GIS

The Geographic Approach for the Nation

ESRI Federal User Conference

Washington, D.C. • February 17–19, 2010

Introduction to Spatial Statistics

Lauren Rosenshein

Presentation Outline

- **Spatial Statistics Overview**
- **Spatial Pattern Analysis**
  - Descriptive spatial statistics
  - Global and local spatial autocorrelation statistics
    - What is a *z* score? What is a *p*-value
  - Spatial weights
- **Additional Resources**

DEMOS

- **The spatial pattern of piracy**
- **Exploring Childhood Obesity using Hot Spot Analysis**
- **Putting it all together: The Pocketman**
- **Tips for navigating online resources**

What are spatial statistics in a GIS environment?

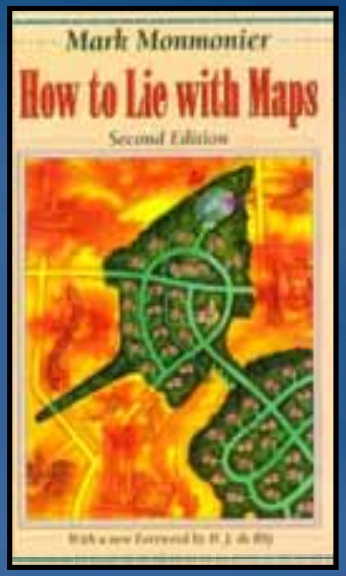
- **Software-based** tools, methods, and techniques developed specifically for **use with geographic data**.
- Spatial statistics:
  - **Describe and model** spatial distributions, spatial patterns, spatial processes, and spatial relationships.
  - **Incorporate space** (area, length, proximity, orientation, and/or spatial relationships) **directly into** their **mathematics**.

In many ways spatial statistics extend what the eyes and mind do intuitively to assess spatial patterns, trends and relationships.

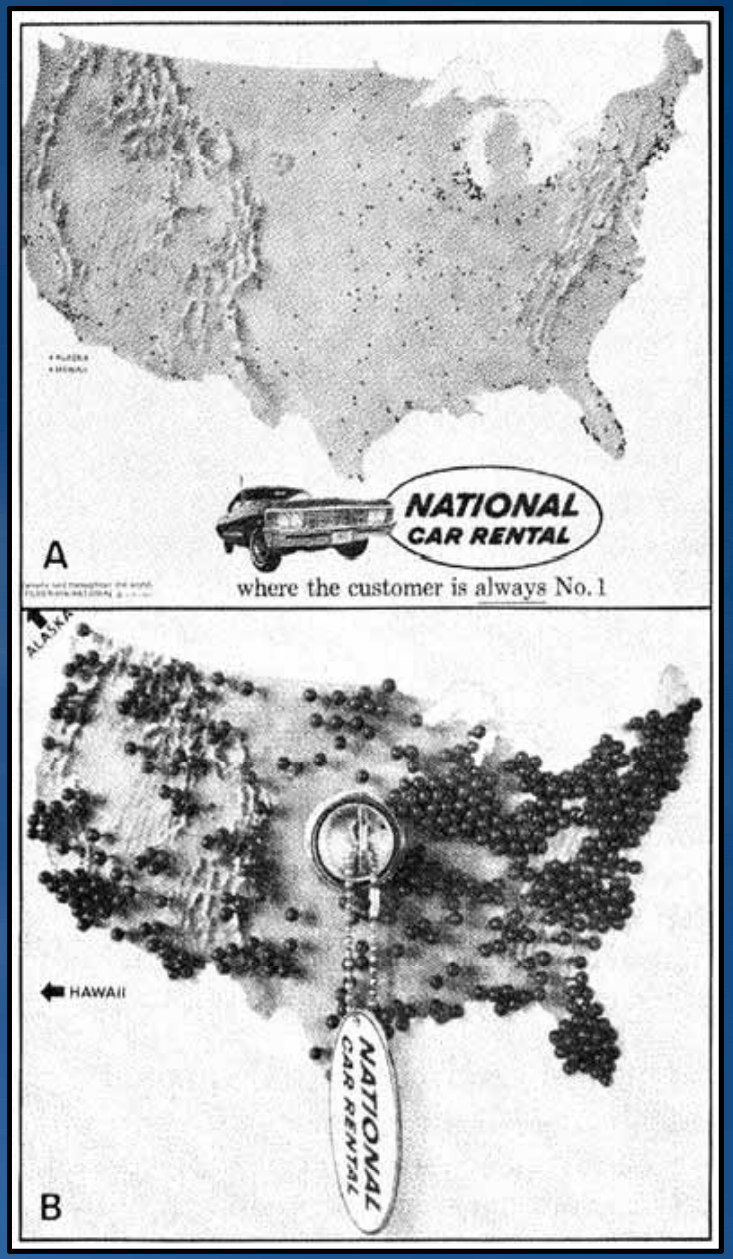
Why use spatial statistics?

Spatial Statistics help us assess:

- Patterns
- Relationships
- Trends



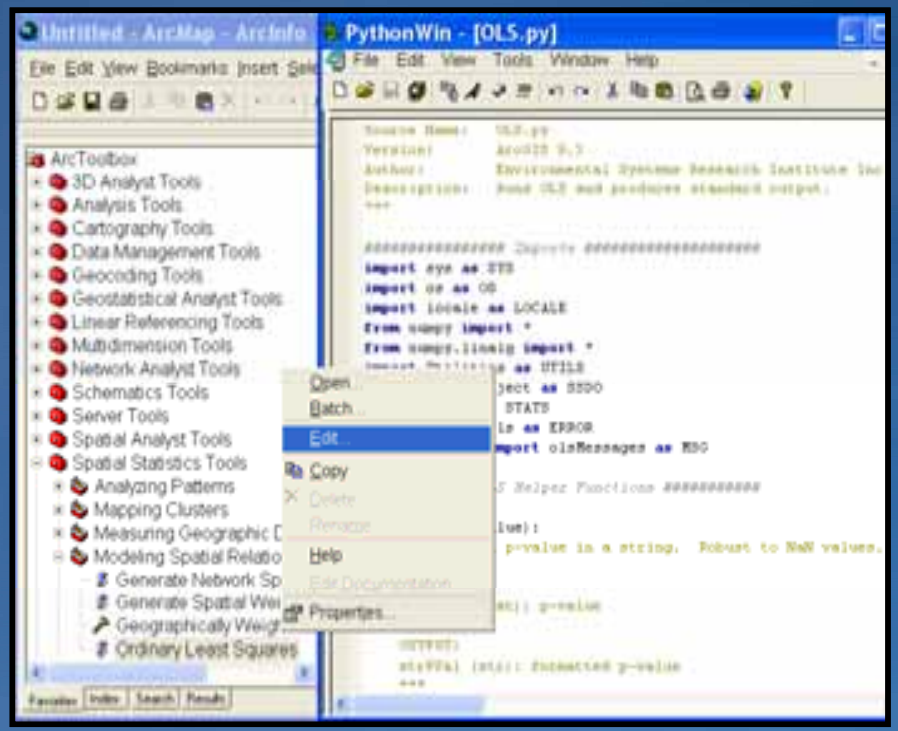
How we present our results (colors, class breaks, symbols…) can either enhance or obscure communication.

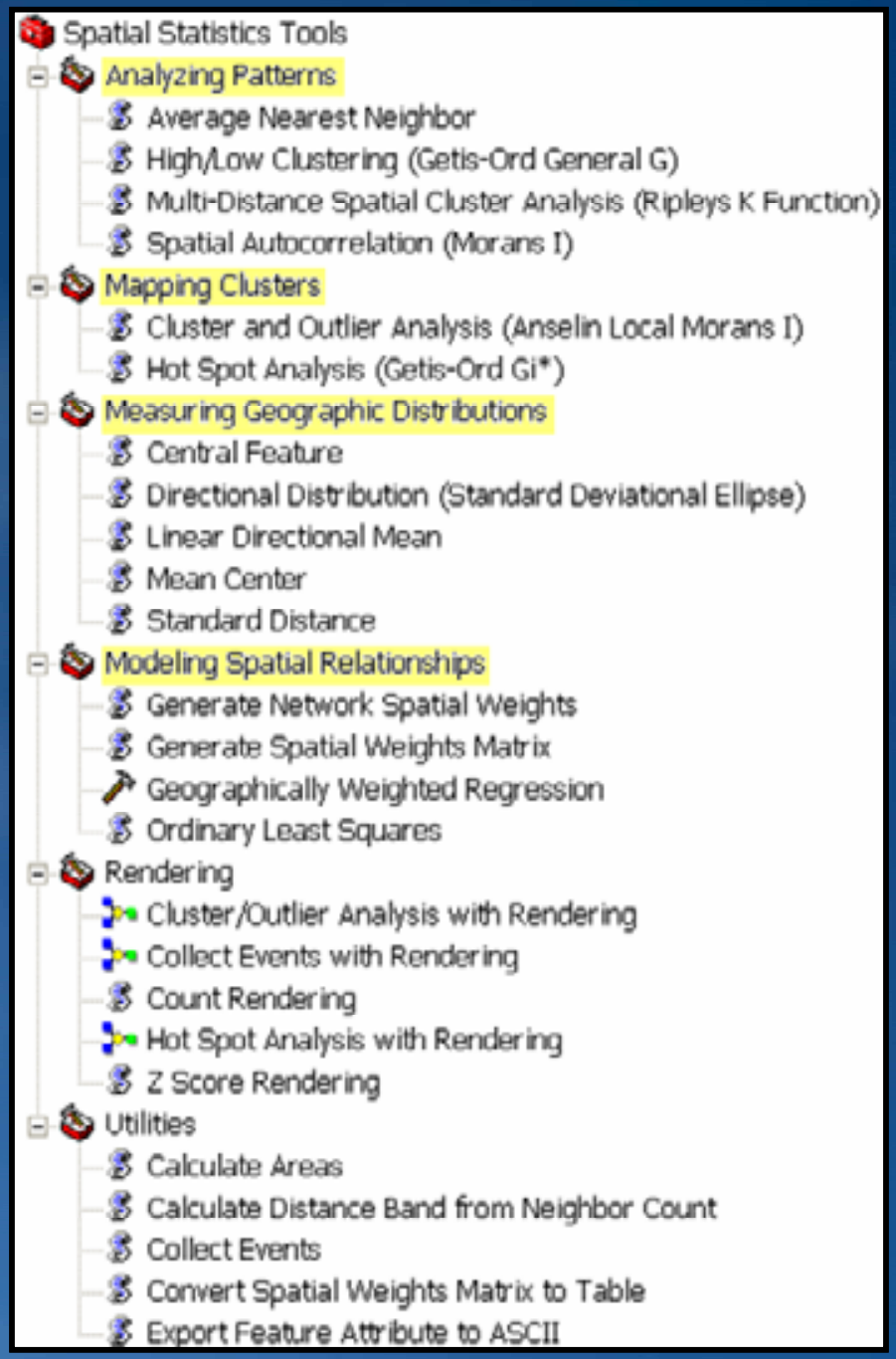


Two ads for National Car Rental. The lower map ad replaced the upper map ad a year later.

Spatial Statistics Toolbox in ArcGIS

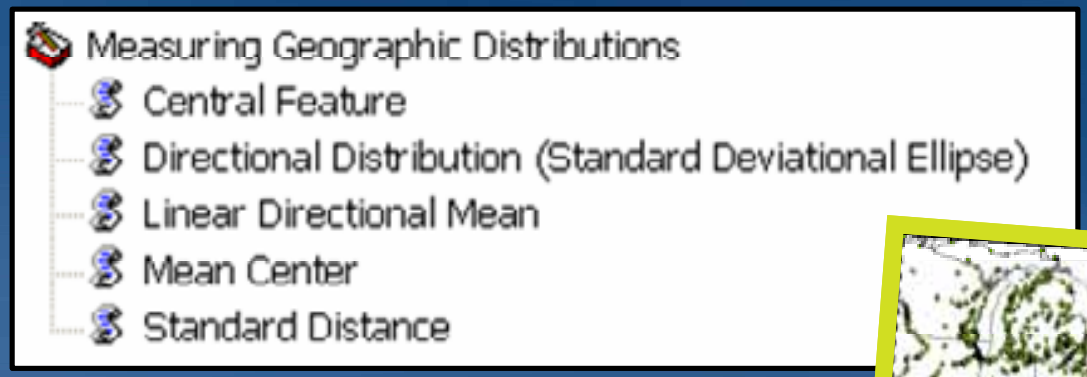
- **Core functionality with ArcGIS (not an extension).**
- **Most tools delivered with their source code.**
- **Most tools available at all license levels.**

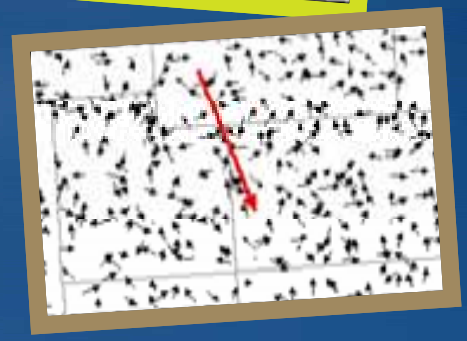


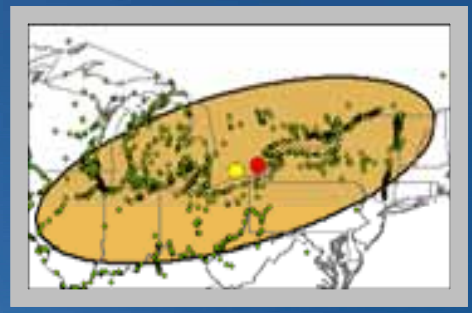


Spatial Statistics Toolbox

- Measuring Geographic Distribution
- Analyzing Patterns
- Mapping Clusters
- Modeling Spatial Relationships





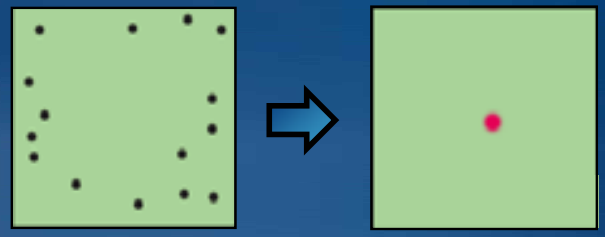


- **Questions**
  - **Which site is most accessible?**
  - **Is there a directional trend to the spatial distribution of the incidents?**
  - **What is the primary wind direction for this region in the winter?**
  - **Where is the population center?**
  - **Which gang has the broadest territory?**

Mean Center

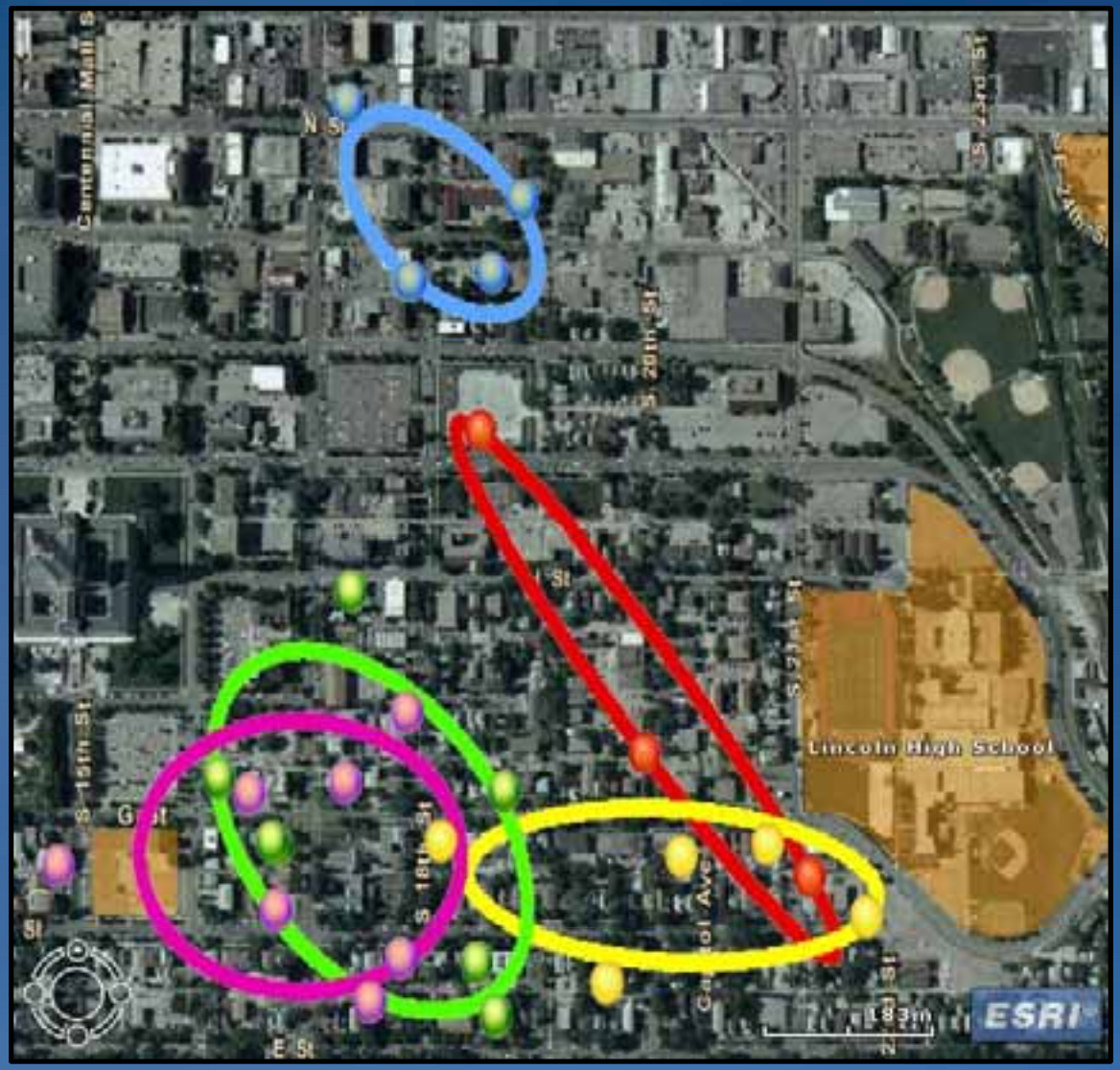
- **Computes the average x and y coordinate, based on all features in the study area.**

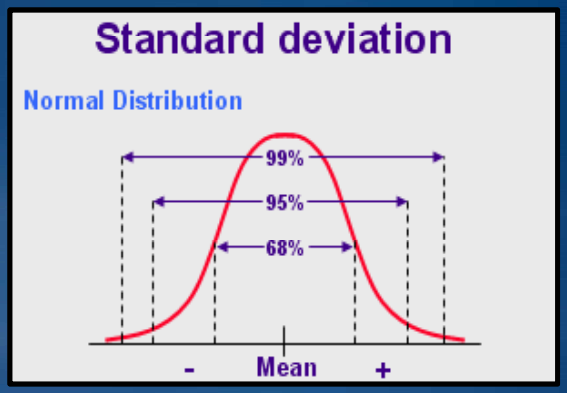
$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$



Directional Distribution
(Standard Deviational Ellipse)

- Abstracting spatial trends in a distribution of features
- Comparing distributions over time





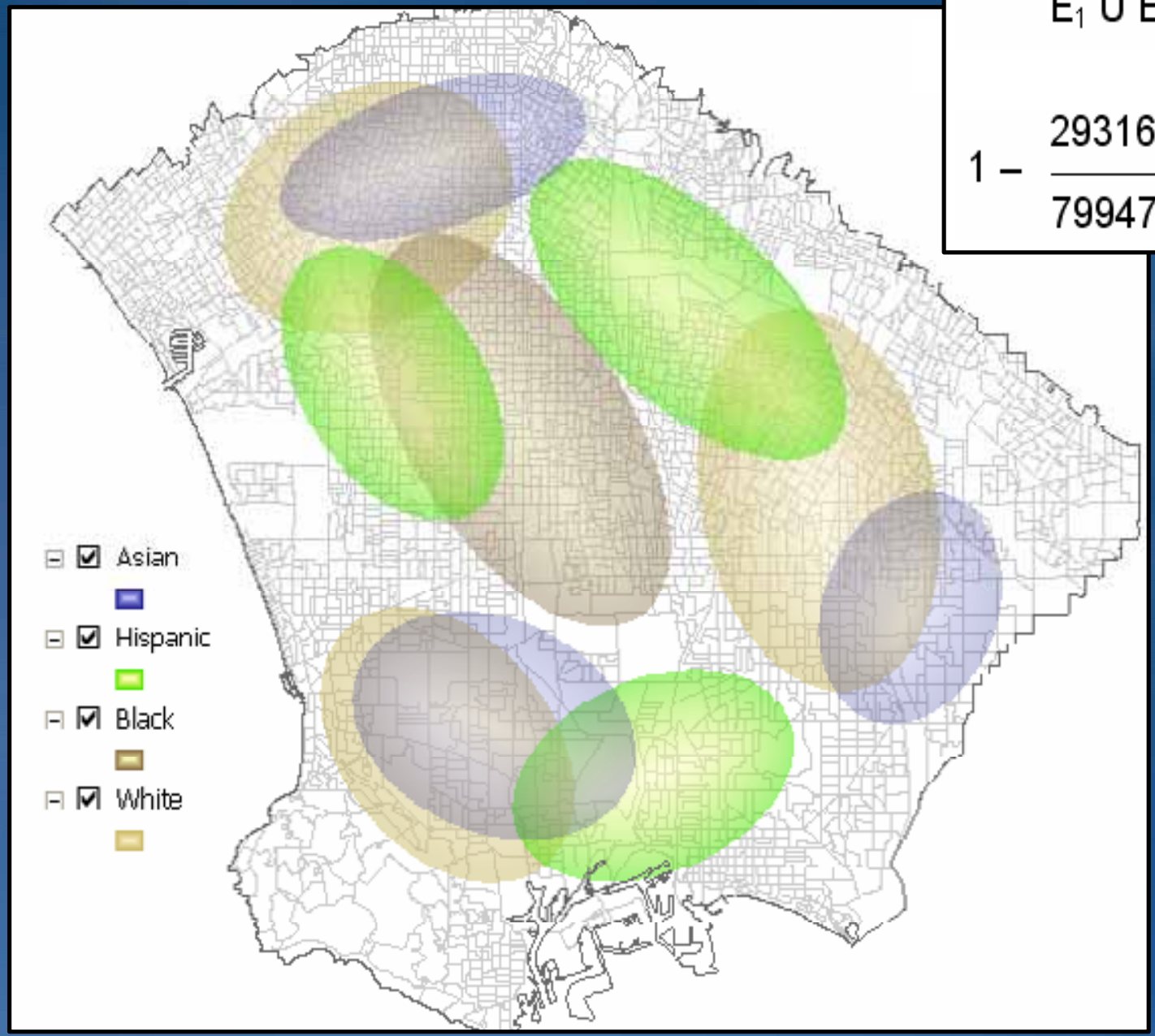
- 1 = 68% of features
- 2 = 95% of features
- 3 = 99% of features

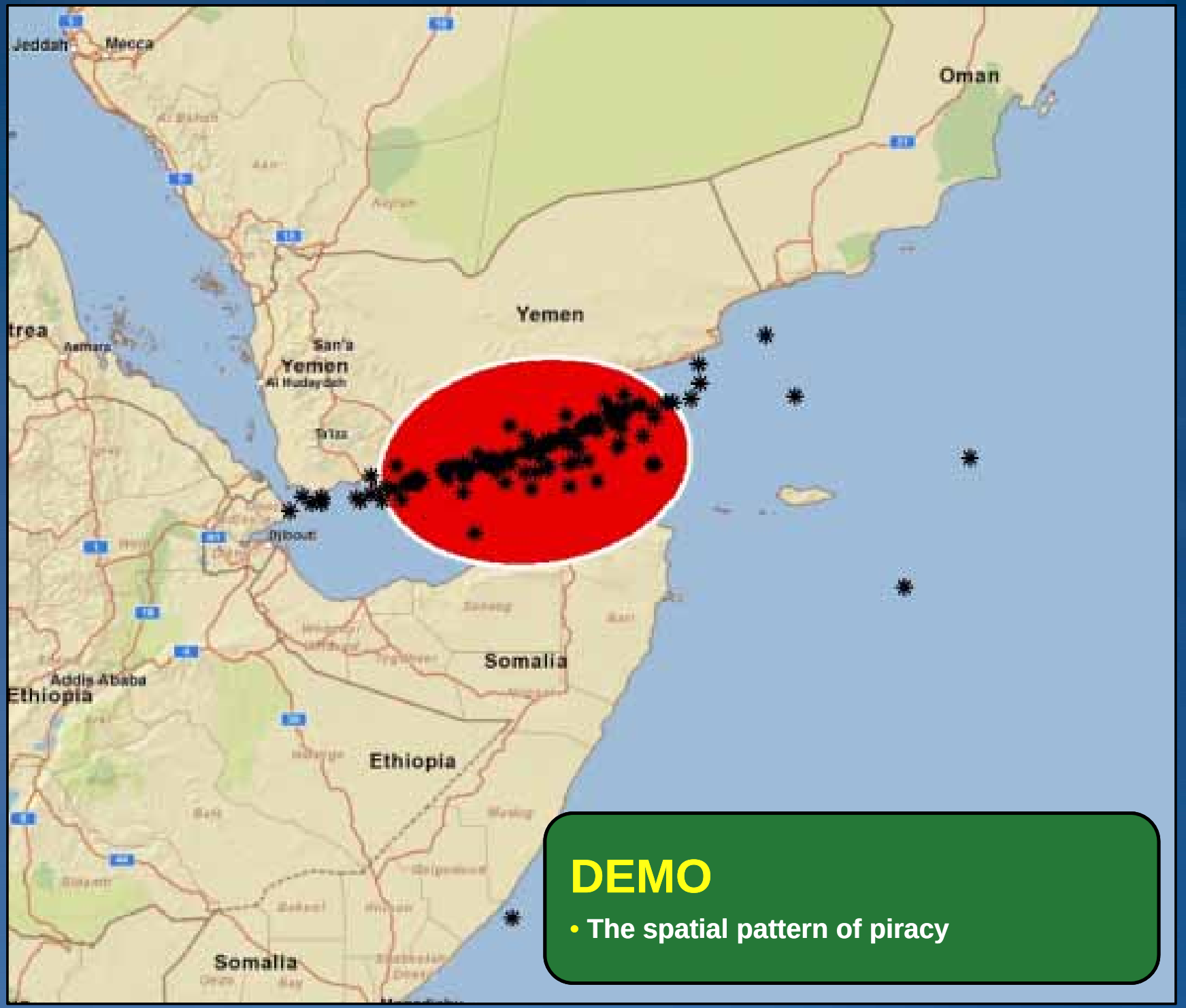
Directional Distribution
(Standard Deviational Ellipse)

Segregation Index =

$$1 - \frac{E_1 \cap E_2 \cap E_3 \cap \ldots E_n}{E_1 \cup E_2 \cup E_3 \cup \ldots E_n} =$$

$$1 - \frac{2931680545.83}{7994760004.92} = \mathbf{0.63}$$





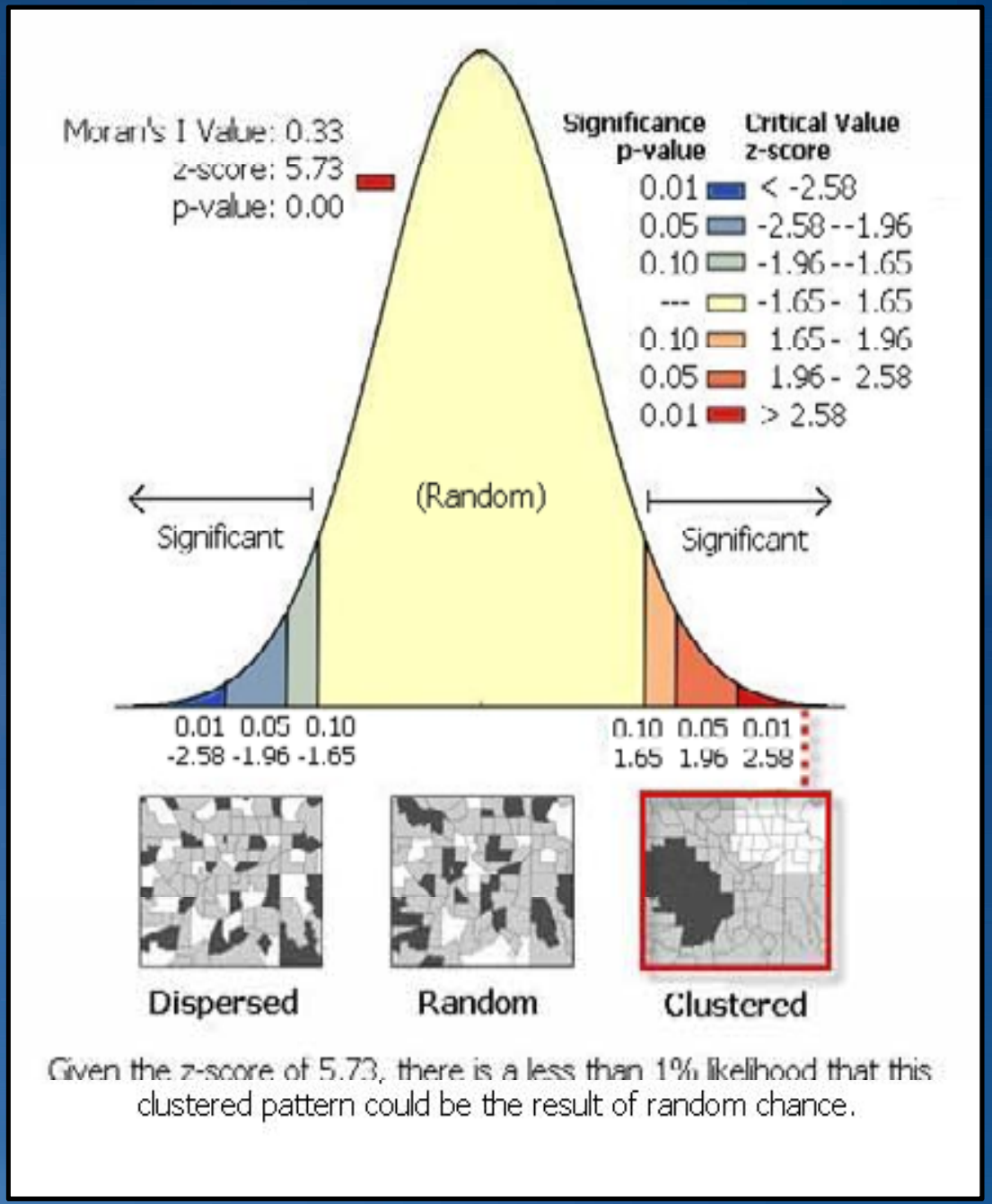
DEMO

- The spatial pattern of piracy

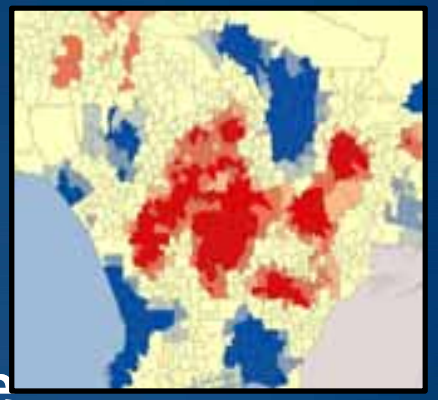
What is a *z* score? What is a *p*-value?

- The null hypothesis for the ArcGIS Spatial Pattern Analysis tools is CSR: Complete Spatial Randomness
- Reject the null hypothesis if the result (the p-value/z score) is statistically significant

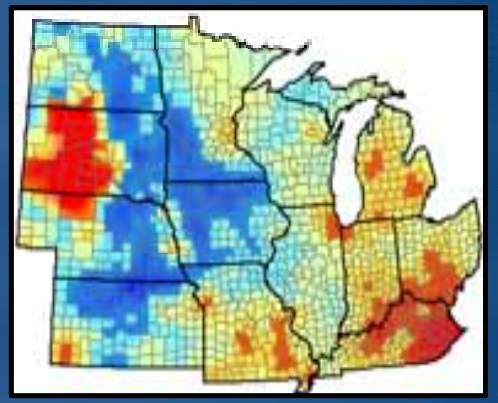
Z Score (Standard Deviations)	P-Value (Probability)	Confidence Level
+/-1.65	0.10	90%
+/-1.96	0.05	95%
+/-2.58	0.01	99%

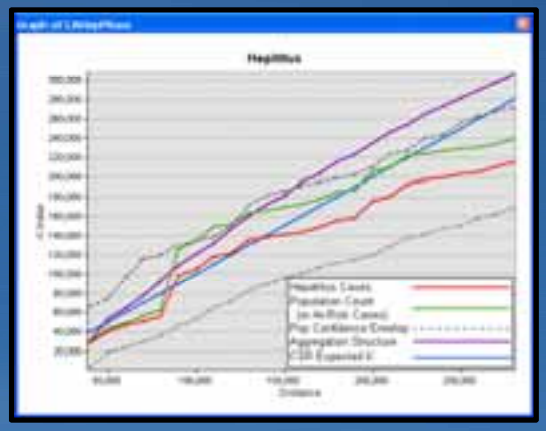


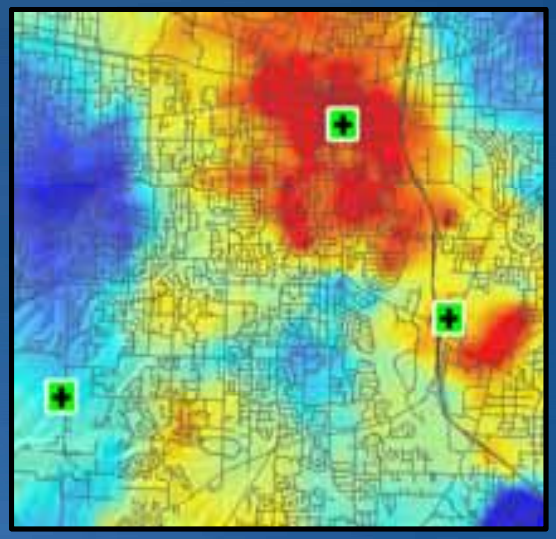
Is CSR useful?



- Raising the bar:
  - Normalize the analysis field to create a rate
  - Analyze average values
  - Compare *z* score magnitudes
    - Across space
    - Over time
    - Among control spatial distributions

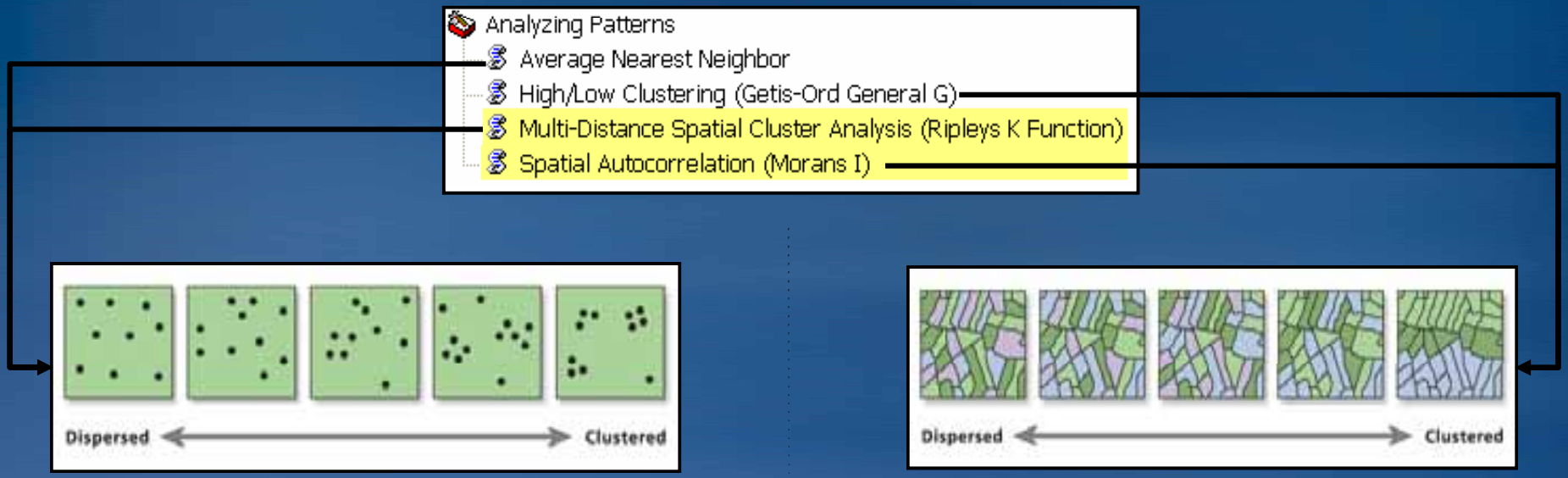






Spatial Statistics Toolbox

- Measuring Geographic Distribution
- **Analyzing Patterns**
- Mapping Clusters
- Modeling Spatial Relationships

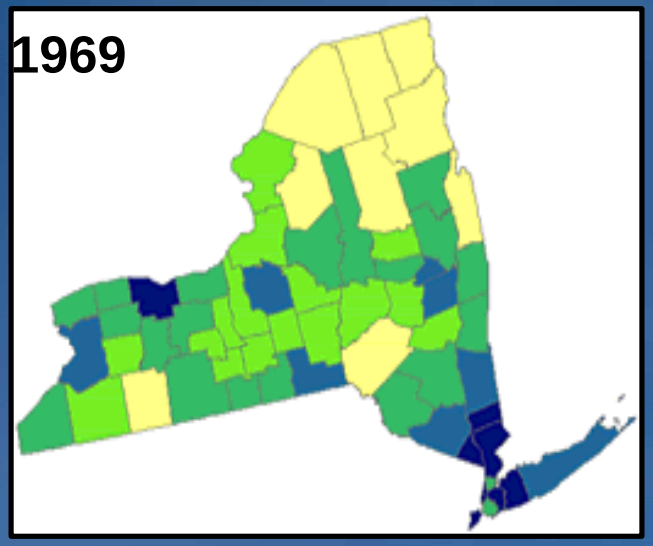


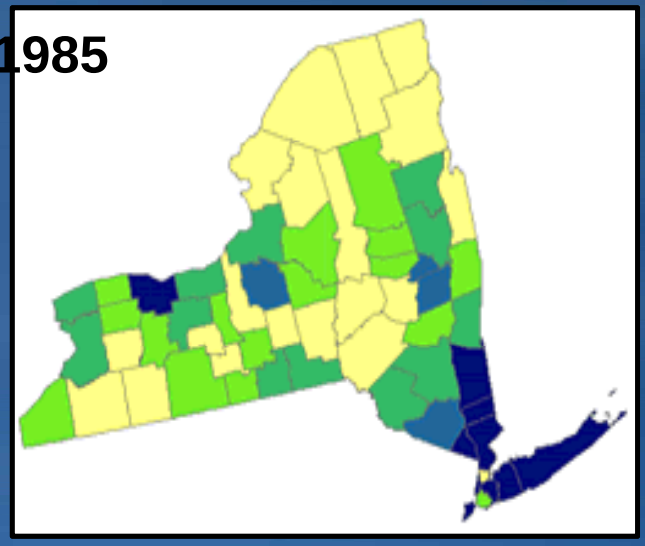
- **Which plant species is most concentrated?**
- **Does the spatial pattern of the disease mirror the spatial pattern of the population at risk?**
- **Is there an unexpected spike in pharmaceutical purchases?**
- **Are new AIDs cases remaining Geographically fixed?**

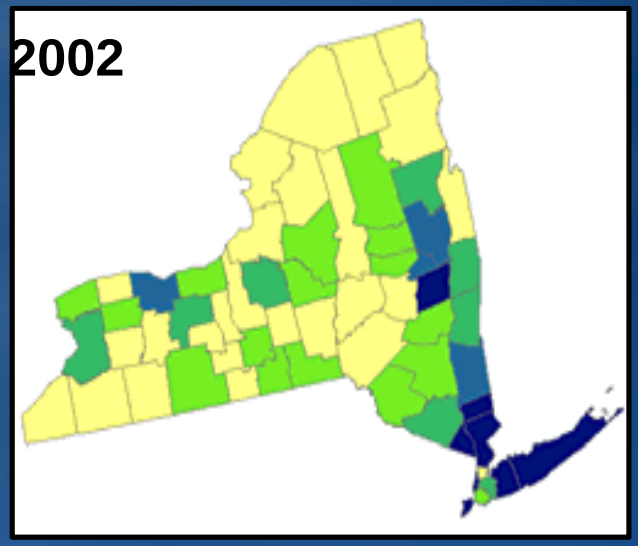
Spatial Autocorrelation (Global Moran's I)

- This tool measures spatial clustering/dispersion
- Results are based on both feature locations and attributes

Thematic Maps showing Relative Per Capita Income for New York, 1969 to 2002





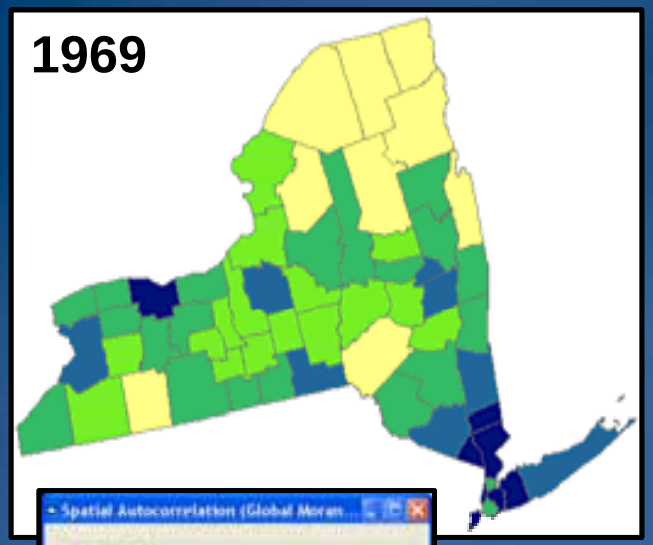


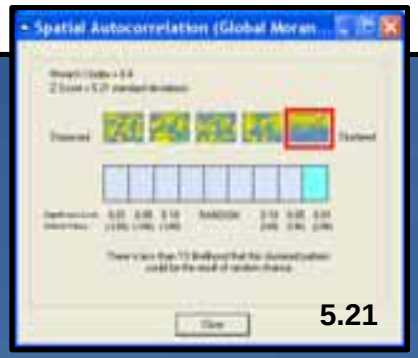
Are rich and poor becoming more or less spatially segregated?

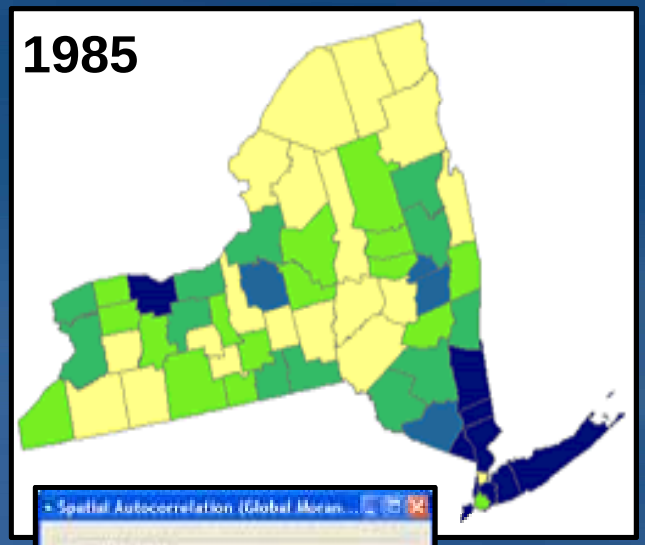
(It's difficult to answer this question looking at thematic maps alone).

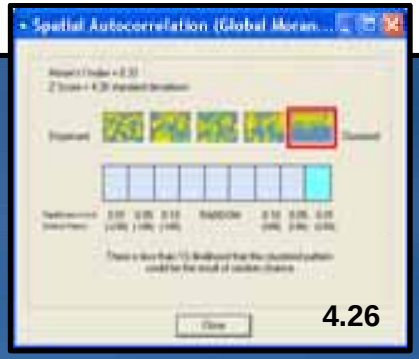
Spatial Autocorrelation (Global Moran's I)

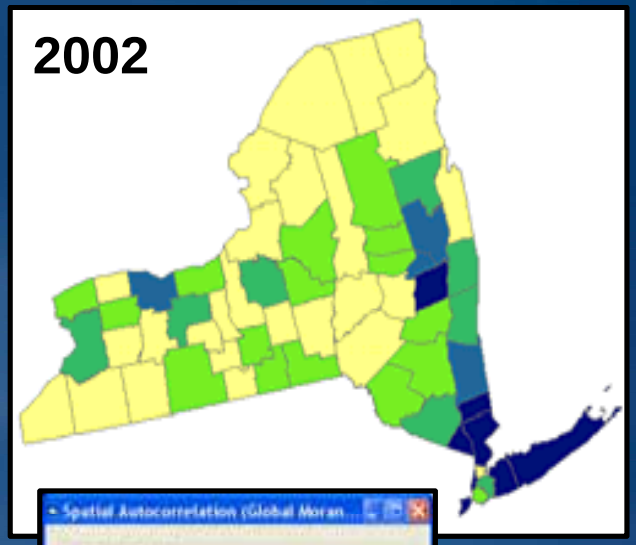
Relative Per Capita Income for New York, 1969 to 2002



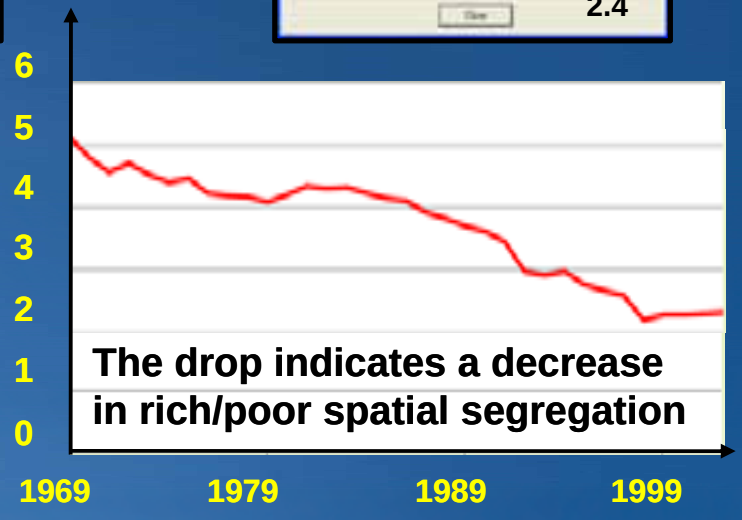






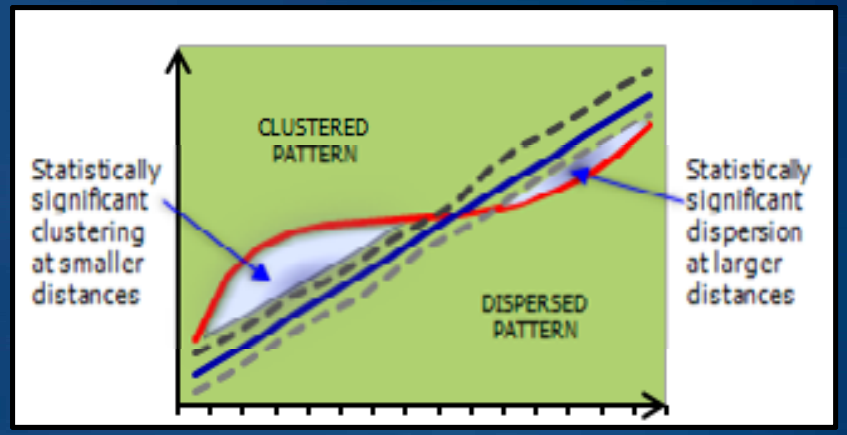


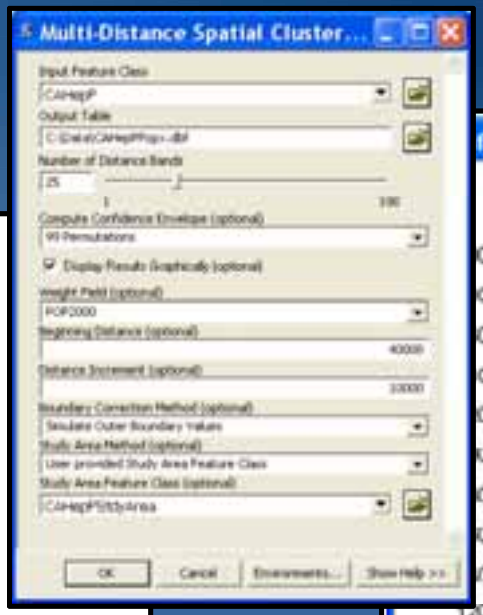
Plot the Z Score from the Global Spatial Autocorrelation tool to reveals broad trends over time.

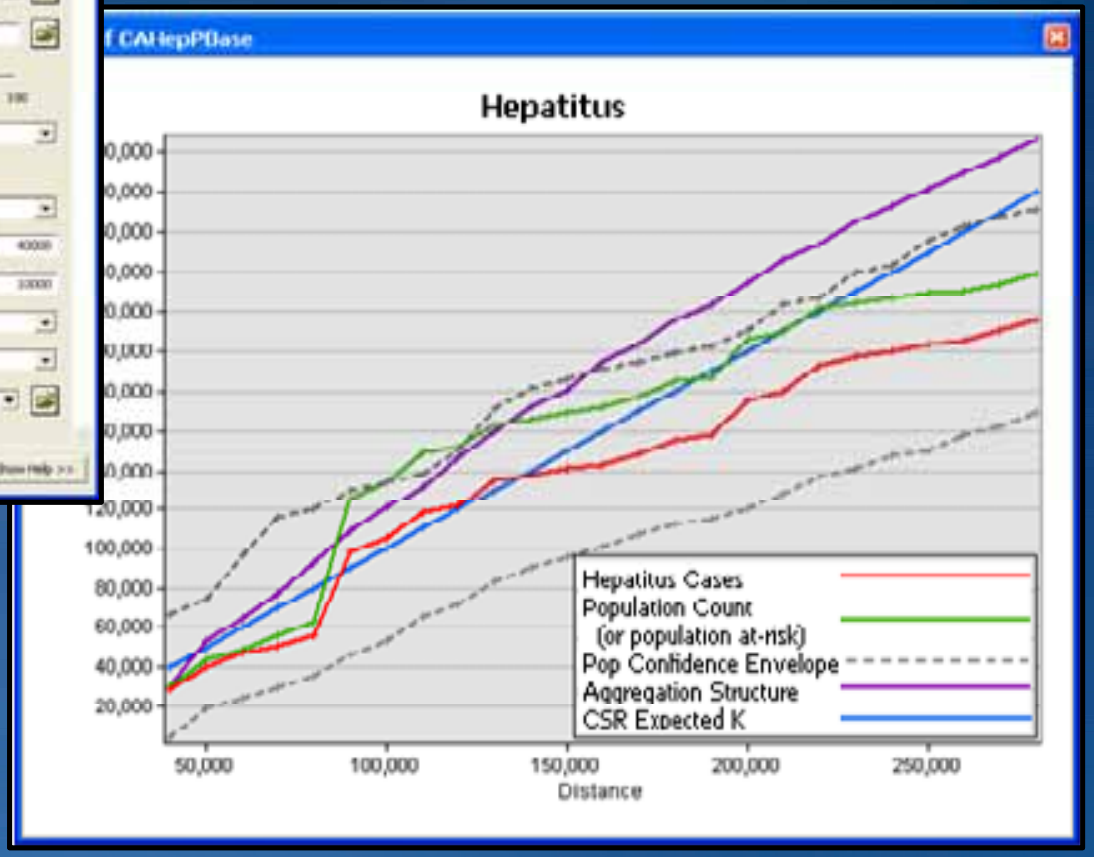


K Function

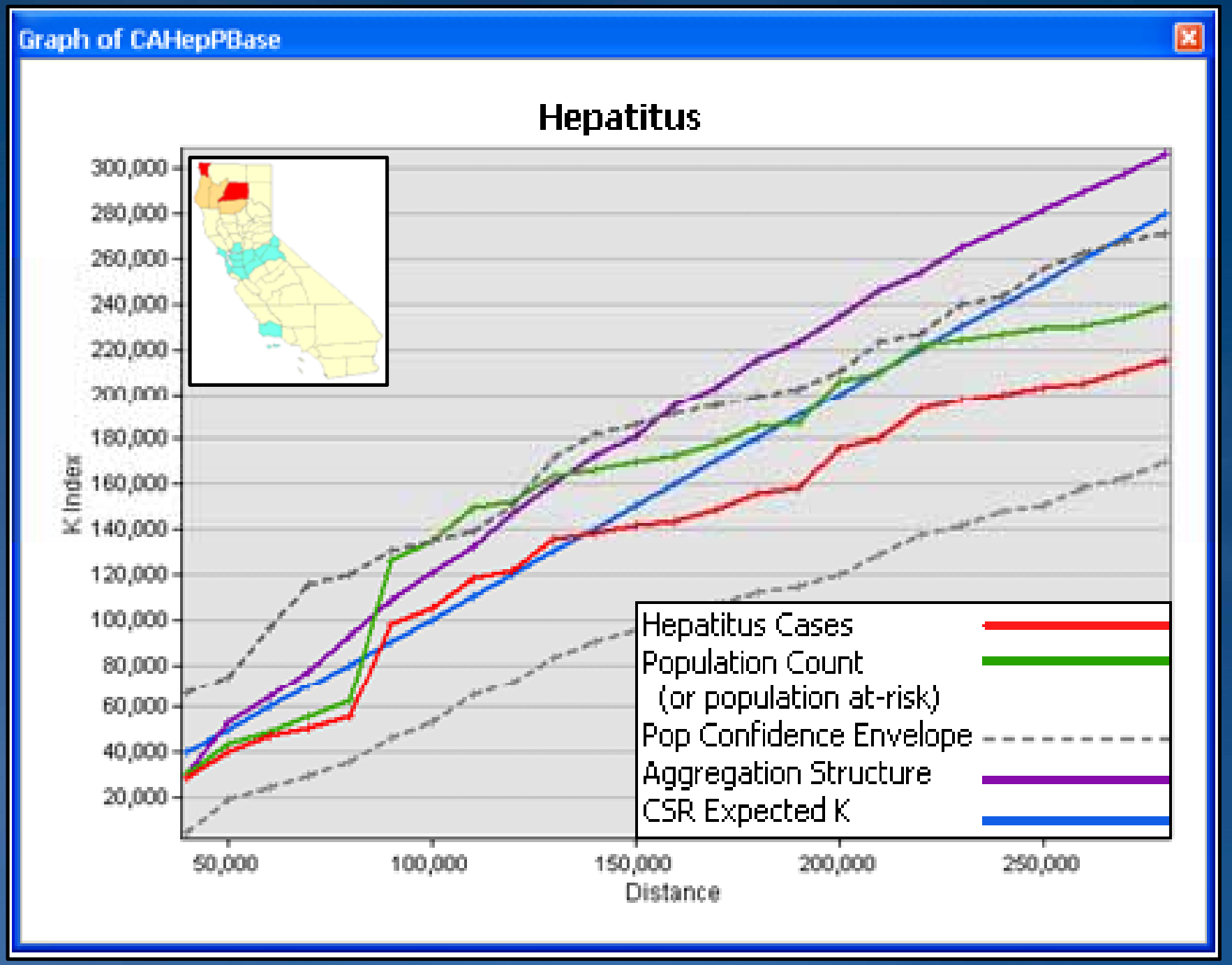
- Counts number of pairs within distance *d* of each feature.







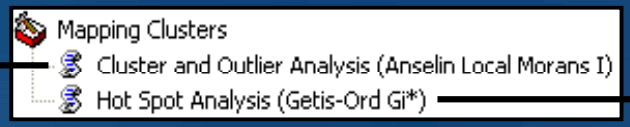
The processes promoting hepatitis (red line) are strongly influenced by the spatial pattern of population (green line).

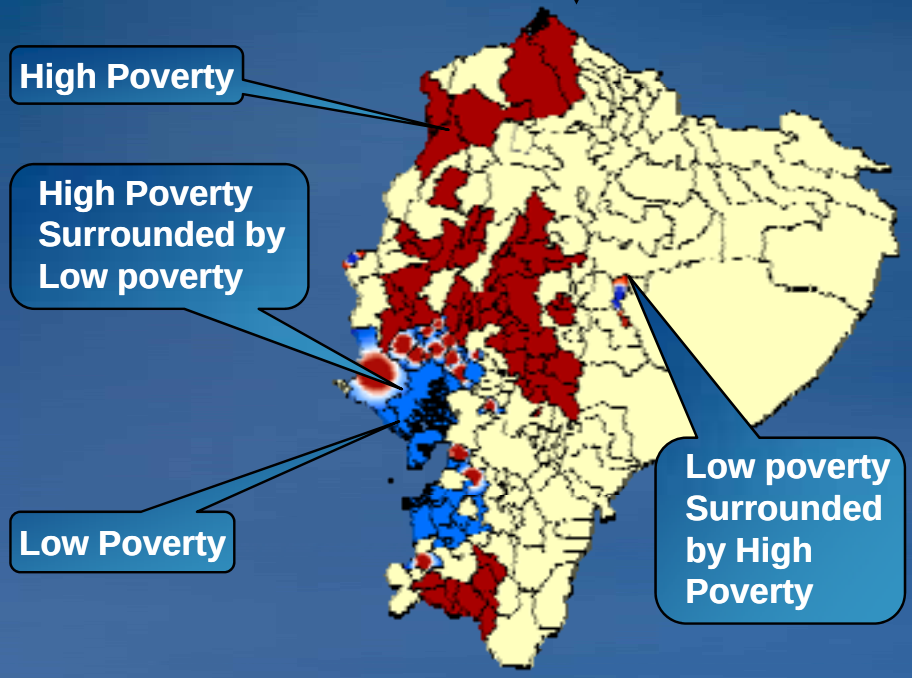


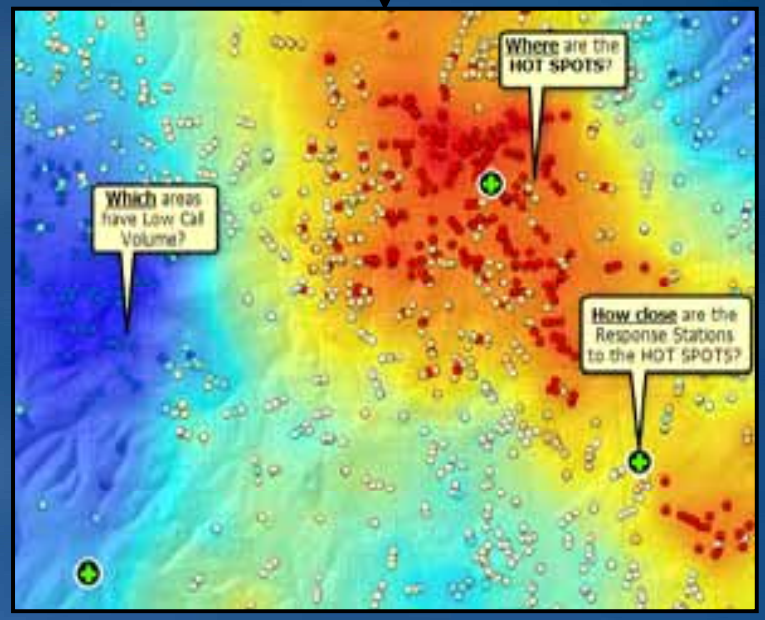
The processes promoting hepatitis (red line) are strongly influenced by the spatial pattern of population (green line).

Spatial Statistics Toolbox

- Measuring Geographic Distribution
- Analyzing Patterns
- **Mapping Clusters**
- Modeling Spatial Relationships



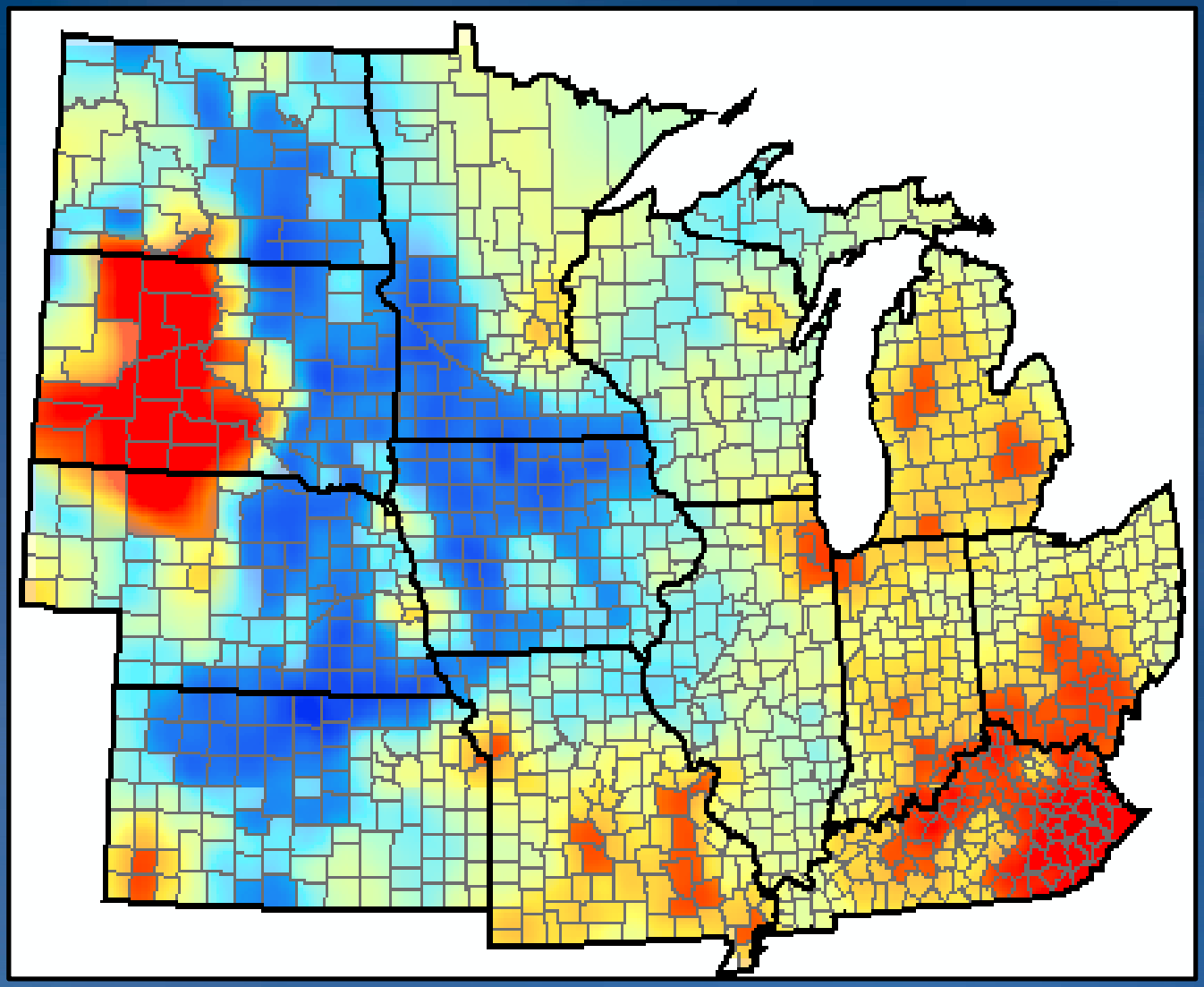




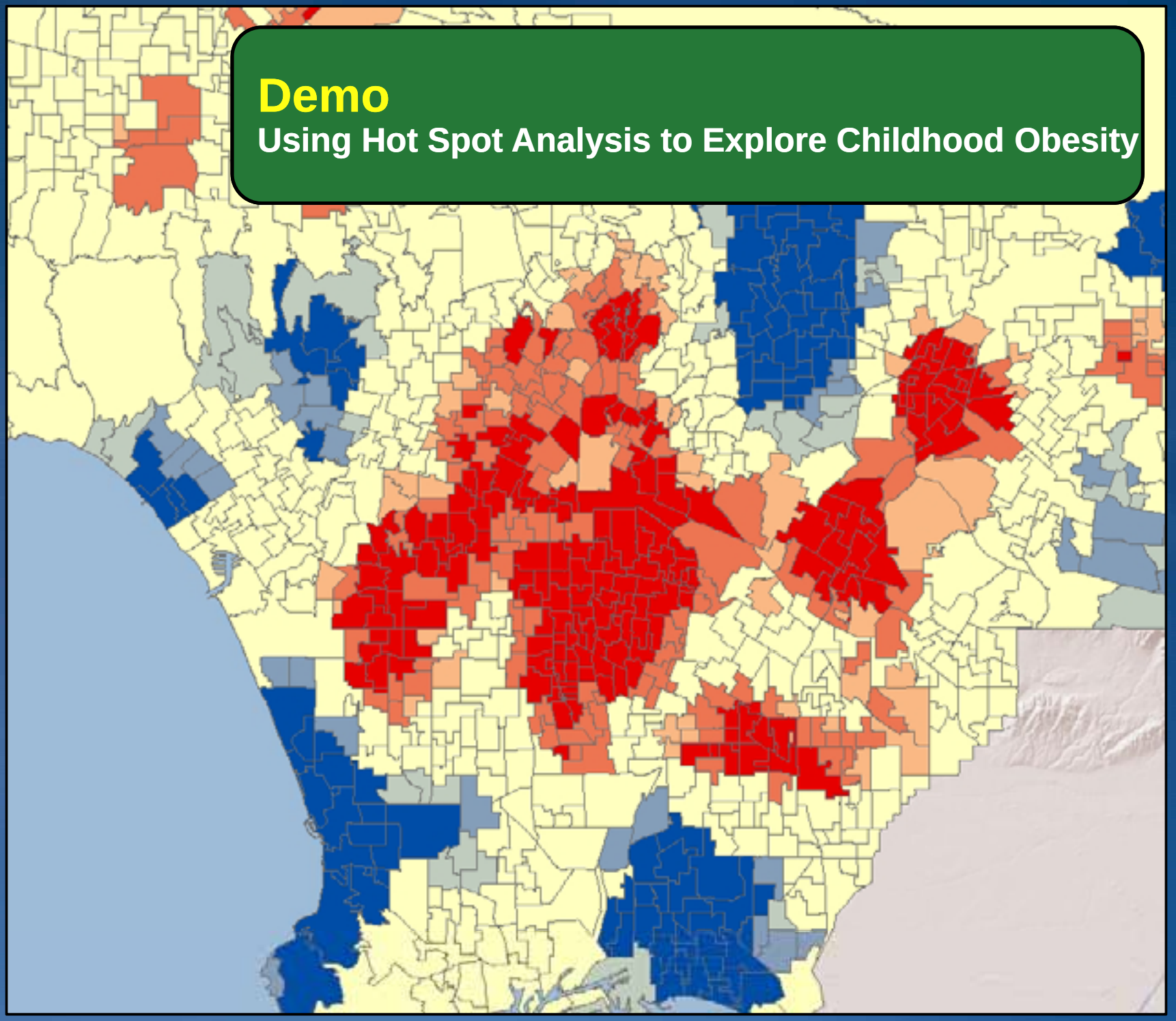
- Where are there sharp boundaries between affluence and poverty in Ecuador?
- Where do we find anomalous spending patterns in Los Angeles?

- Where are the 911 Call Hot Spots?
- Where do we see unexpectedly high rates of diabetes?

Average Age of Death Hot Spot Analysis

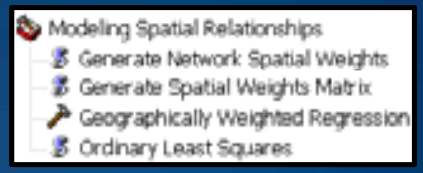


GiZScore	GiPValue
1.09548	0.273308
3.49006	0.000483
1.047	0.295098
1.159	0.246455
1.32213	0.186124
-2.30853	0.02097
-1.23803	0.215706
-1.59006	0.111821
-2.26428	0.023557
-0.665581	0.505679
-1.87229	0.061167
-2.25941	0.023858
-4.01195	0.00006
-2.85555	0.004296
-2.41466	0.01575


Demo
Using Hot Spot Analysis to Explore Childhood Obesity

Spatial Statistic

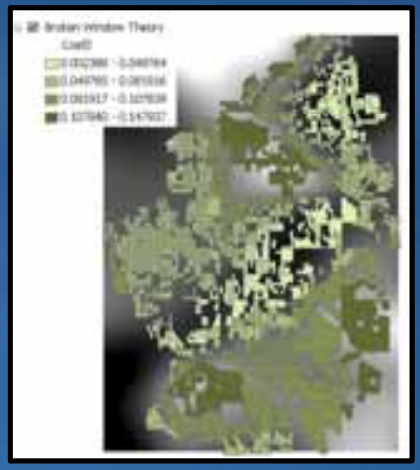
Measuring Geographic Distribution | Analyzing Patterns | Mapping Clusters | **Modeling Spatial Relationships**

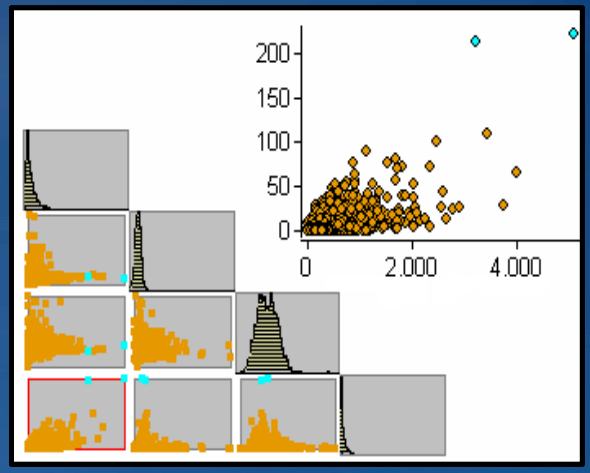


- Can I model spatial relationships based on a real road network?
- Are spatial weights matrix files editable, sharable, re-usable?
- Can I create a custom spatial weights matrix file?

Construct spatial weights matrix files

- What is the relationship between educational attainment and income?
- Is there a relationship between income and public transportation usage? Is that relationship consistent across the study area?
- Where are real estate values likely to go up?

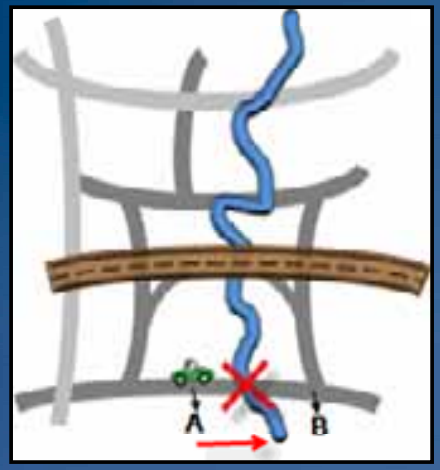




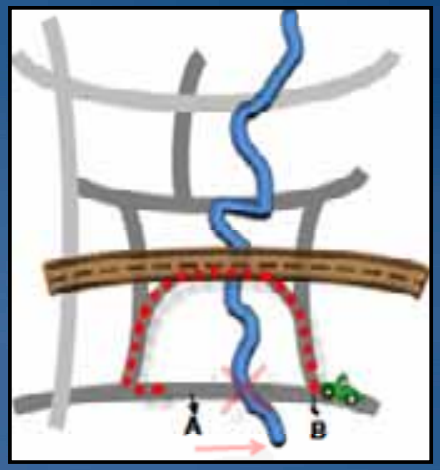
Ordinary Least Squares
Geographically Weighted Regression

Constructing spatial weights matrix files

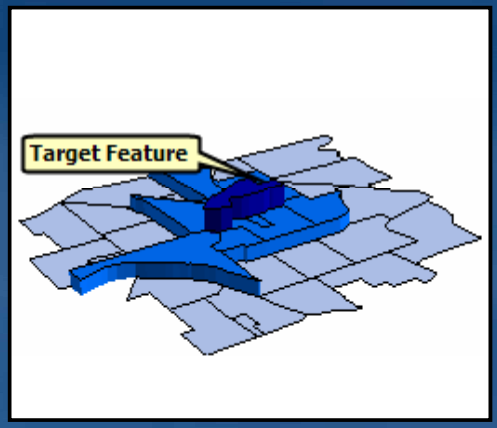
Generate Network Spatial Weights
Generate Spatial Weights Matrix



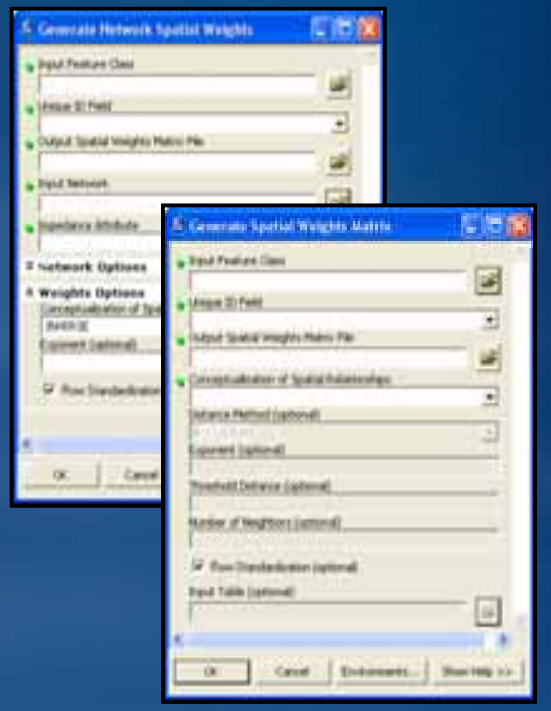
Spatial relationships based on Road Network Distances.



A and B are close in Euclidean space but far in network space due to the barrier.



Spatial relationships based on polygon contiguity.



- Improves tool performance, particularly with large data sets.
- Models spatial relationships using new options: K nearest neighbors, Queen's case polygon contiguity, or Delaunay Triangulation.
- Includes the option to force at least *n* neighbors for all features.

Pocket Man Analysis

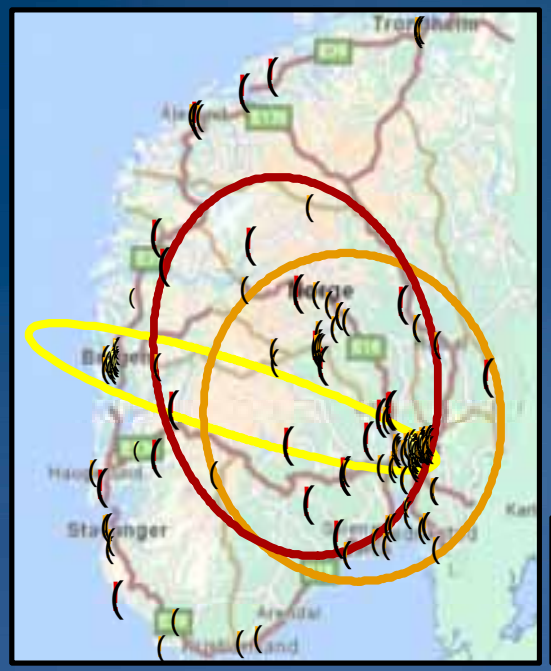
- Potentially over 300 cases.
- Data from 1990 to 2006: 156 cases.
- Suspect apprehended in January, 2008.

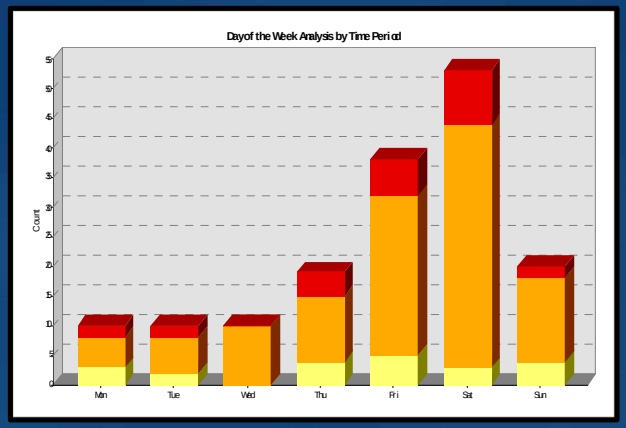
Animation

Distinct spatial regimes

- Bergen
- Oslo
- Around Norway

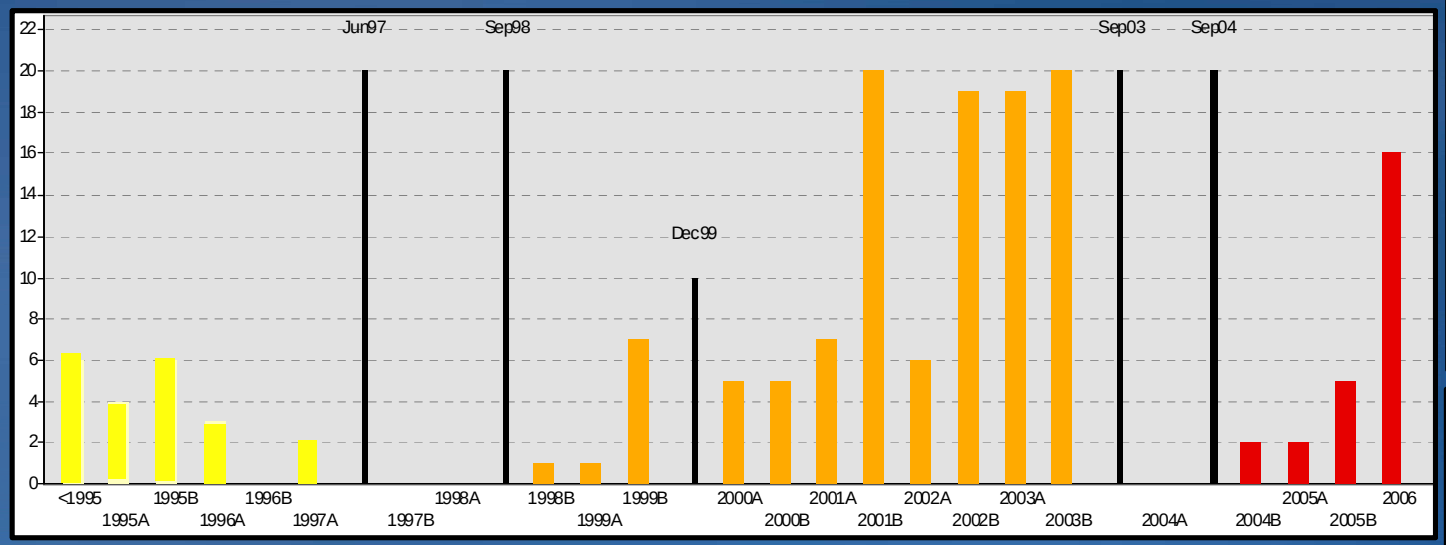
Temporal patterns





DOW analysis shows most incidents occur on Friday or Saturday. Weekend incidents are more common outside of Bergen and Oslo.

Clustering is strongest prior to 1977 and the range of incidents is narrow.



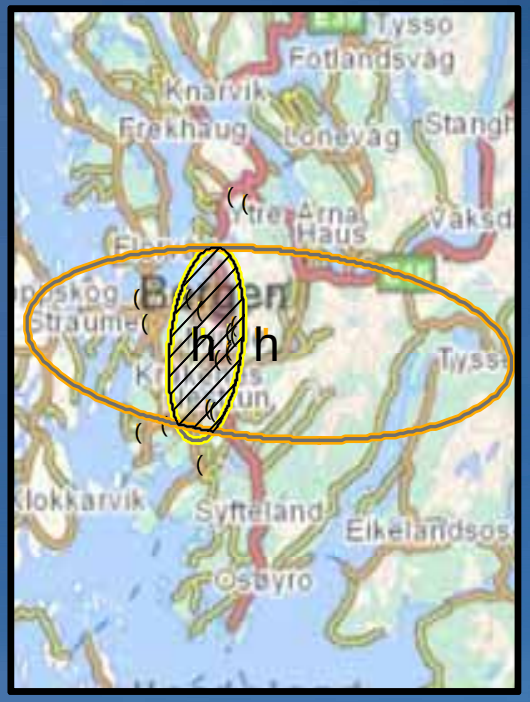
Incidents become increasingly random.

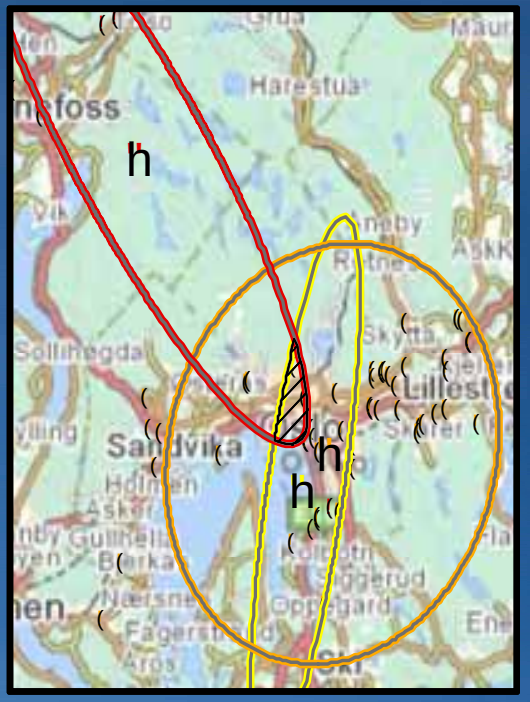
- **Incidents over time indicate 3 distinct time periods.**
- **Most of the incidents prior to Dec 99 are in Bergen.**

Space/Time patterns

Hot Spot Analysis

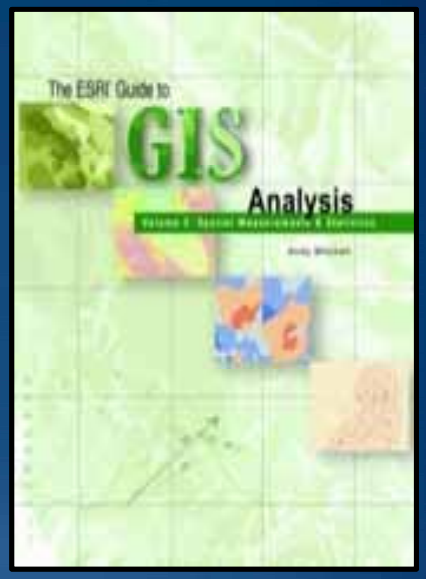
Mean Center & Standard Deviational Ellipse Analyses





Tips for Navigating Online Resources

Resources for learning more…



- Hot Spot and Regression Analysis Tutorials: http://resources.esri.com/geoprocessing/
- Instructor-Led ESRI Training: Performing Analysis with ArcGIS Desktop
- Virtual campus free web seminars: http://campus.esri.com/
- 911 emergency call analysis demo: http://www.esri.com/software/arcgis/arcinfo/about/demos.html
- Articles (keyword search: “Spatial Statistics”): http://www.esri.com/news/arcuser/0405/ss_crimestats1of2.html
- The ESRI Guide to GIS Analysis, Volume 2 by Andy Mitchell
- Online help

QUESTIONS?

LRosenshein@ESRI.com