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: imsocio.org

Credit: google.com

Credit: Hint.fm

MAP CROWD-SOURCING
Geographic Information Systems in everyday life

Credit: SSSL & ESRI

Credit: New York Times

Credit: electoral-vote.com

Credit: NPR
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).
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)."
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**

- The raster view of the world
- Happy Valley spatial entities: Points, lines, areas, networks, surfaces

**Vector**

- The vector view of the world
- Points, lines, areas, networks, surfaces
Location, location, location!!
Earth Reference Systems

Geographic Coordinate Systems
Projected Coordinate Systems

Latitude, Longitude: Mercator, UTM, State Plane, Albers Equal Area, Equidistant

Always ask for the Datum!
The Importance of Projections
The Importance of Projections - cont
Always work in Projected Coordinate Systems!

Which Projection?
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

<table>
<thead>
<tr>
<th>Extent</th>
<th>World</th>
<th>Hemisphere</th>
<th>Continent or Ocean</th>
<th>Region or Sea</th>
<th>Country</th>
<th>Locality</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extent of a single continent</td>
<td>Extent of a country</td>
<td>Extent of an ocean</td>
<td>Extent of a sea</td>
<td>Extent of a region</td>
<td>Extent of a locality</td>
<td>Conformal</td>
</tr>
<tr>
<td></td>
<td>Albers Equal Area Conic</td>
<td>Albers Equal Area Conic</td>
<td>Albers Equal Area Conic</td>
<td>Albers Equal Area Conic</td>
<td>Albers Equal Area Conic</td>
<td>Albers Equal Area Conic</td>
<td>Equal Area</td>
</tr>
<tr>
<td></td>
<td>Equidistant Conic</td>
<td>Equidistant Conic</td>
<td>Equidistant Conic</td>
<td>Equidistant Conic</td>
<td>Equidistant Conic</td>
<td>Equidistant Conic</td>
<td>Equidistant</td>
</tr>
<tr>
<td></td>
<td>Azimuthal Equal Area Conic</td>
<td>Azimuthal Equal Area Conic</td>
<td>Azimuthal Equal Area Conic</td>
<td>Azimuthal Equal Area Conic</td>
<td>Azimuthal Equal Area Conic</td>
<td>Azimuthal Equal Area Conic</td>
<td>True Direction</td>
</tr>
<tr>
<td></td>
<td>Conformal</td>
<td>Conformal</td>
<td>Conformal</td>
<td>Conformal</td>
<td>Conformal</td>
<td>Conformal</td>
<td>Compromise</td>
</tr>
<tr>
<td></td>
<td>Equal Area</td>
<td>Equal Area</td>
<td>Equal Area</td>
<td>Equal Area</td>
<td>Equal Area</td>
<td>Equal Area</td>
<td>Straight Rhumbs</td>
</tr>
<tr>
<td></td>
<td>Equidistant</td>
<td>Equidistant</td>
<td>Equidistant</td>
<td>Equidistant</td>
<td>Equidistant</td>
<td>Equidistant</td>
<td>Perspective</td>
</tr>
<tr>
<td></td>
<td>True Direction</td>
<td>True Direction</td>
<td>True Direction</td>
<td>True Direction</td>
<td>True Direction</td>
<td>True Direction</td>
<td>Suitable Orientation</td>
</tr>
<tr>
<td></td>
<td>Compromise</td>
<td>Compromise</td>
<td>Compromise</td>
<td>Compromise</td>
<td>Compromise</td>
<td>Compromise</td>
<td>or Latitude</td>
</tr>
<tr>
<td></td>
<td>Straight Rhumbs</td>
<td>Straight Rhumbs</td>
<td>Straight Rhumbs</td>
<td>Straight Rhumbs</td>
<td>Straight Rhumbs</td>
<td>Straight Rhumbs</td>
<td>North-South</td>
</tr>
<tr>
<td></td>
<td>Perspective</td>
<td>Perspective</td>
<td>Perspective</td>
<td>Perspective</td>
<td>Perspective</td>
<td>Perspective</td>
<td>East-West</td>
</tr>
<tr>
<td></td>
<td>Suitable Orientation</td>
<td>Suitable Orientation</td>
<td>Suitable Orientation</td>
<td>Suitable Orientation</td>
<td>Suitable Orientation</td>
<td>Suitable Orientation</td>
<td>Oblique</td>
</tr>
<tr>
<td></td>
<td>or Latitude</td>
<td>or Latitude</td>
<td>or Latitude</td>
<td>or Latitude</td>
<td>or Latitude</td>
<td>or Latitude</td>
<td>Equatorial</td>
</tr>
<tr>
<td></td>
<td>North-South</td>
<td>North-South</td>
<td>North-South</td>
<td>North-South</td>
<td>North-South</td>
<td>North-South</td>
<td>Middle Latitudes</td>
</tr>
<tr>
<td></td>
<td>East-West</td>
<td>East-West</td>
<td>East-West</td>
<td>East-West</td>
<td>East-West</td>
<td>East-West</td>
<td>Polar/Circular</td>
</tr>
<tr>
<td></td>
<td>Oblique</td>
<td>Oblique</td>
<td>Oblique</td>
<td>Oblique</td>
<td>Oblique</td>
<td>Oblique</td>
<td></td>
</tr>
</tbody>
</table>
# GIS Functions (I): Data Gathering

Credit: [www.crimereports.com](http://www.crimereports.com)

## Crime Reports

### Crime List 2012.10.25 00:00:00 - 2012.11.08 23:59:59

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Date/Time</th>
<th>Address</th>
<th>Identifier</th>
<th>Desc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breaking &amp; Entering</td>
<td>10/25/2012 00:00:00</td>
<td>500 Block SAPPHIRE ST</td>
<td>112100591</td>
<td>Burgl</td>
</tr>
<tr>
<td>Theft</td>
<td>10/25/2012 00:00:00</td>
<td>600 Block PINE ST</td>
<td>112100613</td>
<td>Gmd</td>
</tr>
<tr>
<td>Breaking &amp; Entering</td>
<td>10/25/2012 03:00:00</td>
<td>500 Block LANCASTER WY</td>
<td>112100586</td>
<td>Burgl</td>
</tr>
<tr>
<td>Breaking &amp; Entering</td>
<td>10/25/2012 12:30:00</td>
<td>1100 Block CLEVELAND ST</td>
<td>112100614</td>
<td>Burgl</td>
</tr>
<tr>
<td>Breaking &amp; Entering</td>
<td>10/25/2012 15:30:00</td>
<td>1100 Block VALOTA RD</td>
<td>112100603</td>
<td>Burgl</td>
</tr>
</tbody>
</table>

Credit: [www.crimereports.com](http://www.crimereports.com)
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

Which parcels are within 60m of the road?

Intersection

Hot Spot Analysis

Average Nearest Neighbor

Dispersed → → → → Clusters
GIS Functions (III): Visualization

**Greed**
Average income compared with number of people living below the poverty line.

**Envy**
Total thefts (robbery, burglary, larceny, and grand theft auto) per capita.

**Wrath**
Number of violent crimes (murder, assault, and rape) per capita.

**Sloth**
Expenditures on art, entertainment, and recreation compared with employment.

**Gluttony**
Number of fast-food restaurants per capita.

**Lust**
Number of STD cases reported per capita.

**Pride**
Aggregate of the other six offenses—because pride is the root of all sin.
Demonstration

The Ambiguity of Map Symbolization
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

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

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

[Map showing crime events in Redwood City, June – September 2012]
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 =

\[
1 - \frac{E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_n}{E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n} = \frac{2931680545.83}{7994760004.92} = 0.63
\]
Directional distribution

Normal distribution

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

<table>
<thead>
<tr>
<th>Z Score (Standard Deviations)</th>
<th>P-Value (Probability)</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/-1.65</td>
<td>0.10</td>
<td>90%</td>
</tr>
<tr>
<td>+/-1.96</td>
<td>0.05</td>
<td>95%</td>
</tr>
<tr>
<td>+/-2.58</td>
<td>0.01</td>
<td>99%</td>
</tr>
</tbody>
</table>

Given the $z$-score of 5.73, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.
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

- 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?

Analyzing Patterns
- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
- Spatial Autocorrelation (Morans I)
Global Spatial Autocorrelation Tools: Spatial Autocorrelation

Spatial Statistics Toolbox

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

1969

1985

2002

Spatial Autocorrelation

- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Multi-Distance Spatial Cluster Analysis (Ripley's K Function)
- Spatial Autocorrelation (Morans I)
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 Report

Moran's I Index: -0.619737
Z-score: -1.01326
p-value: 0.019

Significance Levels:
-0.1: > 2.58
-0.05: > 1.96
-0.10: > 1.65
-0.15: > 1.45
-0.20: > 1.25
-0.25: > 0.95
-0.30: > 0.65
-0.35: > 0.35
-0.40: > 0.15
-0.45: > 0.05
-0.50: > 0.01

Critical Values:
-0.1: 2.58
-0.05: 1.96
-0.10: 1.65
-0.15: 1.45
-0.20: 1.25
-0.25: 0.95
-0.30: 0.65
-0.35: 0.35
-0.40: 0.15
-0.45: 0.05
-0.50: 0.01

Global Moran's I Summary

Given the z-score of -0.10, the pattern does not appear to be significantly different than random.
Local Spatial Autocorrelation Tools: Hot Spots & Outliers

Spatial Statistics Toolbox

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

Local Spatial Autocorrelation Tools: Hot Spots & Outliers

- 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
Exercise: Hot Spot Analysis

Spatial statistics tools:
- Analyzing Patterns
  - Average Nearest Neighbor
  - High/Low Clustering (Getis-Ord General G)
  - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
  - Spatial Autocorrelation (Morans I)
- Mapping Clusters
  - Cluster and Outlier Analysis (Anselin Local Morans I)
  - Hot Spot Analysis (Getis-Ord Gi*)

Hot Spot Analysis (Getis-Ord Gi*):
- Input Feature Class: Redwood_City_crimes_collected_50meters
- Input Field: ICOUNT
- Output Feature Class: E:\Spatial_Statistics_Law_Conference\Crime_data_Peninsula\Projected_data.gdb\Hot_Spot_1
- Conceptualization of Spatial Relationships: FIXED_DISTANCE_BAND
- Distance Method: EUCLIDEAN_DISTANCE
- Standardization: NONE
- Distance Band or Threshold Distance (optional): 1500
- Self Potential Field (optional): 
- Weights Matrix File (optional):
Regression Analysis:
How much is crime explained by population density?

**Ordinary Least Squares**

**Output Feature Class:** OLS_pop

**Coefficient Output Table:** (empty)

**Diagnostic Output Table:** (empty)

### Summary of OLS Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient StdError</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>Robust_SE</th>
<th>Robust_t</th>
<th>Robust_Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.712047 4.757335</td>
<td>-0.359860</td>
<td>0.720501</td>
<td>0.397028</td>
<td>-0.431215</td>
<td>0.688205</td>
</tr>
<tr>
<td>TOTPOP_CY</td>
<td>0.007088 0.002158</td>
<td>3.284786</td>
<td>0.001890*</td>
<td>0.001933</td>
<td>3.656909</td>
<td>0.000605*</td>
</tr>
</tbody>
</table>

### OLS Diagnostics

- **Number of Observations:** 51
- **Akaike's Information Criterion (AICs):** 427.630402
- **Multiple R-Squared:** 0.180462
- **Adjusted R-Squared:** 0.163737

### Notes on Interpretation

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

**WARNING:** 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
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://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](http://gis.stanford.edu)