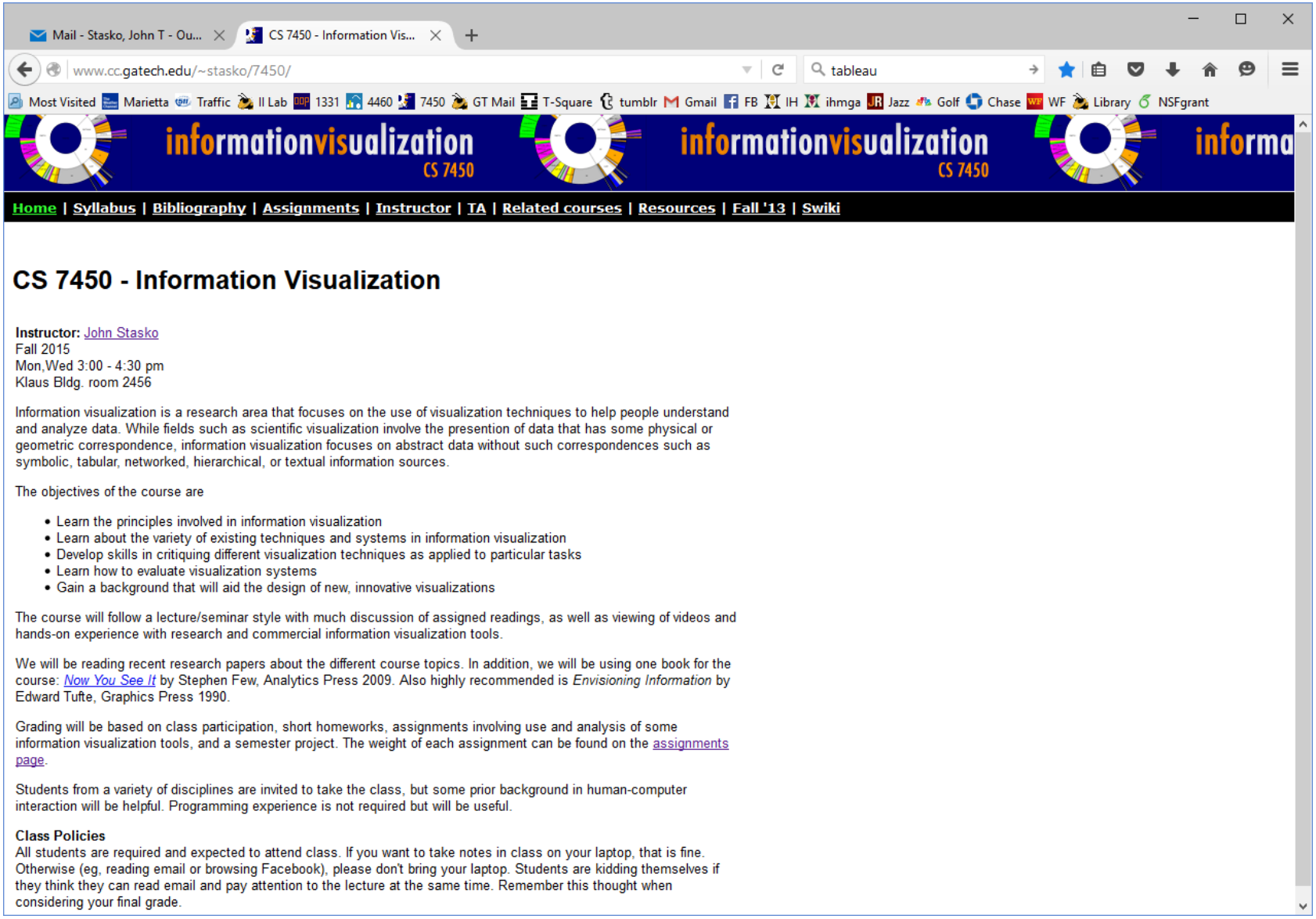
- Which cars have the best miles-per-gallon?
- How much does a Ford Taurus cost?
- Is there a correlation between car weight and torque?
- How many different countries manufacture cars?
- What is the range of car's horsepower?

Important for usability, but can get those answers in other ways

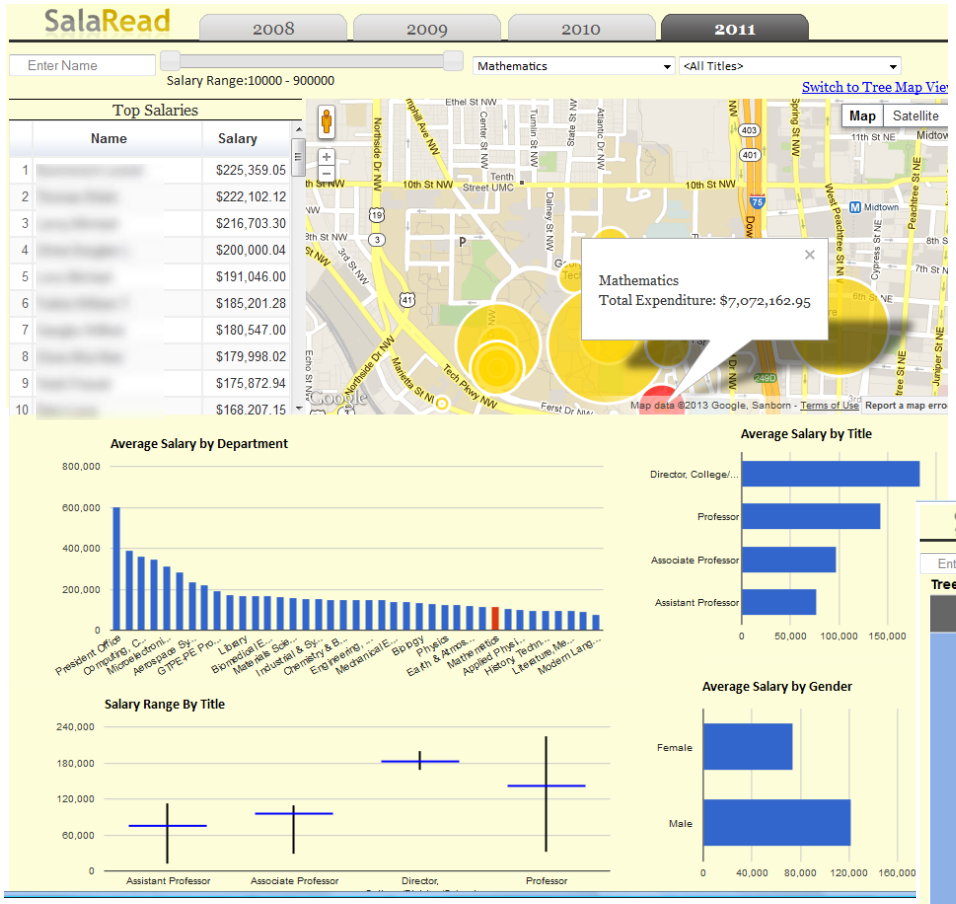
Simply presenting data ***visually*** can have a profound impact

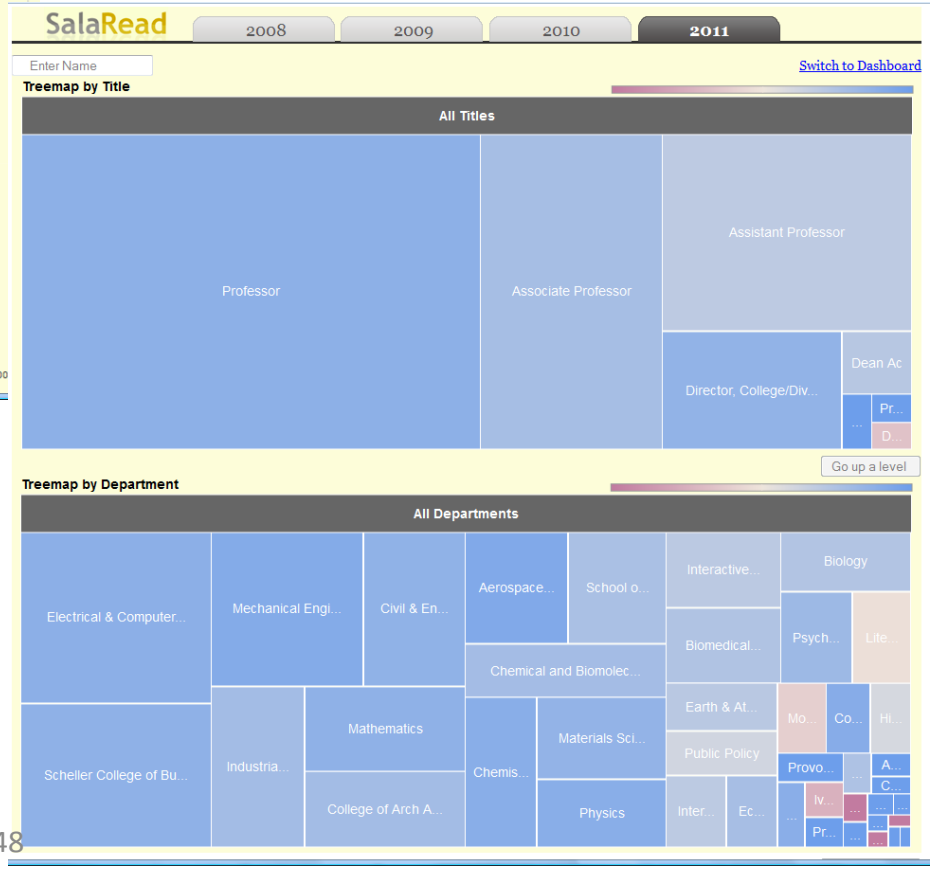
My Class

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



CS 7450
Fall '12
Design project

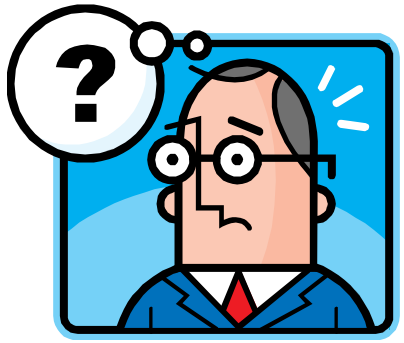




Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

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



Rethink evaluation as identifying a 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?

$$\text{Value} = T + \mathbf{I} + E + C$$

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

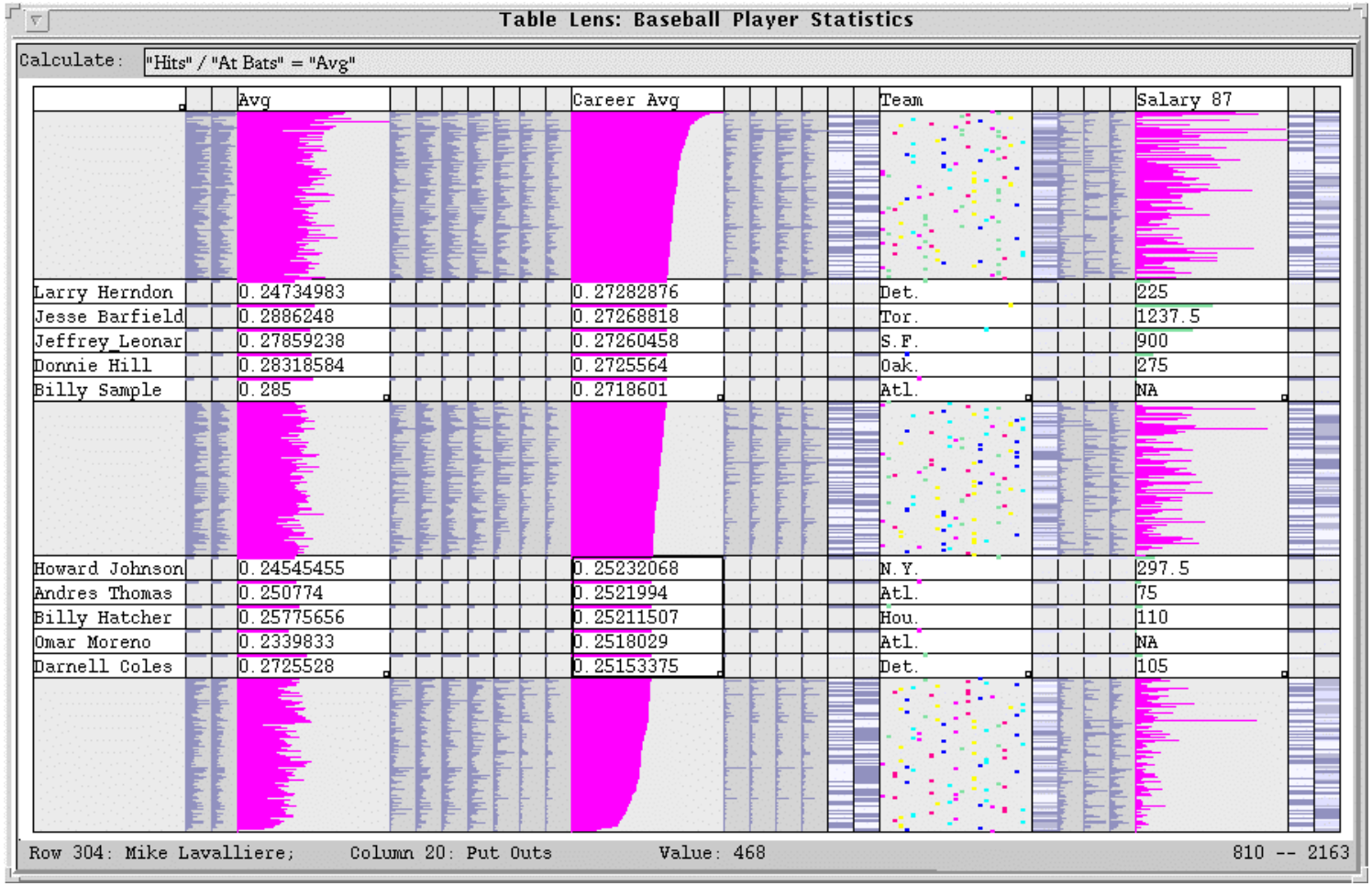


Table Lens

Rao & Card
CHI ‘94

Visualization's Value?

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

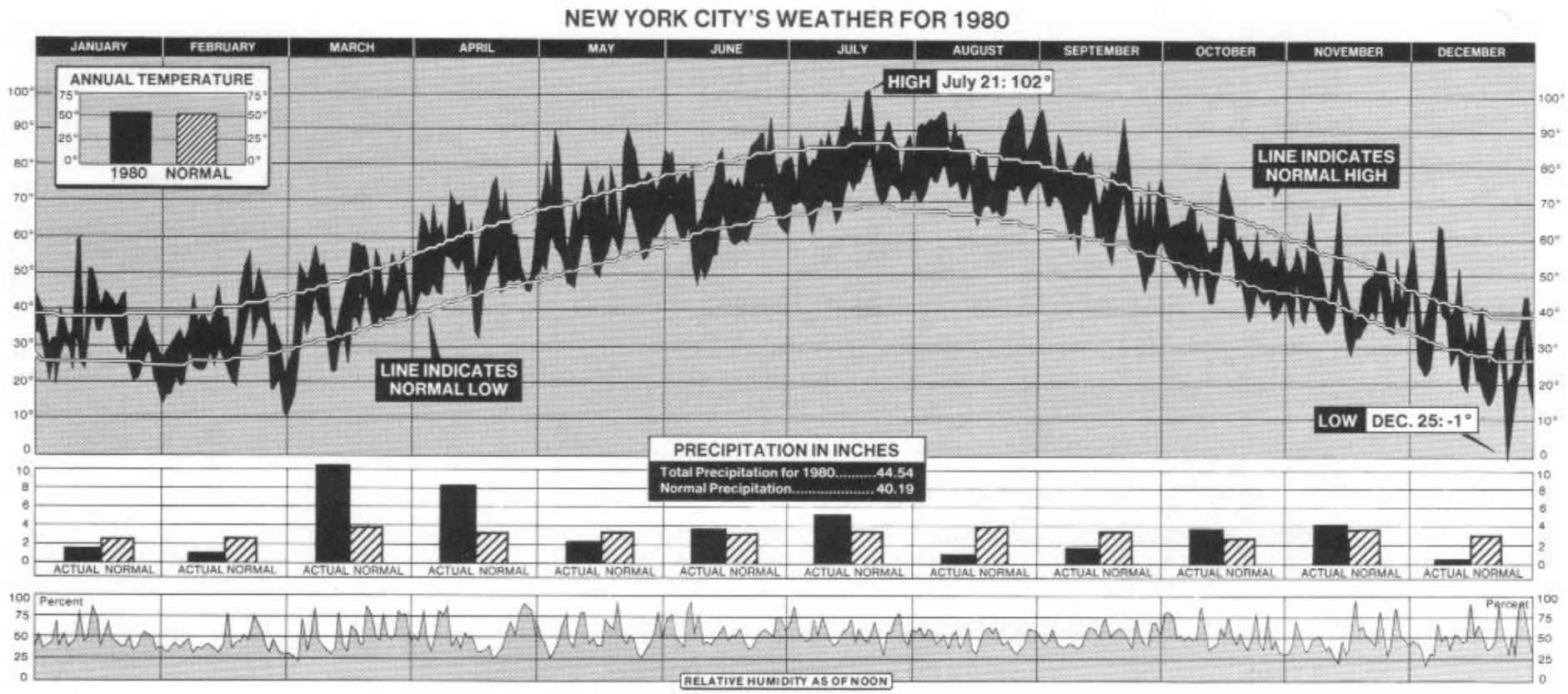
Ability to generate **confidence** and trust about the data, its domain and context

(Beneficial data analysis process side effects)

Some examples

1.

2220 numbers



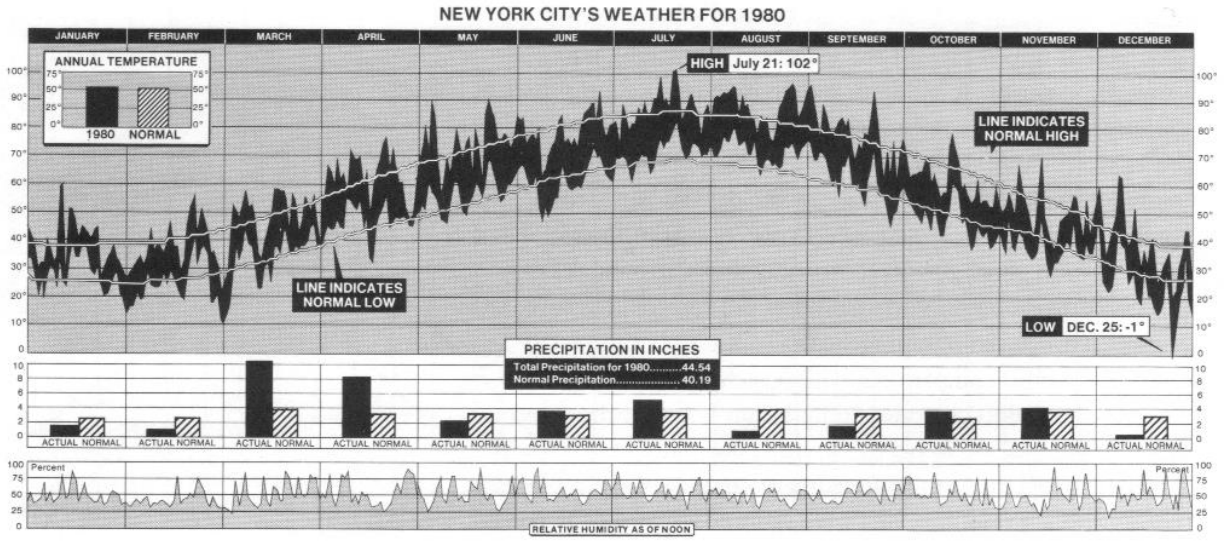
New York Times, January 11, 1981, p. 32.

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

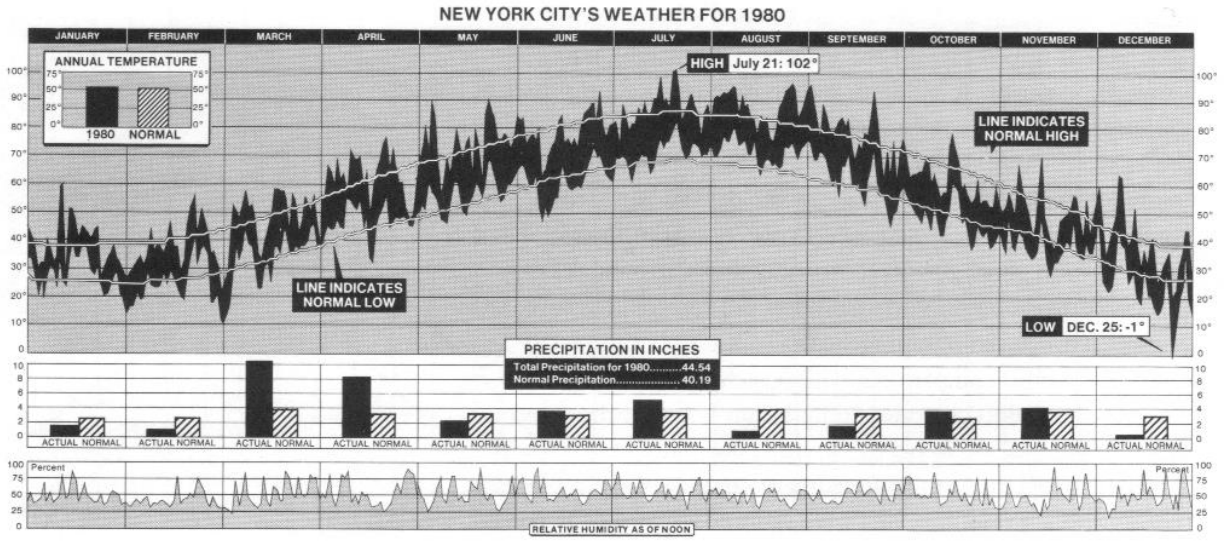
Visualization's Value



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

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

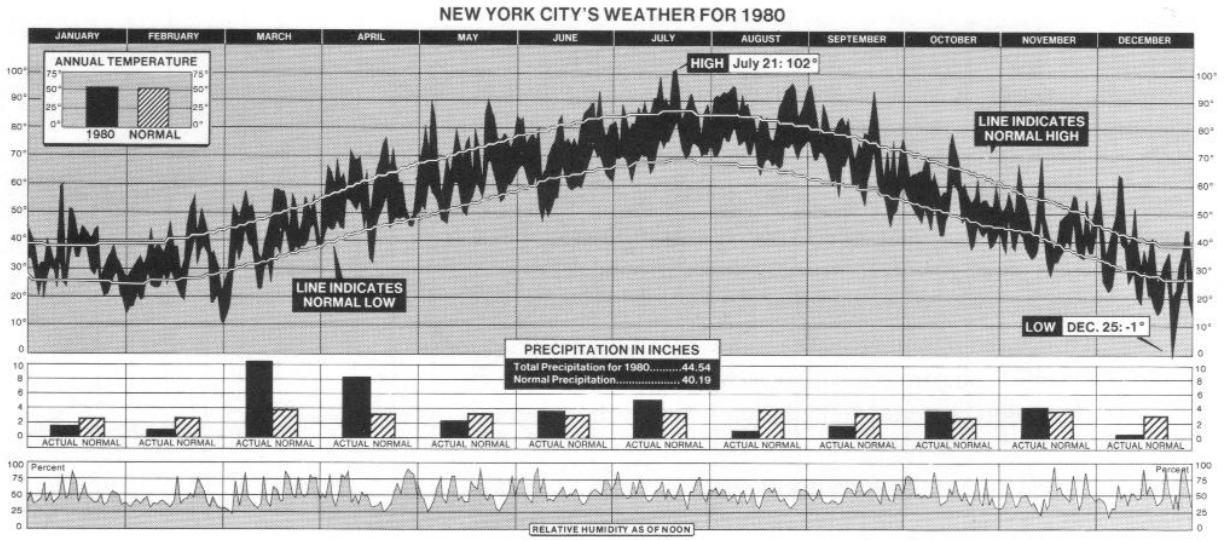
Visualization's Value



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

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

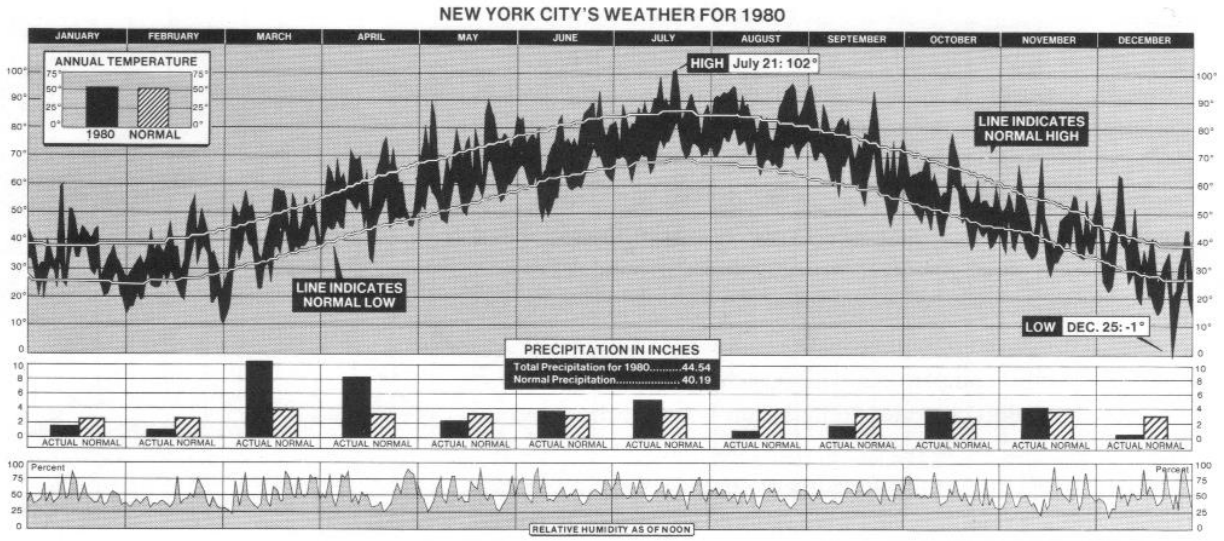
Visualization's Value



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

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

Visualization's Value



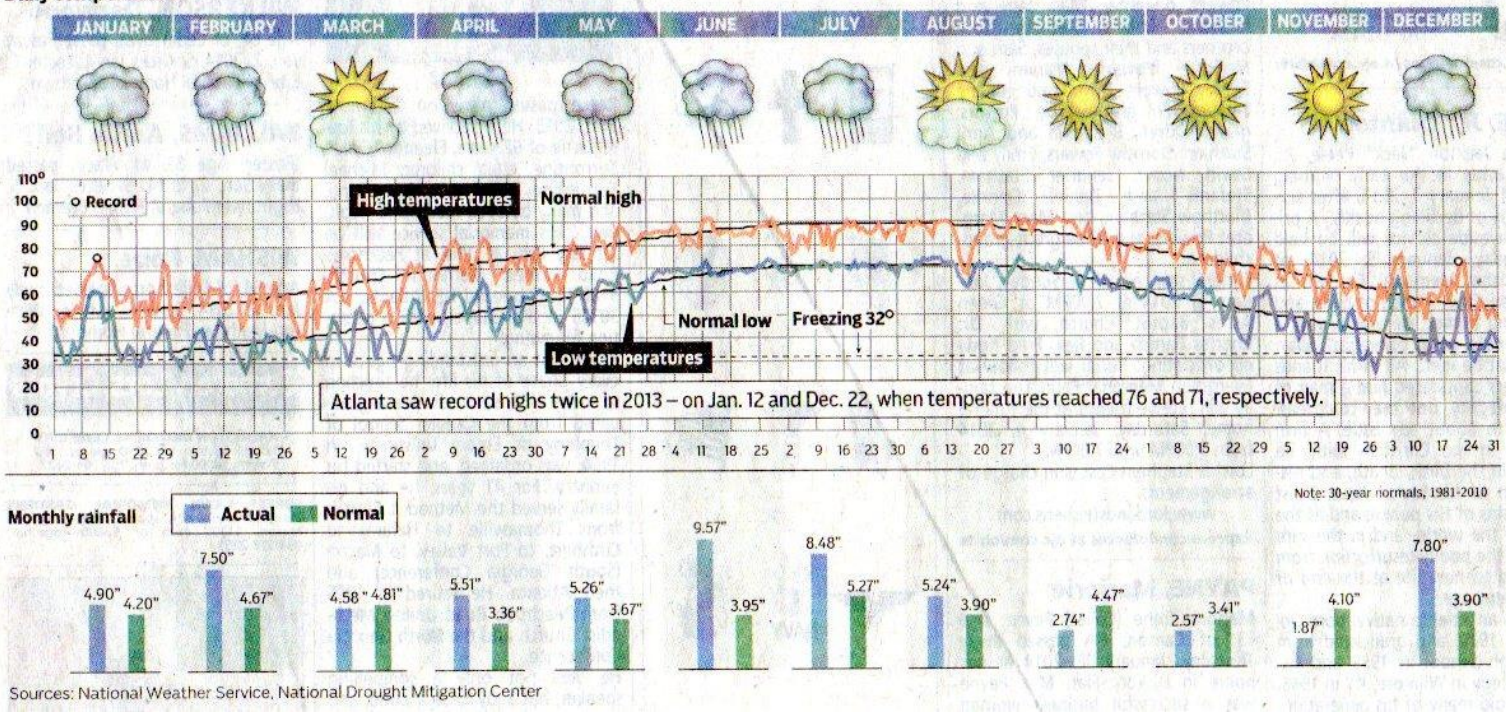
New York Times, January 11, 1981, p. 32.

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

A month-by-month look at Atlanta's weather in 2013

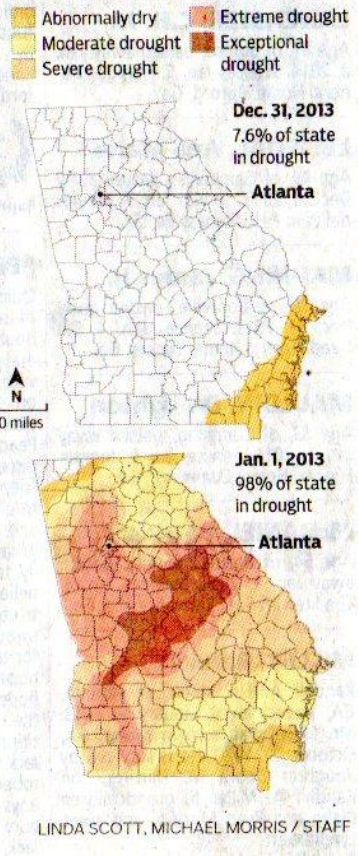
Atlanta's 2013 weather can be summed up in one word – soggy. The city ended the year 16.31 inches above normal in rainfall, with much of that rain falling during the waterlogged summer months, which had about twice as many rainy days as dry ones. Rain fell into the city's official gauge at Hartsfield-Jackson International Airport on 17 days in June, 19 days in July and 24 days in August. The rainfall total for those three months was 23.29 inches, making the period the city's fourth wettest summer on record. Daily rainfall records were set on April 28 (1.73 inches), May 4 (3.40 inches), June 5 (4.14 inches), Aug. 8 (2.60 inches) and Oct. 6 (2.12 inches). Summer 2013 will also be remembered for a lack of extreme heat, particularly compared to the sweltering summer of 2012. Temperatures climbed into the 90s only 20 times, topping out at 92 on June 28 and Aug. 29. That's 32 less days of 90+ heat than in 2012, when the all-time record high of 106 was reached.

Daily temperatures



Sources: National Weather Service, National Drought Mitigation Center

Georgia drought fades during 2013



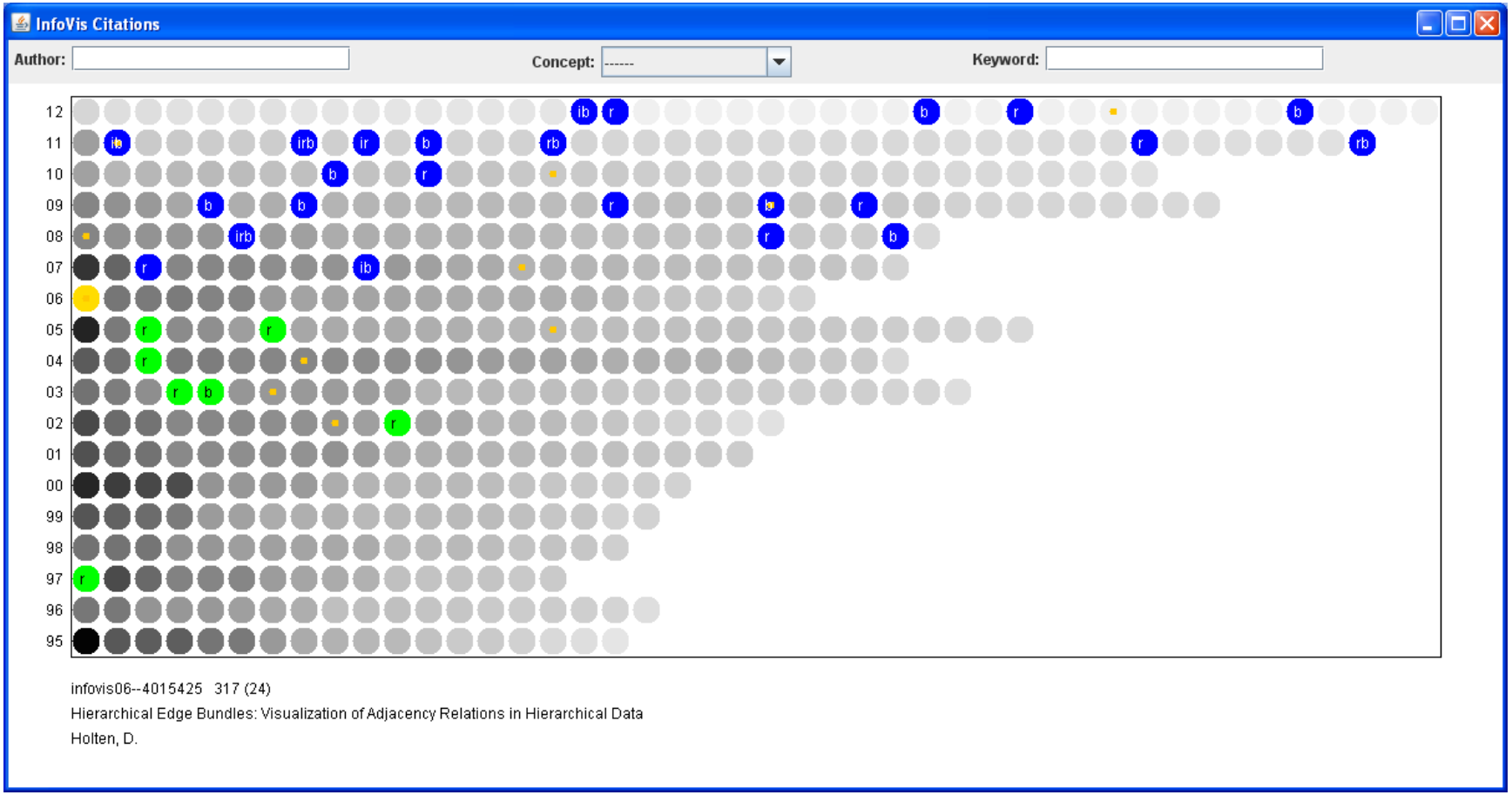
LINDA SCOTT, MICHAEL MORRIS / STAFF

Atlanta Journal Constitution
Jan. 3, 2014

2.

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

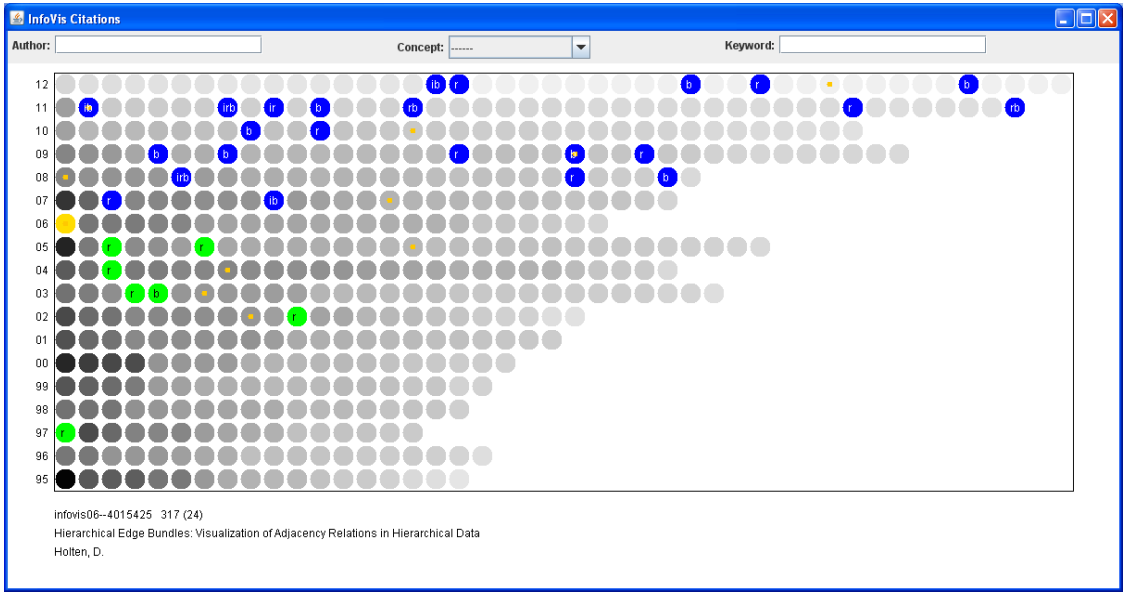
CiteVis



Earlier demo

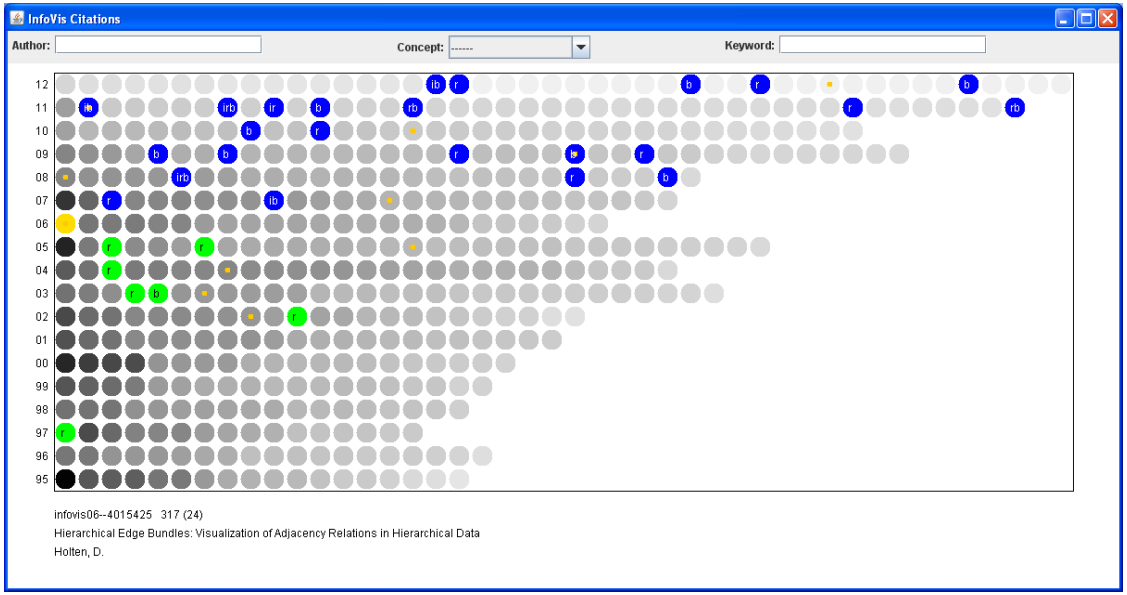
Stasko, Choo, Han, Hu, Pileggi, Sadana & Stolper
InfoVis poster '13

Visualization's Value



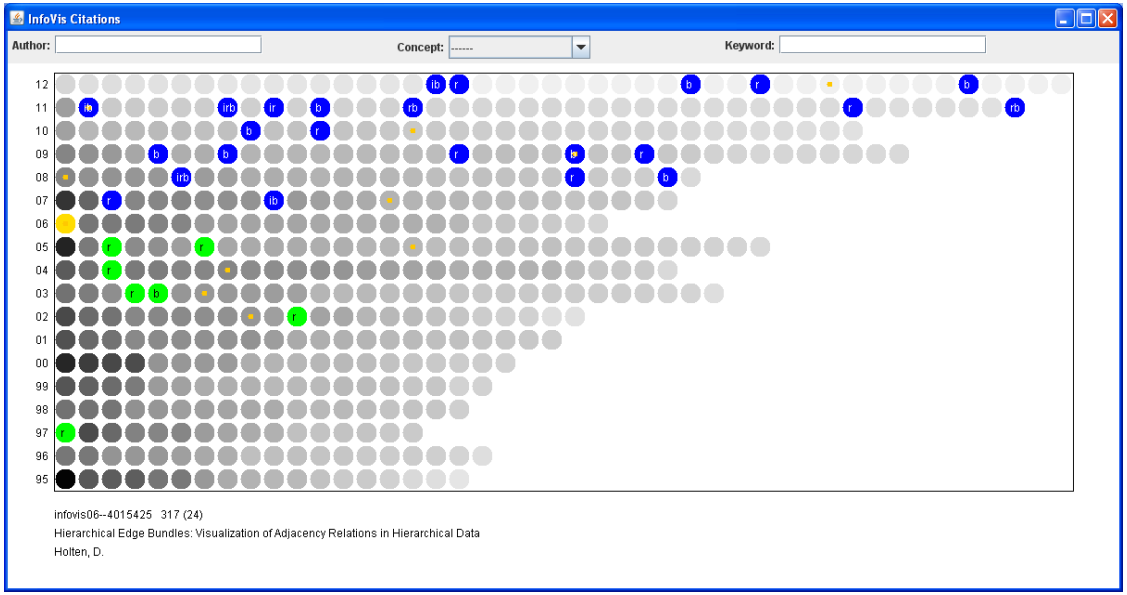
$$V_{alue} = T + I + E + C$$

Visualization's Value



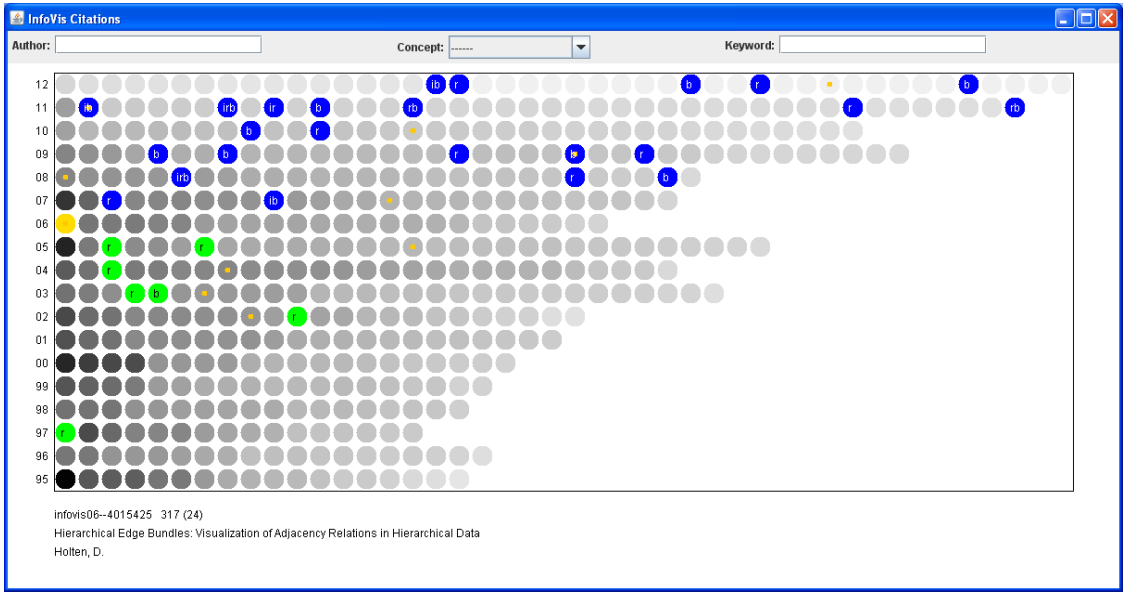
$$V_{alue} = T + I + E + C$$

Visualization's Value



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

Visualization's Value



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

Takeaway 3

Identify value

Key Open Problems in Visualization

Key Open Problems in Visualization

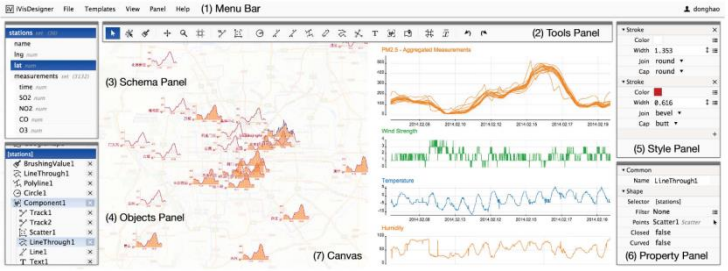
1. Assess and communicate **value**

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

Key Open Problems in Visualization

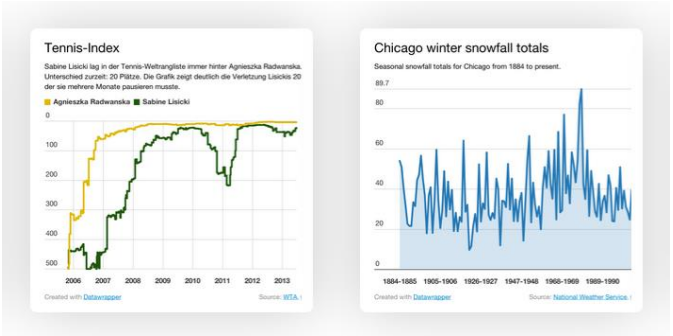
1. Assess and communicate value

2. Make the construction of visualizations easier



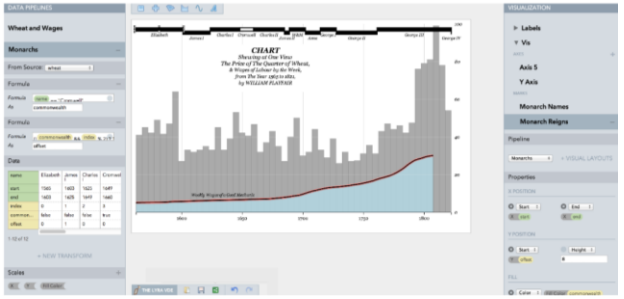
iVisDesigner
Ren, Hollerer, & Yuan
TVCG (InfoVis) '14

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)



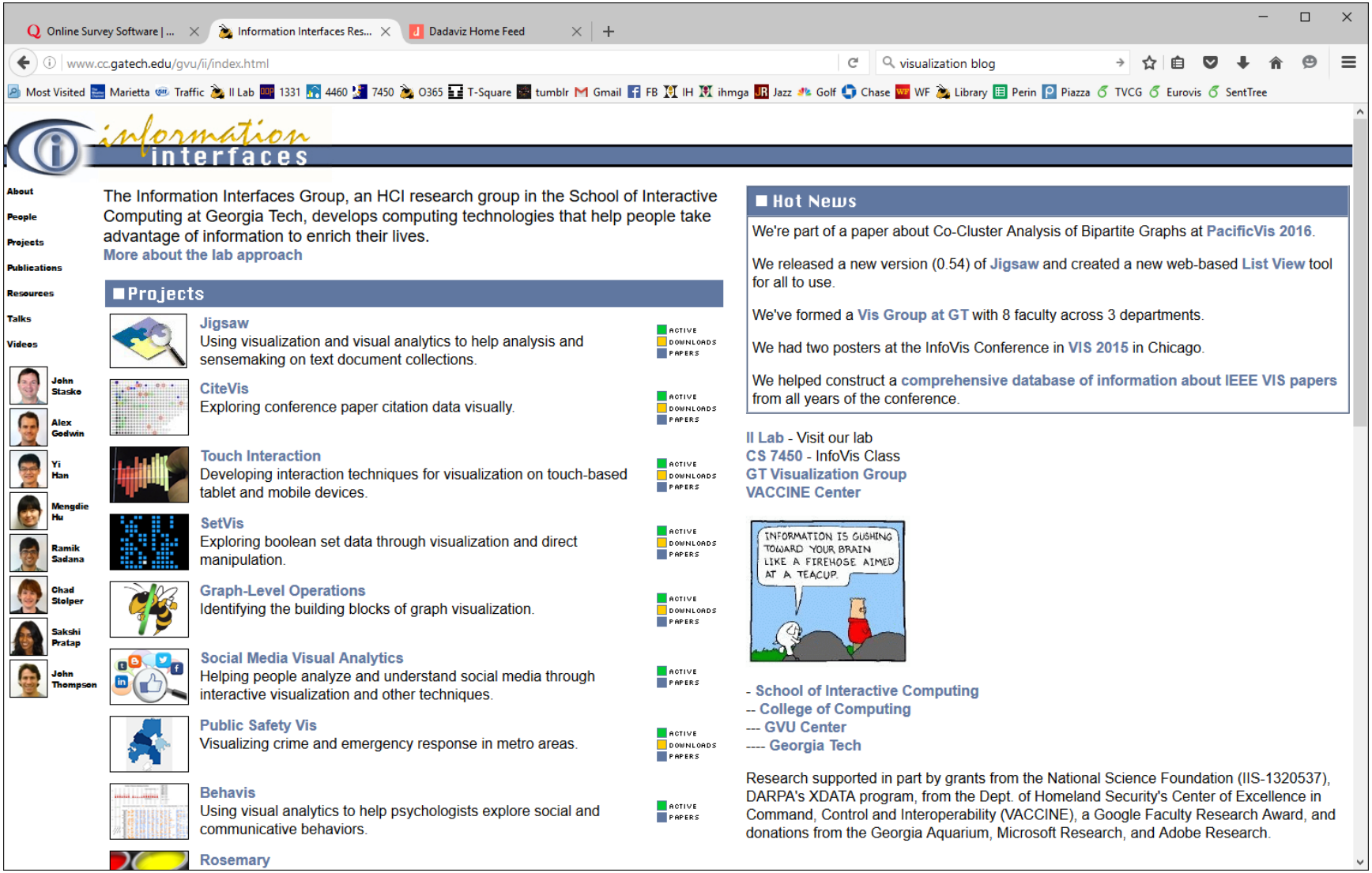


- Supported by CCF-0808863 (FODAVA lead), NSF IIS-0915788, NSF IIS-1320537

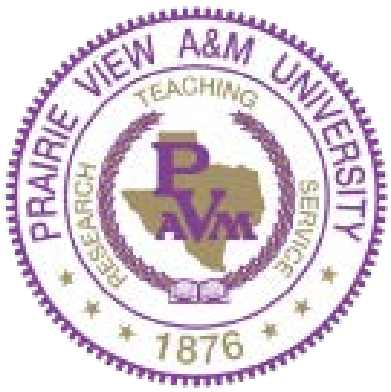


- Supported by gifts from Microsoft, Google, Nokia, and Adobe

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



Prairie View
A & M University



Applied Curriculum Integration Overview :

Crisis Informatics with NodeXL



Louis Ngamassi, Ph.D.
Assistant Professor of MIS
College of Business
Prairie View A&M University

Outline



- Brief Introduction of the College of Business at PVAMU
- *Crisis Informatics* course
  - Motivations
  - Content
  - Students evaluation
  - Challenges
    - Designing
    - Teaching
  - Lessons learned


PRAIRIE VIEW A&M UNIVERSITY
FOUNDED 1876



PRAIRIE VIEW A&M UNIVERSITY

College of Business

College of Business
Fast Facts



- **AACSB Accredited**
  - **"Gold" standard for Business Schools**
  - **Less than 5% of the more than 16,000 schools worldwide** (http://bestbizschools.aacsb.edu/aacsb-accredited#sthash.VZuLfCuD.dpuf)

College of Business

Fast Facts



- AACSB Accredited
- Enrollment
  - Approximately 1,000 undergraduate students
  - 200 graduate students (MBA, MSA)

College of Business

Fast Facts



- AACSB Accredited
- Enrollment
  - Approximately 1,000 undergraduate students
  - 200 graduate students (MBA, MSA)
- **Faculty & Staff**
  - **40 full-time faculty (90% have earned doctoral degrees)**
  - **Eight full-time and one part-time professional staff**

College of Business

Majors



- 5 Undergraduate Majors
  - Accounting
  - Finance
  - Management Information Systems
  - Management
  - Marketing

College of Business

Majors



- 5 Undergraduate Majors
  - Accounting
  - Finance
  - **Management Information Systems**
  - Management
  - Marketing

Sample courses

- Information Systems Analysis & Design
- Data Communication
- Java Programming
- Database
- Introduction to Object-Oriented programming
- IT Strategic Management

College of Business

Majors



- 5 Undergraduate Majors
  - Accounting
  - Finance
  - Management Information Systems
  - Management
  - Marketing
- **2 Graduate Majors**
  - **MBA (Online, Traditional, Executive)**
  - **MSA (MS in Accounting)**

College of Business

Who's Hiring Our Students



Crisis Informatics Course

Motivations



- **AACSB accreditation renewal**
  - **Need for innovations**
    - **Houston area prone to disaster**
    - **Houston big hub for Oil & Gas and health industries**
    - **Other colleges are engaged in disaster related issues**

Motivations



- AACSB accreditation renewal
  - Need for innovations
- **Changes in the MIS curriculum**
  - **SAP**
  - **Spreadsheet skills**
  - **Big data (analytics)**
  - **Non business major**

Motivations



- AACSB accreditation renewal
  - Need for innovations
- Changes in the MIS curricular
  - SAP
  - Spreadsheet skills
  - Big data (analytics)
- **Research into classroom**

Content

Course overview



- The course examines the inter-connectedness of information, people, and technologies in a crisis.
- It explores how information is collected, organized, processed, and used for crisis management.
- The course includes hands-on class activities involving the use of NODEXL to explore crisis related information generated through social media.

Content

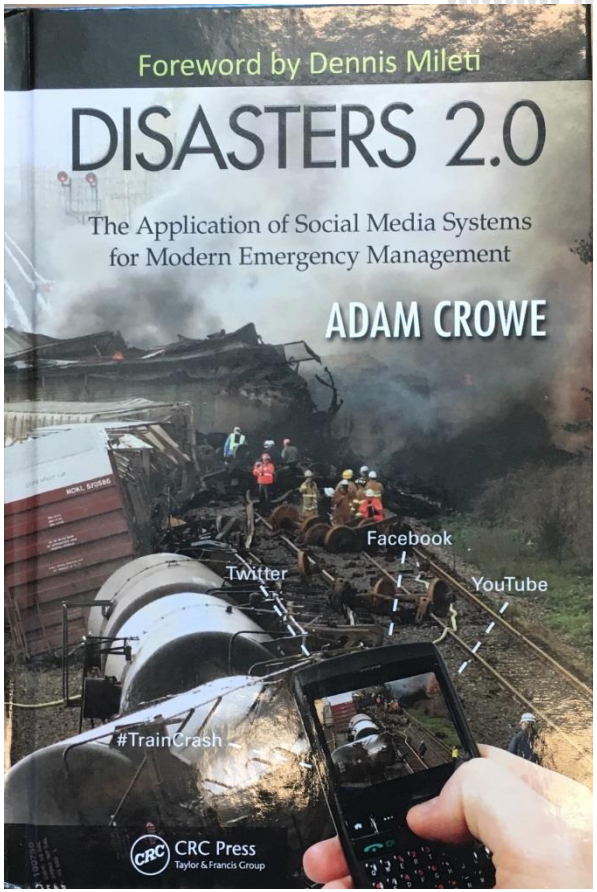
Course learning objectives



- Understand the importance of information in crisis management
- Understand the current level of the impact of social media on crisis management
- Analyze complex information needs and information seeking behavior during a crisis
- Explore the use of social network analysis in crisis management
- Explore the use of visual analytic tools in crisis management.

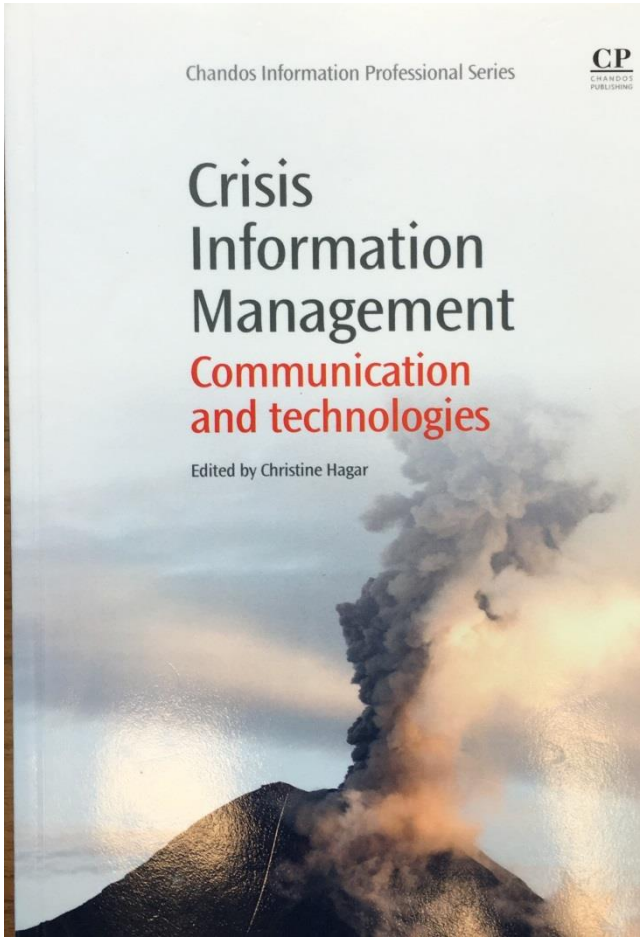
Content

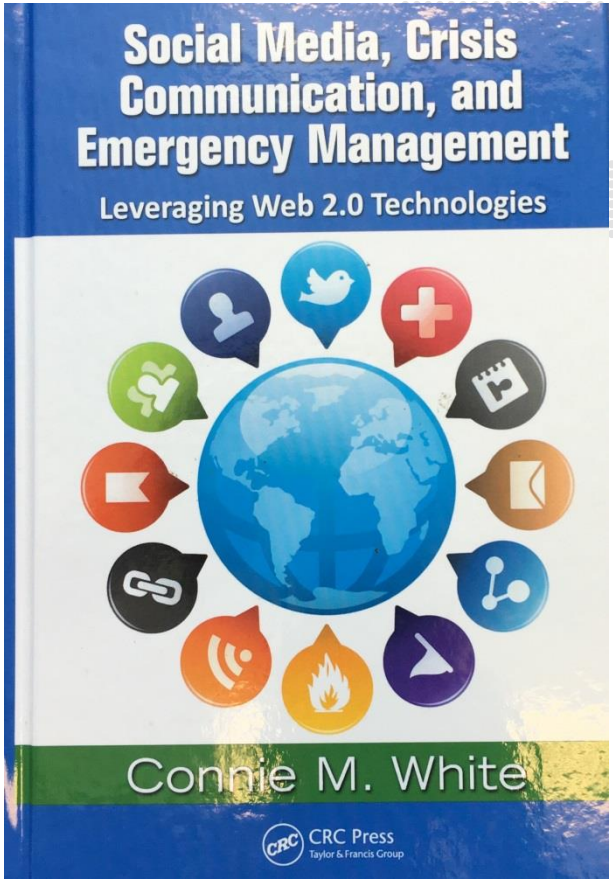
Required text book

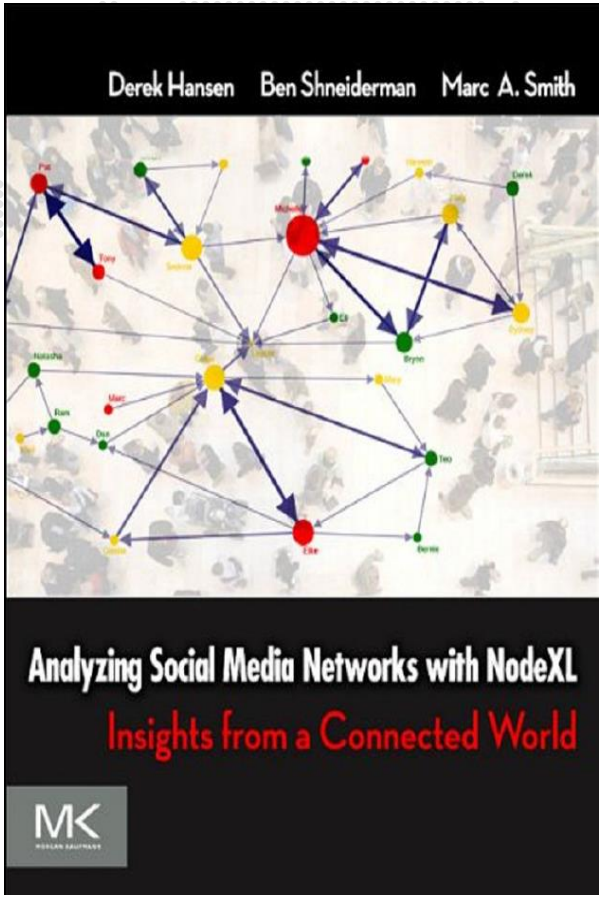


Content

Supplementary readings







Content

Why NODEXL?





- Excel add-in
- Relational data analysis
- Feedback from the Industry
- Capturing social media data
- Visual analytic component

Content

Course schedule

Course Schedule

Week	F2F/ Online	Activities and Readings	Deliverable
#1 Aug 24	F2F	Class Introduction and Overview Introduction to Emergency Management	
#2 Aug 31	F2F	**Chap1:** Introduction to the Application of Social Media in Modern Emergency Management **Chap2:** Social Media Systems: Overview and Purpose	
#3 Sep07		**No Class (Labor Day)**	
#4 Sep14	F2F	Social Network Analysis Fundamentals **Hands-on:** Introduction to NodeXL - 1	Project Team Members HW#1
#5 Sep21	Online	**Chap4:** Engagement Challenges in the Application of Social Media	
#6 Sep28	Online	**Chap6:** Monitoring and Analysis of Social Media Information	HW#2
#7 Oct05	F2F	**Hands-on:** Introduction to NodeXL - 2	Team Project Topic
#8 Oct12	F2F	**Midterm Exam**	HW#3
#9 Oct19	Online	**Chap7:** Aggregation and Validation of Social Media Information	
#10 Oct26	Online	**Chap8:** Modern Integrated Emergency Warning and Notification Strategies	HW#4
#11 Nov02	F2F	**Hands-on:** Using NodeXL to capture, analyze and visualize social media data	
#12 Nov09	Online	**Chap9:** How Social Media Has Revolutionized the Management and Recruitment of People and Supplies	HW#5
#13 Nov16	Online	**Chap10:** The Fundamental Flaw within Formal Response Systems	
#14 Nov23	Online	**Chap14:** Available Tools and Systems to Improve Emergency Management Functions	Team Project Paper
#15 Nov30	F2F	Review for Final Exam Team project presentation	Team Project Presentation
#16	F2F	**Final Exam**	

Content

Course schedule

Hands-on

Course Schedule

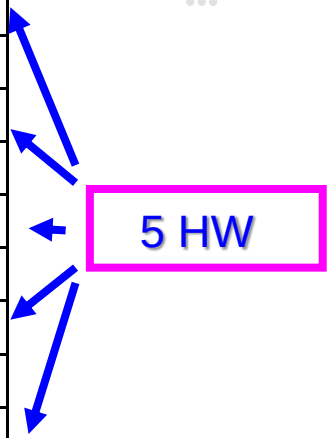
Week	F2F/ Online	Activities and Readings	Deliverable
#1 Aug 24	F2F	Class Introduction and Overview Introduction to Emergency Management	
#2 Aug 31	F2F	**Chap1:** Introduction to the Application of Social Media in Modern Emergency Management **Chap2:** Social Media Systems: Overview and Purpose	
#3 Sep07		**No Class (Labor Day)**	
#4 Sep14	F2F	Social Network Analysis Fundamentals **Hands-on:** Introduction to NodeXL - 1	Project Team Members HW#1
#5 Sep21	Online	**Chap4:** Engagement Challenges in the Application of Social Media	
#6 Sep28	Online	**Chap6:** Monitoring and Analysis of Social Media Information	HW#2
#7 Oct05	F2F	**Hands-on:** Introduction to NodeXL - 2	Team Project Topic
#8 Oct12	F2F	**Midterm Exam**	HW#3
#9 Oct19	Online	**Chap7:** Aggregation and Validation of Social Media Information	
#10 Oct26	Online	**Chap8:** Modern Integrated Emergency Warning and Notification Strategies	HW#4
#11 Nov02	F2F	**Hands-on:** Using NodeXL to capture, analyze and visualize social media data	
#12 Nov09	Online	**Chap9:** How Social Media Has Revolutionized the Management and Recruitment of People and Supplies	HW#5
#13 Nov16	Online	**Chap10:** The Fundamental Flaw within Formal Response Systems	
#14 Nov23	Online	**Chap14:** Available Tools and Systems to Improve Emergency Management Functions	Team Project Paper
#15 Nov30	F2F	Review for Final Exam Team project presentation	Team Project Presentation
#16	F2F	**Final Exam**	

Content

Course schedule

Course Schedule

Week	F2F/ Online	Activities and Readings	Deliverable
#1 Aug 24	F2F	Class Introduction and Overview Introduction to Emergency Management	
#2 Aug 31	F2F	**Chap1:** Introduction to the Application of Social Media in Modern Emergency Management **Chap2:** Social Media Systems: Overview and Purpose	
#3 Sep07		**No Class (Labor Day)**	
#4 Sep14	F2F	Social Network Analysis Fundamentals **Hands-on:** Introduction to NodeXL - 1	Project Team Members HW#1
#5 Sep21	Online	**Chap4:** Engagement Challenges in the Application of Social Media	
#6 Sep28	Online	**Chap6:** Monitoring and Analysis of Social Media Information	HW#2
#7 Oct05	F2F	**Hands-on:** Introduction to NodeXL - 2	Team Project Topic
#8 Oct12	F2F	**Midterm Exam**	HW#3
#9 Oct19	Online	**Chap7:** Aggregation and Validation of Social Media Information	
#10 Oct26	Online	**Chap8:** Modern Integrated Emergency Warning and Notification Strategies	HW#4
#11 Nov02	F2F	**Hands-on:** Using NodeXL to capture, analyze and visualize social media data	
#12 Nov09	Online	**Chap9:** How Social Media Has Revolutionized the Management and Recruitment of People and Supplies	HW#5
#13 Nov16	Online	**Chap10:** The Fundamental Flaw within Formal Response Systems	
#14 Nov23	Online	**Chap14:** Available Tools and Systems to Improve Emergency Management Functions	Team Project Paper
#15 Nov30	F2F	Review for Final Exam Team project presentation	Team Project Presentation
#16	F2F	**Final Exam**	



Content

Course assessment

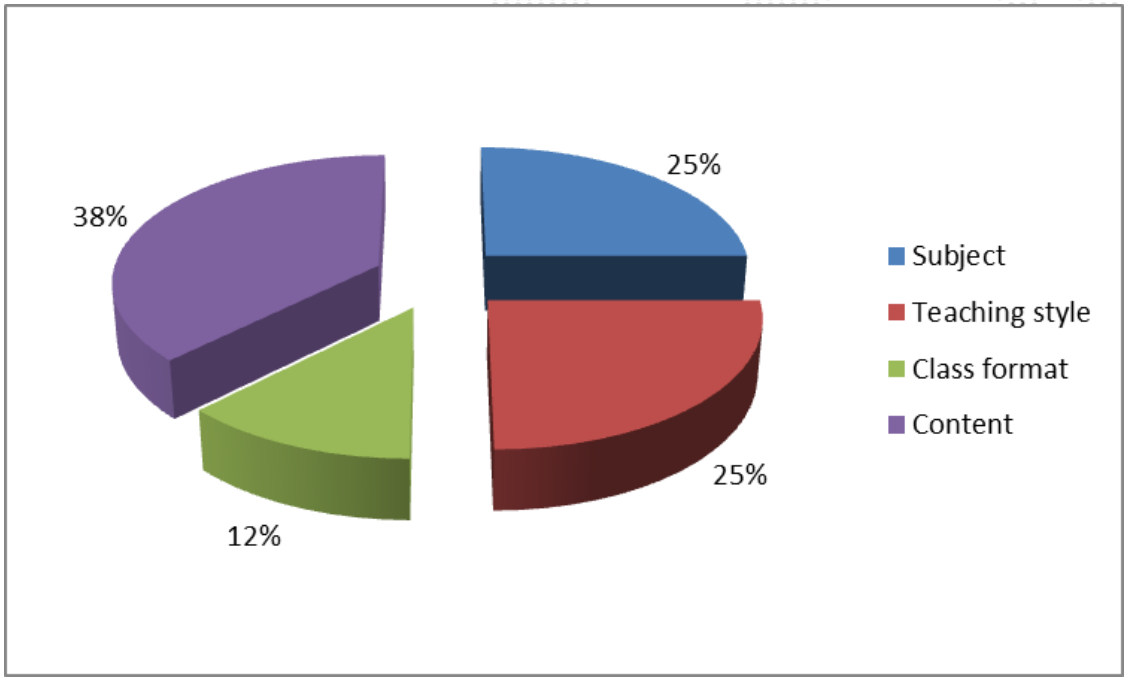
Instrument	Percentage	Points
Participation, Class Discussions & Quizzes	10%	100
Assignments	20%	200
Midterm	20%	200
Project (Team)	20%	200
Final Exam	30%	300
Total	**100%**	**1000**

Student's Evaluation

Student Evaluation



- **What Did you like the best about this class?**



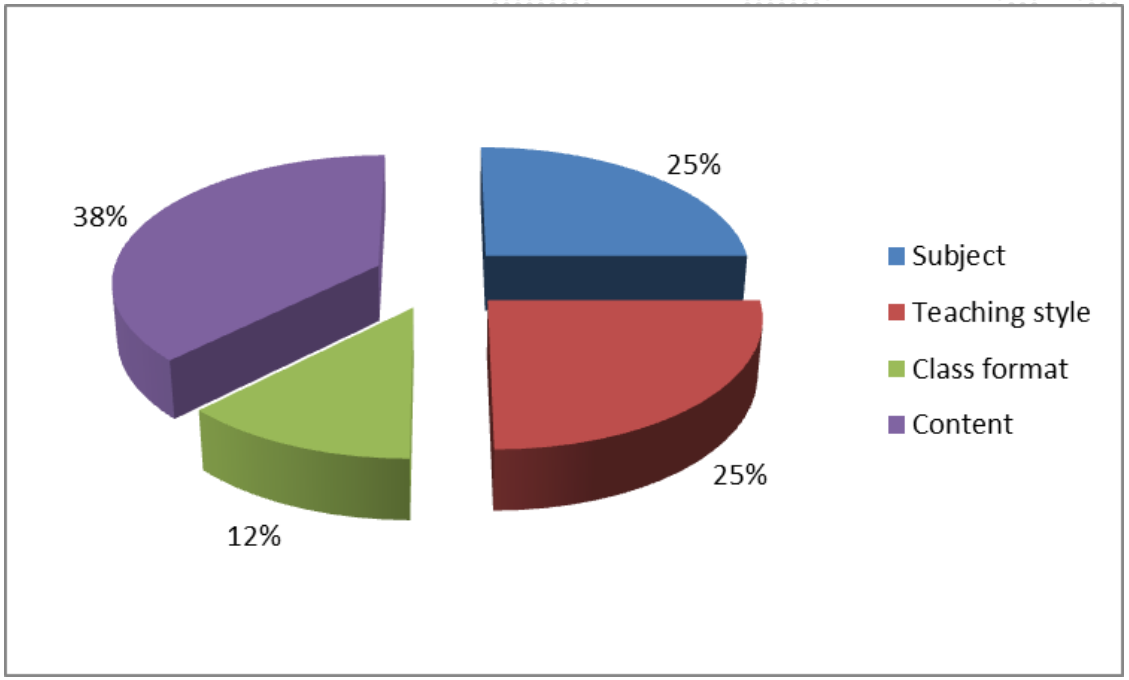
Student Evaluation



"I enjoyed all parts, the writing and the handson in NodeXL"

"New material and new software – NodeXL"

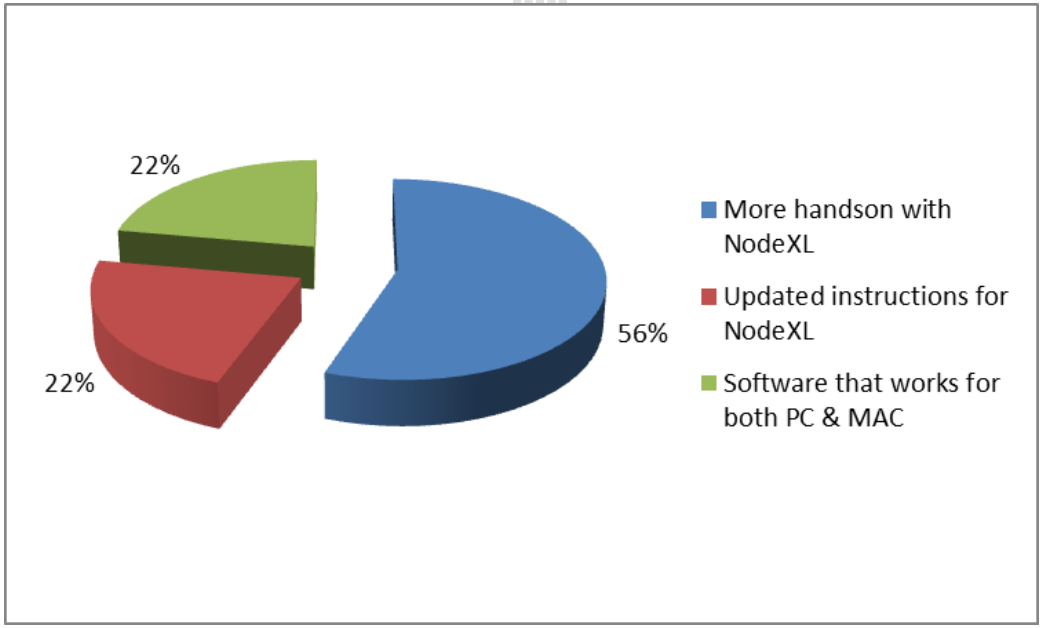
- **What Did you like the best about this class?**



Student Evaluation



- **What changes would you recommend to improve learning in this class?**



Student Evaluation

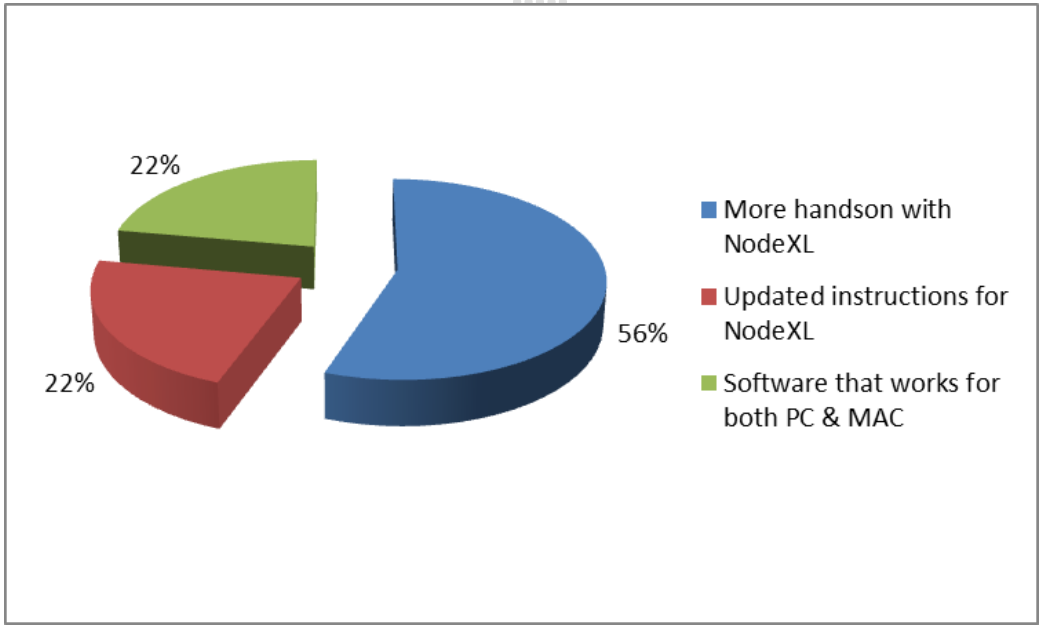


"The overall class was well. More hands-on to keep NodeXL as a focus"

"Ability to use MAC computer with NodeXL"

"Update the learning material in NodeXL"

- **What changes would you recommend to improve learning in this class?**



Some Challenges

Challenges

Designing the course



- Business School curriculum
- Hybrid
- Graduate & Undergraduate
- Appropriate text book



Challenges

Teaching the course





- **Elective course (wide rage of background)**
- **Promoting the course**
- **Leveling skills in spreadsheet**
- **Leveling skills in social network analysis**
- **Installing NodeXL in the computer lab**
- **Issues with NodeXL (http://nodexl.codeplex.com/discussions/646660)**
- **Instructor resources**

Lessons Learned

Lessons Learned





- Have adequate preparation time
  - Before launching
  - During course delivery
- Be aware about the tools specific skills required
- Have another tool ready as plan “B” (e.g. Tableau, R etc…)
- Involve colleagues in the promotion of the course.

Thank You!



Purdue University



Automatic and Manual Tattoo Localization

Joonsoo Kim[a] He Li[a] Jiaju Yue[a] Javier Ribera[a]
Edward J. Delp[a] Landis Huffman[b]

a. Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

b. The MITRE Corporation
Bedford, MA

Acknowledgement

- **This work was partially supported by the U.S. Department of Homeland Security's VACCINE Center and the MITRE Cooperation**

- **We also acknowledge the use of the NIST Tatt-C dataset**

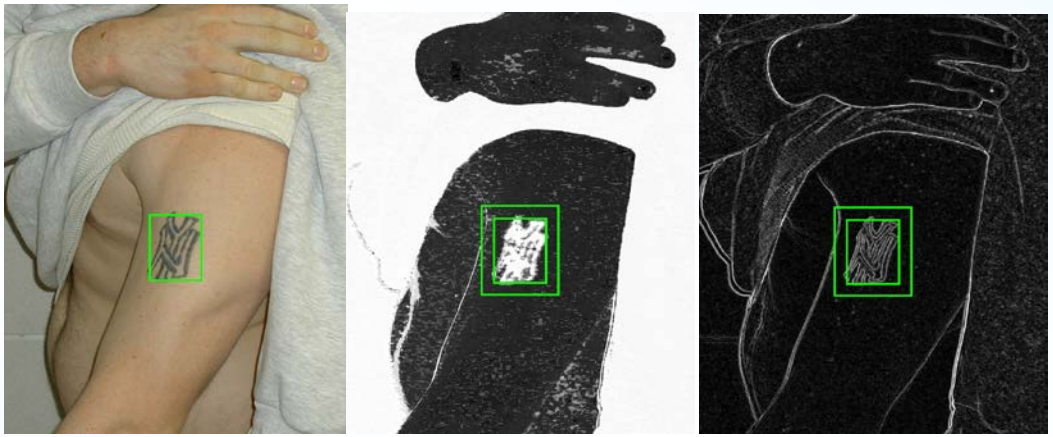
Motivation

- **In a retrieval system tattoo localization is an important step to remove background information in the tattoo image**

- **Tattoo image retrieval accuracy is highly depends on the accuracy of tattoo localization**

Center-Surround Feature Localization

Center-Surround Feature Localization



(a) (b) (c)

(a): Raw image with estimated tattoo bounding box overlaid in green
(b) and (c): Color features and edge features with optimal center-surround filter boundaries overlaid in green

- **Locate a tattoo by searching for regions containing high pixel densities of large edge features and color features between a rectangular image patch and its surrounding area**
- **Two types of features are used:**
  - **Color features: average over 9 skin detectors**
  - **Edge features: spatial gradient magnitude of the RGB pixel intensity**

Center-Surround Feature Localization

- **Color features:**

  Average over 9 skin detectors

Skin Color Domain
$0.30y_R + 0.59y_G + 0.14y_B - \max(y_R, y_G, y_B) \in [6.4, 30.0]$
$y_H \in [0°, 25°] \cup [335°, 360°]$, $y_S \in [0.20, 0.60]$, $y_V \in [0.40, 1.00]$
$y_R > 95$, $y_G > 40$, $y_B > 20$, $\max(y_R, y_G, y_B) - \min(y_R, y_G, y_B) > 15$, $y_R - y_G > 15$, $y_R > y_B$
$y_T \in [0.4, 0.6]$, $y_{\tilde{S}} \in [0.038, 0.25]$, $y_L > 80$
$y_{R'} \in [0.36, 0.465]$, $y_{G'} \in [0.28, 0.363]$, $y_H \in [0°, 50°]$, $y_S \in [0.20, 0.68]$, $y_V \in [0.35, 1.00]$
$y_H \in [0°, 50°]$, $y_S \in [0.23, 0.68]$
$y_{R'} \in [0.38, 0.55]$, $y_{G'} \in [0.25, 0.38]$, $y_H \in [0°, 50°] \cup [340°, 360°]$, $y_S \in [0.20, 1.00]$, $y_V \in [0.35, 1.00]$
$y_{R'} \in [0.38, 0.55]$, $y_{G'} \in [0.25, 0.38]$, $y_H \in [0°, 50°] \cup [340°, 360°]$, $y_S \in [0.20, 1.00]$, $y_V \in [0.35, 1.00]$, $y_T \in [0.4, 0.6]$, $y_{\tilde{S}} \in [0.038, 0.25]$, $y_L > 80$
$y_H \in [0°, 50°] \cup [340°, 360°]$, $y_S \in [0.20, 1.00]$, $y_V \in [0.35, 1.00]$

9 skin detectors

$$f_c(m,n) = 1 - \frac{1}{9}\sum_{j=1}^{9} I_j(y(m,n))$$

$f_c(m,n)$ **: the color feature**

$y(m,n)$ **: the pixel color data**

with *m* and *n as* the spatial coordinates

Center-Surround Feature Localization

- **Edge features:**

  Spatial gradient magnitude of the RGB pixel intensity

$$f_e(m,n) \underline{\underline{\Delta}} \left| \nabla \left| y(m,n) \right| \right|$$

$f_e(m,n)$ **: the edge feature**

$y(m,n)$ **: the pixel color data with *m* and *n* as the spatial coordinates**

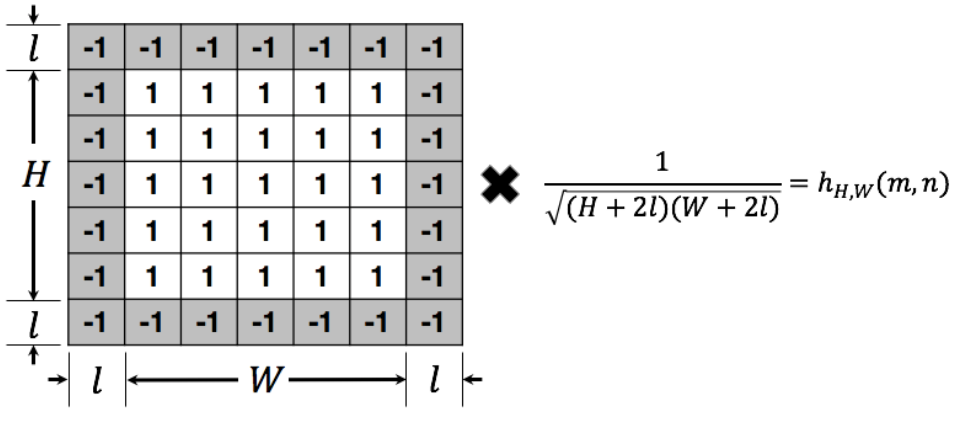
Center-Surround Feature Localization

- **Bounding box selection**

$$h_{H,W}(m,n) \triangleq \begin{cases} \frac{1}{Z}, & |m| \le \frac{H}{2} \text{ and } |n| \le \frac{W}{2} \\ -\frac{1}{Z}, & |m| \in \left(\frac{H}{2}, \frac{H}{2}+l\right] \text{ or } |n| \in \left(\frac{W}{2}, \frac{W}{2}+l\right] \\ 0, & \text{otherwise,} \end{cases}$$

$$Z = \sqrt{(H+2l)(W+2l)}$$

$$l = \left\lfloor \frac{1}{2}\sqrt{H^2+W^2+6HW} - \frac{1}{2}(H+W) \right\rfloor$$



Impulse response of center-surround filter

$$g_c(m,n;H,W) \triangleq h_{H,W} * f_c(m,n)$$

$$g_e(m,n;H,W) \triangleq h_{H,W} * f_e(m,n)$$

$$(\hat{m},\hat{n},\hat{H},\hat{W}) = \underset{m,n,H,W}{\operatorname{argmax}} \left[g_c(m,n;H,W)g_e(m,n;H,W)\right]$$

H **and** ***W*** **are the height and width of the tattoo image,** ***Z*** **is the normalization factor,** ***l*** **is the width of the surround region,** ***g*** **is the filter response,** $(\hat{m},\hat{n},\hat{H},\hat{W})$ **is the estimated bounding box**

Graph-Cut Tattoo Localization

Graph-Cut Tattoo Localization

- **Two important assumptions**
  - **Tattoos are surrounded by skin**
  - **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**

Graph-Cut Tattoo Localization

Overview



(a) Original Image



(b) Possible Segmentation Regions



(c) Graph-Cut Segmentation



(d) BBR



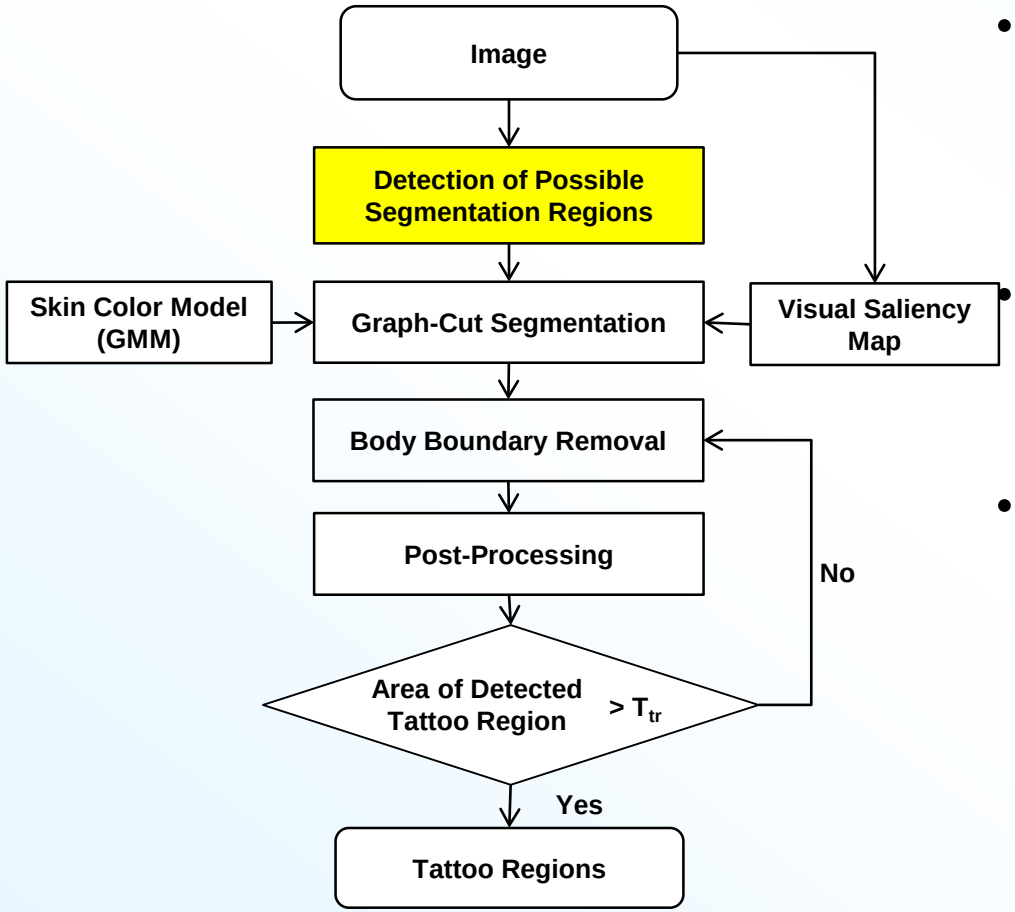
(e) Tattoo Detection



(f) Tattoo Localization

Graph-Cut Tattoo Localization

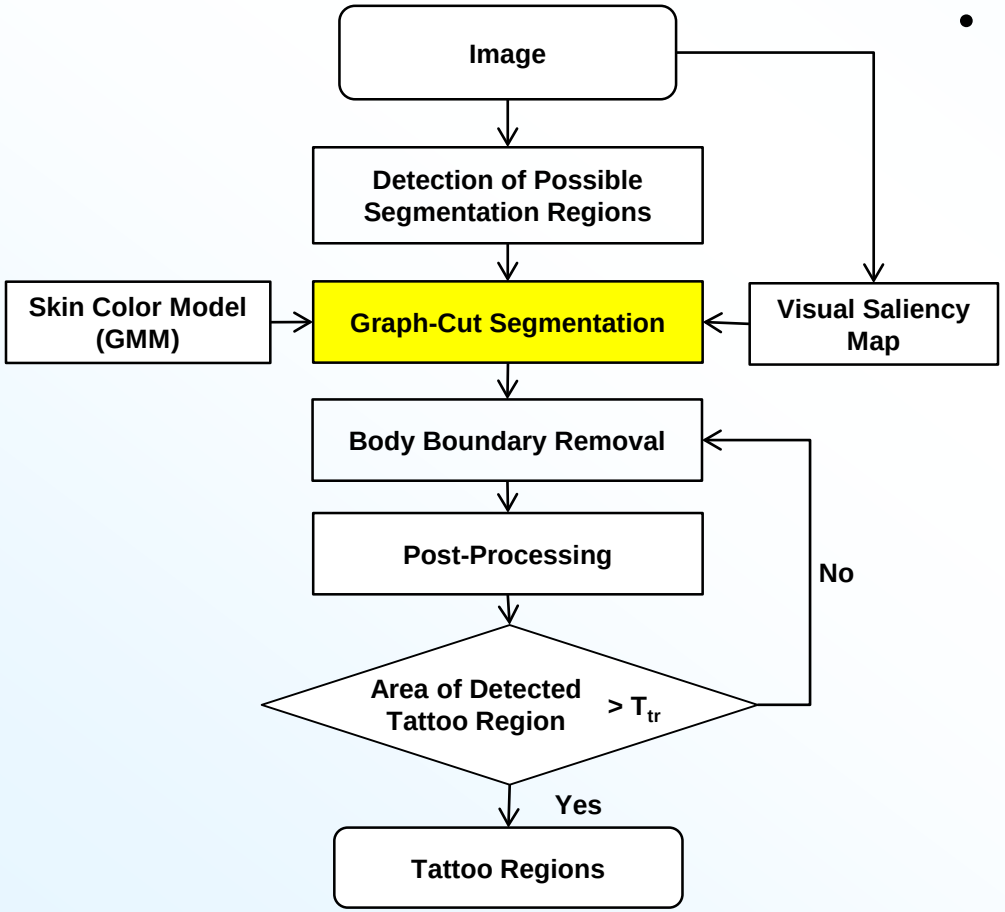
Possible Segmentation Region Detection



- 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

Graph-Cut Tattoo Localization

Graph-Cut 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

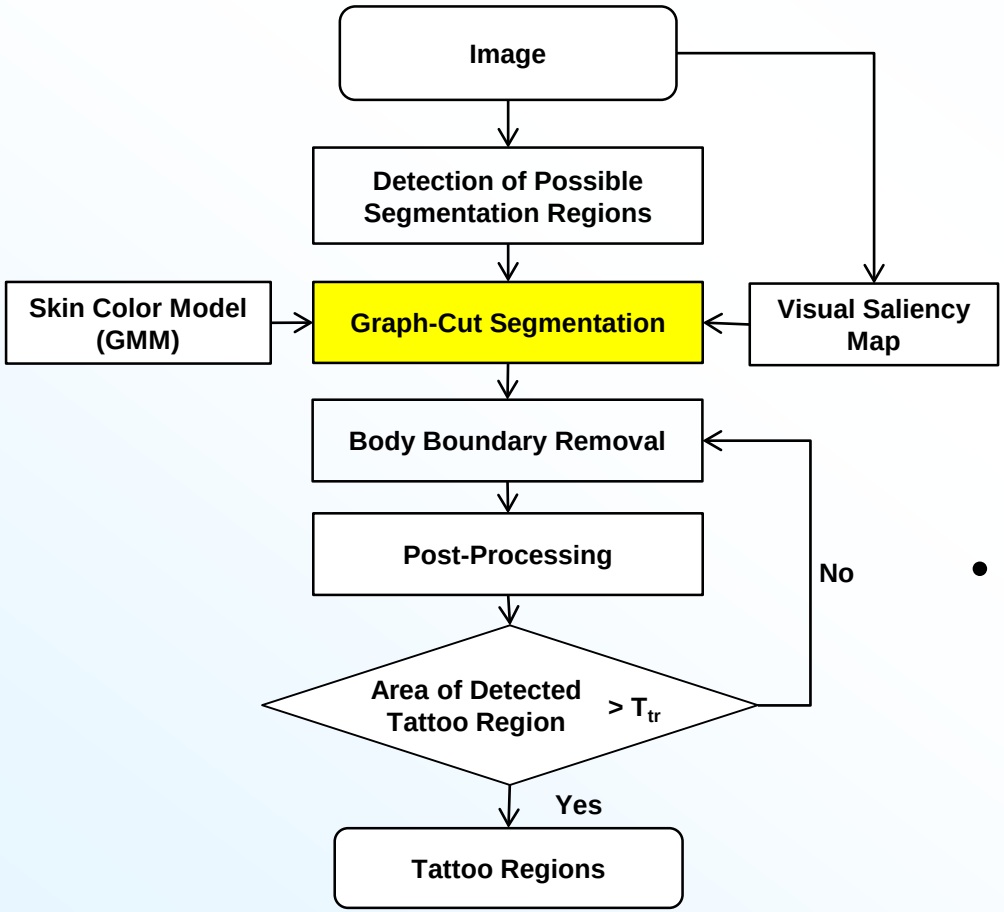
I : image

N_i : neighborhood pixels of pixel I

$x_i \in$ { 0(background), 1(foreground) }

Graph-Cut Tattoo Localization

Graph-Cut 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

- **$V(x_i,x_j)$: the smoothness term**

$$V(x_i, x_j) = |x_i - x_j| * f(C_{ij})$$

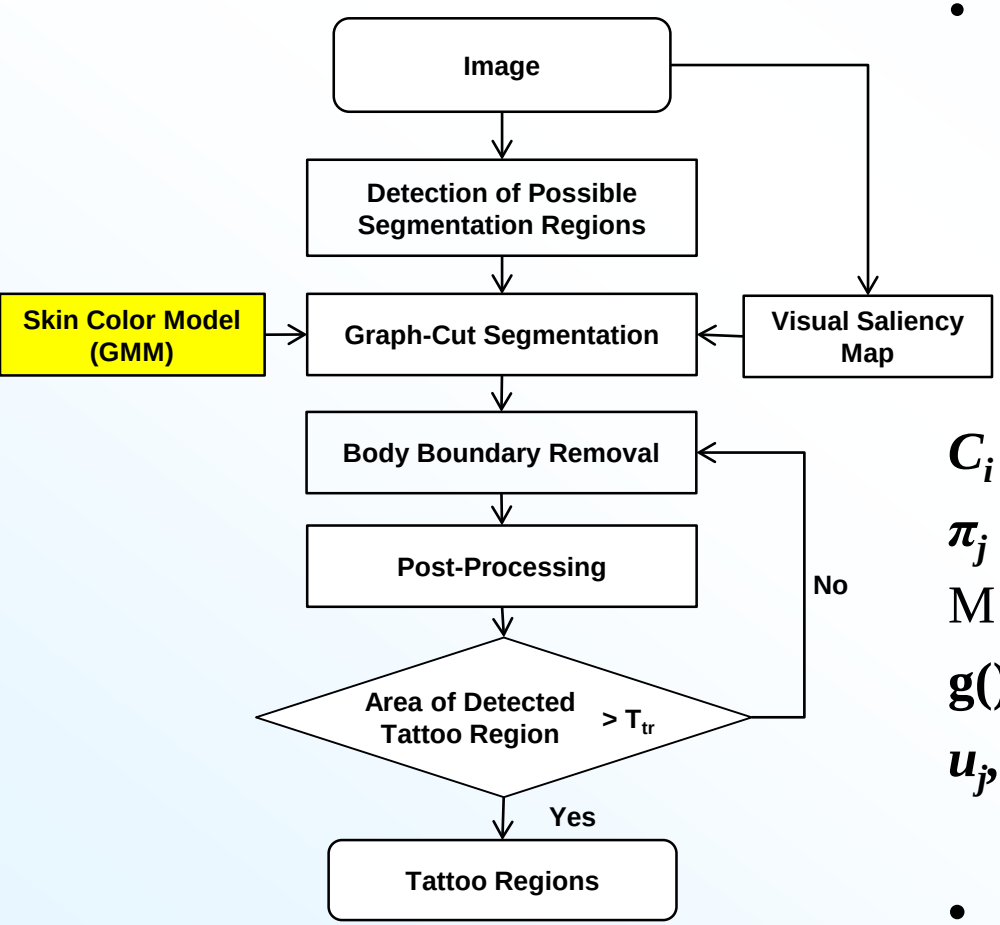
$f(\varepsilon) = 1/(1+\varepsilon)$,

$C_{ij} = ||C_i - C_j||^2$

C_i : YCbCr color of pixel i

Graph-Cut Tattoo Localization

Skin Color Model



- **GMM skin model - train on diverse skin tones**

$$D_1(x_i = 1) = -log(p_1(x_i = 1))$$

$D_1(x_i)$: energy for skin/non-skin

$$p(x_i = 1) = \sum_{j=1}^{M} \pi_j * g(C_i | u_j, \Sigma_j)$$

***C_i* : YCbCr color of *i^{th}* pixel**

***π_j* : *j^{th}* mixture weight**

M : number of mixture components

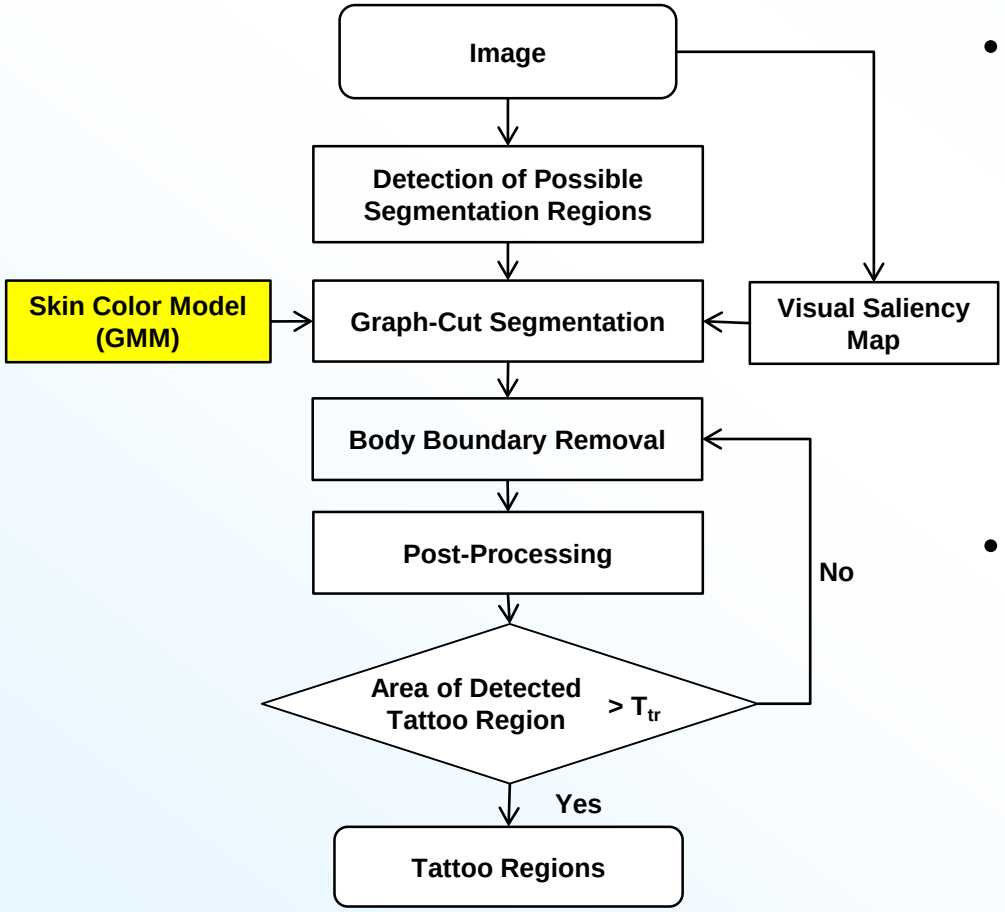
g() : Gaussian density function

***u_j*, *Σ_j* : Gaussian parameters**

- ***π_j* *u_j*, *Σ_j* are estimated using EM**

Graph-Cut Tattoo Localization

Skin Color Model



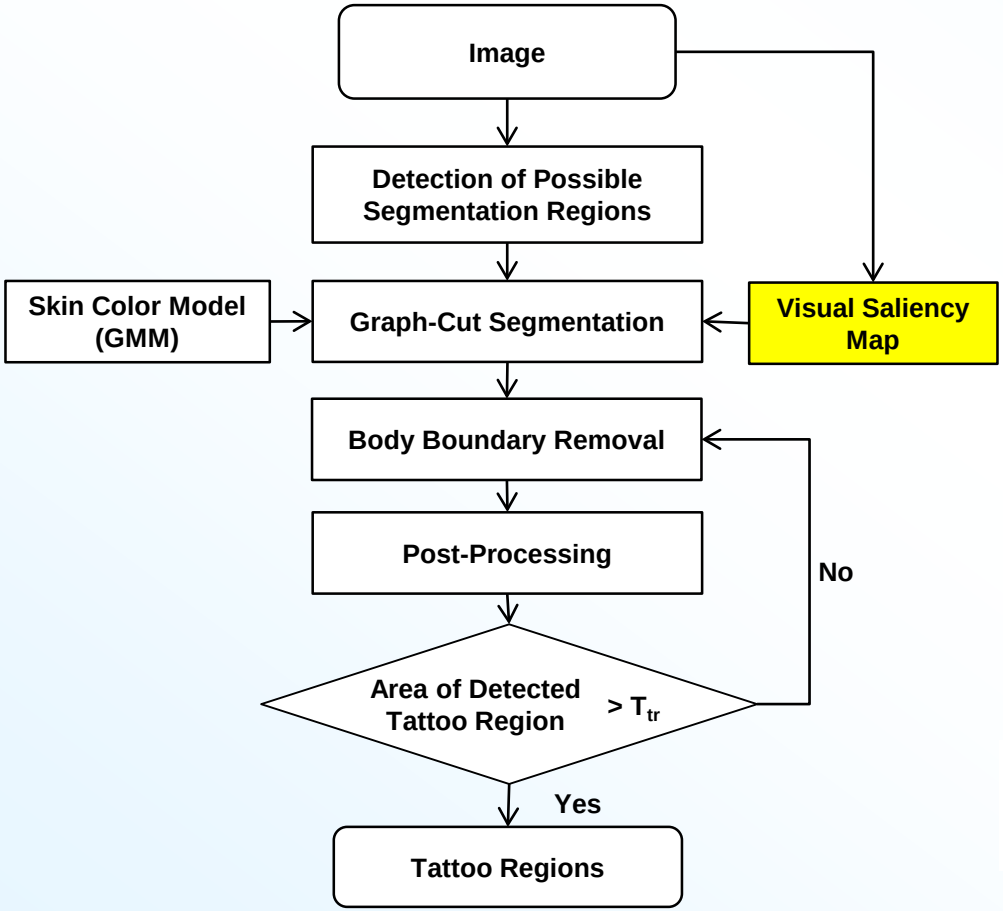
- **Cannot estimate all non-skin colors - use adaptive threshold for non-skin model**

$$D_1(x_i = 0) = -log(wt_1 * \frac{1}{n_I} \sum_{i \in I} p_1(x_i = 1))$$

- **Adaptive threshold helps to segment skin when the skin colors in an image do not fit the skin model**

Graph-Cut Tattoo Localization

Saliency



- **Case for color similar to skin in background, use visual saliency map to focus on skin near tattoo**

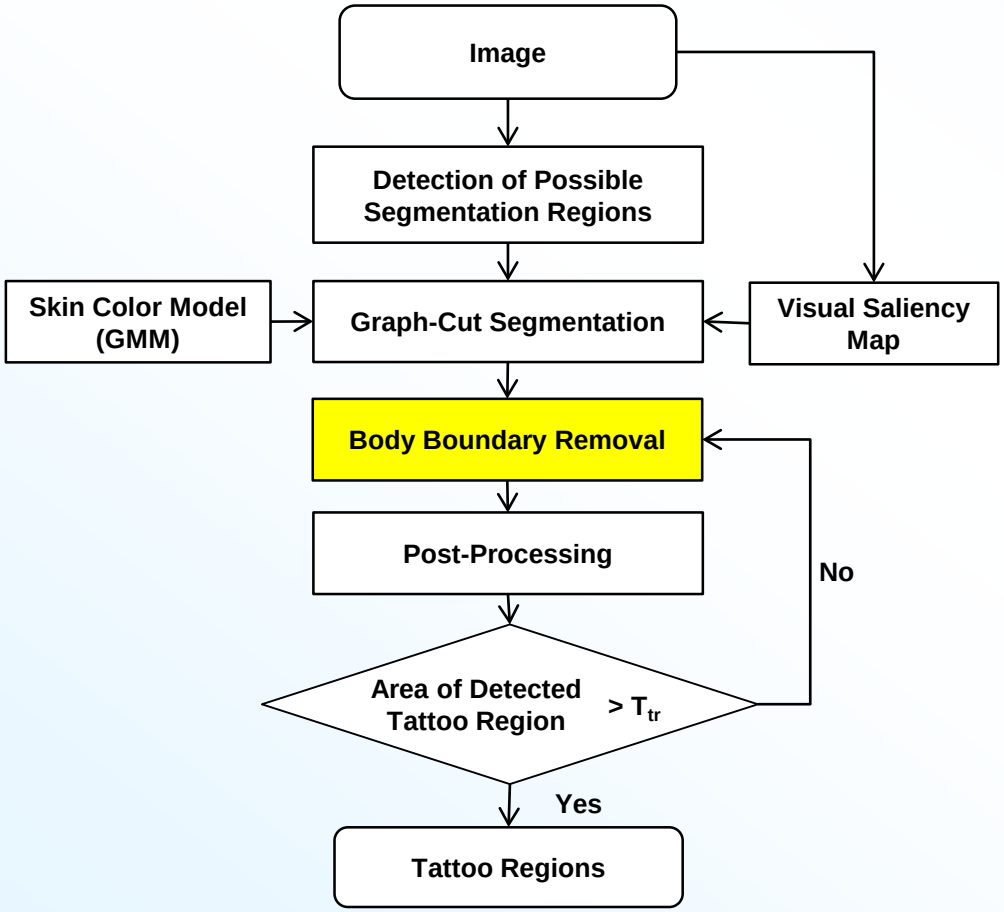
$$D_2(x_i = 1) = -log(p_2(x_i = 1))$$

- **$D_2(x_i)$: energy for normalized saliency map**
- **Normalized GBVS saliency map is used for $p_2(x_i=1)$**
- **Adaptive threshold is used for p2(xi=0)**

$$D_2(x_i = 0) = -log(wt_2 * \frac{1}{n_I} \sum_{i \in I} p_2(x_i = 1))$$

Graph-Cut Tattoo Localization

Body Boundary Removal



- Our previous tattoo segmentation method finds tattoo regions incorrectly when tattoo regions are very close to boundaries of the body
- We observe that the regions near a body boundary mostly have simpler textures than the regions near the boundary of a tattoo
- Our BBR method remove a pixel that has simple textures in a window region centered on the pixel
- Histogram of the gradient orientation, weighted by gradient magnitude, was used to measure the degree of texture complexity

Graph-Cut Tattoo Localization

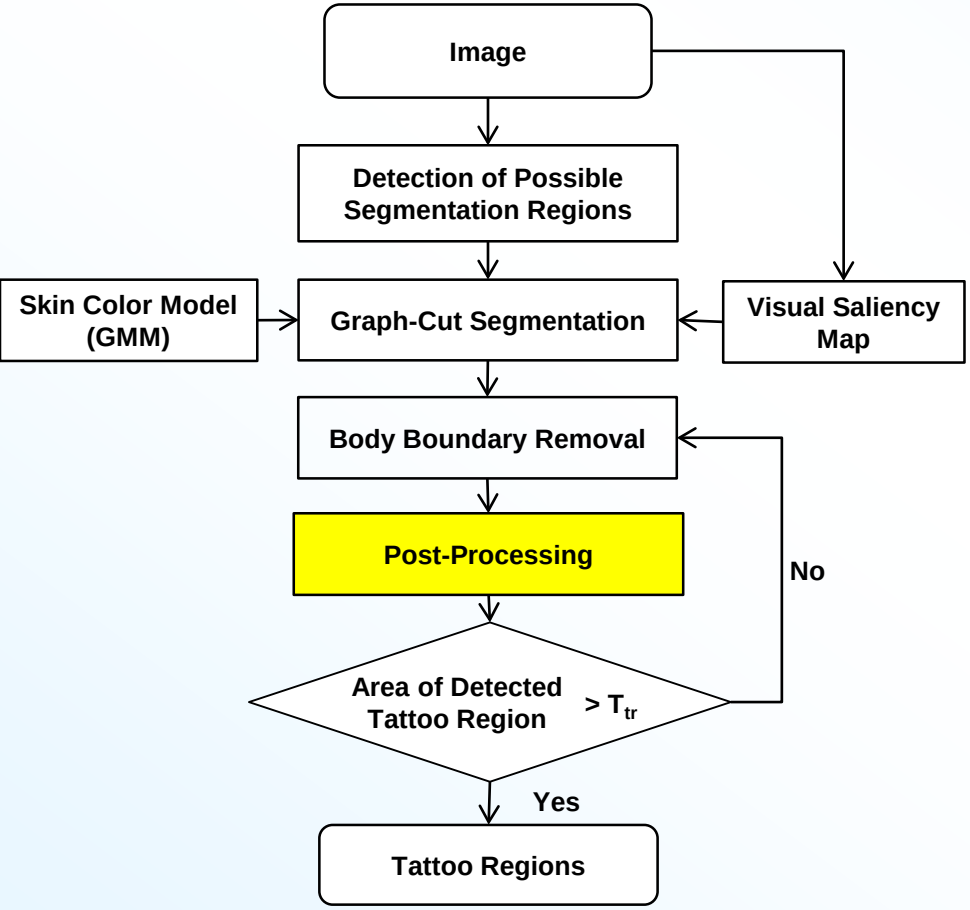
Body Boundary Removal

1. Initialize t_{st}
2. A moving window, W_i is centered on the pixel i resulting from graph-cut segmentation
3. Compute histogram of the orientation of the gradients weighted by gradient magnitudes ($h_i(m)$) for W_i.
4. If following condition is satisfied, then the pixel i is removed from the segmented pixels
   : $\max(h_i(m_{max}) + h_i(m_{max} + j)) > t_{st}$
   where $j=-1, +1$ and $m_{max} = \text{argmax}_m\ h_i(m)$

5. Iterate step 2 ~4 for all the segmented pixels
6. Post-Processing
7. If the detected tattoo region is smaller than T_{tr}, reduce t_{st} and go back to step 2
   Otherwise finish tattoo segmentation

Graph-Cut Tattoo Localization

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 in a 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

Manual Tattoo Localization

Manual Tattoo Localization



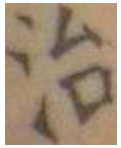
(a) Window



(b) Multi-Bounding Boxes



(c) Crop 1



(d) Crop2

- **The cropping tool is a supportive software developed by VIPER lab to help the user to crop sub images from original image**
- **User can crop the sub image by drawing rectangle to select the target area**
- **User can crop out multiple sub images at one time and save the sub images to local machine**

Experimental Results

- **Experiment Overview**
  - **Goal : Find minimum box regions including tattoos in an image instead of segmenting tattoo objects to evaluate our segmentation methods quantitatively**
  - **Tattoo dataset**
    - **Dataset 1 : 6308 images from the NIST Tattoo Challenge dataset (Tatt-C Dataset)**
    - **Dataset 2 : 1105 images from the eviltattoo.com**
  - **Ground Truth : tattoo regions (minimum box regions including tattoos) are manually cropped using our image cropping tool (manual tattoo localization)**
  - **Comparison : Our previous method [1], CSFL, Graph-cut Method**
  - **Two evaluation metrics : Recall and Accuracy**

Experimental Results

	Recall		Accuracy	
	Tatt-C	Evil Tatto	Tatt-C	Evil Tattoo
[1]	36.16%	63.44%	69.75%	69.30%
Method 1	25.18%	42.90%	66.20%	63.26%
Method 2	41.25%	66.97%	70.46%	69.91%

Method 1 : Center-Surround Feature Localization(CSFL)
Method 2 : Graph-Cut Localization

$$\text{Accuracy} = \frac{n_{OF} + n_{OB}}{n_T} \qquad \text{Recall} = \frac{n_{OF}}{n_{GDF}}$$



: minimum bounding box in automatic detected segments

: minimum bounding box in ground truth

- n_{OF} : the number of overlapped pixels in the minimum bounding boxes including tattoo between automatic detected segments and ground truth
- n_{OB} : the number of overlapped pixels outside of the minimum bounding boxes between automatic detected segments and ground truth
- n_T : the total number of pixels in a tattoo image
- n_{GDF} : the number of pixels in the minimum bounding boxes including tattoo in a ground truth image

[1]. J. Kim, A. Parra, H. Li, and E. J. Delp, "Efficient graph-cut tattoo segmentation," Proceedings of the SPIE/IS&T Conference on Visual Information Processing and Communication VI, vol. 9410, pp. 94100H–1–8, February 2015, San Francisco, CA

Experiment Results





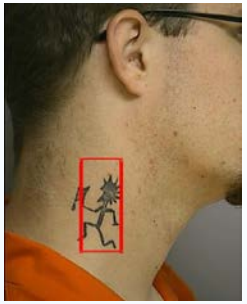

Ground truth | [1] | CSFL Method | Graph-Cut Method






Ground truth | [1] | CSFL Method | Graph-Cut Method

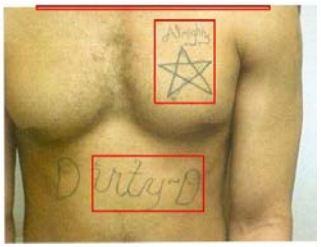
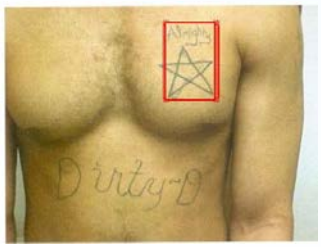




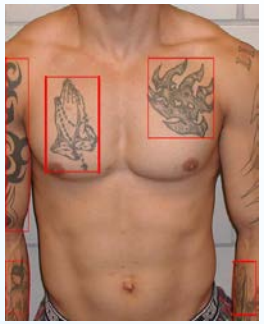
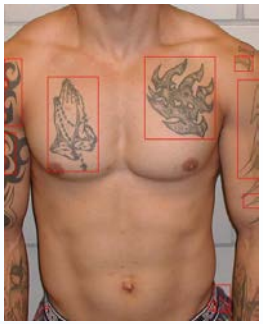

Ground truth | [1] | CSFL Method | Graph-Cut Method

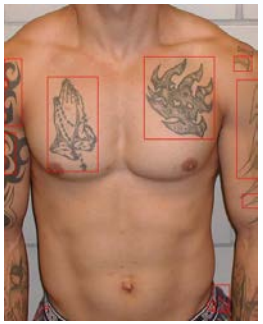
- **Both Methods have good accuracies when tattoo edges are strong**
- **Graph-cut method has better localization performance when tattoo edges are weaker than edges of background clutters (clothes, hair)**

Experiment Results


Ground truth | [1] | CSFL Method | Graph-Cut Method



Ground truth | [1] | CSFL Method | Graph-Cut Method

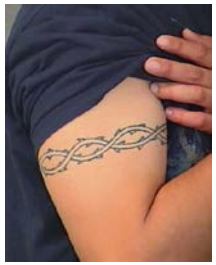





Ground truth | [1] | CSFL Method | Graph-Cut Method

- **Graph-Cut method has the ability to separately localize multiple tattoo symbols in the image while CSFL method does not have it**

Experiment Result



Ground truth [1] CSFL Method Graph-Cut Method






Ground truth [1] CSFL Method Graph-Cut Method






Ground truth [1] CSFL Method Graph-Cut Method

- **Graph-Cut method has better accuracy than [1] when a tattoo region is close to body boundaries**
- **CSFL method is not affected in above case**

Conclusion and Future Work

- **Graph-Cut achieves better performance than CSFL**
- **Graph-Cut sometimes fails to extract tattoo regions correctly when there are tattoo regions close to boundaries of the body and/or near “hair regions”**
- **We will investigate techniques to remove the “hair regions”**

Automatic and Manual Tattoo Localization

Joonsoo Kim[a] He Li[a] Jiaju Yue[a] Javier Ribera[a]
Edward J. Delp[a] Landis Huffman[b]

a. Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

b. The MITRE Corporation
Bedford, MA

City of Chicago Long-Term Evolution (LTE) Broadband Pilot

Khalid Tahboub and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

We acknowledge support from the Office of Emergency Management and Communications (OEMC), the City of Chicago Police Department, the U.S. Department of Homeland Security, the DHS Center of Excellence and Motorola

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

CITY OF CHICAGO – INCORPORATED 4th MARCH 1837 – URBS IN HORTO

MOTOROLA SOLUTIONS

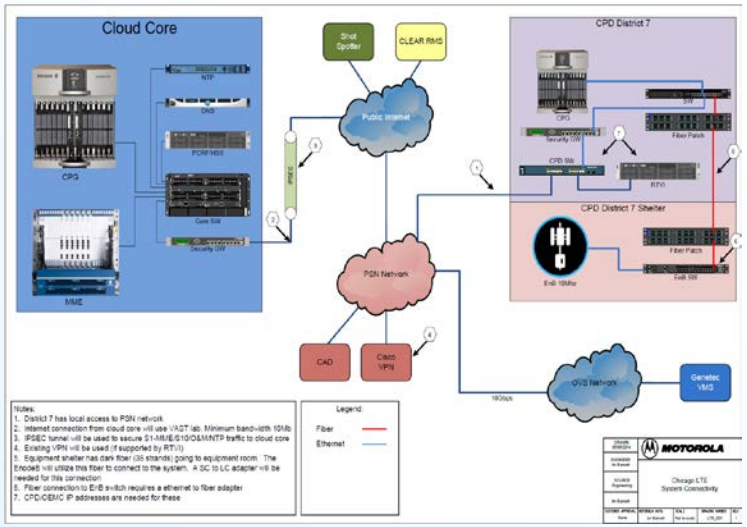
PURDUE UNIVERSITY

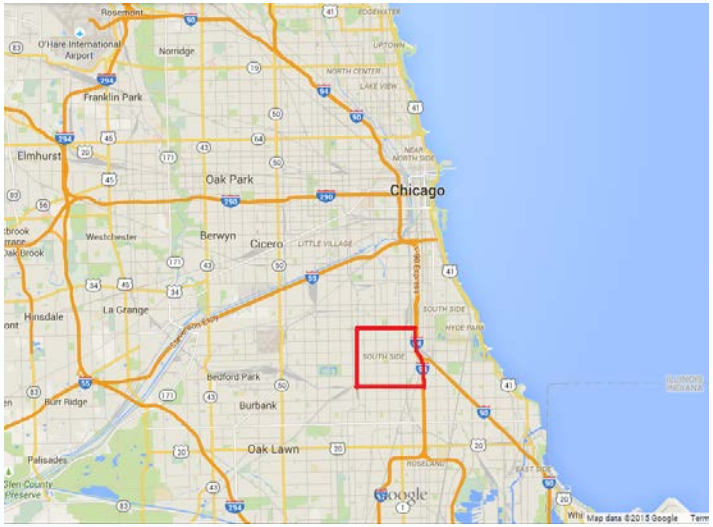
Motivation

- **The Federal Communications Commission (FCC) identified 3GPP Long-Term Evolution (LTE) as the network technology for the National Public Safety Broadband Network (NPSBN)**
- **A pilot LTE network consisting of one cell was deployed by Motorola to provide LTE coverage in the Chicago Police Department (CPD)'s seventh district**
- **Our objective is test the viability and performance of the LTE NPSBN with respect to transport of video imagery**

System Architecture

- **The high-level network architecture of LTE is comprised of two main parts:**
  - **The Evolved Packet Core (EPC)**
  - **The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN)**





Testing Tools



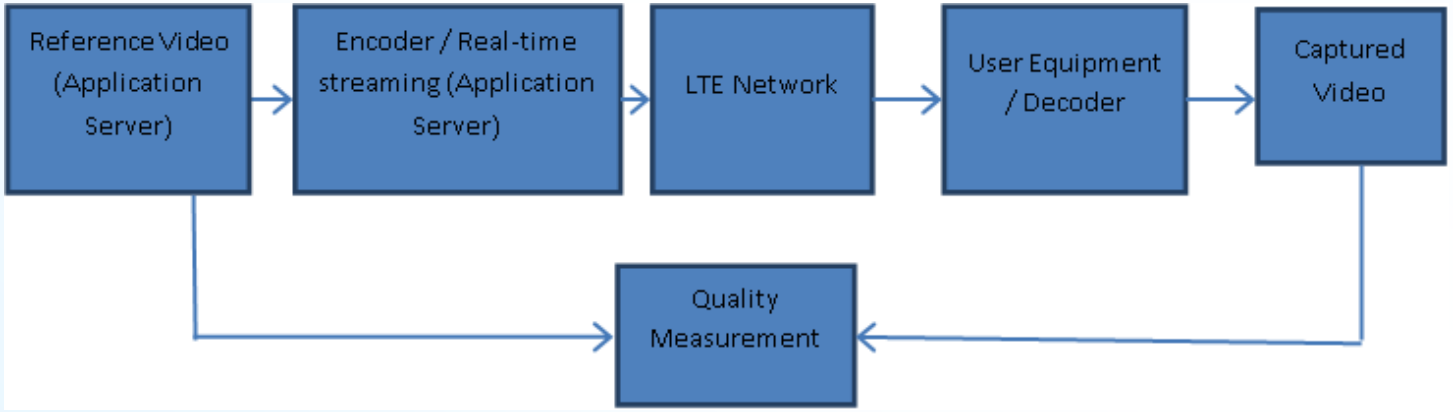
1. **LTE USB dongles**
2. **Police Laptops equipped with RTVI application**
3. **Application server used to host some of the services required for testing purposes such an FTP server and video streaming server**
4. **FFmpeg, iPerf softwares and standard laptops**
5. **Video database**

   The Public Safety Communication Research (PSCR) video samples with applications in public safety shared in the Consumer Digital Video Library (CDVL)

Test Plan

The test procedure consisted of three parts:

1. **Objective perceptual video quality tests designed to measure the video quality when video is streamed in real-time over the LTE network**



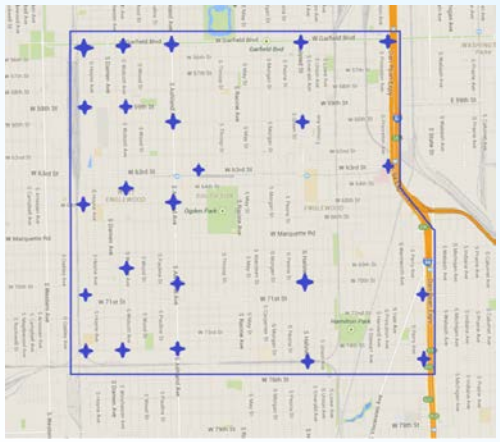
Test Plan

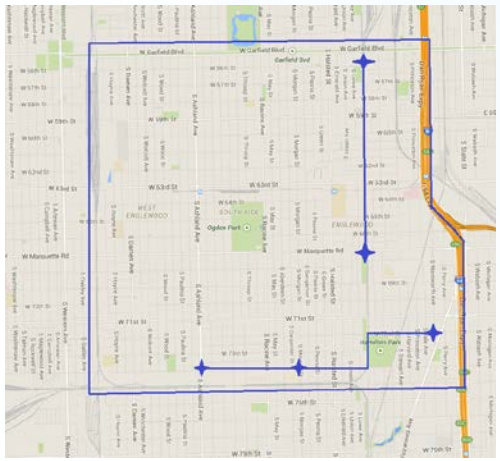
2. **Subjective measurements designed to characterize the performance of the Real Time Video Intelligence (RTVI) under various test conditions, performance metrics:**
   - **Clarity of the picture**
   - **Fluidity of the motion**
3. **Network performance measurements to test the key performance indicators associated with the network:**
   - **Throughputs**
   - **Delays**
   - **Signal strength**
   - **Attachment success rate**

Test Plan

The test procedure included various testing conditions:

1. **Multiple video resolutions and frame rates**
2. **Various levels of network traffic and multi-user scenarios**
3. **Static/dynamic conditions**
4. **Uplink/downlink testing**



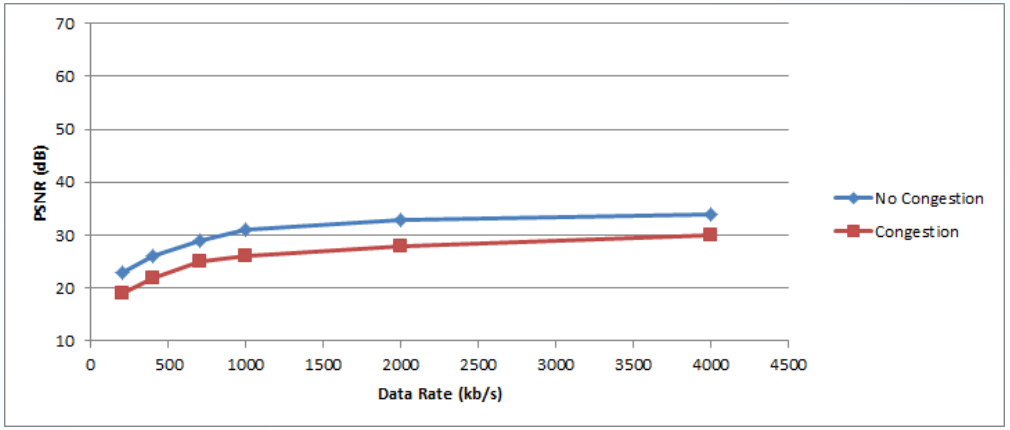


Test Results

1. **Objective Video Quality Measurement**

   Introducing congestion in the network or adding more users when a connection is established from the cell edge degrades the streamed video quality

   Example:



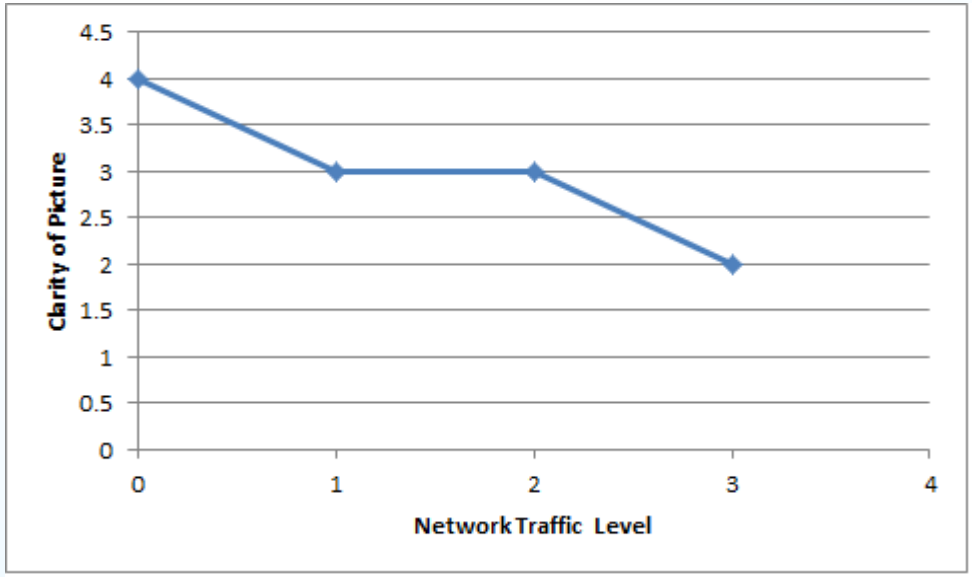
   Observations and results are discussed in the final report

Test Results

2. **Subjective Quality Assessment of CPD Applications**

   RTVI clarity of picture along the cell edges are impacted by network traffic

   Example:

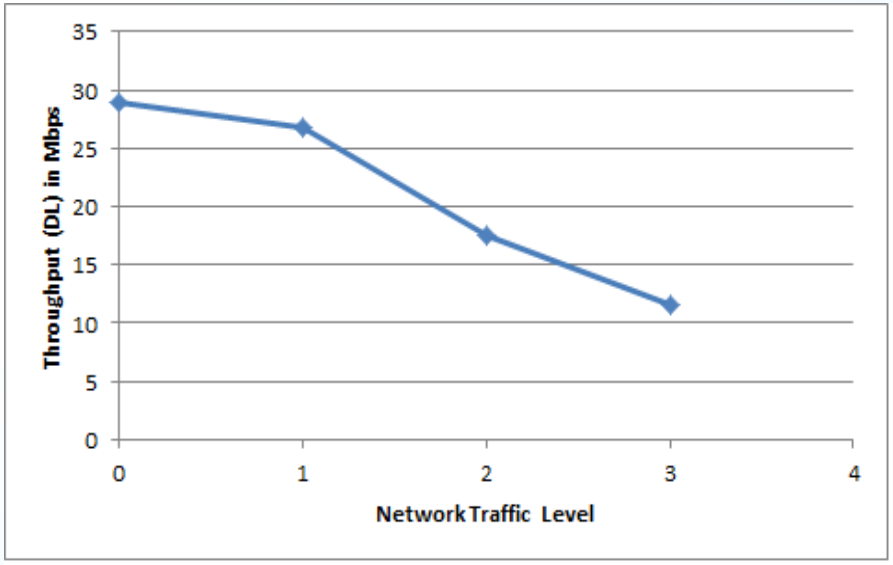


Test Results

3. **Network Performance Measurement**

   Impact of network congestion close to the cell tower

   Example:



   Observations and results are discussed in the final report

Conclusions and Recommendations

1. **The 700MHz National Public Safety Broadband Network (NPSBN) will provide unprecedented resources in terms of network capacity and communication speeds**
2. **Quality of Service (QoS) and prioritization should be taken into consideration**
3. **Careful RF planning is required**
4. **The number of IP cameras connected to the network using an LTE connection should be properly determined**
5. **Video quality requirements should be based on the task**

Future Work

1. **Objective video quality metric with applications in video analytics in public safety**

2. **Quality of service (QoS) and prioritization**

3. **Proximity services and device to device communication**

City of Chicago Long-Term Evolution (LTE) Broadband Pilot

Khalid Tahboub and Edward J. Delp

Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

Superpixels Shape Analysis For Carried Object Detection

Blanca Delgado, Khalid Tahboub, and Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

Introduction

- **Re-identification of subjects in video surveillance systems plays a significant role in public safety**

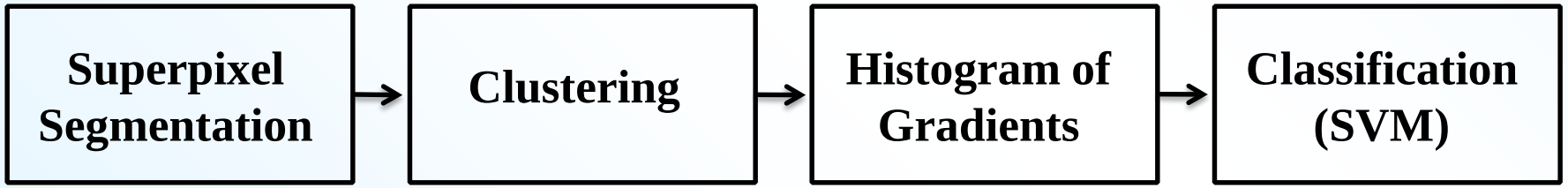
- **Atttributes based re-identification methods can help address this problem by incorporating semantic attributes such as carried objects**

Objective

- **Use a supervised approach based on the contours and shapes of superpixels and histogram of oriented gradients in order to detect attributes with applications to carried objects detection**

Proposed Method – Block Diagram

- **Superpixel Segmentation + Clustering**
- **Histogram of Gradients**
- **Classification – Support Vector Machine**



Proposed Method - Segmentation

- **Superpixel Segmentation:**
  - **Generated by clustering pixels based on their color similarity (LAB Color Space - Perceptually Uniform) and proximity in the image plane**
  - **Enforce regular and smooth shapes, removing smaller superpixels by using morphological operations**

Proposed Method - Clustering

- **DBSCAN (Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise)**
  - **Forming clusters of superpixels that leads to the final segmentation**

Segmentation Results



Figure 1. Segmentation and Clustering results for no carried objects

Segmentation Results



Figure 2. Segmentation and Clustering results for carried objects

Feature Vector

- **Histogram of gradients (HOG)**
  - **Compute gradient on windows of 8x8 pixels**
  - **Magnitude (normalized to 1) + Phase (0^o - 180^o)**
  - **Histogram with bins of width 20**
  - **Contribution to the histogram given by the magnitude of the vector (will always be 1)**
  - **Split contribution between the two closest bins**

Feature Vector

- **The location of carried objects might exist to the left side or the right side of the human subject**
- **To account for this variation, each image is reversed across the vertical axis at the image midpoint**
- **The descriptor consists of the concatenation of the histograms of orientated gradients over each region at two different resolution levels**

Feature Vector

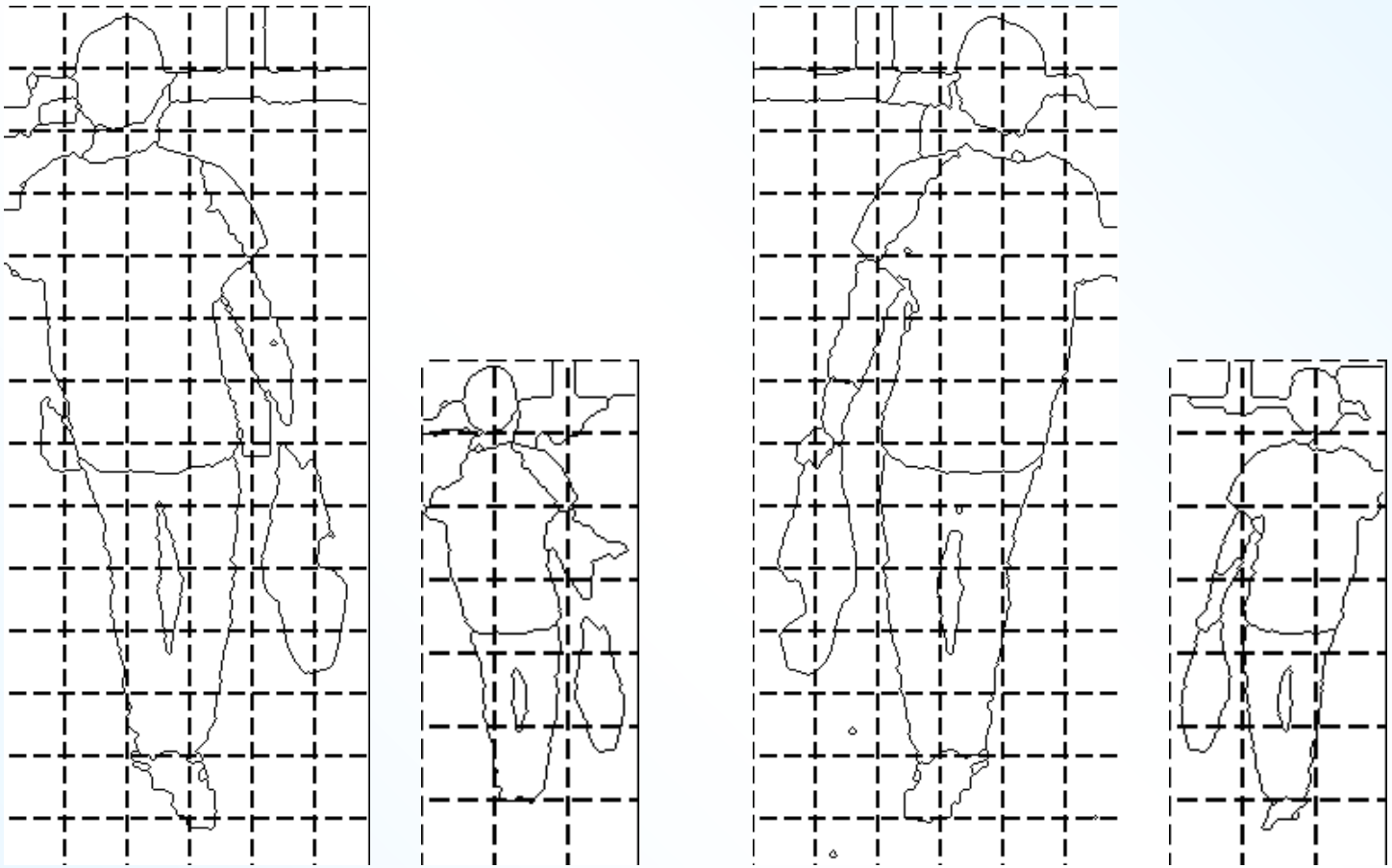


Figure 3. Image Representation

Classification

- **Supervised Classification - Support Vector Machine**
  - **Define two classes:**
    - **Carried Object**
    - **No Carried Object**

- **We will look for the *maximum-margin hyperplane* that divides a group of samples from one class to a group of samples from another class**

Experimental Results - Dataset

- **Experimental evaluation was conducted using a dataset recorded at the Greater Cleveland Regional Transit Authority (GCRTA)**

- **Data was collected over several days at various times**

- **Videos were recorded in HD (1280 × 1024) and SIF resolutions (352 × 240)**

Experimental Dataset





Figure 4. Dataset snapshots



Figure 5. Low lateral bag, shoulder bag, backpack, no carried object

Experimental Results - Training Dataset

Type of Carried Object	Number of Training Images		
	Positives	Negatives	Total
Low Lateral Bag	75	112	187
Shoulder Bag	39	112	149
Shoulder Bag	35	51	86

Experimental Results - Testing Dataset

Type of Carried Object	Number of Testing Images		
	Positives	Negatives	Total
Low Lateral Bag	29	152	181
Shoulder Bag	11	138	149
Shoulder Bag	12	21	33

Experimental Results

- **A true positive (TP) is a correctly identified event, such as a person carrying a backpack**

- **A false positive (FP) is an incorrectly identified event and a false negative (FN) is a missed detection.**

$$\text{Precision} = TP/(TP+FP)$$

$$\text{Recall} = TP/(TP+FN)$$

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

Experimental Results

Type of Carried Object	Precision (%)	Recall (%)	Accuracy (%)
Low Lateral Bag	88	81.48	95.53
Shoulder Bag	100	81.81	98.66
Shoulder Bag	90.90	83.33	90.90

Conclusions

- **Experimental results demonstrate that our proposed method is able to detect carried objects**

- **The precision is constantly higher than the recall value due to the fact that we have more false negatives than false positives**

Future Work

- **As future work we plan to:**
  - **Add texture features in our segmentation model**
  - **Enlarge the dataset and the set of experiments**
  - **Extend method to other semantic attributes such color clothing, presence of hats, …**
  - **Incorporate temporal information following a multi-shot approach**













We acknowledge support from the U.S. Department of Homeland Security and the DHS Centers of Excellence



Superpixels Shape Analysis For Carried Object Detection

Blanca Delgado, Khalid Tahboub, and Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana, USA

Tattoo Image Retrieval For Region of Interest

Joonsoo Kim, He Li, Jiaju Yue, Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

Acknowledgement

•**This work was supported by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001**

•**We also acknowledge the use of the NIST Tatt-C dataset**

NIST Tattoo Recognition Technology-Challenge (Tatt-C)

- **The NIST Tattoo Recognition Technology-Challenge (Tatt-C) : a competition to test image-based tattoo matching technology**

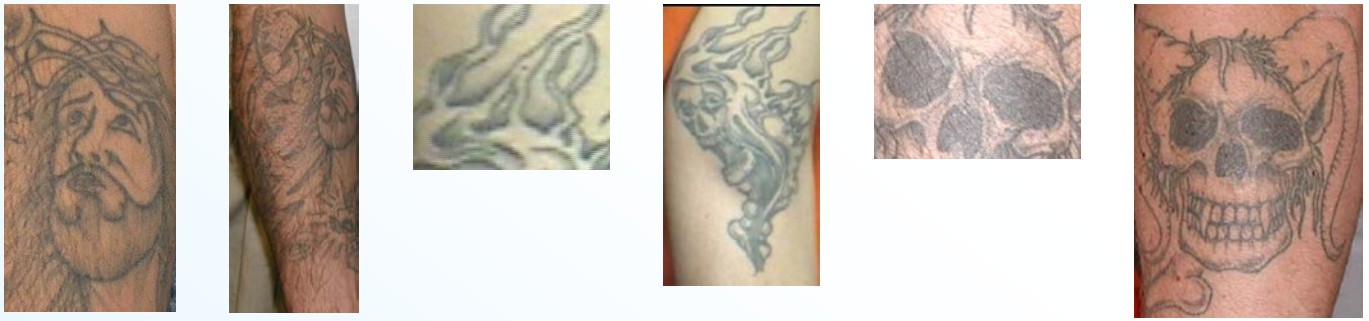
- **This challenge assessed the capability of tattoo recognition methods to detect and retrieve tattoos**

- **The tatt-c [1] database consists of five different datasets focused on five primary use cases: tattoo identification, region of interest, mixed media, tattoo similarity, and tattoo detection**

[1]. M. Ngan and P. Grother, "Tattoo recognition technology - challenge (tatt-c) :An open tattoo database for developing tattoo recognition research," ***Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis,*** **pp. 1–6, March 2015, Hong Kong**

NIST Tattoo Recognition Technology-Challenge (Tatt-C)

- **The goal of region of interest is to test image matching methods that allow sub-regions of interest from one tattoo image to be matched to larger tattoo images**
- **Examples of region of interest images**



Definitions

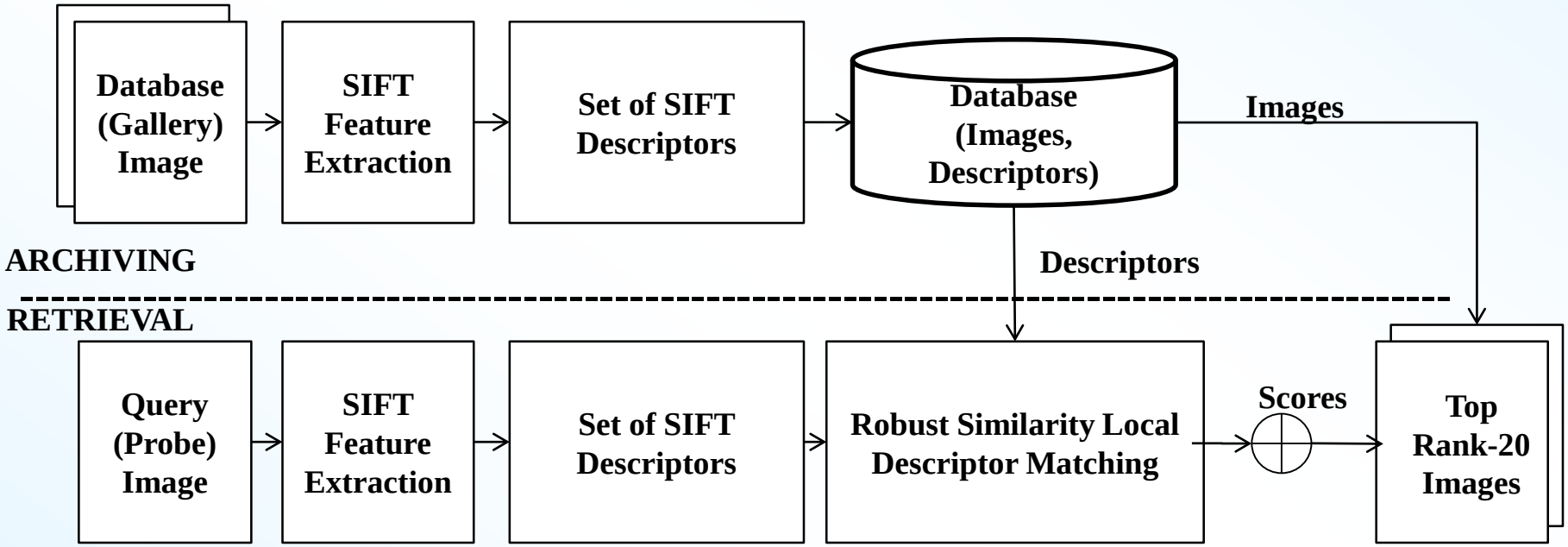
- ***Probe image*: A query image used to search against the dataset. Ideally we would like to find the true matches for a probe image**

- ***Gallery images*: 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 gallery to test the robustness of the method, also known as "distractor"**

Existing Method (MSU)

- **Block Diagram**

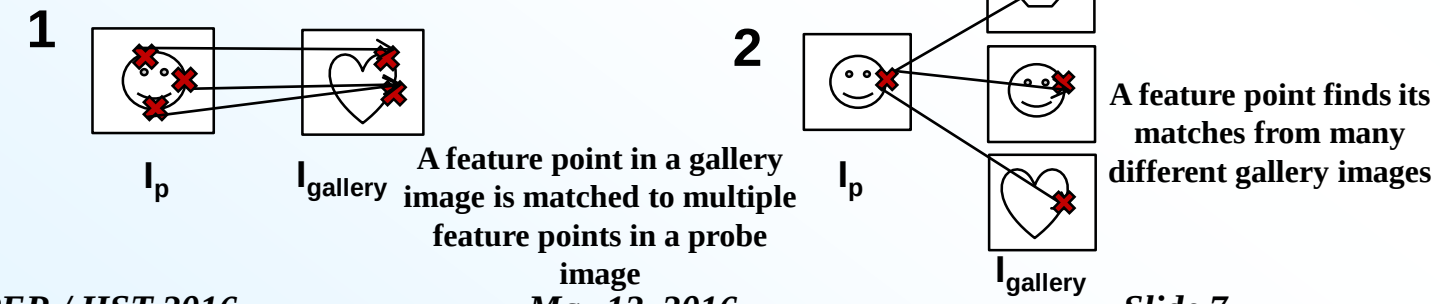


[2] (MSU). J. Lee, R. Jin, A. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," *Proceedings of the IEEE Transactions on Multimedia*, vol. 19, no. 1,pp. 40–49, January 2012

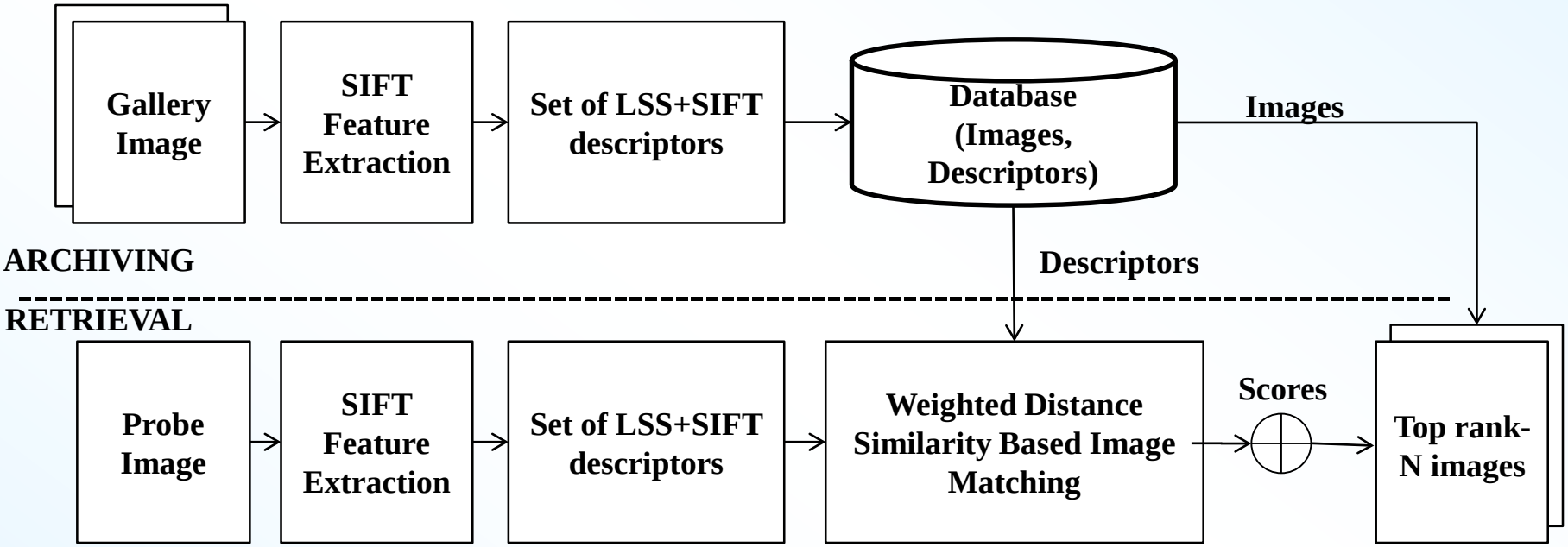
Existing Method (MSU)

(Robust Similarity Local Descriptor Matching)

- **Robust Similarity between two images :** $$S_W(I_p, I_k) = \sum_{i=1}^{N_f} \frac{x_i}{m^i(I_k)} log \frac{N_G}{n^i}$$
  - **where $m^i(I_k)$: the number of feature points including the i^{th} feature point in I_p (probe image) that are matched to the same feature point in I_k (gallery image)**
  - **n^i : the number of gallery images that have feature points matched to i^{th} feature point in I_p**
  - **N_G : the total number of gallery images**
  - **x_i : 1 if the ratio of the closest distance(d_1^i) to the second closest distance(d_2^i) between i^{th} feature descriptor in I_p and all feature descriptors in I_k is less than pre-defined threshold. Otherwise, 0**
  - **The principle of the robust similarity is that a low weight is assigned to the feature point in probe image in the following cases**



Proposed Method



LSS : Local Self Similarity [3]

[3]. E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," ***Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition,*** **pp. 1–8, June 2007, Minneapolis, MN**

SIFT (Scale Invariant Feature Transform)



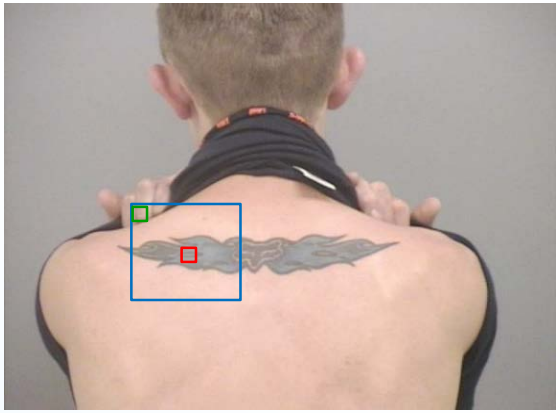
x : Extracted SIFT feature point

x : i^{th} Extracted SIFT feature point

- Each SIFT feature point is identified as a local minima and maxima of the Difference of Gaussians (DOG)
- To refine the features, we reject the feature point for which the absolute values of the DOG is less than the peak value threshold (t_p=1.5)
- SIFT descriptor (*sd*) is generated for each refined feature point
- Each SIFT descriptor (*nsd*) is then normalized
  - Example : normalized SIFT descriptor (nsd_{f_i}) for the i^{th} refined feature point (f_i)

$$nsd_{f_i} = \frac{sd_{f_i}}{||sd_{f_i}||}$$

LSS (Local Self Similarity)

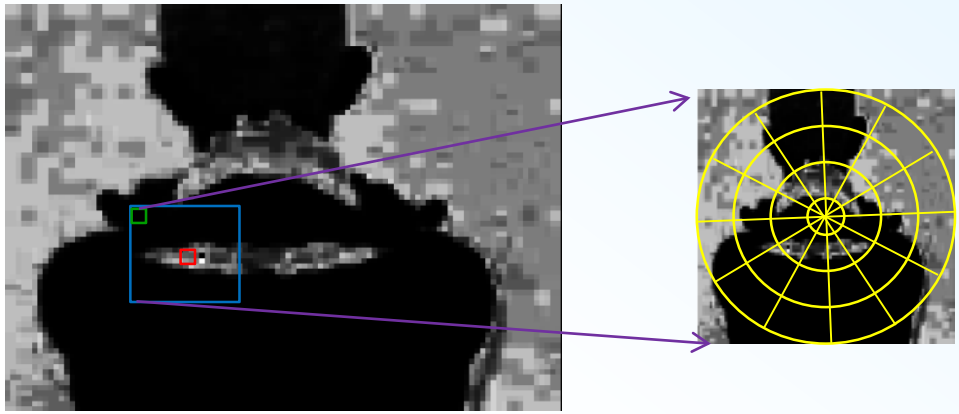


a) Input image

: 5x5 color image patch centered on the i[th] feature point (f_i)
(r_{mnk} is a pixel within this patch where i,j = 1,2,3,4,5 and k=R,G,B)

: 41x41 search range centered on the f_i

: 5x5 color image patch centered on *(x,y)*



b) Correlation Surface

By moving green patch to all the coordinates *(x,y)* within search range, compute $S_{f_i}(x,y)$

$$S_{f_i}(x,y) = exp(-\frac{SSD_{f_i}(x,y)}{C_n})$$

$SSD_{f_i}(x,y)$: sum of square difference of between all r_{mnk} and g_{mnk}.

$$SSD_{f_i}(x,y) = \sum_{k\in\{R,G,B\}} \sum_{m=1}^{5} \sum_{n=1}^{5} (r_{mnk} - g_{mnk})^2$$

S_{f_i} : Correlation surface

$nlss_{f_i}$: Normalized LSS

c) Generate Descriptor

: 2D polar histogram centered on f_i

The maximum value among all the $S_{f_i}(x,y)$ within each bin of 2D polar histogram is computed and is used for value for each bin

Then, the histogram is normalized and vectorized

$$nlss_{f_i} = \frac{lss_{f_i}}{||lss_{f_i}||}$$

Image Descriptor

- **An image descriptor for sub-region of a tattoo image should be generated using several image properties such as color and gradient because only one image property is not enough to represent the sub-region of a tattoo**
- **SIFT : a histogram of pixel gradients is used**
- **LSS : a similar color pattern is used**
- **Our image descriptor is generated by combining these two descriptors**
- **Example : our image descriptor (ld_{f_i}) for f_i**

$$ld_{f_i} = \begin{bmatrix} w_1 nlss_{f_i} \\ w_2 nsd_{f_i} \end{bmatrix}$$

w_1, w_2 : weights for LSS and SIFT respectively

$nlss_{f_i}$ **: normalized LSS for f_i**

nsd_{f_i} **: normalized SIFT for f_i**

Weighted Distance Similarity (WDS) Image Matching

$$S_{WDS}(I_p, I_k) = \sum_{i=1}^{N_f} \frac{1}{m^i(I_k)} log \frac{N_G}{n^i} \left(\frac{x_i}{1 + e^{d_1^i}} \right)$$

- $m^i(I_k)$ **: the number of feature points including the** i^{th} **feature point in** I_p **that are matched to the same feature point in** I_k
- n^i **: the number of gallery images that have feature points matched to** i^{th} **feature point in** I_p
- N_G **: the total number of gallery images**
- x_i **: 1 if the ratio of the closest distance (**d_1^i**) to the second closest distance (**d_2^i**) between** i^{th} **feature descriptor in** I_p **and all feature descriptors in** I_k **is less than pre-defined threshold. Otherwise, 0**

Region of Interest Dataset

- **Number of gallery images: 157**
- **Number of probe images: 297**
- **Number of background images: 4332**

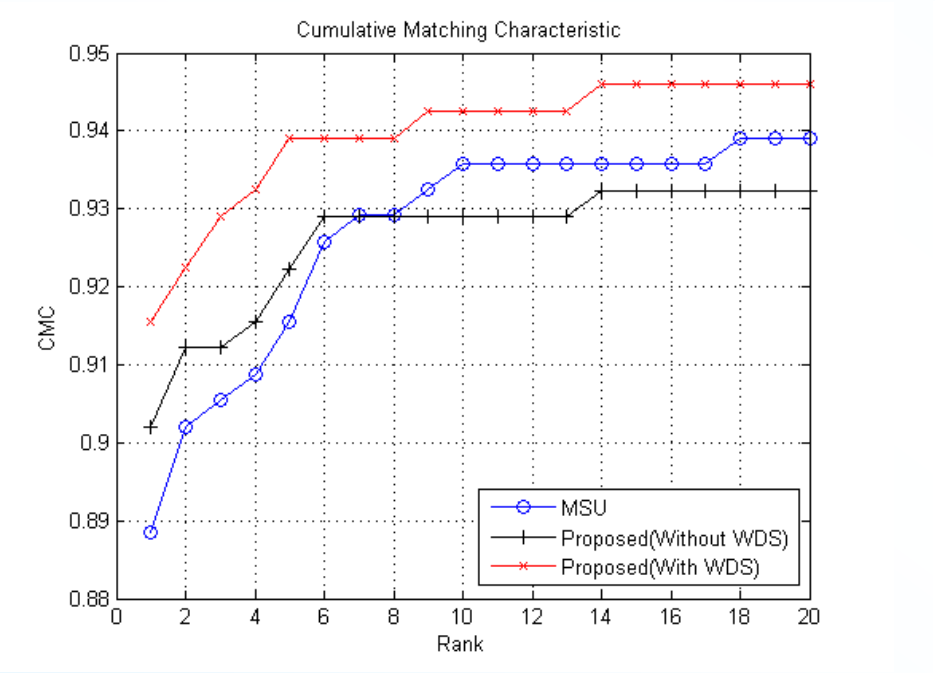
Experimental Results

- **The tattoo region of interest dataset is split into 5 sub-datasets for 5 fold cross validation**
- **Evaluation metric**
  - **CMC (Cumulative Match Characteristic)**
  - **Precision and Recall Curve**
  - **MAP (Mean Average Precision)**

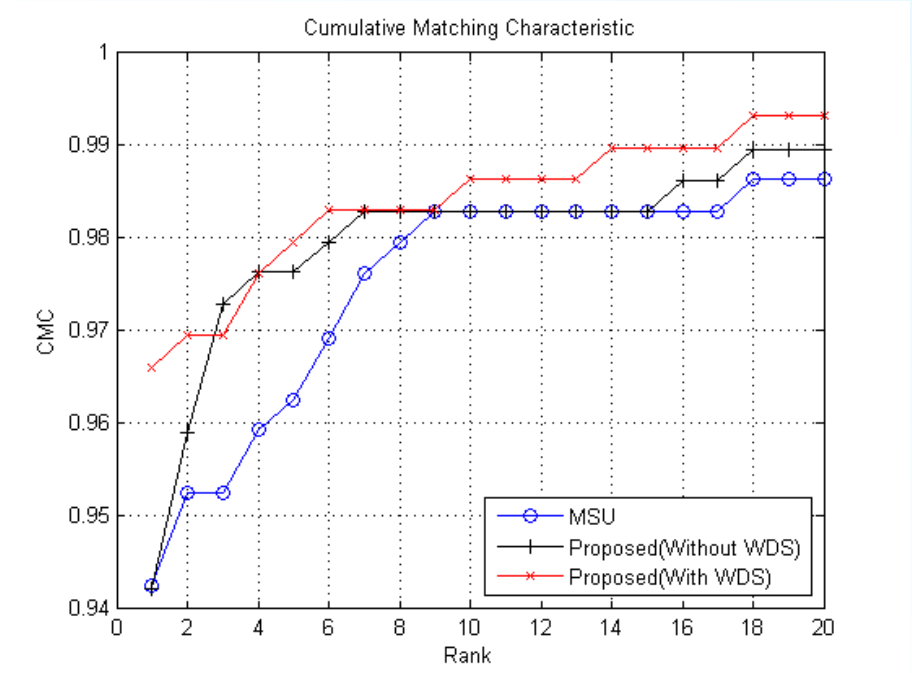
    : The mean of the average area under the precision and recall curve
- **We compared our method against one existing method**
- **We also show the results of our submission for NIST tattoo challenge (tattoo region of interest)**

Experimental Results (CMC)

- **CMC Curve**



(a) With Background Image



(b) Without Background Image

MSU : SIFT Descriptor + MSU Image Similarity
Proposed (Without WDS) : Our Image Descriptor (LSS+SIFT) + MSU Image Similarity
Proposed (With WDS) : Our Image Descriptor (LSS+SIFT) + WDS Similarity

[2] (MSU). J. Lee, R. Jin, A. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," *Proceedings of the IEEE Transactions on Multimedia*, vol. 19, no. 1,pp. 40–49, January 2012

Experimental Results (CMC & MAP)

Method	CMC					MAP
	Rank 1	Rank 5	Rank 10	Rank 15	Rank 20	
MSU	0.8886	0.9156	0.9358	0.9358	0.9391	0.9015
Proposed (Without WDS)	0.9021	0.9223	0.9289	0.9323	0.9323	0.9068
Proposed (With WDS)	0.9156	0.9391	0.9425	0.9460	0.9460	0.9240

(a) With Background Image

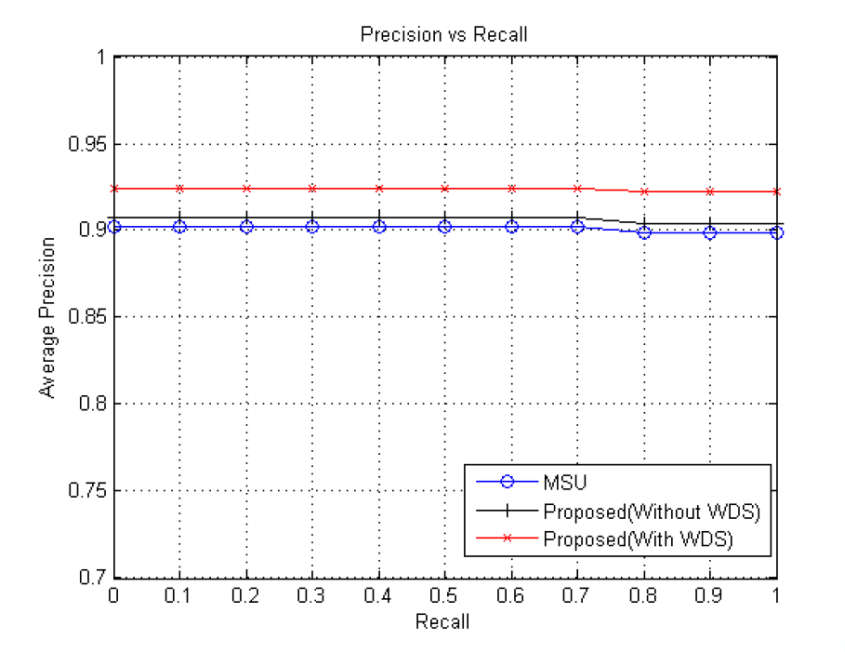
Method	CMC					MAP
	Rank 1	Rank 5	Rank 10	Rank 15	Rank 20	
MSU	0.9426	0.9626	0.9830	0.9830	0.9865	0.9538
Proposed (Without WDS)	0.9423	0.9763	0.9830	0.9830	0.9896	0.9574
Proposed (With WDS)	0.9661	0.9796	0.9865	0.9898	0.9932	0.9720

(b) Without Background Image

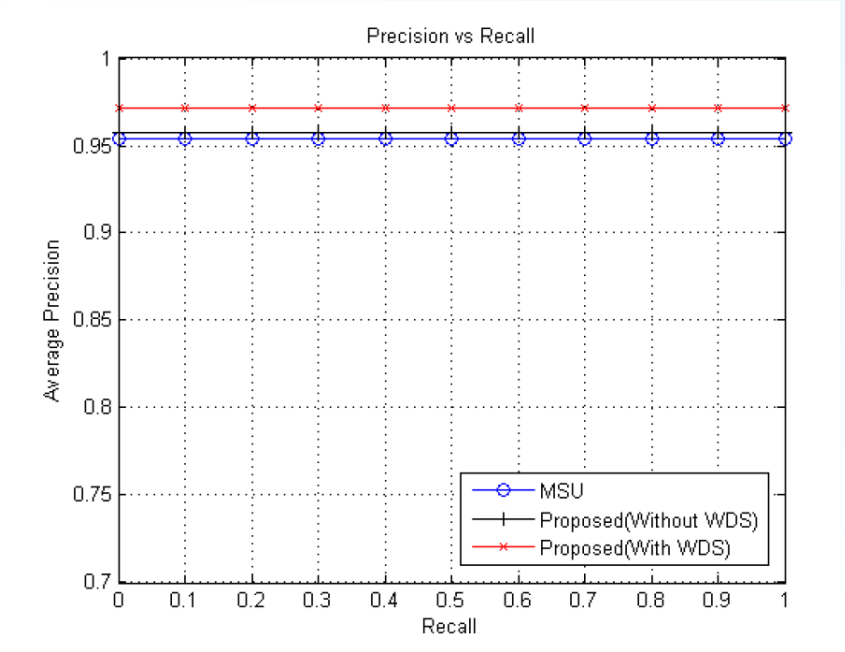
[4] (MSU) : J. Lee, R. Jin, A. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," *IEEE Transactions on Multimedia, vol. 19, no. 1*,pp. 40–49, January 2012

Experimental Results (Precision and Recall Curve)

- **Precision and Recall Curve**



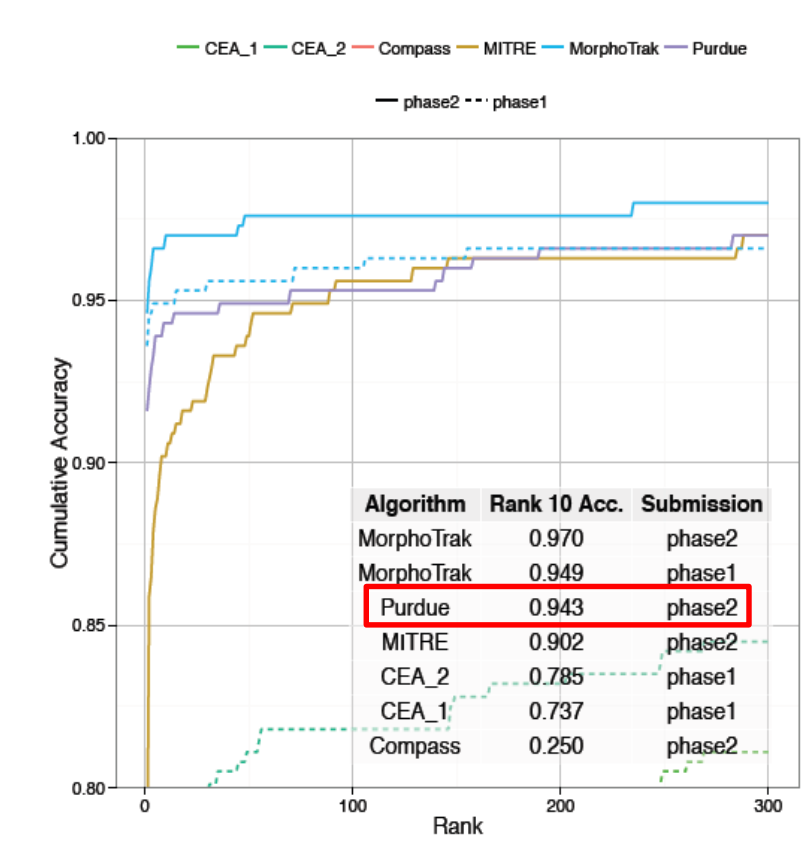
(a) With Background Image



(b) Without Background Image

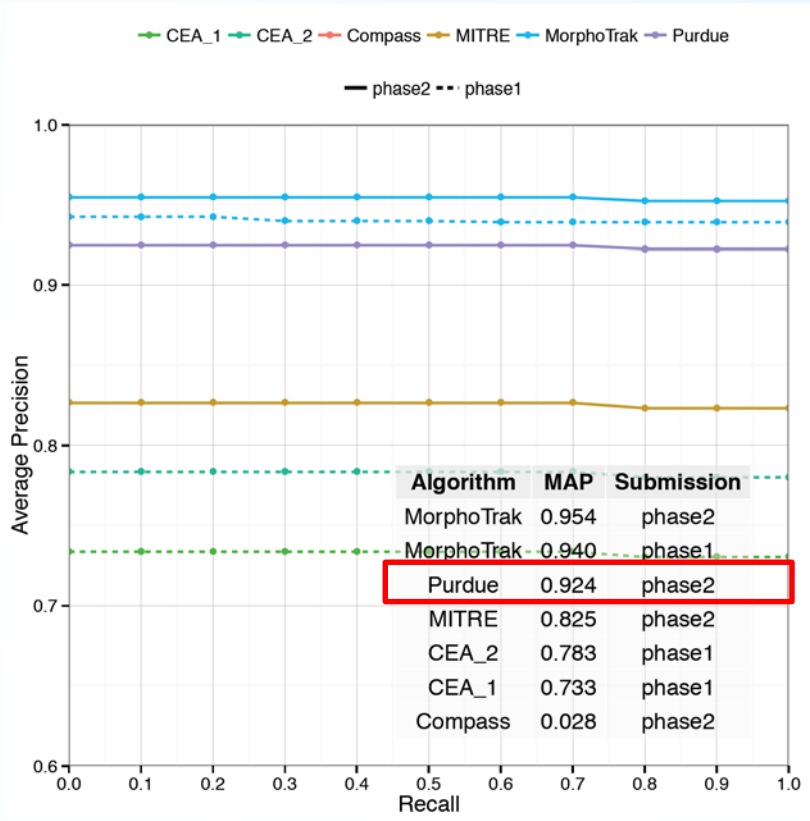
[4] (MSU) : J. Lee, R. Jin, A. Jain, and W. Tong, "Image retrieval in forensics: Tattoo image database application," ***IEEE Transactions on Multimedia, vol. 19, no. 1,*****pp. 40–49, January 2012**

Experimental Results (NIST Tattoo Challenge Results)



Algorithm	Rank 10 Acc.	Submission
MorphoTrak	0.970	phase2
MorphoTrak	0.949	phase1
Purdue	0.943	phase2
MITRE	0.902	phase2
CEA_2	0.785	phase1
CEA_1	0.737	phase1
Compass	0.250	phase2

(a) CMC Curve reported in [4]



Algorithm	MAP	Submission
MorphoTrak	0.954	phase2
MorphoTrak	0.940	phase1
Purdue	0.924	phase2
MITRE	0.825	phase2
CEA_2	0.783	phase1
CEA_1	0.733	phase1
Compass	0.028	phase2

(b) Precision and Recall reported in [4]

[4]. M. Ngan, G. Quinn, and P. Grother, "Tattoo recognition technology challenge (tatt-c) outcomes and recommendations," *NIST Internal Report, September* 2015, National Institute of Standards and Technology, Gaithersburg, MD

Experimental Results
(NIST Tattoo Challenge Results)

Method	Submission	CMC				MAP
		Rank 1	Rank 10	Rank 100	Rank 300	
CEA1	phase 1	0.731	0.737	0.771	0.811	0.733
CEA2	phase 2	0.781	0.785	0.818	0.845	0.783
Compass	phase 2	0.175	0.250	0.325	0.475	0.028
MITRE	phase 2	0.771	0.902	0.956	0.970	0.825
Morpho Track	phase 1	0.936	0.949	0.960	0.966	0.940
Morpho Track	phase 2	0.946	0.970	0.976	0.980	0.954
Purdue	phase 2	0.916	0.943	0.953	0.970	0.924

[4]. M. Ngan, G. Quinn, and P. Grother, "Tattoo recognition technology challenge (tatt-c) outcomes and recommendations," *NIST Internal Report, September* 2015, National Institute of Standards and Technology, Gaithersburg, MD

Examples of Matched Results

- **With Background Images**

Top 10 ranks (left to right, top to bottom)

Matched

Probe Image





Examples of Matched Results

- **With Background Images**

Top 10 ranks (left to right, top to bottom)

Matched



Probe Image





Examples of Matched Results

- **With Background Images**

Top 10 ranks (left to right, top to bottom)

No Matching

Probe Image





Examples of Matched Results

- **Without Background Images**

Top 10 ranks (left to right, top to bottom)

Matched

Probe Image





Examples of Matched Results

- **Without Background Images**

Top 10 ranks (left to right, top to bottom)

Probe Image



Matched





Examples of Matched Results

- **Without Background Images**

Top 10 ranks (left to right, top to bottom)

Matched



Probe Image



Conclusions and Future Work

- **We describe our submission to the NIST Tatt-C region of interest challenge**
- **Our image descriptor is based on the combination of SIFT and LSS**
- **Weighted Distance Similarity (WDS) Image Matching improved matching accuracy**
- **Our method sometimes fails when an image has very low resolution or it is blurred**
- **We will investigate similarity metrics and features that work well with low resolution or blurred images**

Tattoo Image Retrieval For Region of Interest

Joonsoo Kim, He Li, Jiaju Yue, Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

Video-Based Re-Identification of Subjects for The Greater Cleveland Regional Transit Authority

vBOLO

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

vBOLO 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 vBOLO list by manually highlighting them; use computer vision and image processing techniques to generate the “vBOLO fingerprint”

vBOLO Goals

- Use vBOLO 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 vBOLO system would be more effective

vBOLO 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

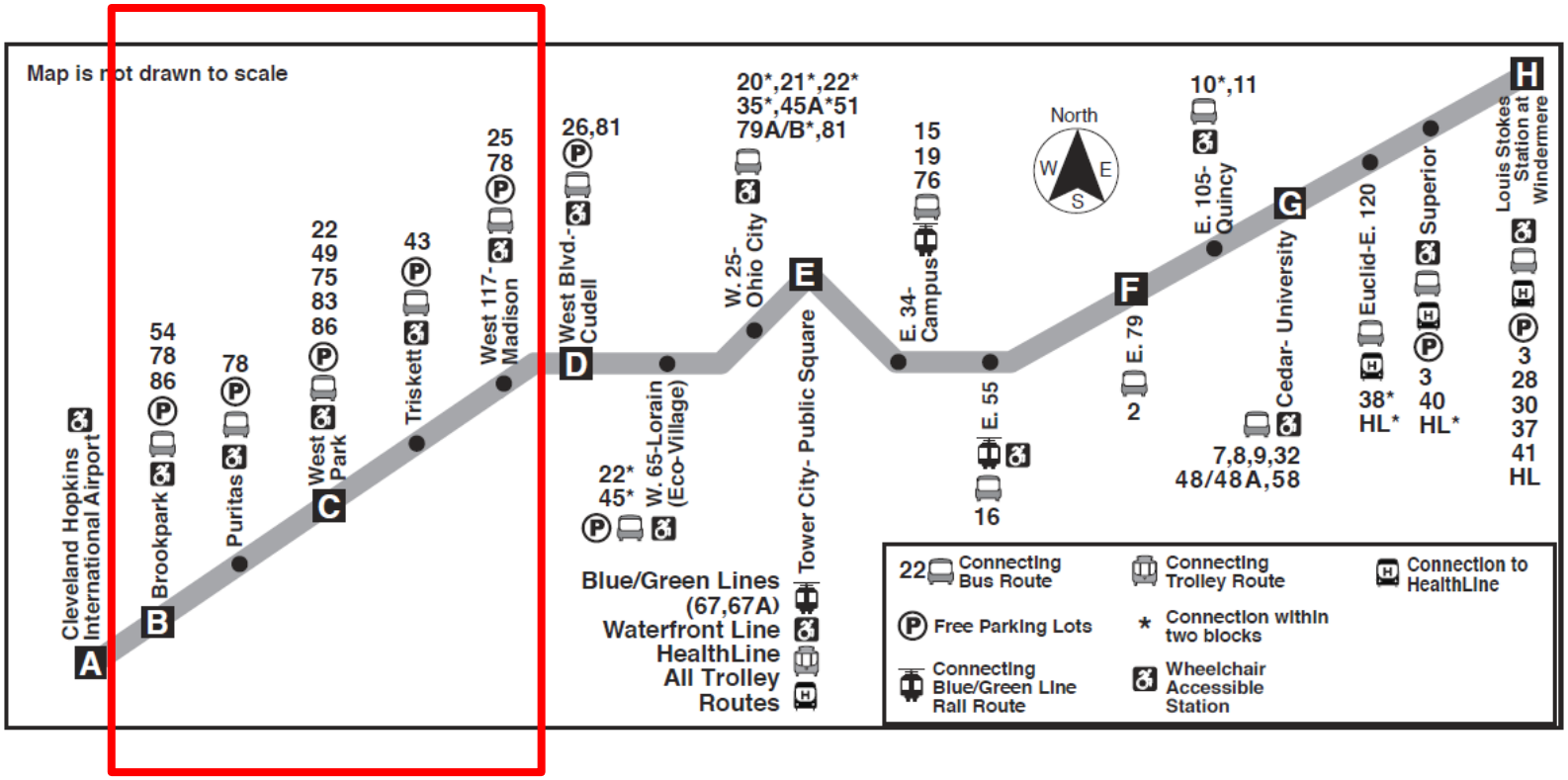
- **Joel Brogan, Amy Li, Nikita Amelchenko**

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

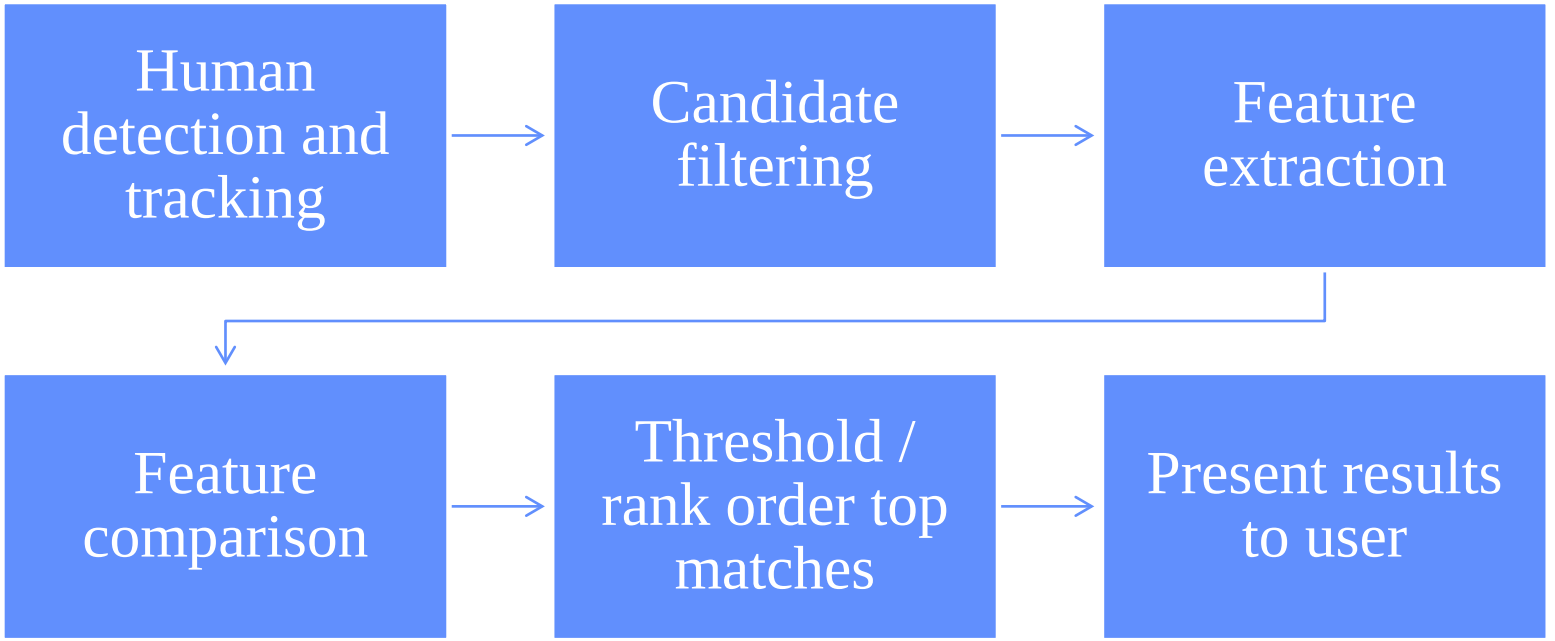
Student 7	19 appearances
Non-students	11 appearances

Body Matching Re-Identification Pipeline

Target specification:



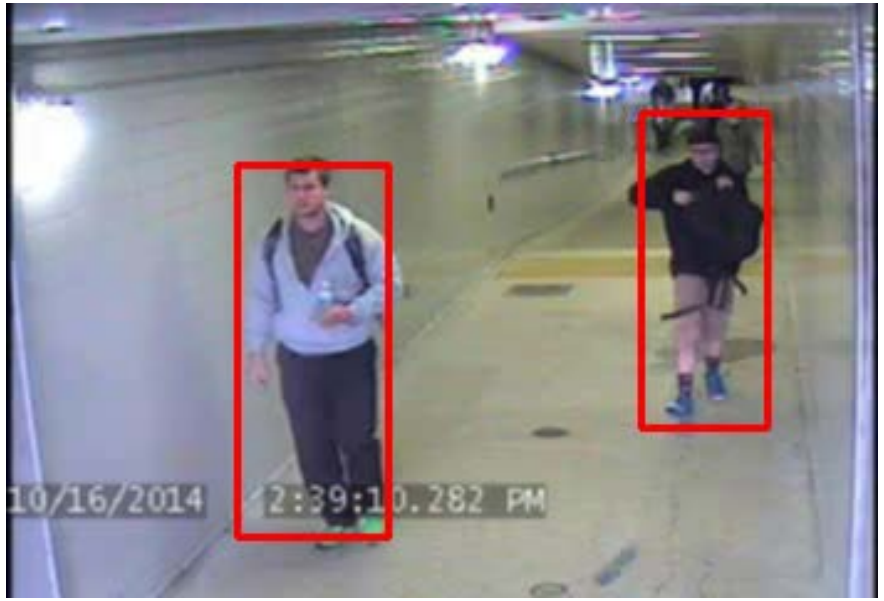
Target re-identification:



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

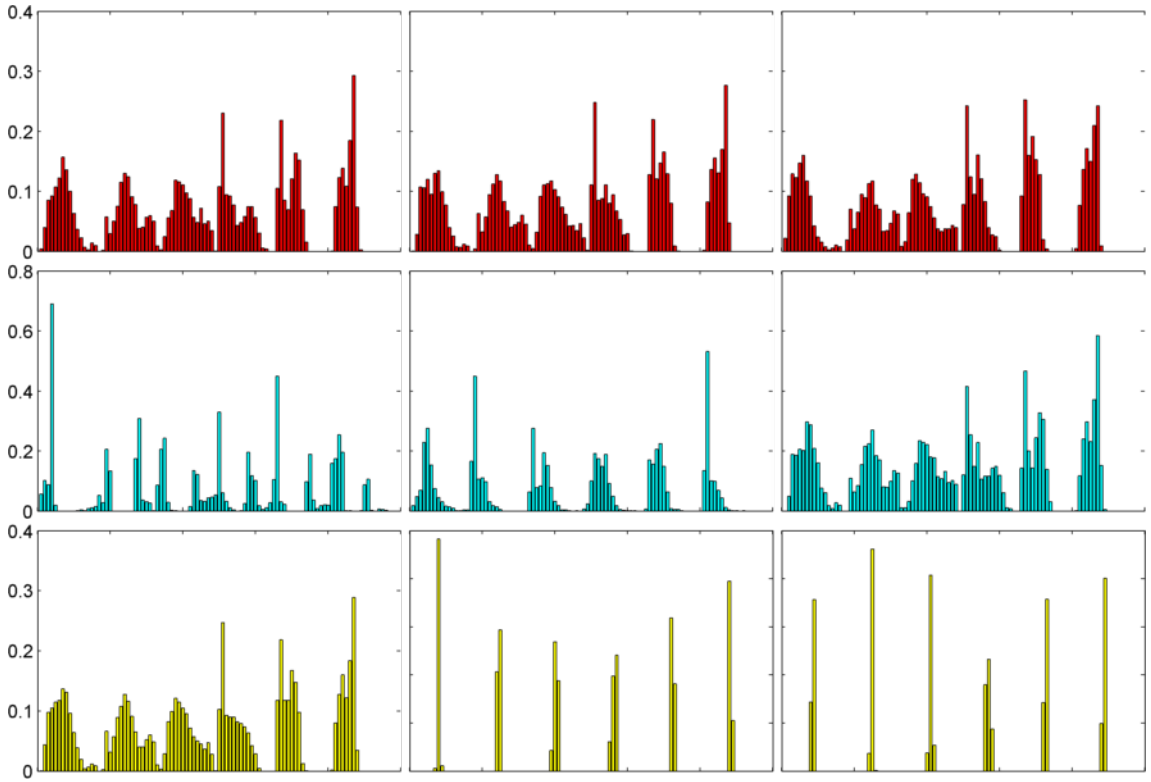
Each target and candidate is divided into several horizontal strips. Features are extracted from each strip based on the distribution of colors



RGB

HSV

YCbCr

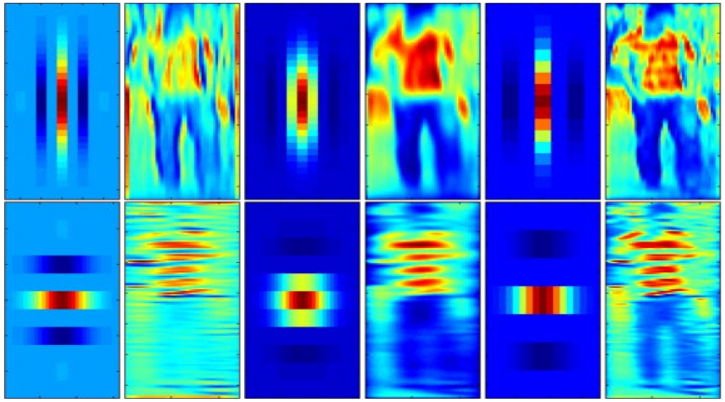


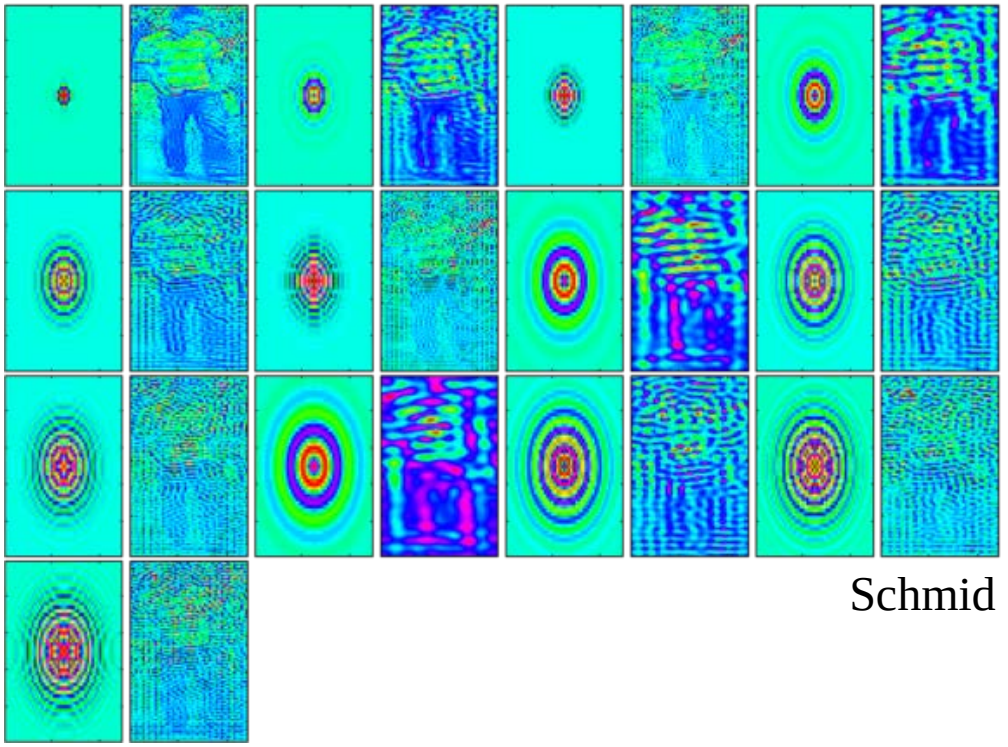
Texture Features

Texture features based on the patterns of intensity in each strip are also calculated



Gabor



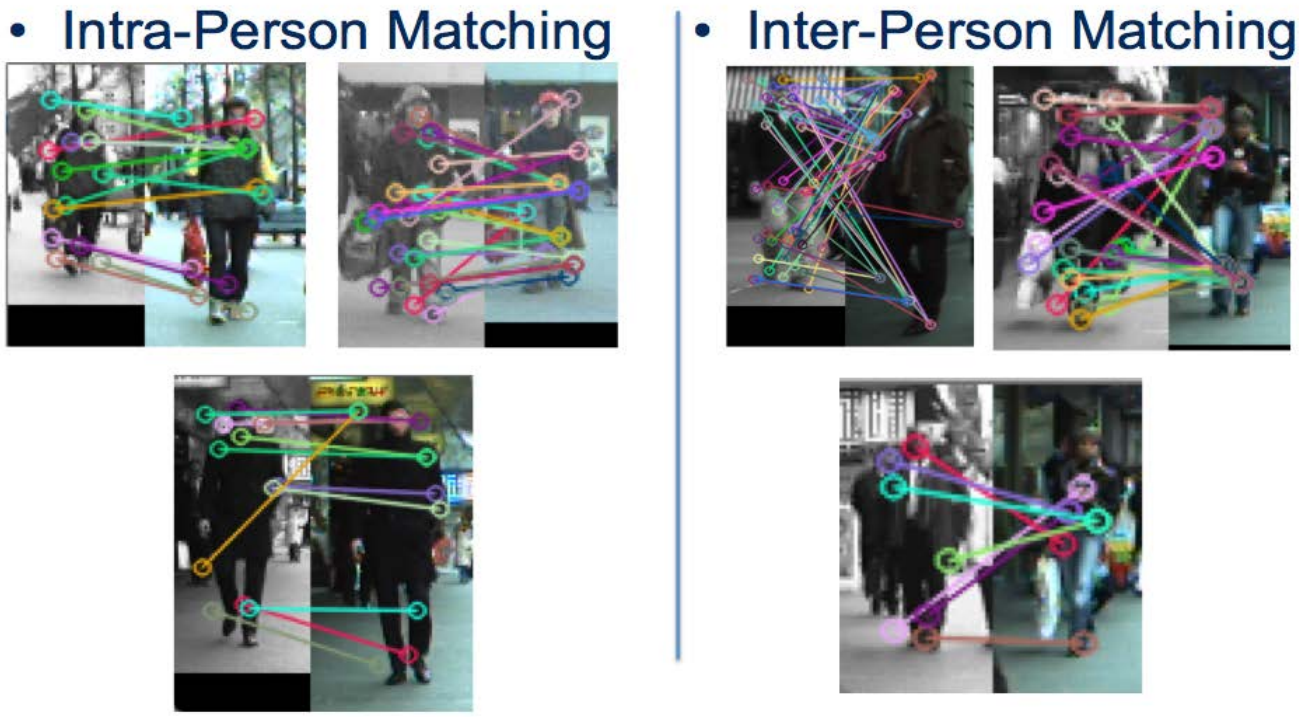


Schmid

We are currently investigating the combinations of color and texture feature that yield the best performance

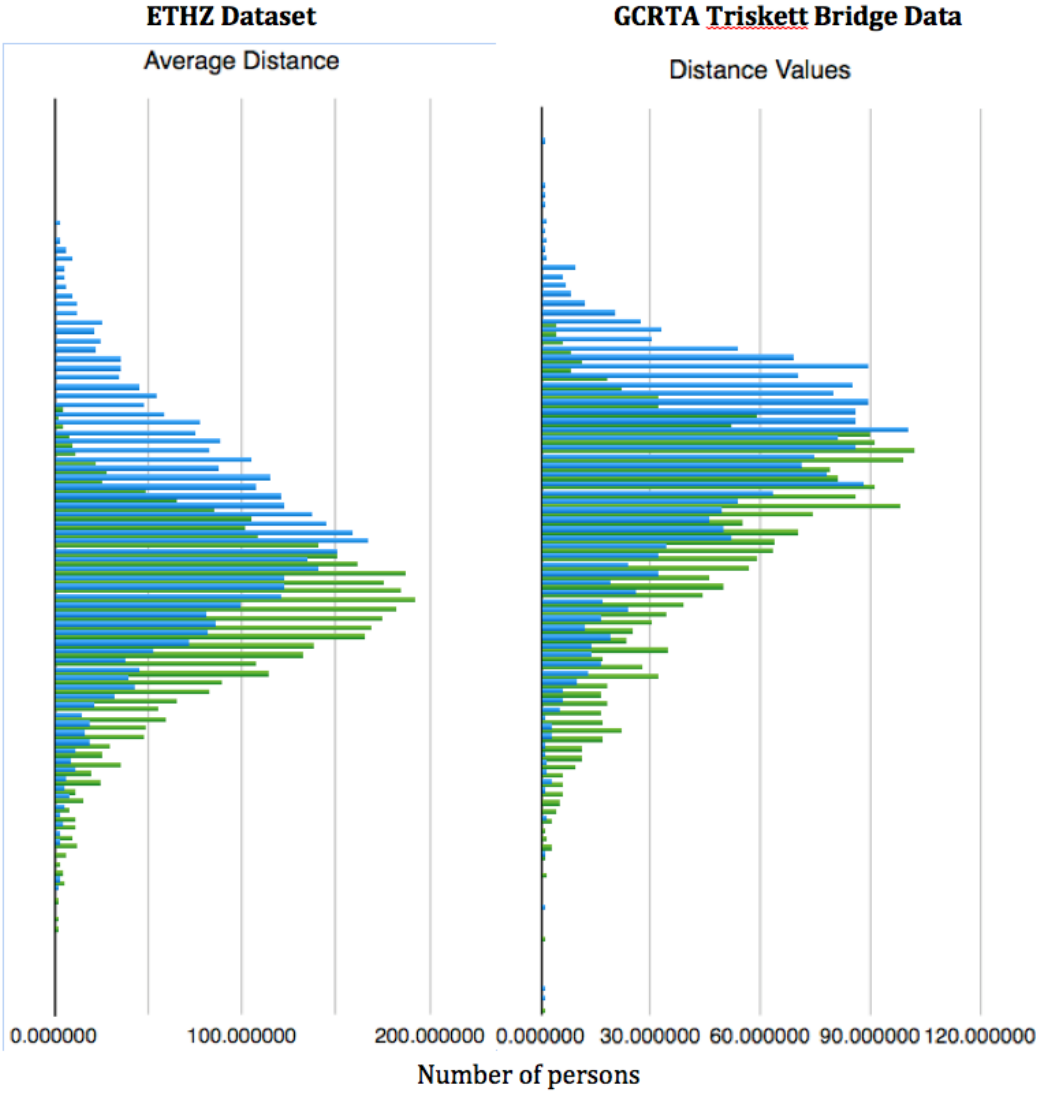
SURF Features

- Spectral analysis performed using SURF features on ETHZ dataset and annotated version of Triskett Bridge video (matching descriptor Cartesian distances ("Global Distance")



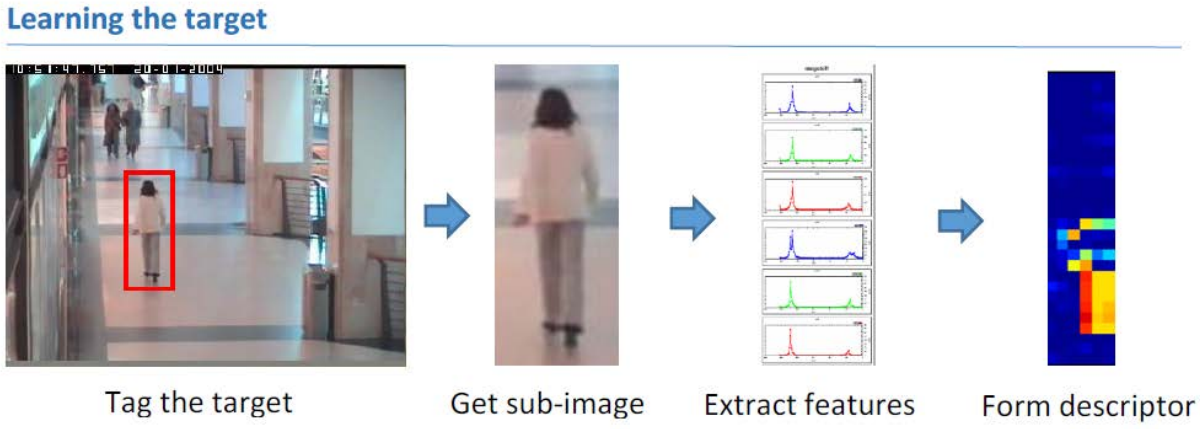
Examples of Intra-Person (true) matching, and Inter-Person (false matching) on pedestrians

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’



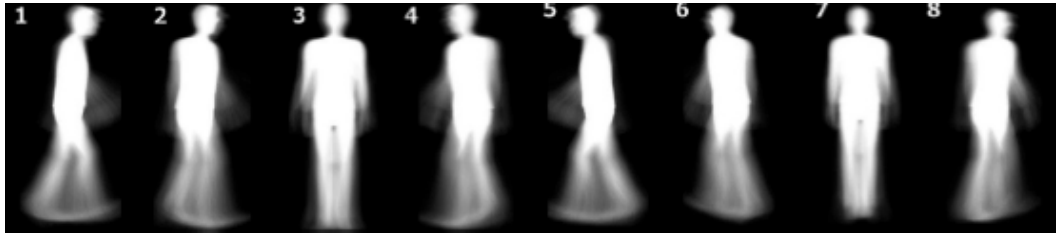
Recognition Of Person-Based Attributes

- 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

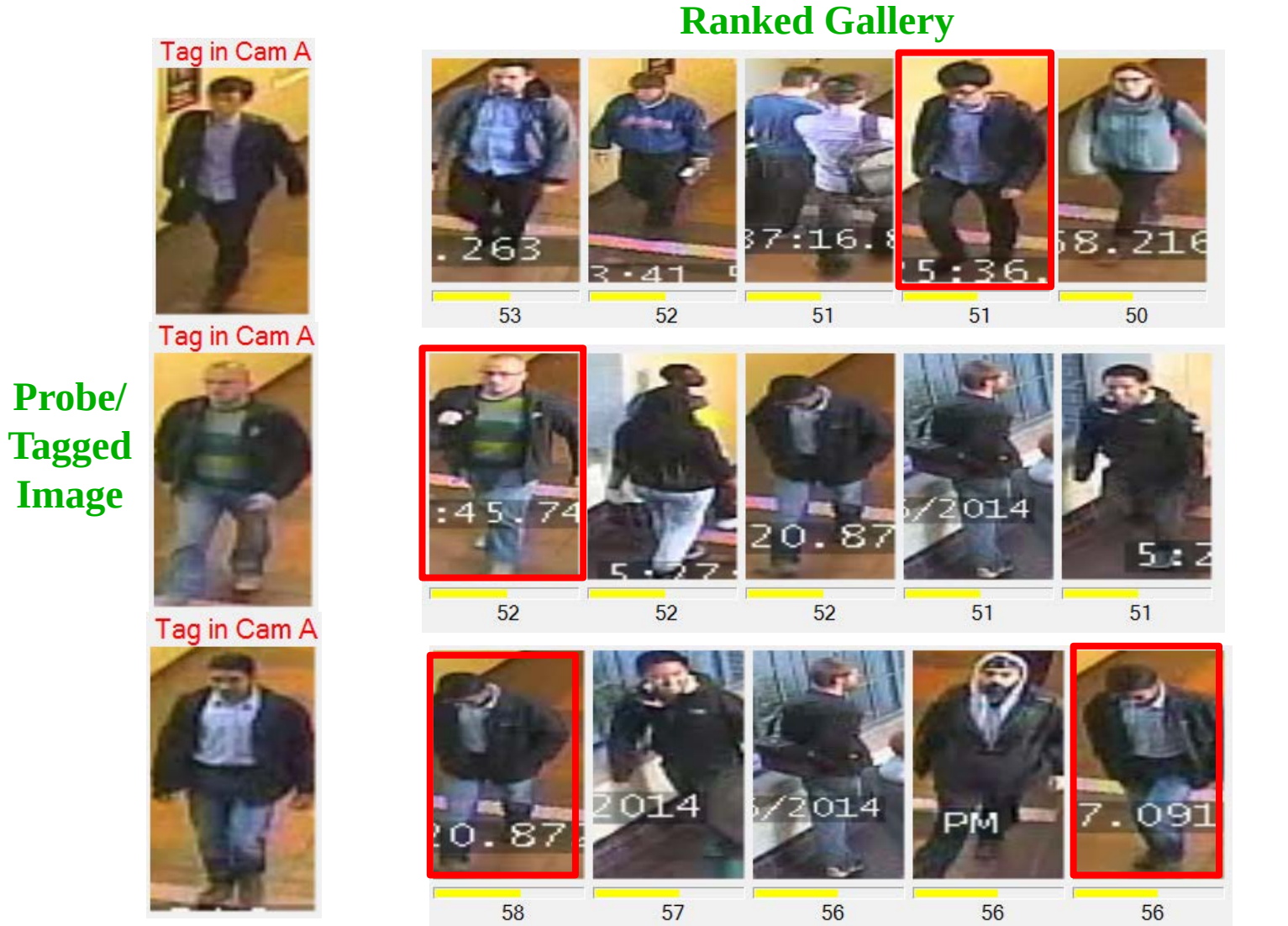
Probe/ Tagged Image



Ranked Gallery



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

Ranked Gallery

(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

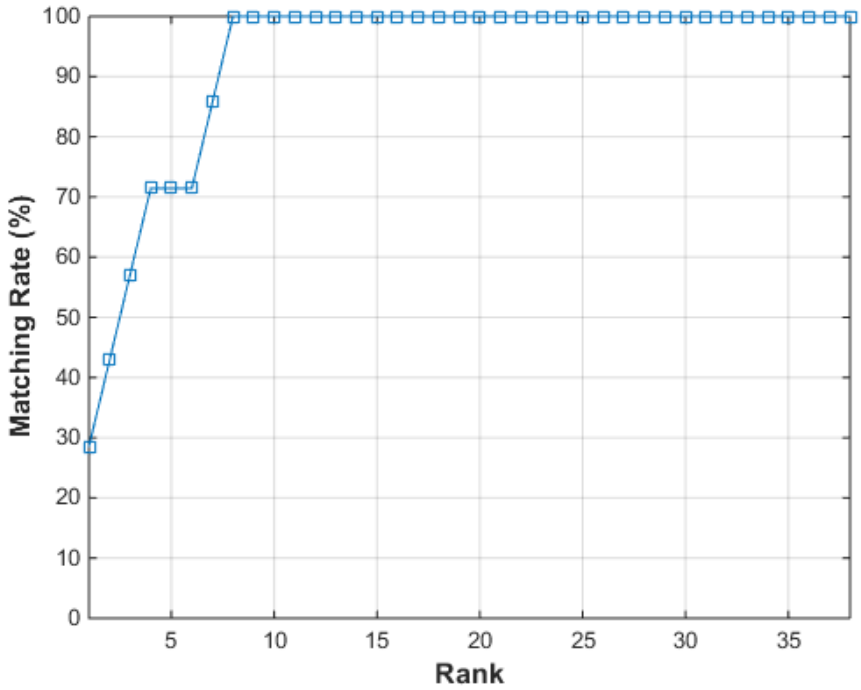


Robust to appearance change!



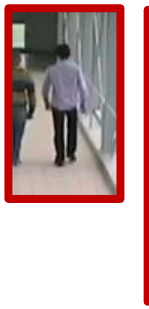
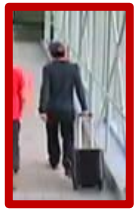
Preliminary results: NEU group

Gallery size: 38

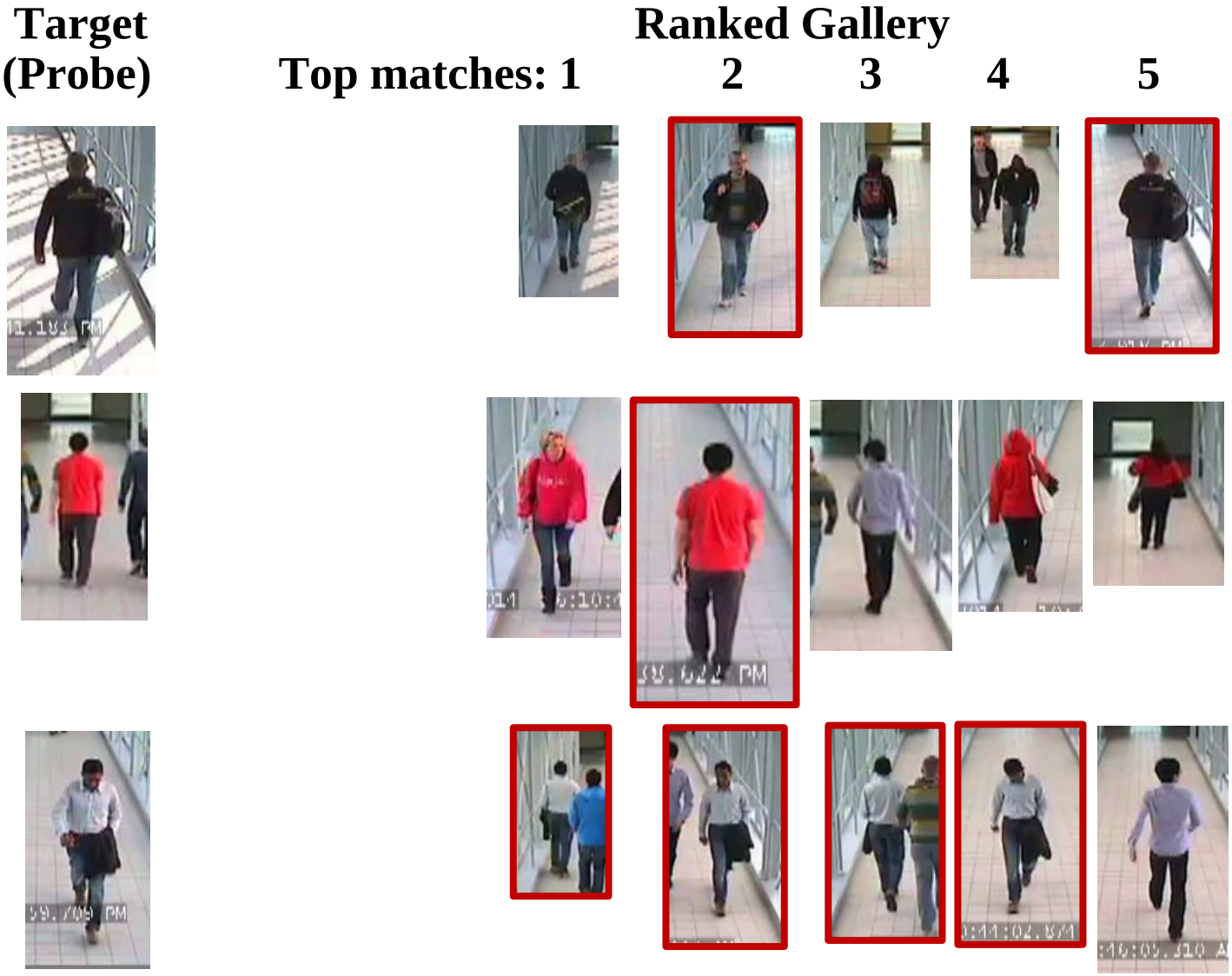


Rank	1	2	3	5	10
Matching Rate (%)	28.6	42.9	57.1	71.4	100

Preliminary Results: Purdue Group

Target (Probe)	Ranked Gallery Top matches: 1	2	3	4	5
					
					
					

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%

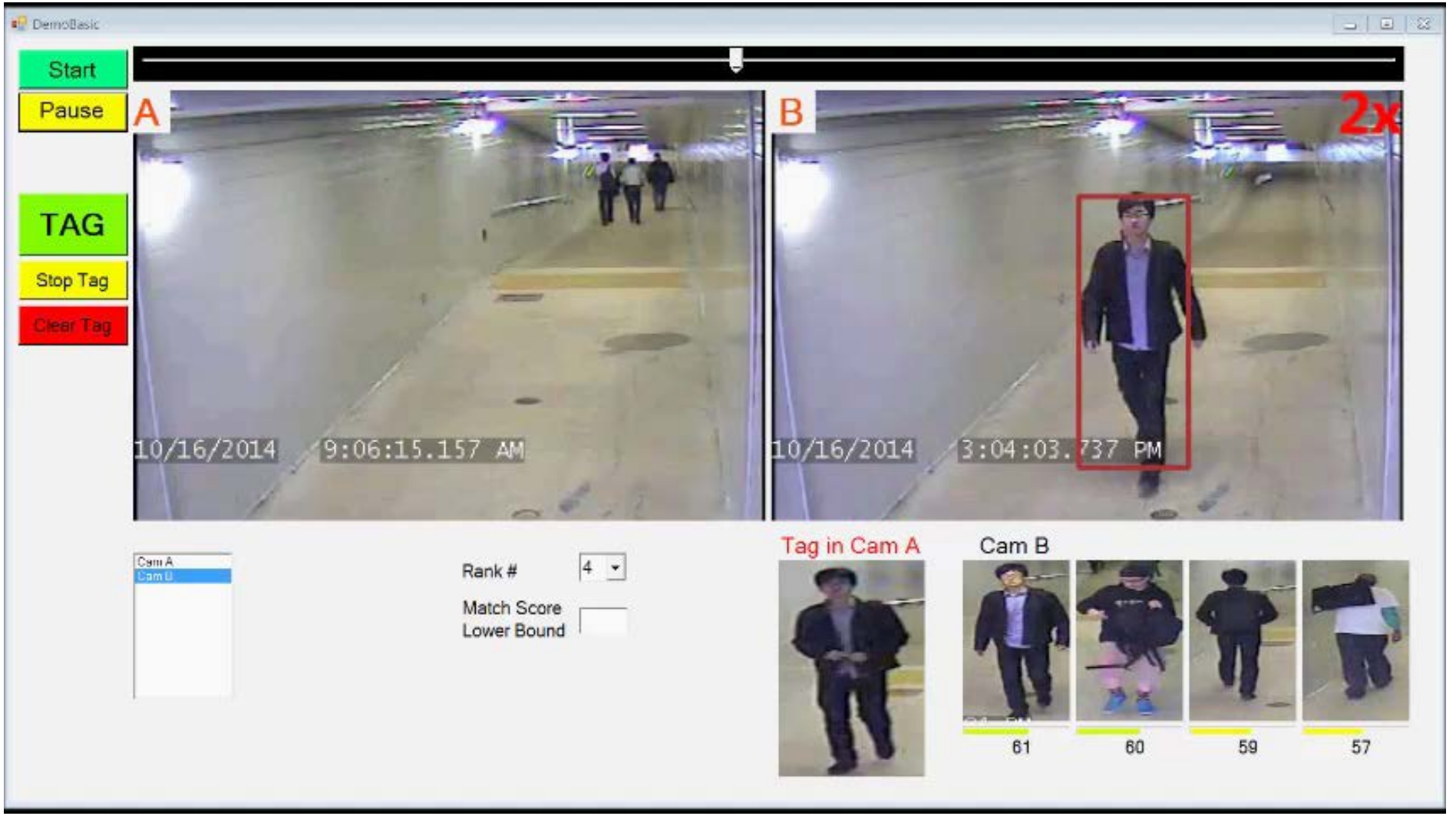


West 177 - Tunnel

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

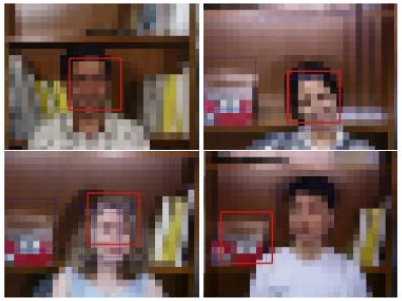
- Face detection in low-resolution images is challenging
- Detection becomes harder for both machine and human from resolutions of 24×24 to 6×6
- Results below are produced by combination of “Viola-Jones detector” + color heuristics



Resolution of faces
24 × 24



16 × 16



8 × 8



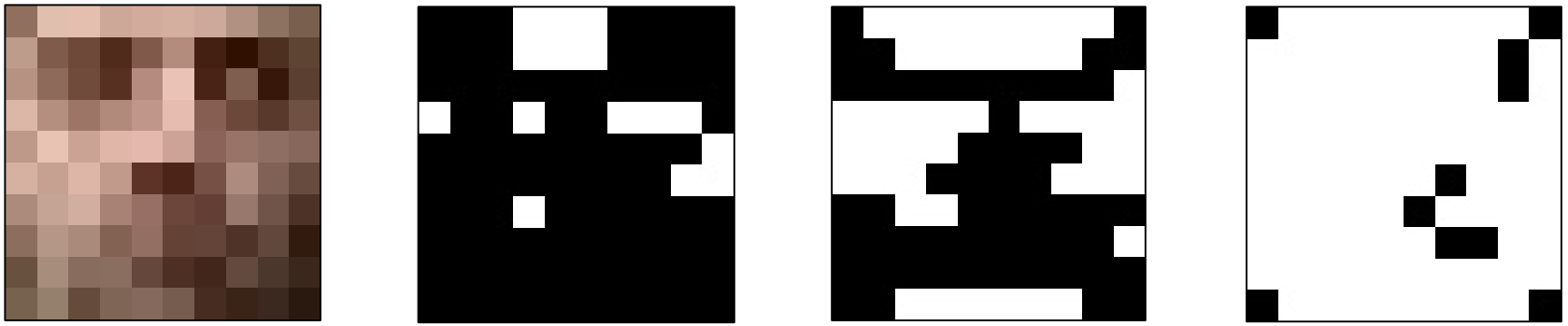
6 × 6

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

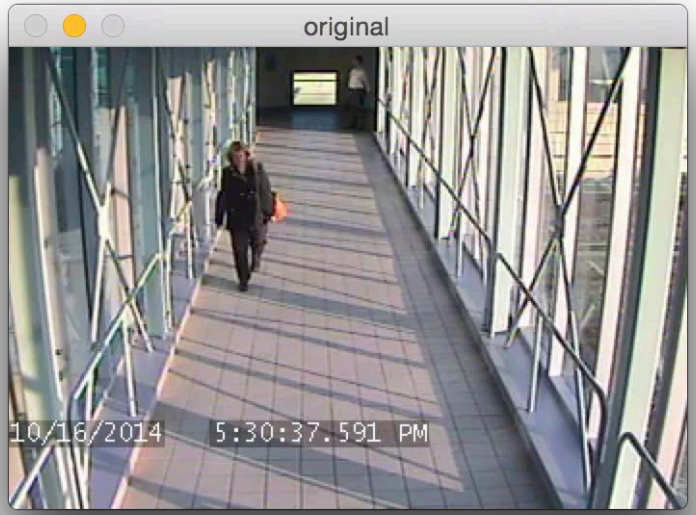
Face Detection: Notre Dame

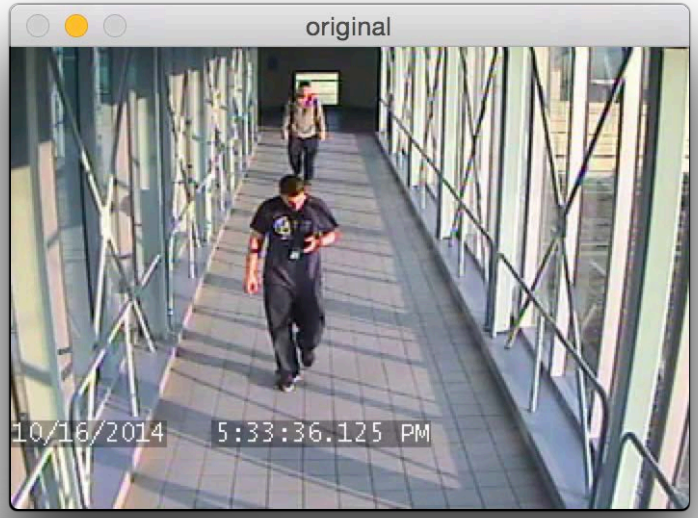
- Custom face detector implemented based on literature
- 3 stage detector, robust to illumination changes
- To prevent false negatives, we allow false positives to contribute to the final face detection
- A density filter of face detection points allows us to robustly filter out the actual facial area 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

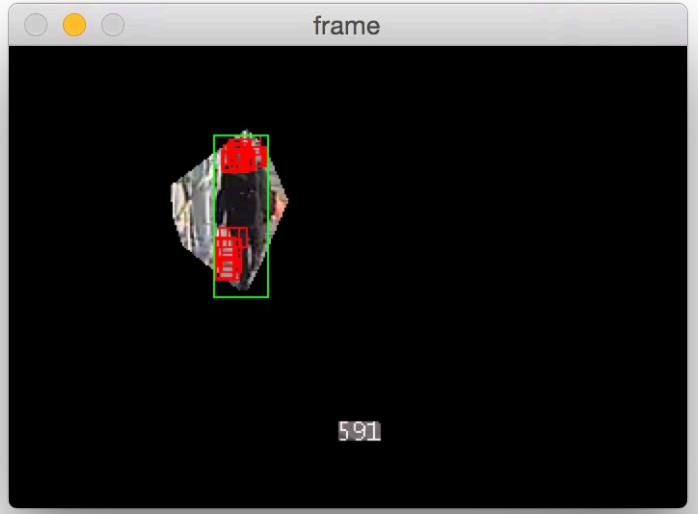
Preliminary Results: Notre Dame

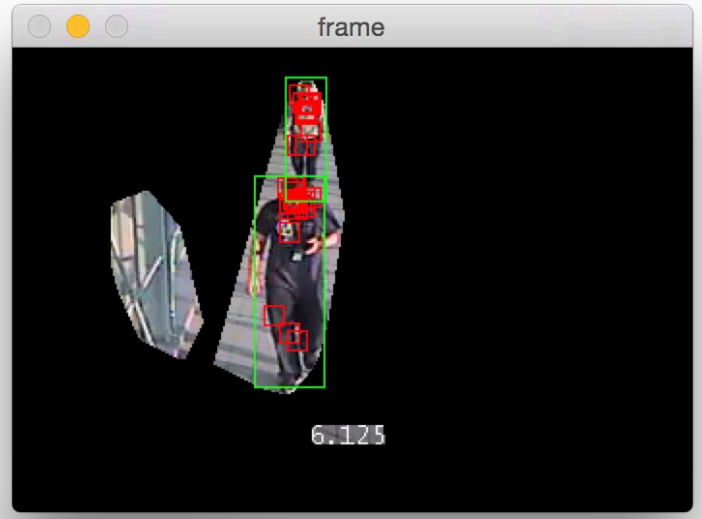






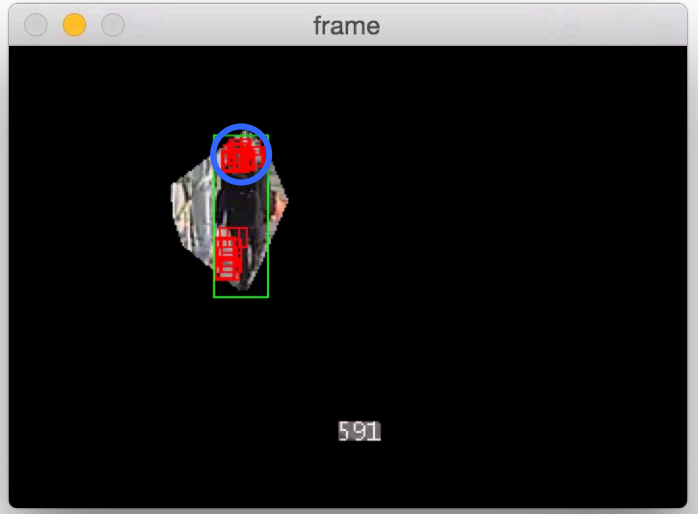
Preliminary Results: Notre Dame

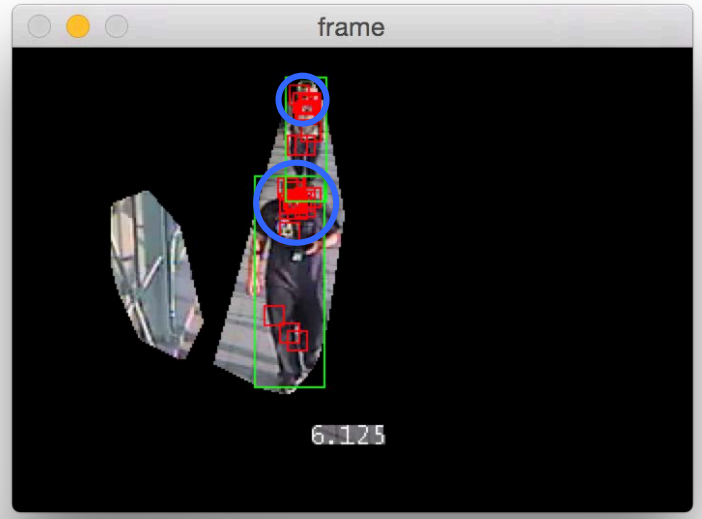






Preliminary Results: Notre Dame

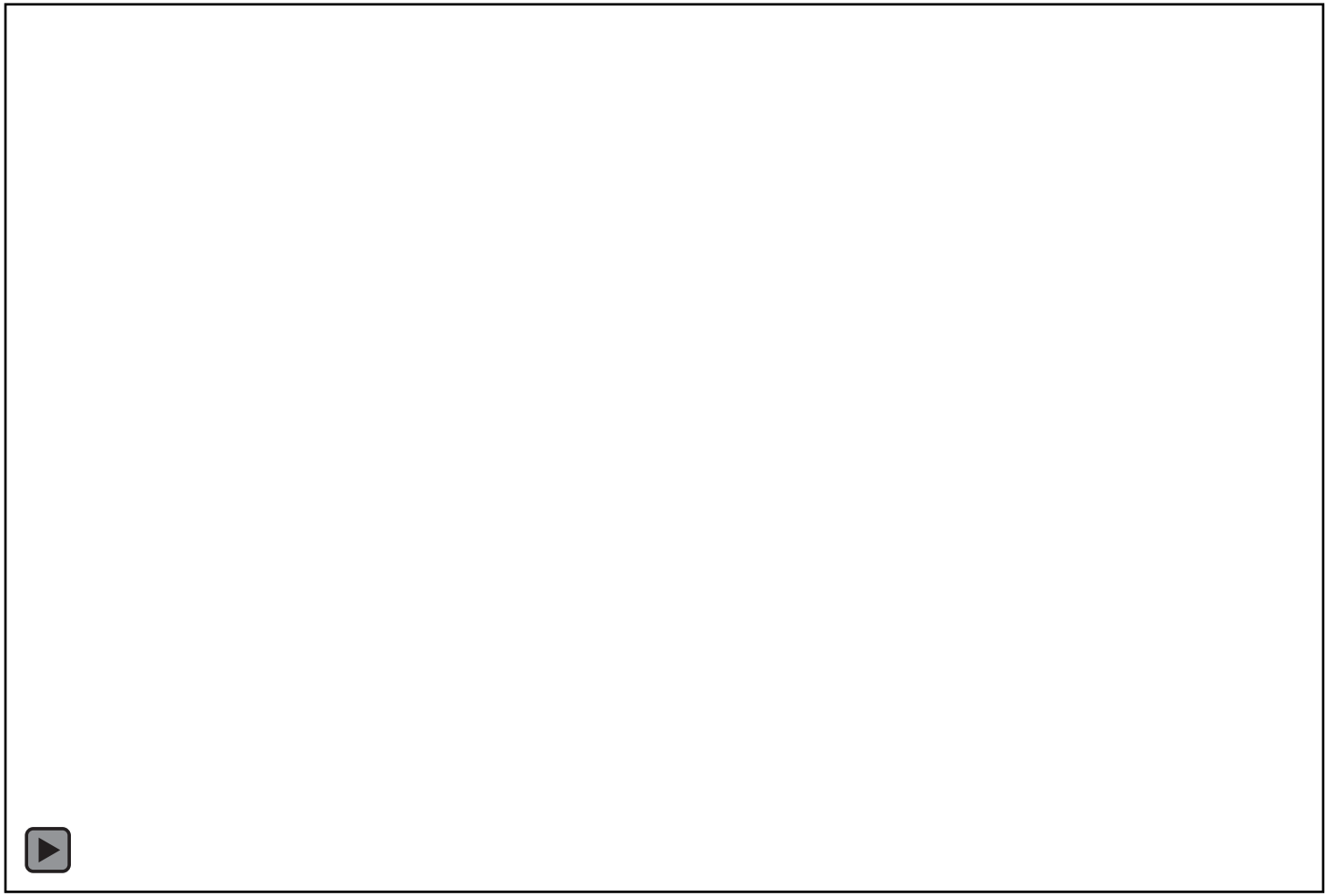






Preliminary Results: Notre Dame

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





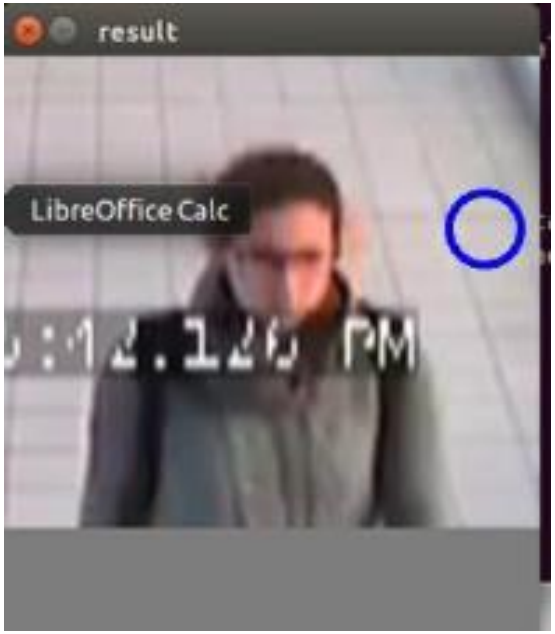
Preliminary Results: NEU Group

- 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)





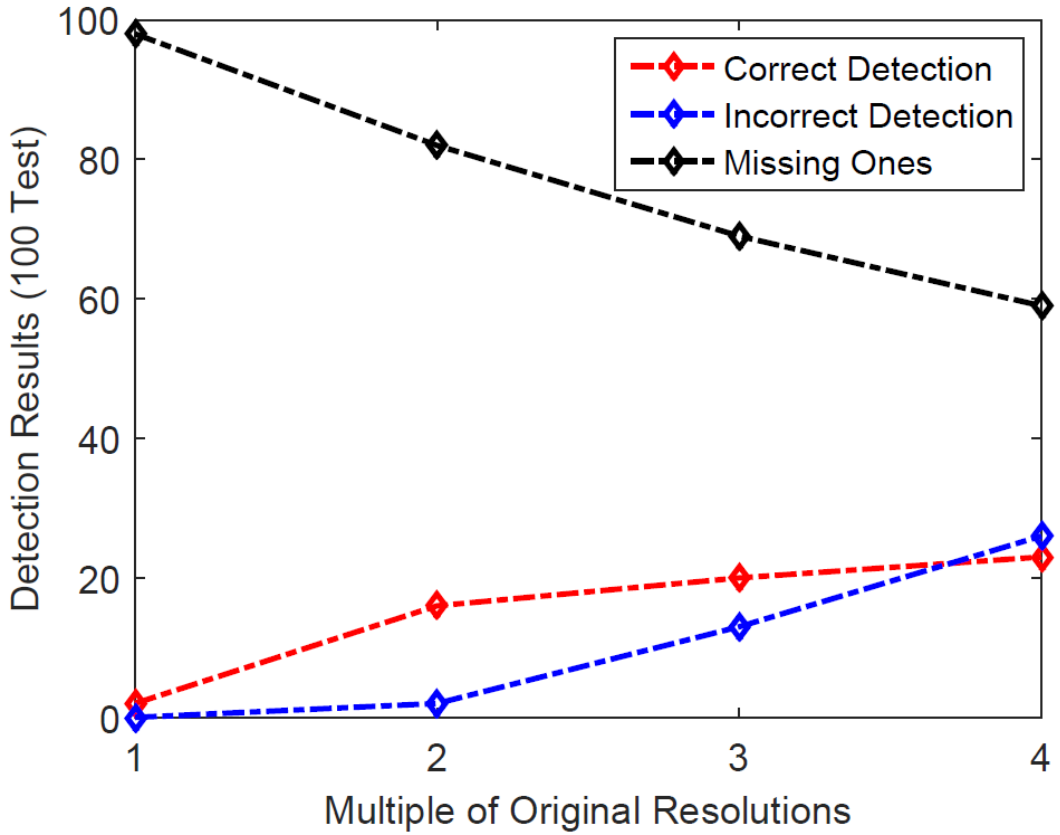






Preliminary Results: NEU Group

- 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

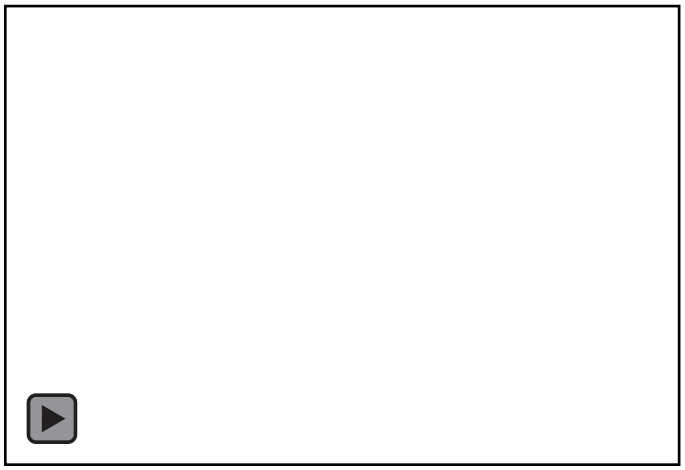
Face Recognition

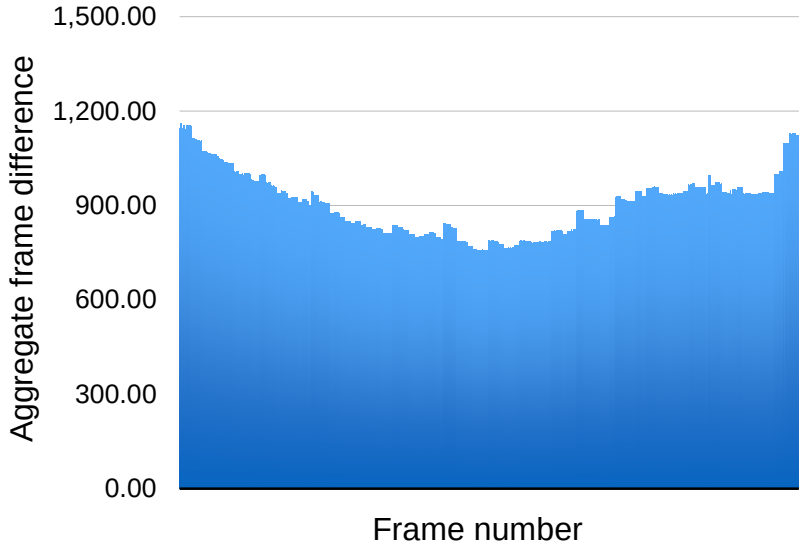
- Face Hallucination
  - Constructing higher-resolution faces for better feature extraction and matching
  - Currently determining viability for the Very Low Resolution case

Video Summarization: Notre Dame

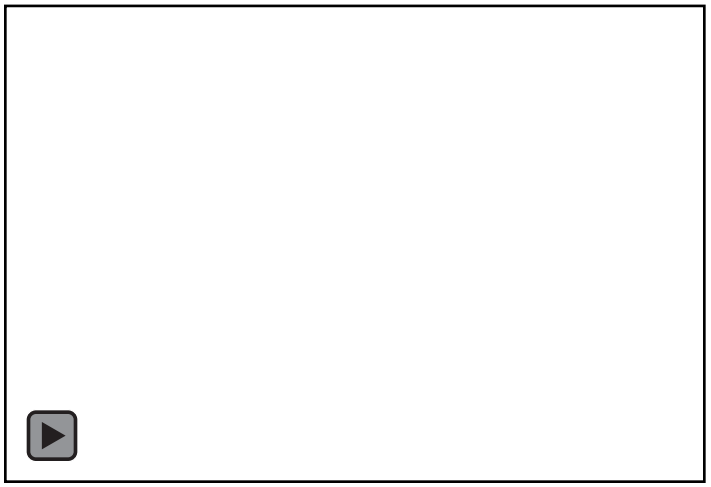
- Personnel: Nikita Amelchenko (CS undergraduate)
- Concept: remove video where nothing is happening
- Challenge: something is always happening
  - Noise
  - Timestamp/annotation
  - Other nonstationarities (sunlight)
- Approach
  - Baseline noise distribution
  - Masking
  - Thresholding difference

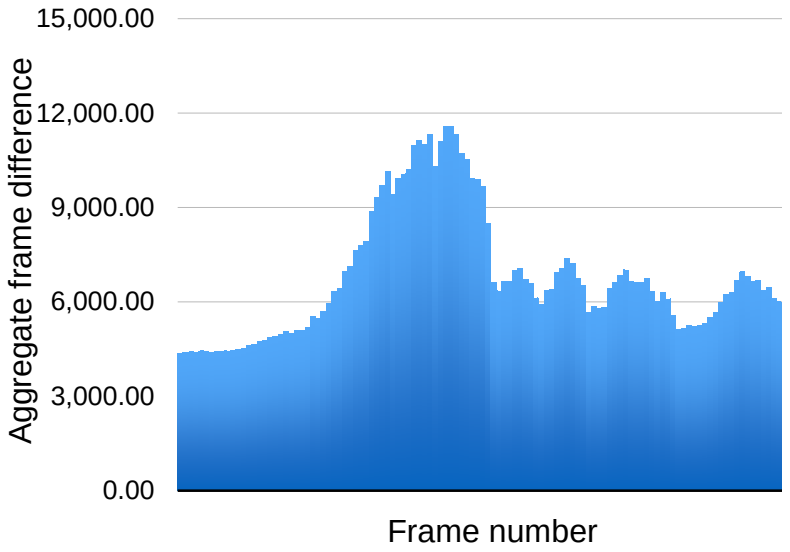
Baseline



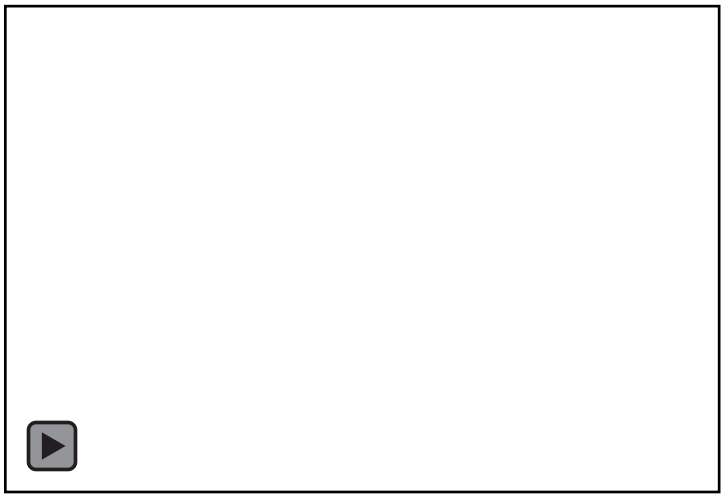

1,500.00
1,200.00
900.00
600.00
300.00
0.00
Aggregate frame difference
Frame number

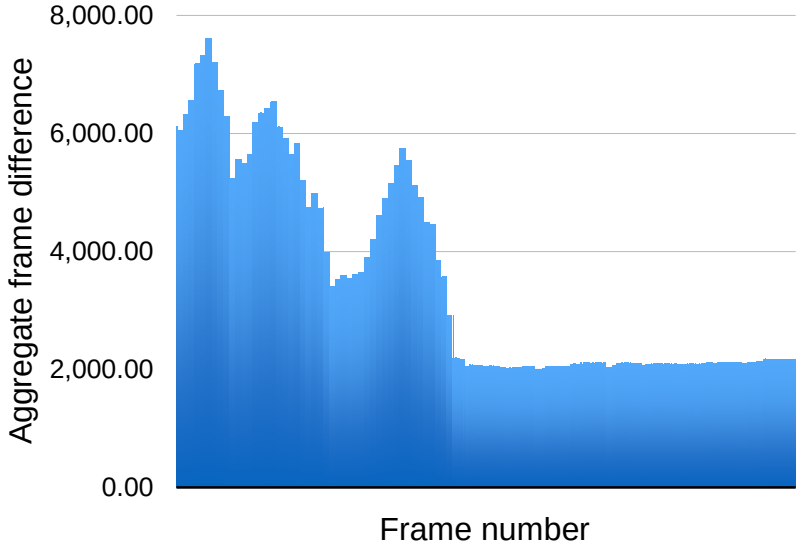
Motion




15,000.00
12,000.00
9,000.00
6,000.00
3,000.00
0.00
Aggregate frame difference
Frame number

Strong motion





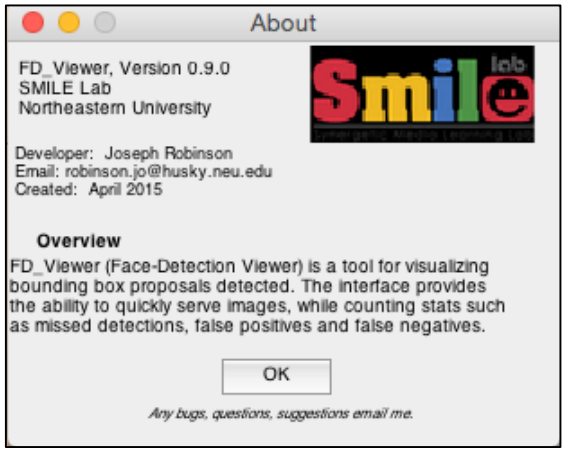
Annotation Tools: NEU Group

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:

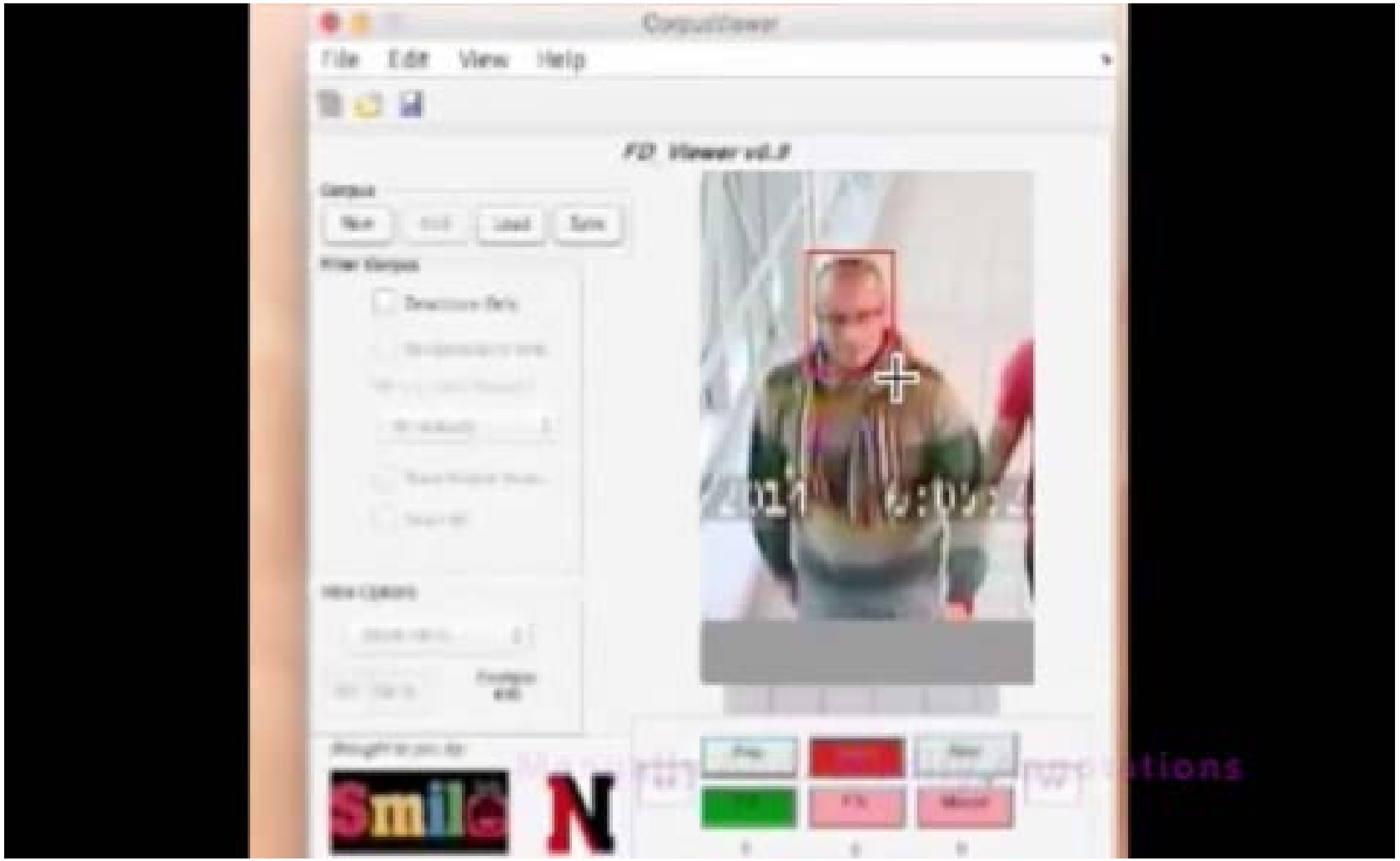
    <image_name, numFN, numFN, numMissed>
- **Easy** to create, save and load Corpus.





Annotation Tools

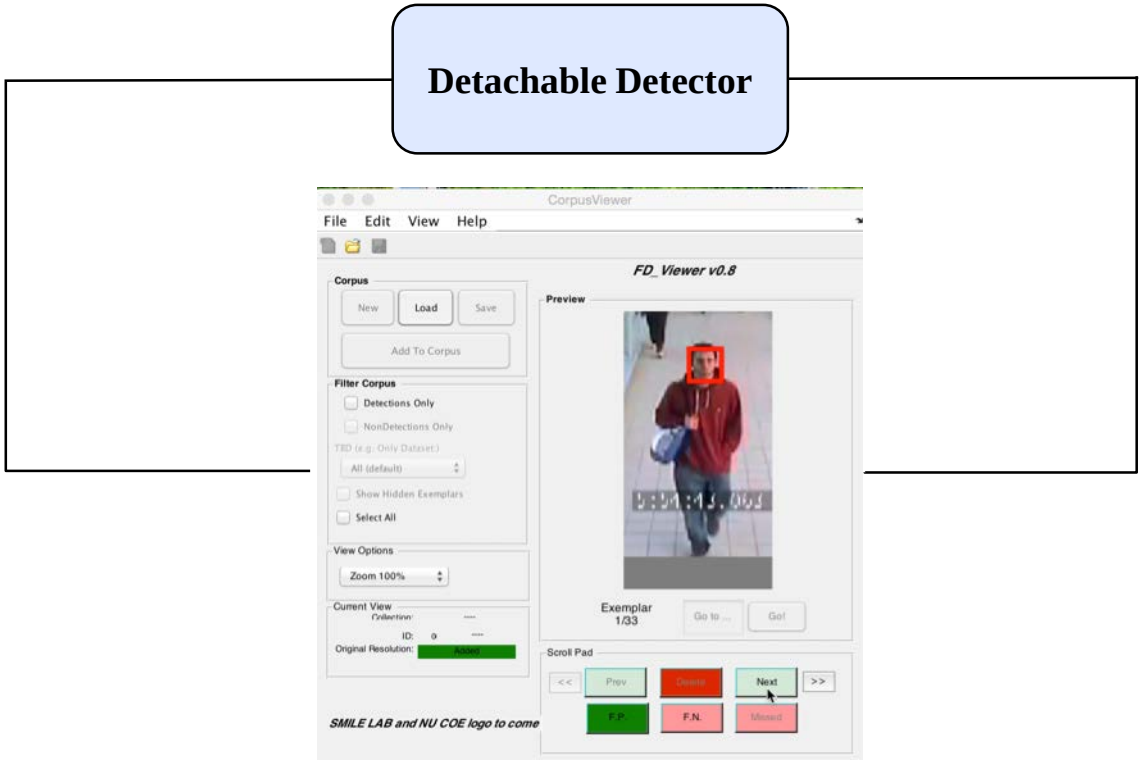
FDViewer: Video Demo



Annotation Tools

FDViewer: An Extendable (Modular) Framework

- **Easily attach** <u>any object detector of choice</u>.
  - **Change** detectors via dropdown menu.
  - **Analyze** and **Compare** different detectors.



Annotation Tools

Future Work

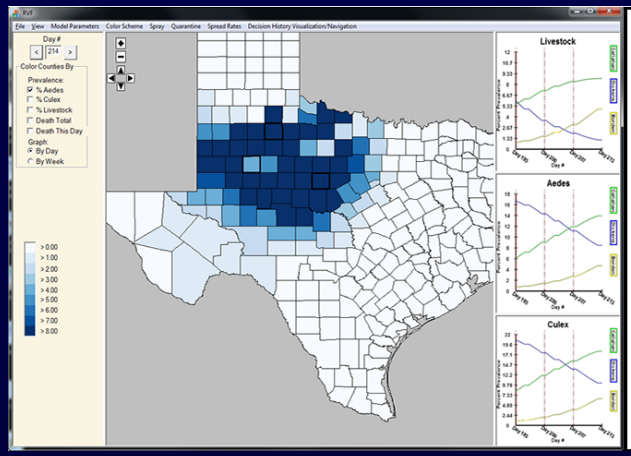
1. Enable Corpus merging.
   - Add user bookmarks.
2. Provide views (exemplar) filter.
   - View type specific results.
3. Allow images to be rescaled.
   - Reprocess scaled images.
4. Add ‘Delete’ to drop unwanted samples.

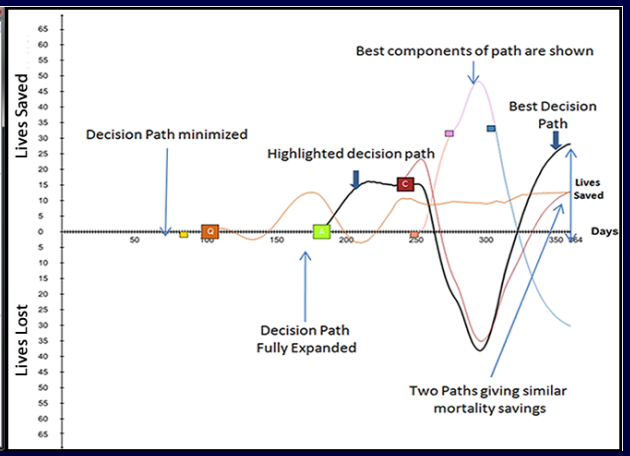


Future Work (face re-id)

- 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 vBOLO application
- Integrate face algorithms with body-based algorithms to improve ranking of correct candidates

Temporal/Geo-spatial Analytics of Epidemic Data and Forecasting Epidemic Spread





Shehzad Afzal

VACCINE - DHS Center of Excellence

Purdue University

(http://www.purdue.edu/discoverypark/vaccine/)

VACCINE

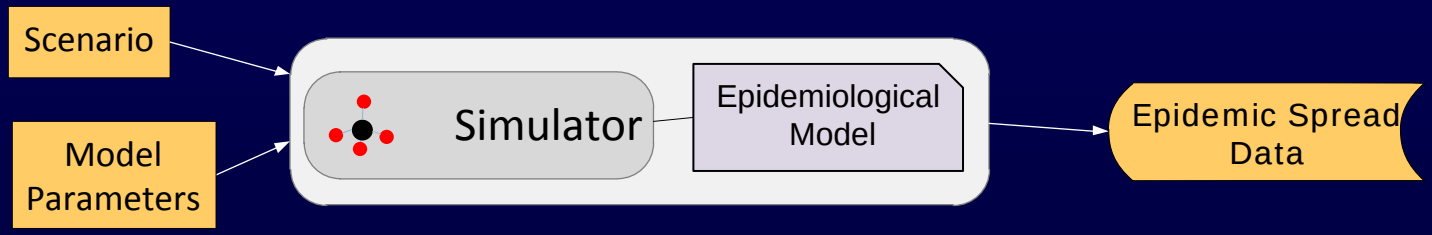
Motivation

- Officials must prepare complex plans to deal with Epidemic Outbreaks like:
  - Rift Valley fever (RVF), Pandemic Influenza, Dengue Fever
- Epidemic Response Planning
  - Table top exercises
  - Past epidemic response histories
  - Computer based Simulations
- Table top exercises & Past response histories have limitations

VACCINE

Motivation

- Computer based simulations



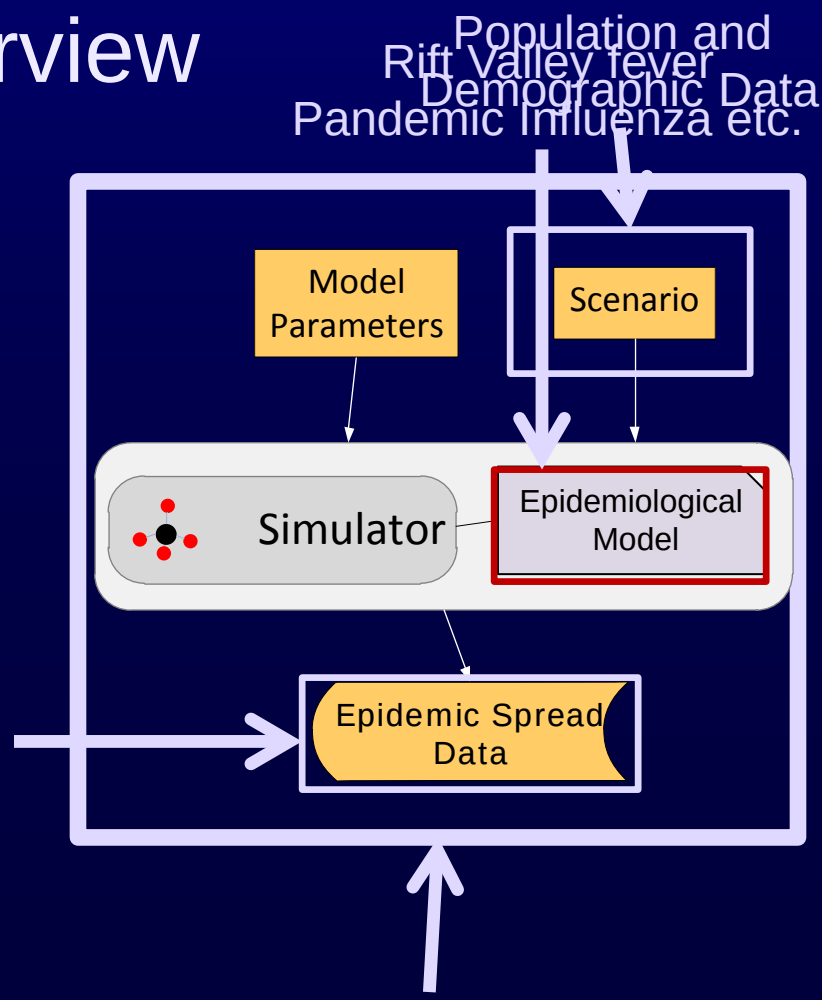
- Mitigative measures modify the input parameters
  - Examples: Pesticides, Quarantine etc.
- Simulation ‘response’ needs to be compared across ‘space & time’

Contribution

Interactive Decision Support Environment:

- Epidemic Models Exploration
- Interactively utilize mitigative response measures
- Decision History Visualization and Navigation View
  - links to multiple simulation runs
- Simulation replay and path exploration visualization
- Design and Evaluate Mitigative Response Strategy

System Overview

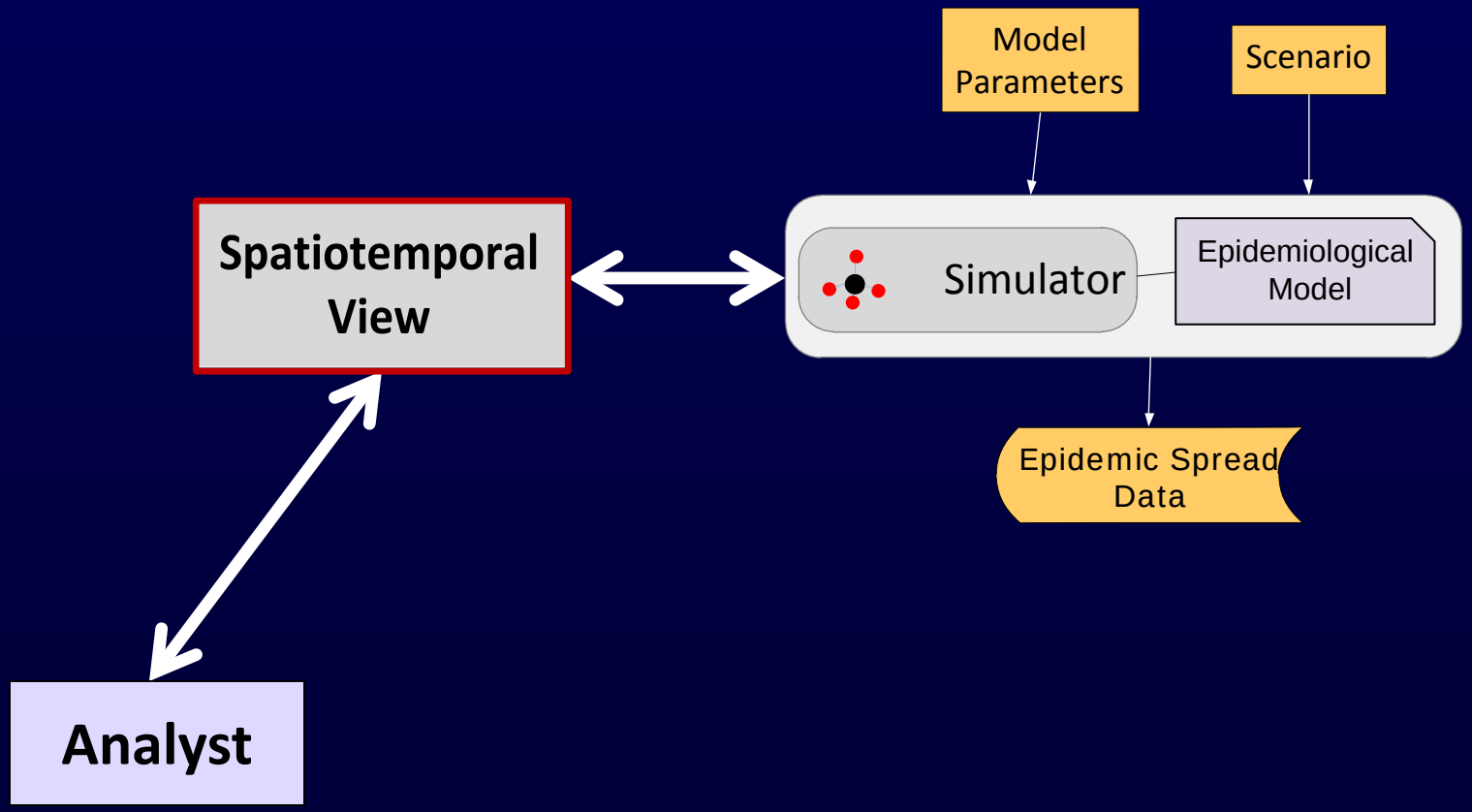


# of sick/dead
per population unit

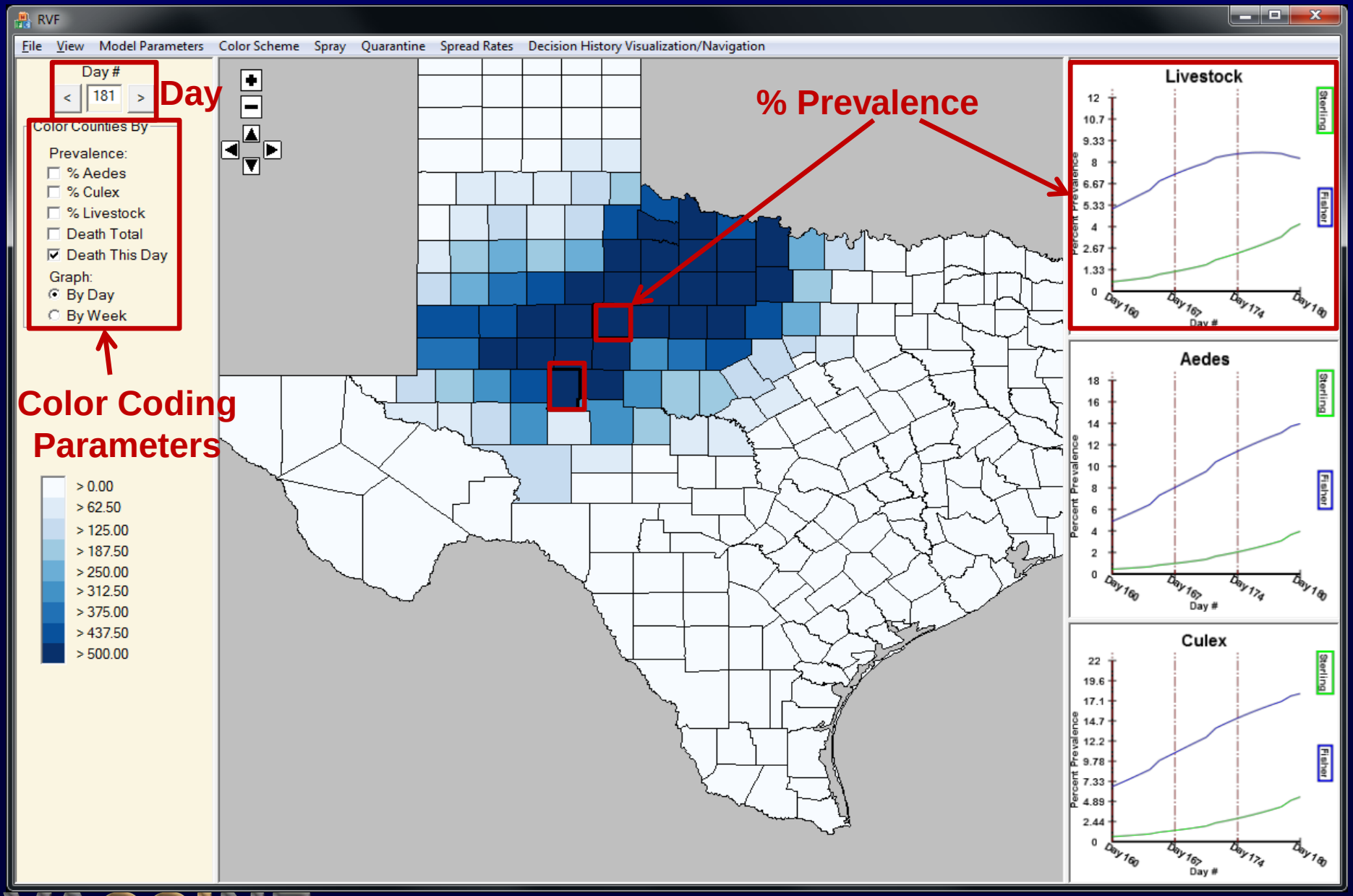
- Interchangeable
- Pre-Implemented

VACCINE

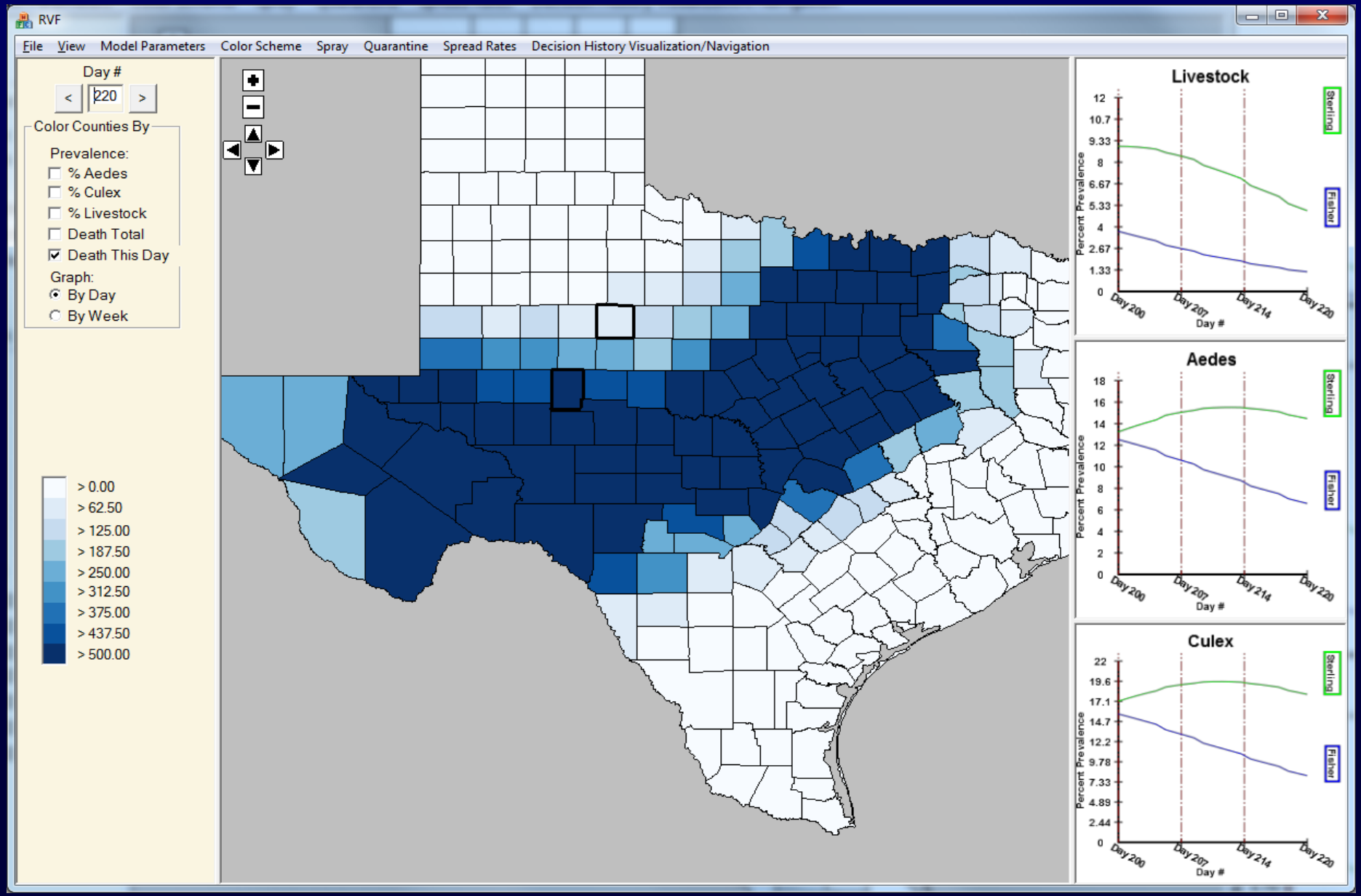
System Overview



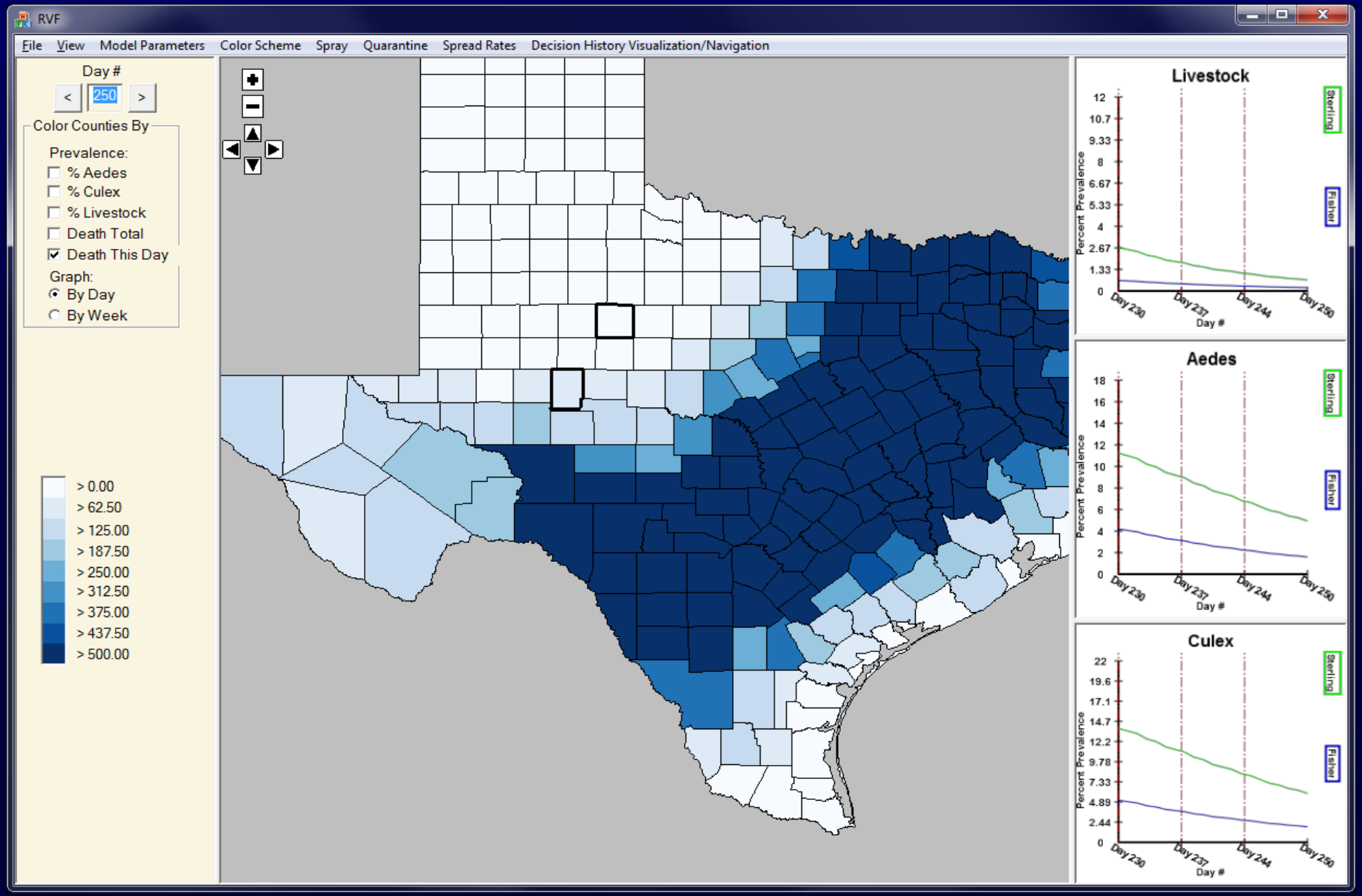
Spatiotemporal View (Rift Valley Fever)



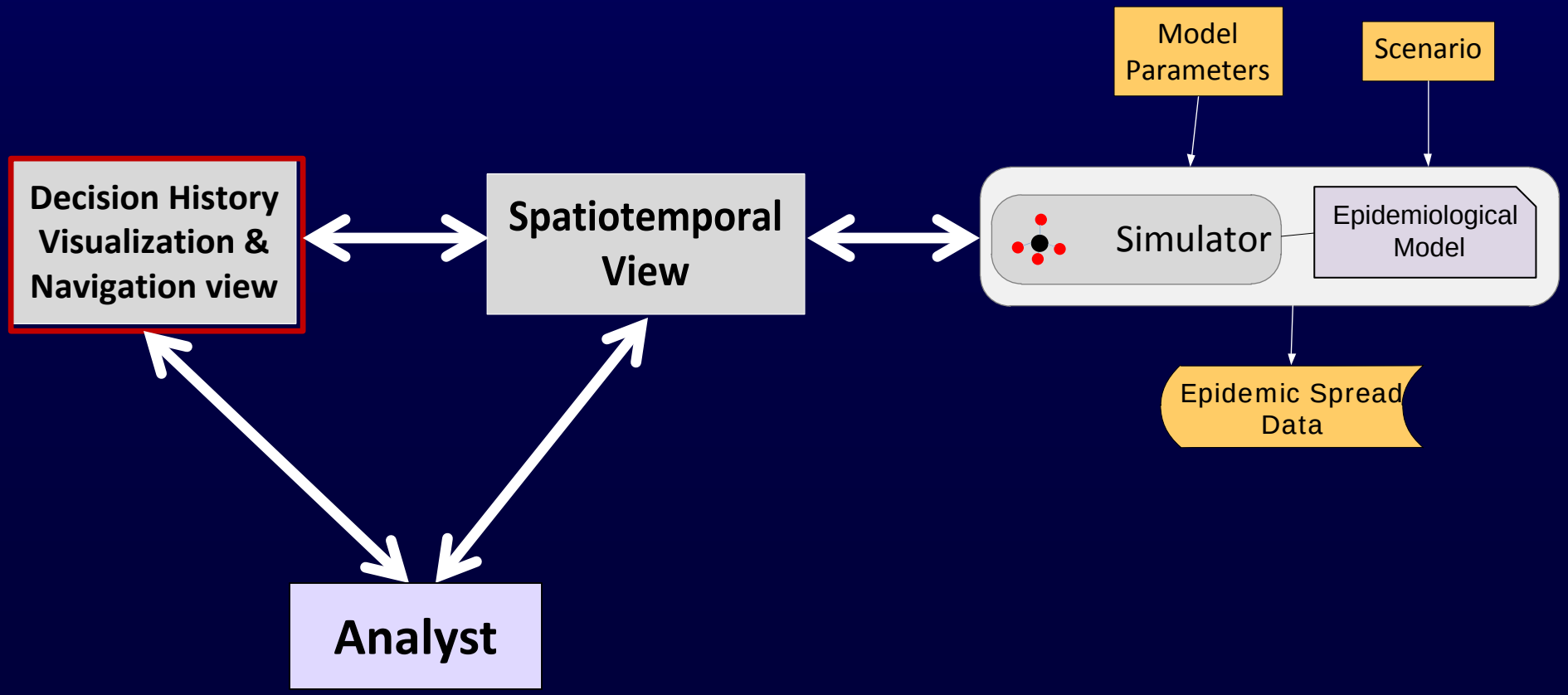
Spatiotemporal View (Rift Valley Fever)



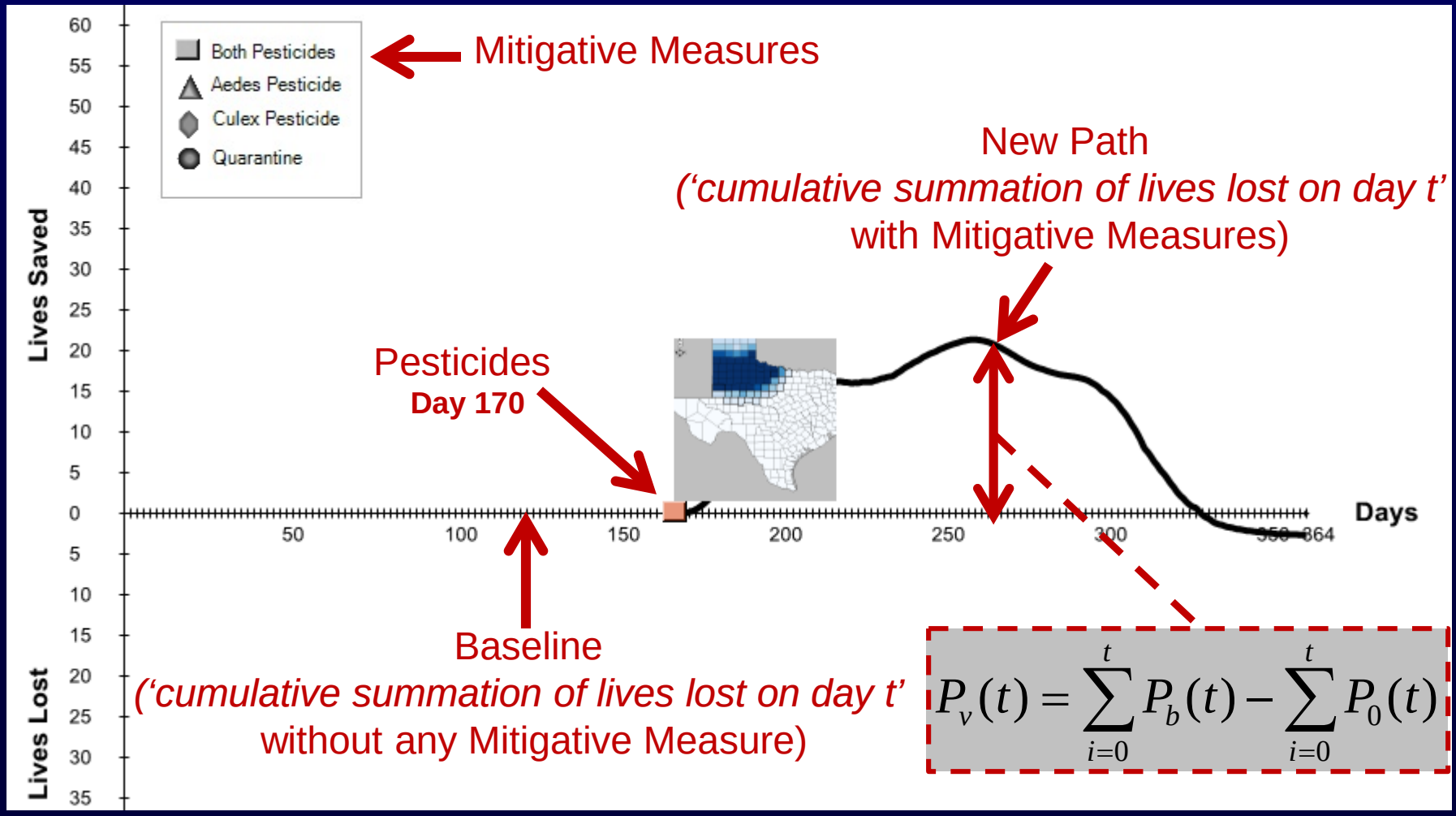
Spatiotemporal View (Rift Valley Fever)



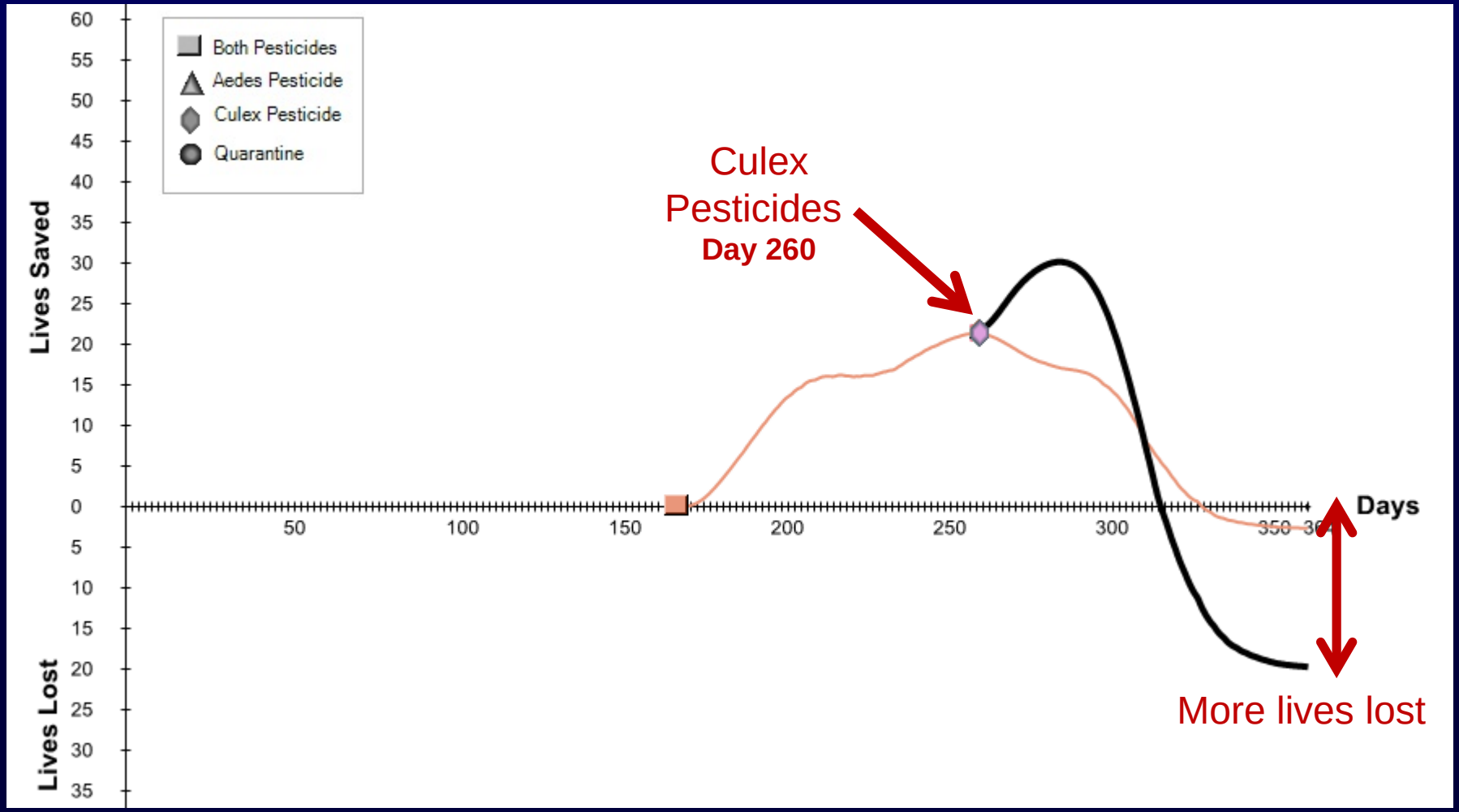
System Overview



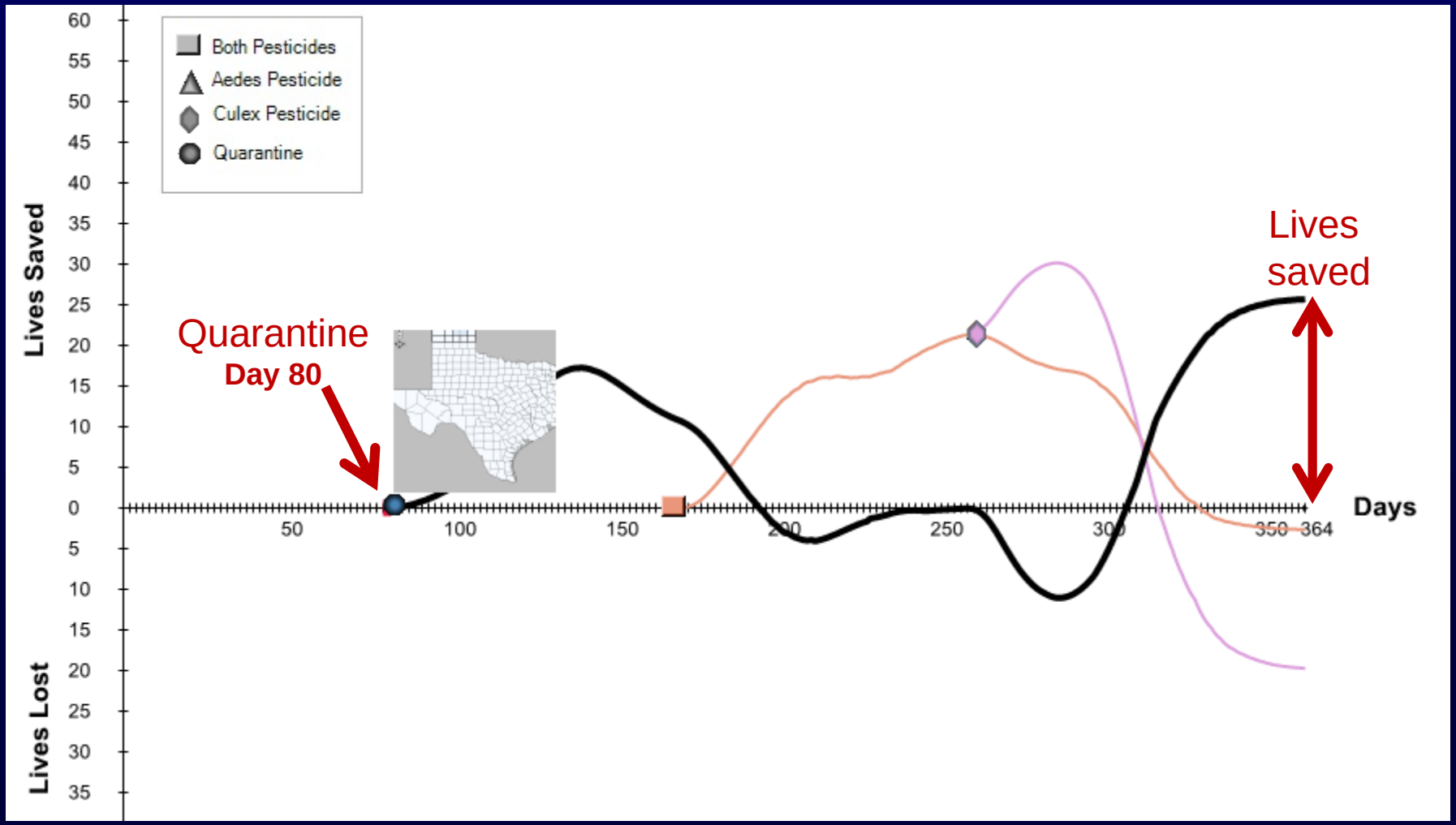
Decision History Visualization & Navigation View
Rift Valley Fever



Decision History Visualization & Navigation View
Rift Valley Fever



Decision History Visualization & Navigation View
Rift Valley Fever



VACCINE

Case Studies

Case Study 1 – Rift Valley Fever

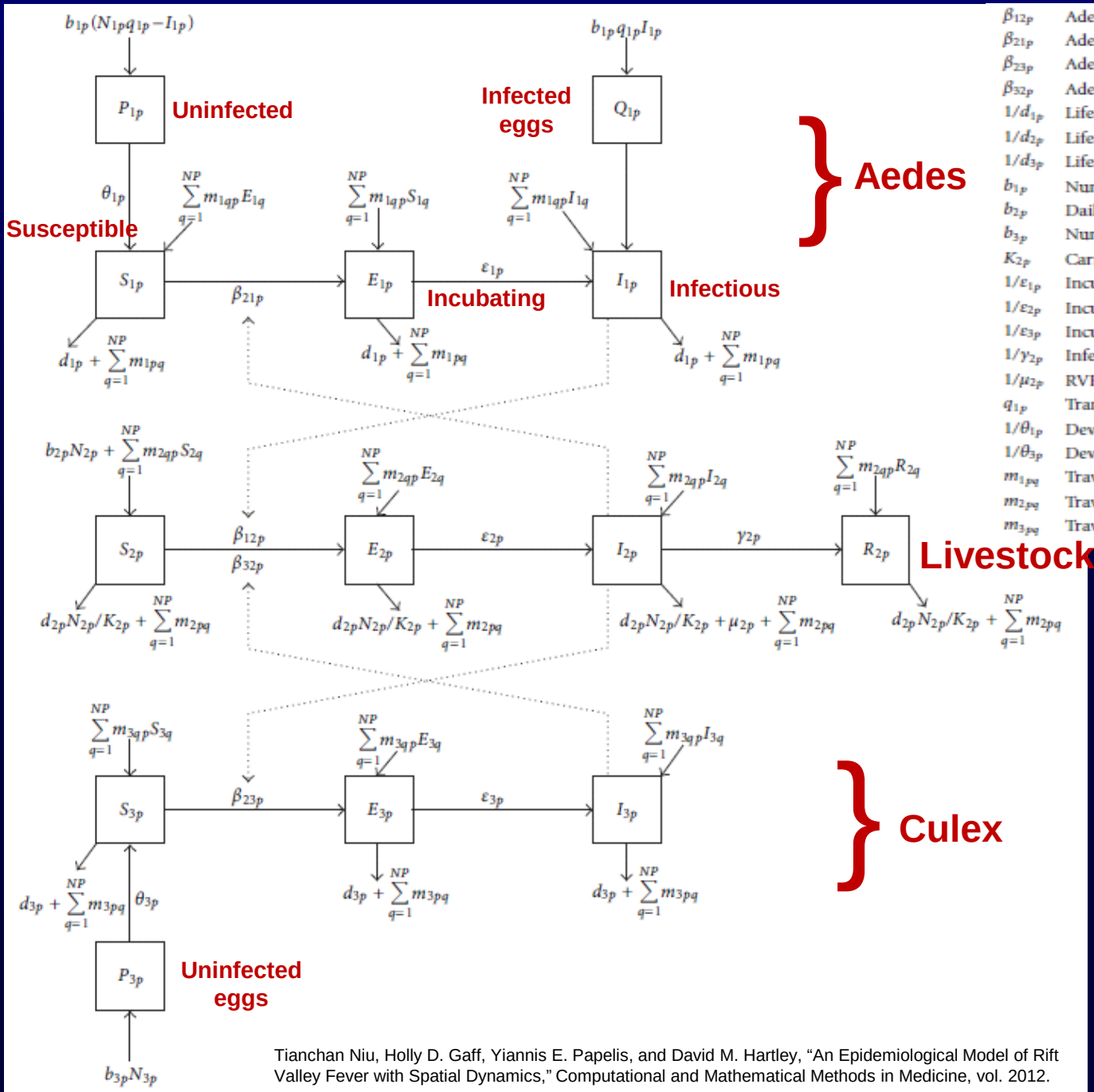
Description:

- Epidemiological model that simulates Rift Valley Fever spread in Texas[2]
- Mosquito Species (Aedes & Culex)
- Infects Cattle Population

Mitigative Response Measures:

1) Pesticides (kill Aedes and Culex Mosquito population)
2) Quarantine (Block transport of cattle into or out of the selected region)

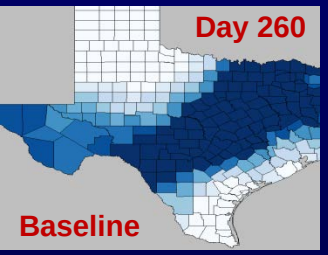
[2]GAFF H. D., HARTLEY D. M., LEAHY N. P.: An epidemiological model of Rift Valley Fever. Electron. J. Diff. Eqns. 115 (2007)

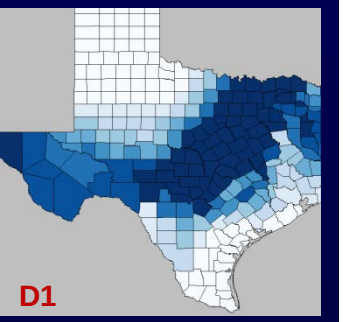


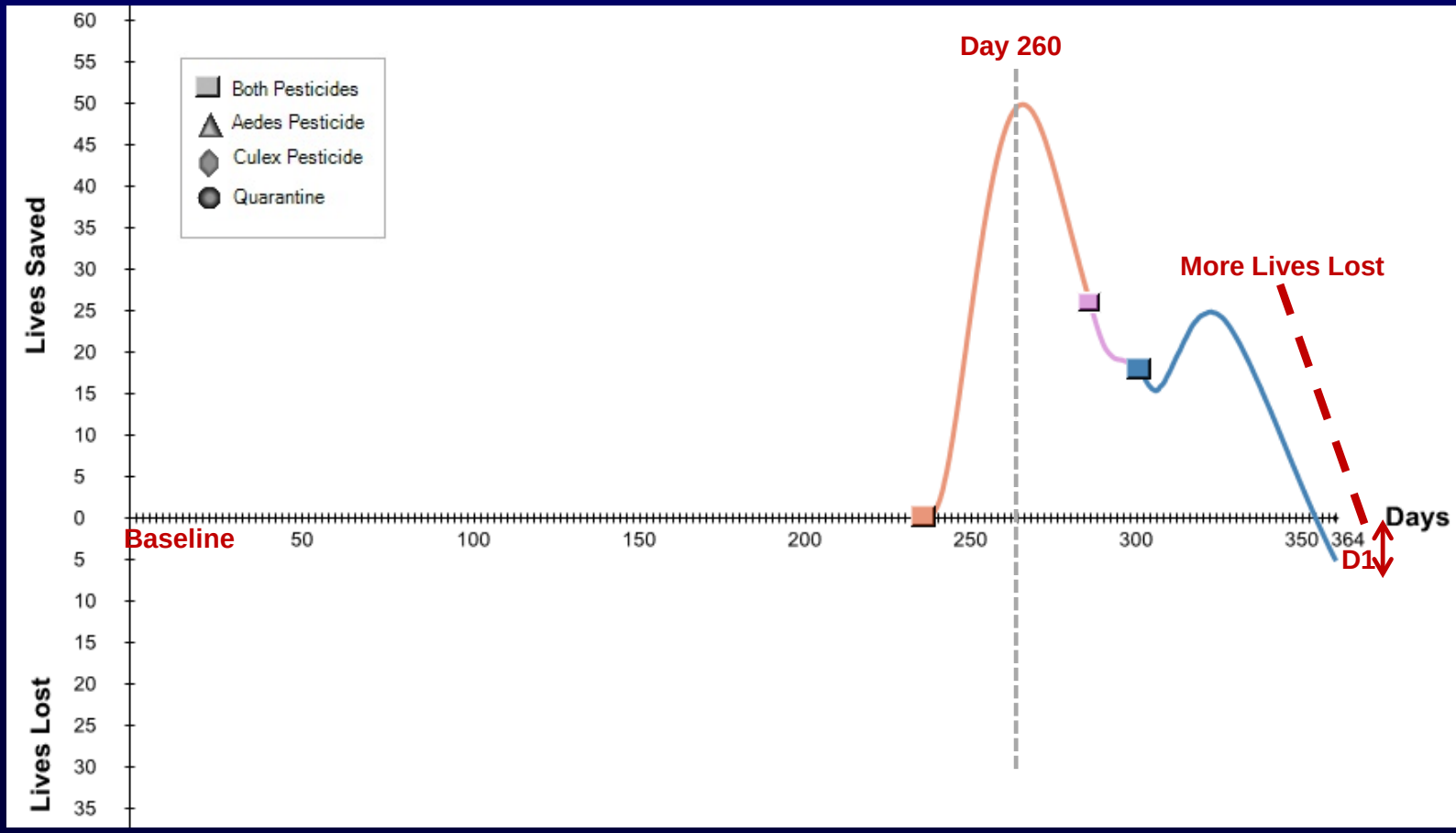
β_{12p}	Adequate contact rate from *Aedes* to livestock on patch p
β_{21p}	Adequate contact rate from livestock to *Aedes* on patch p
β_{23p}	Adequate contact rate from livestock to *Culex* on patch p
β_{32p}	Adequate contact rate from *Culex* to livestock on patch p
$1/d_{1p}$	Lifespan of *Aedes* mosquitoes on patch p
$1/d_{2p}$	Lifespan of livestock on patch p
$1/d_{3p}$	Lifespan of *Culex* mosquitoes on patch p
b_{1p}	Number of *Aedes* eggs laid per day on patch p
b_{2p}	Daily birthrate of livestock on patch p
b_{3p}	Number of *Culex* eggs laid per day on patch p
K_{2p}	Carrying capacity of livestock on patch p
$1/\varepsilon_{1p}$	Incubation period in *Aedes* mosquitoes on patch p
$1/\varepsilon_{2p}$	Incubation period in livestock on patch p
$1/\varepsilon_{3p}$	Incubation period in *Culex* mosquitoes on patch p
$1/\gamma_{2p}$	Infectiousness period in livestock on patch p
$1/\mu_{2p}$	RVF mortality rate in livestock on patch p
q_{1p}	Transovarial transmission fraction in *Aedes* on patch p
$1/\theta_{1p}$	Development time of *Aedes* on patch p
$1/\theta_{3p}$	Development time of *Culex* on patch p
m_{1pq}	Travel rate of *Aedes* mosquitoes from patch p to patch q
m_{2pq}	Travel rate of livestock from patch p to patch q
m_{3pq}	Travel rate of *Culex* mosquitoes from patch p to patch q

Tianchan Niu, Holly D. Gaff, Yiannis E. Papelis, and David M. Hartley, "An Epidemiological Model of Rift Valley Fever with Spatial Dynamics," Computational and Mathematical Methods in Medicine, vol. 2012.

Case Study 1 – Rift Valley Fever





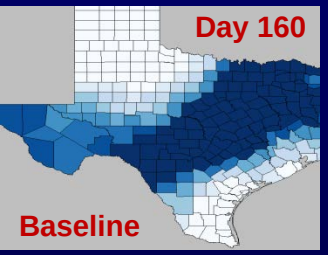


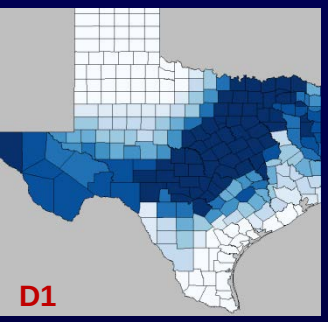
Decision Path D1

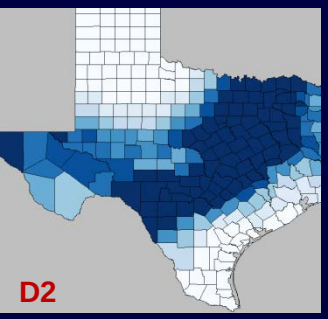
Spray Both Pesticides **Day 240, 290, 305**

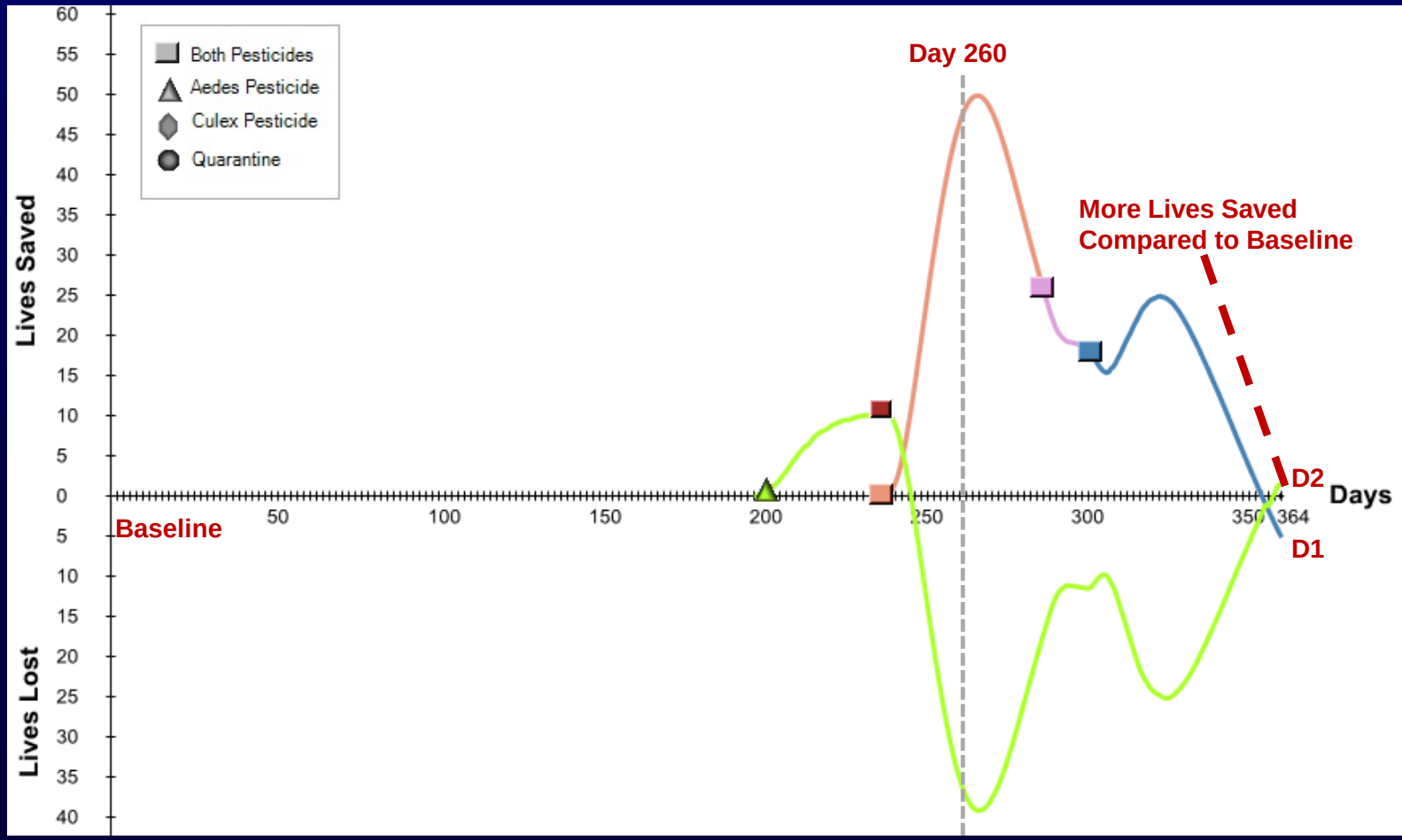
VACCINE

Case Study 1 – Rift Valley Fever







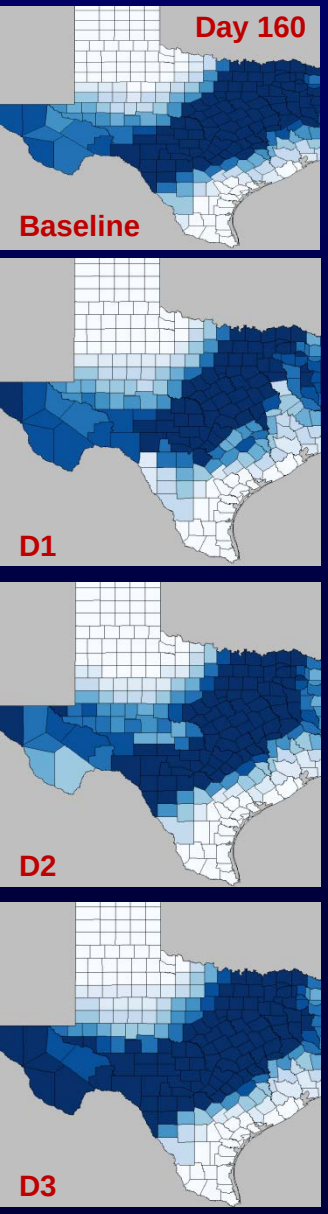


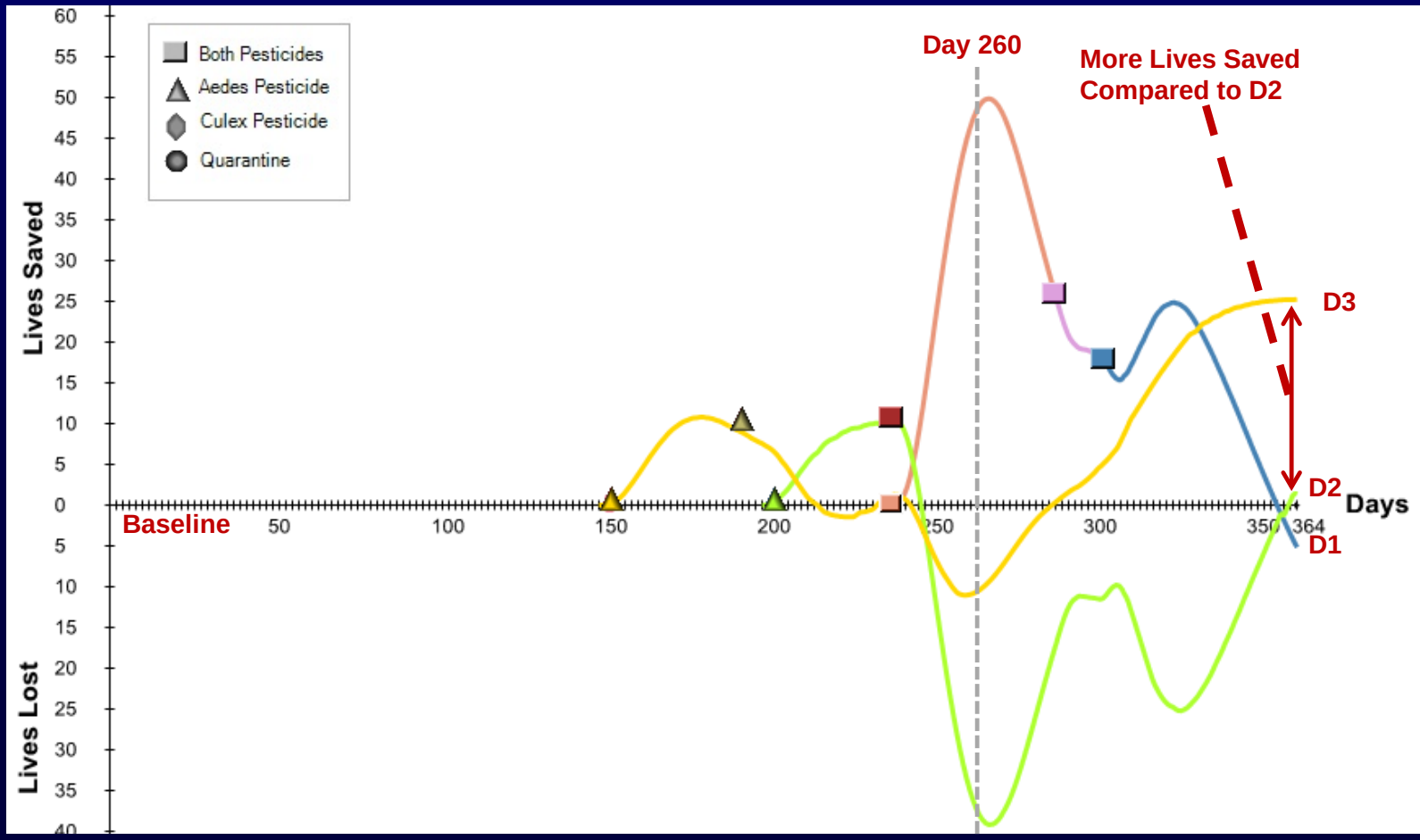
Decision Path D2

Spray	**Aedes Pesticides**	**Day 200**
	Both Pesticides	**Day 240**

VACCINE

Case Study 1 – Rift Valley Fever



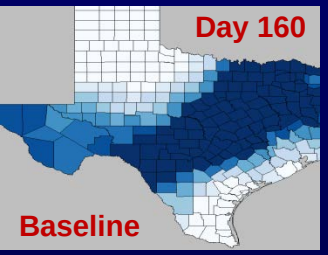


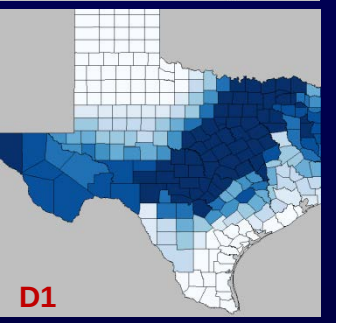
Decision Path D3

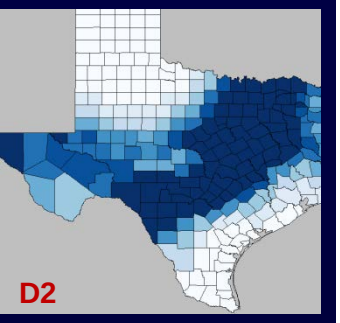
Spray	Aedes Pesticides	Day 150
	Aedes Pesticides	Day 190
	Culex Pesticides	Day 320

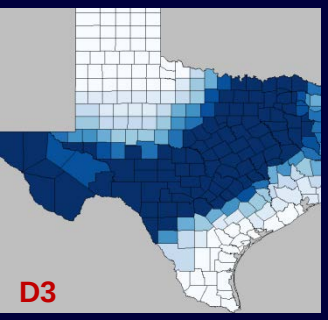
VACCINE

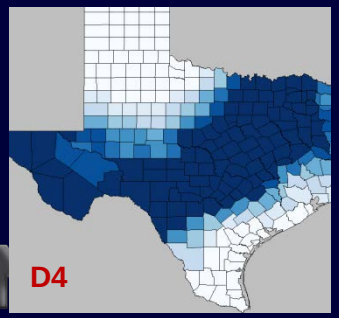
Case Study 1 – Rift Valley Fever

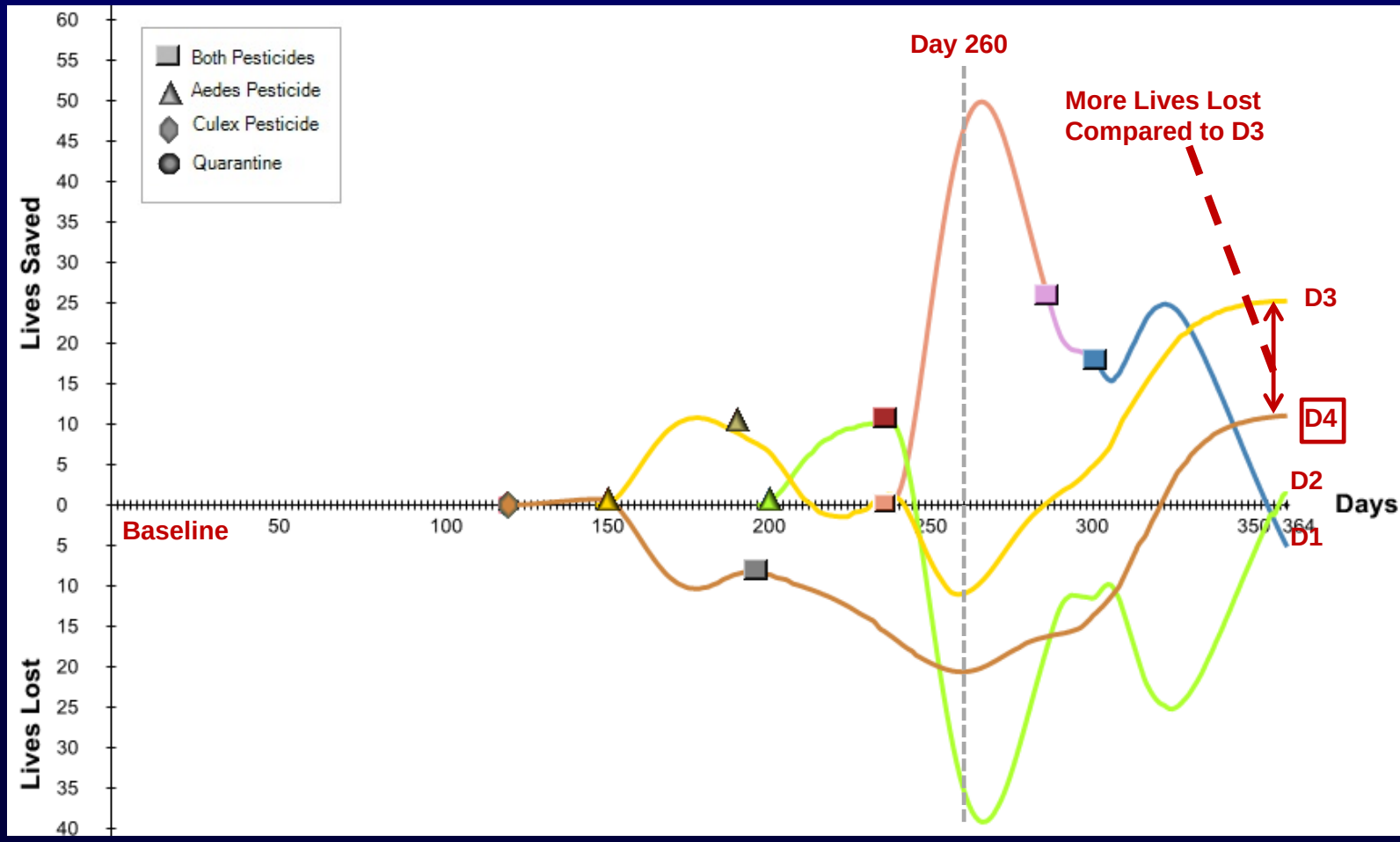








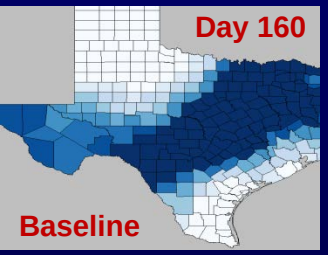


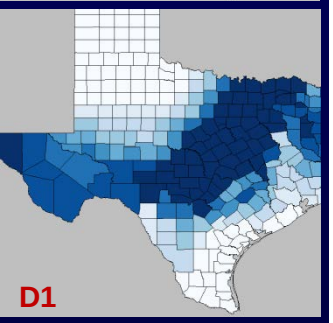


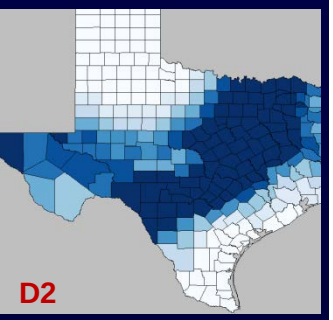
Decision Path D4

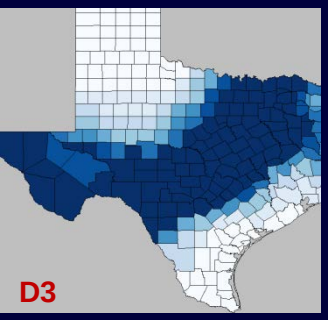
Spray	Culex	Pesticides	Day 120
	Both	Pesticides	Day 200

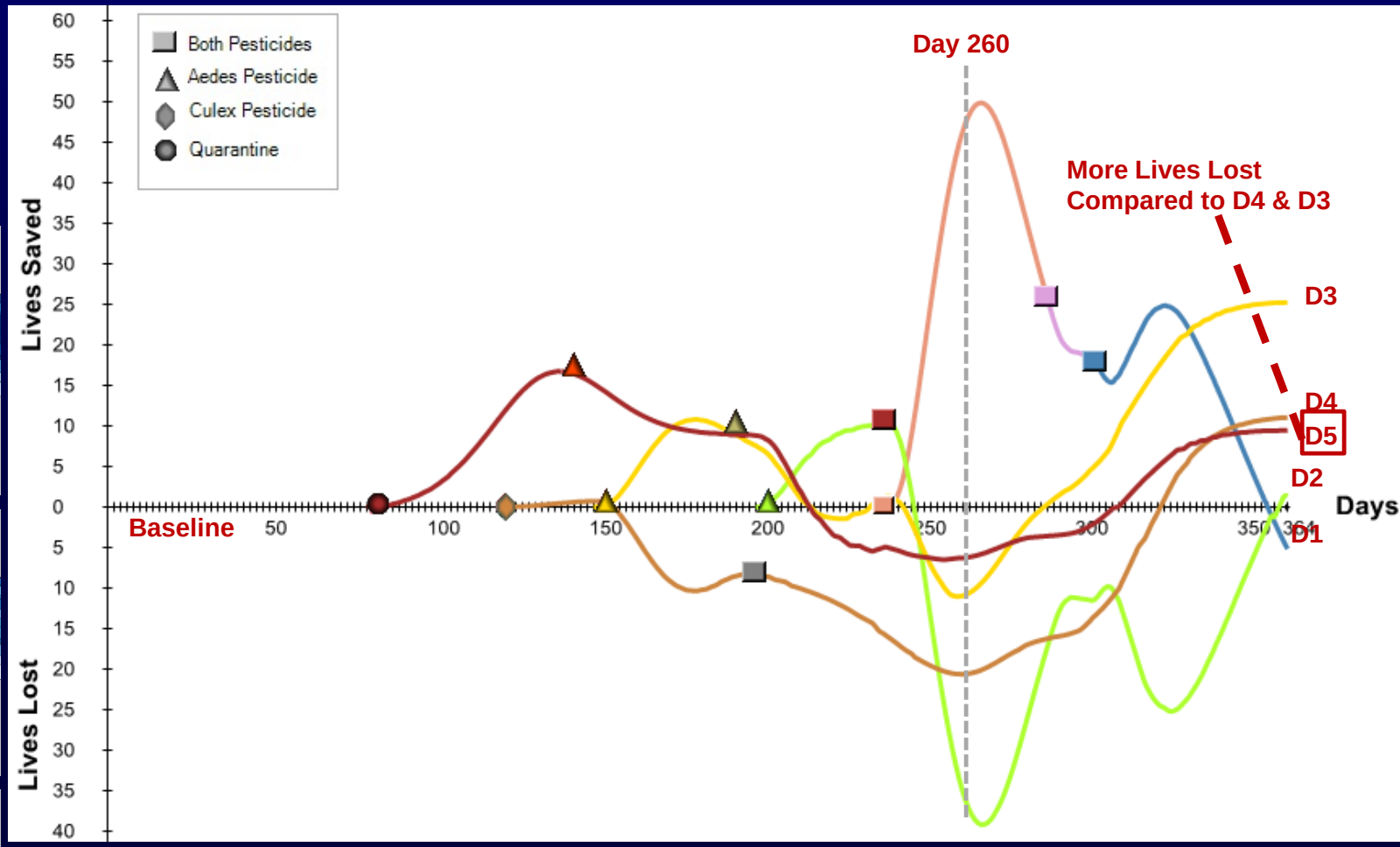
Case Study 1 – Rift Valley Fever

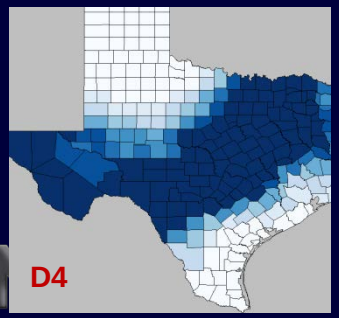






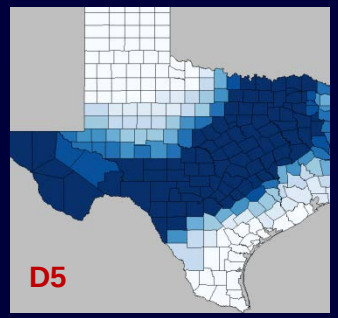




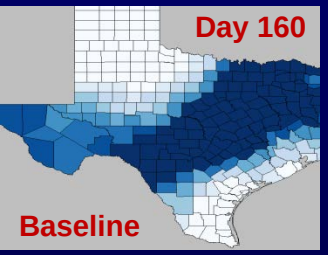


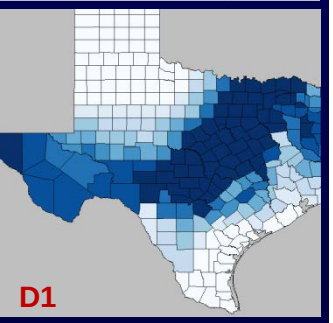
Decision Path D5

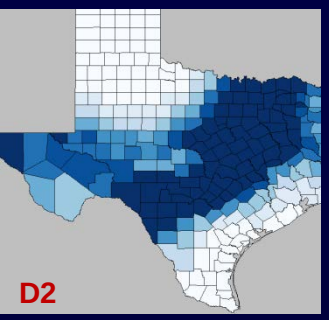
Quarantine	Day 80
Spray Aedes Pesticides	Day 140
Spray Both Pesticides	Day 260

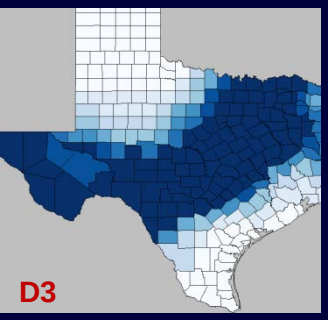


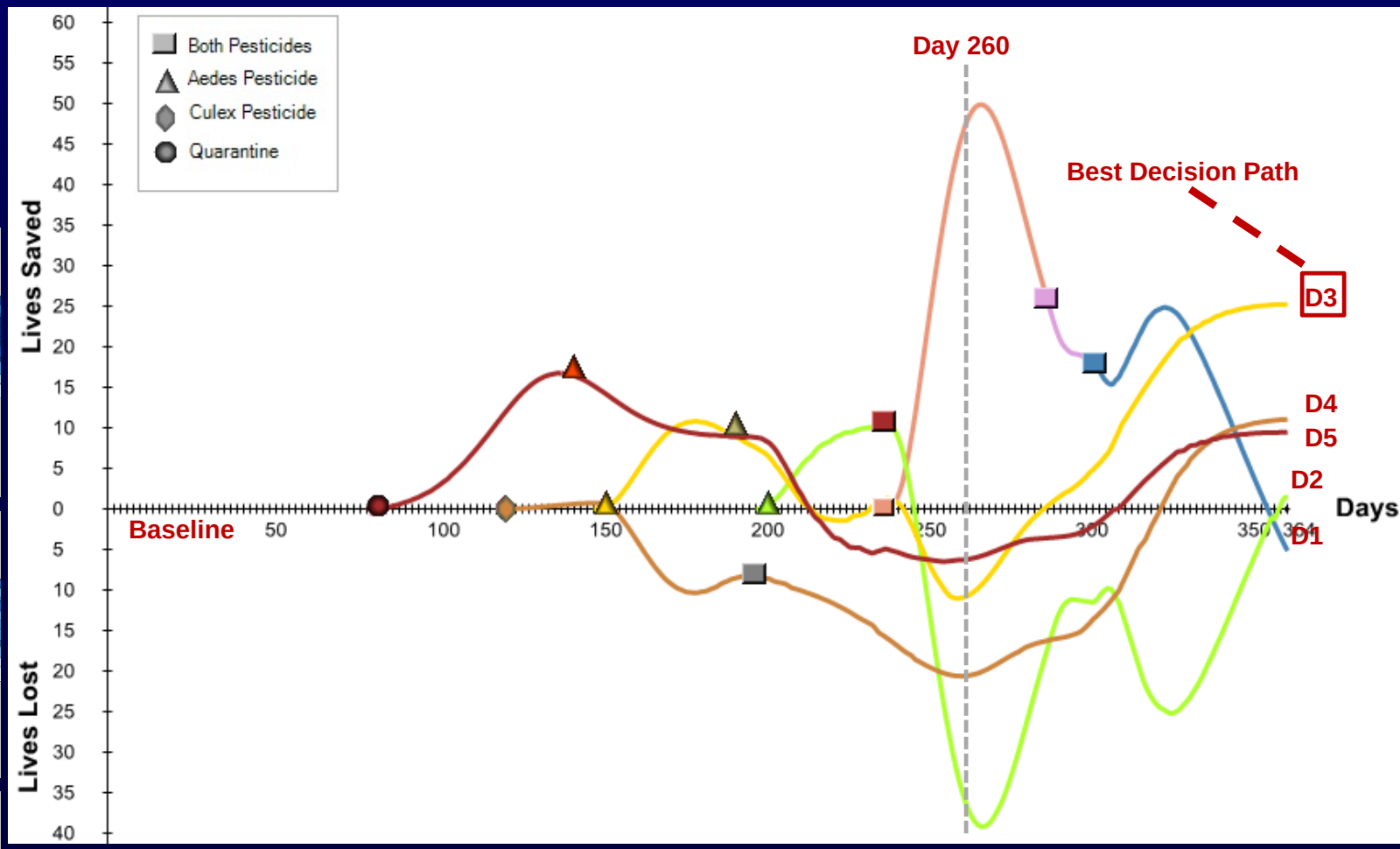
Case Study 1 – Rift Valley Fever

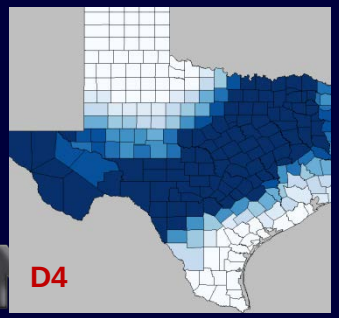






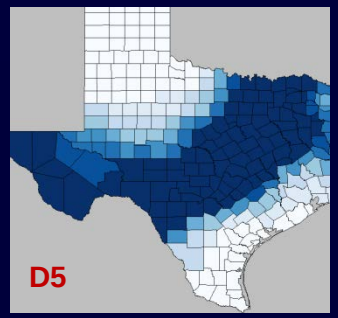






Best Decision Path D3

Spray	Aedes Pesticides	Day 150
	Aedes Pesticides	Day 190
	Culex Pesticides	Day 320



VACCI

Demo Video

Case Study 2 – Pandemic Influenza

Description:

- Epidemic Model that simulates the spread of a pandemic influenza across the US[1]
- User defined point source location
- Incorporates airport traffic models

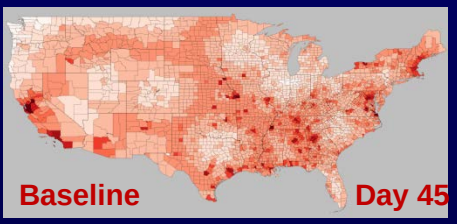
Mitigative Response Measures:

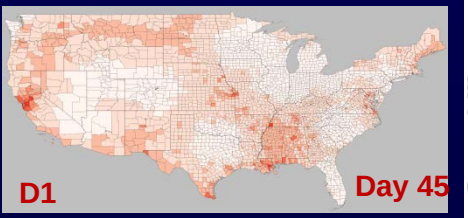
1) Media Alerts
2) Strategic national stockpile deployment (SNS)
3) School Closure

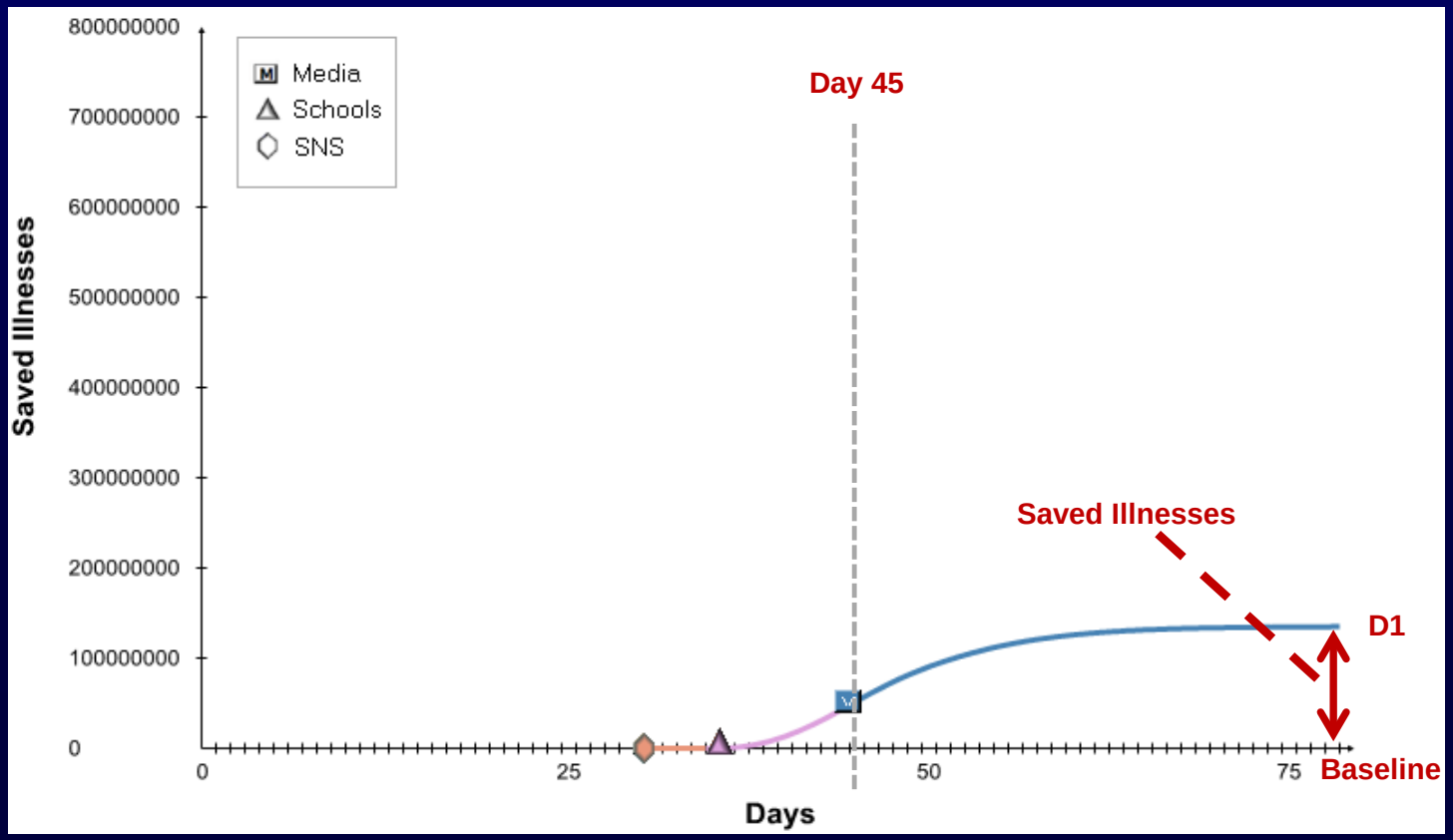
---

[1]R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders. A pandemic influenza modeling and visualization tool. Journal of Visual Languages and Computing (2011).

Case Study 2 – Pandemic Influenza





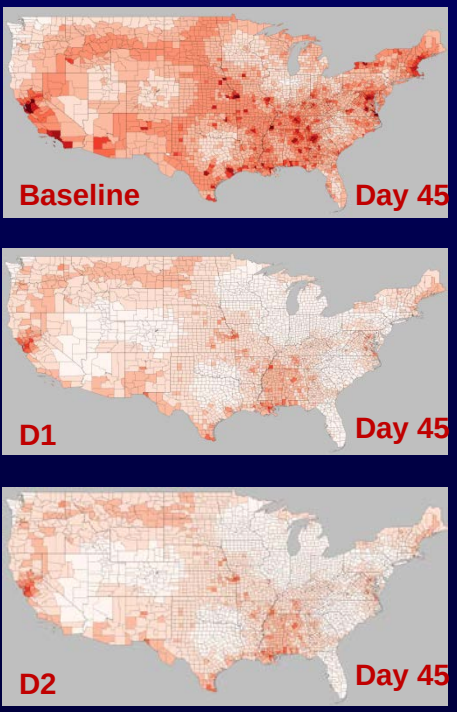


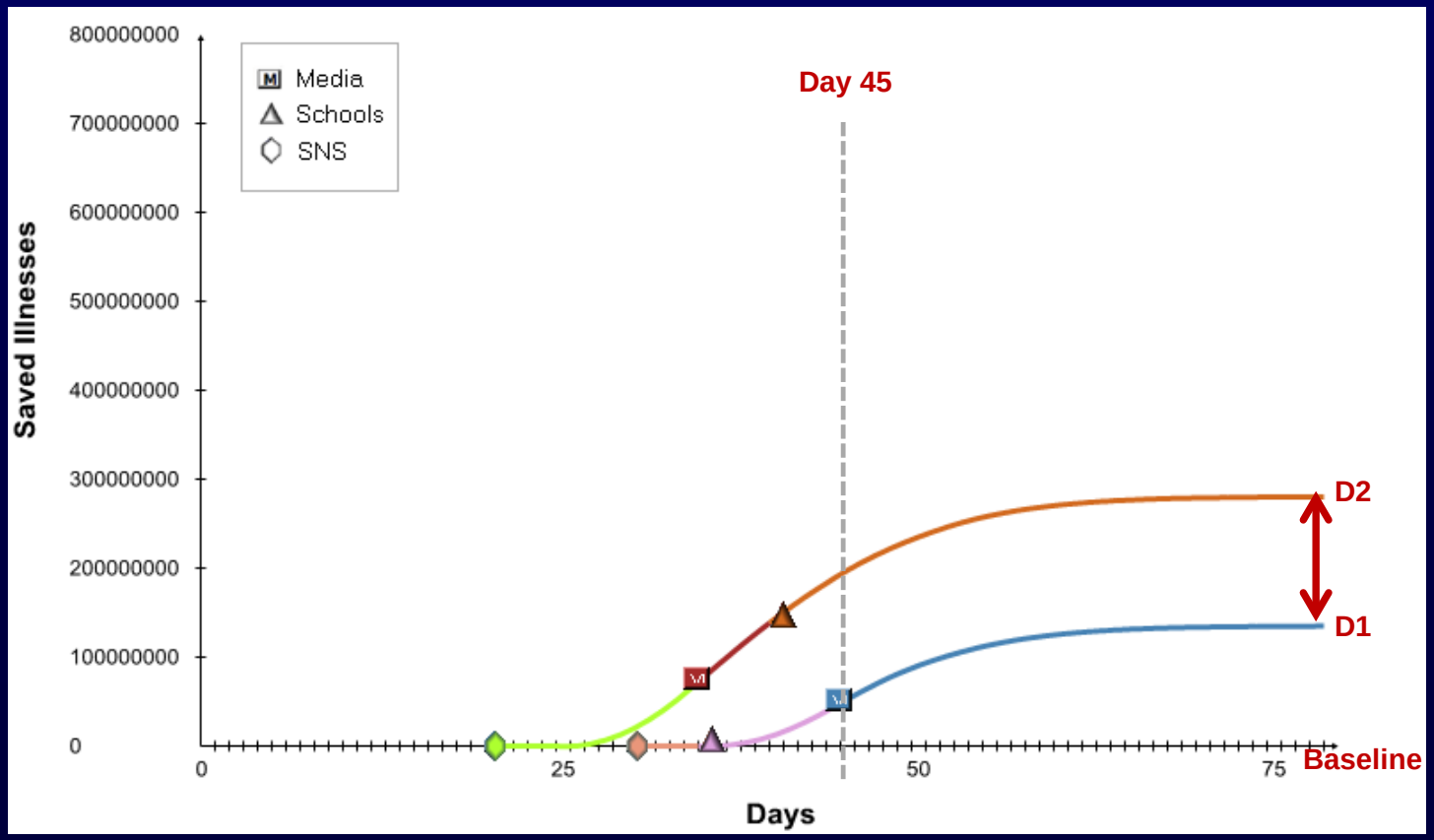
Decision Path D1

SNS Day 30 **School Closure Day 35** **Media Day 45**

VACCINE

Case Study 2 – Pandemic Influenza



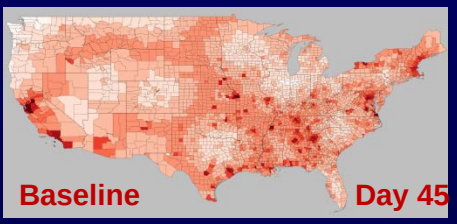


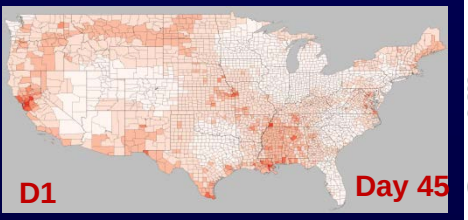
Decision Path D2

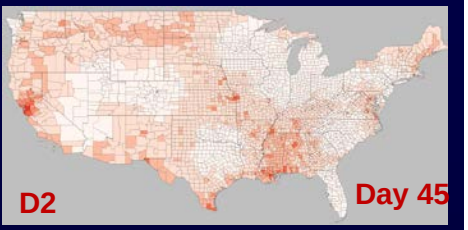
SNS Day 20 **Media Day 34** **School Closure Day 40**

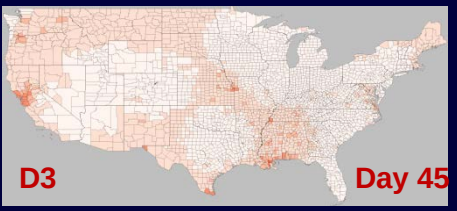
VACCINE

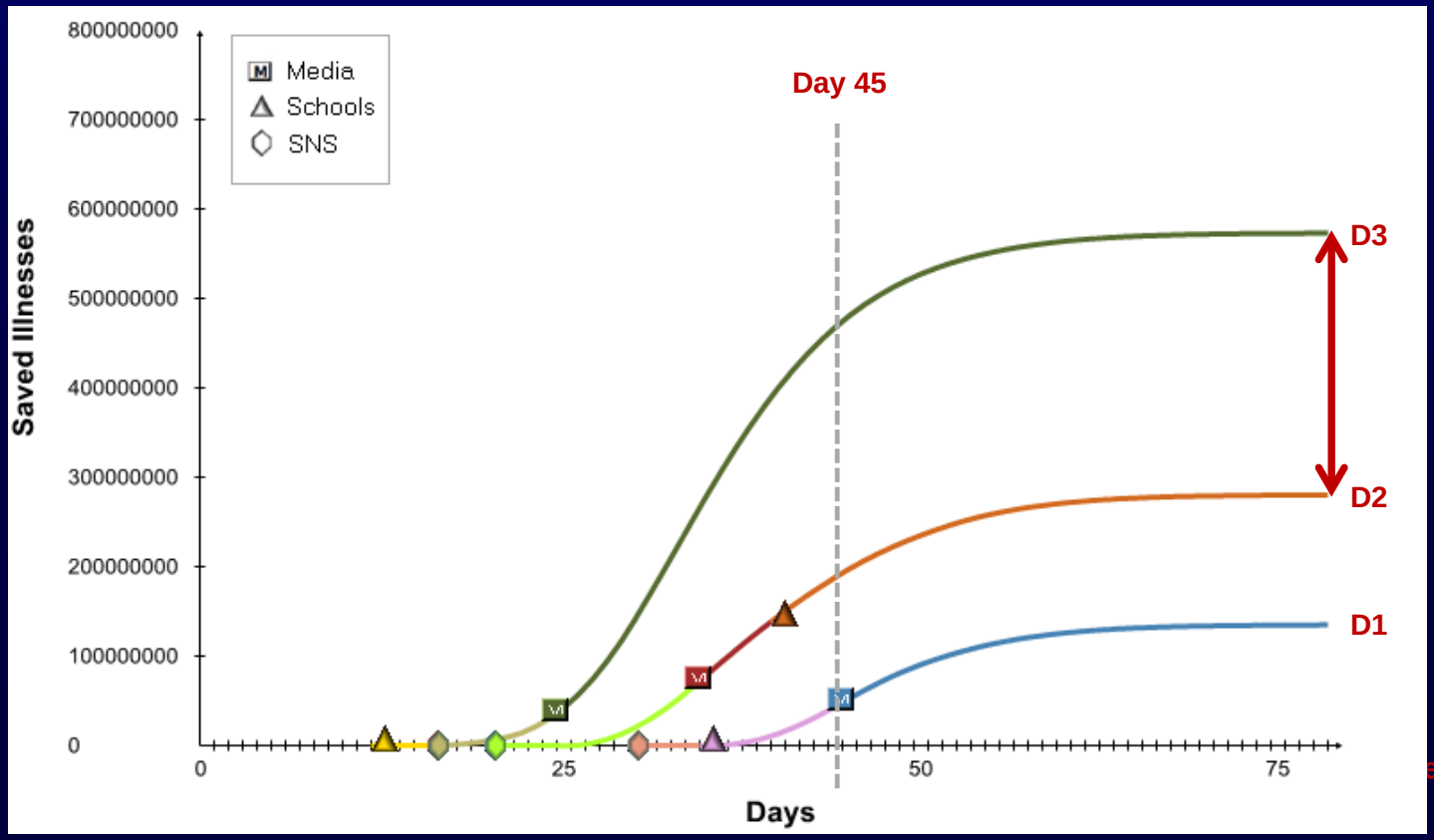
Case Study 2 – Pandemic Influenza









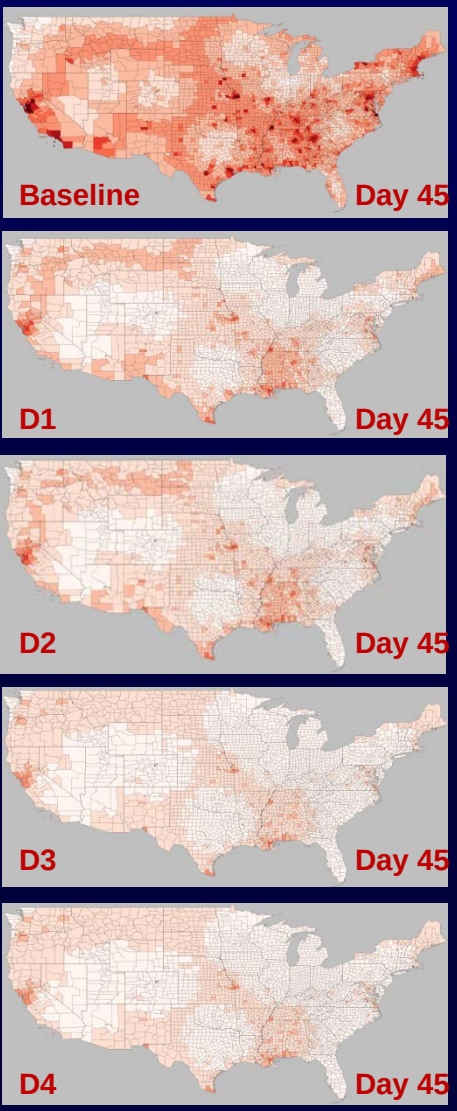


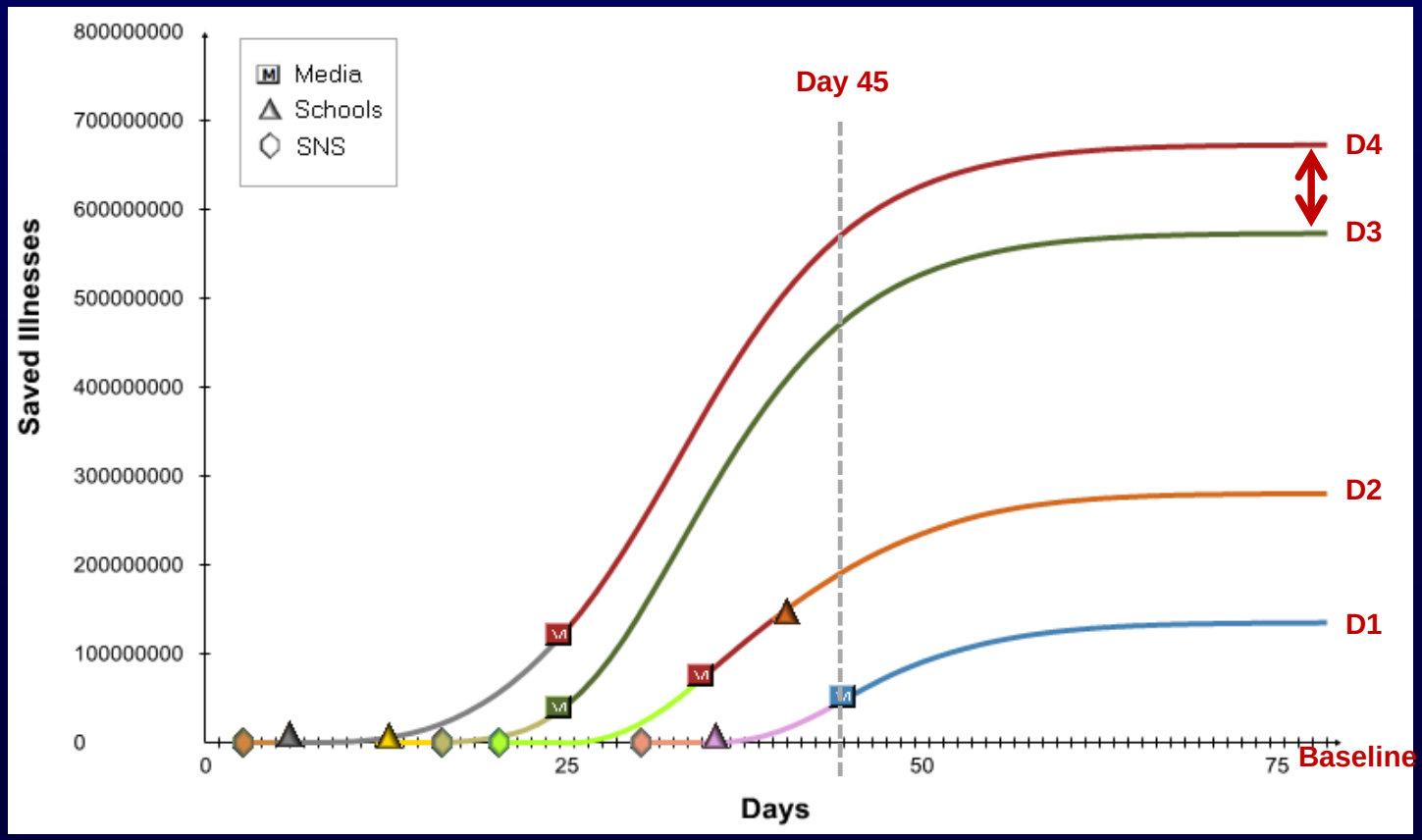
Decision Path D3

School Closure Day 12 **SNS Day 16** **Media Day 25**

VACCINE

Case Study 2 – Pandemic Influenza





Decision Path D4

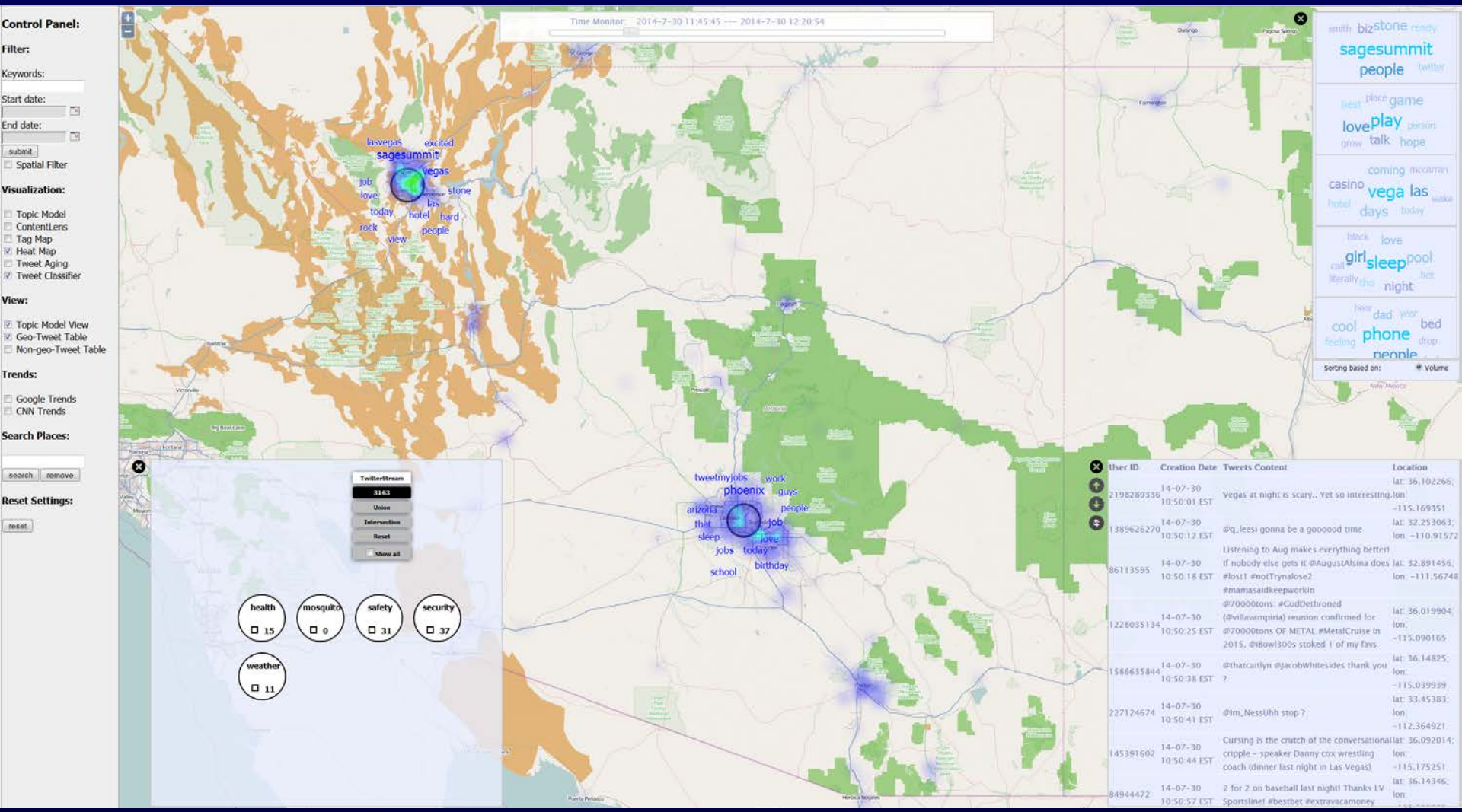
SNS Day 2 **School Closure Day 6** **Media Day 25**

VACCINE

Future Extensions

- Adaptation of RVF Model for Dengue Fever
- Integrate financial models to study the impact of disease outbreaks on local economy
- Combine data form other sources like social media data and actual outbreak data

Social Media Analytics and Reporting Toolkit (SMART)



VACCINE

SMART Features

- Interactive analysis and visualization of large volumes of Twitter Data
- Load live streaming data as well as historical data for specific geographical region
- Geospatial map view can visualize geo-tagged tweet locations
- Automatic Identification of emerging topics and trends
- STL based anomalous event detection

SMART Features

- Create custom filters to classify tweets related to a certain category e.g., epidemic outbreaks, weather and safety.
- Examine tweet clusters using ‘Content Lens’
- Drill down capabilities to examine actual tweet contents
- Integration of external news sources like Google and CNN trends
- Web based front end capable of running in a web browser

Applications in Epidemic Modeling

- Early detection of disease outbreaks and their locations
- Combine model simulation output or chief complaints data with social media data to improve outbreak prediction
- Study the impact of different factors like weather conditions, flooding and health facilities on disease outbreaks
- Identifying spread patterns and identifying modes of transmission

Demo

Acknowledgements

- David Hartley and Tianchan Niu for providing details on the Rift Valley fever simulation

- Min Chen, Harold Bosch, Denis Thom and Sherry Towers for their helpful discussions

- This work was supported in part by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0002, the Foreign Animal and Zoonotic Disease Center, and the Defense Threat Reduction Agency under Award Number HDTRA 1-10-083

VACCINE

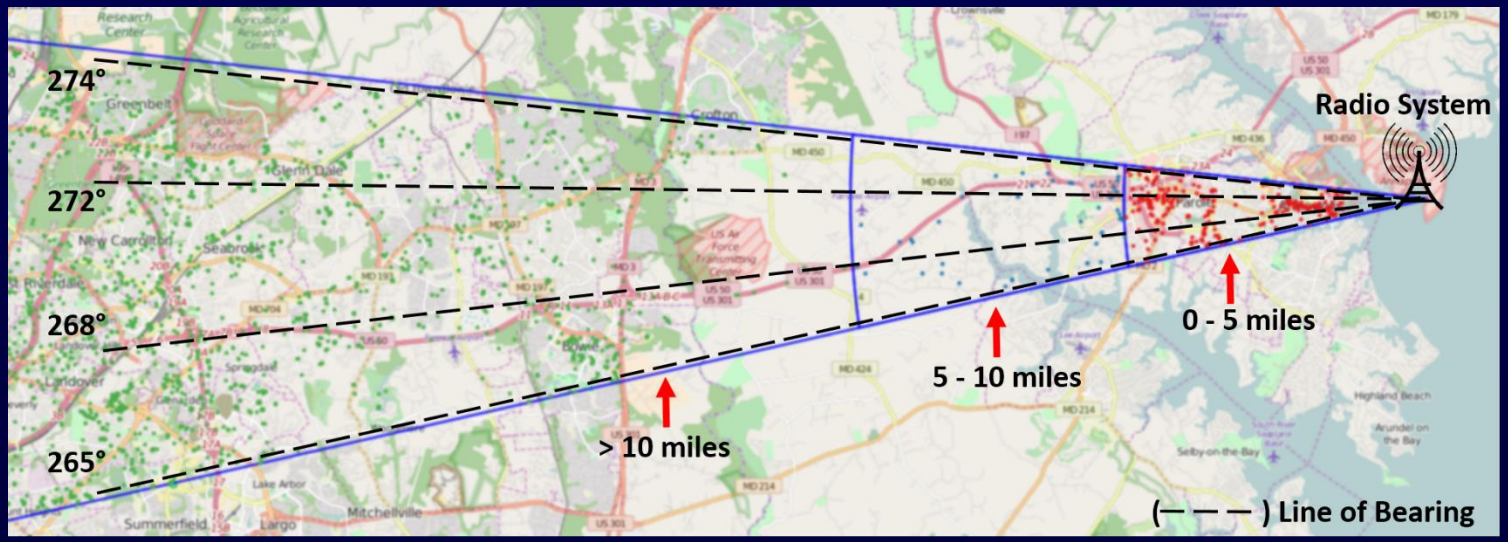
Thank you

Shehzad Afzal
safzal@purdue.edu

http://www.purdue.edu/discoverypark/vaccine

VACCINE

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Junghoon Chae[1], Jiawei Zhang[1], Sungahn Ko[2], Abish Malik[1], Heather Connell[1], David S. Ebert[1]

Purdue University[1], Ulsan National Institute of Science and Technology[2]

May 11, 2016

VACCINE

Introduction

- The U.S. Coast Guard (USCG) receives 139 hoax distress calls nationally per year
- Response to false calls wastes considerable resources
  - Put Coast Guard rescue personnel in dangerous situations
  - Waste valuable search and rescue assets that should be used for real emergency situations

Challenge

- Hard to identify hoax callers
- Rescue 21 (USCG radio communication system)
  - Utilized for identification of suspected hoax calls
  - Provide not enough information for finding hoax callers
- Need more intelligent analytics support

Goals

- A social media visual analytic driven by the Rescue 21 system
- Integrating the tool into SMART (Social Media Analytics Reporting Toolkit)
- Improving investigation of hoax calls by identifying suspicious social media

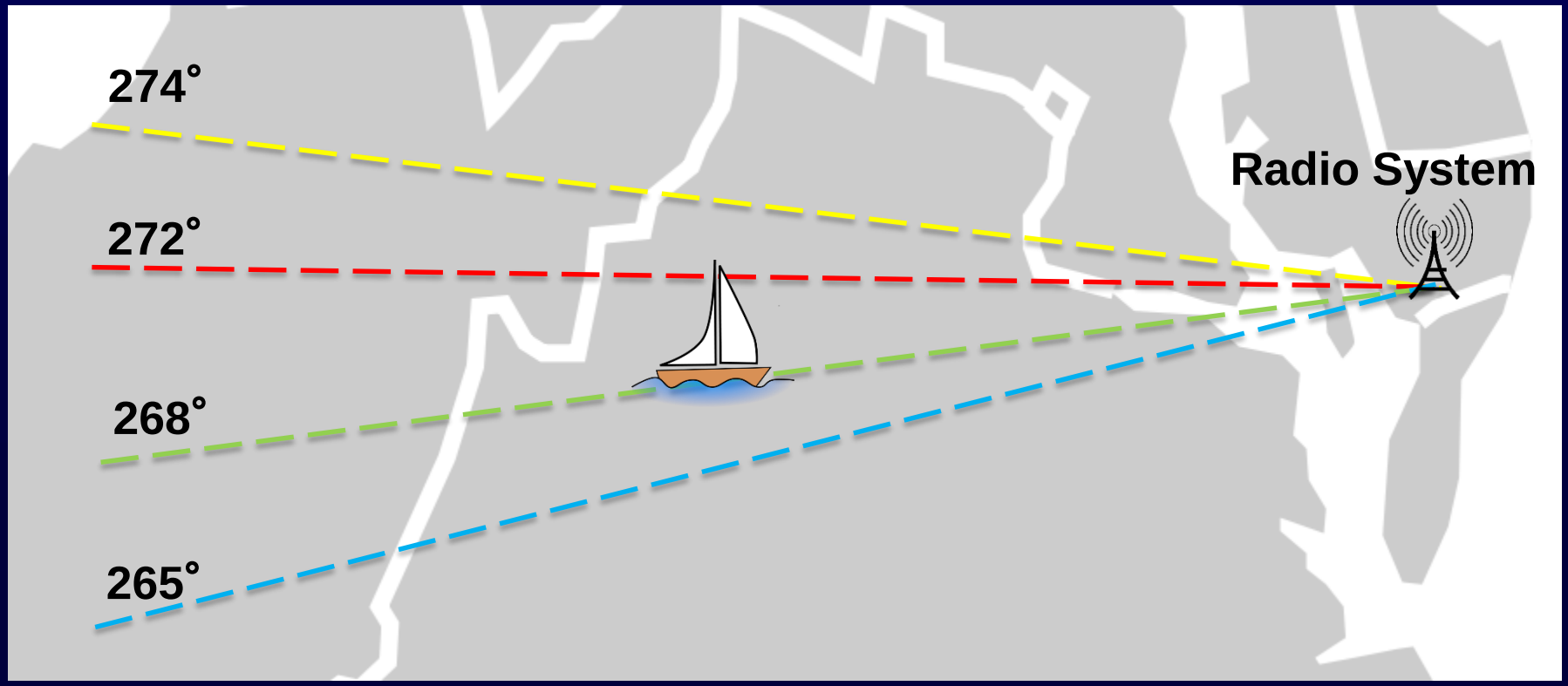
Approach

- Interactive visual analytics using social media
- Keyword collection & Data enrichment
- Identification of suspicious users & messages

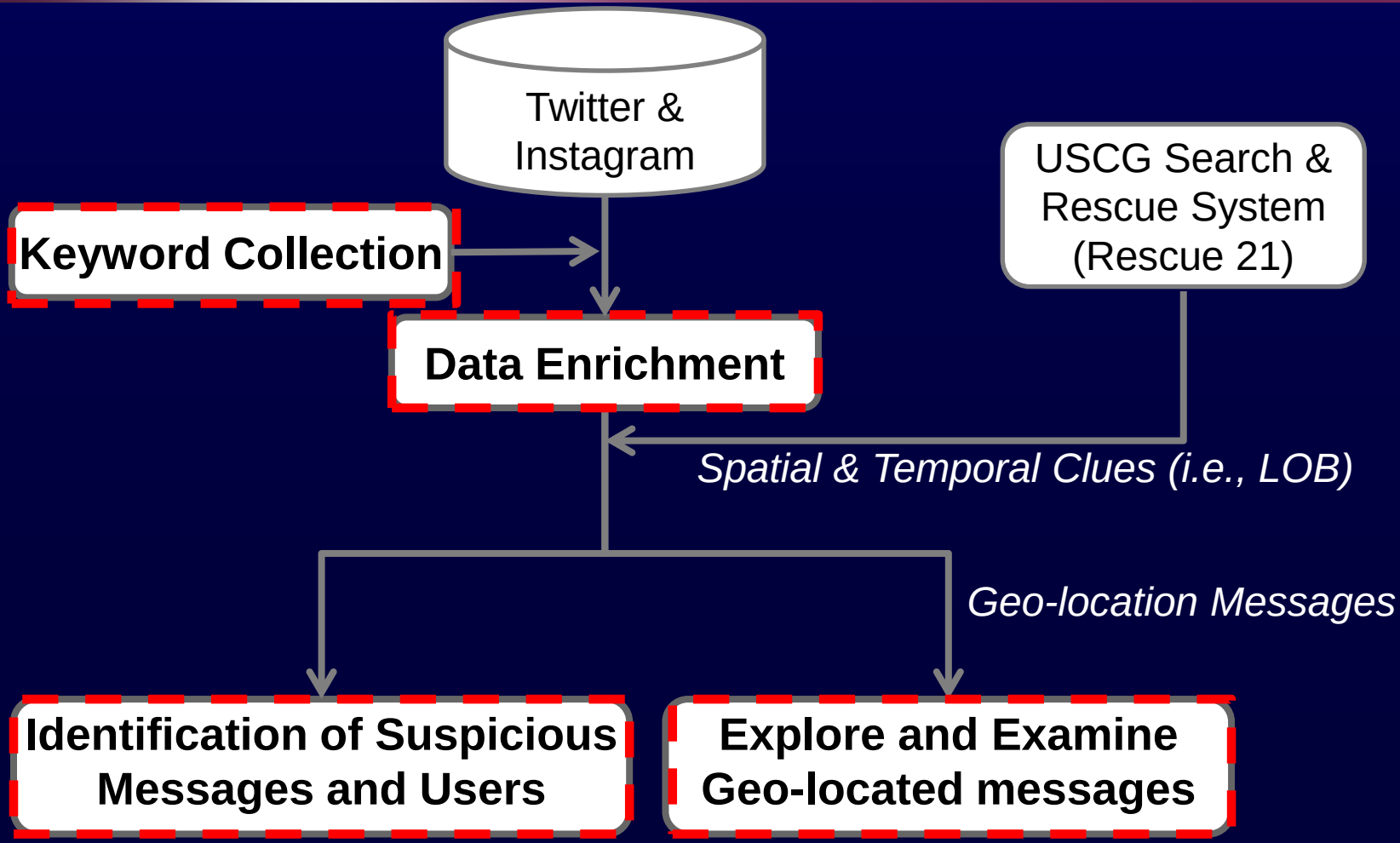
Background: Rescue 21

- USCG's radio communication system with direction-finding capability for search and rescue operations
- Providing lines of bearing (LOB) from a radio tower to the source of radio calls that it receives
- LOB is used for locating transmitters which are not able to specify their locations

Example of LOB



Analysis Process



Keyword Collection

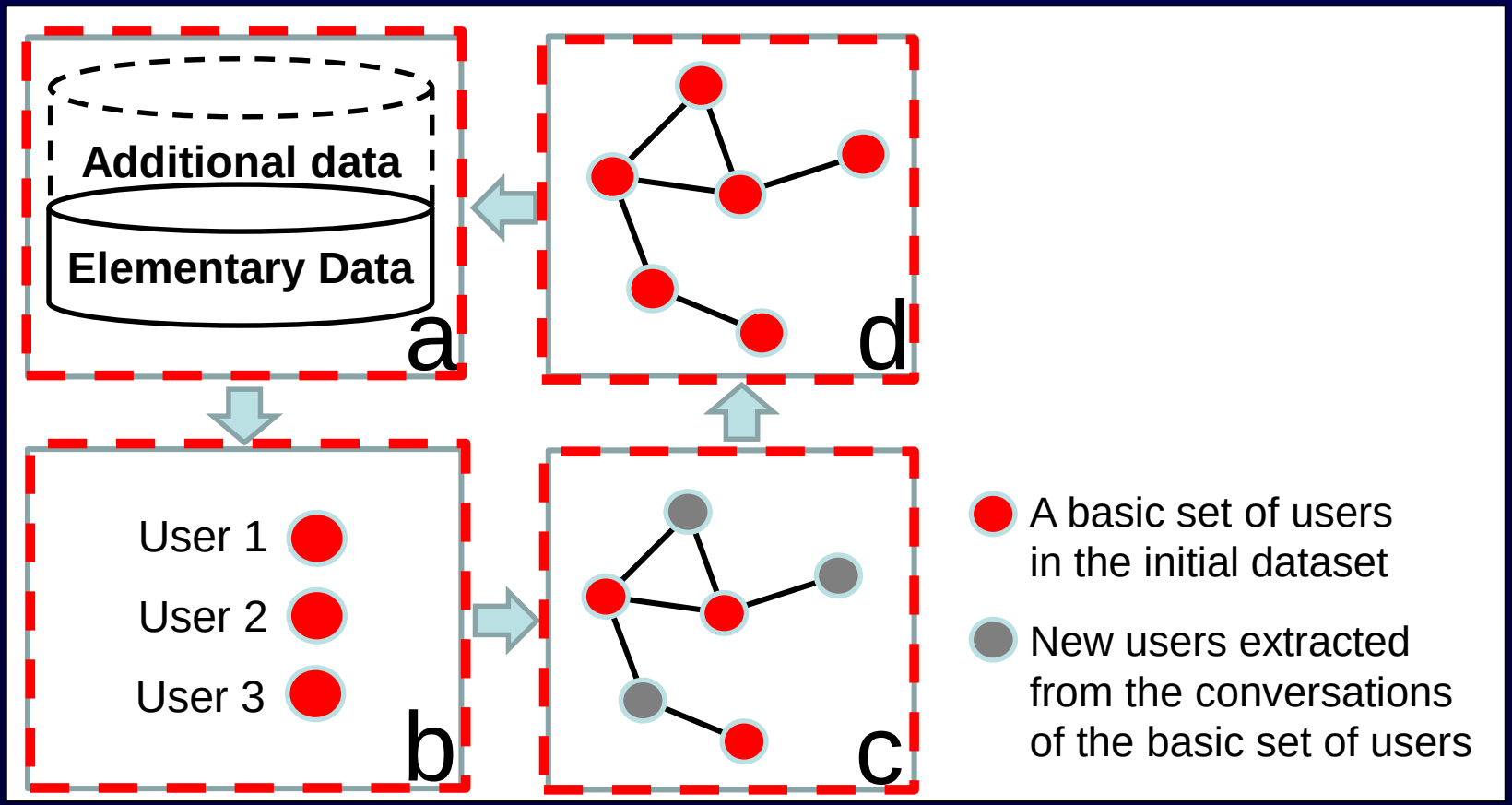
- Building a keyword corpus related to USCG and hoax distress calls
  - Retrieve highly relevant social media data
  - Reduce the analysis space and enable effective analysis tasks
- Adding words with semantic relationship to the initial keywords using WordNet
- USCG related online articles and historical social media data

Keyword Collection

Category	Initial Keywords	Extended Keywords
Coast Guard	USCG, vessel, ship, boat, ferry, shore, coastline ...	#cg, #coastguard, maritime, navy, fleet, lifeboat, craft ...
Hoax Call	**prank call, distress call, mayday, crank call, hoax ...**	**siren, save, false call, bogus, stunt, prankster ...**
Search and Rescue	helicopter, seek, search, rescue, flare, flashing ...	#sos, copter, airliner, navigate, 911, warning ...
Accident	explosion, accident, sink, crash, damaged, burning ...	disaster, hazard, collision, shot, smack, hit, break ...
Radio Channel	broadcast, radio, transmit, communicate ...	cable, on the air, airwaves, wireless, phone, report ...
Illegal Immigration	immigrant, immigration, panga, smuggling, drug ...	foreigner, exodus, migration, crossing, resettlement ...

Data Enrichment

- Data enrichment via conversation networks of users



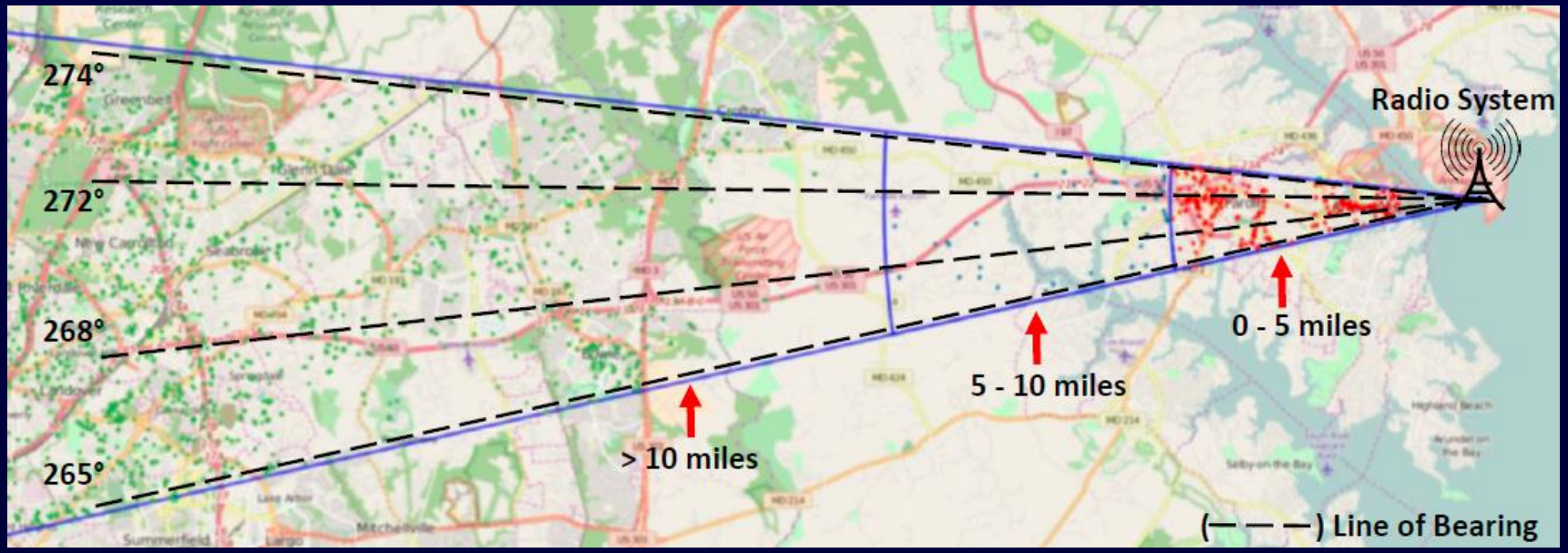
Identification of Suspicious Users

- Multiple hoax distress calls by one caller during different time frames
- Users who post messages within a time window when a hoax call was received are potentially more suspicious than others

For user i, if i posted a message
in a time window of a hoax call
Frequency score[i] += 1

Combination of Visual Analytics and Rescue 21

- Analysis of suspicious geo-tagged social media messages based on spatial and temporal clues driven by a Rescue 21 system



Case Study

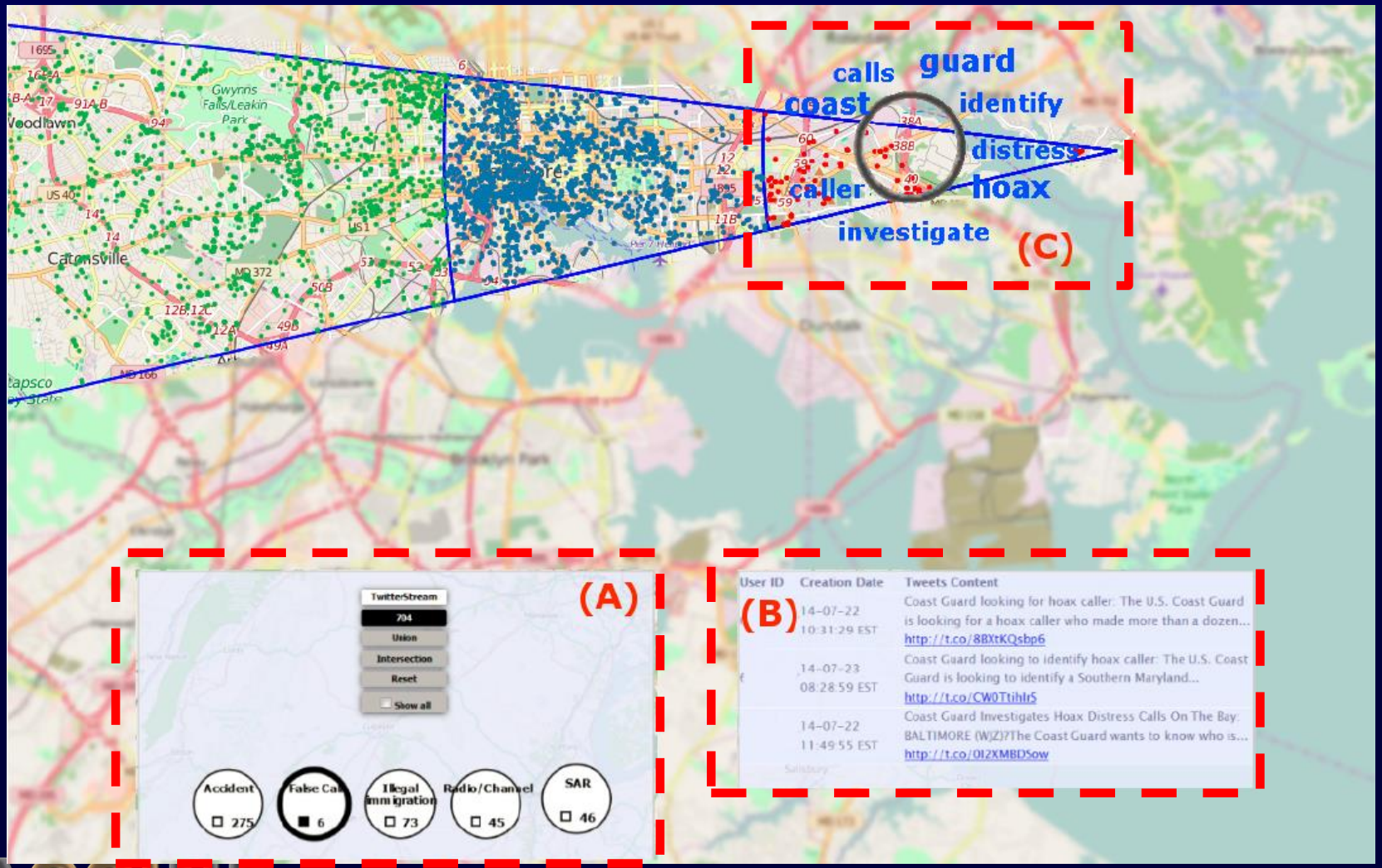
Case	Time	Number of Calls
1	January, 2014	1
2	July and August, 2014	10
3	May, 2014	3

<Information of the hoax distress call cases>

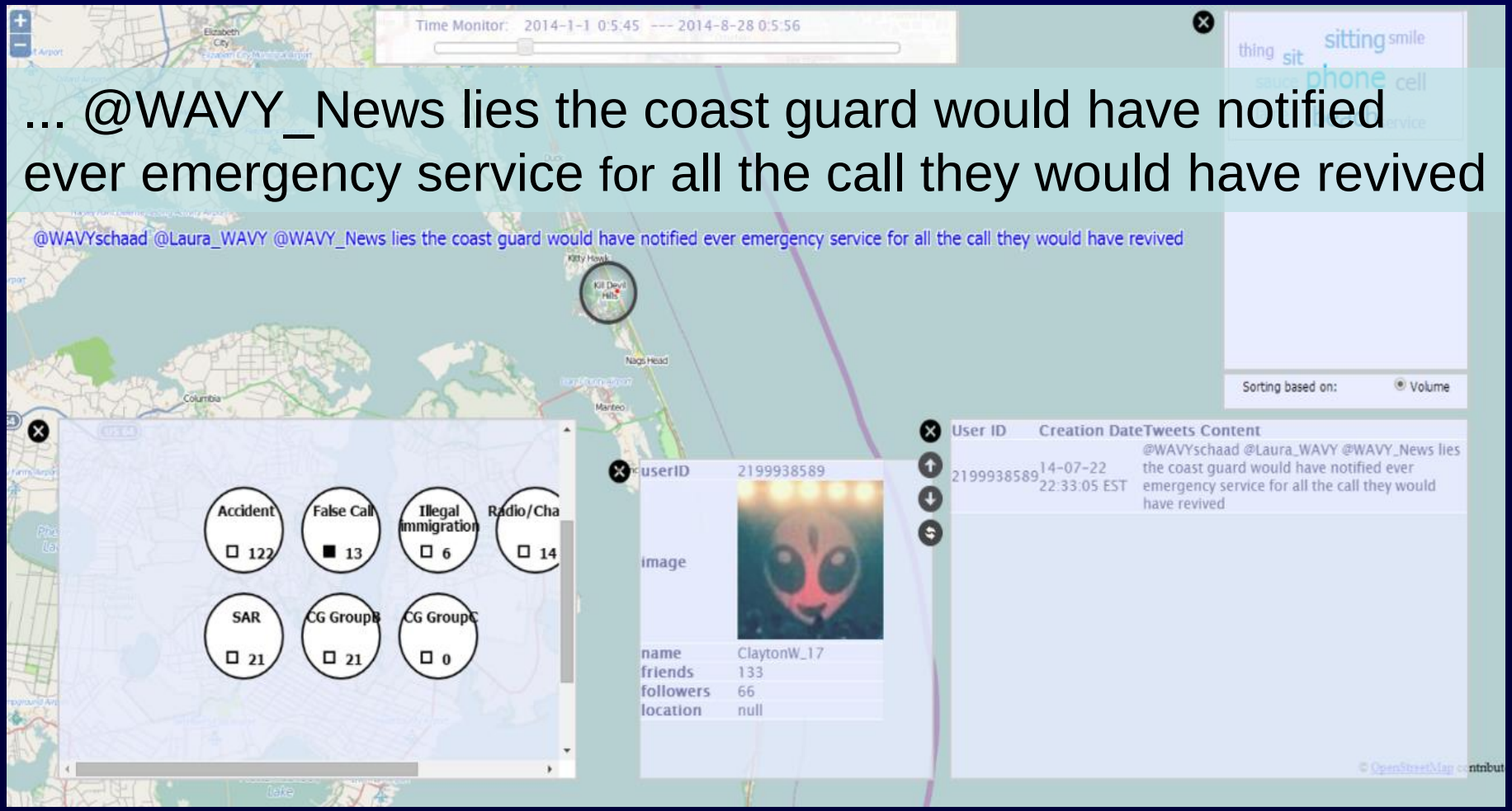
Data Type	Frequency Score						
	6	5	4	3	2	1	total
Twitter	1	4	4	15	60	269	353
Instagram			2	3	20	312	337

<Distribution of the users with respect to the frequency scores of each dataset for case 2>

Case Study


calls
guard
coast
identify
distress
caller
hoax
investigate
(C)
(A)
TwitterStream
704
Union
Intersection
Reset
Show all
Accident
275
False Call
6
Illegal Immigration
73
Radio/Channel
45
SAR
46
(B)
User ID
Creation Date
Tweets Content
14-07-22 10:31:29 EST
Coast Guard looking for hoax caller: The U.S. Coast Guard is looking for a hoax caller who made more than a dozen... http://t.co/8BXtKQsbp6
14-07-23 08:28:59 EST
Coast Guard looking to identify hoax caller: The U.S. Coast Guard is looking to identify a Southern Maryland... http://t.co/CW0TtihlrS
14-07-22 11:49:55 EST
Coast Guard Investigates Hoax Distress Calls On The Bay: BALTIMORE (WJZ)?The Coast Guard wants to know who is... http://t.co/0I2XMBDSow

Case Study


Time Monitor: 2014-1-1 0:5:45 --- 2014-8-28 0:5:56
... @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
@WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
Sorting based on: Volume
Accident 122
False Call 13
Illegal Immigration 6
Radio/Cha 14
SAR 21
CG GroupB 21
CG GroupC 0
userID 2199938589
image
name ClaytonW_17
friends 133
followers 66
location null
User ID Creation Date Tweets Content
2199938589 14-07-22 22:33:05 EST @WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived

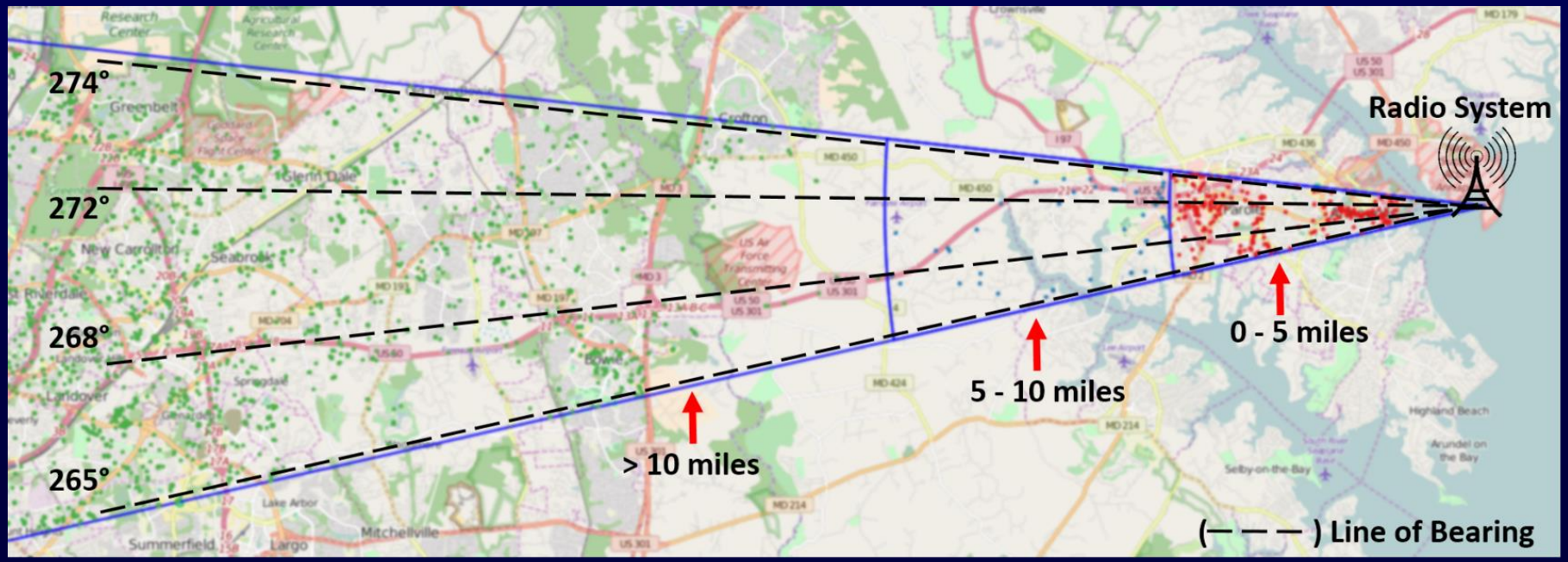
Conclusion

- A visual social media analytics solution driven by the Rescue 21 system for supporting the investigation of hoax distress calls
- Identifying suspicious Twitter and Instagram messages based on the spatial and temporal clues from the system

Acknowledgement

- U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0003

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media

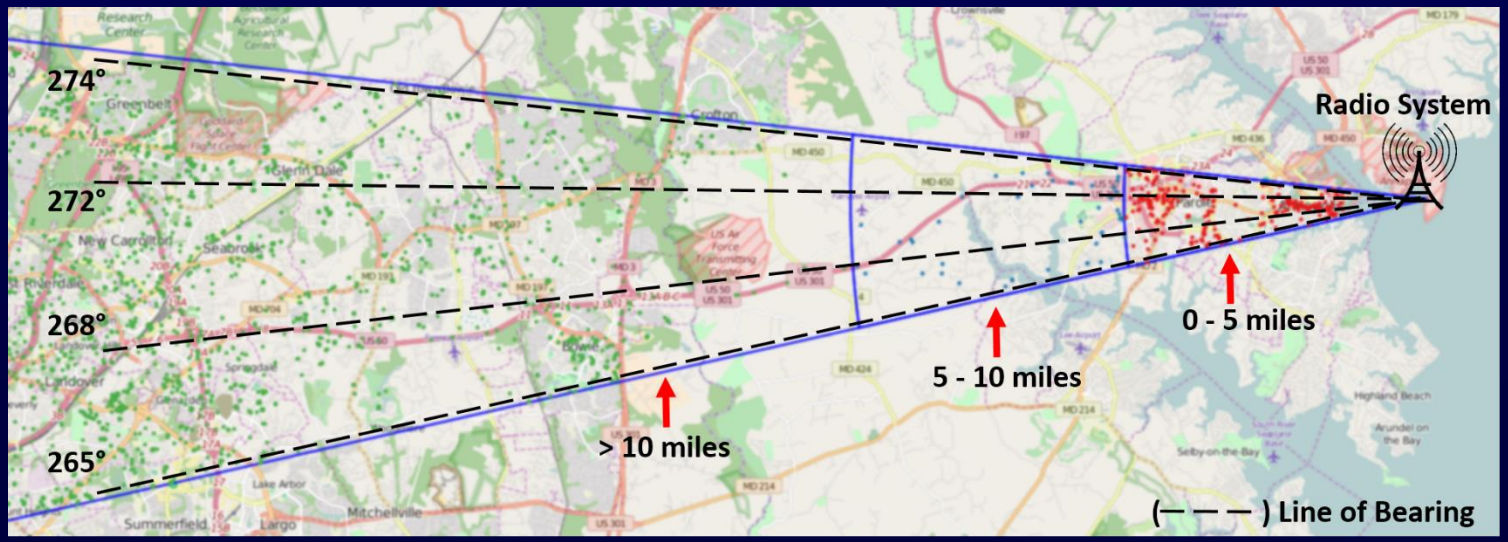


Thank you!

Junghoon Chae
jchae@purdue.edu

VACCINE

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Junghoon Chae[1], Jiawei Zhang[1], Sungahn Ko[2], Abish Malik[1], Heather Connell[1], David S. Ebert[1]

Purdue University[1], Ulsan National Institute of Science and Technology[2]

May 11, 2016

VACCINE

Introduction

- The U.S. Coast Guard (USCG) receives 139 hoax distress calls nationally per year
- Response to false calls wastes considerable resources
  - Put Coast Guard rescue personnel in dangerous situations
  - Waste valuable search and rescue assets that should be used for real emergency situations

Challenge

- Hard to identify hoax callers
- Rescue 21 (USCG radio communication system)
  - Utilized for identification of suspected hoax calls
  - Provide not enough information for finding hoax callers
- Need more intelligent analytics support

Goals

- A social media visual analytic driven by the Rescue 21 system
- Integrating the tool into SMART (Social Media Analytics Reporting Toolkit)
- Improving investigation of hoax calls by identifying suspicious social media

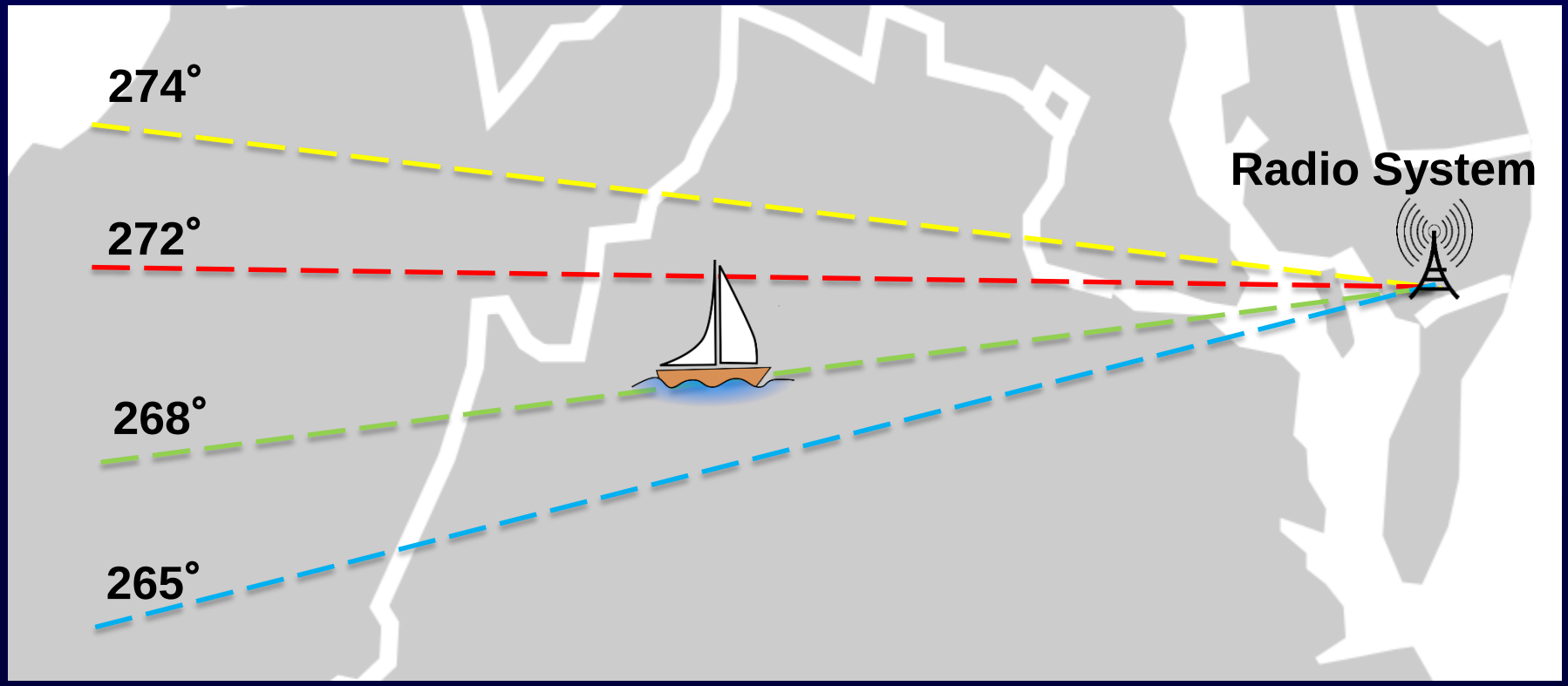
Approach

- Interactive visual analytics using social media
- Keyword collection & Data enrichment
- Identification of suspicious users & messages

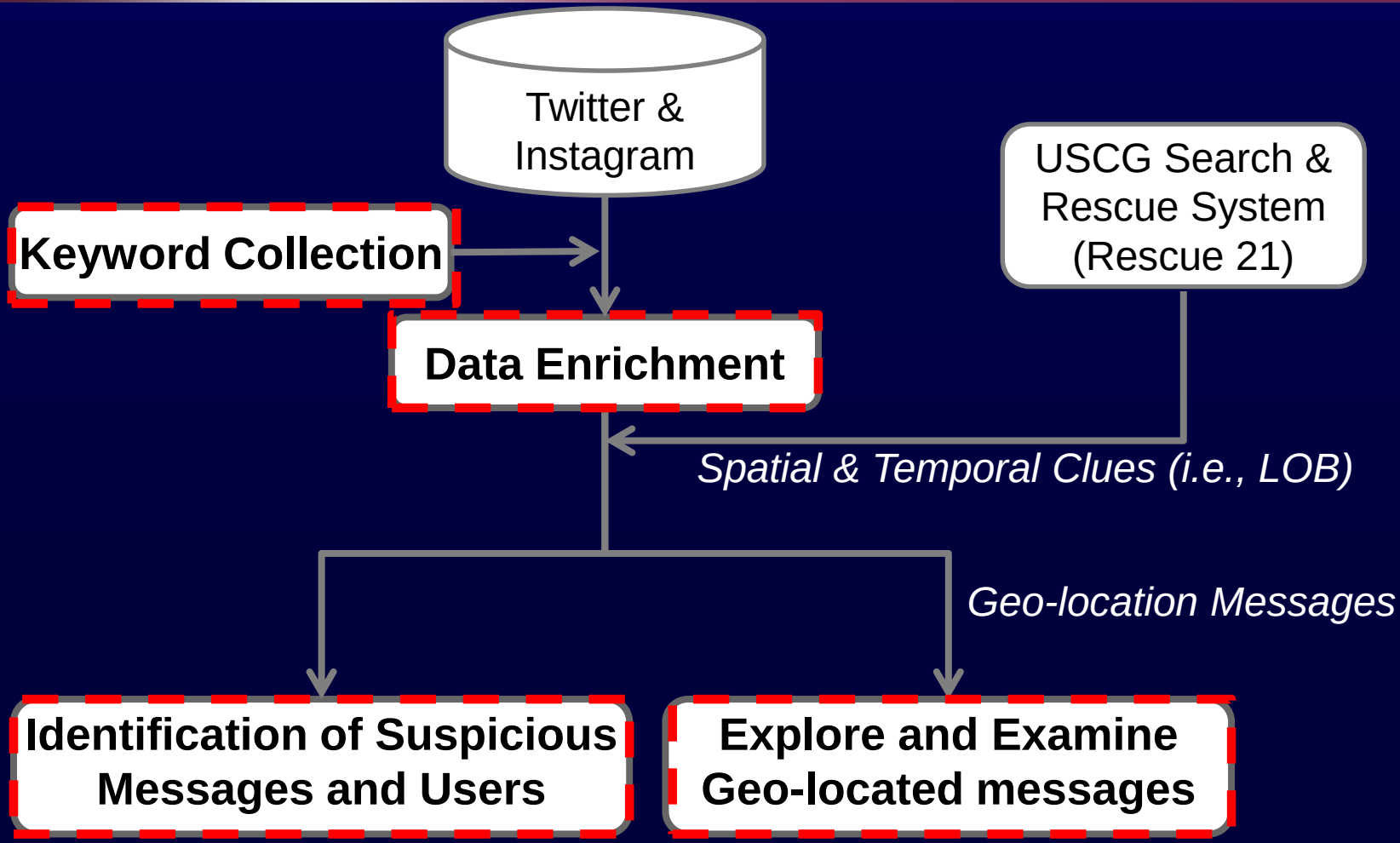
Background: Rescue 21

- USCG's radio communication system with direction-finding capability for search and rescue operations
- Providing lines of bearing (LOB) from a radio tower to the source of radio calls that it receives
- LOB is used for locating transmitters which are not able to specify their locations

Example of LOB



Analysis Process



Keyword Collection

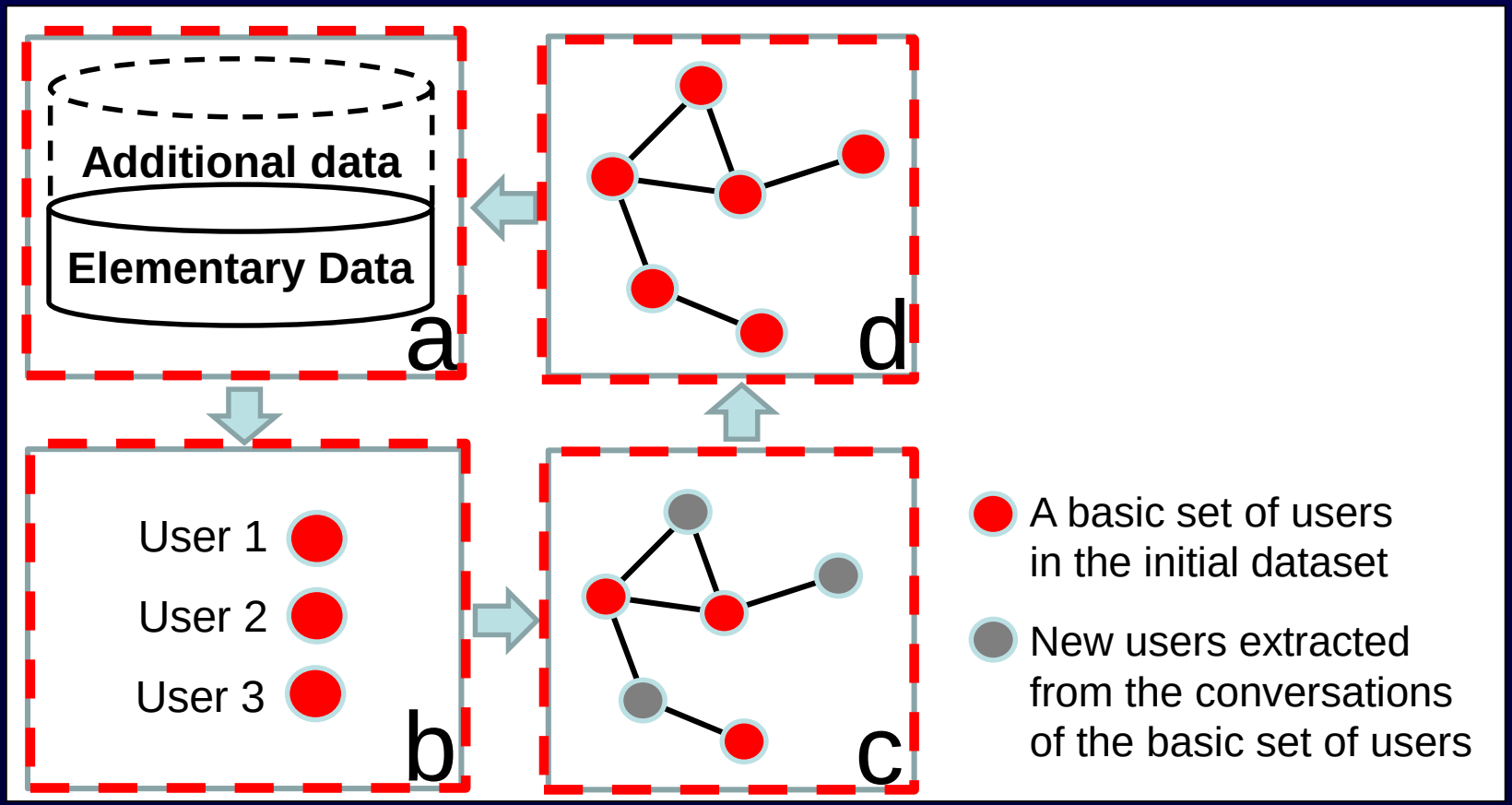
- Building a keyword corpus related to USCG and hoax distress calls
  - Retrieve highly relevant social media data
  - Reduce the analysis space and enable effective analysis tasks
- Adding words with semantic relationship to the initial keywords using WordNet
- USCG related online articles and historical social media data

Keyword Collection

Category	Initial Keywords	Extended Keywords
Coast Guard	USCG, vessel, ship, boat, ferry, shore, coastline ...	#cg, #coastguard, maritime, navy, fleet, lifeboat, craft ...
Hoax Call	**prank call, distress call, mayday, crank call, hoax ...**	**siren, save, false call, bogus, stunt, prankster ...**
Search and Rescue	helicopter, seek, search, rescue, flare, flashing ...	#sos, copter, airliner, navigate, 911, warning ...
Accident	explosion, accident, sink, crash, damaged, burning ...	disaster, hazard, collision, shot, smack, hit, break ...
Radio Channel	broadcast, radio, transmit, communicate ...	cable, on the air, airwaves, wireless, phone, report ...
Illegal Immigration	immigrant, immigration, panga, smuggling, drug ...	foreigner, exodus, migration, crossing, resettlement ...

Data Enrichment

- Data enrichment via conversation networks of users



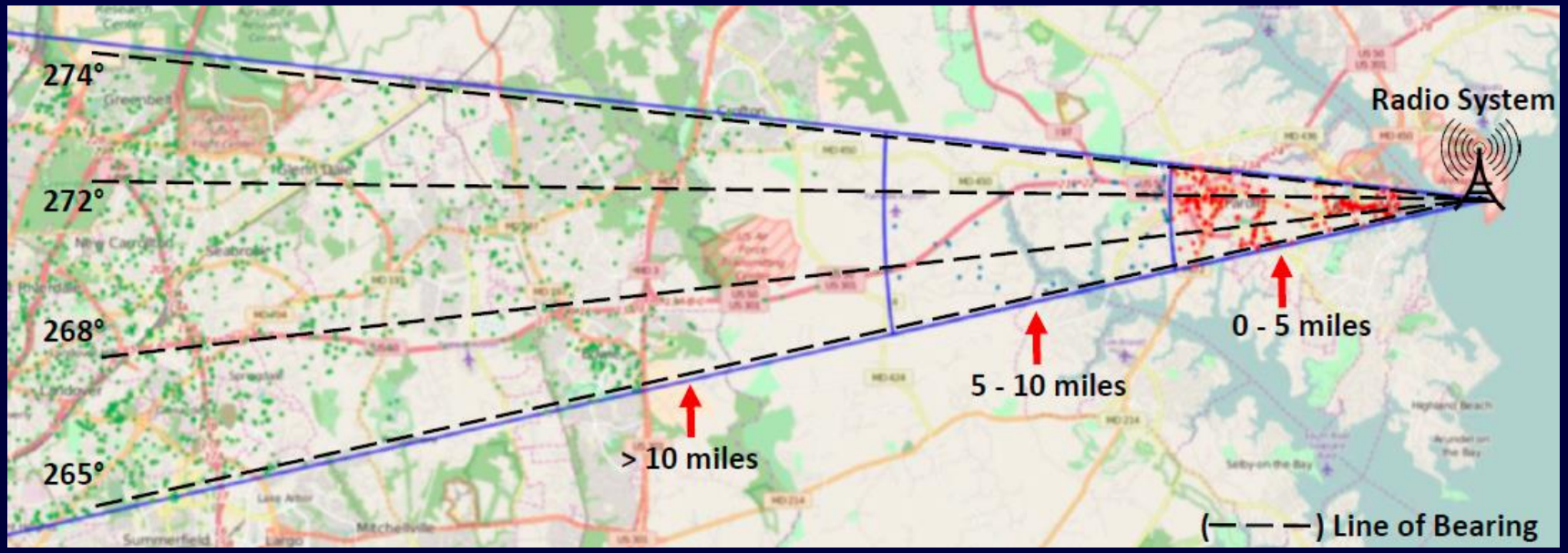
Identification of Suspicious Users

- Multiple hoax distress calls by one caller during different time frames
- Users who post messages within a time window when a hoax call was received are potentially more suspicious than others

For user i, if i posted a message
in a time window of a hoax call
Frequency score[i] += 1

Combination of Visual Analytics and Rescue 21

- Analysis of suspicious geo-tagged social media messages based on spatial and temporal clues driven by a Rescue 21 system



Case Study

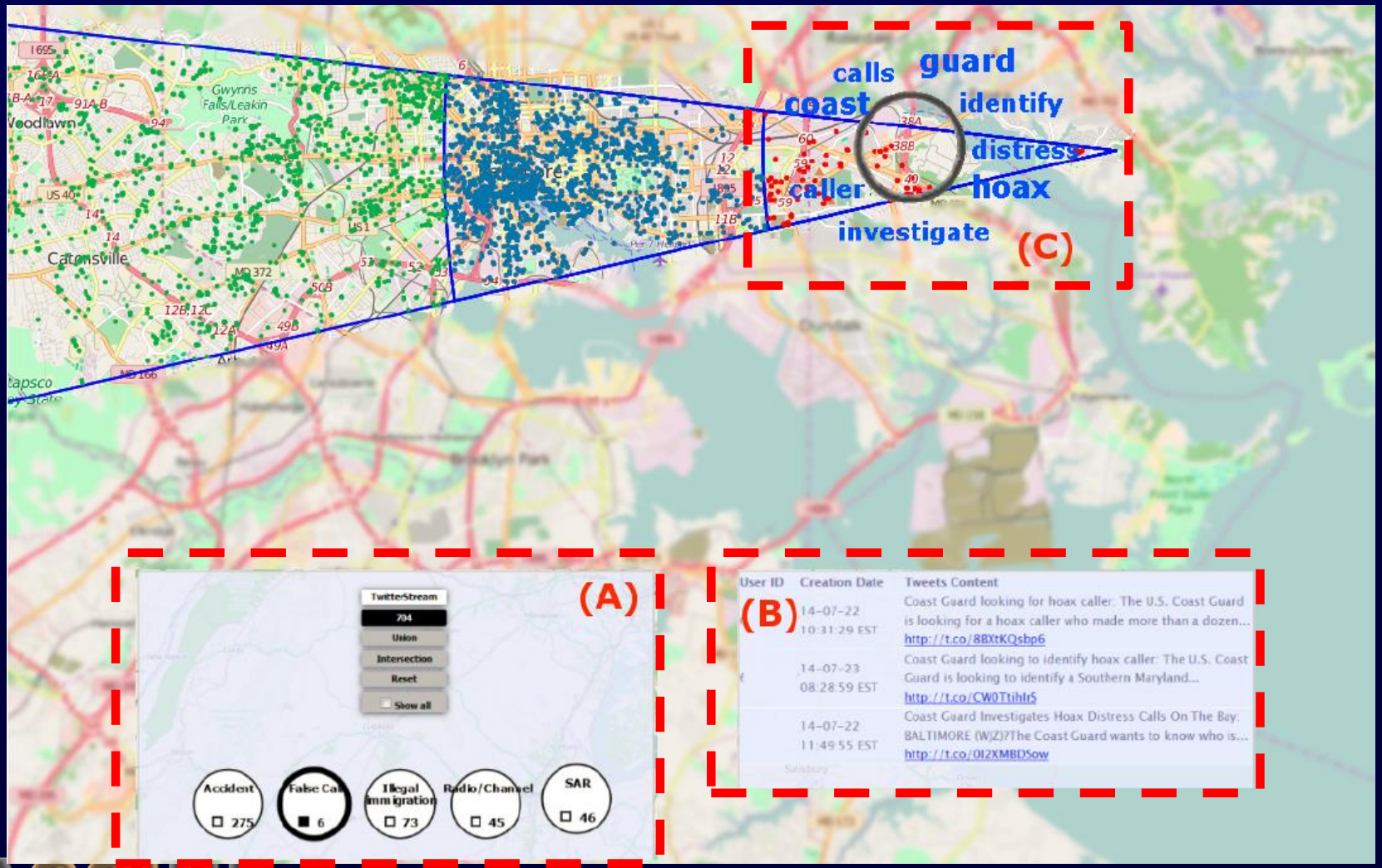
Case	Time	Number of Calls
1	January, 2014	1
2	July and August, 2014	10
3	May, 2014	3

<Information of the hoax distress call cases>

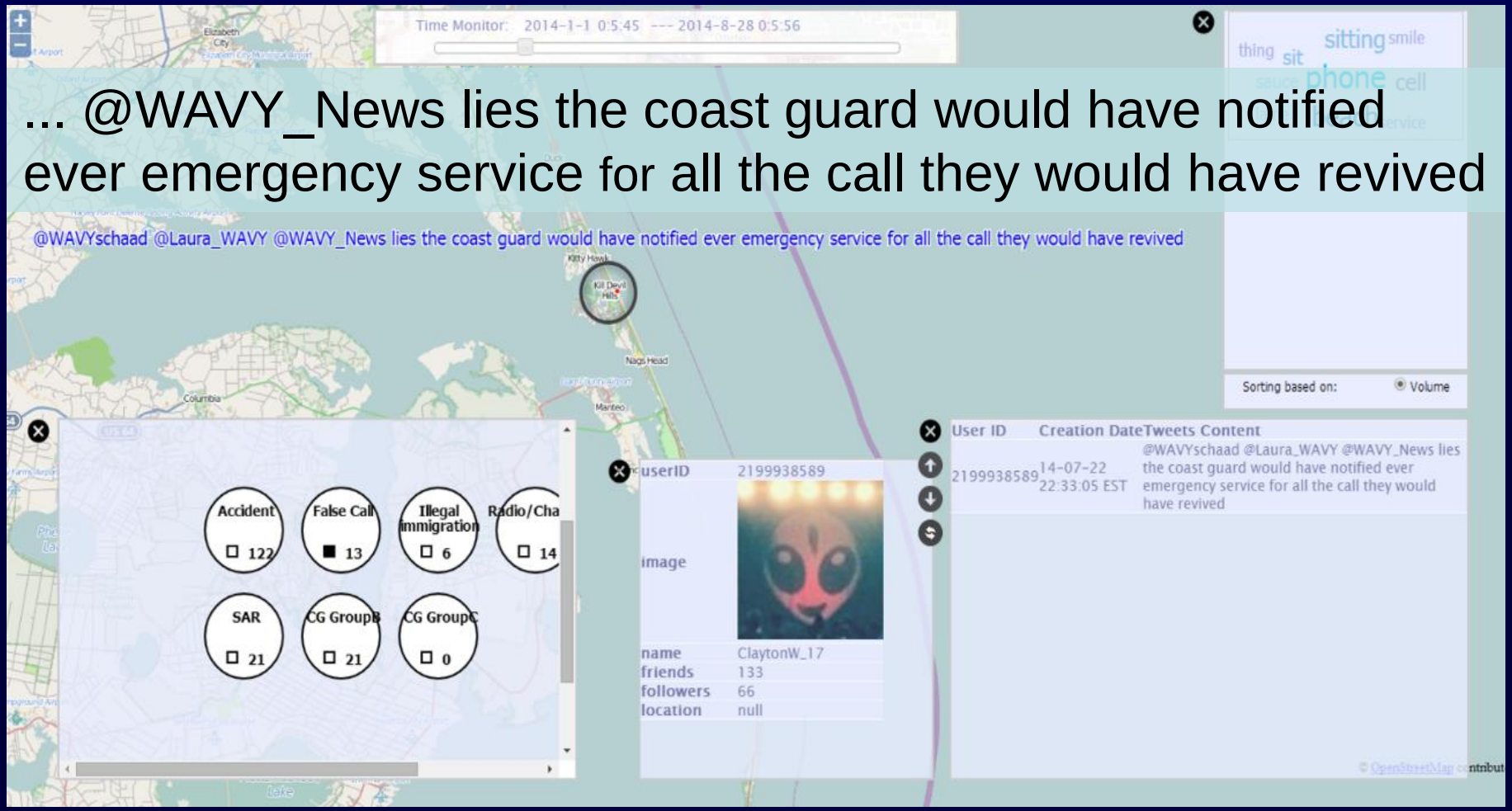
Data Type	Frequency Score						
	6	5	4	3	2	1	total
Twitter	1	4	4	15	60	269	353
Instagram			2	3	20	312	337

<Distribution of the users with respect to the frequency scores of each dataset for case 2>

Case Study


calls
guard
coast
identify
distress
caller
hoax
investigate
(C)
(A)
TwitterStream
704
Union
Intersection
Reset
Show all
Accident
275
False Call
6
Illegal Immigration
73
Radio/Channel
45
SAR
46
(B)
User ID
Creation Date
Tweets Content
14-07-22 10:31:29 EST
Coast Guard looking for hoax caller: The U.S. Coast Guard is looking for a hoax caller who made more than a dozen... http://t.co/8BXtKQsbp6
14-07-23 08:28:59 EST
Coast Guard looking to identify hoax caller: The U.S. Coast Guard is looking to identify a Southern Maryland... http://t.co/CW0TtihlrS
14-07-22 11:49:55 EST
Coast Guard Investigates Hoax Distress Calls On The Bay: BALTIMORE (WJZ)?The Coast Guard wants to know who is... http://t.co/0I2XMBDSow

Case Study


Time Monitor: 2014-1-1 0:5:45 --- 2014-8-28 0:5:56
... @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
@WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
Sorting based on: Volume
Accident 122
False Call 13
Illegal Immigration 6
Radio/Cha 14
SAR 21
CG GroupB 21
CG GroupC 0
userID 2199938589
image
name ClaytonW_17
friends 133
followers 66
location null
User ID Creation Date Tweets Content
2199938589 14-07-22 22:33:05 EST @WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived

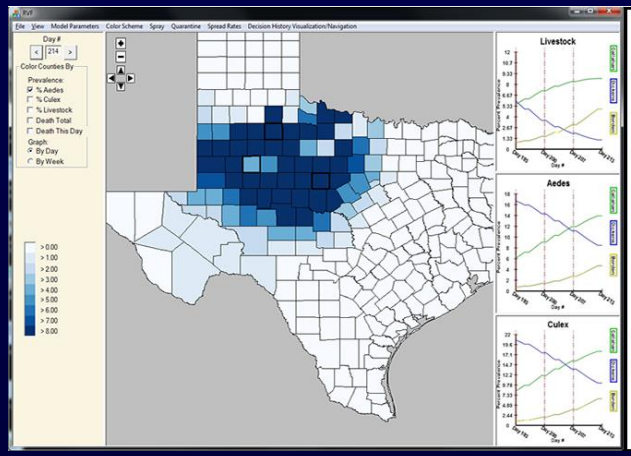
Conclusion

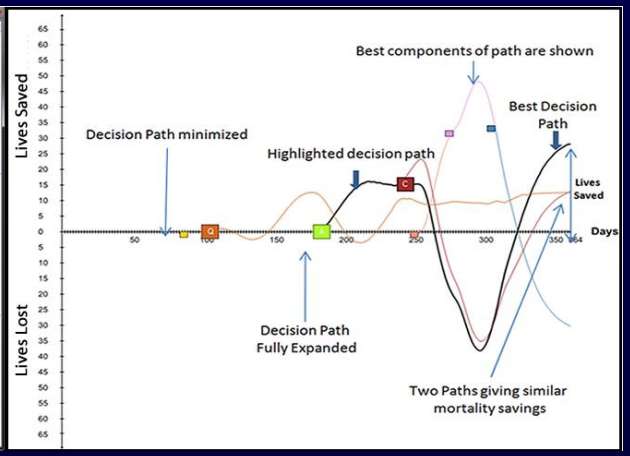
- A visual social media analytics solution driven by the Rescue 21 system for supporting the investigation of hoax distress calls
- Identifying suspicious Twitter and Instagram messages based on the spatial and temporal clues from the system

Acknowledgement

- U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0003

Temporal/Geo-spatial Analytics of Epidemic Data and Forecasting Epidemic Spread





Shehzad Afzal

VACCINE - DHS Center of Excellence

Purdue University

(http://www.purdue.edu/discoverypark/vaccine/)

VACCINE

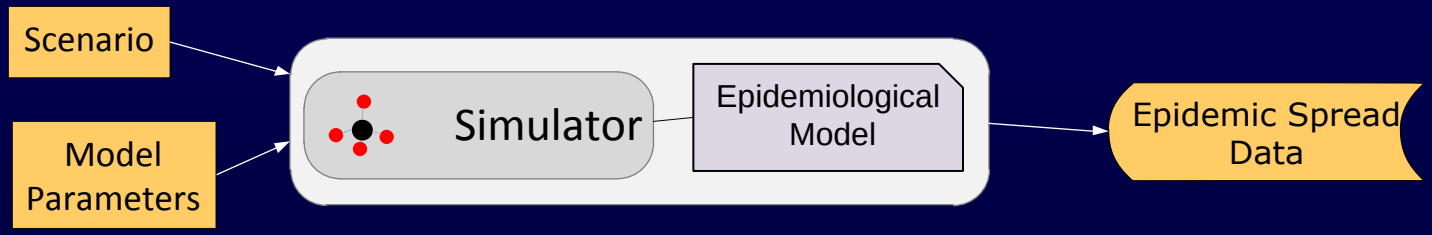
Motivation

- Officials must prepare complex plans to deal with Epidemic Outbreaks like:
  - Rift Valley fever (RVF), Pandemic Influenza, Dengue Fever
- Epidemic Response Planning
  - Table top exercises
  - Past epidemic response histories
  - Computer based Simulations
- Table top exercises & Past response histories have limitations

VACCINE

Motivation

- Computer based simulations



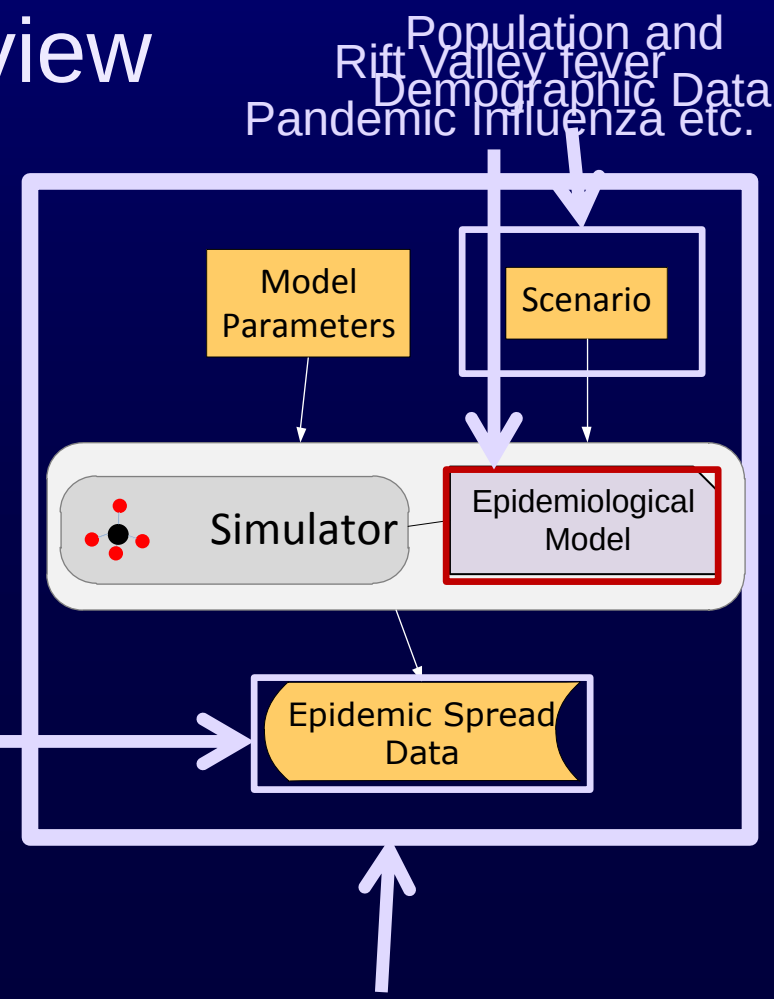
- Mitigative measures modify the input parameters
  - Examples: Pesticides, Quarantine etc.
- Simulation ‘response’ needs to be compared across ‘space & time’

Contribution

Interactive Decision Support Environment:

- Epidemic Models Exploration
- Interactively utilize mitigative response measures
- Decision History Visualization and Navigation View
  - links to multiple simulation runs
- Simulation replay and path exploration visualization
- Design and Evaluate Mitigative Response Strategy

System Overview

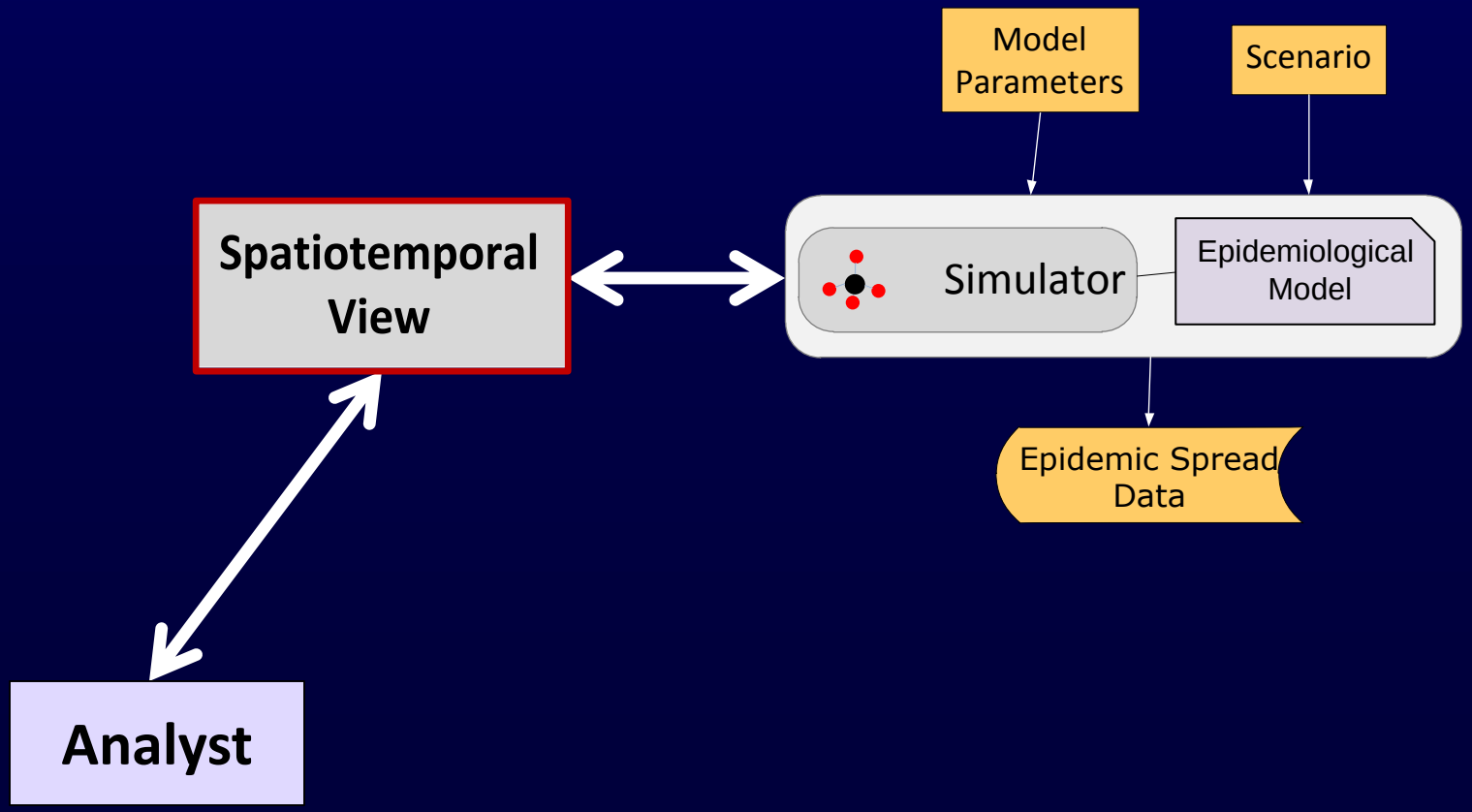


# of sick/dead
per population unit

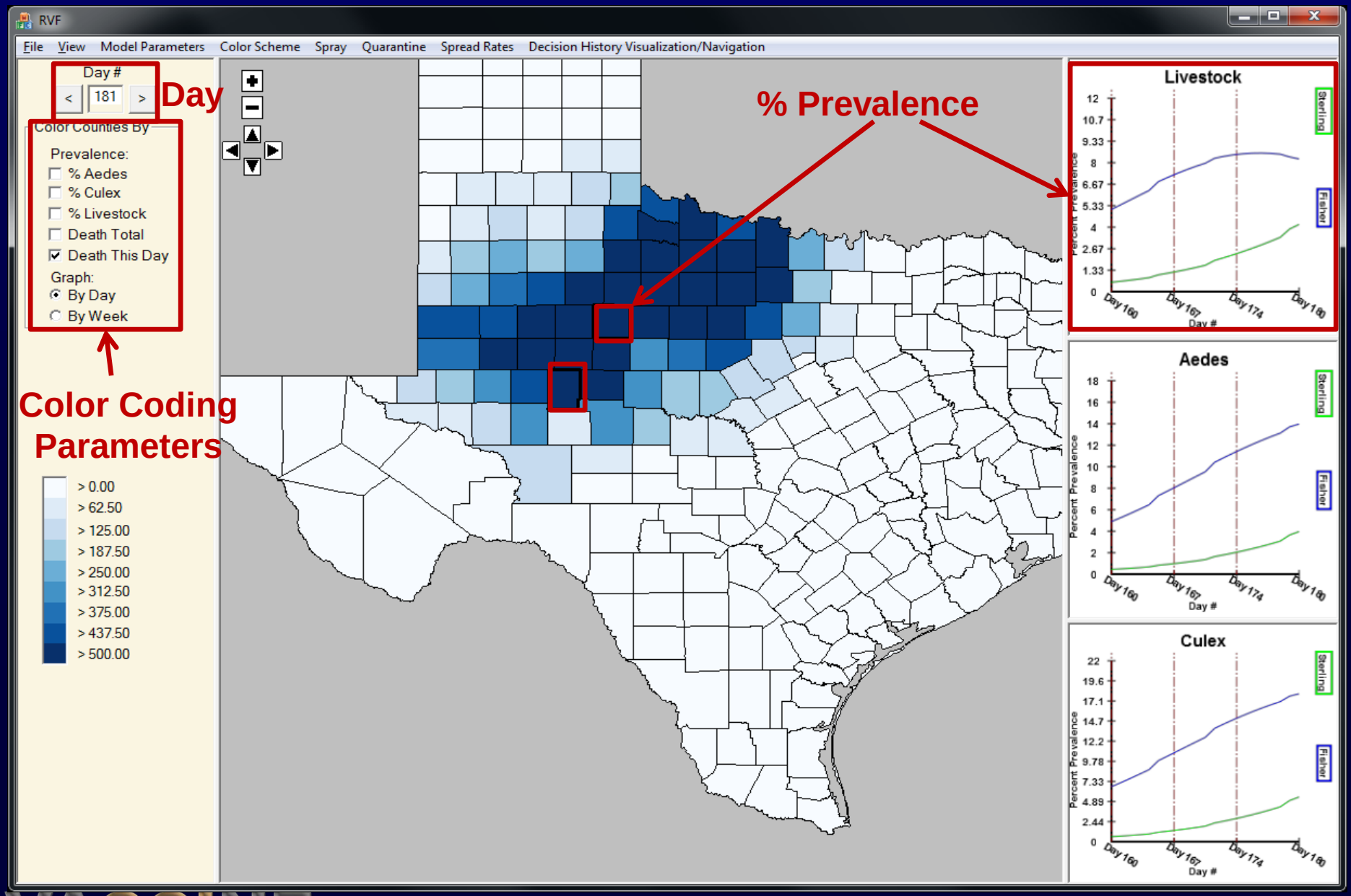
- Interchangeable
- Pre-Implemented

VACCINE

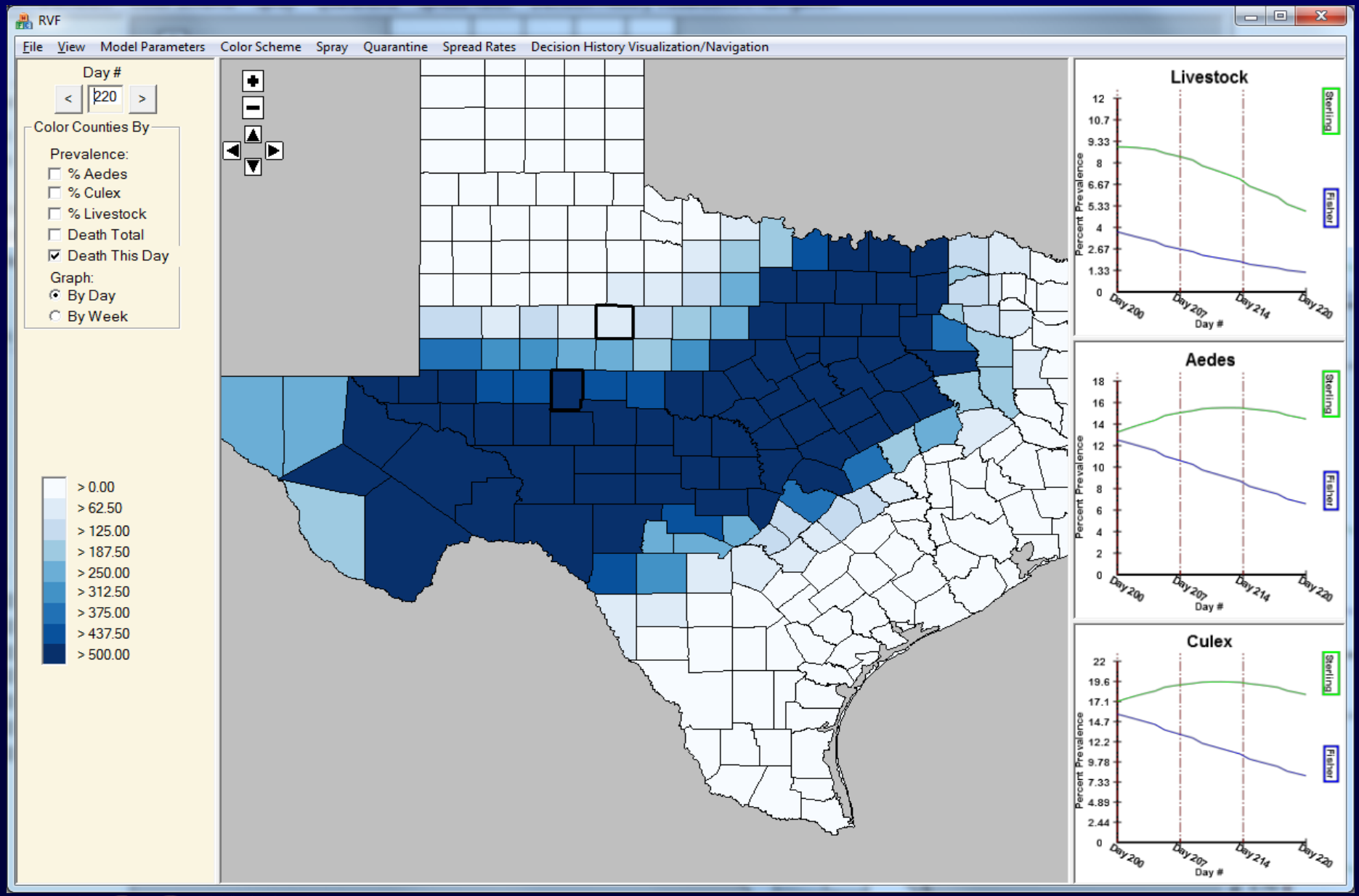
System Overview



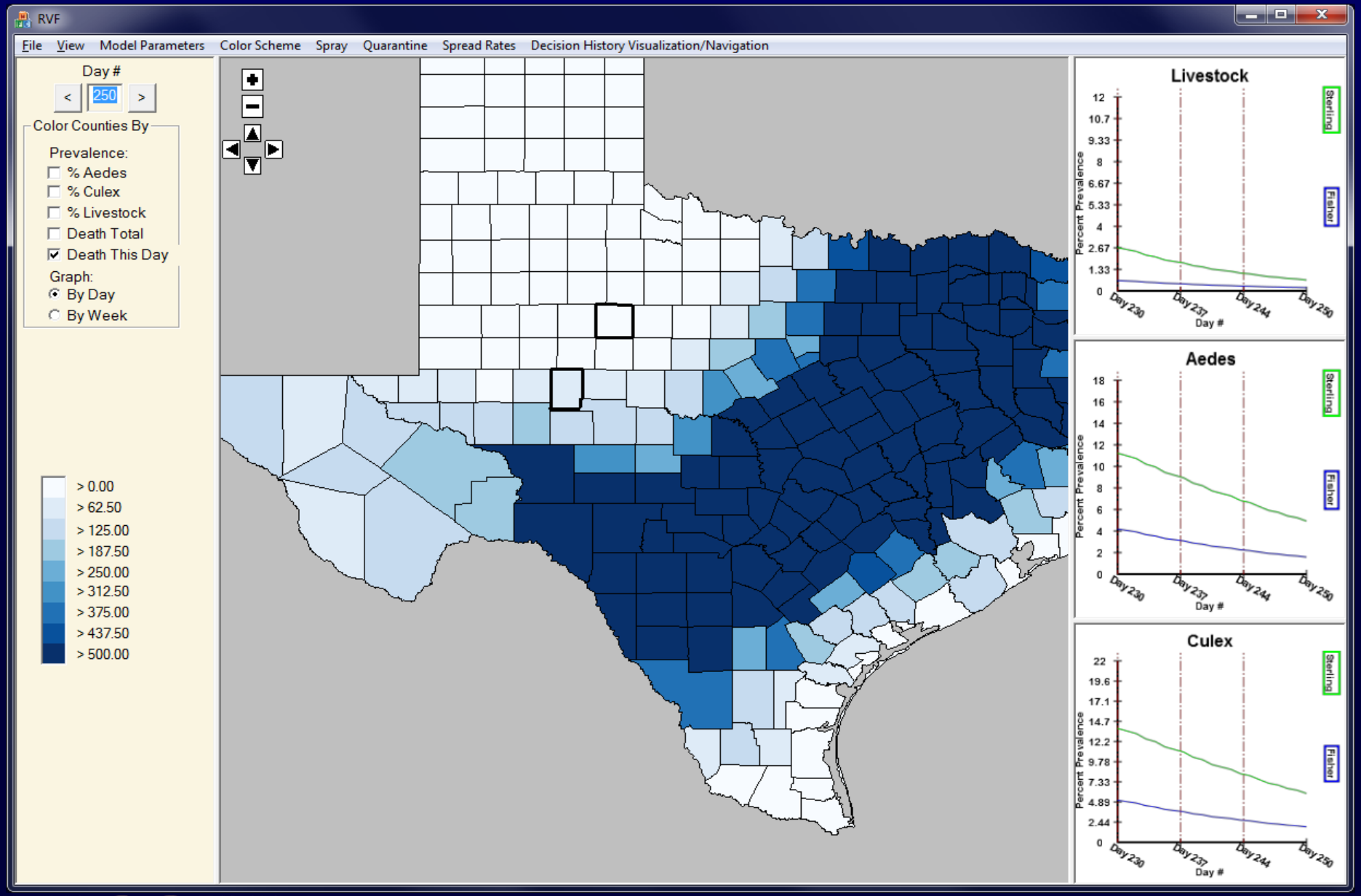
Spatiotemporal View (Rift Valley Fever)



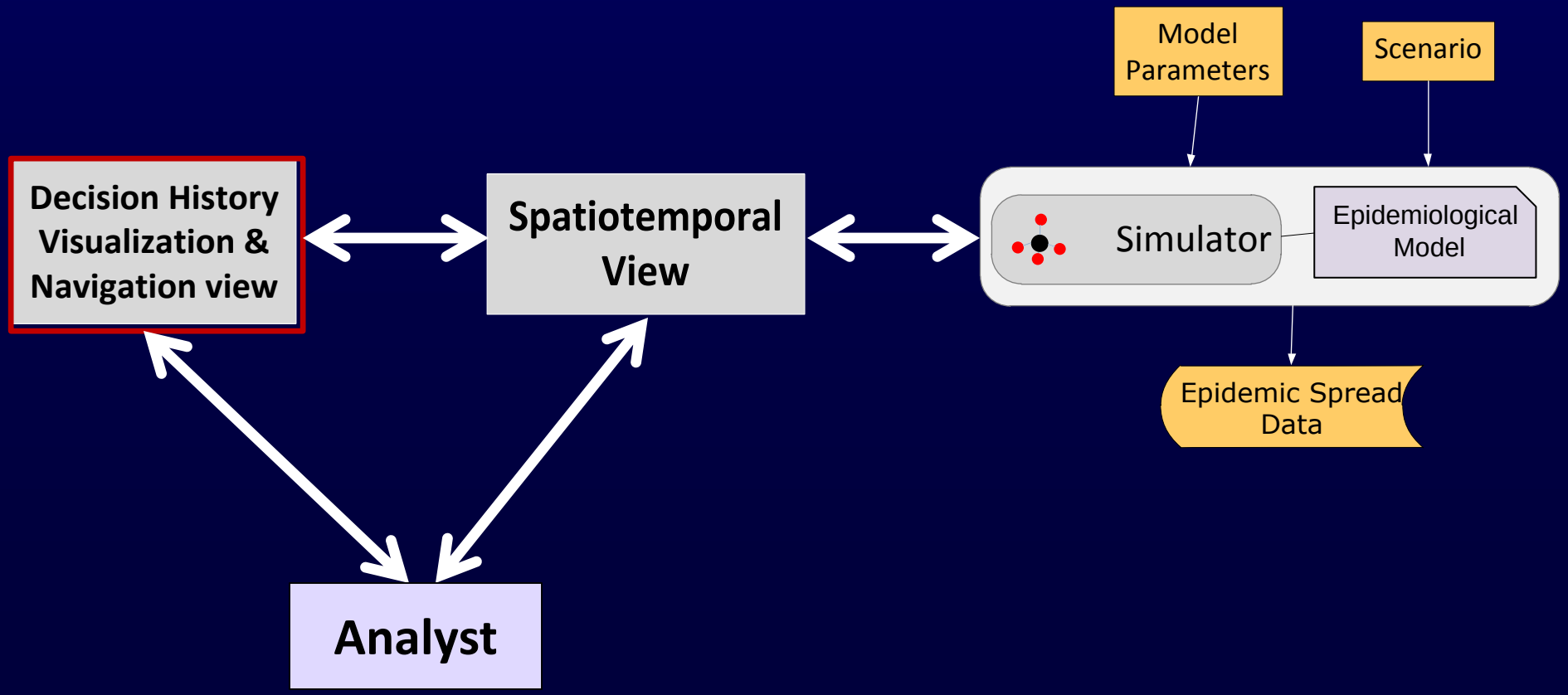
Spatiotemporal View (Rift Valley Fever)



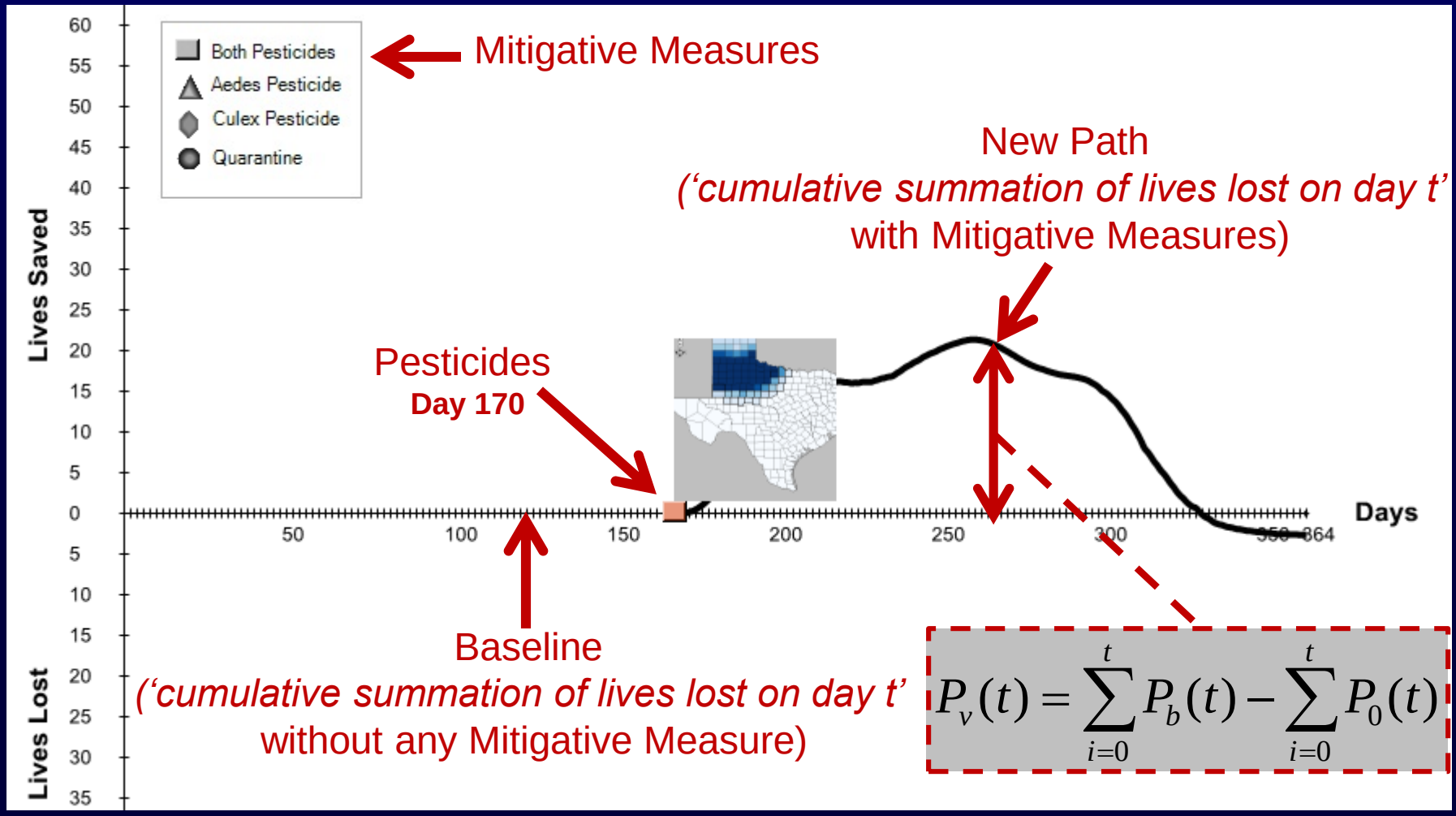
Spatiotemporal View (Rift Valley Fever)



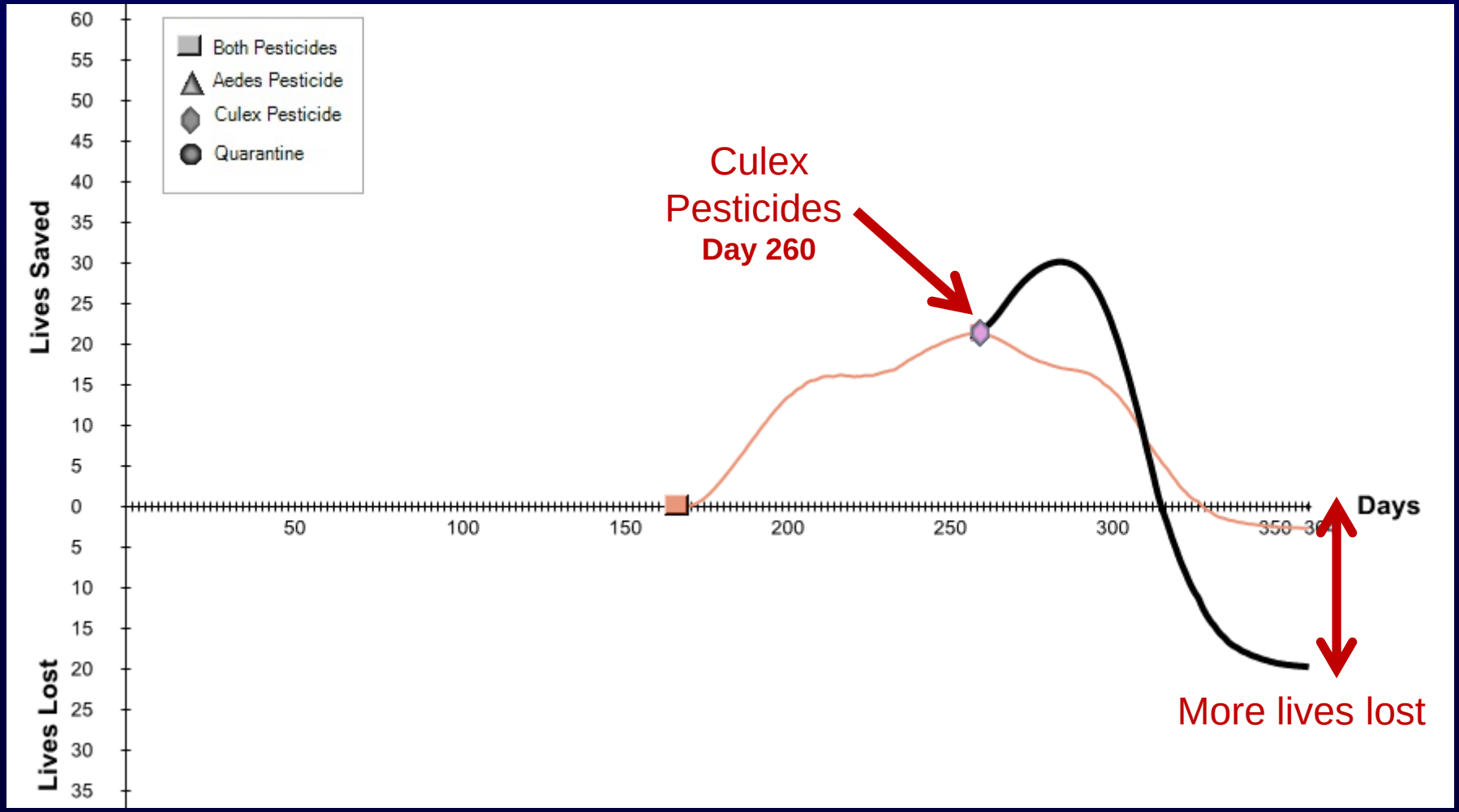
System Overview



Decision History Visualization & Navigation View
Rift Valley Fever

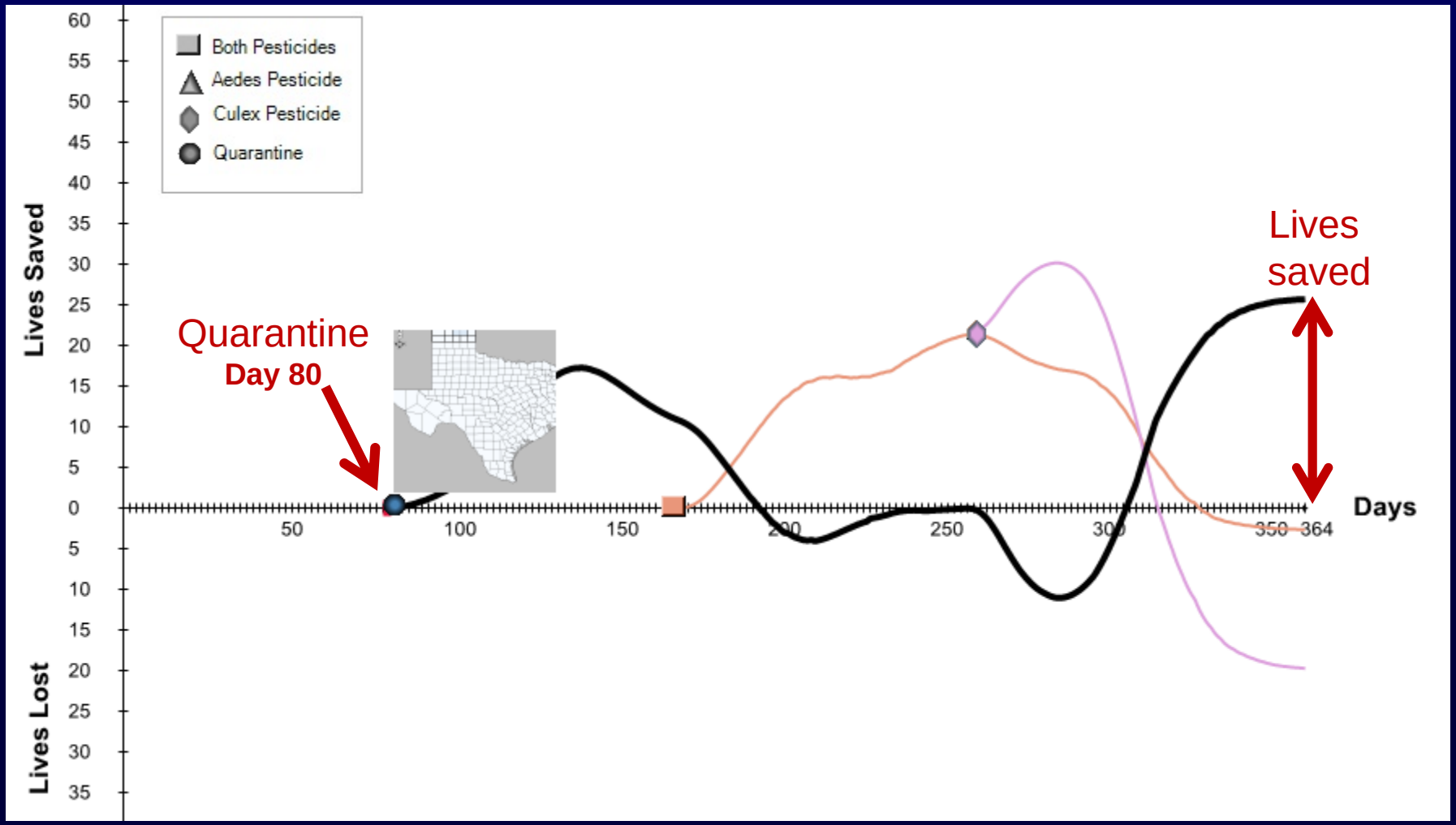


Decision History Visualization & Navigation View
Rift Valley Fever



VACCINE

Decision History Visualization & Navigation View
Rift Valley Fever



VACCINE

Case Studies

Case Study 1 – Rift Valley Fever

Description:

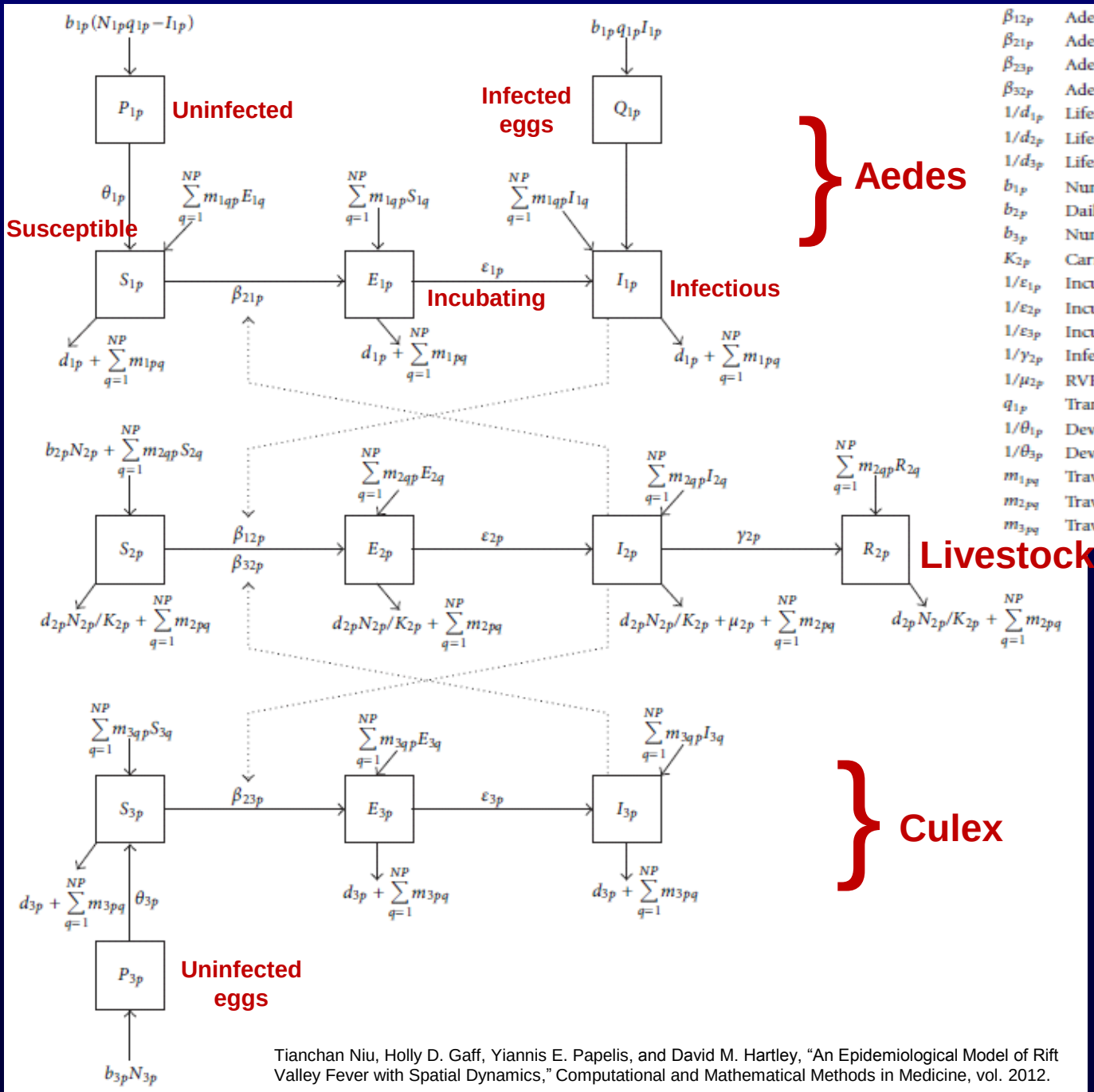
- Epidemiological model that simulates Rift Valley Fever spread in Texas[2]
- Mosquito Species (Aedes & Culex)
- Infects Cattle Population

Mitigative Response Measures:

1) Pesticides (kill Aedes and Culex Mosquito population)
2) Quarantine (Block transport of cattle into or out of the selected region)

---

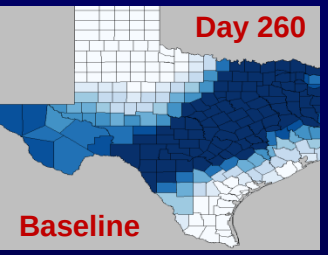
[2]GAFF H. D., HARTLEY D. M., LEAHY N. P.: An epidemiological model of Rift Valley Fever. Electron. J. Diff. Eqns. 115 (2007)

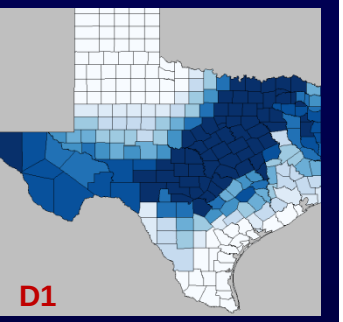


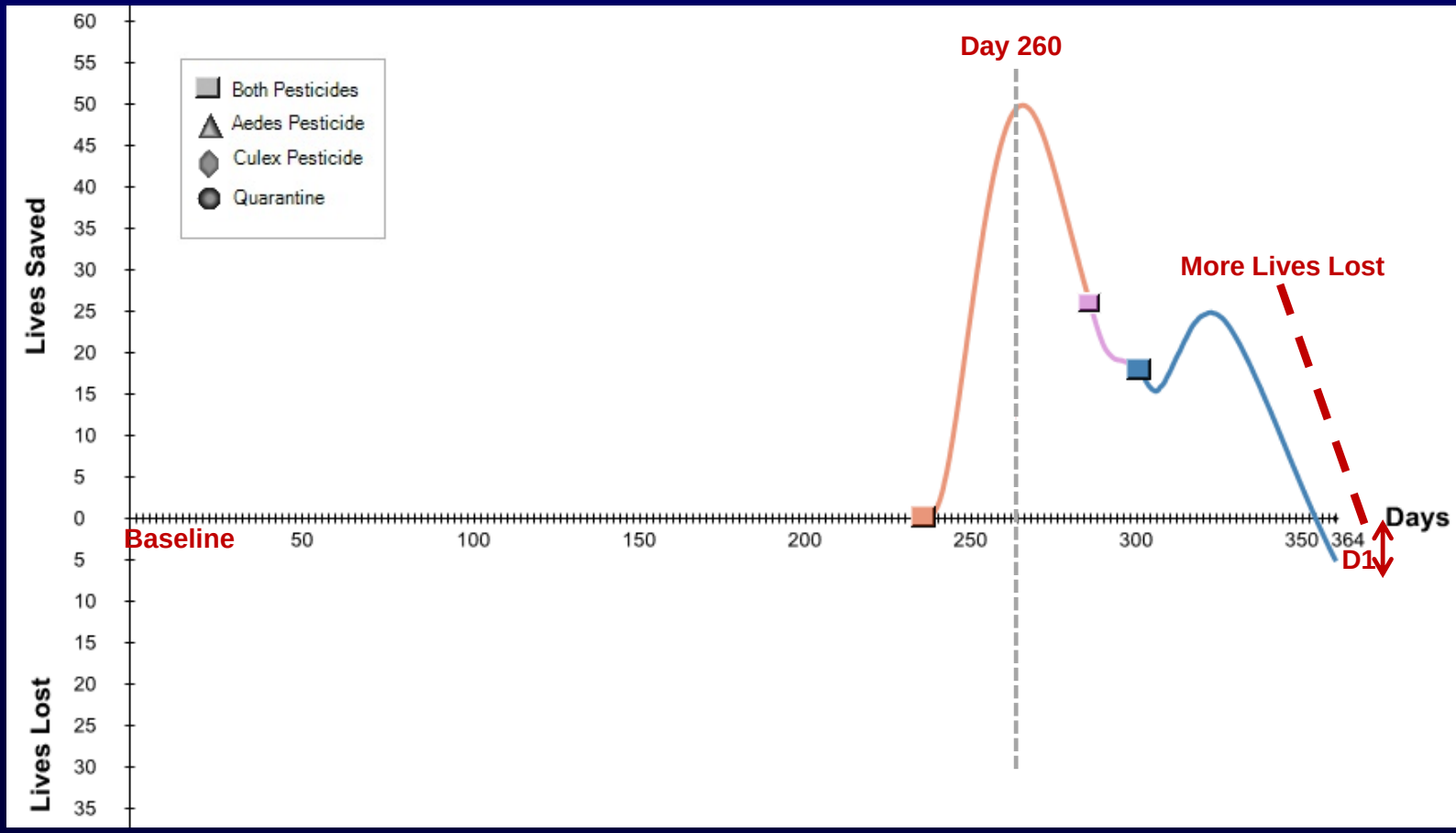
β_{12p}	Adequate contact rate from *Aedes* to livestock on patch p
β_{21p}	Adequate contact rate from livestock to *Aedes* on patch p
β_{23p}	Adequate contact rate from livestock to *Culex* on patch p
β_{32p}	Adequate contact rate from *Culex* to livestock on patch p
$1/d_{1p}$	Lifespan of *Aedes* mosquitoes on patch p
$1/d_{2p}$	Lifespan of livestock on patch p
$1/d_{3p}$	Lifespan of *Culex* mosquitoes on patch p
b_{1p}	Number of *Aedes* eggs laid per day on patch p
b_{2p}	Daily birthrate of livestock on patch p
b_{3p}	Number of *Culex* eggs laid per day on patch p
K_{2p}	Carrying capacity of livestock on patch p
$1/\varepsilon_{1p}$	Incubation period in *Aedes* mosquitoes on patch p
$1/\varepsilon_{2p}$	Incubation period in livestock on patch p
$1/\varepsilon_{3p}$	Incubation period in *Culex* mosquitoes on patch p
$1/\gamma_{2p}$	Infectiousness period in livestock on patch p
$1/\mu_{2p}$	RVF mortality rate in livestock on patch p
q_{1p}	Transovarial transmission fraction in *Aedes* on patch p
$1/\theta_{1p}$	Development time of *Aedes* on patch p
$1/\theta_{3p}$	Development time of *Culex* on patch p
m_{1pq}	Travel rate of *Aedes* mosquitoes from patch p to patch q
m_{2pq}	Travel rate of livestock from patch p to patch q
m_{3pq}	Travel rate of *Culex* mosquitoes from patch p to patch q

Tianchan Niu, Holly D. Gaff, Yiannis E. Papelis, and David M. Hartley, "An Epidemiological Model of Rift Valley Fever with Spatial Dynamics," Computational and Mathematical Methods in Medicine, vol. 2012.

Case Study 1 – Rift Valley Fever





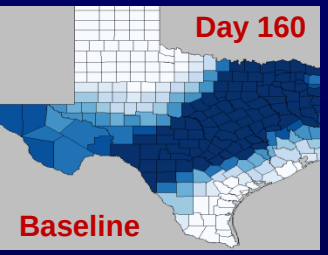


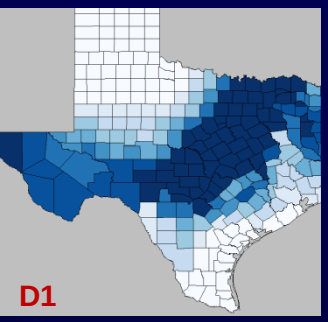
Decision Path D1

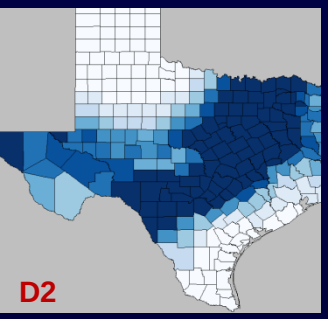
Spray Both Pesticides **Day 240, 290, 305**

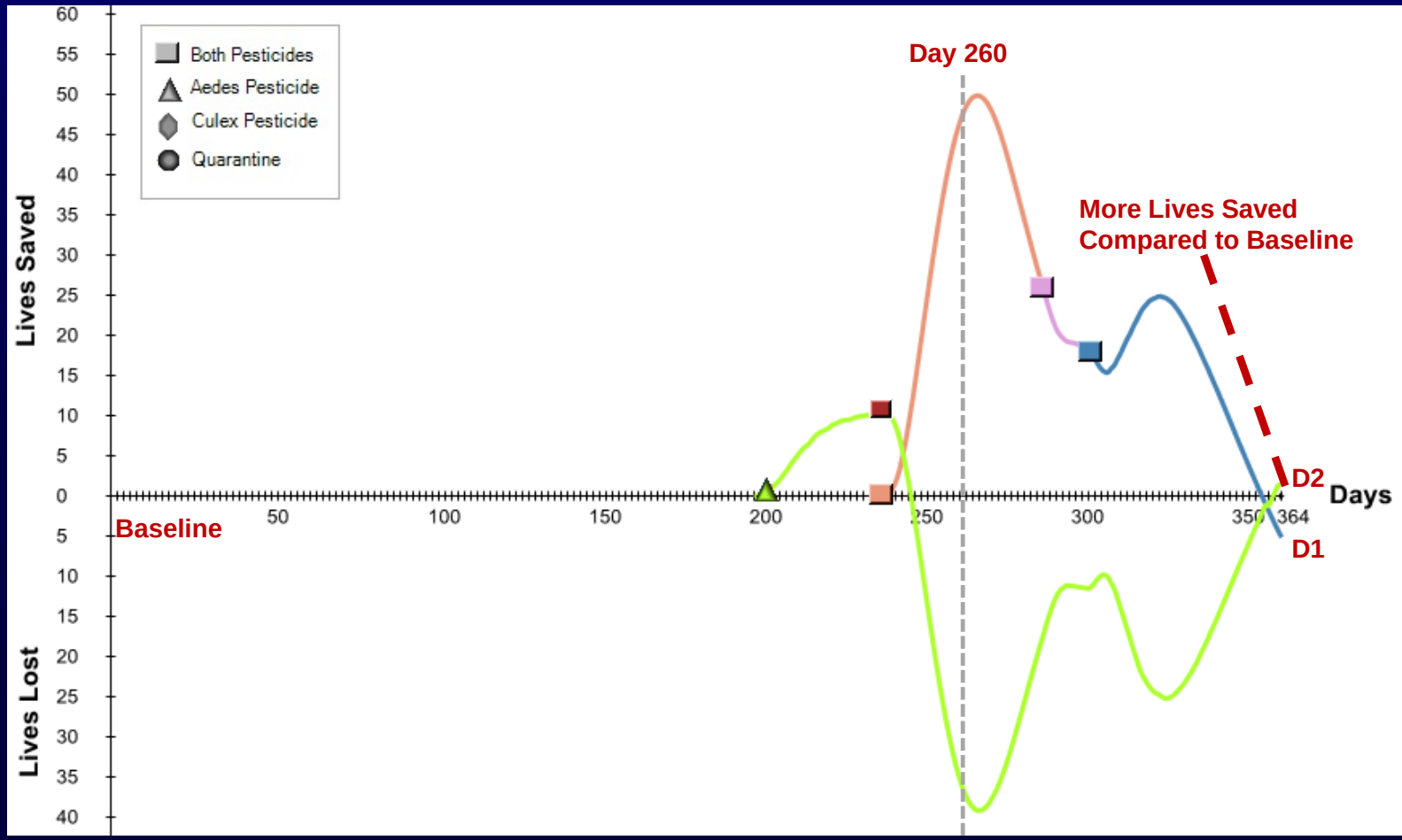
VACCINE

Case Study 1 – Rift Valley Fever







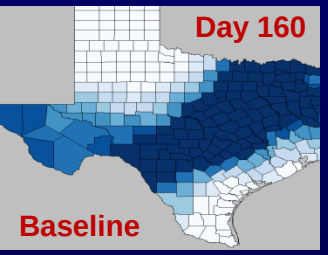


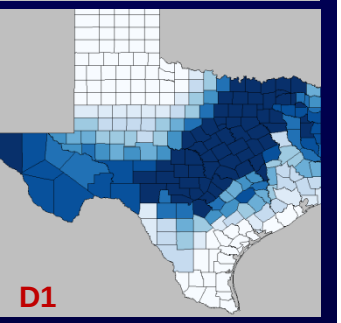
Decision Path D2

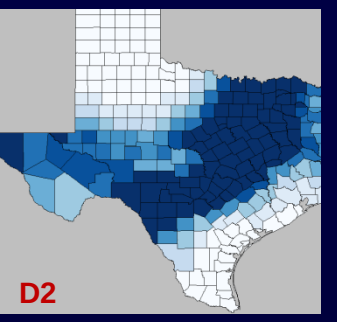
Spray	**Aedes Pesticides**	**Day 200**
	Both Pesticides	**Day 240**

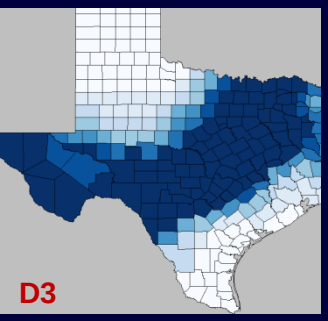
VACCINE

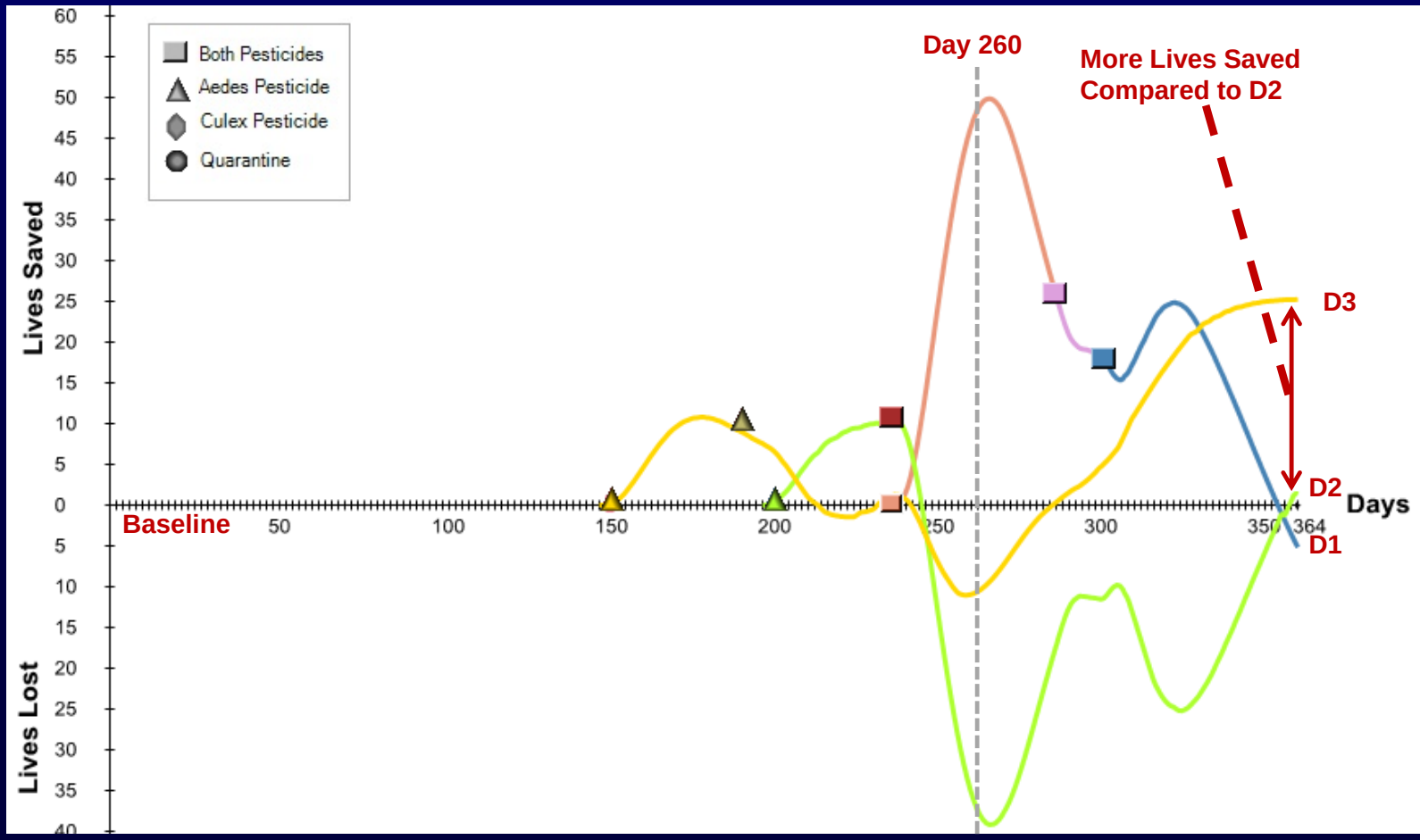
Case Study 1 – Rift Valley Fever









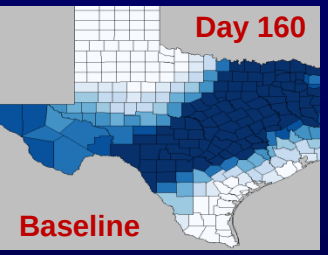


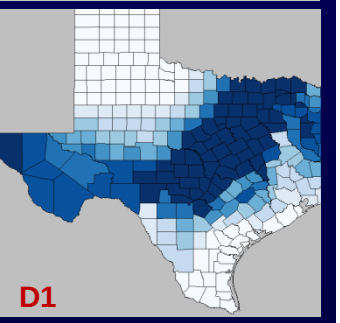
Decision Path D3

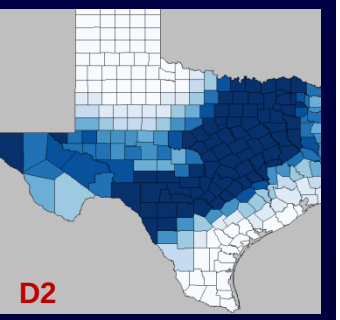
Spray	Aedes Pesticides	Day 150
	Aedes Pesticides	Day 190
	Culex Pesticides	Day 320

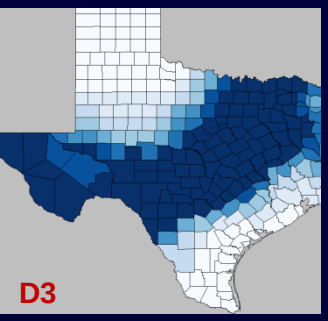
VACCINE

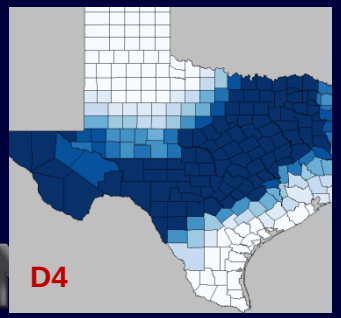
Case Study 1 – Rift Valley Fever

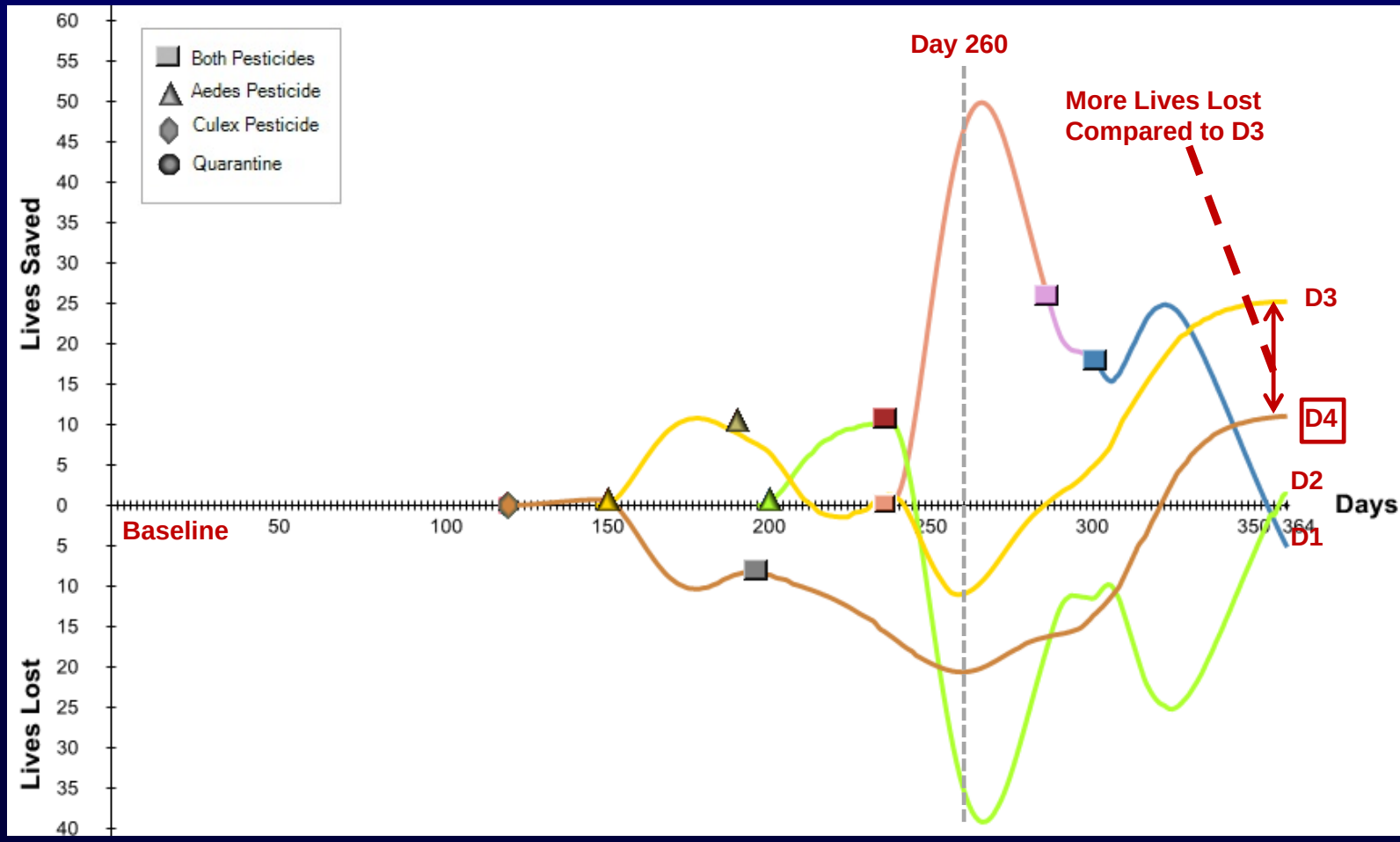








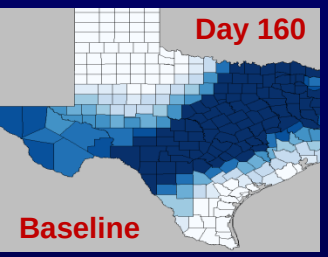


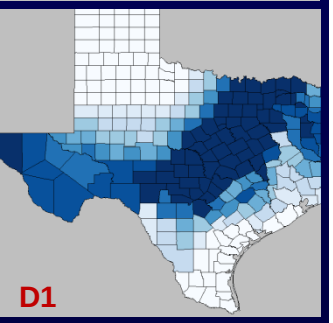


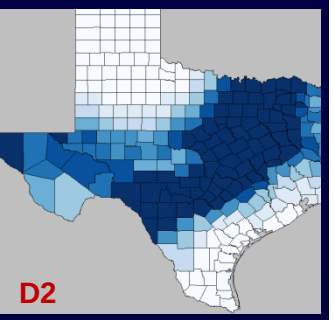
Decision Path D4

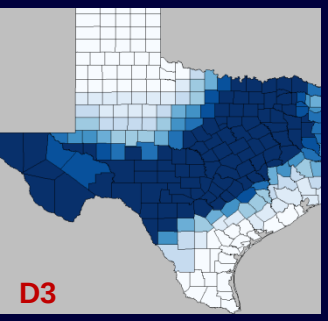
Spray	Culex	Pesticides	Day 120
	Both	Pesticides	Day 200

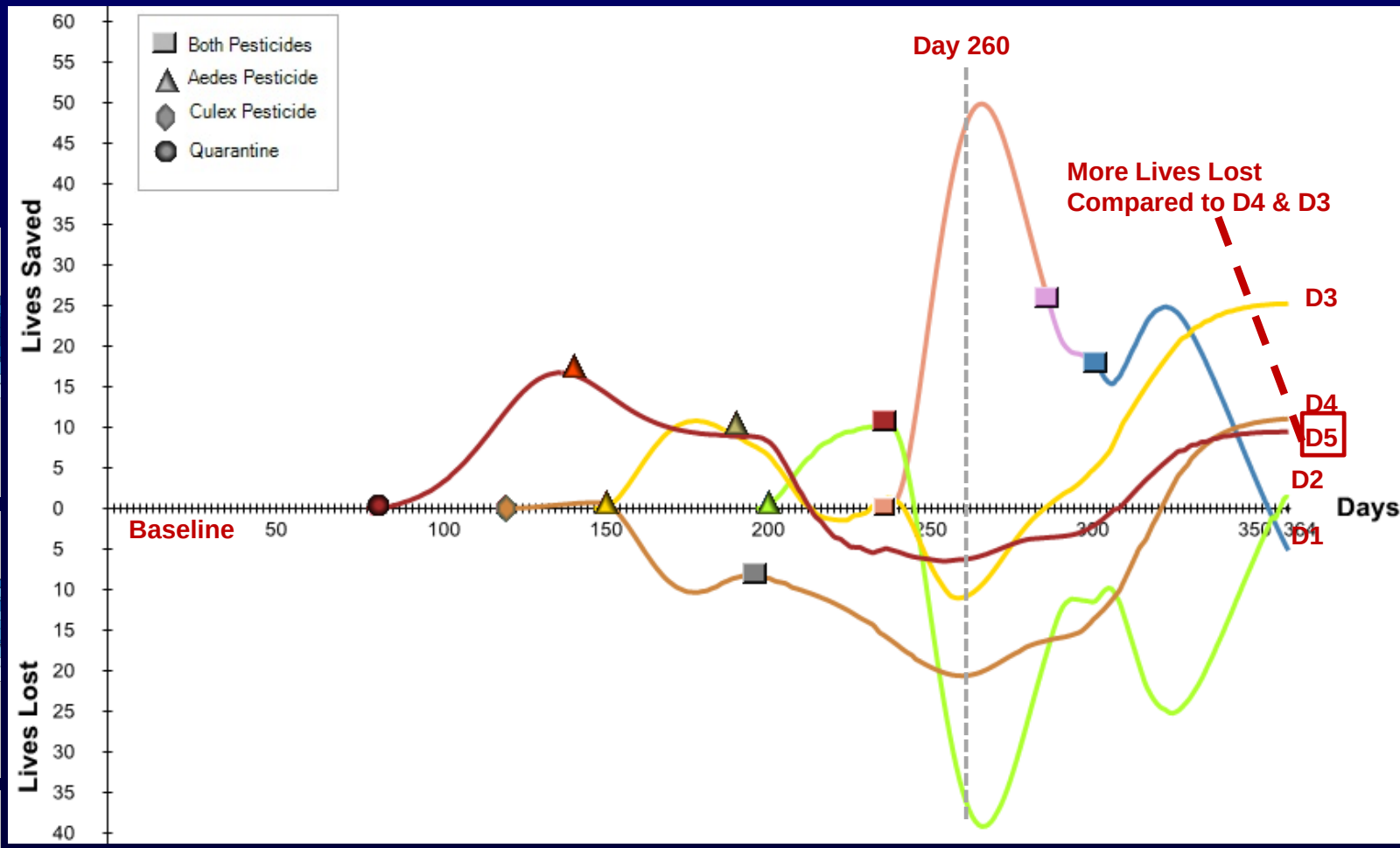
Case Study 1 – Rift Valley Fever

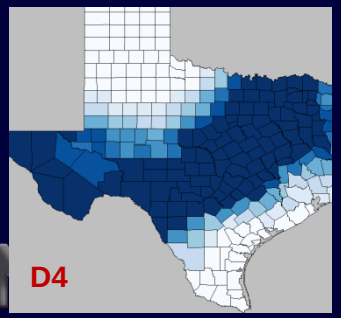






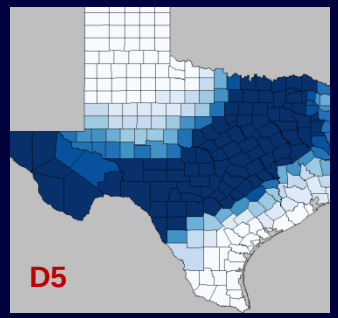




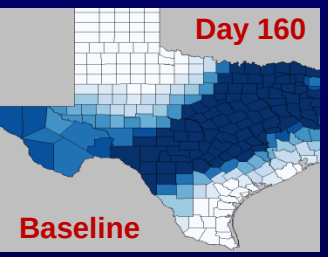


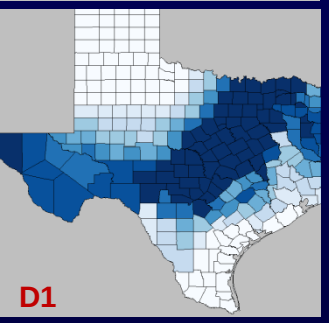
Decision Path D5

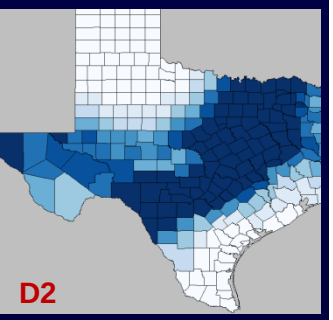
Quarantine	Day 80
Spray Aedes Pesticides	Day 140
Spray Both Pesticides	Day 260

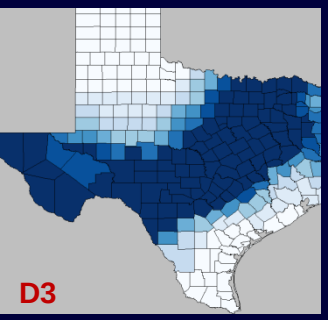


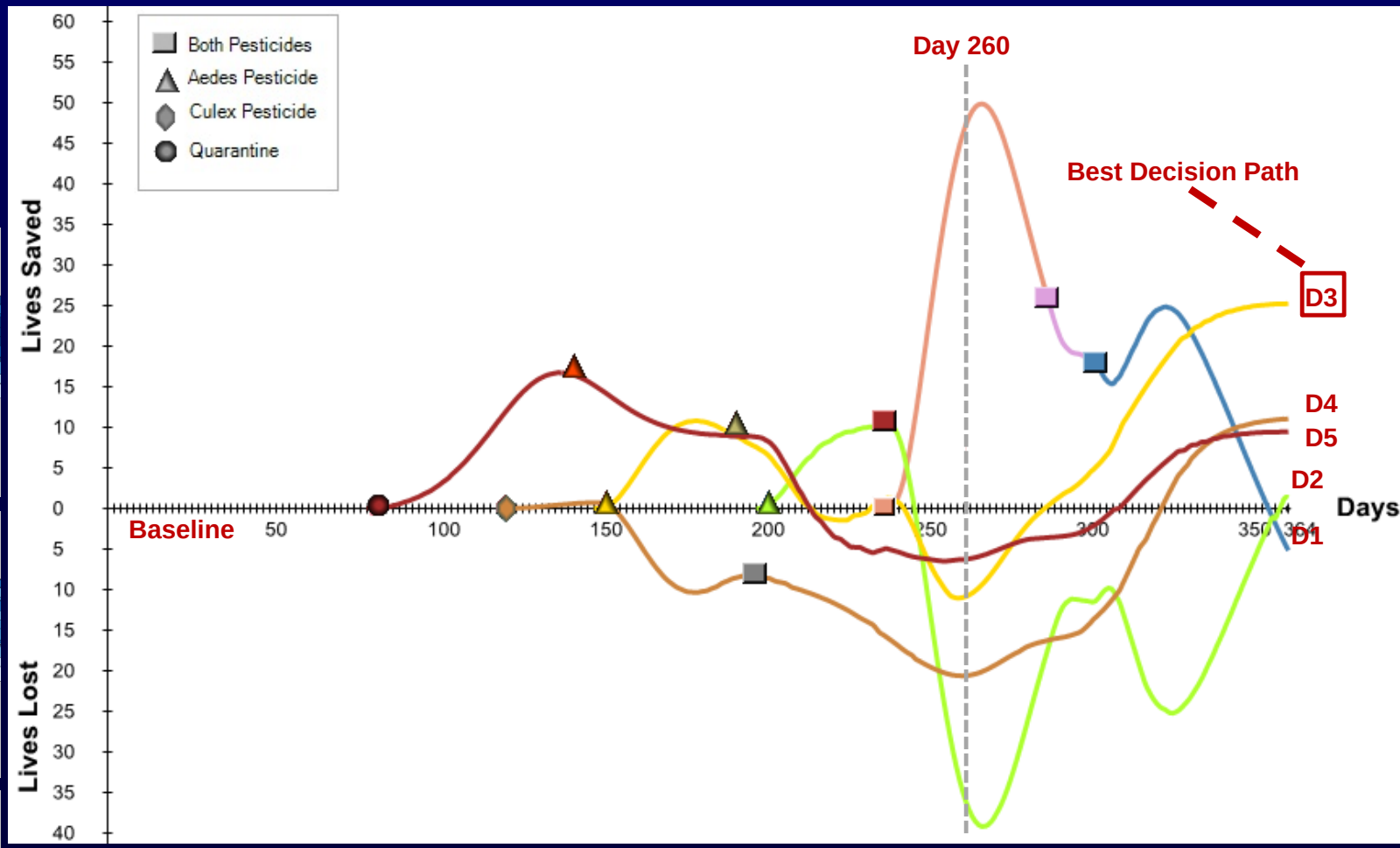
Case Study 1 – Rift Valley Fever

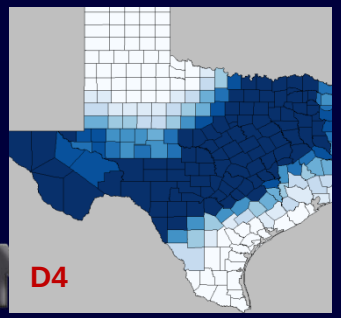






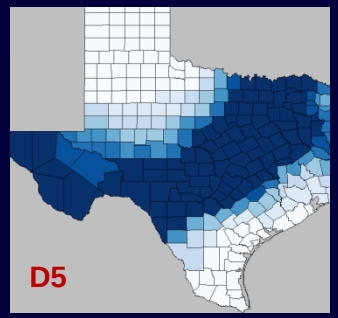






Best Decision Path D3

Spray	**Aedes Pesticides**	**Day 150**
	Aedes Pesticides	**Day 190**
	Culex Pesticides	**Day 320**



Demo Video

Case Study 2 – Pandemic Influenza

Description:

- Epidemic Model that simulates the spread of a pandemic influenza across the US[1]
- User defined point source location
- Incorporates airport traffic models

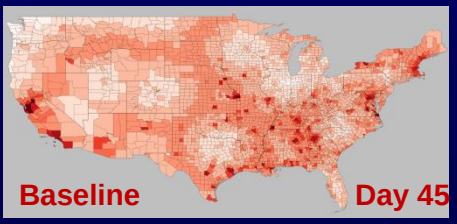
Mitigative Response Measures:

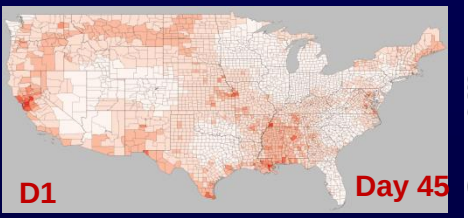
1) Media Alerts
2) Strategic national stockpile deployment (SNS)
3) School Closure

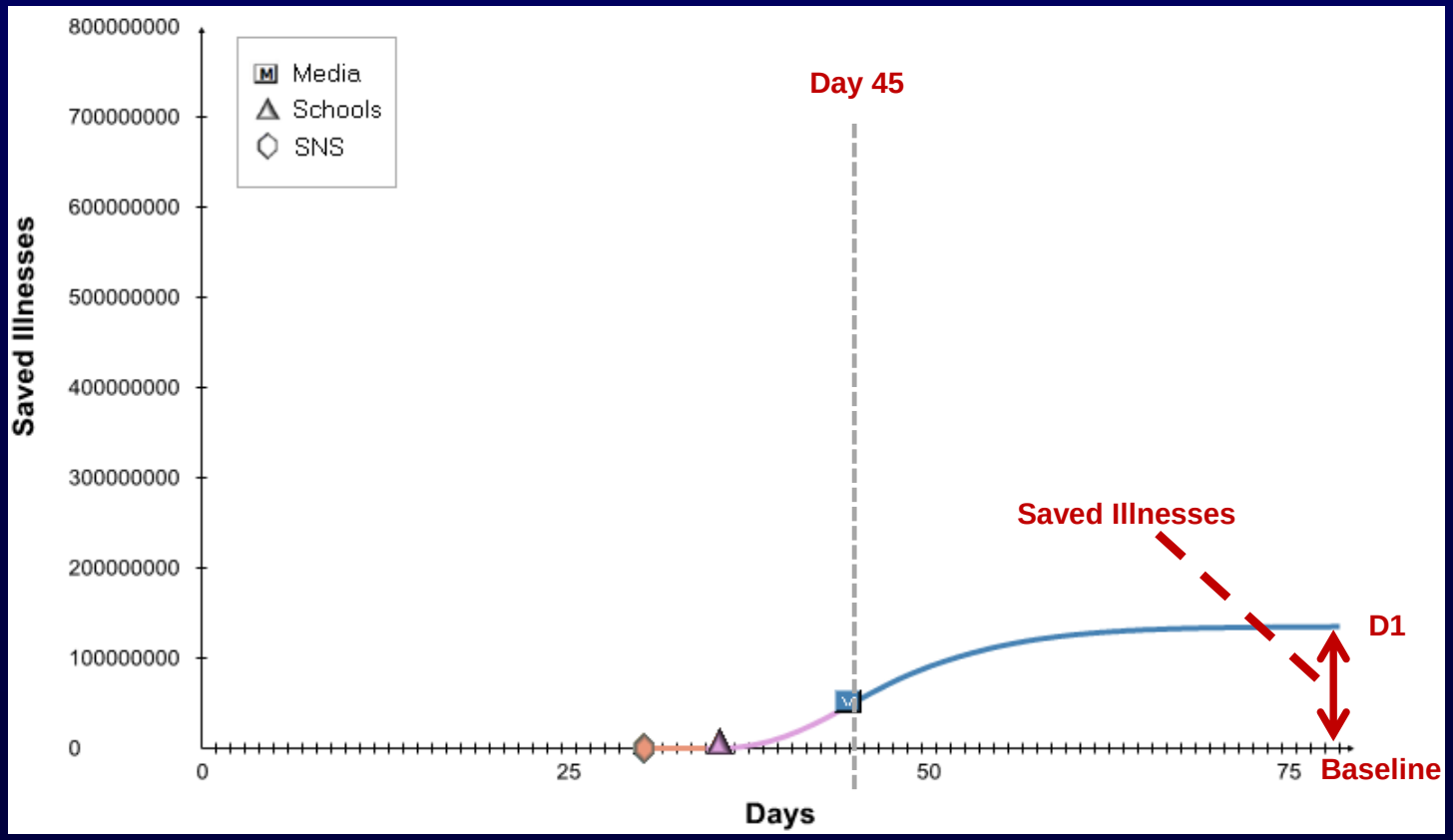
---

[1]R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders. A pandemic influenza modeling and visualization tool. Journal of Visual Languages and Computing (2011).

Case Study 2 – Pandemic Influenza





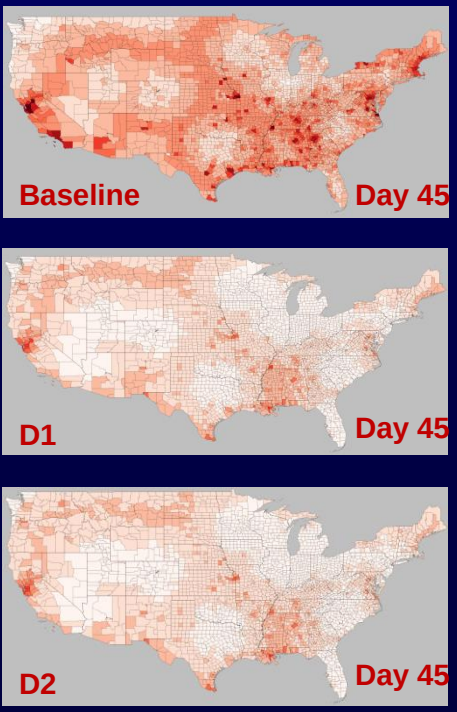


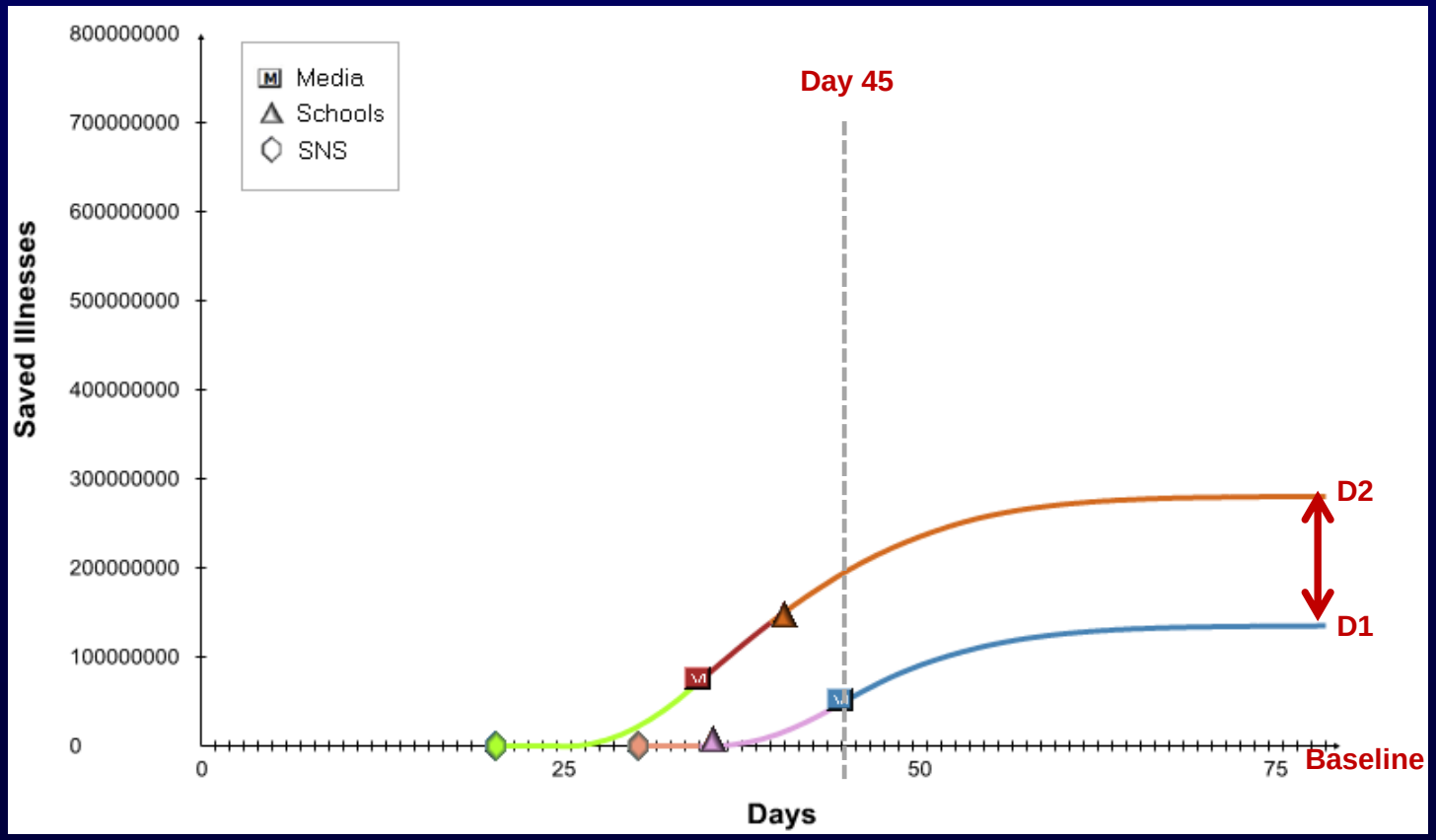
Decision Path D1

SNS Day 30 **School Closure Day 35** **Media Day 45**

VACCINE

Case Study 2 – Pandemic Influenza



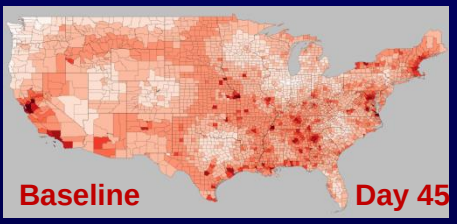


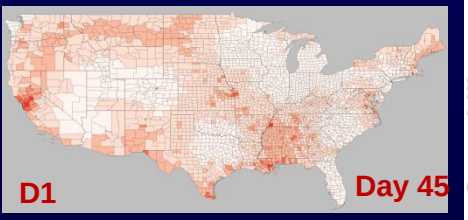
Decision Path D2

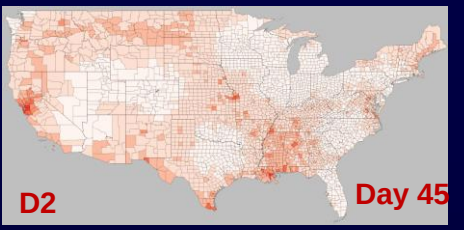
SNS Day 20 **Media Day 34** **School Closure Day 40**

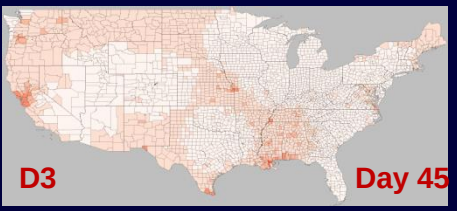
VACCINE

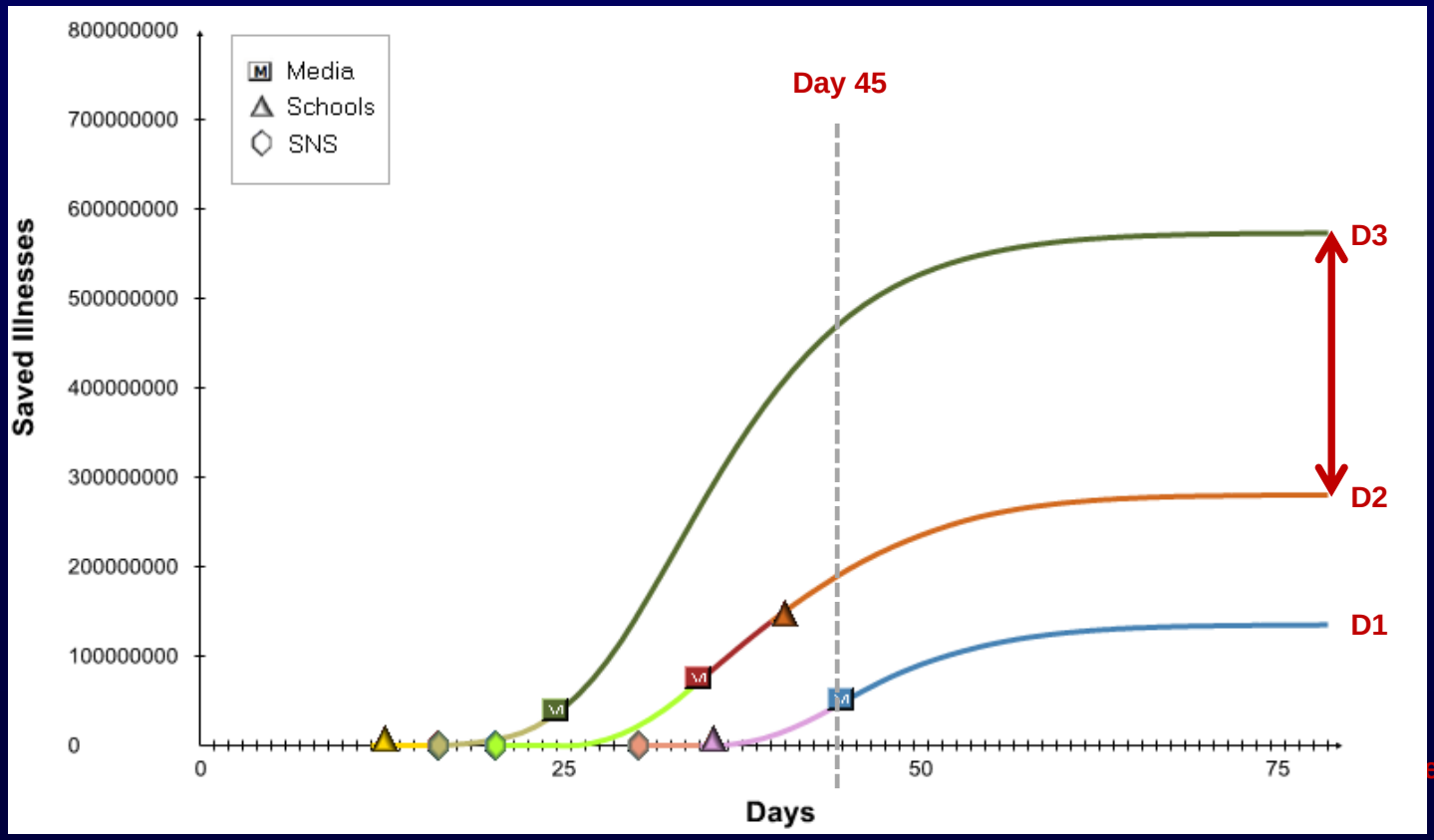
Case Study 2 – Pandemic Influenza









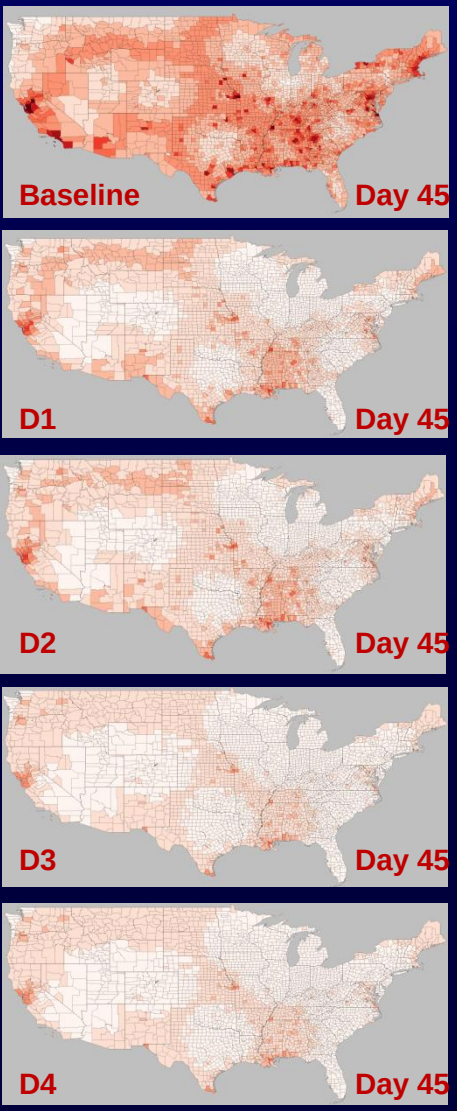


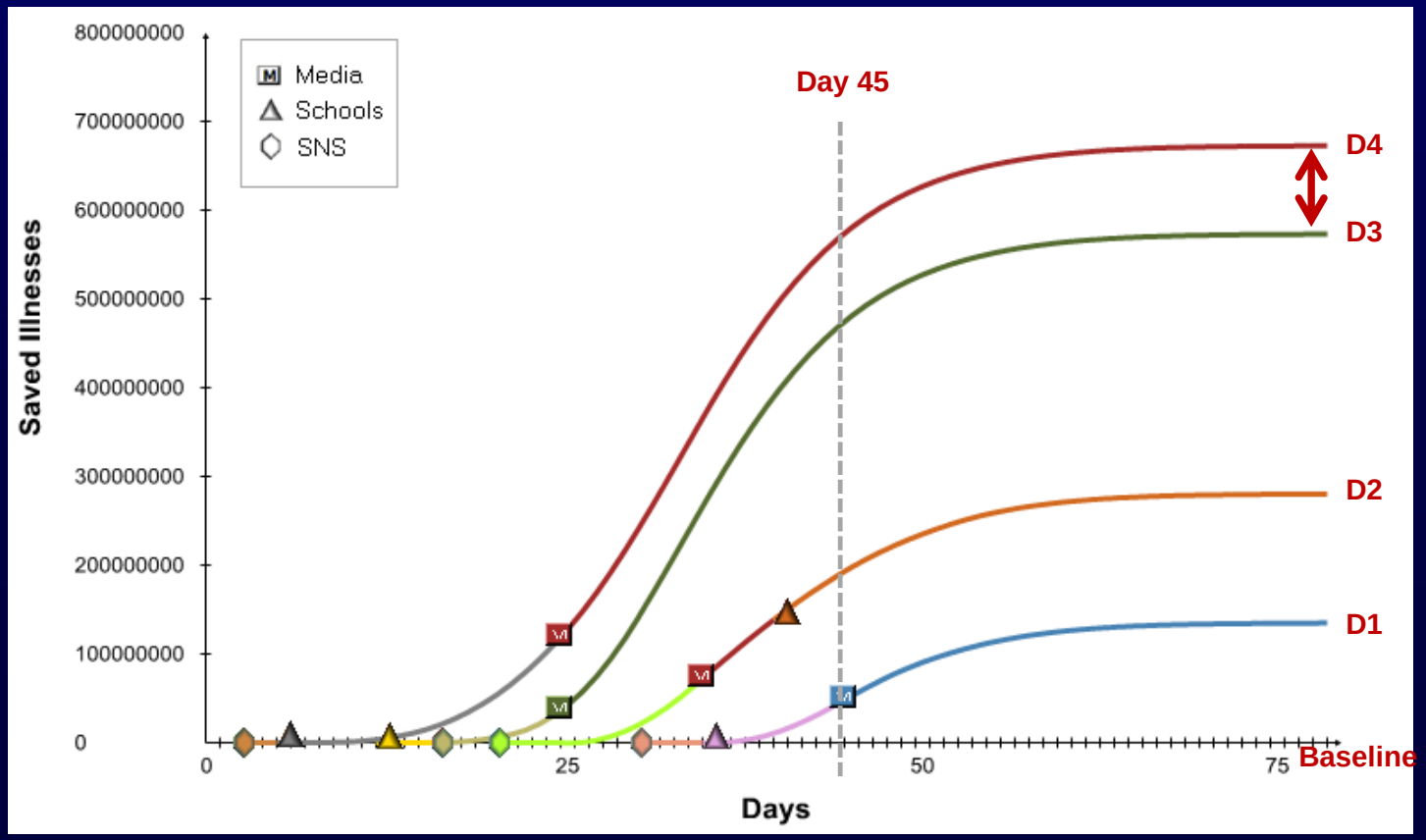
Decision Path D3

School Closure Day 12 **SNS Day 16** **Media Day 25**

VACCINE

Case Study 2 – Pandemic Influenza





Decision Path D4

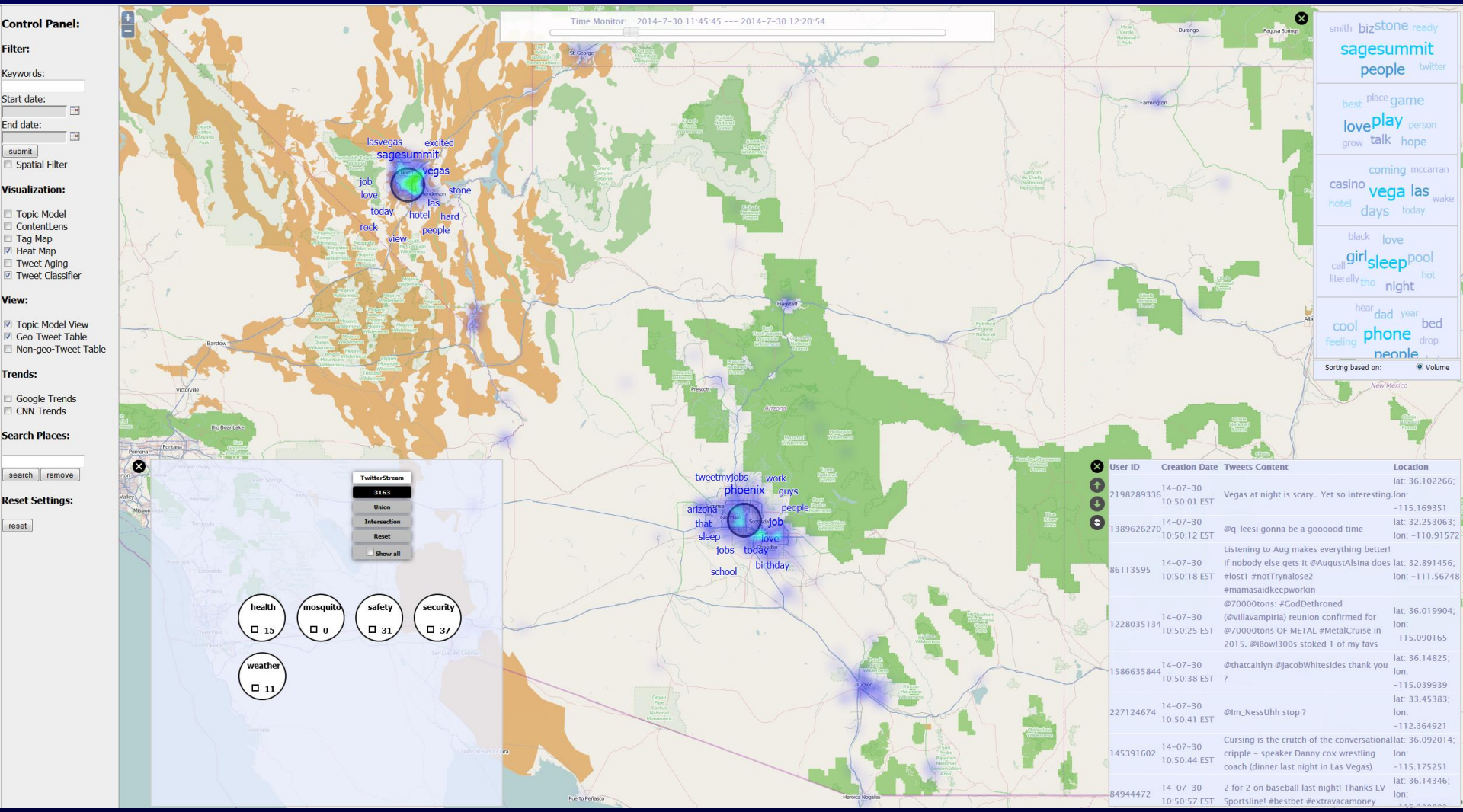
SNS Day 2 **School Closure Day 6** **Media Day 25**

VACCINE

Future Extensions

- Adaptation of RVF Model for Dengue Fever
- Integrate financial models to study the impact of disease outbreaks on local economy
- Combine data form other sources like social media data and actual outbreak data

Social Media Analytics and Reporting Toolkit (SMART)



VACCINE

SMART Features

- Interactive analysis and visualization of large volumes of Twitter Data
- Load live streaming data as well as historical data for specific geographical region
- Geospatial map view can visualize geo-tagged tweet locations
- Automatic Identification of emerging topics and trends
- STL based anomalous event detection

VACCINE

SMART Features

- Create custom filters to classify tweets related to a certain category e.g., epidemic outbreaks, weather and safety.
- Examine tweet clusters using ‘Content Lens’
- Drill down capabilities to examine actual tweet contents
- Integration of external news sources like Google and CNN trends
- Web based front end capable of running in a web browser

Applications in Epidemic Modeling

- Early detection of disease outbreaks and their locations
- Combine model simulation output or chief complaints data with social media data to improve outbreak prediction
- Study the impact of different factors like weather conditions, flooding and health facilities on disease outbreaks
- Identifying spread patterns and identifying modes of transmission

Demo

Acknowledgements

- David Hartley and Tianchan Niu for providing details on the Rift Valley fever simulation

- Min Chen, Harold Bosch, Denis Thom and Sherry Towers for their helpful discussions

- This work was supported in part by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0002, the Foreign Animal and Zoonotic Disease Center, and the Defense Threat Reduction Agency under Award Number HDTRA 1-10-083

VACCINE

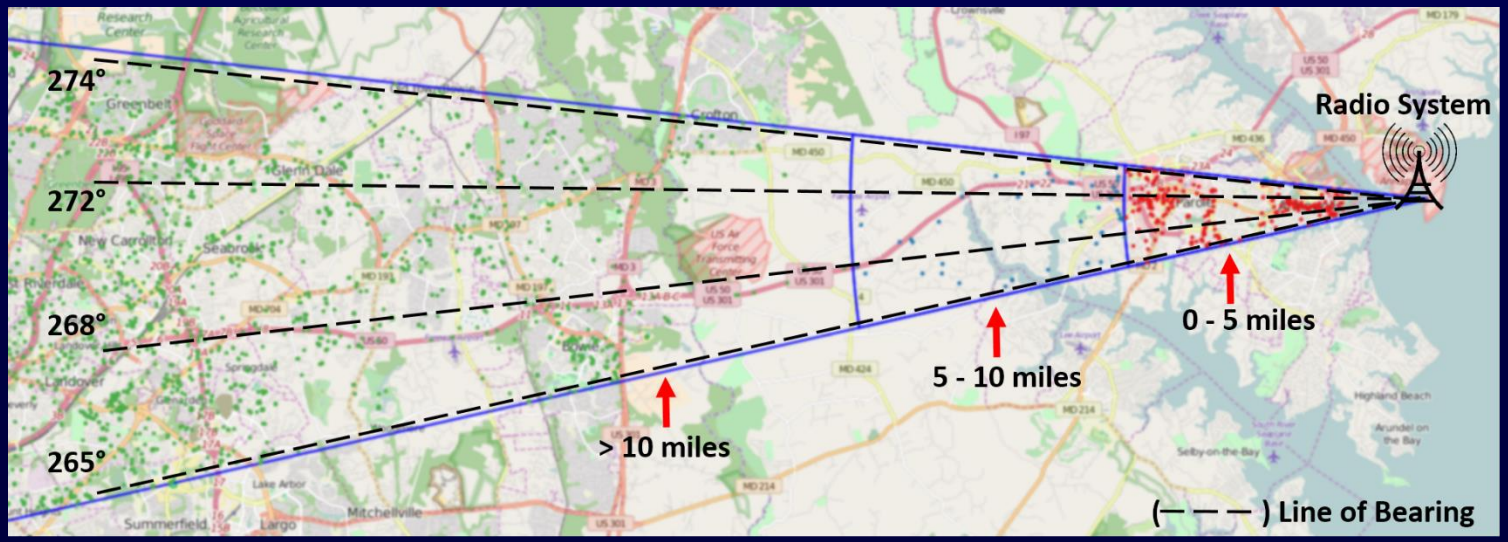
Thank you

Shehzad Afzal
safzal@purdue.edu

http://www.purdue.edu/discoverypark/vaccine

VACCINE

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Junghoon Chae[1], Jiawei Zhang[1], Sungahn Ko[2], Abish Malik[1], Heather Connell[1], David S. Ebert[1]

Purdue University[1], Ulsan National Institute of Science and Technology[2]

May 11, 2016

VACCINE

Introduction

- The U.S. Coast Guard (USCG) receives 139 hoax distress calls nationally per year
- Response to false calls wastes considerable resources
  - Put Coast Guard rescue personnel in dangerous situations
  - Waste valuable search and rescue assets that should be used for real emergency situations

Challenge

- Hard to identify hoax callers
- Rescue 21 (USCG radio communication system)
  - Utilized for identification of suspected hoax calls
  - Provide not enough information for finding hoax callers
- Need more intelligent analytics support

Goals

- A social media visual analytic driven by the Rescue 21 system
- Integrating the tool into SMART (Social Media Analytics Reporting Toolkit)
- Improving investigation of hoax calls by identifying suspicious social media

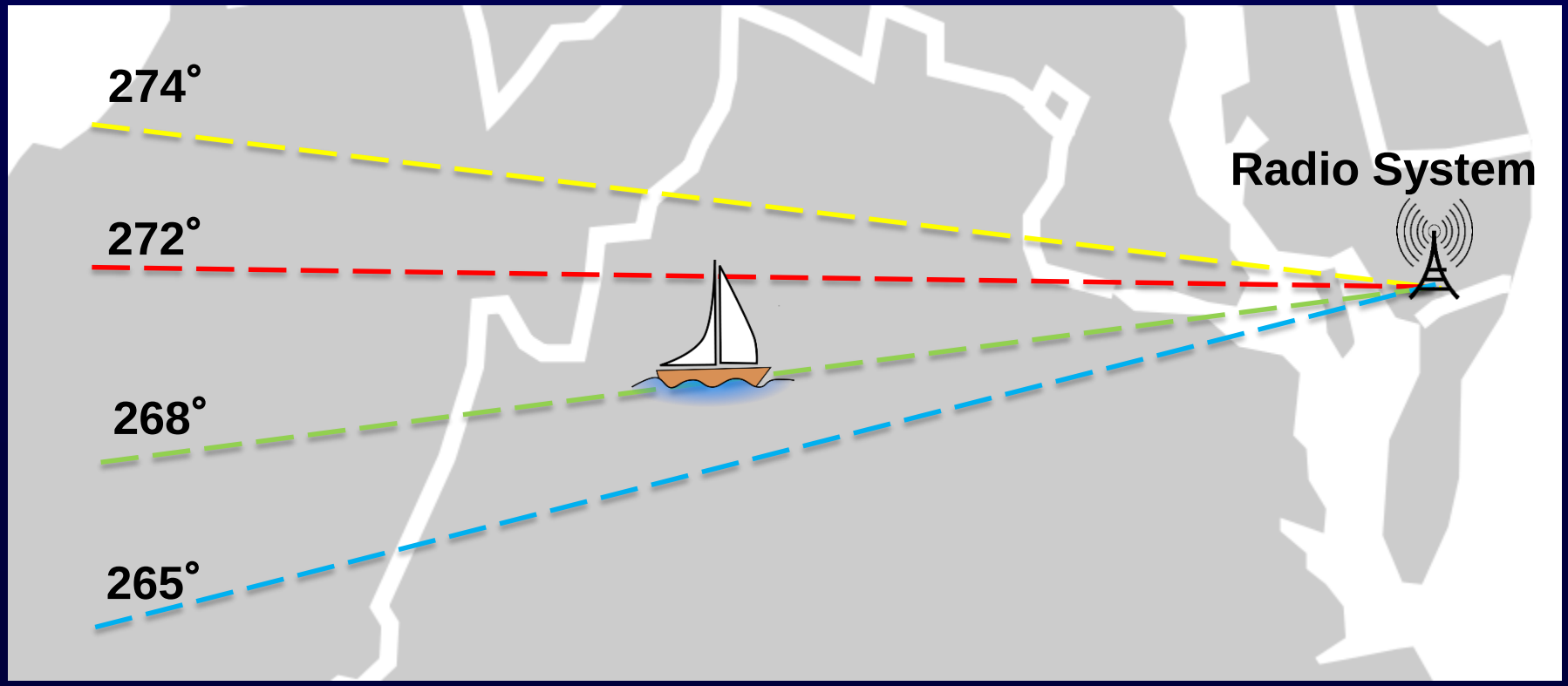
Approach

- Interactive visual analytics using social media
- Keyword collection & Data enrichment
- Identification of suspicious users & messages

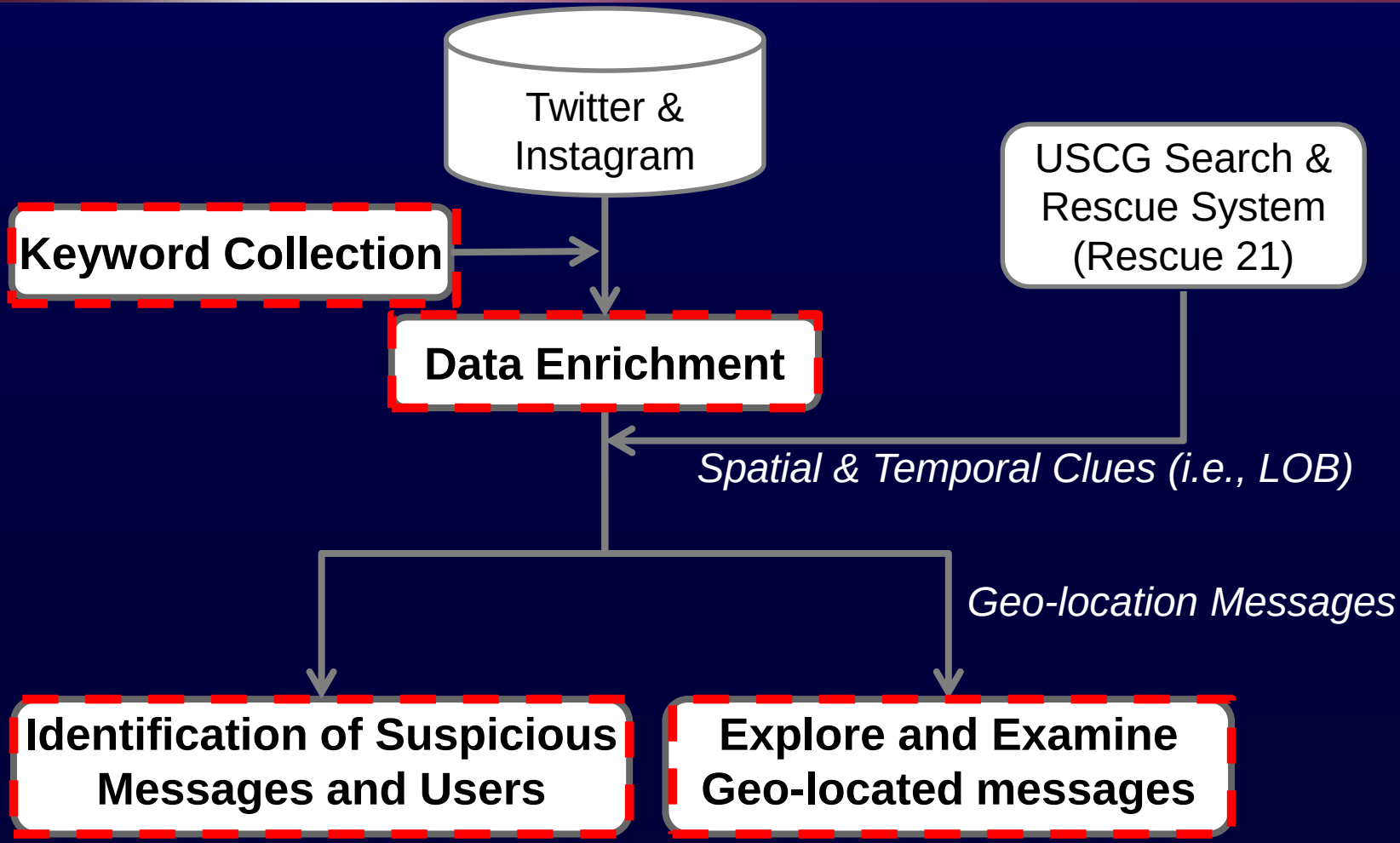
Background: Rescue 21

- USCG's radio communication system with direction-finding capability for search and rescue operations
- Providing lines of bearing (LOB) from a radio tower to the source of radio calls that it receives
- LOB is used for locating transmitters which are not able to specify their locations

Example of LOB



Analysis Process



Keyword Collection

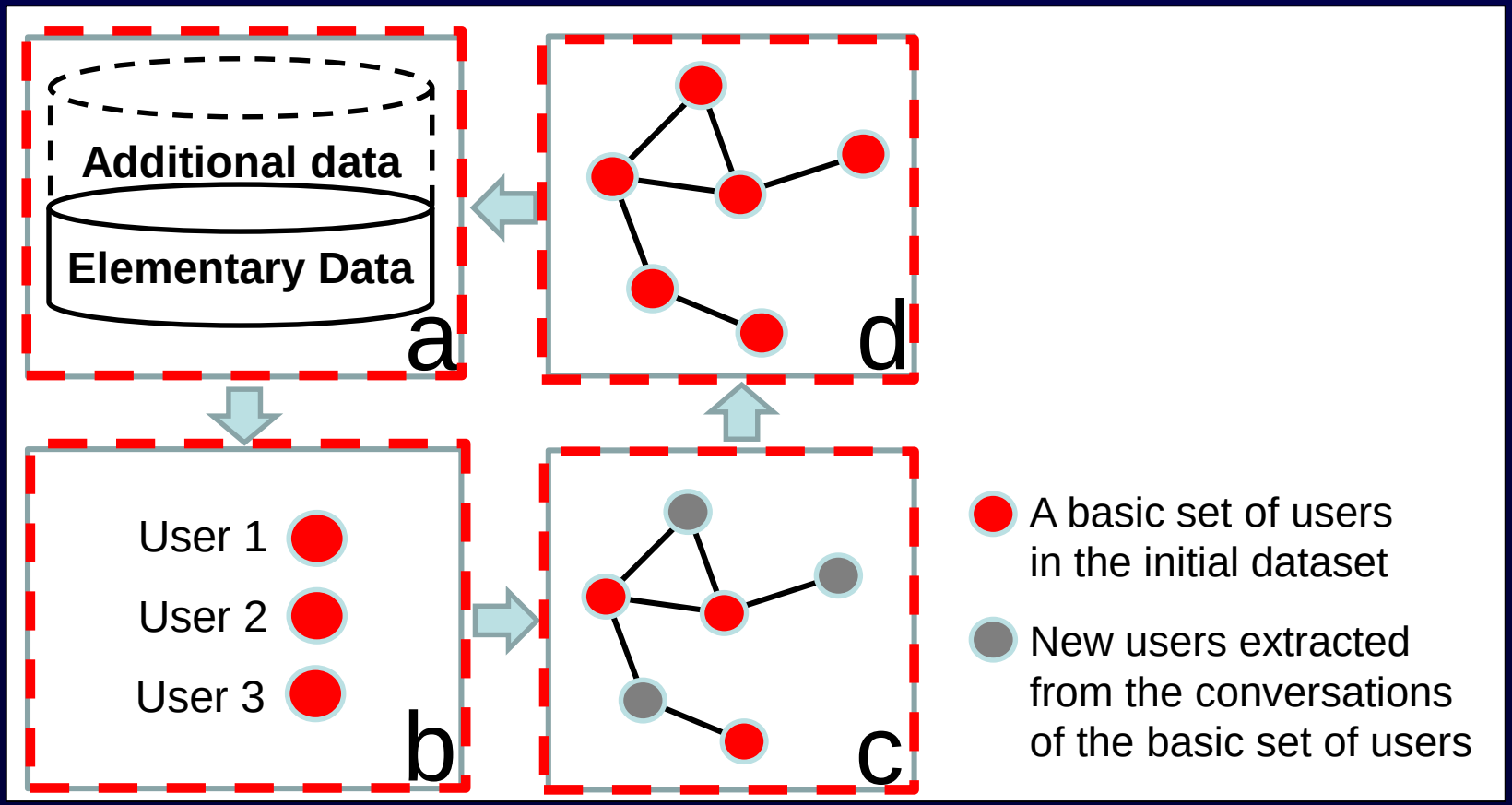
- Building a keyword corpus related to USCG and hoax distress calls
  - Retrieve highly relevant social media data
  - Reduce the analysis space and enable effective analysis tasks
- Adding words with semantic relationship to the initial keywords using WordNet
- USCG related online articles and historical social media data

Keyword Collection

Category	Initial Keywords	Extended Keywords
Coast Guard	USCG, vessel, ship, boat, ferry, shore, coastline ...	#cg, #coastguard, maritime, navy, fleet, lifeboat, craft ...
Hoax Call	**prank call, distress call, mayday, crank call, hoax ...**	**siren, save, false call, bogus, stunt, prankster ...**
Search and Rescue	helicopter, seek, search, rescue, flare, flashing ...	#sos, copter, airliner, navigate, 911, warning ...
Accident	explosion, accident, sink, crash, damaged, burning ...	disaster, hazard, collision, shot, smack, hit, break ...
Radio Channel	broadcast, radio, transmit, communicate ...	cable, on the air, airwaves, wireless, phone, report ...
Illegal Immigration	immigrant, immigration, panga, smuggling, drug ...	foreigner, exodus, migration, crossing, resettlement ...

Data Enrichment

- Data enrichment via conversation networks of users



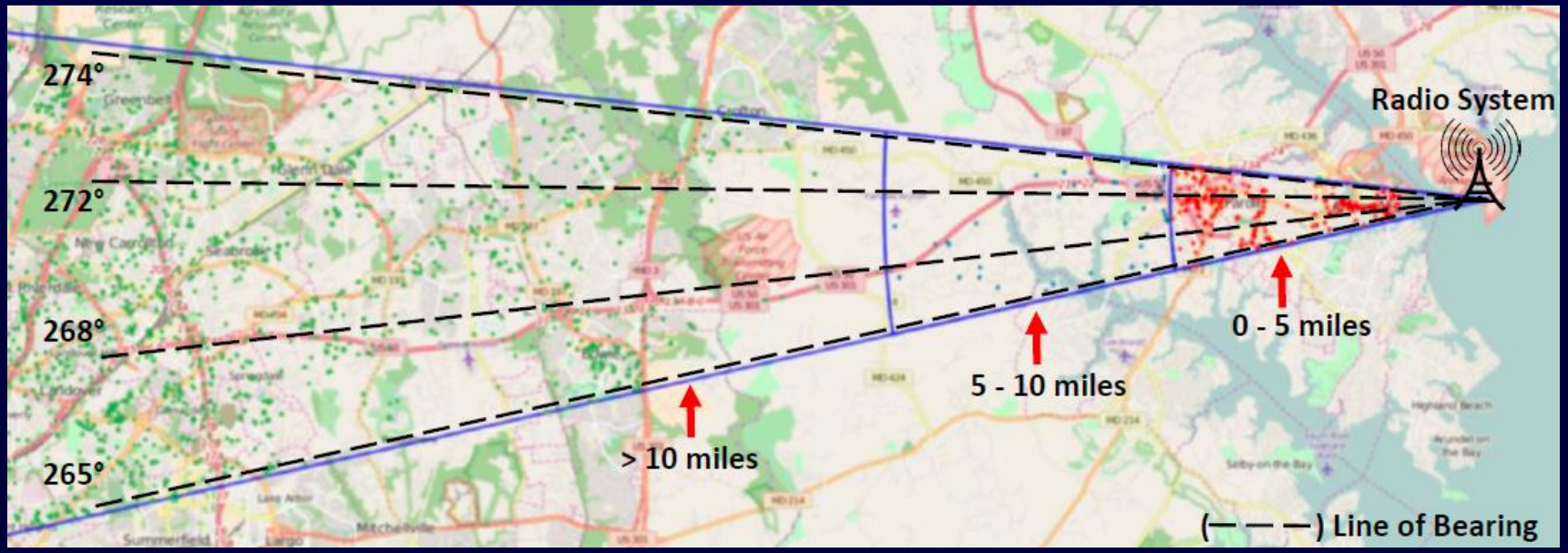
Identification of Suspicious Users

- Multiple hoax distress calls by one caller during different time frames
- Users who post messages within a time window when a hoax call was received are potentially more suspicious than others

For user i, if i posted a message
in a time window of a hoax call
Frequency score[i] += 1

Combination of Visual Analytics and Rescue 21

- Analysis of suspicious geo-tagged social media messages based on spatial and temporal clues driven by a Rescue 21 system



Case Study

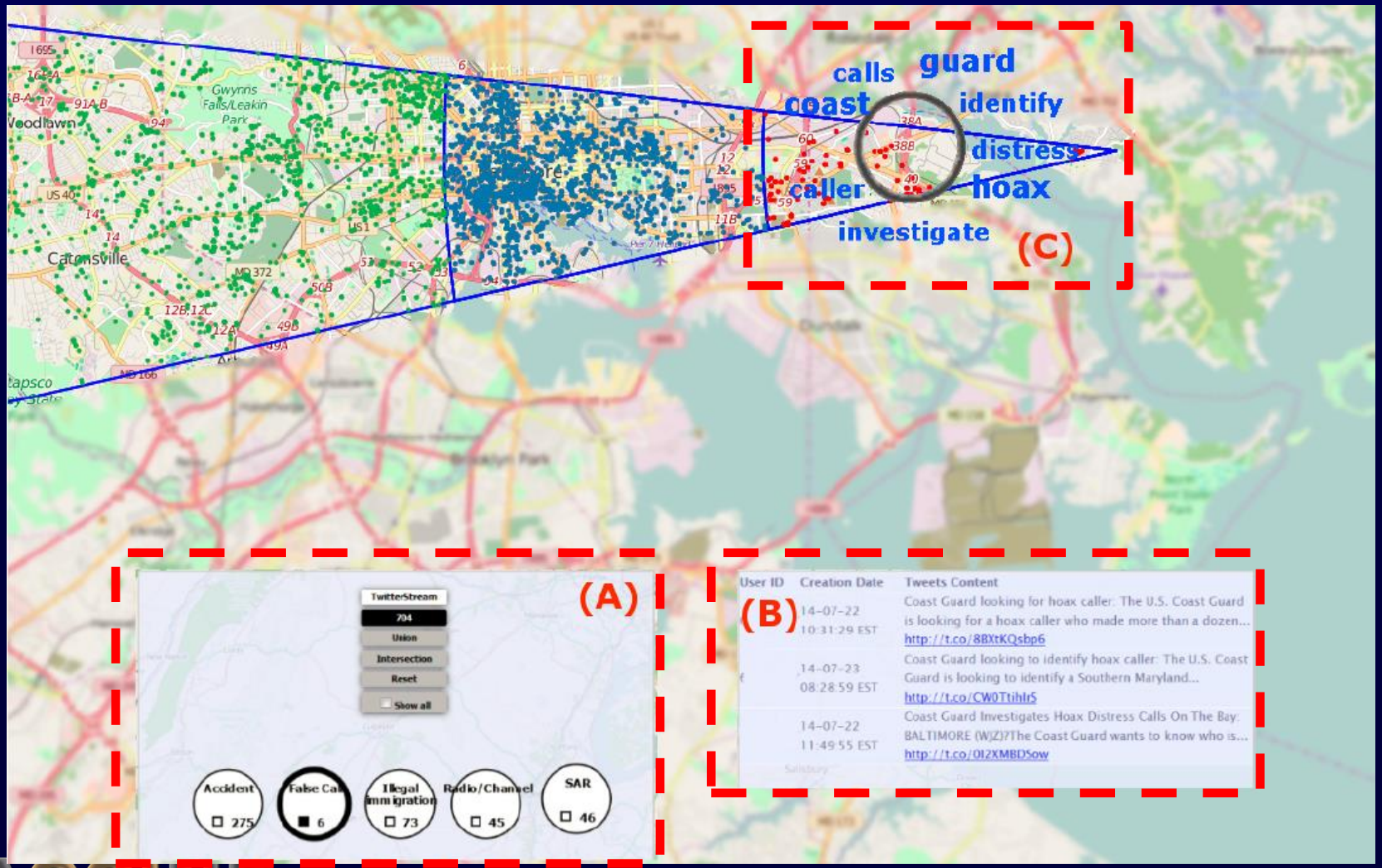
Case	Time	Number of Calls
1	January, 2014	1
2	July and August, 2014	10
3	May, 2014	3

<Information of the hoax distress call cases>

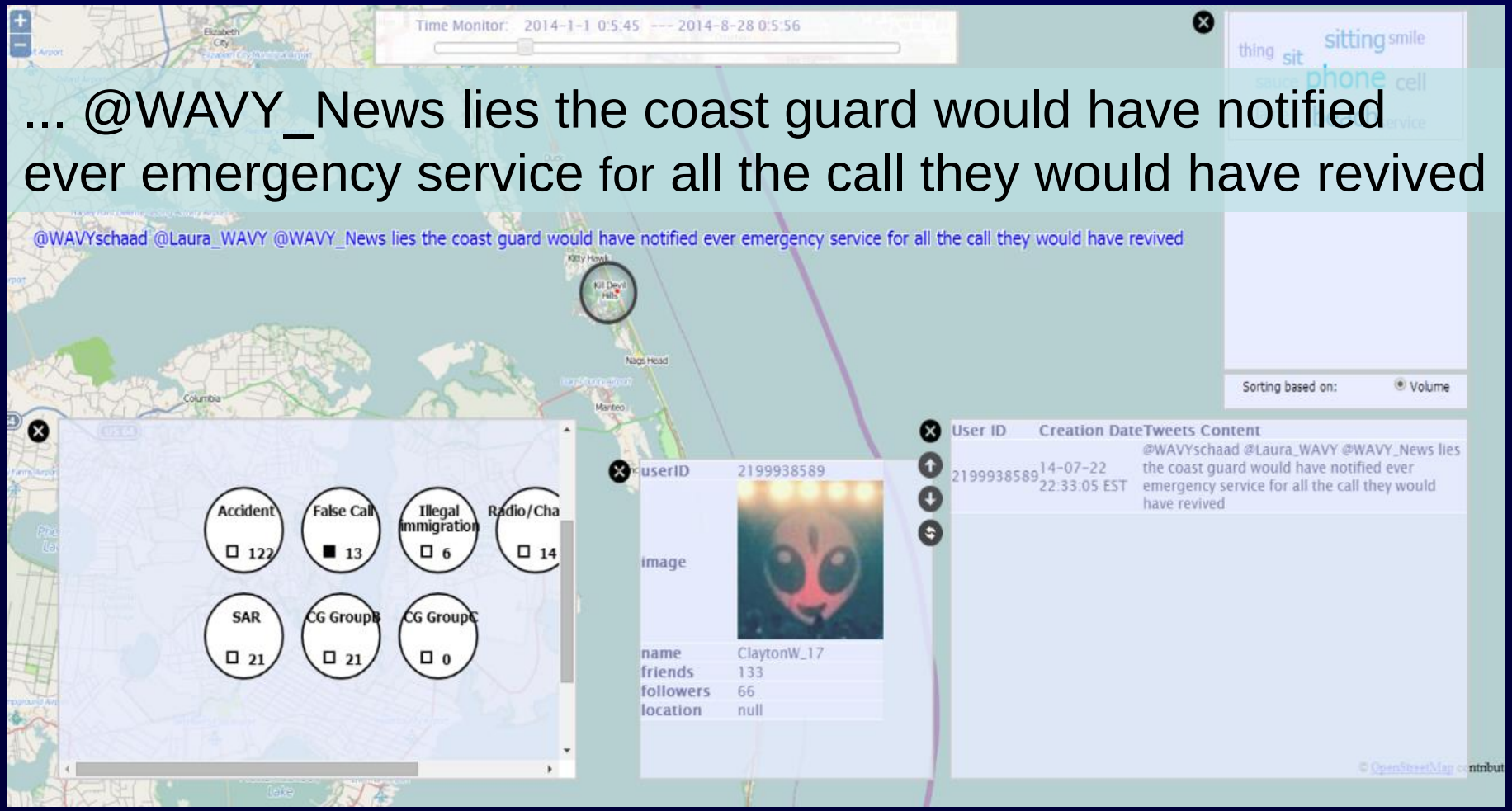
Data Type	Frequency Score						
	6	5	4	3	2	1	total
Twitter	1	4	4	15	60	269	353
Instagram			2	3	20	312	337

<Distribution of the users with respect to the frequency scores of each dataset for case 2>

Case Study


calls
guard
coast
identify
distress
caller
hoax
investigate
(C)
(A)
TwitterStream
704
Union
Intersection
Reset
Show all
Accident
275
False Call
6
Illegal Immigration
73
Radio/Channel
45
SAR
46
(B)
User ID
Creation Date
Tweets Content
14-07-22 10:31:29 EST
Coast Guard looking for hoax caller: The U.S. Coast Guard is looking for a hoax caller who made more than a dozen... http://t.co/8BXtKQsbp6
14-07-23 08:28:59 EST
Coast Guard looking to identify hoax caller: The U.S. Coast Guard is looking to identify a Southern Maryland... http://t.co/CW0TtihlrS
14-07-22 11:49:55 EST
Coast Guard Investigates Hoax Distress Calls On The Bay: BALTIMORE (WJZ)?The Coast Guard wants to know who is... http://t.co/0I2XMBDSow

Case Study


Time Monitor: 2014-1-1 0:5:45 --- 2014-8-28 0:5:56
... @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
@WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
Sorting based on: Volume
Accident 122
False Call 13
Illegal Immigration 6
Radio/Cha 14
SAR 21
CG GroupB 21
CG GroupC 0
userID 2199938589
image
name ClaytonW_17
friends 133
followers 66
location null
User ID Creation Date Tweets Content
2199938589 14-07-22 22:33:05 EST @WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived

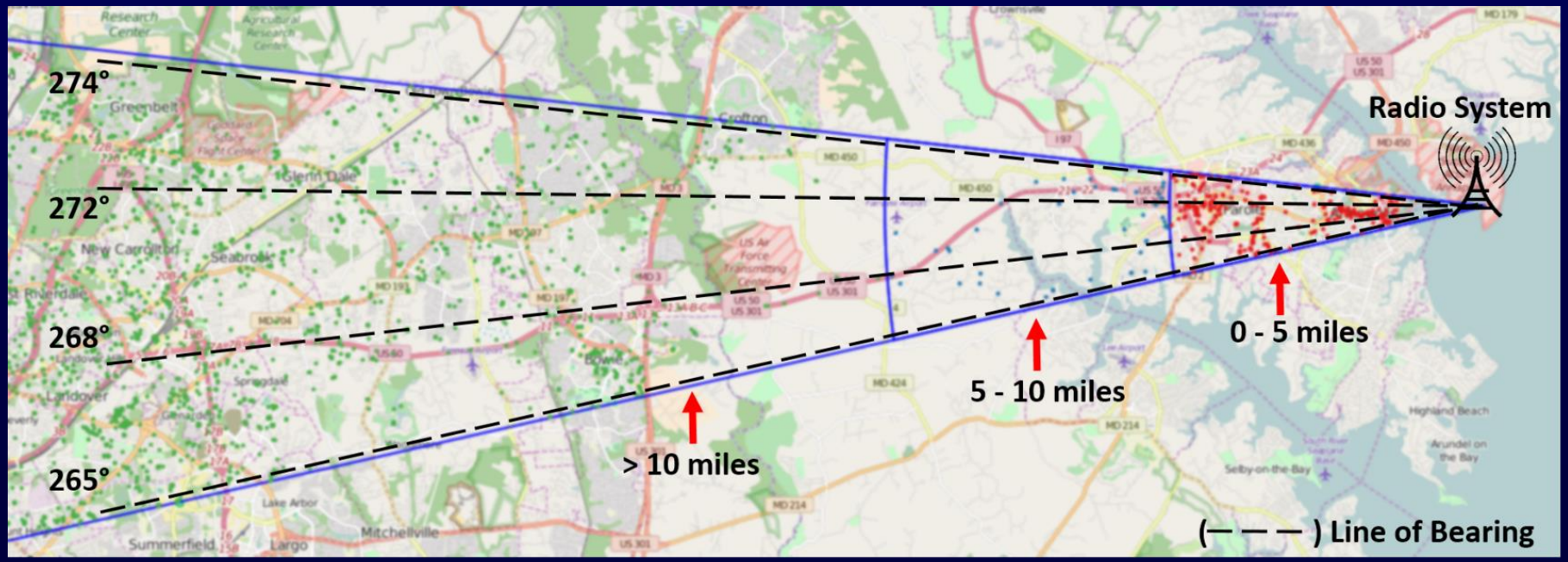
Conclusion

- A visual social media analytics solution driven by the Rescue 21 system for supporting the investigation of hoax distress calls
- Identifying suspicious Twitter and Instagram messages based on the spatial and temporal clues from the system

Acknowledgement

- U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0003

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Thank you!

Junghoon Chae
jchae@purdue.edu

VACCINE

VACCINE

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

Evolving Foundations and Understanding of Visual Analytics Theory

The Past, Present And Future

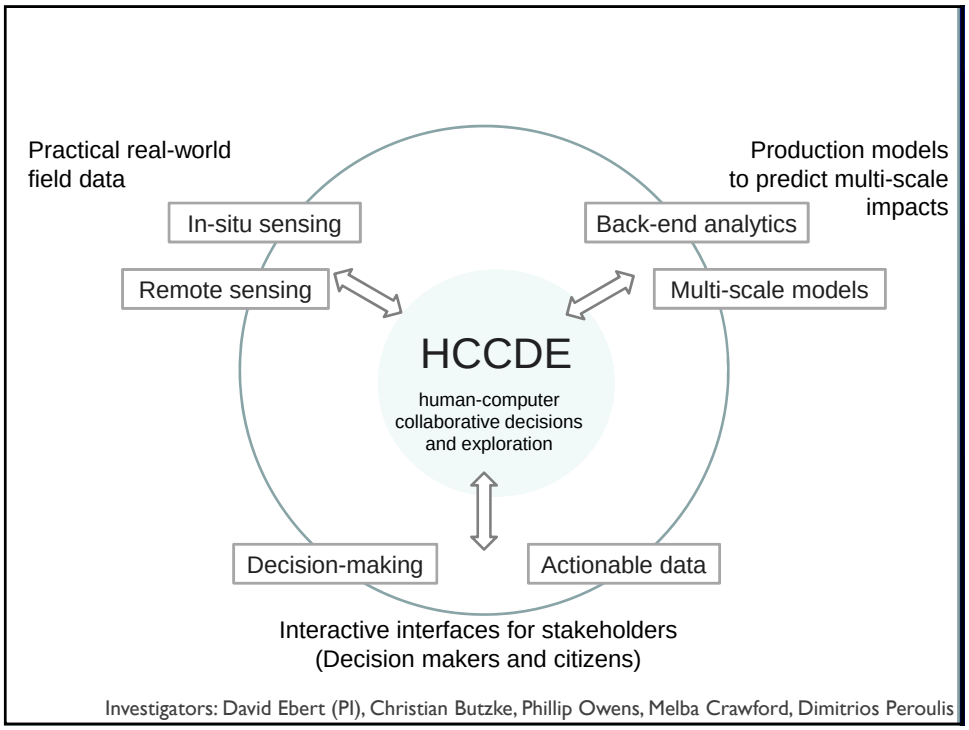


David S. Ebert
Purdue University
ebertd@purdue.edu

April 2016

A Guided Journey

- A glimpse ahead
- A review of the evolution
- The foundations of theory
- Theory evolution and future



Original Working Definition of VA (Thomas et al 2004)

- As a starting point, we offer the following definition:
  - Visual Analytics is the formation of abstract visual metaphors in combination with a human information discourse (interaction) that enables *detection of the expected and discovery of the unexpected* within massive dynamically changing information spaces.[1] The goal of Visual Analytics is to stimulate deep analytical insight from massive, complex, and often conflicting information in order to yield an assessment having quantified certainty.

VACCINE

April 2016

My Position Statement in July 2004 to Attend the Workshops to Define VA

- The main principle of our work is that visualization can be a key component of analyzing and understanding data only if it is interactive, relevant, easy to use, and integrated into a complete analysis system. My expertise that will be of value to this panel can be categorized into three areas:

My 2005 Goals

- Effectively communicate information
- Ease cognitive burden of user
- Create effective external cognitive artifacts
- Semi-automatically adapt to task at hand
- Scale across problem size and display size
- Enable insight and analysis

Background: Visual Analytics Process (adapted from Bertin)

- Define the problem / question
- Determine the data: process / extract the relevant data
  - Abstraction
- Determine
  - Characteristics of the relevant data
  - Types of data: nominal, ordinal, interval, ratio
  - Quality of the data
  - Size, dimensionality & number of data items per sample
- Map the data characteristics to an appropriate visual representation for the task at hand
- Results: Enable rapid, actionable, defensible, reasoned decisions and action

VACCINE

April 2016

Effective Visual Representations for Real-World Problems

- Understand the problem or question - task
- Determine data needed for solution and its characteristics
- Determine effective visual representation
  - Utilize design, cognitive perception, illustration, and advanced rendering techniques



VACCINE

April 2016

Our Guiding Principles for Effective Visual Representations (Norman, Tversky)

- Appropriateness principle:
  - Visual representation should provide neither more nor less information than that needed for the task at hand
- Naturalness principle:
  - Experiential cognition most effective when representation most closely matches the information being represented
  - New visual metaphor's must match the user's cognitive model of the information

Our Guiding Principles for Effective Visual Representations (Norman, Tversky)

- Matching principle
  - Representations much match the task to be performed
  - Visual representations should present affordances suggestive of appropriate actions
- Principle of apprehension:
  - The content of the representation should be accurately and easily perceived

Creating Visual Representations Using These Principles

- Illustrative visualization and analytics
- Key components
  - Appropriate visual representation selected based on data characteristics
  - Design principles for layout and composition
  - Illustration-based techniques for compact representations of massive data – **enabling abstraction**
  - Advanced 3D perceptual cues for aiding cognition

Must Address: Accuracy, Reliability and Reproducibility

- Accuracy – avoid miscommunication of information
- Reliability – dependable for decision making?
- Reproducibility – consistent from dataset to dataset?

Must Address: Interactivity and Usability

- Interactivity
  - Inherent understanding from direct manipulation
- Usability
  - Appropriate interface level
  - Understandable controls
  - Understandable data processing

VACCINE

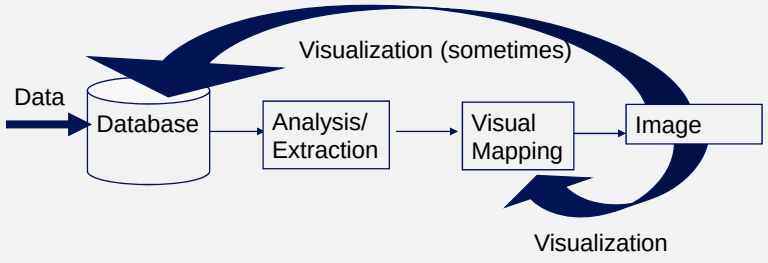
April 2016

My View of VA in 2008

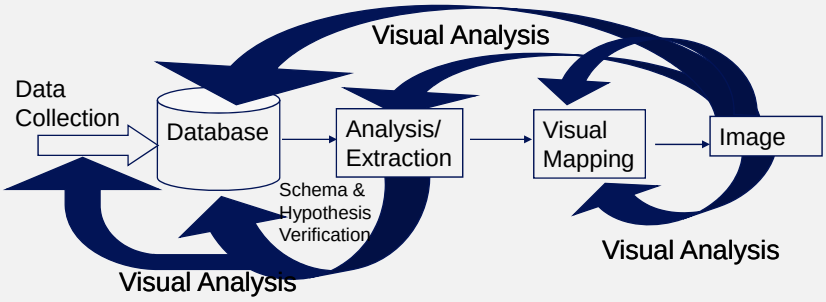
April 2016

Problem

- Every field is faced with a data deluge
  - More data isn't necessarily better
- Traditionally, graphics and visualization are just post-processes



Visual Analysis Solution

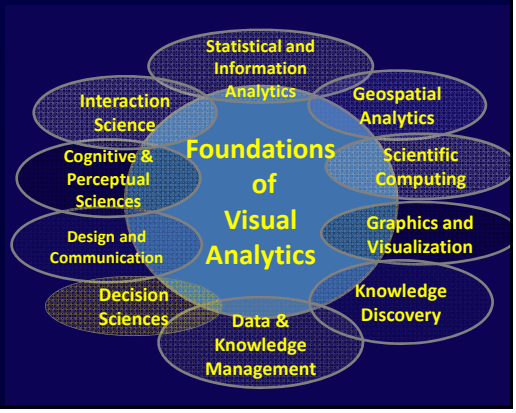


What We Do:Visual Analytics



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

Essential Features for Visual Analytics

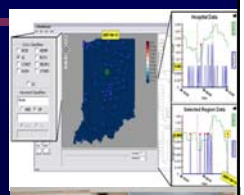
- Interactive, integrated discovery environment
- User-guided and perceptually-guided
- Balance of human cognition and automated computerized analysis to amplify human cognition
- Enable discovery and decision making
- Provide quantitative, reliable, reproducible evidence

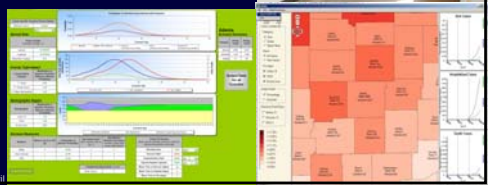
VACCINE

April 2016

Example Real World User Driven Visual Analytic Solutions

- Syndromic Surveillance
- Infield Emergency Response
- Pandemic Preparedness





VACCINE

April

The Depth of the Research Challenges

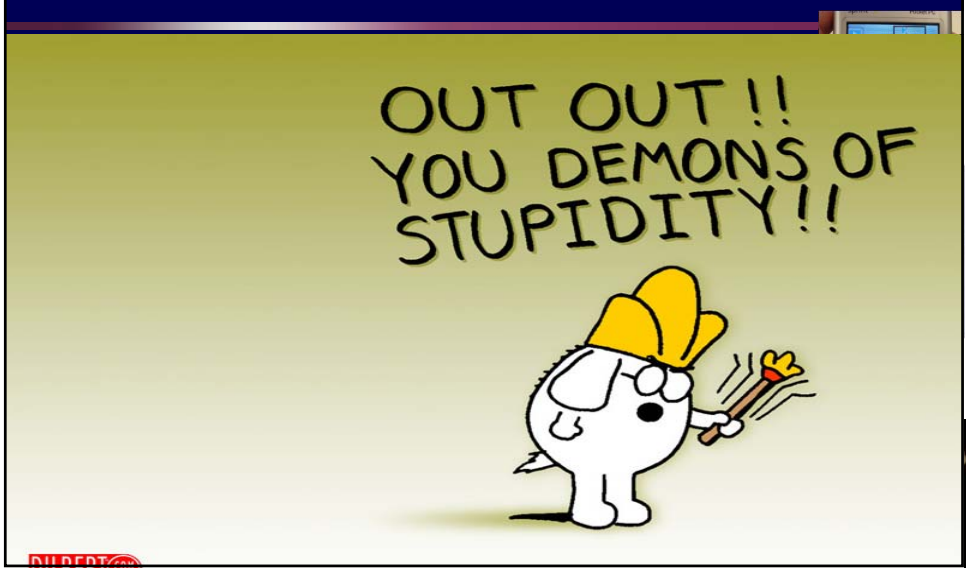
- Long-term fundamental challenges!
  - Mobility; scale; multisource, cross-media analysis; numerical to visual analysis
- How do we effectively integrate and adapt solutions to the
  - User
  - Task
  - Environment
  - Interactive discovery environment
  - Data



VACCINE

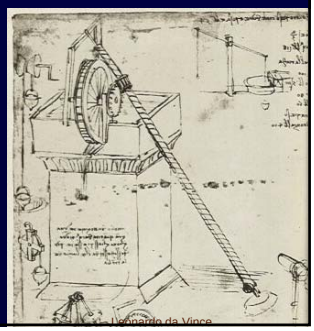
April 2016

The User



The User

- Decision making methods – customize to their approach
  - What external aids and representations are they comfortable with?
  - What is the best balance of interaction guidance and computerized analysis
  - How do they track their thought process?

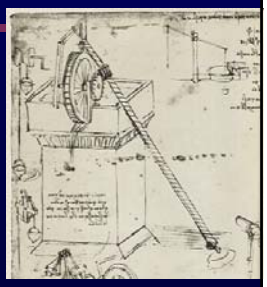


VACCINE

April 2016

The Task

- Adapt visual analysis environment and representation based on the task for more effective decision making
- Great principles and previous work
  - Cognitive task analysis
  - Design
  - Industrial design





Leonardo da Vince

VACCINE

April 2016

The Environment

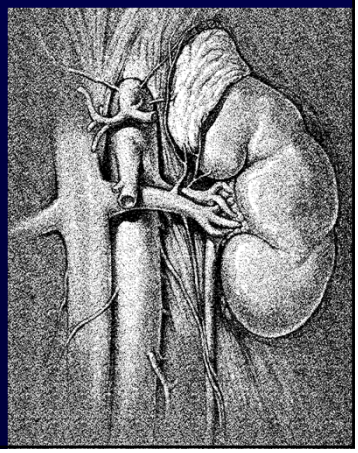
- Physical environment
  - Hands busy, eyes busy
  - Physical gear
  - Visibility, temperature, lighting
- Mobility
- Temporal environment
  - Immediate knowledge transfer for critical decisions?
  - Extended use
  - Occasional vs. routine use
- Psychological environment
  - Leisurely exploration
  - Stress level
  - Ability for attentive focus, concentration





The Interactive Discovery Environment

- Goal: Creating information discourse integrating semi-automated data analysis and visual analysis
- Critical components:
  - Perception, cognition, art, and design
  - Interaction design
  - Graphical/visual layout
- Design for effective:
  - Communication
  - Information transfer
  - Knowledge discovery
  - Decision making



VACCINE

April 2016

Joe Kielman's View in 2009

April 2016

Big/Deep History – What We Mean

Decision-making and judgment formation:

Automatic, image-based systems

Fast
Computationally frugal
Reliable

Pragmatic, embodied heuristics

Modal basis for knowledge

Analog = image
Digital = amodal symbol

Successive paradigms tell us different things about the population of the universe and about that population's behavior

Thomas Kuhn – **The Structure of Scientific Revolutions**

Homeland Security

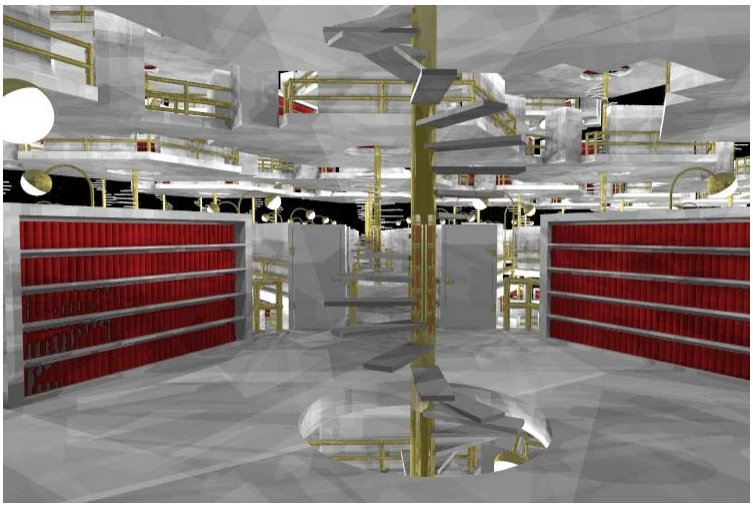
Precise Data

Necessary and Sufficient Conditions for Precision Information Environments:

- Synthetic world and real world co-exist, are co-equal
- Physical and information spaces are co-existent
- Modeling and simulation capabilities apply to both
- Streaming, dynamic data are available, accessible in real-world time
- Environment is tailored, suited for multiple user groups on a dynamic basis
- Information spaces are multimodal
- Multiple users and manifold uses are managed simultaneously

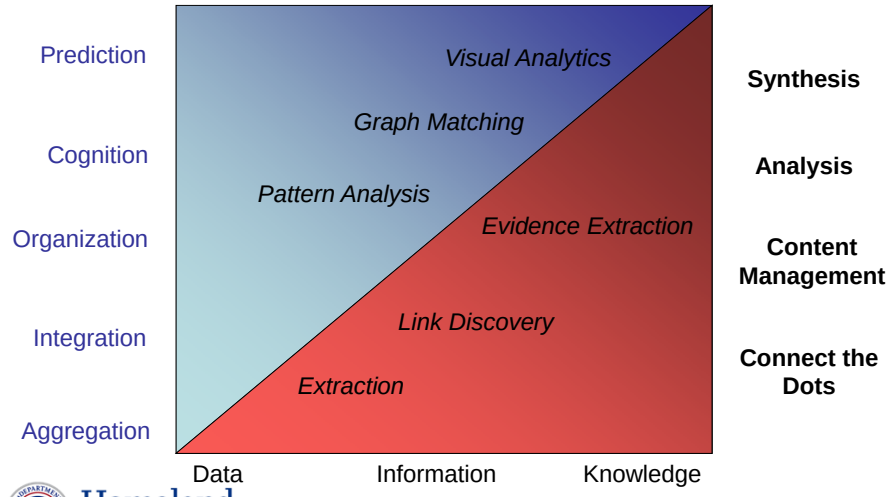
Homeland Security

The Library of Babel



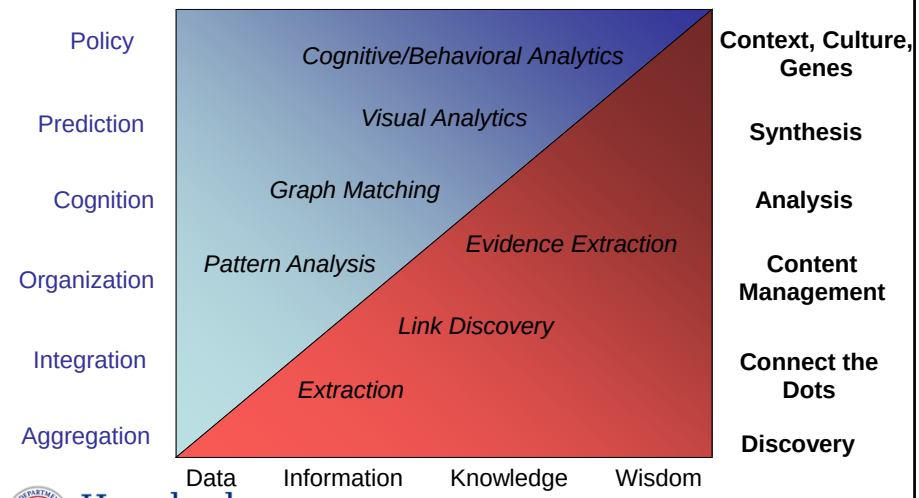
Homeland Security

Analytics



Homeland Security

Predictive, Prospective



Homeland Security

The Future of Analytics

a
World-wide Enterprise
in
Science and Technology
for
Cognitive Analytics (Philosophical Analytics, Cognitive Philosophy),
Which provides
Precisely Tuned
Knowledge
for
Assuring
the
[Quality of] Existence
of
Individuals, Communities, Societies, Cultures, Countries,
and their
Physical World

Homeland Security

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

What I Talk About Today As a VA Researcher:

Human-Computer Collaborative Decision Making

David S. Ebert
Purdue University
ebertd@purdue.edu

April 2016

Why VA and HCCD For Big Data?

(Or What Big Data Analytics Can't Do) [1]

- 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

VACCINE

[1] Inspired and adapted from David Brooks, New York Times, 2/18/2013)

April 2016

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

Approach: Partner-driven solutions and research

VACCINE

April 2016

Learning From Others

Data Science

April 2016

Data Science Is Also Finding Its Place…

ASA Statement on *The Role of Statistics in Data Science*

August 8, 2015

The rise of *data science*, including *big data* and *data analytics,* has recently attracted enormous attention in the popular press for its spectacular contributions in a wide range of scholarly disciplines and commercial endeavors. These successes are largely the fruit of the innovative and entrepreneurial spirit that characterize this burgeoning field. Nonetheless, its interdisciplinary nature means that a substantial collaborative effort is needed for it to realize its full potential for productivity and innovation. While there is not yet a consensus on what precisely constitutes data science, three professional communities, all within computer science and/or statistics, are emerging as foundational to data science: (i) *Database Management* enables transformation, conglomeration, and organization of data resources; (ii) *Statistics and Machine Learning* convert data into knowledge; and (iii) *Distributed and Parallel Systems* provide the computational infrastructure to carry out data analysis.

Data Science Is Also Finding Its Place…

Certainly, data science intersects with numerous other disciplines and areas of research. Indeed it is difficult to think of an area of science, industry, commerce, or government that is not in some way involved in the data revolution. But it is databases, statistics, and distributed systems that provide the core pipeline. At its most fundamental level, we view data science as a mutually beneficial collaboration among these three professional communities, complemented with significant interactions with numerous related disciplines. For data science to fully realize its potential requires maximum and multifaceted collaboration among these groups.

Statistics and machine learning play a central role in data science. Framing questions statistically allows us to leverage data resources to extract knowledge and obtain better answers. The central dogma of statistical inference, that there is a component of randomness in data, enables researchers to formulate questions in terms of underlying processes and to quantify uncertainty in their answers. A statistical framework allows researchers to distinguish between causation and correlation and thus to identify interventions that will cause changes in outcomes. It also allows them to establish methods for prediction and estimation, to quantify their degree of certainty, and to do all of this using algorithms that exhibit predictable and reproducible behavior. In this way, statistical methods aim to focus attention on findings that can be reproduced by other researchers with different data resources. Simply put, statistical methods allow researchers to accumulate knowledge.

How Do We Deve

- Find the solution onli

VACCINE

Part 1 **Conceiving a Theory**

1 **Wonder "why?"** Look for patterns between seemingly unrelated things. Explore the root causes behind everyday events, and try to predict what will happen next. If you already have the seed of a theory in your head, observe the subjects of that idea and try to gather as much information as possible. Write down the "hows," the "whys," and the links between things as you piece them together.[1]

- If you don't have a theory or a hypothesis in mind, you can begin by making connections. If you walk through the world with a curious eye, you may be suddenly struck by an idea.

2 **Develop a theory to explain a law.** In general, a scientific law is the description of an observed phenomenon. It doesn't explain *why* the phenomenon exists or what causes it. The explanation of the phenomenon is called a scientific theory. It is a common misconception that theories turn into laws with enough research.[2]

- For instance: Newton's Law of Gravity was the first to mathematically describe how two different bodies in the universe interact with each other. However, Newton's law doesn't explain why there is gravity, or how gravity works. It wasn't until three centuries after Newton, when Albert Einstein developed his Theory of Relativity, that scientists began to understand how and why gravity works.

3 **Research the academic precedents to your theory.** Learn what has already been tested, proven, and refuted. Find out everything that you can about your subject, and determine whether anyone has asked the same questions before. Learn from the past so that you don't make the same mistakes.[3]

- Use existing knowledge to better understand your subject. This includes equations, observations, and existing theories. If you are addressing a new phenomenon, try to build upon related theories that have already been proven.
- Find out whether anyone has already developed your theory. Before you go any further, try to make reasonably sure that no one else has already explored this topic. If you can't find anything, feel free to develop your theory. If someone has already made a similar theory, read through their work and see if you can build on it.

4 **Build a hypothesis.** A hypothesis is an educated guess or proposition that aims to explain a set of facts or natural phenomena. Propose a possible reality that follows logically from your observations – look for patterns, and think about what might cause those things to happen. Use an "if, then" form: "*If* [X] is true, *then* [Y] is true," or "*If* [X] is true, *then* [Y] is untrue." Formal hypotheses contain an "independent" and a "dependent" variable. The independent variable is a potential cause that you can tweak and control, while the dependent variable is a phenomenon that you observe or measure.

- If you are going to use the scientific method to develop your theory, then your hypothesis must be measurable. You cannot prove a theory without numbers to back it up.
- Try to come up with several hypotheses that might explain your observations. Compare these hypotheses. Consider where they overlap and where they split.
- Example hypotheses: "*If* skin cancer is related to ultraviolet light, *then* people with a high exposure to UV light will have a higher frequency of skin cancer." or "*If* leaf color change is related to temperature, *then* exposing plants to low temperatures will result in changes in leaf color."[4]

5 **Know that every theory starts as a hypothesis.** Be careful not to confuse the two. A theory is a well-tested explanation for why a pattern exists, while a hypothesis is only a predicted reason for this pattern. A theory is always backed by evidence. A hypothesis, however, is only a suggested possible outcome, and it may or may not hold true.[5]

April 2016

How Do We Deve

- Find the solution on

VACCINE

Part 2 **Testing Hypotheses**

1 **Design an experiment.** According to the scientific method, your theory must be testable.[6] Develop a way to test whether each hypothesis holds true. Be sure to perform your test in a controlled environment: try to isolate the event and your proposed cause (the dependent and independent variable) from anything that might complicate the results. Be precise, and look out for external factors.

- Make sure that your experiments are repeatable. In most cases, it is not enough to simply prove a hypothesis once. Your peers should be able to recreate your experiment themselves and get the same results.
- Have peers or advisers review your testing procedure. Ask someone to look over your work and confirm that your logic is sound. If you are working with partners, make sure that everyone gives their input.

2 **Find support.** Depending on your field of study, it may be hard to run complex experiments without access to certain equipment and resources. Scientific gear can be expensive and tricky to procure. If you are enrolled in a university, speak with any professors and researchers who might be able to help.

- If you aren't in school, consider reaching out to professors or graduate students at a local university. For instance, contact the physics department if you want to explore a theory of physics. If you find a far-away university that is doing a lot of exciting research in your field, consider emailing them despite the distance.

3 **Keep precise records.** Again, experiments must be reproducible: other people must be able to set up a test in the same way that you did and get the same result. Keep accurate records of everything you do in your test. Be sure to keep all your data.

- If you're in academia, there are archives which store the raw data gathered in the process of scientific research. If other scientists need to find out about your experiment, they can consult these archives or ask you for your data. Make sure that you can provide all the details.

4 **Evaluate the results.** Compare your predictions against each other and against the outcomes of your experiments. Look for patterns. Ask yourself whether the results suggest anything new, and consider whether there's anything that you've forgotten. Whether or not the data confirms the hypothesis, look out for hidden or "exogenous" variables that may have influenced the results.

5 **Establish certainty.** If the results do not support your hypothesis, reject the prediction as incorrect. If you are able to prove the hypothesis, then the theory is one step closer to being confirmed. Always document your results with as much detail as possible. If a test procedure and its results cannot be reproduced, it will be much less useful.

- Make sure that the results do not change each time you do the experiment. Repeat the tests until you're sure.
- Many theories get abandoned after being refuted by experiment. However, if your new theory explains something that previous theories can't, it may be an important scientific advance.

April 2016

How Do We Develop a Theory?

- Find the solution onli

Part 3 **Accepting and Expanding a Theory**

1 **Draw a conclusion.** Determine whether your theory is valid, and make sure that your experimental results are repeatable. If you accept the theory, you should not be able to disprove it with the tools and information at your disposal. Do not, however, try to spin your theory into absolute fact. [7]

2 **Share your results.** You will likely amass a lot of information in your quest to prove your theory. When you are confident that your results are repeatable and your conclusions are valid, try to distill your theory into a form that others can study and understand. Lay out your process in a logical order: first, write an "abstract" that summarizes your theory; then, lay forth your hypothesis, your experimental procedure, and your results. Try to distill your theory into a series of points or arguments. Finally, end the paper with an explanation of your conclusions.

- Explain how you defined your question, the approach you took, and how you tested it. A proper report will walk the reader through every relevant thought and action that brought you to your conclusion.
- Consider your audience. If you want to share your theory with peers in your field, write an formal paper explaining your results. Consider submitting your work to an academic journal. If you want to make your findings accessible to the general public, try distilling your theory into something more digestible: a book, an article, or a video.

3 **Understand the peer-review process.** In the scientific community, theories are not generally considered valid until they have been peer-reviewed. If you submit your findings to an academic journal, another scientist may decide to peer-review—that is to say, test, consider, and replicate—the theory and process that you have put forward. This will either confirm the theory or leave it in limbo. If the theory survives the test of time, others may eventually try to expand your idea by applying it to other subjects.

4 **Build upon your theory.** Your thought process does not need to end after you share your theory. Indeed, you may find that the act of writing up your ideas forces you to consider factors that you've been ignoring. Don't be afraid to keep testing and revising your theory until you're completely satisfied. This may mean more research, more experiments, and more papers. If your theory is large enough in scope, you may not ever be able to flesh out the implications in their entirety.

- Don't be afraid to collaborate. It can be tempting to keep your intellectual sovereignty, but you may find that your ideas take on new life when you share them with peers, friends, and advisers.

VACCINE

April 2016

So, How Do We Do It?

- Can we adopt/adapt/coalesce theories from our associated disciplines?
  - Theory of Visuali[illegible]on
  - Theory of Huma[illegible]mputer Interaction
  - Theory of Design
  - Theory of Cognitive Science

X

VACCINE

April 2016

What Can We Build Upon?

- Value of Visualization (van Wijke)
- InfoVis Theory work (Liu and Stasko 2010)
  - "In this paper we explore the perceived absence and greater need for theories in information visualization through the perspectives of what, why and how. We discuss five possible forms of theory: law, model, framework, taxonomy and interpretation (the "what"), the reasons and potential benefits of theories (the "why") and possible ways to generate theories (the "how")."
- Infovis theory work (Purchase et al.):
  - Predictive data-centered theory, information theory, Exploration and Transform/Design models, ….

What Can We Build Upon?

- Laws/principles: Fitts, Gestalt, Tobler, Norman …
- Models: Infovis reference model (Card et al.), Mackinlay & Wilkinson's models of visualization, sense-making model, cogsci modell(Fisher & Green)…
- Frameworks: Lam's infovis framework/taxonomy, evaluation frameworks, …
- Taxonomies: financial visualization taxonomy, ….

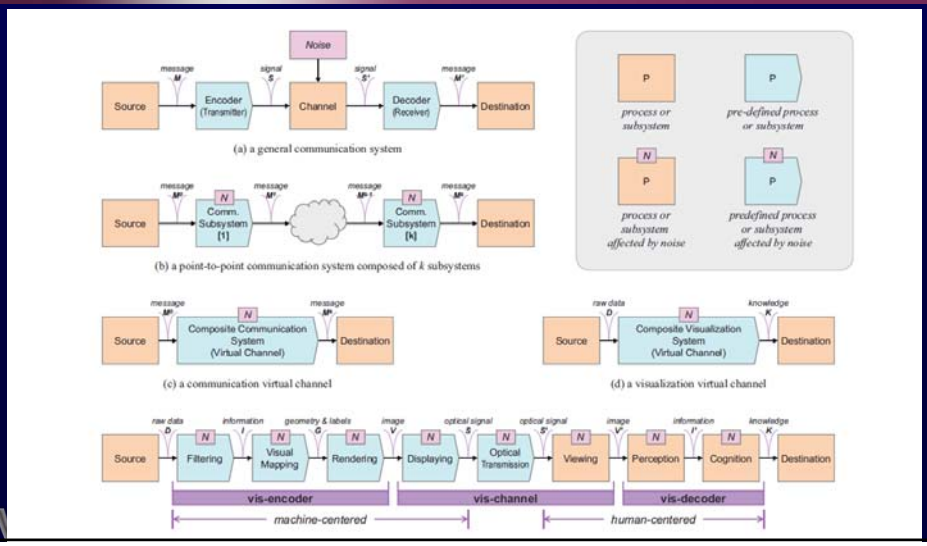
What Can We Build Upon?

- Translational Science:
  - As Fisher points out, “The key innovation in translational research was the emphasis on research methods whose primary goal is the transfer of knowledge between clinical practice and laboratory research and their associated methodologies (Stokes 1997).”
- Information Theory

VACCINE

April 2016

An Information Theoretic View Of Visualization (Jaenicke and Chen)



My View of Where We Are…

- We are developing and science and theory of VA
- We are developing key pieces needed
  - Necessary and sufficient conditions
  - Laws/principles to adopt/adapt
  - Measures for effectiveness – One of the great value of theories
  - Maps of the unknown
  - Problems to be solved and knowledge gaps
  - Grand challenges

Examples:

- Science of Interaction for Visual Analytics challenges (fluidity, collaboration, visual discourse,…)
- Science of Evaluation for VA (frameworks)

Some Key Observations

- Human Centered approach is key (human, task, data)
- Human-computer collaboration is what we are doing
  - Both have advantages, biases/lack of knowledge
- Evaluation and measures of effectiveness are key – the value of VA
- We have started defining VA by necessary and sufficient conditions
- Fluid Interaction is key

Some Key Observations

- We have started defining sub-science components (interaction, evaluation, cognitive principles for HCCD) and role of parent communities
- Do we need a Visual Analytics Turing Test to measure the success of the field?

Direction Forward

- Need frameworks expanded and highlighted
- Build upon our subdisciplines
- Work at 3 levels as mentioned this morning:
  - 1) Guidelines and Principles
  - 2) Conceptual Frameworks and Models
  - 3) Mathematical/social science laws & theory

VACCINE

April 2016

Design and Plan for the Future

- Where will computing and human collaboration be in 2022?
- Where will Cognitive Science be?



VACCINE

April

23 Technologies in 2022

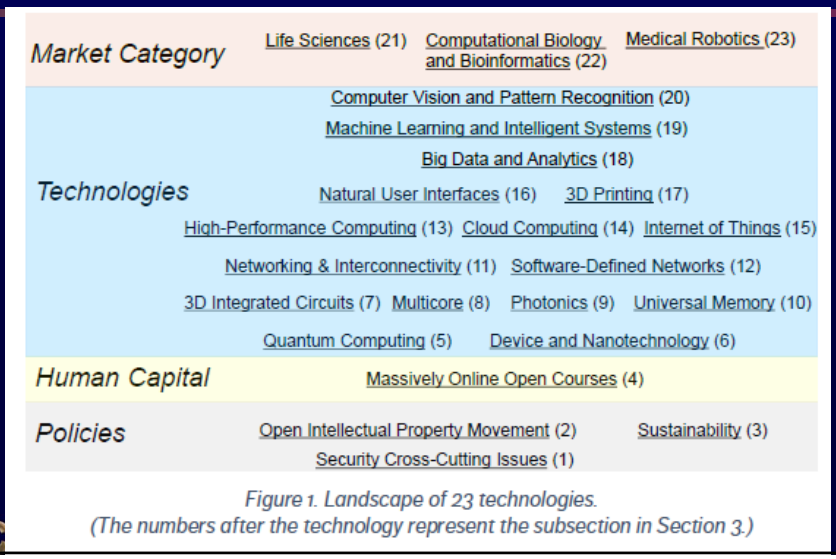


Figure 1. Landscape of 23 technologies.
(The numbers after the technology represent the subsection in Section 3.)

HP's The Machine

- The Machine will fuse memory and storage, flatten complex data hierarchies, bring processing closer to the data, embed security control points throughout the hardware and software stacks, and enable management and assurance of the system at scale
- Will be based around enormous multi-Terabyte universal non-volatile memory that is directly addressable, accessed via fast optical links..
- 320 Terabyte machine delivered in 2016 –all accessible in less than 250nanoseconds
- Memristor coming late in the year

VACCINE

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

Big Data Analytics:
Solution for Data-Driven Decision Making

David S. Ebert
School of ECE
Purdue University
ebertd@purdue.edu

Big Data Challenges

- **Big Data challenges**
- Visual analytics
- VACCINE Center
- Decisions from personal to global scale



"Remember, the other team is counting on Big Data insights based on previous games. So, kick the ball with your other foot."

Big Data



Big Data: Solution to Global Challenges?

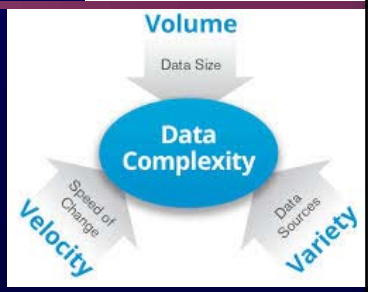


Big Data is often defined along three dimensions:

- **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**





VACCINE

October 2015

These Challenges are Why the U.S. Department of Homeland Security Helped Create the Field of Visual Analytics and Created a Center of Excellence to Solve These Challenges!

VACCINE

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

October 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 Washington
- Arizona State University
- Bethune-Cookman University
- Jackson 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

October 2015

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





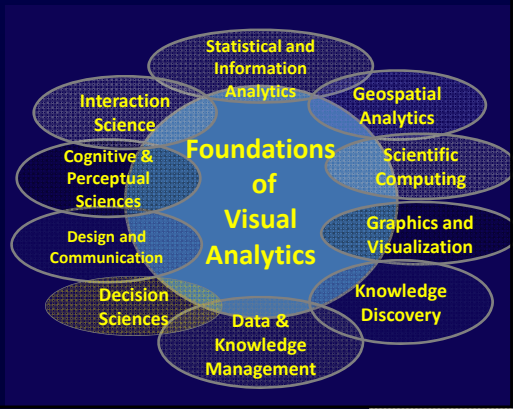
VACCINE

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

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

VACCINE October 2015 PURDUE

What Our Visual Analytic Solutions Offer Decision Makers

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…"

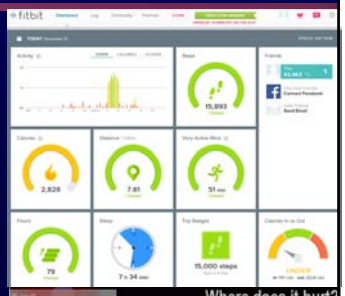
VADM Robert Parker, 2012 MRS Keynote Address

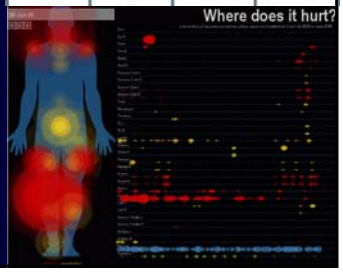


Visual Analytics for Effective Decision Making: Personal to Global

- Visual analytics can help people understand data in their everyday lives







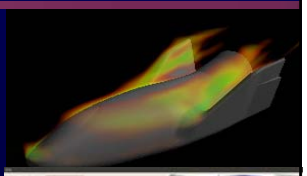
University Scale Decision Making

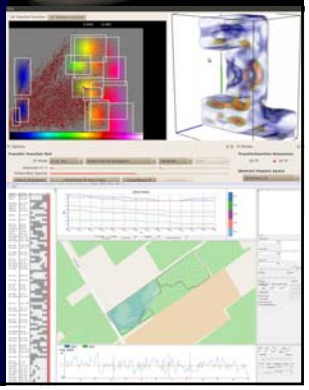
- Enabling scientists and engineers
- Football game force deployment using SMART
- Personal safety across campus
- Student safety
  - Camera locations and safe paths
  - Reducing high-risk alcohol behavior

Scientific Discovery



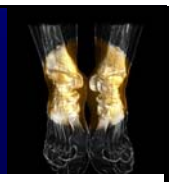
- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality





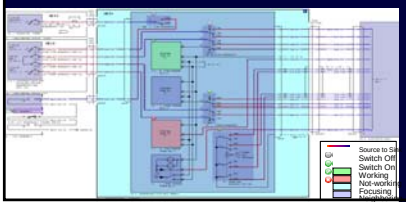
VACCINE

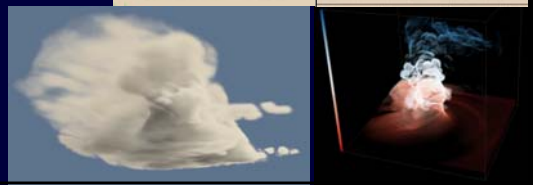
Scientific Discovery



- Illustrative Visualization
  - Mechanical assemblies
  - Schematics and wiring diagrams
  - Illustrative flow visualization
  - Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analytics







VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

SMART: Social Media Analytics and Reporting Tool



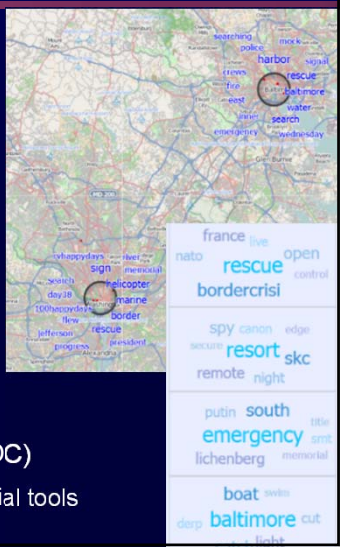
October 2015

Social Media Analytics and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

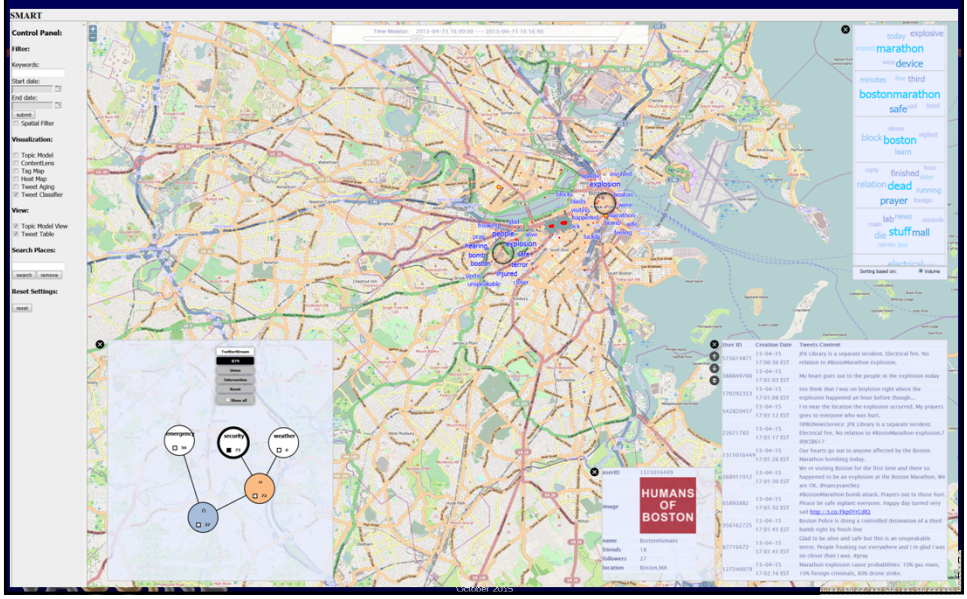
- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games 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

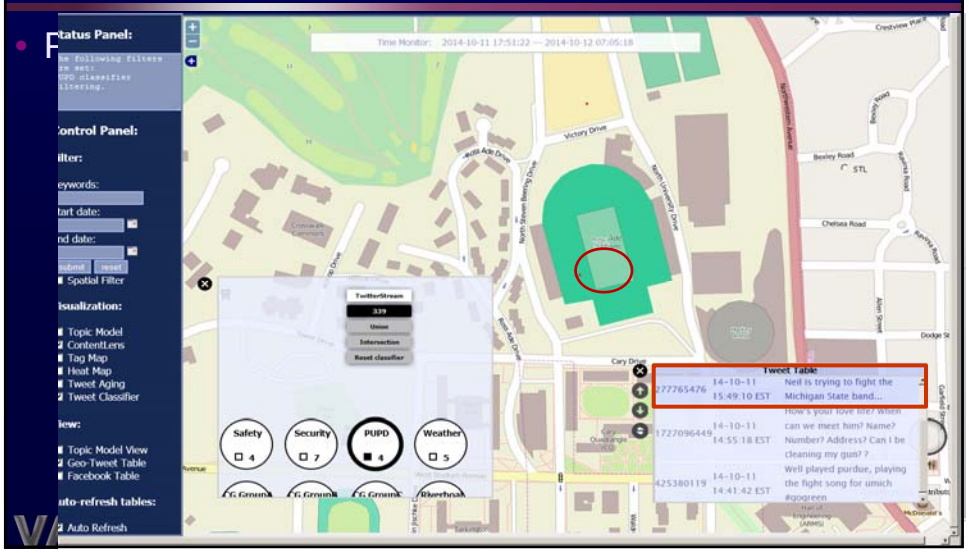
October 2015

Social Media Analytics and Reporting Toolkit (SMART)

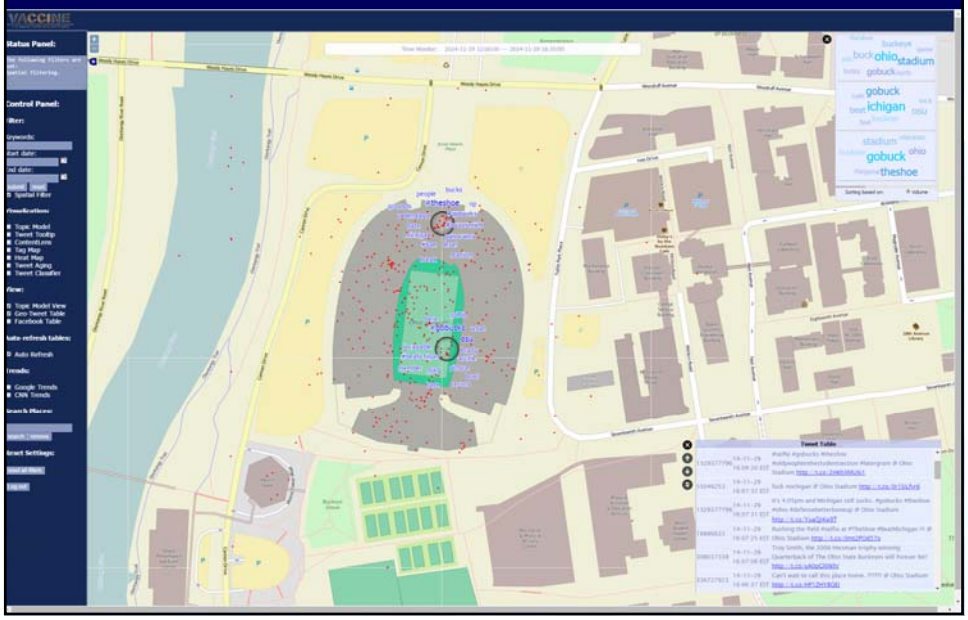


Example From Purdue Football 10/12/14

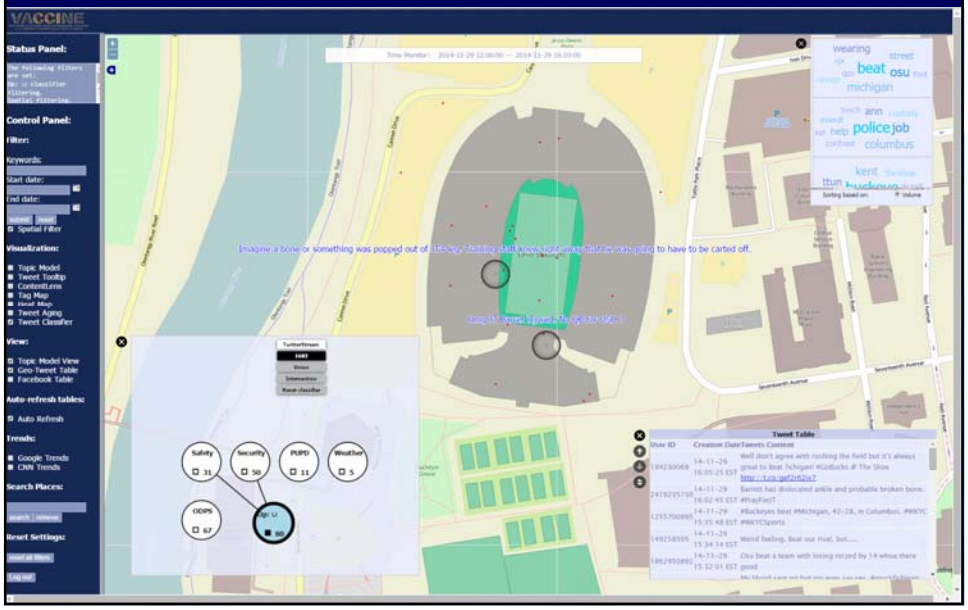
- F



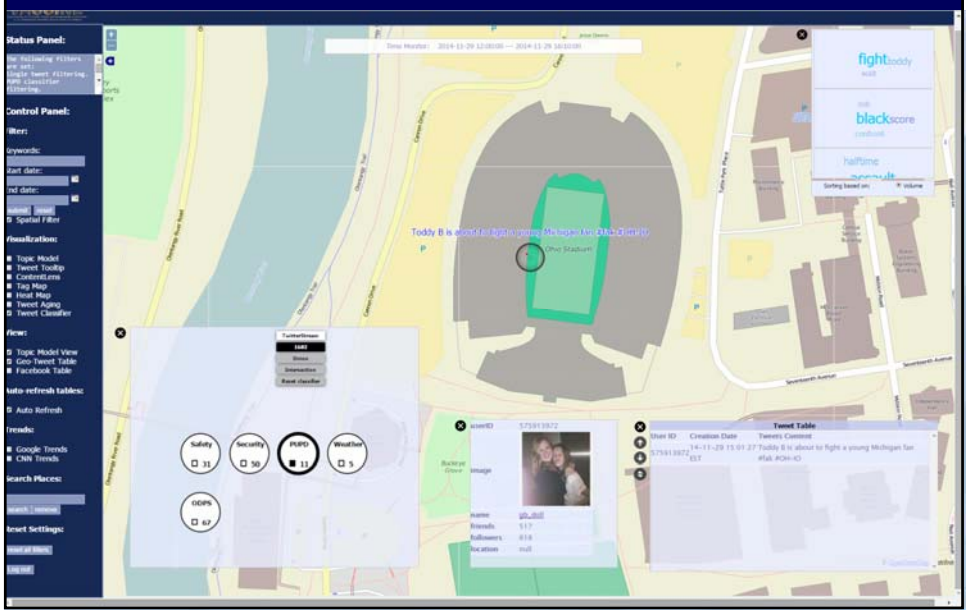
Example From Ohio State Football 11/29/14



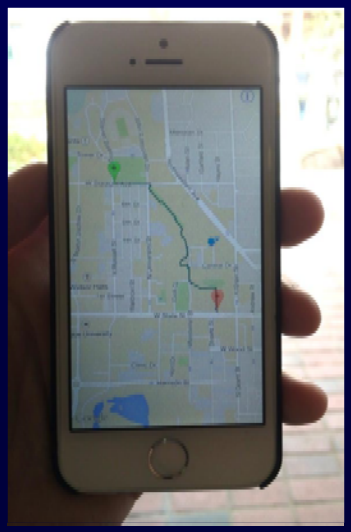
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



Safety in View:
A Public Safety Visual Analytics Tool
Based on CCTV Camera Angles of View



October 2015

Our Technology: Harnessing CCTV Coverage and Crime Data

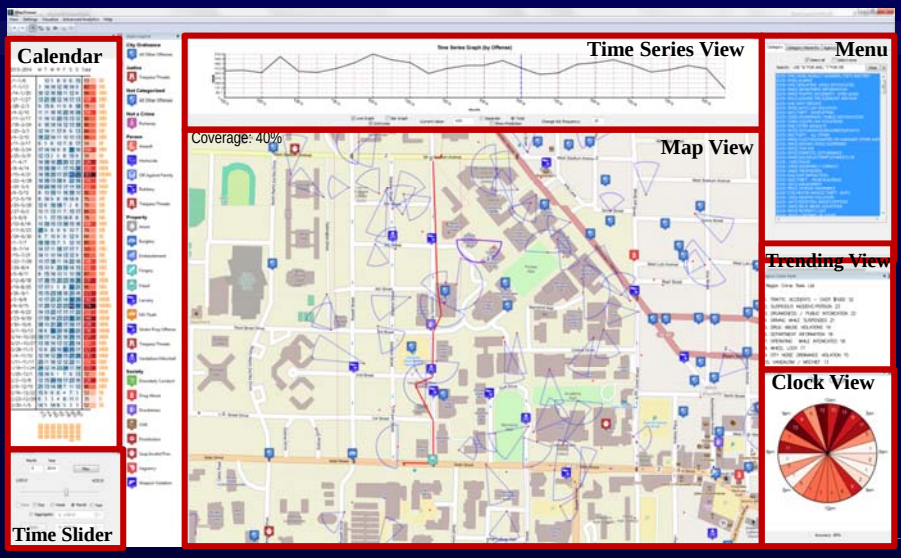
- Law Enforcement: Optimize CCTV coverage using crime, traffic, and civil incidence reports
  - Quickly locate cameras that capture specific incidents
  - Identify critical cameras to monitor
  - Devise data-guided CCTV (re)-deployment strategies
- Pedestrian Safety: Utilize CCTV coverage data to obtain safe walk paths
  - Maximize coverage on walking paths
  - Interactively adjust path solutions as necessary

VACCINE

October 2015

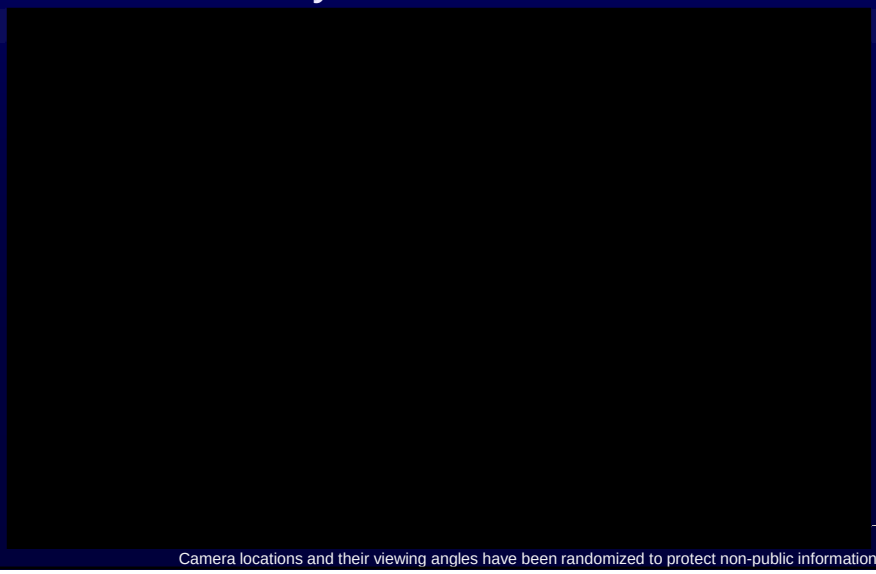
PURDUE

System Overview: Law Enforcement Mode



Camera locations and their viewing angles have been randomized to protect non-public information

Pedestrian Safety: Safe Path Demo

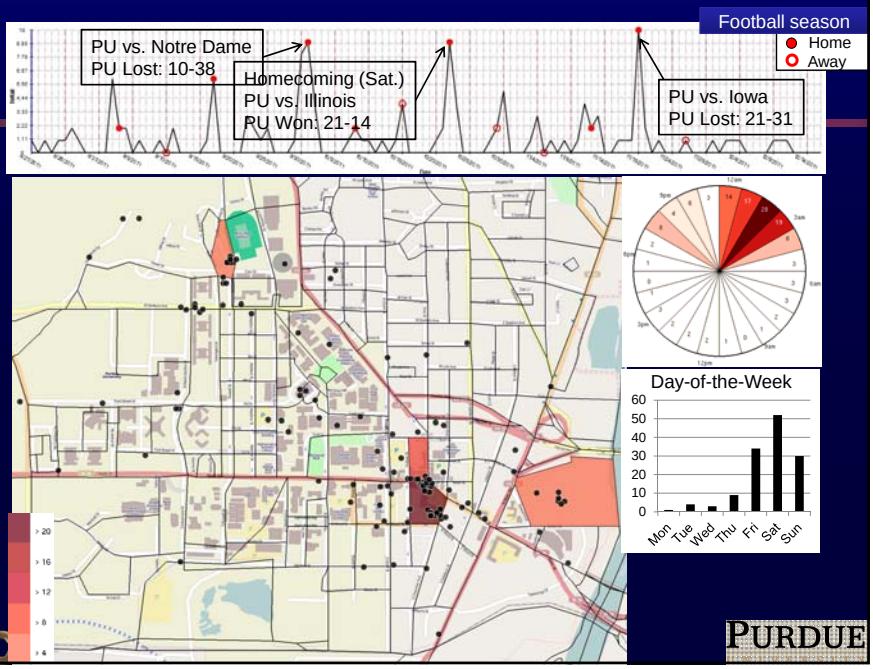


Camera locations and their viewing angles have been randomized to protect non-public information

Reducing High-Risk Alcohol Behavior

- Student Wellness Center Partnership
- Find areas and patterns of high-risk alcohol behavior
- Information campaign to students and dorms
- Increased patrolling

High Risk Behavior : Drunkenness / Public Intoxication



VAC

PURDUE

Community Scale Decision Making

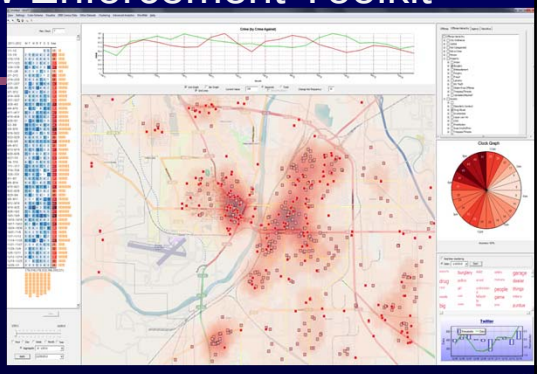
- Public safety – crime analytics (VALET)
- Transit safety – eBOLO

VACCINE

October 2015

PURDUE

Visual Analytics Law Enforcement Toolkit
(VALET, iVALET)



Impacts:

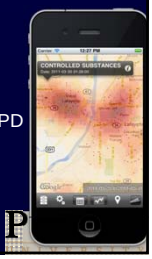
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive 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



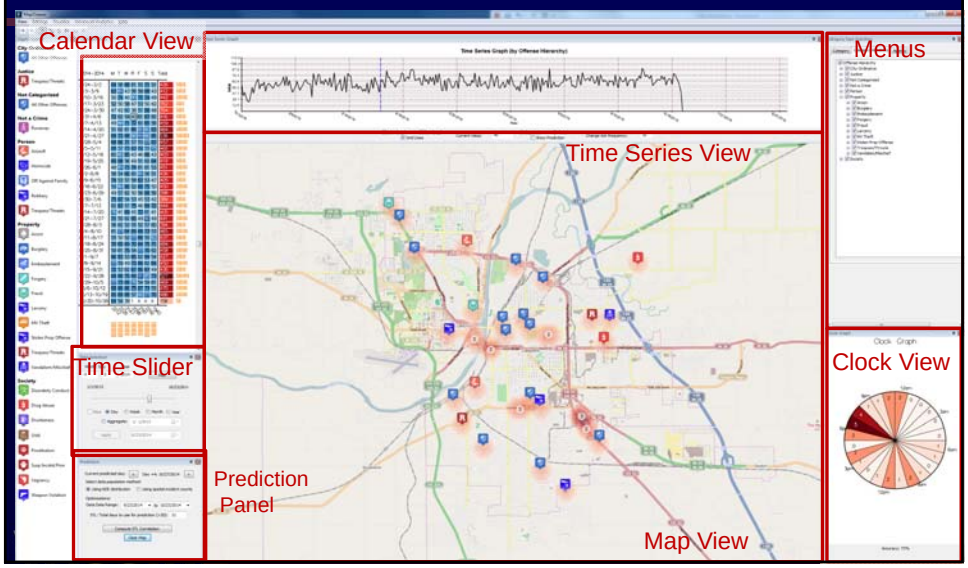
VACCINE

October 2015

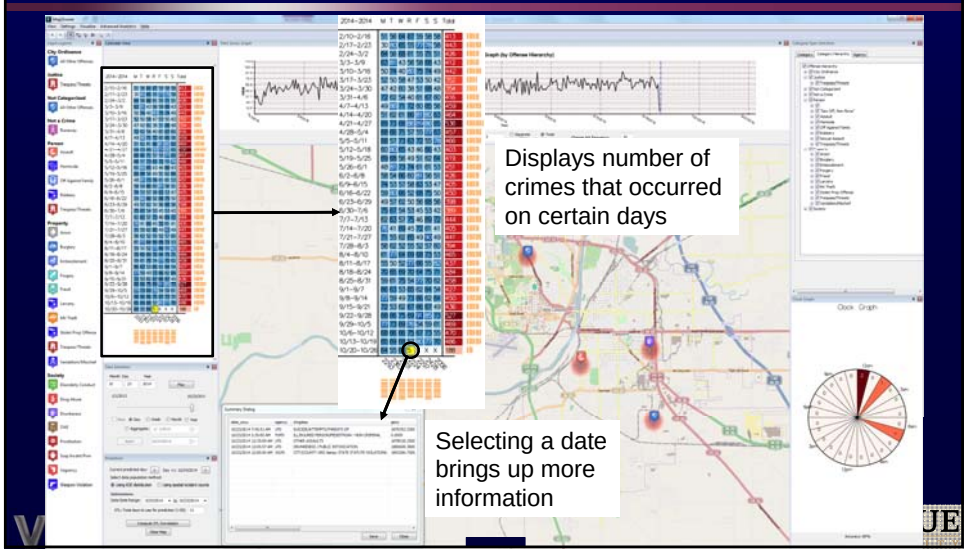
Example Projects: Public Safety and Law Enforcement Applications



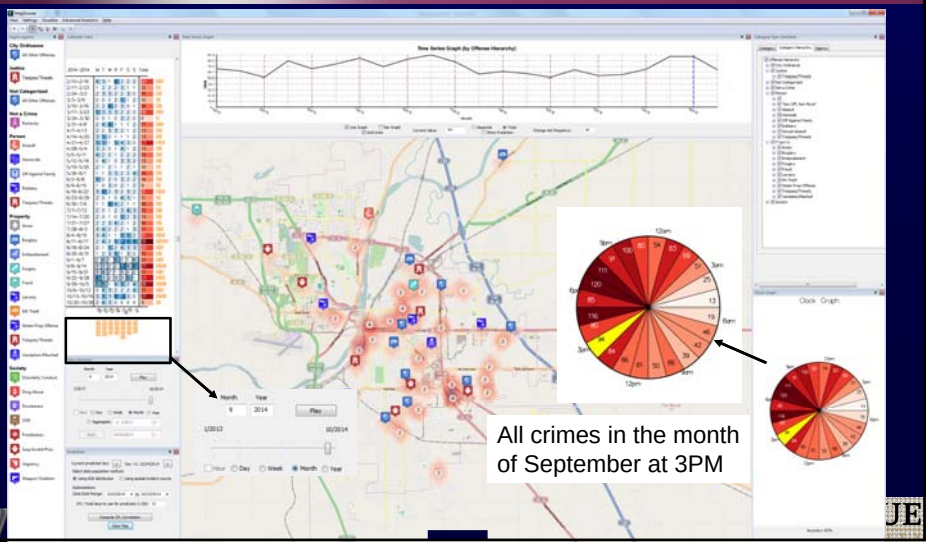
VALET Layout

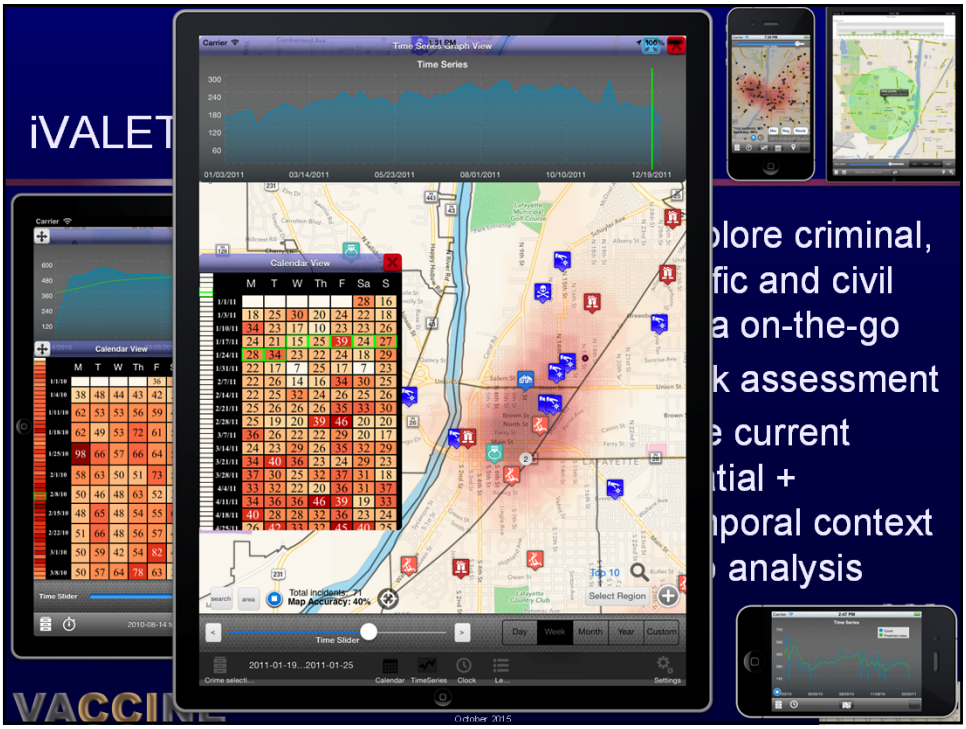


Calendar Display



Crimes During Certain Hours In A Month





eBOLO Goals

- Allow transit system to identify persons of interest from a video and determine if and when they re-appear at a later date
- Use computer vision and image processing techniques to generate the “eBOLO fingerprint” of POI
- 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

VACCINE

October 2015

ALERT
AWARENESS AND LOCALIZATION OF EXPLOSIVES-RELATED THREATS

Preliminary Results: Purdue Group



VACCINE

October 2015

ALERT
AWARENESS AND LOCALIZATION OF EXPLOSIVES-RELATED THREATS

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

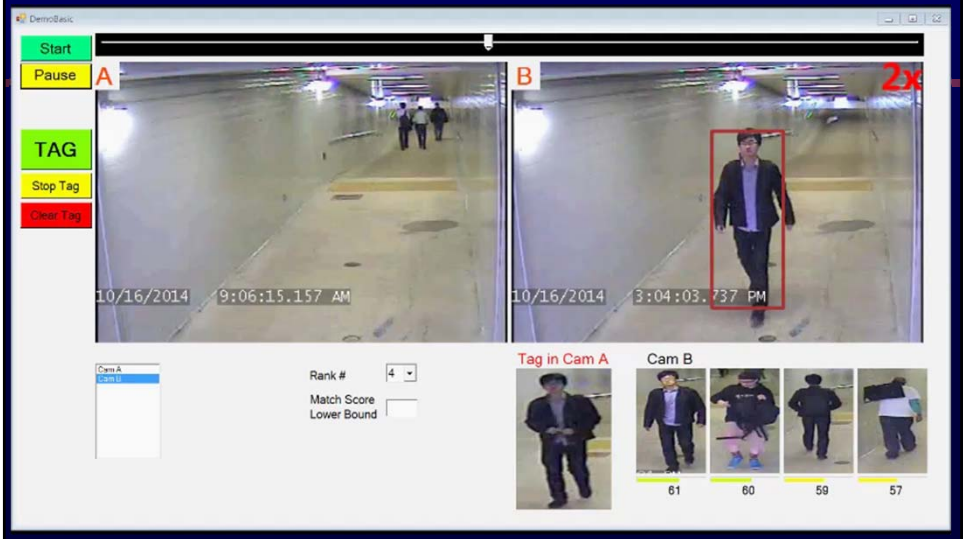
Rank	1	2	3	4	5	6
Matching Accuracy	53%	77%	90%	100%	100%	100%



VACCINE

October 2015

ALERT



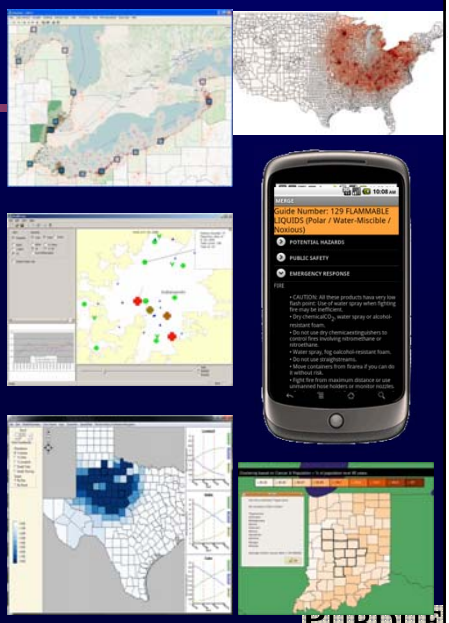
This slide is a playable video

State and Regional Level Decision Making

- Public Health
- Disaster Preparedness and Response
- VASA

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

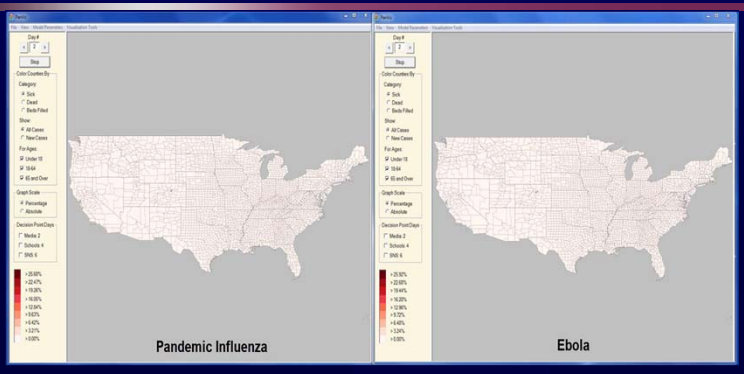


VACCINE

October 2015

PURDUE

Ebola Vs. Pandemic Influenza (H1N1)



Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

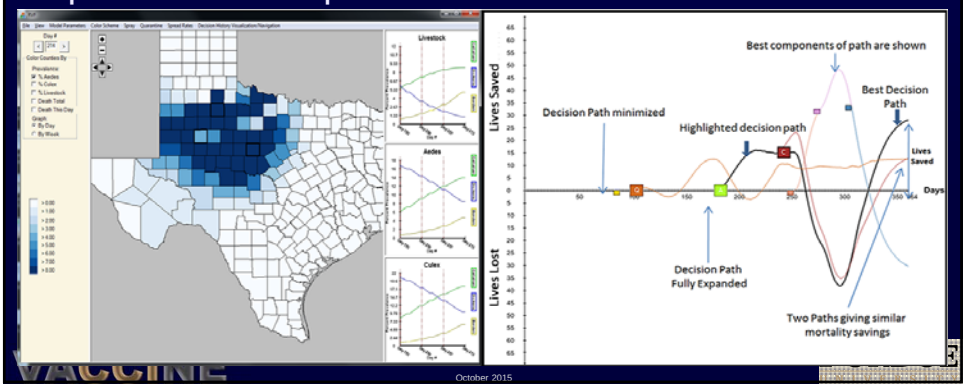
VACCINE

October 2015

PURDUE

Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)

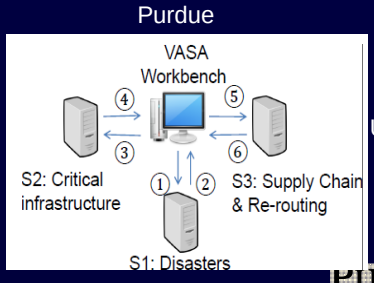
- Integrated simulation models
- Explore alternative courses of action in decision space and real space



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

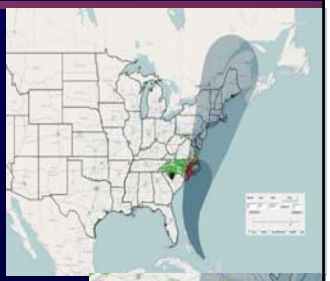


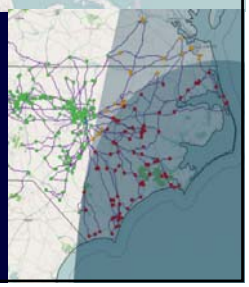
VASA: Visual Analytics for Simulation-Based Action

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 simulation 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





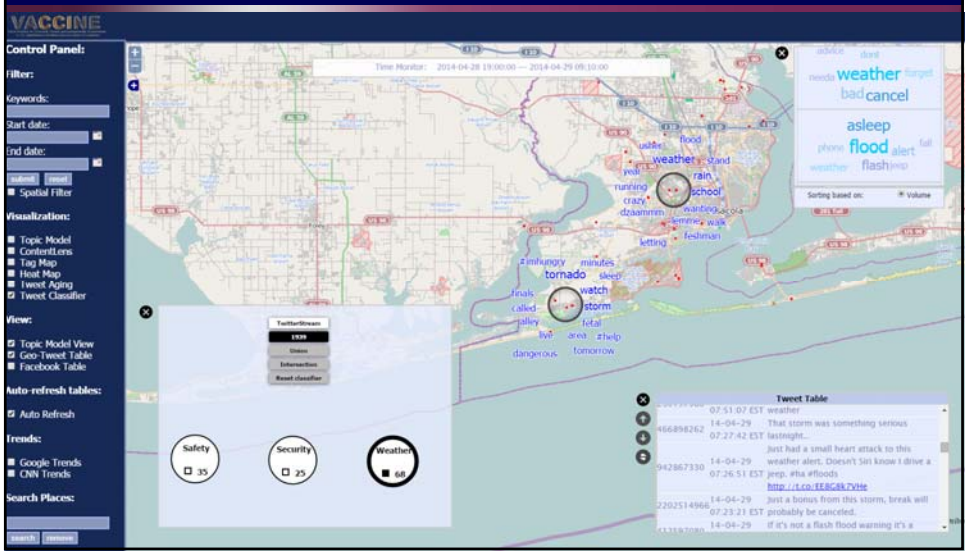
VACCINE

October 2015

SMART for Disaster Response

October 2015

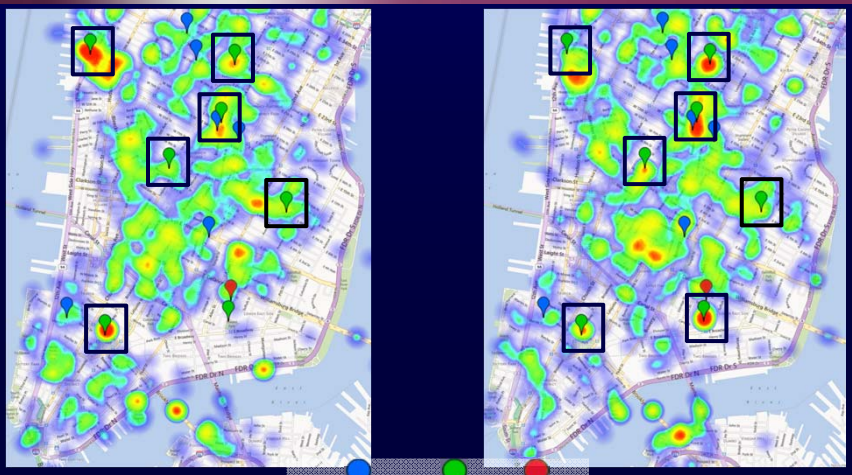
SMART: Disaster Preparedness & Response



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



VACCINE

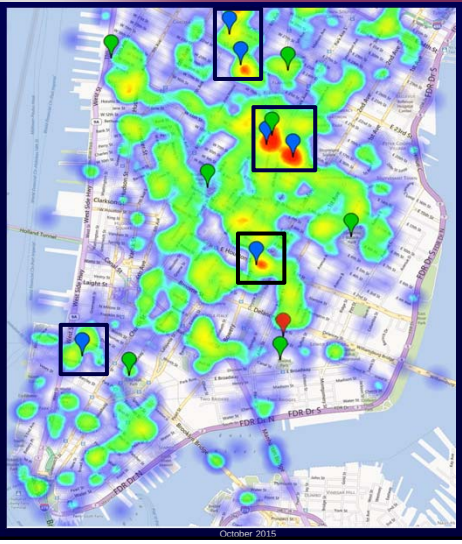
PURDUE

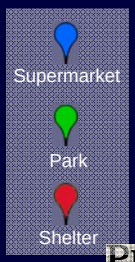
Visual Analytics of Activity During Hurricane Sandy

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

PURDUE

October 2015

SMART Case Study
Hurricane Joaquin

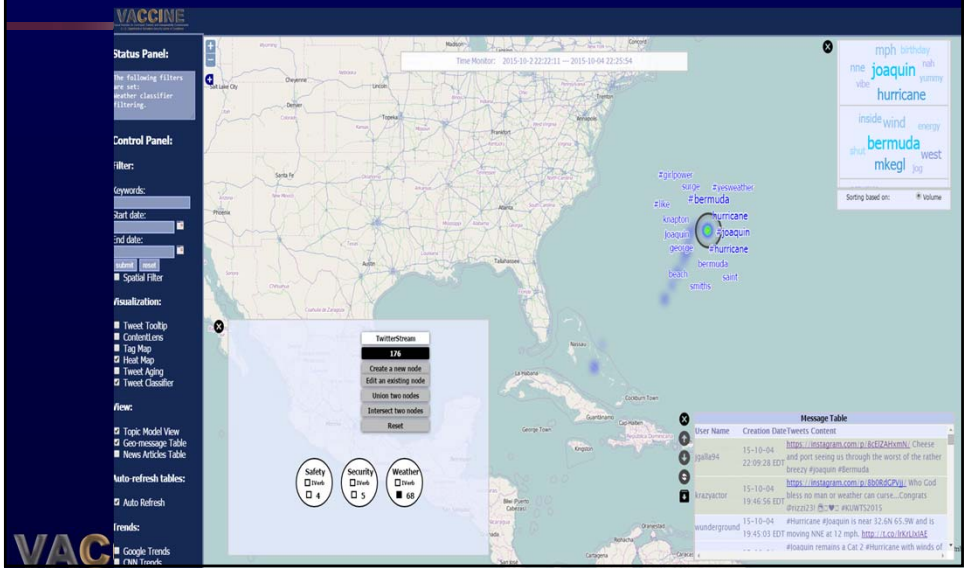
Time range: 2015-10-03 to 2015-10-04

Location: Southeast US

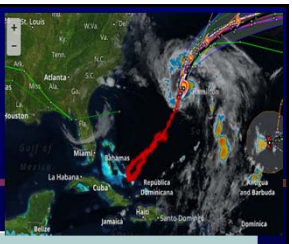
Data source: Twitter & Instagram

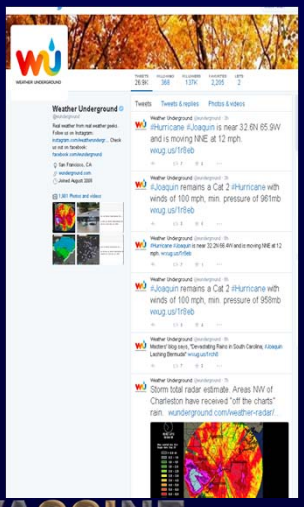
October 2015

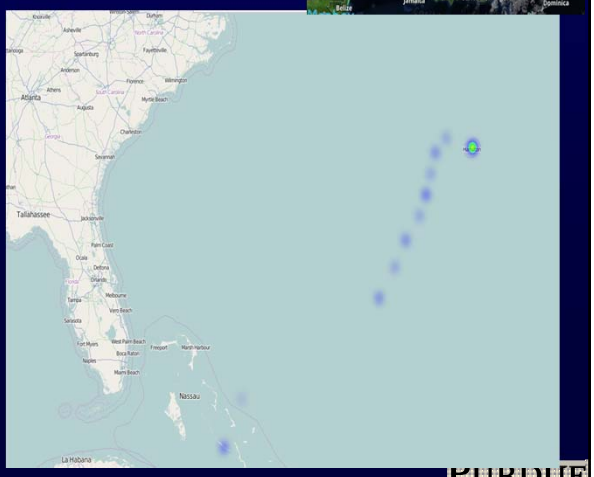
Overview of the event



News media Weather Underground posted tweets over the sea, forming a path similar to the path of Joaquin



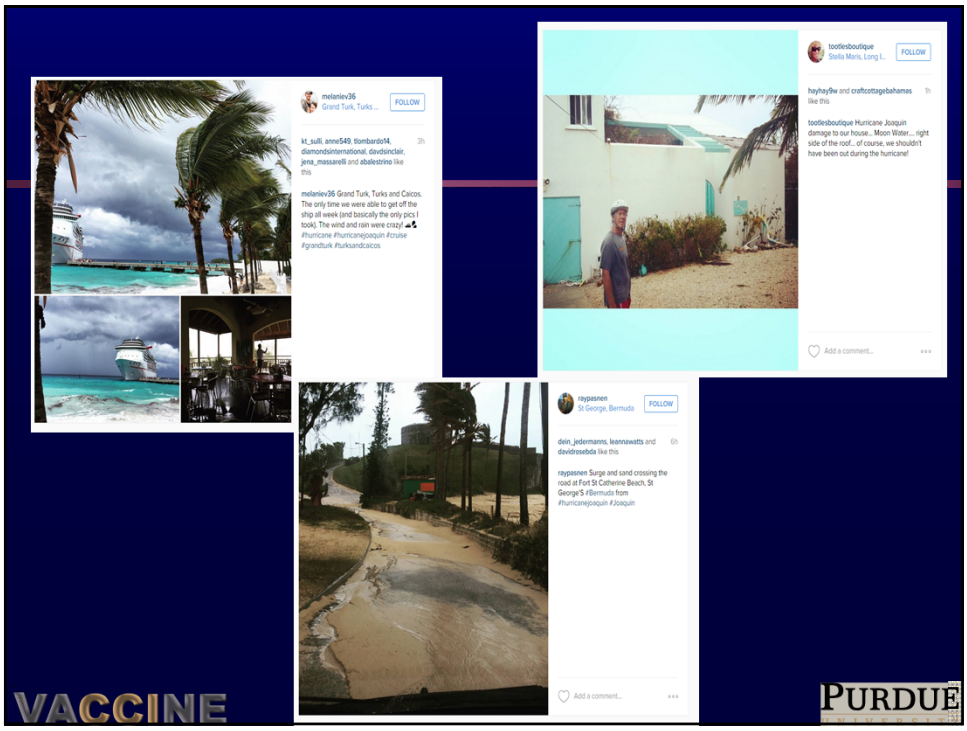


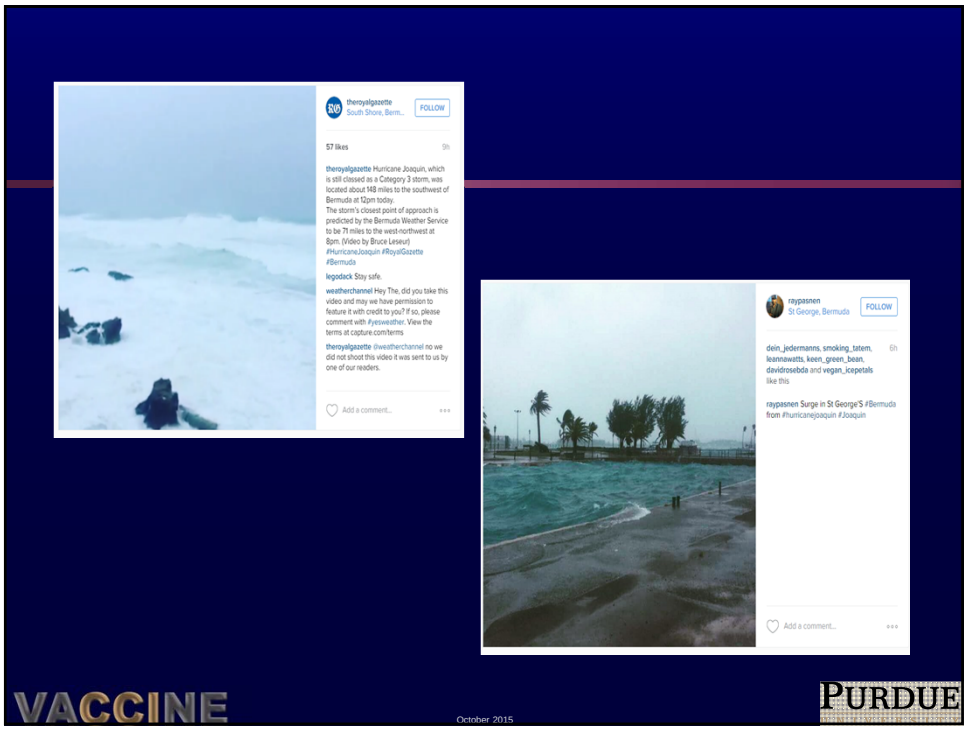


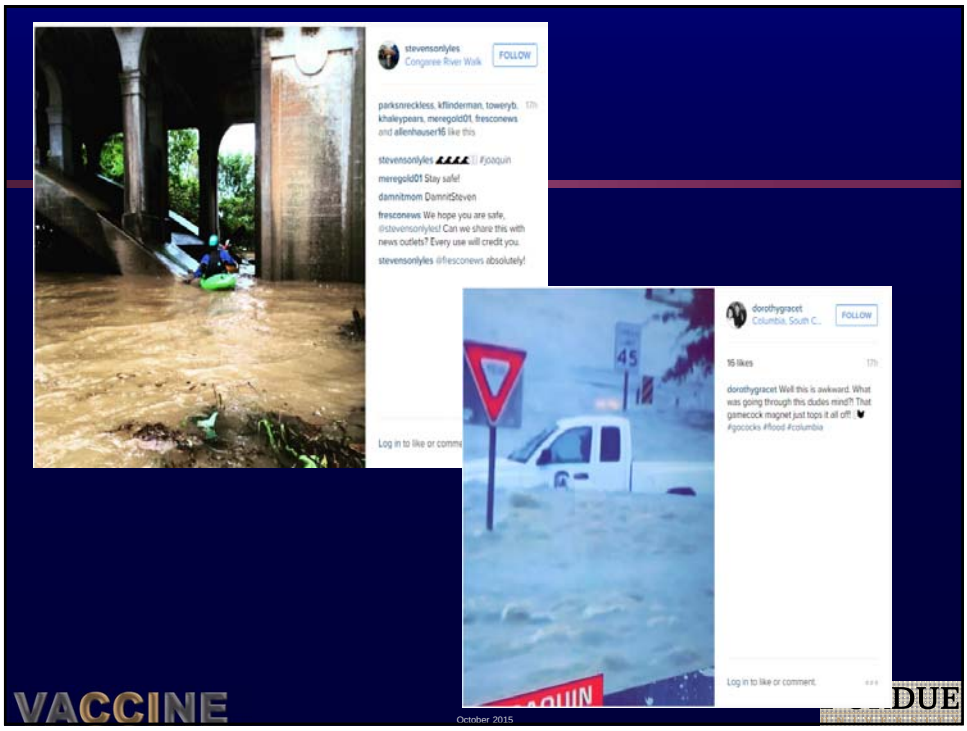
VACCINE

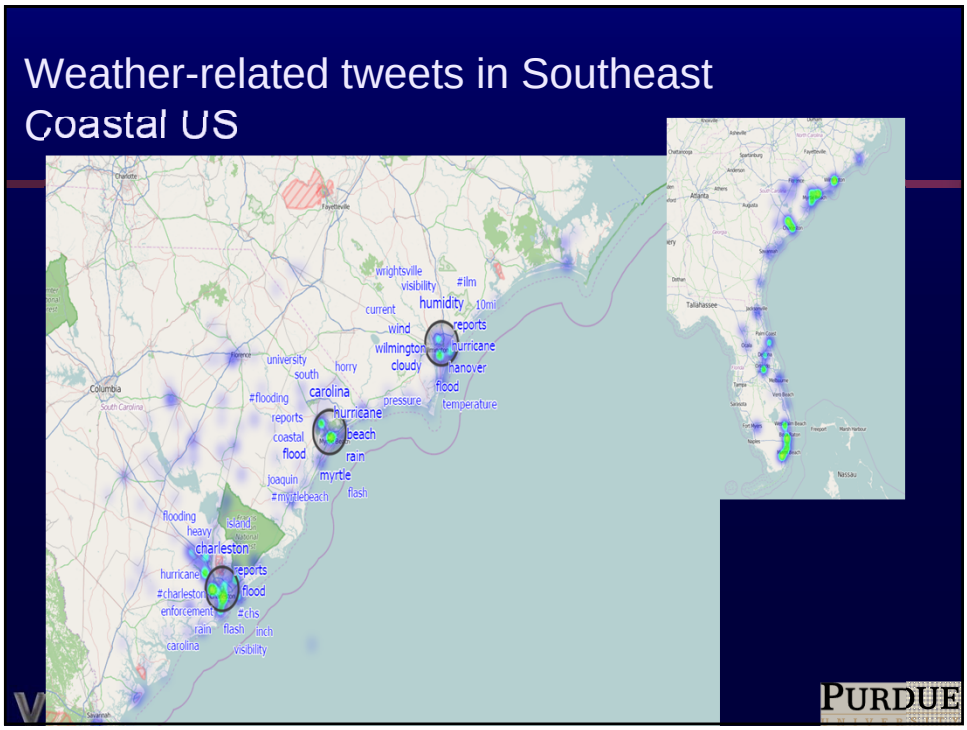
October 2015

PURDUE


melaniev36
Grand Turk, Turks ...
FOLLOW
kt_sulli, anne549, tiombardo14, diamondsinternational, davidsinclair, jena_massarelli and abalestrino like this
melaniev36 Grand Turk, Turks and Caicos. The only time we were able to get off the ship all week (and basically the only pics I took). The wind and rain were crazy! #hurricane #hurricanejoaquin #cruise #grandturk #turksandcaicos
tootiesboutique
Stella Maris, Long I...
FOLLOW
hayhay9w and craftcottagebahamas like this
tootiesboutique Hurricane Joaquin damage to our house... Moon Water... right side of the roof... of course, we shouldn't have been out during the hurricane!
Add a comment...
raypasnen
St George, Bermuda
FOLLOW
dein_jedermanns, leannawatts and davidrosebda like this
raypasnen Surge and sand crossing the road at Fort St Catherine Beach, St George'S #Bermuda from #hurricanejoaquin #Joaquin
Add a comment...
VACCINE
PURDUE


theroyalgazette
South Shore, Berm...
FOLLOW
57 likes
9h
theroyalgazette Hurricane Joaquin, which is still classed as a Category 3 storm, was located about 148 miles to the southwest of Bermuda at 12pm today.
The storm's closest point of approach is predicted by the Bermuda Weather Service to be 71 miles to the west-northwest at 8pm. (Video by Bruce Leseur)
#HurricaneJoaquin #RoyalGazette #Bermuda
legodock Stay safe.
weatherchannel Hey The, did you take this video and may we have permission to feature it with credit to you? If so, please comment with #yesweather. View the terms at capture.com/terms
theroyalgazette @weatherchannel no we did not shoot this video it was sent to us by one of our readers.
Add a comment...
raypasnen
St George, Bermuda
FOLLOW
dein_jedermanns, smoking_tatem, leannawatts, keen_green_bean, davidrosebda and vegan_jcepetals like this
raypasnen Surge in St George'S #Bermuda from #hurricanejoaquin #Joaquin
Add a comment...
VACCINE
October 2015
PURDUE


stevensonlyles
Congaree River Walk
FOLLOW
parksnreckless, kflinderman, toweryb, khaleypears, meregold01, fresconews and allenhauser16 like this.
stevensonlyles #joaquin
meregold01 Stay safe!
damnitmom DamnitSteven
fresconews We hope you are safe, @stevensonlyles! Can we share this with news outlets? Every use will credit you.
stevensonlyles @fresconews absolutely!
Log in to like or comme
dorothygracet
Columbia, South C..
FOLLOW
16 likes
dorothygracet Well this is awkward. What was going through this dudes mind?! That gamecock magnet just tops it all off! #gococks #flood #columbia
Log in to like or comment.
VACCINE
October 2015
PURDUE


Weather-related tweets in Southeast Coastal US
wrightsville
visibility
#ilm
humidity
10mi
current
wind
reports
wilmington
hurricane
cloudy
hanover
flood
university
south
horry
carolina
#flooding
pressure
temperature
reports
hurricane
coastal
beach
flood
rain
joaquin
myrtle
#myrtlebeach
flash
flooding
heavy
island
charleston
hurricane
reports
#charleston
flood
enforcement
#chs
rain
flash
inch
carolina
visibility
PURDUE

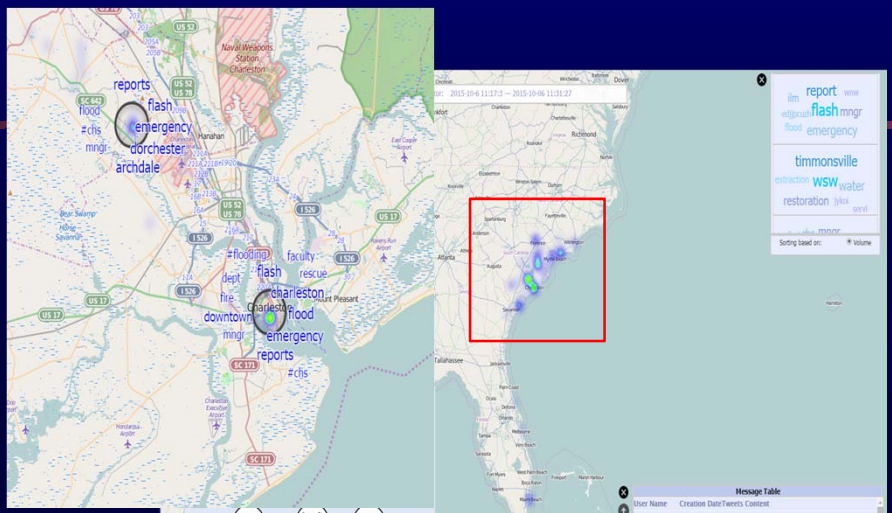
Top Topics related to Weather Classifier



VACCINE

October 2015

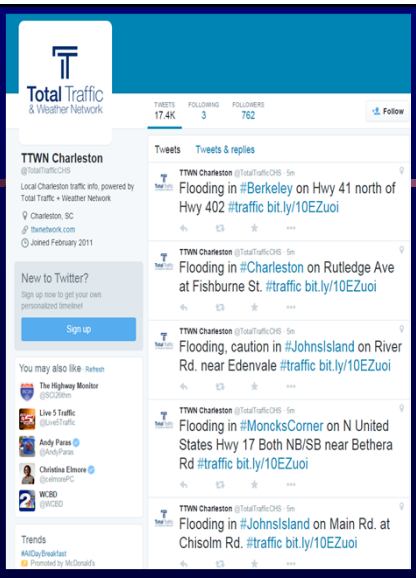
PURDUE



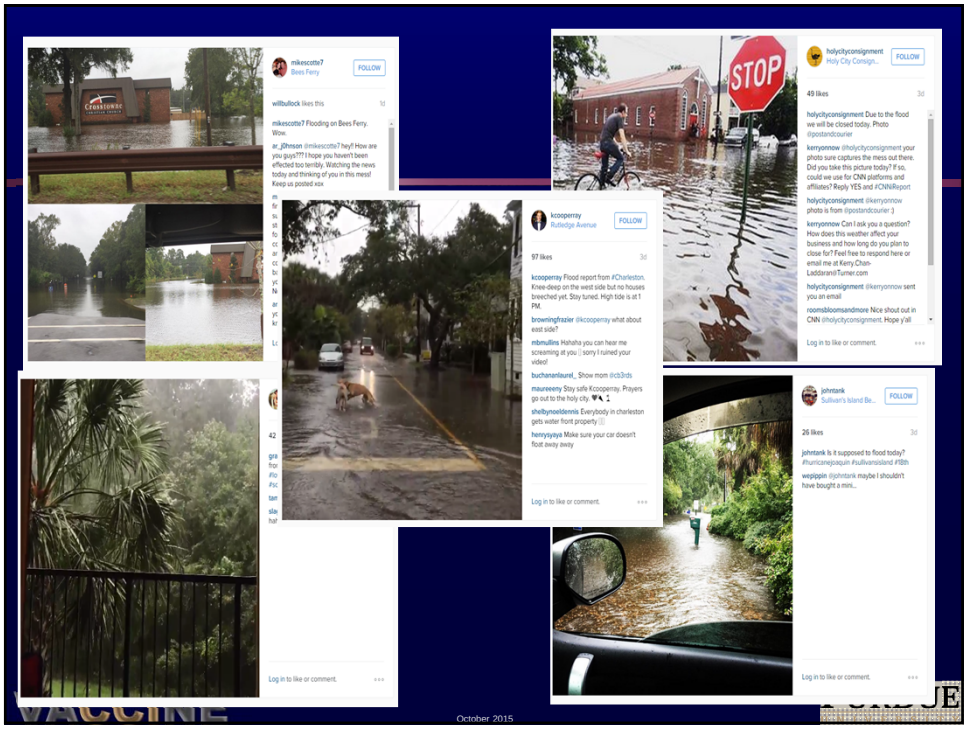
Filtering tweets related to safety/security and weather.

VACCINE

October 2015



These two users(weather media) posted many tweets updating the flooding in Charleston.



Summary of SMART For Event Decision Making

Examples:

- Purdue, Ohio State football games
- Kayaktivists
- MLB All Star Game
- Kentucky Derby
- Belle of Louisville Celebration, Thunder over Louisville
- Fleet Week in San Francisco
- Hurricane Joaquin
- Community College

VACCINE October 2015 PURDUE

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - 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

VACCINE October 2015 PURDUE

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

National / Global Decision Making

- Precision agriculture and droughts
- USCG
- Southern Border Approaches
- Soundscapes

VACCINE

October 2015

PURDUE

Precision Agriculture

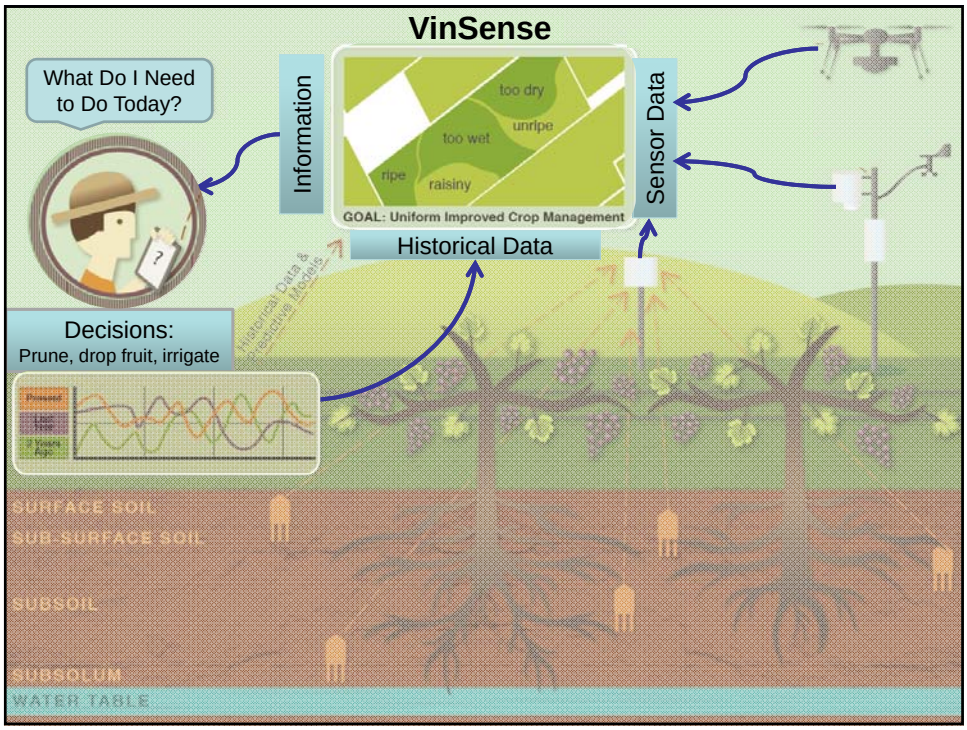
Better Decisions for Wine Grape Crop Management

Highly variable grapes grown with suboptimal decisions

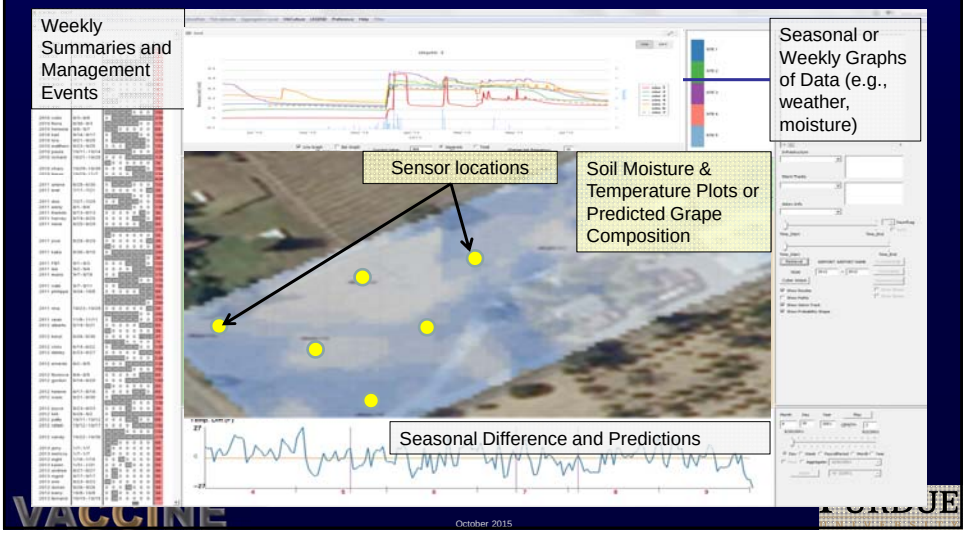


Uniform quality grapes grown with VinSense based decisions

October 2015



Prototype Layout & **Result**



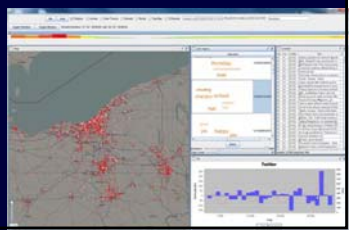
Commercializing Precision Viticulture Work

PURDUE
OFFICE OF TECHNOLOGY COMMERCIALIZATION

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



VACCINE

October 2015

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

October 2015

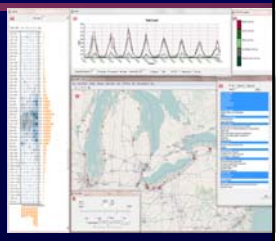
U.S. Coast Guard Search and Rescue VA

(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



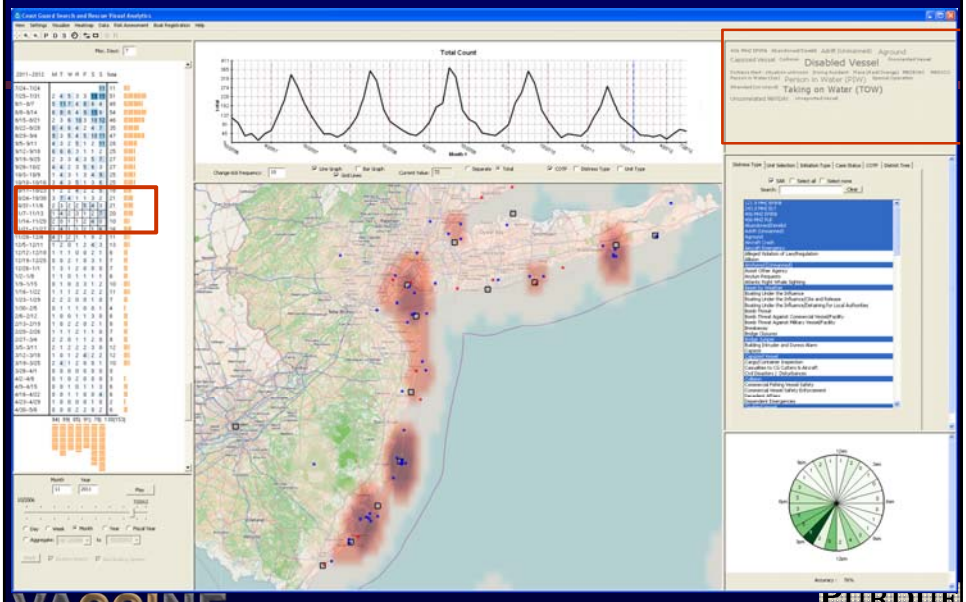


VACCINE

October 2015

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

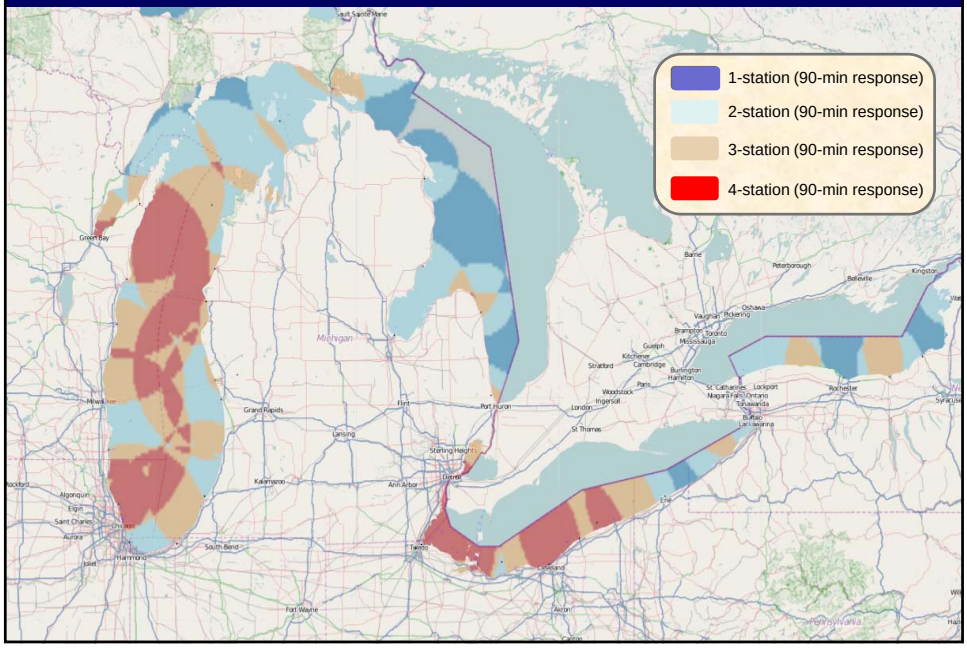


VACCINE

October 2015

PURDUE

Response Efficiency – Potential Future Assets



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

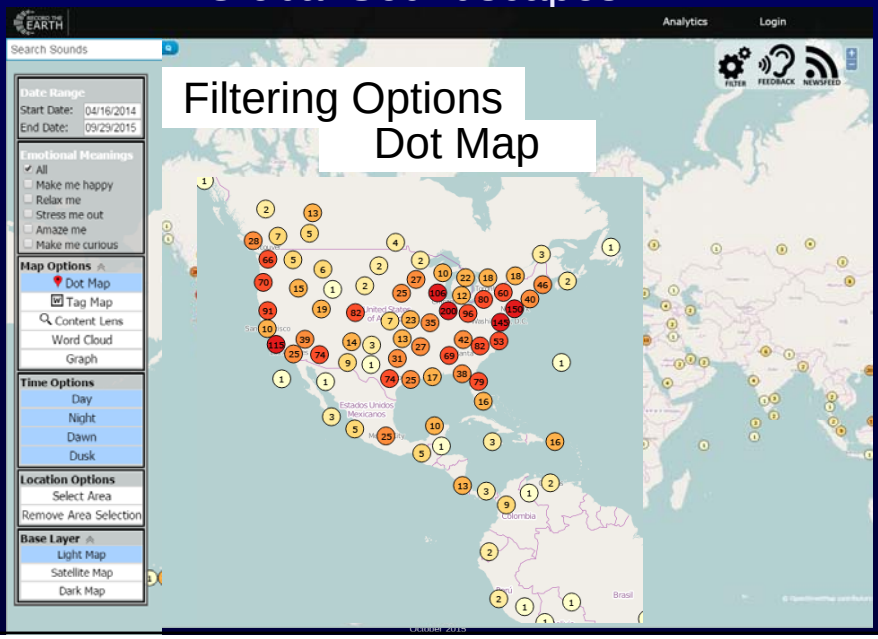
Southern Border Approaches

- Department of Homeland Security has created joint task forces to address Southern Border Approaches security (CBP, ICE, US Coast Guard)
- Joint Task Forces have asked us to assist in visualizing data and metrics to help define performance and evaluate success

Global Soundscapes

- What does the Earth sound like today?
  - Join with researchers involved in one of the newest scientific disciplines - Soundscape Ecology - as they map the sounds of our planet, and what they tell us about the health of the world we live in.
- What kind of sounds make a soundscape?
  - Biophony
    - Sounds of birds and animals
  - Geophony
    - Sounds of wind and water
  - Anthrophony
    - Sounds of vehicles and machines
  - Emotion
    - Sounds that make you happy, stressed, or excited

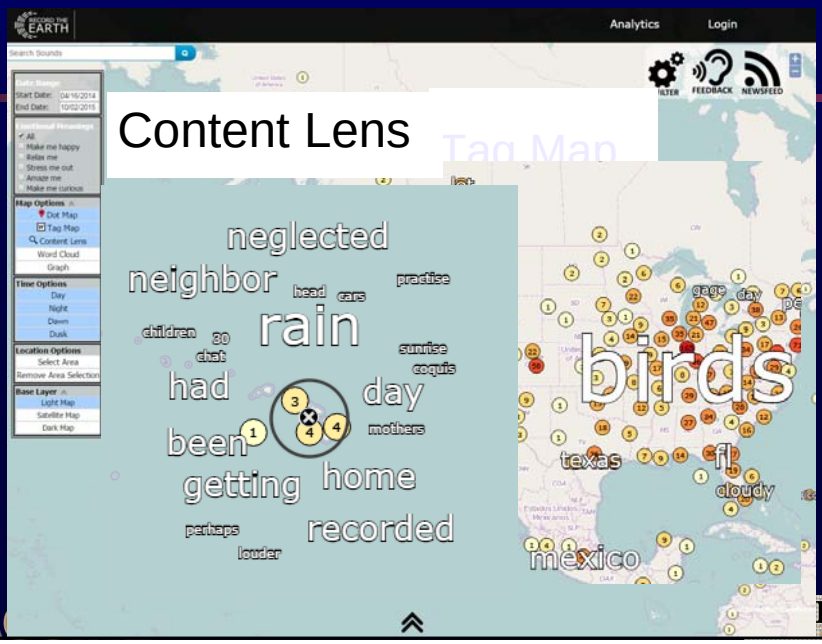
Global Soundscapes



Global Soundscapes



Trending Views



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

Human-Computer Collaborative Decision Making



David S. Ebert
Purdue University
ebertd@purdue.edu

April 2016

Why VA and HCCD For Big Data?

(Or What Big Data Analytics Can't Do) [1]

- 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

VACCINE

[1] Inspired and adapted from David Brooks, New York Times, 2/18/2013)

April 2016

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

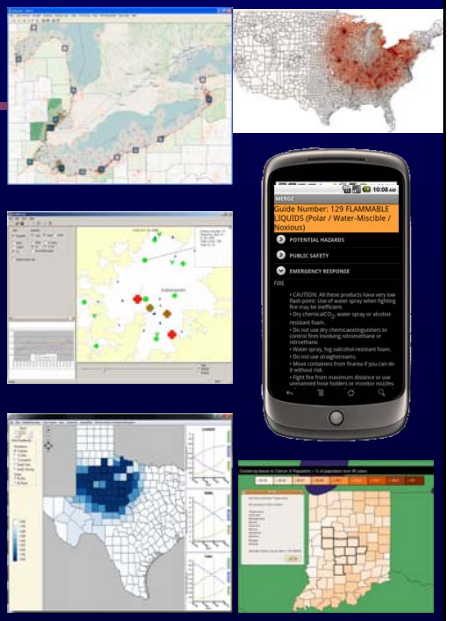
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address



Public Safety & Health

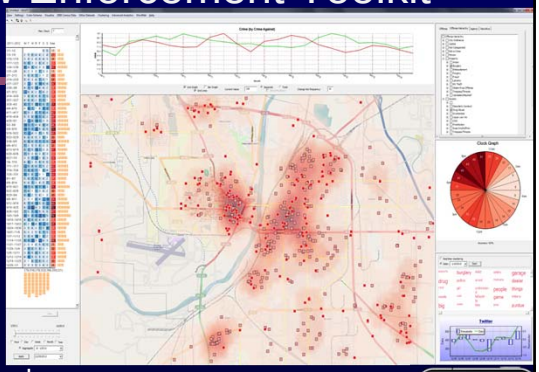
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering



VACCINE

April 2016

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

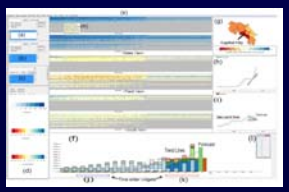


VACCINE

April 2016

Business Visual Analytics





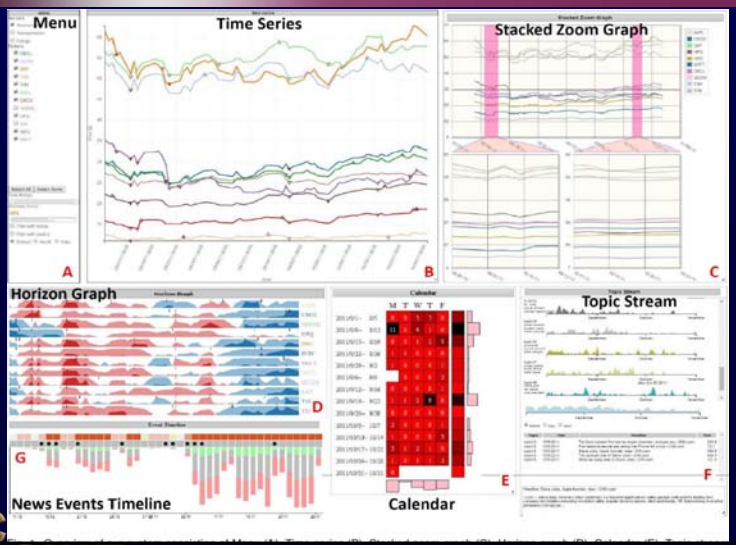


- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation

VACCINE

April 2016

Financial Anomaly HCCD



VAC

TECHNOLOGY VinSense

Respect for tradition, vision for the future

Agricultural HCCD

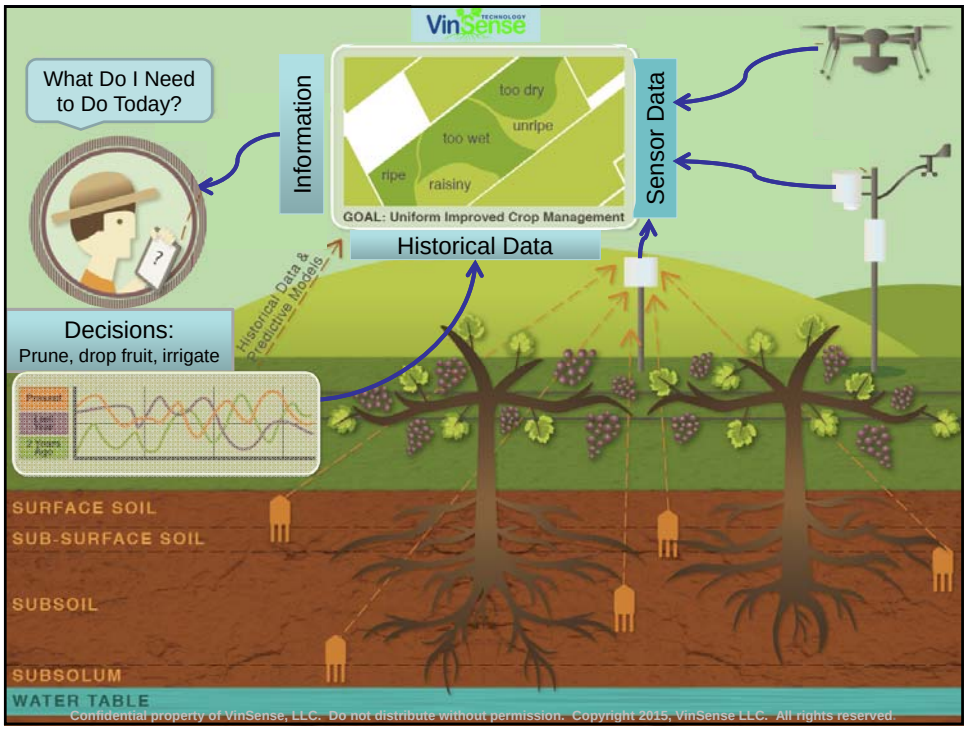
Better Decisions for Wine Grape Crop Management

Highly variable grapes grown with suboptimal decisions

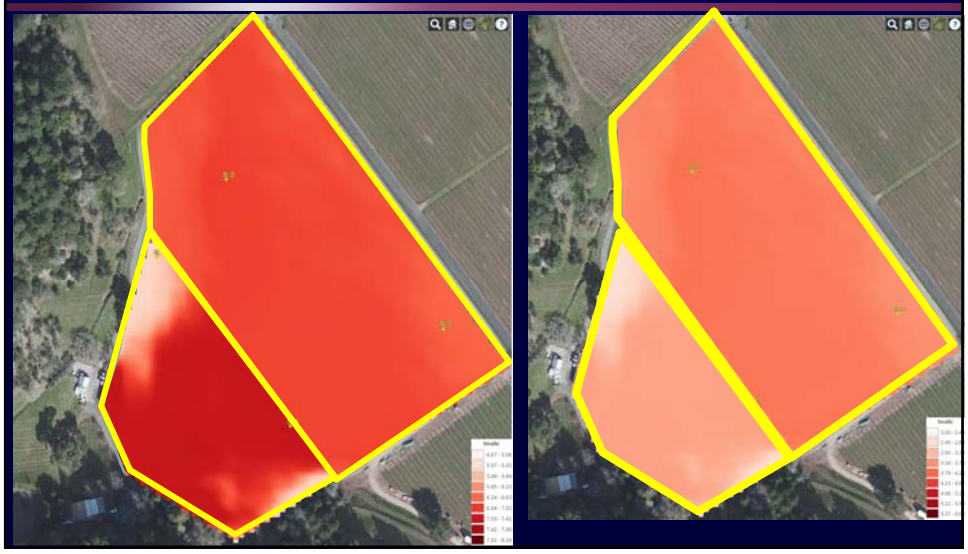


Potential uniform quality grapes grown with VinSense based decisions

Property of VinSense, LLC. Do not distribute without permission. Copyright 2016, VinSense LLC. All rights reserved.



Veraison vs. Harvest Information: L-malic



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

VACCINE

April 2016

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

VACCINE

April 2016

Our Approach: Natural Scale Templates

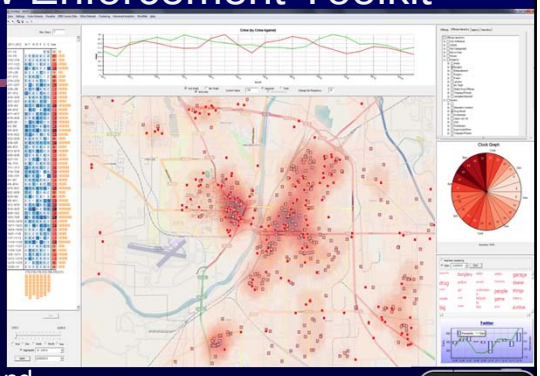
- Overcomes impedance mismatch between data size and human ability to understand [Robertson et al., 2009, Kasik et al., 2009]
- Based on design principle of Naturalness[1]
  - Selecting appropriate mappings natural to human cognition
- Spatial and temporal visualization and analysis templates that balance 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.

VACCINE

April 2016

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

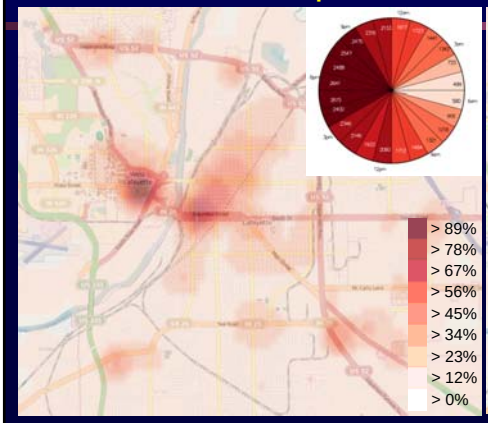


VACCINE

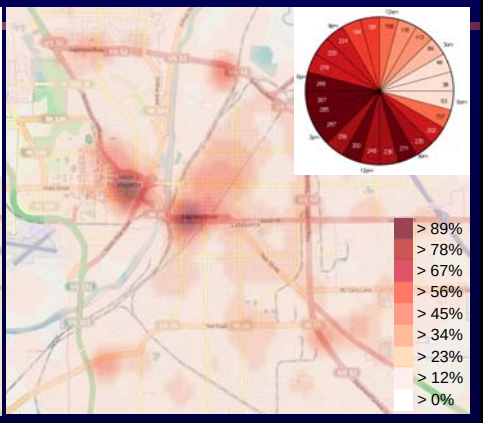
April 2016

Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data



Past 2 Years
3/11/2012 – 3/10/2014

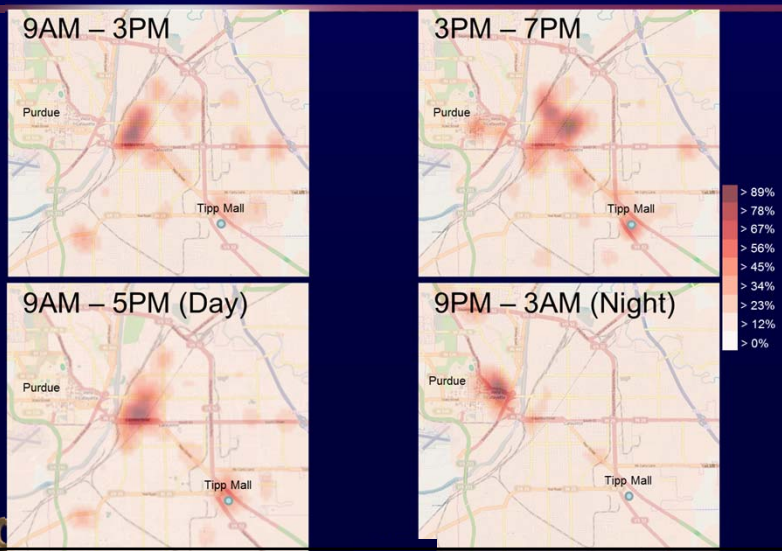


March + Tuesday Filter
(Past 10 Years)

VACCINE

April 2016

Prediction During Certain Hours In The Day

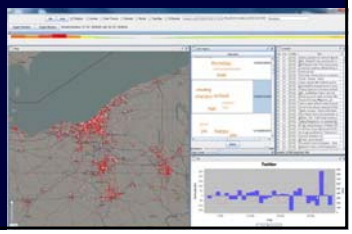


VAC

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



VACCINE

April 2016

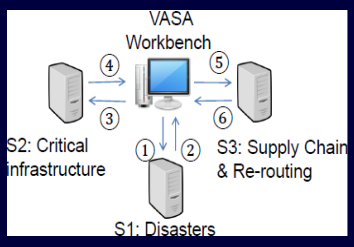
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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

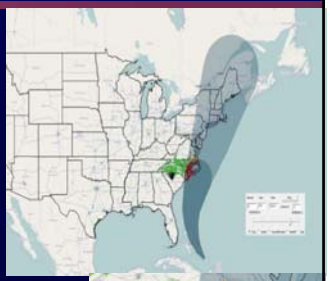
April 2016

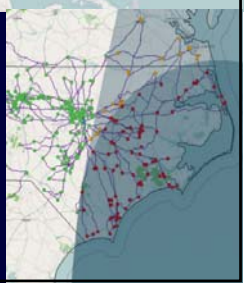
VASA: Visual Analytics for Simulation-Based Action

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

April 2016

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

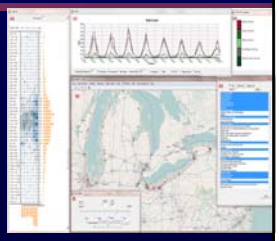
April 2016

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



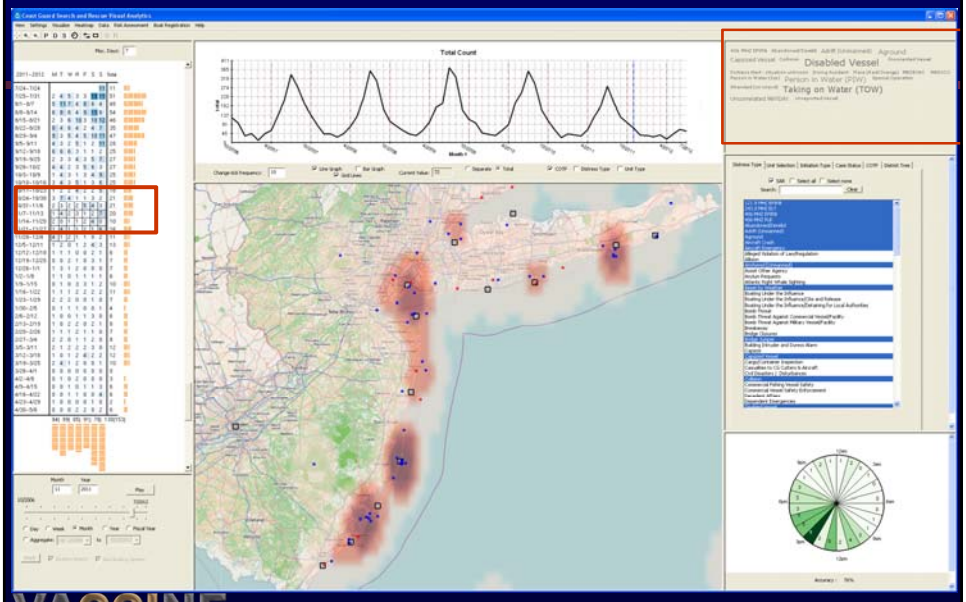


VACCINE

April 2016

Example: Risks and Consequences From Sandy:

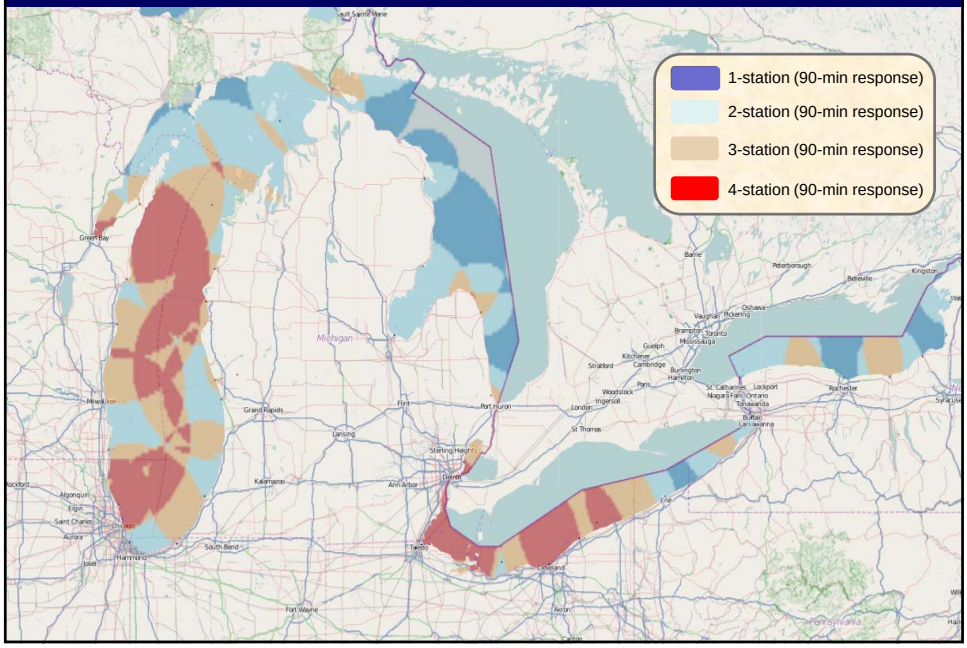
SAR Cases November 2011 NJ/NYC Area



VACCINE

April 2016

Response Efficiency – Potential Future Assets



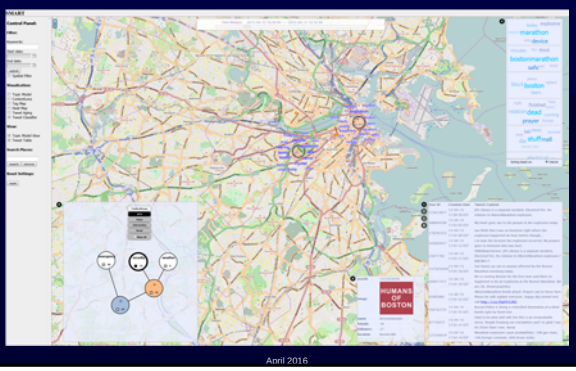
VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media for Preparedness and Emergency Response

(Social Media Analytics and Reporting Tool)



April 2016

Social Media Analytics 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

April 2016

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

April 2016

Anomaly Detection Superbowl Weekend



Where Can It Be Helpful?

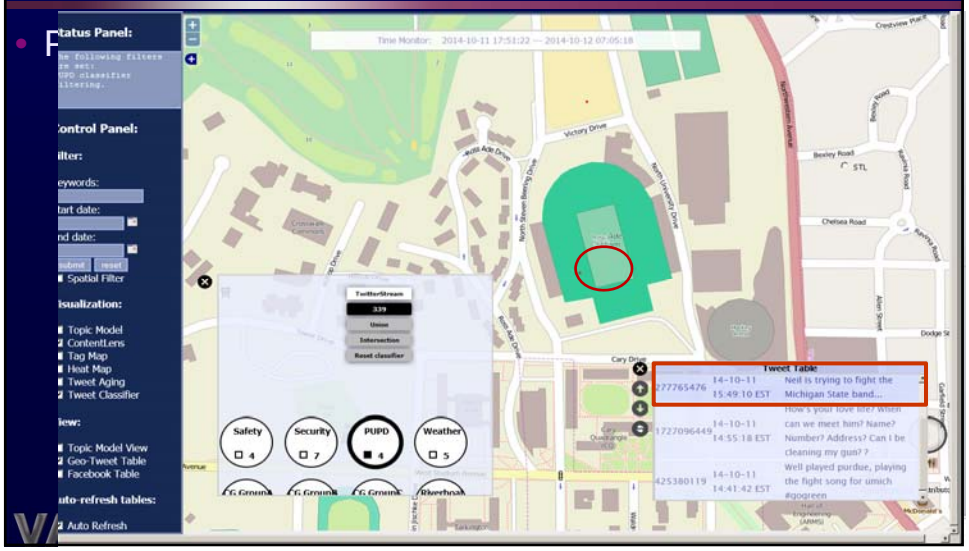
- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

April 2016

Example From Purdue Football 10/12/14

- F



Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day

- Umpqua Community College Shooting



VACCINE

April 2016

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day



kullgreni
Roseburg, Oregon
FOLLOW

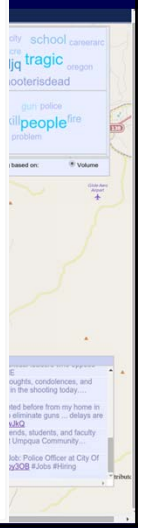
20 likes 3d

kullgreni Prayer chain for UCC shooting victims. The end of a long, tragic day.

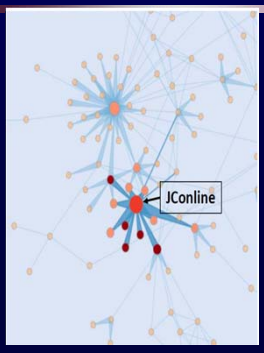
thadizzo84 😔😔😔

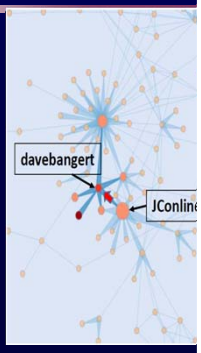
sgonzalez910 Such a tragic thing to cover. Stay strong

Log in to like or comment.



Tracking Information Diffusion and Identifying Influential Users Through User [illegible] Shooting Event)







JConline is a local news media that has a lot

JConline retweeted from **davebangert** about the updates of the shooting event.

davebangert retweeted from **justinlmack**.

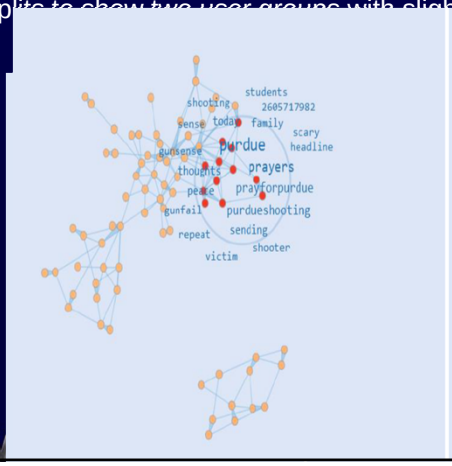
justinlmack posted first-hand updates about the event.

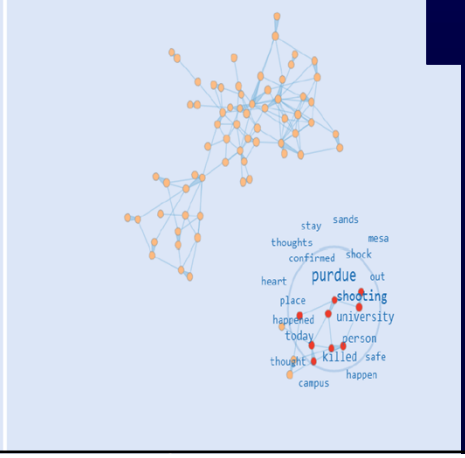
VACCINE

April 2016

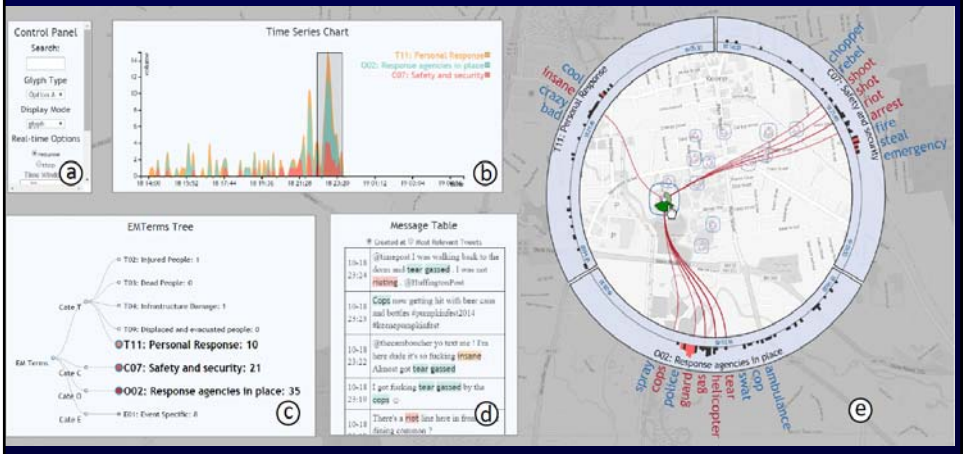
Monitoring Network Evolutions Over Time

Initially, there is a single network containing ‘purdue’, ‘shooting’, ‘custody’. As time passes, the network evolves and new nodes come. The network then spl[illegible]



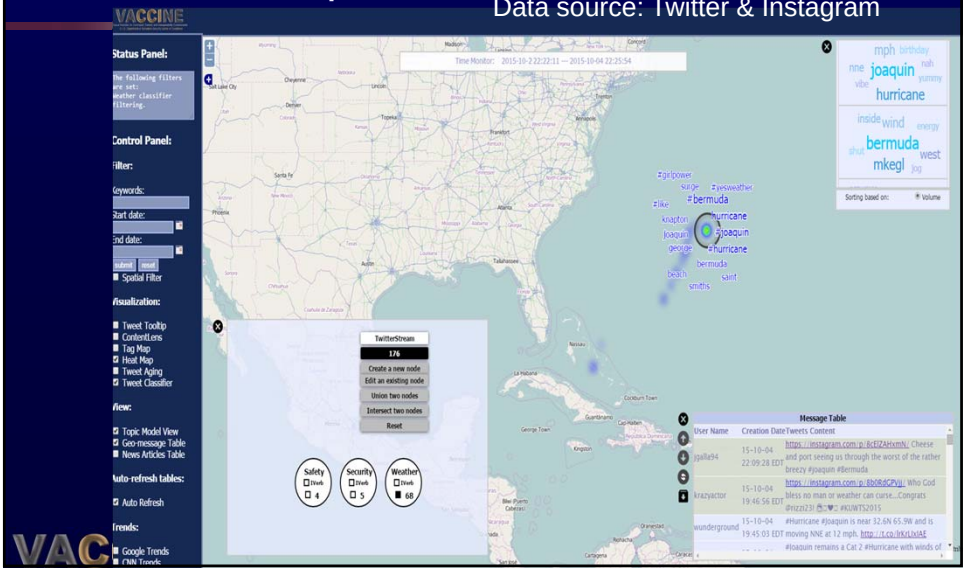


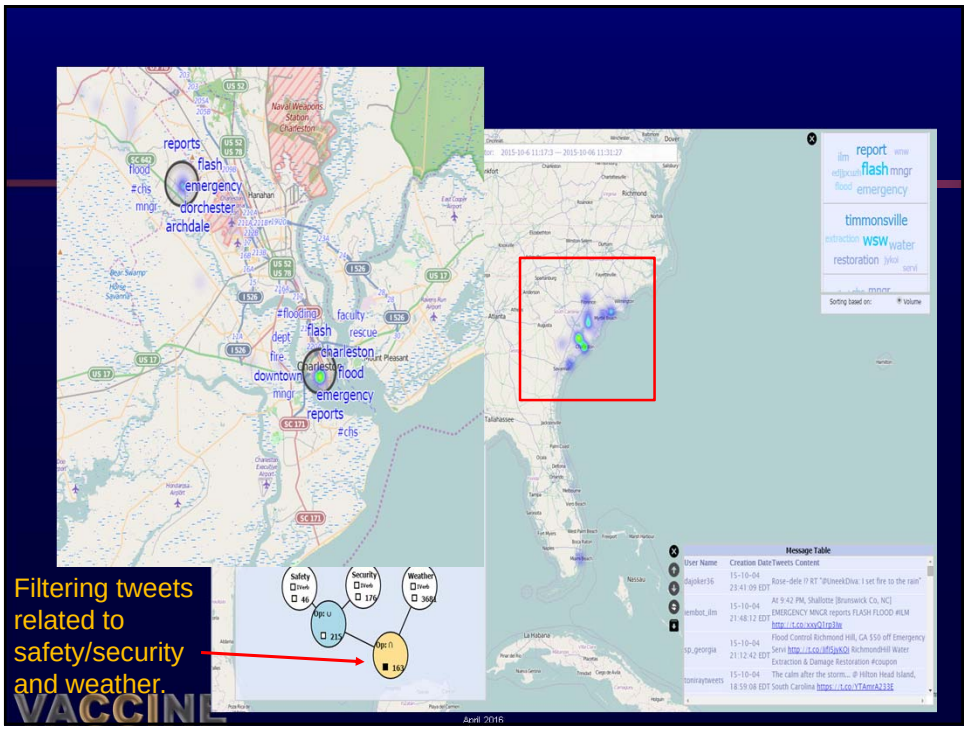
Multi-level Aggregation of Topic Evolution (Eurovis 2016)

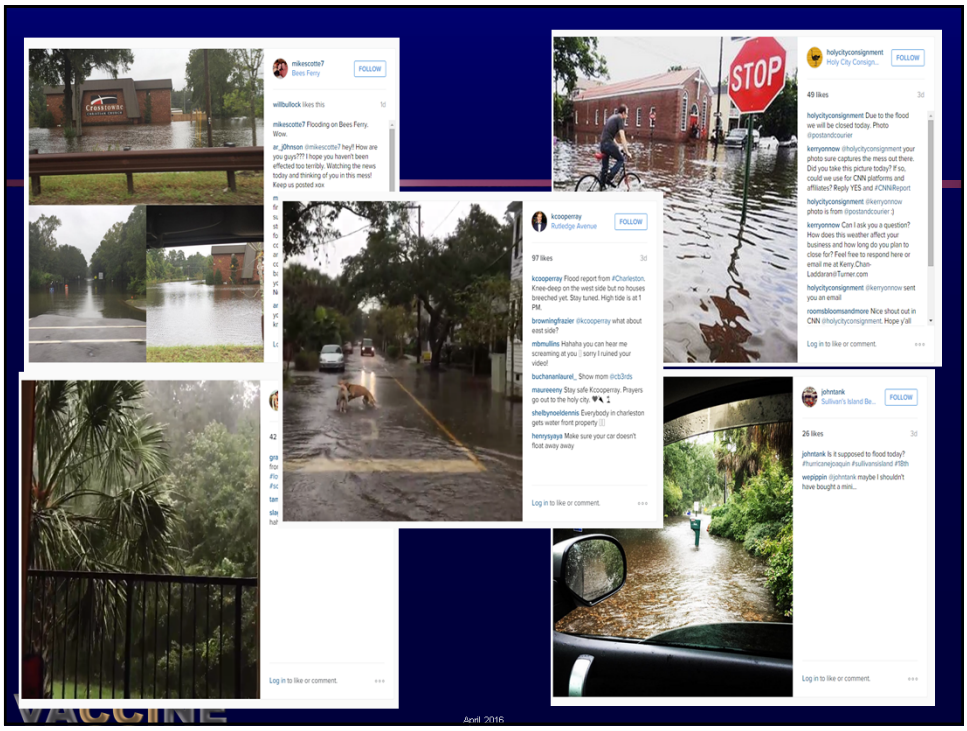


Hurricane Joaquin

Time range: 2015-10-03 to 2015-10-04
Location: Southeast US
Data source: Twitter & Instagram

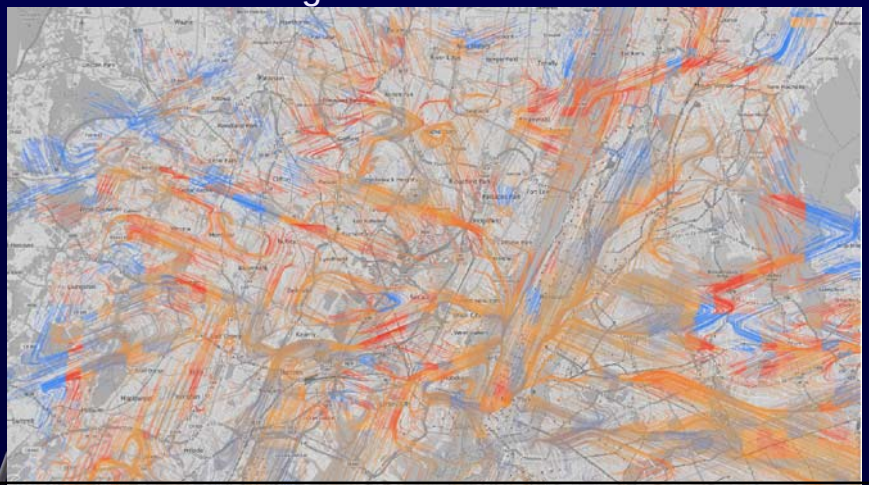



reports
flash
emergency
dorchester
archdale
#flooding
faculty
rescue
dept
flash
charleston
fire
downtown
flood
emergency
reports
#chs
report
flash
flood
emergency
timmonsville
wsw
water
restoration
Message Table
Safety
Security
Weather
Filtering tweets related to safety/security and weather.
April 2016


STOP
April 2016

Movement Patterns from Social Media Data

- Manhattan morning rush hour



Summary of SMART For Event Decision Making

Examples:

- Purdue, Ohio State football games
- Kayaktivists
- MLB All Star Game
- Kentucky Derby
- Belle of Louisville Celebration, Thunder over Louisville
- Fleet Week in San Francisco
- Hurricane Joaquin
- Community College
- RNC Summer 2016

VACCINE

April 2016

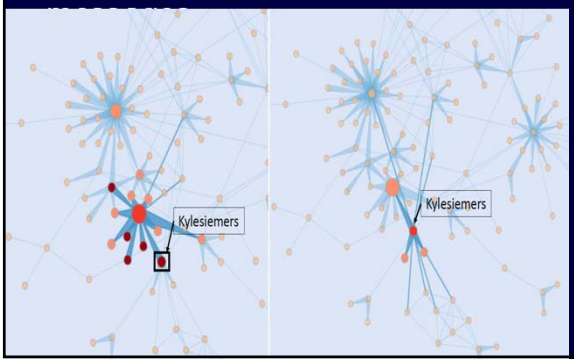
For Further Information

www.VisualAnalytics-CCI.org



Identify Influential Users In Purdue Shooting

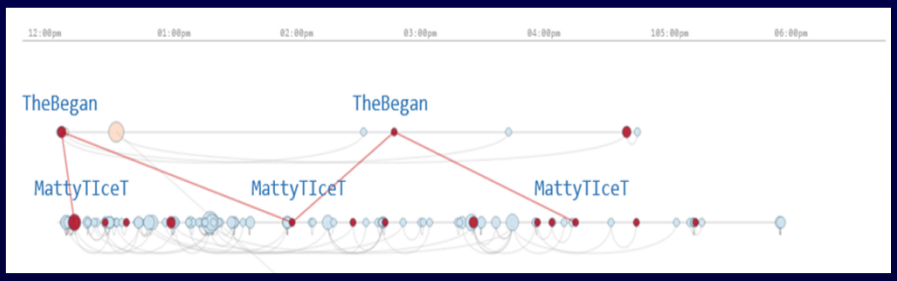
Click **JConline** to show its information traf
darker nodes along the path: **Jconline** ->
Kylesiemers has no darker nodes connec
Kylesiemers as an influential node and m





A timeline of the conversation between **TheBegan** and **MattyTIceT** about the shooting.

@MattyTIceT: White male grey sweatshirt running by pych building is description of a suspect that just came over the scanner.



RT @TheBegan: Groupme texts are saying a TA has been shot in EE.

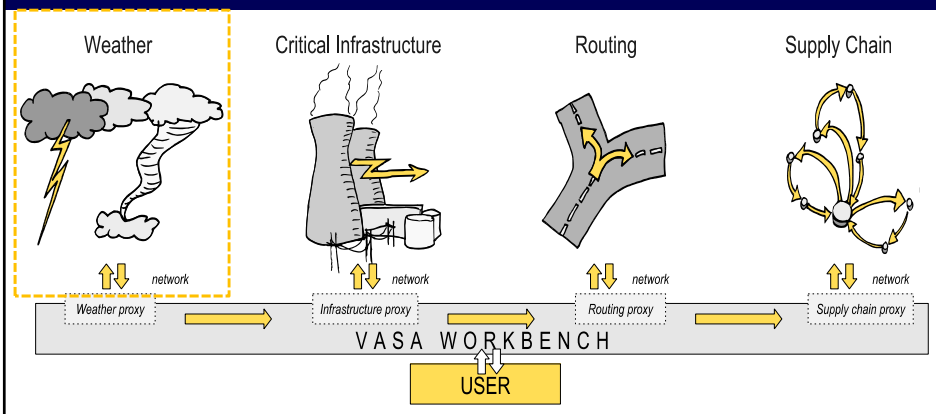
@TheBegan according to the police scanner they have one person in custody

@TheBegan WTHR is airing the audio from the press conference

VACCINE

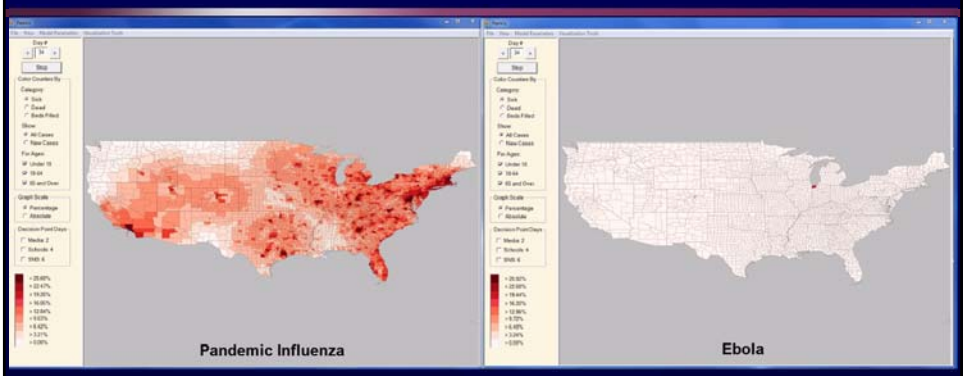
April 2016

VASA COMPONENT: **CRITICAL INFRASTRUCTURE**



Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL



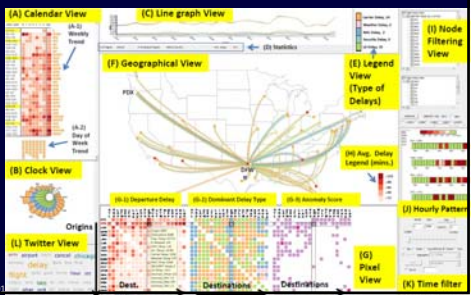
Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

VACCINE

April 2016

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects



VACCINE

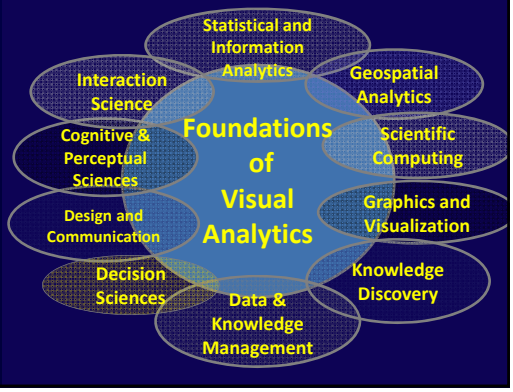
April 201

What We Do:Visual Analytics



Visual Analytics[1] is the science of analytical reasoning facilitated by interactive visual interfaces

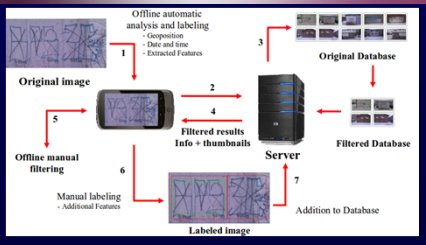
Interactive visualization, data analysis, exploration, and decision making with human in the loop!



1. *Illuminating the Path: The R&D Agenda for Visual Analytics*

VACCINE

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:

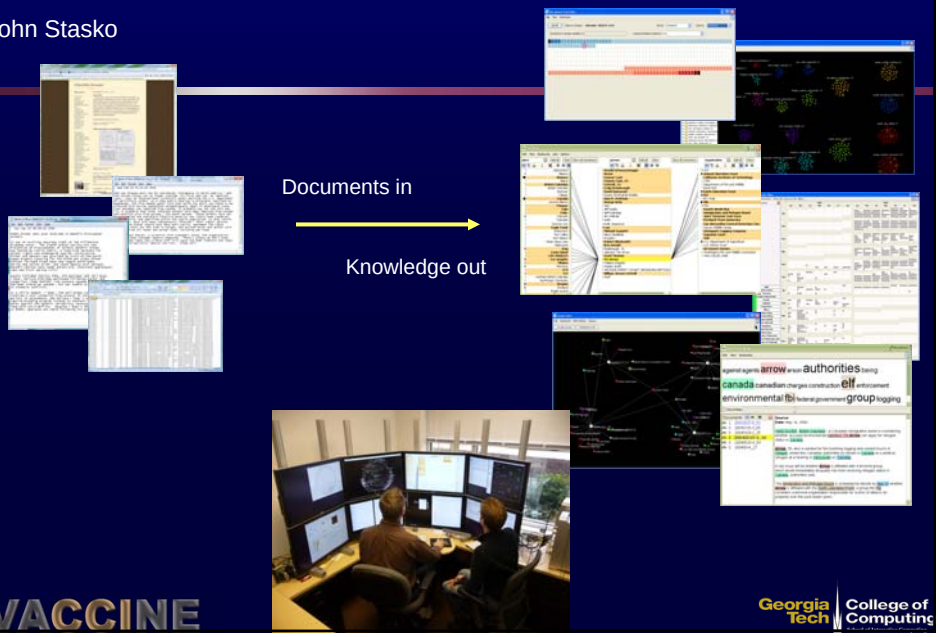
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

April 2016

Jigsaw: Visual Analytics for Investigative Analysis

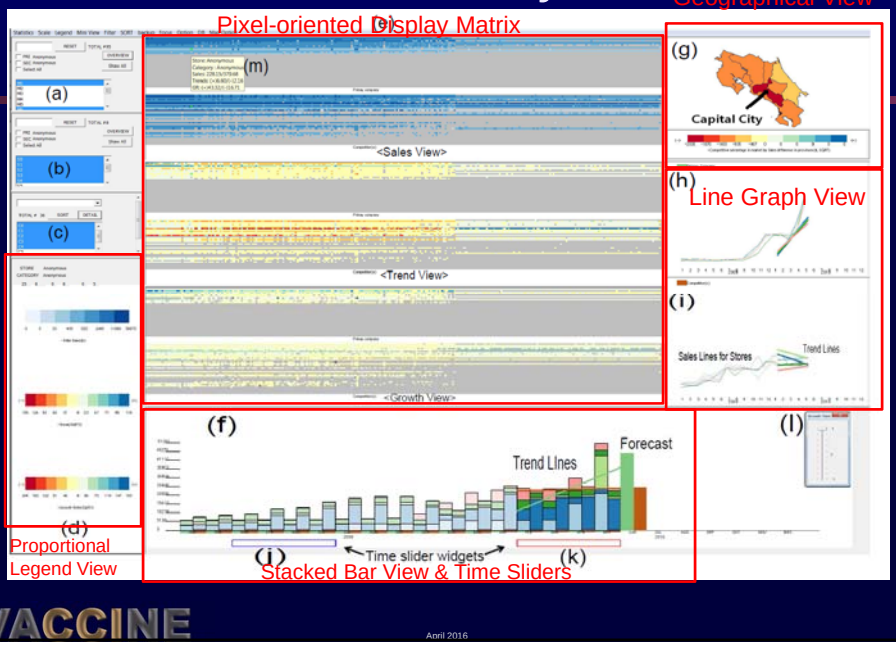
John Stasko



VACCINE

Georgia Tech College of Computing

Market Analyzer



VACCINE

April 2016

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

VACCINE

April 2016

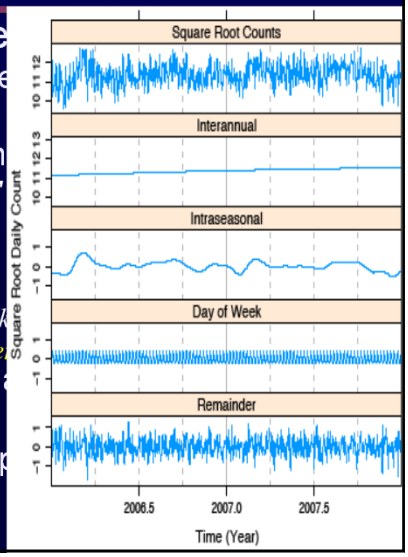
Event/Anomaly Detection - STL

- 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

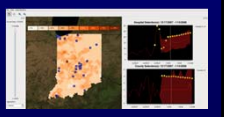
- Large value of *R* indicates substanti
- Filter out the noise component
- If value varies by 2 SD from STL exp

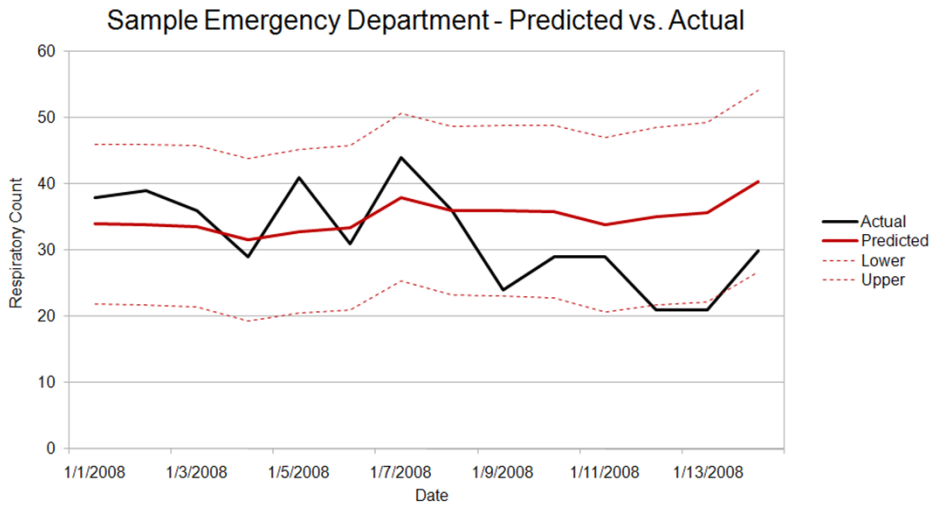


VACCINE

April 2016

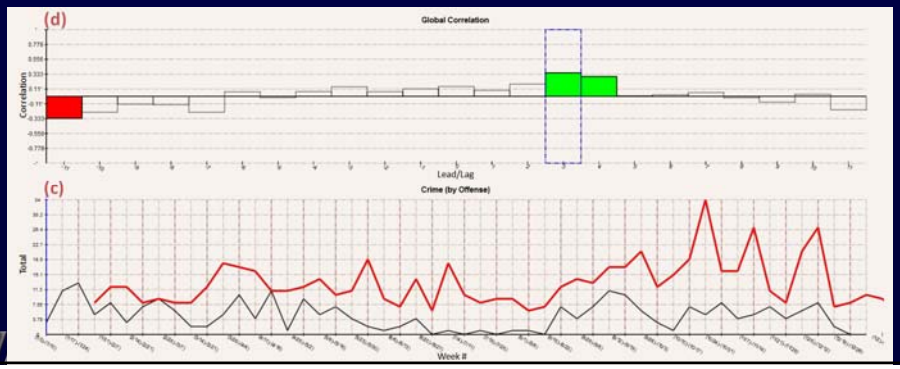
Predictive Visual Analytics



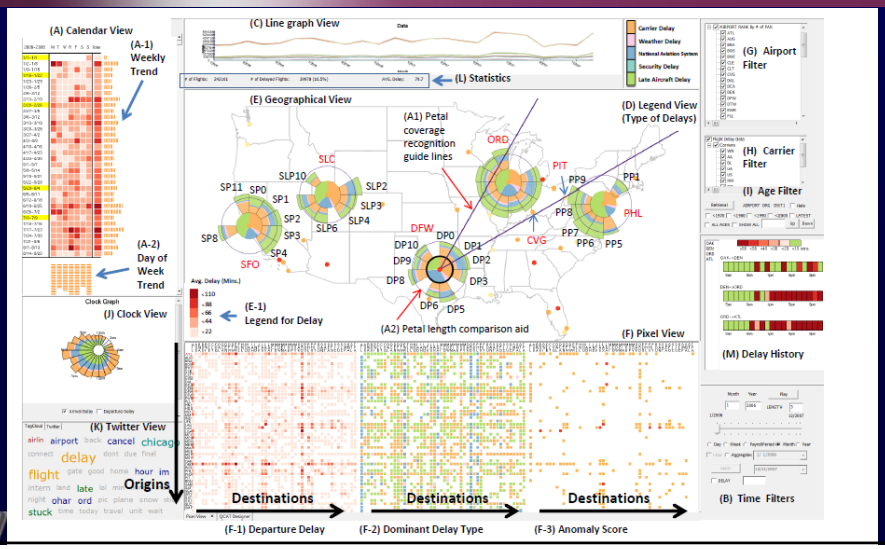


Multivariate Correlative Predictive Analytics: Three Approaches

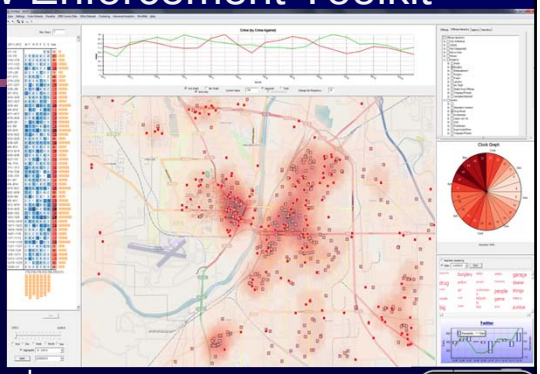
- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space



Multivariate Network Analytics with Information Theory Anomaly Detection



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:

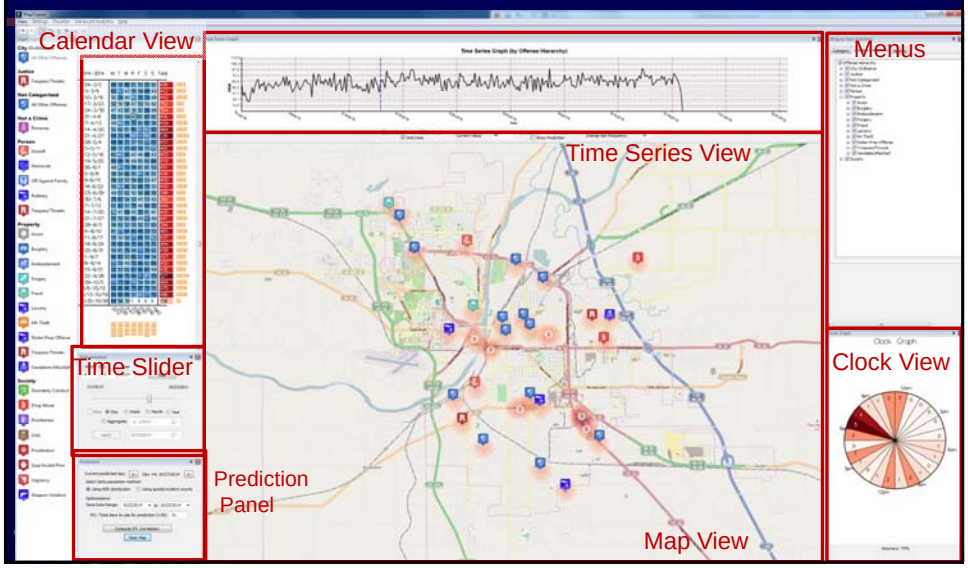
- October 2011: Purdue, WL Police, Lafayette PD



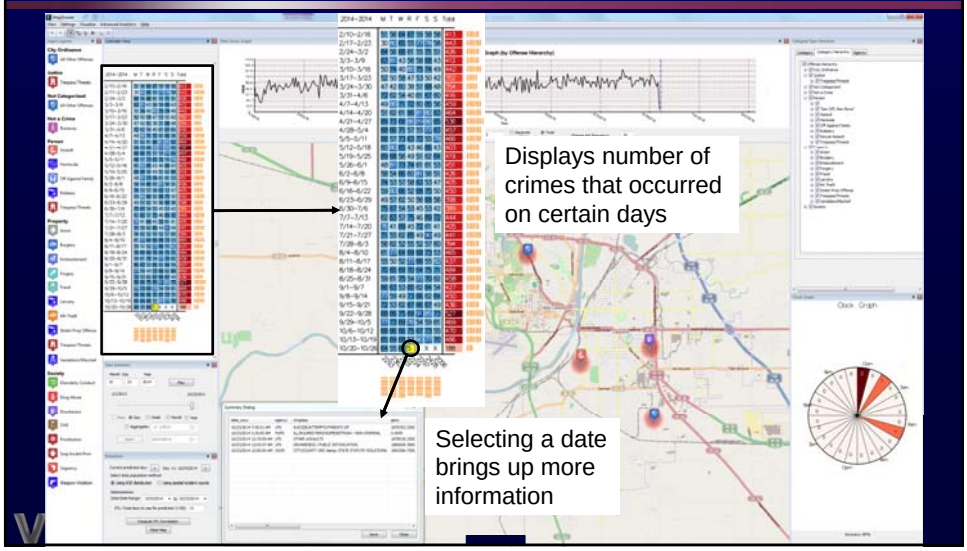
VACCINE

April 2016

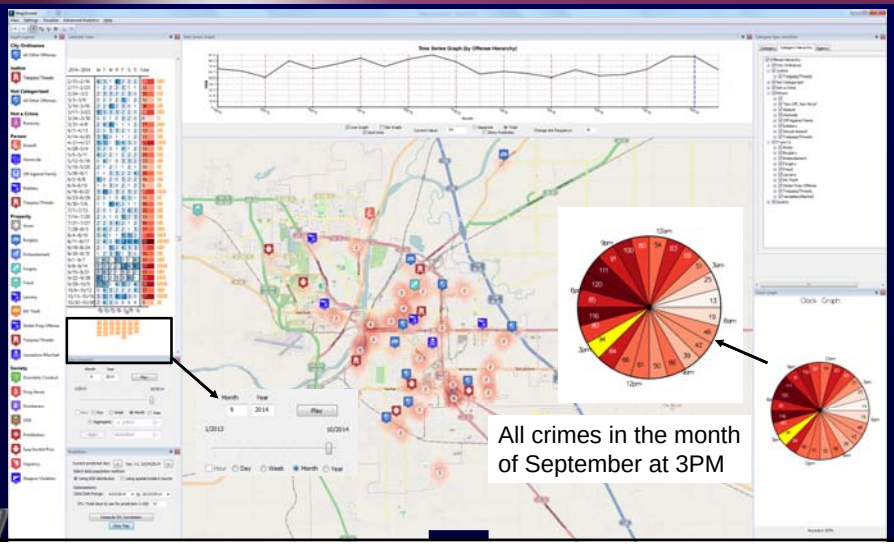
VALET Layout

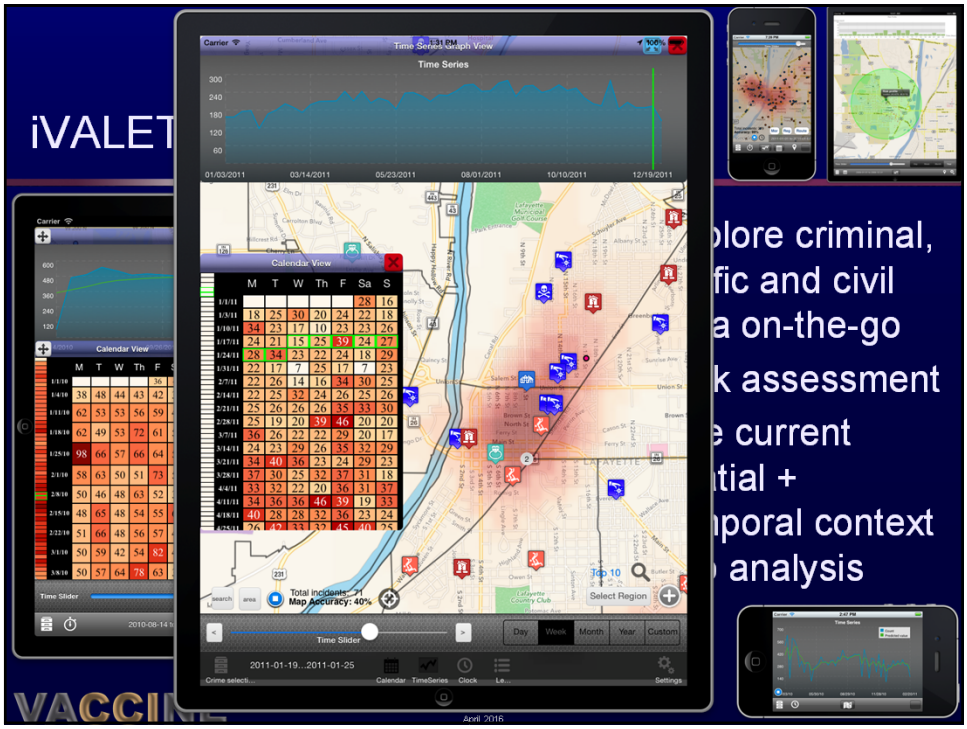


Calendar Display



Crimes During Certain Hours In A Month





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

April 2016

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

April 2016

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

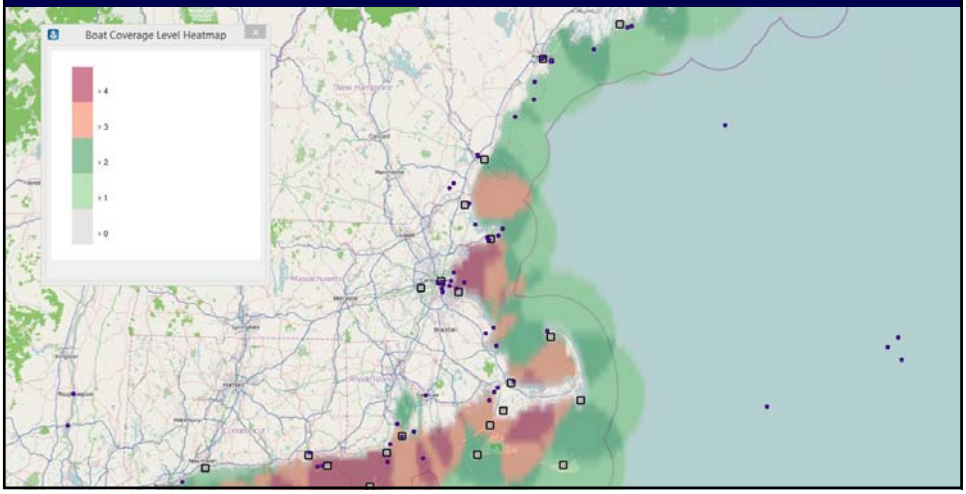




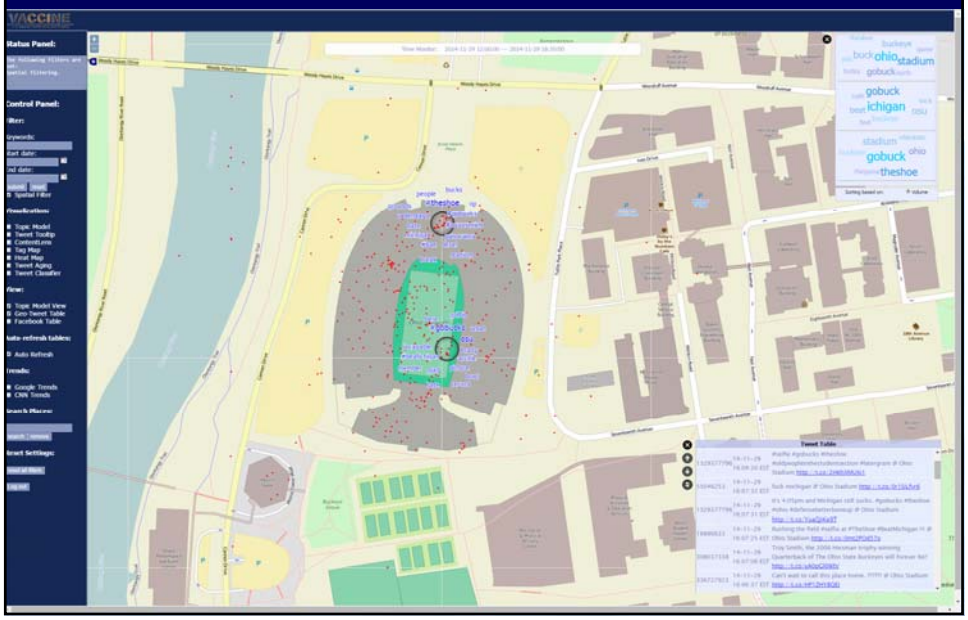
VACCINE

April 2016

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007



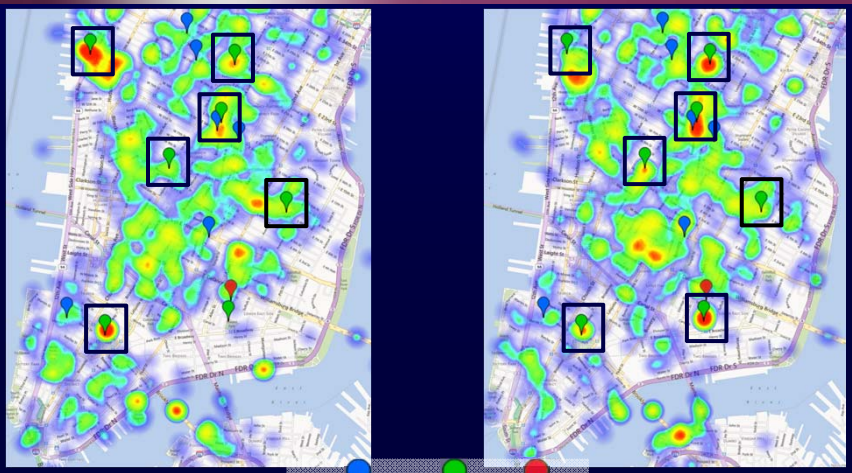
Example From Ohio State Football 11/29/14



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



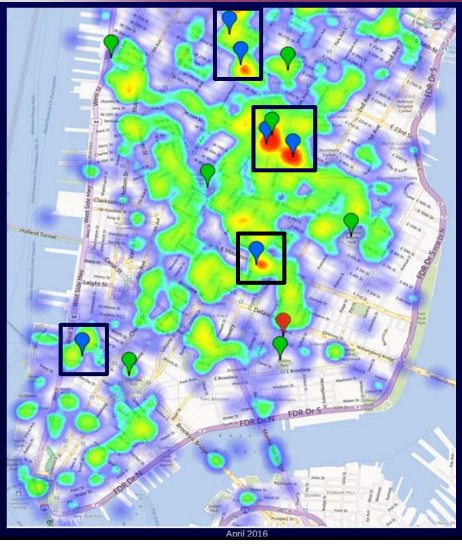
VACCINE

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

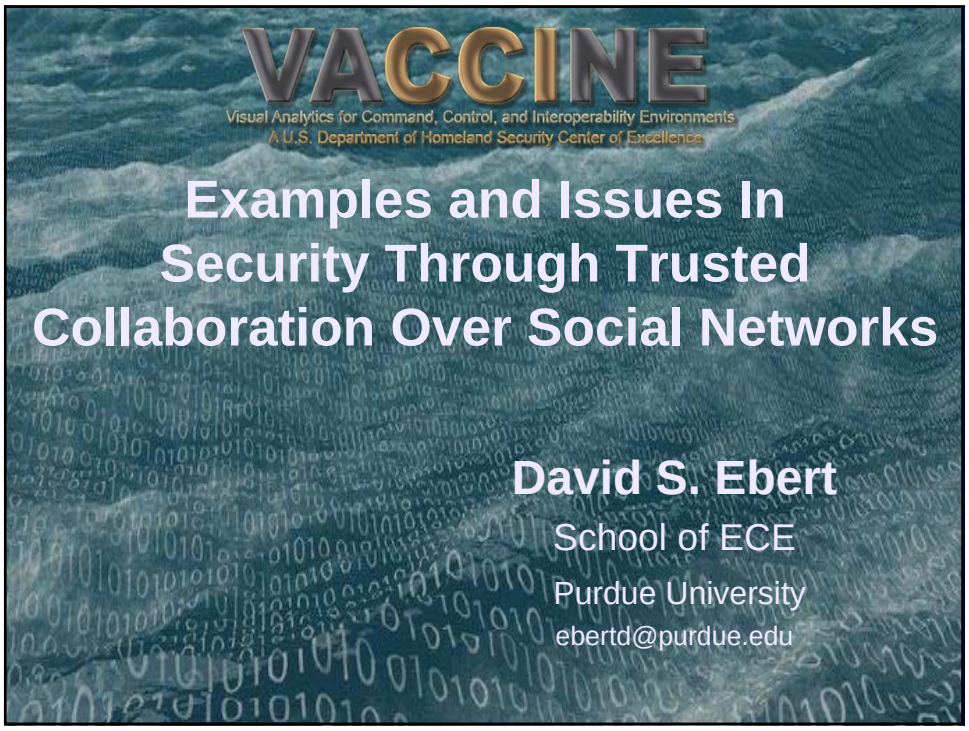


Supermarket

Park

Shelter

VACCINE



Introduction

- Why am I here?
- When can social media help?
- Tools & opportunities
- Panel questions

I Direct a U.S. Department of Homeland Security Center of Excellence Looking at These Issues

VACCINE

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

December 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 Washington
- Arizona State University
- Bethune-Cookman University
- Jackson 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

December 2015

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





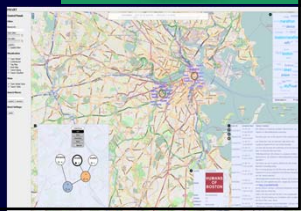
VACCINE

December 2015

How Can LEAs Benefit From Social Media?

- Educate larger audiences on crime prevention tactics and helpful tips
- Promote awareness of important and relevant issues
- Promote awareness of available resources
- Foster a partnership with the community
- Engage with neighborhood watch groups
- Promote crime prevention events & activities
- Establish trust from the community
- Improve situational awareness, crime prevention and tactical response

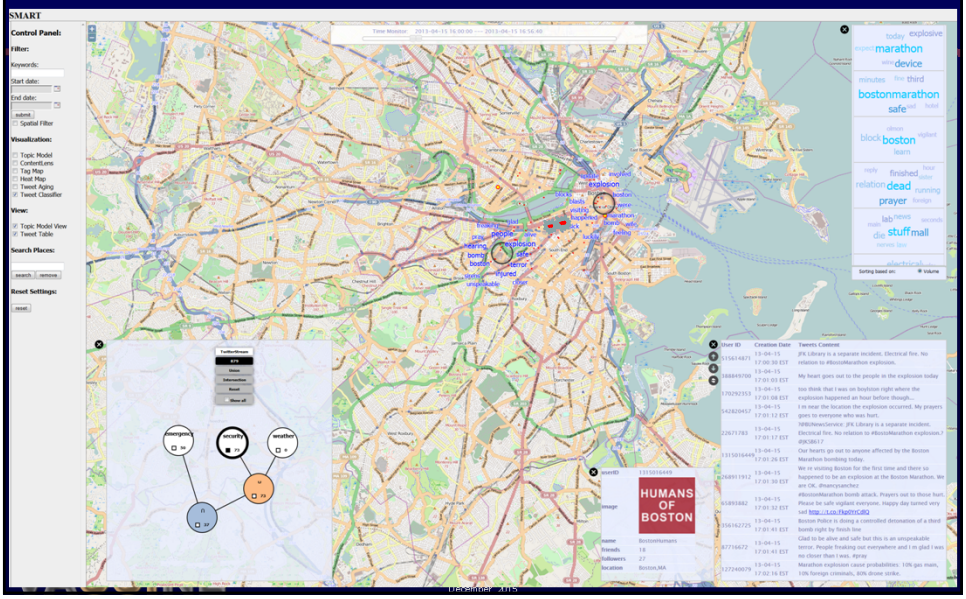




VACCINE

December 2015

Social Media Analytics and Reporting Toolkit (SMART)



Social Media Analytics 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

December 2015

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

December 2015

Anomaly Detection Superbowl Weekend



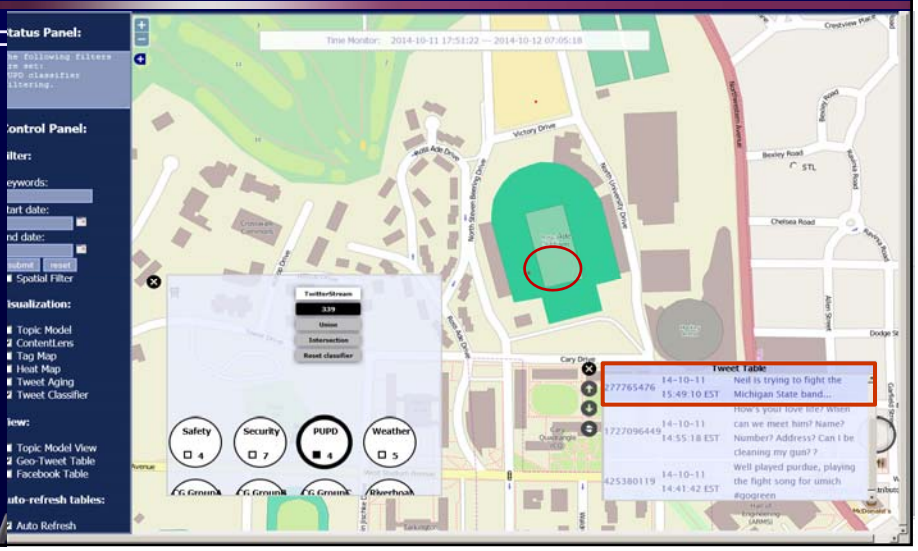
December 2015

Where Can It Be Helpful?

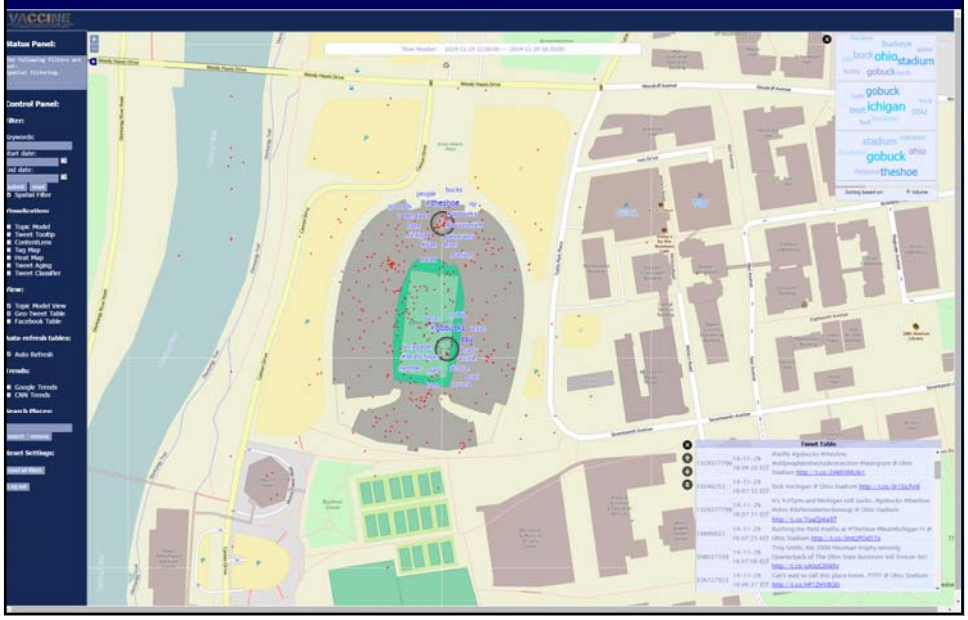
- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Example From Purdue Football 10/12/14

- F



Example From Ohio State Football 11/29/14



Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

December 2015

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day

- Umpqua Community College Shooting



VACCINE

December 2015

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day



kullgreni
Roseburg, Oregon
FOLLOW

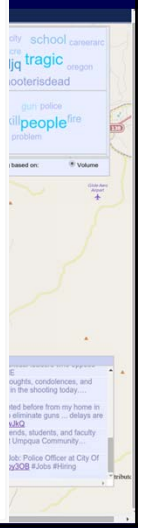
20 likes 3d

kullgreni Prayer chain for UCC shooting victims. The end of a long, tragic day.

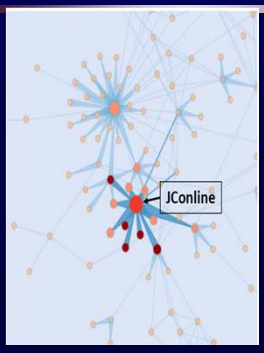
thadizzo84 😔😔😔

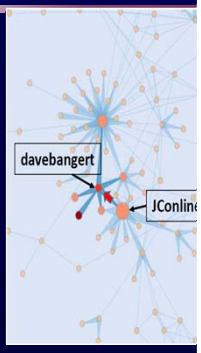
sgonzalez910 Such a tragic thing to cover. Stay strong

Log in to like or comment.



Tracking Information Diffusion and Identifying Influential Users Through User [illegible] Shooting Event)







JConline is a local news media that has a lot

JConline retweeted from **davebangert** about the updates of the shooting event.

davebangert retweeted from **justinlmack**.

justinlmack posted first-hand updates about the event.

VACCINE

December 2015

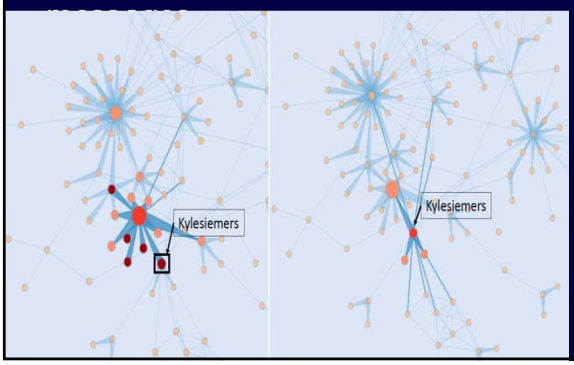
Identify Influential Users In Purdue Shooting

Click **JConline** to show its information traf

darker nodes along the path: **Jconline** ->

Kylesiemers has no darker nodes connec

Kylesiemers as an influential node and m





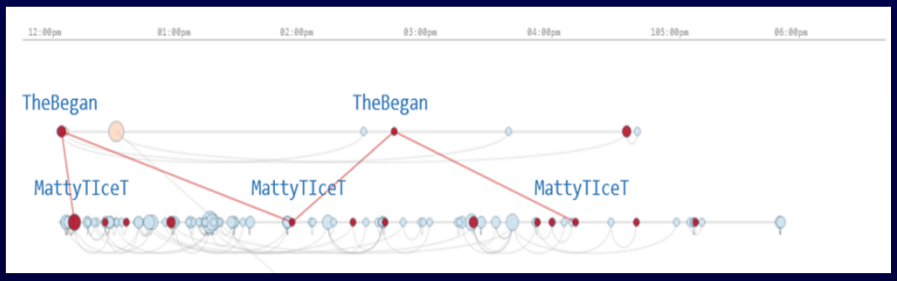
Monitoring Network Evolutions Over Time

Initially, there is a single network containing ‘purdue’, ‘shooting’, ‘custody’. As time passes, the network evolves and new nodes come. The network then splits to show two user groups with slightly different topics.



A timeline of the conversation between **TheBegan** and **MattyTIceT** about the shooting.

@MattyTIceT: White male grey sweatshirt running by pych building is description of a suspect that just came over the scanner.



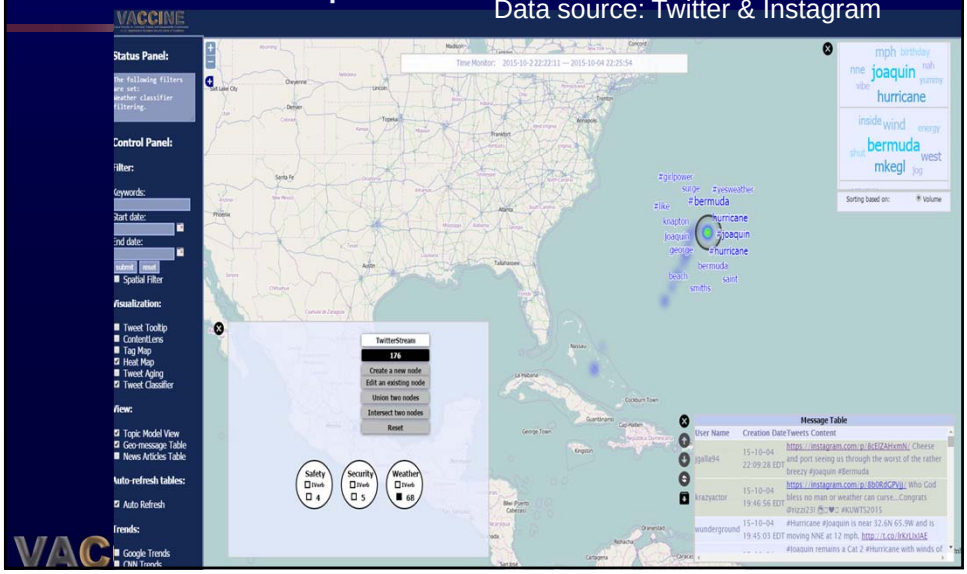
RT @TheBegan: Groupme texts are saying a TA has been shot in EE.

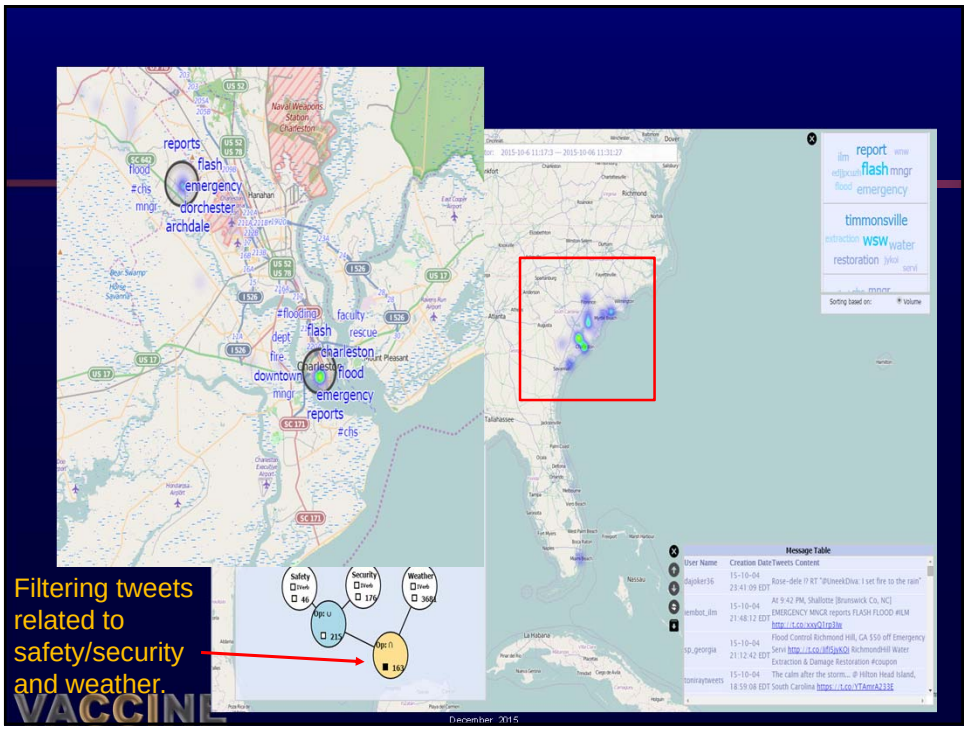
@TheBegan according to the police scanner they have one person in custody

@TheBegan WTHR is airing the audio from the press conference

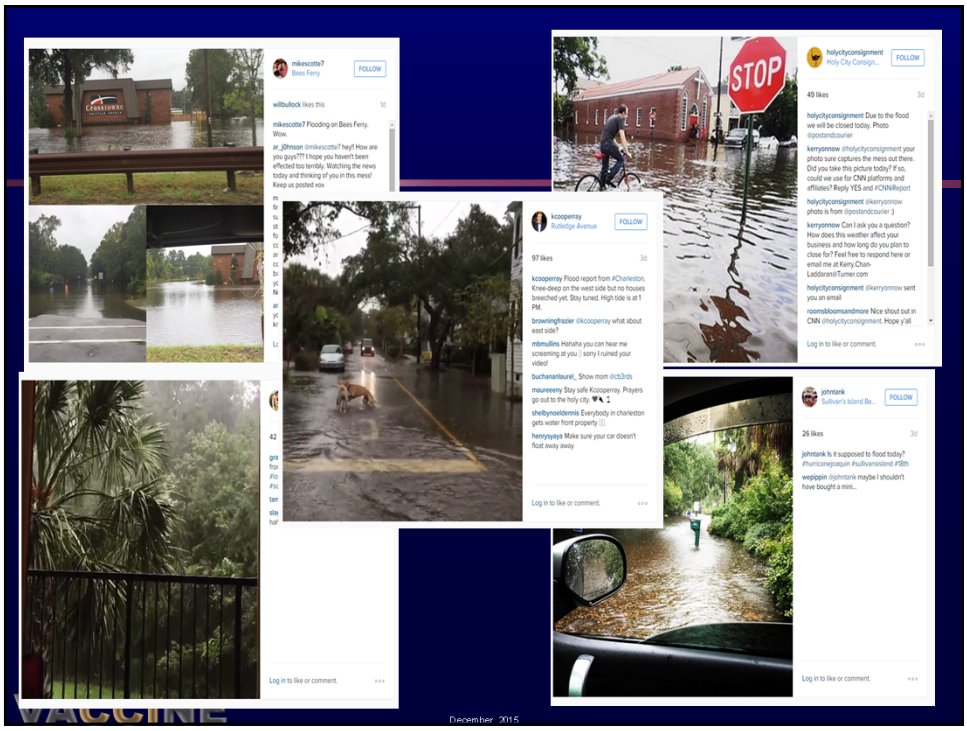
Hurricane Joaquin

Time range: 2015-10-03 to 2015-10-04
Location: Southeast US
Data source: Twitter & Instagram





Filtering tweets related to safety/security and weather.



Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE December 2015

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



Panel Discussion

- Trust:
  - Do they believe you?
  - Can you believe them?
- Reliability: Can we count on this information?
- Information diffusion:
  - Retweets vs. mass interest. How do we tell?
- Crowd Informing vs. Crowd Sourcing:
  - Can we reliably do both?
- Privacy:
  - Will people stop sharing so much information?
  - What is the correct boundary for privacy?

VACCINE

December 2015

VACCINE

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

From Big Data to Data-Driven Policy and Decision Making:

Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
School of ECE
Purdue University
ebertd@purdue.edu

Overview

- Background Visualization vs. Visual Analytics
- Big Data and Visual Analytics
- What Visual Analytics can do (examples)
- Directions for the future



"Remember, the other team is counting on Big Data insights based on previous games. So, kick the ball with your other foot."

Visualization Through the Ages

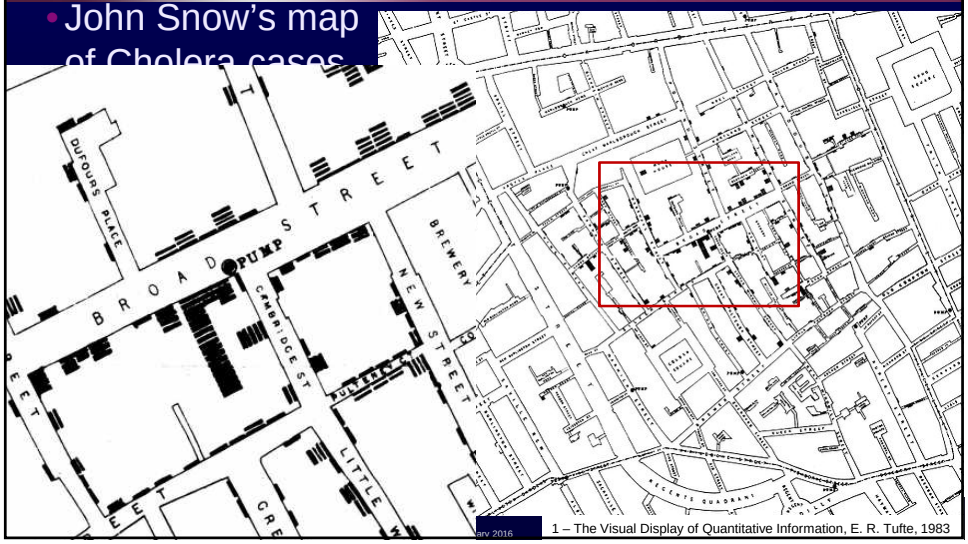
- Hand-drawn illustration[1] of water by Leonardo da Vinci from his studies to determine the processes underlying water flow (1510)



Leonardo da Vinci, Old Man Seated on Rocky Outcrop, Seen in Profile to the Right, with Water Studies, Windsor Castle, Royal Library

Visualization Through the Ages

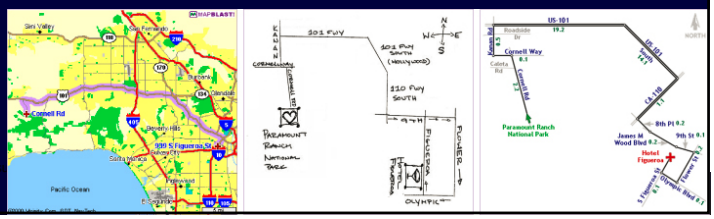
- John Snow's map of Cholera cases



1 – The Visual Display of Quantitative Information, E. R. Tufte, 1983

What is Visualization?

- "The use of computer-supported, interactive visual representations of data to amplify cognition."[1]
- Not simply making a graphic or an image, the goal is to create insight, utilizing cognitive and perceptual principles
- Help people form a mental image of something and internalize their own understanding
- Promote discovery, decision making and explanations



Agrawala, Stolte: Rendering effective route maps: improving usability through generalization. SIGGRAPH 2001: 241-249

VACCINE

Tasks for Visualization

- Searching and browsing - find a specific piece of information, inspect data, seek information
- Analyze – do comparisons, differences, look for outliers, extrema and patterns
- Monitor – look for changes and trends

How to Harness Visual Representations

- Visual representations translate data into a visible form, highlighting important features, such as commonalities and anomalies
- Visual representations make it easy for users to perceive salient aspects of their data quickly
- These visual representations augment the cognitive reasoning process with perceptual reasoning that enhances the analytical reasoning process

1 – J. J. Thomas and K. A. Cook (Ed.). *Illuminating the Path: The R&D Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

The Power of Visualization

- Goal: Transform data into information
- How many terabytes of data doesn't matter, it's how many 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
  - Visualization is key!

Scientific Visualization

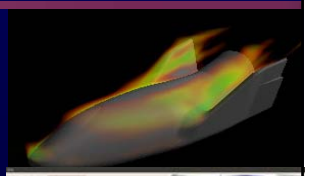
- Primarily relates to and represents something physical or geometrics
- The structure of the data is typically defined or given
- Examples
  - Air flow over a wing
  - Stresses on a girder
  - Organs in the human body
  - Molecular bonding

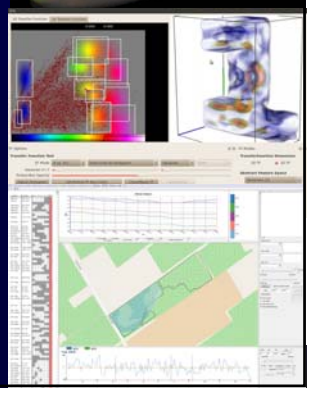
1 – Descriptions on this slide are borrowed from John Stasko's "InfoVis Overview" lecture: http://www.cc.gatech.edu/~stasko/7450/Notes/overview.pdf

Scientific Discovery

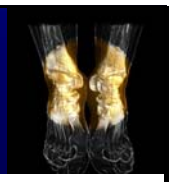


- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality

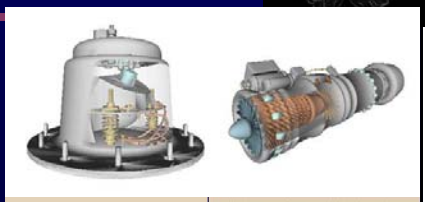


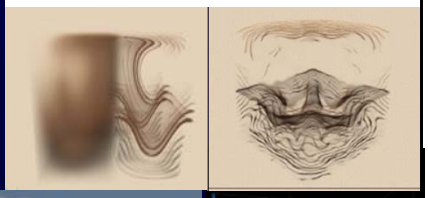


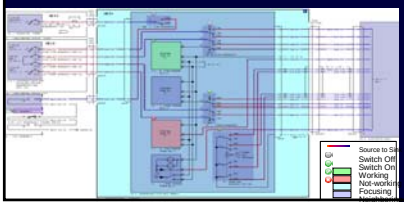
Scientific Discovery

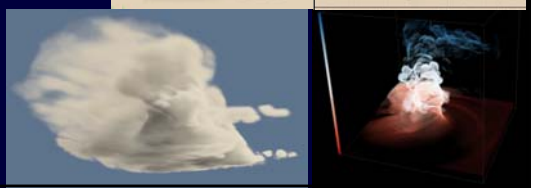


- Illustrative Visualization
  - Mechanical assemblies
  - Schematics and wiring diagrams
  - Illustrative flow visualization
  - Illustrative medical visualization
- Interactive multivariate atmospheric science data visual analanalytics









Information Visualization

- Primarily relates to data that does not have a direct physical correspondence
- Notion of the data is abstract
- Examples
  - Baseball statistics
  - Stock trends
  - My social network

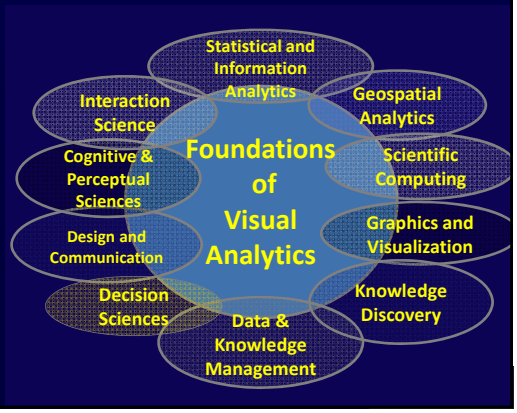
1- Descriptions on this slide are borrowed from John Stasko's "InfoVis Overview" lecture: http://www.cc.gatech.edu/~stasko/7450/Notes/overview.pdf

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

Essential Features for Visual Analytics

- Interactive, integrated discovery environment
- User-guided and perceptually-guided
- Balance of human cognition and automated computerized analysis to amplify human cognition
- Enable discovery and decision making
- Provide quantitative, reliable, reproducible evidence

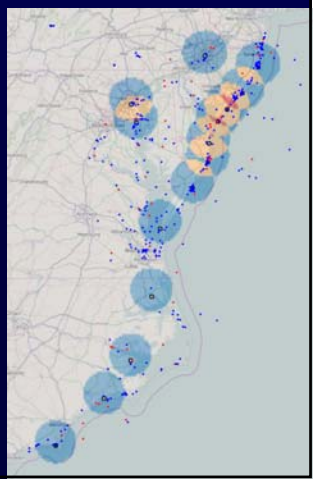
VACCINE

January 2016

TACC

Visualization, Data, and Decision Making

- At some level most decision making is driven by data
- What role does visual representation play?
- Good or bad – people believe charts and graphics – “seeing is believing”



VACCINE

January 2016

Big Data

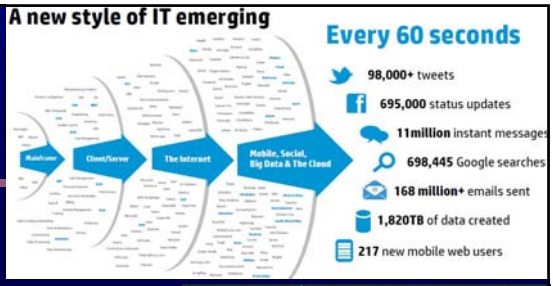
30 ZETTABYTES — The amount of data that will be stored in the world by 2020.
One zettabyte is the equivalent of streaming 36 million years of HDTV video
SOURCE: IDC

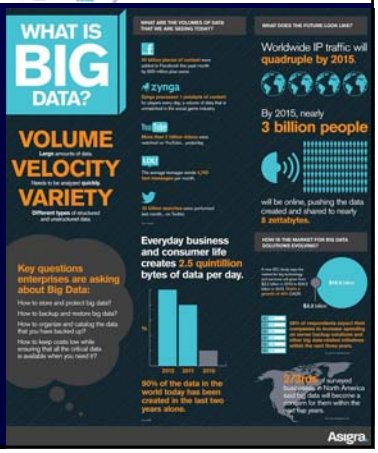
10 TERABYTES — The amount of data Facebook generates daily.

7 TERABYTES — The amount of data Twitter generates daily.
One terabyte would hold about 250,000 songs, or nearly a year's worth of continuous music
SOURCE: IBM

40 PERCENT — The annual growth rate of global data generated per year.
SOURCE: MCKINSEY

10 BILLION+ — The number of mobile-connected computing devices in use by 2016, up from less than 7 billion in 2012.
SOURCE: CISCO





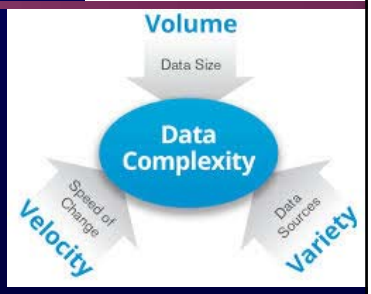
January 2016

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**



VACCINE

January 2016

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

VACCINE

January 2016

TACC

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

VACCINE

January 2016

TACC

What Our Visual Analytic Solutions Offer Decision Makers

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



Visual Analytics for Decision Making

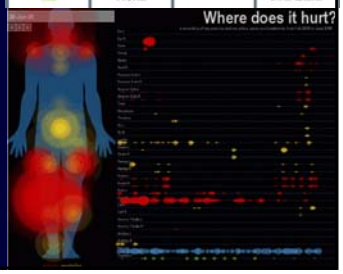
- Personal level
- University level
- Community level
- State and regional level
- National and global scale

Visual Analytics for Effective Decision Making: Personal to Global

- Visual analytics can help people understand data in their everyday lives







University Scale Decision Making

- Football game force deployment using SMART
- Student safety
  - Reducing high-risk alcohol behavior
  - Camera locations and safe paths
- Personal safety across campus
- Enabling scientists and engineers

VACCINE

January 2016

TACC

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

SMART: Social Media Analytics and Reporting Tool



January 2016

Social Media Analytics and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

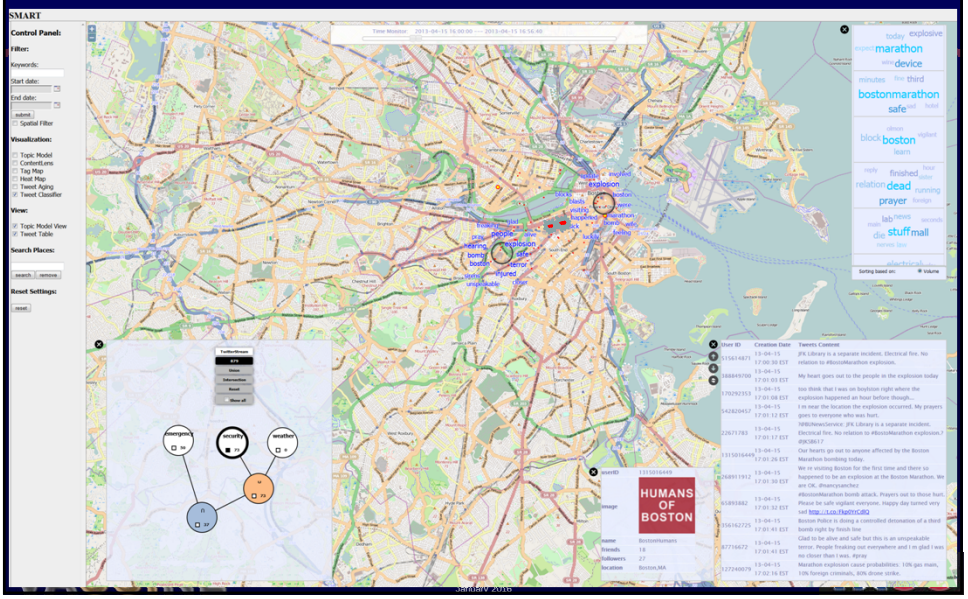
- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games 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

January 2016

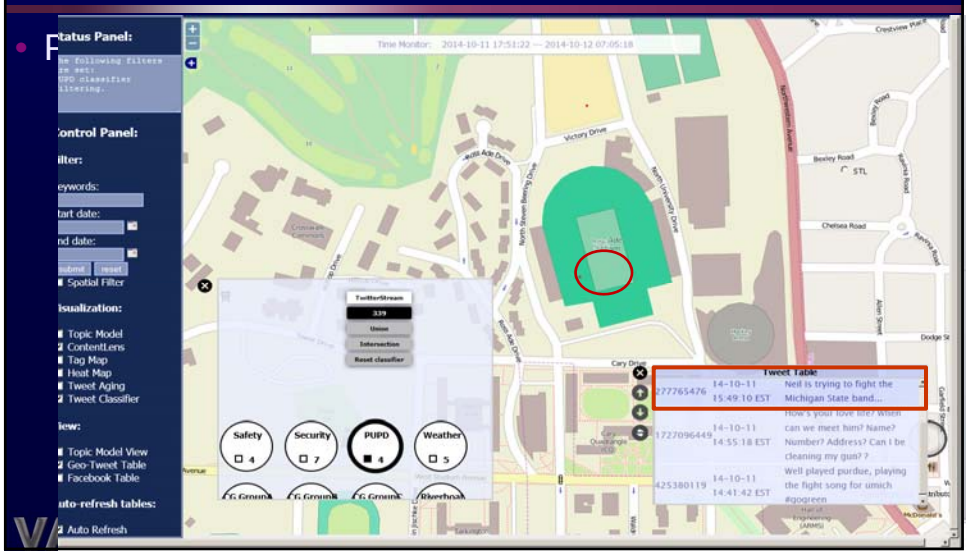
Social Media Analytics and Reporting Toolkit (SMART)



January 2016

Example From Purdue Football 10/12/14

- F



Example From Ohio State Football 11/29/14



Reducing High-Risk Alcohol Behavior

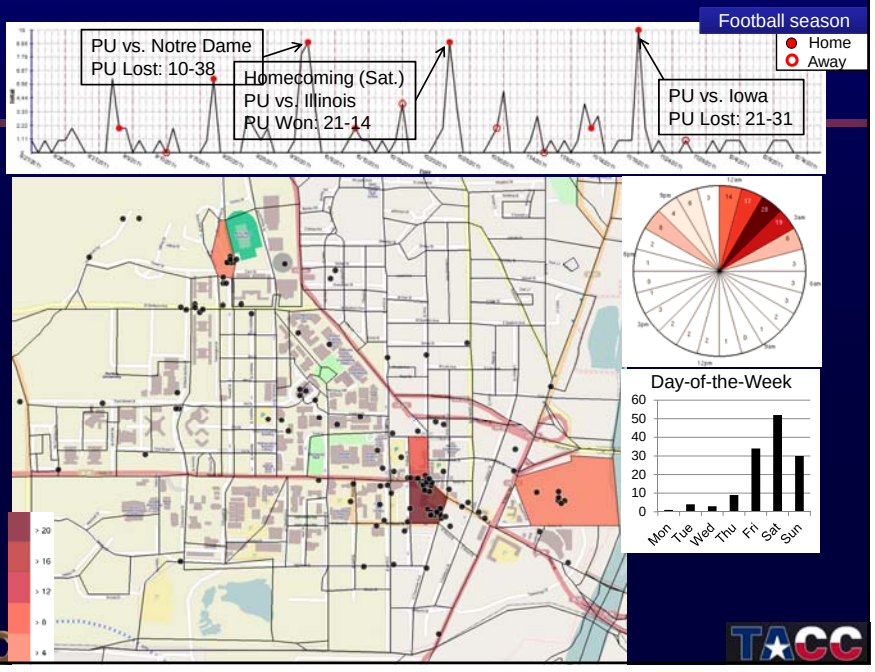
- Student Wellness Center Partnership
- Find areas and patterns of high-risk alcohol behavior
- Information campaign to students and dorms
- Increased patrolling

VACCINE

January 2016

TACC

High Risk Behavior : Drunkenness / Public Intoxication



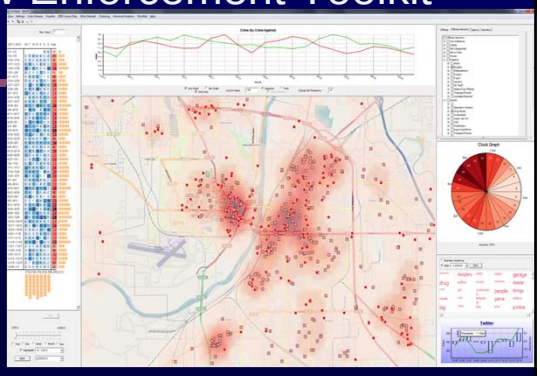
VAC

TACC

Community Scale Decision Making

- Public safety – crime analytics (VALET)
- Transit safety – eBOLO

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)



Impacts:

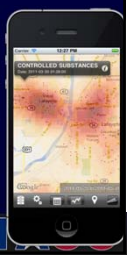
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive 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:

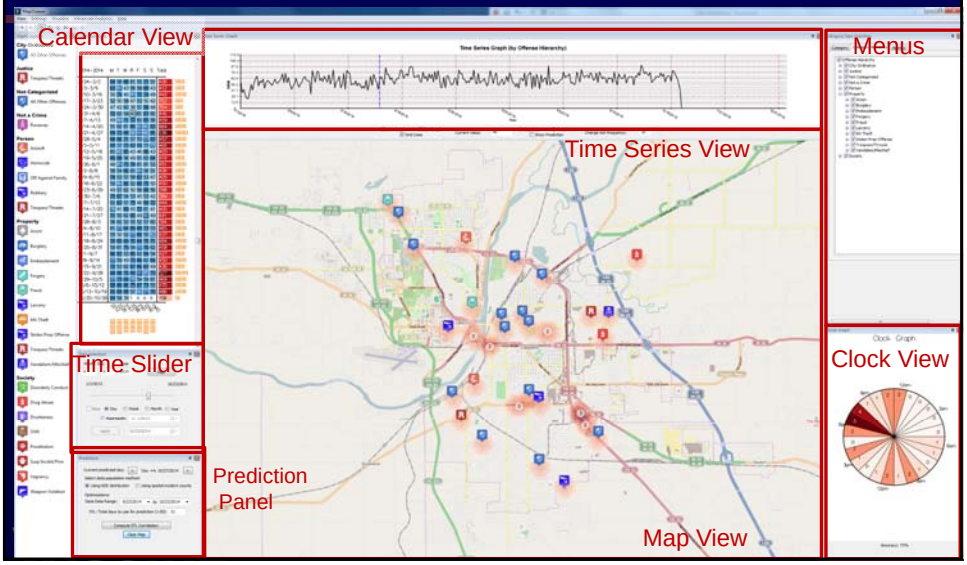
- October 2011: Purdue, WL Police, Lafayette PD



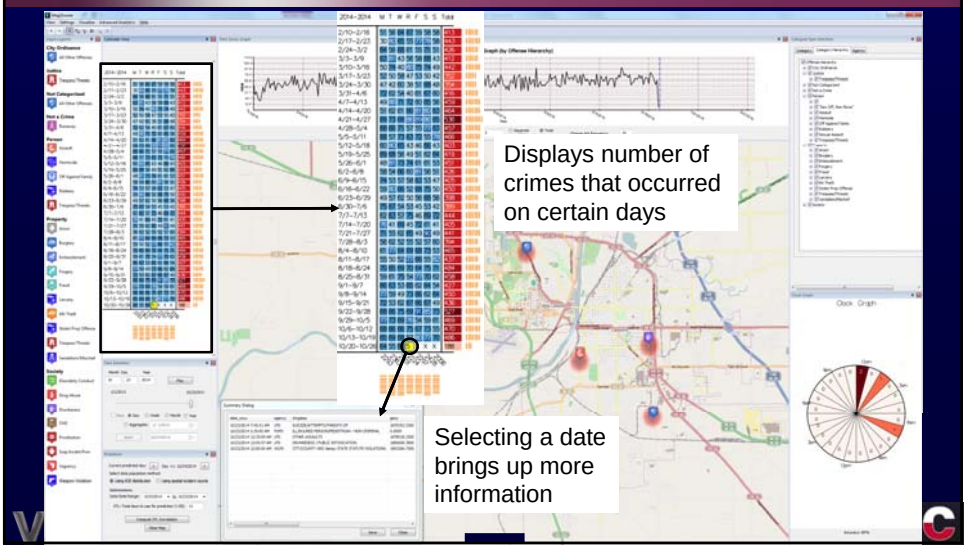
Example Projects: Public Safety and Law Enforcement Applications



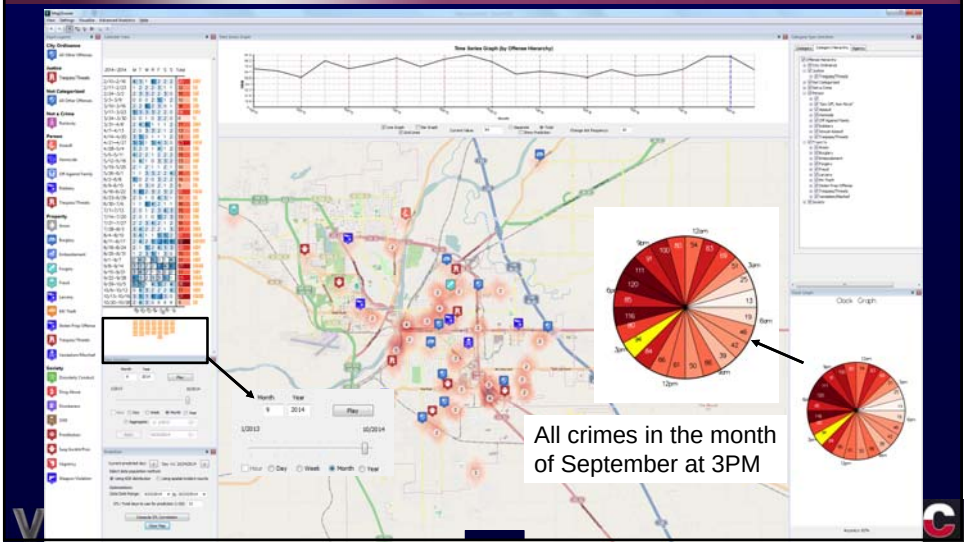
VALET Layout

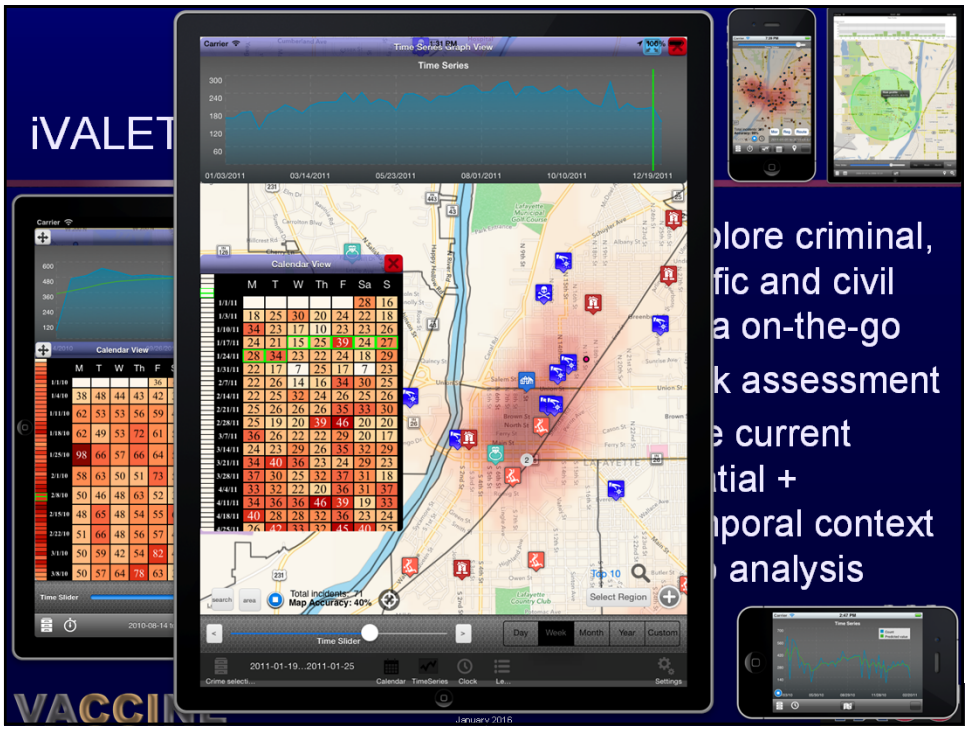


Calendar Display



Crimes During Certain Hours In A Month





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**

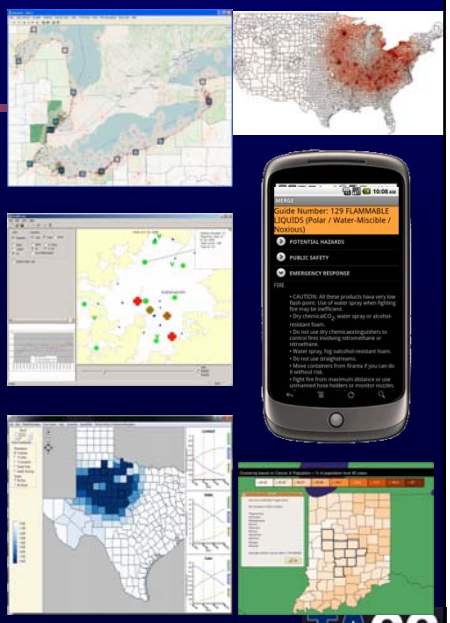
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

State and Regional Level Decision Making

- Public Health
- Disaster Preparedness and Response
- VASA

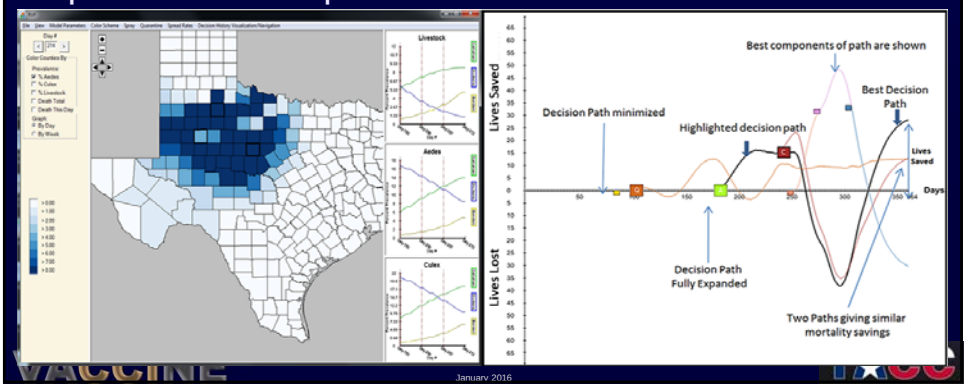
Public Safety & Health

- 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

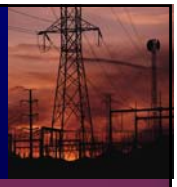


Complex Decision Making: Advanced Decision Support Tools (e.g., Rift Valley Fever)

- Integrated simulation models
- Explore alternative courses of action in decision space and real space



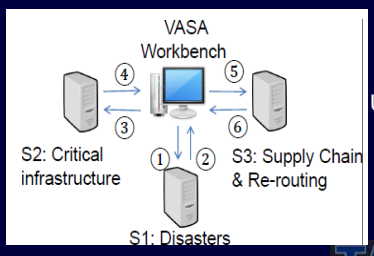
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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

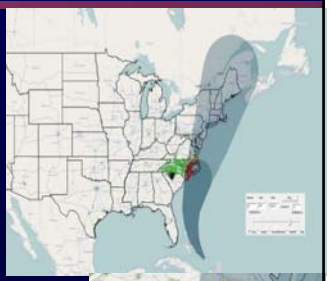
TACC

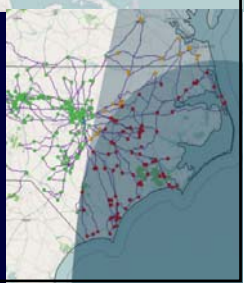
VASA: Visual Analytics for Simulation-Based Action

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 simulation 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

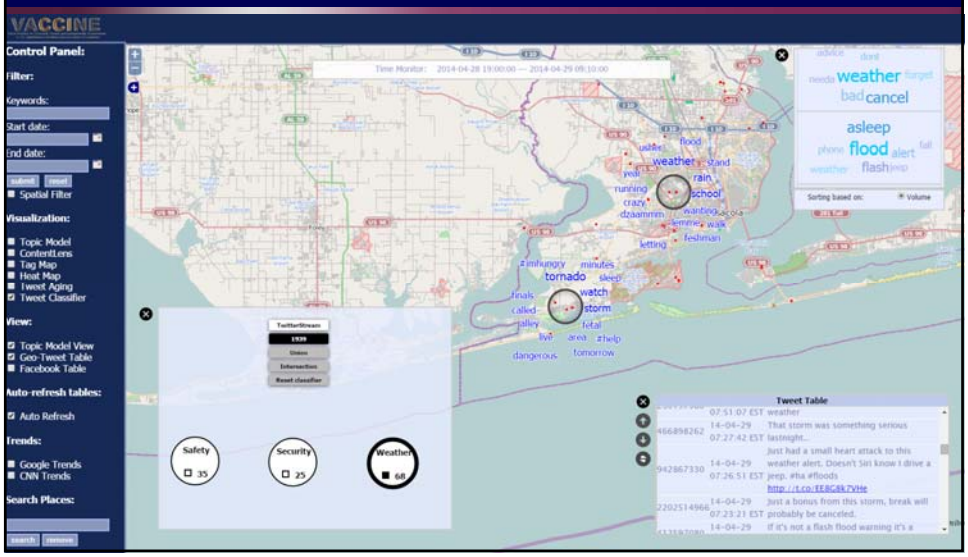




VACCINE

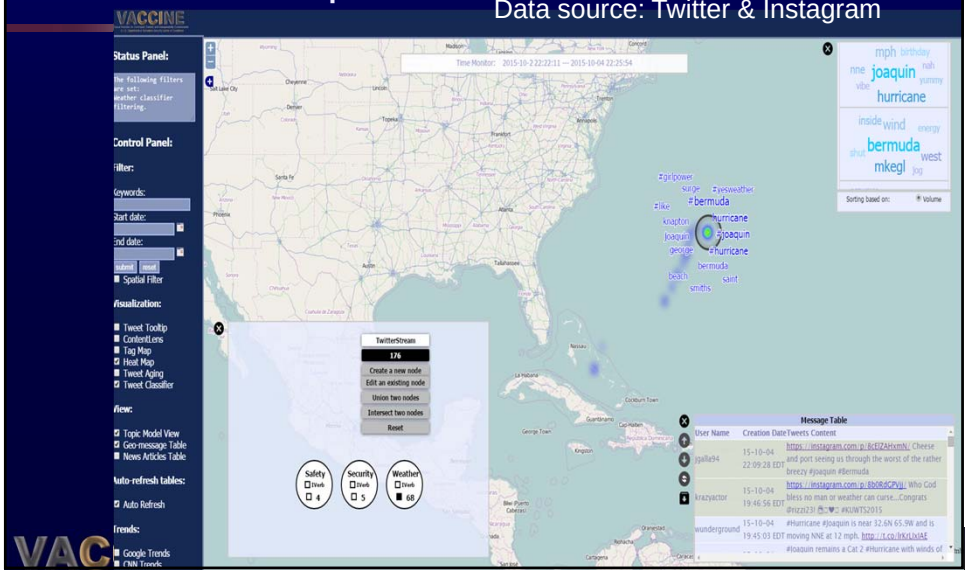
January 2016

SMART: Disaster Preparedness & Response

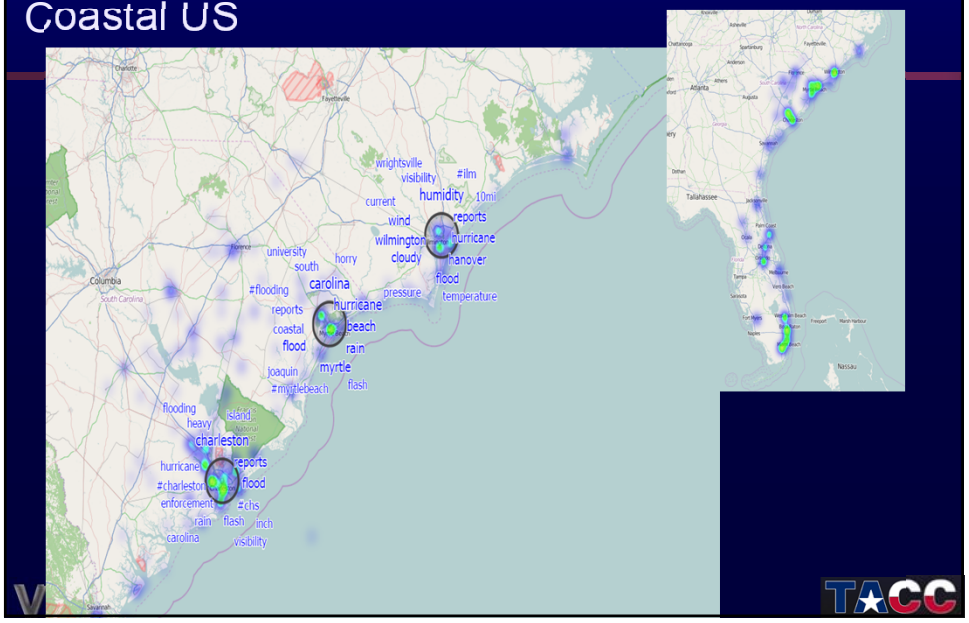


Hurricane Joaquin

Time range: 2015-10-03 to 2015-10-04
Location: Southeast US
Data source: Twitter & Instagram



Weather-related tweets in Southeast Coastal US



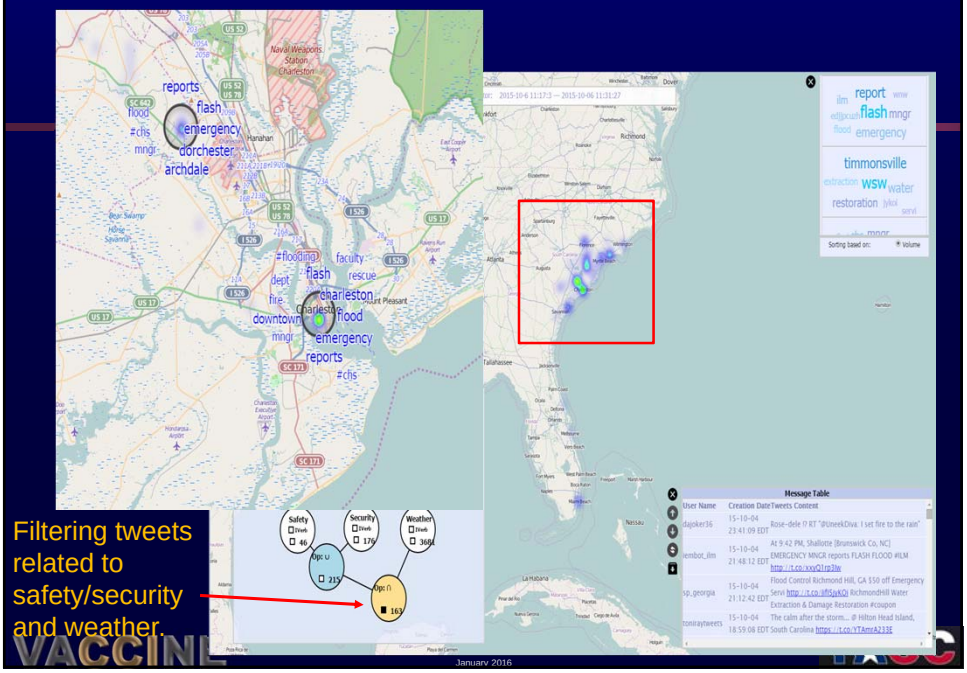
Top Topics related to Weather Classifier



VACCINE

January 2016

TACC

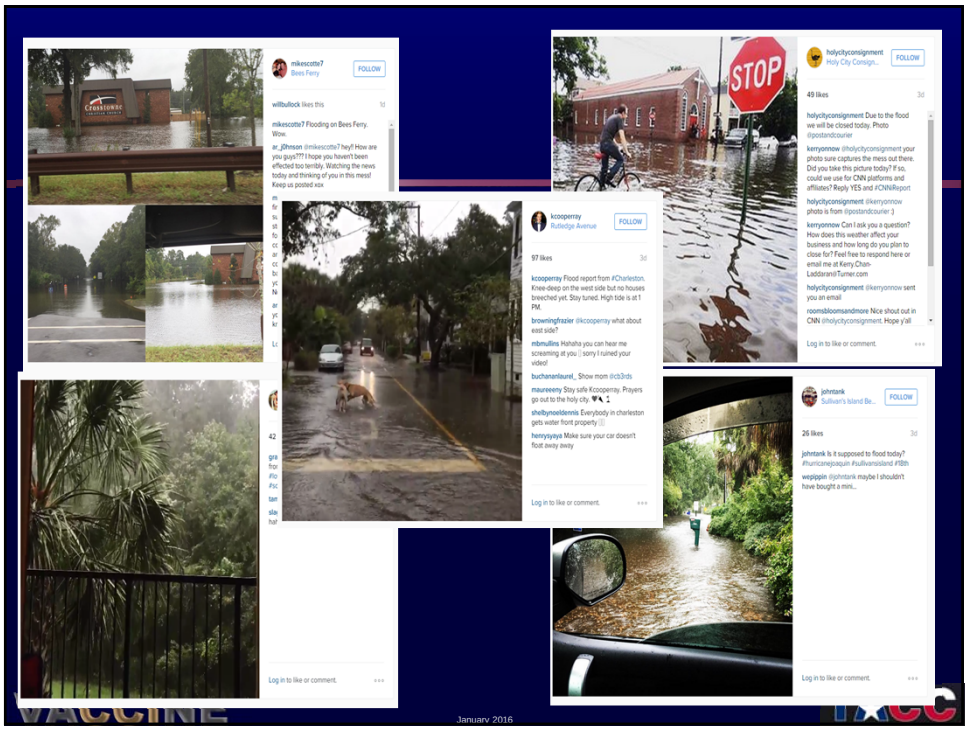


Filtering tweets related to safety/security and weather.

VACCINE

January 2016

TACC



Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - 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

National / Global

- USCG
- Southern Border Approaches
- Precision agriculture and droughts
- Soundscapes

VACCINE

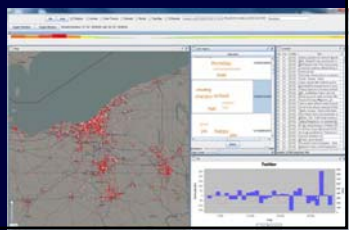
January 2016

TACC

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



VACCINE

January 2016

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

January 2016

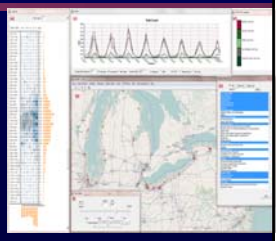
U.S. Coast Guard Search and Rescue VA

(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



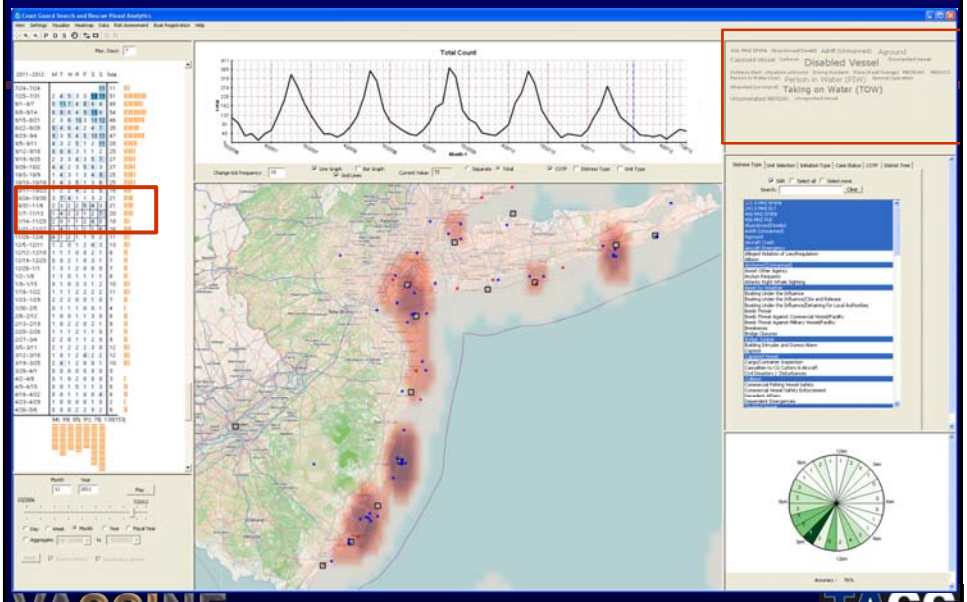


VACCINE

January 2016

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area

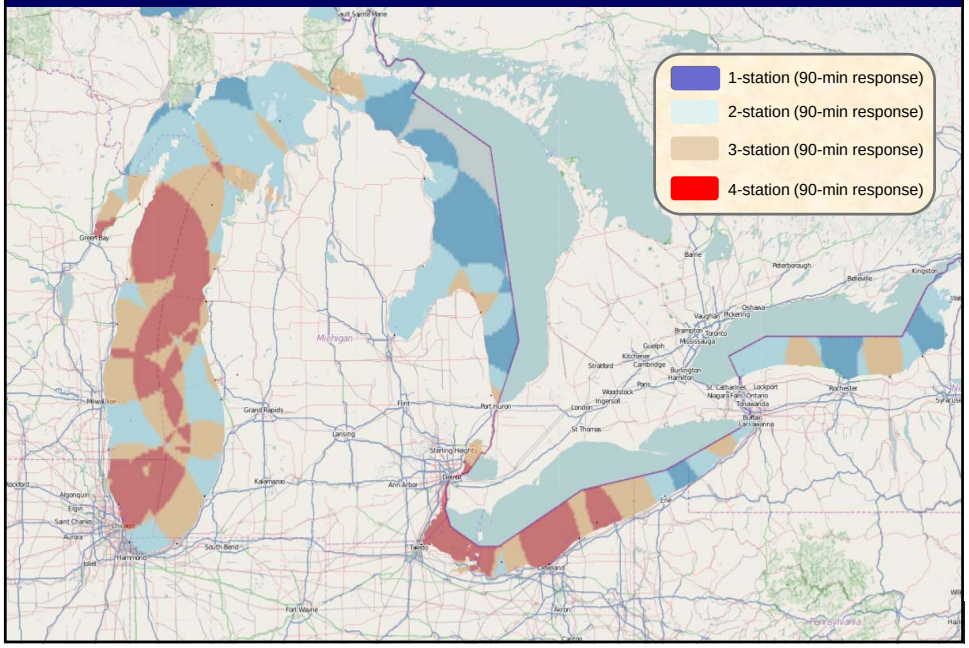


VACCINE

January 2016

TACC

Response Efficiency – Potential Future Assets



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

Southern Border Approaches

- Department of Homeland Security has created joint task forces to address Southern Border Approaches security (CBP, ICE, US Coast Guard)
- Joint Task Forces have asked us to assist in visualizing data and metrics to help define performance and evaluate success

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
- Incorporate analytics and data mining, but let user guide and refine process
  - Understandability of why is the key

- *If you have a problem…*
  we can develop a solution!

VACCINE January 2016 TACC

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

2016 Visual Analytics MSI Faculty Training

Welcome!

David S. Ebert
Purdue University
ebertd@purdue.edu

May 2016

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

May 2016

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





VACCINE

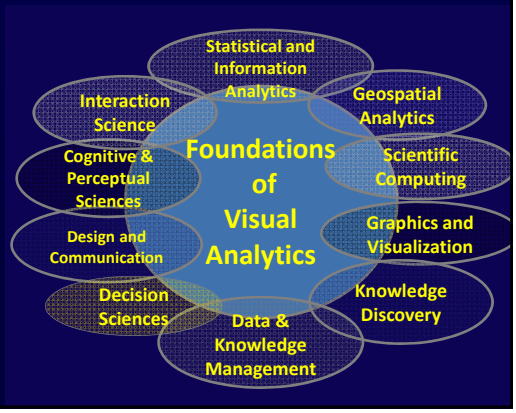
May 2016

Why are We Here? 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

The Depth of the Research Challenges

- Long-term fundamental challenges!
  - Mobility; scale; multisource, cross-media analysis; numerical to visual analysis
- How do we effectively integrate and adapt solutions to the
  - User
  - Task
  - Environment
  - Interactive discovery environment
  - Data



VACCINE

May 2016

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Human-Computer Collaborative Decision Making



David S. Ebert

Purdue University

ebertd@purdue.edu

May 2016

Why VA and HCCD For Big Data?

(Or What Big Data Analytics Can't Do) [1]

- 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

VACCINE

[1] Inspired and adapted from David Brooks, New York Times, 2/18/2013)

May 2016

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

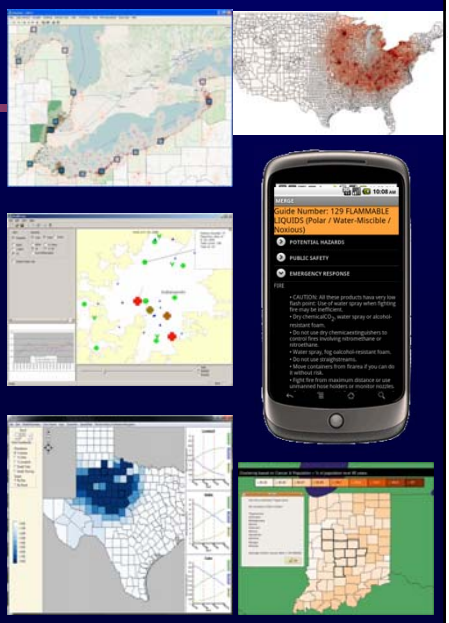
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

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

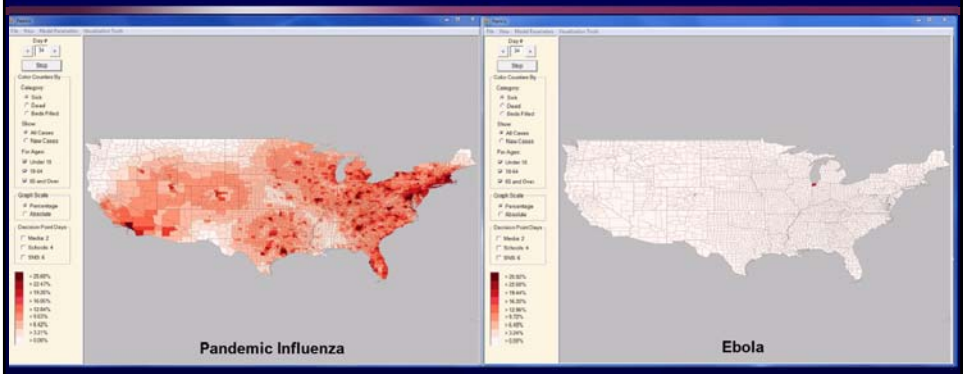


VACCINE

May 2016

Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL

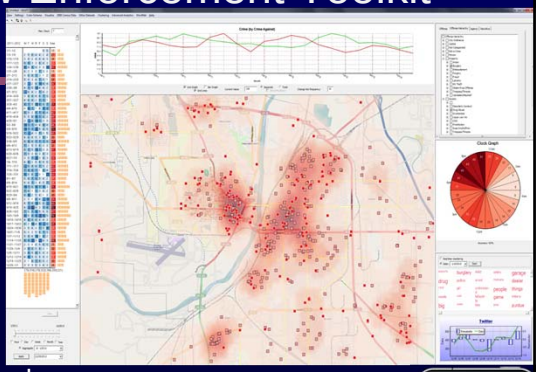


Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

VACCINE

May 2016

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

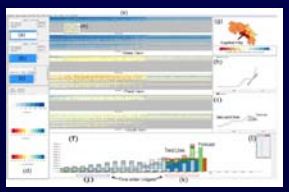


VACCINE

May 2016

Business Visual Analytics





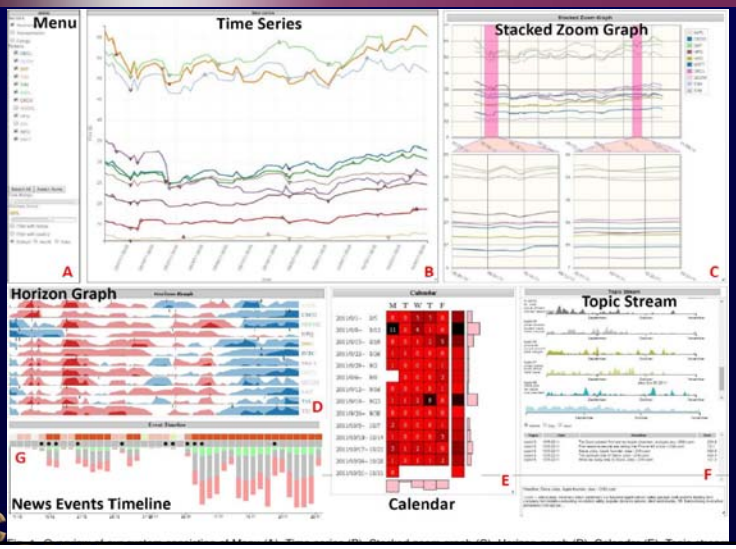


- Risk-based decision making and resource allocation
  - Coast guard operational risk assessment model
  - Helping to prioritize efforts to minimize risk
- Competitive Intelligence
  - Visual analytics system for business intelligence
- EconVIS
  - Visual analytics in various economic problems
  - Improving decision making and identifying key motivations in knowledge creation

VACCINE

May 2016

Financial Anomaly HCCD



VAC

VinSense TECHNOLOGY

Respect for tradition, vision for the future

Agricultural HCCD

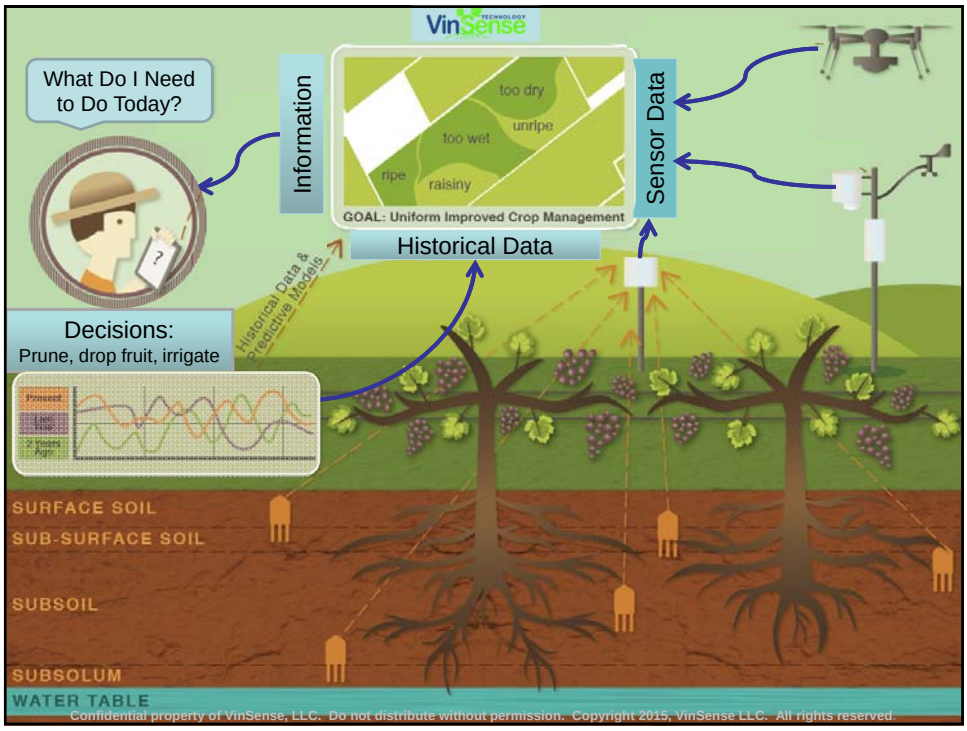
Better Decisions for Wine Grape Crop Management

Highly variable grapes grown with suboptimal decisions

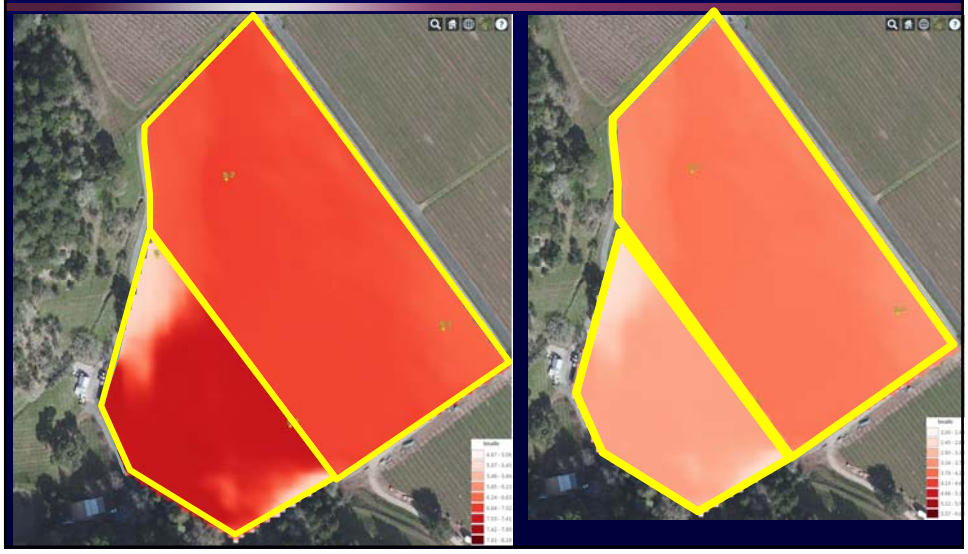


Potential uniform quality grapes grown with VinSense based decisions

May 2016



Veraison vs. Harvest Information: L-malic



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

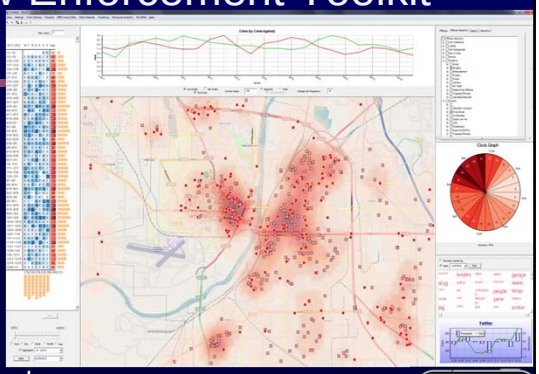
VACCINE

May 2016

Emergency Response and Public Safety Examples

May 2016

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:

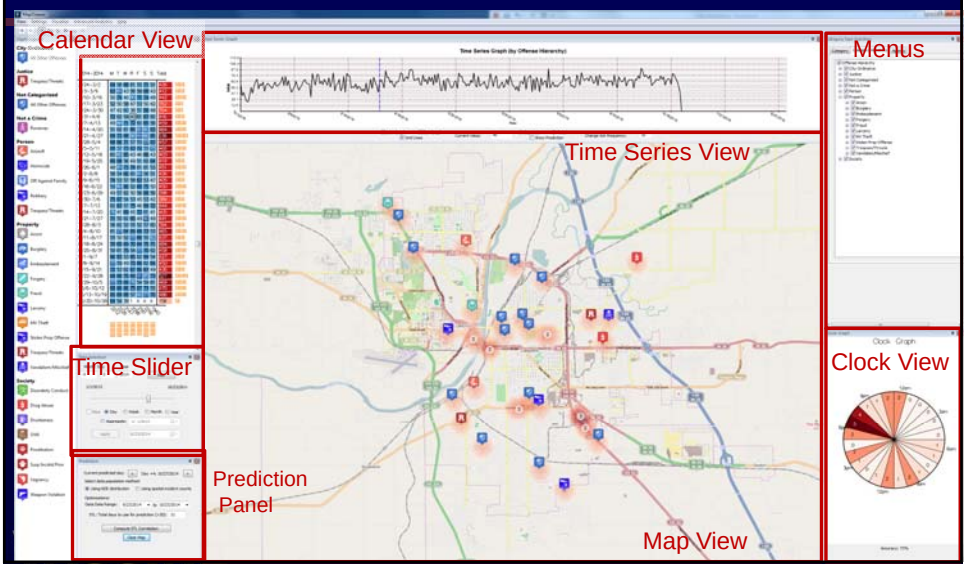
- October 2011: Purdue, WL Police, Lafayette PD



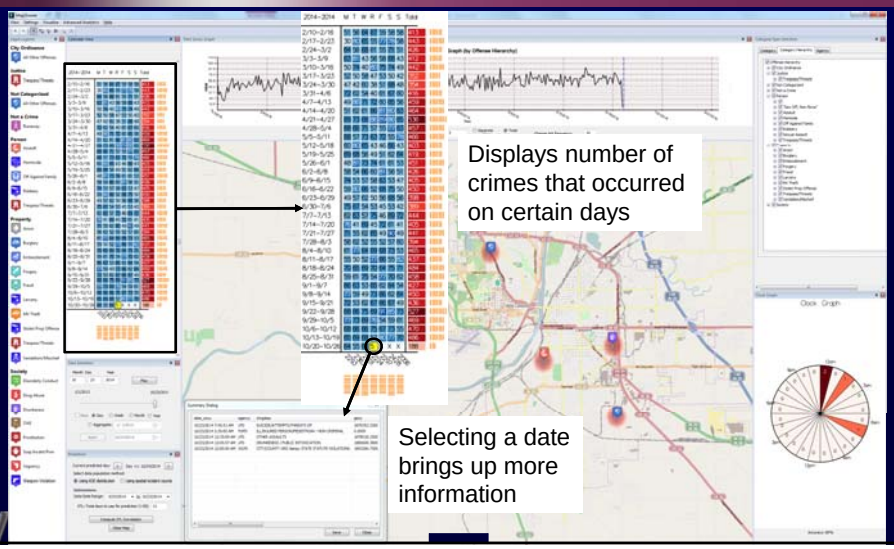
VACCINE

May 2016

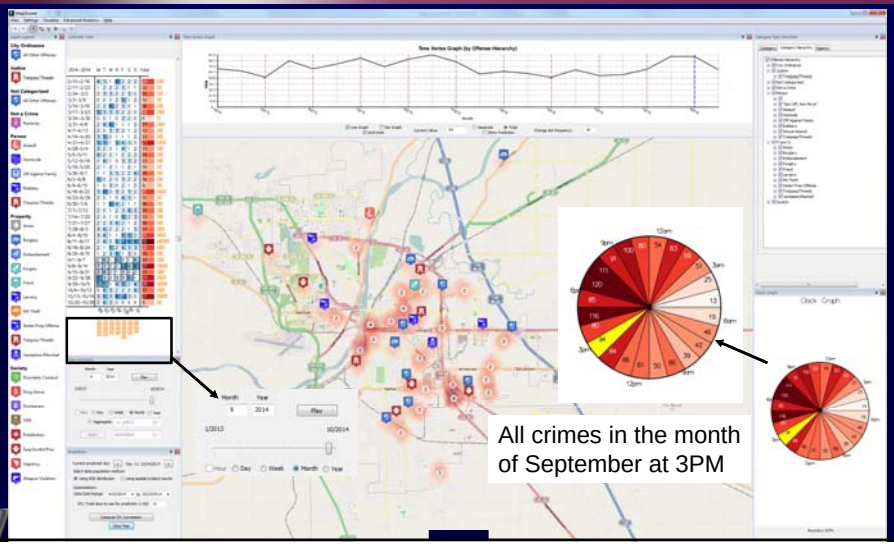
VALET Layout



Calendar Display

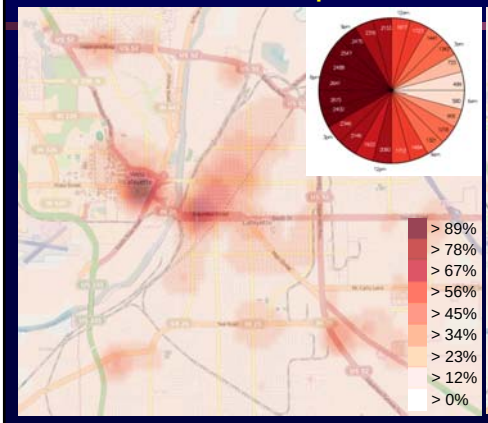


Crimes During Certain Hours In A Month

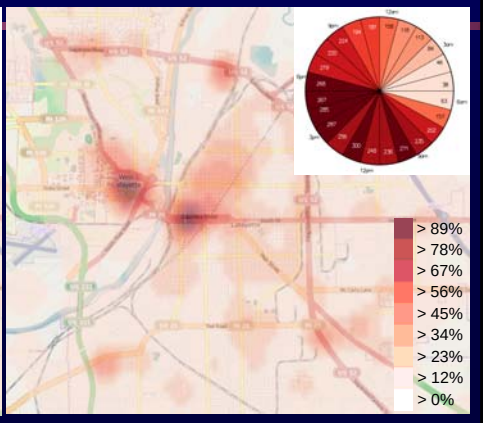


Overall Historical Distribution: Spatial and Temporal Variations

Note areas of sparse data



Past 2 Years
3/11/2012 – 3/10/2014

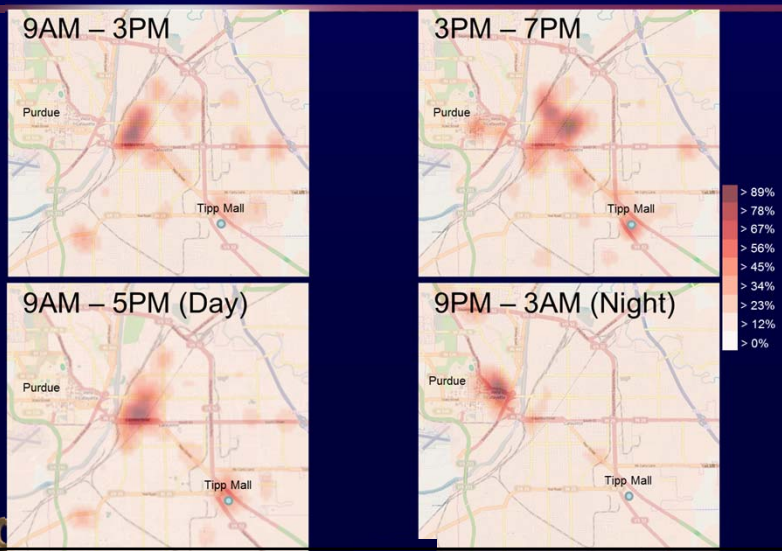


March + Tuesday Filter
(Past 10 Years)

VACCINE

May 2016

Prediction During Certain Hours In The Day

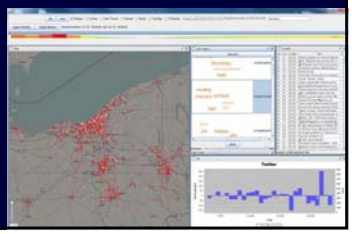


VAC

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



VACCINE

May 2016

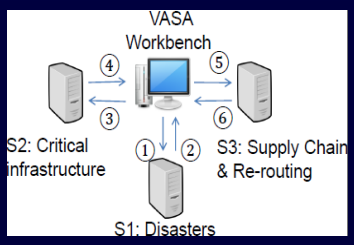
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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

May 2016

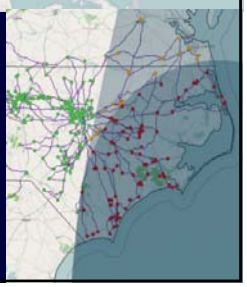
VASA: Visual Analytics for Simulation-Based Action

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

May 2016

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

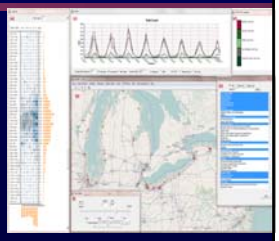
May 2016

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



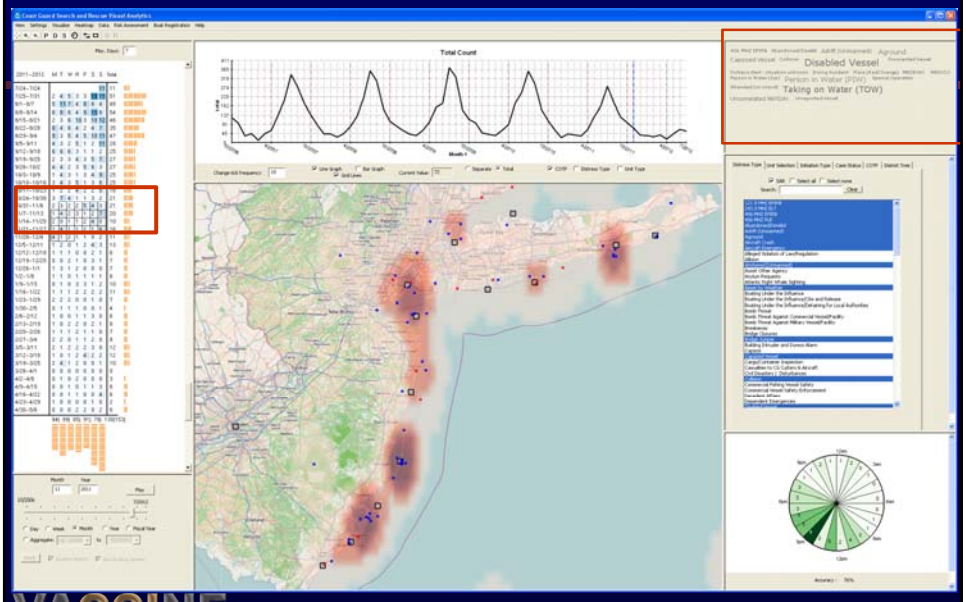


VACCINE

May 2016

Example: Risks and Consequences From Sandy:

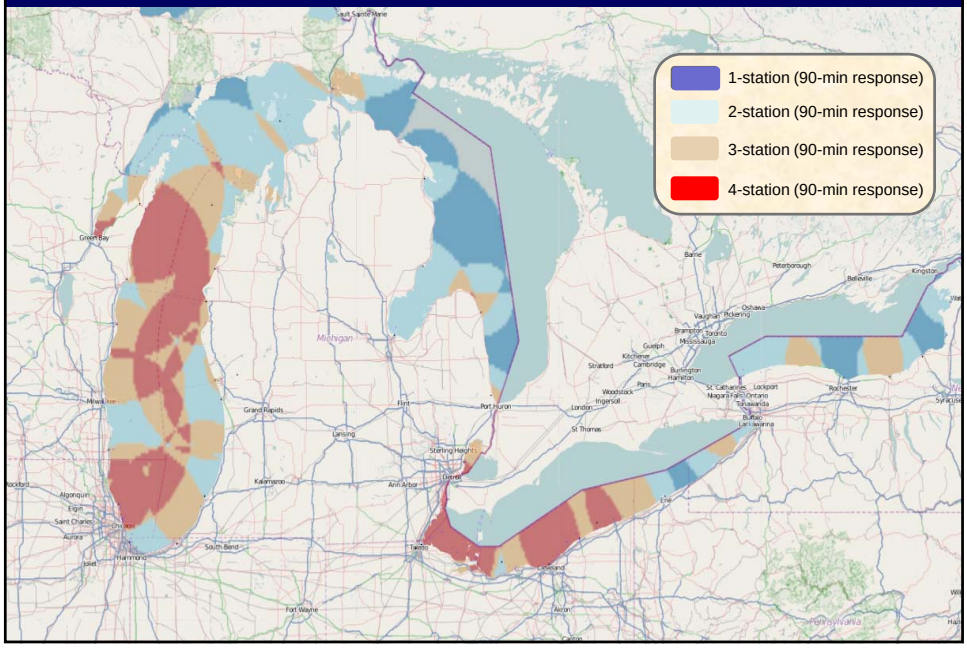
SAR Cases November 2011 NJ/NYC Area



VACCINE

May 2016

Response Efficiency – Potential Future Assets



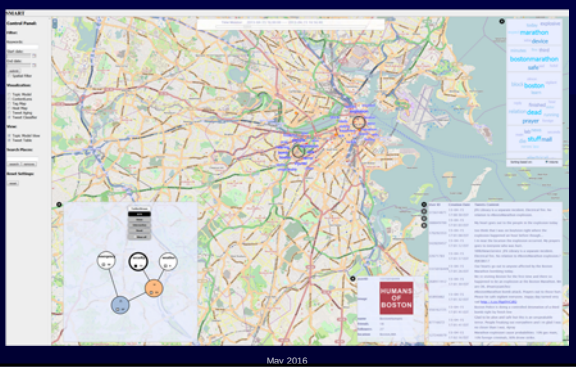
VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

Social Media for Preparedness and Emergency Response

(Social Media Analytics and Reporting Tool)



May 2016

Social Media Analytics 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

May 2016

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

May 2016

Anomaly Detection Superbowl Weekend



Where Can It Be Helpful?

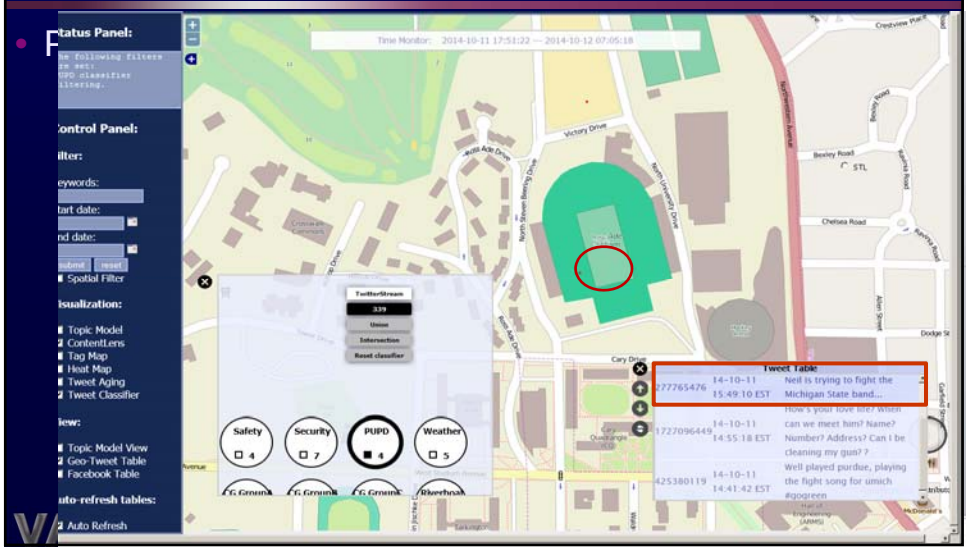
- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

May 2016

Example From Purdue Football 10/12/14

- F



Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - Superbowl 2014 in NJ
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Umpqua Community College Shooting
  - Hurricane Joaquin
  - Superstorm Sandy
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day

- Umpqua Community College Shooting



VACCINE

May 2016

Unexpected Events: Need For Solutions to Help First Responders Shown Every Day



kullgreni
Roseburg, Oregon
FOLLOW

20 likes 3d

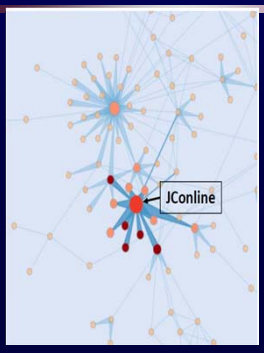
kullgreni Prayer chain for UCC shooting victims. The end of a long, tragic day.

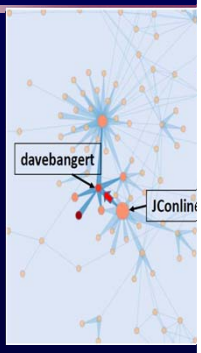
thadizzo84 😔😔😔

sgonzalez910 Such a tragic thing to cover. Stay strong

Log in to like or comment.

Tracking Information Diffusion and Identifying Influential Users Through User [illegible] Shooting Event)







JConline is a local news media that has a lot

JConline retweeted from **davebangert** about the updates of the shooting event.

davebangert retweeted from **justinlmack**.

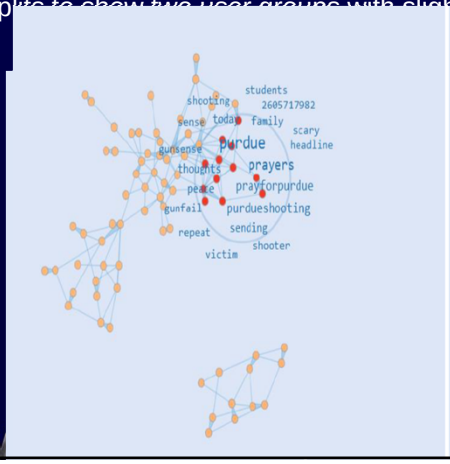
justinlmack posted first-hand updates about the event.

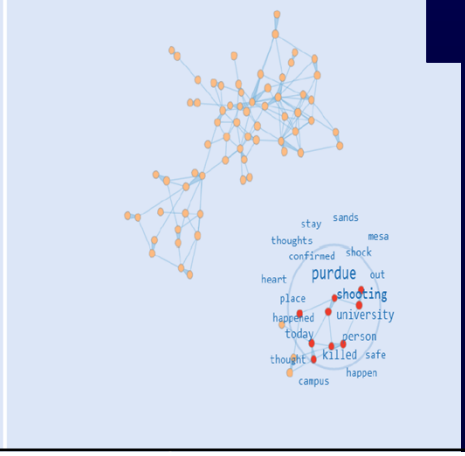
VACCINE

May 2016

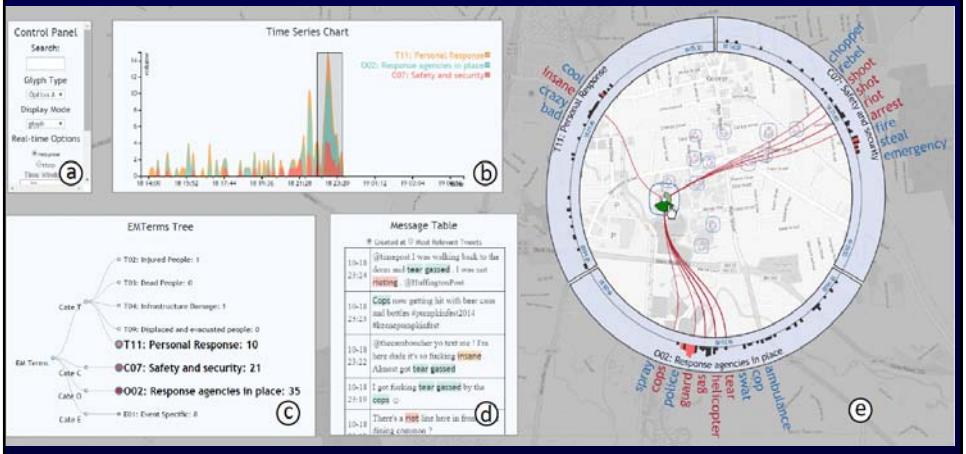
Monitoring Network Evolutions Over Time

Initially, there is a single network containing 'purdue', 'shooting', 'custody'. As time passes, the network evolves and new nodes come. The network then spl[illegible]



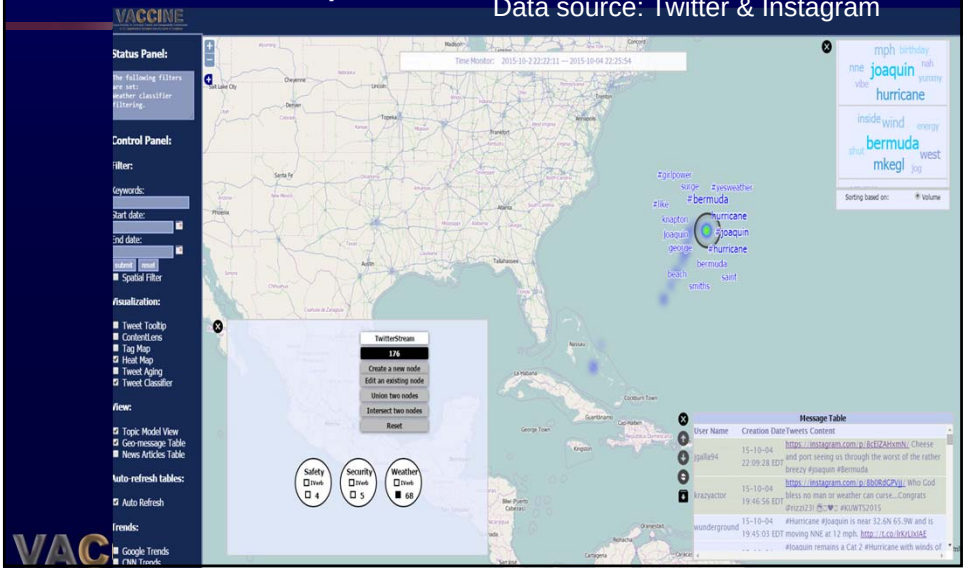


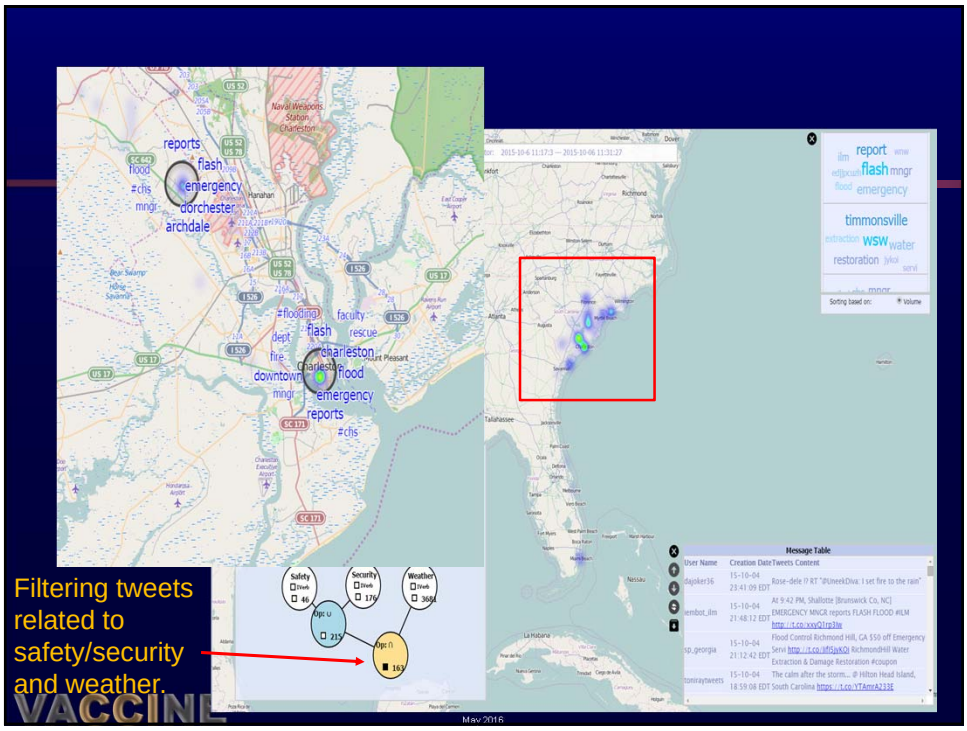
Multi-level Aggregation of Topic Evolution (Eurovis 2016)

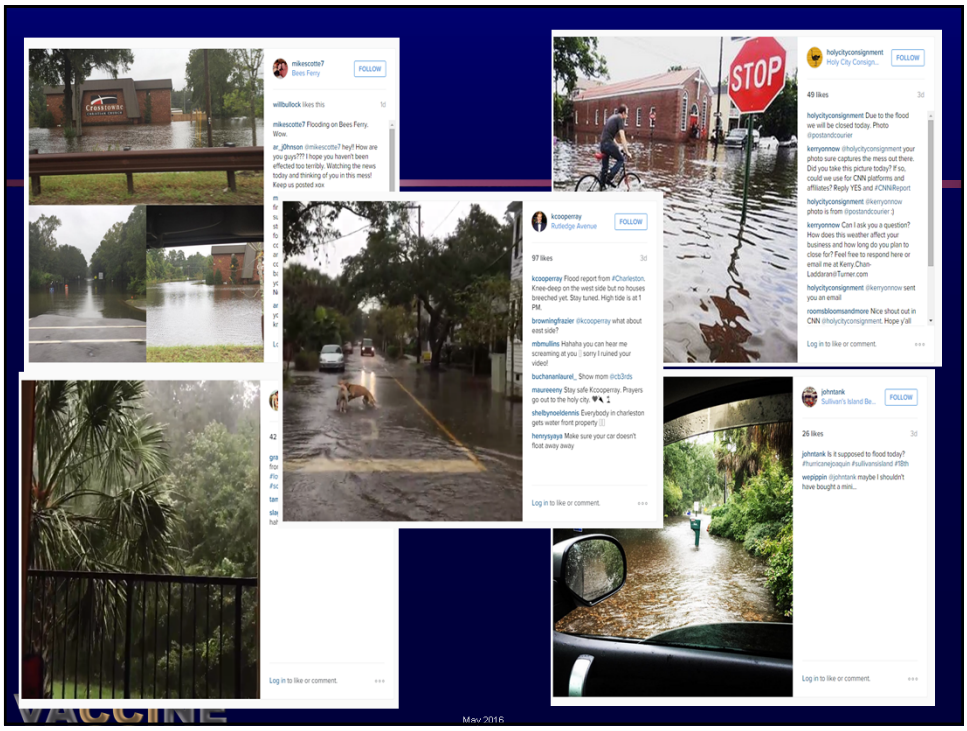


Hurricane Joaquin

Time range: 2015-10-03 to 2015-10-04
Location: Southeast US
Data source: Twitter & Instagram

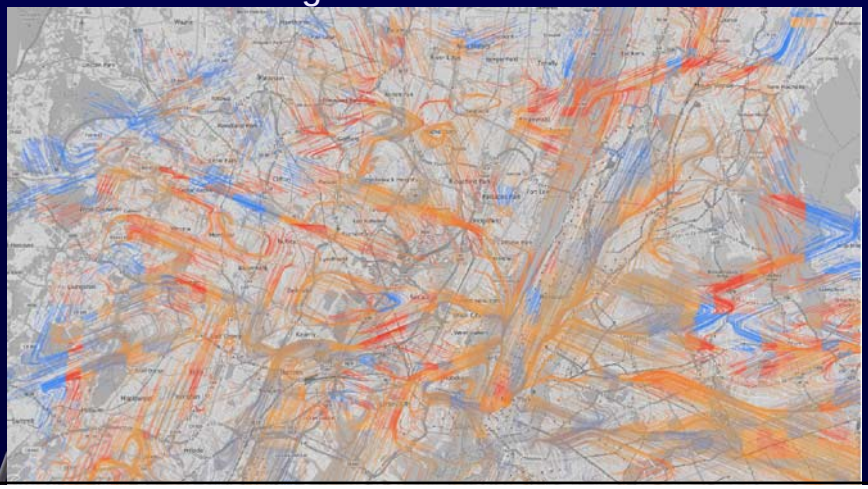



reports
flash
emergency
dorchester
archdale
#flooding
flash
faculty
rescue
dept
charleston
fire
downtown
flood
emergency
reports
#chs
report
flash
mngr
emergency
timmonsville
wsw
water
restoration
Message Table
Filtering tweets related to safety/security and weather.
VACCINE
May 2016


STOP
VACCINE
May 2016

Movement Patterns from Social Media Data

- Manhattan morning rush hour



Summary of SMART For Event Decision Making

Examples:

- Purdue, Ohio State football games
- Kayaktivists
- MLB All Star Game
- Kentucky Derby
- Belle of Louisville Celebration, Thunder over Louisville
- Fleet Week in San Francisco
- Hurricane Joaquin
- Community College
- RNC Summer 2016

VACCINE

May 2016

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

From Big Data to Data-Driven Policy and Decision Making:

Empowering Effective Decision Making Through Visual Analytics

David S. Ebert
School of ECE
Purdue University
ebertd@purdue.edu

Overview

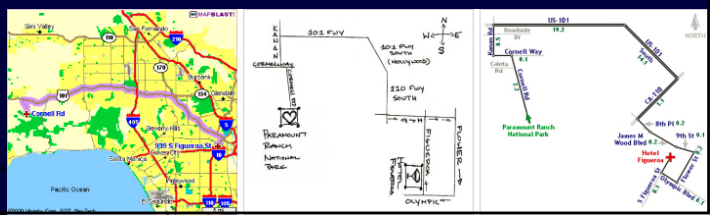
- Background Visualization vs. Visual Analytics
- Big Data and Visual Analytics
- What Visual Analytics can do (examples)
- Directions for the future



"Remember, the other team is counting on Big Data insights based on previous games. So, kick the ball with your other foot."

What is Visualization?

- "The use of computer-supported, interactive visual representations of data to amplify cognition."[1]
- Not simply making a graphic or an image, the goal is to create insight, utilizing cognitive and perceptual principles
- Help people form a mental image of something and internalize their own understanding
- Promote discovery, decision making and explanations



Agrawala, Stolte: Rendering effective route maps: improving usability through generalization. SIGGRAPH 2001: 241-249

Tasks for Visualization

- Searching and browsing - find a specific piece of information, inspect data, seek information
- Analyze – do comparisons, differences, look for outliers, extrema and patterns
- Monitor – look for changes and trends

How to Harness Visual Representations

- Visual representations translate data into a visible form, highlighting important features, such as commonalities and anomalies
- Visual representations make it easy for users to perceive salient aspects of their data quickly
- These visual representations augment the cognitive reasoning process with perceptual reasoning that enhances the analytical reasoning process

1 – J. J. Thomas and K. A. Cook (Ed.). *Illuminating the Path: The R&D Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

The Power of Visualization

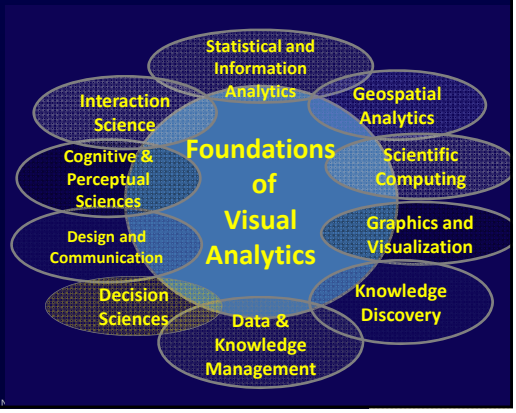
- Goal: Transform data into information
- How many terabytes of data doesn't matter, it's how many insights we can generate from this data
- "The purpose of computing is insight, not numbers."
  Dedication to his book: RW Hamming (1971). Introduction to Applied Numerical Analysis. McGraw Hill.
- We need to make the data understandable to people
  - Visualization is key!

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

Essential Features for Visual Analytics

- Interactive, integrated discovery environment
- User-guided and perceptually-guided
- Balance of human cognition and automated computerized analysis to amplify human cognition
- Enable discovery and decision making
- Provide quantitative, reliable, reproducible evidence

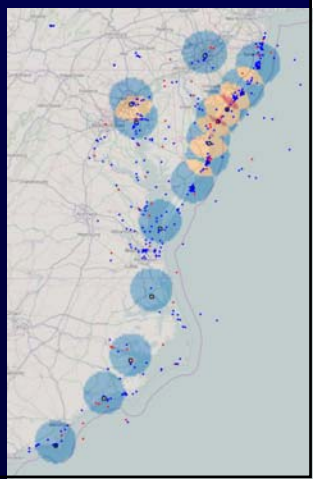
VACCINE

November 2015

PURDUE

Visualization, Data, and Decision Making

- At some level most decision making is driven by data
- What role does visual representation play?
- Good or bad – people believe charts and graphics – “seeing is believing”



VACCINE

Big Data

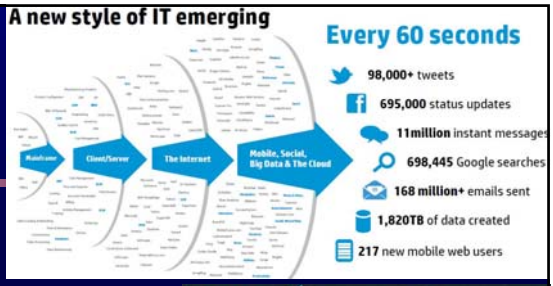
30 ZETTABYTES — The amount of data that will be stored in the world by 2020.
One zettabyte is the equivalent of streaming 36 million years of HDTV video
SOURCE: IDC

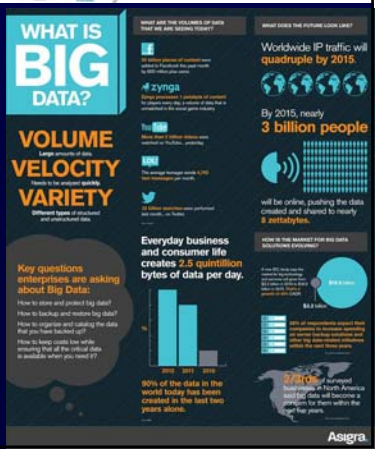
10 TERABYTES — The amount of data Facebook generates daily.

7 TERABYTES — The amount of data Twitter generates daily.
One terabyte would hold about 250,000 songs, or nearly a year's worth of continuous music
SOURCE: IBM

40 PERCENT — The annual growth rate of global data generated per year.
SOURCE: MCKINSEY

10 BILLION+ — The number of mobile-connected computing devices in use by 2016, up from less than 7 billion in 2012.
SOURCE: CISCO



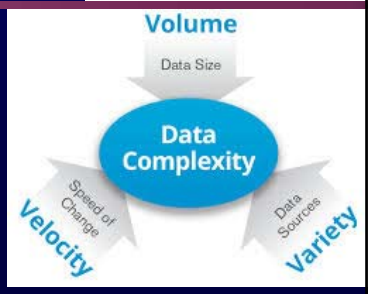


Big Data: Solution to Global Challenges?



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**



VACCINE

November 2015

These Challenges are Why the U.S. Department of Homeland Security Helped Create the Field of Visual Analytics and Created a Center of Excellence to Solve These Challenges!

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

November 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 Washington
- Arizona State University
- Bethune-Cookman University
- Jackson 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

November 2015

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

November 2015

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 Decision Makers

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



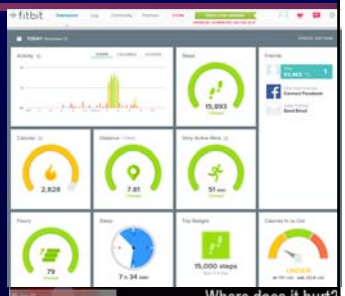
Visual Analytics for Decision Making

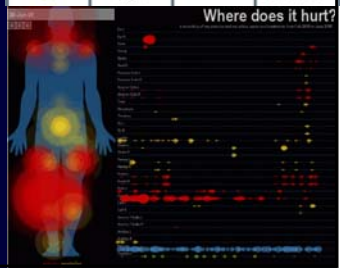
- Personal level
- University level
- Community level
- State and regional level
- National and global scale

Visual Analytics for Effective Decision Making: Personal to Global

- Visual analytics can help people understand data in their everyday lives







University Scale Decision Making

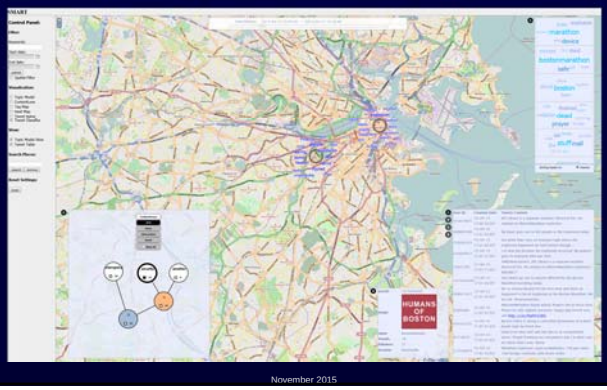
- Football game force deployment using SMART
- Student safety
  - Reducing high-risk alcohol behavior
  - Camera locations and safe paths
- Personal safety across campus
- Enabling scientists and engineers

VACCINE

Visual Analytics for Command, Control, and Interoperability Environments

A U.S. Department of Homeland Security Center of Excellence

SMART: Social Media Analytics and Reporting Tool



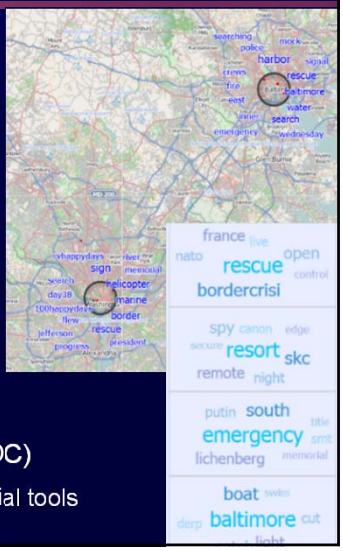
November 2015

Social Media Analytics and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

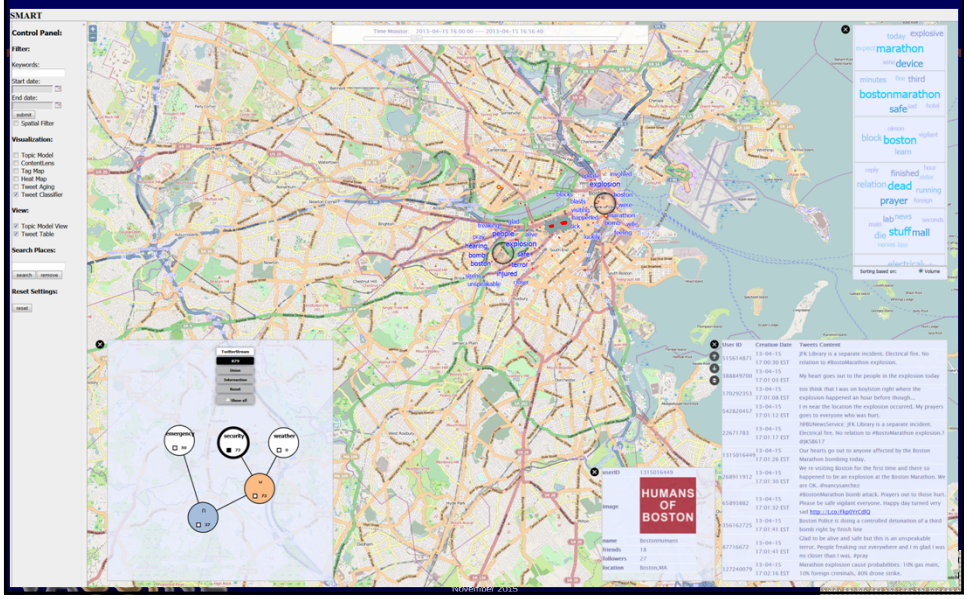
- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games 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

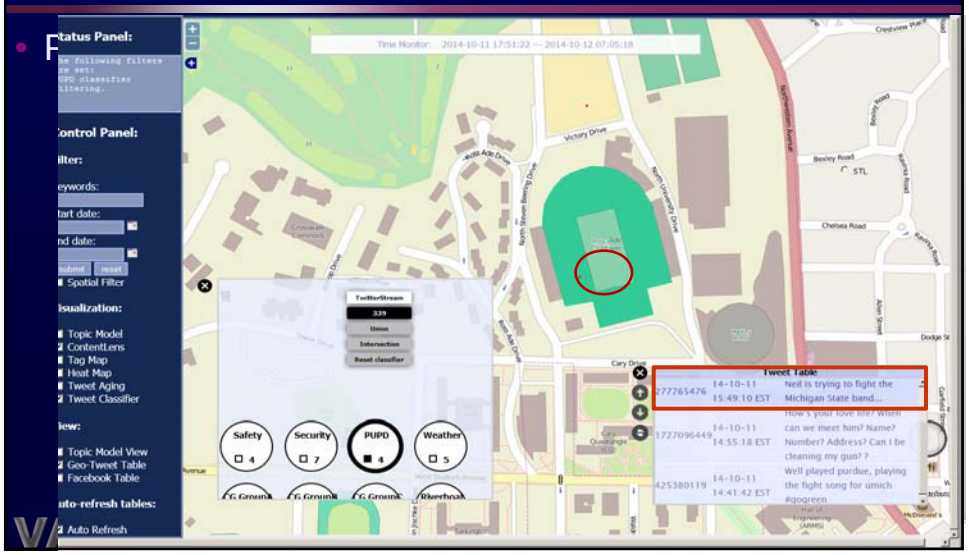
November 2015

Social Media Analytics and Reporting Toolkit (SMART)

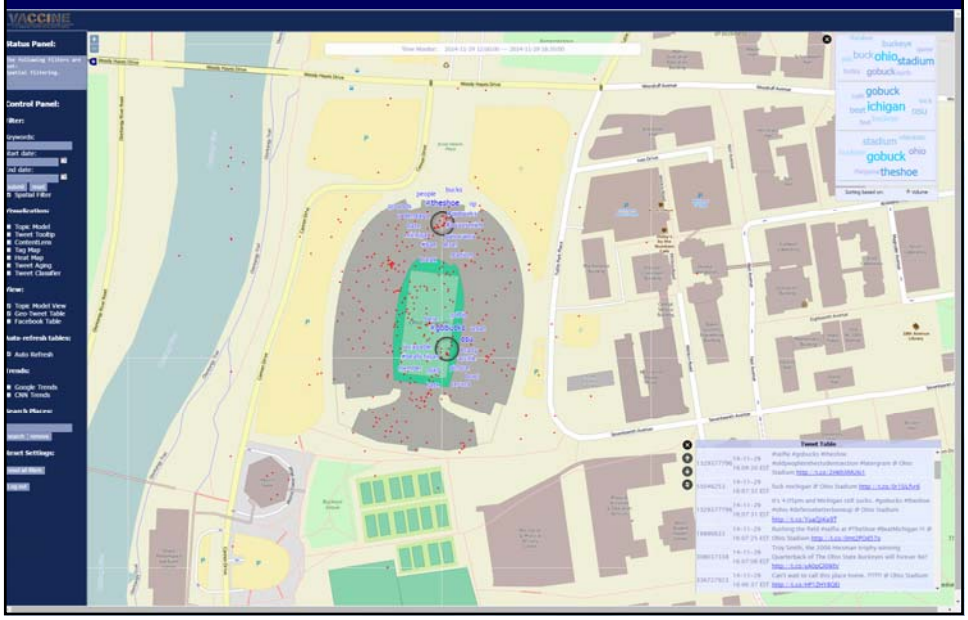


Example From Purdue Football 10/12/14

- F



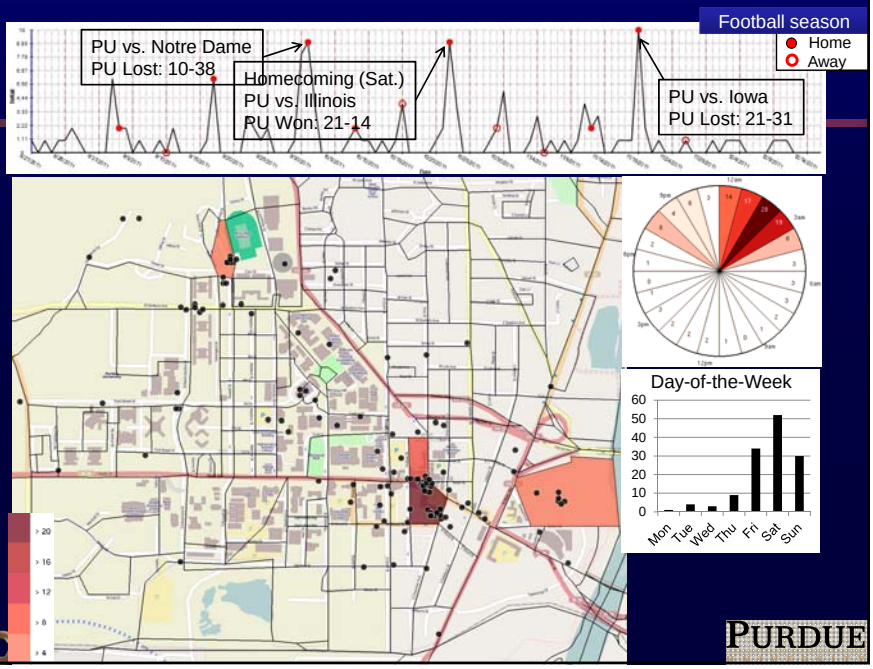
Example From Ohio State Football 11/29/14



Reducing High-Risk Alcohol Behavior

- Student Wellness Center Partnership
- Find areas and patterns of high-risk alcohol behavior
- Information campaign to students and dorms
- Increased patrolling

High Risk Behavior : Drunkenness / Public Intoxication



VAC

PURDUE

Community Scale Decision Making

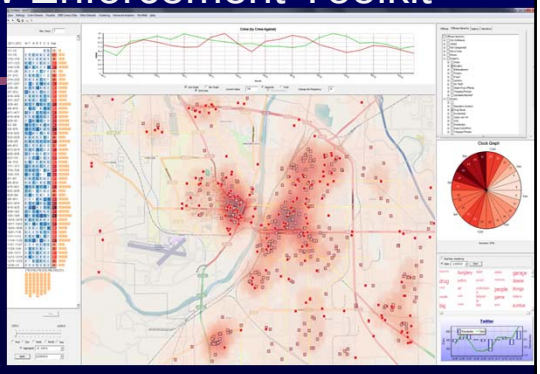
- Public safety – crime analytics (VALET)
- Transit safety – eBOLO

VACCINE

November 2015

PURDUE

Visual Analytics Law Enforcement Toolkit
(VALET, iVALET)



Impacts:

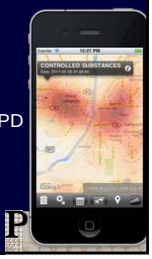
- In use to analyze crime patterns and to connect strings of activities (200+ downloads)
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive 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



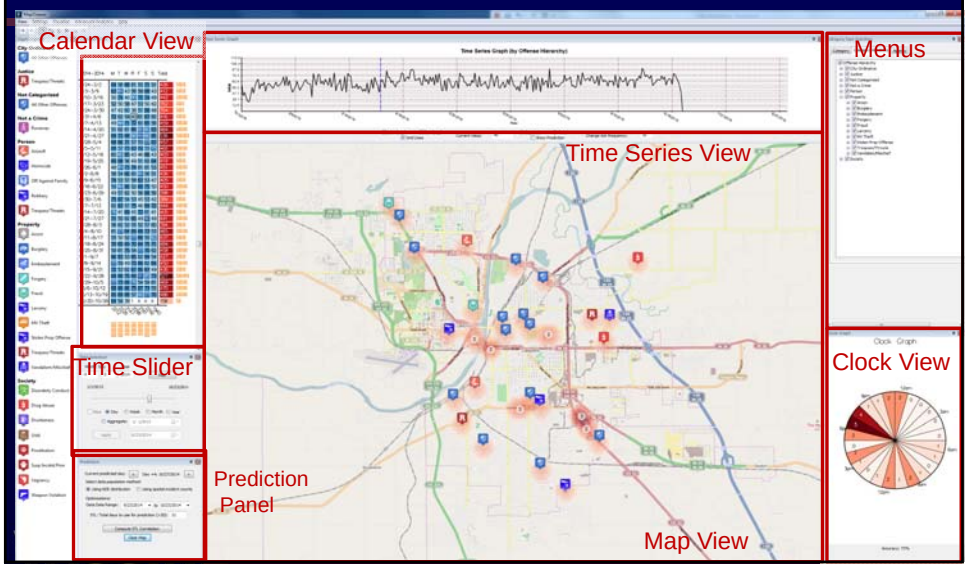
VACCINE

November 2015

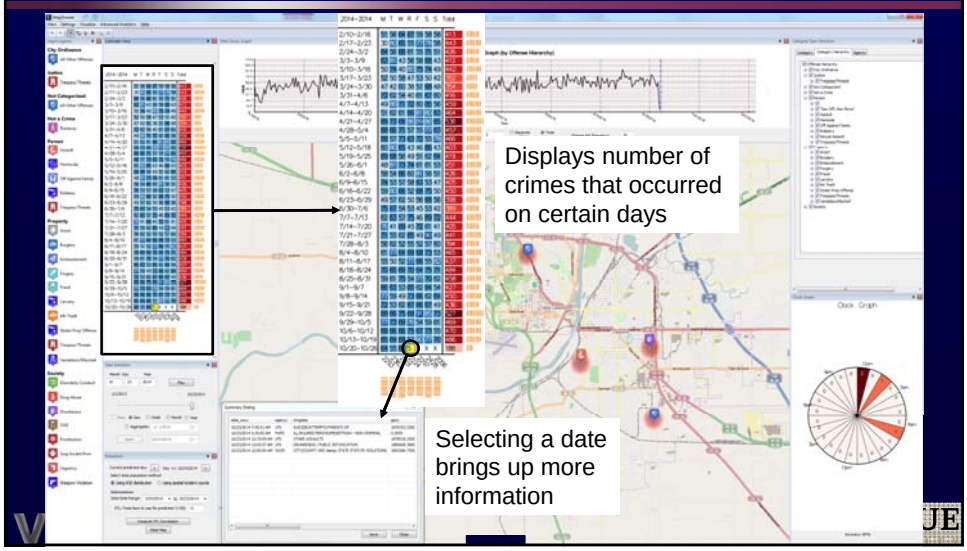
Example Projects: Public Safety and Law Enforcement Applications



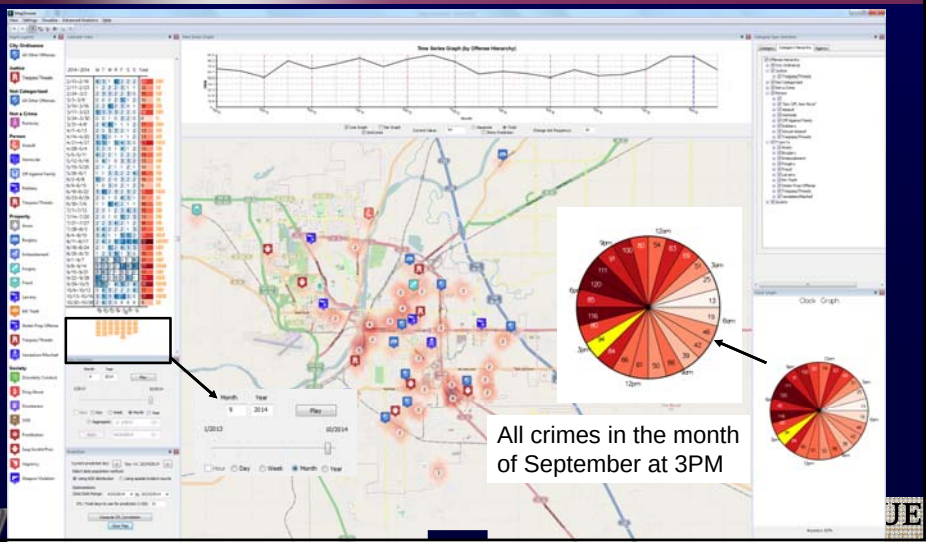
VALET Layout

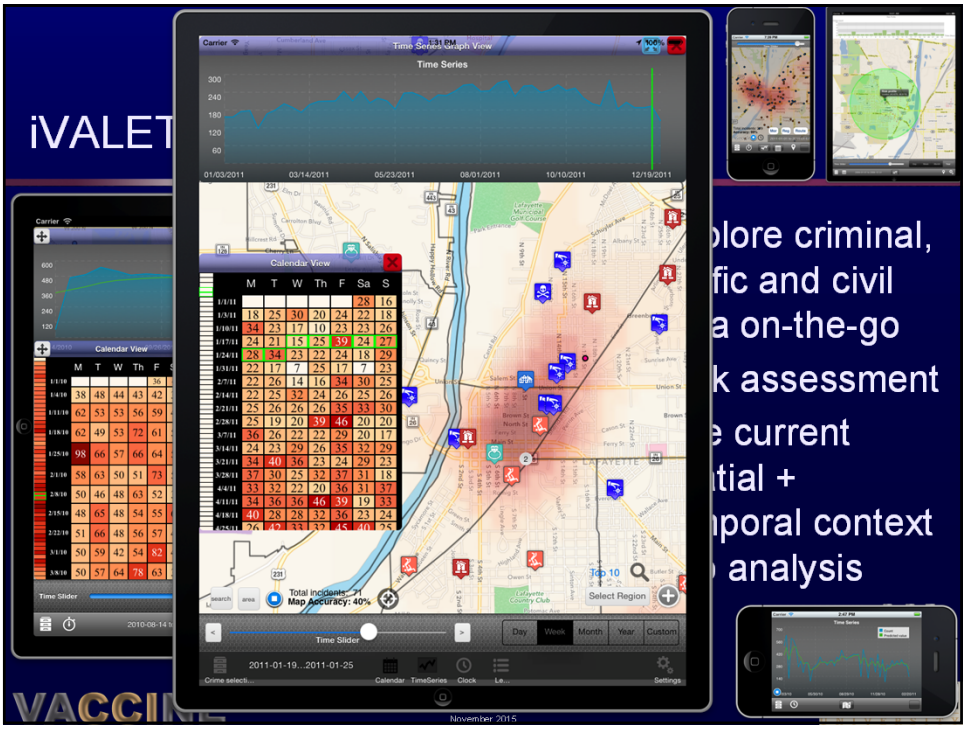


Calendar Display



Crimes During Certain Hours In A Month





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**

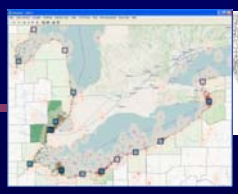
1. D. A. Norman. Things that make us smart: Defending human attributes in the age of the machine. Basic Books, 1993.

State and Regional Level Decision Making

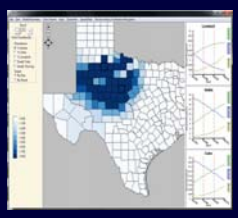
- Public Health
- Disaster Preparedness and Response
- VASA

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





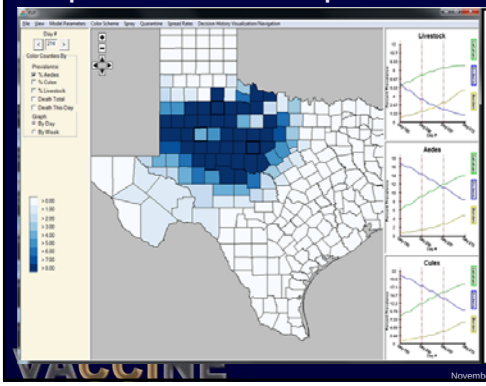

VACCINE

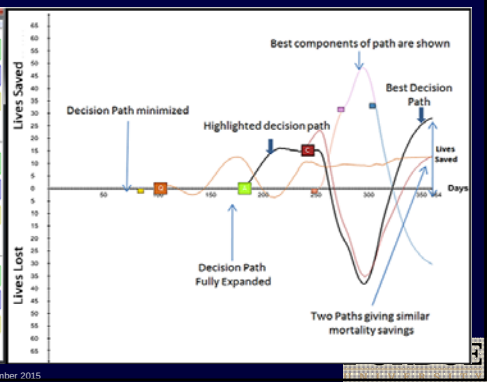
November 2015

PURDUE

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

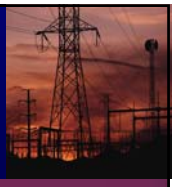




VACCINE

November 2015

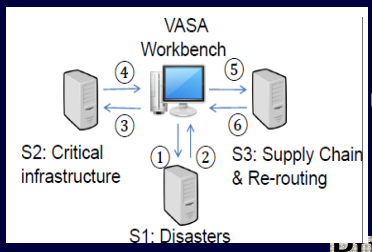
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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

November 2015

PURDUE

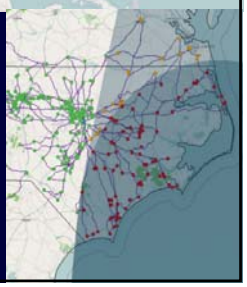
VASA: Visual Analytics for Simulation-Based Action

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 simulation 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

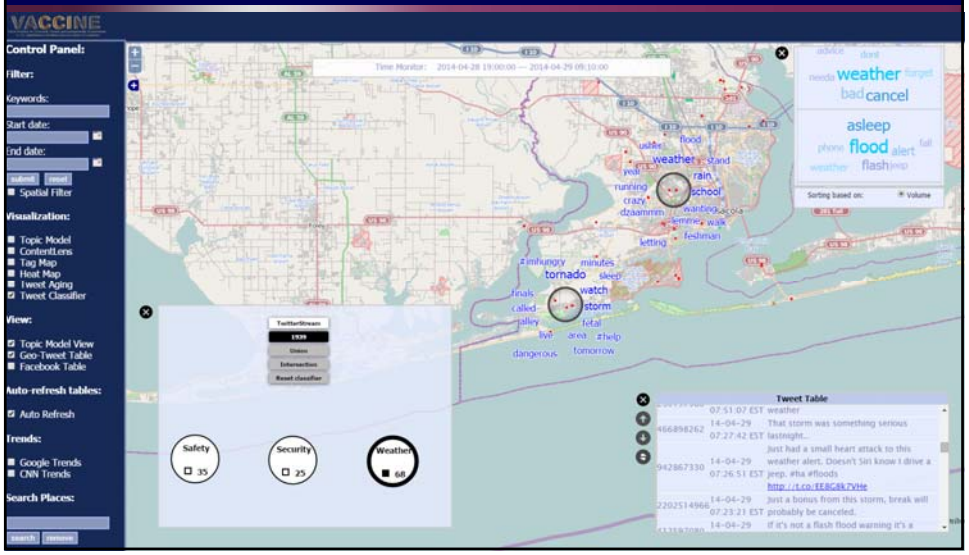




VACCINE

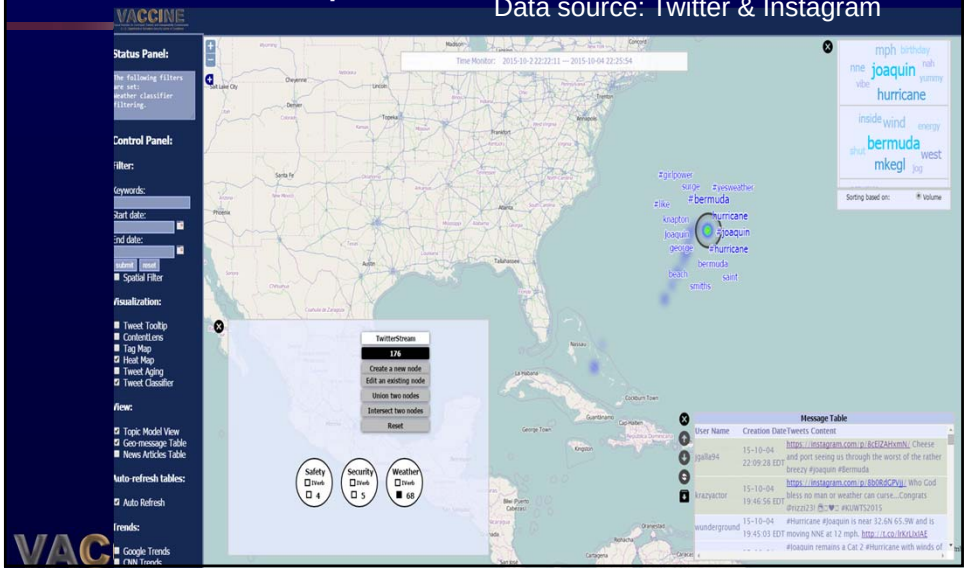
November 2015

SMART: Disaster Preparedness & Response

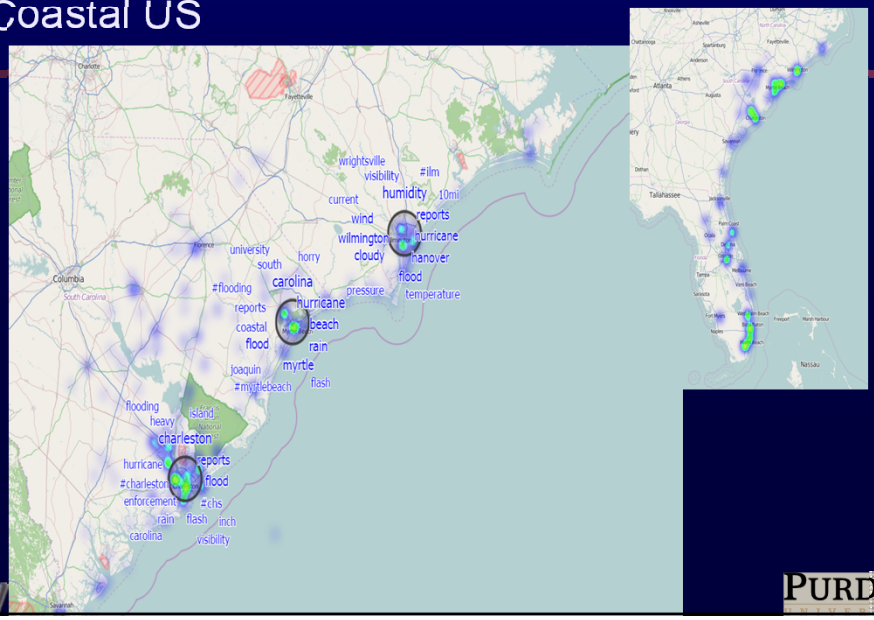


Hurricane Joaquin

Time range: 2015-10-03 to 2015-10-04
Location: Southeast US
Data source: Twitter & Instagram



Weather-related tweets in Southeast Coastal US



PURDUE

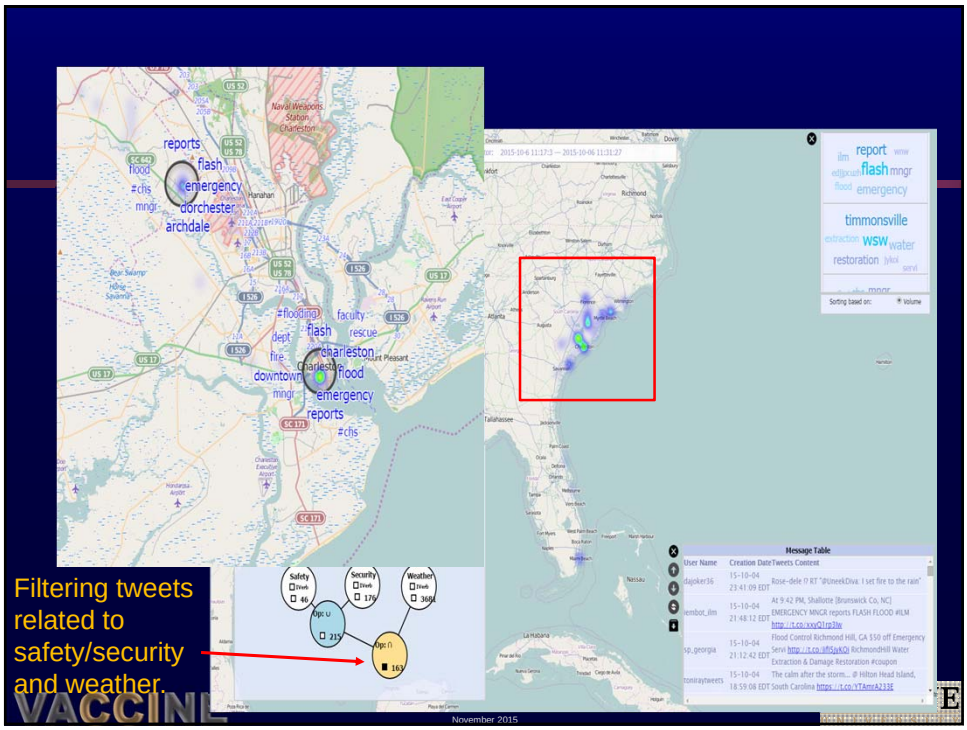
Top Topics related to Weather Classifier

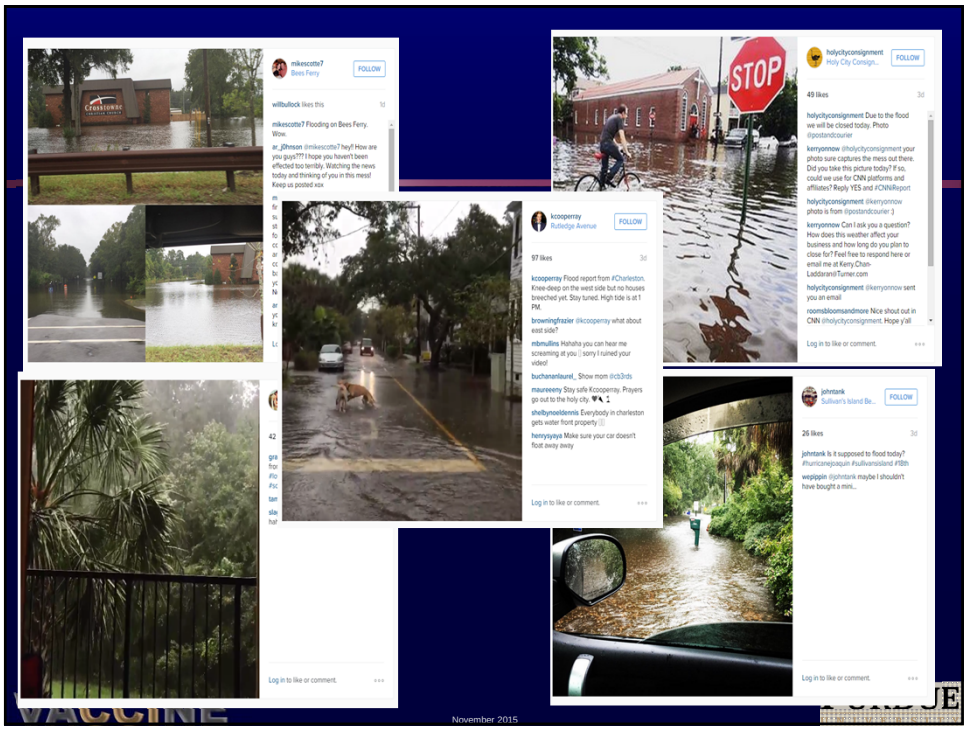


VACCINE

November 2015

PURDUE


reports
flash
emergency
dorchester
archdale
#flooding
faculty
rescue
dept
flash
charleston
fire
downtown
flood
emergency
reports
#chs
report
flash
mngr
flood
emergency
timmonsville
wsw
water
restoration
Message Table
Safety
Security
Weather
Filtering tweets related to safety/security and weather.
November 2015


STOP
November 2015

Where Can It Be Helpful?

- During large-scale planned events, e.g.,
  - MLB All Star Game
  - 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

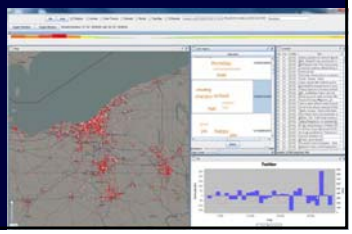
National / Global

- USCG
- Southern Border Approaches
- Precision agriculture and droughts
- Soundscapes

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



VACCINE

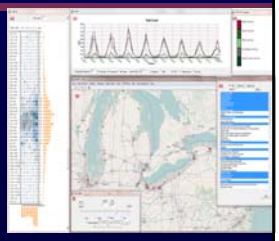
November 2015

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



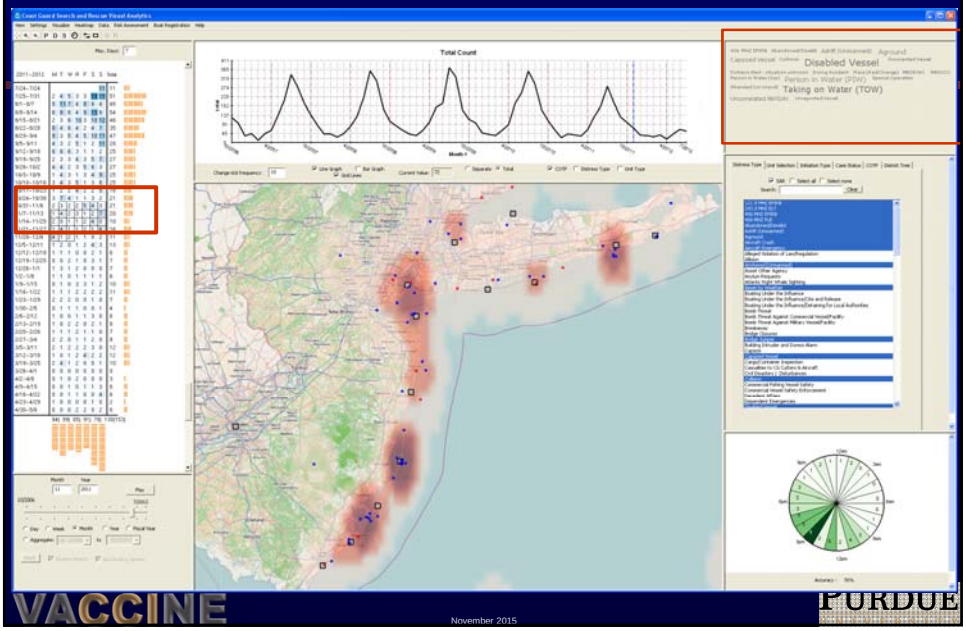


VACCINE

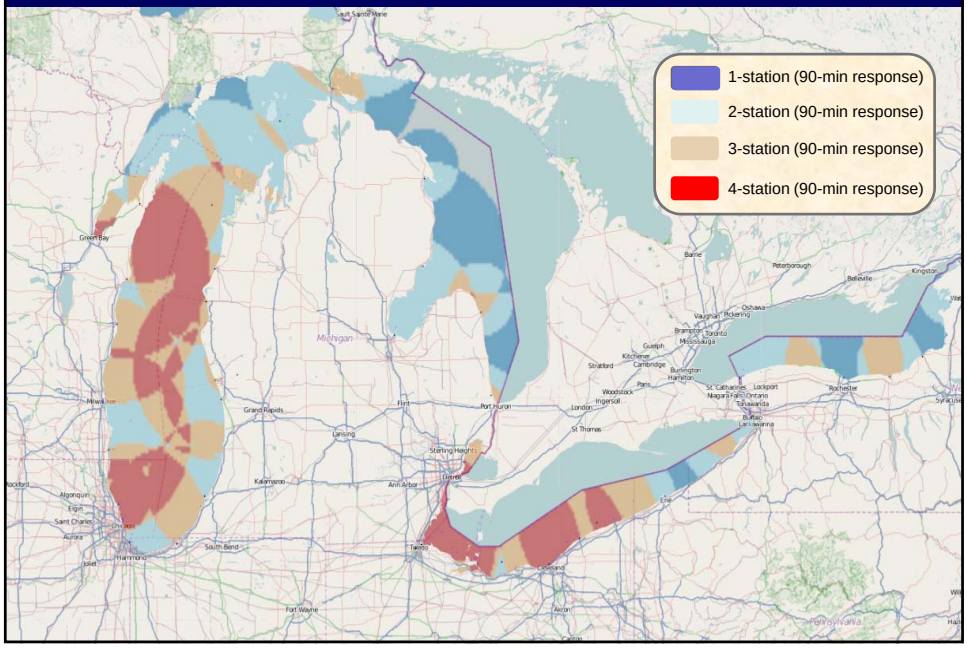
November 2015

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area



Response Efficiency – Potential Future Assets



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

Southern Border Approaches

- Department of Homeland Security has created joint task forces to address Southern Border Approaches security (CBP, ICE, US Coast Guard)
- Joint Task Forces have asked us to assist in visualizing data and metrics to help define performance and evaluate success

VACCINE

November 2015

PURDUE

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
- Incorporate analytics and data mining, but let user guide and refine process
  - Understandability of why is the key

- *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

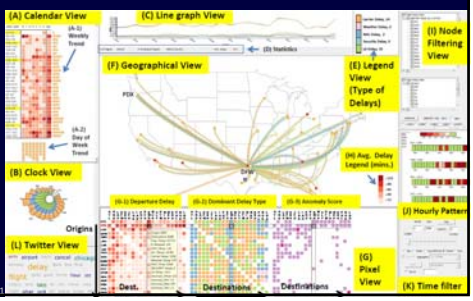
CVADA

Center for Visual Analytics and Data Analytics
A U.S. Department of Homeland Security Center of Excellence

July 2015

Overview

- VACCINE center
- Visual analytics
- Big data challenges
- Example projects



VACCINE

July 201

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

July 2015

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



VACCINE

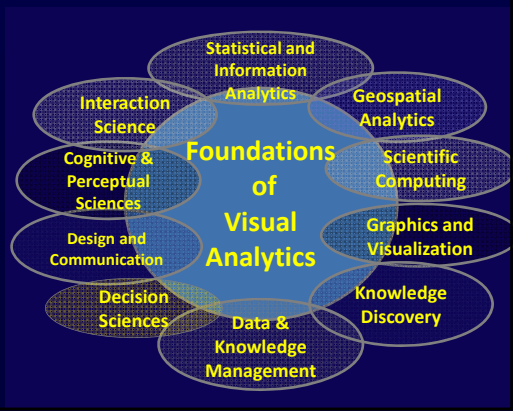
July 2015

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

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

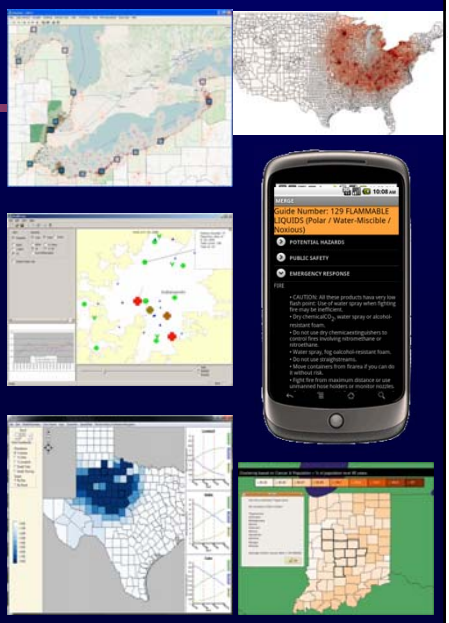
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

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

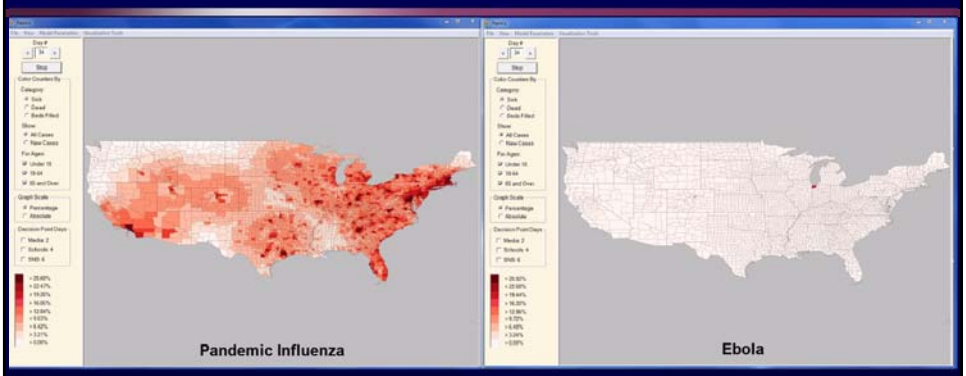


VACCINE

July 2015

Ebola Vs. Pandemic Influenza (H1N1)

Day 34 With Origin of Chicago, IL

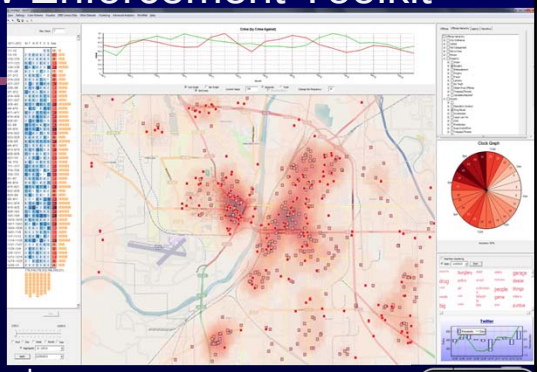


Note: Ebola spread rate based on last 3 outbreaks – U.S. attack rate should be much smaller

VACCINE

July 2015

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:

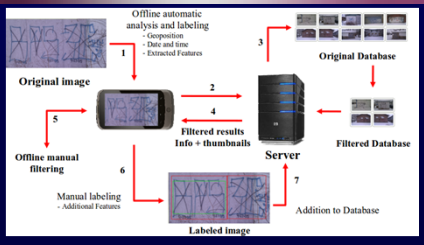
- October 2011: Purdue, WL Police, Lafayette PD



VACCINE

July 2015

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:

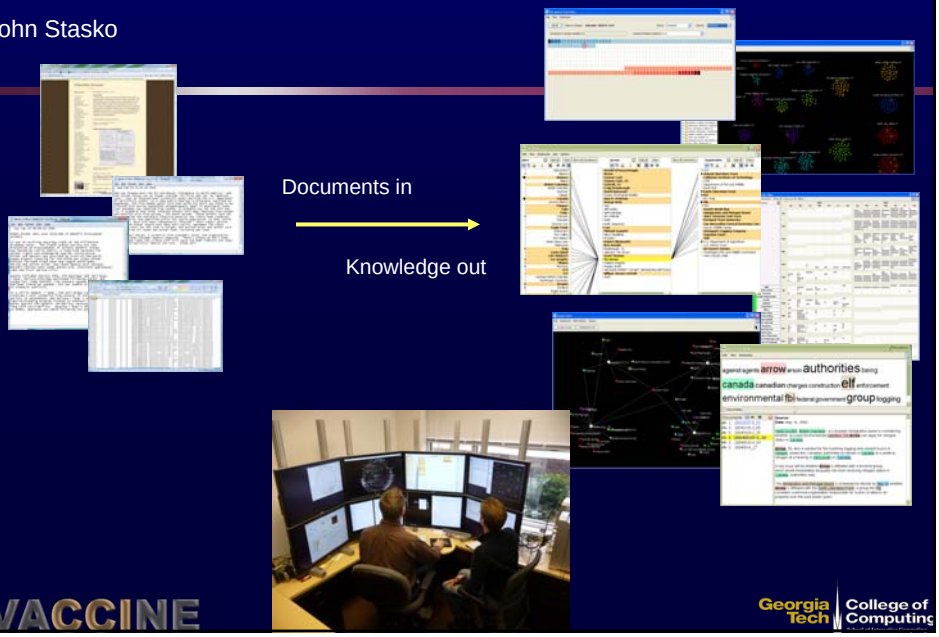
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

July 2015

Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



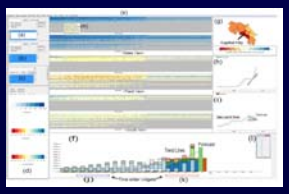
VACCINE

Georgia Tech College of Computing

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



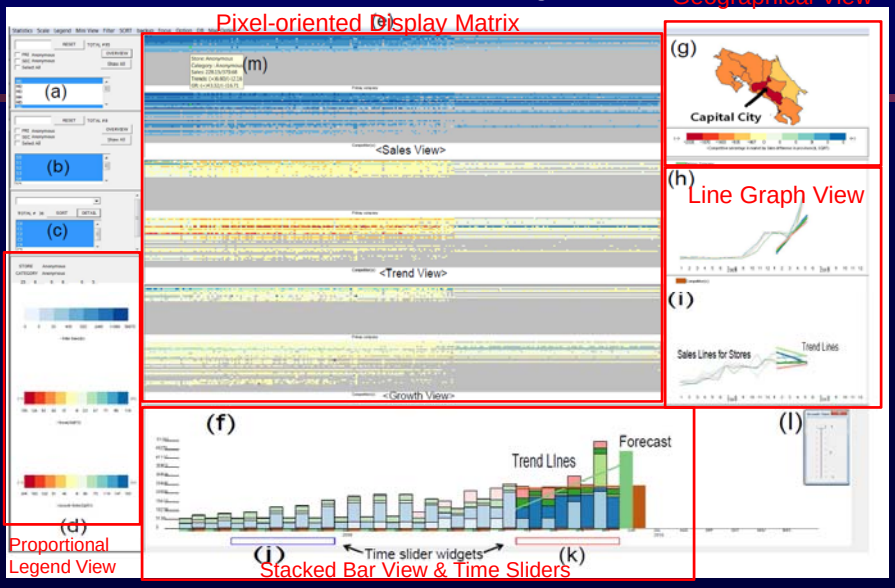




VACCINE

July 2015

Market Analyzer



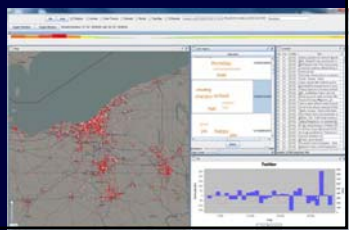
VACCINE

July 2015

Visual Analytics Uses for Public Safety



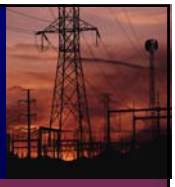
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



VACCINE

July 2015

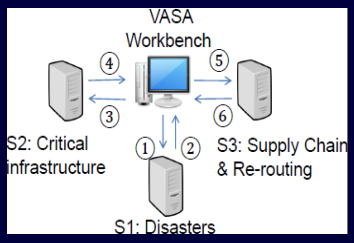
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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



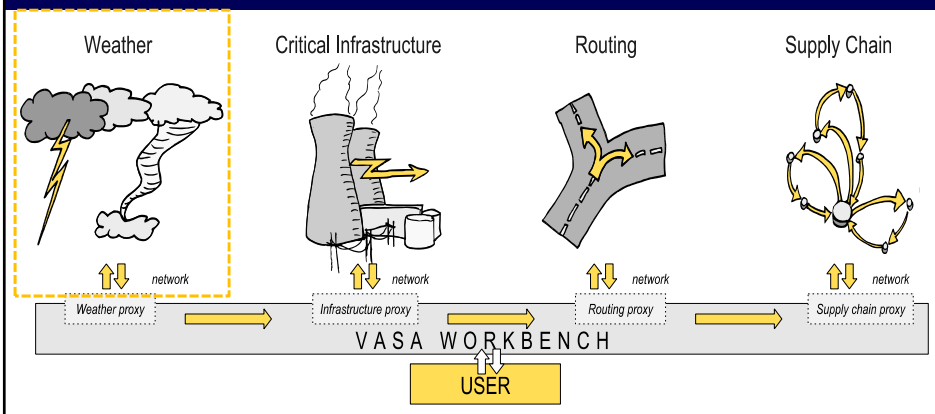
U Minn

TACC

VACCINE

July 2015

VASA COMPONENT: **CRITICAL INFRASTRUCTURE**

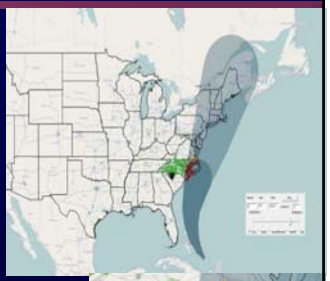


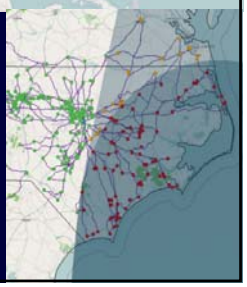
VASA: Visual Analytics for Simulation-Based Action

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

July 2015

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

July 2015

Our Spatiotemporal Visual Analytics Work Contributions

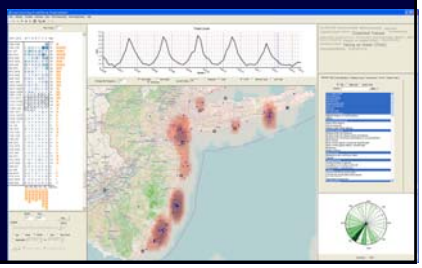
- Techniques:
  - New anomaly and prediction techniques
  - New interactive correlative analysis
  - Natural scale work
  - Social medial visual analytics
- Uses:
  - Anomaly detection
  - Situational awareness
  - Predictive policing
  - Coast Guard visual analytics
    - Search and rescue
    - Risk-based resource allocation
    - Risk-based operational performance

VACCINE

July 2015

Overview

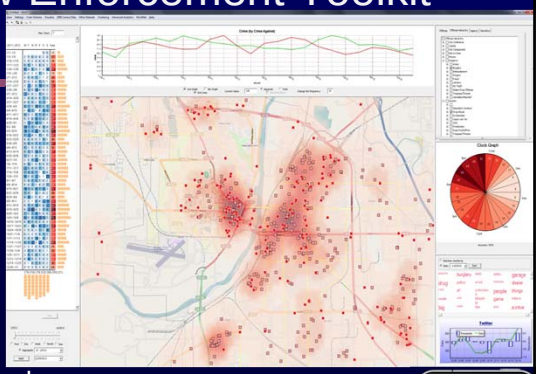
- VACCINE center
- Big Data challenges
- Visual analytics
- Research topics
- Example projects
  - VALET system
  - US Coast Guard
  - Imaging tools
  - Social Media



VACCINE

July 2015

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:

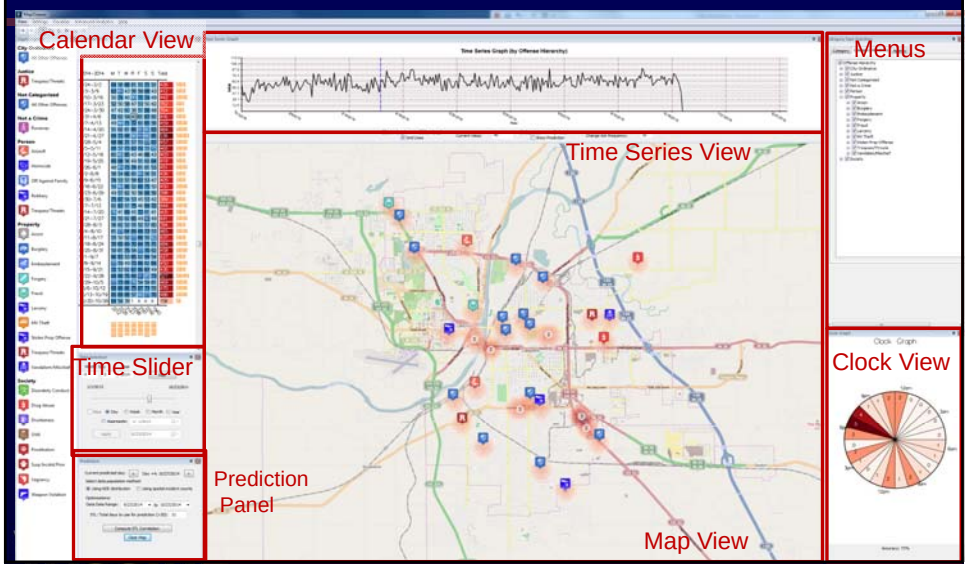
- October 2011: Purdue, WL Police, Lafayette PD



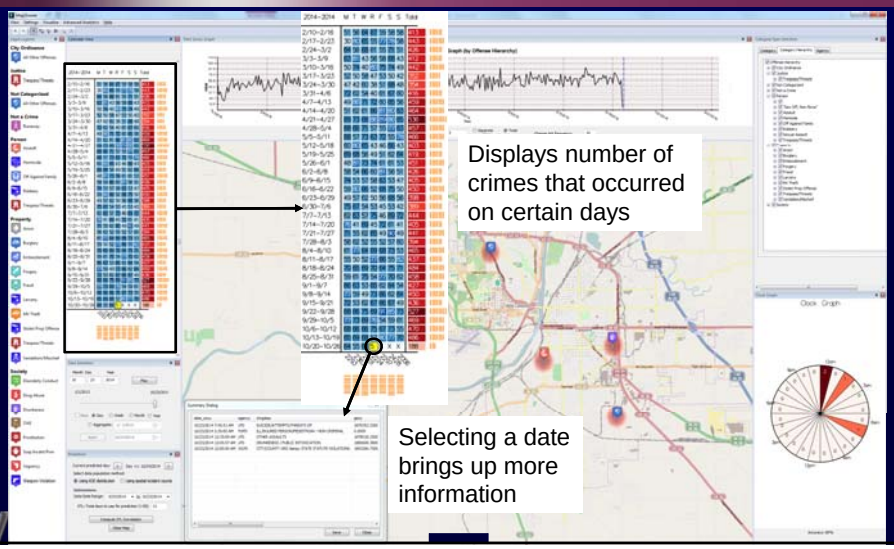
VACCINE

July 2015

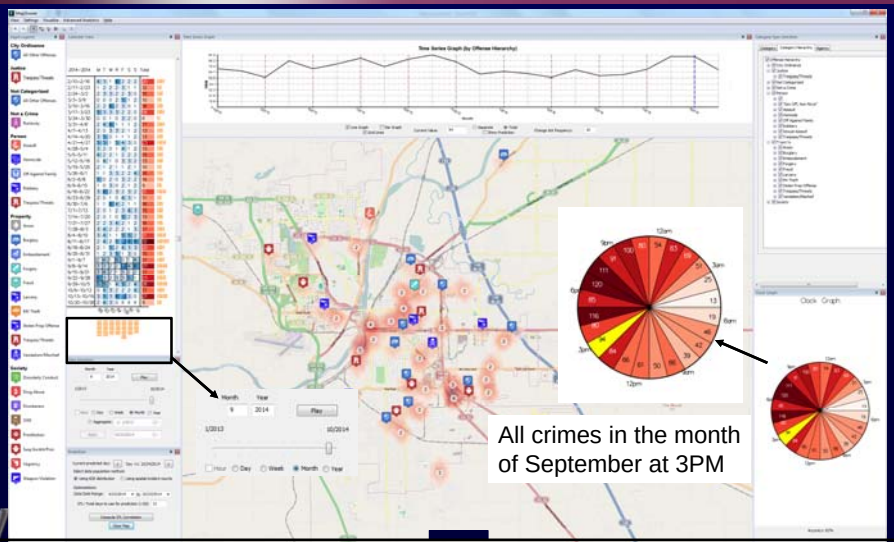
VALET Layout

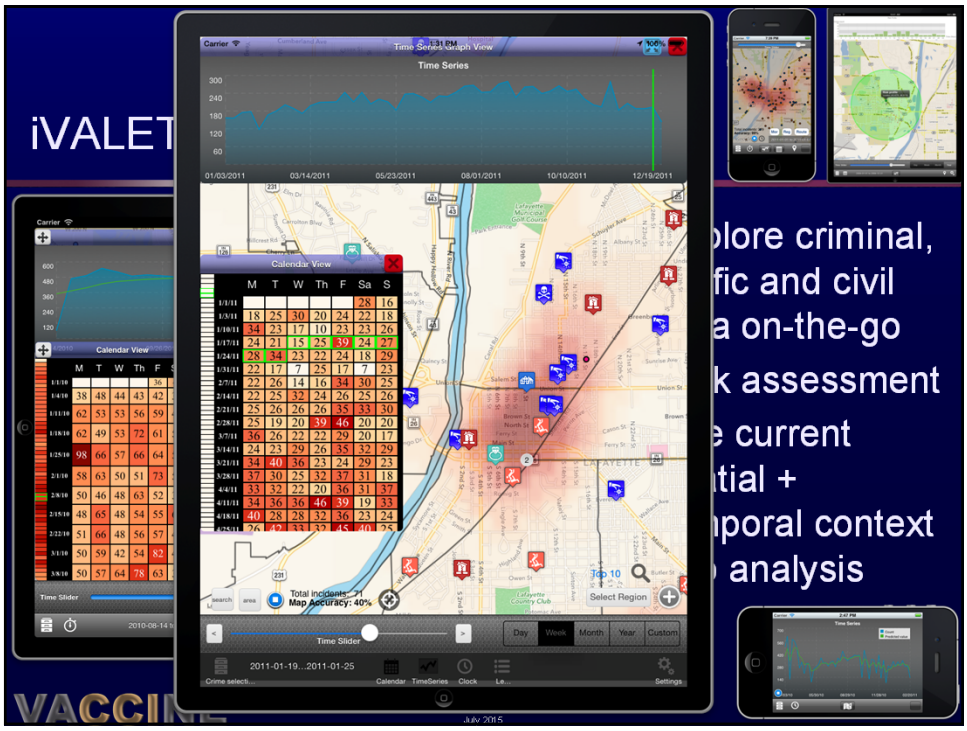


Calendar Display



Crimes During Certain Hours In A Month



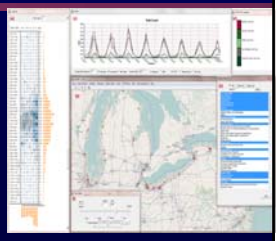


U.S. Coast Guard COAST/ SARVA (cgSARVA)

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



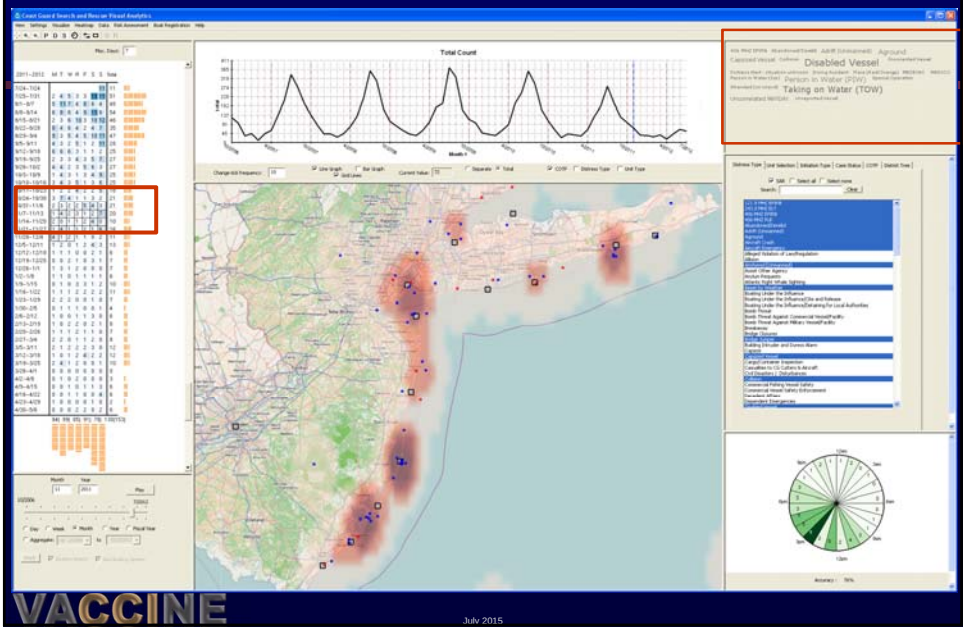


VACCINE

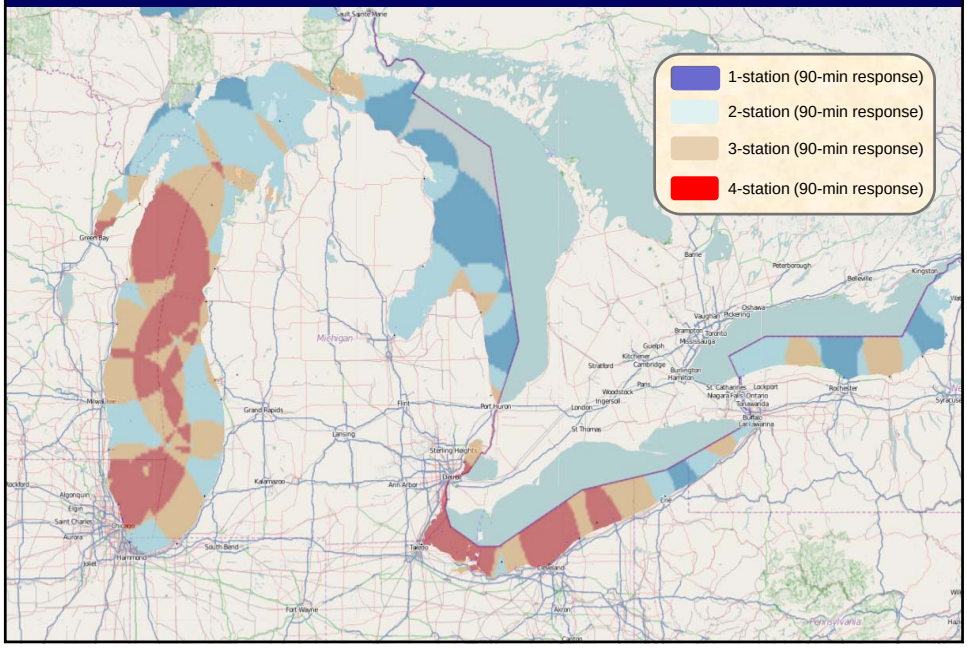
July 2015

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

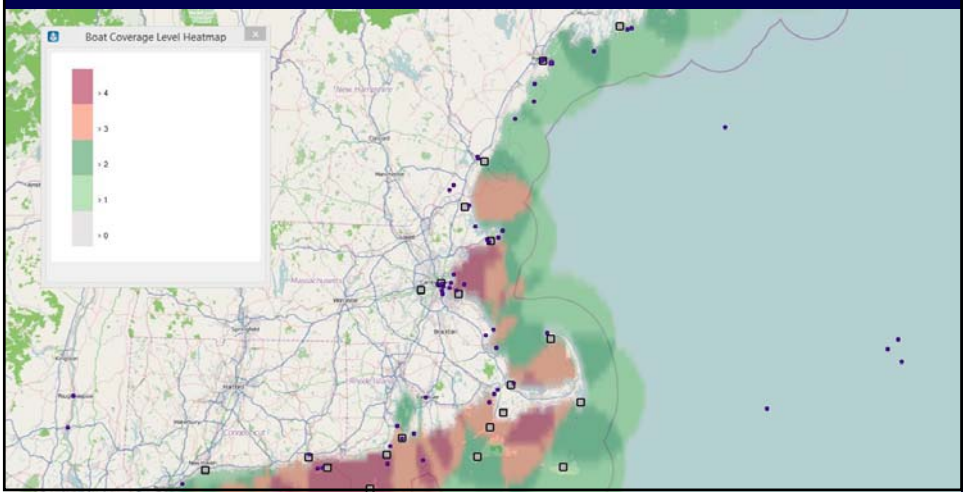


Covered Portions of Mississippi River: District 9, AIRFAC WAUKEGAN
District 8, AIRSTA NEW ORLEANS

VACCINE

July 2015

Coverage vs Lives Lost Aggregated Since Fiscal Year 2007

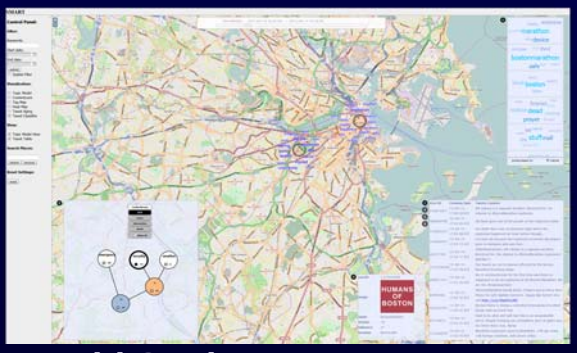


VACCINE

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

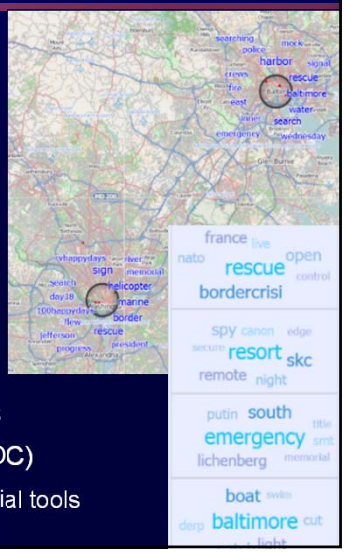
July 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

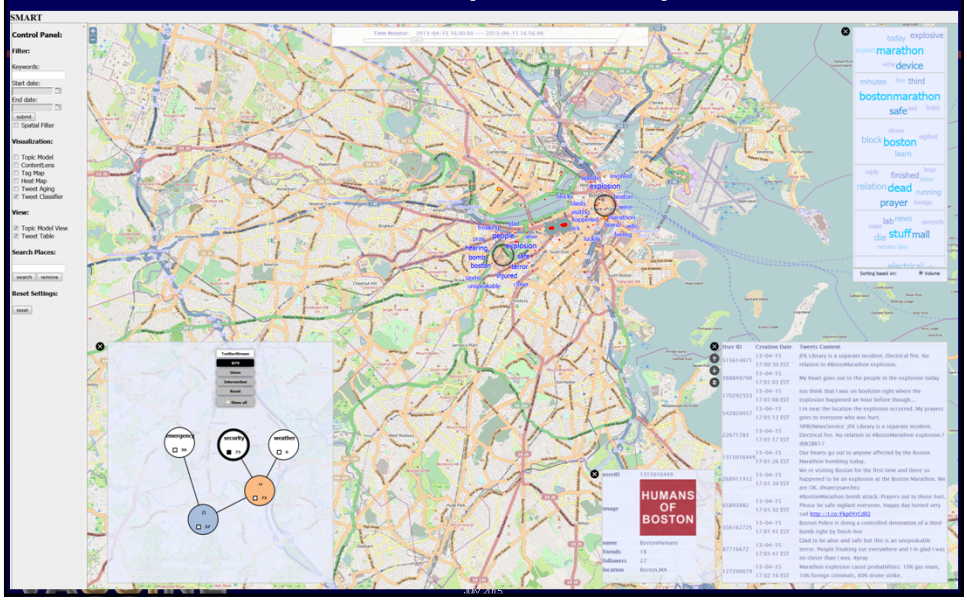
- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games 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

July 2015

Social Media Analysis and Reporting Toolkit (SMART)



Anomaly Detection Superbowl Weekend



Examples From Fleet Wee



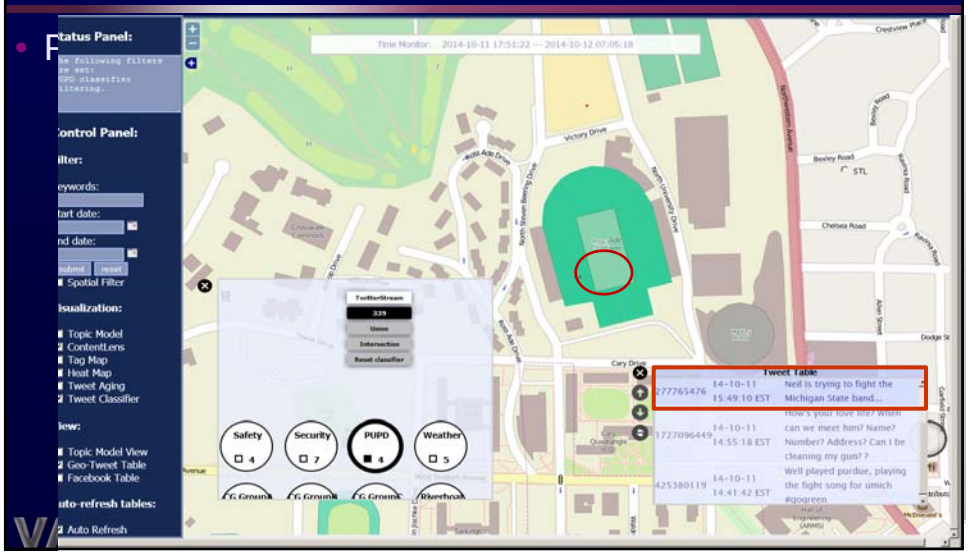
Where Can It Be Helpful?

- 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

VACCINE July 2015

Example From Purdue Football 10/12/14

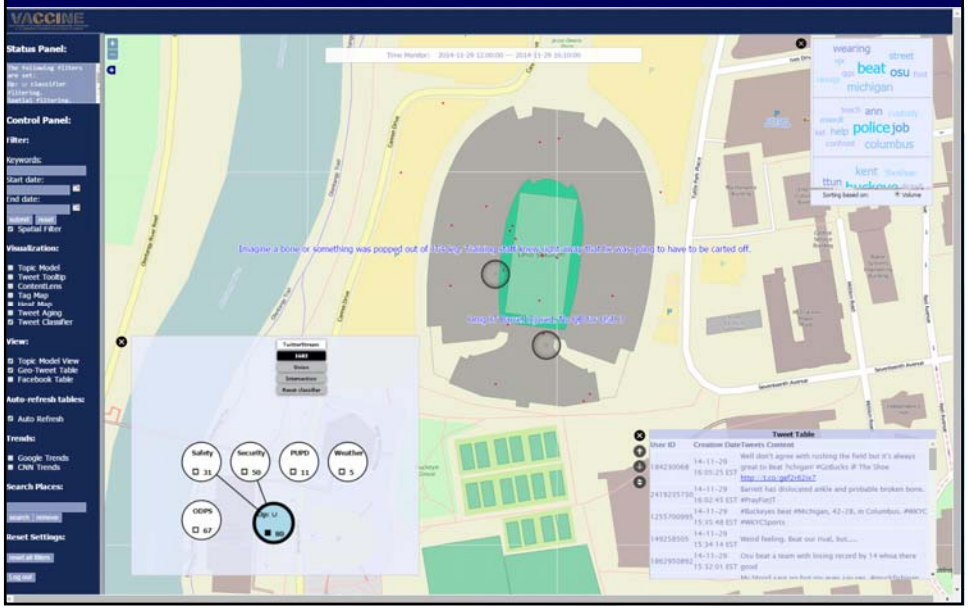
- F



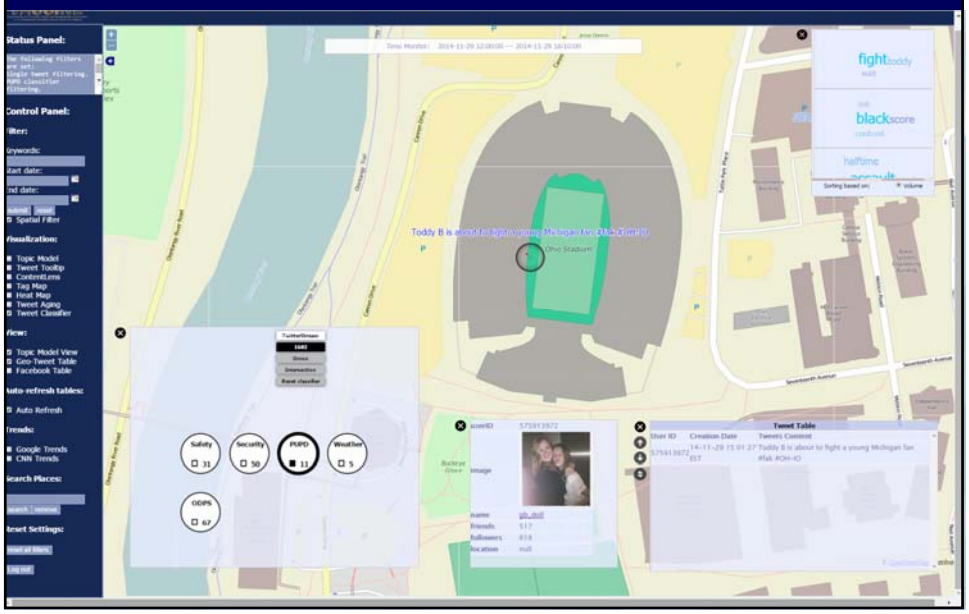
Example From Ohio State Football 11/29/14



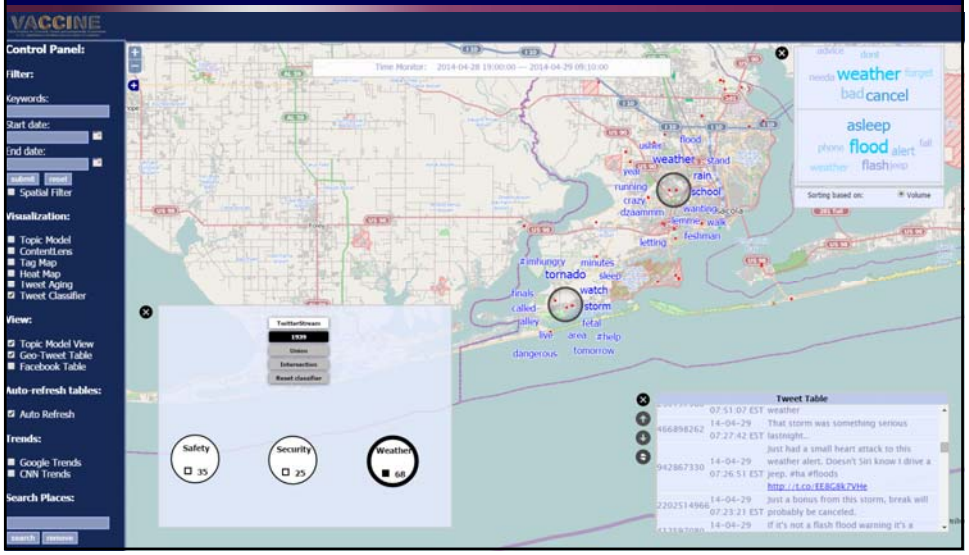
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



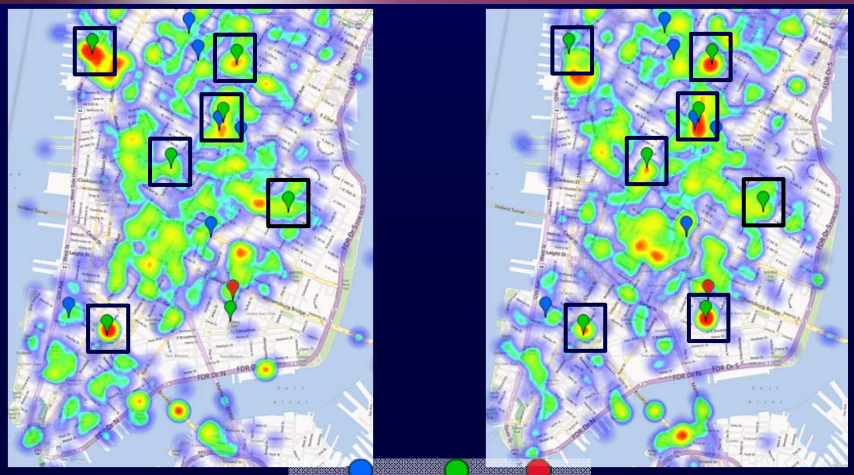
Where Can It Be Helpful?



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



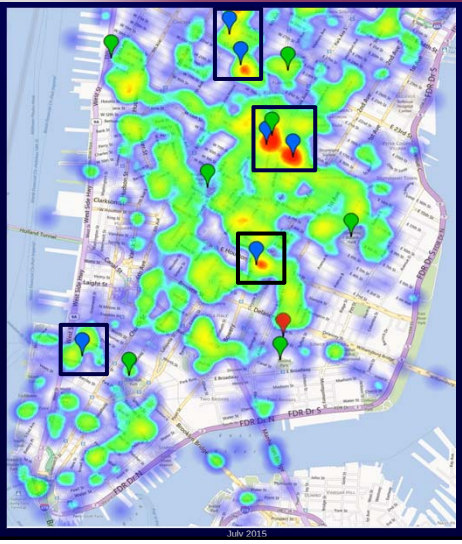
VACCINE

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

July 2015

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

VACCINE

July 2015

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



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

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

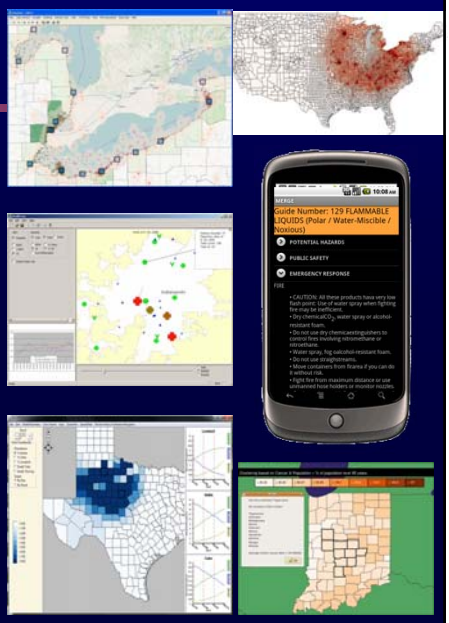
"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address



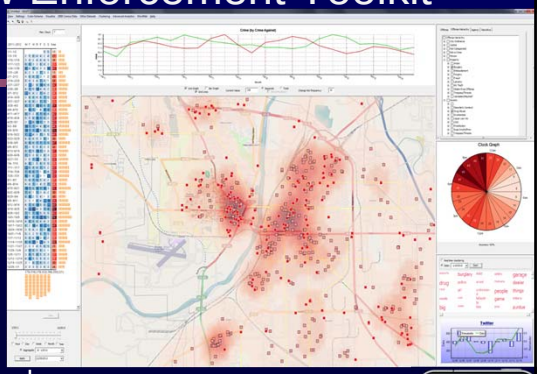
Public Safety & Health

- 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

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

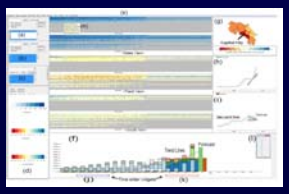


VACCINE

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



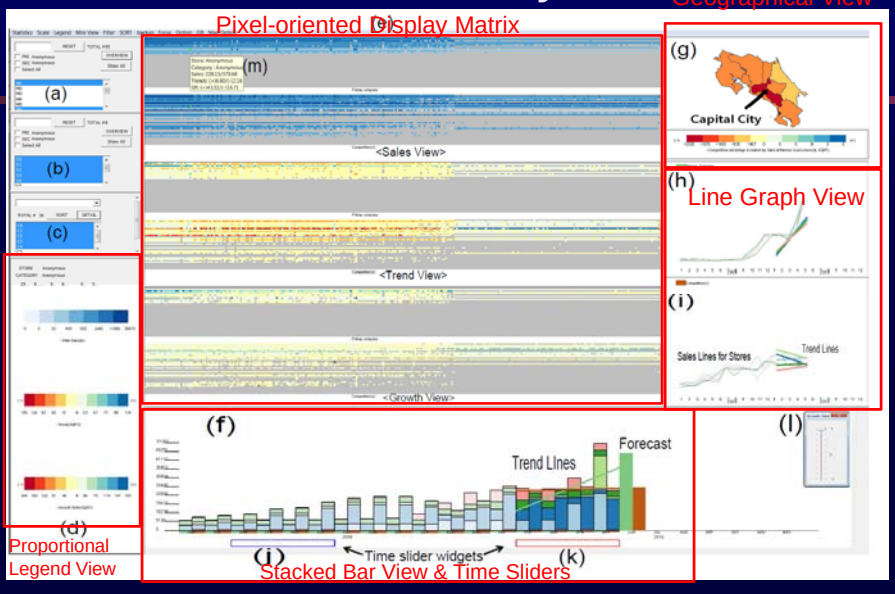




VACCINE

July 2015

Market Analyzer



VACCINE

July 2015

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

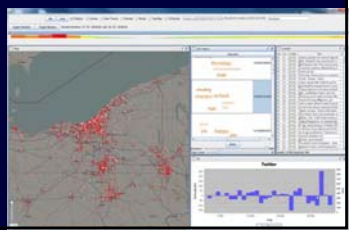


VACCINE

Visual Analytics Uses for Public Safety



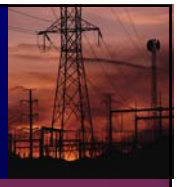
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Multisource, multimedia massive data integration & analysis
- Purpose: Planning for resiliency, training, detection, investigation, response, recovery, remediation



VACCINE

July 2015

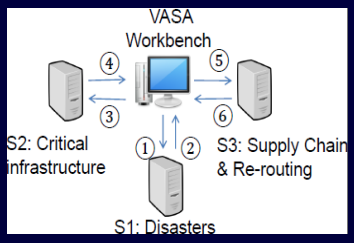
Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



- 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



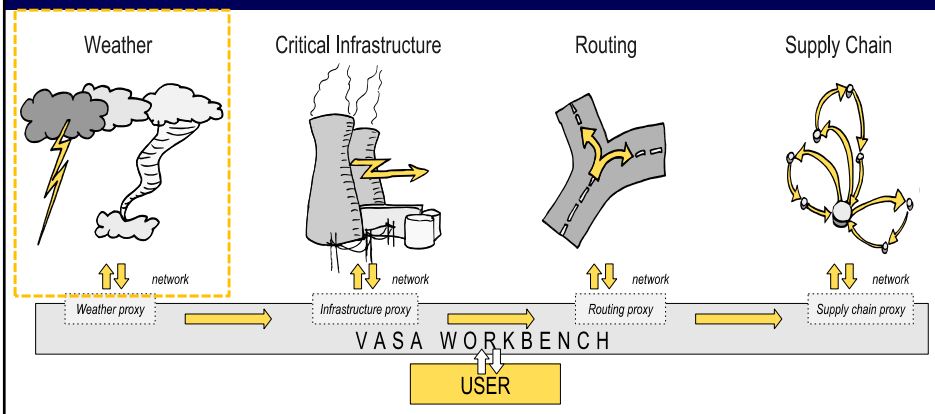
U Minn

TACC

VACCINE

July 2015

VASA COMPONENT: **CRITICAL INFRASTRUCTURE**



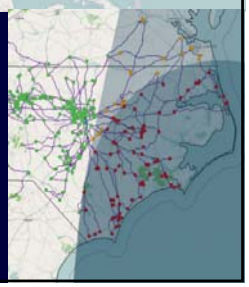
VASA: Visual Analytics for Simulation-Based Action

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



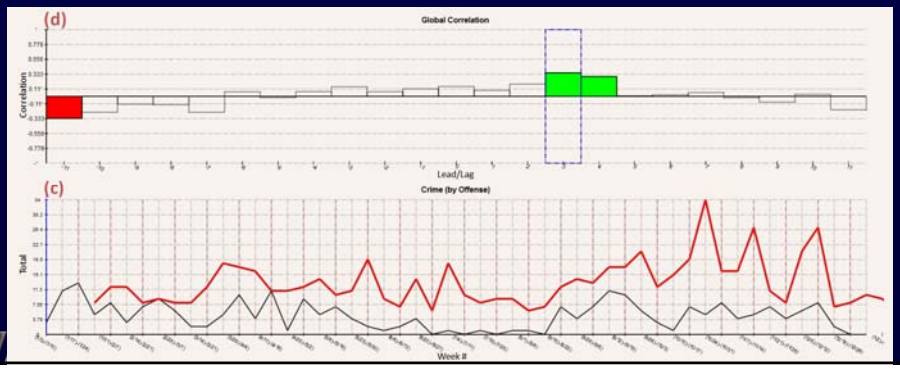


VACCINE

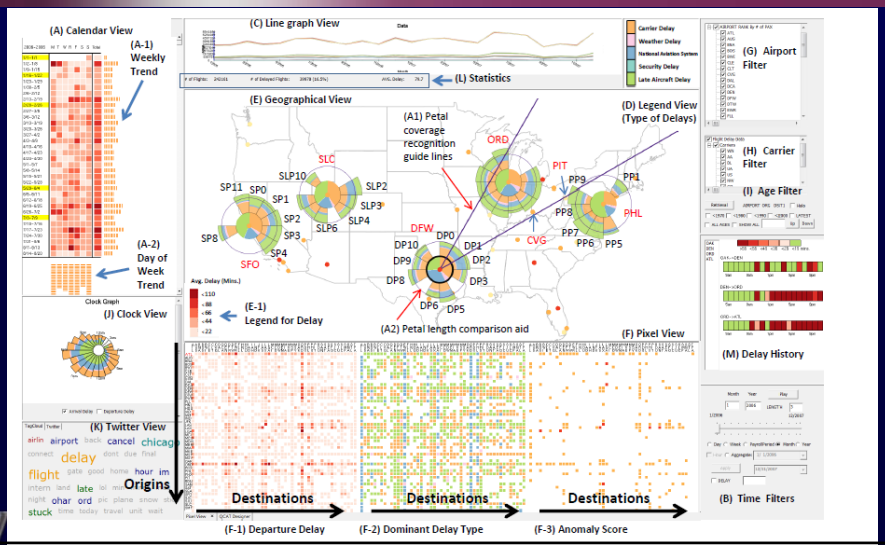
July 2015

Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space



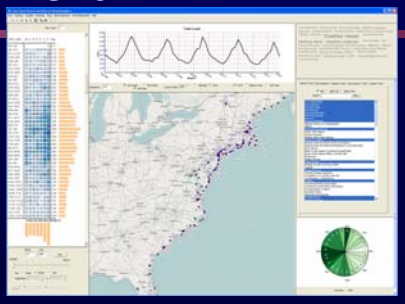
Multivariate Network Analytics with Information Theory Anomaly Detection



Visual Analytics Environment

Design Requirements:

- Interaction with the system.
- Integration with previously used tool.
- Creation of customizable scenarios.



VACCINE

14

Scenario

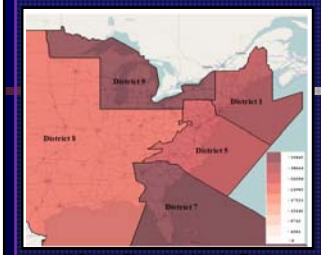
Questions that drive the planning strategy:

- 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?

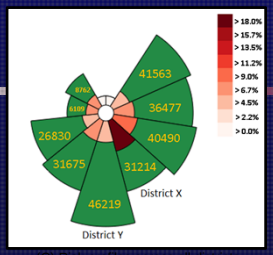
VACCINE

July 2015

Identify and Analyze Potential Risk



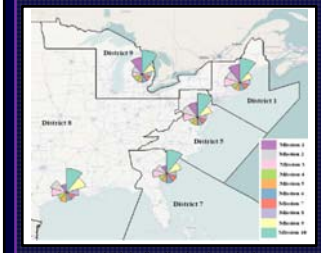
(A) District Choropleth Map for Total Risk across all missions



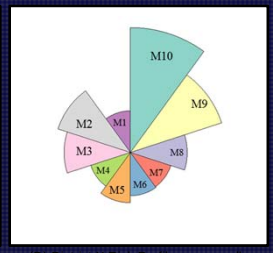
(C) Risk profiles across all districts and mission areas



(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.

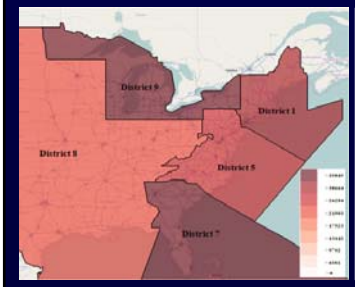
Specific areas can be analyzed by time and the analyst can obtain details by demand.

VACCINE

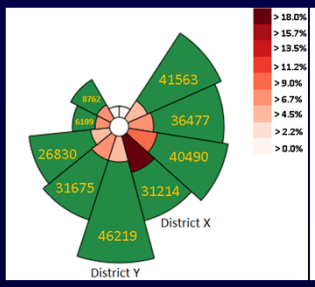
July 2015

Component Overview: Identification of Potential Risks

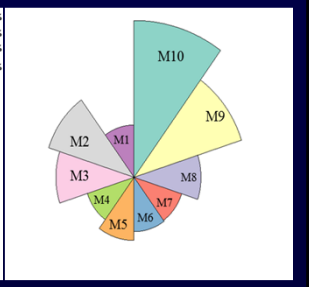
General understanding on how the risk is distributed between the different areas and missions.



(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

VACCINE

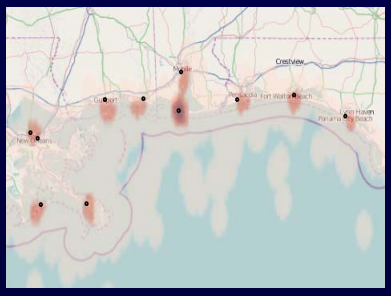
July 2015

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

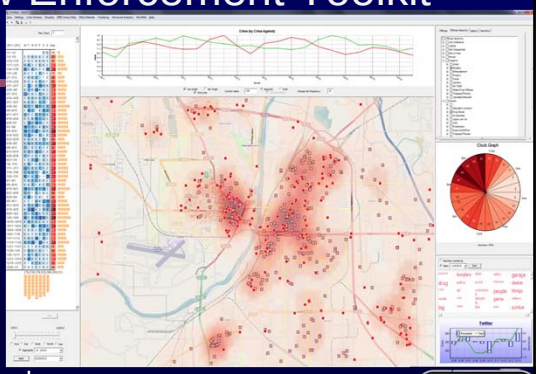


(E) KDE heatmap showing Total M10 Risk values

VACCINE

July 2015

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:

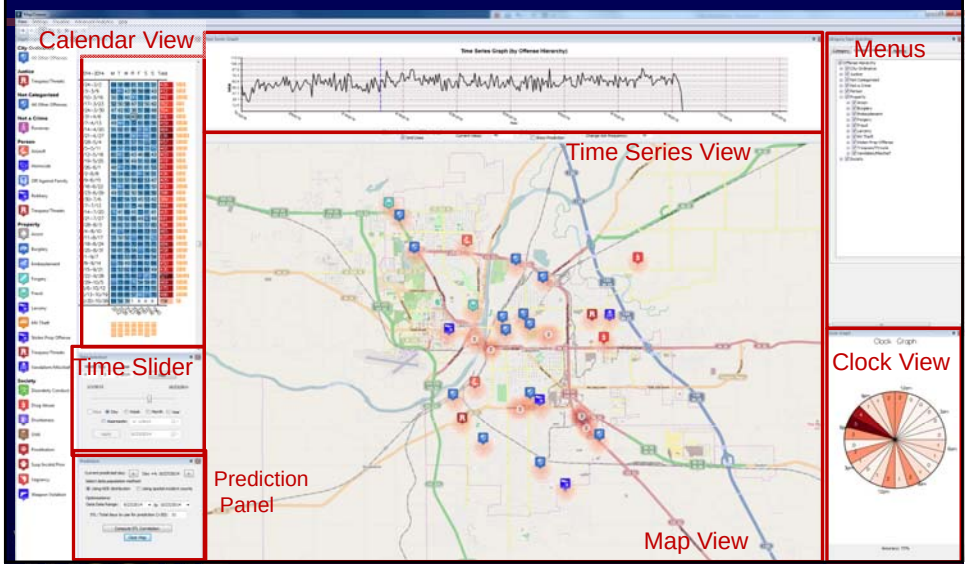
- October 2011: Purdue, WL Police, Lafayette PD



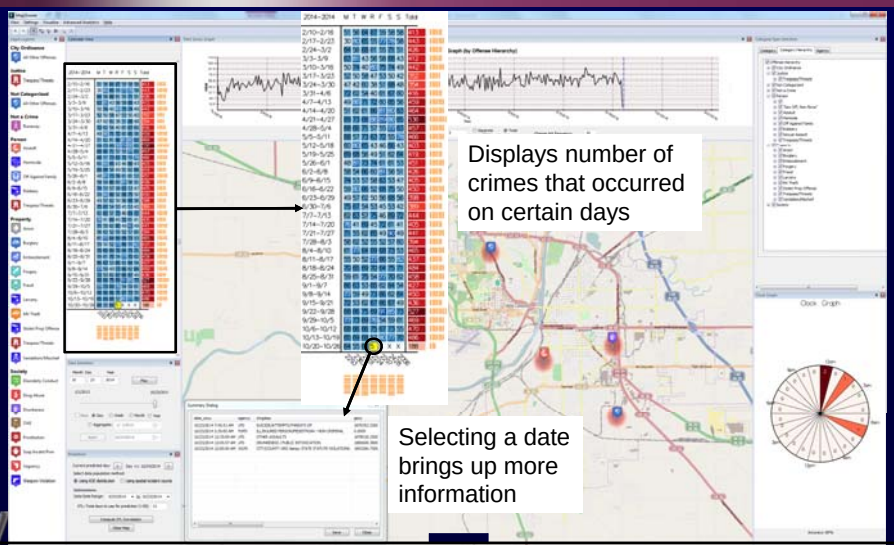
VACCINE

July 2015

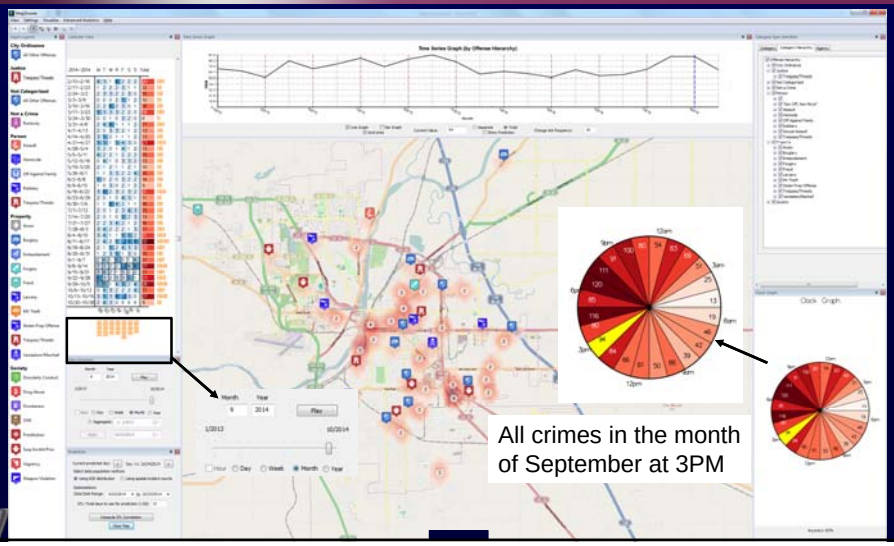
VALET Layout

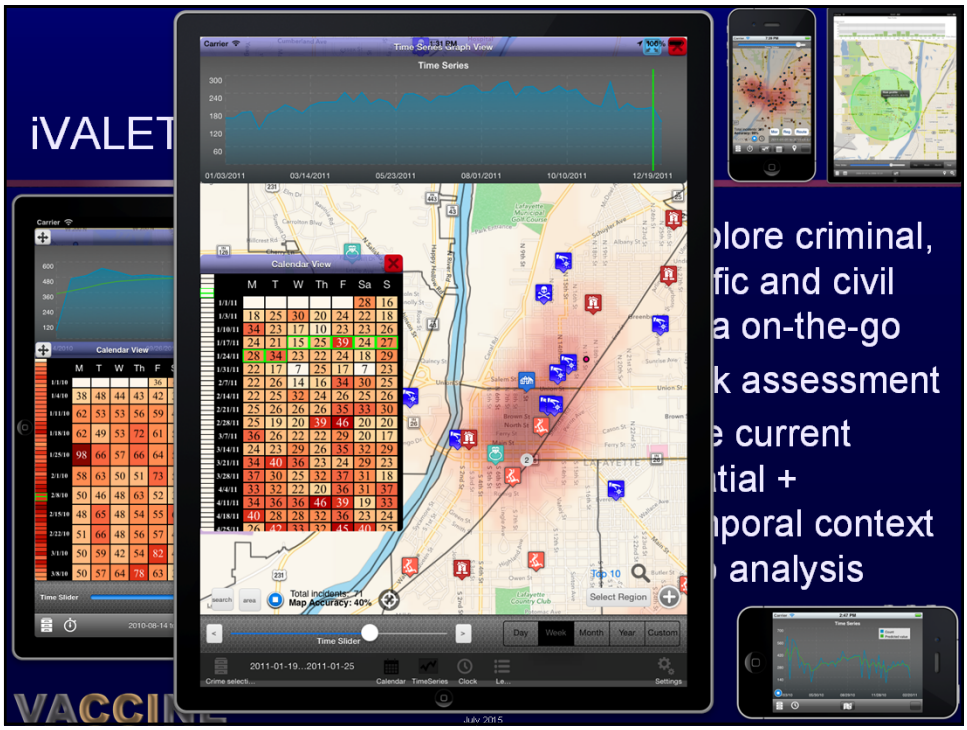


Calendar Display



Crimes During Certain Hours In A Month



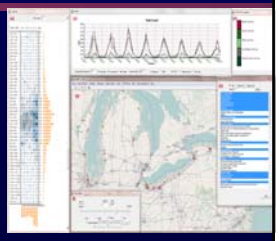


U.S. Coast Guard COAST/ SARVA (cgSARVA)

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



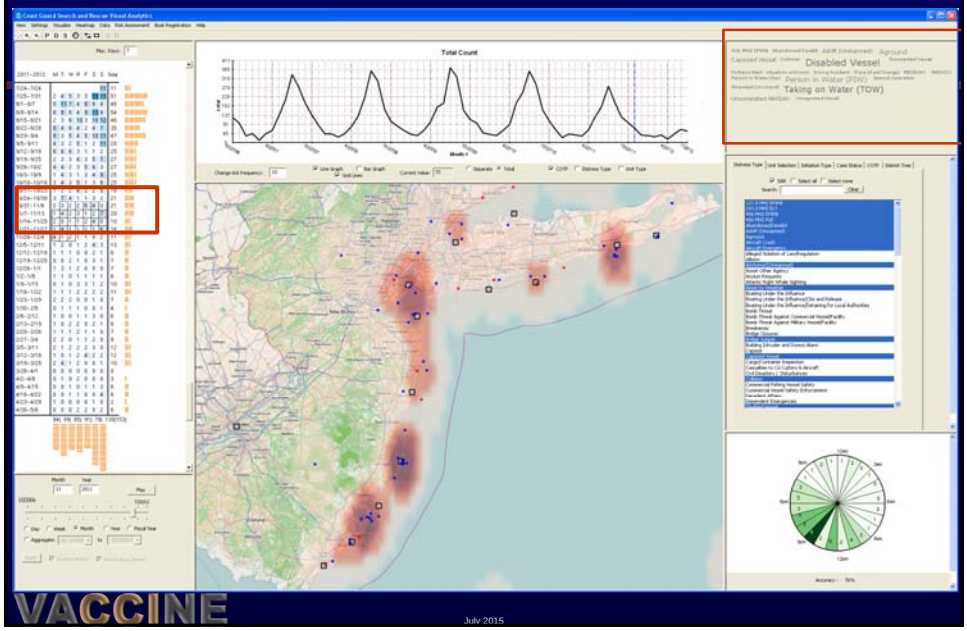


VACCINE

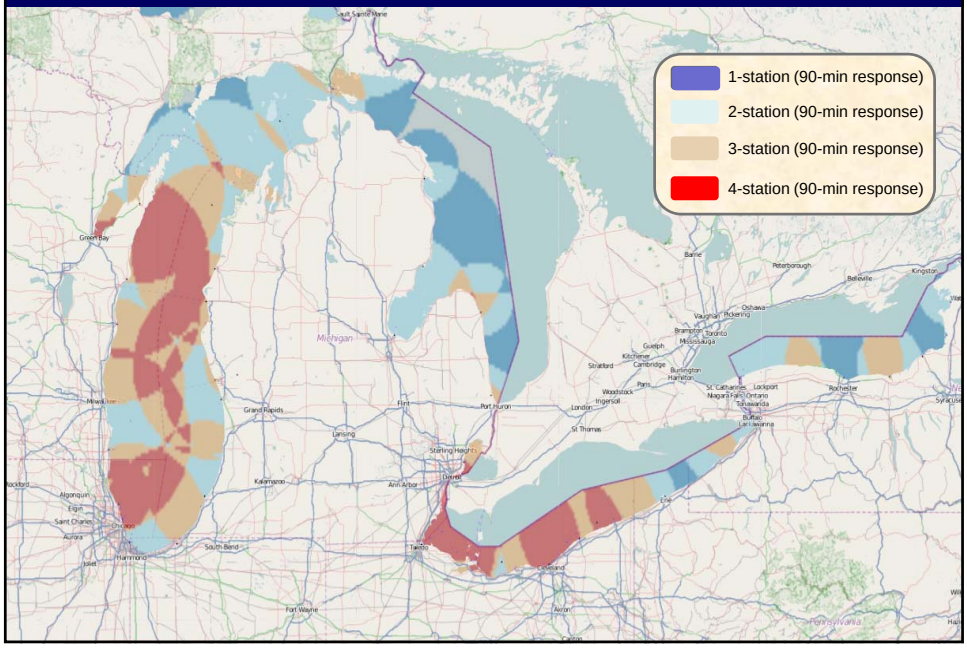
July 2015

Example: Risks and Consequences From Sandy:

SAR Cases November 2011 NJ/NYC Area



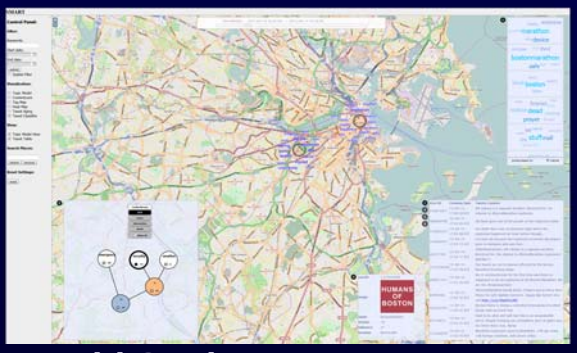
Response Efficiency – Potential Future Assets



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

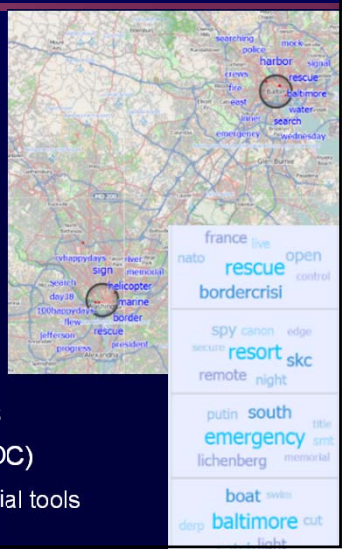
July 2015

Social Media Analysis and Reporting Tool (SMART)

Partners: Purdue, U. Stuttgart, Penn State, USCG LANT, PAC, D8; Purdue Police, IMPD, BSA

IMPACTS:

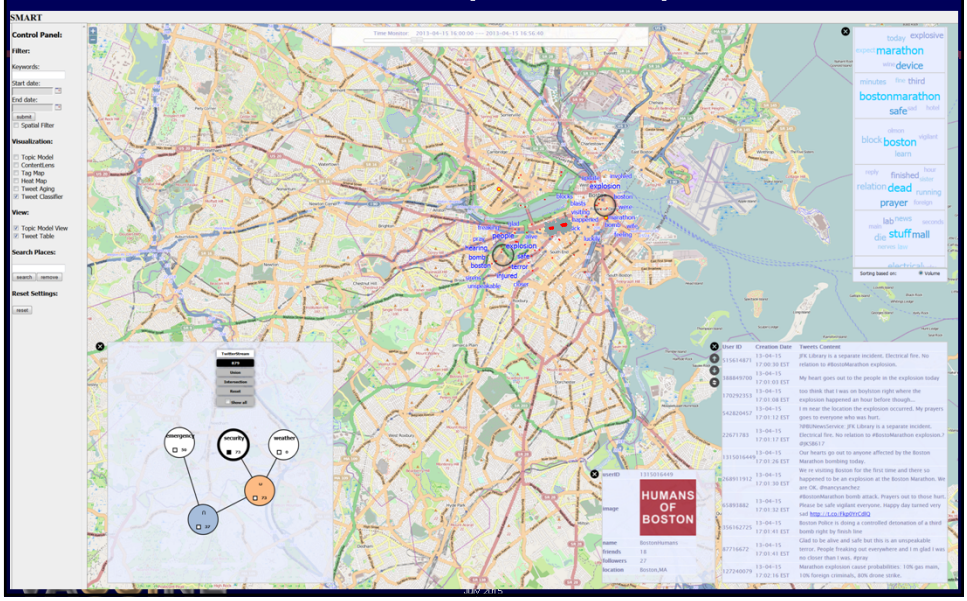
- Used at Boy Scouts of America Jamboree 2013
  - Detected tweet about sniper
- Used by U.S. Coast Guard
  - District 8 for events in 2014 -Detected 3 gang related activities to date
  - PAC for Fleet Week October 2014
  - LANT for SAR hoax call investigations
- Purdue Police for home football games 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

July 2015

Social Media Analysis and Reporting Toolkit (SMART)



Anomaly Detection Superbowl Weekend



Examples From Fleet Wee



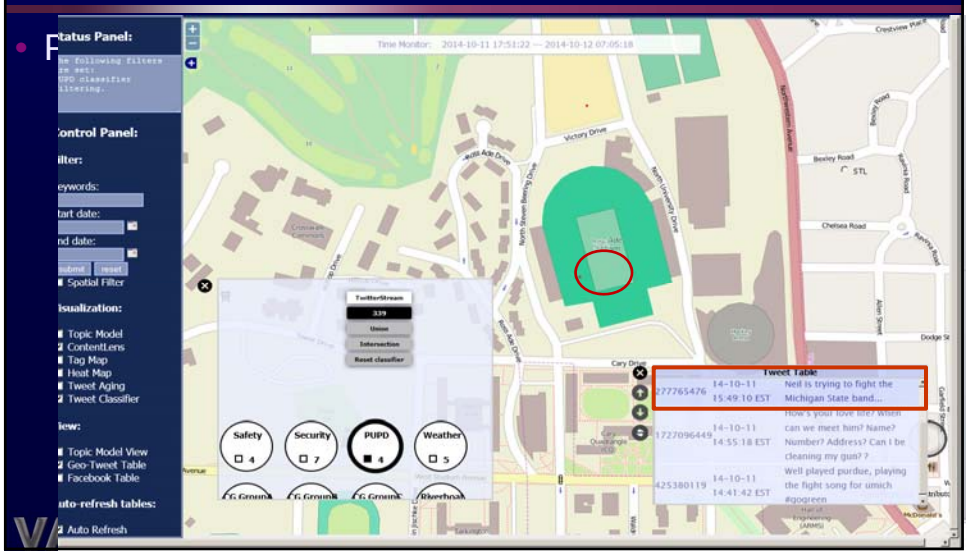
Where Can It Be Helpful?

- 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

VACCINE July 2015

Example From Purdue Football 10/12/14

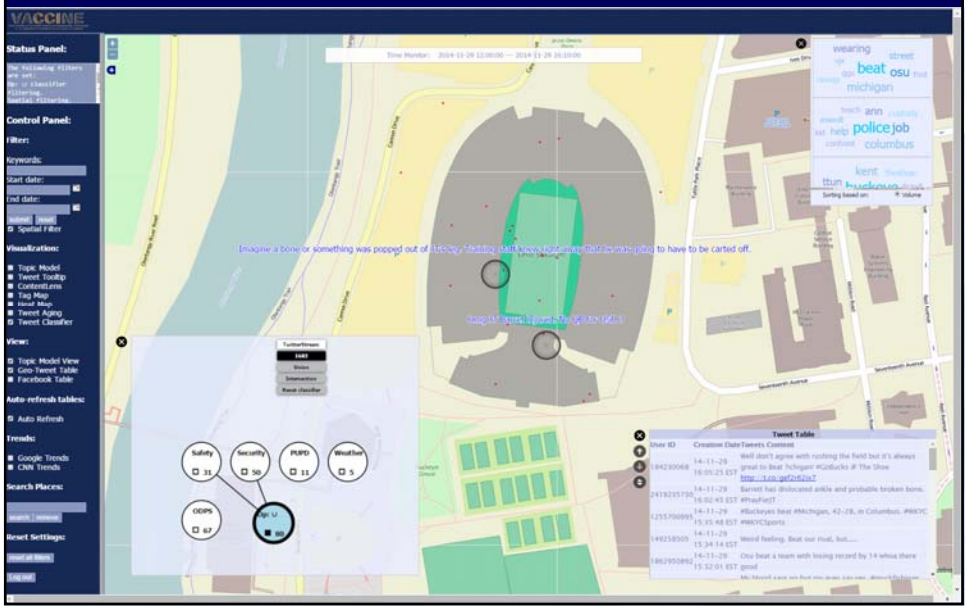
- F



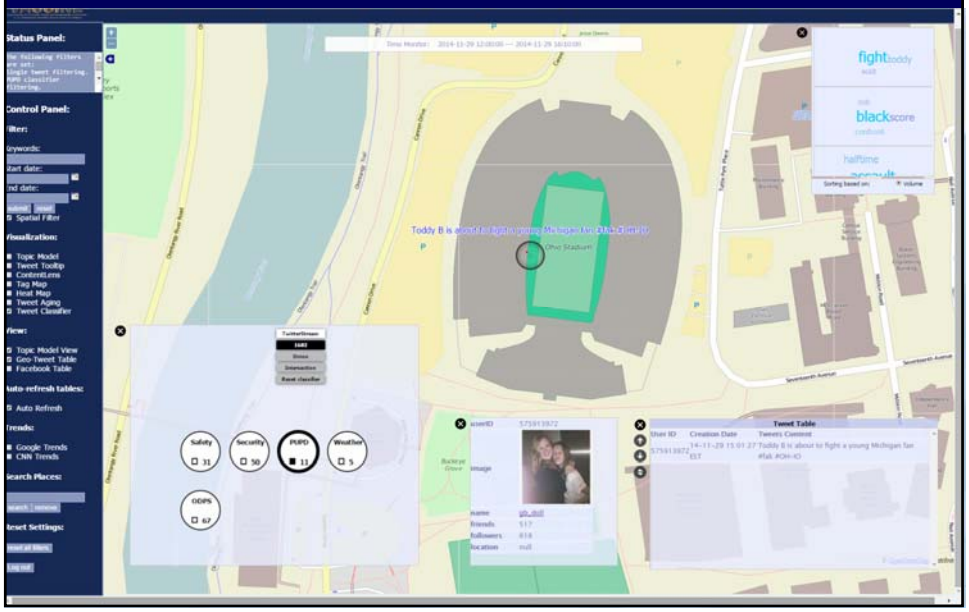
Example From Ohio State Football 11/29/14



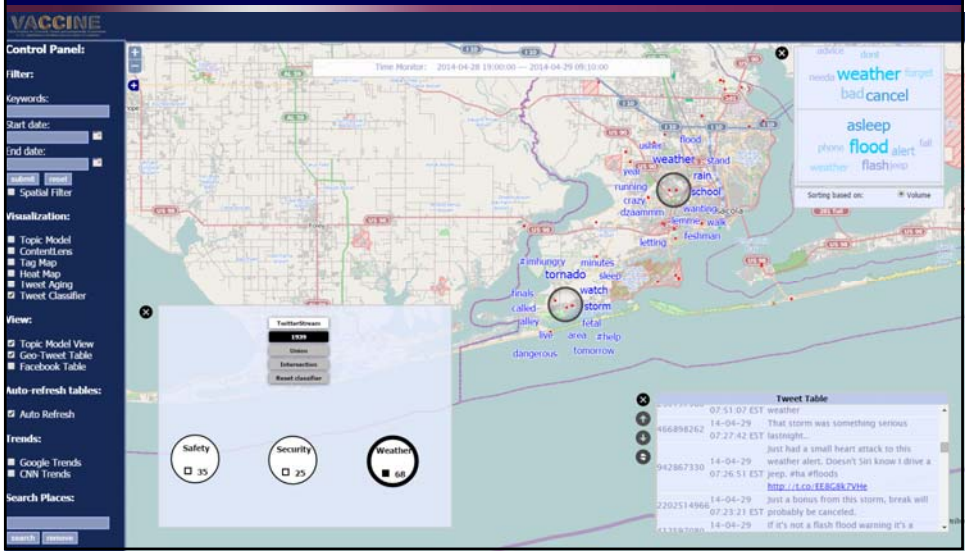
Example From Ohio State Football 11/29/14



Example From Ohio State Football 11/29/14



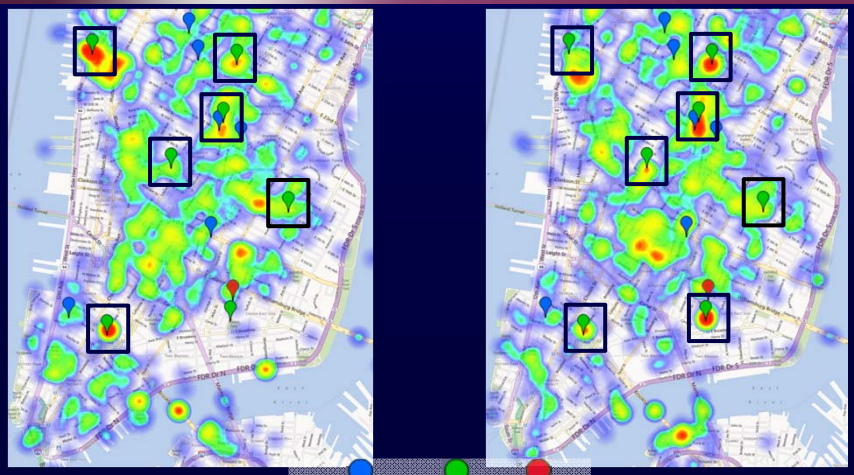
Where Can It Be Helpful?



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



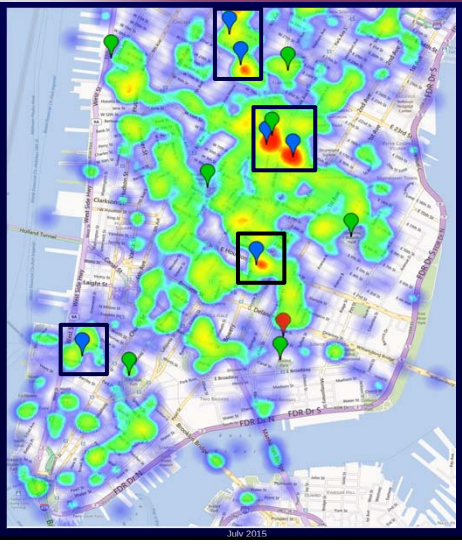
VACCINE

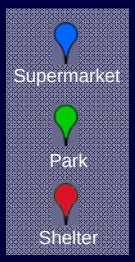
Visual Analytics of Activity During Hurricane Sandy

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

July 2015

Where Can It Be Helpful?

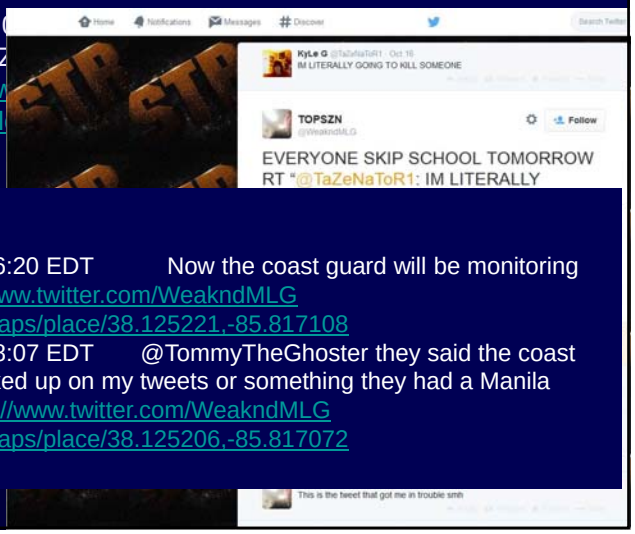
- During large-scale planned events, e.g.,
  - Superbowl 2014 in NJ
  - Riverboat Festival in Louisville October 2014
  - Fleet Week 2014 in San Francisco
- During recurring events
  - Purdue football games – the poor tuba player
- During unexpected events
  - Boston Marathon bombing
  - Pensacola tornado
  - Keene, NH Pumpkin Festival
- Detecting daily issues
  - Teen threatening violence – Louisville, KY

VACCINE

July 2015

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



Operations Performance Assessment Report

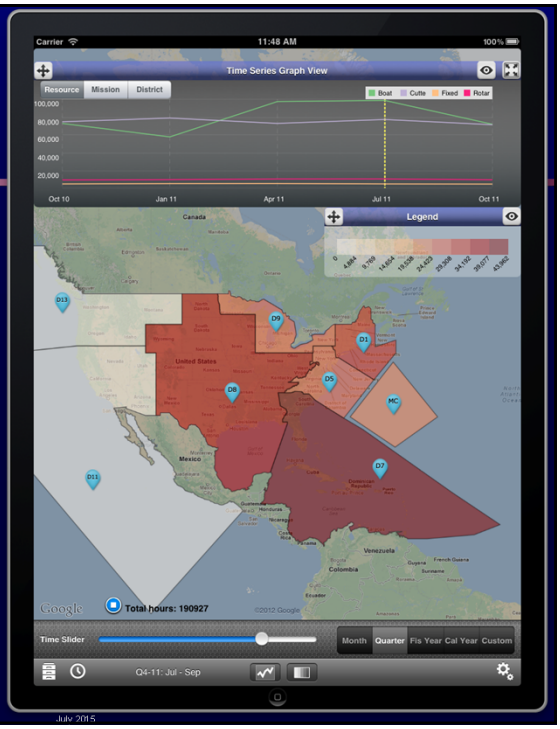
OPAR FEATURES SUMMARY



VACCINE

July 2015

- iOPAR map view and time series view
- Eg: Showing mission hour distribution over districts for Q4-2011
  - One look tells which district is driving the hours for the selected time period
- Time series show hours of resources usage across time
  - (All mission, resources and districts selected)



VACCINE

July 2015

Time series full screen view

- Shows hours of resources usage across time.
- (Can be changed to see usage by missions, districts)



VACCINE

July 2015

Grouped time series

- Group time periods (months, quarters) and compare across years.
- Eg: Here, looking at how each quarter uses the resources across years.



VACCINE

July 2015



- Tap on district to show overview of performance across years and comparing to target
  - Tap on ‘i’ to show more detail





- Comparing hours usage to previous years and the target.
- Target hours can be broken down by resources.

VACCINE

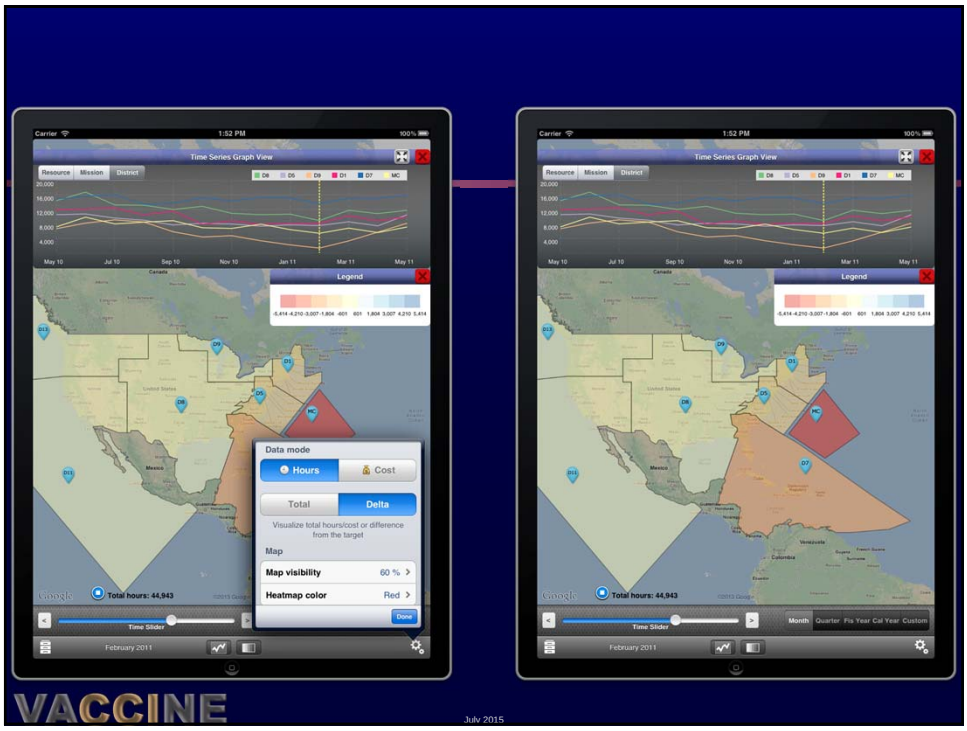
July 2015

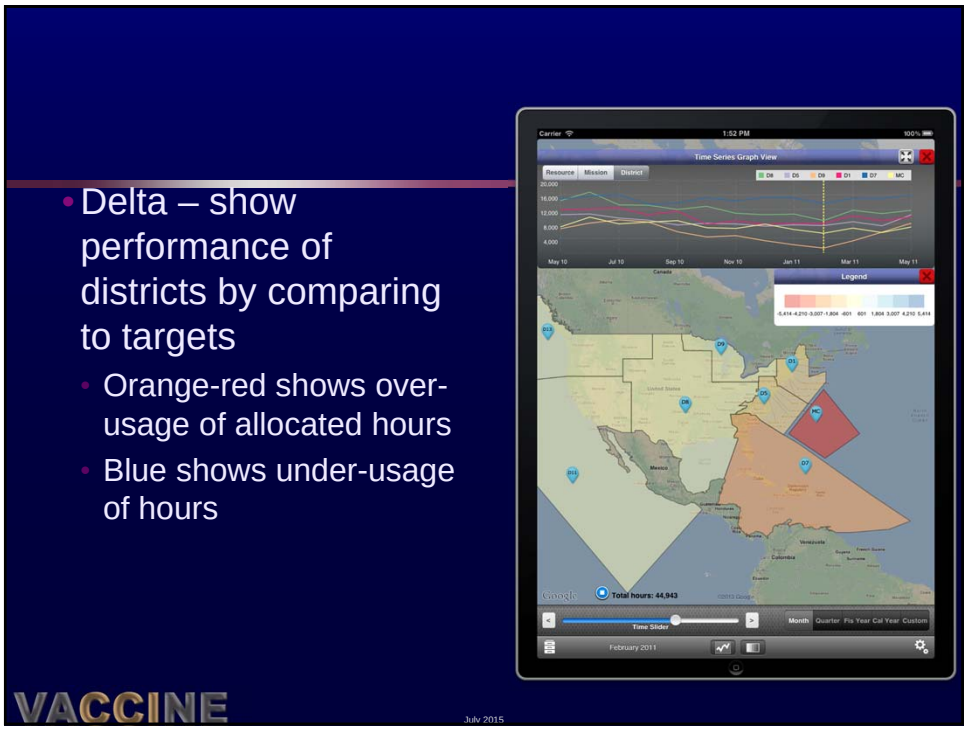
- Target hours can also be broken down by missions
- Each mission is compared against the target for the selected time aggregation level (red bar)



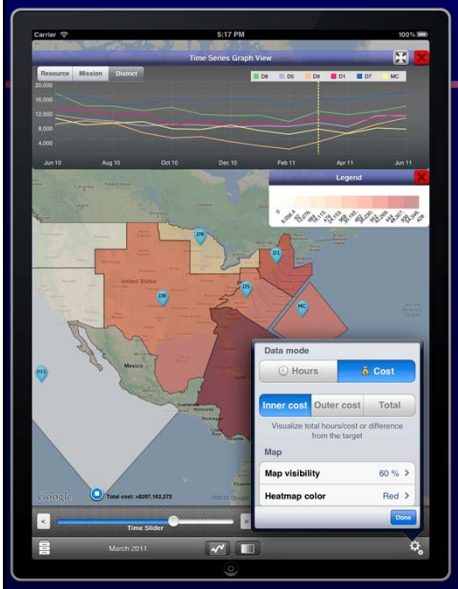
VACCINE

July 2015





- Delta – show performance of districts by comparing to targets
  - Orange-red shows over-usage of allocated hours
  - Blue shows under-usage of hours

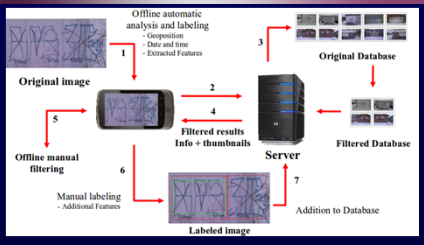


- Cost – calculate the total cost (inner/outer/sum of both) associated with each district

VACCINE

July 2015

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:

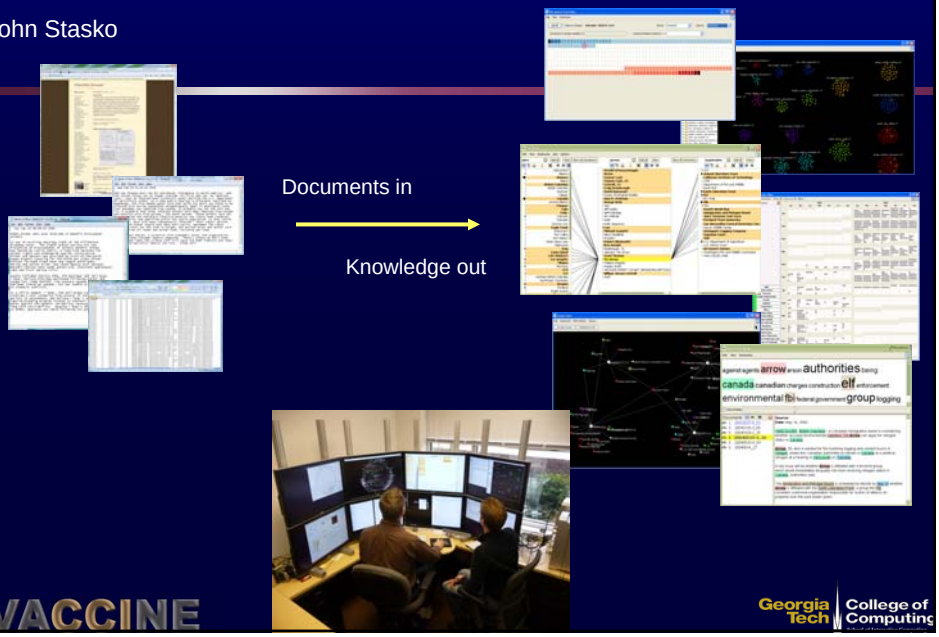
- Summer/Fall 2011:
  - IPD gang detectives
  - IPD at large
  - Ind Fusion Gang Task Force
  - Gang detectives across Indiana
- Spring/Summer 2013:
  - Cook County, IL
  - INGang ownership
  - Tattoo database added

VACCINE

July 2015

Jigsaw: Visual Analytics for Investigative Analysis

John Stasko



VACCINE

Georgia Tech College of Computing

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

July 2015

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

VACCINE

July 2015

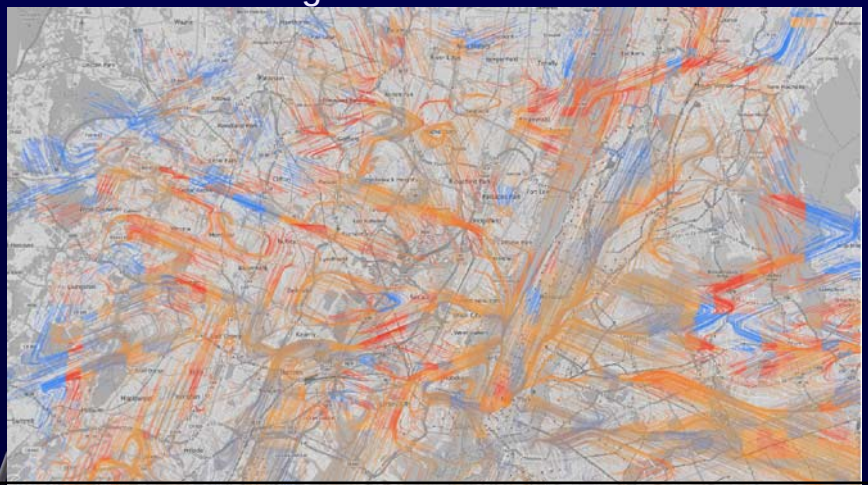
Ideas and Purdue Predictive and Financial VA Work

David Ebert
VACCINE Team

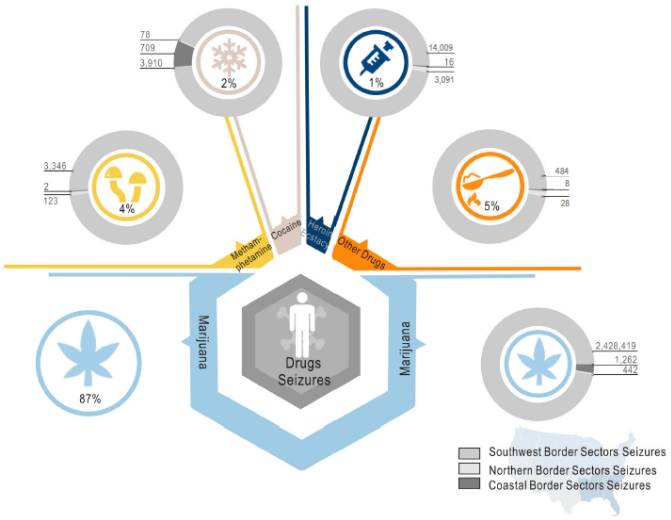
June 2016

Flows and Networks:
Evolving Flow VA (Movement Patterns from Social Media Data)

- Manhattan morning rush hour



State of the Border Examples: Drugs

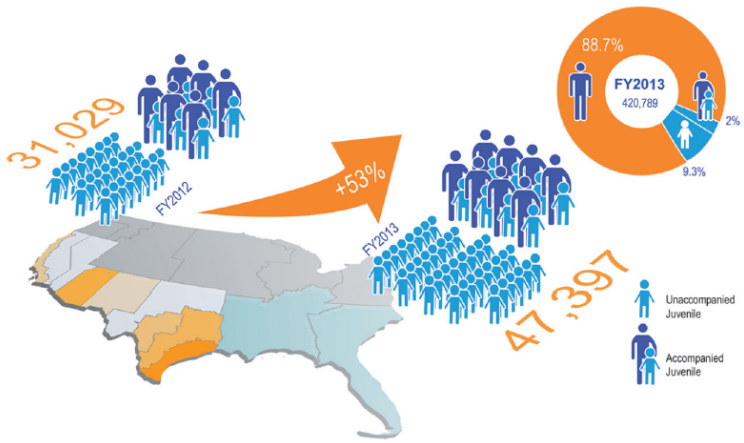


U.S. Border Patrol – State of the Border – Page 5

From FY 2012 to FY 2013, total pounds of illegal drugs and narcotics seized increased by 5 percent, with Marijuana accounting for 6 percent increase from the previous fiscal year. However, total ounces of Heroin seized increased by 34 percent, while pounds of Cocaine seized decreased by 61 percent. Graphic courtesy of VACCINE, Purdue University

VACCI

State of the Border: People

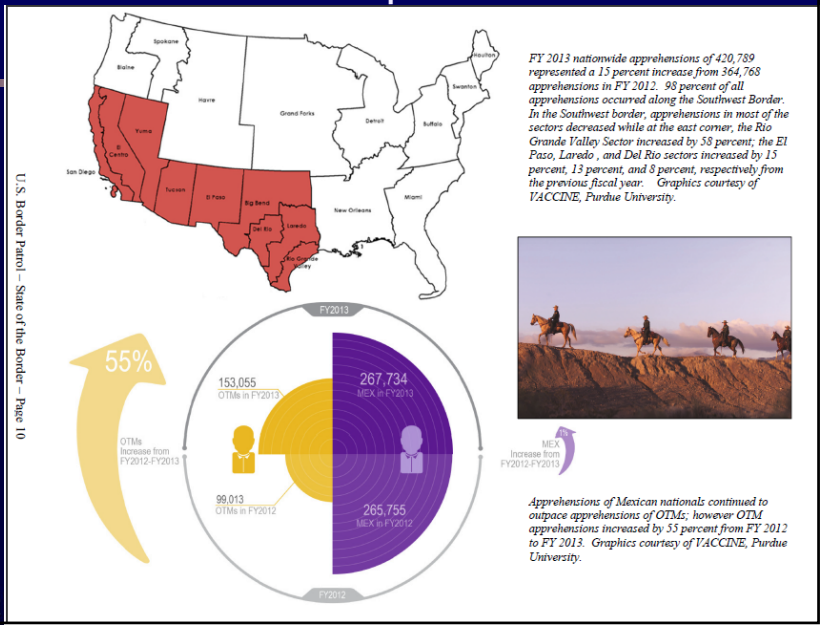


U.S. Border Patrol – State of the Border – Page 6

Juveniles accounted for 11.3 percent of all Border Patrol apprehensions in FY 2013. However, there was a 53 percent increase in juvenile apprehensions from the previous fiscal year. Graphic courtesy of VACCINE, Purdue University.

VACCI

State of the Bortder: People



FY 2013 nationwide apprehensions of 420,789 represented a 15 percent increase from 364,768 apprehensions in FY 2012. 98 percent of all apprehensions occurred along the Southwest Border. In the Southwest border, apprehensions in most of the sectors decreased while at the east corner, the Rio Grande Valley Sector increased by 58 percent; the El Paso, Laredo , and Del Rio sectors increased by 15 percent, 13 percent, and 8 percent, respectively from the previous fiscal year. Graphics courtesy of VACCINE, Purdue University.

Apprehensions of Mexican nationals continued to outpace apprehensions of OTMs; however OTM apprehensions increased by 55 percent from FY 2012 to FY 2013. Graphics courtesy of VACCINE, Purdue University.

VACCINE

Key Themes/Capabilities

- Time, Space, and abstract dimensions equally manipulable, interactive, and analyzed
- Dealing with Scales: Appropriate and varied aggregation and analysis
- Cross dimension and scale analysis
- Multiple visual choices to highlight different types of patterns.

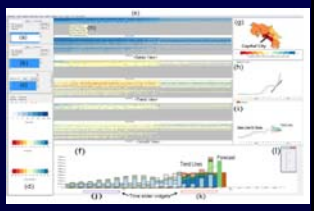
VACCINE

June 2016

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

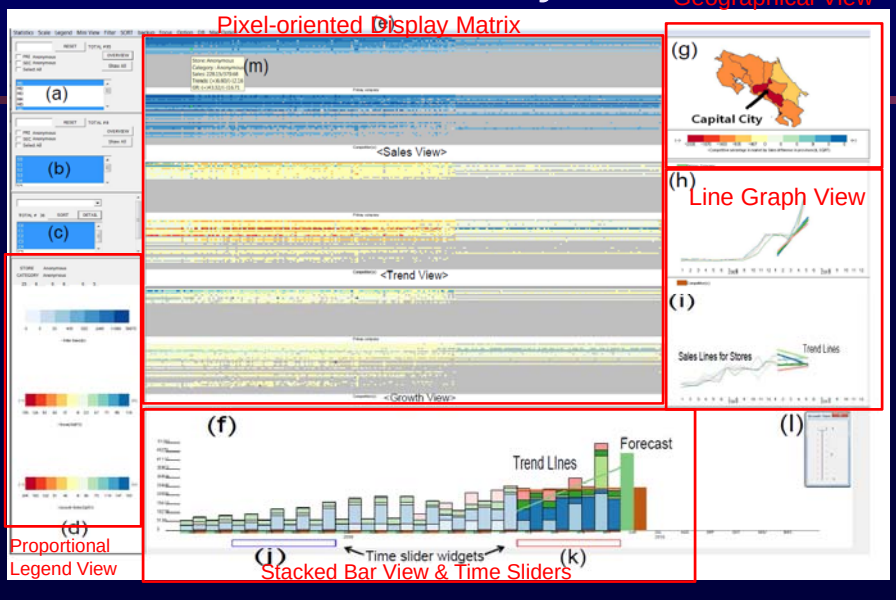






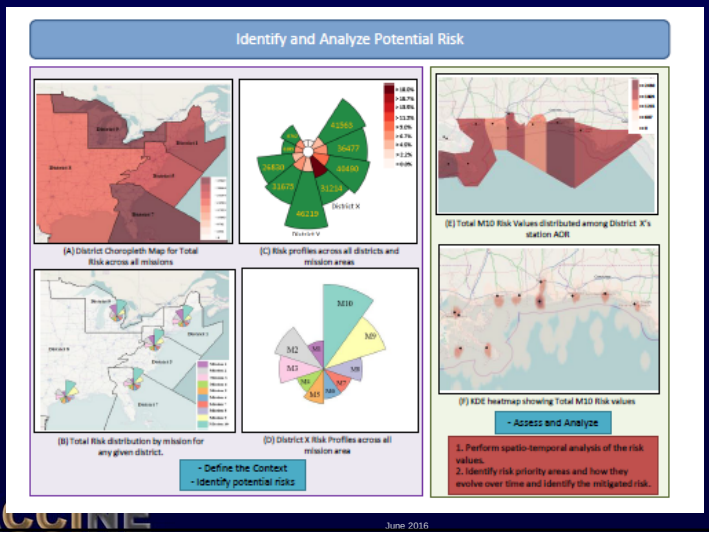
VACCINE

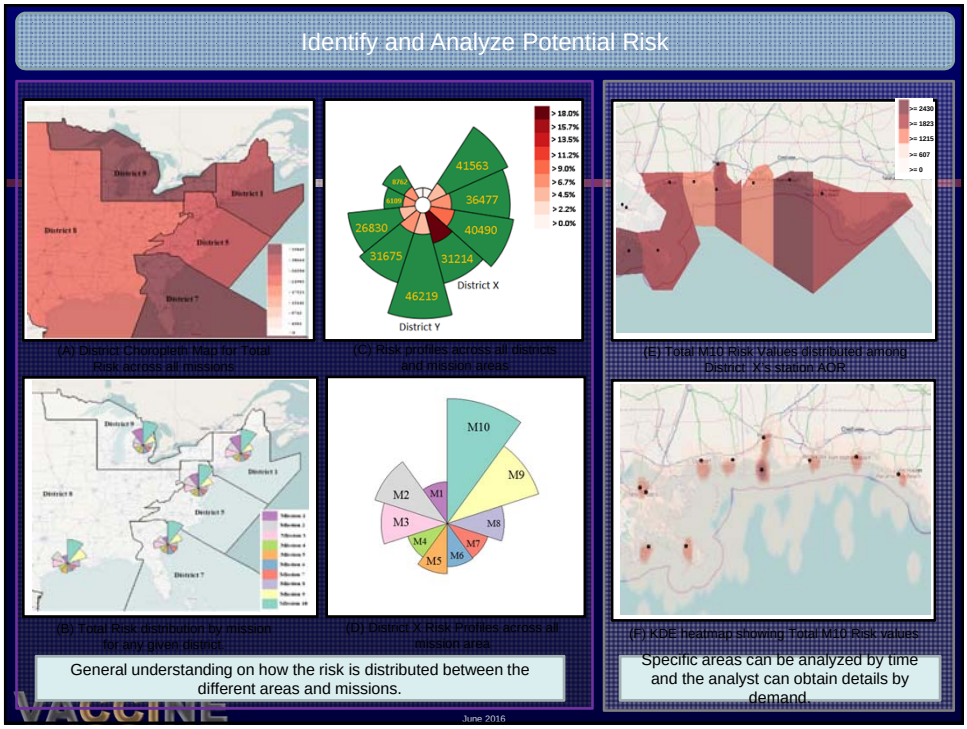
Market Analyzer



VACCINE

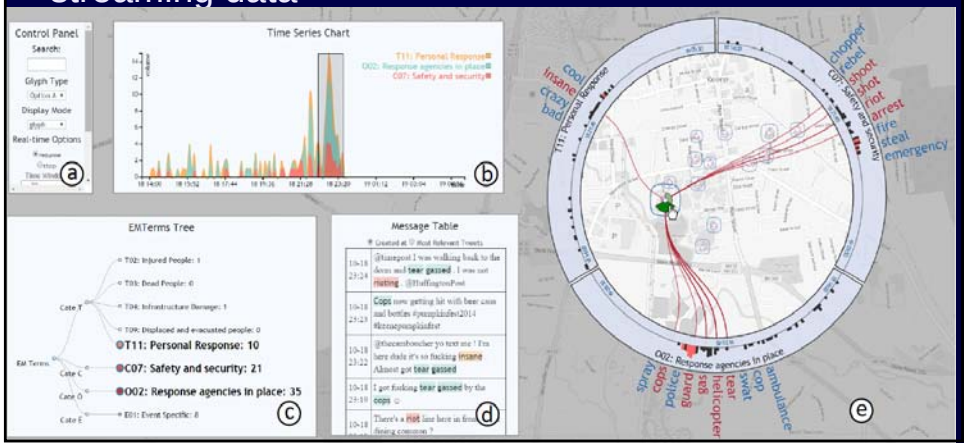
Multi-Visualization, Multivariate VA for Risk-Based Decision Making





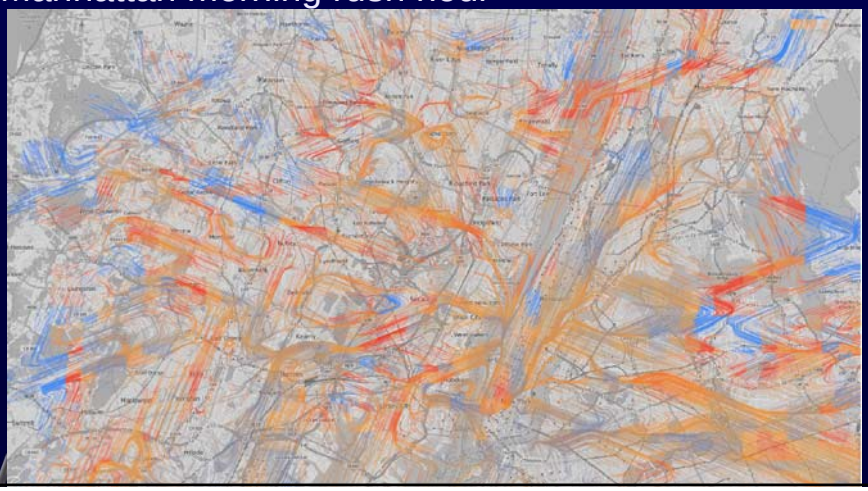
Multi-level Aggregation and VA of Evolving Temporal and Spatial Categorical/Topic Evolution (Eurovis 2016)

- Real-time categorization, analysis, & aggregation of streaming data

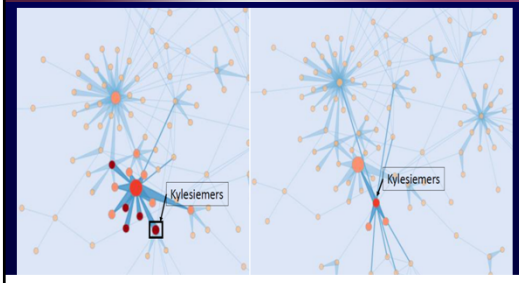


Flows and Networks:
Evolving Flow VA (Movement Patterns from Social Media Data)

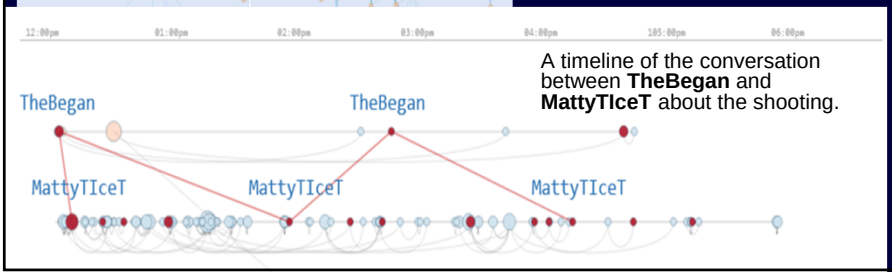
- Manhattan morning rush hour



Flows and Networks:
Evolving and Dynamic Social Networks(Identify Influential Users)



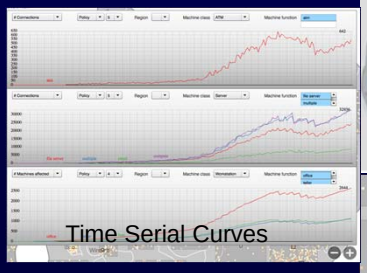
Click **JConline** to show its information traffic, then click the darker nodes along the path: **Jconline** -> **Kylesiemers**. **Kylesiemers** has no darker nodes connected to it. Identify **Kylesiemers** as an influential node and monitor his/her messages.



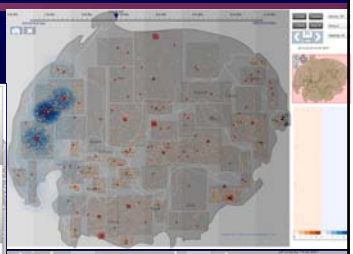
A timeline of the conversation between **TheBegan** and **MattyTIceT** about the shooting.

SemanticPrism: A Multi-aspect View of Large High-dimensional Data (Purdue University)

- **Geo-Temporal**
- **Time-serial**
- **Pixel-based**
- **Semantic Zoom**



Time Serial Curves

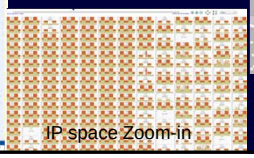




Geo-Temporal Zoom-in



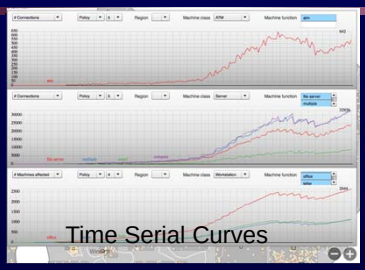
Pixel-based IP space

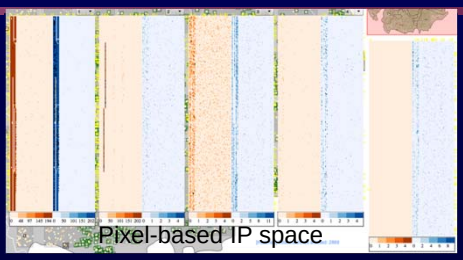


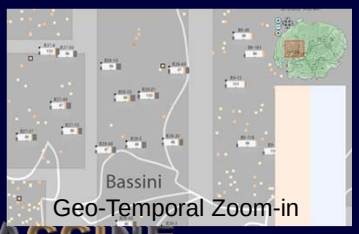
IP space Zoom-in

Victor Yingjie Chen, Ahmad M Razip, Sungahn Ko, Cheryl Zhenyu Qian, David S Ebert

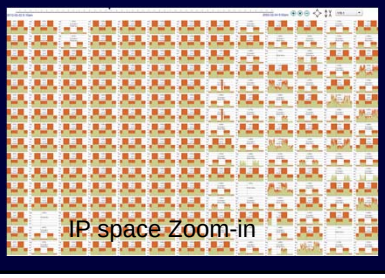
SemanticPrism


Time Serial Curves


Pixel-based IP space



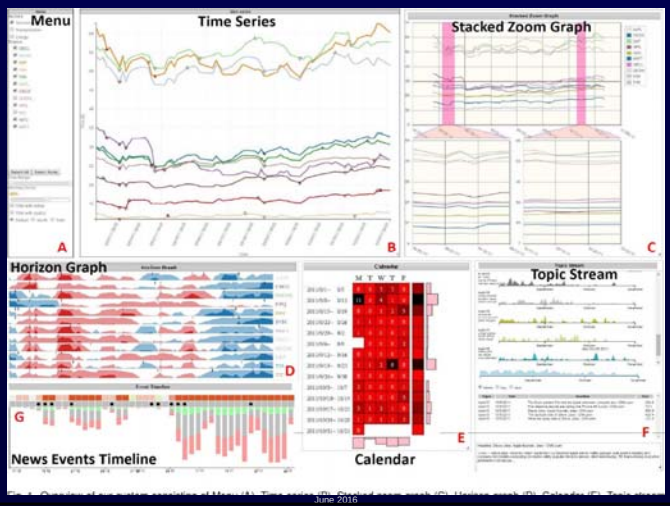
Geo-Temporal Zoom-in


IP space Zoom-in

VACCINE

June 2016

PIKE System: Financial Anomaly VA
(VA by Purdue and UNCC)



June 2016

PIKE System

- Visual Analytics system for exploring and explaining financial data and anomalies



VACCINE

Linking between views



VACCINE

June 2016

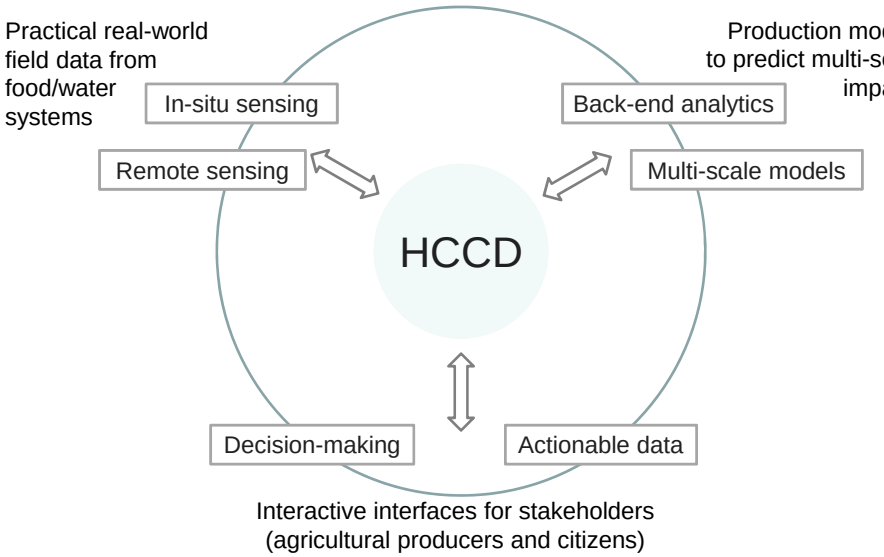
Use Case 1 - Banking Data



Technology and Information Fusion Needs to Address the Food, Energy, Water Systems (FEWS) Nexus Challenges

Napa, California
November 2015

FEW: Technology and Information Fusion Needs to Address the Food, Energy, Water Systems (FEWS) Nexus Challenges



Investigators: David Ebert (PI), Christian Butzke, Phillip Owens, Melba Crawford, Dimitrios Peroulis

Why Are We Here?

CRUSH: "Of Pixels and Palates: Can Geospatial Technologies Help Produce a Better Wine?"

CRUSH Project



Sculptor: Gene Miles
Photo: Brad Lobitz

CRUSH PROJECT

Lee Johnson[1] (Principal Investigator), Daniel Bosch[2], Paul Skinner[3], Philip Freese[4], Kevin Spry[5], Brad Lobitz[6], David Peterson[7]

During and after the 1997 growing season, the NASA/Ames Ecosystem Science and Technology Branch (ECOSAT) (Mountain View CA) teamed with the Robert Mondavi Winery (Oakville CA) and Wine Grow Consulting (Healdsburg CA) to evaluate the technical and economic feasibility of using airborne images to improve crop uniformity in winegrape harvest. In addition, ECOSAT teamed with TerraSpase Vineyard Mapping Services (Napa CA) to provide image-based data and information to clients in the greater North Coast region. Two-meter digital multispectral images were collected in late July 1997 by aircraft over vineyards in Napa and Sonoma Counties by partner CropImage (Salinas). NASA participation in the Canopy Remote-sensing for Uniformly Segmented Harvest (CRUSH) project was sponsored by the Ames Commercial Technology Office under the Commercial Technology Project Fund. The Fund is intended to stimulate the transition of Ames technology to the private sector.

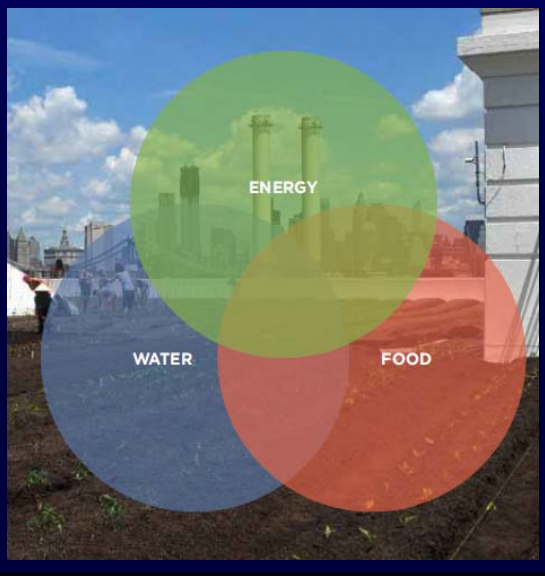
FEWS

Looking at the linked systems and the Nexus

- For instance food and water



FEWS Nexus



FEWS and Technology

- Scientific research in environmental science, energy science, agriculture, computer science, and electrical engineering is resulting in improved models and creating new technology for in-situ and remote sensing for monitoring our environment with much higher spatial and temporal resolution and in novel ways.
- These advances have created unprecedented growth in our capability to generate and access data about the FEWS.

FEWS and Technology

- However, much work is required to focus, adapt, and evolve these research advances into solutions for the FEWS challenges, as well as to develop engineered solutions that will achieve economic and ecologically sustainable impacts.

FEWS, Technology and NSF GEO Potential Areas of Inquiry for Research

- Where do water resources face the greatest stress and uncertainty?
- What technological innovations in engineering and agricultural practices will prove most effective in achieving system resiliency and sustainability?
- How can integrated modeling efforts inform decision making with respect to seasonal high-risk-of-failure water quantity and quality?
- What impact will a changing climate have on the global and regional precipitation pattern?

Focus On The Following Three Major FEWS Challenges:

1. Developing novel targeted remote sensing and in-situ sensing technology that can be practically fielded and used in food and water system management.
2. Developing novel integrated hydrology, soil, microclimate, and plant/agricultural production models that interact accurately and across traditional scales for understanding local, regional, and national impacts.
3. Turning this developing and pending FEWS data deluge into usable, actionable information for agricultural producers, local and regional decision makers, and citizens.

Why Are We Here Version 2

- Help identify funding priorities for 2016: $75M

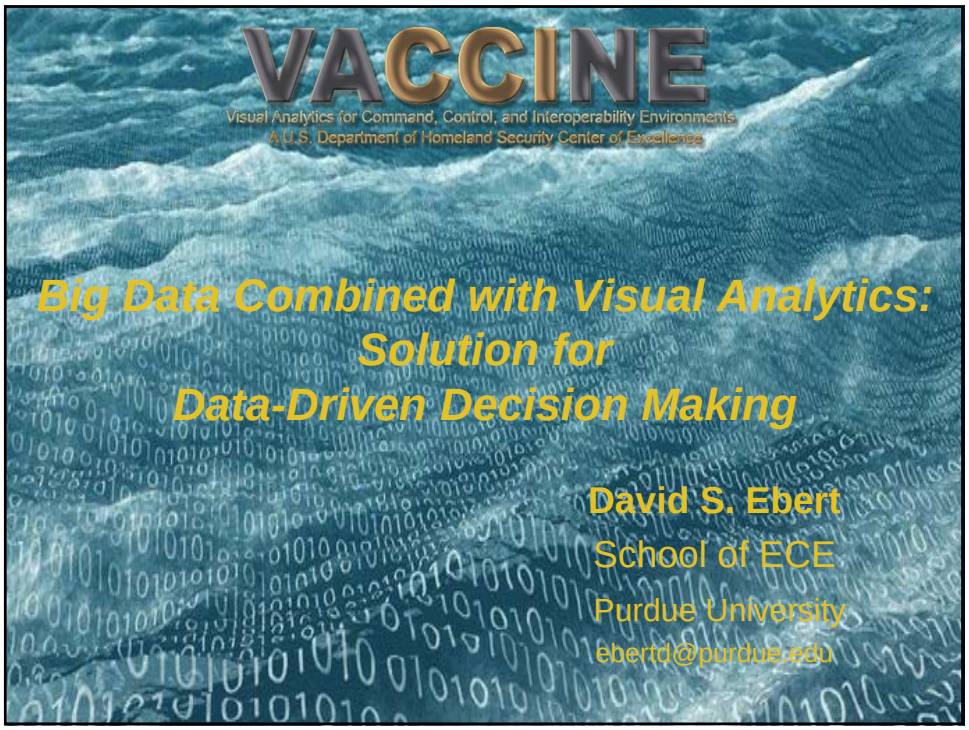
INFEWS Funding by Directorate

(Dollars in Millions)

Dir/Office	FY 2014 Actual	FY 2015 Estimate	FY 2016 Request
BIO	-	-	$7.50
CISE	-	-	13.50
EHR	-	-	6.00
ENG	-	-	13.00
GEO	-	-	14.78
MPS	-	-	8.90
OISE	-	-	1.28
SBE	-	-	5.00
IA	-	-	5.00
Total, INFEWS	**-**	**-**	**$74.96**

Totals may not add due to rounding.

Why Am I Here ?


VACCINE
Visual Analytics for Command, Control, and Interoperability Environments
A U.S. Department of Homeland Security Center of Excellence
Big Data Combined with Visual Analytics:
Solution for
Data-Driven Decision Making
David S. Ebert
School of ECE
Purdue University
ebertd@purdue.edu

Big Data Challenges

- Big data analytics and predictions don't always solve the problem



"Remember, the other team is counting on Big Data insights based on previous games. So, kick the ball with your other foot."

What Our Visual Analytic Solutions Offer Decision Makers

We enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

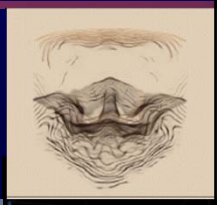
- Provide the right information, in the right format within the right time to solve the problem
- Turn data deluge into a pool of relevant, actionable knowledge
- Enable user to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

Scientific Discovery

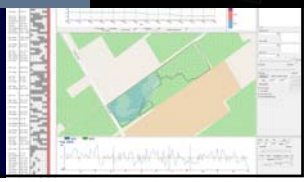


- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations
  - Innovative feature visualization
- Atmospheric Science
  - Multiple sensors, simulations, challenging
  - problems
- Smart Agriculture
  - Predictive analytics and decision making for sustainability, resource management, product quality



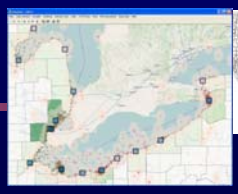




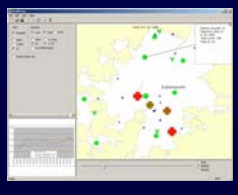


Public Safety & Health

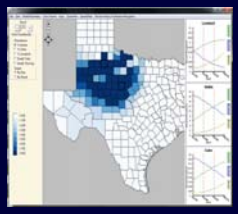
- Public safety visual analytics
  - cgSARVA
    - Coast Guard Search And Rescue Visual Analytics
  - MERGE
    - Mobile Emergency Response Guide
- Public health visual analytics
  - LAHVA
    - Linked Animal-Human Health Visual Analytics
  - RVF
    - Rift Valley Fever
    - Decision support environment for epidemic modeling and responses
  - PanVis
    - Pandemic influenza modeling and visualization tool
  - Cancer Care Engineering





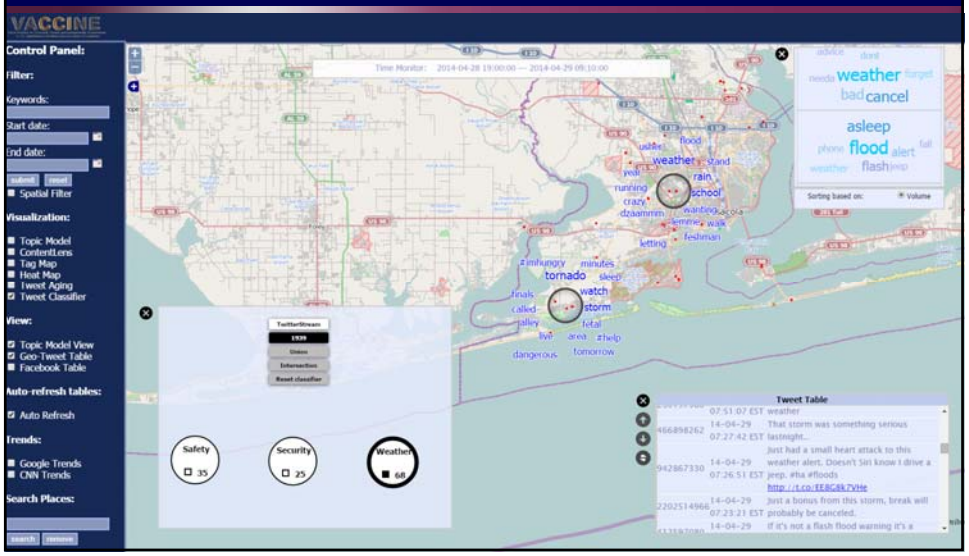








Social Media Analytics
SMART: Disaster Preparedness & Response



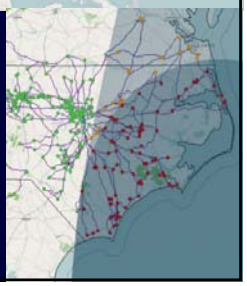
VASA: Visual Analytics for Simulation-Based Action

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

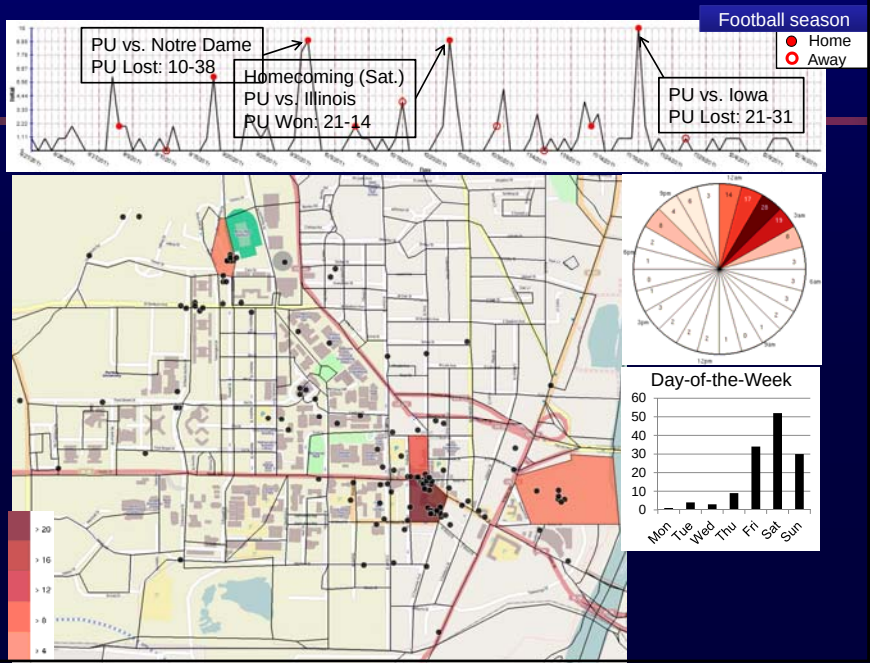
Impacts and Accomplishments:

- 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 simulation 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





High Risk Behavior : Drunkenness / Public Intoxication



Precision Agriculture

Better Decisions for Wine Grape Crop Management

Highly variable grapes grown with suboptimal decisions



Uniform quality grapes grown with VinSense based decisions

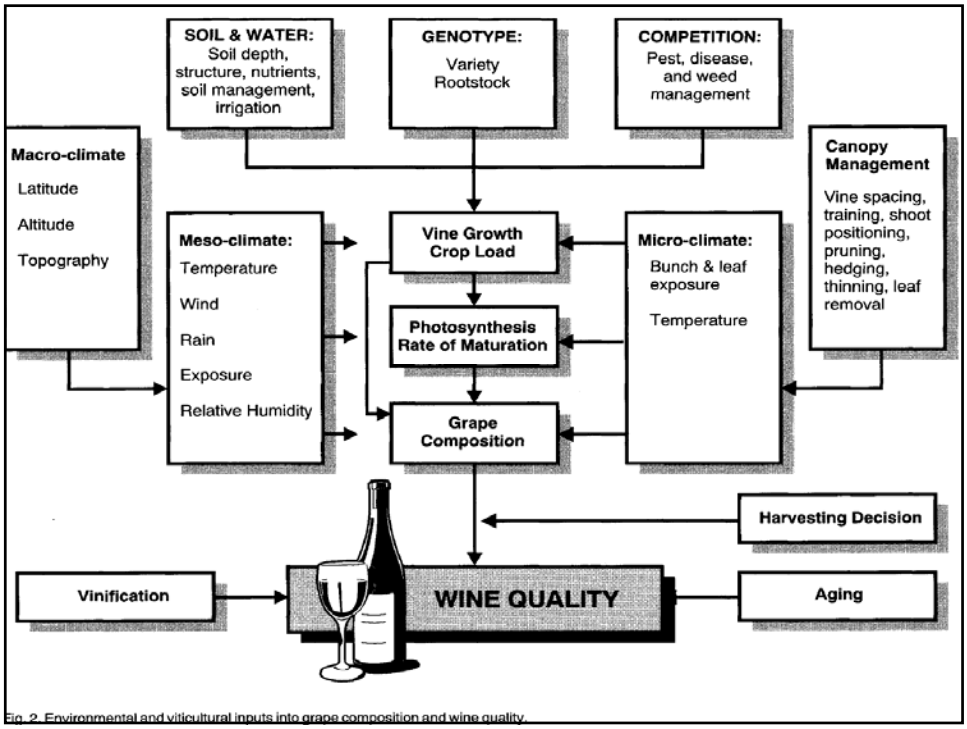
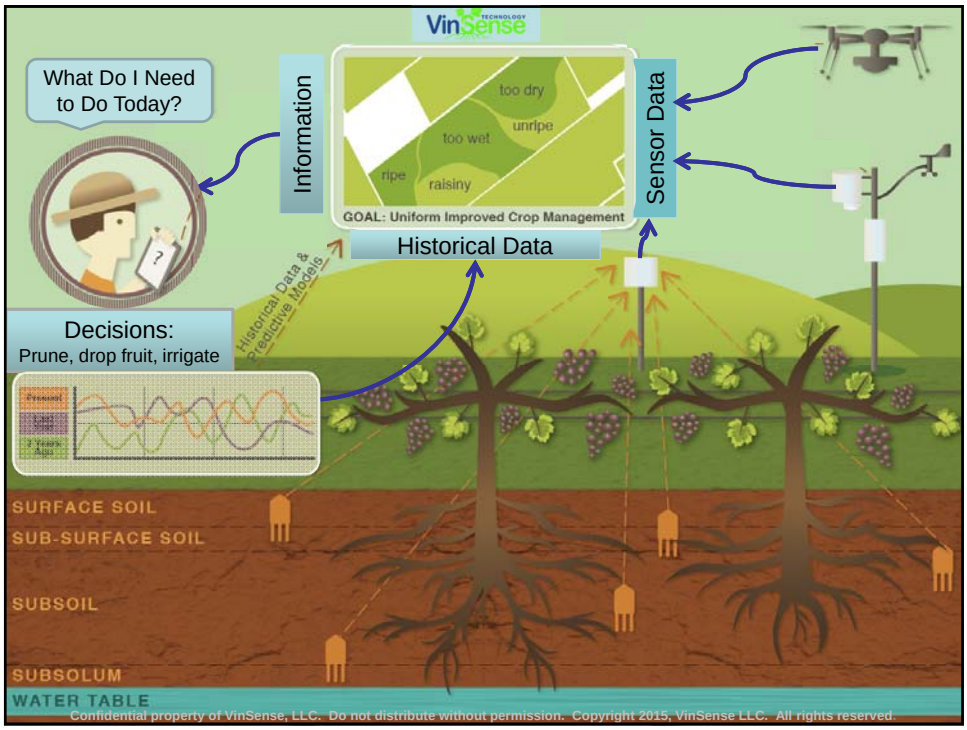


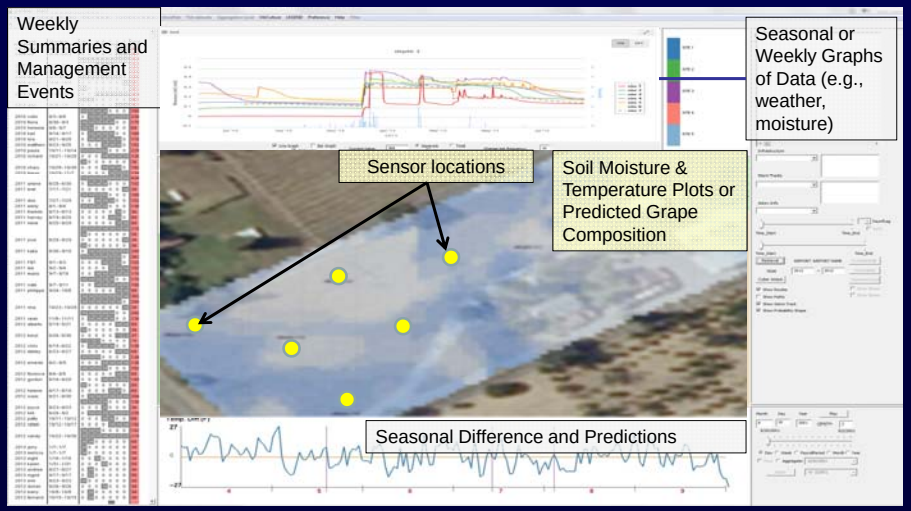
Fig. 2. Environmental and viticultural inputs into grape composition and wine quality.



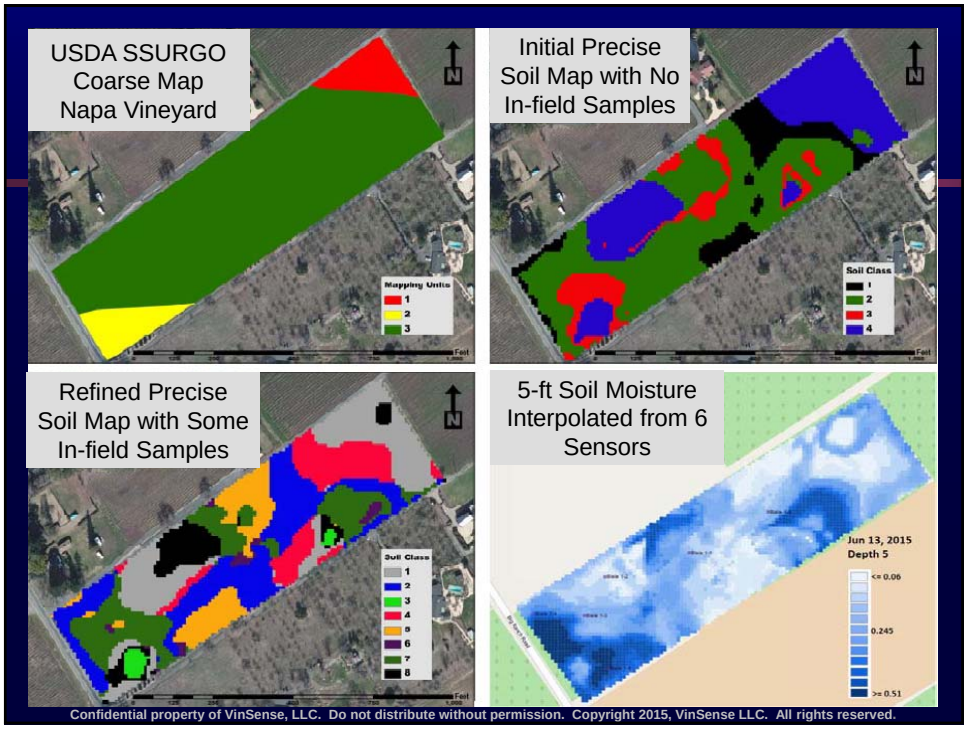
VinSense Technology Differentiators

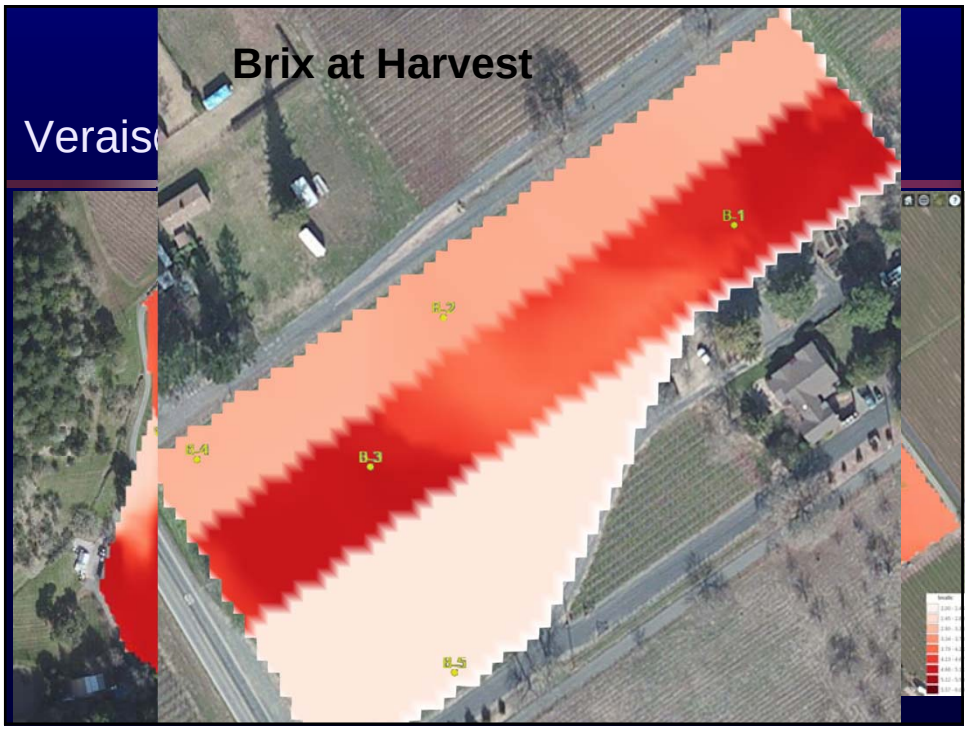
- 4D Soil Mapping and Visualization
  - Continuous time 3D information processing
  - Innovative precision soil mapping
  - Patented visualization analytics that transform high volume data into easy-to-understand decision tools
  - Integrated predictive modeling
  - Tailored to grower or vineyard specific decision points
- Intuitive user interface

Prototype Layout & Result



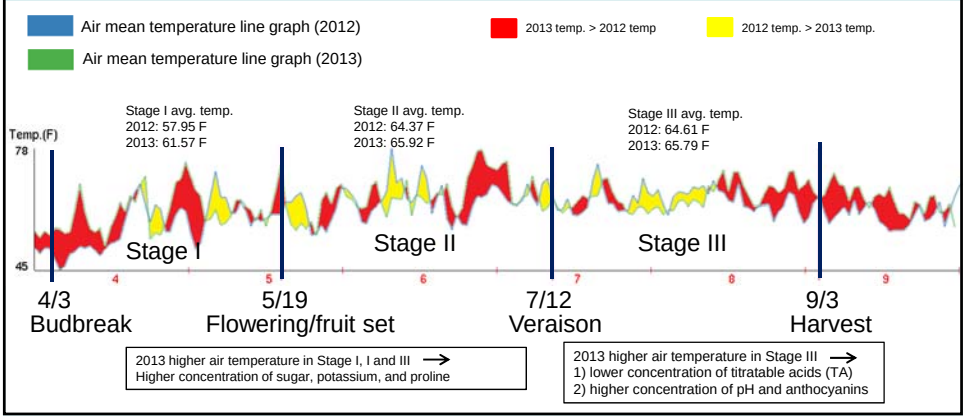
Confidential property of VinSense, LLC. Do not distribute without permission. Copyright 2015, VinSense LLC. All rights reserved.

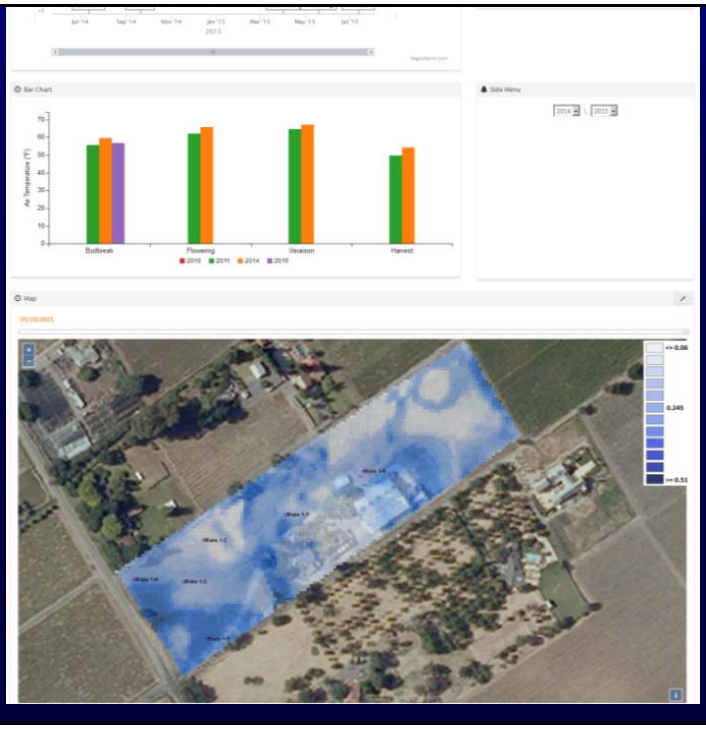

USDA SSURGO Coarse Map Napa Vineyard
Initial Precise Soil Map with No In-field Samples
Refined Precise Soil Map with Some In-field Samples
5-ft Soil Moisture Interpolated from 6 Sensors
Jun 13, 2015
Depth 5
Confidential property of VinSense, LLC. Do not distribute without permission. Copyright 2015, VinSense LLC. All rights reserved.


Brix at Harvest
B-1
B-2
B-3
B-4
B-5

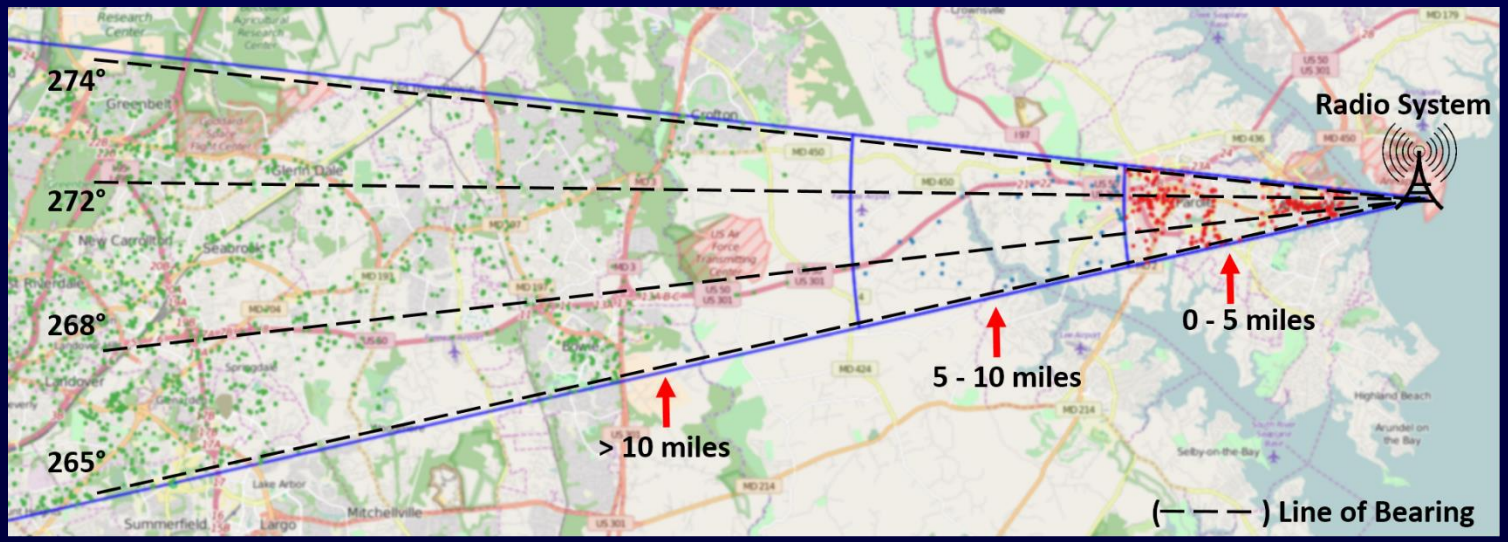
The Technology: Prediction View

- Prediction view based on recent studies showing areas indicating increase in grape component composition





Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Junghoon Chae[1], Jiawei Zhang[1], Sungahn Ko[2], Abish Malik[1], Heather Connell[1], David S. Ebert[1]

Purdue University[1], Ulsan National Institute of Science and Technology[2]

May 11, 2016

VACCINE

Introduction

- The U.S. Coast Guard (USCG) receives 139 hoax distress calls nationally per year
- Response to false calls wastes considerable resources
  - Put Coast Guard rescue personnel in dangerous situations
  - Waste valuable search and rescue assets that should be used for real emergency situations

Challenge

- Hard to identify hoax callers
- Rescue 21 (USCG radio communication system)
  - Utilized for identification of suspected hoax calls
  - Provide not enough information for finding hoax callers
- Need more intelligent analytics support

Goals

- A social media visual analytic driven by the Rescue 21 system
- Integrating the tool into SMART (Social Media Analytics Reporting Toolkit)
- Improving investigation of hoax calls by identifying suspicious social media

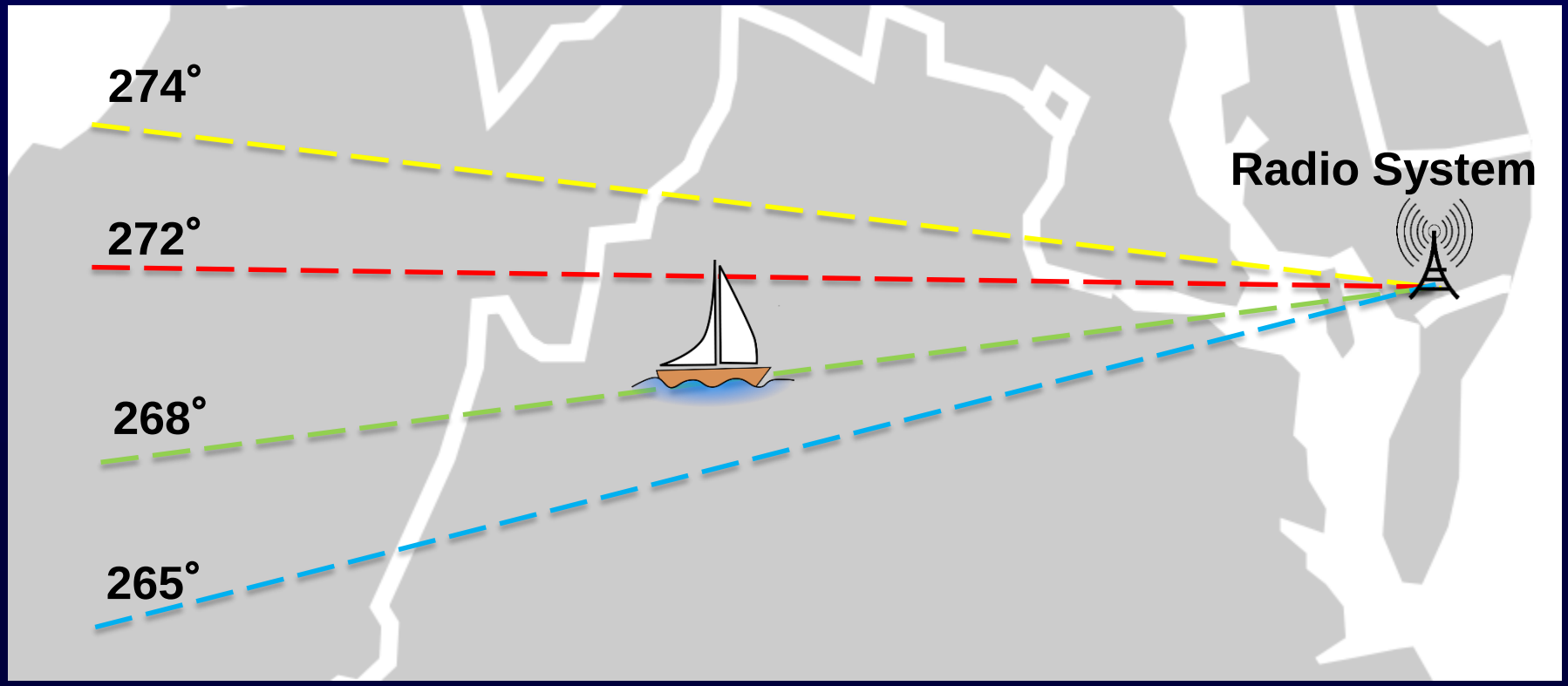
Approach

- Interactive visual analytics using social media
- Keyword collection & Data enrichment
- Identification of suspicious users & messages

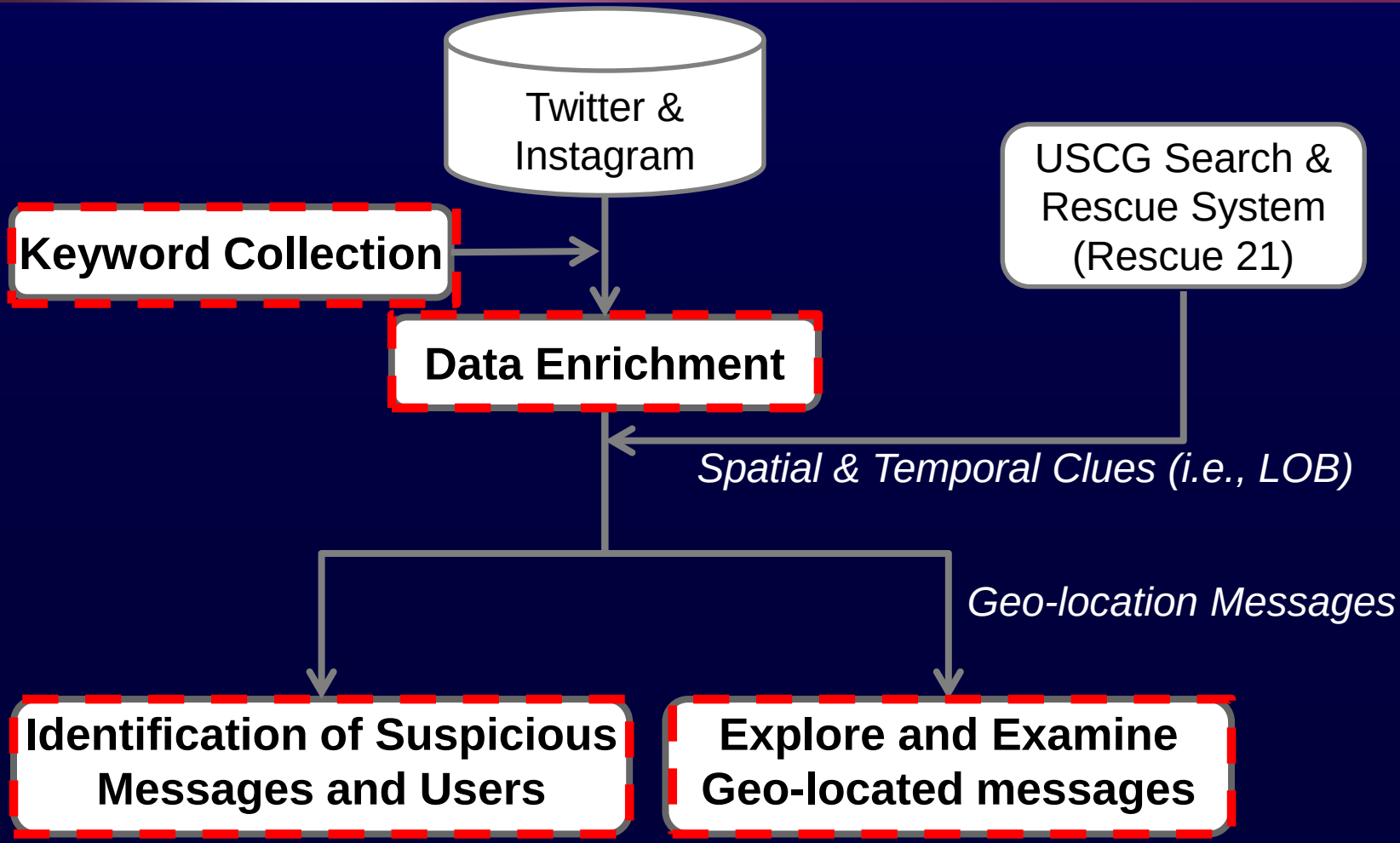
Background: Rescue 21

- USCG's radio communication system with direction-finding capability for search and rescue operations
- Providing lines of bearing (LOB) from a radio tower to the source of radio calls that it receives
- LOB is used for locating transmitters which are not able to specify their locations

Example of LOB



Analysis Process



Keyword Collection

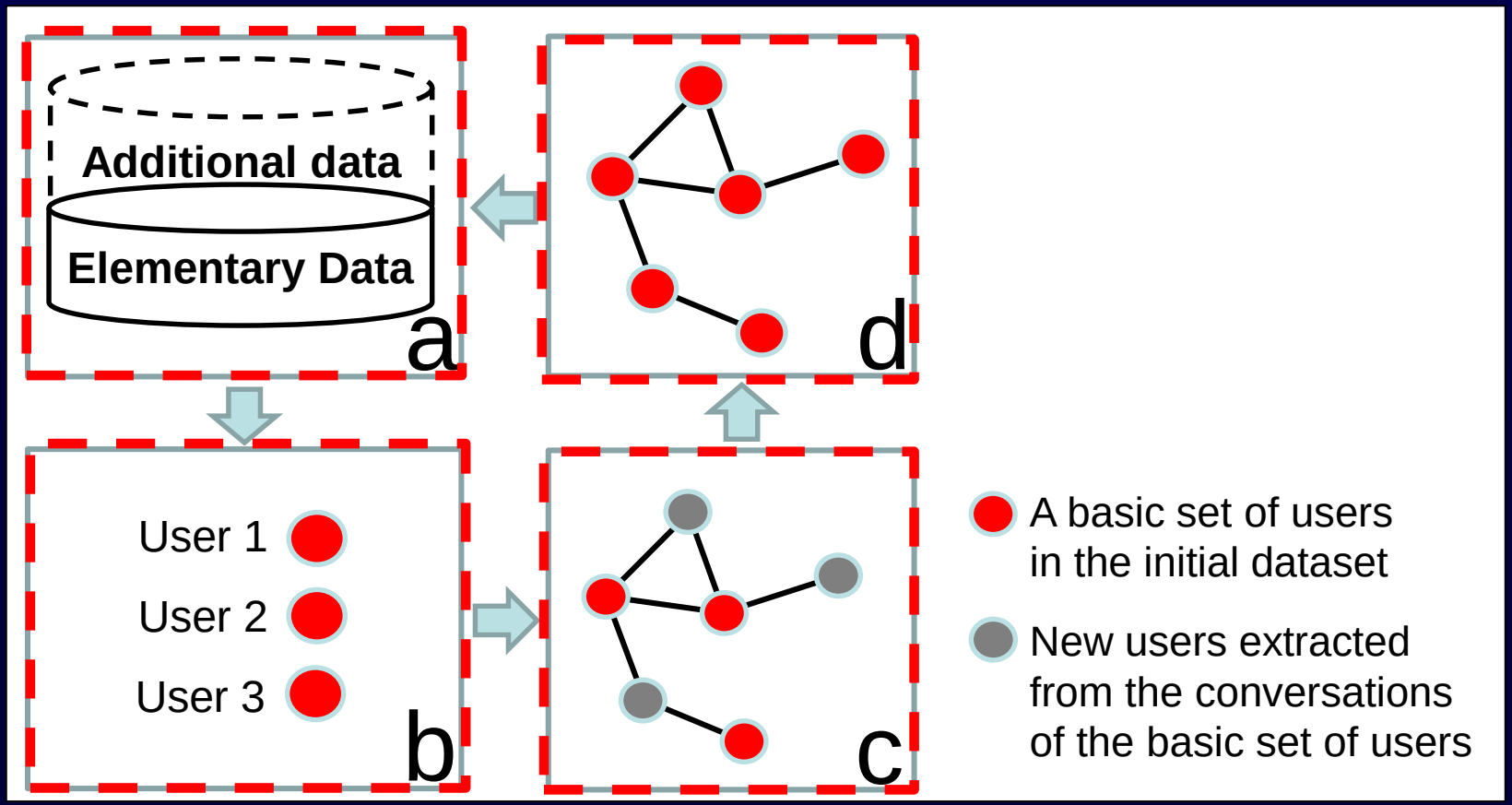
- Building a keyword corpus related to USCG and hoax distress calls
  - Retrieve highly relevant social media data
  - Reduce the analysis space and enable effective analysis tasks
- Adding words with semantic relationship to the initial keywords using WordNet
- USCG related online articles and historical social media data

Keyword Collection

Category	Initial Keywords	Extended Keywords
Coast Guard	USCG, vessel, ship, boat, ferry, shore, coastline ...	#cg, #coastguard, maritime, navy, fleet, lifeboat, craft ...
Hoax Call	**prank call, distress call, mayday, crank call, hoax ...**	**siren, save, false call, bogus, stunt, prankster ...**
Search and Rescue	helicopter, seek, search, rescue, flare, flashing ...	#sos, copter, airliner, navigate, 911, warning ...
Accident	explosion, accident, sink, crash, damaged, burning ...	disaster, hazard, collision, shot, smack, hit, break ...
Radio Channel	broadcast, radio, transmit, communicate ...	cable, on the air, airwaves, wireless, phone, report ...
Illegal Immigration	immigrant, immigration, panga, smuggling, drug ...	foreigner, exodus, migration, crossing, resettlement ...

Data Enrichment

- Data enrichment via conversation networks of users



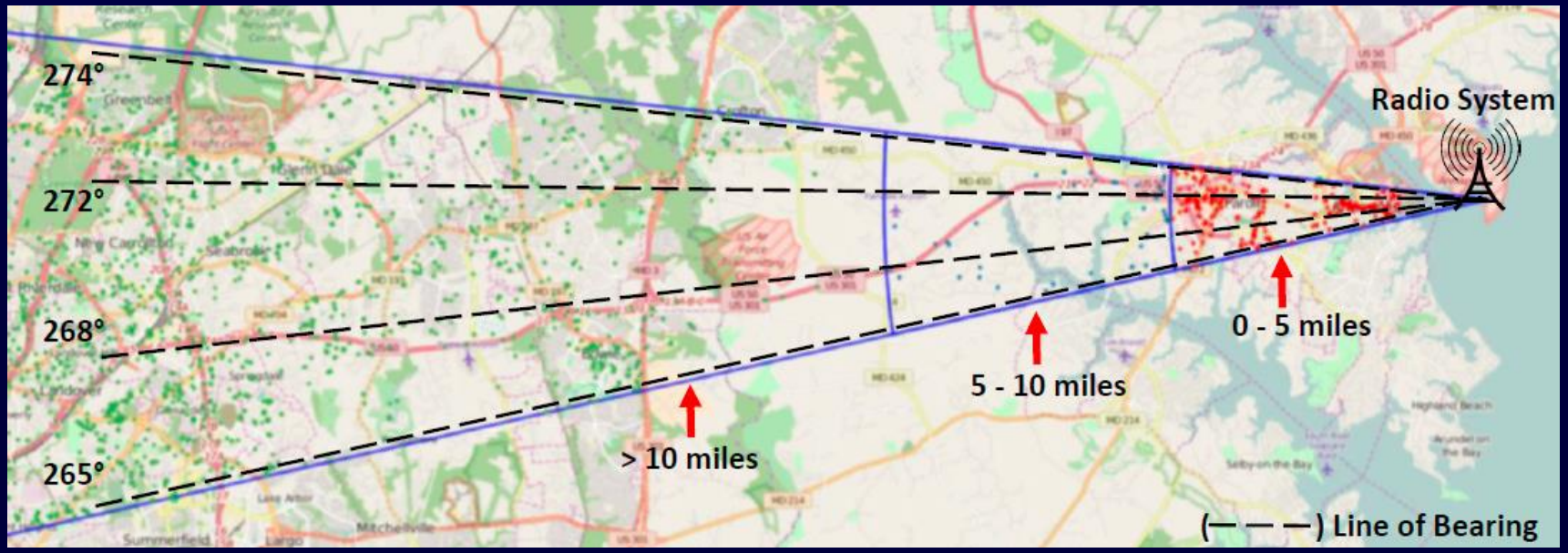
Identification of Suspicious Users

- Multiple hoax distress calls by one caller during different time frames
- Users who post messages within a time window when a hoax call was received are potentially more suspicious than others

For user i, if i posted a message
in a time window of a hoax call
Frequency score[i] += 1

Combination of Visual Analytics and Rescue 21

- Analysis of suspicious geo-tagged social media messages based on spatial and temporal clues driven by a Rescue 21 system



Case Study

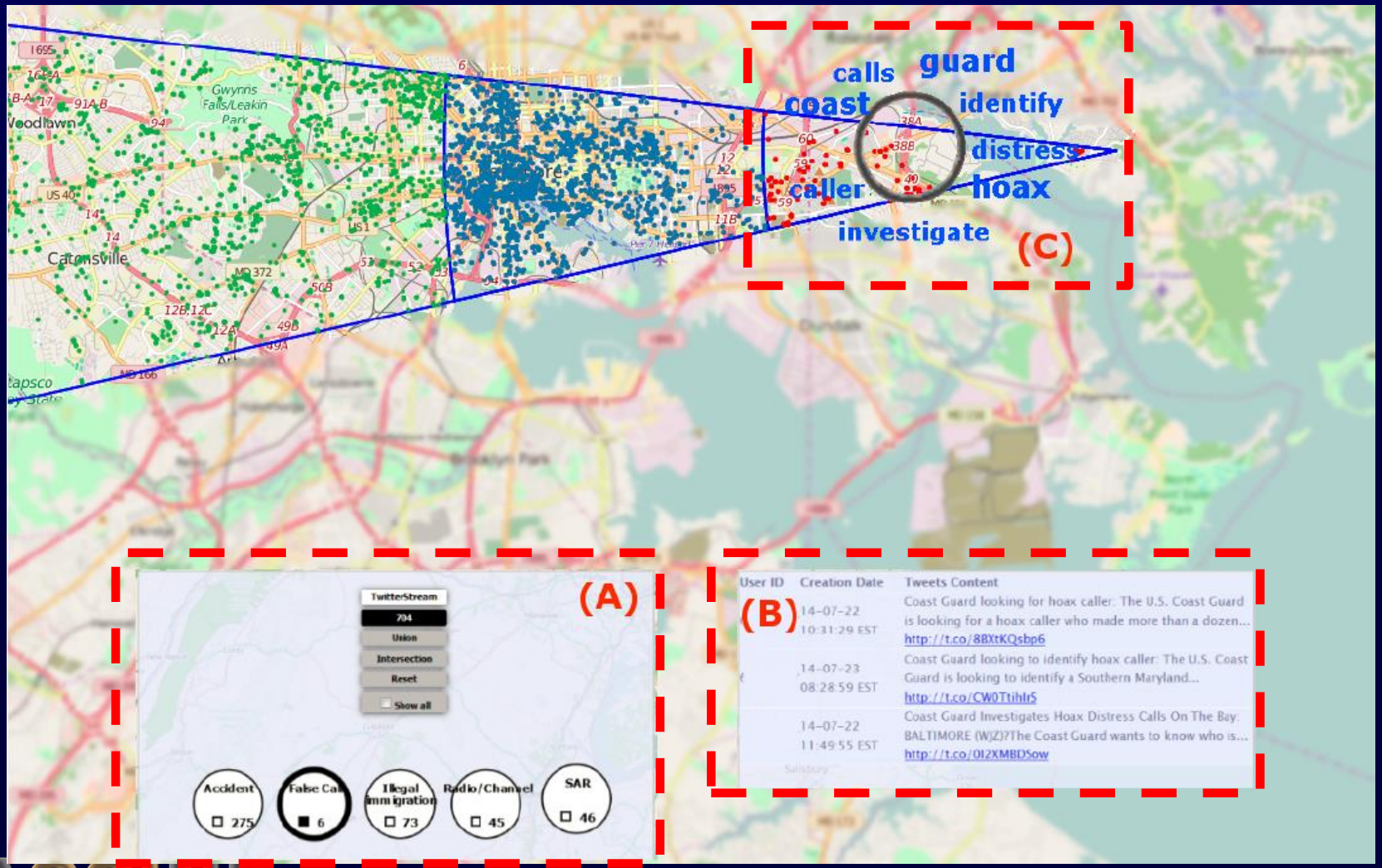
Case	Time	Number of Calls
1	January, 2014	1
2	July and August, 2014	10
3	May, 2014	3

<Information of the hoax distress call cases>

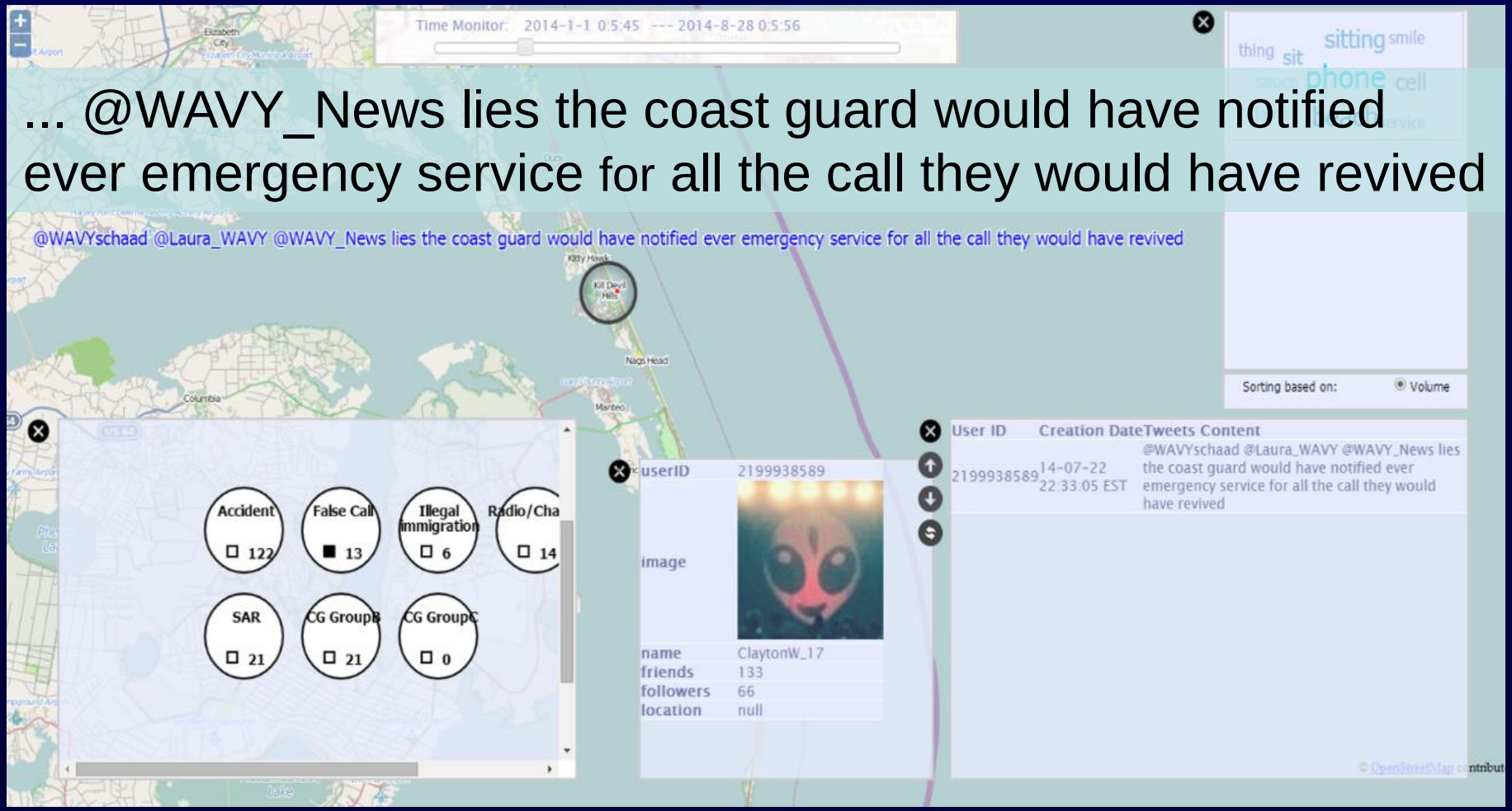
Data Type	Frequency Score						
	6	5	4	3	2	1	total
Twitter	1	4	4	15	60	269	353
Instagram			2	3	20	312	337

<Distribution of the users with respect to the frequency scores of each dataset for case 2>

Case Study


calls
guard
coast
identify
distress
caller
hoax
investigate
(C)
(A)
TwitterStream
704
Union
Intersection
Reset
Show all
Accident
275
False Call
6
Illegal Immigration
73
Radio/Channel
45
SAR
46
(B)
User ID
Creation Date
Tweets Content
14-07-22 10:31:29 EST
Coast Guard looking for hoax caller: The U.S. Coast Guard is looking for a hoax caller who made more than a dozen... http://t.co/8BXtKQsbp6
14-07-23 08:28:59 EST
Coast Guard looking to identify hoax caller: The U.S. Coast Guard is looking to identify a Southern Maryland... http://t.co/CW0TtihlrS
14-07-22 11:49:55 EST
Coast Guard Investigates Hoax Distress Calls On The Bay: BALTIMORE (WJZ)?The Coast Guard wants to know who is... http://t.co/0I2XMBDSow

Case Study


Time Monitor: 2014-1-1 0:5:45 --- 2014-8-28 0:5:56
... @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
@WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived
Sorting based on: Volume
Accident 122
False Call 13
Illegal Immigration 6
Radio/Cha 14
SAR 21
CG GroupB 21
CG GroupC 0
userID 2199938589
image
name ClaytonW_17
friends 133
followers 66
location null
User ID Creation Date Tweets Content
2199938589 14-07-22 22:33:05 EST @WAVYschaad @Laura_WAVY @WAVY_News lies the coast guard would have notified ever emergency service for all the call they would have revived

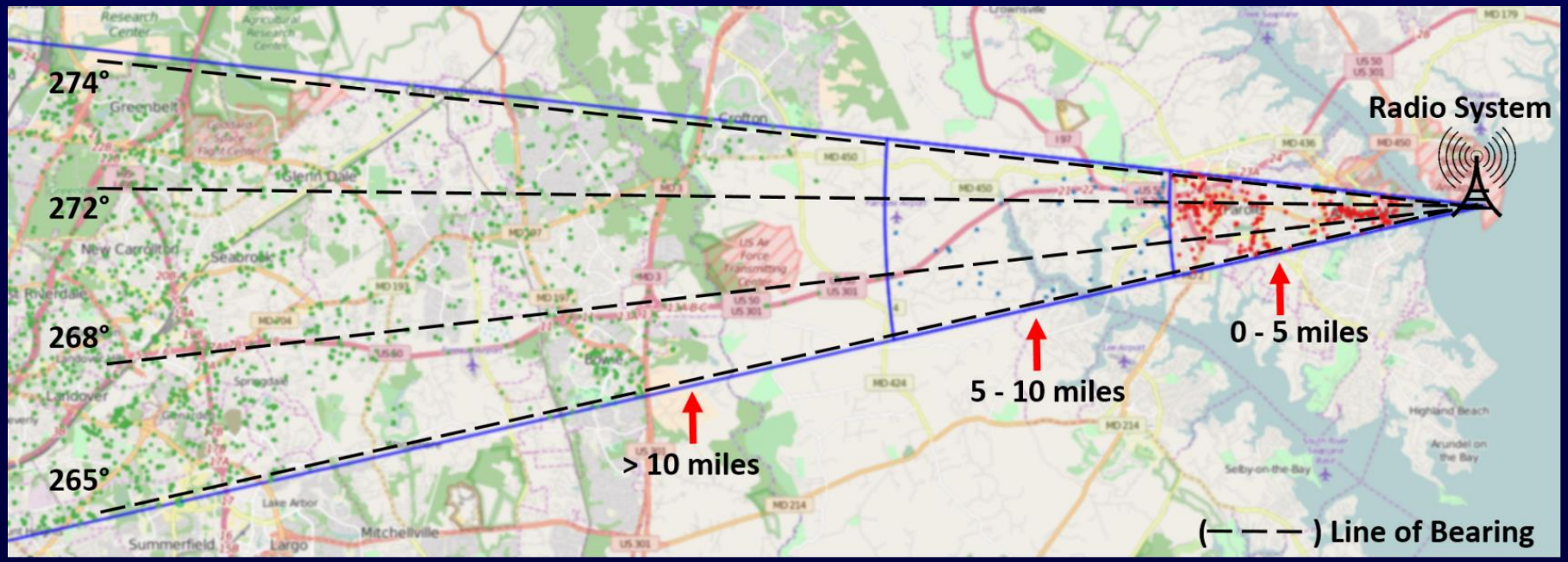
Conclusion

- A visual social media analytics solution driven by the Rescue 21 system for supporting the investigation of hoax distress calls
- Identifying suspicious Twitter and Instagram messages based on the spatial and temporal clues from the system

Acknowledgement

- U.S. Department of Homeland Security VACCINE Center under Award Number 2009-ST-061-CI0003

Visual Analytics for Investigative Analysis of Hoax Distress Calls using Social Media



Thank you!

Junghoon Chae
jchae@purdue.edu

VACCINE