



Exercise 2

Spatial Statistics

Spatial Pattern Analysis

Upon completion of this exercise, you will be familiar with:

The Analyzing Patterns Toolset

Statistical significance and confidence intervals

Several methods to evaluate and statistically validate if features, or the values associated with features, form clustered, dispersed, or random spatial patterns





Table of Contents

Introduction to Exercise 2.....	3
Part 1 – Open the map document and examine the data.....	3
Launch ArcMap and open the SpatialPatternAnalysis.mxd.....	3
Take a look at the R6_BigFoot_Sightings layer.	3
Part 2 – Locate the Analyzing Patterns Toolset.....	4
Launch ArcToolbox and expand Spatial Statistics Tools.....	4
Part 3 – Identify spatial patterns that exist in the data	4
Set up your analysis environment.	4
Execute the Average Nearest Neighbor tool to substantiate clustering or dispersion of BigFoot Sightings.	5
Use the High/Low Clustering (Getis-Ord General G*) tool to verify clustering of BigFoot Sightings increases in areas with more rainfall.	10
Use the Spatial Autocorrelation (Moran’s I) tool to verify correlation between BigFoot Sightings and areas that are known to have black bears.	16



Introduction to Exercise 2

This exercise introduces methods to evaluate and statistically validate if 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.

Part 1 – Open the map document and examine the data

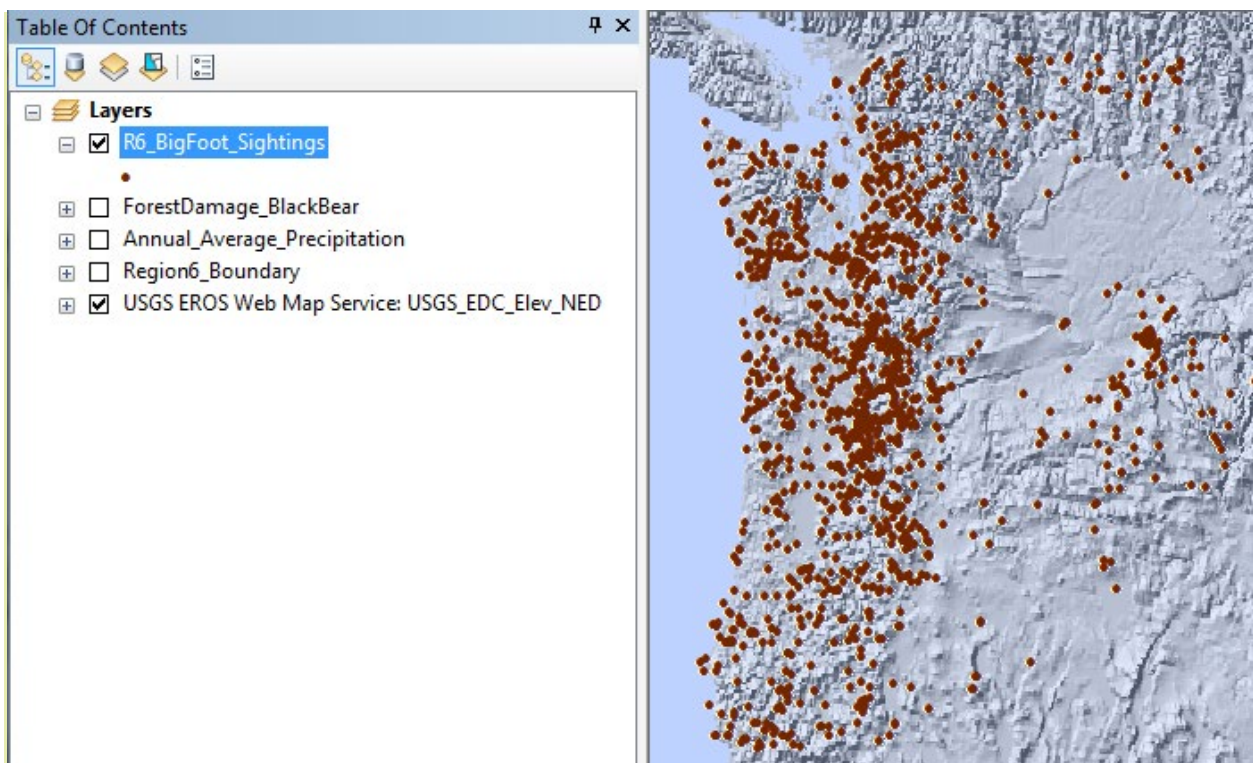
Spatial Pattern Analysis allows us to find clustering, dispersion, or random spatial patterns; and gives us the “DEGREE” to which the features are clustered, dispersed or randomly distributed.

Launch ArcMap and open the SpatialPatternAnalysis.mxd.

1. Start ArcMap.
2. Browse to the ...\\Data directory.
3. Right-click the **Ex2_SpatialPatternAnalysis.mxd** and select **Open**.

Take a look at the R6_BigFoot_Sightings layer.

1. From the Data View, take a moment to examine the R6_BigFoot_Sightings layer.



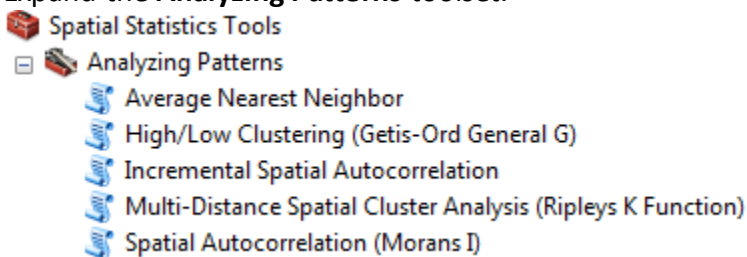
Do any specific spatial patterns such as clustering or dispersion exist? If so, does the clustering/dispersion have statistical significance? We can answer these questions with tools from the Analyzing Patterns toolset.

Part 2 – Locate the Analyzing Patterns Toolset

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

Launch ArcToolbox and expand Spatial Statistics Tools.

1. Click the **Launch ArcToolbox** button on the *Standard Toolbar* or from the *Main Menu*, select **Geoprocessing | ArcToolbox**.
2. Scroll down to **Spatial Statistics** tools and click the **plus (+) sign** adjacent to the toolset to expand the toolset.
3. 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.

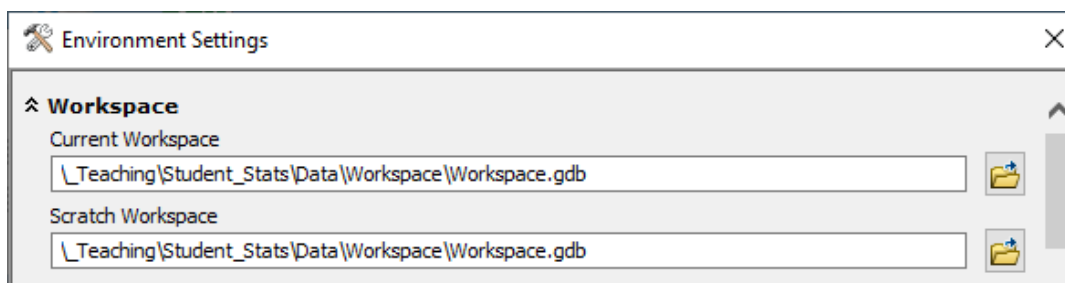
The Analyzing Patterns tools answer questions such as, *"Are the features in the dataset, or the values associated with the features in the dataset, spatially clustered?"* and *"Is the clustering becoming more or less intense over time?"*

Part 3 – Identify spatial patterns that exist in the data

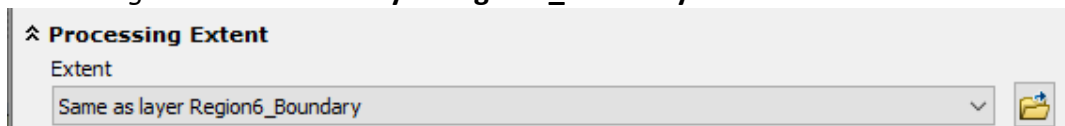
NOTE: Before you begin using tools within the Spatial Statistics toolbox, you should always set up your analysis environment. The analysis environment includes specifying a working directory, establishing an analysis mask and selecting an output coordinate system among other things.

Set up your analysis environment.

1. From the Geoprocessing menu, choose **Environments**.
2. Set both the *Current and Scratch Workspaces* to the **../Data/Workspace/Workspace.gdb**.



3. Expand the *Output Coordinates* section.
4. Choose **Same as Input** for the *Output Coordinate System*.
5. *Processing Extent*: **Same as layer Region6_Boundary**.

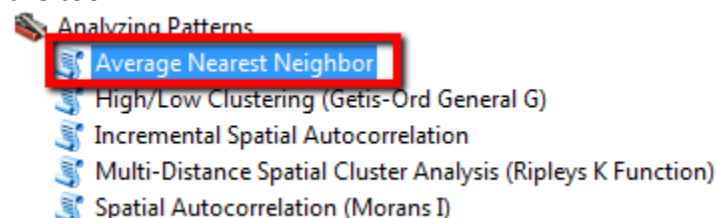


6. Click **OK** to accept the changes and close the *Environment Settings* dialog.
7. Save the Map document.

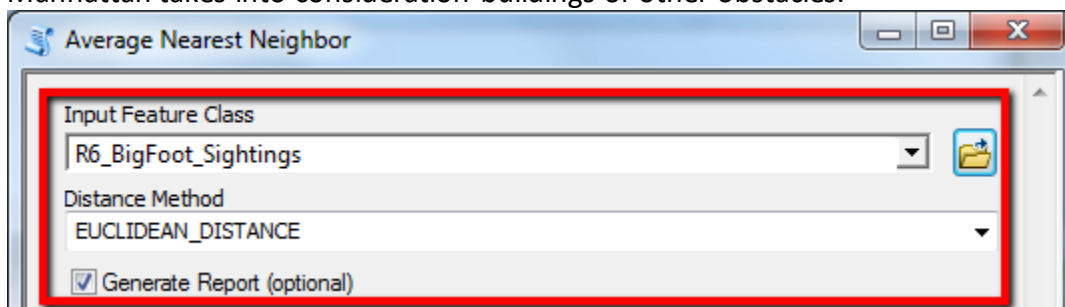
Execute the Average Nearest Neighbor tool to substantiate clustering or dispersion of BigFoot Sightings.

NOTE: 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.

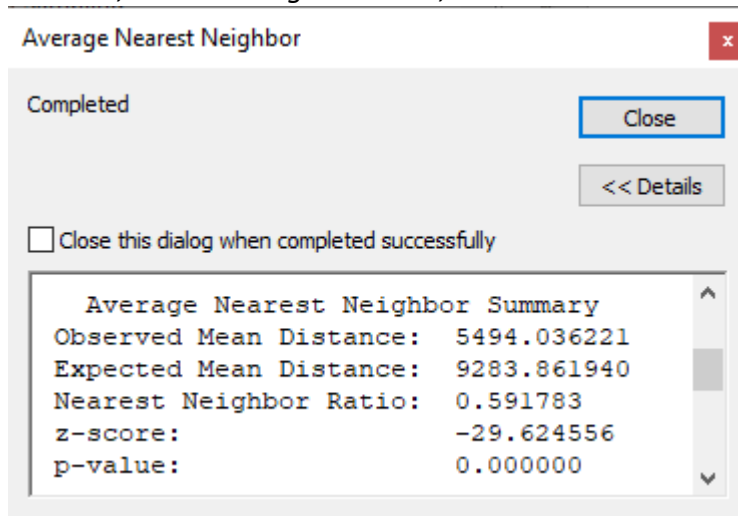
1. From the *Analyzing Patterns* toolset, double-click **Average Nearest Neighbor** to open the tool.



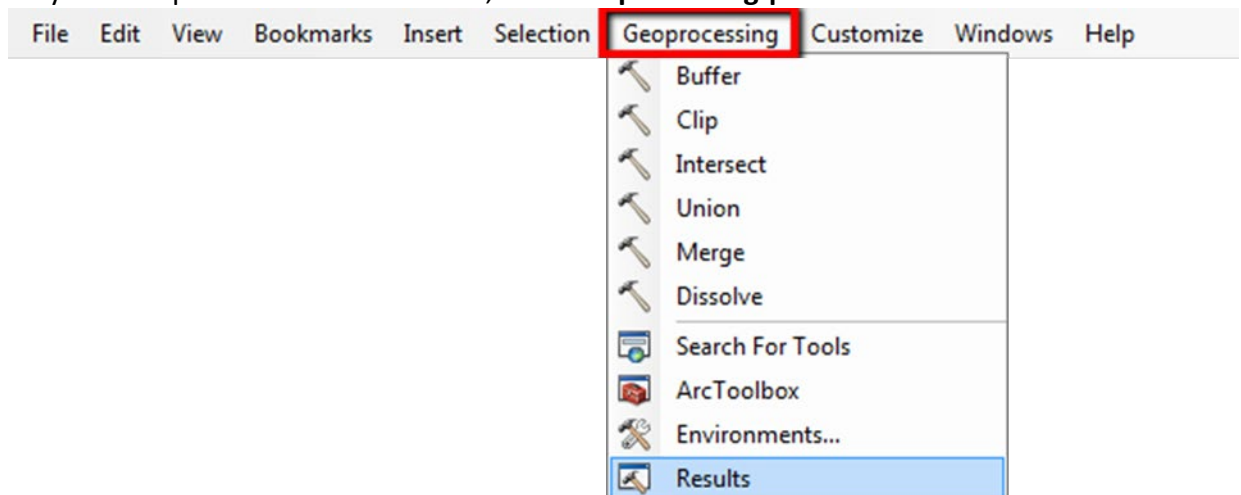
2. For *Input Feature Class*: select **R6_BigFoot_Sightings** from the dropdown.
3. Select **EUCLIDEAN_DISTANCE** for Distance Method. This is a straight line distance. Manhattan takes into consideration buildings or other obstacles.



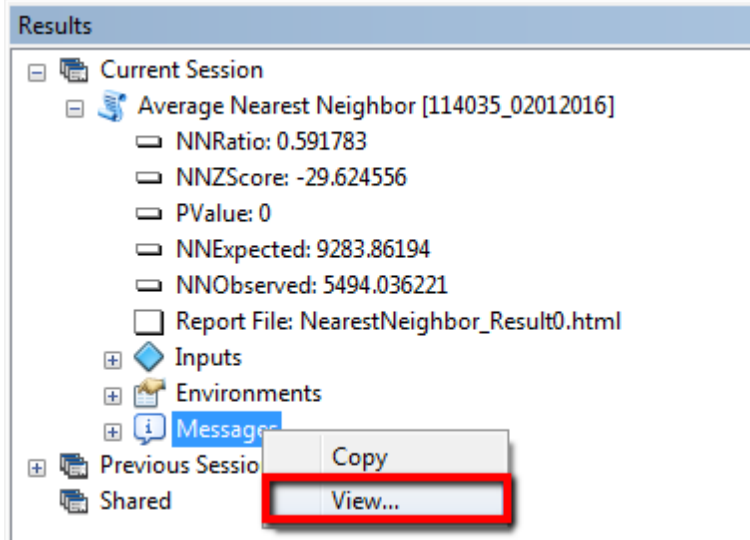
4. *Enable the checkbox adjacent to **Generate Report** to create an HTML file with a graphical summary of the results.*
5. Click **OK** to run the tool. Whenever you execute a tool using its dialog box or in the Python window, information about the execution is written as a result in the Results window. The tool returns five values: *Observed Mean Distance, Expected Mean Distance, Nearest Neighbor Index, Z score and a P value.*



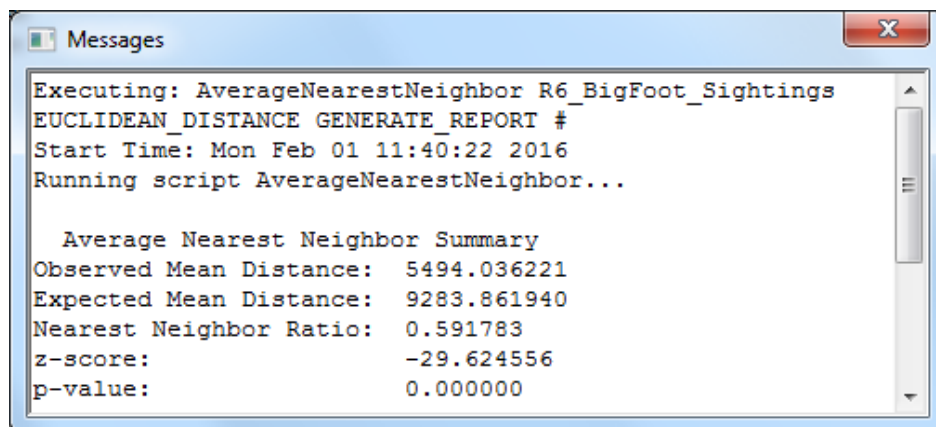
6. Or you can open the Results window, click **Geoprocessing | Results**.



7. Expand, **Current Session**, then double-click **Average Nearest Neighbor**. This will re-open the tool and includes the parameters previously selected. **Close** the tool.
8. Click the **plus sign (+)** adjacent to *Average Nearest Neighbor* to expand its contents.
9. To display the results in a Message dialog box, *right-click Messages* in the Results window and select **View**.



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.

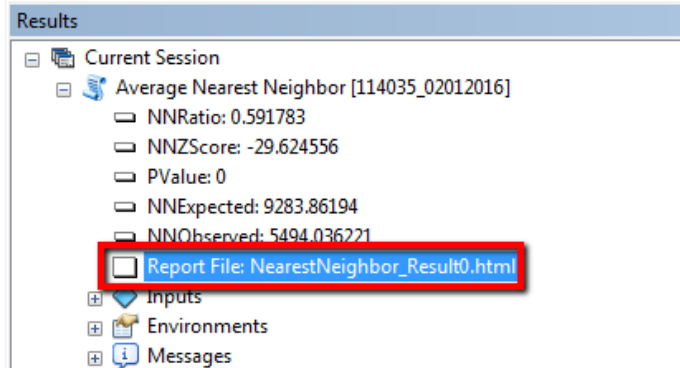


The

Observed Mean

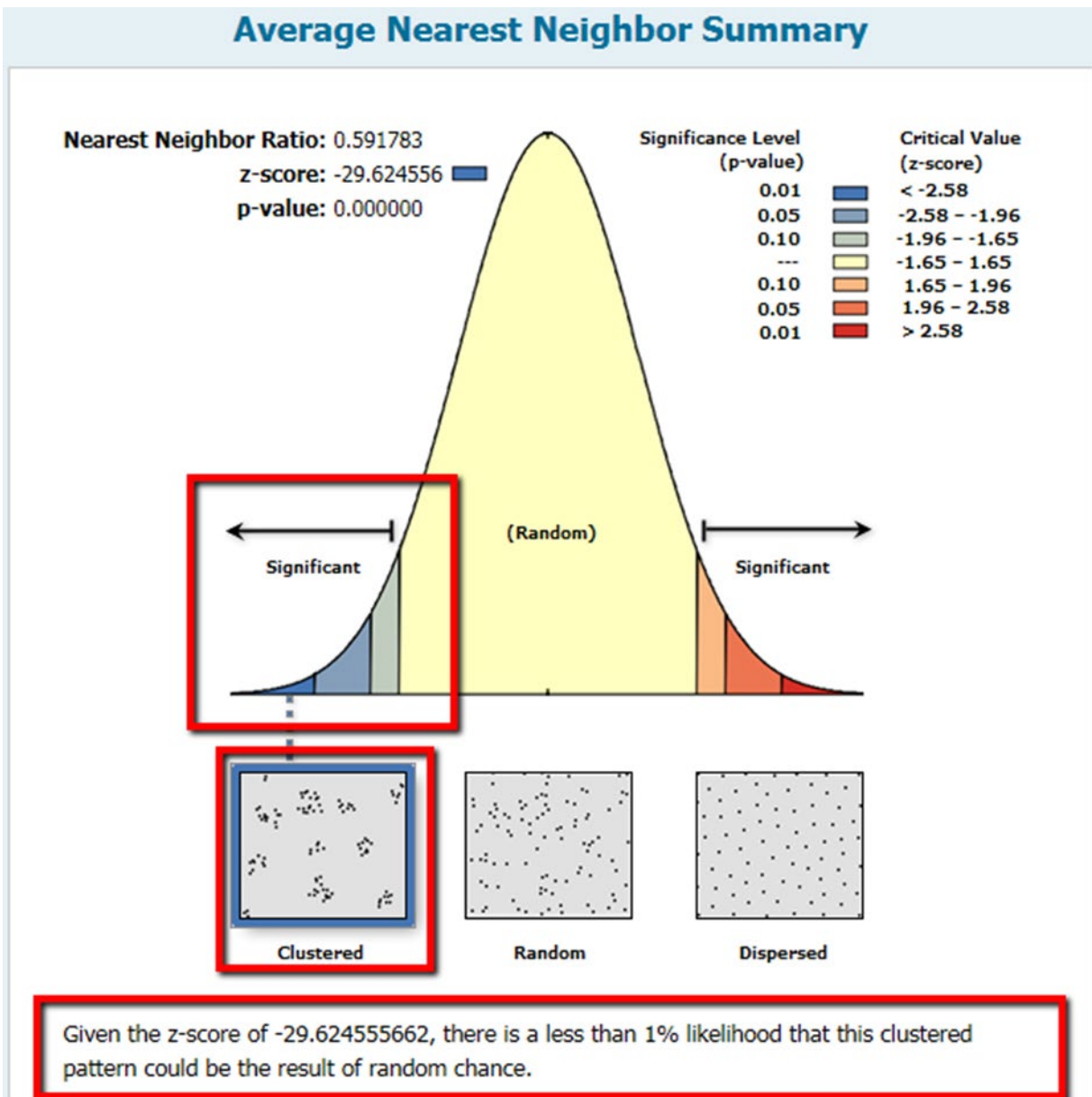
Distance is the actual distance between each point and the closest neighboring point and the **Expected Mean Distance** is the average distance between neighbors in a hypothetical random distribution. The **Nearest Neighbor Index** is expressed as the ratio of the Observed Mean Distance to the Expected 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.

10. Close the **Messages** dialog and then **double-click HTML Report File** in the Results window to open a graphic version of the results



NOTE: The Report should open in an Internet Explorer window. If you are using ArcGIS via Citrix or using a different Internet application the report may not open. A copy of the report is below so you can read it.

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.*



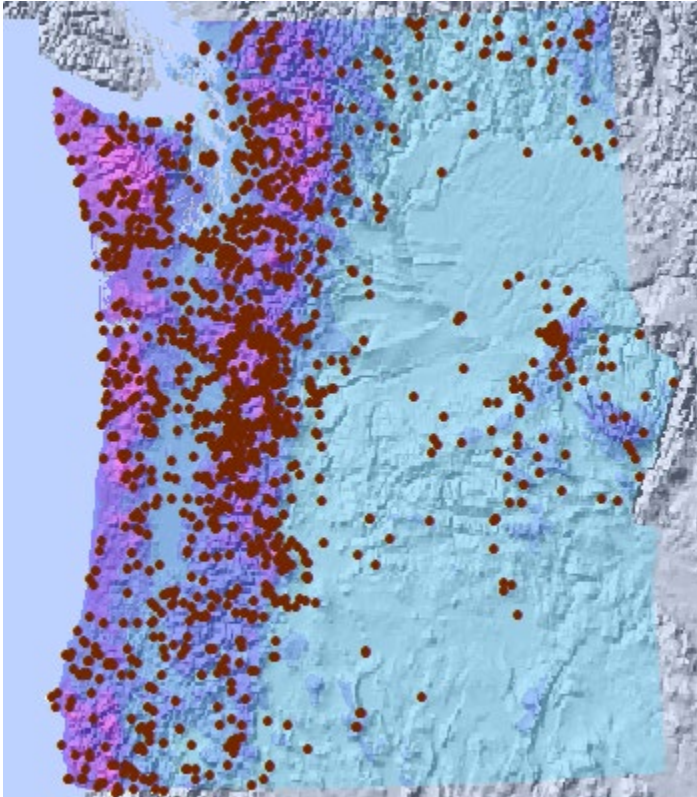
NOTE: Because the index or the average nearest neighbor ratio is less than 1, we know that the pattern exhibits clustering. And since the p-value is very small, it means it is very unlikely that the observed spatial pattern is the result of a random process, so we can reject the null hypothesis.

11. **Close** the Average Nearest Neighbor Summary.

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.

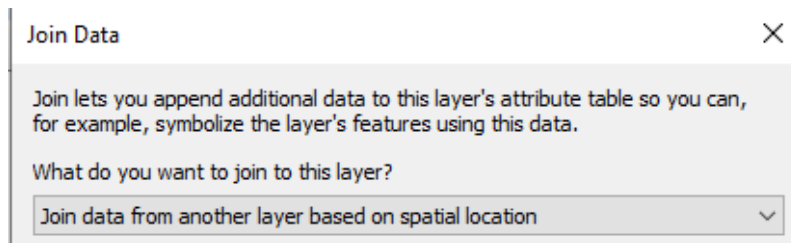
Use the High/Low Clustering (Getis-Ord General G*) tool to verify clustering of BigFoot Sightings increases in areas with more rainfall.

1. From the Table of Contents, turn on **Annual_Average_Precipitation**. BigFoot Sightings 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 BigFoot_Sightings layer based on their spatial relationship. The target features (BigFoot_Sightings) and the attributes from the join features (Annual_Average_Precipitation) will be written to a new output feature class.

2. Open the Attribute Table for **R6_BigFoot_Sightings**.
3. From the *Table Options* dropdown, select **Joins and Relates | Join...**
4. In the Join Data dialog, **Join data from another layer based on spatial location**.



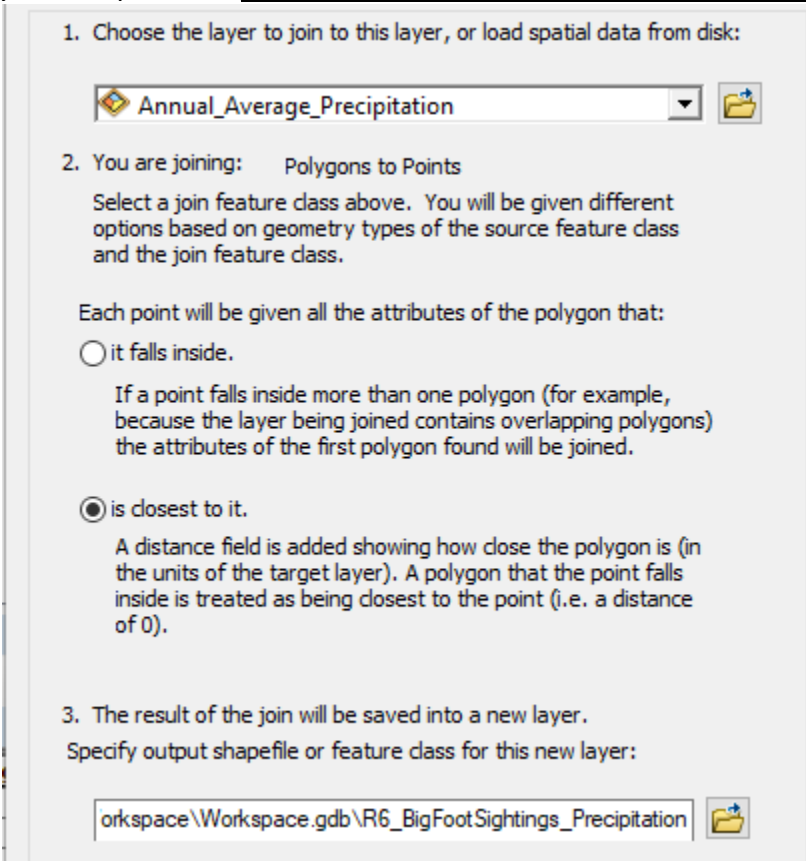
Join Data [X]

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.


What do you want to join to this layer?

Join data from another layer based on spatial location

5. The layer we want to join to this layer is **Annual_Average_Precipitation**.
6. Each point will be given all the attributes of the polygon that **is closest to it**.
7. For the output shapefile or feature class name, browse to the Workspace directory, and then type **R6_BigFootSightings_Precipitation**. *NOTE: Because we are not performing a geoprocessing operation, this process is not included in the Geoprocessing environment you set up earlier; you will need to browse to the output file location.*



1. Choose the layer to join to this layer, or load spatial data from disk:

 Annual_Average_Precipitation

2. You are joining: Polygons to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each point will be given all the attributes of the polygon that:

☐ it falls inside.

If a point falls inside more than one polygon (for example, because the layer being joined contains overlapping polygons) the attributes of the first polygon found will be joined.

☒ is closest to it.

A distance field is added showing how close the polygon is (in the units of the target layer). A polygon that the point falls inside is treated as being closest to the point (i.e. a distance of 0).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

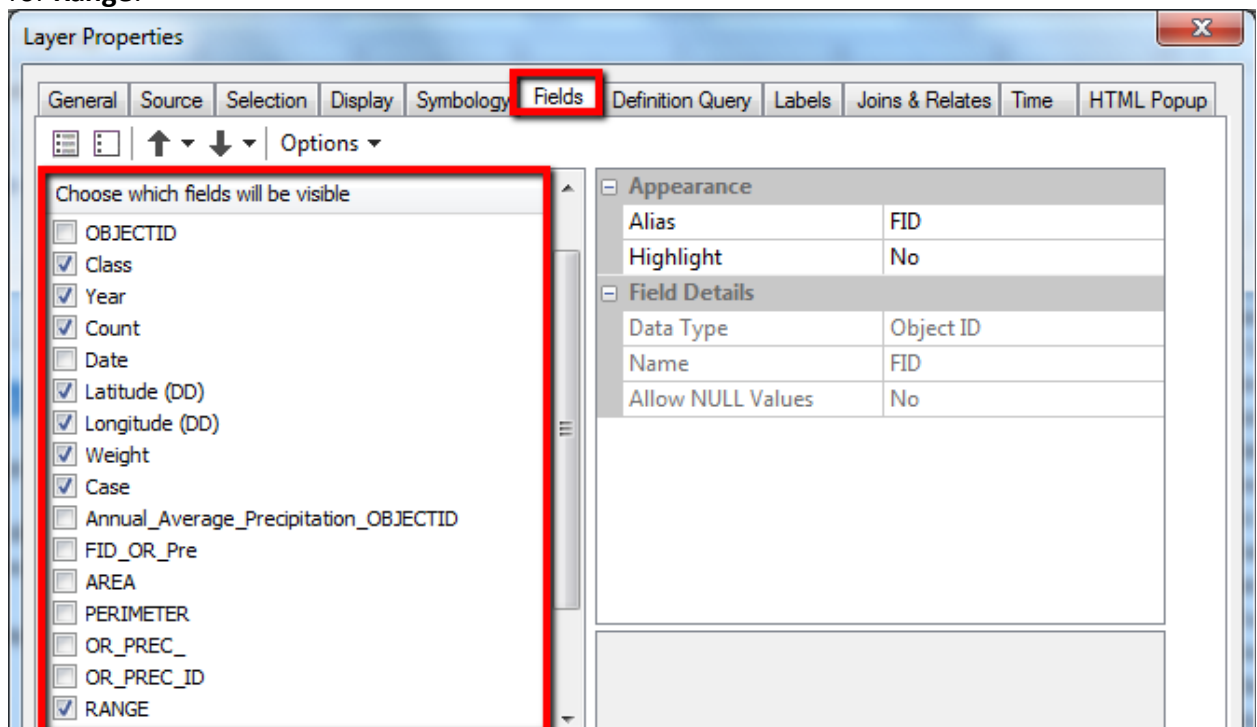
orkspace\Workspace.gdb\R6_BigFootSightings_Precipitation

8. Click **OK** to accept the parameters and perform the join operation and then **close** the attribute table. *Be patient... After a couple of minutes, the new layer will be added to the Table of Contents.*
9. *Open the attribute table* for the new **R6_Big_Foot_Sightings_Precipitation** layer. Scroll through the layers attributes paying close attention to the new attribute columns. We are interested in the Range field which addresses the average amount of annual precipitation in inches. To make the table easier to manage, we can turn off the other precipitation fields. **Close the Attribute Table**.

R6_BigFootSightings_Precipitation							
	FID_OR_Pre	AREA	PERIMETER	OR_PREC_	OR_PREC_ID	RANGE	LEGEND
▶	8	0.024616	7.546867	10	234	81	80-90
	9	0.075227	16.5539	11	244	79	70-80
	11	0.075902	16.205384	13	255	77	70-80
	12	0.062946	13.758155	14	255	75	70-80
	12	0.062946	13.758155	14	255	75	70-80
	15	0.09197	16.940105	17	255	73	70-80

10. Open the **Layer Properties** for *R6_Big_Foot_Sightings_Precipitation*

11. Under the **Fields** tab, disable (uncheck) all the recently added precipitation fields except for **Range**.

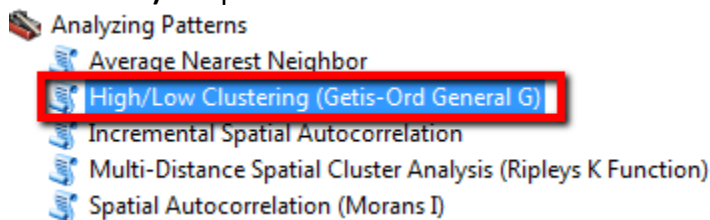


12. Click **Apply** and **OK** to close the layer properties window. *The attribute table now displays only those items selected in the Fields tab.*

Table								
R6_BigFootSightings_Precipitation								
	Class	Year	Count	Latitude (DD)	Longitude (DD)	Weight	Case	RANGE
▶	A	1994	1	-123.56	46.2043	3	6	81
	A	1993	1	-123.49	44.3808	3	6	79
	C	2008	1	-123.486	44.3695	1	6	77
	C	2001	1	-123.346	45.7639	1	6	75
	B	1984	1	-123.368	45.3962	2	5	75
	B	1949	1	-124.023	43.3038	2	4	73
	B	2003	1	-123.314	45.8209	2	6	73

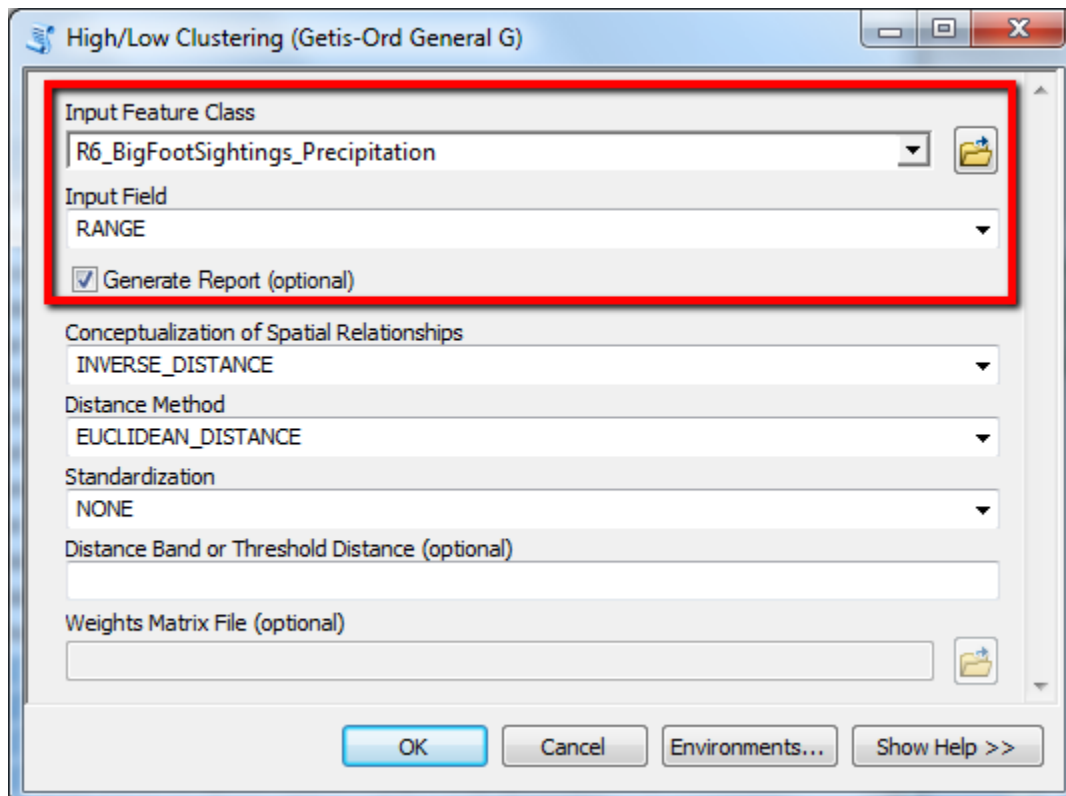
Now that we have added precipitation values to the BigFoot Sightings layer, we can begin our analysis.

- From the *Analyzing Patterns* toolset, double-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.

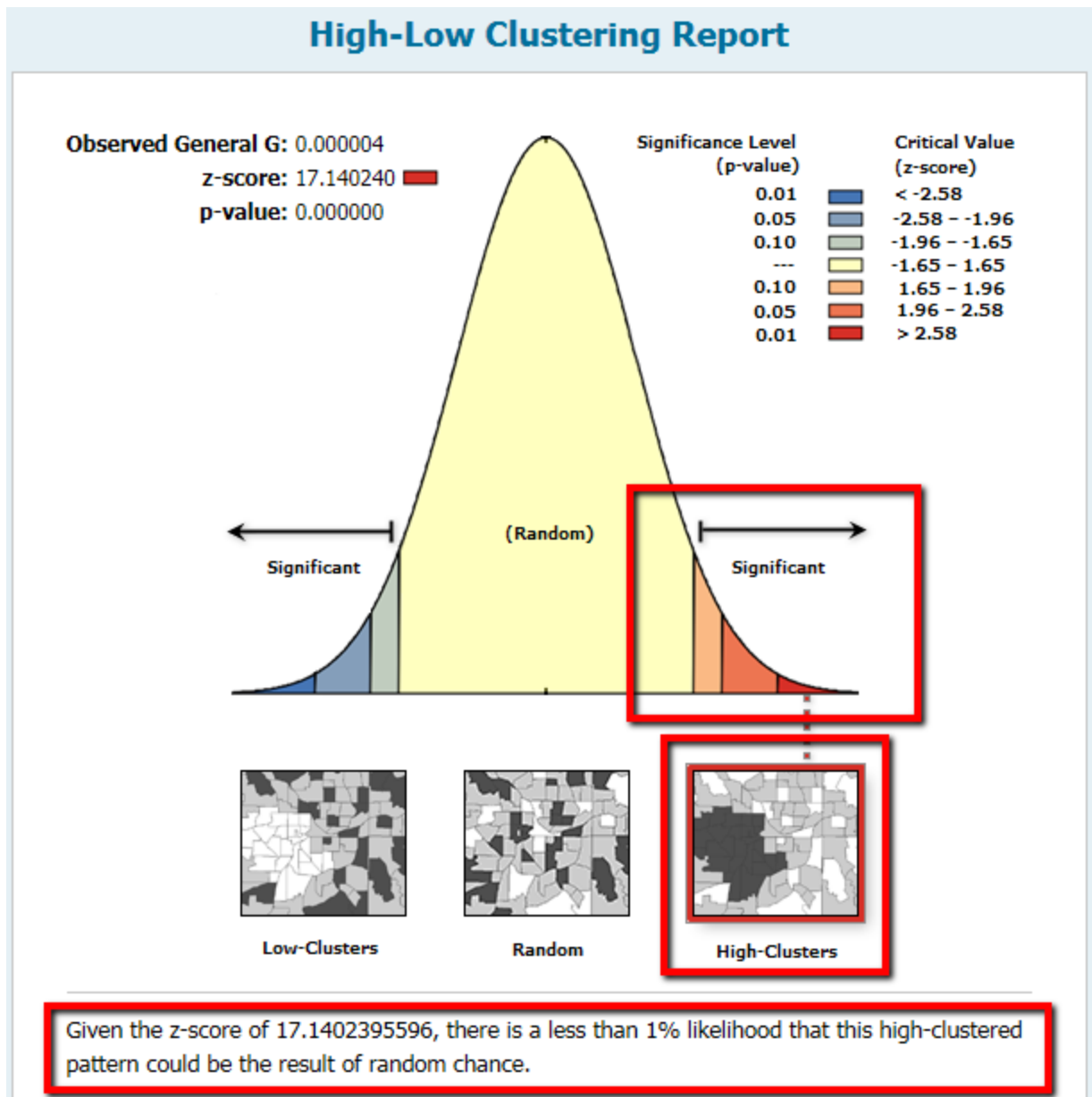
- For *Input Feature Class*: select: **R6_BigFoot_Sightings_Precipitation**.
- For *Input Field*: select **Range**.
- Enable, **Generate Report**.



17. We will accept all other defaults. Click **OK**.
18. Open the Results window. **Geoprocessing | Results**.
19. Under *Current Session*, expand the results for **High/Low Clustering**.
20. View the Messages by right-clicking **Messages** and selecting **View**.

What does the yellow exclamation symbol 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.

21. *Double-click* to open the **HTML** results graphic.



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.

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

1. From the Table of Contents, turn on **ForestDamage_BlackBear**. We need to repeat the steps from before to spatially join the attributes of the *ForestDamage_BlackBear* layer to the *BigFoot sightings* layer.


NOTE: The ForestDamage_BlackBear dataset was obtained from the dataset "Damage to forested areas in the North Pacific LCC" a compilation of forest insect, disease and abiotic damage mapped by aerial detection surveys on forested areas in the United States. Credits: USDA Forest Service, Forest Health Technology Enterprise Team (FHTET)

2. Open the *Attribute Table* for **R6_BigFoot_Sightings**.
3. From the *Table Options* dropdown, select **Joins and Relates | Join...**
4. In the *Join Data* dialog, we want to **Join data from another layer based on spatial location**.
5. The layer we want to join to this layer is **ForestDamage_BlackBear**.
6. Each point will be given all the attributes of the polygon that is **closest to it**.
7. Name the output shapefile or feature class, **R6_BigFootSightings_Black Bear**.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

 ForestDamage_BlackBear
2. You are joining: Polygons to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

Each point will be given all the attributes of the polygon that:

☐ it falls inside.

If a point falls inside more than one polygon (for example, because the layer being joined contains overlapping polygons) the attributes of the first polygon found will be joined.

☒ is closest to it.

A distance field is added showing how close the polygon is (in the units of the target layer). A polygon that the point falls inside is treated as being closest to the point (i.e. a distance of 0).
3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

Workspace\Workspace.gdb\R6_BigFootSightings_Black Bear

8. Click **OK**. After a few minutes, the new layer is added to the Table of Contents.
9. Open the attribute table for the new **R6_BigFootSightings_BlackBear** layer. Scroll through the attributes paying close attention to the *Distance* field.

	HOST	FORTYPE	ACRES	Shape_Leng	Shape_Area	Distance
▶	No data	MIXED CONIFERS	2	319.622627	8119.239763	2710.0252
	No data	MIXED CONIFERS	32.3	1399.218557	130551.729734	1579.972289
	No data	MIXED CONIFERS	2	319.6226	8119.238393	2263.248159
	No data	MIXED CONIFERS	2	319.622628	8119.2398	6162.084796
	No data	MIXED CONIFERS	2	319.62258	8119.23737	1113.912336
	No data	MIXED CONIFERS	31.6	1518.106189	127720.018454	6169.387993

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.

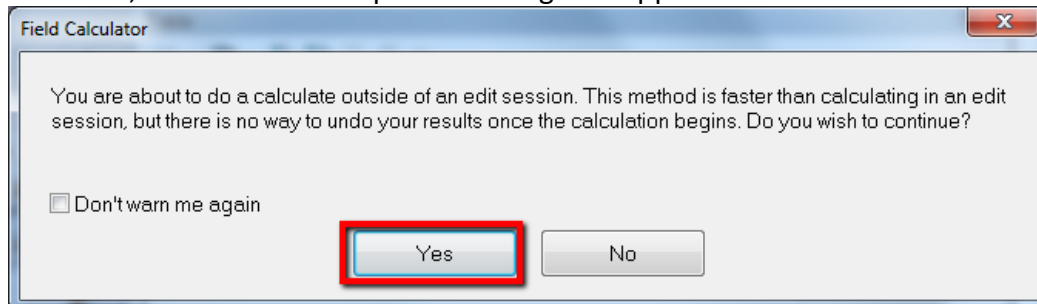
10. In the *R6_BigFootSightings_BlackBear* attribute table, create a new field. **Table options** | **Add field**.
11. In the *Add Field* dialog, for Name enter **Dist_Miles**.
12. Select **Float** for the Type.

13. Click **OK**.
14. Right-click **Dist_Miles** and select **Field Calculator**.

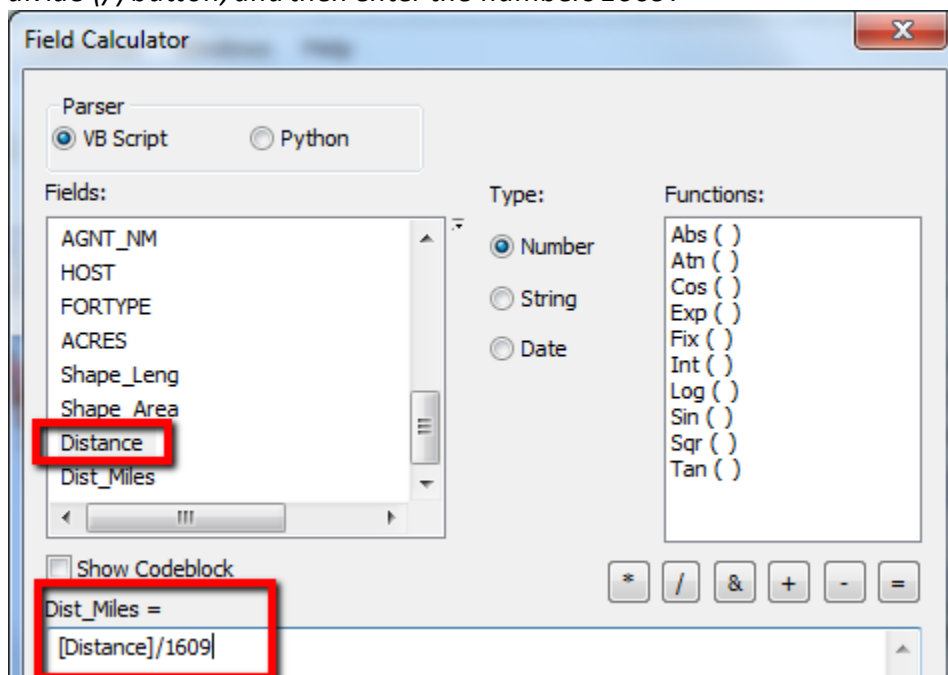
	ACRES	Shape_Leng	Shape_Area	Distance	Dist
	2	319.622627	8119.239763	2710.0252	
	32.3	1399.218557	130551.729734	1579.972289	
	2	319.6226	8119.238393	2263.248159	
	2	319.622628	8119.2398	6162.084796	
	2	319.62258	8119.23737	1113.912336	
	31.6	1518.106189	127720.018454	6169.387993	

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.

15. If needed, select **Yes** to accept the warning that appears.

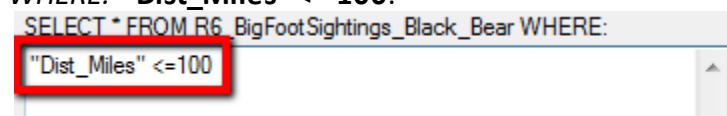


16. In the Field Calculator, enter **[Distance]/1609**. *Hint: double-click Distance, click the divide (/) button, and then enter the numbers 1609.*



17. 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. From the Attribute Table, open the **Select by Attributes** dialog.

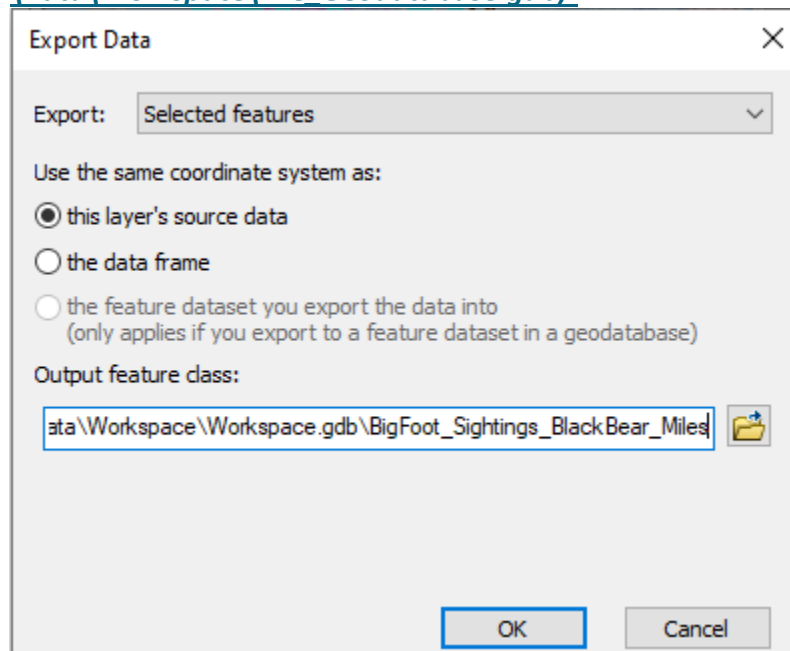
18. The query should be as follows: **SELECT *FROM R6_BigFoot_Sightings_BlackBear WHERE: "Dist_Miles" <= 100.**



19. Click **Apply**. **Close** the Select by Attribute dialog. Only the features that met the 100 mile distance are selected.

20. **Close** the Attribute Table.
21. Export the selected features to a new layer. *Right-click* **R6_BigFoot_Sightings_BlackBear | Data | Export Data**.
22. Export the **Selected features**, using the same coordinate system as “**this layer’s source data**” and name the output file something like **BigFoot_Sightings_BlackBear_Miles**.

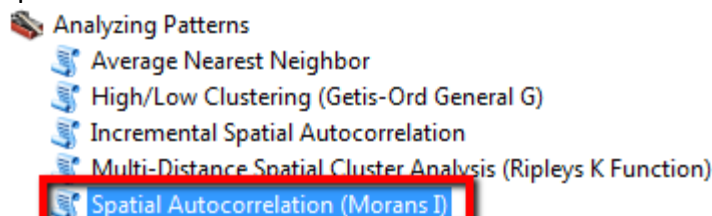
NOTE: The output feature class must be saved to a file geodatabase (e.g... SpatialStatistics\Data\Workspace\File_Geodatabase.gdb).



23. Click **Yes** if prompted to add the new layer to your map. *Clear* any selected features in your map.

LET’S REVIEW: We joined the attributes of the ForestDamage_BlackBear 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.

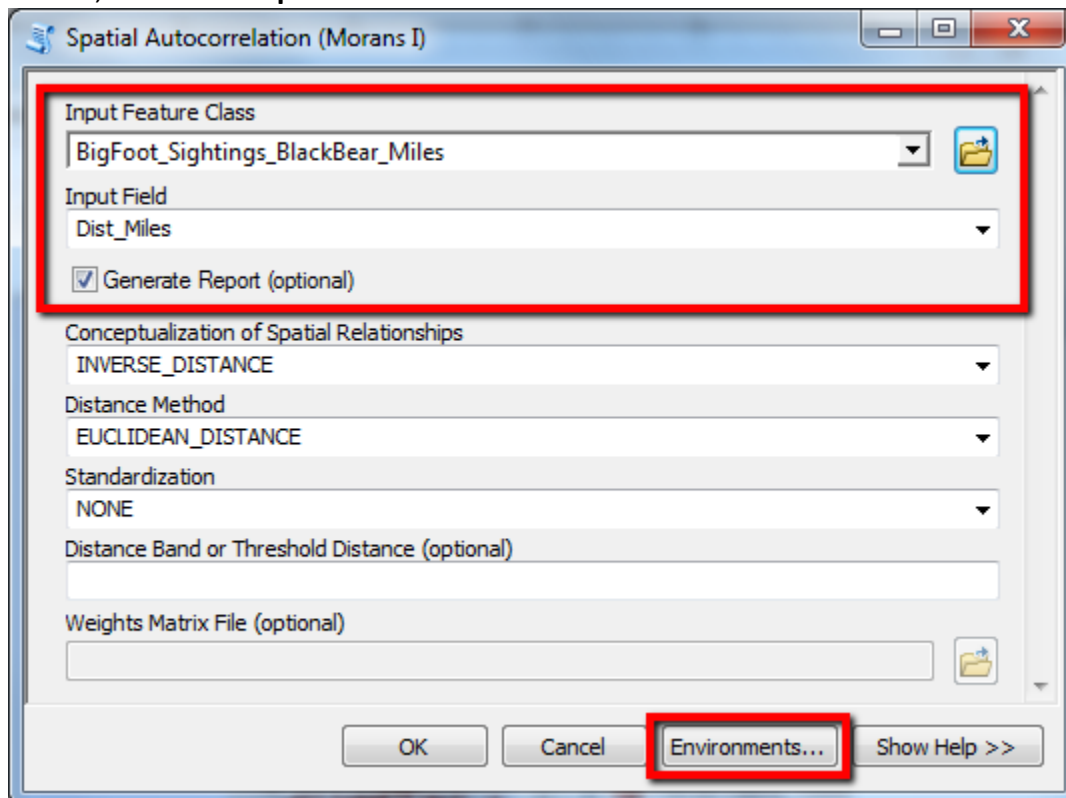
24. From the *Analyzing Patterns* toolset, *double-click* **Spatial Autocorrelation (Moran’s I)** to open the tool.



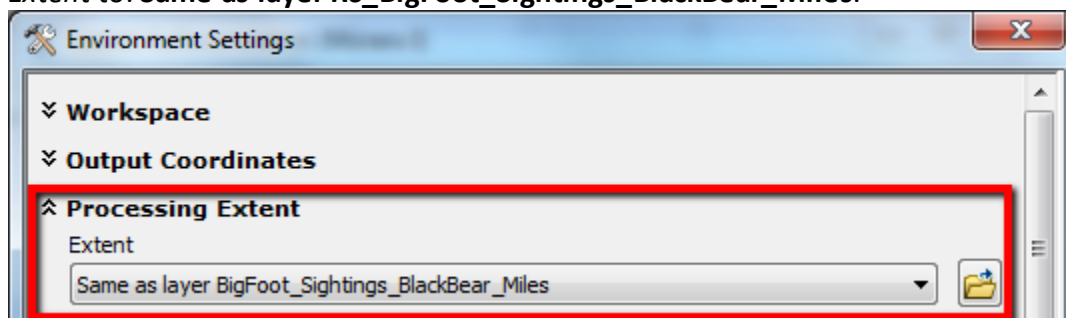
The Spatial Autocorrelation (Moran’s I) tool measures clustering or dispersion based on feature locations and attribute values. The null hypothesis for the Spatial Autocorrelation (Moran’s I) tool

statistic states the attribute being analyzed is randomly distributed among the features in the study area.

25. For *Input Feature Class* select **R6_BigFoot_Sightings_BlackBear_Miles** from the dropdown.
26. For *Input Field* select **Dist_Miles**.
27. Enable, **Generate Report**.



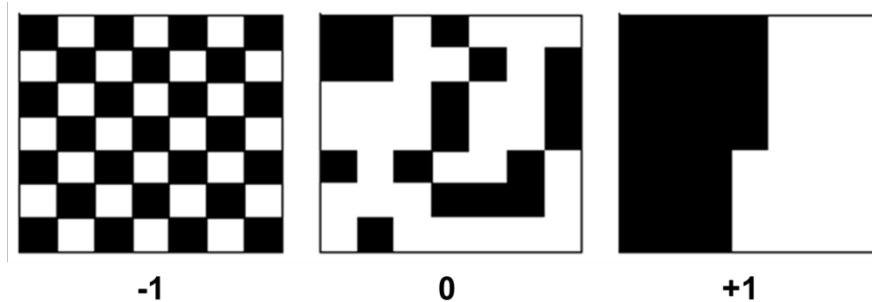
28. Click the **Environments** button. In the Environment Settings dialog, update *Processing Extent* to: **Same as layer R6_BigFoot_Sightings_BlackBear_Miles**.



29. Click **OK**. Accept all other defaults and click **OK** again to run the tool. The tool returns five values: the Moran's I Index, Expected Index, Variance, z-score, and p-value.

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.



STATS 101: 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

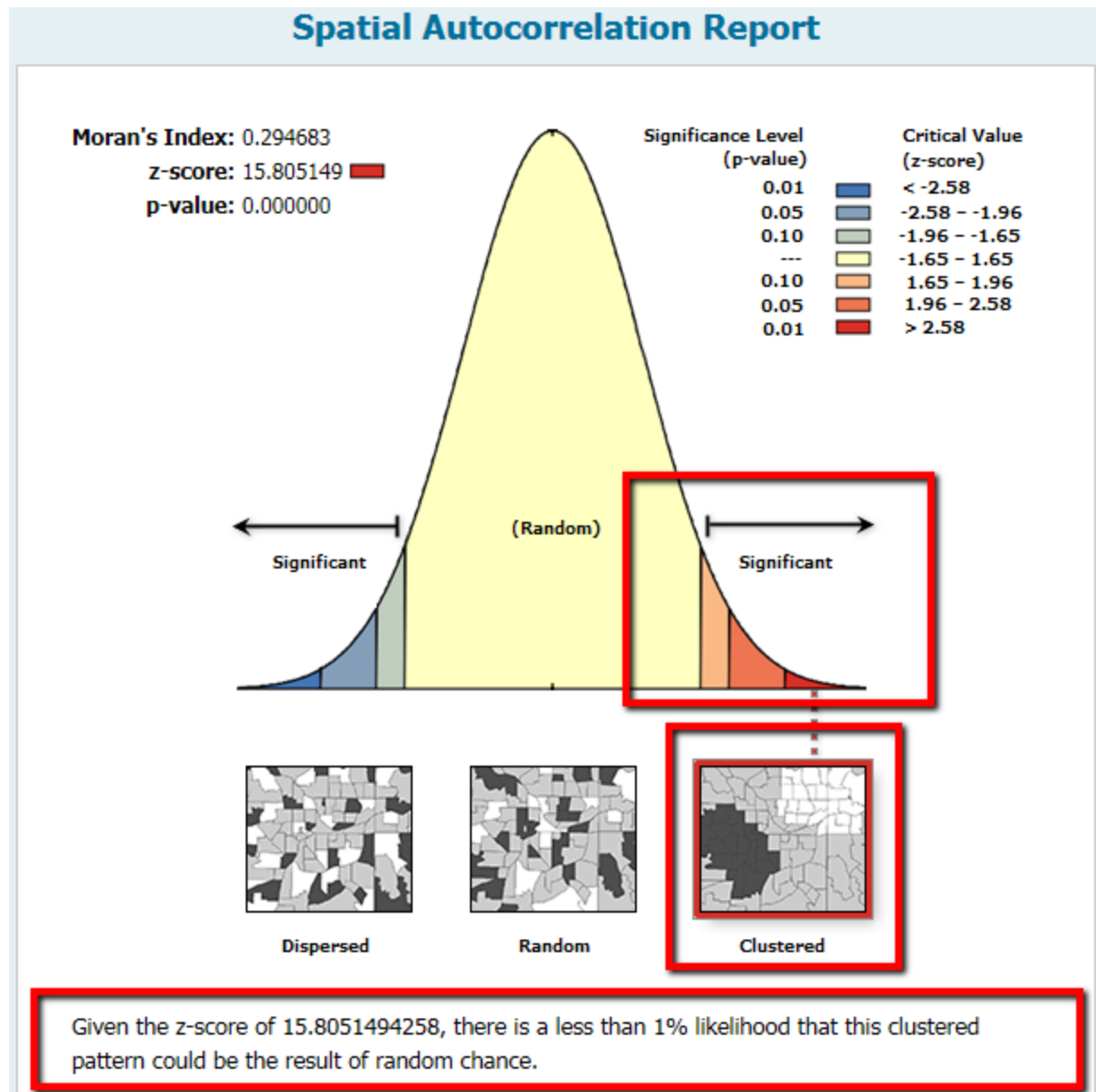
30. Open the *Results* window. **Geoprocessing | Results.**

31. Open the **HTML** results graphic.

INTERPRETATION: the report states very clearly that: Given the z-score of 15.81, 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.



32. **Save** and **Close** your ArcMap project.

You have completed Exercise 2... Please notify your instructor.