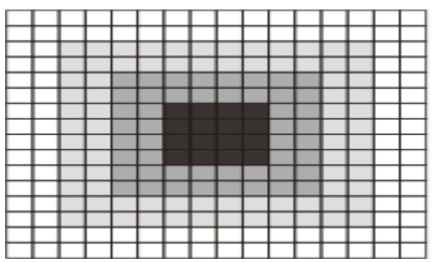
# Applied Spatiotemporal Data Mining

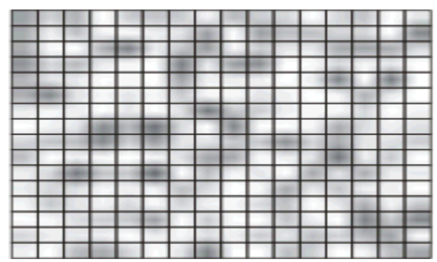
Statistical Analysis of Areal Data

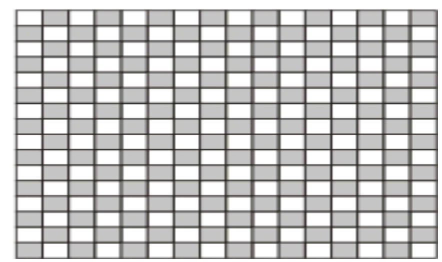
Guofeng Cao
Department of Geosciences
Texas Tech University

# Spatial Autocorrelation

- Tobler's first law of geography
- Spatial auto/cross correlation







If like values tend to cluster together, then the field exhibits high **positive spatial autocorrelation**

If there is no apparent relationship between attribute value and location then there is **zero spatial autocorrelation**

If like values tend to be located away from each other, then there is **negative spatial autocorrelation**

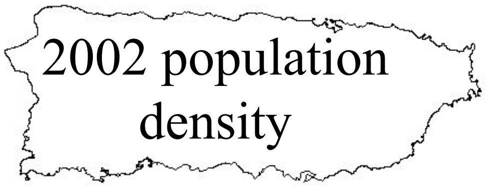
# Spatial Autocorrelation

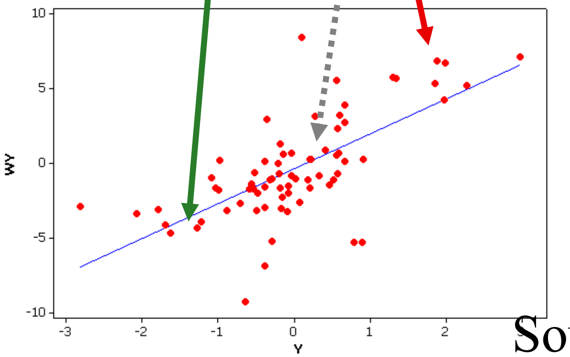
- Spatial autocorrelationship is everywhere
  - Spatial point pattern
    - K, F, G functions
    - Kernel functions
  - Geostatistical data
  - Areal/lattice (this topic)

# Spatial Autocorrelation of Areal Data

# ***Positive spatial autocorrelation***

- high values surrounded by nearby high values
- intermediate values surrounded by nearby intermediate values
- low values surrounded by nearby low values

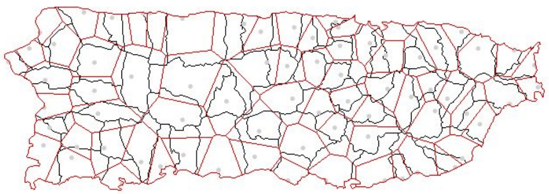



Source: Ron Briggs of UT Dallas

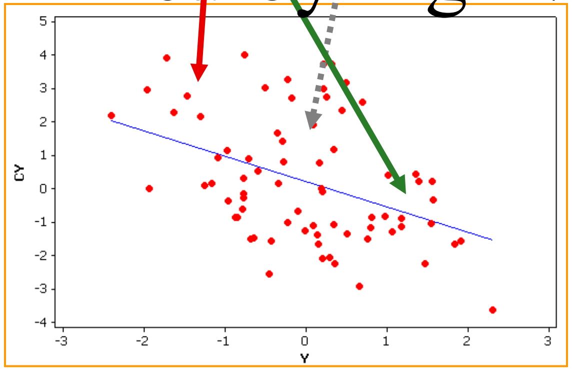
## ***Negative spatial autocorrelation***

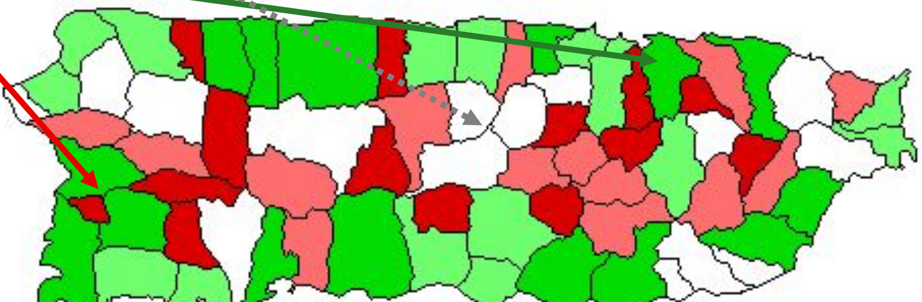
competition for space



Grocery store density

- high values surrounded by nearby low values
- intermediate values surrounded by nearby intermediate values
- low values surrounded by nearby high values



Source: Ron Briggs of UT Dallas

# Spatial Weight Matrix

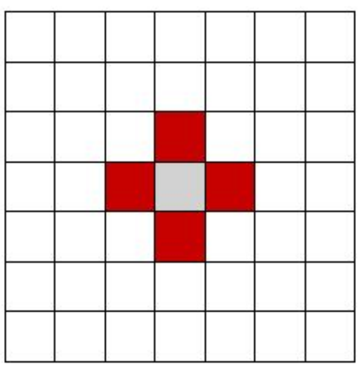
- **Core** concept in statistical analysis of areal data
- Two steps involved:
  - define which relationships between observations are to be given a nonzero weight, i.e., define spatial neighbors
  - assign weights to the neighbors
- Making the neighbors and weights is not easy as it seems to be
  - Which states are near Texas?

# Spatial Neighbors

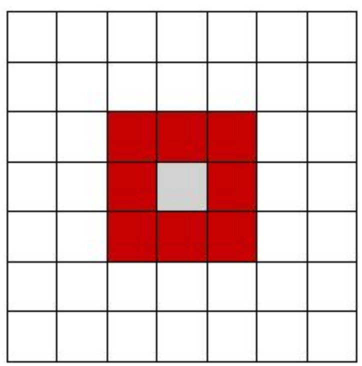
- **Contiguity-based neighbors**
  - Zone *i* and j are neighbors if zone i is contiguity or adjacent to zone *j*
  - But what constitutes contiguity?
- **Distance-based neighbors**
  - Zone i and j are neighbors if the distance between them are less than the threshold distance
  - But what distance do we use?

# Contiguity-based Spatial Neighbors

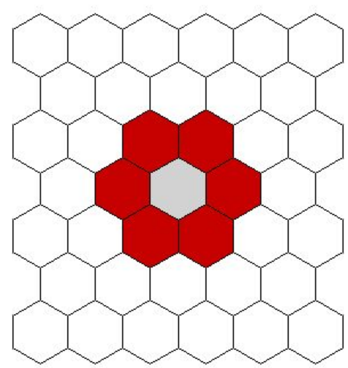
- Sharing a border or boundary
  - Rook: sharing a border
  - Queen: sharing a border <u>or</u> a point



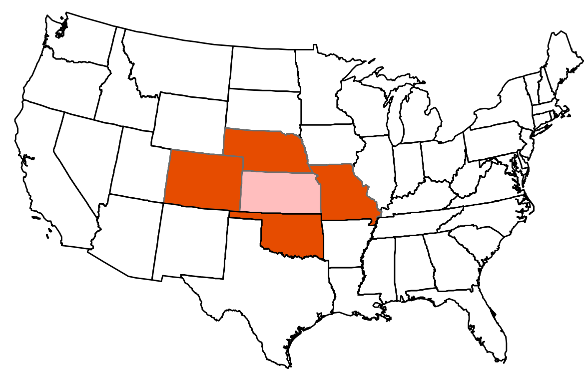
rook



queen



Hexagons



Irregular

Which use?

# Example

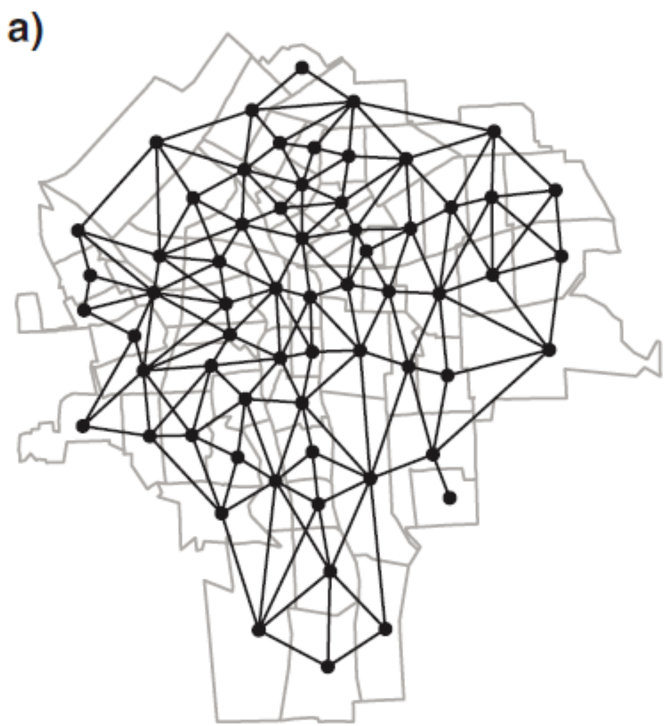
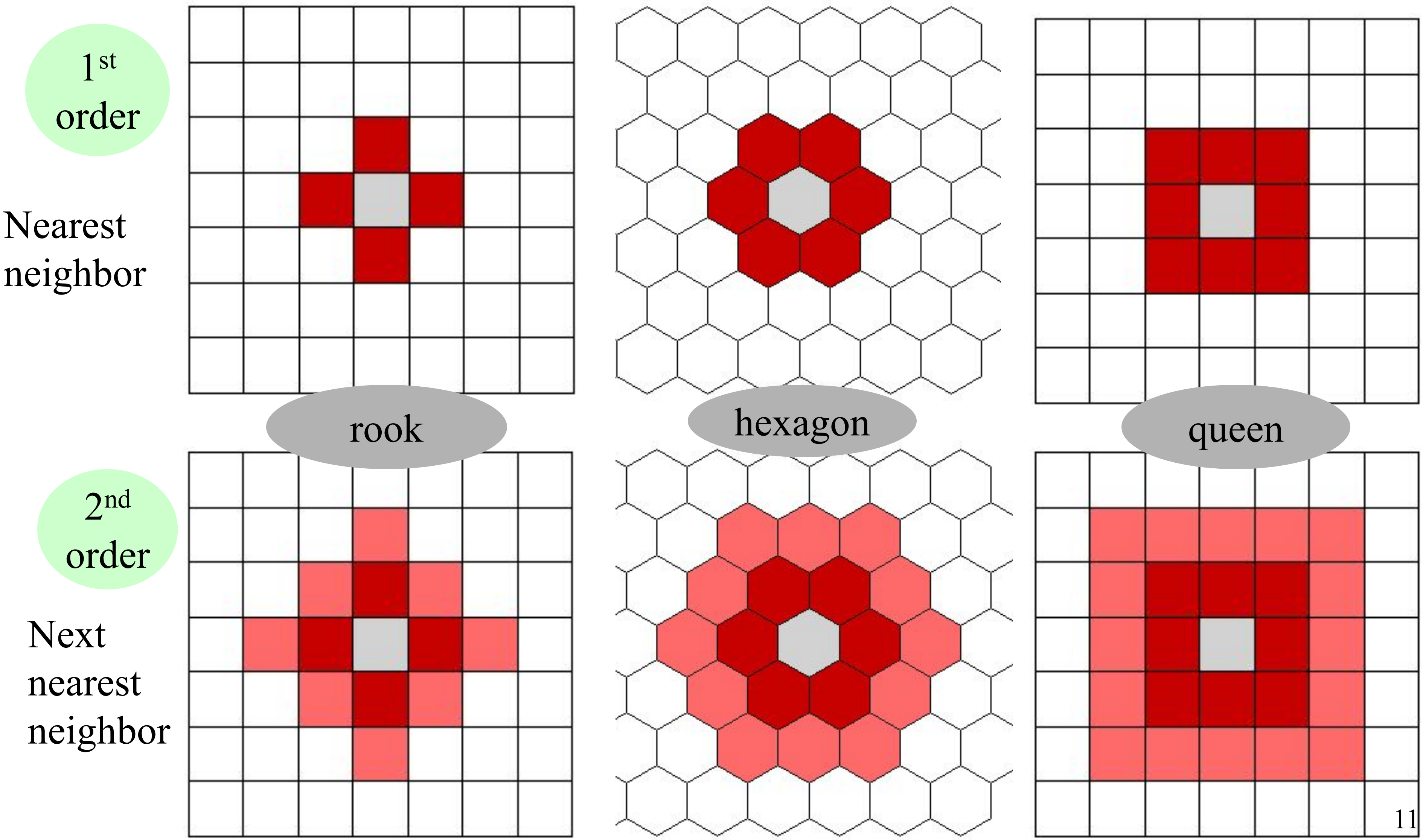



**Fig. 9.3.** (**a**) Queen-style census tract contiguities, Syracuse; (**b**) Rook-style contiguity differences shown as thicker lines

Source: Bivand and Pebesma and Gomez-Rubio

# Higher-Order Contiguity



# Distance-based Neighbors

- How to measure distance between polygons?
- Distance metrics
  - 2D Cartesian distance (projected data)
  - 3D spherical distance/great-circle distance (lat/long data)
    - Haversine formula

*Haversine formula:*

$$a = \sin^2(\Delta\varphi/2) + \cos(\varphi_1).\cos(\varphi_2).\sin^2(\Delta\lambda/2)$$
$$c = 2.\mathrm{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R.c$$

*where $\varphi$ is latitude, $\lambda$ is longitude, $R$ is earth's radius (mean radius = 6,371km)*

# Distance-based Neighbors

- k-nearest neighbors


**Fig. 9.5.** (**a**) $k = 1$ neighbours; (**b**) $k = 2$ neighbours; (**c**) $k = 4$ neighbours

Source: Bivand and Pebesma and Gomez-Rubio

# Distance-based Neighbors

- thresh-hold distance (buffer)

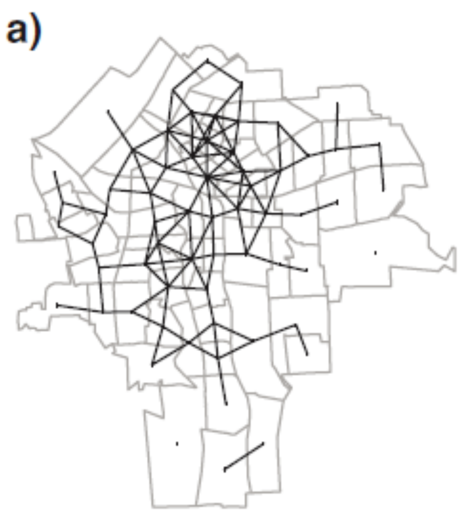
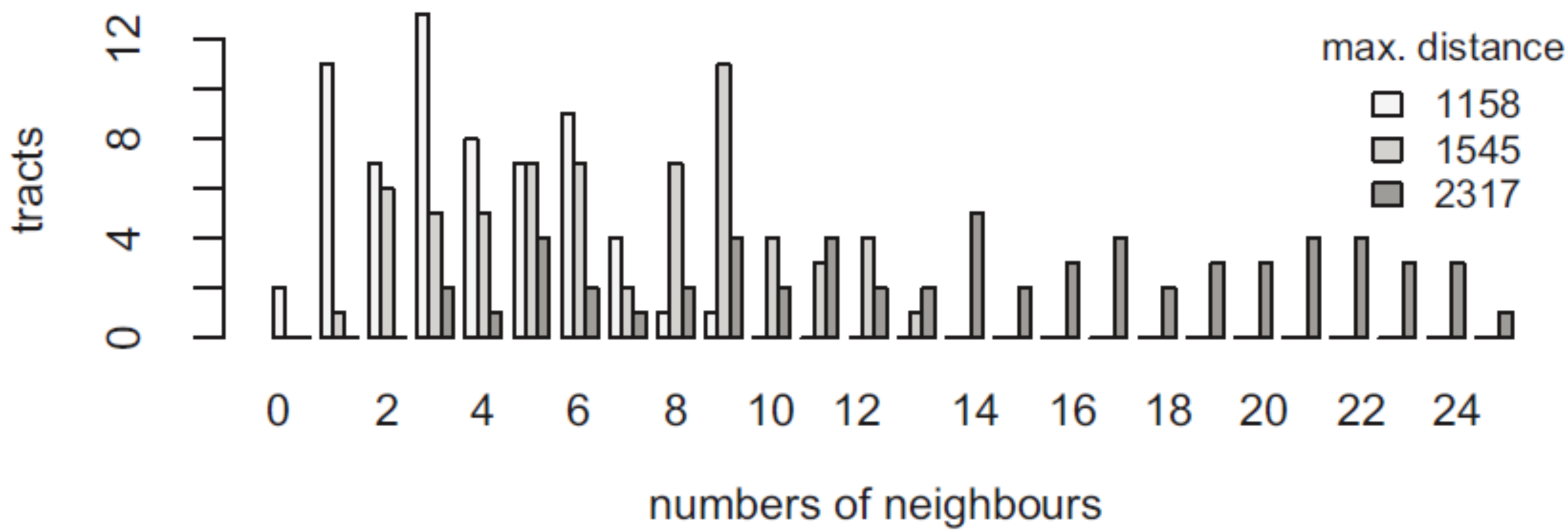



**Fig. 9.6.** (**a**) Neighbours within 1,158 m; (**b**) neighbours within 1,545 m; (**c**) neigh-
bours within 2,317 m

Source: Bivand and Pebesma and Gomez-Rubio

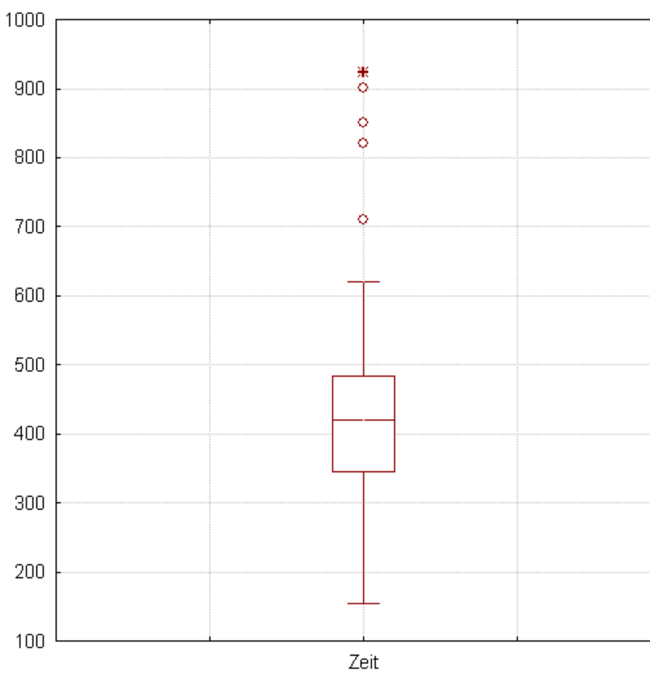
# Neighbor/Connectivity Histogram




Source: Bivand and Pebesma and Gomez-Rubio

# Recap: Box-plot

- Help indicate the degree of dispersion and skewness and identify outliers
  - Non-parametric
  - 25%, 50%, 75% percentiles
  - end of the hinge could mean differently depending on implementation
  - Points outside the range are usually taken as outliers



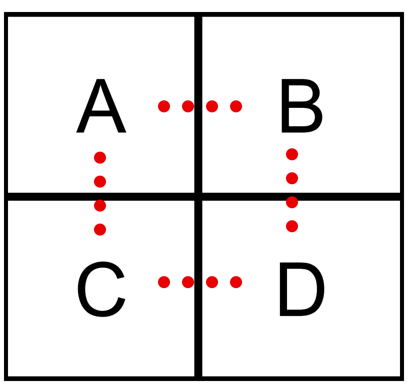
# Spatial Weight Matrix

- Spatial weights can be seen as a list of weights indexed by a list of neighbors
- If zone j is not a neighbor of zone i, weights Wij will set to zero
  - The weight matrix can be illustrated as an image
  - Sparse matrix

# A Simple Example for Rook case

- Matrix contains a:
  - 1 if share a border
  - 0 if do not share a border

4 areal units




Common border

4x4 matrix

$$
\mathbf{W} = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1 & 1 & 0 \\ B & 1 & 0 & 0 & 1 \\ C & 1 & 0 & 0 & 1 \\ D & 0 & 1 & 1 & 0 \end{array}
$$

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Washington | |1| | | |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 Oregon |1| |1| |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 California | |1| |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 Arizona | | |1| |1| | | |1|1| | |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 Nevada | |1|1|1| |1| | |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 Idaho |1|1| | |1| |1|1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 Montana | | | | | |1| |1| | | | | | | |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 Wyoming | | | | | |1|1| |1| | | |1| |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 Utah | | | |1|1|1| |1| |1| | |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 New Mexico | | | |1| | | | |1| |1|1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 Texas | | | | | | | | | |1| |1| | | | | | | | |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 12 Oklahoma | | | | | | | | | |1|1| |1|1| | | | | |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 13 Colorado | | | |1| | | |1|1|1| |1| |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 14 Kansas | | | | | | | | | | | |1|1| |1| | | | |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15 Nebraska | | | | | | | |1| | | | |1|1| |1| | |1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 16 South Dakota | | | | | | |1|1| | | | | | |1| |1|1|1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 17 North Dakota | | | | | | |1| | | | | | | | |1| |1| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 18 Minnesota | | | | | | | | | | | | | | | |1|1| |1| | | | | | | |1| | | | | | | | | | | | | | | | | | | | | | |
| 19 Iowa | | | | | | | | | | | | | | |1|1| |1| |1| | | | | |1|1| | | | | | | | | | | | | | | | | | | | | | |
| 20 Missouri | | | | | | | | | | | |1| |1|1| | | |1| |1| | |1|1|1| | | | | | | | | | | | | | | | | | | | | | | |
| 21 Arkansas | | | | | | | | | | |1|1| | | | | | | |1| |1|1|1| | | | | | | | | | | | | | | | | | | | | | | | | |
| 22 Louisiana | | | | | | | | | | |1| | | | | | | | | |1| |1| | | | | | | | | | | | | | | | | | | | | | | | | | |
| 23 Mississippi | | | | | | | | | | | | | | | | | | | | |1|1| |1| | | | | | | | |1| | | | | | | | | | | | | | | | |
| 24 Tennessee | | | | | | | | | | | | | | | | | | | |1|1| |1| |1| | | | | | | |1|1| |1|1| | | | | | | | | | | | |
| 25 Kentucky | | | | | | | | | | | | | | | | | | | |1| | | |1| |1| | |1|1|1| | | | | |1| | | | | | | | | | | | |
| 26 Illinois | | | | | | | | | | | | | | | | | | |1|1| | | | |1| |1| |1| | | | | | | | | | | | | | | | | | | | |
| 27 Wisconsin | | | | | | | | | | | | | | | | | |1|1| | | | | | | |1| |1| | | | | | | | | | | | | | | | | | | | |
| 28 Michigan | | | | | | | | | | | | | | | | | | | | | | | | | | |1| |1|1| | | | | | | | | | | | | | | | | | | |
| 29 Indiana | | | | | | | | | | | | | | | | | | | | | | | | |1|1| |1| |1| | | | | | | | | | | | | | | | | | | |
| 30 Ohio | | | | | | | | | | | | | | | | | | | | | | | | |1| | |1|1| |1| | | | | | | | | | |1| | | | | | | |
| 31 West Virginia | | | | | | | | | | | | | | | | | | | | | | | | |1| | | | |1| | | | | | |1|1| | | |1| | | | | | | |
| 32 Florida | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1|1| | | | | | | | | | | | | | | |
| 33 Alabama | | | | | | | | | | | | | | | | | | | | | | |1|1| | | | | | | |1| |1| | | | | | | | | | | | | | | |
| 34 Georgia | | | | | | | | | | | | | | | | | | | | | | | |1| | | | | | | |1|1| |1|1| | | | | | | | | | | | | |
| 35 South Carolina | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| |1| | | | | | | | | | | | | |
| 36 North Carolina | | | | | | | | | | | | | | | | | | | | | | | |1| | | | | | | | | |1|1| |1| | | | | | | | | | | | |
| 37 Virginia | | | | | | | | | | | | | | | | | | | | | | | |1|1| | | | | |1| | | | |1| |1| |1| | | | | | | | | |
| 38 Maryland | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| | | | | |1| |1|1| |1| | | | | | | |
| 39 Delaware | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| | |1|1| | | | | | | |
| 40 District of Columbia | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1|1| | | | | | | | | | | |
| 41 New Jersey | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| | |1|1| | | | | | |
| 42 Pennsylvania | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1|1| | | | | | |1|1| |1| |1| | | | | | |
| 43 New York | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1|1| |1| |1| |1| |
| 44 Connecticut | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| |1|1| | | |
| 45 Rhode Island | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| |1| | | |
| 46 Massachussets | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1|1|1| |1|1| |
| 47 New Hampshire | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| |1|1|
| 48 Vermont | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| | |1|1| | |
| 49 Maine | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |1| | |

| Sparse Contiguity Matrix for US States -- obtained from Anselin's web site (see powerpoint for link) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Name** | **Fips** | **Ncount** | **N1** | **N2** | **N3** | **N4** | **N5** | **N6** | **N7** | **N8** |
| Alabama | 1 | 4 | 28 | 13 | 12 | 47 | | | | |
| Arizona | 4 | 5 | 35 | 8 | 49 | 6 | 32 | | | |
| Arkansas | 5 | 6 | 22 | 28 | 48 | 47 | 40 | 29 | | |
| California | 6 | 3 | 4 | 32 | 41 | | | | | |
| Colorado | 8 | 7 | 35 | 4 | 20 | 40 | 31 | 49 | 56 | |
| Connecticut | 9 | 3 | 44 | 36 | 25 | | | | | |
| Delaware | 10 | 3 | 24 | 42 | 34 | | | | | |
| District of Columbia | 11 | 2 | 51 | 24 | | | | | | |
| Florida | 12 | 2 | 13 | 1 | | | | | | |
| Georgia | 13 | 5 | 12 | 45 | 37 | 1 | 47 | | | |
| Idaho | 16 | 6 | 32 | 41 | 56 | 49 | 30 | 53 | | |
| Illinois | 17 | 5 | 29 | 21 | 18 | 55 | 19 | | | |
| Indiana | 18 | 4 | 26 | 21 | 17 | 39 | | | | |
| Iowa | 19 | 6 | 29 | 31 | 17 | 55 | 27 | 46 | | |
| Kansas | 20 | 4 | 40 | 29 | 31 | 8 | | | | |
| Kentucky | 21 | 7 | 47 | 29 | 18 | 39 | 54 | 51 | 17 | |
| Louisiana | 22 | 3 | 28 | 48 | 5 | | | | | |
| Maine | 23 | 1 | 33 | | | | | | | |
| Maryland | 24 | 5 | 51 | 10 | 54 | 42 | 11 | | | |
| Massachusetts | 25 | 5 | 44 | 9 | 36 | 50 | 33 | | | |
| Michigan | 26 | 3 | 18 | 39 | 55 | | | | | |
| Minnesota | 27 | 4 | 19 | 55 | 46 | 38 | | | | |
| Mississippi | 28 | 4 | 22 | 5 | 1 | 47 | | | | |
| Missouri | 29 | 8 | 5 | 40 | 17 | 21 | 47 | 20 | 19 | 31 |
| Montana | 30 | 4 | 16 | 56 | 38 | 46 | | | | |
| Nebraska | 31 | 6 | 29 | 20 | 8 | 19 | 56 | 46 | | |
| Nevada | 32 | 5 | 6 | 4 | 49 | 16 | 41 | | | |
| New Hampshire | 33 | 3 | 25 | 23 | 50 | | | | | |
| New Jersey | 34 | 3 | 10 | 36 | 42 | | | | | |
| New Mexico | 35 | 5 | 48 | 40 | 8 | 4 | 49 | | | |
| New York | 36 | 5 | 34 | 9 | 42 | 50 | 25 | | | |
| North Carolina | 37 | 4 | 45 | 13 | 47 | 51 | | | | |
| North Dakota | 38 | 3 | 46 | 27 | 30 | | | | | |
| Ohio | 39 | 5 | 26 | 21 | 54 | 42 | 18 | | | |
| Oklahoma | 40 | 6 | 5 | 35 | 48 | 29 | 20 | 8 | | |
| Oregon | 41 | 4 | 6 | 32 | 16 | 53 | | | | |
| Pennsylvania | 42 | 6 | 24 | 54 | 10 | 39 | 36 | 34 | | |
| Rhode Island | 44 | 2 | 25 | 9 | | | | | | |
| South Carolina | 45 | 2 | 13 | 37 | | | | | | |
| South Dakota | 46 | 6 | 56 | 27 | 19 | 31 | 38 | 30 | | |
| Tennessee | 47 | 8 | 5 | 28 | 1 | 37 | 13 | 51 | 21 | 29 |
| Texas | 48 | 4 | 22 | 5 | 35 | 40 | | | | |
| Utah | 49 | 6 | 4 | 8 | 35 | 56 | 32 | 16 | | |
| Vermont | 50 | 3 | 36 | 25 | 33 | | | | | |
| Virginia | 51 | 6 | 47 | 37 | 24 | 54 | 11 | 21 | | |
| Washington | 53 | 2 | 41 | 16 | | | | | | |
| West Virginia | 54 | 5 | 51 | 21 | 24 | 39 | 42 | | | |
| Wisconsin | 55 | 4 | 26 | 17 | 19 | 27 | | | | |
| Wyoming | 56 | 6 | 49 | 16 | 31 | 8 | 46 | 30 | | |

# Style of Spatial Weight Matrix

- Row
  - a weight of unity for each neighbor relationship
- Row standardization
  - Symmetry not guaranteed
  - can be interpreted as allowing the calculation of average values across neighbors
- General spatial weights based on distances

# Row vs. Row standardization

| | | |
|---|---|---|
| A | B | C |
| D | E | F |

Divide each number by the **row sum**

Total number of neighbors
--some have more than others

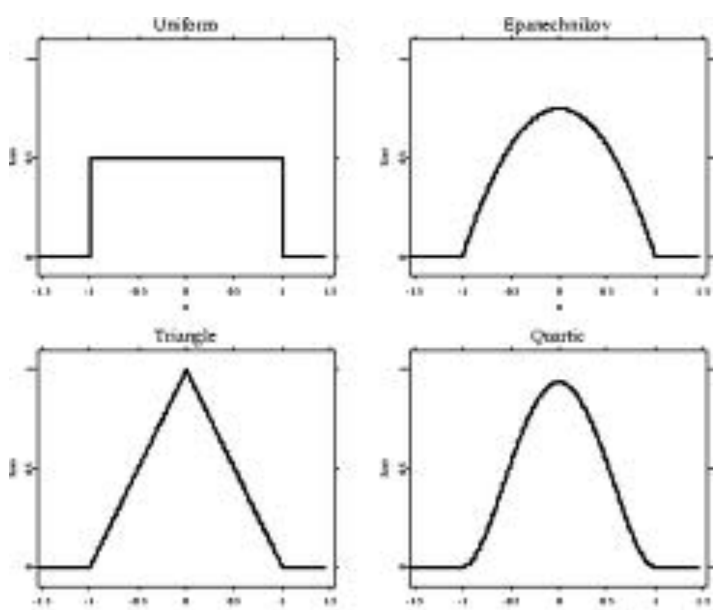
| | **A** | **B** | **C** | **D** | **E** | **F** | **Row Sum** |
|---|---|---|---|---|---|---|---|
| **A** | 0 | 1 | 0 | 1 | 0 | 0 | 2 |
| **B** | 1 | 0 | 1 | 0 | 1 | 0 | ***3*** |
| **C** | 0 | 1 | 0 | 0 | 0 | 1 | 2 |
| **D** | 1 | 0 | 0 | 0 | 1 | 0 | 2 |
| **E** | 0 | 1 | 0 | 1 | 0 | 1 | ***3*** |
| **F** | 0 | 0 | 1 | 0 | 1 | 0 | 2 |

Row standardized
--usually use this

| | **A** | **B** | **C** | **D** | **E** | **F** | **Row Sum** |
|---|---|---|---|---|---|---|---|
| **A** | 0.0 | 0.5 | 0.0 | 0.5 | 0.0 | 0.0 | 1 |
| **B** | 0.3 | 0.0 | 0.3 | 0.0 | 0.3 | 0.0 | 1 |
| **C** | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 | 0.5 | 1 |
| **D** | 0.5 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 1 |
| **E** | 0.0 | 0.3 | 0.0 | 0.3 | 0.0 | 0.3 | 1 |
| **F** | 0.0 | 0.0 | 0.5 | 0.0 | 0.5 | 0.0 | 1 |

# General Spatial Weights Based on Distance

- Decay functions of distance
  - Most common choice is the inverse (reciprocal) of the distance between locations i and j ($w_{ij} = 1/d_{ij}$)
  - Other functions also used
    - inverse of squared distance ($w_{ij} = 1/d_{ij}^2$), or
    - negative exponential ($w_{ij} = e^{-d}$ *or* $w_{ij} = e^{-d^2}$)



# Measure of Spatial Autocorrelation

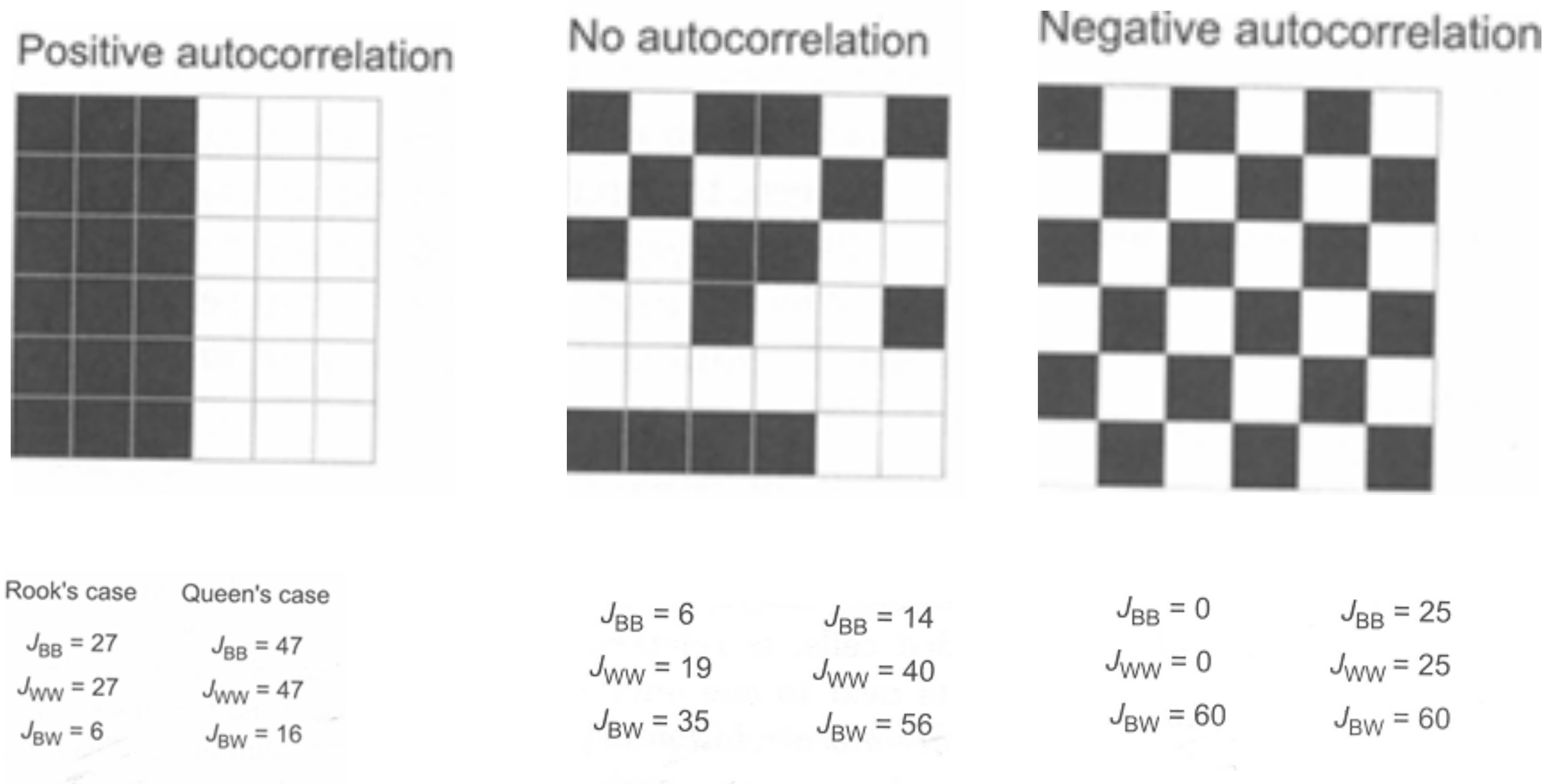
# Global Measures and Local Measures

- Global Measures
  - A single value which applies to the entire data set
    - The same pattern or process occurs over the entire geographic area
    - An average for the entire area
- Local Measures
  - A value calculated for <u>each</u> observation unit
    - Different patterns or processes may occur in different parts of the region
    - A unique number for each location
- Global measures usually can be decomposed into a combination of local measures

# Global Measures and Local Measures

- Global Measures
  - Join Count
  - Moran's I
- Local Measures
  - Local Moran's I

# Join (or Joint or Joins) Count Statistic




– 60 for Rook Case
– 110 for Queen Case

# Join Count: Test Statistic

Test Statistic given by: Z= $\dfrac{\text{Observed - Expected}}{\text{SD of Expected}}$

***Expected*** = random pattern generated by tossing a coin in each cell.

Expected given by:

$$E(J_{BB}) = kp_B^2$$

$$E(J_{WW}) = kp_W^2$$

$$E(J_{BW}) = 2kp_Bp_W$$

Standard Deviation of Expected (standard error) given by:

$$E(s_{BB}) = \sqrt{kp_B^2 + 2mp_B^3 - (k+2m)p_B^4}$$

$$E(s_{WW}) = \sqrt{kp_W^2 + 2mp_W^3 - (k+2m)p_W^4}$$

$$E(s_{BW}) = \sqrt{2(k+m)p_Bp_W - 4(k+2m)p_B^2\,p_W^2}$$

Where: k is the total number of joins (neighbors)

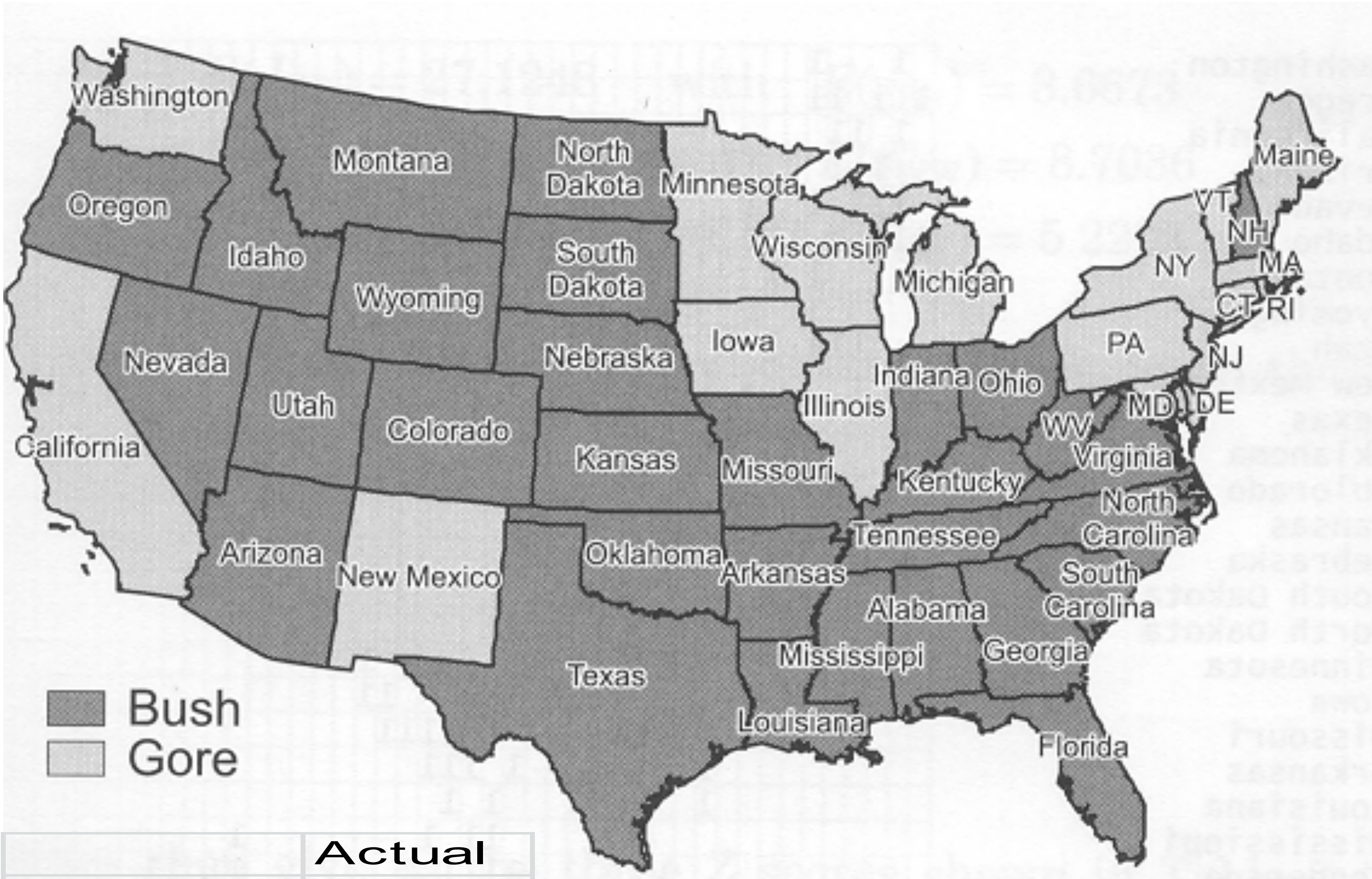
$p_B$ is the expected proportion Black, if random

$p_W$ is the expected proportion White

m is calculated from k according to:

$$m = \frac{1}{2}\sum_{i=1}^{n} k_i(k_i - 1)$$

# Gore/Bush Presidential Election 2000




| | Actual |
|---|---|
| Jbb | 60 |
| Jgg | 21 |
| Jbg | 28 |
| Total | 109 |
| | |

# Join Count Statistic for Gore/Bush 2000 by State

| candidates | probability |
|---|---|
| Bush | 0.49885 |
| Gore | 0.50115 |
| | |

| | Actual | Expected | Stan Dev | Z-score |
|---|---|---|---|---|
| Jbb | 60 | 27.125 | 8.667 | 3.7930 |
| Jgg | 21 | 27.375 | 8.704 | -0.7325 |
| Jbg | 28 | 54.500 | 5.220 | -5.0763 |
| Total | 109 | 109.000 | | |
| | | | | |

- The expected number of joins is calculated based on the proportion of votes each received in the election (for Bush = 109*.499*.499=27.125)
- There are far <u>more</u> Bush/Bush joins (actual = 60) than would be expected (27)
  - Positive autocorrelation
- There are far <u>fewer</u> Bush/Gore joins (actual = 28) than would be expected (54)
  - Positive autocorrelation
- No strong clustering evidence for Gore (actual = 21 slightly less than 27.375)

# Moran's I

- The most common measure of Spatial Autocorrelation
- Use for points <u>or</u> polygons
  - Join Count statistic only for polygons
- Use for a continuous variable (any value)
  - Join Count statistic only for binary variable (1,0)

Patrick Alfred Pierce Moran (1917-1988)

# Formula for Moran's I

$$I = \frac{N\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Where:

$N$ is the number of observations (points or polygons)
$\bar{x}$ is the mean of the variable
$X_i$ is the variable value at a particular location
$X_j$ is the variable value at another location
$W_{ij}$ is a weight indexing location of *i* relative to *j*

# Moran's I

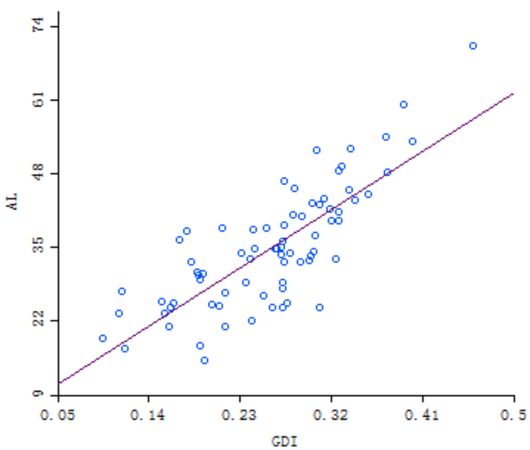
- Expectation of Moran's I under no spatial autocorrelation

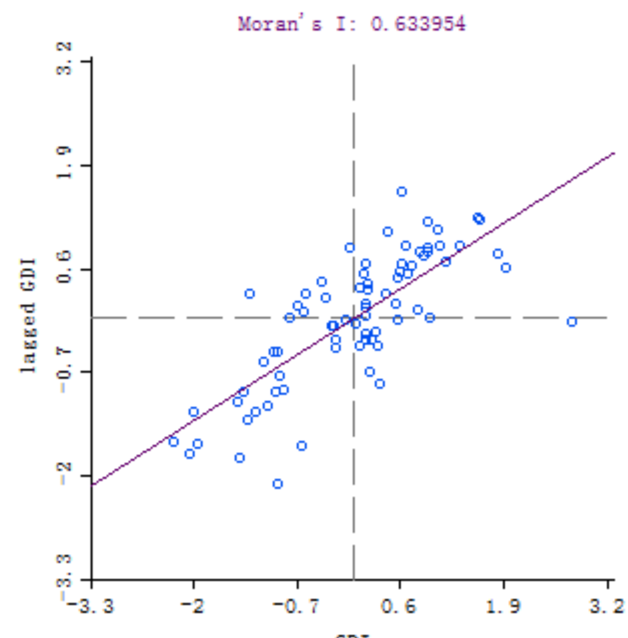
$$E(\mathrm{I}) = -1/(\mathrm{N}-1)$$

- Variance of Moran's is complex and exact equation is given at textbook d&G&L
- [-1, 1]

# Moran's *I* and Correlation Coefficient

- **Correlation Coefficient [-1, 1]**
  - Relationship between <u>two</u> different variables
- **Moran's I [-1, 1]**
- Spatial autocorrelation and often involves <u>one</u> (spatially indexed) variable only
- Correlation between observations of a spatial variable at location X and "spatial lag" of X formed by averaging all the observation at neighbors of X



# Correlation Coefficient

$$\frac{\sum_{i=1}^{n} 1(y_i - \bar{y})(x_i - \bar{x})/n}{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}$$

Note the similarity of the numerator (top) to the measures of spatial association discussed earlier if we view Yi as being the Xi for the neighboring polygon

**(see next slide)**

$$\frac{N\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x}) / \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}$$

# Spatial auto-correlation



Source: Ron Briggs of UT Dallas

## Correlation Coefficient

$$\frac{\sum_{i=1}^{n} 1(y_i - \bar{y})(x_i - \bar{x})/n}{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}$$

Spatial weights

Yi is the Xi for the neighboring polygon

$$\frac{N\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})/\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}$$

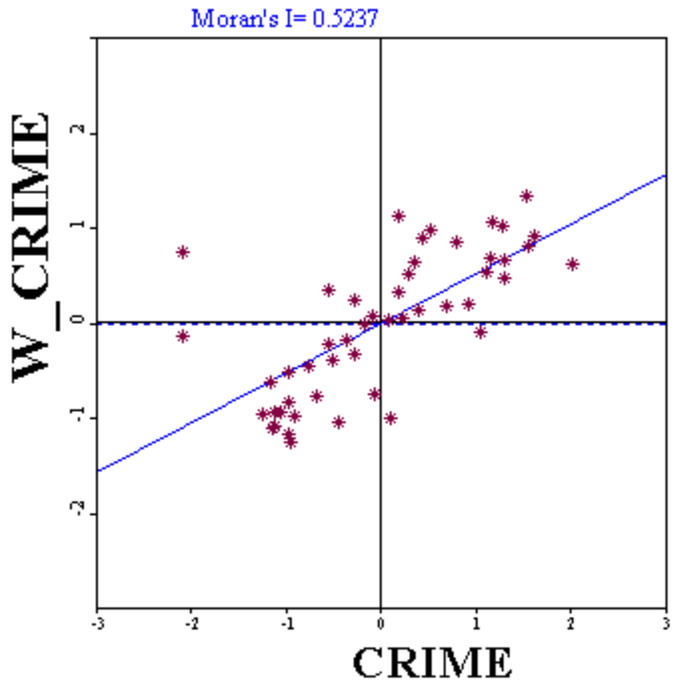
## Moran's I



Source: Ron Briggs of UT Dallas

# Moran Scatter Plots

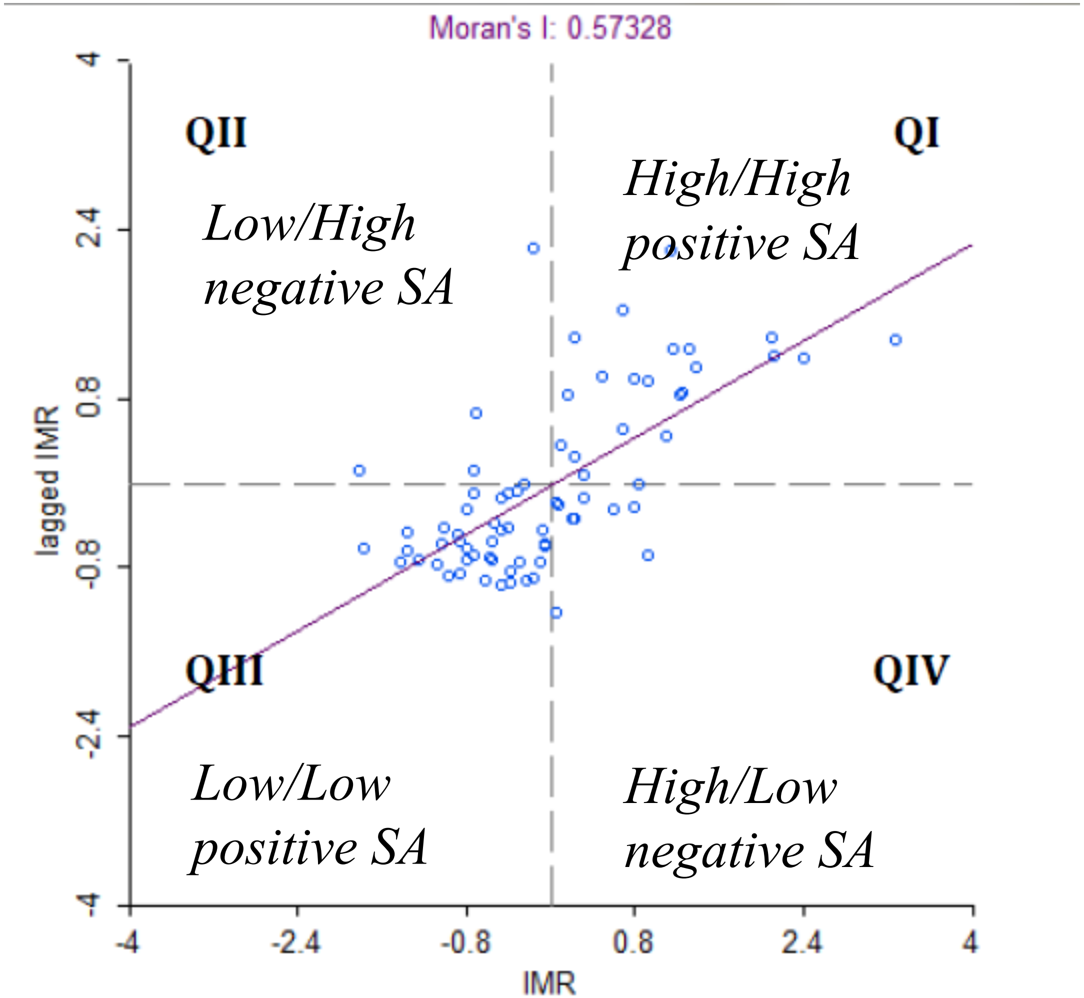
We can draw a scatter diagram between these two variables (in standardized form): **X** and **lag-X** (or W_X)



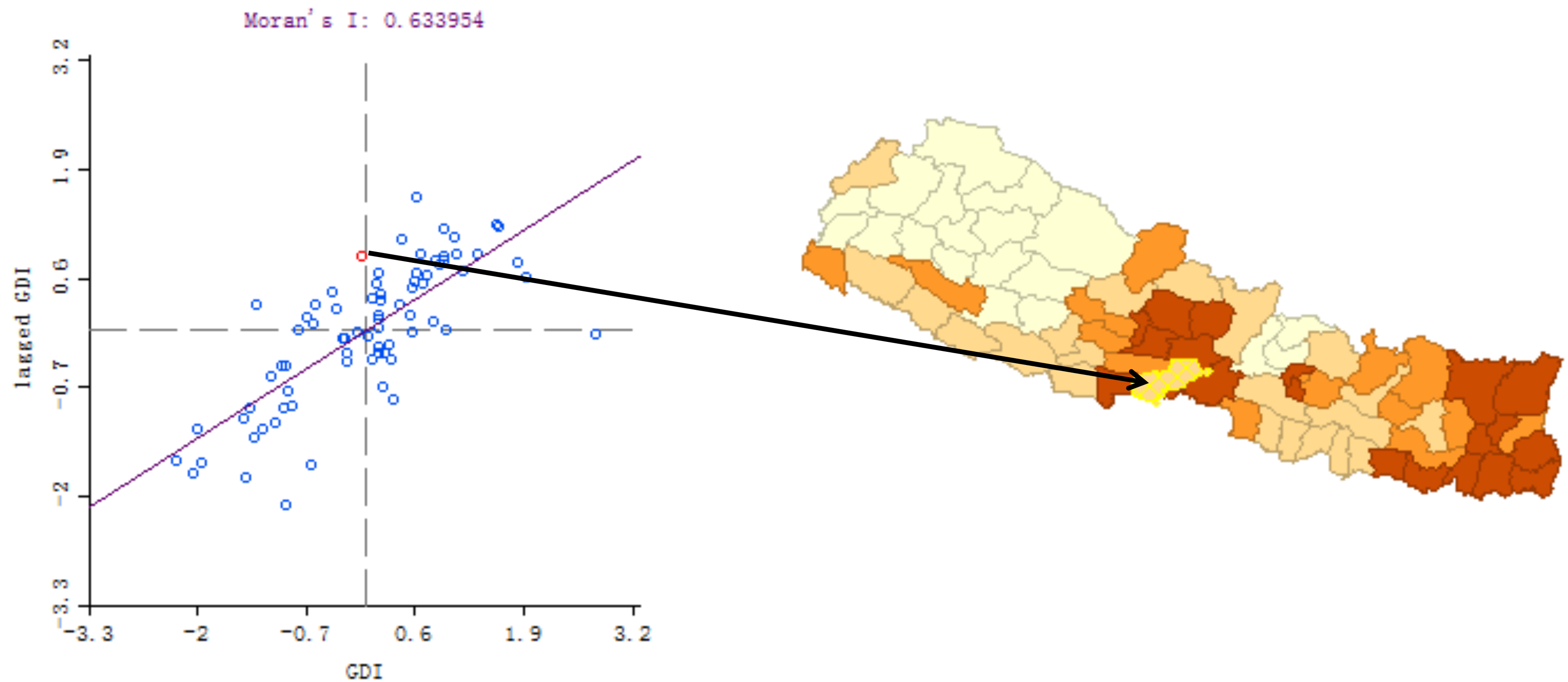


The <u>slope</u> of this *regression line* is Moran's I

# Moran Scatter Plots



# Moran Scatterplot: Example



# Moran’s I for rate-based data

- Moran’s I is often calculated for rates, such as crime rates (e.g. number of crimes per 1,000 population) or infant mortality rates (e.g. number of deaths per 1,000 births)
- An adjustment should be made, especially if the denominator in the rate (population or number of births) varies greatly (as it usually does)
- Adjustment is know as the *EB adjust*ment:
  - see Assuncao-Reis *Empirical Bayes Standardization* Statistics in Medicine, 1999
- *GeoDA* software includes an option for this adjustment

# Statistical Significance Tests for Moran's I

- Based on the normal frequency distribution with

$$Z = \frac{I - E(I)}{S_{error(I)}}$$

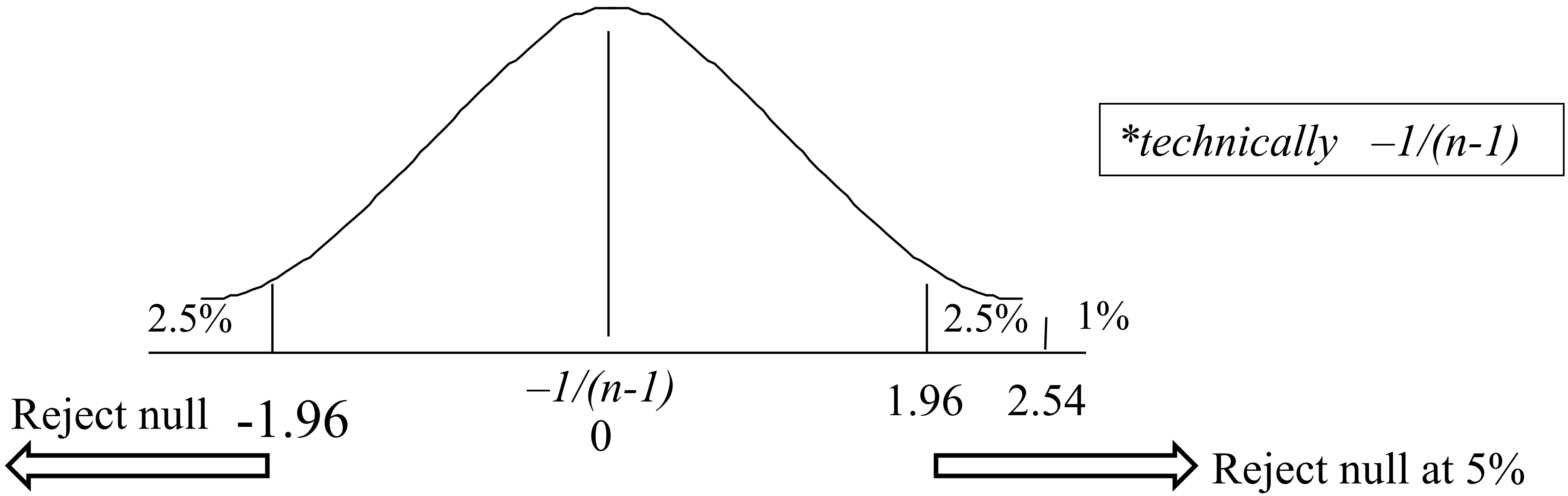
Where: I is the calculated value for Moran's I from the sample

E(I) is the expected value if random

S is the standard error

- Statistical significance test
  - Monte Carlo test, as we did for spatial pattern analysis
  - Permutation test
    - Non-parametric
    - Data-driven, no assumption of the data
    - Implemented in GeoDa

# Test Statistic for Normal Frequency Distribution




Reject null at 1%

*Null Hypothesis:* no spatial autocorrelation

*Moran's I = 0

*Alternative Hypothesis:* spatial autocorrelation exists

*Moran's I > 0

Reject *Null Hypothesis* if Z test statistic > 1.96 (or < -1.96)

---less than a 5% chance that, in the population, there is no spatial autocorrelation

---95% confident that spatial auto correlation exits

*Null Hypothesis:* no spatial autocorrelation

*Moran's I = 0

*Alternative Hypothesis:* spatial autocorrelation exists

*Moran's I > 0

Reject *Null Hypothesis* if Z test statistic > 1.96 (or < -1.96)

---less than a 5% chance that, in the population, there is no spatial autocorrelation

---95% confident that spatial auto correlation exits

# Local Measures of Spatial Autocorrelation

# Local Indicators of Spatial Association (LISA)

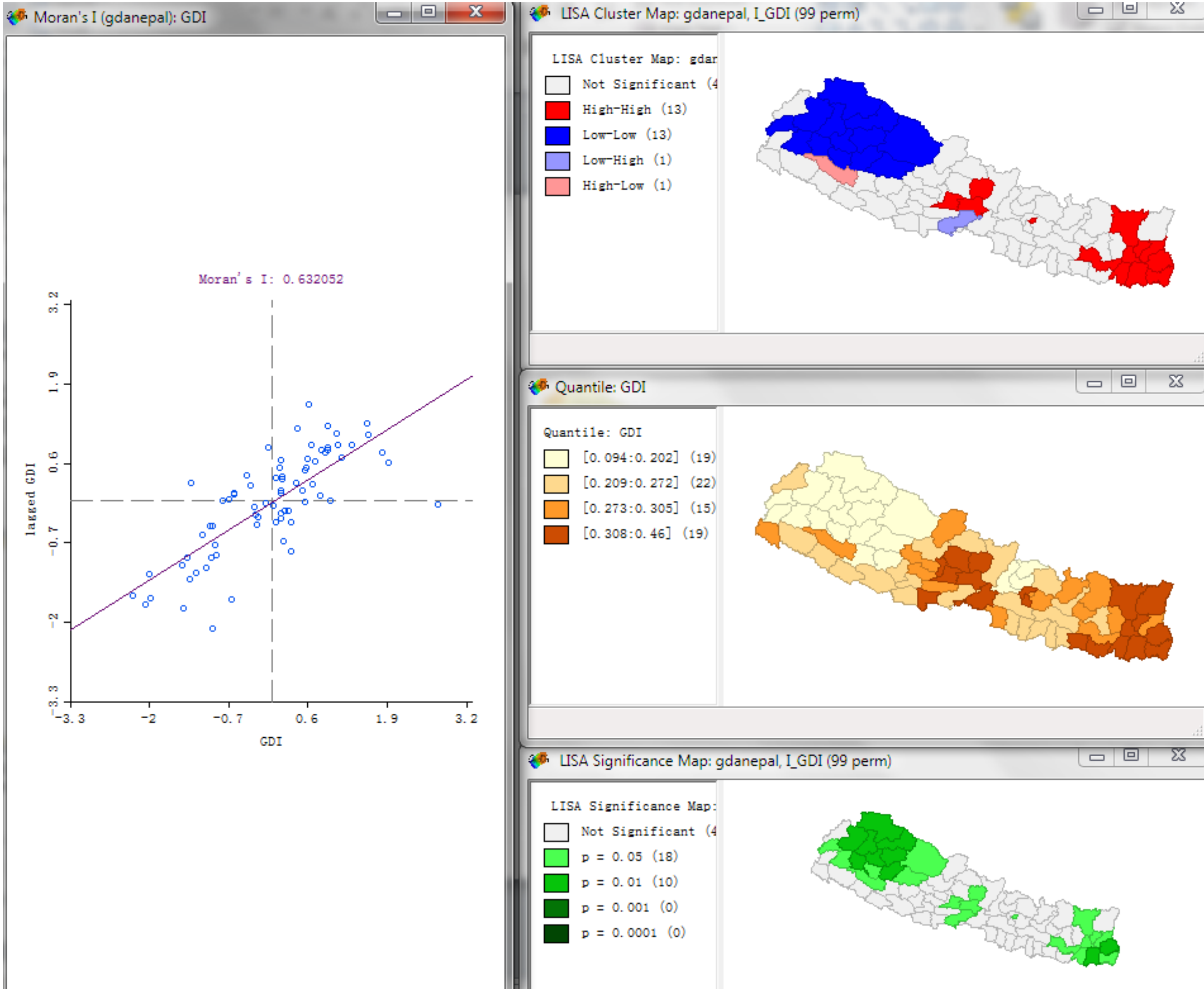
- Local versions of *Moran's I*
- Moran's I is most commonly used, and the local version is often called Anselin's LISA, or just LISA

**See:**
Luc Anselin 1995 *Local Indicators of Spatial Association-LISA* Geographical Analysis 27: 93-115

# Local Indicators of Spatial Association (LISA)

- The statistic is calculated for **each** areal unit in the data
- For each polygon, the index is calculated based on neighboring polygons with which it shares a border
- A measure is available for each polygon, these can be mapped to indicate how spatial autocorrelation varies over the study region
- Each index has an associated test statistic, we can also map which of the polygons has a statistically significant relationship with its neighbors, and show type of relationship

# Example:


Moran's I (gdanepal): GDI
Moran's I: 0.632052
lagged GDI
3.2
1.9
0.6
-0.7
-2
-3.3
-3.3
-2
-0.7
0.6
1.9
3.2
GDI
LISA Cluster Map: gdanepal, I_GDI (99 perm)
LISA Cluster Map: gdar
Not Significant (4
High-High (13)
Low-Low (13)
Low-High (1)
High-Low (1)
Quantile: GDI
Quantile: GDI
[0.094:0.202] (19)
[0.209:0.272] (22)
[0.273:0.305] (15)
[0.308:0.46] (19)
LISA Significance Map: gdanepal, I_GDI (99 perm)
LISA Significance Map:
Not Significant (4
p = 0.05 (18)
p = 0.01 (10)
p = 0.001 (0)
p = 0.0001 (0)

# Calculating Anselin's LISA

- The local Moran statistic for areal unit $i$ is:

$$I_i = z_i \sum_j w_{ij} z_j$$

where $z_i$ is the original variable $x_i$ in "standardized form" $z_i = \frac{x_i - \bar{x}}{SD_x}$

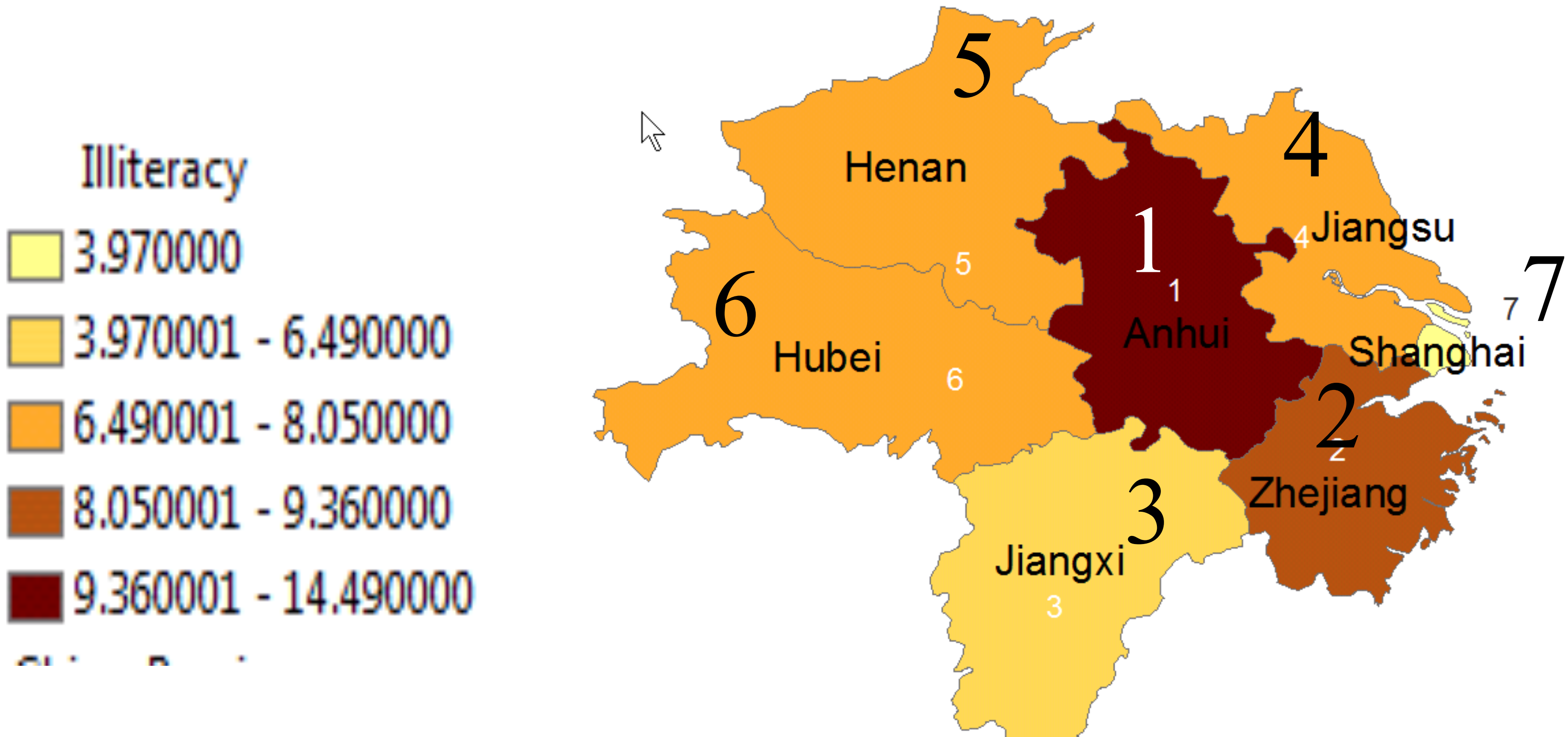
or it can be in "deviation form" $x_i - \bar{x}$

and $w_{ij}$ is the spatial weight

The summation $\sum_j$ is across each <u>row</u> $i$ of the spatial weights matrix.

An example follows

| **Contiguity Matrix** | Code | 1 Anhui | 2 Zhejiang | 3 Jiangxi | 4 Jiangsu | 5 Henan | 6 Hubei | 7 Shanghai | Sum | Neighbors | **Illiteracy** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Anhui | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 | 6 5 4 3 2 | **14.49** |
| Zhejiang | 2 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 | 7 4 3 1 | **9.36** |
| Jiangxi | 3 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 3 | 6 2 1 | **6.49** |
| Jiangsu | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 3 | 7 2 1 | **8.05** |
| Henan | 5 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 6 1 | **7.36** |
| Hubei | 6 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 1 3 5 | **7.69** |
| Shanghai | 7 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 2 4 | **3.97** |




Source: Ron Briggs of UT Dallas

# Contiguity Matrix and Row Standardized Spatial Weights Matrix

**Contiguity Matrix**

| | Code | 1 Anhui | 2 Zhejiang | 3 Jiangxi | 4 Jiangsu | 5 Henan | 6 Hubei | 7 Shanghai | Sum |
|---|---|---|---|---|---|---|---|---|---|
| Anhui | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 |
| Zhejiang | 2 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |
| Jiangxi | 3 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 3 |
| Jiangsu | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 3 |
| Henan | 5 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 |
| Hubei | 6 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 3 |
| Shanghai | 7 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 |

1/3

**Row Standardized Spatial Weights Matrix**

| | Code | Anhui | Zhejiang | Jiangxi | Jiangsu | Henan | Hubei | Shanghai | Sum |
|---|---|---|---|---|---|---|---|---|---|
| Anhui | 1 | 0.00 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.00 | 1 |
| Zhejiang | 2 | 0.25 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.25 | 1 |
| Jiangxi | 3 | 0.33 | 0.33 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | 1 |
| Jiangsu | 4 | 0.33 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 1 |
| Henan | 5 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 1 |
| Hubei | 6 | 0.33 | 0.00 | 0.33 | 0.00 | 0.33 | 0.00 | 0.00 | 1 |
| Shanghai | 7 | 0.00 | 0.50 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 1 |

Source: Ron Briggs of UT Dallas

# Calculating standardized (z) scores

$$z_i = \frac{x_i - \bar{x}}{SD_x}$$

**Deviations from Mean and z scores.**

| | **X** | **X-Xmean** | **X-Mean2** | **z** |
|---|---|---|---|---|
| Anhui | 14.49 | 6.29 | 39.55 | 2.101 |
| Zhejiang | 9.36 | 1.16 | 1.34 | 0.387 |
| Jiangxi | 6.49 | (1.71) | 2.93 | (0.572) |
| Jiangsu | 8.05 | (0.15) | 0.02 | (0.051) |
| Henan | 7.36 | (0.84) | 0.71 | (0.281) |
| Hubei | 7.69 | (0.51) | 0.26 | (0.171) |
| Shanghai | 3.97 | (4.23) | 17.90 | (1.414) |

**Mean and Standard Deviation**

| | | | | | |
|---|---|---|---|---|---|
| Sum | 57.41 | | | 0.00 | 62.71 |
| Mean | 57.41 | / | 7 | = | 8.20 |
| Variance | 62.71 | / | 7 | = | 8.96 |
| SD | √ 8.96 | | | = | 2.99 |



Source: Ron Briggs of UT Dallas

# Calculating LISA

**Row Standardized Spatial Weights Matrix**

| | Code | Anhui | Zhejiang | Jiangxi | Jiangsu | Henan | Hubei | Shanghai |
|---|---|---|---|---|---|---|---|---|
| Anhui | 1 | 0.00 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.00 |
| Zhejiang | 2 | 0.25 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.25 |
| Jiangxi | 3 | 0.33 | 0.33 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 |
| Jiangsu | 4 | 0.33 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 |
| Henan | 5 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 |
| Hubei | 6 | 0.33 | 0.00 | 0.33 | 0.00 | 0.33 | 0.00 | 0.00 |
| Shanghai | 7 | 0.00 | 0.50 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 |

$w_{ij}$

$$I_i = z_i \sum_j w_{ij} z_j$$

**Z-Scores for row Province and its potential neighbors**

| | Zi | Anhui | Zhejiang | Jiangxi | Jiangsu | Henan | Hubei | Shanghai |
|---|---|---|---|---|---|---|---|---|
| Anhui | 2.101 | ***2.101*** | 0.387 | (0.572) | (0.051) | (0.281) | (0.171) | (1.414) |
| Zhejiang | 0.387 | 2.101 | ***0.387*** | (0.572) | (0.051) | (0.281) | (0.171) | (1.414) |
| Jiangxi | (0.572) | 2.101 | 0.387 | ***(0.572)*** | (0.051) | (0.281) | (0.171) | (1.414) |
| Jiangsu | (0.051) | 2.101 | 0.387 | (0.572) | ***(0.051)*** | (0.281) | (0.171) | (1.414) |
| Henan | (0.281) | 2.101 | 0.387 | (0.572) | (0.051) | *(0.281)* | (0.171) | (1.414) |
| Hubei | (0.171) | 2.101 | 0.387 | (0.572) | (0.051) | (0.281) | ***(0.171)*** | (1.414) |
| Shanghai | (1.414) | 2.101 | 0.387 | (0.572) | (0.051) | (0.281) | (0.171) | ***(1.414)*** |

$z_j$

$w_{ij} z_j$

**Spatial Weight Matrix multiplied by Z-Score Matrix (cell by cell multiplication)**

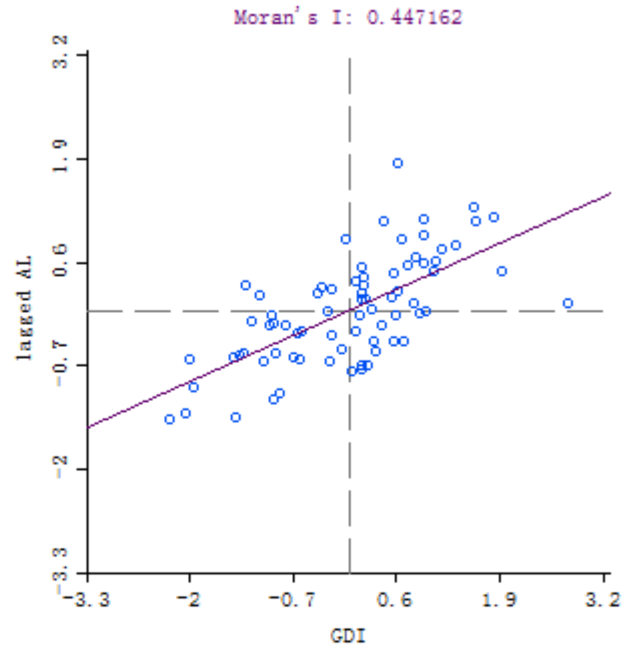
| | **Zi** | **Anhui** | **Zhejiang** | **Jiangxi** | **Jiangsu** | **Henan** | **Hubei** | **Shanghai** | **SumWijZj** | **LISA** | **Lisa from GeoDA** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | **0.000** | | |
| Anhui | **2.101** | - | 0.077 | (0.114) | (0.010) | (0.056) | (0.034) | - | **(0.137)** | **-0.289** | -0.248 |
| Zhejiang | **0.387** | 0.525 | - | (0.143) | (0.013) | - | - | (0.353) | **0.016** | **0.006** | 0.005 |
| Jiangxi | **(0.572)** | 0.700 | 0.129 | - | - | - | (0.057) | - | **0.772** | **-0.442** | -0.379 |
| Jiangsu | **(0.051)** | 0.700 | 0.129 | - | - | - | - | (0.471) | **0.358** | **-0.018** | -0.016 |
| Henan | **(0.281)** | 1.050 | - | - | - | - | (0.085) | - | **0.965** | **-0.271** | -0.233 |
| Hubei | **(0.171)** | 0.700 | - | (0.191) | - | (0.094) | - | - | **0.416** | **-0.071** | -0.061 |
| Shanghai | **(1.414)** | - | 0.194 | - | (0.025) | - | - | - | **0.168** | **-0.238** | -0.204 |

Source: Ron Briggs of UT Dallas

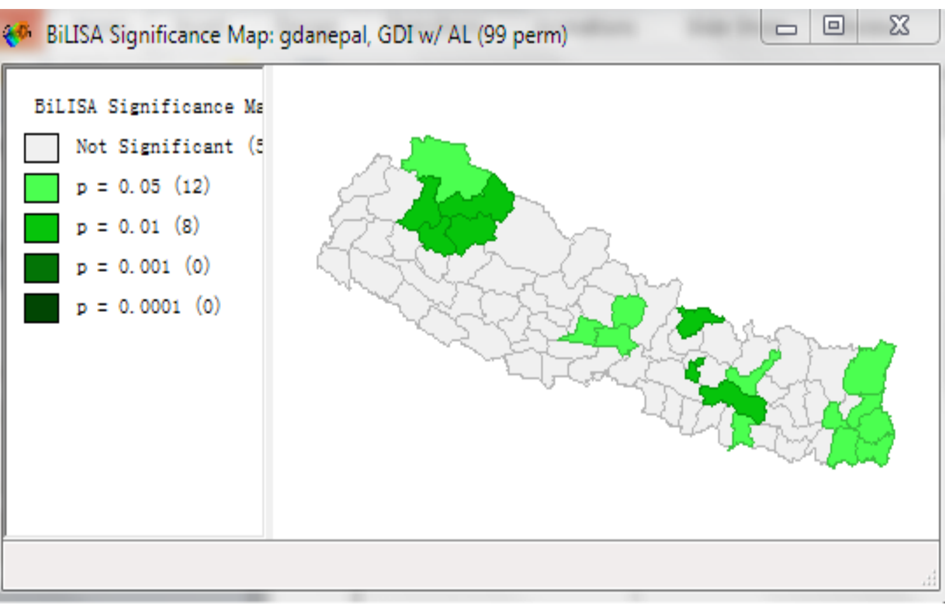
# Bivariate LISA

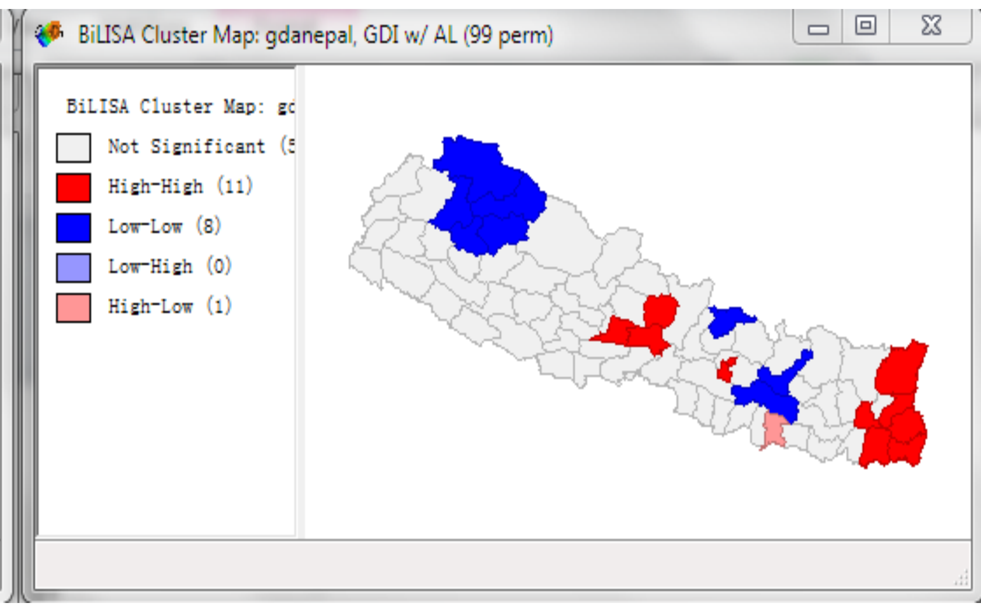
- Moran's I is the correlation between X and Lag-X--the <u>same</u> variable but in <u>nearby</u> areas
  - Univariate Moran's I
- Bivariate Moran's I is a correlation between X and a <u>different</u> variable in <u>nearby</u> areas.

Moran <u>Scatter Plot</u> for GDI vs AL



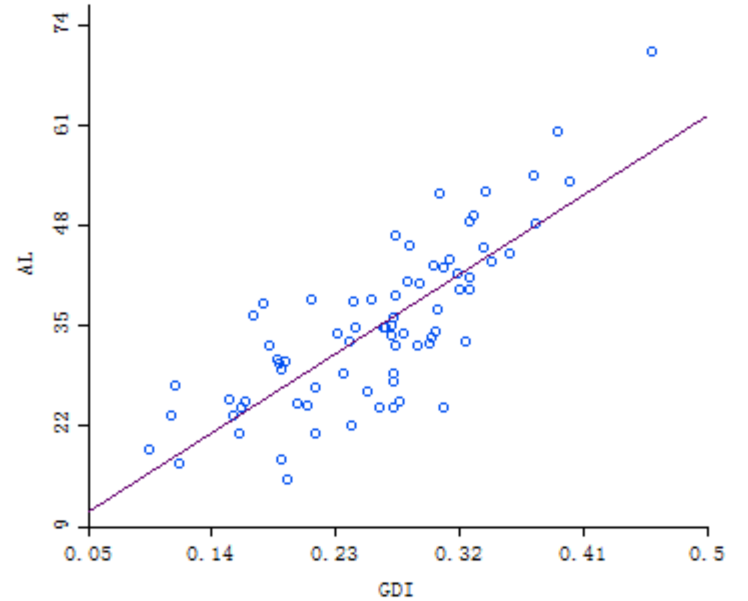

Moran <u>Significance Map</u> for GDI vs. AL

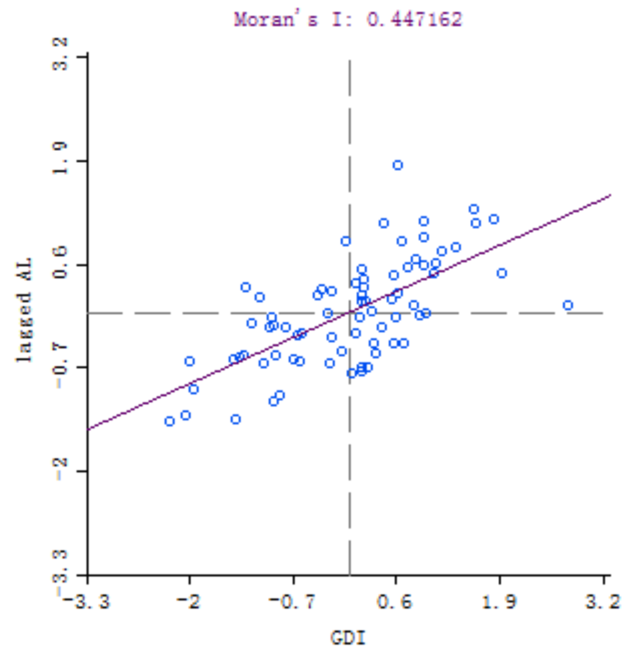


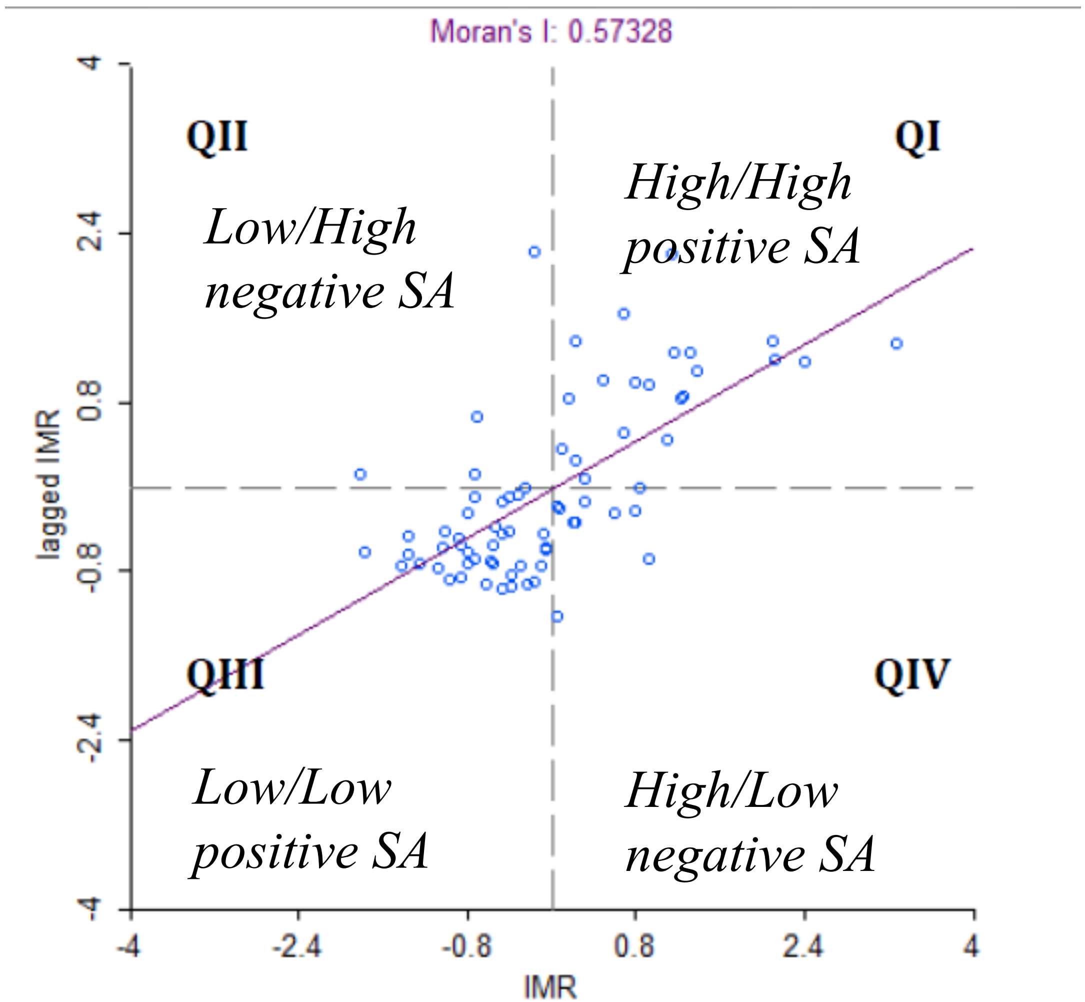
# Bivariate LISA
# and the Correlation Coefficie

- Correlation Coefficient is the relationship between two <u>different</u> variables in the <u>same</u> area
- Bivariate LISA is a correlation between two <u>different</u> variables in an area and in <u>nearby</u> areas.



# Bivariate Moran Scatter Plot



# Summary

- Spatial autocorrelation of areal data
- Spatial weight matrix
- Measures of spatial autocorrelation
- Global Measure
  - Moran's I/General G and G*
- Local
  - LISA: Moran's I/General G and G*
  - Bivariate LISA
  - Significance test

- End of this topic