



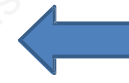
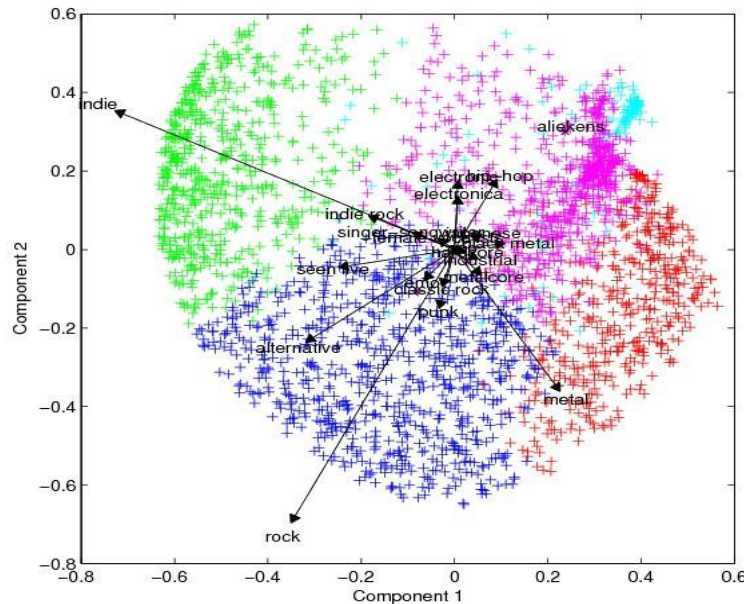
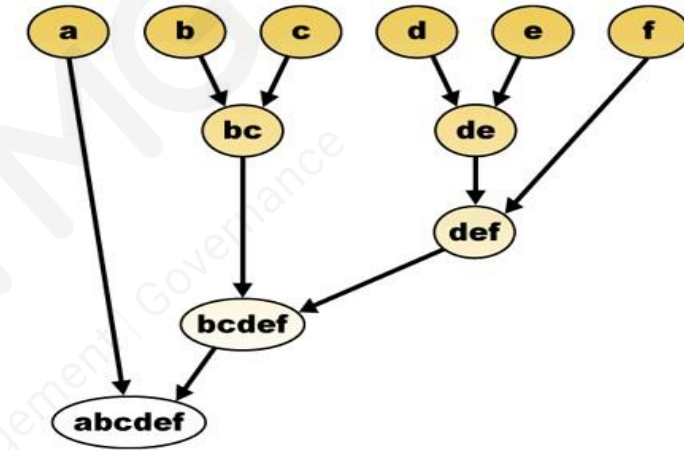
# Clustering / Segmentation

- Cluster Analysis (“*data segmentation*”) is an exploratory method for identifying homogenous groups (“clusters”) of records
- Similar records should belong to the same cluster
- Dissimilar records should belong to different clusters

## Hierarchical methods- Agglomerative:

Begin with  $n$  clusters; sequentially merge similar clusters until 1 cluster is left.

- Useful when goal is to arrange the clusters into a natural hierarchy
- Requires specifying distance measure



## Non-Hierarchical methods:

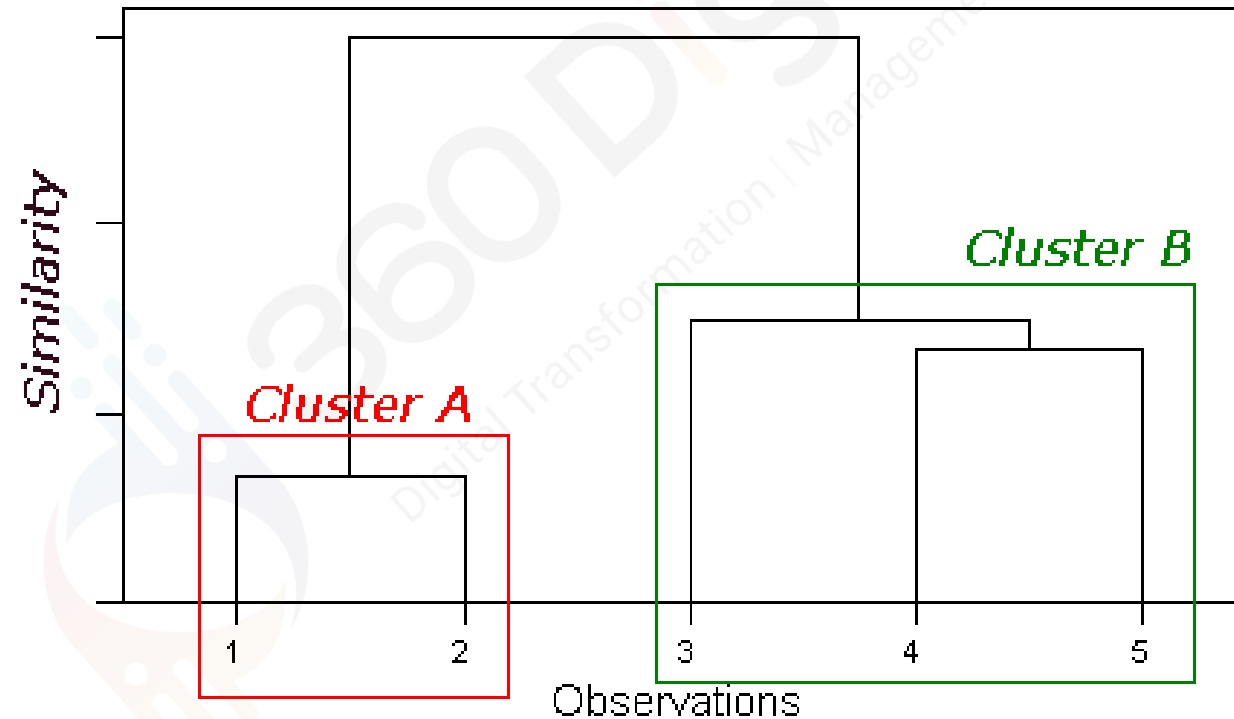
Pre-specified number of clusters, assign records to each of the clusters.

- Requires specifying # of clusters
- Computationally cheap

# Hierarchical Clustering

**Dendrogram:** Tree-like diagram that summarize the clustering process

- Similar records joined by links
- Record location determined by similarity to other records





<https://www.linkedin.com/company/360-digitmg/>

<https://www.facebook.com/pg/360Digitmg/reviews/>