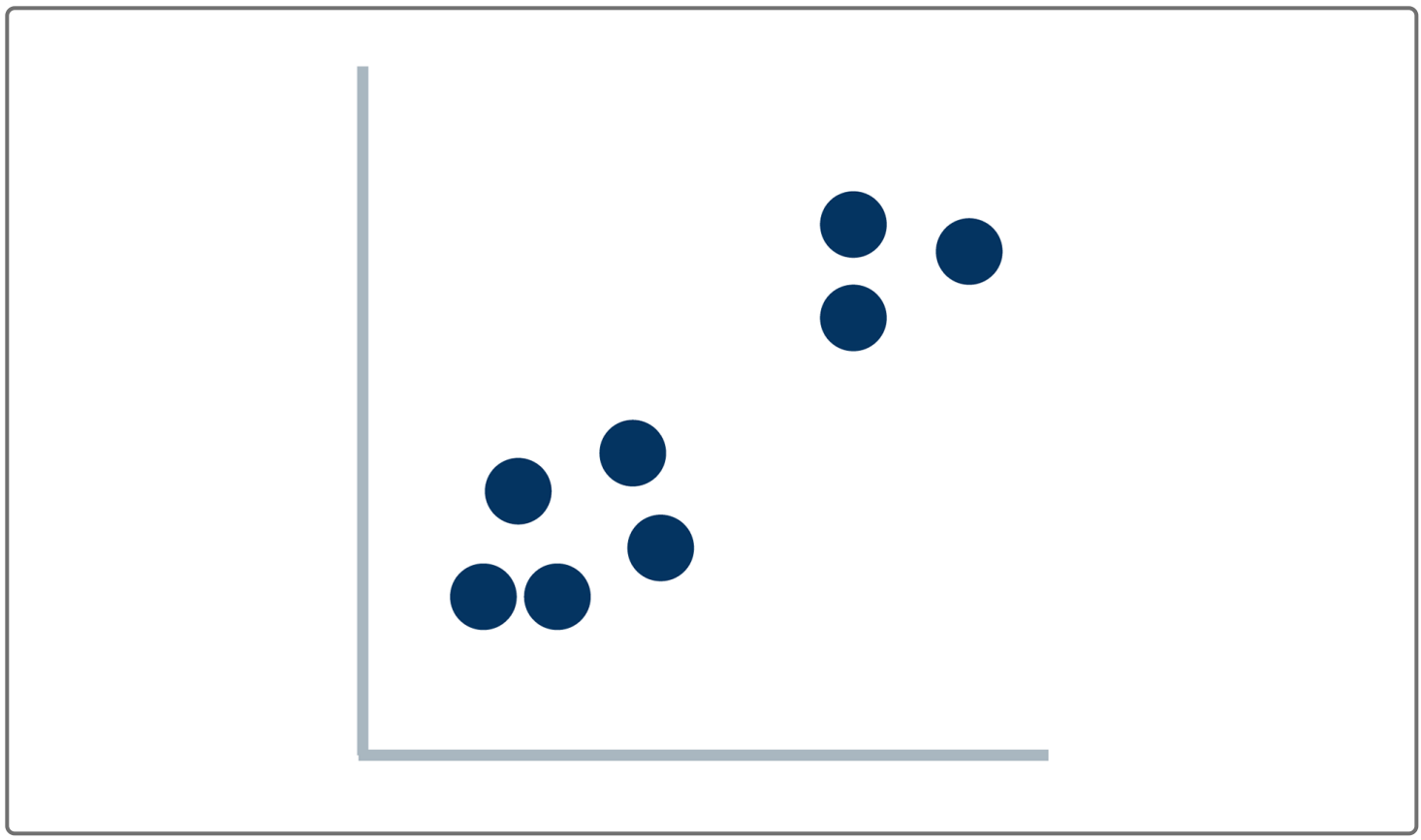
18.6.2

Dendrograms

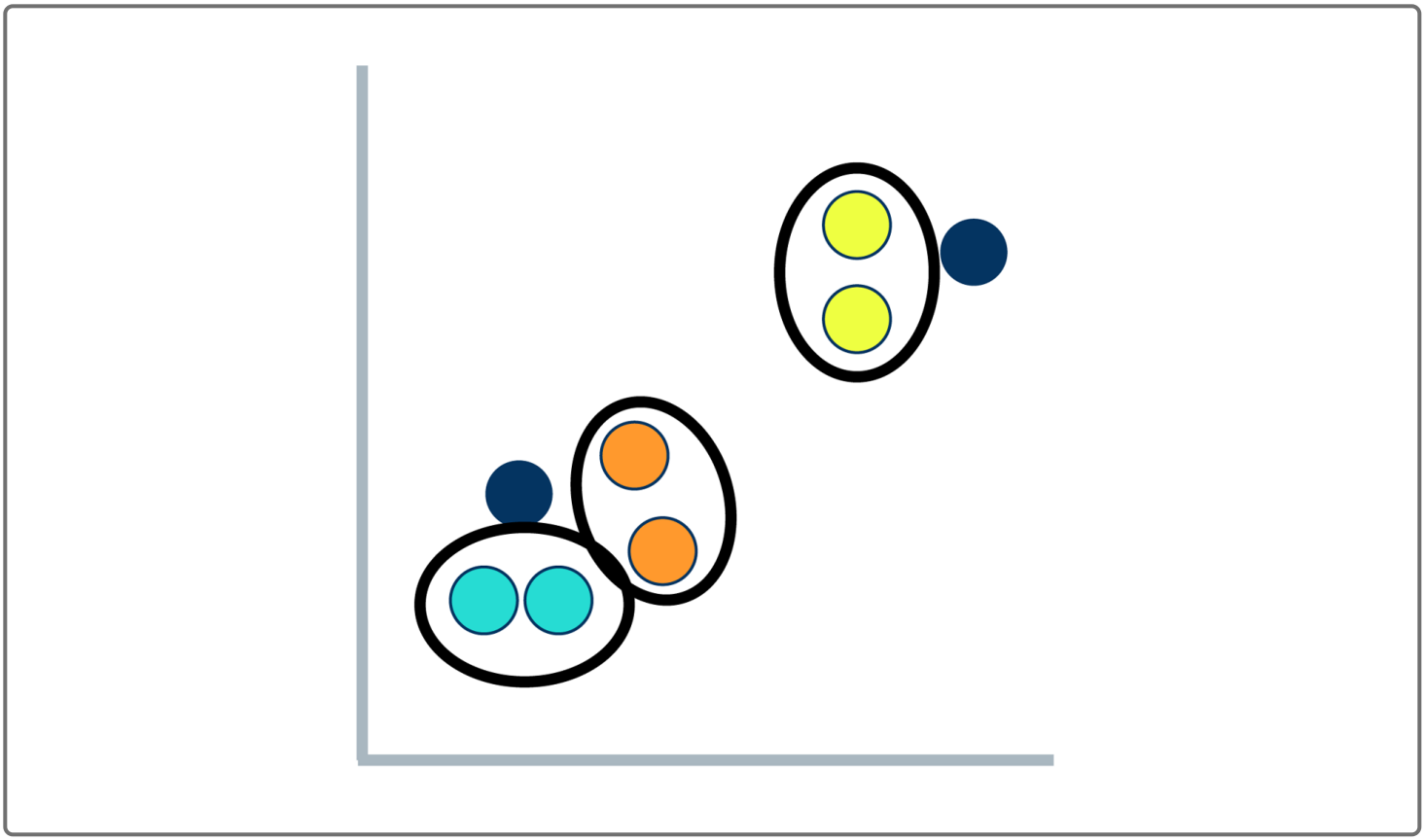
Just like with K-means, trial and error for the amount of clusters is not ideal, even in hierarchical clustering. Luckily, there is a similar method with the use of dendrograms.

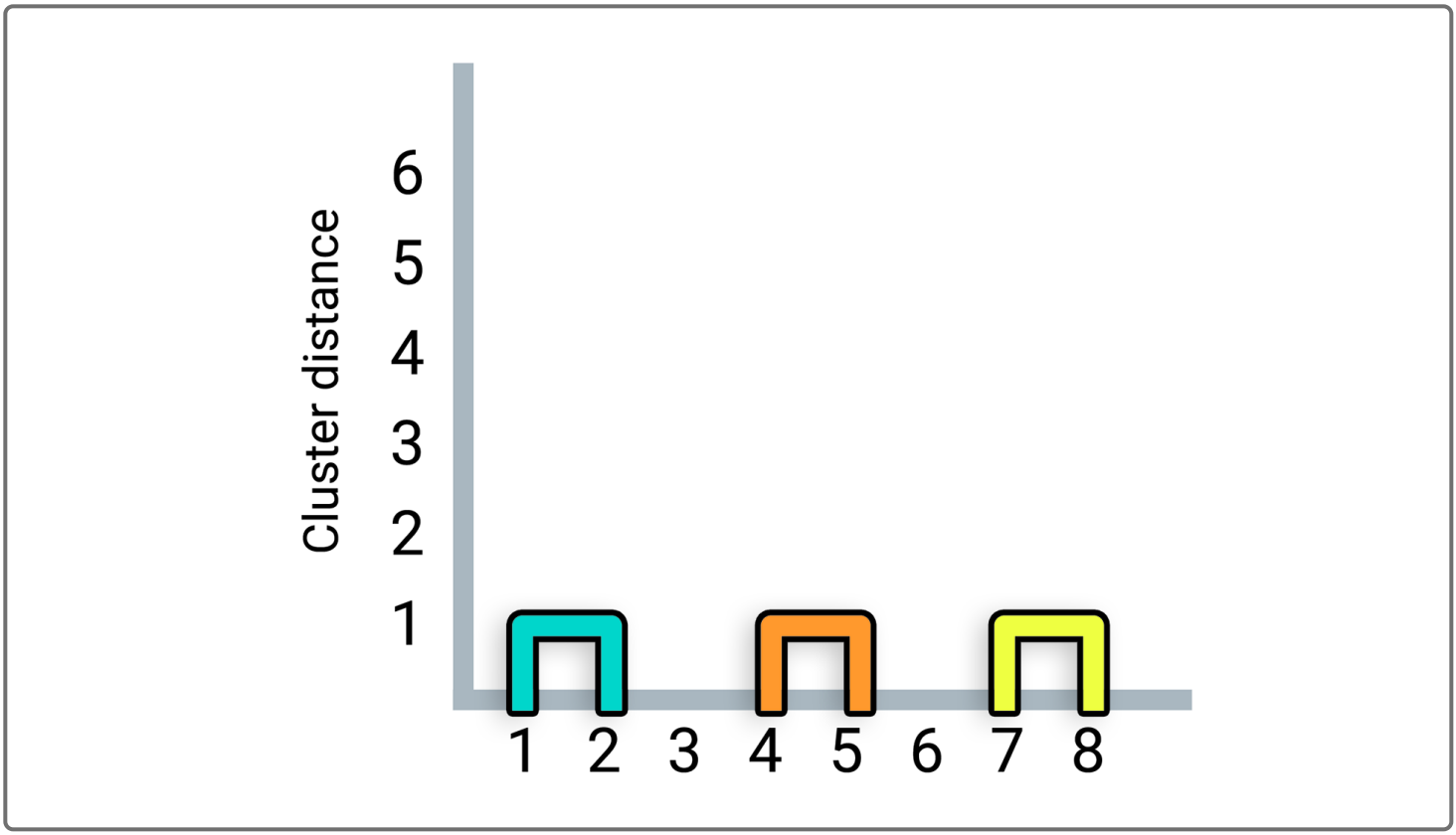
A **dendrogram** is a graph that keeps the values of the points on the x-axis, then connects all the points as they are clustered. This is similar to the elbow curve, as it gives us a better idea of the ideal amount of clusters we want to use. After making a dendrogram, you'll know how many clusters to make based on how refined you want them to be.

Let's build on the clusters from the prior exercise to create our first dendrogram:

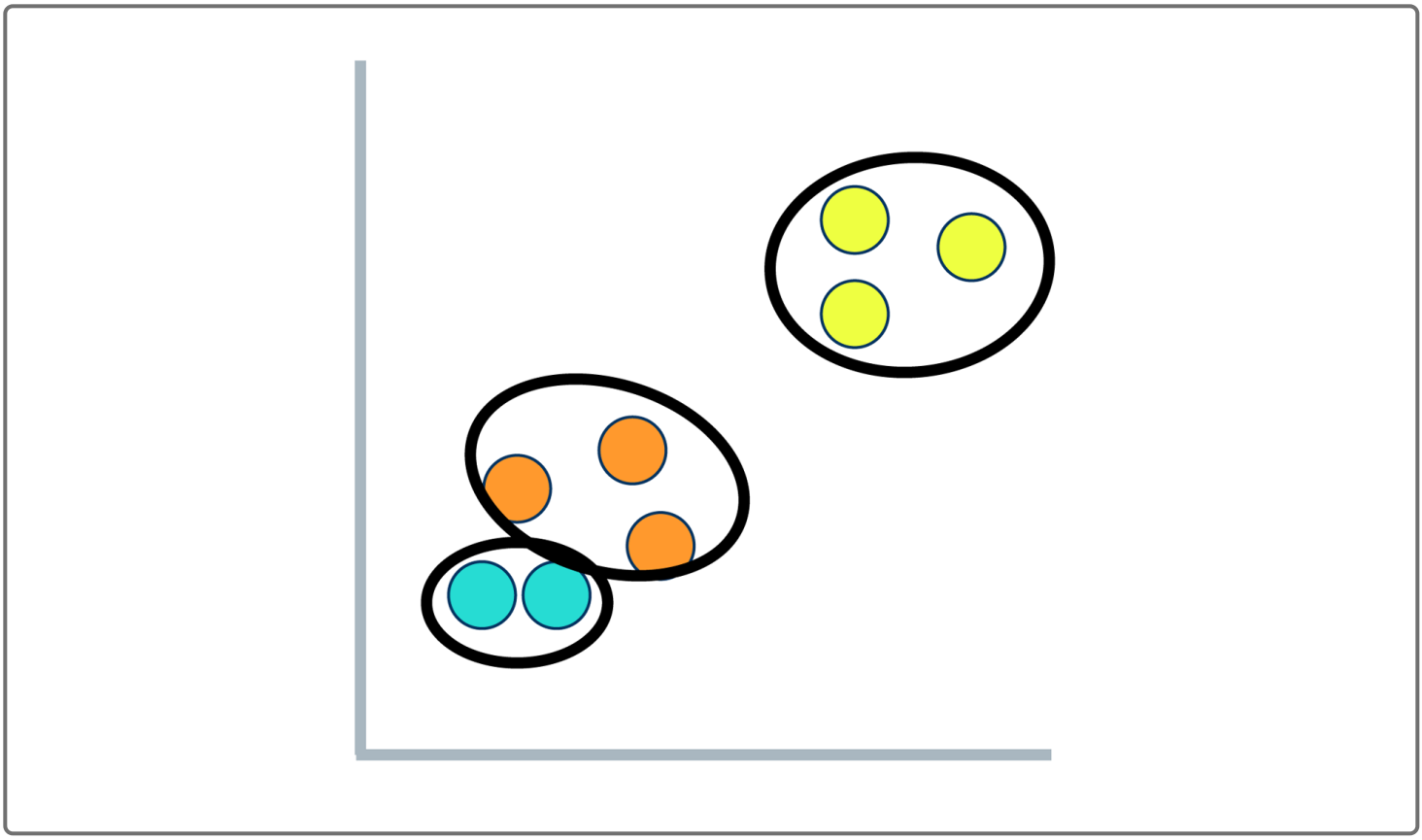


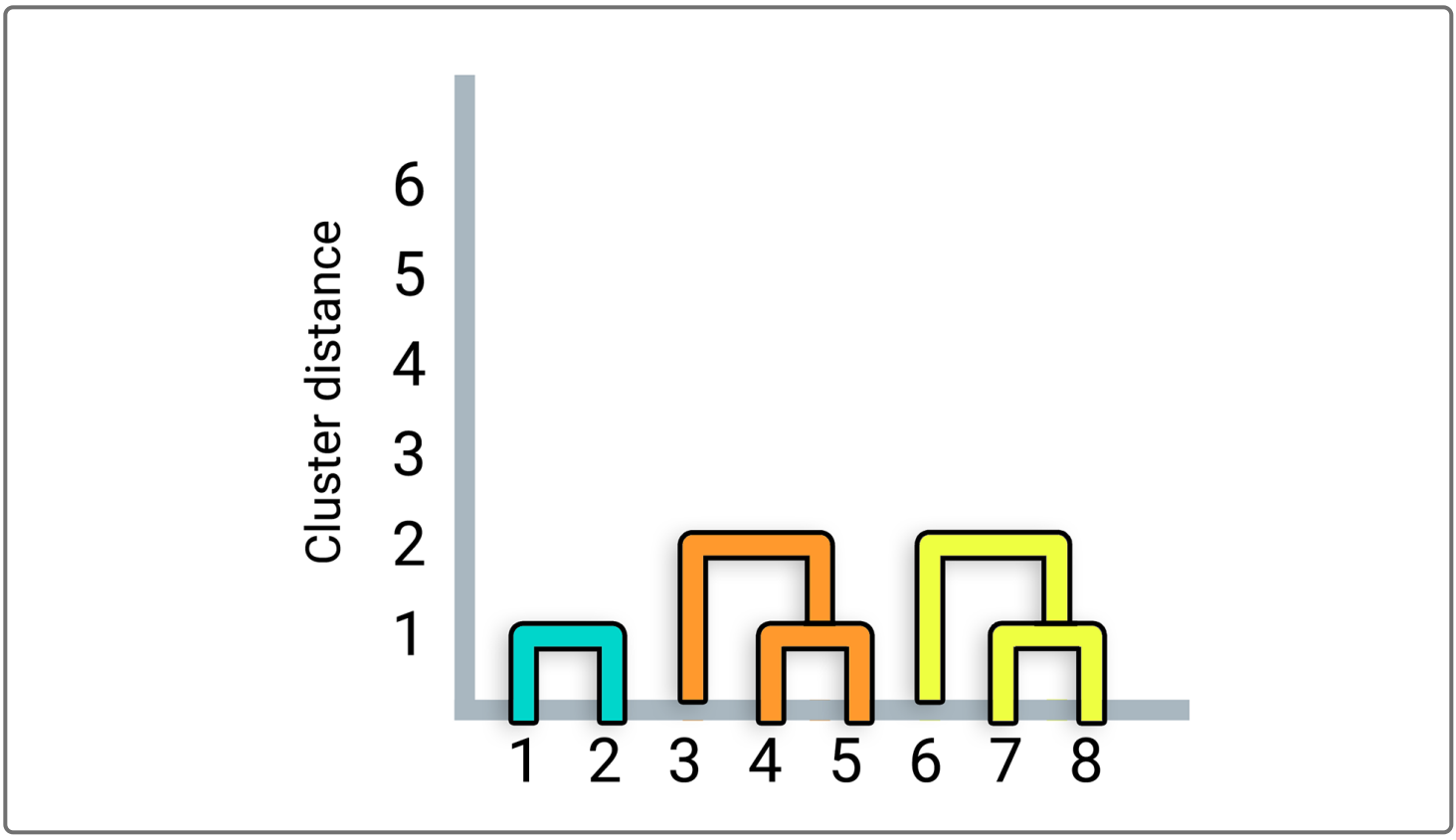
Each point on the axis represents our initial point. Next, we'll start connecting the points with a tree for each cluster that is joined:



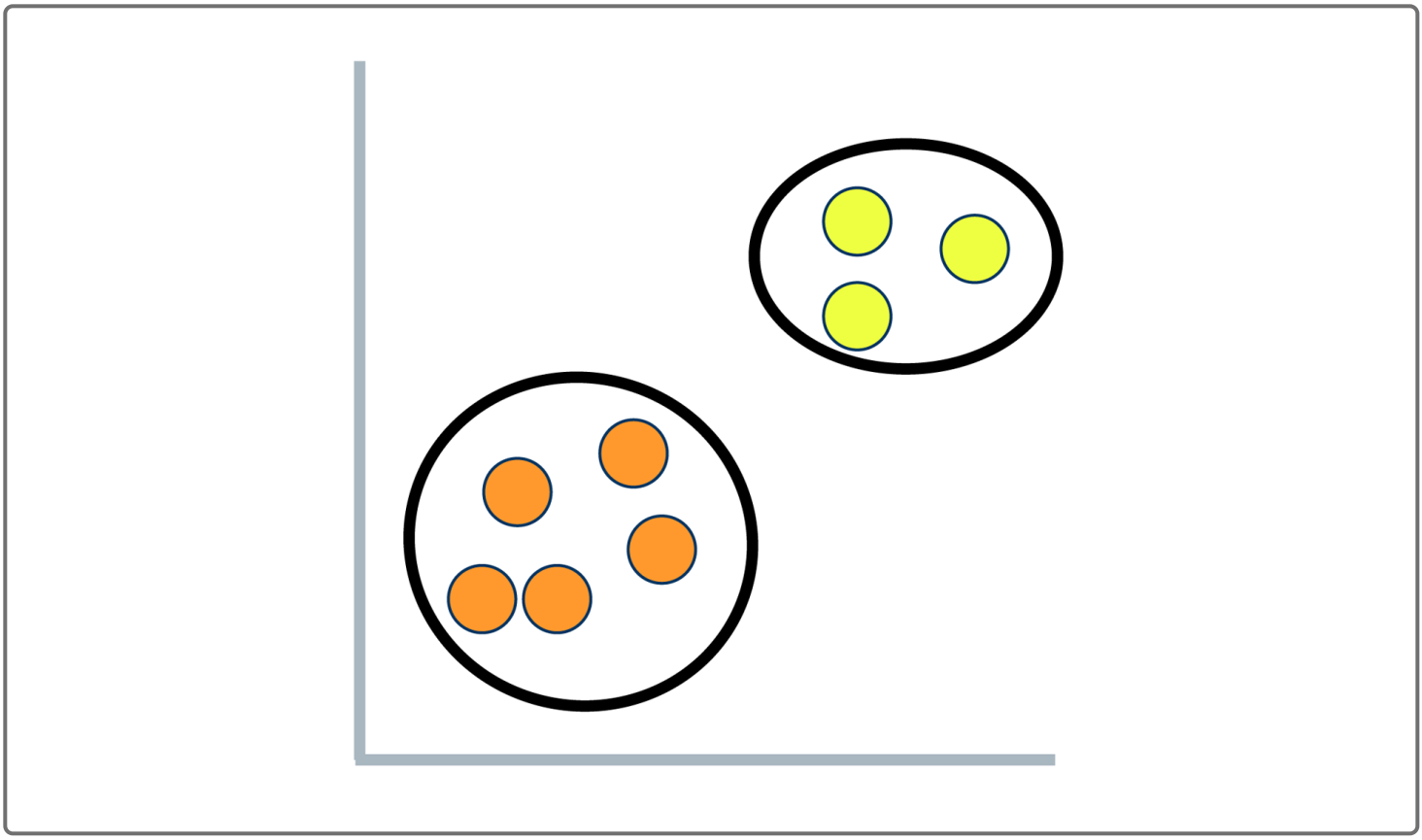


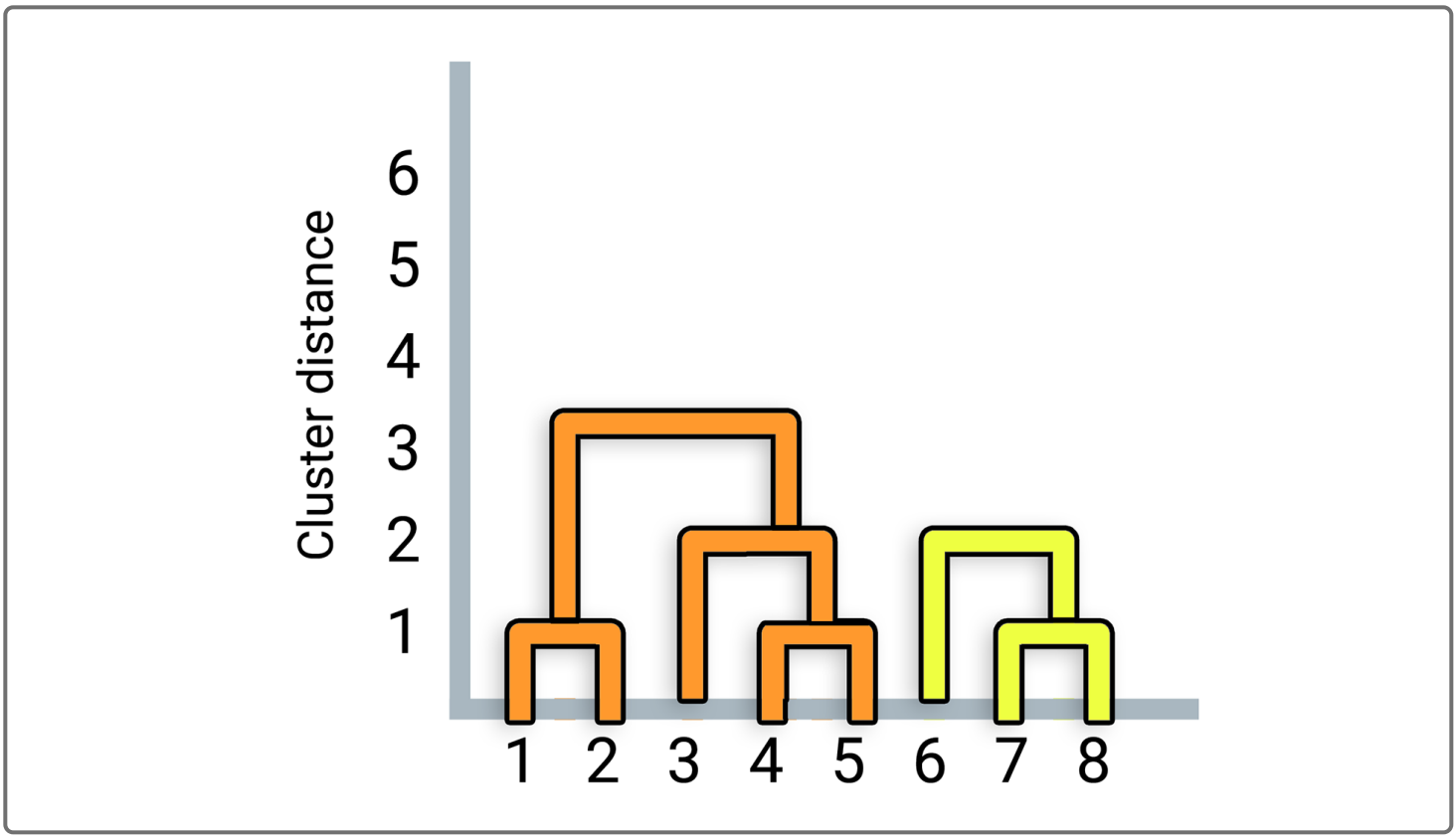
We now have three clusters, with two points unaccounted for, and the points that were merged into a cluster will be connected in the dendrogram:



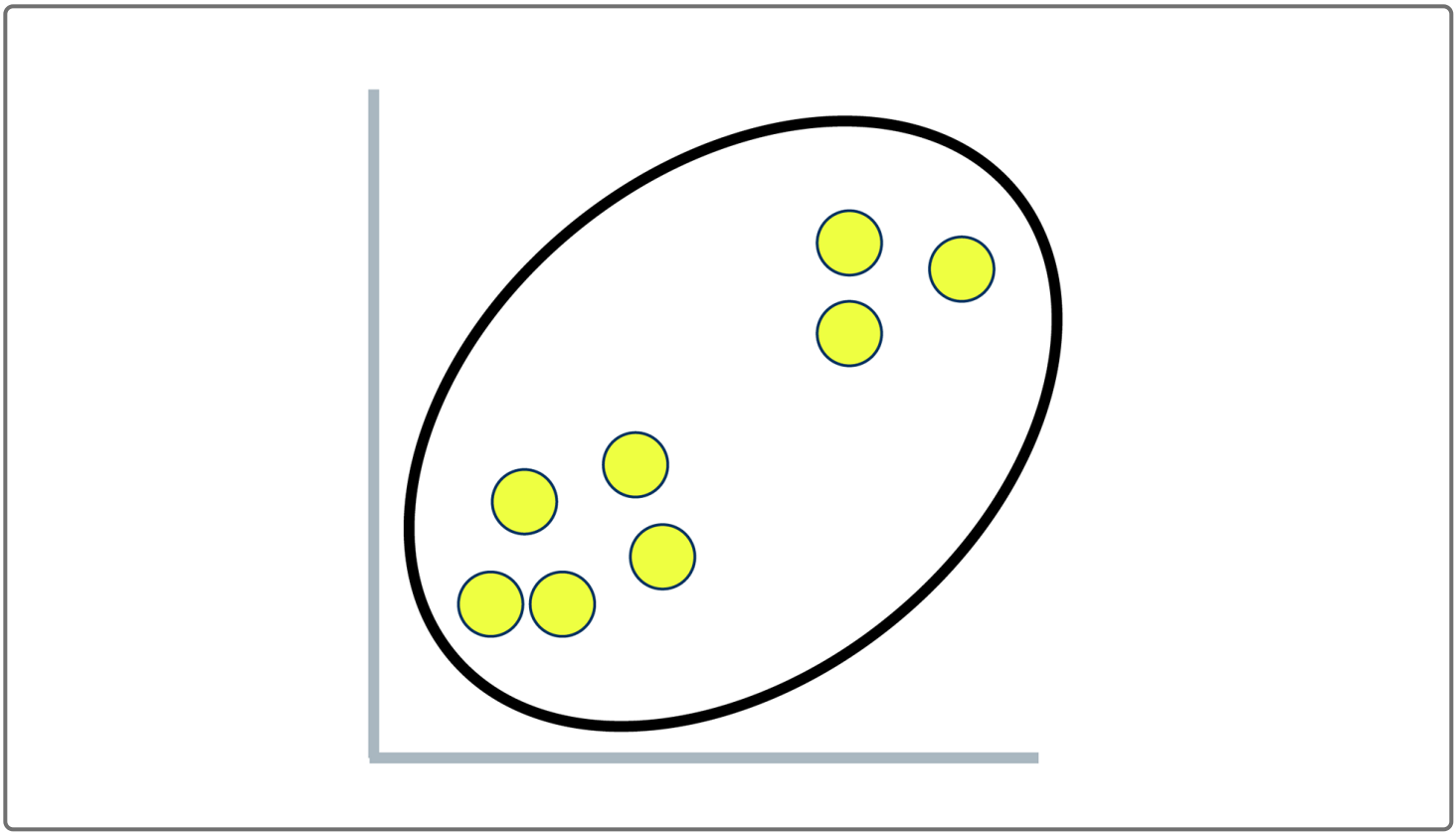


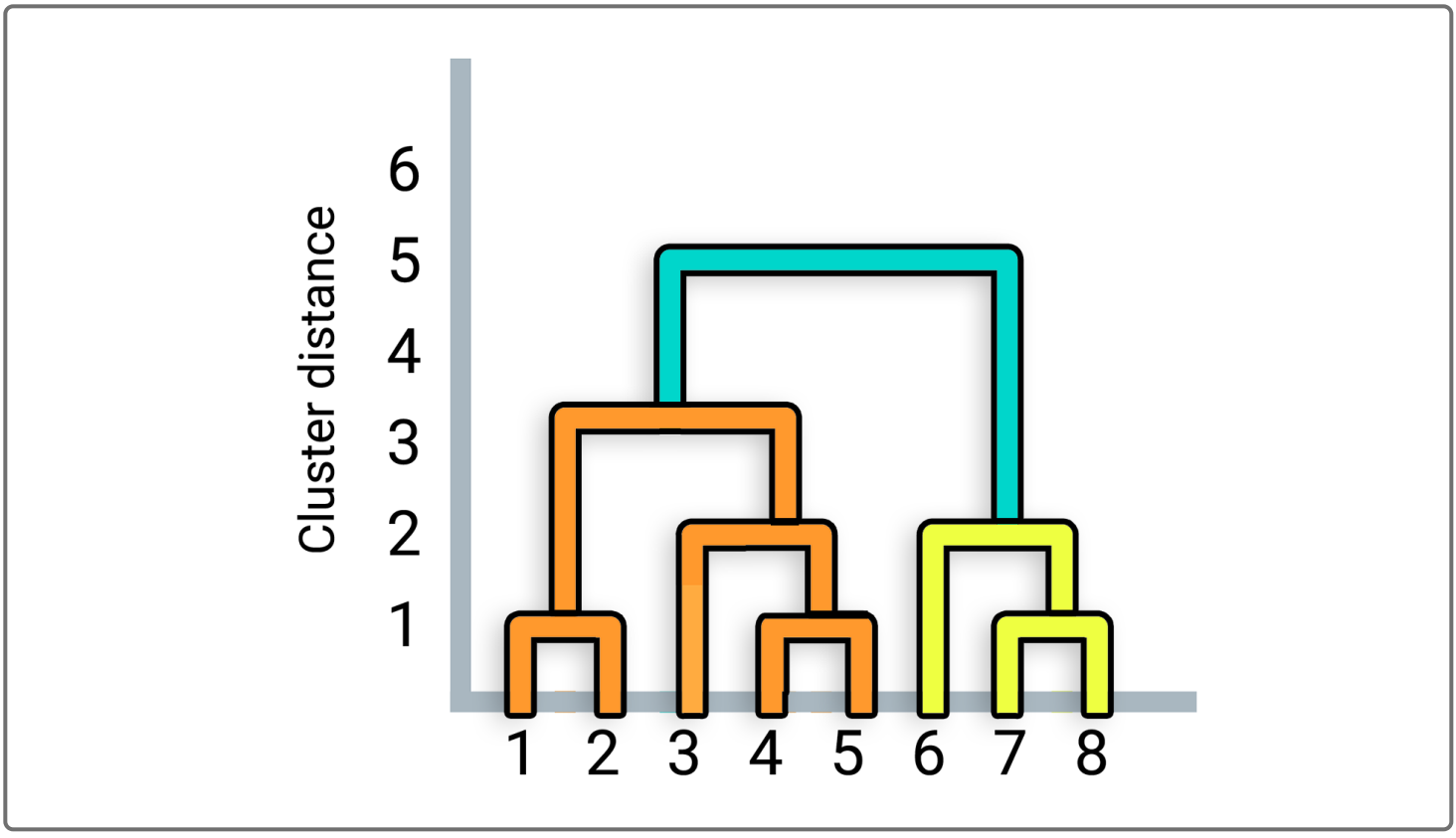
Next, combine the next merger of two clusters:



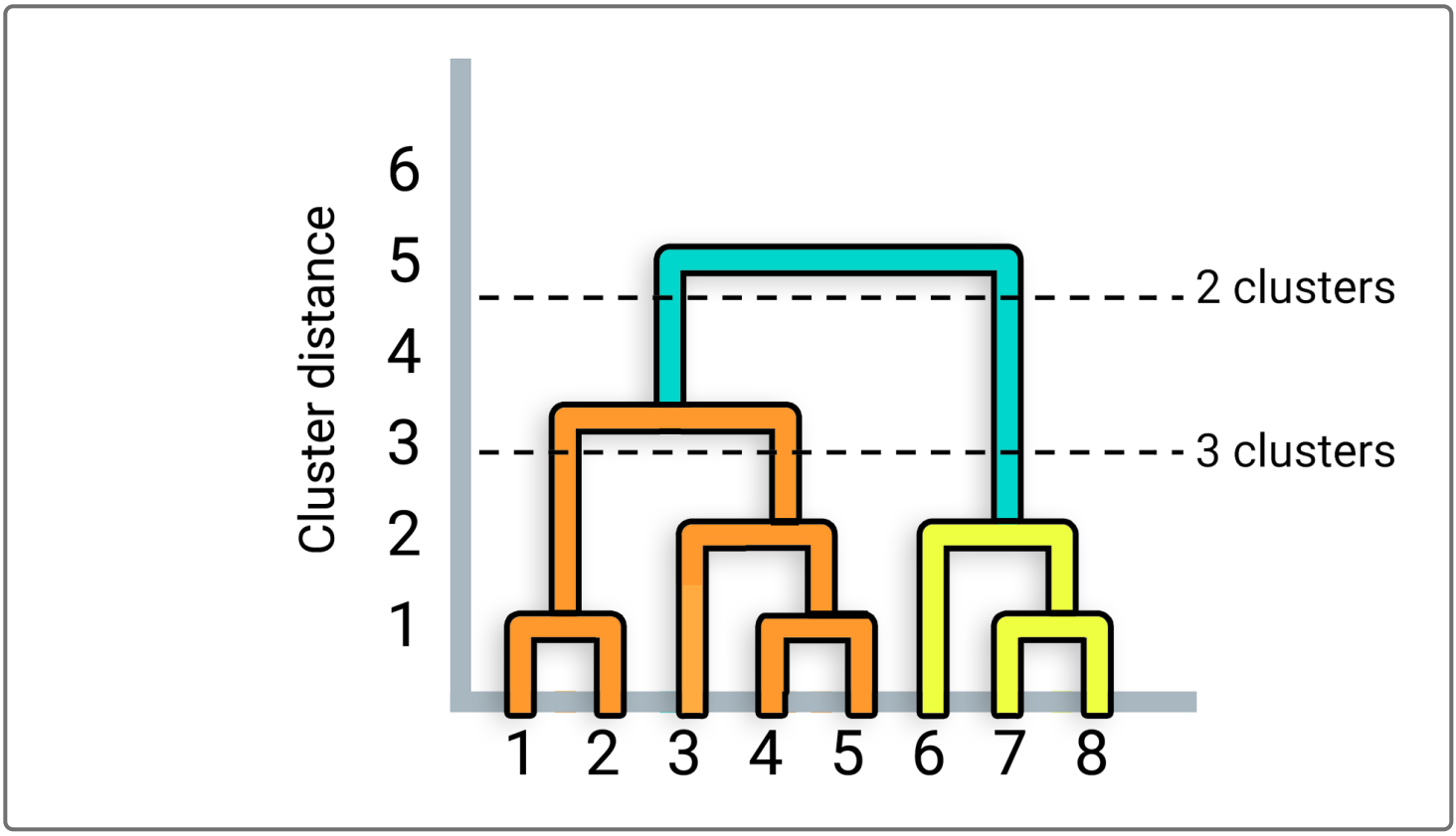

Cluster distance
6
5
4
3
2
1
1 2 3 4 5 6 7 8

In the previous example, we didn't need to merge all points into one cluster. As you'll see, we'll join the remaining two clusters into one and plot it on the dendrogram:





Now we can set the height to determine how many clusters we want:



Choose the tolerance of cluster distance and draw a line at that point. Note how many clusters are beneath that line.

The solid horizontal lines indicate how far apart the clusters are from each other before they're merged. The dashed line for three clusters indicates that the longest horizontal lines below it must be that far away from each other to form three clusters. The dashed line for two clusters indicates that the top horizontal lines are at a greater distance from three clusters to two clusters. Finally, the horizontal line above the dashed line for two clusters would be an even greater distance if we were to form one cluster.

Therefore, the greater the distance, the less similarity exists. There is no correct choice between two and three here. Again, that decision is left up to you, the analyst, to decide how refined you want the clusters to be.

In the next section, we'll return to the iris dataset and see how running hierarchical clustering works.