esri HEALTH AND HUMAN SERVICES
GIS CONFERENCE
MAKING PLACE MATTER

Decoding Health Data with Spatial Statistics

Flora Vale
Jenora D'Acosta

Wait a minute...

Wait a minute...

Where is Lauren??

Wait a minute...

Where is Lauren??

DAY 1

What are Spatial Statistics?
Measuring Geographic Distributions
Analyzing Patterns
Multivariate Analysis

DAY 2

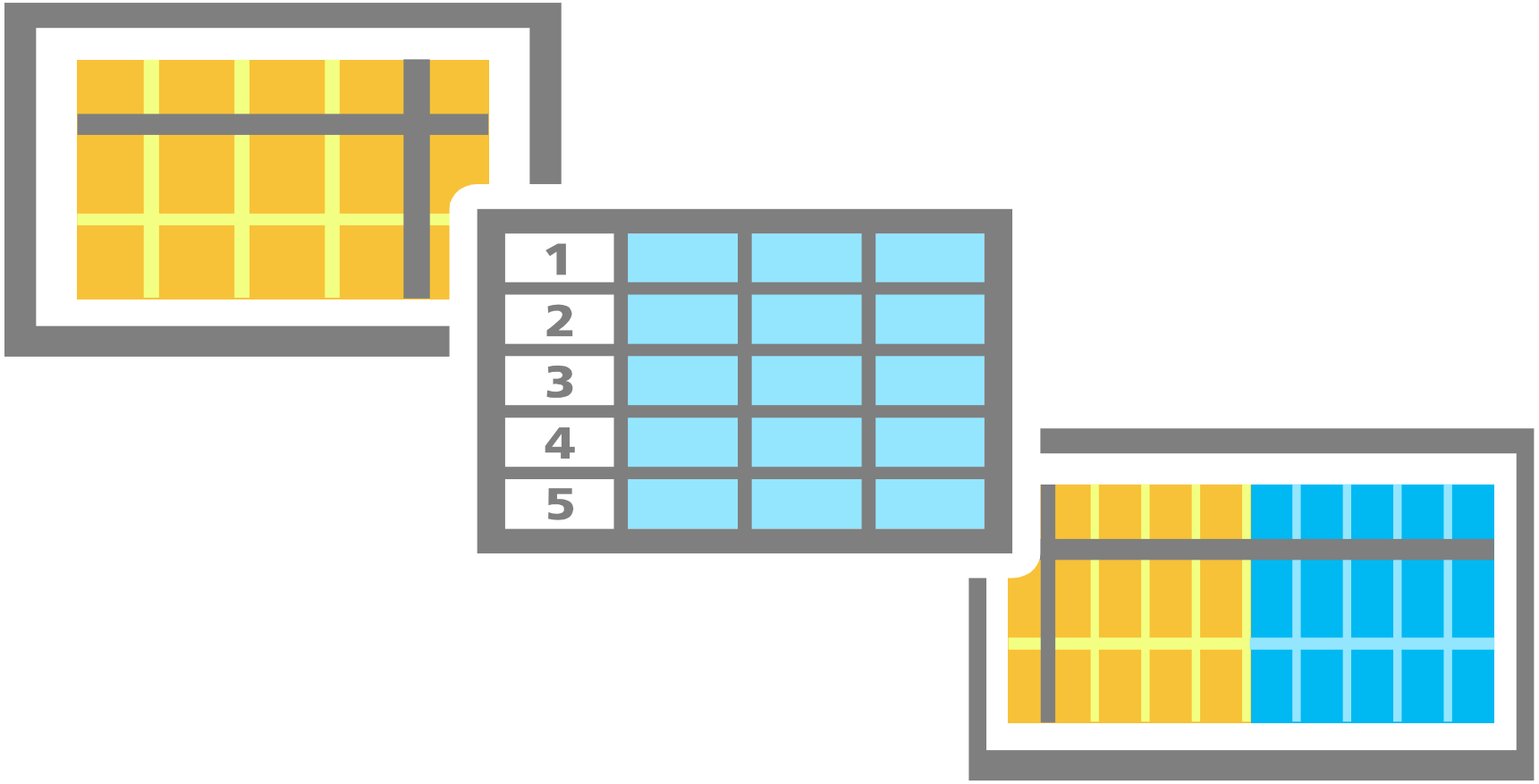
Subjectivity of Maps
Mapping Clusters
Space Time Pattern Mining
Modeling Spatial Relationships

Are you using

Data or Information?

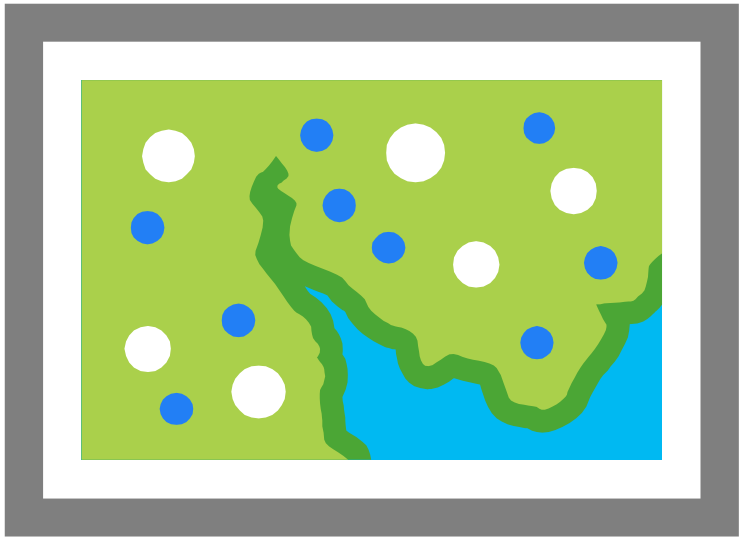
Spreadsheets

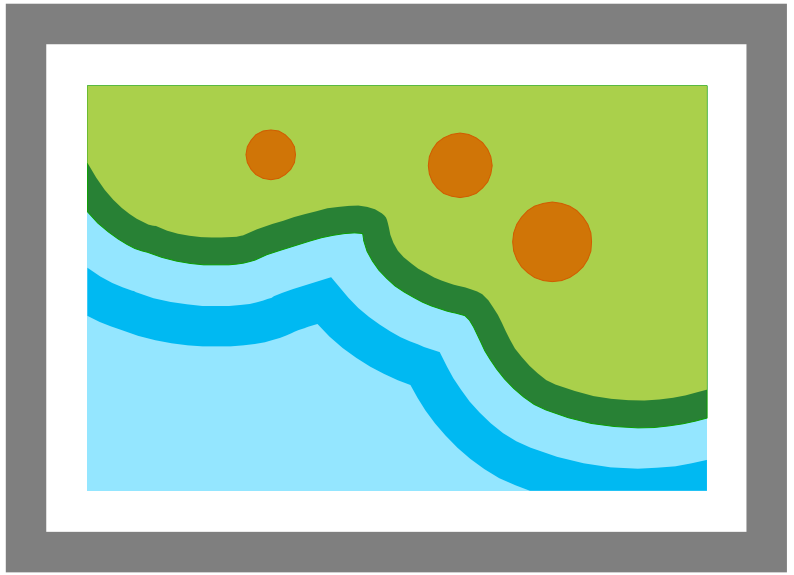
Data or Information?



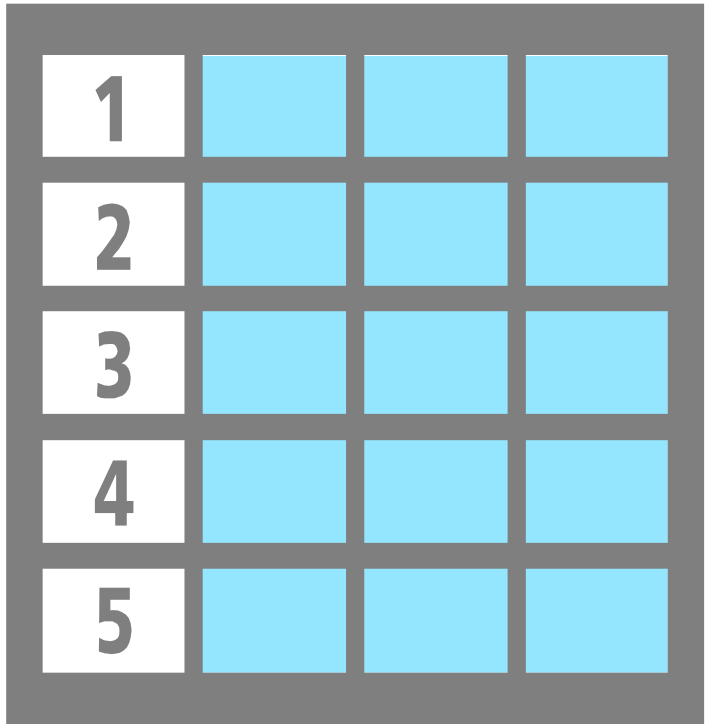
Maps

Data or Information?

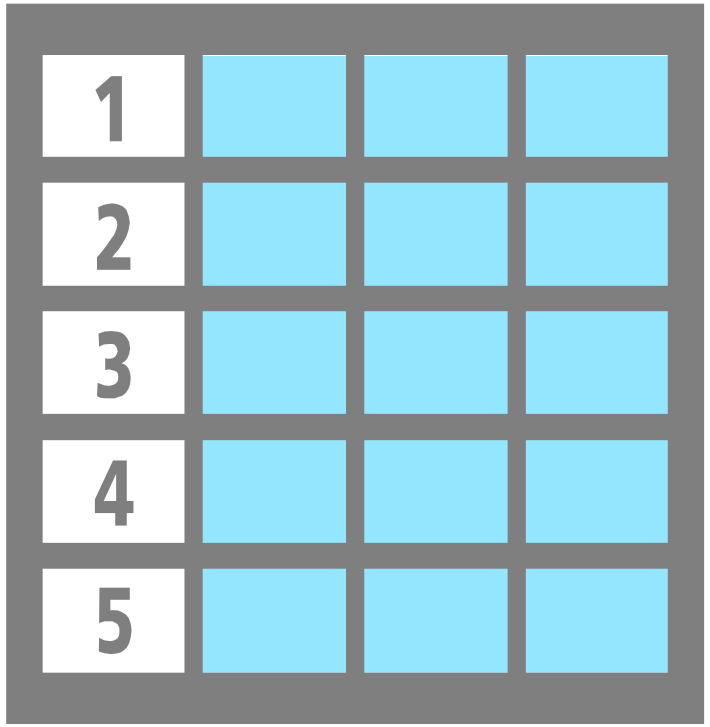




When you look at a spreadsheet…



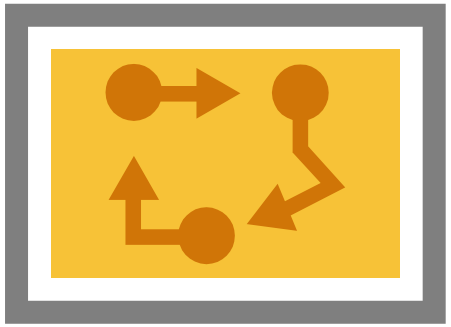
You ask for more

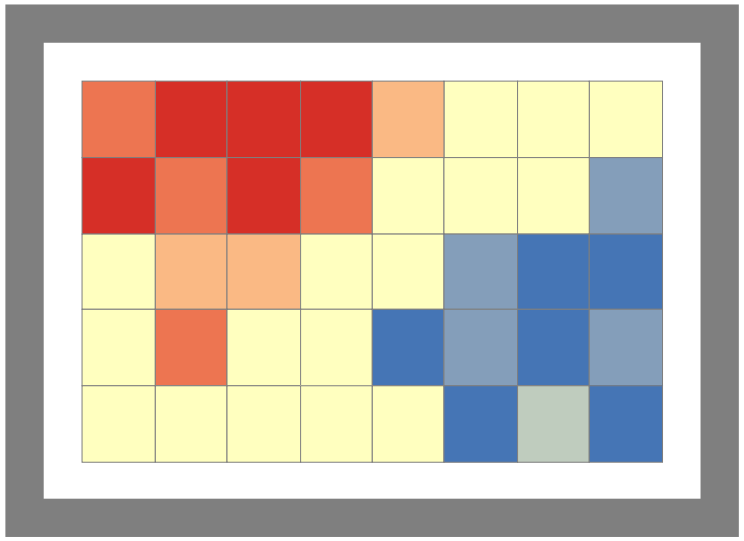


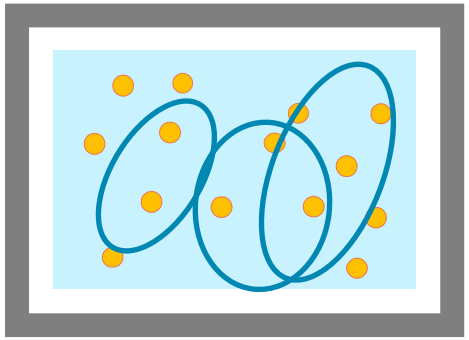
- Mean
- Standard Deviations
- Min and Max
- …

Same goes for maps!

We can do more







What are Spatial Statistics?

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

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

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

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

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

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

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

coincidence

connectivity

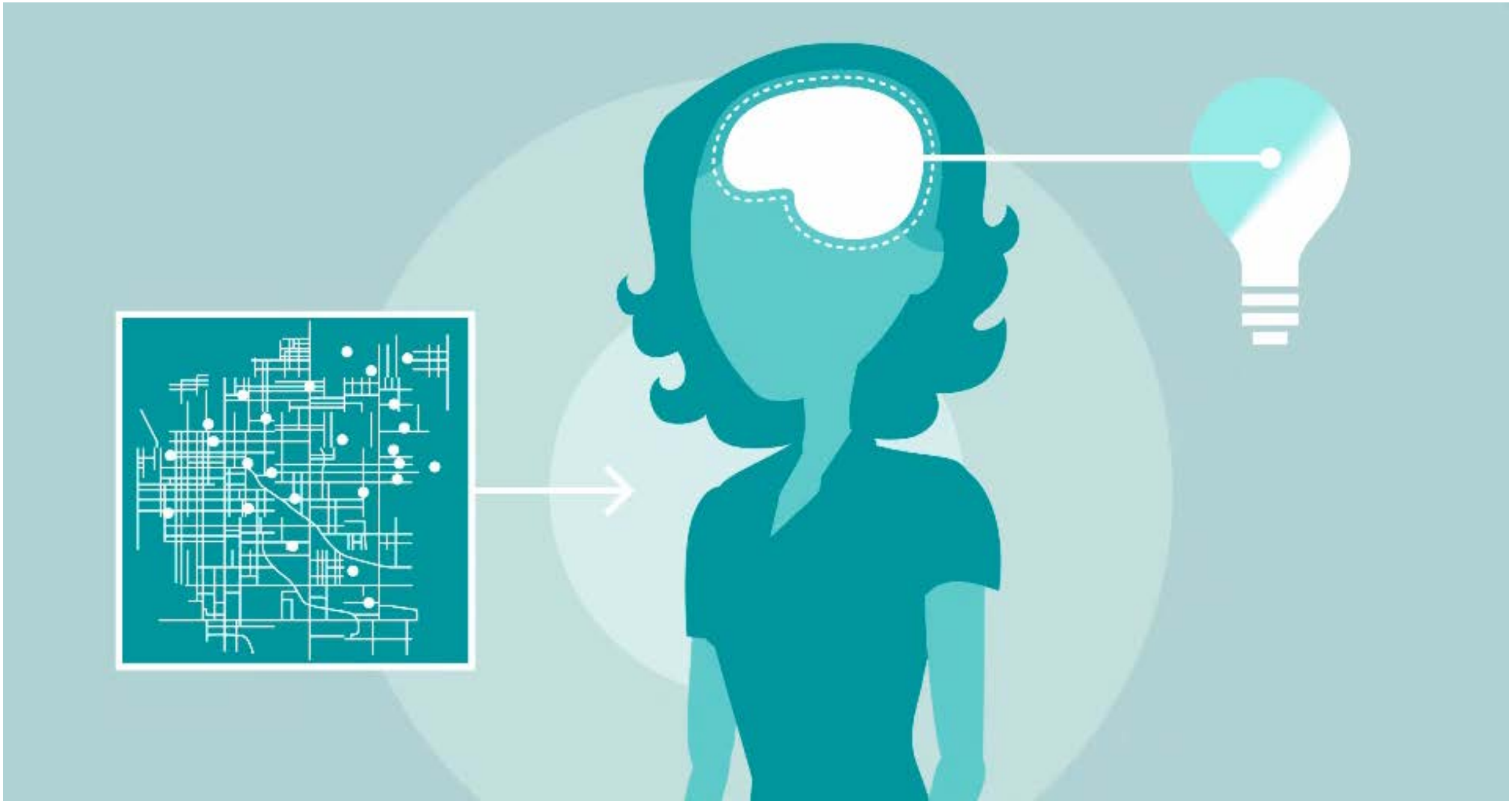
area

proximity

orientation

length

direction



- 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

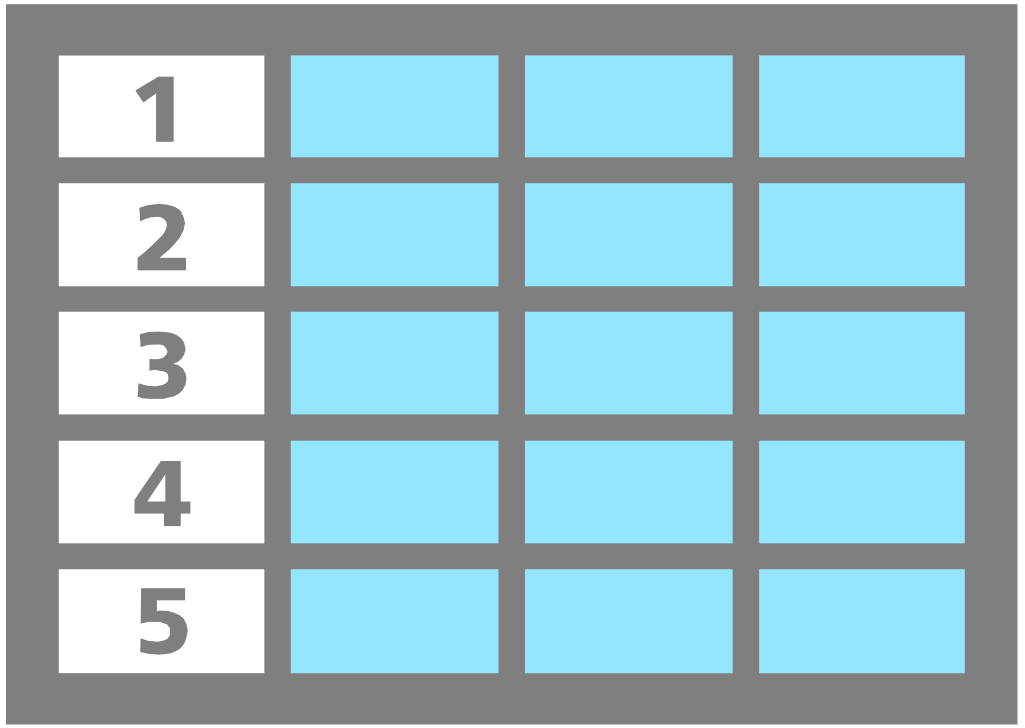
Descriptive

Statistics

Mean

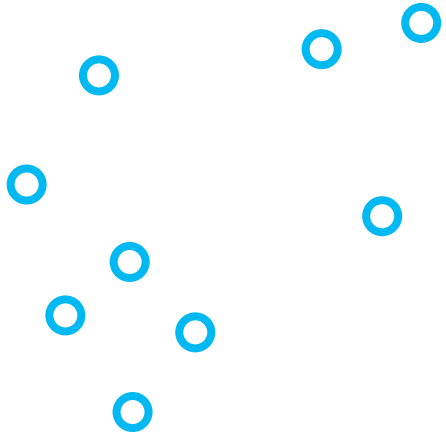
Median

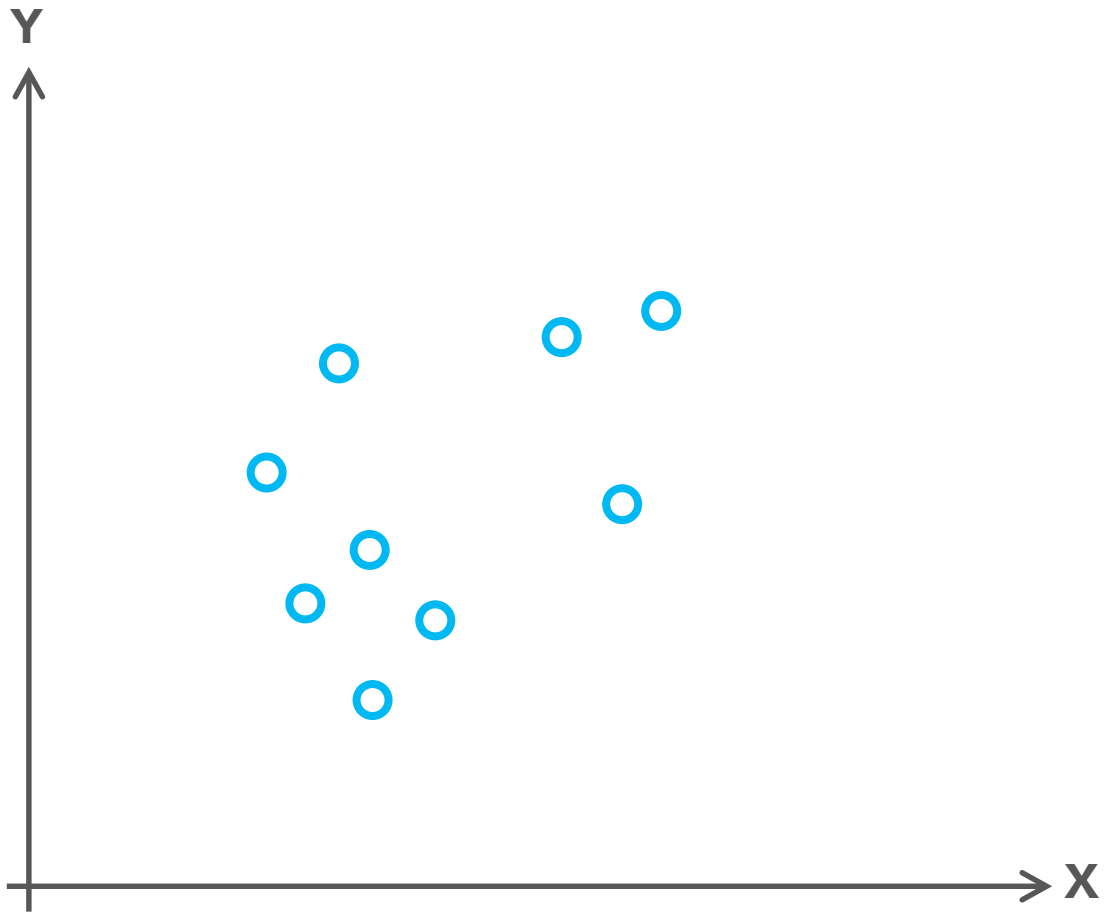
Standard Deviation

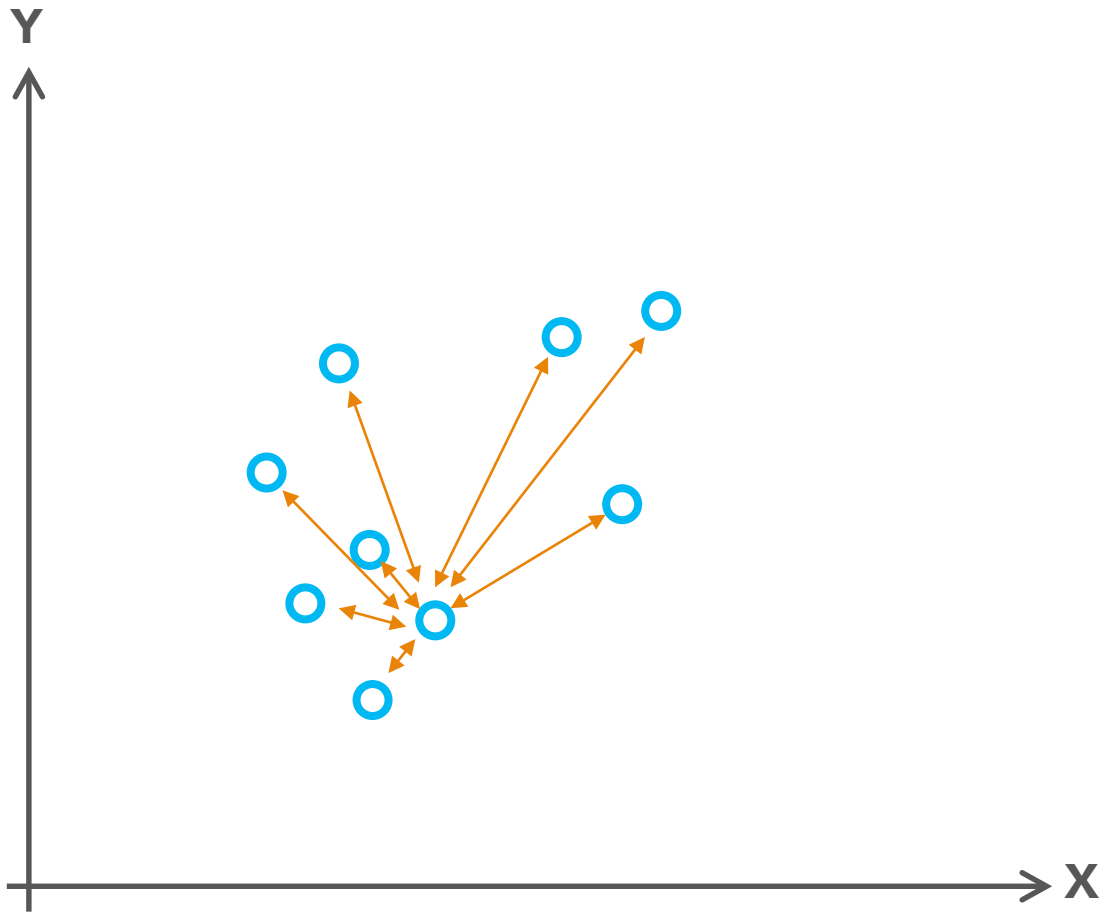


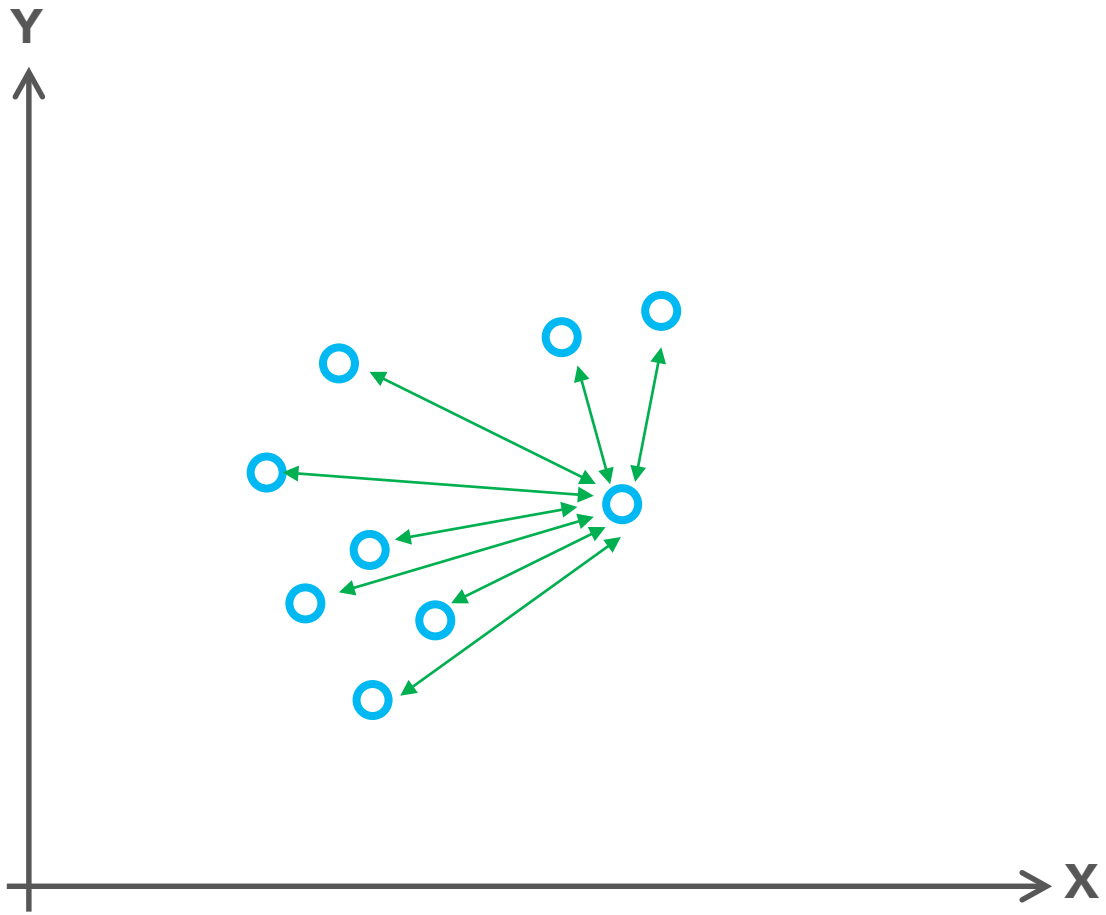
Central Feature

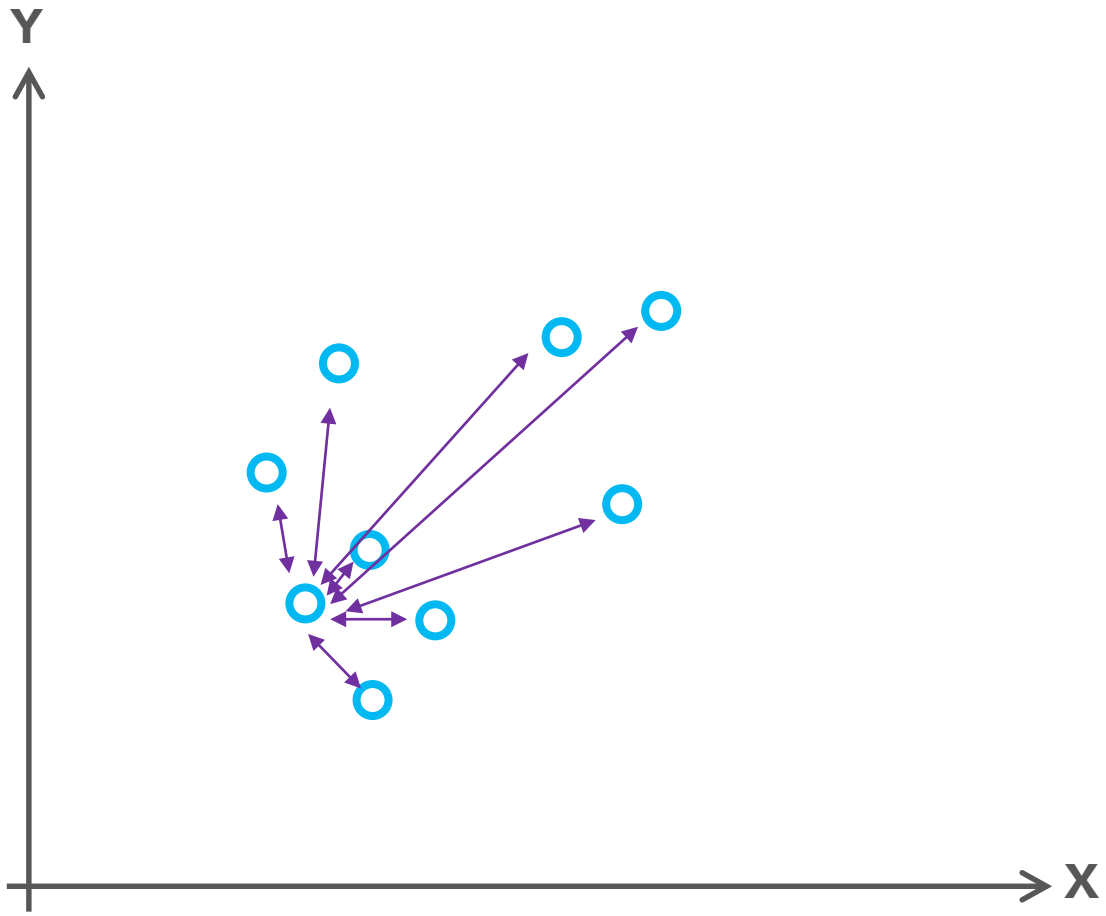
identifies the most centrally located feature in a point, line, or polygon feature class

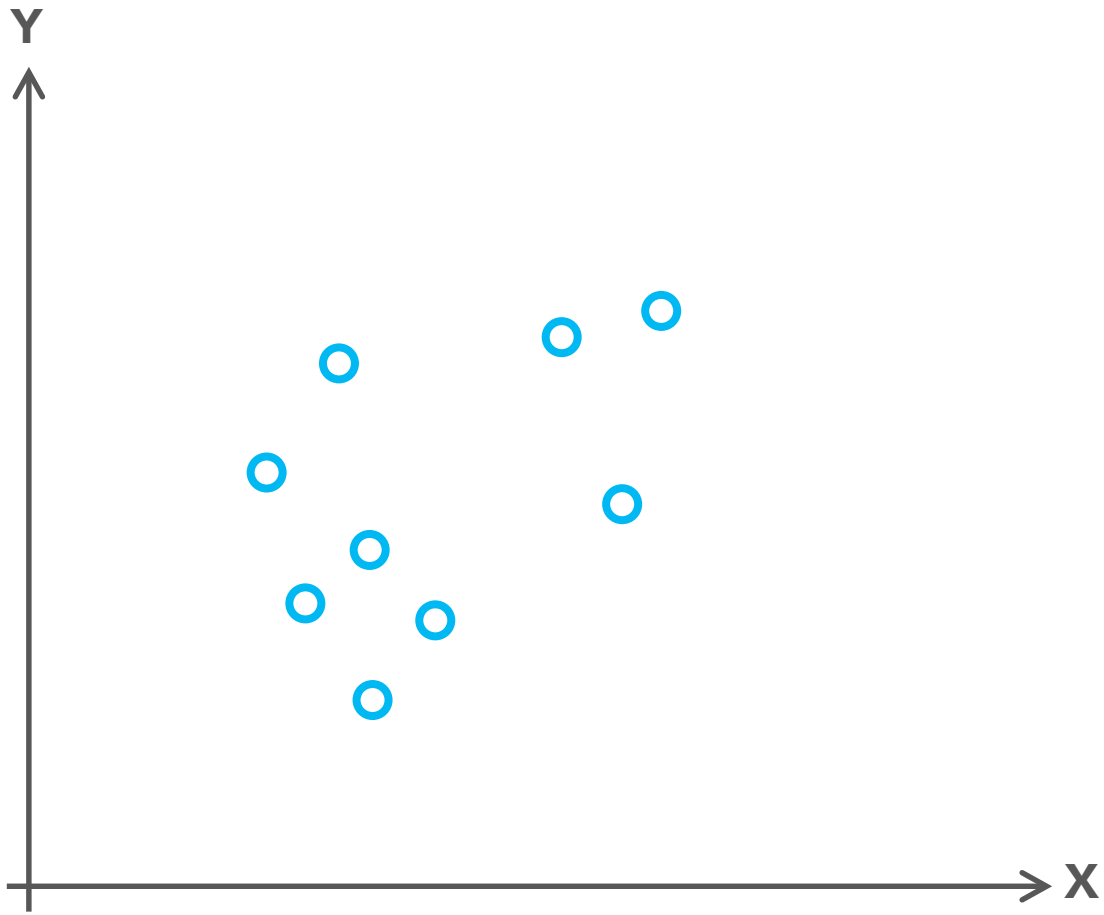


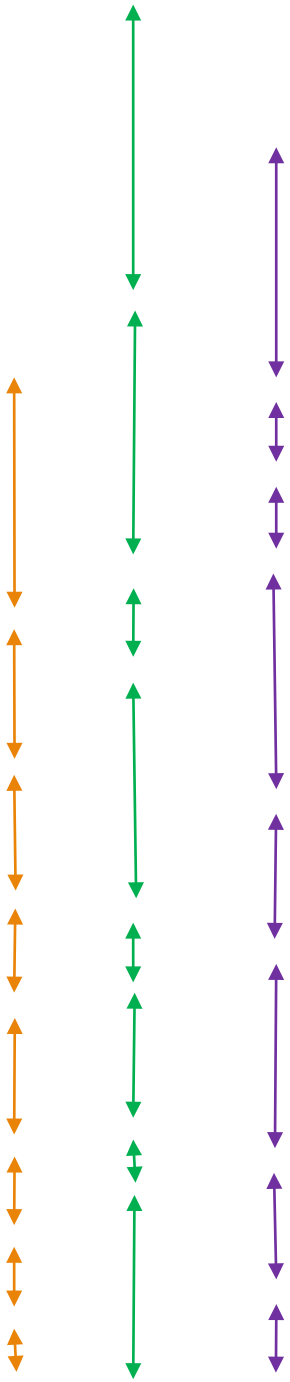

Y
X

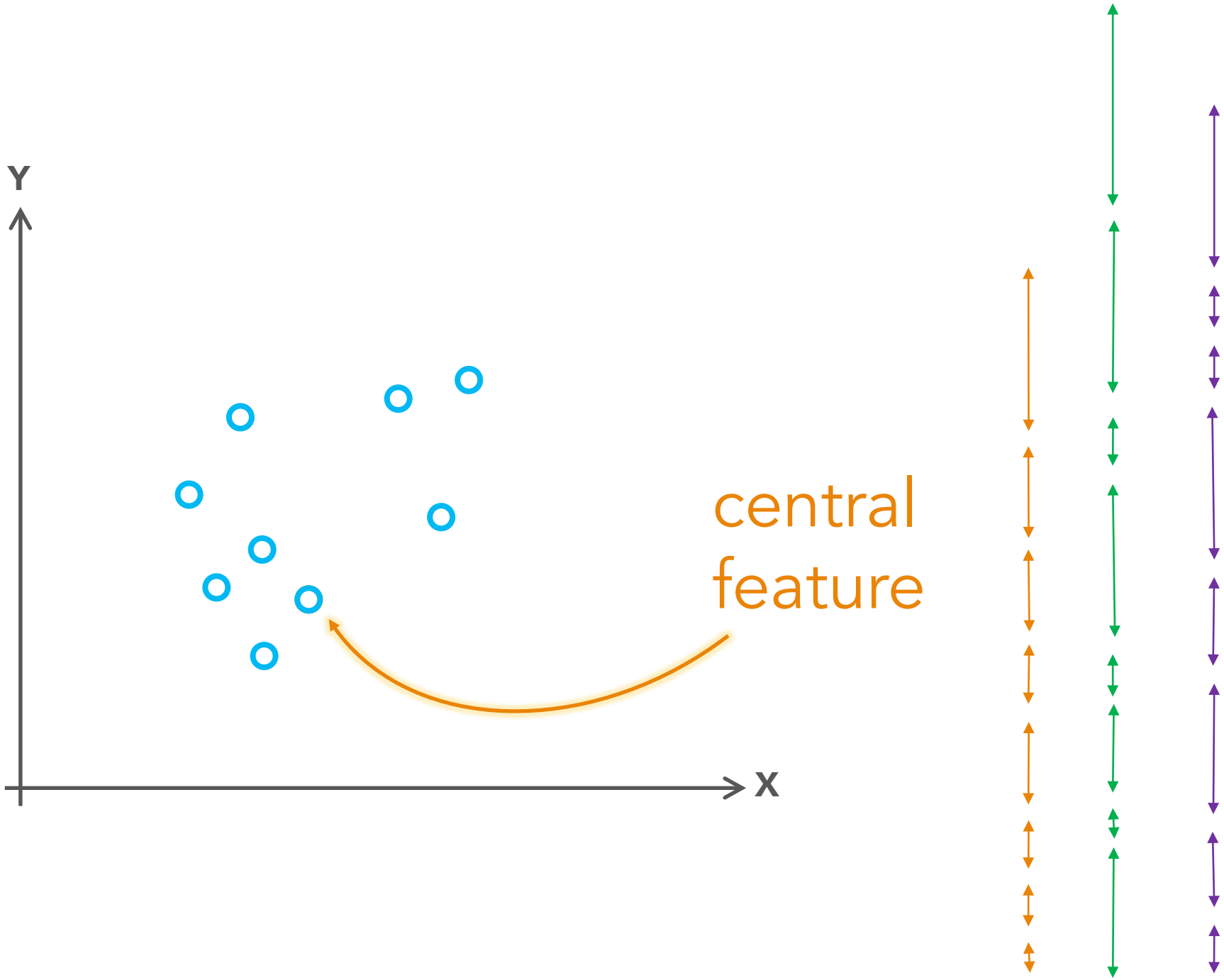

Y
X


Y
X


Y
X

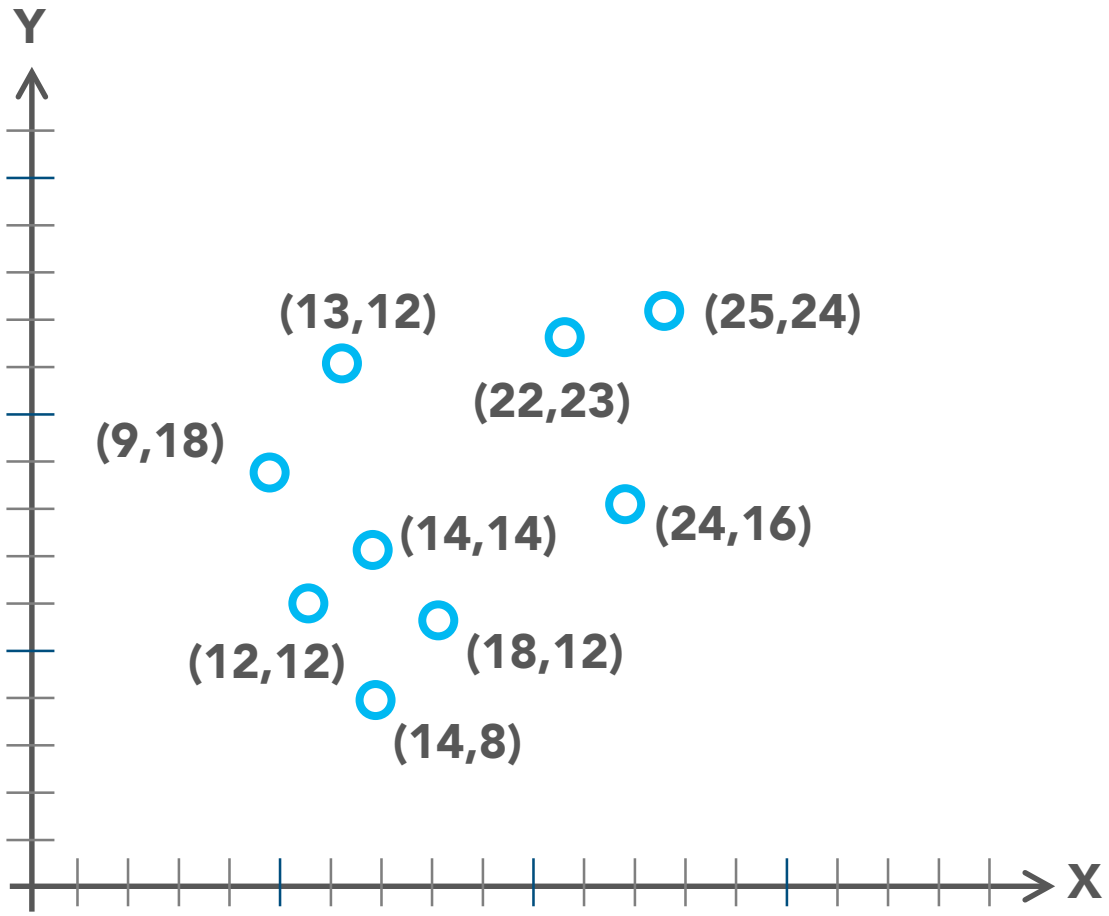

Y
X

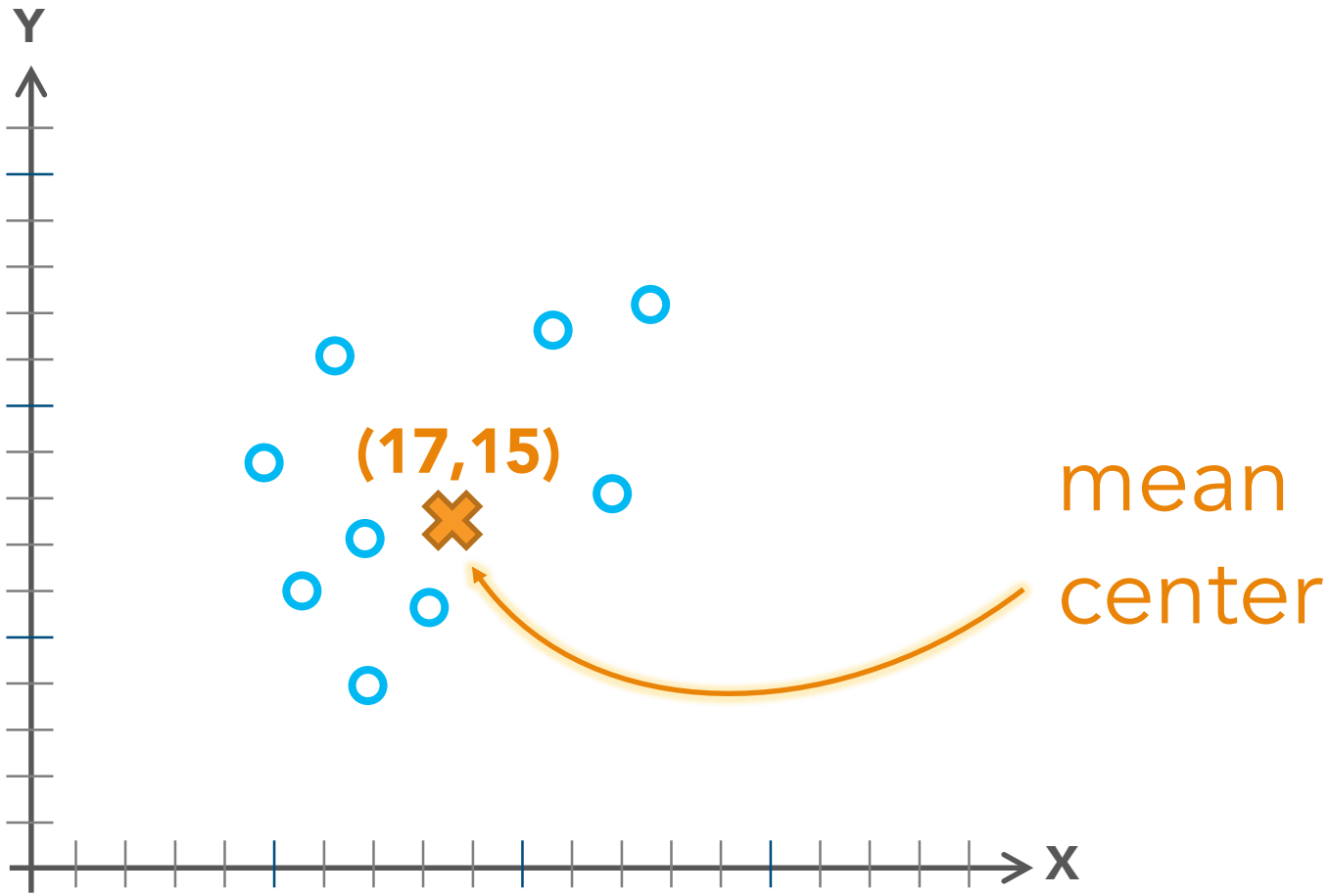



Y
X
central
feature

Mean Center

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


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



(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

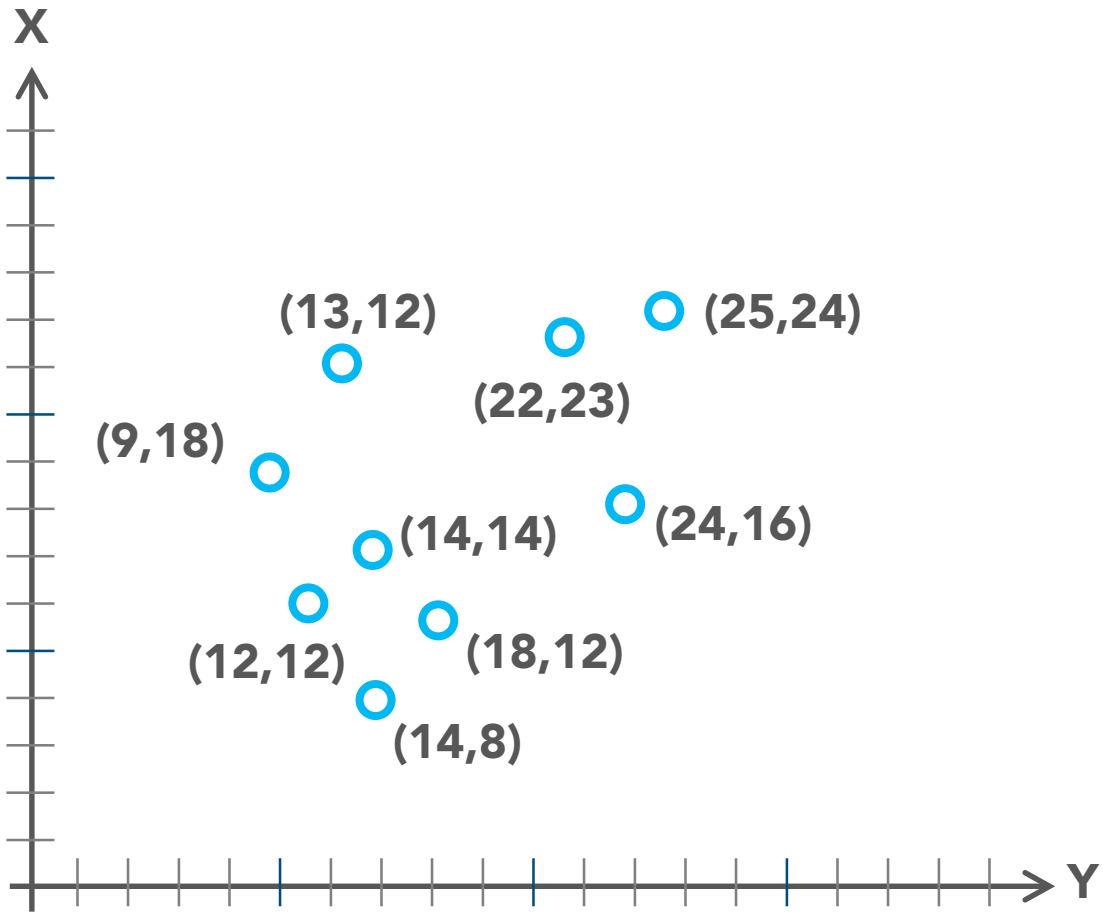
---

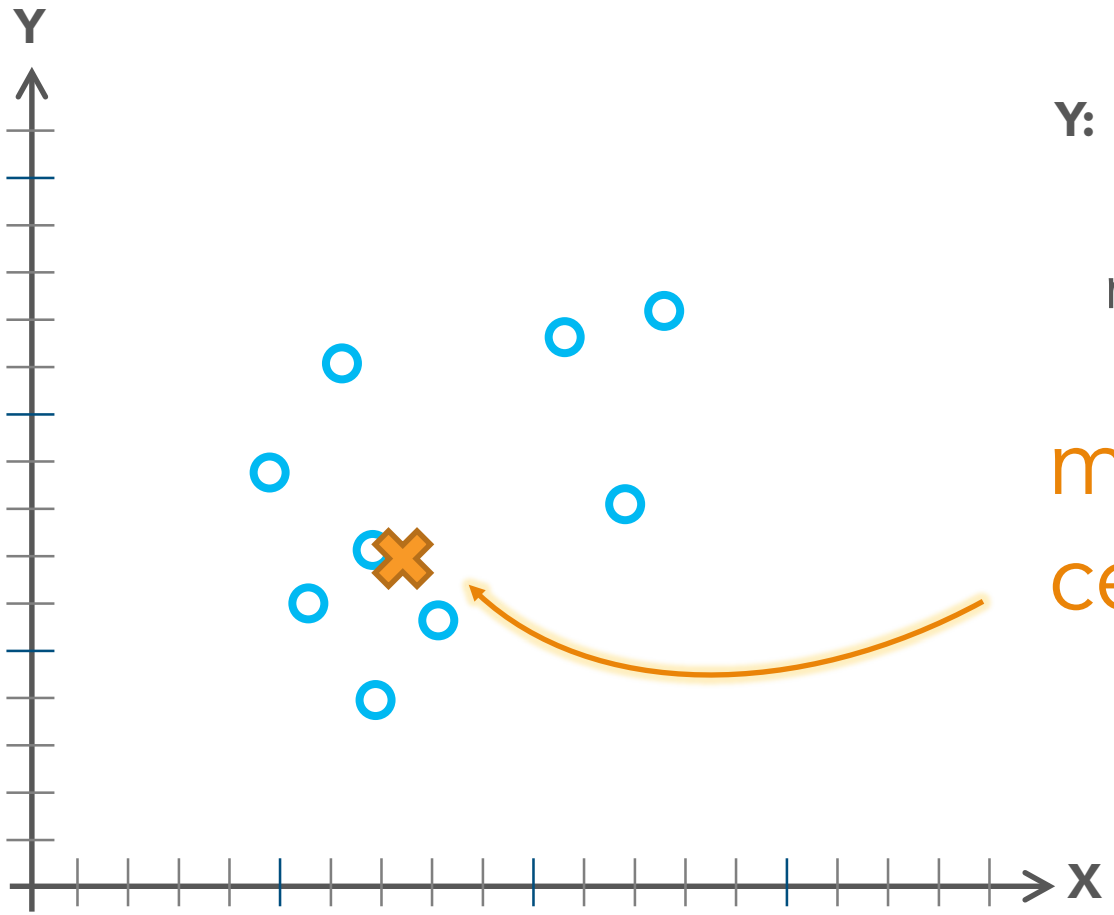
mean =

(17,15)

Median Center

identifies the location that minimizes overall Euclidean distance to the features in a dataset


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



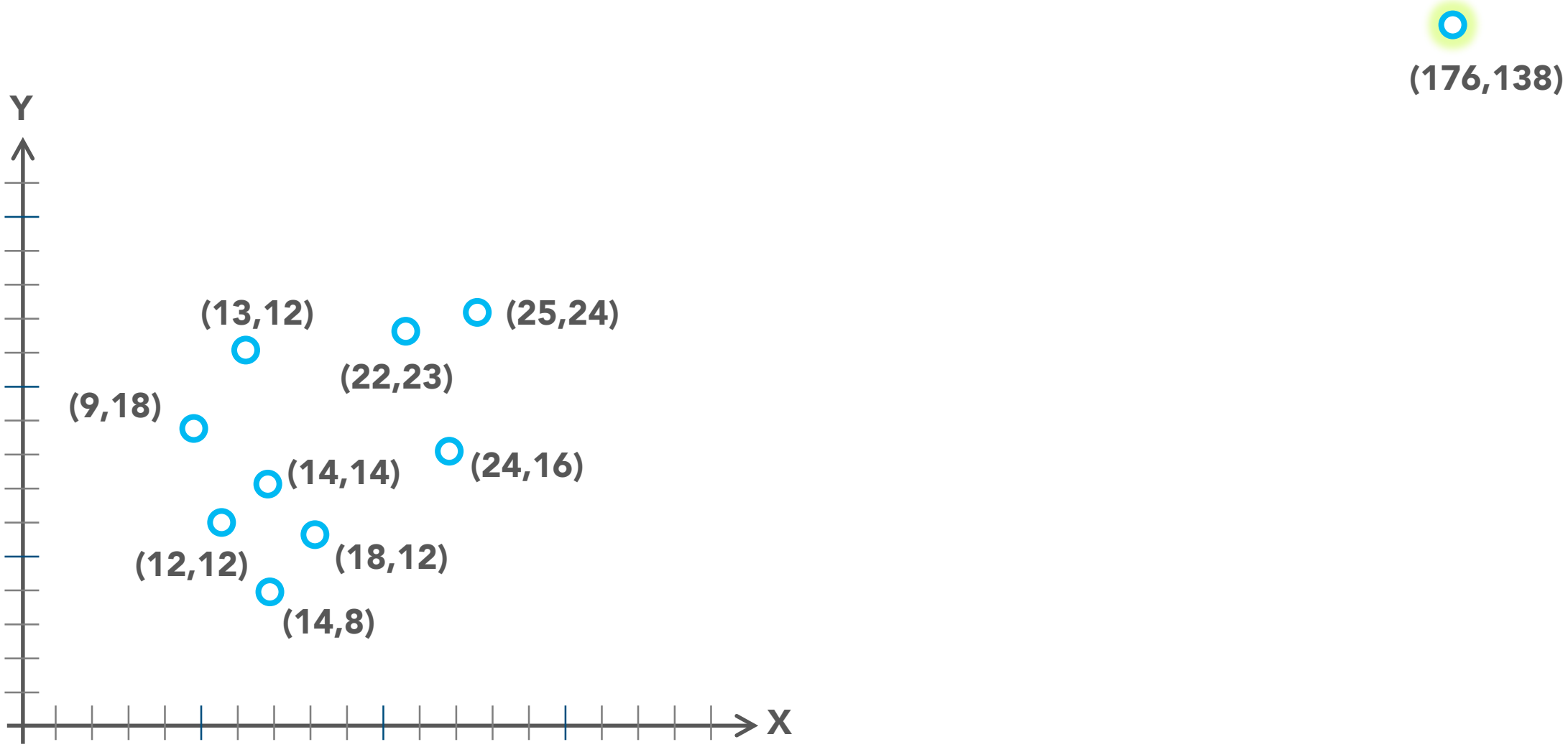
X: ~~25~~ . ~~24~~ . ~~22~~ . ~~18~~ . 14 . ~~14~~ . ~~13~~ . ~~12~~ . ~~9~~

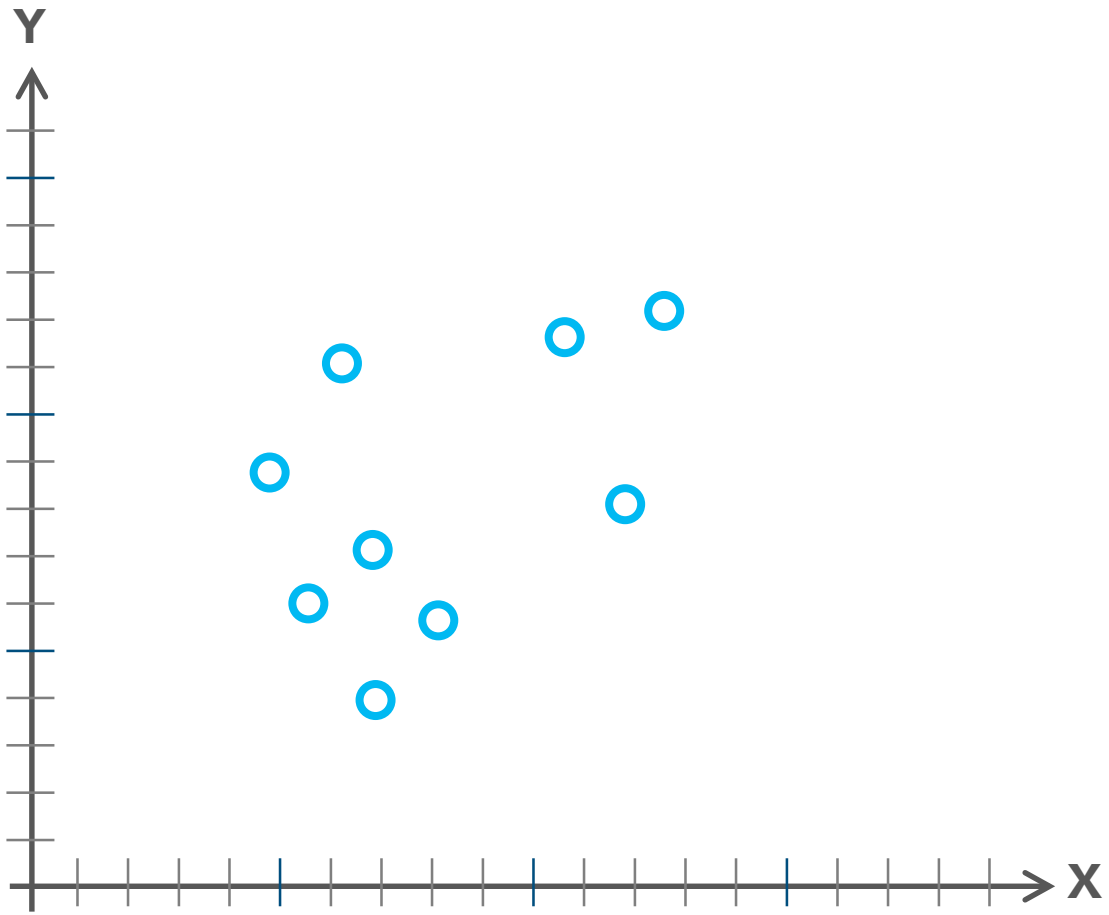
Y: ~~24~~ . ~~23~~ . ~~18~~ . ~~16~~ . 14 . ~~12~~ . ~~12~~ . ~~12~~ . ~~8~~

median = **(14,14)**

median
center

Mean vs Median?


(176,138)
Y
(13,12)
(25,24)
(22,23)
(9,18)
(14,14)
(24,16)
(12,12)
(18,12)
(14,8)
X



(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

(74,38)

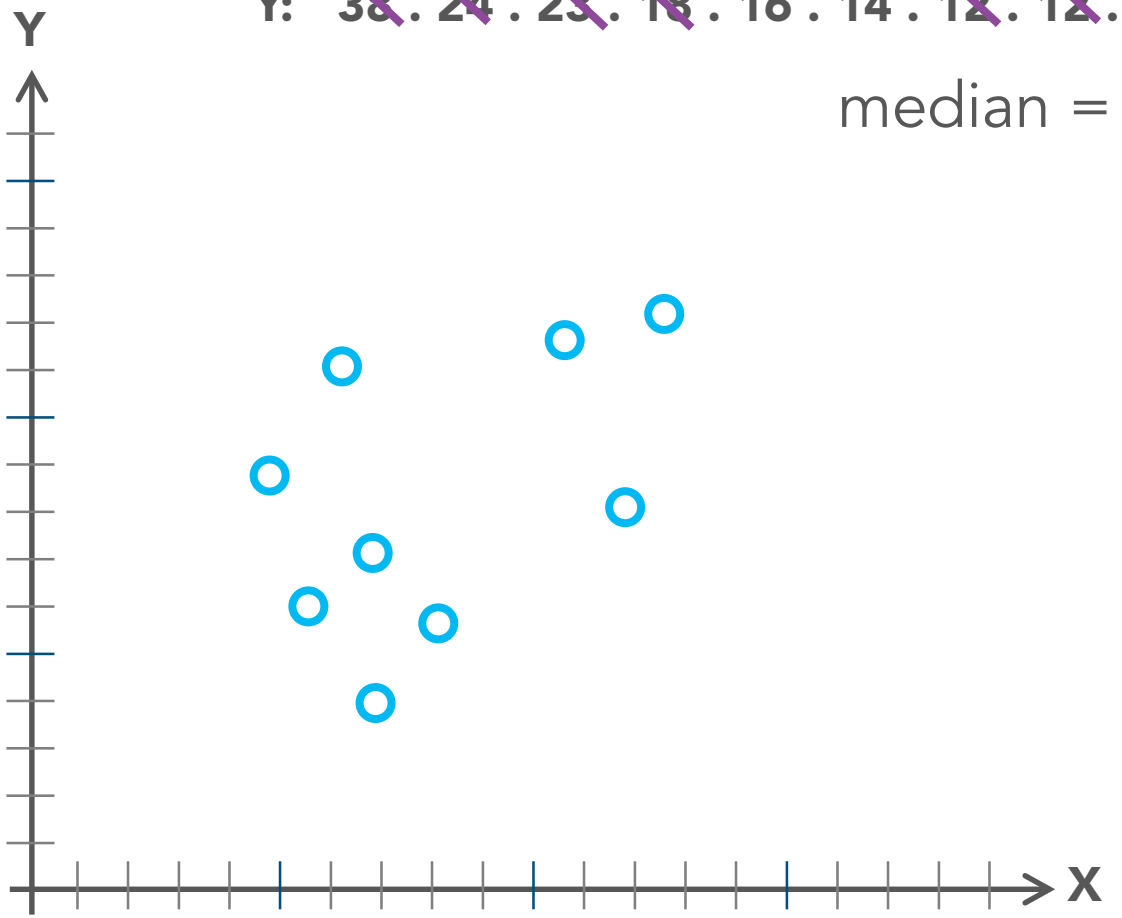
mean =

(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)**



(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

(74,38)

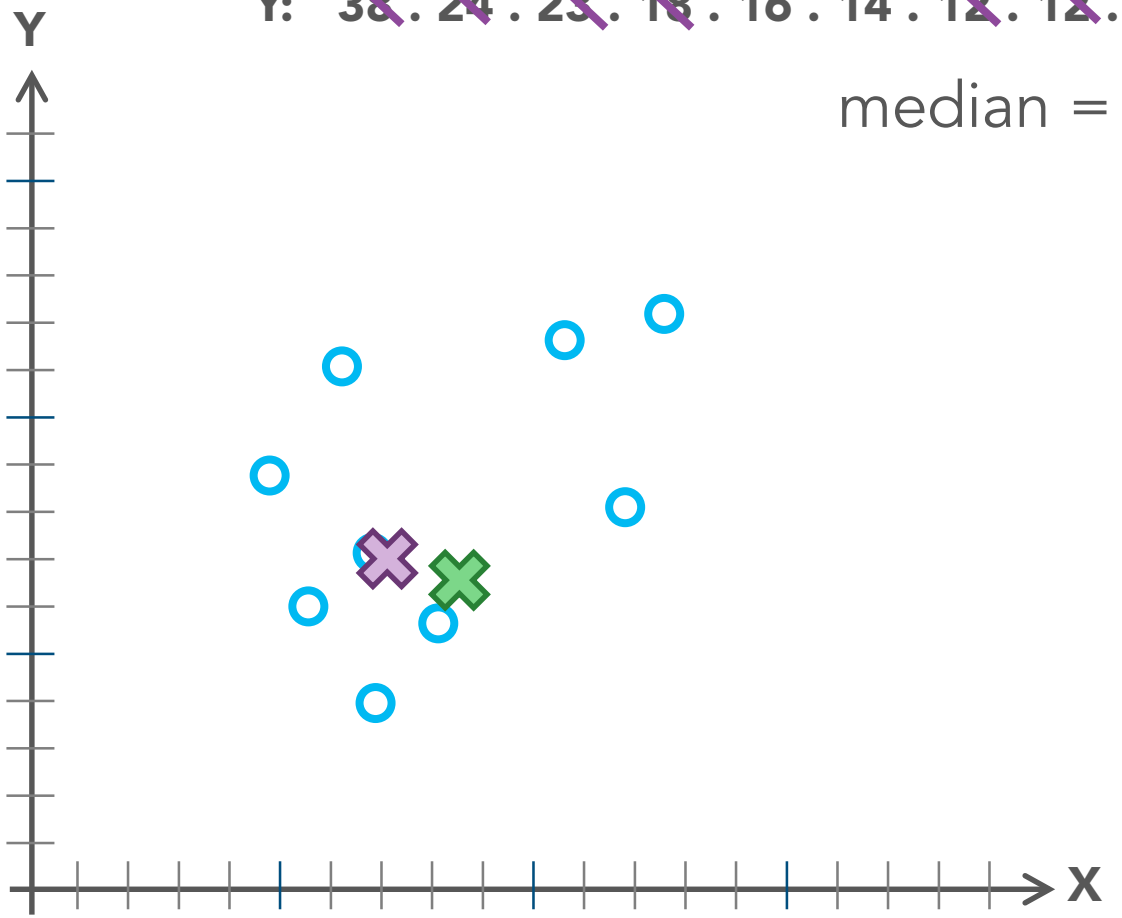
mean =

(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)**



(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

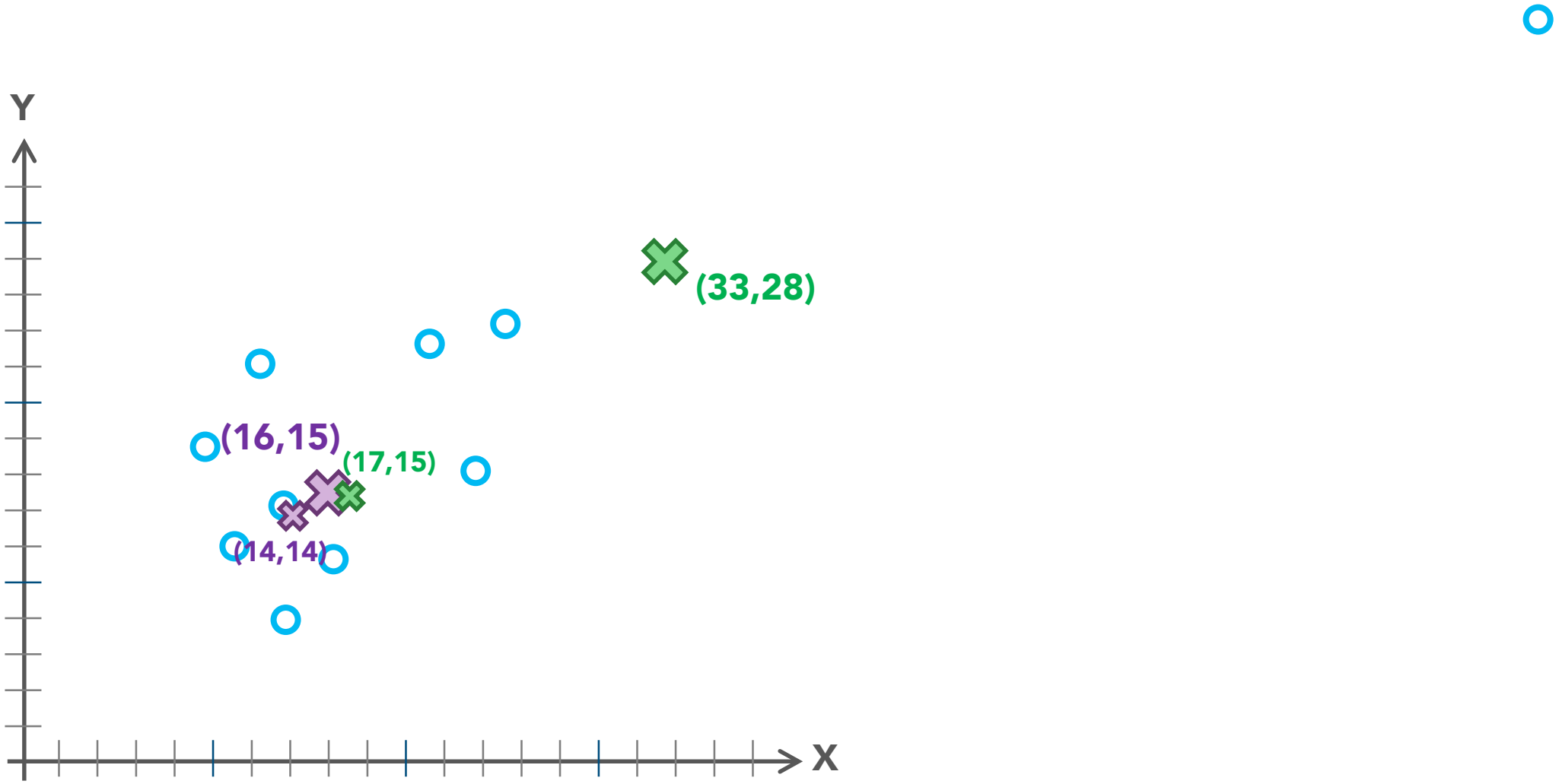
(14,8)

(9,18)

(74,38)

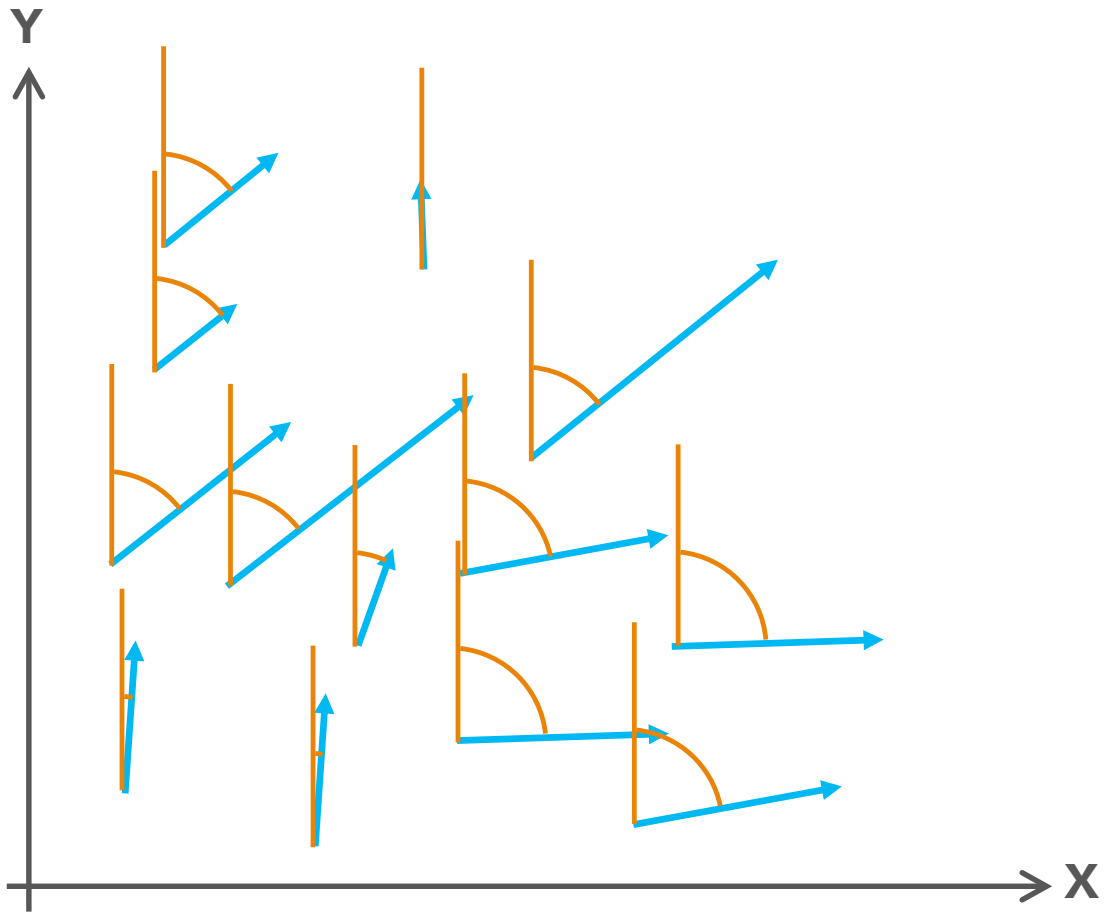
mean =

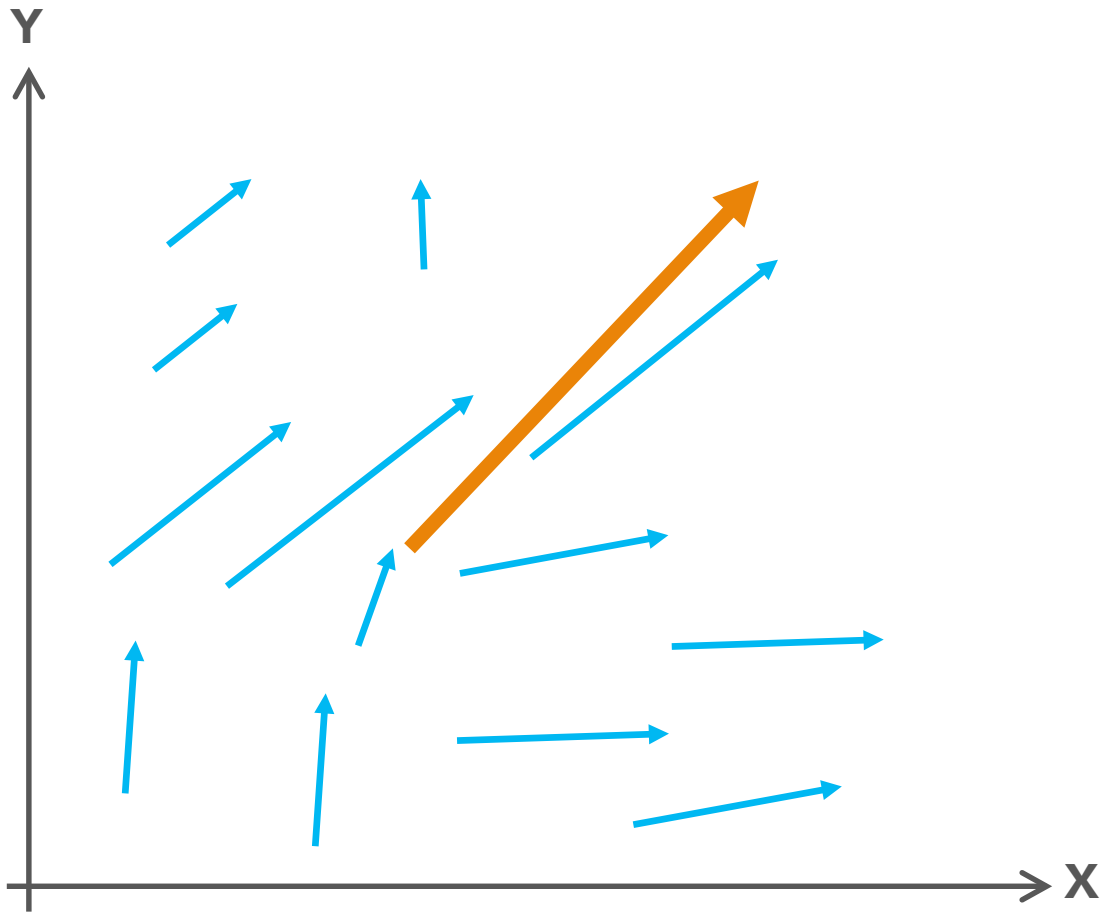
(33,28)


Y
X
(33,28)
(16,15)
(17,15)
(14,14)

Linear Directional Mean

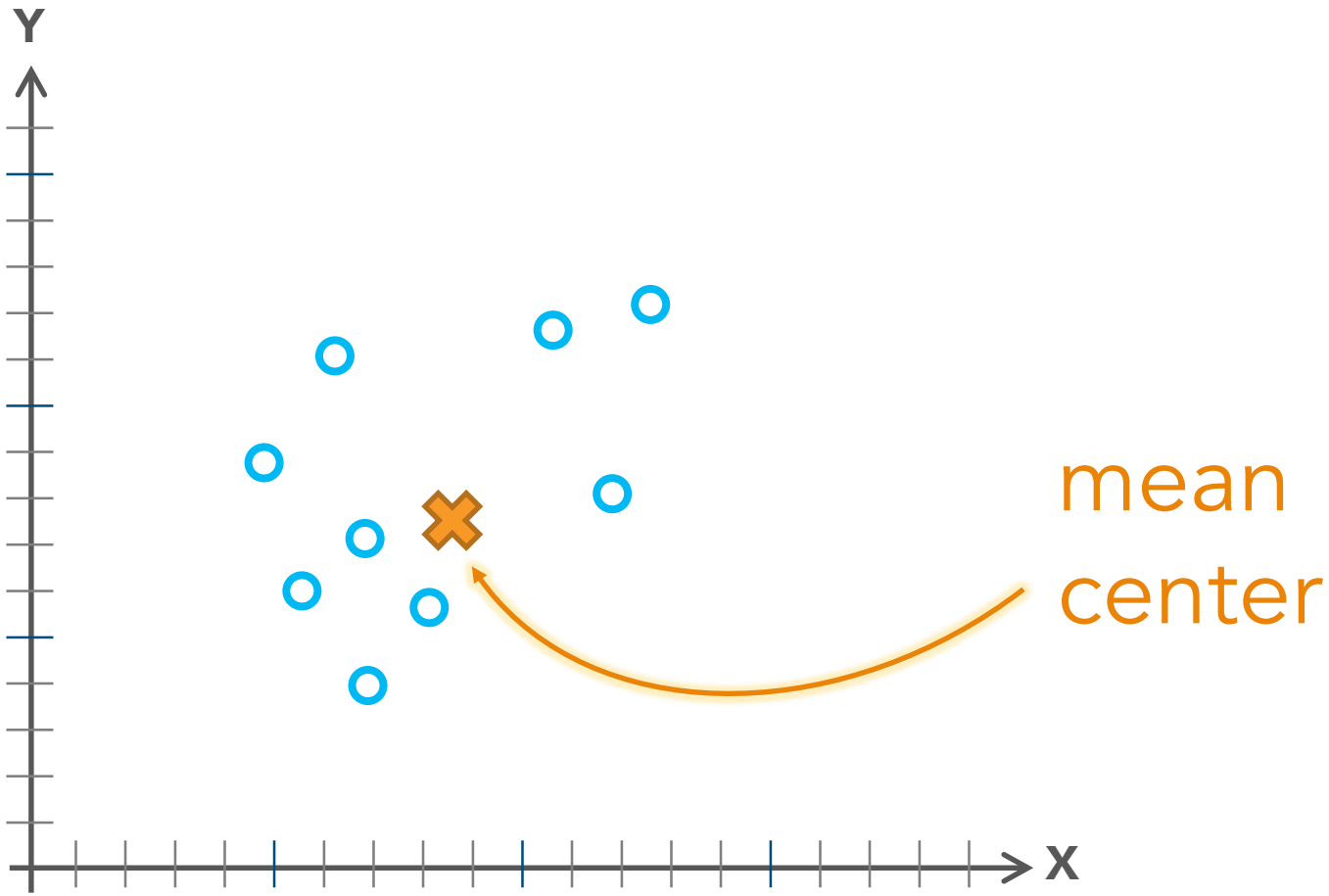
identifies the mean direction, length, and geographic center for a set of lines

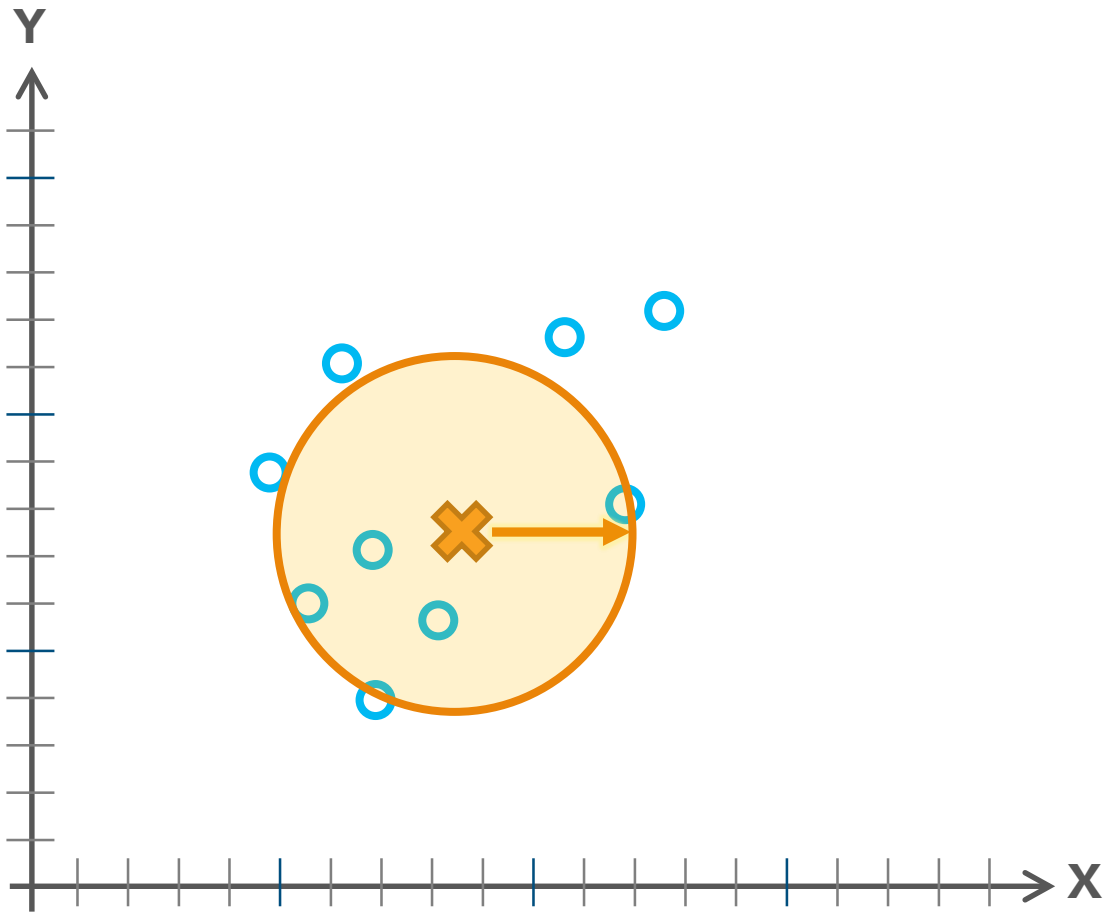

Y
X


Y
X

Standard Distance

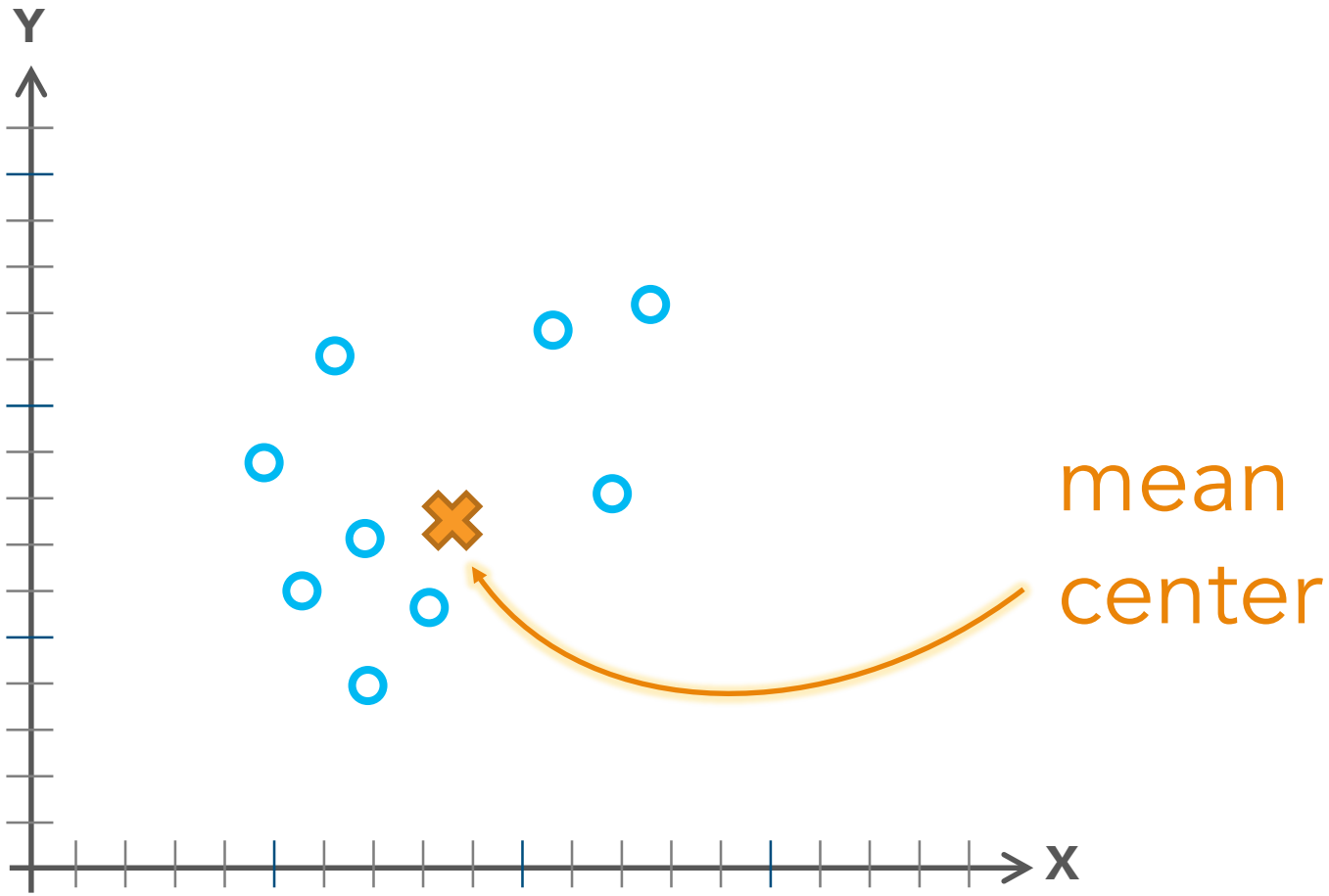
measures the degree to which features are concentrated or dispersed around the geometric mean center

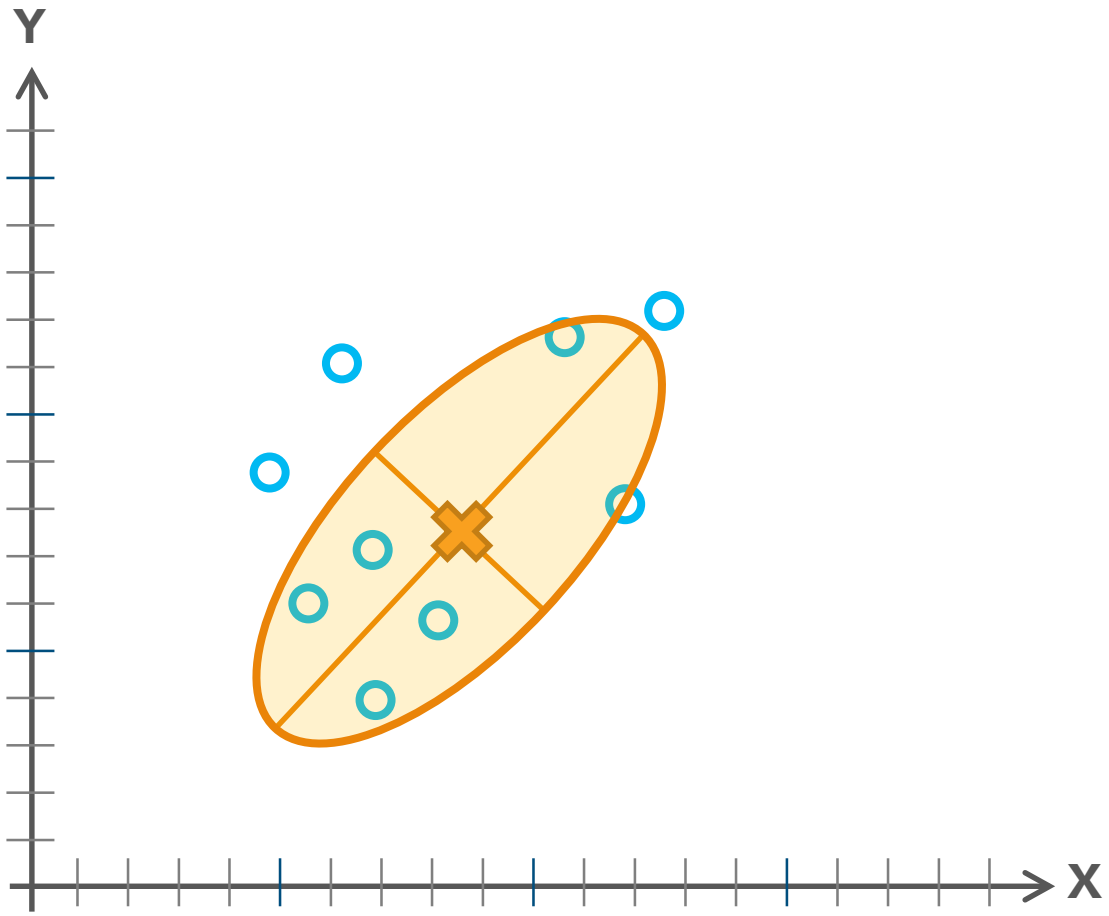

Y
X
mean
center


Y
X

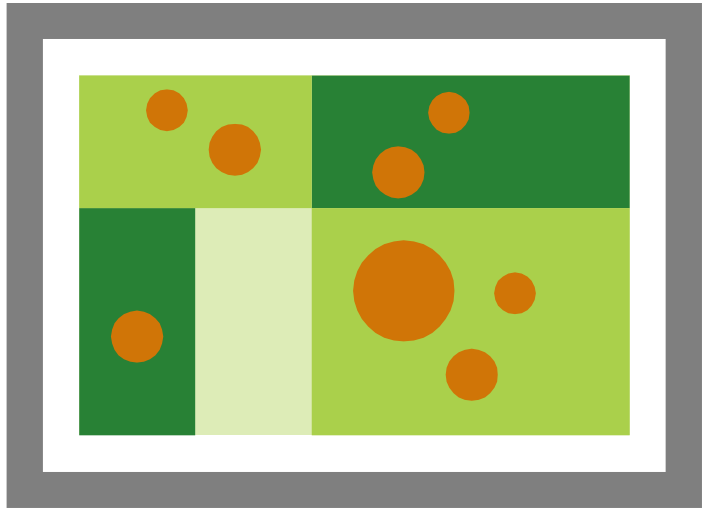
Directional Distribution
(Standard Deviational Ellipse)

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


Y
X
mean
center


Y
X

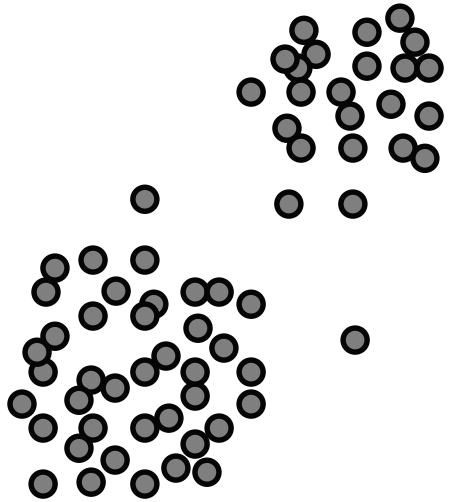
Analyzing Patterns



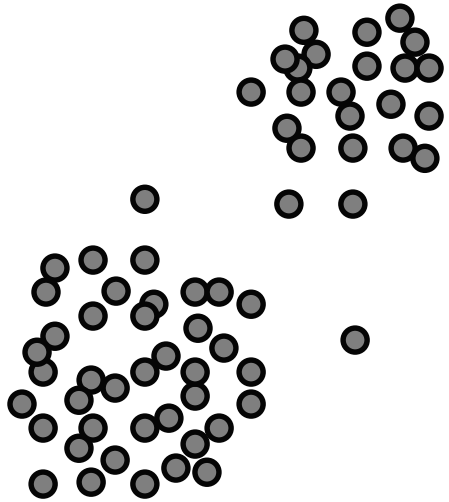
Global
Inferential
Statistics

Is there a PATTERN?

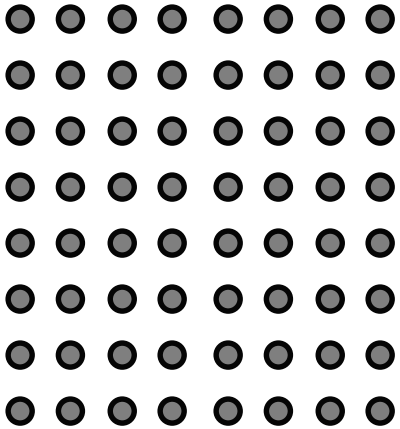
Clustered



Clustered

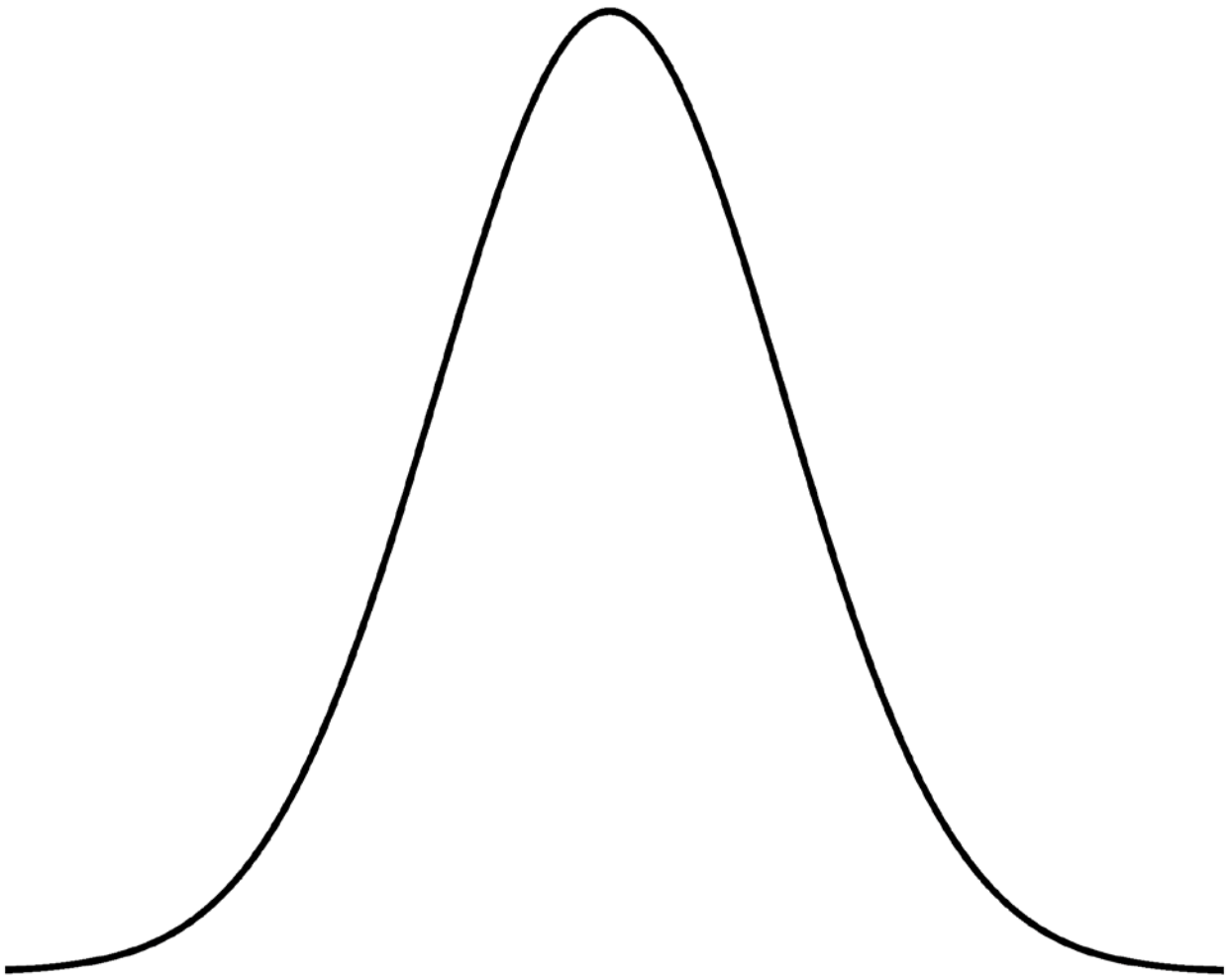


Dispersed



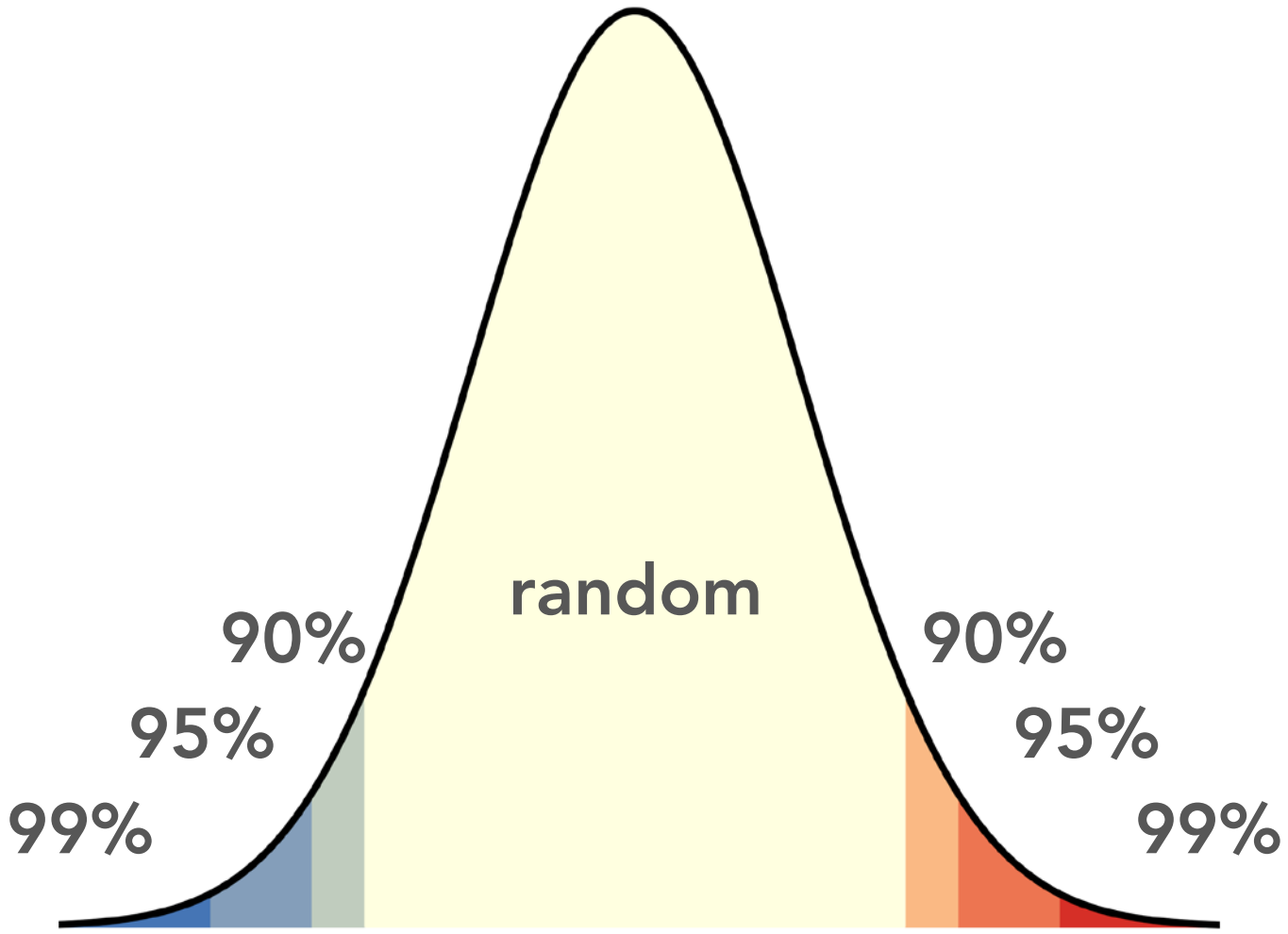
Complete Spatial RANDOMNESS

z-scores and p-values



z-scores	…-2.58	-1.96	-1.65	0	1.65	1.96	2.58…
p-value	…-0.01	-0.05	-0.10	0	0.10	0.05	0.01…

z-scores and p-values



z-scores	…-2.58	-1.96	-1.65	0	1.65	1.96	2.58…
p-values	…0.01	0.05	0.10	0	0.10	0.05	0.01…

How intense is the clustering?

How intense is the clustering?

Compared to what?

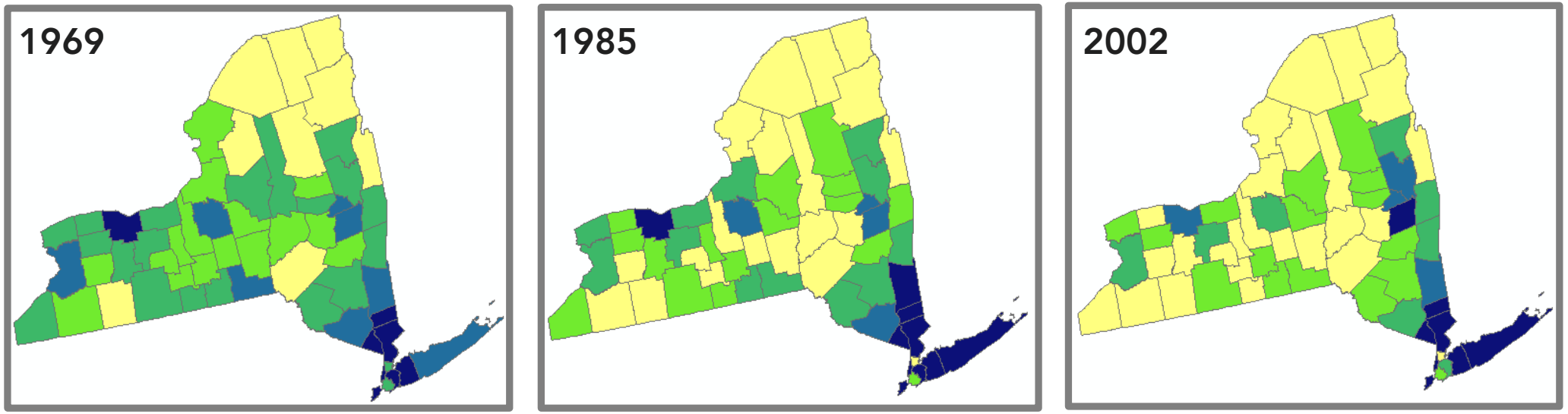
How intense is the clustering?

Compared to where?

How intense is the clustering?

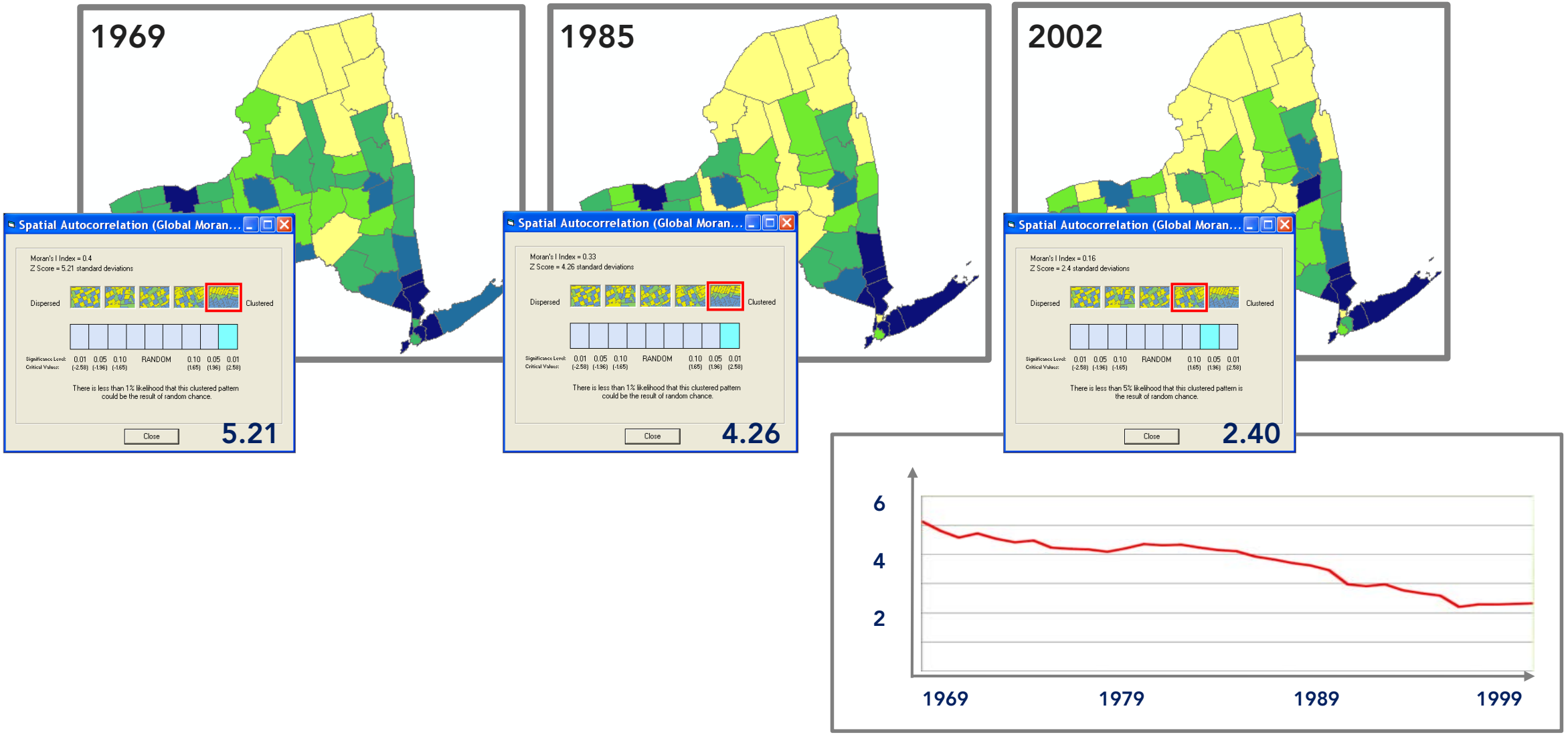
Compared to when?

An Example



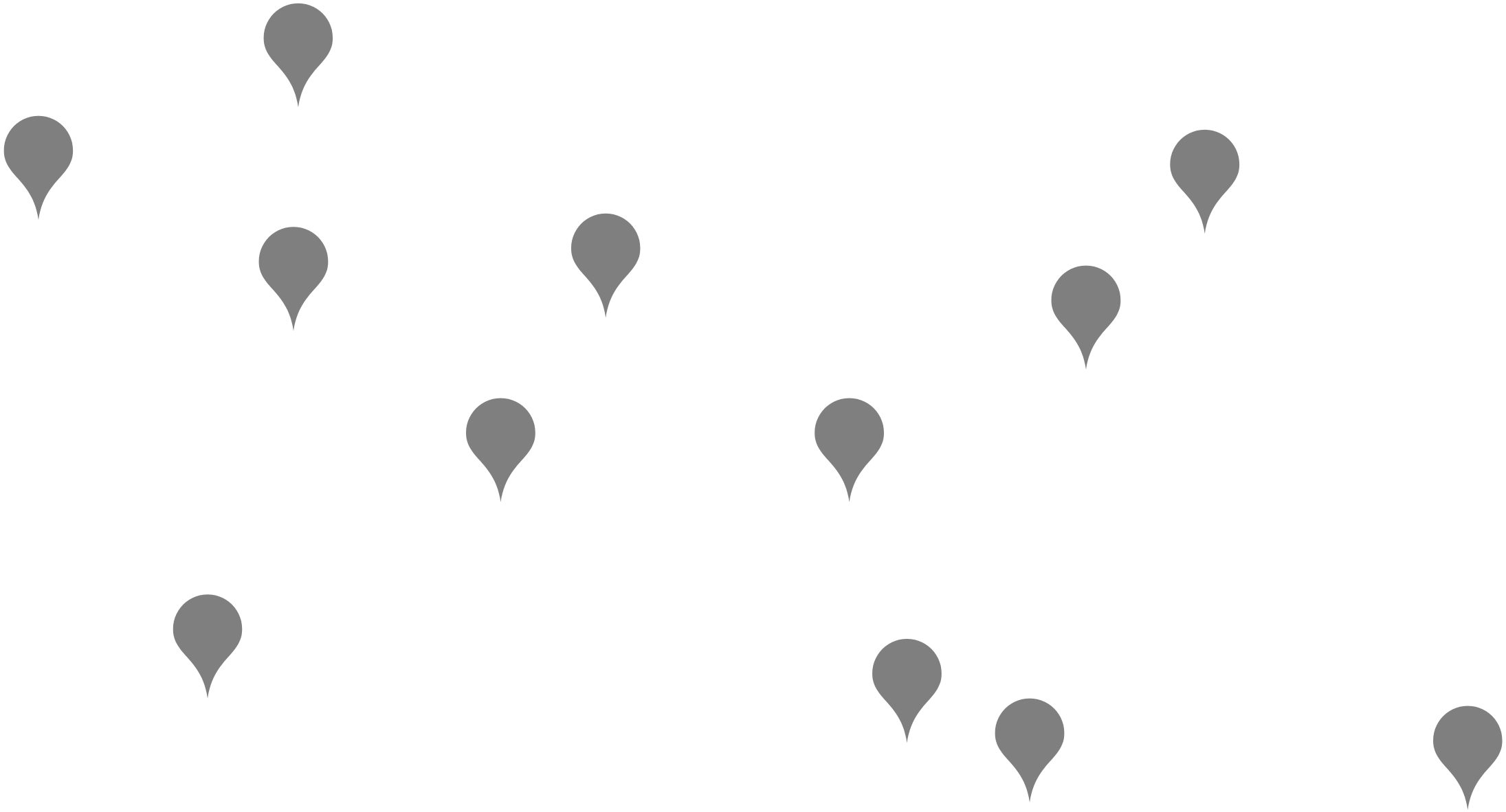
Is the spatial segregation of the rich and the poor increasing or decreasing in New York?

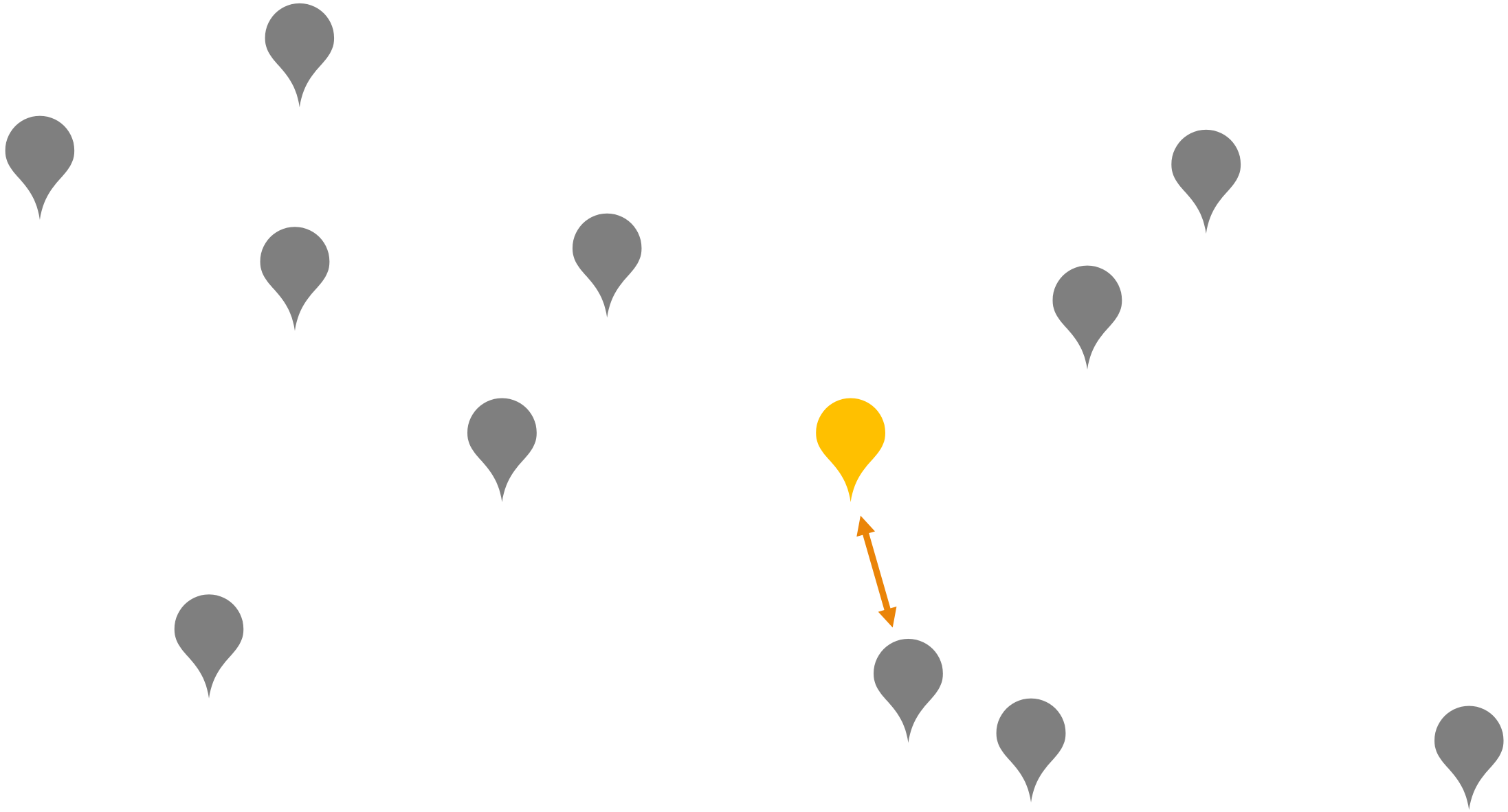
An Example

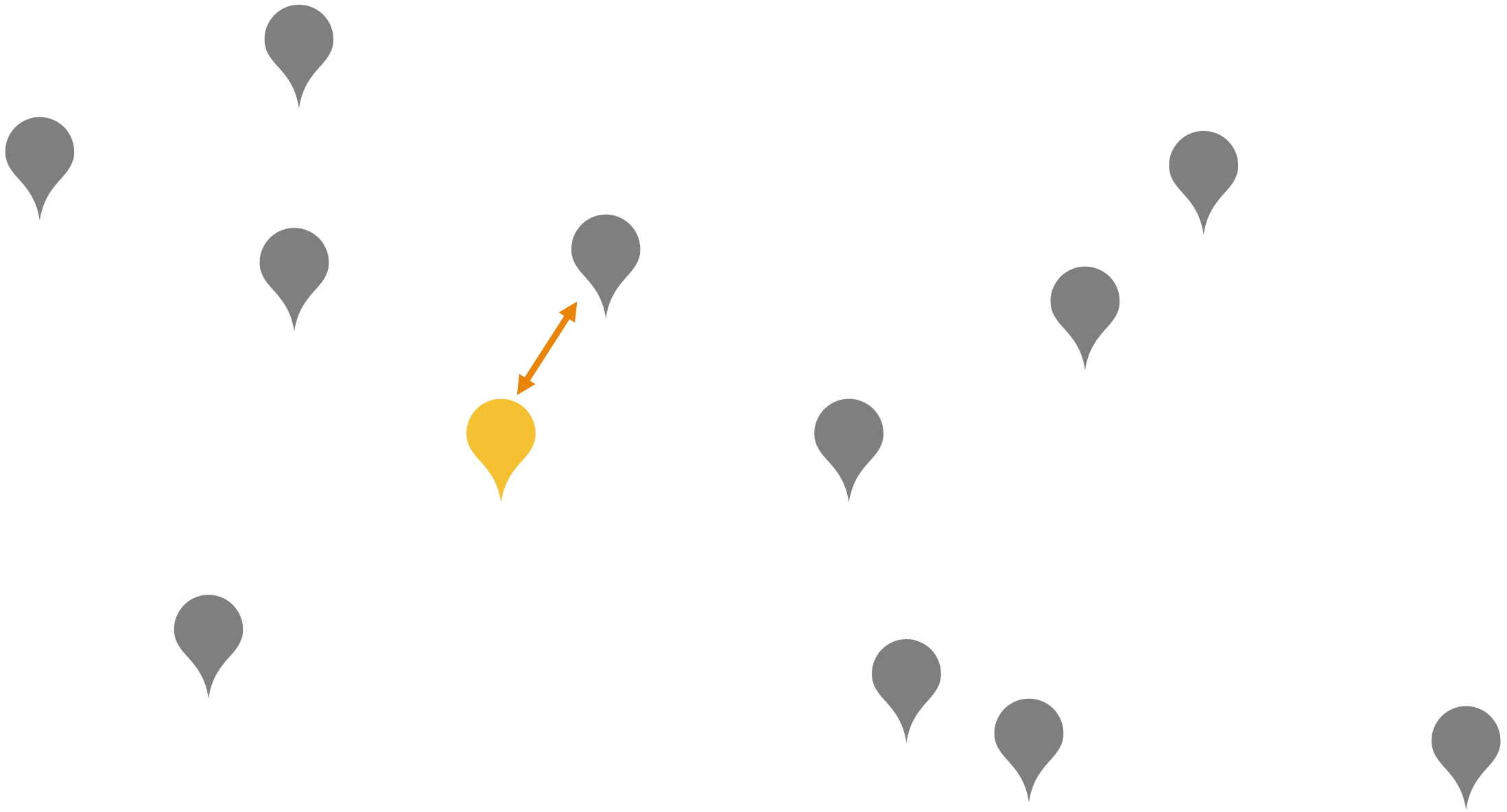

1969
Spatial Autocorrelation (Global Moran...
Moran's I Index = 0.4
Z Score = 5.21 standard deviations
Dispersed
Clustered
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 1% likelihood that this clustered pattern could be the result of random chance.
Close
5.21
1985
Spatial Autocorrelation (Global Moran...
Moran's I Index = 0.33
Z Score = 4.26 standard deviations
Dispersed
Clustered
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 1% likelihood that this clustered pattern could be the result of random chance.
Close
4.26
2002
Spatial Autocorrelation (Global Moran...
Moran's I Index = 0.16
Z Score = 2.4 standard deviations
Dispersed
Clustered
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 5% likelihood that this clustered pattern is the result of random chance.
Close
2.40
6
4
2
1969
1979
1989
1999

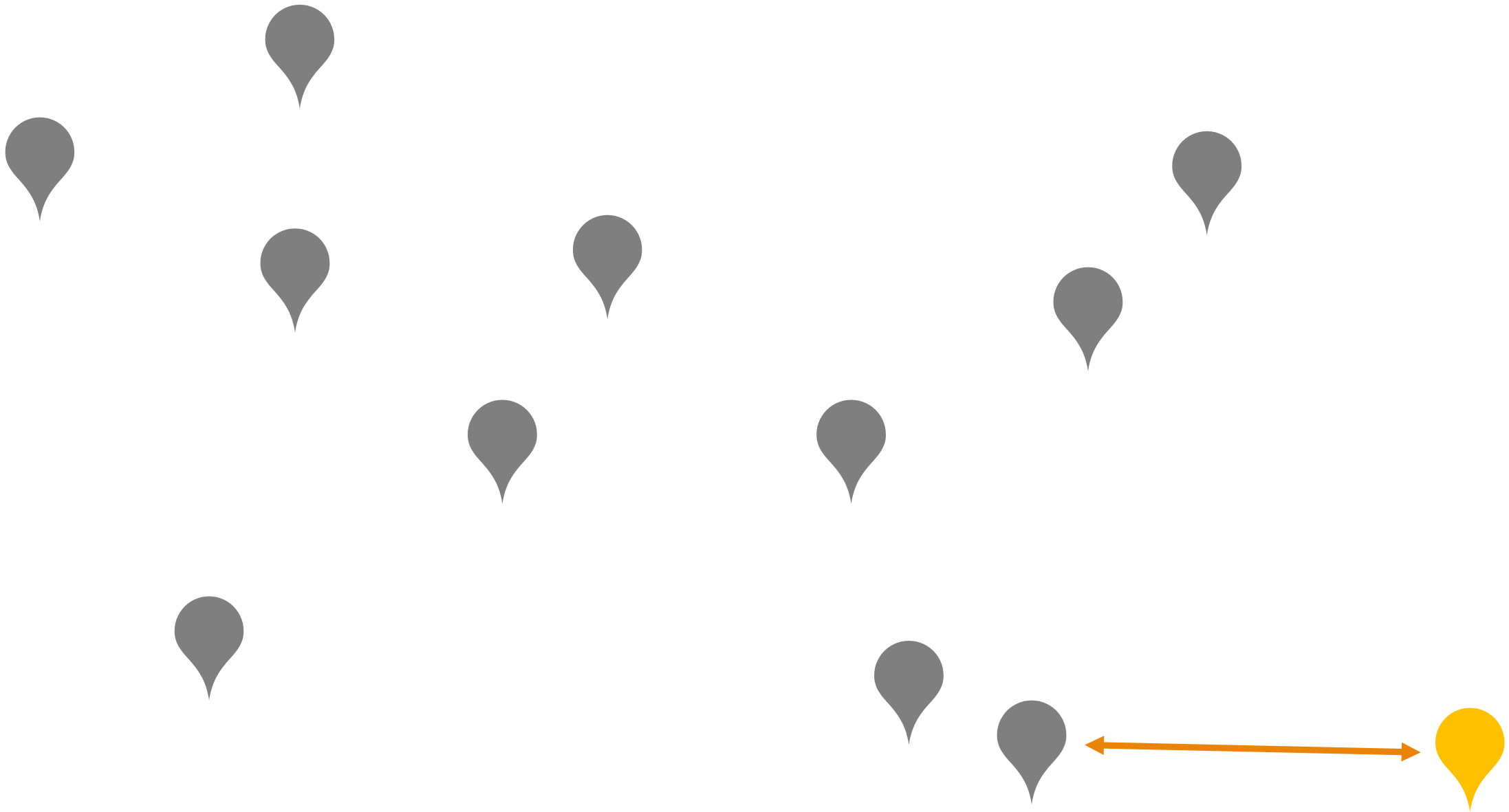
Average Nearest Neighbor

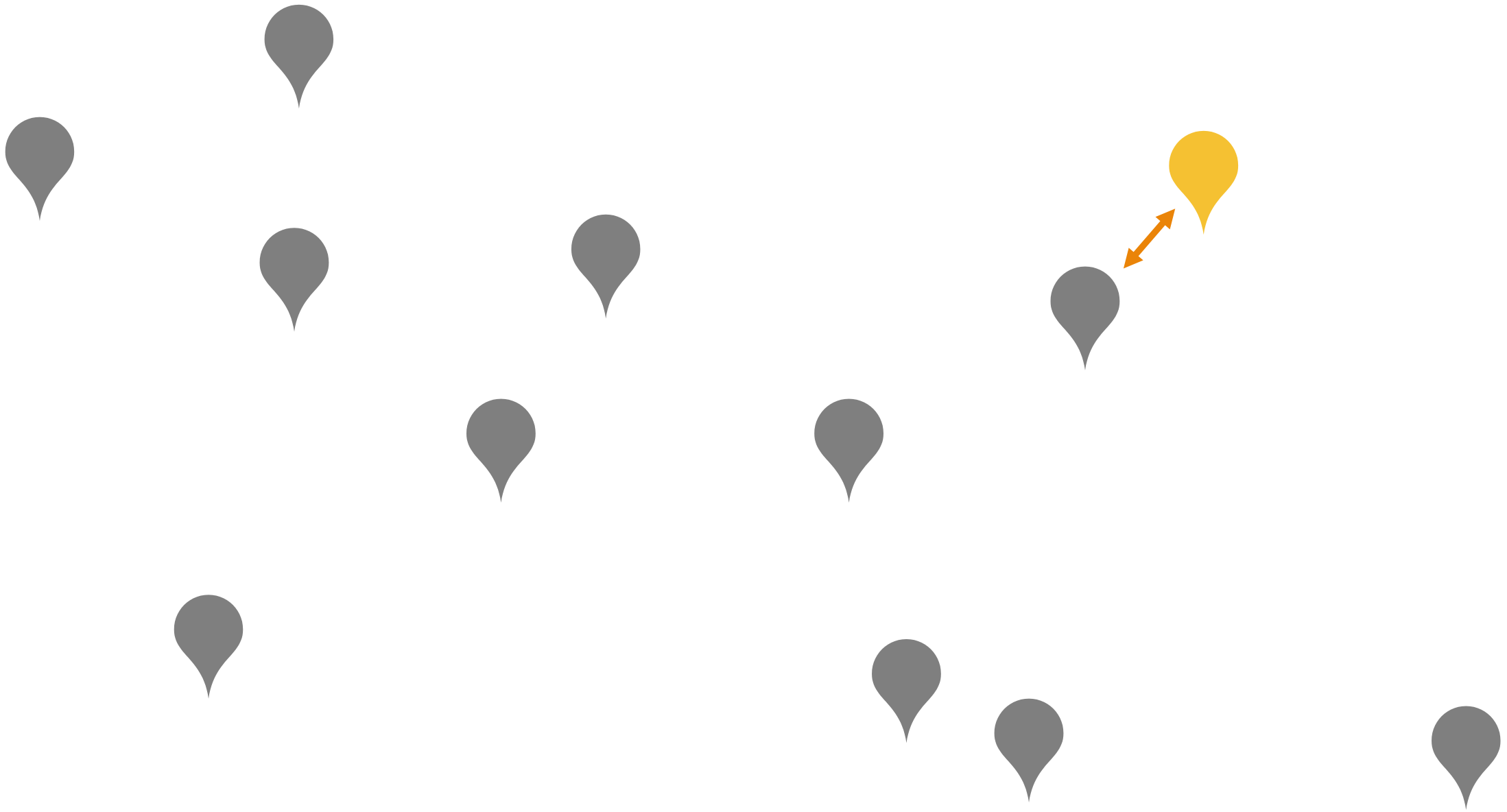
calculates a nearest neighbor index based on the average distance from each feature to its nearest neighboring feature

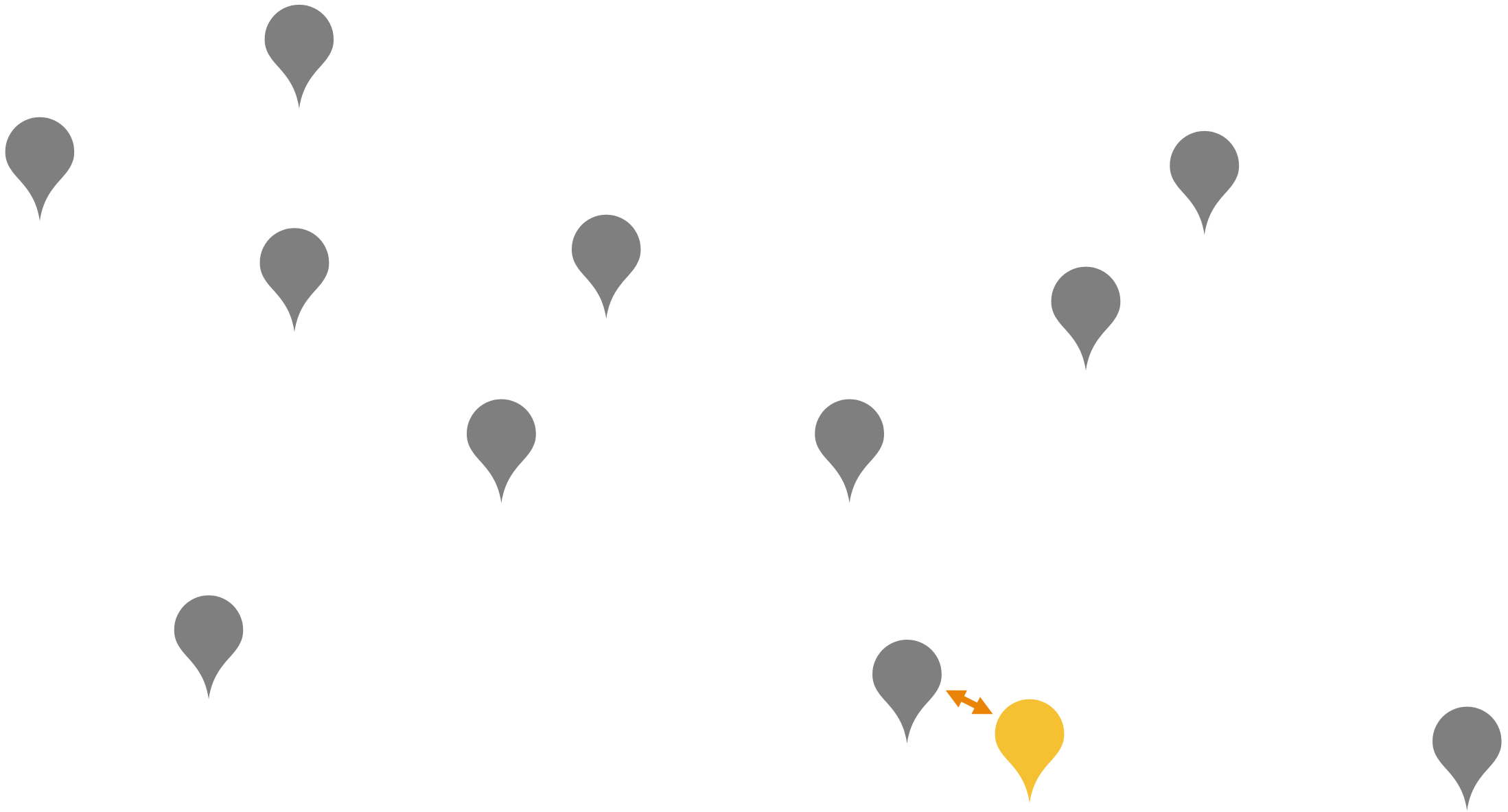


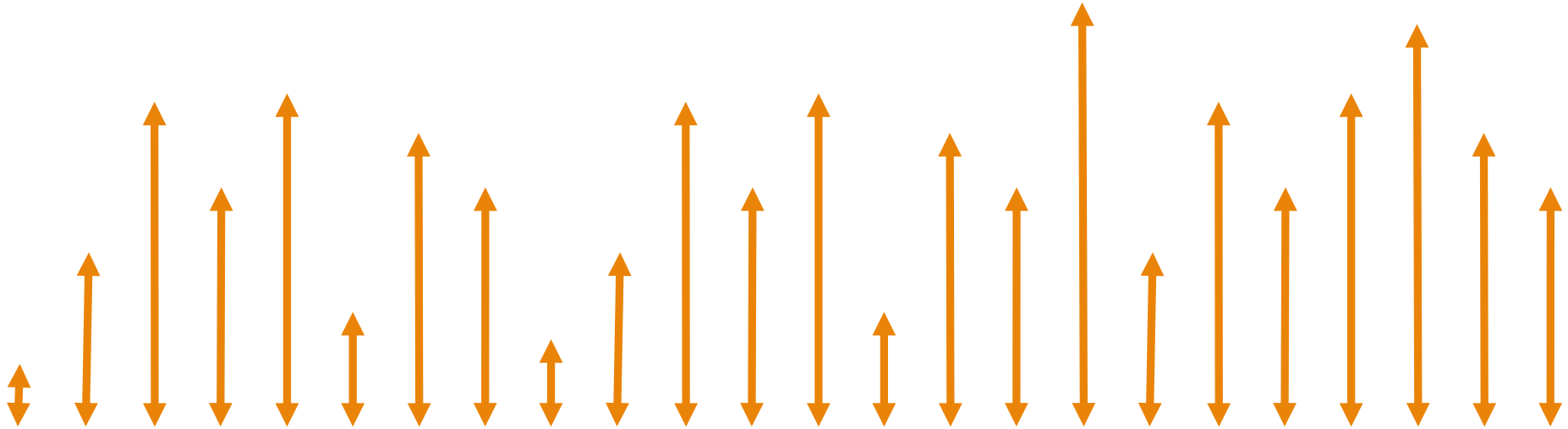








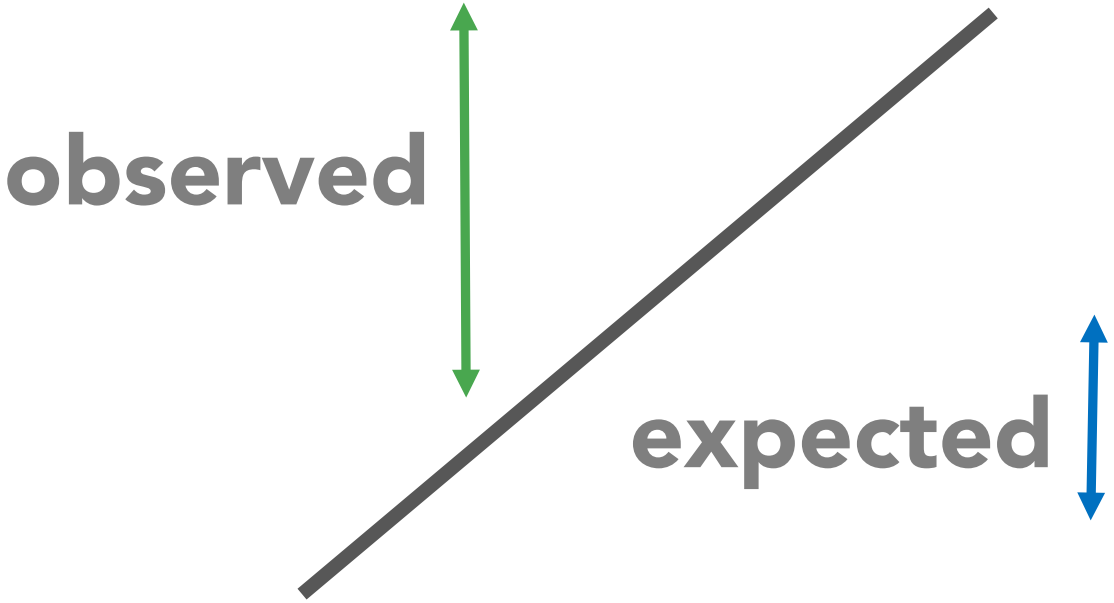




average distance =

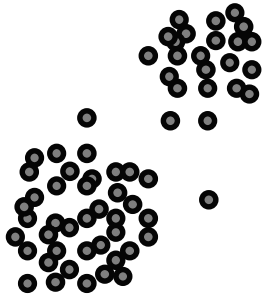
expected average distance =

ANN ratio =



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

Clustered

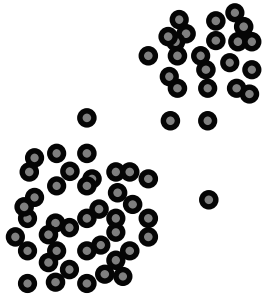


$$< 1$$

ANN ratio =

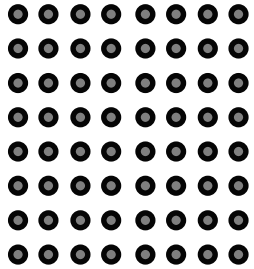
$$\frac{\text{observed}}{\text{expected}}$$

Clustered



$$< 1 <$$

Dispersed

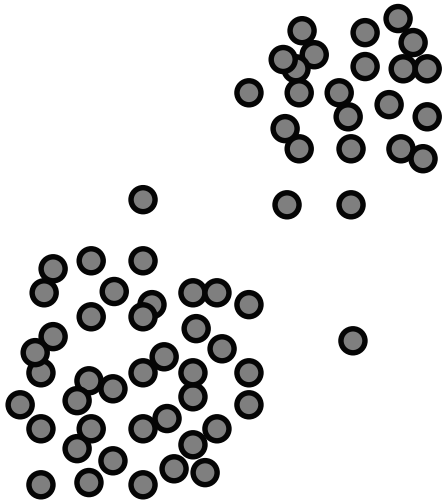


Spatial Autocorrelation
(Moran's I)

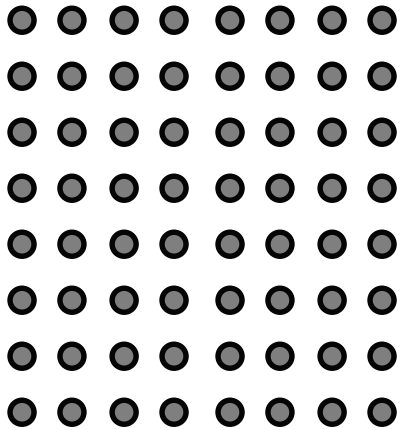
measures spatial autocorrelation based on feature locations and attribute values using the Global Moran's I statistic

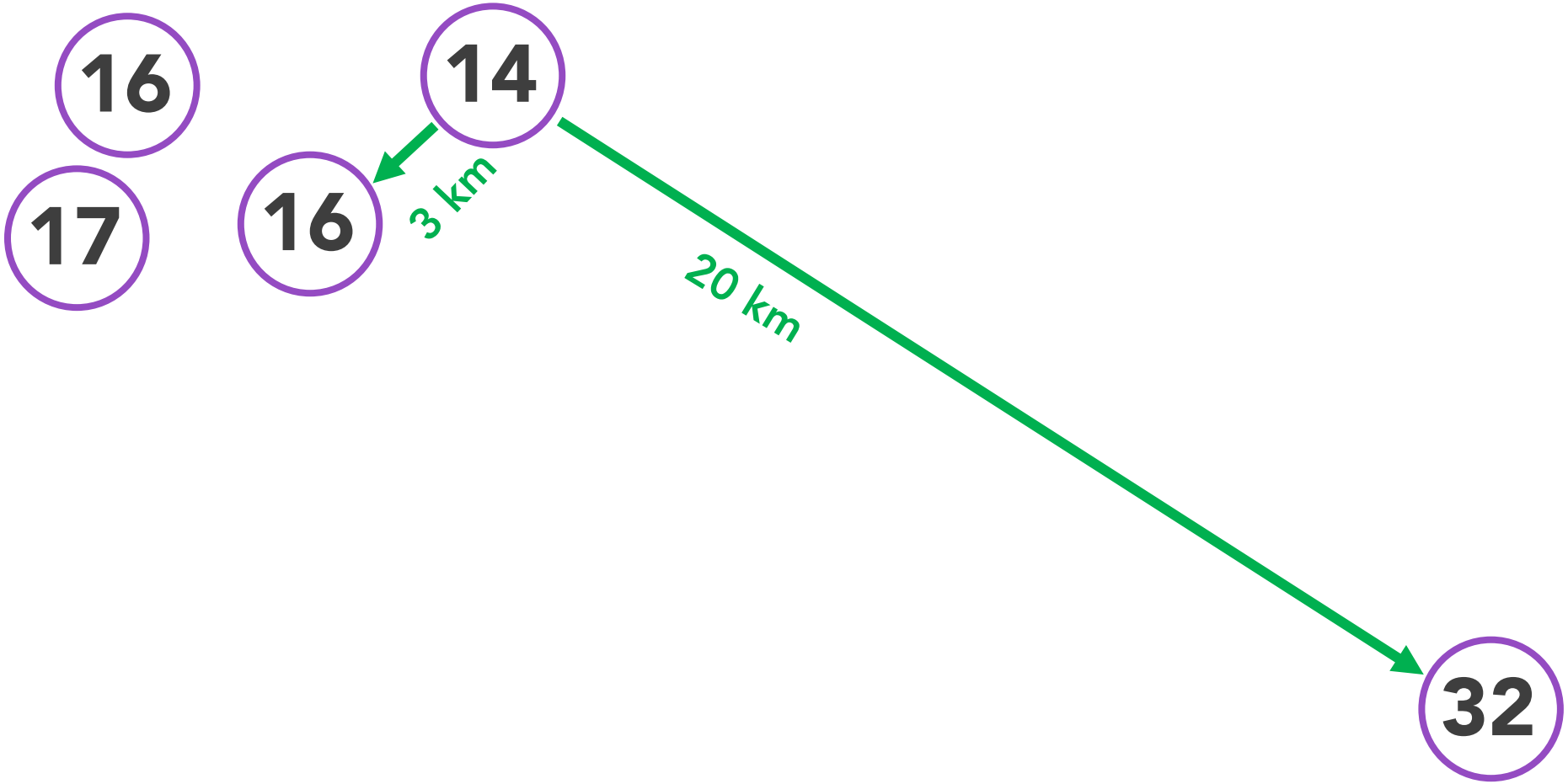
Are distances and values correlated?

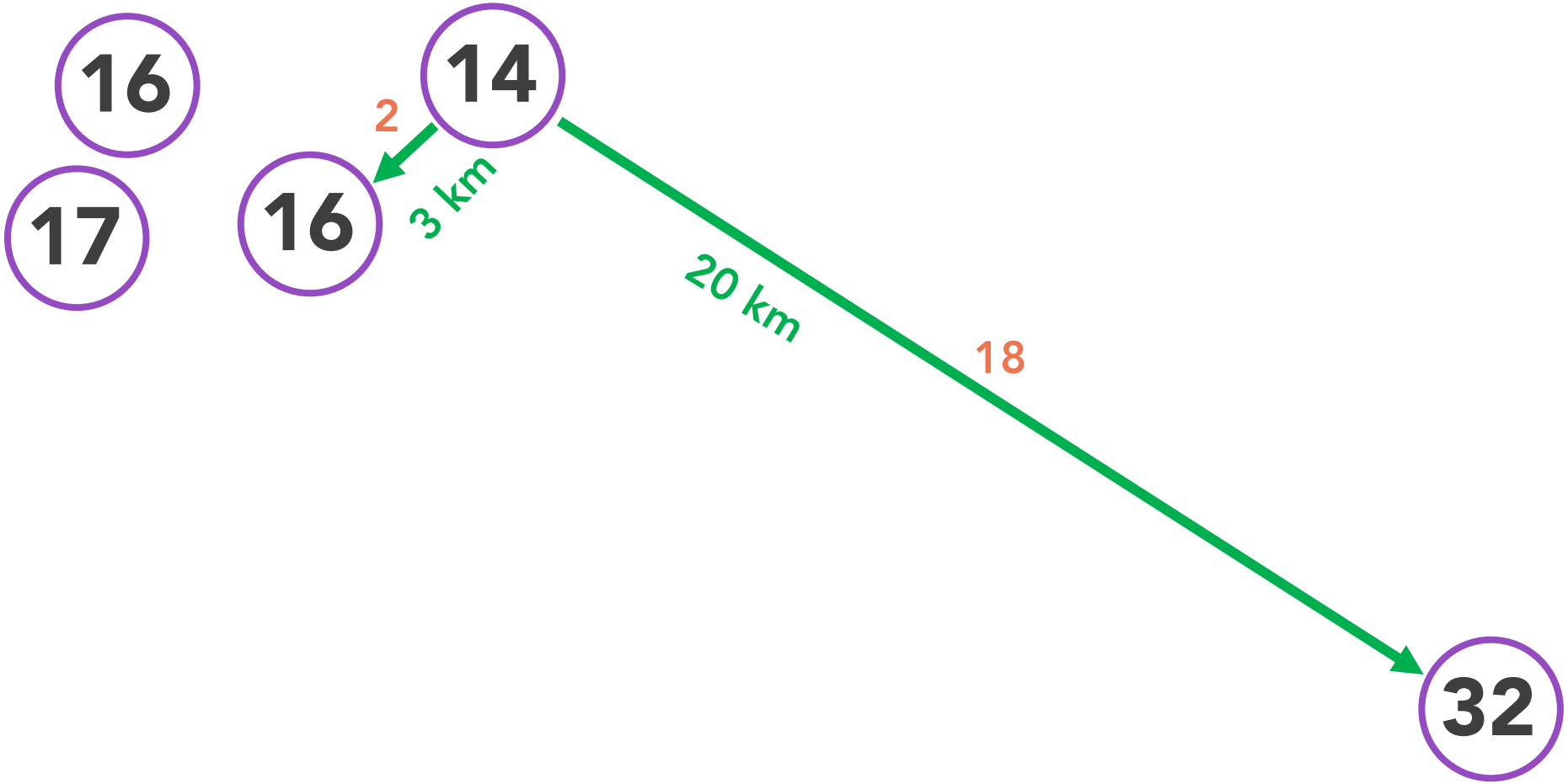
Clustered



Dispersed



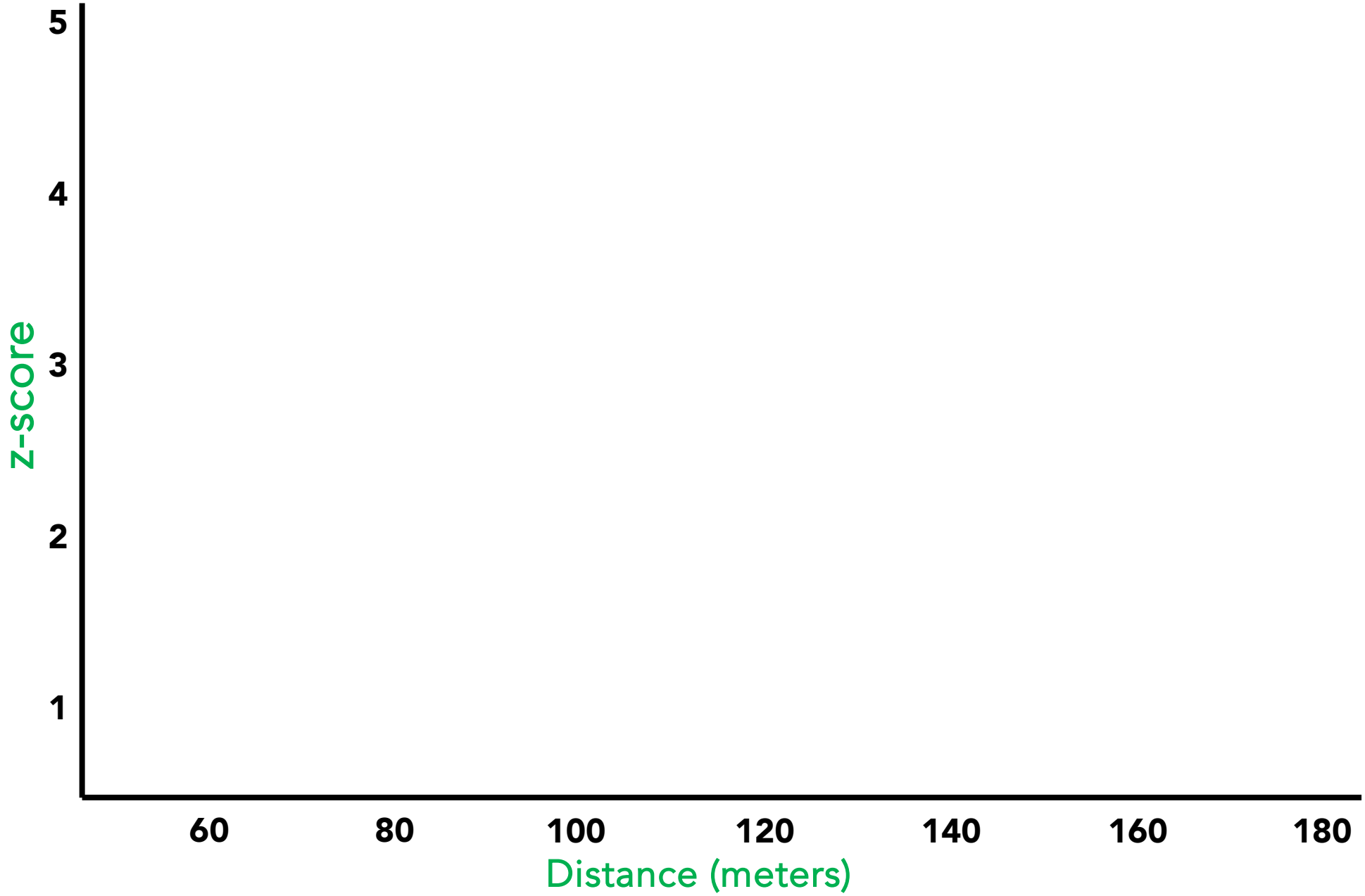

16
14
17
16
3 km
20 km
32


16
17
14
16
2
3 km
20 km
18
32

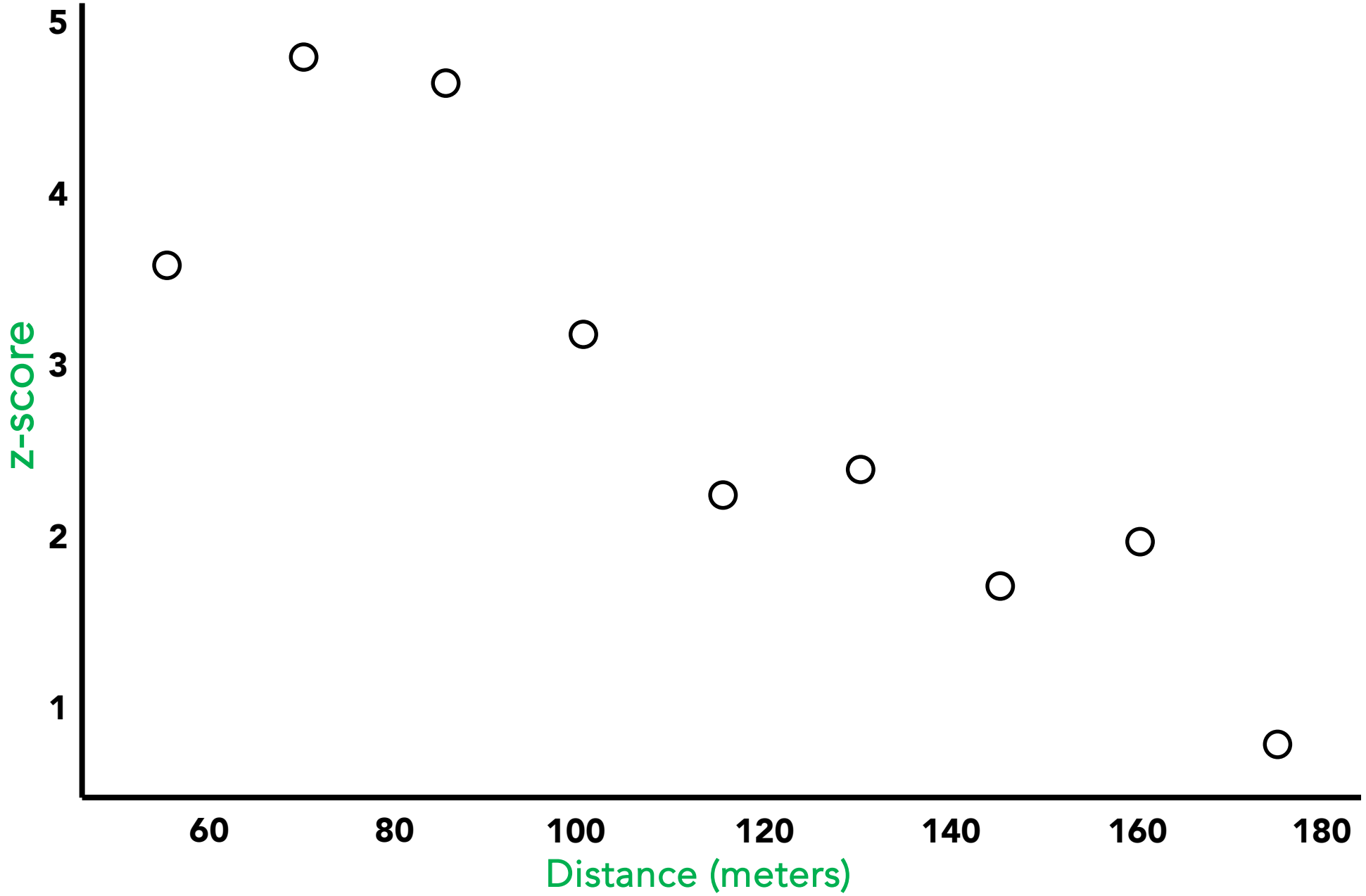
Incremental Spatial Autocorrelation

measures spatial autocorrelation for a series of distances and optionally creates a line graph of those distances and their corresponding z-scores

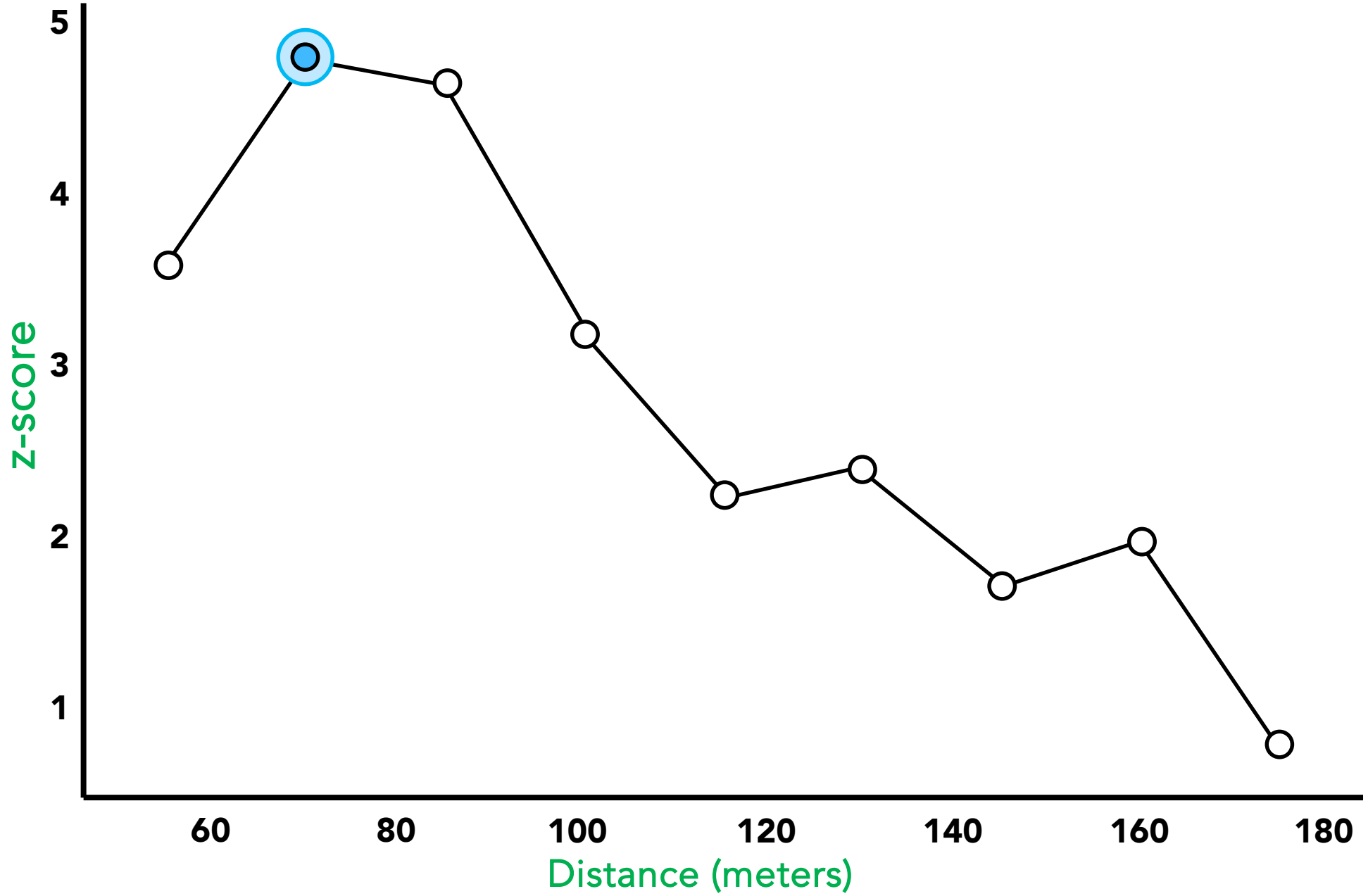
Spatial Autocorrelation by Distance



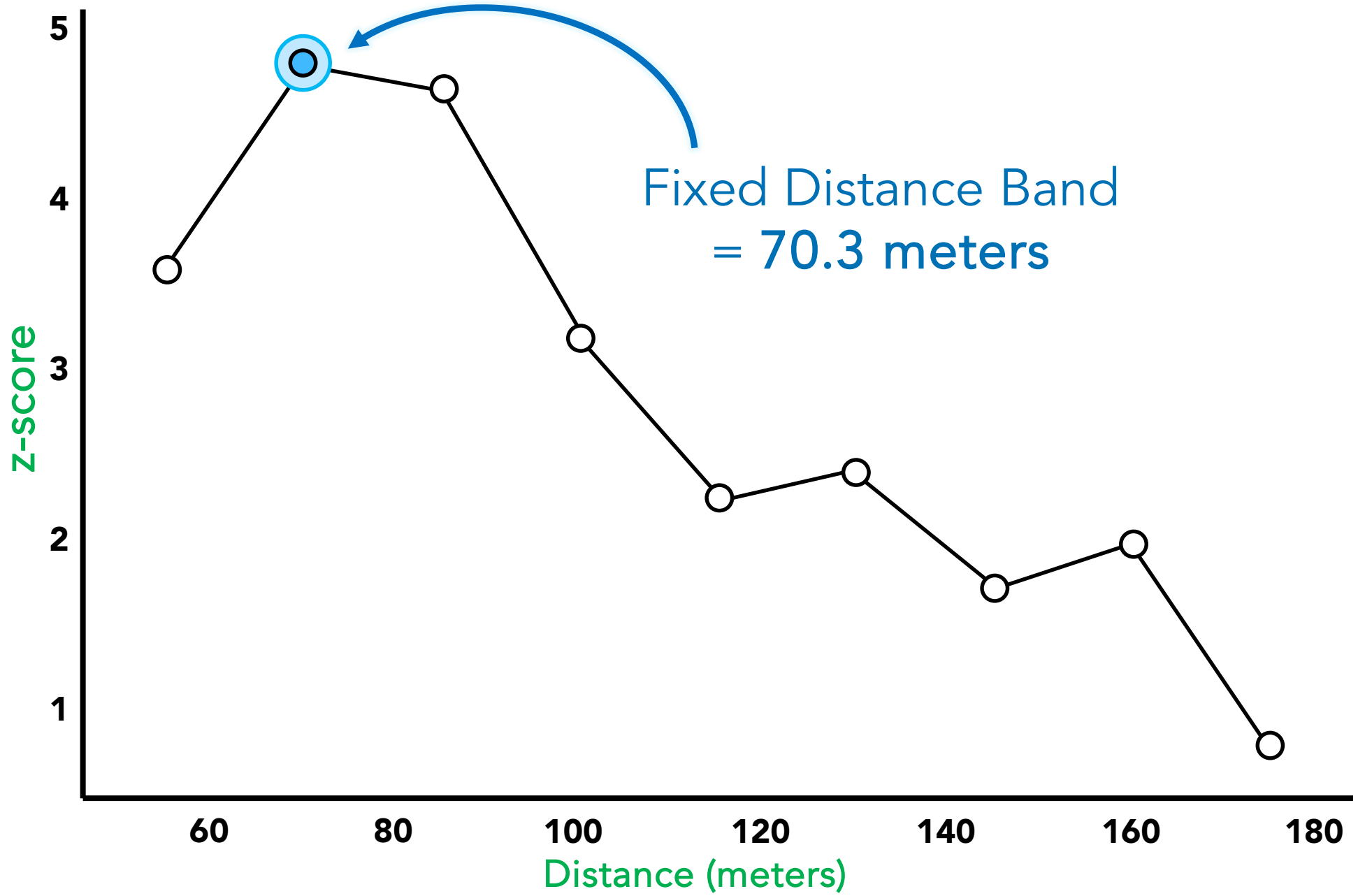
Spatial Autocorrelation by Distance



Spatial Autocorrelation by Distance



Spatial Autocorrelation by Distance

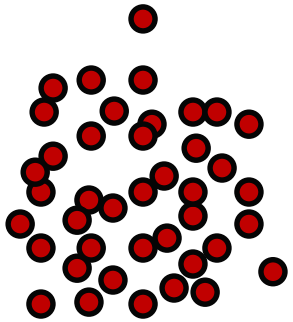


High/Low Clustering
(Getis-Ord General G)

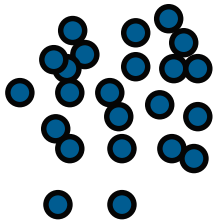
measures the concentration of high or low values for a given study area

What type of clustering is present in the data?

High value clusters



Low value clusters

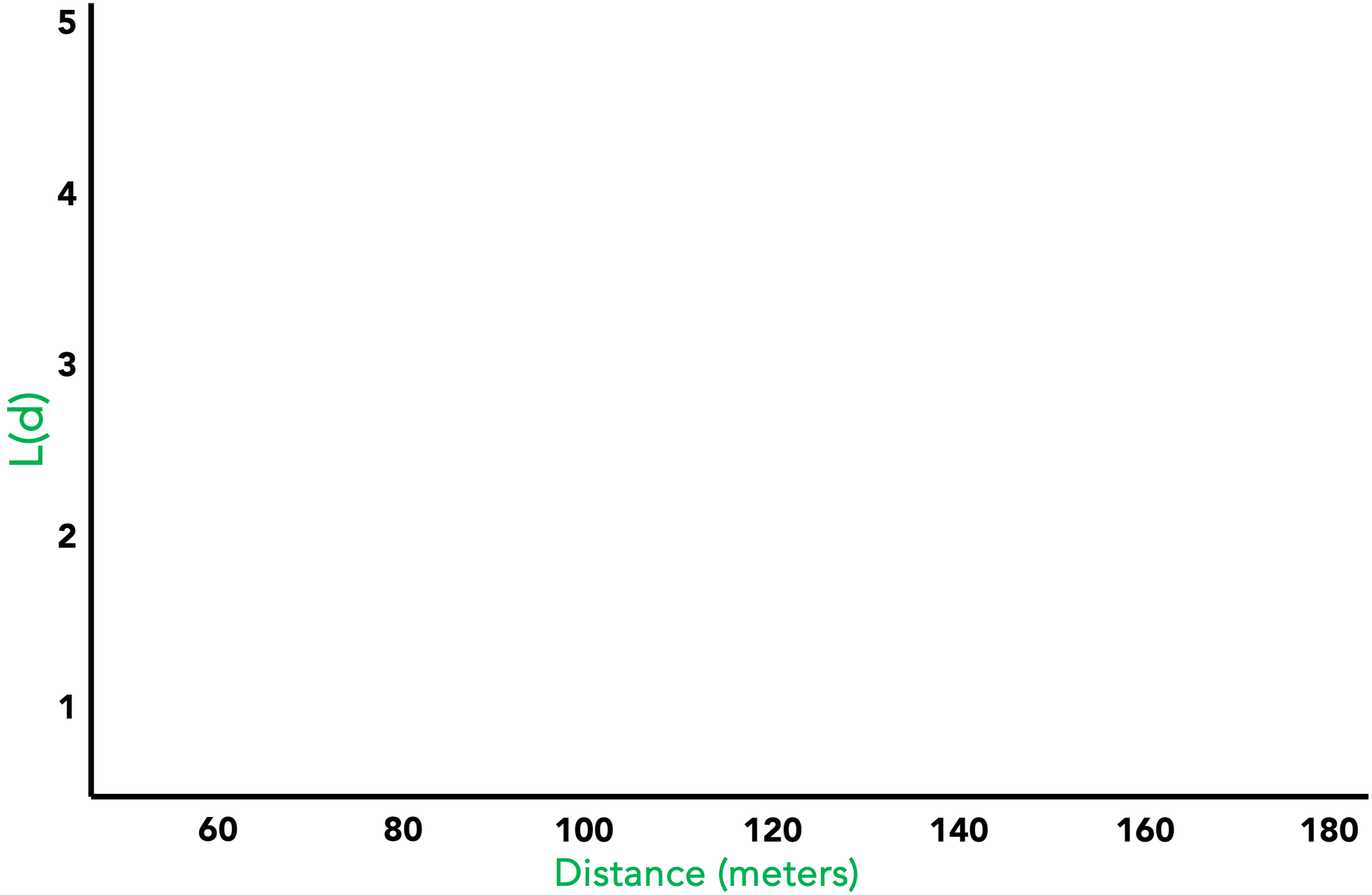


Multi-Distance Spatial Cluster Analysis

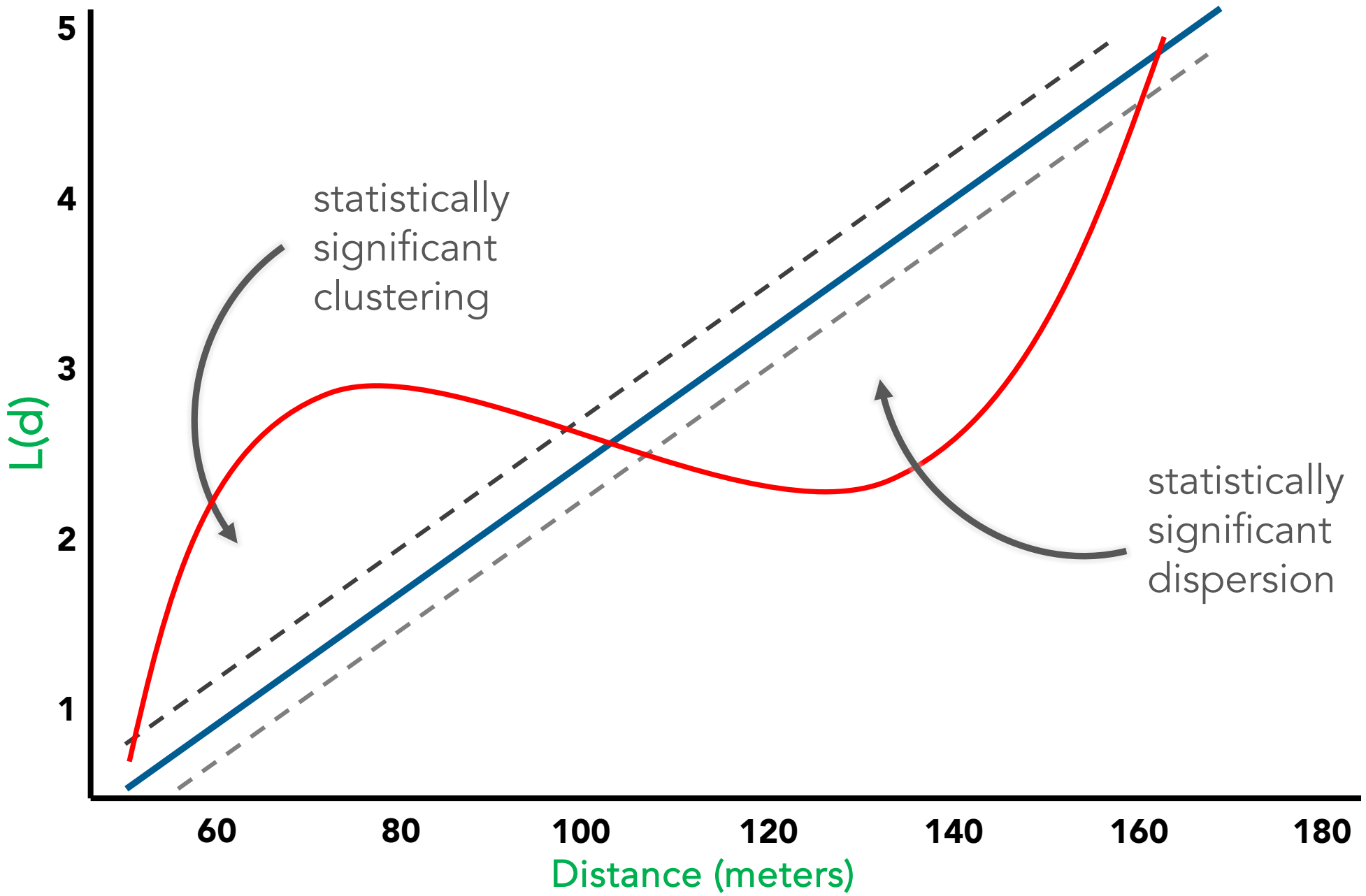
(Ripleys K Function)

determines whether features, or the values associated with features, exhibit statistically significant clustering or dispersion over a range of distances

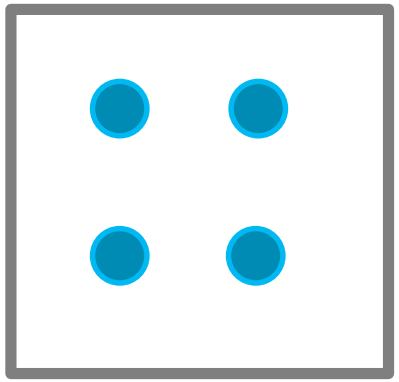
Spatial Clustering by Distance



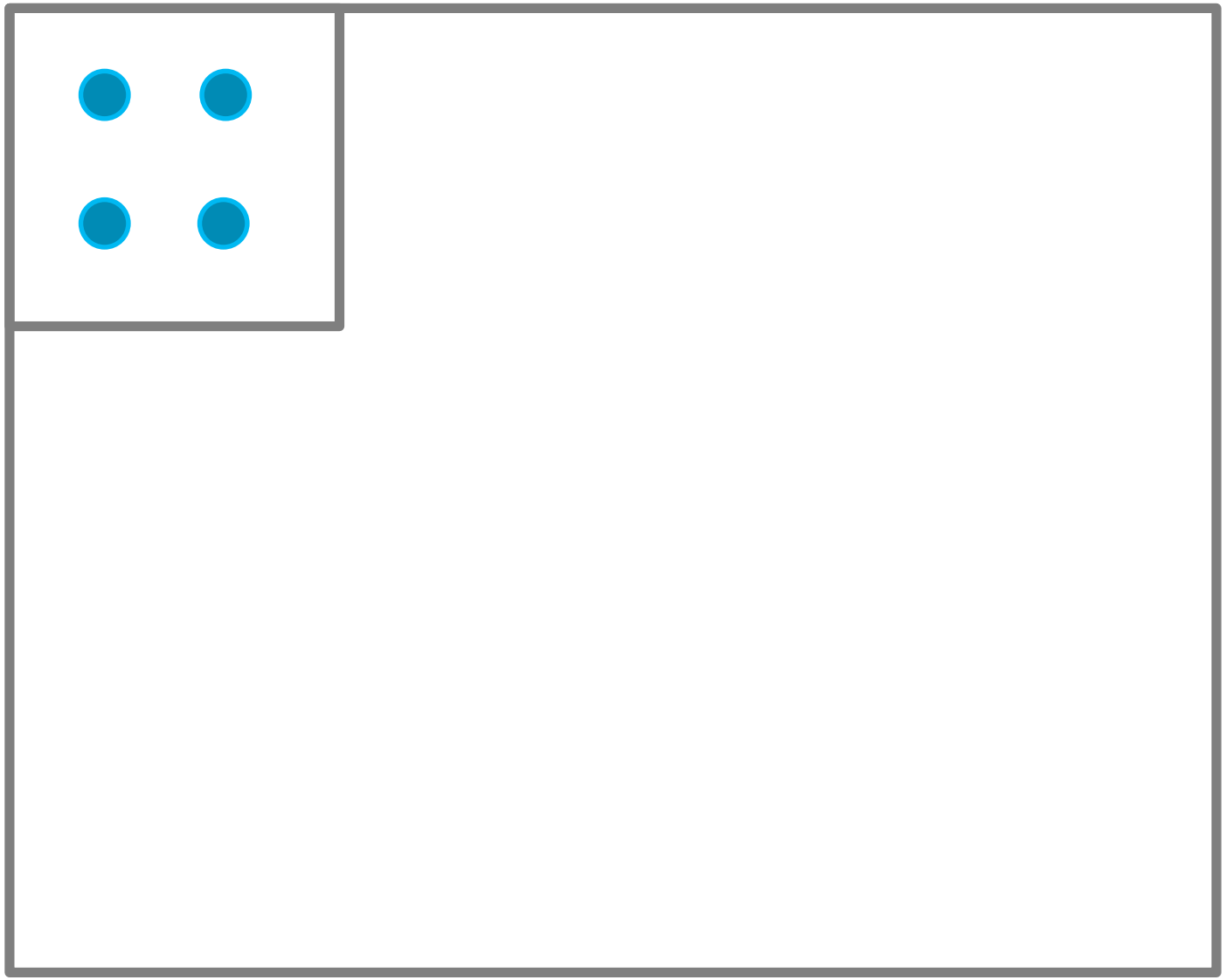
Spatial Clustering by Distance



dispersed

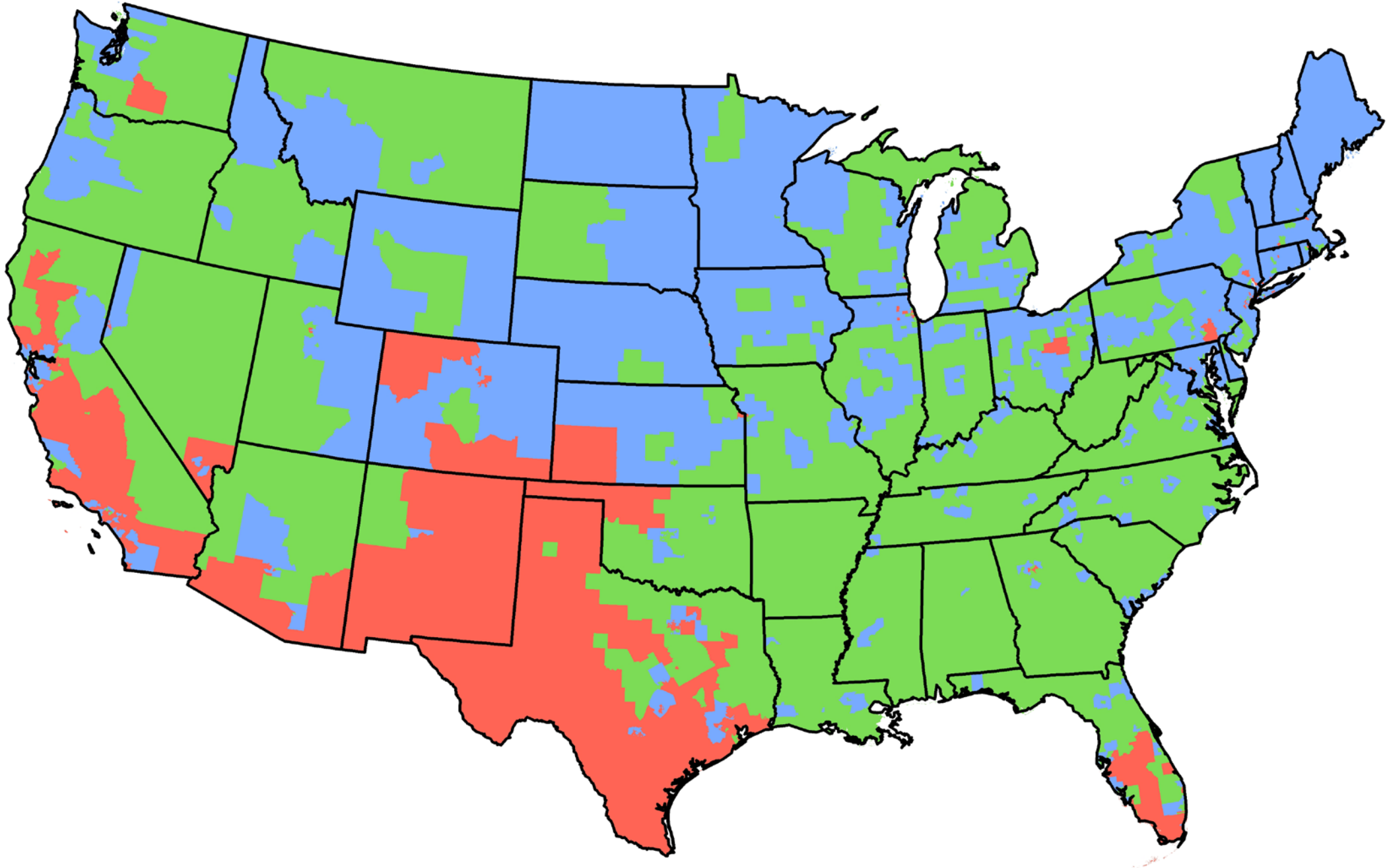


clustered

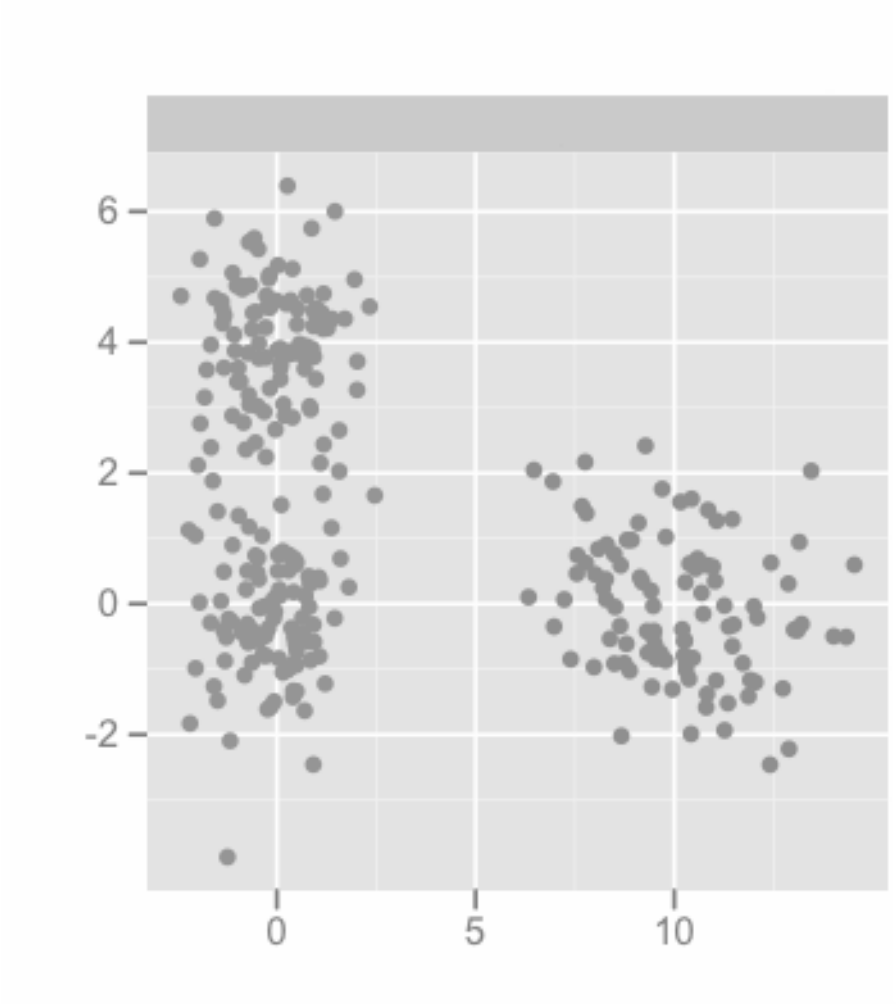


Grouping Analysis

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

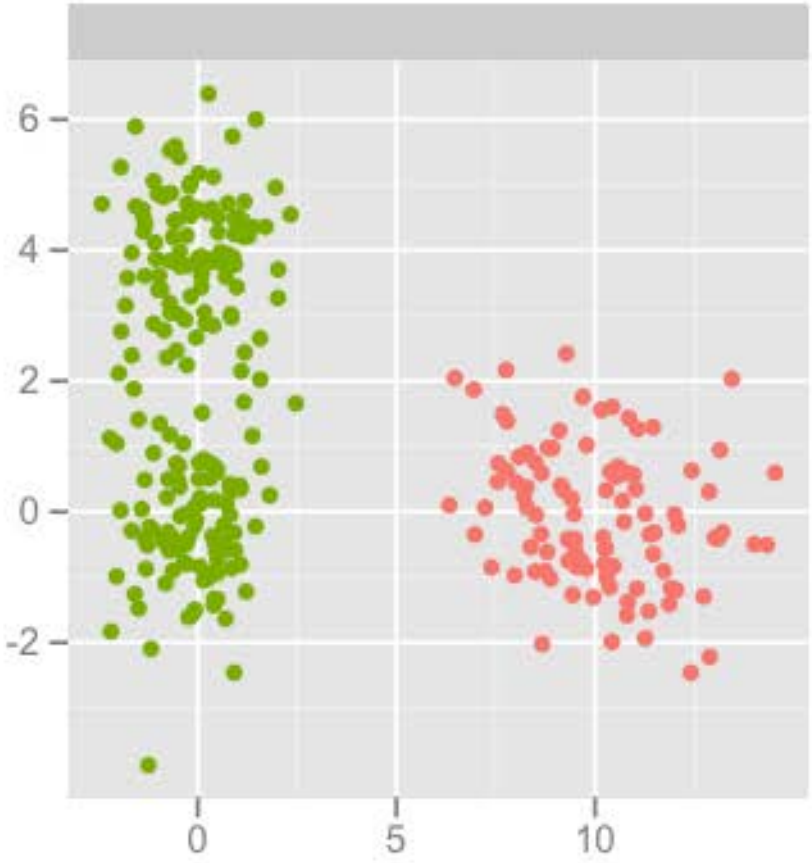


K Means

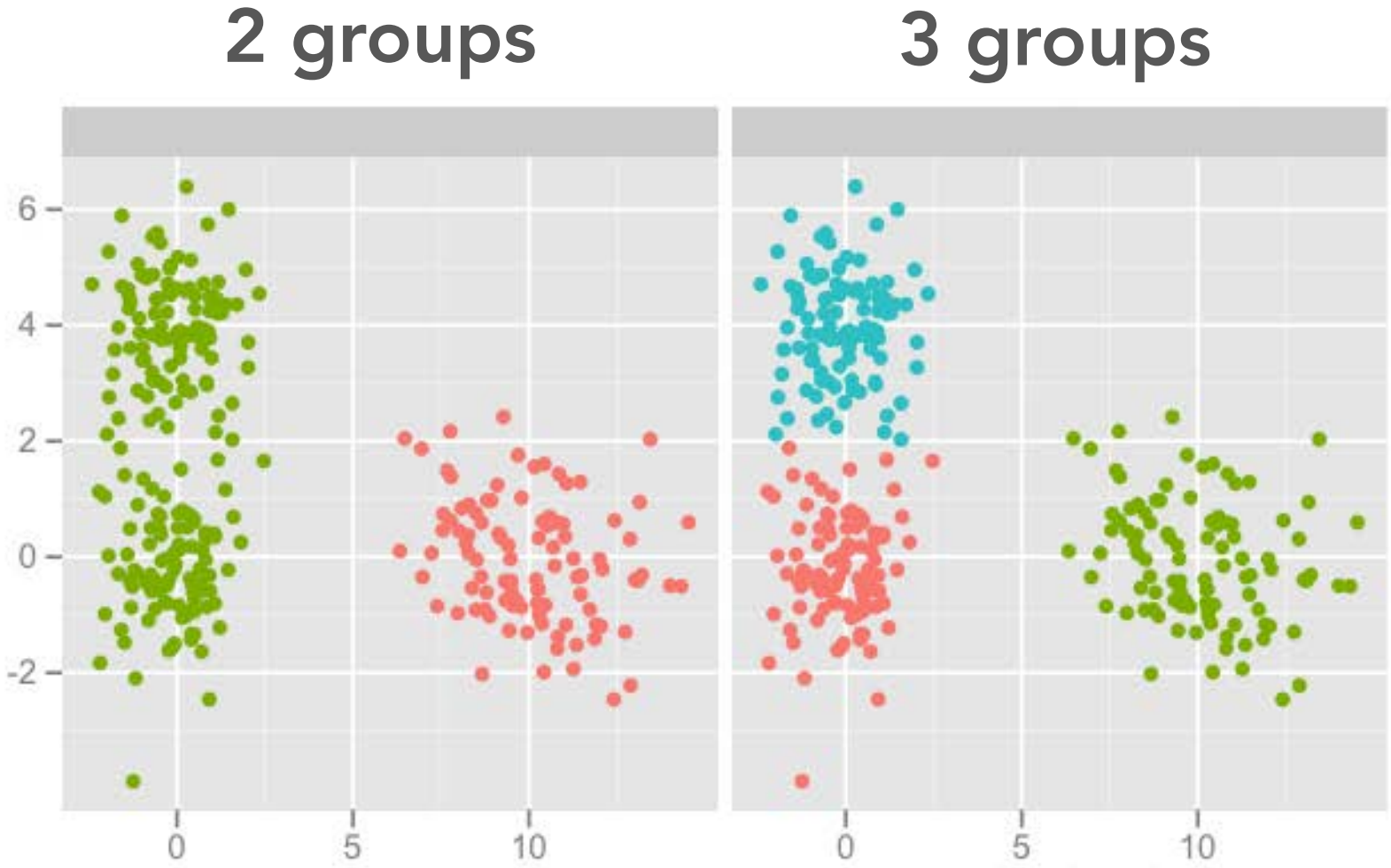


K Means

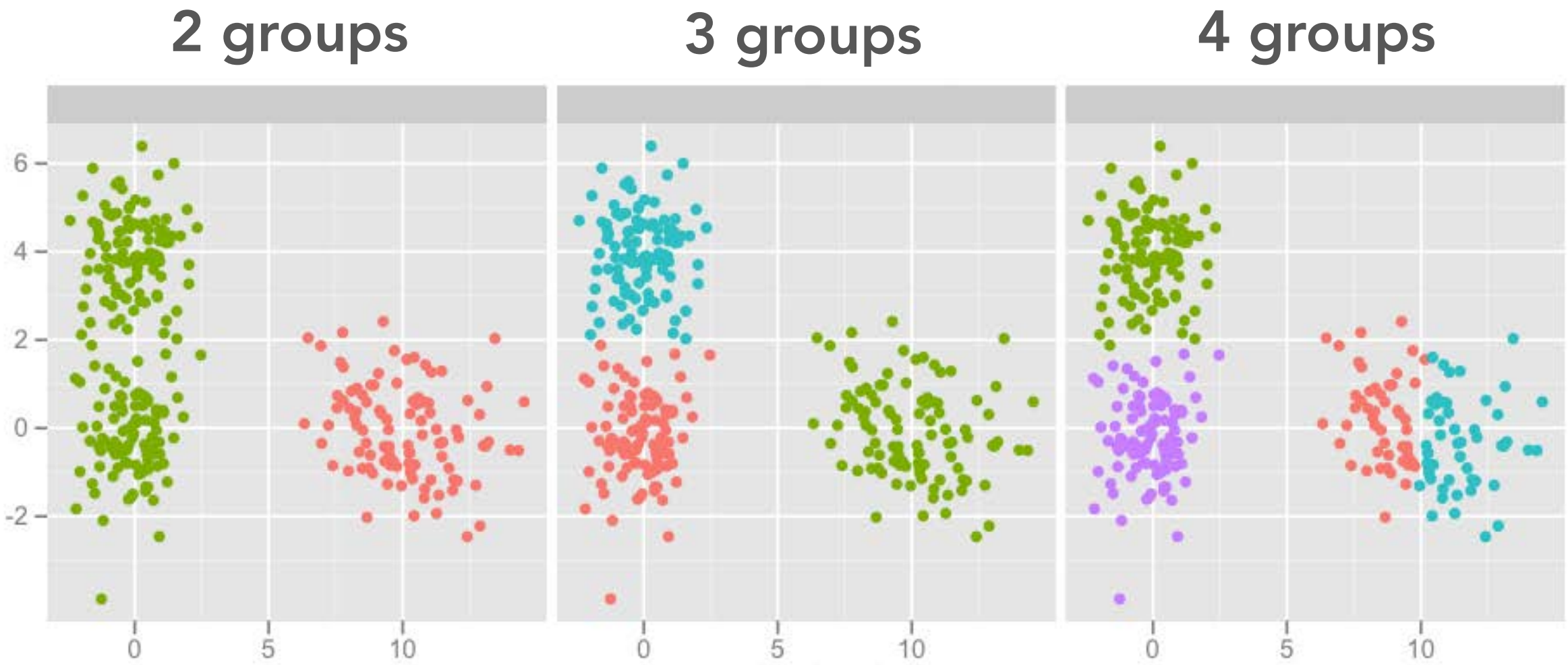
2 groups



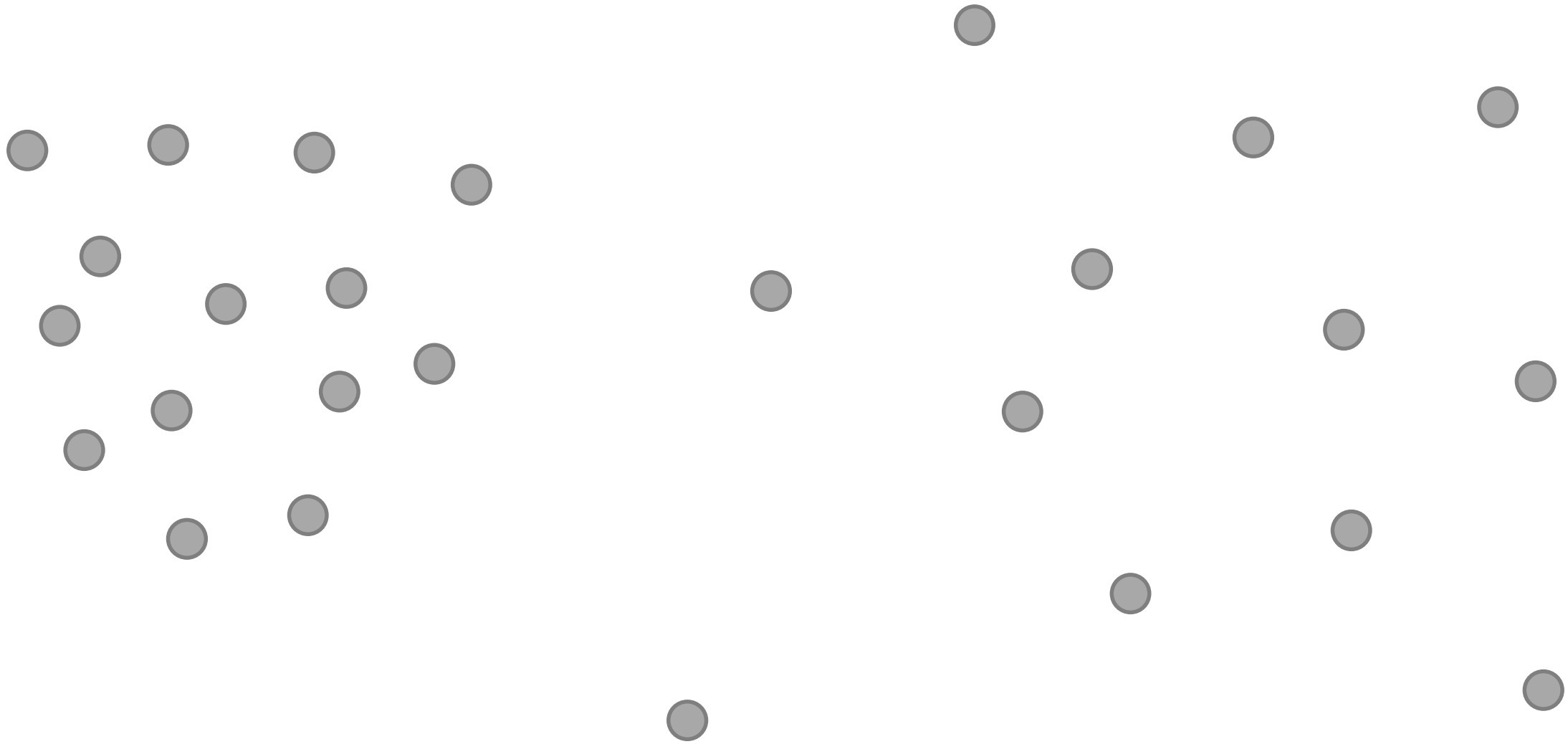
K Means



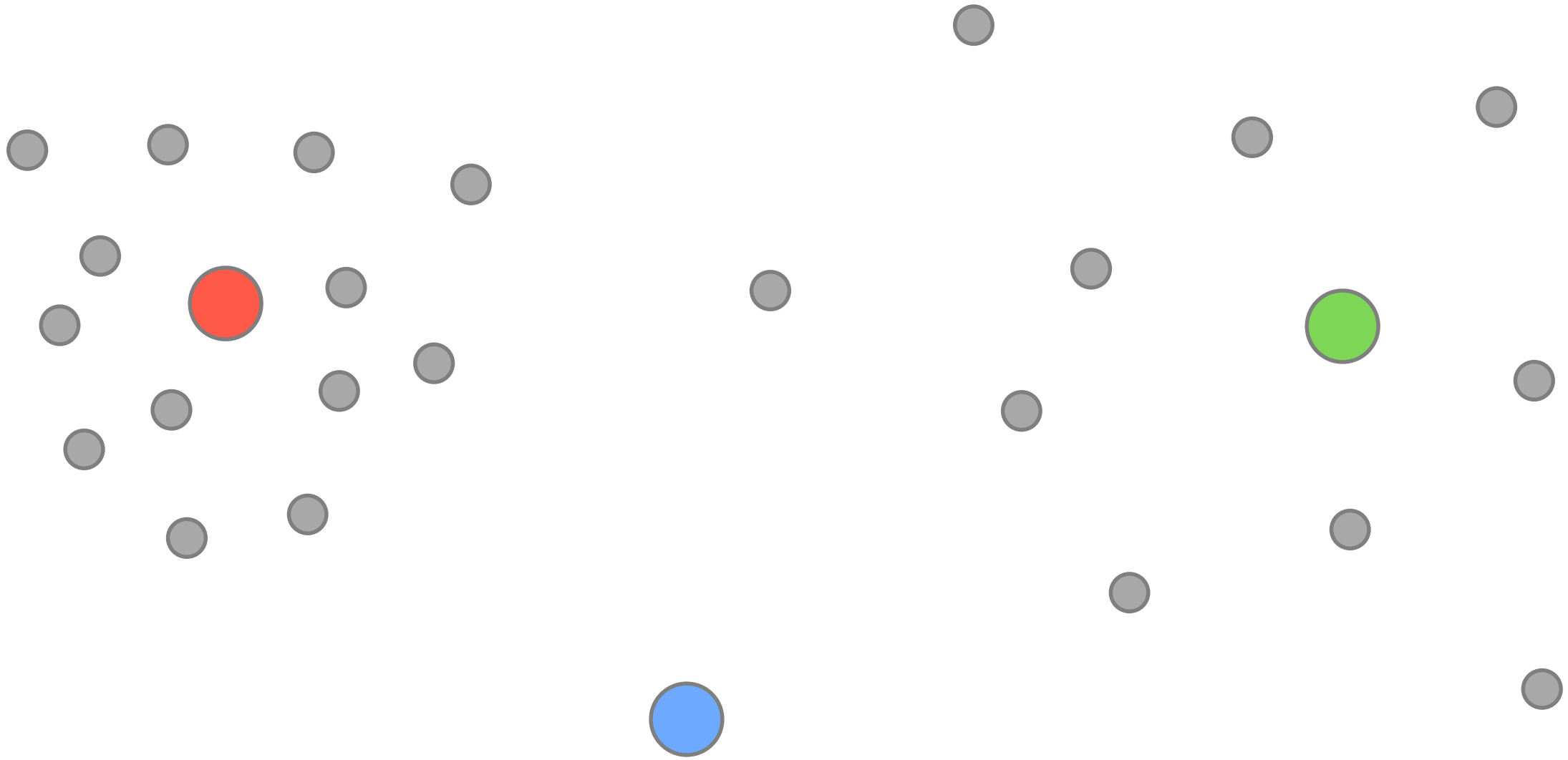
K Means



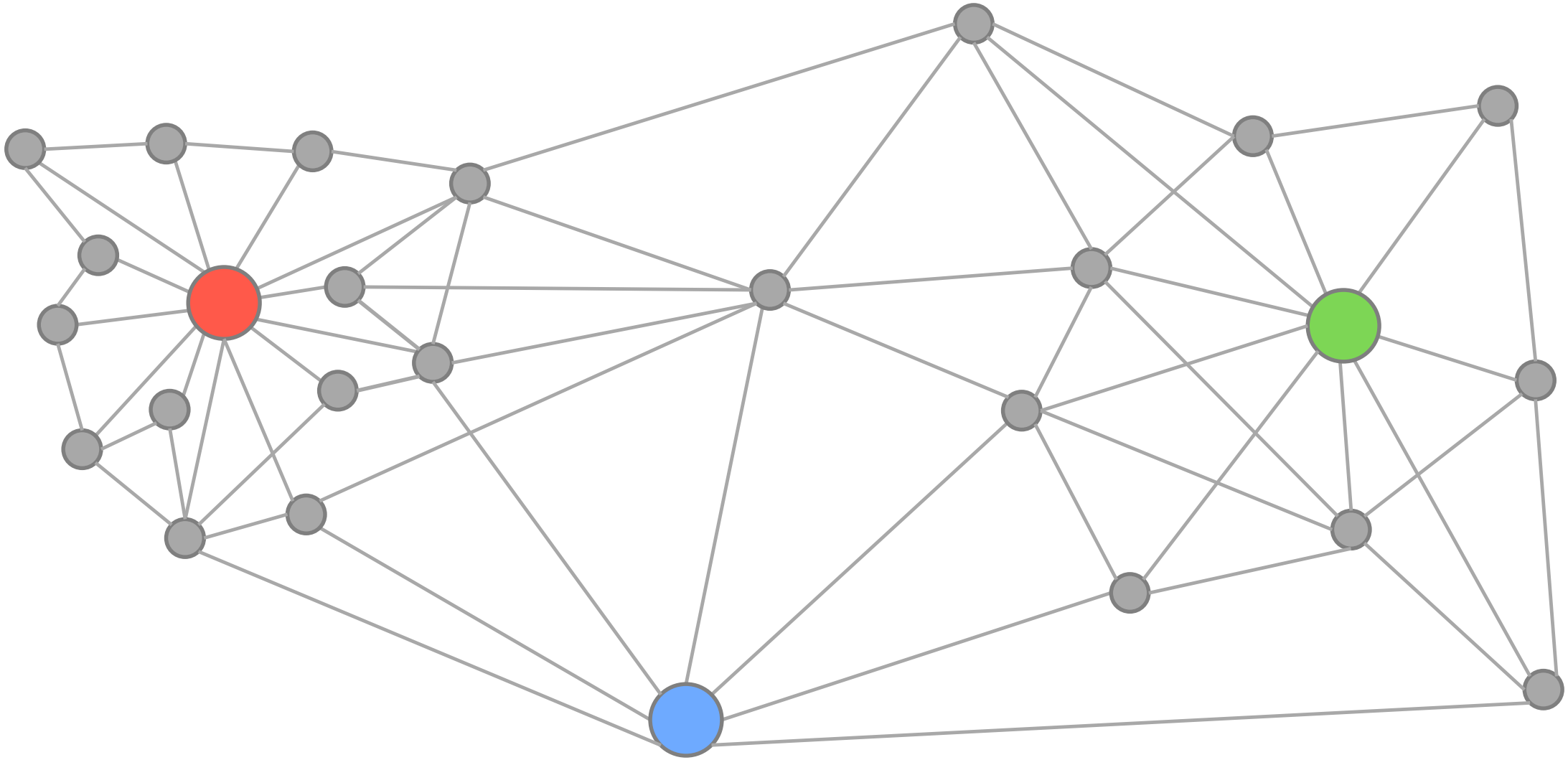
Minimum Spanning Tree



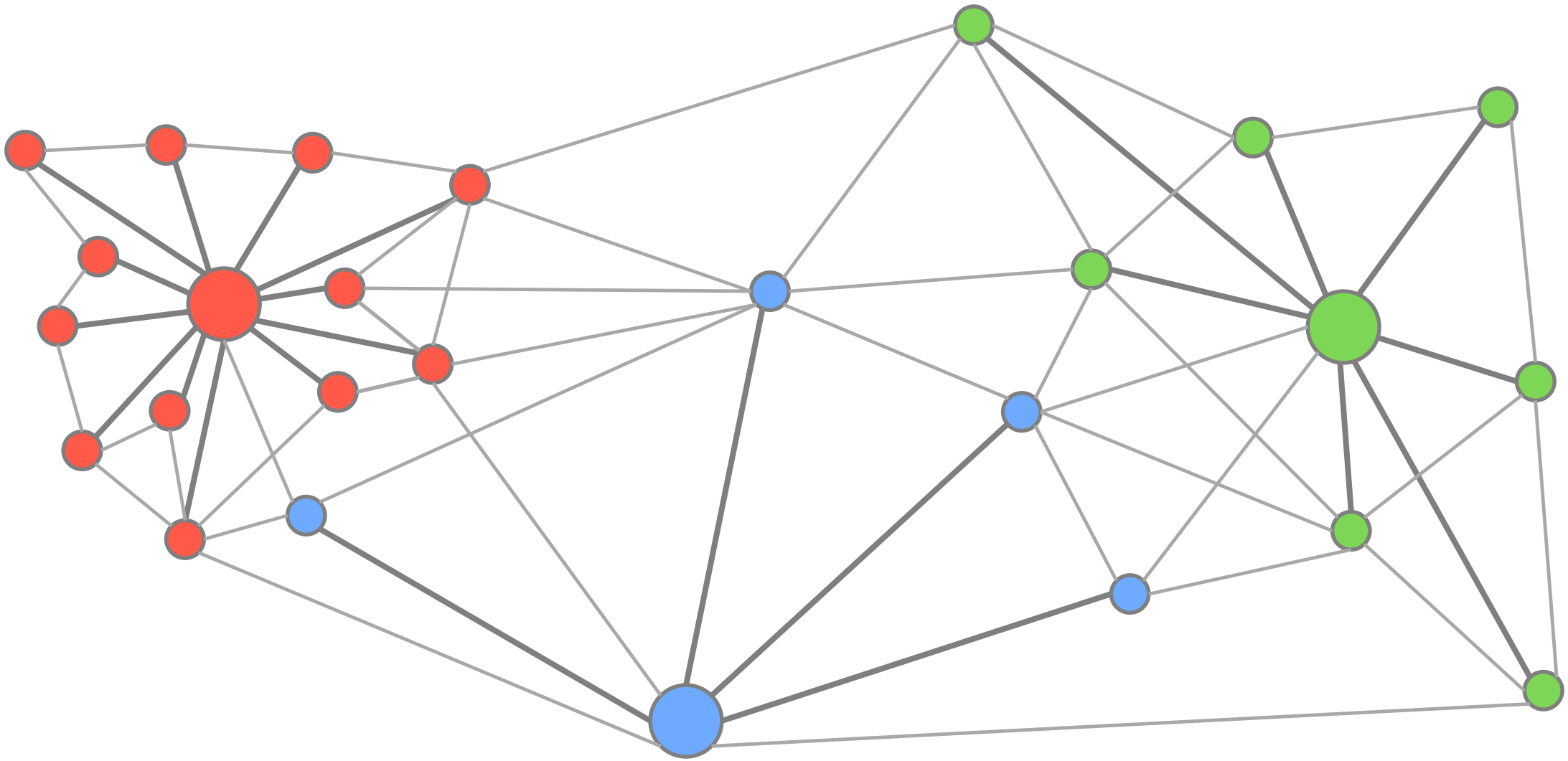
Minimum Spanning Tree



Minimum Spanning Tree

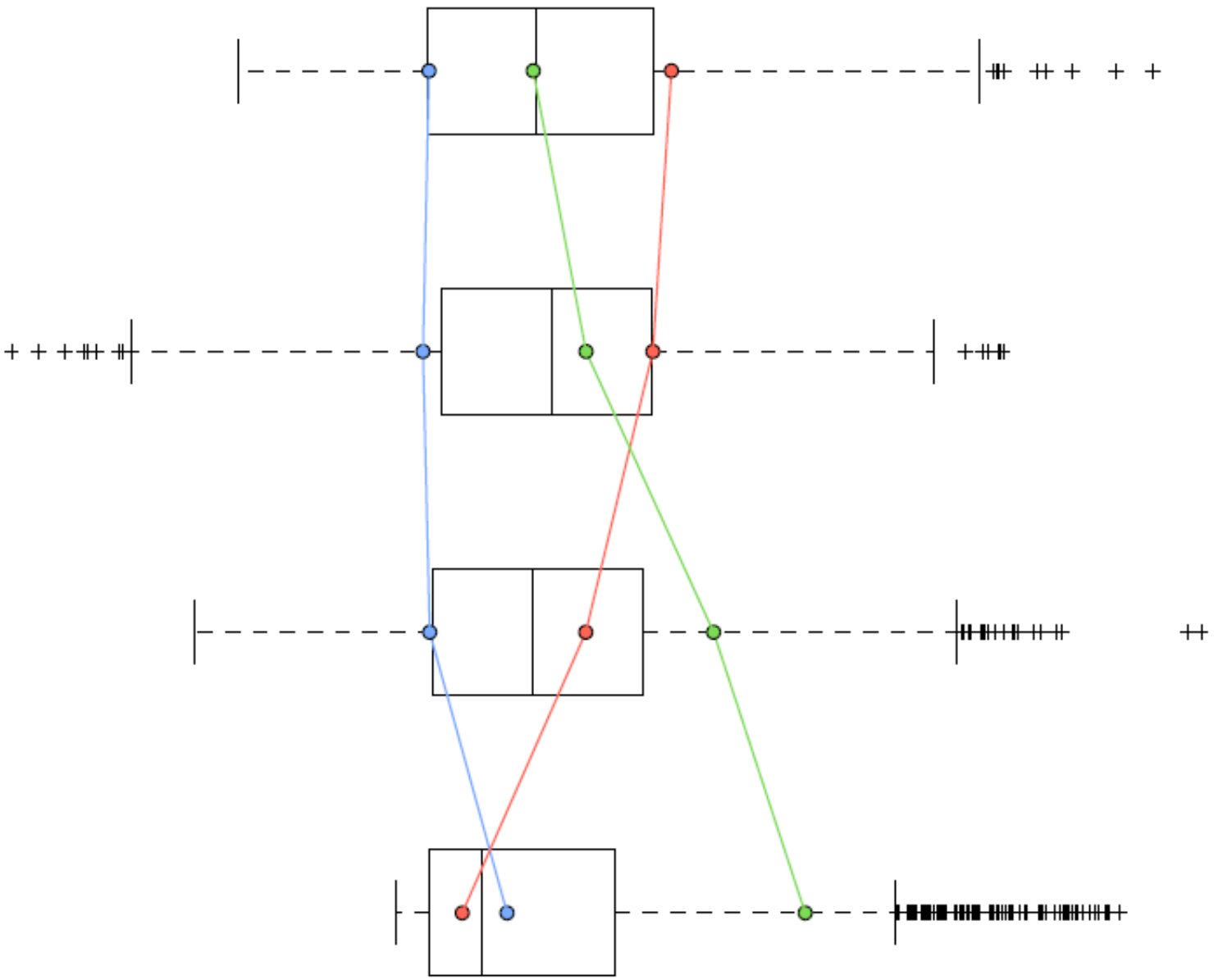


Minimum Spanning Tree



MATH!

interpret results through box plots



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


high performing store
potential store locations

LocID	PopDensity	AvIncome	DistToCompetition

LocID	PopDensity	AvIncome	DistToCompetition

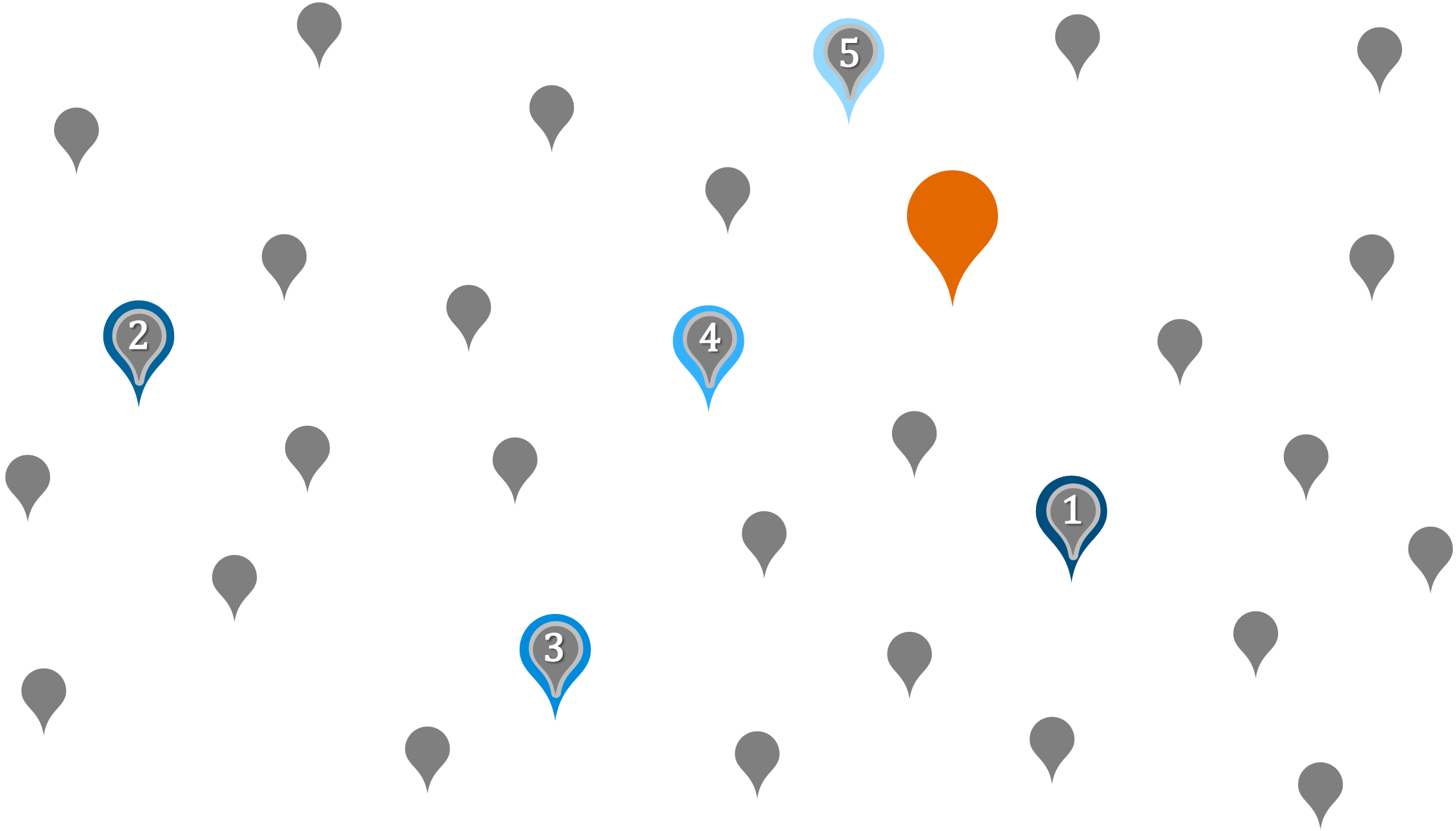
LocID	PopDensity	AvIncome	DistToCompetition

LocID	PopDensity	AvIncome	DistToCompetition

Rank by how similar they are to based on:

- **PopDensity**
- **AvIncome**
- **DistToCompetition**

LocID	PopDensity	AvIncome	DistToCompetition


5
2
4
1
3

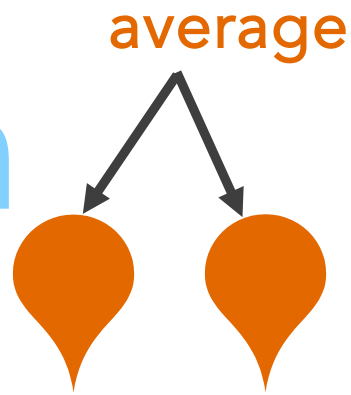
Input Feature(s) to Match

Candidate Features

Attributes of Interest

PopDensity	AvIncome	DistToCompetition

Input Feature(s) to Match



Candidate Features

Attributes of Interest

PopDensity	AvIncome	DistToCompetition

3 Match Methods

3 Match Methods

Attribute Values

3 Match Methods

Attribute Values

Ranked Attribute Values

3 Match Methods

Attribute Values

Ranked Attribute Values

Attribute Profiles

Attribute Values

Attribute Values

standardize attributes

Z-transform:

$$(x - \bar{x}) / SD$$

Attribute Values

standardize attributes

Population = 14,159

% Uninsured = .26

Distance (km) = 535.89

Attribute Values

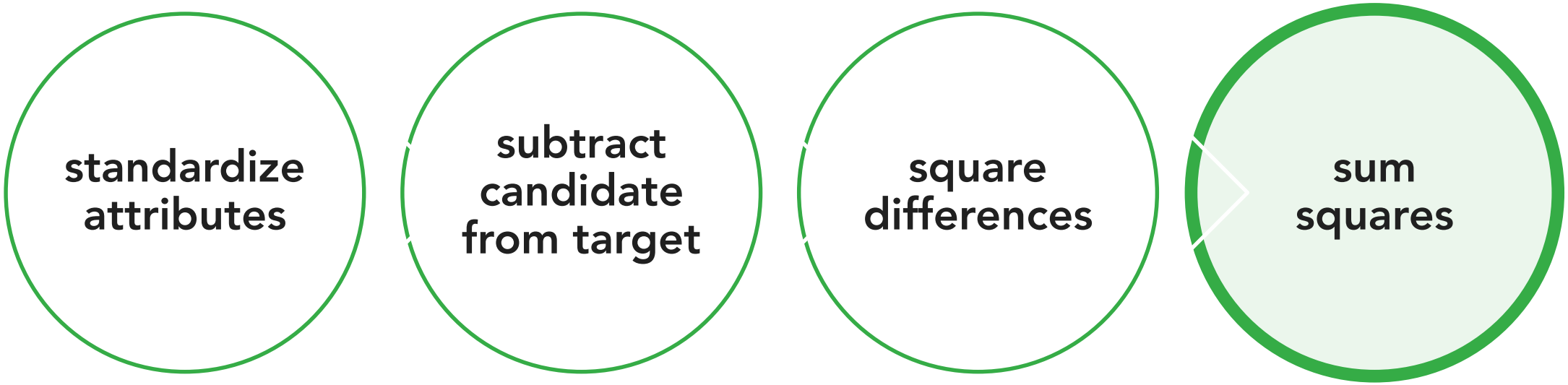
standardize attributes

Population = -.7932

% Uninsured = 3.8462

Distance (km) = .6433

Attribute Values

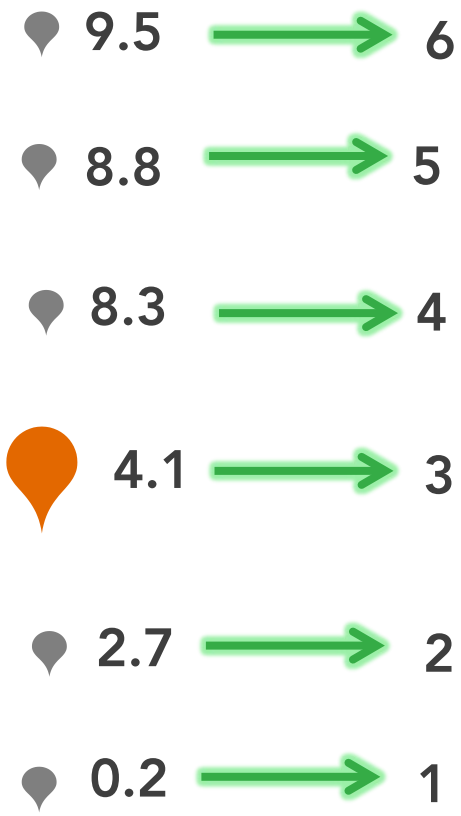


Ranked Attribute Values

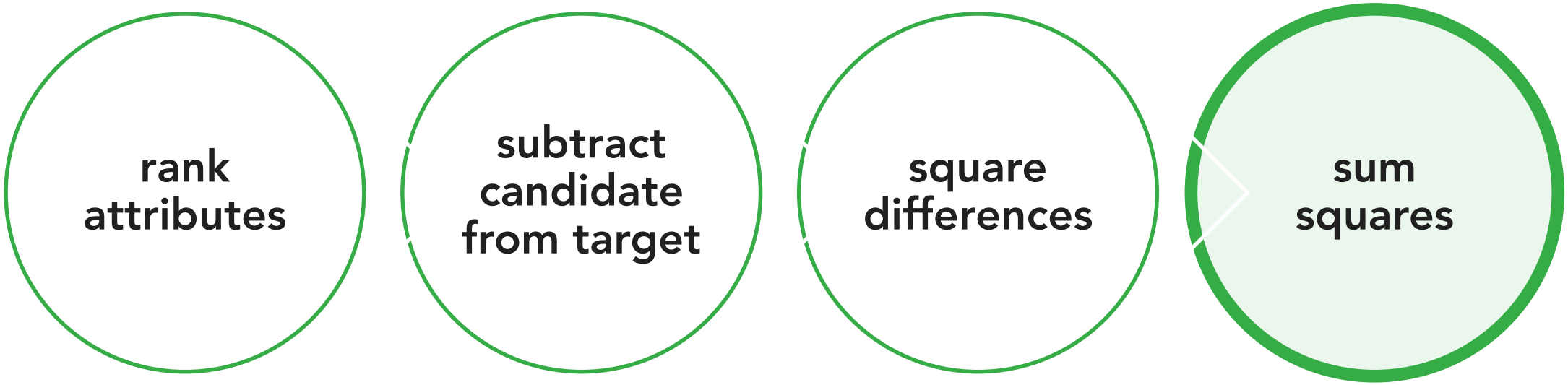
Ranked Attribute Values

Ranked Attribute Values

Ranked Attribute Values



Ranked Attribute Values

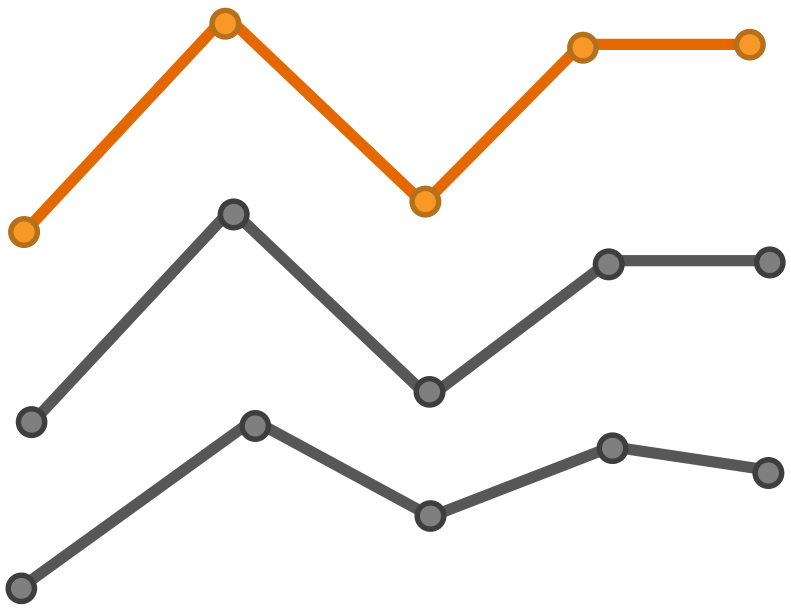


Attribute Profiles

Attribute Profiles

standardize attributes

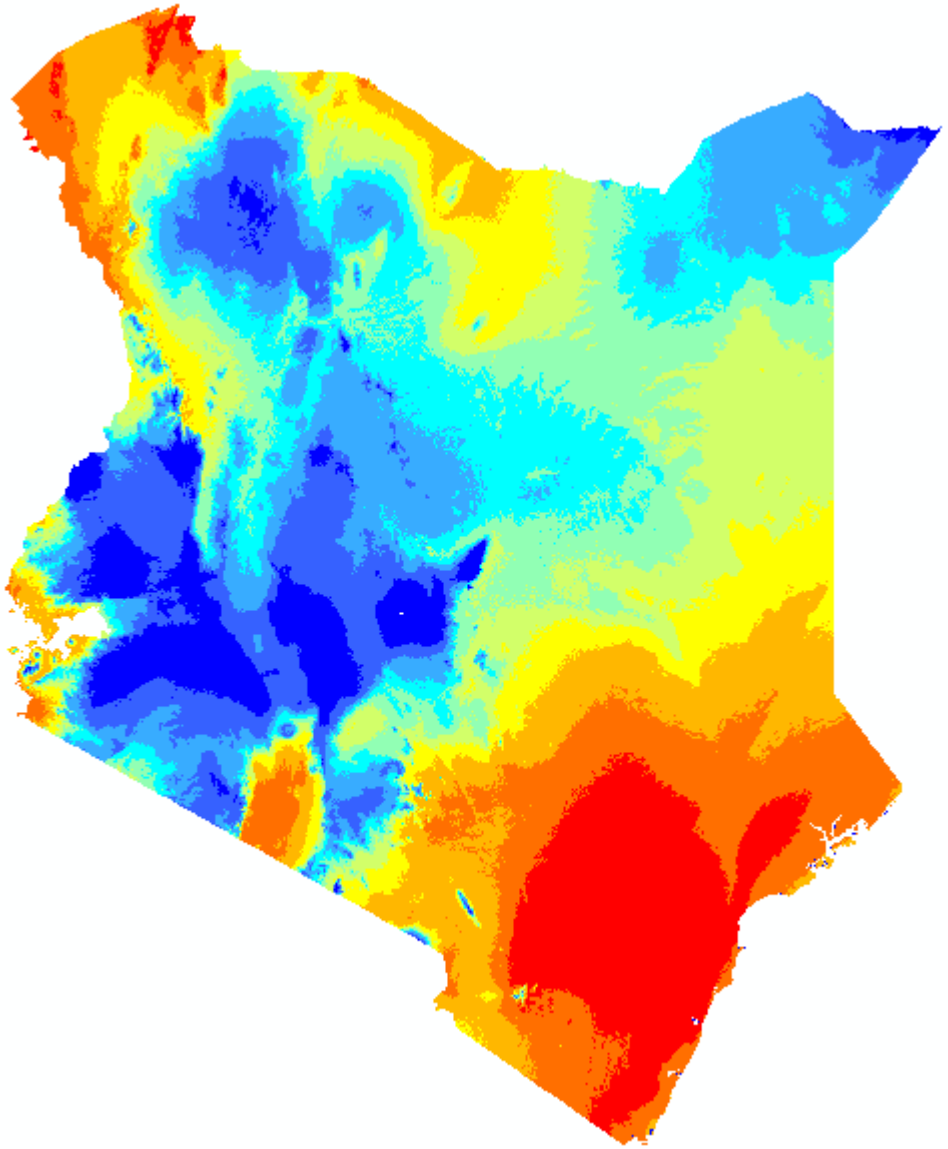
cosine similarity index



$$\text{cosine similarity index} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2}\ \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

* Must have at least 2 attributes of interest

Dengue Fever Risk in Kenya



questions?

fvale@esri.com
jdacosta@esri.com

Want to learn more???

esriurl.com/spatialstats

esri