

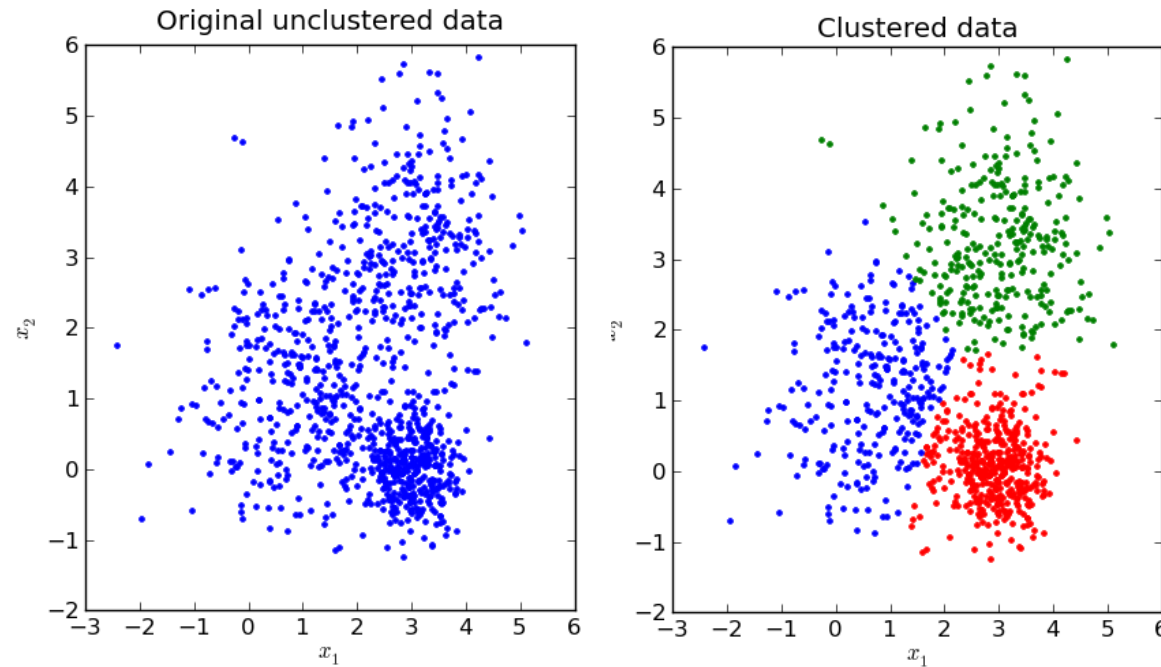
# Class -25

## Clustering (K-Mean & Hierarchical)

---

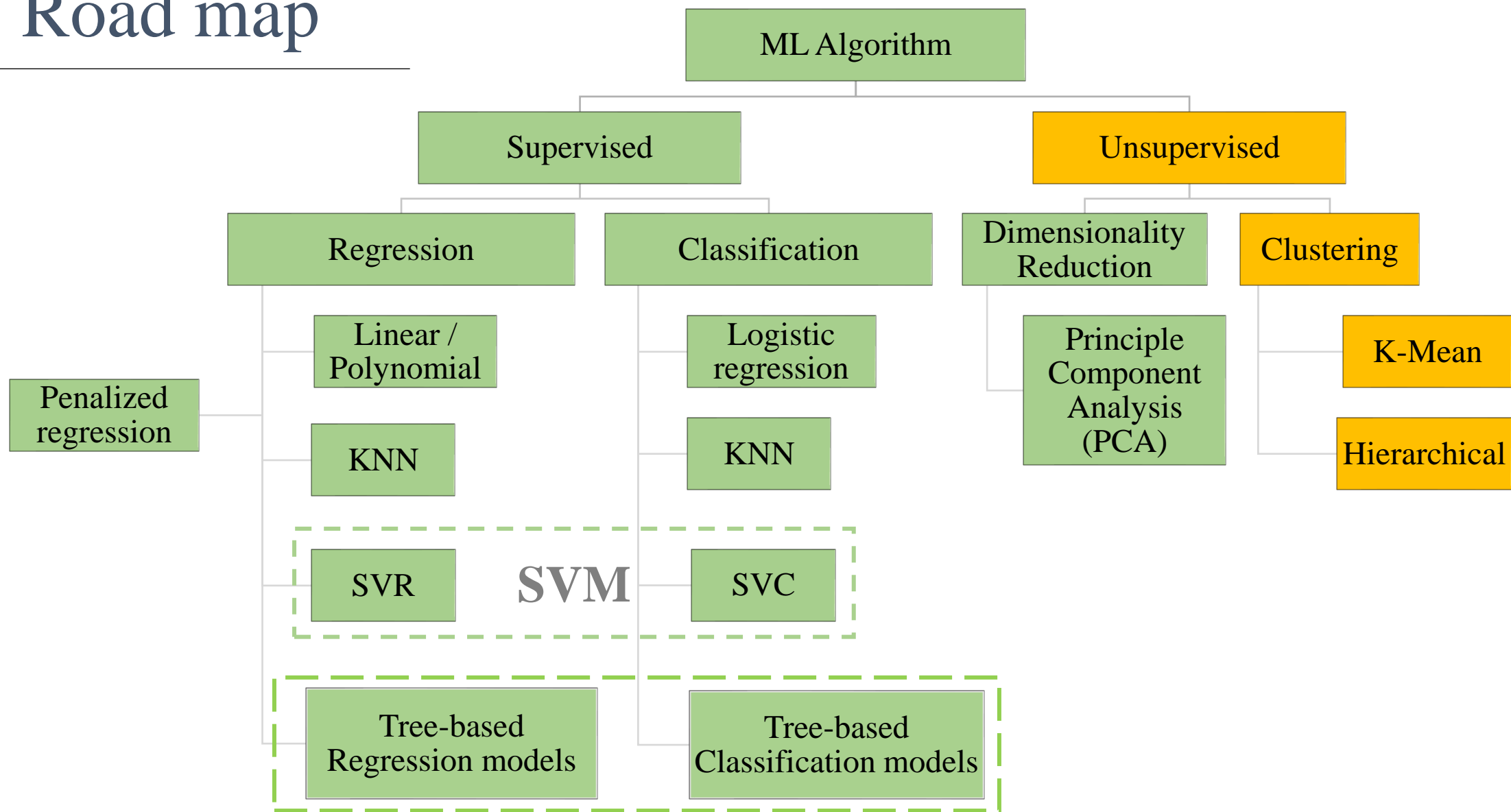


Prof. Pedram Jahangiry





# Road map





# Topics

## Part I

1. What is clustering?
2. Similarity/Dissimilarity metrics
3. PCA vs Clustering

## Part II

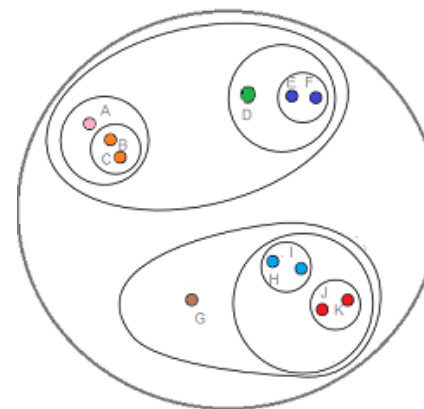
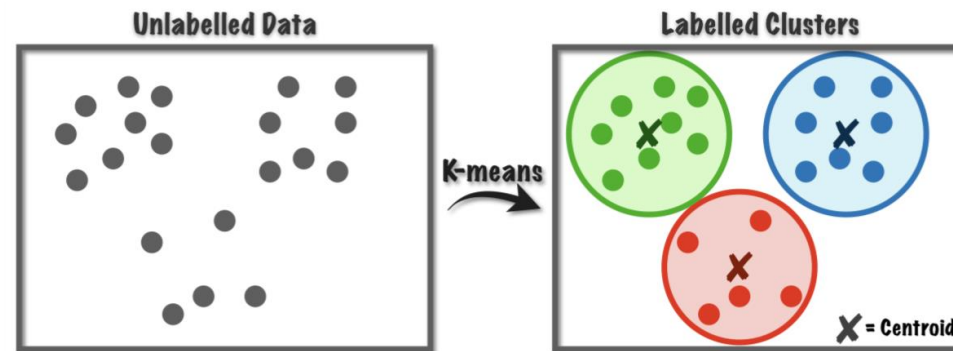
- ✓ K-Mean clustering

## Part III

- ✓ Hierarchical Clustering

## Part IV

- ✓ Applications in finance



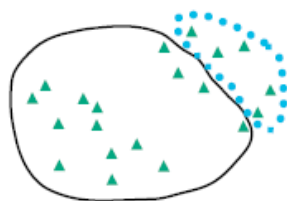
# Part I

1. What is Clustering?
2. Similarity/Dissimilarity metrics
3. PCA vs Clustering

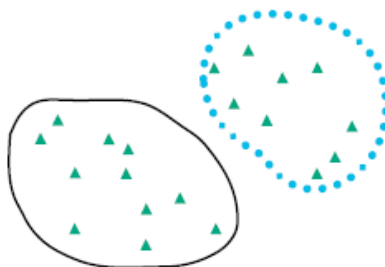
# ➔ What is Clustering?

- Clustering is an **unsupervised** machine learning which is used to organize data points into **similar groups** called **clusters**.
- A cluster contains a subset of observations from the dataset such that all the observations within the same cluster are “**similar**.”
- The goal is to maximize the **intra-clusters** (within) **similarities** or equivalently to maximize the **inter-clusters** (between) **dissimilarities**.

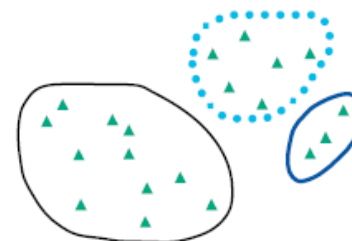
Bad Clustering

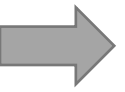


Good Clustering



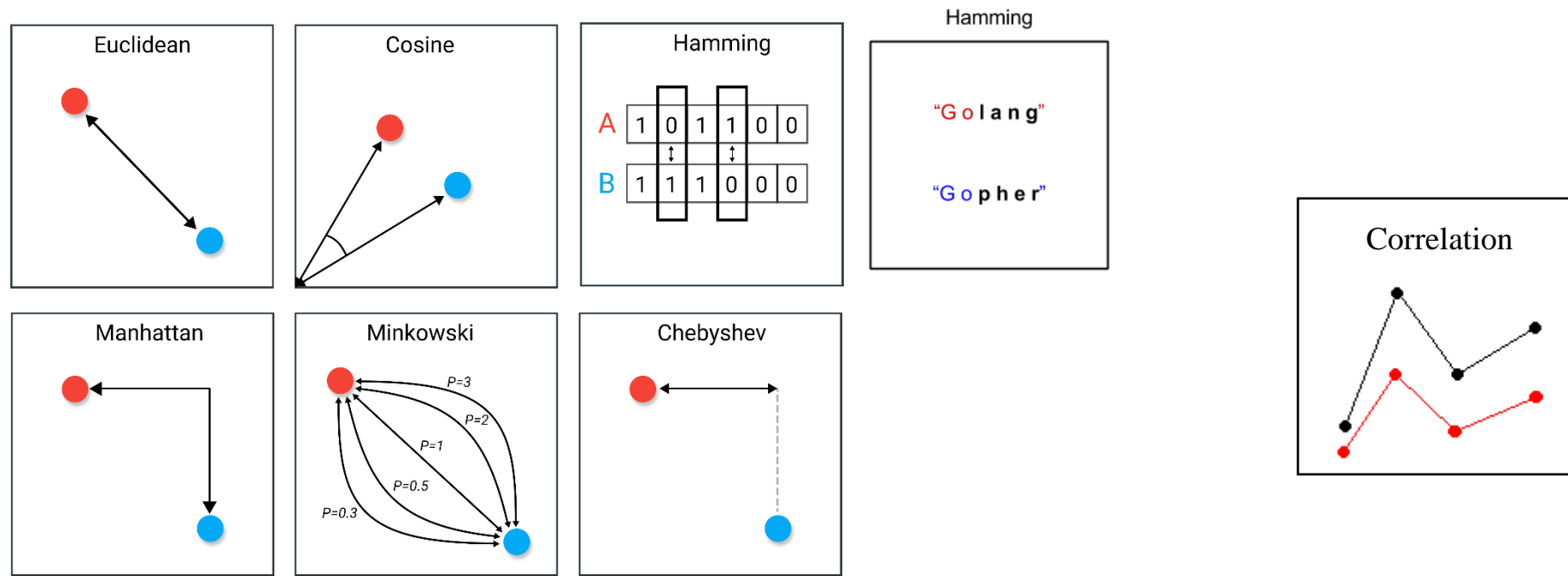
(Maybe) Better Clustering

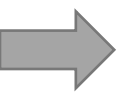




# Similarity/Dissimilarity metrics

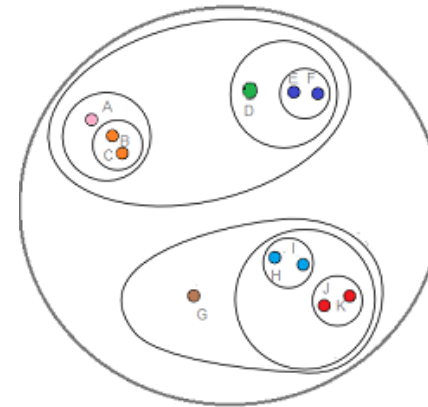
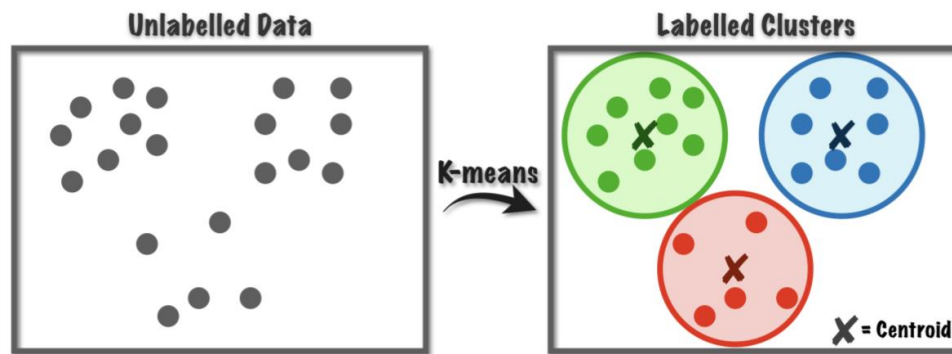
- **Similarity/Dissimilarity** between observations can be thought of as the **distance** between them.
- The smaller the distance, the more similar the observations; the larger the distance, the more dissimilar the observations.





# PCA vs Clustering

- **Principle Component Analysis** (Dimension reduction; feature extraction) looks for a low-dimensional representation of the observations that explains a good fraction of the variance.
- **Clustering** looks for homogeneous subgroups among the observations.
- Two of the more popular clustering approaches are:
  - K-Means clustering
  - Hierarchical clustering



# Part II

□ K-Mean clustering



# K-Means Clustering

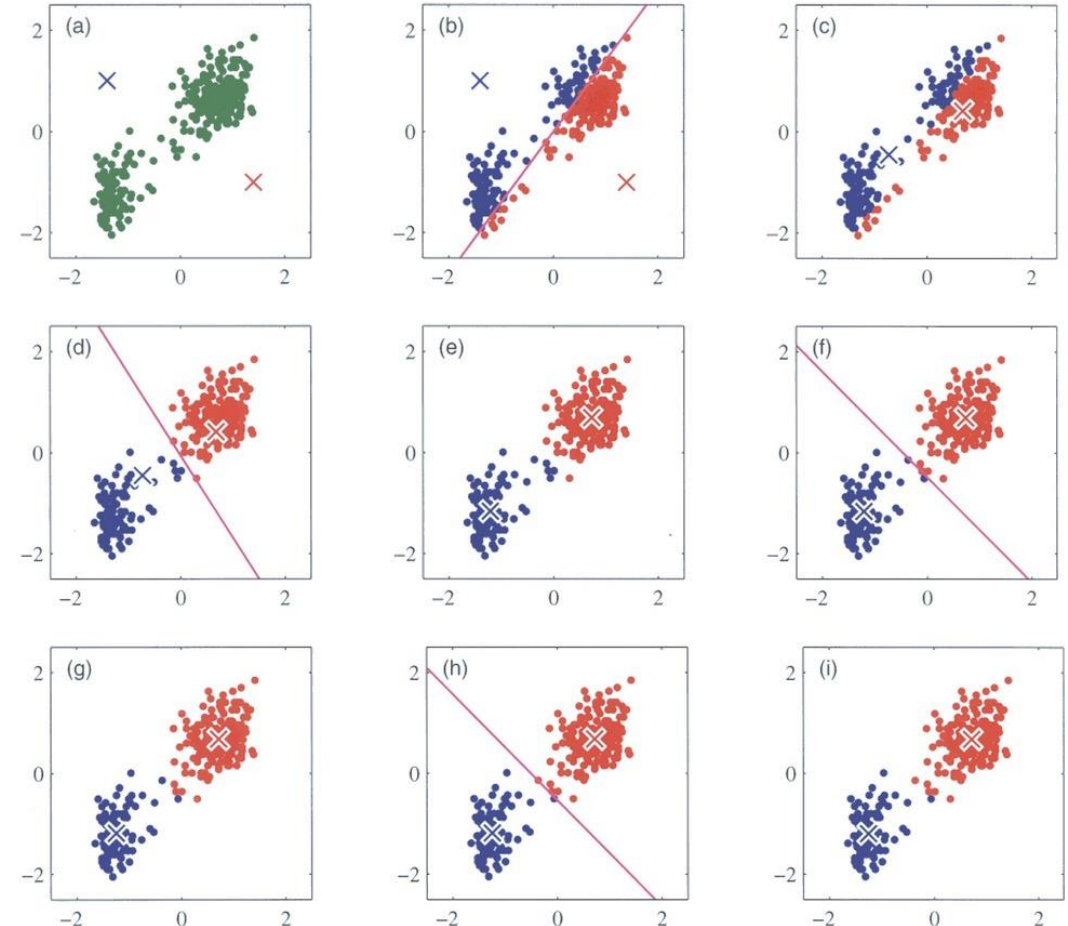
- **K-means** is an algorithm that **repeatedly** partitions observations into a
  - fixed
  - pre-specified and
  - non-overlappingnumber of clusters, **k** (a hyperparameter)
- Each cluster is characterized by its **centroid (arithmetic mean position)**.
- K-means **minimizes** intra-cluster (within-cluster) distance

---

## Algorithm 1 *k*-means algorithm

---

- 1: Specify the number  $k$  of clusters to assign.
  - 2: Randomly initialize  $k$  centroids.
  - 3: **repeat**
  - 4:   **expectation:** Assign each point to its closest centroid.
  - 5:   **maximization:** Compute the new centroid (mean) of each cluster.
  - 6: **until** The centroid positions do not change.
- 



# → K-Means clustering (details)

- **Objective function:** Minimizing the within-cluster variation (WCV)

$$\underset{C_1, \dots, C_K}{\text{minimize}} \left\{ \sum_{k=1}^K \text{WCV}(C_k) \right\}$$

- This optimization says that we want to partition the observations into K clusters such that the **total within-cluster variation** is as small as possible.
- If we use Euclidian distance, then:

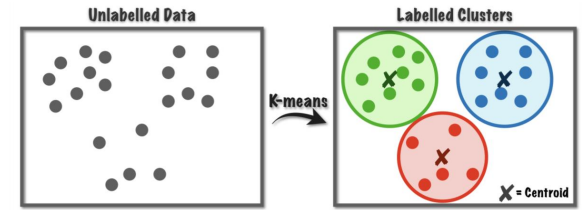
$$\underset{C_1, \dots, C_K}{\text{minimize}} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\}$$



# Initial positioning of the centroids matter!



# K-Means clustering pros and cons



## Pros

- Simple!
- The k-means algorithm is **fast** and works well on **very large datasets**.
- Can help **visualize** the data and facilitate detecting trends or outliers.
- The k-means algorithm is among the most used algorithms in **investment practice**, particularly in data exploration for discovering patterns in high-dimensional data or as a method for deriving alternatives to existing static industry classifications.

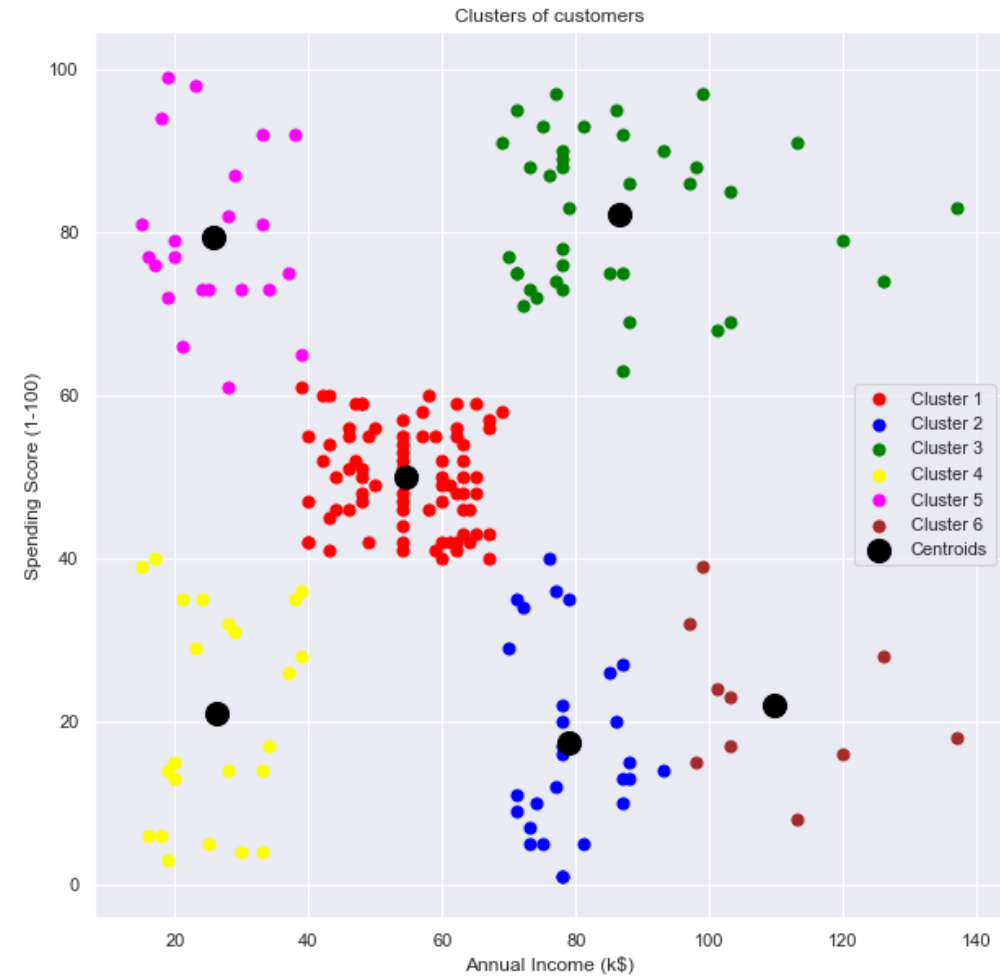
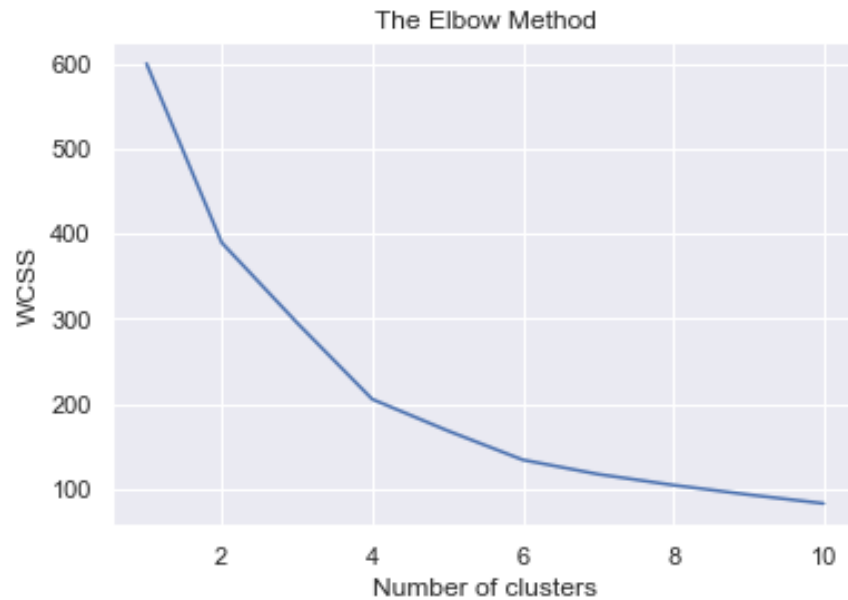
## Cons

- The hyperparameter, **k**, must be decided **before** the algorithm can be run.
- The final assignment of observations to clusters can **depend on the initial location of the centroids**. Local optimum vs global optimum. (should be rerun with several initialization)

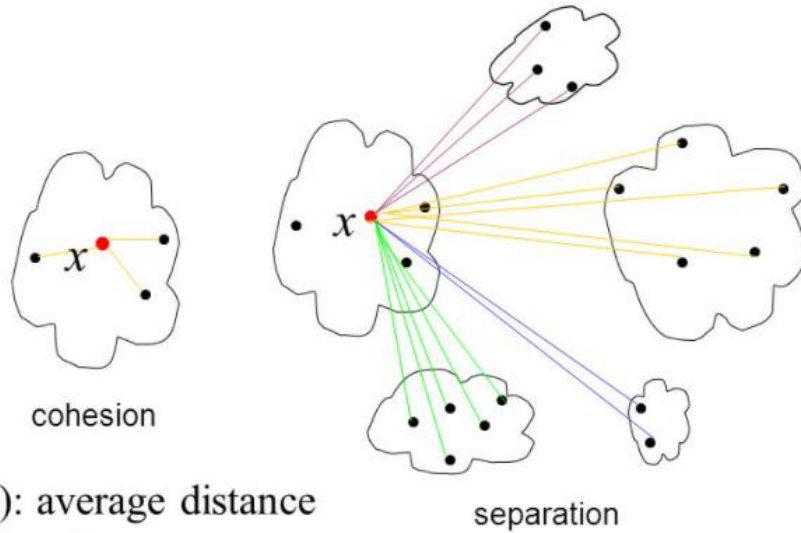


# Optimal number of K (the elbow method)

$$\text{minimize}_{C_1, \dots, C_K} \left\{ \sum_{k=1}^K \text{WCV}(C_k) \right\}$$



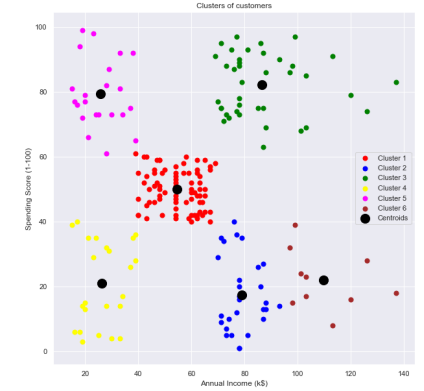
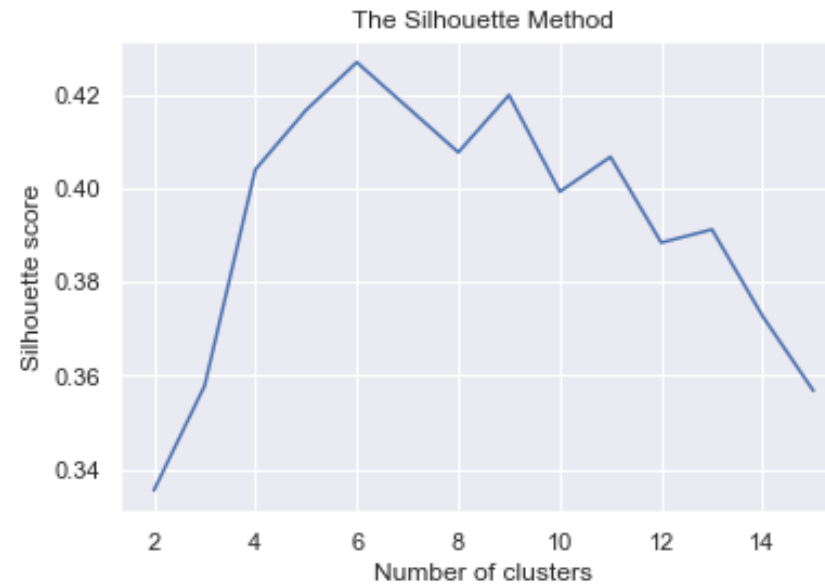
# Optimal number of K (the Silhouette method)



$a(x)$ : average distance  
in the cluster

$b(x)$ : average distances to  
others clusters, find minimal

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad -1 \leq s(i) \leq 1$$



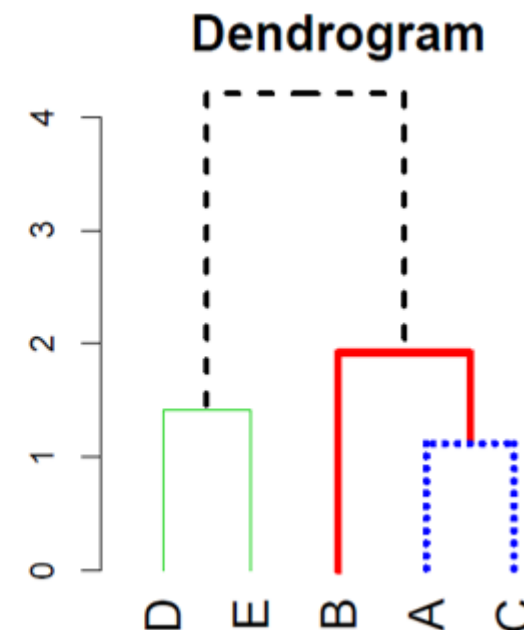
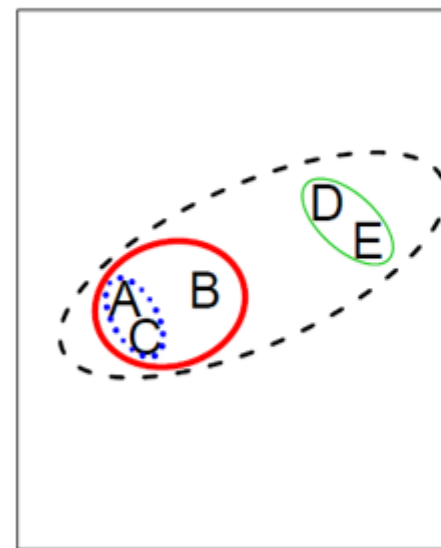
# Part II

## □ Hierarchical Clustering

1. Agglomerative
2. Divisive

# → Hierarchical Clustering

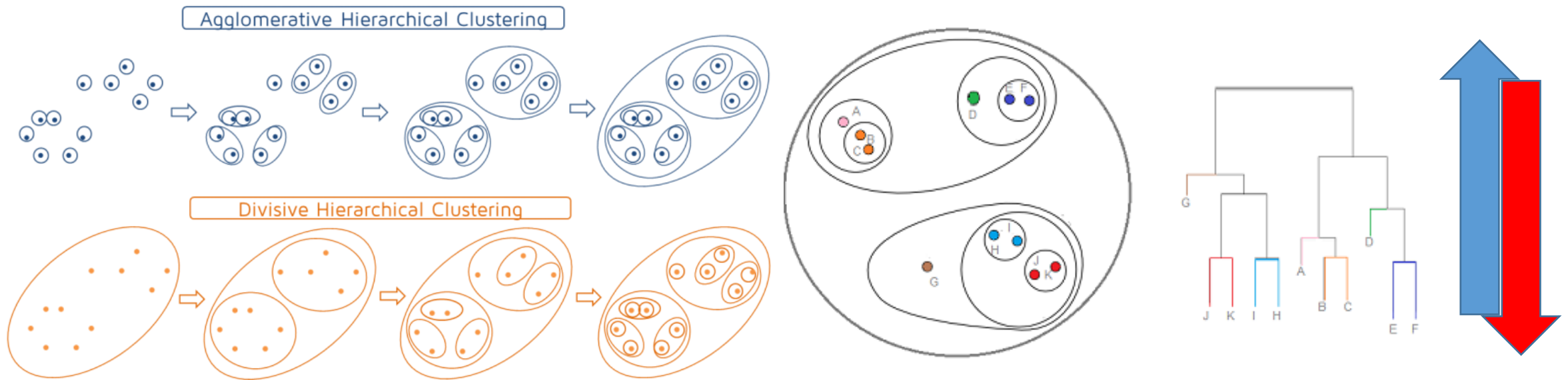
- In **k-means** clustering, the algorithm seeks to partition the data into a **pre-specified** number of clusters  $k$ . All clusters are found **simultaneously**.
- In **hierarchical** clustering, the algorithm **does not require** a pre-specified choice of  $K$ . Clusters are found **sequentially**.
- Hierarchical clustering is an **iterative procedure** used to build a **hierarchy of clusters**.
- Using a **dendrogram** (a type of tree diagram which highlights the hierarchical relationships among the clusters), hierarchical clustering has the advantage of allowing the analyst to examine alternative partitioning of data of **different granularity before** deciding which one to use.

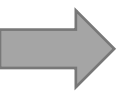




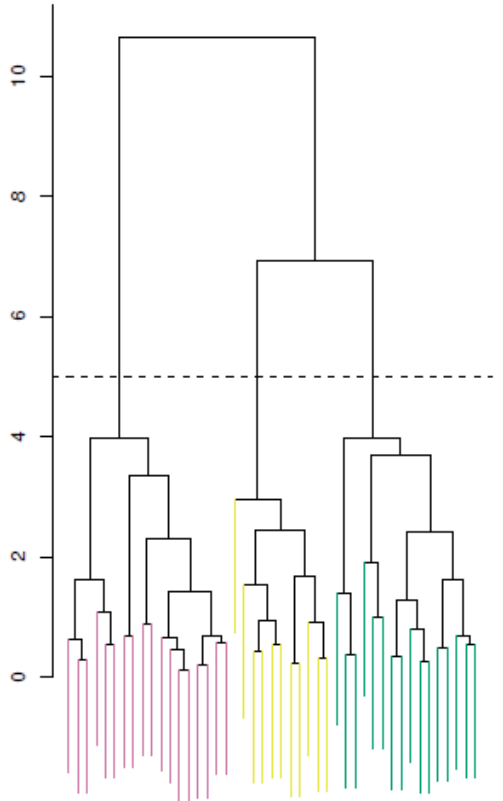
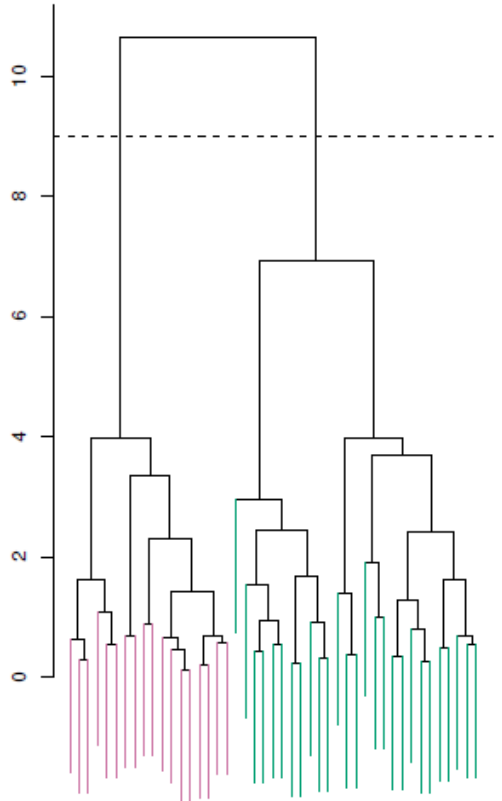
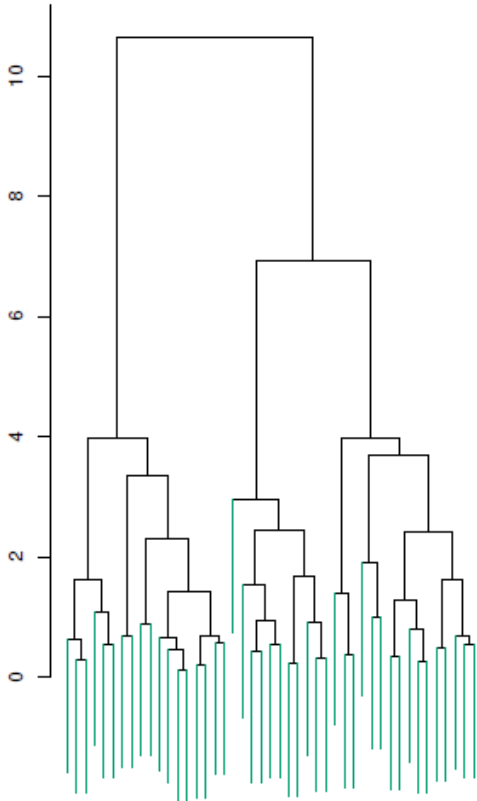
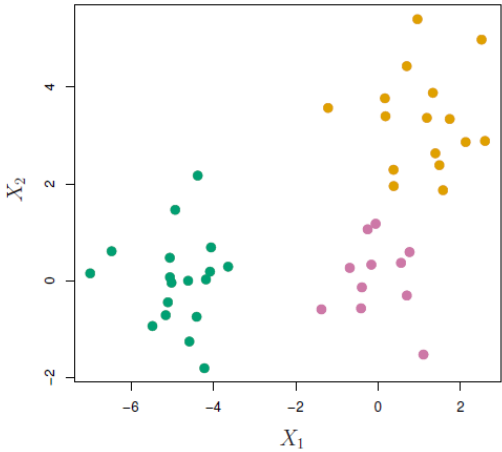
# Agglomerative (bottom-up) vs Divisive (top-down) HCA

- **Agglomerative**: start with each observation being treated as its own cluster
- **Divisive**: starts with all the observations belonging to a single cluster.



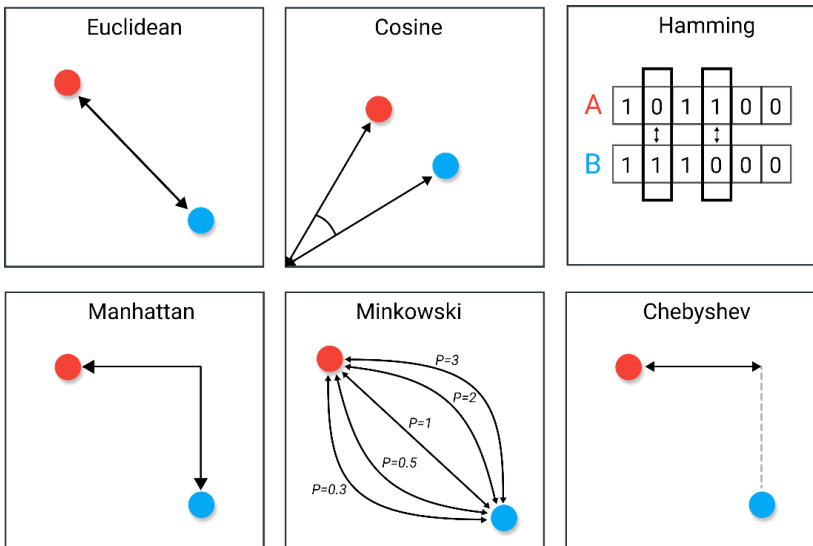


# An example



# Types of Linkage (distance between two clusters)

- To decide on the **closest clusters**, an explicit definition for the **distance** between two clusters is required (linkage)
- Recall: We have already defined the **within-cluster** distance metrics.



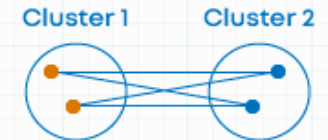
- Single Linkage**  
 $D(c_1, c_2) = \min D(x_i, x_j)$   
Minimum distance or distance between closest elements in clusters



- Complete Linkage**  
 $D(c_1, c_2) = \max D(x_i, x_j)$   
Maximum distance between elements in clusters



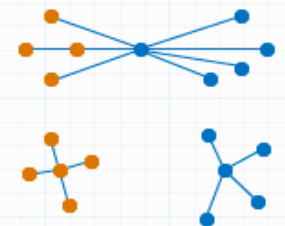
- Average Linkage**  
 $D(c_1, c_2) = \frac{1}{|c_1|} \frac{1}{|c_2|} \sum \sum D(x_i, x_j)$   
Average of the distances of all pairs



- Centroid Method**  
Combining clusters with minimum distance between the centroids of the two clusters



- Ward's Method**
  - Combining clusters where increase in within cluster variance is to the smallest degree.
  - Objective is to minimize the total within cluster variance



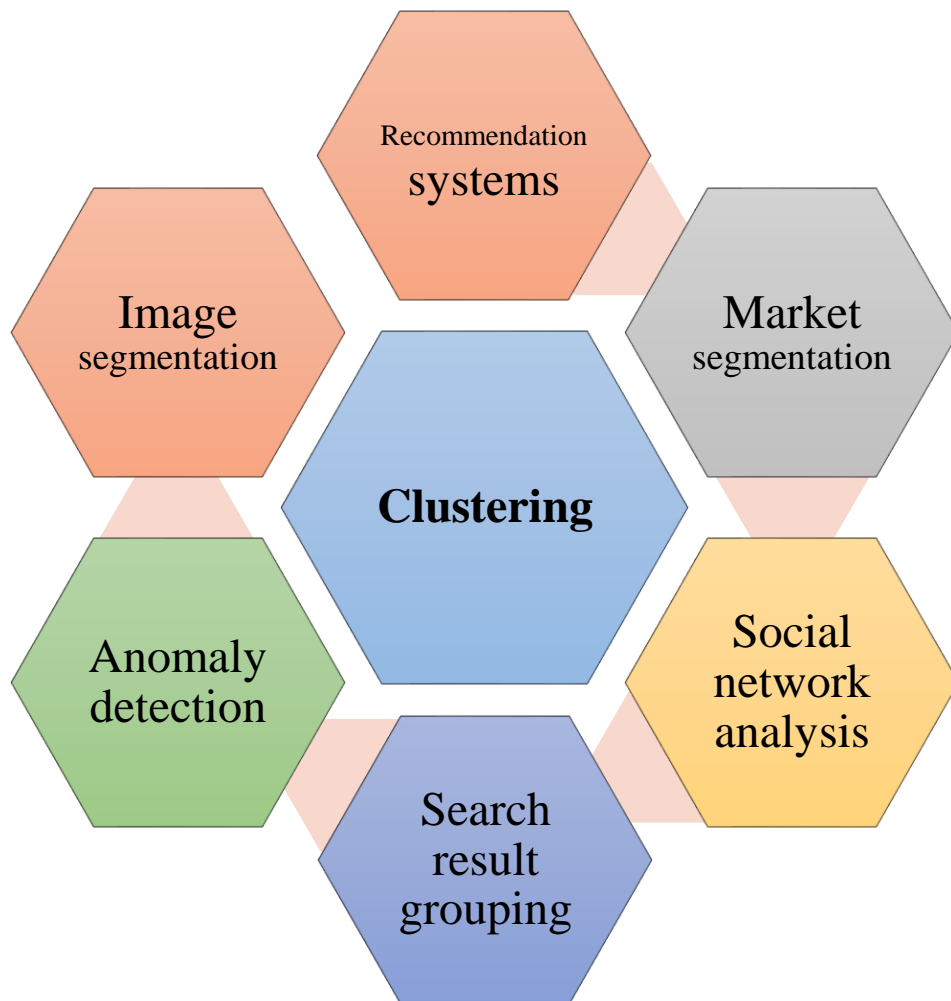
# → Hierarchical Clustering discussion

- The **agglomerative** method is the approach typically used with **large** datasets because of the algorithm's fast computing speed.
- The **agglomerative** clustering algorithm makes clustering decisions based on **local** patterns without initially accounting for the global structure of the data. As such, the agglomerative method is well suited **for identifying small clusters**.
- The **divisive** method starts with a **holistic** representation of the data, so it is designed to account for the global structure of the data and thus is better suited **for identifying large clusters**.
- What **dissimilarity measure** should be used?
- What type of **linkage** should be used?
- There is **no commonly agreed-upon** way to decide where to cut the tree.

# Part III

## Applications in finance

# → Applications of clustering



# ➔ Applications in Finance

- Clustering algorithms are particularly useful in the many investment problems and applications in which the concept of **similarity is important**.
- Applied to grouping companies, for example, clustering may uncover important similarities and differences among companies that are **not captured by standard classifications of companies by industry and sector**.
- In **portfolio management**, clustering methods have been used for improving portfolio diversification by investing in assets from multiple different clusters.





# Clustering stocks based on co-movement similarity

## Exhibit 23 Dataset of Eight Stocks from the S&P 500 Index

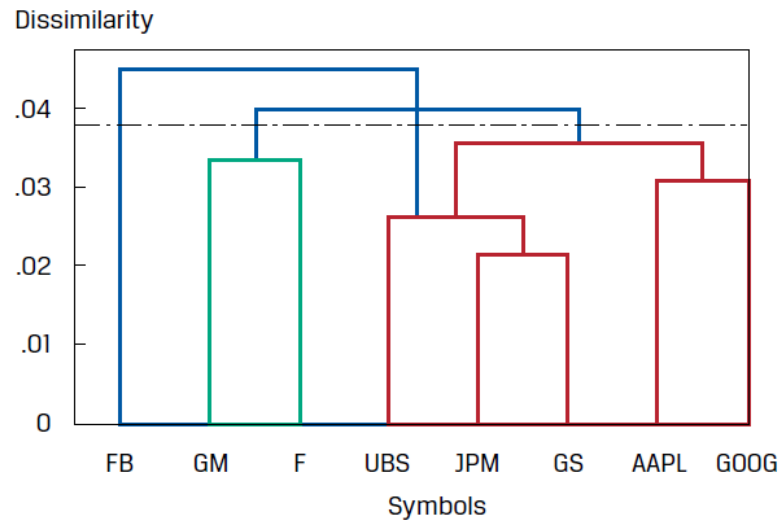
Description: Daily adjusted closing prices of eight S&P 500 member stocks

Trading Dates: 30 May 2017 to 24 May 2019

Number of Observations: 501

Stocks (Ticker Symbols): AAPL, F, FB, GM, GS, GOOG, JPM, and UBS

## Exhibit 26 Dendrogram for Hierarchical Agglomerative Clustering



Developed and written by Matthew Dixon, PhD, FRM.

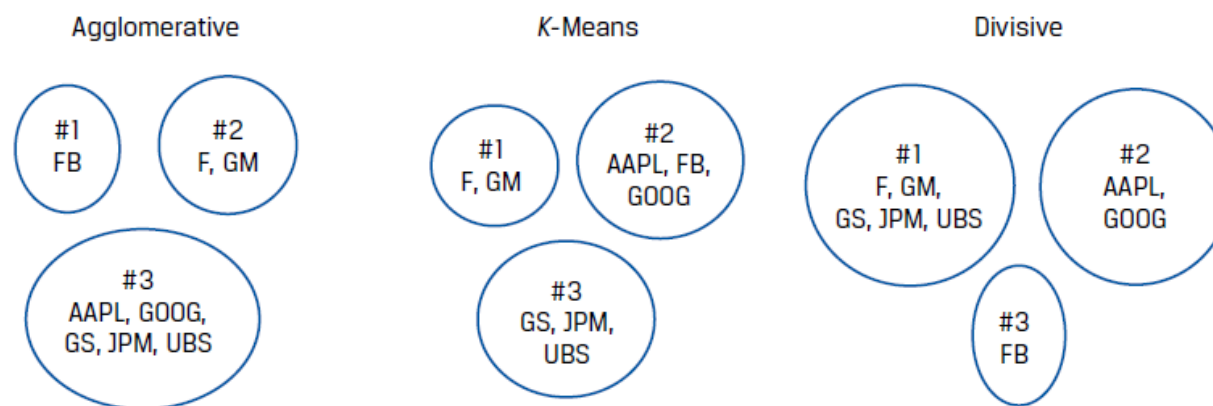




# Comparing clustering methods

**Exhibit 27 Comparison of Results of Different Clustering Algorithms**

	Agglomerative	K-means	Divisive
AAPL	3	2	2
F	2	1	1
FB	1	2	3
GM	2	1	1
GOOG	3	2	2
GS	3	3	1
JPM	3	3	1
UBS	3	3	1



Developed and written by Matthew Dixon, PhD, FRM.

# → Students' questions

---

1. How does K-means clustering assist supervised learning? Add more linearity to data?
2. I don't know how you would know what situations to use K-mean clustering or when it's better to use hierarchical clustering.
3. Does clustering lead to false trends in the data?