



# Unsupervised Learning

## Session-18



# SUMMARY of PREVIOUS CLASS



- **Unsupervised Learning**
- **K-Means Clustering**
- **Clustering Tendency:** Hopkins Test (Randomness)
- **Number of cluster “K”:**
  - Domain Knowledge
  - Data driven (Elbow using Inertia metric-SSD)
- **Clustering Quality**
  - **External metrics (Domain Knowledge-Need some labels):** Adjusted Random Index (ARI),.....
  - **Internal Metrics:** Silhouette Coefficient,....

$$s = \frac{b - a}{\max(a, b)}$$



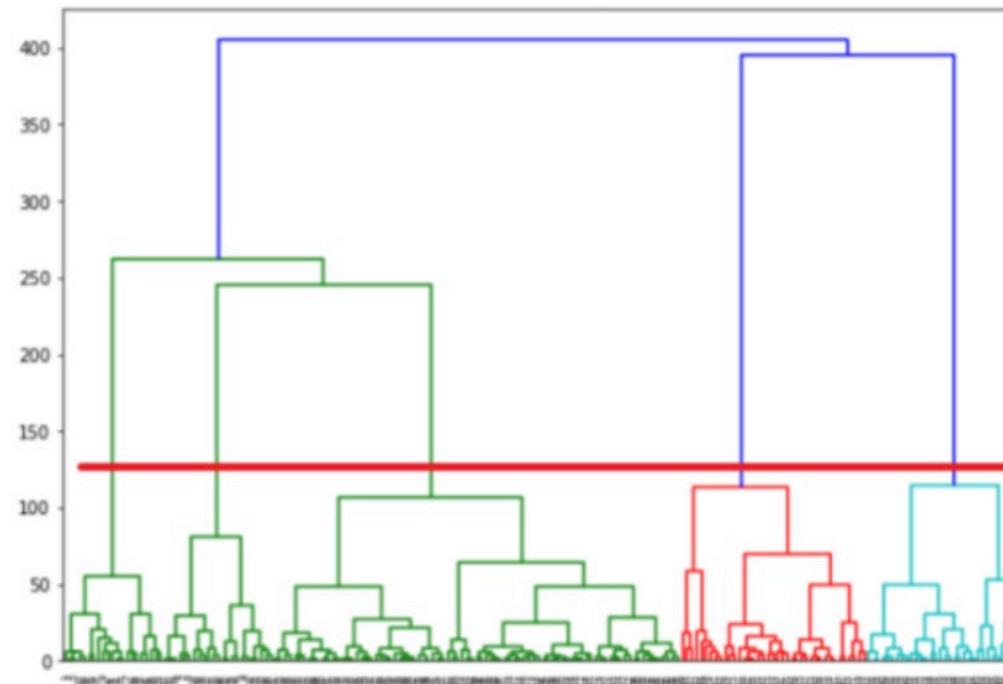
- **Hierarchical Clustering Theory**
- **Hierarchical Clustering with Python**



# Hierarchical Clustering Theory



Hierarchical Clustering is widely used  
**Unsupervised Learning**  
algorithm.



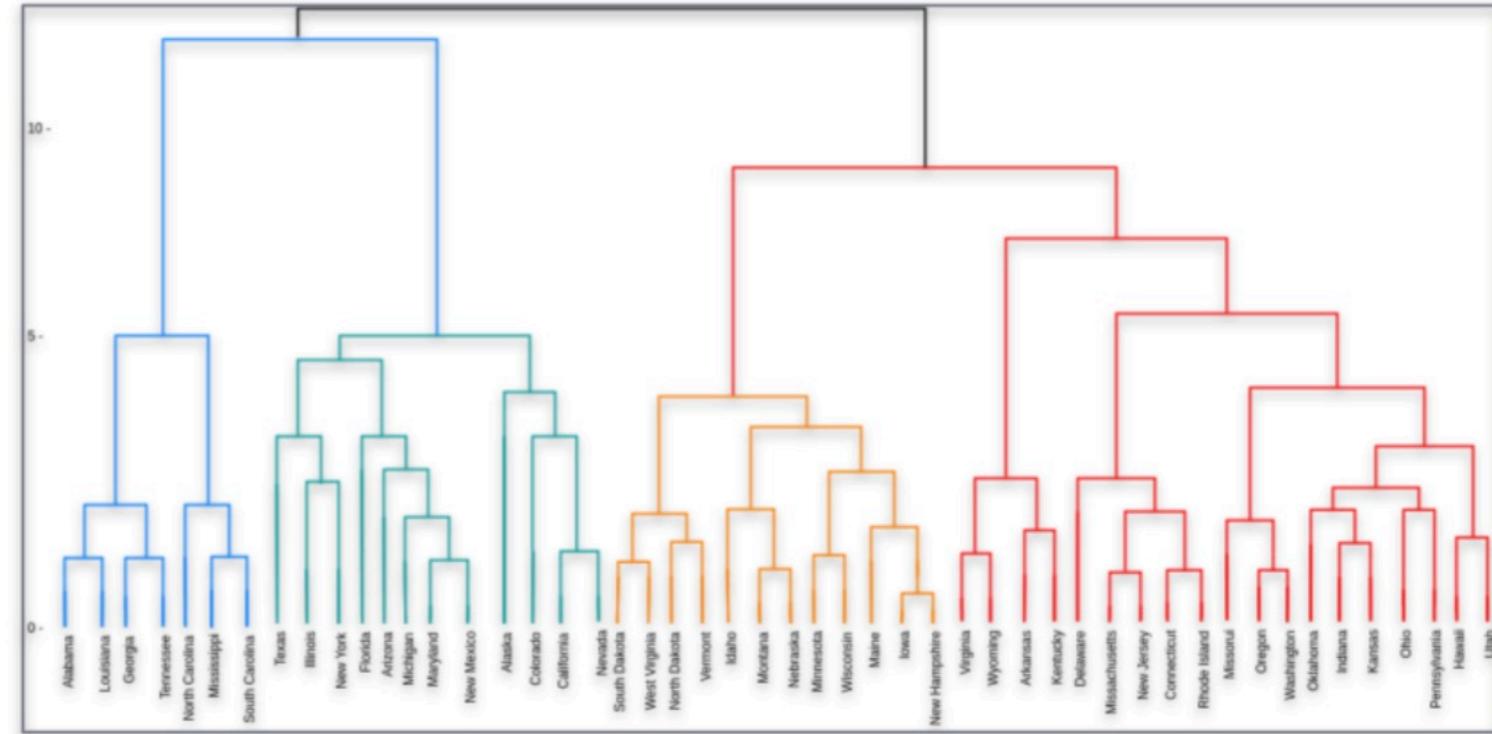
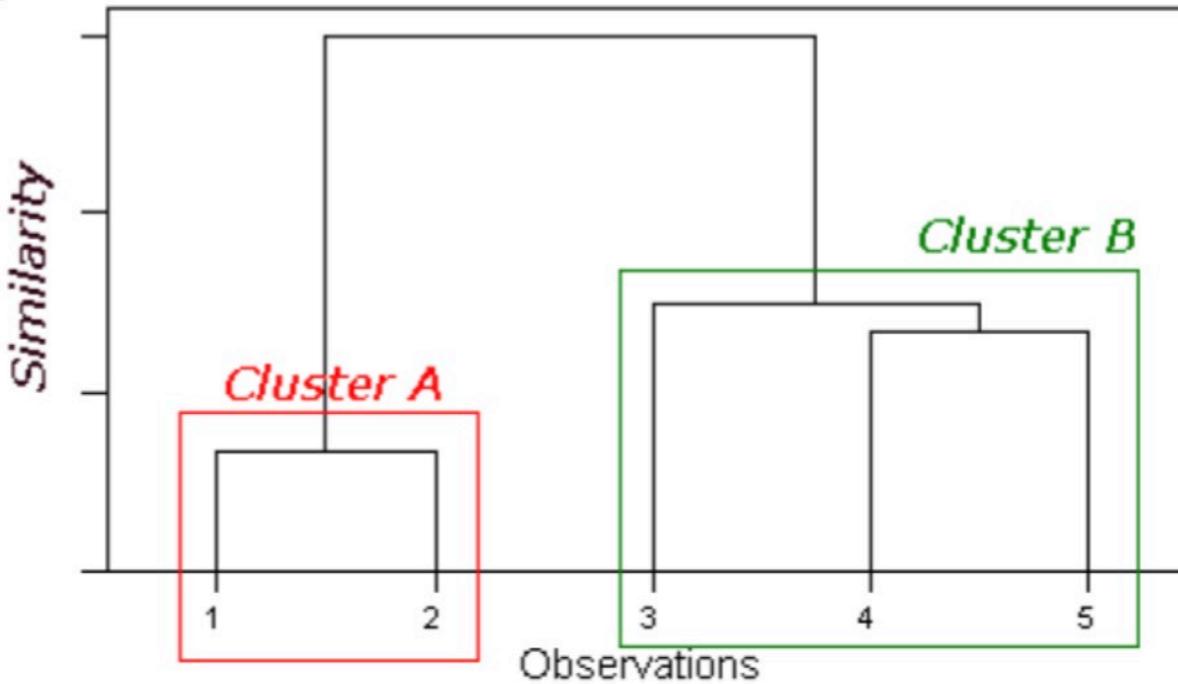
## Some characteristics of Hierarchical Clustering:

- Aims to create a tree-like cluster hierarchy (**dendrogram**) within the data
- No need to specify the number of clusters at the beginning.

# Hierarchical Clustering Theory



## Dendrogram :



Tree-like hierarchical representation of clusters is called a **dendrogram**. It illustrates the arrangement of the clusters produced by the corresponding analyses.

# Hierarchical Clustering Theory



## Types of Hierarchical Clustering

### Step-1:

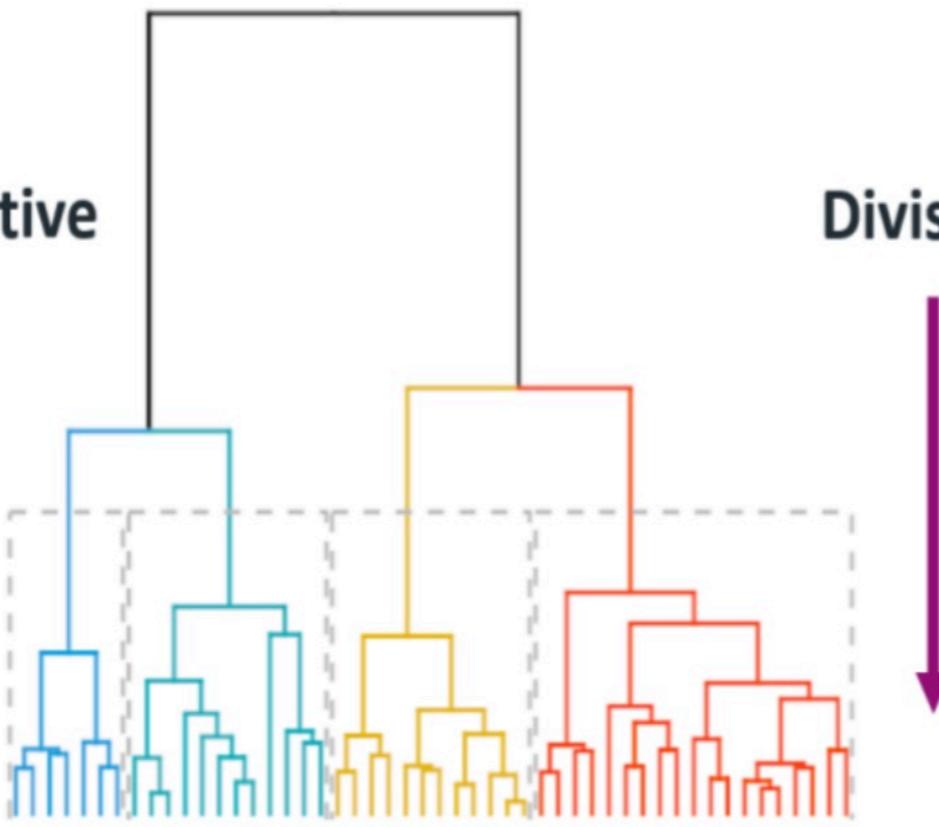
**Combines the two most similar clusters,** thereby reducing the total number of clusters by one.

### Step-2:

Repeats the previous step until only one cluster remains. The figure above shows this approach.

### Agglomerative

widely used



### Step-1:

Uses a parametric clustering algorithm such as **k-mean** to **divide the cluster** into two clusters.

### Step-2:

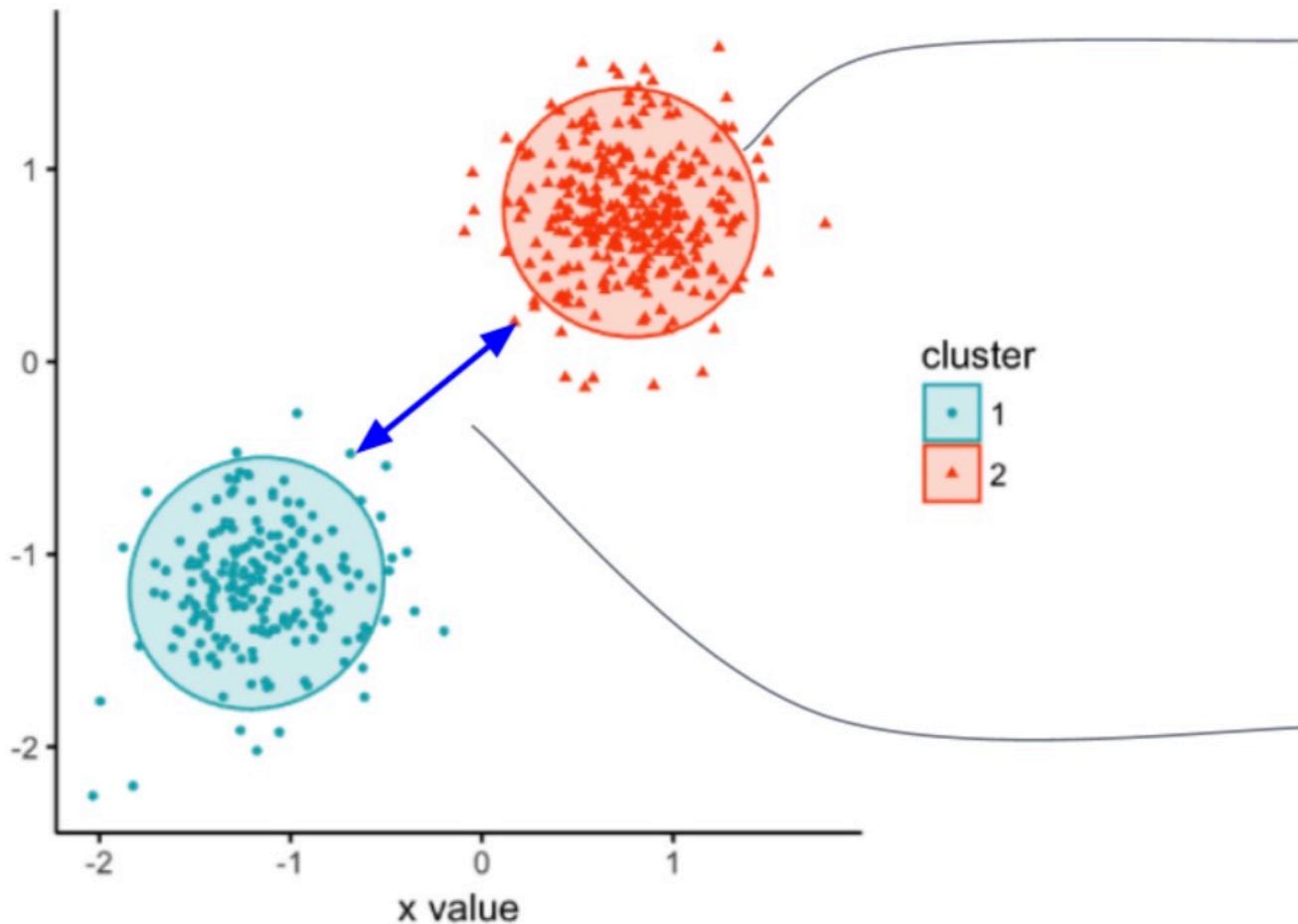
Repeats the above step for each subset until all observations are clustered as independent sets.

# Hierarchical Clustering Theory



Ideal clustering is characterised by  
***minimal intra cluster distance*** and ***maximal inter cluster distance***.

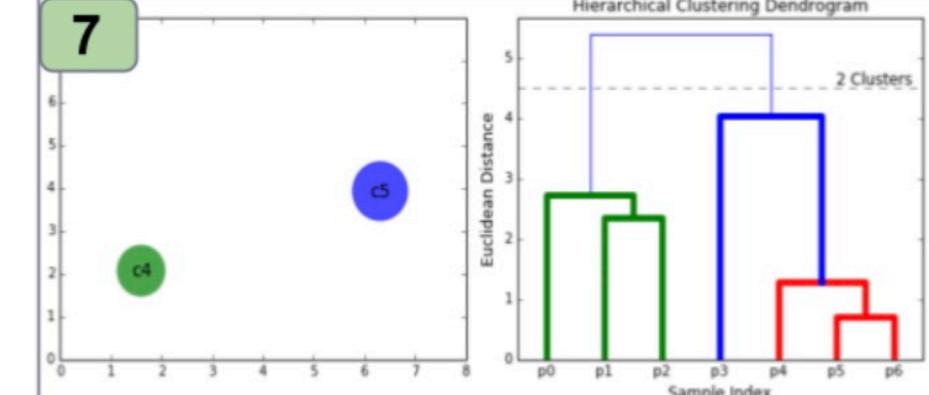
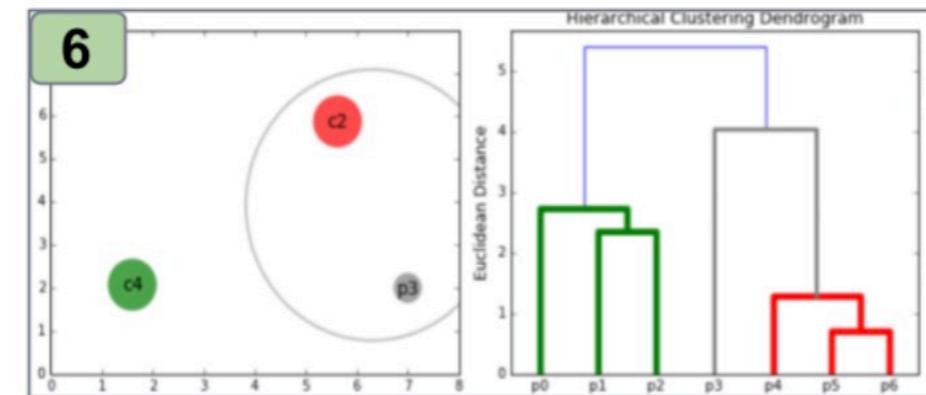
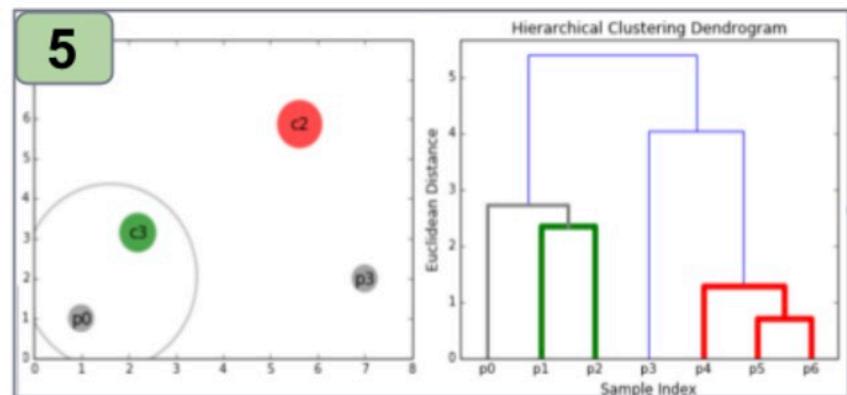
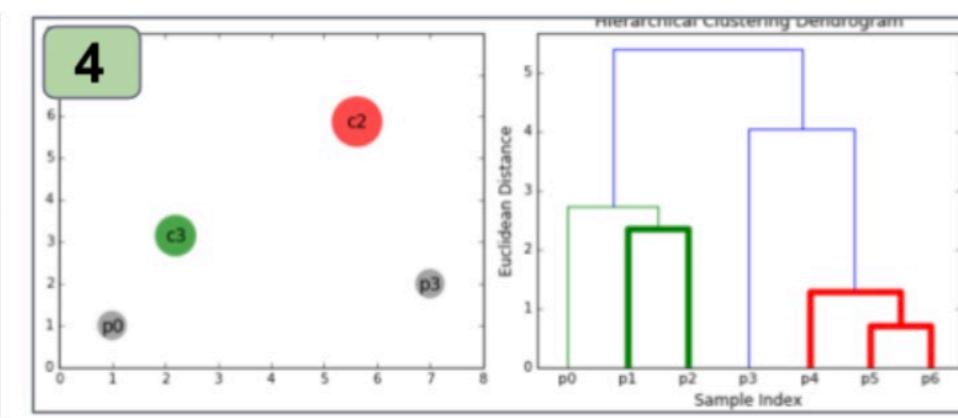
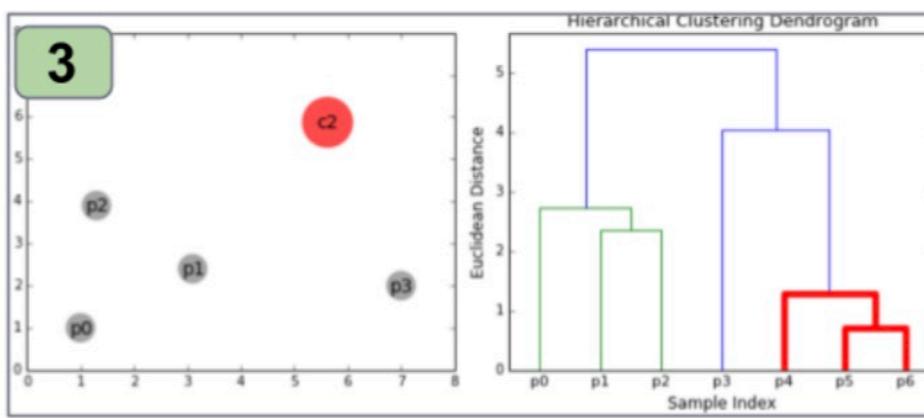
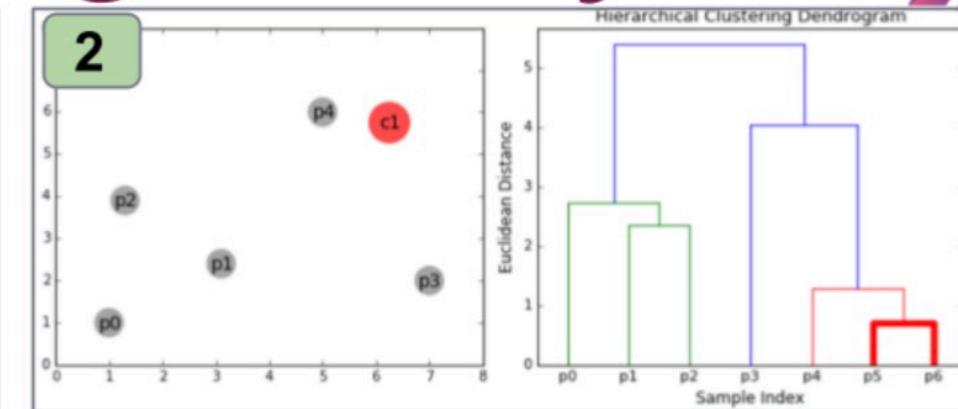
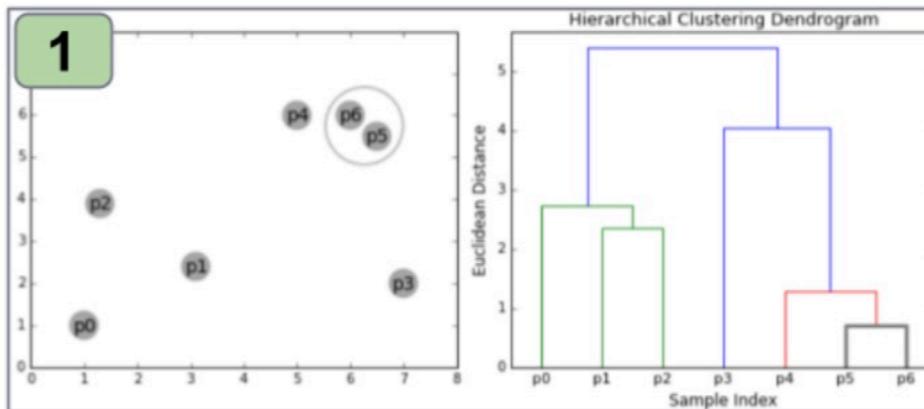
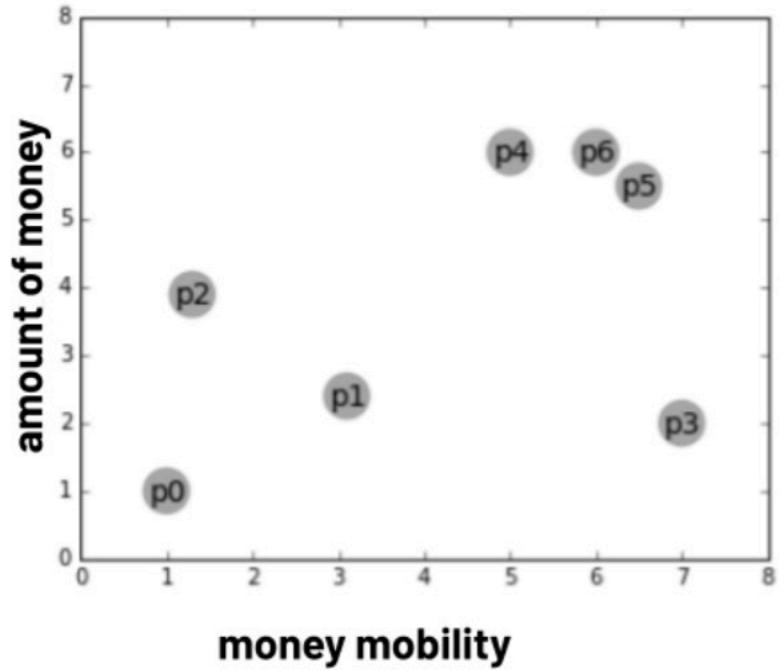
Cluster plot



***minimal intra cluster distance***

***maximal inter cluster distance***

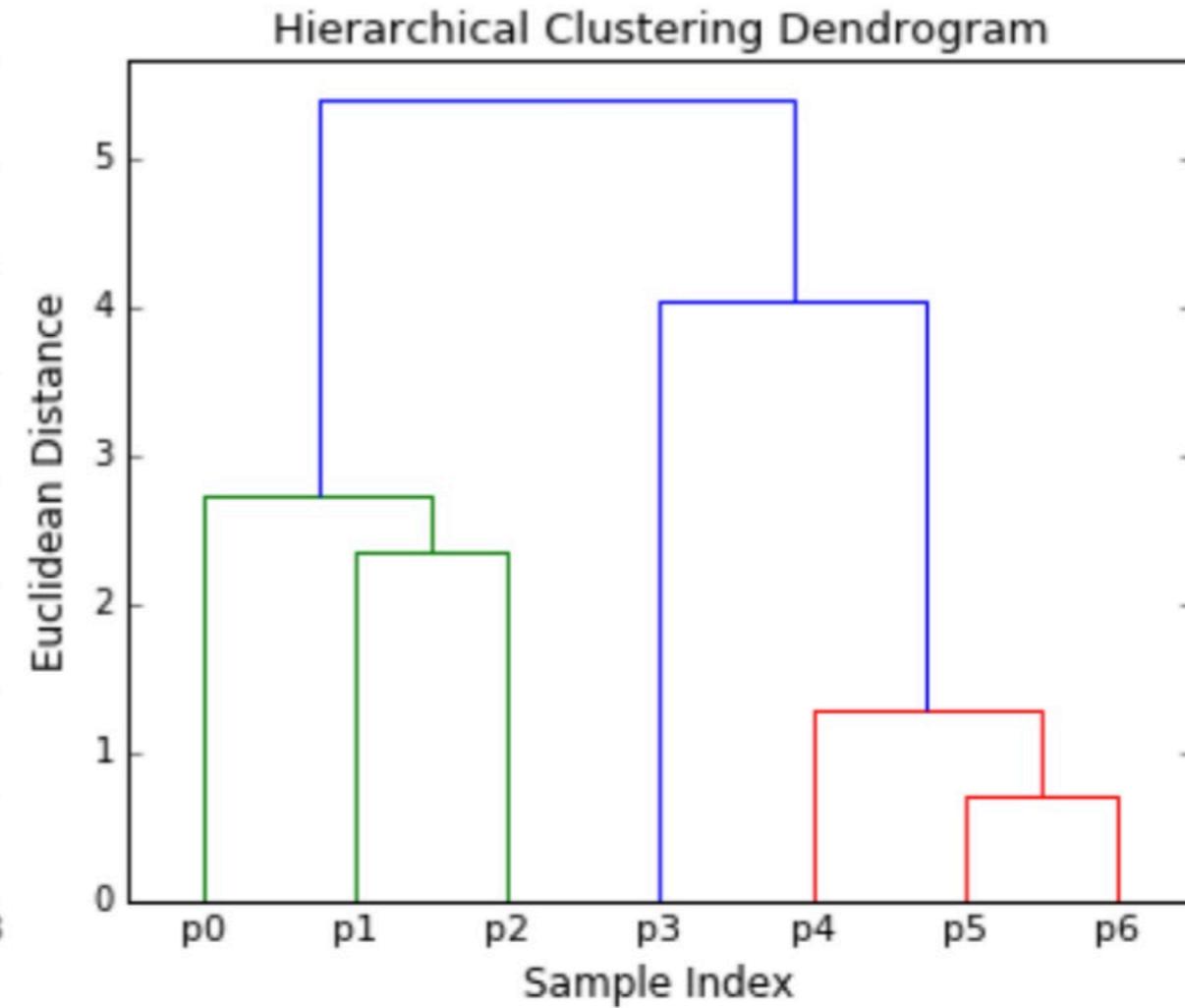
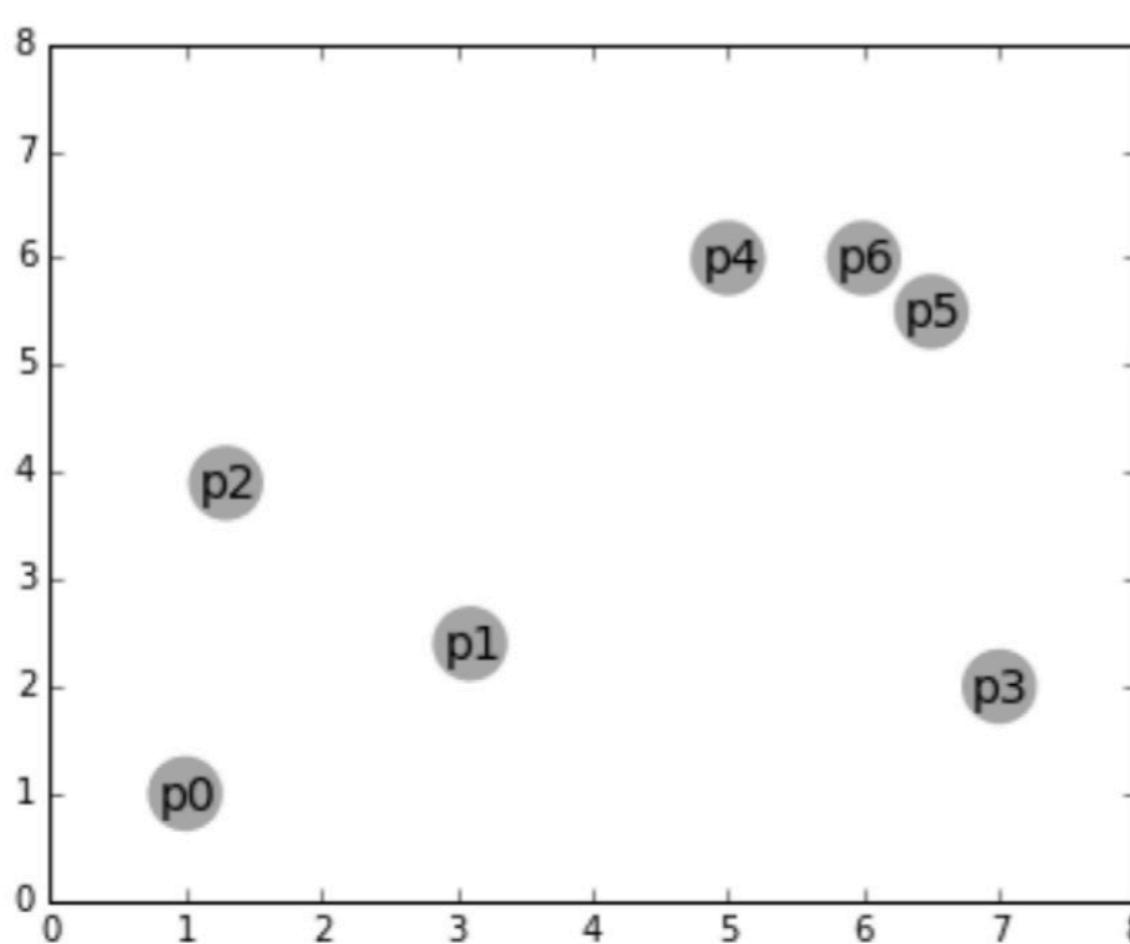
# Hierarchical Clustering Theory



# Hierarchical Clustering Theory



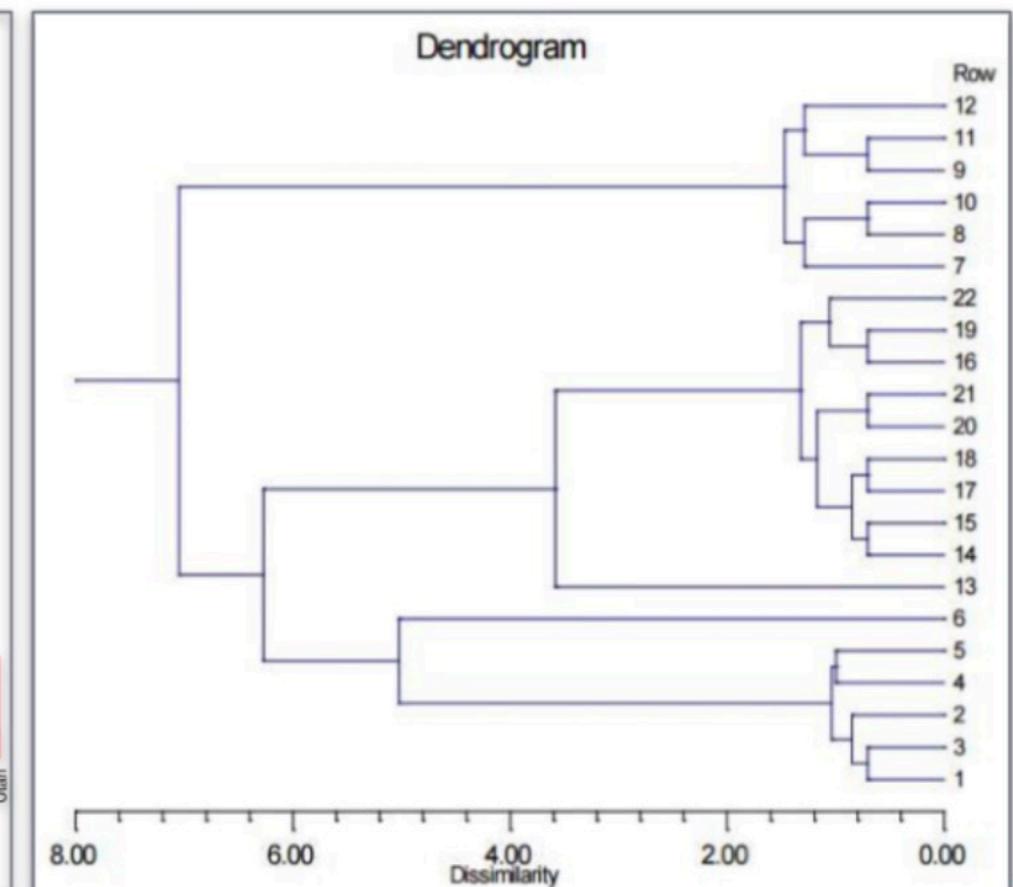
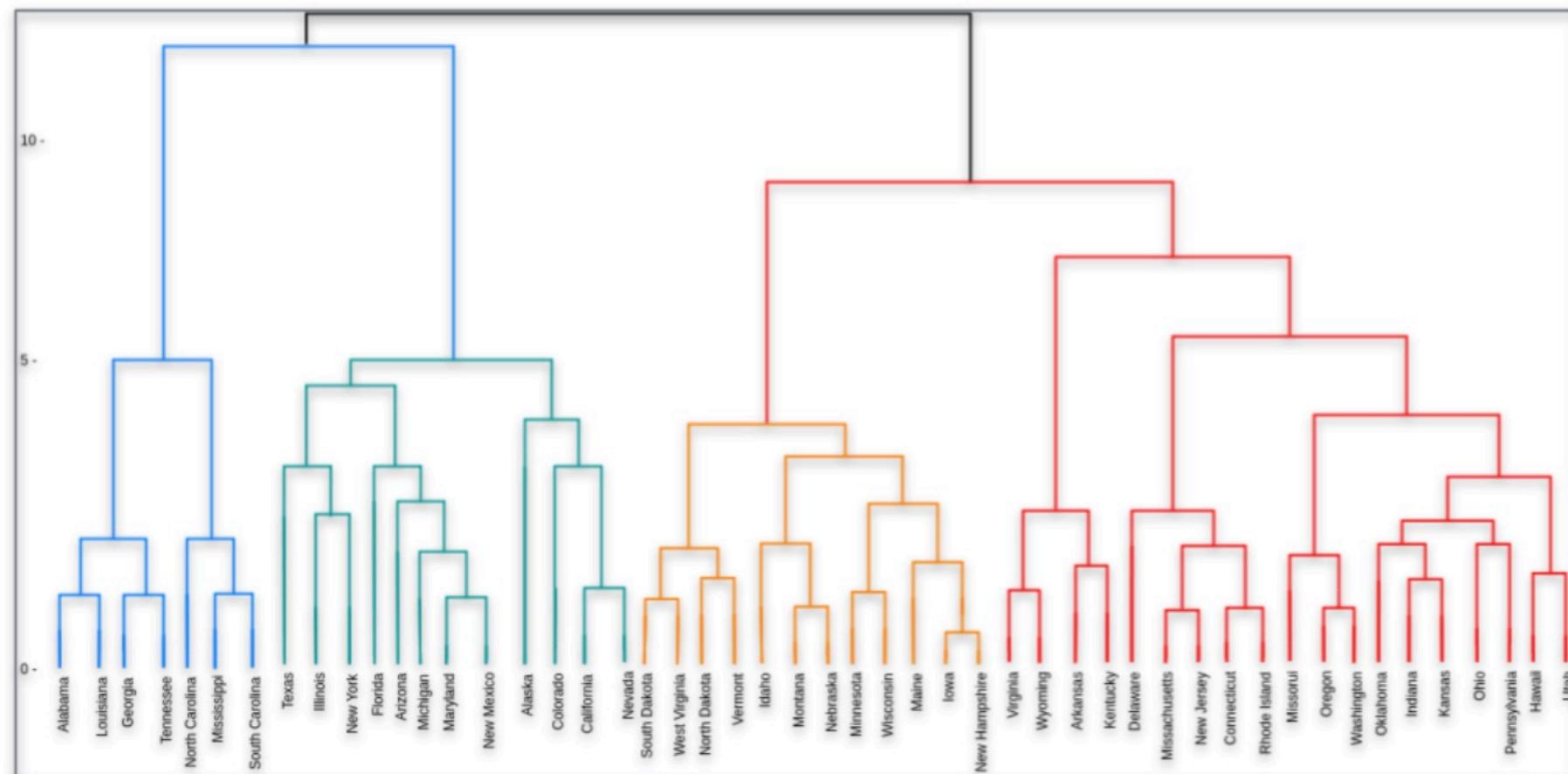
## Types of Hierarchical Clustering



# Hierarchical Clustering Theory



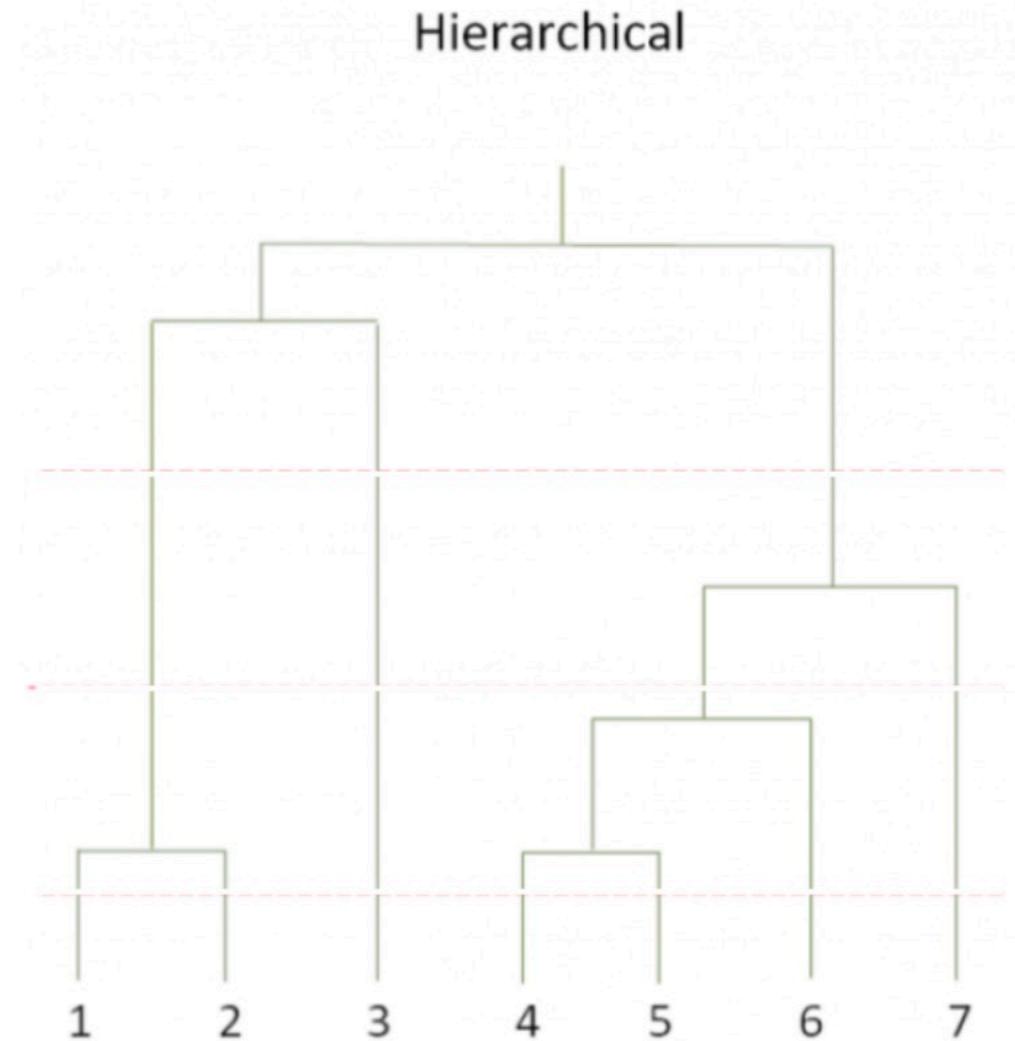
Ideal clustering is characterised by  
***minimal intra cluster distance*** and ***maximal inter cluster distance***.



# Hierarchical Clustering Theory



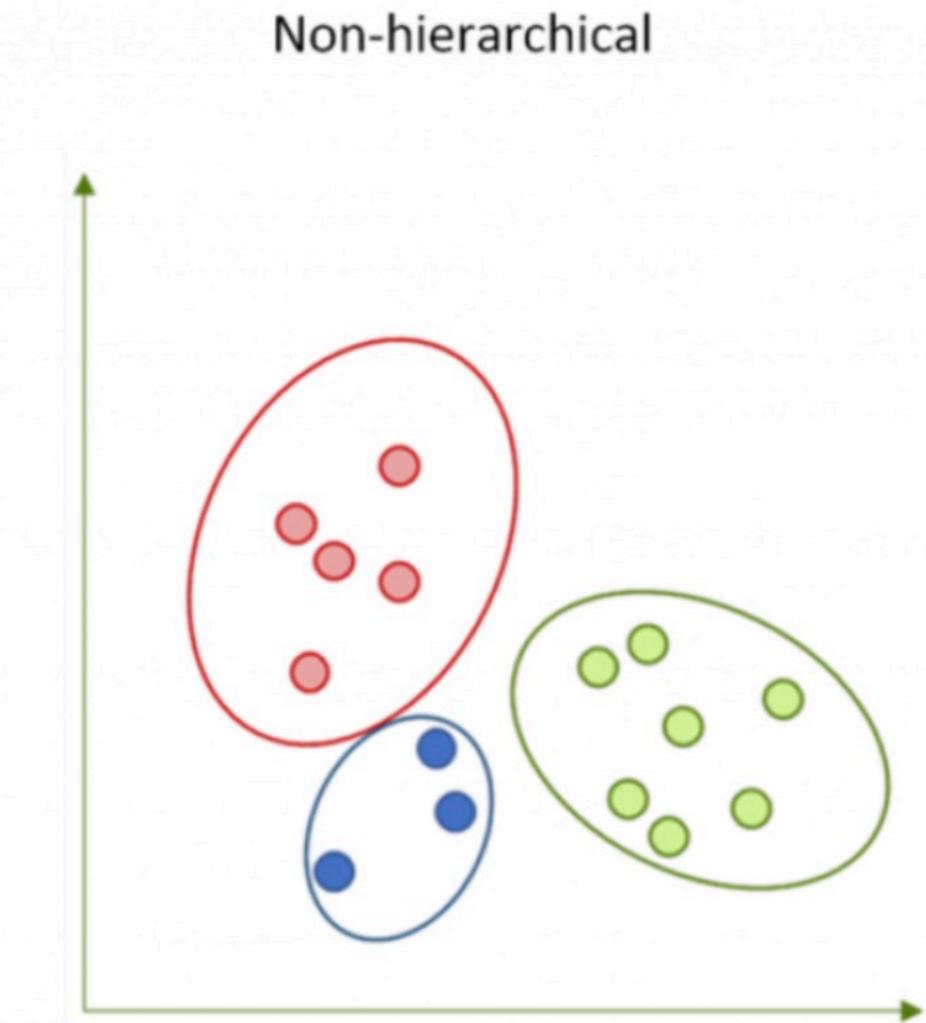
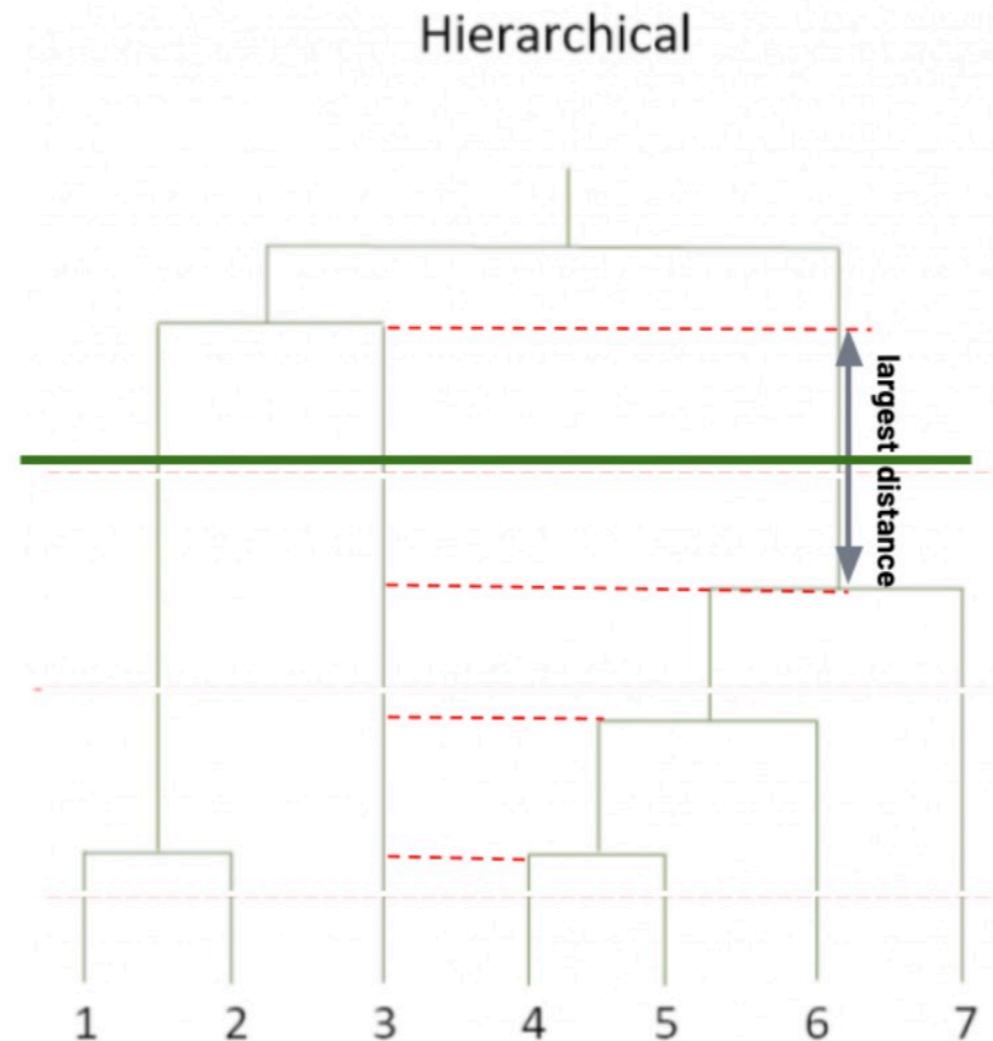
Ideal clustering is characterised by  
***minimal intra cluster distance*** and ***maximal inter cluster distance***.



# Hierarchical Clustering Theory



Ideal clustering is characterised by  
***minimal intra cluster distance*** and ***maximal inter cluster distance***.



# Hierarchical Clustering Theory



## Hierarchical Algorithm:

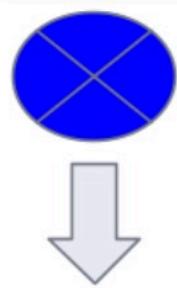
### Hyperparameters

**“linkage” parameter: (default= “ward”)**

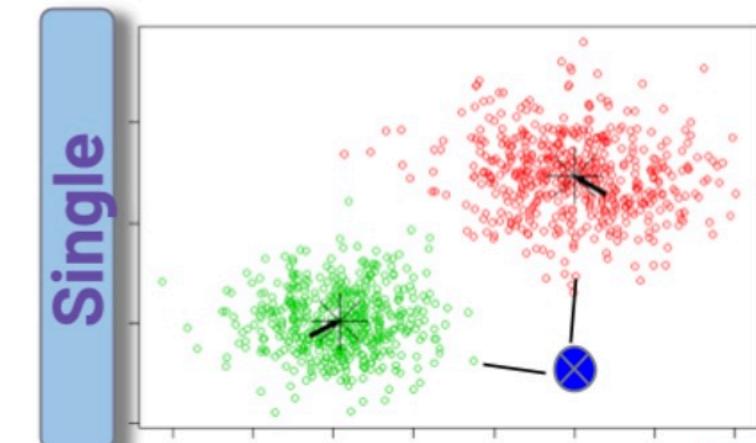
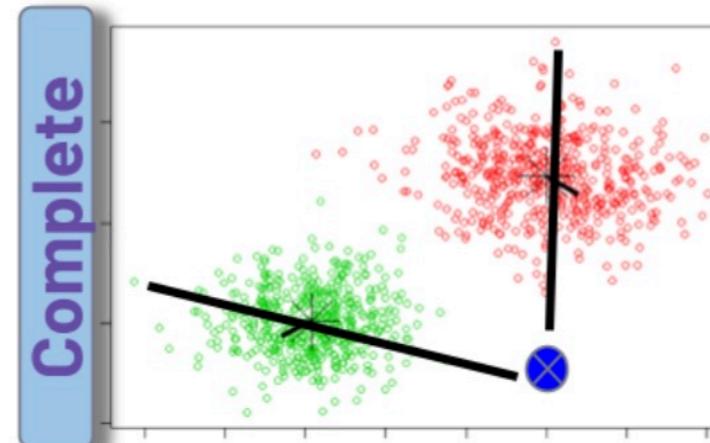
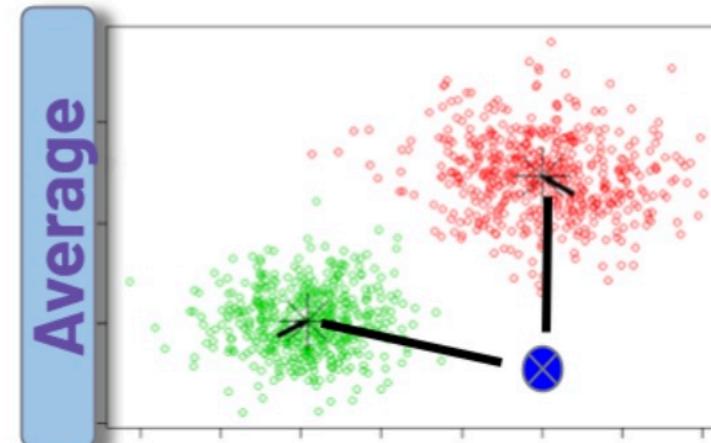
{‘ward’, ‘complete’, ‘average’, ‘single’}

Which **linkage criterion** to use. The linkage criterion determines which distance to use between sets of observation.

- **Ward** minimizes the variance of the clusters being merged.
- **Average** uses the average of the distances of each observation of the two sets.
- **Complete** or **maximum** linkage uses the maximum distances between all observations of the two sets.
- **Single** uses the minimum of the distances between all observations of the two sets.



New Data



# Hierarchical Clustering Theory



## Hierarchical Algorithm:

### Hyperparameters

#### `sklearn.cluster.AgglomerativeClustering`

```
class sklearn.cluster.AgglomerativeClustering(n_clusters=2, *, affinity='euclidean', memory=None,  
connectivity=None, compute_full_tree='auto', linkage='ward', distance_threshold=None) [source]
```

**“affinity” parameter: (default= “euclidean”)**  
{'euclidean', 'manhattan', 'cosine', 'precomputed'}

*Metric used to compute the linkage. If linkage is “ward”, only “euclidean” is accepted.*

# Hierarchical Clustering Theory



*Be ready for*  
**Hierarchical  
Clustering  
Python  
Session**