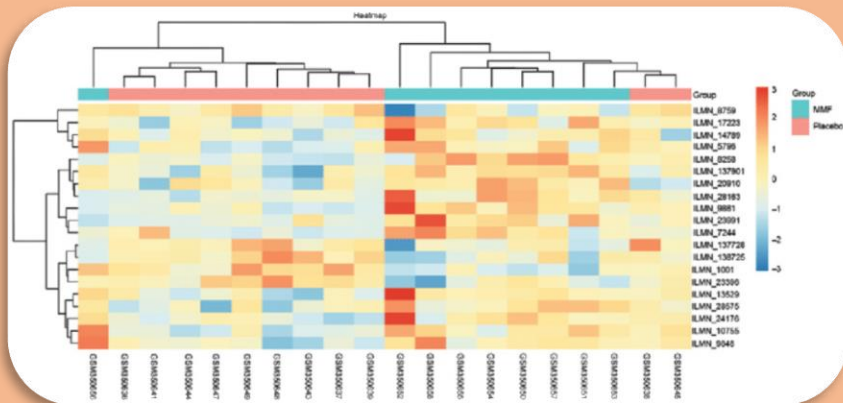


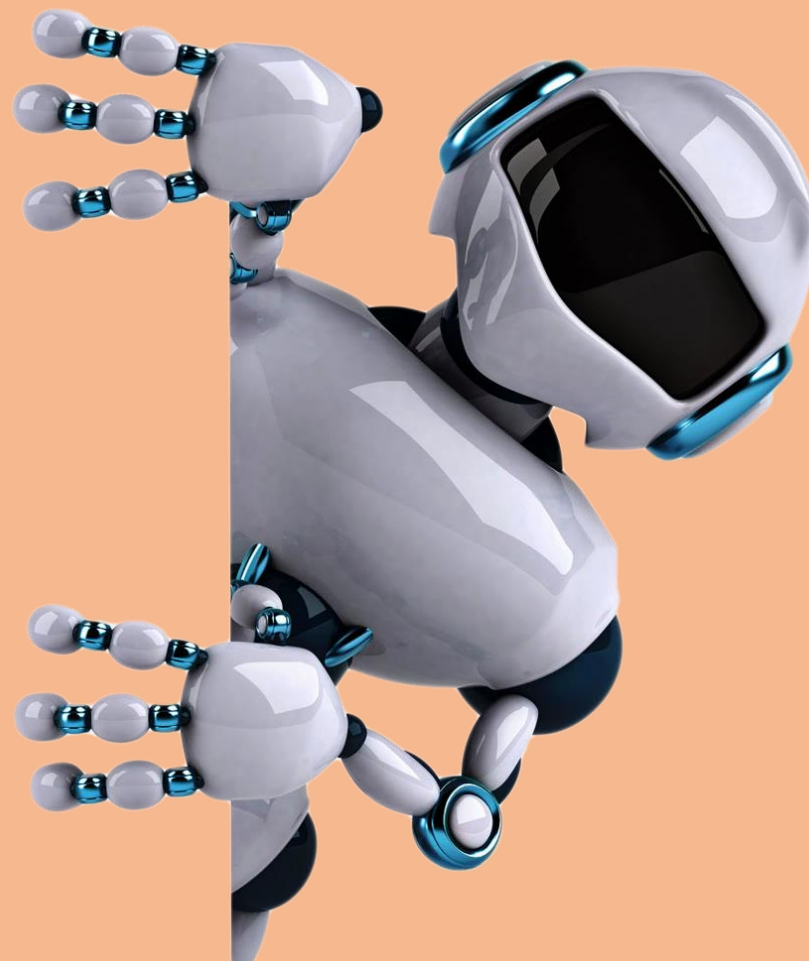
دوره‌ی آموزشی «علم داده»

Data Science Course

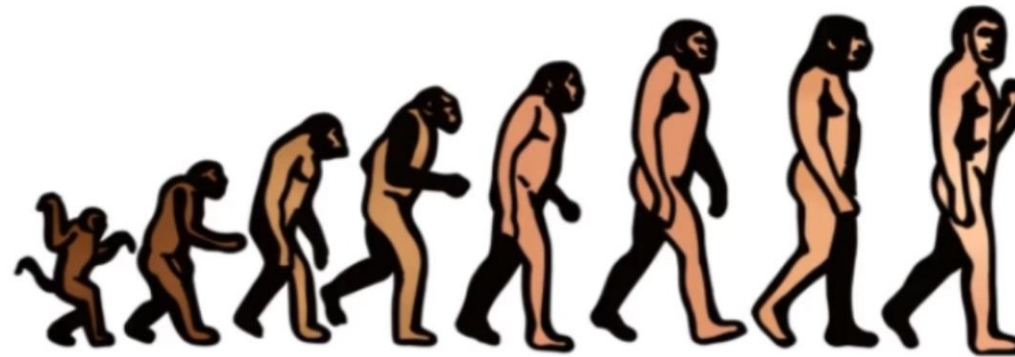
جلسه بیست و چهارم - (بخش اول) خوشه‌بندی سلسله مراتبی و نمودار درختی



مدرس: محمد فزونی
عضو هیئت علمی دانشگاه گنبدکاوس



Anthropologists





- psychology
- intelligence
- others



TYPES OF CLUSTERING

```
graph TD; A[TYPES OF CLUSTERING] --> B[FLAT]; A --> C[HIERARCHICAL]; B --> D[K - MEANS]
```

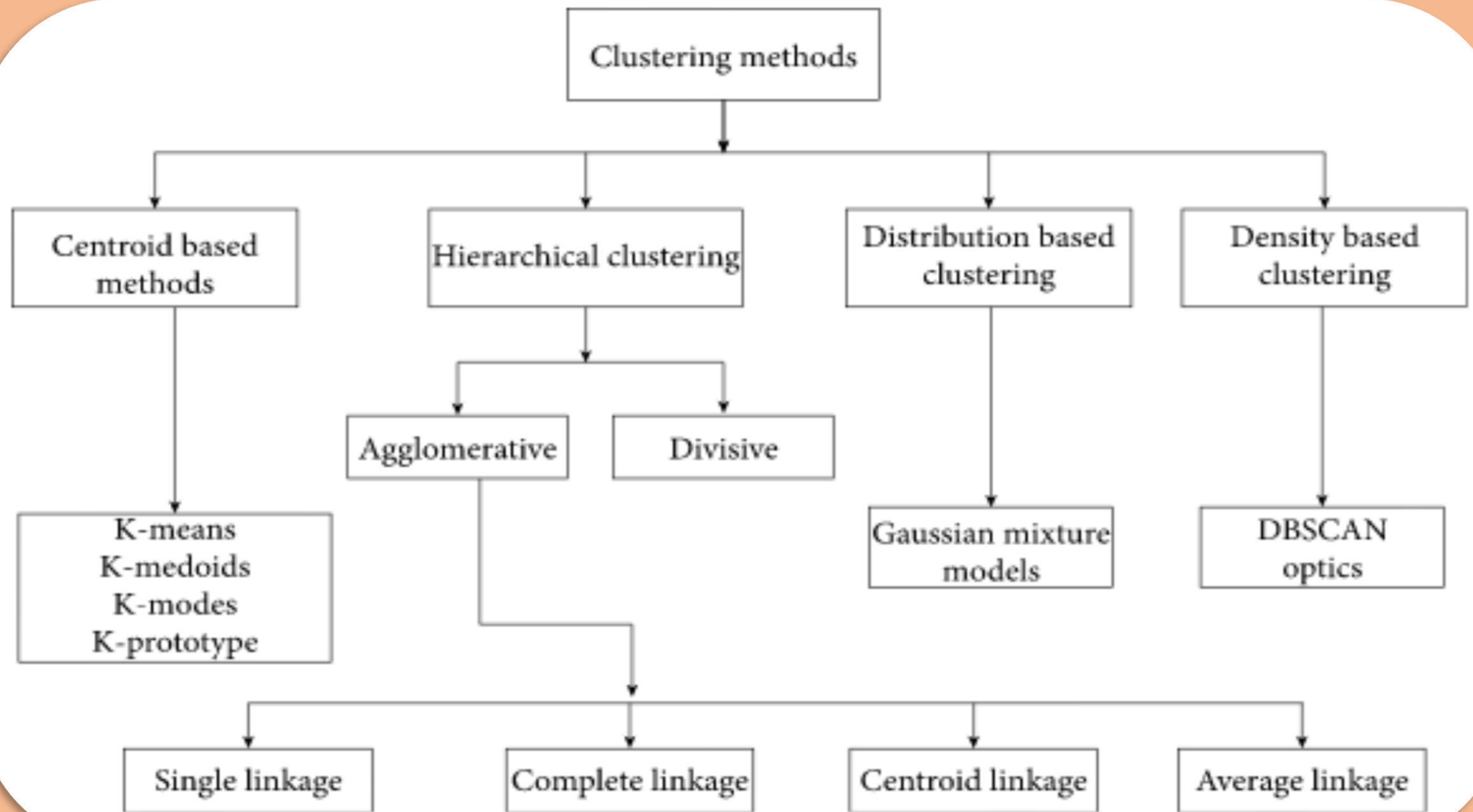
FLAT

HIERARCHICAL

K - MEANS

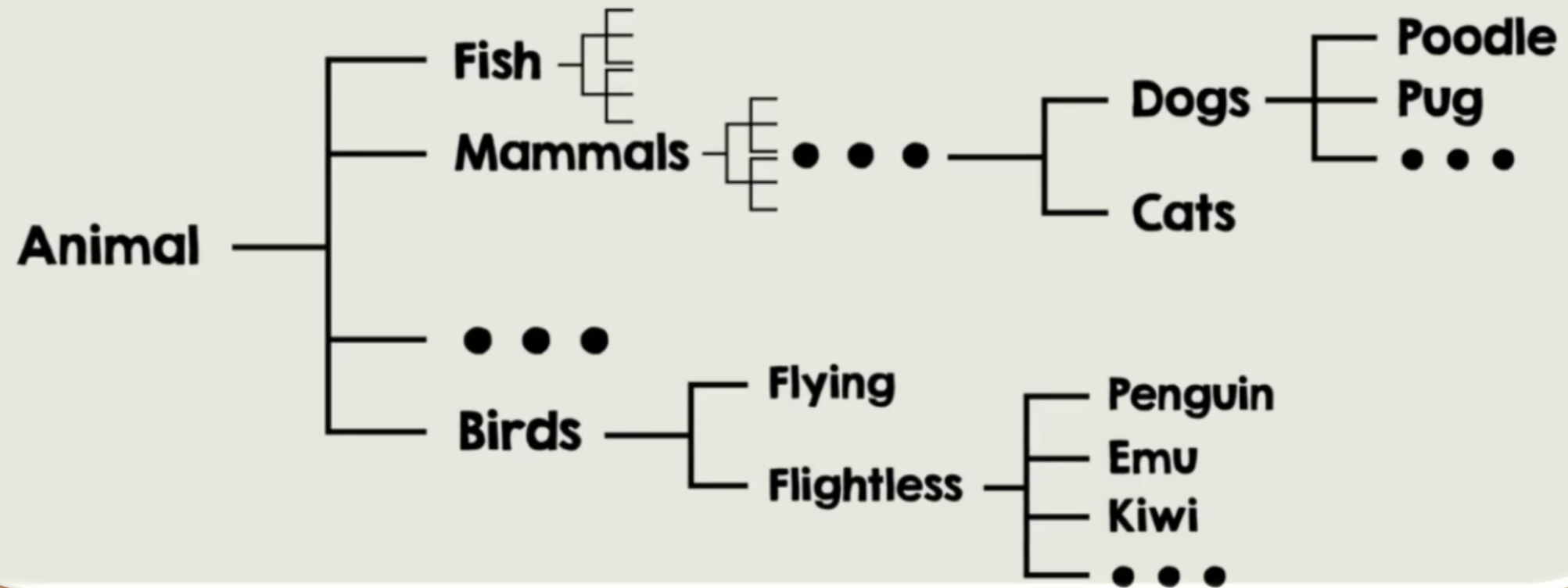
Flat clustering is where the scientist tells the machine how many categories to cluster the data into.

Hierarchical clustering is where the machine is allowed to decide how many clusters to create based on its own algorithms.



HIERARCHICAL

TAXONOMY OF THE ANIMAL KINGDOM



TYPES OF HIERARCHICAL CLUSTERING



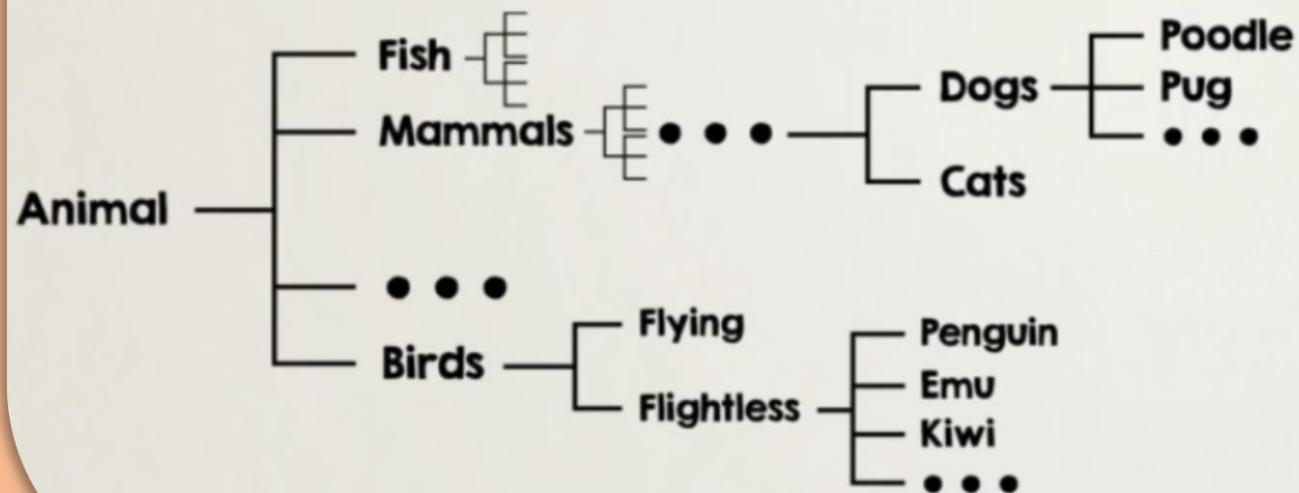
AGGLOMERATIVE
(BOTTOM-UP)



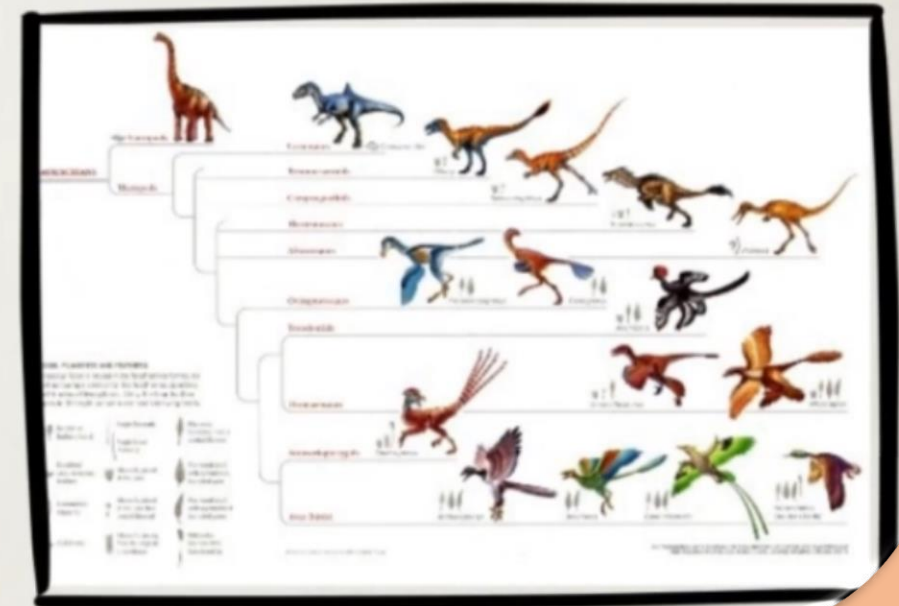
DIVISIVE
(TOP-DOWN)

TYPES OF HIERARCHICAL CLUSTERING

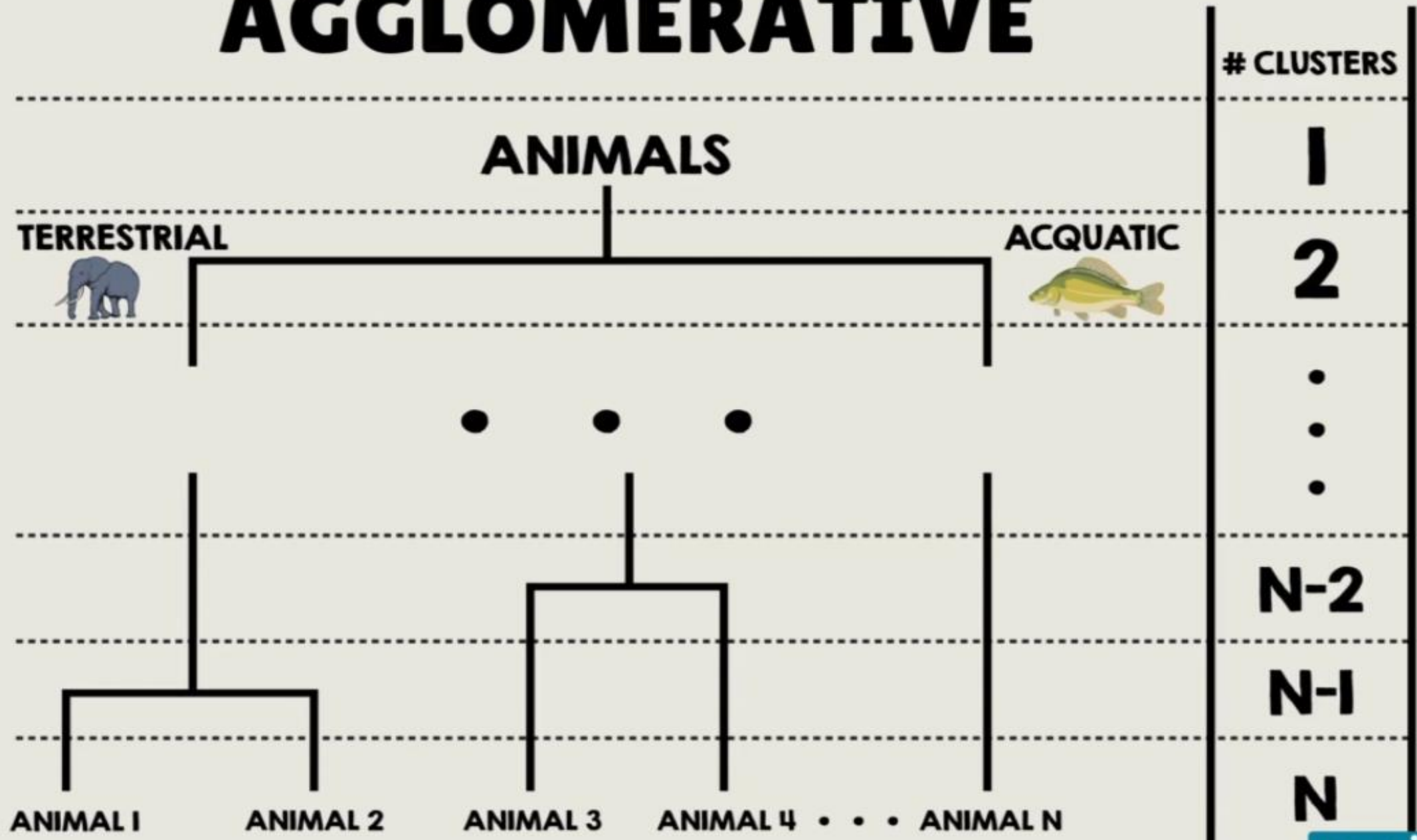
AGGLOMERATIVE (BOTTOM-UP)



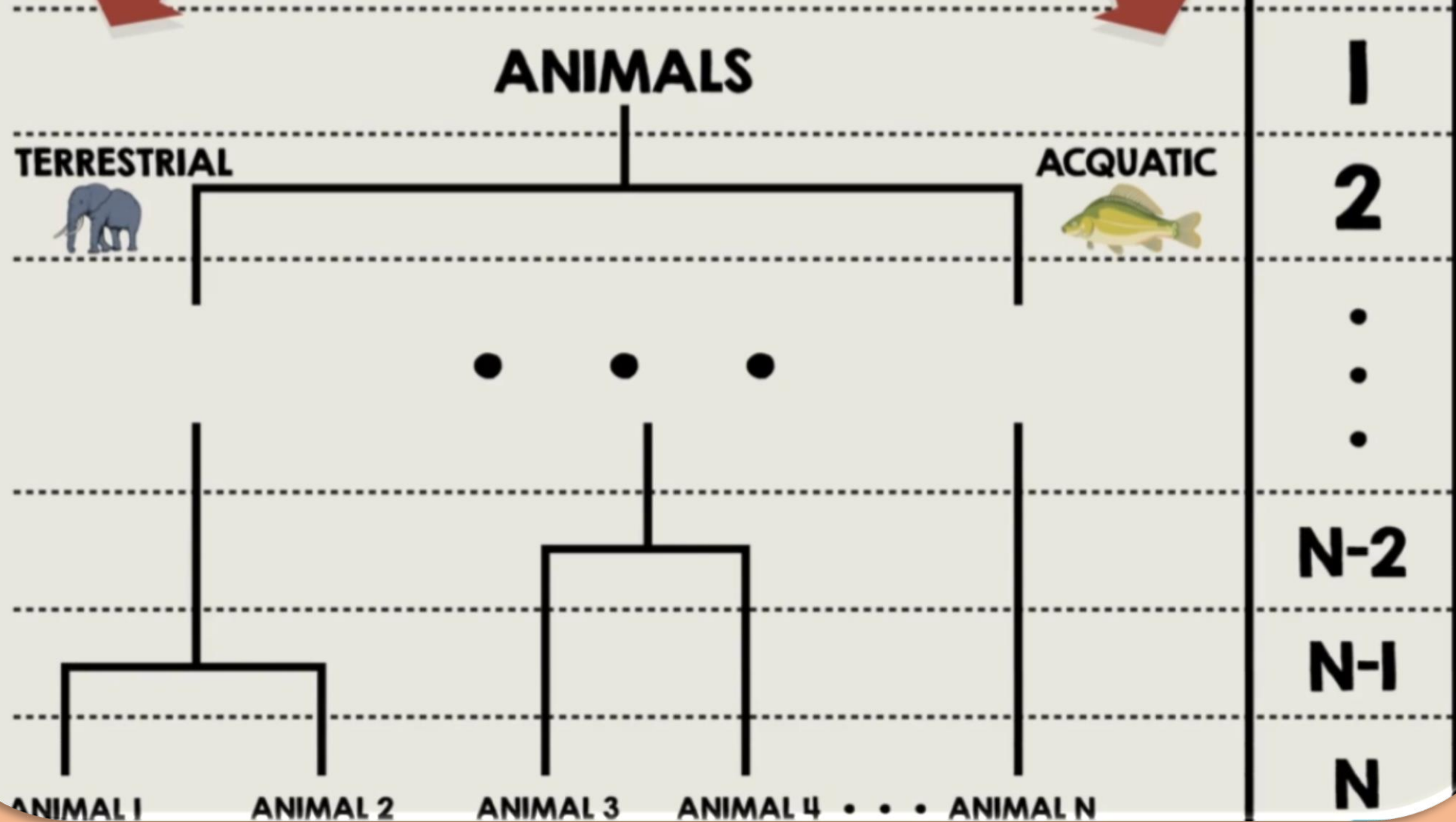
DIVISIVE (TOP-DOWN)



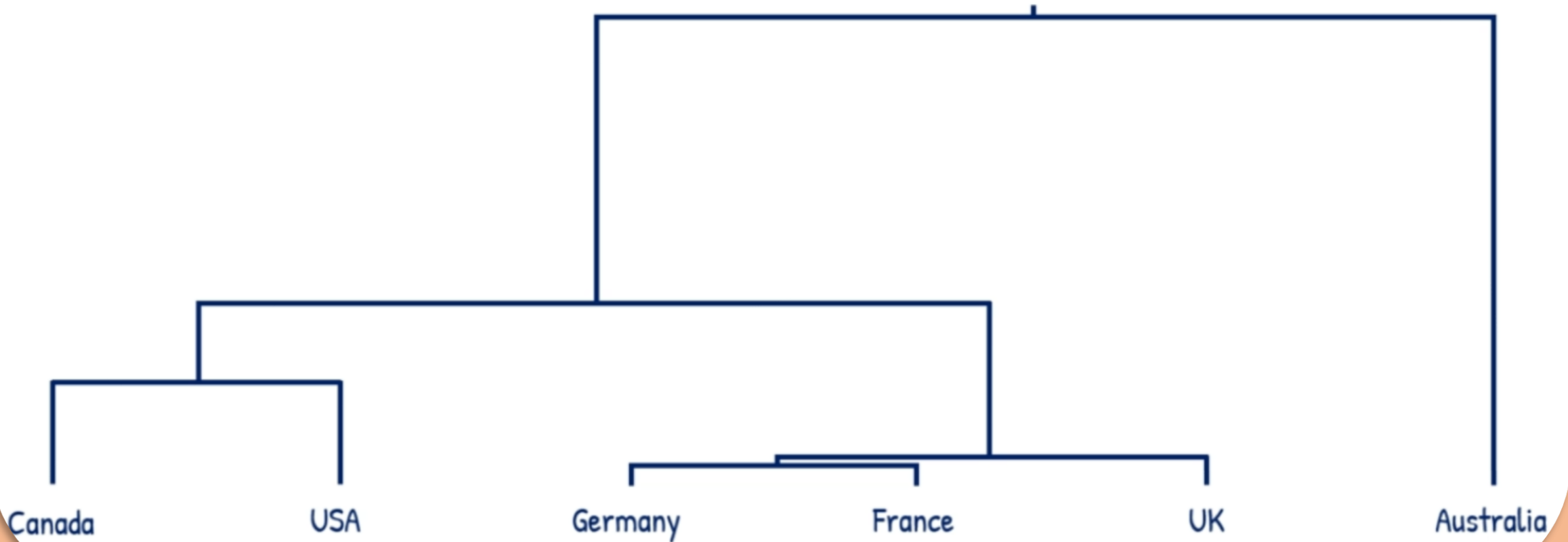
AGGLOMERATIVE



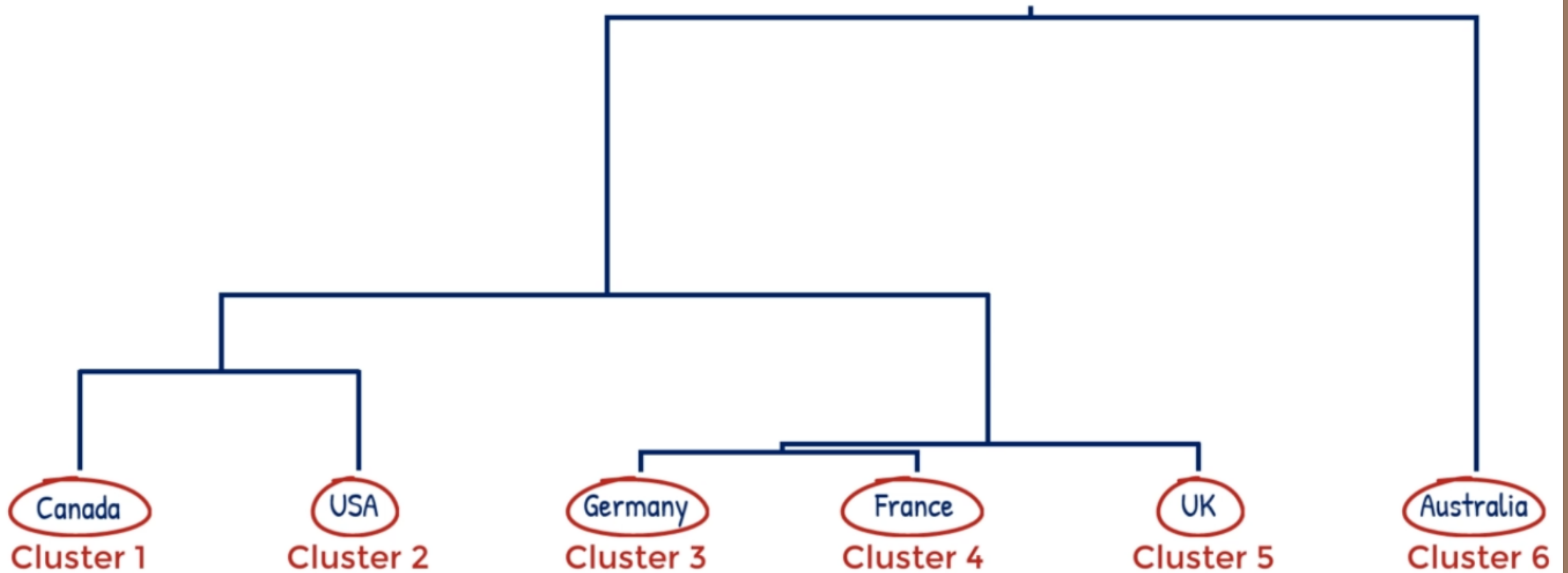
DENDROGRAM



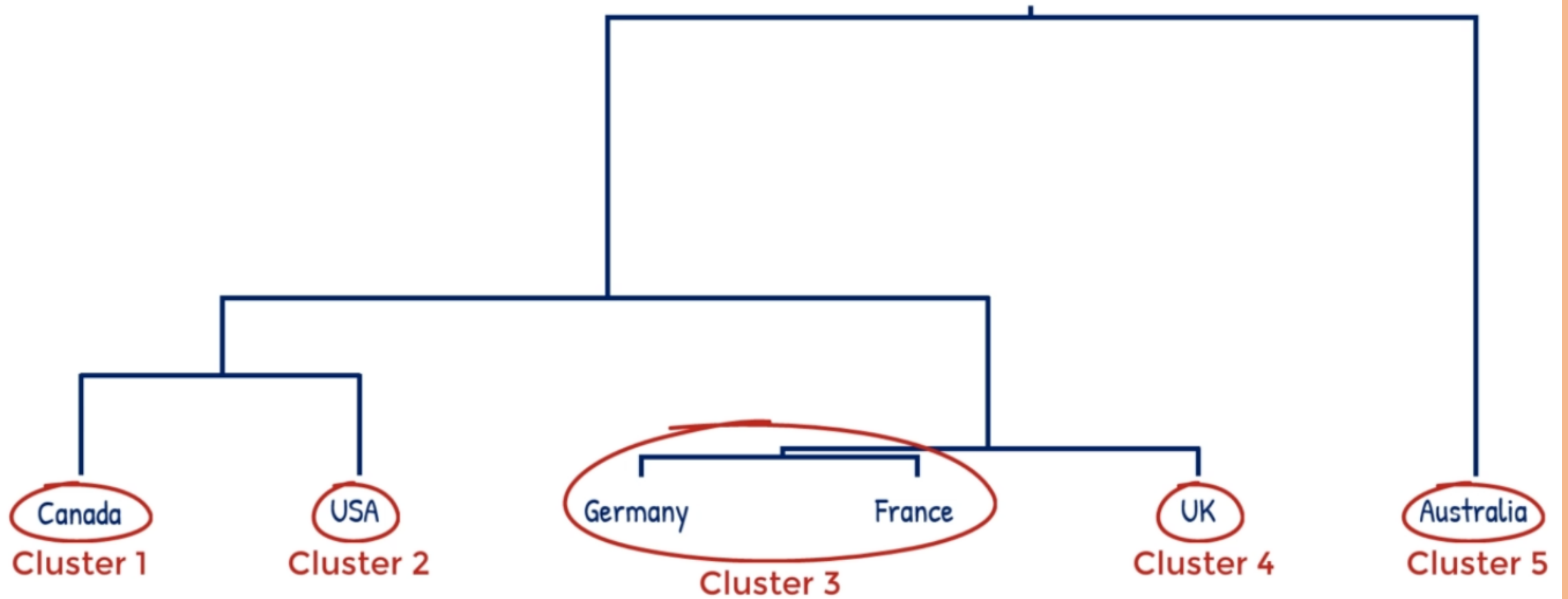
DENDROGRAM



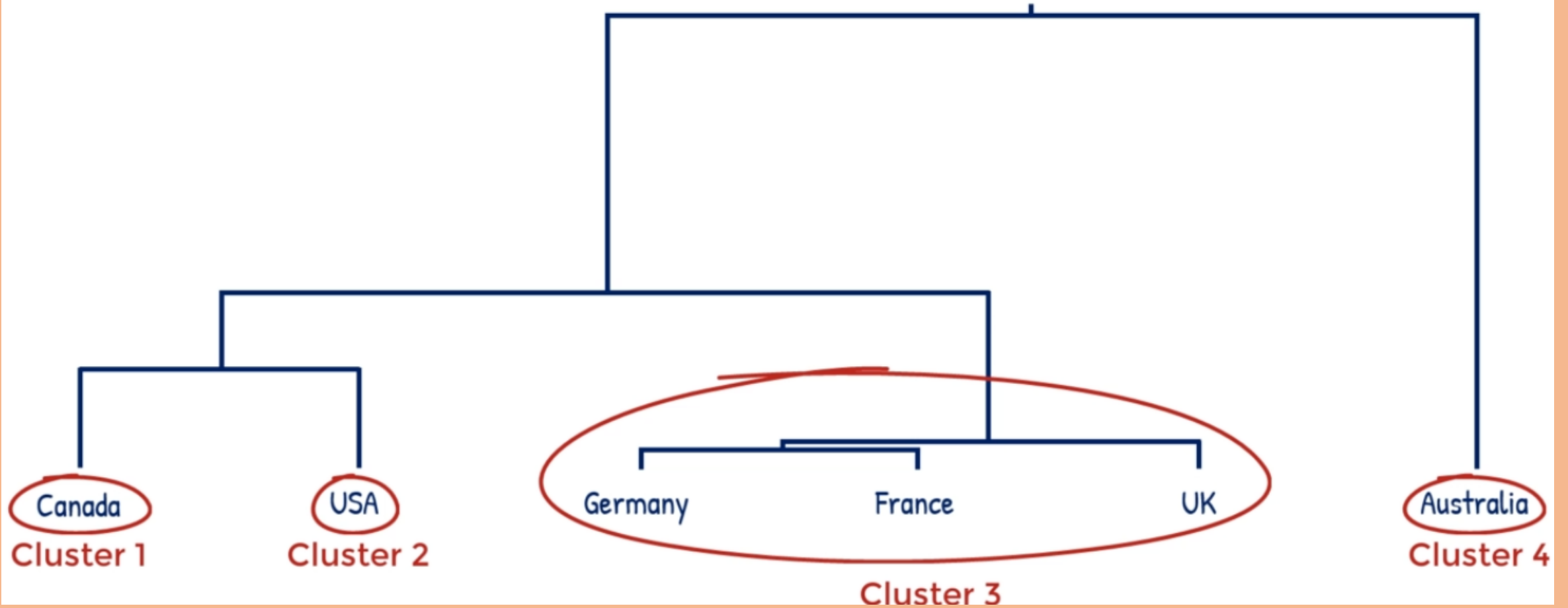
DENDROGRAM



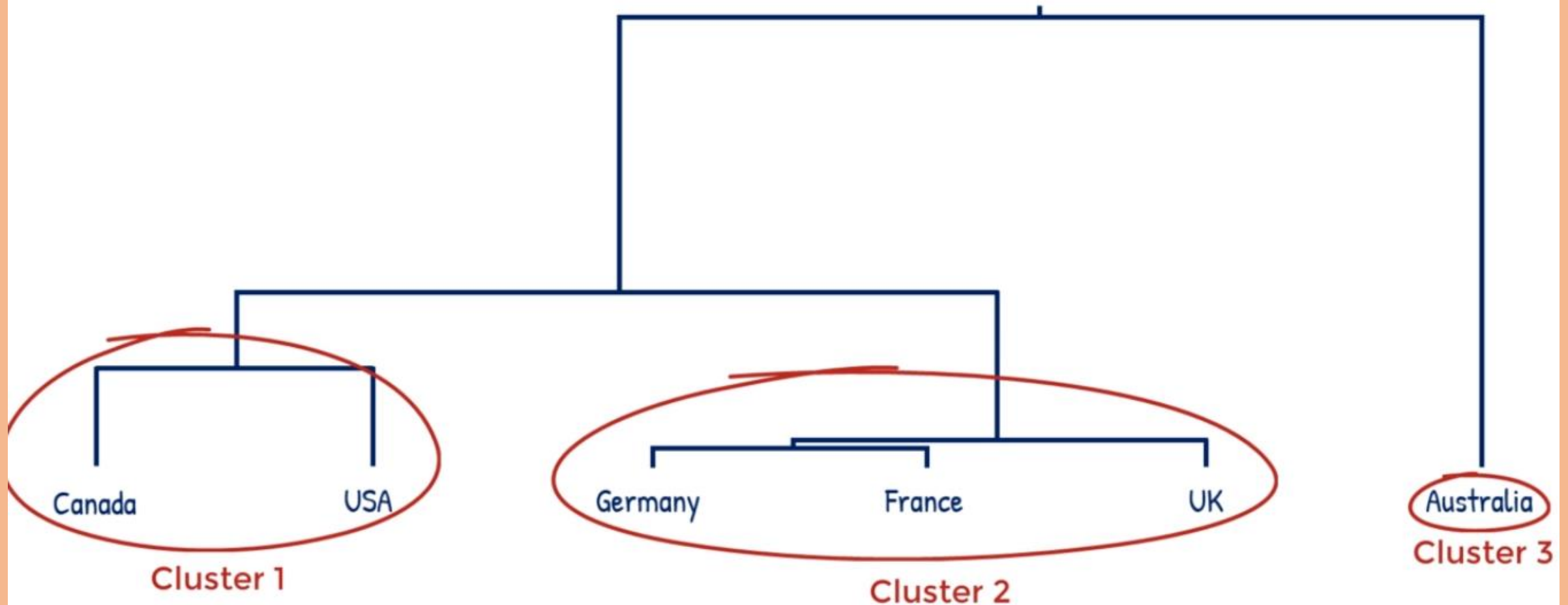
DENDROGRAM



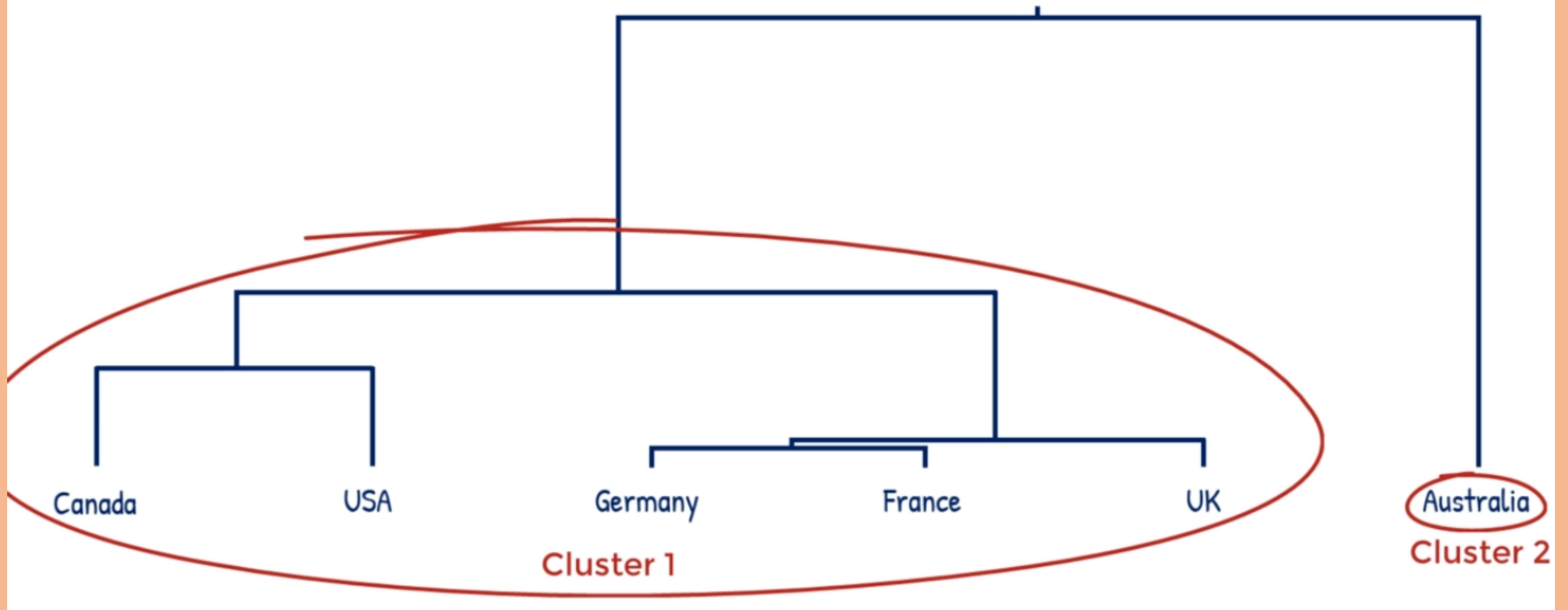
DENDROGRAM



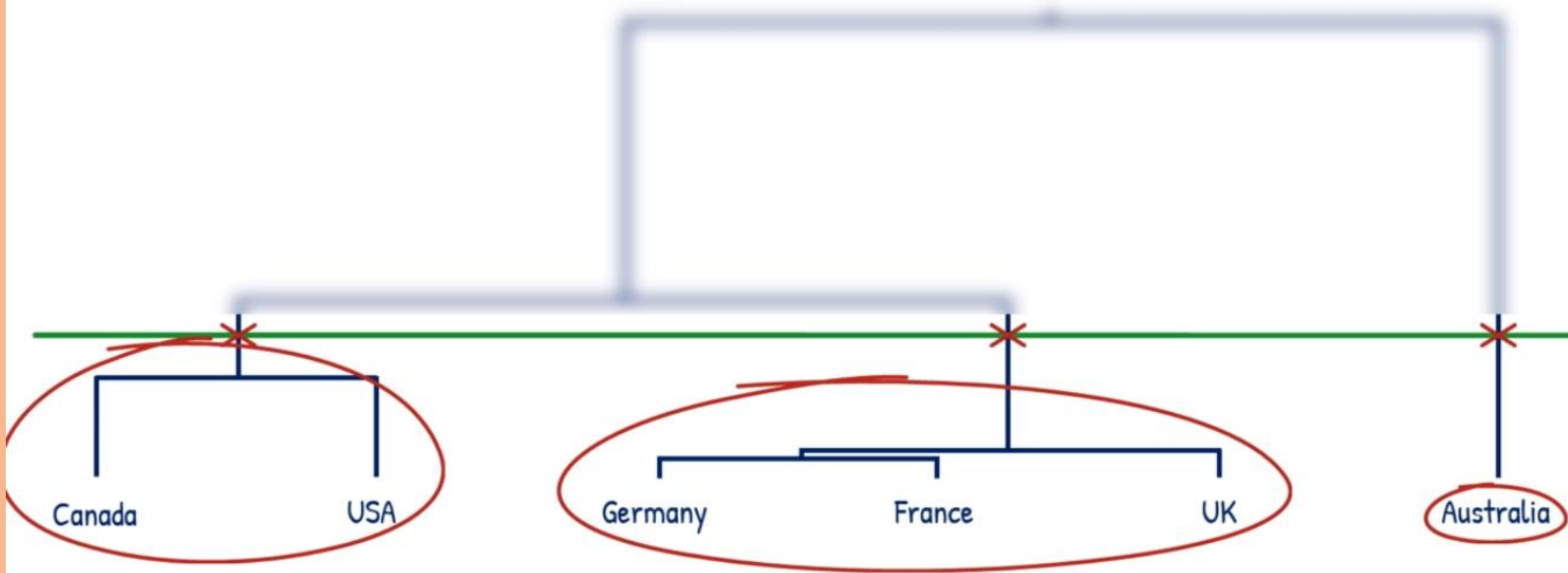
DENDROGRAM



DENDROGRAM



DENDROGRAM



Pros

1. Hierarchical clustering shows all the possible linkages between clusters
2. We understand the data much, much better
3. No need to preset the number of clusters (like with k-means)
4. Many methods to perform hierarchical clustering (Ward!)

In statistics, Ward's method is a criterion applied in hierarchical cluster analysis.

Ward suggested a general agglomerative hierarchical clustering procedure, **where the criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function.**

Time consuming and costly to go through it. So, it's not always a data scientist's cup of tea

Con?

