GEOG 4140/6140 – Winter 2021
Lab 5: Spatial Descriptive Statistics and Spatial Regression
Due Thursday, March 4th by 11:59 PM

Overview
Spatial regression analysis allows for a quantitative approach to understanding correlations between different spatially distributed phenomena. While it may often be possible to visually see and qualitatively describe the ways in which phenomena relate to each other, there are some scenarios where a statistical approach is more appropriate. In this lab, you will analyze the distribution of sightings of Sasquatch, one of North America's largest mammals, in the state of Washington. You will use multiple spatial regression techniques to better understand the distribution of this elusive species in relation to human activity and the natural environment.

The following resource may be of use in completing this assignment:
An overview of the Measuring Geographic Distributions toolset
How Mean Center works
Mean Center
How Median Center works
Median Center
Standard Distance
Directional Distribution (Standard Deviational Ellipse)
Spatial Join
Regression analysis basics
How OLS regression works
Ordinary Least Squares (OLS)
How Geographically Weighted Regression (GWR) works
Geographically Weighted Regression (GWR)

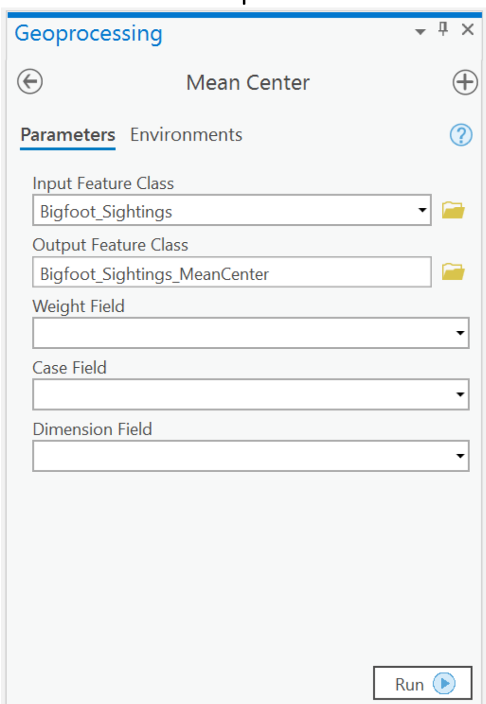
Required Data (*Sources*)
1. Bigfoot Sightings feature class (*Lab data on Canvas*)
2. Washington Counties feature class (*Lab data on Canvas*)
3. Forested_Area_GEOG4140_Lab05.xlsx (*Lab data on Canvas*)
4. Total_Population_GEOG4140_Lab05.csv (*Lab data on Canvas*)

Workflow
1. Open ArcGIS Pro and create a new project titled *Lab05YourLastName* in your working directory for this lab.
    a. Load the reported Sasquatch sightings and Washington counties feature classes into your map and export the two files into your Lab 05 geodatabase.

2. Sasquatch live in densely forested areas so sighting frequency will likely be well predicted by the percentage of the county that is forested.
    a. Add the *Forested Area* csv file from Canvas to your Map, and **join** this data to the counties layer.

3. Despite their general behavior of avoiding humans, some researchers believe that Sasquatch are drawn to human population centers as it gives them some feeling of companionship to be near other bipeds. Due to this suspected behavior, we will look at the relationship between population density and Sasquatch sightings in each county.
    a. Add the *Total Population* csv file containing population data for Washington counties to your map. **Join** this data to your counties layer.

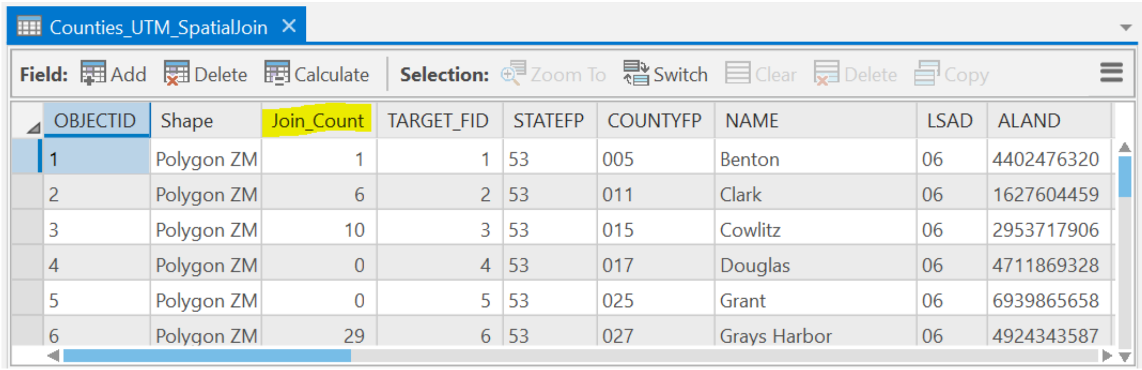
4. Next, calculate descriptive spatial statistics for the spatial distribution of the Sasquatch sightings.
    a. Calculate the **Mean Center**, **Median Center**, **Standard Distance**, and **Standard Deviational Ellipse** of the dataset.
        i. Leave the optional fields (Weights, etc.) empty when running these calculations. Example for the Mean Center tool:



    b. A WARNING will arise when running these calculations. Use the information contained in the warning description, and your knowledge of GIS, to fix the problem and continue working. (Hint: Use UTM Zone 10N for the state of Washington.)

5. Create one layout map showing these four spatial statistics, as well as the Bigfoot Sightings point feature used in the calculations. Be sure to include required map elements and the county feature as a reference. (Deliverable 1)

6. Now that we have a basic idea of the overall spatial distribution of Sasquatch sightings, we will conduct a quantitative analysis of how this distribution might be explained by *population density* and *forested area*. To do this, we will use Ordinary Least Squares (OLS) analysis and Geographically Weighted Regression (GWR).
    a. Before we run the analysis, we need to create a **Spatial Join** to obtain the number of Bigfoot sightings for each county. After running the spatial join, the layer's new attribute table should contain a new field called "Join_Count" as shown below:



OBJECTID	Shape	Join_Count	TARGET_FID	STATEFP	COUNTYFP	NAME	LSAD	ALAND
1	Polygon ZM	1	1	53	005	Benton	06	4402476320
2	Polygon ZM	6	2	53	011	Clark	06	1627604459
3	Polygon ZM	10	3	53	015	Cowlitz	06	2953717906
4	Polygon ZM	0	4	53	017	Douglas	06	4711869328
5	Polygon ZM	0	5	53	025	Grant	06	6939865658
6	Polygon ZM	29	6	53	027	Grays Harbor	06	4924343587

    b. Next, normalize the *number of sightings* and *population* fields by county area, so that the results are not affected by the size of each county.
        i. First, run the **Add Geometry Attributes** tool (select Area with the Area Unit being Square Miles – United States).
        ii. In the attribute table, create two new fields ("Sight_Normalized_by_POLY_AREA" and "Pop_Normalized_by_POLY_AREA"). Set both fields as a Double type.
        iii. Using the **Calculate Field** tool, divide the sightings ("Join_Count" field) by POLY_AREA to populate the "Sight_Normalized_by_POLY_AREA" field. Then, follow the same pattern to populate "Pop_Normalized_by_POLY_AREA" using the Population field.

7. Use the **Ordinary Least Squares (OLS)** analysis tool three times with GEOID as the Unique ID Field and your normalized sightings field as the dependent variable. Be sure to direct the output report file to save in your Lab 5 project folder for each run, as we will use these later.
    a. For the first run, use *forested area* as the explanatory variable.
    b. For the second, use *population density* as the explanatory variable. (Use the normalized population field.)
    c. For the third use, select both fields as explanatory variables.

8. Next, use the **Geographically Weighted Regression (GWR)** tool with both explanatory variables (population density and forested area) as you did at the end of step 7 with the OLS tool. GWR only needs to be run once.
    a. Optional fields can be left unchanged. Set Neighborhood Type to Number of Neighbors and Neighborhood Selection Method to Golden Search.

9. Finally, create a map with two panels (two map frames), one showing the results of your multivariate OLS analysis and the other showing the results of your GWR analysis. (Deliverable 4)

<u>Deliverables</u>

1. A map showing the spatial statistics results of step four, including the original sightings feature used to create those layers. (8 points)

2. In the regression analyses using both explanatory variables, why do you think that prediction accuracy in King County was so low? (2 points)

3. What percentage of the distribution of Sasquatch sightings were explained in your OLS analysis by each variable and by both variables used together (see reports generated in Step 7). Share a few thoughts on why this might be the case. (4 points)

4. A two-panel map, one showing the results of your final OLS analysis and the other showing the results of your GWR analysis. (4 points)

5. Compare the results obtained from the multivariate OLS and GWR analyses. Why might differences exist between these two analyses? (2 points)