Spatial Autocorrelation: Theory and Concepts

GEOG 215 - April 6, 2020

Today's Agenda

- SPATIAL AUTOCORRELATION
    - Moran's *I*
        - Global Moran's *I*
        - Local Moran's *I*
        - Spatial Correlograms
        - Extensions to traditional Moran's *I*

Recall

- Tobler's first law of Geography
  - *Everything is related to everything else, but near things are more related than distant things*
    - Values at locations near each other tend to be similar, with similarity decreasing with distance
  - Implies that phenomena are not distributed randomly (throughout space)
    - Imagine how the world would appear if everything were randomly distributed

Moran's I

- Oft-used statistc for describing/testing the spatial autocorrelation within a region
  - Global (Considers the whole region)
    - Measures the magnitude of spatial autocorrelation
    - Returns a single result (***I***)
    - In addition, it provides a p-value
    - Probability value to test help statistical significance

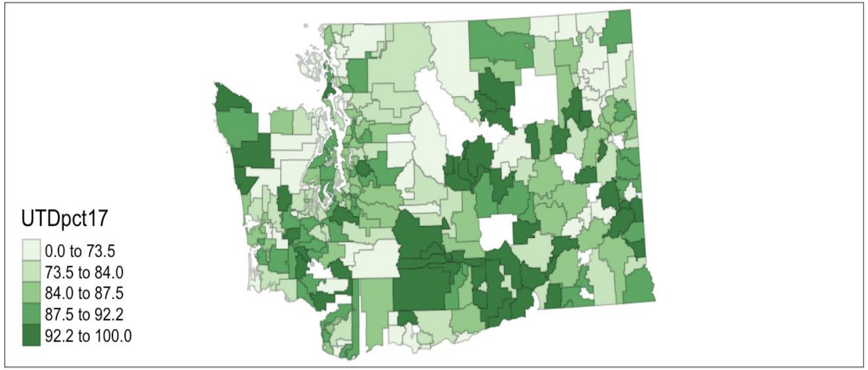
Moran's I

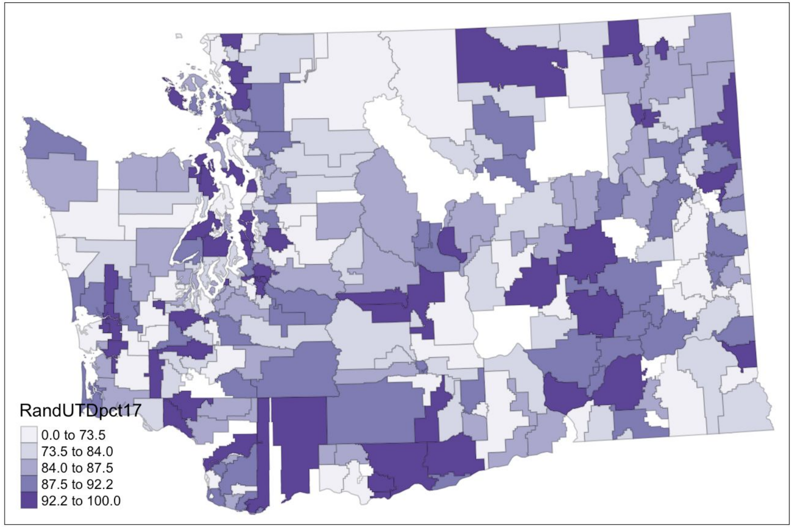
- Global value
    - Moran's ***I*** ranges from -1 to 1 (continuous)
        - Perfectly dispersed/ordered: -1
        - Random: 0
        - Perfectly Clustered: 1 (close to 1 actually)
            - Some exceptions
    - Compares ***I*** of observed data to expected ***I*** (expected is under Complete Spatial Randomness)

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Complete Spatial Randomness (CSR)

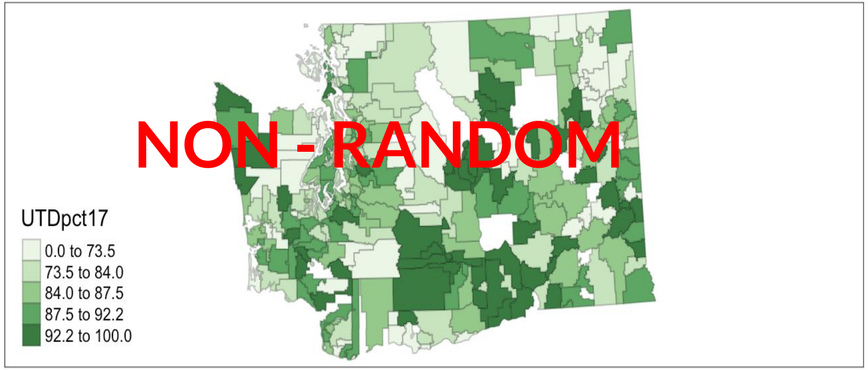
- Imagine that you could pick up the values for the attribute you are analyzing and throw them down onto your features, letting each value fall where it may
  - Random chance spatial process
    - Many many possible realizations (no single pattern)
  - Usual reference point (null hypothesis) for Moran's I

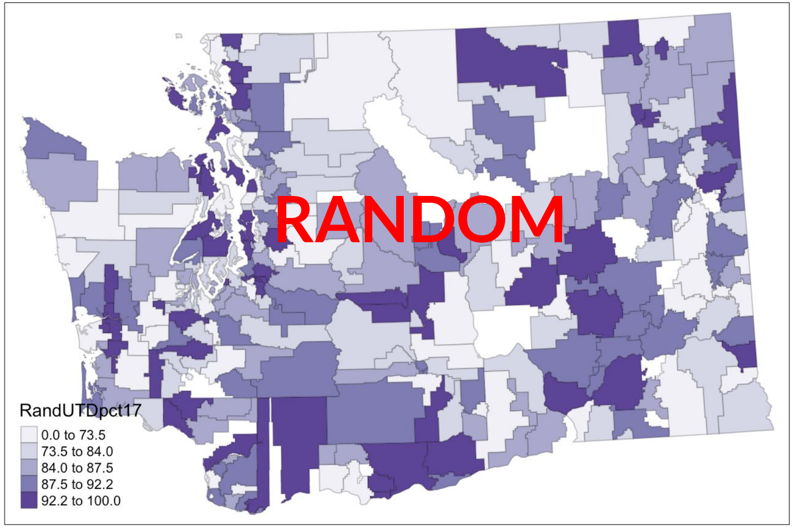




Complete Spatial Randomness (CSR)

- Imagine that you could pick up the values for the attribute you are analyzing and throw them down onto your features, letting each value fall where it may
  - Random chance spatial process
    - Many many possible realizations (no single pattern)
  - Usual reference point (null hypothesis) for Moran's I





Moran's I

- Formula

$$I = \frac{n}{S} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \overline{X})(x_j - \overline{X})}{\sum_{i=1}^{n} (x_i - \overline{X})^2}$$

n = number of areas
w_{ij} = the weight between area i and j
x_i = the value for area i
x_j = the value for area j
$\overline{X}$ = mean of all values
S = sum of all weights

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$$

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Moran's I

- Formula

For every i, compare deviation from mean of self value and neighbors' values

$$I = \frac{n}{S} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \overline{X})(x_j - \overline{X})}{\sum_{i=1}^{n} (x_i - \overline{X})^2}$$

area i

area i's neighbors

n = number of areas
w_{ij} = the weight between area i and j
x_i = the value for area i
x_j = the value for area j
$\overline{X}$ = mean of all values
S = sum of all weights

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$$

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Moran's I

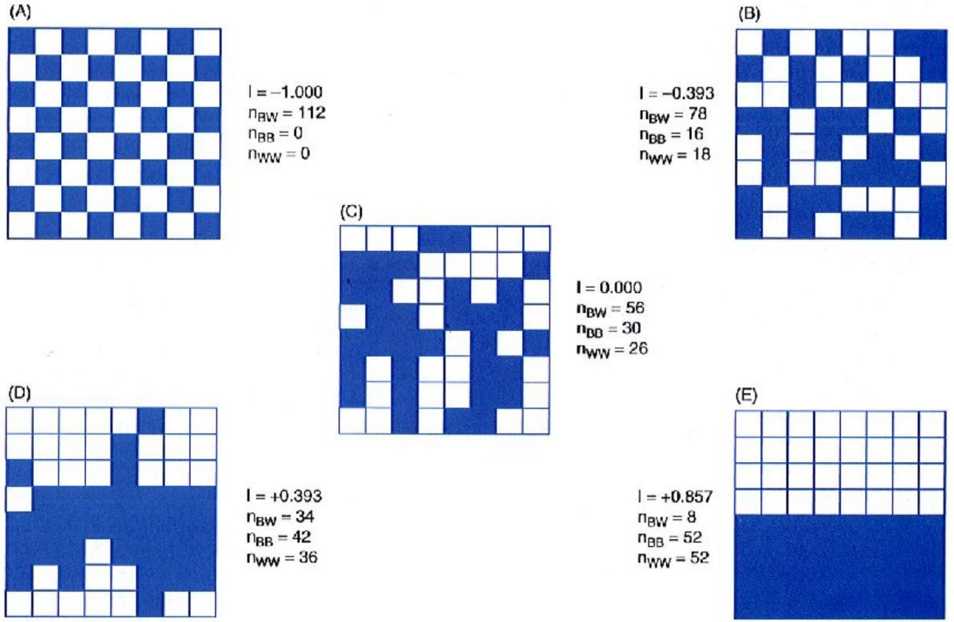


Figure 4.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation; (B) a dispersed arrangement; (C) spatial independence; (D) spatial clustering; and (E) extreme positive spatial autocorrelation. The values of the *I* statistic are calculated using the equation in Section 4.6 (*Source*: Goodchild 1986 CATMOG, GeoBooks, Norwich)

Moran's I

- Interpreting output
  - Magnitude
    - The closer to 1, the more clustered the values are
    - The closer to -1, the more dispersed the values are
    - Close to 0 means high likelihood of complete spatial randomness
  - Significance
    - Statistical
      - Interpret p-value (e.g., <0.05)
    - Substantive
      - Look at the actual values
      - Some phenomena have weaker clustering than others
      - Assess whether local statistics might be better

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm

Moran's I

- Interpreting output
  - Importance/Substantive significance
    - Beware of significant, but unimportant deviation from random pattern
      - For example, I = 0.04, $p < 0.001$
    - p - value is affected by sample size
      - Personal interpretation system:
        - 0 to 0.1, barely clustered (pretty much random)
        - 0.1 to 0.3, slightly clustered
        - 0.3 to 0.5, modertately clustered
        - > 0.5, highly clustered

Moran’s I

- Potential Applications
    - Help identify an appropriate neighborhood distance for a variety of spatial analysis methods by finding the distance where spatial autocorrelation is strongest.
    - Measure broad trends in ethnic or racial segregation over time—is segregation increasing or decreasing?
    - Summarize the diffusion of an idea, disease, or trend over space and time—is the idea, disease, or trend remaining isolated and concentrated, or spreading and becoming more diffuse?
    - Many more applications -
        - Is there an external variable that might be explaining clustering or dispersion patterns?
    - Examples of previous student projects

Moran's I

- Robustness test
  - Multiple neighborhood definitions
    - K-nearest Neighbors
    - Neighbors based on distance/ distance threshold
    - Contiguity based neighbors (queen / rook)
    - Neighbors based on connectivity / thiessen polygons

Table A2. Moran's *I* values for NME rate (%) for block group observations, under ten neighborhood definitions.

YEAR	ID(5)	ID(10)	ID(15)	ID(20)	KNN(5)	KNN(10)	KNN(15)	KNN(20)	CON(Q)	CON(R)
2000	0.073	0.076	0.078	0.079	0.106	0.093	0.083	0.079	0.119	0.104

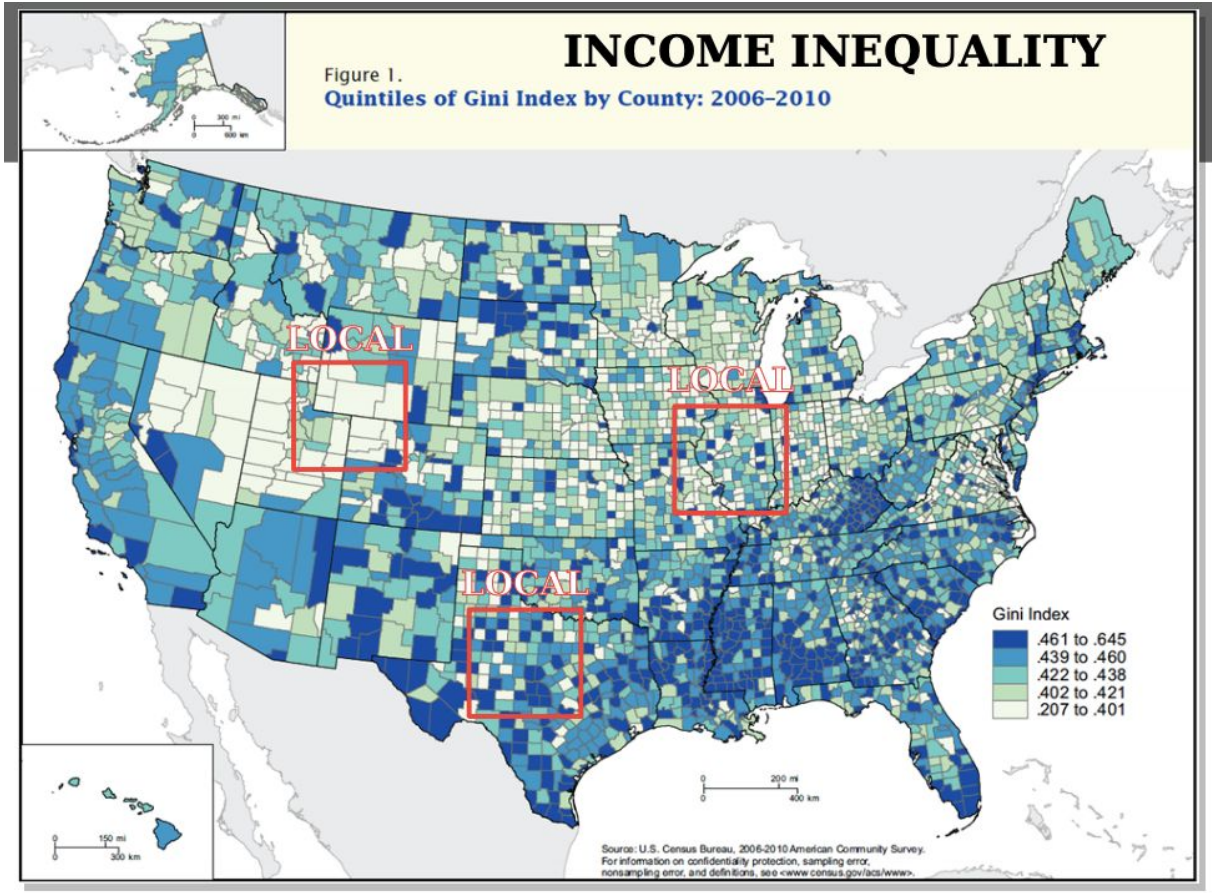
Moran's I

- Other important considerations
    - Works with continuous data
        - There are extensions for binary data
    - Measures autocorrelation - not correlation
        - You can use bivariate Moran' I (with caution)
    - Sensitive to outliers and skewed data
        - Check your variable histograms to see if your data is normal or not
            - If not, consider transforming to normal, run robustness checks
    - Comparisons are only relevant when the study area is fixed
        - And, when your variables are similar (counts, rates, etc)
    - Consider neighborhood definitions
        - Remove areas with no neighbors/ artificial neighbors due to edge effects
        - Neighborhood definition should represent real world process

Stationary vs Non-stationary

- Spatial autocorrelation
  - Global
    - Assumes that autocorrelation is stationary across space
      - Invariant from place to place
        - Similar to thinking about usefulness of an “average”
  - Local
    - Assumes that autocorrelation is non-stationary cross space
      - Varies from place to place

Stationary vs Non-stationary



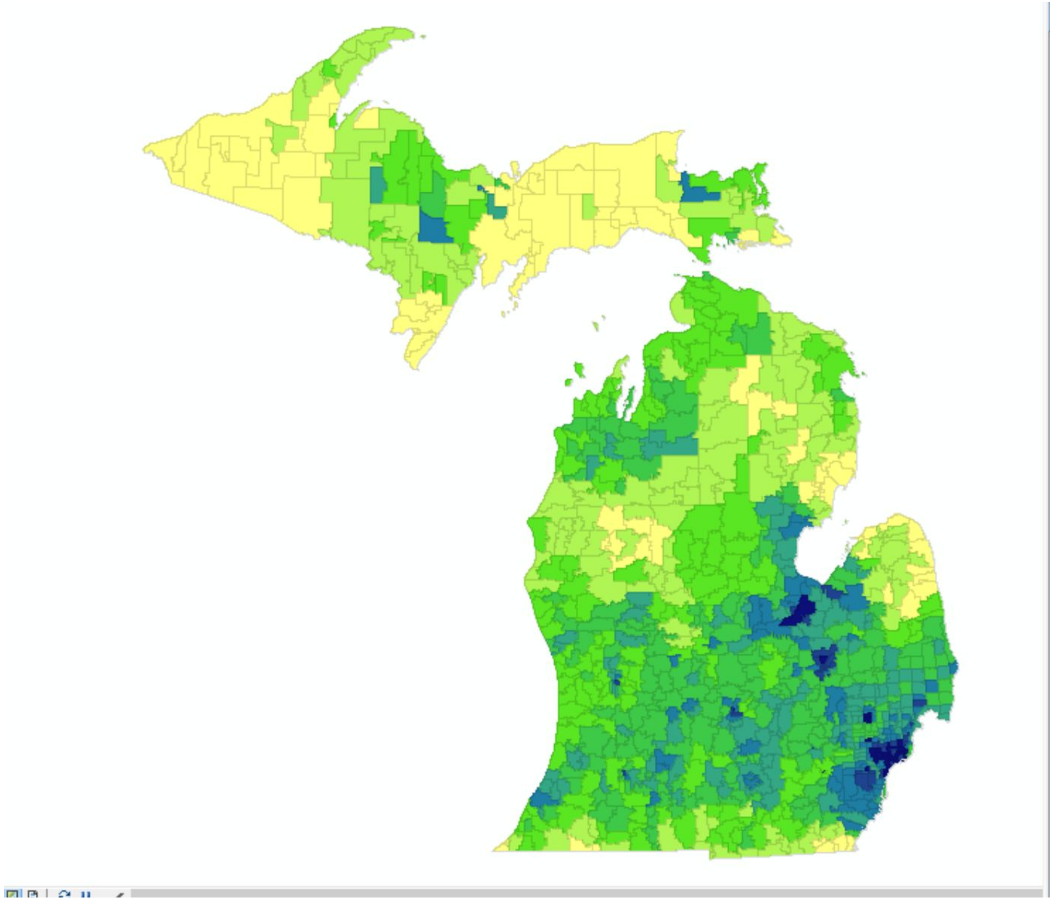
LISA

- Local Indicator of Spatial Association
  - Local version of Moran’s I
    - Iterates through each observation and provides a measure of autocorrelation
      - And, association p-value
    - Unlike global measures, results can be mapped
      - Reveals the nature of spatial autocorrelation throughout the study area

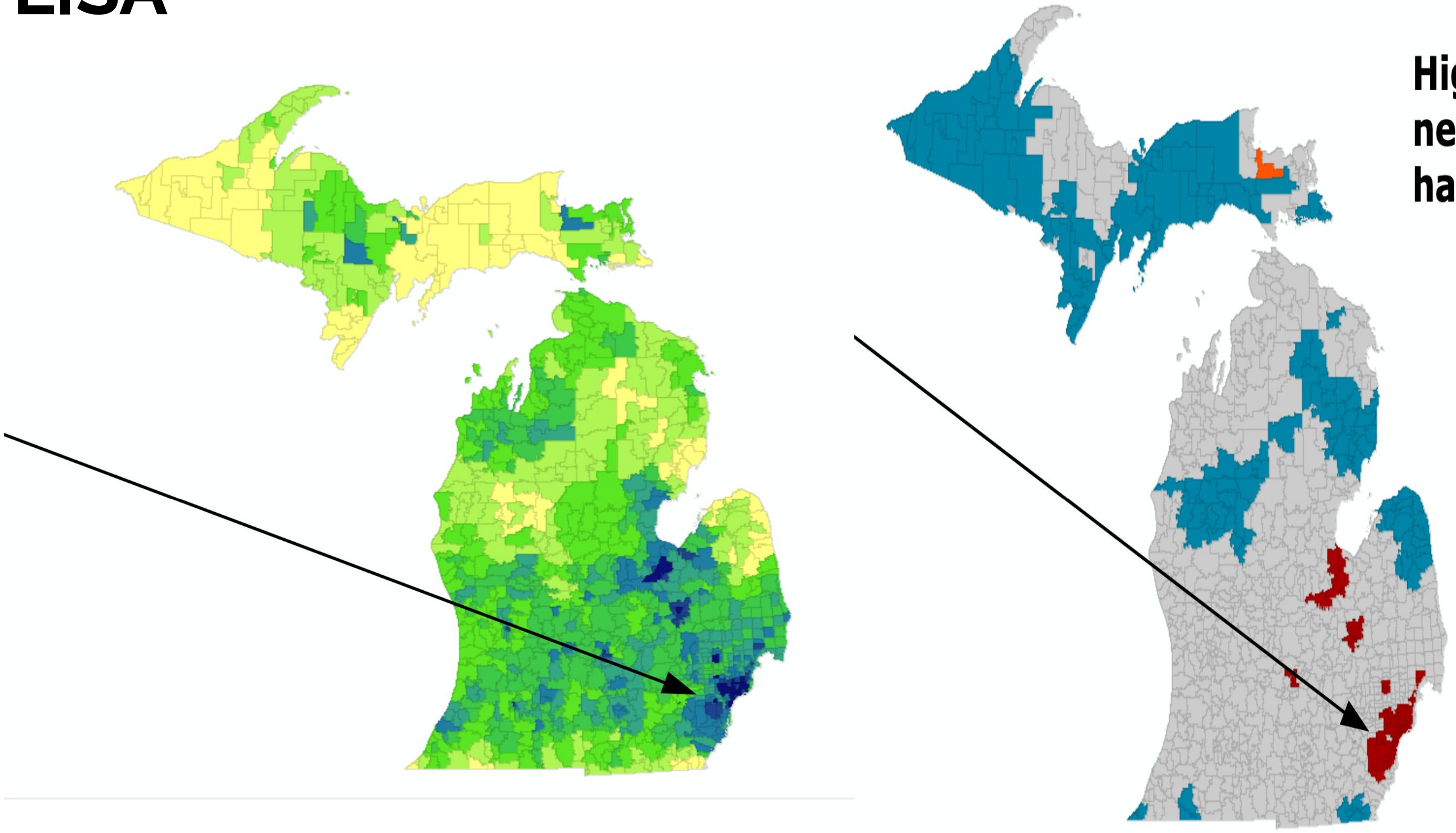
LISA

- Local Indicator of Spatial Association
  - Observations can be “hot” or “cold” spots, high or low outliers, or not significant
    - High-High (observation high, neighbors high)
    - Low-Low (observation low, neighbors low)
    - High outlier (observation high, neighbors low)
    - Low outlier (observation low, neighbors high)
  - Extremely useful for understanding “where” spatial autocorrelation is strong/weak

LISA

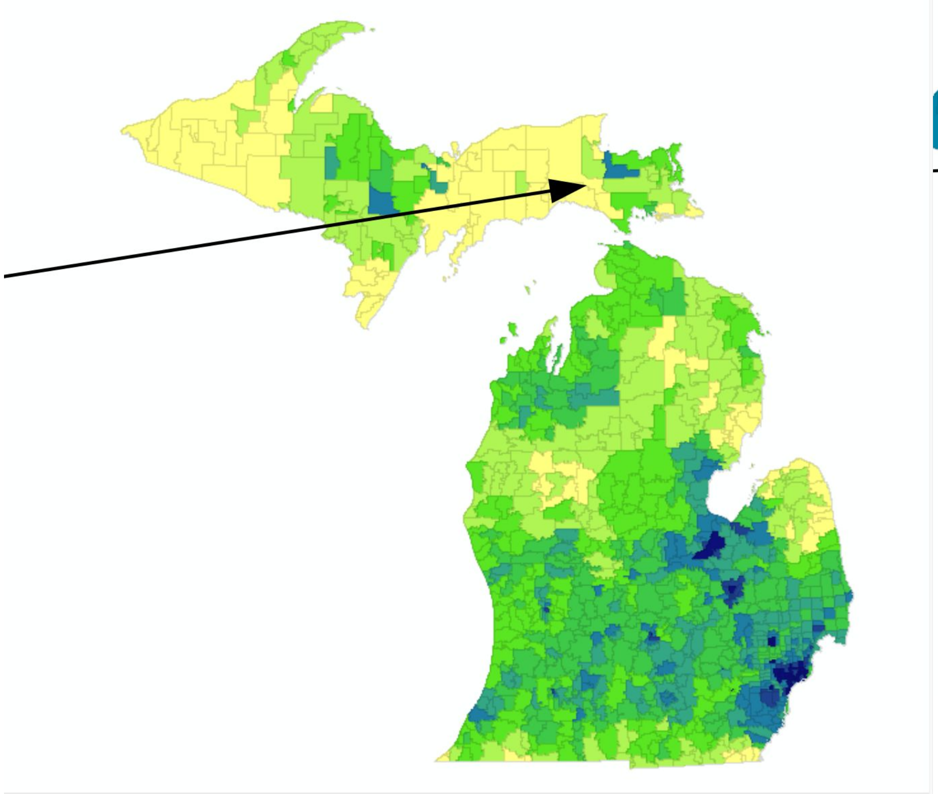


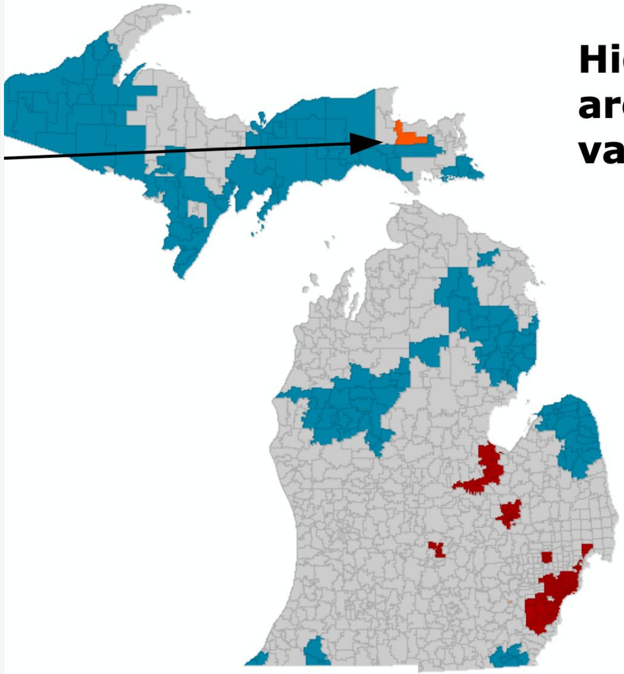
LISA



High values, with neighbors also having high values

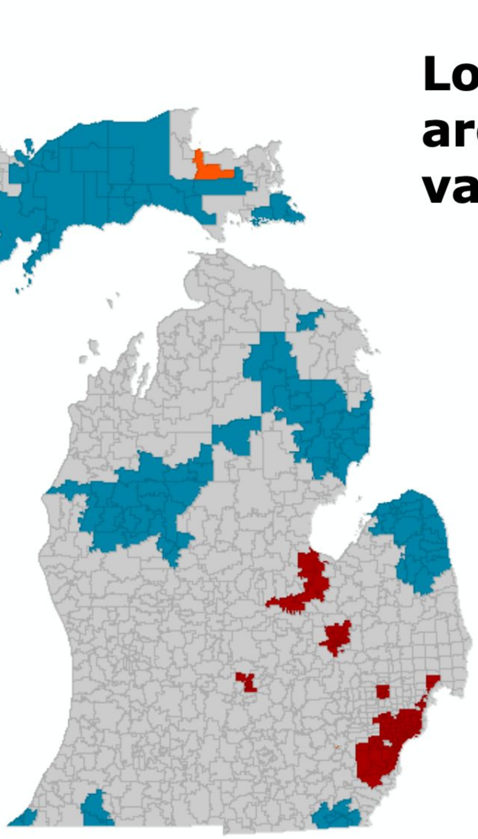
LISA





High value, in an area with low values

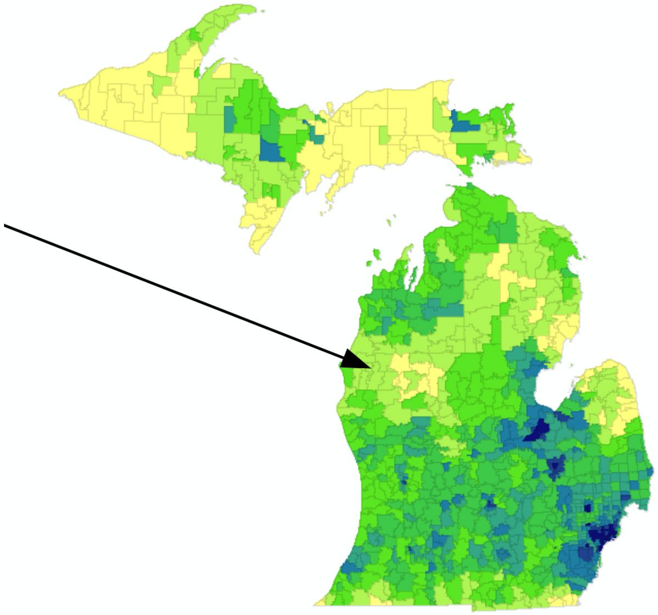
LISA

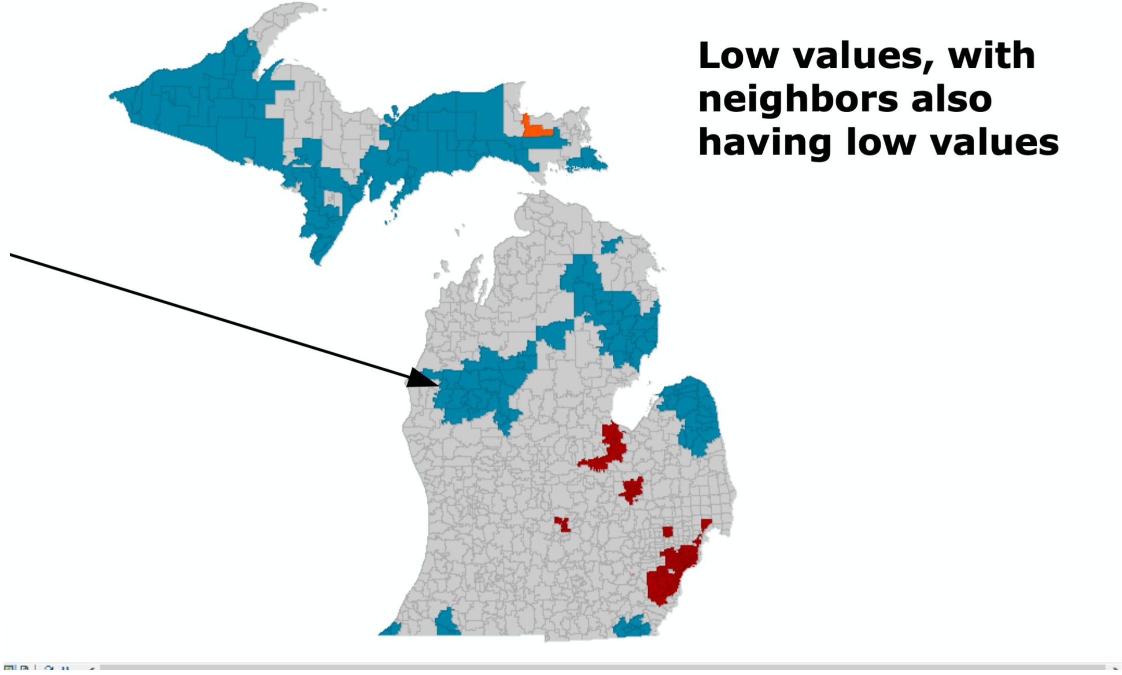


Low value, in an area with high values

NONE IDENTIFIED

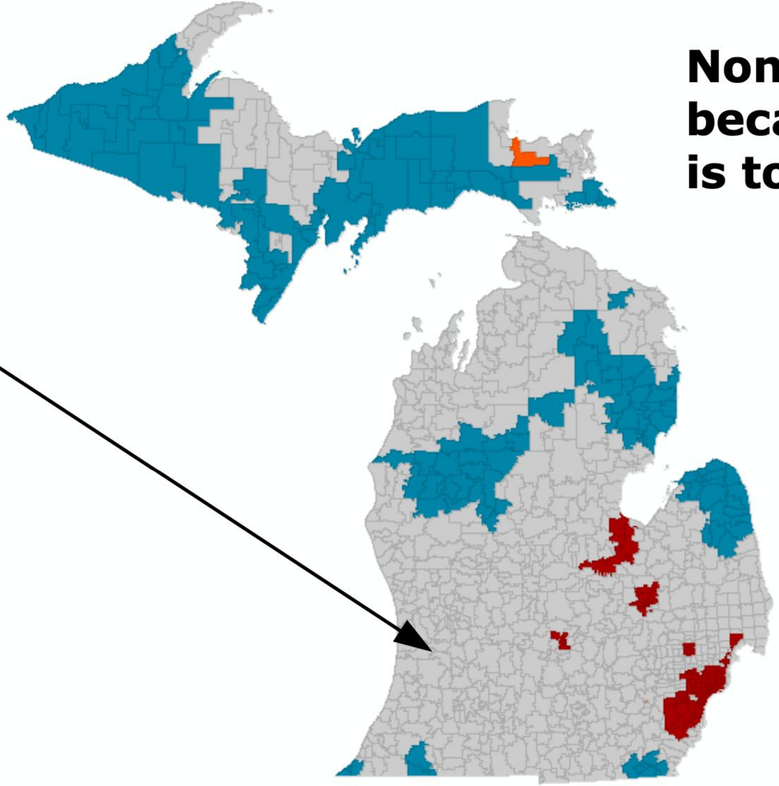
LISA





Low values, with neighbors also having low values

LISA



Non-significant because the p-value is too high (> 0.05)

Spatial Correlograms

- Non-parametric approach to measure Global Spatial Autocorrelation
    - Does not rely on the specification of a spatial weight matrix (parametric approach)
    - Calculates spatial autocorrelation at various distance bands
        - Tells up how spatial autocorrelation varies with distance
            - Generally, spatial autocorrelation decreases with increasing distance (distance decay)

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm#GUID-F063A8F5-9459-42F9-BF41-4E66FBBCC415

Spatial Correlograms

- Non-parametric approach to measure Global Spatial Autocorrelation
    - Does not rely on the specification of a spatial weight matrix (parametric approach)
    - Calculates spatial autocorrelation at various distance bands
        - Tells up how spatial autocorrelation varies with distance
            - Generally, spatial autocorrelation decreases with increasing distance (distance decay)

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm#GUID-F063A8F5-9459-42F9-BF41-4E66FBBCC415

Spatial Correlograms

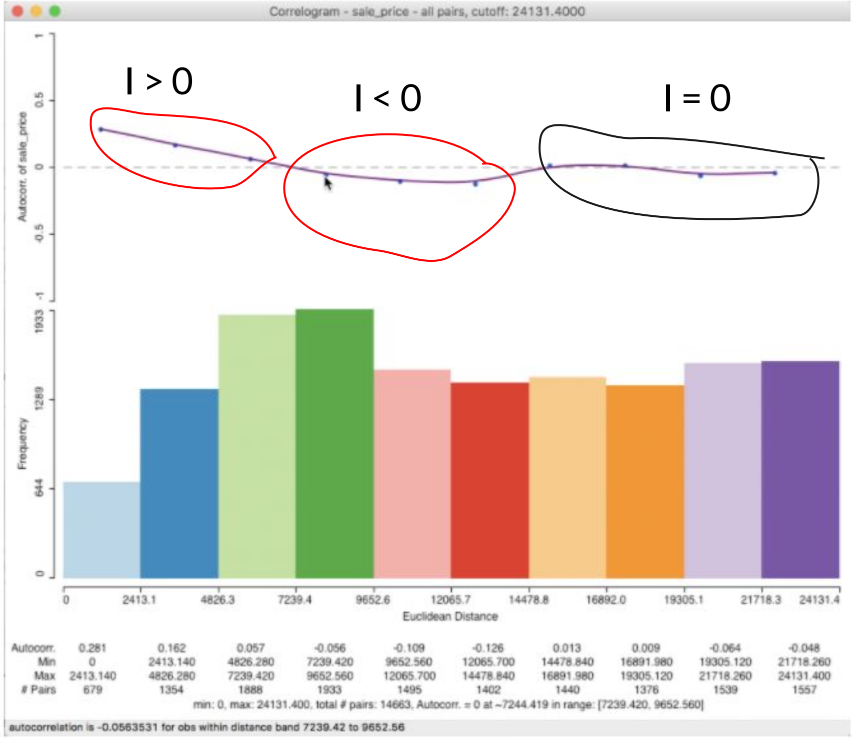


Figure 31: Correlogram with half max distance

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm#GUID-F063A8F5-9459-42F9-BF41-4E66FBBCC415

Extensions

- Bivariate Moran's I
    - Measures the degree to which the value for a given variable at a location is correlated with its neighbors for a *different variable*
        - Does not take into account correlation between variables at the same location
    - Can also extend to time lags
        - Looking at correlation of the value of a variable at one location with values of neighbors for the same variable at a previous time period
        - Differential Moran's I
            - Takes into account time lags but also controls for temporal autocorrelation

Extensions

- Moran's I for rates data
    - Corrects for varying population densities when the variable of interest is a rate or proportion
    - Similar to smoothing
        - Transforms the crude rate to a new variable using empirical bayes techniques (same technique used to smooth data) and then calculate Morans' I
- Moran's I for binary data (0 and 1)
    - You can use join count statistics to calculate spatial autocorrelation for dichotomous data
- Similar extensions to Location Moran's I