



Exercise 3

Spatial Statistics

Mapping Clusters

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

Tools and functionality within the Mapping Clusters, Measuring Geographic Distributions and Analyzing Patterns toolsets

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

Performing cluster analysis to identify the locations of statistically significant hot spots, cold spots, and spatial outliers.





Contents

Introduction to Exercise 3	3
Part 1 – Open the Mapping Clusters project and view the data.....	3
Launch ArcMap and open the MappingClusters.mxd.	3
Take a look at the Historic Earthquakes layer.....	3
Calculate a total yearly count and averages for Historical Earthquakes.	4
Part 2 – Locate the Mapping Clusters Toolset.....	5
Launch ArcToolbox and expand Spatial Statistics Tools.....	5
Part 3 – Explore the Spatial Distribution of the data	5
Set up your analysis environment.	5
Use the Standard Distance tool to help identify trends in earthquakes.....	6
Use the Standard Distance tool to help identify trends in earthquakes based on the year they occurred.	9
Explore the attribute data of earthquakes occurring in 1980.....	10
Part 4 – Identify if the spatial clustering has statistical significance.....	11
Use the Cluster and Outlier Analysis (Anselin Local Moran's I) tool to identify statistically significant spatial clusters of earthquakes within the same Year.....	12
Select the Historical Earthquakes with statistically significant clustering and review their attributes.	13
Part 5 – Utilize the Optimized Hot Spot tool to identify high (hot) & low (cold) value clustering of earthquake magnitudes.....	16
Make sure your data is projected correctly	17
Use the Optimized Hot Spot tool to identify statistically significant spatial clustering or sparsity of point locations.....	17
Use the Optimized Hot Spot Analysis tool to identify statistically significant clustering of historical earthquake magnitudes high values (hot spot) and low values (cold spot).	19
Visualize your results. Use your results to create an interpolated surface and prepare for presentation.....	20



Introduction to Exercise 3

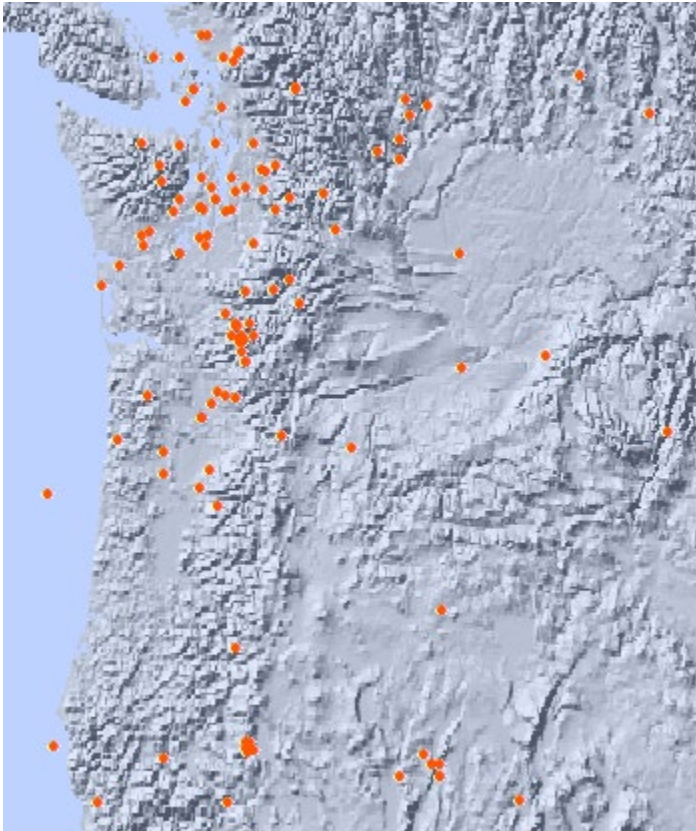
This exercise will walk students through the process of summarizing and evaluating geographic distributions; accessing broad geographic patterns and trends over time and in identifying where spatial clustering occurs and where spatial outliers are located.

Part 1 – Open the Mapping Clusters project and view the data

Cluster Analysis is used to identify the locations of statistically significant hot spots, cold spots, and spatial outliers and allows us to answer questions such as: Where are the clusters? And where are features with similar attributes are clustered

Launch ArcMap and open the Ex3_MappingClusters.mxd.

1. Start ArcMap.
2. *Browse* to the ...**Data** directory.
3. *Right-click* the **Ex3_MappingClusters.mxd** and click **Open**.



Take a look at the Historic Earthquakes layer.

The Historical Earthquake dataset is an ESRI dataset that presents the locations of significant, historic earthquakes that caused deaths, property damage, and geological effects, or were otherwise experienced by residents in the area.

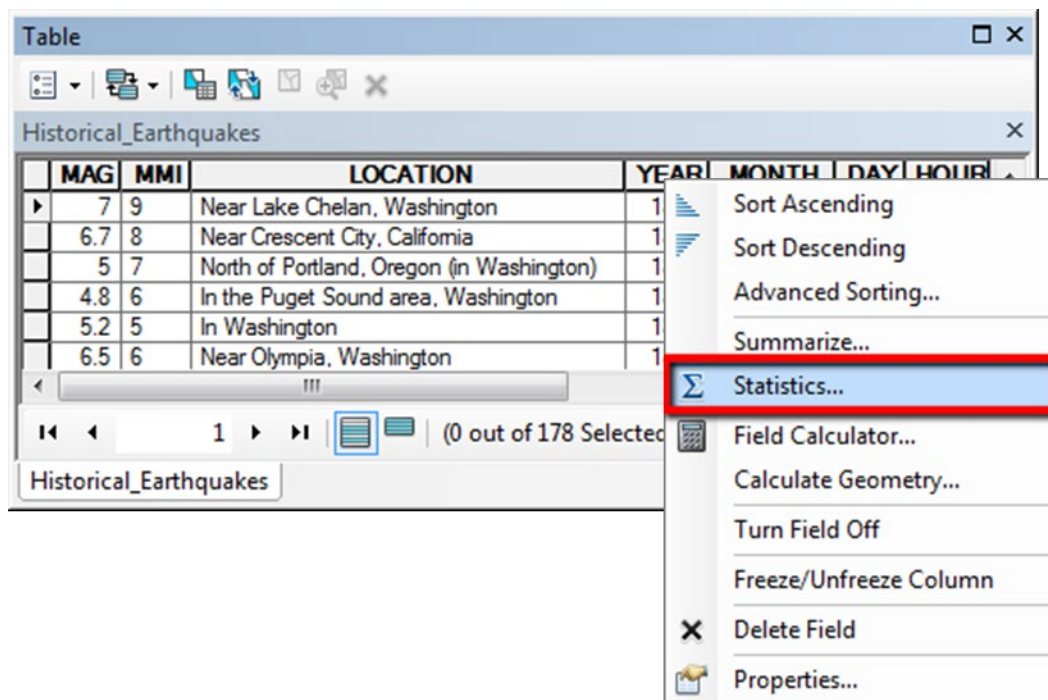
1. From the Data View, take a moment to examine the **Historical Earthquakes** features and attributes.

- Is there any noticeable clustering or dispersion amongst the features? Is there any noticeable clustering or dispersion amongst the features' attributes? These are questions we can answer using tools from the Mapping Clusters toolset.
- Open the **Attribute Table** for *Historical Earthquakes*.

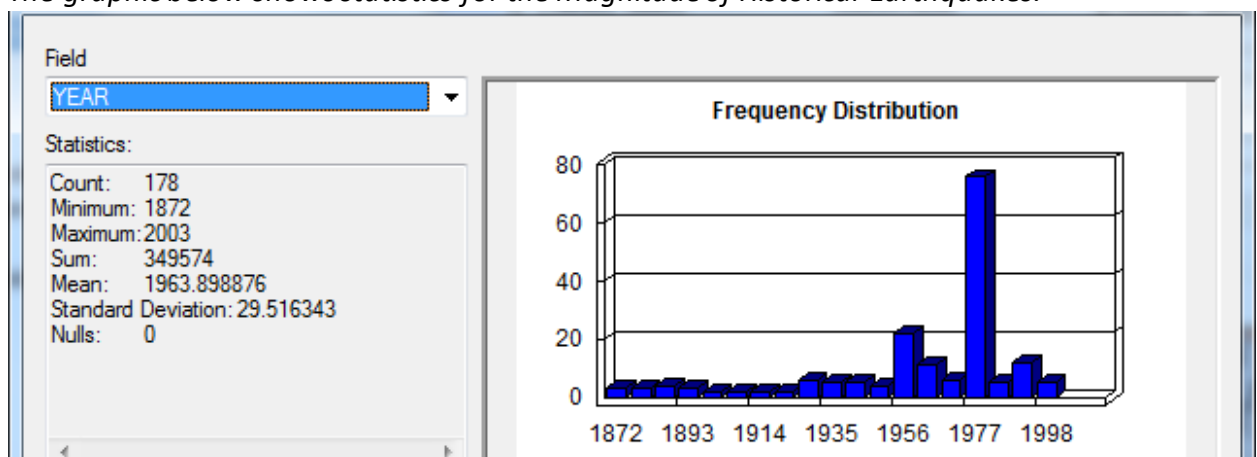
Calculate a total yearly count and averages for Historical Earthquakes.

You can easily get descriptive statistics about your data by opening the attribute table, right clicking a column heading... and choosing Statistics. Statistics give us measures of central tendency and dispersion such as the count, mean, median, sum and standard deviation

- From the attribute table, *right-click* the column heading **Year**, and then choose **Statistics**.



The graphic below shows statistics for the Magnitude of Historical Earthquakes.



QUESTION– Are there any trends? _____

QUESTION– What year did the most earthquakes occur? _____

QUESTION– What is the total amount of earthquakes in this area for all years? _____


QUESTION– Is there a trend such as an increase or decrease of earthquakes over time? _____

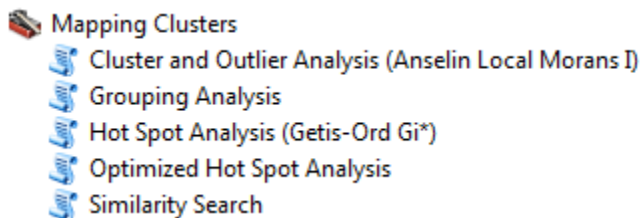
2. Close the **Statistics** and **Attribute** windows.

Part 2 – Locate the Mapping Clusters Toolset

The Mapping Clusters tools perform cluster analysis to allow visualization of cluster locations and their extent.

Launch ArcToolbox and expand Spatial Statistics Tools.

1. Click the **ArcToolbox** icon on the *Standard Toolbar* or **from ArcMap's Main Menu**, select **Geoprocessing | ArcToolbox**. 
2. Scroll down to **Spatial Statistics** tools and click the **plus (+) sign** to expand the **Mapping Clusters** toolset.



Part 3 – Explore the Spatial Distribution of the data

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.

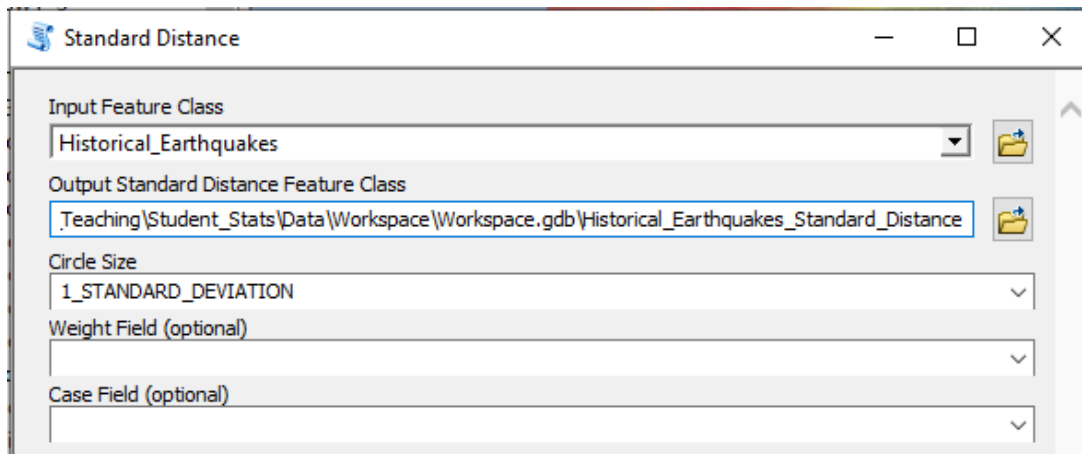
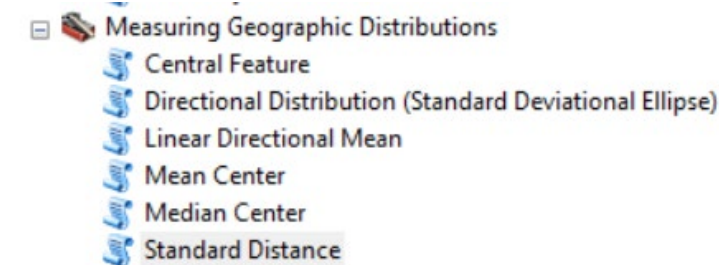
1. From the **Geoprocessing** menu, choose **Environments**.
2. Set both the *Current Workspace* and *Scratch Workspace* to **...\Data\Workspace\Workspace.gdb**
3. Expand the *Output Coordinates* section. Choose **Same as Input** for the *Output Coordinate System*.

4. *Processing Extent: Same as layer Region6_Boudary.*
5. Click **OK** to accept the changes and then **Save** the map. Now that we have the Geoprocessing Environments set, we can begin our analysis.

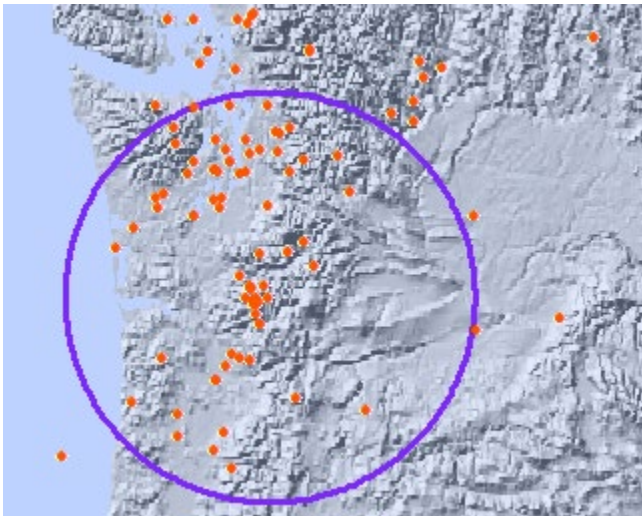
Use the Standard Distance tool to help identify trends in earthquakes.

The Standard Distance tool will show us the distribution of the data. We used this tool earlier to determine the distribution of BigFoot sightings.

1. *Double-click* to open the **Standard Distance** tool under the Measuring Geographic Distributions toolset.
2. In the Standard Distance enter the following parameters: For *Input Feature Class*: select **Historical_Earthquakes**.
3. Include the words **Standard Distance** in the output file name.
4. Select **1 Standard Deviation** for *Circle Size*.

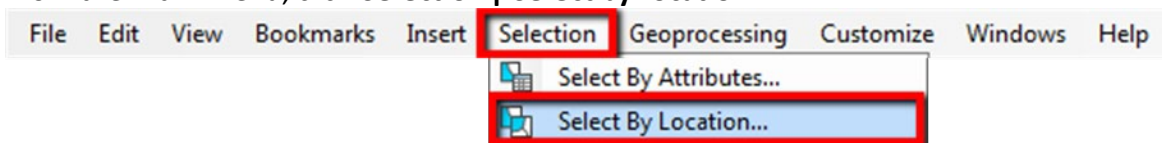


5. Click **OK** to run the tool.
6. The output will look like the screenshot below.

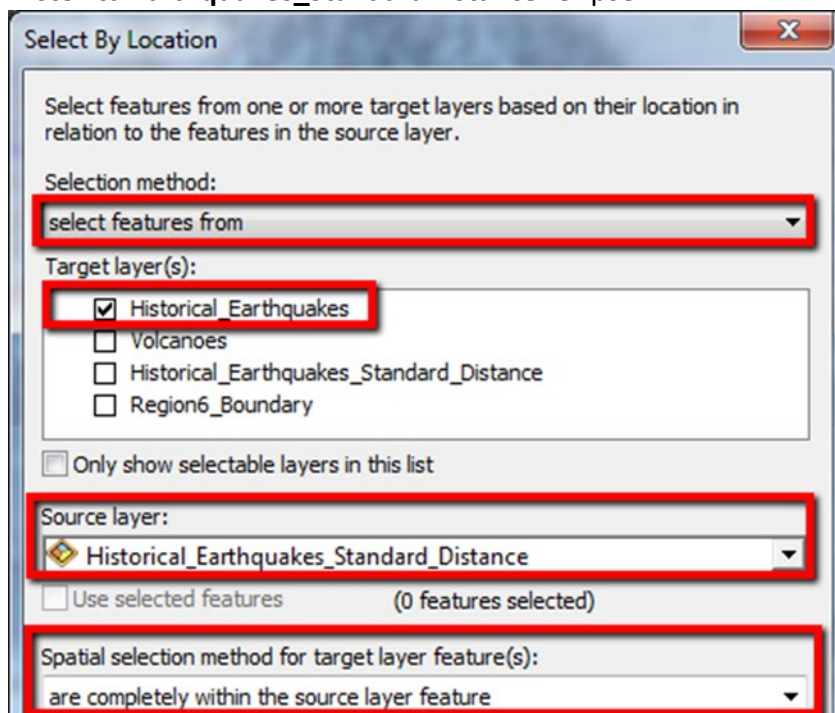


In the standard distance ellipse we can see that a majority of earthquakes occurred within a specific region which indicates possible clustering. Let's look at statistics for those earthquakes. But first we must select the earthquakes that fall within the standard distance ellipse.

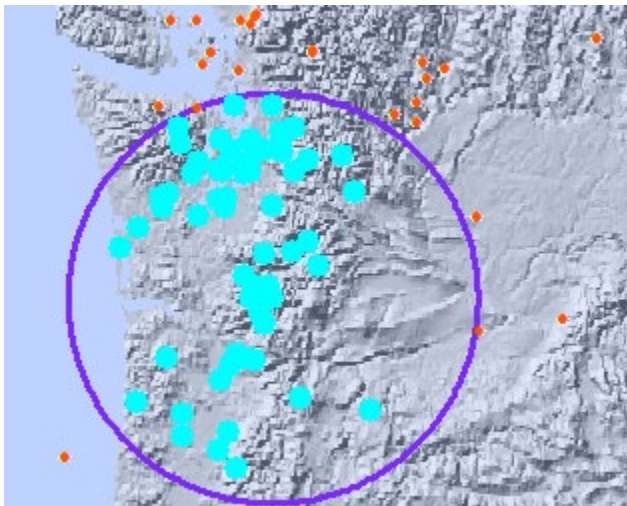
- From the **Main** menu, click **Selection | Select by Location**.



- Select features from the **Historical_Earthquakes** layer that are **completely within the HistoricalEarthquakes_StandardDistance** ellipse.



9. Click **Apply** | **OK**. The earthquakes existing within the boundaries of the ellipse are selected.

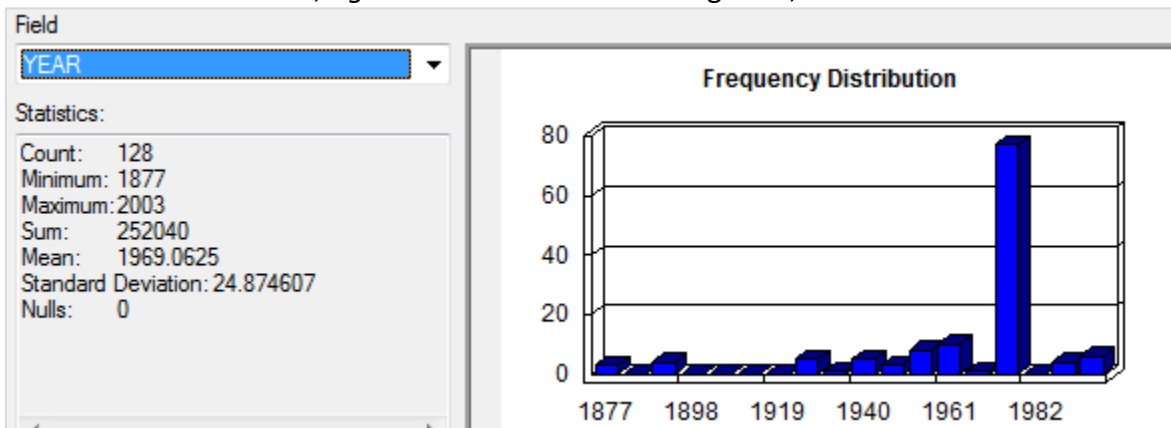


10. Open the **Attribute Table** for *Historical Earthquakes*. 128 out of the 178 Earthquakes are selected.

Table							
Historical_Earthquakes							
	MAG	MMI	LOCATION	YEAR	MONTH	DAY	HOUR
	7	9	Near Lake Chelan, Washington	1872	12	15	5
	6.7	8	Near Crescent City, California	1873	11	23	5
	5	7	North of Portland, Oregon (in Washington)	1877	10	12	21

(128 out of 178 Selected)

11. From the attribute table, *right click* the column heading **Year**, and select **Statistics**.



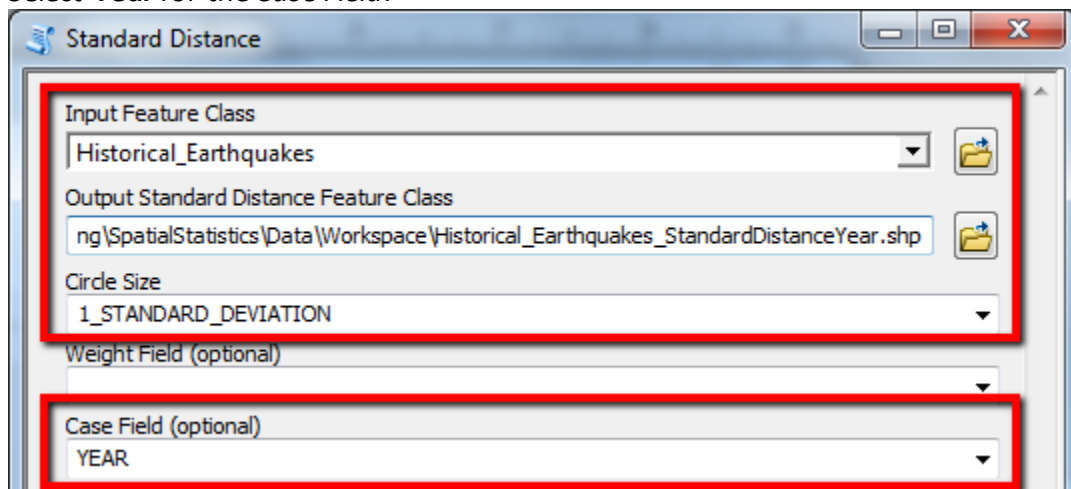
INTERPRETATION: Of the earthquakes that fell within our Standard Distance ellipse more than half occurred during 1980. Maybe we should look a little closer at the earthquakes that occurred during that year.

12. Close the *Statistics* window, **clear** the selected features.

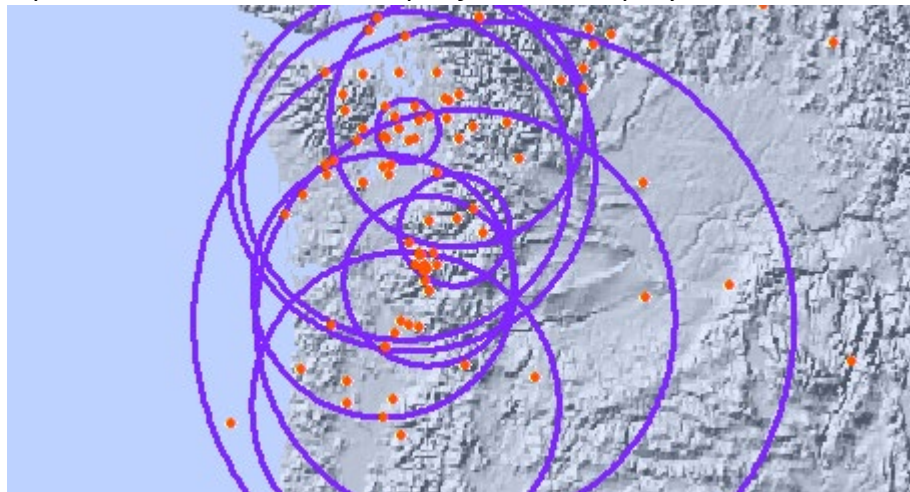
Use the Standard Distance tool to help identify trends in earthquakes based on the year they occurred.

Let's run the standard distance tool again only this time we will examine the annual spatial distribution of the earthquakes by adding YEAR as the optional Case Field. This will create separate distance ellipses for each unique year.

1. From the *Measuring Geographic Distributions* toolset, open the **Standard Distance** tool.
2. Enter the following parameters: For *Input Feature Class*: select **Historical_Earthquakes**. Include the words **Standard Distance Year** in the output file name.
3. Select **1 Standard Deviation** for *Circle Size*.
4. Select **Year** for the *Case Field*.

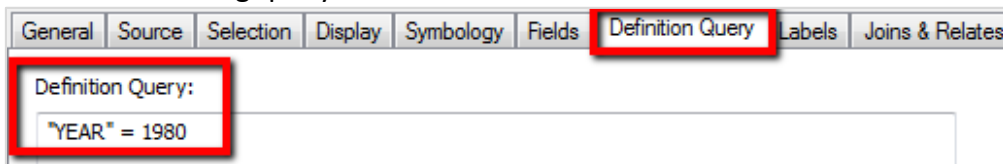


5. Click **OK** to run the tool. The tool runs and a layer is added to the data view displaying a separate Standard Distance Ellipse for each unique year in the dataset.



Remember, we are only interested in the standard distance ellipse for earthquakes that occurred during the year 1980. We will set a definition query in the layer's properties to only display the ellipse for that year.

6. Open the **Layer Properties** for **HistoricalEarthquakes_StandardDistanceYear**. *Right-click the layer, then select Properties.*
7. Select the **Definition Query** tab.
8. Click the **Query Builder** button.
9. Enter the following query: **"YEAR" = 1980**

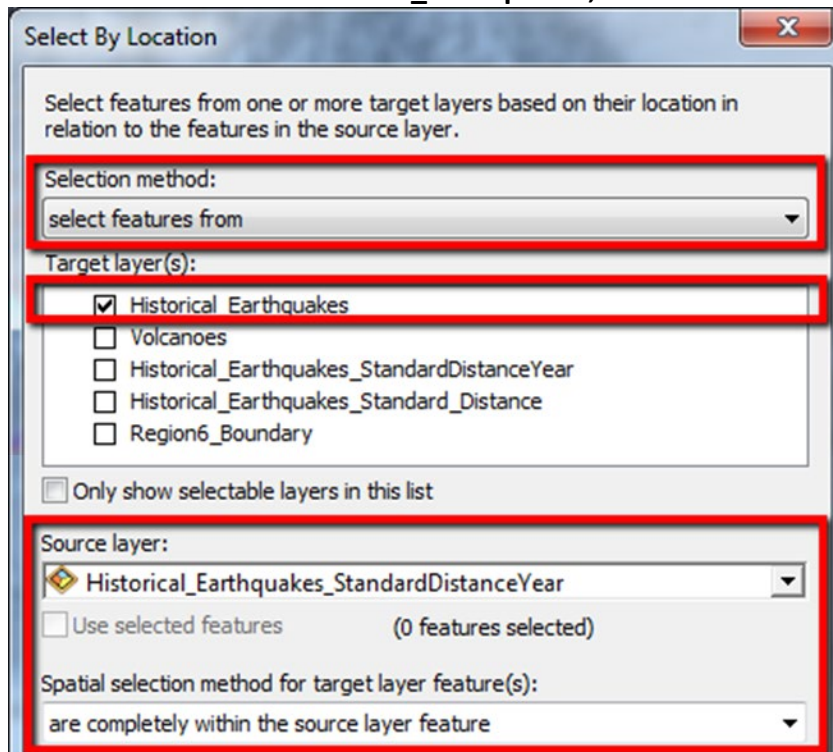


10. Click **OK** then click **OK** again in the Layer Properties dialog to accept the changes.
Change the symbology of the layer if desired.

Evaluating the attributes for earthquakes occurring within the 1980 Standard Distance ellipse, may help us identify a possible cause of the increase of earthquakes during that year.

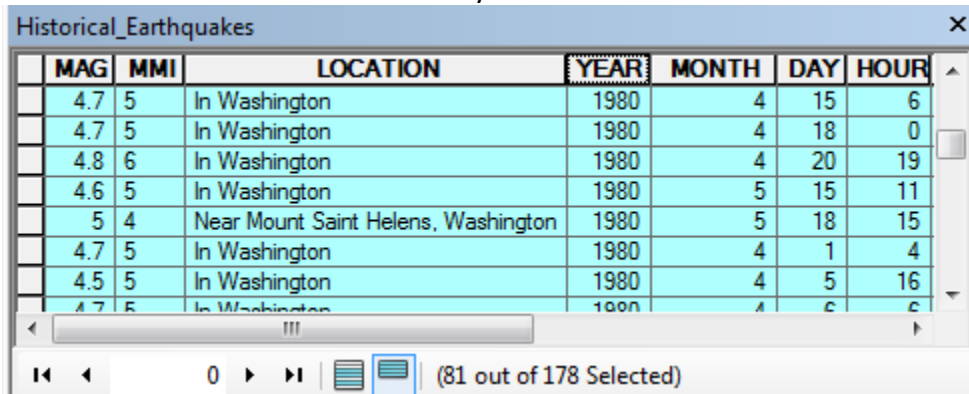
Explore the attribute data of earthquakes occurring in 1980.

1. Open the **Select by Location** tool.
2. Select features from: **Historical_Earthquakes**; where: **"YEAR" = 1980**.



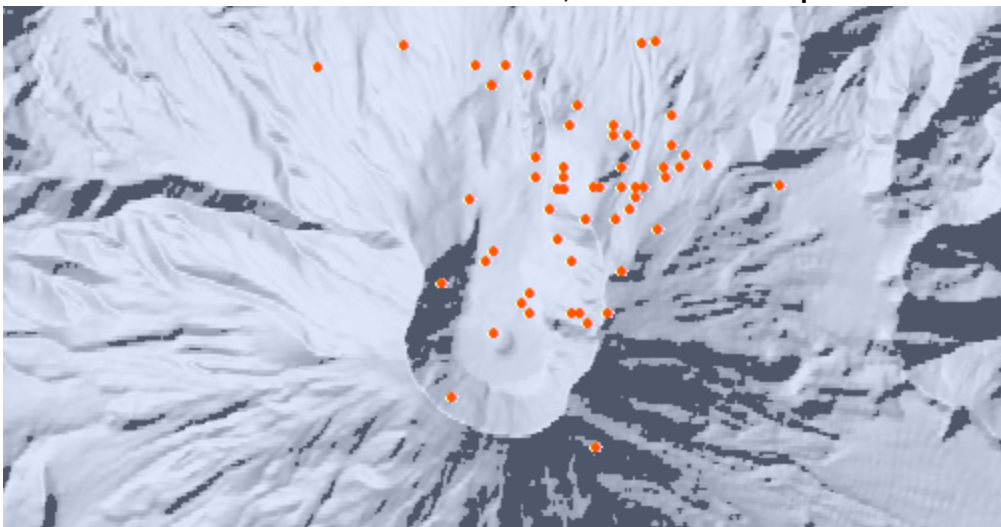
3. Click **Apply | OK**.

- Now let's look at the attributes of the selected earthquake. From the *Table of Contents*, right-click **Historical_Earthquakes** to open the **Attribute Table**.
- Click the **Show Selected Records** only button near the bottom of the window.



MAG	MMI	LOCATION	YEAR	MONTH	DAY	HOUR
4.7	5	In Washington	1980	4	15	6
4.7	5	In Washington	1980	4	18	0
4.8	6	In Washington	1980	4	20	19
4.6	5	In Washington	1980	5	15	11
5	4	Near Mount Saint Helens, Washington	1980	5	18	15
4.7	5	In Washington	1980	4	1	4
4.5	5	In Washington	1980	4	5	16
4.7	5	In Washington	1980	4	6	6

- Clear the selected features and close the attribute table.
- Zoom to the *Mount Saint Helens Bookmark*, click **Bookmarks | Mount Saint Helens**.



As you can see, in May of 1980 Mount Saint Helens erupted. This could be a major factor in the increase of earthquakes that year. We can verify this assumption with further analysis.

NOTE: 62 of the earthquakes that occurred in 1980 are within two miles of the volcano's center.

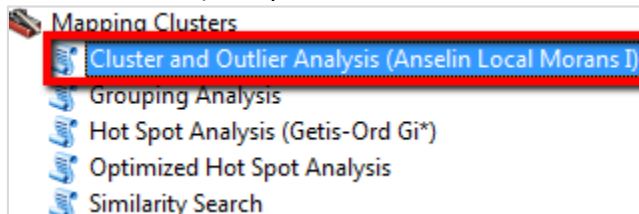
- Zoom to the extent of the **Region6 Boundary**.

Part 4 – Identify if the spatial clustering has statistical significance

We have already identified clustering amongst some of the earthquakes, but is there clustering of earthquakes based on the year they occurred or their magnitudes? And is the clustering statistically significant? These questions can be answered with the Cluster and Outlier Analysis tool.


Use the Cluster and Outlier Analysis (Anselin Local Moran's I) tool to identify statistically significant spatial clusters of earthquakes within the same Year.

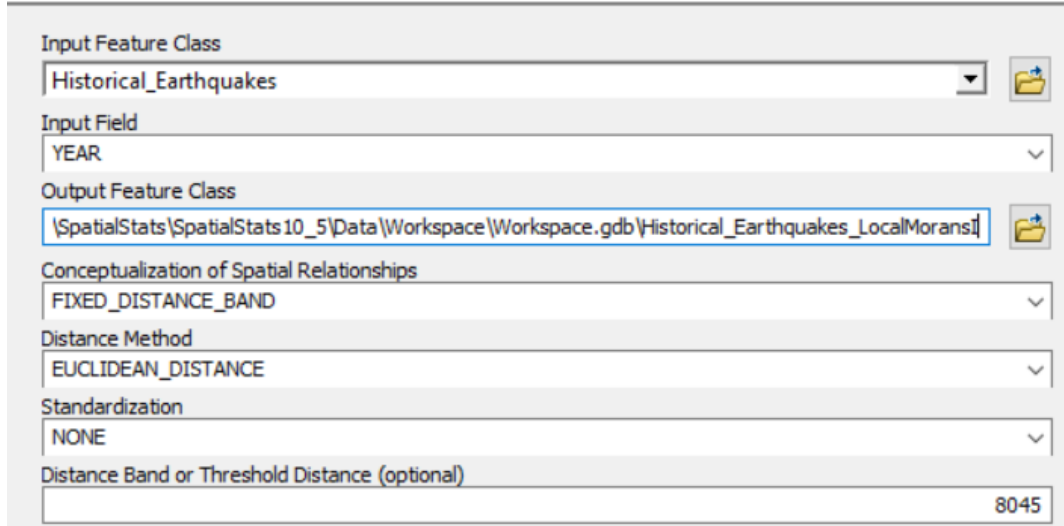
1. From the Mapping Clusters toolset, double click Cluster and Outlier Analysis (Anselin Local Moran's I) to open the tool.



Given a set of weighted features, the Cluster and Outlier Analysis tool identifies statistically significant hot spots, cold spots, and spatial outliers using the Anselin Local Moran's statistic.

2. For *Input Feature Class*, select **Historical_Earthquakes**.
3. For *Input Field*, select **YEAR**.
4. Include the words **Local Moran's I** in the *Output Feature Class Name*.
5. Choose **FIXED_DISTANCE_BAND** for *Conceptualization of Spatial Relationships*.
6. **EUCLIDEAN_DISTANCE** for *Distance Method*.
7. And **8045** for *Distance Band or Threshold Distance*. (Note: our data is in meters; 1 mile = 1609 meters, we want a distance band of 5 miles).


 Cluster and Outlier Analysis (Anselin Local Moran's I)




The screenshot shows the tool's configuration interface with the following settings:

- Input Feature Class:** Historical_Earthquakes
- Input Field:** YEAR
- Output Feature Class:** \\SpatialStats\\SpatialStats10_5\\Data\\Workspace\\Workspace.gdb\\Historical_Earthquakes_LocalMoransI
- Conceptualization of Spatial Relationships:** FIXED_DISTANCE_BAND
- Distance Method:** EUCLIDEAN_DISTANCE
- Standardization:** NONE
- Distance Band or Threshold Distance (optional):** 8045

8. Click **OK** to run the tool.
9. Open **Geoprocessing | Results** to view the two "WARNING" Messages.

 WARNING 000846:

 WARNING 000847:

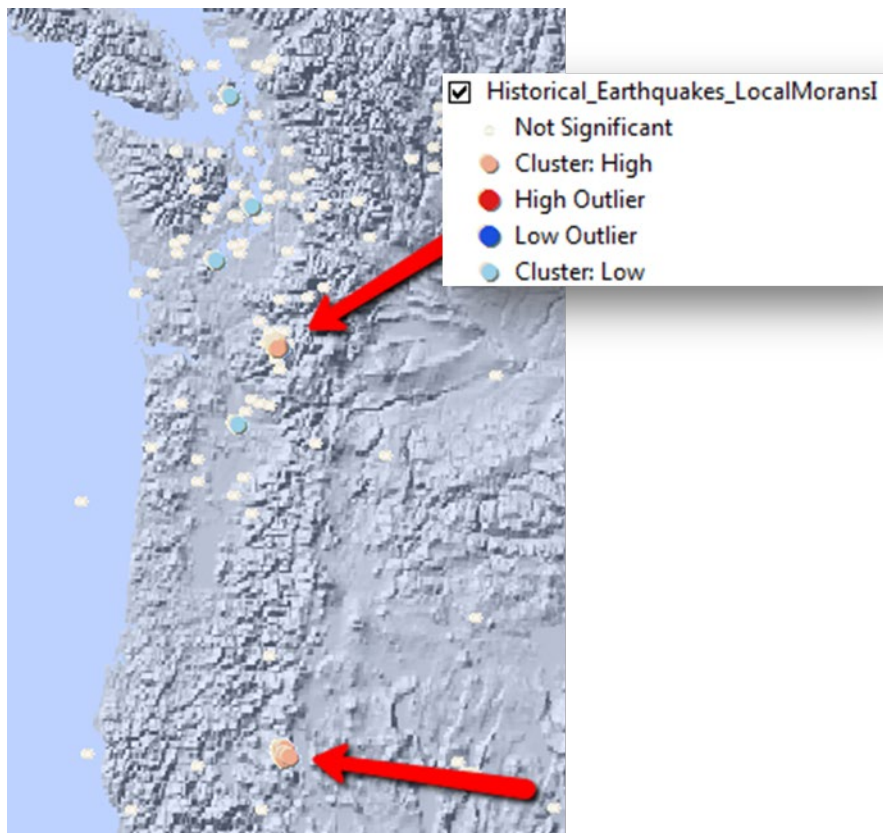
The warning tells us that some of the features in the dataset have no neighboring features within the given specified distance threshold.

10. Close the **Results** window and focus on the output **LocalMoransI** feature class that was added to ArcMap.

INTERPRETATION: The output feature class has the following attributes for each feature in the input feature class: local Moran's I index, z-score, p-value, and cluster/outlier type (COType).

The output field, COType, distinguishes between a statistically significant cluster of high values (HH), a cluster of low values (LL), an outlier in which a high value is surrounded primarily by low values (HL), or an outlier in which a low value is surrounded primarily by high values (LH). Statistical significance is set at the 95 percent confidence level for this tool.

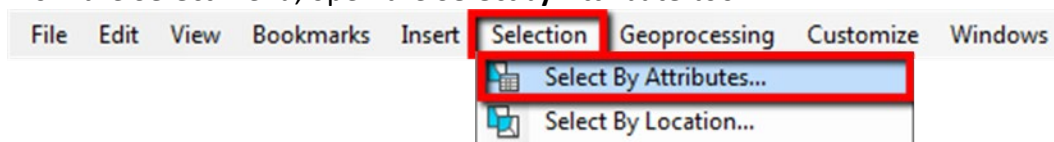
Our results show statistically significant clustering in two locations of our study area.



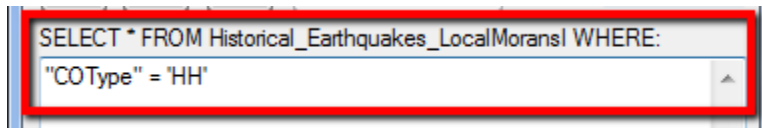
Select the Historical Earthquakes with statistically significant clustering and review their attributes.

Again, evaluating the attributes for earthquakes occurring in locations with statistical significance may help us identify trends in the data.

1. From the **Select** menu, open the **Select by Attribute** tool.



2. Select features from **HistoricalEarthquakes_LocalMoransI** where: "COType" = HH.



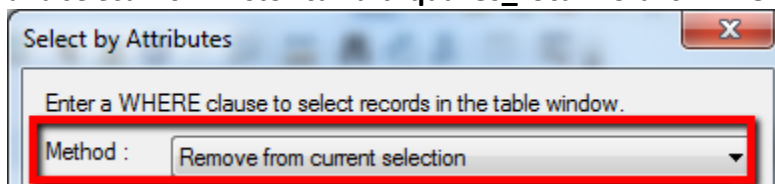
3. Click **Apply | OK**.
4. Open the **Attribute Table** for **HistoricalEarthquakes_LocalMoransi**.
5. From the *Attribute Table*, click the **Show Selected Features** button. *There are 68 features selected that have statistically significant clustering.*

FID	Shape	SOURCE_ID	YEAR	LMiIndex	LMiZScore	LMiPValue	COType
4	Point	5	1993	3.866408	1.980638	0.047632	HH
5	Point	6	1993	3.866408	1.980638	0.047632	HH
6	Point	7	1993	5.832827	2.453519	0.014147	HH
7	Point	8	1994	3.999269	2.048303	0.04053	HH
70	Point	71	1980	17.753819	2.87897	0.00399	HH
77	Point	78	1980	17.753819	2.87897	0.00399	HH
78	Point	79	1980	18.049716	2.914971	0.003557	HH
79	Point	80	1980	17.753819	2.87897	0.00399	HH
80	Point	81	1980	17.753819	2.87897	0.00399	HH
81	Point	82	1980	17.753819	2.87897	0.00399	HH
82	Point	83	1980	17.753819	2.87897	0.00399	HH
83	Point	84	1980	17.753819	2.87897	0.00399	HH
84	Point	85	1980	17.753819	2.87897	0.00399	HH
85	Point	86	1980	17.753819	2.87897	0.00399	HH

6. Locate the **YEAR** field.

There are three different years having statistical significance; 6 in 1993, 1 in 1994, and 61 in 1980. We have already determined the probable cause of earthquakes for 1980, so we will eliminate those and look closer at the others.

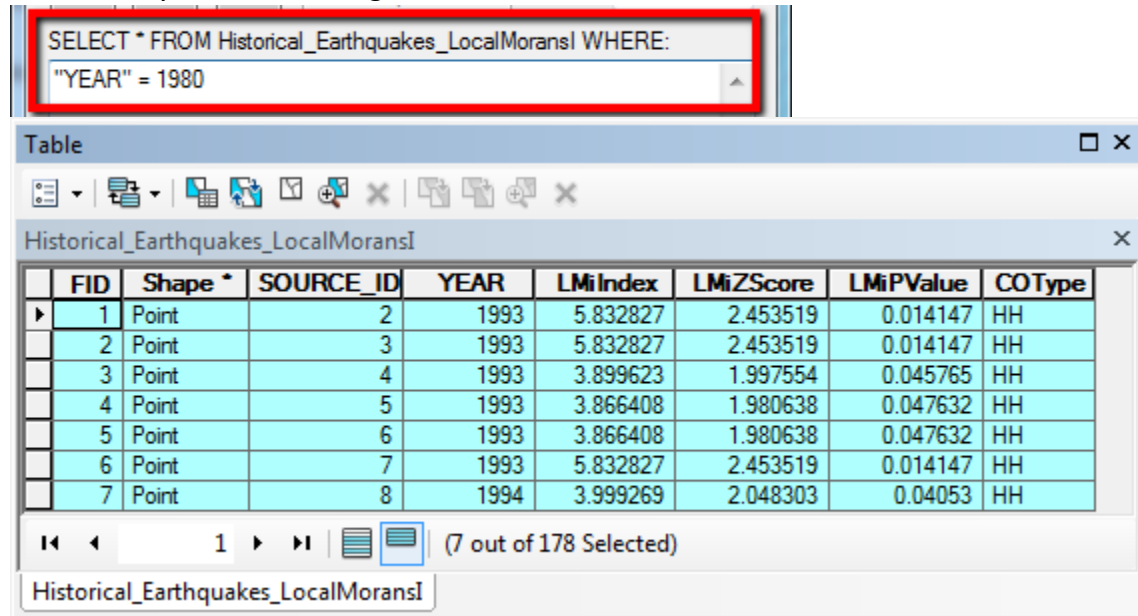
7. From the *Select by Attribute* tool, update *Method* to **Remove from current selection**; and select from **HistoricalEarthquakes_LocalMoransi** where: **"YEAR" = 1980**.



The earthquakes occurring between the years 1993-1994 are all clustered above the southern boundary of our study area. The increase of earthquakes during 1980 appears to be caused by volcanic activity; can we make the same assumption with clustering of earthquakes during 1993-1994?

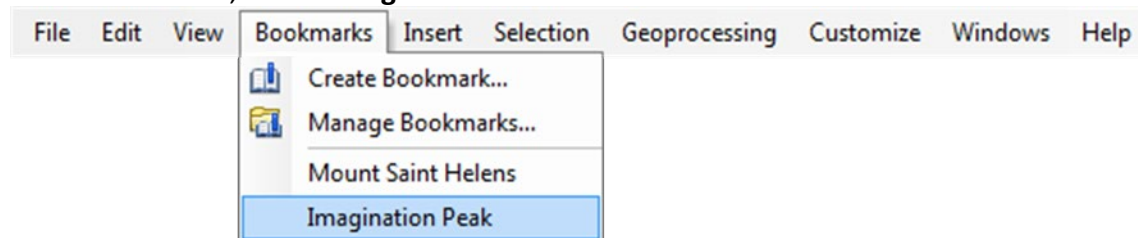
8. Click **Apply & OK**. Any feature with the Year 1980 is removed from the selection. Close

the *Select by Attribute* dialog.

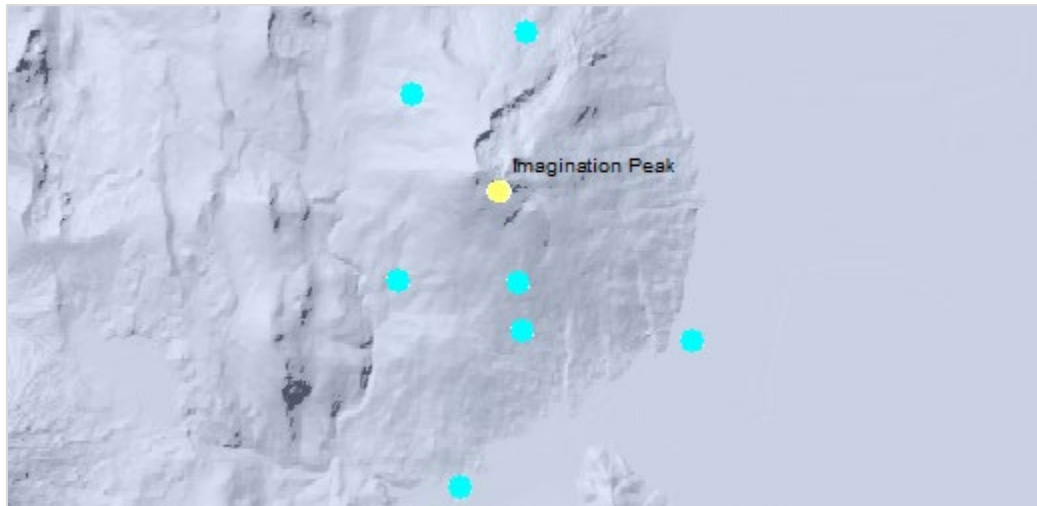


9. From the *Table of Contents*, turn on the **Volcanoes** layer.

10. From **Bookmarks**, select **Imagination Peak**.



Because the date of an eruption from Imagination Peak is unknown, we cannot confirm volcanic activity as an explanation for the increase in earthquakes during these years, however we do know there is less than a 5% chance that the clustering of the earthquakes in this area occurred by random chance.



D1	Last known eruption 1964 or later
D2	Last known eruption 1900-1963
D3	Last known eruption 1800-1899
D4	Last known eruption 1700-1799
D5	Last known eruption 1500-1699
D6	Last known eruption A.D. 1-1499
D7	Last known eruption B.C. (Holocene)
U	Undated, but probable Holocene eruption
Q	Quaternary eruption(s) with the only known activity being hydrothermal
?	Uncertain Holocene eruption

Table						
Volcanoes						
	NAME	ELEVATION	TIME_CODE	TYPE	STATUS	KNOWN_ER
▶	Imagination Peak	1986	?	Pyroclastic cones	Holocene?	0
1 (1 out of 164 Selected)						
Volcanoes						

Part 5 – Utilize the Optimized Hot Spot tool to identify high (hot) & low (cold) value clustering of earthquake magnitudes.

This tool identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots). It automatically aggregates incident data, identifies an appropriate scale of analysis, and corrects for both multiple testing and spatial dependence. This tool interrogates your data in order to determine

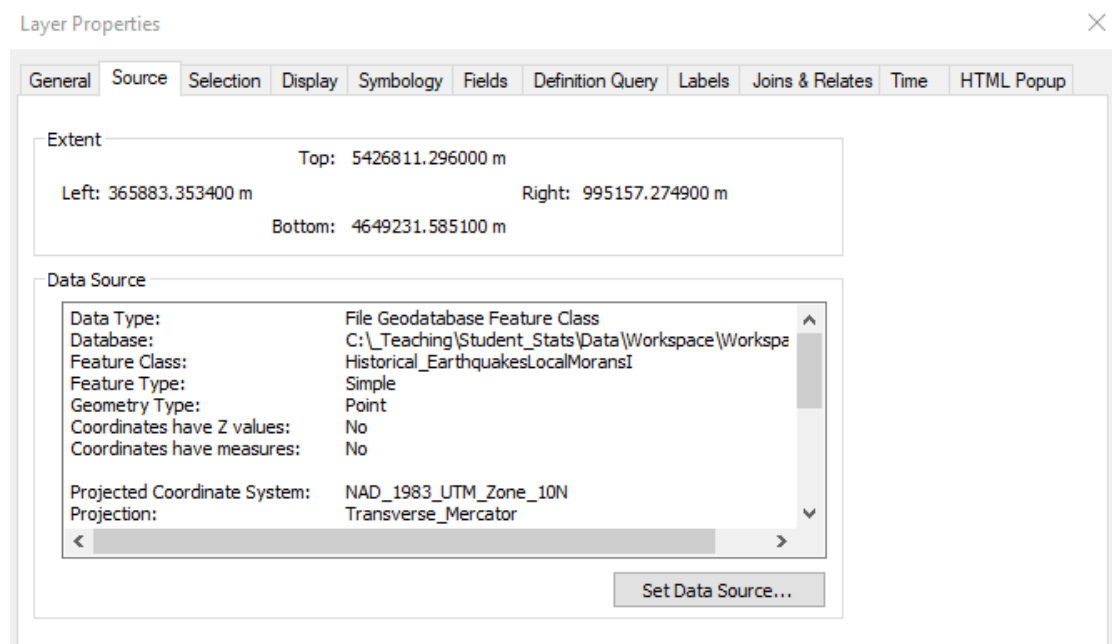
settings that will produce optimal hot spot analysis results. If you want full control over these settings, use the Hot Spot Analysis tool instead.

Make sure your data is projected correctly

Make sure your input and output have the same projected coordinate not geographic coordinate systems. Otherwise distances are computed using chordal measurements. Finding a projection that preserves either distance or area is preferred. You can learn more about projections [here](#).

1. Double click **Historical_Earthquakes** and click on the **Source** tab in the Layer Properties window.
2. Ensure that the layer is set to **NAD_1983_UTM_Zone_10N**.

For this example we have the correct projected coordinate system, if they weren't you would have to re-project it using the Project tool (data management).



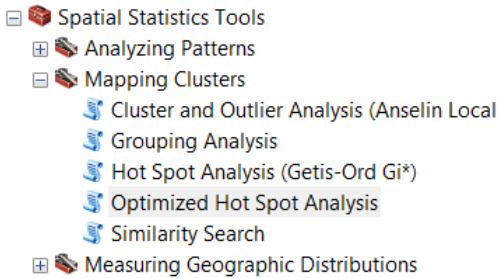
3. **Close** the Layer Properties dialog.

Use the Optimized Hot Spot tool to identify statistically significant spatial clustering or sparsity of point locations.

With point data you will sometimes be interested in analyzing data values associated with each point feature and will consequently provide an Analysis Field. In other cases you will only be interested in evaluating the spatial pattern (clustering) of the point locations or point incidents. The decision to provide an Analysis Field or not will depend on the question you are asking.

When the Input Features you provide represent incident data (when you don't provide an Analysis Field), the tool will aggregate the incidents and the incident counts will serve as the values to be analyzed.

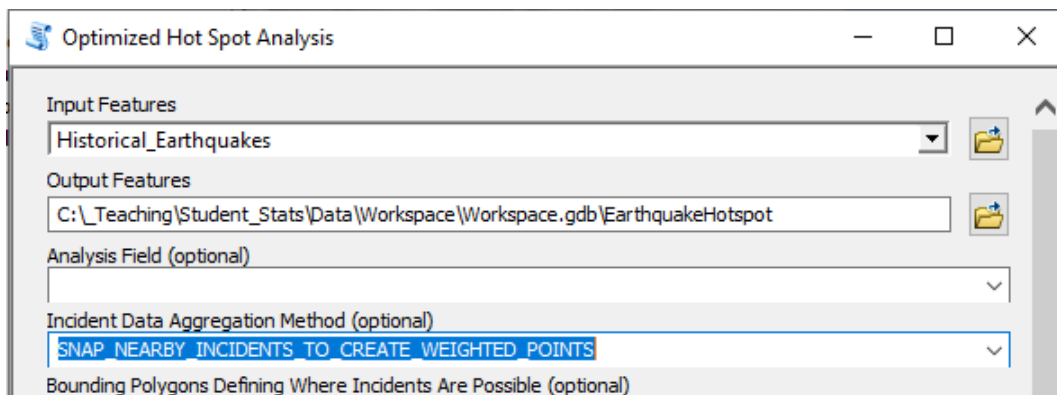
1. Launch the **Optimized Hot Spot Analysis** tool from the Mapping Clusters toolset.



2. For input Features select **Historical_Earthquakes**
3. Name the **Output** ../Workspace/Workspace.gdb/**Earthquake_HotSpot**
For this part of the class we will save the outputs as Feature Classes within a geodatabase.
4. Leave the **Analysis Field** blank.

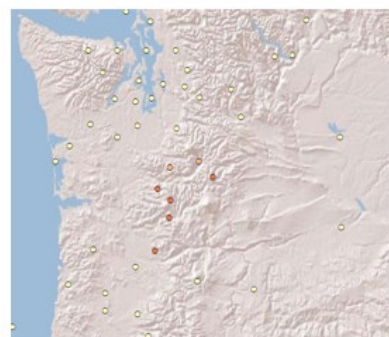
Analyzing point features when there is **no Analysis Field** allows you to identify where point clustering is unusually (statistically significant) intense or sparse. This type of analysis answers questions like: Where are there many points? Where are there very few points? When you **do not** provide an Analysis Field the tool will aggregate your points in order to obtain point counts to use as an analysis field.

5. For Incident Data Aggregation Method insert
SNAP_NEARBY_INCIDENTS_TO_CREATE_WEIGHTED_POINTS



6. Click **OK** to run tool creating a new feature class.

- ☒ EarthquakeHotSpots3
 - Gi_Bin
 - Cold Spot - 99% Confidence
 - Cold Spot - 95% Confidence
 - Cold Spot - 90% Confidence
 - Not Significant
 - Hot Spot - 90% Confidence
 - Hot Spot - 95% Confidence
 - Hot Spot - 99% Confidence



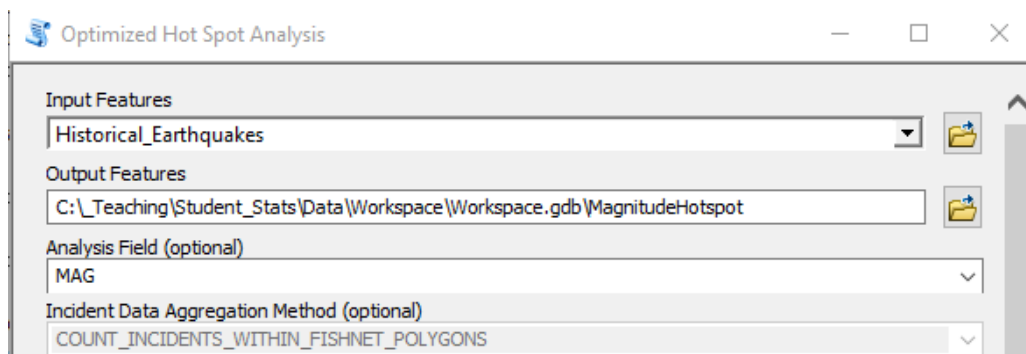
Looking at the results we can see some statistically significant clustering of historical earthquakes (hot spots) centered on Mt. Saint Helens. We also see many “not significant” values. No cold spots were identified.

Use the Optimized Hot Spot Analysis tool to identify statistically significant clustering of historical earthquake magnitudes high values (hot spot) and low values (cold spot).

Analyzing point features **with an Analysis Field** allows you to answer questions like: Where do high and low values cluster? The points are not aggregated when using the Analysis Field like they are in the previous step when your input features represented incident data.

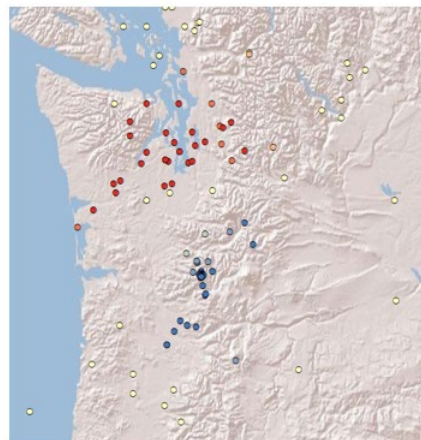
1. Double click to **open the Optimized Hot Spot Analysis Tool** and enter the fields as shown below.
2. For **Input Features** enter **Historical_Earthquakes** from the dropdown
3. Name the **Output Feature** ...\\Data\\Workspace\\Workspace.gdb**Magnitude_HotSpot**
4. For **Analysis Field** select **MAG for Magnitude**

Note: Incident Data Aggregation Method is grayed out because Analysis Field is being utilized.



5. Click **OK** to run the tool.
The result should look like the screengrab below.

- ☒ Magnitude_HotSpot
 - Gi_Bin
 - Cold Spot - 99% Confidence
 - Cold Spot - 95% Confidence
 - Cold Spot - 90% Confidence
 - Not Significant
 - Hot Spot - 90% Confidence
 - Hot Spot - 95% Confidence
 - Hot Spot - 99% Confidence



- Right click the Magnitude_Hotspot layer and select **Open Attribute Table** to examine the results.

Magnitude_HotSpot

	FID	Shape *	SOURCE_ID	MAG	GiZScore	GiPValue	Gi_Bin
▶	0	Point	1	4.6	1.760759	0.078279	0
■	1	Point	2	4.4	0.758721	0.448019	0
■	2	Point	3	6	0.758721	0.448019	0
■	3	Point	4	4.5	0.746907	0.45512	0
■	4	Point	5	6	0.758721	0.448019	0
■	5	Point	6	4.8	0.758721	0.448019	0
■	6	Point	7	5.4	0.758721	0.448019	0
■	7	Point	8	4	0.758721	0.448019	0
■	8	Point	9	4	0.175262	0.860873	0
■	9	Point	10	4.6	1.240513	0.311477	0

1 (0 out of 178 Selected)

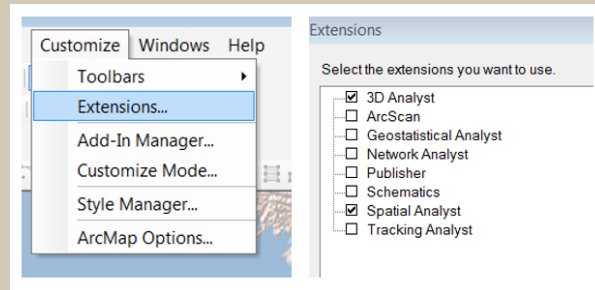
INTERPRETATION: The result of the Hot Spot Analysis tool is a new feature class where every feature in your dataset is symbolized based on whether it is part of a statistically significant hot spot, a statistically significant cold spot, or is not statistically significant. As expected there is clustering around Mt. Saint Helens again but this time as cold spots. Meaning that there is a clustering of lower magnitude earthquakes in this area compared to the rest of the data. Surrounding the Puget Sound we have a number of hot spots meaning that area has a clustering of earthquakes with a higher magnitude relative to the rest of the Pacific Northwest.

Visualize your results. Use your results to create an interpolated surface and prepare for presentation.

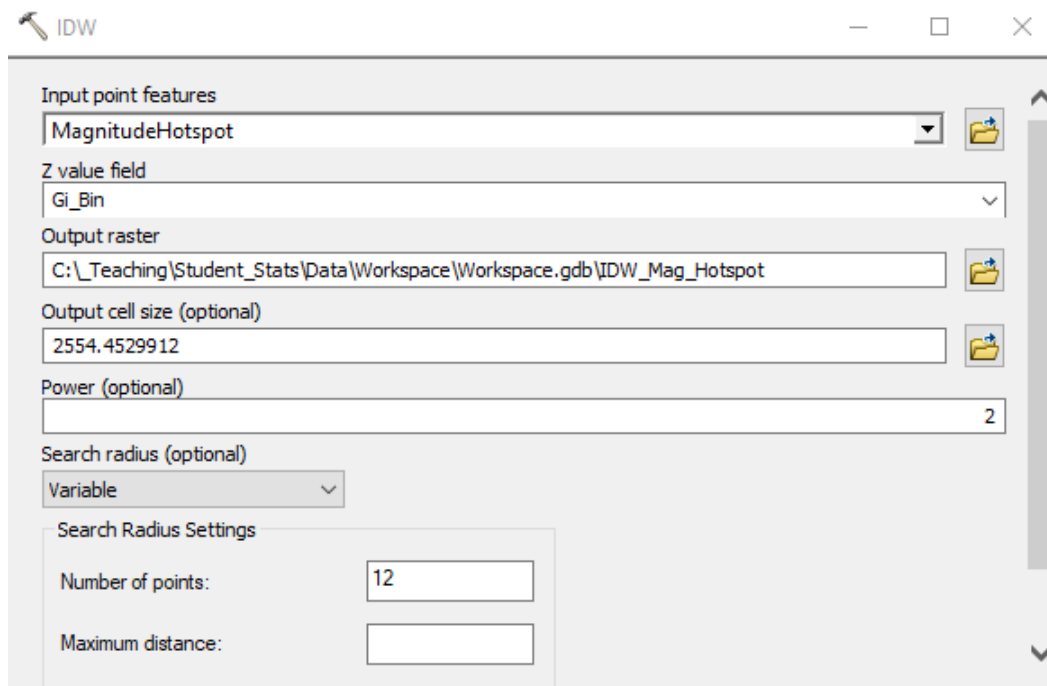
While you as the analyst understand that each location has a particular z-score and a particular p-value that determine the symbology and can guide decision-making, you may be required to share your analysis with a broader audience who may be expecting the results to look like the heat maps people have become accustomed to in the media. Thus, the next step is to create a continuous surface to visualize the results. Remember that this surface is for visualization purposes only, and the true statistical analysis happens feature by feature. Showing both the surface and the true results of the hot spot analysis at the same time is a great way to present both the statistical results and the more approachable visualization. The method that you will use in this tutorial, [IDW \(Spatial Analyst\)](#), interpolates a raster surface from points using an inverse distance weighted technique.

- Using the search tool search **IDW** and **open the IDW (Spatial Analyst) tool**.
Or open by expanding the *Spatial Analyst* toolbox, then expand the *Interpolation* tool set and select *IDW*.

Note: The Spatial Analyst extension is required to run IDW. To add the Spatial Analyst extension click on Customize from the main menu and select Extensions. In the Extensions window that pops up, check Spatial Analyst and click close.

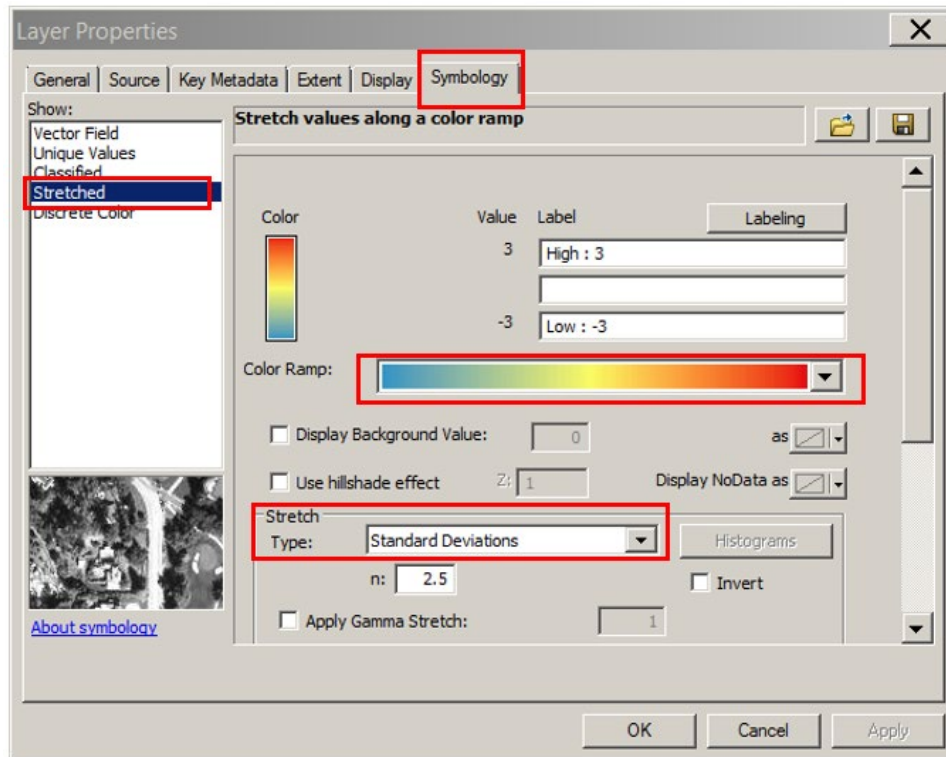


2. Fill out the IDW tool as shown below, for **Input point features** select **Magnitude_HotSpot** from the dropdown
3. For **Z value field** select **Gi_Bin**
4. For **Output raster** name enter **..\Workspace\Worspace.gdb\IDW_Mag_HotSpot**
5. **Output cell size** leave as **default** (it will self-populate when you add the value field)
6. Accept the rest as default and **click OK**



7. Double click the new **IDW_Mag_HotSpot** layer to **open the Layer Properties** window.
8. Select the **Symbology** tab and change the symbology as shown below, chose the **Stretched** option
9. Select the **ramp that is colored from blue to red** (called Prediction). *To match our Magnitude_HotSpot layer color scheme - high values are red and low are blue.*

10. For **Type** select **Standard Deviations** from the dropdown



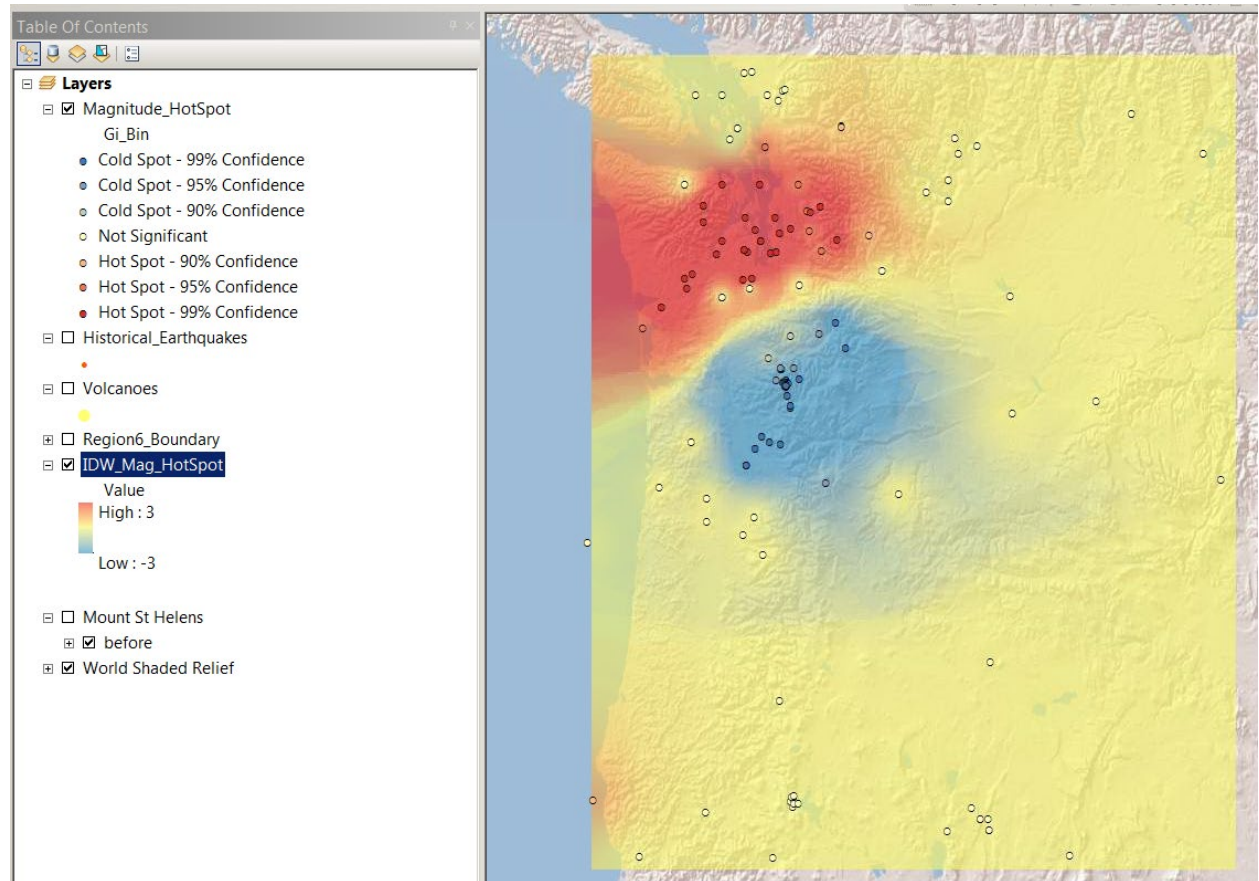
11. Click on the **Display** tab.

12. Enter **40% transparency**

13. Click **OK**

14. **Move IDW_Mag_HotSpot under Magnitude_HotSpot** and turn off all other layers besides the World Shaded Relief.

The Display should resemble the screen grab below.



Note: The best results from IDW are obtained when sampling is sufficiently dense with regard to the local variation you are attempting to simulate.

One application of spatial statistics and the Hot Spot Analysis tool is to help to resolve resource allocation problems. Where would you allocate additional earthquake emergency response resources?

End of Exercise