VACCINE ANNUAL REPORT – YEAR 4
Addendum A - Presentations
April 1, 2012 – June 30, 2013
Cooperative Agreement No. 2009-ST-061-CI0001
Addendum A – Presentations

Interactive Table of Contents

ARIZONA STATE UNIVERSITY
Analytical Brushing in Spatiotemporal Analysis .................................................................................................................. 4
Applied Visual Analytics for Exploring the National Health and Nutrition Examination Survey ........................................... 45

GEORGIA INSTITUTE OF TECHNOLOGY
Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts ........................................................... 94
The Value of Visualization for Exploring and Understanding Data ..................................................................... 121
The Value of Visualization for Exploring and Understanding Data ..................................................................... 148

MORGAN STATE UNIVERSITY
iLEAPS Radio Interview Presentation ........................................................................................................... 173

PENNSYLVANIA STATE UNIVERSITY
Defining the Typical DHS Map .......................................................................................................................... 174
Design & Use Guidelines For Interactive Maps: A Case Study ........................................................................ 189
Geovisualization to Geovisual Analytics: Visual Reasoning with Big & Messy Data ........................................... 226
Mobile Symbol Design and Event Reporting for Mobile Devices ........................................................................ 235
Sharing Map Symbology for Emergency Management ..................................................................................... 241
SymbolStore.org – An Open Resource for Map Symbols .................................................................................. 269
The New Cartography – Current states of science and technology in Interactive Cartography & Geovisualization ..................................................................................................................... 297
The SymbolStore: Expanding to Provide a Social Forum for the Creation, Sharing, and Evaluation of Symbols ..................................................................................................................... 346
Understanding Spatial and Social Relationships in International Trade Network: A Geovisual Analytic Approach ..................................................................................................................... 359
Visualizing Spatial, Temporal, and Social Graph Information with the GeoViz Toolkit ........................................ 376
Visualizing Uncertainty and Decision-Making ................................................................................................. 385

PURDUE UNIVERSITY
CVADA – Predictive Visual Analytics ..................................................................................................................... 440
Fundamentals of Visual Analytics ..................................................................................................................... 470
How Can Visual Analytics Help Cartography? ................................................................................................. 612
Innovation with Impact ................................................................................................................................. 675
Overview of VACCINE Projects ..................................................................................................................... 712
Spatiotemporal Visualization and Analysis of Gang-related Criminal Activity Using Mobile Imaging........753
Teaching Visual Analytics: Leveraging Multidisciplinarity...........................................................................789
Transitioning GARI to Law Enforcement and ICE.....................................................................................804
VACCINE A DHS Center of Excellence USCG Innovation Showcase.........................................................811
Visual Analytic Applications for Law Enforcement..................................................................................817
Visual Analytics: A Lifeboat in The Big Data Ocean..................................................................................841
Visual Analytics: From Situational Awareness to Risk-Based Decision Making.......................................856
Visual Analytics: Powering Discovery, Innovation, and Decision Making.................................................884
Visual Analytics & Big Data Analytics.......................................................................................................943
Visual Analytics A Lifeboat in the Big Data Ocean....................................................................................977
Visual Analytics and Decision Making......................................................................................................996
Visual Analytics for Decision Making........................................................................................................1004
Visual Analytics for Effective Planning, Analysis, and Decision Making.....................................................1054
Visual Analytics for Financial Analytics, Fraud, and Decision Making.......................................................1104
Visual Analytics for Risk-Based Decision Making: Anytime and Anywhere............................................1121
Visual Analytics in Public Health................................................................................................................1149
Visual Analytics for Decision Making – ICE............................................................................................1178

SIMON FRASER UNIVERSITY, CA

Opportunities in Data Visualization and Visual Analytics for Behavioral and Social Science Research......1228
Analytical Brushing in Spatiotemporal Analysis

Ross Maciejewski

E-mail: rmacieje@asu.edu
http://rmaciejewski.faculty.asu.edu/
Much of the data being collected and analyzed today is spatiotemporal

- Twitter¹,²
- Emergency Departments³
- Law Enforcement⁴
- Mobile Phones⁵,⁶
- Animal Migration⁷
- Flow Simulations⁸


SensePlace2: Place-Time-Concept Analytics

- Grid map of tweet frequency matching query; graduated circles depict 500 most relevant tweets; support for spatial filter by distance from point and time filter.
- Ranked, sortable 500 most relevant tweets.
Distribution of Surname Cutchin
Coordinated Multiple Views in Spatiotemporal Analysis

To facilitate answering these questions, we typically rely on interactive coordinated multiple views

- Geographic Maps
- Scatterplots
- Parallel Coordinate Plots

- Dendogram
- Histogram
- Time Series Plots

- Tabular View
- Box-and-Whisker

Facilitating Spatial Analysis: Focus+Context Brushing

Visual Analytics

Visualization is good for exploring data, but we can do more than just explore

“Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces”¹

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.²

“A graphic display has many purposes but it achieves its highest value when it forces us to see what we were not expecting.”³

2 – Keim et al. chapter in Information Visualization: Human-Centered Issues and Perspectives, 2008.
Spatiotemporal Analysis

*Key spatial questions tend to be about causal relationships*

- Do neighboring spatial regions have similar values?
- Do neighboring spatial regions have similar temporal trajectories?
- Are disparate spatial regions showing similar trajectories or values?
- Are temporal trends lagging or leading in neighboring spatial regions?

*Key Temporal questions tend to focus on change*¹

- What is the magnitude of the change?
- What is the shape of the changes?
- How fast is the change occurring (what is the velocity of the change)?
- What is the direction of the change?

Challenges in Spatiotemporal Analysis

Exploring trends and patterns in spatiotemporal data

• Correlations among multi-variate data
• Explore possible predictive links among different variables

Challenges:

• End users understanding underlying statistical algorithms
• Applying algorithms at appropriate spatial and temporal scales
• Dealing with noisiness of the real-world data
Spatial Analysis: Linked Multivariate Filtering

Link complex data queries directly to an interactive logic tree to facilitate multivariate queries.

Problem is this is time consuming and tree gets complex quickly as the number of variables and constraints increase.
Spatial Analysis: Multivariate Clustering

K-means clustering

- Goal is to partition n-observations into k-clusters where each observation belongs to the cluster with the nearest means.
- This provides information on how the data groups together.
- Drawbacks are the use of Euclidian distance and the fact that k is user-defined.
Spatial Analysis: Multivariate Clustering

Hierarchical Clustering
- Produces a set of nested clusters organized as a hierarchical tree
- Can be visualized as a dendrogram

No assumptions on the number of clusters
- Any desired number of clusters can be obtained by ‘cutting’ the dendrogram at the proper level
Spatial Analysis: Multivariate Distance Function

In these high-dimensional clusters, we don’t necessarily have a concept of how far away two points are.

We want to know how spatial units are related in high-dimensional space and observe these patterns in the geographic projection.

We can directly link a distance function to the brush tip.

<table>
<thead>
<tr>
<th>Metric Names</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>( dist = \sqrt{\sum_i (v_i^1 - v_i^2)^2} )</td>
</tr>
<tr>
<td>Manhattan</td>
<td>( dist = \sum_i</td>
</tr>
<tr>
<td>Canberra</td>
<td>( dist = \sum_i \frac{</td>
</tr>
<tr>
<td>Binary</td>
<td>( dist = \sum_i v_i^1 = v_i^2 )</td>
</tr>
</tbody>
</table>
Indices of Industrial Diversity

Indices of industrial diversity measure the relative concentration of a particular economic variable across a specified number of industries.

Industrial diversity is measured by the number of firms as reported by the North American Industrial Classification System (NAICS).

<table>
<thead>
<tr>
<th>N11</th>
<th>Agriculture</th>
<th>N51</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>N21</td>
<td>Mining</td>
<td>N52</td>
<td>Finance and Insurance</td>
</tr>
<tr>
<td>N22</td>
<td>Utilities</td>
<td>N53</td>
<td>Real Estate</td>
</tr>
<tr>
<td>N23</td>
<td>Construction</td>
<td>N54</td>
<td>Scientific and Technical Service</td>
</tr>
<tr>
<td>N31</td>
<td>Manufacturing</td>
<td>N55</td>
<td>Management of Companies</td>
</tr>
<tr>
<td>N42</td>
<td>Wholesale Trade</td>
<td>N56</td>
<td>Administrative Waste Services</td>
</tr>
<tr>
<td>N44</td>
<td>Retail Trade</td>
<td>N61</td>
<td>Educational Services</td>
</tr>
<tr>
<td>N48</td>
<td>Transportation and Warehousing</td>
<td>N71</td>
<td>Arts and Recreation</td>
</tr>
<tr>
<td>N62</td>
<td>Health Care</td>
<td>N72</td>
<td>Accommodation and Food Service</td>
</tr>
</tbody>
</table>
Distance in terms of Educational and Scientific Services (N54 and N61)
Distance in terms of Educational and Scientific Services (N54 and N61)

Tippecanoe County
(Purdue University)
K-means clustering (k=5) in terms of Food and Entertainment Services (N71 and N72)
Spatial Analysis: Interactive Clustering

Adapt k-means so that each spatial unit has a unique weight (initially set so all weights are equal).

Analysts may now modify weights based on original k-means output, this is done by assigning spatial units to different clusters.

Upon cluster reassignment, the weight of that spatial unit is updated to be \( w_i = 1 + \sqrt{s_j} \), where \( s_j \) is the size of cluster \( j \).

Now the cluster centroids are recomputed and k-means rerun such that cluster centroids are
Spatial Analysis: Interactive Clustering
Spatiotemporal Analysis

**Key spatial questions tend to be about causal relationships**
- Do neighboring spatial regions have similar values?
- Do neighboring spatial regions have similar temporal trajectories?
- Are disparate spatial regions showing similar trajectories or values?
- Are temporal trends lagging or leading in neighboring spatial regions?

**Key Temporal questions tend to focus on change**
- What is the magnitude of the change?
- What is the shape of the changes?
- How fast is the change occurring (what is the velocity of the change)?
- What is the direction of the change?

---

Facilitating Spatiotemporal Analysis: Linked Time Series

Facilitating Spatiotemporal Analysis: Animation and Small Multiples

As data grows, accuracy becomes an issue
• Traces and animation get cluttered
• Small multiples get tiny

How do we classify/segment/bin the data?
• Classification over each individual time step?
• Classification over entire data range?


Facilitating Spatiotemporal Analysis: Identifying Temporal Thresholds

Histogram and map showing distribution of spatiotemporal variables.
Facilitating Spatiotemporal Analysis: Temporal Similarity Brushing

Many of these questions depend on temporal trajectories.

We can apply time series similarity work as a means of determining which regions have similar trajectories.

One common metric is the sequence normalized metric which will focus on matching trends as opposed to exact values.

\[
\text{dist}_\text{SNE} (p, q) = \sqrt{\sum_{i=1}^{n} \left( \frac{q_i - \mu_q}{\sigma_q} - \frac{p_i - \mu_p}{\sigma_p} \right)^2}.
\]
Regions with trajectories in Mining (N21) similar to Reagan Co., TX
(75% Similarity Score)
Regions with trajectories in Mining (N21) similar to Reagan Co., TX (75% Similarity Score)
What Does 75% Similar Look Like?
Applied Analytical Brushing: A Case Study in Industrial Diversity

Performing an AND operation with the last three maps
What is the best way to cluster spatiotemporal data?

- **Allow time to be a vector in the cluster algorithm**
  - Now spatial units can only belong to one cluster

- **Cluster data per time step and then link clusters across time step by space**
  - Now spatial units can belong to multiple clusters, but need to maintain coherence between cluster labels at each time step
Assume that we have some spatial arrangement of data with multivariate measures per time step.

For each time step, we perform k-means clustering.

Clusters at $t_1$:
- Cluster 1: A, B, C
- Cluster 2: D, E
- Cluster 3: F

Clusters at $t_2$:
- Cluster 1: A, B
- Cluster 2: C, D, E
- Cluster 3: F

Label Clusters, but use information from previous time step, minimize the number of elements for which the cluster label will change (Kuhn-Munkres (KM) Algorithm$^1$).

So all spatial units maintain the same label from $t_1$ to $t_2$ except for element C.

Temporally Coherent Spatial Clusters

What if we ignore the space and instead visualize how the element cluster labels change?

- What time steps should we present?
- Can we globally minimize the label changes?
- Measures of stability based on label changes?
Challenges

**Scalability**

- *Unfortunately, as dimensionality increases, performance of similarity metrics decreases*¹

  - Is the use of fractional distance metrics intuitive for users?
  - Should dimensional reduction/clustering be done spatially, temporally or spatiotemporally?

*How do we move from hypothesis generation to hypothesis testing?*

  - *Traditional techniques such as multiple regression do not scale well to high-dimensions*

---

Questions?
Facilitating Spatial Analysis: Multivariate Representations

Visualization is concerned primarily with a mapping to visual form

\([x,y]\)
- Position

\([z]\)
- Size (Taille)
- Value (Valeur)
- Color (Couleur)
- Texture (Grain)
- Orientation
- Shape (Forme)

J. Bertin (1967), *The Semiology of Graphics*
Bristle Maps: A Multivariate Abstraction Technique for Geovisualization

Typographic Maps: An Illustrative Abstraction Technique
Applied Visual Analytics for Exploring the National Health and Nutrition Examination Survey
The Ongoing Data Deluge

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

---
Research Challenges in the Data Deluge

Long-term fundamental challenges

- Mobility
- Scale
- Adaptation
- Uncertainty
- Multisource, cross-media analysis
- Numerical to visual analysis

“All this data must be fused with relevant contextual or situational information and visualized in such a way to give our forces a clear picture of threats, options and potential consequences.

The science and technology community has been successful in giving our forces a data advantage; now it’s time to give them a decision advantage.”

-Zachary J. Lemnios Director, Defense Research and Engineering
Department of Defense February 17, 2010
Exploiting the Data Deluge

• Credit card companies monitor purchases for fraudulent activity
• Mobile operators analyze call patterns to see if people you frequently call use a rival network
• The oil industry explores seismic data before drilling
• The US government collects tons of population based data that we can exploit for analysis!
National Health and Nutrition Examination Survey (NHANES)

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

- Combines interviews and physical examinations
- Exams a nationally representative sample of about 5,000 persons per year
- Includes demographic, socioeconomic, dietary and health-related questions
Healthy Eating Index (HEI)

Measure of diet quality that assesses conformance to federal dietary guidance

USDA uses HEI to monitor diet quality of the US population

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


### Healthy Eating Index—2005 components and standards for scoring\(^1\)

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

---

\(^1\)Intakes between the minimum and maximum levels are scored proportionately, except for Saturated Fat and Sodium (see note 5).

\(^2\)Legumes counted as vegetables only after Meat and Beans standard is met.

\(^3\)Includes all milk products, such as fluid milk, yogurt, and cheese, and soy beverages.

\(^4\)Includes nonhydrogenated vegetable oils and oils in fish, nuts, and seeds.

\(^5\)Saturated Fat and Sodium get a score of 8 for the intake levels that reflect the 2005 Dietary Guidelines, \(<10\)% of calories from saturated fat and 1.1 grams of sodium/1,000 kcal, respectively.
Playing Data Detective

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

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

What are some easy questions to ask?

• What is the average … (total fruit consumption)?
• What is the average … (milk consumption)?
NHANES

Utilizing Survey Cycle 2-3

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

Other variables:

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

So with more data, maybe we can think of more questions!
What is the average ... ?
What is the average ... ? Meat Consumption
A Century of Meat

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

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

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

Sources: U.S. Department of Agriculture (data); news and company reports; “Putting Meat on the American Table,” by Roger Horowitz
What is the average ... ? Total Fruit Consumption
What is the average ... ? Dark Green and Orange Vegetable Consumption
How Do These Compare?

- Total Fruit
- Whole Fruit
- Total Vegetables
- Dark Greens
- Total Grains
- Whole Grains
- Milk
- Meat and Beans
- Oils
- Saturated Fat
- Sodium
- SOFAS
So, What Can I say About the Population as a Whole?

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

- But what does this say about someone like you?
What if I further segment the data?
Filter by Age, Gender, Etc.?

Let’s look at Females, Age 19-30

18.5 < BMI < 25.0

Total population in our sample = 511
NHANES in Females Age 19-30, Average BMI
Back to the Entire Population!

[Graphs showing data distributions for Total Fruit, Whole Fruit, Total Vegetables, Dark Greens, Total Grains, Whole Grains, Milk, Meat and Beans, Oils, Saturated Fat, Sodium, SOFAS]
Forming Hypotheses

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

- Their SOFA consumption is more evenly distributed
- Their meat and bean consumption is more skewed
What if I further segment the data? Filter by Age, Gender, Etc.?

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

Total population in our sample = 145
NHANES in Adolescent Underweight Males
Back to the Entire Population!

Table of Contents

- Total Fruit
- Whole Fruit
- Total Vegetables
- Dark Greens
- Total Grains
- Whole Grains
- Milk
- Meat and Beans
- Oils
- Saturated Fat
- Sodium
- SOFAS
Forming Hypotheses

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

- Their vegetable consumption is more evenly distributed
- Their meat and bean consumption is more skewed
- Their SOFAs are more evenly distributed
What if I want to ask:
How is ... related to ... ?

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

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

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

We can try a scatterplot!
Group: Underweight Adolescent Males

I don’t see any obvious patterns in this 😞
Now what do I do?
What if I further question the data? All relationships by Age, Gender?

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

These histograms were made in Excel

One can also use SAS, JMP, MatLab

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

We can’t directly manipulate our variables

So, being a data detective can be very tedious

And these were just simple comparisons!
So what do we do now?

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

A term coined by Bellman in 1961

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

Sparsity is a problem for any method that requires statistical significance

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

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

The Curse of Dimensionality

So, what can we do?

- We can incorporate prior knowledge of the data
- We can smooth the target function
- We can reduce the dimensionality
The Practicality of the Curse of Dimensionality

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

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

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

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

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

Overall, NHANES is actually a relatively modest set of data.

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

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

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

Another is to try dimensional reduction!
Dimensional Reduction

Two approaches are available to reduce dimensionality

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

**Feature Selection:** choosing a subset of all the features

- Well, that’s what we were already doing
- We use our own expert knowledge to select features of interest
- Note we could also use machine learning and statistics here too if we were so inclined
What Does Dimensional Reduction Mean For You?

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

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

One of the most commonly applied dimension reduction techniques

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

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

Principle Component Analysis

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

- PCA simply performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance
- There is no guarantee that the directions of maximum variance will contain the most interesting/important features
That sounded really hard!

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

- But, how does principal component analysis help me?
- What can principal component analysis tell me?
Principal Component Analysis

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

• Ok, you’ve convinced me to run PCA on NHANES and I get 8 components, but now I’m confused!

• The output is such that your first component accounts for as much of the variability in the data set as possible

• The second component accounts for the second most variability, etc.

• Typically we will plot the data using the first two principal components as axes

• We can also write out each component to analyze it
PCA on NHANES

Underweight Adolescent Males

\[ FV1 \approx 0.14 \text{HEI}1 + 0.07 \text{HEI}2 + 0.02 \text{HEI}5 + 0.02 \text{HEI}6 - 0.09 \text{HEI}7 + 0.19 \text{HEI}8 + 0.05 \text{HEI}3 + 0.03 \text{HEI}4 + 0.20 \text{HEI}9 + 0.20 \text{HEI}10 + -0.15 \text{HEI}11 + 0.91 \text{HEI}12 \]

\[ FV2 \approx 0.13 \text{HEI}1 + 0.06 \text{HEI}2 + 0.05 \text{HEI}5 + 0.01 \text{HEI}6 + 0.70 \text{HEI}7 + -0.27 \text{HEI}8 + -0.05 \text{HEI}3 + 0.02 \text{HEI}4 + -0.48 \text{HEI}9 + -0.30 \text{HEI}10 + -0.18 \text{HEI}11 + 0.25 \text{HEI}12 \]
Females Age 19-30, Average BMI

\[ FV1 \approx 0.13 \text{HEI}_1 + 0.11 \text{HEI}_2 + 0.04 \text{HEI}_5 + 0.03 \text{HEI}_6 + -0.00 \text{HEI}_7 + 0.11 \text{HEI}_8 + 0.08 \text{HEI}_3 + 0.07 \text{HEI}_4 + 0.17 \text{HEI}_9 + 0.10 \text{HEI}_{10} + -0.12 \text{HEI}_{11} + 0.94 \text{HEI}_{12} \]

\[ FV2 \approx -0.01 \text{HEI}_1 + 0.01 \text{HEI}_2 + 0.02 \text{HEI}_5 + 0.03 \text{HEI}_6 + 0.72 \text{HEI}_7 + -0.07 \text{HEI}_8 + 0.00 \text{HEI}_3 + 0.00 \text{HEI}_4 + -0.18 \text{HEI}_9 + -0.62 \text{HEI}_{10} + -0.22 \text{HEI}_{11} + 0.08 \text{HEI}_{12} \]
Putting it all together

• *If I know things about you* …
  • Age, gender, BMI

• *Can I guess your typical daily diet?*
  • Probably not very accurately

• *But, if I ask you a few questions, maybe I can get closer!*
Putting it all together

This is the NHANES data. I know everyone’s diet.

I can first split the data by demographics.

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

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

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

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

Maybe if you ate a salad today tells me more!

Now my demographics groups having groups based on diet!

Now I know your diet! (Well sort of)
Summary

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

• Right now, I have a Facebook app that can guess your income based on your location and your friends (still beta testing so it is not publicly available 😅)

• Soon, we’ll have a Facebook app that will start narrowing down your diet
Dr. Ross’s Shameless Plug

- *You can play with my other app though*

- Search for hedcutr on Facebook and you can make this:

- I’m collecting data on the rendering parameters to see what people like!
Questions?
(http://rmaciejewski.faculty.asu.edu)
Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts

Youn-ah Kang and John Stasko

Georgia Institute of Technology
We evaluate systems because..
- We hope our technologies are making an impact
- We hope they are helping people gain value from their information

Evaluation is challenging and not very common
- Usability testing and controlled experiments remain crucial
- Actual case studies of prolonged system use by analysts working with their own data are still rare
Case Studies

- Case studies can provide valuable findings and insights
  - Yield a description of how a tool was used and where users had problems
  - Difficult to achieve through controlled lab studies

- Multi-dimensional in-depth long-term case studies (MILC)
System of Study

- Jigsaw (http://www.cc.gatech.edu/gvu/ii/jigsaw/)
Motivation and Goals

• Is Jigsaw helping analysts with their tasks and problems?

• For what types of documents and analyses does Jigsaw help?

• What are particularly useful features/capabilities as well as missing or problematic ones?
Recruitment and Study Protocol

- 6 working/practicing investigators who were using the system
  - 3 intelligence analysts, 2 academic researchers, and 1 business analyst
- Used Jigsaw for a range of 2-14 months
- Semi-structured interviews
- Follow-up email conversations
Case Studies
Task: Compare two major air traffic control programs and examine their compatibility

**Description:** Departures are sequenced and staged to maintain throughput. Air Navigation Service Provider (ANSP) automation uses departure-scheduling tools to flow surface traffic at high-density airports. Automation provides surface sequencing and staging lists for departures and average departure delay (current and predicted). ANSP automated decision support tools integrate surveillance data. This includes weather data, departure queues, aircraft flight plan information, runway configuration, expected departure times, and gate assignments. Automation provides surface sequencing and staging lists for departures and average departure delay (current and predicted). Local collaboration between ANSP and airport stakeholders improves information flow to decision support as well as the ability for aircraft operators to meet their operational and business objectives.

**Functional Drivers:** The use of improved departure scheduling and surface management will reduce delays and environmental impacts resulting in more efficient operations.

**SOPR:** FAA

**SOPR Unique Reference:** 104209

**SOCR:**

**Primary Supported OIs:** OI-0322, OI-0321

**OI Group:** Trajectory Management - Surface Operational Improvements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OI</td>
<td>O</td>
<td>E 2010</td>
<td>E 2011</td>
<td>P 2008</td>
</tr>
<tr>
<td>OI</td>
<td>O 2012</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**P1: Aerospace Engineering Researcher**
Goal: Identify similarities/differences and create a mapping between the two programs

- Does a concept or capability suggested in one program also appear in the other program?

<table>
<thead>
<tr>
<th>Program A</th>
<th>Program B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OI-0320: Initial Surface Traffic</strong></td>
<td><strong>L07-02 TS-0201</strong>: Basic Departure Management (DMAN)</td>
</tr>
<tr>
<td></td>
<td><strong>L07-02 TS-0202</strong>: Departure Management Synchronized with Pre-Departure Sequencing</td>
</tr>
<tr>
<td></td>
<td><strong>L07-02 TS-0203</strong>: Integration of Surface Management Constraint into Departure Management</td>
</tr>
<tr>
<td></td>
<td><strong>L07-02 TS-0306</strong>: Optimized Departure Management in the Queue Management Process</td>
</tr>
<tr>
<td></td>
<td><strong>L10-02 AO-0205</strong>: Automated Assistance to Controller for Surface Movement Planning and Routing</td>
</tr>
<tr>
<td></td>
<td><strong>L10-08 AO-0501</strong>: Improved Operations in Adverse Conditions through Airport Collaborative Decision Making</td>
</tr>
</tbody>
</table>
Originally done manually using MS word and search
  - Search for descriptions of program A -> identify keywords -> review descriptions of program B containing matching keywords one by one

Jigsaw helped: Review and compare the huge document collection and complete the mapping between the two
  - Published the work at *Aviation Technology Integration and Operations ATIO Conference*
P2: Business Analyst at an Accounting Firm

- **Task:** Analyze unstructured data and identify any linkages between people/companies relevant to a financial fraud

- **Goal:** Find evidence for a financial fraud

- **Before Jigsaw:** Put all documents into an Excel spreadsheet, search for keywords, and read all returned documents

- **Jigsaw helped:** Reveal connections between people & companies that were not easily identifiable
  - Found evidence of a financial fraud after analyzing 100,000 emails
P3: Industrial & Systems Engineering Researcher

- **Task**: Validate her model about company transformation by combining historical company data (5,000+ announcements and news articles of 9 IT firms for 10 years)

- **Goal**: Make sense of the documents and extract keywords for the next step – data mining

- **Jigsaw helped**: Attain a clear understanding of the documents in a short amount of time
P4: Intelligence Analyst at a Police Department

• **Task:** Make sense of daily incident reports and identify patterns, trends, and any top issues in the city

• **Goal:** Find **connections between individuals, places, and other incidents** within accumulated crime reports

• Originally read all the reports individually and tried to remember different connections using printed copies of the documents

• **Jigsaw helped:** Develop a repository of important connections
  • Helped the police arrest a criminal by identifying where he might be
P5: Intelligence Analyst at a National Lab

- **Task**: Review resumes and find a good candidate with a certain specialty

- **Goal**: Examine connections in candidate info and find an expert in a specialized area
  - Skills, publications, co-authors, education, employment history.

- Performed using Analyst’s Notebook but felt limited

- **Jigsaw helped**: Identify possible connections between people and technology
P6: Intelligence Analyst with the US Air Force

- **Task**: Examine budget summaries of R&D programs in the Department of Defense and identify common themes (10,000+ documents from 20+ agencies such as Air Force, Navy, and DARPA)

- **Goal**: Highlight what programs/topics are similar, what makes them similar, and who are working on similar topics

- **Jigsaw helped**: Effectively search for similar tools and technologies that required further investigation
Types of Tasks

- Relationship / connection between entities
  - Targeted investigation rather than seeing the big picture
- Search / comparison
  - If the documents contain specific keywords
- Understanding
  - Getting an overview of documents
- As a communication aid / shared understanding of data
  - Persuasive power in communication
Learning the System

• Learning curve existed
  • “How to better analyze my data using this tool”

• Constructing a frame
  • Which views are most appropriate for my data and task?
  • What entity types do I want to put in this column?

➔ Finding the optimal approach in their own way
Unexpected Use of the System

- Views for evidence/output generation

<table>
<thead>
<tr>
<th>NextGen Operational Improvements</th>
<th>SESAR Operational Improvements and Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>L07-02 TS-0201: Basic Departure Management (DMAN)</td>
<td>L07-02 TS-0202: Departure Management Synchronized with Pre-Departure Sequencing</td>
</tr>
<tr>
<td>L07-02 TS-0202: Departure Management Synchronized with Pre-Departure Sequencing</td>
<td>L07-02 TS-0203: Integration of Surface Management</td>
</tr>
<tr>
<td>Constraint into Departure Management</td>
<td>L07-02 TS-0203: Optimized Departure Management in the Queue Management Process</td>
</tr>
<tr>
<td>L16-03 AO-0501: Improved Operations in Adverse Conditions through Airport Collaborative Decision Making</td>
<td>L10-03 AO-0602: Collaborative Pre Departure Sequencing</td>
</tr>
</tbody>
</table>

A mapping created manually (top) and by Jigsaw (bottom)
Unexpected Use of the System

- Information-dense documents
- Separate docs into several projects
- Merge new incoming documents with an existing Jigsaw project
- Build a historical dataset
Issues and problems

• **Technical issues in the preparation stage**
  - Importing data into Jigsaw
  - Identifying entities

• **Limited filtering options**
  - Not being able to easily select a subset of data in the views
Design Implications

- Supplement automatic entity identification
- Allow flexible data (document) management
  - Provide an ability to easily select a subset of documents
- Empower with numbers
  - Degree centrality, betweenness, closeness
- Consider allowing visualization modification
  - Limit user interaction vs. give more power
- Invest in tutorial
  - Break down into subtopics with use-cases and examples
Design Implications

• Supplement automatic entity identification

• Allow flexible data (document) management
  • Provide an ability to easily select a subset of documents

• Empower with numbers
  • Degree centrality, betweenness, closeness

• Consider allowing visualization modification
  • Limit user interaction vs. give more power

• Invest in tutorial
  • Break down into subtopics with use-cases and examples
Design Implications

• Supplement automatic entity identification

• Allow flexible data (document) management
  • Provide an ability to easily select a subset of documents

• Empower with numbers
  • Degree centrality, betweenness, closeness

• Consider allowing visualization modification
  • Limit user interaction vs. give more power

• Invest in tutorial
  • Break down into subtopics with use-cases and examples
Design Implications

• Supplement automatic entity identification

• Allow flexible data (document) management
  • Provide an ability to easily select a subset of documents

• Empower with numbers
  • Degree centrality, betweenness, closeness

• Consider allowing visualization modification
  • Limit user interaction vs. give more power

• Invest in tutorial
  • Break down into subtopics with use-cases and examples
Design Implications

• Supplement automatic entity identification
• Allow flexible data (document) management
  • Provide an ability to easily select a subset of documents
• Empower with numbers
  • Degree centrality, betweenness, closeness
• Consider allowing visualization modification
  • Limit user interaction vs. give more power
• **Invest in tutorial**
  • Break down into subtopics with use-cases and examples
Contributions

- Identified real-world cases of how an interactive visual system for investigative analysis assisted document sensemaking in various domains and tasks
- Discussed issues and findings that emerged upon the use of the visual analytic system
- Provided design recommendations for the system and future visual analytics tools.
Acknowledgements

• We thank our six professionals for sharing their experience with Jigsaw.

• This work was supported by the National Science Foundation under awards IIS-0915788, CCF-0808863, and the VACCINE Center, a Department of Homeland Security’s Center of Excellence in Command, Control and Interoperability.
The Value of Visualization for Exploring and Understanding Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu
Visualization

“The use of computer-supported, interactive visual representations of data to amplify cognition”
Visualization
-Making pretty pictures-
A cognitive process
   Internalize an understanding

Visuals help us think
Provides a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization
Presentation
Analysis

1. Presentation
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Simply presenting data **visually** can have a profound impact.
Nate Osborne
Nitya Noronha
Ameya Zambre
Pratik Zaveri

Gun ownership in New York counties

http://www.lohud.com/apps/pbcs.dll/article?AID=201212230054&nclick_check=1
Map-Where-gun-permits-your-neighborhood-?nclick=1ouc1sck=1
Frequent presentation goals
Clarify
Focus
Highlight
Simplify

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

2. Analysis
understand, compare, decide, judge, evaluate,
assess, determine, ...
Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning

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

W. Cleveland
*The Elements of Graphing Data*
Visualization Techniques for Analysis
When the standard charts work, use them

Frequent analysis goals
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first

Preconceptions about Visualization Utility
Answering specific questions and accomplishing specific analytic tasks
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for
1. **Visualization** is more than just answering a specific question (as is often the case for automated analysis methods)
It also is about the investigative analysis process, which helps us to learn about, develop awareness of, and generate trust in the data, its domain, and its context.

**Learning, awareness, trust, context**

---

**Many Data Analysis Approaches**
- Statistics
- Database & information retrieval
- Data mining
- Machine learning

Use them when they get the job done!
2. **Visualization**, primarily through its interactive capabilities, promotes a dialog between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

**Visualization**

“The use of computer-supported, interactive visual representations of data to amplify cognition”
3. **Visualization** rapidly and efficiently facilitates flexible exploration to foster both broad and deep understanding of the information in a data collection.

Broad and deep understanding quickly

Visualization most useful in **exploratory data analysis**
Don’t know what you’re looking for
Don’t have a priori questions
Want to know what questions to ask
insights

spontaneous aha! moments vs.
knowledge-building & model confirmation

Chang et al, *IEEE CG&A* '09

SellTrend

Liu, Stasko, Sullivan, *TVCG (InfoVis)* '09
SnapShot

Pileggi, Stolper, Boyle, Stasko
TVCG (InfoVis) ’12

“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
Example

Reviews of wines from Tuscany from ‘02, ‘03, ‘07-’10

Text: review narrative

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

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

1657 reviews
from database of 150,000 reviews
1. Visualization is more than just answering a specific question (as is often the case for automated analysis methods)
It also is about the investigative analysis process, which helps us to learn about, develop awareness of, and generate trust in the data, its domain, and its context.

Learning, awareness, trust, context

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

**Q & A dialog through interaction**

Repeated interactive queries

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

**Broad and deep understanding quickly**

Varieties, colors, reviewers, producers, concepts, ...
Insights
No great aha! – More learning & knowledge building
Domain knowledge (wines) matters

More visualization benefits
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions
Take Aways
Presentation & analysis

Interaction provides the power
Exploring & developing questions

Acknowledgments

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

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

This material is supported in part by:
National Science Foundation
Console

![Console Image]

Document View

![Document View Image]
List View

Graph View
Document Cluster View

![Document Cluster View](image)

Document Grid View

![Document Grid View](image)
Calendar View

WordTree View
Circular Graph View

Scatterplot View
The Value of Visualization for Exploring and Understanding Data

John Stasko
School of Interactive Computing
Georgia Institute of Technology
stasko@cc.gatech.edu

Graphics Interface 2013
Visualization

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

- Making pretty pictures -
A cognitive process
  Internalize an understanding

Visuals help us think
Provides a frame of reference, temporary storage area
Cognition → Perception
Pattern matching
Applications of Visualization

Presentation

Analysis

1. **Presentation**
Communicate data and ideas
Explain and inform
Influence and persuade
Provide evidence and support
Tufte, Vol. 1

The exploding internet
2008

http://www.newscientist.com/data/images/ns/cms/mg20227062.200/mg20227062.200-6_1000.jpg
THE ALMIGHTY DOLLAR
MAPPING DISTRIBUTION OF INCOME BY RELIGIOUS BELIEF

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

All of the goals which took the teams to Euro 2012

Group A

Group B

Simon Scarr
South China Morning Post

Click on any image below to see full description

http://thebeermongers.com/beers/

Simply presenting data *visually* can have a profound impact
Gun ownership in New York counties

Frequent presentation goals
- Clarify
- Focus
- Highlight
- Simplify

May just show a few variables and/or a subset of the data cases
2. Analysis
understand, compare, decide, judge, evaluate,
assess, determine, ...

Many Data Analysis Approaches
Statistics
Database & information retrieval
Data mining
Machine learning
“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland
*The Elements of Graphing Data*

**Visualization Techniques for Analysis**
When the standard charts work, use them
Frequent analysis objectives
Show many variables
Illustrate overview and detail
Facilitate comparison

Display may not be easy to interpret at first
Preconceptions about Visualization Utility
Answering a specific question and accomplishing a specific analytic task
Generating unexpected, serendipitous discoveries and insights
“Finding a needle in a haystack”

Yes, but not what it’s best for

So what is visualization most useful for?
1. **Visualization** is more than just answering a specific question (as is often the case for automated analysis methods). It also is about the investigative analysis process, which helps us to learn about, develop awareness of, and generate trust in the data, its domain, and its context.

**Learning, awareness, trust, context**

### Many Data Analysis Approaches
- Statistics
- Database & information retrieval
- Data mining
- Machine learning

Use them when they get the job done!
2. Visualization, primarily through its interactive capabilities, promotes a dialog between analysts and their data by allowing a diverse and flexible set of questions to be asked and answered about a data collection and by spurring the generation of new questions.

Q & A dialog through interaction

Visualization
“The use of computer-supported, interactive visual representations of data to amplify cognition”
3. **Visualization** rapidly and efficiently facilitates flexible exploration to foster both broad and deep understanding of the information in a data collection.

**Broad and deep understanding quickly**

Visualization most useful in *exploratory data analysis*
- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask
insights

spontaneous aha! moments
vs.
knowledge-building & model confirmation

Chang et al
*IEEE CG&A* ’09

SellTrend

Liu, Stasko, Sullivan
*TVCG (InfoVis)* ’09
SnapShot

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

“Putting the pieces together”
Example

Reviews of wines from Tuscany from ’02, ’03, ’07-’10

Text: review narrative

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

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

1657 reviews from database of 150,000 reviews
1. **Visualization** is more than just answering a specific question (as is often the case for automated analysis methods). It also is about the investigative analysis process, which helps us to learn about, develop awareness of, and generate trust in the data, its domain, and its context.

**Learning, awareness, trust, context**

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

**Q & A dialog through interaction**

Repeated interactive queries

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

**Broad and deep understanding quickly**

Varieties, colors, reviewers, producers, concepts, ...
**Insights**

No great aha! – More learning & knowledge building
Domain knowledge (wines) matters

**Evaluation**

Perhaps key open problem/challenge in visualization

By definition, extremely difficult
Proposed approaches:
- Insight-based (Saraiya et al, TVCG ’05)
- Case study (Shneiderman & Plaisant, BELIV ’06)
More visualization benefits
Ease of specifying queries
Opportunistic discovery of relevant data
Spurs the generation of new questions

Take Aways
Presentation & analysis

Interaction provides the power
Exploring & developing questions
Acknowledgments

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

- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788
http://pixel.ecn.purdue.edu:8080/~purpl/VACCINE_Resources/iLEAPS_WEAA_%20Interview.zip
DEFINING THE TYPICAL DHS MAP
Raechel A. Bianchetti
Robert E. Roth
Justine I. Blanford
Anthony C. Robinson
Alan M. MacEachren

OVERVIEW
- Mission
- Background
- Method
- Results
- Conclusions
MISSION

- To determine whether a common map design existed across a range of DHS maps.
  - Evaluate the use of symbol standards across maps
  - Evaluate the types of symbols used on those maps

BACKGROUND

- Interviews
- Discussions
- Communication
- Card sorting
- Content analysis
- Action
EXISTING STANDARDS

METHODS

- Content Analysis (Berelson 1952)
  - A method used to uncover themes within a collection of non-numerical documents
    - Literature
    - Art
    - Maps
  - Quantitative and Qualitative methods for evaluating the code
CONTENT ANALYSIS & MAPS

- National Geographic Magazine (Lutz and Collins 1993)
- provincial travel maps in Canada (Grant and Keller 1999)
- state-level public health maps depicting cancer rates (Parrott et al. 2007)
- maps generated in support of political discourse (Edsall 2007, Muehlenhaus 2010)
- national topographic maps (Kent 2009, Kent and Vujakovic 2009, Kent and Vujakovic 2011)
- reference maps of European urban centers (Ciolkosz-Styk 2011a, Ciolkosz-Styk 2011b)
- Goode’s World Atlas maps (Muehlenhaus 2011)

METHOD OF CONTENT ANALYSIS

1. Gather documents
   - 76 static maps provided by DHS at interviews or through HSIN
   - Maps are from 5 agencies: OIP, FEMA, IICD.
2. Develop a set of codes relating to the key research theme
   - Chose cartographic design choices
3. Apply the codes to the maps
   - Codes were recorded using excel spreadsheet
LIMITATIONS

- The map must
  - include point symbols and the array of point symbols must be visually discriminable in order to allow for comparison to the ANSI standard and the other emergency and hazards symbol standards reviewed in Section 2.
  - cannot be a screenshot from a web portal in order to ensure that all map content was visible in the document (e.g., all included feature types/map symbols, map elements, etc.).

- The contents of inset maps included as part of a larger map product were not considered in order to eliminate double counting of map features and to eliminate redundancy of thematic information; the presence/absence of inset maps instead was included as a code in the coding scheme for subsequent description of inset map contents.

CODING SCHEME

THemes
- Map Purpose
- Map Content
- Map Design

Categories
- Map objective (o)
- Map source (s)
- Feature type (f)
- Feature condition (c)
- Iconicity (i)
- Symbol Visual Variable (v)
- Basemap (b)
- Map Extent (x)
- Map Elements (e)
CATEGORIES

Map Objective
- Preparedness
- Response
- Recovery
- mitigation

Map Source
- Customs and Border Control
- Federal Emergency Management Agency
- National Operations Center
- Office of Infrastructure Protection
- US Coast Guard

CATEGORIES

Feature Type
- ANSI
- IAL
- MISC

Feature Condition
- Operational Status
- Level of Danger
- Information Trustworthiness
CATEGORIES

Symbol Visual Variables
- Color Hue
- Pattern
- Shape

Iconicity
- Pictorial
- Associative
- Geometric

CATEGORIES

Basemap
- Vector
- Remotely Sensed Imagery
- Shaded Relief

Map Extent
- City
- City-to-State
- Regional
- National
CATEGORIES

Map Elements
- Indication of Orientation
- Indication of Scale
- Inset Maps
- Map Legend

THE MAPS: SPORTING EVENTS
THE MAPS: HURRICANE

THE MAPS: NUCLEAR DETONATION
THE MAPS: SPORTING EVENT

Not all of the maps aligned with the emergency management cycle

Thematic maps vs. reference maps

Map objectives

Most maps were OIP (n=42)
Map context indicated the agency

Map source

RESULTS: MAP OBJECTIVES
# RESULTS

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature Type Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting with 198 ANSI symbols we ended with only 26</td>
<td>Only 7 examples of levels being used</td>
</tr>
<tr>
<td>Additional 84 feature types added</td>
<td>2-5 levels per map</td>
</tr>
</tbody>
</table>

## RESULTS

<table>
<thead>
<tr>
<th>Iconicity</th>
<th>Visual Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric was most common</td>
<td>Only 8 levels instances</td>
</tr>
<tr>
<td>Star Symbols for “everything”</td>
<td>Most common variable used is size</td>
</tr>
</tbody>
</table>
RESULTS

Base Map
- Vector Maps
- Only FEMA and OIP used shaded relief maps
- MODIS Case

Map Extent
- City (n=40) and City-State (n=27) were most prominent
- City had much fewer points per map than City-State

RESULTS

MAP ELEMENTS
- 73/76 used map elements
- Orientation n=71
- Legend n=73
- Scale bar n=69
- Inset Maps n=58**
RESULTS: MAP ELEMENTS

<table>
<thead>
<tr>
<th></th>
<th>SCALE</th>
<th>ORIENTATION</th>
<th>LEGENDS</th>
<th>INSET MAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEMA</td>
<td>87%</td>
<td>93%</td>
<td>93%</td>
<td>93%</td>
</tr>
<tr>
<td>CBP</td>
<td>100%</td>
<td>83%</td>
<td>100%</td>
<td>33%</td>
</tr>
<tr>
<td>USCG</td>
<td>91%</td>
<td>64%</td>
<td>91%</td>
<td>18%</td>
</tr>
</tbody>
</table>

CONCLUSIONS

- This content analysis is a method for critically analyzing the maps created at DHS to support a bottom-up process of symbol standard development.
- There is no typical DHS map (surprised!?)
- The ANSI standard
  - Not typically used
  - Not used as they were meant to be
THANK YOU.

- U.S. Department of Homeland Security
Design & Use Guidelines

For Interactive Maps: A Case Study
http://www.slideshare.net/réroth

@RobertERoth
Robert E. Roth | réroth@wisc.edu

@alanGeoVISTA
Alan M. MacEachren | maceachren@psu.edu

#aag2013
Los Angeles, CA | April 12th, 2013
REPRESENTATION creates uncertainty

default
INTERACTION OVERCOMES UNCERTAINTY(?)
# BACKGROUND: INTERACTION SCIENCE

**Roth (2013)**

*Cartographic Interaction: What We Know and What We Need to Know*

<table>
<thead>
<tr>
<th>Question</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What?</strong></td>
<td>interaction vs. interface • digital vs. analog • stages of interaction • system response time • interactive maps vs. mapping systems vs. map mashups</td>
</tr>
<tr>
<td><strong>Why?</strong></td>
<td>visual thinking • insight • exploration/confirmation/synthesis/presentation • discussion / debate • decision making</td>
</tr>
<tr>
<td><strong>When?</strong></td>
<td>interface complexity • interface freedom &amp; flexibility • convention &amp; standardization • work productivity • efficiency and effectiveness</td>
</tr>
<tr>
<td><strong>Who?</strong></td>
<td>user-centered design • user ability • user expertise • user motivation • adaptive cartography • role-based interfaces • geocollaboration</td>
</tr>
<tr>
<td><strong>Where?</strong></td>
<td>input capabilities • bandwidth size • processing power • display capabilities • interoperability • mobile mapping • open source development</td>
</tr>
<tr>
<td><strong>How?</strong></td>
<td>interaction primitives • stages of interaction • interaction strategies • objectives, operators, operands • interface styles • interface design</td>
</tr>
</tbody>
</table>

---

#aag2013 Los Angeles, CA | April 12th, 2013

[@RobertERoth](https://twitter.com/RobertERoth)  [@alanGeoVISTA](https://twitter.com/alanGeoVISTA)

Robert E. Roth  Alan M. MacEachren
**BACKGROUND: HOW?**

Extant Taxonomies of Interaction Primitives

- Yi et al. (2007)
- Buja et al. (1996)
- Haber & McNabb (1990)
- Crampton (2002)
- Andrienko et al. (2003)
- MacEachren et al. (1999)
- Chi & Riedl (1998)
- Amar et al. (2005)
- Shepherd (1995)
- Wehrend & Lewis (1990)
- Peuquet (1994)
- Shneiderman (1996)
- Becker & Cleveland (1987)
- Keim (2002)
- Masters & Edsall (2000)
- Wehrend (1993)
- Dykes (1997)
- Becker & Cleveland (1987)
- Persson et al. (2006)
- Dix & Ellis (1998)
- Chuah & Roth (1996)
- Edsall et al. (2008)

*Cartographic Interaction Primitives: Framework & Synthesis*
OBJECTIVES: the intention of the user in completing the interaction
OPERATORS:
the functions provided by the interface to support the user’s objectives
OPERANDS:
characteristics of the recipient of the operator
zoom to the extent of the project area

download shapefiles of the storm event

select a map feature or group of features and get details

change the orientation of the map from north as up

select map features through direct manipulation

switch among multiple map representation strategies

select a point to change the attribute data

string together a series of maps into a slideshow

click on the icon indicating where the picture was taken and see the metadata

save a map as a project file to allow other people to collaborate on it

adjust the display scale to restrict the extent of the view

save the data on the map as a table

change the map to show the distribution of colorectal cancer for a different race

create a .kml file in ArcMap to view in Google Earth

select an animal and show on the map the extent of its habitat

manipulate the attribute values of a map feature

display more than one spatial data set in the same coordinate space
draw a boundary to
BACKGROUND: PRIMITIVE TAXONOMY

objectives

identify
compare
rank
associate
delineate

operators

reexpress
arrange
sequence
resymbolize
overlay
reproject

operands

pan
zoom
filter
search
retrieve
calculate

space-alone
attributes-in-space
space-in-time

Roth (2012)
An Empirically-Derived Taxonomy of Cartographic Interaction Primitives
CASE STUDY: GEOVISTA CrimeViz
http://www.geovista.psu.edu/CrimeViz

Project Personnel:
Kevin S. Ross | kevin@kross.com
Benjamin G. Finch | bgf111@psu.edu
Wei Luo | wul132@psu.edu
Craig A. McCabe | CMcCabe@esri.com
Ryan Mullins | ryanmullins@psu.edu
Scott Pezanowski | spezanowski@psu.edu
Camilla Robinson | clr281@psu.edu

Contact:
Alan M. MacEachren | maceachren@psu.edu
Robert E. Roth | reroth@wisc.edu
ASSAULT

case number: 20091110435
date: 2009-11-23
district: 5
description: Terroristic threats abab
address: FIRST BLK S 16TH ST

Click here for Street View

#aag2013
Los Angeles, CA | April 12th, 2013

@RobertERoth @alanGeoVISTA
Robert E. Roth  Alan M. MacEachren
incident filtering
METHOD: INTERACTION STUDY
identifying prototypically successful interaction strategies

Participants:
- \( n=10 \)
- law enforcement personnel at Harrisburg Bureau of Police

Interaction Study Protocol:
- **15** tasks: **5** objectives x **3** operands
- closed-ended: questions had “**correct**” answers
- **3** minute time limit per question

Interaction Analysis
- interaction logging by operator
## Operator Logging

### Filter (F)

| Space                  | Menu Selection Numerical Stepper by 'District'
|                       | Form Fill-in by 'District'
|                       | Menu Selection for 'Reset Advanced Features'
| Attribute             | Menu Selection by 'UCR Primary,' 'UCR Secondary' and 'MO'
|                       | Form Fill-in by 'UCR Primary,' 'UCR Secondary' and 'MO'
|                       | Menu Selection for 'Reset Basic Filters'
| Time                  | Menu Selection Numerical Stepper for 'From' and 'To' Linear Filtering
|                       | Form Fill-in 'From' and 'To' Linear Filtering
|                       | Menu Selection Shortcuts for Linear Filtering ('Week,' 'Month,' 'Year,' 'All')
|                       | Direct Manipulation of 'Hours,' 'Months,' and 'Days' Widgets for Cyclical
|                       | Menu Selection Shortcuts for Cyclical Filtering ('All,' 'None,' 'Winter,' etc.)
|                       | Menu Selection for 'Reset Temporal Parameters'

### Search (S)

| Space                  | Form Fill-in Search by 'Address'
| Attribute             | Form Fill-in Search by 'Report #'

### Retrieve (R)

| Attribute             | Direct Manipulation Mouse-Over of Hexagon Bin
|                       | Direct Manipulation Mouse-Over of Crime Incident
|                       | Direct Manipulation Click of Crime Incident
|                       | Menu Selection to Activate Street View
| Time                  | Direct Manipulation Mouse-Over of Data Layer Element
|                       | Direct Manipulation Mouse-Over of Histogram Bin
## Operator Logging

### Reexpress (X)

| Time | Menu Selection for Linear versus Composite Time |

### Sequence (Q)

| Time | Direct Manipulation Click of the 'Play' (Loop) and 'Pause' VCR Controls |

### Overlay (O)

| Space | Menu Selection for Basemap Type ('Map', 'Sat', 'Terrain') |
| Space | Menu Selection Checkboxes for Point/Line Data Layers ('Schools', etc.) |
| Attribute | Menu Selection Radio Buttons for Polygonal Data Layers ('Districts', etc.) |
| Attribute | Menu Selection 'Reset' Additional Data Layers |

### Pan (P)

| Space | Direct Manipulation Click+Drag on Map |
| Space | Direct Manipulation 'Resent Extent' Control |
| Time | Direct Manipulation Click on Histogram Bin |
| Time | Direct Manipulation Click on 'Back' and 'Step' VCR Controls |
| Time | Direct Manipulation of Histogram Slider Bar (When Entirety is Not Displayed) |

### Zoom (Z)

<p>| Space | Direct Manipulation Double-Click on Map |
| Space | Direct Manipulation Click on Hexagon Grid |
| Space | Direct Manipulation Click on Data Layer Element |
| Space | Direct Manipulation '+' and '-' Controls |
| Space | Direct Manipulation 'Resent Extent' Control |
| Time | Menu Selection for Binning Unit |</p>
<table>
<thead>
<tr>
<th>Task</th>
<th>Complete</th>
<th>Correct</th>
<th>Avg Time</th>
<th>Freq.</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By Objective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>30 (100%)</td>
<td>26 (87%)</td>
<td>0:31</td>
<td>3.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Compare</td>
<td>30 (100%)</td>
<td>25 (83%)</td>
<td>0:58</td>
<td>7.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Rank</td>
<td>29 (96.7%)</td>
<td>28 (93%)</td>
<td>0:56</td>
<td>9.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Associate</td>
<td>30 (100%)</td>
<td>24 (80%)</td>
<td>1:06</td>
<td>7.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Delineate</td>
<td>27 (90%)</td>
<td>20 (67%)</td>
<td>1:27</td>
<td>9.8</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>By Operand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>50 (100%)</td>
<td>47 (94%)</td>
<td>0:34</td>
<td>5.9</td>
<td>2.9</td>
</tr>
<tr>
<td>Attribute</td>
<td>46 (92%)</td>
<td>41 (82%)</td>
<td>1:16</td>
<td>8.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Time</td>
<td>50 (100%)</td>
<td>35 (70%)</td>
<td>1:09</td>
<td>8.2</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>146 (97%)</td>
<td>123 (82%)</td>
<td>1:00</td>
<td>7.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Task</td>
<td>Complete</td>
<td>Correct</td>
<td>Avg Time</td>
<td>Freq.</td>
<td>Diversity</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>By Objective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>30 (100%)</td>
<td>26 (87%)</td>
<td>0:31</td>
<td>3.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Compare</td>
<td>30 (100%)</td>
<td>25 (83%)</td>
<td>0:58</td>
<td>7.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Rank</td>
<td>29 (96.7%)</td>
<td>28 (93%)</td>
<td>0:56</td>
<td>9.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Associate</td>
<td>30 (100%)</td>
<td>24 (80%)</td>
<td>1:06</td>
<td>7.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Delineate</td>
<td>27 (90%)</td>
<td>20 (67%)</td>
<td>1:27</td>
<td>9.8</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>By Operand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>50 (100%)</td>
<td>47 (94%)</td>
<td>0:34</td>
<td>5.9</td>
<td>2.9</td>
</tr>
<tr>
<td>Attribute</td>
<td>46 (92%)</td>
<td>41 (82%)</td>
<td>1:16</td>
<td>8.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Time</td>
<td>50 (100%)</td>
<td>35 (70%)</td>
<td>1:09</td>
<td>8.2</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>146 (97%)</td>
<td>123 (82%)</td>
<td>1:00</td>
<td>7.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Task</td>
<td>Complete</td>
<td>Correct</td>
<td>Avg Time</td>
<td>Freq.</td>
<td>Diversity</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>By Objective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>30 (100%)</td>
<td>26 (87%)</td>
<td>0:31</td>
<td>3.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Compare</td>
<td>30 (100%)</td>
<td>25 (83%)</td>
<td>0:58</td>
<td>7.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Rank</td>
<td>29 (96.7%)</td>
<td>28 (93%)</td>
<td>0:56</td>
<td>9.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Associate</td>
<td>30 (100%)</td>
<td>24 (80%)</td>
<td>1:06</td>
<td>7.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Delineate</td>
<td>27 (90%)</td>
<td>20 (67%)</td>
<td>1:27</td>
<td>9.8</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>By Operand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>50 (100%)</td>
<td>47 (94%)</td>
<td>0:34</td>
<td>5.9</td>
<td>2.9</td>
</tr>
<tr>
<td>Attribute</td>
<td>46 (92%)</td>
<td>41 (82%)</td>
<td>1:16</td>
<td>8.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Time</td>
<td>50 (100%)</td>
<td>35 (70%)</td>
<td>1:09</td>
<td>8.2</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>146 (97%)</td>
<td>123 (82%)</td>
<td>1:00</td>
<td>7.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Task</td>
<td>Frequency</td>
<td>Extensiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>By Operator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reexpress (X)</td>
<td>78</td>
<td>50 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence (Q)</td>
<td>32</td>
<td>14 (9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overlay (O)</td>
<td>57</td>
<td>46 (31%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan (P)</td>
<td>121</td>
<td>41 (27%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom (Z)</td>
<td>127</td>
<td>66 (44%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filter (F)</td>
<td>240</td>
<td>90 (60%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search (S)</td>
<td>42</td>
<td>29 (19%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrieve (R)</td>
<td>395</td>
<td>106 (71%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1092</td>
<td>150 (100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>By Operand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>175</td>
<td>49 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>248</td>
<td>117 (78%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>543</td>
<td>94 (63%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>Frequency</td>
<td>Extensiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>By Operator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reexpress (X)</td>
<td>78</td>
<td>50 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence (Q)</td>
<td>32</td>
<td>14 (9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overlay (O)</td>
<td>57</td>
<td>46 (31%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pan (P)</td>
<td>121</td>
<td>41 (27%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoom (Z)</td>
<td>127</td>
<td>66 (44%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filter (F)</td>
<td>240</td>
<td>90 (60%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search (S)</td>
<td>42</td>
<td>29 (19%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrieve (R)</td>
<td>395</td>
<td>106 (71%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1092</td>
<td>150 (100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>By Operand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>175</td>
<td>49 (33%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>248</td>
<td>117 (78%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>543</td>
<td>94 (63%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RESULTS: SUBOPTIMAL USE PROFILES

Persona: Excessive-Filterer; search vs. filter by sophistication

IDENTIFY BY SPACE-ALONE

<table>
<thead>
<tr>
<th>Time</th>
<th>Participant D</th>
<th>Participant H</th>
<th>Participant I</th>
<th>Participant J</th>
<th>Participant A</th>
<th>Participant F</th>
<th>Participant C</th>
<th>Participant B</th>
<th>Participant G</th>
<th>Participant E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
</tr>
<tr>
<td>0:30</td>
<td></td>
<td></td>
<td></td>
<td>S yes</td>
<td>R yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IDENTIFY BY ATTRIBUTES-IN-SPACE

<table>
<thead>
<tr>
<th>Time</th>
<th>Participant F</th>
<th>Participant B</th>
<th>Participant A</th>
<th>Participant I</th>
<th>Participant H</th>
<th>Participant D</th>
<th>Participant G</th>
<th>Participant J</th>
<th>Participant C</th>
<th>Participant E</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
<td>S yes</td>
</tr>
<tr>
<td>0:30</td>
<td></td>
<td>S yes</td>
<td></td>
<td>S yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Space-Alone  Attributes-in-Space  Space-in-Time
RESULTS: SUBOPTIMAL USE PROFILES

Persona: Mistaken-Reexpresser, Unsure-Retriever

COMPARE BY SPACE-IN-TIME

<table>
<thead>
<tr>
<th>Time</th>
<th>Participant A</th>
<th>Participant B</th>
<th>Participant C</th>
<th>Participant D</th>
<th>Participant E</th>
<th>Participant F</th>
<th>Participant G</th>
<th>Participant H</th>
<th>Participant I</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Space-Alone  Attributes-in-Space  Space-in-Time

#aag2013
Los Angeles, CA | April 12th, 2013

@RobertERoth  @alanGeoVISTA
Robert E. Roth  Alan M. MacEachren
RESULTS: SUBOPTIMAL USE PROFILES
Persona: Blind-Sequencer

ASSOCIATE BY ATTRIBUTES-IN-SPACE

Space-Alone  Attributes-in-Space  Space-in-Time

Participant J  Participant B  Participant H  Participant C  Participant E  Participant I  Participant F  Participant G  Participant A  Participant D

0:00  F  F  F  F  F  F  F  F  F  F
0:30  F  F  F  F  F  F  F  F  F  F
1:00  F  F  F  F  F  F  F  F  F  F
1:30
2:00
2:30

@RobertERoth  @alanGeoVISTA
Robert E. Roth  Alan M. MacEachren
Design and Use Guidelines for Interactive Maps: A Case Study

@RobertERoth
Robert E. Roth | reroth@wisc.edu

@alanGeoVISTA
Alan M. MacEachren | maceachren@psu.edu

#aag2013
Los Angeles, CA | April 12th, 2013
## Task (Objective+Operand) Protocol

<table>
<thead>
<tr>
<th>Identify</th>
<th><strong>Space</strong></th>
<th>On what street did incident #20101100945 occur?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Attribute</strong></td>
<td>What type of crime is incident #20101100894, which occurred at 200 Herr Street?</td>
</tr>
<tr>
<td></td>
<td><strong>Time</strong></td>
<td>How many total crime incidents occurred in District #5 on September 1st, 2010?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compare</th>
<th><strong>Space</strong></th>
<th>Are Fire Station #2 and Fire Station #8 in the same police district?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Attribute</strong></td>
<td>Is incident #20101100945, which occurred on Market Street, the same type of crime as incident #20101100608, which occurred on 3rd Street?</td>
</tr>
<tr>
<td></td>
<td><strong>Time</strong></td>
<td>In October 2010, how many more crime incidents occurred within Harrisburg on Sundays compared to Mondays?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th><strong>Space</strong></th>
<th>What school in Harrisburg is closest to Interstate-83?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Attribute</strong></td>
<td>Which crime type was the most common in District #1 on November 5th, 2010?</td>
</tr>
<tr>
<td></td>
<td><strong>Time</strong></td>
<td>From 2006 through 2010, which month exhibited the highest frequency of crime incidents across Harrisburg?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Associate</th>
<th><strong>Space</strong></th>
<th>Which route should Harrisburg citizens take to get to the west bank of the Susquehanna River during an evacuation related to Three Mile Island?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Attribute</strong></td>
<td>From 2006 through 2010, is the geographic pattern of prostitution (16) related to the geographic pattern of sex offenses (17)?</td>
</tr>
<tr>
<td></td>
<td><strong>Time</strong></td>
<td>From 2006 through 2010, does the trend in crime increase or decrease across Harrisburg from noon (12:00) to midnight (24:00)?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delineate</th>
<th><strong>Space</strong></th>
<th>Which police districts exhibit clusters of increased criminal activity from 2006 through 2010?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Attribute</strong></td>
<td>From 2006 through 2010, how many different ways was fraud (11) committed across Harrisburg?</td>
</tr>
<tr>
<td></td>
<td><strong>Time</strong></td>
<td>The 2008 spike in robbery (03) incidents in Harrisburg spanned across which months?</td>
</tr>
</tbody>
</table>
Geo-Social Visual Analytics with Applications to Catastrophic Risk Management

Wei Luo
Ph.D Candidate
GeoVISTA Center, Department of Geography
Penn State University
Visual Analytics

- Visual analytics is defined as the science of analytical reasoning facilitated by interactive visual interfaces (Thomas and Cook 2005).

- Objective: Reducing America's vulnerability to terrorism and natural disasters.

Future Visual Analytics

Introduction

- Geographical context and social network context demonstrate strong conceptual and observed overlaps, but the integration of geography and social network contexts has not received sufficient attention to develop a comprehensive understanding of their interaction or their impact on outcomes of interest.

Goal

- This research develops theoretical and practical models of geosocial context through geovisual analytics approaches that account for geographical, social network, and conceptual dimensions of that context.
Objectives

- Develop a conceptual model to understand the interaction of geography and social network at the theoretical level.
- Develop a geosocial visual analytics tool at a geographical region level (e.g., county, state, and country) to understand the role of spatial proximity in shaping the international trade network (ITN) across different geographical regions.
- Develop a geosocial visual analytics tool to support decision-making processes in airborne disease control (e.g., influenza) with agent-based epidemic models.

GeoSocial Conceptual Framework

Based on fundamental categories of embeddedness (Hess 2004)
GeoSocial Visual Analytics

The convergence of the iterated correlations (CONCOR) algorithm (Breiger et al. 1975).

GeoSocialApp with International Trade Network in 2005

http://www.geovista.psu.edu/GeoSocialApp/
Selected Results

Potential Applications

- Financial contagion on the international trade network (Kali and Reyes 2010).
- The impact of climate change on human migrants within the U.S.
GeSocial Agent-Based Epidemic Visual Analytics

Does the proposed multi-scale geovisual analytics approach enable new control scenarios compared to agent-based epidemic models alone?

Does the distribution of high-risk individuals enable new control scenarios?

Does the visualization of disease transmission simulation process enable new control scenarios?

To achieve three goals, I will do a task analysis with and without the tools and follow up with post-task interviews of the expert participants.

Expected Insights
Contributions

- The proposed conceptually geo-social model provides a theoretical framework to study geo-social relationships.

- The second case study transforms identified patterns into useful knowledge to directly support decision making in epidemics. This contribution fills one major gap in visual analytics: how to transform knowledge developed through visual exploration into decision-making support directly.

- Geosocial visual analytics presented here are applicable to other geographical and network contexts among geographical regions (e.g., human migrants among different states in the U.S.), as well as at the individual level (e.g., facebook and twitter network over space).
References

Geovisualization to Geovisual Analytics: Visual Reasoning with Big & Messy Data

Alan M. MacEachren
GeoVISTA Center
Department of Geography
Penn State University
maceachren@psu.edu

GI Forum, Salzburg, Austria, July 5, 2012

Outline

- Geovisualization and the roots of Geovisual Analytics:
  - cartography → geovisualization → geovisualization + KDD
  - sample applications
- Geovisual analytics:
  - Definition, Drivers, Conceptual frameworks
  - Challenges
  - scaling humans to cope with big data
  - supporting analytical reasoning to generate and apply the information
  - extracting meaningful information from messy, heterogeneous data
- 2 Case study projects: info foraging and contextualization, sensemaking and situational awareness
- Summary & GIScience challenges

from cartography, through geovisualization, to geovisual analytics:
data → info → knowledge → knowledge application

- (late 1980s) Visualization in Scientific Computing (VisSC): prompting a shift from cartography → geovisualization –
- (1990s) dynamic geovisualization
  - (Cartography + VisSC + EDA + HCI + InfoVis): prompting hypotheses and enabling insight

Exploratory Spatial-Temporal Analysis Toolkit (ESTAT): user-centered design / usability engineering

ESTAT, (exploratory spatio-temporal analysis toolkit) is designed to facilitate visual exploration for cancer research.
18 epidemiologists at NCI and Penn State HMC participated in assessing and refining the ESTAT environment.
Evaluating ESTAT has involved the following knowledge elicitation techniques: focus groups, card-sorting, verbal protocol analysis, case study collaboration … and resulted in a hierarchical approach to user-centered design for exploratory geovis in epi...
Dynamic representation = user and data-driven change

is there a pattern?

is your answer different?

Animation: Fundamental issues in visual cognition
pace: test pattern – growing region

- map readers answer more quickly and identify more patterns correctly when using animated maps than when using static small-multiple maps
- pace and cluster coherence interact so that different paces are more effective for identifying certain types of clusters (none vs. subtle vs. strong)
- there are some gender differences in the animated condition (males faster)

GeoViz Toolkit: Flu Data Analysis

http://www.google.org/about/flutrends/download.html


Use subject to Terms of Service (http://www.google.com/accounts/TOS)


Each week begins on the Sunday indicated for the row

For more information, please visit http://google.org/flutrends

Frank Hardisty
hardisty@psu.edu

http://www.google.org/about/flutrends/download.html

For more information, please visit http://google.org/flutrends

Author: Alan M. MacEachren – Please do not post online
GeoViz Toolkit: Flu Data Analysis

Lancaster Case Study: Transforming views to identify local patterns

Spatial representation of Wards containing LSOAs sized by number of electricity meters and coloured by Mosaic Group... arranged without overlap in a space-filling geographically ordered Treemap... coloured by average consumption... coloured to show the difference between the LSOA ave. consumption and the ave. consumption for Lancaster (red = more than, blue = less than)

http://www.gicentre.org/houseprices/

Data → info → knowledge → knowledge application

from cartography, through geovisualization, to geovisual analytics: data → info → knowledge → knowledge application

visualization is critical here... also here... and here... and here...

- (late '80s) Visualization in Scientific Computing (VisSC), prompting shift from cartography → geovisualization →
- ('90s) dynamic geovisualization = cartography + VisSC + EDA + HCI + InfoVis, prompting hypothesis & enabling insight
- (2000s) geovisualization + computation = (…) + KDD: turning data to info, uncovering patterns and relationships, supporting knowledge construction
Multivariate patterns related to undesirable cancer control situations

developing and applying visual analytic methods and tools

Self-organized map based computational categorization of cancer + covariates
• more similar colors represent more similar places for cancer and covariates
• "local" for cancer diagnosis is better = cancer identified early (localized in body)

Counties grouped due to (a) well below average % of cancer cases diagnosed at "local" stage for two age groups: 40-64 and 65+ and (b) low incomes, low rate of MDs, and low # of hospitals with oncology service

Geovisual analytics: 3 challenges

• scaling humans to cope with big data
• extracting meaningful information from messy, heterogeneous data
• understanding and supporting the analytical reasoning processes needed to utilize the data

Georg Garder (Tues. keynote): maps "enable spatial problem solving (reasoning, planning)"

from cartography, through geovisualization, to geovisual analytics:
data → info → knowledge → knowledge application

• (late ‘80s) Visualization in Scientific Computing (VisSC): prompting shift from cartography → geovisualization →
• (‘90s) dynamic geovisualization – (Cartography + VisSC + EDA + HCI + InfoVis): prompting hypotheses & enabling insight
• (2000s) geovisualization + computation – (…) + KDD: turning data to info, uncovering patterns and relationships, supporting knowledge construction
• (geo)visual analytics →

Some scientific and social challenges

• Big Data
• Complex, often ill-structured problems
• Diverse sources and kinds of information

Geovisual Analytics

• Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces ... thus, the focus is on assembling evidence, generating inferences and explanations from evidence, comparing / assessing those inferences and explanations, and reporting results
• Geovisual analytics focuses on visual interfaces to analytical methods that support reasoning with and about geo-information – to enable insights about something for which place matters

Geographic info is important!

Analytics is attracting attention too!

Author: Alan M. MacEachren – Please do not post online
Geovisual analytics: scaling analysts to cope with information deluge

Big Data

Climate simulation

Economy & real estate

Spatio-Temporal Big Data Analysis

Example of visual analytics: analysis of city traffic

E.g. Milan can movement data:
2,073,216 location records
at 12.244 cars during 1 week.

This is far too much for processing in RAM!

Analysis:
1) aggregate the data in the database using the standard IBM® operations
2) load the aggregates in RAM and visualize for analyzing analysis
Spatial aggregation: by compartments (cells) of a territory division
Temporal aggregation: by time intervals: review time, cyclic time

Information Processing Challenges for Sensemaking (reasoning) & Decision-making

from Zack, 2007 as adapted by Muhren & van de Walle, 2010


Geovisual analytics: supporting analytical reasoning / sensemaking

Information, Concepts

Category, relation, theory

Hypothesis, Concept

Pfad & Contco 2005

Abduction

Induction

Deduction

Model-based

Rhetoric

External data

Evidence

Hypotheses

Evidence file

Schema

Presentation

Abduction

Induction

Deduction

Model-based

Rhetoric

Foraging loop

Sense-making loop

Data

Hypothesis, Concept

Category, relation, theory

Hypothesis, Concept

Pirolli & Card, 2005

after Pirolli & Card (2005)

after Gahegan (2005)
Case study 1: Geovisual Analytics to leverage heterogeneous, messy data

- SensePlace – leveraging news to support infectious disease modeling and analysis

![Image of SensePlace application]

- High Birth Rate
- Mainly subsistence economy
- Mainly nomadic
- Desert
- Mainly low vaccine coverage

Measles in Niger

Case study 2: SensePlace2 – Supporting Situational Awareness w/ Public Social Media

- Traditional map-based situational awareness methods for assessing a situation based on reports from the field

![Image of SensePlace2 application]

Geo-Historical Context Model

- Place
- Time
- Thing
- Individual collaboration

Geo-Social Context

- Tasks and Goals – determine the geographic-social context and information needs


Disease - Dengue

Geo-Social Interface

Visual Analytics and Interfaces

Analytic info is important!

Geographic info is important!

Analytics is attracting attention too!
But, social media is more important!

Turkey turns to Twitter after quake

Okan Bayülgen: “Twitter has worked well. It has been the most effective social web site in terms of organizing help. Several victims were found and rescued thanks to the posts of Twitter.”

Potential of social media for SA

Headling off disaster, one tweet at a time

SensePlace2: Computational process

- Collecting and processing Tweets:
  - crawl Tweets & load in database
  - analyze tweets for where they are from
    - “location” setting – often imprecise and inaccurate
    - time zones – accurate but imprecise
    - GPS coordinates – accurate & precise but uncommon
  - extract named-entities: e.g., locations, organizations, persons, hashtags, URLs, etc (what and where tweets are about) - uses customized version of GATE’s ANNIE named-entity extraction
  - store extracted entities: PostGRES DB w/ GIS extension
  - index tweets: Lucene index provides fast query and relevance ranks based on “boosting” (e.g., for tweets w/ places, hashtags, organizations, time, money, etc.)
Summary: Insights thus far

- **Situational Awareness:**
  - can be effectively conceptualized as a process to which the sensemaking model of visual analytics is applicable

- **Social media & crisis management:**
  - practitioners consider social media a potential SA source
  - social media have uses for both info push and info foraging
  - for info foraging, practitioners want info on: incidents, times, places, people, organizations

- **System design:**
  - combination of spatial relational database and Lucene indexing enables flexible query on 10s of millions of info fragments

### Beyond situational awareness: How are social media changing place and space?


Some specific GIScience challenges for Geovisual Analytics

- develop scientific understanding and related methods/tool that specifically:
  - support analytical reasoning with and about place-based information – this will require deeper understanding of what “reasoning” is in relation to crisis management tasks such as info (geo) foraging, sensemaking, situation assessment, decision-making, etc. and methods / tools for representing and utilizing context
  - support an extended process of work – this will require work domain analysis coupled with structures and tools to support processes rather than tasks (particularly ones dealing with uncertain, complex, ambiguous, and equivocal information).
  - recognize and capture geographic insight and its provenance – this will require attention to knowledge management
That’s all: Thanks
maceachren@psu.edu
www.geovista.psu.edu

Acknowledgements: Parts of the work represented here were supported by funds from: National Science Foundation: EIA-0306845 & 9983451; Gates Foundation funding of the Vaccine Modeling Initiative (https://www.vaccinemodeling.org/); National Visualization and Analytics Center, a U.S. Dept. of Homeland Security program operated by the Pacific Northwest National Laboratory (PNNL), and the Dept of Homeland Security Visual Analytics for Command, Control, and Interoperability Environments Center of Excellence. A large number of people contributed to work represented. These include Justina Blasford, Jin Chan, Ping Da, Mark Ghahgo, Bryson Grenfell, Diansheng Guo, Frank Hardisty, Anuj Jaiswal, Gene Lengerich, Prasenjit Mitra, Donna Peuquet, Scott Pezanowski, ChiChun Pan, Anthony Robinson, Robert Roth, Alexander Savelyev, Michael Stryker, Brian Tomaszewski, Ian Turton, Chaoping, Ya, Ying Wang, Chris Weaver, Xiao Zhang.
Mobile Symbol Design and Event Reporting for Mobile Devices:
A Preview of a Mobile Application and Developer API Based on an Online Symbol Repository

Overview
1. Event reporting + current strategies
2. Preview of our mobile app
3. What makes our app possible
4. Symbol design for mobile devices

Event Reporting
- Interest of emergency management, crisis mapping, and law enforcement
- ...but also the general public
- Played a major role in...
  - Haiti earthquake, 2010
  - Arab Spring, 2010
  - Christchurch earthquake, 2011
  - Public demonstrations since (e.g., #OWS)

Christchurch Recovery Map, Built w/ Ushahidi
The Good: Event maps are popular + utilized
The Bad: Event maps are largely desktop-oriented.

The Ugly: Event maps typically use the same abstract symbology.

The Uglier: Event mapping platforms center on a single geographic area and/or event type.


Streamlined Event Symbolization.

Map Alert
Streamlined Event Reporting

How MapAlert is Possible

What Makes MapAlert Possible:
Symbol Store: An online repository of 2,400+ map symbols

Symbol Store Web Services
- Developers can create their own apps to access the Symbol Store
- Main feature: Search
  - Symbol name
  - Tags
  - Category
  - Agency
  - Org: Set
Extended Features

- Edit
- Upload
- Parse and generate style files, SVG, PNG

How Symbol Store Web Services Work

Your App ➔ Symbol Store

Symbol Store ➔ Your Parser

XML

Search Example using HTML GET

Request
GET /OurServiceURL/OurService.asmx/SearchLucene?value=str&timestamp=string HTTP/1.1
Host: www.geovista.psu.edu

Response
HTTP/1.1 200 OK
Content-Type: text/xml; charset=utf-8
Content-Length: length

<?xml version="1.0" encoding="utf-8"?>
<ArrayOfAnyType xmlns="http://www.geovista.psu.edu">
  <anyType />
  <anyType />
</ArrayOfAnyType>

These services were made for you.

Mobile Symbol Design

- We wanted to learn if mobile symbols were unique (spoiler alert: they are)
- Reviewed close to 80 separate studies and manufacturer guidelines (Apple + Google)
- Boiled the findings down to some “simple” recommendations...brief overview here
- Designed mobile symbology to test against existing DHS HSWG set
Mobile Symbol Recommendation 1:
Use dark symbols on light backgrounds.

Mobile Symbol Recommendation 2:
Size matters. Bigger is (mostly) better.

Mobile Symbol Recommendation 3:
Symbols should be simple.

Mobile Symbol Recommendation 4:
Visually distinguish interactive and non-interactive symbols.

Our Mobile Symbol Design vs HSWG

Evaluation now Planned for November
- Will compare our symbols to DHS:HSWG on a tablet device
- Looking at:
  - accuracy
  - time
  - associative relationships
  - influence of base maps
  - interactive affordance
Thank You - questions & comments are welcome!

Joshua Stevens, UI+UX and Symbol Design  josh.stevens@psu.edu | @jscarto
Scott Pezanowski, Lead Developer  spezanowski@psu.edu | @spezanowski
Elaine Guidero, Symbol Design and Metadata  guidero@psu.edu | @elaineguidero
Anthony C. Robinson, Co-PI  acr181@psu.edu | @A_C_Robinson
Alan M. MacEachren, PI  maceachren@psu.edu | @alanGeoVISTA
Sharing Map Symbology for Emergency Management

Anthony C. Robinson, Ph.D

Co-Authors:
Scott Pezanowski
Raechel Bianchetti
Elaine Guidero
Justine Blanford
Alan M. MacEachren

Lead Faculty for Online Geospatial Education
John A. Dutton e-Education Institute
Assistant Director, GeoVISTA Center
Department of Geography
The Pennsylvania State University
Agenda

symbology @ DHS

symbol store design

current progress

future work
<table>
<thead>
<tr>
<th>Symbology</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial services office</td>
<td>Postal service infrastructure</td>
</tr>
<tr>
<td>Commercial infrastructure</td>
<td>Postal distribution center</td>
</tr>
<tr>
<td>Chemical plant</td>
<td>Print office</td>
</tr>
<tr>
<td>Steam manufacturers</td>
<td>Public service infrastructure</td>
</tr>
<tr>
<td>Steam rail lines</td>
<td>Emergency facility</td>
</tr>
<tr>
<td>Hazardous material production</td>
<td>Open facility</td>
</tr>
<tr>
<td>Hazardous material storage</td>
<td>Industrial facility</td>
</tr>
<tr>
<td>Industrial site</td>
<td>Religious institution</td>
</tr>
<tr>
<td>Landfill</td>
<td>Special needs infrastructure</td>
</tr>
<tr>
<td>Pharmaceutical manufacturer</td>
<td>Adult day care</td>
</tr>
<tr>
<td>Airport</td>
<td>Child day care</td>
</tr>
<tr>
<td>Basic release inventory</td>
<td>Senior care</td>
</tr>
<tr>
<td>Educational facilities</td>
<td>Telecom communications facility</td>
</tr>
<tr>
<td>College/university</td>
<td>Parking/transportation facility</td>
</tr>
<tr>
<td>@ DHS</td>
<td></td>
</tr>
</tbody>
</table>
background

- Diverse DHS organizations produce and use maps
  - Audiences range from geospatial analysts to general public
- No consistent set of map symbols used across DHS
  - Even if we just look at point symbols
- ANSI INCITS 415-2006 intended for emergency management mapping
  - Poorly adopted by practitioners
- Initial Objective: Develop process for symbol standardization
- Secondary Objective: Develop mechanism for symbol interoperability
ANSI Standard

• Point symbol set designed for emergency response
  – Goal was to facilitate common situational awareness, support point symbol interoperability

• Federal/state/local stakeholders took part in the process

• Symbols designed to work in black & white
  – Outline shapes used to distinguish between symbol types (incidents, natural events, operations, infrastructure)

• Evaluation conducted online with first responders
  – Made use of an “accept” or “reject” methodology, partially following the ANSI guidelines
<table>
<thead>
<tr>
<th>INCIDENTS</th>
<th>NATURAL EVENTS</th>
<th>OPERATIONS</th>
<th>INFRASTRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>civil disasters</td>
<td>floods</td>
<td>medical ops</td>
<td>buildings</td>
</tr>
<tr>
<td>civil demonstrations</td>
<td>drought</td>
<td>law enforcement</td>
<td>air traffic</td>
</tr>
<tr>
<td>civil disorders</td>
<td>earthquakes</td>
<td>fire</td>
<td>electricity</td>
</tr>
<tr>
<td>civil hazards</td>
<td>tornadoes</td>
<td>law enforcement</td>
<td>airports</td>
</tr>
<tr>
<td>civil incidents</td>
<td>hurricanes</td>
<td>law enforcement</td>
<td>roads</td>
</tr>
<tr>
<td>civil incidents</td>
<td>wildfires</td>
<td>law enforcement</td>
<td>railroads</td>
</tr>
</tbody>
</table>

*yellow highlight = symbols changed in latest release, green highlight = new symbols in last release*
### ANSI Standard

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td><em>Bird Infestation (Infestation)</em> - A harassing or troublesome invasion of birds. (Source: derived from the definition of “infestation” found in the FactMonster.com dictionary)</td>
</tr>
<tr>
<td>V</td>
<td><em>Insect Infestation (Infestation)</em> - A harassing or troublesome invasion of insects. (Source: derived from the definition of “infestation” found in the FactMonster.com dictionary)</td>
</tr>
<tr>
<td>W</td>
<td><em>Microbial Infestation (Infestation)</em> - A harassing or troublesome invasion of a microbe. (Source: derived from the definition of “infestation” found in the FactMonster.com dictionary)</td>
</tr>
<tr>
<td>X</td>
<td><em>Reptile Infestation (Infestation)</em> - A harassing or troublesome invasion of reptiles. (Source: derived from the definition of “infestation” found in the FactMonster.com dictionary)</td>
</tr>
<tr>
<td>Y</td>
<td><em>Rodent Infestation (Infestation)</em> - A harassing or troublesome invasion of rodents. (Source: derived from the definition of “infestation” found in the FactMonster.com dictionary)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A</td>
<td>Civil Disturbance Incident (Theme) - Human activities resulting in the disrupting of services or requiring varying levels of support, law enforcement or attention.</td>
</tr>
<tr>
<td>B</td>
<td>Civil Demonstrations (Civil Disturbance) - A public display of group feelings toward a person or cause. (Source: Merriam-Webster Online Dictionary definition)</td>
</tr>
<tr>
<td>C</td>
<td>Civil Displaced Population (Civil Disturbance) - Persons or groups of person who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, violations of human rights, or natural or human-made disasters. (Source: United Nations Guiding Principles on Internally Displaced Persons, 1998)</td>
</tr>
<tr>
<td>D</td>
<td>Civil Rioting (Civil Disturbance) - A public disturbance involving (1) an act or acts of violence by one or more persons part of an assemblage of three or more persons, which act or acts shall constitute a clear and present danger of, or shall result in, damage</td>
</tr>
</tbody>
</table>
Understanding the Problem

• We discovered that most mission areas had their own, ‘in house’ standards

• These were developed on an ad hoc basis, usually by one cartographer

• Collections of Esri markers and whatever else they could scrape together

• Such ‘standards’ are passed around to new employees and are promoted as default option

New ‘Standardization’ Process

• Distributed, web-based activities through a customized Drupal site

• Phase 1: Needs Assessment
  – Review current symbology, identify new symbol needs, problems with current symbols

• Phase 2: Initial Standard Development
  – Develop symbol categories, vote on changes to current symbology

• Phase 3: Standard Refinement
  – Discuss, refine & vote on final categories

• Phase 4: Implementation & Quality Control
  – Test new symbology in exercise, submit standard for graphical refinement by cartographers

• Methods feature
  – Round-based discussion & voting (modified Delphi)
  – Card-sorting activities (using websort.com)
  – Anonymized participation
<table>
<thead>
<tr>
<th>Symbol Store</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong>: A bomb threat or other threat involving explosives.</td>
<td></td>
</tr>
<tr>
<td><strong>Keywords</strong>: bomb, threat, explosive, warning, hazard</td>
<td></td>
</tr>
<tr>
<td><strong>Categories</strong>: incident, event, people, attack, terrorism</td>
<td></td>
</tr>
<tr>
<td><strong>Users</strong>: FBI, NGA, USCG, DOD</td>
<td></td>
</tr>
<tr>
<td><strong>Set</strong>: ANSI INCITS 415, MILSPEC 2525B</td>
<td></td>
</tr>
<tr>
<td><strong>Creator</strong>: <a href="mailto:r.roth@psu.edu">r.roth@psu.edu</a></td>
<td></td>
</tr>
<tr>
<td><strong>Uploaded</strong>: May 21, 2010 at 10:32am</td>
<td></td>
</tr>
<tr>
<td><strong>Add to Cart</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol Store</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong>: A significant fire event to a natural or man-made entity</td>
<td></td>
</tr>
<tr>
<td><strong>Keywords</strong>: fire, flame, firefighting, wildfire, forest fire, arson</td>
<td></td>
</tr>
<tr>
<td><strong>Categories</strong>: incident, event, natural hazard, arson</td>
<td></td>
</tr>
<tr>
<td><strong>Users</strong>: CBP, FBI, NGA, USCG</td>
<td></td>
</tr>
<tr>
<td><strong>Set</strong>: ANSI INCITS 415, MILSPEC 2525B</td>
<td></td>
</tr>
<tr>
<td><strong>Creator</strong>: <a href="mailto:a.robinson@psu.edu">a.robinson@psu.edu</a></td>
<td></td>
</tr>
<tr>
<td><strong>Uploaded</strong>: July 2, 2010 at 4:35pm</td>
<td></td>
</tr>
<tr>
<td><strong>Add to Cart</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol Store</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong>: A significant release of toxic chemicals in gas, liquid, or solid form.</td>
<td></td>
</tr>
<tr>
<td><strong>Keywords</strong>: spill, release, toxic, chemical, hazmat, cloud, gas, liquid, solid</td>
<td></td>
</tr>
<tr>
<td><strong>Categories</strong>: incident, event, people, hazard, hazmat</td>
<td></td>
</tr>
<tr>
<td><strong>Users</strong>: DOD, FBI, NGA</td>
<td></td>
</tr>
</tbody>
</table>
Sharing Symbols

• What happens once a group has completed the process?

• How can cartographers in different mission areas check out these ‘in house’ standards?

• What can we do to better understand what these various standards have in common?

• How can we encourage cartographers to share their symbols more widely?
Symbol Store Approach

• Develop web-based solution for sharing / browsing point symbols

• Simple design – aim to support one key task really well

• Symbols that have gone through our process can receive special tags

• Iterative development process -> static mockup, dynamic prototype, refined prototype, final version
Design Goals

Four simple design objectives for Symbol Store:

1. **Search for and retrieve symbols**
   - by keyword

2. **Preview symbols on realistic maps**
   - like ColorBrewer, of course

3. **Browse for symbols**
   - discover new symbols by category, agency, etc...

4. **Share symbols**
   - Contribute your own symbols and add appropriate metadata
Initial Design
Initial Design
Back-end Design

• ArcObjects is used to parse Esri style files and generate preview images of the symbols

• Lucene index for text content
  – Quick text searching and retrieval of large amounts of content
  – Ranks hits based upon user searches
  – Weights search criteria to improve search results

• .NET web service supports read/write to the Lucene index and read/write for .style files

• All content stored in the Lucene index or files on the server (style files and image preview files)
User Interface

• Interface built using Flash Catalyst (converts AI & other art into UI objects)

• Functionality connected to interface using Flash Builder 4 (formerly Flex)

• Flash Builder used to get results from the web service

• Flash plugin required for web browsers, but otherwise works across platforms
current progress
Progress

• Developed interactive prototype based on static mockup + sponsor feedback

• Prototype supports all four of our initial task goals:
  – Search
  – Preview
  – Browse
  – Share

• Public version now available at symbolstore.org
  – Uploading not enabled for public version yet 😞
### Main UI

**Symbol Store**

![Symbol Store Interface](image)

<table>
<thead>
<tr>
<th>Symbol Name</th>
<th>Description</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>DART Station</td>
<td>DART Station</td>
<td>Homeland Security Operations, Emergency Management</td>
</tr>
<tr>
<td>Holocene Volcano</td>
<td>Holocene Volcano</td>
<td>Emergency Management, Natural Hazard</td>
</tr>
<tr>
<td>Significant Event</td>
<td>Significant Event</td>
<td>Emergency Management, Natural Hazard</td>
</tr>
<tr>
<td>Tsunami Event</td>
<td>Tsunami Event</td>
<td>Emergency Management, Natural Hazard</td>
</tr>
</tbody>
</table>

**Browse and Upload**

- Search: 
- Sort by: most recent
- Browse by:

**Cart**

- 3 items in cart
- Remove

**Options**

- Empty Cart
- Use Symbols
- Preview on Map

**Steps**

1. Prepare Symbols
2. Download Symbols
Map Preview
Uploading & Metadata

Your file has been uploaded!

- **Filename:** StyleFile_LOGs_11022010.style
- **Date uploaded:** September 1, 2011
- **Contributing agency:**
- **Categories:**
- **Username:**
- **Set:**

**Edit your symbols here**

- **Add description...**
- **Add keywords...**
- **Add categories...**
- **Add users...**
- **Add rating...**

**Symbols available:**
- Airplane
- Cross
- Flag
- Building
- Bell
- Dolphin
- Tractor
- Golf
- House
- Letter F
- Letter G

**Fields for:**
- Description
- Keywords
- Categories
- Users
Initial Evaluation

• 6 Cartographers from California’s Dept. of Water Resources
  – Primary mission involves flood mapping

• Basic task analysis + follow-up focus group using early Symbol Store prototype

• Tasks focused on our four key design goals, overall satisfaction with the prototype was high for each feature

• Lots of suggested improvements, including:
  – Grid view to show more symbols at once
  – Use style file details to pre-populate metadata
  – Show USGS Topo design as option in Map Preview
future work
Next Steps

• Integrate portions of our symbol standardization process into the Symbol Store

• Develop support for other technical standards (SLD, for example)

• Extend tools to handle dynamic / multi-scale point symbols

• Explore Symbol Store usage patterns to identify frequently used symbols, cross-organizational commonalities, etc...

• Vector export (instead of just Esri .style files)
for more information:
www.geovista.psu.edu/symbology/

try it - www.symbolstore.org

email - arobinson@psu.edu

This work is supported by a contract from the U.S. Department of Homeland Security Science and Technology Directorate, Command, Control and Interoperability Division. The views and opinions expressed here are of the authors, and do not reflect the official positions of the Department of Homeland Security or the Federal Government.
SymbolStore.org: An Open Resource for Map Symbols

Anthony C. Robinson, PhD
Lead Faculty for Online Geospatial Education
John A. Dutton e-Education Institute
Assistant Director, GeoVISTA Center
Department of Geography
The Pennsylvania State University

Project Team:
Scott Pezanowski
Joshua Stevens
Justine Blanford
Alan M. MacEachren

Raechel Bianchetti
Elaine Guidero
Ryan Mullins
Eun-Kyeong Kim
Agenda

Sharing symbols

SymbolStore features

SymbolStore.org progress & demo

Next steps
Sharing symbols
Background

• Our work began with DHS to explore how they produce and use maps
  – Focus on symbol interoperability

• No consistent set of map symbols used across DHS
  – No mechanisms to share them, either

• Standards for various sub-domains exist, but are weakly adopted
  – Symbol discovery happens in ArcGIS and through simple web sharing

• Our goal: Develop mechanism to support symbol interoperability
Understanding the Problem

• Working with DHS, we discovered that most mission areas had their own ‘in house’ standards

• These were developed on an ad hoc basis, usually by one cartographer

• Collections of Esri markers and whatever else they could scrape together

• Such ‘standards’ are passed around to new employees and are promoted as default option

New ‘Standardization’ Process

- Distributed, web-based activities through a customized Drupal site

- Phase 1: Needs Assessment
  - Review current symbology, identify new symbol needs, problems with current symbols

- Phase 2: Initial Standard Development
  - Develop symbol categories, vote on changes to current symbology

- Phase 3: Standard Refinement
  - Discuss, refine & vote on final categories

- Phase 4: Implementation & Quality Control
  - Test new symbology in exercise, submit standard for graphical refinement by cartographers

- Methods feature
  - Round-based discussion & voting (modified Delphi)
  - Card-sorting activities (using websort.com)
  - Anonymized participation
SymbolStore features
Key design goals

• What happens once a group has completed the standardization process?

• How can other cartographers check out others standards?

• What can we do to better understand what these various standards have in common?

• How can we encourage cartographers to share their symbols more widely?
Symbol Store Approach

• Develop web-based solution for sharing / browsing point symbols

• Simple design – aim to support one key task really well

• Symbols that have gone through our process can receive special tags

• Initially just for DHS – today we’re releasing the first public version
Design Goals

Four simple design objectives for Symbol Store:

1. **Search for and retrieve symbols**
   - by keyword

2. **Preview symbols on realistic maps**
   - like ColorBrewer and other map design apps

3. **Browse for symbols**
   - discover new symbols by category, agency, etc...

4. **Share symbols**
   - Contribute your own symbols and add appropriate metadata
Initial Design
User Interface

• Interface built using Flash Catalyst

• Functionality connected to interface using Flash Builder 4 (formerly Flex)

• Flash Builder used to get results from our web service

• Yes, we plan on migrating to a Javascript-based interface in the near future
Back-end Design

• ArcObjects is used to parse Esri .style files and generate preview images of the symbols

• Lucene index for text content
  – Quick text searching and retrieval of large amounts of content
  – Ranks hits based upon user searches
  – Weights search criteria to improve search results

• .NET web service supports read/write to the Lucene index and read/write for .style files

• All content stored in the Lucene index or files on the server (style files and image files)
SymbolStore.org progress
Progress

• Developed interactive prototype based on static mockup + sponsor feedback (only for DHS)

• Initial evaluation with flood mapping users in California (Troedson, NACIS 2011)

• Current prototype supports all four of our initial task goals:
  – Search
  – Preview
  – Browse
  – Share

• UI Reskinned by Josh Stevens

• Public version now available at symbolstore.org
Search

• Lucene enables more flexible search
  – “Park” will also retrieve “Parking”
  – “Firing” will also retrieve “Fire”

• Next step involves using WordNet to expand queries automatically
Map Preview
Symbol Download

Symbols in the cart are exported in a single .zip containing:

- Esri.style file
- PNG images
- Vector graphics (SVG)
Contribute Symbols

- Basic uploading features are enabled
- Automated processing of Style files is difficult given issues with fonts
What’s available

• 2480 point symbols

• Sets from
  – DHS (FEMA, HSIP, IICD…)
  – National Park Service
  – MapBox Maki
  – PSU Mobile Designs

• Support for downloading complete symbol sets coming soon
Next steps
Integration plan

• e-Symbology Portal → Symbol Store

• Support small groups of users who:
  – want to pull in symbols from the Symbol Store
  – refine them
  – contribute those changes back to the Symbol Store

• Also support individual power users who:
  – want to serve as the official curator of a symbol set
  – want to contribute metadata for large numbers of symbols

• Workshops with mapmakers to test and refine the integrated tools
Metadata contribution

• Interface and back-end procedures exist to support flexible metadata creation

• Needs to be reskinned and incorporated into the overall process

• One or more symbols can be easily selected and metadata changes applied to all of them.
Next Steps

• Continue progress on integration between e-Symbology Portal and the Symbol Store

• Develop automated support for other input formats (SVG)

• Collect and post more symbols

• Extend tools to handle dynamic / multi-scale point symbols

• Explore Symbol Store usage/download patterns

• Figure out what to do with duplication issues
  — same symbol, different meaning, for example
for more information:

www.geovista.psu.edu/symbology/

try it now - www.symbolstore.org

email - arobinson@psu.edu

This work is supported by a contract from the U.S. Department of Homeland Security Science and Technology Directorate, Command, Control and Interoperability Division. The views and opinions expressed here are of the authors, and do not reflect the official positions of the Department of Homeland Security or the Federal Government.
The New Cartography

current states of science and technology in Interactive Cartography & Geovisualization

Robert E. Roth
reroth@wisc.edu
The New Cartographers

By Emily Underwood
March 18, 2013

Twenty years ago, a driver lost at night would pull his car over, take out a paper map bought at a gas station, and pore over its folds under a dim light. With luck and some critical thinking, he would eventually get where he was going. Today, he’d be more likely to swipe his finger across a smart phone screen and follow directions using Google Maps.

As maps have changed, so have mapmakers. No longer static images, maps have become active interfaces for information exchange, continuously determining where we are in relation to distant satellites and suggesting where we ought to go, says Seth Spielman, a 38-year-old geography professor at the University of Colorado, Boulder. Today, the global geoservices industry collects, shares, and analyzes data on an unprecedented scale. It’s valued at as much as $270 billion per year and employs 500,000 people in the United States, according to a recent report from Google.

"Future shortages in cartography, photogrammetry, and geodesy seem likely because the number of graduates is too small (tens to hundreds) to give NGA choices or means of meeting sudden demand." — Future U.S. Workforce for Geospatial Intelligence, a report from the National Research Council
1. Applications

2. Science

3. Technology

Interactive mapping

Cartographic interaction

Brushing

Emerging mapping tech
Applications

interactive cartography &
geovisualization
Interactive Cartography
Space-Time Mapping
Exploratory Geovisualization
# Needs Assessment Interviews

**geospatial technology unmet needs & barriers to use**

## Science of Cartographic Interaction (USGIF; Doctoral)

- **Focus:** Interactive Cartography & Geovisualization
- **Participants:** Sampling of application domains, including Emergency Response/Crisis Management (n=4) and Intelligence Analysis (n=3)

## Symbology Development Best Practices (DHS)

- **Focus:** ANSI INCITS 415-2006 symbol standard
- **Participants:** CBP, IICD, USCG, NOC, FEMA, USFS, DNDO (n=14)

## VACCINE (DHS Center of Excellence)

- **Focus:** Spatial Criminology & Crime Mapping
- **Participants:** municipal law enforcement agencies, NIJ (n=9)

---

[http://www.geovista.psu.edu](http://www.geovista.psu.edu)
CrimeViz
spatial criminology

Roth et al. (2013a)
IED visualization
intelligence analysis

NCTC Worldwide Incident Tracking System

Murdock et al. (2012)
Science
towards a science of cartographic interaction
“Recommendation 3.3: Create a new science of interaction to support visual analytics. The grand challenge of interaction is to develop a taxonomy to describe the design space of interaction techniques that supports the science of analytic reasoning. We must characterize this design space and identify under-explored areas that are relevant to visual analytics. Then, R&D should be focused on expanding the repertoire of interaction techniques that can fill those gaps in the design space.”

Thomas et al. (2005)
Visual Variables
how we represent

location  size  color hue  color value  color saturation

orientation  grain  arrangement  shape  fuzziness  transparency

Bertin (1967)
Interaction Primitives
how we interact

Roth (2012)

brushing

focusing

linking

viewer motion

object rotation

delete

label

add

join

create

highlight

same data, changing representation

dynamic re-expression

dynamic comparison

assignment

colormap manipulation

viewpoint manipulation

accessing extra/exact information

overview

manipulate objects

encode data

set-graphical-value

Dykes (1997)

Shepherd (1995)

Becker & Cleveland (1987)

Chuah & Roth (1996)

same representation, changing parameters

same representation, changing parameters

Dix & Ellis (1998)

Ezdall et al. (2008)

Edsall et al. (2008)

Masters & Edsall (2000)

MacEachren et al. (1999)

Keim (2002)


Edsall et al. (2008)

conditional visibility

altering representation type

altering symbolization

panning/re-centering

posing queries

filter

details-on-demand

extract

history

relate

navigation

selection

arranging views

Biirn et al. (1996)

Birn et al. (1996)
ASSAULT

case number: HSG20130501991

date: 2013-05-12

district: 04

description: Resist arrest / other law enforcement
time: 05:30

date: 2013-05-12

Click here for Street View

63 matching records

Temporal Parameters: Linear timeline

Binning Unit: Day

reset temporal parameters

map showing 04 cases for Sunday, May 12, 2013
analysts use Interfaces
analysts use Interfaces

analysts experience Interactions
cartographic interaction primitives for User Experience (UX) Design

**objectives**
(user tasks)
- identify
- compare
- rank
- associate
- delineate

**operators**
(functionality)
- reexpress
- arrange
- sequence
- resymbolize
- overlay
- reproject
- pan
- zoom
- filter
- search
- retrieve
- calculate

**operands**
(geospatial data)
- space
- time
- attribute
The Grand(er) Challenge
a syntactics of cartographic interaction primitives

ASSOCIATE BY ATTRIBUTES-IN-SPACE

<table>
<thead>
<tr>
<th>Time</th>
<th>Participant J</th>
<th>Participant B</th>
<th>Participant H</th>
<th>Participant C</th>
<th>Participant E</th>
<th>Participant I</th>
<th>Participant F</th>
<th>Participant G</th>
<th>Participant A</th>
<th>Participant D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0:30</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>X</td>
<td>R</td>
<td>R</td>
<td>F</td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>X</td>
<td>R</td>
<td>R</td>
<td>F</td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

interaction logging + social science methods

Roth (forthcoming)
3 Technology

emerging technology for browser-based mapping
Flash Player is not available for this device
<table>
<thead>
<tr>
<th>REPRESENTATION</th>
<th>INTERACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing Maps API</td>
<td></td>
</tr>
<tr>
<td>CartoDB</td>
<td></td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>Tiledrawer</td>
<td></td>
</tr>
<tr>
<td>Tiledmill</td>
<td></td>
</tr>
<tr>
<td>Tilesiache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
<tr>
<td>Map vs. Imagery</td>
<td></td>
</tr>
<tr>
<td>Basemap Styling</td>
<td></td>
</tr>
<tr>
<td>Tile Rendering</td>
<td></td>
</tr>
<tr>
<td>Vector Overlays</td>
<td></td>
</tr>
<tr>
<td>Choropleth</td>
<td></td>
</tr>
<tr>
<td>Dot Density</td>
<td></td>
</tr>
<tr>
<td>Isoline/Surface</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>Cartogram</td>
<td></td>
</tr>
<tr>
<td>Bivariate/Multivariate</td>
<td></td>
</tr>
<tr>
<td>Animation</td>
<td></td>
</tr>
<tr>
<td>Graphics/Charts</td>
<td></td>
</tr>
<tr>
<td>Arrange/Linked Views</td>
<td></td>
</tr>
<tr>
<td>Reexpress</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>Resymbolize</td>
<td></td>
</tr>
<tr>
<td>Overlay/Toggle</td>
<td></td>
</tr>
<tr>
<td>Reproject</td>
<td></td>
</tr>
<tr>
<td>Pan</td>
<td></td>
</tr>
<tr>
<td>Zoom</td>
<td></td>
</tr>
<tr>
<td>Filter</td>
<td></td>
</tr>
<tr>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>Retrieve</td>
<td></td>
</tr>
<tr>
<td>Calculate</td>
<td></td>
</tr>
<tr>
<td>Mobile Support</td>
<td></td>
</tr>
<tr>
<td>Location Aware</td>
<td></td>
</tr>
<tr>
<td>REPRESENTATION</td>
<td>INTERACTION</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Bing Maps API</td>
<td></td>
</tr>
<tr>
<td>CartoDB</td>
<td></td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>Tiledrawer</td>
<td></td>
</tr>
<tr>
<td>Tilemill</td>
<td></td>
</tr>
<tr>
<td>Tilestache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
<tr>
<td>Map vs. Imagery</td>
<td></td>
</tr>
<tr>
<td>Basemap Styling</td>
<td></td>
</tr>
<tr>
<td>Tile Rendering</td>
<td></td>
</tr>
<tr>
<td>Vector Overlays</td>
<td></td>
</tr>
<tr>
<td>Proportional Symbol</td>
<td></td>
</tr>
<tr>
<td>Dot Density</td>
<td></td>
</tr>
<tr>
<td>Isoline/Surface</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>Bivariate/Multivariate</td>
<td></td>
</tr>
<tr>
<td>Animation</td>
<td></td>
</tr>
<tr>
<td>Graphics/Charts</td>
<td></td>
</tr>
<tr>
<td>Arrange/Linked Views</td>
<td></td>
</tr>
<tr>
<td>Reexpress</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>Resymbolize</td>
<td></td>
</tr>
<tr>
<td>Overlay/Toggle</td>
<td></td>
</tr>
<tr>
<td>Reproject</td>
<td></td>
</tr>
<tr>
<td>Pan</td>
<td></td>
</tr>
<tr>
<td>Zoom</td>
<td></td>
</tr>
<tr>
<td>Filter</td>
<td></td>
</tr>
<tr>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>Retrieve</td>
<td></td>
</tr>
<tr>
<td>Calculate</td>
<td></td>
</tr>
<tr>
<td>Mobile Support</td>
<td></td>
</tr>
<tr>
<td>Location Aware</td>
<td></td>
</tr>
</tbody>
</table>

Coding Completed 2012

Roth et al. (2013b)
<table>
<thead>
<tr>
<th>REPRESENTATION</th>
<th>INTERACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing Maps API</td>
<td></td>
</tr>
<tr>
<td>CartoDB</td>
<td></td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>Tiledrawer</td>
<td></td>
</tr>
<tr>
<td>Tilemill</td>
<td></td>
</tr>
<tr>
<td>Tilestache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
<tr>
<td>Map vs. Imagery</td>
<td></td>
</tr>
<tr>
<td>Basemap Styling</td>
<td></td>
</tr>
<tr>
<td>Tile Rendering</td>
<td></td>
</tr>
<tr>
<td>Vector Overlays</td>
<td></td>
</tr>
<tr>
<td>Proportional Symbol</td>
<td></td>
</tr>
<tr>
<td>Dot, Density</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>Choropleth</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>Isoline/Surface</td>
<td></td>
</tr>
<tr>
<td>Bivariate/Multivariate</td>
<td></td>
</tr>
<tr>
<td>Animation</td>
<td></td>
</tr>
<tr>
<td>Graphics/Charts</td>
<td></td>
</tr>
<tr>
<td>Arrange/Linked Views</td>
<td></td>
</tr>
<tr>
<td>Reexpress</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>Resymbolize</td>
<td></td>
</tr>
<tr>
<td>Overlay/Toggle</td>
<td></td>
</tr>
<tr>
<td>Reproject</td>
<td></td>
</tr>
<tr>
<td>Pan</td>
<td></td>
</tr>
<tr>
<td>Zoom</td>
<td></td>
</tr>
<tr>
<td>Filter</td>
<td></td>
</tr>
<tr>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>Retrieve</td>
<td></td>
</tr>
<tr>
<td>Calculate</td>
<td></td>
</tr>
<tr>
<td>Mobile Support</td>
<td></td>
</tr>
<tr>
<td>Location Aware</td>
<td></td>
</tr>
</tbody>
</table>

Coding Completed 2012

Roth et al. (2013b)
<table>
<thead>
<tr>
<th>REPRESENTATION</th>
<th>INTERACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing Maps API</td>
<td>Pan, Zoom, Search, Retrieve, Calculate, Mobile Support, Location Aware</td>
</tr>
<tr>
<td>CartoDB</td>
<td>Filter</td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>TileDrawer</td>
<td></td>
</tr>
<tr>
<td>Tilemill</td>
<td></td>
</tr>
<tr>
<td>Tilestache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
<tr>
<td>Map vs. Imagery</td>
<td></td>
</tr>
<tr>
<td>Basemap Styling</td>
<td></td>
</tr>
<tr>
<td>Tile Rendering</td>
<td></td>
</tr>
<tr>
<td>Vector Overlays</td>
<td></td>
</tr>
<tr>
<td>Choropleth</td>
<td></td>
</tr>
<tr>
<td>Dot Density</td>
<td></td>
</tr>
<tr>
<td>Isoline / Surface</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>Bivariate / Multivariate</td>
<td></td>
</tr>
<tr>
<td>Cartesian / Multidimensional</td>
<td></td>
</tr>
<tr>
<td>Animation</td>
<td></td>
</tr>
<tr>
<td>Graphics / Charts</td>
<td></td>
</tr>
<tr>
<td>Arrange / Linked Views</td>
<td></td>
</tr>
<tr>
<td>Reproject</td>
<td></td>
</tr>
</tbody>
</table>

Coding Completed 2012

Roth et al. (2013b)
### Make beautiful interactive maps

Whether you're a journalist, web designer, researcher, or seasoned cartographer, TileMill is the design studio you need to create stunning interactive maps.

OpenLayers: Free Maps for the Web

Get OpenLayers Now!
- 2.12 (Stable): tar.gz | zip
- 2.12 Release Notes
- API Documentation, User documentation
- See examples of OpenLayers Usage:
  - Release Examples (2.12), Development Examples
  - Fork us on GitHub

About OpenLayers
OpenLayers makes it easy to put a dynamic map in any web page. It can display map tiles and markers loaded from any source. OpenLayers has been developed to further the use of geographic information of all kinds. OpenLayers is completely free, Open Source JavaScript, released under the 2-clause BSD License (also known as the FreeBSD).

Toward OpenLayers 3!
We've begun the development effort to make the next major version of

http://openlayers.org
### Data-Driven Documents

D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

Download the latest version here:

http://d3js.org

---

Roth et al. (2013b)
Leaflet is a modern open-source JavaScript library for mobile-friendly interactive maps. It is developed by Vladimir Agapitos with a team of dedicated contributors. Weighing just about 33 KB of JS code, it has all the features most developers ever need for online maps.

Leaflet is designed with simplicity, performance, and usability in mind. It works efficiently across all major desktop and mobile platforms out of the box, taking advantage of HTML5 and CSS3 on modern browsers while still being accessible on older ones. It can be extended with many plugins, has a beautiful, easy to use and well-documented API, and a simple, readable source code that is a joy to contribute to.

Used by: Flickr foursquare craigslist IGN Wikipedia OSM Meetup WSJ Geocaching StreetEasy CloudMade CartoDB GIS Cloud ...
<table>
<thead>
<tr>
<th>REPRESENTATION</th>
<th>INTERACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing Maps API</td>
<td>HTML5</td>
</tr>
<tr>
<td>CartoDB</td>
<td>Mapgeeks</td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>Tiledrawer</td>
<td></td>
</tr>
<tr>
<td>Tilemill</td>
<td></td>
</tr>
<tr>
<td>Tilestache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
<tr>
<td>REPRESENTATION</td>
<td>INTERACTION</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Bing Maps API</td>
<td></td>
</tr>
<tr>
<td>CartoDB</td>
<td></td>
</tr>
<tr>
<td>CartoWeb</td>
<td></td>
</tr>
<tr>
<td>Cloudmade Editor</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td></td>
</tr>
<tr>
<td>deCarta</td>
<td></td>
</tr>
<tr>
<td>GeoEXT</td>
<td></td>
</tr>
<tr>
<td>GeoMoose</td>
<td></td>
</tr>
<tr>
<td>Google Maps API</td>
<td></td>
</tr>
<tr>
<td>Jump</td>
<td></td>
</tr>
<tr>
<td>Ka-Map</td>
<td></td>
</tr>
<tr>
<td>Kartograph</td>
<td></td>
</tr>
<tr>
<td>Leaflet</td>
<td></td>
</tr>
<tr>
<td>MapBender</td>
<td></td>
</tr>
<tr>
<td>Mapnik</td>
<td></td>
</tr>
<tr>
<td>MapQuery</td>
<td></td>
</tr>
<tr>
<td>MapQuest API</td>
<td></td>
</tr>
<tr>
<td>MapServer</td>
<td></td>
</tr>
<tr>
<td>Mapstraction</td>
<td></td>
</tr>
<tr>
<td>Modest Maps</td>
<td></td>
</tr>
<tr>
<td>Nokia</td>
<td></td>
</tr>
<tr>
<td>OpenLayers</td>
<td></td>
</tr>
<tr>
<td>OpenScales</td>
<td></td>
</tr>
<tr>
<td>Polymaps</td>
<td></td>
</tr>
<tr>
<td>Processing</td>
<td></td>
</tr>
<tr>
<td>Processing.js</td>
<td></td>
</tr>
<tr>
<td>Raphael</td>
<td></td>
</tr>
<tr>
<td>ReadyMap</td>
<td></td>
</tr>
<tr>
<td>Tiledrawer</td>
<td></td>
</tr>
<tr>
<td>Tilemill</td>
<td></td>
</tr>
<tr>
<td>Tiledcache</td>
<td></td>
</tr>
<tr>
<td>TimeMap</td>
<td></td>
</tr>
<tr>
<td>ViaMichelin</td>
<td></td>
</tr>
<tr>
<td>Wax</td>
<td></td>
</tr>
<tr>
<td>WorldKit</td>
<td></td>
</tr>
</tbody>
</table>

Cybersecurity
To Conclude

think differently about…
Natural Earth is a public domain map dataset available at 1:10m, 1:50m, and 1:110 million scales. Featuring tightly integrated vector and raster data, with Natural Earth you can make a variety of visually pleasing, well-crafted maps with cartography or GIS software.

Natural Earth was built through a collaboration of many volunteers and is supported by NACIS (North-American Cartographic Information Society), and is free for use in any type of project (see our Terms of Use page for more information).

Convenience
Natural Earth solves a problem: finding suitable data for making small-scale maps. In a time when the web is awash in geospatial data, cartographers are forced to waste time sifting through confusing tangles of poorly attributed data to make clean, legible maps. Because your time is

Neatness Counts
The carefully generalized linework maintains consistent, recognizable geographic shapes at 1:10m, 1:50m, and 1:110m scales. Natural Earth was built from the ground up so you will find that all data layers align precisely with one another. For example, where rivers and

GIS Attributes
Natural Earth, however, is more than just a collection of pretty lines. The data attributes are equally important for mapmaking. Most data contain embedded feature names, which are ranked by relative importance. Other attributes facilitate faster map production.

http://www.naturalearthdata.com
Natural Earth is a public domain map dataset available at 1:10m, 1:50m, and 1:100 tightly integrated vector and raster data, with Natural Earth you can make a variety of well-crafted maps with cartography or GIS software.

Natural Earth was built through a collaboration of many volunteers and is supported by NACIS (North-American Cartographic Information Society), and is free for use in any type of project (see our Terms of Use page for more information).

Convenience

Natural Earth solves a problem: finding suitable data for making small-scale maps. In a time when the web is awash in geospatial data, cartographers are forced to waste time sifting through confusing tangles of poorly attributed data to make clean, legible maps. Because your time is

Neatness Counts

The carefully generalized linework maintains consistent, recognizable geographic shapes at 1:10m, 1:50m, and 1:100 scales. Natural Earth was built from the ground up so you will find that all data layers align precisely with one another. For example, where rivers and

GIS Attributes

Natural Earth, however, is more than just a collection of pretty lines. The data attributes are equally important for mapmaking. Most data contain embedded feature names, which are ranked by relative importance. Other attributes facilitate faster map production.

http://www.naturalearthdata.com
think differently about your Mapping Products

Geovisual Analytics
The New Cartographers

By Emily Underwood

March 18, 2013

Twenty years ago, a driver lost at night would pull his car over, take out a paper map bought at a gas station, and pore over its folds under a dim light. With luck and some critical thinking, he would eventually get where he was going. Today, he'd be more likely to swipe his finger across a smartphone screen and follow directions using Google Maps.

As maps have changed, so have mapmakers. No longer static images, maps have become active interfaces for information exchange, continuously determining where we are in relation to distant satellites and suggesting where we ought to go, says Seth Spielman, a 38-year-old geography professor at the University of Colorado, Boulder. Today, the global geoservices industry collects, shares, and analyzes data on an unprecedented scale. It's valued at as much as $270 billion per year and employs 500,000 people in the United States, according to a recent report from Google. The rapid transformation, which Spielman equates with a "renaissance" in the field, has overturned traditional ideas of what a geographer does.

"Future shortages in cartography, photogrammetry, and geodesy seem likely because the number of graduates is too small (tens to hundreds) to give NGA choices or means of meeting sudden demand." —Future U.S. Workforce for Geospatial Intelligence, a report from the National Research Council
Further Reading:
(open access articles at http://www.geography.wisc.edu/faculty/roth)

Roth RE. (forthcoming) Interactive Maps: What we know and what we need to know. Journal of Spatial Information Science. [open access]


UX Design

Stages of Interaction

1. Forming the Goal
2. Forming the Intention
3. Specifying an Action
4. Executing the Action
5. Perceiving the State of the System
6. Interpreting the State of the System
7. Evaluating the Outcome

Primary Ability
Cognition
Perception
Motor Skills

Gulf of Execution
Visual Affordances

Gulf of Evaluation

Visual Feedback

Interacting-With

Object in the World

Seeing-That

map

operands

Object in the World

computing device

operators

objective

Reasoning-Why

The Design of Everyday Things

Donald A. Norman

Norman (1985); Roth (2011)
## Research Agenda

**towards a science of cartographic interaction**

<table>
<thead>
<tr>
<th><strong>How?</strong></th>
<th>interaction primitives • stages of interaction • interaction strategies • objectives, operators, operands • interface styles • interface design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What?</strong></td>
<td>interaction vs. interface • digital vs. analog • stages of interaction • system response time • interactive maps vs. mapping systems vs. map mashups</td>
</tr>
<tr>
<td><strong>Why?</strong></td>
<td>visual thinking • insight • exploration/confirmation/synthesis/presentation • discussion / debate • decision making</td>
</tr>
<tr>
<td><strong>When?</strong></td>
<td>interface complexity • interface freedom &amp; flexibility • convention &amp; standardization • work productivity • efficiency and effectiveness</td>
</tr>
<tr>
<td><strong>Who?</strong></td>
<td>user-centered design • user ability • user expertise • user motivation • adaptive cartography • role-based interfaces • geocollaboration</td>
</tr>
<tr>
<td><strong>Where?</strong></td>
<td>input capabilities • bandwidth size • processing power • display capabilities • interoperability • mobile mapping • open source development</td>
</tr>
</tbody>
</table>

Roth (forthcoming)
The SymbolStore: Expanding to Provide a Social Forum for the Creation, Sharing, and Evaluation of Symbols

Ryan S Mullins
Scott Pezanowski
Anthony C. Robinson, PhD
Alan M MacEachren, PhD

GeoVISTA Center
Department of Geography
The Pennsylvania State University
Outline

SymbolStore and Reviewer

Social Symbol Creation & Refinement

Features

Interface

Moving Forward
The SymbolStore

- Search and browse symbols
  - By keyword, organization, category, etc.

- Preview symbols on realistic maps
  - Similar to ColorBrewer and others

- Download symbols for use

- Share symbols with others

http://SymbolStore.org
The SymbolStore Reviewer

• Web–based activities via custom Drupal site

• Review Cycle:
  – Phase 1: Needs Assessment
  – Phase 2: Initial Standard Development
  – Phase 3: Standard Refinement
  – Phase 4: Implementation & Quality Control

• Methods
  – Round–based discussion
  – Voting (modified Delphi)

---

Lingering Questions

- How do we create a request for a new symbol, or modifications to an existing symbol?
- How do we track changes to symbols?
- Do we provide a means for discussion of designs, or will that create too many problems?
- How can we leverage community members to help the SymbolStore, and its content, grow?
Social Symbol
Creation and Refinement

Target Audience
• Proactive cartographers from participating organizations
• Estimated 5–10% of users

Needs
• Textual description of issues/needs
• Revision control
• Visual comparison across versions
• Discussion of
• Interfaces for:
  – Requesting changes
  – Contributing new versions
  – Examining and discussing
Features: Defining the Problem

• Sets:
  – Problem symbols
    • Redundant – Two symbols, same meaning
    • Confusing – Visual does not describe referent
    • Multiple meanings – One symbol, multiple meanings
    • Poorly designed – Displays poorly on digital/print media
  – Missing symbols

• Symbols:
  – Track problems with symbol individually
  – Similar to bug trackers for source code

• Universally:
  – Detailed description of set/symbol purpose
Features: Revision Control

• Revision control similar to those used for source code control

• Utilize popular, existing tools (e.g. Git)

• Allow comments for change logs

• Link to individual issues defined in request

• Review/merge functionality for symbol contributor
Features: Discussion

• Forum to record discussion about sets/symbols

• Threaded replies to topic
  – Toggle to show/hide
  – Vote up or down based on relevance
  – Filter by issue, symbol, etc.

• (Semi-)Automated moderation
Interface Concept: Requests
Interface Concept: Examining

Civil Rioting
Organization: HSWG
Set: ANSI
Contributor: HSWG

Categories: Civil Disturbance Feature
Tags: None
Uploaded: 20 Oct 2012

Comments
FakeUser1: This symbol is rather confusing, what exactly is it supposed to be conveying?
FakeUser2: I'm also unsure here.
FakeUser3: I am working on a fix for issue #2, is there any reasoning for the color of the symbol?
Contributor: Symbol color is typically used to denote event state, though any color scheme should work as long as it can be quickly and easily changed.

Issues
<table>
<thead>
<tr>
<th>#1</th>
<th>High</th>
<th>A problem description</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2</td>
<td>Normal</td>
<td>A problem description</td>
<td>Accepted</td>
</tr>
</tbody>
</table>
Moving Forward: Challenges

• Automated generation of symbols in supported formats

• Scalability of hardware/software

• Linking revision control and issue tracking
For more information:

Try it now – www.symbolstore.org
RyanMullins@psu.edu
@RyanMullins

This work is supported by a contract from the U.S. Department of Homeland Security Science and Technology Directorate, Command, Control and Interoperability Division. The views and opinions expressed here are of the authors, and do not reflect the official positions of the Department of Homeland Security or the Federal Government.
Understanding Spatial and Social Relationships in International Trade Network: A Geovisual Analytic Approach

Wei Luo
Peifeng Yin
Frank Hardisty
Alan M. MacEachren

GeoVISTA Center, Department of Geography
Penn State University
Motivation

The role of geographical proximity in shaping the International Trade Network (ITN) has not been explored, especially across different regions.

World trade among some countries in 1992 (Krempel and Pluemper 1999)
Research Questions

- What regions in the world are more influenced by the geographical proximity relationship?
- What regions in the world are more influenced by the international trade network relationship?
- What regions in the world on which both relationships have the approximately similar influence?
It is a challenge to distinguish the role of the social relationship and spatial relationship in the ITN, because both relationships always interact with each other.
Geographical Space

Social Space

Attribute Space

Modified from (Hess 2004)

http://www.geovista.psu.edu/GeoSocialApp/
Imports–exports relationship

<table>
<thead>
<tr>
<th>Years</th>
<th>importer1</th>
<th>importer2</th>
<th>flow1</th>
<th>flow2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>United States</td>
<td>Canada</td>
<td>291944</td>
<td>195151</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Bahamas</td>
<td>726.3</td>
<td>1945.79</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Cuba</td>
<td>0</td>
<td>397.87</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Haiti</td>
<td>458.5</td>
<td>756.91</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Dominican R</td>
<td>4721.4</td>
<td>5179.24</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Jamaica</td>
<td>410.9</td>
<td>1962.2</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Trinidad and</td>
<td>8342.2</td>
<td>1583.01</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Barbados</td>
<td>33.4</td>
<td>595.28</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Dominica</td>
<td>3.8</td>
<td>67.43</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Grenada</td>
<td>6</td>
<td>90.64</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>St. Lucia</td>
<td>33.6</td>
<td>148.06</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>St. Vincent a</td>
<td>15.7</td>
<td>79.98</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>Antigua &amp; B</td>
<td>4.6</td>
<td>189.9</td>
</tr>
<tr>
<td>2005</td>
<td>United States</td>
<td>St. Kitts and</td>
<td>55.8</td>
<td>138.82</td>
</tr>
</tbody>
</table>

Binary matrix for 0% Threshold in 2005

<table>
<thead>
<tr>
<th></th>
<th>AFGHANISTAN</th>
<th>ALBANIA</th>
<th>ALGERIA</th>
<th>AMERICA</th>
<th>ANDORRA</th>
<th>ANGOLA</th>
<th>ANTIGUA</th>
<th>ARGENTINA</th>
<th>ARS</th>
<th>ARMENIA</th>
<th>AUSTRALIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFGHANISTAN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ALBANIA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ALGERIA</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AMERICAN SAM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ANDORRA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ANGOLA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ANTIGUA AND E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ARGENTINA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARMENIA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AUSTRALIA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Results: The First Split (A Core–Periphery Structure)

The convergence of the iterated correlations (CONCOR) algorithm (Breiger et al. 1975).
The Second Split (Core Group)
The Third Split (Core Group)
The Second Split (Periphery Group)
The Third Split (Periphery Group)
The Third Split (Periphery Group)
Summary

- Global trade is hierarchical with a core–periphery structure in terms of international trade network.

- The complex interaction between the social relationship and spatial relationship always exists in the ITN, but the spatial relationship are more influential to developing countries, and the social relationship are more influential to developed countries.

- The interaction of spatial proximity and social closeness in terms of the ITN cross different regions also indicates that the development of countries experiences a process of loosened spatial constraints (Temporal implications).
Conclusion

- Spatial analysis and social network analysis can give insights on the same datasets from different perspectives, so it is necessary to consider simultaneously.

- New approaches should be developed to study the interaction between social relationships and spatial relationships.
Future Work

- Use the tool to explore the international trade data for multiple years.
Questions?
References


Visualizing Spatial, Temporal and Social Graph Information With the GeoViz Toolkit

Frank Hardisty
GeoVISTA Center, Penn State

Outline
- GeoViz Toolkit
- STempo
- GeoTxt.org
- GeoViz.js
- Cloud and Server GIS

GeoViz Toolkit

The Process of Inference

Starplot Overview
- Starplots
  - What and why of Starplots
  - My innovations

Why Starplots?
- Multivariate Exploratory Mapping (Geovisualization)
- Good for: a large number of attributes
  - Think of students who struggle to make a thematic map with two variables
- Bad for: a large number of observations
What are Starplots?

New thing: Starplot Maps

What is a parallel coordinate plot?

Slides by Linda Pickle at NCI
Parallel Coordinate Plot showing U.S. Counties

Demo
Need for Flow Theory: Edge Assignment

Network Dependency and User Agency Characteristics of Social Flow Transactions

- Pedestrian
- Car
- Bus or Taxi
- Train or Subway
- Airplane
- Communication

Lack of GIS Infrastructure

- No sparse matrices unless for OD logistics.
- No spatial selection for flows.
- No summary statistics.
- No singular entities.
- Need for social network prediction, modulation or partitioning.

Need for ESDA Tools

- R
- Build your own with D3.js
- Geometry: JTS, GEOS, Shapely

Open Alternatives

- STempo
- Factory Pattern in STempo for Polyglot Programming

Java
Python

Helper Methods
TPattern Implementation
<<interface>> TPattern
STempo GUI

Java
Python
Demo

Open Alternatives

• Same as GeoViz Toolkit

GeoTxt.org

Demo

Open Alternatives

• Two steps: NER and Geocoding
  • NER
    – Stanford NLP
    – GATE
  • Geocoding
    – GeoNames local
    – GeoNames webservice
  • NOT Google, Yahoo APIs if you need to respect TOS
GeoViz.js

SensePlace2

Open Alternatives

- Tons of JavaScript frameworks that do coordination of some kind
  - jQuery, Backbone, Ember, Angular, Tangle….
- Write your own

Demo

GEOG 897C: Cloud and Server GIS
Network accessibility
On-demand service
Resource pooling
Elasticity
Metered service

Demo

Open Alternatives

- Cloud providers
  - Amazon
  - Heroku

- MOOCs rock the world
  - Coursera
  - edX
  - Udacity
- Coursera Demo
Pep Talk
Spatial Data is Hot!

Collaborators and Sponsor

Research:  
Donna Peuquet  
Alan MacEachren  
Alex Klippel  
Anthony Robinson  
Jim Kroon  
Sam Stehle

Teaching:  
Sterling Quinn  
Ryan Baxter  
Jim Sloan  
Jim Detwiler  
Marty Gutowski  
Mark Wherley

Sponsors:  
NGA -- STempo  
DHS -- VACCINE  
Amazon

Thank you!
Visualizing Uncertainty and Decision-Making

Alan M. MacEachren
GeoVISTA Center
Department of Geography
Penn State University
maceachren@psu.edu

University of Zurich, March 21, 2013

Outline

- What is uncertainty (in the context of geographical information)?
- How can we depict uncertainty visually (in static and interactive displays)?
- How should we depict uncertainty? What do we know about which methods work?
- What happens when we depict uncertainty? What are the implications for reasoning and decisions?

Outline

Categorizing Uncertainty: GIScience

<table>
<thead>
<tr>
<th>Category</th>
<th>Spacial</th>
<th>Temporal</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accurate error</td>
<td>coordinates, buildings</td>
<td>+/- 1 day</td>
<td>counts, magnitudes</td>
</tr>
<tr>
<td>Precision</td>
<td>1 degree</td>
<td>once per day</td>
<td>nearest 1000</td>
</tr>
<tr>
<td>Lineage</td>
<td>geographic sources/transforms</td>
<td>time sources/transforms</td>
<td>attribute sources/transforms</td>
</tr>
<tr>
<td>Consistency</td>
<td>from / for a place</td>
<td>5 say Mon, 2 say Tues</td>
<td>multiple classifiers</td>
</tr>
<tr>
<td>Currency/ timing</td>
<td>age of maps</td>
<td>C = present - Time</td>
<td>census data</td>
</tr>
<tr>
<td>Reliability</td>
<td>knowledge of place</td>
<td>reliability of model</td>
<td>U.S. analyst vs. informant</td>
</tr>
<tr>
<td>Subjectivity</td>
<td>local ↔ outsider</td>
<td>expert ↔ trainee</td>
<td>fast ↔ guess</td>
</tr>
<tr>
<td>Interrelatedness</td>
<td>source independence</td>
<td>source independence</td>
<td>source independence</td>
</tr>
</tbody>
</table>

Categorizing Uncertainty: Cartography

<table>
<thead>
<tr>
<th>Category</th>
<th>Space</th>
<th>Time</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>position of vegetation boundaries</td>
<td>state birth rate</td>
<td>soil order</td>
</tr>
<tr>
<td>Attributes</td>
<td>total HIV positive case/county</td>
<td>mean monthly rainfall</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>date of the last glacier</td>
<td>mean monthly rainfall</td>
<td>mean monthly rainfall</td>
</tr>
</tbody>
</table>
Outline

- What is uncertainty (in the context of geographical information)?
- How can we depict uncertainty visually (in static and interactive displays)?

error bounds to depict uncertainty

“Visualisation of uncertainty using a cross-sectional transect (Lowell et al. 2007). The confidence layer depicted here can be represented by the statistical 90% confidence interval.”


depicting error range and agreement: IPCC Fourth Assessment Report: Climate Change 2007

Figure SPM.5. Solid lines are multi-model global average of surface warming (relative to 1980–1999) for the scenarios A2, A1B and B1, shown as continuations of the 20th century simulations. Shading denotes the ±1 standard deviation range of individual model annual averages. The orange line is for the experiment where concentrations were held constant at year 2000 values. The grey bars at right indicate the best estimate (solid line within each bar) and the likely range assessed for the six SRES marker scenarios. The assessment of the best estimate and likely ranges in the grey bars includes the AOGCMs in the left part of the figure, as well as results from a hierarchy of independent models and observational constraints. [Figures 10.4 and 10.29]

Uncertainty via visual variables: uncertainty visualization of snow avalanche intensities (buffer of 10%)


Uncertainty via visual and dynamic variables: Landcover classification

Uncertainty via sonic variables: Landcover classification


Uncertainty as a data layer: Meteo. forecast uncertainty


Visualizing uncertainty & ensemble data

Adjacent display

Coincident display overlay

Coincident display: visual variables

Figure A. Visualizations of mean and standard deviations. (Left) Mean and standard deviation are overlaid and proportions using color ramps. (Center) Mean is connected through a line ramp, and standard deviations is shown as an overlaid contour. (Right) Standard deviations is mapped to a height field and shown in color ramp.


Outline

- What is uncertainty (in the context of geographical information)?
- How can we depict uncertainty visually (in static and interactive displays)?
- How should we depict uncertainty? What do we know about which methods work?

Semiotics: the science of “signs” and sign systems

triadic model of signs as relations

interpretant

(signing)

vehicle

symbolization

(referent)
Visual Semiotics: which visual variable imply uncertainty?

- Visual variables

Experiment 1: Assessing intuitiveness

- Participants (31 in pilot + 72 in main study) rate the logic (intuitiveness) of the symbol set

Series 1: Results

- Series 1: Results
- Color saturation was not highly rated
- Fuzziness, location, and value rate high; arrangement is marginally successful
- Color hue and shape rate low (as expected), as do inverted fuzziness, location, color

NCHS atlas color/reliability schemes

- Cindy Brewer and Alan MacEachren researched the color and reliability map symbols for the Atlas of United States Mortality by the National Center for Health Statistics (CDC).

colors/textures to represent reliability

- Sample Test Maps:
  - cause: 16
  - color scheme: purple-green
  - reliability schemes map pairs color change texture overlay
Outline

- What is uncertainty (in the context of geographical information)?
- How can we depict uncertainty visually (in static and interactive displays)?
- How should we depict uncertainty? What do we know about which methods work?
- What happens when we depict uncertainty? What are the implications for reasoning and decisions?

Decision-making

- The impacts of uncertainty on analytical reasoning and decision-making are not well understood. Only a few studies have been carried out, e.g.
  - Deitrick and Edsall (2006) present evidence that uncertainty information has an influence on decision-making.
  - Leitner and Buttenfield (2000) found that uncertainty representation for facility location can clarify the map and lead to more accurate decisions.
  - Severtson and Meyers (2012), found that geo-uncertainty representation factors (“focus” / crispness of map contours and verbal-relative versus numeric risk expression) and personal characteristics (prior beliefs and numeracy) interacted to influence risk beliefs related to modeled cancer risk from air pollution.
  - Cliburn, et al (2002) “Decisions cannot be made from uncertain data, it only leads decision-makers to discount the results. Unfortunately, not considering uncertainty may lead to inappropriate decisions. A potential collaborator, who viewed the application in its later stages, suggested incorporating a reasoning network of potential actions to problems presented by the visualizations.”

Communicating ambiguity in risk

“Overall, results indicate incremental shading effectively conveys a dose-response message, and contour focus and risk expression show promise for improving decision makers’ understanding of information uncertainty.”


Findings:
• uncertainty influences decision-making
• coincident display better for complex tasks
• uncertainty via coincident depiction does not decrease data interpretation accuracy
• no performance difference for 2.5D vs 2D data-uncertainty depictions

Mapping Climate Change Uncertainty:

A pilot study of effects on risk perceptions and decision making

David Retchless, PhD candidate
dpr173@psu.edu

fallmeeting.agu.org/2012/eposters/eposter/gc43b-1024

Outline

- What is uncertainty (in the context of geographical information)?
- How can we depict uncertainty visually (in static and interactive displays)?
- How should we depict uncertainty? What do we know about which methods work?
- What happens when we depict uncertainty? What are the implications for reasoning and decisions?
Research Questions

- Which maps of temperature change and uncertainty are easiest to understand?
- Do map users combine magnitude and certainty of change when:
  - assessing risks?
  - making decisions?
- Which maps do users prefer?

Methods

- Survey: Mechanical Turk & Survey Gizmo
- 4 Ranking Questions, each ranking 7 map regions
  - Temperature change
  - Certainty (signal/noise, described as precision)
  - Harm to environment
  - Suitability for nature reserve (given temperature requirements)

Methods

- 274 respondents, randomly assigned 1 of 20 maps:
  - 10 types of maps
  - 2 emissions scenarios (high & low)
  - ~14 respondents/map
  - Data from CMIP5

Control (Temperature Only)

Conclusions: Understanding Magnitude & Certainty

- Temperature ranking was easy, uncertainty ranking was hard.
- Consistent with MacEachren et al. (1998), texture outperformed color for uncertainty ranking.
- Best maps for uncertainty ranking:

  - Control with small map
  - Texture – Lines
  - Texture – Spots

- Magnitude was primary driver of risk assessment and decisions.
Challenges

- understand components of uncertainty and their relationships to domains, users, and information needs
- develop representation methods for depicting multiple kinds of uncertainty
- develop methods and tools for interacting with uncertainty depictions
- understand how (or whether) uncertainty visualization aids exploratory analysis
- understand how knowledge of uncertainty influences information analysis, decision making, and outcomes
- develop methods to capture and encode analysts’ or decision makers’ uncertainty
- assess usability and utility of uncertainty capture, representation, and interaction methods and tools.

That’s all: Thanks
maceachren@psu.edu
www.geovista.psu.edu

Acknowledgements: Parts of the work represented here were supported by funds from the Dept of Homeland Security Visual Analytics for Command, Control, and Interoperability Environments Center of Excellence. A large number of people contributed to work represented. These include Mark Gahegan, Anthony Robinson, Robert Roth, ...
Gang Graffiti Automatic Recognition and Interpretation (GARI)

Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu
Acknowledgements

• This work is funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001

• Indiana Gang Intelligence Network (INGangNetwork)
GARI System Overview

1. Original image
   - Offline automatic analysis and labeling
     - Geoposition
     - Date and time
     - Extracted Features

2. Filtered results
   - Info + thumbnails

3. Original Database

4. Server

5. Offline manual filtering

6. Manual labeling
   - Additional Features

7. Addition to Database

Original image

Filtered results

Labeled image

Filtered Database

Manual labeling

Labeled image
GARI System Goals

• System that allows first responders to:
  – Collect gang graffiti and gang tattoos images
  – Analyze and interpret the images
  – Image retrieval
  – Interact with a database of images
  – Browse database using an interactive map

• Platforms
  – Android application
  – Web-based interface
  – iPhone
Gang Graffiti
Gang Tattoos
Gang Graffiti Interpretation

SHAPE

SYMBOLS

NUMBERS/LETTERS

TIME

COLOR
Image Analysis
Image Analysis

GANG RIVALRY TRACKING

- **Date:** 08/19/2010
- **Time:** 3.25 PM
- **Geo:** 41.387917, 2.169919

- **Date:** 01/03/2011
- **Time:** 5.11 PM
- **Geo:** 41.387917, 2.169919
Image Retrieval
User Filtering

**KNOWN GANG**
18th Street Gang

**CLEAR TOKENS**
Ruthless

**SPECIAL FONT TYPE**

**KNOWN SYMBOLS**
Down arrows
Mobile App

Main Screen

Capture Image

Send to Server

Possible Options

Browse Database

Trace Color

Find Similar Images
Gang Graffiti Recognition and Analysis Using a Mobile Telephone

Gangs are a serious threat to public safety throughout the United States. Gang members are continuously migrating from urban cities to suburban areas. They are responsible for an increasing percentage of crime and violence in many communities.

According to the National Gang Threat Assessment, approximately 1 million gang members belonging to more than 20,000 gangs were criminally active within all 50 states and the District of Columbia as of September 2008. Criminal gangs commit as much as 80 percent of the crime in many communities according to law enforcement officials throughout the nation.

Street gang graffiti is their most common way to communicate messages, including challenges, warnings or intimidation to rival gangs. It is, however, an excellent way to track gang affiliation and growth, or even sometimes to obtain membership information.
Web-Based Interface

Upload Image

Filename: IMG_0103.JPG

First responder ID: 000101

Date and Time: 2011:01:28 10:57:05

Upload image

Not the right image? Choose another one:

File: Choose File No file chosen Submit
Installation Process

• Your identity is authenticated by InGang you will receive a “welcome” email from gari2ind@ecn.purdue.edu
  – email contains your GARI credentials (id and password)
  – Need to change password on first use (you must also accept the Disclaimer)

• Android app – need Gmail email associated with you’re the phone – the app is then emailed to you (should be supplied when you register for access)

• iPhone – the link is in the “welcome” email (you MUST open the email on the iPhone)

• Questions or problems – gari2ind@ecn.purdue.edu
One Issue/Bug

- Web-Based Interface
  - https://108.59.50.111/~gari/
Gang Graffiti Automatic Recognition and Interpretation (GARI)

Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
Automatic Recognition and Interpretation of Gang Graffiti (GARI)

Edward J. Delp

Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA
ace@ecn.purdue.edu
Acknowledgements

• This work is funded by the U.S. Department of Homeland Security’s VACCINE Center under Award Number 2009-ST-061-CI0001

• Faculty and graduate students involved have included:
  – Professor Mireille Boutin
  – Andrew Haddad
  – Albert Parra
  – Kevin Lorenz
  – Bin Zhao
  – Joonsoo Kim
GARI System Overview

1. Offline automatic analysis and labeling
   - Geoposition
   - Date and time
   - Extracted Features

2. Filtered results
   - Info + thumbnails

3. Original Database

4. Server

5. Offline manual filtering

6. Manual labeling
   - Additional Features

7. Addition to Database
Project Status

• Deployed on Android Phone
• Main phone application is image database system and browsing (acquisition tool) with some limited analysis capabilities
• Desktop backend
  – Browse by radius
  – Upload images
  – View and edit image details
  – Interact with map
  – Searching for similar images
  – Integration with social media (Facebook and Twitter)
System Goals

• Network connected phone:
  – Image Analysis
  – User input
  – Interacts with a database of gang symbols
Gang Graffiti Interpretation

SHAPE
Simple, Straightforward

NUMBERS
42nd street gang

SYMBOLS
6-point star, pitchforks

COLOR
Goon Squad: Red/Black
Gang Graffiti Interpretation

**LETTERS**
East Side Gang

**POSITION/ALIGNMENT**
Letters at star points
Numbers in the middle
Letters at the bottom
Pitchforks upright

**TIME**
Black: 18 ST (18\textsuperscript{th} Street Gang)
Red: 13 SUR (Sureños 13)
Image Analysis – Color Recognition

MEXICANOS MALDITOS SUREÑOS 13

18 STREET GANG

SUREÑOS 13
Image Analysis – Scene Analysis

GANG RIVALRY TRACKING

- **Date:** 08/19/2010
  - **Time:** 3.25 PM
  - **Geo:** 41.387917, 2.169919

- **Date:** 01/03/2011
  - **Time:** 5.11 PM
  - **Geo:** 41.387917, 2.169919
Database Structure Descriptions

Database Fields: automatically populated

- Camera EXIF data
  - Make
  - Model
  - Focal length
  - ISO equivalent speed
  - Exposure time
  - Exposure bias
  - Exposure mode
  - Aperture
  - X/Y resolution

- White balance
- Compressed bit per pixel
- Metering mode
- Flash
- F number
- Interoperability offset
- Sensing method
- Custom rendered
- Digital zoom ratio
- YCbCr positioning
Database Structure Descriptions

Database Fields: automatically

- Image properties
  - File size
  - Resolution
Database Structure Descriptions

Database Fields: automatically/manually populated

- Geoposition
  - Country
  - State
  - County
  - City
  - ZIP code
  - Address

- Date and time
Database Structure Descriptions

Database Fields: manually populated

- First responder
  - Name/ID
  - Affiliation

- Tattoo (only GARI)
  - Tattoo?
  - Gang?
  - Prison

- Gang information
  - Gang Name
  - Gang Member
GARI App Features
Gang Graffiti – User Interface

App Version 2.1

Main menu

User options

Analysis results

Image matching
Gang Graffiti – User Interface

Color tracking
Gang Graffiti – Browse Database

• Browse by radius from current position
• Download images and information from server (image EXIF tags, gang related information)
  – Compare results
  – Track graffiti
• Browse map (Google Maps API)

• Network connection required
Gang Graffiti – Browse Database

Browse general results

Show graffiti on map

Inspect specific graffiti
Gang Graffiti – Browse Database
Desktop Version
Gang Graffiti – Browse Database

Desktop Version

Date and Time: 2011-01-28 09:17:21
GPS latitude: 40.429325
GPS longitude: -86.9126441667
File Size: 664501 bytes
Height: 1552 px
Width: 2592 px
Focal length: 4.31 mm
Camera make: HTC
Camera model: HTC Desire
Image id: 1111
Show in map
Gang Graffiti – Browse Database
Desktop Version

Specific image
All images in radius
Gang Graffiti – Browse Database

Desktop Version

More Information

Lat, Lon: 39.7671784, -86.1164571
Gang Graffiti – Color Recognition Device

Finger tracking

Color detection

Related gangs:
- Crips
- MS-13
- 18th Street
- Sureños 13
- Ñetas
Project Status

• Android App Version
  – Browse by radius
  – Upload images
  – View image details
  – Find similar images
  – Track color
  – Interact with map
  – Uses SSL security
  – Password protected
Project Status

• Desktop Database Version
  – Browse by radius/date/address
  – Upload images
  – View and edit image details
  – Interact with map
  – Uses SSL security
  – Password protected
How to Get The App?

- Contact gari@ecn.purdue.edu
  - It requires Android phone (app version 2.1)
  - We assign user IDs and initial passwords
  - Database server
System Requirements for GARI server

- **Hardware**
  - Intel Core i7 Processor (or similar)
  - Multithreading
  - 6M Cache
  - 2.20 GHz
  - at least 8 GB RAM

- **Image storage**
  - Taken with phone: 500 KB/image x 30,000 = 15 GB
  - Taken with camera: 5 MB/image x 30,000 = 150 GB
System Requirements for GARI server

• Web-server software
  – Server version: Apache/2.2.9 (Debian)
  – PHP Version: 5.2.6-1+lenny13
  – OpenSSH_5.1p1 Debian-5, OpenSSL 0.9.8g 19 Oct 2007

• Database management system
  – psql (PostgreSQL) 8.1.11
  – contains support for command-line editing
Predictive Visual Analytics

David S. Ebert
Who We Are: International Team of Experts - 75+ Faculty, 25 institutions

<table>
<thead>
<tr>
<th>Universities</th>
<th>Universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Purdue University</td>
<td>- University of Houston, Downtown</td>
</tr>
<tr>
<td>- Georgia Institute of Technology</td>
<td>- Virginia Tech</td>
</tr>
<tr>
<td>- Pennsylvania State University</td>
<td>- Indiana University</td>
</tr>
<tr>
<td>- Stanford University</td>
<td>- Florida International University</td>
</tr>
<tr>
<td>- University of North Carolina at Charlotte</td>
<td>- University of Texas at Austin</td>
</tr>
<tr>
<td>- University of Washington</td>
<td>- Morgan State University</td>
</tr>
<tr>
<td>- Simon Fraser University</td>
<td>- Navajo Technical College</td>
</tr>
<tr>
<td>- University of British Columbia</td>
<td>- University of Stuttgart</td>
</tr>
<tr>
<td>- Justice Institute of British Columbia</td>
<td>- University of Swansea, UK</td>
</tr>
<tr>
<td>- Ontario Institute of Technology</td>
<td>- Oxford</td>
</tr>
<tr>
<td>- Dalhousie University</td>
<td>- University of Calgary</td>
</tr>
<tr>
<td></td>
<td>- University of Manitoba</td>
</tr>
<tr>
<td></td>
<td>- Carleton University</td>
</tr>
<tr>
<td></td>
<td>- University of Victoria</td>
</tr>
</tbody>
</table>
VACCINE Multidisciplinary Team

- Experts in computer science, image analysis, signal processing, cognitive science, decision theory, management science, geography and GIS, HCI, visualization, HPC, statistics, political science …

- Partners in nursing, public health, economics, databases, command and control, emergency management
VACCINE’s Value

**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
Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Federal Emergency Management Agency
  • US Customs and Border Patrol
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies

• Fusion Centers
  • Ohio and Indiana
Example Projects

- Predictive Analytics
- Financial Visual Analytics
- TSA
  - Integrating staffing, financial, performance metric, legal data into a unified VA system
- Law Enforcement
- USCG – Risk-based decision making
Predictive Analytics

• Use STL per data item to develop predicted values with confidence intervals
• Applied to health data and crime data
• Adapting to appropriate spatial resolution for most accurate predictions
• Spatial prediction based on population adapted kernel density estimation
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Respiratory Count

Date

Linked Predictive Crime Models by Type
Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space
Example: Drug Abuse Violations Vs. Burglaries – Focus on Geospatial Hotspots

- High positive correlation in same neighborhood with zero lag
WireVis – Streaming Data - Multiple Linked Views

- Temporal, geospatial, theme, cluster, list views with association linkages between views

- **Search by Example**
  - (Find Similar Accounts)

- **Keyword Network**
  - (Keyword Relationships)

- **Strings and Beds**
  - (Relationships over Time)
Competitive Business Intelligence Based On Point of Sales Data

- Characteristics of point of sale data
  - Multivariate
    - Large # of dimensions
      - E.g., 38 categories(products) in 288 stores
  - Temporal: 18 months
  - Spatial
    - Stores located all over Costa Rica
- Requirements
  - Supporting easy comparison among companies
    - E.g., visualize all data at once, sorting by importance
    - Enable geographical comparison
  - Easy recognition of any change in sales
    - E.g., proportional legends
  - Forecast for decision-making
Market Analyzer

Pixel-oriented Display Matrix

Geographical View

Proportional Legend View

Stacked Bar View & Time Sliders

Line Graph View

Sales View

Trend View

Growth View

Capital City

Sales Lines for Stores

Trend Lines

Forecast

Time slider widgets
RiskVA: Key Credit Risk Analytics Challenges

- Consumer credit data is large, temporal, and related across multiple investments and financial markets.
- The data are heterogeneous, not clean, have missing values, may be misleading and inefficient to explore.
- The data contains important behaviors and relations/groupings that change over time.
- Analysts need to have customer-centric interactive systems to achieve full analysis of risks.
RiskVA Data Integration

- Large Credit Incidence Data (500,000 data points tested)
- Statistical Data Transformation
- Entity Heatmap
- Trend Analysis
- Product Comparison

Variables

Time

Entity

Data Cube

System Components
RiskVA Overview: Interactive Exploratory Visualizations

Identification of Emerging Risks
Law Enforcement

• Law enforcement visual analytics
  • VALET, iVALET (iPhone/iPad), CrimeViz
    • Visual Analytics Law Enforcement Toolkit
    • Analyzing crime patterns and time of day problems
• Gang activity analytics
  • GARI
    • Gang Graffiti Recognition and Interpretation using a mobile telephone
    • Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
• Document visual analytics
  • JIGSAW
    • Visualization for investigative analysis
    • Discovery of hidden relationship and threats across documents
We Have an App For That!

Our Mobile Tools

• VALET
• Evacuation Planning
• Rosetta Phone
• Hazmat app
• Gang graffiti app
• Tatoo app
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
USCG: Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Superstorm Sandy: Used for resource allocation in response and in rebuilding
• Used for Hurricane Irene resource allocation decisions
• Informed Commandant’s budget testimony and recommendations to Congress
• Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Risk-Based Allocations

- Comparative visual analysis of mission cases/hours vs. staffing hours
- Comparative visualization of resources vs. risk
- Trend visual analytics
  - Increase/decrease in resource allocation
  - Increase/decrease in risk (total, mitigated, residual)
  - Increase/decrease in incidents
- Exploration of alternatives and effect on risk
- Predictive analytics based on historical data (STL and EWMA)
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
**Visual Analytics Law Enforcement Toolkit (VALET, iVALET)**

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (September 2012) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

**VALET delivered:**
- Spring 2011: WL, Lafayette Police

**iVALET delivered:**
- October 2011: Purdue, WL Police
VALET Overview
Explore criminal, traffic and civil data on-the-go
Risk assessment
Use current spatial + temporal context into analysis
Fundamentals of Visual Analytics

Ross Maciejewski and David S. Ebert
The Ongoing Data Deluge

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

---

The Ongoing Data Deluge

• FedEx’s ships more than 8.5 million packages per day (http://www.fedex.com/us/about/today/companies/corporation/facts.html)

• Consumers carry more than 1 billion Visa cards worldwide. More than 450 million of those cards are in the United States (http://www.creditcards.com/credit-card-news/credit-card-industry-facts-personal-debt-statistics-1276.php)
Data Overload

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

• The goal is to take all of this data and transform it into information.
• How many terabytes of data we have collected doesn’t matter, it’s how many petaflops of insights we can generate from this data.
• We need to make the data understandable to people and a key way of doing this is through visualization.
<table>
<thead>
<tr>
<th></th>
<th>Animal</th>
<th>Brain Weight (kg)</th>
<th>Body Weight (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Animal</td>
<td>Brain Weight (kg)</td>
<td>Body Weight (kg)</td>
</tr>
<tr>
<td>2</td>
<td>Mountain beaver</td>
<td>1.35</td>
<td>465</td>
</tr>
<tr>
<td>3</td>
<td>Cow</td>
<td>465</td>
<td>423</td>
</tr>
<tr>
<td>4</td>
<td>Grey wolf</td>
<td>36.33</td>
<td>119.5</td>
</tr>
<tr>
<td>5</td>
<td>Goat</td>
<td>27.66</td>
<td>115</td>
</tr>
<tr>
<td>6</td>
<td>Guinea pig</td>
<td>1.04</td>
<td>5.5</td>
</tr>
<tr>
<td>7</td>
<td>Diplodocus</td>
<td>11700</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>Asian elephant</td>
<td>2547</td>
<td>4603</td>
</tr>
<tr>
<td>9</td>
<td>Donkey</td>
<td>187.1</td>
<td>419</td>
</tr>
<tr>
<td>10</td>
<td>Horse</td>
<td>521</td>
<td>655</td>
</tr>
<tr>
<td>11</td>
<td>Potar monkey</td>
<td>10</td>
<td>115</td>
</tr>
<tr>
<td>12</td>
<td>Cat</td>
<td>3.3</td>
<td>25.6</td>
</tr>
<tr>
<td>13</td>
<td>Giraffe</td>
<td>529</td>
<td>680</td>
</tr>
<tr>
<td>14</td>
<td>Gorilla</td>
<td>207</td>
<td>406</td>
</tr>
<tr>
<td>15</td>
<td>Human</td>
<td>62</td>
<td>1320</td>
</tr>
<tr>
<td>16</td>
<td>African elephant</td>
<td>6654</td>
<td>5712</td>
</tr>
<tr>
<td>17</td>
<td>Triceratops</td>
<td>9400</td>
<td>70</td>
</tr>
<tr>
<td>18</td>
<td>Rhesus monkey</td>
<td>6.8</td>
<td>179</td>
</tr>
<tr>
<td>19</td>
<td>Kangaroo</td>
<td>35</td>
<td>56</td>
</tr>
<tr>
<td>20</td>
<td>Golden hamster</td>
<td>0.12</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>Mouse</td>
<td>0.023</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Which animal weights the least/most? Is there a relationship between brain weight and body weight? If so, are there any outliers?
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)

\(^1\) – Leonardo da Vinci. *Old Man Seated on Rocky Outcrop, Seen in Profile to the Right, with Water Studies*. Windsor Castle, Royal Liberty, RL 12579r, c. 1510-1513.
Visualization Through the Ages

• John Snow’s map of Cholera cases in the London epidemic of 1854

---

Visualization Through the Ages

- Charles Minard’s 1869 flow map\(^1\) showing Napoleon’s disastrous Russian campaign of 1812.

1 – Charles Joseph Minard: Mapping Napoleon’s March, 1861 by John Corbett, Center for Spatially Integrated Social Science
Visualization Through the Ages

• James Maxwell’s\(^1\) thermodynamic surface sculpture (1874)
• Three-dimensional plot of the states of a fictitious water-like substance
• Coordinates are volume (x), entropy (y), and energy (z)

---

Why is Visualization Helpful?

• It utilizes the high bandwidth of the human visual systems
• The human mind is fast and parallel
• Humans are great at visual pattern recognition
• We have pre-attentive visual phenomena
• Visual tools can extend memory and cognitive capacity
• We think visually!
• “A picture is worth a thousand words.” – *Printers Ink*, pp. 96-97
  December 8, 1921
Why Is Visualization So Helpful?

• Amplifies cognition
  – Expands our working memory by allowing us to offload results
  – Reduces the search time
  – Pattern detection and recognition can be improved through perceptual cues and inference
  – Visualizations can be designed to control attention interaction for improving cognition

Readings in Information Visualization: Using Vision to Think, SK Card, J Mackinlay and B. Shneiderman, 1999
What is Visualization?

• “The use of computer-supported, interactive visual representations of data to amplify cognition.”¹

• This is not simply the process of making a graphic or an image, the goal is to create insight, not pretty pictures

• We want to help people form a mental image of something and internalize their own understanding

• We want to promote discovery, decision making and explanations

¹Readings in Information Visualization: Using Vision to Think, SK Card, J Mackinlay and B.Shneiderman, 1999
What is Visualization

- We want to find and utilize cognitive and perceptual principles
- We want to optimize our visualizations and our interactions with the visualization according to these principles

Maneesh Agrawala, Chris Stolte: Rendering effective route maps: improving usability through generalization. SIGGRAPH 2001: 241-249
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
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
Purpose of Visualization

• Analysis – Understand your data better and act upon that understanding
  – Given a data set, compare, contrast, assess, evaluate
  – Solve a problem!
• Presentation – Communicate and inform others more effectively
• Visualization is most useful in exploratory data analysis

1 – J. W. Tukey. Exploratory Data Analysis
The Information Seeking Mantra

Overview first,
then zoom and filter,
details on demand

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

W. S. Cleveland

*The Elements of Graphing Data*
Is This the Total Solution?

• Traditional visualization misses several key factors in how people solve difficult problems
• Visual analytics aims to overcome these shortfalls and create an interactive human-computer exploration and decision making environment
Visual Analytics

- Visualization is good for exploring data, but we can do more than just explore.
- “Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.”
- 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.
- “A graphic display has many purposes but it achieves its highest value when it forces us to see what we were not expecting.”

Example Systems And Applications
Example: Public Health Visual Analytics

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

Pandemic spread on day 37 with no decision measures implemented

Pandemic spread on day 37 with all decision measures implemented
Syndromic Surveillance

– Syndromic surveillance is the detection of adverse health events focusing on pre-diagnosis information to improve response time

– Pre-diagnosis information can consist of multiple data sources:
  • Over the counter medicine sales
  • News reports on emerging diseases
  • Pro-med news feeds
  • Emergency department chief complaints
Syndromic Surveillance


<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Date</th>
<th>Chief Complaints</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>9398</td>
<td>4/16/09</td>
<td>Ear pain</td>
<td>Karachi</td>
</tr>
<tr>
<td>10816</td>
<td>4/16/09</td>
<td>Stuffy nose</td>
<td>Lebanon</td>
</tr>
<tr>
<td>1491</td>
<td>4/16/09</td>
<td>Fever</td>
<td>Allepo</td>
</tr>
<tr>
<td>16237</td>
<td>4/16/09</td>
<td>Head bleed</td>
<td>Yemen</td>
</tr>
</tbody>
</table>

1) Respiratory
2) Gastro-intestinal
3) Neurological
4) Botulinic
5) Constitutional
6) Rash
7) Hemmoraghic
8) Other
Hypothesis Generation and Exploration

Time Series Modeling

- Seasonal-Trend Decomposition Based on Loess
  - Time series can be viewed as the sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components

Hafen, R., Anderson, D., Cleveland, W., Maciejewski, R., Ebert, D., Abusalah, A., Yakout Monitoring Disease Counts,” *BMC Medical Informatics and Decision Making*, 9(21), 2009
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- **Actual**
- **Predicted**
- **Lower**
- **Upper**

Respiratory Count

Date

Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

- Pandemic Influenza Planning Tool
- Models user specified:
  - Pandemic influenza characteristics
  - County population, demographics, hospital beds
- Decision measures
  - Strategic National Stockpile deployment
  - School Closures
  - Media Alerts

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

Pandemic spread on day 37 with no decision measures implemented
Pandemic spread on day 37 with all decision measures implemented
Visual Analytics Uses for Crisis Management and Response

• Trend analysis, clustering, anomaly detection
• Interactive, multi-day, month, type investigation
• Multisource, multimedia data integration & analysis

• Purpose:
  – Planning for resiliency
  – Long-term analysis
  – Predictive analytics
  – Training
  – Detection
  – Investigation
  – Response
  – Recovery, remediation
Visual Analytics Situational Awareness \(\leftrightarrow\) Sensemaking (courtesy Alan MacEachren)

Situational awareness model (after Endsley, 1995; top) compared to model of sense making (after Pirolli & Card, 2005; bottom)
Situational Awareness, Planning, Investigation, Response Visual Analytics

• Situational Awareness and Assessment from massive data
  – In-field and desktop historical and current incident/crime information
  – Innovative spatiotemporal visualization, analysis, prediction
  – Image analysis tools
  – Integrated statistical forecasting and simulation models
  – Advanced social media analytics

• Tools for Intelligence analysis

• Planning and response support
  – Integrated simulation, resource allocation, census, weather data
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
- Interactive, multi-day, month, type investigation
- Multisource, multimedia massive data integration & analysis
Example Maritime Safety Visual Analytics Environment

- Supports decision making and risk assessment
- Interactive exploration and analysis of trends, patterns and anomalies
- Allows analysis of risks associated with
  - Closing one or more Coast Guard stations
    - Find optimal stations that absorb work load of the closing station
  - Allocating new resources
    - Impact on safety and efficiency of operation
- Currently being used by analysts at the U.S. Ninth District, Fifth District, HQ, and Atlantic Commands
Example: USCG D9 Search & Rescue Operational Analysis
Total Risk Based on SAR Cases
System Features:
SAR Risk Profile

Time taken by CG stations to deploy an asset to the Great Lakes to respond to an incident.
Response Efficiency – Current Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
D5: Cost based on Resources used for the Open-Suspended Sorties over 5 years (2006-2011)
Resource Allocation and Risk-Based Decision Making

• Explore risk-based decision making and utilize historical data for analysis and prediction
  – Total Risk, Mitigated Risk, Residual Risk
  – Explore 11 different USCG missions
  – Explore allocation of assets with different capabilities
  – Explore staffing, utilization, assets vs. risk measures
  – Perform What-If scenarios
Visual Analytics Law Enforcement Toolkit (i)VALET
Interesting VA features:
• Time a first class citizen
• Temporal exploration and comparisons
• Integrated predictive models
• Interactive multivariate correlation and visual analytics
• Novel statistical predictive model incorporated for planning
• Novel interactive anomaly detection from social media

VALET delivered:
• Spring 2011: WL, Lafayette Police

iVALET delivered:
• October 2011: Purdue, WL Police
Linked Predictive Crime Models by Type

2008 (red) vs. 2007 (blue background)
Example:
Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame
  PU Lost: 10-38

- Homecoming (Sat.)

- PU vs. Illinois
  PU Won: 21-14

- PU vs. Iowa
  PU Lost: 21-31
VALET Feature: Correlating Datasets at Multiple Granularities

- Correlative analysis framework for exploring correlations among multi-variate/-source data
- Determining forces that determine nature of relationships among datasets
- Detection of periodic properties among different data variables
- Allows correlations at multiple spatial and temporal granularities
Case Study: Drug Abuse
Violations vs. Burglaries

- Over geospatial hotspots
VALET Feature:
Real-time Twitter Monitoring and Integration

- Topic extraction using novel STL based remainder estimation technique

- Dynamically linked views providing options to monitor emerging / emergent twitter feeds

- Topics extracted shown as a dynamic word cloud
Spatiotemporal Social Media Analytics for Abnormal Event Detection
Correlation Between Multiple Social Media

<table>
<thead>
<tr>
<th>Date</th>
<th>Twitter (z-score)</th>
<th>Flickr (z-score)</th>
<th>Date</th>
<th>YouTube (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/17, 2011</td>
<td>2.53</td>
<td>0.97</td>
<td>9/18, 2011</td>
<td>5.98</td>
</tr>
<tr>
<td>9/30, 2011</td>
<td>3.20</td>
<td>3.50</td>
<td>10/2, 2011</td>
<td>1.38</td>
</tr>
<tr>
<td>10/5, 2011</td>
<td>3.17</td>
<td>3.36</td>
<td>10/6, 2011</td>
<td>2.75</td>
</tr>
</tbody>
</table>
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

• Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Jigsaw:
Visual Analytics for Investigative Analysis
and Exploration of Document Collections

Goal: Assist investigators with understanding, sense-making, and analysis of large, unstructured and structured document collections

Approach: Provide multiple visual perspectives on the documents and entities within them, highlighting connections between entities

Users: Over 150 downloads – active users include the following:

• Seattle PD
• Rock Hill, South Carolina PD
• Indianapolis PD
• West Lafayette, PD (student embedded)
• TSA
Gang Graffiti Recognition and Analysis 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 40 users and 450 graffiti images acquired

GARI delivered:
• Summer 2011:
  • IMPD gang detectives
• August 2011:
  • IMPD at large
  • Ind Fusion Gang Task Force
• October 2011:
  • Gang detectives across Indiana
Crime Analytics & SA Issues and Techniques

• Fuse data from a variety of sources
  – Law enforcement records management
  – Weather and phases of the moon
  – Street light locations, bus routes
  – Tracking release data of offenders
  – Civil court data
  – Social Media
  – Local event calendar

• Reliable predictive models

• Understandability and trust of predictions

• Main Question: What helps officers, detectives, chief do their variety of jobs?
Data, Models, Reductions, and Visual Representations
Data Models

- Data models are structured forms suitable for computer-based transformations.
- These structures exist in the original data or are derivable from the original data.
- Structures retain the information and knowledge content and the related context within the original data.
- These structures are transformable into lower-dimensional representations for visualization and analysis.

Data Models vs. Conceptual Models

• Data models are mathematical abstractions
  – We can perform numerical operations, addition, subtraction, etc.

• Conceptual models are our mental constructs
  – These contain semantic structure and support reasoning
  – Think of giving directions to someone using landmarks
Data Types

• **Nominal**
  – Data whose categories have no implied ordering
  – Examples include political affiliations of a population

• **Ordinal**
  – Data that has a specified order, but no specified distance metric
  – Examples include beverage sizes at McDonalds (Small, medium, large)

• **Interval**
  – Data that has measurable distances
  – Examples include periods of time (second, minute, etc.) – the zero point is arbitrary

• **Ratio**
  – Same as interval, but include a zero point
  – Example include Celsius scale, height above sea level

The Visual Analytics Pipeline

- We want to take these different data types and map them to an appropriate visual representation
- **Data Analysis** – data are prepared for visualization (smooth, interpolate, transform)
- **Filtering** – A subset of the data (usually user defined) is selected for visualization
- **Mapping** – Data are mapped to geometric primitives and their attributes
- **Rendering** – Geometric data are transformed to image data

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 which enhances the analytical reasoning process
Bertin’s Visual Variables

• Visualization is concerned primarily with a mapping to visual form
• \([x,y]\)
  – Position
• \([z]\)
  – Size (Taille)
  – Value (Valeur)
  – Color (Couleur)
  – Texture (Grain)
  – Orientation
  – Shape (Forme)

J Bertin (1967), *The Semiology of Graphics*
<table>
<thead>
<tr>
<th>Form</th>
<th>Point</th>
<th>Line</th>
<th>Area</th>
<th>Surface</th>
<th>Solid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
</tr>
<tr>
<td>Shape</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
</tr>
<tr>
<td>Rotation</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
</tr>
</tbody>
</table>

| Color     |       |     |      |         |       |
| Brightness| ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
| Hue       | ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
| Saturation| ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |

| Texture   |       |     |      |         |       |
| Granularity| ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
| Pattern   | ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
| Orientation| ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |

| Optics    |       |     |      |         |       |
| Blur      | ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
| Transparency | ⬤      | ⬤    | ⬤    | ⬤      | ⬤    |
How Many Variables to Use?

- The number of visual variables necessary for the representation is at least equal to the number of components in the information.
- With three components, the information can be perceived as a single image.
- Otherwise, we need several images (this is often termed *small multiples* and will be discussed in a future lecture).
- The number of components is the best basis for a classification of graphic constructions.

J Bertin (1967), *The Semiology of Graphics*
How Do I Choose the Encoding?

Expressing Data With Color

• Results from research on visual attention can be used to assign visual features to data values
• One of the key components of visually representing data is choosing the appropriate color scale
• There is no “best” color scale
• Choice depends on\(^1,^2\)
  – data type
  – problem domain
  – visual representation
  – Questions the analyst is asking of the data
• While there is no “best” choice, there are design principles

Design Principles for Color Schemes

• **Order**\(^1\) – Given a univariate data type, the color scale that is chosen to map the data must represent a perceived ordering

• **Separation**\(^2\) – Important differences between ranges of the variable should be represented by colors that can be perceived as being different
  – Not only should they be perceived as different, but also equal

• **Aesthetics**\(^3\) – color map should be aesthetically pleasing, contain a maximum perceptual resolution, and ordering should be intuitive

---

Univariate Color Schemes

• Qualitative scheme
  – Rainbow color scale is one of the most commonly used, but it is a poor color map in a large variety of domain problems
  – Ordering of the hues is unintuitive
  – Nominal data types can use this scale as no ordering is implied

Univariate Color Schemes

• **Sequential color scheme**
  – Simplest is the gray scale map where variable is mapped to brightness
  – Sequential maps represent ordered data
  – Dark colors typically represent high ranges, bright, low
  – Benefits are that the scale is intuitive
  – Weakness is that limited number of distinguishable colors can be represented

Univariate Color Schemes

• Divergent color scheme
  – Provides means for variable comparisons
  – Best suited for ratio data where there is some meaningful zero point
  – Scale lacks a natural ordering of colors
  – Careful choices must be made in choosing high and low ends
  – Can use concept of cool (blues) and warm (reds and yellow) colors

Multivariate Color Schemes

Cleveland’s Heirarchy

- Cleveland evaluated elements when isolated
- Tasks were restricted to magnitude and ratio comparisons
- Research indicates this hierarchy may be best in pre-attentive stages or when focusing only on portions of a graphic

1. Position along a common scale
2. Position along nonaligned scales
3. Length
4. Angle/Slope
5. Area
6. Volume
7. Color

1 – WS Cleveland, *The Elements of Graphing Data*, 1985
What if I Combine Encodings?

• So, what happens when we combine several scales in a single display?
• Can we represent one quantitative dimension with color and another with orientation and expect a perceiver to respond to both dimensions?
• Do these things make psychological sense?
Integral Versus Separable Dimensions

• “A configuration has properties that have to be expressed as some form of interaction or interrelation between the components, be they features or dimensions.”

• Integral dimensions are not as easily decomposable by perceivers as separable dimensions

• Separating hue from brightness in a color (integral dimensions) are harder to decompose than say size and texture (separable dimensions)

---

Combinatorics of Encodings

• Challenge:
  – Pick the best encodings from the exponential number of possibilities \((n+1)^8\)

• Principle of Consistency: The properties of the image should match the properties of the data

• Principle of Importance Ordering: Encode the most important information in the most effective way

1 – This slide is borrowed from Pat Hanrahan’s “From Data to Image” lecture: http://graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/
The Curse of Dimensionality

• A term coined by Bellman in 1961
• Refers to the problems associated with multivariate data analysis as the dimensionality increases the available data becomes sparse
• Sparsity is a problem for any method that requires statistical significance
• Sometimes data dimensions are redundant and can be reduced with minimal information loss
• In visualization we are also limited with screen space and the number of available visual variables, so choosing the most appropriate dimensions is key

The Curse of Dimensionality

• So, what can we do?
  – We can incorporate prior knowledge of the data
  – We can smooth the target function
  – We can reduce the dimensionality
The Practicality of the Curse of Dimensionality

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

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

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

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

- Two approaches are available to reduce dimensionality
  - **Feature extraction**: creating a subset of new features by combinations of the existing features
  - **Feature Selection**: choosing a subset of all the features

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_N
\end{bmatrix} \rightarrow \begin{bmatrix}
  x_{i_1} \\
  x_{i_2} \\
  x_{i_M}
\end{bmatrix} \quad \begin{bmatrix}
  x_1 \\
  x_2 \\
  x_N
\end{bmatrix} \rightarrow \begin{bmatrix}
  y_1 \\
  y_2 \\
  y_M
\end{bmatrix} = f\left(\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_N
\end{bmatrix}\right)
\]

- Given a feature space \( x_i \in R^N \) find a mapping \( y = f(x) : R^N \rightarrow R^M \) with \( M<N \) such that the transformed feature vector \( y \in R^M \) preserves (most of) the information or structure in \( R^N \)
- An optimal mapping is one that does not increase error
Principle Components Analysis

• One of the most commonly applied dimension reduction techniques
• PCA is a deterministic analytical procedure that utilizes an orthogonal transformation to reduce a set of sample observations with potentially correlated variables into a set of uncorrelated variables called principal components
• The number of principal components will always be less than or equal to the original number of variables in the sample set

Principle Component Analysis

• The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector
  – PCA simple performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance
  – There is no guarantee that the directions of maximum variance will contain the most interesting/important features
K-Means Clustering

- K-means clustering
  - Goal is to partition n-observations into k-clusters where each observation belongs to the cluster with the nearest mean.
  - This provides information on how the data groups together.
  - Drawbacks are the use of Euclidean distance and the fact that k is user-defined.
  - The relaxed solution of K-means clustering is given by the PCA, and the PCA subspace spanned by the principal directions is identical to the cluster centroid subspace.

Visual Analytics

• We want to know how to best create images of data effectively in order to facilitate advanced data analysis
  – Spotting outliers
  – Discriminating clusters
  – Checking distributional and other assumptions
  – Examining relationships
  – Comparing mean differences
  – Observing time-based processing
• What are the two main ways of presenting multivariate data sets?
  – Directly (textually) – Tables
  – Symbolically (pictures) – Graphs

• How do we decide which to use, and when?

1 – Descriptions on this slide are borrowed from John Stasko’s “Multivariate Data & Tables and Graphs” lecture: http://www.cc.gatech.edu/~stasko/7450/Notes/data.pdf
Tables?

• Use tables when
  – The document will be used to look individual value
  – The document will be used to compare individual values
  – Precise values are required
  – The quantitative info to be communicated involves more than one unit of measure

• Use graphs when
  – The message is contained in the shape of the values
  – The document will be used to reveal relationships among values

1 – Descriptions on this slide are borrowed from John Stasko’s “Multivariate Data & Tables and Graphs” lecture: http://www.cc.gatech.edu/~stasko/7450/Notes/data.pdf
2 – S. Few, Show Me the Numbers: Designing Tables and Graphs to Enlighten,
Graphs?

- **Graph**
  - Visual display that illustrates one or more relationships among entities
  - Shorthand way to present information
  - Allows a trend, pattern or comparison to be easily comprehended

- **Critical to remain task-centric**
  - Why do you need a graph?
  - What questions are being answered?
  - What data is needed to answer those questions?
  - Who is the audience?

1 – Descriptions on this slide are borrowed from John Stasko’s “Multivariate Data & Tables and Graphs” lecture: http://www.cc.gatech.edu/~stasko/7450/Notes/data.pdf
Histograms

• Typical first look visualization method
  – Shows the shape of the data distribution
• The choice of the histogram bin width greatly impacts the resultant visualization
• There is no “best” number of bins, instead, different bin sizes can reveal different features of the data
• Some methods work to determine an optimal number of bins, but these methods make assumptions on the underlying data distribution
Histogram Binning

• Number of bins (k) can be user specified or chosen from a suggested bin width (h) such that

\[ k = \left[ \frac{\text{max } x - \text{min } x}{h} \right] \]

• Common choices for k include the square-root choice where \( k = \sqrt{N} \)

• Sturge’s formula\(^1\) where \( k = \lceil \log N + 1 \rceil \)

• Scott’s choice\(^2\) where \( h = \frac{3.5\sigma}{N^{\frac{1}{3}}} \)

• Freedman-Diaconis rule\(^3\) \( h = 2IQR(x)N^{-\frac{1}{3}} \)

Frequency count of batting average

User defined bin width

Batting average

.287 .292 .297 .302 .307 .312 .316 .321 .326 .331 .336
Frequency count of batting average

Sturges’ Formula

Batting average
• Drawbacks
  – Density estimate depends on starting position of the bins
  – Discontinuities of the estimate are not due to the underlying density, only artifacts of the chosen bin location
  – As the number of dimensions grow, need many samples or else most bins would be empty
• Histograms are best suited for quick visualizations in 1 or 2 dimensions
Bivariate Case – Stacked Bar Graph

Bivariate Case - Scatterplot

- Visualizes discrete data values along two axes
- Used as a means of analyzing bivariate relationships
- Quick means of assessing outliers, clusters and distributions
- Putting a line through the data can help assess trends, but can also mislead viewer
Time Series Visualization

- We can think of a line-graph time series visualization as a form of a histogram where the x-axis bins the data by some increment of time.
- Depending on the binning used, temporal data patterns can emerge.

Number of Coast Guard Search and Rescue Missions by Day
Number of Coast Guard Search and Rescue Missions by Week
Non-Data Components of Graphs

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


Graph Aspect Ratios

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

Data Distributions

• The type of data distribution affects not only the way the data should be analyzed, but also the way it should be visualized!
• Pre-conditioning data through transformations is a key step for appropriate analysis and visualization.
The Normal Distribution

• The Normal (Gaussian) distribution has many features which make it popular

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

– It can be fully characterized with two parameters
– The probability of any value can be obtained by knowing how many standard deviations it is away from the mean
– Many statistical measures and tests are well defined for this distribution

• Must be careful not to characterize non-Normal data as Normal
Skewness

• Measure of the asymmetry of the probability distribution
• Can be positive, negative or undefined
• Negative skew indicates that the tail on the left side is longer than on the right
• Positive skew indicates that the tail on the right side is longer than on the left
• A zero value indicates the tails are relatively evenly distributed on both sides of the mean, but does not necessarily imply symmetry
Visualizing Skewed Data

• Plotting skewed data compresses data values into smaller regions, resulting in lower visual fidelity\(^1\)

• How to overcome skewness?
  – Remove data outliers
  – Interactive techniques (zoom, pan, brush)

• This can help for the visual analysis, but not always!
  – What if I want to compare data sets with different skews?
  – What if I want to perform statistical analyses?

---

1 - W. S. Cleveland. *Visualizing Data*. Hobart Press, 1993
The Power Transformation

• We can transform data to a normal approximation by a power transformation\(^1\)

• Such transformations reduce the effects of:
  - Skewness (transformations changes the range of data values and helps fit data onto the display)
  - Random noise (transformations can help show global trends)
  - Monotone spread

---

The Power Transformation

• The Box-Cox\textsuperscript{1} Power Transformation
  – x is the observed (recorded) data and \( \lambda \) is the power

\[
x(\lambda) = \frac{x^\lambda - 1}{\lambda}, \quad (\lambda \neq 0)
\]
\[
= \ln(x), \quad (\lambda = 0)
\]

• The power (\( \lambda \)) is chosen by maximizing the logarithm of the likelihood function

Tag Clouds and Wordles

- Visual representation for text data where words are placed and scaled based on some statistical measures.
- Font size is typically determined by the number of instances a word is used.

Text cloud comparing 2002 State of the Union Address by U.S. President Bush and 2011 State of the Union Address by President Obama. Created at TagCrowd.

Multivariate Case - Scatterplot Matrix
Parallel Coordinate Plots

Issues With Parallel Coordinate Plots

• Different variables can take different values with very different ranges
  – Need to normalize data ranges (maybe do a power transformation and then a normalization?)

• Order of the parallel coordinate plots has a major impact on the resultant visualization

• The more variables we plot, the more lines we get and the more clutter that we get

Star Plot

- Lay out axes in a radial layout, length of a ray emanates from a central point
- Rays are then joined together by a polyline drawn around the outside

Geovisualization

• Geovisualization primarily denotes tools and techniques that are designed to support analyses that focus on datasets with a geographic component
• Visual representations are designed and built utilizing cartographic principles
• Looking for trends over geographic regions
Choropleth Maps

Areas of the map are shaded in proportion to a measured variable.

Coloring is based on a classification (histogram binning) of the distribution of the measured variable.


http://www.nass.usda.gov
Coloring Choropleth Maps¹

• The number of colors depends on the number of classes (bins)
• Too many classes can overwhelm the user and distract them from seeing trends
• Too many classes can compromise legibility as colors become difficult to distinguish
• Typical cartographic rule of thumb is 5-7 classes
• Typical coloring schemes² include sequential, divergent and qualitative
• For more details on mapping as a visual representation, see ²

Isopleth Map

• Isarithmic map or contour map is created by interpolating a set of isolines between sample points of known values.
• Created from collection of point data and then interpolating unknown values between them.
• Interpolation methods:
  – Kernel density estimation (see earlier lectures)
  – Inverse distance
  – Kriging

Proportional Symbol Mapping

- Represents numerical data associated with point locations.
- Utilizes a symbol and scales the size to show data.
- Proportional symbols can be used for two forms of point data:
  - True point data: actual measurements at point locations (number of calls made from a phone booth).
  - Conceptual point data: data collected over an area, but data is conceived as being located at points (for example, the centroid).

Coordinated Multiple Views

• Instead of trying to make the “best” visualization of all of our data, perhaps we can use multiple views
• Data can be expressed in a variety of ways
• Given the multivariate nature of data, a single statistical graphic may not be enough
• Interactive graphics systems can provide multiple representations of the data
• These representations can be linked or coordinated

Interaction Types

- Highlighting and focus
- Drill-down and hyperlinks
- Overview and context
- Changing parameters
- Changing representations
- Temporal fusion

Interaction Types

• How do we think of interacting with data?
  – Comparing and sorting
    • Provide a selection of data graphics to compare values and sort these values based on information
  – Adding/filtering variables
  – Highlighting
  – Aggregating/combining variables
  – Zoom and Pan
  – Rescale
  – Annotate
  – Bookmark

S. Few, *Now You See It*.
Interaction

• Really, the key is user intent
  – Focusing on what a user wants to achieve through a specific interaction technique

• Interaction is done by a person for a purpose
  – Seeking more information
  – Solving a problem
  – Exploratory analysis
  – Analytic discourse

* - Material borrowed from John Stasko’s lecture on Interaction: http://www.cc.gatech.edu/~stasko/7450/Notes/interaction.pdf
Interaction in Visualization

• Broken down into 7 categories
  – Select
  – Explore
  – Reconfigure
  – Encode
  – Abstract/Elaborate
  – Filter
  – Connect

Visual Analysis vs. Visual Analytics

• The choice of visual representation will influence the user’s analysis
• Massaging the data can help improve the visual representation
• It also makes some data analysis methods easier
• Adding interaction allows the user to form and explore their own hypotheses
• But, how do we find “important” features in the data?
Data Mining

- Data mining domain has techniques for examining data, looking for patterns, looking for anomalies
- Can be used to enhance the visualizations, show us what is important
- Can also be used in exploratory analysis, “I think this looks interesting, show me similar trends.”
Temporal Analysis

Methods

- For temporal data, we can find anomalies using control chart methods.
- Control charts consist of a statistic representing some measurement in time (number of patients in a hospital, value of a stock, etc.).
- The mean and standard deviation of the statistic is calculated given all the available samples.
- If the current value is greater than some pre-set number of standard deviations from the mean, then an alert is generated (Shewart suggested three standard deviations).
Finding Similarities in Time Series

- Indexing problem
  - Find all stocks whose fluctuations are similar to stock X
- Subsequence Similarity
  - Find out other days in which stock X had similar movements as today
- Clustering
  - Group regions that have similar temporal patterns
- Rule Discovery
  - Find rules like if X goes up, Y goes up, etc.
Scan Statistics\(^1\)

The idea is to exhaustively scan the search space to check for all possible statistically significant clusters.

For each case and control, the algorithm iteratively draws circles containing more and more data samples until the largest circle contains \(N/2\) samples.

Scan Statistics

- For each circle (window) compute the likelihood function
- For Bernoulli distributions, the function is:

\[
\left( \frac{c}{n} \right)^c \left( \frac{n-c}{n} \right)^{n-c} \left( \frac{C-c}{N-n} \right)^{c-c} \left( \frac{(N-n)-(C-c)}{N-n} \right)^{(N-n)-(C-c)} I() \]

- \( c \) is the number of cases in the window, \( C \) is the total number of cases in the dataset, \( n \) is the number of cases + controls in the window, \( N \) is the total number of cases + controls, \( I() \) is an indicator function that is one if the number of cases is more than expected based on the null hypothesis and zero otherwise.
Scan Statistics

- case
- control

$L_0$
Scan Statistics

- case
- control

L₀
Scan Statistics

[L_0]

L_1

- control
- case
Scan Statistics

\[ [L_0, L_1] \]

- control
- case
Scan Statistics

\[ [L_0, L_1, L_2, \ldots, L_M]_1 \]

\[ [L_5, L_2, L_{88}, L_0, \ldots, L_{M-22}] \]

\[ P_{\text{value}} = \frac{\text{Sorted position of } L_0 \text{ in list}}{M} \]

- control
- case
Scan Statistics

• Keeping the window in the same location, randomly redistribute the cases and controls
• Calculate the likelihood function for the new distribution
• Repeat 99, 999, xxx times and determine a p-value for the original distribution by sorting the likelihood functions

• Related work includes the Geographical Analysis Machine\(^1\)


* - [http://www.ccg.leeds.ac.uk/software/gam/](http://www.ccg.leeds.ac.uk/software/gam/)
Summary

• Visual Analytics combines a wide breadth of topics
  – Visual representations
  – Perception
  – Statistics
  – Human Computer Interaction
  – Machine Learning and Data Mining
  – Cognition and Many More

• This tutorial only glossed over some of the more commonly used methods and functions as an introduction
How Can Visual Analytics Help Cartography?
Path: Where Do We Want to Go?

- How can visual analytics help cartography?
- How can cartography help visual analytics?
How Can Cartography Help Visual Analytics?

- Inspiring new compact visual representations
How Can Cartography Help Visual Analytics?

• An amazing amount of information has a spatial context
• Spatialization is a natural mental model, but isn’t always natural for visualization…

George Robertson
Path: Where Do We Want to Go?

Answers and challenges
An example
Illustrative examples
Science visual analytics
Health visual analytics
Public safety and law enforcement
Visual analytic challenges and future
How Can Visual Analytics Help Cartography: Answer

• Interactive, evolving, analytics and simulation driven analytical, correlative, predictive, decision making environments

• Of course!

• The interactive human analysis through interaction is key
The Science of Interaction

• Definition: The study of methods by which humans create knowledge through the manipulation of an interface.

• …interaction and inquiry are inextricable. It is through the interactive manipulation of a visual interface – the analytic discourse – that knowledge is constructed, tested, refined and shared.”
  Stasko et al., IV2009

5 Challenges:

1. Visual Discourse - Interactive visual thinking tools for the exploration, understanding, collaboration, description, explanation, decision, dissemination, persuasion of concepts & data.

2. Multi-modal Sensemaking In The Large - Enable large-scale, distributed teams to interactively make sense of big multi-modal data and problems.

3. Fluid Interaction - Designing fluid, high-bandwidth, and powerful interaction models and paradigms for the purpose of individuals reaching their full potential in viewing, analyzing, and understanding large and complex data.

4. Collaboration – beyond space and time- synergizing technologies and human users

5. Mixed Initiative Data Discovery and Manipulation - Balancing active user input with systematic guidance to enable visual data manipulation and analysis.
Answers:

- Novel integrated, interactive visualization and analytics approaches for
  - Multisource, potentially conflicting data (more than layers)
  - Big data
  - Uncertain data
  - Multiscale data
  - Multimodal data
  - Temporal data
- Risk-based decision making
- Interactive hypothesis testing
- Integrated simulations and predictive models
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”
Big Data Is The Rage
Visual Analytics Is More Than Big Data Analytics
VA Approaches to Big Data

• Don’t display all the data!!
• Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space
• User interaction to guide above process
• Should we use big iron or cell phones?
Visual Analytics at Real-World Scale

• Utilize advanced HPC techniques to enable interactive spatiotemporal analysis (spatiotemporal clustering, prediction)
• Cluster-based and cloud-based solutions
• GPGPU solutions

• Develop easily usable HPC visual analytic environments

Example: Longhorn Exascale Visual Analytic Platform

- 2048 compute cores (Nehalem quad-core)
- 512 GPUs (128 NVIDIA Quadro Plex S4s, each containing 4 NVIDIA FX 5800s)
- 13.5 TB of distributed memory
- 210 TB global file system
Multisource, Multiscale, Evolving, Uncertain Data

• Challenges
  • Signature extraction for visual cognition, analysis, correlation
  • Common scales and alignment (e.g., temporal, spatial, aggregation level, …)
  • Temporal reasoning
  • Natural representations of temporal aspects
  • Natural representations of confidence, certainty, risk
  • Natural integration of spatial and nonspatially (scaled) data
Path: Where Do We Want to Go?

Answers and challenges
An example
Illustrative examples
Science visual analytics
Health visual analytics
Public safety and law enforcement

Visual analytic challenges and future
Example Challenging Problem - Atmospheric Science: Multi-scale Interactions (in the words of a cloud physicist)

No observing platform can measure the quantities of interest over all needed spatial and temporal scales needed.

No numerical model can simulate the quantities of interest over all needed spatial and temporal scales.

→ We observe/simulate over a subset of the pertinent scales, using different instruments/models, and must assimilate these results to understand the “big picture.”

Visual analytics is crucial for this task.

Issues: Multi-scale, multi-system, multisource, massive, data & simulations
One Solution in Use: Our Atmospheric Visual Analytic Environment

Utilize multiple rendering styles
Provide interactive data exploration and user directed analysis
Allow user specified analysis and queries on the fly
Allow interactive correlative analysis of multisource data
Aug 2012

- Cloud ICE
- Cloud Water
- Graupel
- Frozen rain
- Rain water
- Snow
- TKE
- Water vapor
No lighting

Lighting + GradMag[0..max]
Water vapor

Lighting + GradMag[0.x ..max]

No Lighting + GradMag[0.x ..max]
Path: Where Do We Want to Go?

Answers and challenges
An example
Illustrative examples
Science visual analytics
Health visual analytics
Public safety and law enforcement
Visual analytic challenges and future
Example: Public Health Visual Analytics

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

Health Surveillance

Simulation and Planning

Planning and Response

Pandemic spread on day 37 without decision measures implemented

Pandemic spread on day 37 with all decision measures implemented
Visual Analytics For Syndromic Surveillance

- Syndromic surveillance: detection of adverse health events focusing on pre-diagnosis information to improve response time
- Pre-diagnosis information can consist of multiple data sources:
  - Over the counter medicine sales
  - News reports on emerging diseases
  - Pro-med news feeds
  - Emergency department chief complaints
Time Series Modeling

- Seasonal-Trend Decomposition Based on Loess
  - Time series can be viewed as the sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

Respiratory Count

Date

Actual
Predicted
Lower
Upper
Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

• Pandemic Influenza Planning Tool
• Models user specified:
  • Pandemic influenza characteristics
  • County population, demographics, hospital beds
• Decision measures
  • Strategic National Stockpile deployment
  • School Closures
  • Media Alerts

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.
Visual Analytics Uses for Crisis Management and Response

- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

Purpose:
- Planning for resiliency
- Long-term analysis
- Predictive analytics
- Training
- Detection
- Investigation
- Response
- Recovery, remediation
Visual Analytics vs. Situational Awareness

Situational Awareness ↔ Sensemaking

situational awareness as a sensemaking process (courtesy Alan MacEachren)

Situational awareness model (after Endsley, 1995; top) compared to model of sense making (after Pirolli & Card, 2005; bottom)

State of environment

Perception of elements in current situation

Comprehension of current situation

Projection of future status

Decision

Performance of actions

External data source

Shoebox

Evidence file

Schema

Hypotheses

Presentation

Foraging loop

Sense-making loop

goals & objectives; preconceptions

Search & filter

Read & extract

Schematize

Build case

Tell story

← Search for information

← Search for relations

← Search for evidence

← Search for support

← Reevaluate
Situational Awareness, Planning, Investigation, Response Visual Analytics

- Situational Awareness and Assessment from massive data
  - In-field and desktop historical and current incident/crime information
- Innovative spatiotemporal visualization, analysis, prediction
- Image analysis tools
- Integrated statistical forecasting and simulation models
- Advanced social media analytics

- Tools for Intelligence analysis

- Planning and response support
  - Integrated simulation, resource allocation, census, weather data
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
- Interactive, multi-day, month, type investigation
- Multisource, multimedia massive data integration & analysis
Example Maritime Safety Visual Analytics Environment

- Supports decision making and risk assessment
- Interactive exploration and analysis of trends, patterns and anomalies
- Allows analysis of risks associated with
  - Closing one or more Coast Guard stations
    - Find optimal stations that absorb work load of the closing station
  - Allocating new resources
    - Impact on safety and efficiency of operation
- Currently being used by analysts at the U.S. Ninth District, Fifth District, HQ, and Atlantic Commands
Example: USCG D9 Search And Rescue Operational Analysis

- Interactive visual analytics of multivariate performance metrics for each unit's activities
- Interactive linked spatial temporal display, calendar view, and timeline views
Total Risk Based on SAR Cases
System Features: SAR Risk Profile

Time taken by CG stations to deploy an asset to the Great Lakes to respond to a SAR incident.
Response Efficiency – Current Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
D5: Cost based on Resources used for the Open-Suspended Sorties over 5 years (2006-2011)
Resource Allocation and Risk-Based Decision Making

• Explore risk-based decision making and utilize historical data for analysis and prediction
  • Total Risk, Mitigated Risk, Residual Risk
• Explore 11 different USCG missions
• Explore allocation of assets with different capabilities
• Explore staffing, utilization, assets vs. risk measures
• Perform What-If scenarios
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Interesting VA features:
• Time a first class citizen
• Temporal exploration and comparisons
• Integrated predictive models
• Interactive multivariate correlation and visual analytics
• Novel statistical predictive model incorporated for planning
• Novel interactive anomaly detection from social media

VALET delivered:
• Spring 2011: WL, Lafayette Police

iVALET delivered:
• October 2011: Purdue, WL Police
VALET Overview
Linked Predictive Crime Models by Type
Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31
VALET Feature:
Correlating Datasets at Multiple Granularities

- Correlative analysis framework for exploring correlations among multi-variate/-source data

- Determining forces that determine nature of relationships among datasets

- Detection of periodic properties among different data variables

- Allows correlations at multiple spatial and temporal granularities
Case Study: Drug Abuse Violations vs. Burglaries

- Over geospatial hotspots
VALET Feature: Real-time Twitter Monitoring and Integration

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud

Grand Prix Weekend, Purdue University
Spatiotemporal Social Media Analytics for Abnormal Event Detection

VACCINE
Correlation Between Multiple Social Media

- Occupy Wall Street event

<table>
<thead>
<tr>
<th>Date</th>
<th>Twitter (z-score)</th>
<th>Flickr (z-score)</th>
<th>Date</th>
<th>YouTube (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/17, 2011</td>
<td>2.53</td>
<td>0.97</td>
<td>9/18, 2011</td>
<td>5.98</td>
</tr>
<tr>
<td>9/30, 2011</td>
<td>3.20</td>
<td>3.50</td>
<td>10/2, 2011</td>
<td>1.38</td>
</tr>
<tr>
<td>10/5, 2011</td>
<td>3.17</td>
<td>3.36</td>
<td>10/6, 2011</td>
<td>2.75</td>
</tr>
</tbody>
</table>
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Jigsaw: Visual Analytics for Investigative Analysis and Exploration of Document Collections

**Goal:** Assist investigators with understanding, sense-making, and analysis of large, unstructured and structured document collections

**Approach:** Provide multiple visual perspectives on the documents and entities within them, highlighting connections between entities

**Users:** Over 150 downloads – active users include the following:

- Seattle PD
- Rock Hill, South Carolina PD
- Indianapolis PD
- West Lafayette, PD (student embedded)
- TSA
Gang Graffiti Recognition and Analysis Using a Mobile Telephone (GARI)

Edward J. Delp, Mireille Boutin
Collaborating Institution(s): Purdue University
End-User(s): Gary Coons, Chief/Indianapolis Department of Public Safety Division of Homeland Security and IMPD, Indiana Fusion Center Gang Task Force

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 40 users and 450 graffiti images acquired

GARI delivered:
- Summer 2011:
  - IMPD gang detectives
- August 2011:
  - IMPD at large
  - Ind Fusion Gang Task Force
- October 2011:
  - Gang detectives across Indiana
Crime Analytics & SA Issues and Techniques

- Fuse data from a variety of sources
  - Law enforcement records management
  - Weather and phases of the moon
  - Street light locations, bus routes
  - Tracking release data of offenders
  - Civil court data
  - Social Media
  - Local event calendar

- Reliable predictive models

- Understandability and trust of predictions

- Main Question: What helps officers, detectives, chief do their variety of jobs?
Directions Forward

Challenges
Six 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
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
What We Do With Visual Analytics

• Enable effective decision making through interactive visual analytic environments
• Enable effective communication of information
• Provide quantitative, reliable, reproducible evidence
• Enable user to be more effective from planning to detection to response to recovery
• Enable proactive and predictive visual analytics
• Enable effective situational awareness (perception, comprehension, projection)
What’s Needed for Proactive and Predictive Visual Analytics?

• Reliable and reproducible models and simulation
• Understanding of the data
  • Distribution and skewness, errors, appropriate analysis techniques
• Understanding of the sources and types of data
• Comparable or Correlative sources of data
  • Appropriate transformations applied to enable meaningful comparison and correlation
• Understanding of the use and problem to be solved!
About VACCINE

• What we do
  • Help people make effective decisions from massive data
  • **Big Data Visual Analytics**

• Who we can help
  • Law enforcement
  • Public safety and health
  • Scientific discovery
  • Business production, resource allocation, risk-based decision making

• Types of technologies
  • Novel visualization techniques and analytics
  • Mobile graphics, volume graphics, procedural abstraction of complex and massive data
Law Enforcement

• Law enforcement visual analytics
  • VALET & iVALET (iPhone/iPad)
    • Visual Analytics Law Enforcement Toolkit
    • Analyze crime patterns and time of day problems

• Gang activity analytics
  • GARI
    • Gang Graffiti Recognition and Interpretation using a mobile telephone
    • Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities

• Document visual analytics
  • JIGSAW
    • Visualization for investigative analysis
    • Discovery of hidden relationship and threats across documents
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
Scientific Discovery

- Flow dynamics visualization
  - Providing insights on large flow data
  - Visualization linked with simulations

- Nanohub
  - Information-assisted data analysis and visualization of nanoelectronic models

- Cancer biomarker
  - Visual analysis suite for exploring multiple samples of data from cancer care engineering
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:
Who We Are
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

• Purdue University
• Georgia Institute of Technology
• Pennsylvania State University
• Stanford University
• University of North Carolina at Charlotte
• University of Washington
• Simon Fraser University
• University of British Columbia
• Justice Institute of British Columbia
• Ontario Institute of Technology
• Dalhousie University

• University of Houston, Downtown
• Virginia Tech
• Indiana University
• Florida International University
• University of Texas at Austin
• Morgan State University
• Navajo Technical College
• University of Stuttgart
• University of Swansea, UK
• Oxford
• University of Calgary
• University of Manitoba
• Carleton University
• University of Victoria
VACCINE’s Value

Our Value / Solution: 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:

Some of Our Projects
Example Projects

- Health Applications
- TSA
- Public Safety and Law Enforcement
- USCG
- Cybersecurity
Visual Analytics for Syndromic Surveillance: Hypothesis Generation and Exploration

Predictive Visual Analytics

- Time Series Modeling:
  - Seasonal-Trend Decomposition Based on Loess
    - Time series ~ sum of multiple trend components
    - For each data signal, components are extracted
    - Can then analyze correlation between components
  - Predictive Visual Analytics using STL
Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

• Pandemic Influenza Planning Tool
  • Models user specified:
    • Pandemic influenza characteristics
    • County population, demographics, hospital beds

• Decision measures
  • Strategic National Stockpile deployment
  • School Closures
  • Media Alerts

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.
TSA Projects

• Visual Analytics for multifaceted operational analysis
  • Analysis and correlative analysis of
    • Operational performance (throughput per lane, use of AIT, targets)
    • Security performance (incidents)
    • Financial performance (expenses vs. performance metrics)
  • Spatial, temporal, spoke, hub, and regional analysis

• Customer satisfaction visual analytics
  • Jigsaw use to understand and find patterns in compliment/complaint database
  • Potential linkage to above data

• MERGE for rail inspectors
TSA Visual Analysis Prototype
Public Safety & Law Enforcement

- Law enforcement visual analytics
  - VALET, iVALET (iPhone/iPad), CrimeViz
    - Visual Analytics Law Enforcement Toolkit
    - Analyzing crime patterns and time of day problems
- Gang activity analytics
  - GARI
    - Gang Graffiti Recognition and Interpretation using a mobile telephone
    - Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
- Document visual analytics
  - JIGSAW
    - Visualization for investigative analysis
    - Discovery of hidden relationship and threats across documents
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview

Time Series View

Clock View

Twitter monitoring

Map View

Calendar View

Time Slider

Menus
Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame
  PU Lost: 10-38

- Homecoming (Sat.)
  PU vs. Illinois
  PU Won: 21-14

- PU vs. Iowa
  PU Lost: 21-31

Day-of-the-Week

- Home
- Away
Social Media: Real-time Twitter Monitoring and Integration into Tools
(Purdue, Stuttgart, Penn St.)

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis
We Have an App For That!

Our Mobile Tools

• VALET
• Evacuation Planning
• Rosetta Phone
• Hazmat app
• Gang graffiti app
• Tatoo app
Visual Analytics Uses for Risk-Based Decision Making

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Interactive, multi-day, month, type investigation
• Multisource, multimedia data integration & analysis
USCG: 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
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
Risk-Based Allocations

- Comparative visual analysis of mission cases/hours vs. staffing hours
- Comparative visualization of resources vs. risk
- Trend visual analytics
  - Increase/decrease in resource allocation
  - Increase/decrease in risk (total, mitigated, residual)
  - Increase/decrease in incidents
- Exploration of alternatives and effect on risk
- Predictive analytics based on historical data (STL and EWMA)
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
- Hurricane Irene resource allocation decision based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony and recommendations to Congress
- Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Jones Beach, Sandy Hook and Manasquan Inlet

Closest Cases (not necessarily reported under them)

All Cases November 2011
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Example Work in Cybersecurity

- Corporate Insider Threat Detection (Oxford, Leicester, Cardiff)
- Sensor Forensics (Purdue)
- SemanticPrism (Purdue)
- Multiscreen, Multiview Interactive Cyber Investigation (VaTech, PNNL)
- Log Visualization (Purdue)
Overview of VACCINE Projects
COAST
USCG Risk-Based Resource Allocation for Coastal Mission
Built on Coast Guard Search and Rescue Visual Analytics

VACCINE
May, 2012
System Overview
Capabilities and advantages

• Geo-temporal visualization of incidents on a map
• Calendar view of incidents
• Word Cloud of Incident Types
• Filter data by distress type, unit (stations), initiation types, cases status, and COTP owner.
• Heatmaps for: incidents, risk numbers, weighted incidents and cost.
• Risk assessment capabilities for resource allocation and station closures
• Boat and air assets
Recent Updates

• Addition of detailed filters for case initiation and case status, along with their respective heat maps.
• Expansion to D1 and D5 districts.
Recent Updates

- Addition of custom response rings
- Addition of redundancy areas in coverage
- New station utility metrics
System Overview

1. State selector
2. Data browse button
3. Data list view
4. Individual color selector
5. Color scheme selector
6. Visualization options
7. Refresh button
8. Table view
9. Map view
Capabilities and Advantages

- Geocoding
- Routes
- Reachable areas
- Linked table and map views
- Temporal cues
- Various map types
- Various visualization items
- Multiple data comparison
- Selective visualization
Multiple Data Comparison

- Example of 5 individual data

- Living in same or close county
  - Person02 (Green) and Person04 (Red)
  - Person03 (Blue) and Person05 (Pink)

- Shared reachable areas
  - Person02 (Green) and Person04 (Red)

- Recent movement into Indianapolis
  - Person01 (Yellow)
Crime Analytics: VALET – Features and Benefits

• Analysis of spatiotemporal law enforcement data
• Criminal data exploration and analysis
• Linked statistical and geographic visualizations
• Comparison of datasets for potential correlations
• Detection of patterns at different granularity levels
• Predictive analysis
• Resource allocation strategy development
• Twitter integration for real-time data monitoring
System – Overview

Map View

Time Series View

Clock View

Twitter monitoring

Calendar View

Menues

Time Slider

Map View
VALET Feature Highlight: Correlating Datasets at Multiple Granularities

• Correlative analysis framework for exploring correlations among multi-variate/-source data

• Determining forces that determine nature of relationships among datasets

• Detection of periodic properties among different data variables

• Allows correlations at multiple spatial and temporal granularities
Comparing Drug Abuse Violations & Burglaries: Within and Across Neighborhoods

- Tippecanoe County, IN
Drunkenness/public intoxication and noise complaints vs. traffic accidents

Drunkenness/public intoxication/noise complaints and traffic accidents repeat periodically every 7 days
VALET Feature Highlight: Real-time Twitter Monitoring and Integration

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud

Grand Prix Weekend, Purdue University
iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Linked views to explore the multivariate spatiotemporal dataset
iVALET – Risk profile – Where & When

- Use current spatial & temporal context of user into situational assessment and risk analysis
• Display a path between 2 (user-specified) points to show risks along intended route
iVALET – ALOHA plume model integration

• The chemical release plume model identifies census tracts with the highest number of expected people affected.
iVALET

- Works on iPhone/iPad
- Readily downloadable at iVALET website
Jigsaw

Visualization for Investigative Analysis across Document Collections

- Law enforcement & intelligence community
- Fraud (finance, accounting, banking)
- Academic research
- Journalism & reporting
- Consumer research

“Putting the pieces together”
Problem Addressed

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

- Articles & reports
- Spreadsheets
- XML documents
- Blogs
Example Documents

1) Report Date: 12 July, 2004. A routine customs inspection was performed on a package that was sent by a person named Pieter Dopple, 22 Hoveniersstraat, Antwerp, Belgium. This package was addressed to A. Hijazi, 1212 Lyons Ave, Newark, NJ. The customs form stated that this package contained two decorative clocks having a commercial value of $250.00. The package instead contained 111 polished diamonds, whose value is estimated to be $47,000. Discussions with the FBI resulted in a decision not to pursue a customs violation charge against Hijazi. The reason is that Hijazi is currently under FBI surveillance. This package was resealed and delivered to Hijazi by USPS.

2) Report Date: 22 February, 2003. Surveillance report on Cesar Arze, whose residence is 77 Avenue Francis, Santo Domingo, Dominican Republic. Arze, who moved from Havana, Cuba to Santo Domingo in 1992, works as a medical technician in Santo Domingo. Arze is under surveillance because of information that he is associated with Cuban intelligence services. Arze was photographed in company with a man identified as Hector Lopez in Bogota, Columbia on 23 January, 2003. Lopez, a known representative of FARC, has conducted narcotics distribution activities throughout South and Central America and the Caribbean.
Analytic Focus

• Entities within the documents
  – Person, place, organization, phone number, date, license plate, etc.

• Thesis: A story/narrative/plot/threat within the documents will involve a set of entities in coordination
VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.
Entities Identified

Source:
Date: May 14, 2004

VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.
Connections

• Entities relate/connect to each other to make a larger “story”

• Connection definition:
  – Two entities are connected if they appear in a document together
  – The more documents they appear in together, the stronger the connection
Jigsaw

• Computational analysis of document text
  – Entity identification, document similarity, clustering, summarization, sentiment

• Multiple visualizations (views) of documents, analysis results, entities and their connections

• Views are highly interactive and coordinated
System Views
Console

Entity types

Color Legend:
- author (1017)
- conference (2)
- journal (17)
- year (16)
- concept (77)
- indexterm (1790)
- keyword (1202)

Workspace: no active workspace
List View

Lists of entities by type
Connections highlighted
WordTree View

Context of a word in the collection

110 matches

interaction

and

techniques

with

displayed nodes

systems were made

and

the

a 2D modified scatterplot computed from two different metrics calculated over the elements of a network

information from text documents. The paper describes an approach to N that involves visualizing text content for enhanced visual browsing and analysis.

by digging down through many successive layers.

It is hard to provide visual feedback at interactive rates for datasets containing millions of entries.

temporal flexibility, spatial flexibility, and changing collaboration styles.

and through tests with security experts, several ameliorations over the standard techniques have been provided.

financial visual analytics has been widely accepted that interactive visualization techniques enable users to more effectively form hypotheses and identify areas requiring both operationally and strategically.

an important part of information visualization (InfoVis), it has garnered a relatively low level of attention from the InfoVis community.

between visualizations, system and tool type were found

for dimensionality manipulation.

tools has been developed.

are not immediately clear.

Costs in information visualization is an important but poorly understood factor in visualization design.

DataSpace temporal interactions (or flows), such as population migration and disease spread, naturally form a weighted location-to-location network (graph).

in complex social networks is a critical component of visual tools for intelligence analysis, consumer behavior analysis, and human geography.

on a desktop computer.

technique for expanding or collapsing subtrees to any depth with a single mouse drag.

environment by providing direct visual and algorithmic support for the coordination of data analysis actions over shared large displays.

for graphs with visual nodes. Graph and tree visualization techniques enable interactive exploration of complex relations while communicating topology.

techniques for manipulating visualizations of networks by selecting subgraphs and then applying various commands to modify their layout and topological properties.

capabilities in SelfTrend provides an innovative approach to this problem and to other similar types of multivariate, temporally driven transactional databases.

context

Showing 36 / 100 (36.00%) leaves in current branch, of 100 leaves in tree. Prune Percentage: 36

information interfaces
Document Cluster View

Clustered by document text or by entities

Summarized by three words
Document Grid View

User controls order and color

Sentiment analysis shown here
Calendar View

Showing connections between entities and dates.
To Learn More about Jigsaw & Availability

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

Available for (free) trial use

Send email to: stasko@cc.gatech.edu
Spatiotemporal Visualization and Analysis of Gang-related Criminal Activity Using Mobile Imaging

David S. Ebert
Purdue University
Outline

• Background or “Why am I here?”
  • What is VACCINE?
• Public health visual analytics examples
• Extending syndromic surveillance to law enforcement:
  • Crime analytics
  • Mobile gang graffiti app
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
VACCINE Benefit

Enable users to be more effective through innovative interactive visualization, analysis, and decision making tools

• Provide the right information, in the right format within the right time to solve the problem
• Turn data deluge into a pool of relevant, actionable knowledge
• Enable user to be more effective from planning to detection to response to recovery
• Enable effective communication of information

Approach: Partner-driven solutions and research
Visual Analytics for Syndromic Surveillance: Hypothesis Generation and Exploration

Predictive Visual Analytics

- **Time Series Modeling:**
  - Seasonal-Trend Decomposition Based on Loess
  - Time series \( \sim \) sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components

- **Predictive Visual Analytics using STL**
Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

- Pandemic Influenza Planning Tool
- Models user specified:
  - Pandemic influenza characteristics
  - County population, demographics, hospital beds
- Decision measures
  - Strategic National Stockpile deployment
  - School Closures
  - Media Alerts
Visual Analytics Uses for Law Enforcement

- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

**Purpose:**
- Planning for resiliency
- Long-term analysis
- Predictive analytics
- Training
- Detection
- Investigation
- Response
- Recovery, remediation
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (September 2012) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview

Map View

Time Series View

Calendar View

Time Slider

Twitter monitoring

Clock View

Menus

Map View
Linked Predictive Crime Models by Type
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame
  - PU Lost: 10-38
- Homecoming (Sat.)
- PU vs. Illinois
  - PU Won: 21-14
- PU vs. Iowa
  - PU Lost: 21-31

Home vs. Away

- Day-of-the-Week
- Football season
VALET Feature: Real-time Twitter Monitoring and Integration

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Gang Graffiti Recognition and Analysis (GARI)
PI: Ed Delp
There’s An App For That

- What is GARI?
  - A mobile device application that analyzes gang graffiti
- How does GARI work?
  - User takes an image of the graffiti
  - User receives an analysis
  - Geographic locations
  - Graffiti colors, shapes, meaning
  - Database search of similar tags
Why It Works

- Catalogs and analyzes graffiti images
  - More than 500 graffiti images acquired (plus 506 for research testing)
- Search of related image
- Search of local area
- Partially interprets graffiti images
  - Colors, symbols
  - Recorded gang analyst notes
- Future analysis
  - Image meaning
  - Which gang
- Helps law enforcement track and identify gang activity throughout a region
GARI On The Street

- Over 50 law enforcement officials are using or field testing GARI
- Delivered to:
  - IMPD gang detectives (Summer 2011)
  - IMPD at large, Indiana Fusion Gang Task Force (August 2011)
  - Gang detectives across Indiana (October 2011)
- Homeland Security applications
GARI System Overview

1. Offline automatic analysis and labeling
   - Geoposition
   - Date and time
   - Extracted Features

2. Filtered results
   Info + thumbnails

3. Original Database

4. Server

5. Offline manual filtering

6. Manual labeling
   - Additional Features

7. Addition to Database
Project Status

• Deployed on Android Phone
• Main phone application is image database system and browsing (acquisition tool) with searching and some limited analysis capabilities
• Desktop backend
  • Browse by radius
  • Upload images
  • View and edit image details
  • Interact with map
  • Searching for similar images
Gang Graffiti Interpretation

SHAPE
Simple, Straightforward

NUMBERS
42nd street gang

SYMBOLS
6-point star, pitchforks

COLOR
Goon Squad: Red/Black
Gang Graffiti Interpretation

LETTERS
East Side Gang

POSITION/ALIGNMENT
Letters at star points
Numbers in the middle
Letters at the bottom
Pitchforks upright

TIME
Black: 18 ST (18th Street Gang)
Red: 13 SUR (Sureños 13)
Image Analysis – Color Recognition

- MEXICANOS MALDITOS SUREÑOS 13
- 18 STREET GANG
- SUREÑOS 13
Image Analysis – Scene Analysis

GANG RIVALRY TRACKING

- Date: 08/19/2010
- Time: 3.25 PM
- Geo: 41.387917, 2.169919

- Date: 01/03/2011
- Time: 5.11 PM
- Geo: 41.387917, 2.169919
GARI App Features
Gang Graffiti – User Interface

Send to server

Analysis

Capture Image  Browse Database  Browse Image

About  Settings

Main menu

5 Point Star: Symbolic to gangs within the People Nation

October 2012
Gang Graffiti – Browse Database

- Browse by radius from current position
- Download images and information from server (image EXIF tags, gang related information)
  - Compare results
  - Track graffiti
- Browse map (Google Maps API)
- Network connection required
Gang Graffiti – Browse Database

- Browse general results
- Show graffiti on map
- Inspect specific graffiti
GARI Progress and Directions

- Detective Steve Shafer, IPD adding annotations to database
- Shape and tag analysis always improving
- Transitioning to Indiana State Police / IIFC Gang Task Force
- Creating linked, satellite database and tool for Navajo Nation (using Navajo Technical College)
- Investigating use by ICE, CBP, FBI
- Two-way linkage with VALET / iVALET
How to Get The App?

- Contact gari@ecn.purdue.edu
  - We assign user IDs and initial passwords
  - Database server
For Further Information

www.VisualAnalytics-CCI.org

cvaccine@purdue.edu
ebertd@purdue.edu
Teaching Visual Analytics: Leveraging Multidisciplinarity
Definition

Visual Analytics\(^1\) is the science of analytical reasoning facilitated by interactive visual interfaces.

People use visual analytics tools and techniques to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, understandable assessments
- Communicate assessment effectively for action

Motivation

To solve today’s and tomorrow’s problems requires exploring, analyzing, and reasoning with massive, multisource, multiscale, heterogeneous, streaming data.
VA Goals – Solution Driven

- Enable effective decision making through interactive visual analytic environments
- Enable effective communication of information
- Provide quantitative, reliable, reproducible evidence
- Enable user to be more effective from planning to detection to response to recovery
- Enable proactive and predictive visual analytics
- Enable effective situational awareness (perception, comprehension, projection)
Foundations of Visual Analytics

- Statistical and Information Analytics
- Geospatial Analytics
- Scientific Computing
- Graphics and Visualization
- Decision Sciences
- Knowledge Discovery
- Data Management & Knowledge
- Design and Communication
- Cognitive & Perceptual Sciences
- Interaction Science
## Course Outline

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>Analytical exercise</td>
</tr>
<tr>
<td>2–3</td>
<td>Analytical reasoning</td>
<td>The analysis process, critical thinking, sensemaking, and situation awareness</td>
</tr>
<tr>
<td>4</td>
<td>Perception</td>
<td>Human perception, preattentiveness, color, shape, and texture</td>
</tr>
<tr>
<td>5</td>
<td>Cognition</td>
<td>Cognitive theory</td>
</tr>
<tr>
<td>6–7</td>
<td>Data management</td>
<td>Representations, transformations, and statistics (temporal and spatial)</td>
</tr>
<tr>
<td>9</td>
<td>Visual representations</td>
<td>Visualization techniques</td>
</tr>
<tr>
<td>11</td>
<td>Interaction</td>
<td>Interaction techniques</td>
</tr>
<tr>
<td>12</td>
<td>Communication</td>
<td>Production, presentation, and dissemination</td>
</tr>
<tr>
<td>13</td>
<td>Collaboration</td>
<td>Collaborative VA</td>
</tr>
<tr>
<td>14</td>
<td>Evaluation</td>
<td>Evaluating VA</td>
</tr>
<tr>
<td>15</td>
<td>Advanced topics</td>
<td>Conducting VA research, novel computing platforms, and mobile VA</td>
</tr>
</tbody>
</table>
Jigsaw: Visual Analytics for Investigative Analysis and Exploration of Document Collections

Goal: Assist investigators with understanding, sense-making, and analysis of large, unstructured and structured document collections

Approach: Provide multiple visual perspectives on the documents and entities within them, highlighting connections between entities
# Course Outline

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>Analytical exercise</td>
</tr>
<tr>
<td>2–3</td>
<td>Analytical reasoning</td>
<td>The analysis process, critical thinking, sensemaking, and situation awareness</td>
</tr>
<tr>
<td>4</td>
<td>Perception</td>
<td>Human perception, preattentiveness, color, shape, and texture</td>
</tr>
<tr>
<td>5</td>
<td>Cognition</td>
<td>Cognitive theory</td>
</tr>
<tr>
<td>6–7</td>
<td>Data management</td>
<td>Representations, transformations, and statistics (temporal and spatial)</td>
</tr>
<tr>
<td>9</td>
<td>Visual representations</td>
<td>Visualization techniques</td>
</tr>
<tr>
<td>11</td>
<td>Interaction</td>
<td>Interaction techniques</td>
</tr>
<tr>
<td>12</td>
<td>Communication</td>
<td>Production, presentation, and dissemination</td>
</tr>
<tr>
<td>13</td>
<td>Collaboration</td>
<td>Collaborative VA</td>
</tr>
<tr>
<td>14</td>
<td>Evaluation</td>
<td>Evaluating VA</td>
</tr>
<tr>
<td>15</td>
<td>Advanced topics</td>
<td>Conducting VA research, novel computing platforms, and mobile VA</td>
</tr>
</tbody>
</table>
VA Course Challenges

1. Broad field to cover in 15 weeks
2. Students with diverse backgrounds
3. Teacher expertise in normally only portion of course topics
To Program or Not to Program…. 

- We design course without programming expertise required
- Challenges:
  - Evaluating students results fairly across projects of different types (design studies, application use, new techniques)
  - Overcoming faculty biases
Current Results

• Graduate class offered twice with 10-15 students per semester
• Variety of student backgrounds
• At least 5 projects have led to conference submissions
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
iVALET

- Linked views to explore multivariate spatiotemporal dataset
- Analytical tools to help explore data
VA Course Challenges

1. Broad field to cover in 15 weeks
2. Students with diverse backgrounds
3. Teacher expertise in normally only portion of course topics
For Further Information

ebertd@purdue.edu
elm@purdue.edu
Transitioning GARI to Law Enforcement and ICE

Dr. David S. Ebert
Director, Visualization Sciences Co-Lead (VACCINE)
CVADA
Purdue University
GARI - Gang Graffiti Automatic Recognition and Interpretation

Enabling tracking, analysis, and interpretation of gang graffiti / activity

- **Problem:** How to use gang graffiti for improved policing and safety
  - Complex language difficult to interpret
  - Encodes threats, information, gang activity
  - Most officers aren’t trained to interpret it

- **Solution:** A mobile device application that analyzes gang graffiti
  - User takes a picture and receives some analysis
  - Finds similar images, locations, and officer notes on the images

- **Key Users:** Indianapolis PD, Indiana Gang Intelligence Network (INGANG), Indiana State Police, Indiana Intelligence Fusion Center, Illinois State Police (in process), Cook County, IL Sherriff’s department (in process)
When and Where:

- Summer 2012 – Indianapolis - Pilot – deployed to gang detectives IPD, lead detective Steve Shafer
  - “GARI can really help the street officers because it's available 24/7--it's quick and instantaneous” - Detective Steve Schafer, Criminal Gang Unit, Indianapolis Police
- Share gang expertise / force multiplier
- Fall 2012 – Indiana – deployed to multiple agencies across Indiana – hundreds of images uploaded and analyzed
  - Over 800 images, 72 users
- Winter 2013 – INGANG deployment – all Indiana law enforcement agencies – server will be transferred to Indiana State Police (April 2013)
  - Adding prison tattoos to linked database for Indiana Department of Corrections

- Spring 2013 - Illinois State Police
  - “This does more than the RFP we are about to issue for software to capture images,” “Sign us up now,” Aaron Kustermann, ISP Chief of Intelligence

- Spring 2013 – Cook County, IL
  - Will upload thousands of images
  - Adding capability to record graffiti removal, background type (e.g., garage door, picket fence, brick)
How: Transitioning to Sustainable Deployment

- **Key:** Close interaction with officers from the start – Purdue police, Tippecanoe police departments, Indiana police
- Tight Integration with Indiana State Police INGANG network (Capt. Scott Beamon)
- Server transitioning to ISP
- ISP will vet law enforcement officers nationwide who want GARI and agree to their INGANG terms of use
- Midwest regional deployment is underway with Ohio State Highway Patrol, Ohio DHS, and Illinois partners

GARI app interface, showing multiple tags
Transition Cycle to Date

- Question from Chief John Cox at monthly public safety consortium meeting – “Can the Rosetta Phone be changed to help with this picture I just took of graffiti in the parking garage?”
- Prototype with a handful of officers and 1 gang graffiti expert
- Statewide deployment
- Midwest regional deployment
VACCINE

A DHS S&T Center of Excellence

USCG Innovation Showcase

June 2012

Dr. David Ebert
VACCINE, Director
VACCINE Mission Statement
Create methods and tools to analyze and manage vast amounts of information for all mission areas of homeland security.

Mission Relevance to USCG
VACCINE develops tools to enable more effective and efficient decision making both tactically, operationally, and strategically.

We have developed a visual analytic tool allowing USCG to analyze and understand their SAR, LE, etc. data for D9, D5 and D1.

Enabling formulation of various strategies to balance operational staffing levels with the risk/reward impact on response time, potential lives saved and property lost.
## Principal Partners

<table>
<thead>
<tr>
<th>University</th>
<th>Areas of Expertise/Core Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purdue University</td>
<td>Visualization of Structured, Unstructured and Streaming Data; Scalable Filtering and Dissemination; Video and Image Processing; Mobile, Lightweight Information Analytics; Public Safety Operations; Undergraduate/Graduate Education; Professional Training</td>
</tr>
<tr>
<td>Georgia Institute of Technology</td>
<td>Information Visualization; Visual Analytics; Peripheral Awareness; Undergraduate/Graduate Education;</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>Spatial cognition; Geo-information Representation, Spatial Analysis, Cartography; Undergraduate/Graduate Education; Professional Training</td>
</tr>
<tr>
<td>University of NC-Charlotte</td>
<td>3D Multimodal Interaction; Bioinformatics Visualization; Virtual Environments; Visual Reasoning; Interactive Visualization of Large-scale Information Spaces; Undergraduate/Graduate Education; Professional Training</td>
</tr>
</tbody>
</table>

## Extended Partner Network

Florida International University*; University of Houston-Downtown*; University of Texas, Austin; California State University Dominguez Hills*; Morgan State University*; Virginia Polytechnic Institute and State University; Navajo Technical College*; Simon Fraser University, Canada; University of British Columbia, Canada; Oxford University, UK; Swansea University, UK; Stuttgart University, Germany; Stanford University; University of Washington; Tennessee State University; Middlesex University, UK; Dalhousie University, Canada; University of Manitoba; University of Victoria; Indiana University; Jackson State University*; University of Calgary, Canada; Carleton University; Justice Institute of British Columbia, Canada; Ontario Institute of Technology, Canada.

* Indicates MSI
# VACCINE Research Areas

<table>
<thead>
<tr>
<th>Research Area</th>
<th>Approaches</th>
<th>Expected Uses</th>
<th>Customers</th>
<th>COE Partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Indiana and Ohio State Law Enforcement and Public Safety</td>
<td>- Integrate and improve public safety’s analysis of data</td>
<td>- Large-scale field tests on technologies including VALET, TRIP, iVALET, MERGE and GARI.</td>
<td>- Law Enforcement, Emergency Management</td>
<td>Seed Proposals to be considered from current and extended partners that meet objectives of this E2E project.</td>
</tr>
<tr>
<td></td>
<td>- Promote improved operations in the law enforcement, fire, emergency management and EMS fields</td>
<td>- Uses include intelligence analysis, social media fusion, asset and resource management, hazmat operations and investigations.</td>
<td>- TSA Rail Inspectors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Community Watch</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Operations and Analysis Visual Analytics Tools: Tactical and Strategic Decision Making</td>
<td>- Phase I will be the extension and adaption of cgSARVA to meet the needs of the USCG at the tactical and strategic level, including their proposed Coastal Operations and Analysis Suite of Tools (COAST).</td>
<td>- To inform senior CG leaders’ decisions regarding asset capabilities, acquisitions, and allocations; unit locations; policies and concepts of operations; and mission tradeoffs.</td>
<td>- USCG Headquarters and District Operations and Planning Personnel</td>
<td>Seed Proposals to be considered from current and extended partners that meet objectives of this E2E project.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- COAST is intended to measure the operational effectiveness of all 11 Coast Guard missions.</td>
<td>- D1, D5, D9</td>
<td></td>
</tr>
<tr>
<td>3. TSA Airport Database Evaluation Visualization</td>
<td>- Analysis, assessment and visualization of TSA databases across airport environment</td>
<td>- Phase I of this project will provide TSA with an understanding of the database linkages, redundancies, efficiencies and inefficiencies across their airport environment.</td>
<td>- TSA</td>
<td>Seed Proposals to be considered from current and extended partners that meet objectives of this E2E project.</td>
</tr>
</tbody>
</table>
cgSARVA/Visual Analytics

- Coast Guard Search and Rescue Visual Analytics - Presents a risk analysis system that enables the interactive visualization, analysis, and assessment of SAR and other missions completed by each Coast Guard station. Through use of advanced GIS visualization and statistical research, cgSARVA provides a spatial and temporal overview of the USCG operations as well as a comprehensive maritime risk assessment and analysis on a strategic and operational level.
- Leveraged for COAST project

MERGE/Visual Analytics

- Mobile Emergency Response GuidE - The MERGE system is an electronic version of the Emergency Response Guidebook with many new features and capabilities. These new capabilities include the use of image analysis methods to automatically determine the type of Hazardous Materials present based on an image taken of the sign/placard.
Visual Analytic Applications for Law Enforcement

David S. Ebert
Edward Delp

www.VisualAnalytics-CCI.org
Who We Are: International Team of Experts - 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
What We Offer

- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community

- **Improved Effectiveness:** We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information

- **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students
  - Actively Engaged Partners – We define and extend the new science of visual analytics driven by real-world, real-scale problems of engaged partners (local, state, federal)

“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 Uses for Crisis Management and Response

- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

Purpose:
- Planning for resiliency
- Long-term analysis
- Predictive analytics
- Training
- Detection
- Investigation
- Response
- Recovery, remediation
Visual Analytics Uses for Public Safety

• Risk visualization and analysis
• Predictive analytics
• Uncertain decision making
• Alternative evaluation and consequence investigation
• Trend analysis, clustering, anomaly detection
• Interactive and multi-day investigations
• Multisource, multimedia massive data integration & analysis
• Visual standardization and interoperability
Visual Analytics Uses for Law Enforcement

- Situational Awareness and Assessment from massive data
  - In-field and desktop historical and current incident/crime information
- Innovative spatiotemporal visualization, analysis, prediction
- Image analysis tools
- Integrated statistical forecasting and simulation models
- Advanced social media analytics

- Tools for intelligence analysis
- Planning and response support
  - Integrated simulation, resource allocation, census, weather data
Current Public Safety Projects: Law Enforcement

Desktop crime analytics and intelligence analytics

- Jigsaw
  - Delivered to West Lafayette PD, IMPD 09/2010, TSA 06/2012
- TRIP
  - Delivered to Indiana Intelligence Fusion Center May 2012
- VALET & CrimeViz
  - Delivered to Lafayette PD 2011, WL PD Spring 2012
  - Delivered to Harrisburg Bureau of Police Spring 2011

In-field crime analytics

- Gang Graffiti Mobile Application (Gari)
  - Delivered to IPD 09/2011, in use by 15 different agencies in Indiana
- iValet
  - Delivered to WLPD, LPD, Purdue PD Fall 2011
TRIP: Travel Response Investigative Profiler

- Understanding movement behavior and spatiotemporal patterns of individuals
- Integrated spatiotemporal visualization, exploration and analysis of multiple individuals’ movement history
- Geocoding
- Routes
- Reachable areas
- Temporal cues
- Various visualization items
- Support various address formats
- International address visualization
System Overview

1. State selector
2. Data browse button
3. Data list view
4. Individual color selector
5. Infrastructure data list view
6. Color scheme selector
7. Visualization options
8. Refresh & Save PDF
9. Table view
Various Visualization Items

- Locations
- Oldest / newest locations
- County
- Routes
- Reachable areas
- Infrastructures
Routes

- Driving routes
- Route thickness
- Temporal cue from the oldest to the newest
Reachable Areas

- Polygonal geofences
- Along the routes
- Given a distance that can be traveled.
Visual Analytics Law Enforcement Toolkit (i)VALET
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview
Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

Football season

- PU vs. Notre Dame
  PU Lost: 10-38
- Homecoming (Sat.)
- PU vs. Illinois
  PU Won: 21-14
- PU vs. Iowa
  PU Lost: 21-31

Day-of-the-Week

PU vs. Notre Dame
PU Lost: 10-38
Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14
PU vs. Iowa
PU Lost: 21-31
Social Media: Real-time Twitter Monitoring and Integration into Tools
(Purdue, Stuttgart, Penn St.)

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
Top 10 Hot Incidents

- Identify unusual localized high-frequency patterns of crimes (near repeats)
- Each data entry is checked for other crimes with similar properties within a 1 block radius of the incident location and a 14-day time period;
- Top 10 incidents with the most number of related incidents in this space-time window are highlighted
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 75 users and 800 graffiti images acquired

GARI delivered:
- Summer 2011: IPD gang detectives
- August 2011: IPD at large, Ind Fusion Gang Task Force
- September 2011: Gang detectives across Indiana
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Public Safety & Law Enforcement

- Law enforcement visual analytics
  - VALET, iVALET (iPhone/iPad), CrimeViz
    - Visual Analytics Law Enforcement Toolkit
    - Analyzing crime patterns and time of day problems
- Gang activity analytics
  - GARI
    - Gang Graffiti Recognition and Interpretation using a mobile telephone
    - Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
- Document visual analytics
  - JIGSAW
    - Visualization for investigative analysis
    - Discovery of hidden relationship and threats across documents
Big Data Is The Rage
Big Data Challenges

• Heterogeneity
• Unprecedented volume and/or complexity
• Multivariate, Multidimensional
• Streaming
• Qualitative, Quantitative, Categorical, Temporal,…
• Distributed Sources: sensors, text, images, maps, …
• Inconsistent and Uncertain
• Different Scales and Cross Scale
The Curse of Dimensionality (Bellman 1961)

- Problems associated with multivariate data analysis as the dimensionality increases the available data becomes sparse.
- Sparsity is a problem for any method that requires statistical significance.
- In visualization, we are also limited with screen space and the number of available visual variables, so choosing the most appropriate dimensions is key.
Big Data Isn’t Just Large Data!

- All of the above characteristics lead to big, complex data
- Example 1: A few million crime records with uncertain, missing data, dozens of fields, associated risks, related social media data, images, …
- Example 2: Basic question of understanding why and when a cumulus cloud will produce precipitation…
Example Challenging Problem - Atmospheric Science: Multi-scale Interactions (in the words of a cloud physicist)

No observing platform can measure the quantities of interest over all needed spatial and temporal scales needed.

No numerical model can simulate the quantities of interest over all needed spatial and temporal scales.

→ We observe/simulate over a subset of the pertinent scales, using different instruments/models, and must assimilate these results to understand the “big picture.”

Visual analytics is crucial for this task.

Issues: Multi-scale, multi-system, multisource, massive, data & simulations.
Data Overload

- How can we make sense of this data?
- How can we harness this data in the decision making process?
- How do we avoid being overwhelmed by all of this?
- How to turn data into actionable knowledge?
The Power of Visualization

• The goal is to take all of this *data* and transform it into *information and knowledge*

• How many terabytes of data we have collected doesn’t matter, it’s how many petaflops of *insights* we can generate from this data

• “The purpose of computing is insight, not numbers.” RW Hamming (1971)

• We need to make the data understandable to people and a key way of doing this is through *visualization*
VA Approaches to Big Data

• Don’t display all the data!!
• Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space
• User interaction to guide above process
• Should we use big iron or cell phones?
Six Challenges for Proactive & Predictive Big Data 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
Path: Where Do We Want to Go?

• Solving these challenges will enable dramatic changes in information-driven risk-based decision making, discoveries, science, and engineering!
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Path: Where Do We Want to Go?

Answers and challenges
An example
Illustrative examples
Science visual analytics
Health visual analytics
Public safety and law enforcement
Visual analytic challenges and future
What We Do With Visual Analytics

• Enable effective decision making through interactive visual analytic environments
• Enable effective communication of information
• Provide quantitative, reliable, reproducible evidence
• Enable user to be more effective from planning to detection to response to recovery
• Enable proactive and predictive visual analytics
• Enable effective situational awareness (perception, comprehension, projection)
What’s Needed for Proactive and Predictive Visual Analytics?

- Reliable and reproducible models and simulation
- Understanding of the data
  - Distribution and skewness, errors, appropriate analysis techniques
- Understanding of the sources and types of data
- Comparable or Correlative sources of data
  - Appropriate transformations applied to enable meaningful comparison and correlation
- Understanding of the use and problem to be solved!
Visual Analytics: From Situational Awareness to Risk-Based Decision Making

David S. Ebert
ebertd@purdue.edu
Who We Are: International Team of Experts - 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some VACCINE Regional and Corporate Partners

- Indiana Department of Homeland Security
- Indiana State Department of Health
- Indianapolis Public Safety
- Charlotte Mecklenburg Police Department
- Lafayette, West Lafayette, Purdue Police Departments
- Tippecanoe County EMA & Sherriff’s Department
- Harrisburg, PA Police
- Coast Guard Sector Boston, Seattle, LA
- Joint Harbor Operations Command Center
- Port of Seattle
- Indianapolis Public Safety
- USCG District 9, D5, D1
- Motorola
- Harris Corporation
- Boeing
- Next Wave Systems, LLC
- Banfield, The Pet Hospital
- Raytheon
- MacDonald, Dettwiler and Associates
- Oculus Info Inc.
- Kx Systems
- Bank of America
- Duke Energy
- World Vision International
- Gates Foundation
- Kimberly Clark
- Hallmark
Some VACCINE Government Partners

- U.S. Coast Guard
- Federal Emergency Management Agency
- Customs and Border Patrol
- National Geospatial Intelligence Agency
- National Science Foundation
- Army Research Office
- Department of Defense
- Department of Health and Human Services
- U.S. Dept. of State
- National Institute of Justice
- National Maritime Intelligence Center
- Defence Research & Development Canada
- Foreign Broadcast Information Service
- US-CERT
- Oak Ridge National Laboratories
- US Army Corps of Engineers-ERDC
- Pacific Northwest National Laboratory
- ARL CERDEC & ARDEC
- US Attorney’s Office (David Capp)
Visual Analytics Uses for Public Safety and Situational Awareness

• Trend analysis, clustering, anomaly detection
• Interactive, multi-day, month, type investigation
• Multisource, multimedia data integration & analysis
• Purpose:
  • Planning for resiliency
  • Long-term analysis
  • Predictive analytics
  • Training
  • Detection
  • Investigation
  • Response
  • Recovery, remediation
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

End-Users: Lafayette, WL Police, Indiana Fusion Center, Indiana State Police, Ohio Fusion Center (initiated)

Impacts:
• In use to analyze crime patterns in Lafayette, Indiana and connect strings of activities
• Mobile version being released to public (November 2011) for community-based policing
• Investigating correlation of bus routes and crime, street lights and crime
• Analyzing time of day problems and improving accuracy of police record management system
• Novel statistical predictive model incorporated for planning

VALET delivered:
• Spring 2011: WL, Lafayette Police

iVALET delivered:
• October 2011: Purdue, WL Police
Linked Predictive Crime Models by Type
Drunkenness / Public Intoxication

May 2012
May 2012

Drunkenness / Public Intoxication

- PU vs. Notre Dame
  - PU Lost: 10-38

- Homecoming (Sat.)
  - PU vs. Illinois
  - PU Won: 21-14

- PU vs. Iowa
  - PU Lost: 21-31
Example: Public Health Visual Analytics

Health Surveillance

Pandemic Simulation and Planning

Zoonoses Planning and Response

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

Pandemic spread on day 37 with no decision measures implemented.

Pandemic spread on day 37 with all decision measures implemented.
Advanced Decision Support Tools: Rift Valley Fever
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
Risk-Based Allocations

- Comparative visual analysis of mission cases/hours vs. staffing hours
- Comparative visualization of resources vs. risk
- Trend visual analytics
  - Increase/decrease in resource allocation
  - Increase/decrease in risk (total, mitigated, residual)
  - Increase/decrease in incidents
- Exploration of alternatives and effect on risk
- Predictive analytics based on historical data (STL and EWMA)
Example: USCG D9 Search And Rescue Operational Analysis

- Interactive visual analytics of multivariate performance metrics for each unit's activities
- Interactive linked spatial temporal display, calendar view, and timeline views
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7 (Operational Analysis), USCG D9, USCG D5

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided evidence of temporal and spatial patterns used in planning – new insights to SAR mission
• Hurricane Irene resource allocation decision based on cgSARVa analysis and visualization
  • Highest SAR workload that weekend for D9
U.S. Coast Guard
Swimmer Death Analysis

Impact:
• Analyzed spatial and temporal patterns of shore-based and boat-based swimmer deaths to understand death dramatic increase in D9 in Summer 2010
• Provided information and visualizations used for public information campaign 2011 and for patrols 2011
• Significant decrease in deaths in 2011

Findings:
• Swimmer deaths
  • August highest frequency
  • Late afternoon highest frequency
  • Lake Michigan (south and west shore) have high concentration
• Boating deaths
  • Fri, Sat, Sun account for almost all deaths
  • Mid July to Mid August have highest frequency (only 1 week significantly high)
• 2009-2010 from MISLE Data
  • Large increase on Mon, Thu, Fri, Sun
  • Early and late season increase
Resource Allocation and Risk-Based Decision Making

- Explore risk-based decision making and utilize historical data for analysis and prediction
  - Total Risk, Mitigated Risk, Residual Risk
- Explore 11 different USCG missions
- Explore allocation of assets with different capabilities
- Explore staffing, utilization, assets vs. risk measures
- Perform What-If scenarios
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
Financial Risk VA
Competitive Business Intelligence Based On Point of Sales Data

- Characteristics of point of sale data
  - Multivariate
    - Large # of dimensions
      - E.g., 38 categories(products) in 288 stores
  - Temporal: 18 months
  - Spatial
    - Stores located all over Costa Rica
- Requirements
  - Supporting easy comparison among companies
    - E.g., visualize all data at once, sorting by importance
    - Enable geographical comparison
  - Easy recognition of any change in sales
    - E.g., proportional legends
  - Forecast for decision-making
FinVis: Applied Visual Analytics for Personal Financial Planning

• First visual analytics work to address portfolio planning for casual users
  • General public could benefit from this tool
• Decisions are better when using FinVis
  • Quantified using experimental economics
  • Improved
  – Decisions
  – Learning
  – exploration,
  – confidence
WireVis – Streaming Data - Multiple Linked Views

• Temporal, geospatial, theme, cluster, list views with association linkages between views

Search by Example (Find Similar Accounts)

Keyword Network (Keyword Relationships)

Strings and Beds (Relationships over Time)
Introduction to our Center - VACCINE
Motivation and introduction to visual analytics
Example solutions we’ve developed
  Health visual analytics
  Public safety
  Resource allocation

Visual analytic challenges and future
Five Six 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
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu

ebertd@purdue.edu
Visual Analytics: Powering Discovery, Innovation, and Decision Making
(Much more than Big Data Analytics plus Visualization)
Path

Introduction to our Center - VACCINE
Motivation and introduction to visual analytics

Example solutions we’ve developed
- Health visual analytics
- Maritime public safety and law enforcement
- Public safety
- Resource allocation

Visual analytic challenges and future
VACCINE Mission

- Provide visual analytic and scalable solutions for all 2.3 million extended homeland security personnel
  - 185,000 DHS personnel, 350,000 law enforcement personnel, 750,000 homeland security practitioners, 2.3 million extended personnel

- Achieve excellence in visual analytics and visualization sciences

- Educate homeland security stakeholders and the next generation of talent
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University

- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
Some VACCINE Regional and Corporate Partners

- Indiana Department of Homeland Security
- Indiana State Department of Health
- Indianapolis Public Safety
- Charlotte Mecklenburg Police Department
- Lafayette, West Lafayette, Purdue Police Departments
- Tippecanoe County EMA & Sherriff’s Department
- Harrisburg, PA Police
- Coast Guard Sector Boston, Seattle, LA
- Joint Harbor Operations Command Center
- Port of Seattle
- Indianapolis Public Safety
- USCG District 9, D5, D1
- Motorola
- Harris Corporation
- Boeing
- Next Wave Systems, LLC
- Banfield, The Pet Hospital
- Raytheon
- MacDonald, Dettwiler and Associates
- Oculus Info Inc.
- Kx Systems
- Bank of America
- Duke Energy
- World Vision International
- Gates Foundation
- Kimberly Clark
- Hallmark
Some VACCINE Government Partners

- U.S. Coast Guard
- Federal Emergency Management Agency
- Customs and Border Patrol
- National Geospatial Intelligence Agency
- National Science Foundation
- Army Research Office
- Department of Defense
- Department of Health and Human Services
- U.S. Dept. of State
- National Institute of Justice
- National Maritime Intelligence Center
- Defence Research & Development Canada
- Foreign Broadcast Information Service
- US-CERT
- Oak Ridge National Laboratories
- US Army Corps of Engineers-ERDC
- Pacific Northwest National Laboratory
- ARL CERDEC & ARDEC
- US Attorney’s Office (David Capp)
What We Do

• Enable effective decision making through interactive visual analytic environments
• Enable effective communication of information
• Provide quantitative, reliable, reproducible evidence
• Enable user to be more effective from planning to detection to response to recovery
• Enable proactive and predictive visual analytics
• Enable effective situational awareness (perception, comprehension, projection)
Path

Introduction to our Center - VACCINE

Motivation and introduction to visual analytics

Example solutions we’ve developed
  Health visual analytics
  Maritime public safety and law enforcement
  Public safety
  Resource allocation

Visual analytic challenges and future
Motivation

To solve today’s and tomorrow’s problems requires exploring, analyzing, and reasoning with massive, multisource, multiscale, heterogeneous, streaming data.
Definition

Visual Analytics\(^1\) is the science of **analytical reasoning facilitated by interactive visual interfaces**

People use visual analytics tools and techniques to

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, understandable assessments
- Communicate assessment effectively for action

---

1. *Illuminating the Path: The R&D Agenda for Visual Analytics*, Editors: Thomas and Cook

---

April 2012
What’s Needed for Proactive and Predictive Visual Analytics?

- Reliable and reproducible models and simulation
- Understanding of the data
  - Distribution and skewness, errors, appropriate analysis techniques
- Understanding of the sources and types of data
- Comparable or Correlative sources of data
  - Appropriate transformations applied to enable meaningful comparison and correlation
- Understanding of the use and problem to be solved!
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 which enhances the analytical reasoning process.

Why Was I Invited?
Visual Analytics Uses for Crisis Management and Response?

- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

Purpose:
- Planning for resiliency
- Long-term analysis
- Predictive analytics
- Training
- Detection
- Investigation
- Response
- Recovery, remediation
Visual Analytics vs. Situational Awareness

Situational Awareness ↔ Sensemaking

Situational awareness as a sensemaking process (courtesy Alan MacEachren)

Situational awareness model (after Endsley, 1995; top) compared to model of sense making (after Pirolli & Card, 2005; bottom)
Path

Introduction to our Center - VACCINE
Motivation and introduction to visual analytics

Example solutions we’ve developed

Health visual analytics
Maritime public safety and law enforcement
Public safety
Resource allocation and business intelligence

Visual analytic challenges and future
Example: Public Health Visual Analytics

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

- Pandemic spread on day 37 with no decision measures implemented
- Pandemic spread on day 37 with all decision measures implemented

Health Surveillance

Pandemic Simulation and Planning

Zoonoses Planning and Response
Visual Analytics For Syndromic Surveillance

• Syndromic surveillance: detection of adverse health events focusing on pre-diagnosis information to improve response time
• Pre-diagnosis information can consist of multiple data sources:
  – Over the counter medicine sales
  – News reports on emerging diseases
  – Pro-med news feeds
  – Emergency department chief complaints
Hypothesis Generation and Exploration

Time Series Modeling

- Seasonal-Trend Decomposition Based on Loess
  - Time series can be viewed as the sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Date:

Respiratory Count
0, 10, 20, 30, 40, 50, 60

April 2012
Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

- Pandemic Influenza Planning Tool
- Models user specified:
  - Pandemic influenza characteristics
  - County population, demographics, hospital beds
- Decision measures
  - Strategic National Stockpile deployment
  - School Closures
  - Media Alerts

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.
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
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
Example Maritime Public Safety Projects

- Search and rescue resource allocation
- Swimmer death analysis
- PWCS analysis
- Economic impact analysis
- Resource allocation and risk-based decision making
Example: Visual Analytics Environment

- Supports decision making and risk assessment
- Interactive exploration and analysis of trends, patterns and anomalies
- Allows analysis of risks associated with
  - Closing one or more Coast Guard stations
    - Find optimal stations that absorb work load of the closing station
  - Allocating new resources
    - Impact on safety and efficiency of operation
- Currently being used by analysts at the U.S. Ninth District, HQ, and Atlantic Commands
Example: USCG D9 Search And Rescue Operational Analysis

- Interactive visual analytics of multivariate performance metrics for each unit's activities
- Interactive linked spatial temporal display, calendar view, and timeline views
System Features: SAR Risk Profile

Time taken by CG stations to deploy an asset to the Great Lakes to respond to a SAR incident.
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7 (Operational Analysis), USCG D9, USCG D5

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided evidence of temporal and spatial patterns used in planning – new insights to SAR mission
• Hurricane Irene resource allocation decision based on cgSARva analysis and visualization
  • Highest SAR workload that weekend for D9
U.S. Coast Guard
Swimmer Death Death Analysis

Impact:
- Analyzed spatial and temporal patterns of shore-based and boat-based swimmer deaths to understand death dramatic increase in D9 in Summer 2010
- Provided information and visualizations used for public information campaign 2011 and for patrols 2011
- Significant decrease in deaths in 2011

Findings:
- Swimmer deaths
  - August highest frequency
  - Late afternoon highest frequency
  - Lake Michigan (south and west shore) have high concentration
- Boating deaths
  - Fri, Sat, Sun account for almost all deaths
  - Mid July to Mid August have highest frequency (only 1 week significantly high)
- 2009-2010 from MISLE Data
  - Large increase on Mon, Thu, Fri, Sun
  - Early and late season increase
Resource Allocation and Risk-Based Decision Making

• Explore risk-based decision making and utilize historical data for analysis and prediction
  • Total Risk, Mitigated Risk, Residual Risk
  • Explore 11 different USCG missions
  • Explore allocation of assets with different capabilities
  • Explore staffing, utilization, assets vs. risk measures
  • Perform What-If scenarios
Example Public Safety Projects:

• Evacuation planning and management
• Hazardous Materials - mobile imaging
• Law enforcement visual analytics
• Additional projects
  • GARI – gang graffiti
  • Jigsaw
Large Scale Evacuation Model: Decision Support (Ribarsky et al. UNCC)
Mobile Interface
MERGE: Mobile Emergency Response Guide – System Overview

1. Image Capture
2. Image Analysis
3. Database Query
4. Display Information
• Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Geovisual Analytics Support For Sensemaking
In Public Health And Crisis Management
(GeoVista Center, PSU)

• SensePlace – leveraging news to support infectious disease modeling and delivery of services in a developing country context
• SensePlace2 – leveraging social media to support situational awareness for health and crises

VACCINE

April 2012
SensePlace2: Place-Time-Concept Analytics

query window

grid map of tweet frequency matching query; graduated circles depict 500 most relevant tweets; support for spatial filter by distance from point

time filter & freq display

ranked, sortable 500 most relevant tweets

history view

selected place

http://www.geovista.psu.edu/SensePlace2/

NEW PHOTOS FROM VAN AND ABOUT NEW EARTHQUAKES http://t.co /qIAH6Uoq Van Armenia Yerevan Earthquake Turkey 10/24 2011
Scatterblogs: Geo-Spatial Document Analysis
Spatiotemporal Social Media Analytics for Abnormal Event Detection
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

End-Users: Lafayette, WL Police, Indiana Fusion Center, Indiana State Police, Ohio Fusion Center (initiated)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and connect strings of activities
- Mobile version being released to public (November 2011) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

**VALET delivered:**
- Spring 2011: WL, Lafayette Police

**iVALET delivered:**
- October 2011: Purdue, WL Police
Linked Predictive Crime Models by Type
Drunkenness / Public Intoxication

- PU vs. Notre Dame
  - PU Lost: 10-38

- Homecoming (Sat.)
  - PU vs. Illinois
  - PU Won: 21-14

- PU vs. Iowa
  - PU Lost: 21-31

Day-of-the-Week

- Mon
- Tue
- Wed
- Thu
- Fri
- Sat
- Sun

Football season
- Home
- Away
iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
iVALET

- Linked views to explore multivariate spatiotemporal dataset
- Analytical tools to help explore data
GeoVISTA CrimeViz

... an extensible web-based geovisualization application that supports exploration of & sensemaking about criminal activity in space & time

Project Personnel:

Alan M. MacEachren
Robert E. Roth
Kevin S. Ross
Scott Pezanowski

GeoVISTA Center
Department of Geography
Penn State University
maceachren@psu.edu

http://www.geovista.psu.edu/ CrimeViz
GeoVISTA CrimeViz: Year/Month View

Aggrevated Assault, by month, for a year; map depicts freq/cell in Nov.
Crime Analytics & SA Issues and Techniques

- Fuse data from a variety of sources
  - Law enforcement records management
  - Weather and phases of the moon
  - Street light locations, bus routes
  - Tracking release data of offenders
  - Civil court data
  - Social Media
  - Local event calendar

- Reliable predictive models

- Understandability and trust of predictions

- Main Question: What helps officers, detectives, chief do their variety of jobs?
Jigsaw: Visual Analytics for Investigative Analysis and Exploration of Document Collections

**Goal:** Assist investigators with understanding, sense-making, and analysis of large, unstructured and structured document collections

**Approach:** Provide multiple visual perspectives on the documents and entities within them, highlighting connections between entities
Gang Graffiti Recognition and Analysis Using a Mobile Telephone (GARI)
Edward J. Delp, Mireille Boutin
Collaborating Institution(s): Purdue University
End-User(s): Gary Coons, Chief/Indianapolis Department of Public Safety Division of Homeland Security and IMPD, Indiana Fusion Center Gang Task Force

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 40 users and 450 graffiti images acquired

GARI delivered:
• Summer 2011:
  • IMPD gang detectives
• August 2011:
  • IMPD at large
  • Ind Fusion Gang Task Force
• October 2011:
  • Gang detectives across Indiana
Path

Introduction to our Center - VACCINE
Motivation and introduction to visual analytics
Example solutions we’ve developed
  Health visual analytics
  Public safety
  Resource allocation

Visual analytic challenges and future
Five Six 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
Visual Analytics At Real-World Scale

- Utilize advanced HPC techniques to enable interactive spatiotemporal analysis (spatiotemporal clustering, prediction)
- Cluster-based and cloud-based solutions
- GPGPU solutions
- Develop easily usable HPC visual analytic environments

Example: Longhorn Exascale Visual Analytic Platform

- 2048 compute cores (Nehalem quad-core)
- 512 GPUs (128 NVIDIA Quadro Plex S4s, each containing 4 NVIDIA FX 5800s)
- 13.5 TB of distributed memory
- 210 TB global file system
The Science of Interaction

• Definition: The study of methods by which humans create knowledge through the manipulation of an interface.

  • …interaction and inquiry are inextricable. It is through the interactive manipulation of a visual interface – the analytic discourse – that knowledge is constructed, tested, refined and shared.”

Stasko et al., IV2009

5 Challenges:

1. **Visual Discourse** - Interactive visual thinking tools for the exploration, understanding, collaboration, description, explanation, decision, dissemination, persuasion of concepts & data.

2. **Multi-modal Sensemaking In The Large** - Enable large-scale, distributed teams to interactively make sense of big multi-modal data and problems.

3. **Fluid Interaction** - Designing fluid, high-bandwidth, and powerful interaction models and paradigms for the purpose of individuals reaching their full potential in viewing, analyzing, and understanding large and complex data.

4. **Collaboration** – beyond space and time- synergizing technologies and human users

5. **Mixed Initiative Data Discovery and Manipulation** - Balancing active user input with systematic guidance to enable visual data manipulation and analysis.
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
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Visual Analytics ≠ Big Data Analytics

David S. Ebert
ebertd@purdue.edu

Path

Introduction
Motivation and introduction to visual analytics
Example solutions we’ve developed
Science visual analytics
Health visual analytics
Maritime public safety and law enforcement
Public safety
Resource allocation
Visual analytic challenges and future
Big Data Is The Rage

The Promise of Big Data:

• Commercial Zoo:

IBM Big Data Platform

Analytic Applications

<table>
<thead>
<tr>
<th>Visualization &amp; Discovery</th>
<th>Application Development</th>
<th>Systems Management</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accelerators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holistic System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stream Computing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data Warehouse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information/Intelligence</td>
<td></td>
</tr>
</tbody>
</table>

Big Data Landscape

<table>
<thead>
<tr>
<th>Vertical Apps</th>
<th>Ad/Media Apps</th>
<th>Business Intelligence</th>
<th>Analytics &amp; Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Data Apps</td>
<td></td>
<td>Data as a Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operational Infrastructure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Infrastructure as a Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Structured Databases</td>
<td></td>
</tr>
</tbody>
</table>

July 2012
Premise of Talk:

• Visual Analytics Is More Than Big Data Analytics

What We Do With Visual Analytics

• Enable effective decision making through interactive visual analytic environments
• Enable effective communication of information
• Provide quantitative, reliable, reproducible evidence
• Enable user to be more effective from planning to detection to response to recovery
• Enable proactive and predictive visual analytics
• Enable effective situational awareness (perception, comprehension, projection)
Motivation

To solve today’s and tomorrow’s problems requires exploring, analyzing, and reasoning with massive, multisource, multiscale, heterogeneous, streaming data.

Path

Introduction

Motivation and introduction to visual analytics

Example solutions we’ve developed
  - Science visual analytics
  - Health visual analytics
  - Maritime public safety and law enforcement
  - Public safety
  - Resource allocation

Visual analytic challenges and future
Definition

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

People use visual analytics tools and techniques to
- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, understandable assessments
- Communicate assessment effectively for action

---

1. *Illuminating the Path: The R&D Agenda for Visual Analytics, Editors: Thomas and Cook*
Research Motivation:

- Solving these real-world problems requires
  - Novel theories, techniques, approaches, and adaptations of algorithms
  - Integration of cross-disciplinary expertise
- Solving these real-world problems provides
  - Compelling, publicly understandable value for your research
  - Advances in CS and in other disciplines
  - New publication opportunities
  - Great collaboration partners and proponents
  - Opportunities for new adventures

Example Challenging Problem - Atmospheric Science: Multi-scale Interactions (in the words of a cloud physicist)

No observing platform can measure the quantities of interest over all needed spatial and temporal scales needed

No numerical model can simulate the quantities of interest over all needed spatial and temporal scales

→ We observe/simulate over a subset of the pertinent scales, using different instruments/models, and must assimilate these results to understand the “big picture”

Visual analytics is crucial for this task

Issues: Multi-scale, multi-system, multisource, massive, data & simulations
One Solution in Use: Our Atmospheric Visual Analytic Environment

Utilize multiple rendering styles
Provide interactive data exploration and user directed analysis
Allow user specified analysis and queries on the fly
Allow interactive correlative analysis of multisource data

What’s Needed for Proactive and Predictive Visual Analytics?

• Reliable and reproducible models and simulation
• Understanding of the data
  • Distribution and skewness, errors, appropriate analysis techniques
• Understanding of the sources and types of data
• Comparable or Correlative sources of data
  • Appropriate transformations applied to enable meaningful comparison and correlation
• Understanding of the use and problem to be solved!
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 which enhances the analytical reasoning process.

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”
Path

Introduction to our Center - VACCINE

Motivation and introduction to visual analytics

Example solutions we’ve developed

Science visual analytics
Health visual analytics
Maritime public safety and law enforcement
Public safety
Resource allocation and business intelligence

Visual analytic challenges and future
Example: Public Health Visual Analytics

Visual Analytics For Syndromic Surveillance

- Syndromic surveillance: detection of adverse health events focusing on pre-diagnosis information to improve response time
- Pre-diagnosis information can consist of multiple data sources:
  - Over the counter medicine sales
  - News reports on emerging diseases
  - Pro-med news feeds
  - Emergency department chief complaints
Hypothesis Generation and Exploration

Time Series Modeling

- Seasonal-Trend Decomposition Based on Loess
  - Time series can be viewed as the sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components

(Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., Cleveland, W., Grannis, S., Wade, M., Ebert, D.)

Hafen, R., Anderson, D., Cleveland, W., Maciejewski, R., Ebert, D., Abusalah, A., Yakout, M., Ouzzani, M., Grannis, S., Wade, M., Ebert, D.
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

• Pandemic Influenza Planning Tool
  • Models user specified:
    • Pandemic influenza characteristics
    • County population, demographics, hospital beds

• Decision measures
  • Strategic National Stockpile deployment
  • School Closures
  • Media Alerts

Visual Analytics Uses for Crisis Management and Response

• Trend analysis, clustering, anomaly detection
• Interactive, multi-day, month, type investigation
• Multisource, multimedia data integration & analysis

• Purpose:
  • Planning for resiliency
  • Long-term analysis
  • Predictive analytics
  • Training
  • Detection
  • Investigation
  • Response
  • Recovery, remediation
Visual Analytics vs. Situational Awareness

Situational Awareness ↔ Sensemaking
situational awareness as a sensemaking process (courtesy Alan MacEachren)

Situational awareness model (after Endsley, 1995; top) compared to model of sense making (after Pirolli & Card, 2005; bottom)

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
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
Example Maritime Public Safety Projects

- Search and rescue resource allocation
- Swimmer death analysis
- PWCS analysis
- Economic impact analysis
- Resource allocation and risk-based decision making

Example: Visual Analytics Environment

- Supports decision making and risk assessment
- Interactive exploration and analysis of trends, patterns and anomalies
- Allows analysis of risks associated with
  - Closing one or more Coast Guard stations
    - Find optimal stations that absorb work load of the closing station
  - Allocating new resources
    - Impact on safety and efficiency of operation
- Currently being used by analysts at the U.S. Ninth District, HQ, and Atlantic Commands
Example: USCG D9 Search And Rescue Operational Analysis

System Features: SAR Risk Profile

Time taken by CG stations to deploy an asset to the Great Lakes to respond to a SAR incident.
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7 (Operational Analysis), USCG D9, USCG D5

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided evidence of temporal and spatial patterns used in planning – new insights to SAR mission
• Hurricane Irene resource allocation decision based on cgSARva analysis and visualization
  • Highest SAR workload that weekend for D9

U.S. Coast Guard Swimmer Death Analysis

Impact:
• Analyzed spatial and temporal patterns of shore-based and boat-based swimmer deaths to understand death dramatic increase in D9 in Summer 2010
• Provided information and visualizations used for public information campaign 2011 and for patrols 2011
• Significant decrease in deaths in 2011

Findings:
• Swimmer deaths
  • August highest frequency
  • Late afternoon highest frequency
  • Lake Michigan (south and west shore) have high concentration
• Boating deaths
  • Fri, Sat, Sun account for almost all deaths
  • Mid July to Mid August have highest frequency (only 1 week significantly high)
• 2009-2010 from MISLE Data
  • Large increase on Mon, Thu, Fri, Sun
  • Early and late season increase
Resource Allocation and Risk-Based Decision Making

- Explore risk-based decision making and utilize historical data for analysis and prediction
  - Total Risk, Mitigated Risk, Residual Risk
  - Explore 11 different USCG missions
  - Explore allocation of assets with different capabilities
  - Explore staffing, utilization, assets vs. risk measures
  - Perform What-If scenarios

USCG PWCS PROTECT Project
Partners: USC CREATE, USCG RDC, USCG D1, USCG LANT, USCG PAC

Impact:
- Provided insight and analysis into historical security patrols in Boston
- Analyzed developing PROTECT model and provided pattern analysis to improve model during operational deployment in Boston and for improvements for deployment in NYC
- Developed end-user tool for evaluation and validation of patrol routes
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

Example Public Safety Projects:
• Evacuation planning and management
• Hazardous Materials - mobile imaging
• Law enforcement visual analytics
• Additional projects
  • GARI – gang graffiti
  • Jigsaw
MERGE: Mobile Emergency Response Guide – System Overview

Image Capture → Image Analysis

Display Information → Database Query

MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

• Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Scatterblogs: Geo-Spatial Document Analysis

Spatiotemporal Social Media Analytics for Abnormal Event Detection
Correlation Between Multiple Social Media

- Occupy Wall Street event

<table>
<thead>
<tr>
<th>Date</th>
<th>Twitter (z-score)</th>
<th>Flickr (z-score)</th>
<th>Date</th>
<th>YouTube (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/17, 2011</td>
<td>2.53</td>
<td>0.97</td>
<td>9/18, 2011</td>
<td>5.98</td>
</tr>
<tr>
<td>9/30, 2011</td>
<td>3.20</td>
<td>3.50</td>
<td>10/2, 2011</td>
<td>1.38</td>
</tr>
<tr>
<td>10/5, 2011</td>
<td>3.17</td>
<td>3.36</td>
<td>10/6, 2011</td>
<td>2.75</td>
</tr>
</tbody>
</table>

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

End-Users: Lafayette, WL Police, Indiana Fusion Center, Indiana State Police, Ohio Fusion Center (initiated)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and connect strings of activities
- Mobile version being released to public (November 2011) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
Visual Analytics Law Enforcement Toolkit (VALET) - Tippecanoe County Example

Linked Predictive Crime Models by Type

Day vs. Night Thefts

July 2012

2008 (red) vs. 2007 (blue background)

Aggravated Theft

Drunkenness / Public Intoxication

Day-of-the-Week
Drunkenness / Public Intoxication

Football season

- Home
- Away

PU vs. Notre Dame
PU Lost: 10-38

Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Case Study 2: Drug abuse violations vs. burglaries (IV)

- Over geospatial hotspots
Case Study 3: Drunkenness/public intoxication and noise complaints vs. traffic accidents (II)

By Hour

iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
iVALET

- Linked views to explore multivariate spatiotemporal dataset
- Analytical tools to help explore data

Crime Analytics & SA Issues and Techniques

- Fuse data from a variety of sources
  - Law enforcement records management
  - Weather and phases of the moon
  - Street light locations, bus routes
  - Tracking release data of offenders
  - Civil court data
  - Social Media
  - Local event calendar

- Reliable predictive models
- Understandability and trust of predictions
- Main Question: What helps officers, detectives, chief do their variety of jobs?
**Jigsaw:**
Visual Analytics for Investigative Analysis and Exploration of Document Collections

**Goal:** Assist investigators with understanding, sense-making, and analysis of large, unstructured and structured document collections

**Approach:** Provide multiple visual perspectives on the documents and entities within them, highlighting connections between entities.

---

**Gang Graffiti Recognition and Analysis Using a Mobile Telephone (GARI)**
Edward J. Delp, Mireille Boutin
Collaborating Institution(s): Purdue University
End-User(s): Gary Coons, Chief/Indianapolis Department of Public Safety Division of Homeland Security and IMPD, Indiana Fusion Center Gang Task Force

**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 40 users and 450 graffiti images acquired

GARI delivered:
- Summer 2011: IMPD gang detectives
- August 2011: IMPD at large
- October 2011: Gang detectives across Indiana
Path

Introduction to our Center - VACCINE
Motivation and introduction to visual analytics
Example solutions we’ve developed
Science visual analytics
Health visual analytics
Public safety
Resource allocation

Visual analytic challenges and future

Six 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
Visual Analytics At Real-World Scale

- Utilize advanced HPC techniques to enable interactive spatiotemporal analysis (spatiotemporal clustering, prediction)
- Cluster-based and cloud-based solutions
- GPGPU solutions
- Develop easily usable HPC visual analytic environments

Example: Longhorn Exascale Visual Analytic Platform

- 2048 compute cores (Nehalem quad-core)
- 512 GPUs (128 NVIDIA Quadro Plex S4s, each containing 4 NVIDIA FX 5800s)
- 13.5 TB of distributed memory
- 210 TB global file system

The Science of Interaction

- Definition: The study of methods by which humans create knowledge through the manipulation of an interface.
  - …interaction and inquiry are inextricable. It is through the interactive manipulation of a visual interface – the analytic discourse – that knowledge is constructed, tested, refined and shared.”
  - Stasko et al., IV2009

5 Challenges:

1. **Visual Discourse** - Interactive visual thinking tools for the exploration, understanding, collaboration, description, explanation, decision, dissemination, persuasion of concepts & data.

2. **Multi-modal Sensemaking In The Large** - Enable large-scale, distributed teams to interactively make sense of big multi-modal data and problems.

3. **Fluid Interaction** - Designing fluid, high-bandwidth, and powerful interaction models and paradigms for the purpose of individuals reaching their full potential in viewing, analyzing, and understanding large and complex data.

4. **Collaboration** – beyond space and time- synergizing technologies and human users

5. **Mixed Initiative Data Discovery and Manipulation** - Balancing active user input with systematic guidance to enable visual data manipulation and analysis.
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
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu

ebertd@purdue.edu
VACCINE Mission

• Provide visual analytic and scalable solutions for all 2.3 million extended homeland security personnel
  - 185,000 DHS personnel, 350,000 law enforcement personnel, 750,000 homeland security practitioners, 2.3 million extended personnel

• Achieve excellence in visual analytics and visualization sciences

• Educate homeland security stakeholders and the next generation of talent
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

• Purdue University
• Georgia Institute of Technology
• Pennsylvania State University
• Stanford University
• University of North Carolina at Charlotte
• University of Washington
• Simon Fraser University
• University of British Columbia
• Justice Institute of British Columbia
• Ontario Institute of Technology
• Dalhousie University

• University of Houston, Downtown
• Virginia Tech
• Indiana University
• Florida International University
• University of Texas at Austin
• Morgan State University
• Navajo Technical College
• University of Stuttgart
• University of Swansea, UK
• Oxford
• University of Calgary
• University of Manitoba
• Carleton University
• University of Victoria

Some VACCINE Regional and Corporate Partners

• Indiana Department of Homeland Security
• Indiana State Department of Health
• Indianapolis Public Safety
• Charlotte Mecklenburg Police Department
• Lafayette, West Lafayette, Purdue Police Departments
• Tippecanoe County EMA & Sheriff’s Department
• Harrisburg, PA Police
• Coast Guard Sector Boston, Seattle, LA
• Joint Harbor Operations Command Center
• Port of Seattle
• Indianapolis Public Safety
• USCG District 9, D5, D1
• Motorola
• Harris Corporation
• Boeing
• Next Wave Systems, LLC
• Banfield, The Pet Hospital
• Raytheon
• MacDonald, Dettwiler and Associates
• Oculus Info Inc.
• Kx Systems
• Bank of America
• Duke Energy
• World Vision International
• Gates Foundation
• Kimberly Clark
• Hallmark
Some VACCINE Government Partners

- U.S. Coast Guard
- Federal Emergency Management Agency
- Customs and Border Patrol
- National Geospatial Intelligence Agency
- National Science Foundation
- Army Research Office
- Department of Defense
- Department of Health and Human Services
- U.S. Dept. of State
- National Institute of Justice
- National Maritime Intelligence Center
- Defence Research & Development Canada
- Foreign Broadcast Information Service
- US-CERT
- Oak Ridge National Laboratories
- US Army Corps of Engineers-ERDC
- Pacific Northwest National Laboratory
- ARL CERDEC & ARDEC
- US Attorney’s Office (David Capp)
Visual Analytics
A Lifeboat In The Big Data Ocean

Interpreting Visualized Data for Defense and Government Operations

David S. Ebert
ebertd@purdue.edu

Big Data

• **Big Data Definition:** A phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex and diverse types of data. (TechAmerica Foundation)
  
  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**
Big Data Challenges

• Heterogeneity
• Unprecedented volume and/or complexity
• Multivariate, Multidimensional
• Streaming
• Qualitative, Quantitative, Categorical, Temporal,…
• Distributed Sources: sensors, text, images, maps, …
• Inconsistent and Uncertain
• Different Scales and Cross Scale

Big Data Analytics – The Solution?

• Advanced data analytics for big data can help manage big data
  • Machine learning
  • Advanced statistical analysis
  • Hadoop-deployed analytics
• However, it falls short in several aspects
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 does it incorporate all relevant information for decision making
  • Humans have extensive information and experience that never make it into the collected data

What Big Data Analytics Can’t Do
(inspired and adapted from David Brooks, New York Times, 2/18/2013)

• 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
The Power of Visualization

- Turn the abstract into the concrete
- Enable rapid perception and visual cognition of information to increase efficient understanding and decision making
- Goal: Transform data into information and knowledge
- Data size doesn’t matter, size of insights matters
- Visualization is a key method to increase understandability of data and analytical results

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”
Beyond Visualization to Visual Analytics

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

Interactive visualization, data analysis, exploration, and decision making with human in the loop!

People use visual analytics tools and techniques to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, understandable assessments
- Communicate assessment effectively for action

VA Approaches to Big Data

- Don’t display all the data!
- Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space
- User interaction to guide above process
- Adapt to available hardware for the problem
  - Should we use big iron or cell phones?
Visual Analytics’ Value

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: End User-driven solutions and research

Examples And Visual Analytic Solutions

• Example 1: Basic question of understanding why and when a cumulus cloud will produce precipitation
  • Data: Multiscale, multisource, incomplete data

• Example 2: Predictive crime analytics – when and where will crimes occur today
  • Data: A few million crime records with uncertain, missing data, dozens of fields, associated risks, related social media data, images

• Example 3: US Coast Guard decision making for efficiency and effectiveness
  • Data: Mission record, risk models, costs, weather data
Atmospheric Science: Multi-scale Interactions (in the words of a cloud physicist)

No observing platform can measure the quantities of interest over all needed spatial and temporal scales needed.

No numerical model can simulate the quantities of interest over all needed spatial and temporal scales.

→ We observe/simulate over a subset of the pertinent scales, using different instruments/models, and must assimilate these results to understand the “big picture.”

**Visual analytics is crucial for this task.**

**Issues:** Multi-scale, multi-system, multisource, massive, data & simulations

One Solution in Use: Our Atmospheric Visual Analytic Environment

- Utilize multiple rendering styles
- Provide interactive data exploration and user directed analysis
- Allow user specified analysis and queries on the fly
- Allow interactive correlative analysis of multisource data
Crime Analytics Issues And Techniques

- **Requirement:** Fuse data from a variety of sources
  - Law enforcement records management
  - Weather and phases of the moon
  - Street light locations, bus routes
  - Tracking release data of offenders
  - Civil court data
  - Social Media
  - Local event calendar

- **Need:**
  - Reliable predictive models
  - Understandability and trust of predictions

- **Main Question:** What helps officers, detectives, chief do their variety of jobs?
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police
- October 2011: Purdue, WL Police

iVALET delivered:
- October 2011: Purdue, WL Police

VALET Overview
Social Media: Real-time Twitter Monitoring and Integration into Tools
(Purdue, Stuttgart, Penn St.)

- Topic extraction using novel STL based remainder estimation technique

- Dynamically linked views providing options to monitor emerging / emergent Twitter feeds

- Topics extracted shown as a dynamic word cloud

Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

Home
Away

Football season

PU vs. Notre Dame
PU Lost: 10-38
Homecoming (Sat.)
PU vs. Illinois
PU Won: 21-14

PU vs. Iowa
PU Lost: 21-31

Day-of-the-Week

0 10 20 30 40 50 60

iVALET
• Explore criminal traffic and civil data on the go
• Provides risk assessments
• Use current spatial + temporal context into analysis
• Deployed to four police departments in beta testing

February 2013
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected.

---

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 60 users and 500 graffiti images acquired.

GARI delivered:
- Summer 2011:
  - IPD gang detectives
  - August 2011:
  - IPD at large
  - Ind Fusion Gang Task Force
  - September 2011:
  - Gang detectives across Indiana.
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

USCG: 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 at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Superstorm Sandy: Used for resource allocation in response and in rebuilding
• Used for Hurricane Irene resource allocation decisions
• Informed Commandant’s budget testimony and recommendations to Congress
• Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

Example Screenshot: cgSARVA
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area

Jones Beach, Sandy Hook and Manasquan Inlet
Closest Cases (not necessarily reported under them)
All Cases November 2011
Risk-Based Allocations

• Comparative visual analysis of mission cases/hours vs. staffing hours
• Comparative visualization of resources vs. risk
• Trend visual analytics
  • Increase/decrease in resource allocation
  • Increase/decrease in risk (total, mitigated, residual)
  • Increase/decrease in incidents
• Exploration of alternatives and effect on risk
• Predictive analytics based on historical data
• Effectiveness and efficiency measures, cost measures

Cost Based on Resources Used for the Open-Suspended Sorties Over 5 years (2006-2011)
Response Efficiency – Notional Future Assets

Interactive Operational Performance

- OPAR
- Standard report within the Coast Guard Business Intelligence (CGBI) system
- Displays resource use and performance by core CG mission areas
- iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making

- Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
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
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu

February 2013
Visual Analytics and Decision Making

David S. Ebert
ebertd@purdue.edu
What We Do

- Enable effective decision making through interactive visual analytic environments
- Enable effective communication of information
- Provide quantitative, reliable, reproducible evidence
- Enable user to be more effective from planning to detection to response to recovery
- Enable proactive and predictive visual analytics
Some Example Projects

- **Science Visual Analytics** *(weather, nanotechnology, cancer care)*
- **Scalable Law Enforcement Projects**
- **USCG Projects**
  - Search and rescue and swimmer deaths
  - PWCS analysis and economic impact analysis
  - Resource allocation and risk-based decision making
- **Predictive spatiotemporal analytics – Indiana Intelligence Fusion Center**
- **Multivariate visual analytics (business analytics)**
- **Decision support with linked simulations**
- **Investigative analytics (text and streaming data)**
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 which enhances the analytical reasoning process.

---

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"
Visualization, Data, and Decision Making

- Unfortunately, it is easy to mislead with visual representations, even unintentionally!
- Fortunately, visual perception is the largest bandwidth into the human brain
- Visualization is an effective, efficient communication mechanism and a common language
  - “Even a simple sailor like me can understand this”
    RDM Mike Parks, USCG D9
Current: Utilization and Effectiveness

- At some level, most policies and decisions are data driven
- Beginning to explore conveyance of decisions
- Ease and effectiveness across scales of
  - Decision makers
  - Affected population
  - Time for consideration decreases exponentially with increase in management level

Need: Interactive VA with integrated consequence modeling
Caveats

• Reliability and certainty must be conveyed
• Someone must understand and verify the data interpretation
VACCINE:

Who We Are & What We Do
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
What We Do: Visual Analytics

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

People use visual analytics tools and techniques to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, understandable assessments
- Communicate assessment effectively for action

1. Illuminating the Path: The R&D Agenda for Visual Analytics, Editors: Thomas and Cook
VACCINE Value to DHS/ HSE

- **Innovative Fielded Solutions**: We provide innovative visual analytic and scalable solutions to the extended homeland security community

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information

- **People and Partnerships**
  - Interdisciplinary world-leading team of researchers and students
  - Actively Engaged Partners – We define and extend the new science of visual analytics driven by real-world, real-scale problems of engaged partners (local, state, federal)

“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
Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Customs and Immigration Service
  • US Federal Emergency Management Agency
  • US Customs and Border Patrol
  • US CERT
  • US ICE (in progress)

• Law Enforcement
  • Over 40 local and state agencies

• Fusion Centers
  • Ohio (SAIC)
  • Indiana (IIFC)
VACCINE:

Our Work
Visual Analytics Uses for Crisis Management and Response

- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

Purpose:
- Planning for resiliency
- Long-term analysis
- Predictive analytics
- Training
- Detection
- Investigation
- Response
- Recovery, remediation
Situational Awareness, Planning, Investigation, Response Visual Analytics

- Situational Awareness and Assessment from massive data
  - In-field and desktop historical and current incident/crime information
- Innovative spatiotemporal visualization, analysis, prediction
- Image analysis tools
- Integrated statistical forecasting and simulation models
- Advanced social media analytics

- Tools for Intelligence analysis
- Planning and response support
  - Integrated simulation, resource allocation, census, weather data
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
- Interactive, multi-day, month, type investigation
- Multisource, multimedia massive data integration & analysis
- Visual standardization and interoperability
Current Public Safety Coalition Projects: Law Enforcement

Desktop crime analytics and intelligence analytics

- VALET & CrimeViz
  - Delivered to Lafayette PD 2011, WL PD Spring 2012
  - Delivered to Harrisburg Bureau of Police Spring 2011
- TRIP
  - Delivered to Indiana Intelligence Fusion Center May 2012
- Jigsaw
  - Delivered to West Lafayette PD, IMPD 09/2010, TSA 06/2012

In-field crime analytics

- Gang Graffiti Mobile Application (Gari)
  - Delivered to IMPD 09/2011, in use by 15 different police agencies in Indiana
- iValet
  - Delivered to WLPD, LPD, Purdue PD Fall 2011

Desktop

• Miami-Dade EOC Image-based Disaster Assessment
  • Delivered to EOC Spring 2011, students deployed 2011, 2012
• Flooding damage assessment (cooperation with CHC COE)
  • Used by CHC for assessment and by FEMA/DHS
• Visual Analytics in EOCs (partnership with DRDC Canada)
  • Used in Richmond, BC EOC and for exercises

In-field

• HazMat Visual Analytic Application (MERGE)
  • Delivered to Carroll County EMA, Purdue Police, Ind. Fire, 08/2011
  • Delivered to TSA July 2012
A Few Examples of Working with Stakeholders

• GARI – initiated by Chief John Cox, Purdue Police
• MERGE – initiated by FLETC meeting
• VALET – initiated by Sheriff Tracy Brown, Tippecanoe County
• Jigsaw – initiated by briefs with intelligence analysts
• TRIP – initiated by Indiana Intelligence Fusion Center
• cgSARVA – initiated by Dr. Joe DiRenzo, Chief Operations Analysis, USCG LANT
VACCINE:

Our Projects
Example Projects

• VASA
• TSA
• Public Safety and Law Enforcement
• USCG
• Financial Fraud
• Cybersecurity
VASA: Visual Analytics for Security Applications
Cascading Critical Infrastructure Resiliency Modeling and Analytics

• Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.

• Collaborating Institution(s):
Purdue, UNC Charlotte, U. Minn. (NCFPD), U. 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
CI/KR’s: Transportation, Food Safety, Cyber Networks and Power Grids at the national levels.

Key Outcome: Visual analytics prototype to analyze large ensembles of simulations (massive, dependant data) for a range of situations to fully understand areas of stability and key failure points.

Year 1:
- Develop QSR model based on measurements and link in basic VA system prototype
- Develop utility model and powergrid interactions and link in basic VA system prototype
TSA Projects

• Visual Analytics for multifaceted operational analysis
  • Analysis and correlative analysis of
    • Operational performance (throughput per lane, use of AIT, targets)
    • Security performance (incidents)
    • Financial performance (expenses vs. performance metrics)
  • Spatial, temporal, spoke, hub, and regional analysis

• Customer satisfaction visual analytics
  • Jigsaw use to understand and find patterns in compliment/complaint database
  • Potential linkage to above data

• MERGE for rail inspectors
TSA Visual Analysis Prototype
Flight Delay Visual Analysis Prototype

(A) Calendar View
(A-1) Weekly Trend

(A-2) Day of Week Trend

(B) Time Filters

(C) Line graph & Correlation View

(D) Legend View (Type of Delays)

(E) Geographical View
(E-1) Legend for Pixel and Map View

(F) Pixel View

(G) Airport Filter

(H) Carrier Filter

(I) Age Filter

(J) Clock View

(K) Twitter View

Origins

Destinations

(F-1) Departure Delay

(F-2) Type of Delay & Arrival Delay

(F-3) Airlines

(F-4) Ages of Aircrafts
Flight Delay Visual Analysis Prototype

• Delay at DFW causing delays at destinations by delay type
Public Safety & Law Enforcement

- Law enforcement visual analytics
  - VALET, iVALET (iPhone/iPad), CrimeViz
  - Visual Analytics Law Enforcement Toolkit
  - Analyzing crime patterns and time of day problems
- Gang activity analytics
  - GARI
  - Gang Graffiti Recognition and Interpretation using a mobile telephone
  - Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
- Document visual analytics
  - JIGSAW
  - Visualization for investigative analysis
  - Discovery of hidden relationship and threats across documents
Jigsaw

Visualization for Investigative Analysis across Document Collections

- Law enforcement & intelligence community
- Fraud (finance, accounting, banking)
- Academic research
- Journalism & reporting
- Consumer research

“Putting the pieces together”
Jigsaw: Problem Addressed

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

- Articles & reports
- Spreadsheets
- XML documents
- Blogs
Jigsaw: Visual Analytics for Investigative Analysis

Documents in

Knowledge out

John Stasko
Jigsaw List View

Lists of entities by type
Connections highlighted

<table>
<thead>
<tr>
<th>Concept</th>
<th>Year</th>
<th>Conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>1995</td>
<td>VAST</td>
</tr>
<tr>
<td>Evaluation</td>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>Insight</td>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>Visual Analytics</td>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>Case Study</td>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>Navigation</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>Animation</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>Document</td>
<td>2005</td>
<td></td>
</tr>
<tr>
<td>Dynamic Query</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>Filter</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Focus+Context</td>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>Hierarchy</td>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>Intelligence Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software Visualization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freeware</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aesthetics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioinformatics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brushing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Mining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geospatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph Hardware</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-dimensional Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VACCINE
"Recently, the Washington Joint Analytical Center (WAJAC) assisted a local police department in analyzing 345 incident reports pertaining to graffiti. Jigsaw was an invaluable tool when it came to examining reports. Without the use of Jigsaw, examining reports would have been tedious and time consuming, and connections between entities may not have been seen.

The local police department was very impressed with Jigsaw and stated that the program will become very useful in tracking graffiti activity, and possibly other criminal activity.

WAJAC uses Jigsaw on a continuous basis to examine information in our intelligence database. Jigsaw is also used to analyze incidents relating to the security of the Washington State Ferry System. It provides a searchable database and quick analytical tool for tracking potential threats to the largest ferry transportation system in the U.S."

CHAD R MELTON
Criminal Intelligence Analyst
WAJAC/FBI-FIG
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:

- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (February 2013) for community-based policing
- Investigating correlation factors
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning
- Incorporating predictive alerts

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview

Map View

Time Series View

Clock View

Twitter monitoring

Time Slider

Calendar View

Menu
Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame: PU Lost: 10-38
- Homecoming (Sat.): PU vs. Illinois
  - PU Won: 21-14
- PU vs. Iowa: PU Lost: 21-31

Football season:
- Home
- Away

Day-of-the-Week
- Mon
- Tue
- Wed
- Thu
- Fri
- Sat
- Sun

Map of the area with locations marked for football games and the number of occurrences.
Social Media: Real-time Twitter Monitoring and Integration into Tools
(Purdue, Stuttgart, Penn St.)

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
SensePlace2: Place-Time-Concept Analytics

- **query window**: search for: earthquake Turkey
- **time filter & freq display**: grid map of tweet frequency matching query; graduated circles depict 500 most relevant tweets; support for spatial filter by distance from point
- **history view**: ranked, sortable 500 most relevant tweets
- **selected place**

For more information on SensePlace2, please see the project site: [http://www.geovista.psu.edu/SensePlace2/](http://www.geovista.psu.edu/SensePlace2/)
Explore criminal, traffic and civil data on-the-go

• Risk assessment
• Use current spatial + temporal context into analysis
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected.
We Have an App For That!

Our Mobile Tools

• VALET
• Evacuation Planning
• Rosetta Phone
• Hazmat app
• Gang graffiti app
• Tatoo app
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 75 users and 800 graffiti images acquired

**GARI delivered:**
- Summer 2011:
  - IPD gang detectives
- August 2011:
  - IPD at large
  - Ind Fusion Gang Task Force
- September 2011:
  - Gang detectives across Indiana
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
USCG: 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
Risk-Based Allocations

• Comparative visual analysis of mission cases/hours vs. staffing hours
• Comparative visualization of resources vs. risk
• Trend visual analytics
  • Increase/decrease in resource allocation
  • Increase/decrease in risk (total, mitigated, residual)
  • Increase/decrease in incidents
• Exploration of alternatives and effect on risk
• Predictive analytics based on historical data (STL and EWMA)
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 and recommendations to Congress
• Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area
RiskVA: Key Credit Risk Analytics Challenges

- Consumer credit data is large, temporal, and related across multiple investments and financial markets.
- The data are heterogeneous, not clean, have missing values, may be misleading and inefficient to explore.
- The data contains important behaviors and relations/groupings that change over time.
- Analysts need to have customer-centric interactive systems to achieve full analysis of risks.
RiskVA Overview: Interactive Exploratory Visualizations

Identification of Emerging Risks

Rule 1: Bank A & Market C & Credit Score > 740
Rule 2: Bank A & Market C & Credit Score 680-739
Rule 3: Bank A & Market C & Credit Score 620-679
Rule 4: Bank A & Market C & Credit Score < 620

Investments in dollar amount

Indicator for market stability
WireVis – Streaming Data - Multiple Linked Views

- Temporal, geospatial, theme, cluster, list views with association linkages between views

Search by Example (Find Similar Accounts)

Keyword Network (Keyword Relationships)

Strings and Beds (Relationships over Time)
Example Work in Cybersecurity

- Corporate Insider Threat Detection (Oxford, Leicester, Cardiff)
- Sensor Forensics (Purdue)
- SemanticPrism (Purdue)
- Multiscreen, Multiview Interactive Cyber Investigation (VaTech, PNNL)
- Log Visualization (Purdue)
Visual Analytics for Effective Planning, Analysis, and Decision Making

David S. Ebert
Who We Are & What We Do
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
# VACCINE Multidisciplinary Team

<table>
<thead>
<tr>
<th>Principal Partners</th>
<th>Areas of Expertise/Core Capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purdue University</td>
<td>Visualization of Structured, Unstructured and Streaming Data; Scalable Filtering and Dissemination; Video and Image Processing; Mobile, Lightweight Information Analytics; Public Safety Operations; Undergraduate/Graduate Education; Professional Training</td>
</tr>
<tr>
<td>Georgia Institute of Technology</td>
<td>Information Visualization; Visual Analytics; Peripheral Awareness; Undergraduate/Graduate Education;</td>
</tr>
<tr>
<td>Pennsylvania State University</td>
<td>Spatial cognition; Geo-information Representation, Spatial Analysis, Cartography; Undergraduate/Graduate Education; Professional Training</td>
</tr>
<tr>
<td>University of NC-Charlotte</td>
<td>3D Multimodal Interaction; Bioinformatics Visualization; Virtual Environments; Visual Reasoning; Interactive Visualization of Large-scale Information Spaces; Undergraduate/Graduate Education; Professional Training</td>
</tr>
</tbody>
</table>

**Extended Partner Network**

- Florida International University*
- University of Houston-Downtown*
- University of Texas, Austin
- California State University Dominguez Hills*
- Morgan State University*
- Virginia Polytechnic Institute and State University
- Navajo Technical College*
- Simon Fraser University, Canada
- University of British Columbia, Canada
- Oxford University, UK
- Swansea University, UK
- Stuttgart University, Germany
- Stanford University
- University of Washington
- Tennessee State University
- Middlesex University, UK
- Dalhousie University, Canada
- University of Manitoba
- University of Victoria
- Indiana University
- Jackson State University*
- University of Calgary, Canada
- Carleton University
- Justice Institute of British Columbia, Canada
- Ontario Institute of Technology, Canada.

* Indicates MSI
What We Do: Visual Analytics

Visual Analytics\(^1\) is the science of \textit{analytical reasoning} facilitated by \textit{interactive visual interfaces}.

People use visual analytics tools and techniques to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.
- Detect the expected and discover the unexpected.
- Provide timely, defensible, understandable assessments.
- Communicate assessment effectively for action.

\(^1\)Illuminating the Path: The R&D Agenda for Visual Analytics, Editors: Thomas and Cook
VACCINE’s Value

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 Research Themes To Solve Real World Problems

- **Visual Analytics** – integrated, interactive visual exploration, analysis, and decision making environments to enable effective decisions from massive, time-evolving multimedia data
- **Interactive scalable analytics techniques** – statistical, spatiotemporal, image, video, signal, machine learning
- **Science of Interaction** for visual analytics
- **Evaluation** in real world environments
VACCINE Research Topics/Projects

1. Big Data visual analytics
   • Both in size and complexity

2. Event/anomaly detection
   • Health and crime STL, spatial anomalies

3. Predictive analytics
   • Incorporating statistical analytical models (STL)

4. Social media analysis
   • Scatterblogs, SensePlaceII, …

5. Combining complex data (Structured, unstructured, streaming, variable)
   • Multi-source police data (Valet, iValet)
   • Multiple text sources (Jigsaw)
VACCINE Research Topics/Projects

6. Visual analytics solutions for risk-based decision making
   • US Coast Guard projects

7. Making "data science" available to non-data scientists
   • iVALET, iOPAR

8. Active Data
   • Alert lists based on unusual recent activity

9. Video analytics

10. Infield visual analytics and image analytics
    • MERGE, GARI, RosettaPhone
Topic 1: Big Data Challenges

- Heterogeneity
- Unprecedented volume and/or complexity
- Multivariate, Multidimensional
- Streaming
- Qualitative, Quantitative, Categorical, Temporal,…
- Distributed Sources: sensors, text, images, maps,…
- Inconsistent and Uncertain
- Different Scales and Cross Scale
The Curse of Dimensionality (Bellman 1961)

• Problems associated with multivariate data analysis as the dimensionality increases the available data becomes sparse
• Sparsity is a problem for any method that requires statistical significance
• In visualization, we are also limited with screen space and the number of available visual variables, so choosing the most appropriate dimensions is key
Big Data Isn’t Just Large Data!

• All of the above characteristics lead to big, complex data
• Example: A few million crime records with uncertain, missing data, dozens of fields, associated risks, related social media data, text reports, images, video...
VA Approaches to Big Data

• Don’t display all the data!!
• Extract relevant information at the appropriate natural scales using analytics and statistical models to reduce data space so that it is mappable to the visual space
• User interaction to guide above process
• Should we use big iron or cell phones?
  • Solution should adapt to the available hardware
Topic 2: Event/Anomaly Detection - STL

- Temporal trend analysis and prediction
- Seasonal Trend decomposition based on locally weighted regression (STL)
- Time series data viewed as the sum of multiple components:
  \[ Y = T + S + R \]

  \( Y \): original time series
  \( T \): trend component
  \( S \): seasonal (daily/weekly) component
  \( R \): remainder component

- Large value of \( R \) indicates substantial variation
- Filter out the noise component
- If value varies by 2 SD from STL expectation, generate Alert
Topics 2: Spatial Anomaly/Hot Spot Detection

- Use Kernel Density Estimation to generate population adaptive hotspot detection
Topic 3: Predictive Analytics

- Use STL per data item to develop predicted values with confidence intervals
- Applied to health data and crime data
- Adapting to appropriate spatial resolution for most accurate predictions
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Respiratory Count vs. Date

Dates: 1/1/2008 to 1/13/2008
Linked Predictive Crime Models by Type
Multivariate Correlative Predictive Analytics: Three Approaches

- Automatic correlation computation against lead/lags
- Temporal and spatial windowing
- Data category parameter space
Example: Drug Abuse Violations Vs. Burglaries – Focus On Geospatial Hotspots
Example: Drug Abuse Violations Vs. Burglaries – Focus on Geospatial Hotspots

- High positive correlation in same neighborhood with zero lag
Topic 4: Social Media Analytics : Real-time Twitter Monitoring and Integration into Tools
(Purdue, Stuttgart, Penn St.)

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
SensePlace2: Place-Time-Concept Analytics (PSU)

grid map of tweet frequency matching query; graduated circles depict 500 most relevant tweets; support for spatial filter by distance

ranked, sortable 500 most relevant tweets

query window

time filter & freq display

history view

selected place

http://www.geovista.psu.edu/SensePlace2/
Deep Data Analytics: Temporal Event Structuring from Streaming Media (UNCC)

Analyzing Event Stages for Occupy Movement using topic modeling and entity extraction

**Predictive Event Structuring:** Picking up key indicators (location, time, organizers) for a given topic—in this case the seminal OWS event on Sept. 17, 2011.
Events (bursts of activity for a given topic)
Example: Story of Occupy Wall Street

On-Going Event Structuring: Grouping events over related topics and over time to build a topical story
Topic 5: Complex Data

- Deep Data Analytics tool
  - Multisource news feeds and twitter feeds

- Scatterblogs
  - Twitter, Flickr, YouTube with anomaly detection and prediction
  - Developed predictive capability of each vs. event characteristics

- VALET – multisource data
  - Crime case data, dispatch data, EMS data
  - Weather data
  - Social media
  - Gang graffiti
  - Public civil/legal court data to be added

- Jigsaw

- Cybersecurity VA
VALET Overview

Time Series View

Map View

Calendar View

Time Slider

Twitter monitoring

Clock View

Menus
Jigsaw: Visual Analytics for Investigative Analysis

John Stasko

Documents in

Knowledge out
Example Work in Cybersecurity

- Corporate Insider Threat Detection (Oxford, Leicester, Cardiff)
- Sensor Forensics (Purdue)
- SemanticPrism (Purdue)
- Multiscreen, Multiview Interactive Cyber Investigation (VaTech, PNNL)
- Log Visualization (Purdue)
Topic 6: Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
USCG: Effective Risk-based Decision Making and Resource Allocation Visual Analytics

• Evaluate current and historical mission area:
  • Demands
  • Risks (total, mitigated, residual)
  • Resource allocation
  • Return on investment

• Evaluate courses of action

• Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Used for response to SuperStorm Sandy
• Hurricane Irene resource allocation decision based on cgSARVA analysis and visualization
• Informed Commandant’s budget testimony and recommendations to Congress
• Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example: Risks and Consequences From Sandy: SAR Cases November 2011 NJ/NYC Area
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)
Topic 7: Non Data Scientists
Interactive Operational Performance Assessment Report (iOPAR)

• **OPAR**
  - Standard report within the Coast Guard Business Intelligence (CGBI) system
  - Displays resource use and performance by core CG mission areas

• **iOPAR** – interactive iPad version to allow interactive visualization and analysis and inform decision making
  - Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
iOPAR Alpha Version: Spatiotemporal View

- May 2012
Chemical release plume modeling identifies census tracts with the highest number of expected people affected.
Topic 8: Active Data

• Incorporating automatic predication and relation algorithms to incoming data
• Exploring methods for visualizing/alerting user:
  • Passive visualizations on screen (e.g., related tweets, related incidents and how related)
  • Periodic alerts based on some criteria
    • Top 10 crimes that occurred in areas with unusual volume
    • All crimes that occurred in anomalous (>2SD) areas
Topic 9: Video and Sensor Analytics
Sensor Forensics – Based on Signatures

• Forensic characterization
  • Observe device output ➔ which device produced it?
  • Exploit how the device “makes” its output

• Device authentication
  • Performed using forensic characterization
  • Identify device type, make, model, configuration
  • Can the sensor be trusted?

• Detection of data forgery or alteration

• Fingerprint and trace

• Devices: Printers, Cameras, Scanners, Sensors
  Nodes, RF Devices
Visual Surveillance

• Video contains enormous information
• Systems “easy” to obtain/install
• Allows monitoring from a remote location
• Passive operation, traditionally a forensic tool
• Manned operations prone to human error
  • Fatigue, distraction, information overload
• Automated surveillance systems:
  • Model/understand/summarize the scene
  • Assist human operator by filtering information
Visual Surveillance

**Object level**
- Tracking
- Combine
- Detection

**Semantic level**
- **Persons**
  - Gesture, Activity
  - Trajectory
  - Crowd, Group

- **Vehicles**
  - Static
  - Motion
    - Make, Tire, Type
    - Trajectory, Velocity
    - Bounce

**Behavioral level**
- Anomaly detection
- Anomaly detection

**Miscellaneous**
- Data collection
  - Public datasets
  - New video
  - CCTV video
- Mobile devices
- Multi-camera

February 2013
Visual Surveillance Topics:

• Moving object detection and tracking
• Video-analysis based video summarization
• Crowd Density and Flow Estimation
  • Crowd flow, congestion, unusual patterns
Topic 10: In-Field Analytics: We Have an App for That!

- iVALET
- Evacuation Planning
- Rosetta Phone
- Hazmat app
- Gang graffiti app
- Tatoo app
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 60 users and 500 graffiti images acquired

**GARI delivered:**
- Summer 2011:
  - IPD gang detectives
- August 2011:
  - IPD at large
  - Ind Fusion Gang Task Force
- September 2011:
  - Gang detectives across Indiana
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Visual Analytics for Financial Analytics, Fraud, and Decision Making

David S. Ebert
Who We Are:
International Team of Experts
- 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria

VACCINE Multidisciplinary Team

- Experts in computer science, image analysis, signal processing, cognitive science, decision theory, management science, geography and GIS, HCI, visualization, HPC, statistics, political science …

- Partners in nursing, public health, economics, databases, command and control, emergency management
VACCINE’s Value

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

Engaged End-Users

- Federal Operating Components:
  - US Coast Guard
  - US Transportation Security Agency
  - US Federal Emergency Management Agency
  - US Customs and Border Patrol
  - US CERT
- Law Enforcement
  - Over 40 local and state agencies
- Fusion Centers
  - Ohio and Indiana
Example Projects

- Financial Visual Analytics
- TSA
  - Integrating staffing, financial, performance metric, legal data into a unified VA system
- Law Enforcement
- USCG – Risk-based decision making

The Pipeline for Financial Anomaly Analysis

1. Identify
2. Prioritize
3. Investigate
4. Interactive Visualization
5. Google
6. Report
WireVis – Streaming Data - Multiple Linked Views

- Temporal, geospatial, theme, cluster, list views with association linkages between views

### Competitive Business Intelligence Based On Point of Sales Data

- Characteristics of point of sale data
  - Multivariate
    - Large # of dimensions
      - E.g., 38 categories/products in 288 stores
    - Temporal: 18 months
    - Spatial
      - Stores located all over Costa Rica
  - Requirements
    - Supporting easy comparison among companies
      - E.g., visualize all data at once, sorting by importance
      - Enable geographical comparison
    - Easy recognition of any change in sales
      - E.g., proportional legends
    - Forecast for decision-making
FinVis: Applied Visual Analytics for Personal Financial Planning

- First visual analytics work to address portfolio planning for casual users
  - General public could benefit from this tool
- Decisions are better when using FinVis
  - Quantified using experimental economics
- Improved
  - Decisions
  - Learning
  - exploration,
  - confidence
RiskVA: Key Credit Risk Analytics Challenges

- Consumer credit data is large, temporal, and related across multiple investments and financial markets.
- The data are heterogeneous, not clean, have missing values, may be misleading and inefficient to explore.
- The data contains important behaviors and relations/groupings that change over time.
- Analysts need to have customer-centric interactive systems to achieve full analysis of risks.

RiskVA Data Integration
RiskVA Overview: Interactive Exploratory Visualizations

Identification of Emerging Risks

Sensor Forensics – Based on Signatures

- Forensic characterization
  - Observe device output ➔ which device produced it?
  - Exploit how the device “makes” its output

- Device authentication
  - Performed using forensic characterization
  - Identify device type, make, model, configuration
  - Can the sensor be trusted?

- Detection of data forgery or alteration
- Fingerprint and trace
- Devices: Printers, Cameras, Scanners, Sensors
  Nodes, RF Devices

February 2013
Law Enforcement

- Law enforcement visual analytics
  - VALET, iVALET (iPhone/iPad), CrimeViz
    - Visual Analytics Law Enforcement Toolkit
    - Analyzing crime patterns and time of day problems
  - Gang activity analytics
    - GARI
      - Gang Graffiti Recognition and Interpretation using a mobile telephone
      - Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
  - Document visual analytics
    - JIGSAW
      - Visualization for investigative analysis
      - Discovery of hidden relationship and threats across documents

Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

Impacts:
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (September 2012) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview

Menus

Time Series View

Clock View

Twitter monitoring

Map View

Time Slider

Calendar View

February 2013

iVALET

• Explore criminal, traffic and civil data on-the-go
• Risk assessment
• Use current spatial + temporal context into analysis
We Have an App For That!

Our Mobile Tools

• VALET
• Evacuation Planning
• Rosetta Phone
• Hazmat app
• Gang graffiti app
• Tatoo app

Gang Graffiti Recognition and Analysis Using a Mobile Telephone (GARI) - There’s An App For That

• What is GARI?
  • A mobile device application that analyzes gang graffiti
• How does GARI work?
  • User takes an image of the graffiti
  • User receives an analysis
    • Geographic locations
    • Graffiti colors, shapes, meaning
    • Database search of similar tags
GARI On The Street

- Over 50 law enforcement officials are using or field testing GARI
- Delivered to:
  - IMPD gang detectives (Summer 2011)
  - IMPD at large, Indiana Fusion Gang Task Force (August 2011)
  - Gang detectives across Indiana (October 2011)
- Homeland Security applications

GARI System Overview

1. Offline automatic analysis and labeling:
   - Decomposition
   - Date and time
   - Extracted Features
2. Filtered results Info + thumbnails
3. Original Database
4. Filtered Database
5. Offline manual filtering
6. Manual labeling:
   - Additional Features
7. Addition to Database
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis

USCG: Effective Risk-based Decision Making and Resource Allocation Visual Analytics

- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level
U.S. Coast Guard Search and Rescue VA (cgSARVA)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:
• Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
• Used for resource allocation for SAR
• Provided new insights to SAR mission
• Superstorm Sandy: Used for resource allocation in response and in rebuilding
• Used for Hurricane Irene resource allocation decisions
• Informed Commandant's budget testimony and recommendations to Congress
• Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
• Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

Example Screenshot: cgSARVA
Risk-Based Allocations

- Comparative visual analysis of mission cases/hours vs. staffing hours
- Comparative visualization of resources vs. risk
- Trend visual analytics
  - Increase/decrease in resource allocation
  - Increase/decrease in risk (total, mitigated, residual)
  - Increase/decrease in incidents
- Exploration of alternatives and effect on risk
- Predictive analytics based on historical data (STL and EWMA)

Response Efficiency – Current Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
Response Efficiency – Potential Future Assets

1-station (90-min response)
2-station (90-min response)
3-station (90-min response)
4-station (90-min response)

February 2013

For Further Information
www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Visual Analytics for Risk-Based Decision Making: Anytime and Anywhere

David S. Ebert
Purdue University
Abish Malik, Silvia Oliveros, Ahmad Razip, Sungahn Ko, Yang Yang
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (September 2012) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police

Map View
- Time Series View
- Clock View
- Twitter monitoring
- Map View
- Menus
Anytime and Anywhere

• Adaptation to mobile technology
• On-device analysis and data collection
• Real-time data analysis and integration (e.g., weather, resources, Twitter, Youtube, Flickr)

Examples:
• Law enforcement and public safety
  - VALET/iVALET, GARI, MERGE
• Maritime domain
  - cgSARVA
• ORAM
• iOPAR
MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected
Summary

• Advanced analytics running at real-time rates
• Real-time date feeds and analysis
• Scenario and course of action planning
• Integration of multiple mobile tools for decision making

• Now let’s look at maritime applications
USCG: 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
Risk-Based Allocations

- Comparative visual analysis of mission cases/hours vs. staffing hours
- Comparative visualization of resources vs. risk
- Trend visual analytics
  - Increase/decrease in resource allocation
  - Increase/decrease in risk (total, mitigated, residual)
  - Increase/decrease in incidents
- Exploration of alternatives and effect on risk
- Predictive analytics based on historical data (STL and EWMA)
U.S. Coast Guard Missions VA (cgSARVA/COAST)
Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

- Resource allocation and realignment analysis
- Risk-based analysis
- Economic analysis
- Examples:
  - 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
  - Hurricane Irene resource allocation decision based on cgSARVA analysis and visualization
Mission-based Risk Analysis Capabilities

- Efficiencies in CG station coverage
  - Resource allocation for SAR
  - Temporal and spatial patterns used in planning
  - Hypothetical station closure analysis
- Operational risk analysis and assessment (total risk, mitigated risk, residual risk)
- Effects of different mitigation strategies
- Balancing costs of different decisions
Response Efficiency: Hypothetical Asset Allocation

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
Example: Risks and Consequences From Sandy:
SAR Cases November 2011 NJ/NYC Area
Jones Beach, Sandy Hook, and Manasquan Inlet cases recorded as responded by these stations.

All Cases November 2010
Jones Beach, Sandy Hook and Manasquan Inlet closest cases (not necessarily reported under them).

All Cases November 2010
Jones Beach, Sandy Hook, and Manasquan Inlet cases recorded as responded by these stations.
All Cases November 2011
Jones Beach, Sandy Hook and Manasquan Inlet closest cases (not necessarily reported under them).

All Cases November 2011
Closure for Jones Beach (with all the regular operations for the past 5 years taken into account).
Resource Allocation and Risk-Based Decision Making (e.g., ORAM)

- Explore risk-based decision making and utilize historical data for analysis and prediction
- Total Risk, Mitigated Risk, Residual Risk
- Explore 11 different USCG missions
- Explore allocation of assets w/ different capabilities
- Explore staffing, utilization, assets vs. risk
- Perform What-If scenarios
ORAM Risk Pie Graph by Mission Type

- SAR Residual
- SAR Total
- MARSAFE Residual
- AMIO Residual
- AMIO Total
- CD Residual
- CD Total
- OLE Residual
- MEP Residual
- LMR Residual
- LMR Total
- ATON Residual
- ATON Total
- DOMICE Residual
- DOMICE Total

Controls and Filters:
- D1
- D5
- D7
- D8
- D9
- JATFS-CARIB
- D11
- D13
- D14
- D17
- JATFS-EPAC
ORAM Parallel Coordinate Analysis
ORAM Top 2 Districts based on Residual SAR Risk

Top 2 Residual SAR districts highlighted (D7 and D9)
Interactive Operational Performance Assessment Report (iOPAR)

• OPAR
  • Standard report within the Coast Guard Business Intelligence (CGBI) system
  • Displays resource use and performance by core CG mission areas

• iOPAR – interactive iPad version to allow interactive visualization and analysis and inform decision making
  • Analyze/visualize performance, targets, seasonal trends, predictions by boat, aircraft, cutter type, region levels
iOPAR Background

• Provide an interactive, mobile display of OPAR information
• Enable real time or near real time assessment of hourly expenditure and performance
• For senior level decision makers to have access to OPAR information while away from a workstation
• Assist operational decision making
iOPAR goals

• Provide interactive filtering tools to filter data by:
  • Resources and subclasses (e.g., boats, aircrafts, cutters, types of boats)
  • Region (e.g., area, district, sector)
  • Mission area (e.g., SAR, MEP)

• Display hourly expenditure across missions, regions, resources
  • Drill down using interactions

• Measure performance

• Analysis capabilities
  • Seasonal trends
  • Predictive utilization rates
iOPAR

- iOPAR map & time series view
- 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)
iOPAR Grouped Time Series

- Group time periods (months, quarters) and compare across years.
- Here, looking at how each quarter uses the resources across years.
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.

Pandemic spread on day 37 with no decision measures implemented

Pandemic spread on day 37 with all decision measures implemented
Background on Our Work

Purdue leads the Visual Analytics for Command, Control, and Interoperability Environments (VACCINE) Center of Excellence in Visualization Sciences for the U.S. Department of Homeland Security

26 schools across the globe

Research, education, and engagement to solve real world problems
Definition

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

People use visual analytics tools and techniques to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.
- Detect the expected and discover the unexpected.
- Provide timely, defensible, understandable assessments.
- Communicate assessment effectively for action.

1. Illuminating the Path: The R&D Agenda for Visual Analytics, Editors: Thomas and Cook
Spatiotemporal Proactive and Predictive Visual Analytics

Most activities involve detecting patterns, clusters, and anomalies in space, time, people, and events.

Existing analytical methods remain poor for identifying real-world place-time features as well as for understanding the space-time dynamics.

The methods are also infeasible computationally for most real-time scenarios.

 Syndromic Surveillance data: Emergency Department complaints reported to the Indiana Public Health Emergency Surveillance System.
Syndromic Surveillance

- Syndromic surveillance is the detection of adverse health events focusing on pre-diagnosis information to improve response time.
- Disease surveillance enables public health officials and their partners to achieve such purposes by monitoring changes in the population’s health status.
- Pre-diagnosis information can consist of multiple data sources:
  - *Over the counter medicine sales*
  - *News reports on emerging diseases*
  - *Pro-med news feeds*
  - *Emergency department chief complaints*
Syndromic Surveillance

• How do we conduct early identification of adverse health events with high potential to cause high morbidity and mortality?

• In response to the need for earlier recognition of significant health events, public health institutes have developed modern surveillance applications based on the world wide web.

Basic ESSENCE Functions:

• Early event detection (Alert List):
  • Analyzes time-sensitive data for the purpose of detecting and flagging outbreaks as early as possible.

• Situational awareness (Query Portal):
  • Queries data sources to obtain size, location, spread of an imminent health event.
  • Tracks ongoing health events to assess impact in terms of time, geography, and demography.
Alert Generation

- Alerts are detected by ESSENCE using specially developed statistical algorithms.
- **Red flags**: the difference in count and expected frequency to be statistically significant at $p$-value $\leq 0.01$.
- **Yellow flag**: difference at statistical significance of a $p$-value between 0.01 and 0.05.
Data Sources

• Data sources:
  • Emergency department data
  • Over-the-counter drug sales (non-traditional data source)

• Why does this work?
  • In USA, Emergency Departments are the primary care provider for the uninsured population
ESSENCE Syndromes

Syndromes (subsyndromes):
- Respiratory (cough, pneumonia, influenza)
- Gastrointestinal (vomiting, diarrhea)
- Neurological (meningitis, altered mental status, dizzy)
- Fever
- Rash (vesicular rash, chicken pox)
- Botulism-like (weakness, blurred vision, speech)
- Shock/Coma (syncope)
- Hemorrhagic Illness (bleeding)
Primary Benefits of ESSENCE for Hospitals

• Offers excellent “slicing and dicing” capabilities for analysis.
• Delivers valuable hospital view of data for Infection Control Practitioners.
• Allows collaboration with public health while protecting hospital/patient confidentiality.
Primary Benefits for Public Health Officials

• Early event detection.
• Outbreak case investigation and follow-up management.
• Assessing impacts of natural disasters or severe weather.
• Exposure contact tracing.
• Exposure source investigation and linking of cases and contacts to exposure sources.
• Good to know that no alerts have been generated.
Indiana’s Public Health Emergency Surveillance System (PHESS)

- 77 Emergency Departments
- 6,500 – 7,000 visits per day ~ 15 MB data per day
- Records contain home address, age, gender, chief complaint and syndrome
  - 8 Syndrome Categories:
    - Respiratory
    - Gastro-intestinal
    - Neurological
    - Constitutional
    - Hemorrhagic
    - Botulinic
    - Rash
    - Other

4 Patients!
### Patient Information

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Date</th>
<th>Chief Complaints</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>9398</td>
<td>4/16/09</td>
<td>Ear pain</td>
<td>Karachi</td>
</tr>
<tr>
<td>10816</td>
<td>4/16/09</td>
<td>Stuffy nose</td>
<td>Lebanon</td>
</tr>
<tr>
<td>1491</td>
<td>4/16/09</td>
<td>Fever</td>
<td>Allepo</td>
</tr>
<tr>
<td>16237</td>
<td>4/16/09</td>
<td>Head bleed</td>
<td>Yemen</td>
</tr>
</tbody>
</table>

**CoCo Classifier**

Improving Syndromic Surveillance

Interactive visual analytic environment for effective syndromic surveillance and response

- System designed based on collaboration and feedback with state epidemiologists
- Integrated temporal, geospatial, multi-source, multi-scale analytic capability
- Density estimation for data exploration
- Syndromic control charts for temporal alerts
- Demographic filter controls for advanced analysis
Improving Syndromic Surveillance

Benefits/ impact

- Enhanced hypothesis testing capabilities
- Linked views allow quicker cross validation of hypothesis
- Less time investigating false positives
- Systemic biological pandemic, syndromic, chem/bio/nuclear surveillance, management, and response
Spatiotemporal Hypothesis Generation and Visual Analysis
Syndromes

Lag = 21 days
Mortality rate = 8.864%
Infection rate = 23.591%

Hemorrhagic percentage = 7.286%
Neurological percentage = 2.914%
Rash percentage = 1.457%
Respiratory percentage = 1.639%
Other percentage = 36.521%
Situational Surveillance and Predictive Visual Analytics

- Focus is on categorical spatiotemporal event data
- Utilizing time series and density estimations we want to create an interactive environment for predicting future event magnitudes and locations
- We utilize seasonal trend decomposition with Loess smoothing
- 3D Kernel density estimation for spatiotemporal probability distributions
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Black line: Actual
- Red line: Predicted
- Dotted red line: Lower
- Dotted red line: Upper

Date:
- 1/1/2008
- 1/3/2008
- 1/5/2008
- 1/7/2008
- 1/9/2008
- 1/11/2008
- 1/13/2008

Respiratory Count:
- 0
- 10
- 20
- 30
- 40
- 50
- 60
Advanced Decision Support Tools: Rift Valley Fever
Modeling a Pandemic

- **Pandemic Influenza Planning Tool**
- **Models user specified:**
  - Pandemic influenza characteristics
  - County population, demographics, hospital beds
- **Decision measures**
  - Strategic National Stockpile deployment
  - School Closures
  - Media Alerts

Map of the United States showing the estimated percent of the population ill based on a simulated pandemic influenza model originating in Chicago, IL.
Case Study – Pandemic Influenza

Decision Path D4

SNS Day 2

School Closure Day 6

Media Day 25
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
Time Series Modeling

- Seasonal-Trend Decomposition Based on Loess
  - Time series can be viewed as the sum of multiple trend components
  - For each data signal, components are extracted
  - Can then analyze correlation between components
Advanced Decision Support Tools: Rift Valley Fever

Decision Path minimized

Best Decision Path

Best components of path are shown

Highlighted decision path

Lives Saved

Decision Path Fully Expanded

Two Paths giving similar mortality savings

Lives Lost

Day #
Who We Are: International Team of Experts - 75+ Faculty, 25 institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea, UK
- Oxford
- University of Calgary
- University of Manitoba
- Carleton University
- University of Victoria
VACCINE Benefit

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
Engaged End-Users

• Federal Operating Components:
  • US Coast Guard
  • US Transportation Security Agency
  • US Federal Emergency Management Agency
  • US Customs and Border Patrol
  • US CERT

• Law Enforcement
  • Over 40 local and state agencies

• Fusion Centers
  • Ohio and Indiana
What We Do in a Nutshell

• Help people find relevant, useful information from massive, messy, conflicting, complicated data
• Enable effective communication of information
• Provide quantitative, reliable, reproducible evidence
• Enable user to be more effective from planning to detection to response to recovery
Example Projects

• TSA
• Law Enforcement
• USCG
Law Enforcement

- Law enforcement visual analytics
  - VALET, iVALET (iPhone/iPad), CrimeViz
    - Visual Analytics Law Enforcement Toolkit
    - Analyzing crime patterns and time of day problems
- Gang activity analytics
  - GARI
    - Gang Graffiti Recognition and Interpretation using a mobile telephone
    - Allowing police to catalog and analyze gang graffiti images, better track and determine gang activities
- Document visual analytics
  - JIGSAW
    - Visualization for investigative analysis
    - Discovery of hidden relationship and threats across documents
Temporal Predictive Analysis

- Temporal trend analysis and prediction:
  - Seasonal Trend decomposition based on \textit{loess} (locally weighted regression) (STL)
  - Time series data viewed as the sum of multiple components
  - Filter out the noise component
  - Predict in the future
Predictive Visual Analytics

Sample Emergency Department - Predicted vs. Actual

- Actual
- Predicted
- Lower
- Upper

Date:
- 1/1/2008
- 1/3/2008
- 1/5/2008
- 1/7/2008
- 1/9/2008
- 1/11/2008
- 1/13/2008

Respiratory Count:
- 20
- 30
- 40
- 50
- 60
Visual Analytics Law Enforcement Toolkit (VALET, iVALET)

**Impacts:**
- In use to analyze crime patterns in Lafayette, Indiana and to connect strings of activities
- Mobile version being released to public (September 2012) for community-based policing
- Investigating correlation of bus routes and crime, street lights and crime
- Analyzing time of day problems and improving accuracy of police record management system
- Novel statistical predictive model incorporated for planning

VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police
VALET Overview

- Map View
- Time Series View
- Calendar View
- Time Slider
- Clock View
- Twitter monitoring
- Map View

Menus
Linked Predictive Crime Models by Type

Day vs. Night Thefts

2008 (red) vs. 2007 (blue background)
Example: Drunkenness / Public Intoxication
Example: Drunkenness / Public Intoxication

- PU vs. Notre Dame: Home Loses 10-38
- Homecoming (Sat.): PU vs. Illinois: Home Wins 21-14
- PU vs. Illinois: Away Loses 21-31

Day-of-the-Week

Home and Away Games
- Home games are marked with a red circle,Away games are marked with a red triangle.
VALET Feature: Real-time Twitter Monitoring and Integration

- Topic extraction using novel STL based remainder estimation technique
- Dynamically linked views providing options to monitor emerging / emergent twitter feeds
- Topics extracted shown as a dynamic word cloud
Correlation b/w Multiple Social Media (1/2)

• Occupy Wall Street
  • Protest movement starting on September 17, 2011, in New York City
• Utilization of social media
  • Communication and reports about the movement in forms of text, images and videos
Correlation b/w Multiple Social Media (2/2)

-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00

9/14
9/19
9/24
9/29
10/4
10/9
10/14
10/19

Date

Abnormality z-score

Twitter
Flickr
YouTube

: periods of actual events during the protests
### Discussion

- **Different temporal trends according to event types**

<table>
<thead>
<tr>
<th></th>
<th>Abrupt short event (e.g., Earthquake)</th>
<th>Social and planned event (e.g., Occupy wall street)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Text type</strong></td>
<td>Strong change</td>
<td>Gradual increase and decrease</td>
</tr>
<tr>
<td><strong>Twitter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multimedia type</strong></td>
<td>No significant signal</td>
<td>Strong change, but delay</td>
</tr>
<tr>
<td><strong>YouTube, Flickr</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Data limitation**
  - Only geo-located messages by GPS enabled devices, **but fast increase**
  - The **most relevant group** for situational awareness scenarios
October 2012

iVALET

- Explore criminal, traffic and civil data on-the-go
- Risk assessment
- Use current spatial + temporal context into analysis
We Have an App For That!

Our Mobile Tools

- VALET
- Evacuation Planning
- Rosetta Phone
- Hazmat app
- Gang graffiti app
- Tatoo app
Gang Graffiti Recognition and Analysis 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 60 users and 500 graffiti images acquired

GARI delivered:
- Summer 2011:
  - IPD gang detectives
- August 2011:
  - IPD at large
  - Ind Fusion Gang Task Force
- September 2011:
  - Gang detectives across Indiana
Gang Graffiti Recognition and Analysis (GARI)  
PI: Ed Delp
There’s An App For That

• What is GARI?
  • A mobile device application that analyzes gang graffiti

• How does GARI work?
  • User takes an image of the graffiti
  • User receives an analysis
    • Geographic locations
    • Graffiti colors, shapes, meaning
  • Database search of similar tags
Why It Works

- Catalogs and analyzes graffiti images
  - More than 500 graffiti images acquired (plus 506 for research testing)
- Search of related image
- Search of local area
- Partially interprets graffiti images
  - Colors, symbols
  - Recorded gang analyst notes
- Future analysis
  - Image meaning
  - Which gang
- Helps law enforcement track and identify gang activity throughout a region
GARI On The Street

- Over 50 law enforcement officials are using or field testing GARI
- Delivered to:
  - IMPD gang detectives (Summer 2011)
  - IMPD at large, Indiana Fusion Gang Task Force (August 2011)
  - Gang detectives across Indiana (October 2011)
- Homeland Security applications
GARI System Overview

1. Original image
   - Offline automatic analysis and labeling
     - Geoposition
     - Date and time
     - Extracted Features

2. Server
   - Filtered results
     - Info + thumbnails

3. Original Database

4. Filtered Database

5. Offline manual filtering

6. Manual labeling
   - Additional Features

7. Addition to Database

Labeled image
Project Status

- Deployed on Android Phone
- Main phone application is image database system and browsing (acquisition tool) with searching and some limited analysis capabilities

- Desktop backend
  - Browse by radius
  - Upload images
  - View and edit image details
  - Interact with map
  - Searching for similar images
Gang Graffiti Interpretation

SHAPE
Simple, Straightforward

NUMBERS
42nd street gang

SYMBOLS
6-point star, pitchforks

COLOR
Goon Squad: Red/Black
Gang Graffiti Interpretation

LETTERS
East Side Gang

POSITION/ALIGNMENT
Letters at star points
Numbers in the middle
Letters at the bottom
Pitchforks upright

TIME
Black: 18 ST (18th Street Gang)
Red: 13 SUR (Sureños 13)
Image Analysis – Color Recognition

MEXICANOS MALDITOS
SUREÑOS 13

18 STREET GANG

SUREÑOS 13
Image Analysis – Scene Analysis

GANG RIVALRY TRACKING

- Date: 08/19/2010
- Time: 3.25 PM
- Geo: 41.387917, 2.169919

- Date: 01/03/2011
- Time: 5.11 PM
- Geo: 41.387917, 2.169919
GARI App Features
Gang Graffiti – User Interface

Send to server

Analysis

Main menu

Capture Image
Browse Image
Browse Database
About
Settings

IMAGE SUCCESSFULLY UPLOADED:
File name: 000000-20110812114657.jpg
File Size: 721062 bytes
Height: 1552 px
Width: 2592 px
Date and Time: 2011:08:12 11:47:09
Camera Model: HTC Desire
Metering mode: 65535
White balance: 65535
Focal Length: 4.31 mm
Flash: 65535
Camera Make: HTC
GPS Latitude (dms): N 40d 25m 46.249s
GPS Latitude: 40.4295136111
GPS Longitude (dms): W 86d 54m 45.033s
GPS Longitude: -86.9125091667
First Responder ID: 000000
Uploaded from web: 0

5 Point Star: Symbolic to gangs within the People Nation
Gang Graffiti – Browse Database

- Browse by radius from current position
- Download images and information from server (image EXIF tags, gang related information)
  - Compare results
  - Track graffiti
- Browse map (Google Maps API)
- Network connection required
Gang Graffiti – Browse Database

Browse general results

Show graffiti on map

Inspect specific graffiti
GARI Progress and Directions

- Detective Steve Shafer, IPD adding annotations to database
- Shape and tag analysis always improving
- Transitioning to Indiana State Police / IIFC Gang Task Force
- Creating linked, satellite database and tool for Navajo Nation (using Navajo Technical College)
- Investigating use by ICE, CBP, FBI
- Two-way linkage with VALET / iVALET
How to Get The App?

- Contact gari@ecn.purdue.edu
  - We assign user IDs and initial passwords
  - Database server
Visual Analytics Uses for Risk-Based Decision Making

- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis
USCG: 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
## Effective Risk-based Decision Making and Resource Allocation Visual Analytics (including finished projects)

<table>
<thead>
<tr>
<th>Project Components</th>
<th>USCG Mission Areas</th>
<th>Level</th>
<th>Resources</th>
<th>End-users</th>
<th>Interesting questions</th>
</tr>
</thead>
</table>
| cgSARVA / COAST    | SAR, LE, others in future | Tactical | Boats, Aircraft | USCG HQ 771, USCG LANT-7 D9, D5 | • Effect of closing a boat station on SAR mission  
• Outcome of SAR case based on initiation-type  
• Cost of SAR mission by case type & station  
• Effect of next-gen boats on station utility |
| ORAM VA, iOPAR     | All                | Strategic | Boats, aircraft, personnel hours | USCG LANT-7 | Operational performance across mission areas by district and sector. Performance target evaluation  
Strategic resource reallocation evaluation |
| PROTECT (Completed) | PWCS               | Tactical | Boat patrols | USCG RDC, LANT-7, D1 | • What predictable patterns in PWCS patrols existed in Boston and does the new randomized PROTECT system remove patterns and optimize PWCS schedules? |
| Port Closure VA (Completed) | Port Security | Strategic | Ports, industry by sector, local and national | USCG RDC, LANT-7 | • Evaluate economic impact of port closures by economic sector at local and national level  
• Evaluate effect of different mitigation strategies |
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
- Hurricane Irene resource allocation decision based on cgSARVA analysis and visualization
- Informed Commandant’s budget testimony and recommendations to Congress
- Key component of USCG D9 reallocation plan for 2011-2012 based on decreased budget
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ
Example Screenshot: cgSARVA

- Interactive visual analytics of multivariate performance metrics for each unit's activities
- Interactive linked spatial temporal display, calendar view, and timeline views
Risk-Based Allocations

• Comparative visual analysis of mission cases/hours vs. staffing hours
• Comparative visualization of resources vs. risk
• Trend visual analytics
  • Increase/decrease in resource allocation
  • Increase/decrease in risk (total, mitigated, residual)
  • Increase/decrease in incidents
• Exploration of alternatives and effect on risk
• Predictive analytics based on historical data (STL and EWMA)
Response Efficiency – Current Assets

- 1-station (90-min response)
- 2-station (90-min response)
- 3-station (90-min response)
- 4-station (90-min response)
For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu
USCG PWCS PROTECT Project
Partners: USC CREATE, USCG RDC, USCG D1, USCG LANT, USCG PAC

Impact:

• Provided insight and analysis into historical security patrols in Boston
• Analyzed developing PROTECT model and provided pattern analysis to improve model during operational deployment in Boston and for improvements for deployment in NYC
• Developed end-user tool for evaluation and validation of patrol routes
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
U.S. Coast Guard
Swimmer Death Analysis

PI: Ebert, Maciejewski
End-User(s): USCG District 9
Delivered: May 2010

Impact:
• Analyzed spatial and temporal patterns of shore-based and boat-based swimmer deaths to understand death dramatic increase in D9 in Summer 2010
• Provided information and visualizations used for public information campaign 2011 and for patrols 2011
• Significant decrease in deaths in 2011

Findings:
• Swimmer deaths
  • August highest frequency
  • Late afternoon highest frequency
  • Lake Michigan (south and west shore) have high concentration

• Boating deaths
  • Fri, Sat, Sun account for almost all deaths
  • Mid July to Mid August have highest frequency (only 1 week significantly high)

• 2009-2010 from MISLE Data
  • Large increase on Mon, Thu, Fri, Sun
  • Early and late season increase

video online at http://videocast.nih.gov/summary.asp?Live=10949