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

Visual Analytics for Security, Safety, and Privacy:

Approaches, Lessons Learned, Opportunities, and Challenges

David S. Ebert

CVADA
Center for Visual Analytics and Data Analytics
A U.S. Department of Homeland Security Center of Excellence

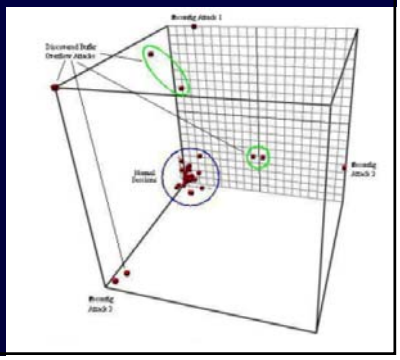
October 2013

Overview

- Background: Why am I here?
- Challenges in developing effective deployed solutions
- Approaches: which one to choose?
- Some examples and lessons
- Path forward

Why Am I Here?

- My seminal paper from VisSym 2001?
  - Atkinson, T., Pensy, K., Nicholas, C., Ebert, D., Atkinson, A., Morris, C., "Case Study: Visualization and Information Retrieval Techniques for Network Intrusion Detection," *VisSym 2001:* Joint Eurographics - IEEE TCCG Symposium on Visualization, May 2001.
  - Interactive volume visualization of network attacks projected onto know attacks
- Or for my experience leading VACCINE?
  - Different safety and security (in general)
  - Cybersecurity enters many projects



VACCINE

October 2013

Who We Are:
International Team of Experts
75+ Faculty, 26 Institutions

- Purdue University
- Georgia Institute of Technology
- Pennsylvania State University
- Stanford University
- University of North Carolina at Charlotte
- University of Washington
- Arizona State University
- Simon Fraser University
- University of British Columbia
- Justice Institute of British Columbia
- Ontario Institute of Technology
- Dalhousie University
- University of Victoria
- University of Houston, Downtown
- Virginia Tech
- Indiana University
- Florida International University
- University of Texas at Austin
- Morgan State University
- Navajo Technical College
- University of Stuttgart
- University of Swansea
- Oxford University
- University of Calgary
- University of Manitoba
- Carleton University

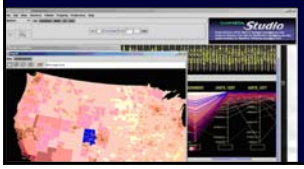
VACCINE

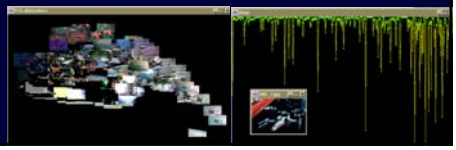
VACCINE's Role

Problem: To solve current and future homeland security problems requires exploring, analyzing, and reasoning with massive, multi-source, multi-scale, heterogeneous, streaming data –**BIG DATA**

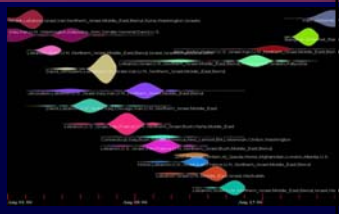
- Cuts across entire spectrum of homeland security needs

We provide tools to enable end users to get the relevant information they need during any situation to make a decision or take action





VACCINE Mission



- Provide visual analytic and scalable solutions to 2.3 million extended homeland security personnel
  - 185,000 DHS personnel, 350,000 law enforcement personnel, 750,000 homeland security practitioners
- Achieve excellence in visual analytics and visualization sciences
- Educate homeland security stakeholders and the next generation of talent

VACCINE 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 users to be more effective from planning to detection to response to recovery
- Enable effective communication of information

Approach: Partner-driven solutions and research

VACCINE Value Part II

Our people and partnerships

- Interdisciplinary world-leading team of researchers
- Defining and extending the new science of visual analytics driven by real-world, real-scale problems of engaged partners (local, state, federal)

Visual Analytic Solutions: What We Offer

- **Improved Effectiveness**: We enable users to be more effective through innovative, interactive visualization, analysis, and decision making tools
  - Provide the right information, in the right format, within the right time to solve the problem
  - Enable user to be more effective from planning to detection to response to recovery
  - Enable effective communication of information
- **Innovative Fielded Solutions:** We provide innovative visual analytic and scalable solutions to the extended homeland security community
- **People and Partnerships**

"cgSARVA has proven its worth time and again, providing key analytic information for decision makers for large scale projects…"

VADM Robert Parker, 2012 MRS Keynote Address

VADM Robert Parker with VACCINE student researchers (cgSARVA, COAST, iOPAR)

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 (IN, IL, OH, SC, PA, NC, NY)
- Fusion Centers
  - Ohio (SAIC)
  - Indiana (IIFC)

VACCINE

October 2013

Challenges in Developing Effective Deployed Solutions: Crossing the Chasm

Challenges in Developing Effective Deployed Solutions

1. Understanding the situation
   - Task/problem
     - What are they trying to find, analyze, explore?
     - What is the final product of the system and task?
   - User
     - How do they conceptualize the problem?
     - What are the natural scales/aggregation levels, features?
   - Environment - time frame, solitary vs. collaborative, equipment
   - Language - developing a common language

Challenges in Developing Effective Deployed Solutions

2. Changing requirements
   - How to be effective in an agile software development environment
   - Avoiding feature creep
   - Clear end state, goals, and deliverables

Challenges in Developing Effective Deployed Solutions

3. Trust, polices, lack of standards
   - Trust
     - Will you deliver and follow-through?
     - Or, do you just want my data?
     - What can an academic really know about what I do?
   - Polices
     - Legal agreements and delays
     - Data and privacy
   - Standards - everyone has a different schema, RMS, etc.

One Potentially Useful Approach: Application-Driven Research & Development

- A full contact sport
- Increases rate of VA advances and application deployment to effectiveness
- Increases rate of application domain advances

Our Application-Driven Research Approach and Plan

- Evolving, effective, and enduring research by tight integration with stakeholders
- Driven by stakeholders – from initiation, through iterative development (agile software development), to deployment
- Visual Analytics research integrated:
  - Interactive visual/cognitive analytic environments based on novel research in visual analytics, algorithms, information transformation, cognitive and interaction science, creating precise information environments

Full-scale exercise February 2008

Research Motivation:

- Solving these real-world problems requires
  - Novel theories, techniques, approaches, and adaptations of algorithms
  - Integration of cross-disciplinary expertise
  - Overcoming the chasm from academic idea to deployed solution
- 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

Examples of Overcoming the Chasm

- Public health syndromic surveillance
- Crime analytics
- US Coast Guard solutions

Solutions for Spatial Temporal Decision Making Environments – A Progression: The Long and Winding Road[1]

- Public health surveillance
  - Fusing apparently similar data that isn't (health data)
  - Dual domain decision making and real-world visualization and analysis for disease spread and interdiction
- Spatial and temporal visual analytics for law enforcement
- Search and rescue (SAR) and risk based visual analysis

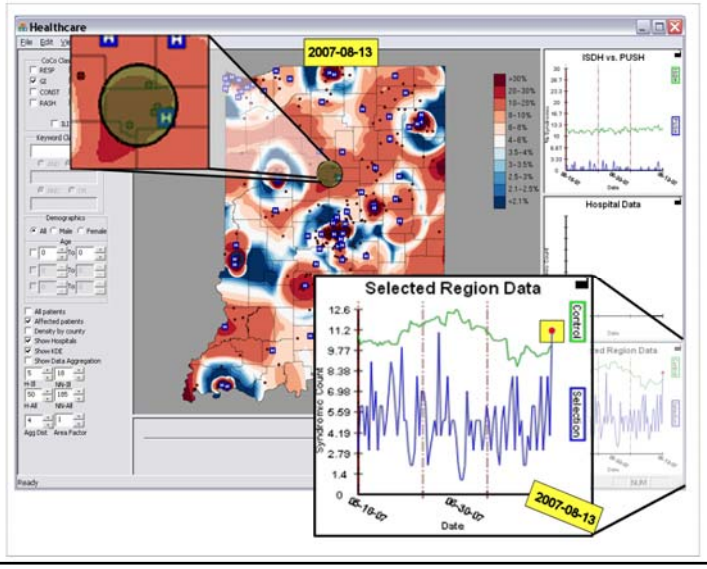
"The long and winding road
That leads to your door
Will never disappear" – P. McCartney

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

Visual Analytics for Syndromic Surveillance: Hypothesis Generation and Exploration

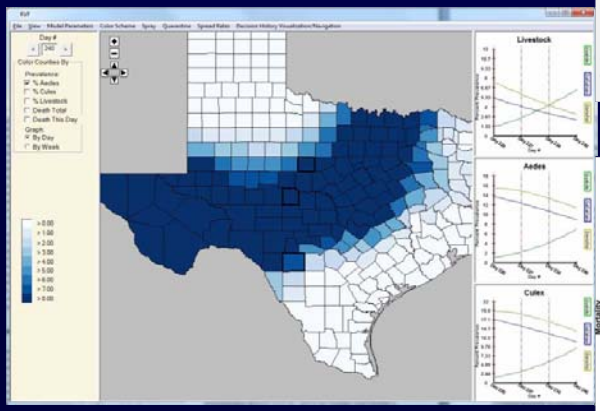


Project Design & Workflow Impetus: Indiana State Epidemiologist, EHR researcher

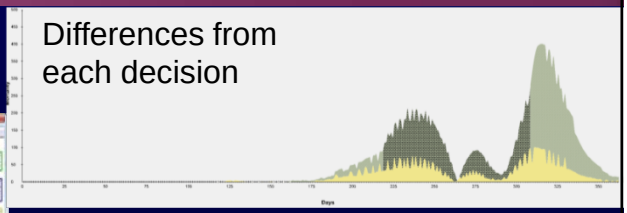
Best Paper Nominee, *IEEE Symposium on Visual Analytics Science and Technology (VAST),* October 2008, for "Understanding Syndromic Hotspots – A Visual Analytics Approach," (Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., Cleveland, W., Grannis, S., Wade, M., Ebert, D.).

Example Decision Analysis Linked Displays – Example with 3 Decisions

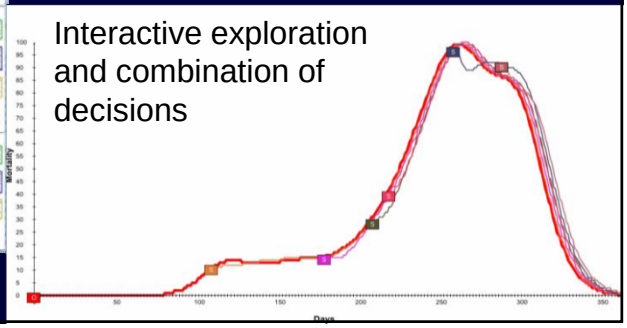
Partnership with FAZD



Differences from each decision



Interactive exploration and combination of decisions

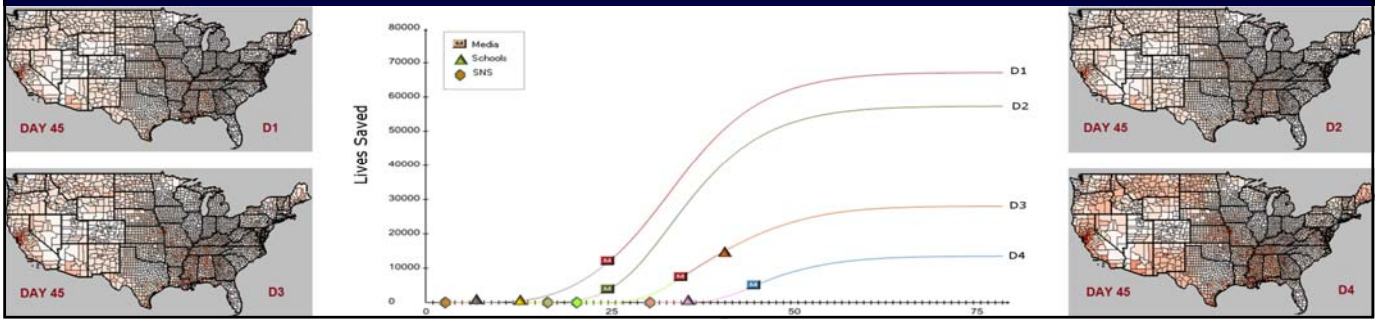


VACCINE

October 2013

Integrated Interactive Simulations and Analysis

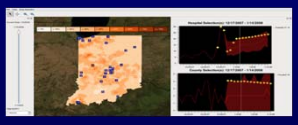
- Analysis and simulation must be interactive for integration into interactive environment
- Need novel computational & statistical models
- Goal: enable improved discovery, decision making,

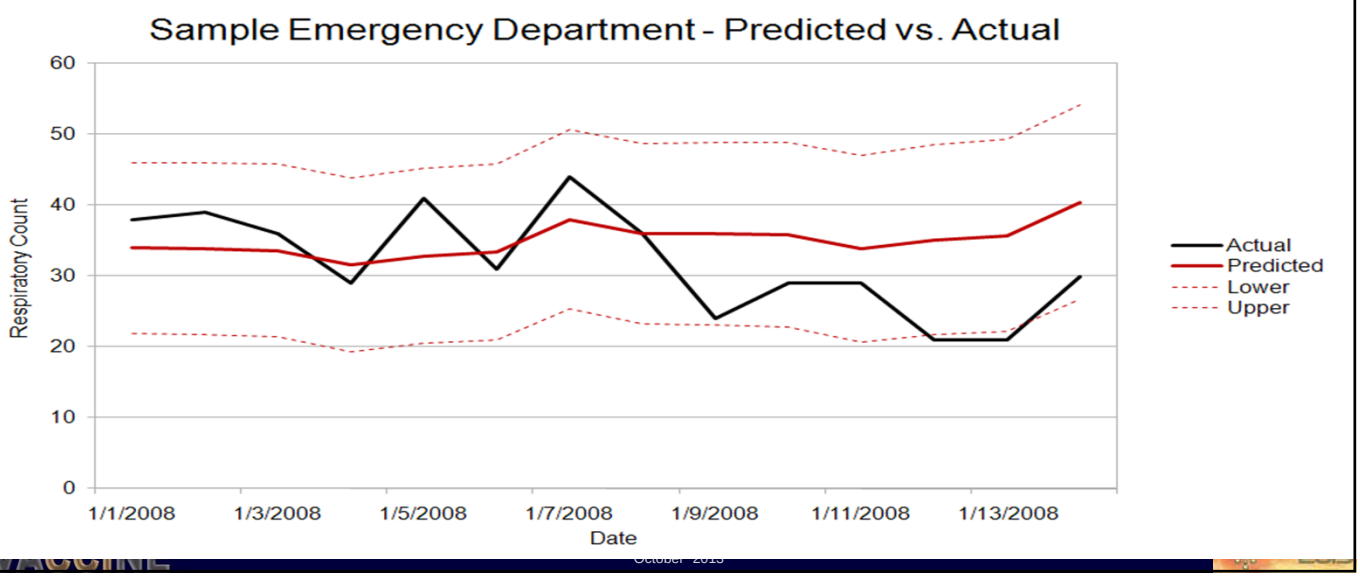


Situational Surveillance and Predictive Visual Analytics

- 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





Crime VA – The Next Curve

- Sheriff wanted to know how to use integrated data across the county to see
  - If they are being more effective
  - If crime is being reduced
  - How officers to top-level officials can use this data for proactive and predictive policing
- Frequent meetings and continuous refinement of tools
- Now being tested by agencies in 4 states
  - NYPD, OSHP, IISP, LPD, WLPD, PUPD, TCSD

Visual Analytics Law Enforcement Toolkit (i)VALET

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 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

iVALET
Debriefing Example

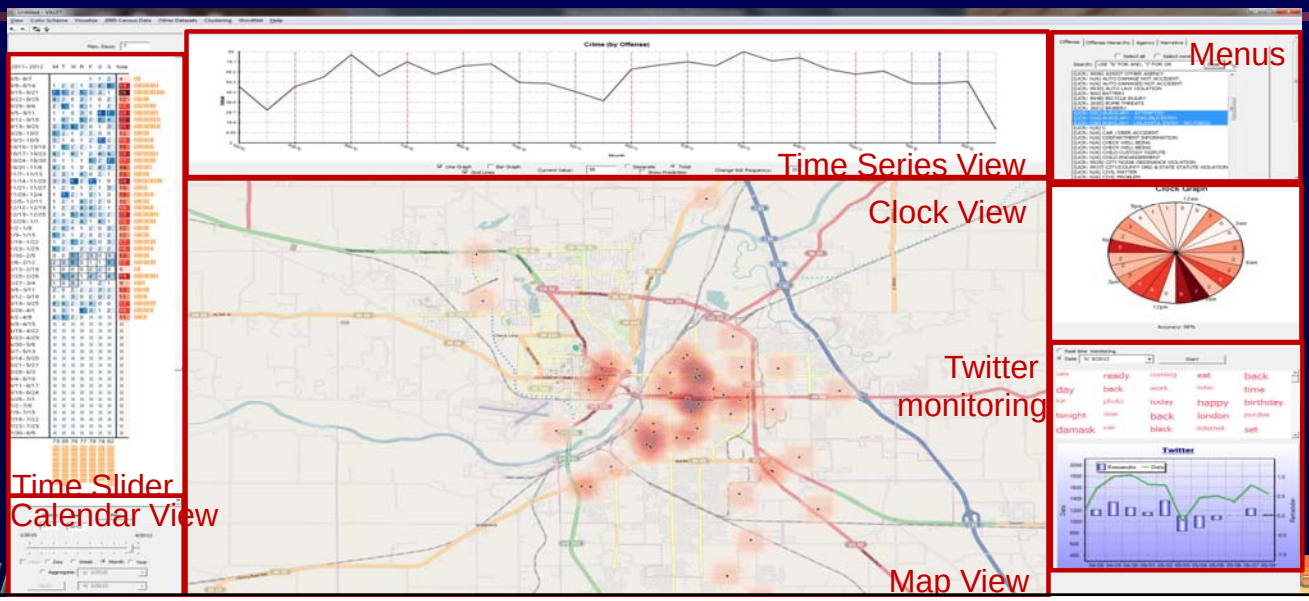
VALET delivered:
- Spring 2011: WL, Lafayette Police

iVALET delivered:
- October 2011: Purdue, WL Police

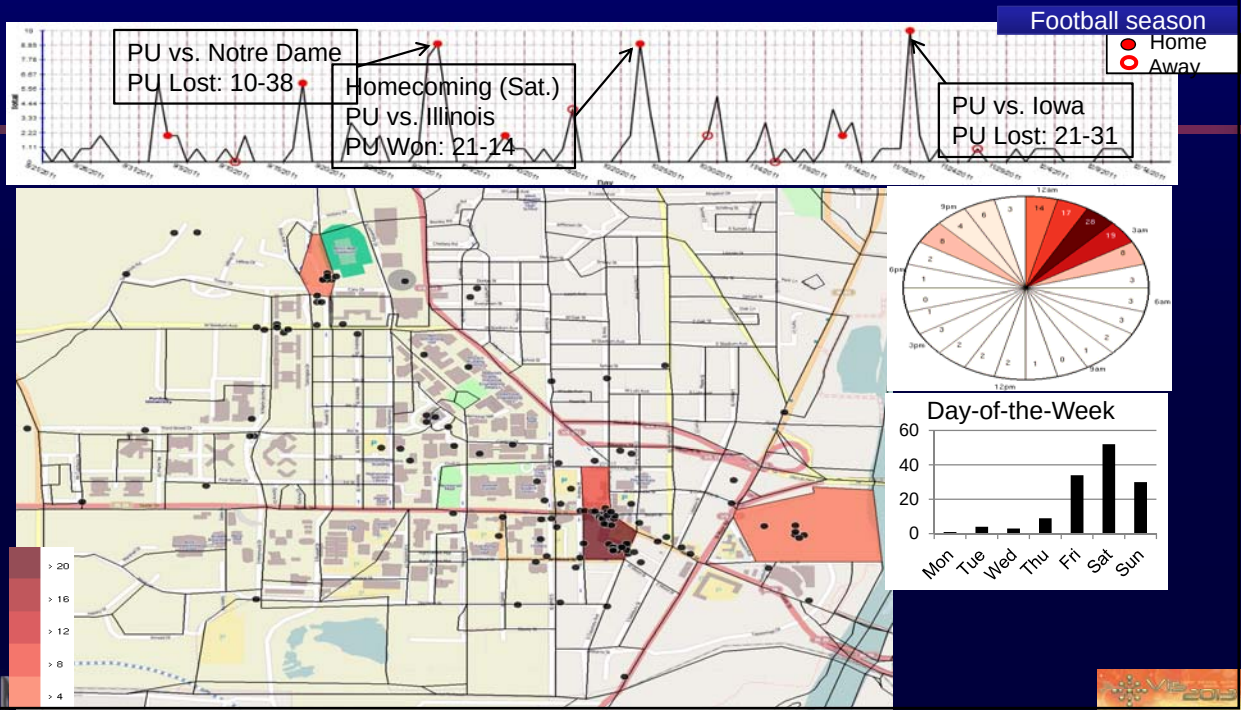
VACCINE

October 2013

VALET Overview

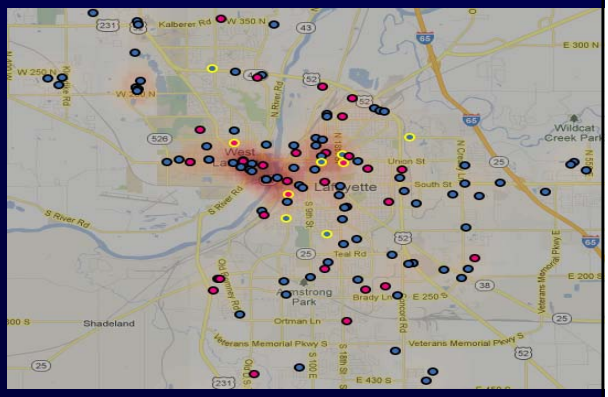


Example: Drunkenness / Public Intoxication



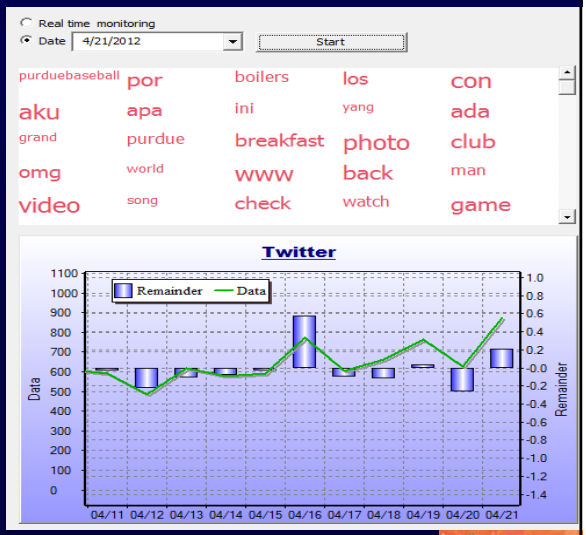
Top 10 Hot Incidents

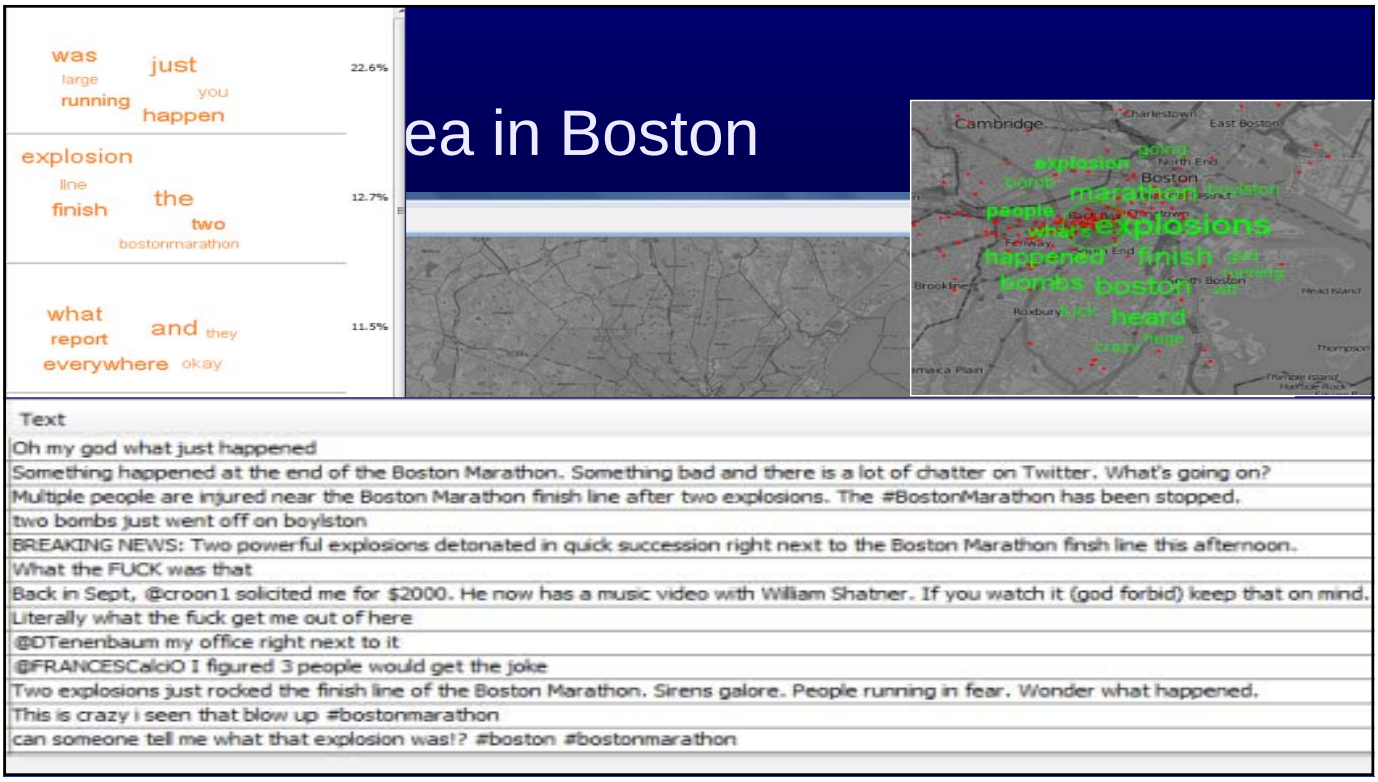
- Identify unusual localized high-frequency patterns of crimes (near repeats)
- Each data entry is checked for other crimes with similar properties within a 1 block radius of the incident location and a 14-day time period
- Top 10 incidents with the most number of related incidents in this space-time window are highlighted



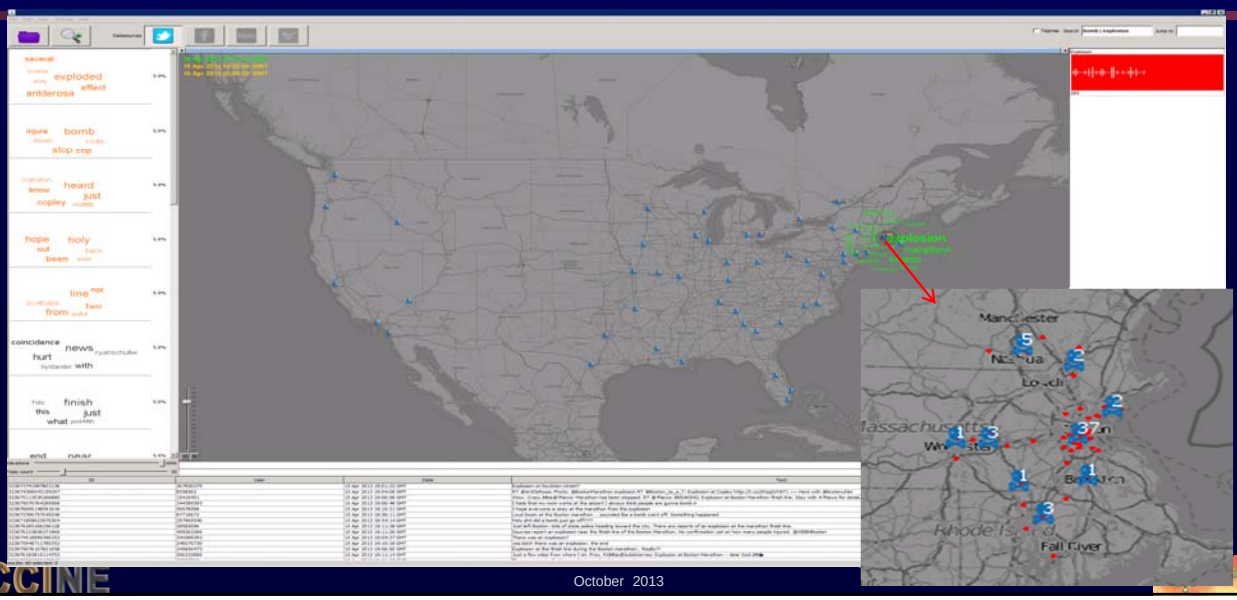
Social Media: Real-time Twitter Monitoring and Integration into Tools (Purdue, Stuttgart, Penn St.)

- 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





Detection using the Explosion Classifier



First Response (Tweet & Picture)

Visual Analytics of Activity During Hurricane Sandy

Two weeks before Sandy
10/14 (Sunday), 12:00 ~ 16:00

One week before Sandy
10/21 (Sunday), 12:00 ~ 16:00

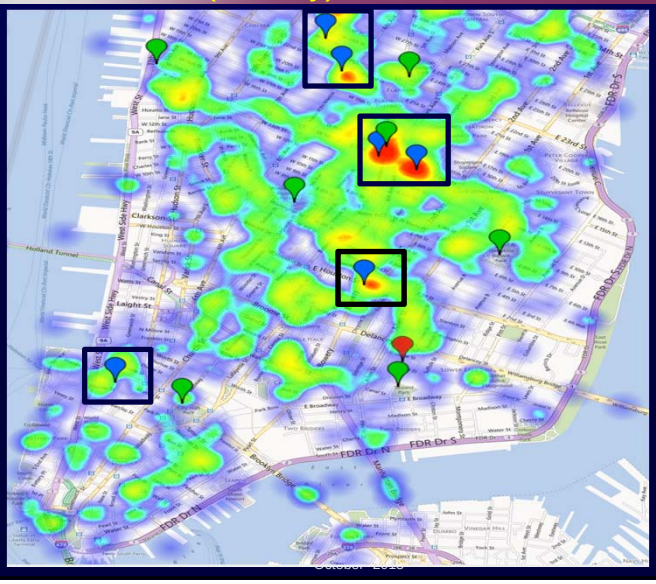
Supermarket Park Shelter

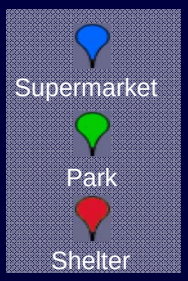
Visual Analytics of Activity During Hurricane Sandy

After the evacuation order
10/28 (Sunday), 12:00 ~ 16:00

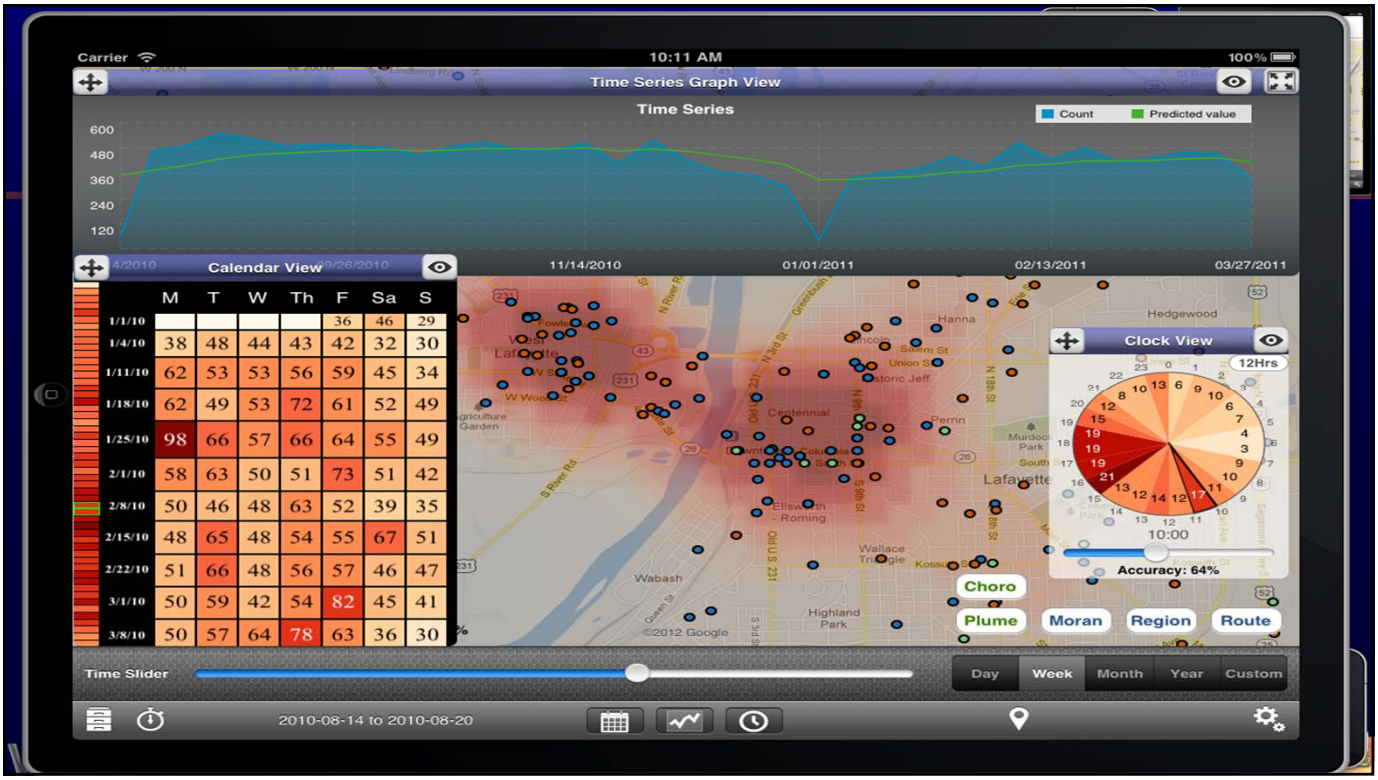
Evacuation order:
10/28, 10:30 AM

Hurricane Sandy's
Arrival at NYC:
10/29, 8:00 PM



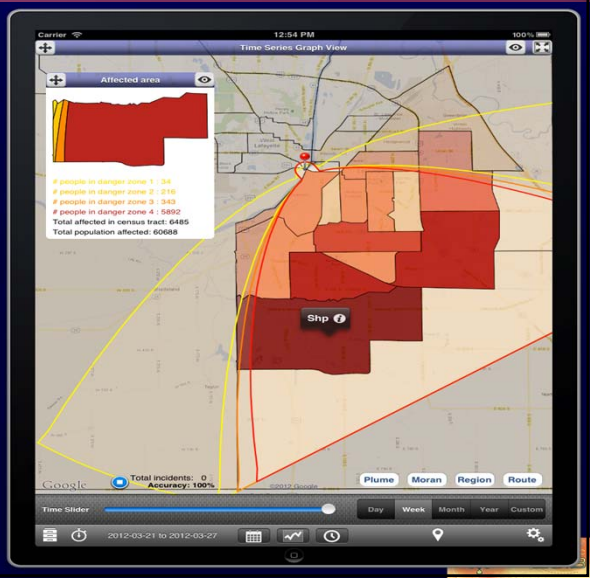


VACCINE



MERGE – iVALET Interactive Plume Visualization and Evacuation Planning

- Chemical release plume modeling identifies census tracts with the highest number of expected people affected



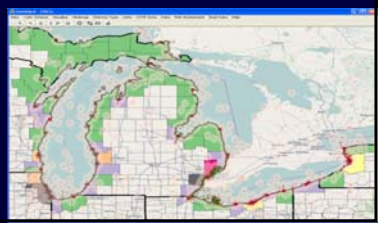
VACCINE

October 2013

The Next Bend: US Coast Guard

Visual Analytics Uses for Risk-Based Decision Making

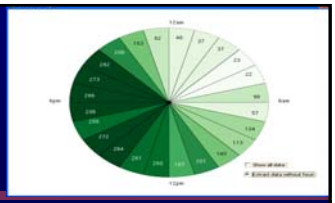
- Risk visualization and analysis
- Predictive analytics
- Uncertain decision making
- Alternative evaluation and consequence investigation
- Trend analysis, clustering, anomaly detection
- Interactive, multi-day, month, type investigation
- Multisource, multimedia data integration & analysis



USCG: Effective Risk-based Decision Making and Resource Allocation Visual Analytics

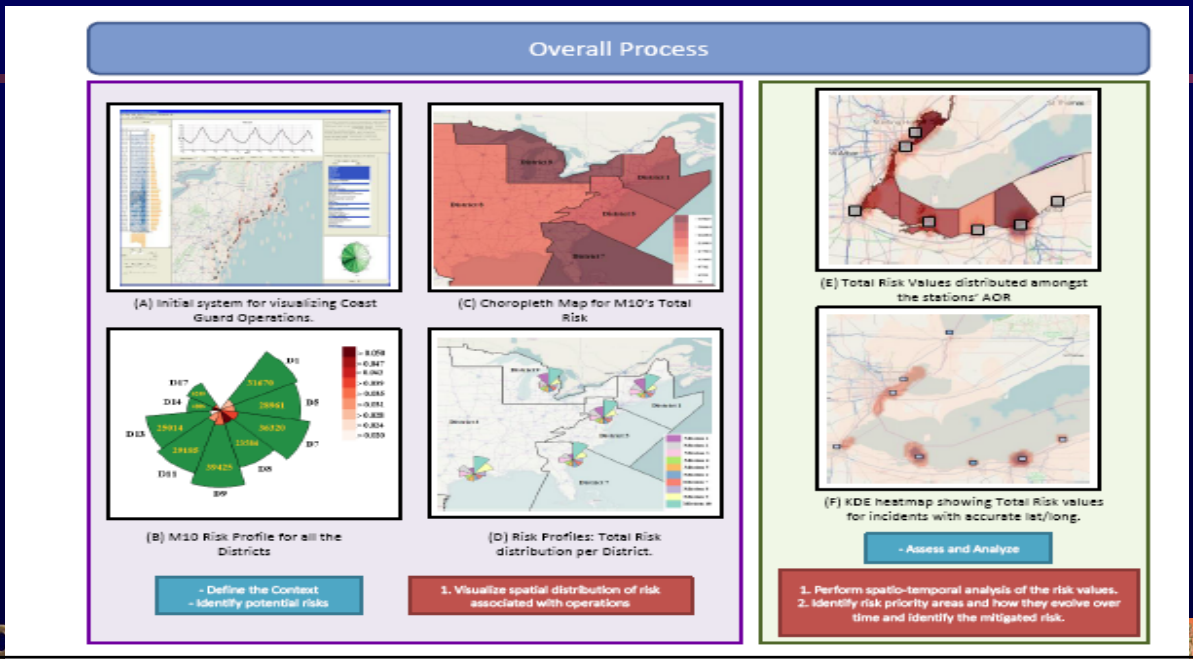
- Evaluate current and historical mission area:
  - Demands
  - Risks (total, mitigated, residual)
  - Resource allocation
  - Return on investment
- Evaluate courses of action
- Evaluate above at both Strategic and Tactical/Operational level

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)

VA For Risk-Based Decision Making Process

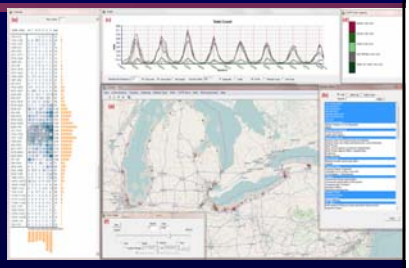


U.S. Coast Guard Search and Rescue VA (cgSARVA)

Partners: USCG LANT 7, USCG D9, USCG D5, USCG HQ 771

IMPACTS:

- Analyzed impact of CG auxiliary stations on search and rescue mission in Great Lakes
- Used for resource allocation for SAR
- Provided new insights to SAR mission
- Hurricanes Sandy and Irene resource allocation decisions based on cgSARVA analysis and visualization
- Informed Commandant's budget testimony to Congress
- Key component of USCG D9 reallocation plan for 2011-12
- Key component of Coastal Operations Allocation Suite of Tools (COAST) – USCG HQ

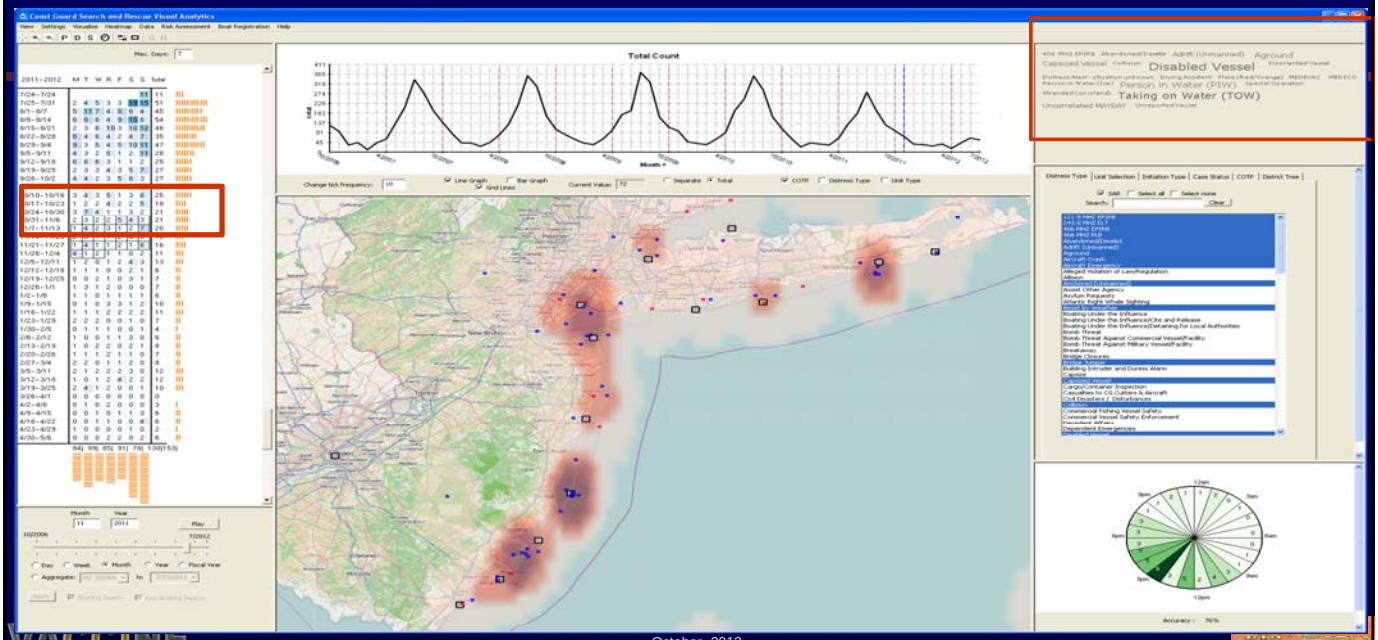


VACCINE

October 2013

Example: Risks and Consequences From Sandy:

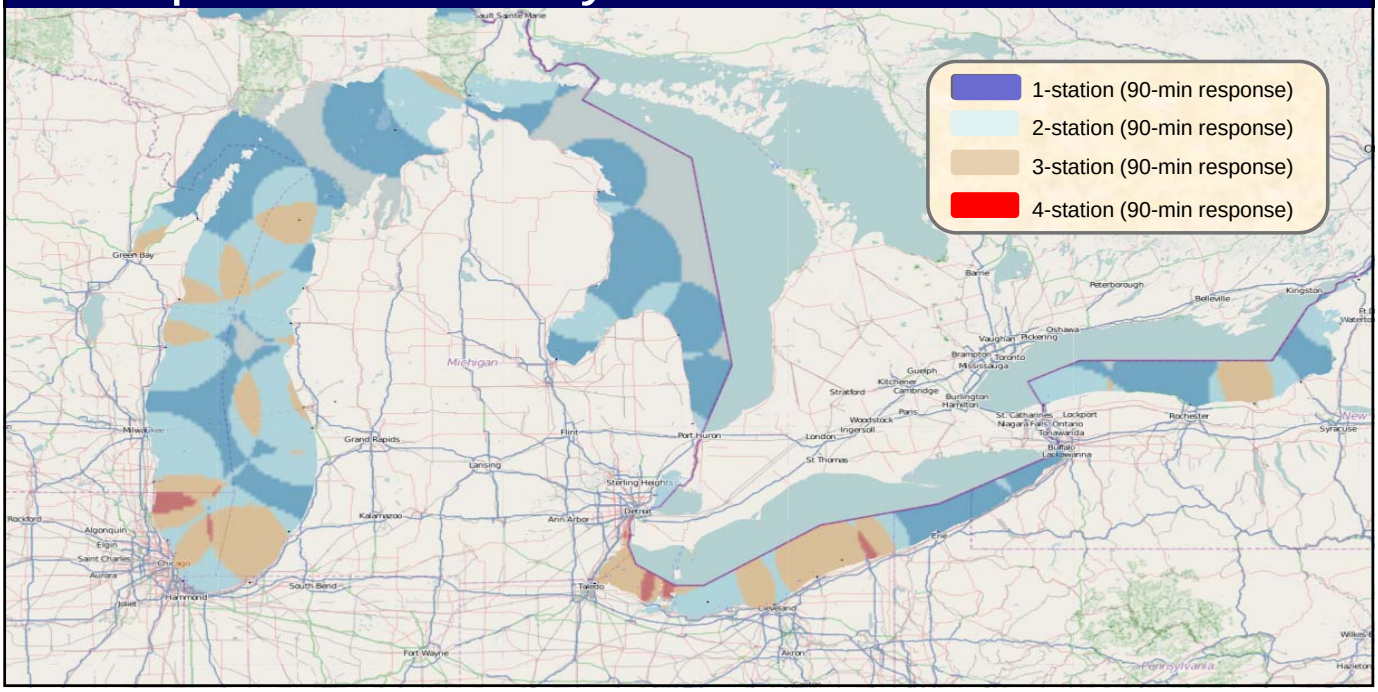
SAR Cases November 2011 NJ/NYC Area



VACCINE

October 2013

Response Efficiency – Possible Asset Allocation



Software Accredited for Decision Making

- April 22, 2013 cgSARVA VV&A'd for US Coast Guard system-wide use

VACCINE

Chasm Update – Crossed And Survived

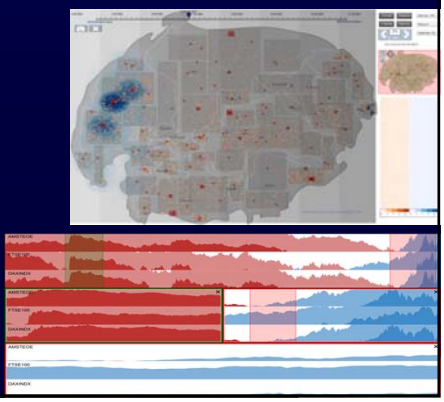
-

Lessons Learned

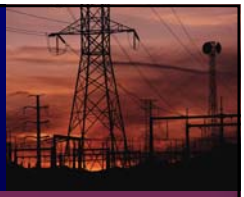
- Extremely worthwhile
- Communication and interaction are key
- Continually ask questions
- Many surprises around each turn (e.g., we need to VV&A the software)
- A growth and learning experience for everyone – a lot of acquired wisdom

Example VACCINE Team Work in Cybersecurity

- Visual Analytics for Security Application (VASA)
- Corporate Insider Threat Detection (Oxford, Leicester, Cardiff)
- Sensor Forensics (Purdue)
- SemanticPrism (Purdue)
- Multiscreen, Multiview, Interactive Cyber Investigation (VaTech, PNNL)
- Log Visualization (Purdue)



Cascading Critical Infrastructure Resiliency Modeling and Analytics (VASA)



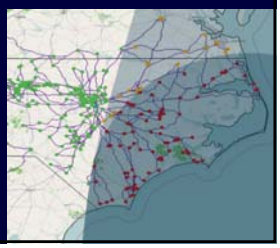
- Purpose: Apply visual analytics to the problem of monitoring and understanding cyber networks and critical infrastructures during detrimental cascading effects, and to the management of the ensuing crisis response.
- Collaborating Institution(s): Purdue, UNC Charlotte, U. Minn. (NCFPD), U. Konstanz, U. Stuttgart, Fraunhofer IGD, Siemens, German utilities
- End-User(s): Power Suppliers (e.g., Duke Energy), Cyber Community (e.g., Cisco), Quick Service Restaurants and suppliers, food supply

VASA: Visual Analytics for Security Applications

Collaborating Institution(s): Purdue, Minnesota, UTexas, UNCC + German universities
End-User(s): Fast-food restaurant chain, emergency management and planning personnel

Impacts and Accomplishments:

- Support decision-making for extreme weather and disaster (natural, man-made) scenarios
  - Combine real and simulation data
  - Allow "what-if" exploration
- **System of systems**: binds together multiple simulations models from collaborators into coherent whole
  - **Minnesota**: food distribution model
  - **Texas**: simulated and historical weather (hurricanes, storms)
  - **UNCC**: critical infrastructure
  - **Purdue**: roads + interaction visual analytics tool
- **Challenge**: combine interactive VA with complex simulation models for effective decision making



Corporate Insider Threat Detection: Cyber Security Inside and Out

(Universities of Oxford, Leicester, and Cardiff)

- **Sponsor:** Centre for the Protection of National Infrastructure
- **Academics:** Sadie Creese (PI), Min Chen, Michael Goldsmith, Michael Levi, David Upton and Monica Whitty
- **Combined Expertise** in cyber security, psychology, criminology, visual analytics, enterprise operations management and executive education
- **Objectives:**
  - Develop a model,
  - Understand psychological indicators
  - Identify the most effective algorithms
  - Understand enterprise culture and common practices
  - Provide a **visual analytical** interface
  - Develop an understanding of both the various organisational roles and awareness raising and educational methods
- **URL:** http://www.cs.ox.ac.uk/projects/CITD/index.html
- **Oxford Cybersecurity Centre:** http://www.cybersecurity.ox.ac.uk/index.html

Sensor Forensics
(Purdue – Delp)

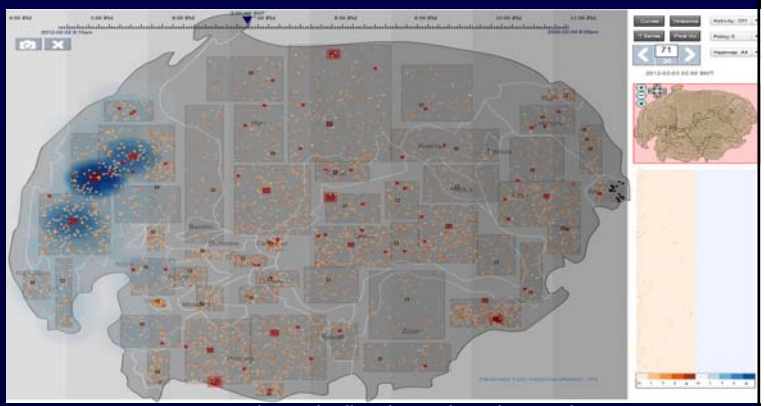
- Forensic characterization
  - Observe device output ➔ which device produced it?
  - Exploit how the device "makes" its output
- Device authentication
  - Performed using forensic characterization
  - Identify device type, make, model, configuration
  - Can the sensor be trusted?
- Detection of data forgery or alteration
- Fingerprint and trace

SemanticPrism: A Multi-aspect View of Large High-dimensional Data (Purdue University)

- **VAST 2012 Mini Challenge 1 Award: Outstanding Integrated Analysis and Visualization**

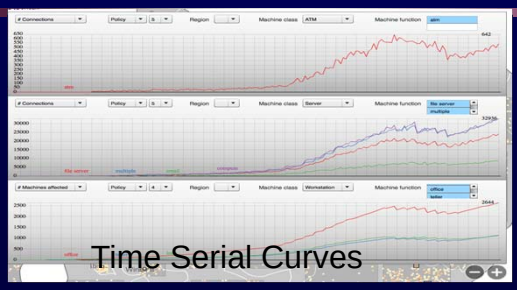
- **Geo-Temporal**
- **Time-serial**
- **Pixel-based**

- **Semantic Zoom**

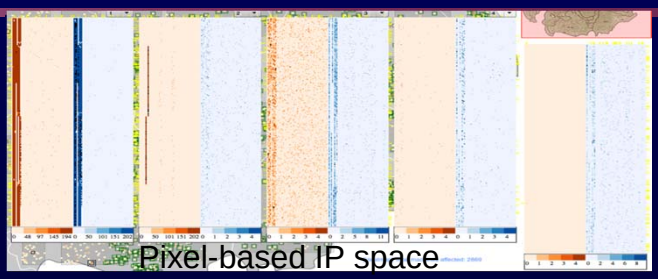


Victor Yingjie Chen, Ahmad M Razip, Sungahn Ko, Cheryl Zhenyu Qian, David S.Ebert

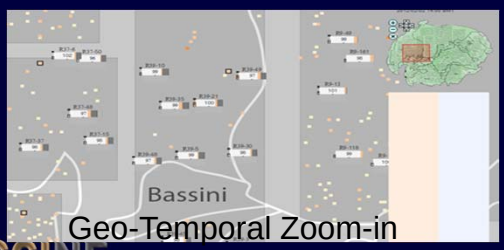
SemanticPrism



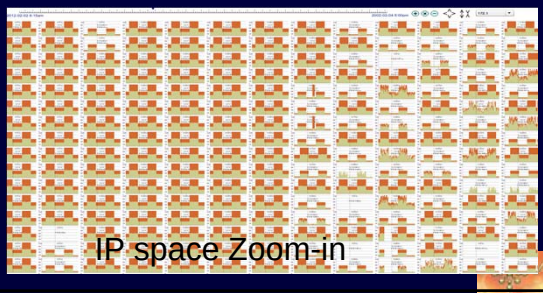
Time Serial Curves



Pixel-based IP space



Geo-Temporal Zoom-in



IP space Zoom-in

VACCINE

October 2013

VA for Cybersecurity Analysts

(VaTech – North, Endert)

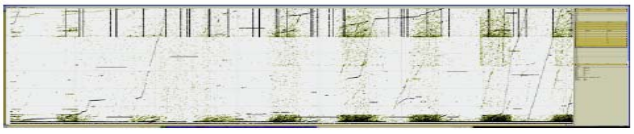
Large High-Res Workspaces for Analysts

History and Traceability

Multiple, Simultaneous Investigation cases

Large, High-Resolution Visualizations

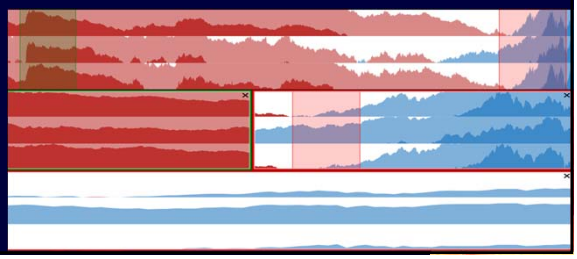
De-Aggregate Vital Information



Server Cluster Log File Visualization

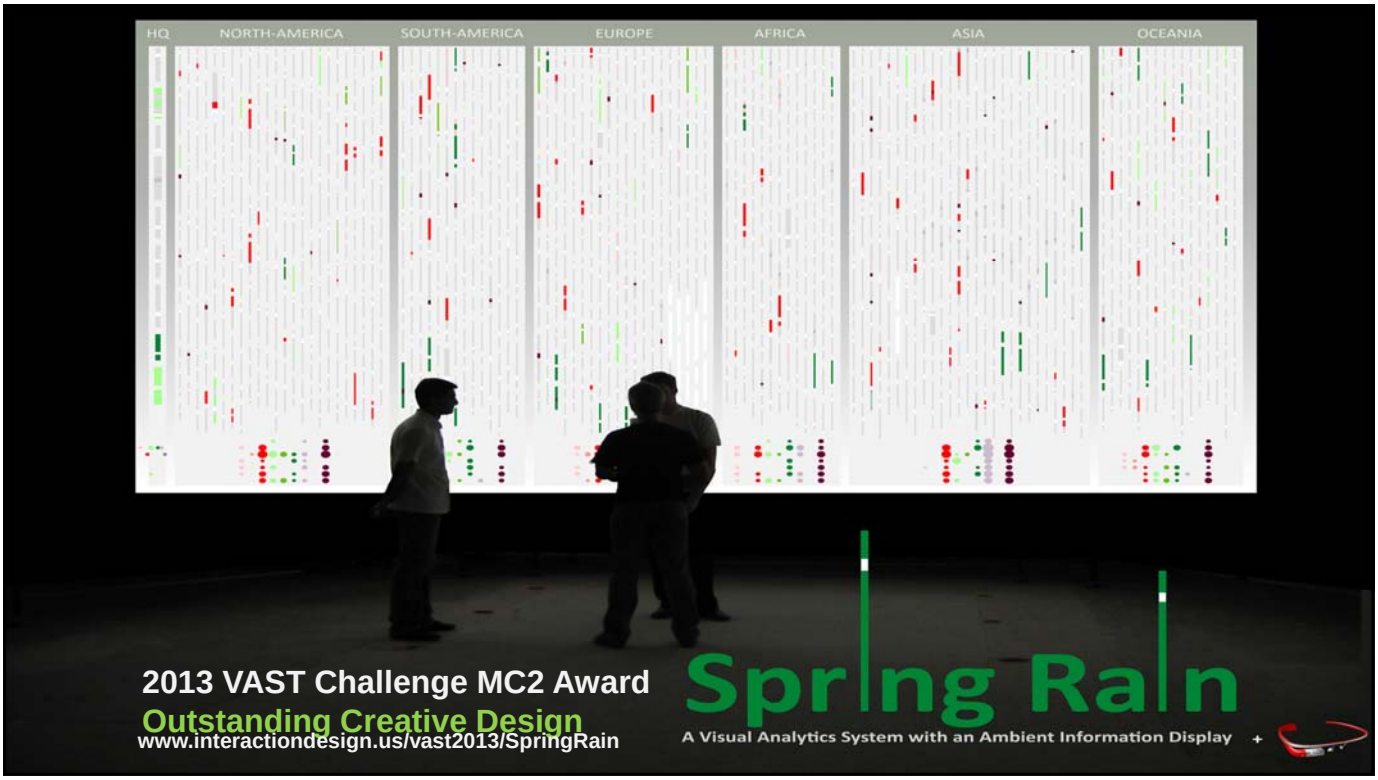
(Elmqvist, Purdue)

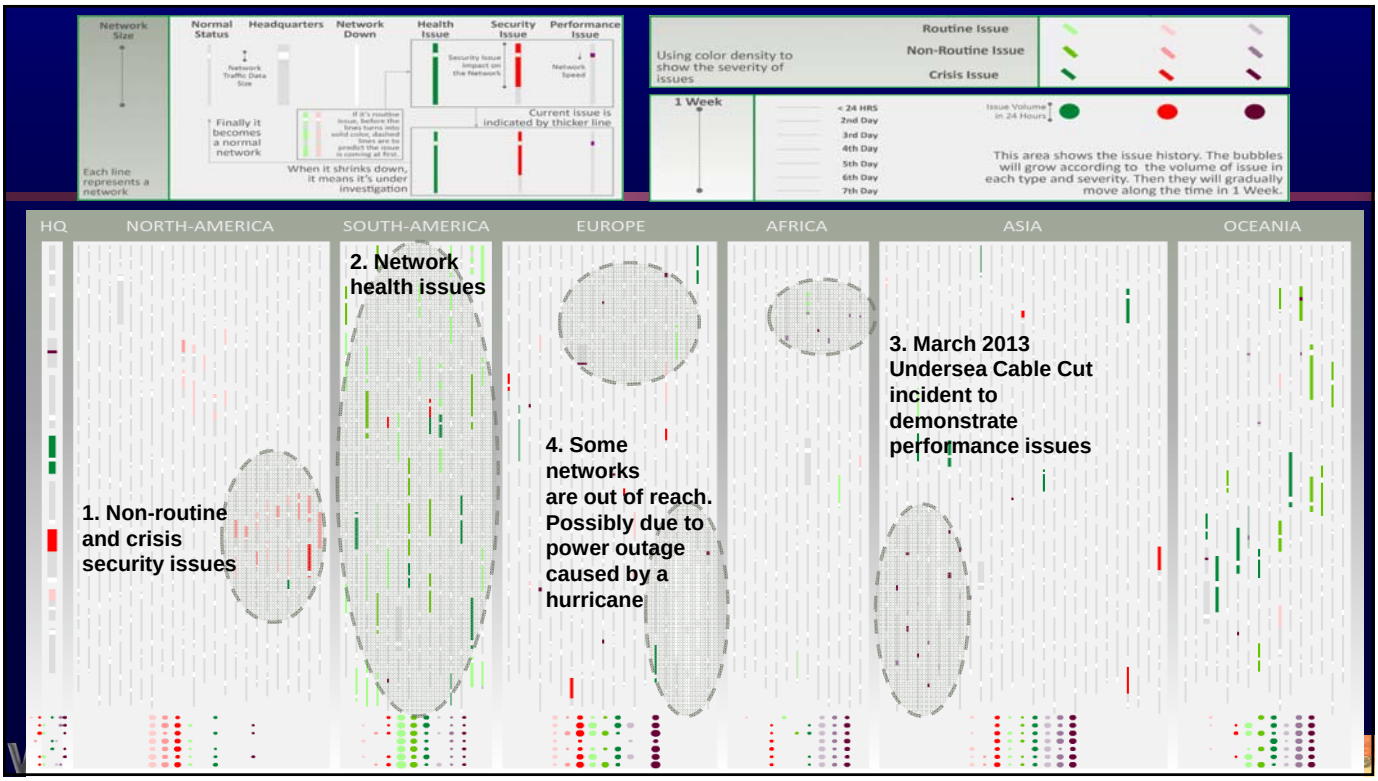
- Log file visualization for Purdue's ECN group
  - 200+ servers, 30 TB of storage, 6 million hits per month
  - User-centered, interview-based design
- Applied **stack zooming** to quantitative log data
  - CPU load, network hits, storage usage,
  - Users navigate in data
  - Very long time periods
- Limited deployment in Fall 2010
  - Very positive, powerful



VACCINE

October 2013





Application-Driven Visualiza Cybersecurity

- Should we be driving the research different users' goals are?
- Interesting survey article
  - Taxonomy of use-case classes:
    - Host-server monitoring
    - Internal/external monitoring
    - Port activity
    - Attack patterns
    - Routing behavior

Shiravi, et al., " A survey of visualization systems for networ security, *IEEE TVCG* 2012.

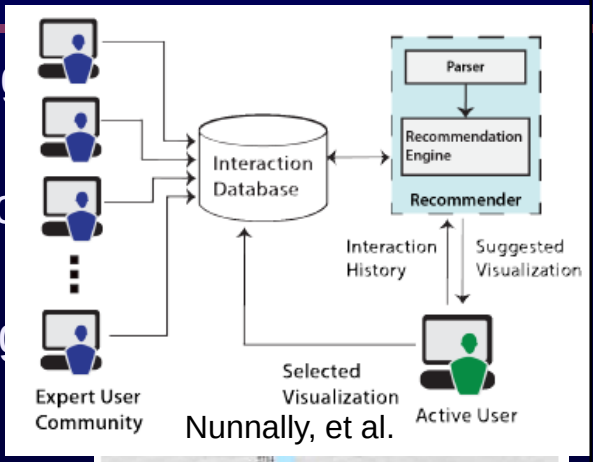
Visualization System	Visualization Technique(s)	Data Source(s)	Number of Citations
Host / Server Monitoring			
Erbacher et al. [4][5]	Glyph	Server Logs	106 | 7
Tudumi [6]	3D Node Link	Server Logs	38
NVisionIP [7,8]	Scatter Plot	NetFlows	145 | 20
Portall [9]	Node Link	Packet Traces	21
HoNe [10]	Node Link	Packet Traces	8
Perlman et al. [11]	Node Link | Glyph	Packet Traces	5
Radial Traffic [12]	Radial Panel	Packet Traces	23
Mansmann et al. [13]	Node Link	Packet Traces	2

Do citation counts show real-world value?

Cube of Doom [19]	3D Scatter Plot	Packet Traces	99
PortVis [20]	Scatter Plot	NetFlows	112
NetBytes Viewer [21]	3D Scatter Plot	NetFlows	7
Existence Plots [22]	Scatter Plot	Packet Traces	3
Attack Patterns			
Giardin [29]	Color Map	Packet Traces	60
NIVA [30]	Node Link | Glyph	Intrusion Alerts	51
Snort View [31]	Scatter Plot | Glyph	Intrusion Alerts	67
IDGraphs [32]	Scatter Plot	NetFlows	29
IP Matrix [33]	Scatter Plot | Color	Intrusion Alerts	21
Visual Firewall [34]	Scatter Plot	Packet Traces	24
IDS Rainstorm [35]	Scatter Plot	Intrusion Alerts	60
Vizalert [36][37][38]	Radial Panel	Intrusion Alerts	38 | 35 | 29
Rumint [39][40]	Parallel Coordinates	Packet Traces	15 | 35
Ren et al. [41]	Flying Term	DNS Traces	10
Xiao et al. [42]	Scatter Plot	Packet Traces	23
Svision [43]	3D Scatter Plot	Packet Traces	9
Mansmann et al. [44]	Treemap	Packet Traces	20
SpiralView [45]	Radial Panel	Intrusion Alerts	5
NFlowVis [46]	Treemap	NetFlows	17
Avisa [49]	Radial Panel	Intrusion Alerts	2
Routing Behavior			
BGPlay [50]	Node Link	BGP Traces	22
Wong et al. [51]	Node Link	BGP Traces	9
LinkRank [52]	Node Link	BGP Traces	16
Teoh et al. [53][54][55]	Histogram | Node Link	BGP Traces	54 | 28 | 35
BGP Eye [56]	Color Map	BGP Traces	8

VizSec Papers - My Analysis

- Expert users engaged: 3 of 9 (e.g. network analysts)
- Evaluation – performance most c informal user studies
- Data – 4 with actual data, 1 using
- Users involved from the start – 1 paper
- Training of Novices based on experts



Nunnally, et al.

Hao, et al.

Directions Forward, Keys to Success & Challenges

Cybersecurity Education

- How do we train practitioners in the field?
  - Varied backgrounds, varied tasks, communication to public
  - Short time to learn
  - Visualization is key
- Approaches:
  - How people learn framework
  - Personalizing learning - Community of practice train
- How do we educate the public?
  - Again, visualization is key

Some Challenges for Cybersecurity

1. Understanding the task and workflow, access to expert users, actual problems, environments
2. Creating decision making environments for analysts with realtime data and decision making constraints at real-world scale (computer-human visual cognition environments)
3. Solving specific scale issues (scalability) and cross-scale issues (machine, intranet, internet)
4. Managing uncertainty and time
5. Enabling risk-based decision making environments

Keys for Success

- User and problem driven
- Balance human cognition and automated analysis and modeling
- Interactivity and easy interaction
  - Intuitive and scalable solutions vital
- Understandability
- Intuitive visual cognition
- Not overloaded with features

For Further Information

www.VisualAnalytics-CCI.org

vaccine@purdue.edu
ebertd@purdue.edu