Spatial Pattern Analysis: Mapping Trends and Clusters

Lauren M. Scott, PhD
Lauren Rosenshein Bennett, MS
Presentation Outline

- Spatial statistics overview
- Describing spatial patterns
- Quantifying spatial patterns
- Mapping spatial clusters
  - Hot and cold spots
  - Spatial outliers
  - Similar features

DEMOS
- Analyzing Ushahidi data
- Exploring regional variations in health care
- Multi-dimensional data analysis
What are Spatial Statistics in ArcGIS?

• A set of exploratory methods and techniques
• Specifically developed for use with geographic data
  - They incorporate space (area, length, proximity, orientation) directly into their mathematics
• They describe and model:
  - Spatial Distributions
  - Spatial Patterns
  - Spatial Processes
  - Spatial Relationships

_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.
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.
Describing Spatial Data

- Which site is most accessible?
- What is the primary wind direction in the winter months?
- Where is the population center and how it is changing over time?
- Which strain of the disease has the broadest spatial distribution? Which are the most spatially integrated?
Finding the center

- The Mean Center tool computes the average x and y coordinate, based on all features in the study area.
Measuring central tendency

- **Mean Center**
  - Computes average X and Y

- **Median Center**
  - Finds the location that minimizes Euclidean distance to all features
  - More robust to outliers than the Mean Center tool

- **Central Feature**
  - Find the feature that is most central to all other features
Distribution and Direction
(Standard Deviational Ellipse)

- Abstracting spatial and temporal trends in a distribution of features

Dengue Fever Outbreak
Salmon Distributions
Ellipse size

Normal distribution

- Mean +

99% 95% 68%

1 = 68% of features
2 = 95% of features
3 = 99% of features
Measuring segregation and integration
(Standard Deviational Ellipse)

Segregation Index =

\[
1 - \frac{E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_n}{E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n} = 1 - \frac{2931680545.83}{7994760004.92} = 0.63
\]
Ushadidi data analysis

Lauren Rosenshein Bennett, MS
Quantifying Spatial Patterns

Which pattern is most clustered?

Hospital Readmission Rates 2010
Quantifying Spatial Patterns
Which pattern is most clustered?

Hospital Readmission Rates 2010

Equal Interval

Natural Breaks

Quantile
Quantifying Spatial Patterns

You say it’s clustered

- Compared to what?
- Says who?
Compared to what?
Complete Spatial Randomness

- Inferential statistics start with a null hypothesis
  - Random distribution of features
  - Random distribution of values within fixed features
Is CSR useful?
How Random is That?

- What are the chances that we are looking at one of many possible instances of a random pattern?
- Are these just random acts?
- Is it just bad luck?

Crime in Lincoln, Nebraska

Random
Says who?

Probability theory

- Inferential statistics report a p-value and z-score
  - z-scores are standard deviations
  - p-values are probabilities

- z-scores can be mapped to specific p-values
  - p-value 0.01 = z-score +/- 2.58
  - p-value 0.05 = z-score +/-1.96
  - p-value 0.10 = z-score +/-1.65
Quantifying spatial patterns over time
Spatial Autocorrelation (Global Moran’s I)

Thematic maps showing relative per capita Income for New York by county, 1969 to 2002

Is the spatial “gap” between rich and poor increasing or dissipating?
Quantifying spatial patterns over time

Spatial Autocorrelation (Global Moran’s I)

1969: 5.21
1985: 4.26
2002: 2.4

Analyzing Patterns
- Average Nearest Neighbor
- High/Low Clustering (Getis-Ord General G)
- Incremental Spatial Autocorrelation
- Multi-Distance Spatial Cluster Analysis (Ripley’s K Function)
- Spatial Autocorrelation (Moran’s I)
Incremental Spatial Autocorrelation

- Identifies appropriate scales for analysis
  - Distances where processes promoting clustering are most pronounced
- Helpful for tools requiring a distance or radius including:
  - Point Density
  - IDW
  - Hot Spot Analysis
  - Cluster/Outlier Analysis

Data mining reduces trial and error
Mapping Spatial Clusters

- Mapping spatial clusters
  - Hot and cold spots
  - Spatial outliers
  - Similar features

Where are the 911 call hot spots?

Which homes sold for more than expected?

Which countries face similar challenges?
## Mapping Spatial Clusters

<table>
<thead>
<tr>
<th>Questions</th>
<th>Tools</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where are the hot spots? Where are the cold spots? How intense is the clustering?</td>
<td>Hot Spot Analysis, Cluster and Outlier Analysis</td>
<td>Where are the sharpest boundaries between affluence and poverty? Where are biological diversity and habitat quality highest?</td>
</tr>
<tr>
<td>Where are the spatial outliers?</td>
<td>Cluster and Outlier Analysis</td>
<td>Where do we find anomalous spending patterns in Los Angeles?</td>
</tr>
<tr>
<td>How can resources be most effectively deployed?</td>
<td>Hot Spot Analysis</td>
<td>Where do we see unexpectedly high rates of diabetes? Where are kitchen fires a higher-than-expected proportion of residential fires?</td>
</tr>
<tr>
<td>Which locations are farthest from the problem?</td>
<td>Hot Spot Analysis</td>
<td>Where should evacuation sites be located?</td>
</tr>
<tr>
<td>Which features are most alike? What does the spatial fabric of the data look like?</td>
<td>Grouping Analysis</td>
<td>Which crimes in the database are most similar to the one just committed? Which disease incidents are likely part of the same outbreak?</td>
</tr>
</tbody>
</table>
Exploring regional variations in health care

Lauren Rosenshein Bennett, MS
Examining the spatial pattern of poverty in Kenya
(Cluster and Outlier Analysis: Anselin Local Moran’s I)
Space-Time Cluster Analysis

Defining the space-time window

- Create the space-time weights matrix file (.swm)
- Use the .swm to perform cluster analyses
Space-time vs Time Snapshots

- Time-slices can disconnect related features
- Space-time relationships are relative to each feature’s date stamp
Time Snapshots
Analyzing Data in both Space and Time
Space-Time Cluster Analysis

Visualization

- Default symbology at 10.1 is a 2D display
- To see the temporal patterns:
  - Animate the results or
  - Use 3D symbols in ArcGlobe or ArcScene
Grouping Analysis

Group 1: low income, low education, high unemployment, high crime
Group 2: middle values for all variables
Group 3: higher income and education, lower unemployment and crime

Multi-dimensional data analysis simplifies complex data
Multidimensional Data Analysis

Lauren Rosenshein Bennett, MS

The Africa vulnerability model data was kindly provided by the Robert S. Strauss Center for International Security and Law.
Resources for learning more...

- **www.esriurl.com/spatialstats**
  - Short videos
  - Articles and blogs
  - Online documentation
  - Supplementary model and script tools
  - Hot spot, Regression, and ModelBuilder tutorials

- **resources.arcgis.com**

QUESTIONS?

LBennett@Esri.com
LScott@Esri.com