



United States Department of Agriculture

# Spatial Statistics

---



Geospatial Technology  
and Applications Center















Forest Service

# Overview

- Focus areas:
  - Basics of spatial statistics
    - Descriptive Statistics
    - Inferential Statistics
  - Spatial Statistics toolbox
  - Determining spatial relationships
  - Spatial pattern analysis

# Lesson Overview

## Measuring Geographic Distributions

- [-]  Spatial Statistics Tools
  - [+]  Analyzing Patterns
  - [+]  Mapping Clusters
  - [-]  Measuring Geographic Distributions
    -  Central Feature
    -  Directional Distribution (Standard Deviation Ellipse)
    -  Linear Directional Mean
    -  Mean Center
    -  Median Center
    -  Standard Distance
  - [+]  Modeling Spatial Relationships
  - [+]  Utilities

# Spatial Statistics Described

- Measuring spatial relationships
  - Summarize complex datasets
- Spatial Pattern Analysis
  - Identify patterns, trends, relationships
  - Locate potential geographic cause / effect relationships  
(*e.g., water pollution/wildlife illness*)

*Different from what you learned in STATS 101 because the results vary based on the physical locations of the data*



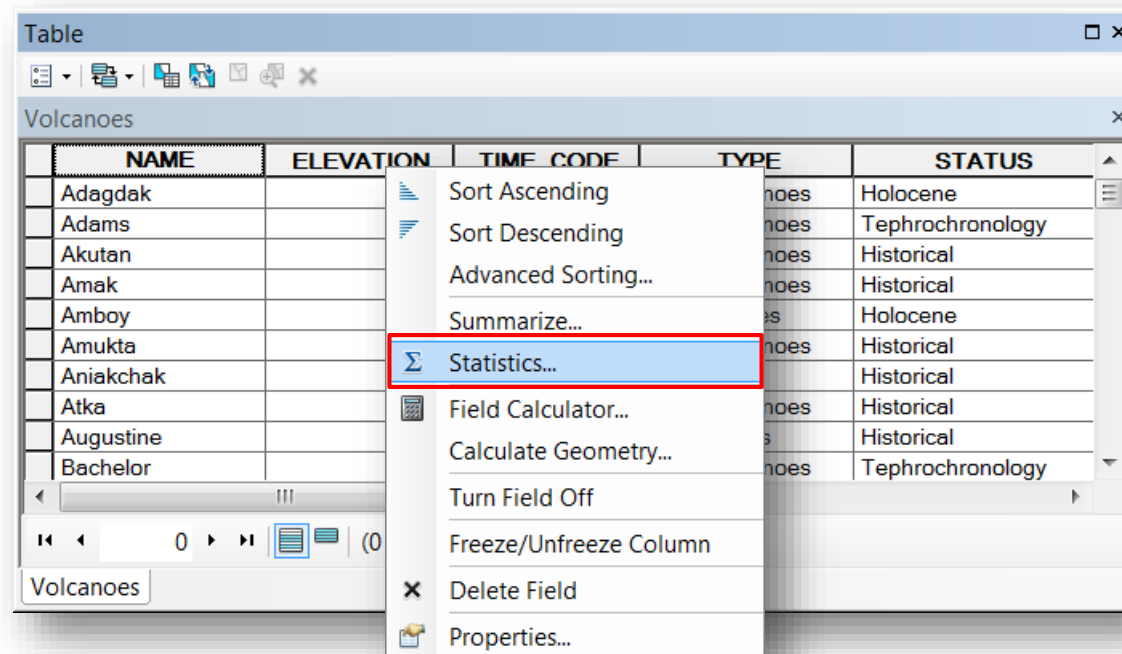
# Statistics Described

- Descriptive Statistics
  - measures of central tendency and dispersion
    - *count, mean, median, sum, standard deviation, null values*
- Inferential Statistics
  - descriptive stats + probability theory to generalize results of a sample to some larger group like
    - *hypothesis testing, confidence intervals, linear regression*



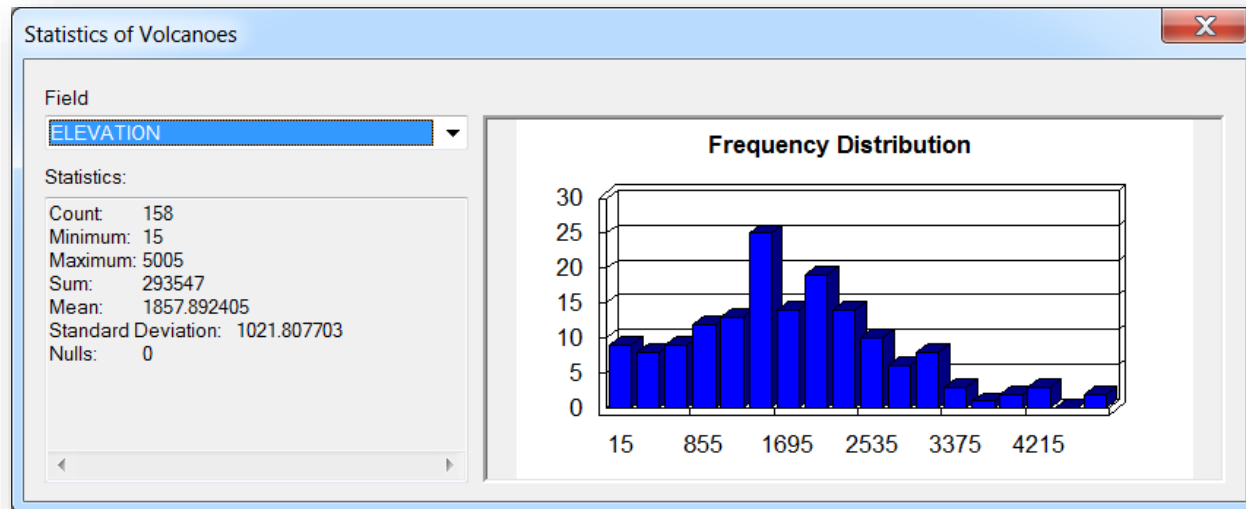
# Descriptive Statistics in ArcGIS

- Descriptive Statistics
  - measures of central tendency and dispersion
    - *count, mean, median, sum, standard deviation, null values*



# Descriptive Statistics in ArcGIS

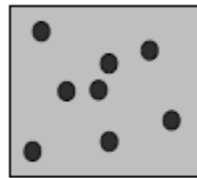
- Descriptive Statistics
  - measures of central tendency and dispersion
    - *count, mean, median, sum, standard deviation, null values*



# Inferential Statistics in ArcGIS

- Inferential Statistics

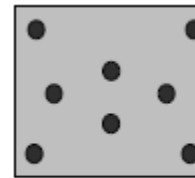
- Assumes values in each sample are independent of one another
  - Geographic data are often not independent but spatially autocorrelated



Random



Clustered



Dispersed



# Inferential Statistics in ArcGIS

- Inferential Statistics

- Assumes values in each sample are independent of one another
  - Geographic data are often not independent but spatially autocorrelated

## The First Law of Geography

*Everything is related to everything else,  
but near things are more related than distant things.*

*-Tobler*

**SPATIAL AUTOCORRELATION**



# Inferential Statistics in ArcGIS

- Inferential Statistics

- Statistics are measures calculated from samples which are estimates of population parameters
- Principal question with statistics: Are your results due to the sampling process or is there a real difference in the underlying population(s)?
  - Hypothesis testing
- Statistical significance does not always equate to scientific significance



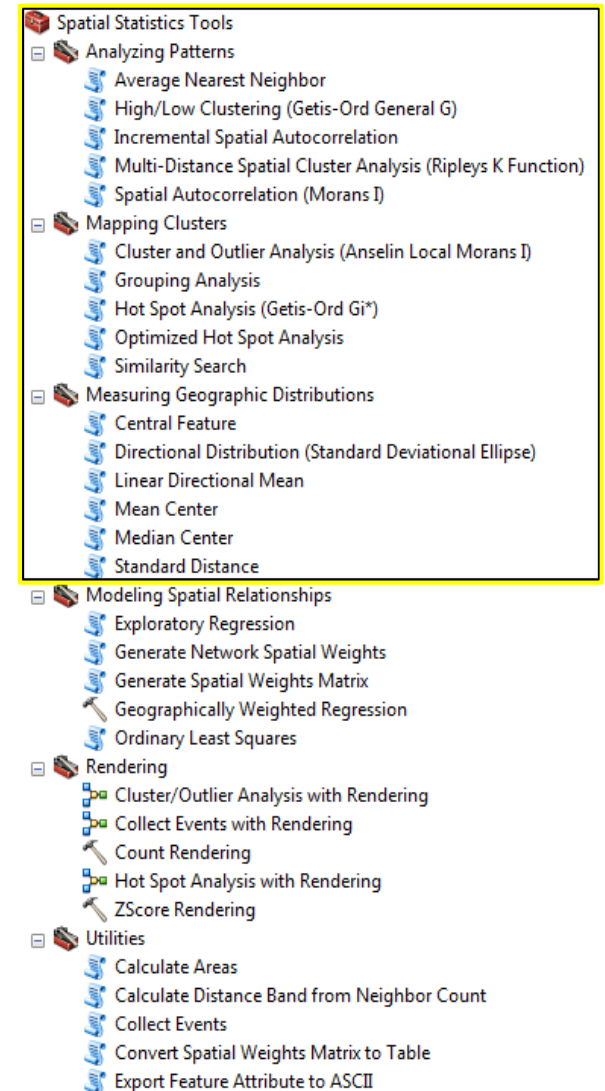
# Why Spatial Statistics?

- To help assess patterns, trends, and relationships
  - Better understand behaviors
  - Isolate geographic patterns
- Make decisions with higher levels of confidence
  - Statistically significant results
- Summarize your data distributions with a single numerical value



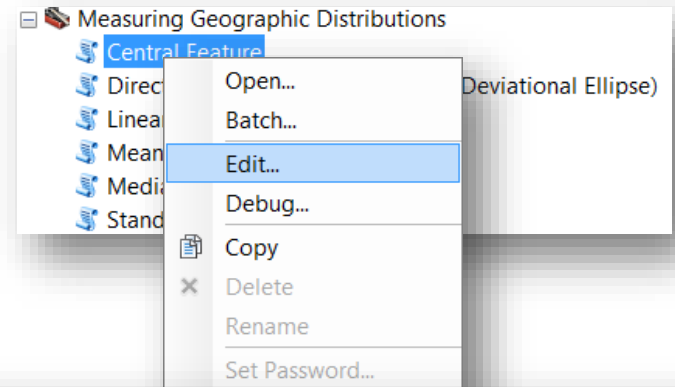
# Spatial Statistics in ArcGIS

- Spatial Statistics Tools
  - Included with ArcGIS Desktop
    - Not an extension
  - Available at all license levels
    - ArcView, ArcEditor and ArcInfo
  - Source code provided
    - Scripts
    - Models



# Spatial Statistics in ArcGIS

- Spatial Statistics Tools
  - Included with ArcGIS Desktop
    - Not an extension
  - Available at all license levels
    - ArcView, ArcEditor and ArcInfo
  - Source code provided
    - Scripts
    - Models



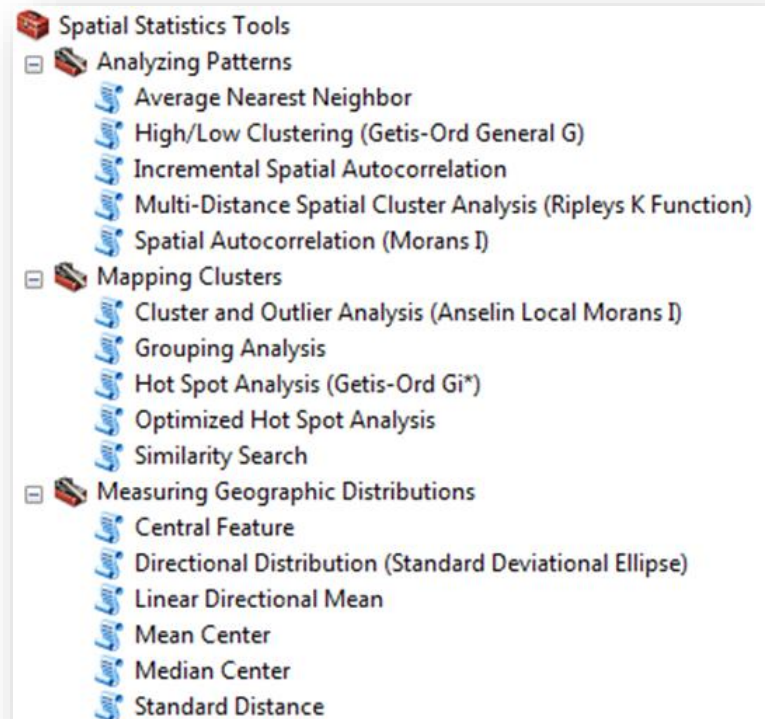
```
File Edit Format View Help
"""
Tool Name: Central Feature
Source Name: CentralFeature.py
Version: ArcGIS 10.1

This script will identify the most centrally located
feature in a point,
line, or polygon feature class (with weighting and
grouping optional).
"""

##### Imports #####
import os as OS
import collections as COLL
import numpy as NUM
```

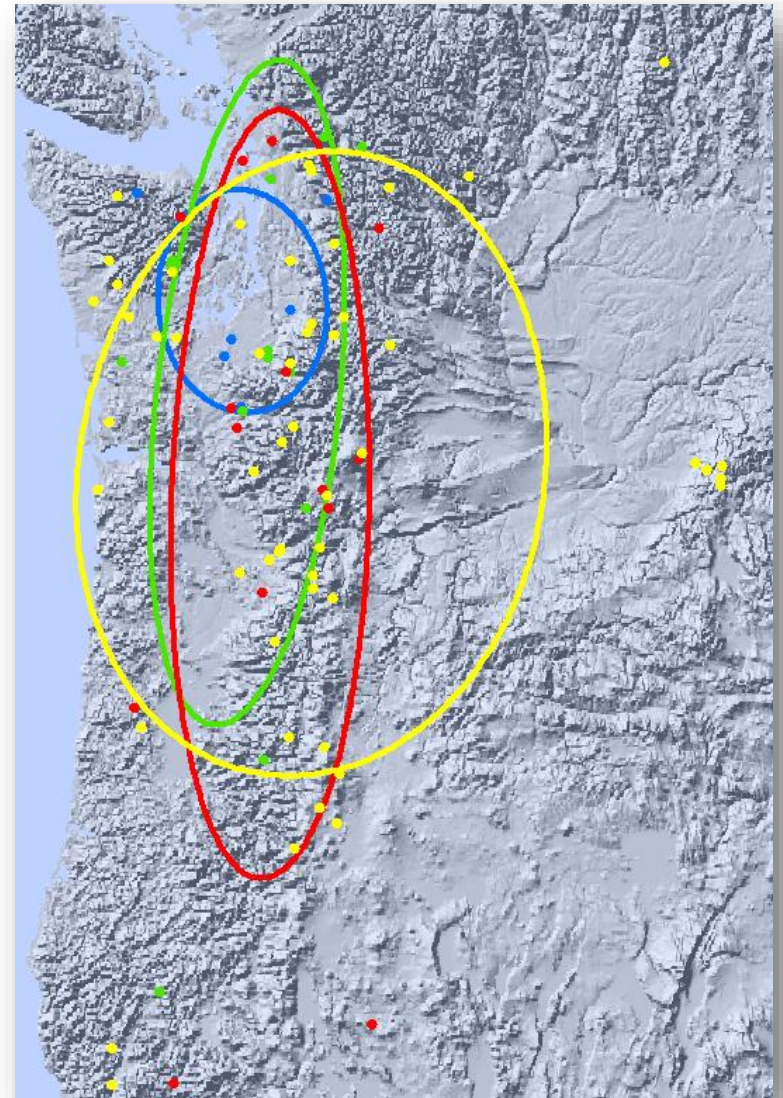
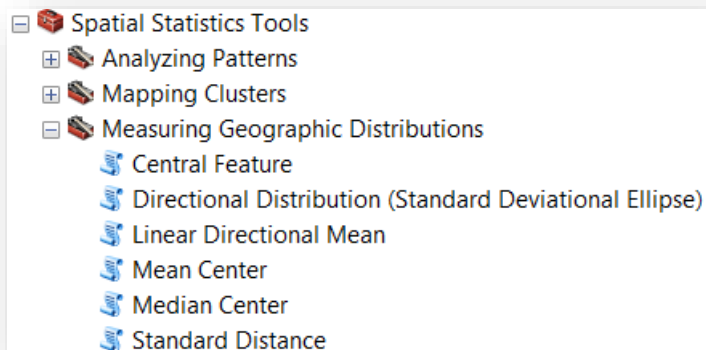
# Spatial Statistics Tools

- Spatial Statistics Tools
  - Measuring Geographic Distributions
  - Analyzing Patterns
  - Mapping Clusters



# Spatial Statistics Tools

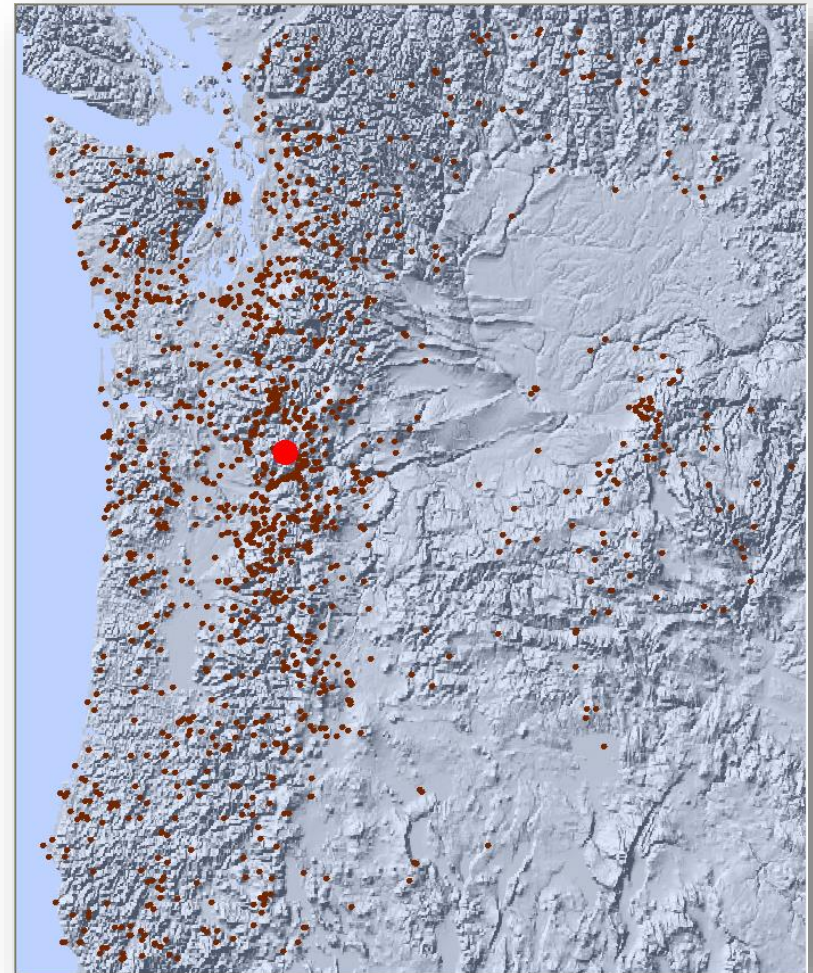
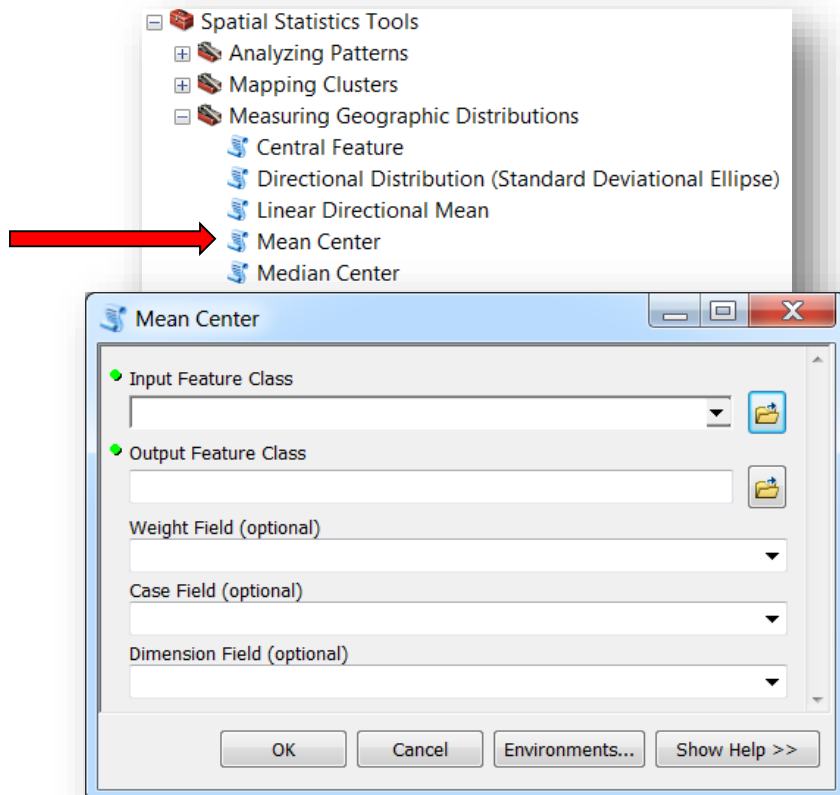
- *Measuring Geographic Distributions Toolset*
  - *Where's the center?*
  - *What's the shape and orientation of the data?*
  - *How dispersed are the features?*





# Measuring Geographic Distributions

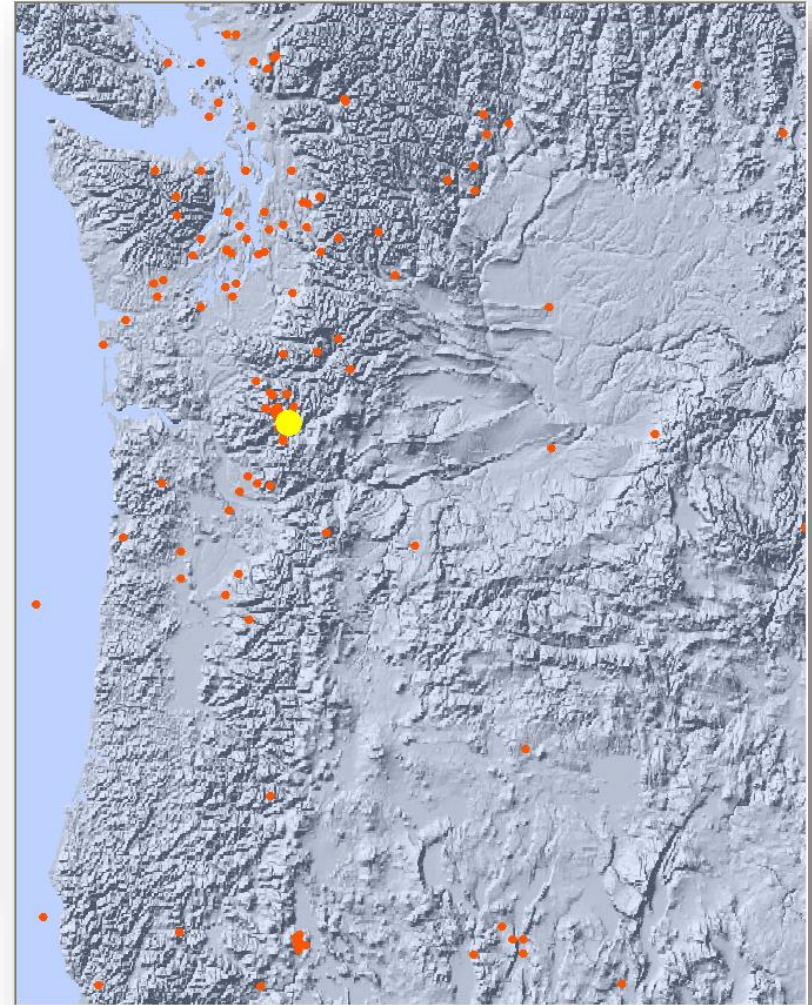
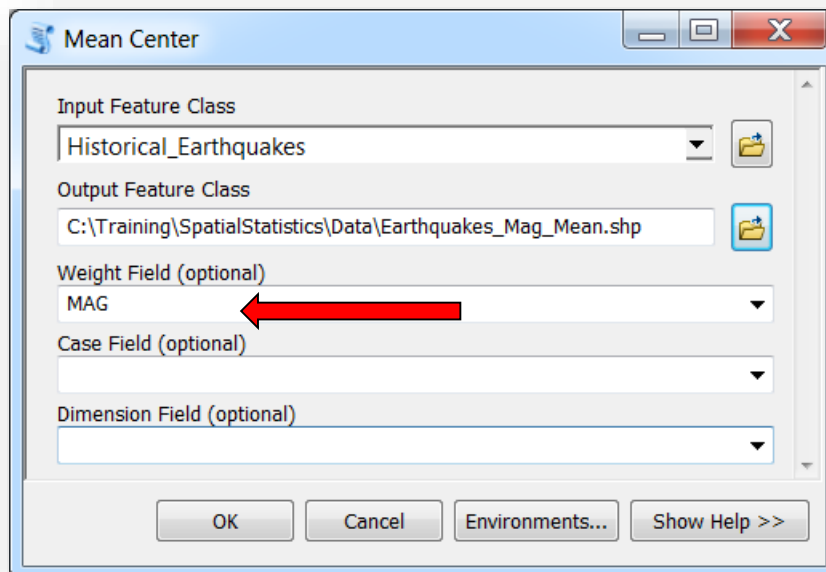
- Mean Center tool
  - *Where's the center?*





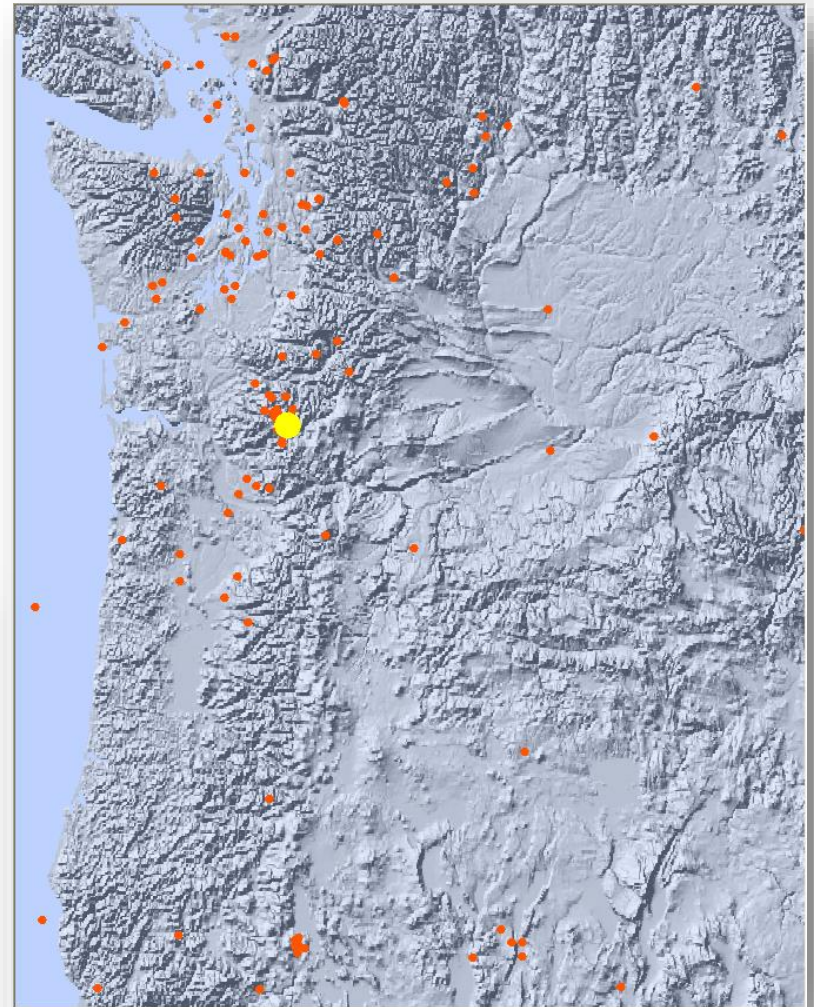
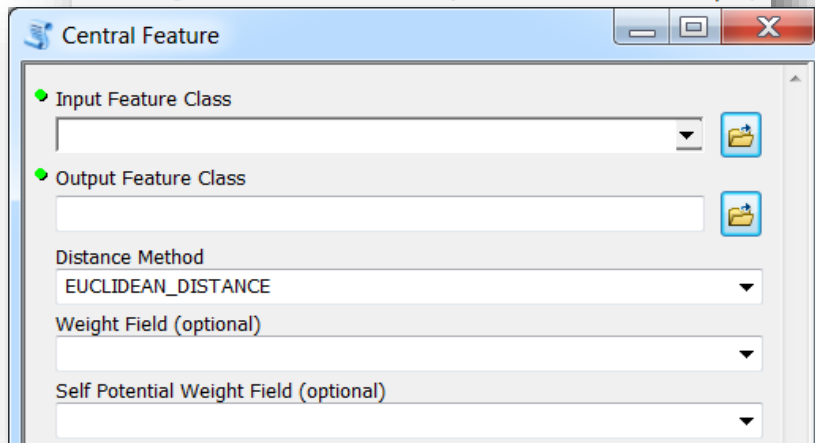
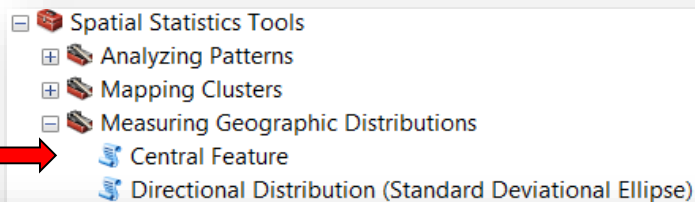
# Measuring Geographic Distributions

- Mean Center tool
  - Where's the *weighted* mean center?



# Measuring Geographic Distributions

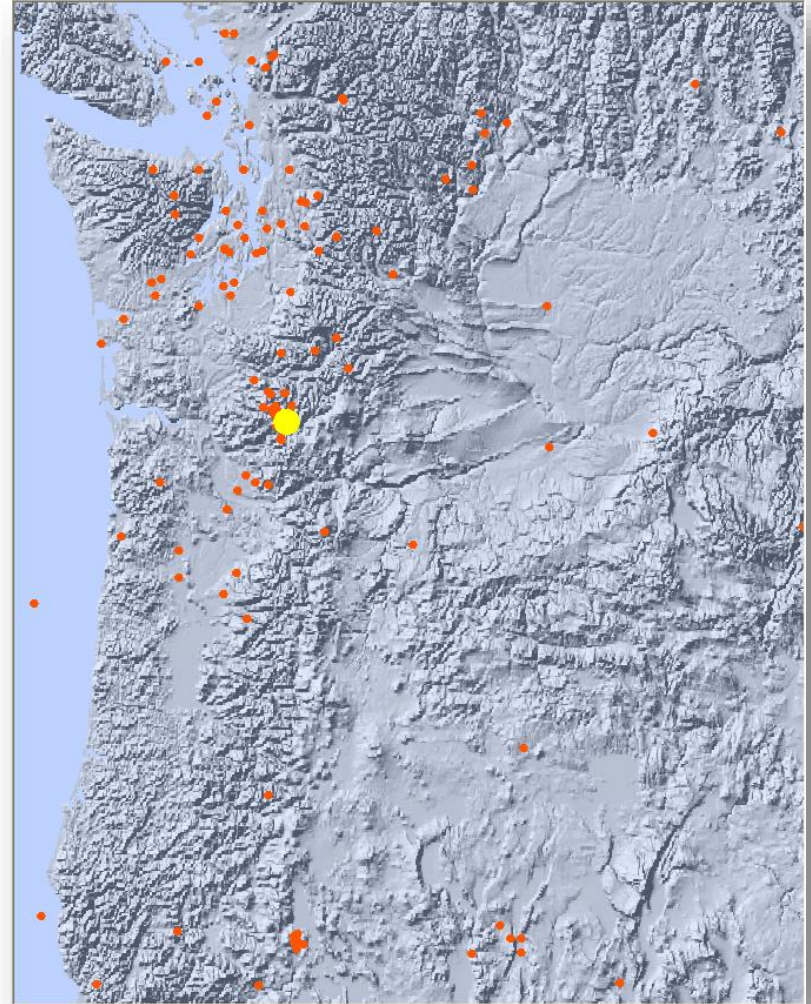
- Central Feature tool
  - *What feature is centrally located?*
  - *What feature is closest to all other features?*





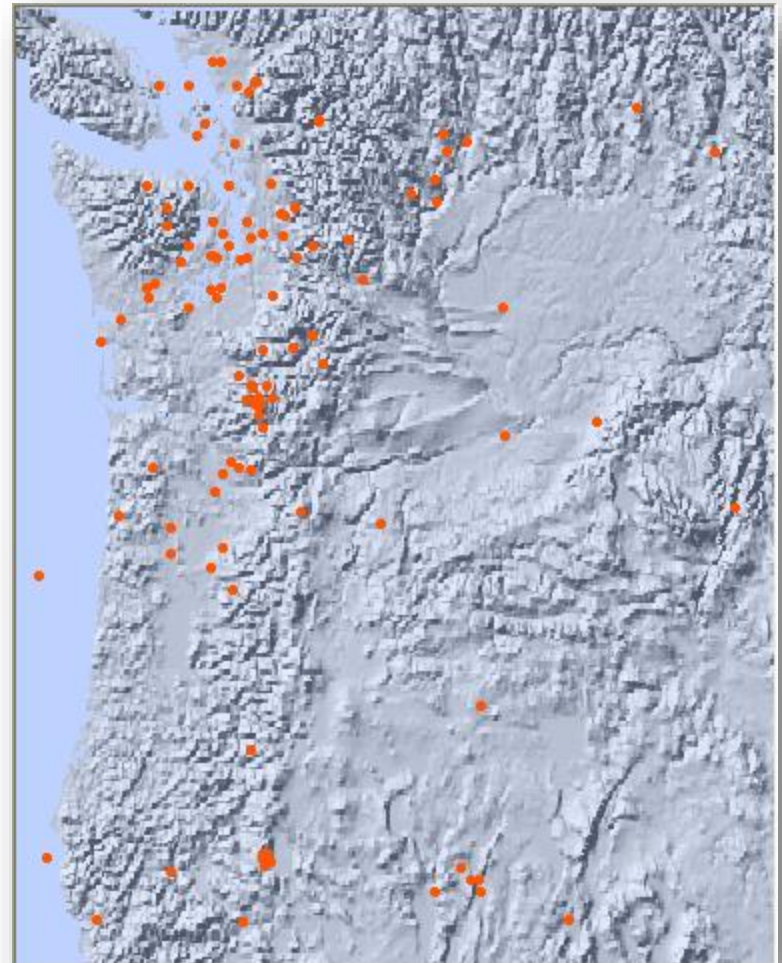
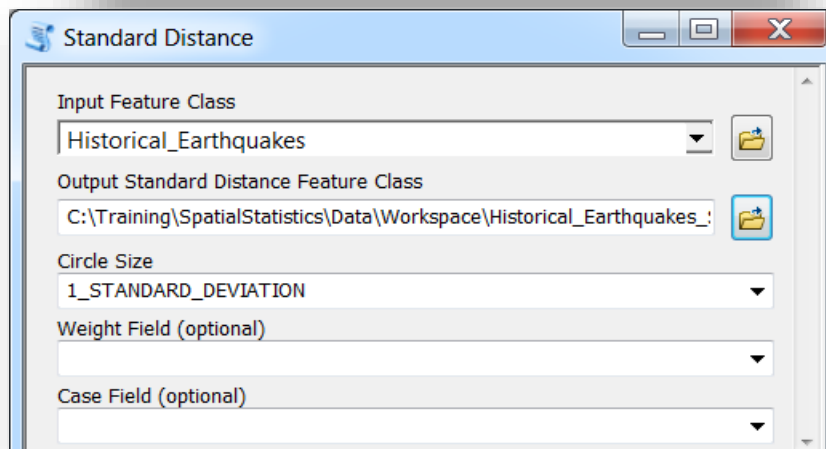
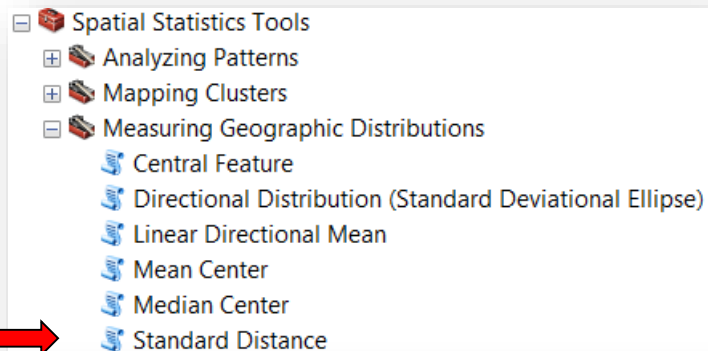
# Measuring Geographic Distributions

- Central Feature tool
  - *What feature is centrally located?*
  - *What feature is closest to all other features?*
  - *Where is the most accessible location to the greatest number of people?*



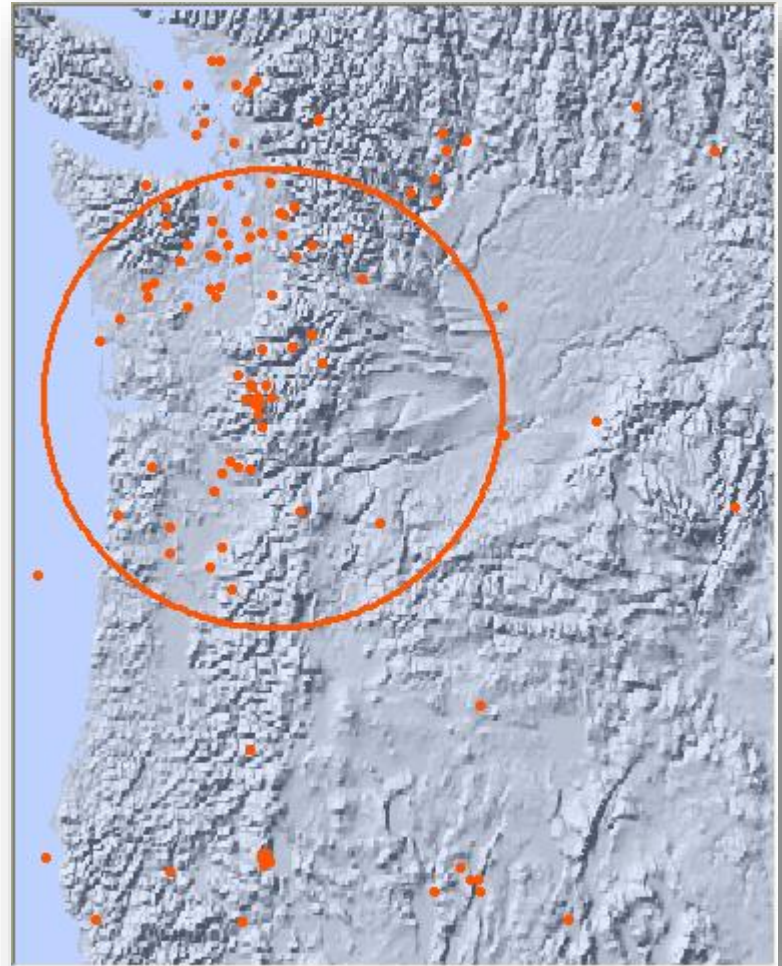
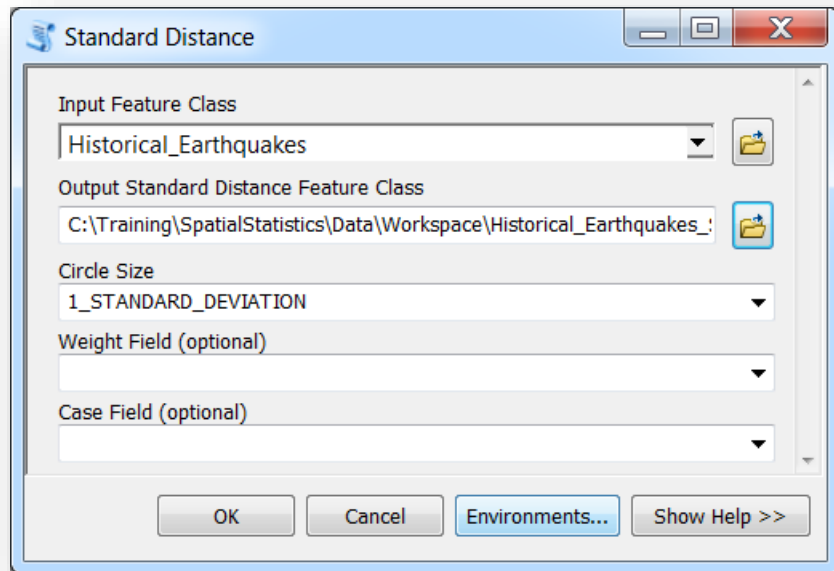
# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*

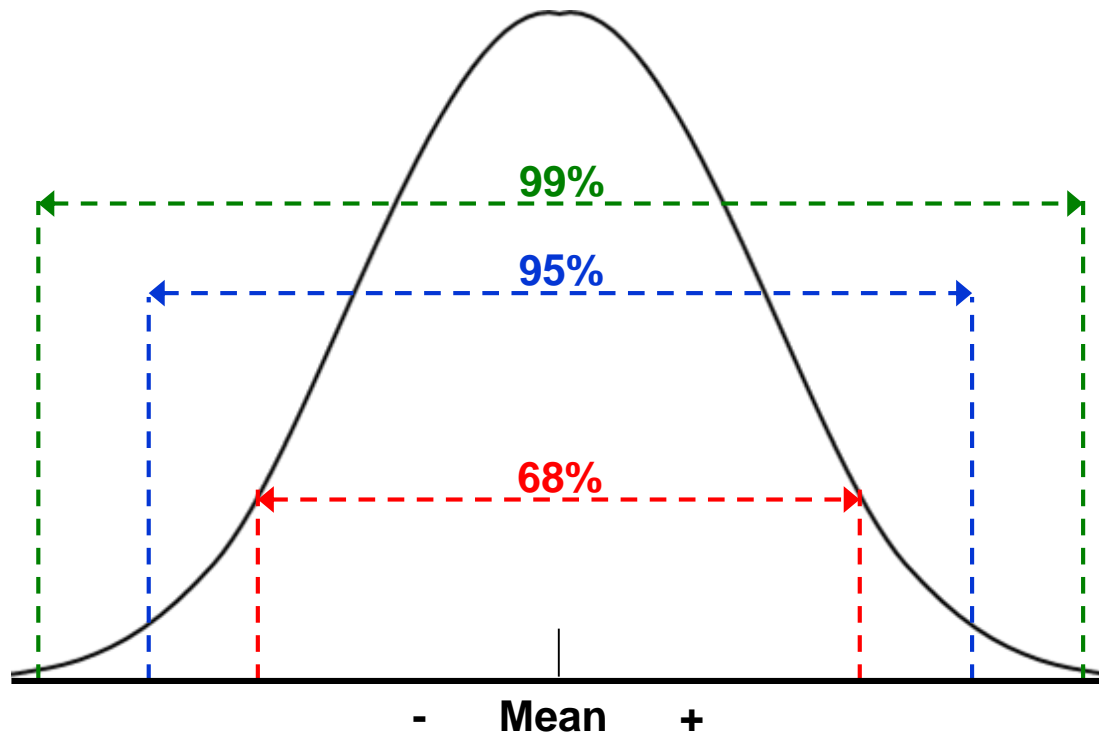


# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*



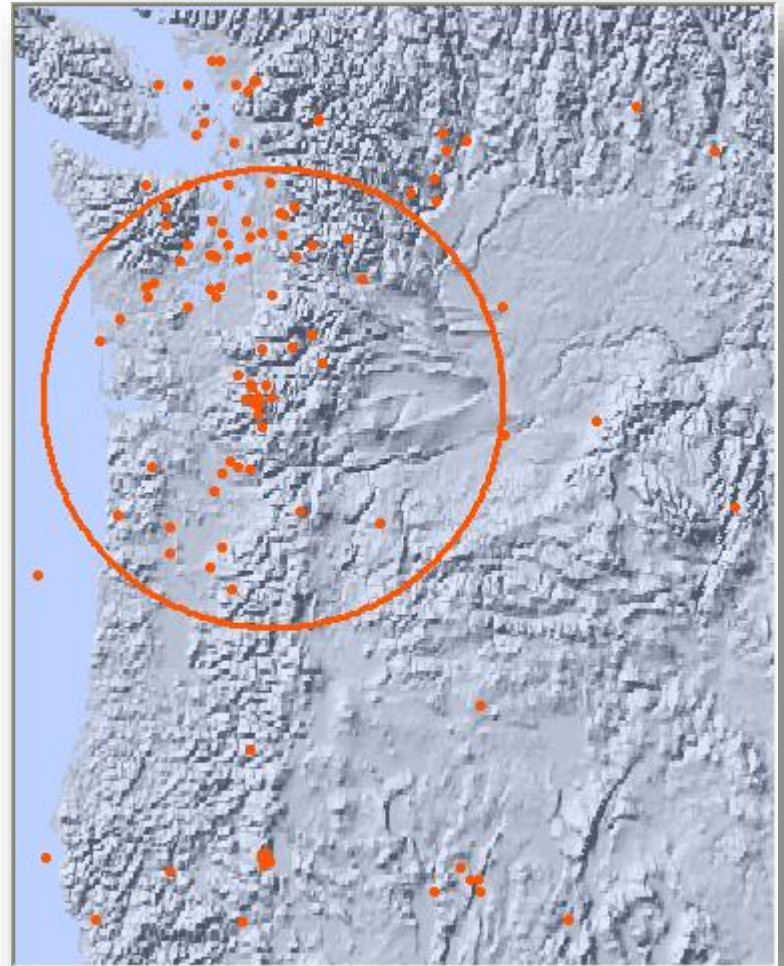
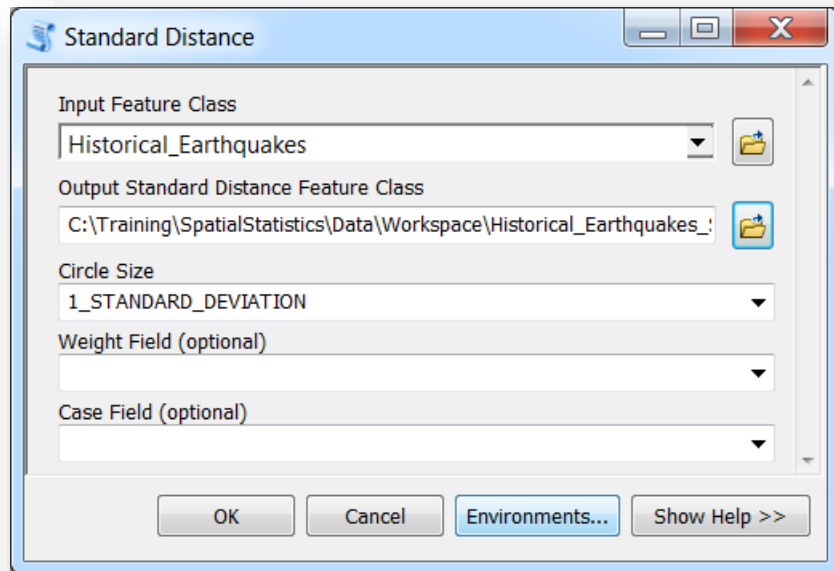
# Standard Deviations





# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*



*Distribution of earthquakes along the pacific northwest*

# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*

Table

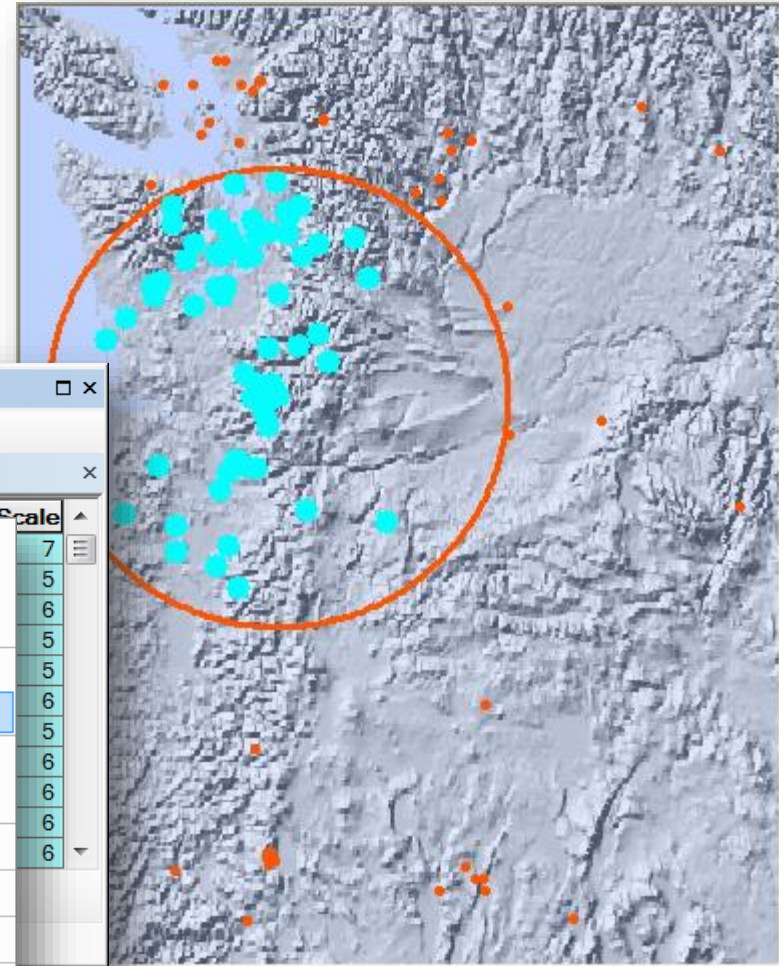
Historical\_Earthquakes

	MAG	MMI	LOCATION	YEAR	MONTH	DAY	HOUR	MMI Scale
▶	5	7	North of Portland, Oregon (in Washington)	187				7
▶	5.2	5	In Washington	188				5
▶	6.5	6	Near Olympia, Washington	188				6
▶	5	5	In Washington	189				5
▶	5	5	WASHINGTON-OREGON BORDER REGION	189				5
▶	5.9	6	Near Olympia, Washington	189				6
▶	5	5	WASHINGTON-OREGON BORDER REGION	189				5
▶	5.6	6	Near Startup, southern Snohomish County,	192				6
▶	5.4	6	Near Perrydale, Oregon	193				6
▶	4.8	6	In the Puget Sound area, Washington	193				6
▶	5.2	6	Near the forks of the Tolt River, Washington	193				6

1 (128 out of 178 Selected)

Historical\_Earthquakes

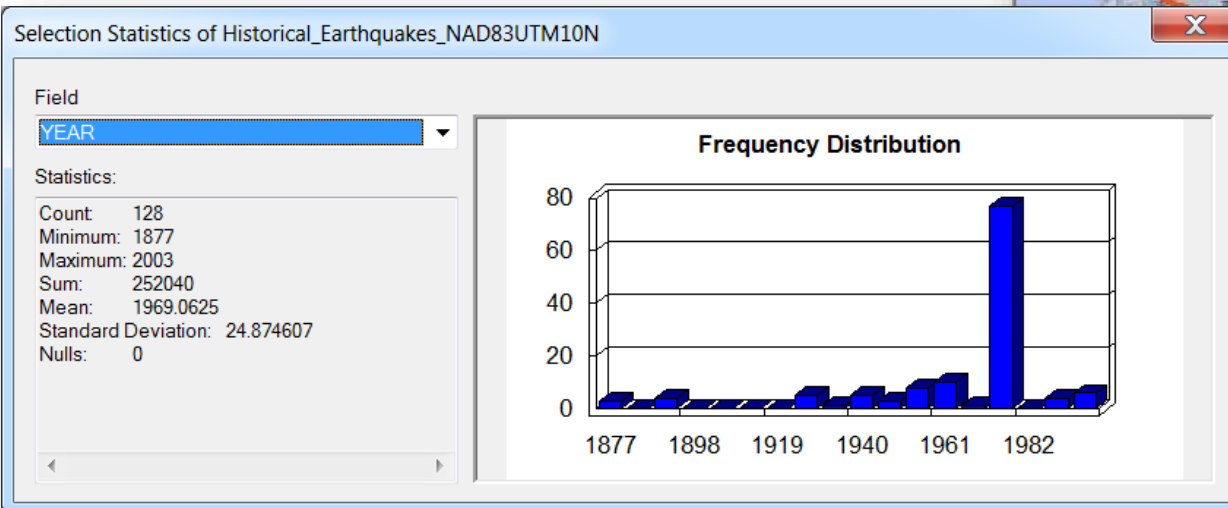
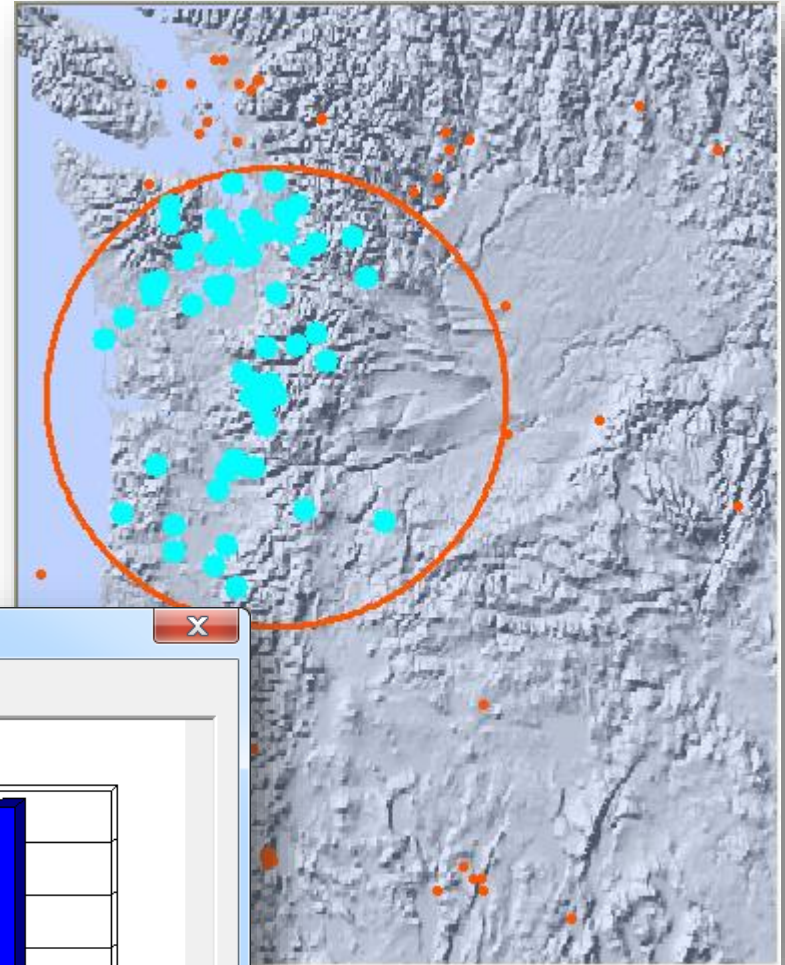
- Sort Ascending
- Sort Descending
- Advanced Sorting...
- Summarize...
- Σ Statistics...
- Field Calculator...
- Calculate Geometry...
- Turn Field Off
- Freeze/Unfreeze Column
- Delete Field
- Properties...





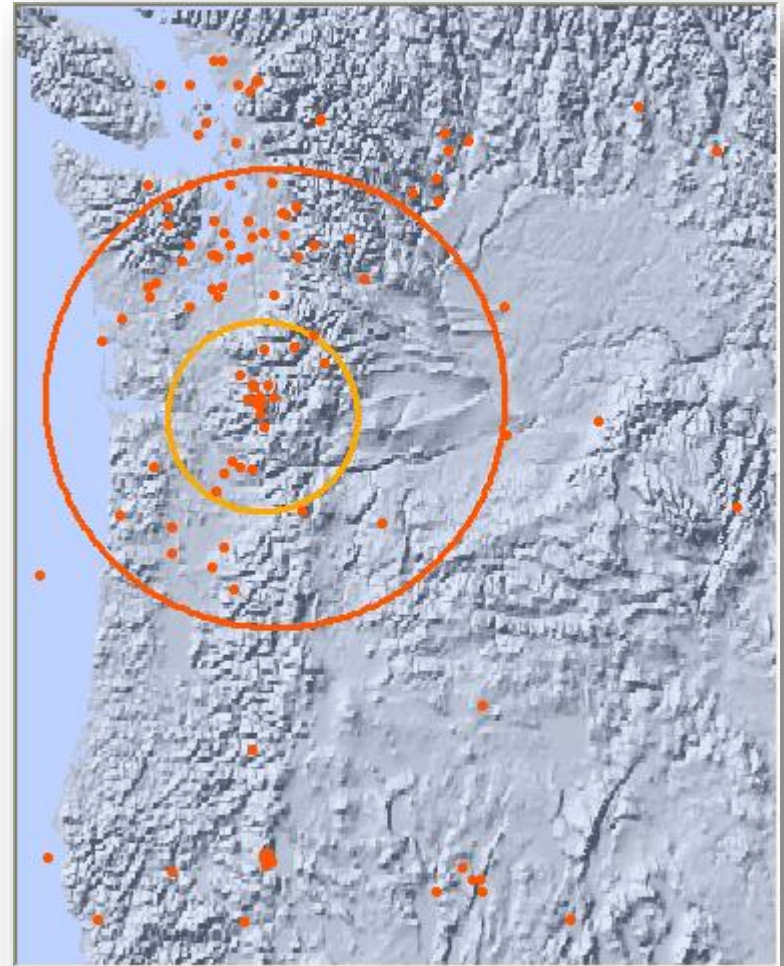
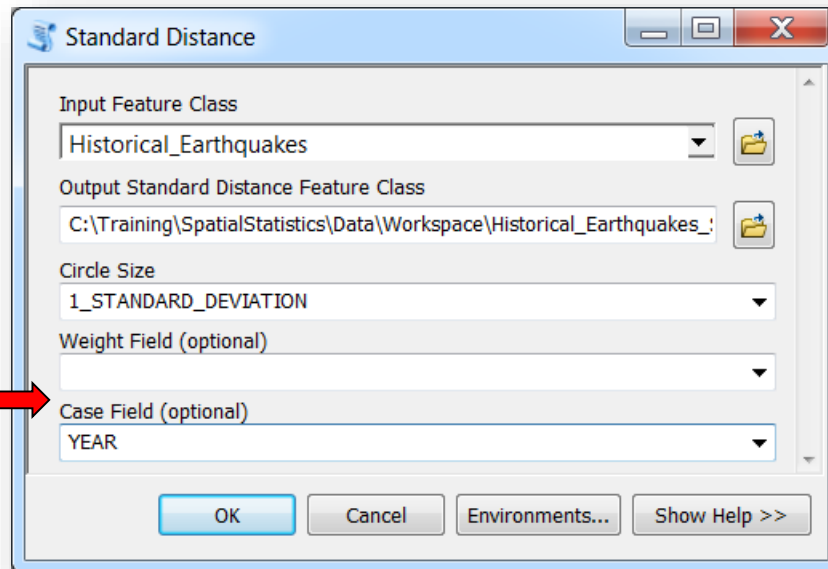
# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*



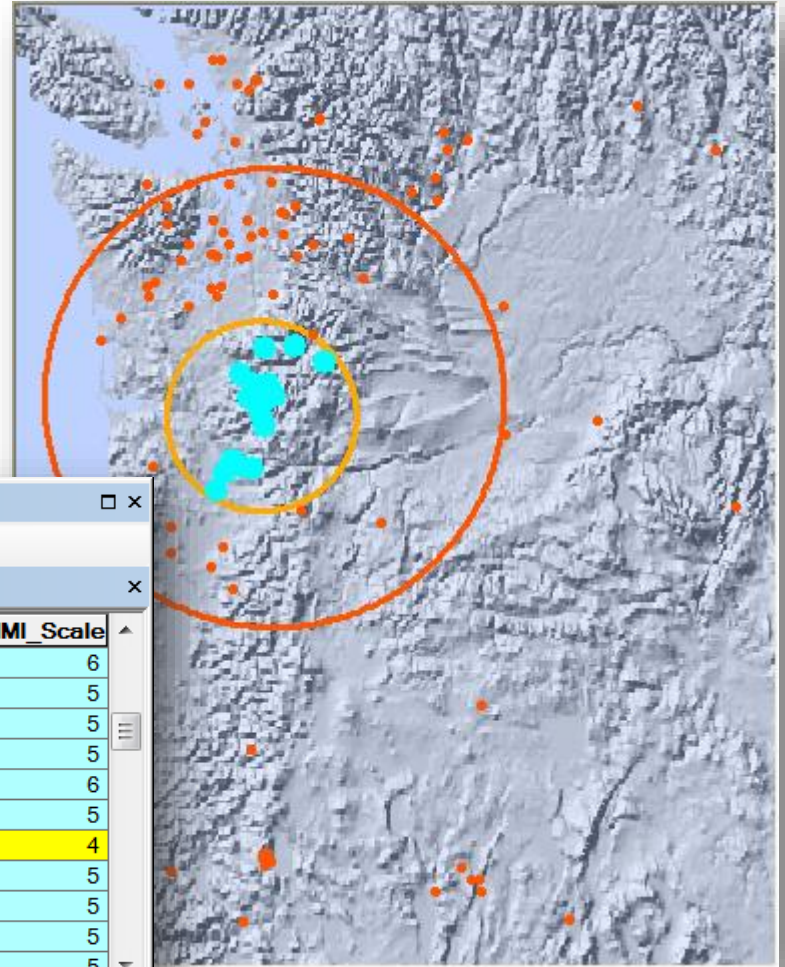
# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*



# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*



Table

Historical\_Earthquakes

	MAG	MMI	LOCATION	YEAR	MONTH	DAY	HOUR	MMI_Scale
	4.5	6	In Washington	1980	4	3	23	6
	4.5	5	In Washington	1980	4	9	18	5
	4.7	5	In Washington	1980	4	15	6	5
	4.7	5	In Washington	1980	4	18	0	5
	4.8	6	In Washington	1980	4	20	19	6
	4.6	5	In Washington	1980	5	15	11	5
	5	4	Near Mount Saint Helens, Washington (earth	1980	5	18	15	4
	4.7	5	In Washington	1980	4	1	4	5
	4.5	5	In Washington	1980	4	5	16	5
	4.7	5	In Washington	1980	4	6	6	5
	4.7	5	In Washington	1980	4	7	15	5

(81 out of 178 Selected)

Historical Earthquakes



# Measuring Geographic Distributions

- Standard Distance Tool
  - *How is the data dispersed around the mean?*

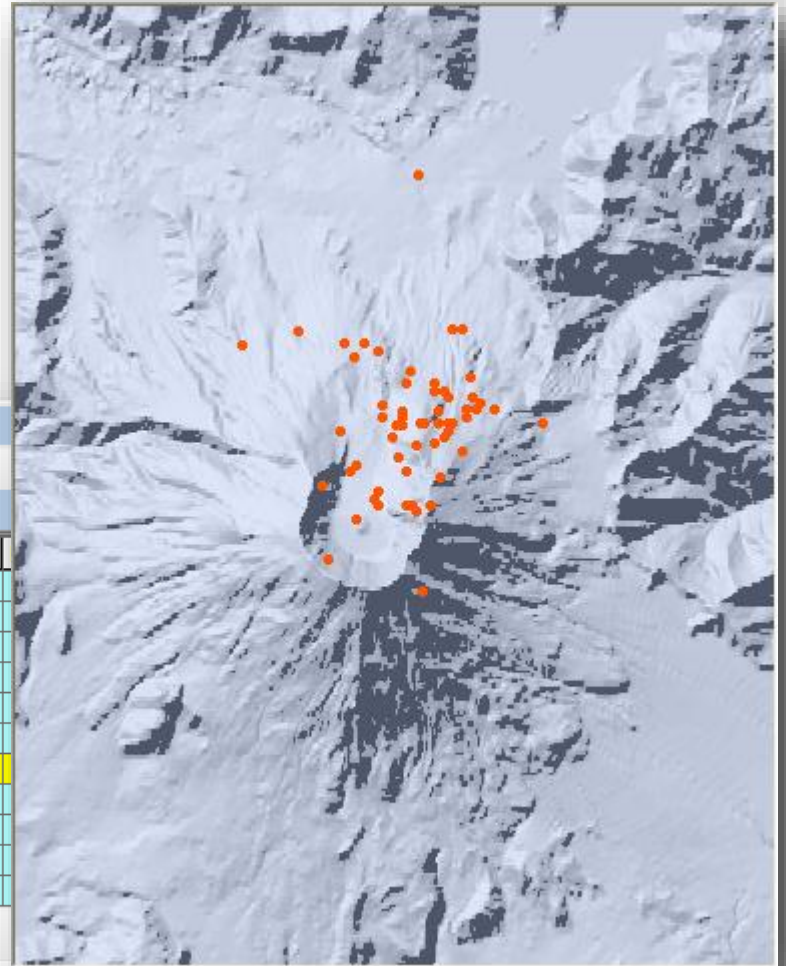
Table

Historical\_Earthquakes

	MAG	MMI	LOCATION	YEAR	MONTH	DAY	HOUR
	4.5	6	In Washington	1980	4	3	23
	4.5	5	In Washington	1980	4	9	18
	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 (earth	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
	4.7	5	In Washington	1980	4	7	15

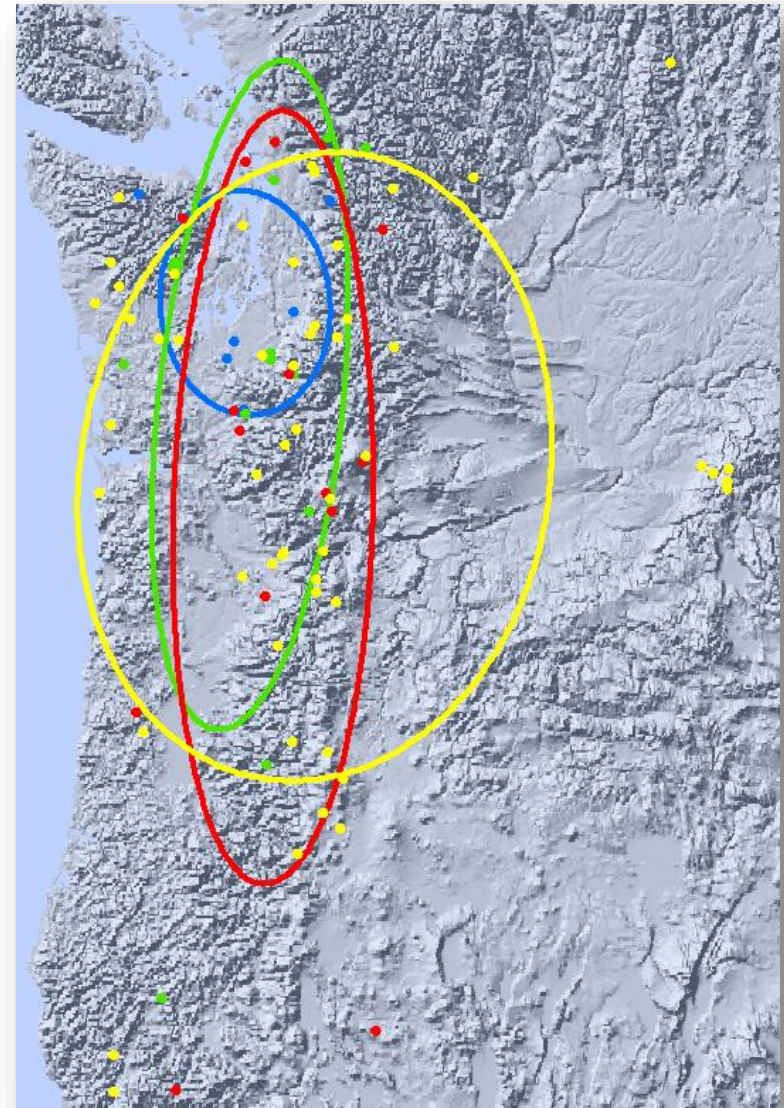
0 (81 out of 178 Selected)

Historical\_Earthquakes



# Measuring Geographic Distributions

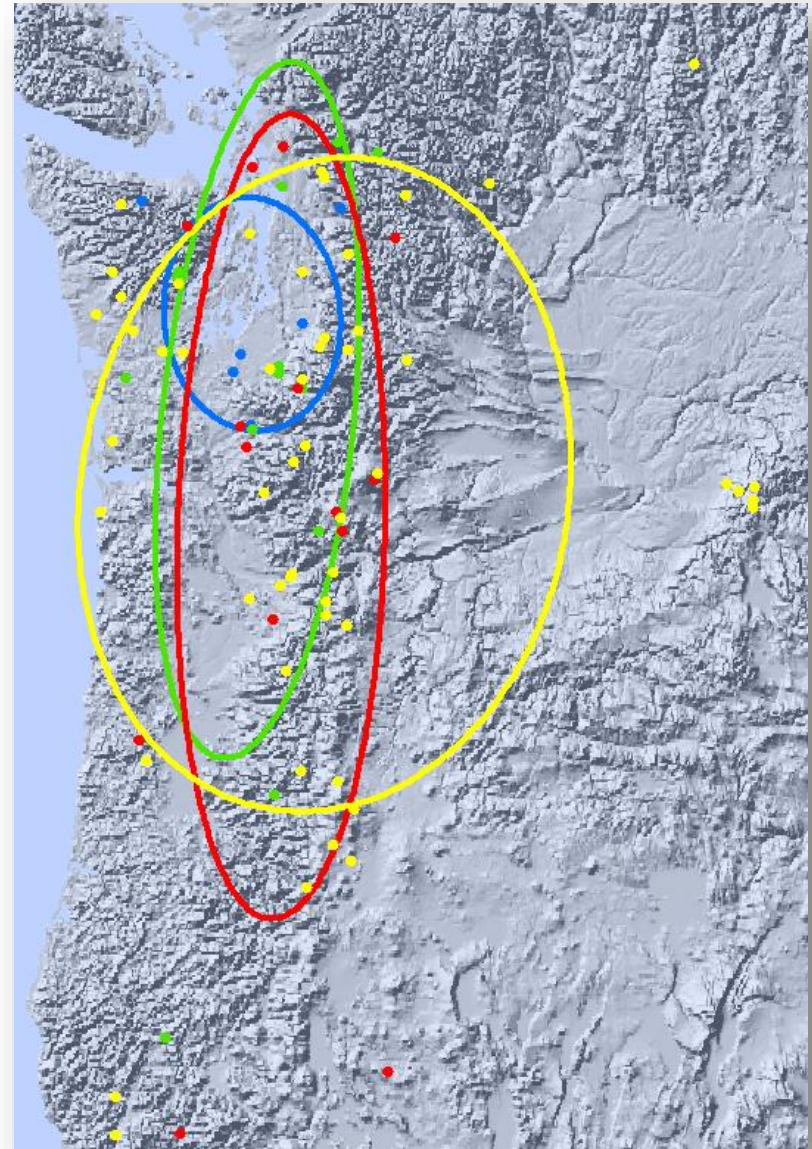
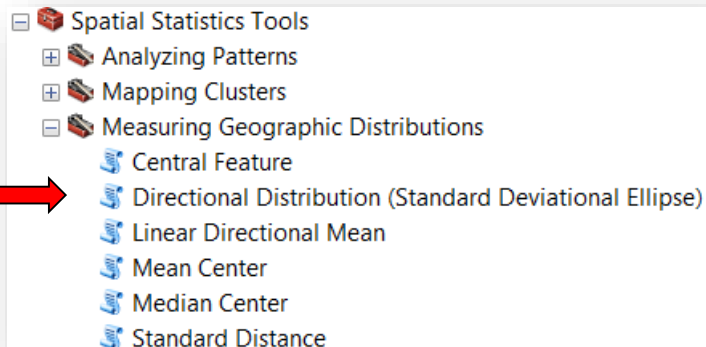
- Directional Distribution (Standard Deviation Ellipse) Tool
  - *Identify spatial trends in the distribution of the data*
  - *Compare distributions*
  - *Examine data over different time periods*
  - *Show compactness and orientation*





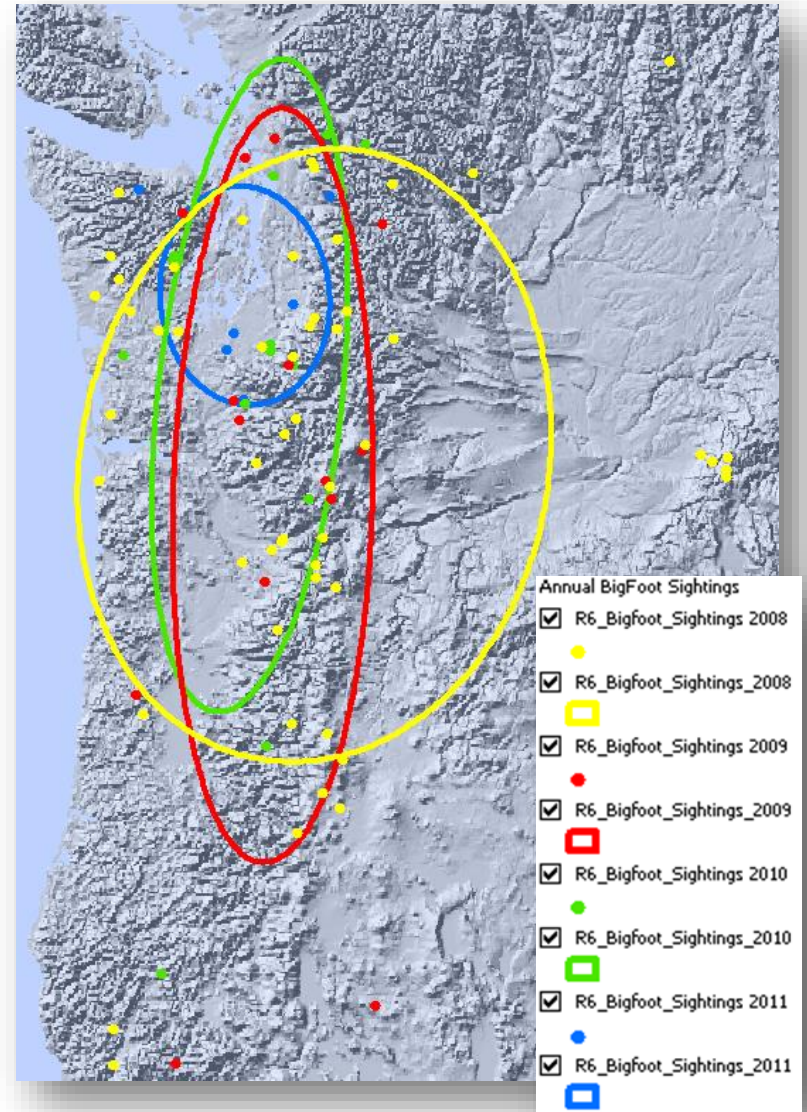
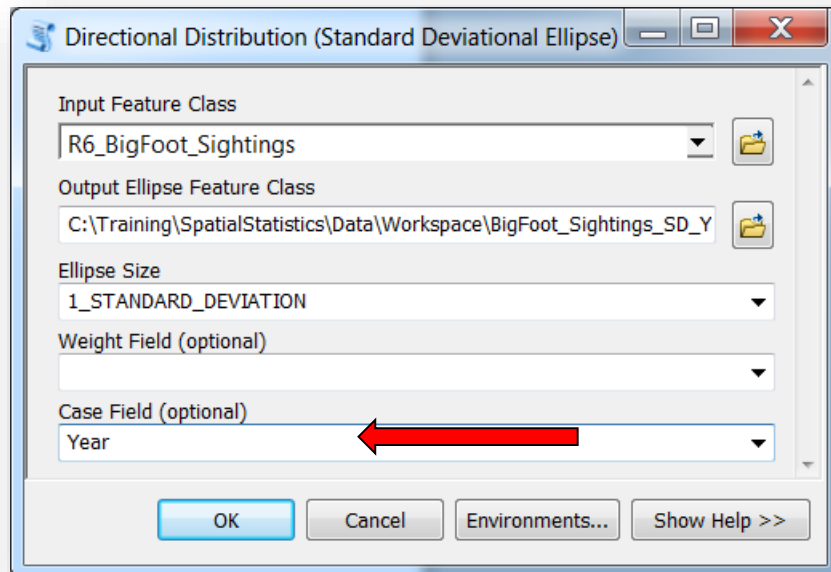
# Measuring Geographic Distributions

- Directional Distribution (Standard Deviation Ellipse) Tool
  - *What is the distribution and orientation of the data?*
  - *What is the migration pattern of wildlife*
  - *Track forest fire movement*



# Measuring Geographic Distributions

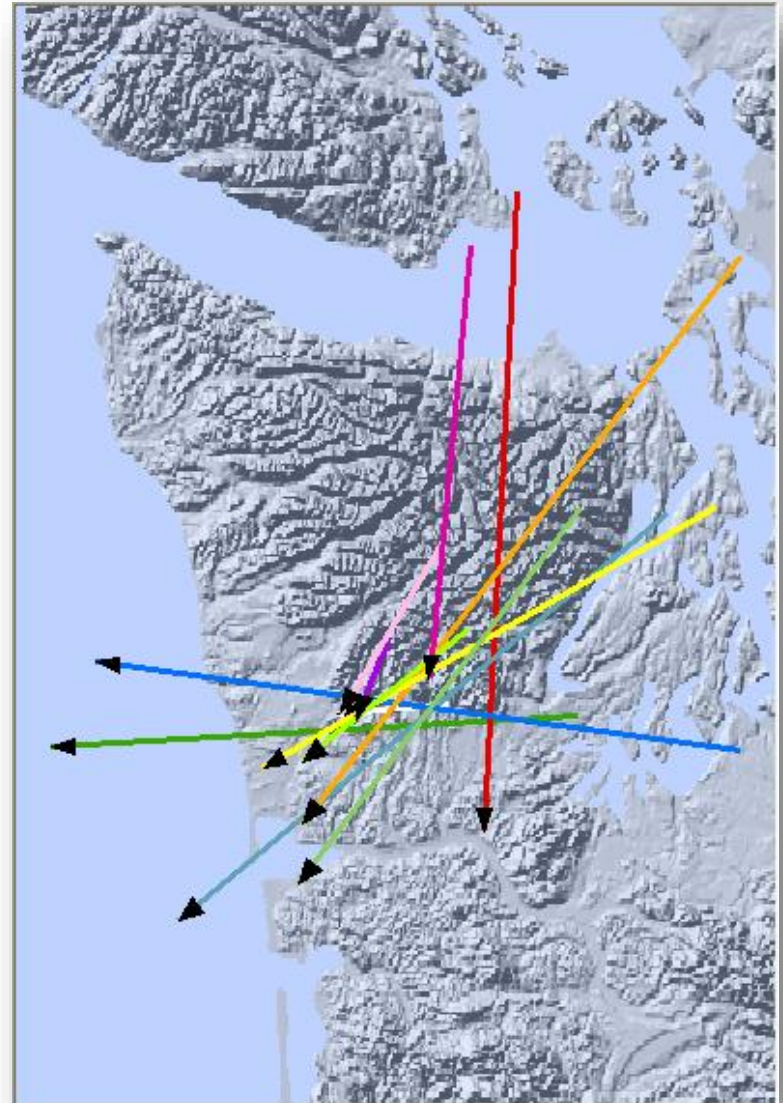
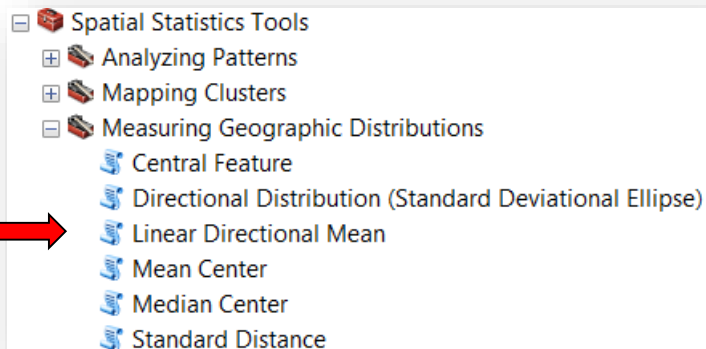
- Directional Distribution (Standard Deviation Ellipse) Tool





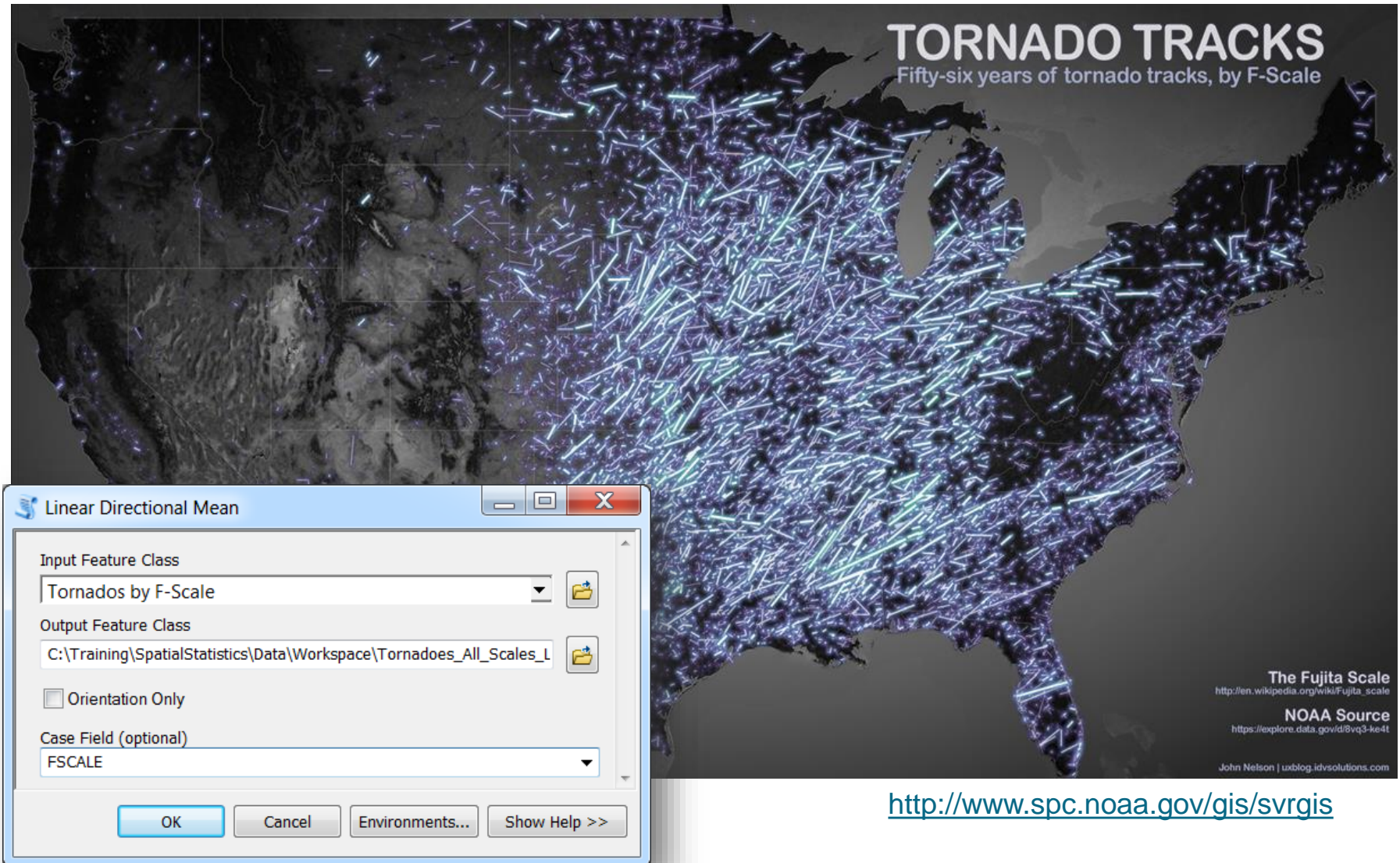
# Measuring Geographic Distributions

- Linear Directional Mean Tool
  - Compare directions of wildlife migration routes and determine their distances
  - Analyze directional trends and patterns during different seasons/years





# Tornado Paths from 1950-2011



# Measuring Geographic Distributions

- Linear Directional Mean Tool
  - Output represents mean center and direction or orientation of inputs
    - Orientation Only if option selected

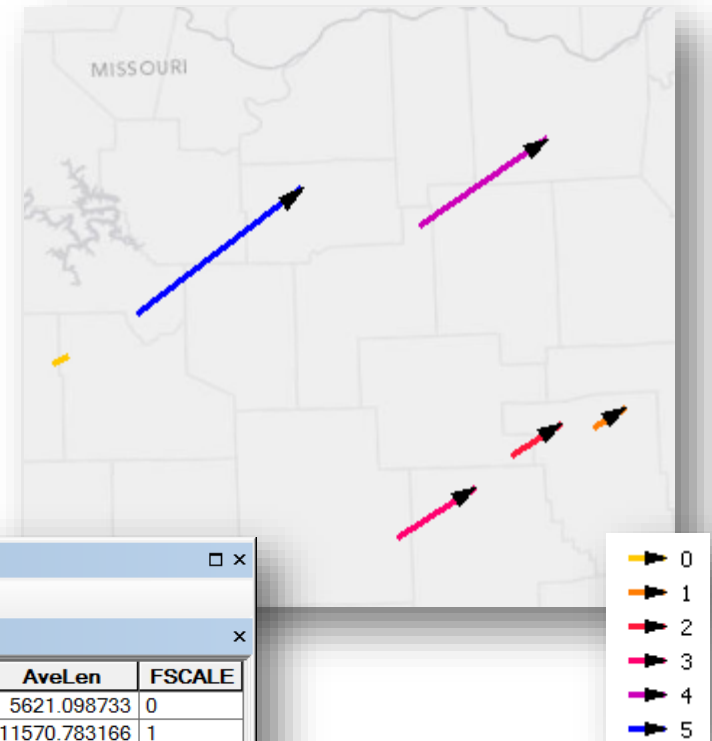


Table									
Tornadoes_All_Scales_LDM									
	FID	Shape *	Id	CompassA	DirMean	CirVar	AveX	AveY	AveLen
▶	0	Polyline	0	62.047508	27.952492	0.353151	275494.908482	-131297.509528	5621.098733
	1	Polyline	0	59.087297	30.912703	0.246899	444207.092687	-149023.408738	11570.783166
	2	Polyline	0	56.215931	33.784069	0.195643	422008.40825	-155962.268278	18841.978024
	3	Polyline	0	56.960521	33.039479	0.161706	391196.48778	-178418.322703	28978.136392
	4	Polyline	0	55.684091	34.315909	0.140319	405459.431312	-76370.772921	47629.591529
	5	Polyline	0	52.599961	37.400039	0.108384	324469.795157	-97669.740754	63624.946579

1 (0 out of 6 Selected)

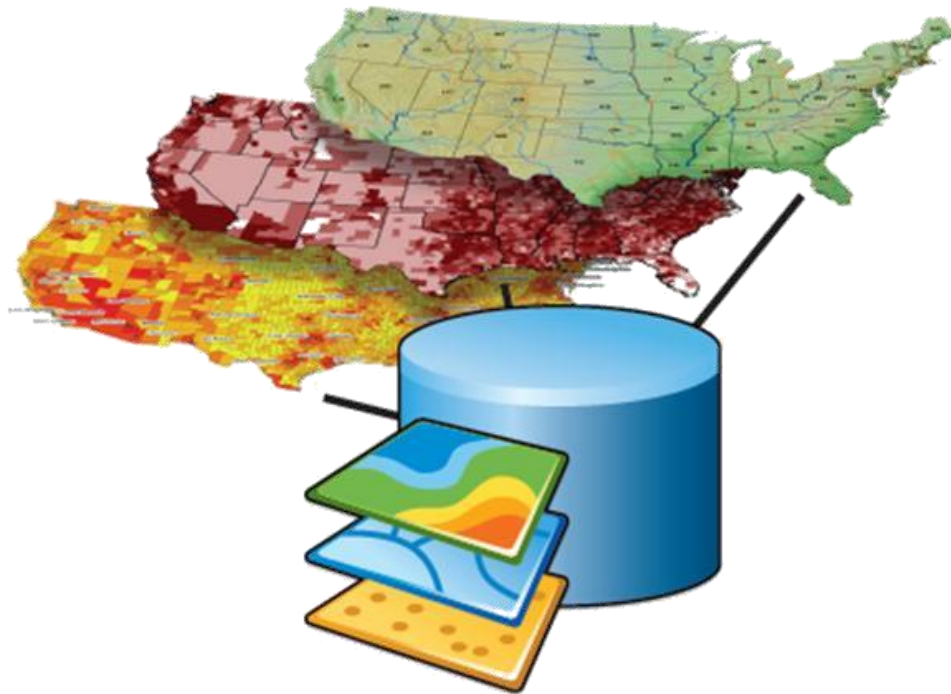
Tornadoes\_All\_Scales\_LDM

# Review / Q&A

- What are spatial statistics?
  - *A measure of what's going on spatially*
- Why use spatial statistics?
  - *Where are the spatial outliers?*
- Spatial Statistics tools
  - *Not an extension*
- Measuring Geographic Distributions Toolset
  - *Mean Center*
  - *Central feature*
  - *Standard Distance*
  - *Directional Distribution (Standard Deviation Ellipse)*

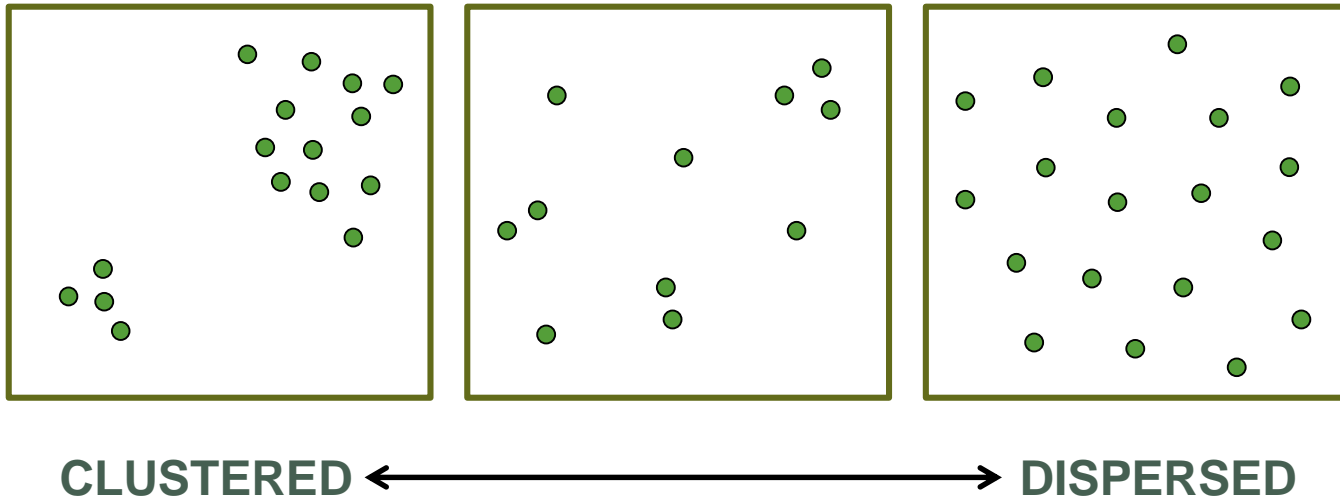


# Demonstration: Measuring Geographic Distributions



# Spatial Pattern Analysis

- Measures the “*degree*” features are clustered, dispersed or randomly distributed



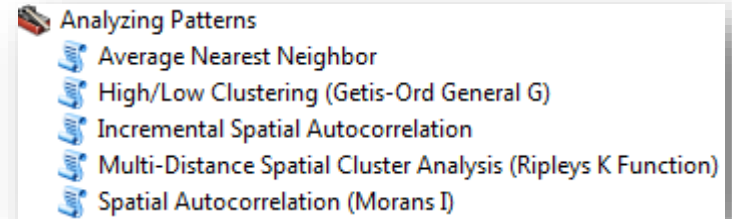


# Spatial Pattern Analysis

- Analyzing Patterns toolset

- Global Calculations

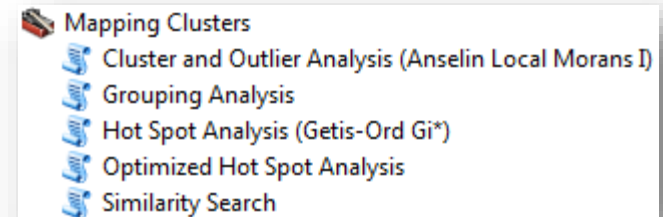
- Summarizes the overall spatial pattern of a dataset
    - Identifies statistically significant clustering/dispersion
    - *Is there spatial clustering?*



- Mapping Clusters toolset

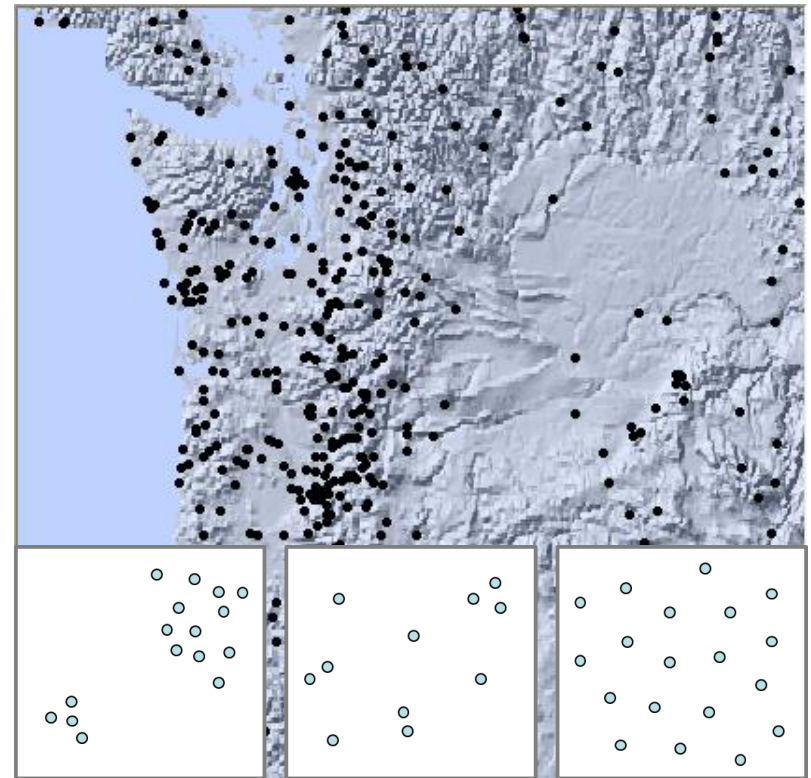
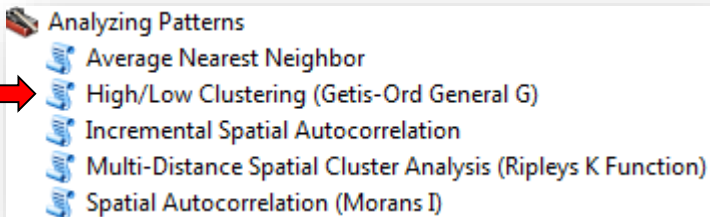
- Local Calculations

- Identifies the extent and location of clustering
    - *Where are the clusters (hot spots/cold spots)?*
    - *Where are the spatial outliers?*



# Analyzing Patterns

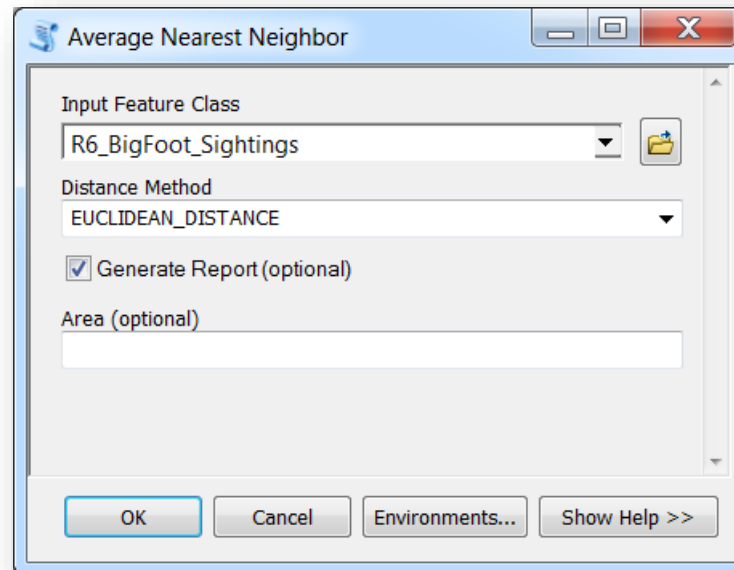
- Average Nearest Neighbor tool
  - Identifies statistically significant clustering



CLUSTERED ← → DISPERSED

# Analyzing Patterns

- Average Nearest Neighbor tool
  - Identifies statistically significant clustering
    - Observed Mean Distance, Expected mean Distance, Nearest Neighbor Index, Z Score, P Value



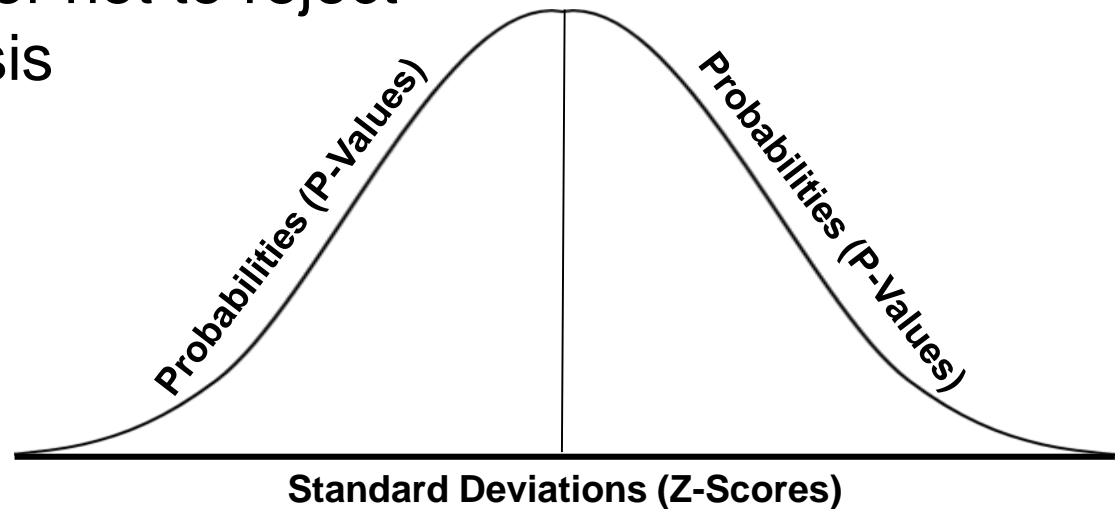
# Statistics 101

- Observed Mean Distance
  - Actual average distance between neighbors
- Expected Mean Distance
  - Hypothetical average distance between neighbors in a normal random distribution
    - $< 1$  = clustering
    - $> 1$  = dispersion
- Nearest Neighbor Index
  - Observed Mean Distance : Expected Mean Distance



# Statistics 101

- Average Nearest Neighbor Hypothesis
  - “complete spatial randomness”
- Z Score and P Value
  - measures of statistical significance
  - tell you whether or not to reject the null hypothesis



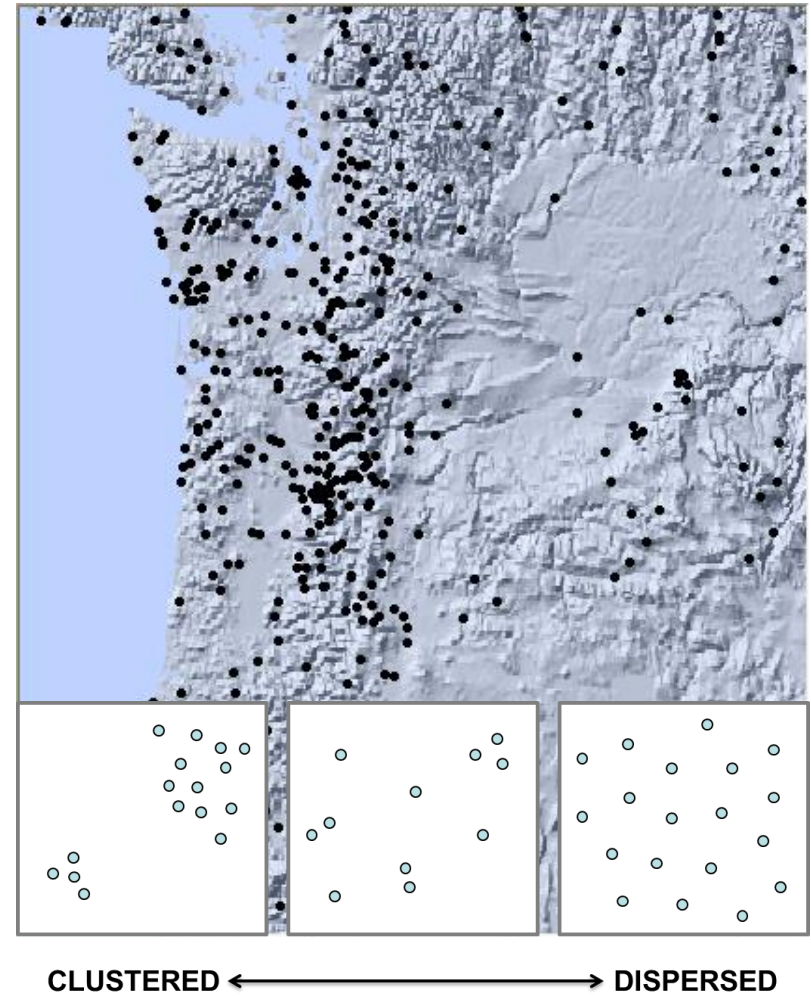
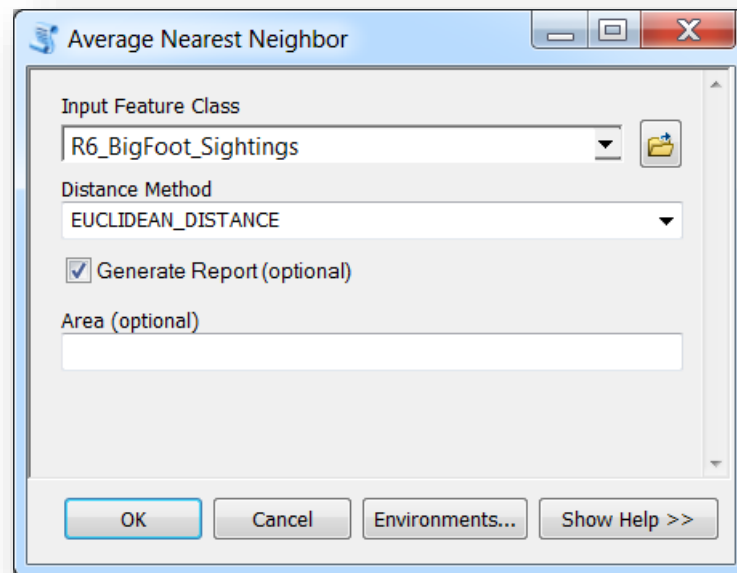
# Statistics 101

- Critical Z Scores and P Values

z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%

# Analyzing Patterns

- Average Nearest Neighbor tool
  - Identifies statistically significant clustering



# Average Nearest Neighbor Results

Results

- Current Session
  - Average Nearest Neighbor [103840\_04012014]
    - NNRatio: 0.591783
    - NNZScore: -29.624556
    - PValue: 0
    - NNExpected: 9283.86194
    - NNObserved: 5494.036221
    - Report File: NearestNeighbor\_Result.html

Inputs

Environments

Messages

Previous Sessions

Shared

Copy

View...

DOUBLE  
CLICK

RIGHT  
CLICK

Messages

Running script AverageNearestNeighbor...

Average Nearest Neighbor Summary

Observed Mean Distance: 5494.036221

Expected Mean Distance: 9283.861940

Nearest Neighbor Ratio: 0.591783

z-score: -29.624556

p-value: 0.000000

Distance measured in Meters

Writing html report....

C:\SpatialStatistics\Data\Workspace

\NearestNeighbor\_Result.html

Completed script AverageNearestNeighbor.

Succeeded at Tue Apr 01 10:38:40 2014 (E

3.75 seconds)

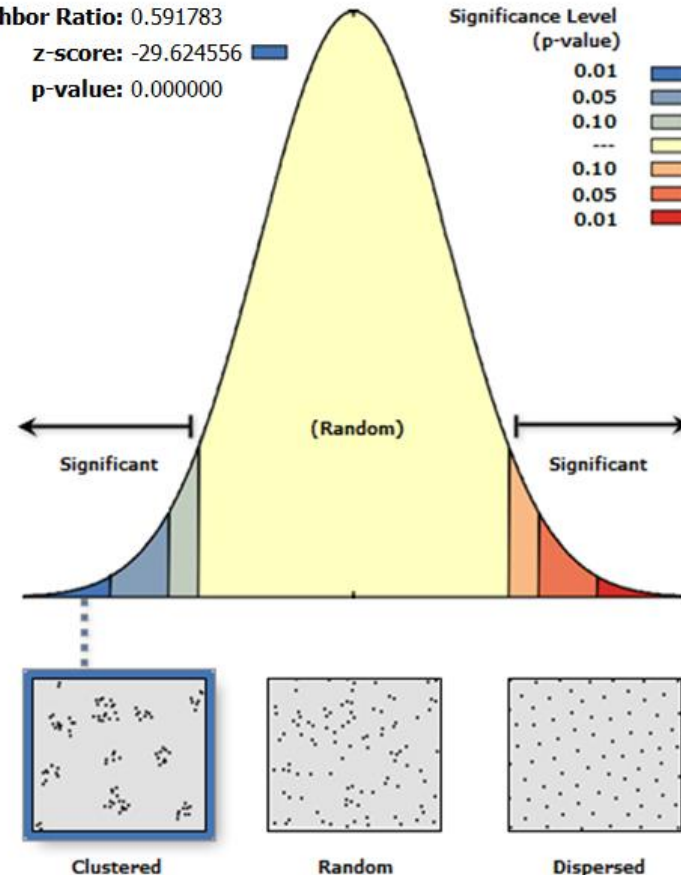
Nearest Neighbor Ratio: 0.591783

z-score: -29.624556

p-value: 0.000000

Significance Level  
(p-value)

Significance Level (p-value)	Critical Value (z-score)
0.01	< -2.58
0.05	-2.58 - -1.96
0.10	-1.96 - -1.65
---	-1.65 - 1.65
0.10	1.65 - 1.96
0.05	1.96 - 2.58
0.01	> 2.58

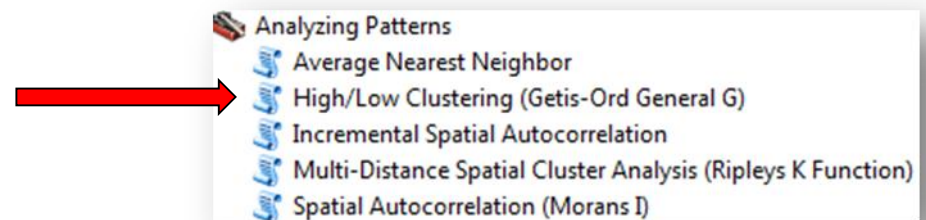
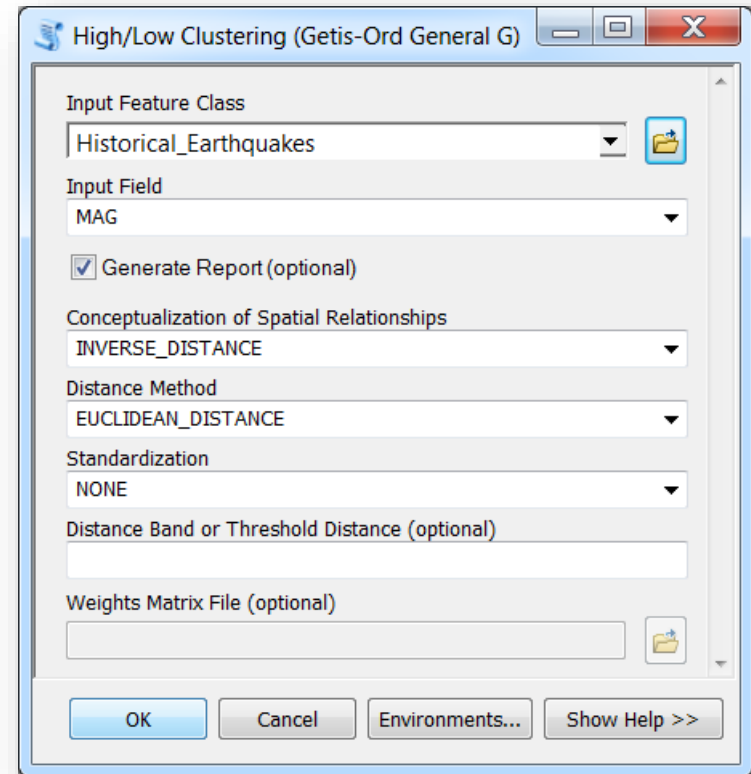


Given the z-score of -29.62, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.



# Analyzing Patterns

- High/Low Clustering tool (Getis-Ord General G\*)
  - Measures concentrations of high or low attribute values
  - Most effective when spatial pattern is consistent
  - Hypothesis: no spatial clustering of feature values

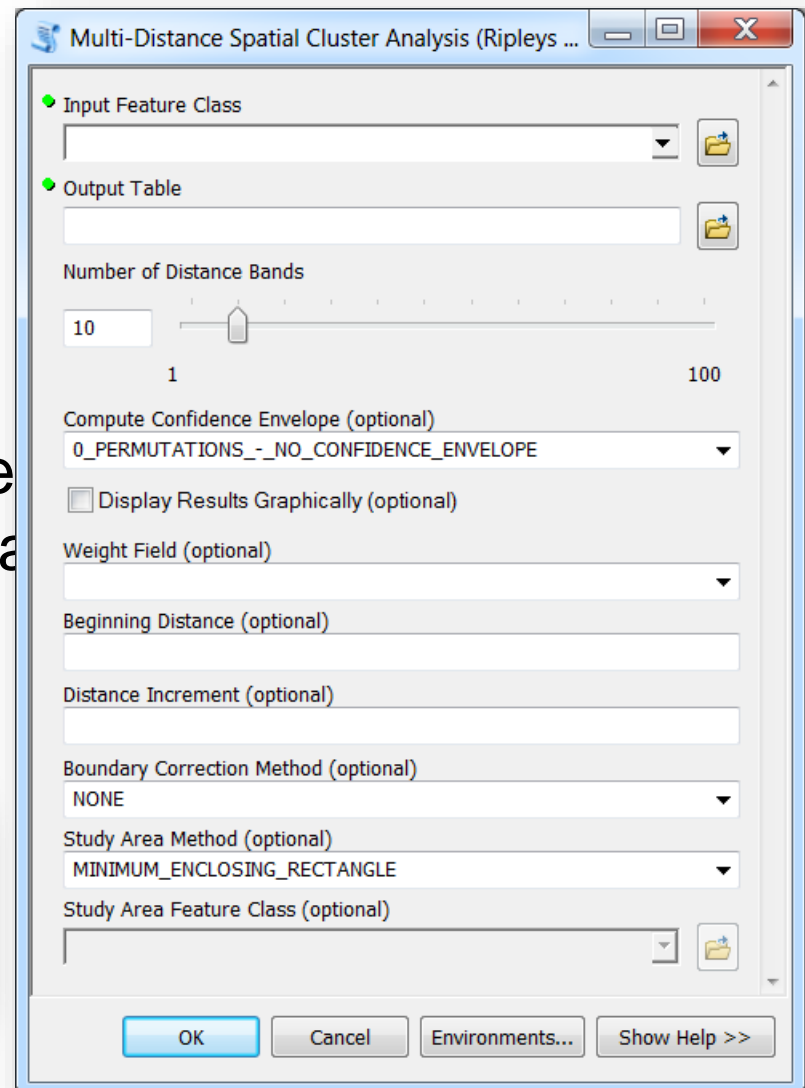
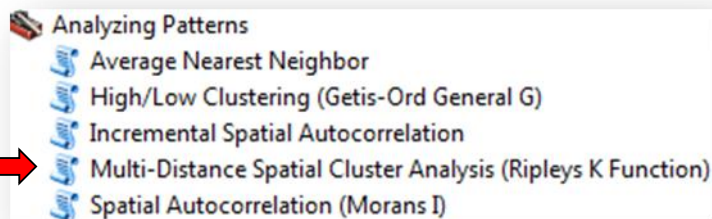


# High/Low Clustering Results

Result	High/Low Clustering
The p-value is <b>not</b> statistically significant.	You cannot reject the null hypothesis. It is quite possible that the spatial distribution of feature attribute values is the result of random spatial processes. Said another way, the observed spatial pattern of values could well be one of many, many possible versions of complete spatial randomness.
The p-value <b>is</b> statistically significant, and the z-score is positive.	You may reject the null hypothesis. The spatial distribution of <b>high</b> values in the dataset is more spatially clustered than would be expected if underlying spatial processes were truly random.
The p-value <b>is</b> statistically significant, and the z-score is negative.	You may reject the null hypothesis. The spatial distribution of <b>low</b> values in the dataset is more spatially clustered than would be expected if underlying spatial processes were truly random.

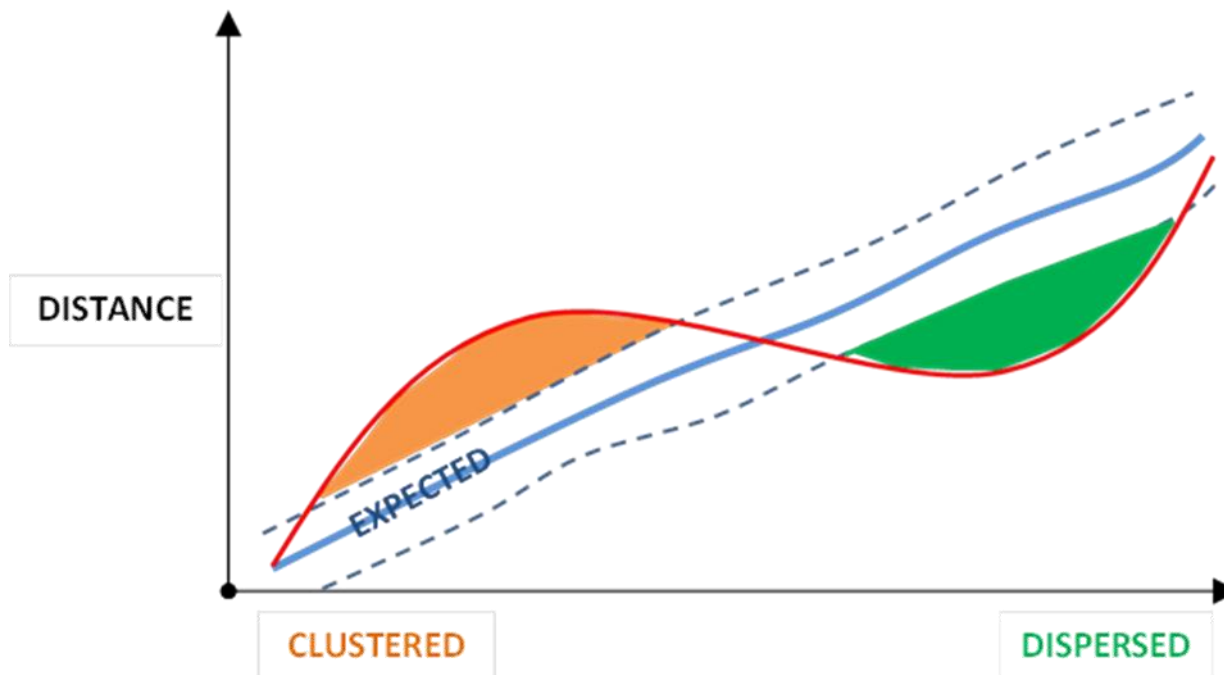
# Analyzing Patterns

- Multi-Distance Spatial Cluster Analysis tool (Ripley's K Function)
  - Assesses spatial clustering/dispersion for a set of geographic features over a range of distances



# Analyzing Patterns

- Multi-Distance Spatial Cluster Analysis (Ripleys K Function) tool
  - Assesses spatial clustering/dispersion for a set of geographic features over a range of distances

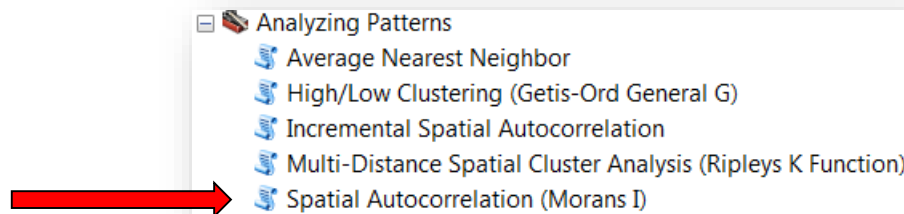


# Analyzing Patterns

- Spatial Autocorrelation (Moran's I) tool
  - Measures the similarity of neighboring features
  - Identifies if features are clustered or dispersed
  - Based on location and attribute values

## The First Law of Geography

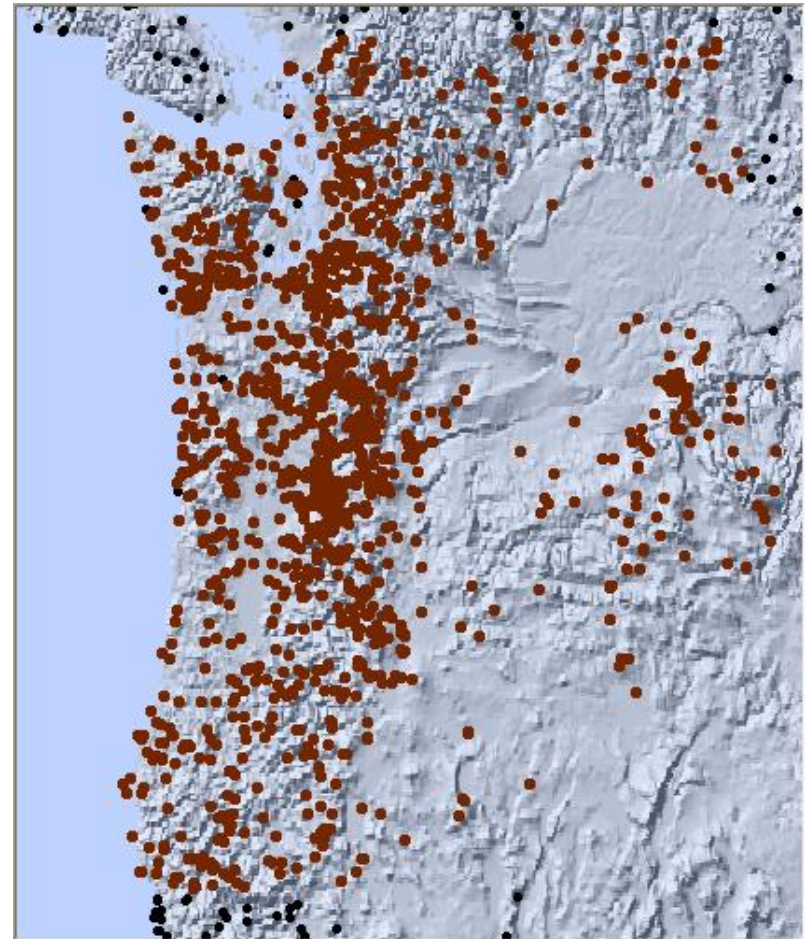
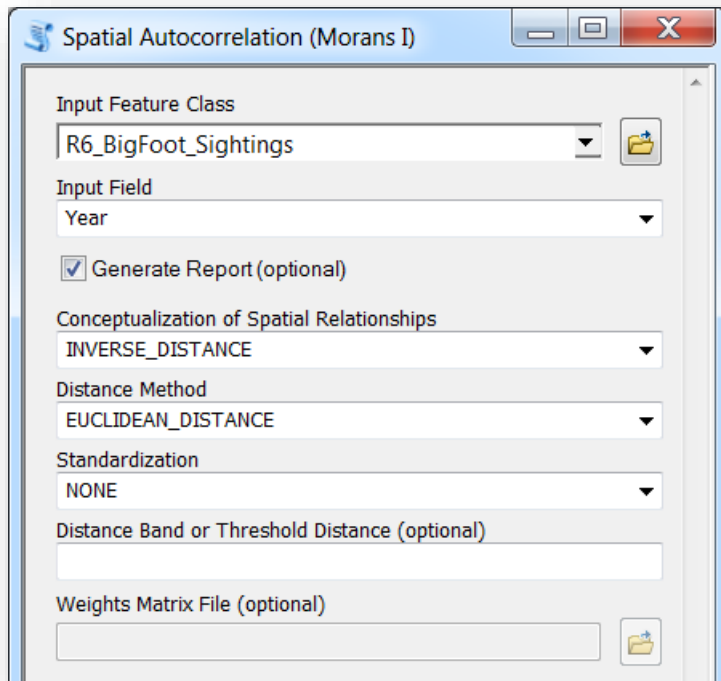
*Everything is related to everything else,  
but near things are more related than distant things.*





# Analyzing Patterns

- Spatial Autocorrelation (Moran's I) tool



*Null Hypothesis: the attribute being analyzed is randomly distributed among the features in the study area*

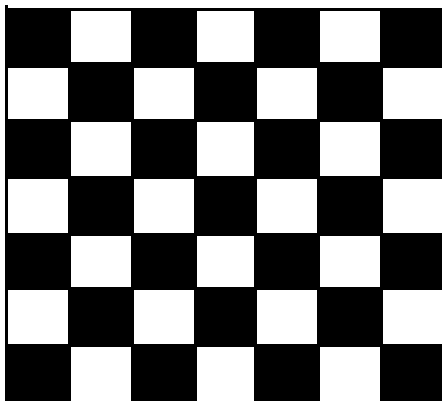
# Statistics 101

- Moran's I Index

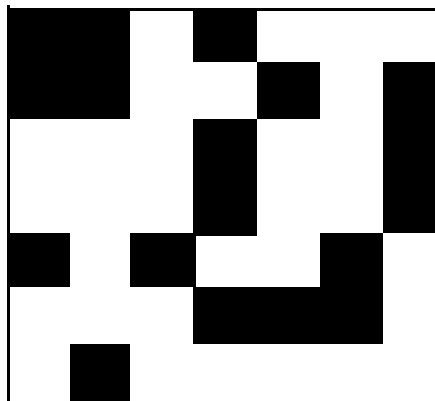
$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

- A measure of spatial autocorrelation

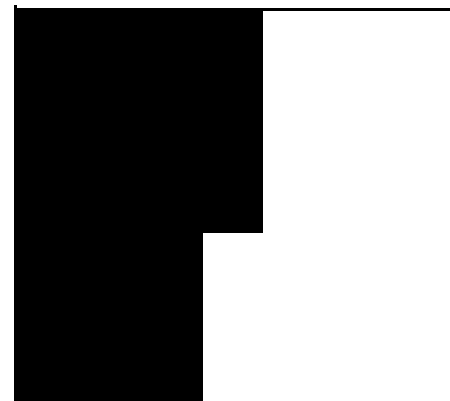
- Values range from  $-1$  (indicating perfect dispersion) to  $+1$  (perfect correlation). A zero value indicates a random spatial pattern.



**-1**



**0**



**+1**

# Statistics 101

- Expected Index

$$E(I) = \frac{-1}{N-1}$$

- Hypothetical random distribution

- Variance

$$Var(I) = \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2}$$

- Any difference from the expected index

- Z Score and P Value

- measures of statistical significance
  - tell you whether or not to reject the null hypothesis

# Spatial Autocorrelation Results

Spatial Autocorrelation (Morans I) [104947\_04012014]

Index: 0.08321

ZScore: 4.842798

PValue: 0.000001

Report File: MoransI\_Result0.html

Inputs

Environments

Messages

DOUBLE  
CLICK

RIGHT  
CLICK

Messages

```
Executing: SpatialAutocorrelation
R6_BigFoot_Sightings Year
GENERATE_REPORT INVERSE_DISTANCE
EUCLIDEAN_DISTANCE NONE # #
Start Time: Tue Apr 01 10:49:42
2014
Running script
SpatialAutocorrelation...
WARNING 000853: The default
neighborhood search threshold was
62519.7903 Meters.
```

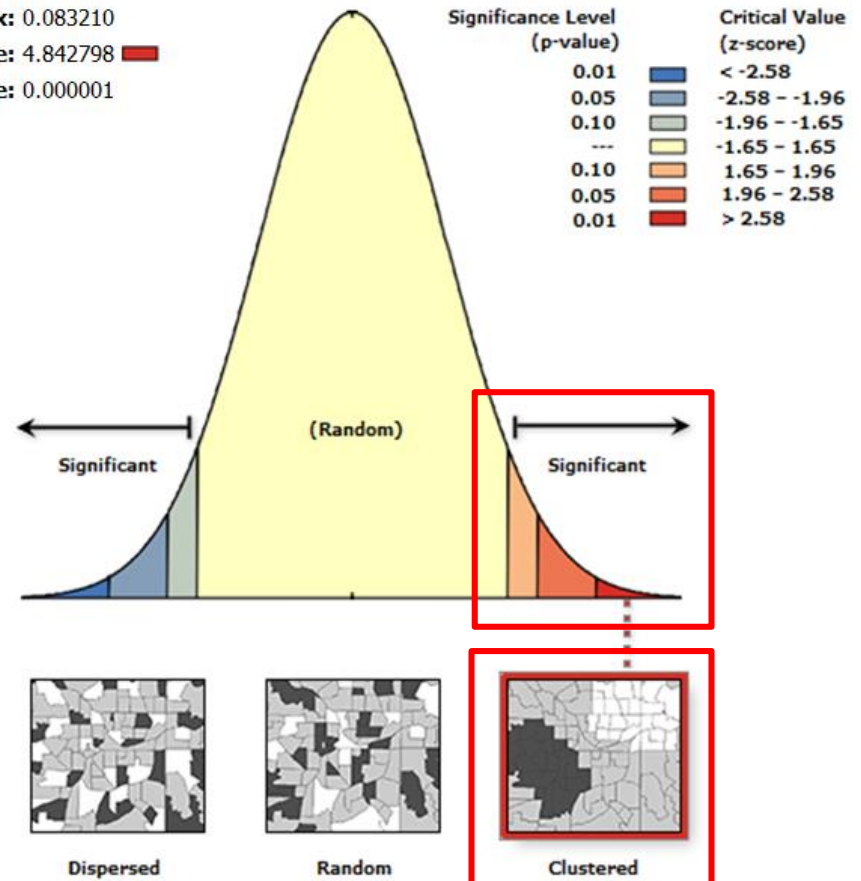
Global Moran's I Summary

Moran's Index:	0.083210
Expected Index:	-0.000695
Variance:	0.000300
z-score:	4.842798
p-value:	0.000001

Moran's Index: 0.083210

z-score: 4.842798

p-value: 0.000001



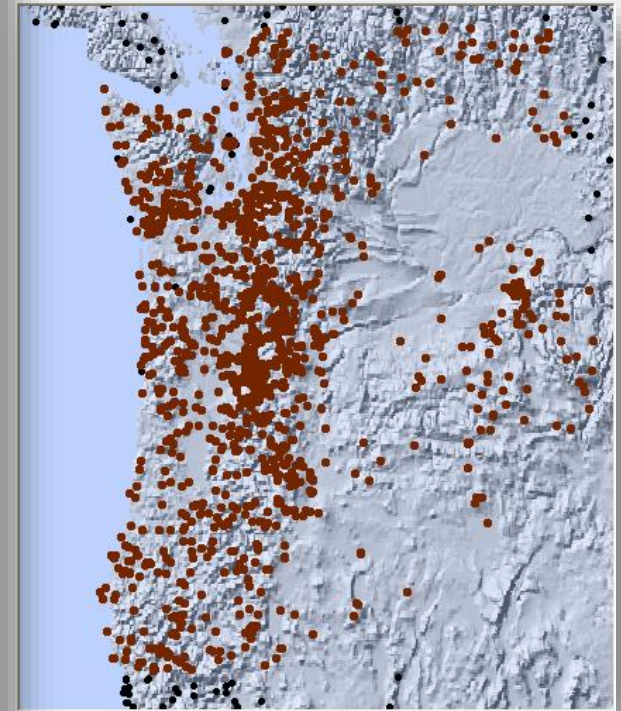
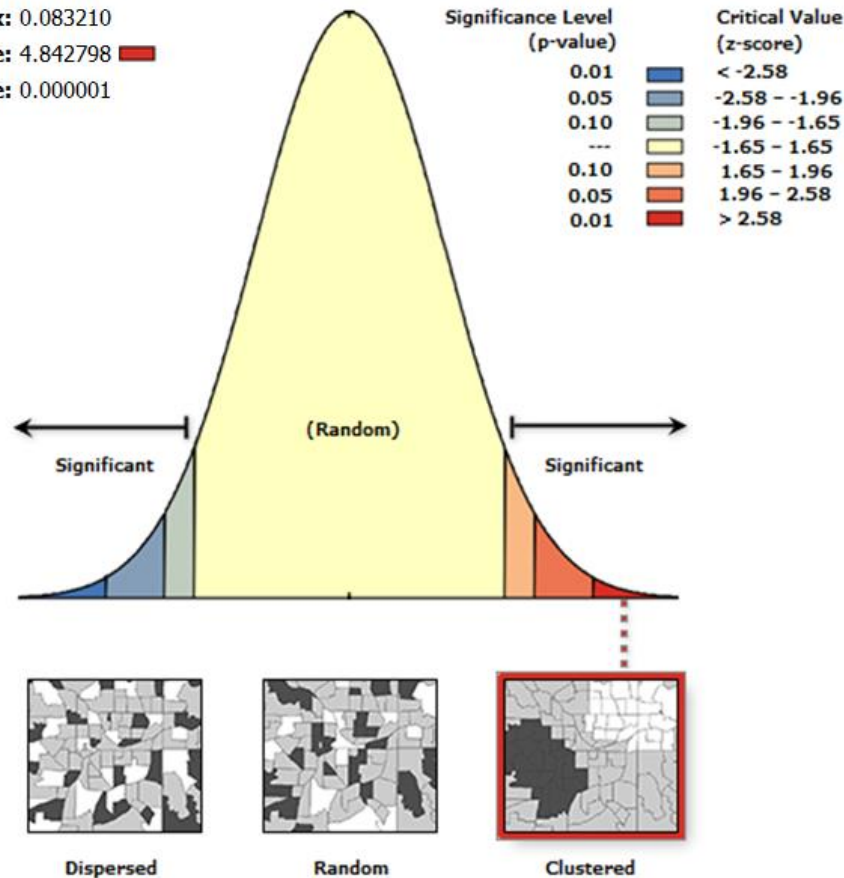
Given the z-score of 4.84, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

# Autocorrelation (Moran's I) tool

Moran's Index: 0.083210

z-score: 4.842798

p-value: 0.000001



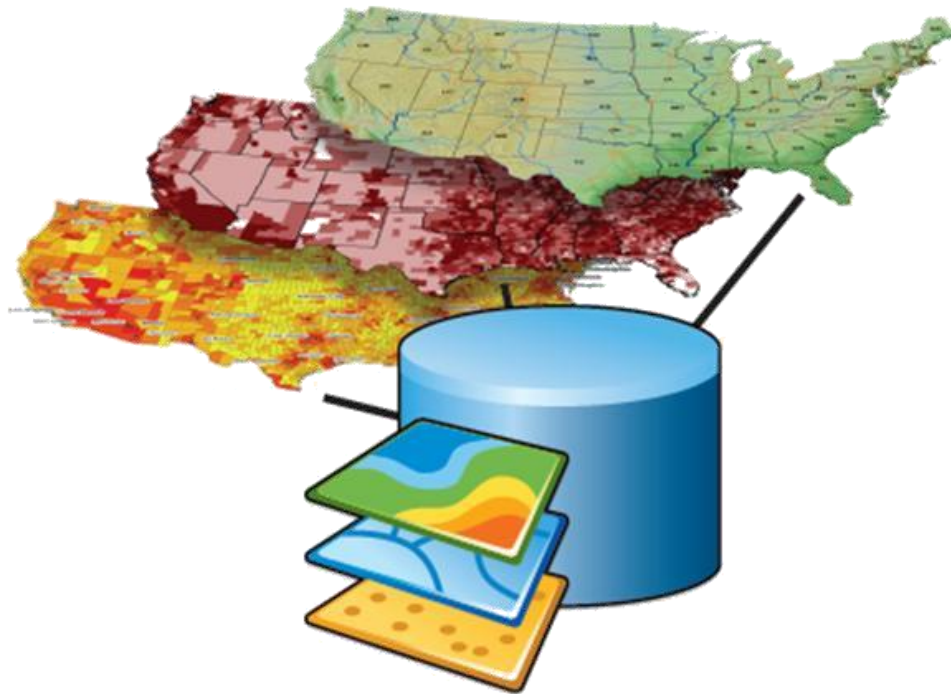
Given the z-score of 4.84, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.



# Spatial Autocorrelation Results

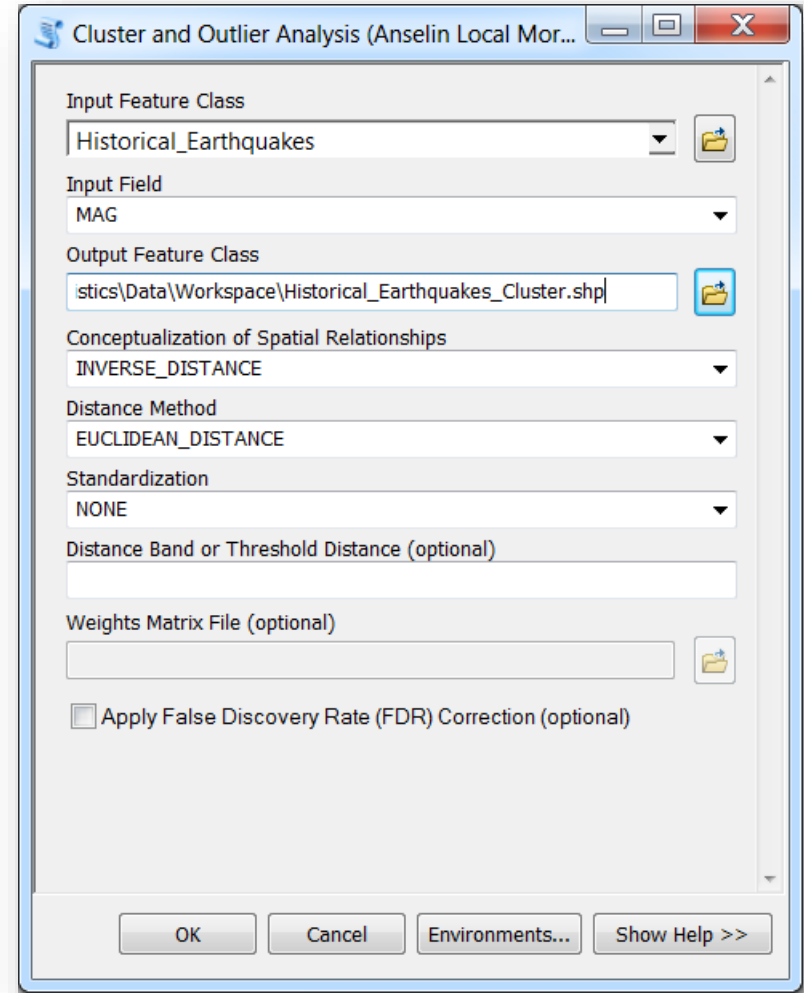
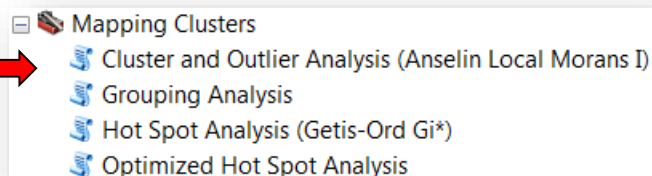
Result	Spatial Autocorrelation
The p-value is <b>not</b> statistically significant.	You cannot reject the null hypothesis. It is quite possible that the spatial distribution of feature attribute values is the result of random spatial processes. Said another way, the observed spatial pattern of values could well be one of many, many possible versions of complete spatial randomness.
The p-value <b>is</b> statistically significant, and the z-score is positive.	You may reject the null hypothesis. The spatial distribution of high values and/or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were truly random.
The p-value <b>is</b> statistically significant, and the z-score is negative.	You may reject the null hypothesis. The spatial distribution of high values and low values in the dataset is more spatially <b>dispersed</b> than would be expected if underlying spatial processes were truly random. A dispersed spatial pattern often reflects some type of competitive process: a feature with a high value repels other features with high values; similarly, a feature with a low value repels other features with low values.

# Demonstration: Spatial Pattern Analysis



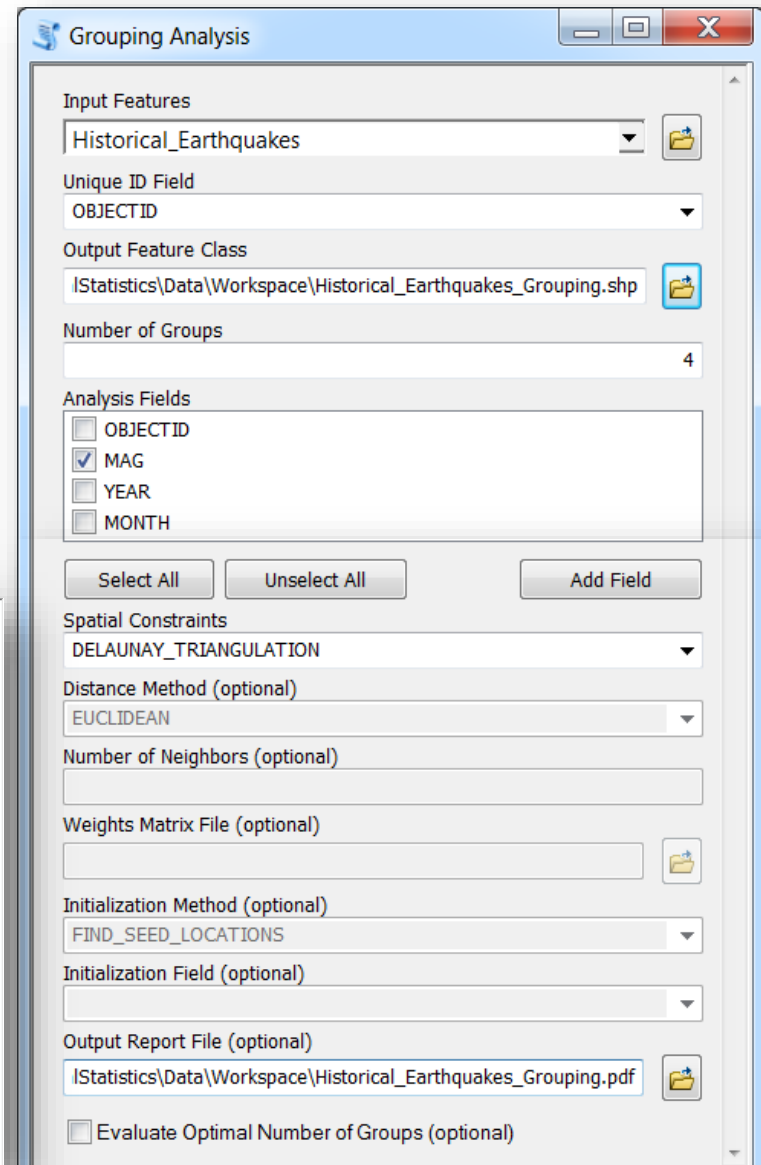
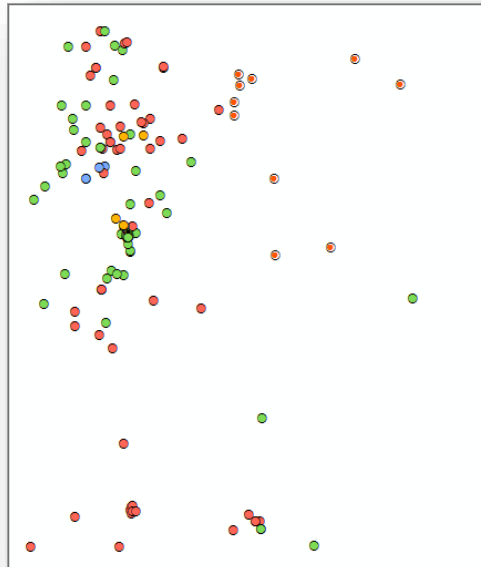
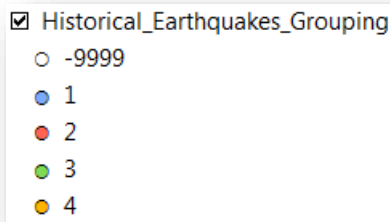
# Mapping Clusters

- Cluster and Outlier Analysis
  - Where are the clusters?
  - Where are features with similar attributes are clustered?
  - Used to identify the locations of statistically significant hot spots, cold spots, and spatial outliers



# Grouping Analysis

- Groups features based on feature attributes and optional spatial/temporal constraints.
- Identify Spawning beds
- Classify different types of soils
- Identify demographic patterns/trends



Grouping Analysis

Input Features: Historical\_Earthquakes

Unique ID Field: OBJECTID

Output Feature Class: IStatistics\Data\Workspace\Historical\_Earthquakes\_Grouping.shp

Number of Groups: 4

Analysis Fields:

- ☐ OBJECTID
- ☒ MAG
- ☐ YEAR
- ☐ MONTH

Select All Unselect All Add Field

Spatial Constraints: DELAUNAY\_TRIANGULATION

Distance Method (optional): EUCLIDEAN

Number of Neighbors (optional):

Weights Matrix File (optional):

Initialization Method (optional): FIND\_SEED\_LOCATIONS

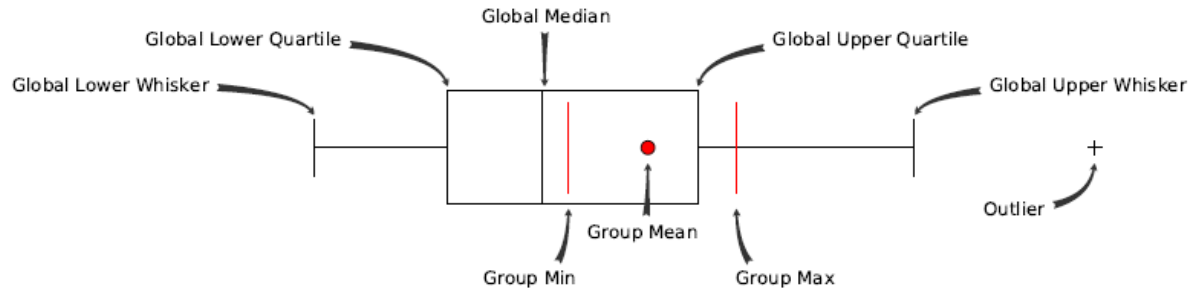
Initialization Field (optional):

Output Report File (optional): IStatistics\Data\Workspace\Historical\_Earthquakes\_Grouping.pdf

☐ Evaluate Optimal Number of Groups (optional)

# Grouping Analysis

## Group-Wise Summary



**Overall Variable Statistics: Count = 168, Std. Distance = 0.6002, SSD = 101.5051**

Variable	Mean	Std. Dev.	Min	Max	R2	
MAG	4.8137	0.5831	3.3000	7.0000	0.4297	+++ --- --- +++ + ++

**Group 1: Count = 4, Std. Distance = 0.3577, SSD = 1.5056**

Variable	Mean	Std. Dev.	Min	Max	Share	
MAG	6.4875	0.3577	5.9000	6.8000	0.2432	+++ --- --- +++ ● ++

**Group 2: Count = 57, Std. Distance = 0.6183, SSD = 64.0923**

Variable	Mean	Std. Dev.	Min	Max	Share	
MAG	5.0935	0.6183	3.8000	7.0000	0.8649	+++ --- --- +++ ● ++

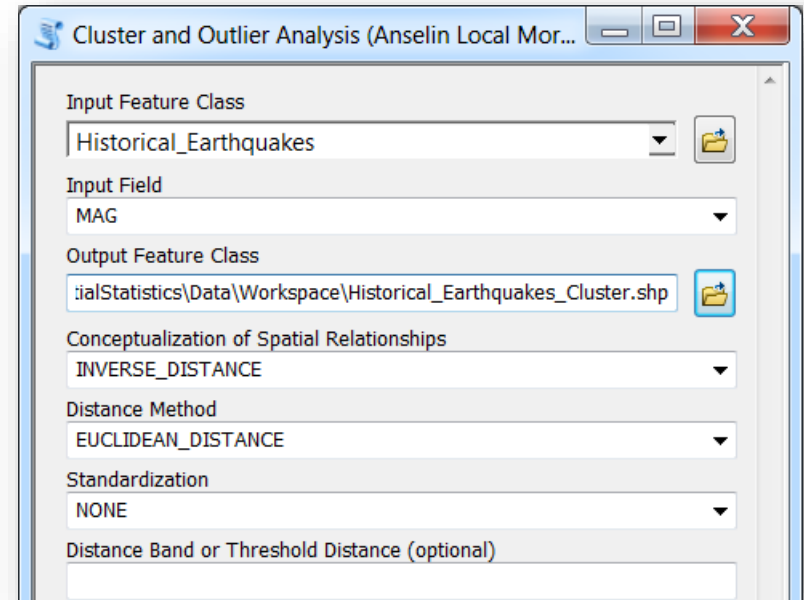
**Group 3: Count = 101, Std. Distance = 0.3417, SSD = 34.6913**

Variable	Mean	Std. Dev.	Min	Max	Share	
MAG	4.6691	0.3417	3.3000	5.8000	0.6757	+++ --- --- +++ ● ++

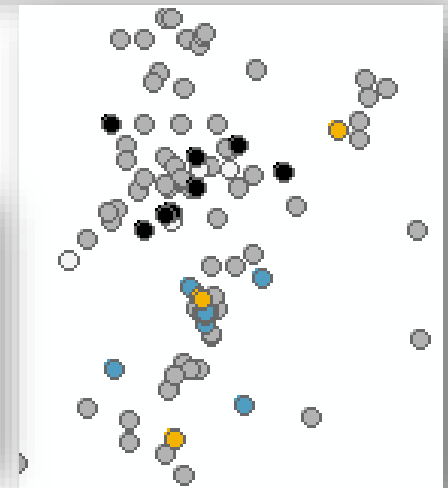


# Mapping Clusters

- Cluster and Outlier Analysis
  - COType
    - Cluster of high values (HH)
    - Cluster of low values (LL)
    - Outlier in which a high value is surrounded primarily by low values (HL)
    - Outliers in which a low value is surrounded primarily by high values (LH)

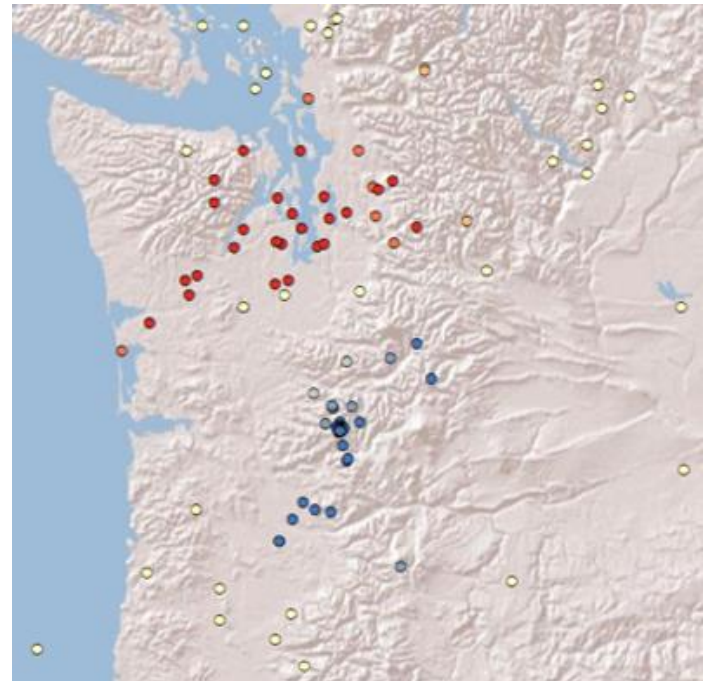


- ☒ Historical\_Earthquakes\_Cluster
  - Not Significant
  - High-High Cluster
  - High-Low Outlier
  - Low-High Outlier
  - Low-Low Cluster



# Hot Spot Analysis

- Optimized Hot Spot Analysis
  - High positive Z-score, Small P-value – “**HOT**” spots
  - Low negative Z-score, Small P-value – “**COLD**” spots



# Hot Spot Analysis

- Hot Spot Analysis (Getis-Ord  $G_i^*$ )
  - Identifies clustering in both high and low values
  - The  $G_i$ \_Bin field identifies statistically significant hot and cold spots.
  - Displays confidence level +/- 3 bins, 99% 95% and 90%



# Review / Q&A

- Spatial Pattern Analysis
  - Analyzing Patterns toolset
    - What is the overall pattern of my dataset?
    - Global analysis
  - Mapping Clusters toolset
    - Where are the clusters?
    - Local analysis

# Demonstration: Mapping Clusters

