ERSA Summer School 2006

Spatial Analysis with GeoDa

2. Spatial Weights

Luc Anselin
University of Illinois, Urbana-Champaign

http://sal.uiuc.edu/

Outline

- Concepts
- Contiguity Weights
- Distance-Based Weights
- Characteristics of Weights
- Spatially Lagged Variables

Concepts

Spatial Autocorrelation

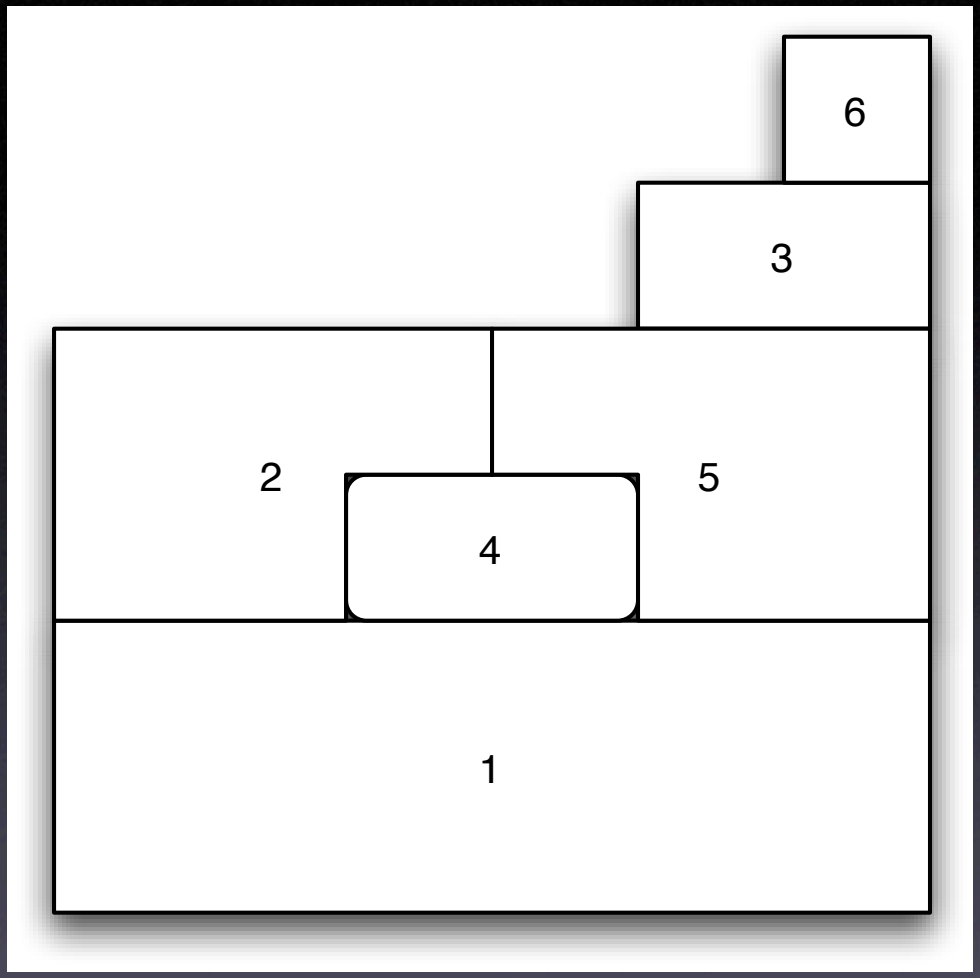
- Match
  - value similarity
  - locational similarity
- How to Specify Locational Similarity

Why Spatial Weights

- Identification Problem
  - spatial covariance $Cov[y_i, y_j]$
  - total number of interactions is $N(N-1)/2$
  - only N observations in a cross-section
- Incidental Parameter Problem
  - number of parameters increases with sample size, $O(N^2)$

Solution

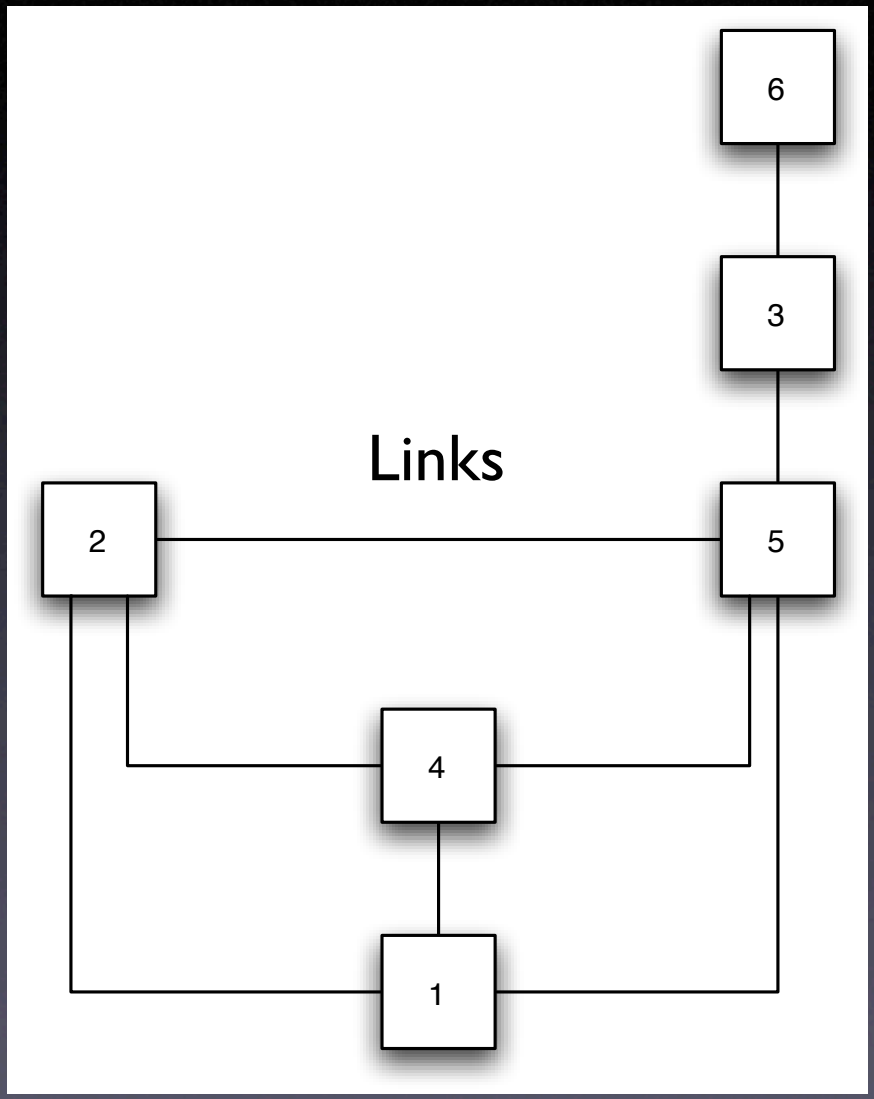
- Impose Structure on the Problem
  - set some interactions to zero
  - only "neighbors" interact directly
  - constrain the number of neighbors
- Assume a Single Parameter
  - spatial autocorrelation coefficient



6 Polygon Layout
Neighbor as Having Common Boundary

Neighbor Structure as a Graph

Nodes



Spatial Weights Matrix

- N by N positive matrix W
  - elements are w_{ij}
  - w_{ij} nonzero for i,j neighbors, w_{ij} = 0 not
  - w_{ii} = 0 no self-neighbors

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

Binary Contiguity

- Weights are 0 or 1
  - for i, j neighbors, $w_{ij} = 1$
- Example, using common boundary

$$\mathbf{w} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Types of Spatial Weights

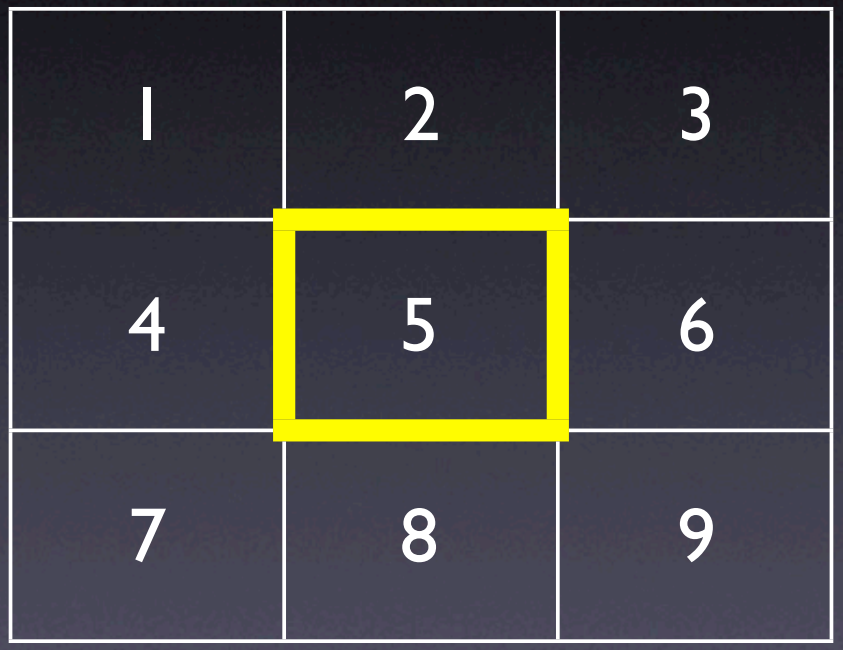
- Geographic Weights
  - contiguity
  - distance
  - general weights
  - graph-based weights
- Socio-Economic Weights

Contiguity Weights

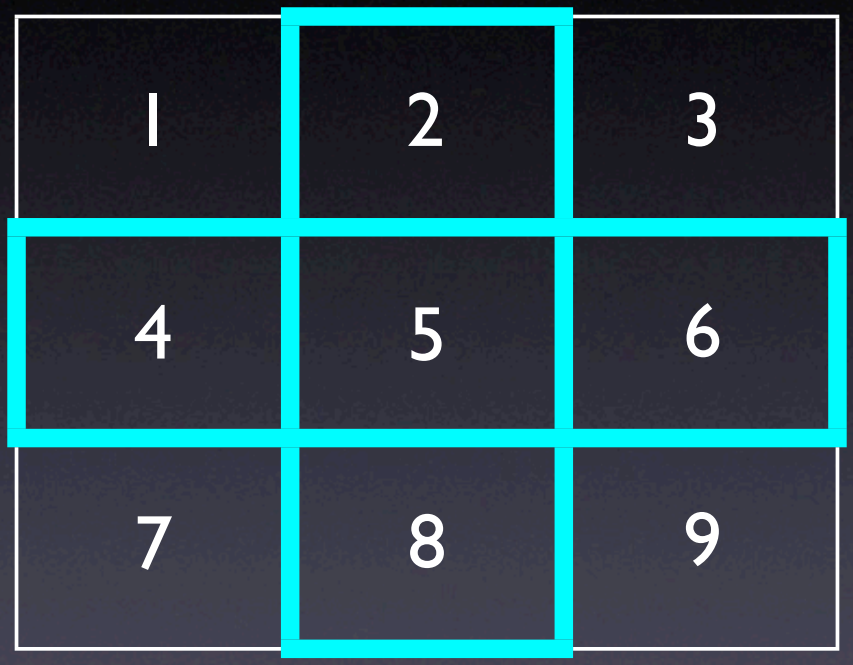
Contiguity

- Common Boundary
  - boundary of non-zero length
- What is a Non-zero Boundary
- Three Views
  - rook
  - bishop
  - queen

Regular Grid Contiguity



Rook Contiguity



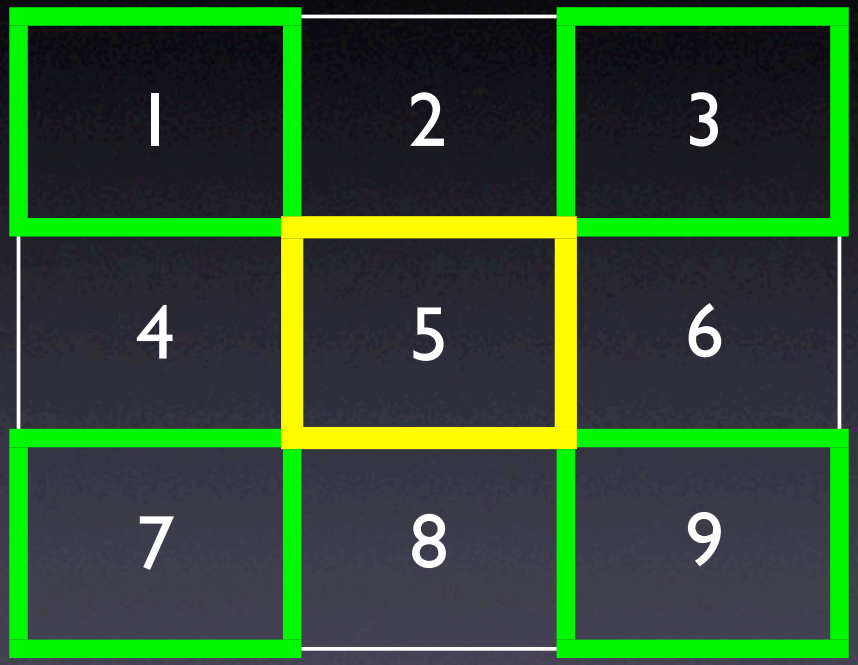
Neighbors for 5: 2, 4, 6, 8

Common Border

Rook Weights

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Bishop Contiguity



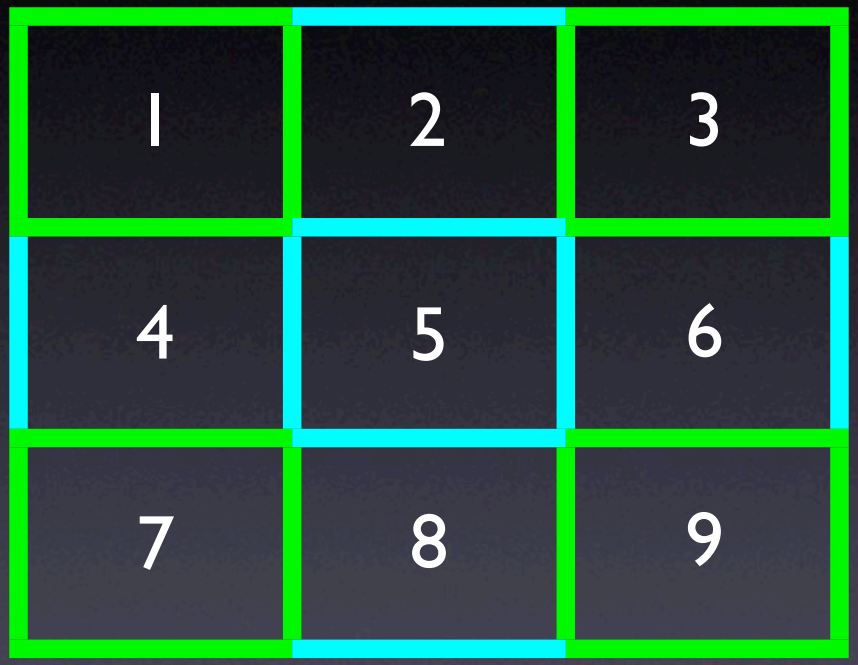
Neighbors for 5: 1, 3, 7, 9

Common Vertex

Bishop Weights

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Queen Contiguity



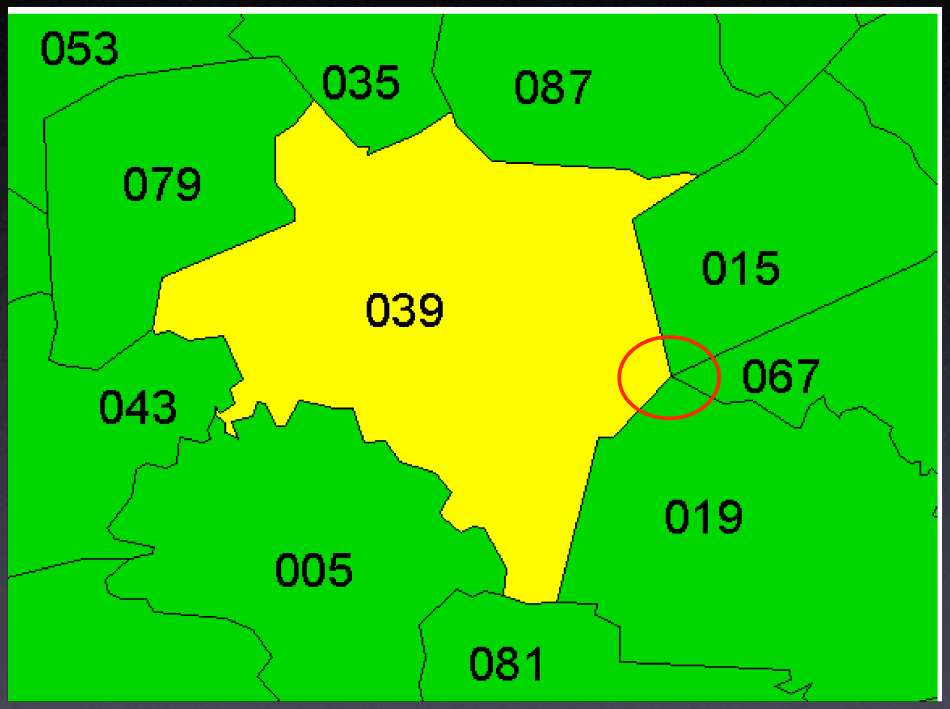
8 Neighbors for 5

Both Border and Vertex

Queen Weights

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

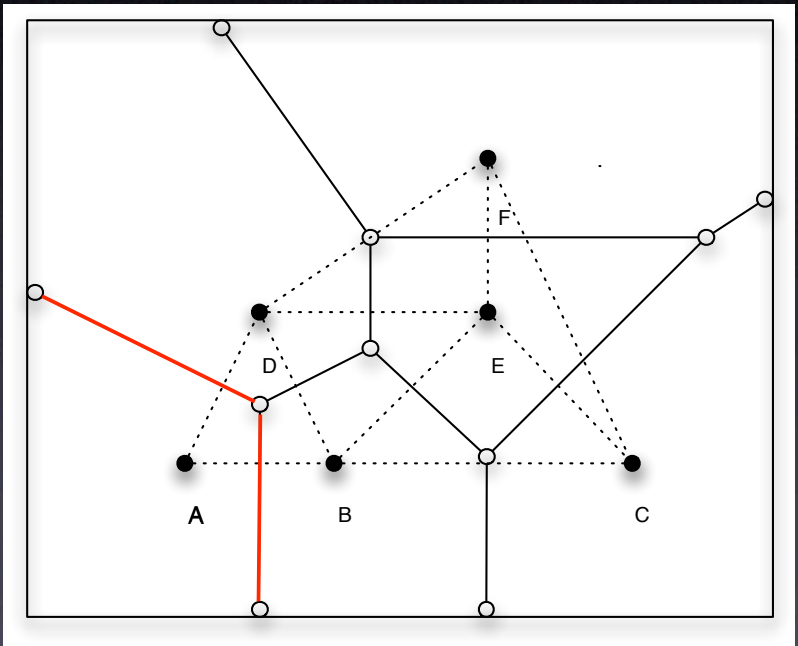
Irregular Unit Contiguity



Rook Contiguity: common border only

Queen Contiguity: border and vertices - 039 and 067

Contiguity for Points



Thiessen Polygons

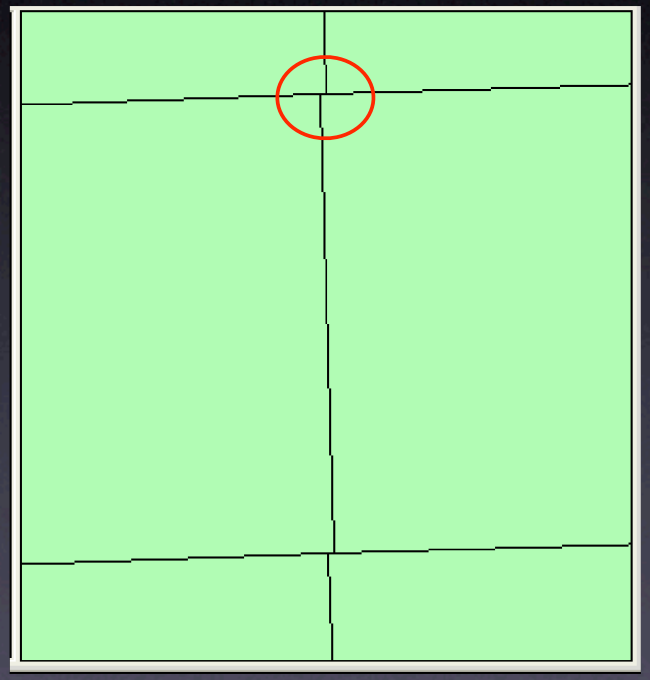
$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Contiguity Weights

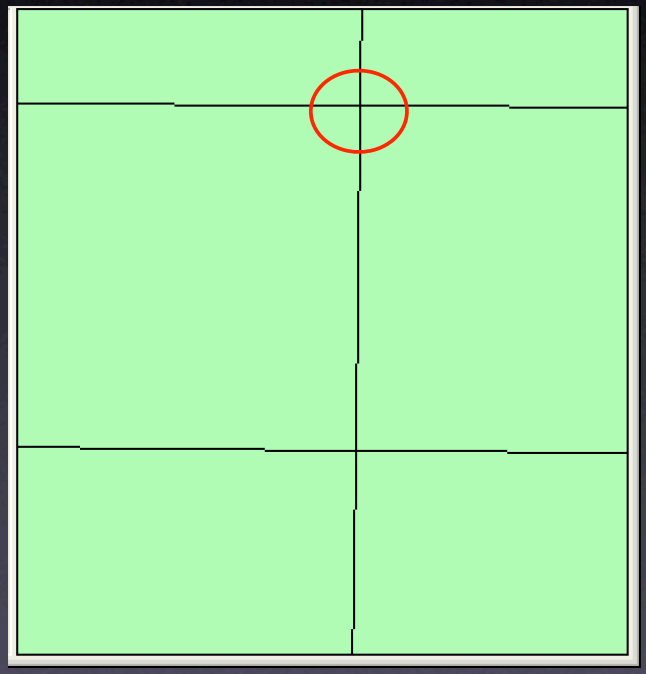
Contiguity Gotchas

- Contiguity Derived from GIS
  - checking boundary file for common arcs and/or points
- Regular Grid Polygons
  - lack of precision can cause misalignment
  - artificial neighbors
- Street Centerline
  - different conventions can cause islands

Misaligned Grid Cells

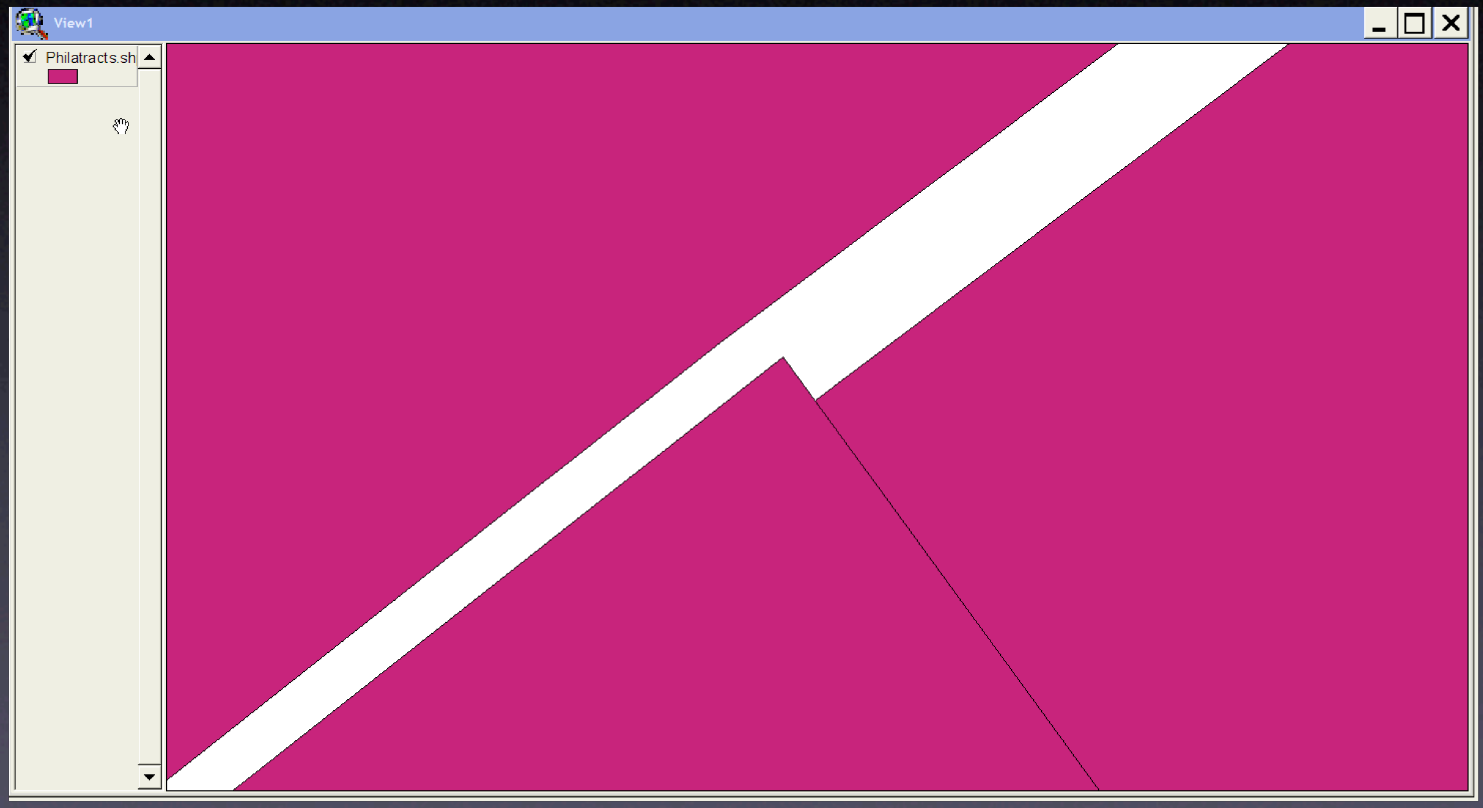


misaligned



properly aligned

Street Centerline Islands



Distance-Based Weights

Distance Metrics

- For Projected Points
  - planar coordinates
- Euclidean Distance
  - $d_{ij} = [(x_i - y_i)^2 + (y_i - y_j)^2]^{1/2}$
  - as the crow flies

Distance Metrics (2)

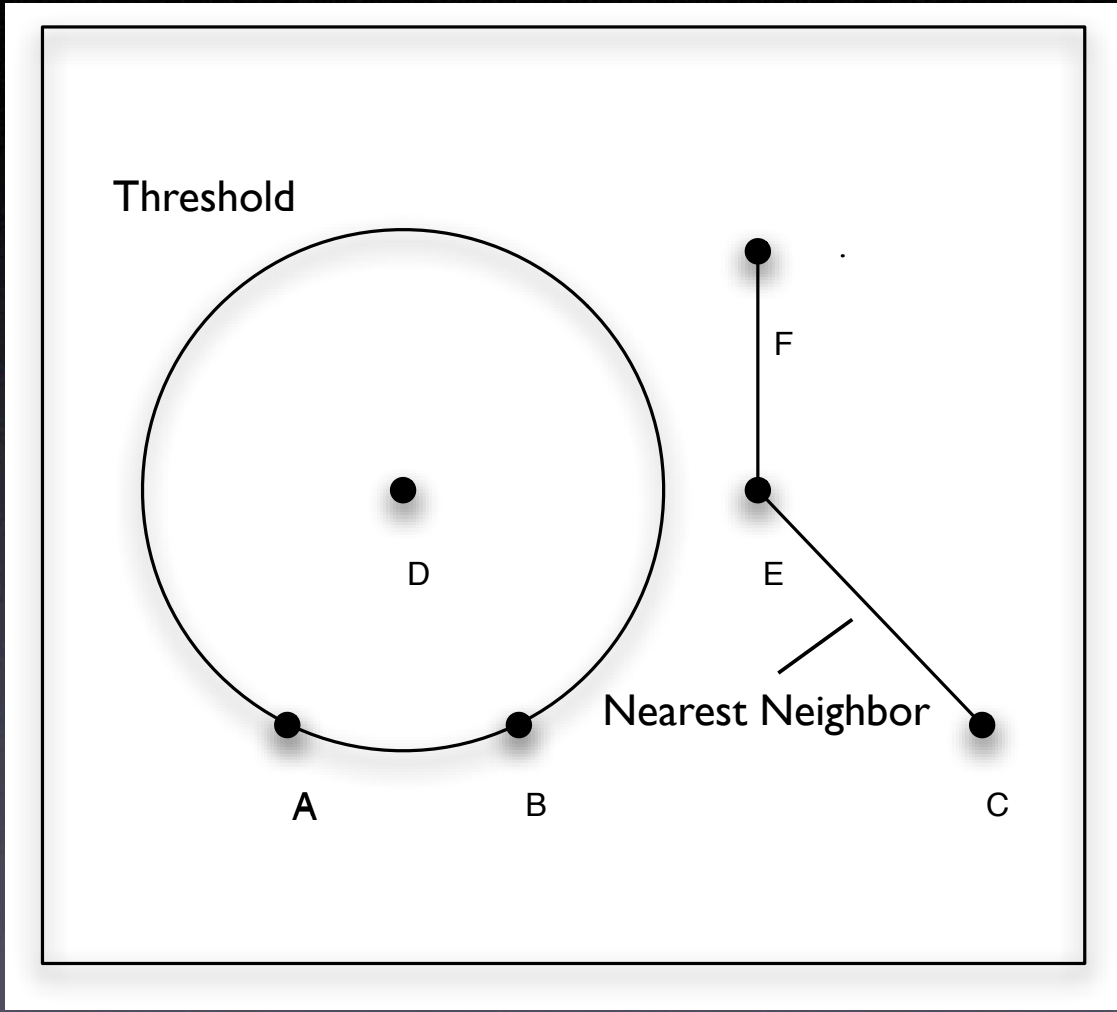
- Manhattan Block Distance
  - $d_{ij} = | x_i - x_j | + | y_i - y_j |$
  - right angles only
- Minkowski Metric
  - $d_{ij} = (| x_i - x_j |^p + | y_i - y_j |^p)^{(1/p)}$
  - general distance metric

Great Circle Distance

- For Unprojected Points (on a sphere)
  - lat-lon, but note: lat is y and lon is x
  - distance on a sphere (spheroid)
- Arc Distance
  - $d_{ij} = R.\arccos[\sin(y_i).\sin(y_j) + \cos(x_i).\cos(x_j).\cos(x_i-x_j)]$
  - R = radius of earth 6378 km or 3963 mi

Distance-Based Weights

- Distance Cut-Off
  - critical distance δ
  - $w_{ij} = 1$ for $d_{ij} < \delta$, $w_{ij} = 0$ otherwise
  - avoid islands → $\max_i(\min_j d_{ij})$
- K-Nearest Neighbors
  - no islands, always k neighbors
  - problems with ties
  - weights are asymmetric



Distance Matrix

	B	C	D	E	F
A	10.0	30.0	11.2	22.4	28.3
B		20.0	11.2	14.1	22.4
C			26.9	14.1	22.4
D				15.0	18.0
E					10.0

Distance Weights

symmetric

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

critical distance threshold $\delta = 11.2$ (radius around D)
C is an island -- no neighbors

Distance Matrix
min neighbor distance

	B	C	D	E	F
A	10.0	30.0	11.2	22.4	28.3
B		20.0	11.2	14.1	22.4
C			26.9	14.1	22.4
D				15.0	18.0
E					10.0

Distance Weights

symmetric

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

critical distance threshold $\delta = 14.1$
max of nearest neighbor distances

K-Nearest Neighbors

k = 3

	B	C	D	E	F
A	10.0	30.0	11.2	22.4	28.3
B		20.0	11.2	14.1	22.4
C			26.9	14.1	22.4
D				15.0	18.0
E					10.0

Distance Weights

asymmetric

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

w_{ij} = 1 for 3 nearest neighbors

Characteristics of Weights

Connectivity

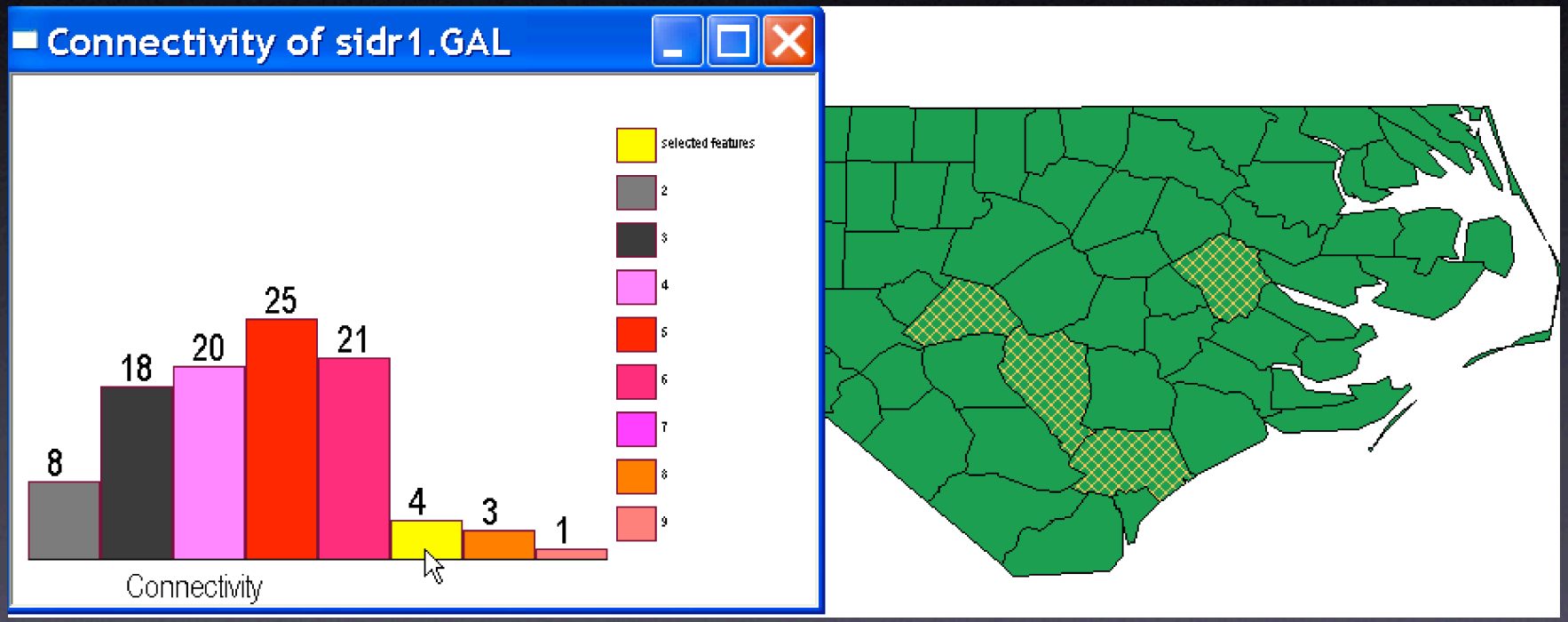
- Measures of Overall Connectedness
  - % non-zero weights: sparseness
  - distribution/average number of links
  - average weight
- Location-Specific Measures
  - most/least connected observations
  - unconnected observations = islands

Spatial Weights Characteristics in R

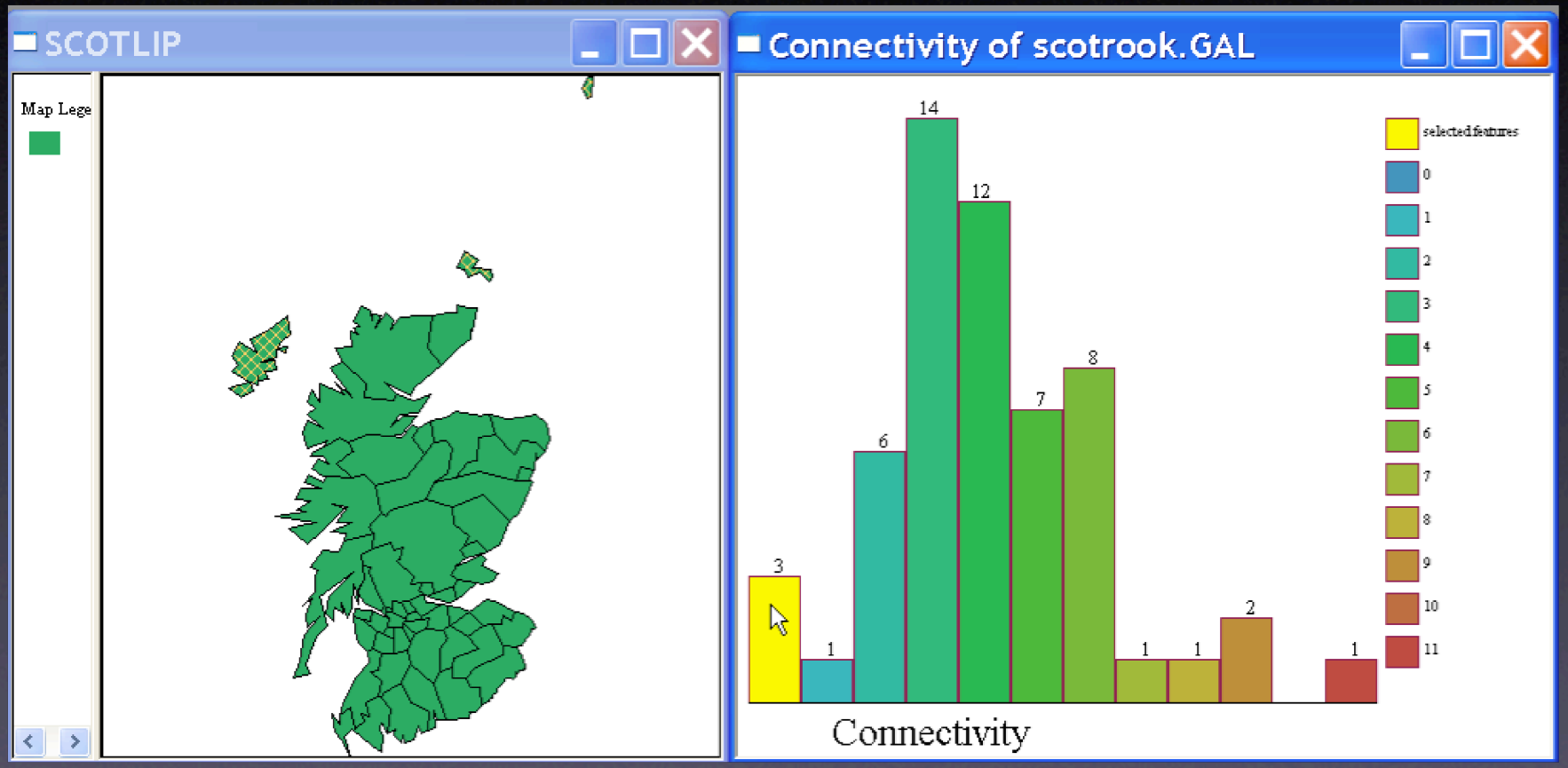
```
> summary.nb(col.gal.nb)
Neighbour list object:
Number of regions: 49
Number of nonzero links: 230
Percentage nonzero weights: 9.579342
Average number of links: 4.693878
Link number distribution:

 2  3  4  5  6  7  8  9 10
 7  7 13  4  9  6  1  1  1
7 least connected regions:
1005 1008 1045 1047 1049 1048 1015 with 2 links
1 most connected region:
1017 with 10 links
```

Connectivity Histogram



Island Connectivity

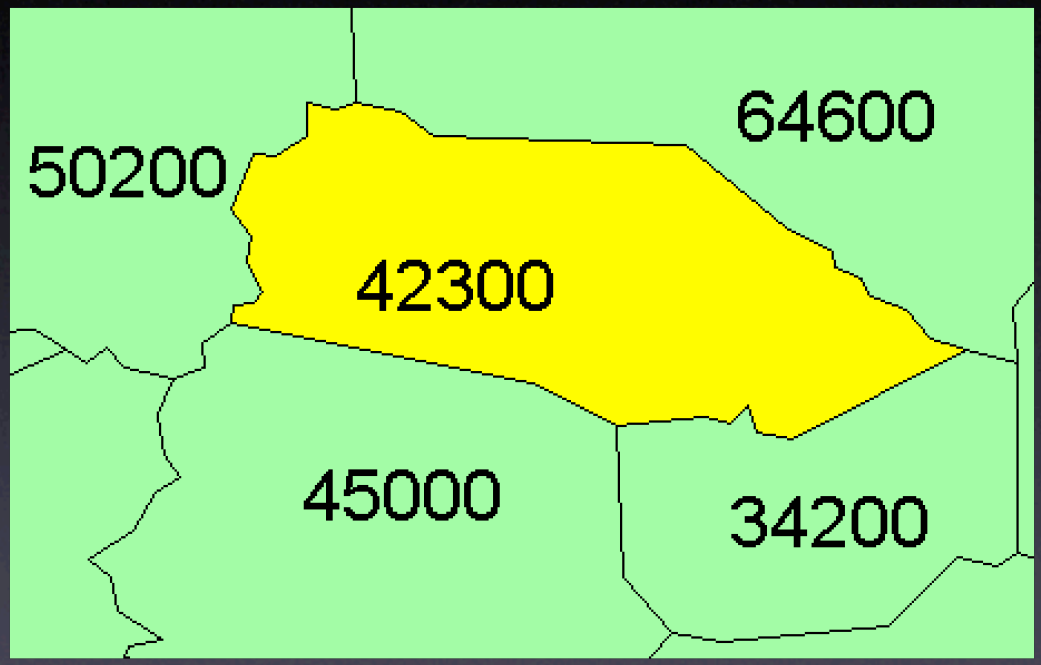


Spatially Lagged Variables

Spatial Lag

- Weighted Average of Neighbors
  - $y_{iL} = w_{i1}.y_1 + w_{i2}.y_2 + \ldots + w_{iN}.y_N$
  - $y_{iL} = \Sigma_j w_{ij}y_j$, $y_L = Wy$
  - most w_{ij} are zero, only non-zero count
- Spatial Lag is a Smoother
  - Wy has smaller variance than y

Spatial Lag Example

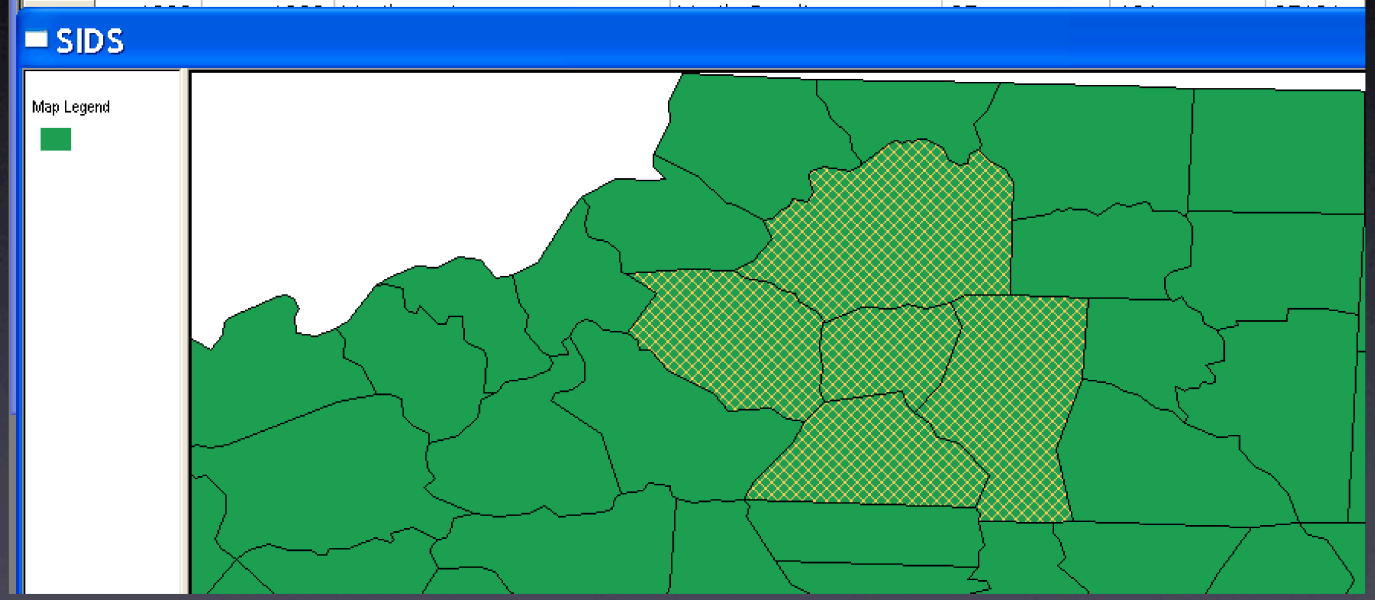


Average Neighbor Land Values

1/4x50200 + 1/4x45000 + 1/4x34200 + 1/4x64600

Table : SIDS

	CNTY_	CNTY_ID	NAME	STATE_NAME	STATE_FIPS	CNTY_FIPS	FIPS
52	1986	1986	Catawba	North Carolina	37	035	37035
39	1947	1947	Iredell	North Carolina	37	097	37097
41	1950	1950	Alexander	North Carolina	37	003	37003
34	1932	1932	Caldwell	North Carolina	37	027	37027
18	1874	1874	Wilkes	North Carolina	37	193	37193



Alexander County and its First Order Neighbors

SIDR74	SIDR79	NWR74	NWR79	W_SIDR74
0.868961	3.050995	137.295794	132.790934	1.527684
0.966417	0.925926	276.395265	241.666667	1.198714
0.000000	1.188354	96.024006	89.126560	1.192336
1.662510	2.118145	85.619285	84.725818	0.715943
1.271456	1.879195	63.572791	59.597315	0.832222

Spatial Lag for SIDR74
Computed as Average of 4 Neighbors