Outline of This Week

• Last week, we learned:
  – spatial point pattern analysis (PPA)
  – focus on location distribution of ‘events’

• This week, we will learn:
  – spatial autocorrelation
  – global measures of spatial autocorrelation
  – local measure of spatial autocorrelation
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 autocorrelation is everywhere
  – Spatial point pattern
    • K, F, G functions
    • Kernel functions
  – Areal/lattice (this topic)
  – Geostatistical data (next 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

- 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 or a point

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

1\textsuperscript{st} order
Nearest neighbor
rook

2\textsuperscript{nd} order
Next nearest neighbor
hexagon
queen
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

\[
\text{Haversine formula: } a = \sin^2(\Delta \varphi/2) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2(\Delta \lambda/2) \\
\text{c = 2.atan2(}\sqrt{a}, \sqrt{1-a}) \\
d = \text{R.c}
\]

*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) neighbours within 2,317 m

Source: Bivand and Pebesma and Gomez-Rubio
Neighbor/Connectivity Histogram

Source: Bivand and Pebesma and Gomez-Rubio
Side Note: 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 \( W_{ij} \) 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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4x4 matrix

\[
W = \begin{pmatrix}
A & 0 & 1 & 1 & 0 \\
B & 1 & 0 & 0 & 1 \\
C & 1 & 0 & 0 & 1 \\
D & 0 & 1 & 1 & 0 \\
\end{pmatrix}
\]
<table>
<thead>
<tr>
<th>Name</th>
<th>Fips</th>
<th>Ncount</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>N5</th>
<th>N6</th>
<th>N7</th>
<th>N8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>1</td>
<td>4</td>
<td>28</td>
<td>13</td>
<td>12</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>4</td>
<td>5</td>
<td>35</td>
<td>8</td>
<td>49</td>
<td>6</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>5</td>
<td>6</td>
<td>22</td>
<td>28</td>
<td>48</td>
<td>47</td>
<td>40</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>32</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>8</td>
<td>7</td>
<td>35</td>
<td>4</td>
<td>20</td>
<td>40</td>
<td>31</td>
<td>49</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>9</td>
<td>3</td>
<td>44</td>
<td>36</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>10</td>
<td>3</td>
<td>24</td>
<td>42</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>11</td>
<td>2</td>
<td>51</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>12</td>
<td>2</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>13</td>
<td>5</td>
<td>12</td>
<td>45</td>
<td>37</td>
<td>1</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td>16</td>
<td>6</td>
<td>32</td>
<td>41</td>
<td>56</td>
<td>49</td>
<td>30</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>17</td>
<td>5</td>
<td>29</td>
<td>21</td>
<td>18</td>
<td>55</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>18</td>
<td>4</td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>19</td>
<td>6</td>
<td>29</td>
<td>31</td>
<td>17</td>
<td>55</td>
<td>27</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>20</td>
<td>4</td>
<td>40</td>
<td>29</td>
<td>31</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>21</td>
<td>7</td>
<td>47</td>
<td>29</td>
<td>18</td>
<td>39</td>
<td>54</td>
<td>51</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>22</td>
<td>3</td>
<td>28</td>
<td>48</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>23</td>
<td>1</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>24</td>
<td>5</td>
<td>51</td>
<td>10</td>
<td>54</td>
<td>42</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>25</td>
<td>5</td>
<td>44</td>
<td>9</td>
<td>36</td>
<td>50</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>26</td>
<td>3</td>
<td>18</td>
<td>39</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>27</td>
<td>4</td>
<td>19</td>
<td>55</td>
<td>46</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td>28</td>
<td>4</td>
<td>22</td>
<td>5</td>
<td>1</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td>29</td>
<td>8</td>
<td>5</td>
<td>40</td>
<td>17</td>
<td>21</td>
<td>47</td>
<td>20</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>Montana</td>
<td>30</td>
<td>4</td>
<td>16</td>
<td>56</td>
<td>38</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td>31</td>
<td>6</td>
<td>29</td>
<td>20</td>
<td>8</td>
<td>19</td>
<td>56</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>32</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>49</td>
<td>16</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>33</td>
<td>3</td>
<td>25</td>
<td>23</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>34</td>
<td>3</td>
<td>10</td>
<td>36</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>35</td>
<td>5</td>
<td>48</td>
<td>40</td>
<td>8</td>
<td>4</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>36</td>
<td>5</td>
<td>34</td>
<td>9</td>
<td>42</td>
<td>50</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>37</td>
<td>4</td>
<td>45</td>
<td>13</td>
<td>47</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td>38</td>
<td>3</td>
<td>46</td>
<td>27</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>39</td>
<td>5</td>
<td>26</td>
<td>21</td>
<td>54</td>
<td>42</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>40</td>
<td>6</td>
<td>5</td>
<td>35</td>
<td>48</td>
<td>29</td>
<td>20</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>41</td>
<td>4</td>
<td>6</td>
<td>32</td>
<td>16</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>42</td>
<td>6</td>
<td>24</td>
<td>54</td>
<td>10</td>
<td>39</td>
<td>36</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>44</td>
<td>2</td>
<td>25</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>45</td>
<td>2</td>
<td>13</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>46</td>
<td>6</td>
<td>56</td>
<td>27</td>
<td>19</td>
<td>31</td>
<td>38</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>47</td>
<td>8</td>
<td>5</td>
<td>28</td>
<td>1</td>
<td>37</td>
<td>13</td>
<td>51</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>Texas</td>
<td>48</td>
<td>4</td>
<td>22</td>
<td>5</td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>49</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>35</td>
<td>56</td>
<td>32</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>50</td>
<td>3</td>
<td>36</td>
<td>25</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>51</td>
<td>6</td>
<td>47</td>
<td>37</td>
<td>24</td>
<td>54</td>
<td>11</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>53</td>
<td>2</td>
<td>41</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>54</td>
<td>5</td>
<td>51</td>
<td>21</td>
<td>24</td>
<td>39</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>55</td>
<td>4</td>
<td>26</td>
<td>17</td>
<td>19</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>56</td>
<td>6</td>
<td>49</td>
<td>16</td>
<td>31</td>
<td>8</td>
<td>46</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

Total number of neighbors
--some have more than others

Divide each number by the row sum

Row standardized
--usually use this
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 $w_{ij} = e^{-d^2}$)
Due to the Federal government shutdown, NOAA.gov and most associated web sites are unavailable.

Only web sites necessary to protect lives and property will be maintained.

See [Weather.gov](https://weather.gov) for critical weather information or contact [USA.gov](https://usa.gov) for more information about the shutdown.
Example

• Compare three different weight matrix in images
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 each 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, Geary’s C, Getis-Ord’s G

• Local Measures
  – Local Moran’s I, Geary’s C, Getis-Ord’s G
Join (or Joint or Joins) Count Statistic

- 60 for Rook Case
- 110 for Queen Case
Join Count: Test Statistic

Test Statistic given by:  \( Z = \frac{\text{Observed} - \text{Expected}}{\text{SD of Expected}} \)

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

Expected given by:

\[
\begin{align*}
E(J_{BB}) &= kp_B^2 \\
E(J_{WW}) &= kp_W^2 \\
E(J_{BW}) &= 2kp_Bp_W
\end{align*}
\]

Standard Deviation of Expected (standard error) given by:

\[
\begin{align*}
E(s_{BB}) &= \sqrt{kp_B^2 + 2mp_B^3 - (k + 2m)p_B^4} \\
E(s_{WW}) &= \sqrt{kp_W^2 + 2mp_W^3 - (k + 2m)p_W^4} \\
E(s_{BW}) &= \sqrt{2(k + m)p_Bp_W - 4(k + 2m)p_B^2p_W^2}
\end{align*}
\]

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

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jbb</td>
<td>60</td>
</tr>
<tr>
<td>Jgg</td>
<td>21</td>
</tr>
<tr>
<td>Jbg</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
</tr>
</tbody>
</table>
# Join Count Statistic for Gore/Bush 2000 by State

<table>
<thead>
<tr>
<th>candidates</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bush</td>
<td>0.49885</td>
</tr>
<tr>
<td>Gore</td>
<td>0.50115</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Expected</th>
<th>Stan Dev</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jbb</td>
<td>60</td>
<td>27.125</td>
<td>8.667</td>
<td>3.7930</td>
</tr>
<tr>
<td>Jgg</td>
<td>21</td>
<td>27.375</td>
<td>8.704</td>
<td>-0.7325</td>
</tr>
<tr>
<td>Jbg</td>
<td>28</td>
<td>54.500</td>
<td>5.220</td>
<td>-5.0763</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>109.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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 more Bush/Bush joins (actual = 60) than would be expected (27)
  - Positive autocorrelation
- There are far fewer 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 or 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

\[ I = \frac{N \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \overline{x})(x_j - \overline{x})}{(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}) \sum_{i=1}^{n} (x_i - \overline{x})^2} \]

- \( N \) is the number of observations (points or polygons)
- \( \overline{x} \) 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

\[ E(I) = -\frac{1}{(N-1)} \]

• 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 two different variables

- **Moran’s I [-1, 1]**
  - Spatial autocorrelation and often involves one (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 $X$
\[ \sum_{i=1}^{n} 1(y_i - \bar{y})(x_i - \bar{x})/n \]
\[
\frac{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}}{n \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}}
\]

**Correlation Coefficient**

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)

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

**Spatial auto-correlation**

Source: Ron Briggs of UT Dallas
Correlation Coefficient

\[
\sum_{i=1}^{n} \frac{(y_i - \bar{y})(x_i - \bar{x})}{n}
\]

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

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x}) / \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}
\]

\[
\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \quad \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}
\]

Yi is the Xi for the neighboring polygon

Moran’s I

Source: Ron Briggs of UT Dallas
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
Moran Scatter Plots

We can draw a scatter diagram between these two variables (in standardized form): $X$ and $\text{lag-X}$ (or $W_X$)

The slope of this regression line is Moran’s I
Moran Scatter Plots

QII
Low/High
negative SA

QI
High/High
positive SA

QHI
Low/Low
positive SA

QIV
High/Low
negative SA

Moran's I: 0.57328
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 known as the EB adjustment:
  – see Assuncao-Reis Empirical Bayes Standardization Statistics in Medicine, 1999

• GeoDA software includes an option for this adjustment
Geary’s C

• **Calculation** is similar to Moran’s I,
  – For Moran, the cross-product is based on the deviations from the mean for the two location values
  – For Geary, the cross-product uses the actual values themselves at each location
  – Covariance vs. variogram

\[
C = \frac{N \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2}{2 \left( \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \right) \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]
Geary’s C vs. Moran’s I

- **Interpretation** is very different, essentially the opposite!
  - Geary’s C varies on a scale from 0 to 2
    - 0 indicates perfect **positive** autocorrelation/clumped
    - 1 indicates no autocorrelation/random
    - 2 indicates perfect **negative** autocorrelation/dispersed

- Can convert to a -/+1 scale by: calculating $C^* = 1 - C$

- Morain’s I is more often used
Statistical Significance Tests for Geary’s C

- Similar to Moran
- Again, based on the normal frequency distribution with

\[ Z = \frac{C - E(C)}{S_{error(I)}} \]

Where:  
- \( C \) is the calculated value for Geary’s C from the sample 
- \( E(C) \) is the expected value if no autocorrelation 
- \( S \) is the standard error

however, \( E(C) = 1 \)
Hot Spots and Cold Spots

• What is a *hot spot*?
  – A place where *high* values cluster together

• What is a *cold spot*?
  – A place where *low* values cluster together

• Moran’s I and Geary’s C cannot distinguish them
  • They only indicate *clustering*
  • Cannot tell if these are hot spots, cold spots, or both
Getis-Ord General/Global  G-Statistic

- The G statistic distinguishes between hot spots and cold spots. It identifies *spatial concentrations*.
  - G is relatively large if high values cluster together
  - G is relatively low if low values cluster together
- The General G statistic is interpreted relative to its *expected value*
  - The value for which there is no spatial association
  - G > (larger than) expected value → potential “hot spots”
  - G < (smaller than) expected value → potential “cold spots”
- A *Z test statistic* is used to test if the difference is statistically significant
- Calculation of G based on a *neighborhood distance* within which cluster is expected to occur

The General G statistic of overall spatial association is given as:

\[
G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j}, \quad \forall j \neq i
\]  \hspace{1cm} (1)

where \(x_i\) and \(x_j\) are attribute values for features \(i\) and \(j\), and \(w_{i,j}\) is the spatial weight between feature \(i\) and \(j\). \(n\) is the number of features in the dataset and \(\forall j \neq i\) indicates that features \(i\) and \(j\) cannot be the same feature.

The \(z_G\)-score for the statistic is computed as:

\[
z_G = \frac{G - E[G]}{\sqrt{V[G]}}
\]  \hspace{1cm} (2)

where:

\[
E[G] = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \quad \forall j \neq i
\]  \hspace{1cm} (3)

\[
\]  \hspace{1cm} (4)
Comments on General G

• General G will not show negative spatial autocorrelation

• Should only be calculated for ratio scale data
  – data with a “natural” zero such as crime rates, birth rates

• Although it was defined using a contiguity (0,1) weights matrix, any type of spatial weights matrix can be used
  – ArcGIS gives multiple options

• There are two global versions: G and G*
  – G does not include the value of \( X_i \) itself, only “neighborhood” values
  – G* includes \( X_i \) as well as “neighborhood” values

• Significance test on General G and G* follows the similar procedure as used for Moran’s I
Local Measures of Spatial Autocorrelation
Local Indicators of Spatial Association (LISA)

- Local versions of Moran’s I, Geary’s C, and the Getis-Ord G statistic
- 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:
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”

or it can be in “deviation form”

\[
z_i = \frac{x_i - \bar{x}}{SD_x}
\]

and \( w_{ij} \) is the spatial weight

The summation \( \sum_j \) is across each row \( i \) of the spatial weights matrix.

An example follows
## Contiguity Matrix

<table>
<thead>
<tr>
<th>Code</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
<th>Sum</th>
<th>Neighbors</th>
<th>Illiteracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>6 5 4 3 2</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>7 4 3 1</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>6 2 1</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7 2 1</td>
</tr>
<tr>
<td>Henan</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>6 1</td>
</tr>
<tr>
<td>Hubei</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1 3 5</td>
</tr>
<tr>
<td>Shanghai</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2 4</td>
</tr>
</tbody>
</table>

Source: Ron Briggs of UT Dallas
## Contiguity Matrix and Row Standardized Spatial Weights Matrix

### Contiguity Matrix

<table>
<thead>
<tr>
<th>Code</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Henan</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Hubei</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Shanghai</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

### Row Standardized Spatial Weights Matrix

<table>
<thead>
<tr>
<th>Code</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>1.00</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>2.00</td>
<td>0.00</td>
<td>0.25</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>3.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>4.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>1.00</td>
</tr>
<tr>
<td>Henan</td>
<td>5.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Hubei</td>
<td>6.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Shanghai</td>
<td>7.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Source: Ron Briggs of UT Dallas
Calculating standardized (z) scores

Deviations from Mean and z scores.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>X-Xmean</th>
<th>X-Mean2</th>
<th>z</th>
<th>( \text{z}_i = \frac{x_i - x}{SD_x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>14.49</td>
<td>6.29</td>
<td>39.55</td>
<td>2.101</td>
<td></td>
</tr>
<tr>
<td>Zhejiang</td>
<td>9.36</td>
<td>1.16</td>
<td>1.34</td>
<td>0.387</td>
<td></td>
</tr>
<tr>
<td>Jiangxi</td>
<td>6.49</td>
<td>(1.71)</td>
<td>2.93</td>
<td>(0.572)</td>
<td></td>
</tr>
<tr>
<td>Jiangsu</td>
<td>8.05</td>
<td>(0.15)</td>
<td>0.02</td>
<td>(0.051)</td>
<td></td>
</tr>
<tr>
<td>Henan</td>
<td>7.36</td>
<td>(0.84)</td>
<td>0.71</td>
<td>(0.281)</td>
<td></td>
</tr>
<tr>
<td>Hubei</td>
<td>7.69</td>
<td>(0.51)</td>
<td>0.26</td>
<td>(0.171)</td>
<td></td>
</tr>
<tr>
<td>Shanghai</td>
<td>3.97</td>
<td>(4.23)</td>
<td>17.90</td>
<td>(1.414)</td>
<td></td>
</tr>
</tbody>
</table>

Mean and Standard Deviation

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>57.41</td>
<td>0.00</td>
<td>62.71</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>57.41</td>
<td>/ 7   =</td>
<td>8.20</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>62.71</td>
<td>/ 7   =</td>
<td>8.96</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>( \sqrt{8.96} ) =</td>
<td>2.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Calculating LISA

#### Row Standardized Spatial Weights Matrix

<table>
<thead>
<tr>
<th>Code</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>1</td>
<td>0.00</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>2</td>
<td>0.25</td>
<td>0.00</td>
<td>0.25</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>3</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>4</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Henan</td>
<td>5</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Hubei</td>
<td>6</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Shanghai</td>
<td>7</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Province</th>
<th>Zi</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>2.101</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>0.387</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>(0.572)</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>(0.051)</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Henan</td>
<td>(0.281)</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Hubei</td>
<td>(0.171)</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
<tr>
<td>Shanghai</td>
<td>(1.414)</td>
<td>2.101</td>
<td>0.387</td>
<td>(0.572)</td>
<td>(0.051)</td>
<td>(0.281)</td>
<td>(0.171)</td>
<td>(1.414)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Province</th>
<th>Zi</th>
<th>Anhui</th>
<th>Zhejiang</th>
<th>Jiangxi</th>
<th>Jiangsu</th>
<th>Henan</th>
<th>Hubei</th>
<th>Shanghai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>2.101</td>
<td>-</td>
<td>0.077</td>
<td>(0.114)</td>
<td>(0.010)</td>
<td>(0.056)</td>
<td>(0.034)</td>
<td>-</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>0.387</td>
<td>0.525</td>
<td>-</td>
<td>(0.143)</td>
<td>(0.013)</td>
<td>-</td>
<td>-</td>
<td>(0.353)</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>(0.572)</td>
<td>0.700</td>
<td>0.129</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.057)</td>
<td>-</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>(0.051)</td>
<td>0.700</td>
<td>0.129</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.471)</td>
</tr>
<tr>
<td>Henan</td>
<td>(0.281)</td>
<td>1.050</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.085)</td>
<td>-</td>
</tr>
<tr>
<td>Hubei</td>
<td>(0.171)</td>
<td>0.700</td>
<td>-</td>
<td>(0.191)</td>
<td>-</td>
<td>(0.094)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Shanghai</td>
<td>(1.414)</td>
<td>-</td>
<td>0.194</td>
<td>-</td>
<td>(0.025)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Calculating LISA

\[ I_i = z_i \sum_j w_{ij} z_j \]

<table>
<thead>
<tr>
<th>Province</th>
<th>LISA from GeoDA</th>
<th>LISA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anhui</td>
<td>-0.289</td>
<td>-0.137</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>0.016</td>
<td>0.006</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>-0.442</td>
<td>-0.421</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>-0.018</td>
<td>-0.016</td>
</tr>
<tr>
<td>Henan</td>
<td>0.358</td>
<td>0.271</td>
</tr>
<tr>
<td>Hubei</td>
<td>0.965</td>
<td>0.233</td>
</tr>
<tr>
<td>Shanghai</td>
<td>-0.071</td>
<td>-0.061</td>
</tr>
</tbody>
</table>

Source: Ron Briggs of UT Dallas
Local Getis-Ord G and G\* Statistics

- **Local Getis-Ord G**
  - It is the proportion of all \( x \) values in the study area accounted for by the neighbors of location \( I \).
  - \( G^* \) will include the self value.

\[
G_i(d) = \frac{\sum_j w_{ij}x_j}{\sum_j x_j}
\]

\( G \) will be **high** where **high** values cluster
\( G \) will be **low** where **low** values cluster
Interpreted relative to expected value if randomly distributed.

\[
E(G_i(d)) = \frac{\sum_j w_{ij}(d)}{n-1}
\]
Bivariate LISA

- Moran’s I is the correlation between X and Lag-X—the same variable but in nearby areas
  - Univariate Moran’s I
- Bivariate Moran’s I is a correlation between X and a different variable in nearby areas.

Moran Significance Map for GDI vs AL

Moran Scatter Plot for GDI vs AL
Bivariate LISA and the Correlation Coefficient

- Correlation Coefficient is the relationship between two different variables in the same area.
- Bivariate LISA is a correlation between two different variables in an area and in nearby areas.
Bivariate Moran Scatter Plot

Low/High negative SA

High/High positive SA

Low/Low positive SA

High/Low negative SA
Summary

- Spatial autocorrelation of areal data
- Spatial weight matrix
- Measures of spatial autocorrelation
  - Global Measure
    - Moran’s I/Geary’s C/General G and G*
  - Local
    - LISA: Moran’s I/Geary’s C/General G and G*
    - Bivariate LISA
    - Significance test
• End of this topic