What are Spatial Statistics?

Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.

proximity

area

length

orientation
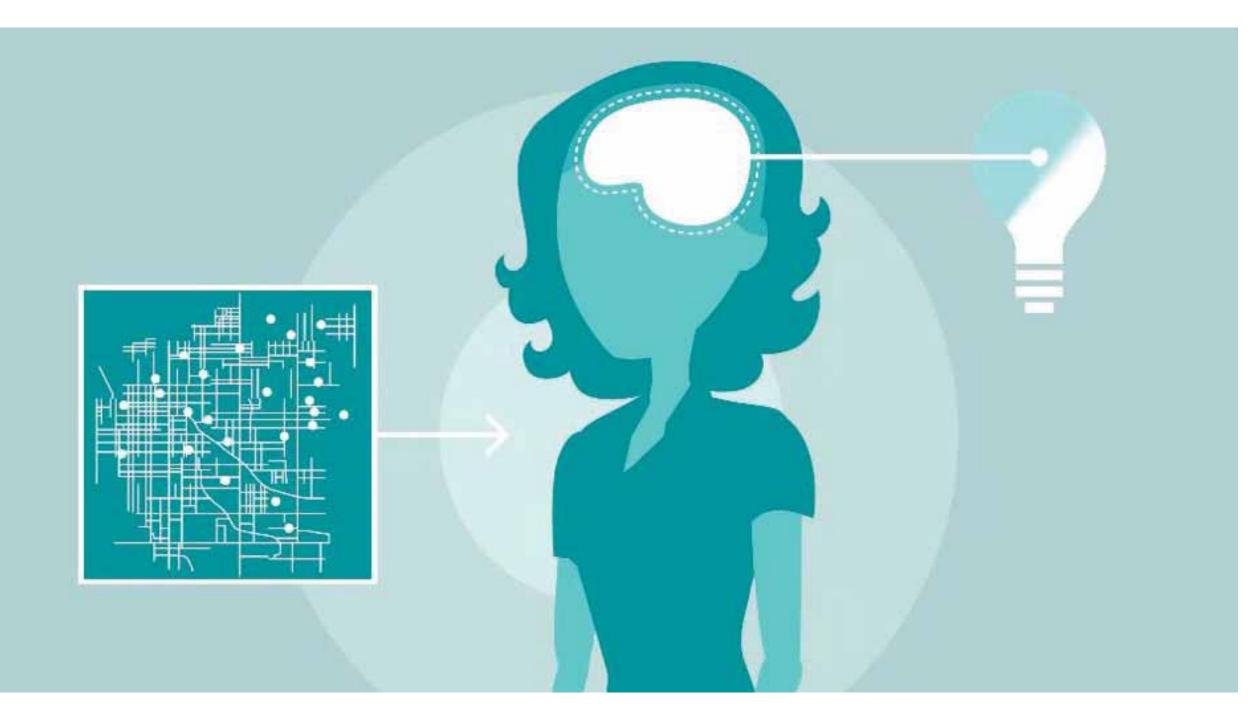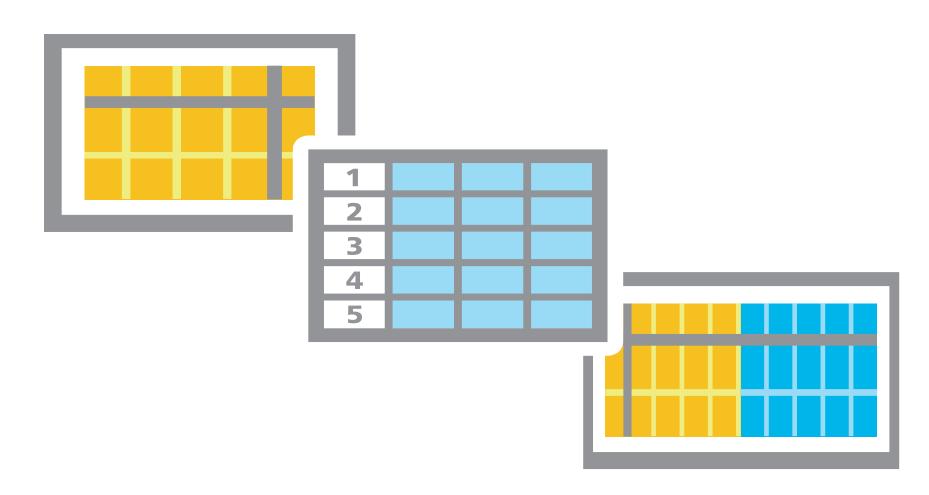
connectivity

coincidence

distance

direction

Spreadsheets

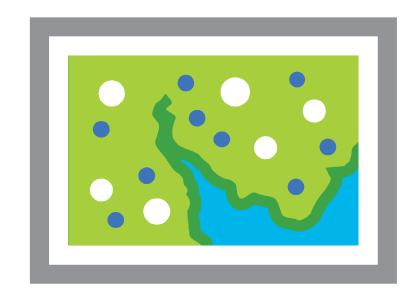## Data or Information?

Maps

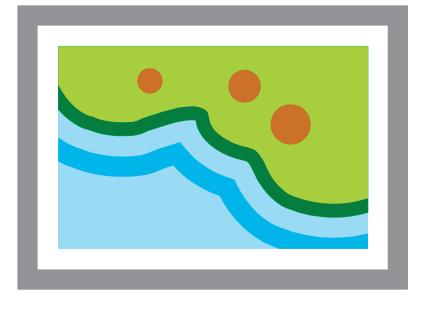## Data or Information?
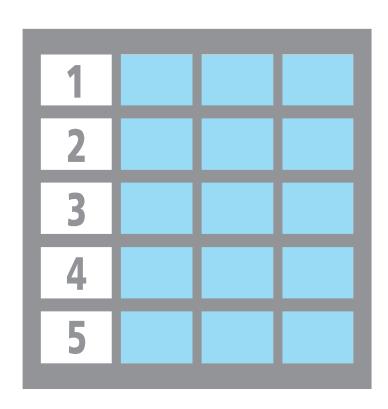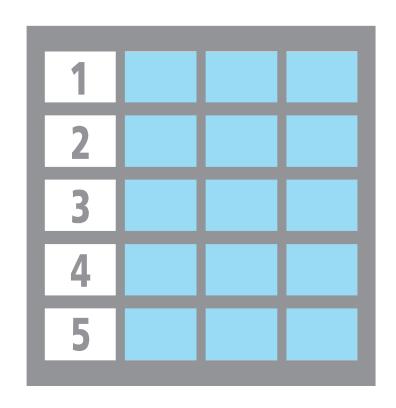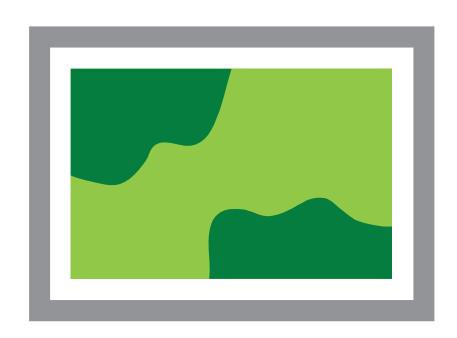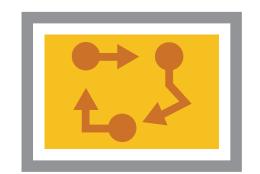
You ask for more
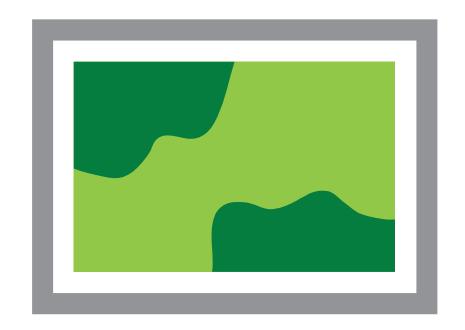
- Mean
- Standard Deviations
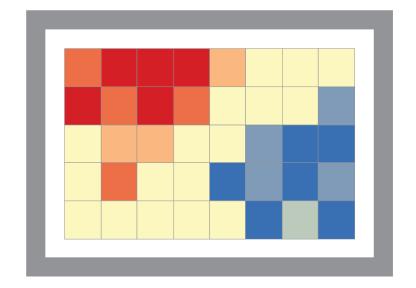- Min and Max
- …

Same goes for maps!
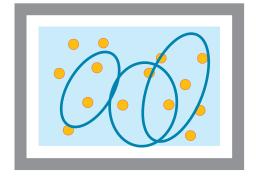
We can do more

- Spatial Statistics Tools
  - Analyzing Patterns
    - Average Nearest Neighbor
    - High/Low Clustering (Getis-Ord General G)
    - Incremental Spatial Autocorrelation
    - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    - Spatial Autocorrelation (Morans I)
  - Mapping Clusters
    - Cluster and Outlier Analysis (Anselin Local Morans I)
    - Grouping Analysis
    - Hot Spot Analysis (Getis-Ord Gi*)
    - Optimized Hot Spot Analysis
    - Similarity Search
  - Measuring Geographic Distributions
    - Central Feature
    - Directional Distribution (Standard Deviational Ellipse)
    - Linear Directional Mean
    - Mean Center
    - Median Center
    - Standard Distance
  - Modeling Spatial Relationships
    - Exploratory Regression
    - Generate Network Spatial Weights
    - Generate Spatial Weights Matrix
    - Geographically Weighted Regression
    - Ordinary Least Squares

- Spatial Statistics Tools
  - Analyzing Patterns
    - Average Nearest Neighbor
    - High/Low Clustering (Getis-Ord General G)
    - Incremental Spatial Autocorrelation
    - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    - Spatial Autocorrelation (Morans I)
  - Mapping Clusters
    - Cluster and Outlier Analysis (Anselin Local Morans I)
    - Grouping Analysis
    - Hot Spot Analysis (Getis-Ord Gi*)
    - Optimized Hot Spot Analysis
    - Similarity Search
  - Measuring Geographic Distributions
    - Central Feature
    - Directional Distribution (Standard Deviational Ellipse)
    - Linear Directional Mean
    - Mean Center
    - Median Center
    - Standard Distance
  - Modeling Spatial Relationships
    - Exploratory Regression
    - Generate Network Spatial Weights
    - Generate Spatial Weights Matrix
    - Geographically Weighted Regression
    - Ordinary Least Squares

python
script!

```
OptimizedHotSpotAnalysis - Notepad

File   Edit   Format   View   Help

        #### Remove Locations Outside Boundary FC ####
        featureLayer = "ClippedPointFC"
        DM.MakeFeatureLayer(tempFC, featureLayer)
        if self.boundaryFC:
            msg = ARCPY.GetIDMessage(84454)
            ARCPY.SetProgressor("default", msg)
            DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                     self.boundaryFC, "#",
                                     "NEW_SELECTION")
            DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                "#", "#", "SWITCH_SELECTION")
            DM.DeleteFeatures(featureLayer)
        else:
            if additionalZeroDistScale == "ALL":
                msg = ARCPY.GetIDMessage(84455)
                ARCPY.SetProgressor("default", msg)
                DM.SelectLayerByAttribute(featureLayer, "NEW_SELECTION",
                                    '"Join_Count" = 0')
                DM.DeleteFeatures(featureLayer)

            else:
                distance = additionalZeroDistScale * fish.quadLength
                distanceStr = self.ssdo.distanceInfo.linearUnitString(distance,
                                                        convert = True)
                nativeStr = self.ssdo.distanceInfo.printDistance(distance)
                msg = "Removing cells further than %s from input pointsd...."
                ARCPY.AddMessage(msg % nativeStr)
                DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                         self.ssdo.inputFC, distanceStr,
                                         "NEW_SELECTION")
                DM.SelectLayerByLocation(featureLayer, "INTERSECT",
                                    "#", "#", "SWITCH_SELECTION")
                DM.DeleteFeatures(featureLayer)

        DM.Delete(featureLayer)
        del collSSDO

        ARCPY.env.extent = oldExtent
```
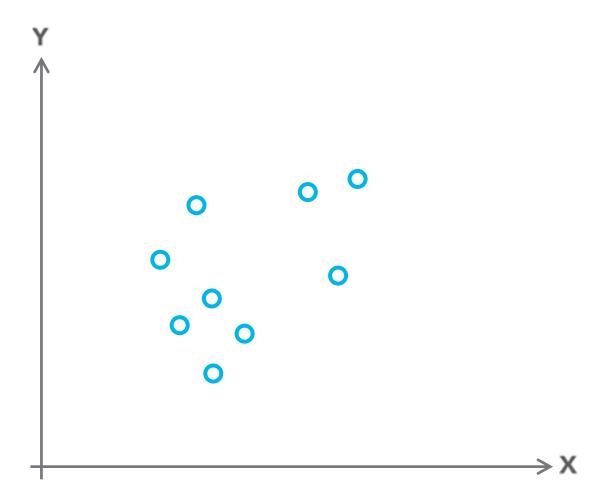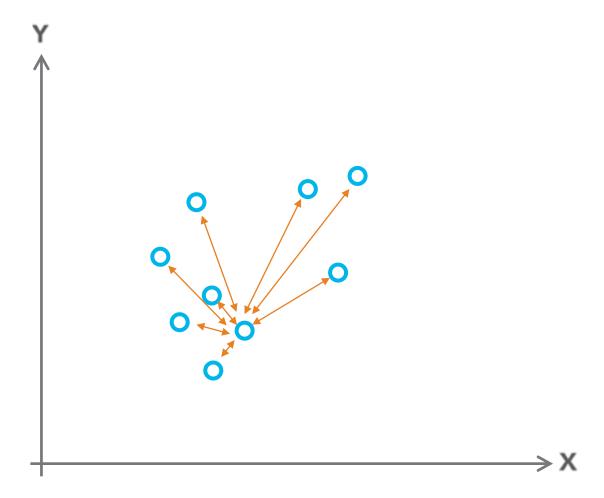
Measuring Geographic Distributions

Spatial Statistics Tools
  Analyzing Patterns
    Average Nearest Neighbor
    High/Low Clustering (Getis-Ord General G)
    Incremental Spatial Autocorrelation
    Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    Spatial Autocorrelation (Morans I)
  Mapping Clusters
    Cluster and Outlier Analysis (Anselin Local Morans I)
    Grouping Analysis
    Hot Spot Analysis (Getis-Ord Gi*)
    Optimized Hot Spot Analysis
    Similarity Search
  Measuring Geographic Distributions
    Central Feature
    Directional Distribution (Standard Deviational Ellipse)
    Linear Directional Mean
    Mean Center
    Median Center
    Standard Distance
  Modeling Spatial Relationships
    Exploratory Regression
    Generate Network Spatial Weights
    Generate Spatial Weights Matrix
    Geographically Weighted Regression
    Ordinary Least Squares
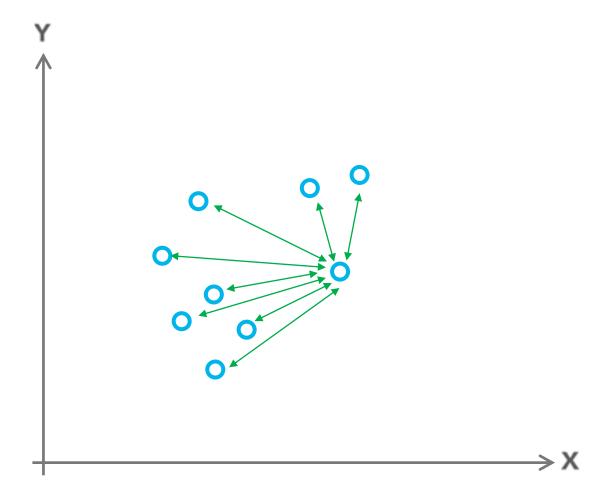
These tools address questions such as Where's the center? What's the shape and orientation? How dispersed are the features?
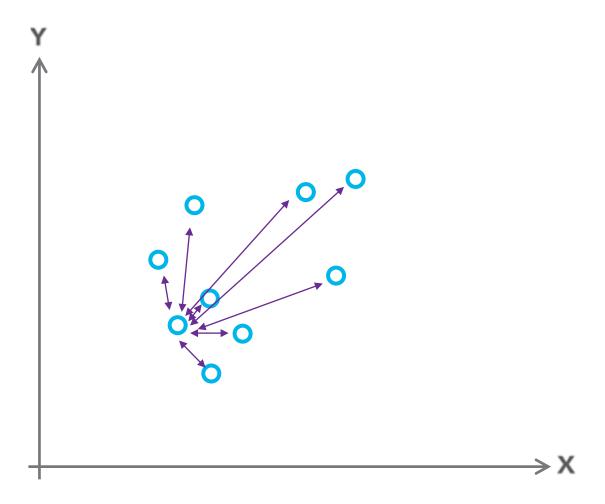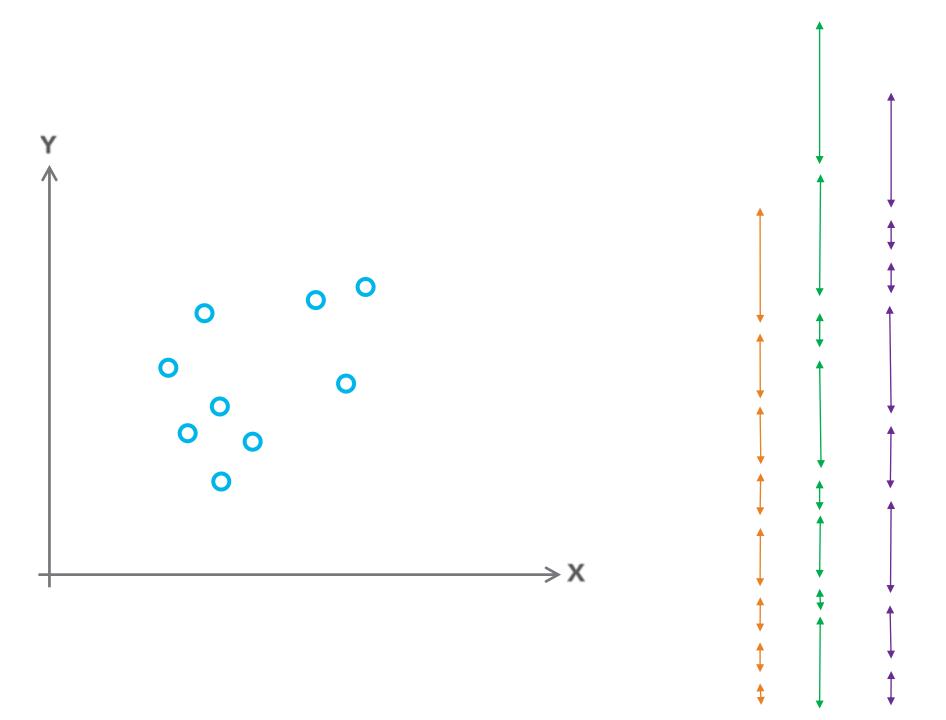
Central Feature

identifies the most centrally located feature
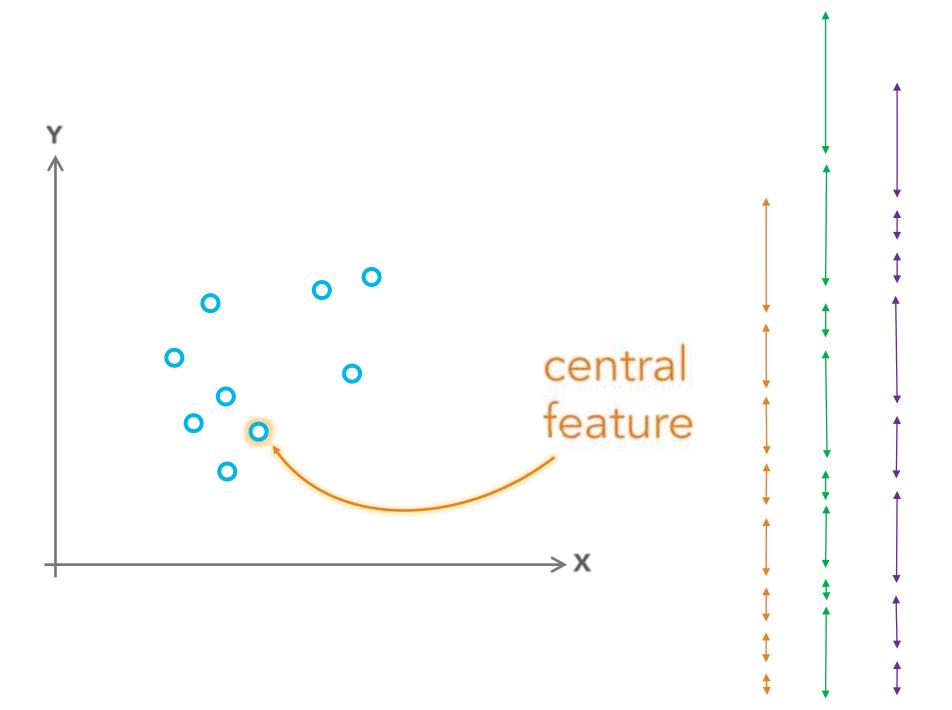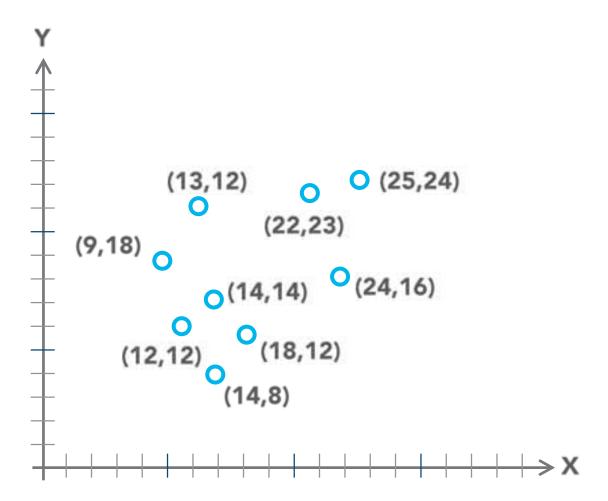in a point, line, or polygon feature class

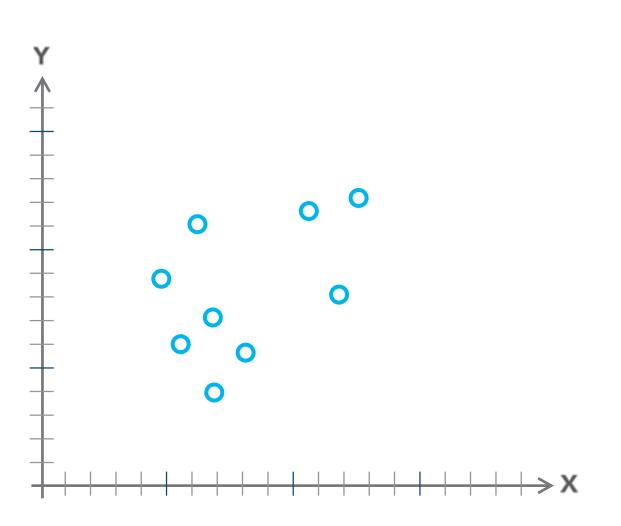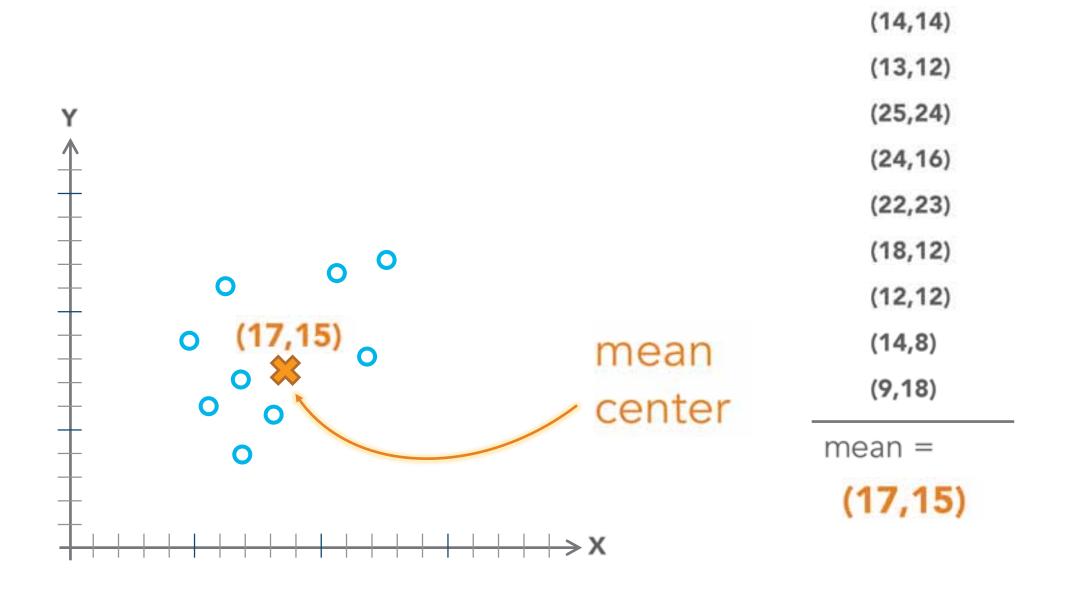central feature

Mean Center

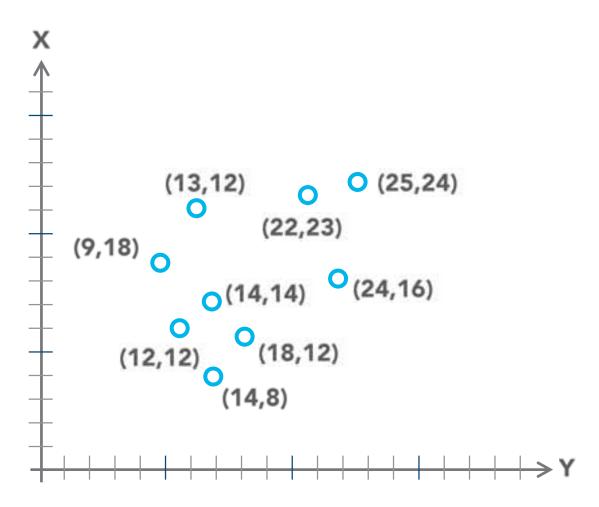identifies the geographic center (or the center of concentration) for a set of features

(14,14)
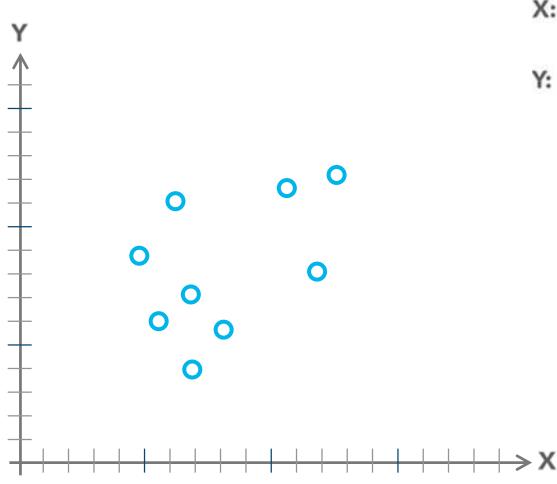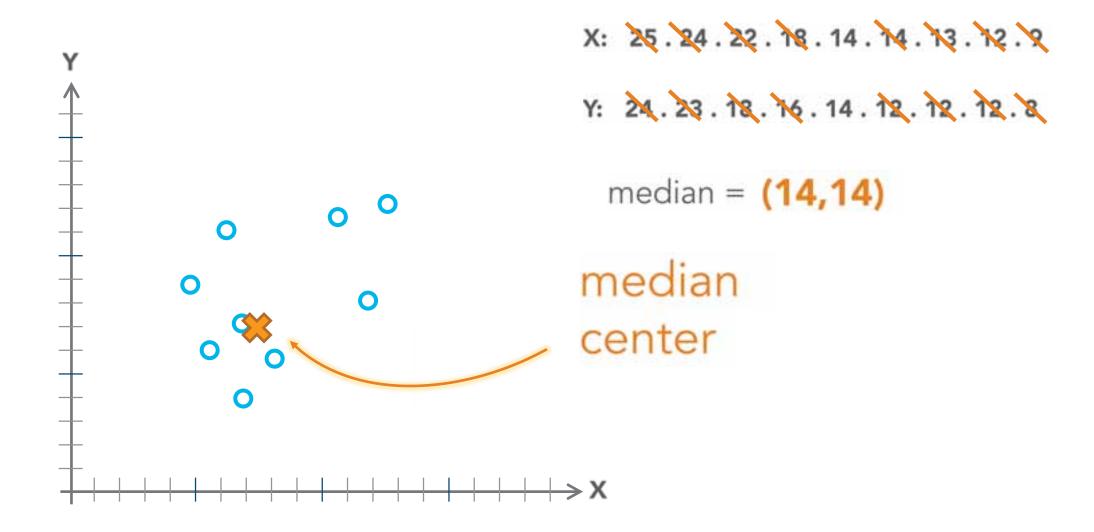
(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

mean =

(17,15)

Y

(17,15)

mean center

(14,14)
(13,12)
(25,24)
(24,16)
(22,23)
(18,12)
(12,12)
(14,8)
(9,18)

mean =

(17,15)

X

Median Center

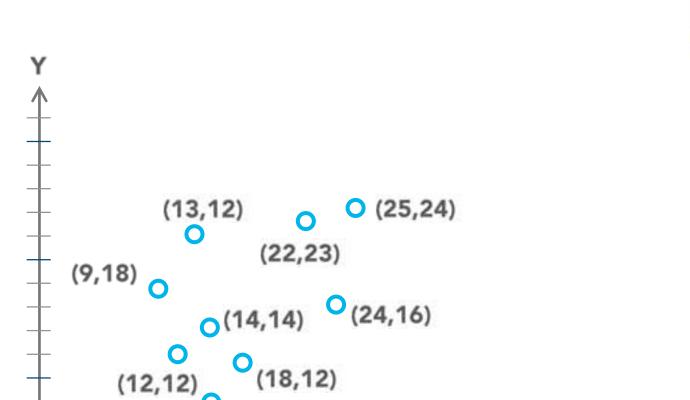identifies the location that minimizes overall Euclidean distance to the features in a dataset
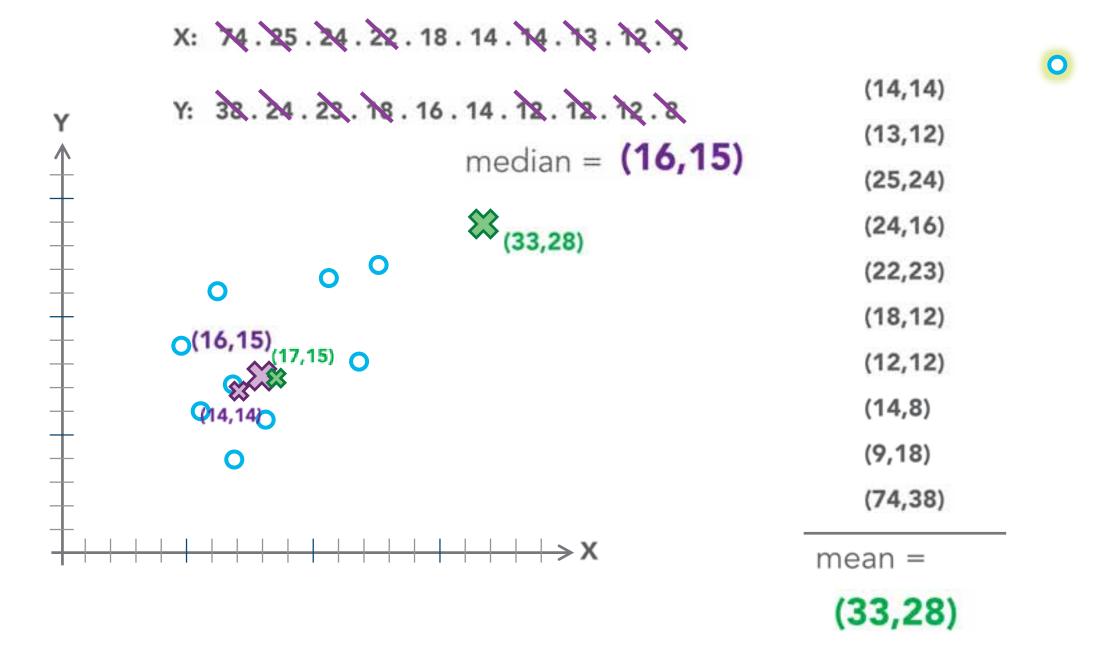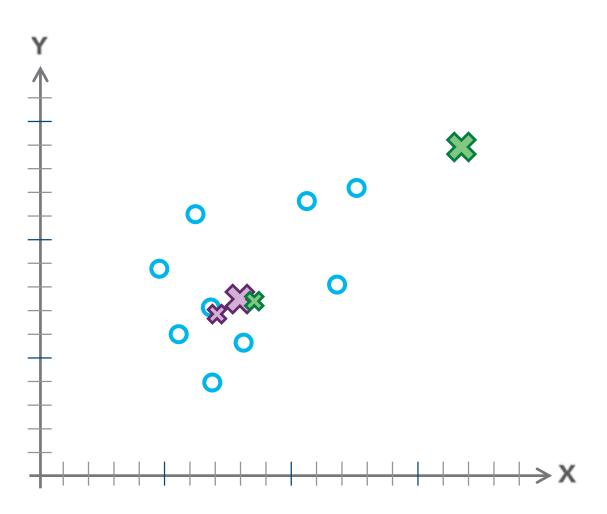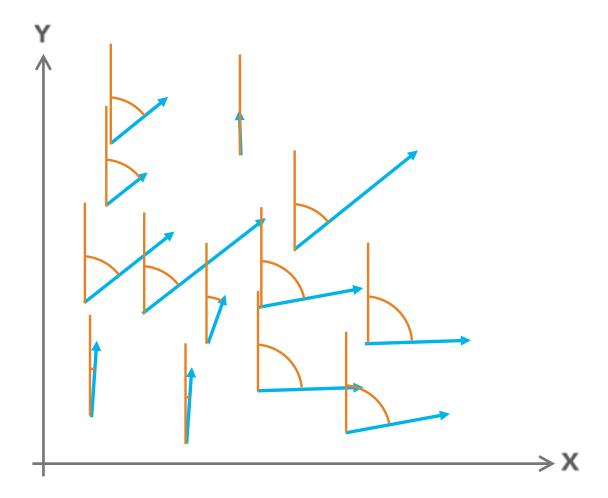
Y
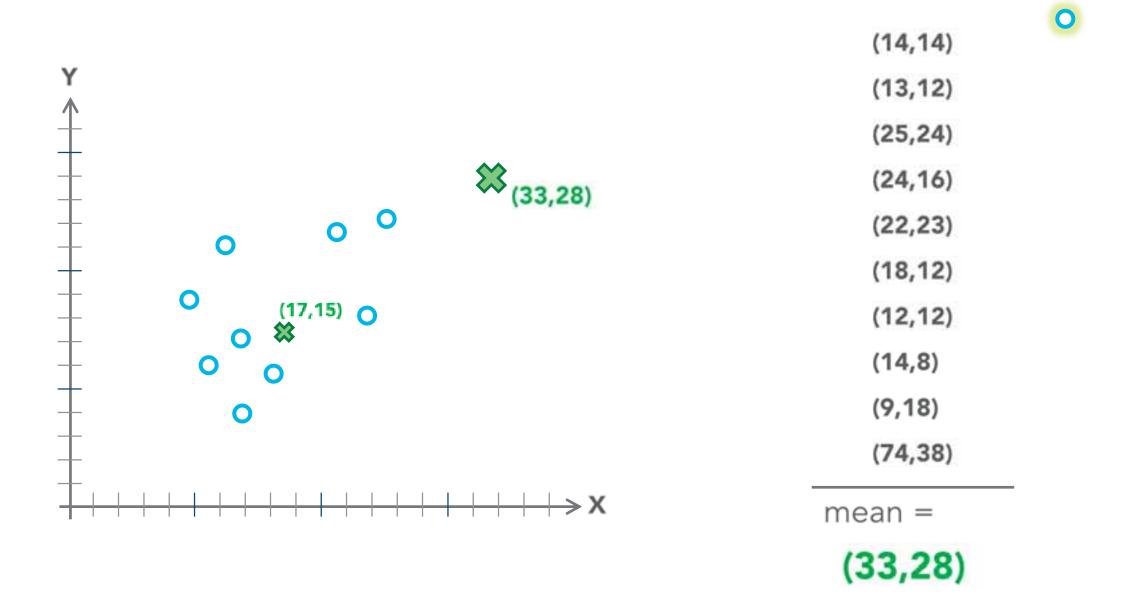
X: 25 . 24 . 22 . 18 . 14 . 14 . 13 . 12 . 9

Y: 24 . 23 . 18 . 16 . 14 . 12 . 12 . 12 . 8

X

X: ~~25~~ . ~~24~~ . ~~22~~ . ~~18~~ . 14 . ~~14~~ . ~~13~~ . ~~12~~ . ~~9~~

Y: ~~24~~ . ~~23~~ . ~~18~~ . ~~16~~ . 14 . ~~12~~ . ~~12~~ . ~~12~~ . ~~8~~

median = **(14,14)**

median center

Mean vs Median?

(14,14)
(13,12)
(25,24)
(24,16)
(22,23)
(18,12)
(12,12)
(14,8)
(9,18)
(74,38)

mean =

(33,28)

X: ~~74~~ . ~~25~~ . ~~24~~ . ~~22~~ . 18 . 14 . ~~14~~ . ~~13~~ . ~~12~~ . ~~9~~

Y: ~~38~~ . ~~24~~ . ~~23~~ . ~~18~~ . 16 . 14 . ~~12~~ . ~~12~~ . ~~12~~ . ~~8~~

median = **(16,15)**

X: ~~74~~ . ~~25~~ . ~~24~~ . ~~22~~ . 18 . 14 . ~~14~~ . ~~13~~ . ~~12~~ . ~~9~~

Y: ~~38~~ . ~~24~~ . ~~23~~ . ~~18~~ . 16 . 14 . ~~12~~ . ~~12~~ . ~~12~~ . ~~8~~

median = **(16,15)**

(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

(74,38)

mean =

(33,28)
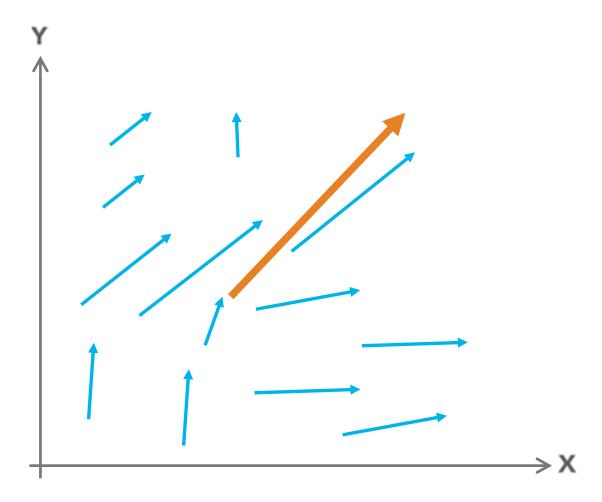
Linear Directional Mean

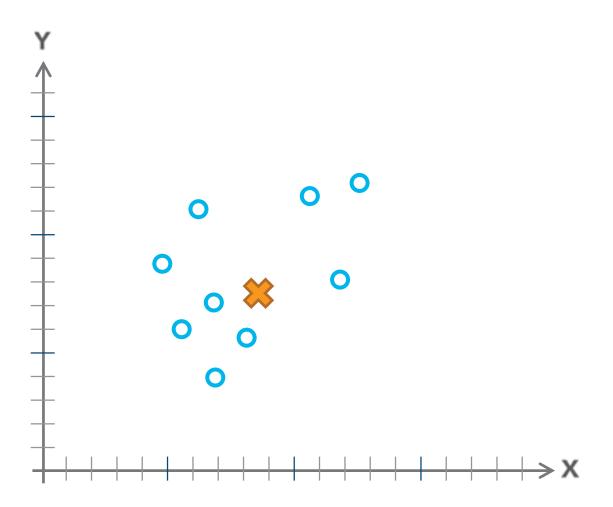identifies the mean direction, length, and geographic center for a set of lines
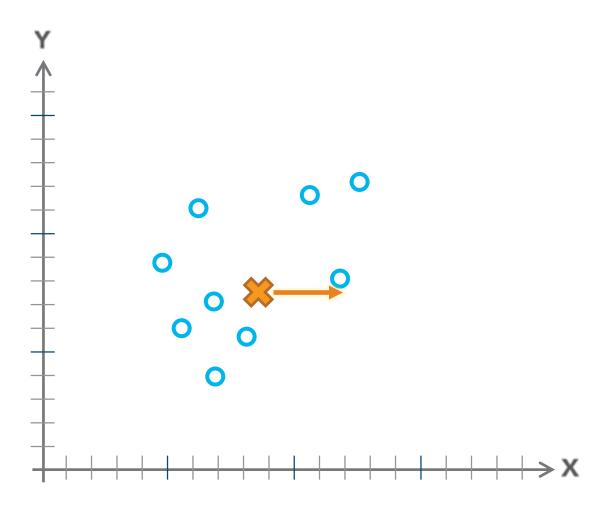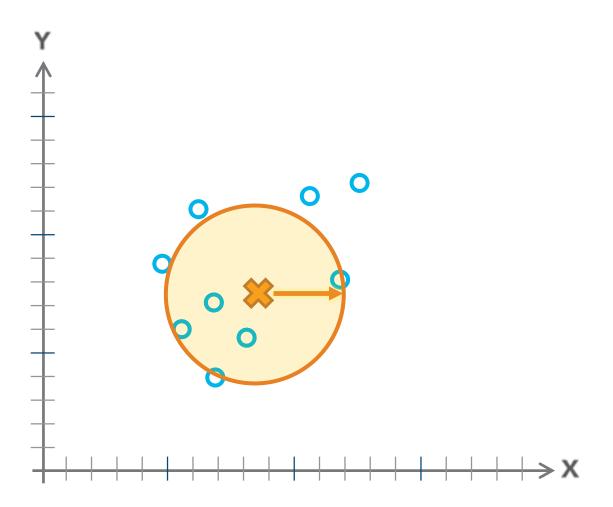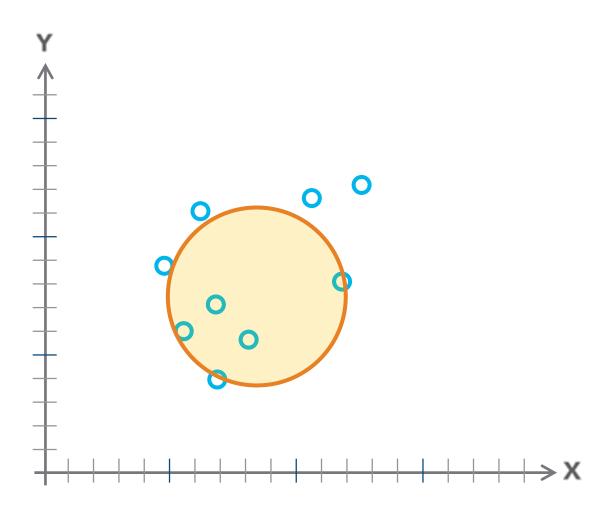
Standard Distance

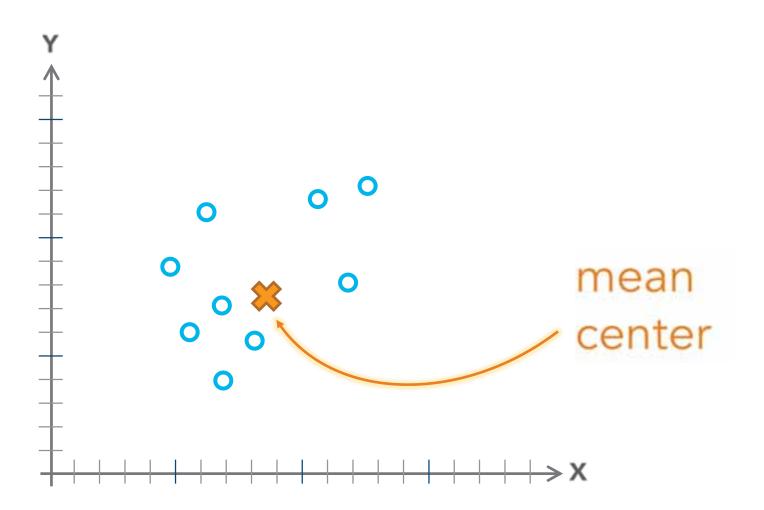measures the degree to which features are concentrated or dispersed around the geometric mean center
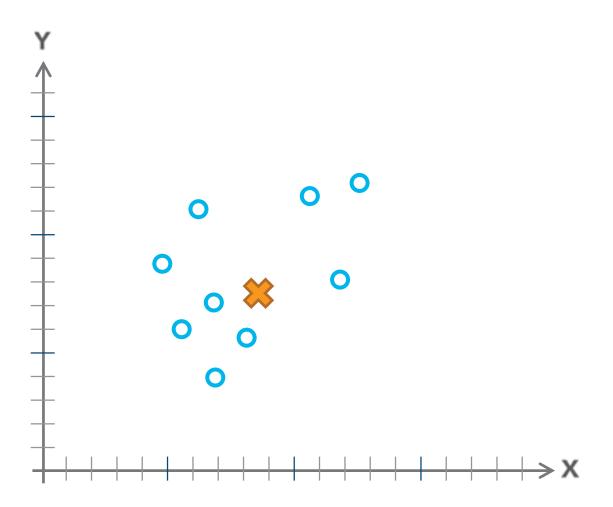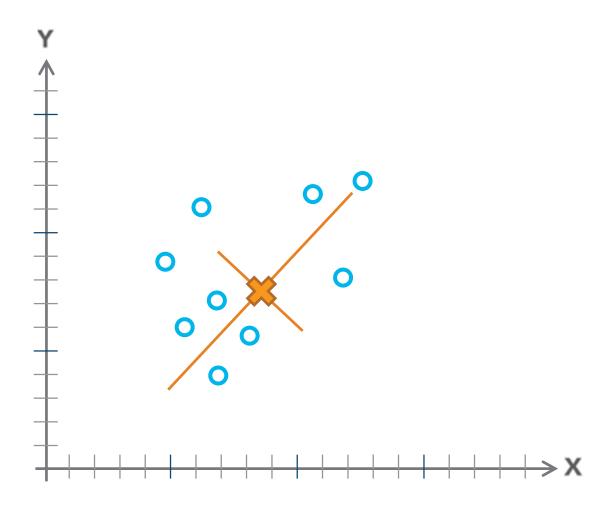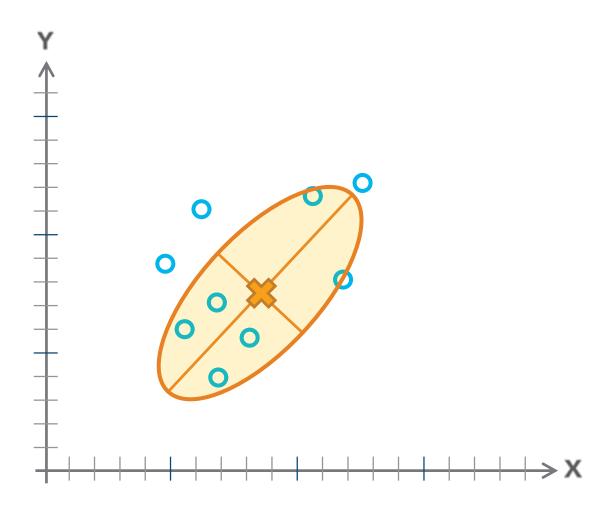
Directional Distribution
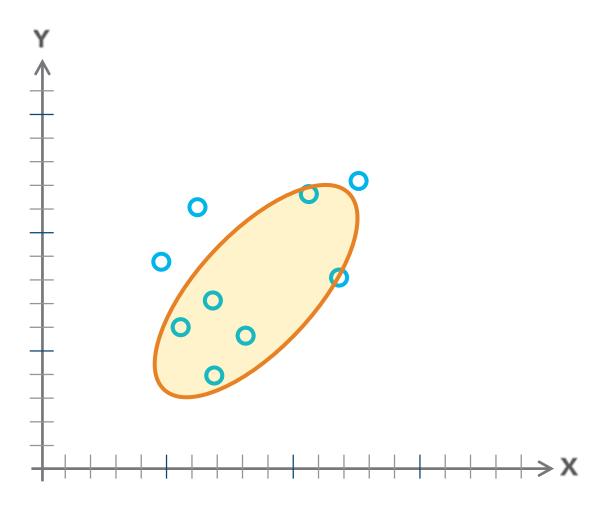## (Standard Deviational Ellipse)

creates standard deviational ellipses to summarize the spatial characteristics of geographic features: central tendency, dispersion, and directional trends
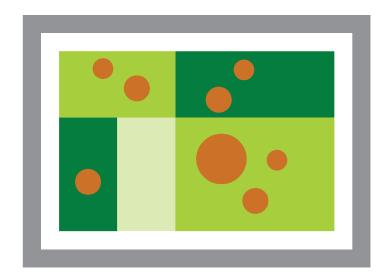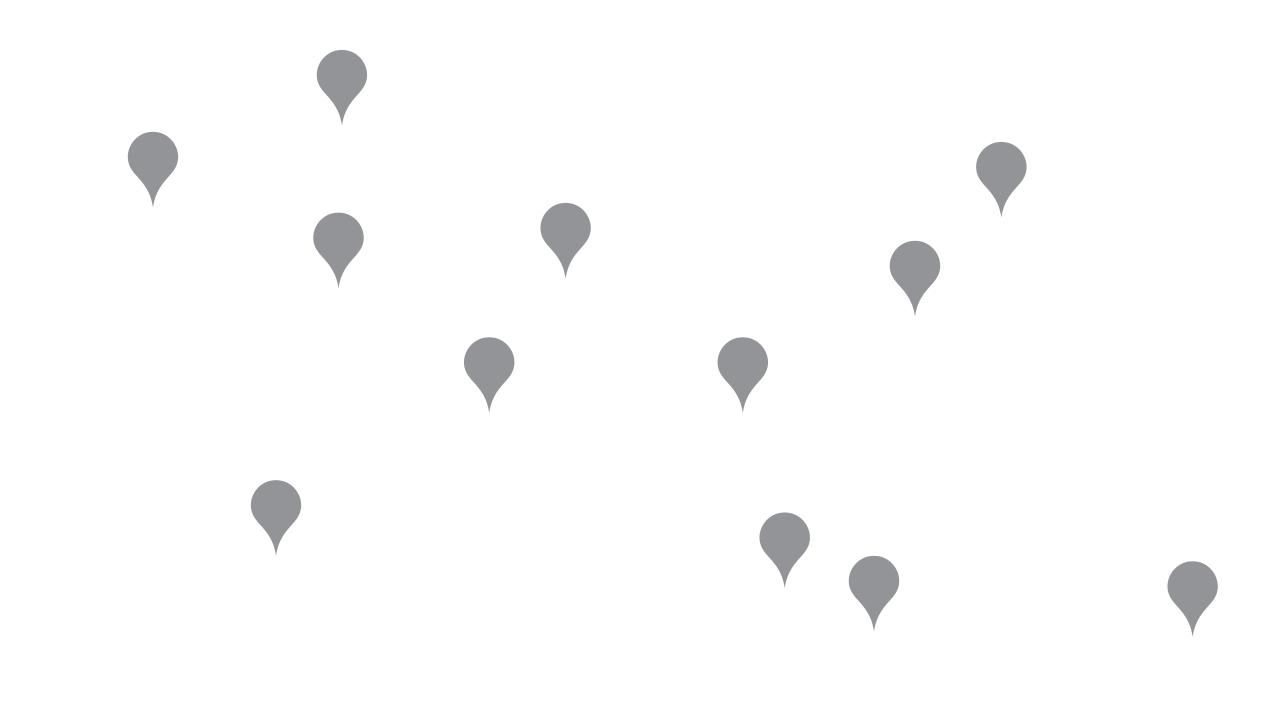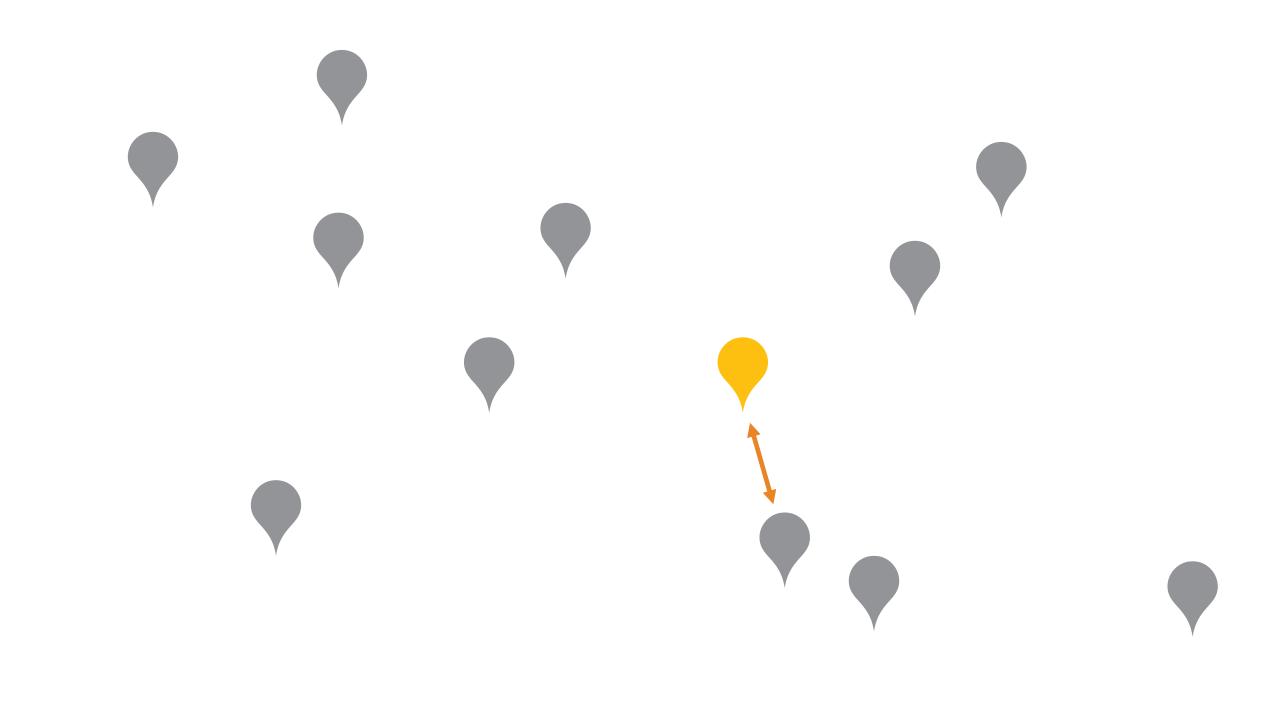
Analyzing Patterns

Spatial Statistics Tools
- Analyzing Patterns
  - Average Nearest Neighbor
  - High/Low Clustering (Getis-Ord General G)
  - Incremental Spatial Autocorrelation
  - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
  - Spatial Autocorrelation (Morans I)
- Mapping Clusters
  - Cluster and Outlier Analysis (Anselin Local Morans I)
  - Grouping Analysis
  - Hot Spot Analysis (Getis-Ord Gi*)
  - Optimized Hot Spot Analysis
  - Similarity Search
- Measuring Geographic Distributions
  - Central Feature
  - Directional Distribution (Standard Deviational Ellipse)
  - Linear Directional Mean
  - Mean Center
  - Median Center
  - Standard Distance
- Modeling Spatial Relationships
  - Exploratory Regression
  - Generate Network Spatial Weights
  - Generate Spatial Weights Matrix
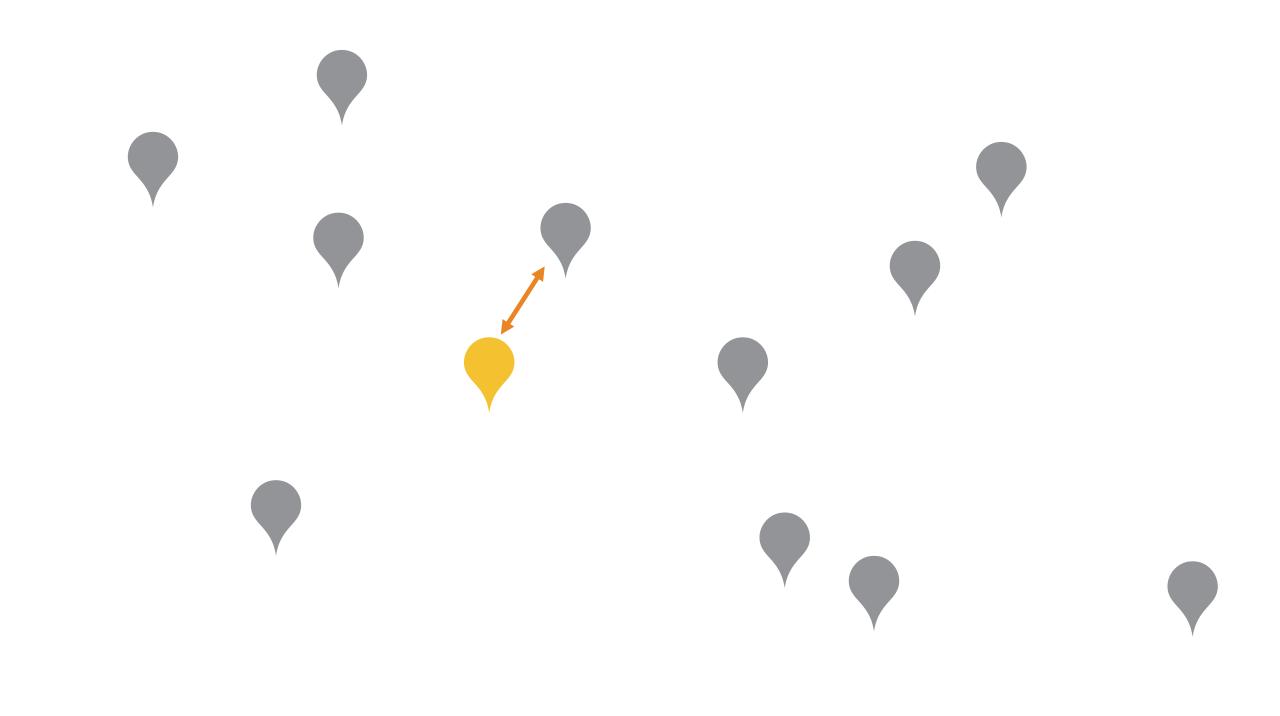 - Geographically Weighted Regression
  - Ordinary Least Squares

These tools evaluate if features, or the values associated with features, form a clustered, dispersed, or random spatial pattern.
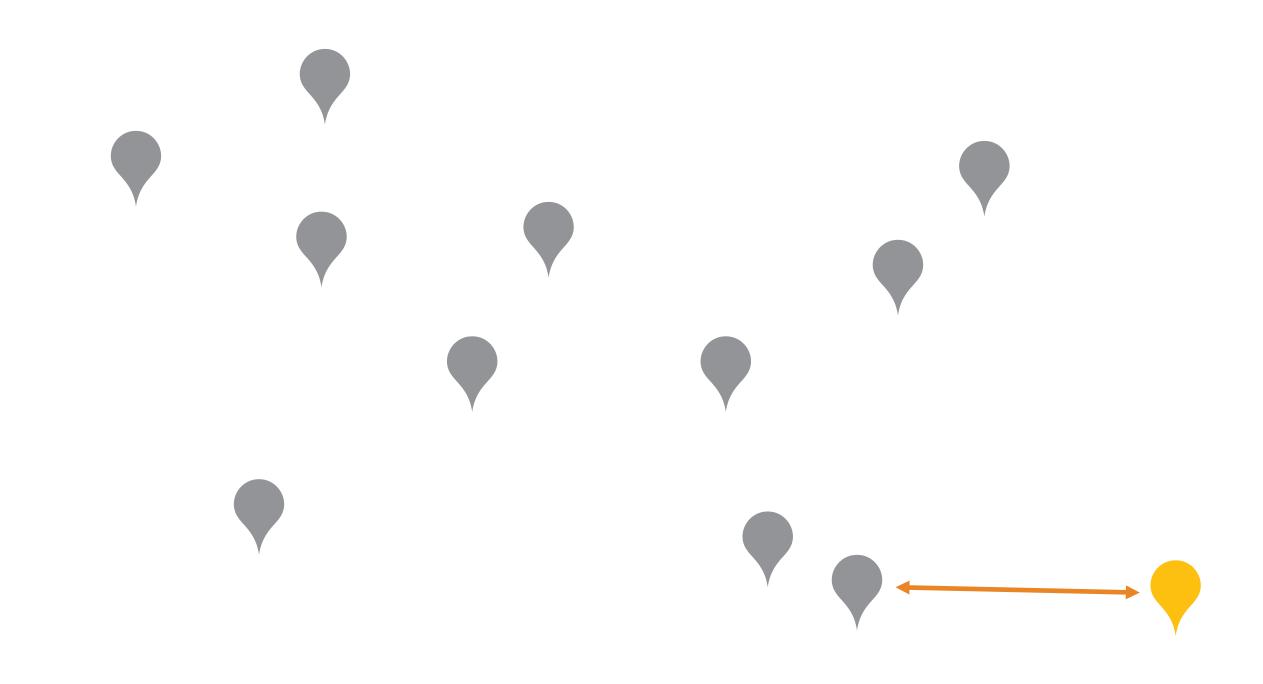
Average Nearest Neighbor
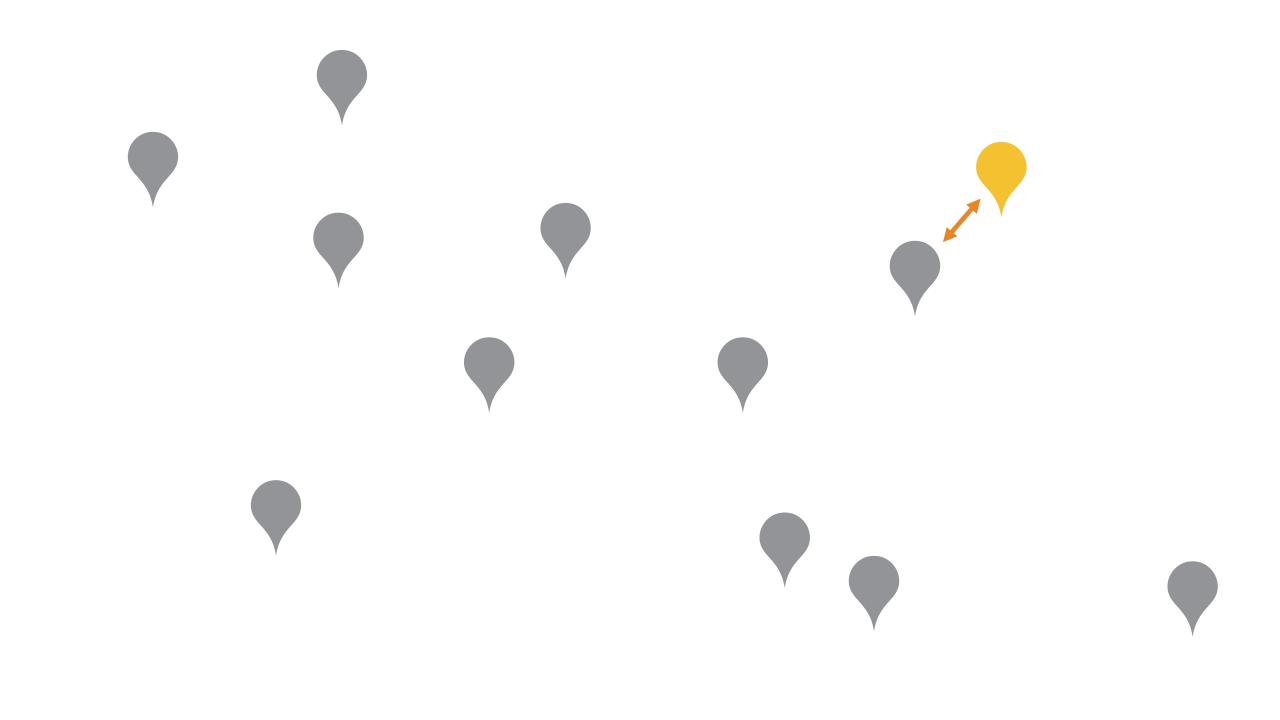
calculates a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature

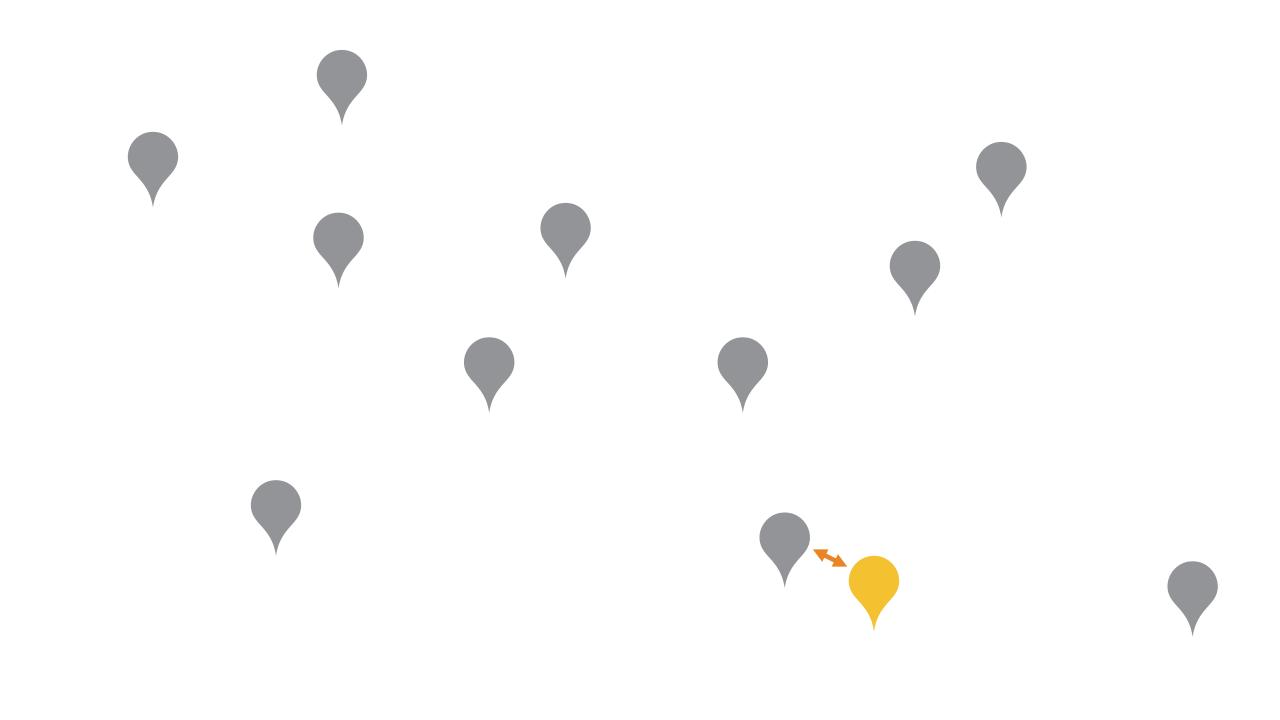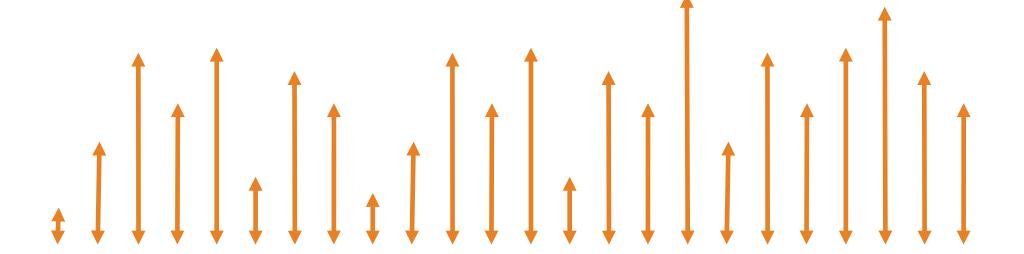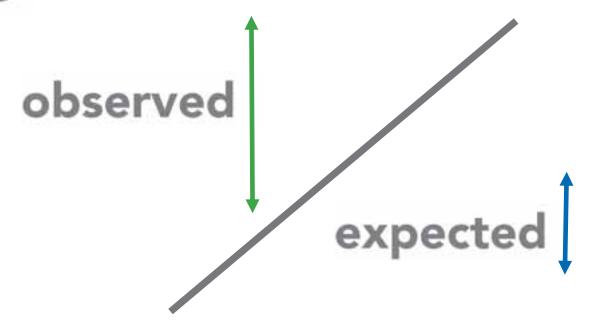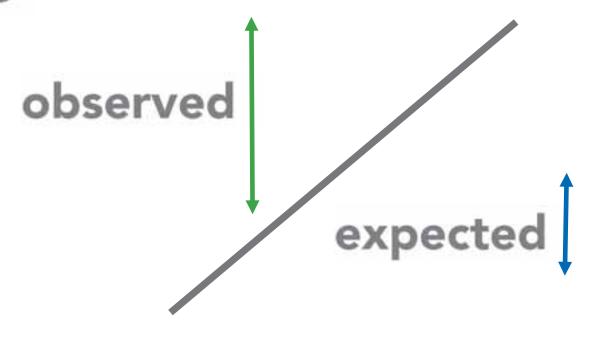average distance =   expected average distance =

ANN ratio = $\dfrac{\text{observed}}{\text{expected}}$

ANN ratio =

observed

expected

Clustered

< 1

ANN ratio =

$$\frac{\text{observed}}{\text{expected}}$$
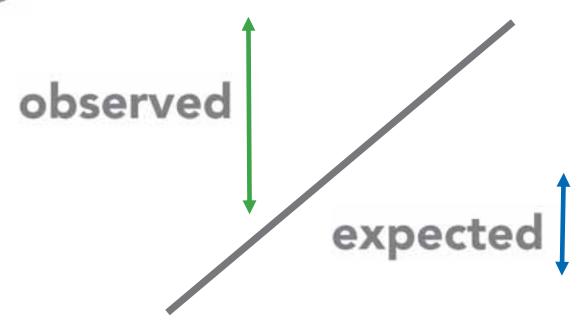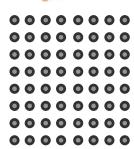
Clustered < **1** < **Dispersed**
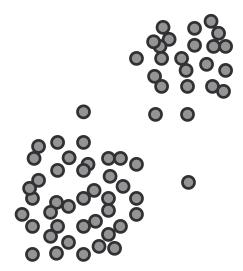
Spatial Autocorrelation
(Moran's I)
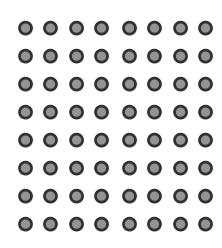
measures spatial autocorrelation based on feature locations and attribute values using the Global Moran's I statistic
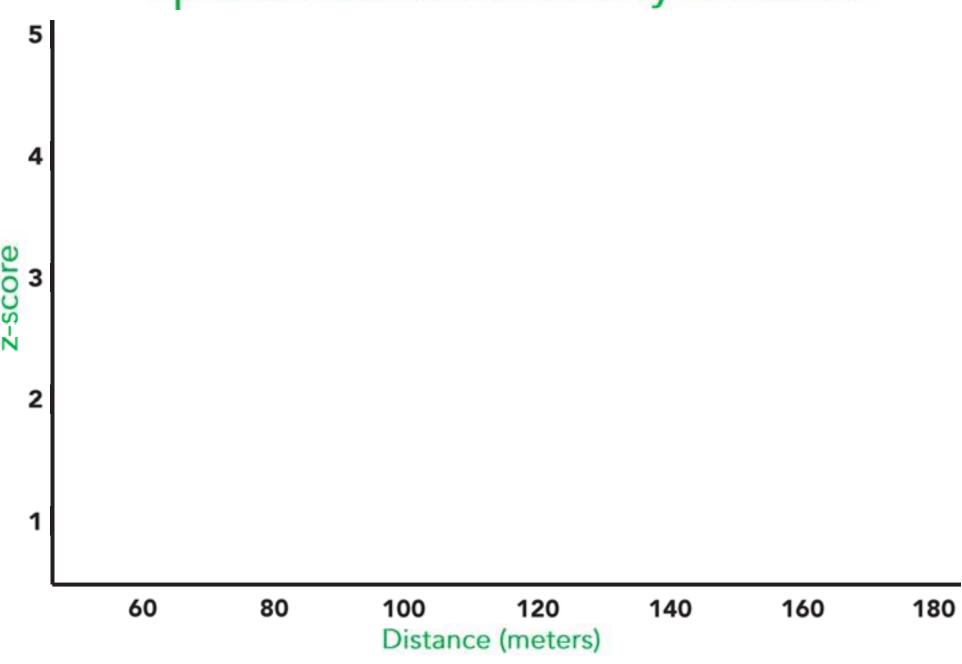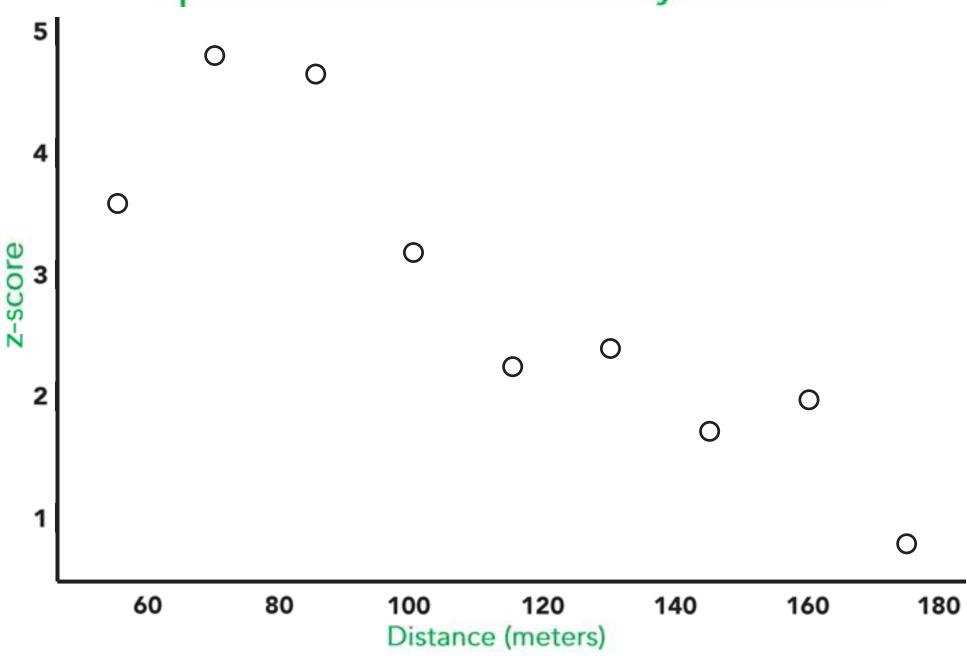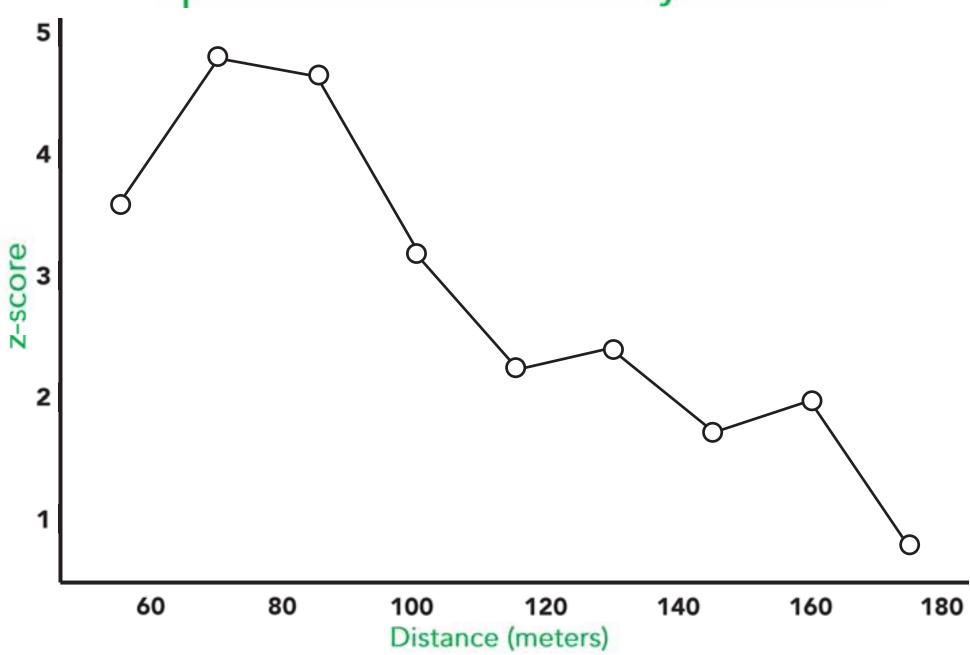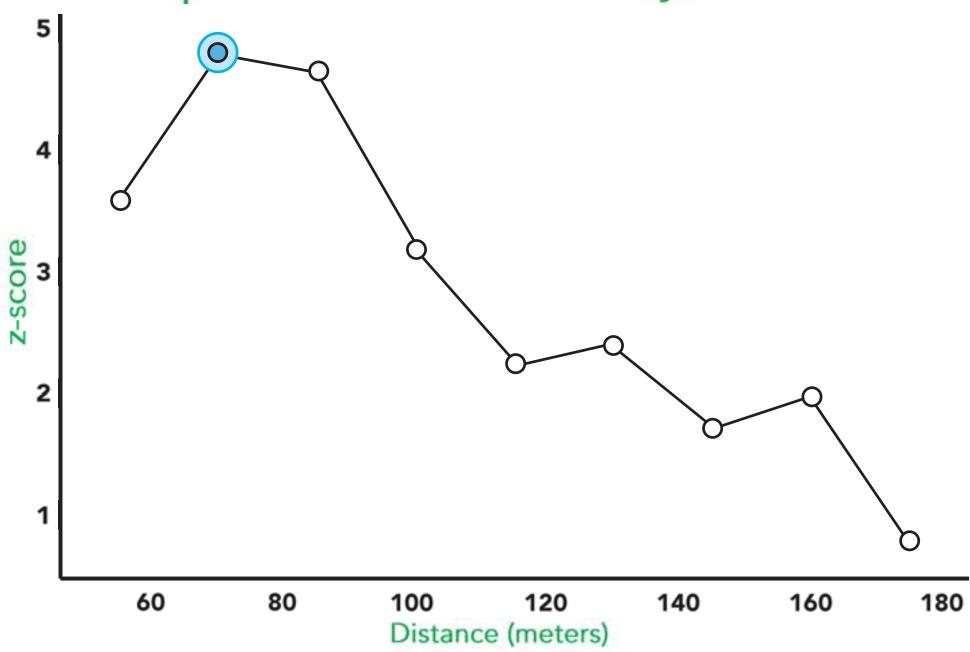
Incremental Spatial Autocorrelation
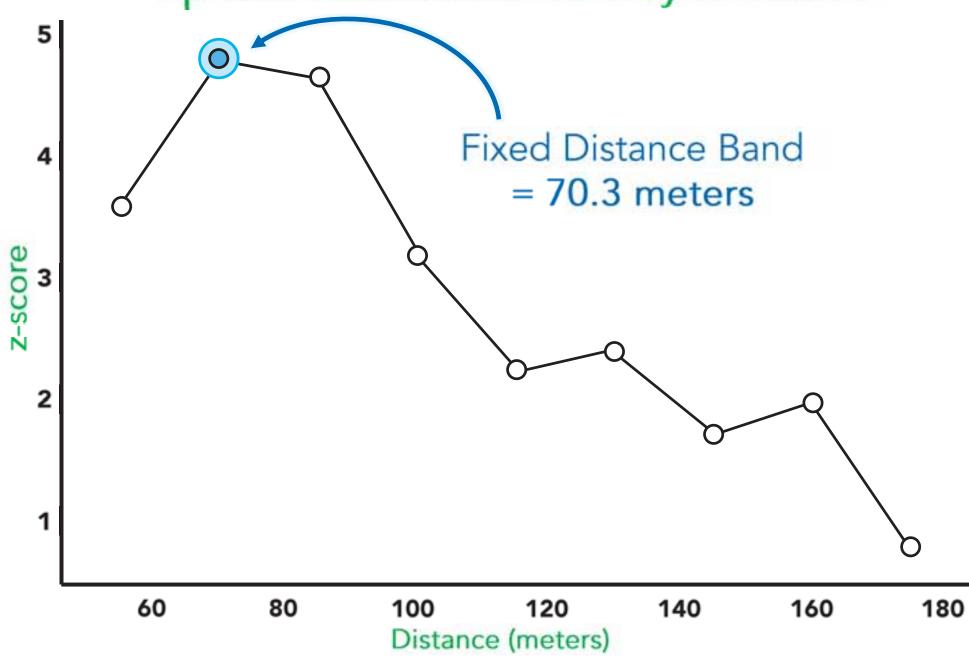
measures spatial autocorrelation for a series of distances and optionally creates a line graph of those distances and their corresponding z-scores

Spatial Autocorrelation by Distance

Spatial Autocorrelation by Distance

Spatial Autocorrelation by Distance

Spatial Autocorrelation by Distance

Fixed Distance Band
= 70.3 meters

z-score

Distance (meters)
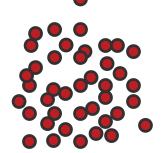
High/Low Clustering
(Getis-Ord General G)

measures the concentration of high or low values for a given study area

What type of clustering is present in the data?

High value clusters
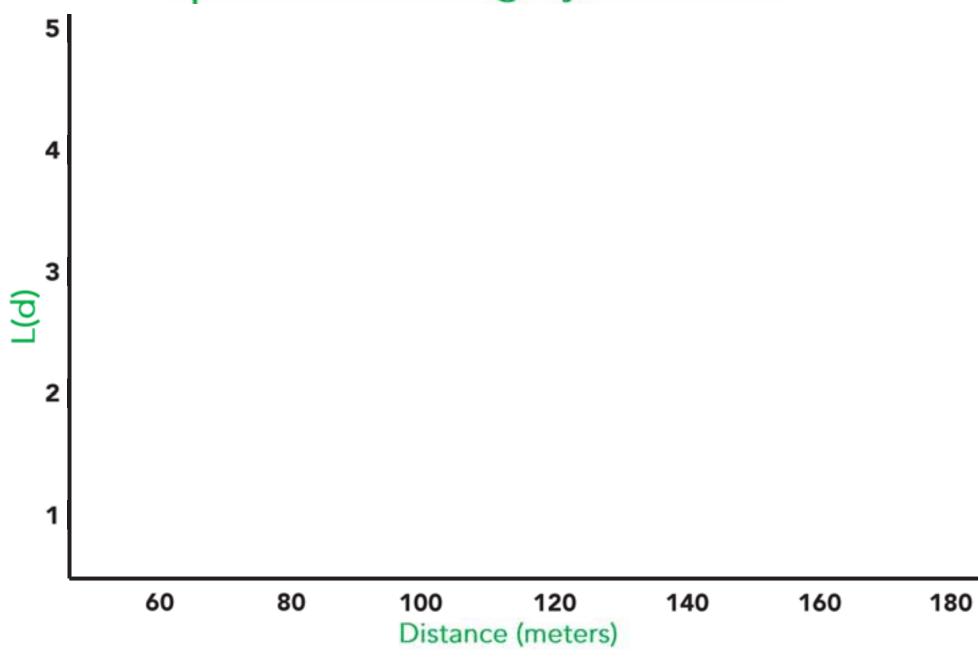
Low value clusters
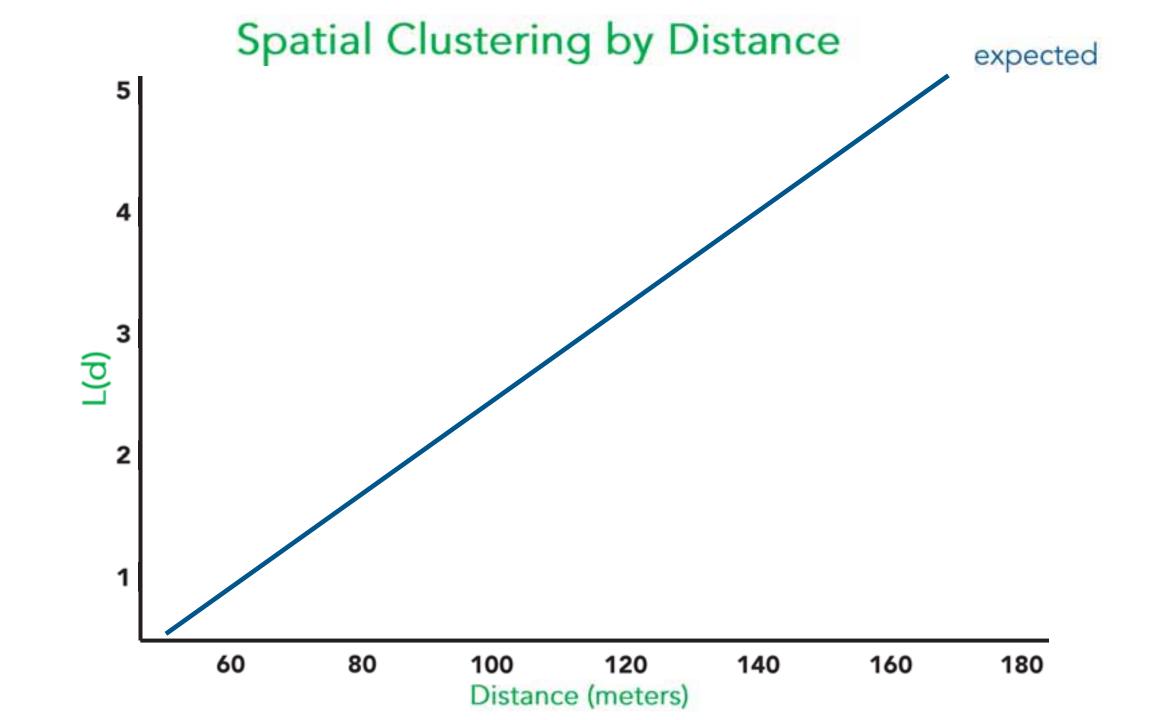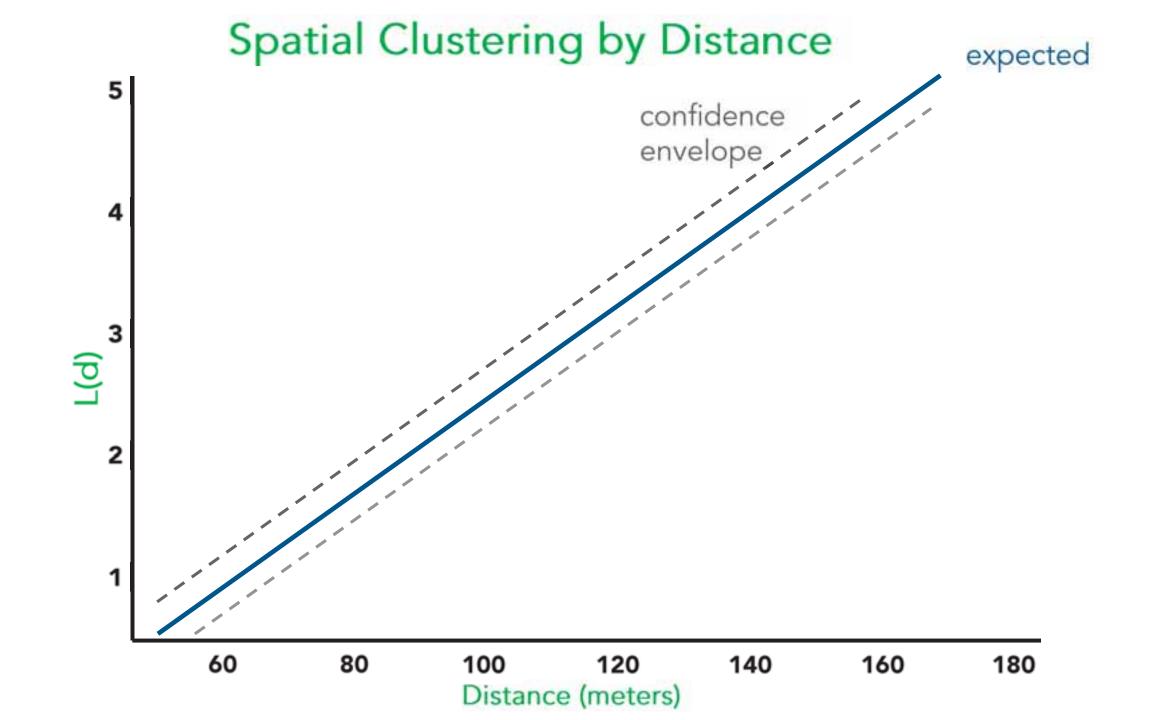
Multi-Distance Spatial Cluster Analysis
(Ripleys K Function)
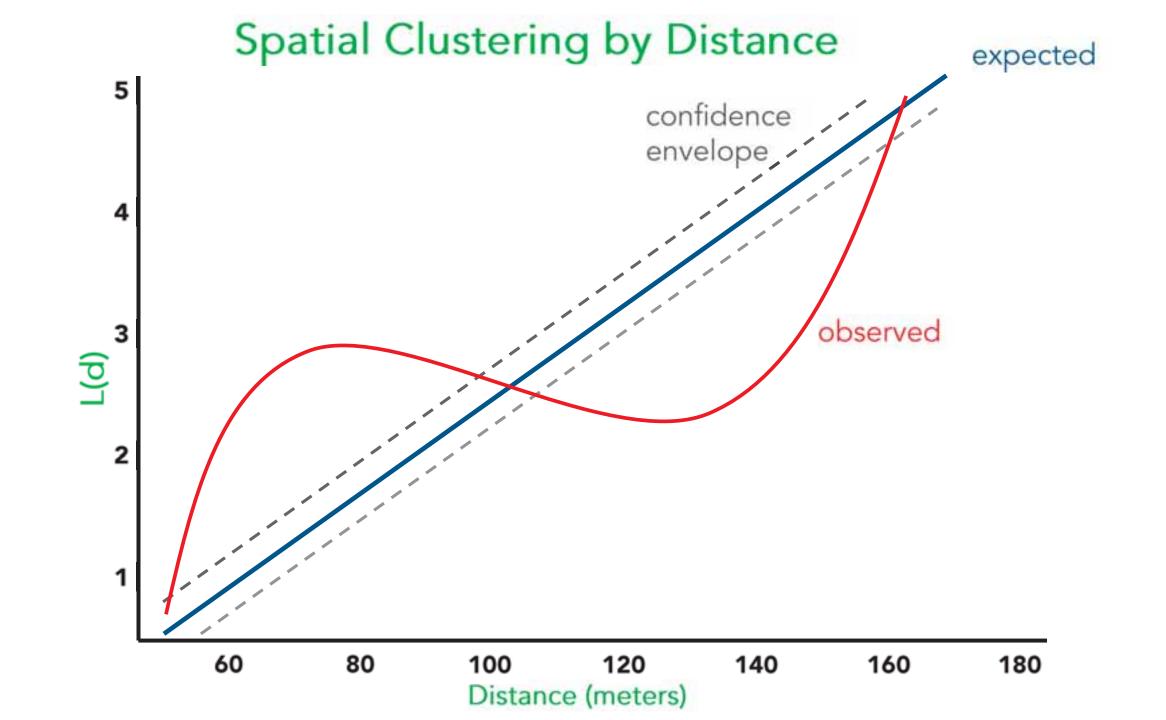
determines whether features, or the values associated with features, exhibit statistically significant clustering or dispersion over a range of distances

Spatial Clustering by Distance

L(d)

Distance (meters)

Spatial Clustering by Distance

expected

L(d)

Distance (meters)

Spatial Clustering by Distance

Spatial Clustering by Distance

Spatial Clustering by Distance

statistically significant clustering

statistically significant dispersion

L(d)

Distance (meters)
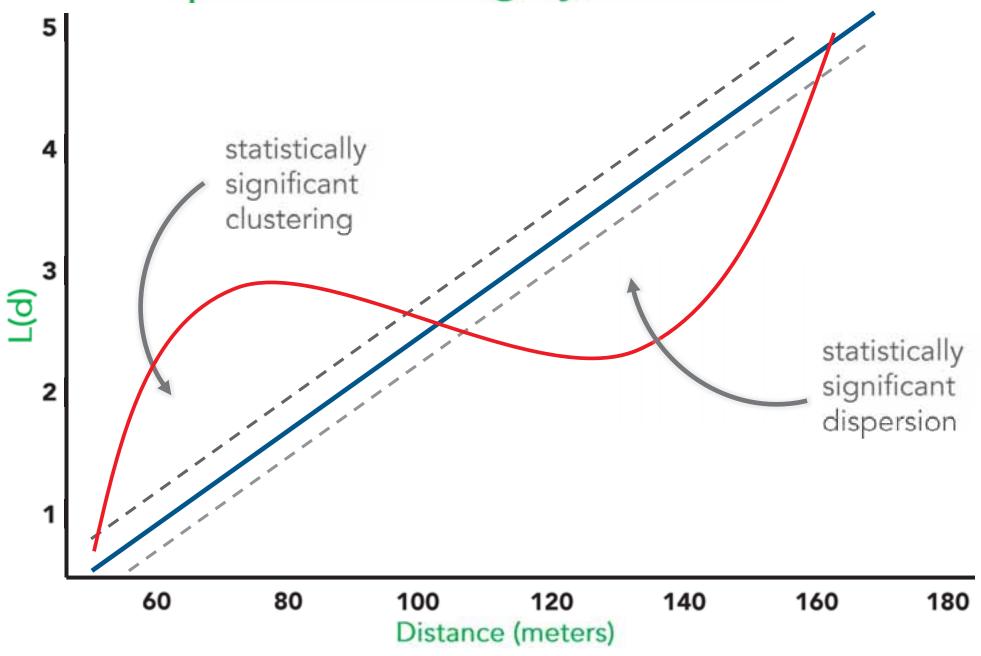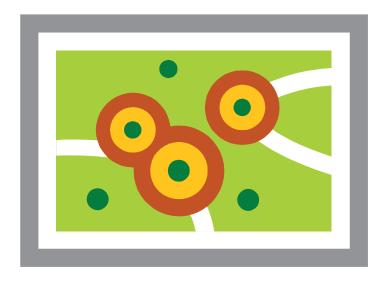
Mapping Clusters

Spatial Statistics Tools
  Analyzing Patterns
    Average Nearest Neighbor
    High/Low Clustering (Getis-Ord General G)
    Incremental Spatial Autocorrelation
    Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    Spatial Autocorrelation (Morans I)
  Mapping Clusters
    Cluster and Outlier Analysis (Anselin Local Morans I)
    Grouping Analysis
    Hot Spot Analysis (Getis-Ord Gi*)
    Optimized Hot Spot Analysis
    Similarity Search
  Measuring Geographic Distributions
    Central Feature
    Directional Distribution (Standard Deviational Ellipse)
    Linear Directional Mean
    Mean Center
    Median Center
    Standard Distance
  Modeling Spatial Relationships
    Exploratory Regression
    Generate Network Spatial Weights
    Generate Spatial Weights Matrix
    Geographically Weighted Regression
    Ordinary Least Squares

These tools may be used to identify statistically significant hot spots, cold spots, or spatial outliers. There are also tools to identify or group features with similar characteristics.

Similarity Search

identifies which candidate features are most similar or most dissimilar to one or more input features based on feature attributes
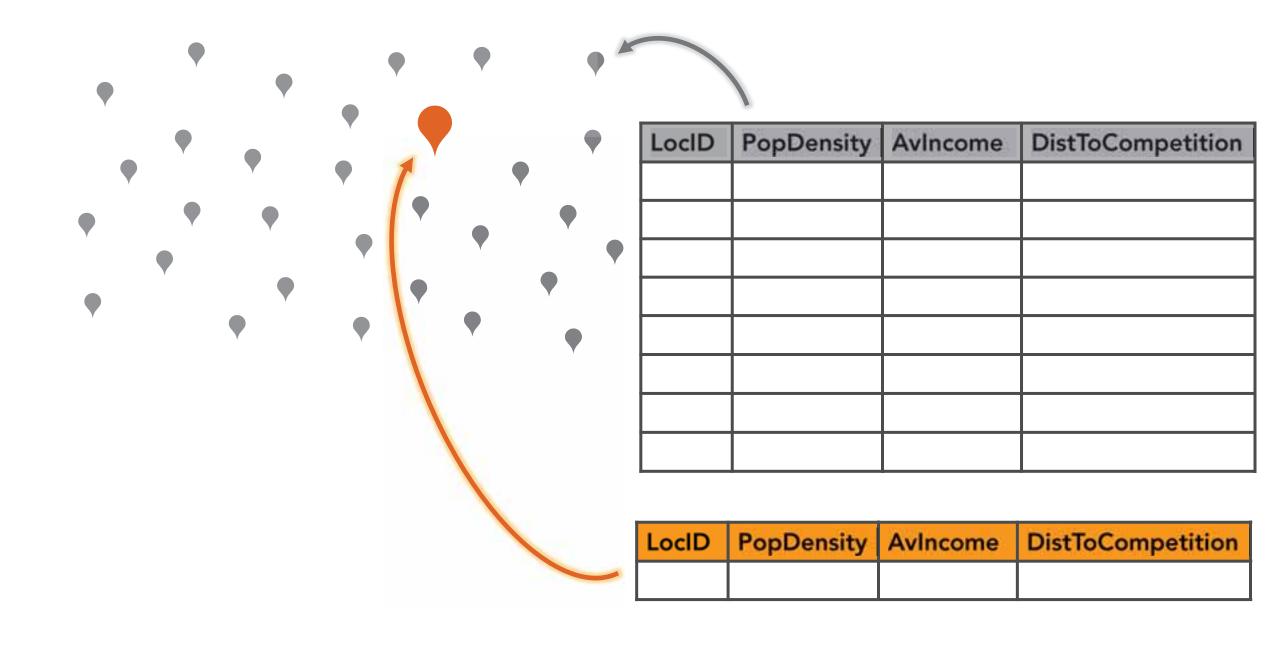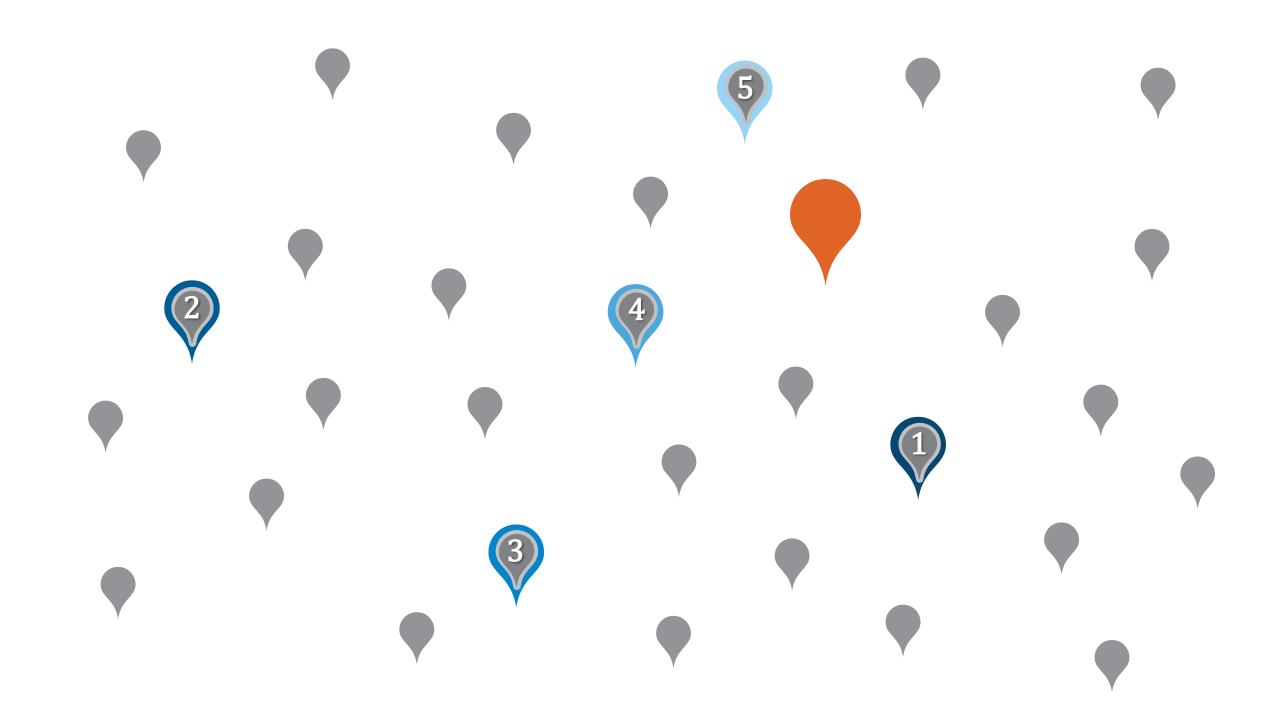
potential
store
locations

performing
store

potential
store
locations

LocID	PopDensity	AvIncome	DistToCompetition

LocID	PopDensity	AvIncome	DistToCompetition

LocID	PopDensity	AvIncome	DistToCompetition

Rank 📍 by how similar they are to 📍 based on:
- **PopDensity**
- **AvIncome**
- **DistToCompetition**

LocID	PopDensity	AvIncome	DistToCompetition
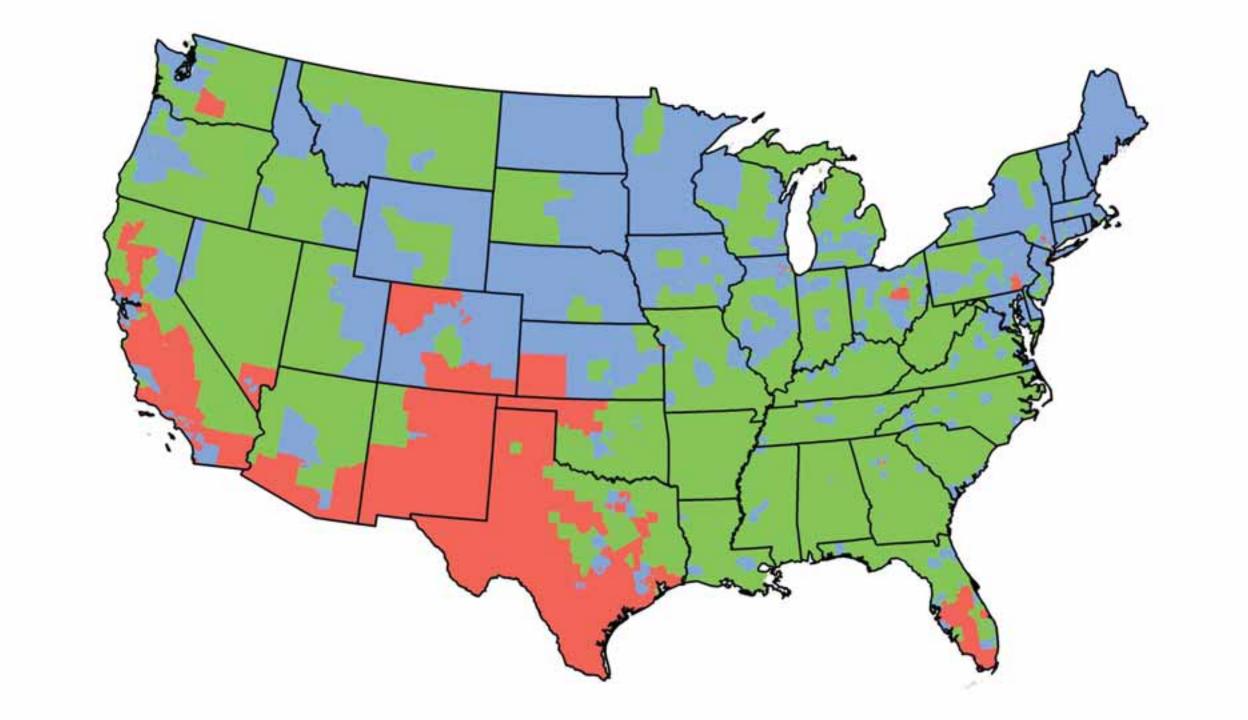
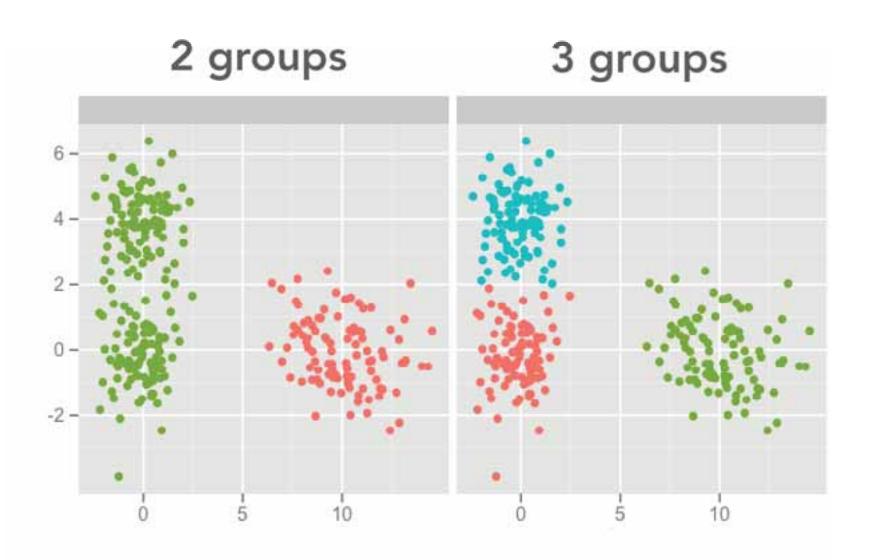| LocID	PopDensity	AvIncome	DistToCompetition
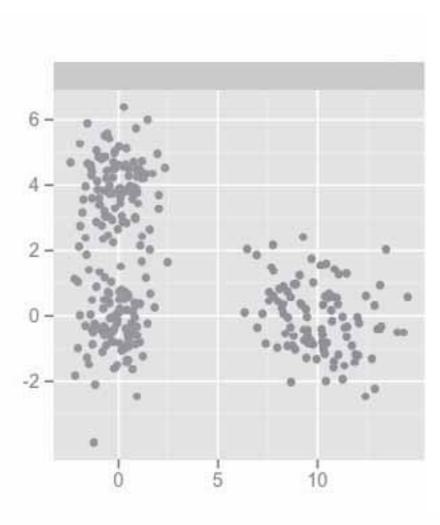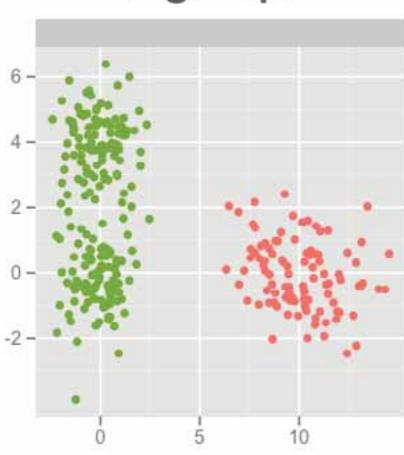
Grouping Analysis

groups features based on feature attributes
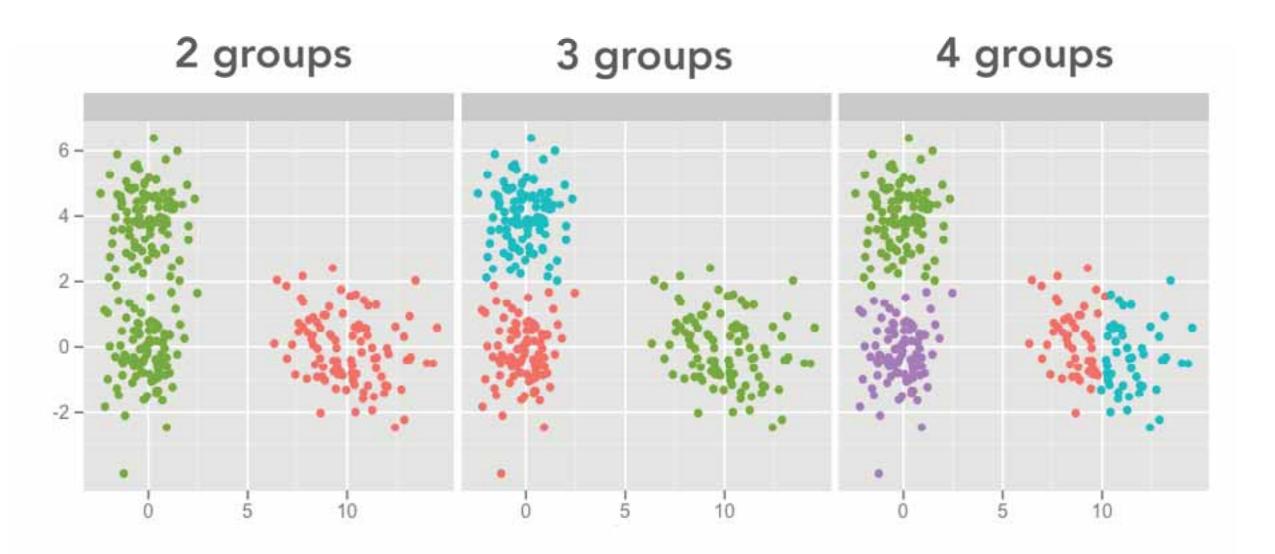and optional spatial/temporal constraints
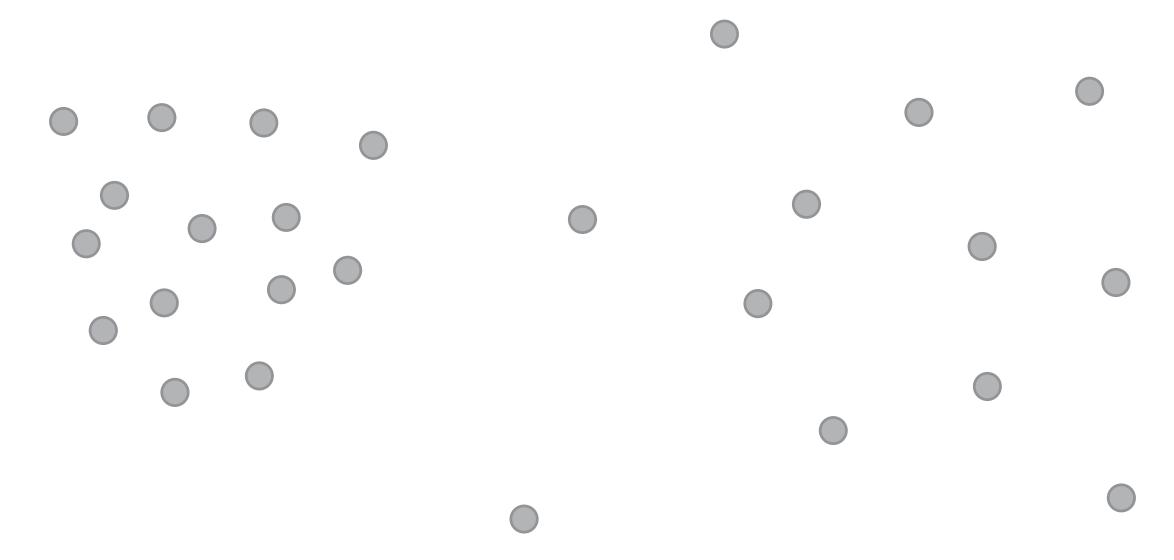
K Means

K Means

2 groups

K Means

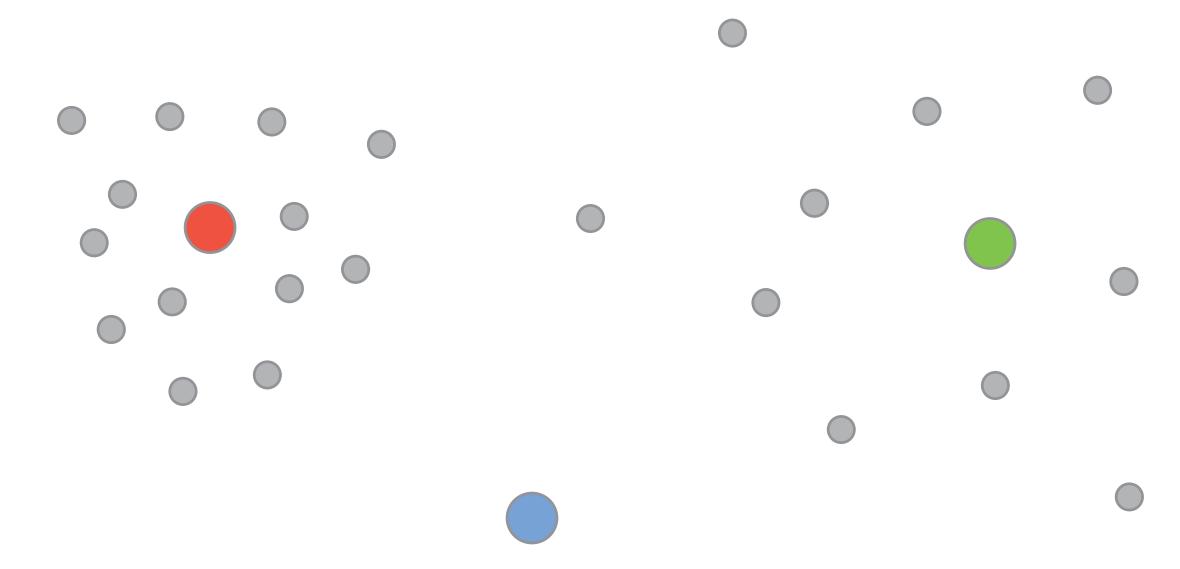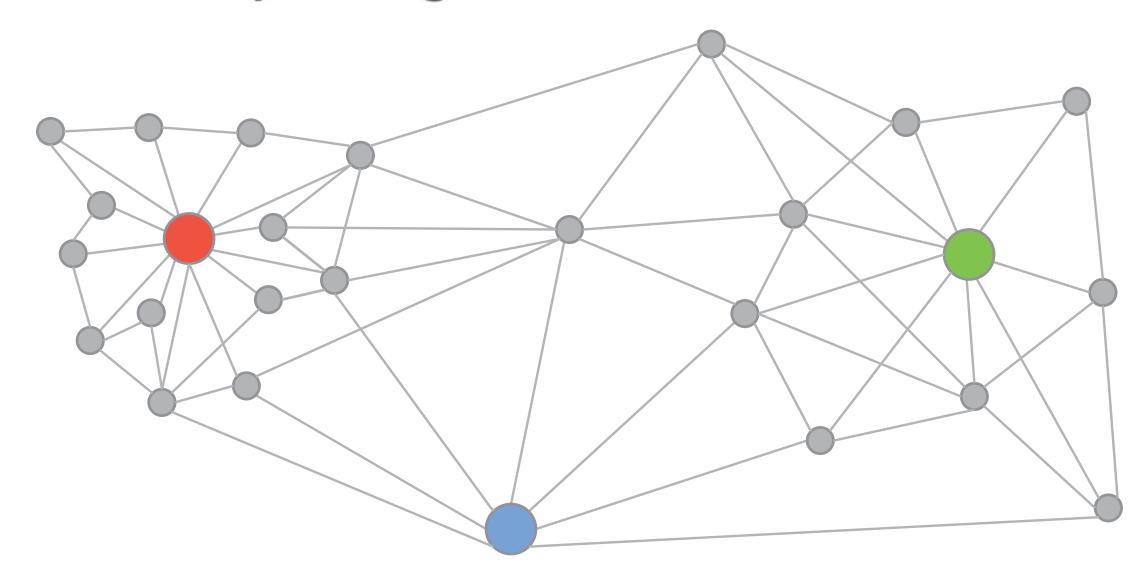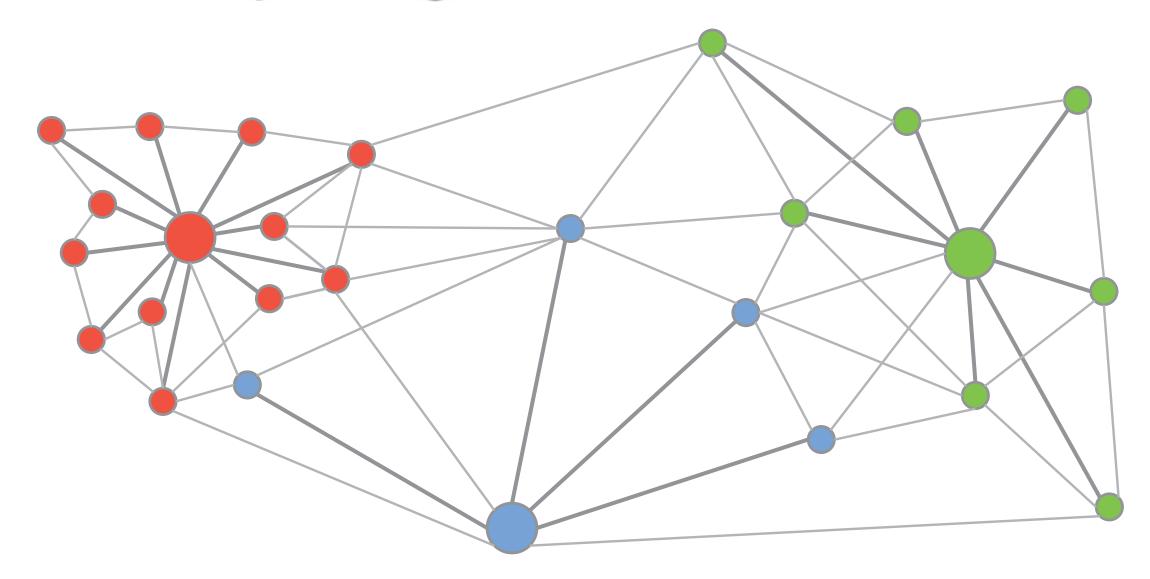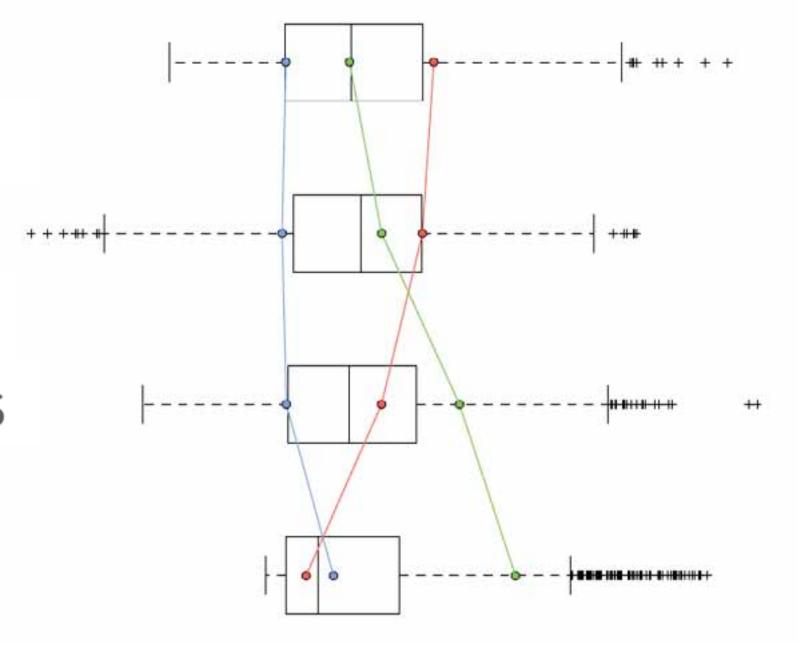# Minimum Spanning Tree

Minimum Spanning Tree
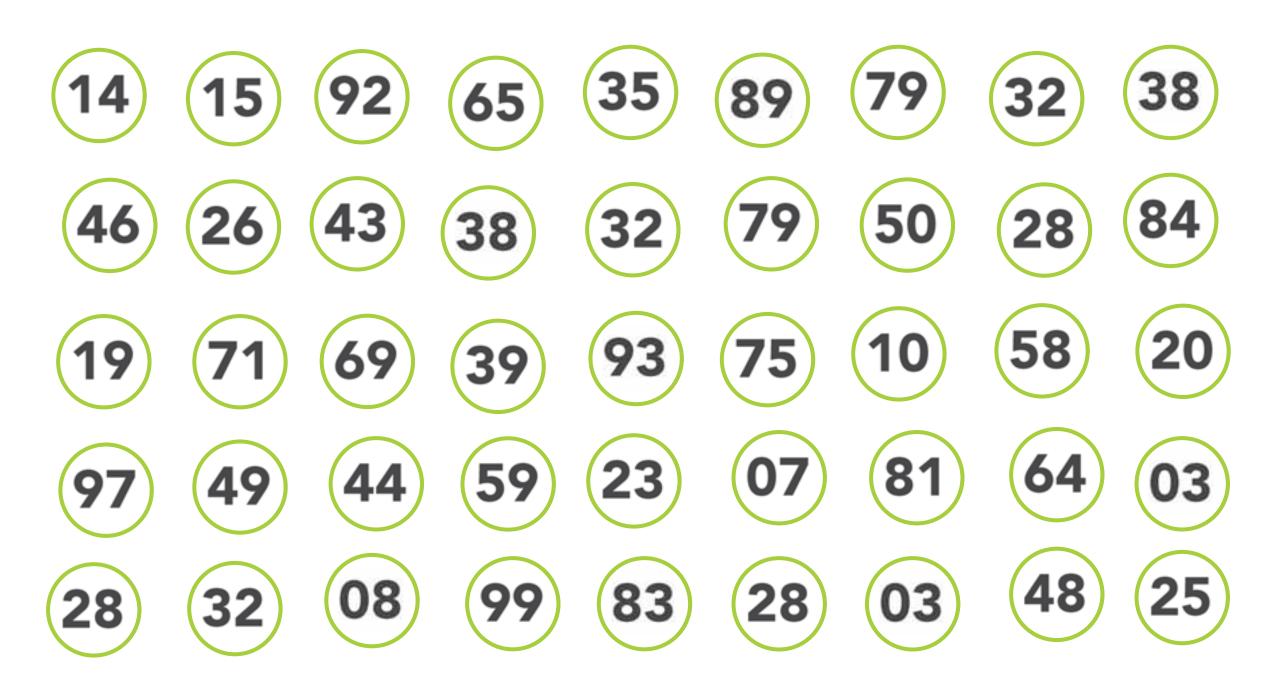
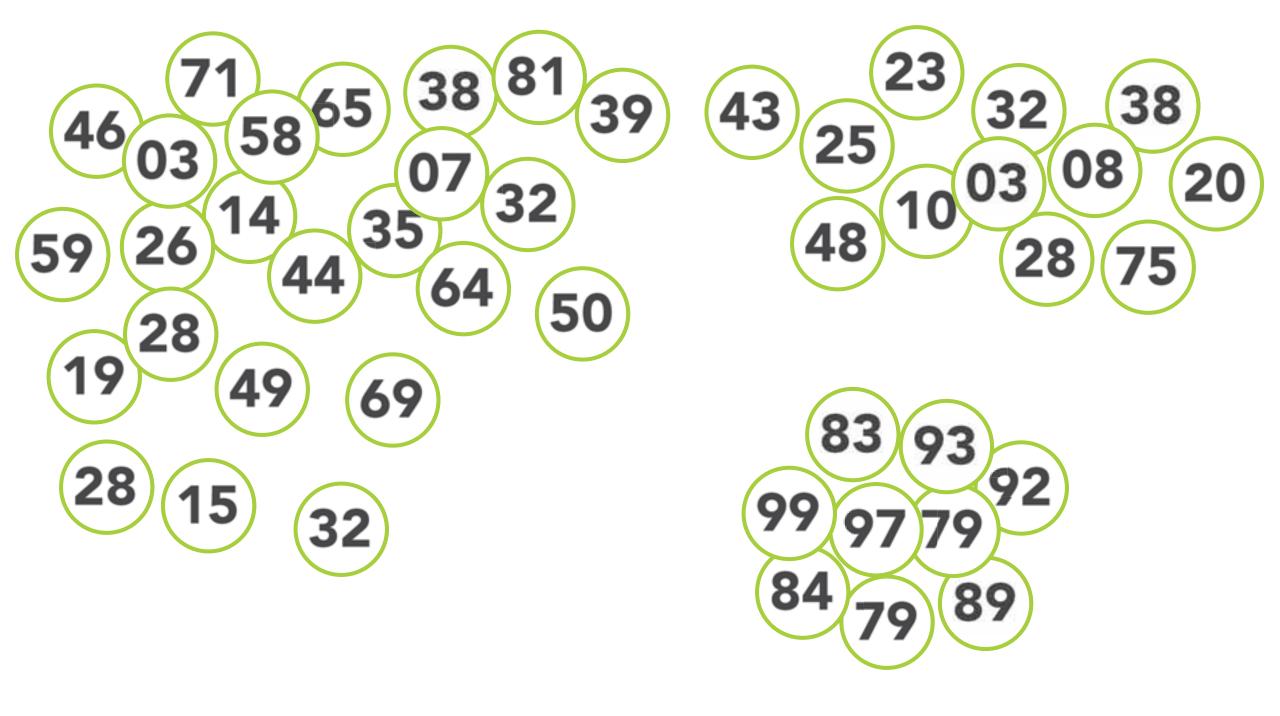# Minimum Spanning Tree

Minimum Spanning Tree
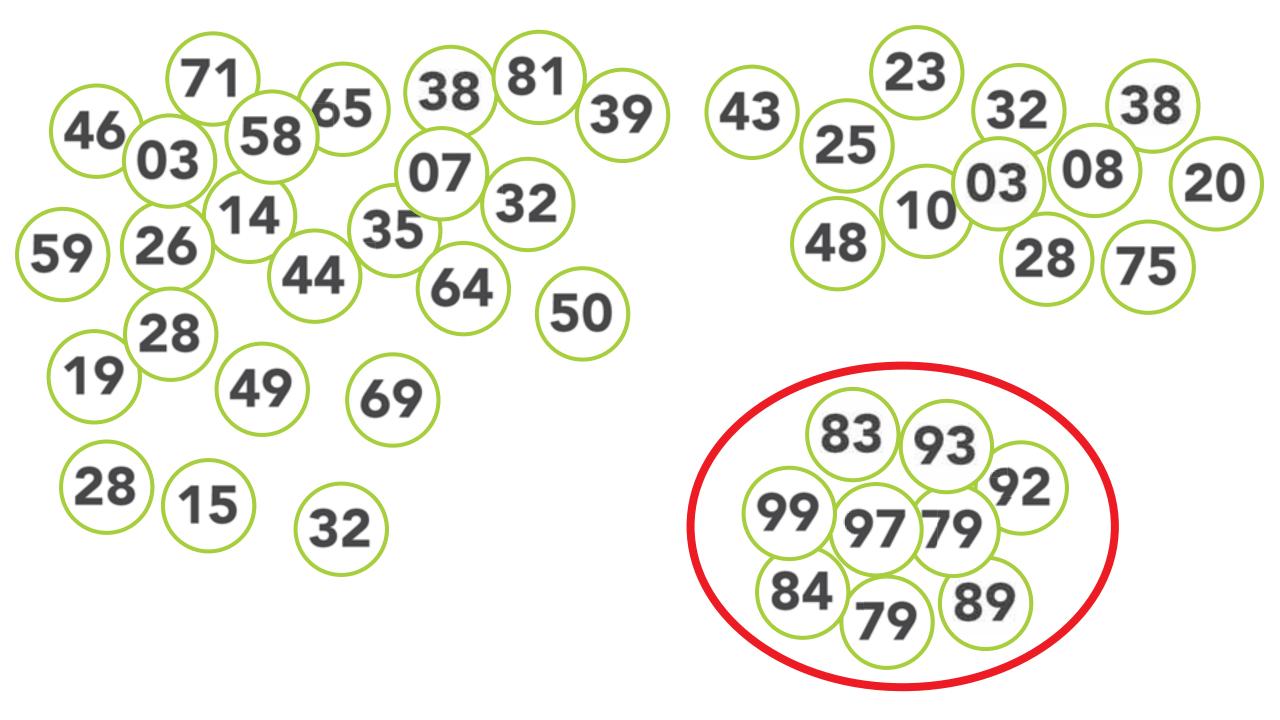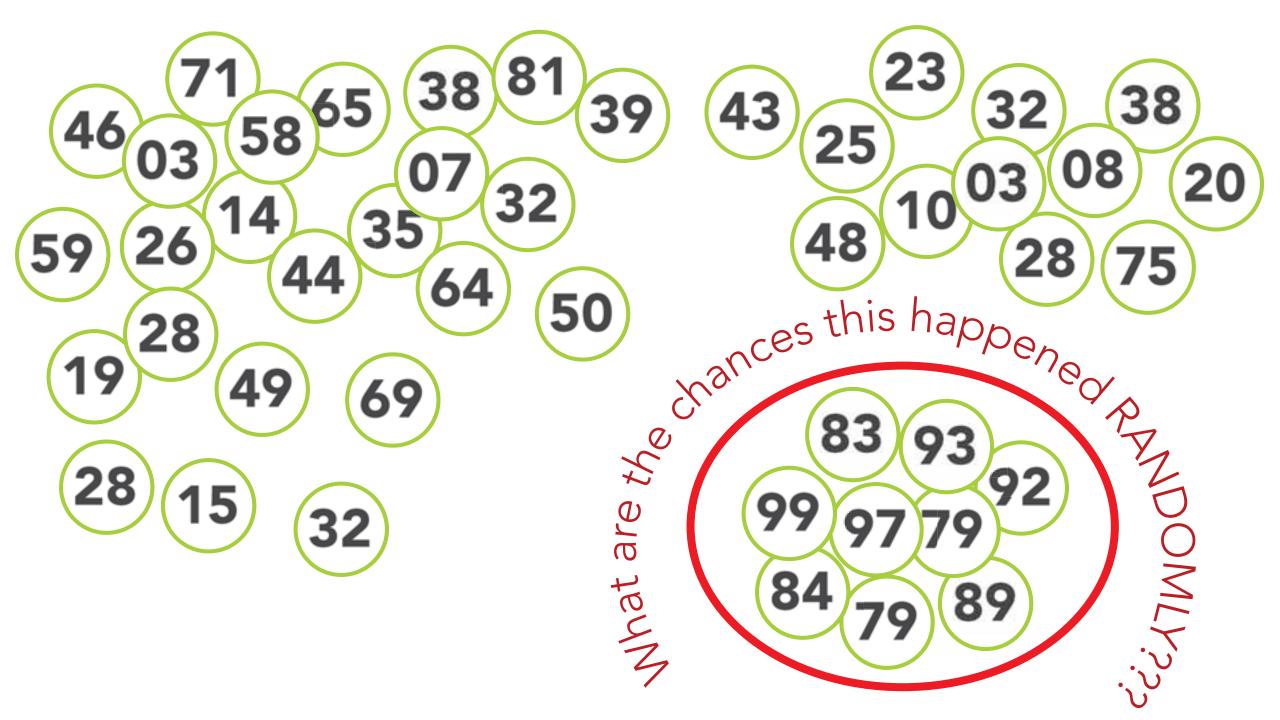
interpret
results
through
box plots

Hot Spot Analysis
(Getis-Ord Gi*)
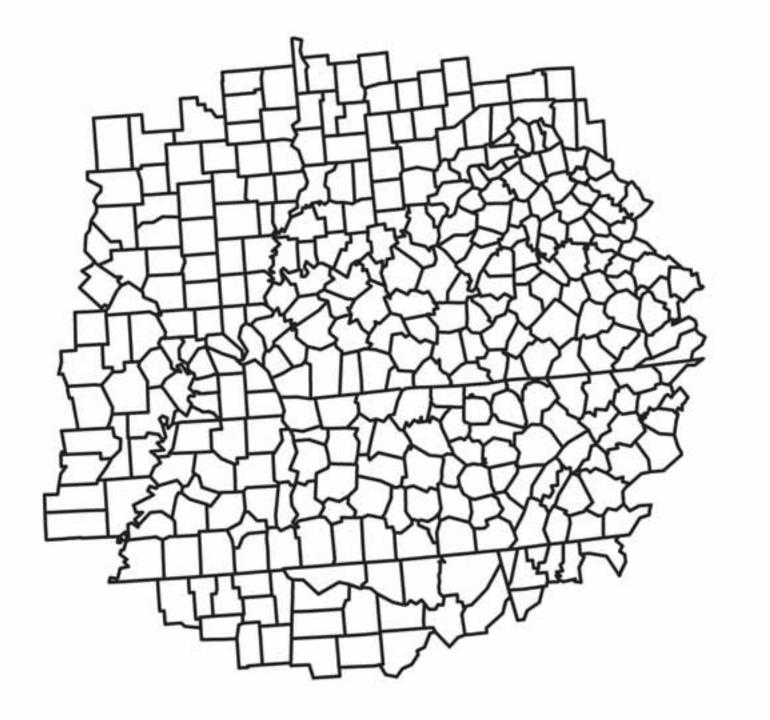
given a set of weighted features, identifies statistically significant hot spots and cold spots using the Getis-Ord Gi* statistic
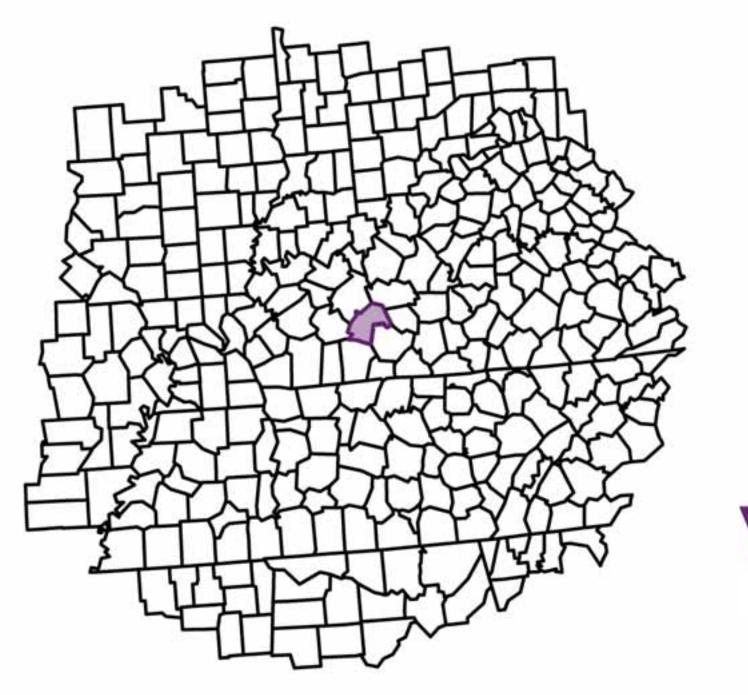
14	15	92	65	35	89	79	32	38
46	26	43	38	32	79	50	28	84
19	71	69	39	93	75	10	58	20
97	49	44	59	23	07	81	64	03
28	32	08	99	83	28	03	48	25

46 71 65 38 81 39
03 58 43 23 32 38
07 25 03 08 20
14 32 10
59 26 35 48 28 75
44 64
28 50
19 49 69
28 15 32
83 93
99 97 79 92
84 79 89

71 46 03 58 65 38 81 39

14 07 32

59 26 35

44 64 50

28

19 49 69

28 15 32

43 25 23 32 38
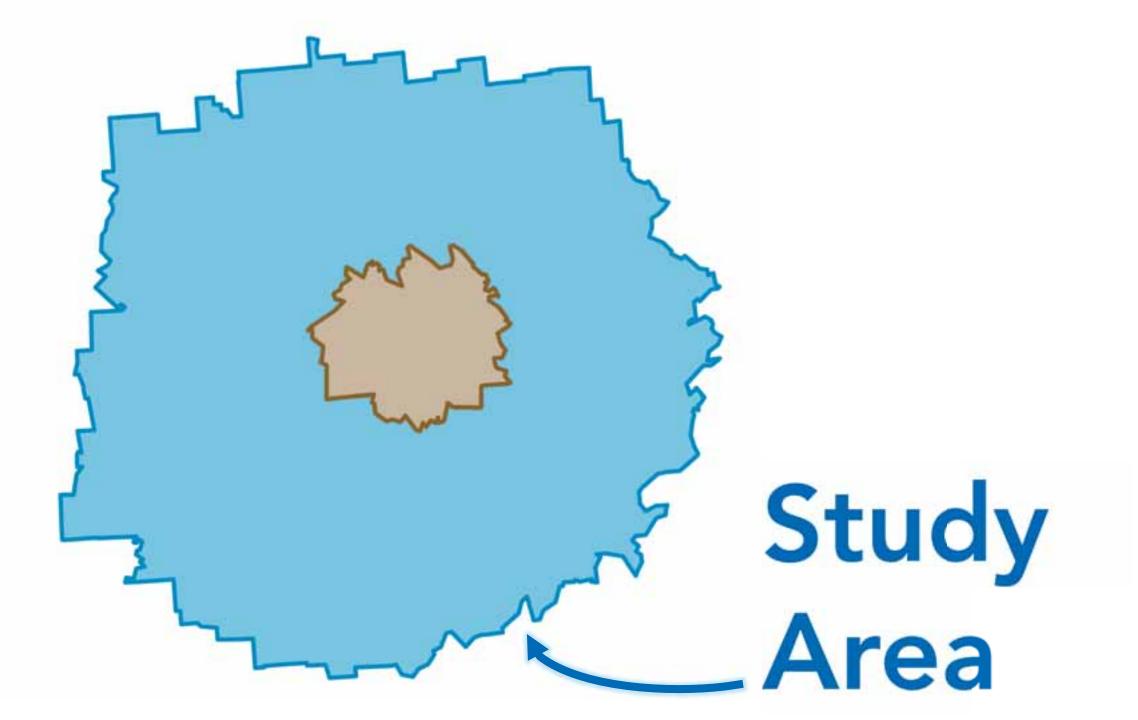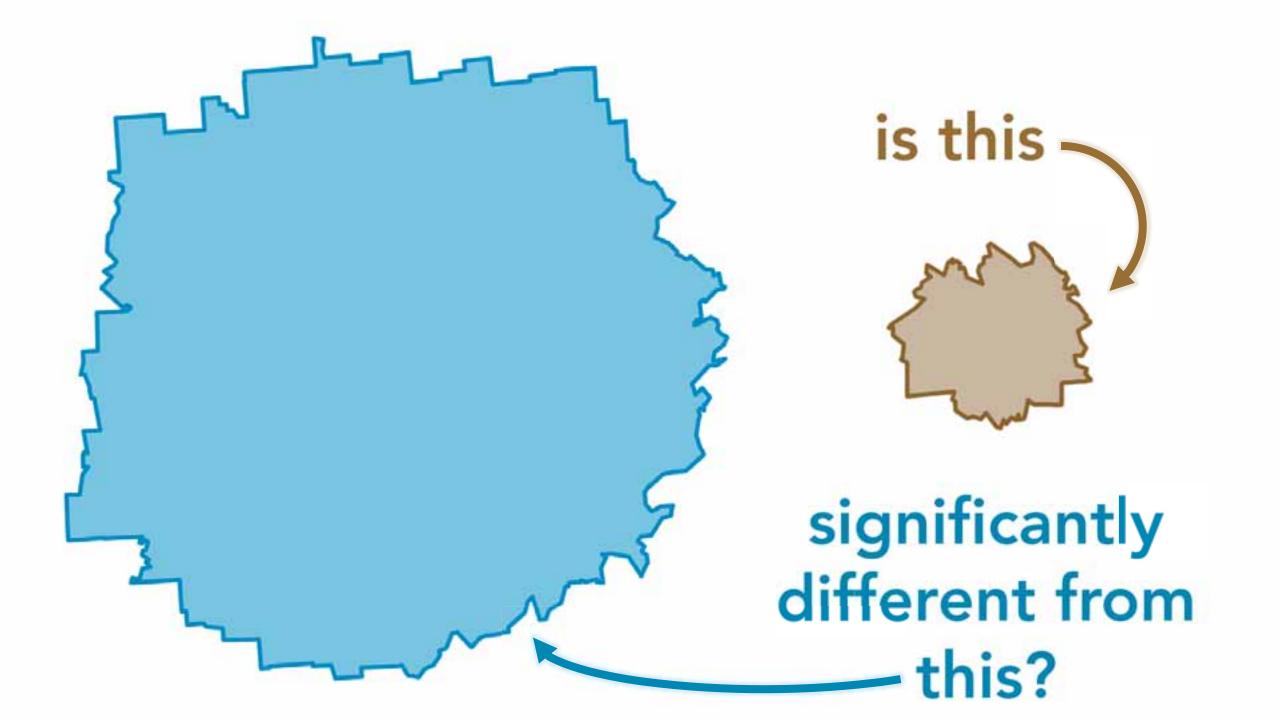
48 10 03 08 20

28 75

What are the chances this happened RANDOMLY???

83 93

99 97 79 92

84 79 89

feature

each
feature
has a
value

Neighborhood

Study Area

is this

significantly different from this?

if significantly
higher...

feature is
marked as a
hot spot!

	Hot Spot - 90% Confidence
	Hot Spot - 95% Confidence
	Hot Spot - 99% Confidence

Cold Spot - 99% Confidence

Cold Spot - 95% Confidence

Cold Spot - 90% Confidence

Not Significant

Hot Spot - 90% Confidence
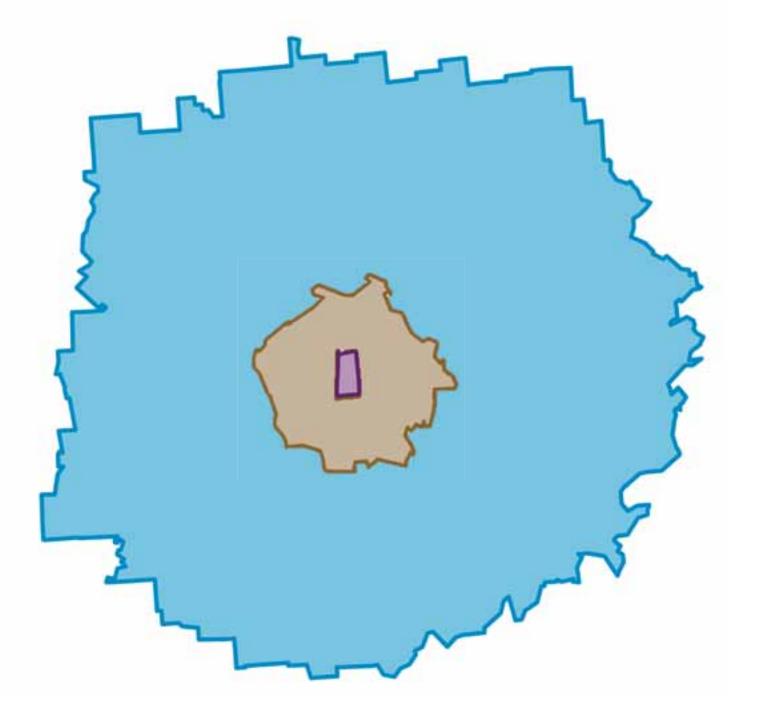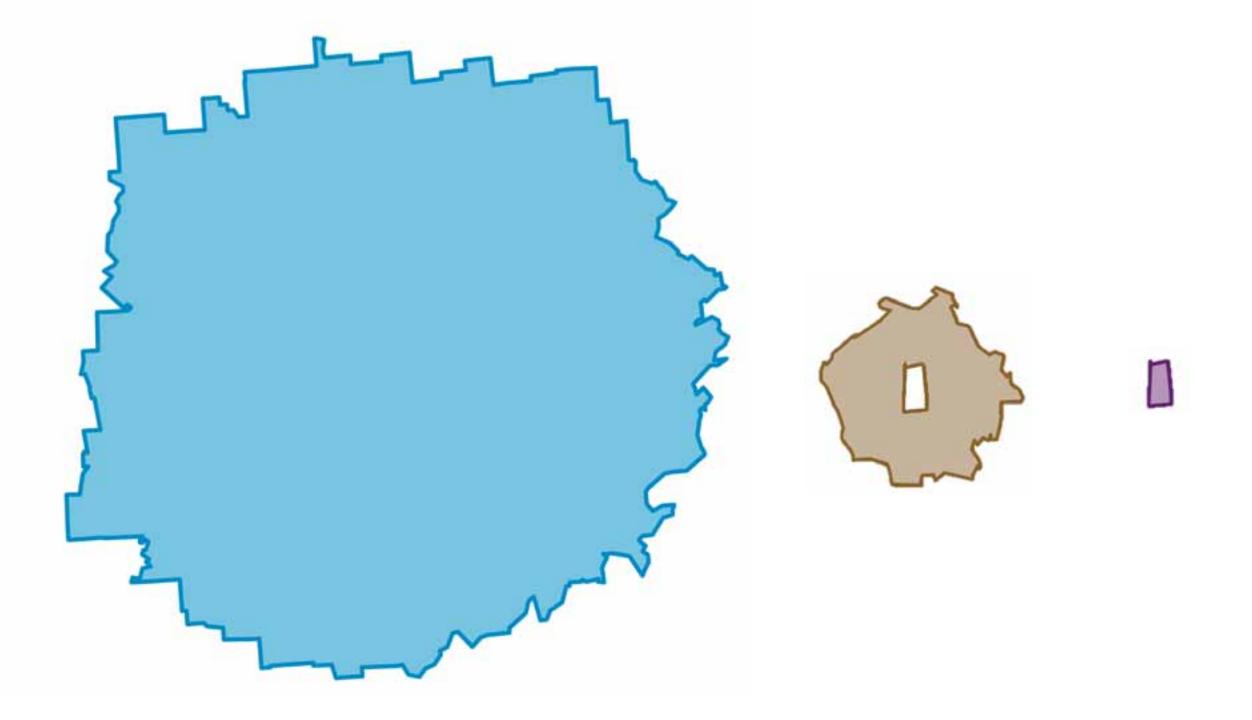
Hot Spot - 95% Confidence

Hot Spot - 99% Confidence
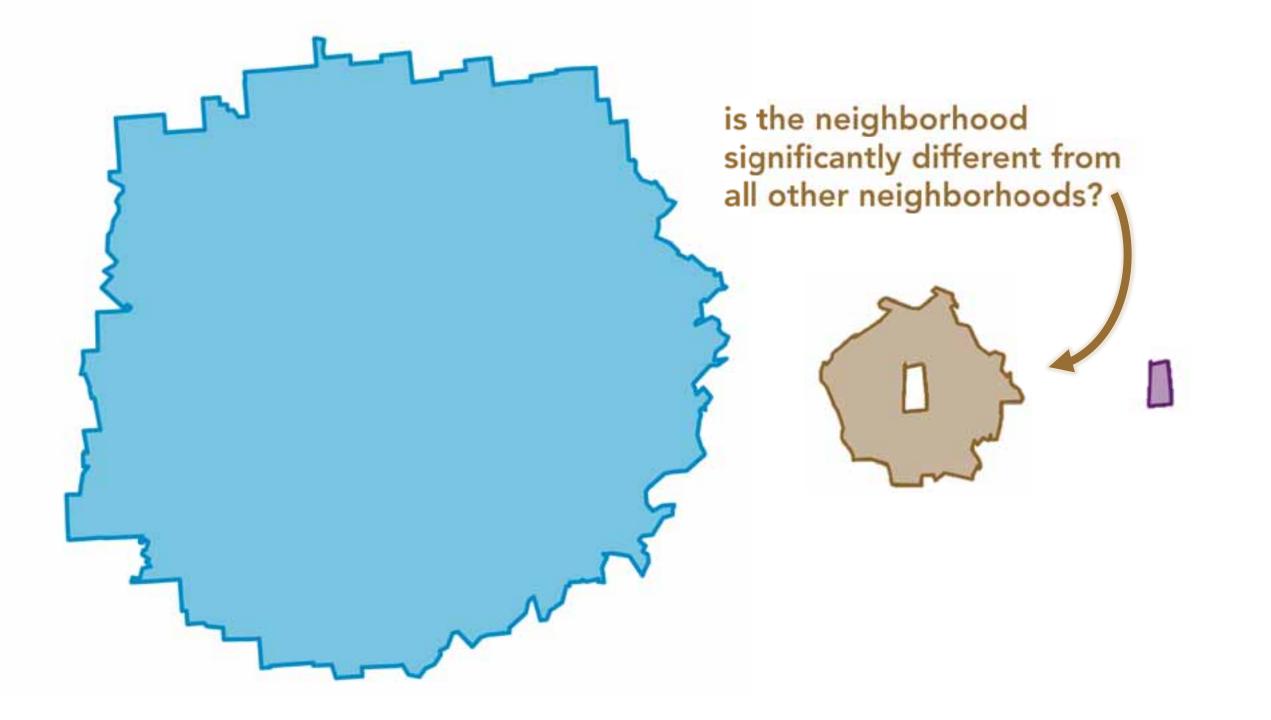
Optimized Hot Spot Analysis

evaluates input feature characteristics to produce optimal Hot Spot Analysis results
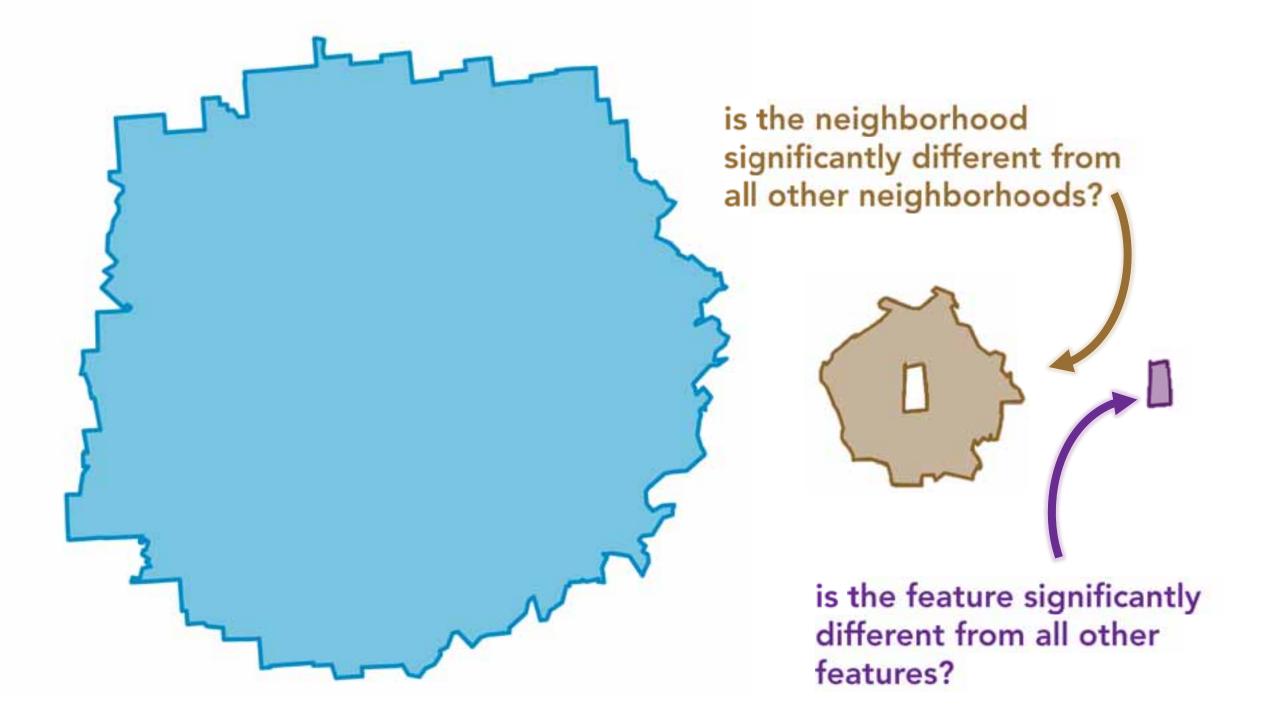
Cluster and Outlier Analysis
(Anselin Local Moran's I)

given a set of weighted features, identifies statistically significant hot spots, cold spots, and spatial outliers using the Anselin Local Moran's I statistic

is the neighborhood
significantly different from
all other neighborhoods?

is the neighborhood significantly different from all other neighborhoods?

is the feature significantly different from all other features?

feature is **higher** than
other features,
neighborhood is **lower**
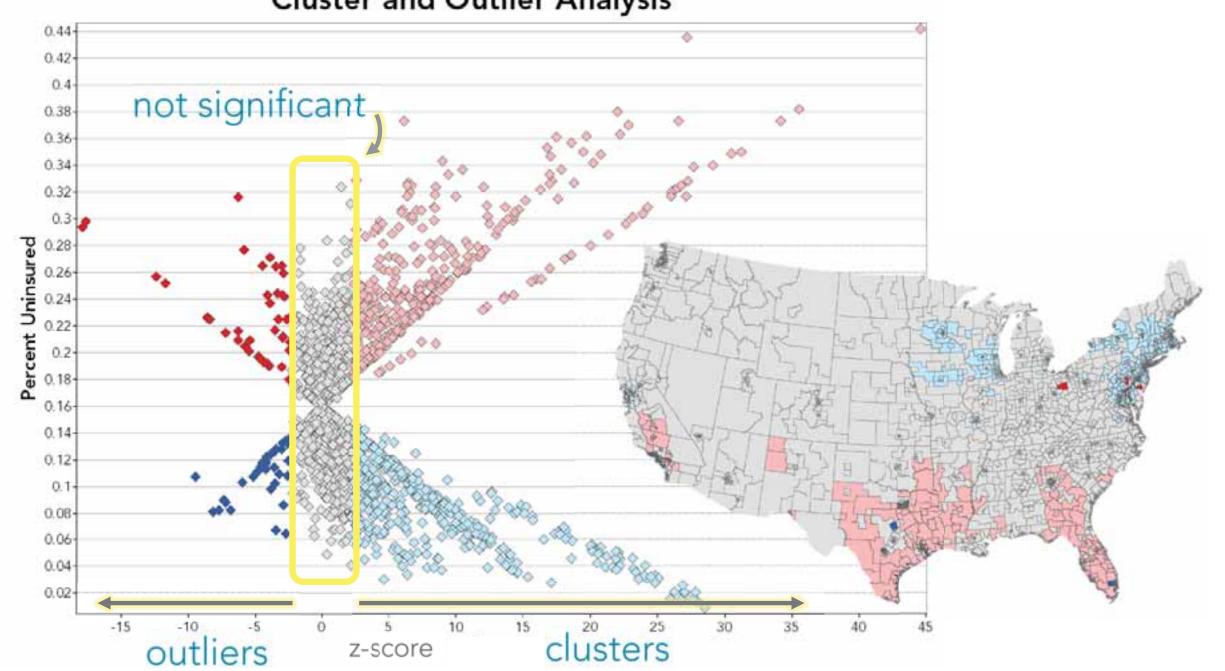than other
neighborhoods

HL

HH

feature is **higher** than
other features,
neighborhood is **higher**
than other
neighborhoods

feature is **lower** than
other features,
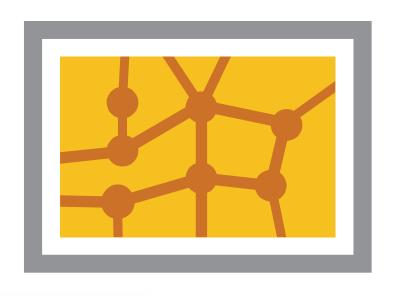neighborhood is **lower**
than other
neighborhoods

LL

LH

feature is **lower** than
other features,
neighborhood is **higher**
than other
neighborhoods

Cluster and Outlier Analysis

Modeling Spatial Relationships
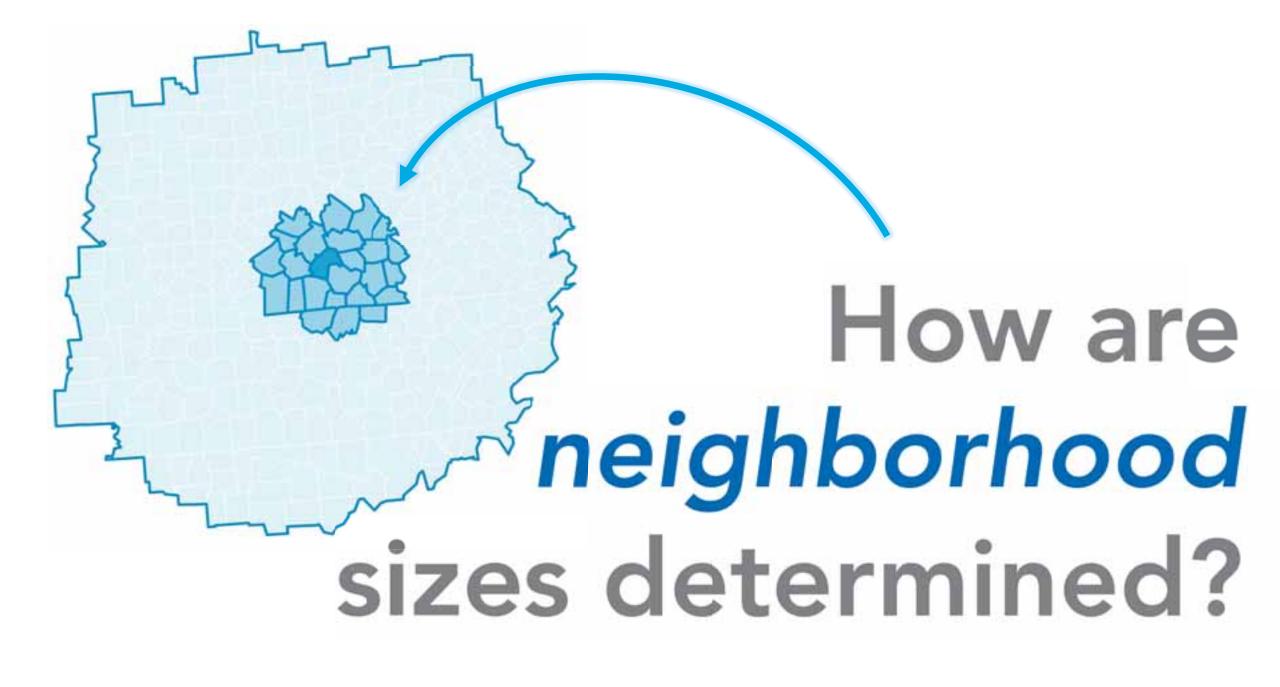
- Spatial Statistics Tools
  - Analyzing Patterns
    - Average Nearest Neighbor
    - High/Low Clustering (Getis-Ord General G)
    - Incremental Spatial Autocorrelation
    - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    - Spatial Autocorrelation (Morans I)
  - Mapping Clusters
    - Cluster and Outlier Analysis (Anselin Local Morans I)
    - Grouping Analysis
    - Hot Spot Analysis (Getis-Ord Gi*)
    - Optimized Hot Spot Analysis
    - Similarity Search
  - Measuring Geographic Distributions
    - Central Feature
    - Directional Distribution (Standard Deviational Ellipse)
    - Linear Directional Mean
    - Mean Center
    - Median Center
    - Standard Distance
  - Modeling Spatial Relationships
    - Exploratory Regression
    - Generate Network Spatial Weights
    - Generate Spatial Weights Matrix
    - Geographically Weighted Regression
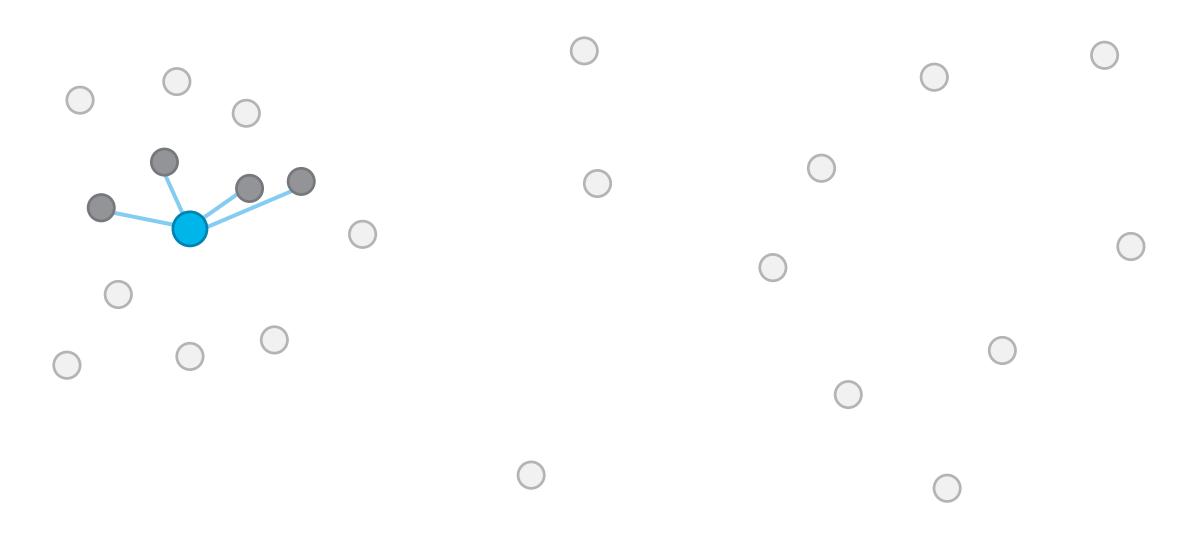   - Ordinary Least Squares

These tools model data relationships using regression analyses or construct spatial weights matrices.
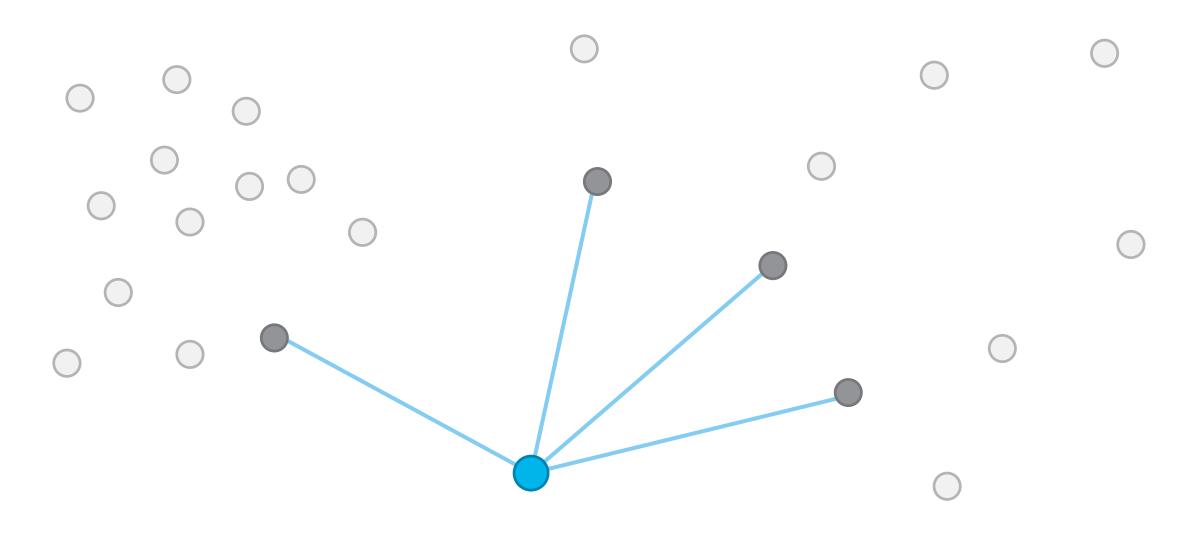
Generate Spatial Weights Matrix

constructs a spatial weights matrix (.swm) file to represent the spatial relationships among features in a dataset
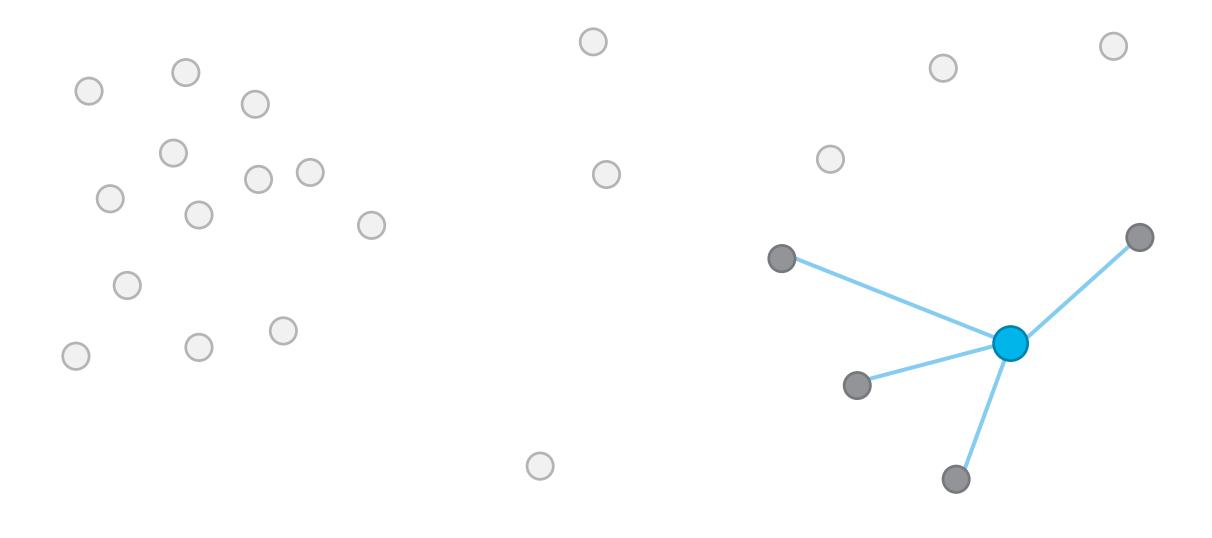
How are *neighborhood* sizes determined?

K Nearest Neighbors

K Nearest Neighbors

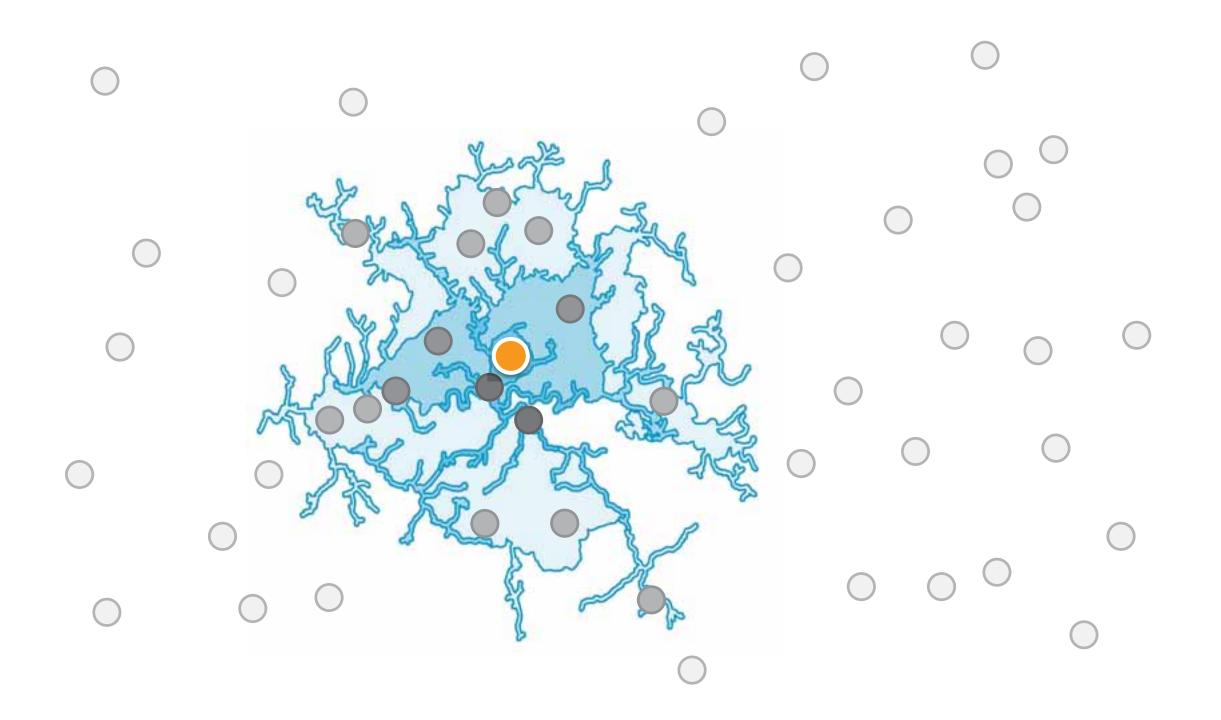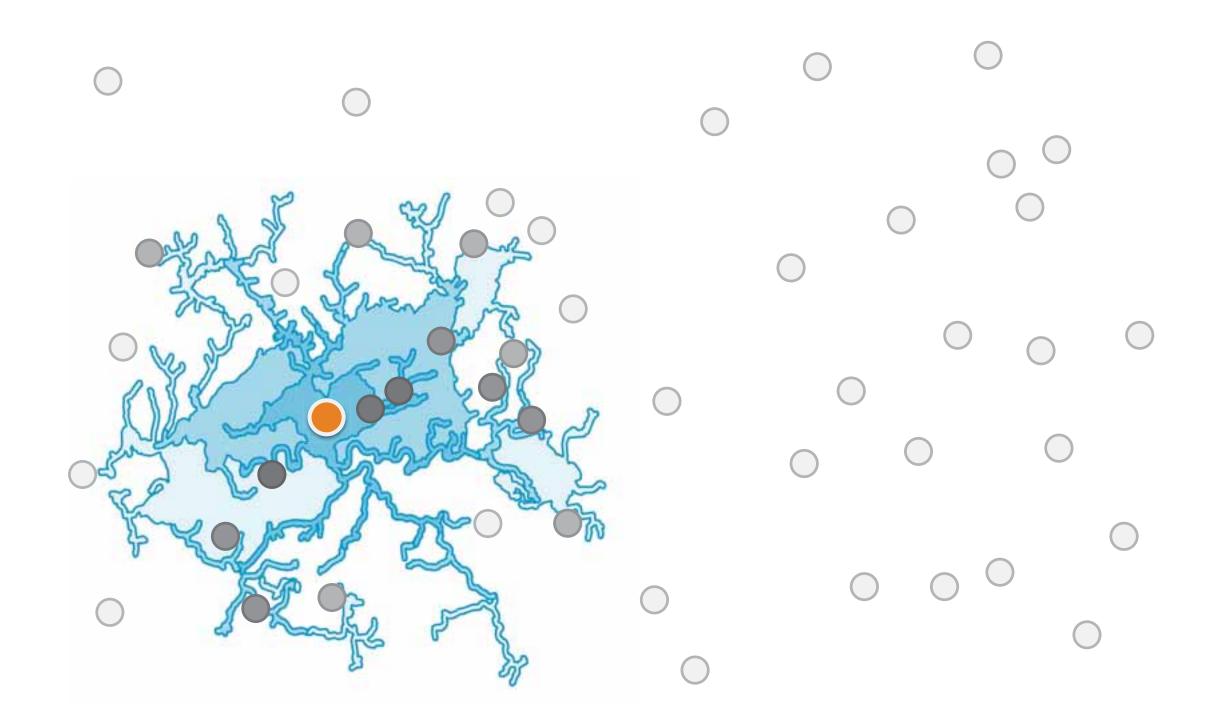# K Nearest Neighbors

Generate Network Spatial Weights

constructs a spatial weights matrix file (.swm) using a Network dataset, defining feature spatial relationships in terms of the underlying network structure
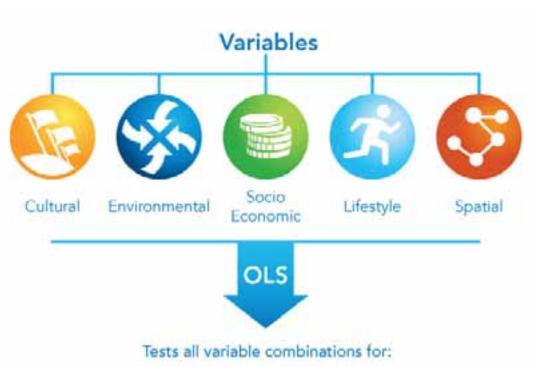
Ordinary Least Squares

performs global linear regression to model a dependent variable in terms of its relationships to a set of explanatory variables

a statistical process
for estimating
the relationships
between variables

Exploratory Regression

evaluates all possible combinations of the input candidate explanatory variables, looking for properly specified OLS models

Variables

Cultural	Environmental	Socio Economic	Lifestyle	Spatial

OLS

Tests all variable combinations for:

1. Redundancy
2. Completeness
3. Significance
4. Bias
5. Performance
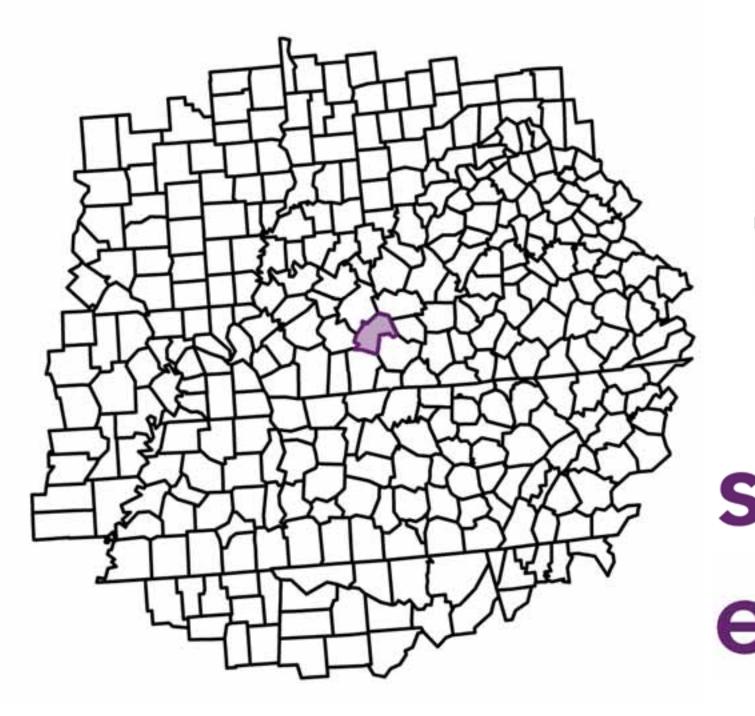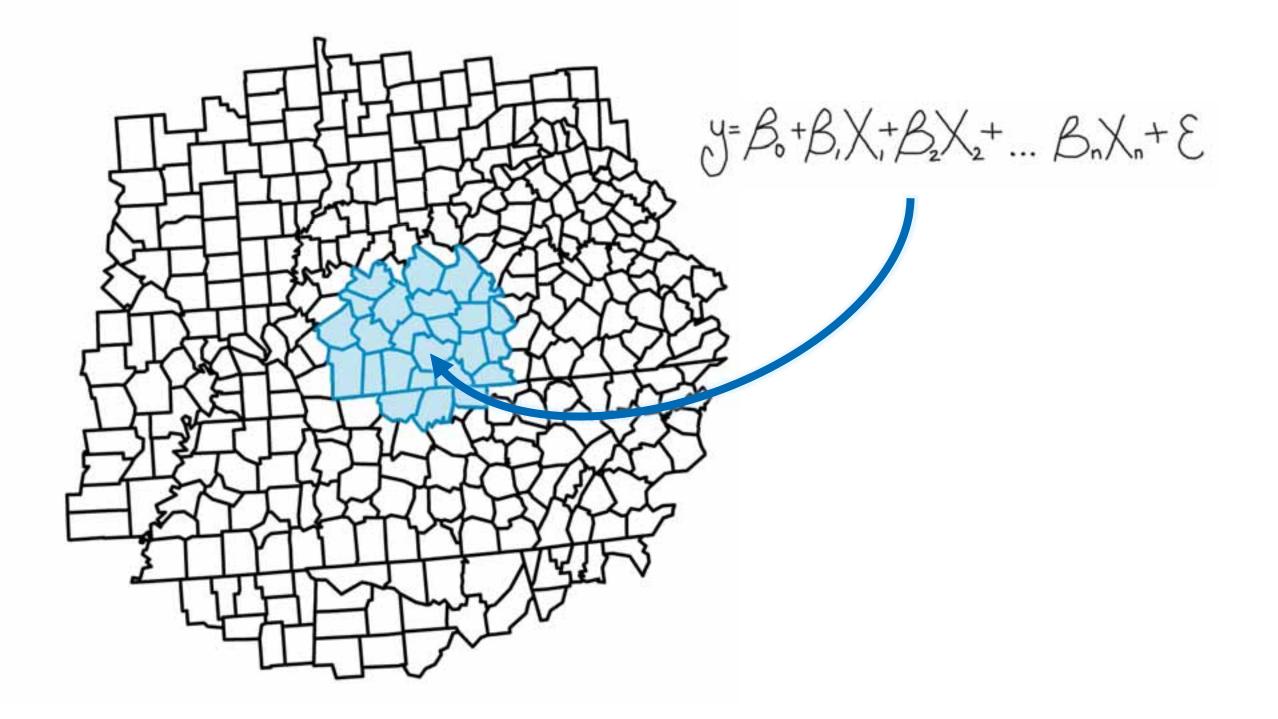
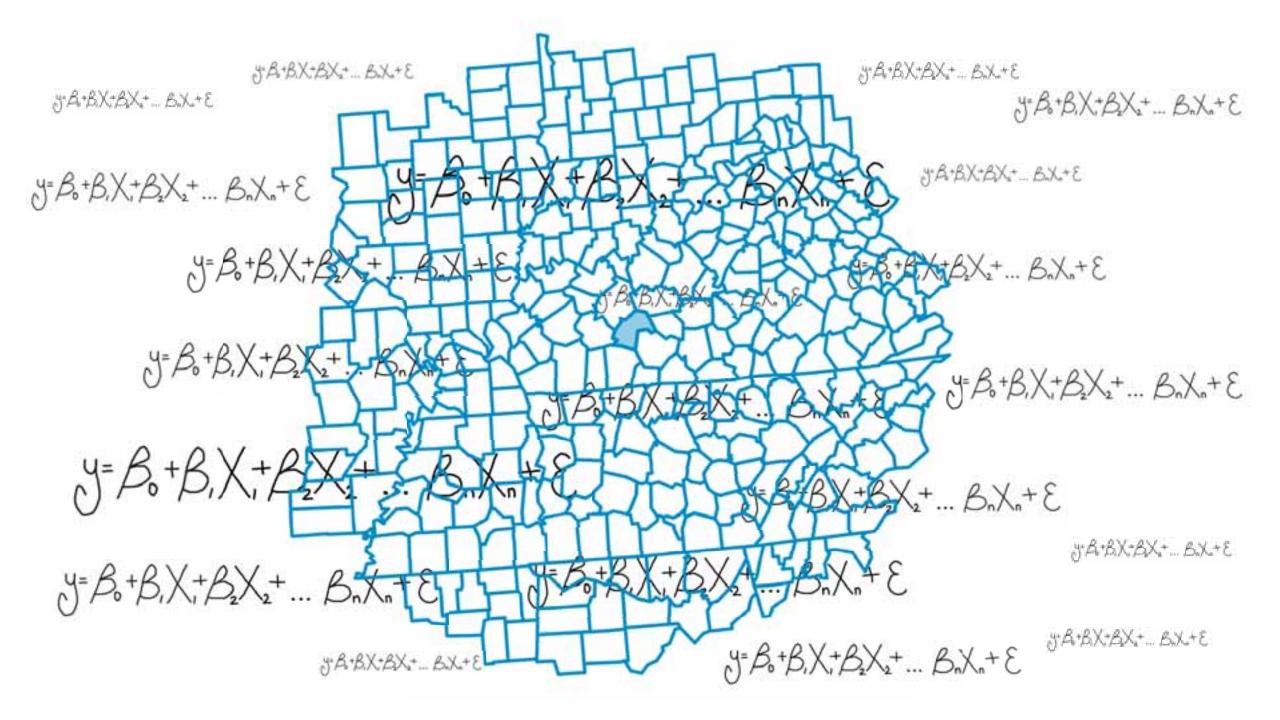Creates Output
Diagnostic Report

Geographically Weighted Regression

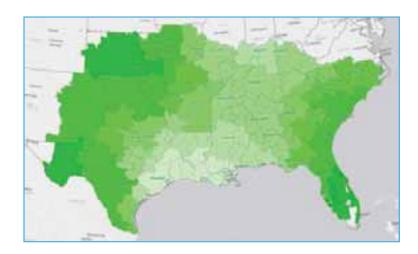performs a local form of linear regression used to model spatially varying relationships
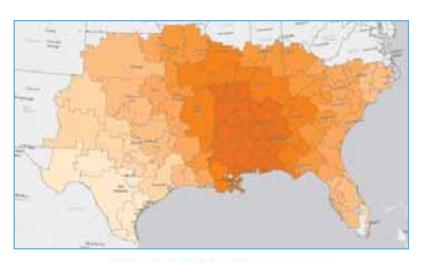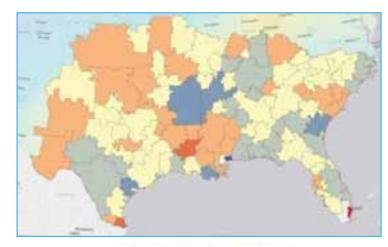
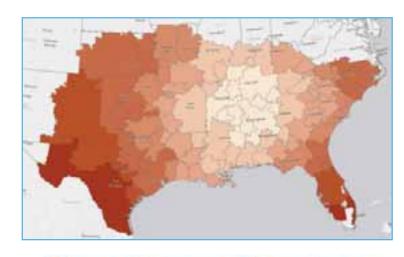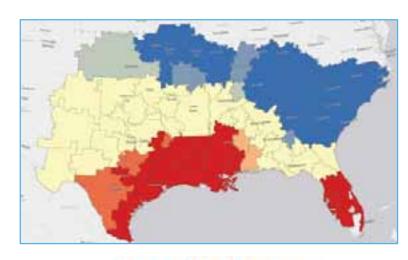each
feature
gets a
separate
equation

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \varepsilon$$

Local R-Squared

Coefficients

Residuals

Condition Number

Predictions

esri

lbennett@esri.com
fvale@esri.com

Want to learn more???
esriurl.com/spatialstats