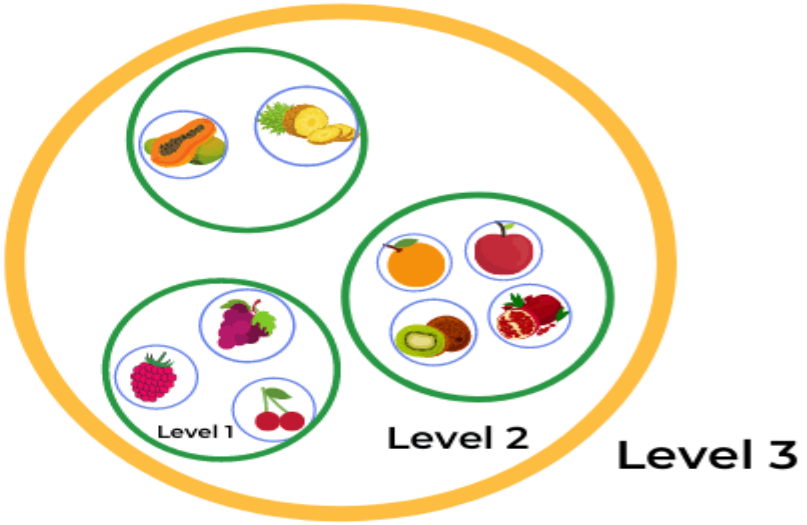


Hierarchal Clustering Algorithm

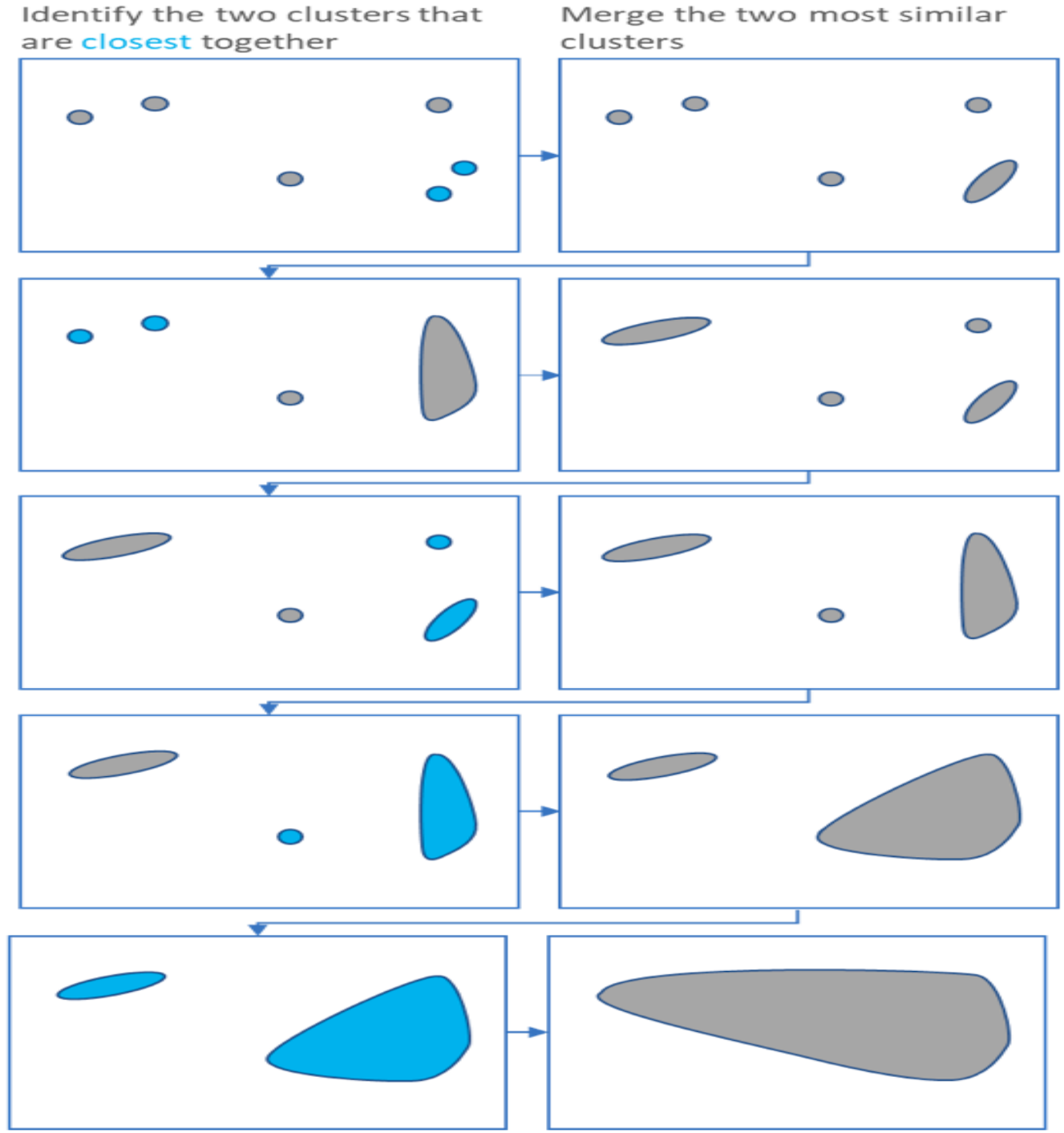
LEC.ASSIST : AHMED YOUSRY

What is Hierarchal clustering

- Is an algorithm that groups similar objects into groups called *clusters*.
- The endpoint is a set of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other.

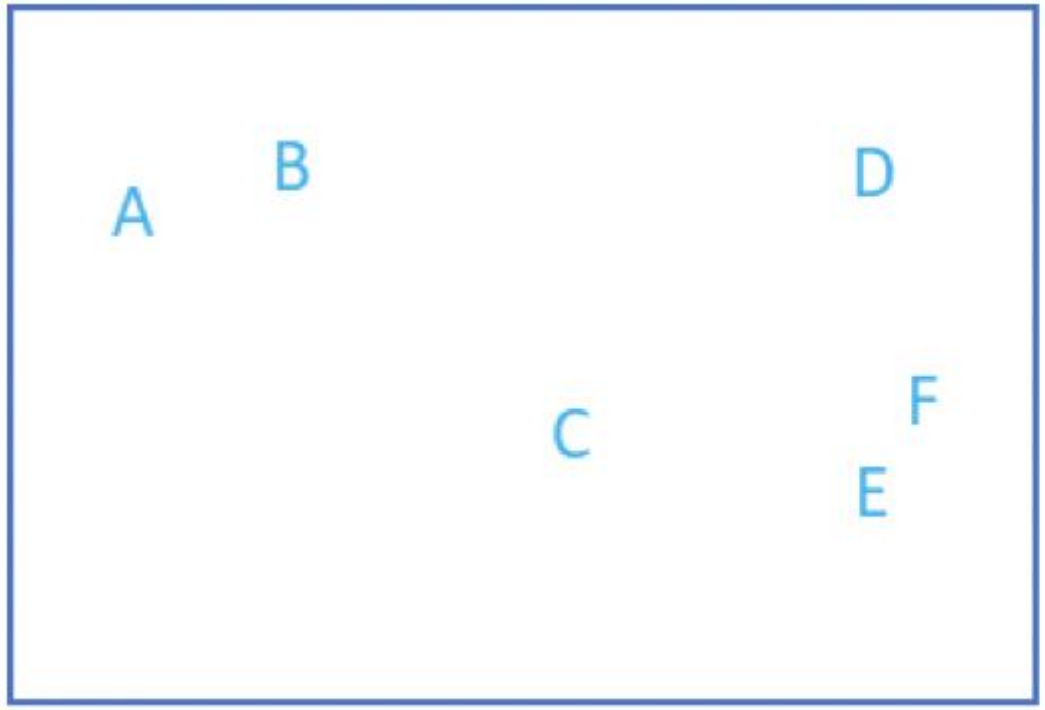
How it works

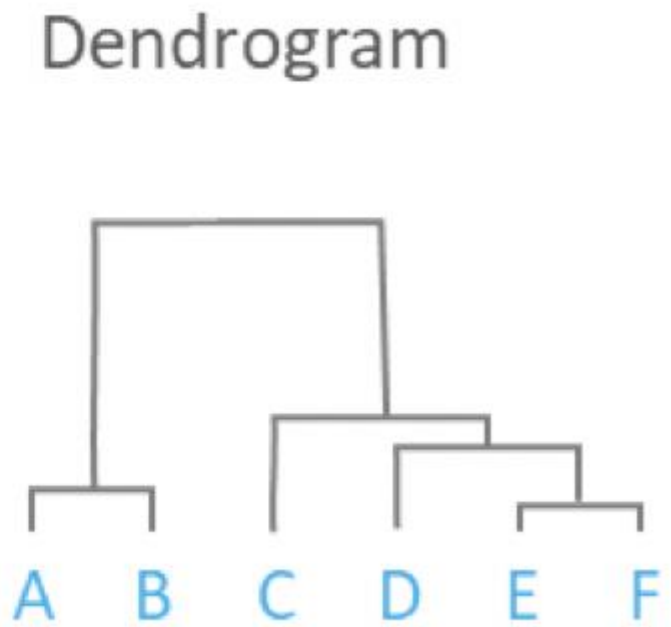
- Hierarchical clustering starts by treating each observation as a separate cluster.
- Then, it repeatedly executes the following two steps:
  - Identify the two clusters that are closest together.
  - Merge the two most similar clusters.
    - This iterative process continues until all the clusters are merged together.


Identify the two clusters that are closest together
Merge the two most similar clusters

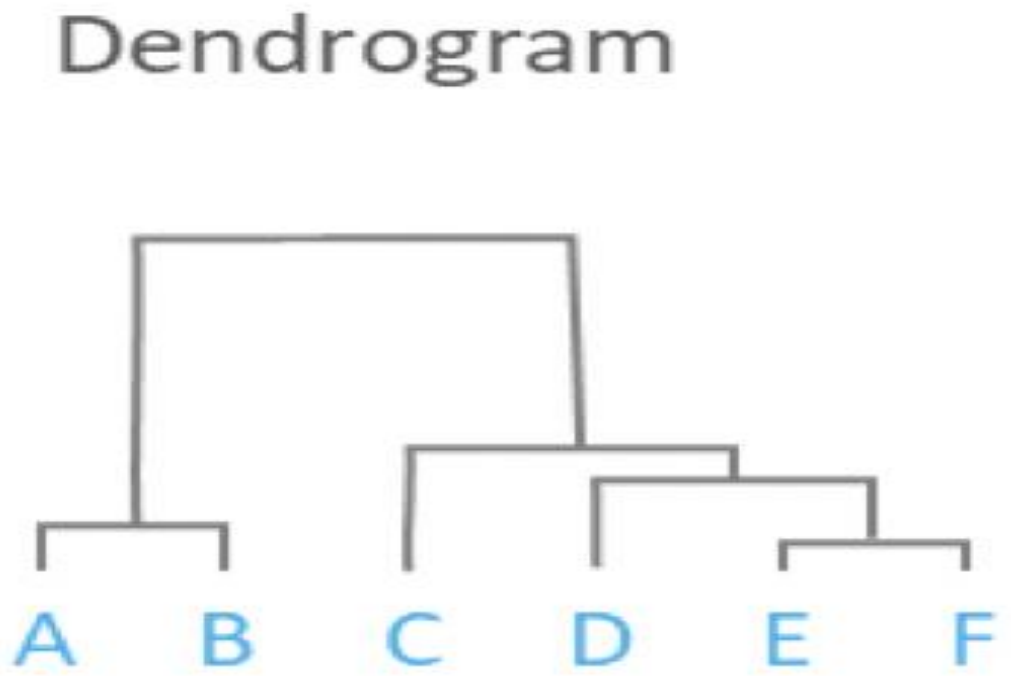
The output

The main output of Hierarchical Clustering is a *dendrogram*, which shows the hierarchical relationship between the clusters:

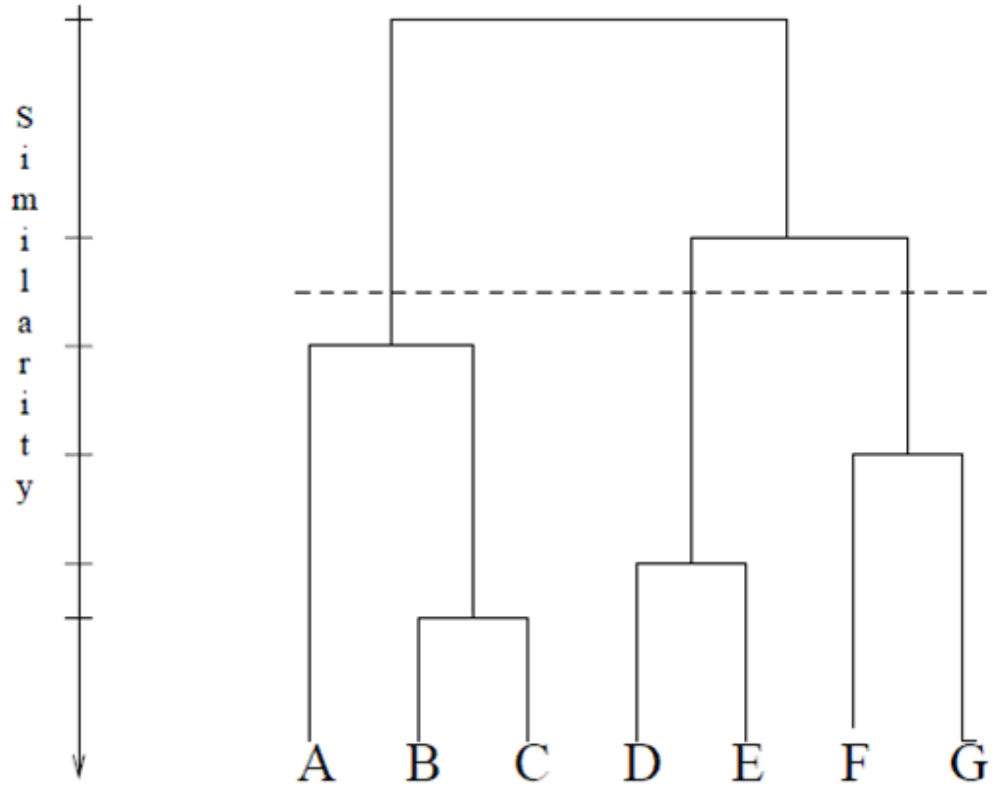




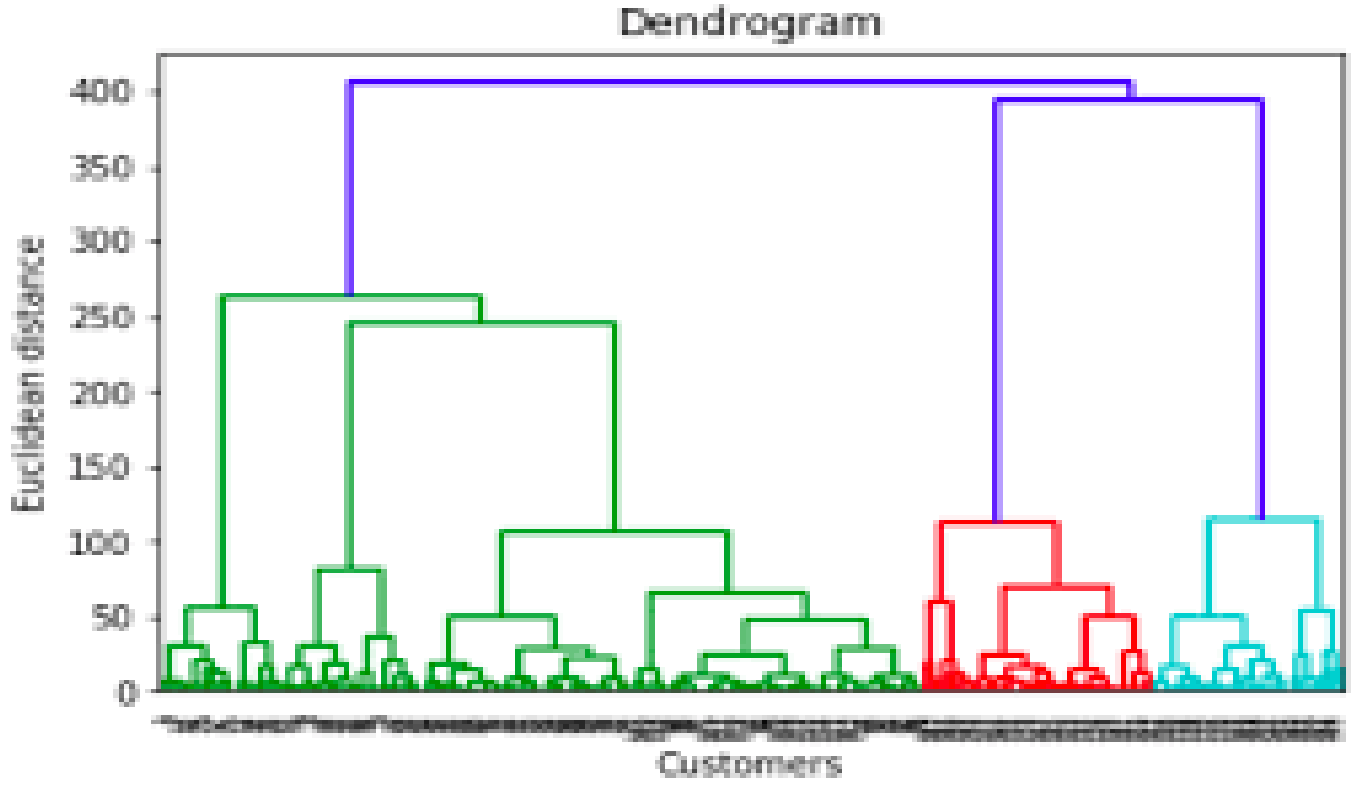
What is Number of clusters

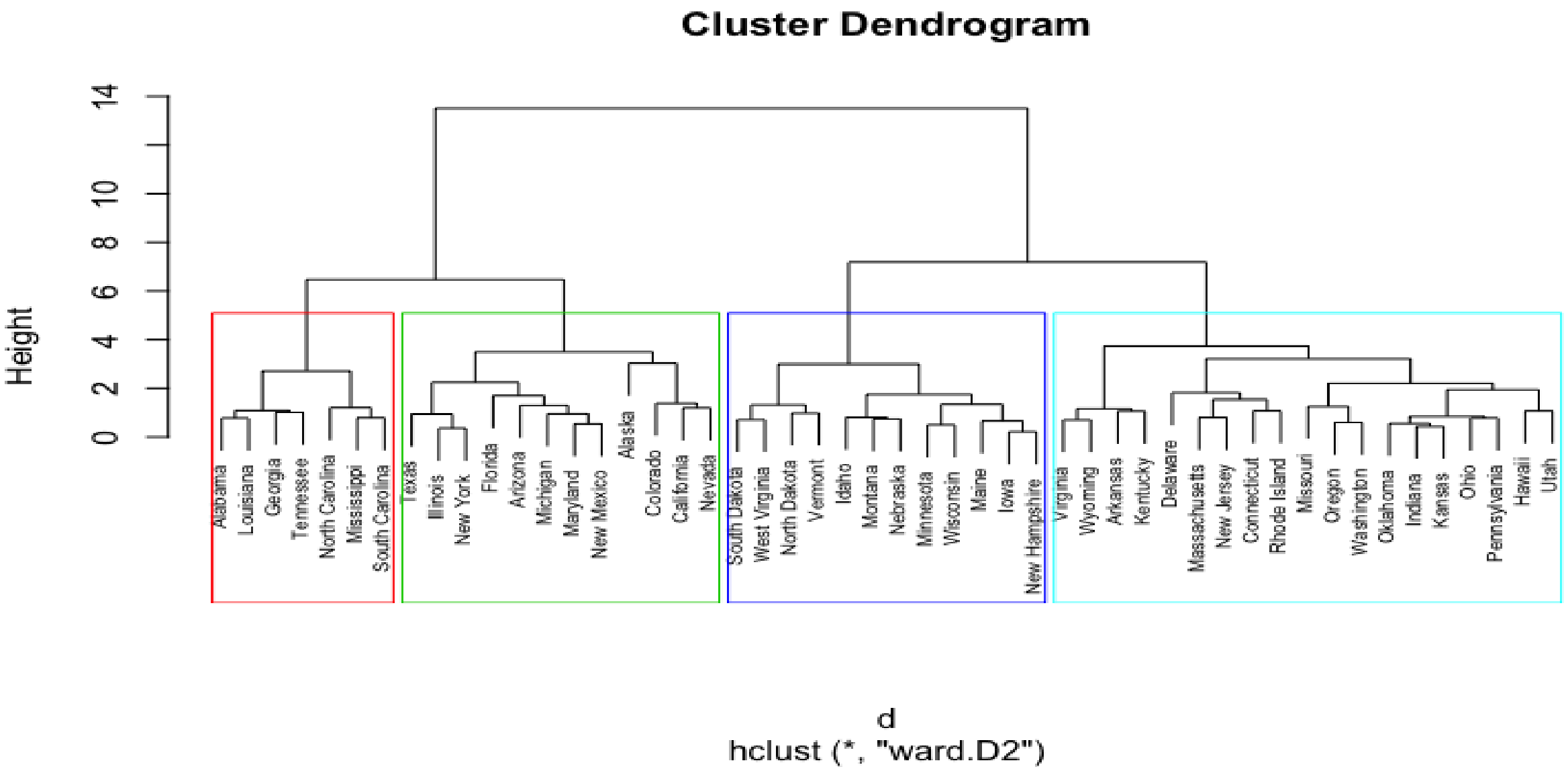


Another example



Another example



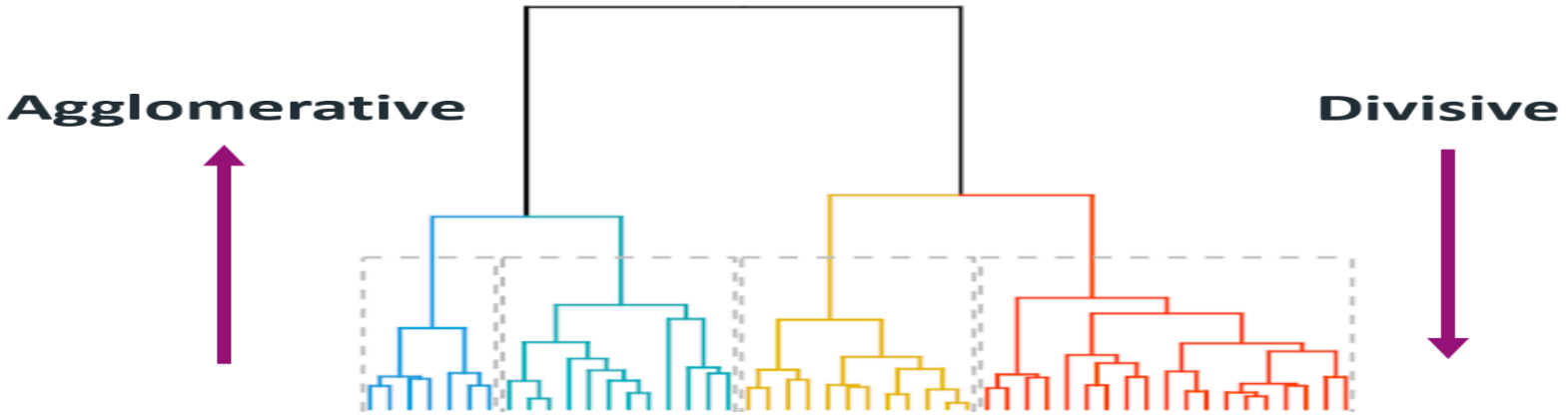

Cluster Dendrogram
Height
0
2
4
6
8
10
14
Alabama
Louisiana
Georgia
Tennessee
North Carolina
Mississippi
South Carolina
Texas
Illinois
New York
Florida
Arizona
Michigan
Maryland
New Mexico
Alaska
Colorado
California
Nevada
South Dakota
West Virginia
North Dakota
Vermont
Idaho
Montana
Nebraska
Minnesota
Wisconsin
Maine
Iowa
New Hampshire
Virginia
Wyoming
Arkansas
Kentucky
Delaware
Massachusetts
New Jersey
Connecticut
Rhode Island
Missouri
Oregon
Washington
Oklahoma
Indiana
Kansas
Ohio
Pennsylvania
Hawaii
Utah
d
hclust (*, "ward.D2")

Measures of distance (similarity)

- In the previous example, the *distance* between two clusters has been computed based on the length of the straight line drawn from one cluster to another.
- This is commonly referred to as the *Euclidean distance*.
- Many other *distance metrics* have been developed.

Agglomerative Vs. divisive algorithms

- Hierarchical clustering typically works by sequentially merging similar clusters, as shown above. This is known as *agglomerative hierarchical clustering*.
- In theory, it can also be done by initially grouping all the observations into one cluster, and then successively splitting these clusters.
- This is known as *divisive hierarchical clustering*.
- Divisive clustering is rarely done in practice.



Thanks