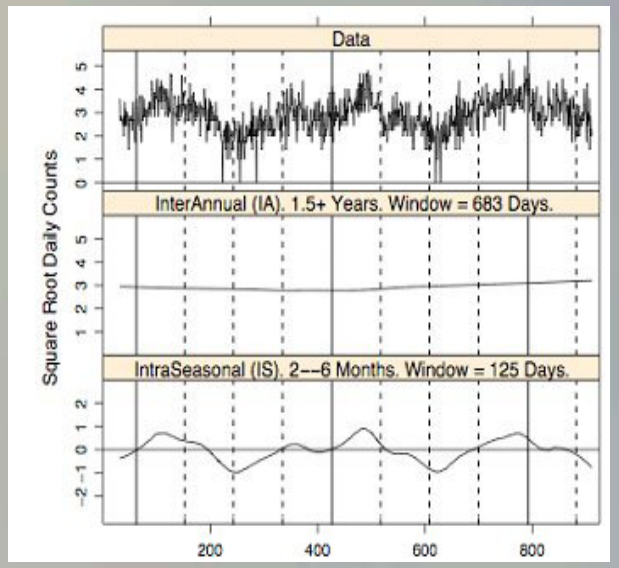
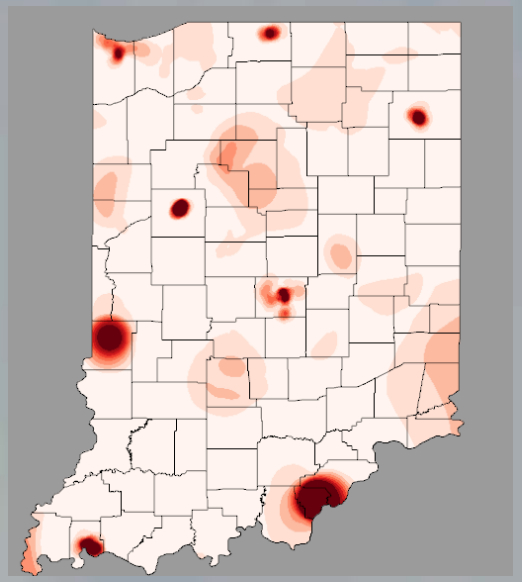
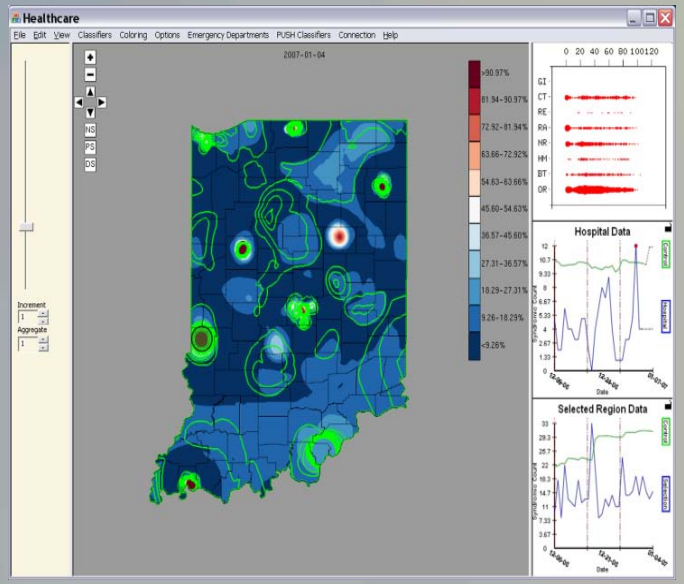
PURVAC

Purdue University Regional Visualization and Analytics Center

Geo-Spatiotemporal Data Applications in Visual Analytics







Ross Maciejewski

www.purvac.org

RVAC Regional Visualization and Analytics Centers

Discovery Park

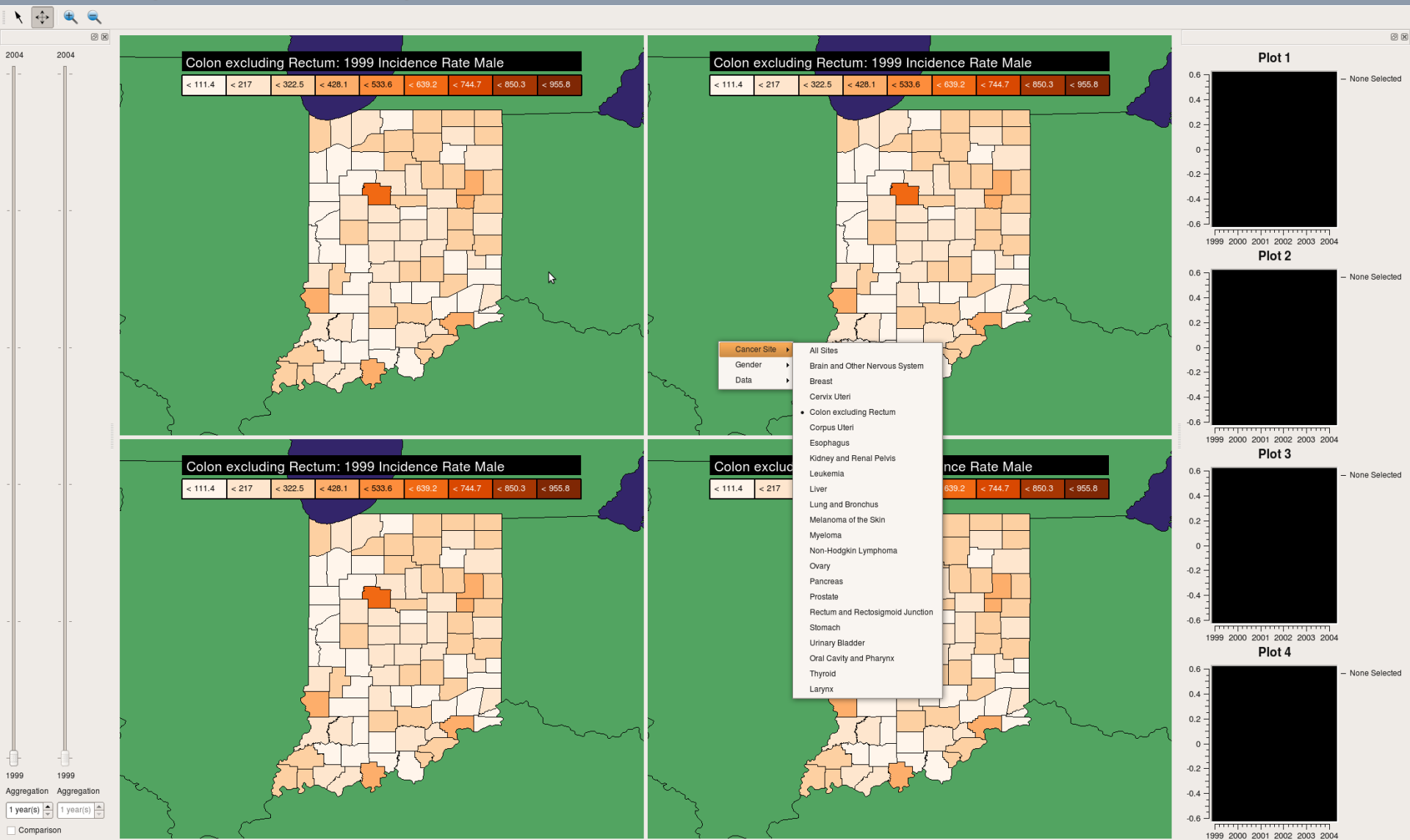
Table 4. Average Indiana Cancer Incidence Rates by County, 2001 - 2003

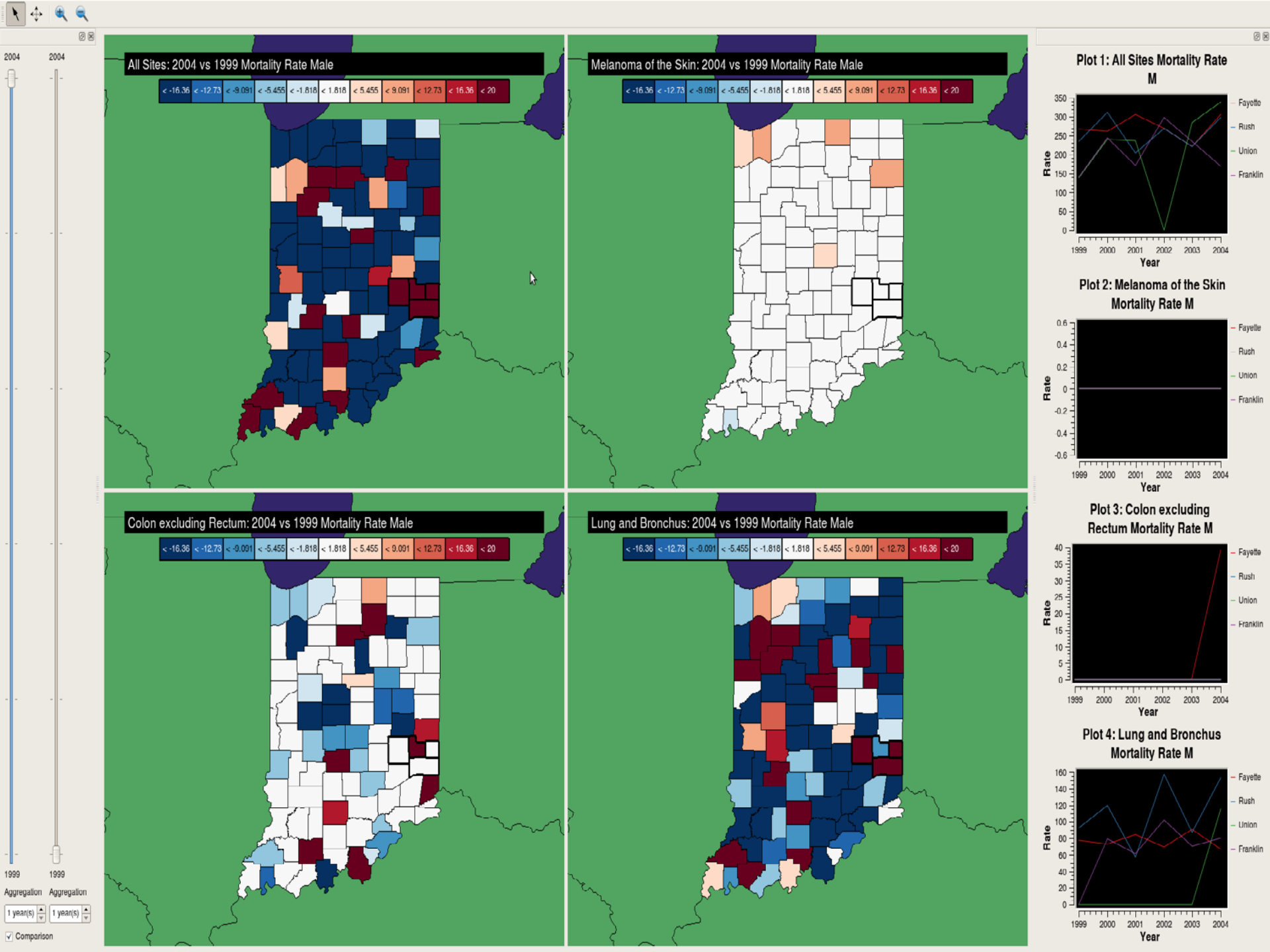
County	All Cancers		Female Breast		Colon & Rectum		Lung		Prostate	
	Count	Rate	Count	Rate	Count	Rate	Count	Rate	Count	Rate
Adams	359	353.2	58	102.0	37	35.2	49	49.2	33	76.2
Allen	4,232	446.4	688	130.5	561	59.5	648	69.6	418	104.0
Bartholomew	1,040	466.7	166	134.7	101	45.6	163	73.1	103	102.9
Benton	192	618.6	20	123.0	17	#	38	123.4	30	216.5
Blackford	315	636.2	41	157.4	41	80.2	66	131.6	34	158.0
Boone	596	424.4	102	129.8	54	38.8	78	58.4	64	106.6
Brown	143	287.9	20	73.3	9	#	26	52.5	19	#
Carroll	259	385.9	34	93.2	36	53.0	36	54.2	37	121.3
Cass	682	510.5	86	120.9	95	69.3	119	88.6	72	120.2
Clark	1,607	532.6	239	141.6	171	57.3	314	104.4	259	202.0
Clay	501	556.8	66	136.8	64	69.9	87	95.4	56	142.4
Clinton	489	442.1	66	111.4	56	49.8	73	66.3	60	128.2
Crawford	172	489.6	15	#	23	68.1	39	110.4	18	#
Daviess	399	414.0	49	93.6	58	59.8	61	62.3	41	100.4
Dearborn	566	413.2	62	81.8	79	59.3	118	88.2	68	110.7
Decatur	346	439.8	35	83.5	36	45.0	59	74.3	32	90.9
DeKalb	514	440.0	85	134.9	67	57.9	73	62.9	45	88.9
Delaware	1,791	484.4	223	113.4	213	56.2	309	82.4	224	144.3
Dubois	505	406.1	84	124.7	72	57.8	52	41.7	82	148.9
Elkhart	2,287	448.9	307	109.9	274	54.1	381	76.4	275	127.5
Fayette	444	495.5	62	129.1	46	50.0	81	88.5	63	153.5
Floyd	1,102	505.3	149	122.1	105	48.5	208	96.1	177	194.9
Fountain	332	523.8	40	117.9	46	70.7	60	92.7	47	165.6
Franklin	248	359.7	40	110.2	31	45.2	47	68.0	24	74.8
Fulton	374	519.9	56	148.4	42	57.3	59	81.4	58	179.8
Gibson	502	443.8	67	110.4	86	74.9	92	80.2	53	105.8
Grant	1,279	507.8	146	110.4	128	50.2	225	87.3	183	163.8
Greene	569	489.9	63	105.2	85	73.2	112	94.4	72	138.6
Hamilton	2,010	429.8	348	126.0	208	46.7	278	67.3	243	119.6
Hancock	846	485.1	117	120.7	97	57.5	148	86.2	90	112.7
Harrison	458	439.3	54	98.0	60	58.7	84	81.4	68	139.0
Hendricks	1,474	491.1	228	137.6	156	52.8	229	80.5	187	140.3
Henry	810	469.0	105	116.6	115	64.7	146	81.8	93	120.6

Discovery Park

PURVAC
Purdue University

Visual Analytics for Cancer Care Engineering




2004
2004
1999
1999
Aggregation
Aggregation
1 year(s)
1 year(s)
Comparison
All Sites: 2004 vs 1999 Mortality Rate Male
Melanoma of the Skin: 2004 vs 1999 Mortality Rate Male
Colon excluding Rectum: 2004 vs 1999 Mortality Rate Male
Lung and Bronchus: 2004 vs 1999 Mortality Rate Male
< -16.36
< -12.73
< -9.091
< -5.455
< -1.818
< 1.818
< 5.455
< 9.091
< 12.73
< 16.36
< 20
Plot 1: All Sites Mortality Rate M
Plot 2: Melanoma of the Skin Mortality Rate M
Plot 3: Colon excluding Rectum Mortality Rate M
Plot 4: Lung and Bronchus Mortality Rate M
Rate
Year
Fayette
Rush
Union
Franklin

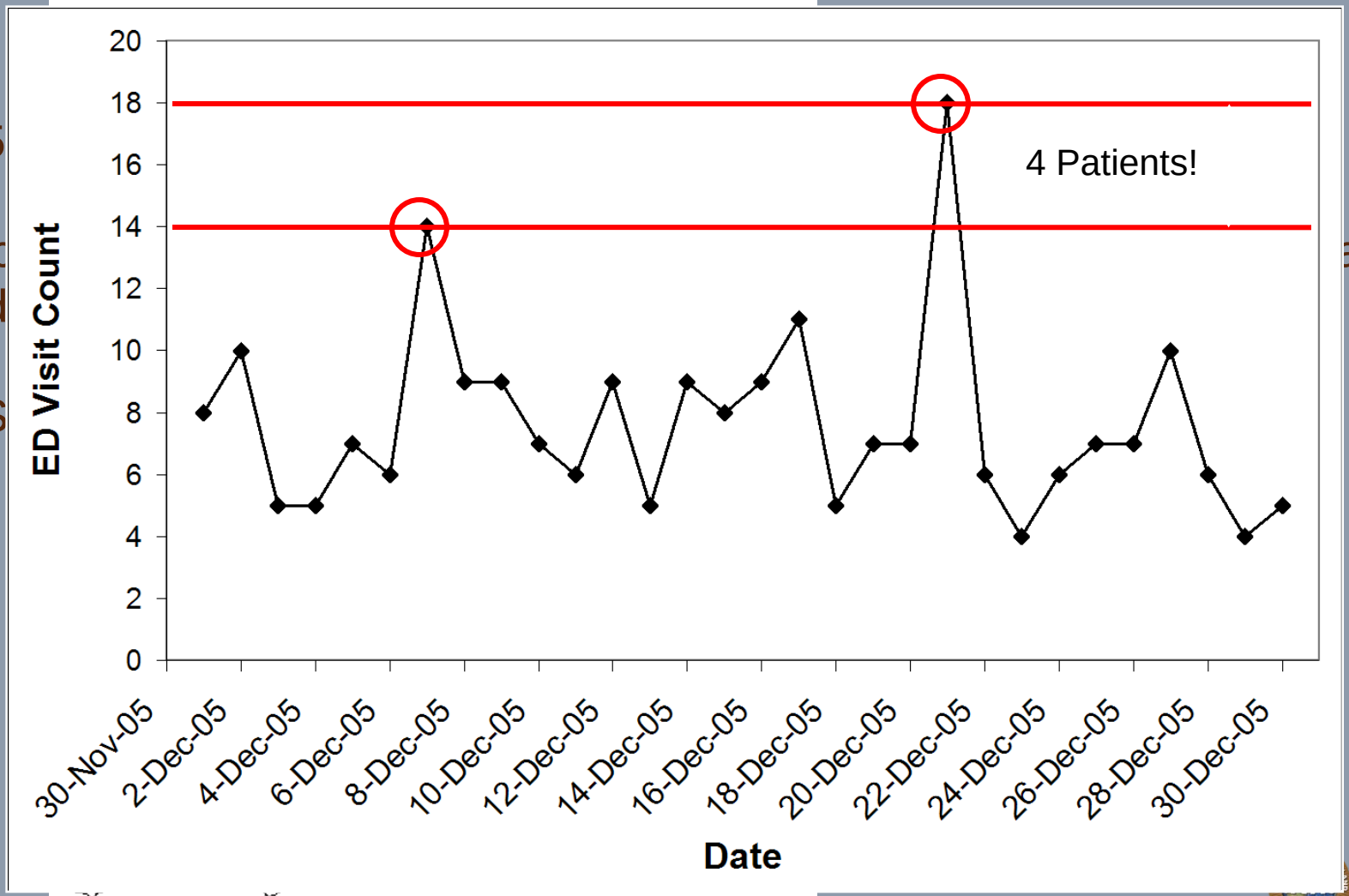
Syndromic Surveillance

- 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*

Indiana's Public Health Emergency Surveillance System (PHESS)

- 77
- 6,5
- Rec and synd
- 8 S



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

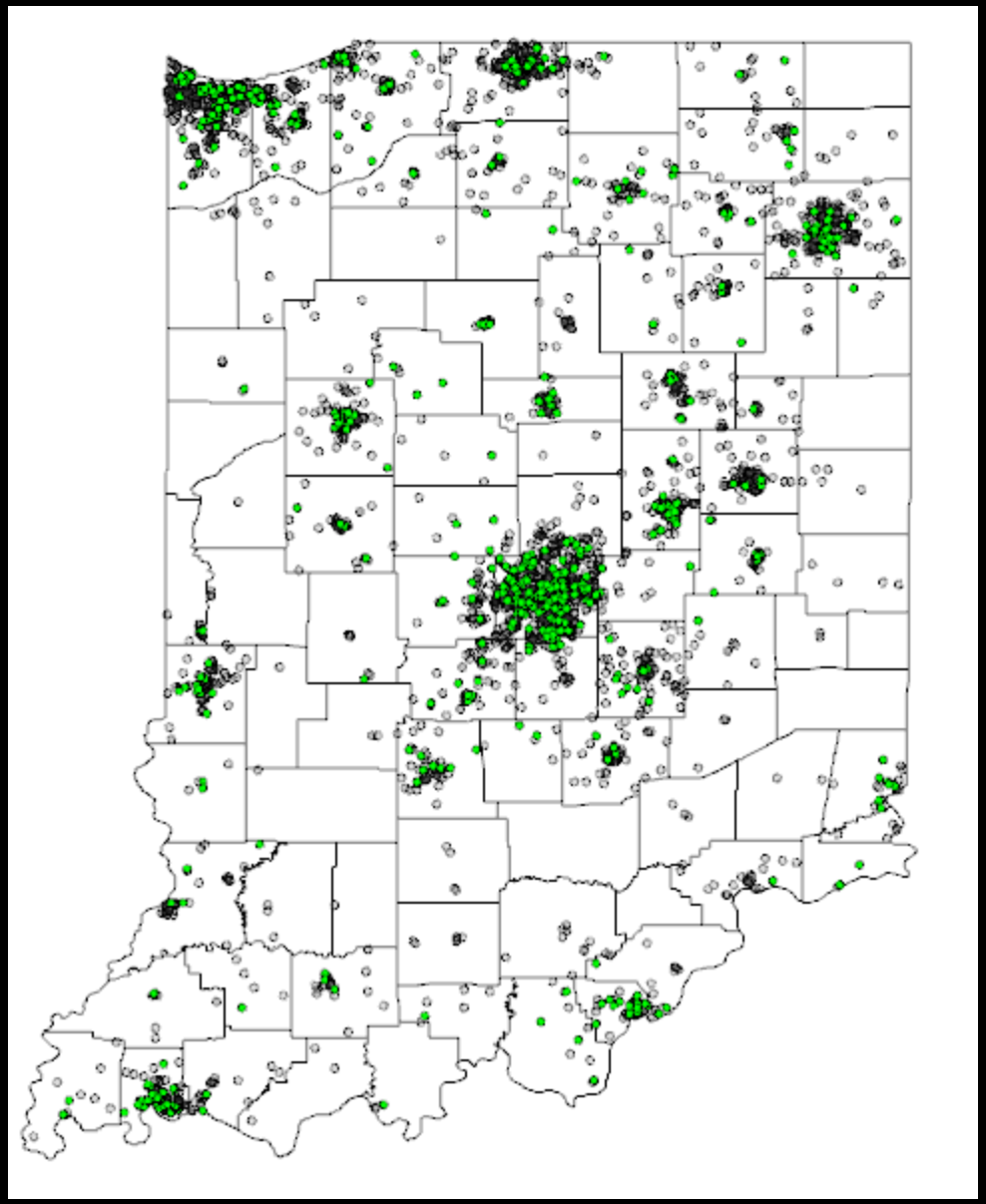
Hotspot – Data Aggregation and Privacy Preservation

Hotspot

- *An area in which an unexpectedly large amount of syndromes are occurring in space and time*

How to display geospatial data?

- *Individual locations give away private information*
- *Aggregation by geographic boundary*
- *Color as a function of percent syndromic?*



Aggregation by Geographic Boundary

Hotspot

- What's expected? What's the population at risk?
  - Total population of county?

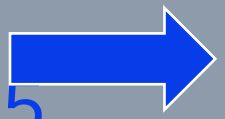
An area in which an unexpectedly large amount of syndromes are occurring in space and time

Aggregation by Geographic Boundary

- Tippecanoe
  - Population – 148,955
  - Number GI visits – 21

.00014%

- Marion
  - Population – 860,454
  - Number GI visits – 135



.00015%

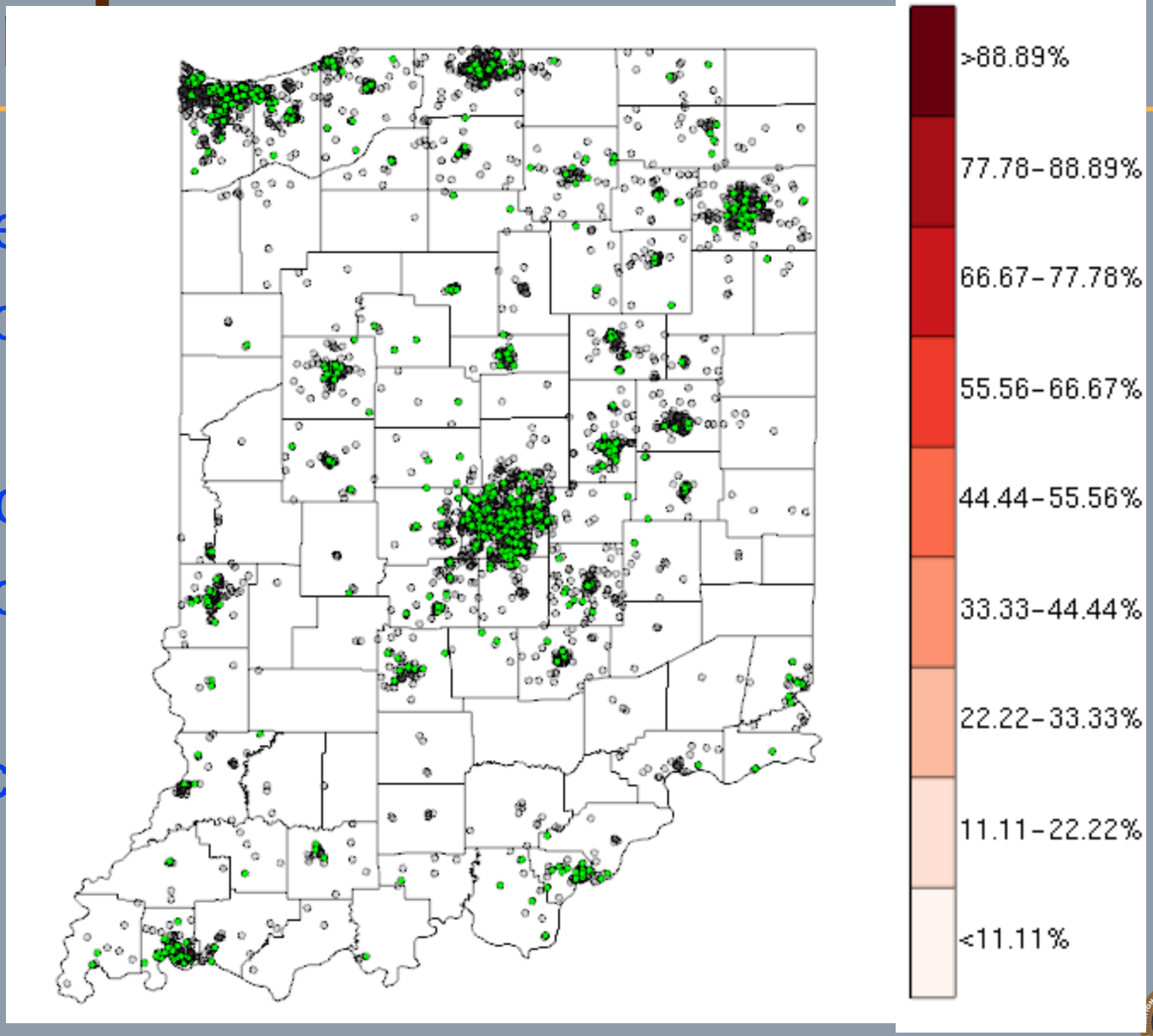
- Which County should be investigated?

Aggregation by Geographic Boundary

- What's expected? What's the population at risk?
  - Visits to the ED?

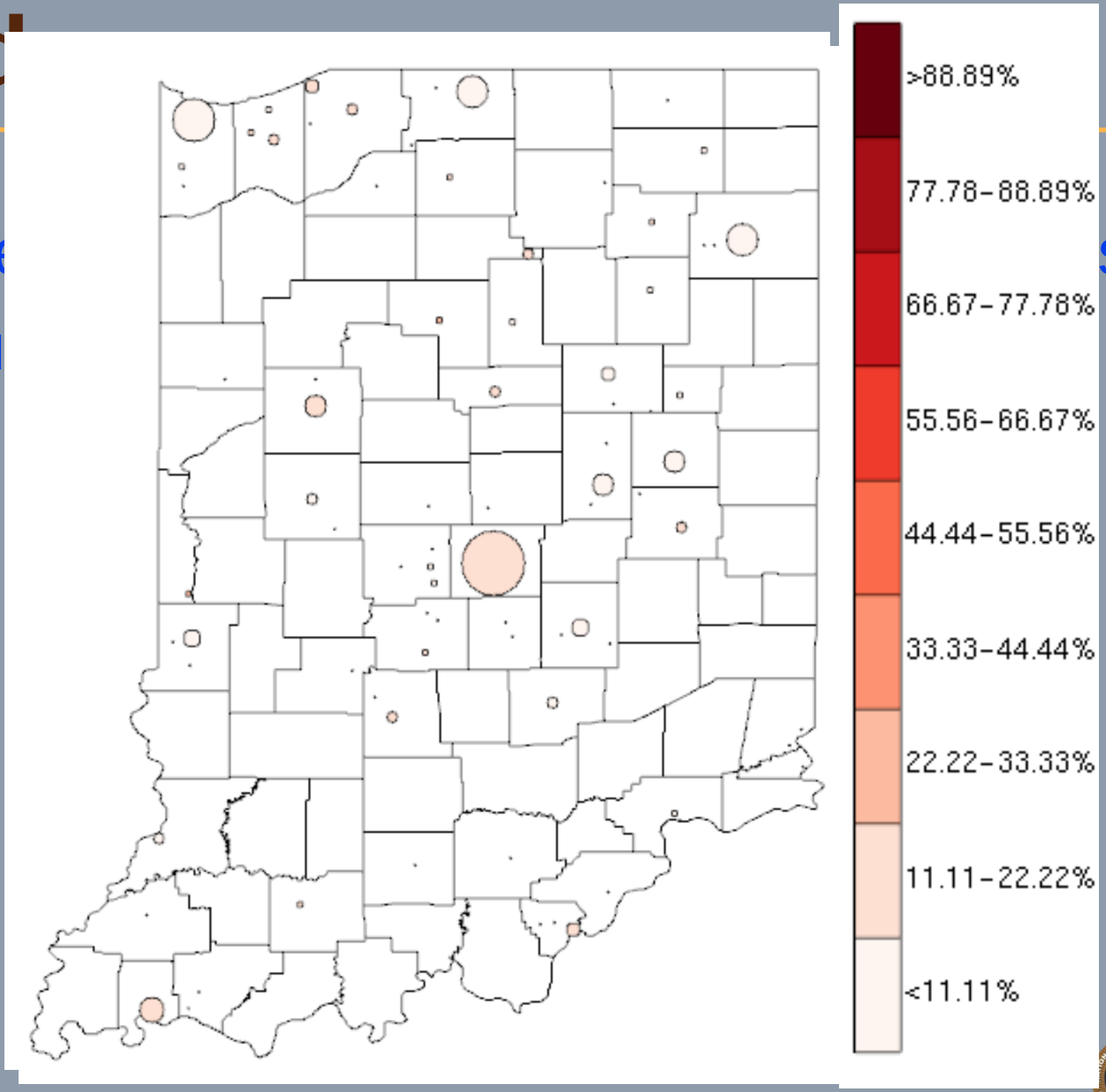
Aggregation by Geographic Bou

- Tippe
  - To
  - G
- Mario
  - To
  - G
- Whic

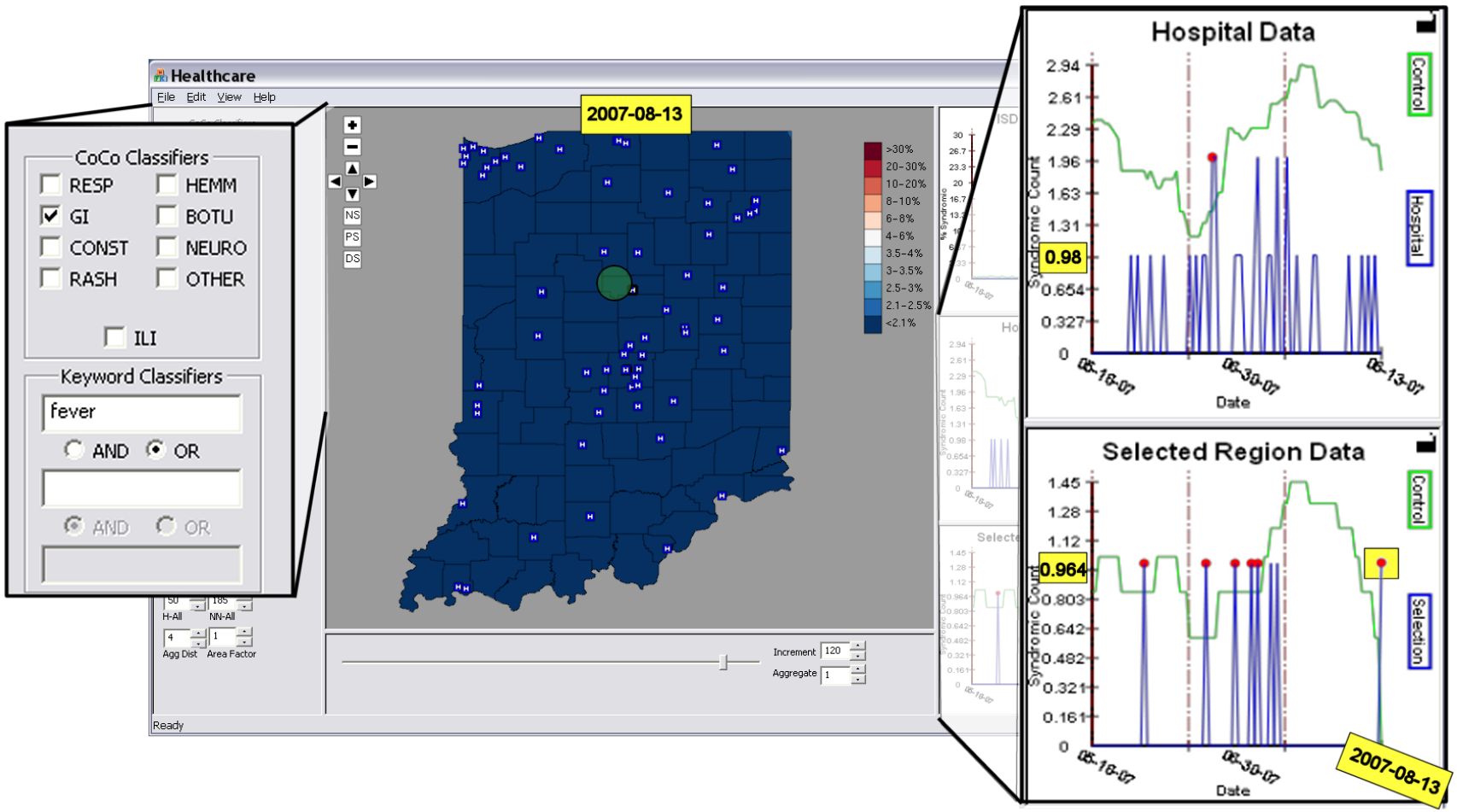


Aggregation by Geographic Bound

- What's e sk?
  - Grou



Integrated Interactive Environment: Hypothesis Testing



Healthcare

File Edit View Help

CoCo Classifiers
- [x] RESP
- [ ] HEMM
- [ ] GI
- [ ] BOTU
- [ ] CONST
- [ ] NEURO
- [ ] RASH
- [ ] OTHER
- [ ] ILI

Keyword Classifiers
( ) AND (•) OR
(•) AND ( ) OR

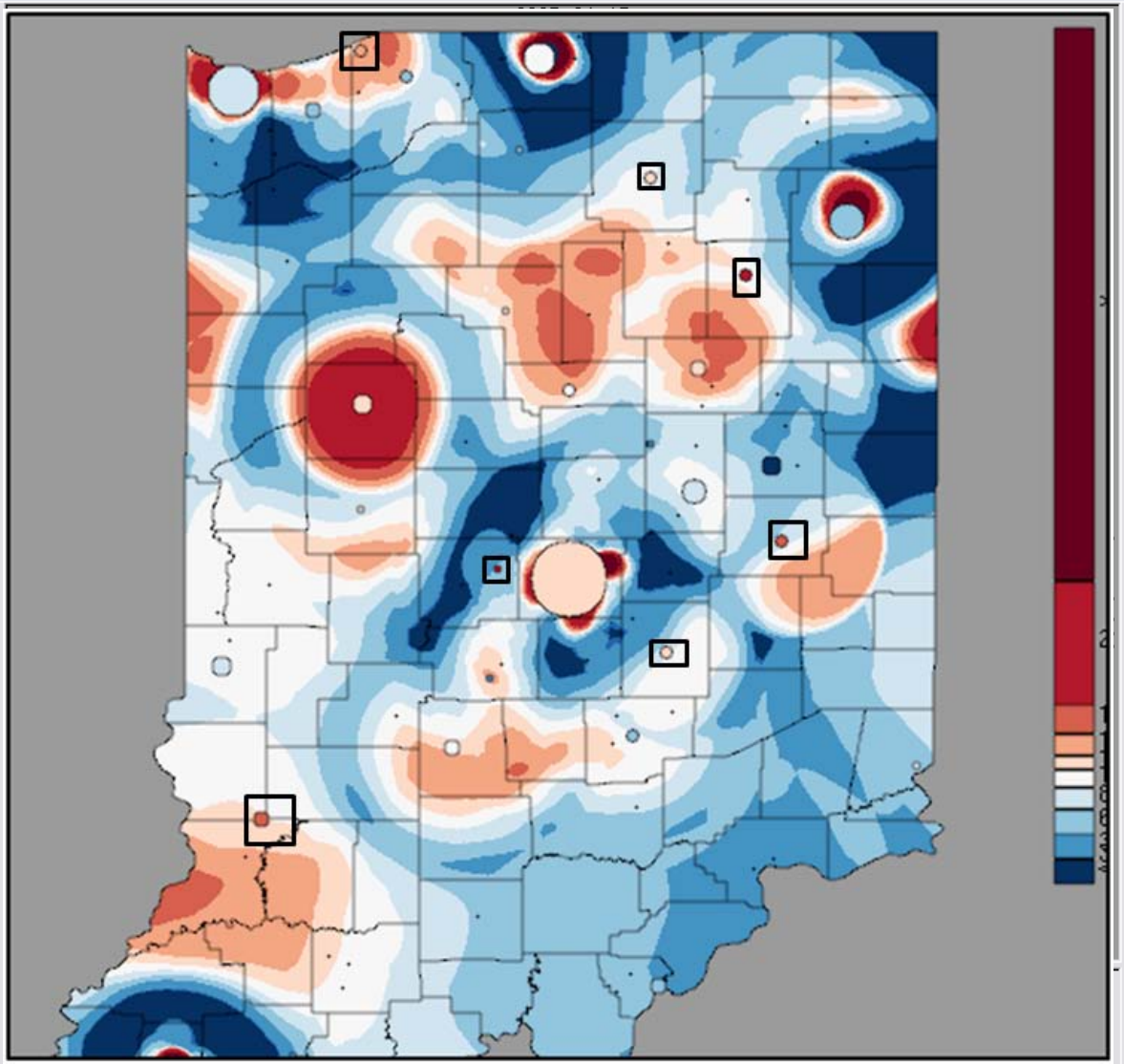
Demographics
(•) All ( ) Male ( ) Female

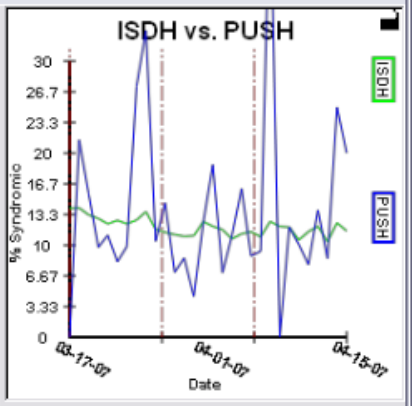
Age
[ ] 0 To 0
[ ] 0 To 0
[ ] 0 To 0

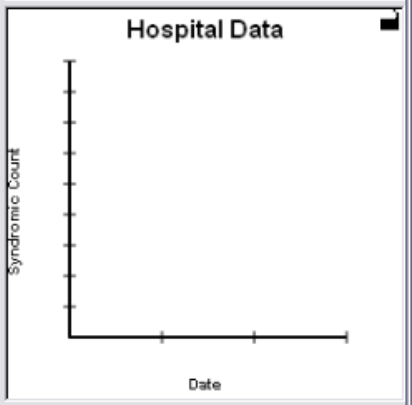
- [ ] All patients
- [ ] Affected patients
- [ ] Density by county
- [ ] Show Hospitals
- [x] Show KDE
- [x] Show Data Aggregation

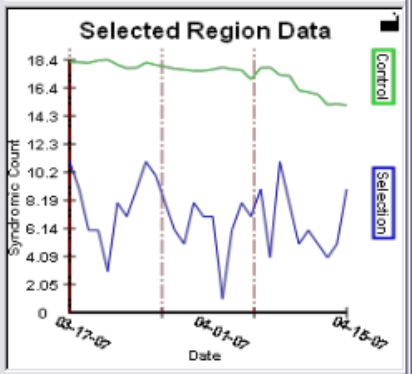
5 H-Ill
18 NN-Ill
50 H-All
185 NN-All
4 Agg Dist
1 Area Factor

( ) Lin ( ) Ramp ( ) Exp
(•) Log ( ) Gaus ( ) Int



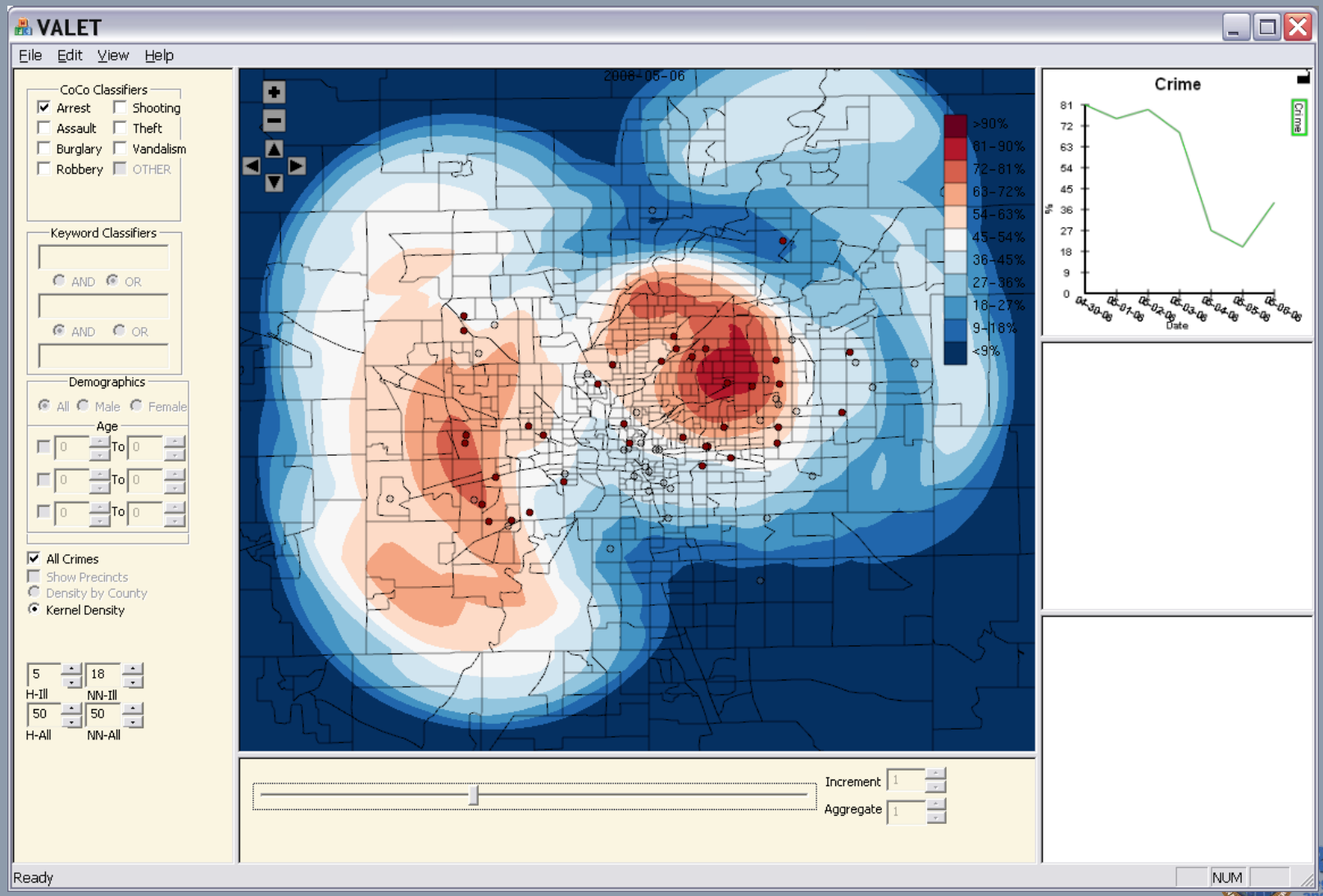


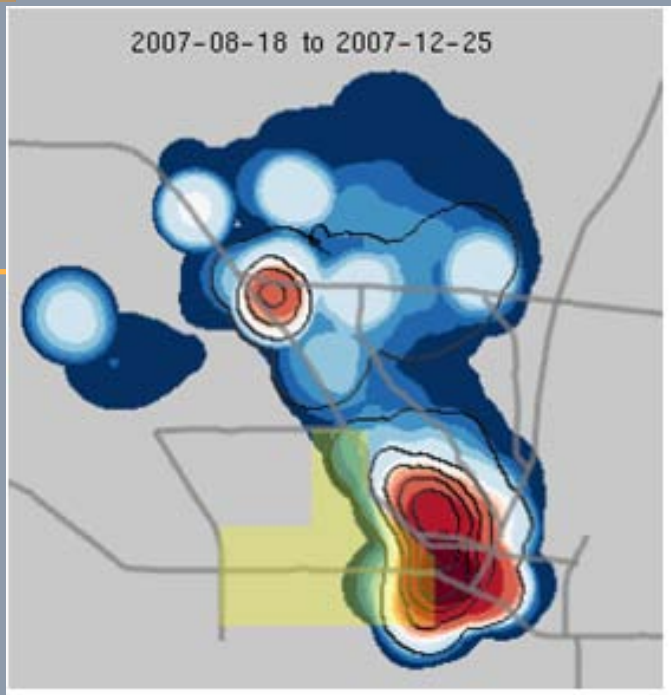


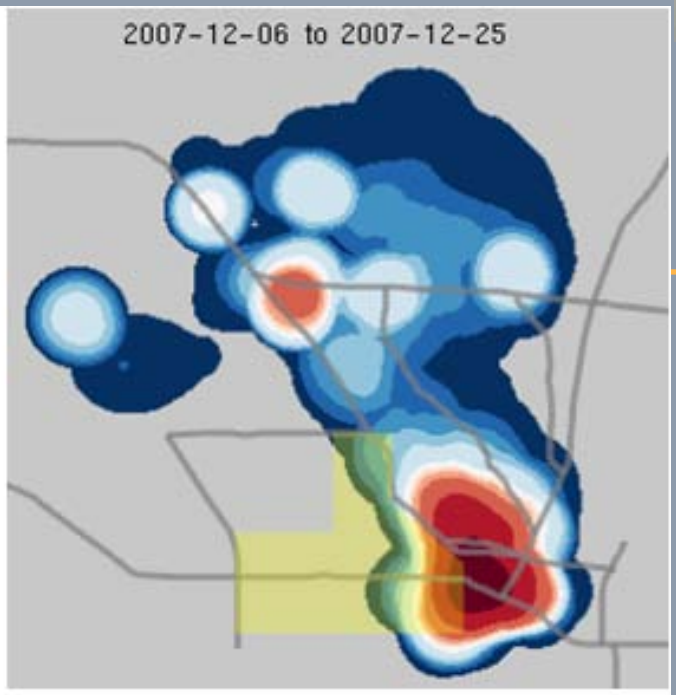


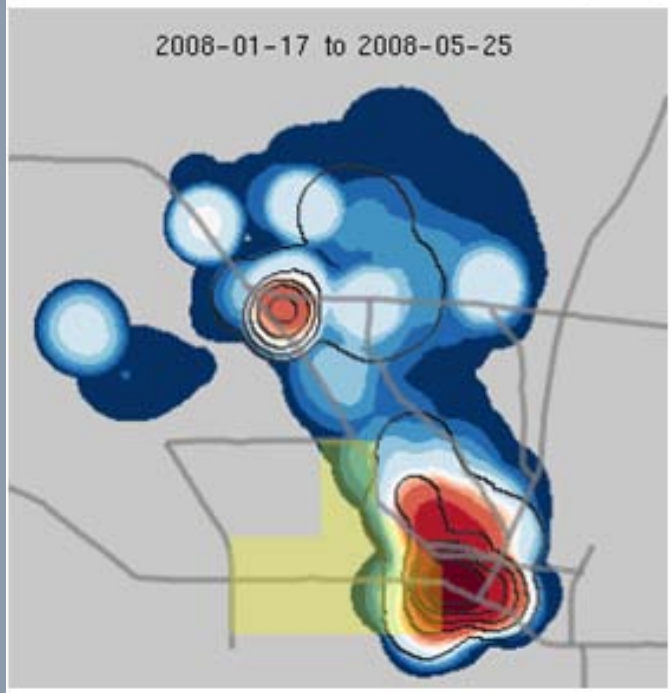
Ready NUM

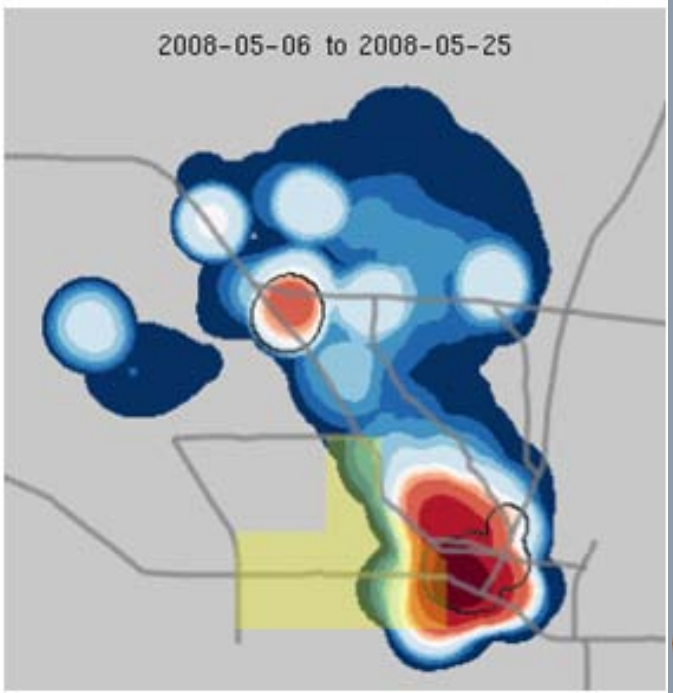
VALET: Visual Analytics Law Enforcement Technology




2007-08-18 to 2007-12-25


2007-12-06 to 2007-12-25


2008-01-17 to 2008-05-25

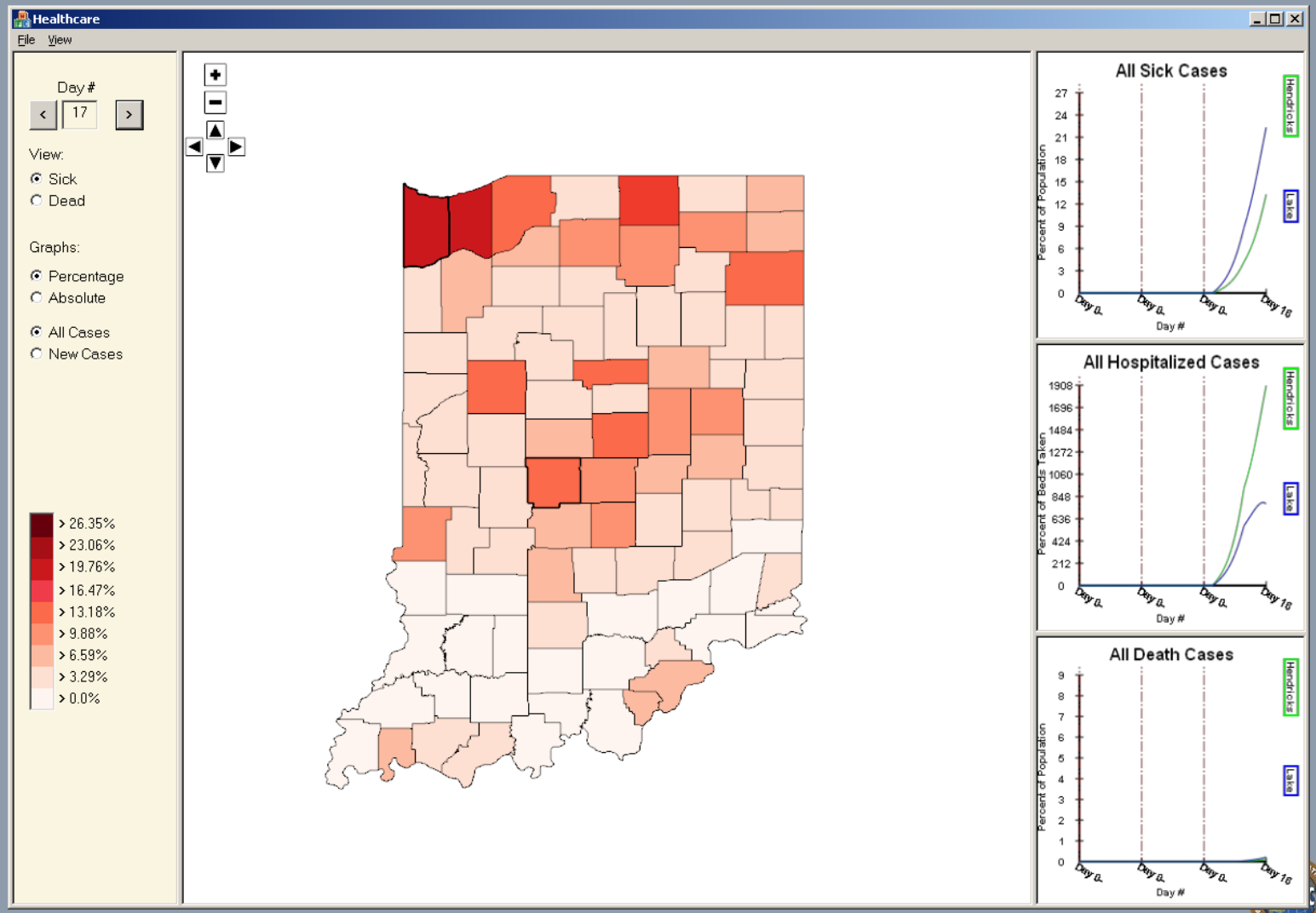

2008-05-06 to 2008-05-25

Pandemic Influenza Preparedness Exercise Visual Analytics

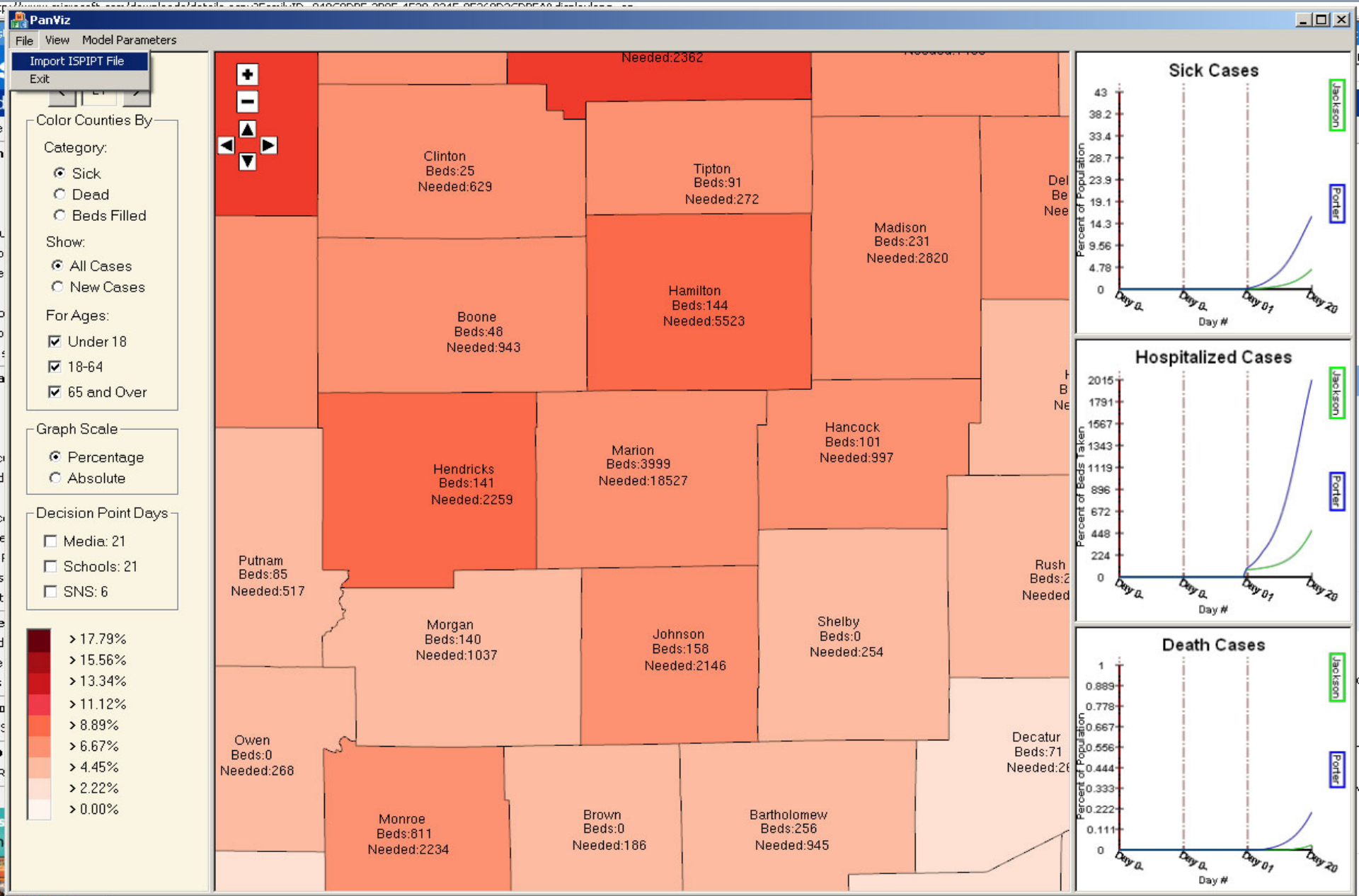
Requirements for system

- Allow decision makers to understand the effects of three types of actions:
  - *Use of Strategic National Stockpile*
  - *Public information announcements*
  - *Closing schools*
- Increase understanding of impact, decisions, capacity needs
- Allow changing of pandemic influenza model parameters

PanViz


Healthcare
File View
Day #
17
View:
Sick
Dead
Graphs:
Percentage
Absolute
All Cases
New Cases
> 26.35%
> 23.06%
> 19.76%
> 16.47%
> 13.18%
> 9.88%
> 6.59%
> 3.29%
> 0.0%
All Sick Cases
Percent of Population
Day #
All Hospitalized Cases
Percent of Beds Taken
All Death Cases
Hendricks
Lake

PanViz


PanViz
File
View
Model Parameters
Import ISPIPT File
Exit
Color Counties By
Category:
Sick
Dead
Beds Filled
Show:
All Cases
New Cases
For Ages:
Under 18
18-64
65 and Over
Graph Scale
Percentage
Absolute
Decision Point Days
Media: 21
Schools: 21
SNS: 6
> 17.79%
> 15.56%
> 13.34%
> 11.12%
> 8.89%
> 6.67%
> 4.45%
> 2.22%
> 0.00%
Needed:2362
Clinton
Beds:25
Needed:629
Tipton
Beds:91
Needed:272
Madison
Beds:231
Needed:2820
Boone
Beds:48
Needed:943
Hamilton
Beds:144
Needed:5523
Hancock
Beds:101
Needed:997
Hendricks
Beds:141
Needed:2259
Marion
Beds:3999
Needed:18527
Putnam
Beds:85
Needed:517
Morgan
Beds:140
Needed:1037
Johnson
Beds:158
Needed:2146
Shelby
Beds:0
Needed:254
Owen
Beds:0
Needed:268
Monroe
Beds:811
Needed:2234
Brown
Beds:0
Needed:186
Bartholomew
Beds:256
Needed:945
Sick Cases
Percent of Population
Day #
Hospitalized Cases
Percent of Beds Taken
Death Cases
Jackson
Porter

Discovery Park

PURVAC
Purdue University

County Specific Scenario Plot for Display	
Selected County:	Adams

Spread Rate

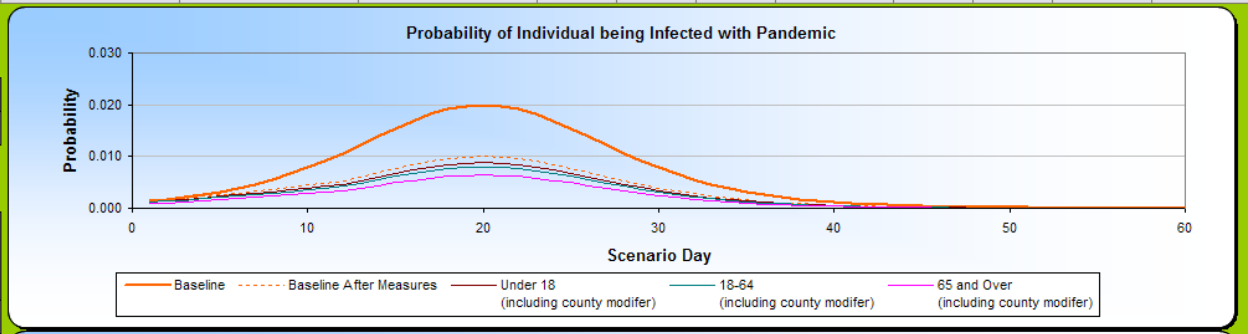
Point of Origin (currently Chicago, IL)	
Latitude	41.879536
Longitude	-87.624333
Miles/Day Traveled:	25.0

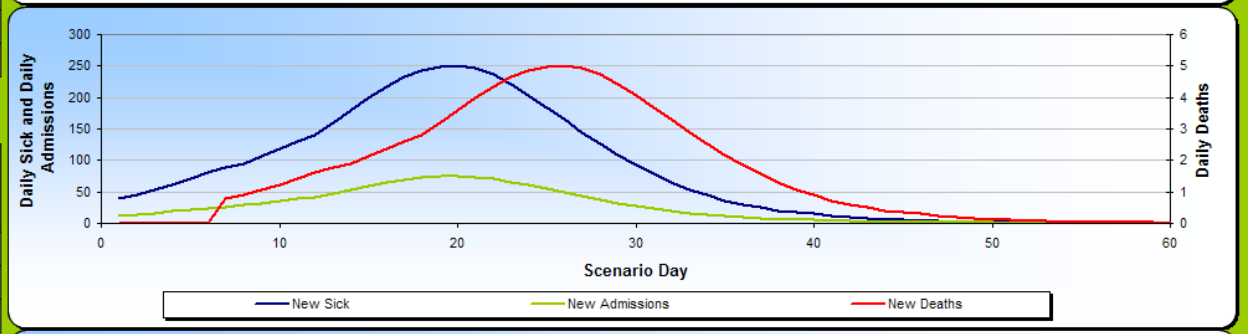
County Type Impact

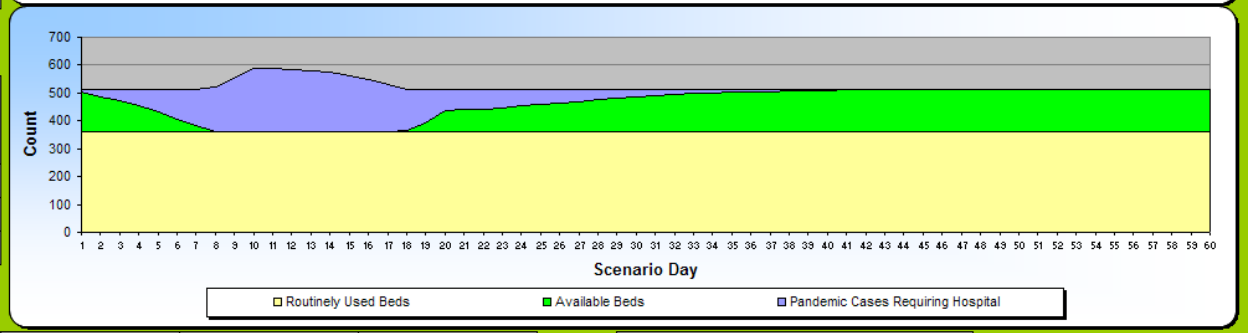
County Density Category	Modification to Baseline Pandemic Influenza Infection Probability
1 = Rural	0.80
2 = Small Towns	1.00
3 = Major Metropolitan	1.20

Demographic Impact

Demographic	Modification to Baseline Pandemic Influenza Infection Probability
Under 18	1.10
18-64	1.00
65 and Over	0.80







Adams
Scenario Summary

Category	30 Day Totals	60 Day Totals
Infected	4,482	4,921
Admitted	1,349	1,483
Deaths	73	98

Output Data for all Counties

Decision Measures

Measure	Meaure On (1) or Off (0)	% Reduction in Infection Probability	Day Implemented (from scenario start on day 1)	Day Measure Reaches Full Impact (number of days after being implemented)
Media	1	10.00	2	2
Close Schools	1	15.00	4	5
SNS	1	25.00	6	7

Hospital Bed Model (Enter 1 or 2)	
Enter Choice:	2

Global Parameters (these can be overriden county by county by going to the specific county tab)	
Mortality Rate:	2.0%
Recover Rate:	98.0%
Hospitalization Rate:	30.0%
Typical Hospital Capacity:	70.0%
Mean Time to Recover (days):	10
Mean Time in Hospital (days):	6
Mean Time to Die (days):	6

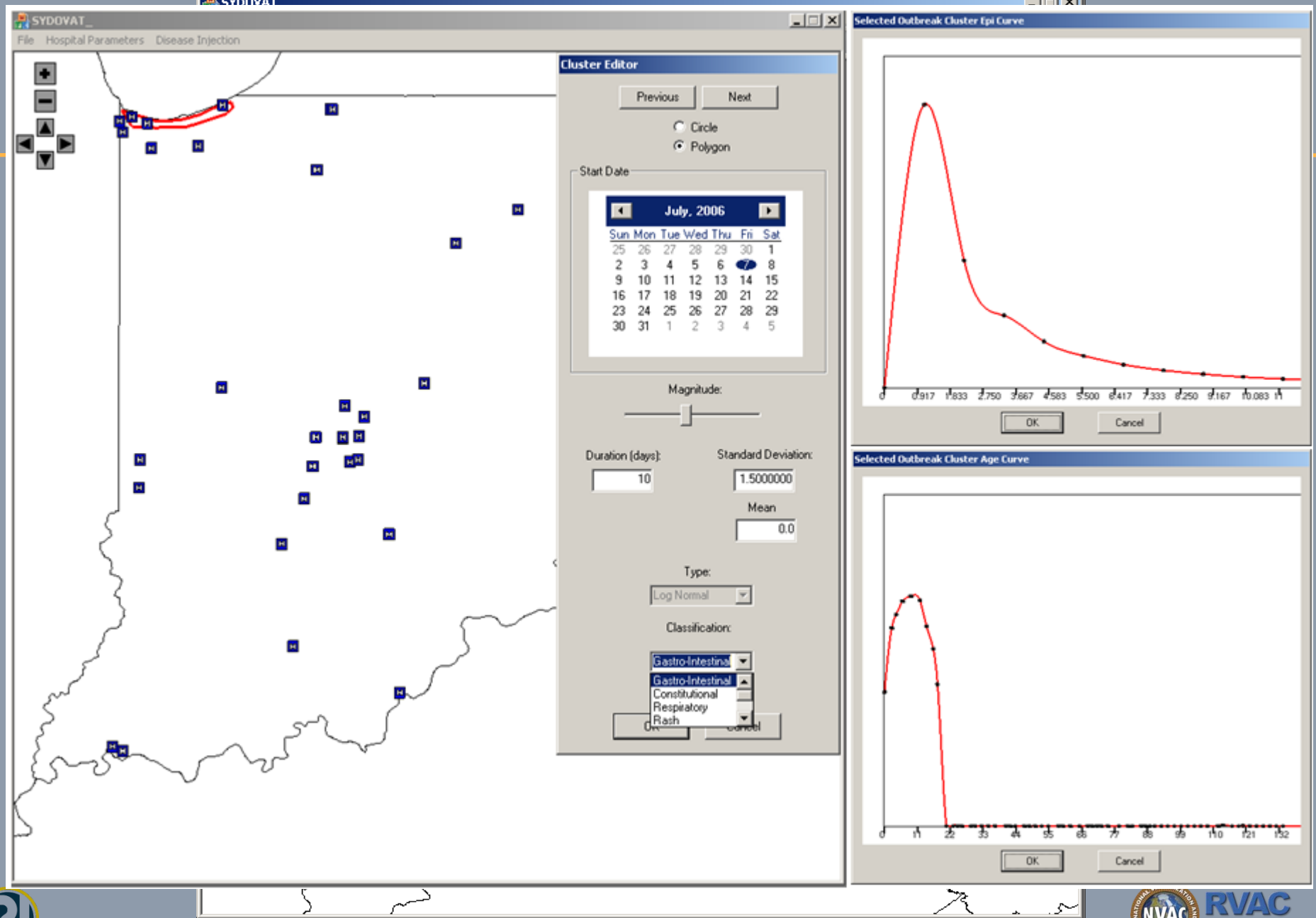
Synthetic Data Generation

System Features:

- User defined ED placement
- Adjustable population probability density control
- Adjustable demographic probability control
- User defined spatiotemporal outbreak controls

Benefits:

- No privacy concerns
- Data is not aggregated
- Known outbreaks to test tools


SYDOVAT_
File
Hospital Parameters
Disease Injection
Cluster Editor
Previous
Next
Circle
Polygon
Start Date
July, 2006
Sun Mon Tue Wed Thu Fri Sat
25 26 27 28 29 30 1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30 31 1 2 3 4 5
Magnitude:
Duration (days):
10
Standard Deviation:
1.5000000
Mean
0.0
Type:
Log Normal
Classification:
Gastro-Intestinal
Gastro-Intestinal
Constitutional
Respiratory
Rash
OK
Cancel
Selected Outbreak Cluster Epi Curve
0 0.917 1.833 2.750 3.667 4.583 5.500 6.417 7.333 8.250 9.167 10.083 11
OK
Cancel
Selected Outbreak Cluster Age Curve
0 11 22 33 44 55 66 77 88 99 110 121 132
OK
Cancel

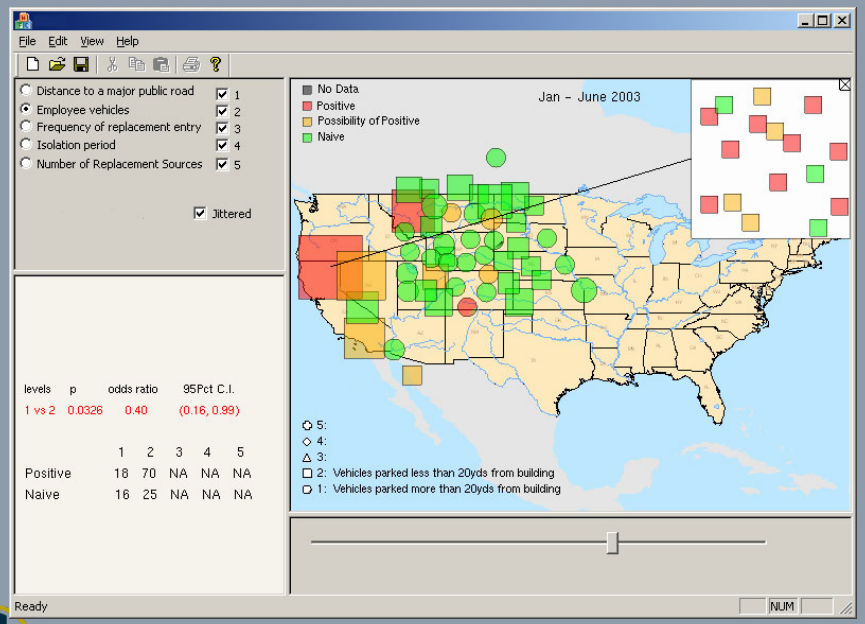
Health Monitoring and Management Visual Analytics: Impact and Relevance

- *Analyzed public health effects from chemical spill in Fairburn, Georgia.*
- *Analyzed syndromic spread factors for national veterinary association to reduce production losses and disease spread.*
- *Evaluated effectiveness of and proposed improvements to Indiana pandemic influenza monitoring and management.*
- *Developing linked animal-human health surveillance system for more timely and accurate health monitoring and response.*
- *Investigated correlation of influenza system in pets and humans for improved influenza planning and response.*
- *Evaluated use of ProMed and VIN data sources for outbreak/event detection and monitoring*
- *Analyzed day of week effect variance across hospitals in PHESS*

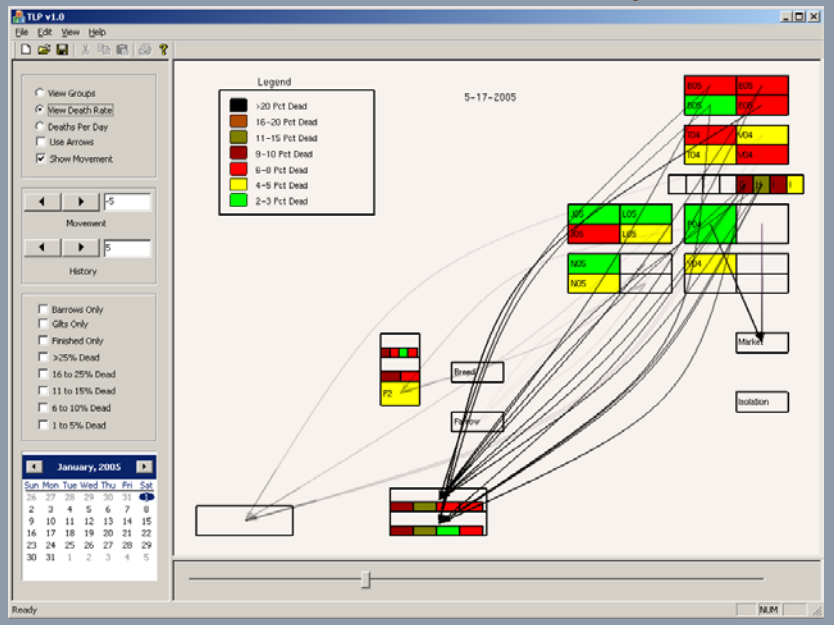
Examples: Zoonotic Syndromic Surveillance

Goals

- Improve disease surveillance and management at local, state and national levels
- Disease/loss tracking in food production sites
- Abstraction/simplification of a large number of events
- Maintain privacy while preserving information coherency
- Develop analytic tools for multisource data, factor correlation, and exploration



Example of nation-wide syndromic surveillance



Example of disease/loss tracking in a food animal production site

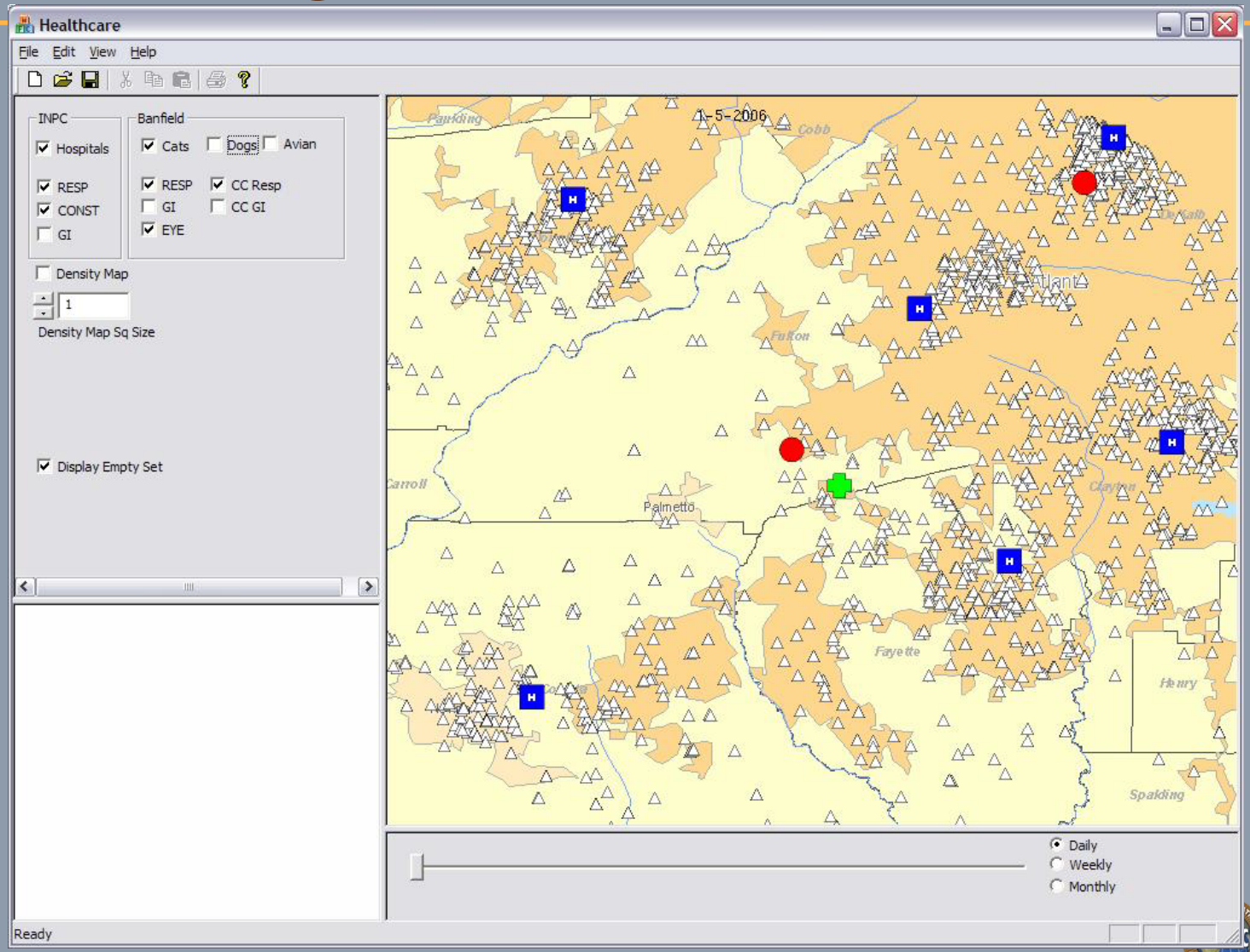
Case Study: Assessing Effects of a Chemical Release (Study Description)

In 2006 industrial wastewater was accidentally released prompting a public health investigation.

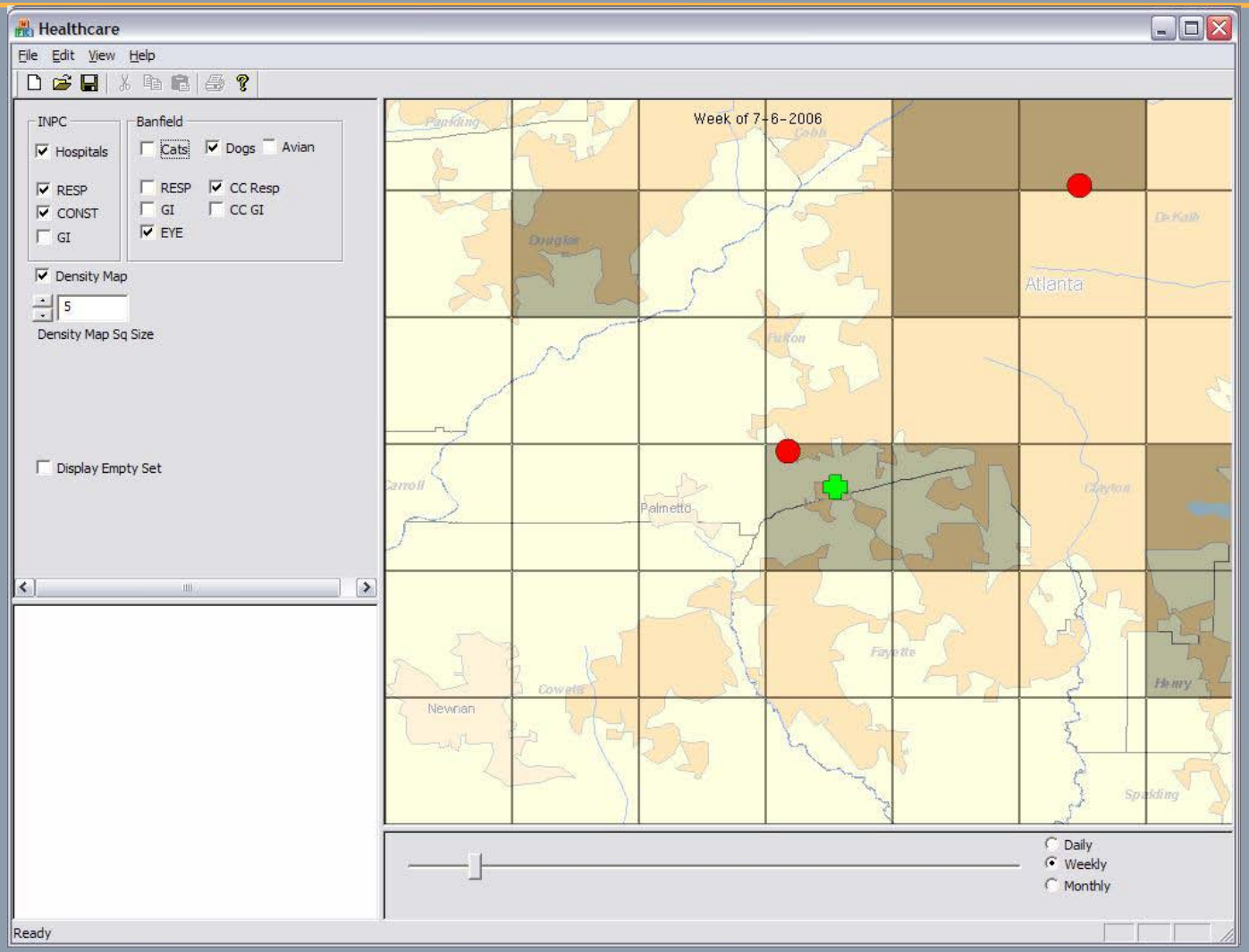
Results of the investigation were self-reported surveys, no emergency room data was available.

Can companion animal health be used to determine the effects of such a release?

Pet Coverage



Dog Eye Inflammation



Case Study: Assessing Effects of a Chemical Release (Study Results)