Hierarchical Clustering

Dr.Mydhili K Nair
&
Dr.Pushpalatha MN
ISE Dept.
M S Ramaiah Institute of Technology

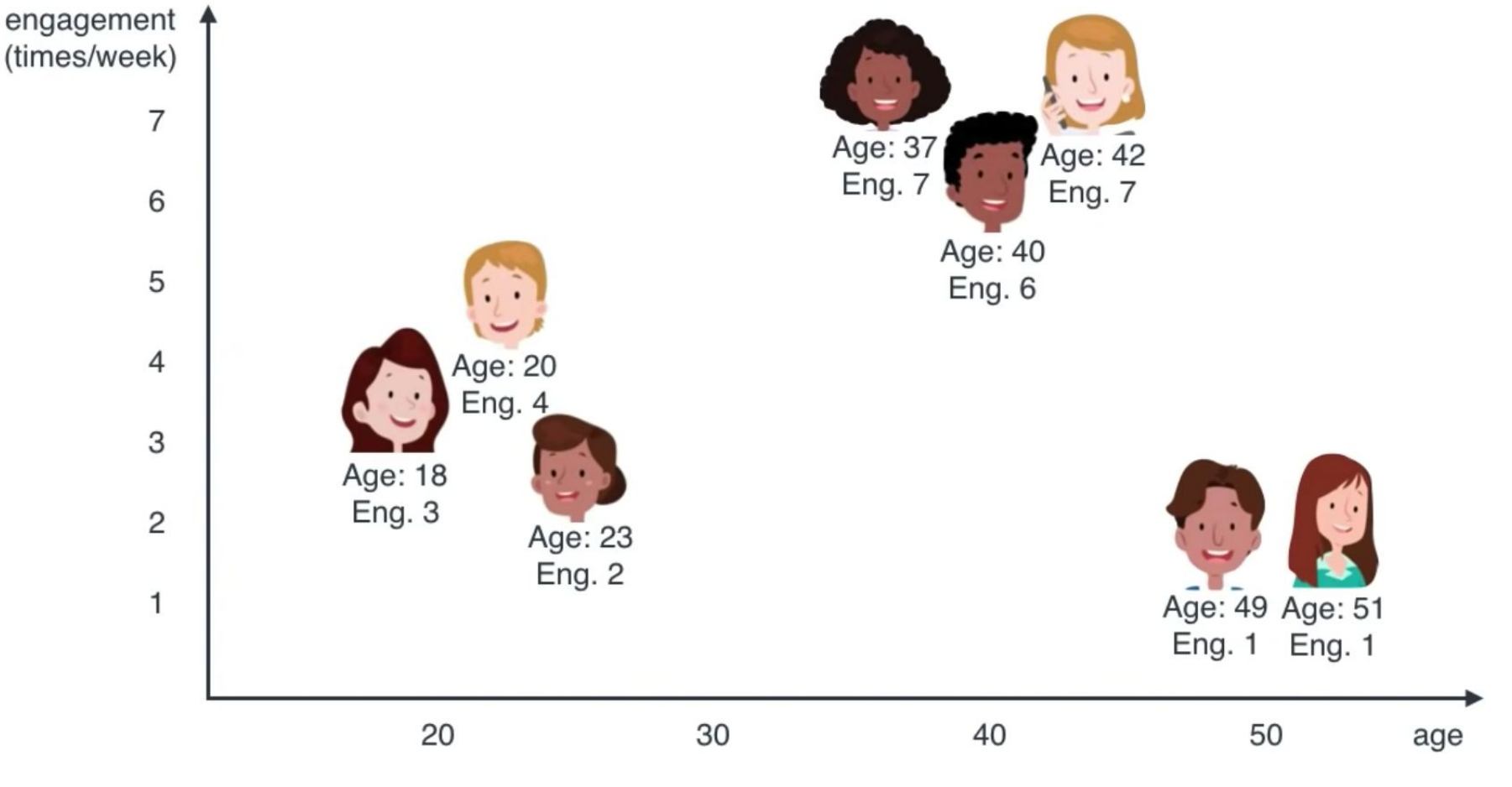
Customer Segmentation

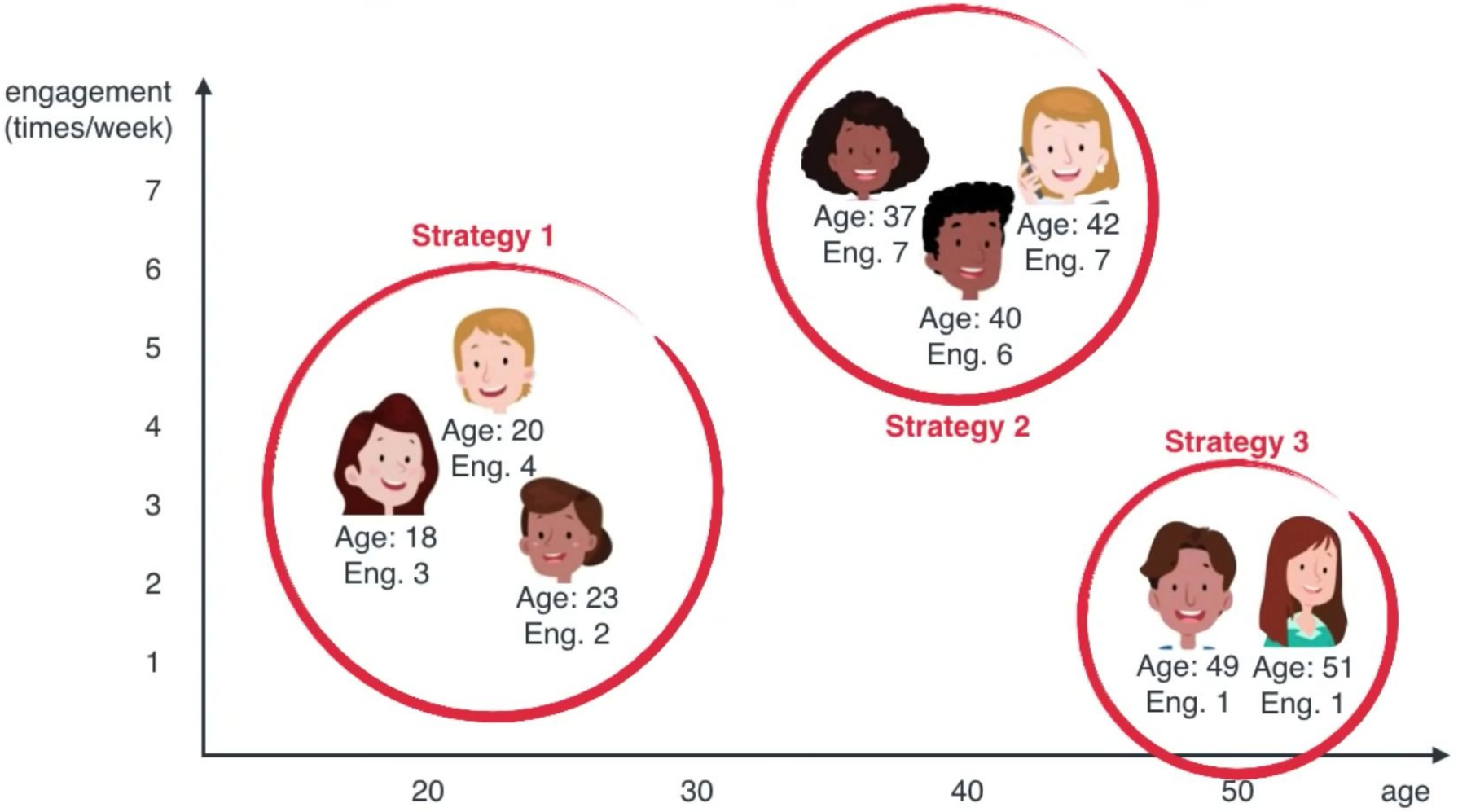
Goal: To make 3 marketing strategies

Age (in years)

Engagement with the page (in days/week)

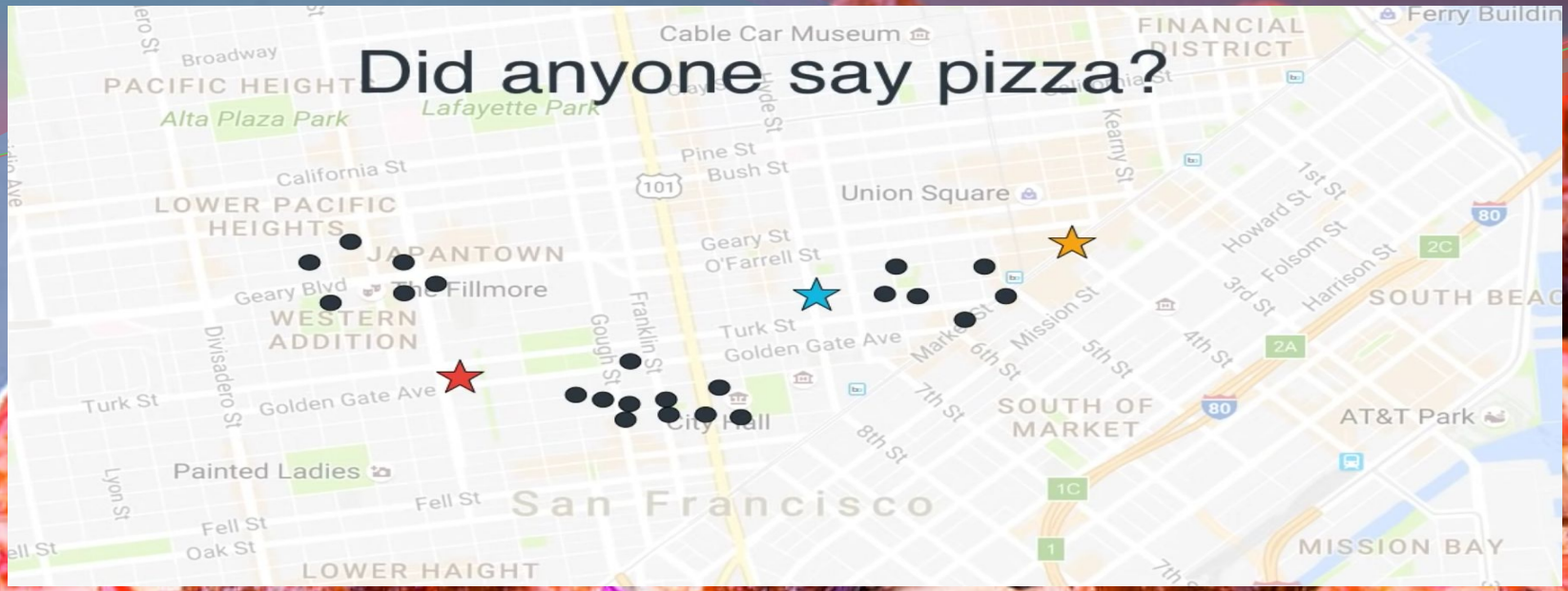
Source: https://www.youtube.com/watch?v=QXOkPvFM6NU


engagement
(times/week)
7
6
5
4
3
2
1
Age: 37
Eng. 7
Age: 42
Eng. 7
Age: 40
Eng. 6
Age: 20
Eng. 4
Age: 18
Eng. 3
Age: 23
Eng. 2
Age: 49
Eng. 1
Age: 51
Eng. 1
20
30
40
50
age


engagement
(times/week)
7
6
5
4
3
2
1
Strategy 1
Age: 20
Eng. 4
Age: 18
Eng. 3
Age: 23
Eng. 2
Age: 37
Eng. 7
Age: 42
Eng. 7
Age: 40
Eng. 6
Strategy 2
Strategy 3
Age: 49
Eng. 1
Age: 51
Eng. 1
20
30
40
50
age

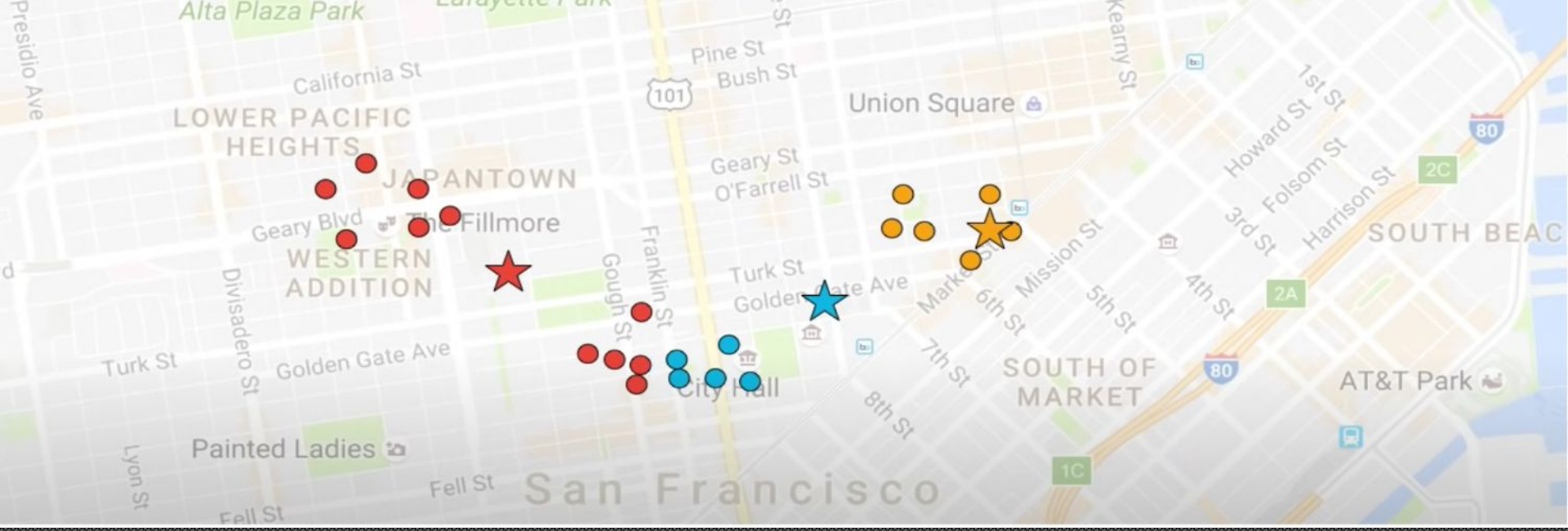
Did anyone say pizza?

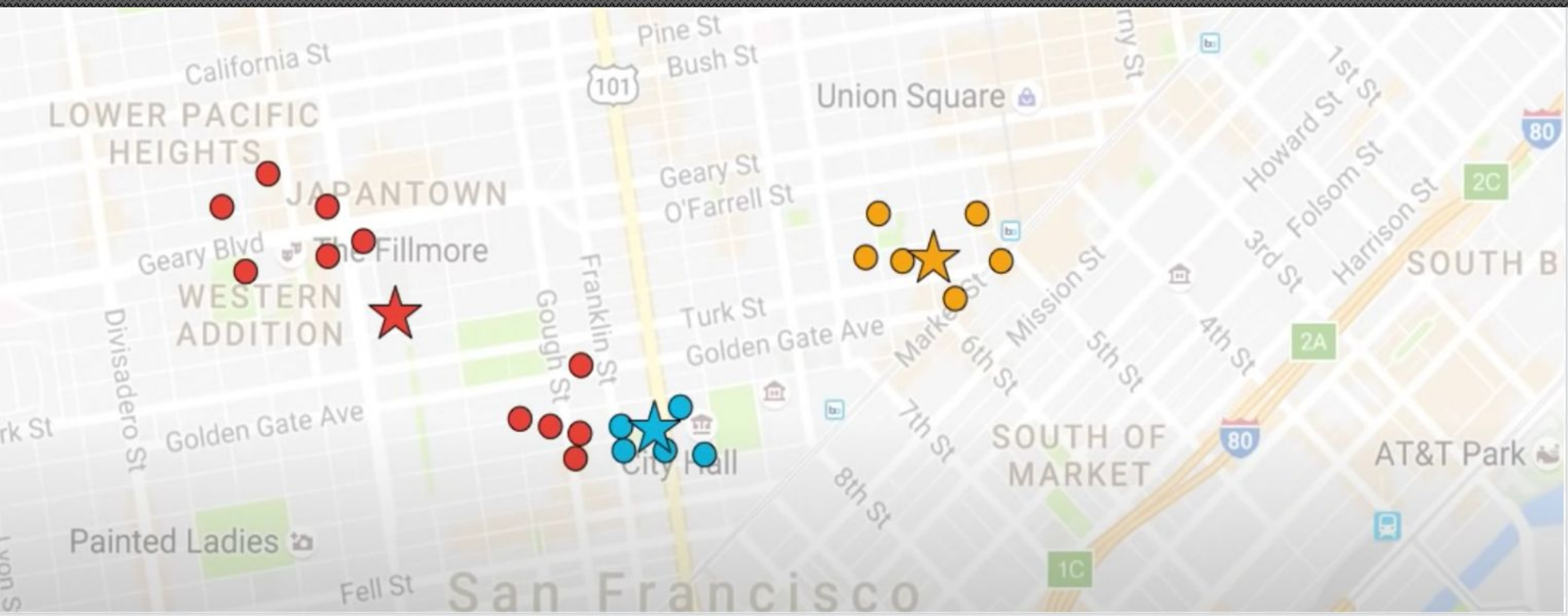
Did anyone say pizza?

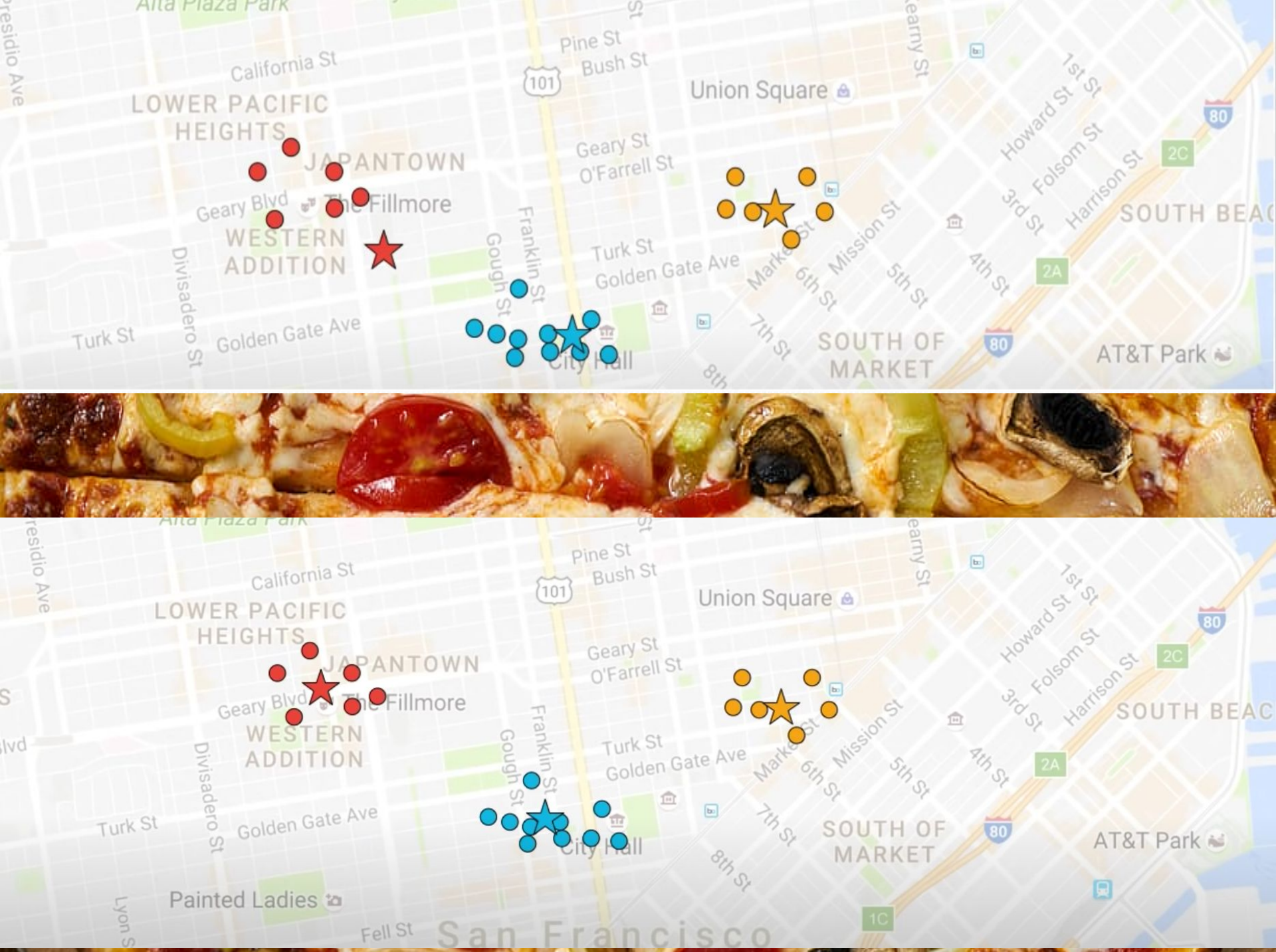

Did anyone say pizza?
Cable Car Museum
FINANCIAL DISTRICT
Ferry Buildin
Broadway
PACIFIC HEIGHTS
Alta Plaza Park
Lafayette Park
California St
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN ADDITION
Divisadero St
Gough St
Franklin St
Turk St
Golden Gate Ave
Pine St
Bush St
Union Square
Geary St
O'Farrell St
Kearny St
1st St
Howard St
Folsom St
3rd St
Harrison St
SOUTH BEAC
Market St
6th St
Mission St
5th St
4th St
7th St
8th St
SOUTH OF MARKET
AT&T Park
City Hall
Painted Ladies
Lyon St
Fell St
Oak St
San Francisco
LOWER HAIGHT
MISSION BAY
101
80
2C
2A
1C
1


Alta Plaza Park
Lafayette Park
Presidio Ave
California St
Pine St
Bush St
101
Union Square
Kearny St
1st St
Howard St
Folsom St
3rd St
Harrison St
80
2C
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary St
O'Farrell St
Geary Blvd
The Fillmore
WESTERN ADDITION
Franklin St
Gough St
Turk St
Golden Gate Ave
Market St
Mission St
6th St
5th St
4th St
2A
SOUTH BEAC
Divisadero St
Turk St
Golden Gate Ave
City Hall
7th St
8th St
SOUTH OF MARKET
AT&T Park
Painted Ladies
Lyon St
Fell St
San Francisco
1C


Did anyone say pizza?
Cable Car Museum
FINANCIAL DISTRICT
Ferry Building
Broadway
PACIFIC HEIGHTS
Alta Plaza Park
Lafayette Park
Presidio Ave
California St
Pine St
Bush St
Union Square
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN ADDITION
Geary St
O'Farrell St
Divisadero St
Gough St
Franklin St
Turk St
Golden Gate Ave
Market St
Mission St
Howard St
Folsom St
Harrison St
1st St
3rd St
4th St
5th St
6th St
7th St
8th St
Kearny St
SOUTH BEACH
SOUTH OF MARKET
City Hall
AT&T Park
Painted Ladies
Fell St
Lyon St
San Francisco
101
80
2C
2A
1C

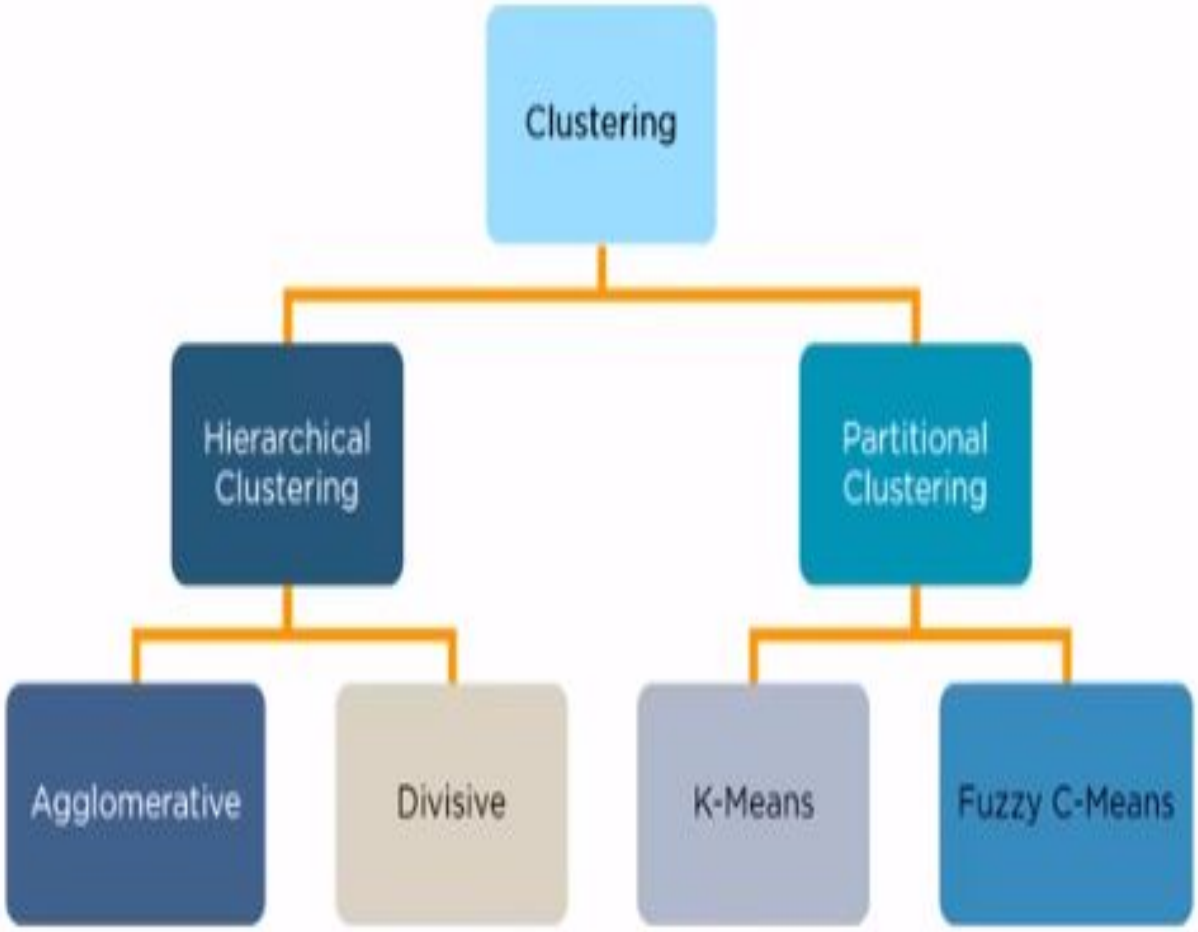

Alta Plaza Park
Presidio Ave
California St
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN ADDITION
Divisadero St
Turk St
Golden Gate Ave
Gough St
Franklin St
Pine St
Bush St
101
Union Square
Geary St
O'Farrell St
Kearny St
1st St
Howard St
Folsom St
3rd St
Harrison St
80
2C
SOUTH BEAC
Mission St
Market St
6th St
5th St
4th St
2A
7th St
SOUTH OF MARKET
AT&T Park
8th St
City Hall
Painted Ladies
Lyon St
Fell St
1C
San Francisco


California St
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN ADDITION
Divisadero St
Golden Gate Ave
Gough St
Franklin St
Pine St
Bush St
101
Union Square
Geary St
O'Farrell St
Howard St
1st St
Folsom St
3rd St
Harrison St
80
2C
SOUTH B
Mission St
Market St
6th St
5th St
4th St
2A
7th St
SOUTH OF MARKET
AT&T Park
8th St
City Hall
Turk St
Golden Gate Ave
Painted Ladies
Fell St
1C
San Francisco


Alta Plaza Park
California St
Pine St
Bush St
LOWER PACIFIC HEIGHTS
JAPANTOWN
Geary Blvd
The Fillmore
WESTERN ADDITION
Geary St
O'Farrell St
Union Square
Howard St
1st St
Folsom St
3rd St
Harrison St
SOUTH BEAC
Divisadero St
Gough St
Franklin St
Turk St
Golden Gate Ave
Market St
Mission St
6th St
5th St
4th St
7th St
8th St
SOUTH OF MARKET
AT&T Park
City Hall
Painted Ladies
Fell St
San Francisco
101
80
2C
2A
1C

Clustering & types

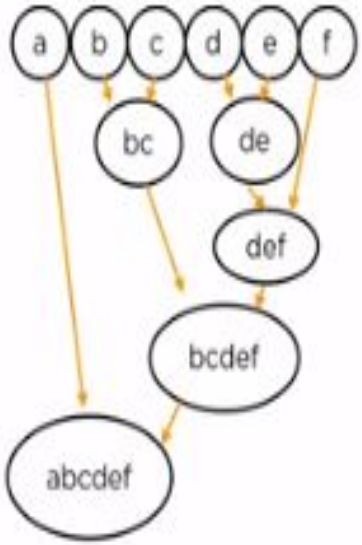
– Unsupervised learning
– Requires data, but no labels
– Detect patterns e.g. in
  • Group emails or search results
  • Customer shopping patterns
  • Regions of images
– Useful when don't know what you're looking for
– But: can get gibberish

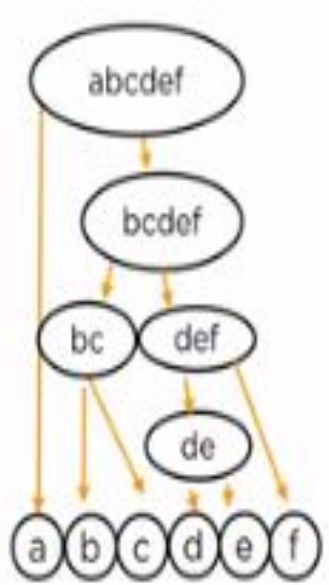


Hierarchical Clustering

✔Clusters have a tree like structure

✔Has a parent child relationship





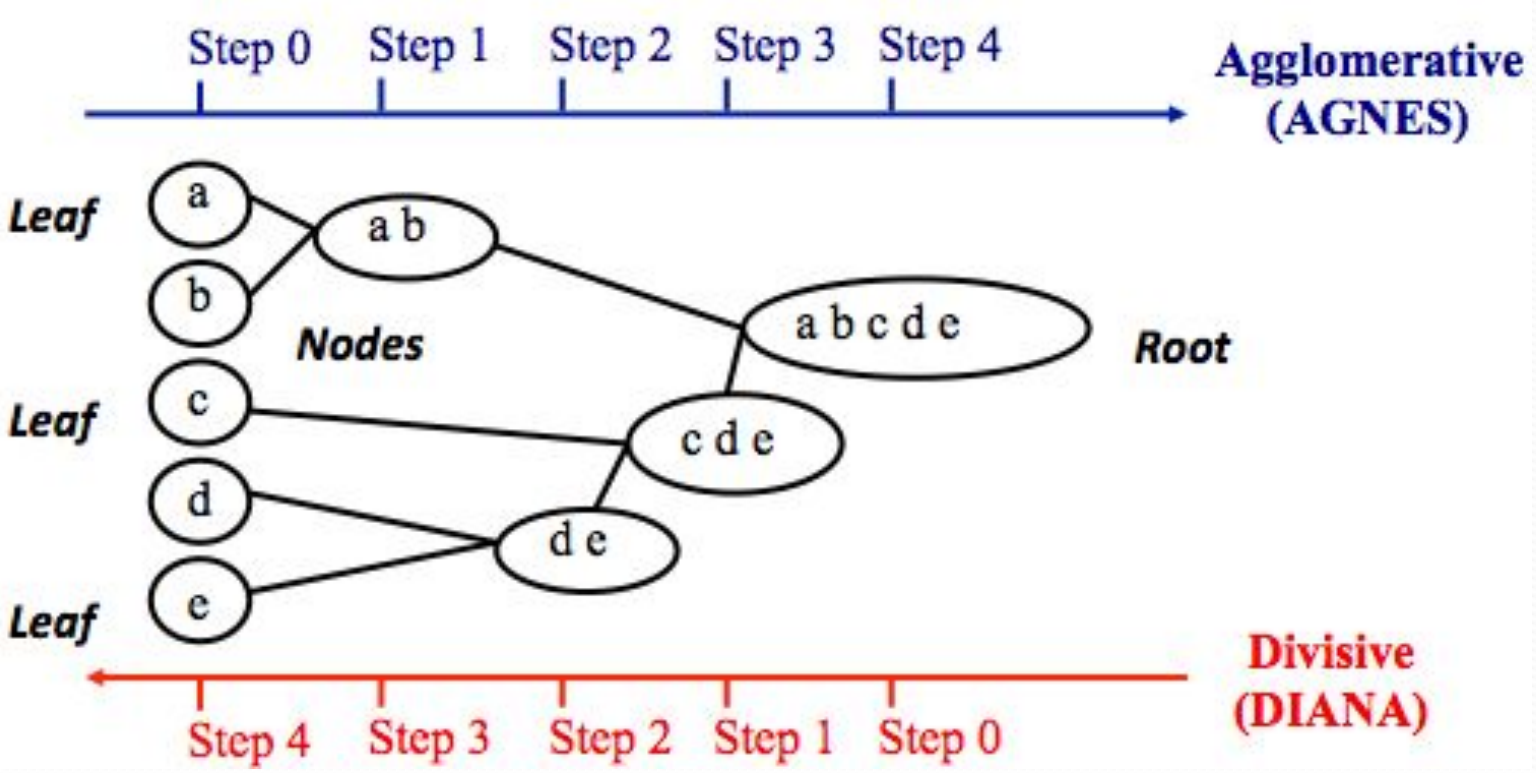
"Bottom up" approach: Begin with each element as a separate cluster and merge them into successively larger clusters

"Top down" approach begin with the whole set and proceed to divide it into successively smaller clusters.

Hierarchical clustering types

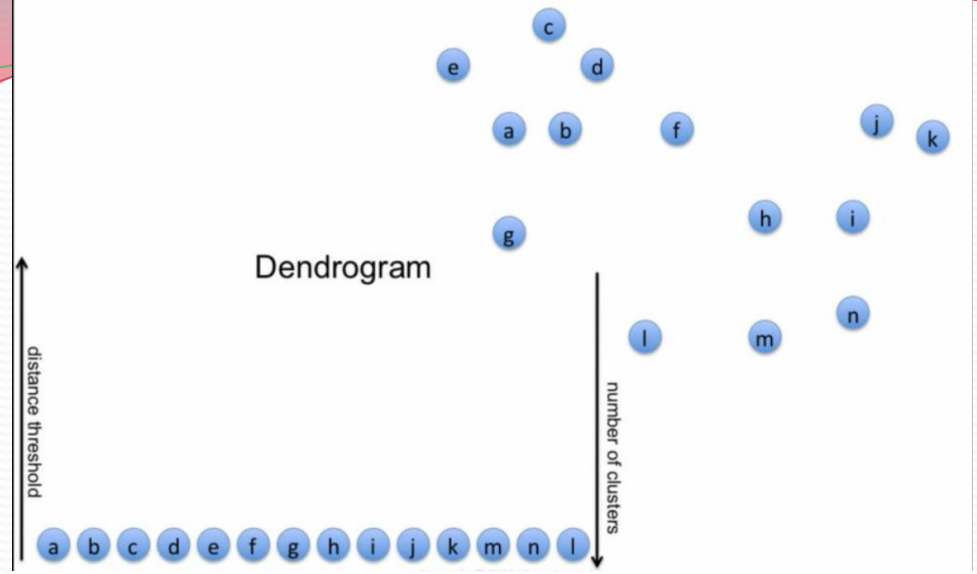
- Hierarchical clustering algorithm is of two types:
- i) Agglomerative Hierarchical clustering algorithm or AGNES (agglomerative nesting) and
- ii) Divisive Hierarchical clustering algorithm or DIANA

Hierarchical clustering

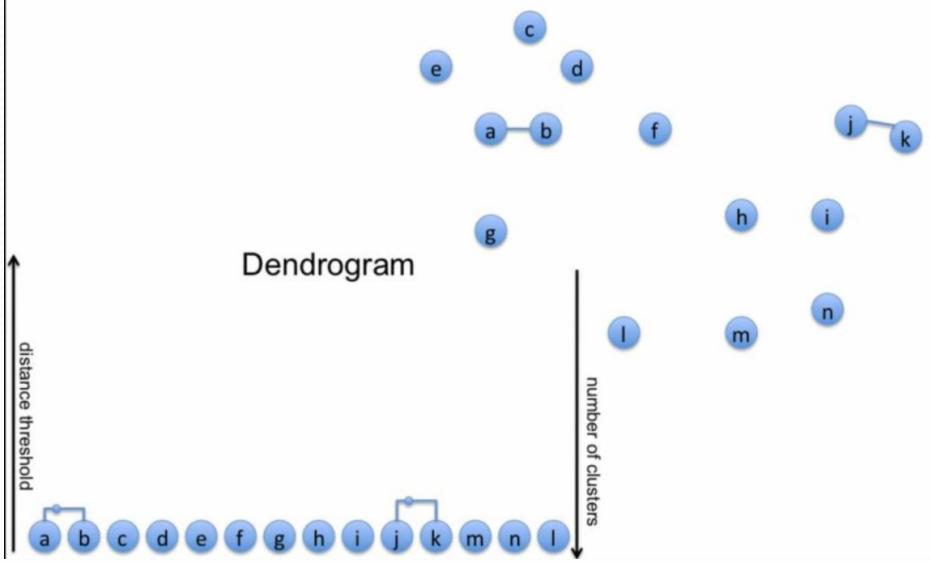


Note that, agglomerative clustering is good at identifying small clusters. Divisive clustering is good at identifying large clusters

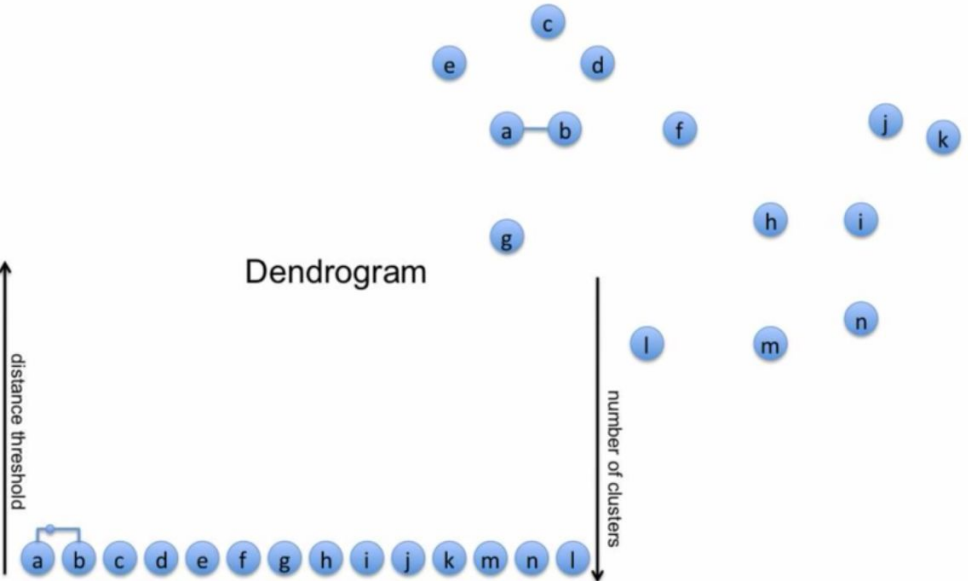
Agglomerative clustering: example



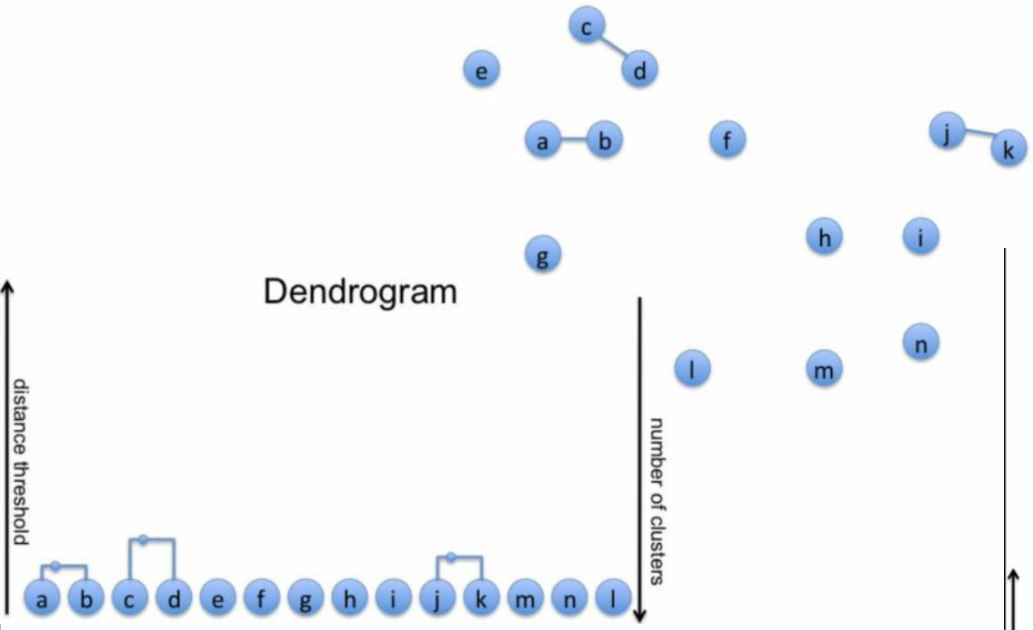
Agglomerative clustering: example



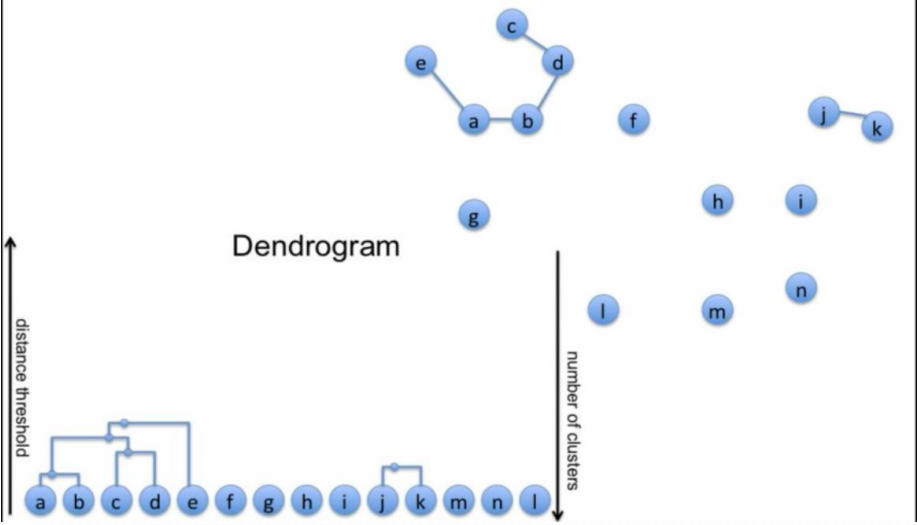
Agglomerative clustering: example



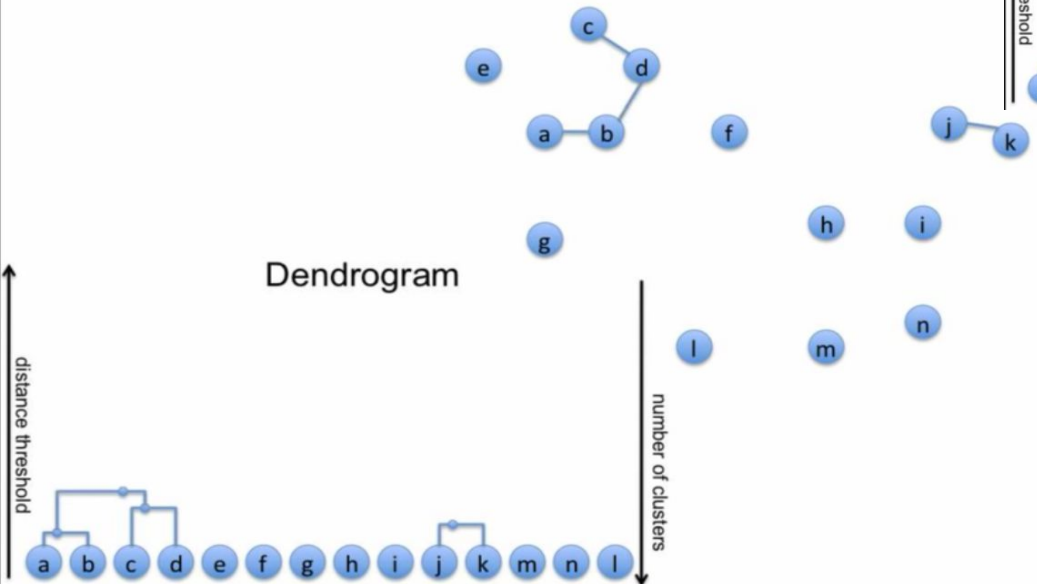
Agglomerative clustering: example



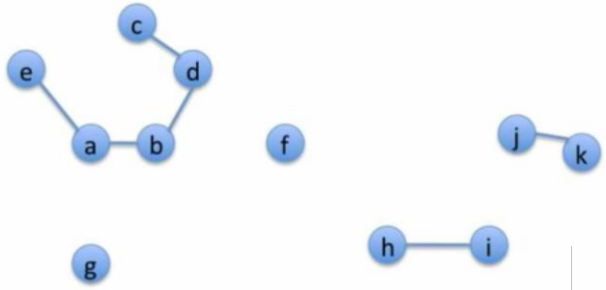
Agglomerative clustering: example



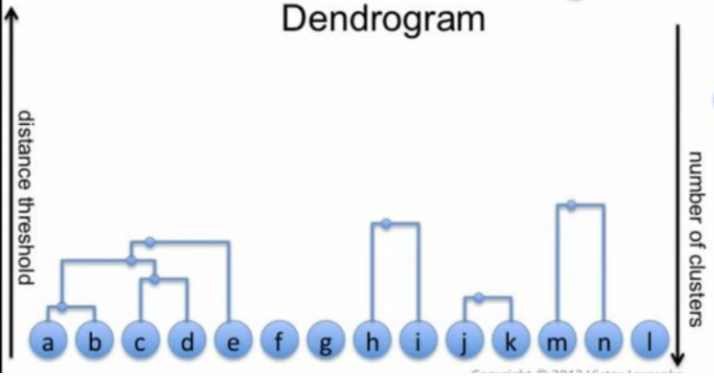
Agglomerative clustering: example



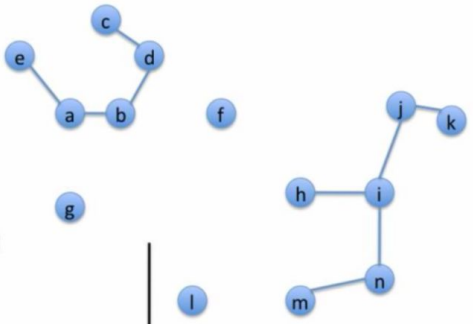
Agglomerative clustering: example



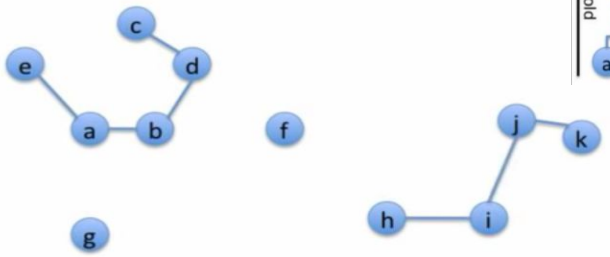
Dendrogram



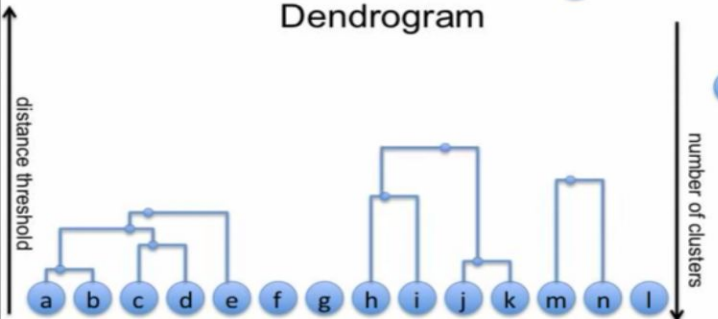
Agglomerative clustering: example



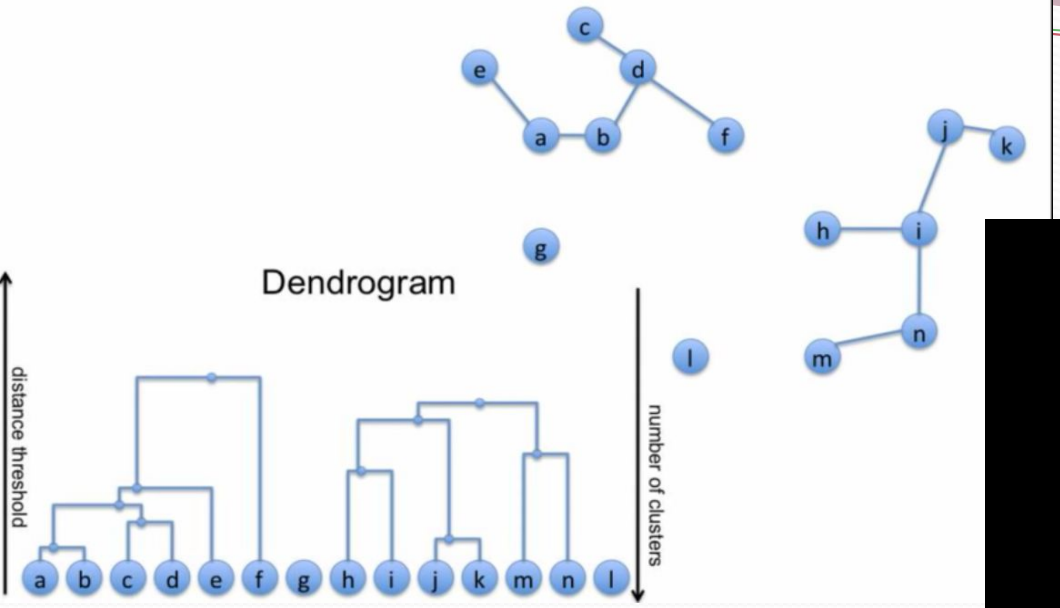
Agglomerative clustering: example



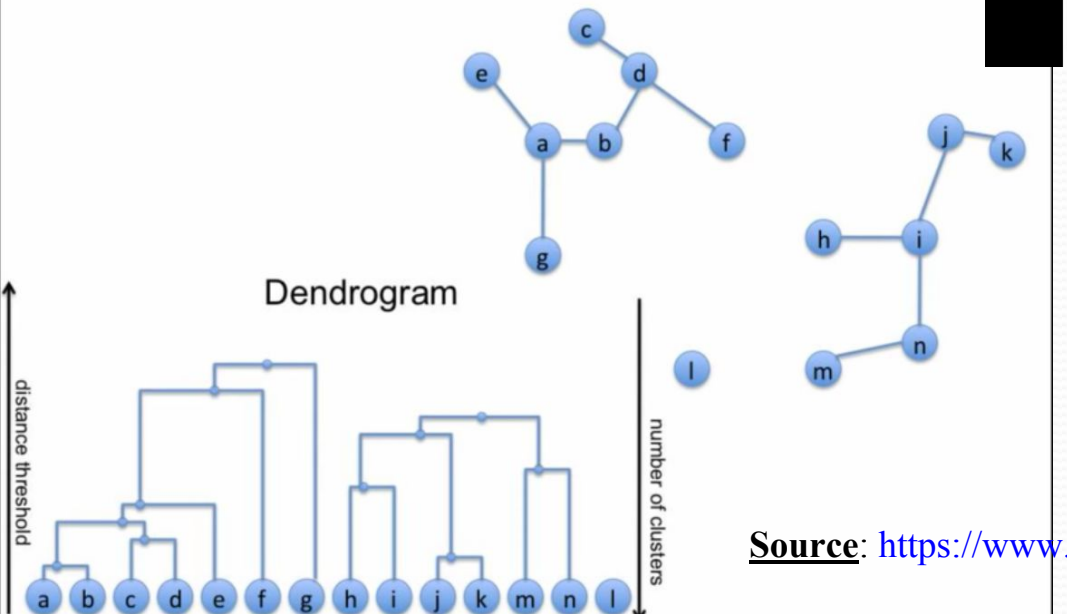
Dendrogram



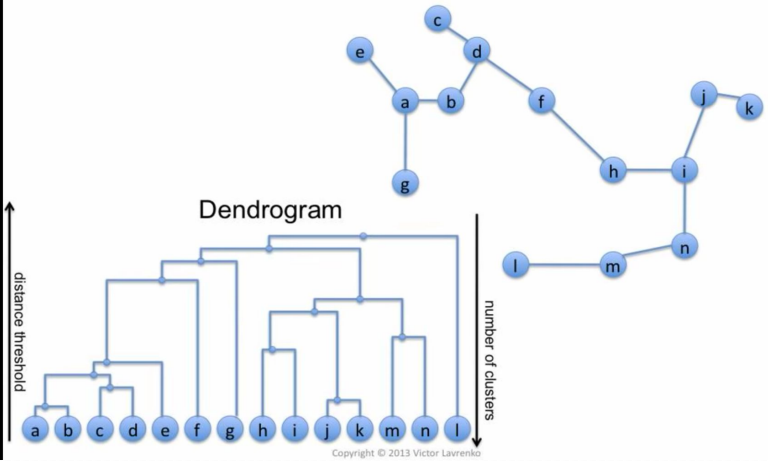
Agglomerative clustering: example



Agglomerative clustering: example

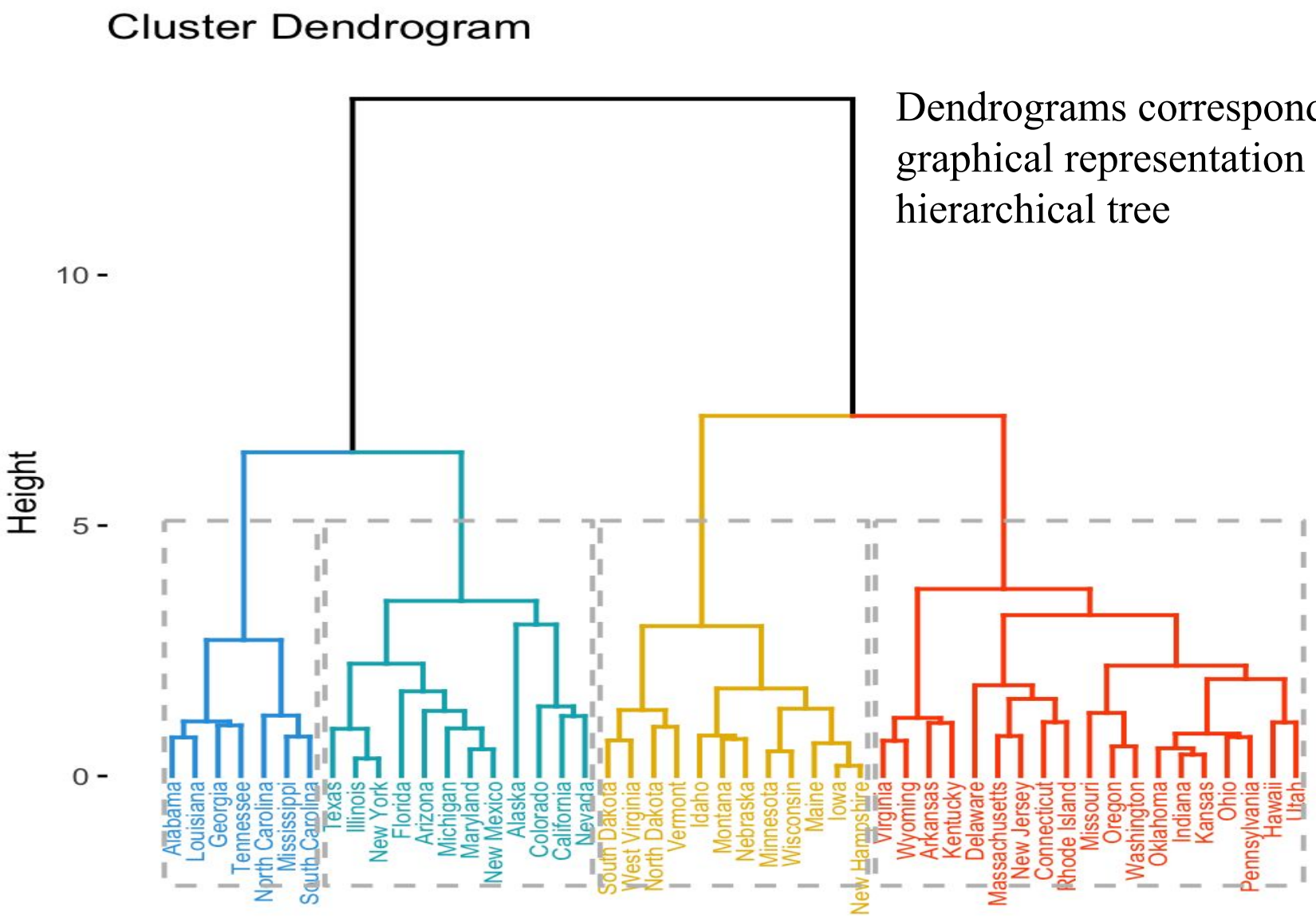


Agglomerative clustering: example

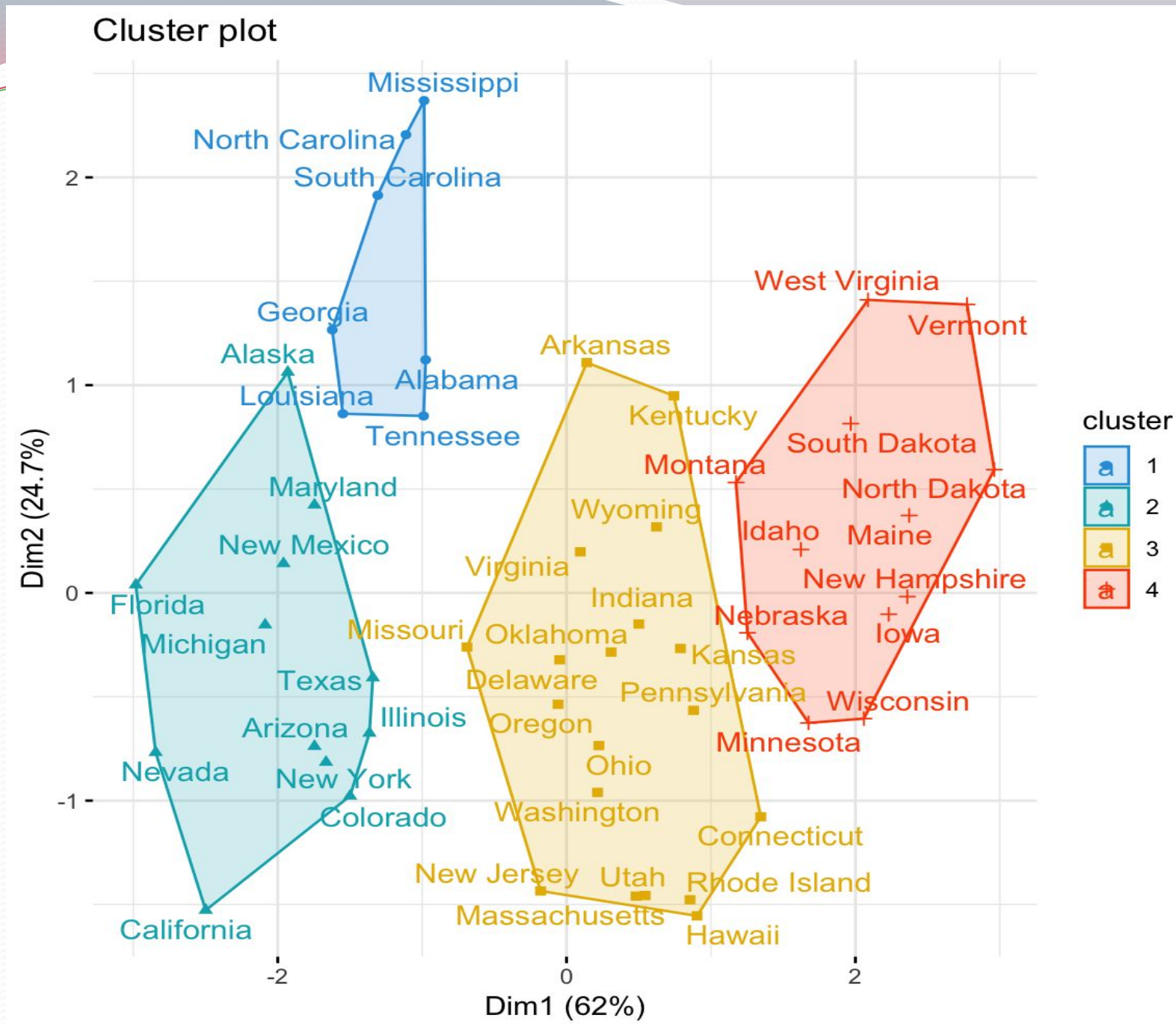


Source: https://www.youtube.com/watch?v=XJ3194AmH40

Dendogram

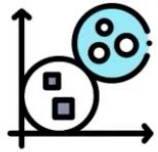


Dendrograms correspond to the graphical representation of the hierarchical tree

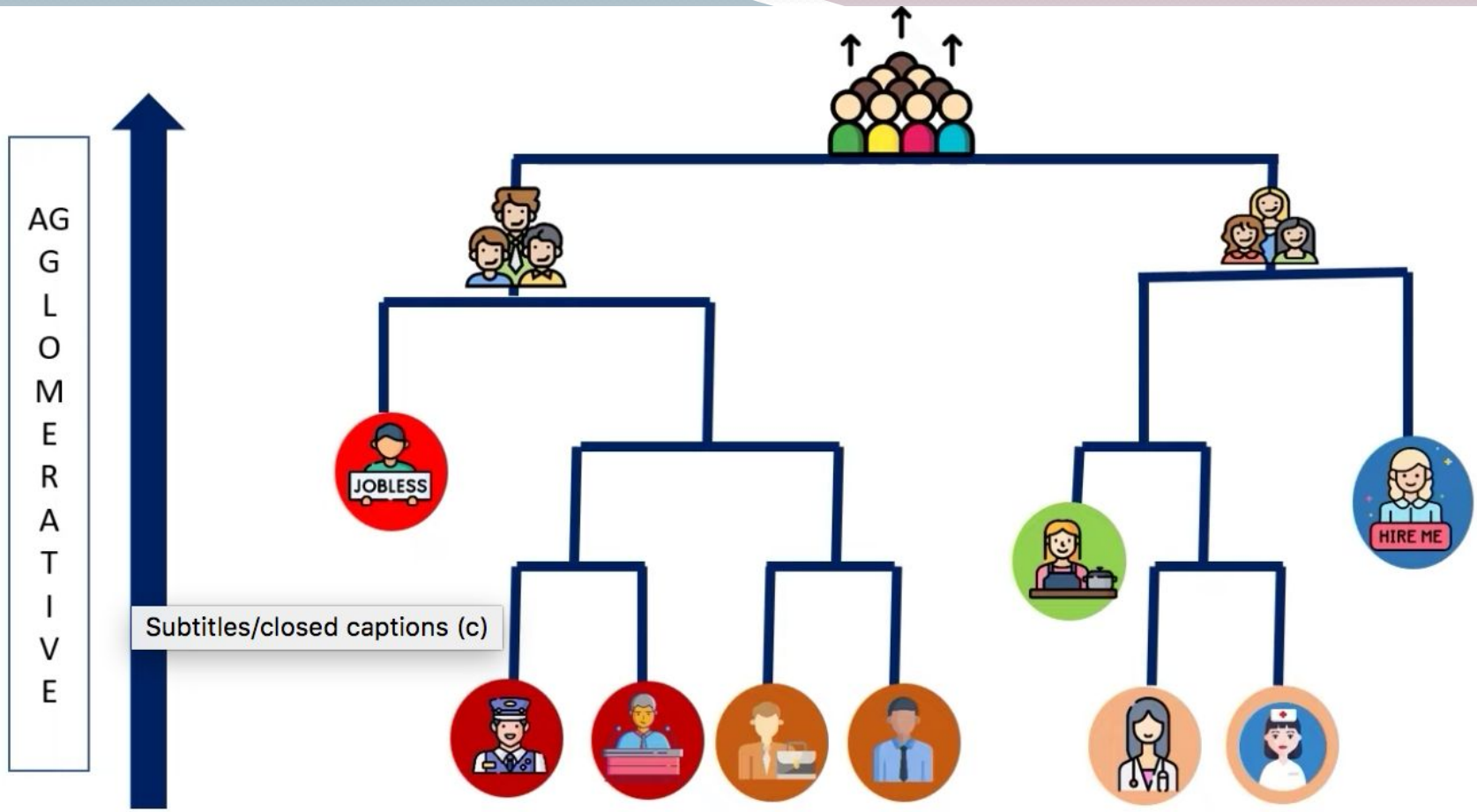

Cluster plot
Mississippi
North Carolina
South Carolina
Georgia
Alaska
Alabama
Louisiana
Tennessee
Maryland
New Mexico
Florida
Michigan
Texas
Illinois
Arizona
Nevada
New York
Colorado
California
Arkansas
Kentucky
Wyoming
Virginia
Indiana
Missouri
Oklahoma
Kansas
Delaware
Pennsylvania
Oregon
Ohio
Washington
Connecticut
New Jersey
Utah
Rhode Island
Massachusetts
Hawaii
West Virginia
Vermont
South Dakota
Montana
North Dakota
Idaho
Maine
New Hampshire
Nebraska
Iowa
Wisconsin
Minnesota
Dim2 (24.7%)
Dim1 (62%)
2
1
0
-1
-2
0
2
cluster
1
2
3
4

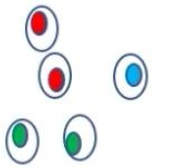
What is Clustering?

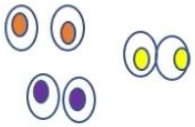
Clustering is an unsupervised Machine Learning technique that forms clusters based on similarity between the data points.

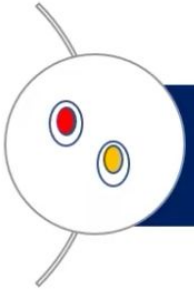


Clustering Algorithms that cluster the data like k-means, DBSCAN, Mean-Shift Clustering, Hierarchical Clustering

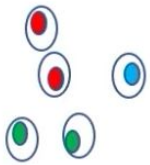

AGGLOMERATIVE
JOBLESS
HIRE ME
Subtitles/closed captions (c)

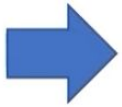


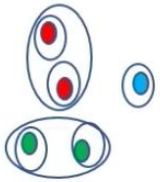


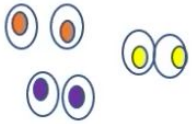


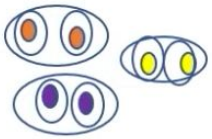
Each observation is treated as a separate cluster.

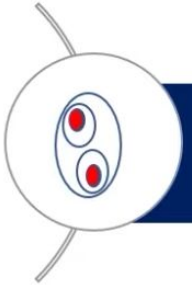




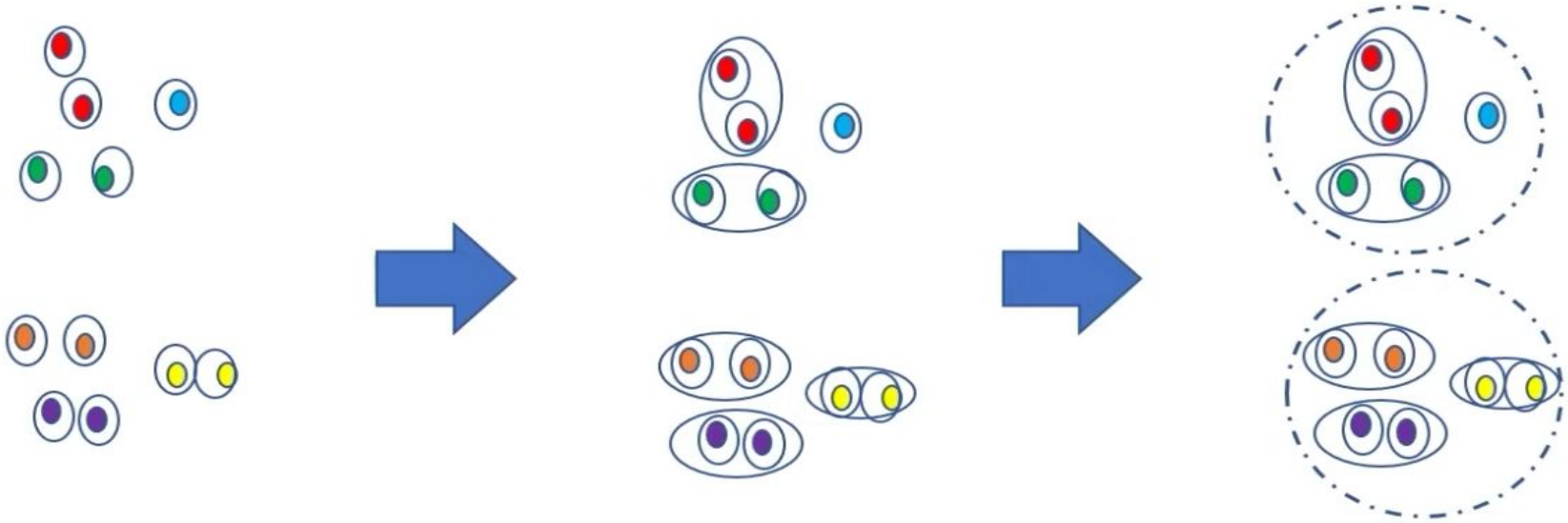


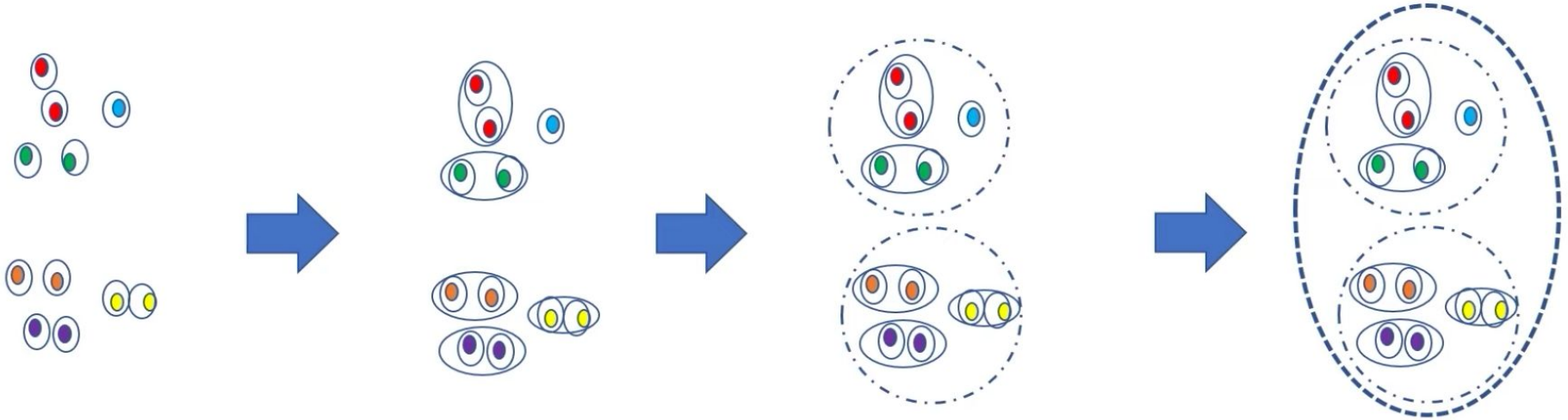




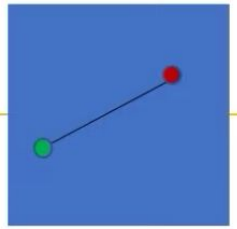


Two Clusters that are closest are identified, and merged.

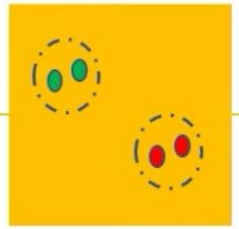




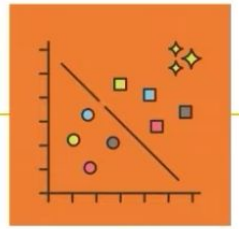
How To Decide The Group Of Clusters?



Based on the similarity, single linkage, complete linkage.



Similarity is obtained by calculating distance between clusters.

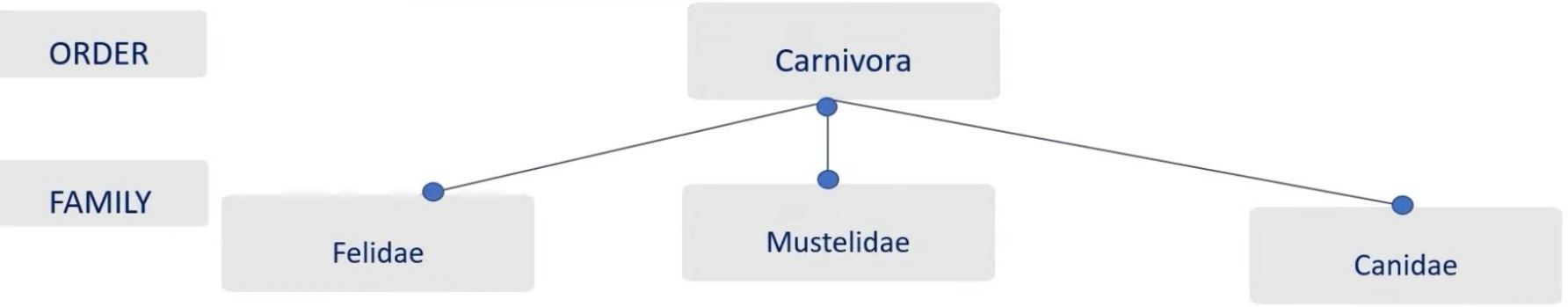


Distance metrics like Euclidian , Manhattan, Correlation distance

Applications of Hierarchical Clustering

ORDER	Carnivora

Applications of Hierarchical Clustering

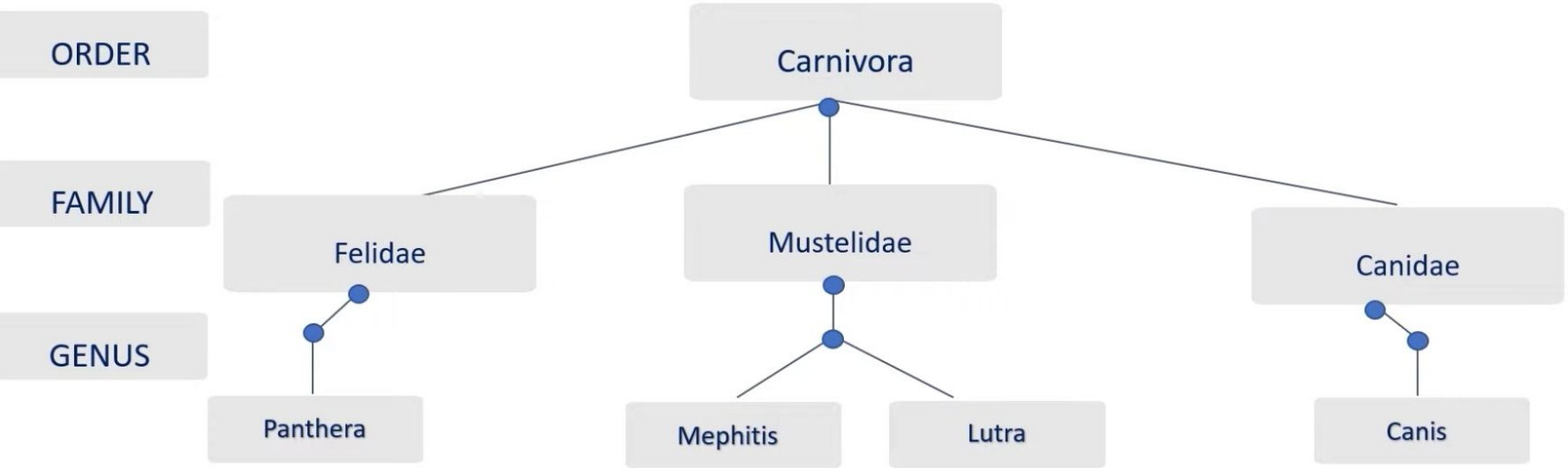


P4	0.37	0.15	0

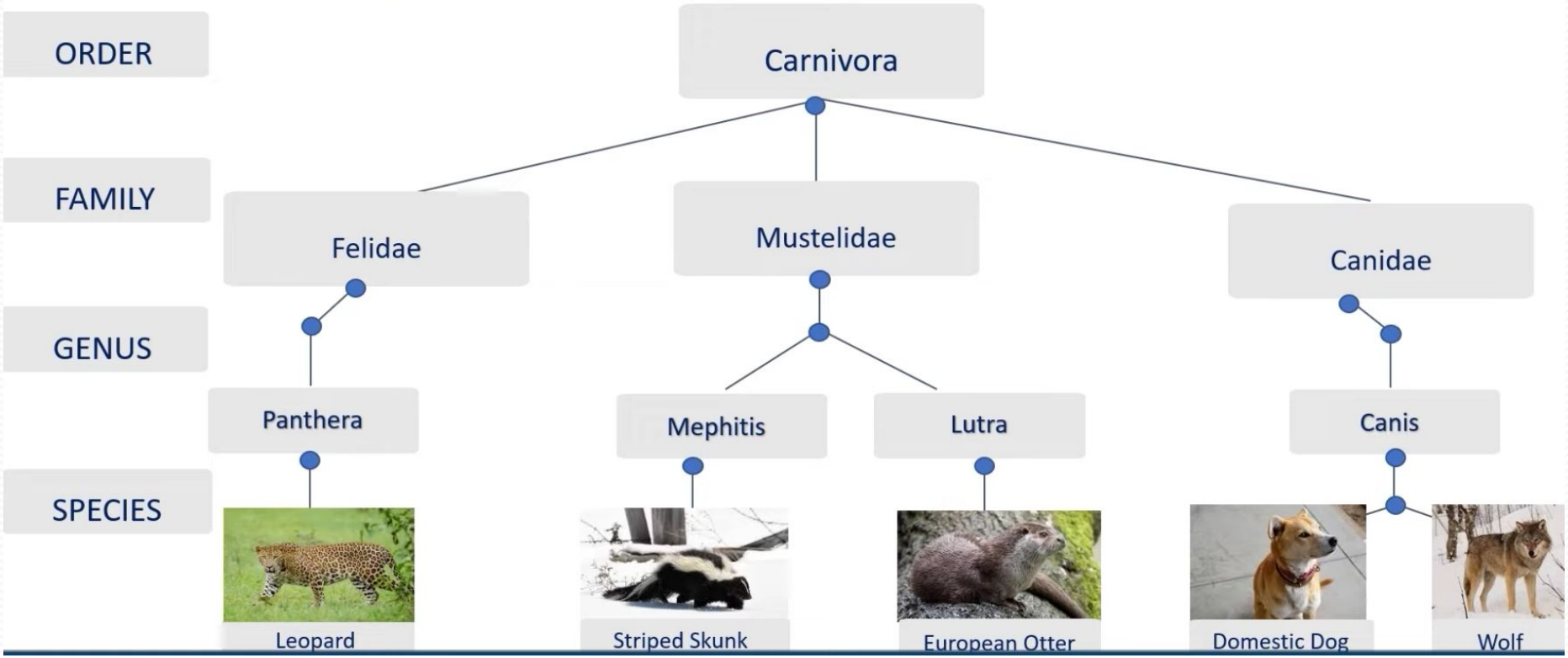
edureka!

www.edureka.co/data-science-python-certification-course

Applications of Hierarchical Clustering



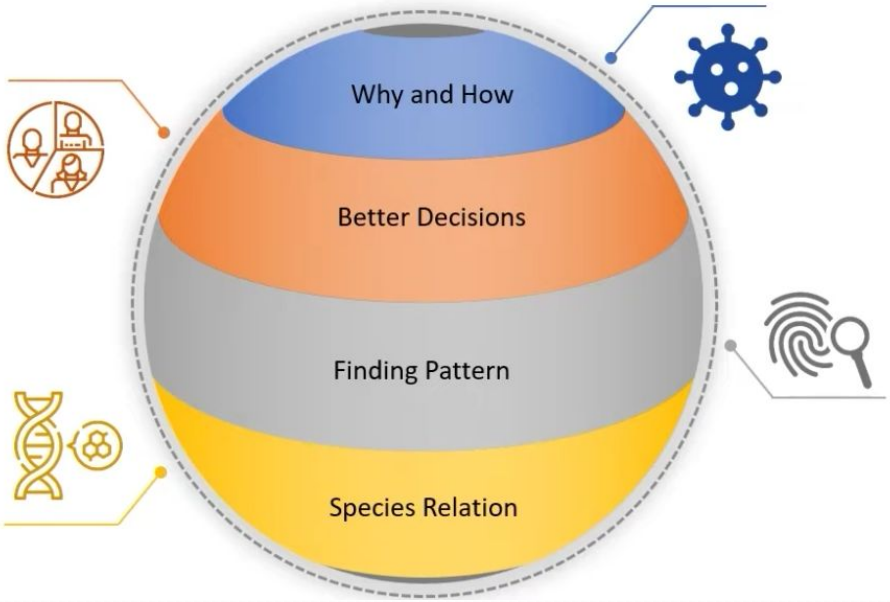
Applications of Hierarchical Clustering



Applications of Hierarchical Clustering

Segmenting the customers

Tracking viral Outbreaks



Clustering Crimes sites in the city

Evolution through Phylogenetic trees

Advantages of Hierarchical Clustering

Easy to implement and understand

No prior information is required about the number of clusters.

Deterministic, more predictable.

Outliers can be detected With the help of dendrogram

Disadvantages of Hierarchical Clustering

It not suitable for large datasets.

Difficulty in handling different sized clusters

It is sensitive to outliers and noise in the dataset

Source: https://www.youtube.com/watch?v=3M1wUK2zCKY