School of Geography and Earth Sciences McMaster University

Applied Spatial Statistics

Area Data
III & IV

This session:

- **Area Data III & IV**
  - Exploration of First Order Effects
  - Exploration of Second Order Effects
    - Spatial correlation
    - Local spatial association
  - Example

Exploration:
Spatial Moving Averages

- Variations on the mean value of a variable

- Weighted average of the values in neighboring areas

Exploration:
Spatial Moving Averages

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} w_{ij} y_j}{\sum_{j=1}^{n} w_{ij}} = \sum_{j=1}^{n} w_{ij}^{st} y_j$$

Exploration:
Spatial Moving Averages

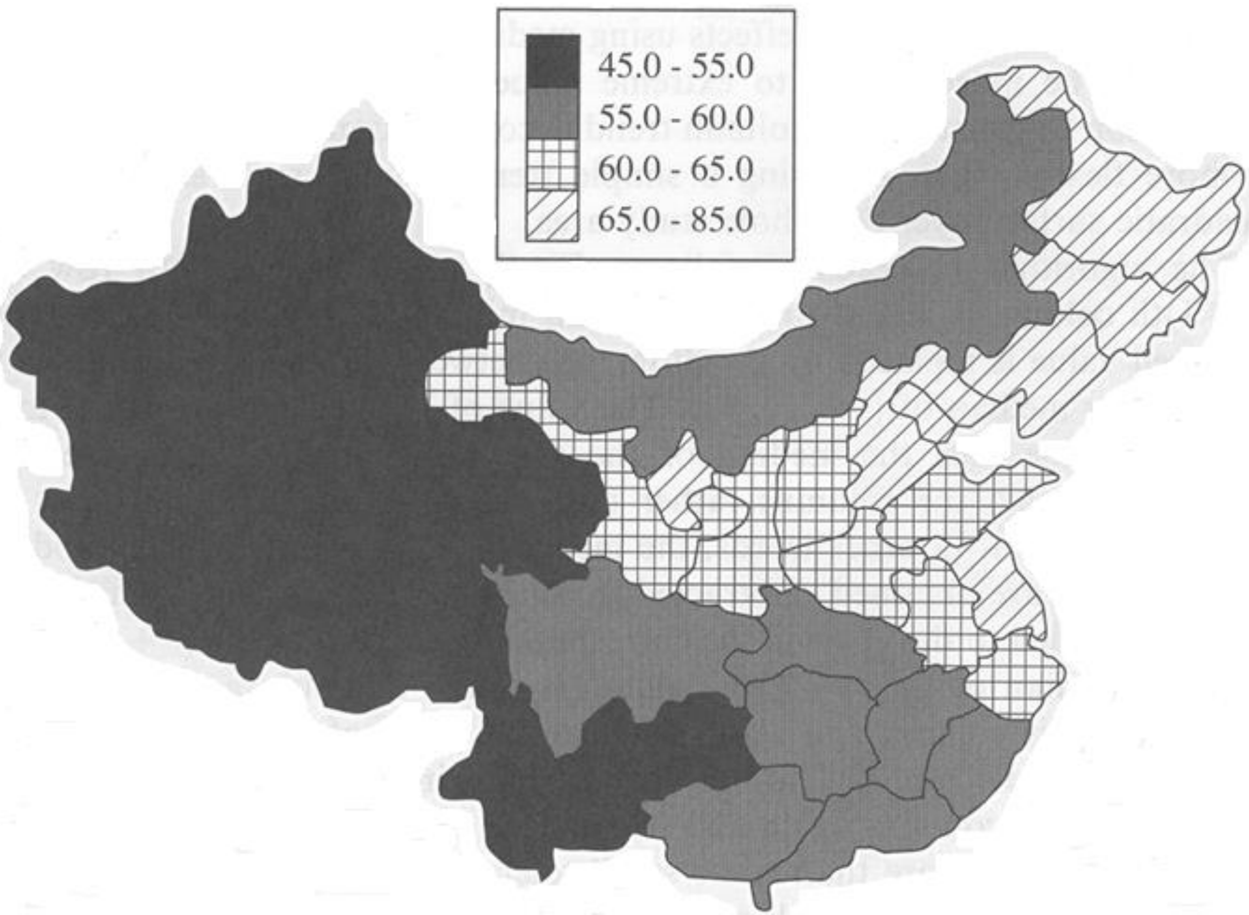


Fig. 7.3 Spatial moving average of gross industrial output in China

Exploration: Kernel Estimation

- Convert areas to points (i.e. use centroids)

Exploration: Kernel Estimation

$$\hat{\mu}(\mathbf{s}) = \sum_{i=1}^{n} w_i(\mathbf{s}) y_i$$

$$w_i(\mathbf{s}) = \frac{k\left(\frac{(\mathbf{s}-\mathbf{s}_i)}{\tau}\right)}{\sum_{i=1}^{n} k\left(\frac{(\mathbf{s}-\mathbf{s}_i)}{\tau}\right)}$$

Exploration: Kernel Estimation

$$\hat{\mu}(\mathbf{s}) = \frac{\sum_{i=1}^{n} k\left(\frac{(\mathbf{s}-\mathbf{s}_i)}{\tau}\right) y_i}{\sum_{i=1}^{n} k\left(\frac{(\mathbf{s}-\mathbf{s}_i)}{\tau}\right)}$$

Second Order Effects

- Spatial dependency/association/correlation

Spatial dependency

- Do attributes in “neighboring” zones show spatial dependency? i.e. Do they co-vary?

Spatial Correlation

- Moran's I

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Spatial Correlation

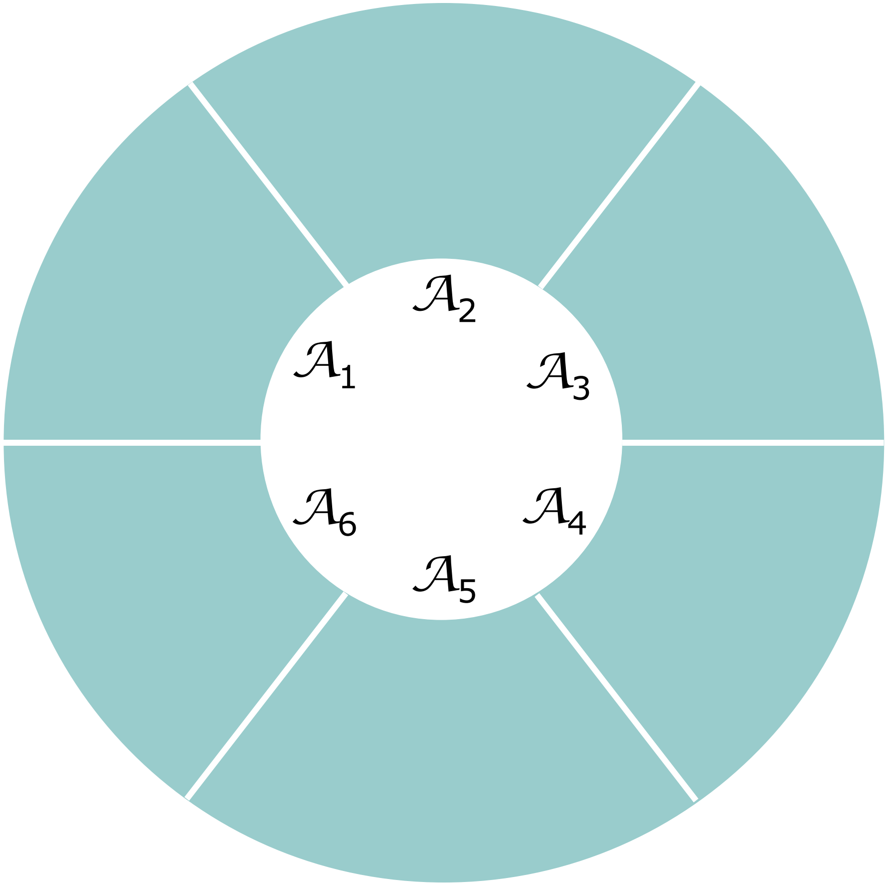
- Moran's I

$$z_i = y_i - \bar{y}$$

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^{\ 2}}$$

*Row standardization?

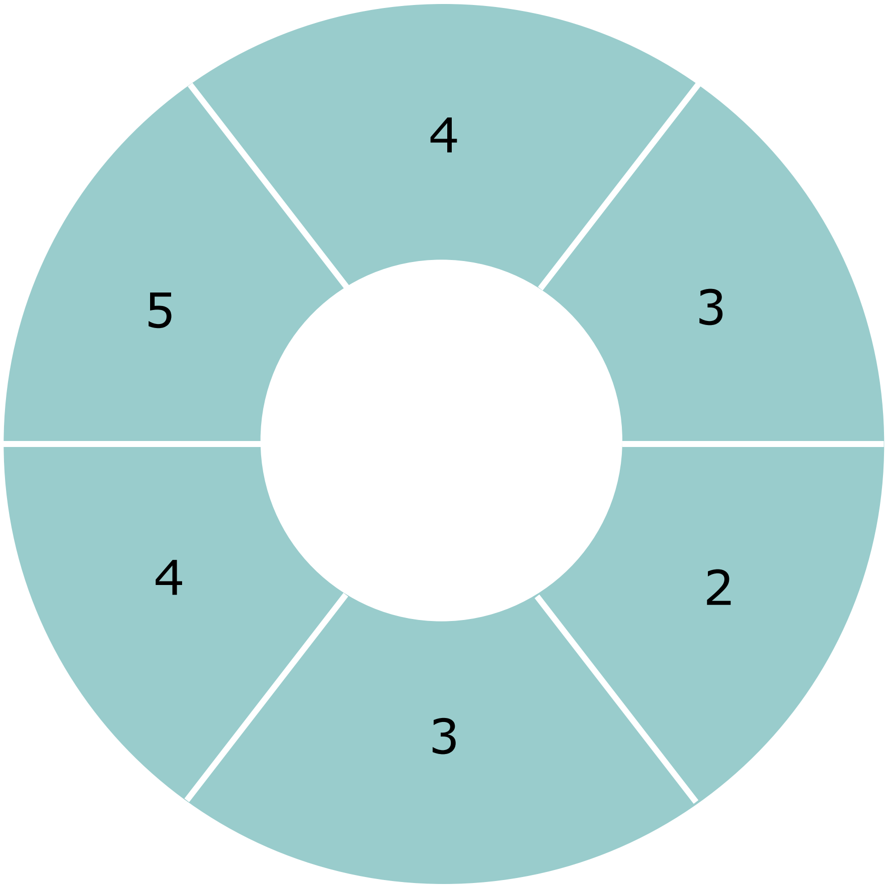
Spatial Correlation



Proximity Matrix

	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$
$\mathcal{A}_1$	0	1	0	0	0	1
$\mathcal{A}_2$	1	0	1	0	0	0
$\mathcal{A}_3$	0	1	0	1	0	0
$\mathcal{A}_4$	0	0	1	0	1	0
$\mathcal{A}_5$	0	0	0	1	0	1
$\mathcal{A}_6$	1	0	0	0	1	0

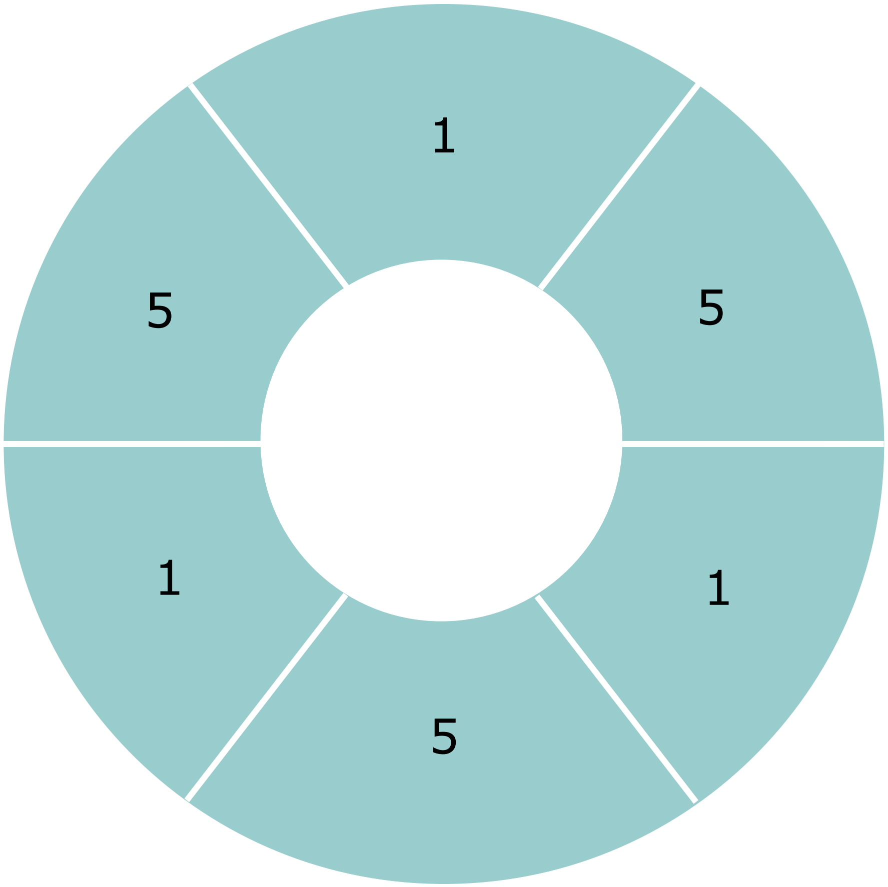
Spatial Correlation



I=0.9091
Z(I)=1.4097

$$\bar{Y} = 3.5$$

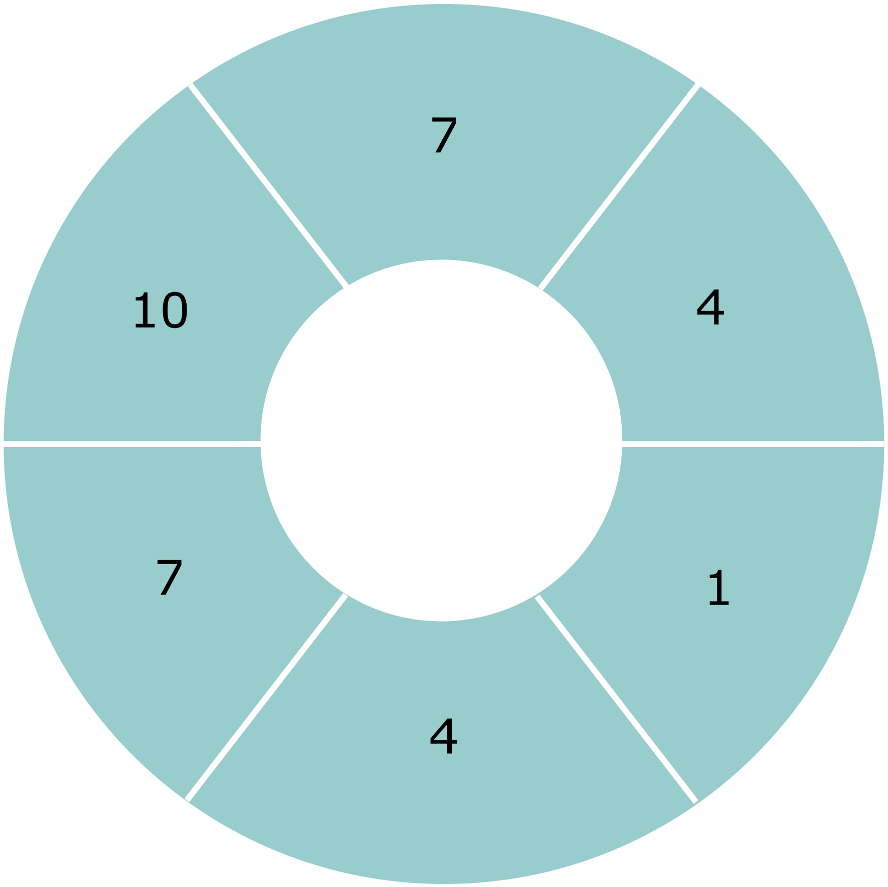
Spatial Correlation



$$\bar{Y} = 3$$

I=-2.000
Z(I)=-3.7417

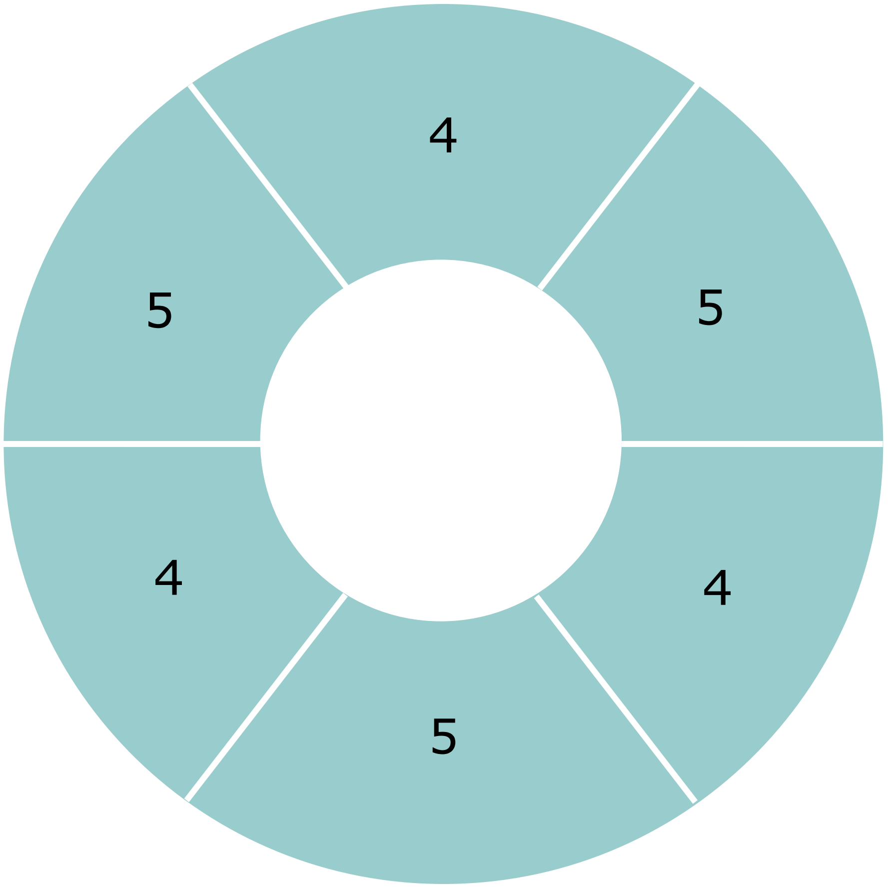
Spatial Correlation



I=0.9091
Z(I)=1.4097

$$\bar{Y} = 5.5$$

Spatial Correlation



I=-2.000
Z(I)=-3.7417

$$\bar{Y} = 4.5$$

Spatial Correlation

- Geary's c (compare to variogram)

$$c = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - y_j)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Moran's I

- Expected value (mean)
  - *Under the assumption of no autocorrelation!*

$$E[I] = -\frac{1}{n-1}$$

* Testing for significance

Moran's I

- Variance

$$V[I] = -\frac{n^2 S_1 - nS_2 + 3(\sum\sum w_{ij})^2}{(\sum\sum w_{ij})^2 (n^2 - 1)}$$

$$S_1 = \sum\sum(w_{ij} + w_{ji})^2$$

$$S_2 = \sum\left(\sum w_{ij} + \sum w_{ji}\right)^2$$

Moran's I

- Testing for significance
  - Asymptotically normally distributed

$$Z[I] = \frac{I - E[I]}{\sqrt{V[I]}}$$

* Testing for significance

Higher Order Neighbors

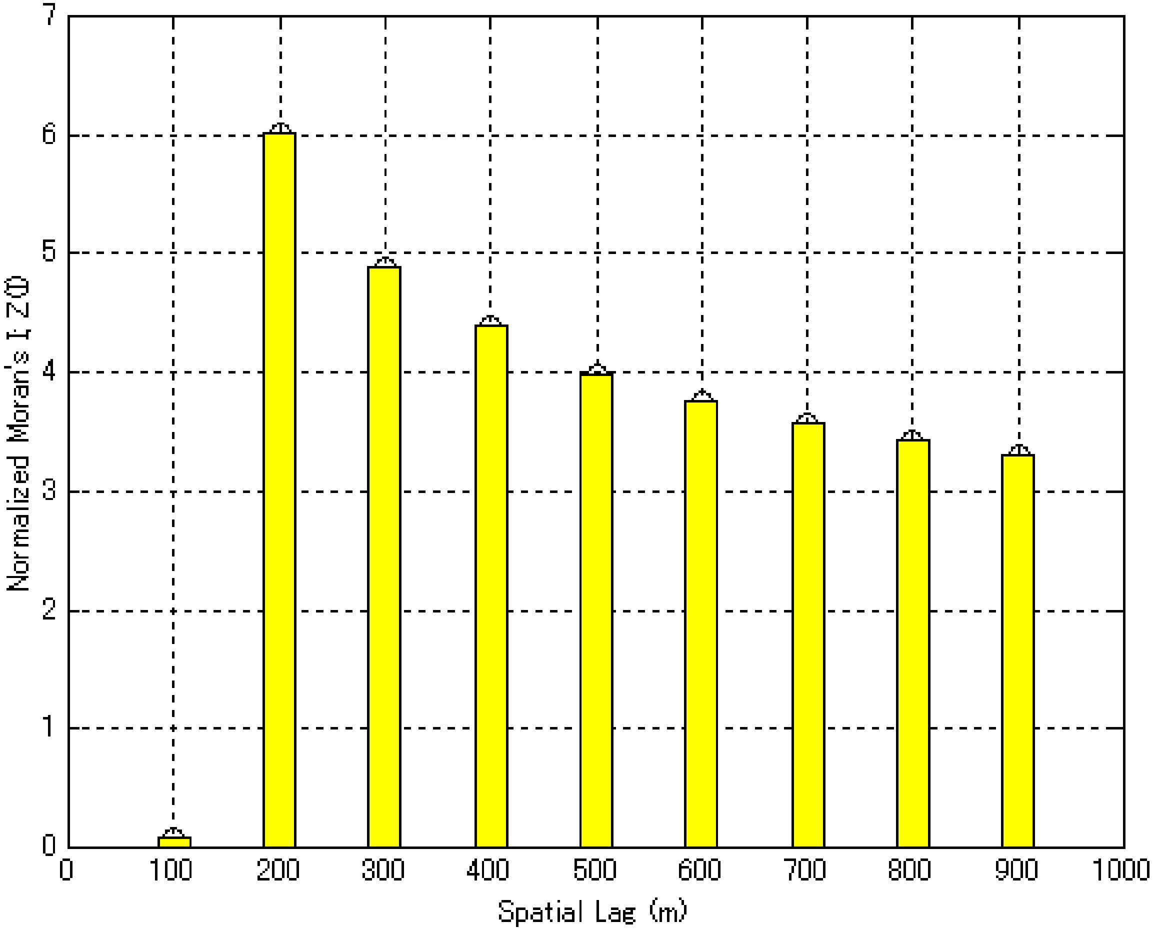
- Moran's I using higher order neighbors

$$I^{(k)} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{(k)}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{(k)}(y_i - \overline{y})(y_j - \overline{y})}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

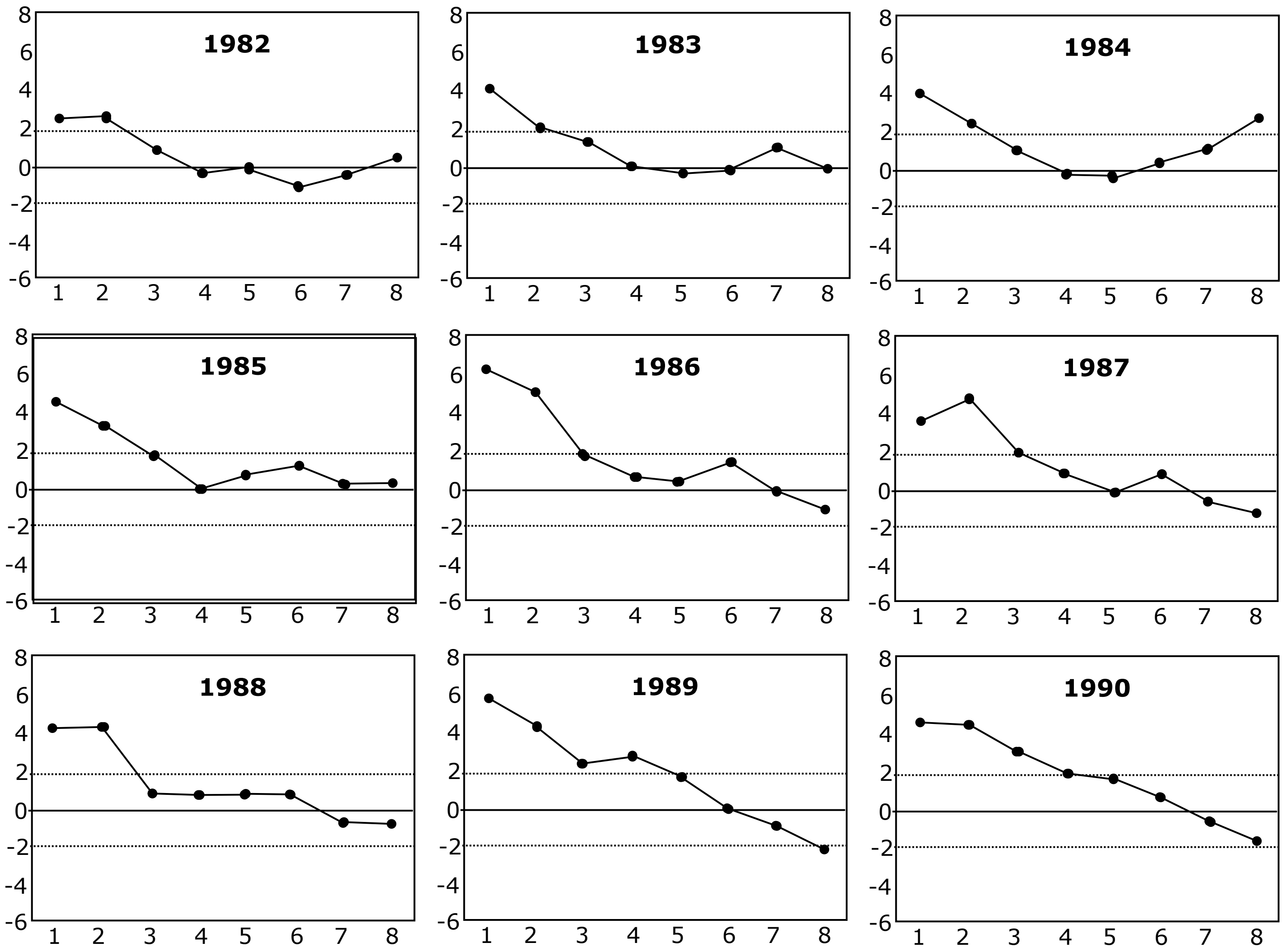
Correlogram

- A graph showing the value of Moran’s I at different distances

Correlogram:
Land Prices in a Japanese City



Example: Spatial-Temporal Spread of AIDS in Florida (1982-1990)

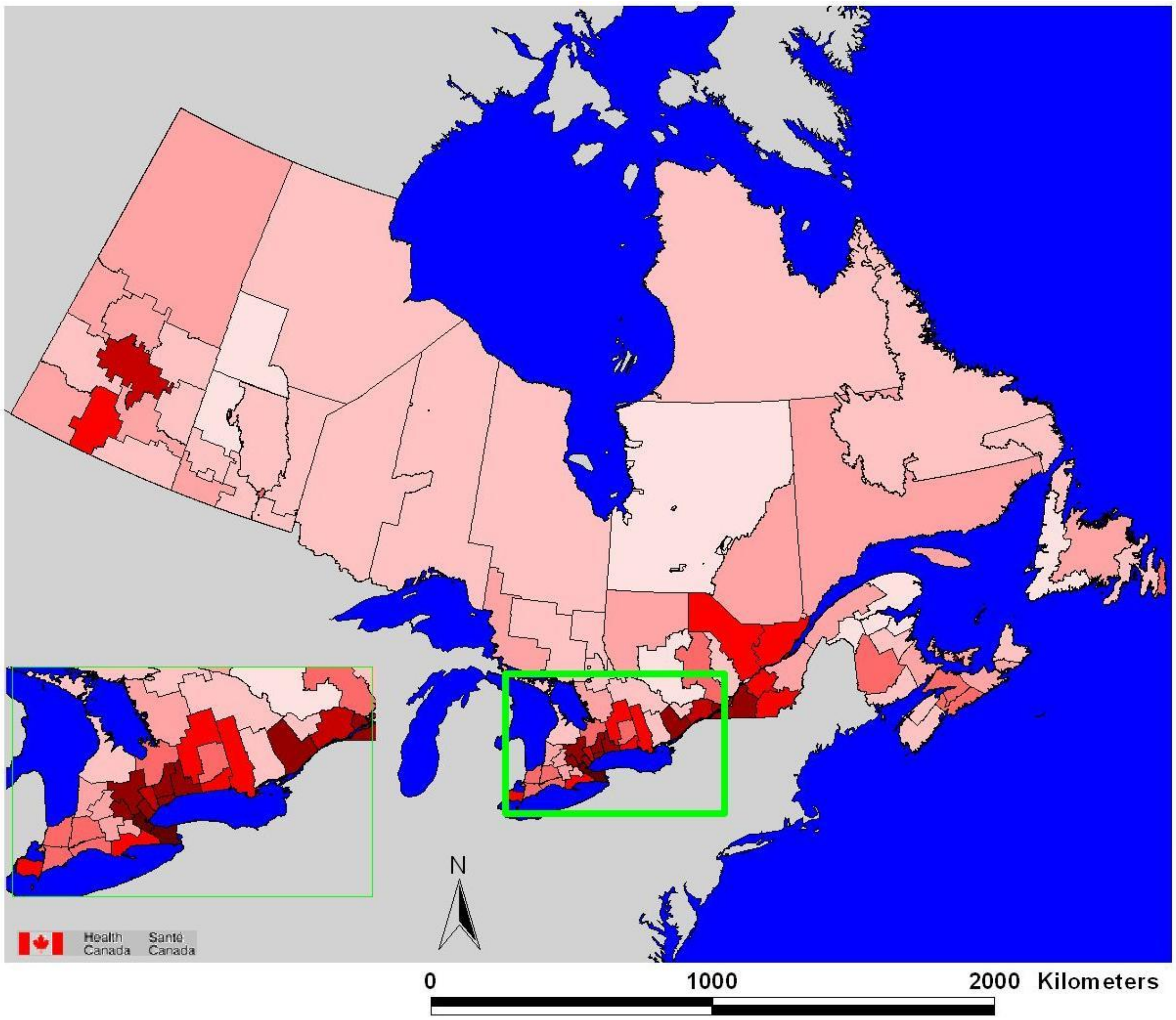


Source: N. Lam, M. Fan and K. Liu (1996)

Global Statistics

- Global patterns
- 1 statistic for all the study area
- What is happening at specific locations?

Global Statistics



Local Statistics

- Local Indicators of Spatial Association (LISA)

- Getis and Ord $G_i(d)$ statistic

LISA

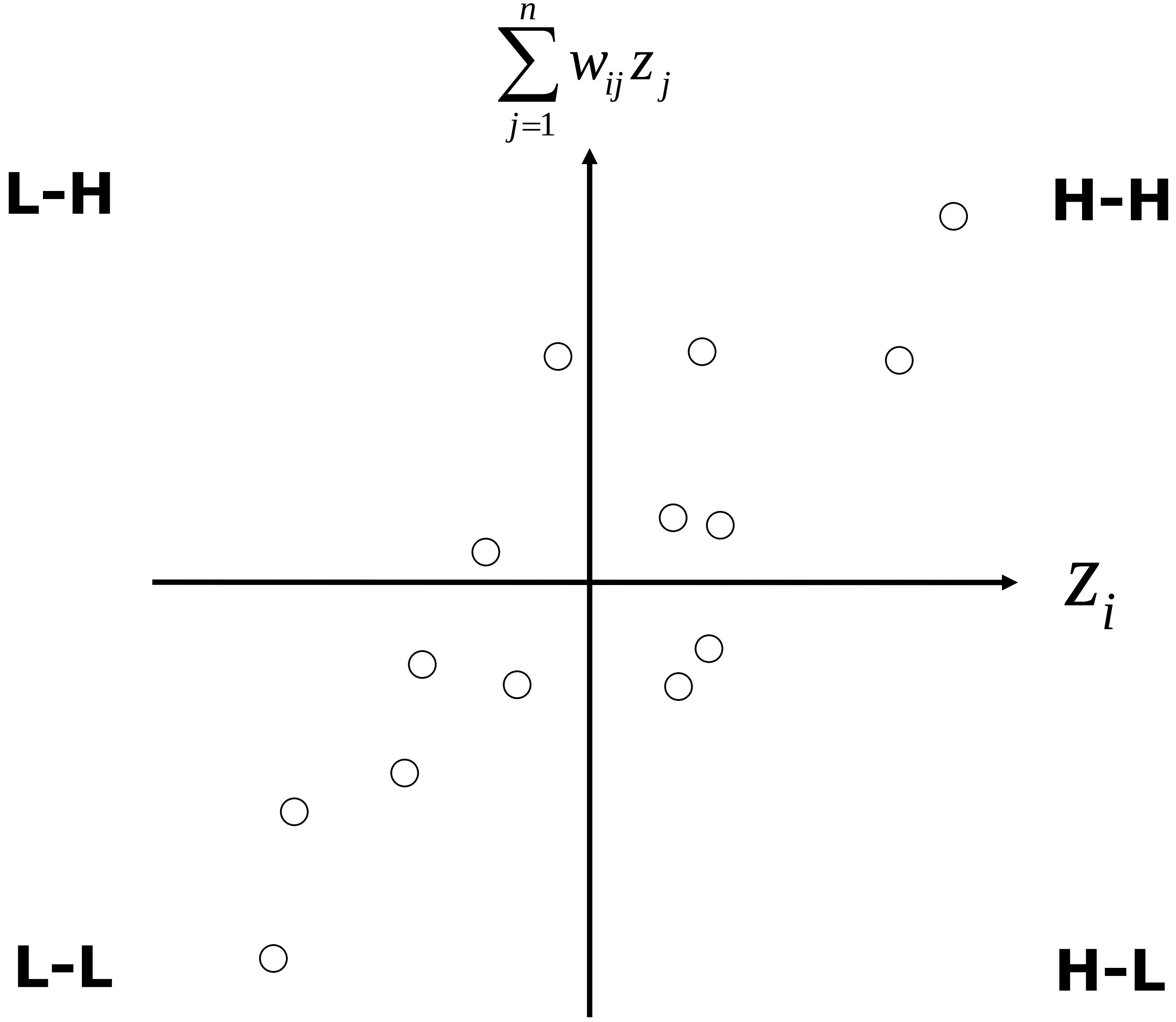
- Decomposition of Moran's I

$$I_i = \frac{z_i}{m_2} \sum_{i=1}^{n} w_{ij} z_j$$

$$m_2 = \frac{\sum_{i=1}^{n} z_i^2}{n}$$

*Patterns of association

Moran's Scatterplot



Moran’s Map:
Population Density Dijon (France)

MAP 6. Moran significance map for population density 1999 (contiguity weight matrix)

NOTES: D62, D66 and Lo4 are potential outliers detected by Bayesian heteroscedastic estimation

Source: C. Baumont, C. Ertur, J. LeGallo (2003)

Moran's Map: Population Density Dijon (France)

MAP 4. Moran significance map for employment density 1999 (contiguity weight matrix)

Source: C. Baumont, C. Ertur, J. LeGallo (2003)

Getis and Ord $G_i(d)$ statistic

- Indication of the concentration or lack of concentration of a variable

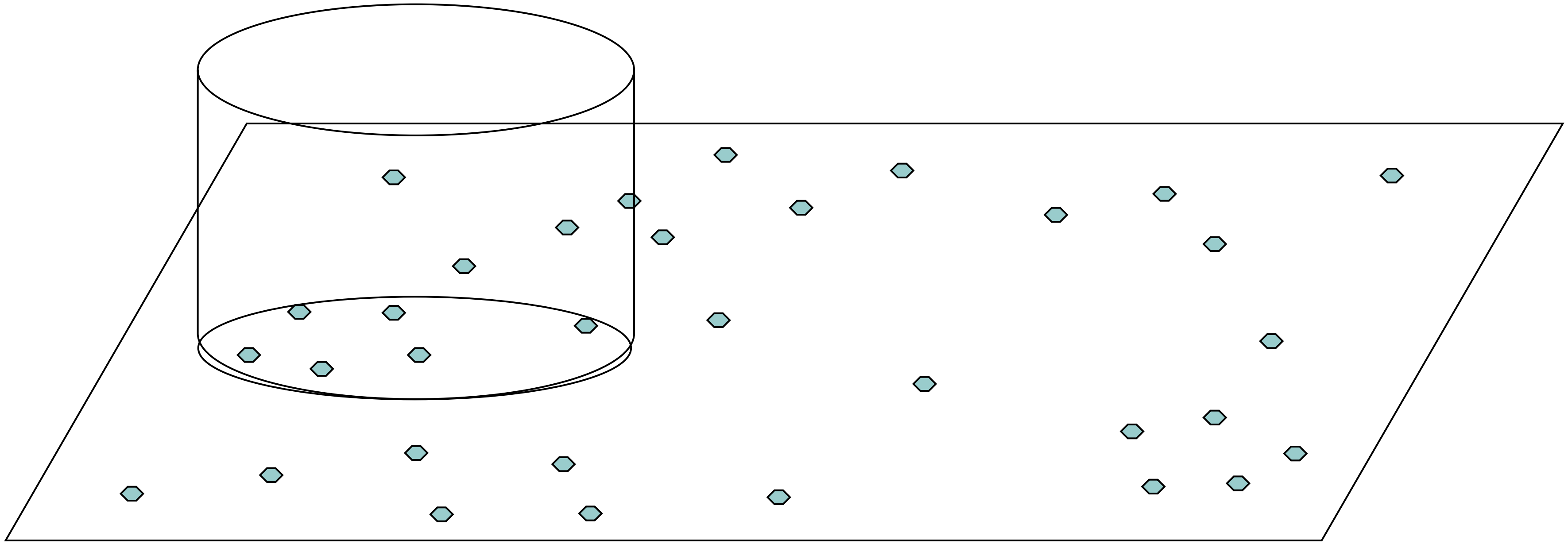
$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{i=1}^{n} x_j}$$

Getis and Ord $G_i(d)$ statistic

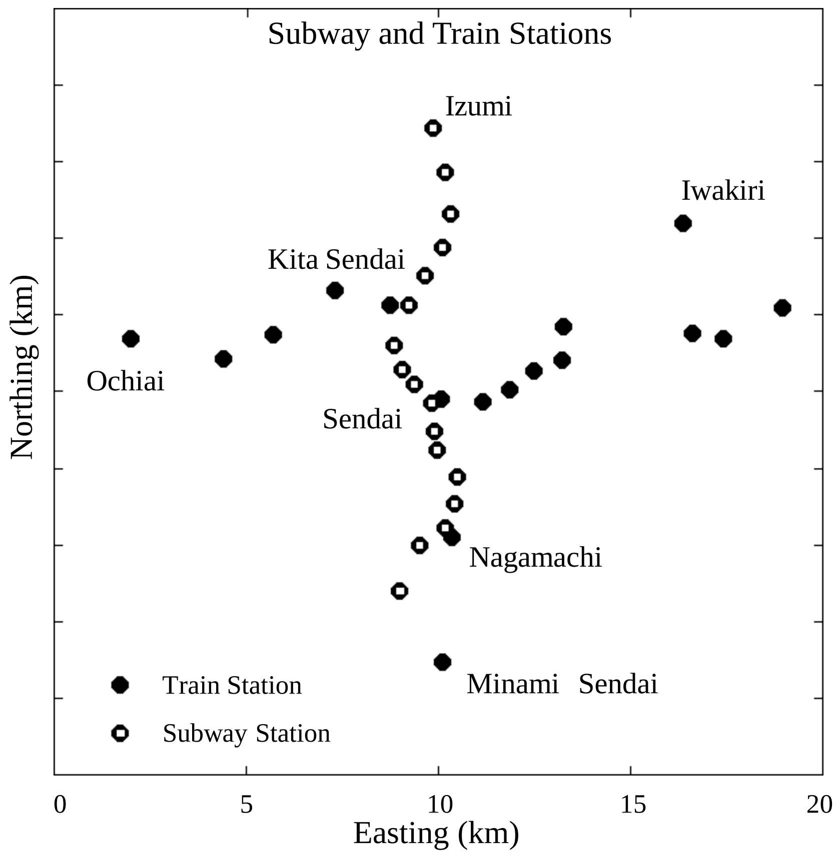
- Binary proximity matrix

$$w_{ij}(d) = \begin{cases} 1 & \text{if } d_{ij} < d \\ 0 & \text{otherwise} \end{cases}$$

Getis and Ord $G_i(d)$ statistic



Getis and Ord $G_i(d)$ statistic



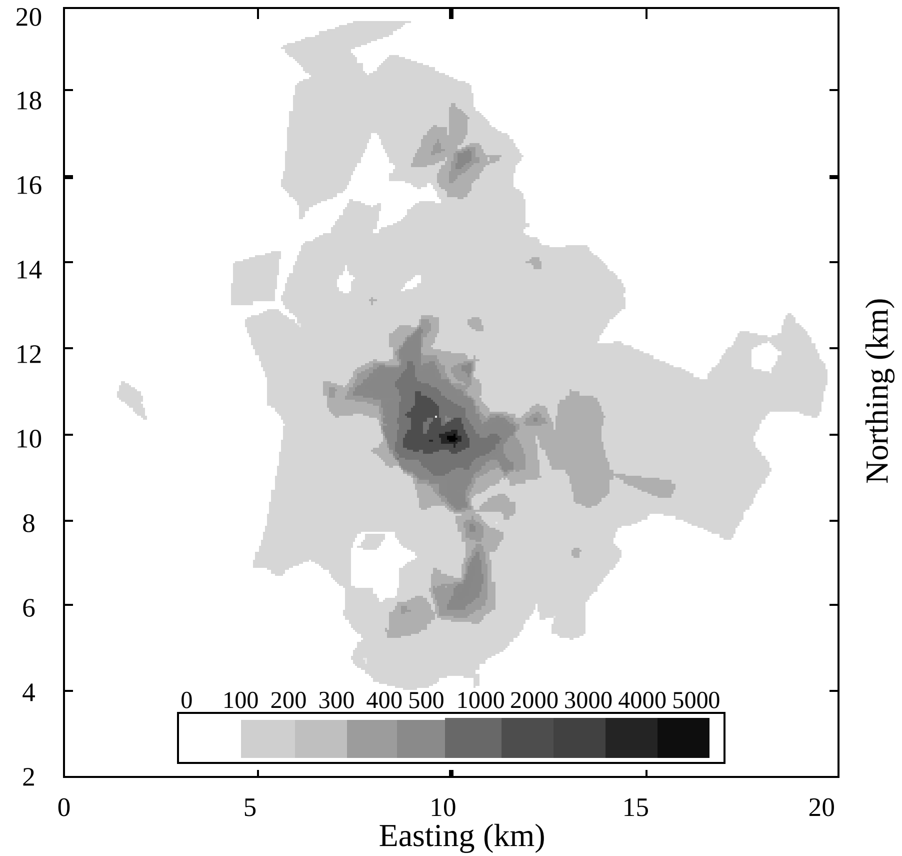
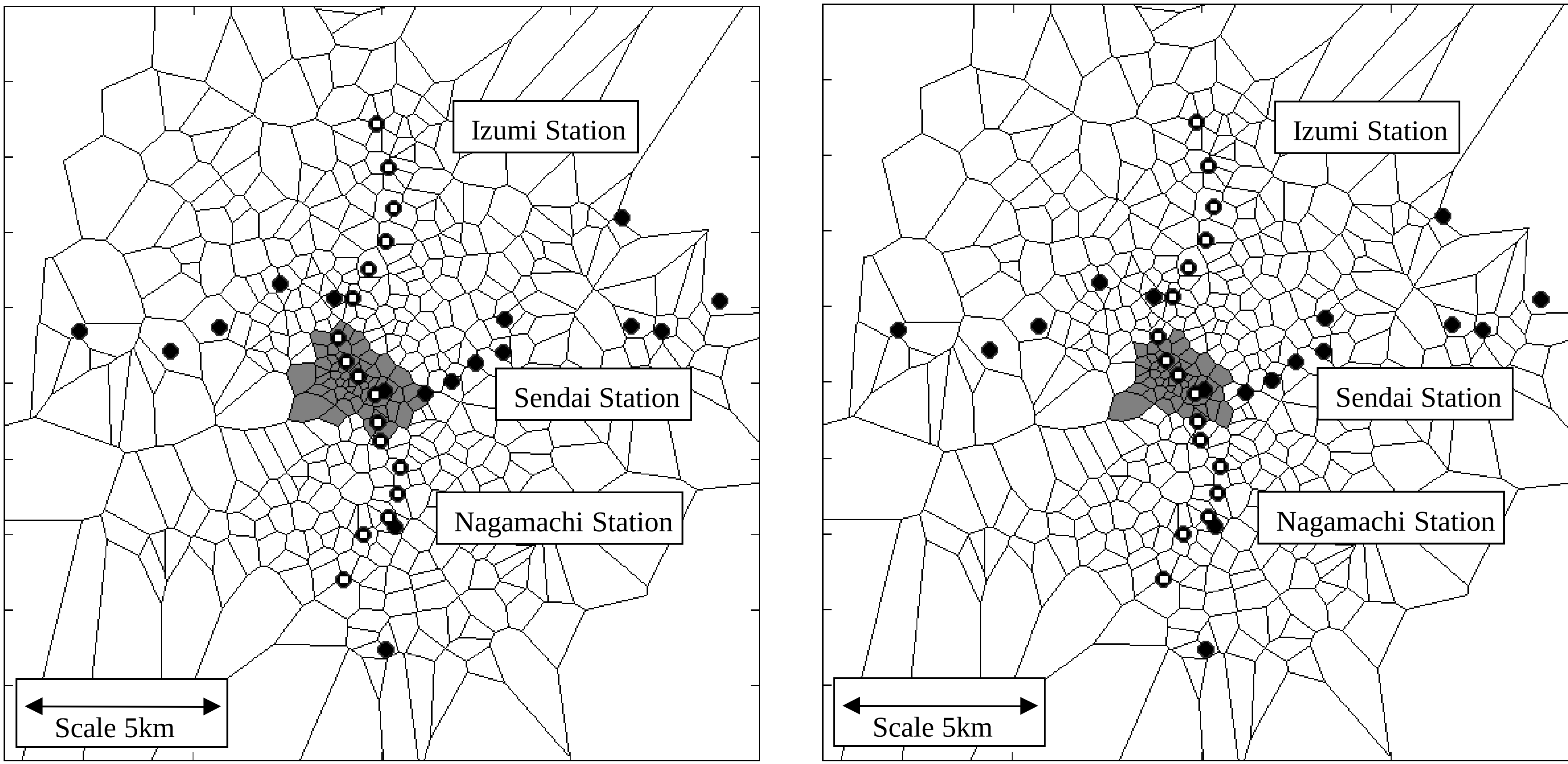


Fig. 1 a) Subway and train stations in Sendai City; b) Land price profiles (unit: \10 000/m^2)

Getis and Ord $G_i(d)$ statistic

Case 1) Significance Test α = 0.05 (Without Correction)

Case 2) Significance Test α = 0.0001 (With Correction)



$Z(G_i)$ Significant Observation (+)

$Z(G_i)$ Non-significant Observation

Fig. 2 Local autocorrelation analysis: a) Without Bonferroni correction; b) With Bonferroni correction

Example

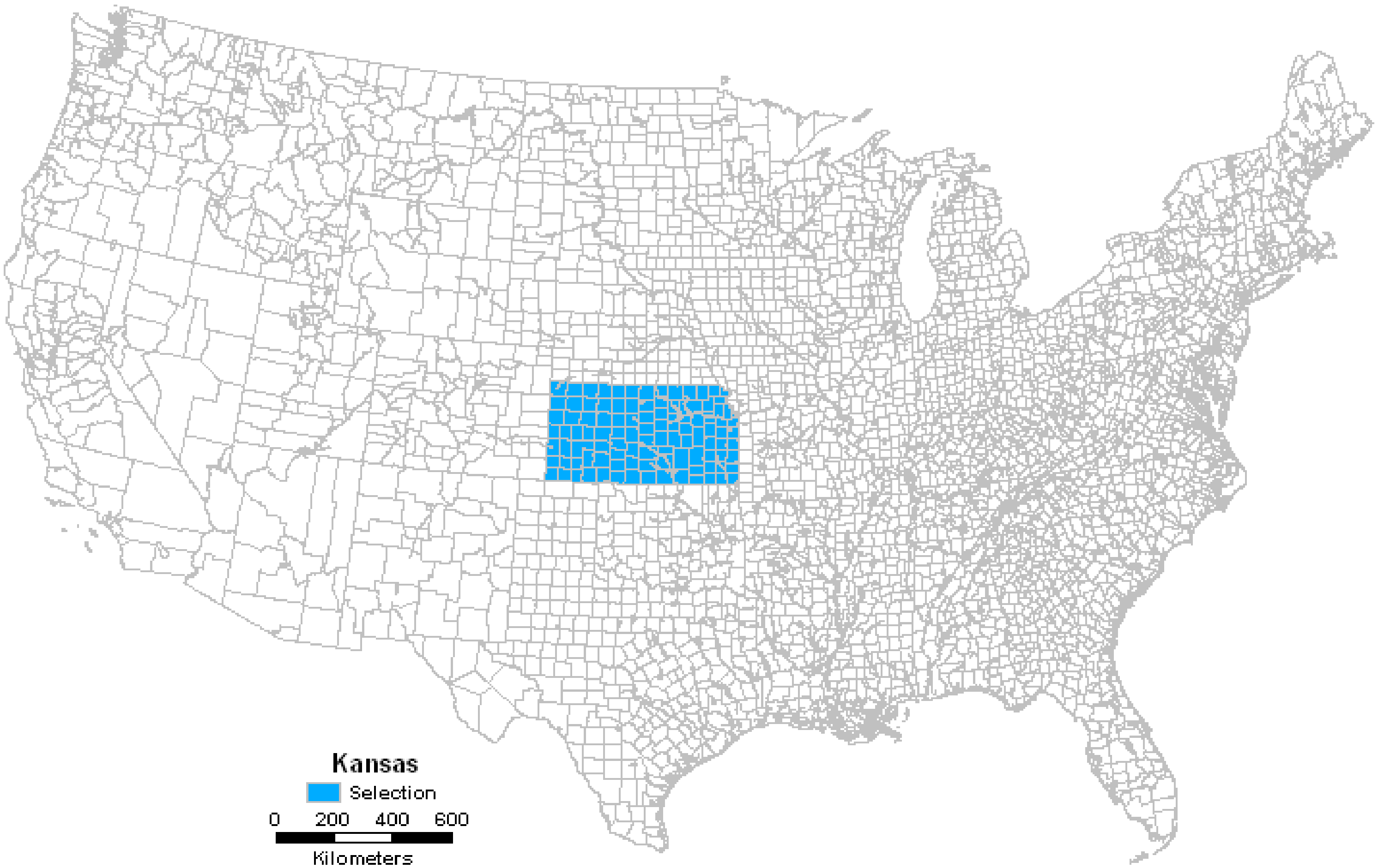
- Exploratory Data Analysis
- Visualization and exploration of area data
  - First order properties
  - Second order properties

Example

- Presidential election voter turnout (Kansas, 1980)

Example

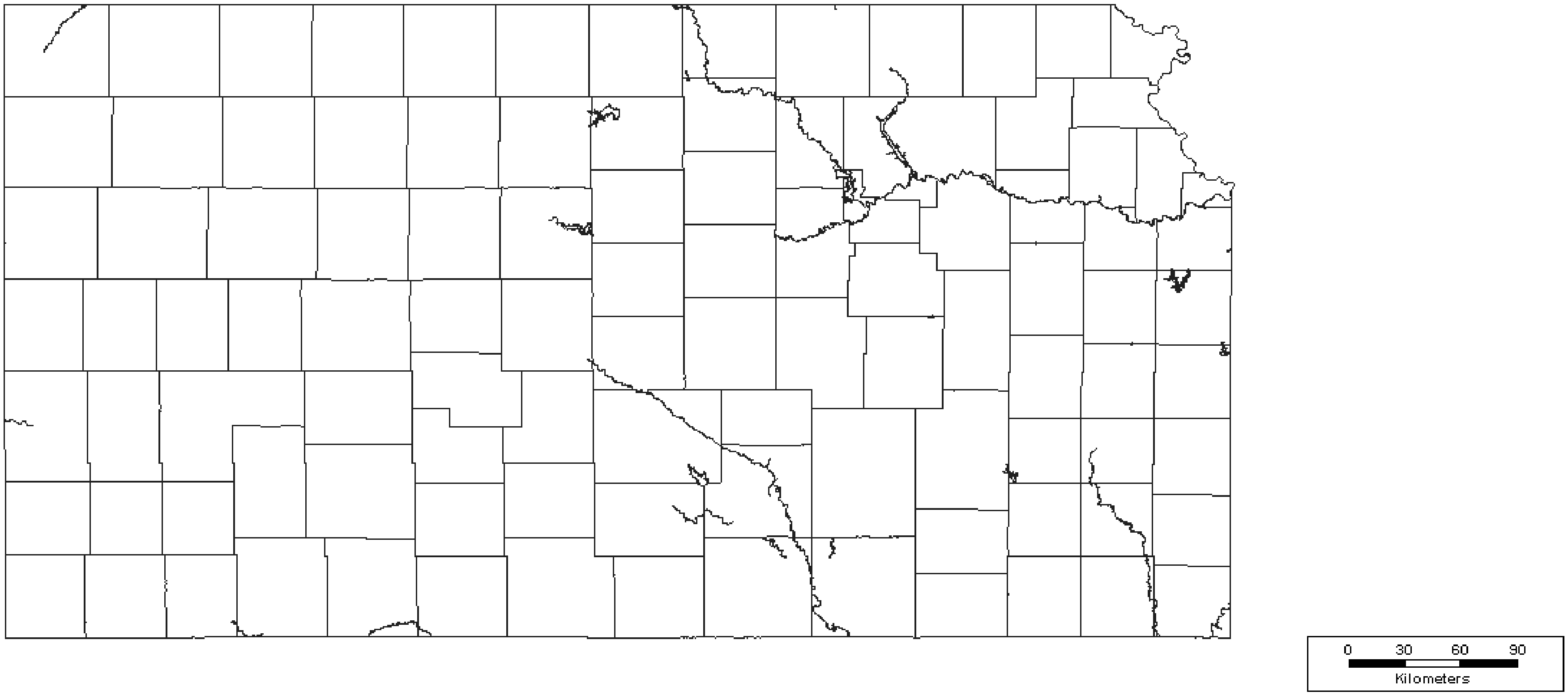
- Kansas



Definitions: Irregular Lattice

- Counties in Kansas State

Produced by Academic TransCAD



Example

- Presidential election voter turnout (Kansas, 1980)
  - Total number of votes cast per county in the 1980 presidential election for both candidates (VOTES)
  - Voting age population (POP)
  - Population in each county with 12^{th} grade or higher education (EDU)
  - Number of owner occupied houses (HOUSE)
  - Aggregate income in the county (INCOME)

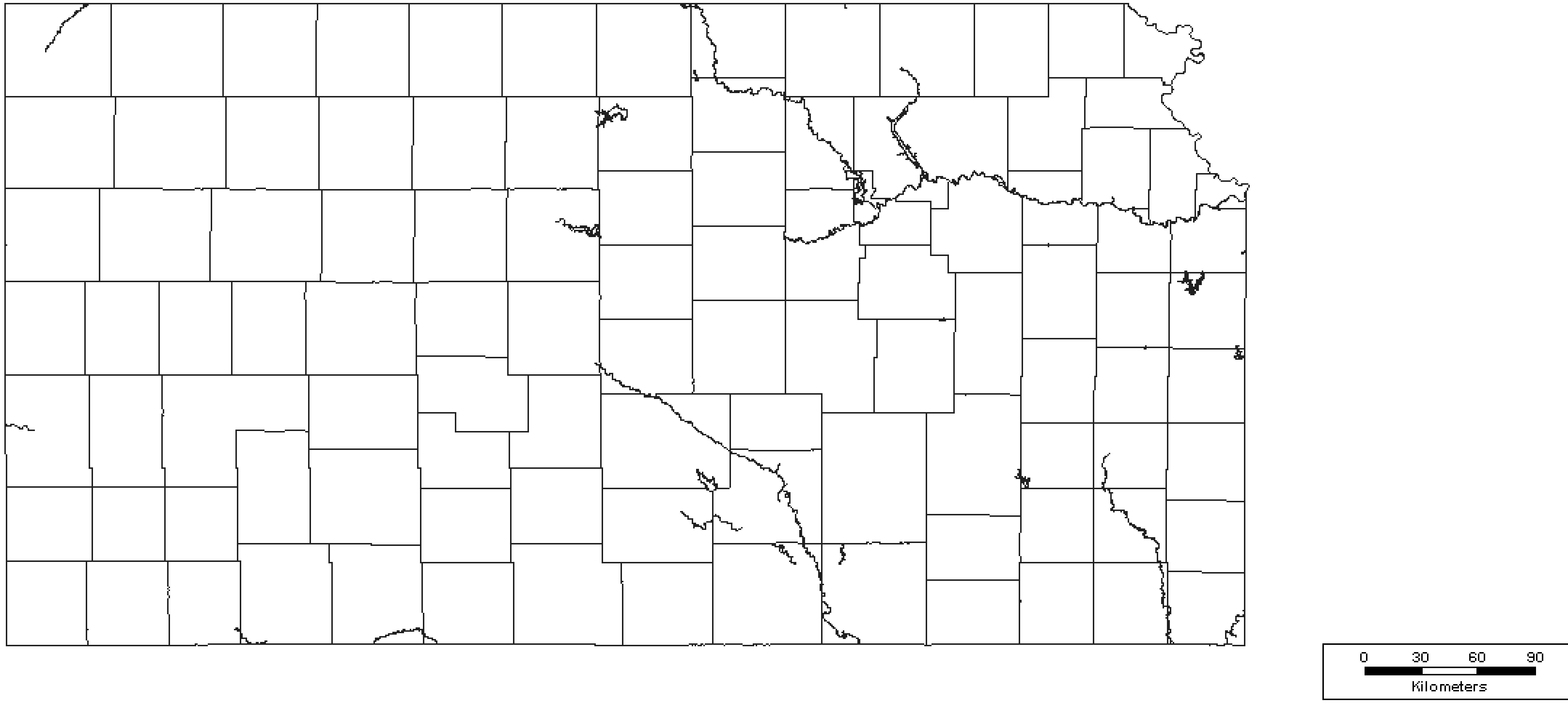
Example

- Presidential election voter turnout (Kansas, 1980)
  - Potential issues
    - What are the factors that influence voter turnout?
    - How can voter turnout be increased?
    - Certain population segments are more commonly identified with a given political party

Example: Voter turnout

- Zoning system: 105 counties

Produced by Academic TransCAD



Example: Voter turnout

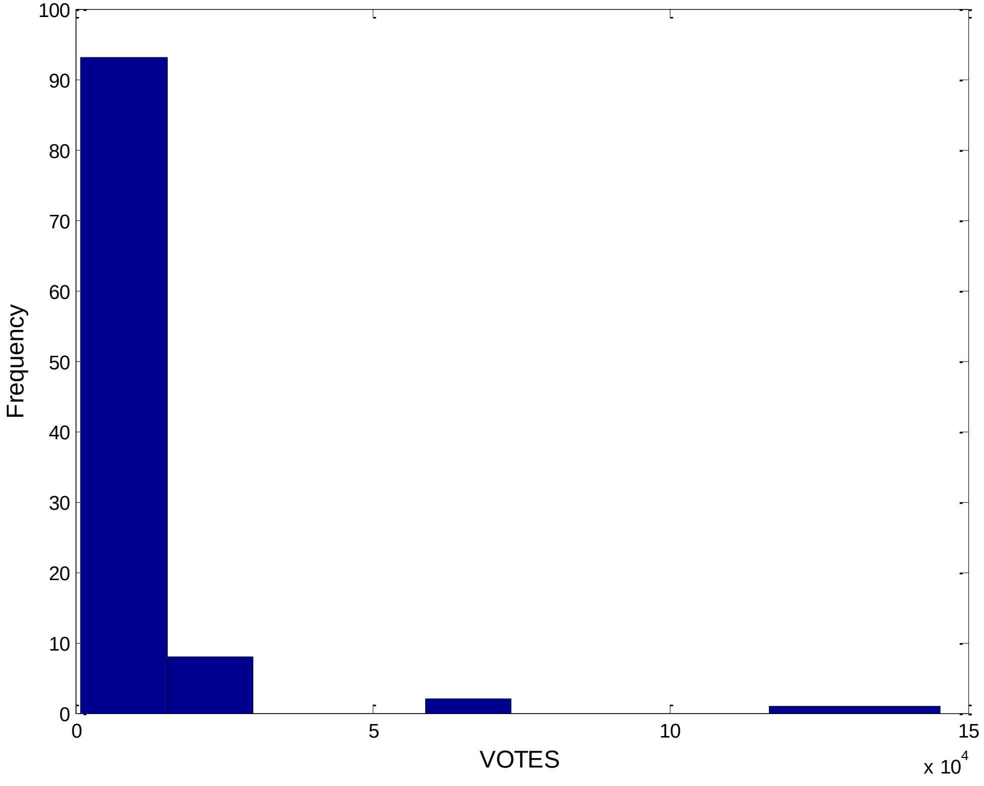
- Zoning system: 105 counties
  - Proximity matrix defined in terms of first order contiguities

Exploratory Data Analysis

- Univariate analysis
  - Histogram
  - Maps
  - Moving averages
  - Autocorrelation

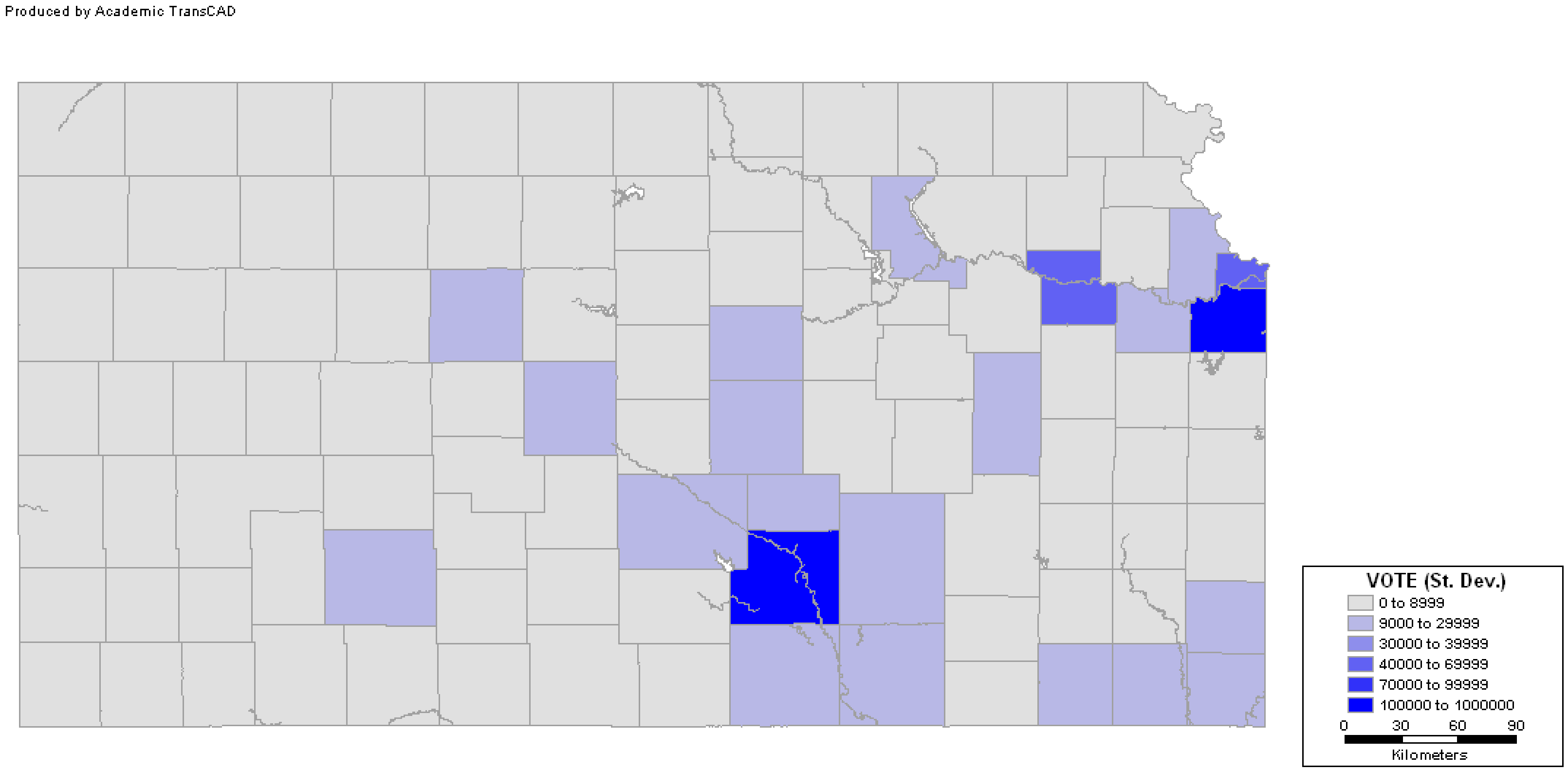
Exploratory Data Analysis

- Histogram: Votes



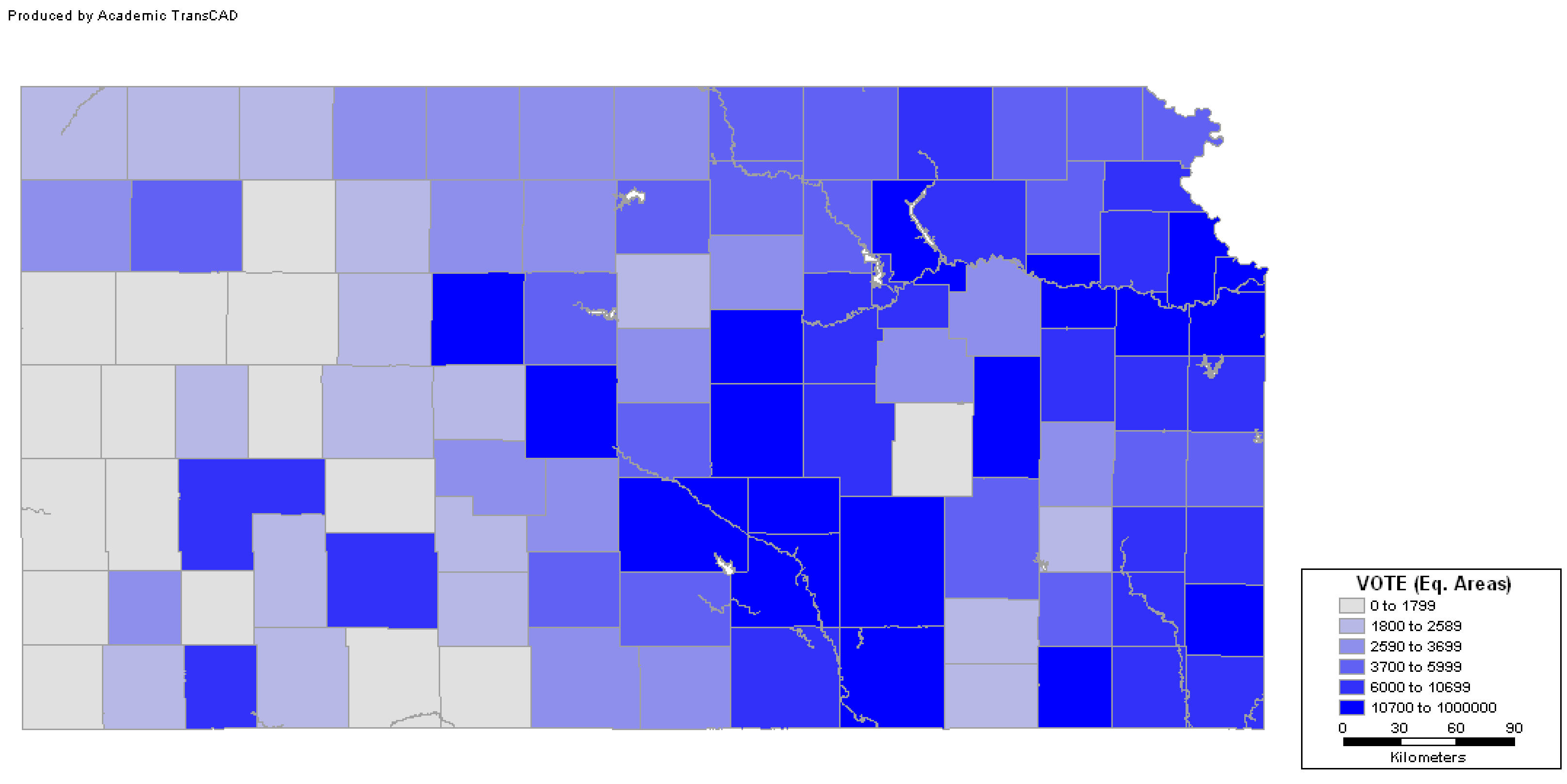
Exploratory Data Analysis

- Choropleth map: Votes (Std. Dev.)



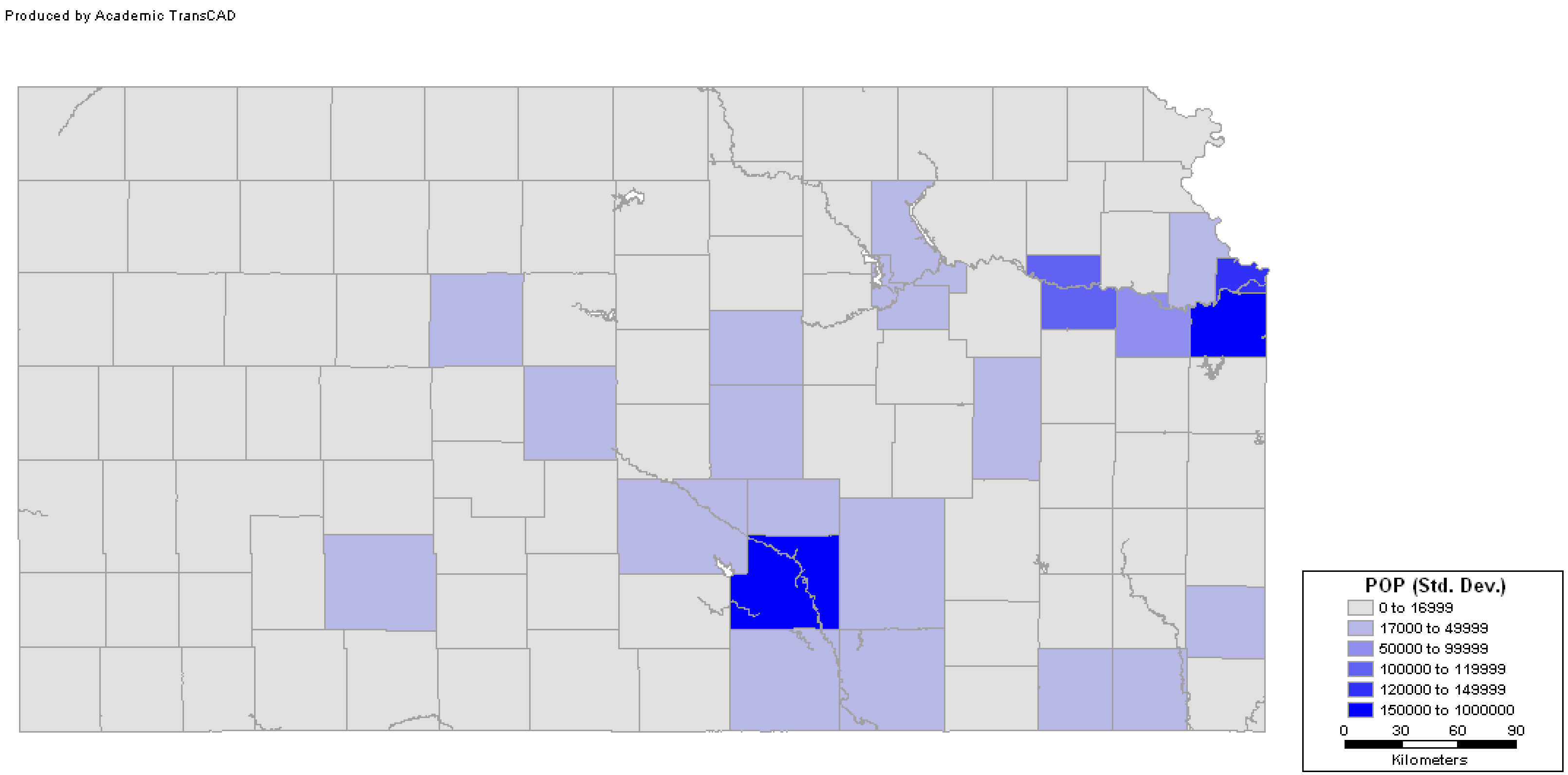
Exploratory Data Analysis

- Choropleth map: Votes (Eq. Areas)



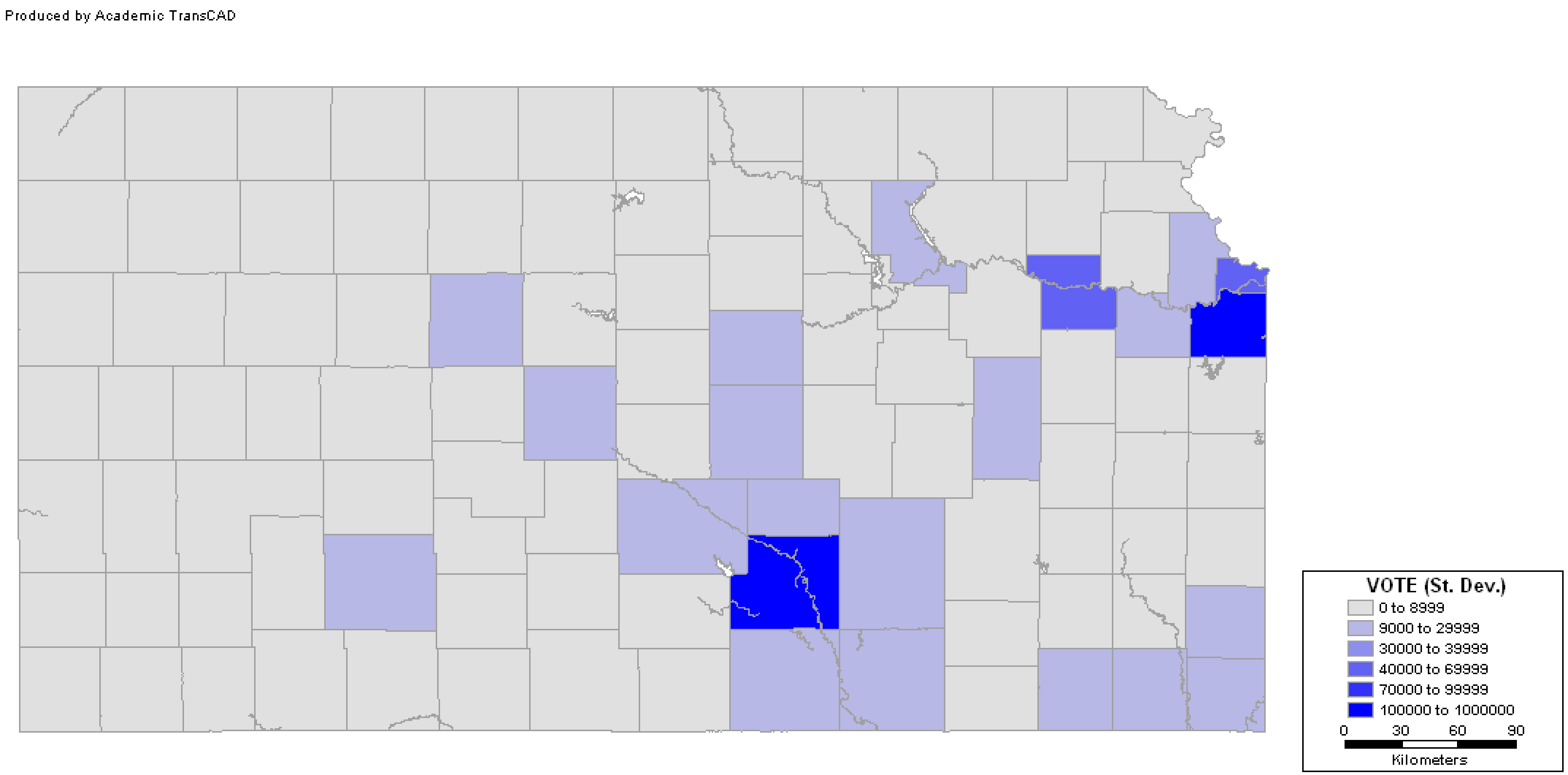
Exploratory Data Analysis

- POP (Std. Dev.)



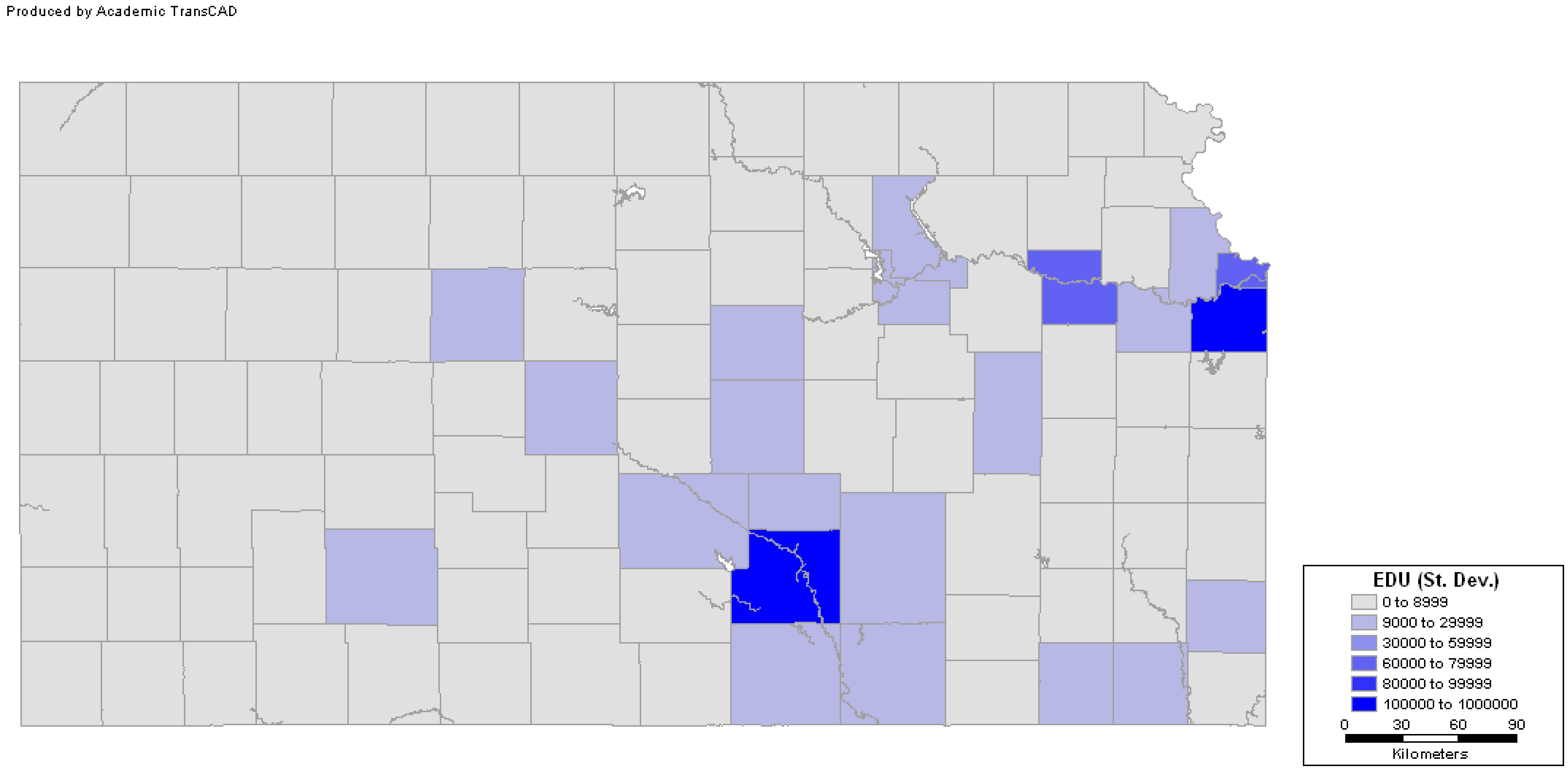
Exploratory Data Analysis

- Votes



Exploratory Data Analysis

- EDU (Std. Dev.)

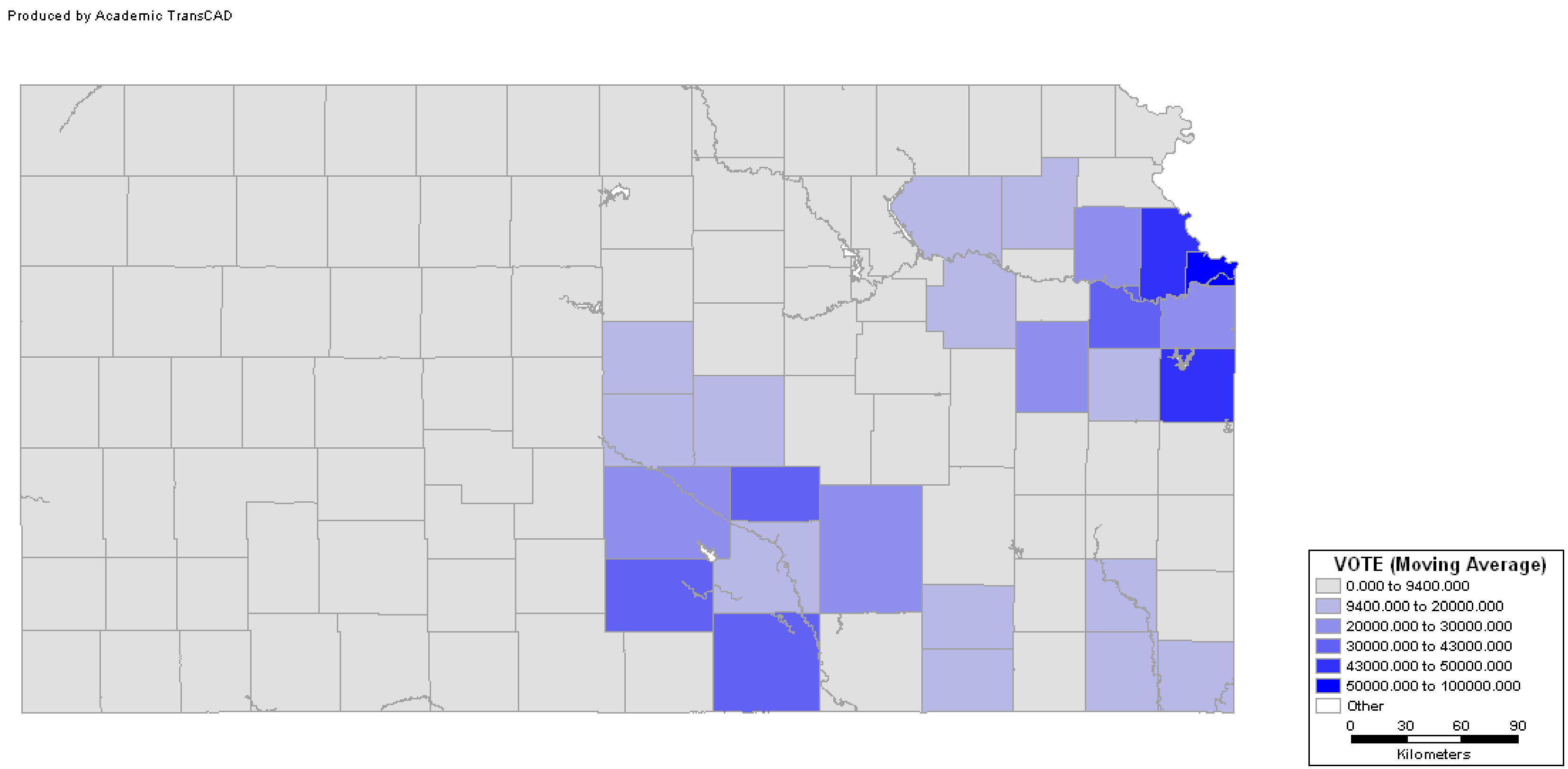


Exploratory SPATIAL Data Analysis

- Spatially moving averages (first order properties)
- Spatial autocorrelation (second order properties)

Exploratory Spatial Data Analysis

- Votes (Moving Average)



Exploratory Spatial Data Analysis

- Spatial autocorrelation
  - Standardized Moran's I (normal distribution)
  - Null hypothesis: no autocorrelation
  - All tests are significant ($p<0.05$)

Variable	$Z(I)$
VOTE	3.47
POP	3.78
EDU	3.36
HOUSE	3.75
INCOME	3.14

Conclusion

- Voting patterns follow a spatial pattern
- The spatial pattern is very clear, and probably relates to urban concentrations
- Voting is probably correlated with population, education and other variables

Next…

- Area data V & VI
  - Non-spatial regression
  - Error autocorrelation and other guidelines for model evaluation
  - Spatial regression