EXERCISE 2
Spatial Statistics: Spatial Pattern Analysis

Introduction

This exercise introduces methods to evaluate and statistically validate features such as identifying statistically significant clustering in the spatial distribution of your data; identifying spatial clustering/dispersion for a set of geographic features over a range of distances; and measuring clustering or dispersion of features based on feature locations and attribute values.

Objectives

- Learn to use the Analyzing Patterns Toolset
- Understand statistical significance and confidence intervals
- Exploring several methods to evaluate and statistically validate if features, or the values associated with features, form clustered, dispersed, or random spatial patterns

Prerequisites

- **ArcGIS Pro installed on your machine**

USDA Non-Discrimination Statement

In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, gender identity (including gender expression), sexual orientation, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotape, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at How to File a Program Discrimination Complaint and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by: (1) mail: U.S. Department of Agriculture, Office of the Assistant Secretary for Civil Rights, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410; (2) fax: (202) 690-7442; or (3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

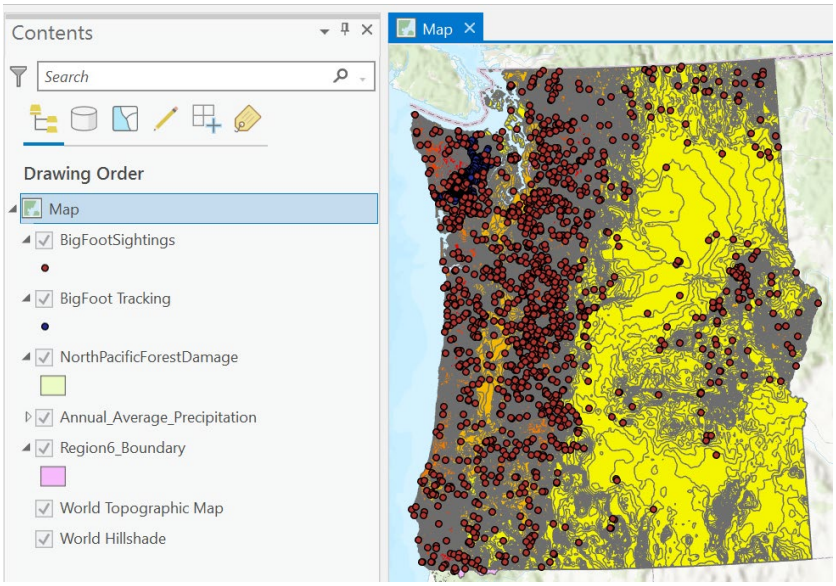
Table of Contents

Part 1: Explore the Project Data 4
A. Open the Exercise 2 project file 4
Part 2: Identify Spatial Patterns that Exist in the Data 4
A. Open the analyzing patterns toolset 4
B. Execute the Average Nearest Neighbor tool to substantiate clustering or dispersion of BigFoot Sightings. 5
C. Use the High/Low Clustering (Getis-Ord General G*) tool to verify clustering of BigFoot Sightings increases in areas with more rainfall. 7
D. Use the Spatial Autocorrelation (Moran's I) tool to verify correlation between BigFoot Sightings and areas that are known to have black bears. 11

Part 1: Explore the Project Data

A. Open the Exercise 2 project file

1. Open Windows Explorer and navigate to your course folder **….\ArcGISProSpatialStatistics\Data\Exercise_2**
   i. Open the **Exercise_2** ArcGIS Project File.
2. Inspect the BigFootSightings points in the map for patterns.



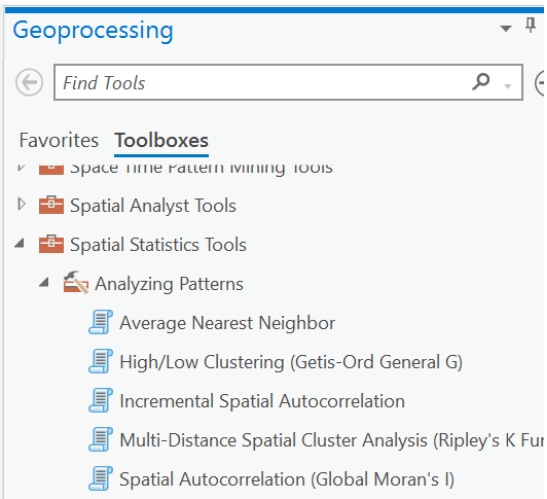
Do any specific spatial patterns such as clustering or dispersion exist in the Bigfoot sighting points? If so, does the clustering/dispersion have statistical significance? We can answer these questions with tools from the Analyzing Patterns toolset.

Part 2: Identify Spatial Patterns that Exist in the Data

The Analyzing Patterns toolset provides tools to help identify overall patterns in the data and identify statistically significant clustering or dispersion.

A. Open the analyzing patterns toolset

1. In the Geoprocessing pane, click the **Toolboxes tab**.
2. Expand the **Spatial Statistics** toolbox then expand the **Analyzing Patterns** toolset.



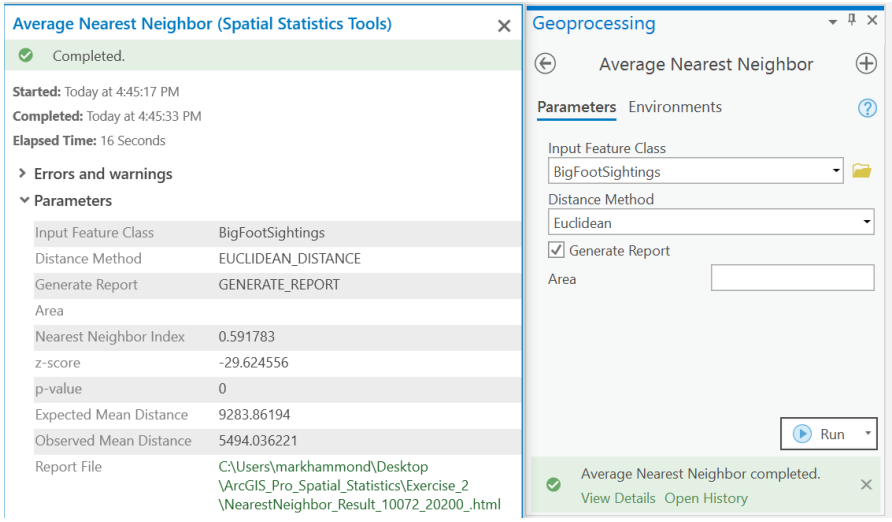
The tools in the Analyzing Patterns toolset are based on inferential statistics and start with the null hypothesis that your features, or the values associated with your features, exhibit a spatially random pattern. They then compute a p-value identifying the probability of whether the null hypothesis is correct. Calculating a probability is important when you need a high level of confidence in a particular decision.

The Analyzing Patterns tools answer questions such as, "Are the features in the dataset, or the values associated with the features in the dataset, spatially clustered?" and "Is the clustering becoming more or less intense over time?"

B. Execute the Average Nearest Neighbor tool to substantiate clustering or dispersion of BigFoot Sightings.

The Average Nearest Neighbor tool identifies statistically significant clustering or dispersion by calculating a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature.

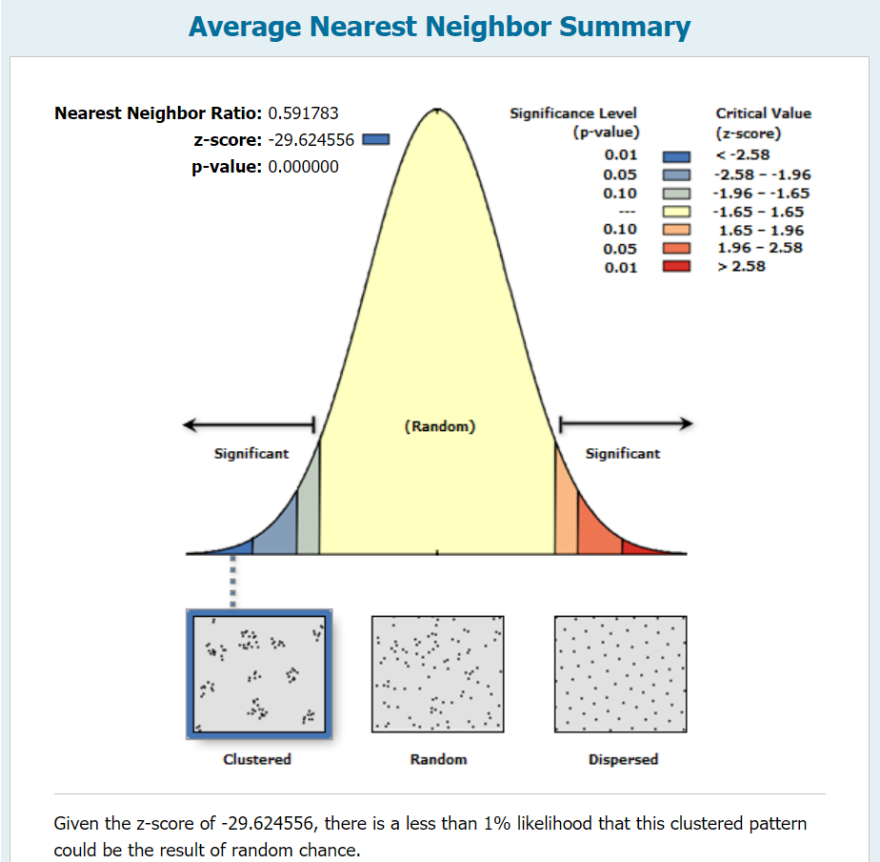
1. From the *Analyzing Patterns toolset*, click **Average Nearest Neighbor** to open the tool.
   i. Input Feature Class: **BigFootSightings**.
   ii. Distance Method: **Euclidian**. This is a straight-line distance. Manhattan takes into consideration buildings or other obstacles
   iii. Generate Report: **checked**.
   iv. Click **Run**.
   v. At the bottom of the tool window, click **View Details** and notice five values in the window that appears: ***Nearest Neighbor Index, z-score, p-value, Expected Mean Distance, and Observed Mean Distance.***



STATISTICS 101: The Z-score and P-value results are measures of statistical significance which tell you whether or not to reject the null hypothesis. For the Average Nearest Neighbor statistic, the null hypothesis states that features are randomly distributed.

The Observed Mean Distance is the actual distance between each point and the closest neighboring point and the Expected Mean Distance is the average distance between neighbors in a hypothetical random distribution. The Nearest Neighbor Index is expressed as the ratio of the Observed Mean Distance to the Expected Mean Distance. If the index is less than 1, the pattern exhibits clustering; if the index is greater than 1, the trend leans toward dispersion.

vi. Click the **Report File** link near the bottom of the View Results window.



INTERPRETATION: the report states very clearly that: Given the z-score of –29.62, there is less than 1% likelihood that this clustered pattern could be the result of random chance. We are 99% confident that the clustering in this dataset is "NOT" random.

2. Close the report.

C. Use the High/Low Clustering (Getis-Ord General G*) tool to verify clustering of BigFoot Sightings increases in areas with more rainfall.

Now that we have established a definitive clustering of Bigfoot sightings; let's determine if clustering also exist in sightings relative to the amount of annual rainfall for that region.

1. Analyze the **Annual_Average_Precipitation** layer in relation to the **BigFootSightings** points in the map. The points appear to be clustered in areas that have greater rainfall averages. We can use the High/Low Clustering (Getis-Ord General G*) tool to verify this belief.

Prior to performing the analysis, we must prepare the data by joining the attributes of the Annual_Average_Precipitation layer to the BigFootSightings layer based on their spatial relationship. The target features (BigFootSightings) and the attributes from the join features (Annual_Average_Precipitation) will be written to a new output feature class.

2. In the Contents pane, right-click **BigFootSightings** then click **Joins and Relates** and select **Spatial Join**.
3. In the Spatial Join window set the following parameters:

i. Target Features: **BigFootSightings**

ii. Join Features: **Annual_Average_Precipitation**

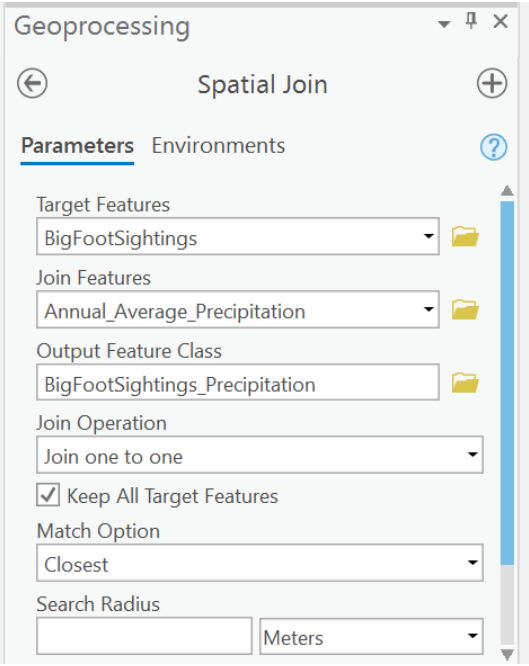
iii. Output Feature Class: **BigFootSightings_Precipitation**

iv. Join Operation: **Join one to one**

v. Keep All Target Features (**checked**)

vi. Match Option: **Closest**

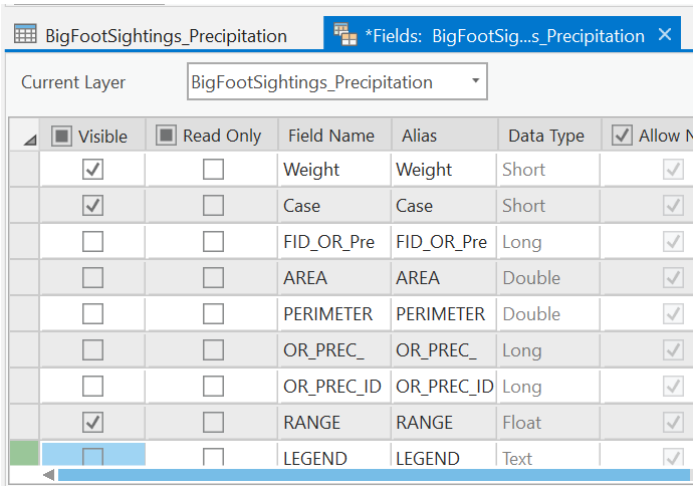
vii. Click **OK.**



4. *Open the attribute table* for the new **BigFootSightings_Precipitation** layer. Scroll through the layer's attributes paying close attention to the new attribute columns. We are interested in the **Range** field which displays the average amount of annual precipitation in inches.

   i. To make the table easier to manage, turn off the other precipitation fields. One way to do this is go to the contextual Table **View** tab and click the **Fields** button (**Field** group).

   ii. In the Fields table, uncheck all the precipitation fields in the **Visible** column except **Range**.



BigFootSightings_Precipitation

Field: Selection:

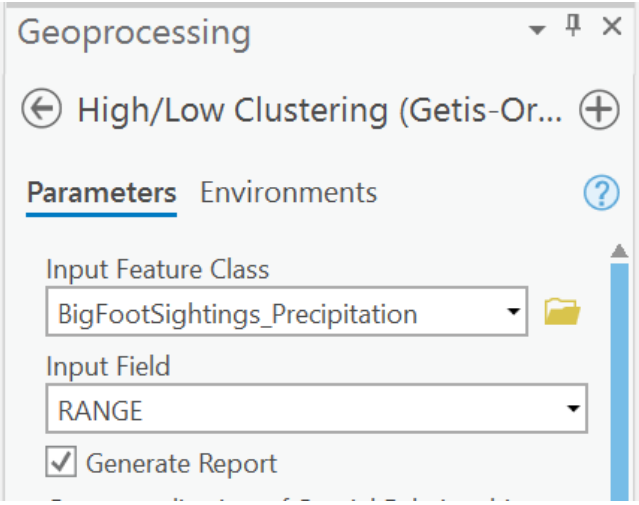
ET_FID	Class	Year	Number	Date	Latitude	Longitude	Weight	Case	RANGE
1	B	2008	1	1/1/2008	-121.6885	45.23679	2	6	55
2	B	2008	1	1/1/2008	-118.2287	48.7375	2	6	23
3	B	2008	1	1/1/2008	-123.3687	46.99385	2	6	69
4	B	2008	1	1/1/2008	-122.0428	47.49865	2	6	71
5	B	2008	1	1/1/2008	-121.8429	47.07024	2	6	89
6	B	2008	1	1/1/2008	-121.796	45.57957	2	6	105
7	B	2008	1	1/1/2008	-121.6423	44.05426	2	6	59

Now that we have added precipitation values to the BigFoot Sightings layer, we can begin our analysis.

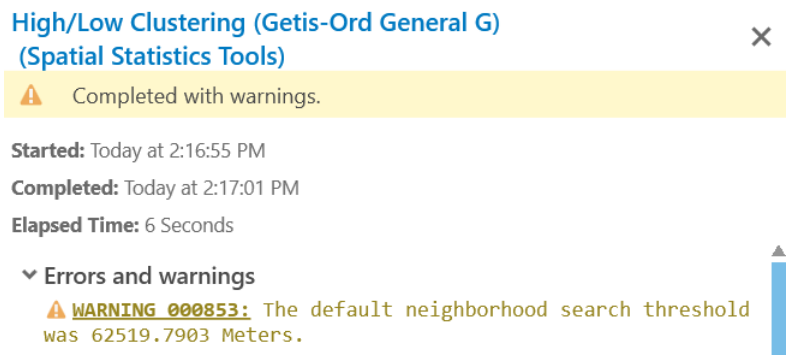
5. From the *Analyzing Patterns* toolset, click **High/Low Clustering (Getis-Ord General G*)** to open the tool.

The High/Low Clustering tool measures the concentration of high or low values for a given study area. The null hypothesis for the High/Low Clustering (General G) statistic states there is no spatial clustering of feature values.

6. In the High/Low Clustering tool window set the following parameters:
   i. Input Feature Class: **BigFootSightings_Precipitation**.
   ii. Input Field: **Range**.
   iii. Generate Report: **checked**.
   iv. Accept the other defaults and click **Run.**

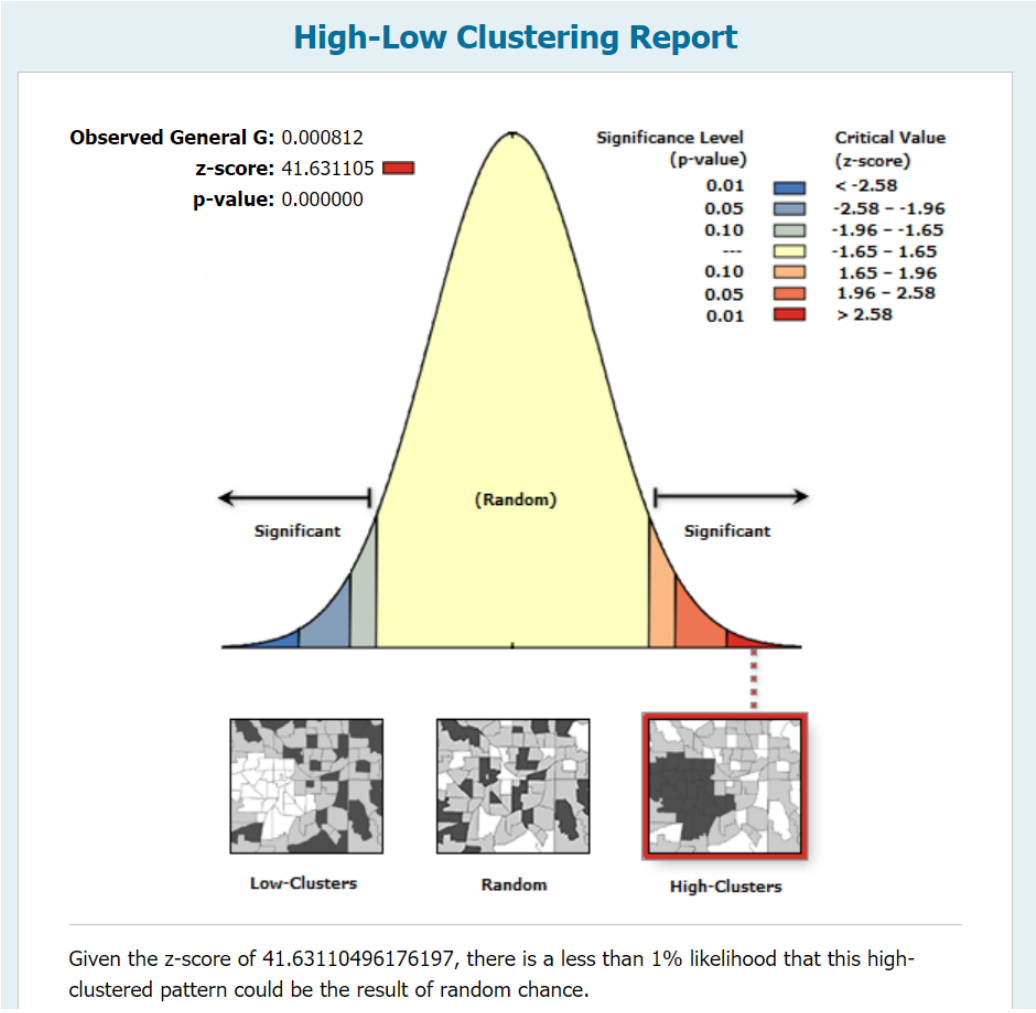


   v. Click **View Details** and notice the warning message that appears**.**



What does that warning message mean? This is a warning provided to notify you that a distance threshold was not used. When no value is specified for the distance threshold, the minimum distance to ensure every feature has at least one neighbor will be used.

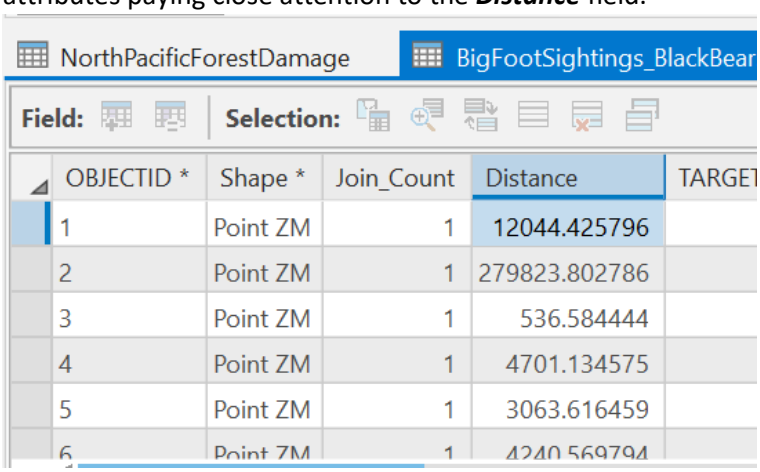
vi. Click the **Report File** link.



INTERPRETATION: *Our data is clustered indicating a correlation between BigFoot sightings and "high" rainfall. When the p-value returned by this tool is small (statistically significant), the null hypothesis can be rejected. If the null hypothesis is rejected, then the sign of the z-score becomes important. If the z-score is positive, the observed index is larger than the expected, indicating high attributes are clustered in the study area. If the z-score is negative, the observed index is smaller than the expected, indicating low attribute values are clustered in the study area. With these results, one might suggest, BigFoot sightings were "mistakenly" identified because of the rain or hazy view of the animal. Could they have been a bear standing on its hind legs? We will need to complete further analysis to decide.*

D. Use the Spatial Autocorrelation (Moran's I) tool to verify correlation between BigFoot Sightings and areas that are known to have black bears.

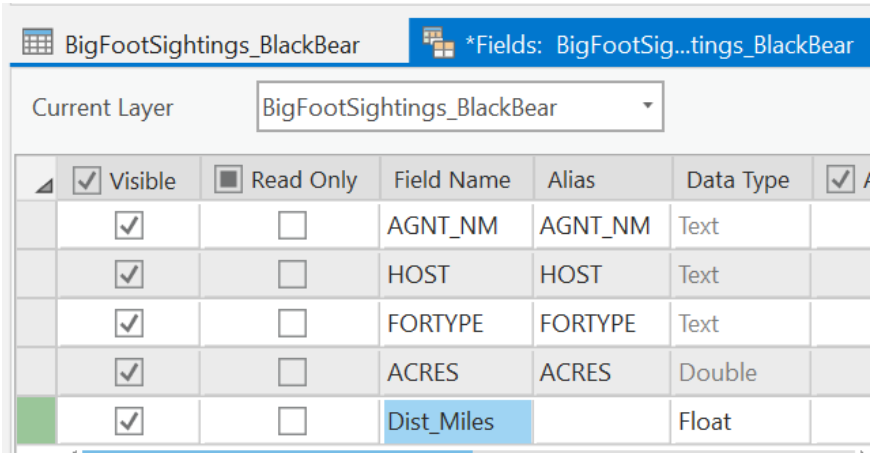
1. Examine the **NorthPacificForestDamage** layer. We need to repeat the steps from before to spatially join the attributes of the NorthPacificForestDamage layer to the BigFoot sightings layer.
2. In the Contents pane, right-click **BigFootSightings** then click **Joins and Relates** and select **Spatial Join**.
3. In the Spatial Join window set the following parameters:
   i. Target Features: **BigFootSightings**
   ii. Join Features: **NorthPacificForestDamage**
   iii. Output Feature Class: **BigFootSightings_BlackBear**
   iv. Join Operation: **Join one to one**
   v. Keep All Target Features (**checked**)
   vi. Match Option: **Closest**
   vii. Distance Field Name: **Distance**
   viii. Click **OK**
4. Open the attribute table for the new **BigFootSightings_BlackBear** layer and scroll through the attributes paying close attention to the ***Distance*** field.



OBJECTID *	Shape *	Join_Count	Distance	TARGET
1	Point ZM	1	12044.425796	
2	Point ZM	1	279823.802786	
3	Point ZM	1	536.584444	
4	Point ZM	1	4701.134575	
5	Point ZM	1	3063.616459	
6	Point ZM	1	4240.569794	

The Distance field (created during the join) gives the distance in meters between the target feature (BigFoot Sightings) and the closest join feature (Forest Damage). Because Black Bears typically travel up to 100 miles away from their home range, we only want to include points that are within that distance. To make this easier, we will create a new field to convert the distance from meters into miles.

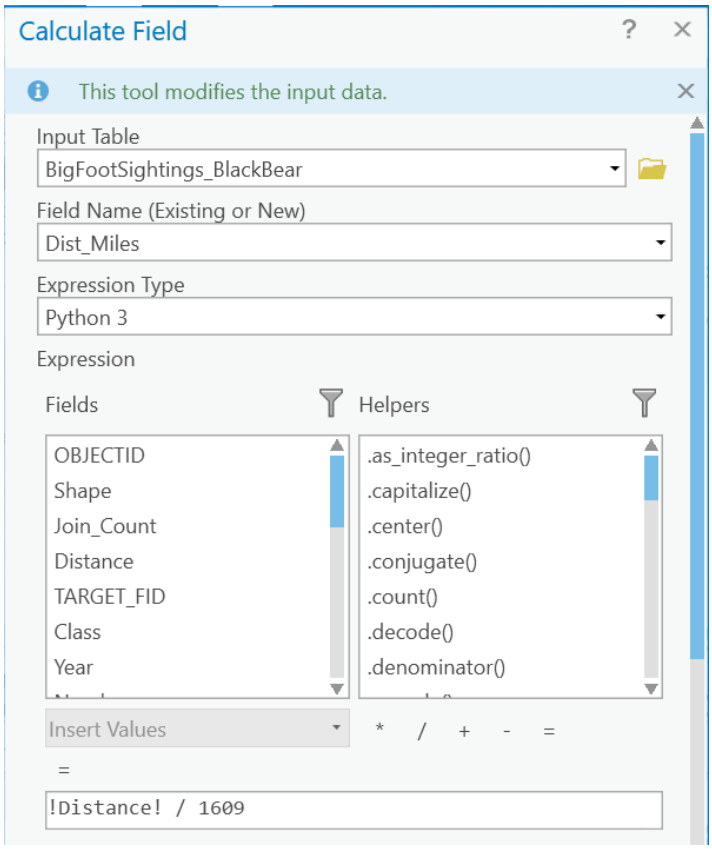
5. In the top left corner of the BigFootSightings_BlackBear attribute table, click **Add Field**.
   i. Field Name: **Dist_Miles**
   ii. Data Type: **Float**



6. Close the Fields table and click **Yes** to Save all changes.
7. Right-click **Dist_Miles** field header and select **Calculate Field.**

There are 1609 meters in a mile; we will create an expression in the Field Calculator to convert meters from the Distance field to miles in the new Dist_Miles field.

   i. In the Field Calculator, enter **!Distance! / 1609**. *Hint: double-click Distance, click the divide (/) button, and then enter the numbers 1609.*

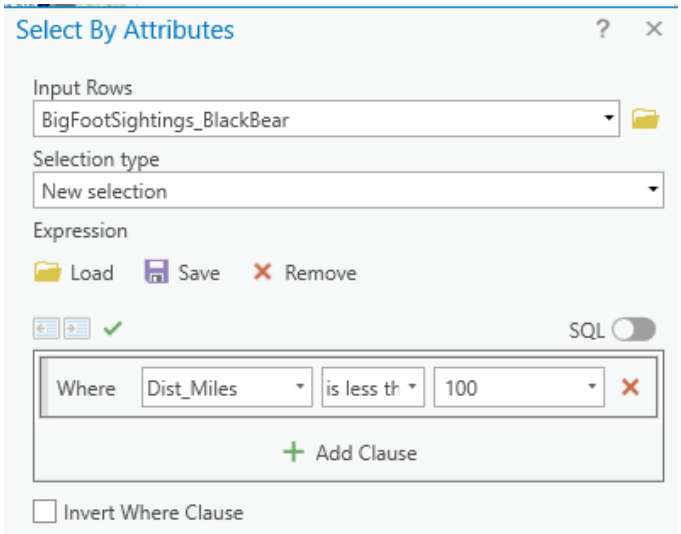


ii. Click **Apply.**

iii. Close the Calculate Field window.

8. Because we only want to include sightings within 100 miles of damaged forest, we will create a query to select only those features meeting that requirement. In the Attribute Table, click the **Select by Attributes** button.
9. In the Select by Attributes window create a **New expression**.

i. Set the following query: *Where **Dist_Miles** is **less than or equal to 100***.



ii. Click **Apply** then close the Select by Attributes window. Only the features that meet the 100 mile distance are selected.

10. Export the selected features to a new layer. In the Contents pane, right-click **BigFoot_Sightings_BlackBear** layer, click **Data,** then click **Export Features**.

i. In the Export Features window, name the output file **BigFoot_Sightings_BlackBear_Miles** and click **OK**.

LET'S REVIEW: We joined the attributes of the NorthPacificForestDamage layer to the BigFoot sightings layer, converted the distance field from meters to miles. Queried and selected distances of 100 miles or less and converted that to a new layer. Whew! We are now ready to begin our analysis.

11. From the *Analyzing Patterns* toolset, *click* **Spatial Autocorrelation (Global Moran's I)** to open the tool.

The Spatial Autocorrelation (Global Moran's I) tool measures clustering or dispersion based on feature locations and attribute values. The null hypothesis for the Spatial Autocorrelation (Moran's I) tool statistic states the attribute being analyzed is randomly distributed among the features in the study area.

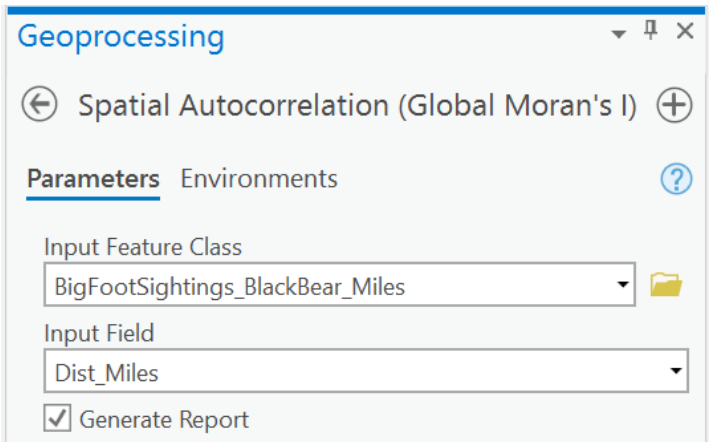
12. Set the following parameters:

i. Input Feature Class: **BigFoot_Sightings_BlackBear_Miles**.

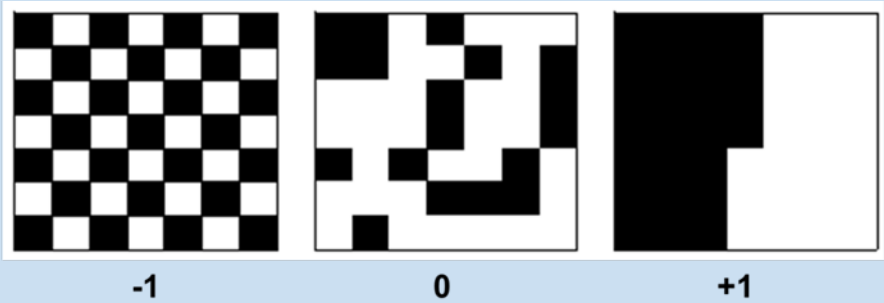
ii. Input Field: **Dist_Miles**.

iii. Generate Report: **checked**.

iv. Accept the other defaults and click **Run**.

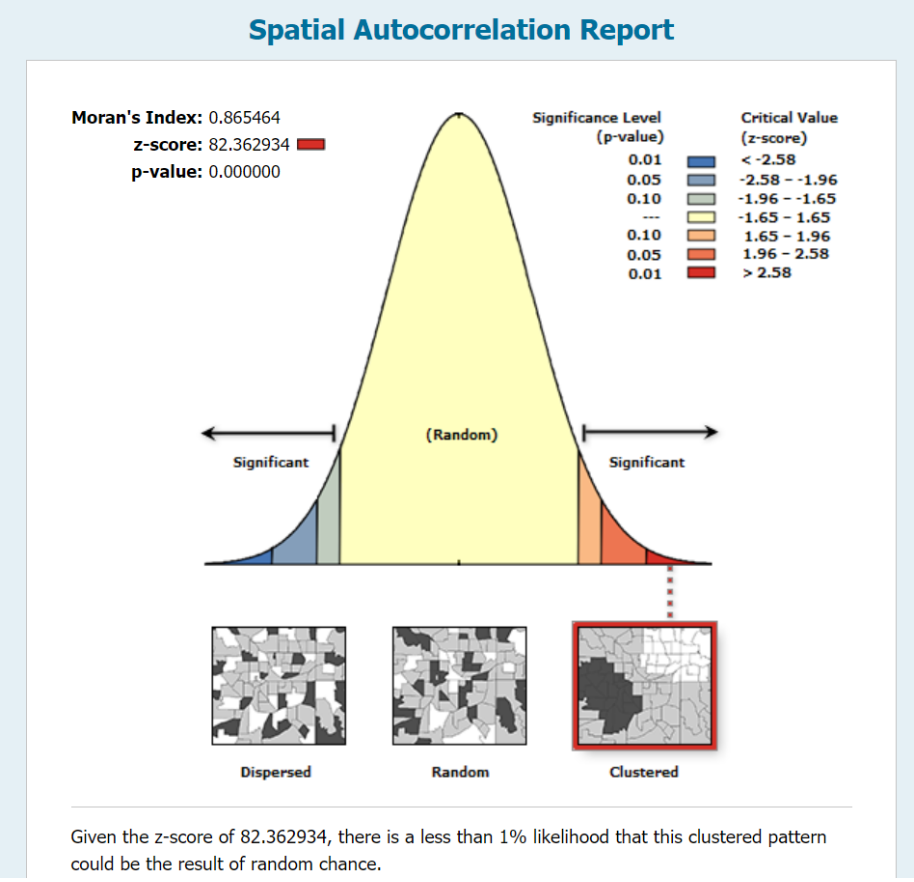


STATS 101: *The Moran's Index is a measure of spatial autocorrelation. In the example below: the white and black squares are perfectly dispersed so Moran's I would be -1. If the white squares were stacked to one half of the board and the black squares to the other, Moran's I would be close to +1. A random arrangement of square colors would give Moran's I a value that is close to 0.*



The expected index is simply a measure of spatial autocorrelation in a hypothetical random distribution indicating a spatial random pattern. A variance gives us a value indicating how far away, or the difference of the actual index, from the expected index. The z-scores and p-values tell you if you can reject the null hypothesis

13. Open the **Report File** and examine the results.



INTERPRETATION: *the report states very clearly that: Given the z-score of 82.36 there is less than 1% likelihood that this clustered pattern could be the result of random chance. We are 99% confident that the clustering in this dataset is "NOT" random*

Our data is clustered indicating a correlation between BigFoot sightings and forest damage by bears. When the p-value returned by this tool is small (statistically significant), the null hypothesis can be rejected. The p-value is statistically significant and the z-score value is positive, indicating high values and low values for the attributes are clustered in the study area. With these results, we can confirm our theory that there is a correlation between BigFoot Sightings and areas known to have bears.

14. **Save** your project.

Congratulations! You have completed Exercise 2.