Mathematics and Geography,

Using Linear Algebra to Determine Spatial Autocorrelation.

Pat McGrath

Objectives

- The First Law of Geography states, "*Everything is related to everything else, but near things are more related to each other*." –Waldo Tobler.

- For this project, I am going to show a method to determine spatial correlation. To do this I will use Moran's Index to find the index values between Salt Lake and eight other cities in Utah.

- Then, I am going to graph the index values versus distance from Salt Lake City to determine if the first law of geography holds true.

- Spatial autocorrelation is a statistical method to determine how related locations are to each other.

- Moran’s Index is an equation to determine the spatial autocorrelation between two locations. It is different than the covariance because it takes into account the space, or distance between two subjects.

- Here is Moran’s Index, developed by Australian statistician Patrick Moran in the early 1950’s.

$$\text{Moran's Index } I = \frac{N}{\sum_i \sum_j W_{ij}} \cdot \frac{\sum_i \sum_j W_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2}$$

- The Moran’s Index values range from -1 to 1. The value 1 indicates perfect correlation, or location j is of interest to location i. Zero indicates random correlation, and -1 indicates perfect dispersion. Both zero and -1 imply location J not of interest to Location i.

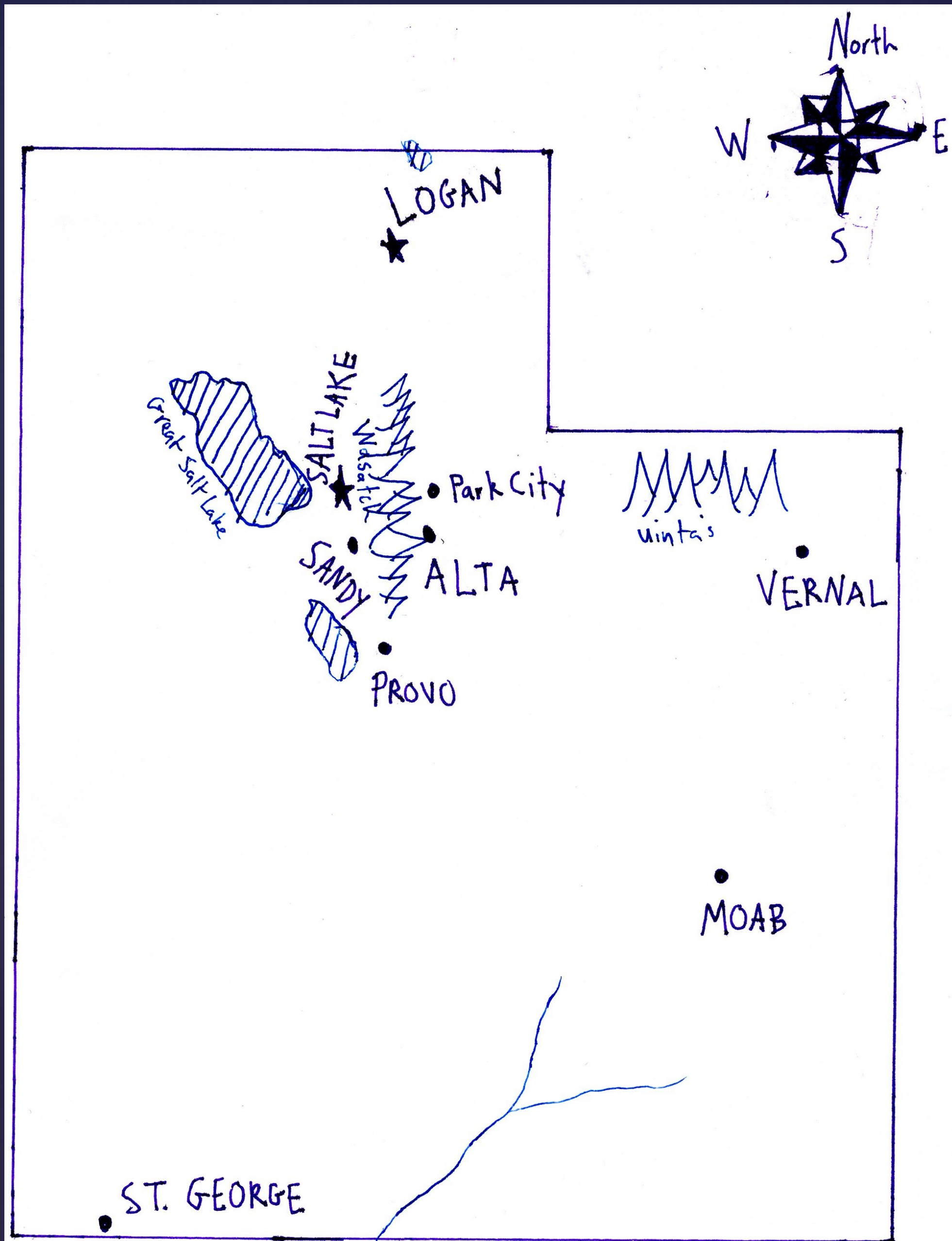
Method

- 1st each city is going to have an (x,y) pair. The x will be the location, and y will be the variables we are testing.
- The x will be a 3x1 matrix consisting of longitude, latitude, and altitude.
- The y will also be 3x1 matrix consisting of temperature, annual snowfall, and base depth. All of the y variables in this example are of significant interest to skiers and snowboarders.
- Here is what each locations coordinate pair will look like,

$$\text{Location } i = (x, y) = \left(\begin{pmatrix} \text{longitude}, x \\ \text{latitude}, y \\ \text{Altitude}, z \end{pmatrix}, \begin{pmatrix} \text{Temperature} \\ \text{Annual Snowfall} \\ \text{Base Depth} \end{pmatrix} \right)$$

y i_y Vector

Location



#	city	x coordinate (West) x longetide	Y coord N Y latitude	z elevation
1	Logan	111°49'51"	41°44'16"	4534'
2	Saltlake	111°53'0"	40°45'0"	4226'
3	Sandy	111°51'35"	40°34'21"	4450'
4	Alta	111°38'14"	40°34'51"	8530'
5	Park City	111°29'59"	40°39'34"	7000'
6	Provo	111°39'39"	40°14' [illegible]"	4551'
7	Vernal	109°32'8"	40°27'17"	5328'
8	Moab	109°32'59"	38°34'21"	4026'
9	St. George	113°34'41"	37°5'42"	2860'

Above is the x, y, and z coordinates for each city. It will be used to calculate distance.

- Moran's Index;

$$\text{Moran's Index } I = \frac{N}{\sum_i \sum_j W_{ij}} \cdot \frac{\sum_i \sum_j W_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2}$$

- i and j subscripts indicate the two separate cities, i and j. Y bar is the average temperature, annual snowfall, and base depth of the two cities. Here it is with an example.

$$\bar{Y} = \frac{(Y_i + Y_j)}{n}$$

$$\bar{Y} = \left[\begin{pmatrix} 40 \\ 392 \\ 0 \end{pmatrix} + \begin{pmatrix} 44 \\ 301 \\ 0 \end{pmatrix} \right] \cdot \frac{1}{2} = \boxed{\begin{pmatrix} 42 \\ 346.5 \\ 0 \end{pmatrix} = \bar{Y}} \text{ in this case}$$

- W_{ij} is the weighted matrix. It is the matrix with every locations distance from all of the other locations. In this example w_{ij} is 9x9 because of the 9 cities.

$$W_{ij} = \begin{array}{c} \text{cities} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{pmatrix} 0 & & & & & & & & \\ & 0 & & & & & & & \\ & & 0 & & & & & & \\ & & & 0 & & & & & \\ & & & & 0 & & & & \\ & & & & & 0 & & & \\ & & & & & & 0 & & \\ & & & & & & & 0 & \\ & & & & & & & & 0 \end{pmatrix} \end{array} = \text{Distance Matrix}$$

construct the W_{ij} matrix, for that purple circle, $W_{(2,8)}$

location 2 = Salt Lake City $= \begin{pmatrix} x \\ y \\ z \end{pmatrix}$
$x = 111° 53' 0''$ = longitude (West)
$y = 40° 45' 0''$ = latitude (North)
$z = 4226'$ = altitude

1^{st}. longitude and latitude are in degrees, minutes, seconds. Convert to decimal degre

$$\underset{[X]}{(111° \;\; 53' \;\; 0'')} \cdot \underset{[A]}{\begin{pmatrix} 1 \\ 1/60 \\ 1/3600 \end{pmatrix}} = 111° \cdot 1 + 53 \cdot \left(\tfrac{1}{60}\right) + 0 \cdot \left(\tfrac{1}{3600}\right) = \underline{111.8833°}$$

- Also multiply $Y \cdot A = \underline{40.75°}$
- $Z = Z = 4226'$

Repeat for city 8 = Moab. $x = 109.54°$
$y = 38.57°$
$z = 4026'$

- New matrix for SLC $= [P] = \begin{pmatrix} 111.8833° \\ 40.75° \\ 4226' \end{pmatrix}$ for Moab $= [Q] = \begin{pmatrix} 109.54° \\ 38.57° \\ 4026' \end{pmatrix}$

- Change in longitude = T, change in latitude = U, change in elevation = V

$T = \Delta X \cdot \cos\left(\frac{y_1 + y_2}{2}\right)$

$U = \Delta y$

$V = \Delta Z$

- $\begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = [P] - [Q] = [PQ] = \begin{pmatrix} 2.3433° \\ 2.18° \\ 200' \end{pmatrix}$ = Change Matrix. But I need to multiply $\Delta x \cdot \cos\left(\frac{y_1+y_2}{2}\right)$ to account for the way longitudes distance changes north of the equator.

$$\begin{pmatrix} \cos(y_1 + y_2/2) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot [PQ] = \begin{pmatrix} 1.8039 \\ 2.18 \\ 200' \end{pmatrix}$$

$95.821 + \overset{150.42}{246.24} \; (.03787)$

- Now, I want to turn degrees to feet. I will use following facts.
  - 60 nautical miles in 1 degree of latitude.
  - 1.15 statute miles in 1 nautical mile.
  - $\therefore$ $\underline{1° \text{ of latitude} = 69 \text{ statute miles}}$ $Y \cdot 69$

- for longitude (X) $= \cos(\text{lat}_1 + \text{lat}_2/2) \cdot 69$ miles. $= x$

- for altitude feet $\cdot \frac{1}{5280}$ = miles.

So now

$$\begin{pmatrix} \overset{\text{original } y\text{'s}}{\cos(y_1+y_2/2) \cdot 69} & 0 & 0 \\ 0 & 69 & 0 \\ 0 & 0 & 1/5280 \end{pmatrix} \begin{pmatrix} 1.8039 \\ 2.18 \\ 200 \end{pmatrix} = \begin{pmatrix} 95.821 \text{ miles} \\ 150.42 \text{ miles} \\ .03787 \text{ miles} \end{pmatrix} \begin{matrix} x_\Delta \\ y_\Delta \\ z_\Delta \end{matrix}$$

Finally, use distance formula

$d = \sqrt{x_\Delta^2 + y_\Delta^2 + z_\Delta^2} = 178$ miles between SLC + M

That Answer is as the crow flies, it is also the way to calculate Great Circle Path between 2 points.

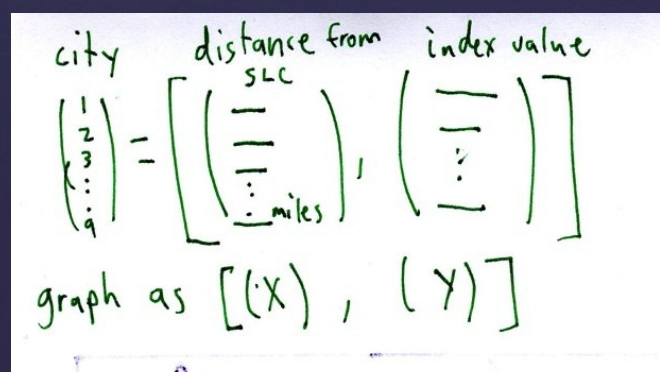
- Because this is distance between Salt Lake (2) and MOAB (8) it goes into matrix W at $W_{2,8}$ and $W_{8,2}$.

Repeat for all cities.

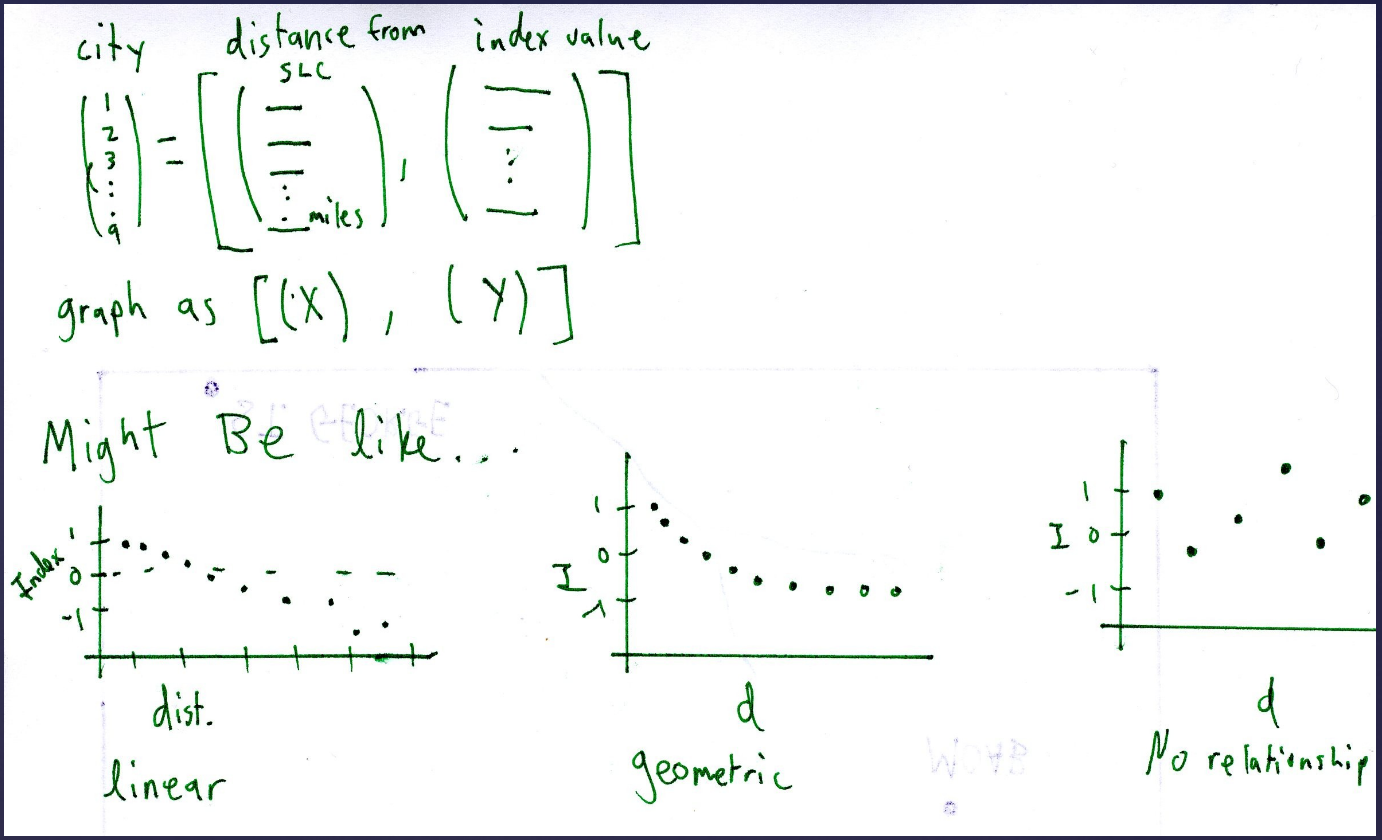
- With the W_{ij} matrix we have everything we need to calculate the correlation value in Moran's Index.

- Now compute each index value with Salt Lake as city i and the remaining seven cities eight cities as location j.

- Then, we have each cities distance from Salt Lake, and its associated Index value. I am going graph each pair, with distance from SLC as the x value, and its index value as the y.



Here is how the graphs might look



- To test whether the first law of geography holds, I am going to do a least squares approximation to determine the relationship between distance and index value. I am assuming it is linear for the sake of simplifying the least squares but it might not be linear.
- Each city will use its coordinate pair used to graph it in the last slide, and then here is the rest

city 1 = [X(distance from SLC), y(Index value)]

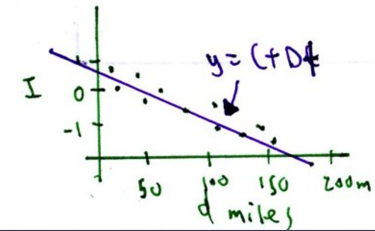
Solve $A^TA\hat{x} = A^Tb$

$x = (A^TA)^{-1}A^Tb$

Matrix $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_9 \end{bmatrix}$ $B = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{pmatrix}$

- if linear, which I am assuming $\hat{x} = \begin{pmatrix} C \\ D \end{pmatrix}$
- Solve $\hat{x} = (A^TA)^{-1}A^TB$

line of Best fit = $\boxed{y = C + Dt}$



- If cities closer to Salt Lake have a score close to 1, and more distant cities have lower index values, then the first law of geography holds when testing temperature, annual snowfall, and base depth. The closer the locations, the more related.

- For skiers and riders in this example, if you know your location you can find your distance to Salt Lake, and determine an estimate for the index score. But for skiers and riders you would want your index score based on the relationship to Alta, not Salt Lake.

accurate way to find distance between locations, and most importantly you can test multiple variables at the same time. You can test all the factors that go into good snow conditions, instead of just how much snow.

- Other examples where matrices would be useful include,
- Farmers interested in not just soil ph, but also acidity in the water, and soil depth.
- Economist wanting to find correlation between cities with the variables being income, years of education, and miles of roads.
- The government interested in the correlation of two cities %Caucasians, %pacific islanders, %Latinos, etc. to study demographics.
- Matrix operations are great for handling all of the different information that any group needs to include to make results

The Finale

- Note, it is not usually necessary to graph distance versus index value. I simply did it in this project because I was interested in testing the first Law of Geography, and it led to including more linear algebra in the project.

- But, I believe this process is a great way to test the first law. The first law hardly seems scientific at all, it is very wishy washy compared to the laws of planetary motion or something similar. So this process adds a little bit of concreteness to the law, one can calculate *how* related different locations are, as opposed to just saying the closer the more related. You could even find a distance threshold, where past a value c, there is no longer a correlation.

- Thank You very much for reading!