esri | THE SCIENCE OF WHERE

Means & Medians to Machine Learning: Spatial Statistics Basics and Innovations

Lauren Bennett, PhD

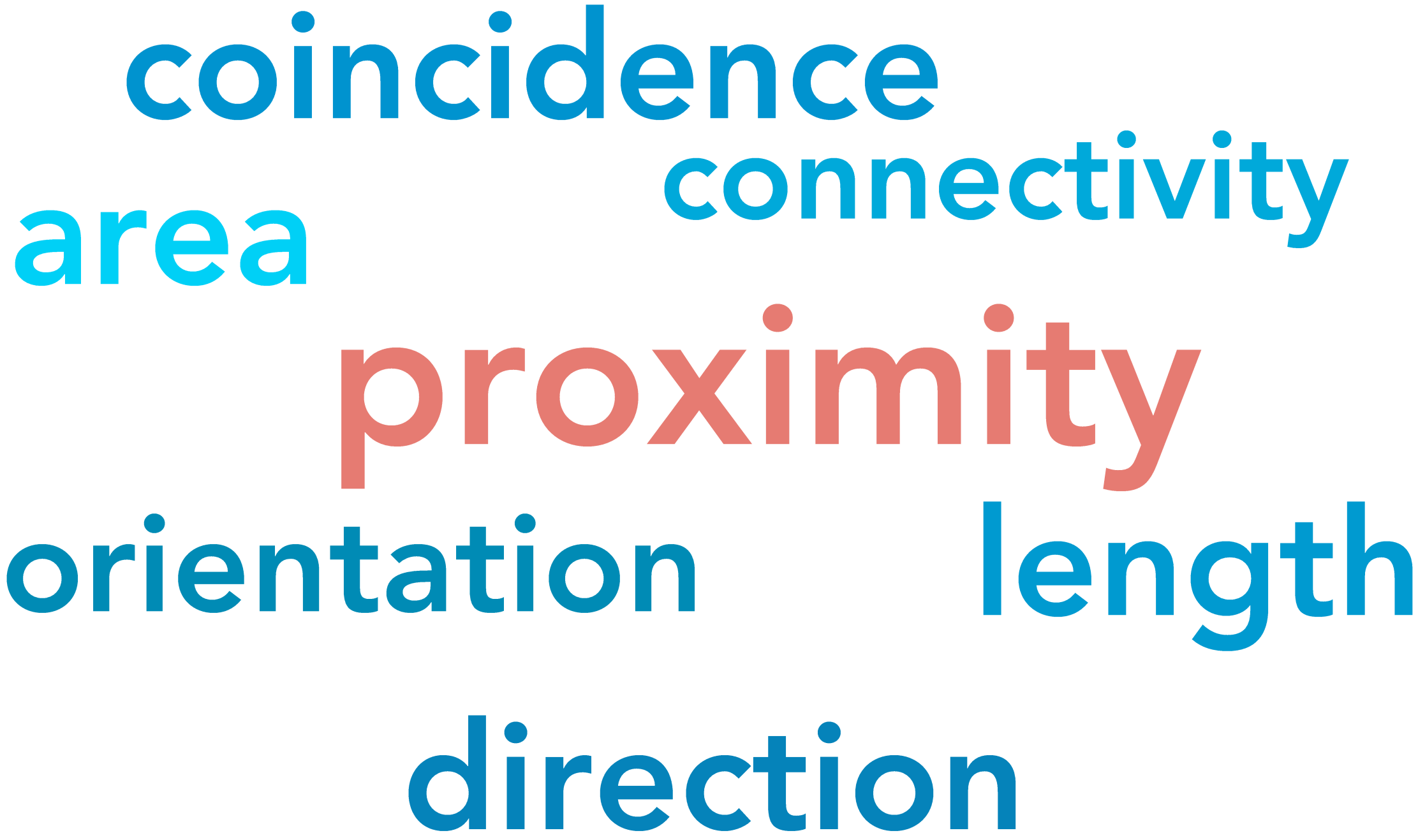
Ryan Ruthart

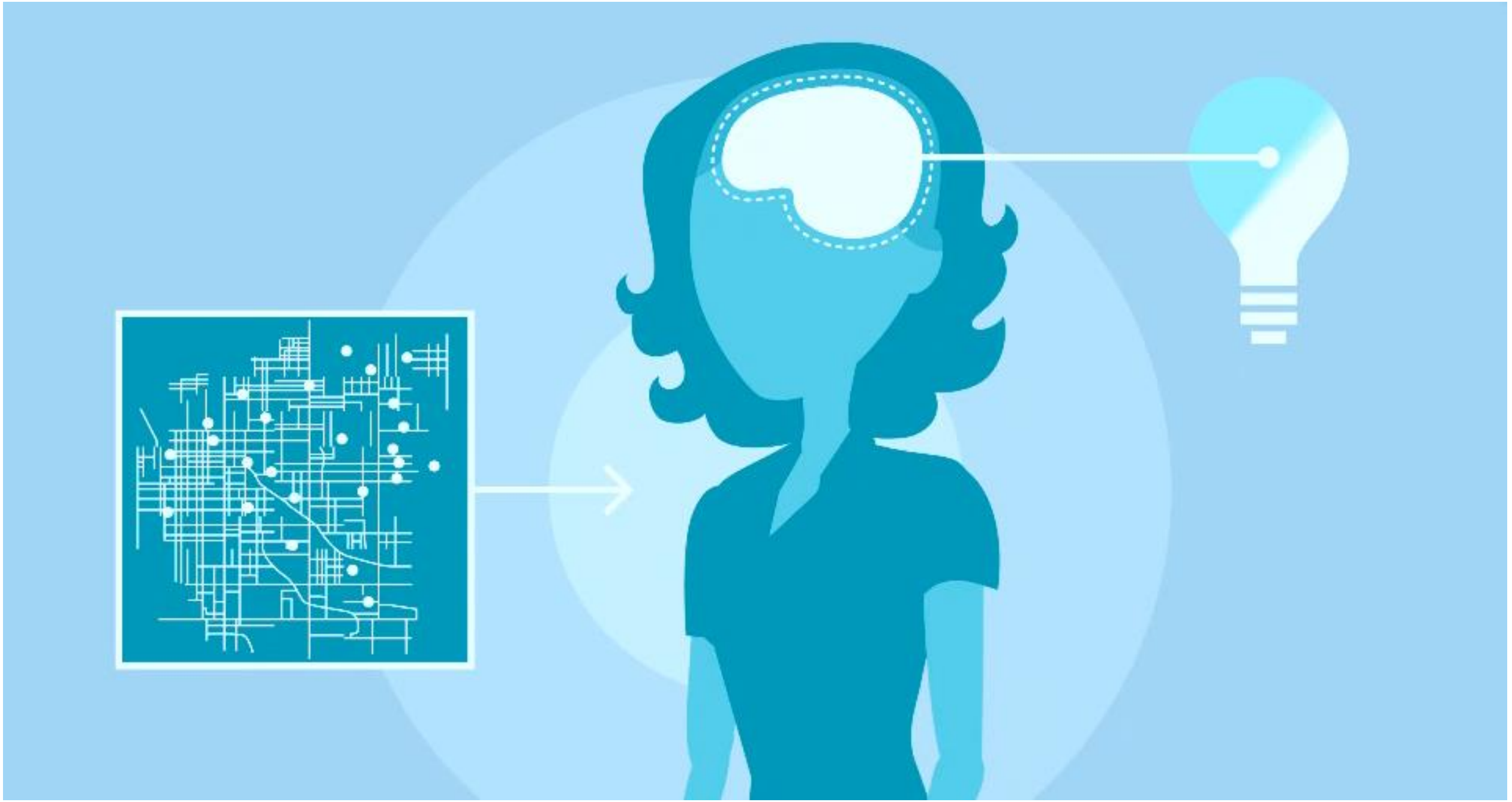
Flora Vale

esriurl.com/spatialstats

What are Spatial Statistics?

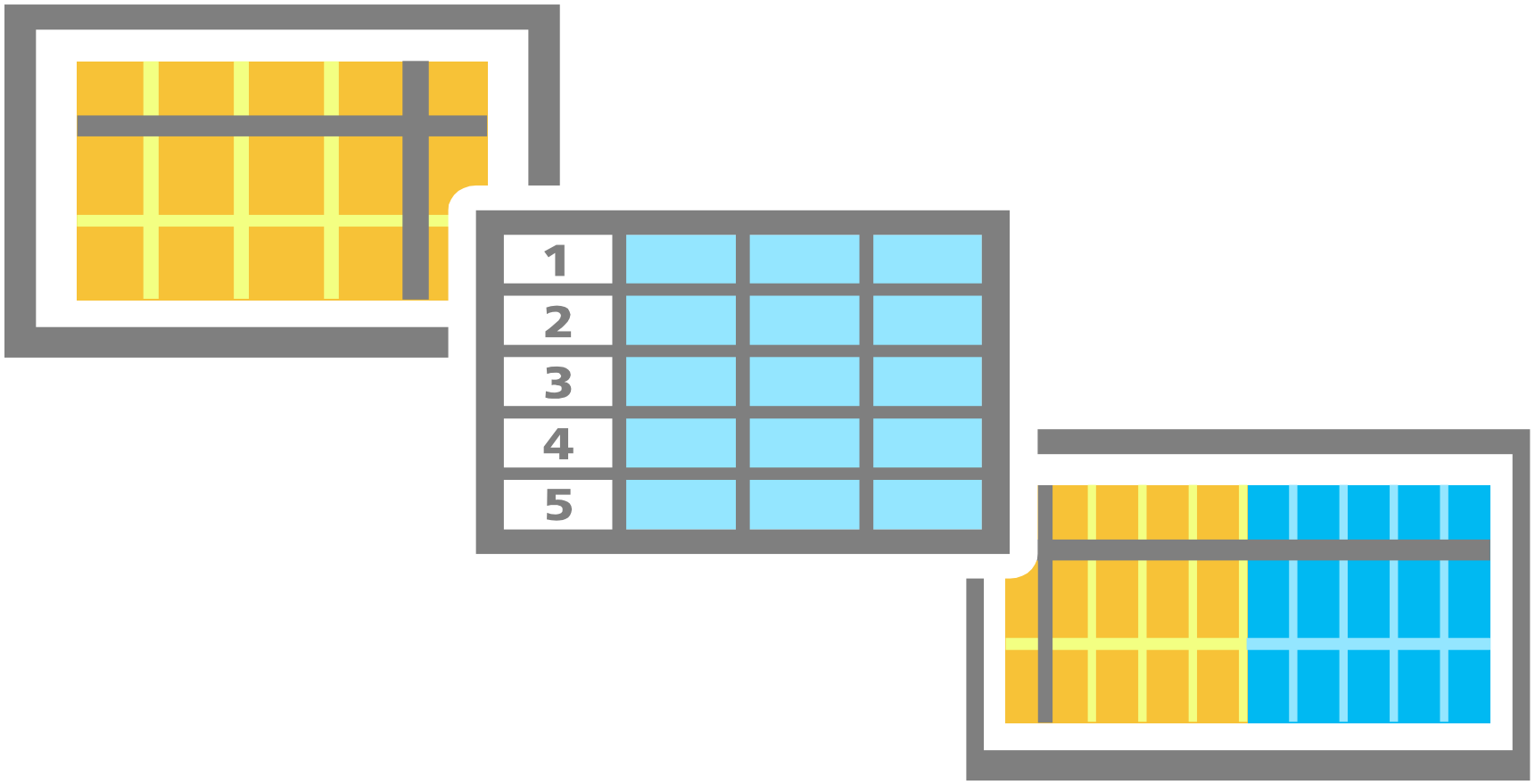
Spatial Statistics are a set of exploratory techniques for describing and modeling spatial distributions, patterns, processes, and relationships.


coincidence
connectivity
area
proximity
orientation
length
direction



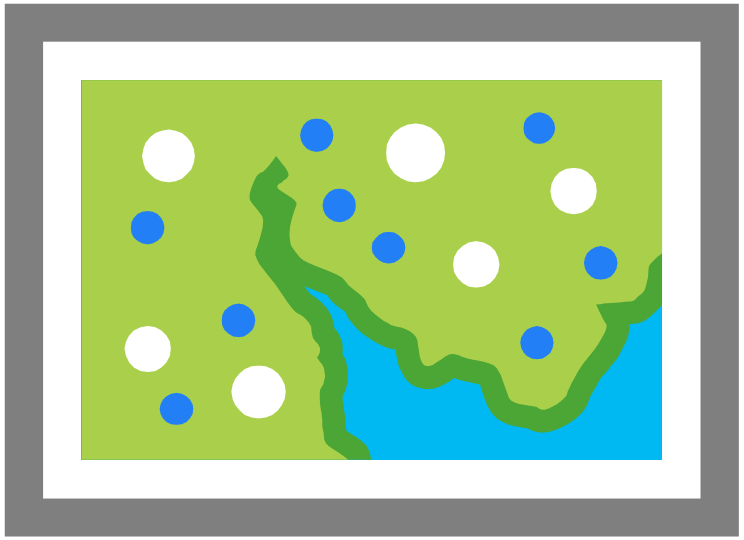
Spreadsheets

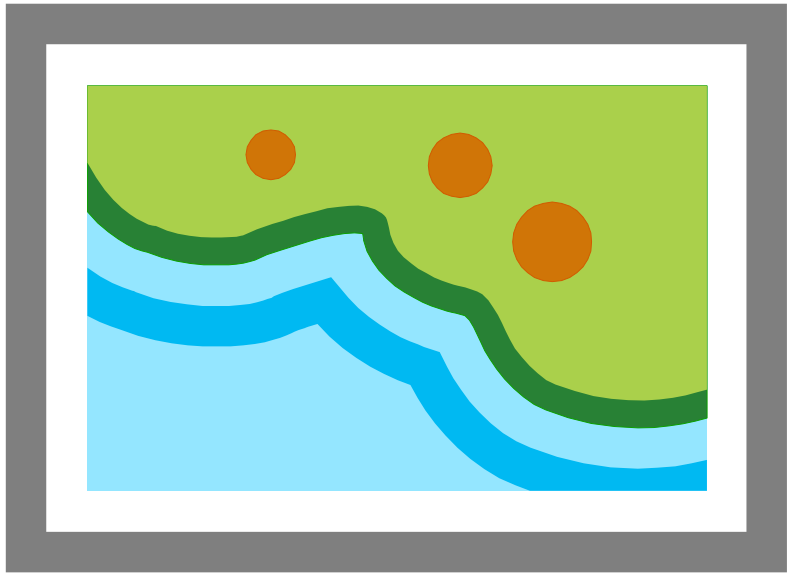
Data or Information?



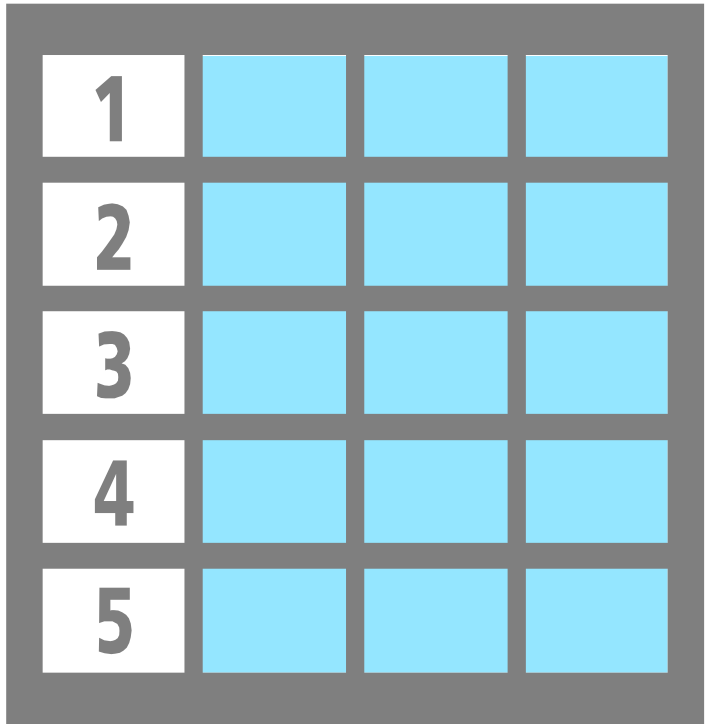
Maps

Data or Information?

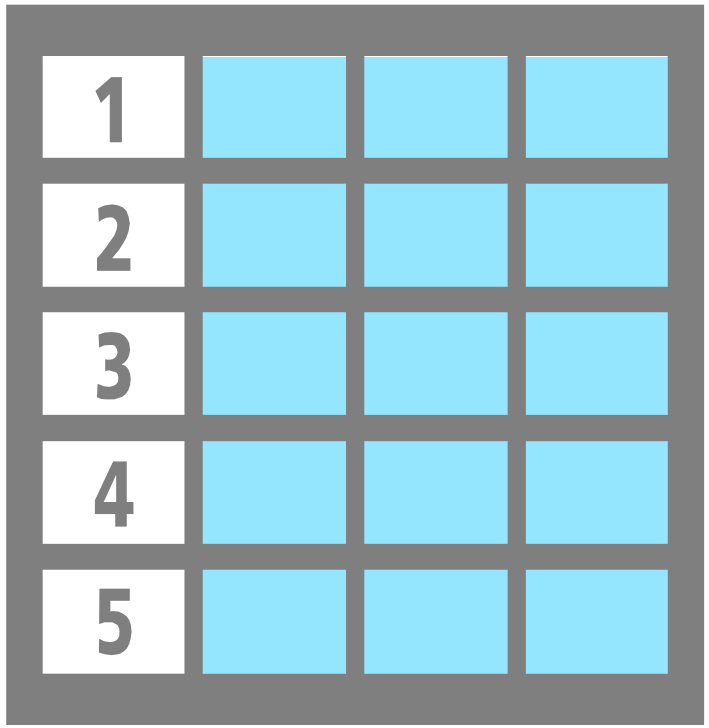


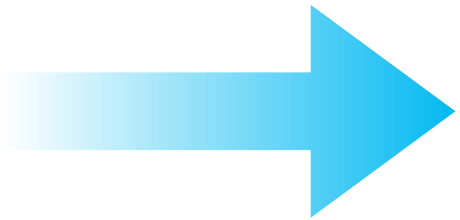


When you look at a spreadsheet...



You ask for more

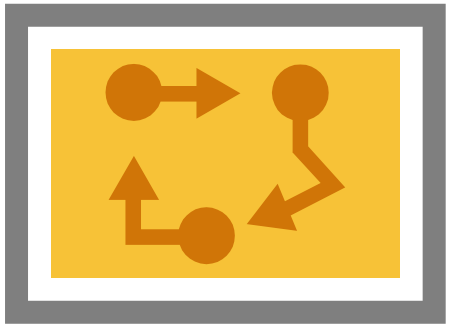


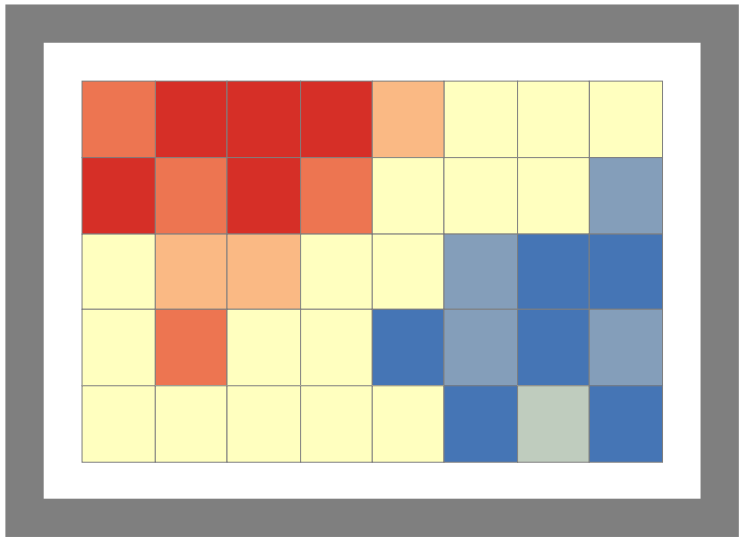


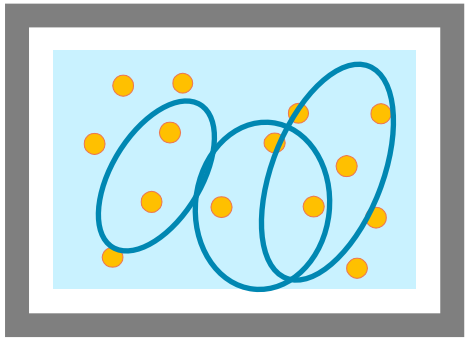
- Mean
- Standard Deviations
- Min and Max
- …

Same goes for maps!

We can do more







Means and Medians

→ summarizing spatial distributions

Machine Learning

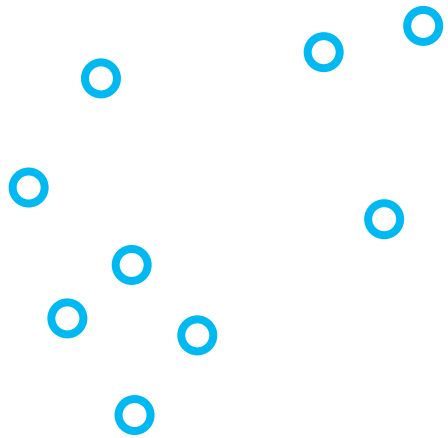
→ clustering methods

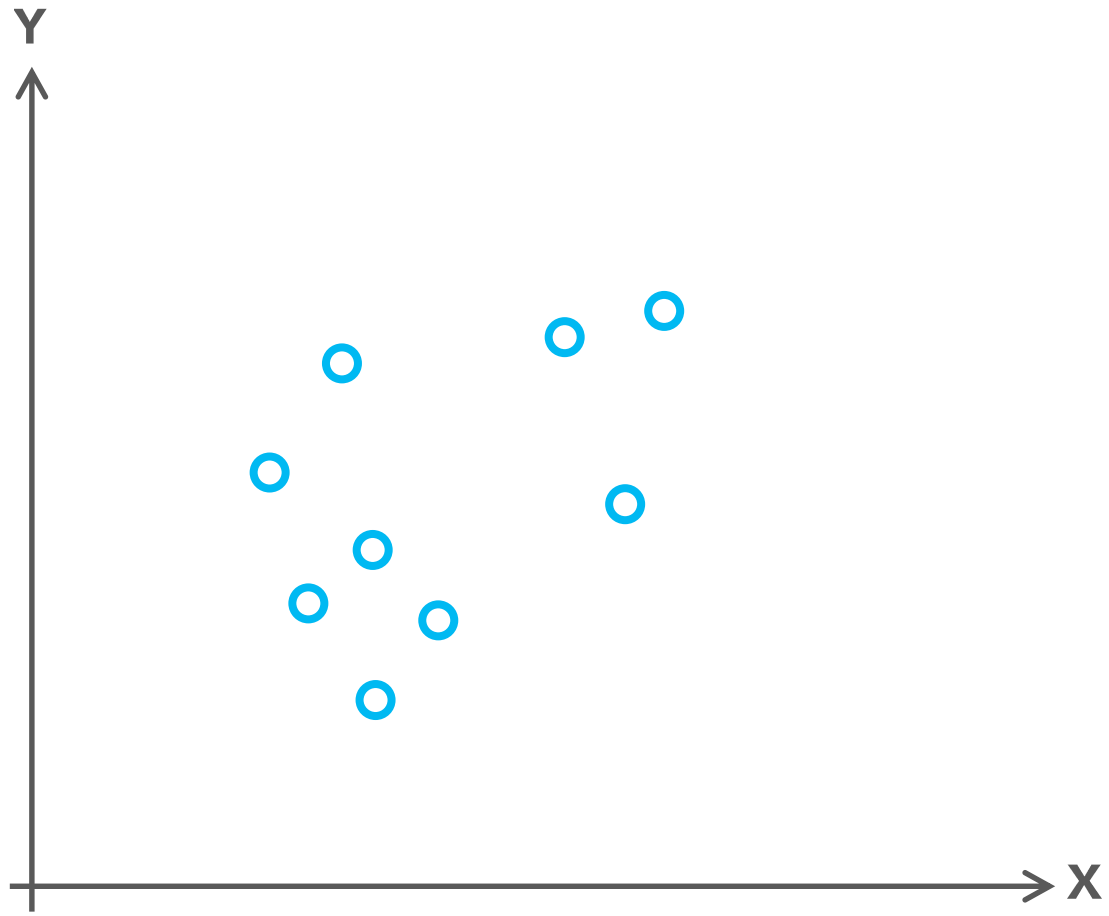
Means and Medians

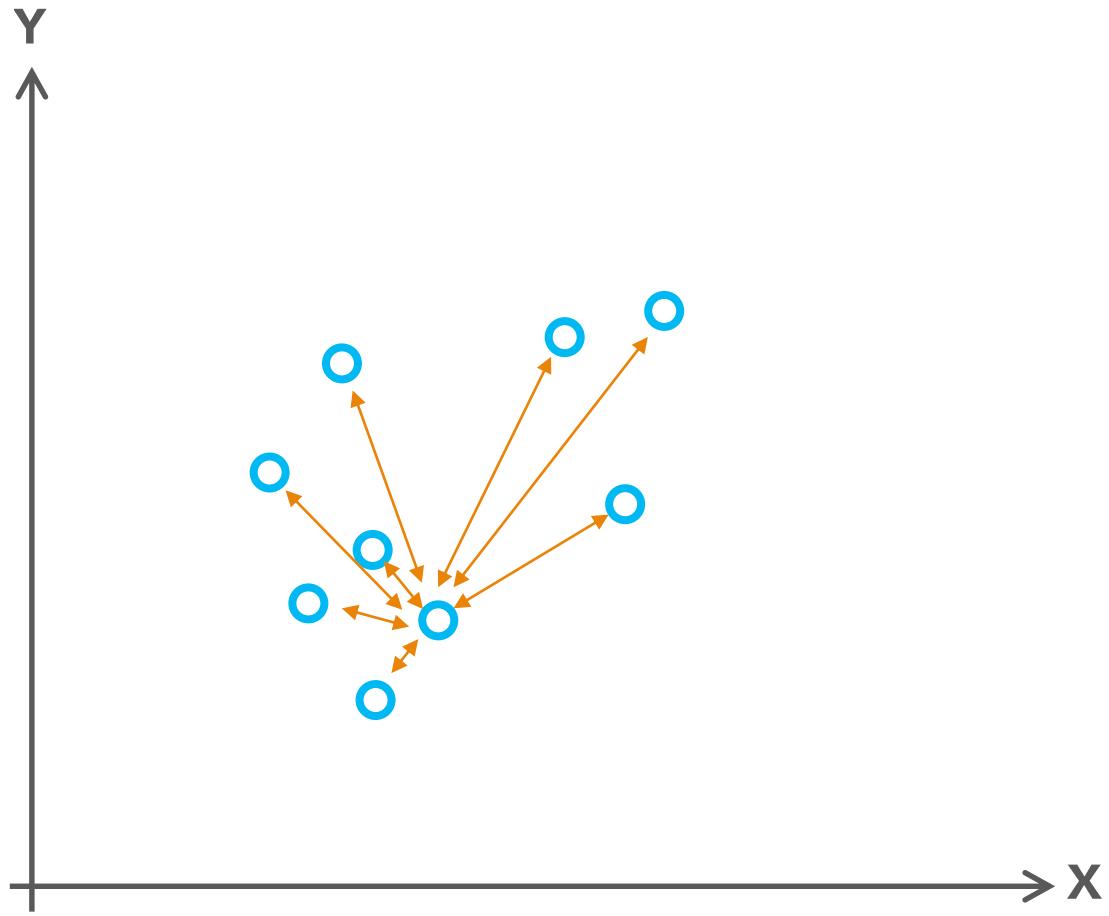
summarizing spatial distributions

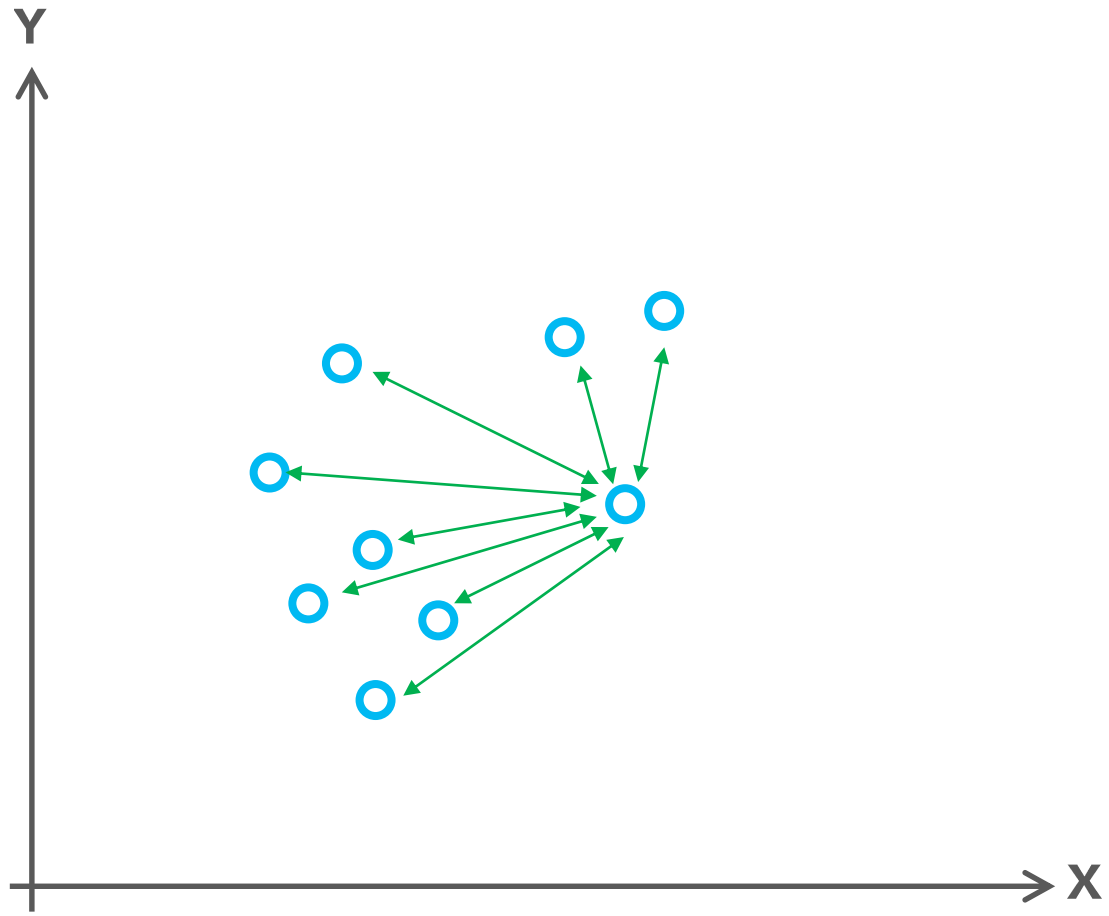
Central Feature

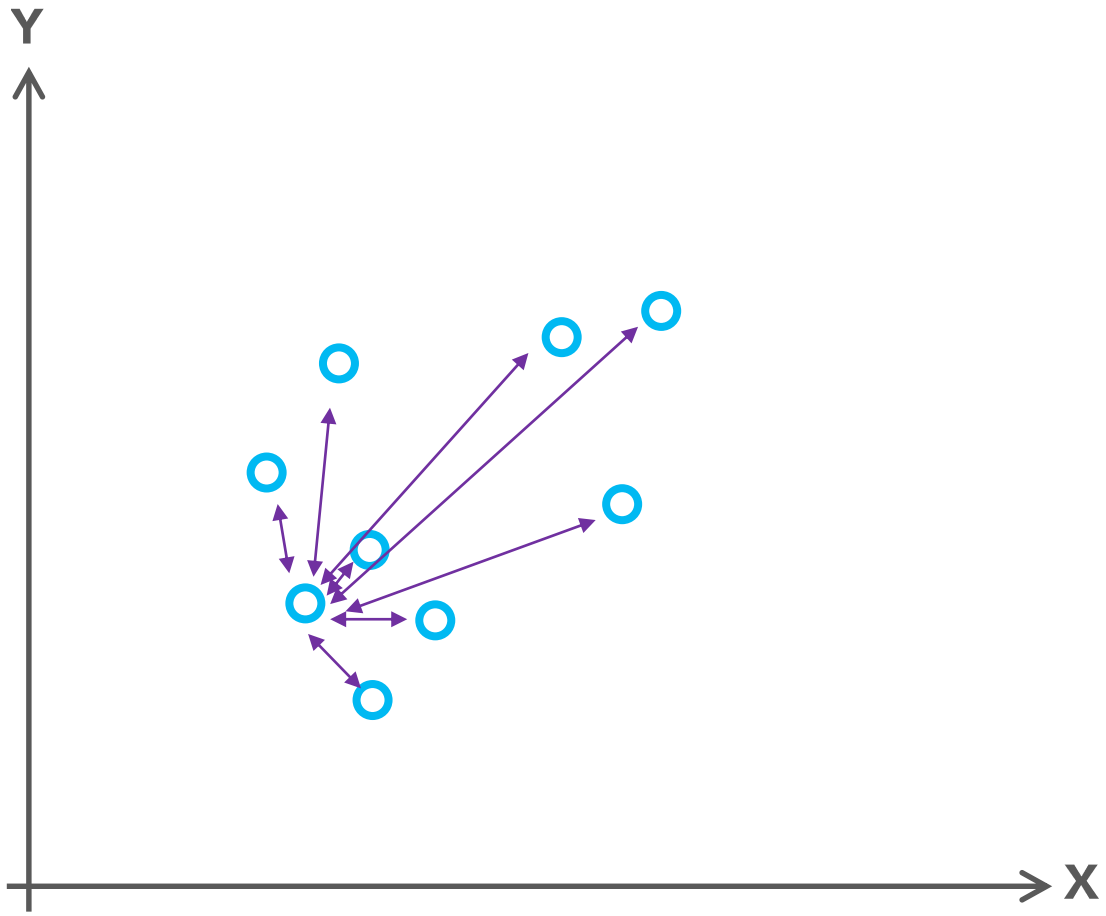
identifies the most centrally located feature in a point, line, or polygon feature class

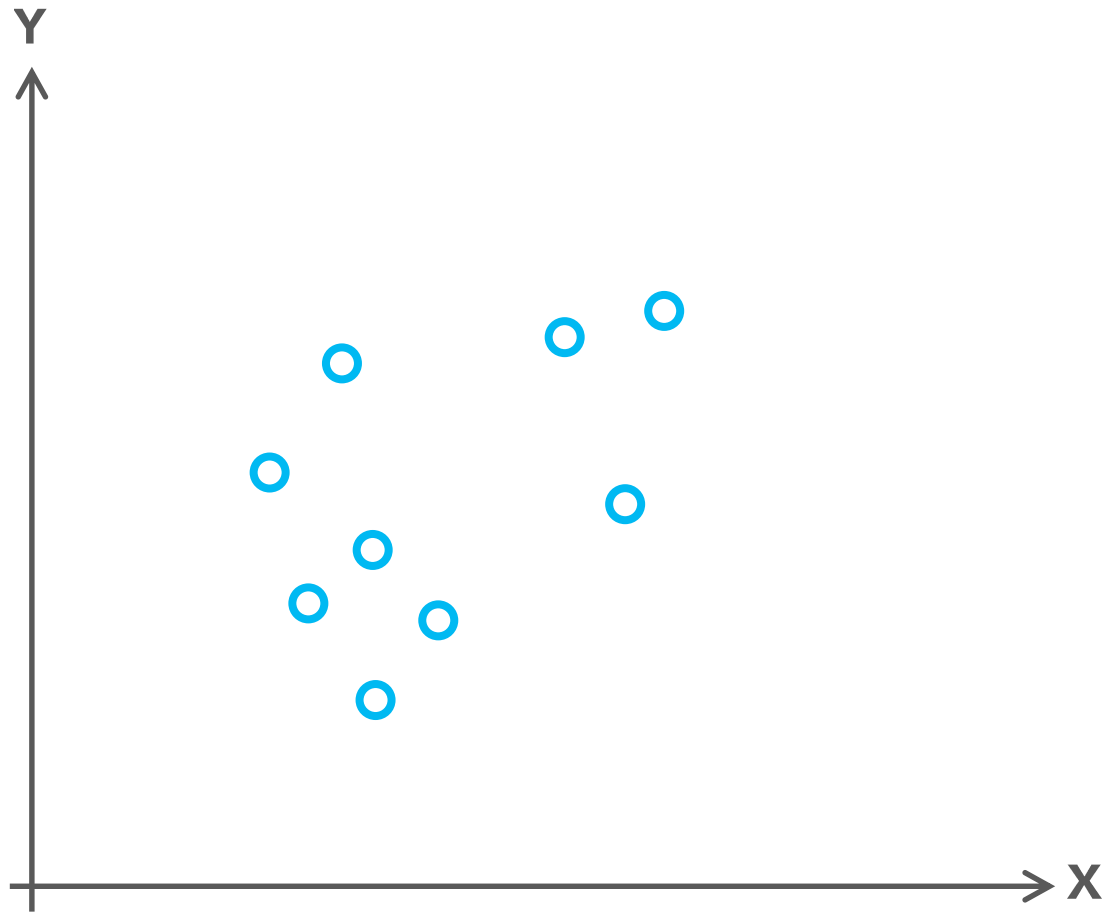


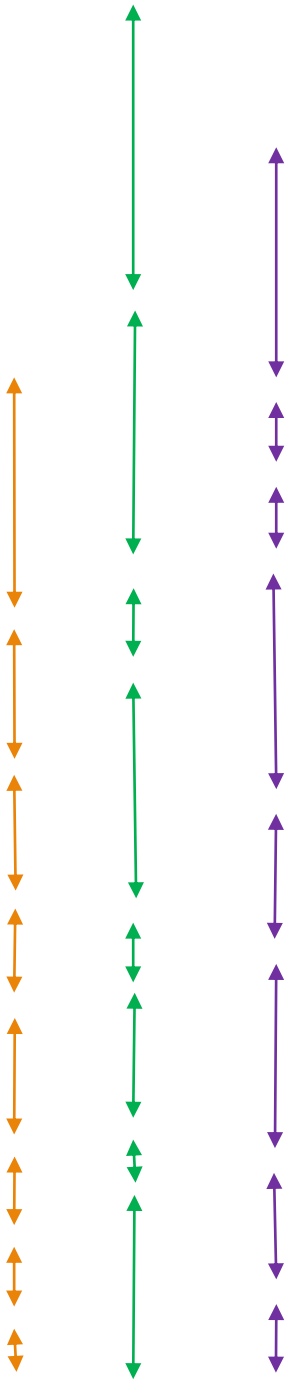

Y
X

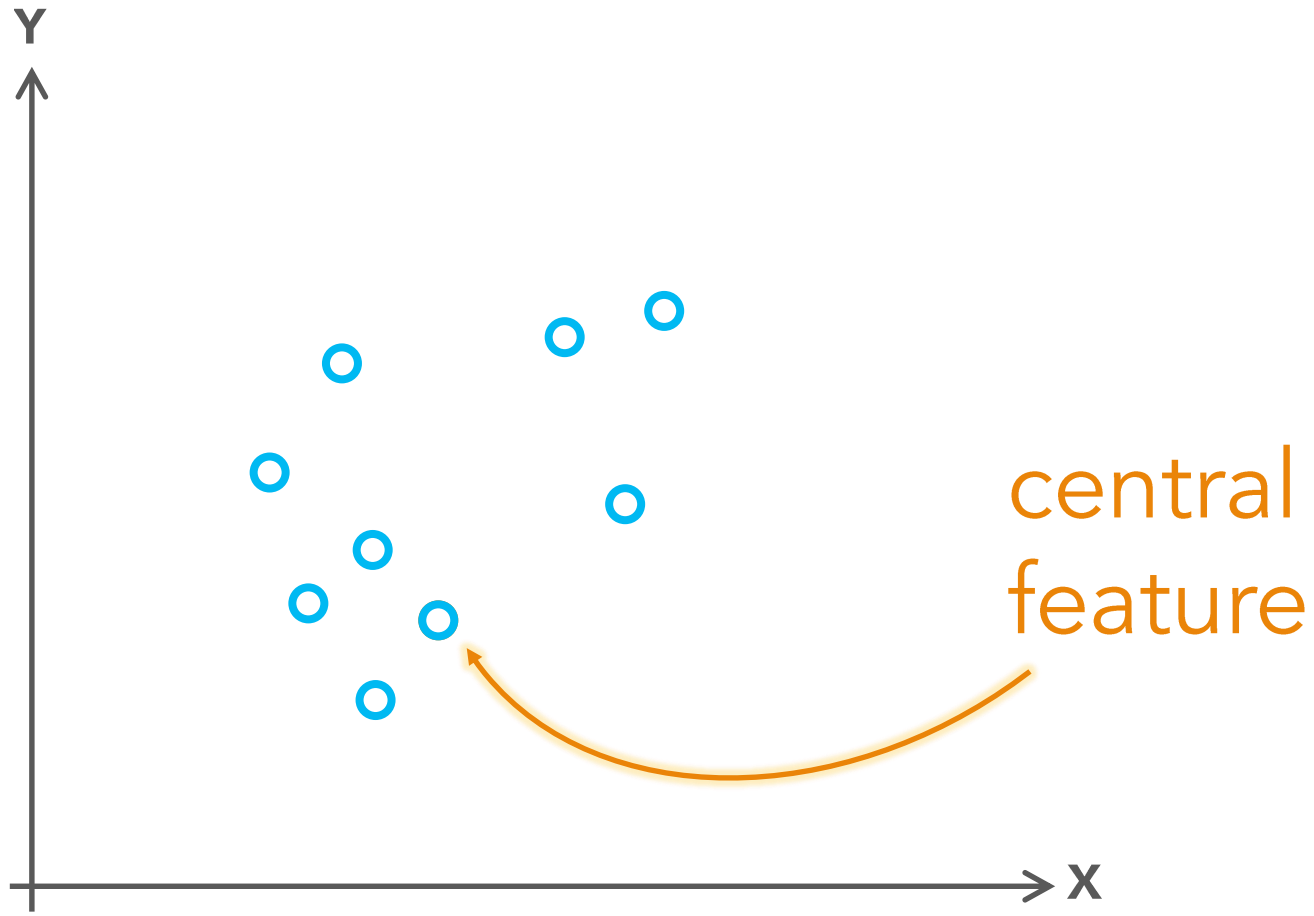

Y
X


Y
X


Y
X

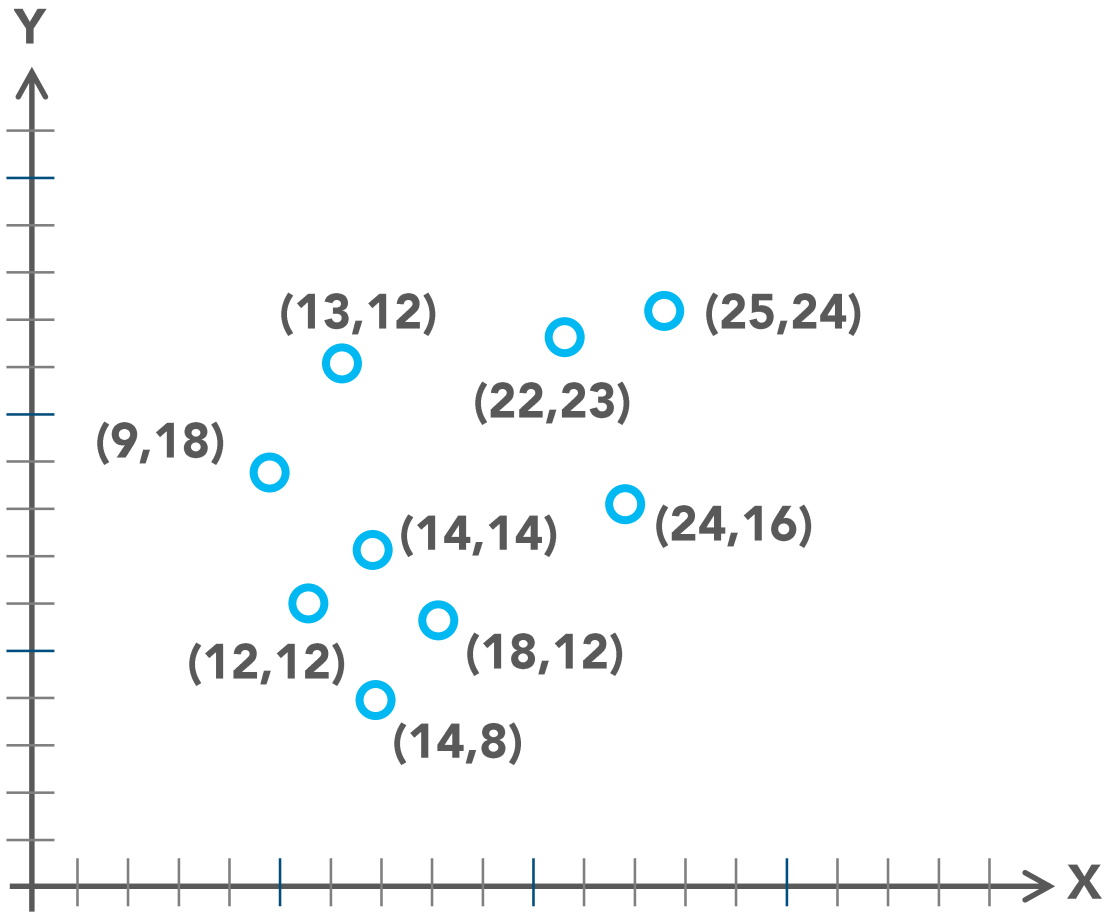

Y
X

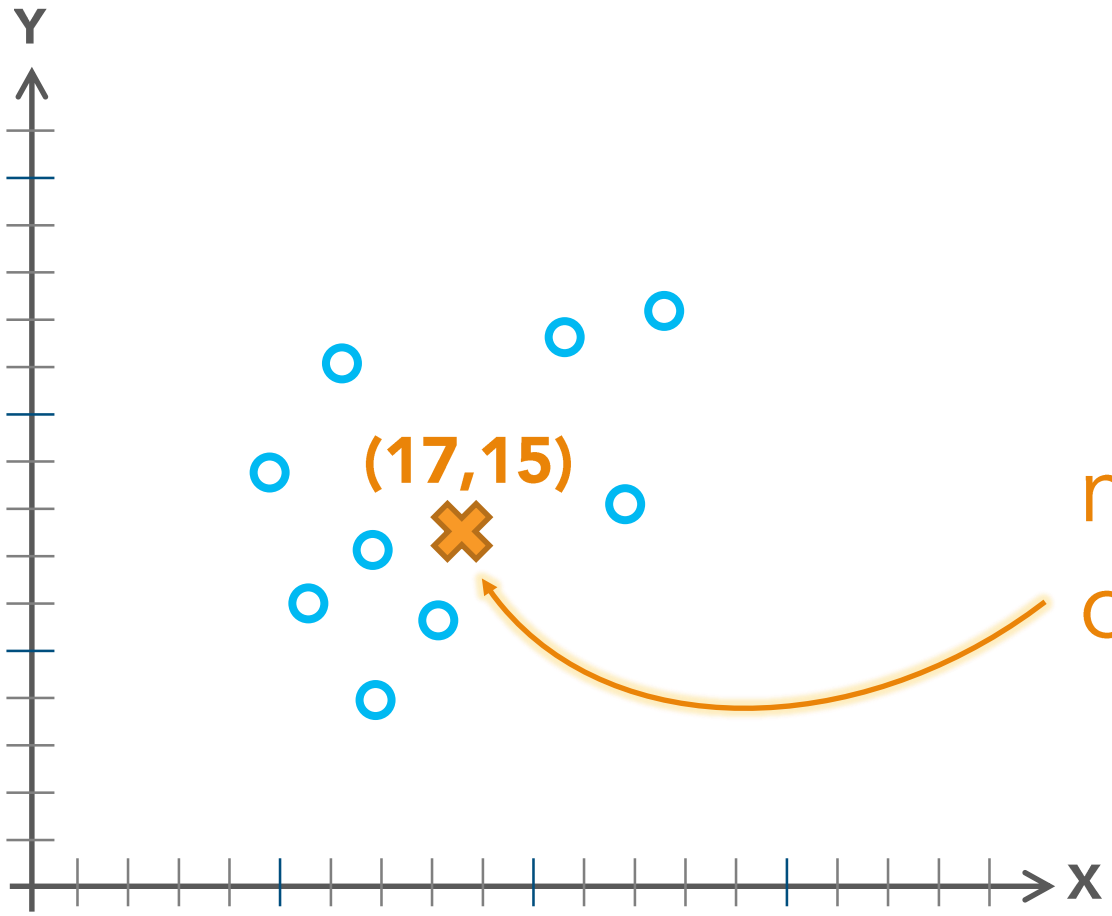



Y
X
central
feature

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)
(14,14)
(24,16)
(12,12)
(18,12)
(14,8)



mean
center

(14,14)

(13,12)

(25,24)

(24,16)

(22,23)

(18,12)

(12,12)

(14,8)

(9,18)

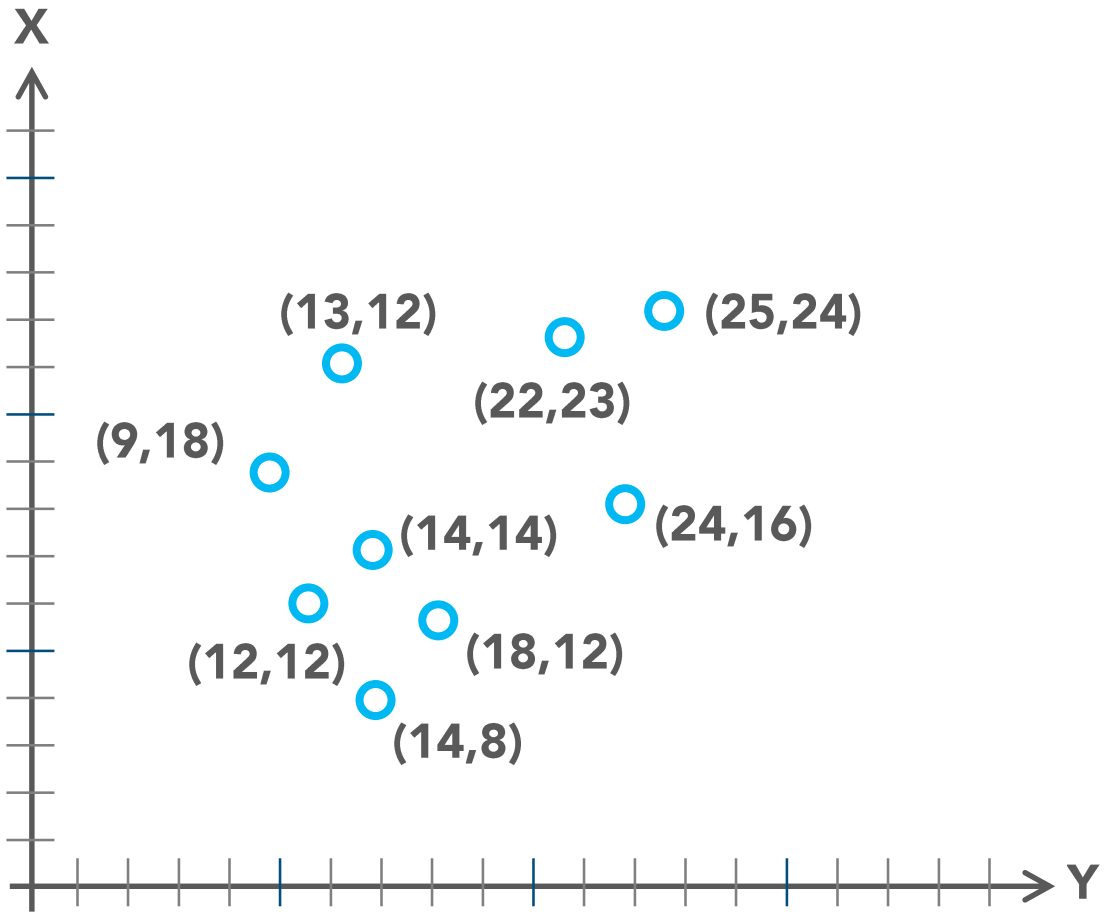
---

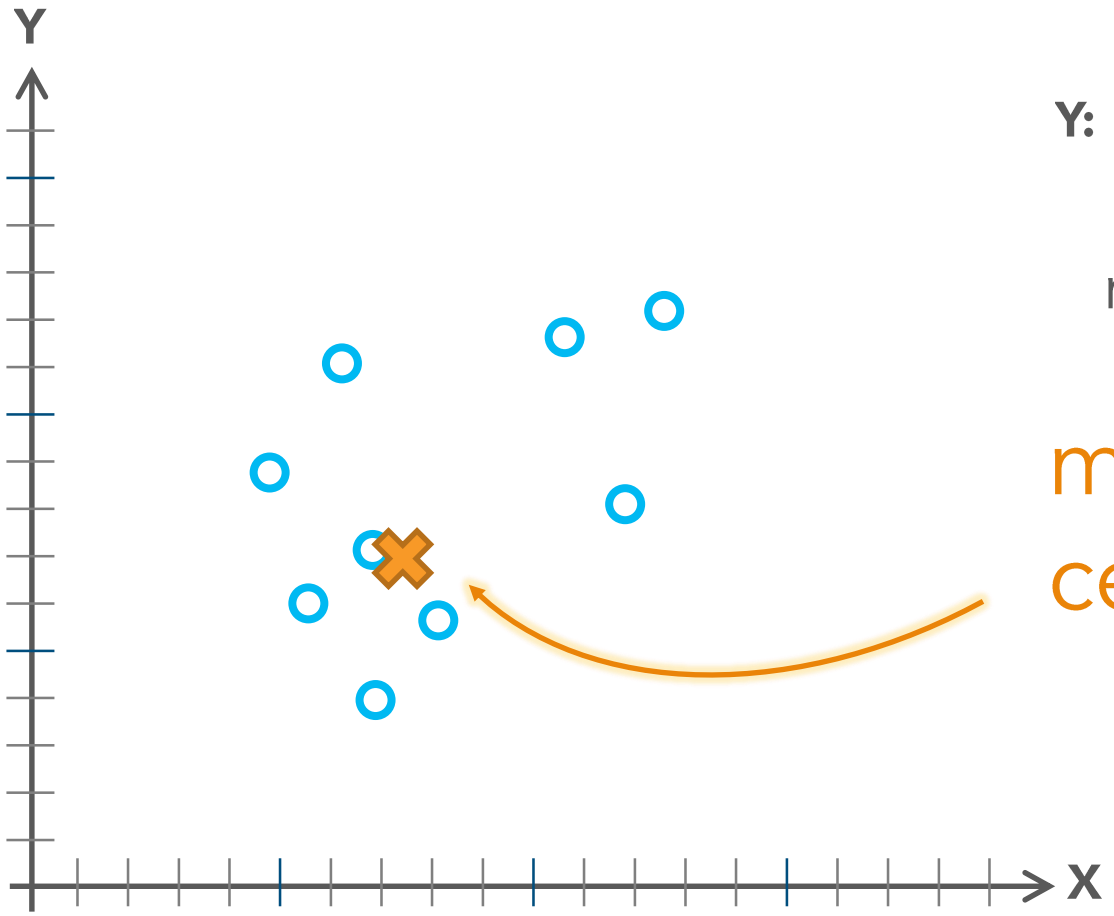
mean =

(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)
(14,14)
(24,16)
(12,12)
(18,12)
(14,8)



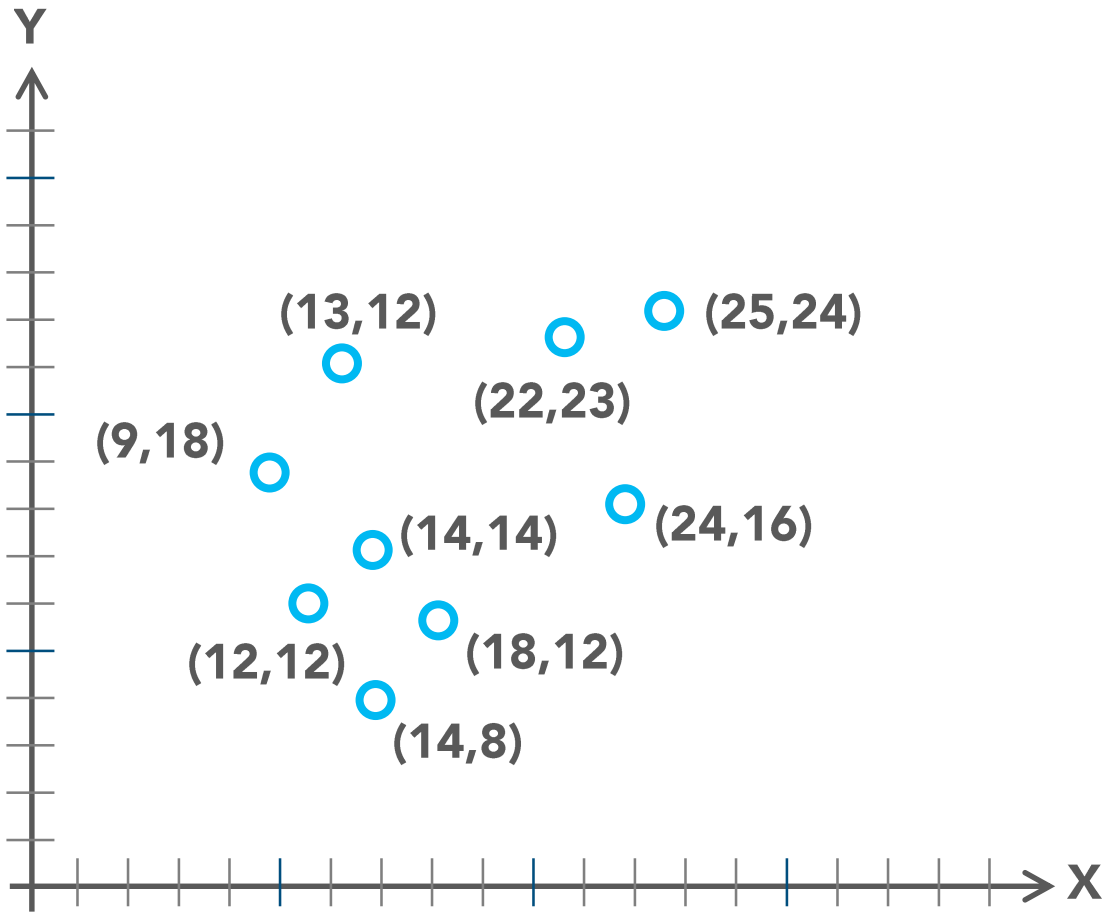
X: ~~25~~ . ~~24~~ . ~~22~~ . ~~18~~ . 14 . ~~14~~ . ~~13~~ . ~~12~~ . ~~8~~

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

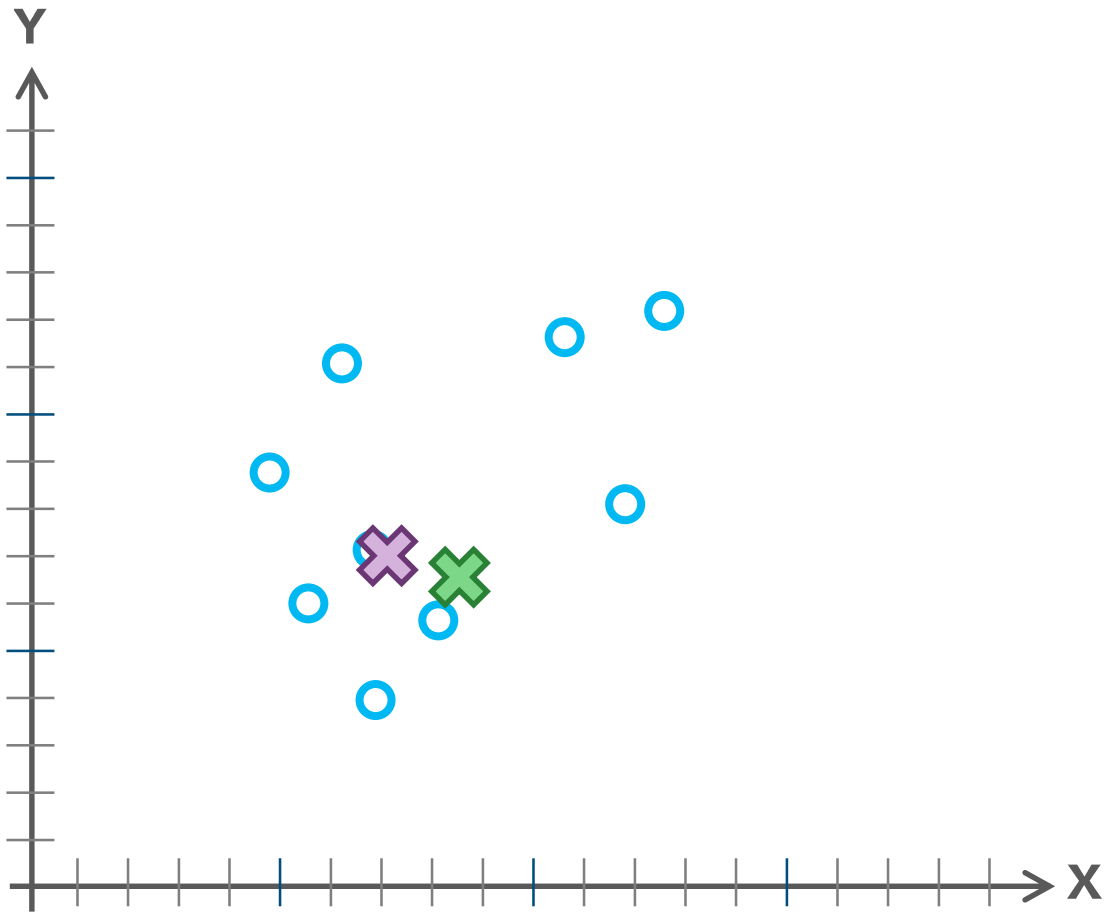
median = **(14,14)**

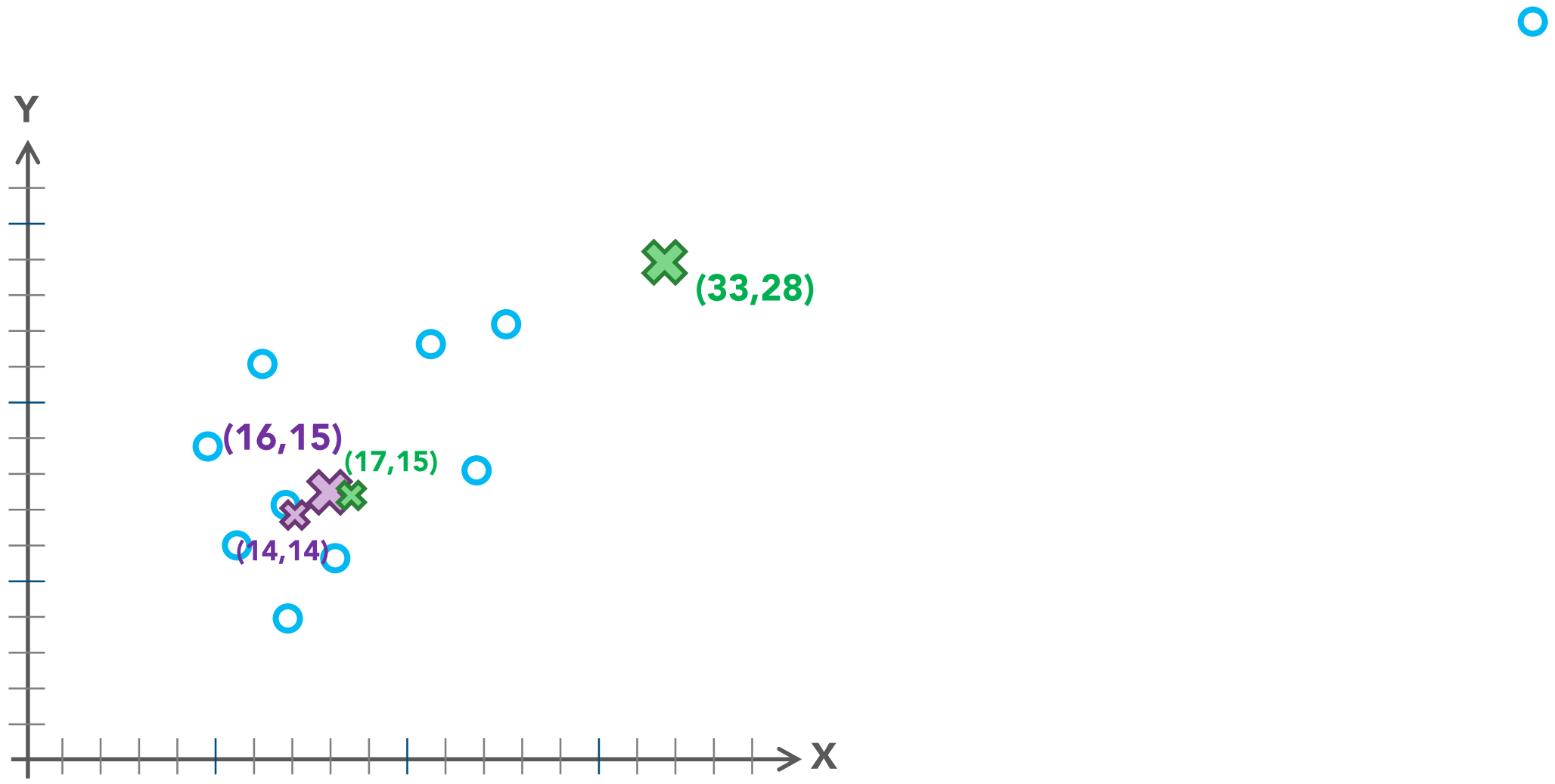
median
center

Mean vs Median?

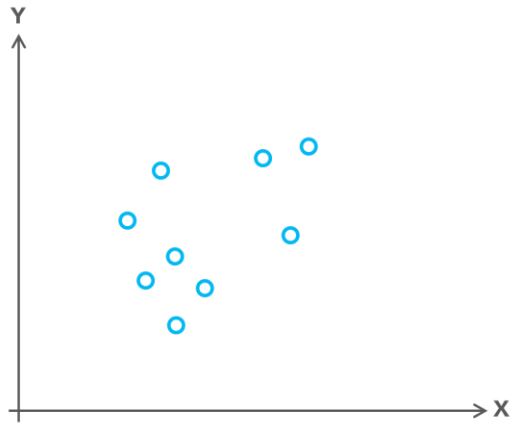


(176,138)


Y
X

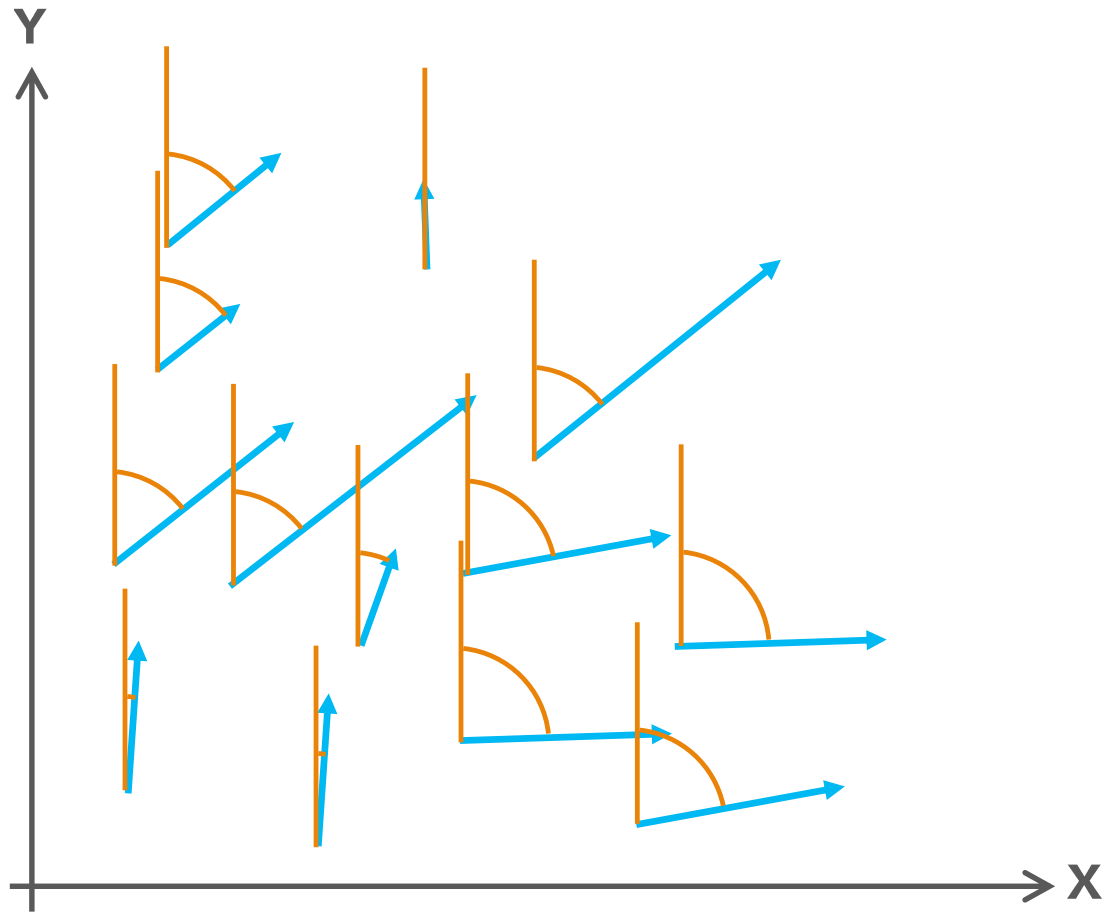

Y
X
(33,28)
(16,15)
(17,15)
(14,14)

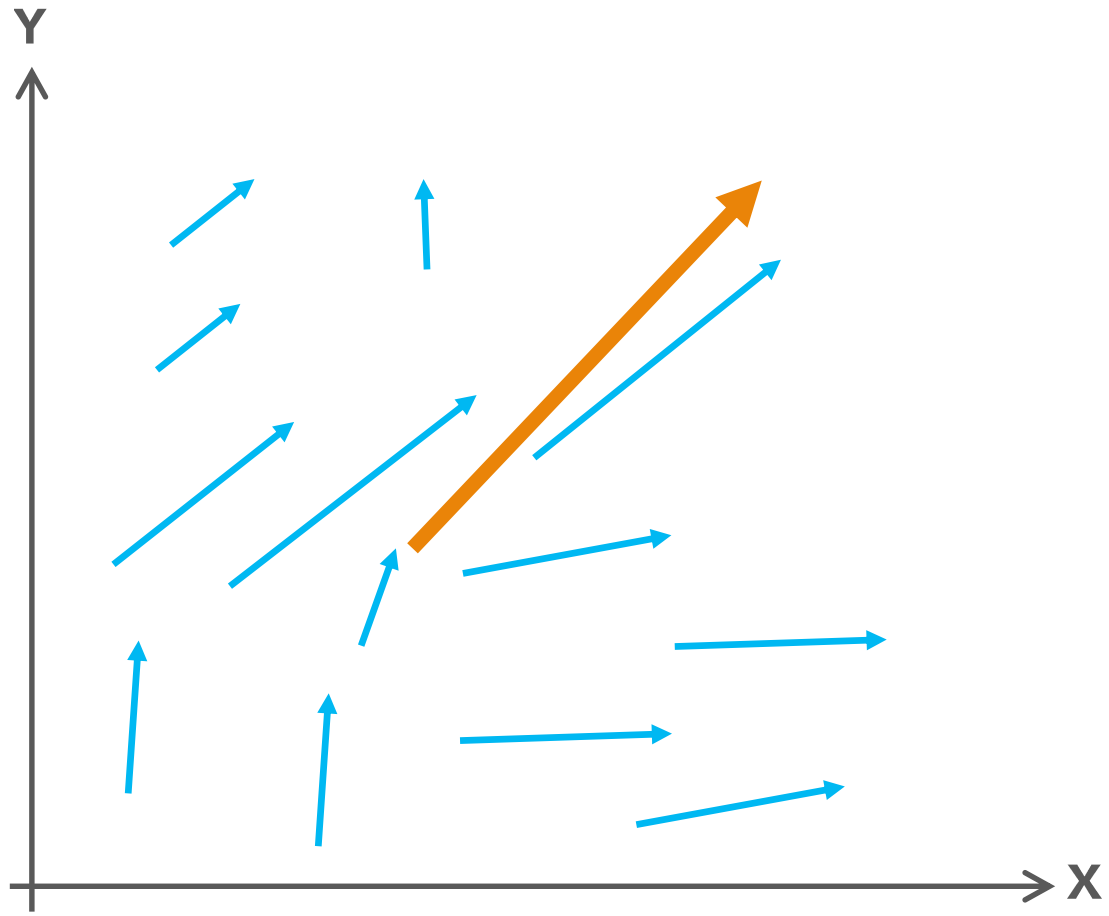
Demo



Linear Directional Mean

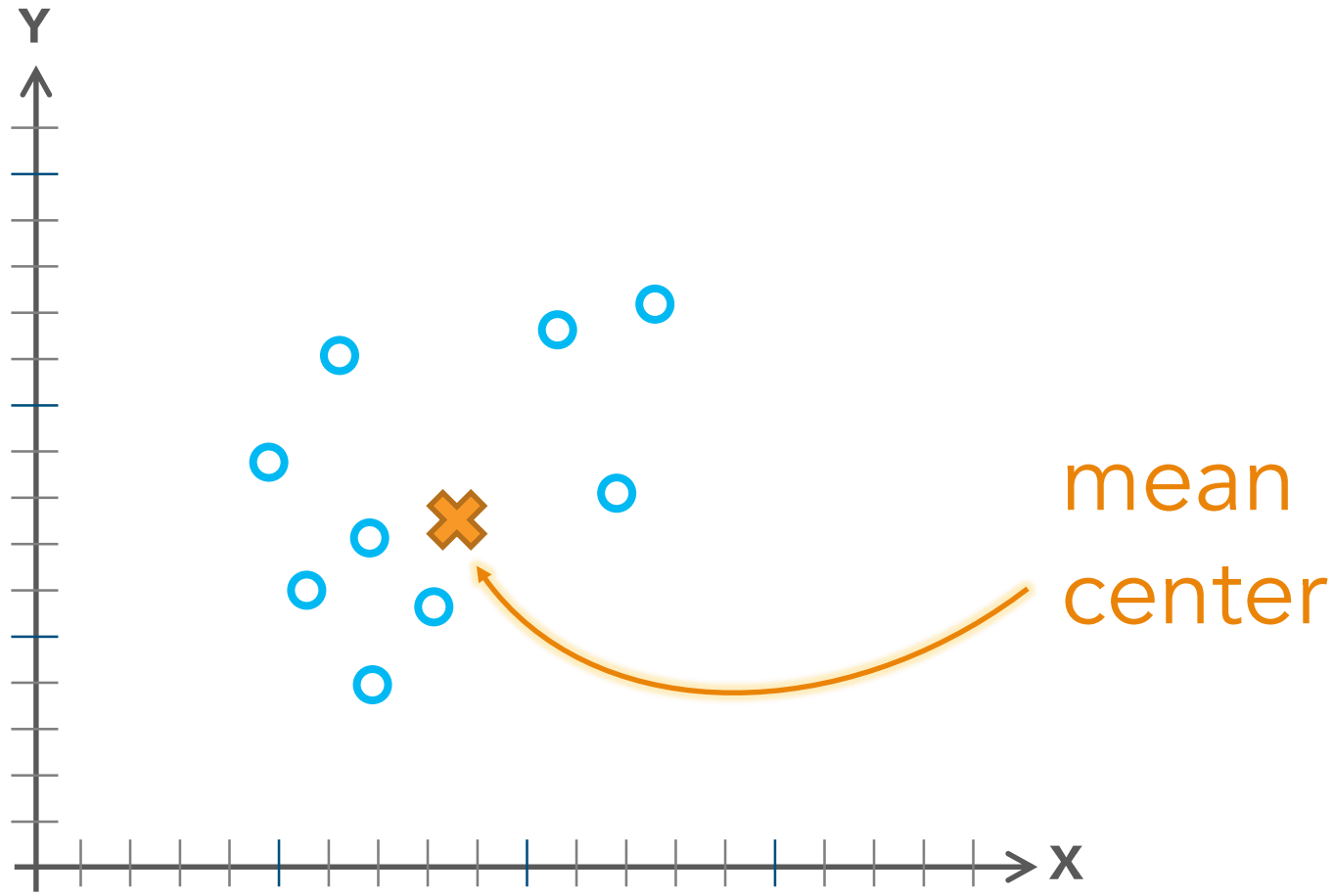
identifies the mean direction, length, and geographic center for a set of lines

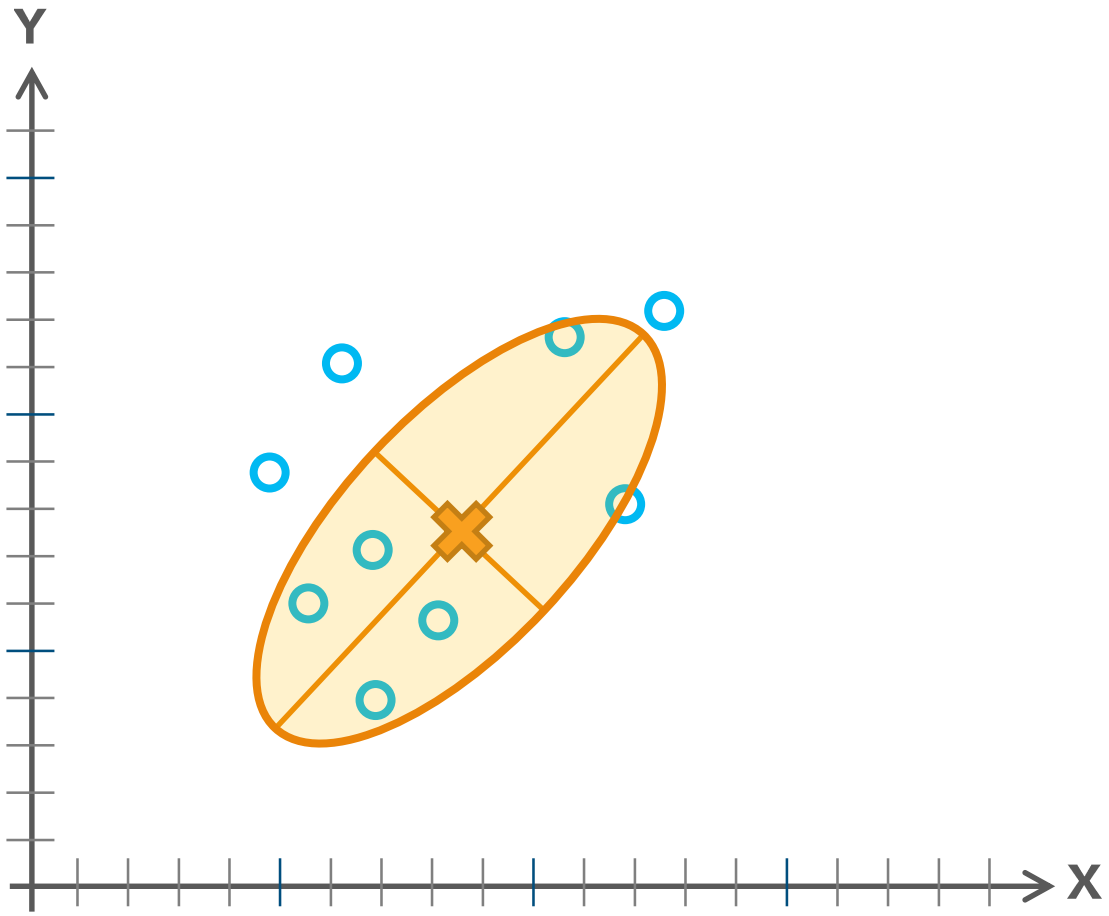

Y
X


Y
X

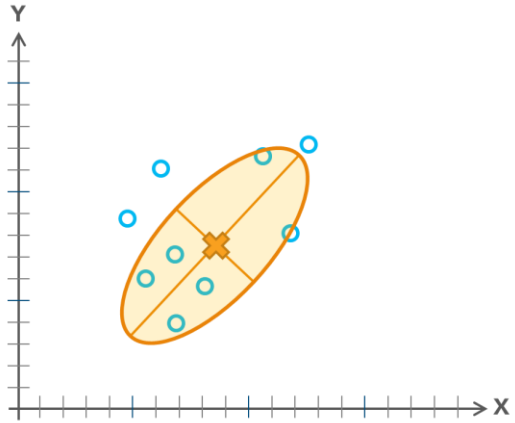
Directional Distribution (Standard Deviational Ellipse)

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


Y
X
mean
center


Y
X

Demo

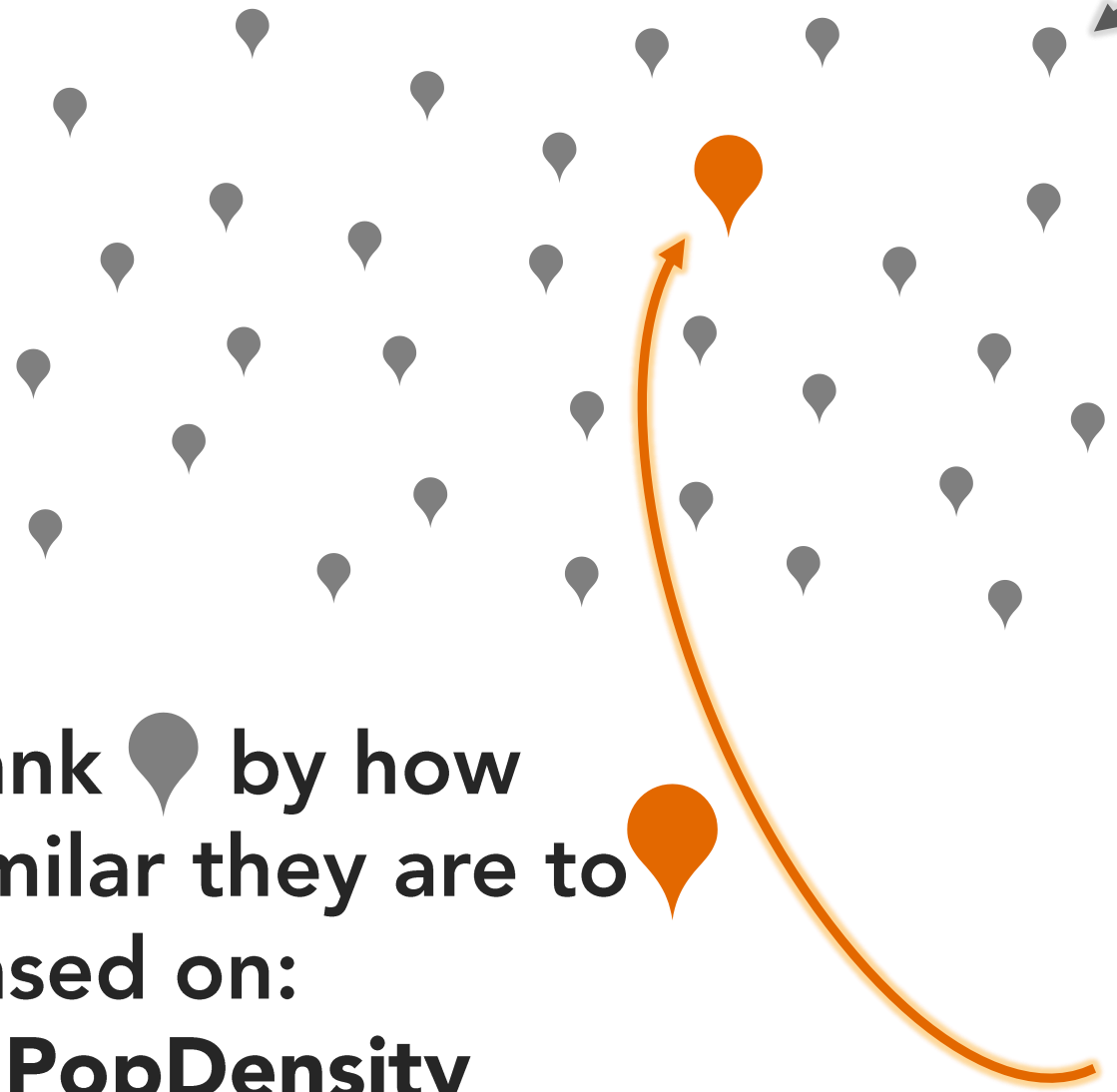


Similarity Search

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

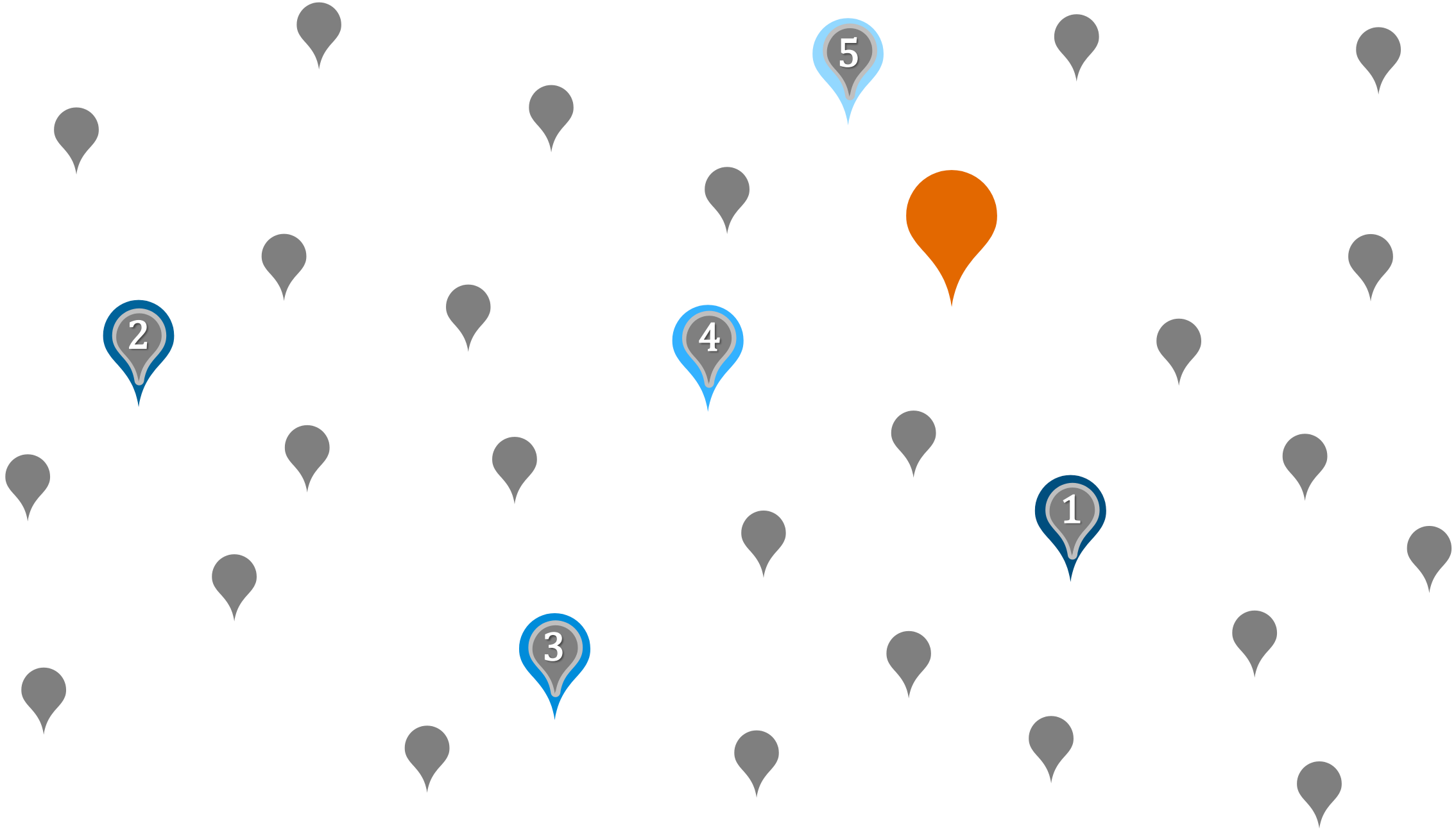


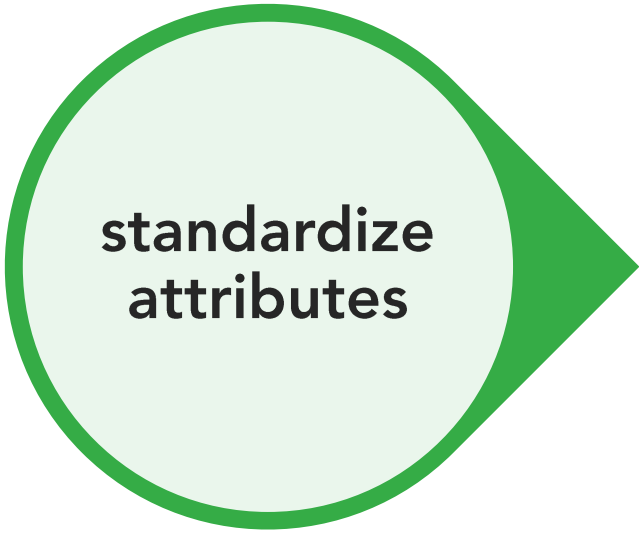
Rank by how similar they are to based on:

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

LocID	PopDensity	AvIncome	DistToCompetition

LocID	PopDensity	AvIncome	DistToCompetition


5
2
4
1
3



Z-transform:

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

standardize attributes

Population = 14,159

% Uninsured = .26

Distance (km) = 535.89

standardize attributes

Population = -.7932

% Uninsured = 3.8462

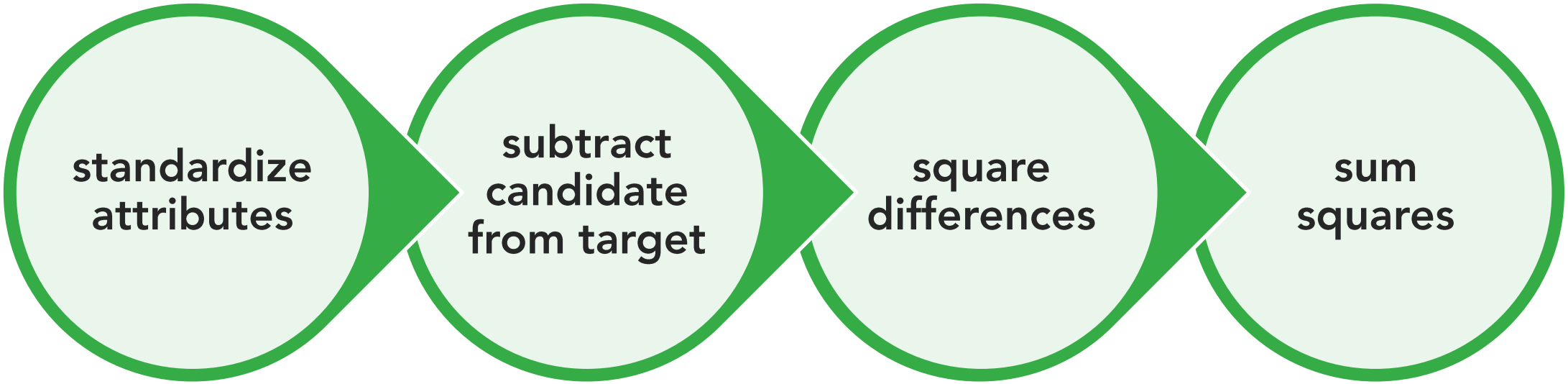
Distance (km) = .6433

standardize attributes

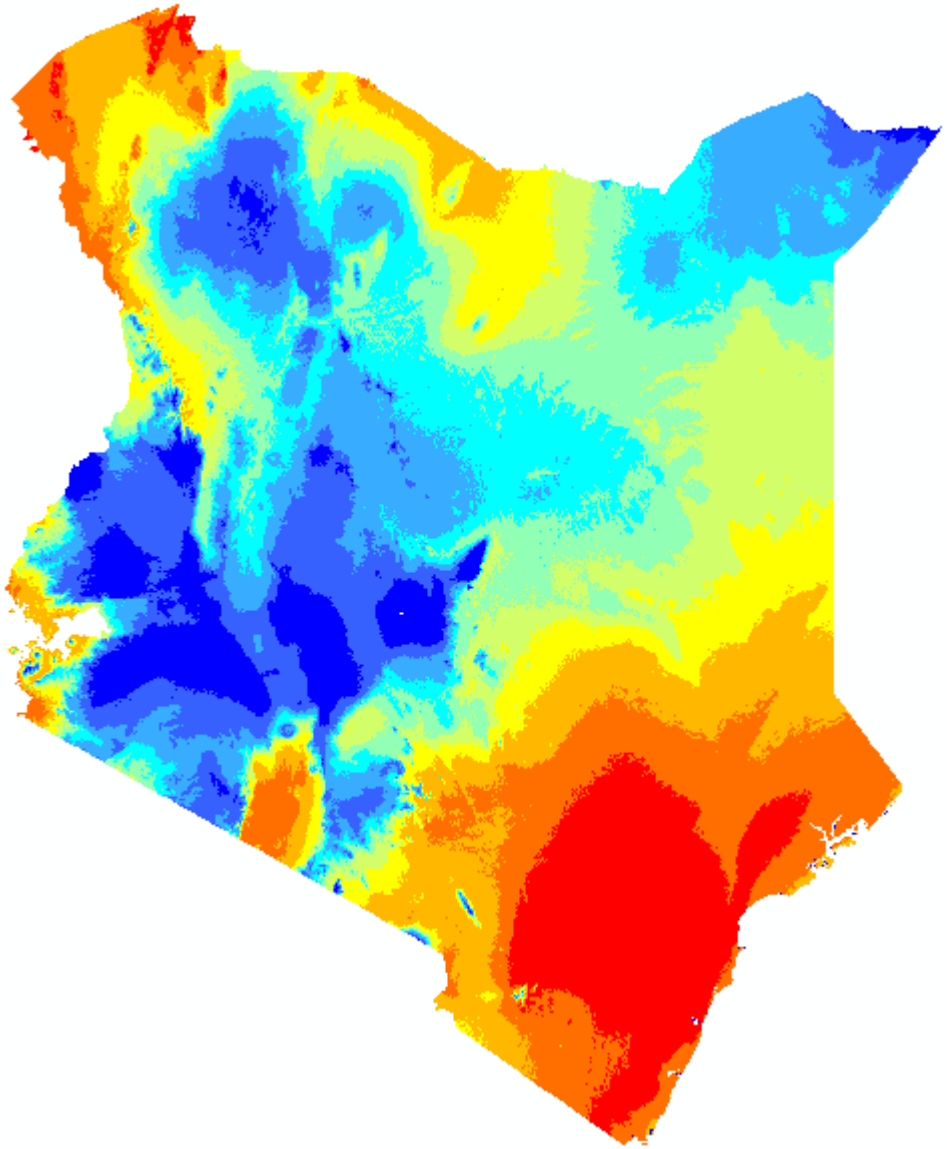
Population = -.7932

% Uninsured = 3.8462

Distance (km) = .6433


standardize attributes
subtract candidate from target
square differences
sum squares

Dengue Fever Risk in Kenya

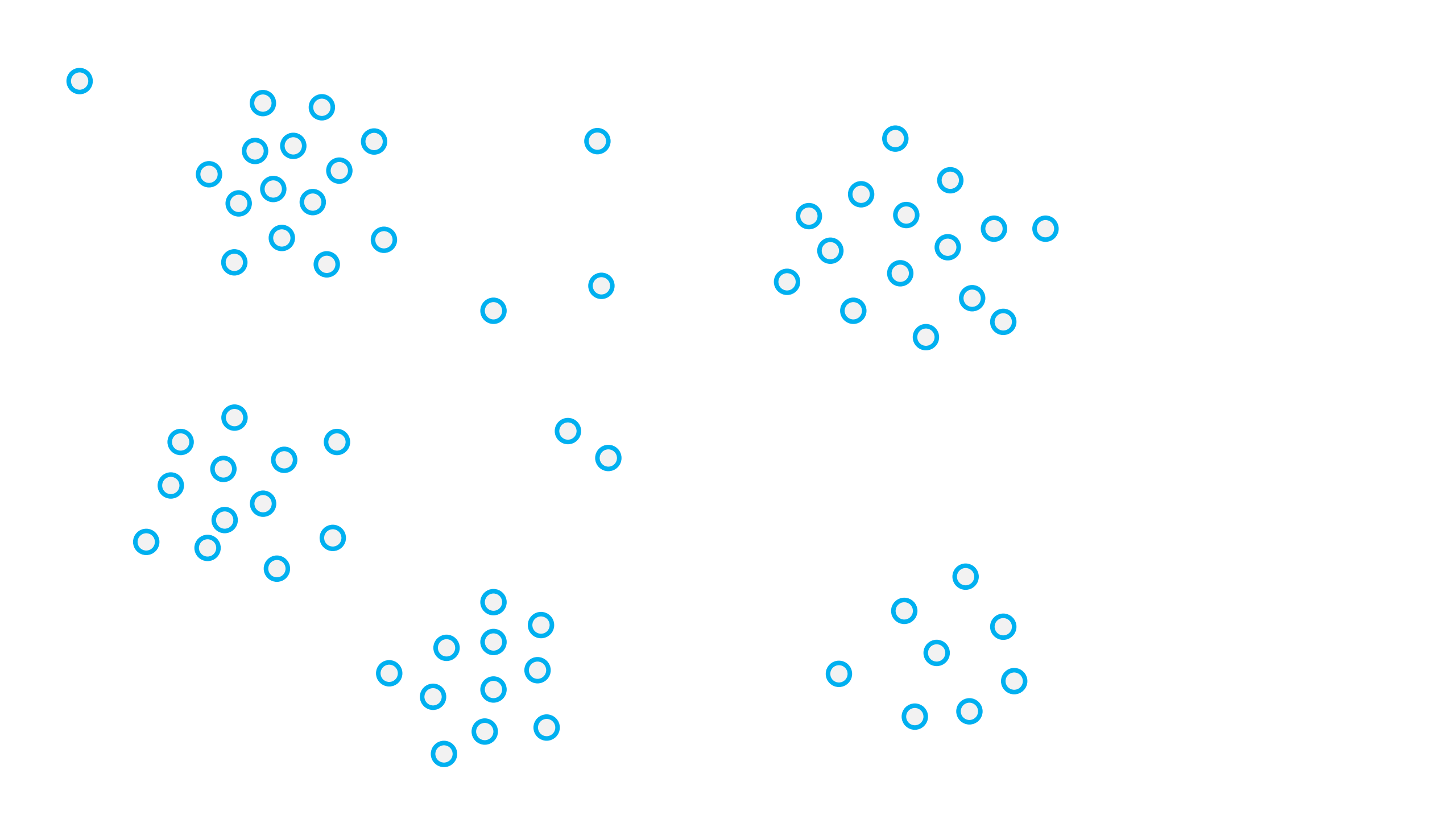


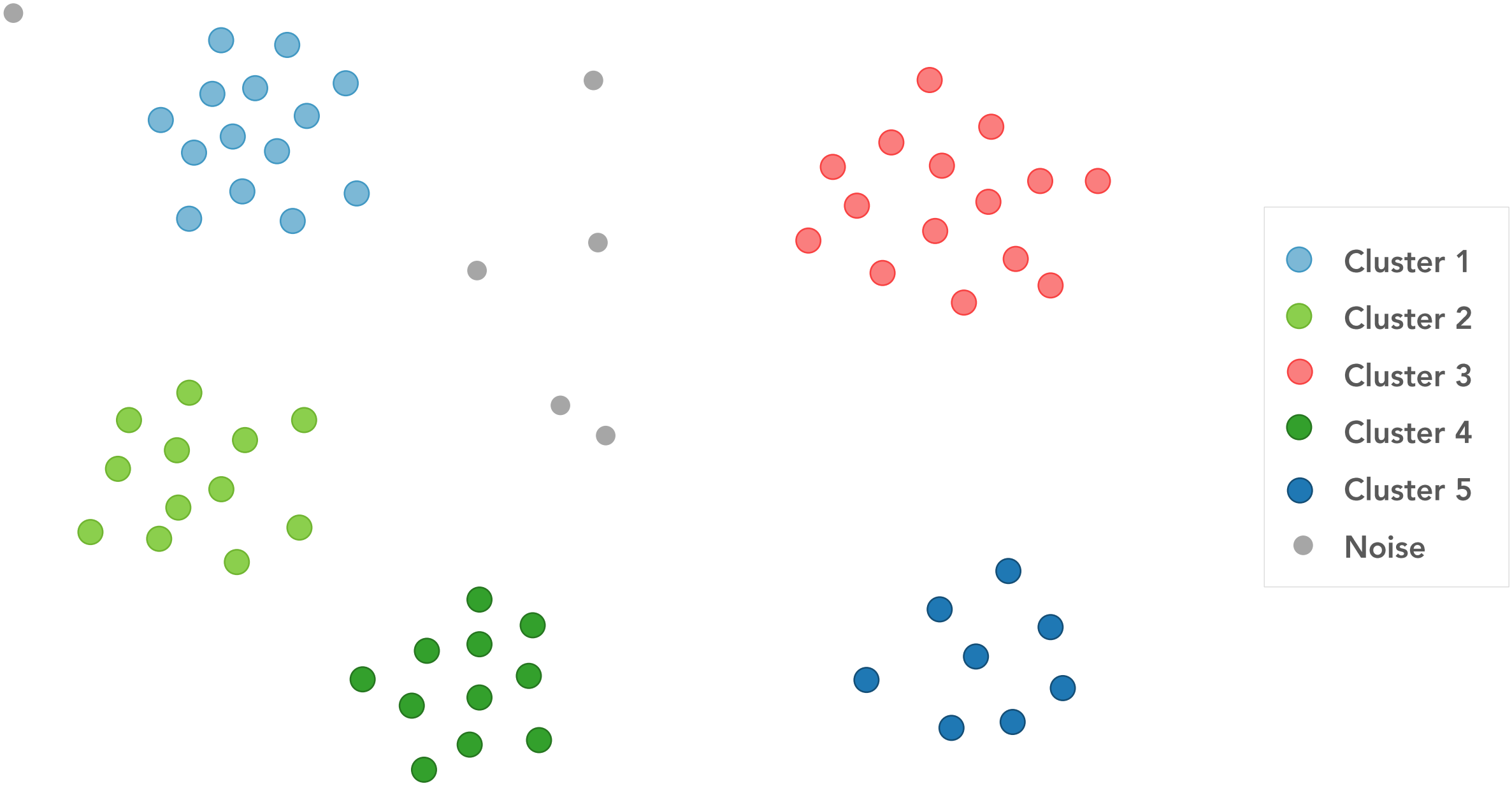
Machine Learning

clustering methods

Density-based Clustering

finds clusters based on feature locations



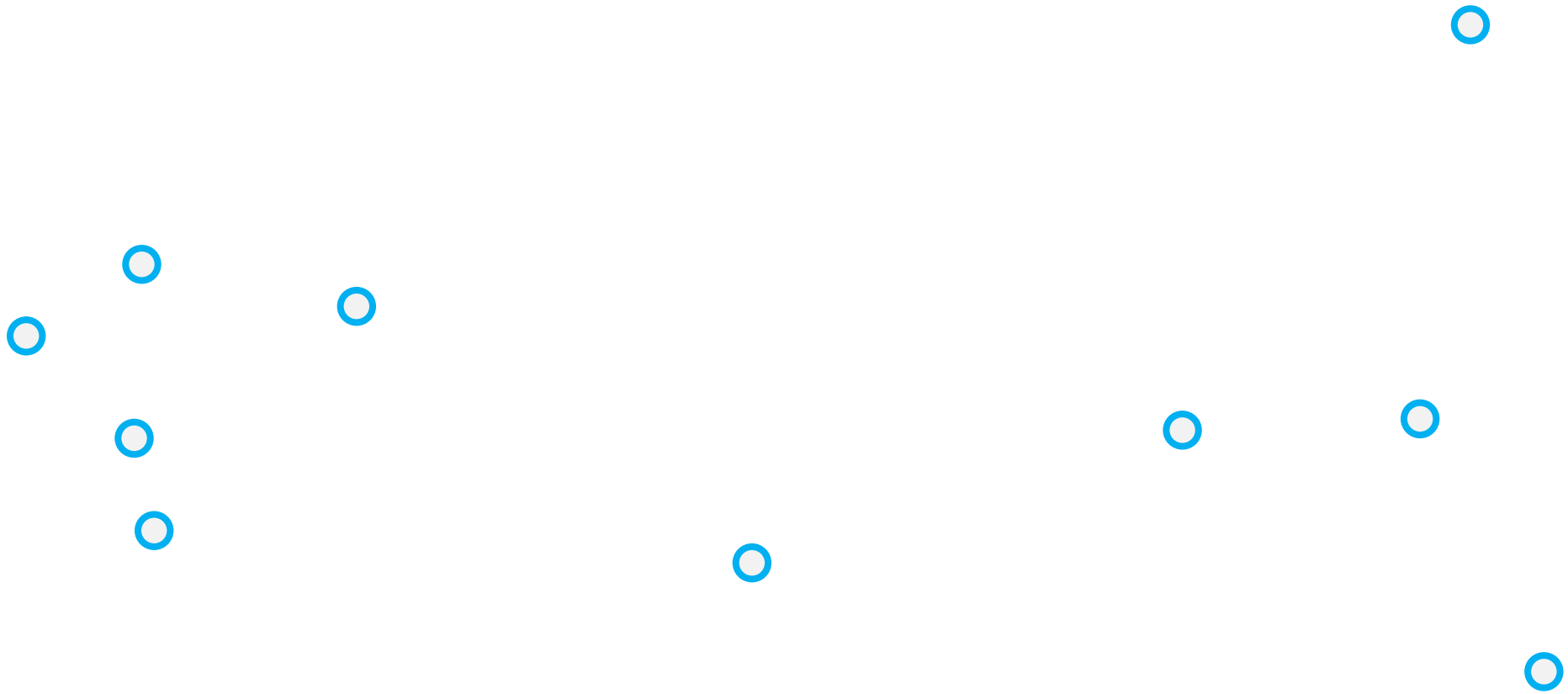

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Noise

DBSCAN – defined distance

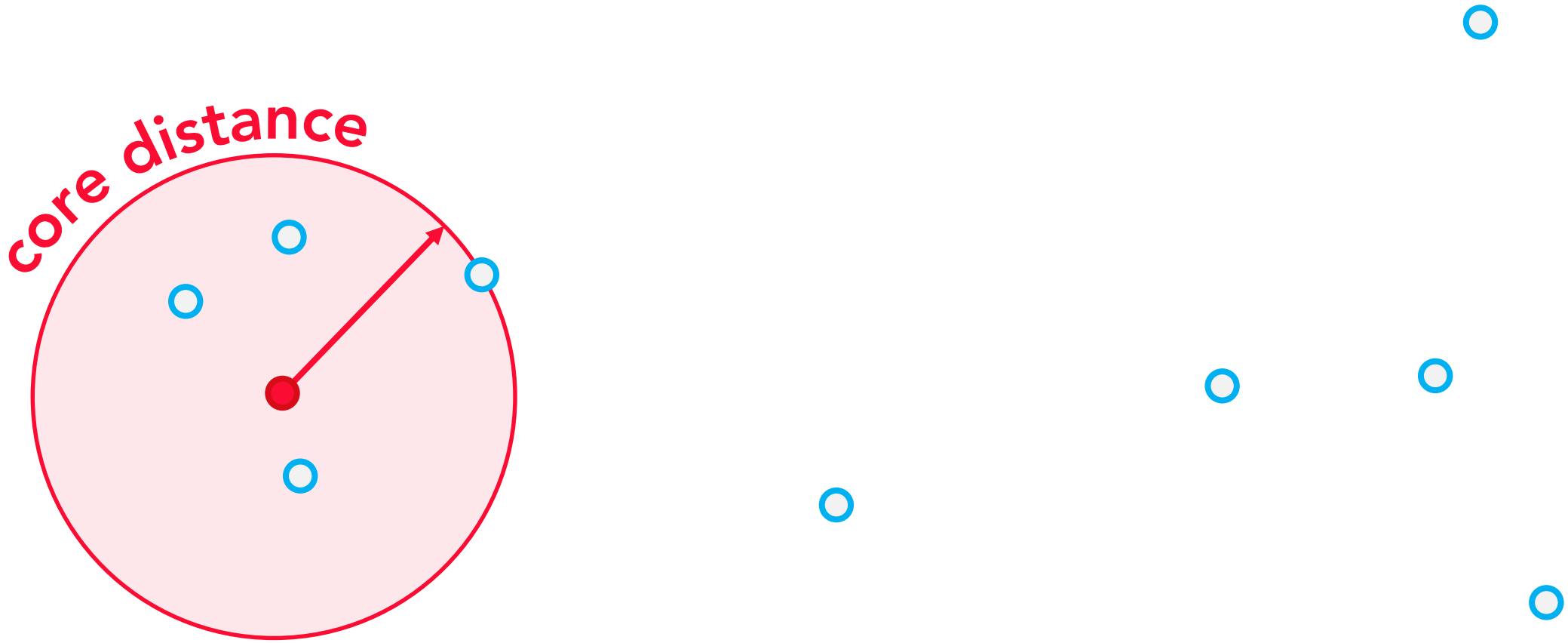
HDBSCAN – self adjusting

OPTICS – multi-scale

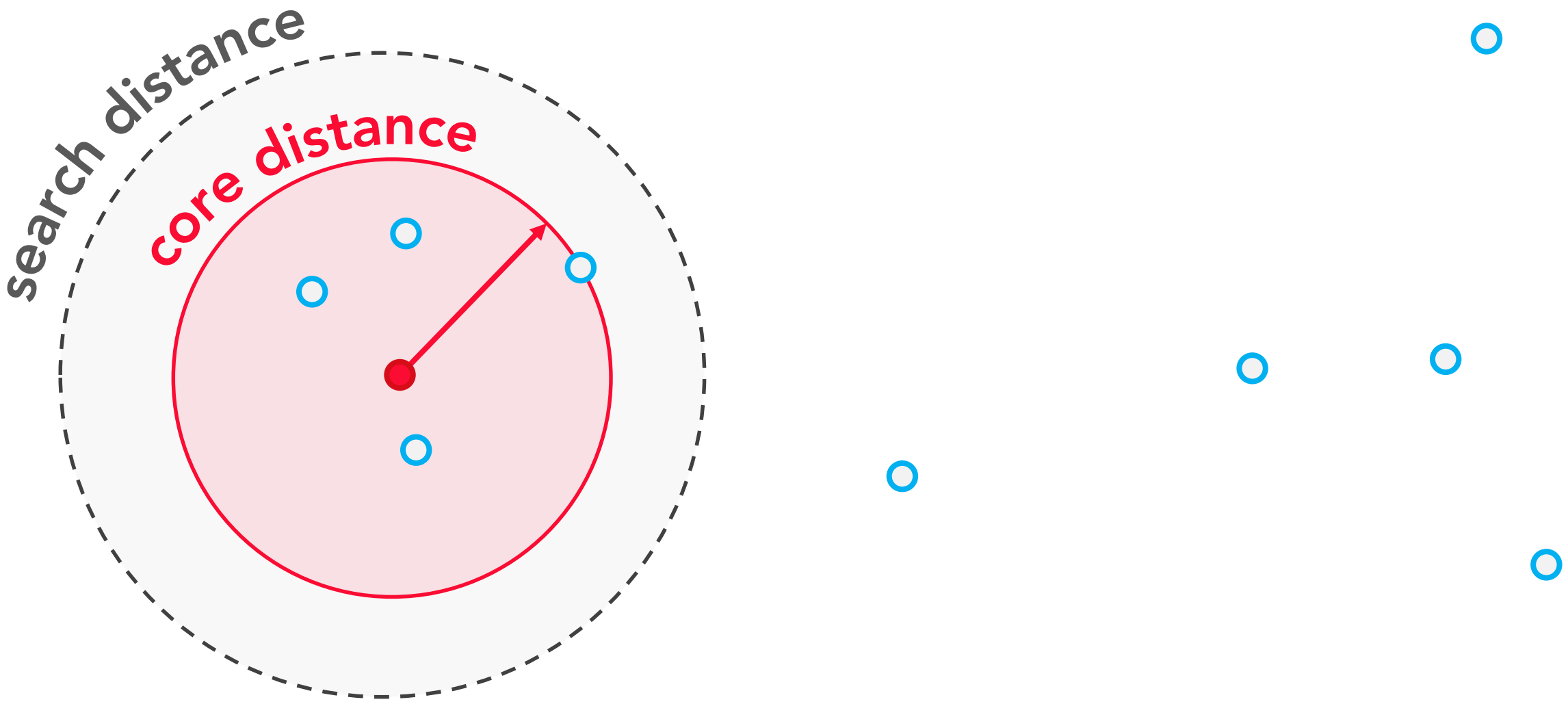
DBSCAN – defined distance



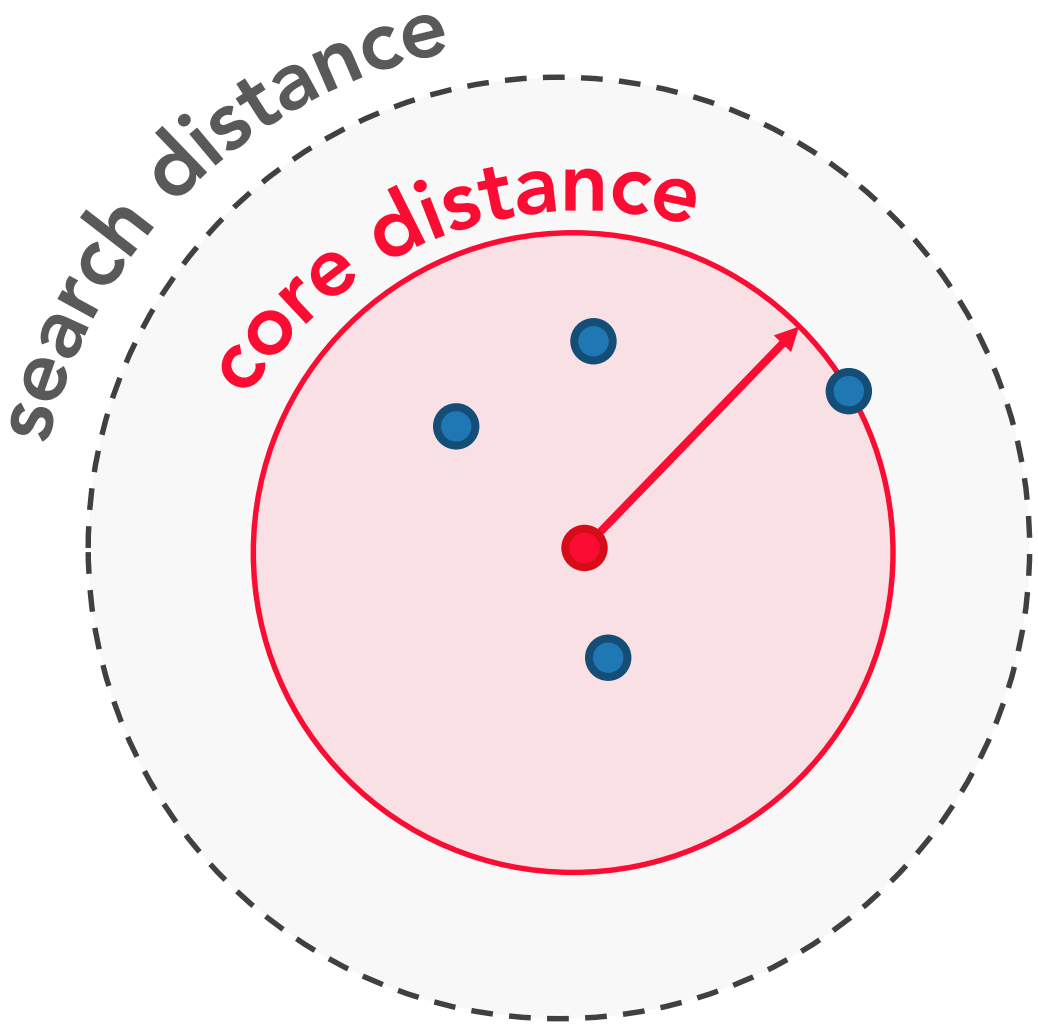
DBSCAN – defined distance



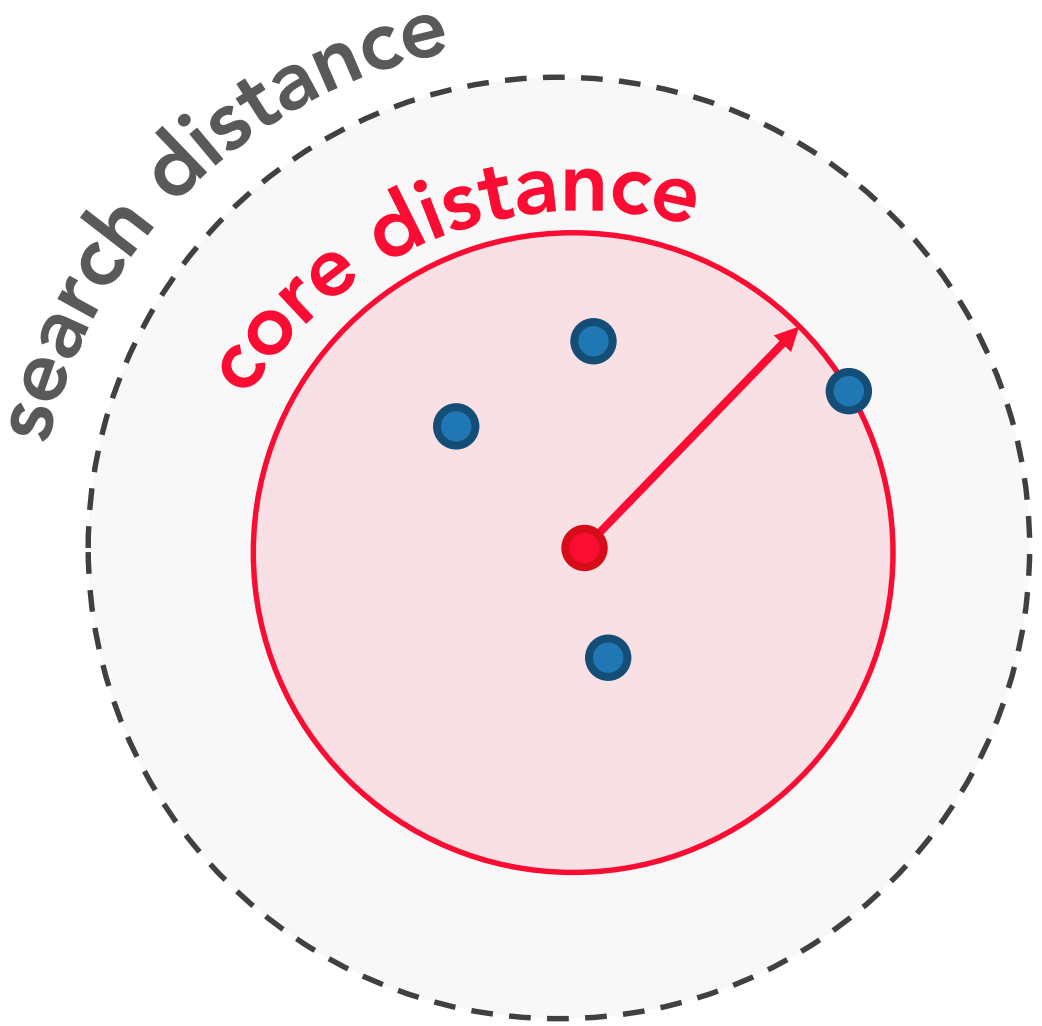
DBSCAN – defined distance

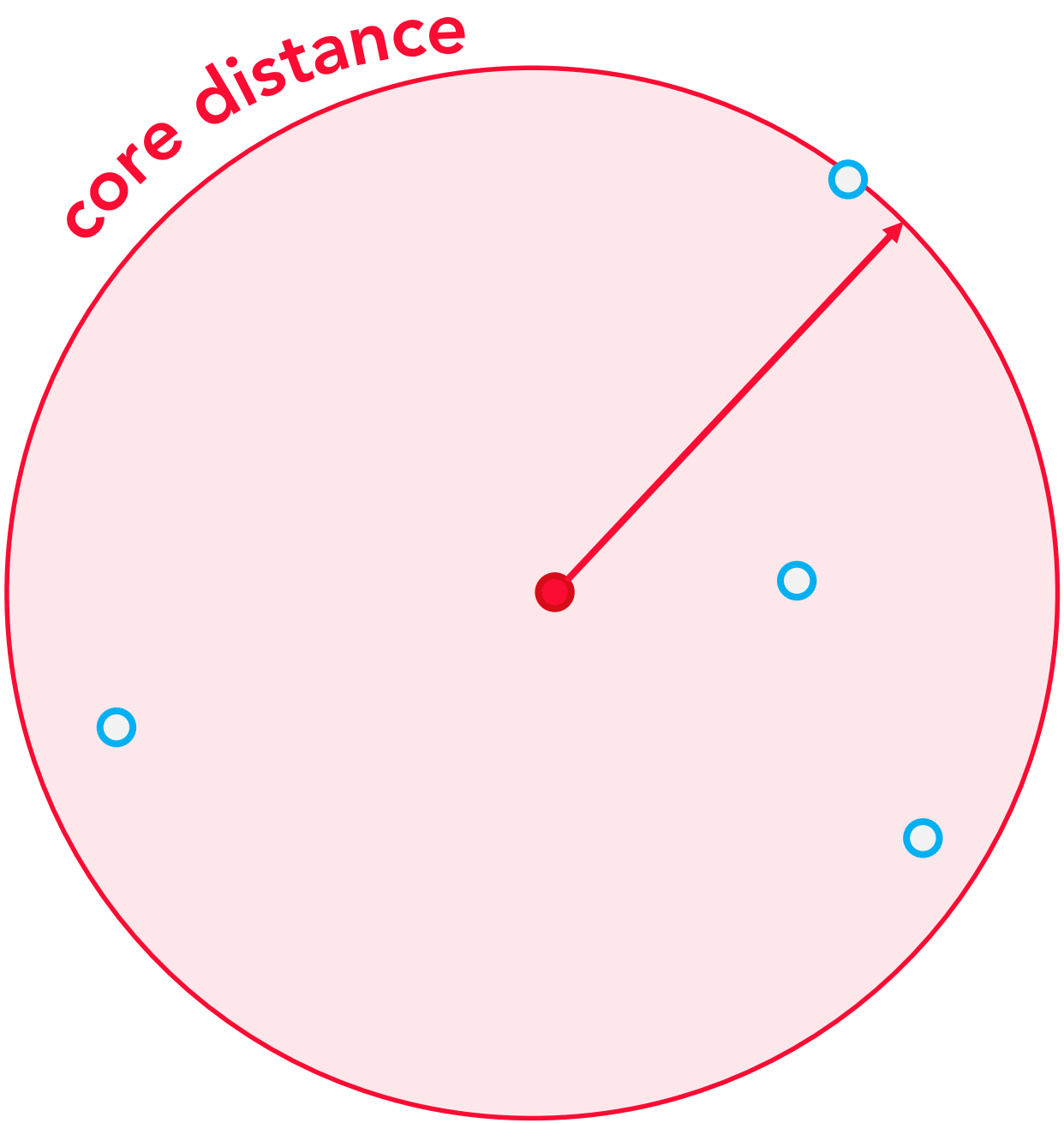


DBSCAN – defined distance

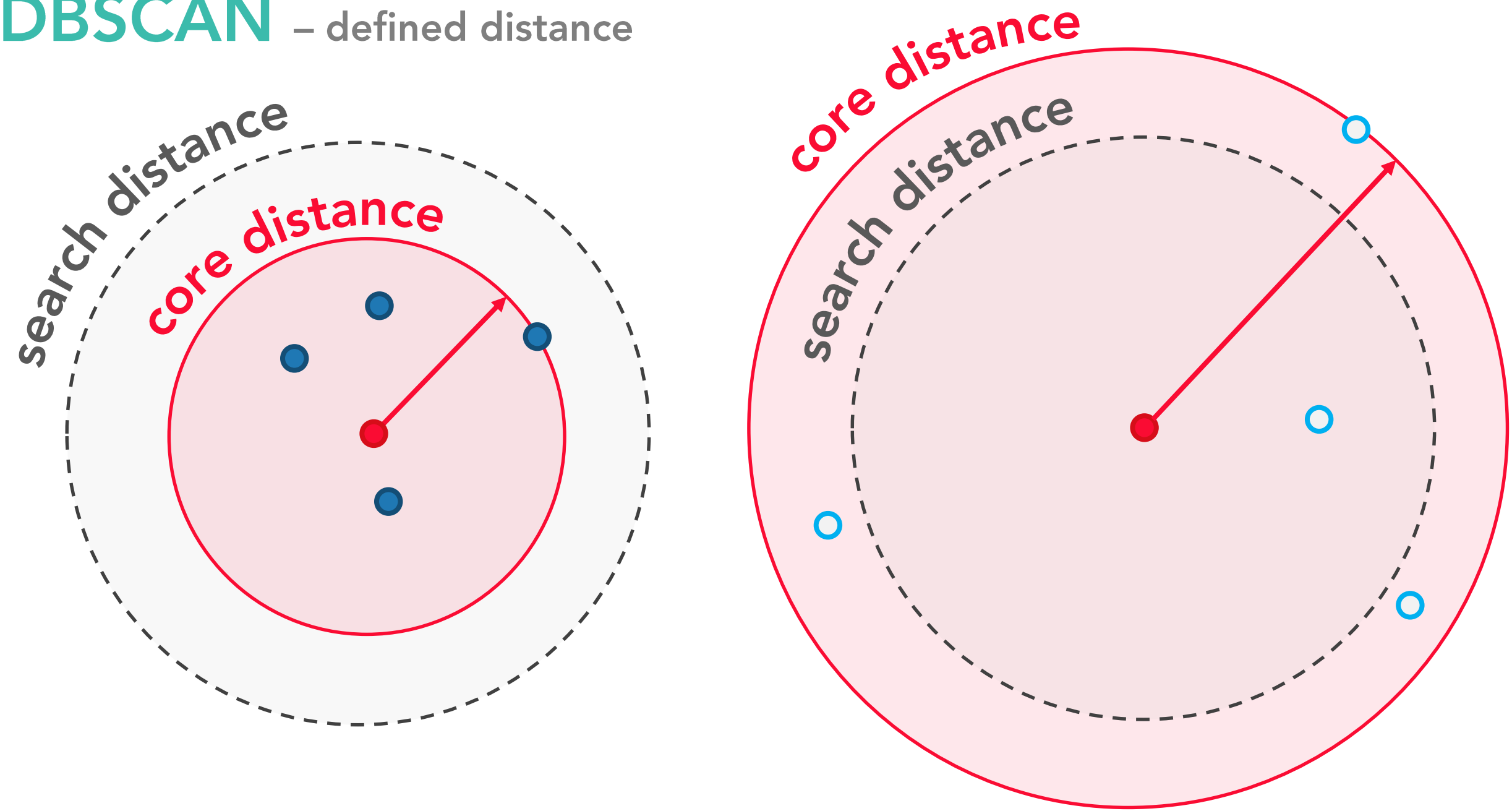


DBSCAN – defined distance

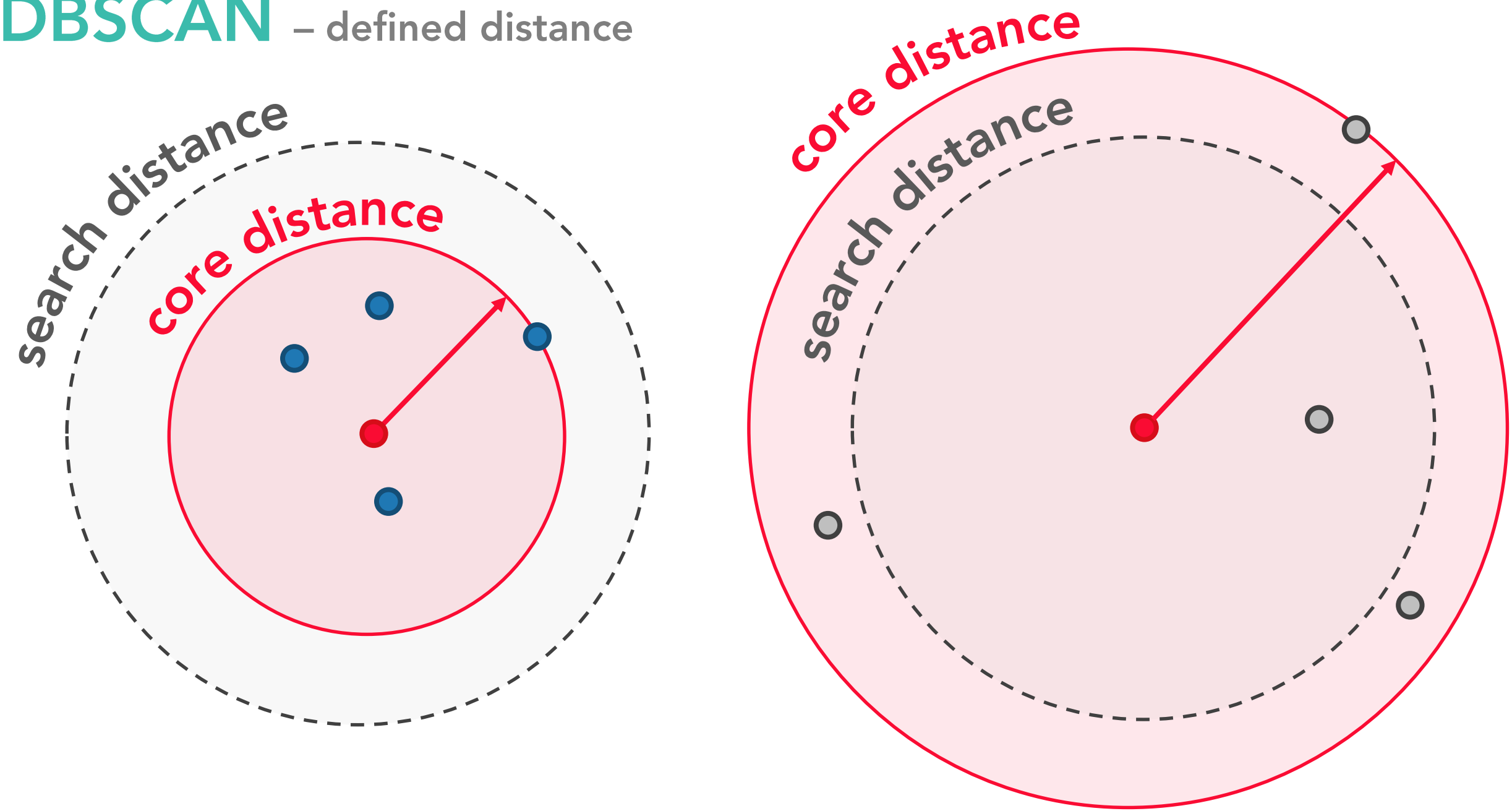




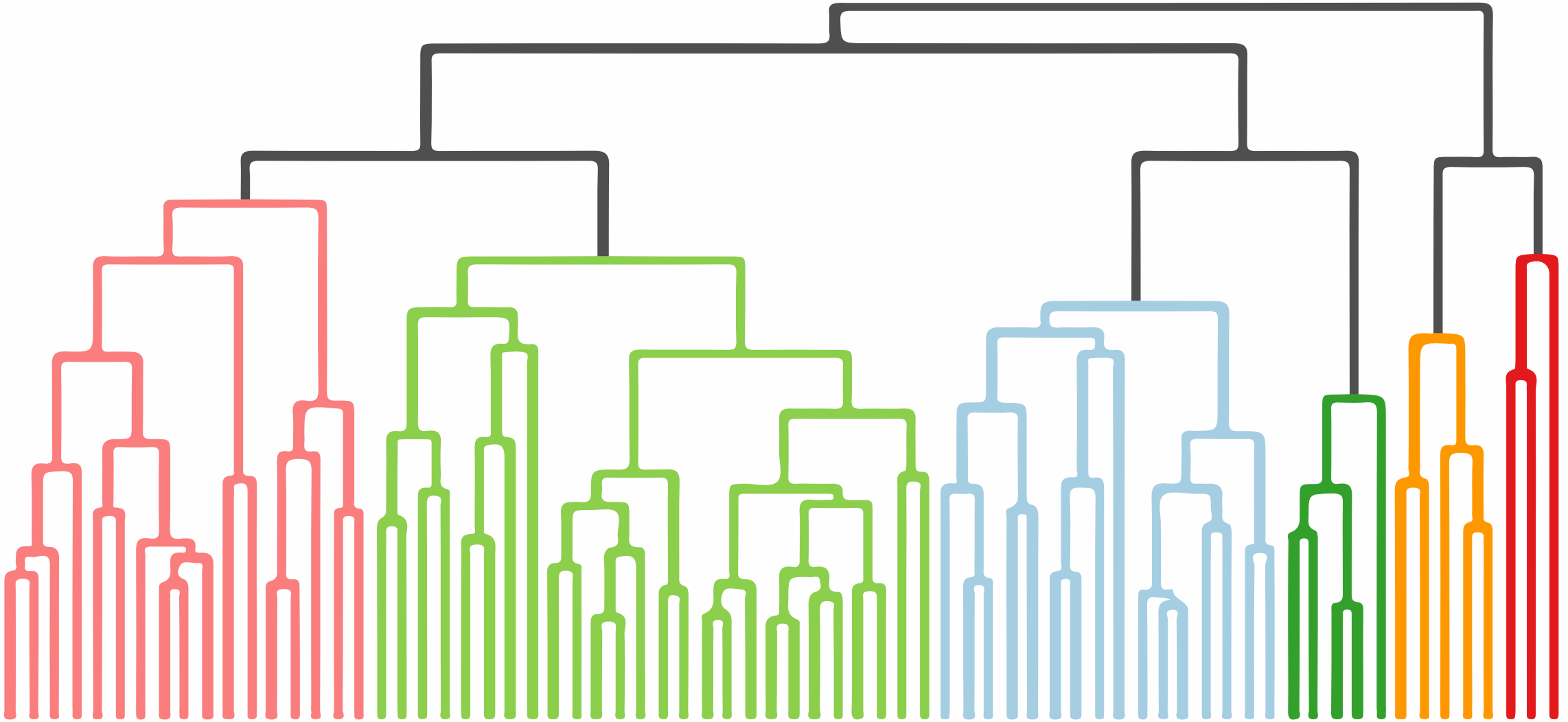
DBSCAN – defined distance



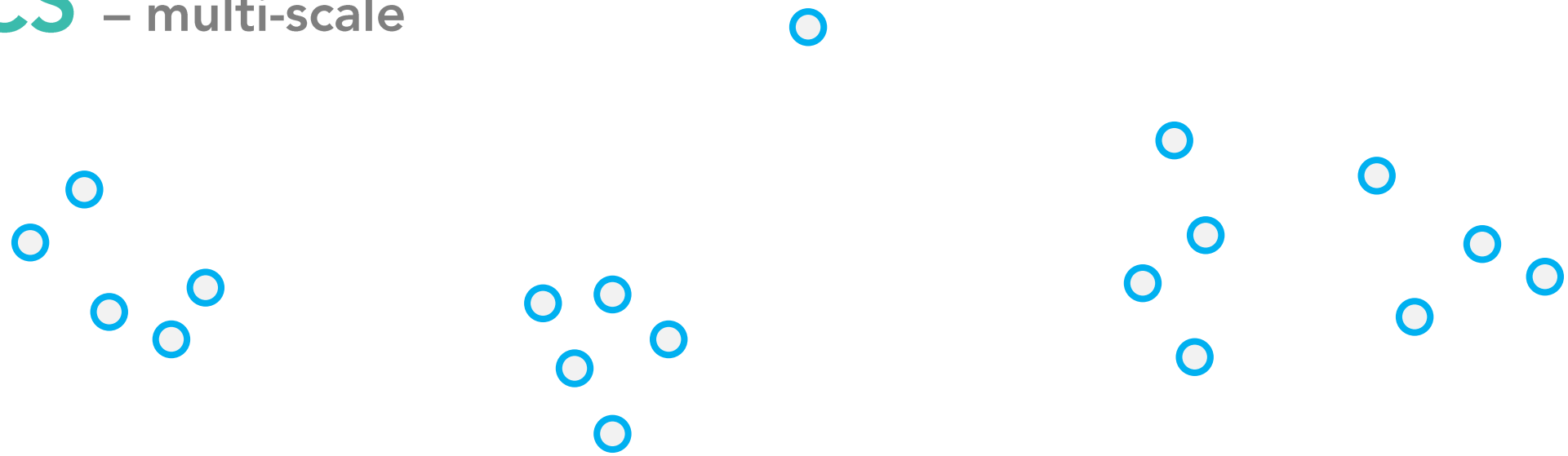
DBSCAN – defined distance



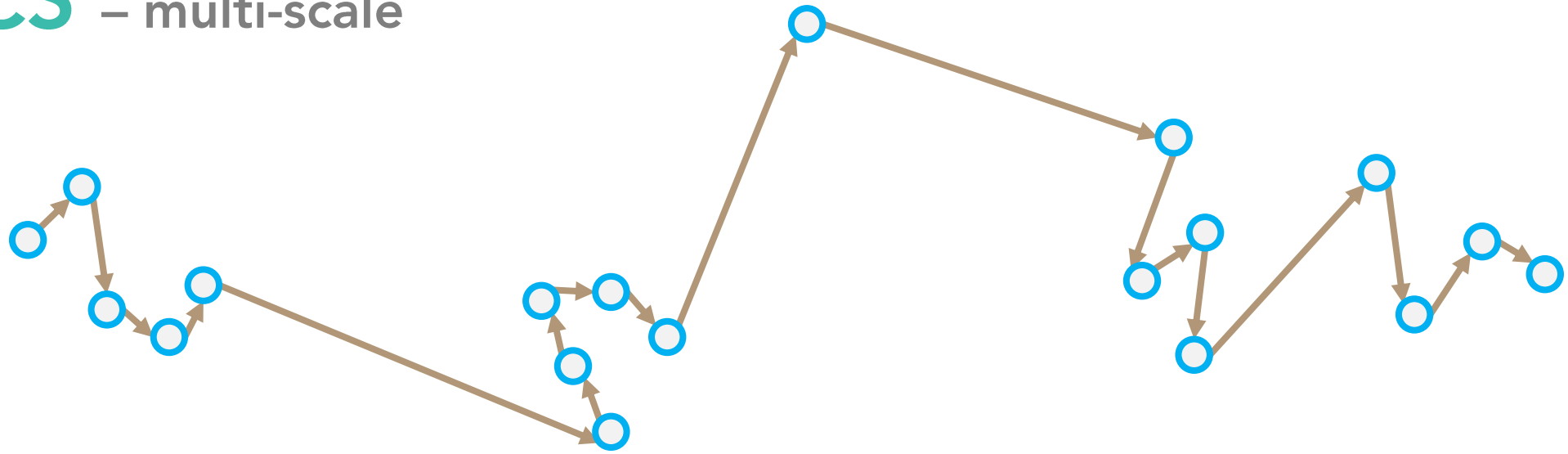
HDBSCAN – self adjusting



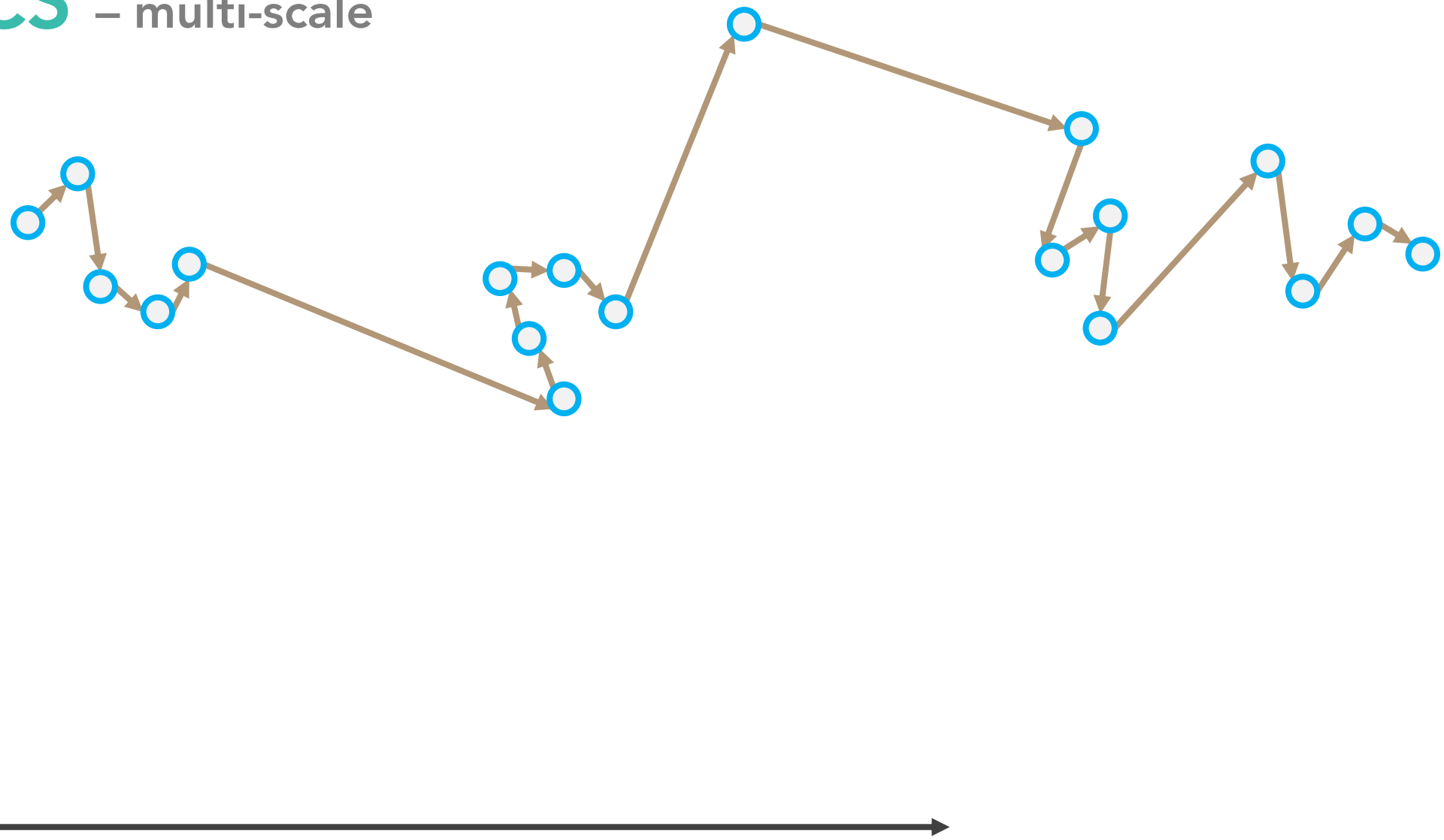
OPTICS – multi-scale



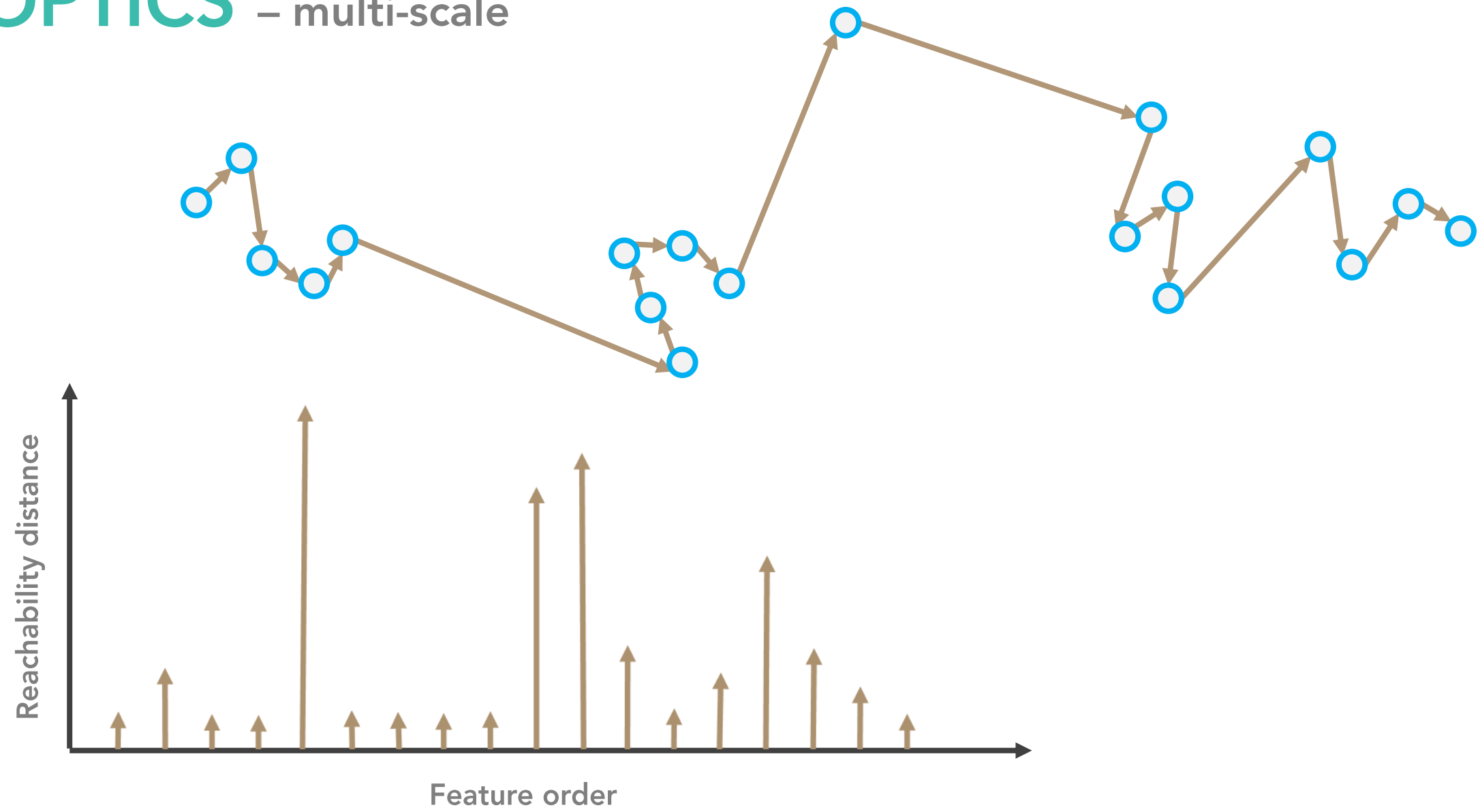
OPTICS – multi-scale



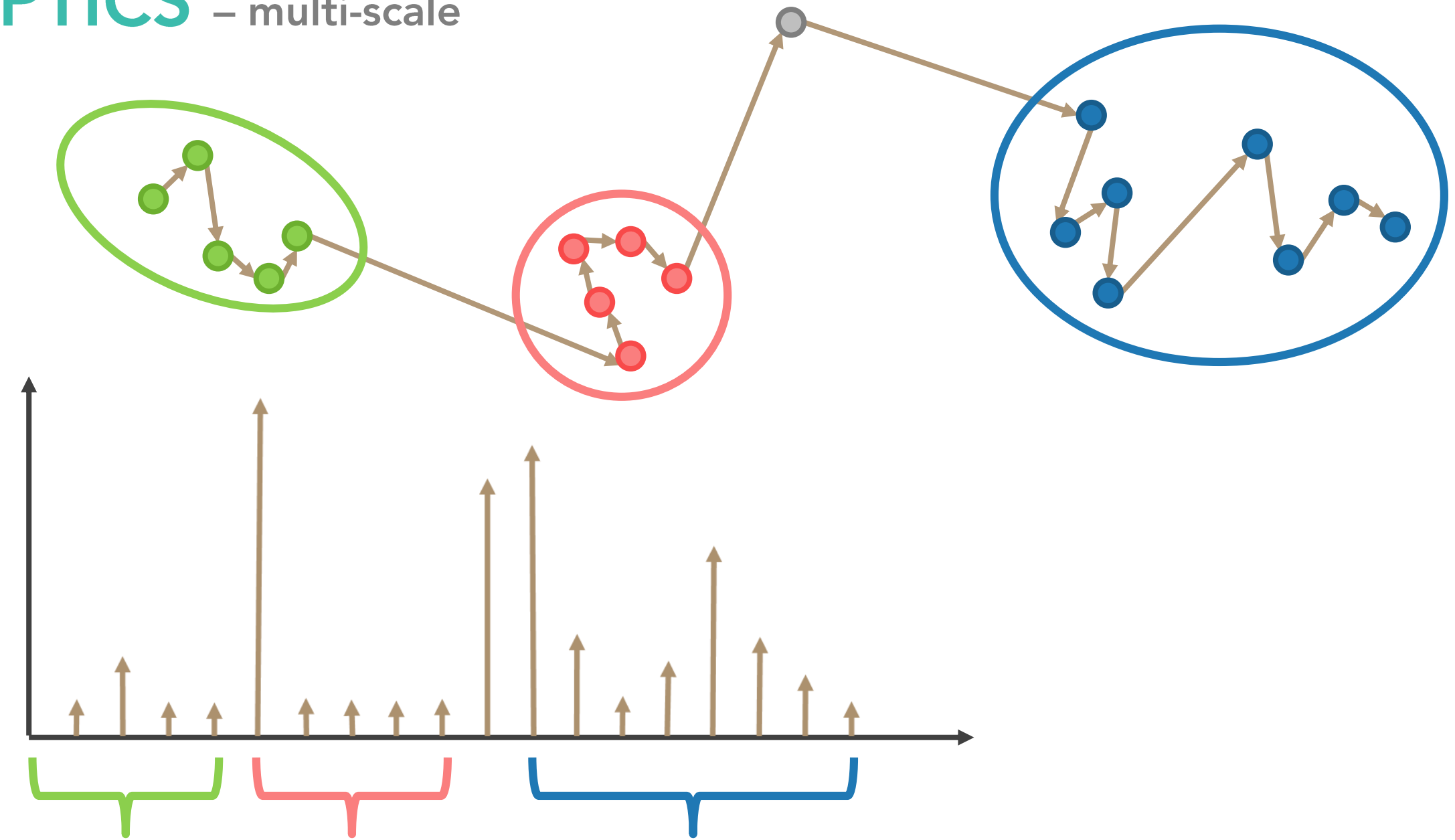
OPTICS – multi-scale



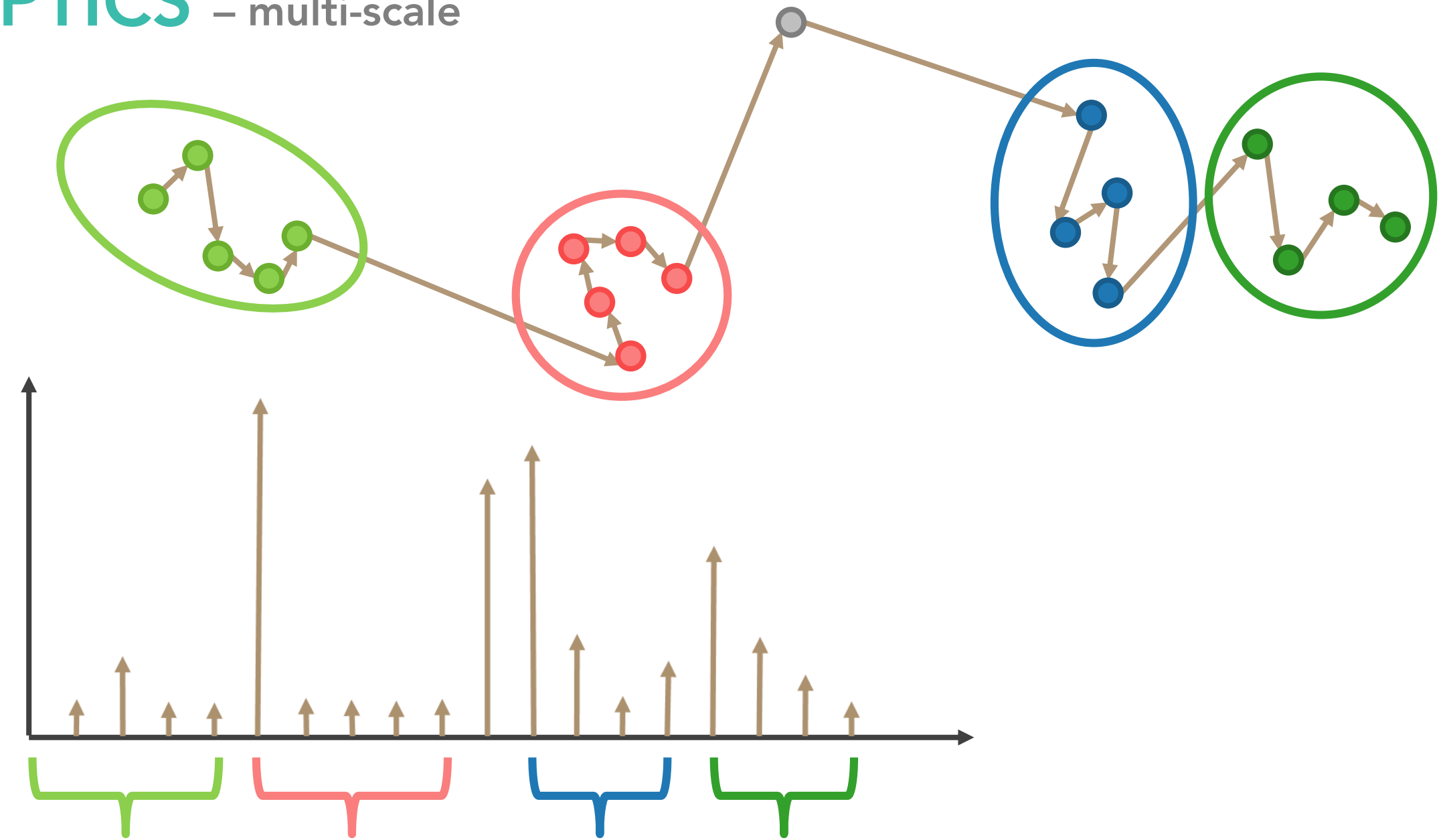
OPTICS – multi-scale



OPTICS – multi-scale



OPTICS – multi-scale



DBSCAN

- Uses fixed search distance
- Clusters of similar densities
- Fast

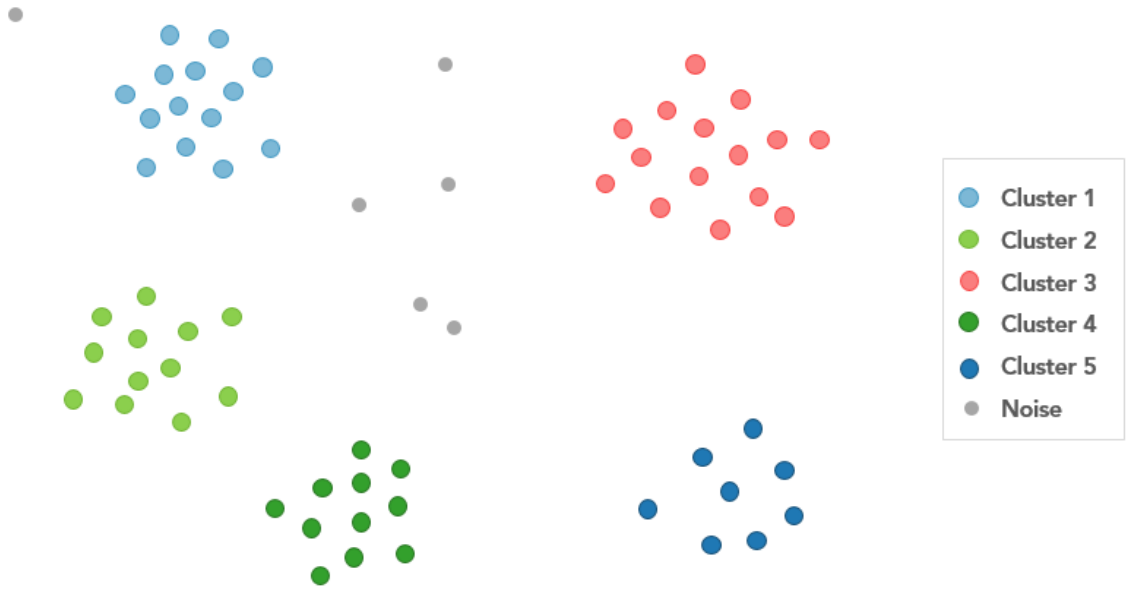
HDBSCAN

- Uses range of search distances to find clusters of varying densities
- Data driven, requires least user input

OPTICS

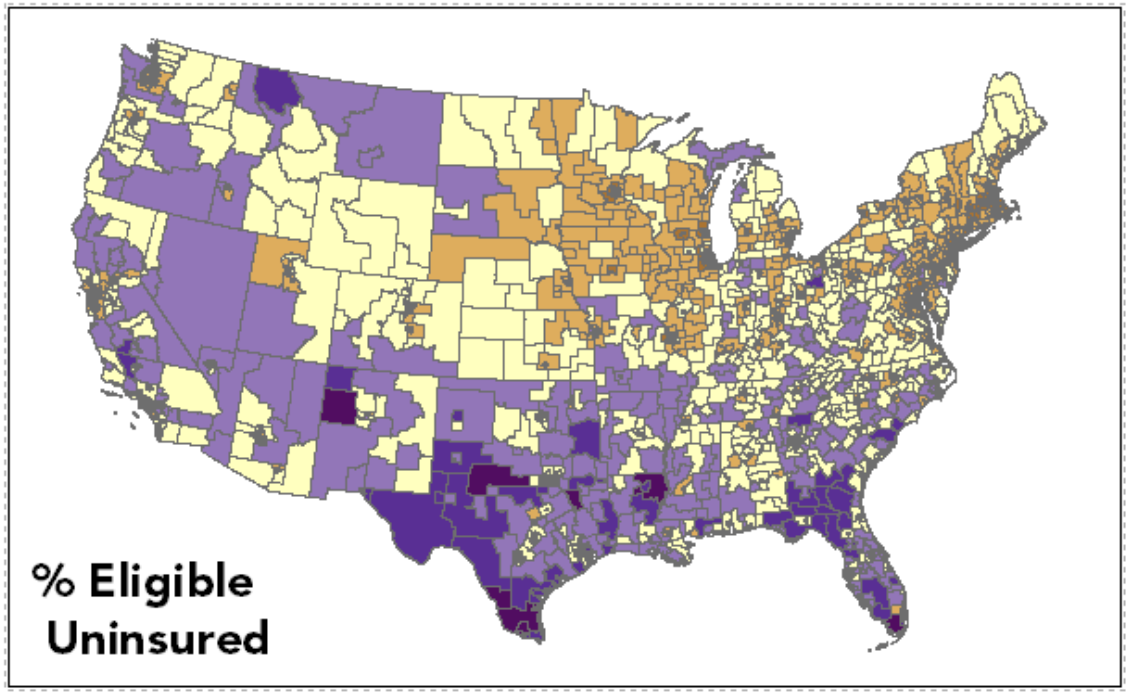
- Uses neighbor distances to create reachability plot
- Most flexibility for fine tuning
- Can be computationally intensive

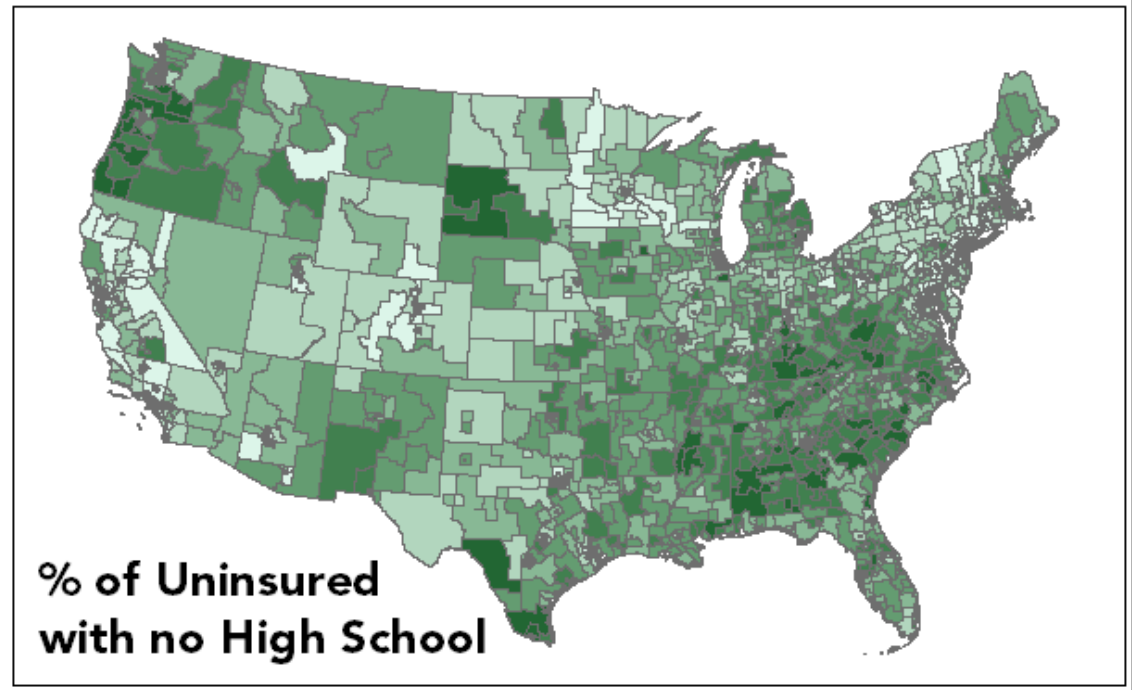
Demo

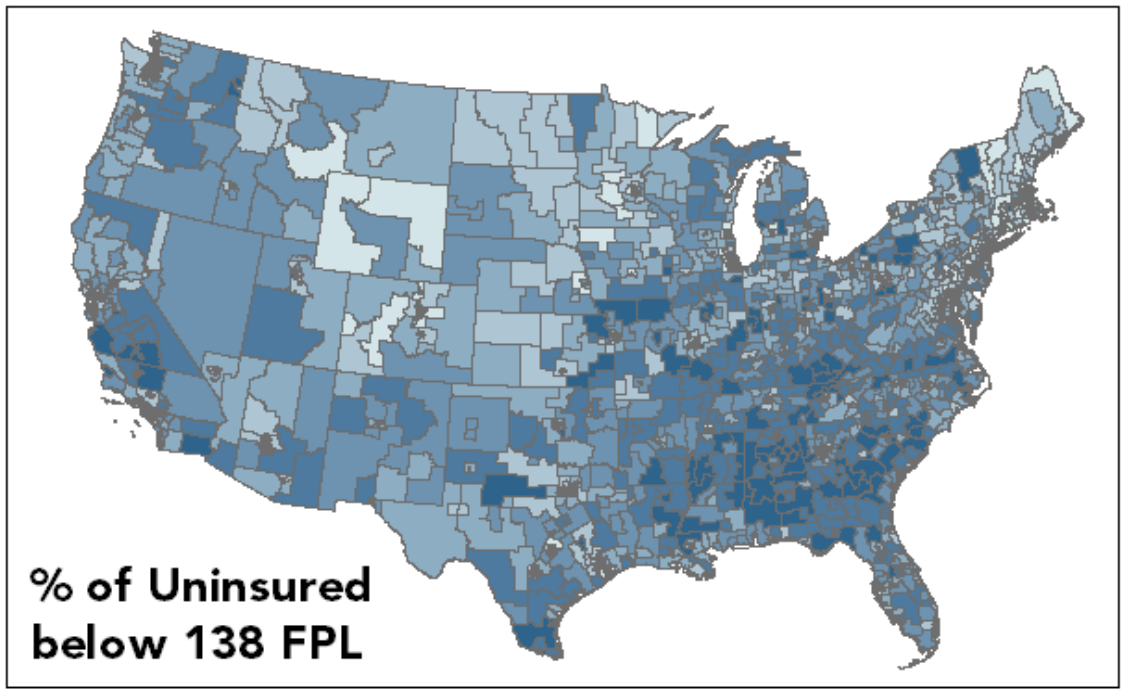


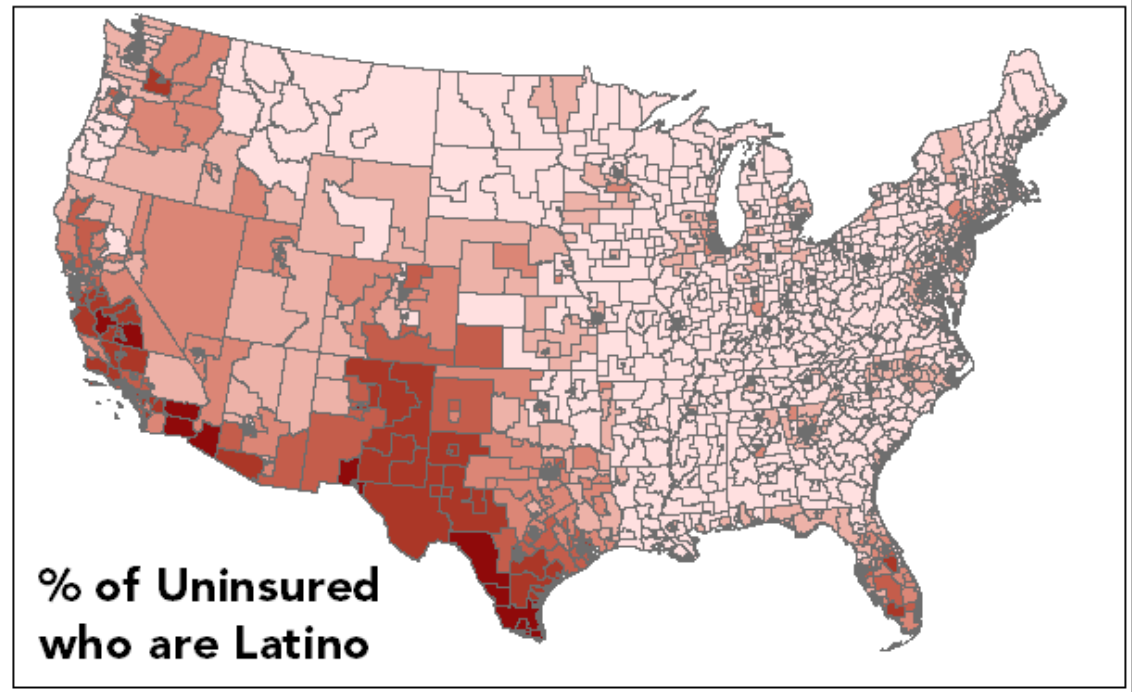
Multivariate Clustering

finds clusters based on feature attributes

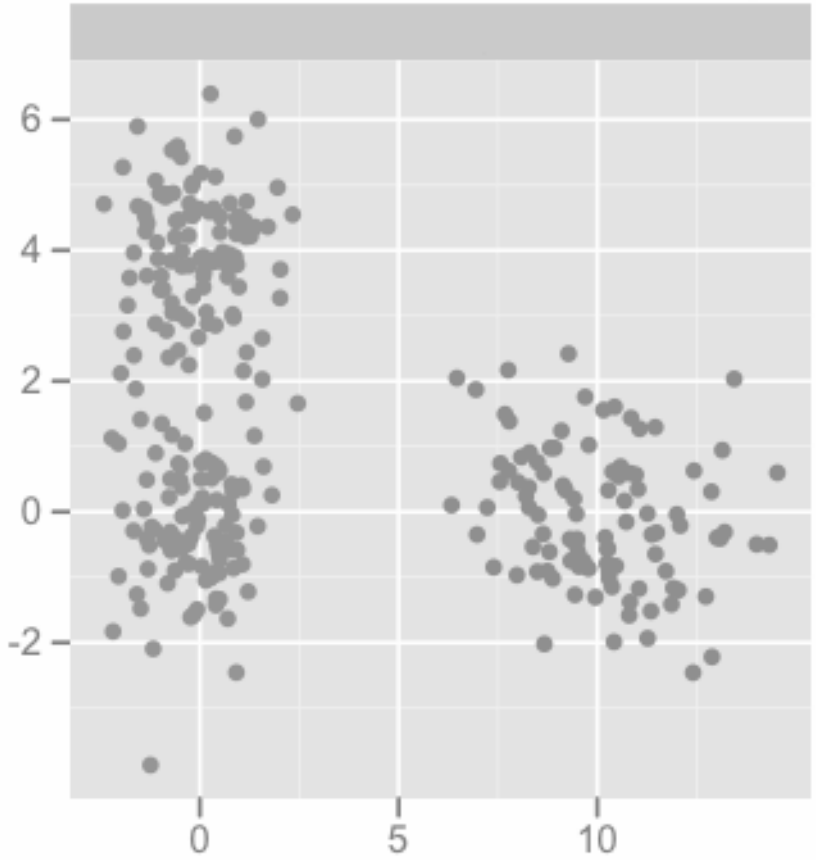

% Eligible
Uninsured


% of Uninsured
with no High School


% of Uninsured
below 138 FPL

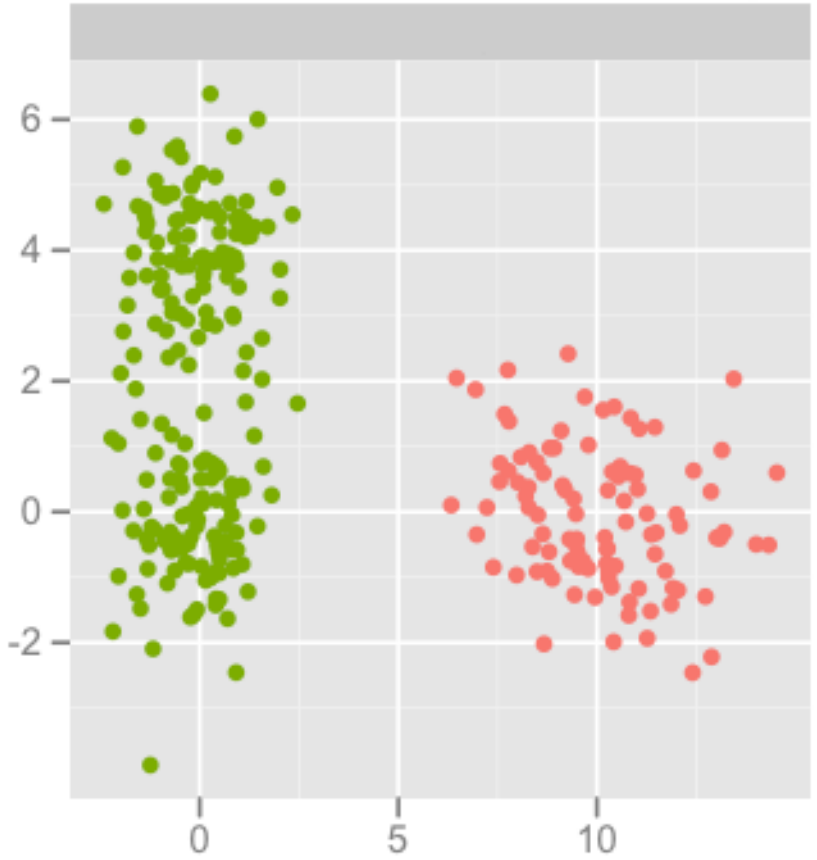

% of Uninsured
who are Latino

K Means

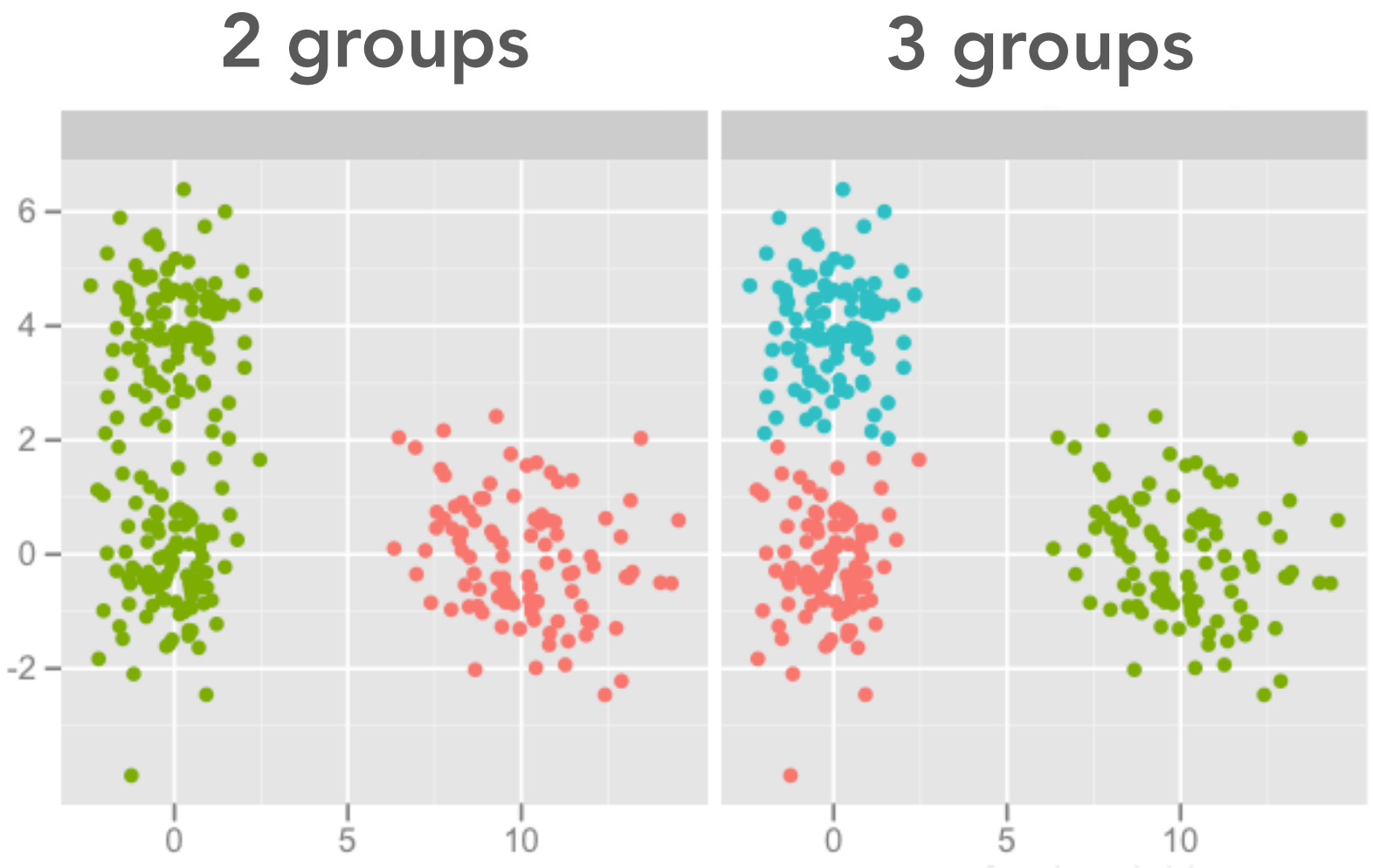


K Means

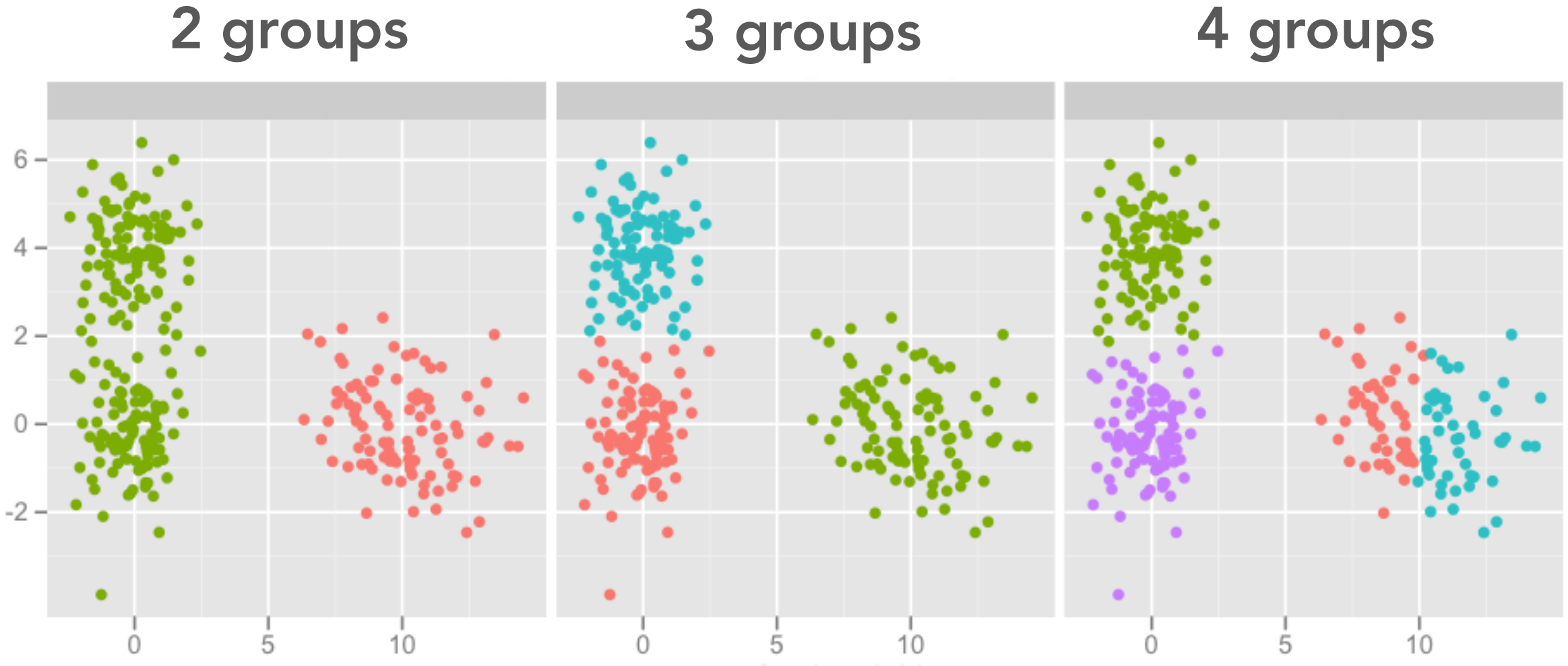
2 groups



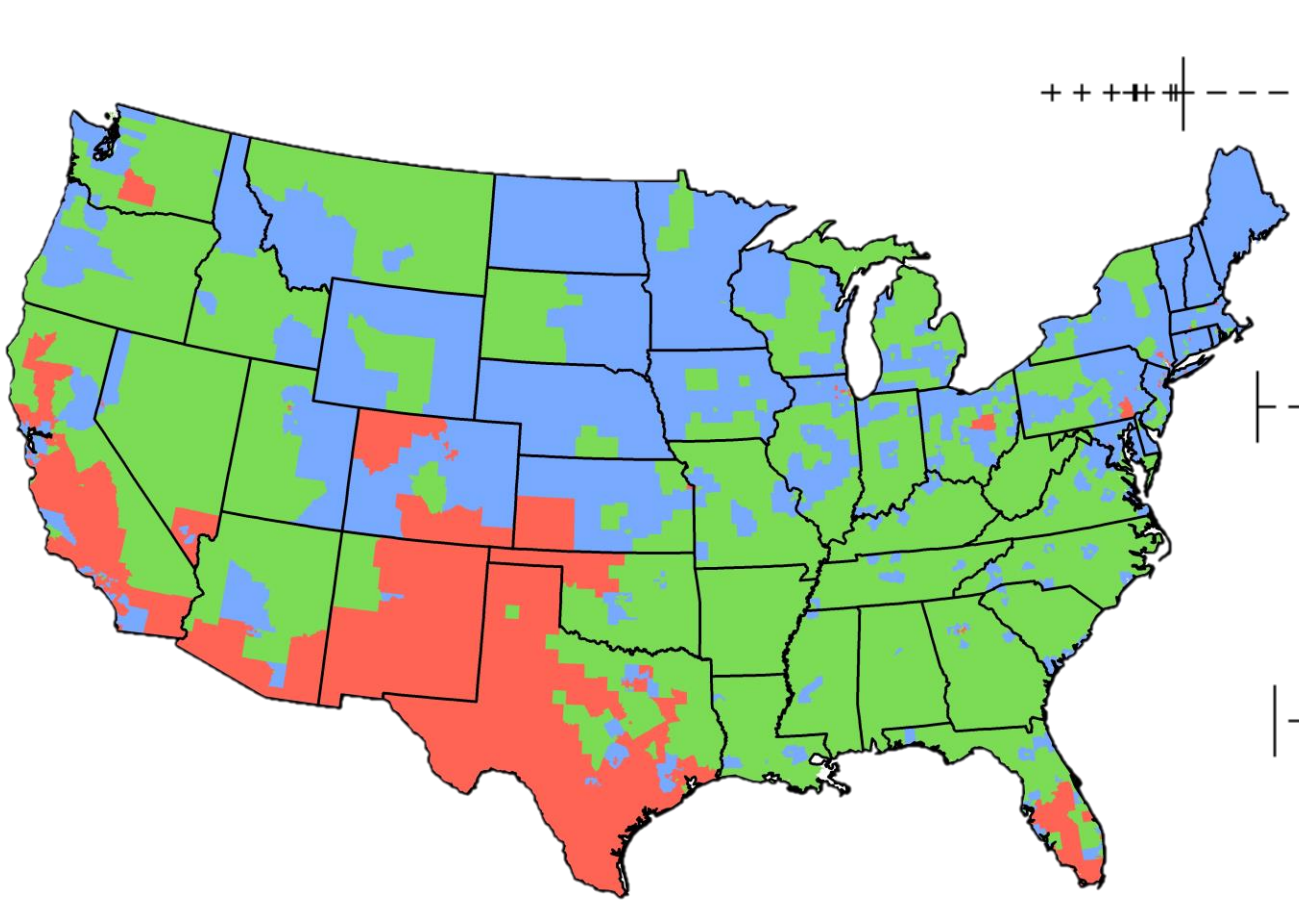
K Means

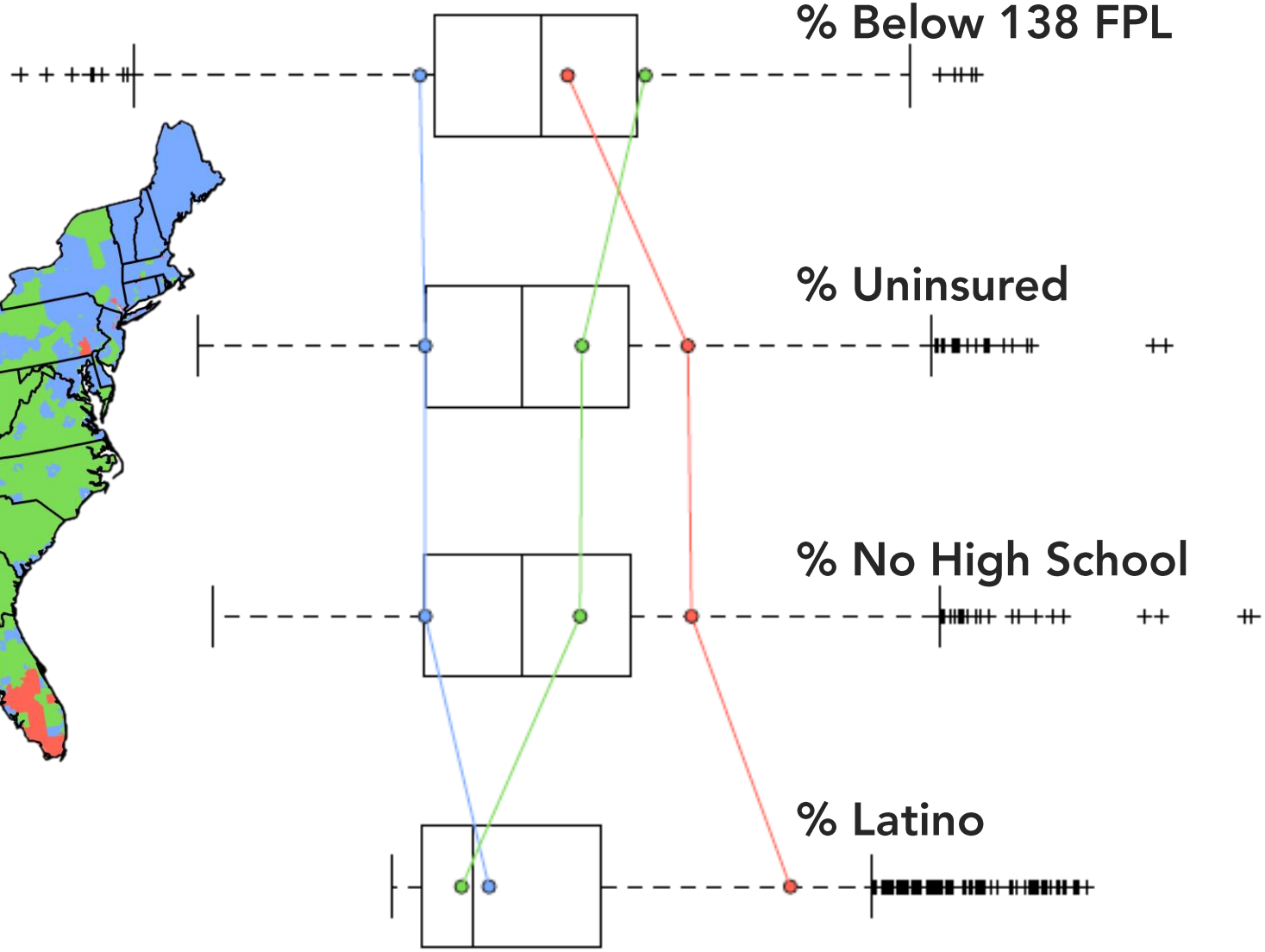


K Means



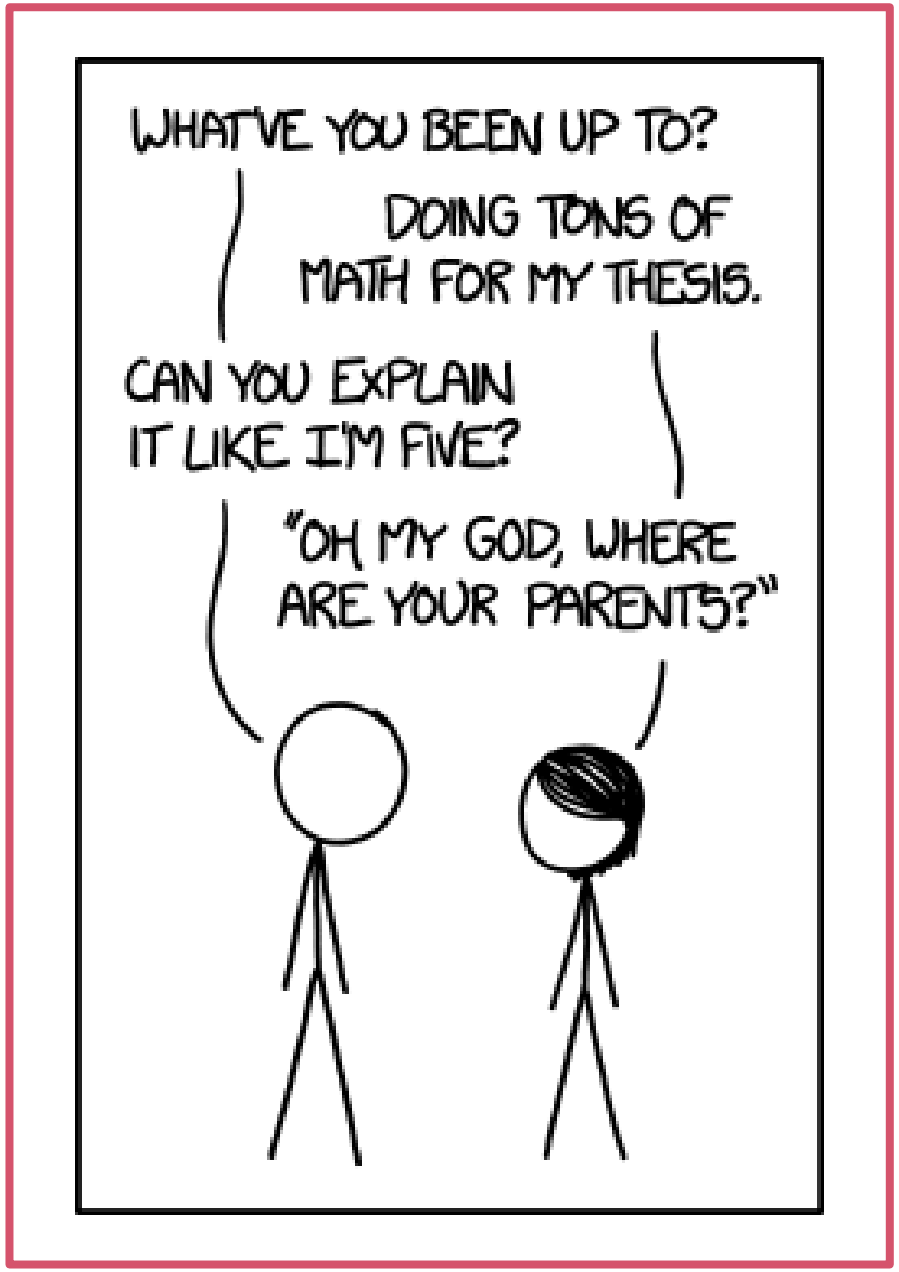
Eligible Uninsured Americans





Spatially Constrained Multivariate Clustering

finds clusters based on feature attributes and proximity



https://xkcd.com/1364/

Minimum Spanning Tree

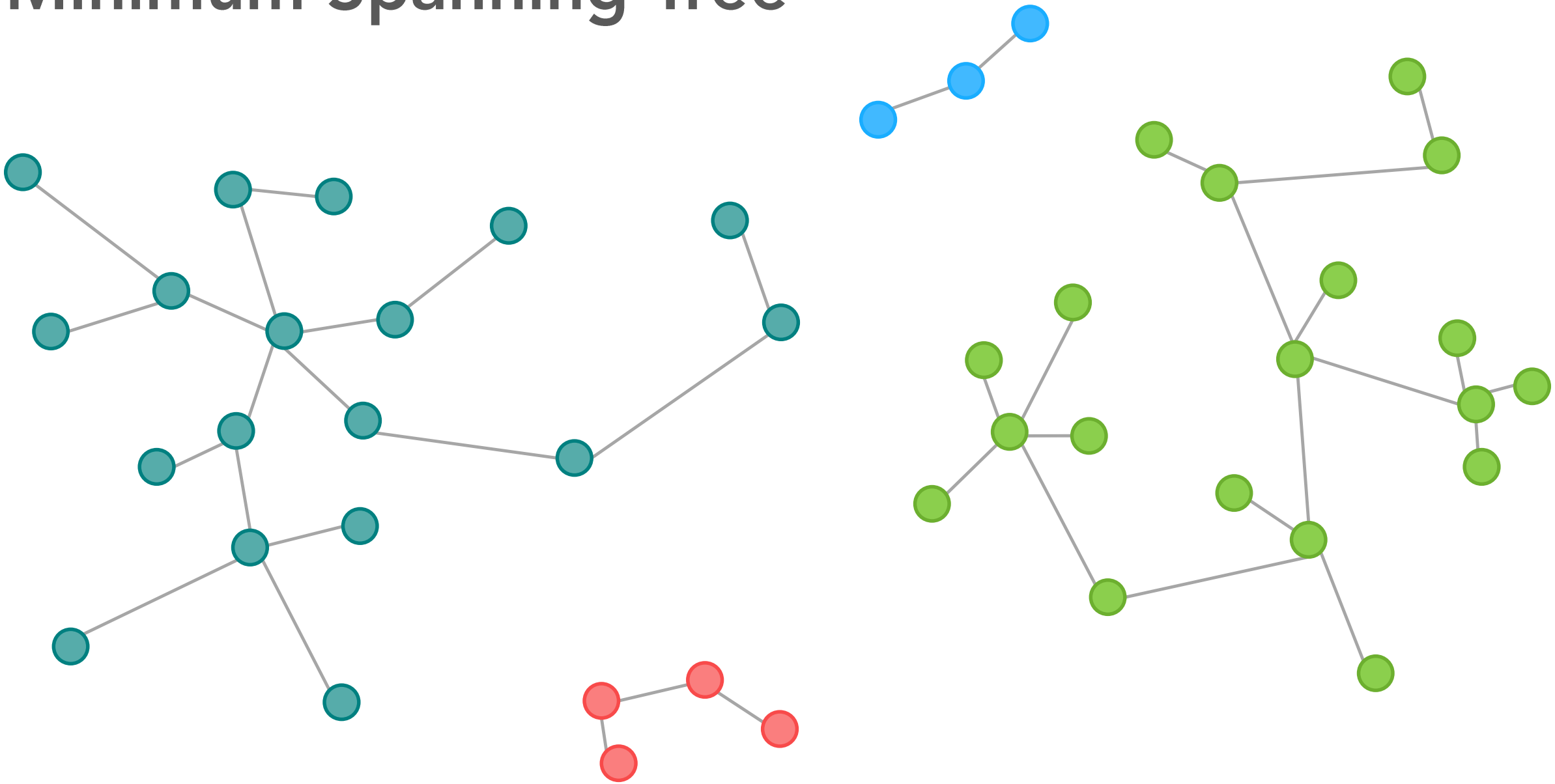
Minimum Spanning Tree

Minimum Spanning Tree

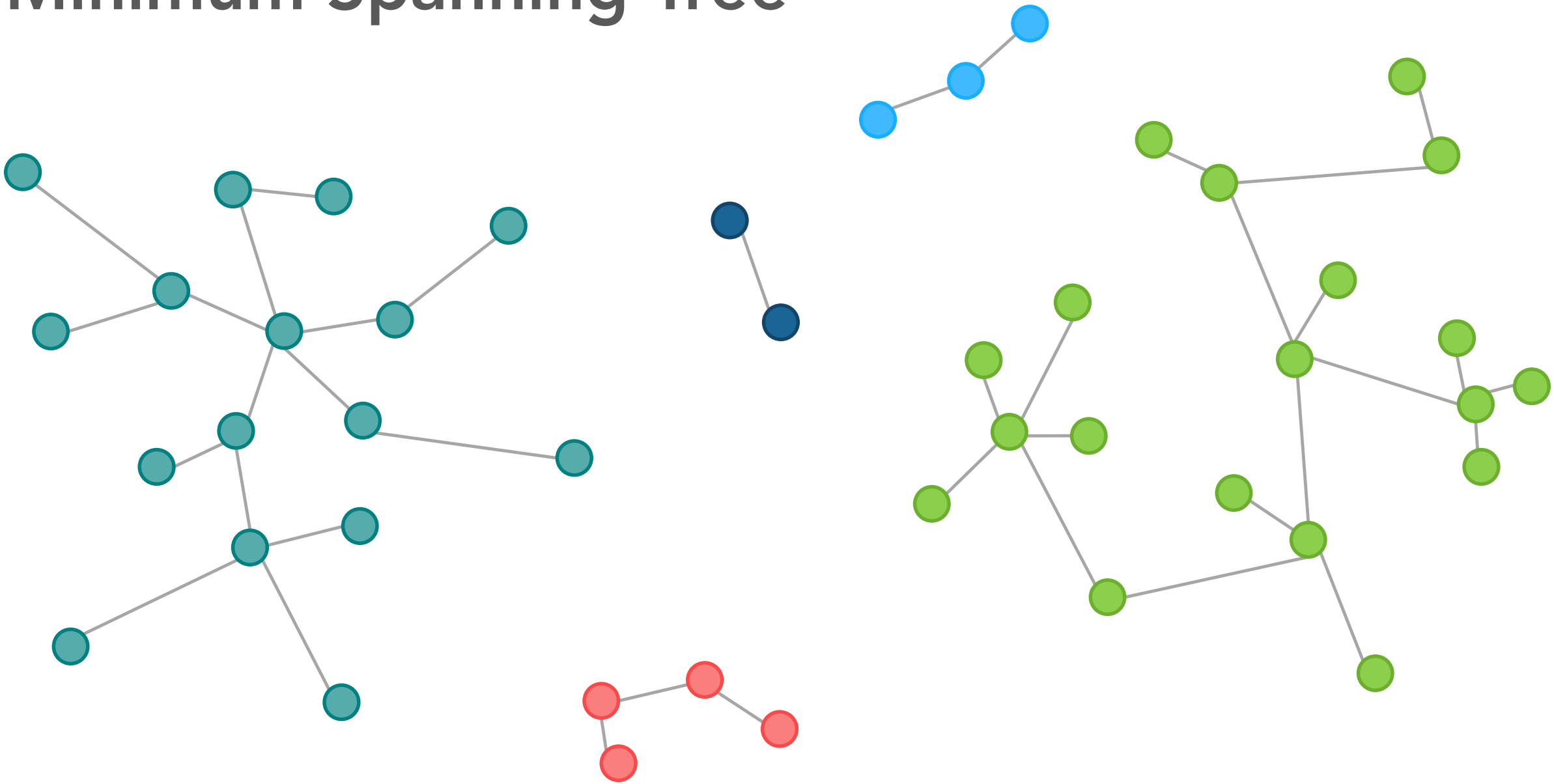
Minimum Spanning Tree

Minimum Spanning Tree

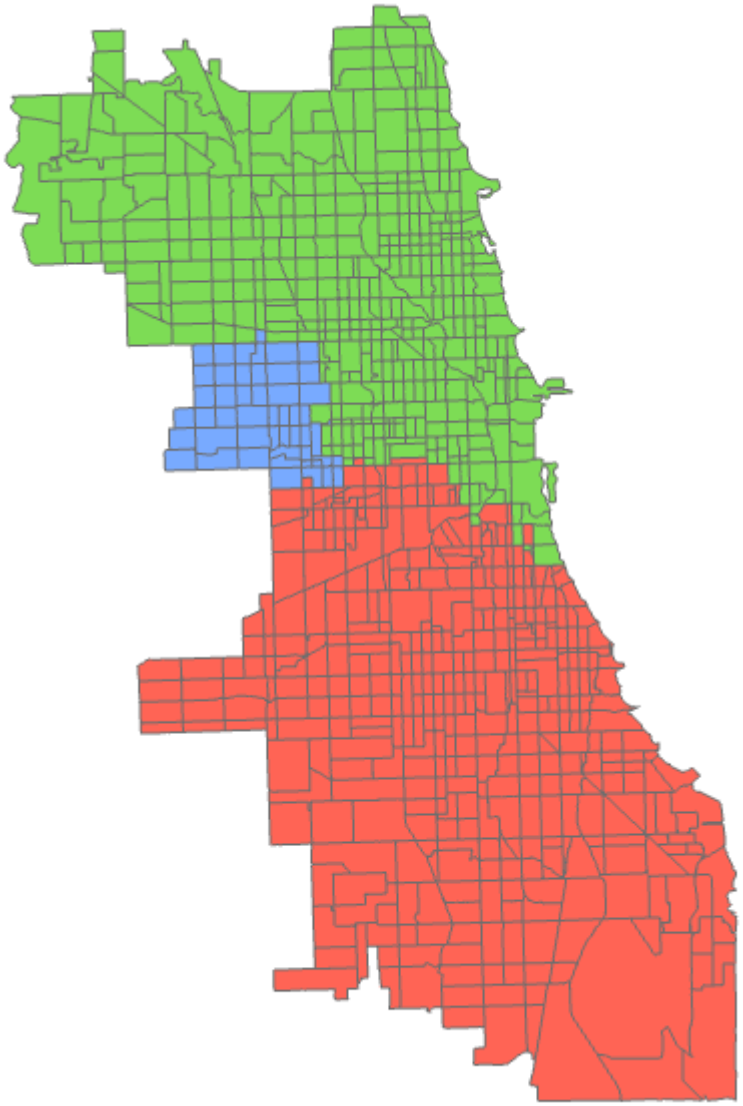
Minimum Spanning Tree



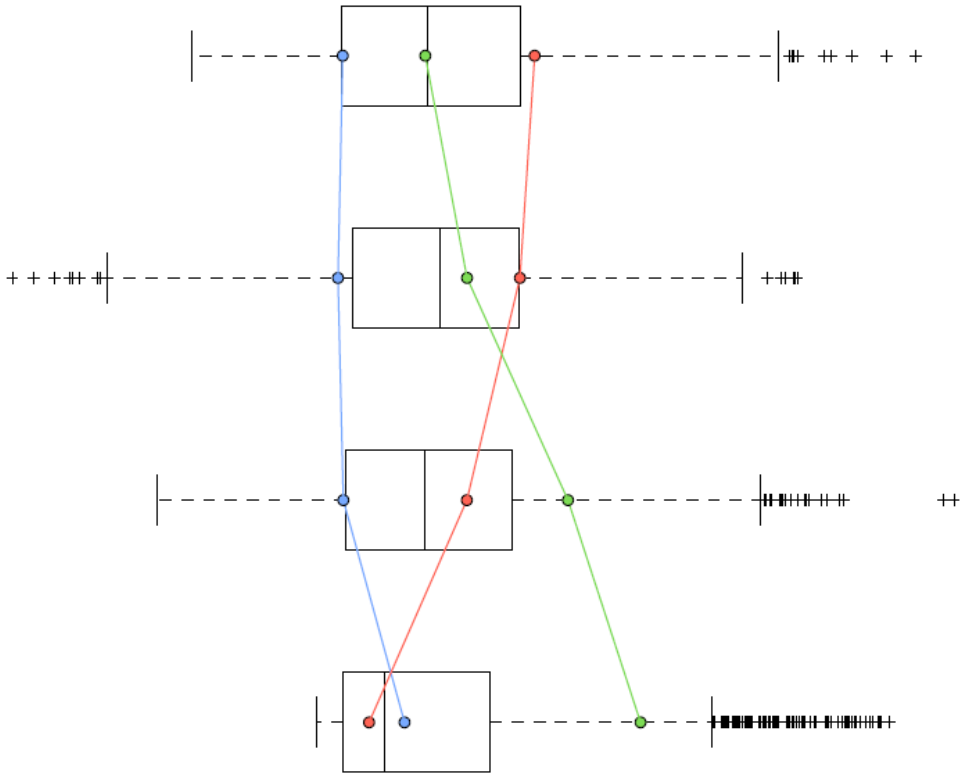
Minimum Spanning Tree



Crime in Chicago



Demo



Want to learn more???

esriurl.com/spatialstats

TUESDAY

~~8:30a From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 15B~~

~~10a Data Visualization for Spatial Analysis 10~~

~~2:30p Spatial Data Mining I: Essentials of Cluster Analysis 15B~~

~~4p From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 14B~~

WEDNESDAY

10a Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 29C

2:30p Spatial Data Mining I: Essentials of Cluster Analysis Room 15A

4p Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 31B

THURSDAY

10a Data Visualization for Spatial Analysis 07A/B

1p Beyond Where: Modeling Spatial Relationships and Making Predictions 17B

4p Beyond Where: Modeling Spatial Relationships and Making Predictions 17A

esri | THE SCIENCE OF WHERE

lbennett@esri.com
rruthart@esri.com
fvale@esri.com

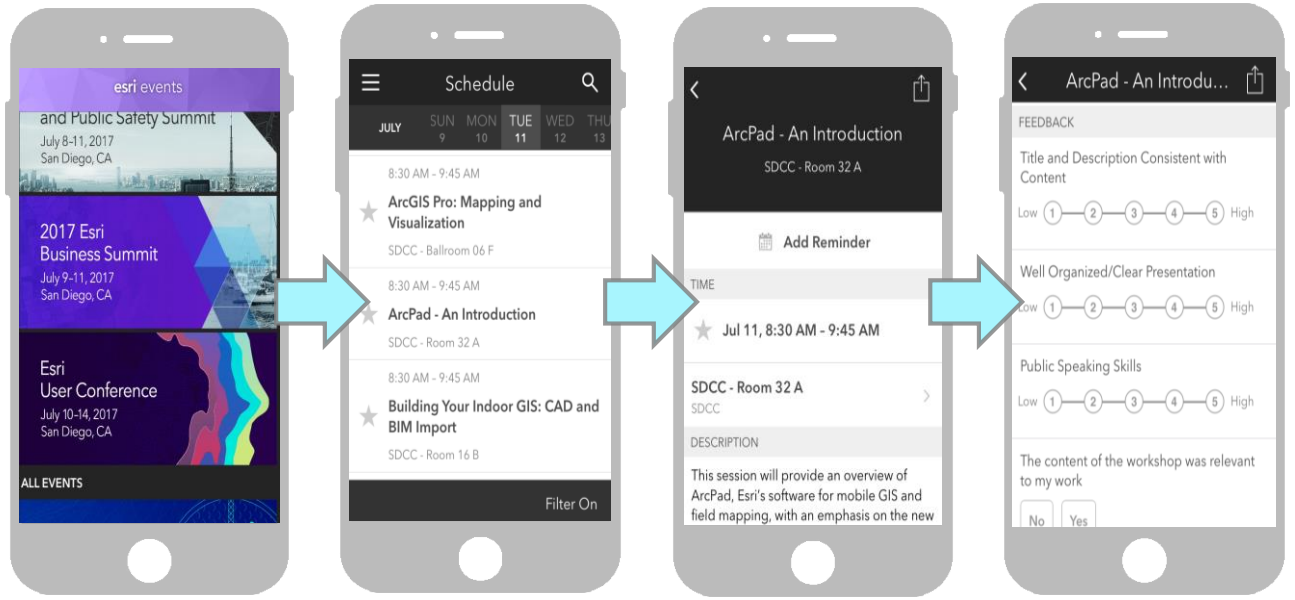
Please fill out a course survey!!

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the survey

Complete Answers and Select "Submit"



Want to learn more???

esriurl.com/spatialstats

TUESDAY

~~8:30a From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 15B~~

10a Data Visualization for Spatial Analysis 10

2:30p Spatial Data Mining I: Essentials of Cluster Analysis 15B

4p From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations 14B

WEDNESDAY

10a Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 29C

2:30p Spatial Data Mining I: Essentials of Cluster Analysis Room 15A

4p Spatial Data Mining II: A Deep Dive Into Space-Time Analysis Room 31B

THURSDAY

10a Data Visualization for Spatial Analysis 07A/B

1p Beyond Where: Modeling Spatial Relationships and Making Predictions 17B

4p Beyond Where: Modeling Spatial Relationships and Making Predictions 17A

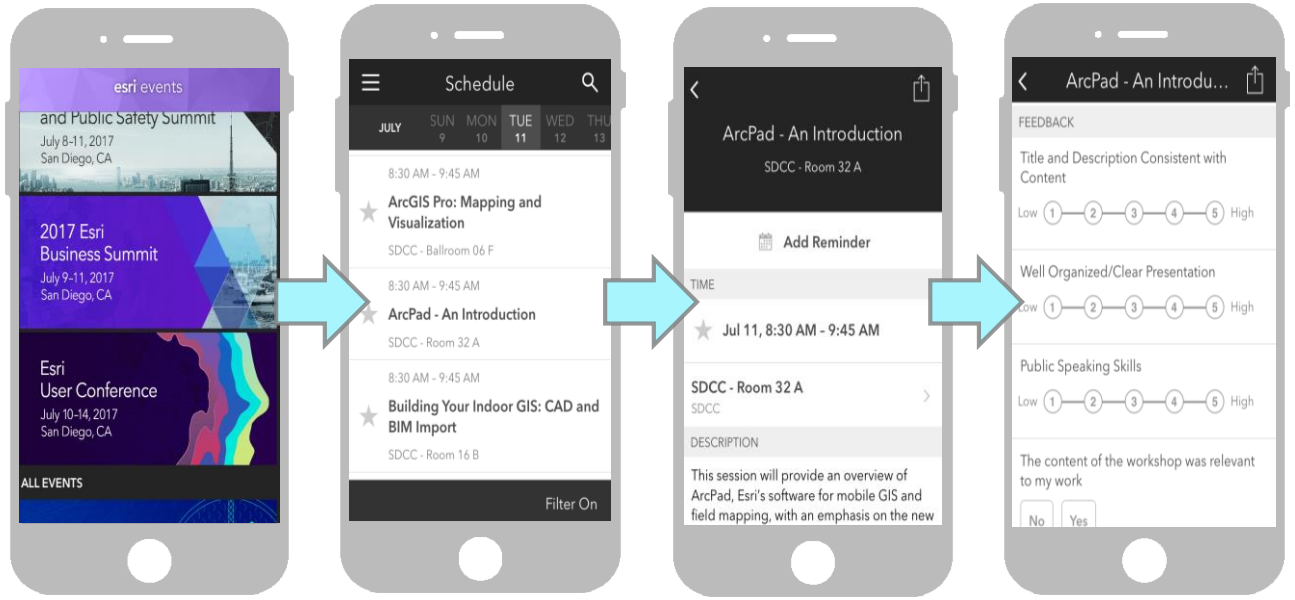
esri | THE SCIENCE OF WHERE

lbennett@esri.com
rruthart@esri.com
fvale@esri.com

Please fill out a course survey!!

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the survey

Complete Answers and Select "Submit"