Measuring Spatial Autocorrelation

Tobler's first law of geography

'Everything is related to everything else, but near things are more related than distant things'

Tobler, W. A computer model simulation of urban growth in the Detroit region. Economic Geography, 46(2): 234-240

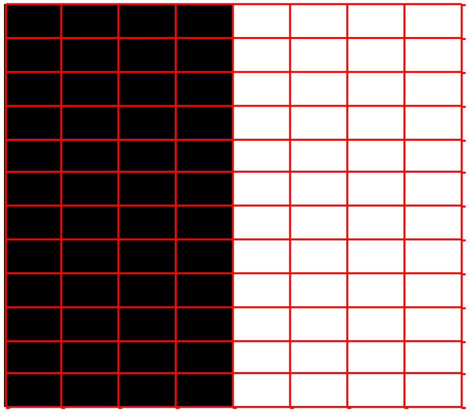
Spatial autocorrelation

'Spatial autocorrelation can be loosely defined as the coincidence of value similarity with locational similarity'

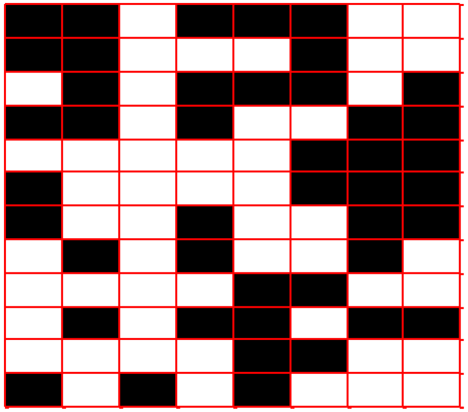
(Anselin and Bera, 1998)

Global autocorrelation

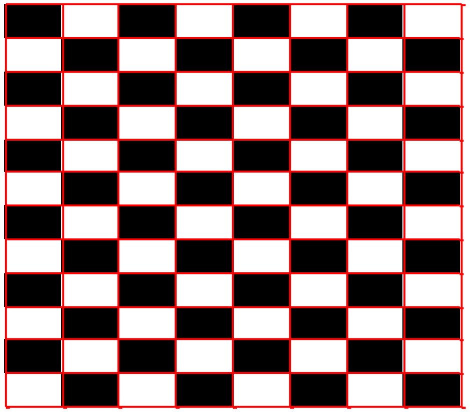
Positive (most geographical processes)



Random



Negative



Operationalizing the spatial autocorrelation concept

$$y_i = f(y_j) \quad , \quad i=1,..,n \quad \text{and } i \neq j$$

What is f() ?

From Tobler, **need a function that includes the « proximity » between i and j**

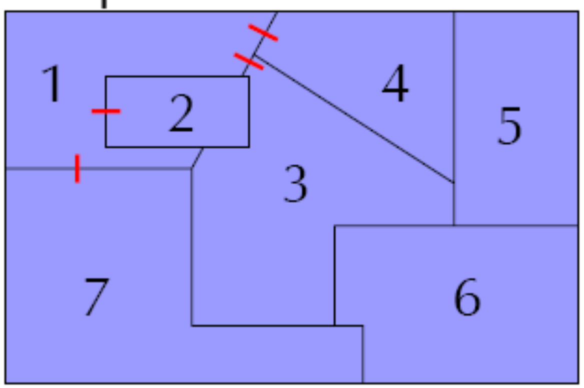
Example:

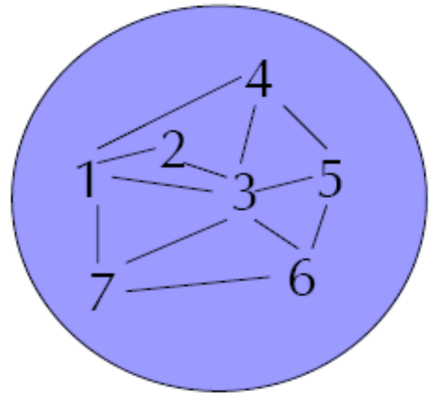
- a 0/1 matrix representing adjacency between each pair of spatial units
- a **weight matrix** w representing how « proximal » are each pair of spatial units

Weight matrix ? Contiguity (adjacency)

Recall:

Connectivity example:
Map of seven areas





Share border:

$\text{Neighbor}(1) = \{2,3,4,7\}$
$\text{Neighbor}(2) = \{1,3\}$
..
$\text{Neighbor}(7) = \{1,3,6\}$

$$\widetilde{W} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

From Sweeney, 2006 ppt

Binary/Standardized matrix

Binary W matrix:

$$\tilde{V} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Row standardized W matrix:

$$W = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{bmatrix}$$

From Sweeney, 2006 ppt

Alternative weight matrix ?

- Second order contiguity, or more

- Distance-based weights
  - Euclidean distance ~potential of spatial interaction $w_{ij} = 1/d_{ij}^{\alpha}$
  - 0/1 matrix based on distance cutoffs

- Transport cost distance functions
  - Network based
  - generalised i.e. time +cost

- Socio-economic distances $w_{ij} = 1/|x_i - x_j|$

Measuring spatial autocorrelation: Moran's I

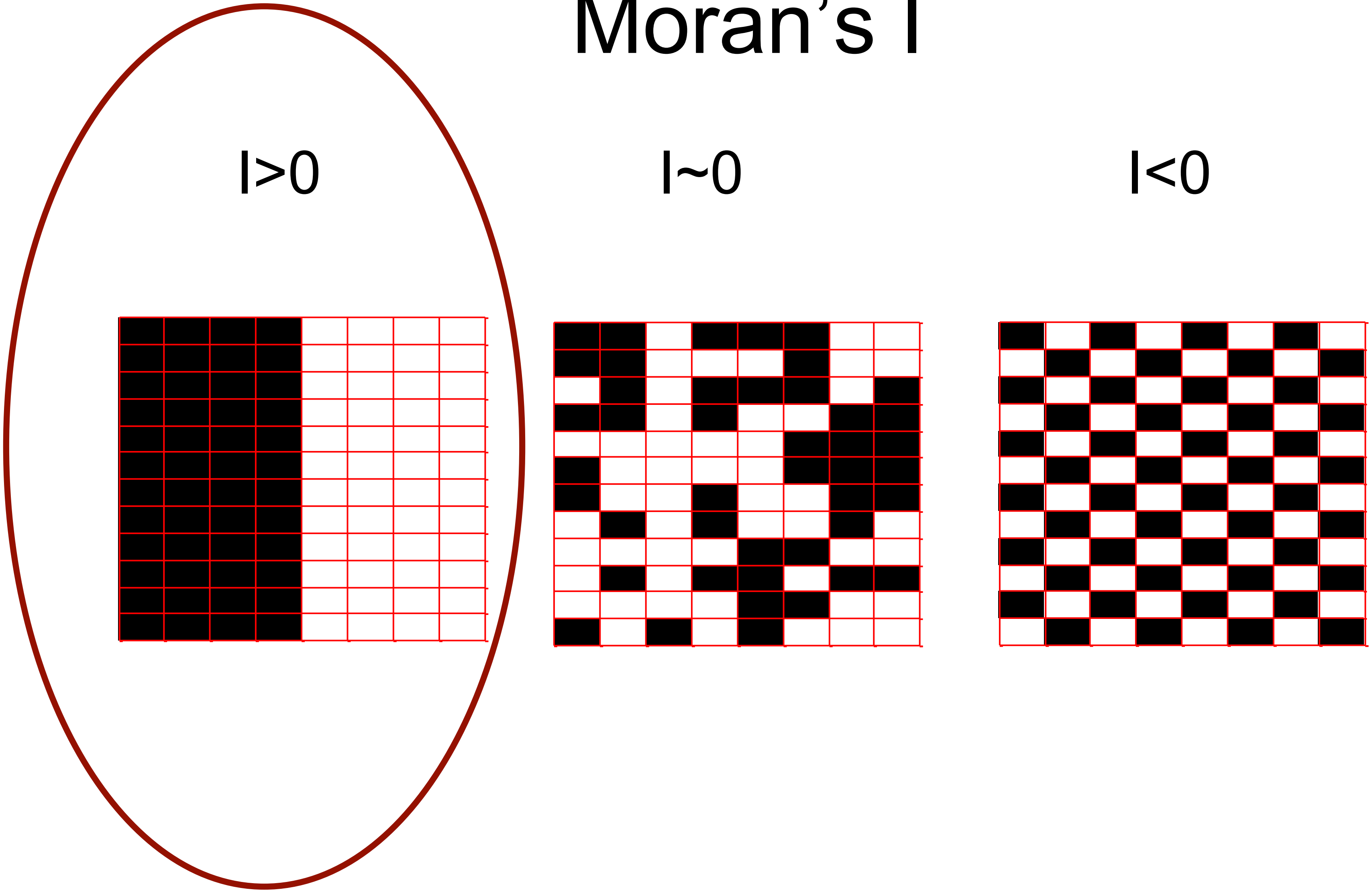
- Most widely used (global) index of autocorrelation

$$I = \frac{1}{p}\frac{\sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2}, \text{ where}$$

$$p = \sum_i \sum_j w_{ij} / n$$

- I>0, positive spatial autocorrelation
- I<0, negative spatial autocorrelation
- Results will depend on specification of *w*

Measuring spatial autocorrelation: Moran's I



ArcToolbox

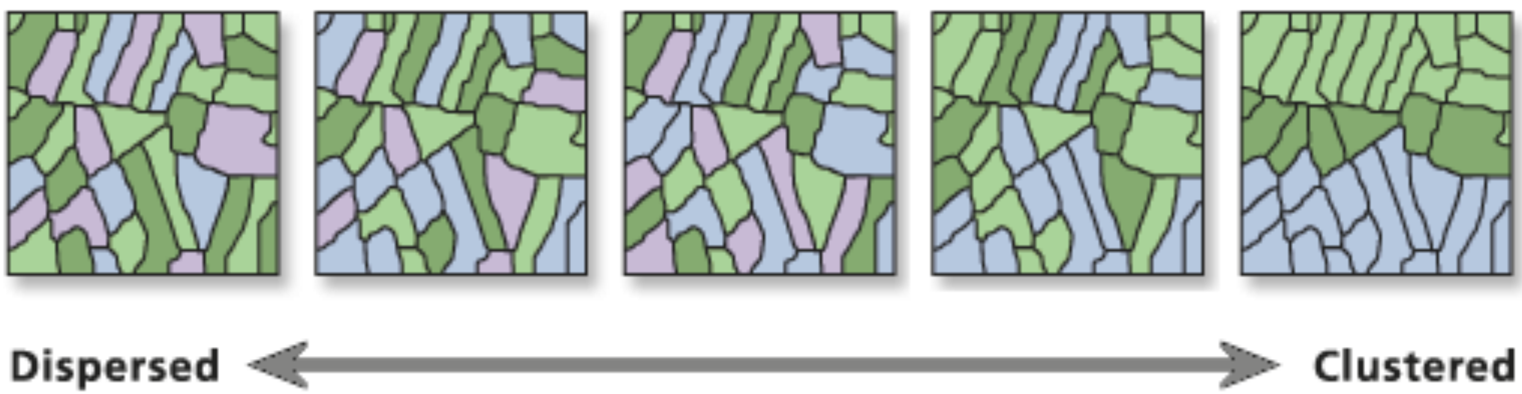
Spatial Autocorrelation (Morans I) (Spatial Statistics)

related topics open tool...

Measures spatial autocorrelation based on feature locations and attribute values.

Learn more about how Spatial Autocorrelation: Moran's I works

Illustration

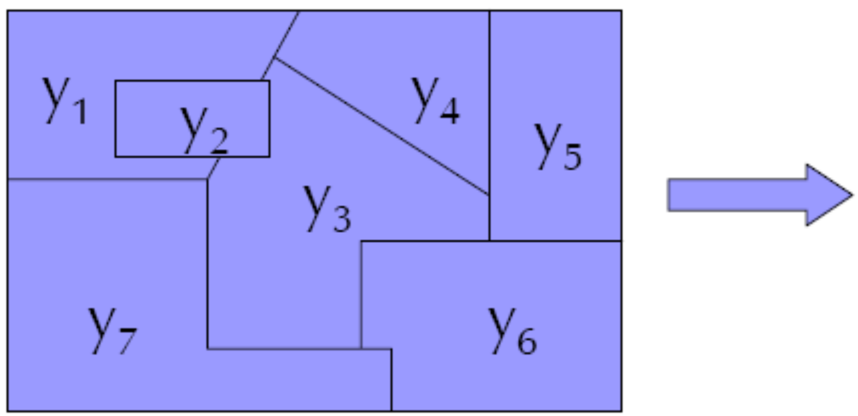


Measuring spatial autocorrelation

Graphical interpretation of Moran and linking with local spatial association (LISA)

Measuring spatial autocorrelation

- Suppose the following Y values (e.g. crime)



$$y = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 8 \\ 9 \\ 4 \\ 2 \end{bmatrix}$$

From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Define the Proximity relationship: weight matrix (row-standardized)

Binary W matrix:

$$\tilde{V} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Row standardized W matrix:

$$W = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{bmatrix}$$

From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Compute

$$y = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 8 \\ 9 \\ 4 \\ 2 \end{bmatrix} \qquad W = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{bmatrix}$$

$$Wy = \begin{bmatrix} 3(\frac{1}{4}) + 1(\frac{1}{4}) + 8(\frac{1}{4}) + 2(\frac{1}{4}) \\ 4(\frac{1}{2}) + 1(\frac{1}{2}) \\ 4(\frac{1}{6}) + 3(\frac{1}{6}) + 8(\frac{1}{6}) + 9(\frac{1}{6}) + 4(\frac{1}{6}) + 2(\frac{1}{6}) \\ 4(\frac{1}{3}) + 1(\frac{1}{3}) + 9(\frac{1}{3}) \\ 1(\frac{1}{3}) + 8(\frac{1}{3}) + 4(\frac{1}{3}) \\ 1(\frac{1}{3}) + 9(\frac{1}{3}) + 2(\frac{1}{3}) \\ 4(\frac{1}{3}) + 1(\frac{1}{3}) + 4(\frac{1}{3}) \end{bmatrix} = \begin{bmatrix} (7/2) \\ (5/2) \\ 5 \\ (14/3) \\ (13/3) \\ 4 \\ 9 \end{bmatrix}$$

« Wy is high in i when neighbours of i have high crime value »

From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Compute

$$y = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 8 \\ 9 \\ 4 \\ 2 \end{bmatrix} \qquad W = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{bmatrix}$$

$$Wy = \begin{bmatrix} 3(\frac{1}{4}) + 1(\frac{1}{4}) + 8(\frac{1}{4}) + 2(\frac{1}{4}) \\ 4(\frac{1}{2}) + 1(\frac{1}{2}) \\ 4(\frac{1}{6}) + 3(\frac{1}{6}) + 8(\frac{1}{6}) + 9(\frac{1}{6}) + 4(\frac{1}{6}) + 2(\frac{1}{6}) \\ 4(\frac{1}{3}) + 1(\frac{1}{3}) + 9(\frac{1}{3}) \\ 1(\frac{1}{3}) + 8(\frac{1}{3}) + 4(\frac{1}{3}) \\ 1(\frac{1}{3}) + 9(\frac{1}{3}) + 2(\frac{1}{3}) \\ 4(\frac{1}{3}) + 1(\frac{1}{3}) + 4(\frac{1}{3}) \end{bmatrix} = \begin{bmatrix} (7/2) \\ (5/2) \\ 5 \\ (14/3) \\ (13/3) \\ 4 \\ 9 \end{bmatrix}$$

Wy is a « spatially lagged » transform of Y

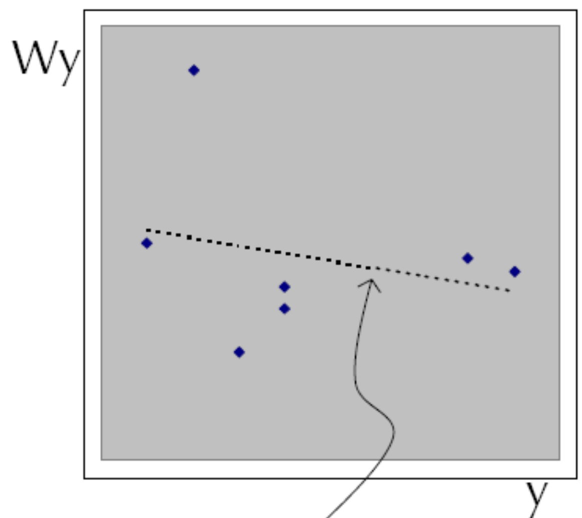
From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Scatterplot of Y and W_Y

$$y = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 8 \\ 9 \\ 4 \\ 2 \end{bmatrix} \quad Wy = \begin{bmatrix} (7/2) \\ (5/2) \\ 5 \\ (14/3) \\ (13/3) \\ 4 \\ 9 \end{bmatrix}$$

Plot Wy, y:



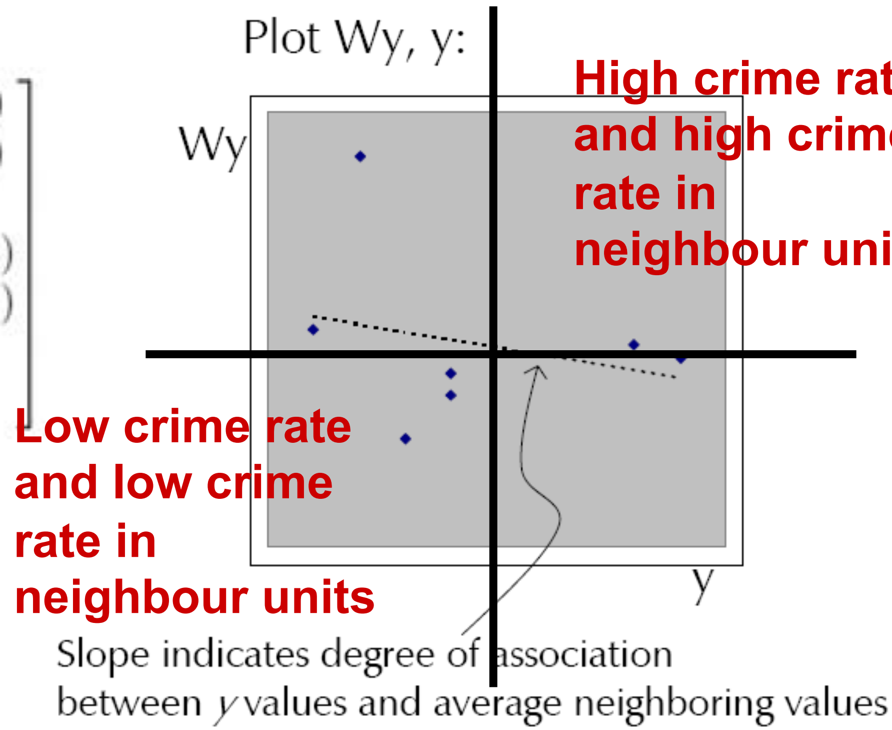
Slope indicates degree of association between *y* values and average neighboring values

From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Scatterplot of Y and W_Y

$$y = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 8 \\ 9 \\ 4 \\ 2 \end{bmatrix} \quad Wy = \begin{bmatrix} (7/2) \\ (5/2) \\ 5 \\ (14/3) \\ (13/3) \\ 4 \\ 9 \end{bmatrix}$$



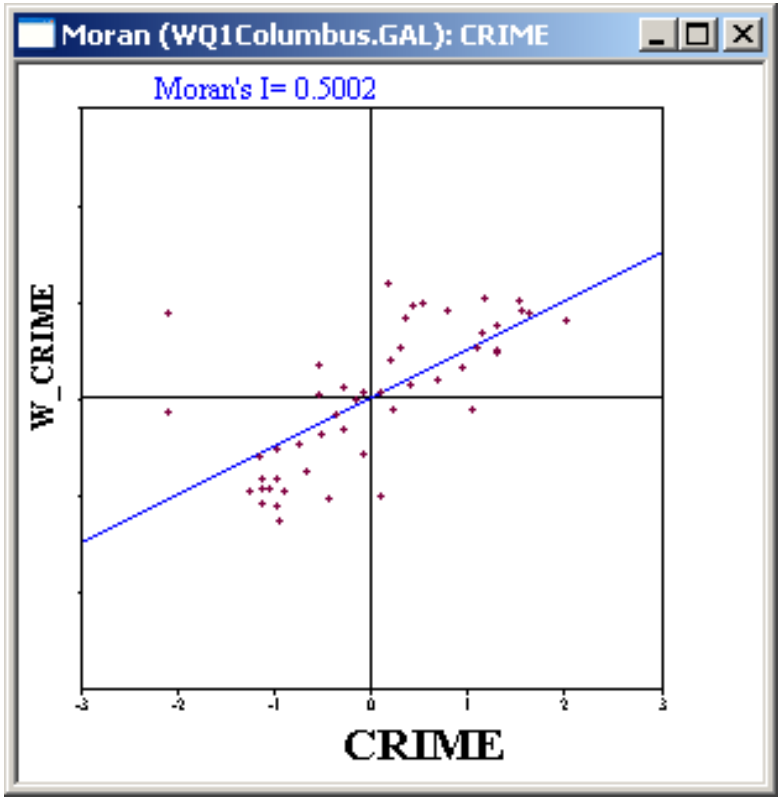
From Sweeney, 2006 ppt

Measuring spatial autocorrelation

- Moran's scatterplot: idem but with standardized Y values
- Index Moran's I is then the slope of the linear curve fitted in the scatterplot
- *If I >0 => positive correlation between crime rate in a place and crime rate in neighbourhing places = positive spatial autocorrelation*

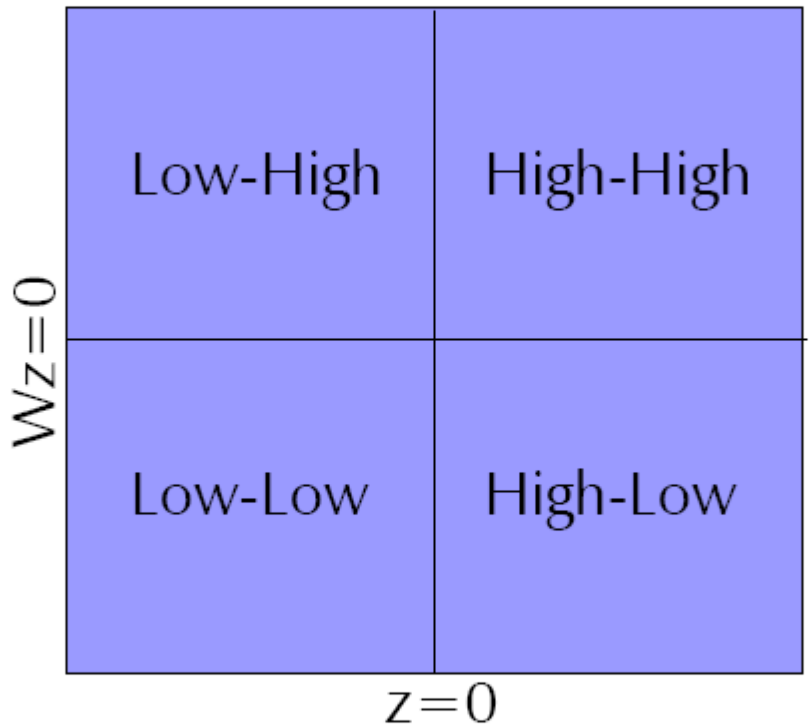
From Sweeney, 2006 ppt

Real example: crime in Columbus, US



Source: see GEODA on Wikipedia

Moran's scatterplot



Global Moran -> Local Moran

- Global autocorrelation analysis (I) yields only one statistic to summarize the whole study area

  => Assumes homogeneity.

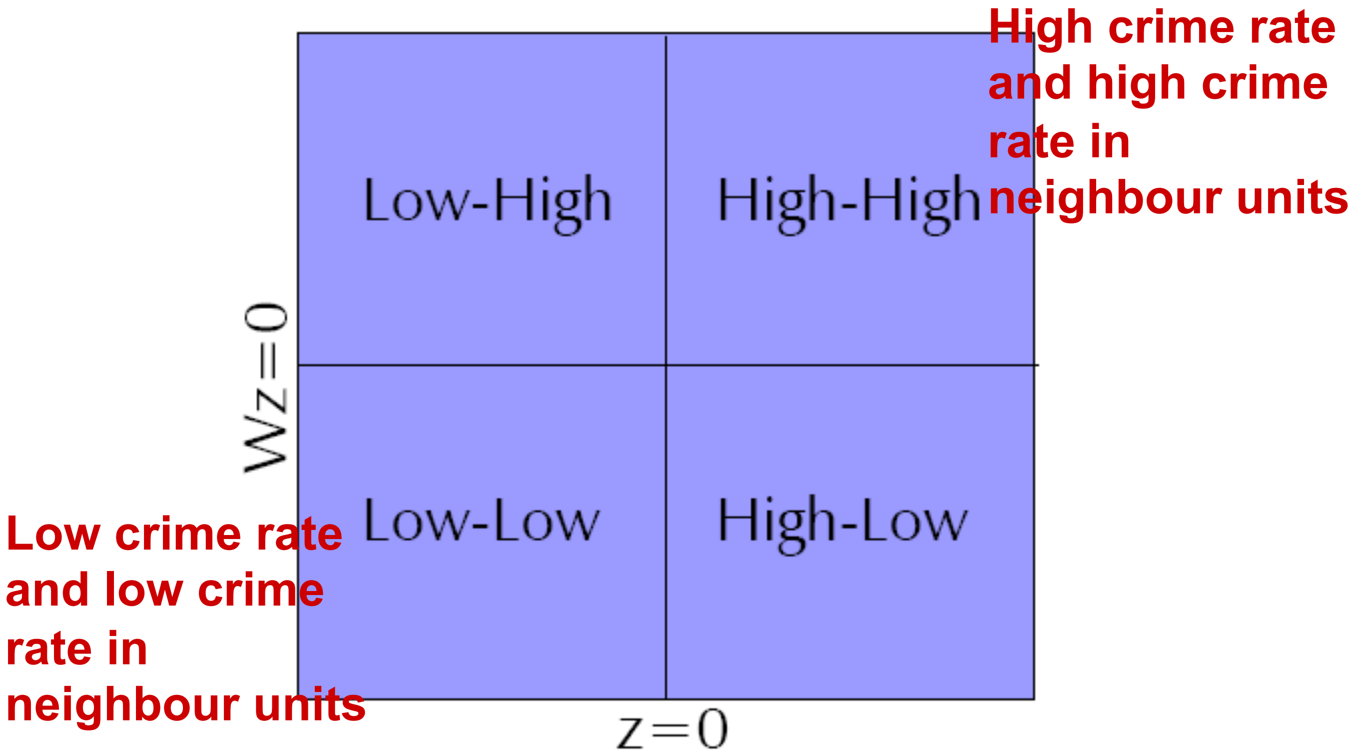
  => If homogeneity assumption does not hold, having only one statistic does not make sense

  => the statistic should differ over space = Local Index of Spatial Association (LISA)

- LISA (or Local Moran) is the crossproduct of the lagged variable and the variable itself (normalized).
- LISA values can be found easily from Moran scatterplot

Moran's scatterplot > Local Moran

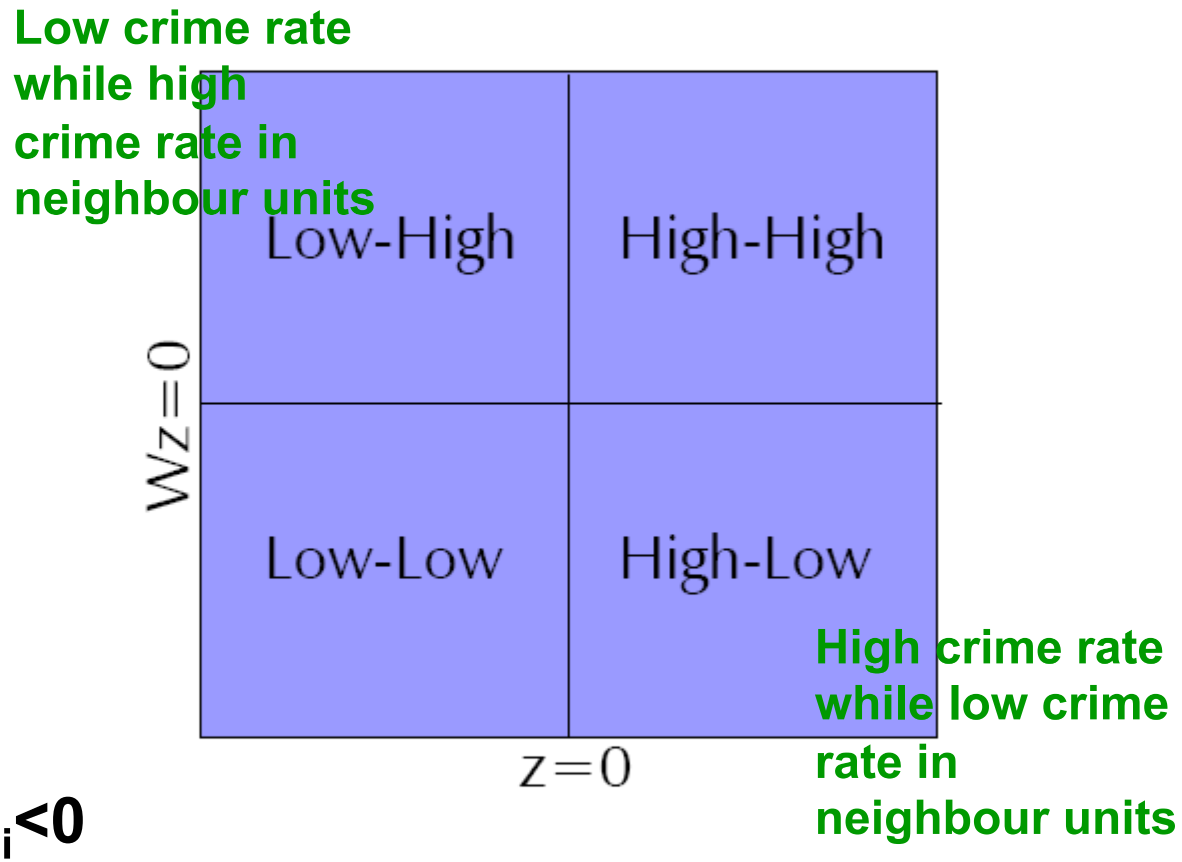
- Places having similar value as their neighbours:



- **LISA: $I_i > 0$**

Moran’s scatterplot > Local Moran

- Places having value different from their neighbours:



- **LISA: I_i<0**

LISA concept

Local indicator of spatial association (LISA)

- *« The LISA for each observation gives an indication of significant spatial clustering of similar values around that observation »*

- *« The sum of LISAs for all observations is proportional to a global indicator of spatial association »*

From: Anselin, L, 1995, Local indicators of spatial association—LISA, Geographical Analysis, 27, 93-115

LISA=>Hot and Cold Spot map

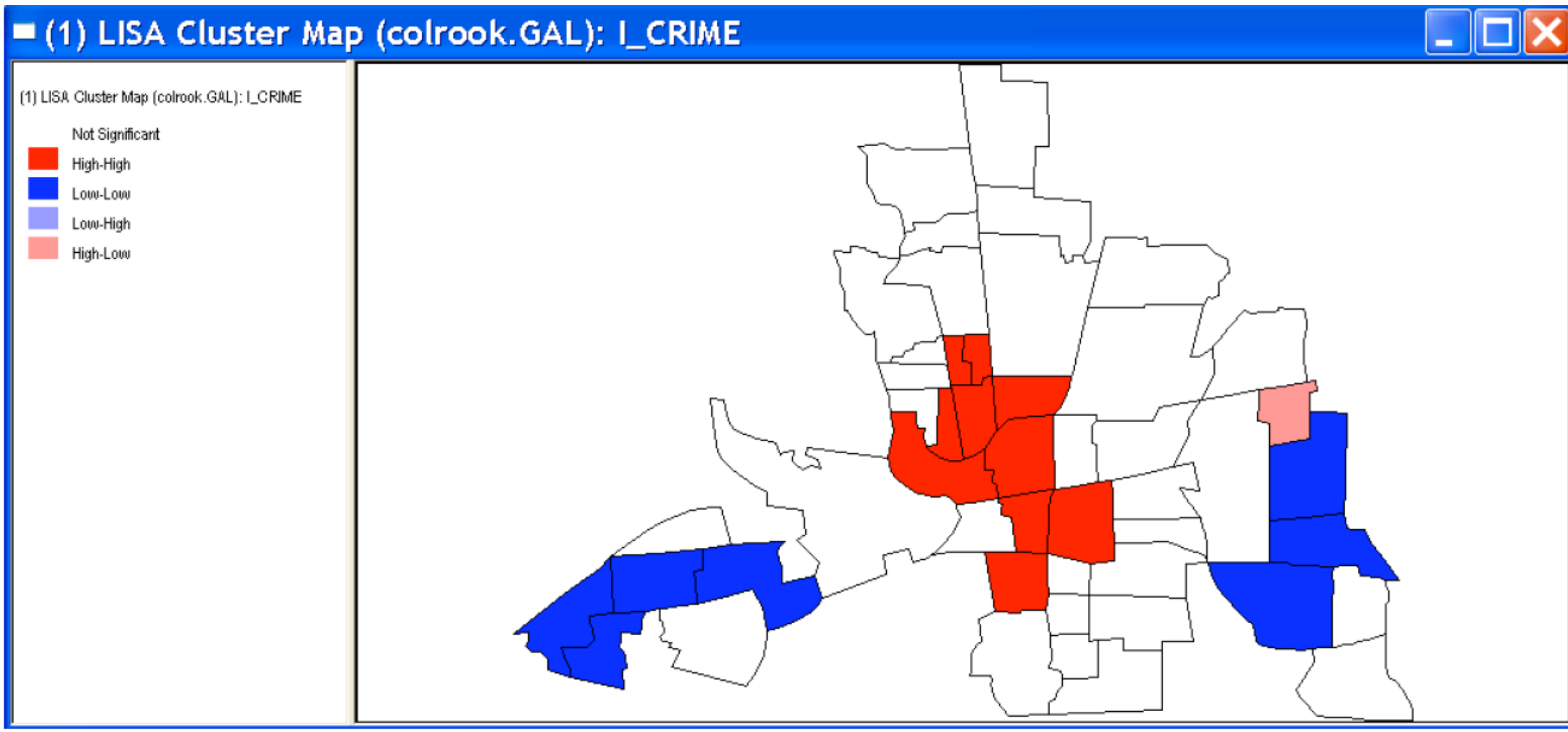


Figure 154. LISA cluster map for Columbus **CRIME**.

ArcToolbox

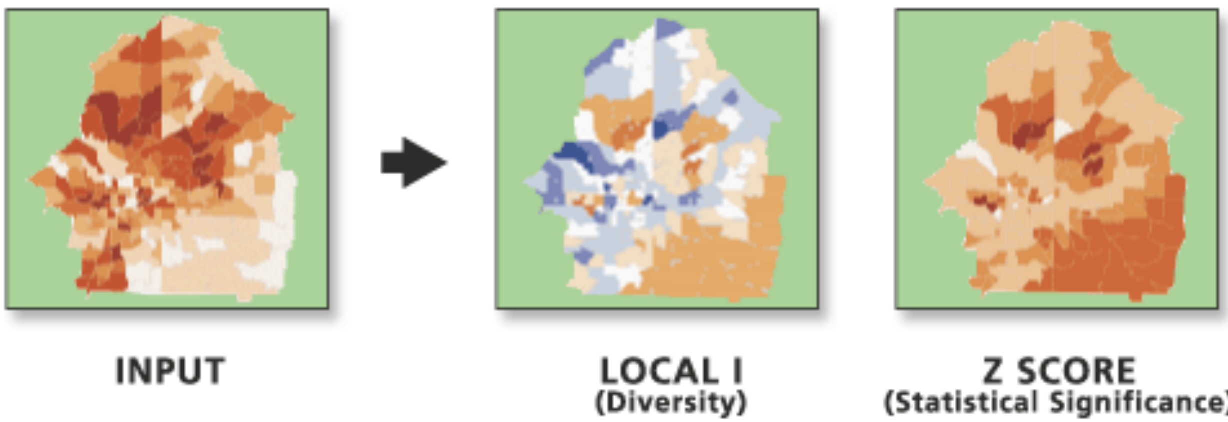
Cluster and Outlier Analysis: Anselin Local Moran's I (Spatial Statistics)

related topics   open tool...

Given a set of weighted data points, identifies those clusters of points with values similar in magnitude and those clusters of points with very heterogeneous values.

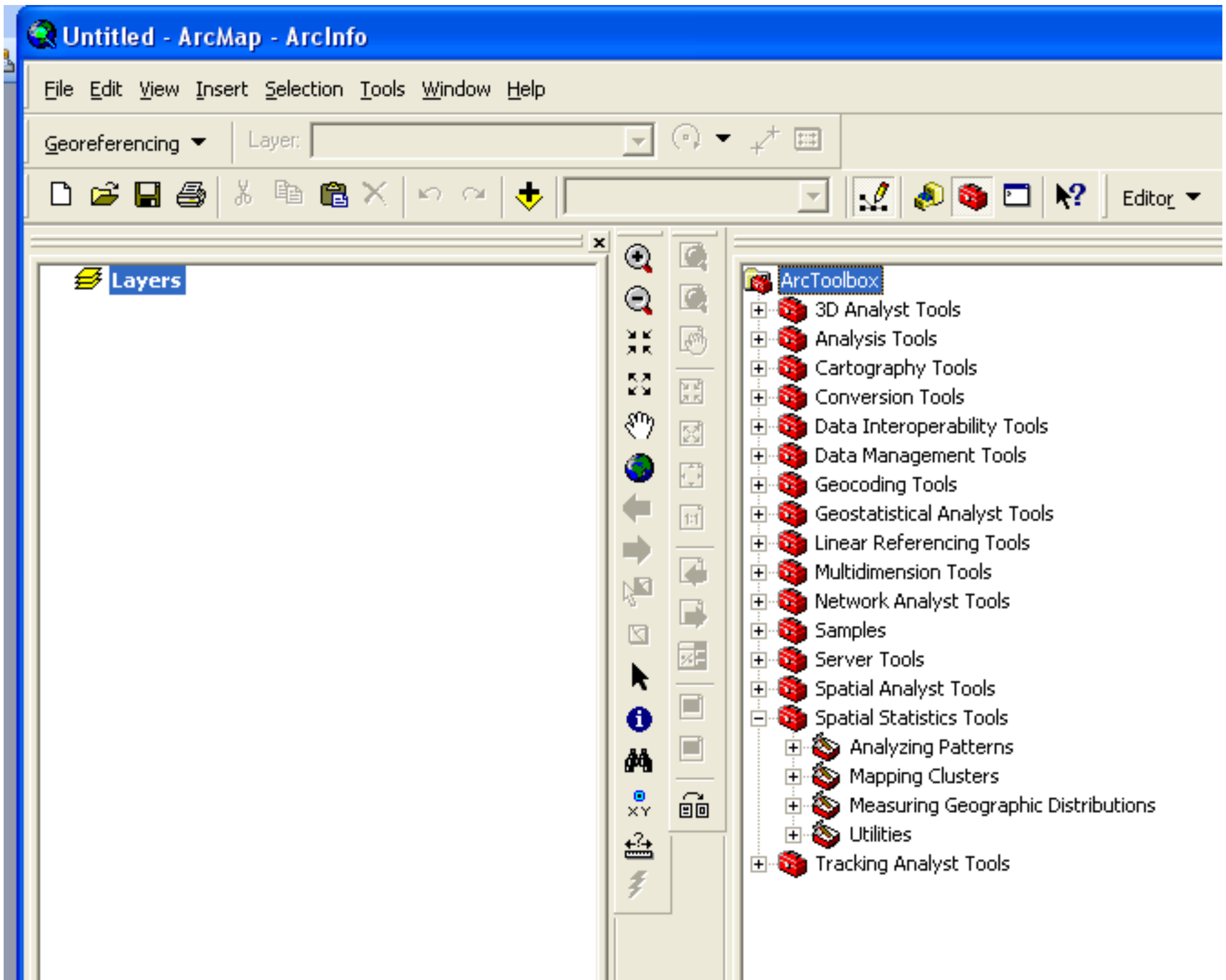
Learn more about how Cluster and Outlier Analysis: Anselin Local Moran's I works

Illustration

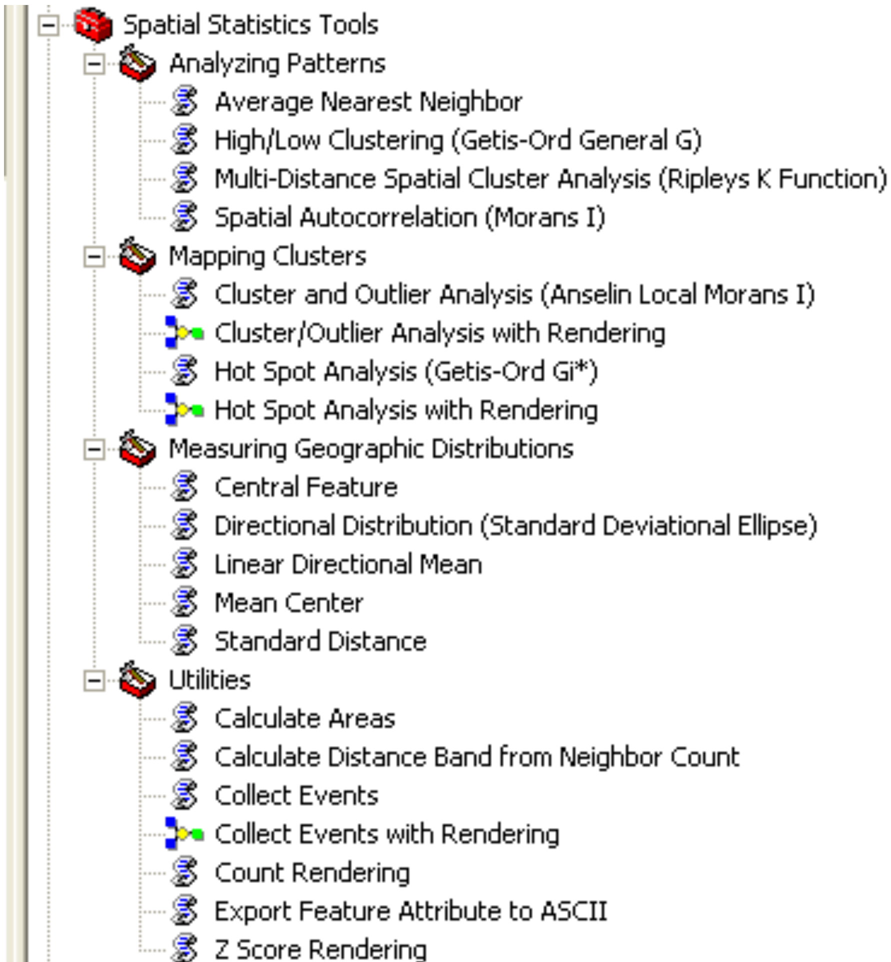


LISA cluster map + significance test (based on p-value)

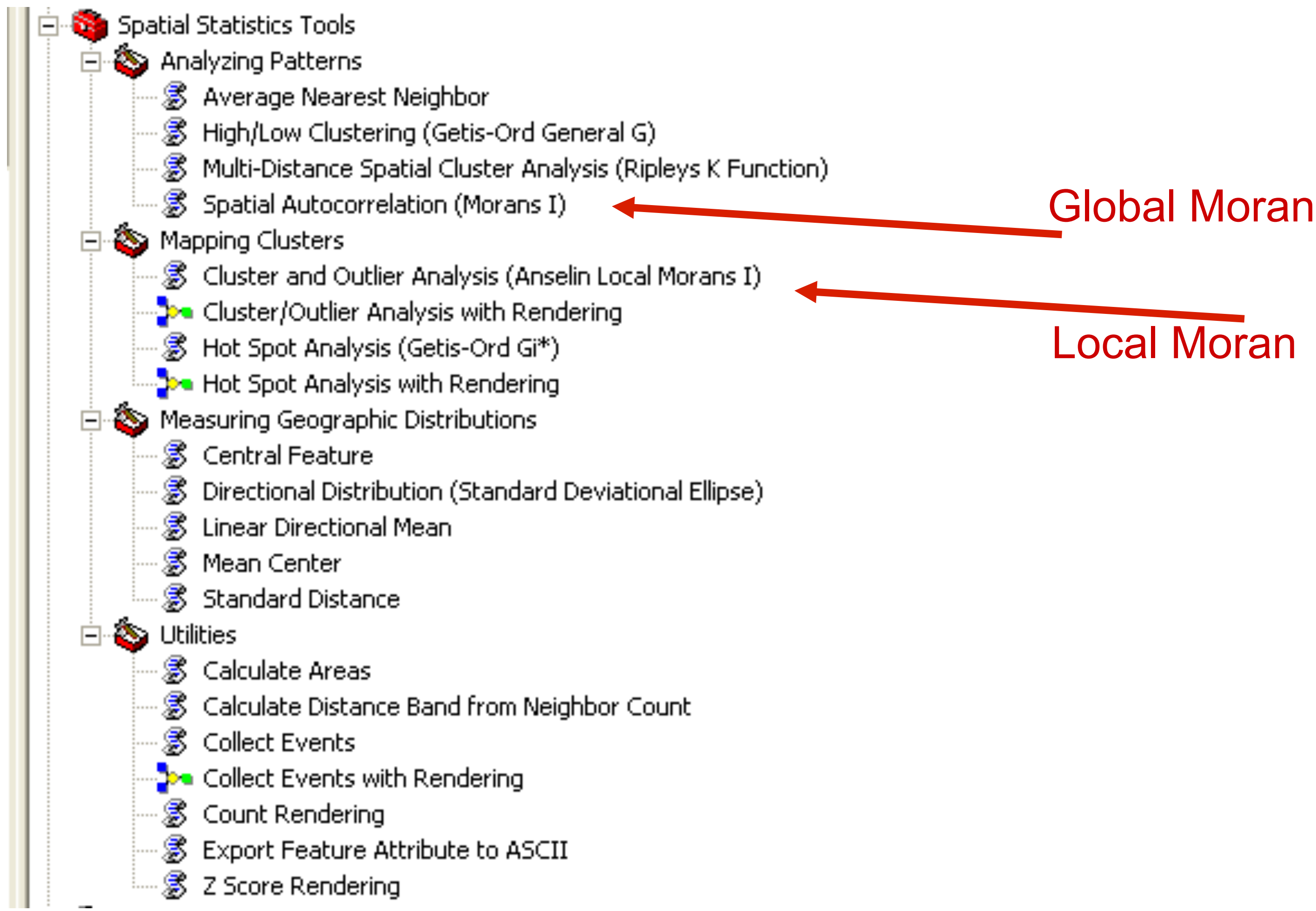
ArcToolbox – Spatial Statistics Tools



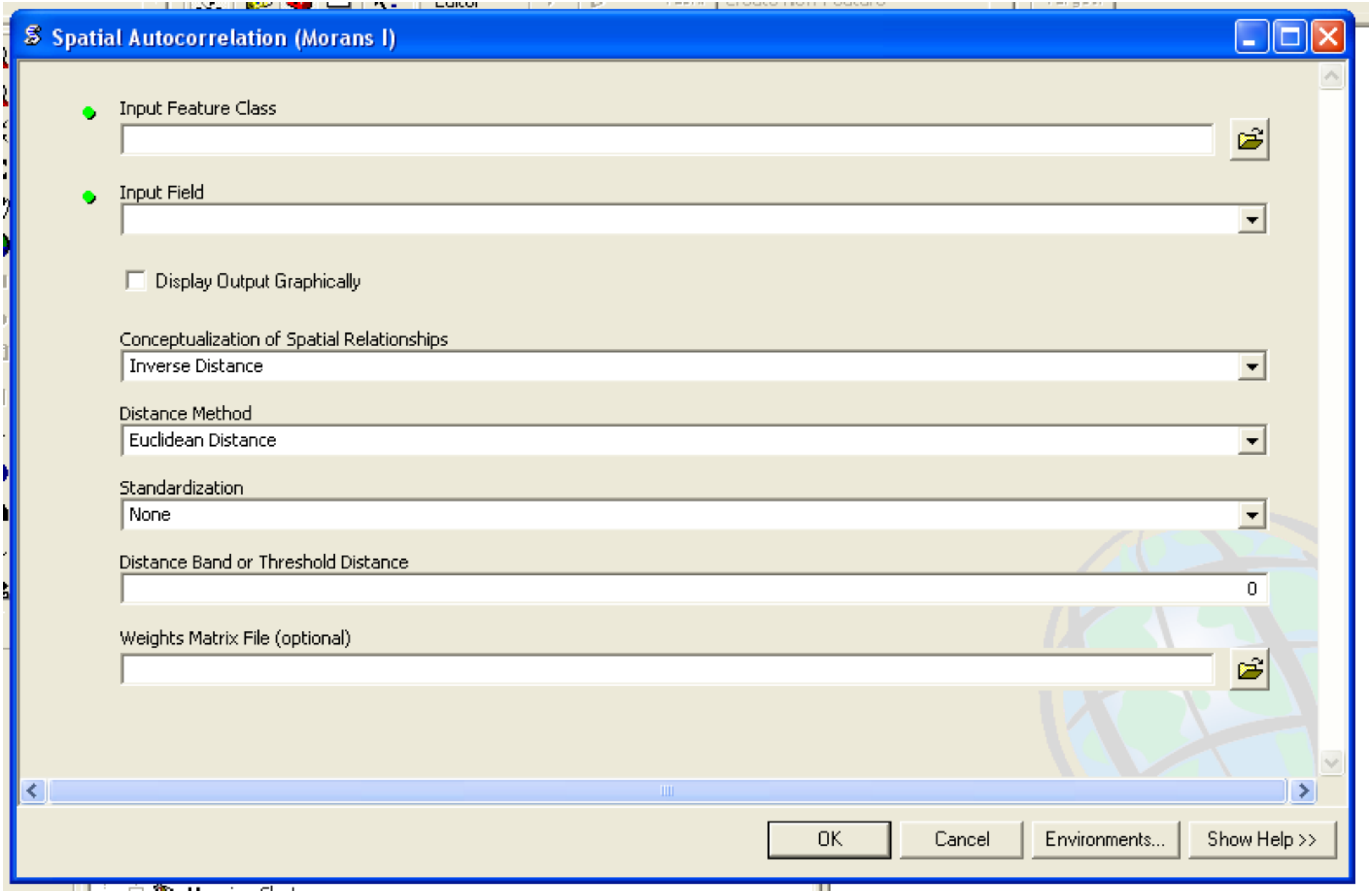
ArcToolbox – Spatial Statistics Tools



ArcToolbox – Spatial Statistics Tools



Moran’s I – dialog box



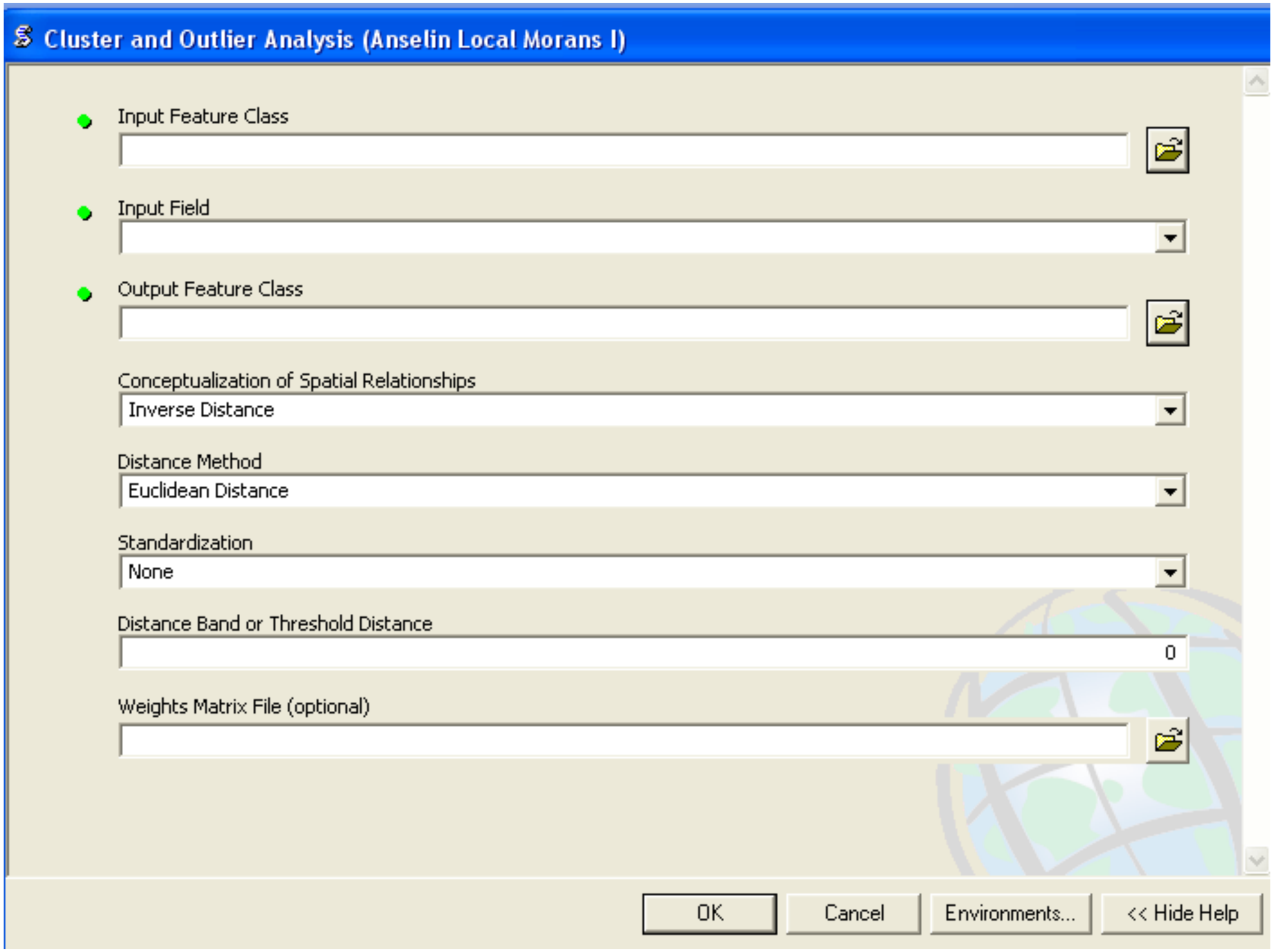
Moran’s I – parameters

<Input_Feature_Class>	The feature class for which spatial autocorrelation will be calculated.	Feature Layer
<Input_Field>	The numeric field used in measuring spatial autocorrelation.	Field
<Display_Output_Graphically>	Specifies whether the tool will display the Moran's I and z score values graphically. • True—The output will be displayed graphically. • False—The output will not be displayed graphically.	Boolean
<Inverse Distance | Inverse Distance Squared | Fixed Distance Band | Zone of Indifference | Polygon Contiguity (First Order) | Get Spatial Weights From File>	Specifies how spatial relationships between features are conceptualized. • Inverse Distance—The impact of one feature on another feature decreases with distance. • Inverse Distance Squared—Same as Inverse Distance, but the impact decreases more sharply over distance. • Fixed Distance Band—Everything within a specified critical distance is included in the analysis; everything outside the critical distance is excluded. • Zone of Indifference—A combination of Inverse Distance and Fixed Distance Band. Anything up to a critical distance has an impact on your analysis. Once that critical distance is exceeded, the level of impact quickly drops off. • Polygon Contiguity (First Order)—The neighbors of each feature are only those with which the feature shares a boundary. All other features have no influence. • Get Spatial Weights From File—Spatial relationships are defined in a spatial weights file. The pathname to the spatial weights file is specified in the Weights Matrix File parameter.	String
<Euclidean Distance | Manhattan Distance>	Specifies how distances are calculated when measuring spatial autocorrelation. • Euclidean (as the crow flies)—The straight-line distance between two points. • Manhattan (city block)—The distance between two points measured along axes at right angles. Calculated by summing the (absolute) differences between point coordinates.	String

Moran’s I – parameters

<table>
<tr><th></th><th></th><th></th></tr>
<tr><td><None | Row ></td><td>The standardization of spatial weights provides more accurate results.
• None—No standardization of spatial weights is applied.
• Row—Spatial weights are standardized by row. Each weight is divided by its row sum.</td><td>String</td></tr>
<tr><td><Distance_Band_or_Threshold_Distance></td><td>Specifies a distance cutoff value. When determining the neighbors for a particular feature, features outside the specified Distance Band or Threshold Distance are ignored in the cluster analysis. The value entered for this parameter should be in the units of the Input Feature Class' coordinate system.A value of zero indicates that no threshold distance is applied. This is only valid with the "Inverse Distance" and "Inverse Distance Squared" spatial conceptualizations.This parameter has no effect when "Polygon Contiguity" and "Get Spatial Weights From File" spatial conceptualizations are selected.</td><td>Double</td></tr>
<tr><td>{Weights_Matrix_File}</td><td>The pathname to a file containing spatial weights that define spatial relationships between features.</td><td>File</td></tr>
</table>

LISA – same dialog box



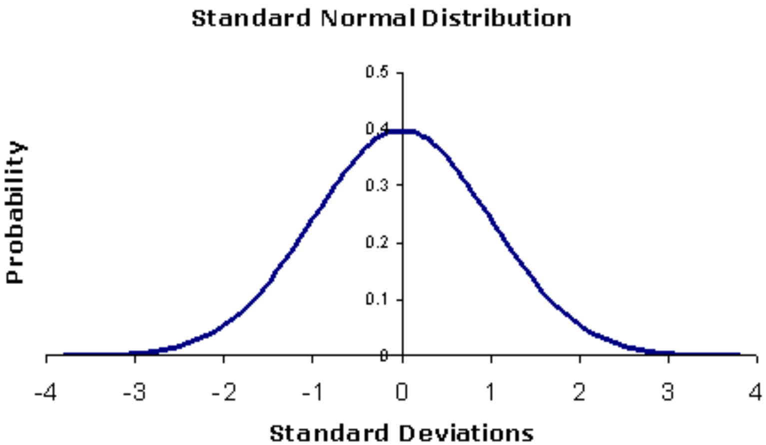
Significance test

ArcToolbox

What is a Z Score?

Most statistical tests begin by identifying a null hypothesis. The null hypothesis for pattern analysis tools essentially states that there is no pattern; the expected pattern is one of hypothetical random chance. The Z Score is a test of statistical significance that helps you decide whether or not to reject the null hypothesis.

Z scores are measures of standard deviation. For example, if a tool returns a Z score of +2.5 it is interpreted as "+2.5 standard deviations away from the mean". Z score values are associated with a standard normal distribution. This distribution relates standard deviations with probabilities and allows significance and confidence to be attached to Z scores.



Very high or a very low Z scores are found in the tails of the normal distribution. From the graph above, it is evident that the probabilities in the tails of the distribution are very low. When you perform a feature pattern analysis and it yields either a very high or a very low Z Score, this indicates it is very UNLIKELY that the observed pattern is some version of the theoretical spatial pattern represented by your null hypothesis.

In order to reject or accept the null hypothesis, you must make a subjective judgment regarding the degree of risk you are willing to accept for being wrong. This degree of risk is often given in terms of critical values and/or confidence level.

To give an example: the critical Z score values when using a 95% confidence level are -1.96 and +1.96 standard deviations. If your Z score is between -1.96 and +1.96 you cannot reject your null hypothsis; the pattern exhibited is a pattern that could very likely be one version of a random pattern. If the Z score falls outside that range(for example -2.5 or +5.4), the pattern exhibited is probably too unusual to be just another version of random chance. If this is the case, it is possible to reject the null hypothesis and proceed with figuring out what might be causing either the statistically significant clustered or statistically significant dispersed pattern.

R - spdep

```
>
> library(spdep)


Le chargement a nécessité le package : tripack
Le chargement a nécessité le package : sp
Le chargement a nécessité le package : maptools
Le chargement a nécessité le package : foreign
Le chargement a nécessité le package : boot
Le chargement a nécessité le package : Matrix
Le chargement a nécessité le package : lattice

Attachement du package : 'lattice'


        The following object(s) are masked from package:boot :

         melanoma

>
```

Example

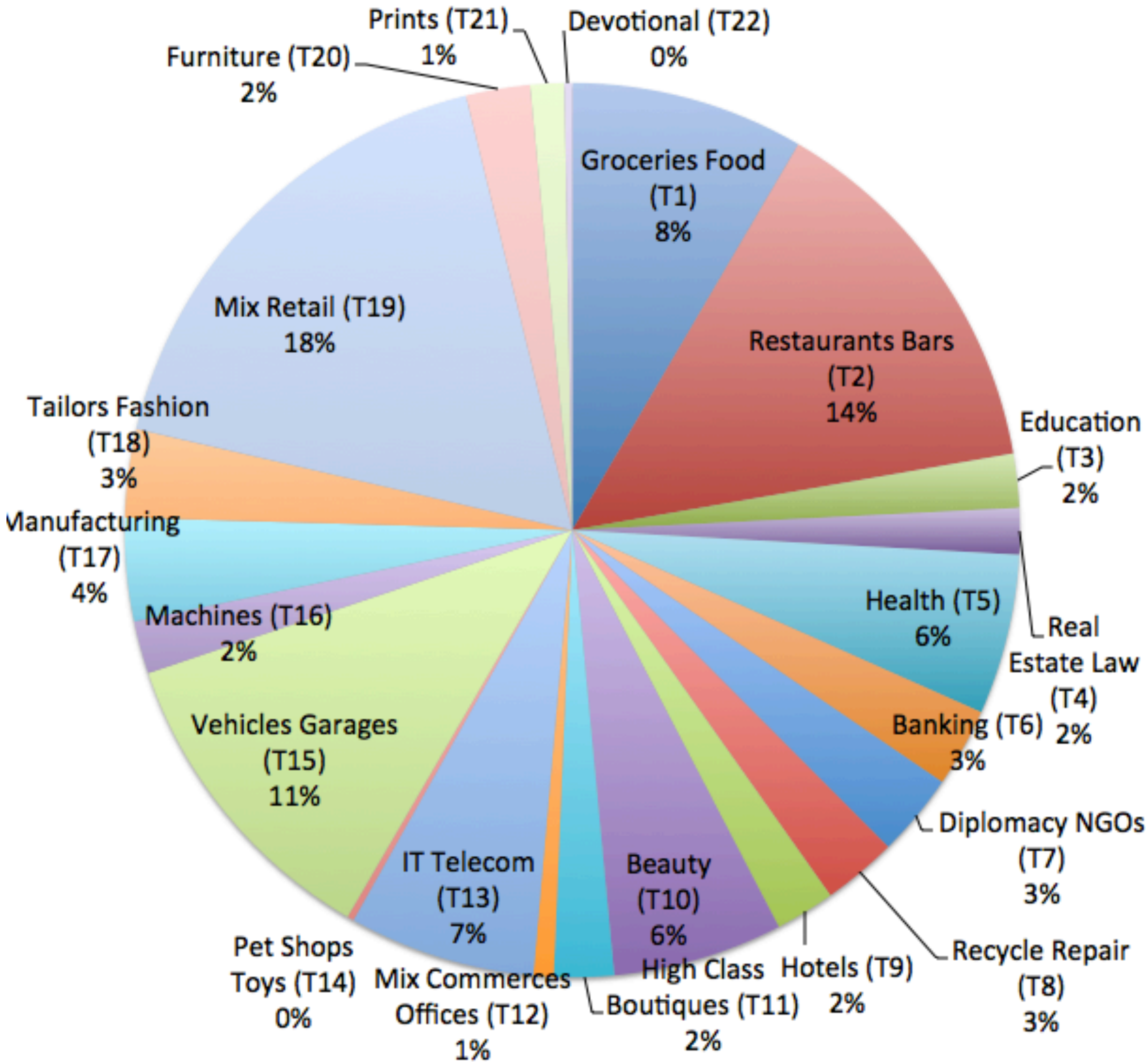
- Caruso, Kolnberger (ongoing)
- Retail location in Phnom Penh

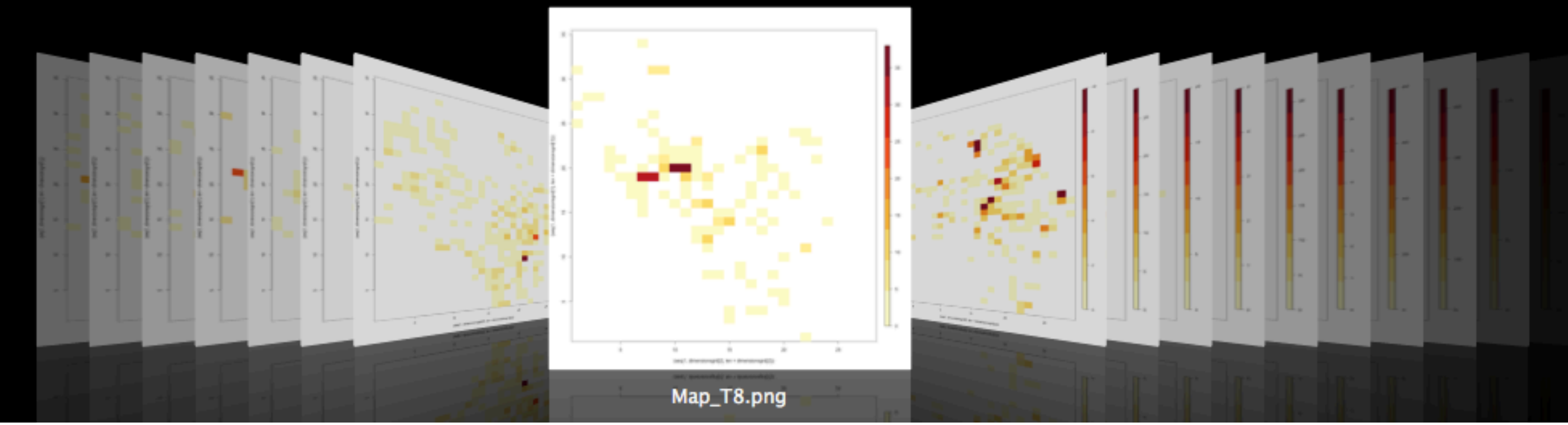

The „pavement economy“
of the *Chinese shophouse*:

Adobe Illustrator - [Phnom Penh Original Juni 2010 zwischengesichert.ai @ 150% (RGB/Preview)]
File Edit Object Type Select Filter Effect View Window Help
Russey
Market
Charles de Gaulle Blvd.
Sok Hok Street (107)
Ung Po Kun S
Baktouk Primary School
PPS - Schule
Camb.-Japanese Center
P K Conversation with Foreigner
Newton Thilay School
Sovannphumi School
Teenfashion
DLA Mod. Sewing SC
ambulanter Frischmarkt
Spanferkel
Propangas
Blumen
Gewürze
Nähmaschinen
Puppenhäuser
Sonnenschirme
Matratzen
Hausaltäre
Vitrinenbau
Landmasch.
Konditorwaren
Folien
Reifen u.a.
Devotional
Weißblech
Schminkzeug
Luxusleuchten
A/C & Waschmasch
Mart
Clinic
150%
Open

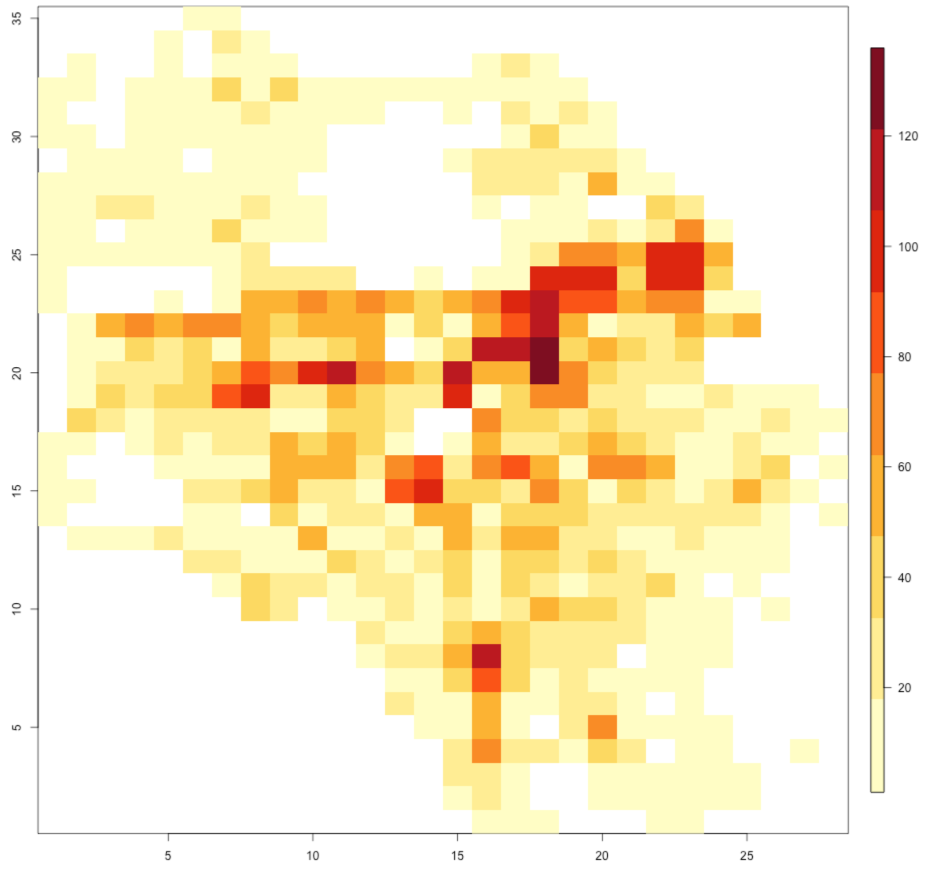
Dataset to explore

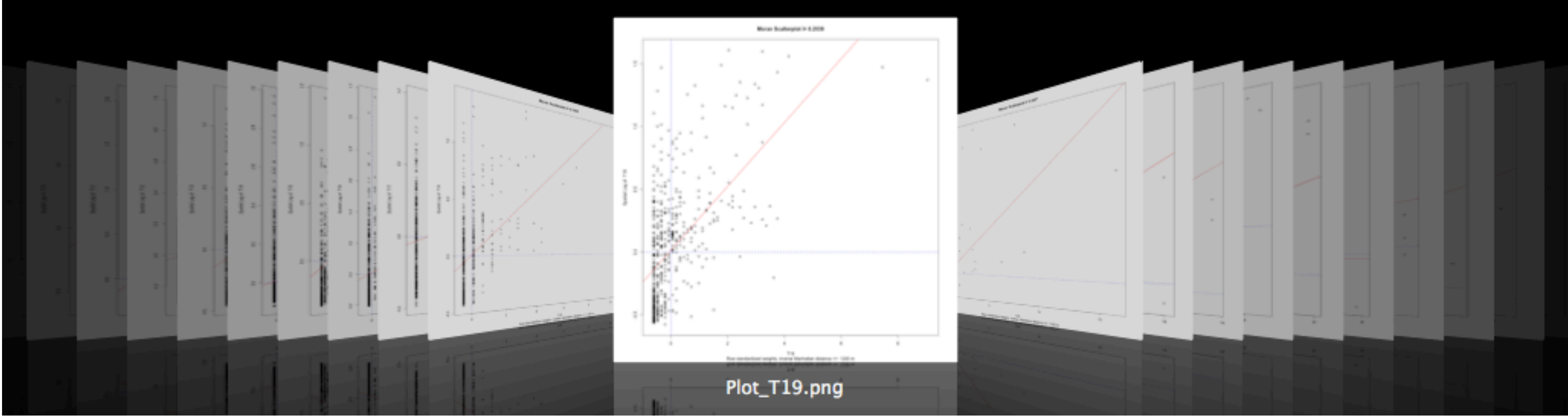
- 14549 businesses
- 558 blocks with at least one shop
- 200 m blocks
- 111 subcategories
- 22 categories



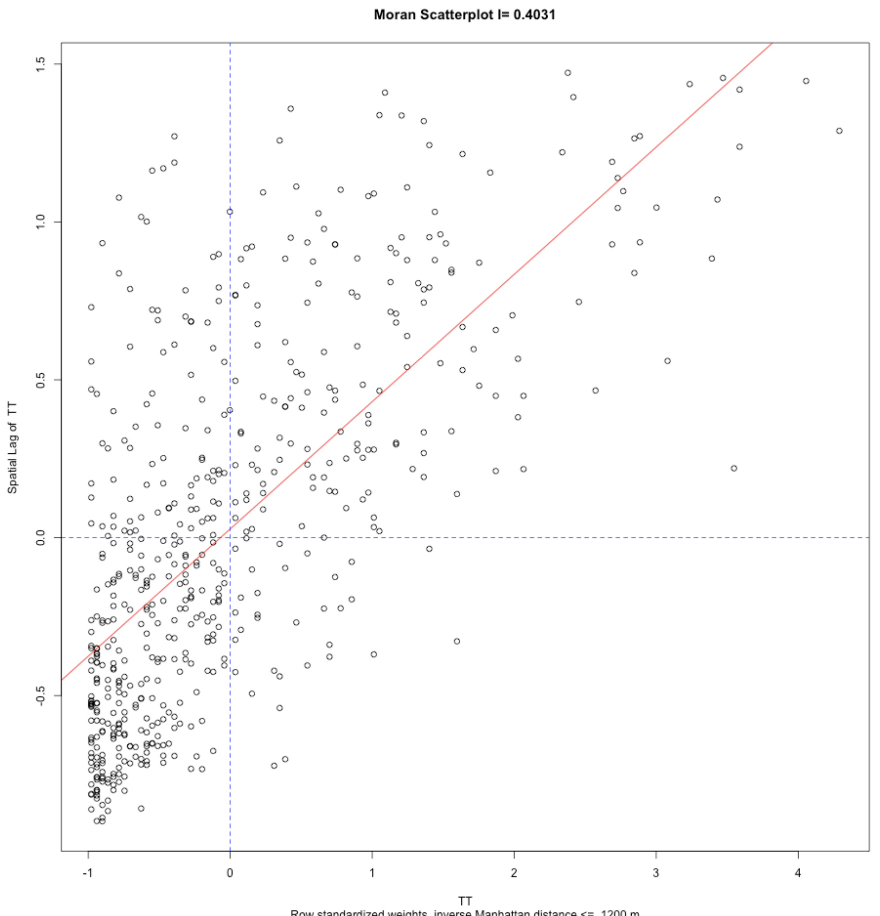


- Total business density

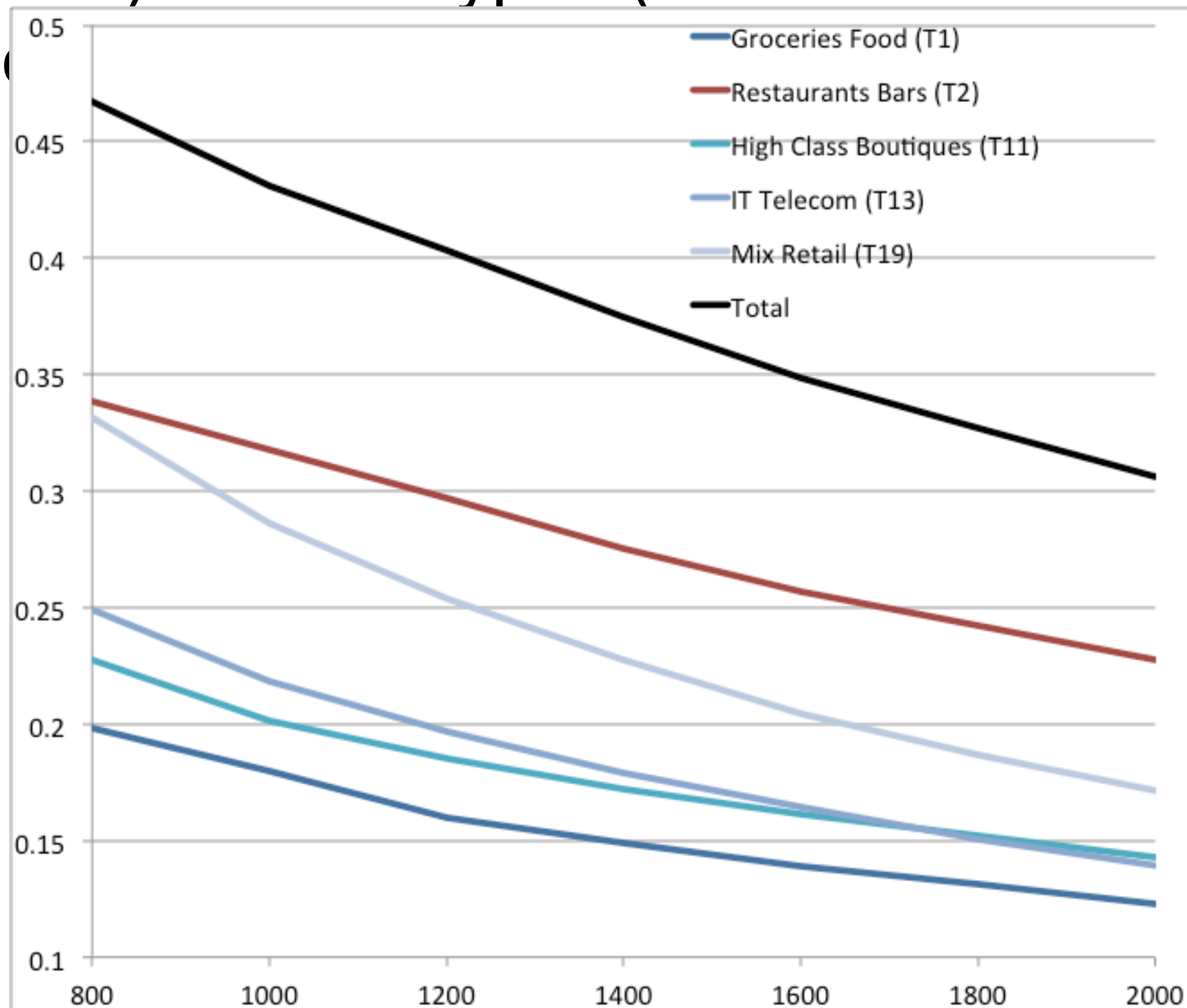


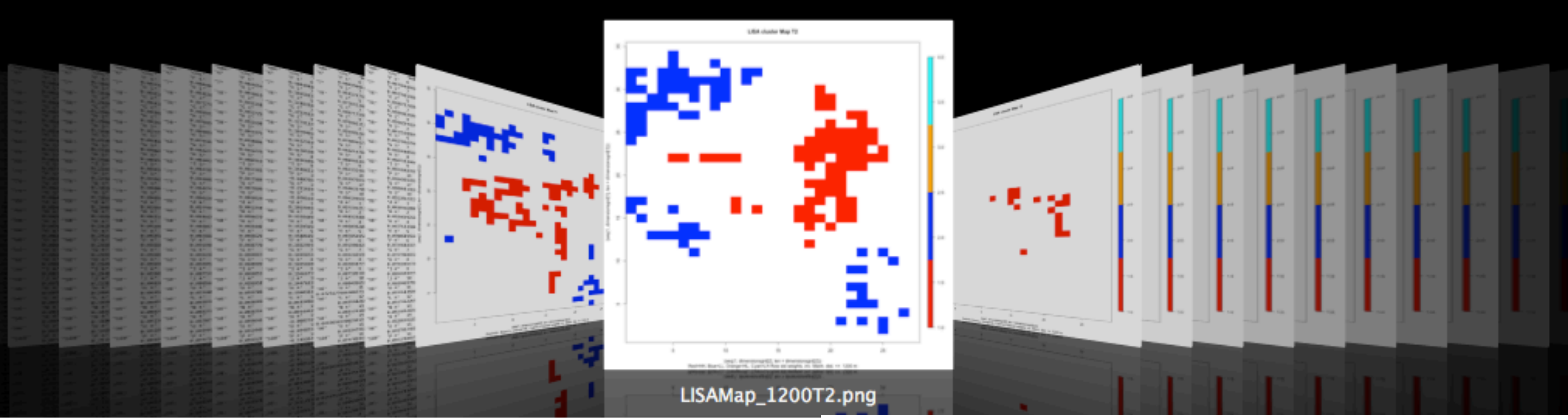


- Total business Moran's scatterplot



- Moran's I vs distance lags
  - Significant positive global autocorrelation of (almost) all retail types (robust across 4 to 8 bl





- Total business LISA clusters
  - 1 strong central HH cluster spreading W and S
  - Significant LL for newer residential dvlpt (NW) and river bank (SE)

- No significant HL LH across categories and distance bands !