**Hierarchical Clustering**

**↙↘**

**Agglomerative Divisive**

**Agglomerative Clustering :**

**Step 1 :** Make each data point a single-point cluster ⇒ That forms **N clusters**.

**Step 2 :** Take two closest data points and combine to make them one cluster

⇒ That forms **N-1** clusters

**Step 3 :** Take the two **closest clusters** and combine them to form one cluster

⇒ That forms **N-2** clusters.

**Step 4 :** Repeat Step 3 until there is only one cluster

**Step 5 : Finish.**

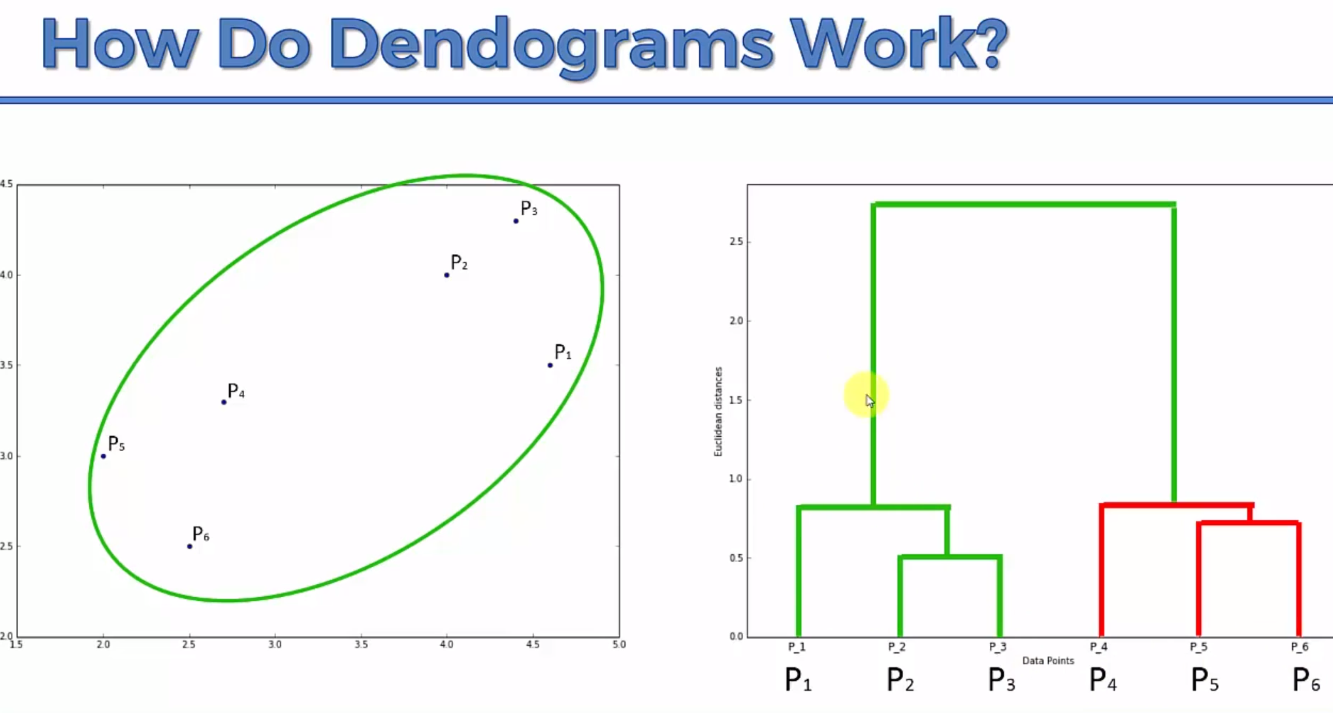
How to find the closest cluster ? ⇐ This really affects the algorithm performance.

**Option 1 :** Closest point

**Option 2 :** Furthest point

**Option 3 :** Average Distance

**Option 4 :** Distance between Centroids

**Dendograms :**

Vertical distance in dendrogram is euclidean distance(decided by user)

And we can set max distance(corresponding to dis-similarity between cluster)

It is data scientist job set that threshold.

**Step 1 :** All datapoints as individual clusters

**Step 2 :** p2 and p3 combined in one cluster

**Step 3 :** p1 and (p2,p3) combined in one cluster

**Step 4 :** p5 and p6 combined one cluster

**Step 5 :** p4 and (p5,p6) combined in one cluster

**Step 6 :** (p1,p2,p3) and (p4,p5,p6) combined in one final big cluster.