Spatial Statistics for Public Health and Safety

Lauren M. Scott, PhD

Nathan Warmerdam, MSc

ESRI

Geoprocessing, Spatial Statistics

Spatial Data Analysis

- Traditional GIS Analyses
  - Spatial/temporal queries
  - Buffer analysis
  - Overlay analysis
- Spatial Statistics Toolbox
  - Identify, analyze and respond to incident hot/cold spots
  - Summarize spatial pattern over time
  - Identify anomalous spatial patterns

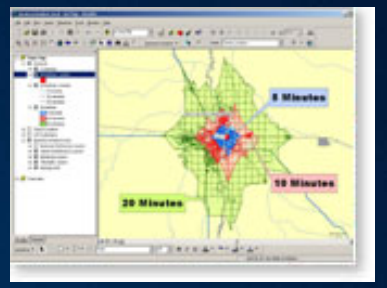
Objectives

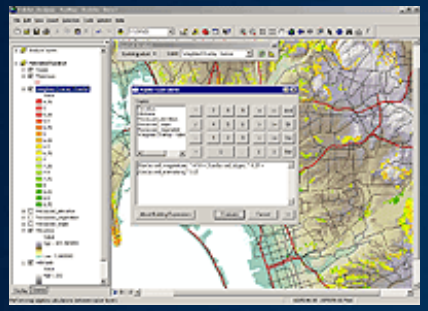
- Introduce core spatial statistics tools available in ArcGIS 9
- Provide examples for how these tools might be used in Public Safety and Public Health
- Be sure you know where to go for additional information

Presentation

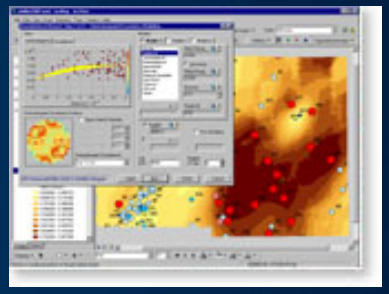
- Basics: What are spatial statistics and why do we care?
  - Maps can lie
- Overview of Spatial Statistics Toolbox
  - Measuring Geographic Distributions
  - Analyzing Spatial Patterns
  - Mapping Clusters and Outliers
    - 911 Emergency Call Analysis Demo
    - SAS/ArcGIS Mortality Data Analysis
- Resources

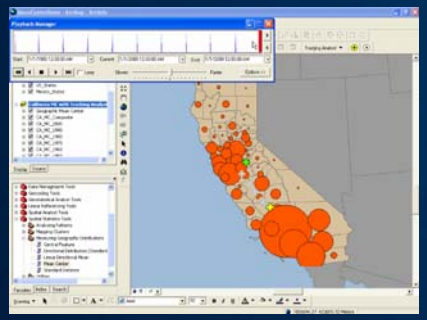
ESRI software and Statistics





- Business Analyst
- Spatial Analyst
- Geostatistical Analyst
- SAS Bridge

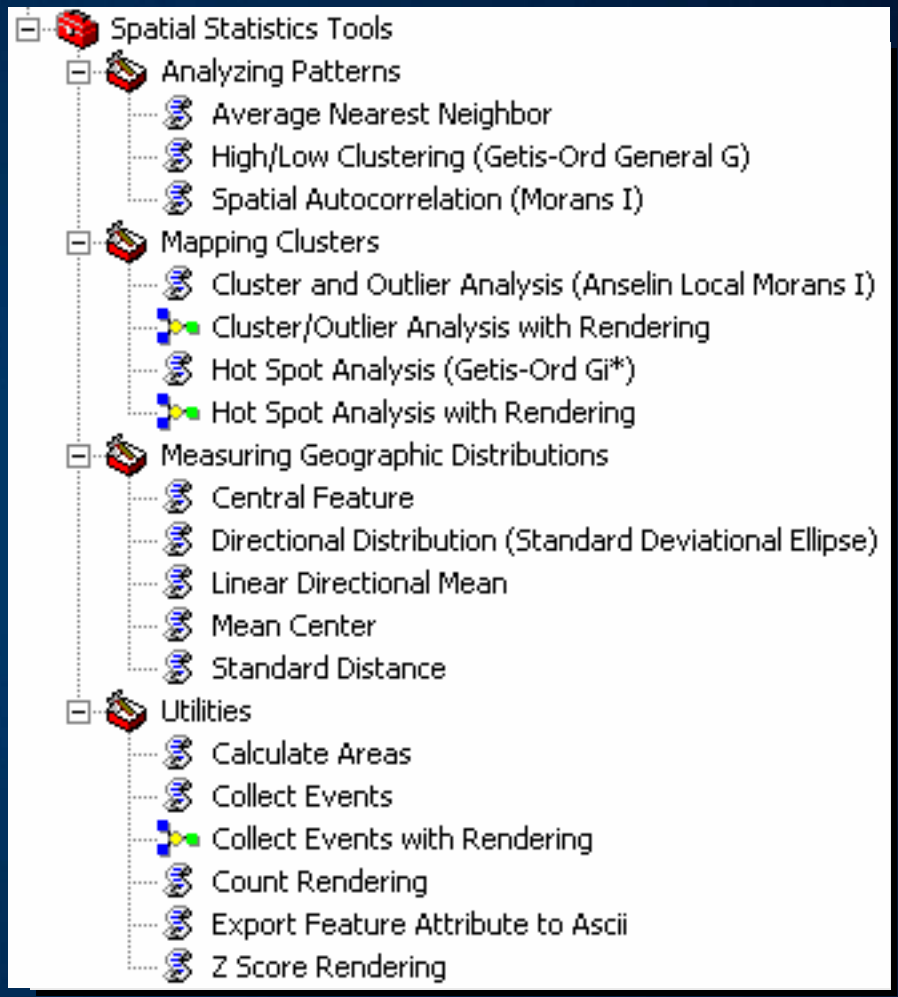




- **Spatial Statistics Tools**

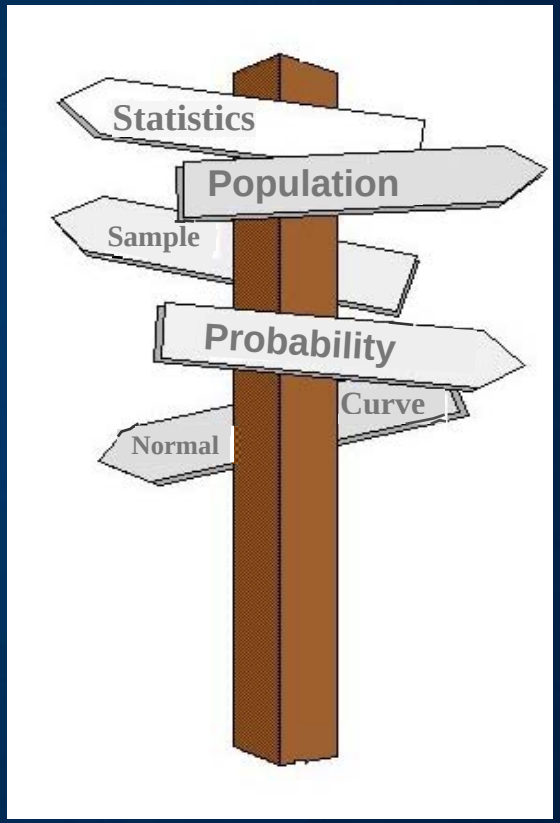
Toolsets and tools

- Core functionality with ArcGIS 9 (not an extension)
- Available with all ArcGIS licenses
- Source code provided
- Use ArcGIS 9.1 with Service Pack 1 or 2
- *The ESRI Guide to GIS Analysis, Volume 2*

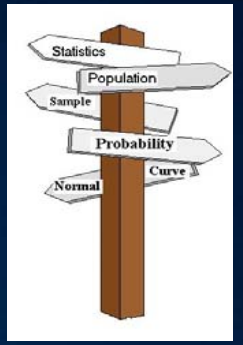


What are spatial statistics?

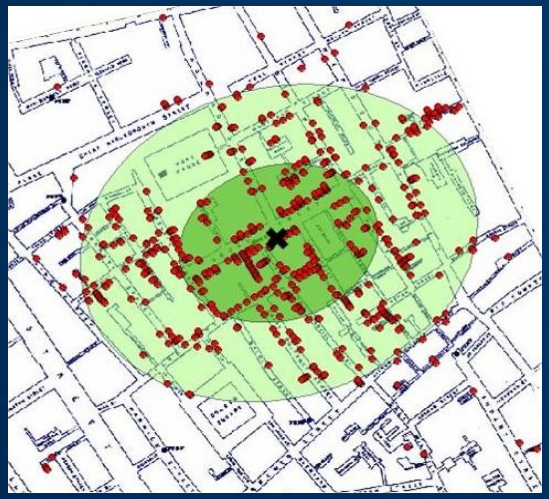
- A measure of what's going on spatially
  - Not the same as traditional (a-spatial) statistics



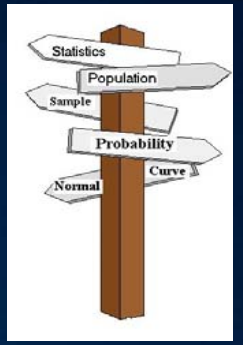
What are spatial statistics?



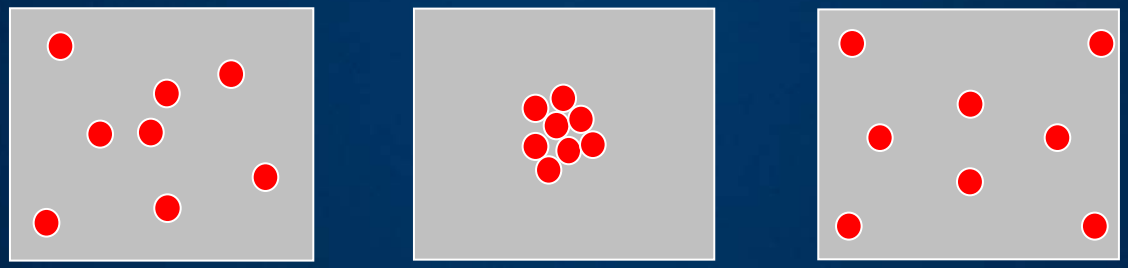
- A measure of what's going on spatially
  - Not the same as a-spatial stats
- Two categories of spatial measurements
  - 1) Identifying characteristics of a distribution



What are spatial statistics?



- A measure of what's going on spatially
  - Not the same as a-spatial stats
- Two categories of spatial measurements
  - 1) Identifying characteristics of a distribution
  - 2) Quantifying geographic pattern

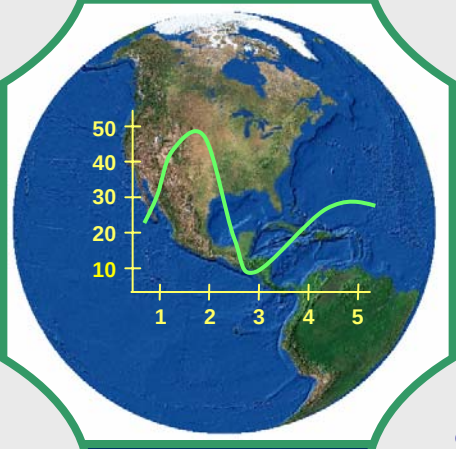


Random **Clustered** **Dispersed**

Why use spatial statistics?

Better understand behavior of geographic phenomena

Make decisions with higher level of confidence



Pinpoint causes of specific geographic patterns

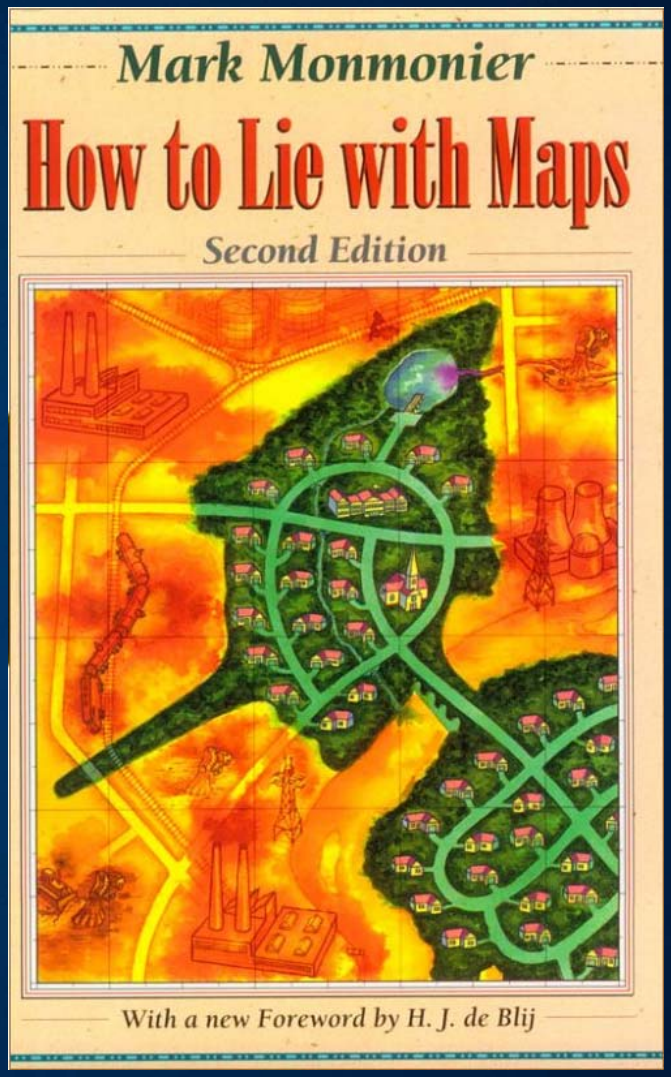
Summarize the distribution in a single number

Why Use Spatial Statistics?

Spatial Statistics help us assess:

- **Patterns**
- **Relationships**
- **Trends**

How we present/map our results (colors, classbreaks) can either enhance or obscure communication



Crime Analysis Demo

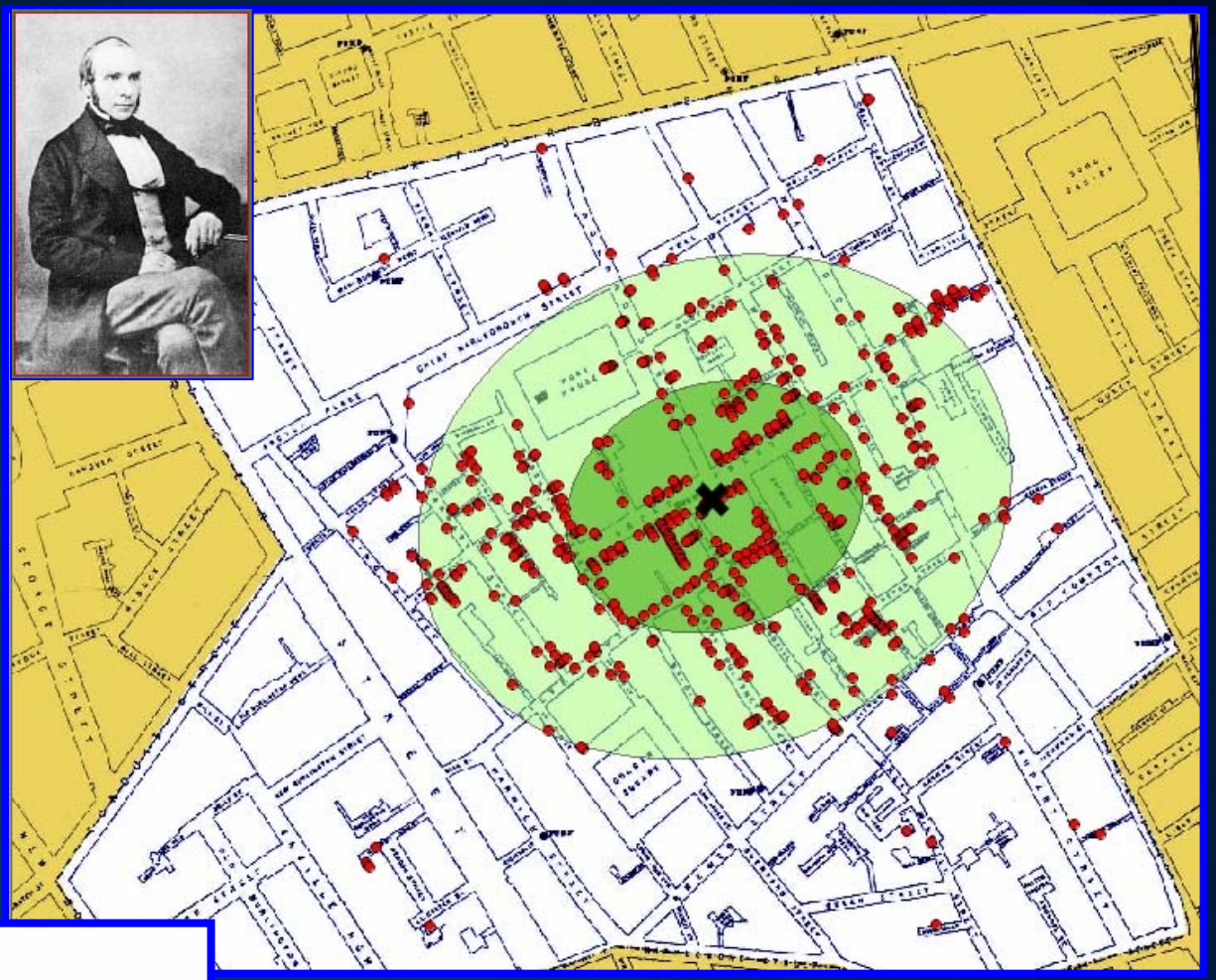
Measuring Geographic Distributions toolset

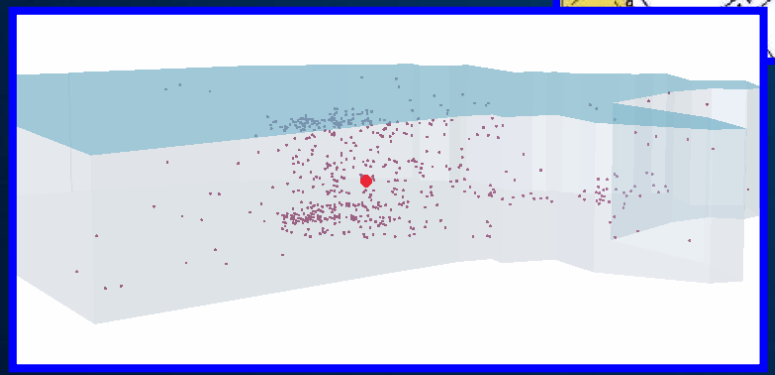
- Identify spatial characteristics of a distribution
  - Where is the center?
  - What is most central?
  - How dispersed are features around the center?
  - Are there any directional trends?
- Often used to:
  - Compare different feature categories
  - Examine changes over time

1. Central Feature
2. Mean Center
3. Linear Directional Mean
4. Standard Distance
5. Standard Deviational Ellipse

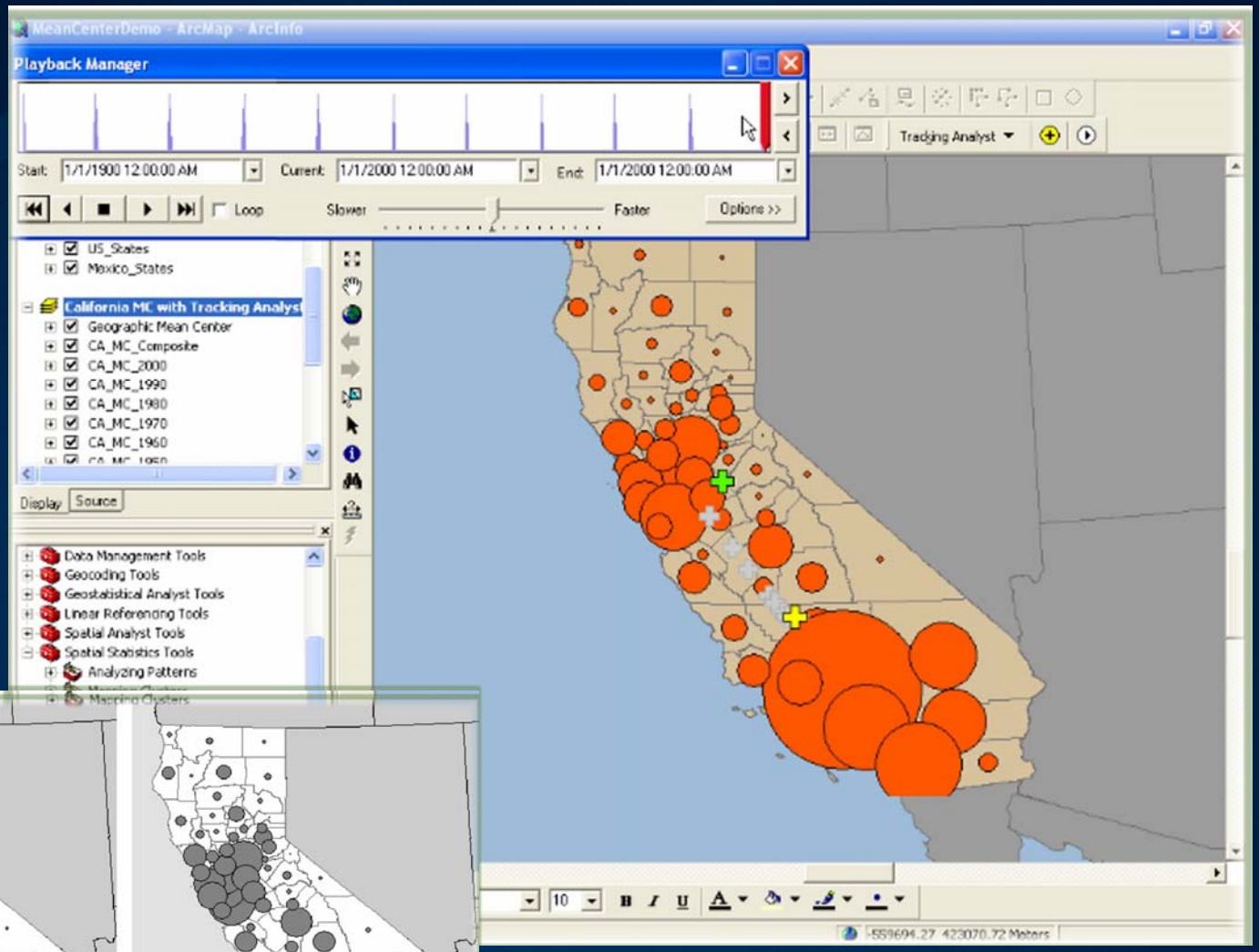
Mean Center

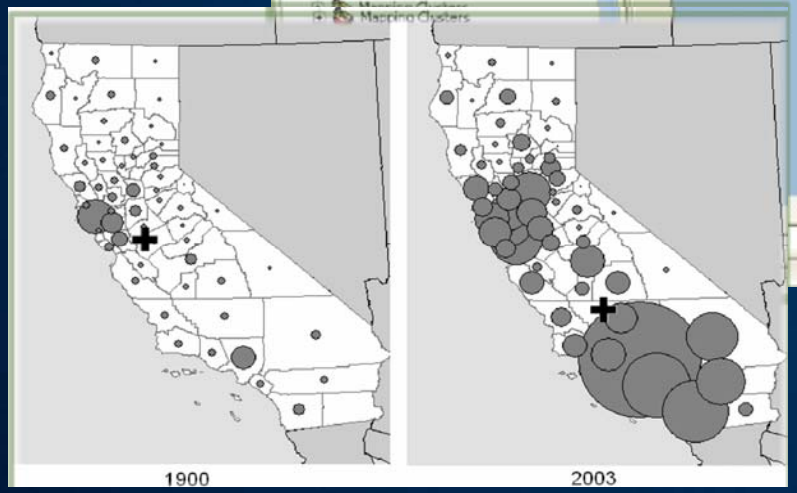
Computes the average x and y coordinate, based on all features in the study area





Spatial Statistics Tools
Mean Center





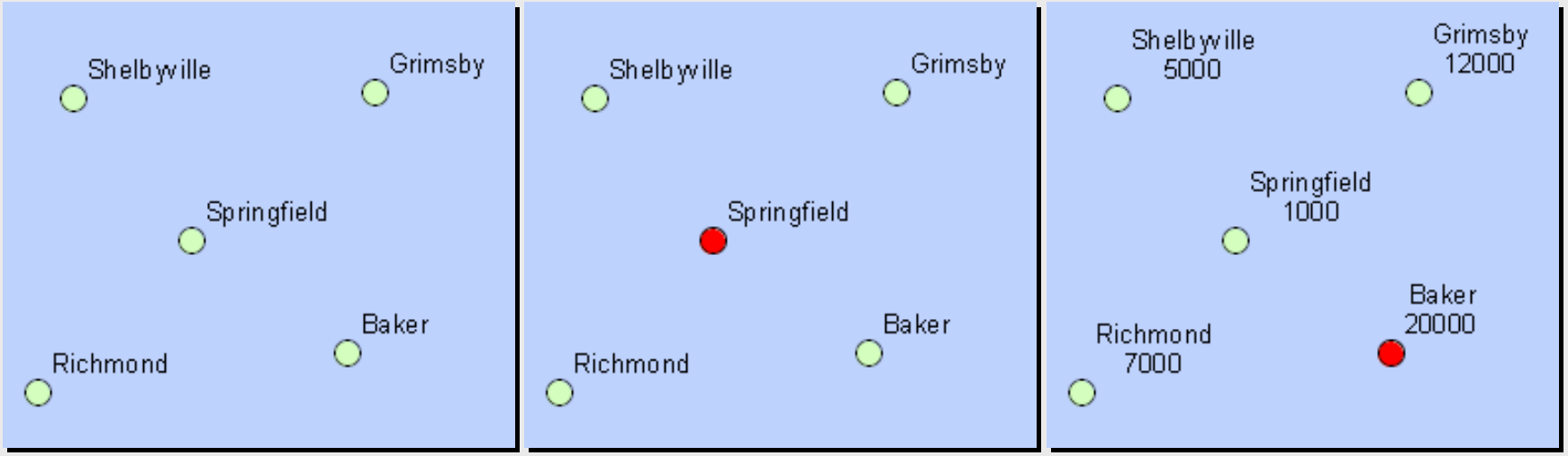
Mean Center of Population
California

Mean Center Demo

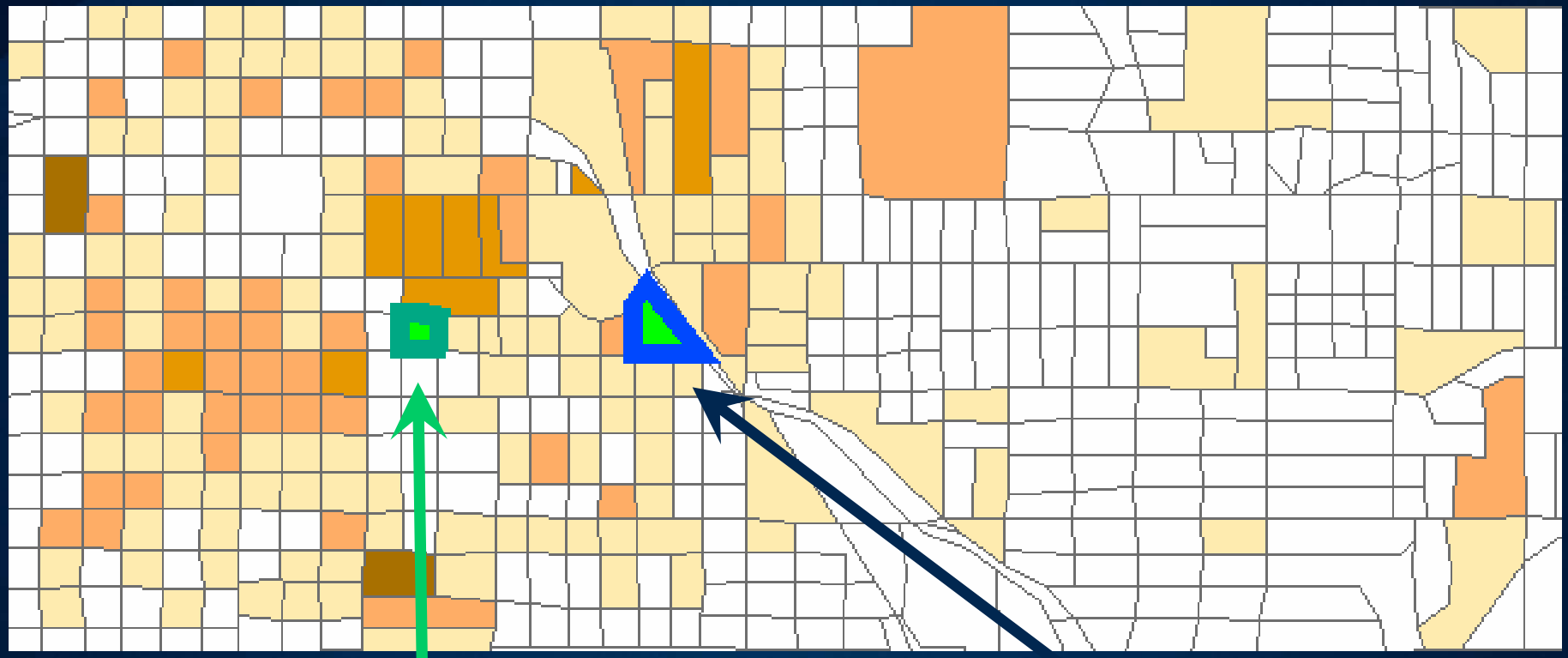
Central Feature

- Identifies the most centrally located feature
  - Feature with the lowest total distance to all other features
- Uses:
  - Finding the most accessible existing feature

What is the best location for a new service center?



Locating services for victims of domestic crime

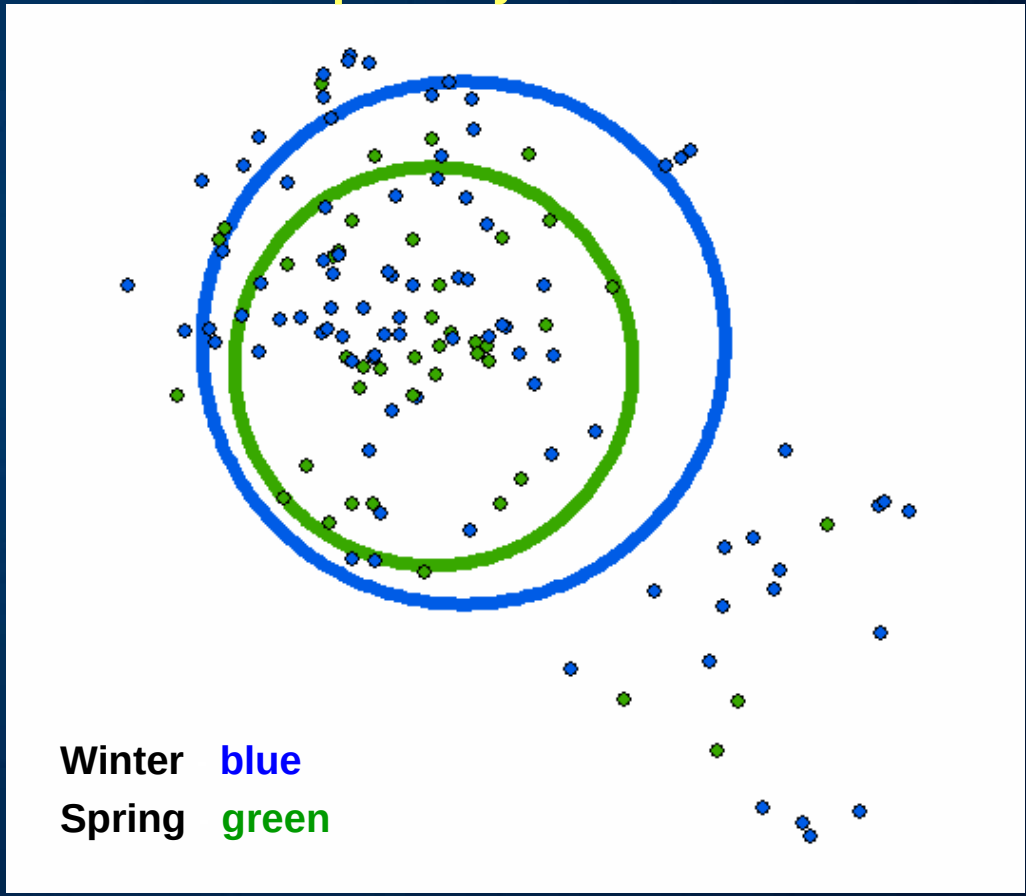


Measuring feature distribution

Standard Distance tool

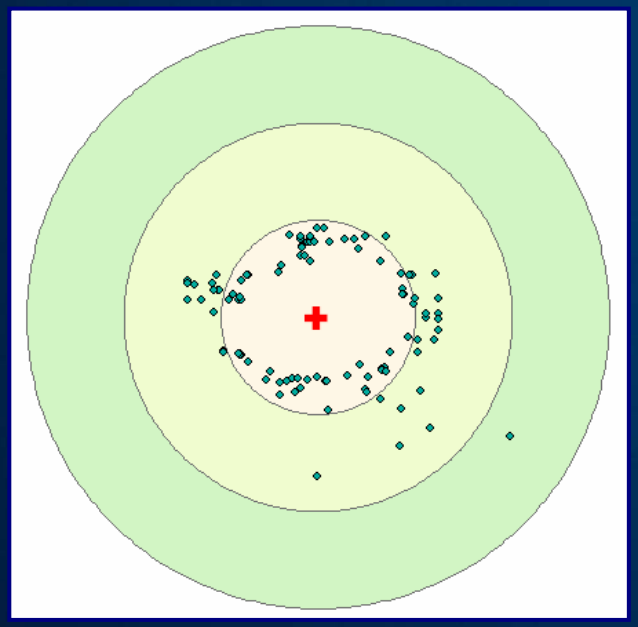
- Measures distribution of features around the mean
- Result is a summary statistic representing distance

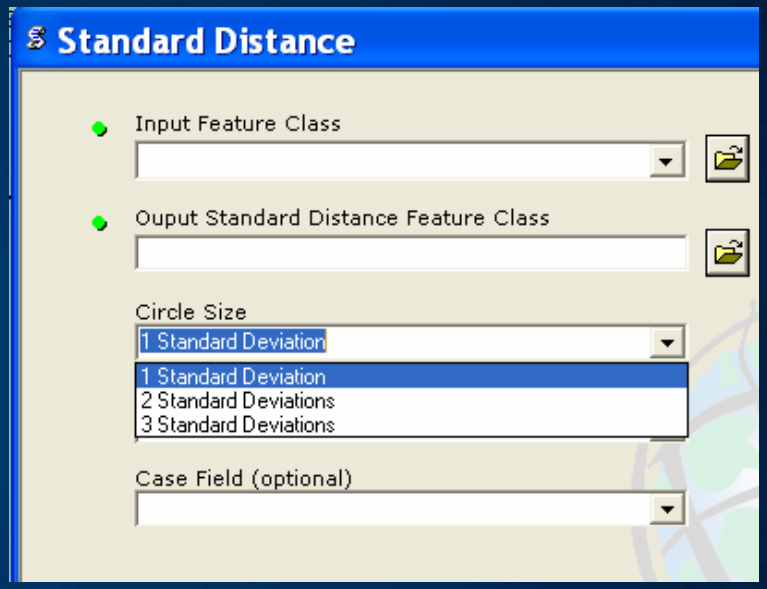
Seasonal effects of childhood respiratory diseases



Interpreting results… standard distance

- Standard distance deviation is the radius of the standard distance circle
- 1, 2, or 3 standard deviations?
- Problem distributions

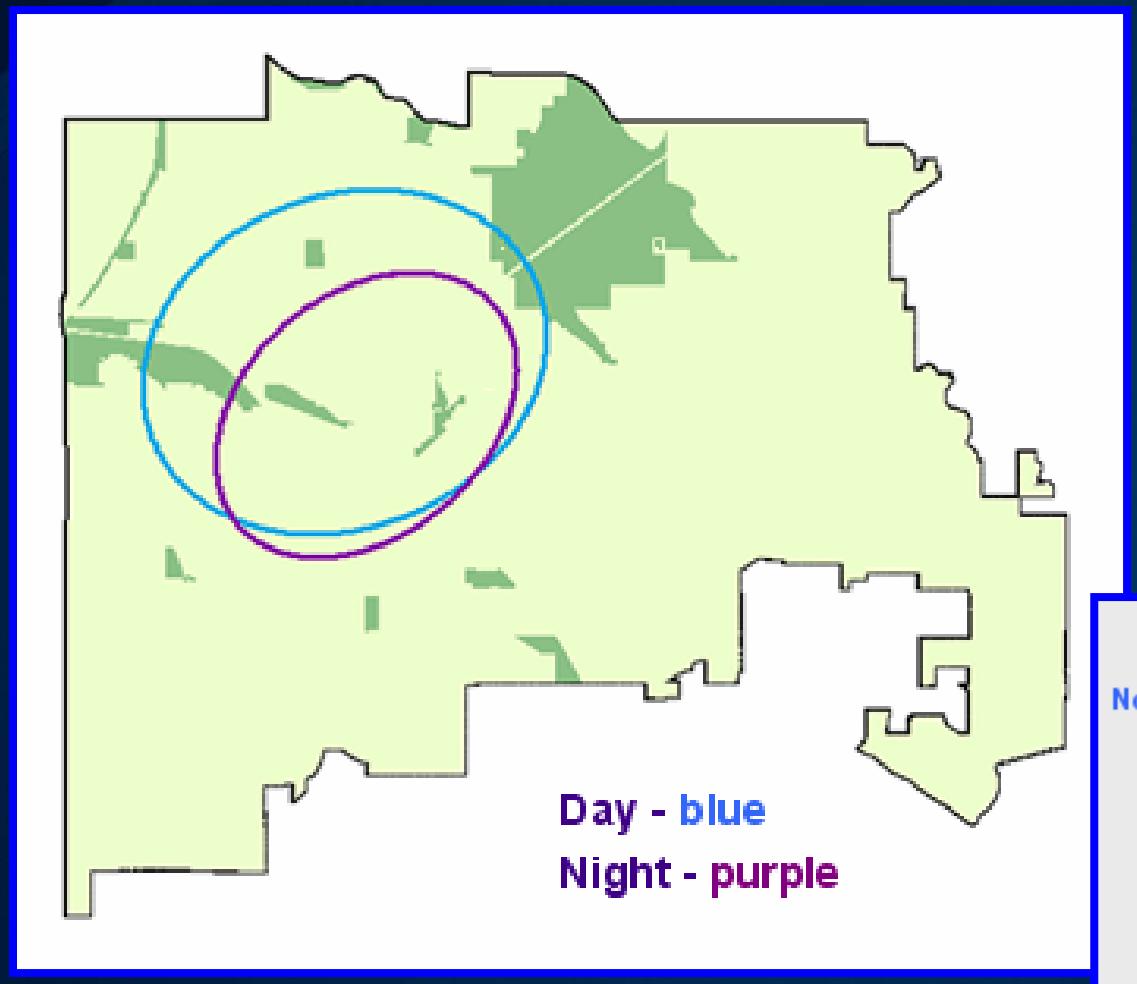




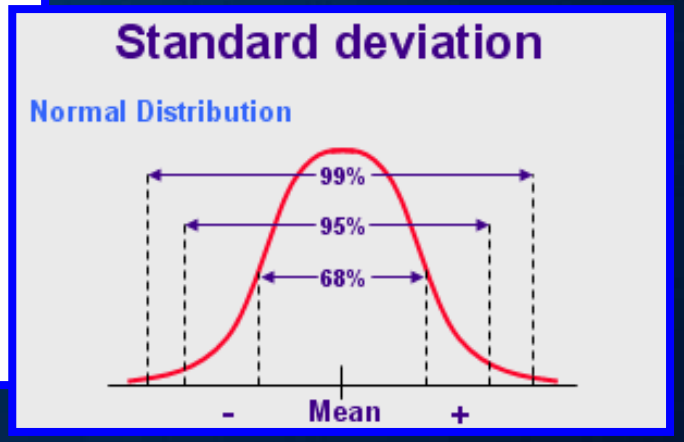
- **1 standard deviation = 68% of features**
- **2 standard deviations = 95% of features**
- **3 standard deviations = 99% of features**

Directional Distribution (Standard Deviational Ellipse)

- Abstracting spatial trends in a distribution of features

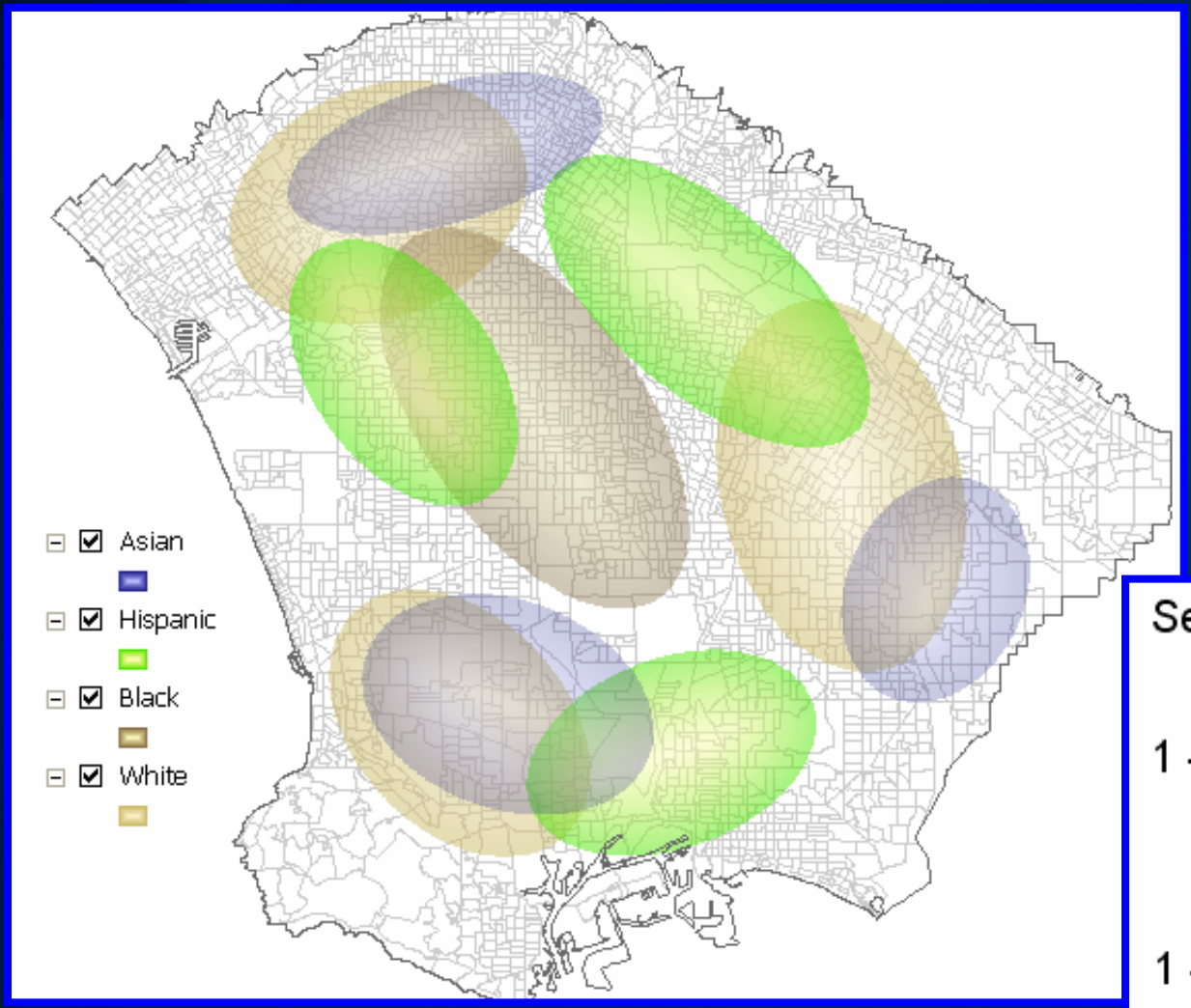


- Uses:
  - Comparing distributions
  - Examining different time periods
  - Showing compactness and orientation



Directional Distribution (Standard Deviational Ellipse)

- Abstracting spatial trends in a distribution of features

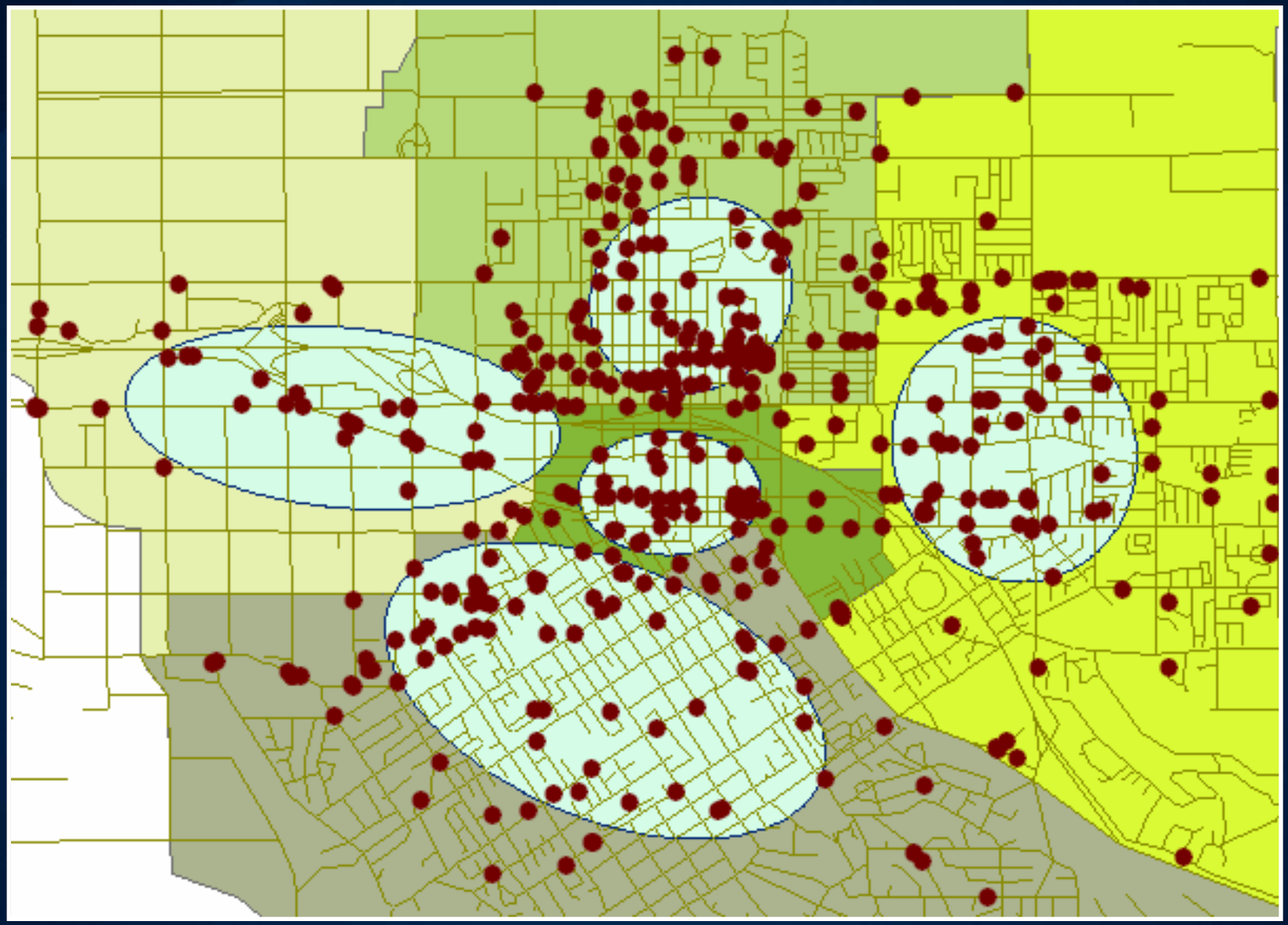


Segregation Index =

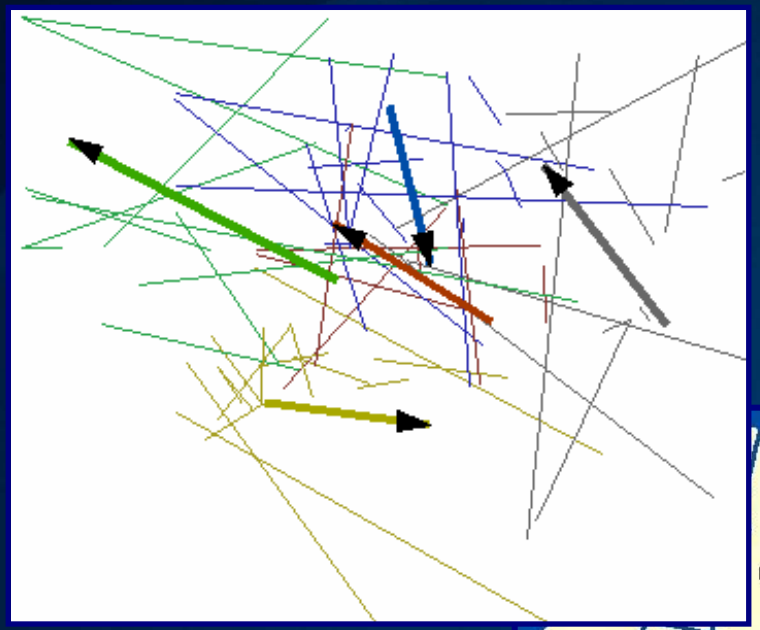
$$1 - \frac{E_1 \cap E_2 \cap E_3 \cap \ldots E_n}{E_1 \cup E_2 \cup E_3 \cup \ldots E_n} =$$

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

Another example… standard deviational ellipse



Linear Directional Mean

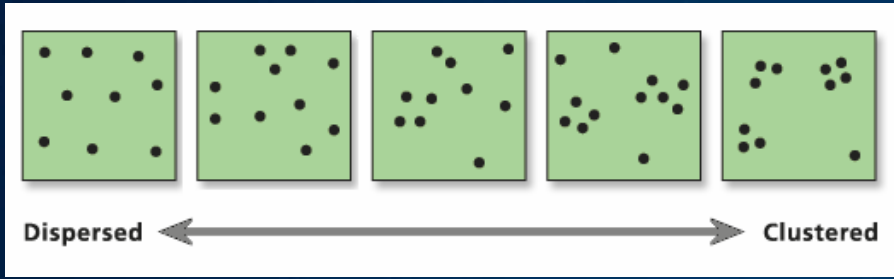


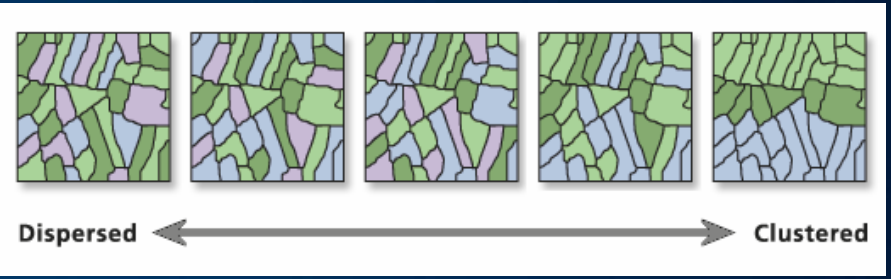
Angle of a line representing mean direction (or orientation) based on all lines in a dataset

Analyzing Patterns toolset

- Determine the extent to which features are clustered, dispersed, or random across study area

1. Average Nearest Neighbor
2. High/Low Clustering (Getis/Ord General G)



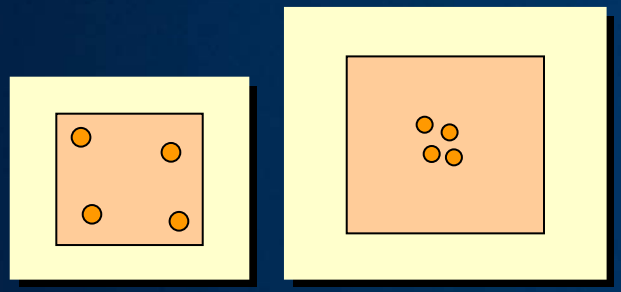


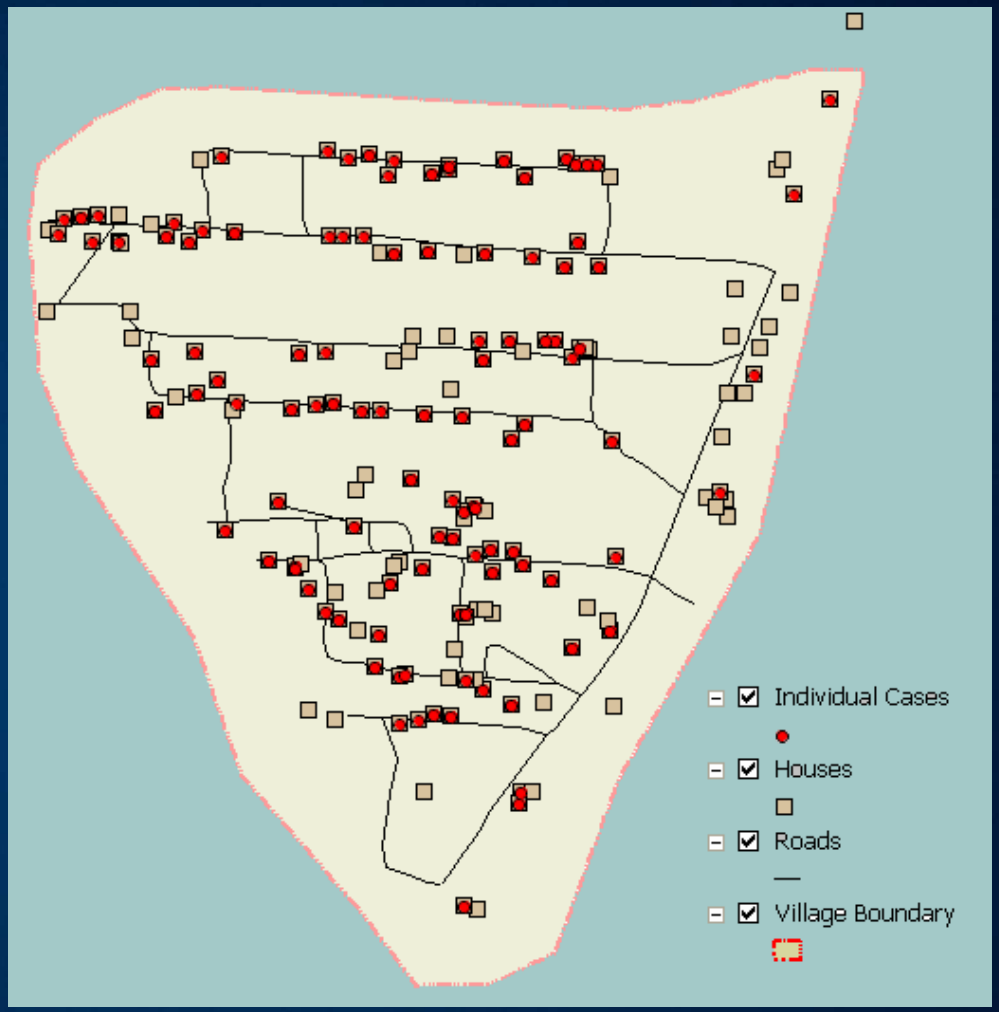
- Global analysis
- Useful for:
  - Gaining a better understanding of feature distribution
    - Degree of clustering or dispersion across study area
  - Comparing different sets of features
  - Tracking changes in patterns (over time)

3. Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
4. Spatial Autocorrelation (Global Moran's I)

Average Nearest Neighbor

- Calculates the average distance between each feature
  - Considers arrangement and distance of features
- Based on area

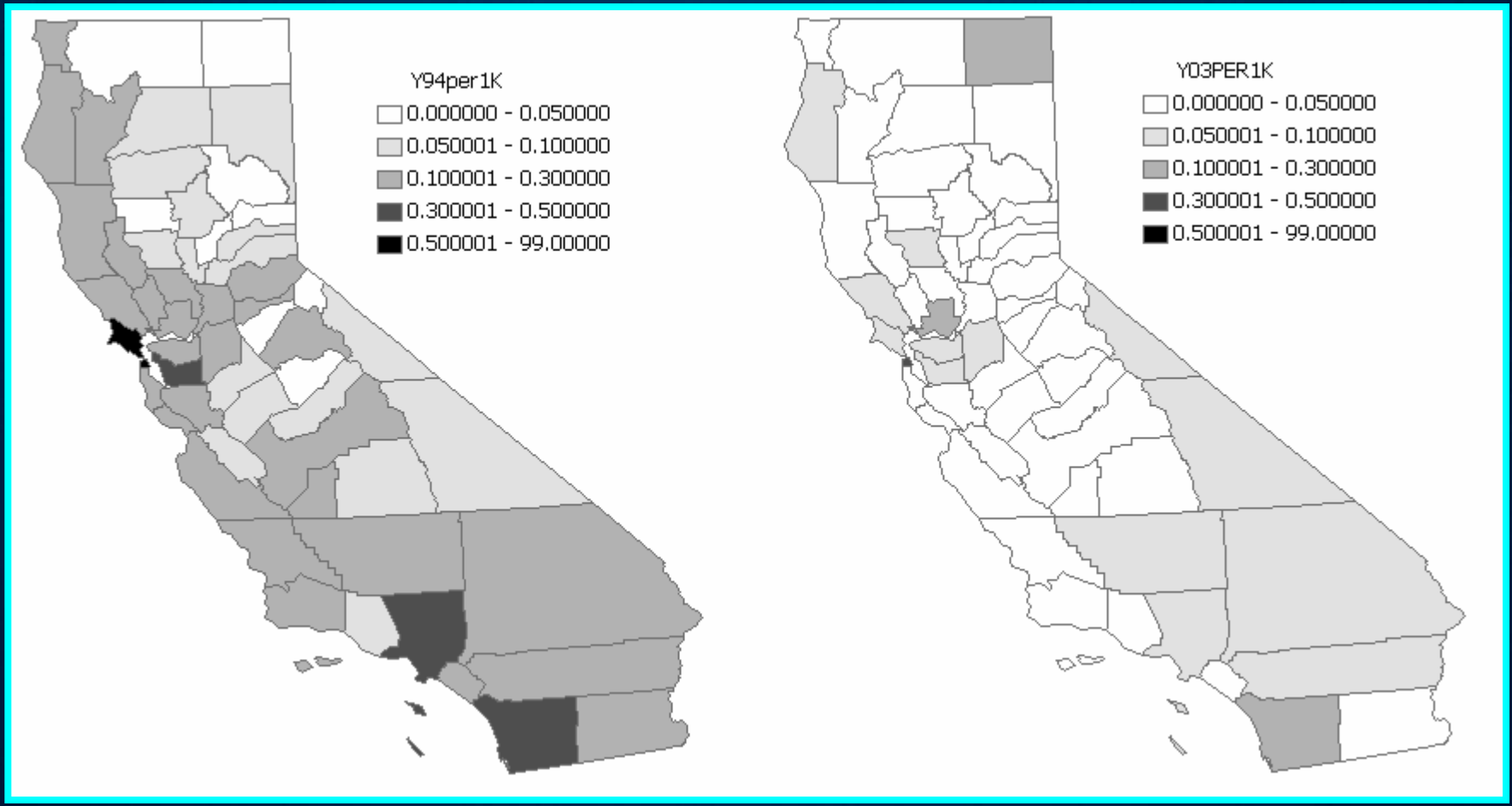




Nearest Neighbor Ratio	Pattern
Ratio = 1	Random
Ratio < 1	Clustered
Ratio > 1	Dispersed

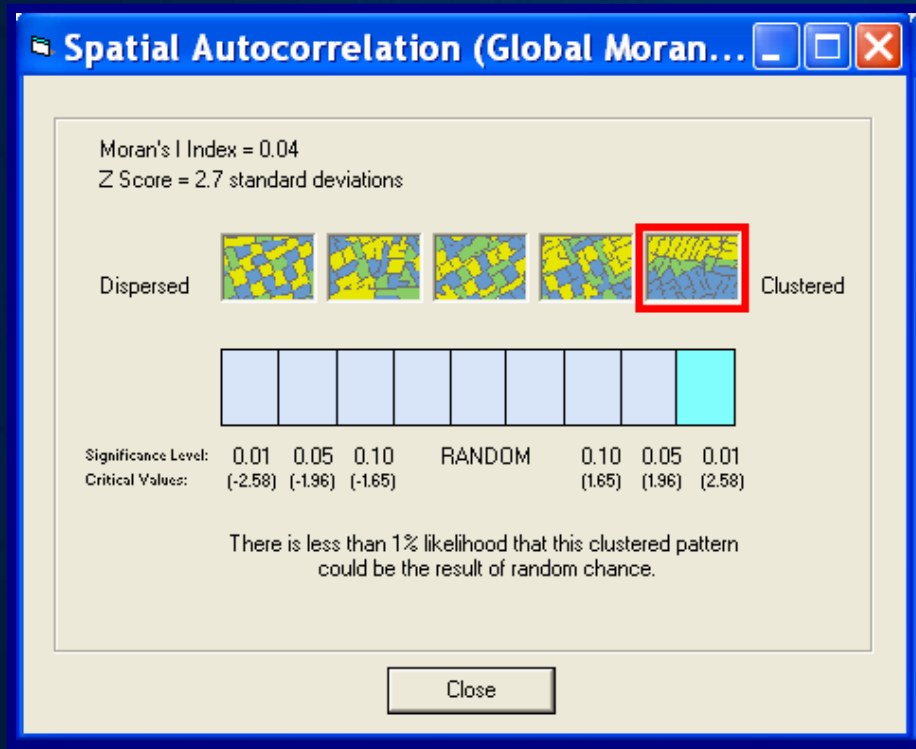
Spatial Autocorrelation (Global Moran's I)

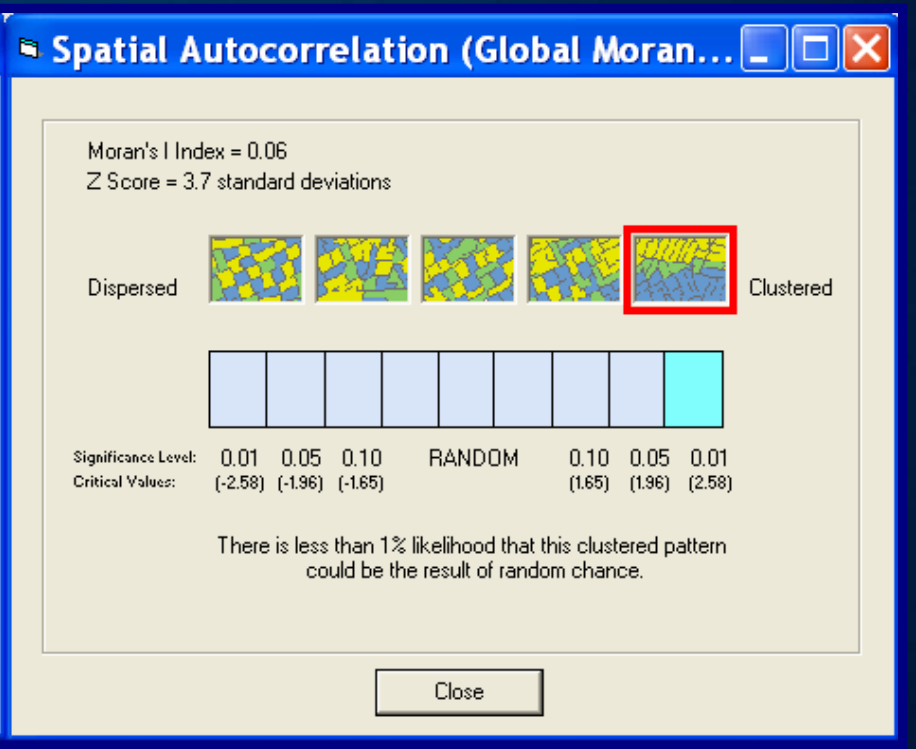
- Measures similarity of neighboring features
  - Things closer are more alike than things that are not
- Based on both feature locations and attribute values



Spatial Autocorrelation (Global Moran's I)

- Measures similarity of neighboring features
  - Things closer are more alike than things that are not
- Based on both feature locations and attribute values



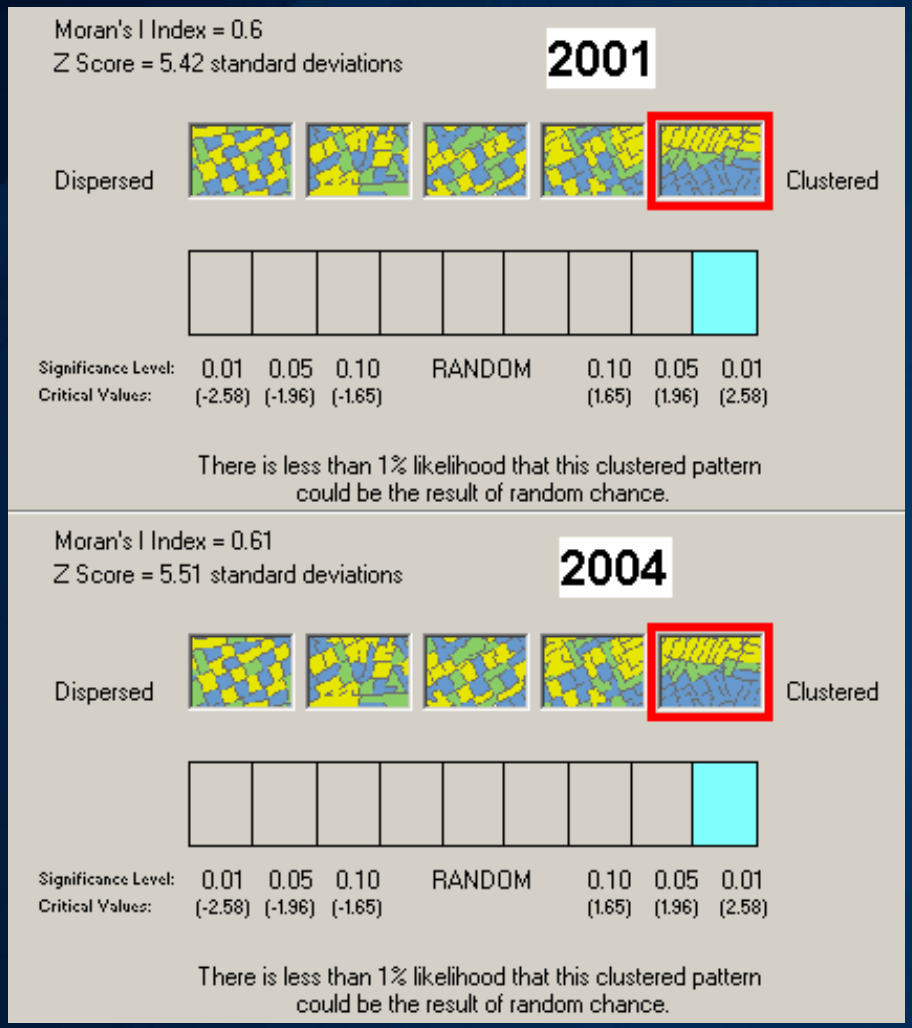


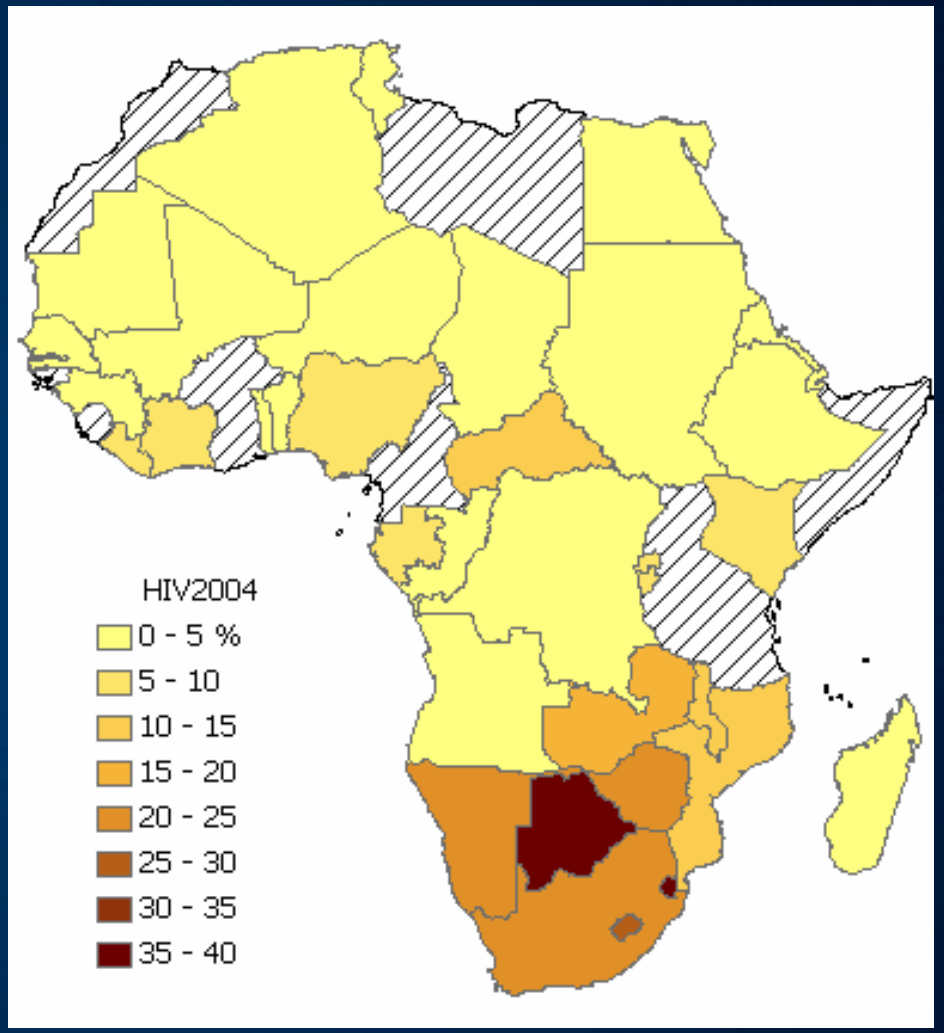
Z Score = 2.7 standard deviations

Z Score = 3.7 standard deviations

Spatial Autocorrelation (Global Moran's I)

Z Score = 5.4 standard deviations

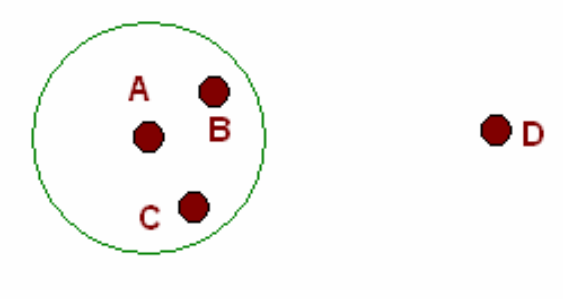




Z Score = 5.5 standard deviations

K Function

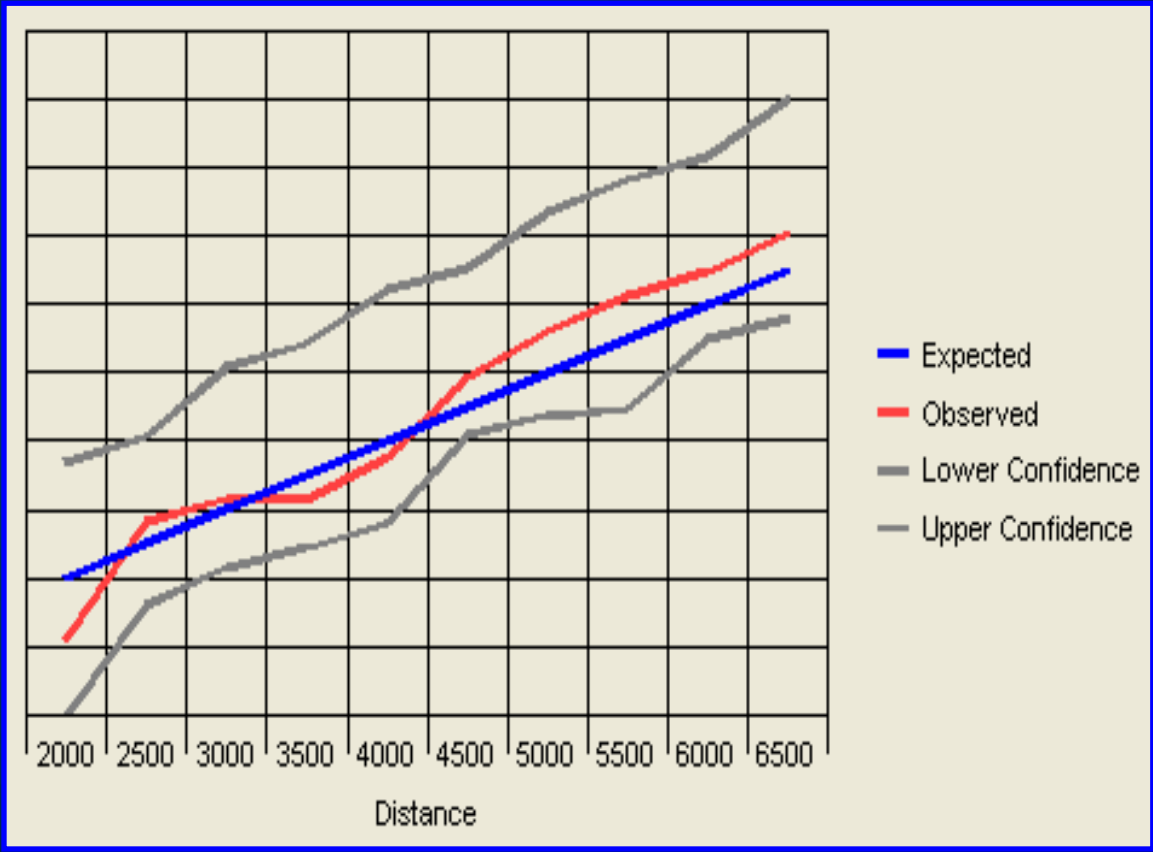
- Counts number of pairs within distance *d* of each feature



- Interpreting results

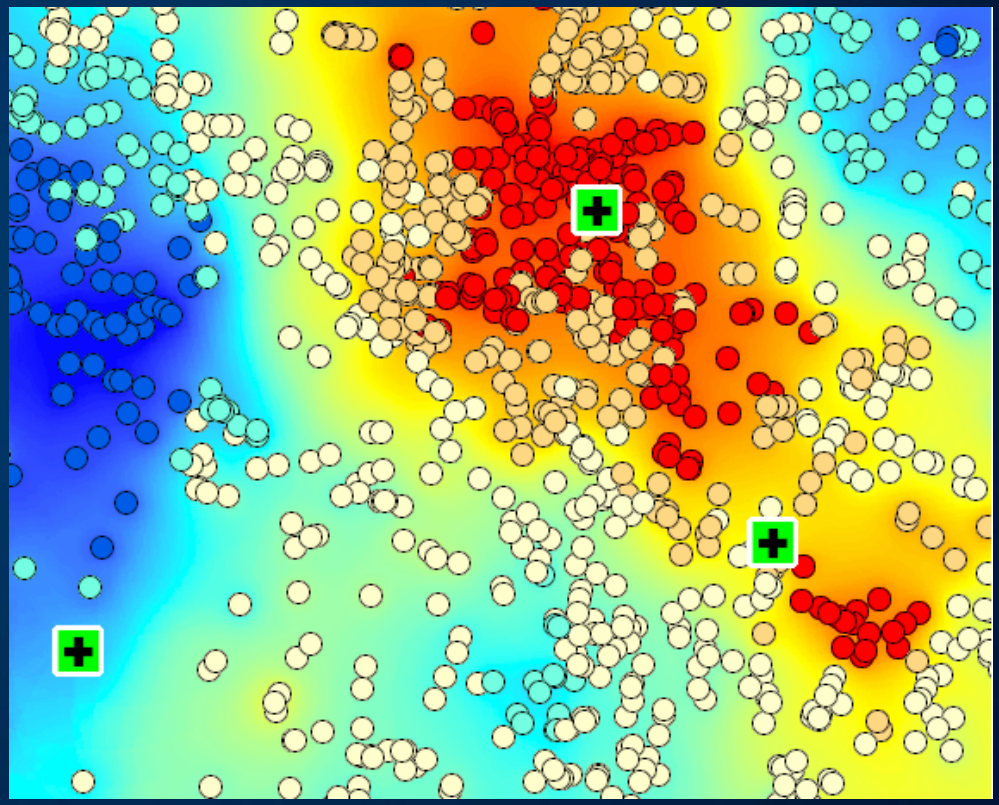
A, B
A, C
B, A
B, C
C, A
C, B

= 6 pairs for A



Mapping Clusters toolset

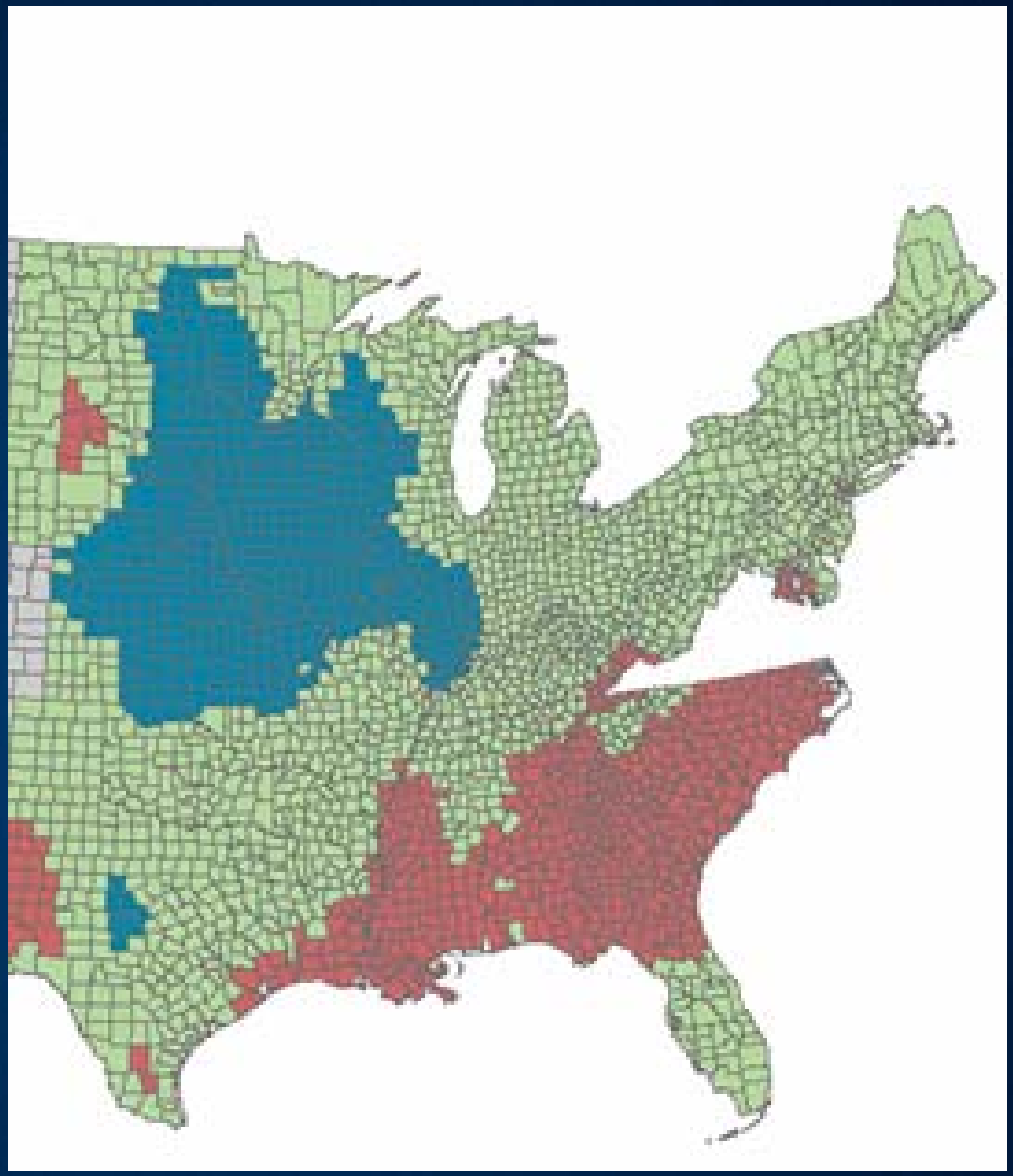
- Comparing clusters to locations of other features can help to better understand why clusters occur
- Reveal clustering and/or outliers within a study area
- Local calculations
  - Considers each feature
  - See differences across the study area
  - Explore how patterns change with changes in spatial scale



1. Hot Spot Analysis (Getis/Ord Gi*)
2. Cluster/Outlier Analysis (Anselin Local Moran's I)

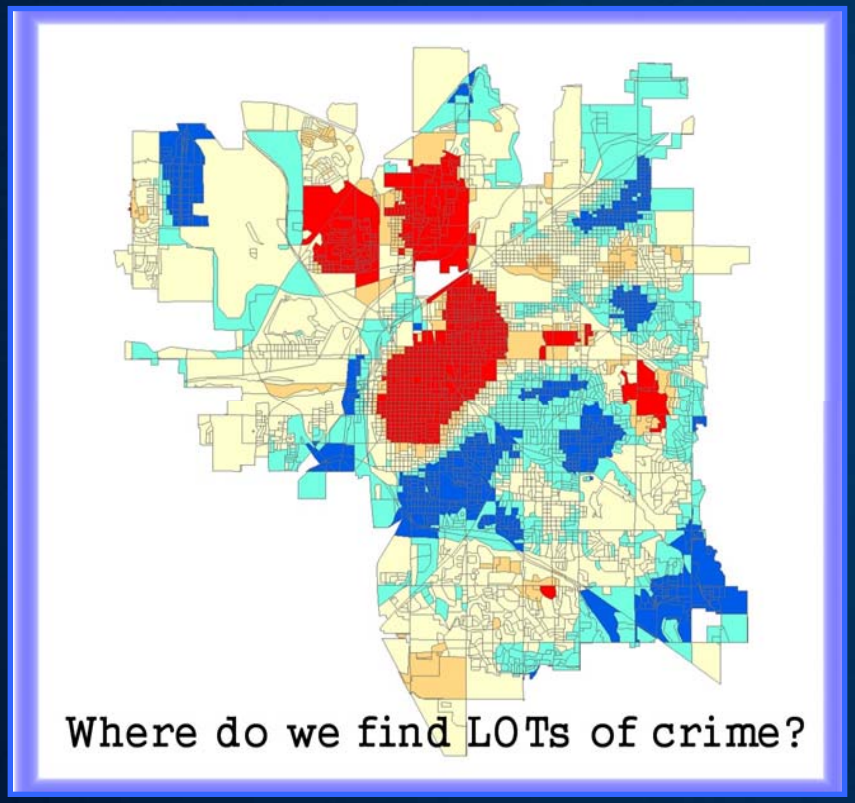
Hot spot analysis

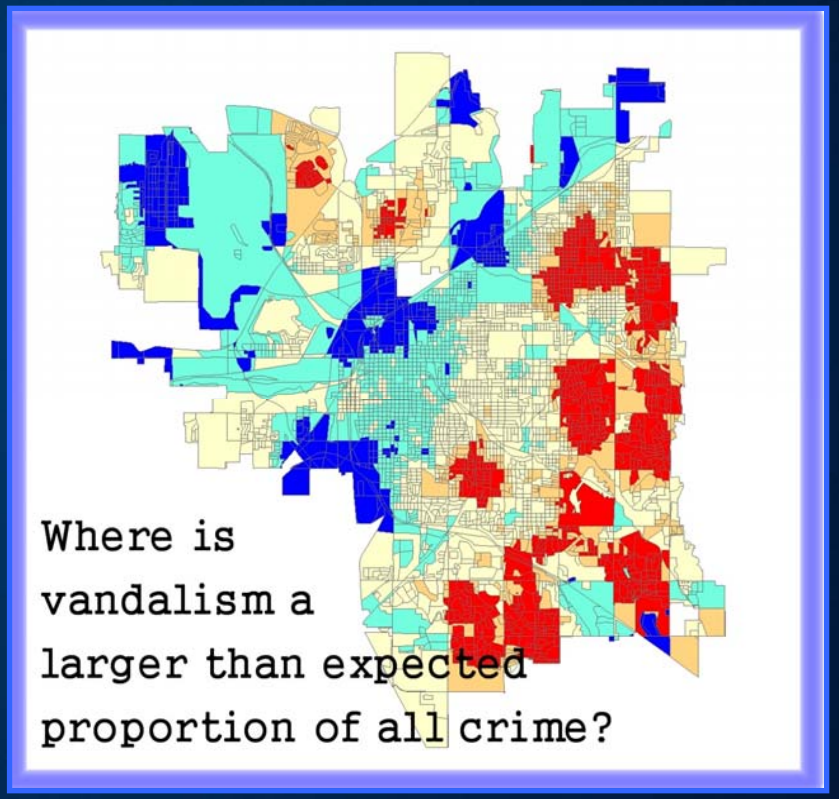
Gi8000
3.36996
-1.28478
-1.1748
-0.53801
-1.9605
0.817222
-1.20506
-2.69827
-0.17904



Hot Spot Analysis (Getis-Ord Gi*)

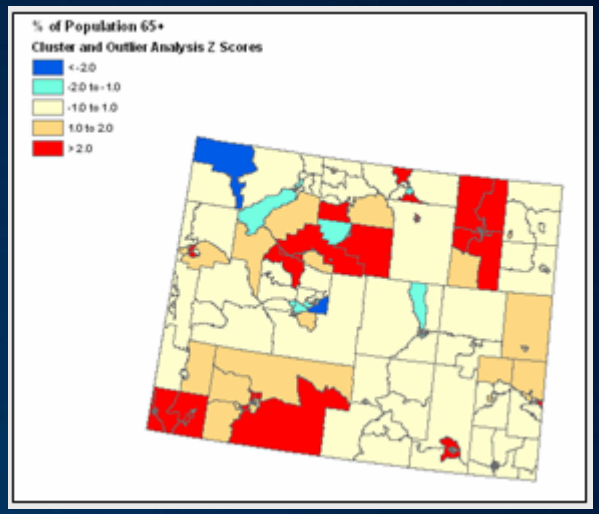
- Identifies where clusters of high values and clusters of low values exist in study area
- Calculates a z score for each feature

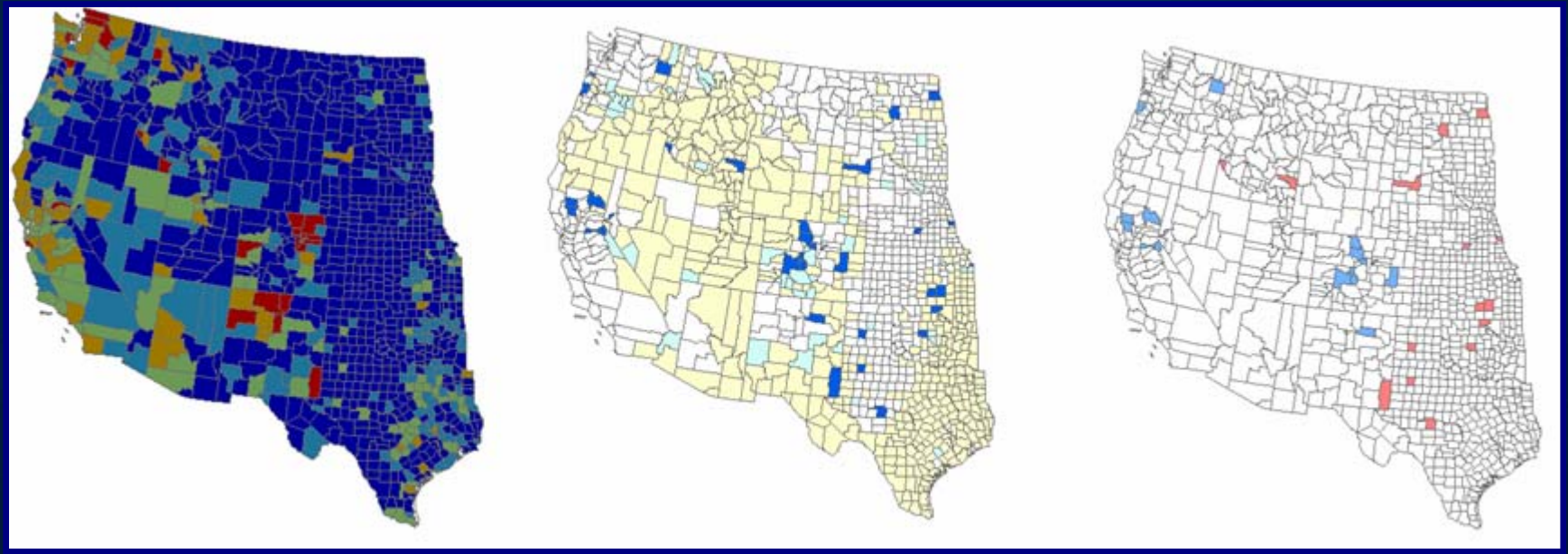




Cluster and Outlier Analysis (Anselin Local Moran's I)

- Feature location and feature attributes
- Emphasizes how feature values differ in the study area
- Helps identify outliers and errors in data
- Calculates index value and z score





911 Emergency Call Demo

Key Concepts

1. What is the weight field?
   a) Collect Events
   b) Aggregate
2. What is the scale of analysis?
   a) Sometimes you know
   b) Not too big, not too small
   c) Find the distance where the underlying spatial processes promoting clustering are strongest/most pronounced
   d) Multi-scale exploratory analysis
3. What is the question?
   a) Where do we have LOTS of <blap>?
   b) Where do we have a higher then expected *proportion* of <blap> ?

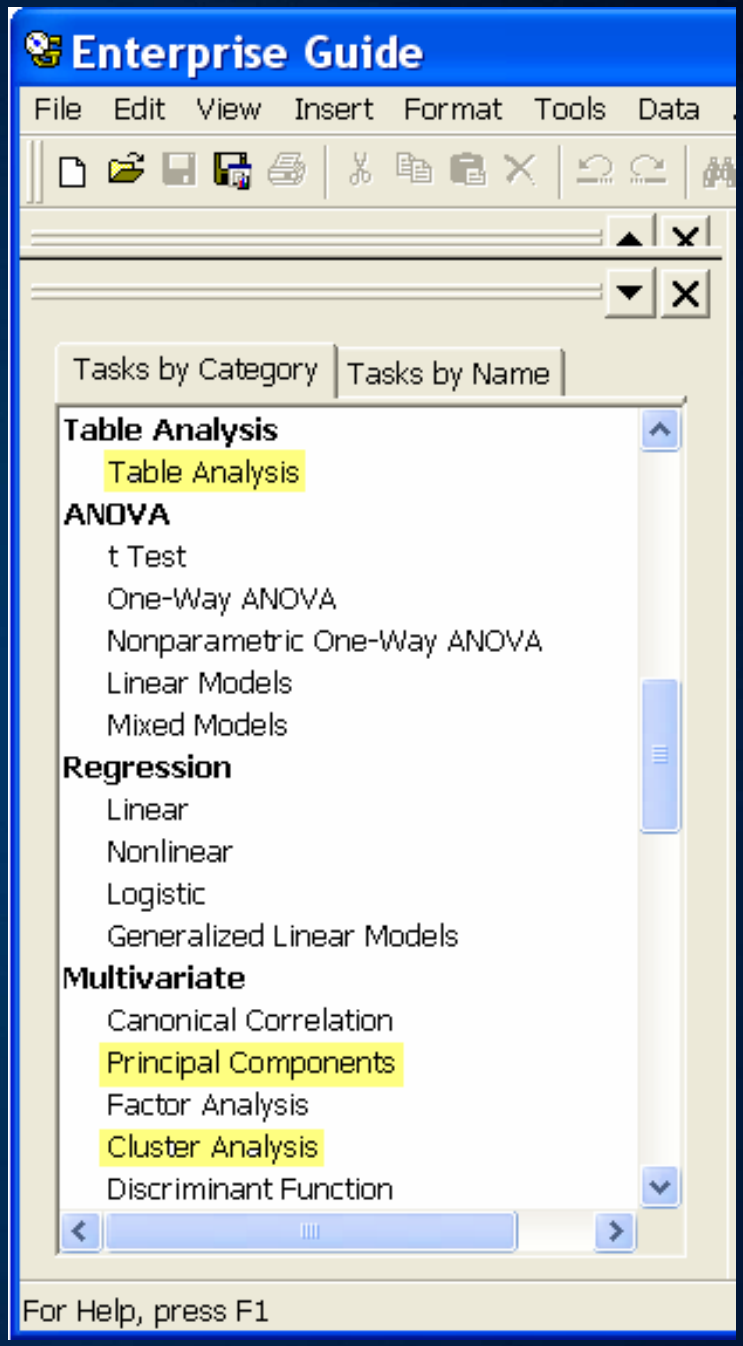
Health Demographics Analysis

Goals:

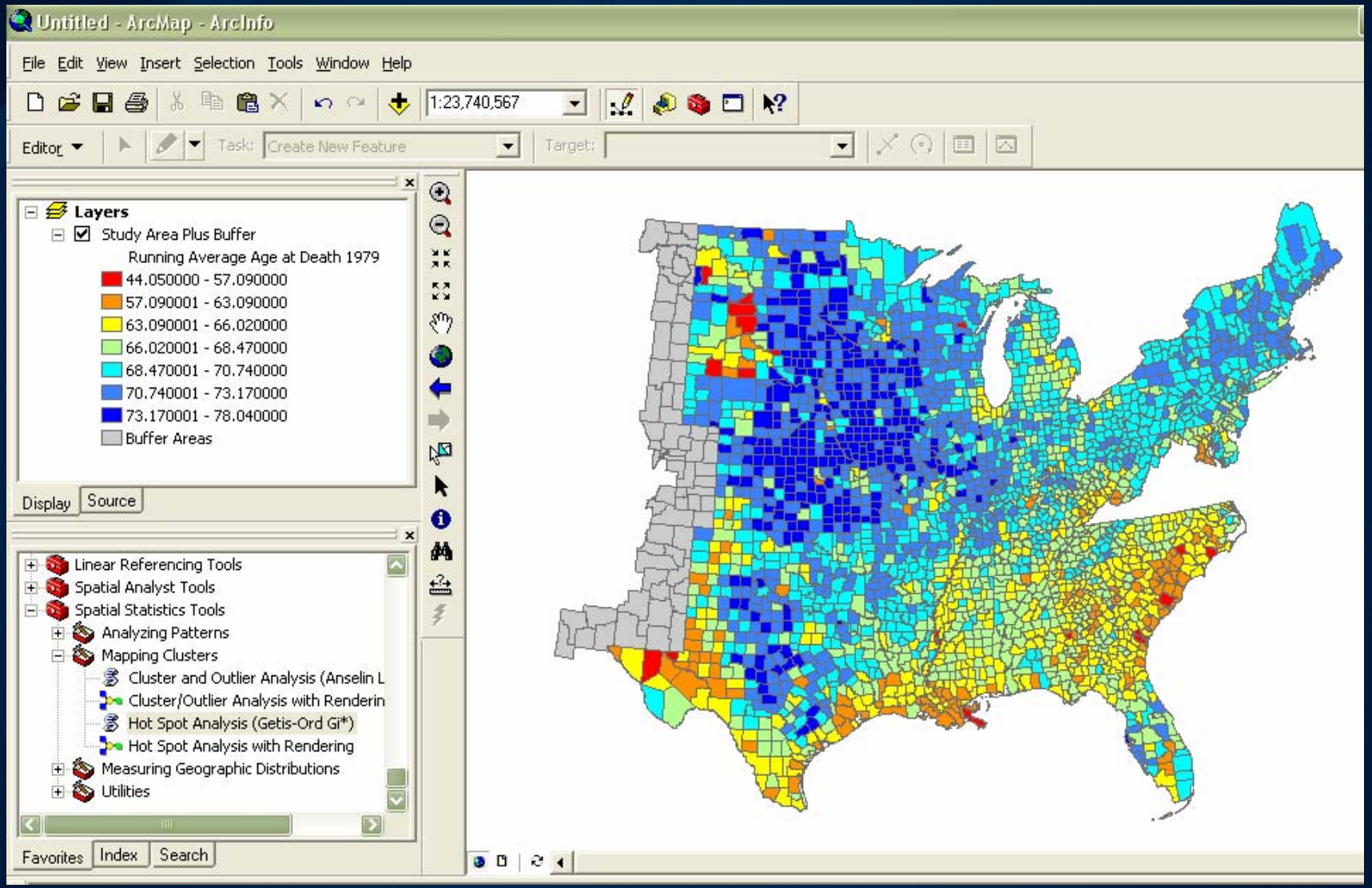
- Analyze US mortality data. Are there regions where the average age at death is persistently younger than expected?
- Explore characteristics of these persistent early death regions. Are there specific factors contributing to early death?

SAS Statistical Tools

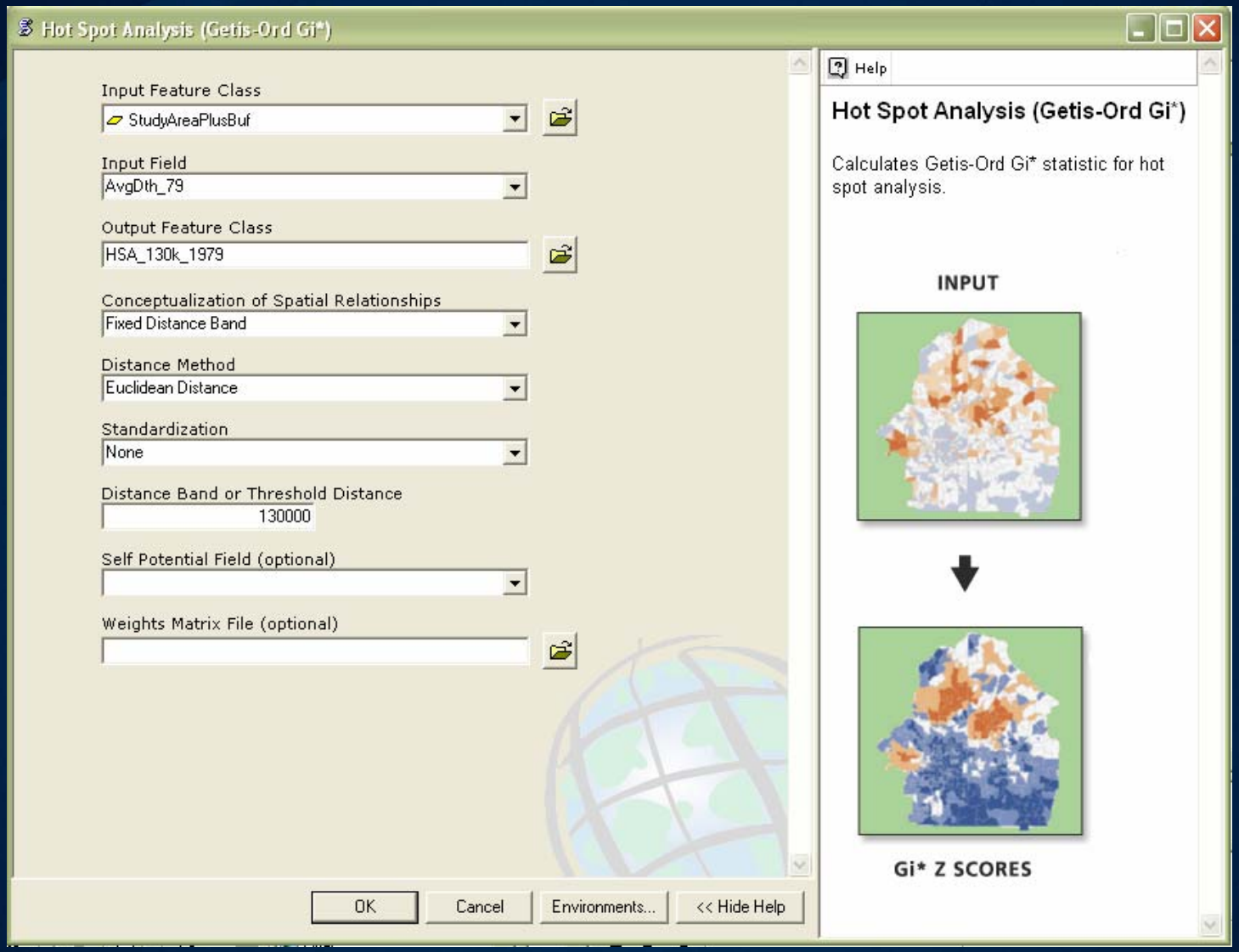
SAS provides traditional (a-spatial) statistics such as Cluster Analysis, Principal Components Analysis and Chi^2



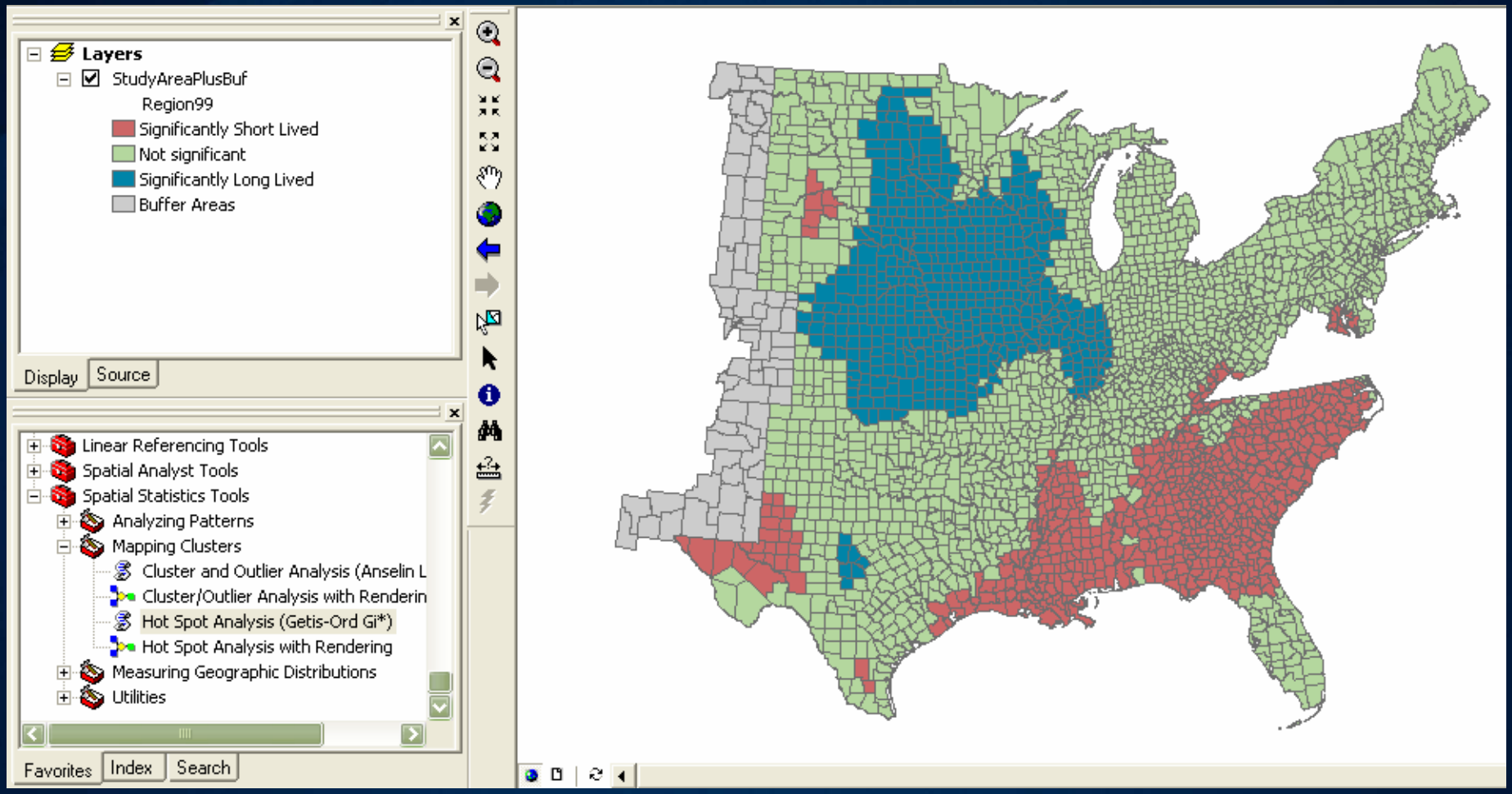
Average Age at Death in 1979



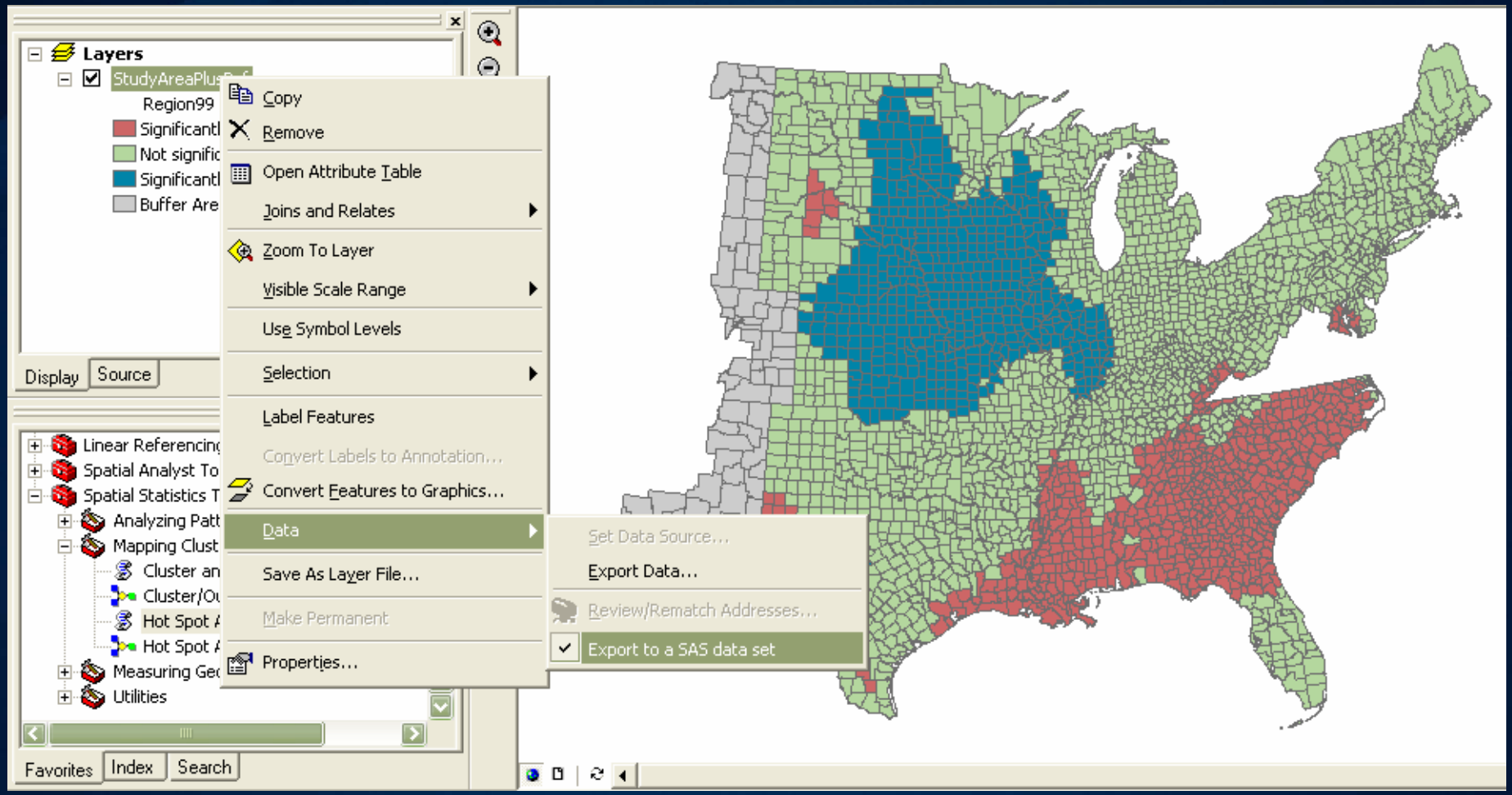
Identify Hot Spots



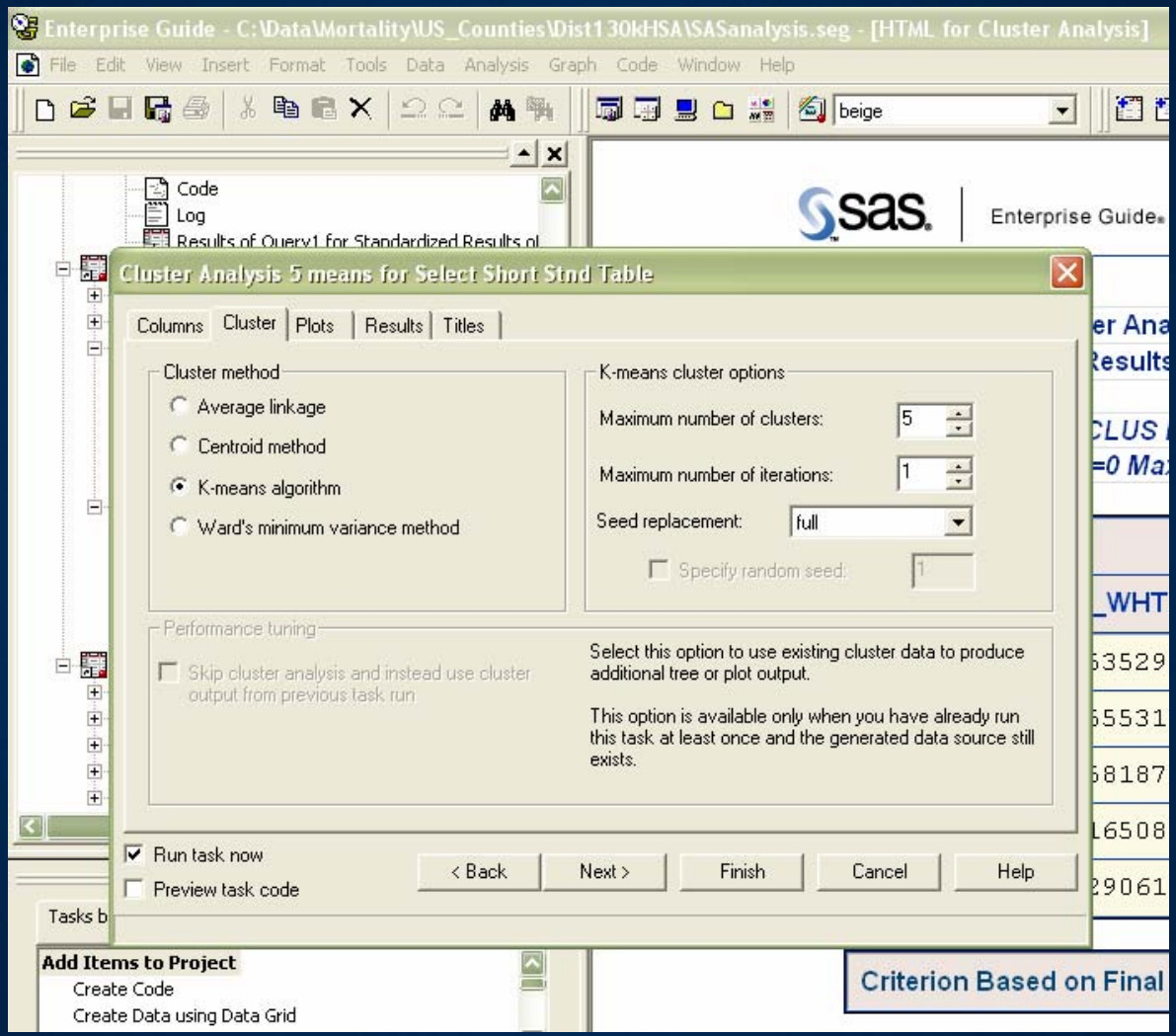
Results from ArcGIS Hot Spot Analysis



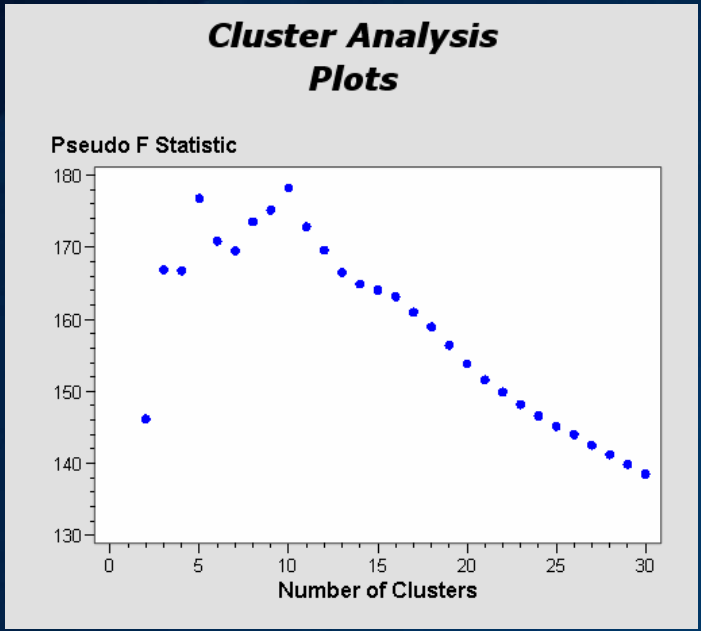
Export Hot Spot Data to SAS

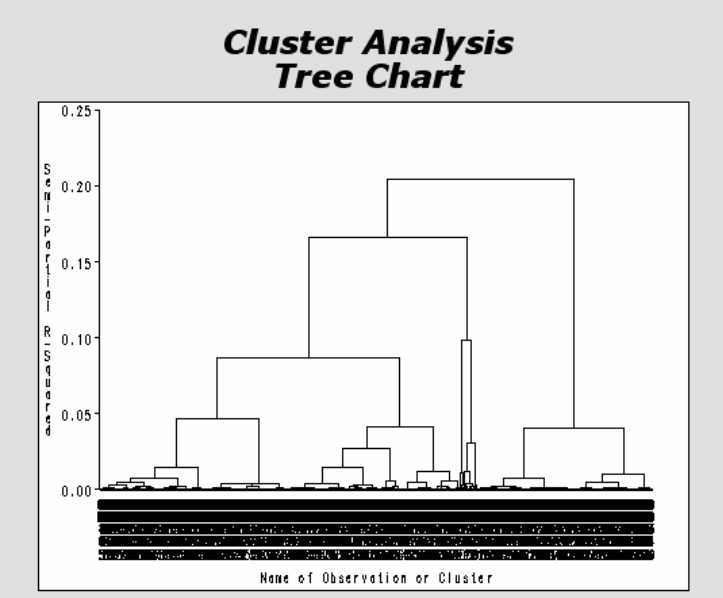


Cluster Analysis in SAS



Exploring Clusters





Correlation

	Pop Density	Docs per 100K	Crime	Asian	Urban	PCI	Med Inc	Med House Inc	Farmland	Beds	White	Black	Amer Ind	Latin
	1	0.38168	0.3779	0.34074	0.32043	0.29355	0.23565	0.21367	-0.1553	0.07876	-0.07075	0.06815	-0.02103	-0.01223
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.0602	0.0915	0.1041	0.6163	0.7707
	Uran	Crime	Asian	Med Fam Income	Doc per 100K	Med House Inc	PCI	Pop Density	Latin	Beds	Amer Ind	Farmland	Black	White
	1	0.69104	0.5865	0.49051	0.45674	0.44445	0.43311	0.32043	0.22046	0.21843	-0.09025	0.05491	-0.04648	0.01343
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0312	0.1905	0.2679	0.7491
	White	Black	Med House Income	Med Fam Income	PCI	Farmland	Laint	Amer Ind	Pop Density	Beds	Docs per 100K	Urban	Asian	Crime
	1	-0.96071	0.41256	0.37378	0.35357	-0.21944	0.10272	-0.09225	-0.07075	-0.06973	-0.01479	0.01343	0.01009	0.00823
		<.0001	<.0001	<.0001	<.0001	<.0001	0.0141	0.0276	0.0915	0.0963	0.7245	0.7491	0.8101	0.8446

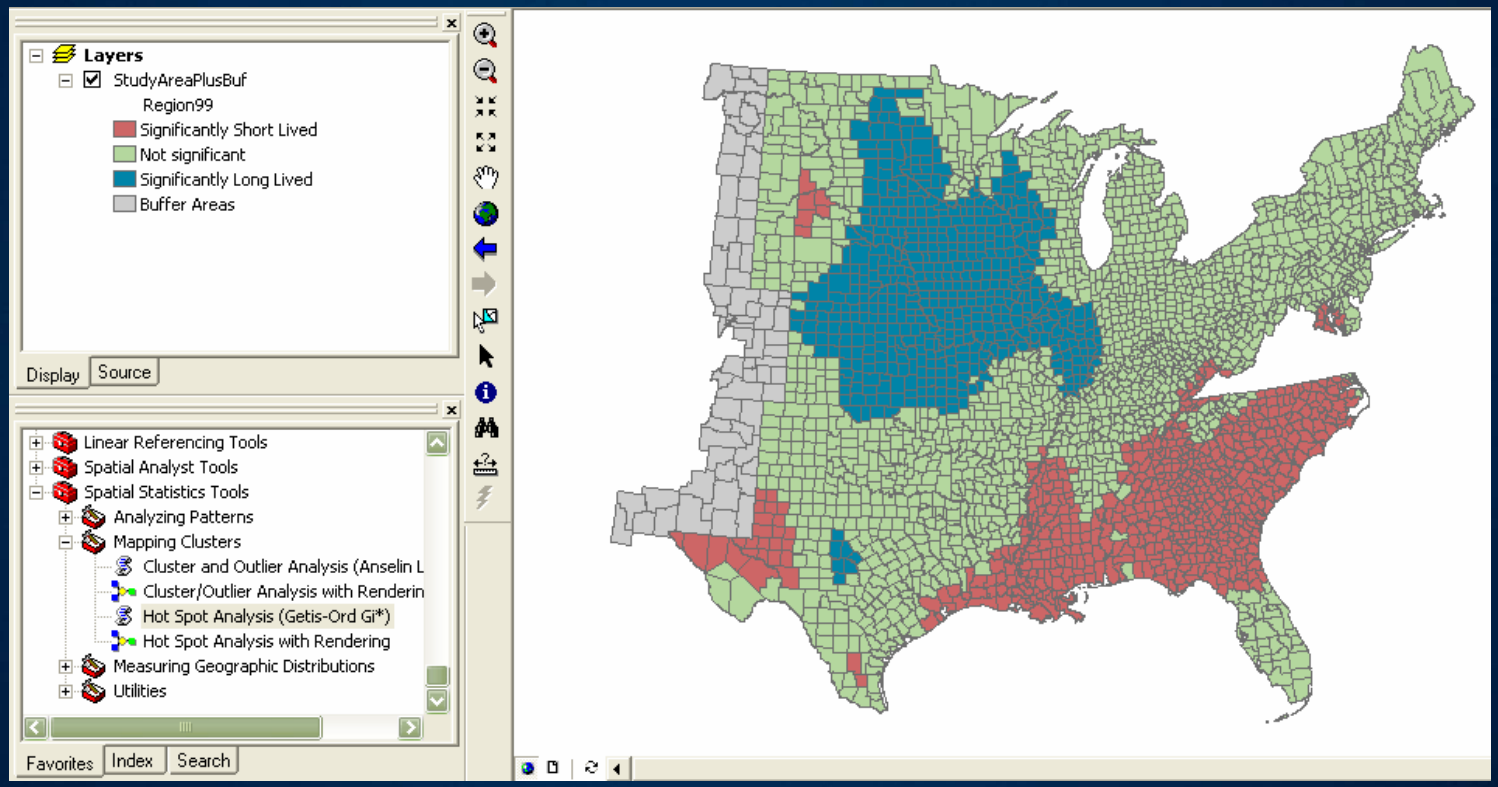
Chi² Analysis

Chi2 Analysis: Test Relationship between Income and Age at Death			
	Age At Death		
Household Income	Younger	Older	Total
Low	364	236	600
	63.86	43.95	
High	206	301	507
	36.14	56.05	
Total	570	537	1107

Analysis Findings

Are there regional trends for average age at death?

- Yes, there are persistent statistically significant regions in the US where average age of death is either younger or older than expected



Are there specific demographic factors for early death “hot spots”?

- Lower incomes
- Fewer physicians and hospital beds

Are there specific demographic factors for late death “hot spots”?

- Low crime rates

Is median income significant?

Average	63.86	43.95	
Above Average	206 36.14	301 56.05	507
Total	570	537	1107

Square	1	44.4402	<.0001
Continuity Adj. Chi-Square	1	43.3645	<.0001
Mantel-Haenszel Chi-Square	1	44.1231	<.0001
Phi Coefficient		0.1997	
Contingency Coefficient		0.1959	
Cramer's V		0.1997	

Is number of physicians significant?

Chi2 Analysis			
Doctors per 100K	Age at Death		Total
	Younger	Older	
Fewer than Average	301 52.81	254 47.3	555
More than Average	269 47.19	283 52.7	552
Total	570	537	1107

Statistic	DF	Value	Prob
Chi-Square	1	3.3545	0.067
Likelihood Ratio Chi-Square	1	3.3562	0.067
Continuity Adj. Chi-Square	1	3.1378	0.0765
Mantel-Haenszel Chi-Square	1	3.3515	0.0671
Phi Coefficient		0.055	
Contingency Coefficient		0.055	
Cramer's V		0.055	

Summary

1. Are there persistent hot/cold spots? Where are they? (Explore data spatially)
2. Do those hot/cold spots have similar characteristics? (Explore the attribute data… a-spatial data analysis)
3. Identify a set of hypotheses
4. Test those hypotheses

Resources

- 911 emergency call analysis demo:
  http://www.esri.com/software/arcgis/arcinfo/about/demos.html
- Virtual campus free seminar
  http://campus.esri.com/
- Articles
  http://www.esri.com/news/arcuser/0405/ss_crimestats1of2.html
- Online help (“Modeling Spatial Relationships”)
- Instructor-led ESRI training
- The ESRI Guide to GIS Analysis, Volume 2 by Andy Mitchell

The ESRI Guide to GIS Analysis

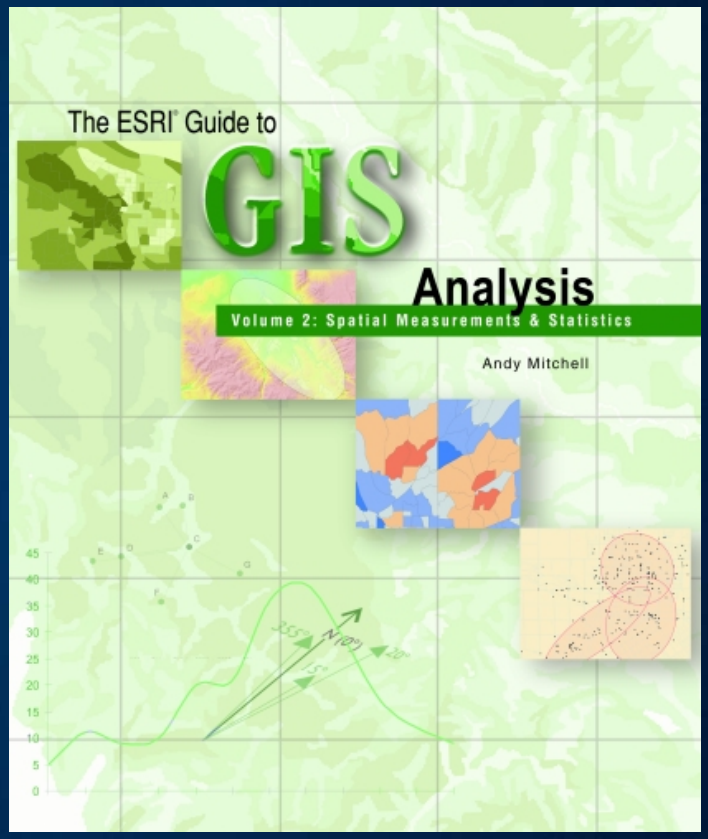
Volume 2

Chapter 2: Measuring Geographic Distributions

Chapter 3: Analyzing patterns

Chapter 4: Mapping clusters

Chapter 5: Analyzing relationships



Questions?

LScott@ESRI.com