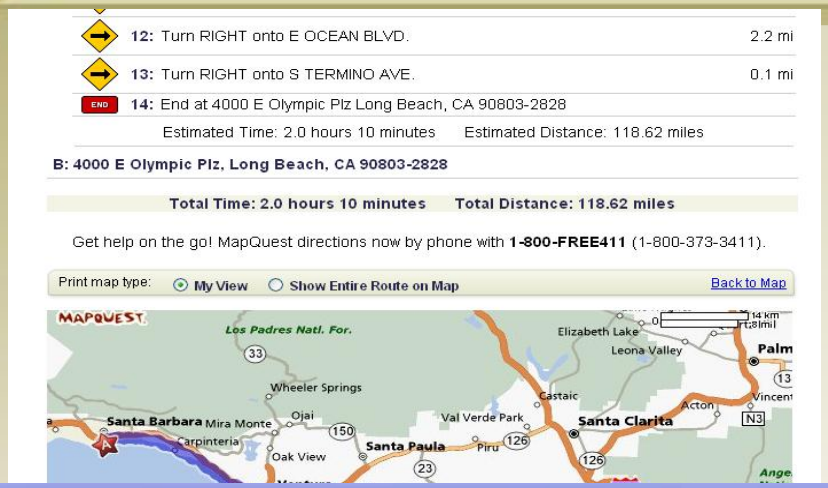
Spatial Statistics & Geographic Information Systems (GIS)

Patricia Carbajales,
Stanford Geospatial Center
carbajales@stanford.edu

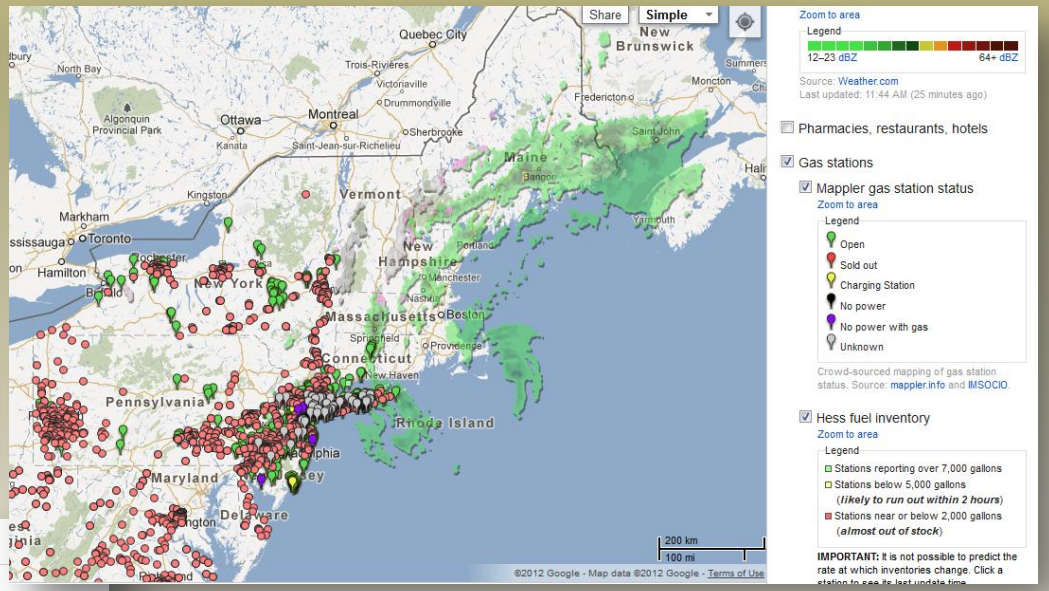
Agenda

- Introduction to GIS
- Location! The importance of choosing your projection
- Spatial Statistics: An Overview
- **Descriptive** spatial statistics
  - Histograms, Standard Deviational Ellipse
- **Inferential** spatial statistics
  - **Global**: Analyzing broad spatial **patterns**
    - **Spatial Autocorrelation**
  - **Local**: Mapping **clusters**
    - **Hot Spot Analysis**

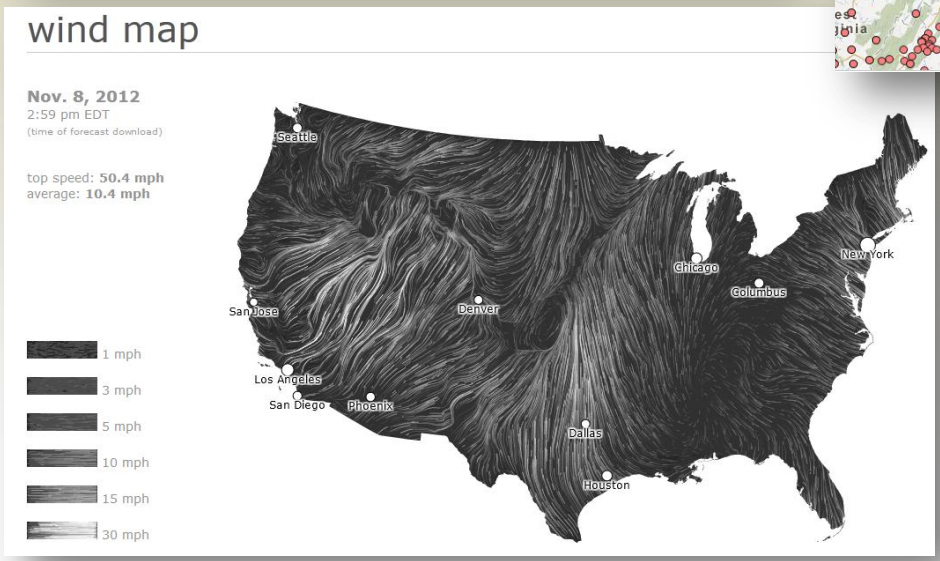
Geographic Information Systems in everyday life



Geographic Information Systems in everyday life



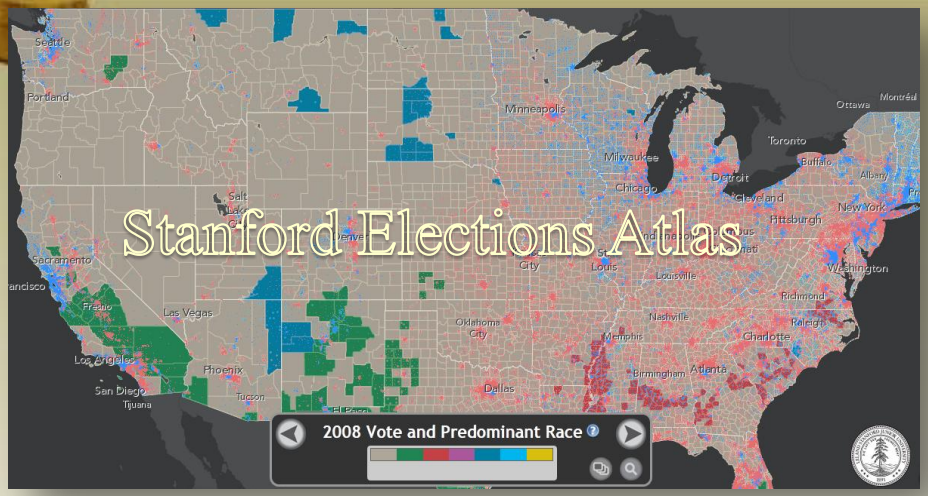
Credit: google.com



Credit: Hint.fm

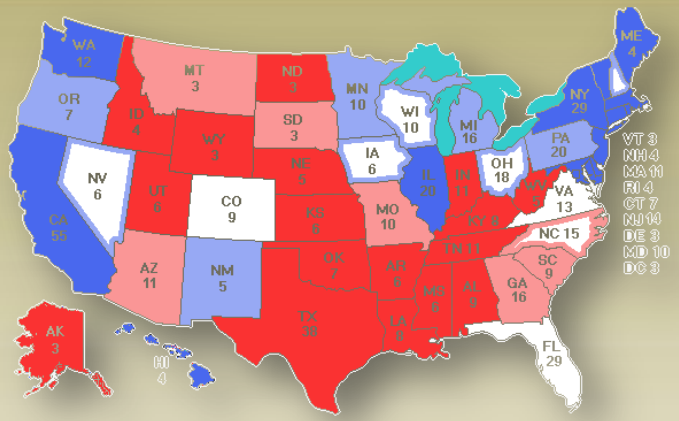
MAP CROWD-SOURCING

Geographic Information Systems in everyday life

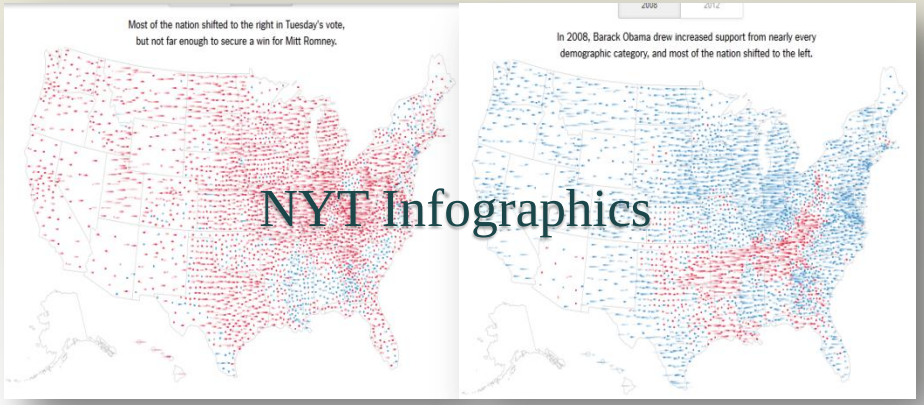


Credit: SSSL & ESRI

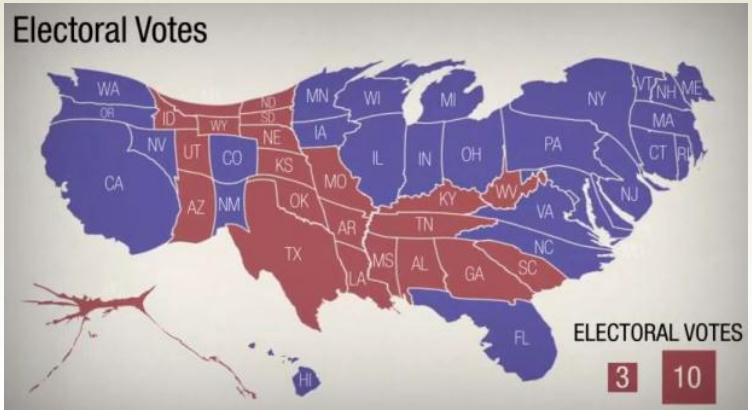
Mash up Version



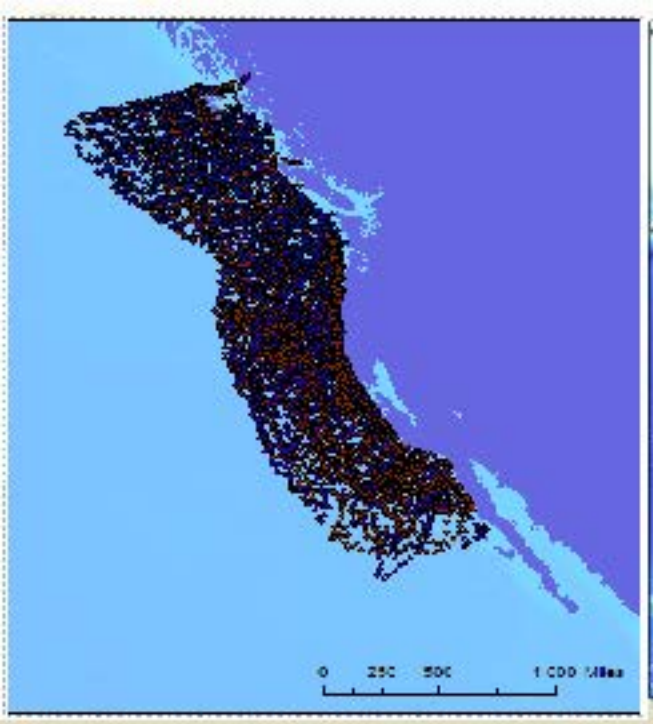
Credit: electoral-vote.com

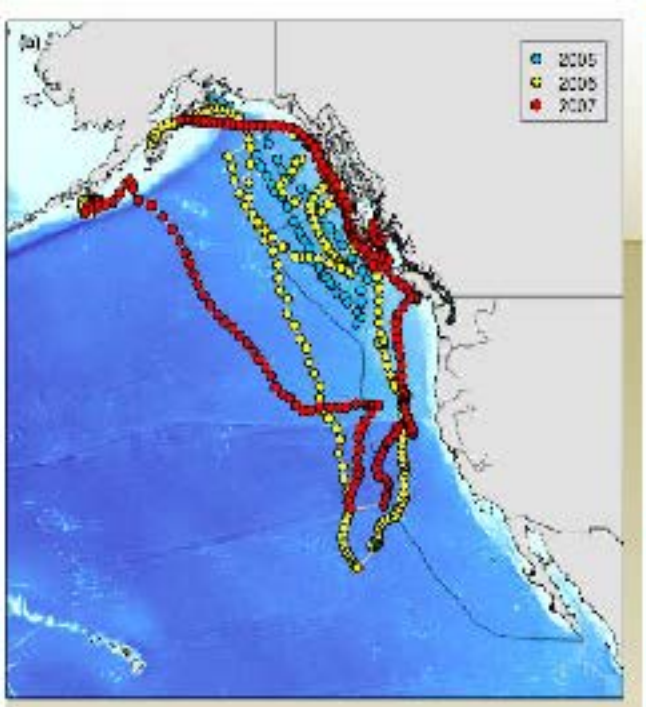


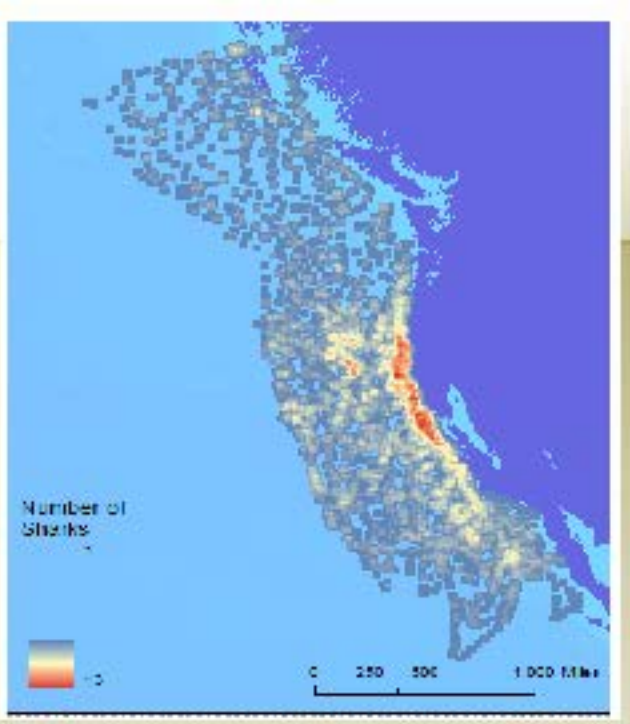
Credit: New York Times

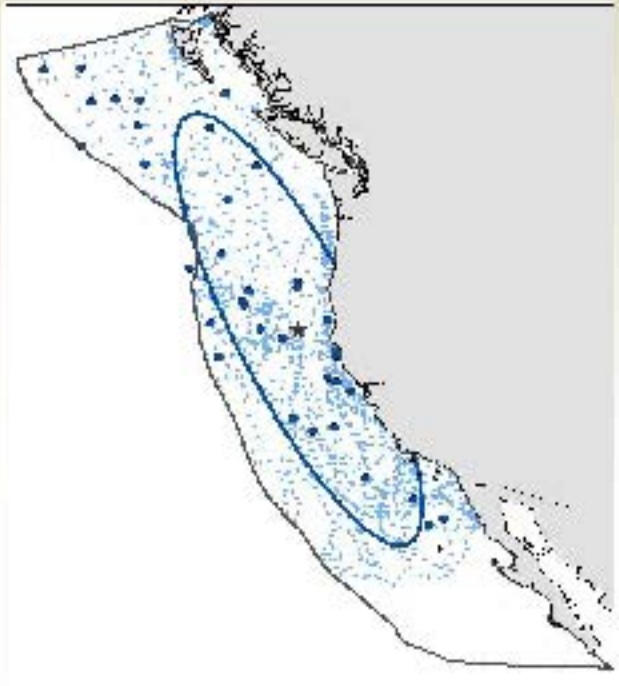


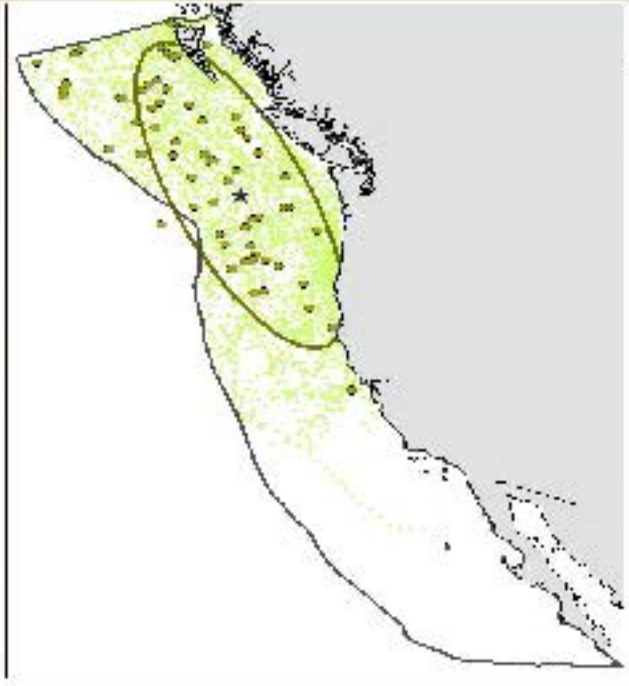
Credit: NPR

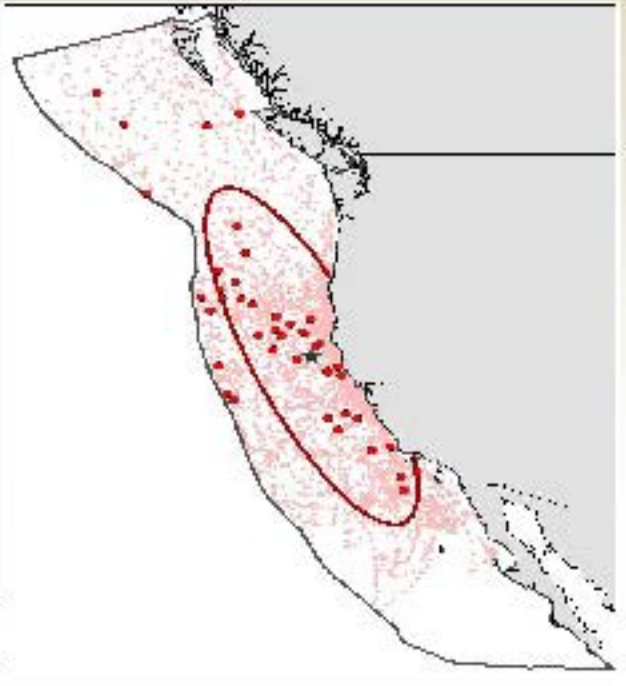

0 250 500 1,000 Miles


(a)
2005
2006
2007


Number of Sharks
0 250 500 1,000 Miles



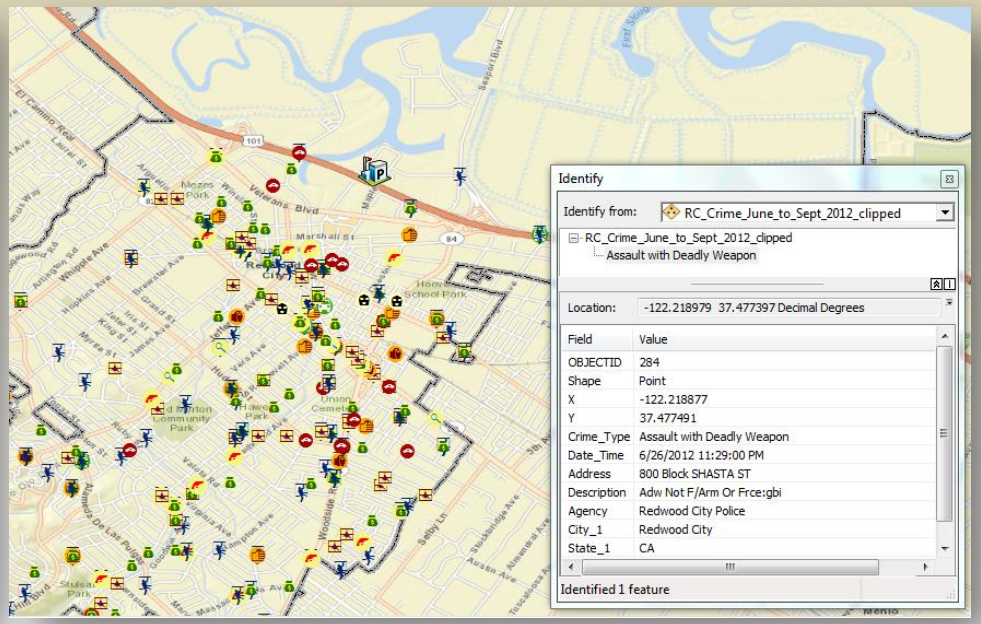




What is GIS?

A geographic information system (**GIS**) is a computer-based tool that **links geographic information** (where things are) with **descriptive information** (what things are).

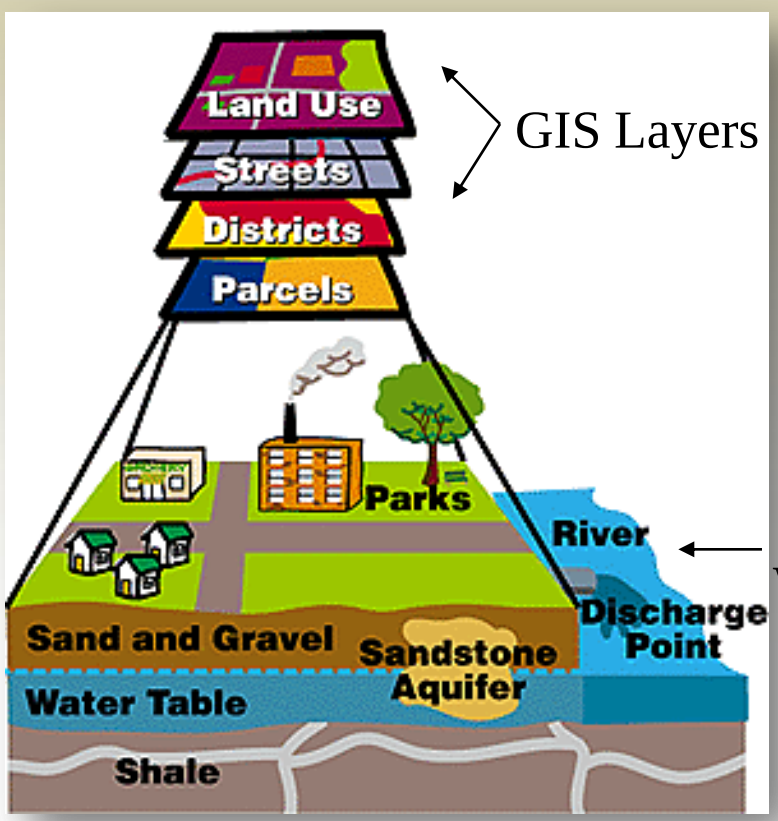
Paper map



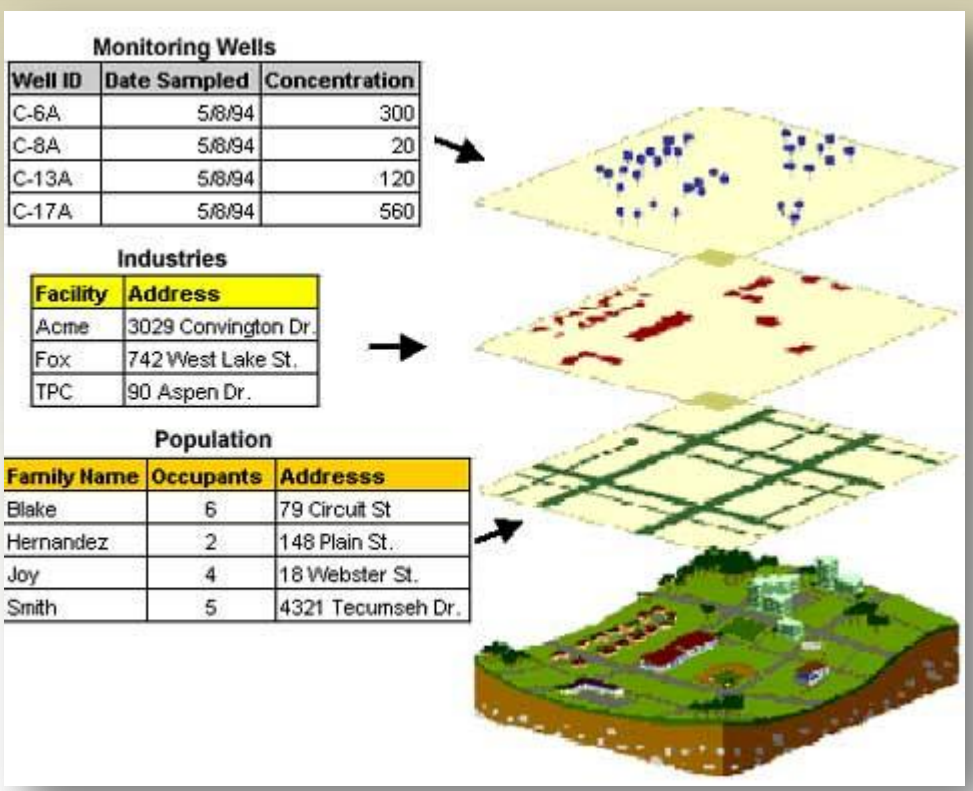
Digital map

What is GIS?

A GIS is: "***A system for capturing, storing, checking, integrating, manipulating, analyzing and displaying data which are spatially referenced to the Earth*** (Chorley, 1987)."



Real World

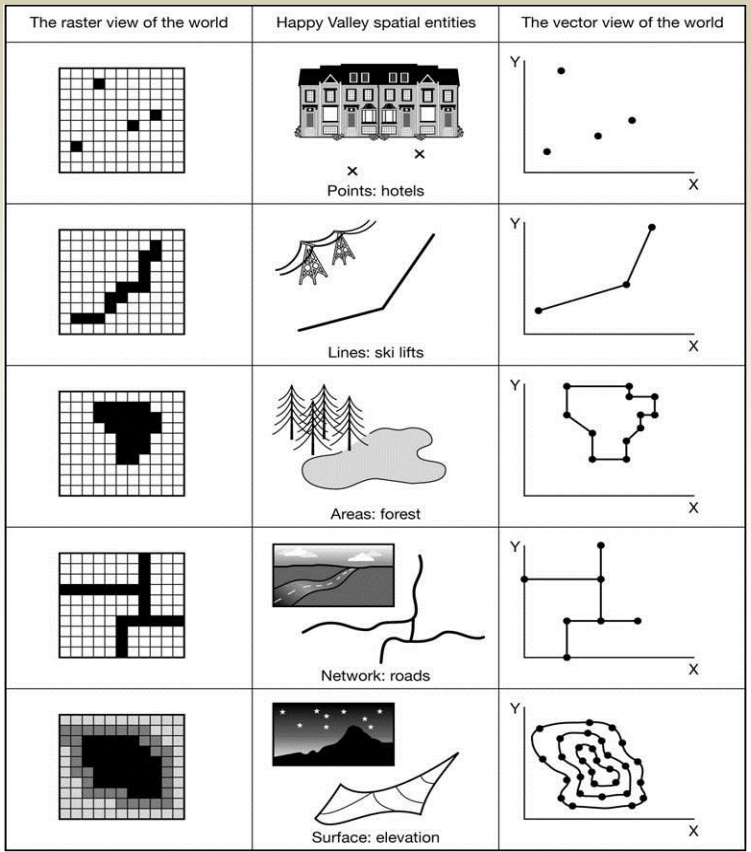


How GIS Works

A GIS stores information about the world as a collection of **thematic layers** that can be linked together by geography

There are 2 basic spatial data types representing the real world:

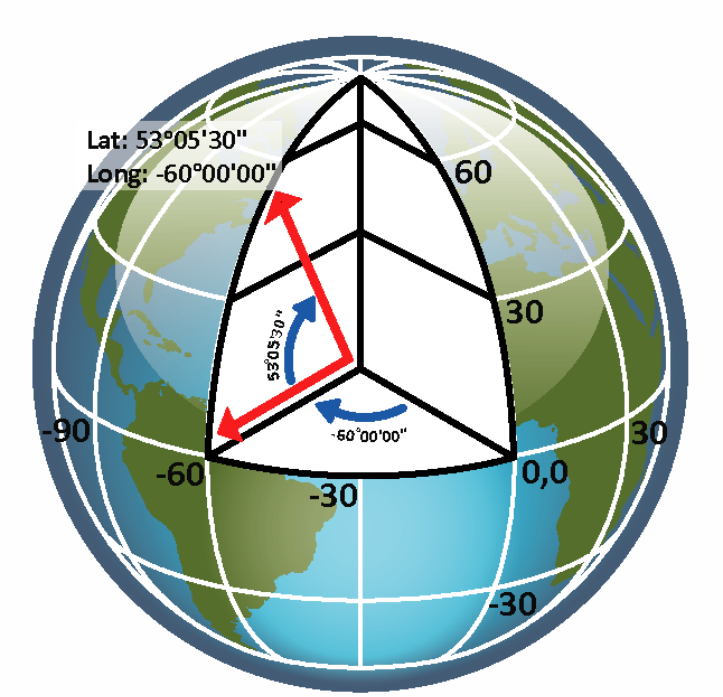
Raster **Vector**

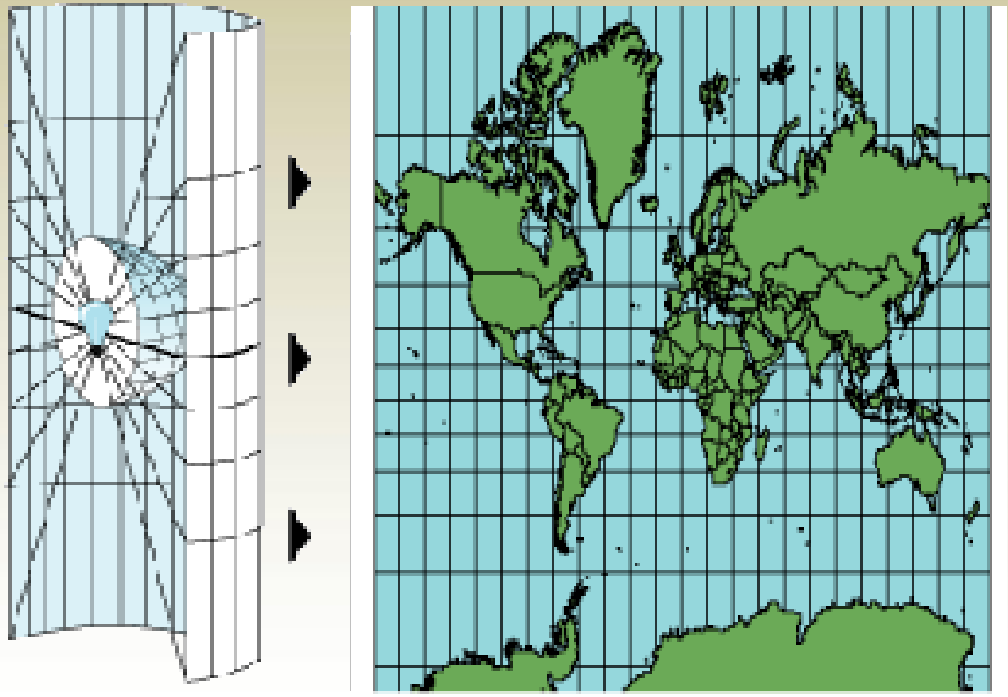


Location, location, location!!
Earth Reference Systems

Geographic Coordinate Systems

Projected Coordinate Systems

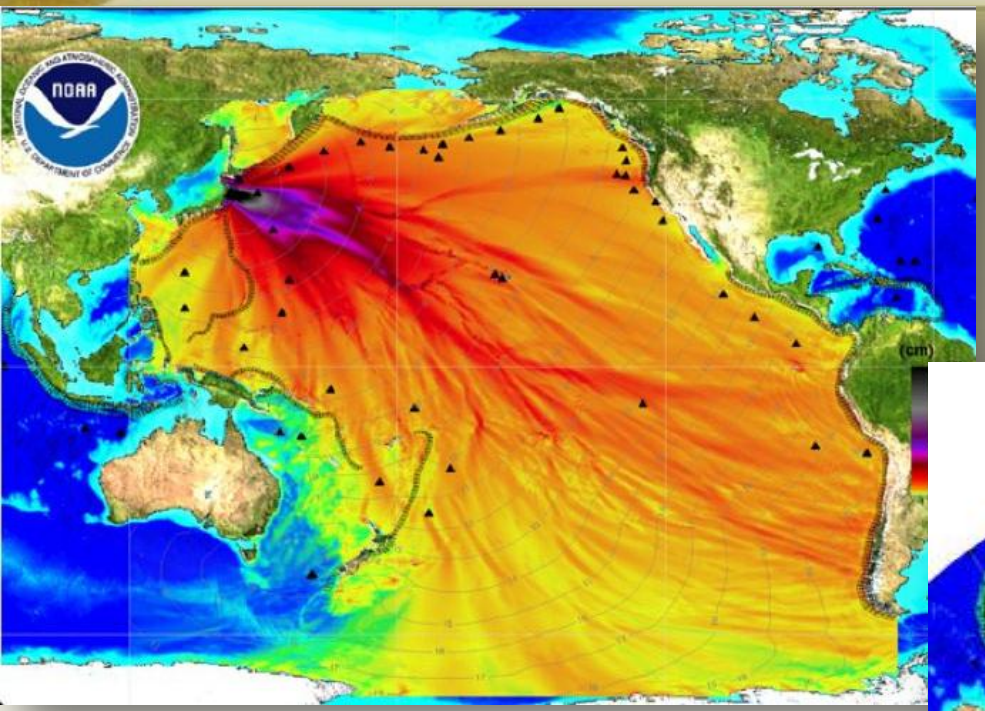


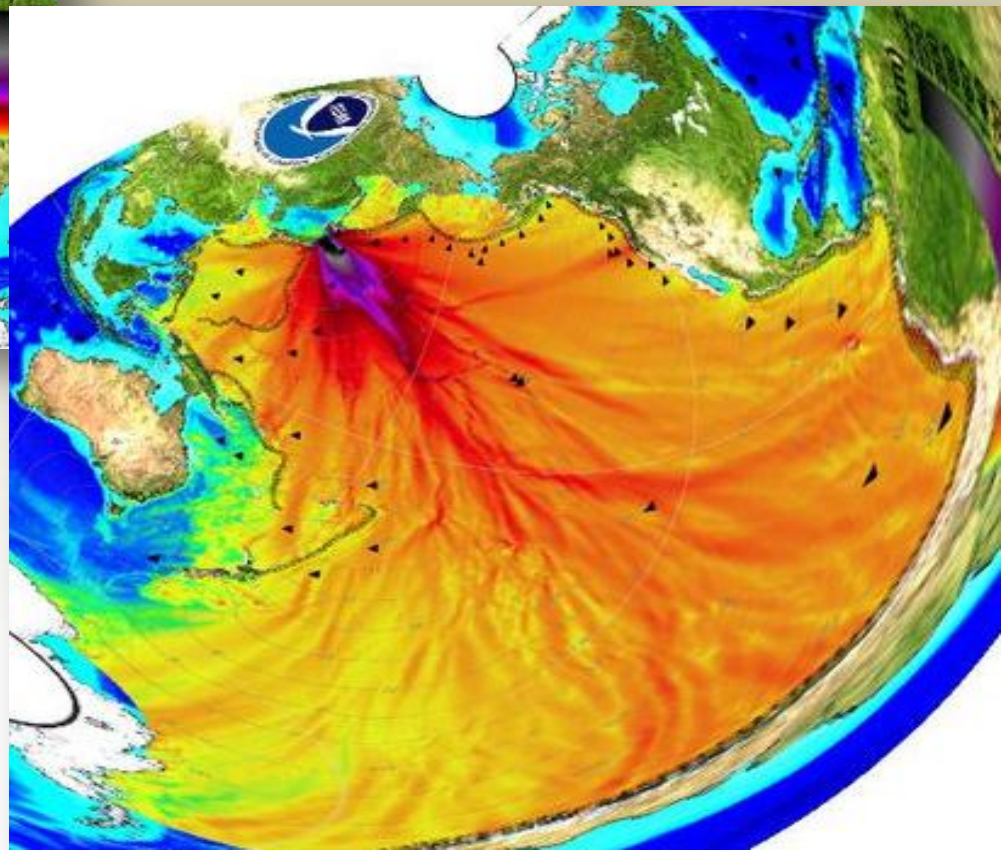


Latitude, Longitude:
Always ask for the Datum!

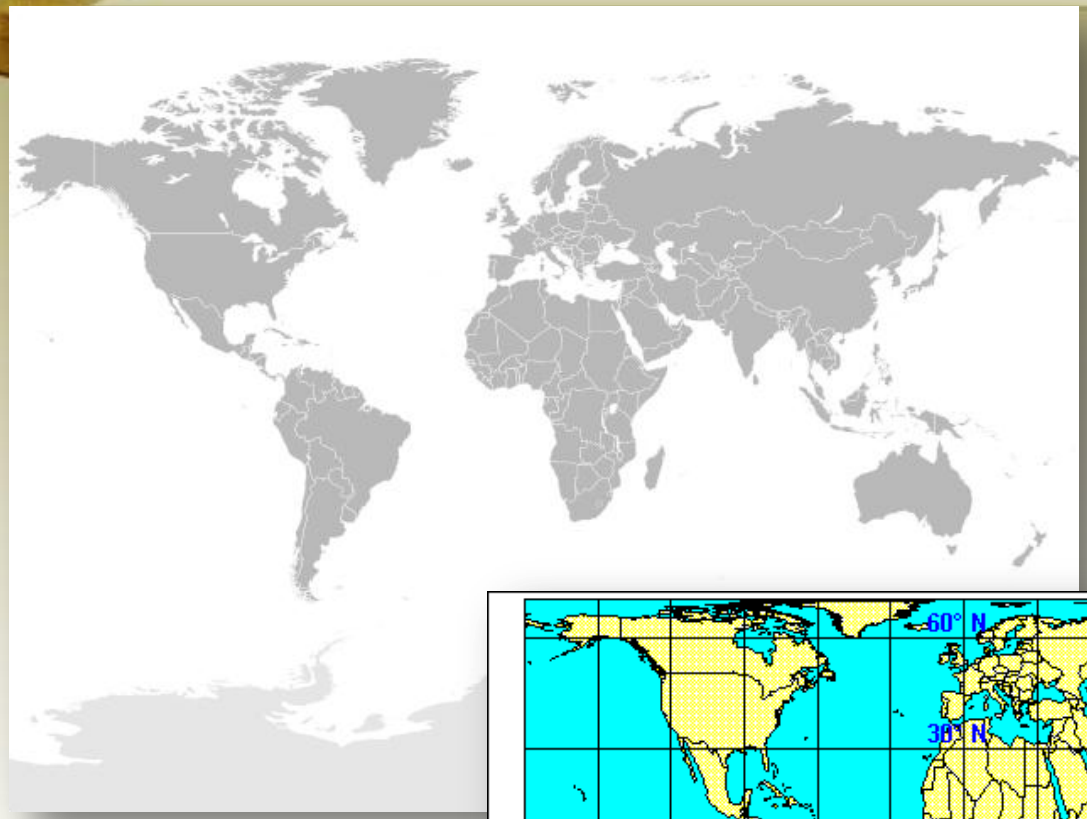
Mercator, UTM, State Plane,
Albers Equal Area, Equidistant

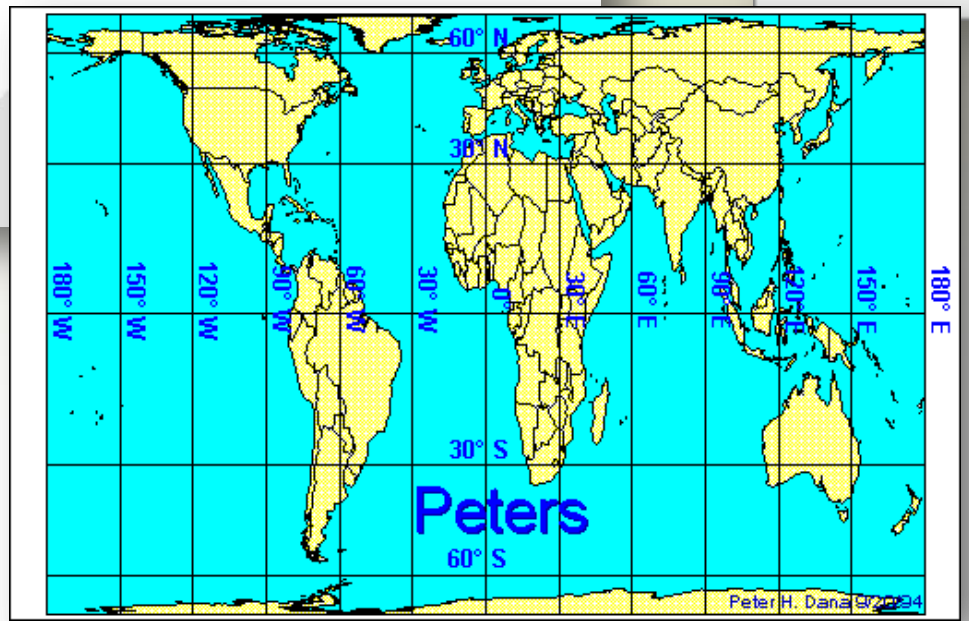
The Importance of Projections



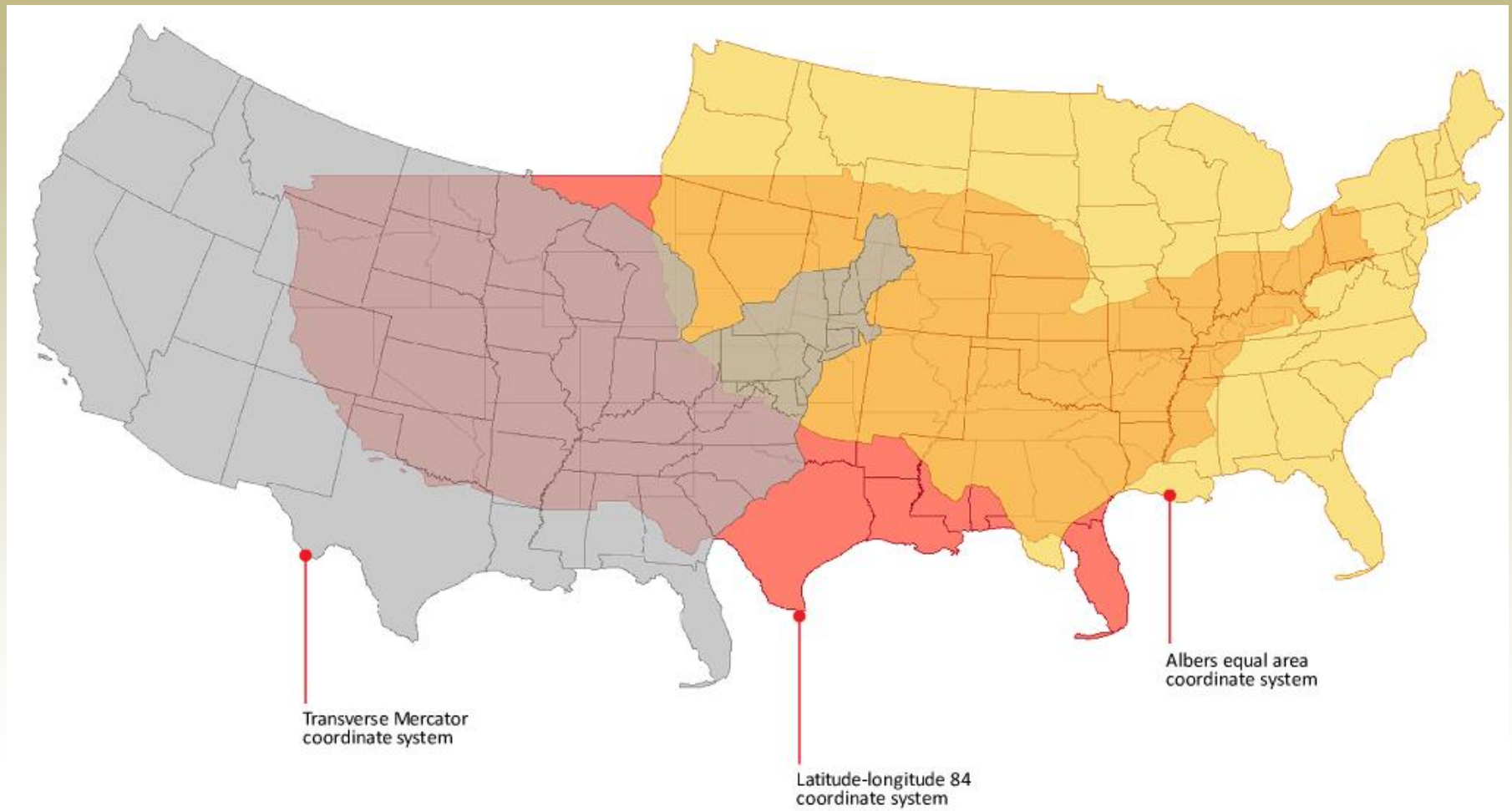


The Importance of Projections - cont





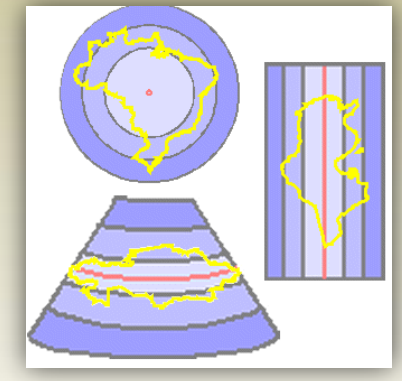
Which Projection?



Always work in Projected Coordinate Systems!

Which projection to use??

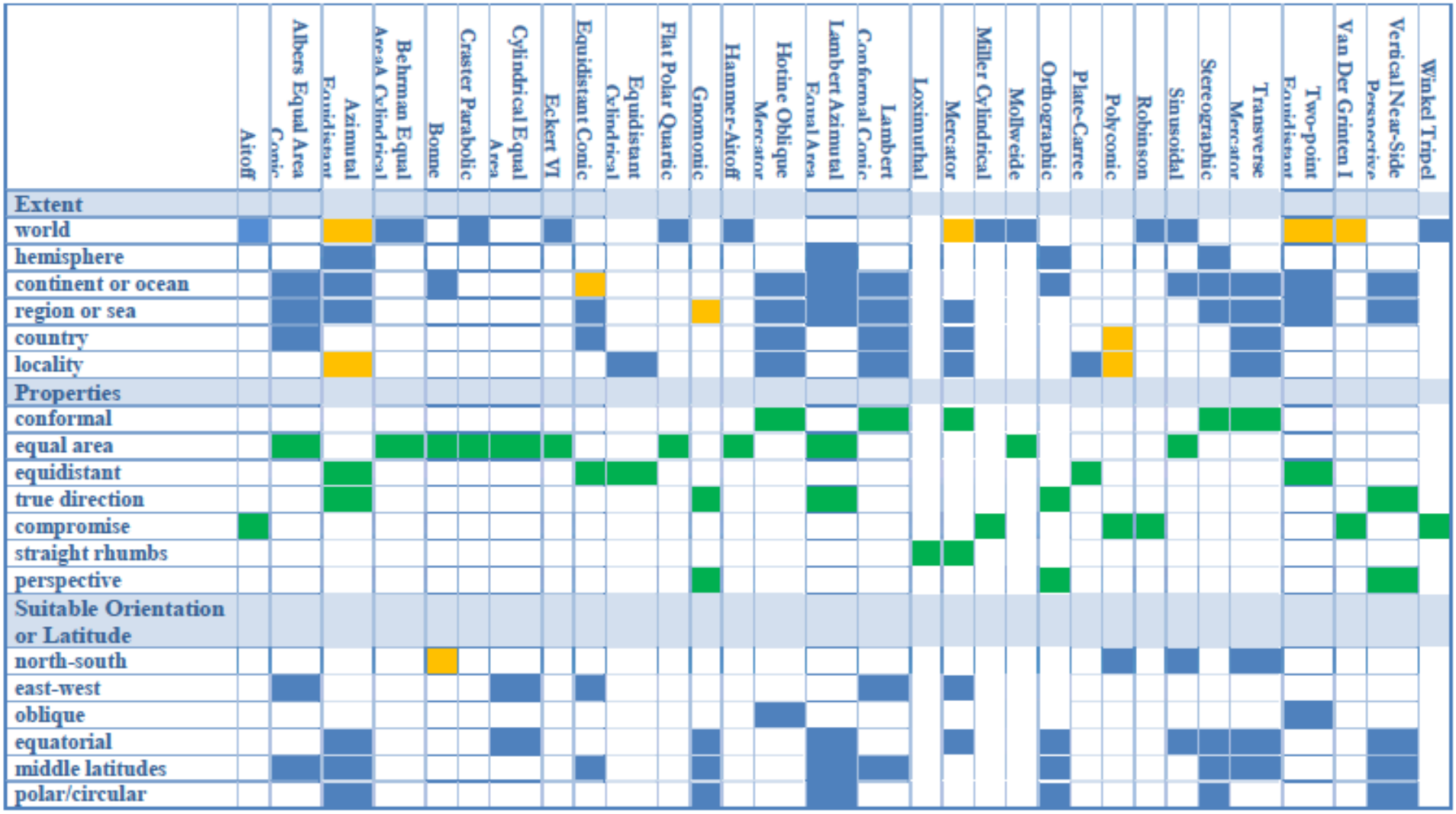
- **What is the map's purpose?**
  - For general reference and atlas maps, you usually want to balance shape and area distortion
  - If your map has a specific purpose, you may need to preserve a certain spatial property— shape, area—to achieve that purpose.



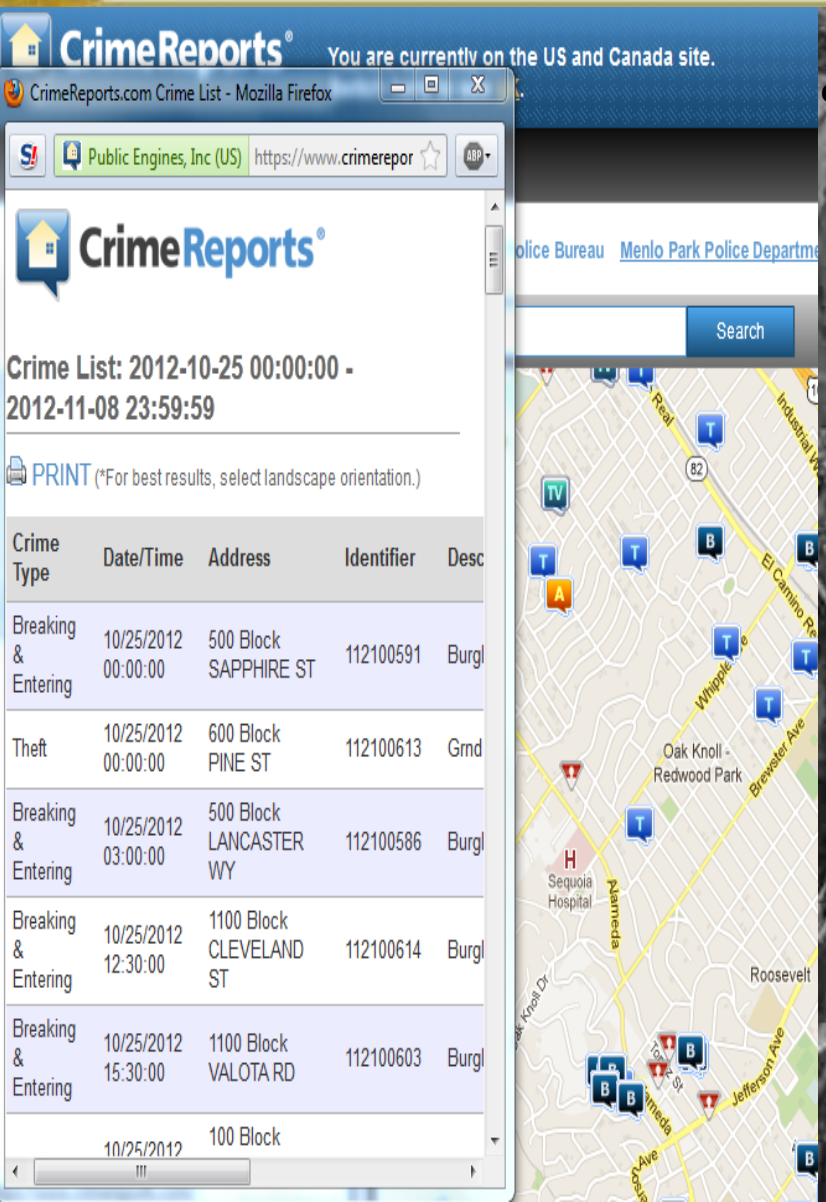
- **What shape is your area of interest?**
  - Areas that extend along a great circle: cylindrical projection
  - Areas that extend along a small circle: conic projection
  - Areas that are approximately circular: azimuthal projection

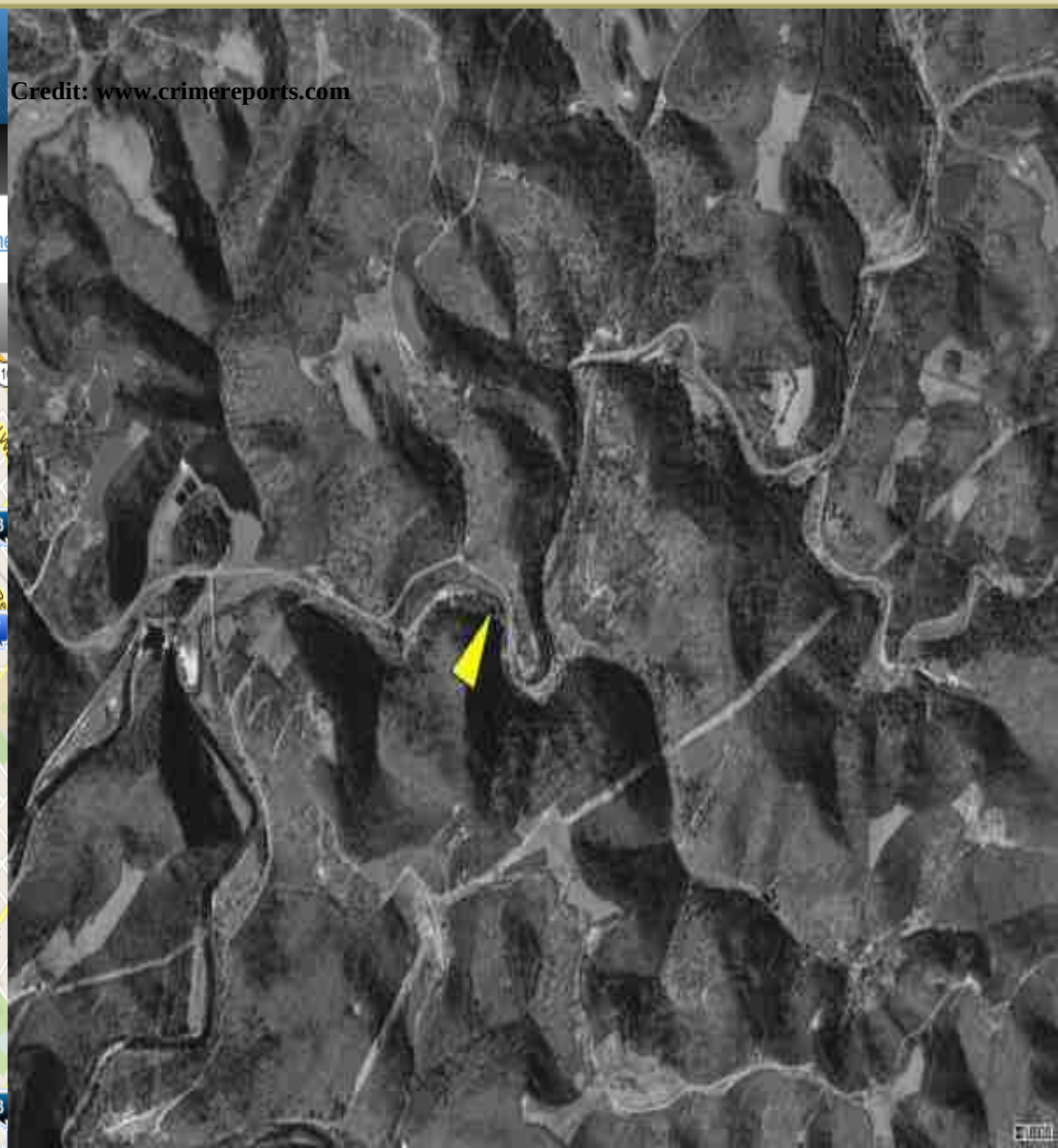
- **Which part of the world does your map show?**
  - Tropical regions: cylindrical projection
  - Middle latitudes: conic projection
  - Polar region: azimuthal projection

PROJECTIONS CHART

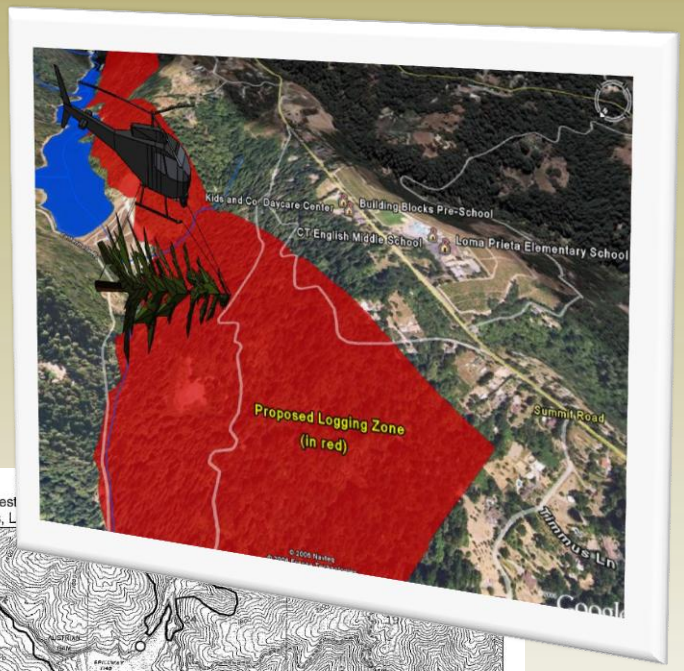


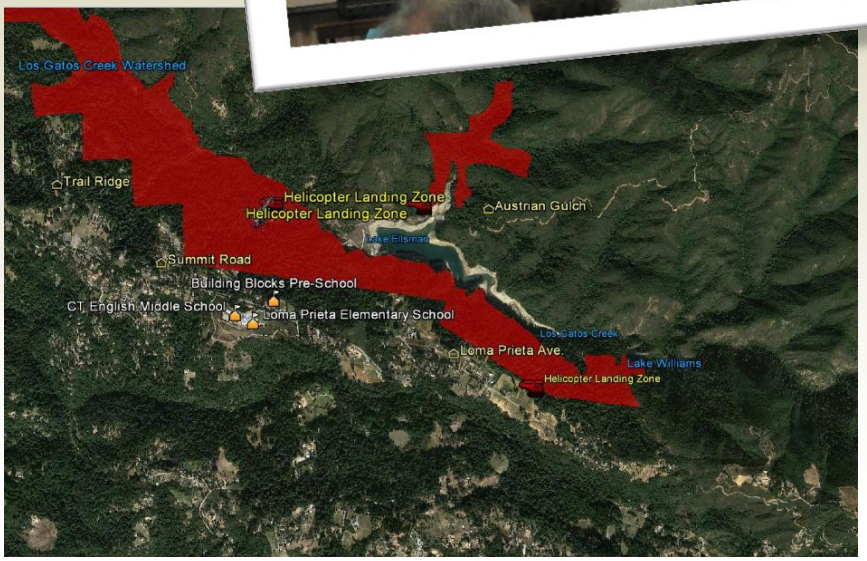
GIS Functions (I): Data Gathering

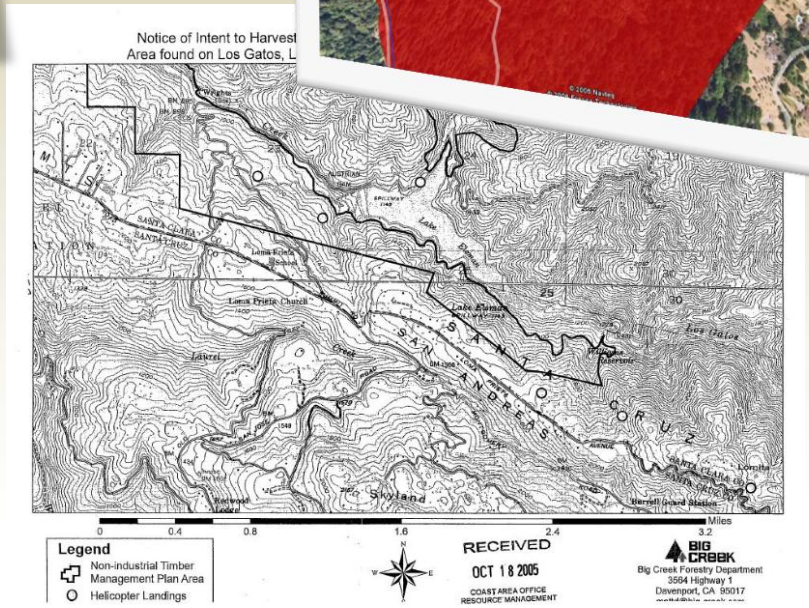




GIS Functions (I): Data Gathering

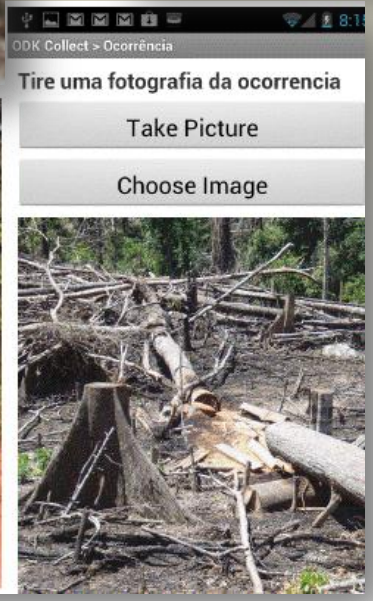






Neighbors Against Irresponsible Logging (NAIL).

GIS Functions (I): Data Gathering



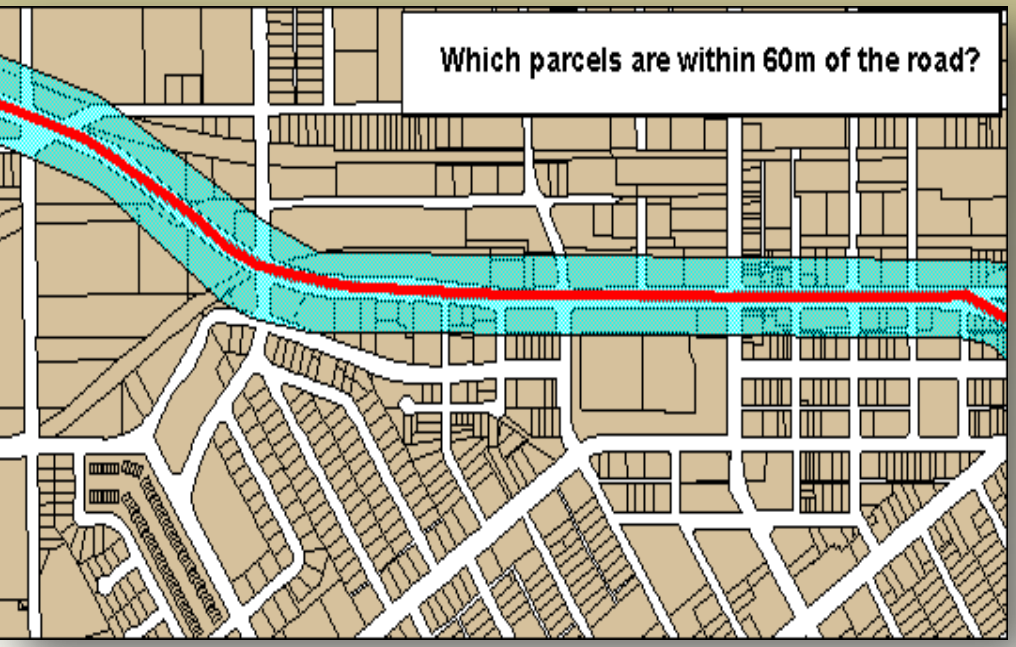
Field data gathering with mobile phones: **Surui Forest Carbon Project**

Assessing the Impact of Land and Forest Laws on Forest Cover in the Brazilian Amazon

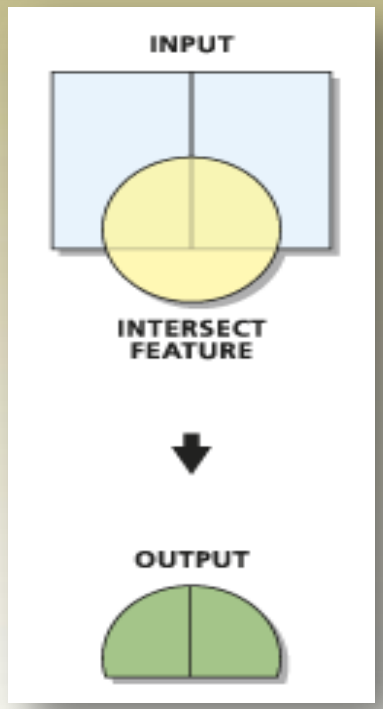
By Brenda Brito do Carmo,
Theo Gibbs-Plessl,
and Tamer Shabani,
Stanford University

GIS Functions (II): Spatial Analysis

Buffer Analysis

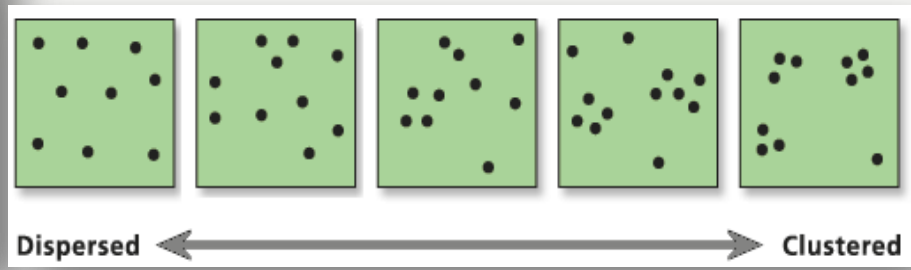


Intersection

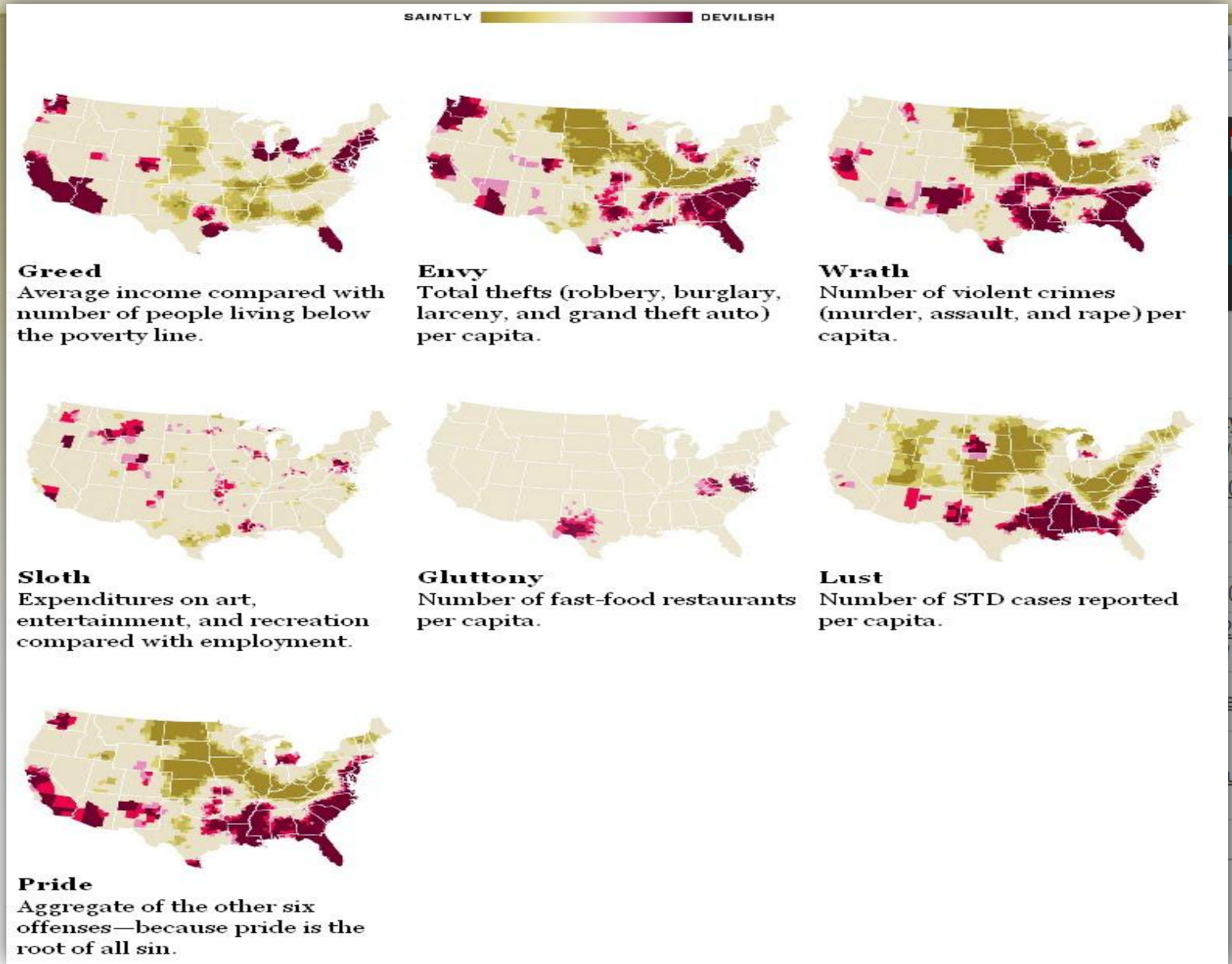


Hot Spot Analysis

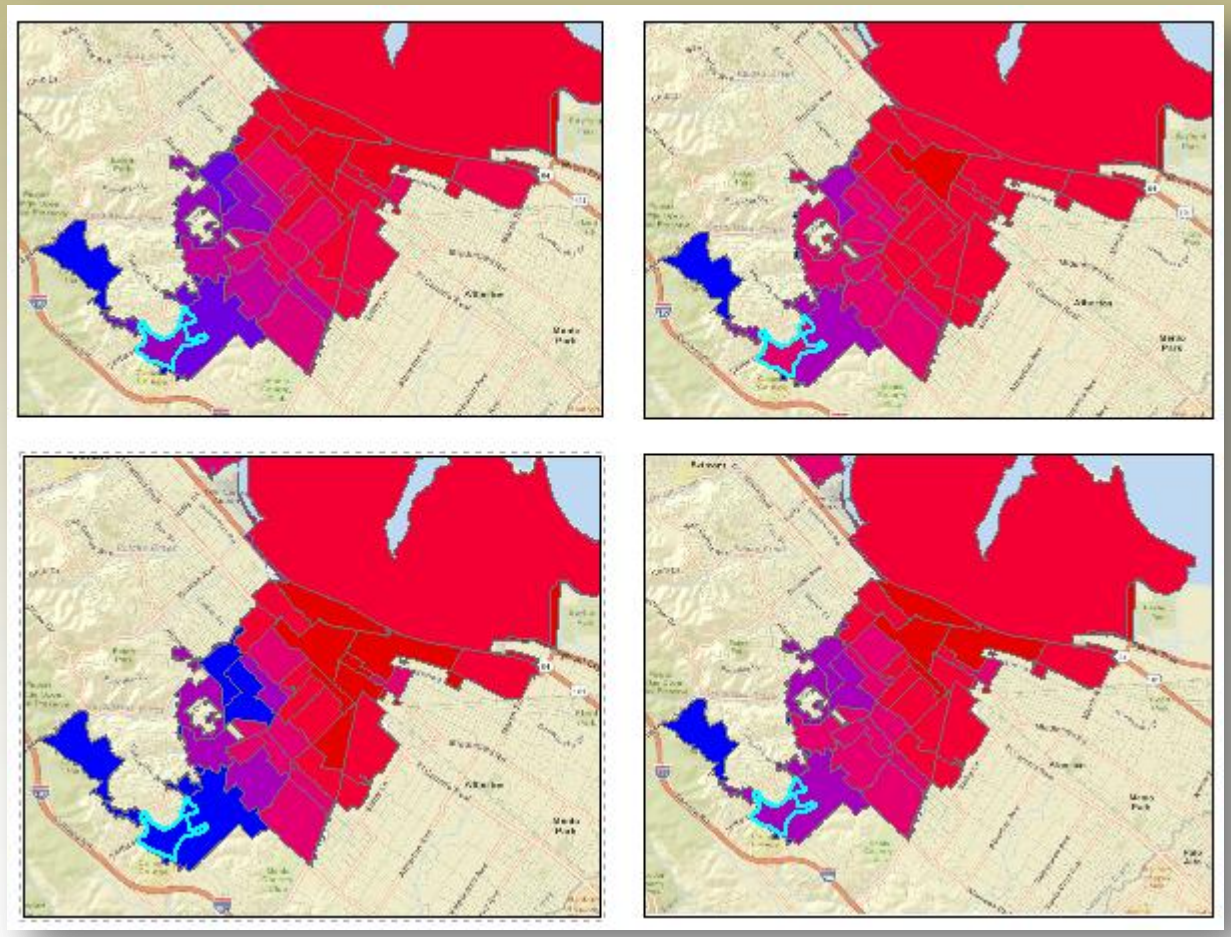
Average Nearest Neighbor



GIS Functions (III): Visualization

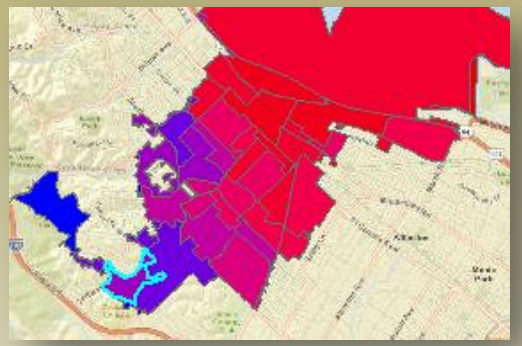


Demonstration



The Ambiguity of Map Symbolization

Spatial Statistics in GIS

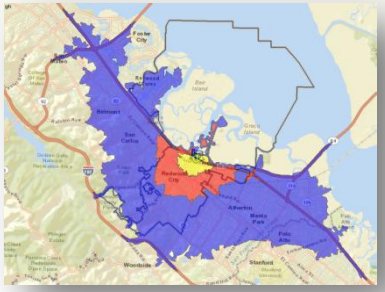


Spatial statistics:

"*Spatial statistics is the collection of statistical methods in which spatial locations play an explicit role in the analysis of data (*Ribeiro and Diggle, 2001)"

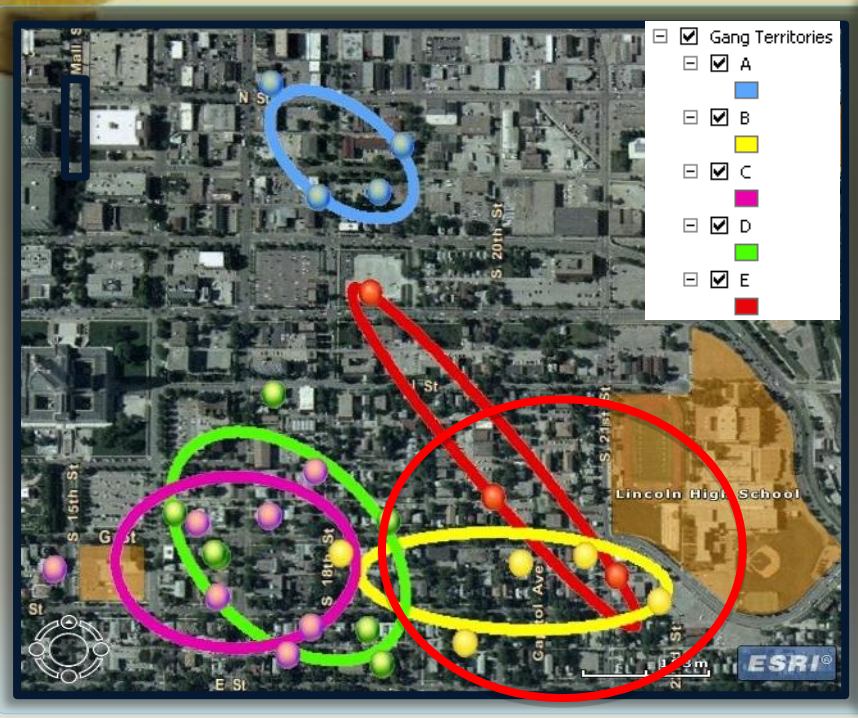
- **Incorporate space** (area, length, proximity, orientation, and/or spatial relationships) **directly into** their **mathematics**

- Set of **exploratory techniques** for describing and modeling spatial data and spatial processes .

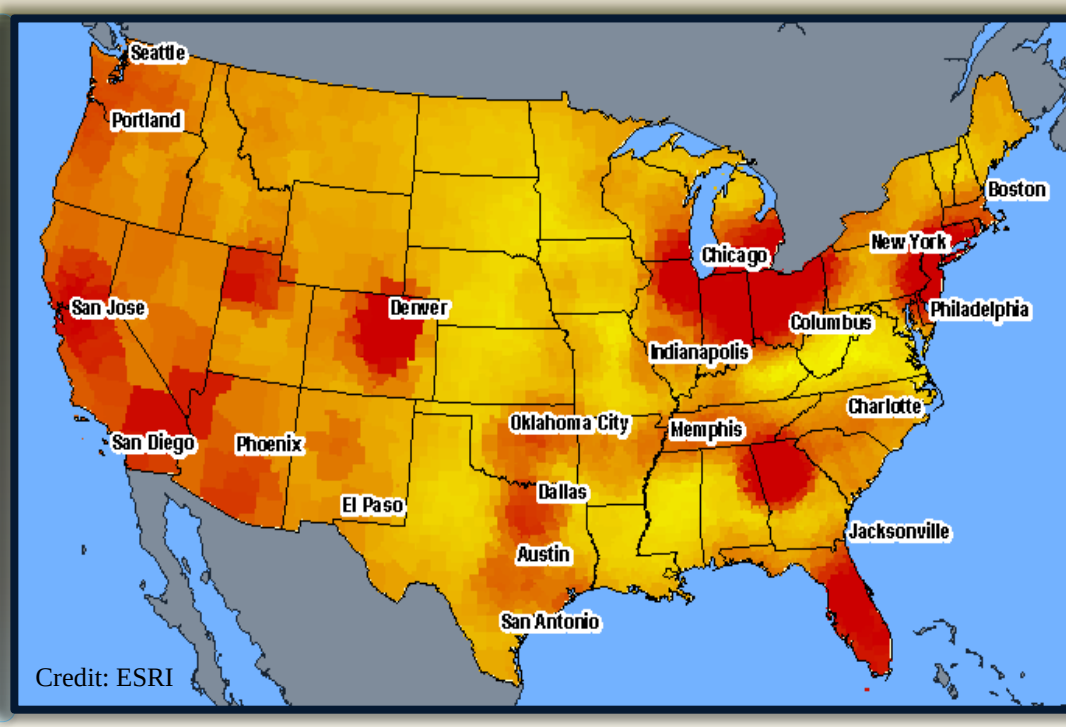


Spatial statistics extend what the eyes and mind do naturally to assess spatial patterns, trends and relationships.

What kind of questions can you answer?

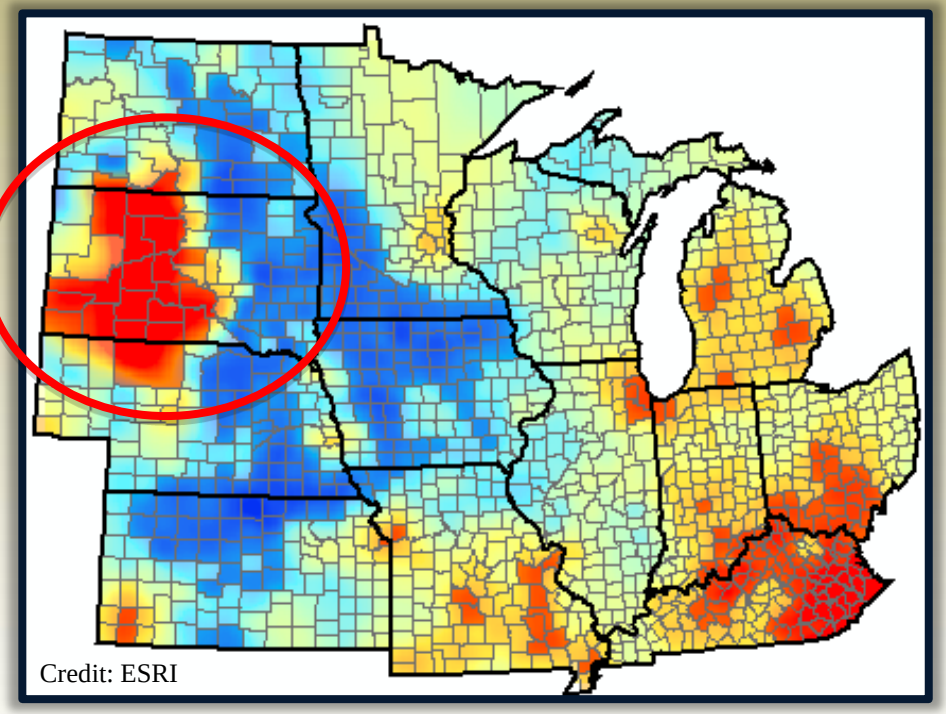


Where are gang territories overlapping in our city?



Where are hot spots of foreclosures in the U.S.?

What kind of questions can you answer?



- Recognize **patterns** and **trends**
- Identify **outliers** and **anomalies**
- Visualize **relationships**, **proximity**, **connectivity**
- **Interpolate** and **extrapolate**

Why are people dying young in South Dakota?

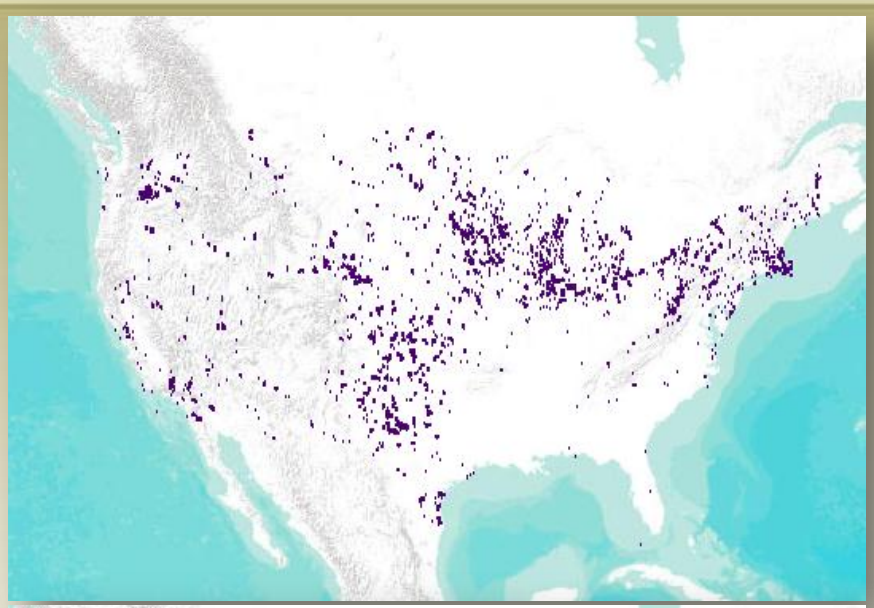
Why use spatial statistics?

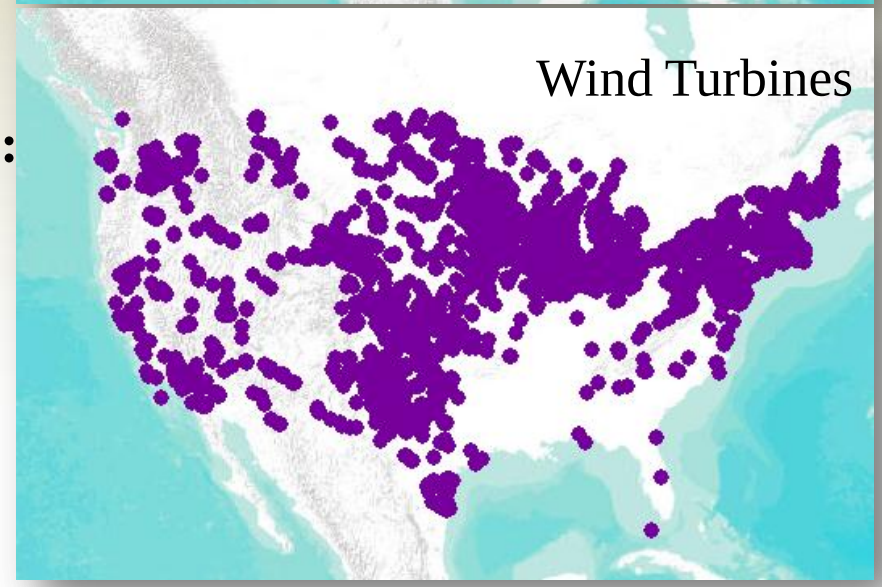
To Assess:

Patterns

Relationships

Trends





How we present our results in a map:

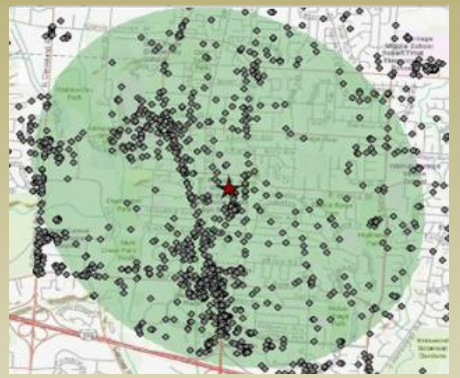
- **Colors**
- **Class Breaks**
- **Symbols Style & Size**

Can either enhance or obscure communication

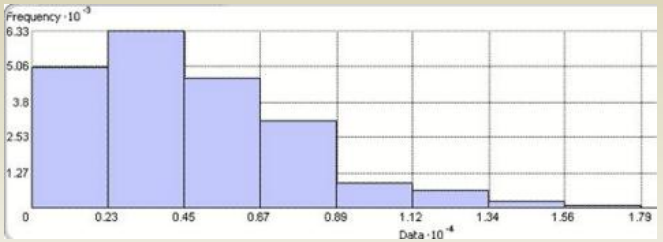
First step always, Explore your data!

- Why explore your data?
  - Allows you to better select an appropriate tool to analyze your data
- Data exploration should always be the **first step** during a data analysis project
- What **Factors** to consider when exploring your data?
  1. What is the Spatial/Geographic location of your data?
  2. What are the most common data values?
  3. How is the value of the data related to its location?
  4. How do I use this information to select an analysis tool?

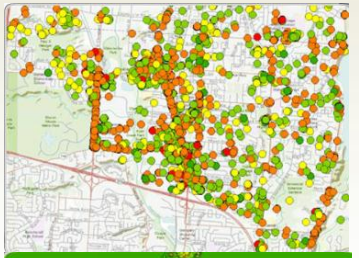
Exploring your data



- Explore the **Location** of your data
  - Factors to consider
  - Where is the data?
  - Spatial Distribution of the data

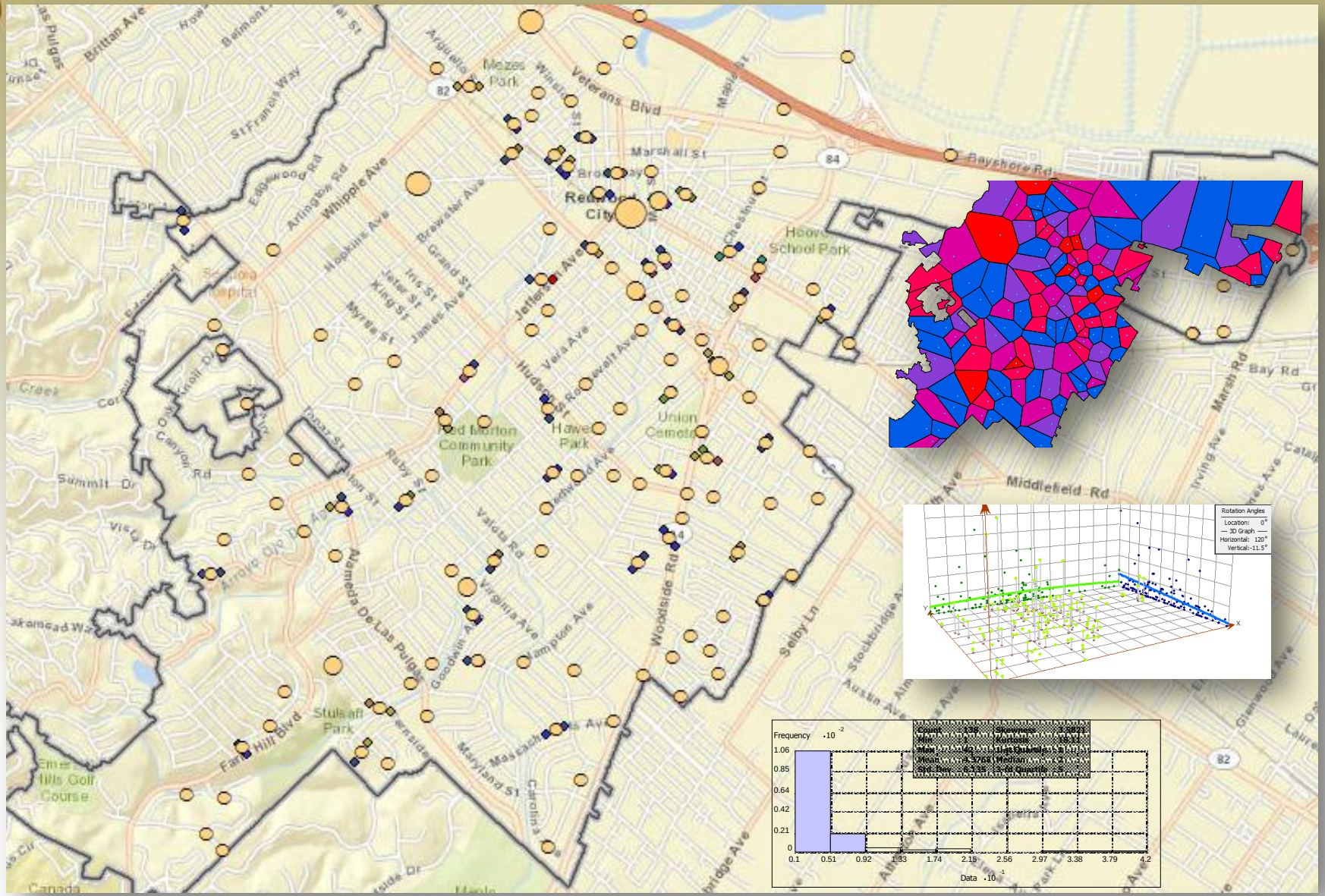


- Explore the **Values** of your data
  - Normally distributed data?
  - Frequency of your data? Distribution?



- Explore **Spatial Relationships** in your data
  - Value of the data related to its location:
  - Tobler's first law of Geography: "*Everything is related to everything else, but near things are more related than distant things*"
  - Variation in your data, Outliers
- Explore **Trends** in your data

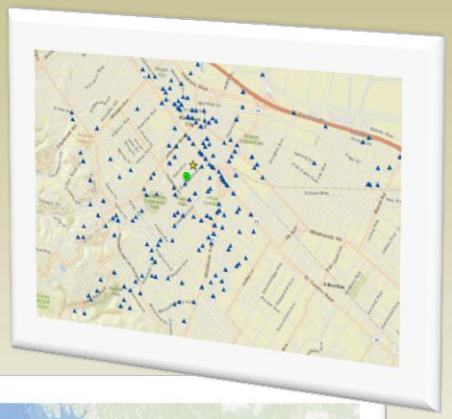
Exploring Crime Data

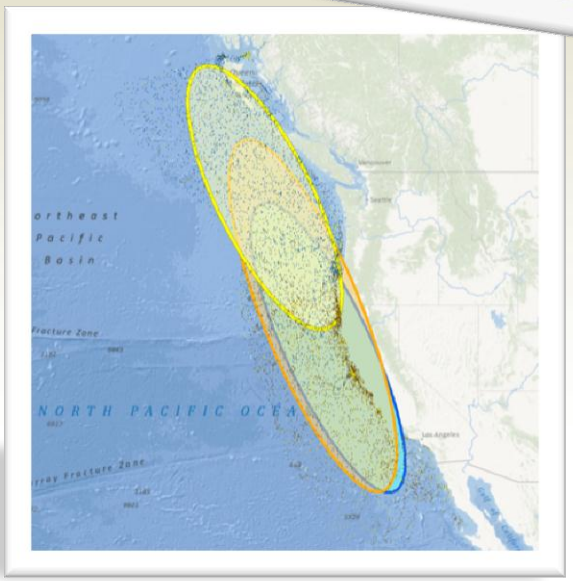


Descriptive Tools: Measuring Geographic Distribution

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

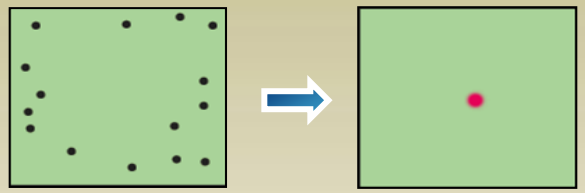
- Questions
  - Which neighborhood is most accessible?
  - Is there a directional trend or bias in incidents?
  - What is the primary direction of urban growth?
  - Where is the crime center?
  - Which gang has the broadest territory?
  - Which type of crime is most concentrated?



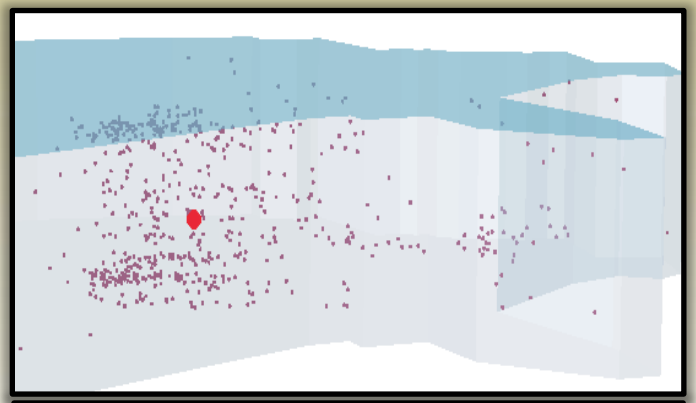


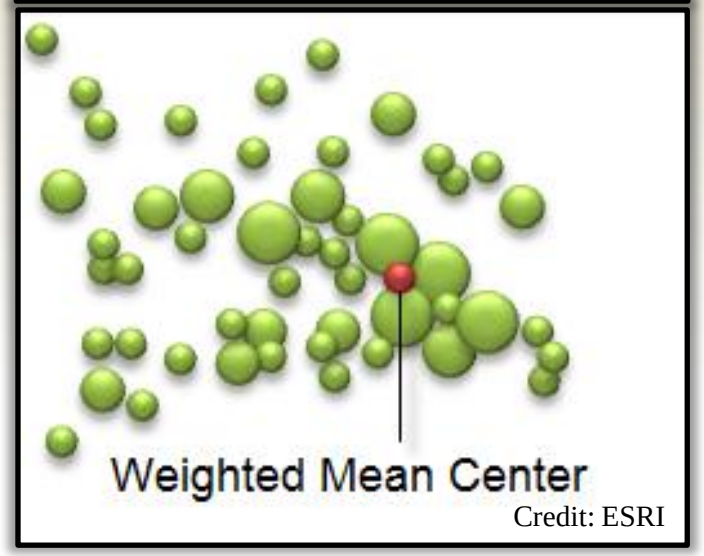
Measuring Geographic Distribution: Finding the Center

The **Mean Center** tool computes the average x and y coordinate, based on all features in the study area



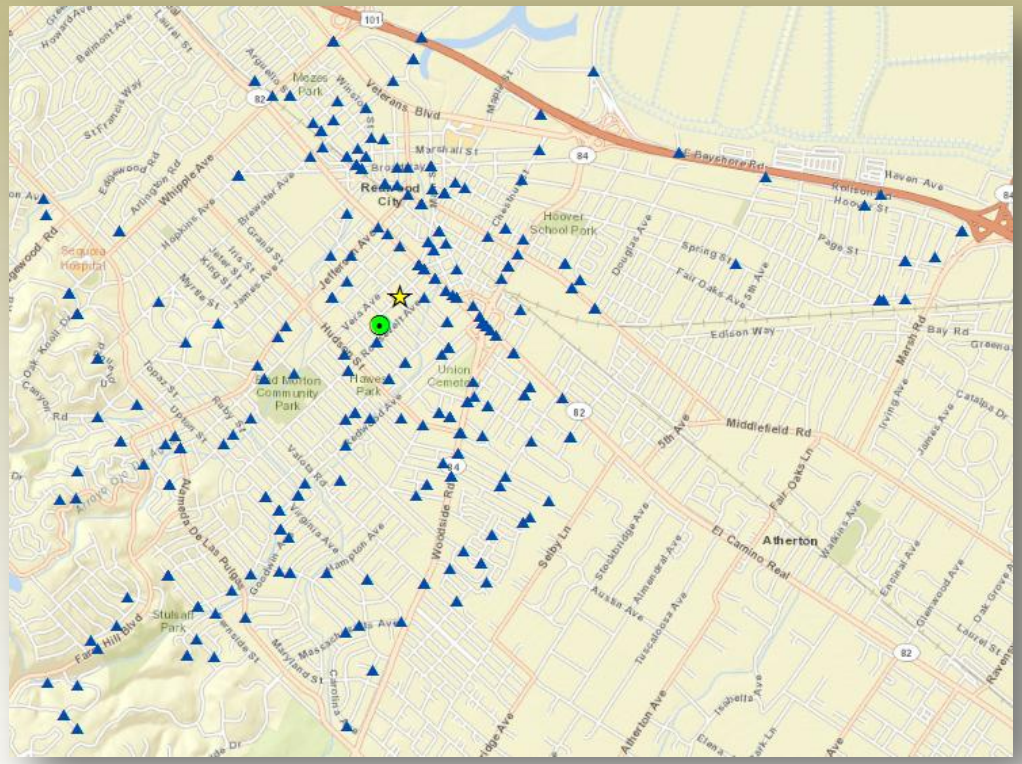
$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$





Demo: Crime events in Redwood City, June – September 2012

Mean and Median Center for all Crimes



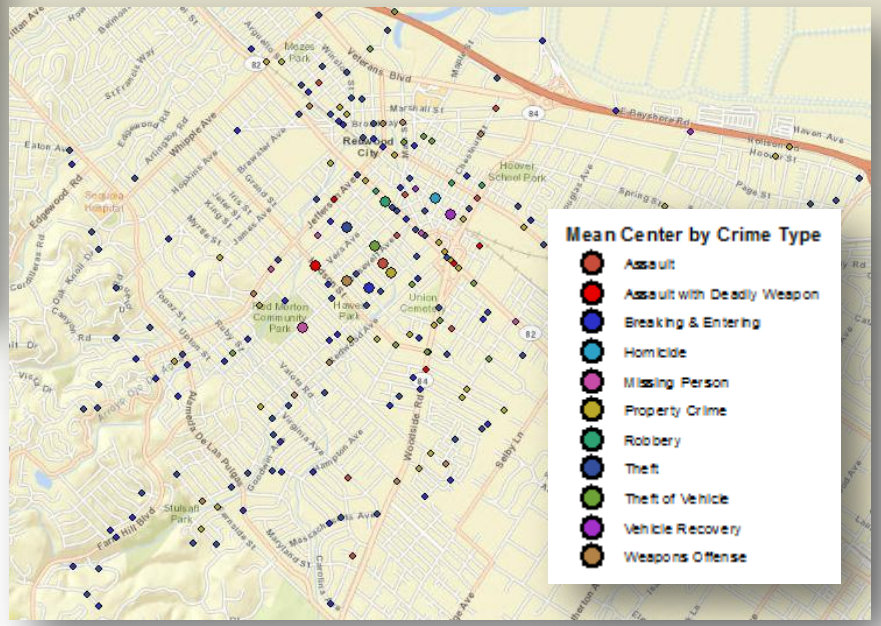
Mean Center = geographic center

Median Center = location that minimizes Euclidean distance

Weight Field: gives more importance to some features

Case Field: groups features for separate calculations

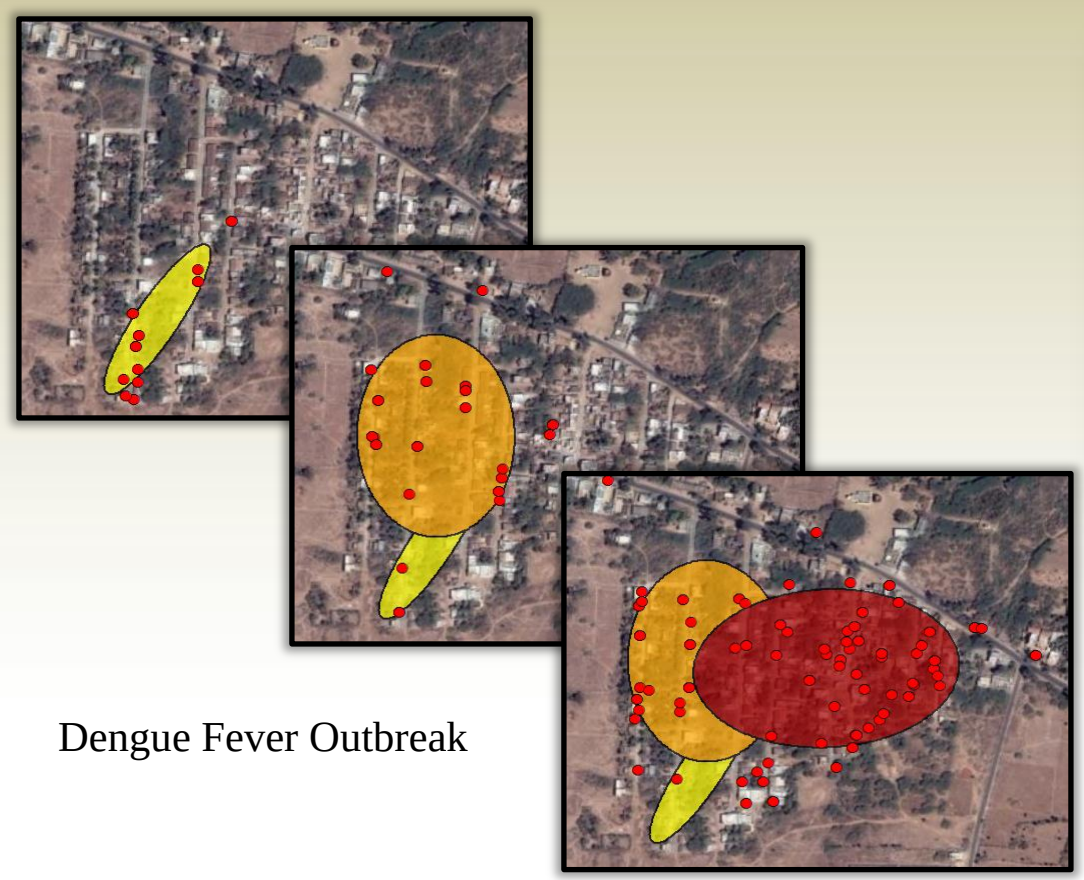
Attribute Field: numeric field to be calculated



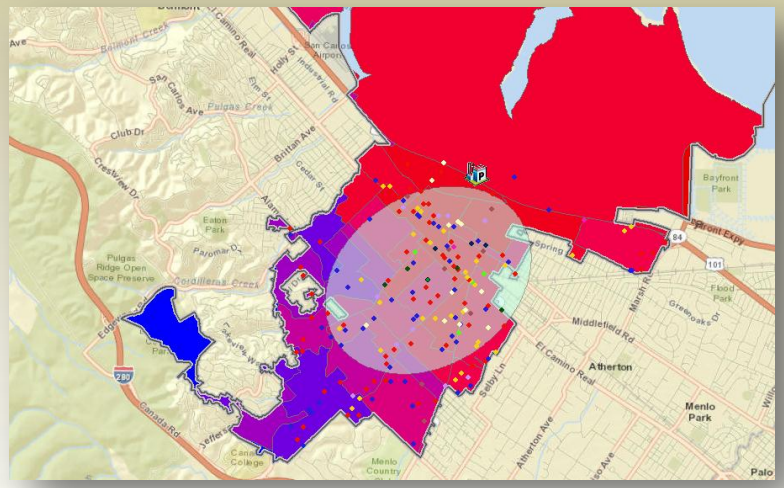
Mean by crime type

Measuring Distribution and Direction: Standard Deviational Ellipse

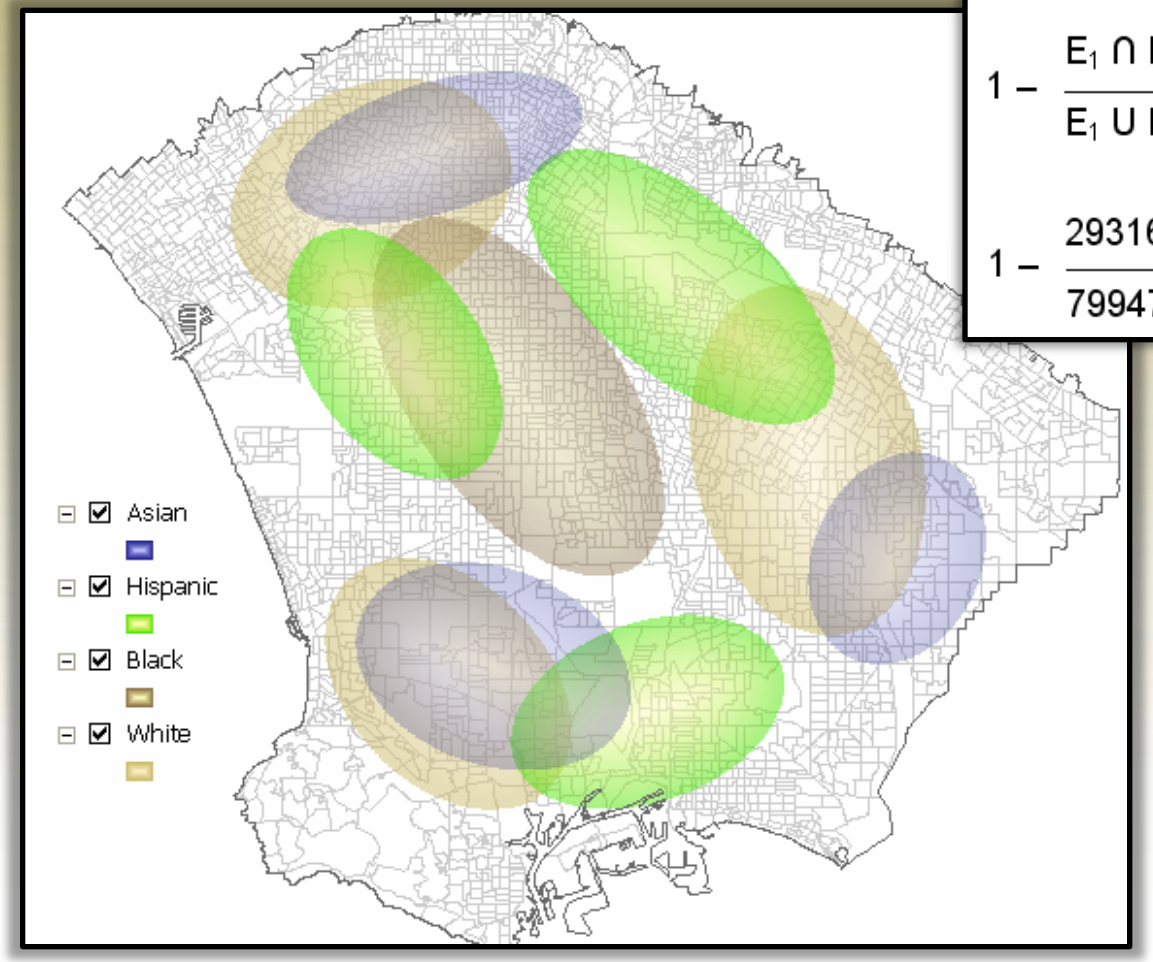
- Central tendency, orientation & dispersion
- Abstracting spatial trends in a distribution of features
- Comparing distributions over time



Dengue Fever Outbreak



Measuring Distribution and Direction: Standard Deviational Ellipse



Segregation Index =

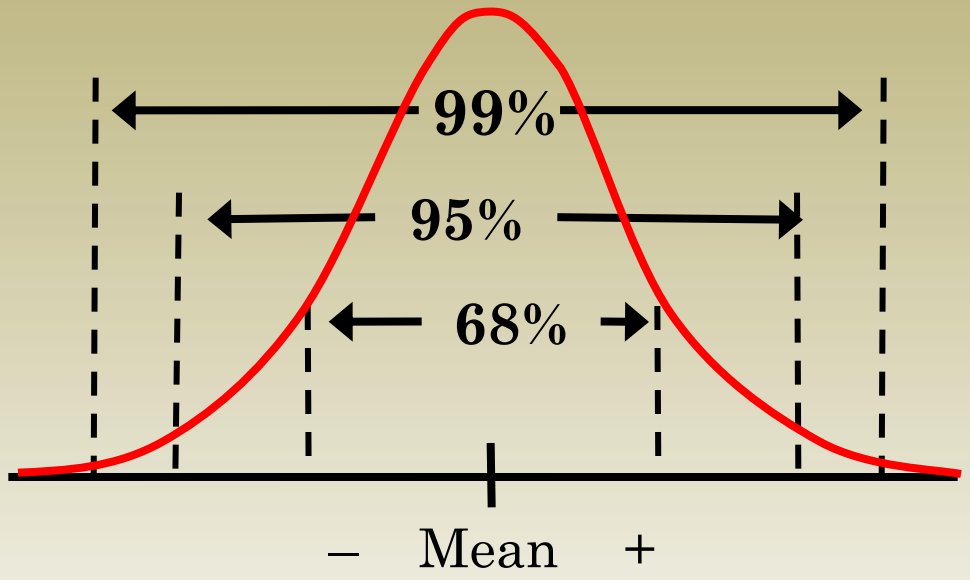
$$1 - \frac{E_1 \cap E_2 \cap E_3 \cap \ldots E_n}{E_1 \cup E_2 \cup E_3 \cup \ldots E_n} =$$

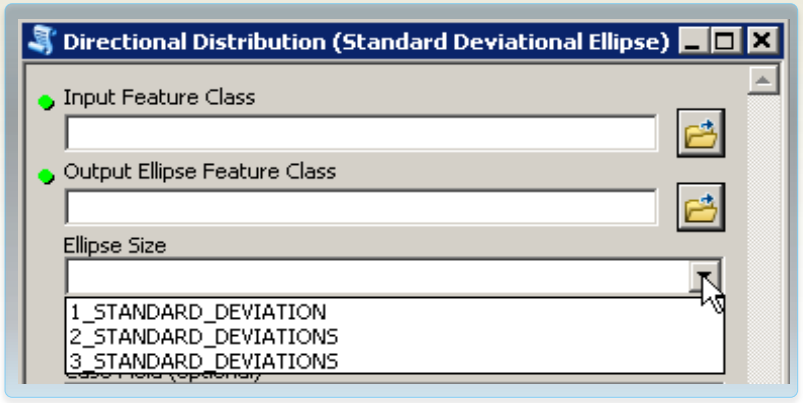
$$1 - \frac{2931680545.83}{7994760004.92} = \mathbf{0.63}$$

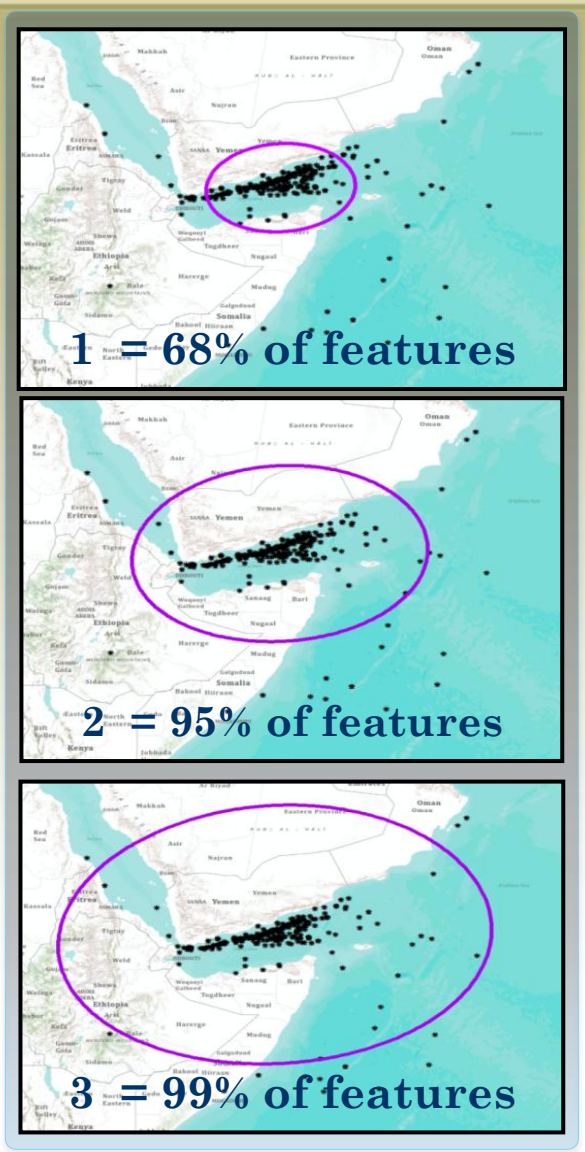
Credit: ESRI

Directional distribution

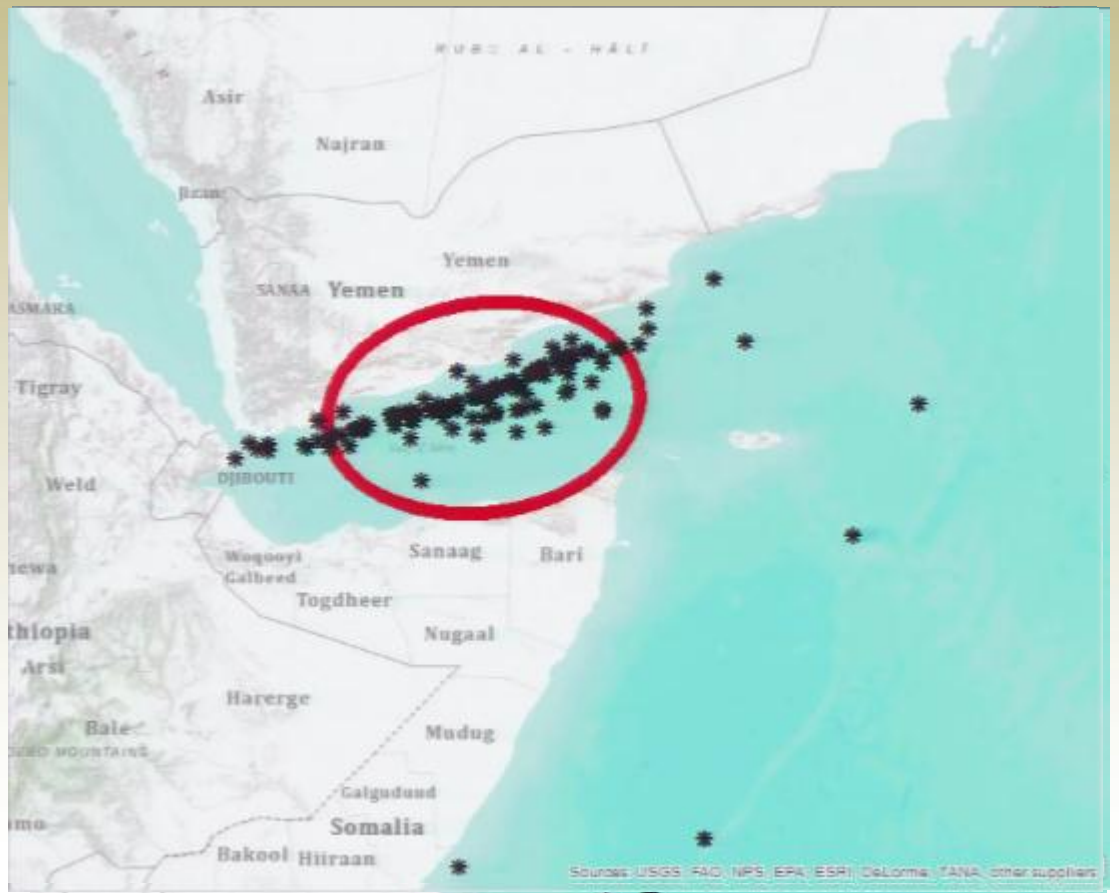
Normal distribution







Spatial Distribution of Piracy



Pirate attacks : March 2007 – August 2007

Pirate attacks : Sept 2007 – February 2008

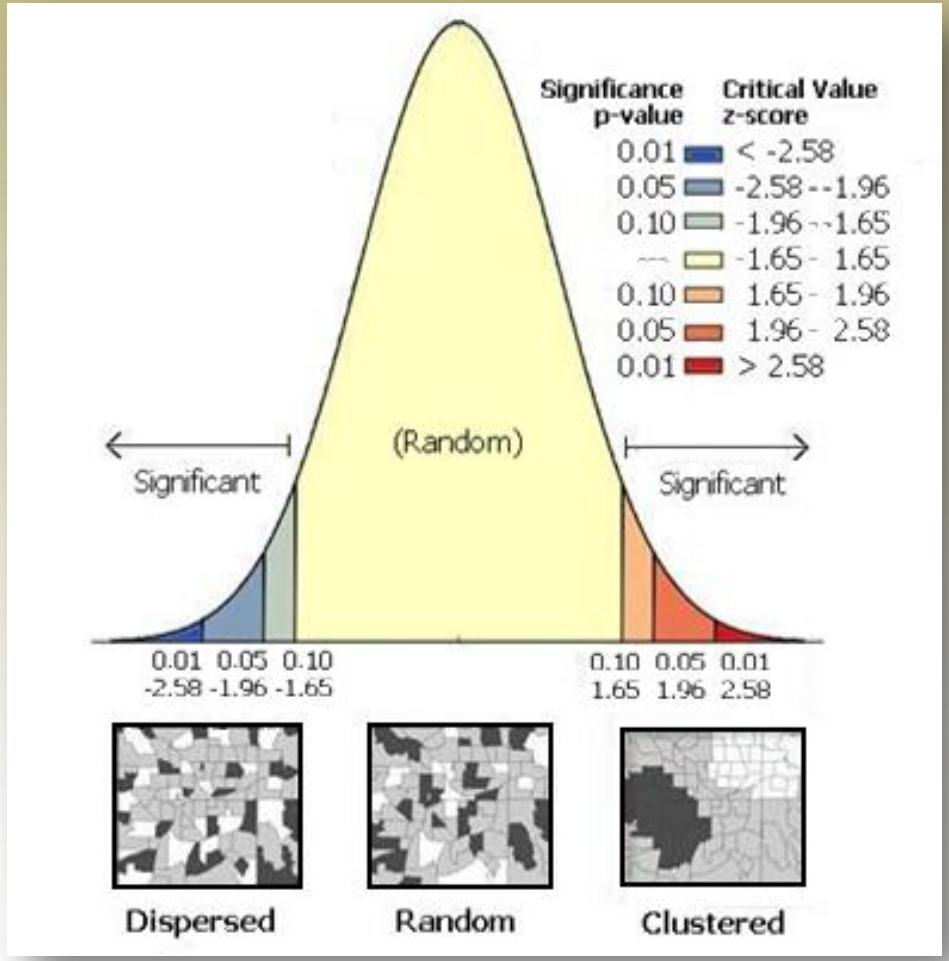
Pirate attacks : March 2008 – August 2008

Pirate attacks : Sept 2008 – February 2009

Inferential Statistics

- Start with a null hypothesis
  - 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

E.g. Sudden Infant Death Syndrome (SIDS)



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

- Reject null hypothesis:

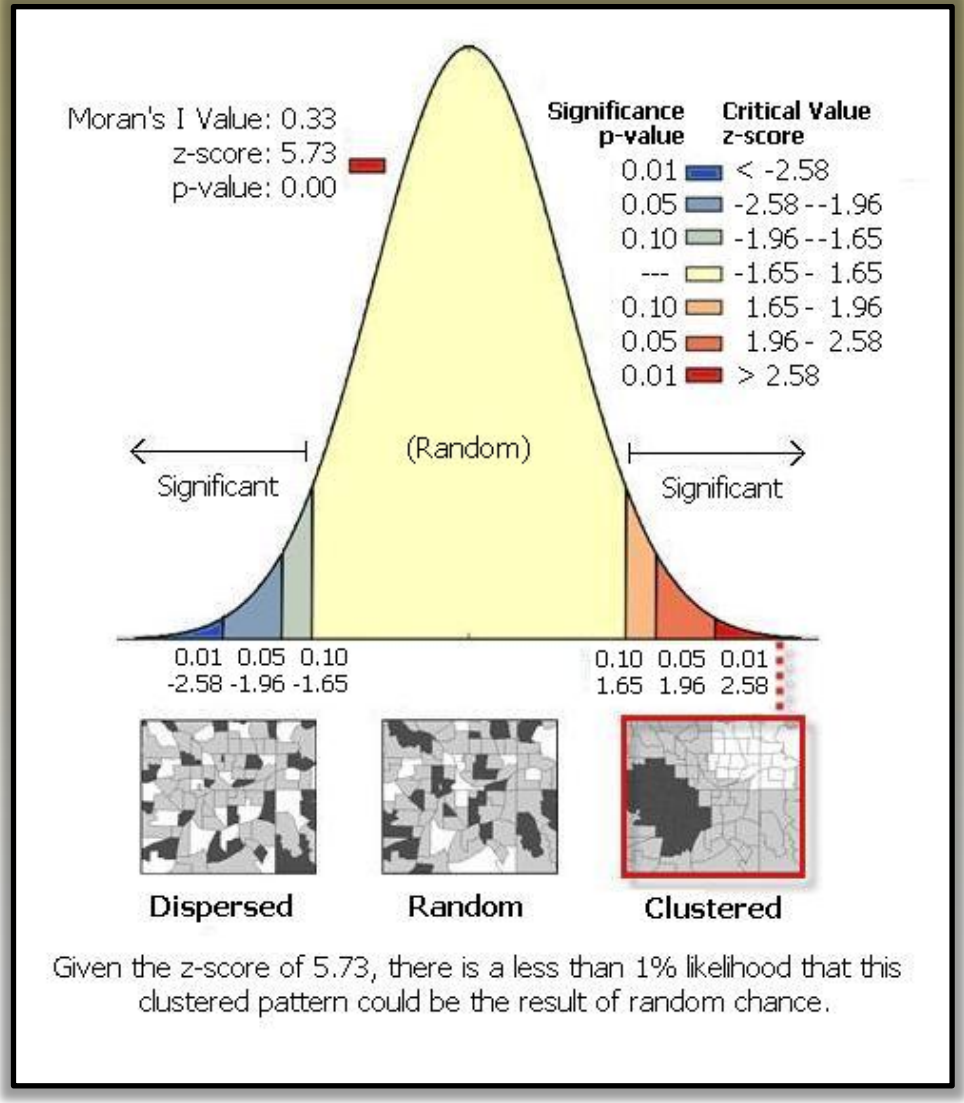
  No random pattern

- P-values are probabilities

  Small p-values, like 0.01, mean it is UNLIKELY the pattern is random

- Z-scores are standard deviations and can be mapped to specific p-values

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



Inferential statistics tools: types

- **Global Spatial Autocorrelation Statistics:**
  Analyzing **broad spatial patterns**
  How intense is the clustering?

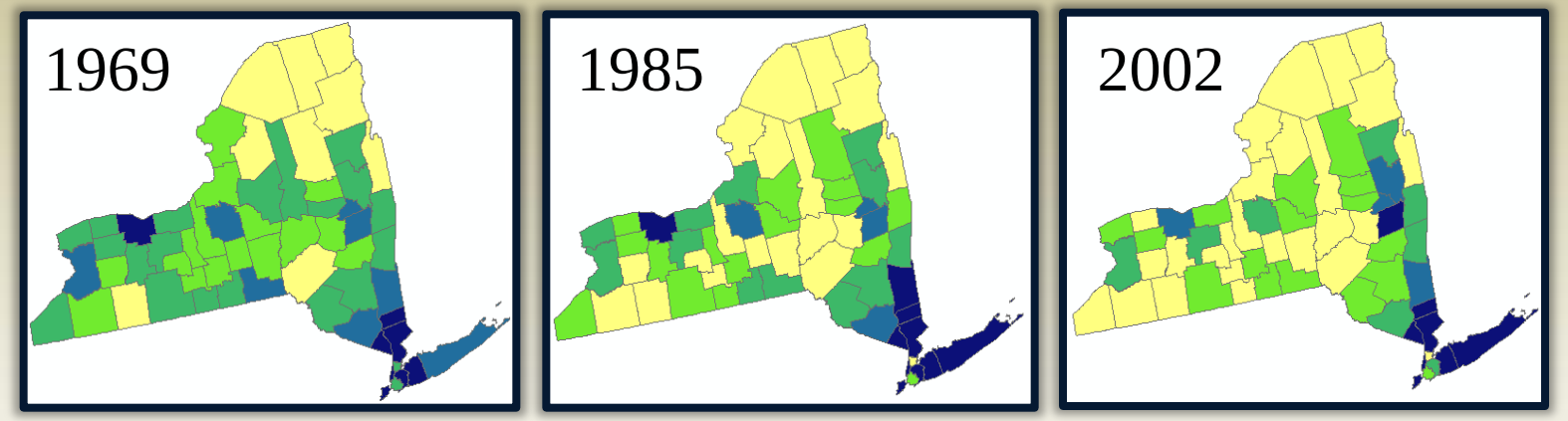
  Spatial Autocorrelation: Moran's I

- **Local Spatial Autocorrelation Statistics:**
  Mapping **clusters, outliers**
  Where is the clustering?

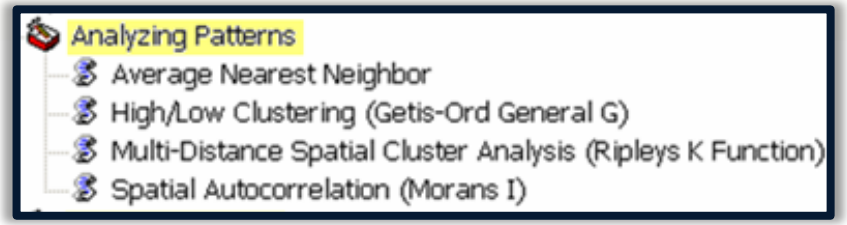
  Hot Spot Analysis: Getis-Ord G

Global Spatial Autocorrelation Tools: Intensity of Clustering

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



- Which crime type is most concentrated?
- Does the spatial pattern of the disease mirror the spatial pattern of the population at risk?

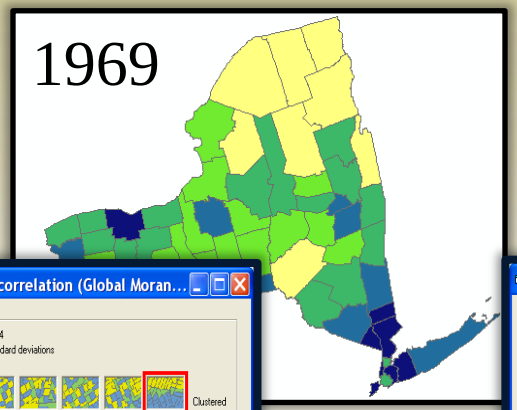


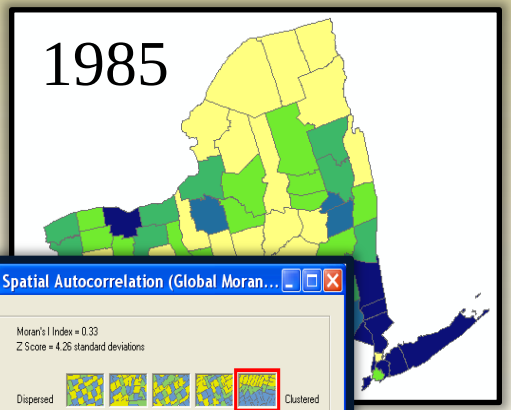
- Is there an unexpected spike in foreclosure rates?
- Are new AIDs cases remaining Geographically fixed or are they spreading to nearby counties?

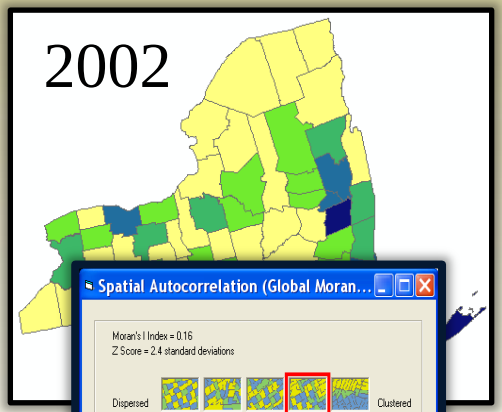
Global Spatial Autocorrelation Tools: Spatial Autocorrelation

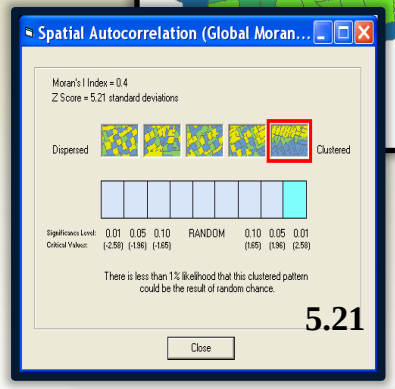
Spatial Statistics Toolbox

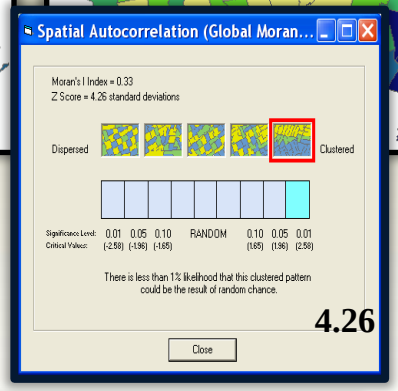
Measuring Geographic Distribution | **Analyzing Patterns** | **Mapping Clusters** | **Modeling Spatial Relationships**

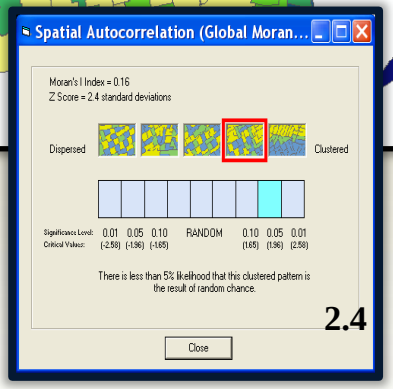


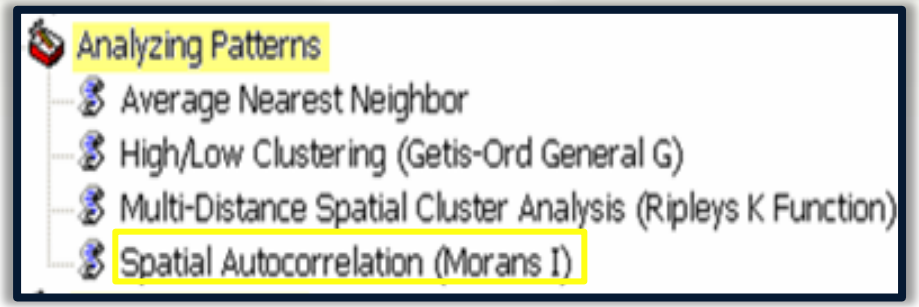


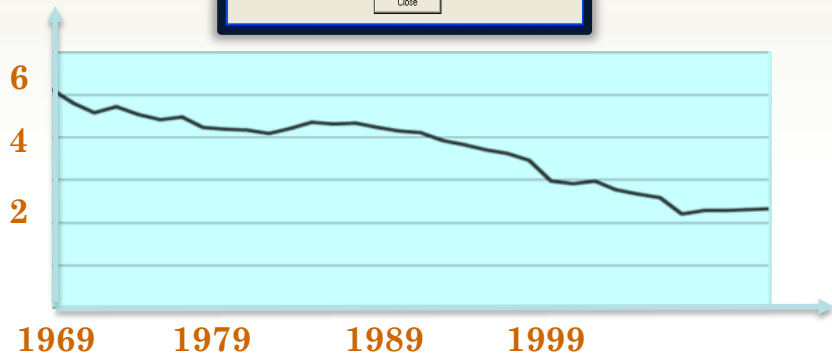






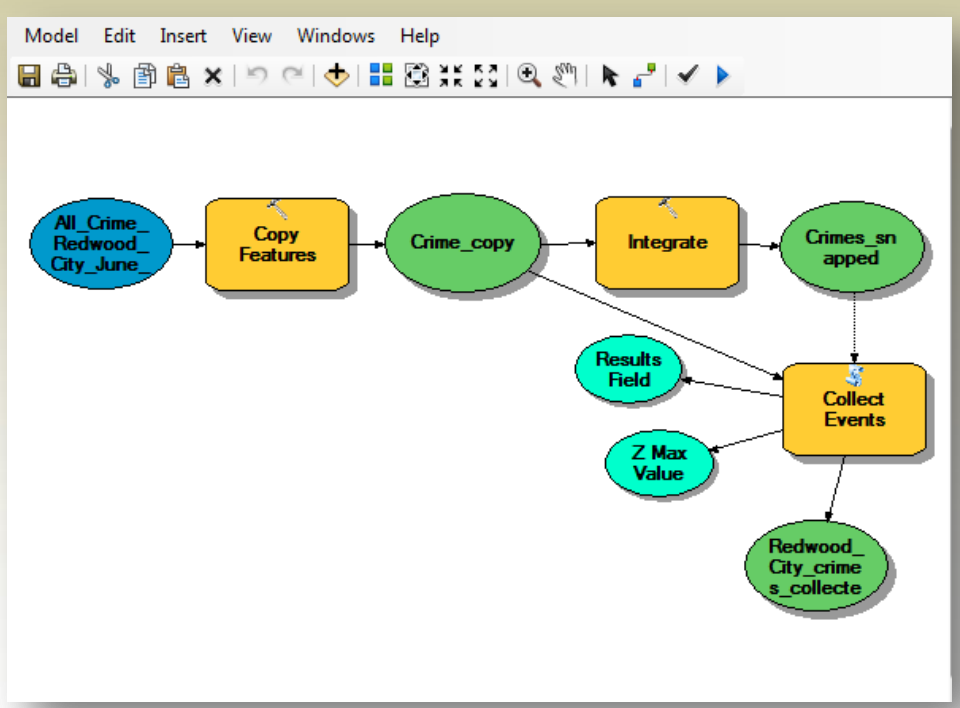




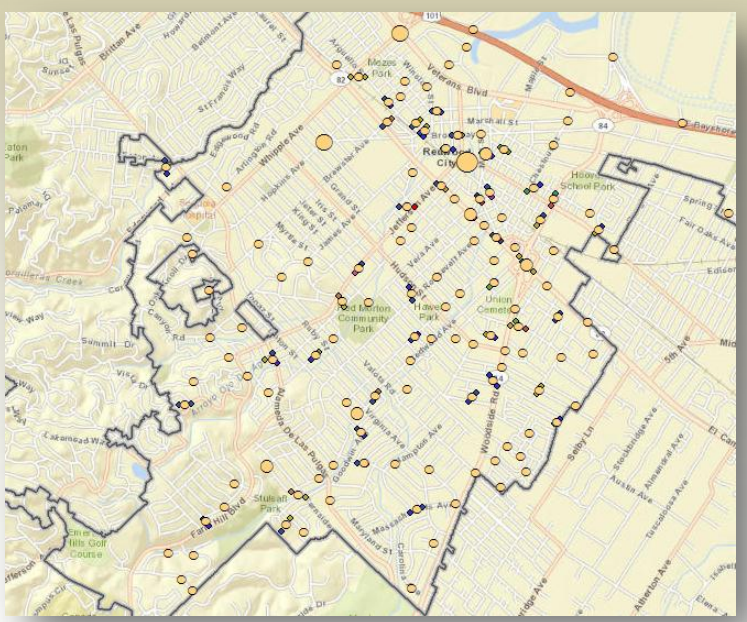


Exercise: Are crime locations random chance? Spatial Autocorrelation

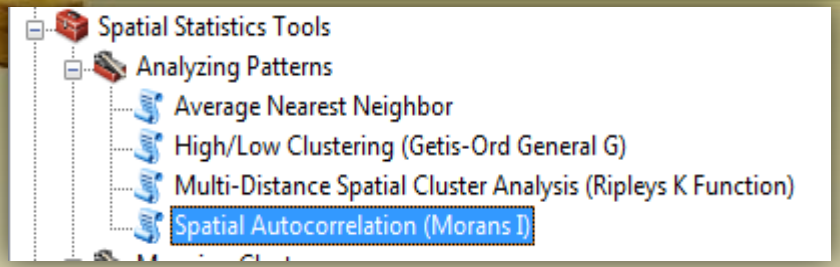
Moran's I tool measures spatial autocorrelation based on both feature locations and feature values simultaneously.
Crimes are events that do not have feature values.
To go from events to values we used the **Integrate** and **Collect** tools.

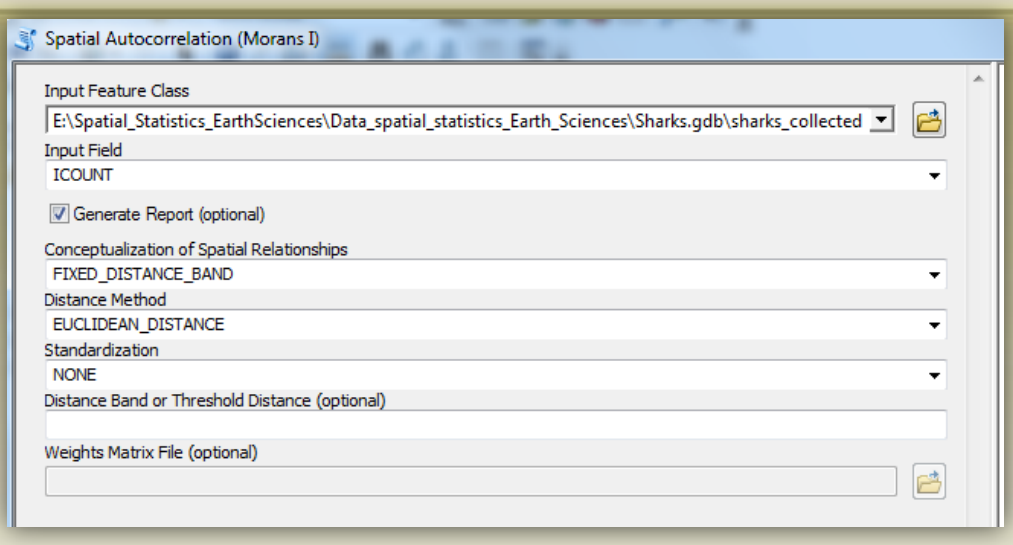


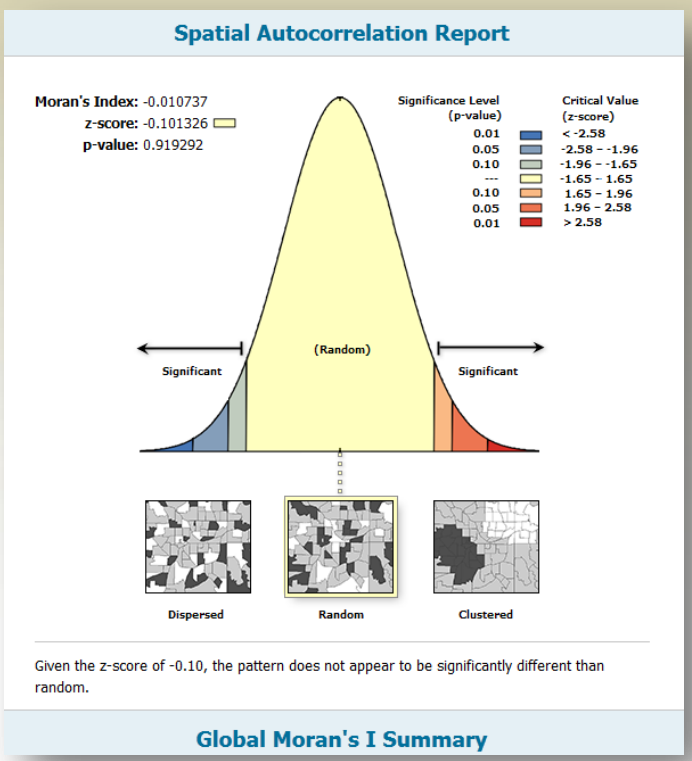
Model Builder example



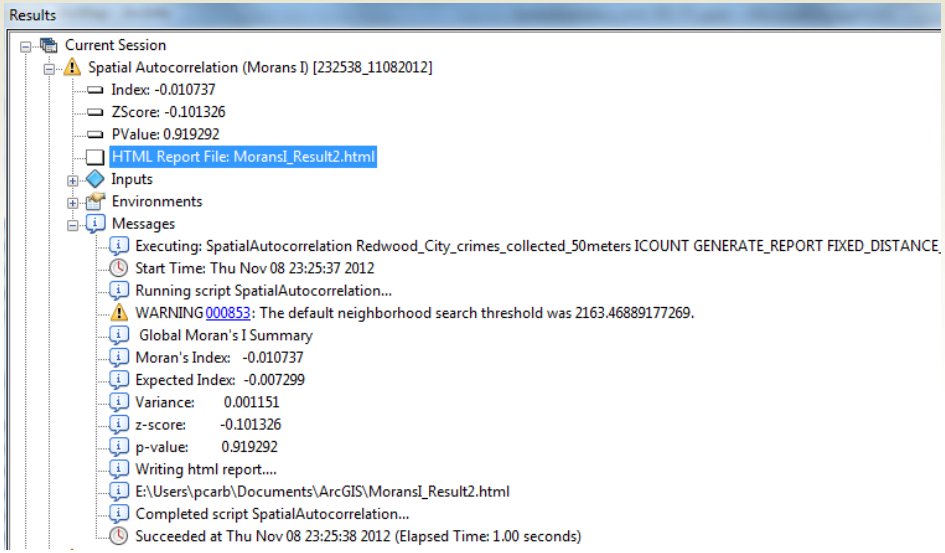
Exercise: Are crime locations random chance?
Spatial Autocorrelation







Spatial Autocorrelation (Morans I) settings

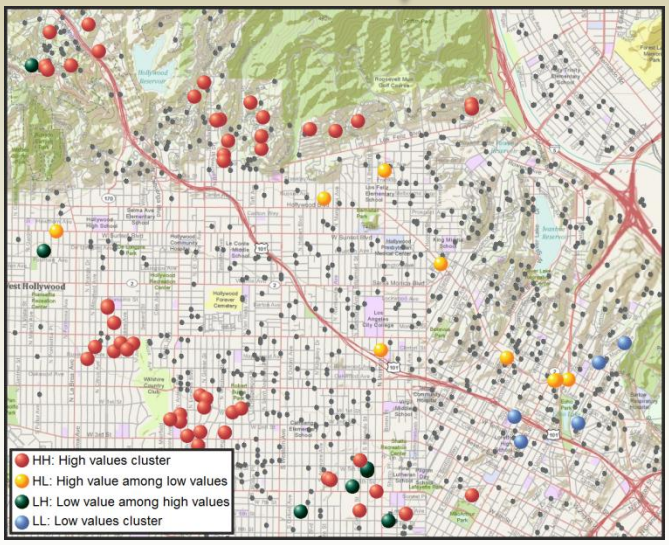


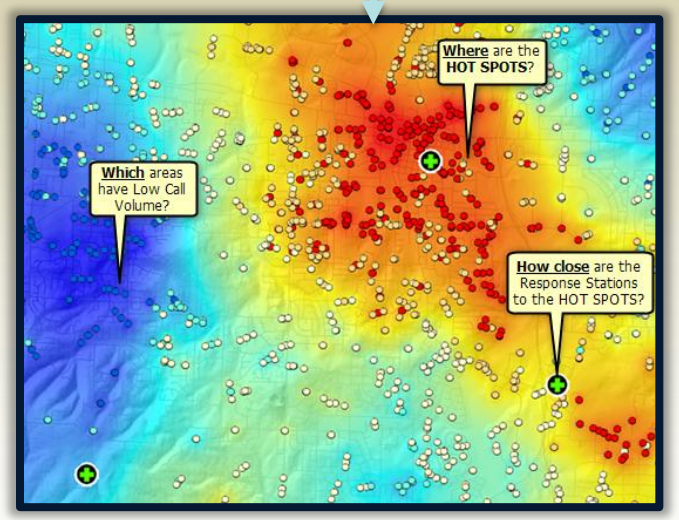
Local Spatial Autocorrelation Tools: Hot Spots & Outliers

Spatial Statistics Toolbox

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

Mapping Clusters
- Cluster and Outlier Analysis (Anselin Local Morans I)
- Hot Spot Analysis (Getis-Ord Gi*)





- Which houses sold for much more than expected?
- Where do we find anomalous spending patterns?

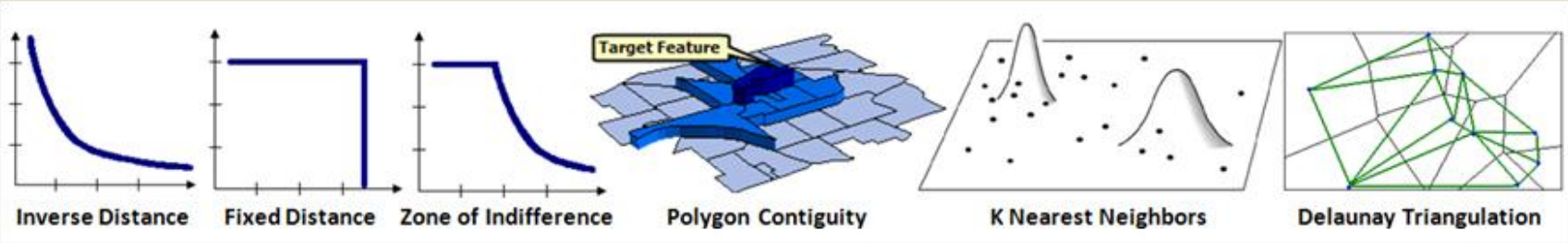
- Where are the 911 call, crime, arson, disease… hot spots?
- Where do we see unexpectedly high rates of urban growth?

Local Spatial Autocorrelation Tools: Hot Spot Analysis

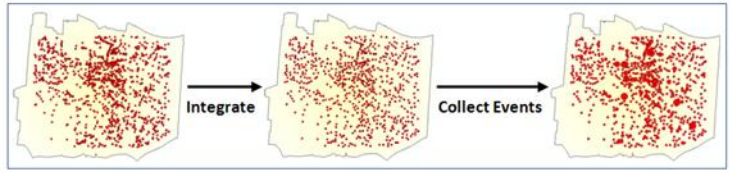
Hot Spot Analysis:

Spatial cluster detection method which identifies **statistically significant spatial concentrations** of the high and of low values associated with a set of geographic features.

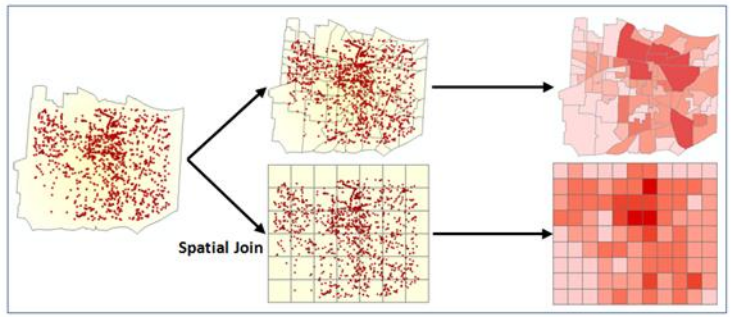
- **Never accept the defaults! 3 Things to Consider:**
  - **Conceptualization of Spatial Relationships** (what constitutes to be a neighbor?)



- **Analysis Weight Field**
  - Integrate & Collect for points

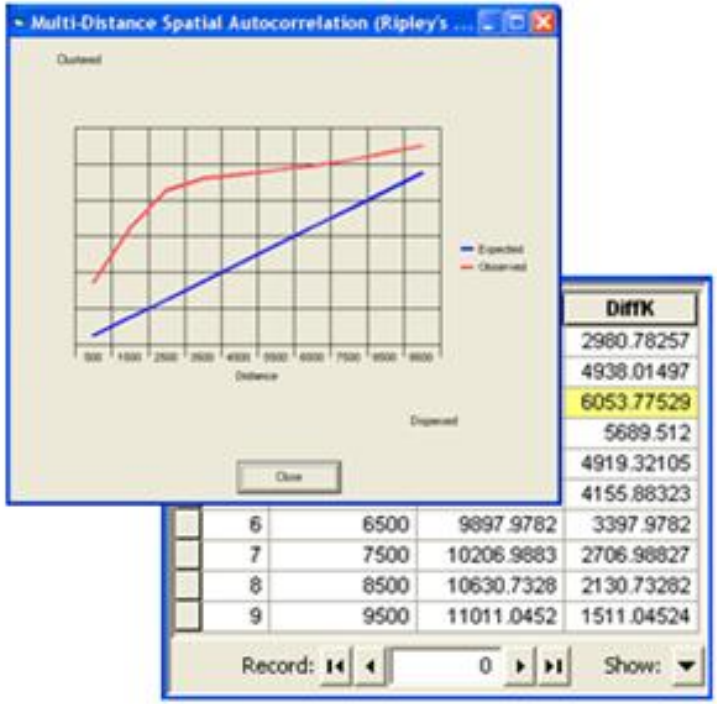


- **Scale of analysis or distance**
  - Geographic extent of the spatial process
  - At least one neighbor
  - Distance band that reflects maximum spatial autocorrelation

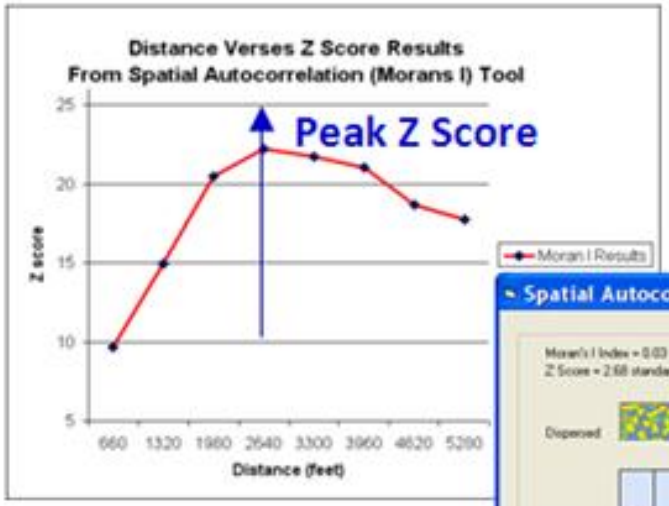


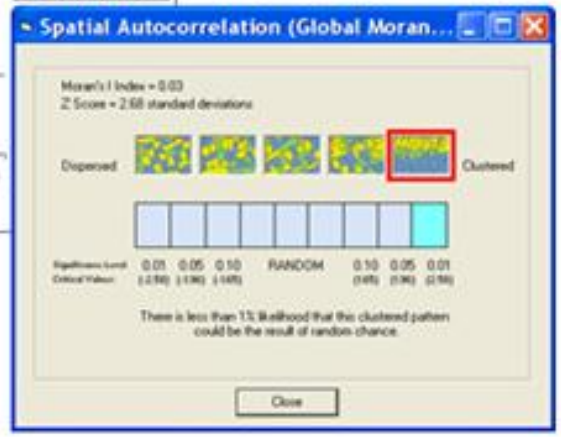
Distance band that reflects maximum spatial autocorrelation

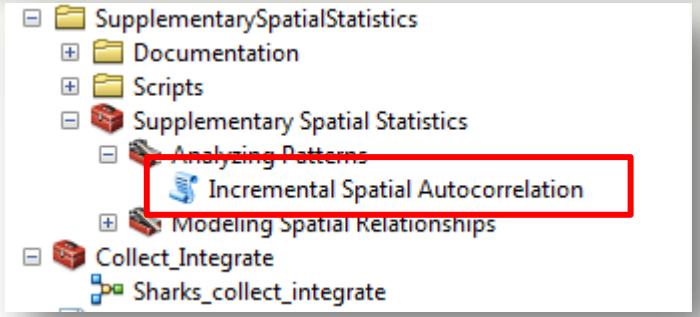
Multi-Scale Exploratory Analysis



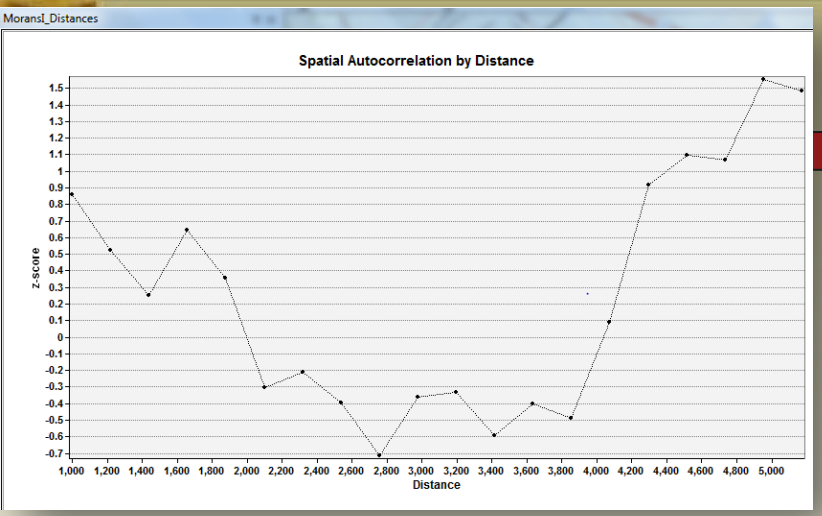
Global Morans I

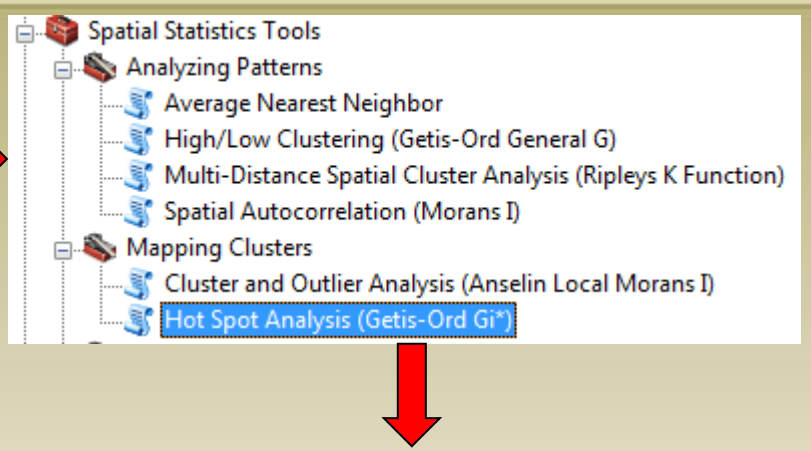


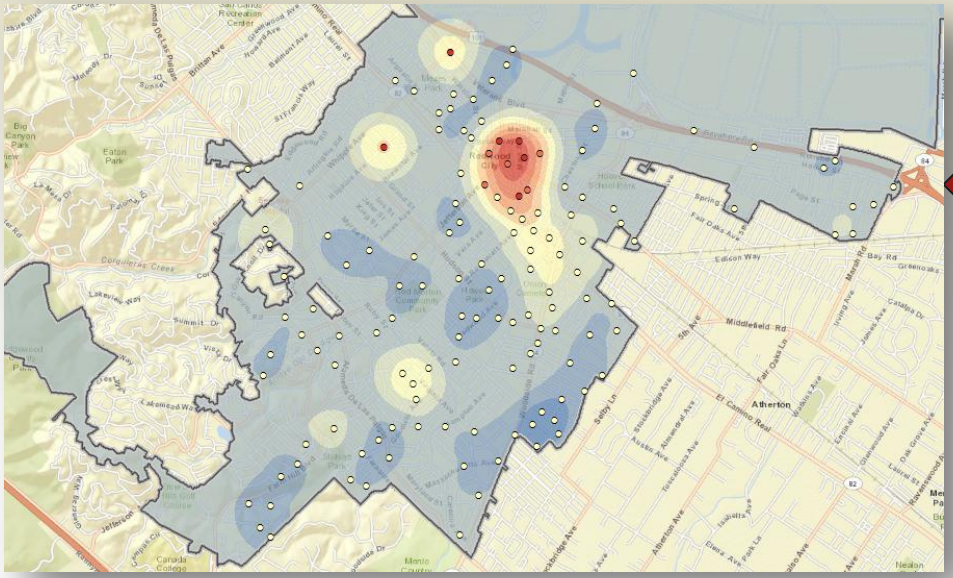


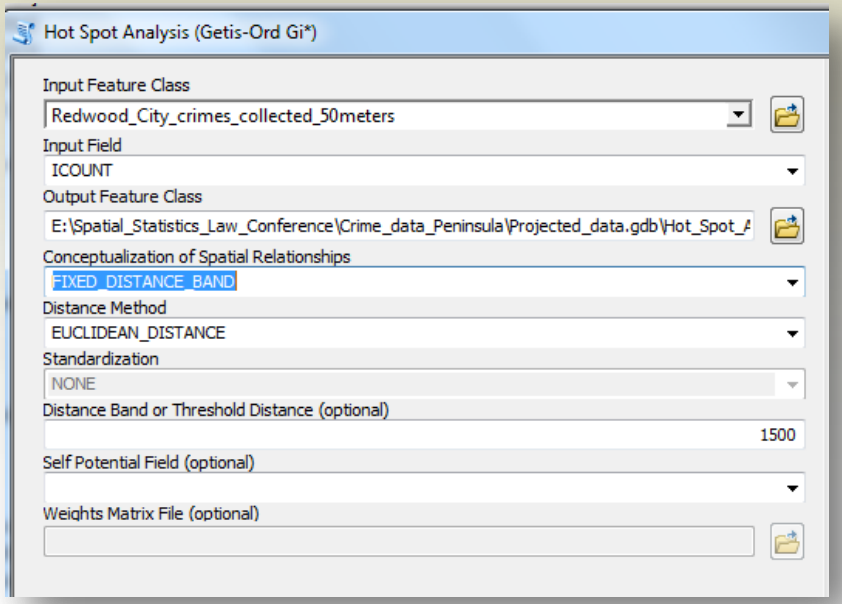


Exercise: Hot Spot Analysis





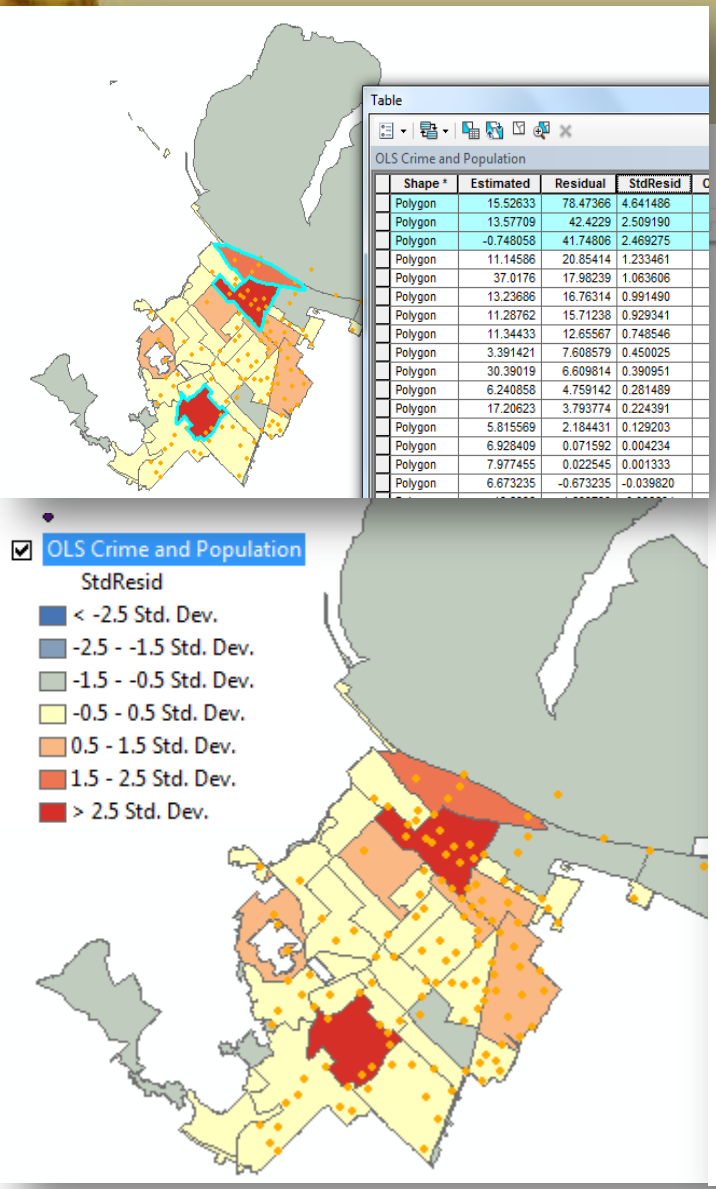




Regression Analysis:

How much is crime explained by population density?

Ordinary Least Squares



Ordinary Least Squares [065759_11092012]
- Output Feature Class: OLS_pop
- Coefficient Output Table: <empty>
- Diagnostic Output Table: <empty>
- Inputs
- Environments
- Messages

Executing: OrdinaryLeastSquares block_groups_with_Crime_projected OBJECTID E:\Spatial_Statistics_Law_Conference\Crir
Start Time: Fri Nov 09 06:40:57 2012
Running script OrdinaryLeastSquares...

Summary of OLS Results

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr
Intercept	-1.712047	4.757535	-0.359860	0.720501	3.970287	-0.431215	0.668205
TOTPOP_CY	0.007088	0.002158	3.284786	0.001890*	0.001933	3.666909	0.000605*

OLS Diagnostics

Input Features:	block_groups_with_Crime_projected	Dependent Variable:	COUNT_
Number of Observations:	51	Akaike's Information Criterion (AICc) [2]:	437.630402
Multiple R-Squared [2]:	0.180462	Adjusted R-Squared [2]:	0.163737
Joint F-Statistic [3]:	10.789819	Prob(>F), (1,49) degrees of freedom:	0.001889*
Joint Wald Statistic [4]:	13.446219	Prob(>chi-squared), (1) degrees of freedom:	0.000246*
Koenker (BP) Statistic [5]:	0.145684	Prob(>chi-squared), (1) degrees of freedom:	0.702695
Jarque-Bera Statistic [6]:	223.772848	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* Statistically significant at the 0.05 level.
[1] Large VIF (> 7.5, for example) indicates explanatory variable redundancy.
[2] Measure of model fit/performance.
[3] Significant p-value indicates overall model significance.
[4] Significant p-value indicates robust overall model significance.
[5] Significant p-value indicates biased standard errors; use robust estimates.
[6] Significant p-value indicates residuals deviate from a normal distribution.

WARNING 000851: Use the Spatial Autocorrelation (Moran's I) Tool to ensure residuals are not spatially autocorrelated.

Gender Equity & Inheritance Reform
Evidence from Rural India

What is gender-equalizing land inheritance reform's impact?
Does it improve gender equity in land inheritance's distribution?

Legal Context

Hindu Successi

Daughters gain inde
land, on par with

State Amendments
19

Growing women's e
ins

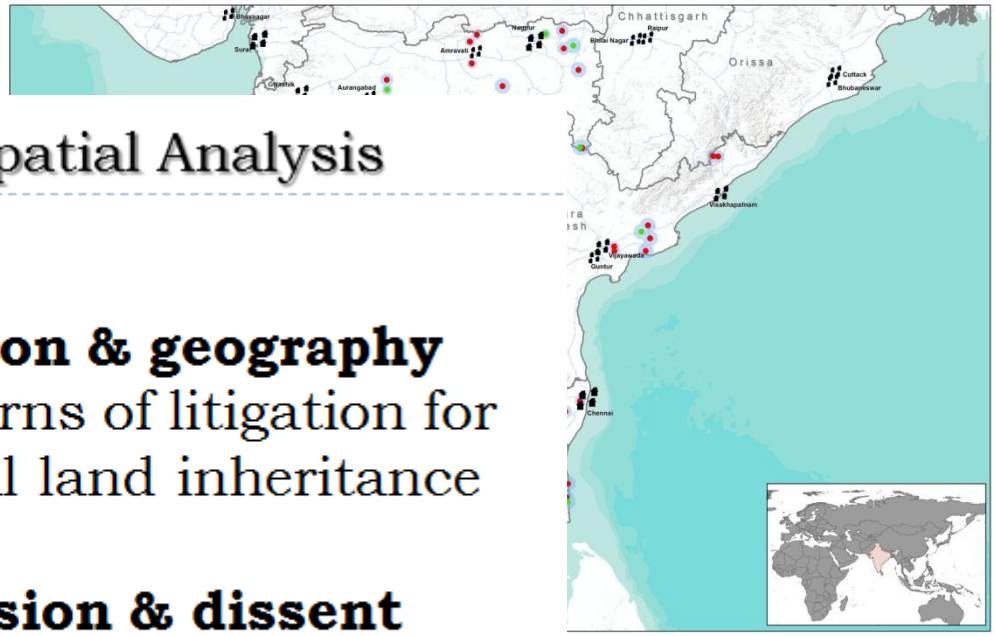
Future Geospatial Analysis

Legal diffusion & geography

- Spatial patterns of litigation for gender-equal land inheritance

Legal diffusion & dissent

- Correlation between geography of litigation & organized social dissent



Rachel Brulé, PhD Candidate in Political Science, Stanford University

Spatial Statistic Software & Tutorials

- Proprietary
  - ArcGIS
  - Matlab
- Open Source
  - R
  - GRASS
  - OpenGeoda
- Tutorials & Other Resources
  - http://blogs.esri.com/esri/arcgis/2010/07/13/spatial-statistics-resources/
  - http://www.ai-geostats.org/
  - http://www.sal.uiuc.edu/
  - http://www.spatial-statistics.com/software_index.htm

Special Thanks to…

- **All of you** for your patience!
- Daniel E. Ho

  Law Professor, Stanford University
- Rachel Brule

  Ph.D. Candidate, Political Science, Stanford University
- Matt Marostica

  Subject Specialist in Economics and Political Science, Stanford University
- Doug Spencer

  Ph.D. Candidate in Jurisprudence and Social Policy, University of California, Berkeley
- Michael Dale

  Post-Doctoral Scholar, GCEP, Stanford University

gis.stanford.edu