Hierarchical Cluster

Enrique J. De La Hoz D.

UTB - Data Science

The closest observation to a Pair

The Closest Observation to a Pair

	1	2	3
2	11.7		
3	16.8	18.0	
4	**10.0**	20.6	15.8

- Is 2 is closest to group 1,4?
- Is 3 is closest to group 1,4?

Figure 1:

Linkage Criteria: Complete

	1	2	3
2	11.7		
3	16.8	18.0	
4	**10.0**	20.6	15.8

Figure 2:

- Is 2 is closest to group 1,4?
  - max(D(2,1), D(2,4)) = 20.6
- Is 3 is closest to group 1,4?
  - max(D(3,1), D(3,4)) = 16.8

Hierarchical Clustering

- Complete Linkage: maximum distance between two sets

Grouping With Linkage & Distance

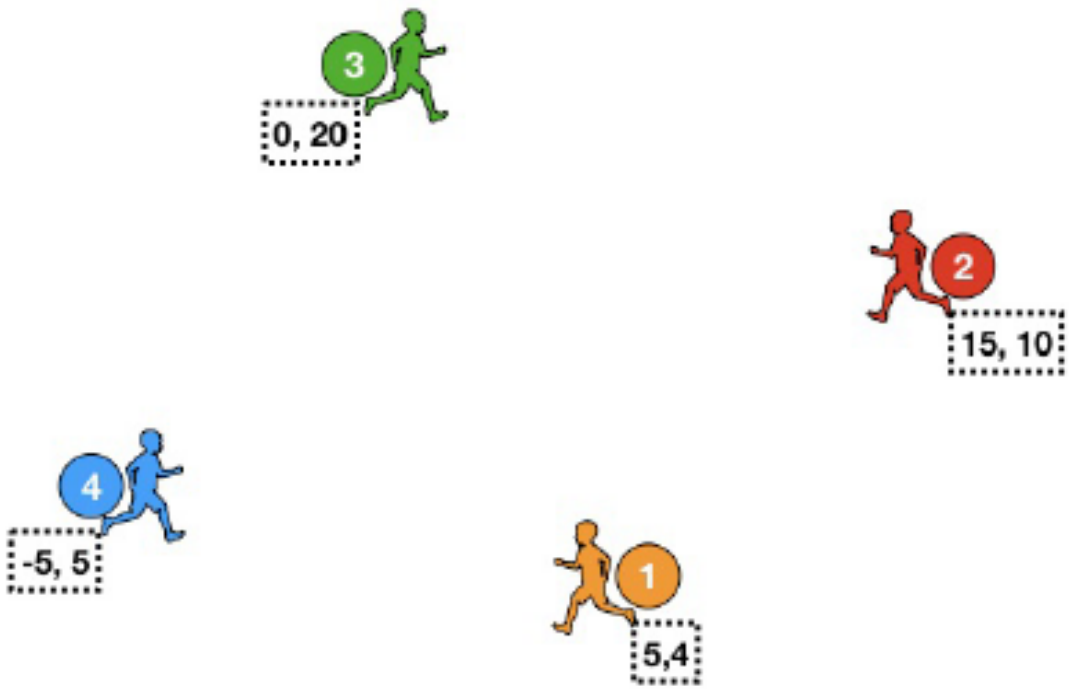


Figure 3:

Grouping With Linkage & Distance

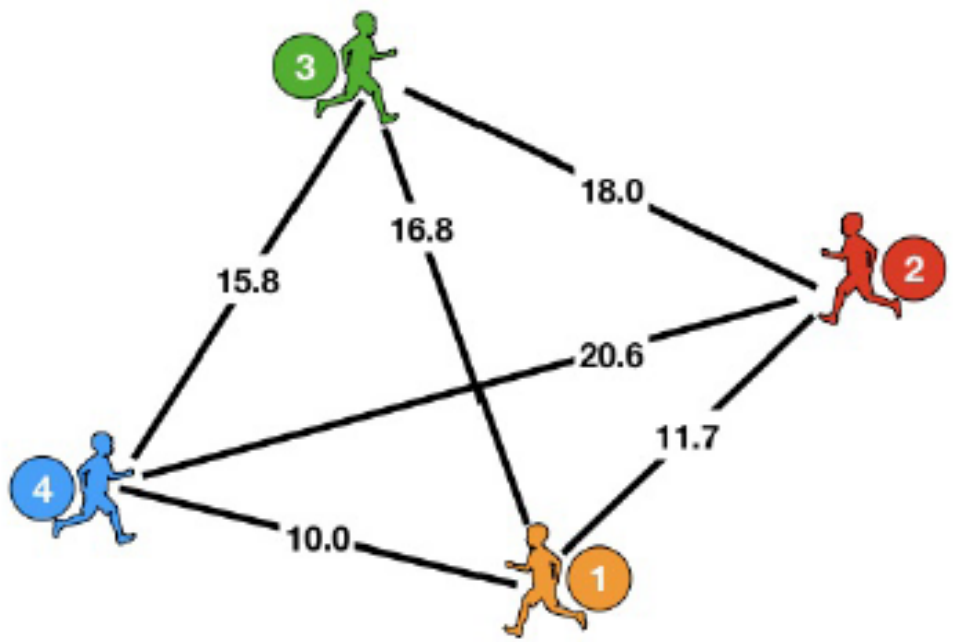


Figure 4:

Grouping With Linkage & Distance

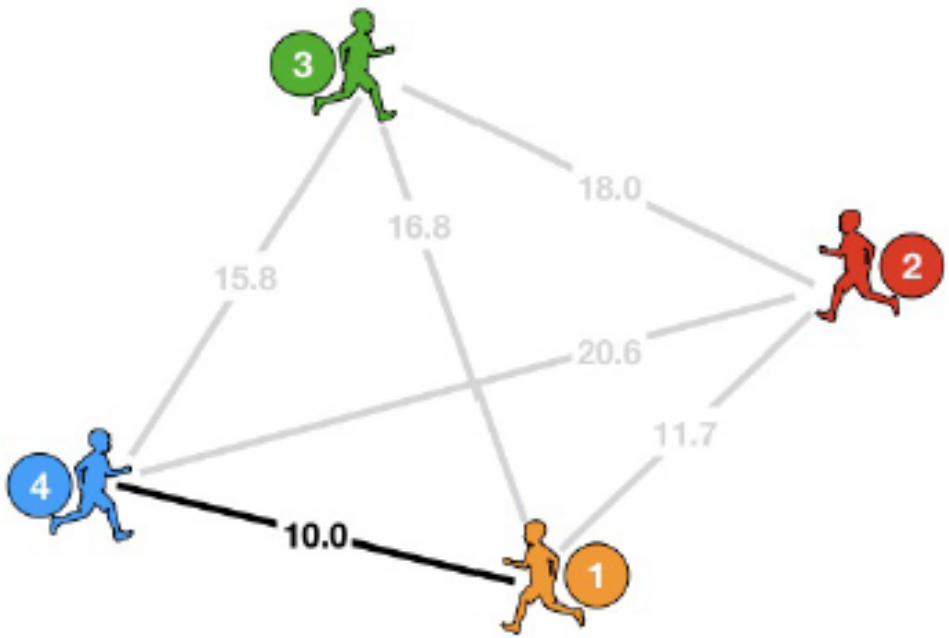


Figure 5:

Grouping With Linkage & Distance

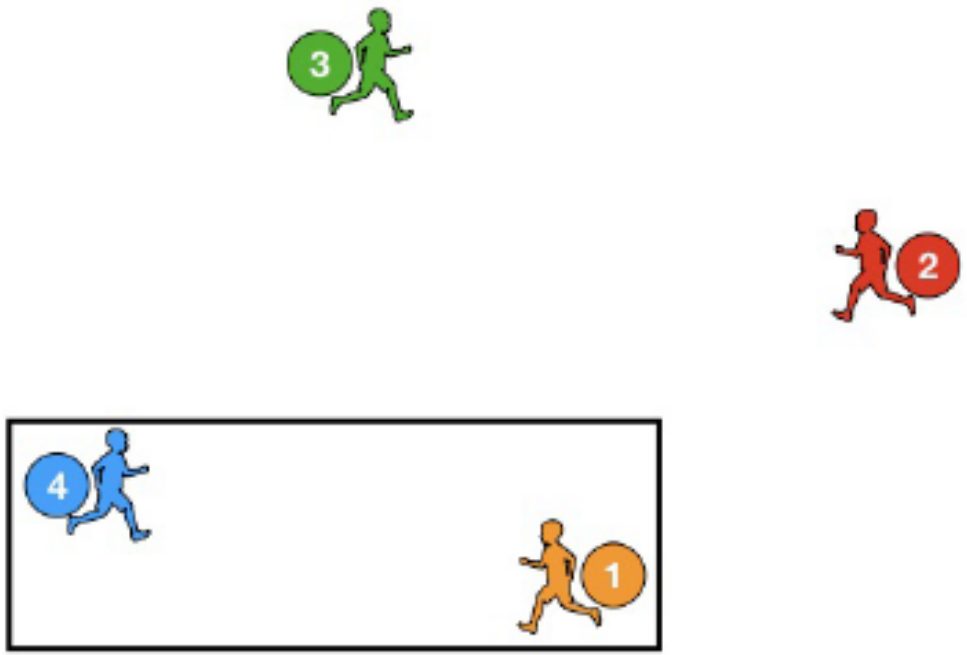


Figure 6:

Grouping With Linkage & Distance

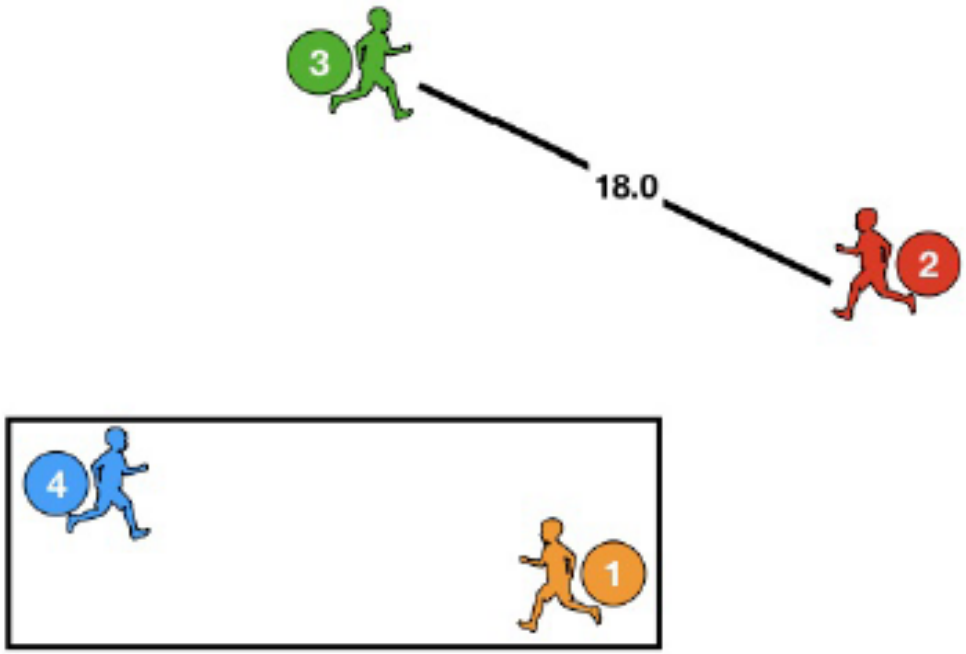


Figure 7:

Grouping With Linkage & Distance

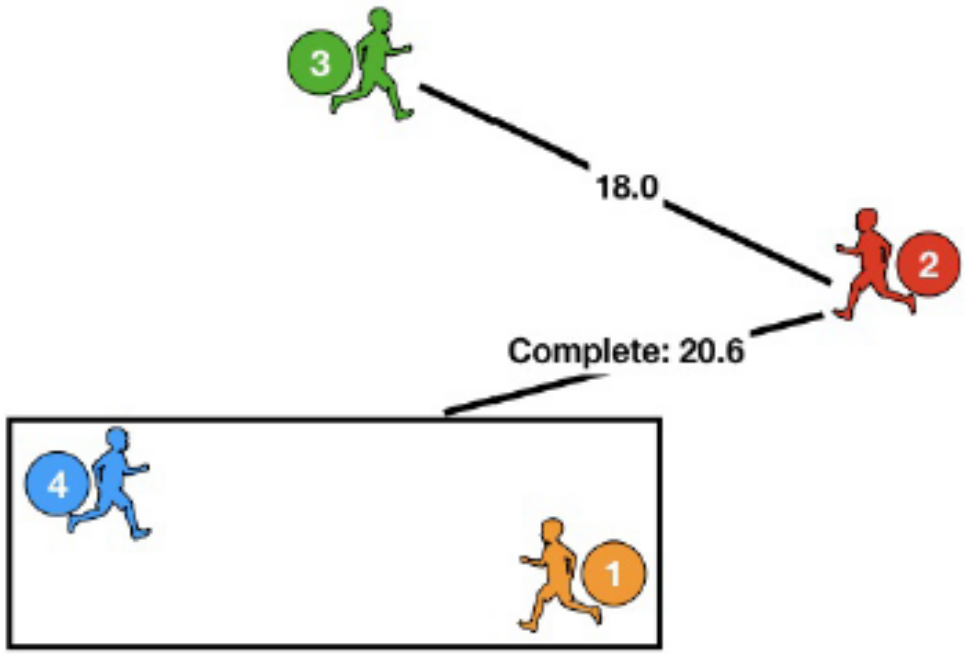


Figure 8:

Grouping With Linkage & Distance

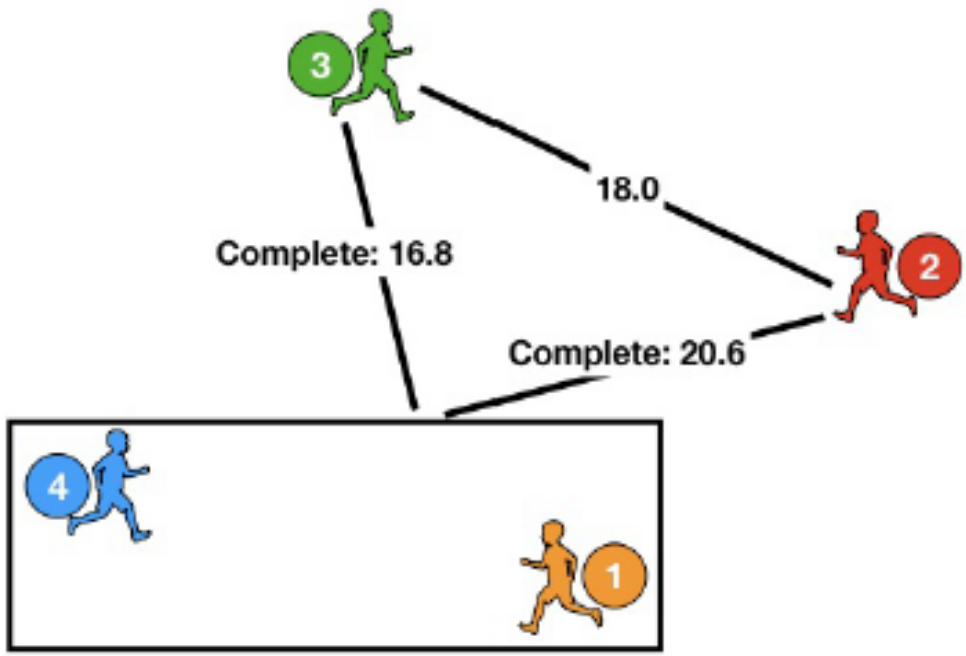


Figure 9:

Grouping With Linkage & Distance

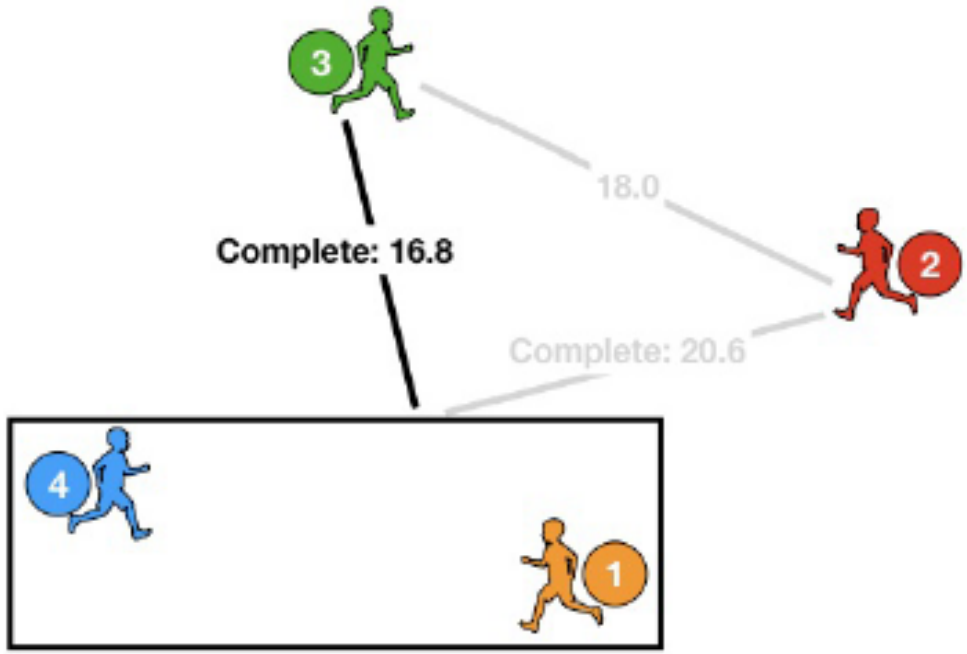


Figure 10:

Grouping With Linkage & Distance

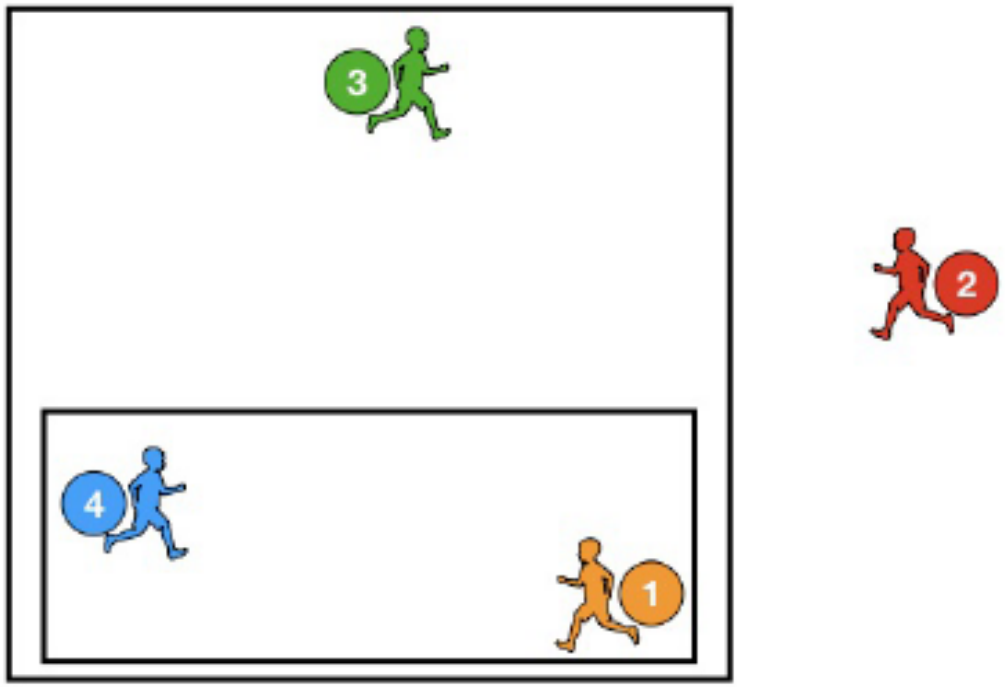


Figure 11:

Grouping With Linkage & Distance

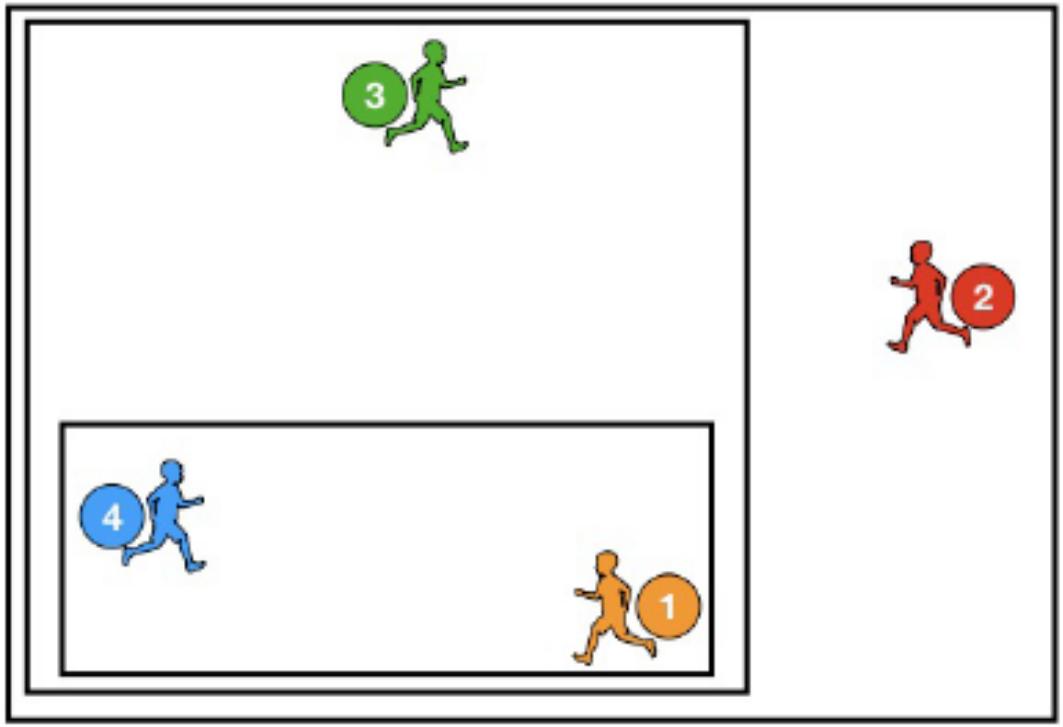


Figure 12:

Linkage Criteria

- Complete Linkage: maximum distance between two sets
- Single Linkage: minimum distance between two sets
- Average Linkage: average distance between two sets

Defining the number of clusters

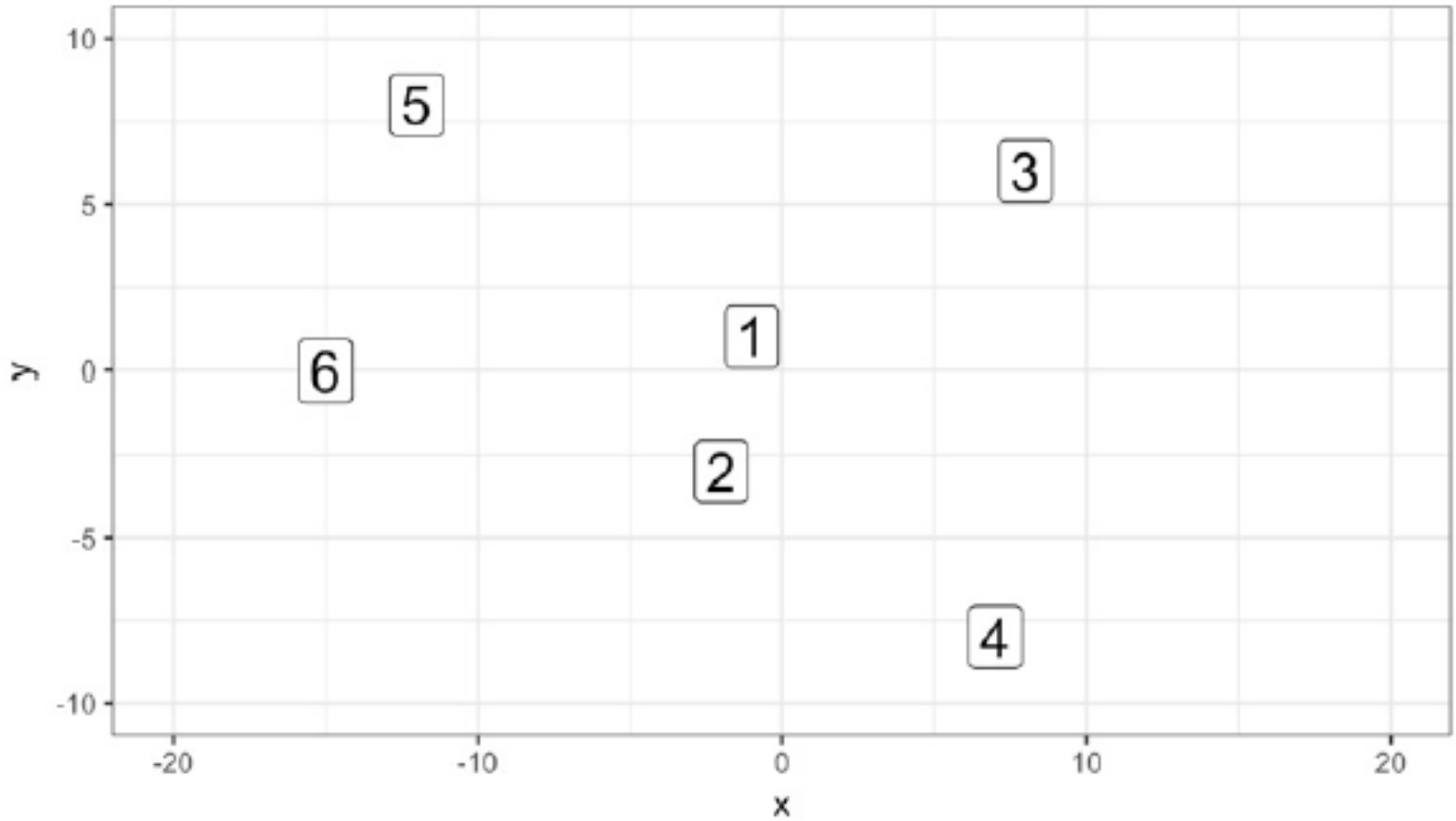


Figure 13:

Defining the number of clusters

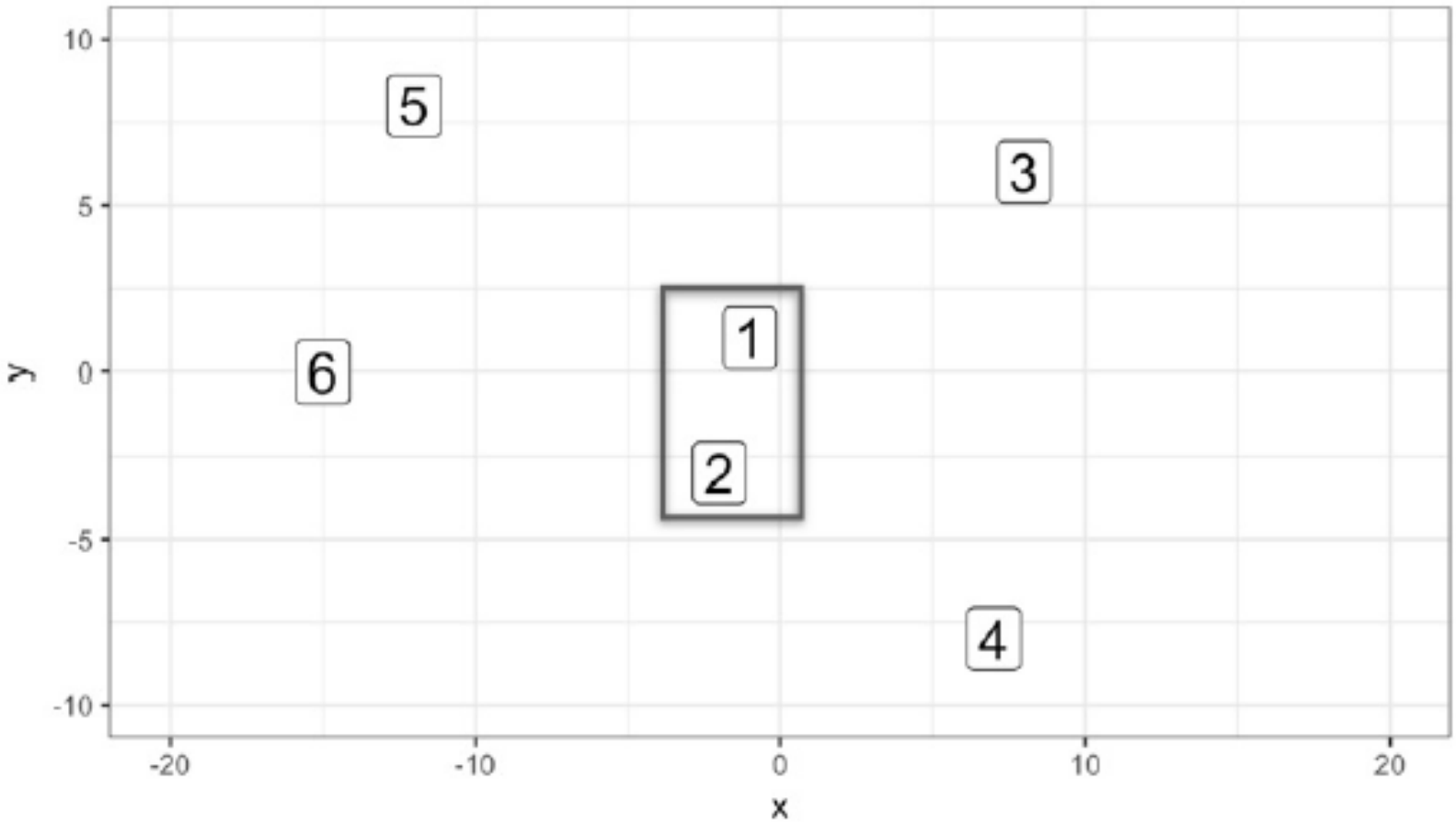


Figure 14:

Defining the number of clusters

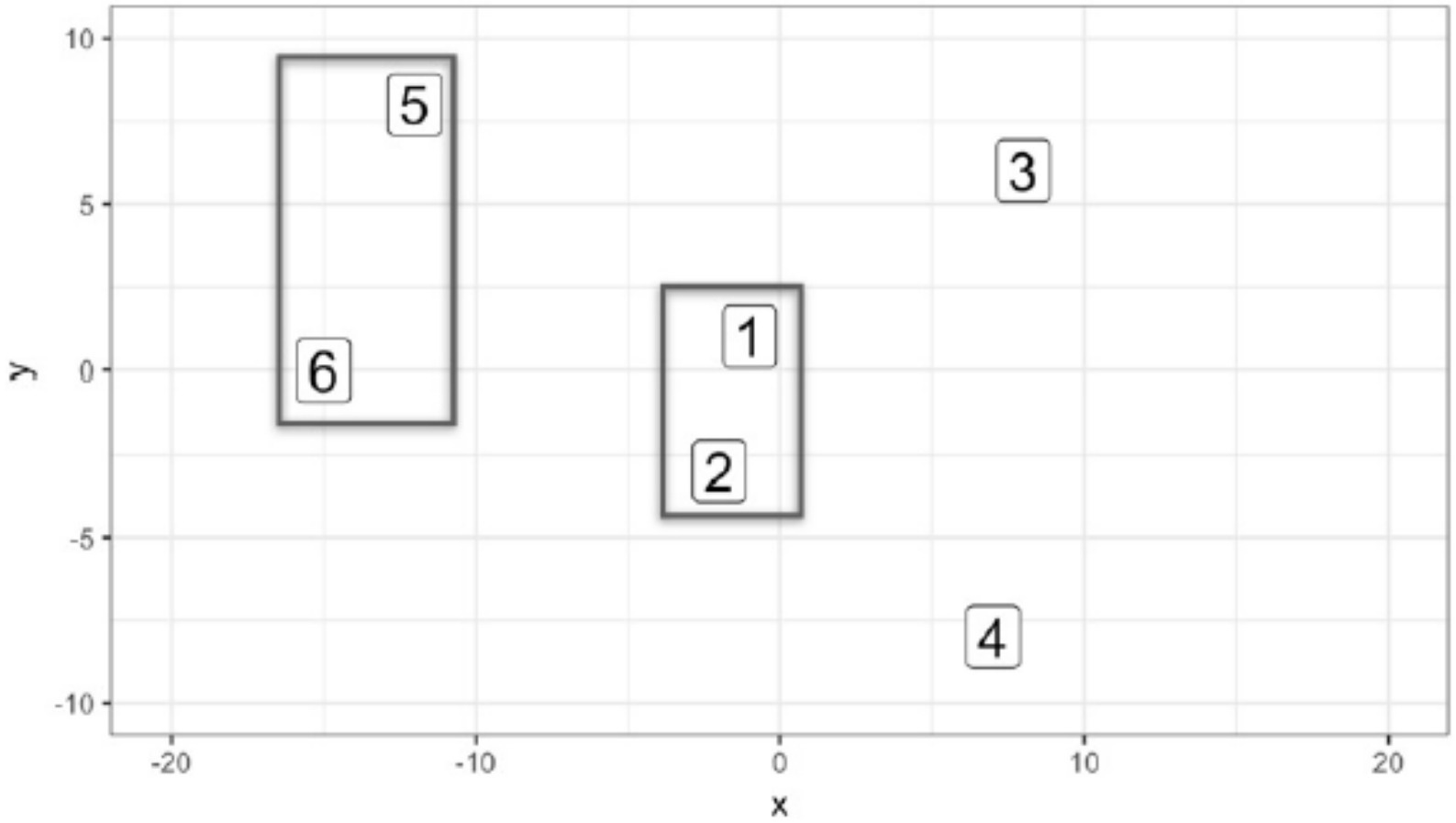


Figure 15:

Defining the number of clusters

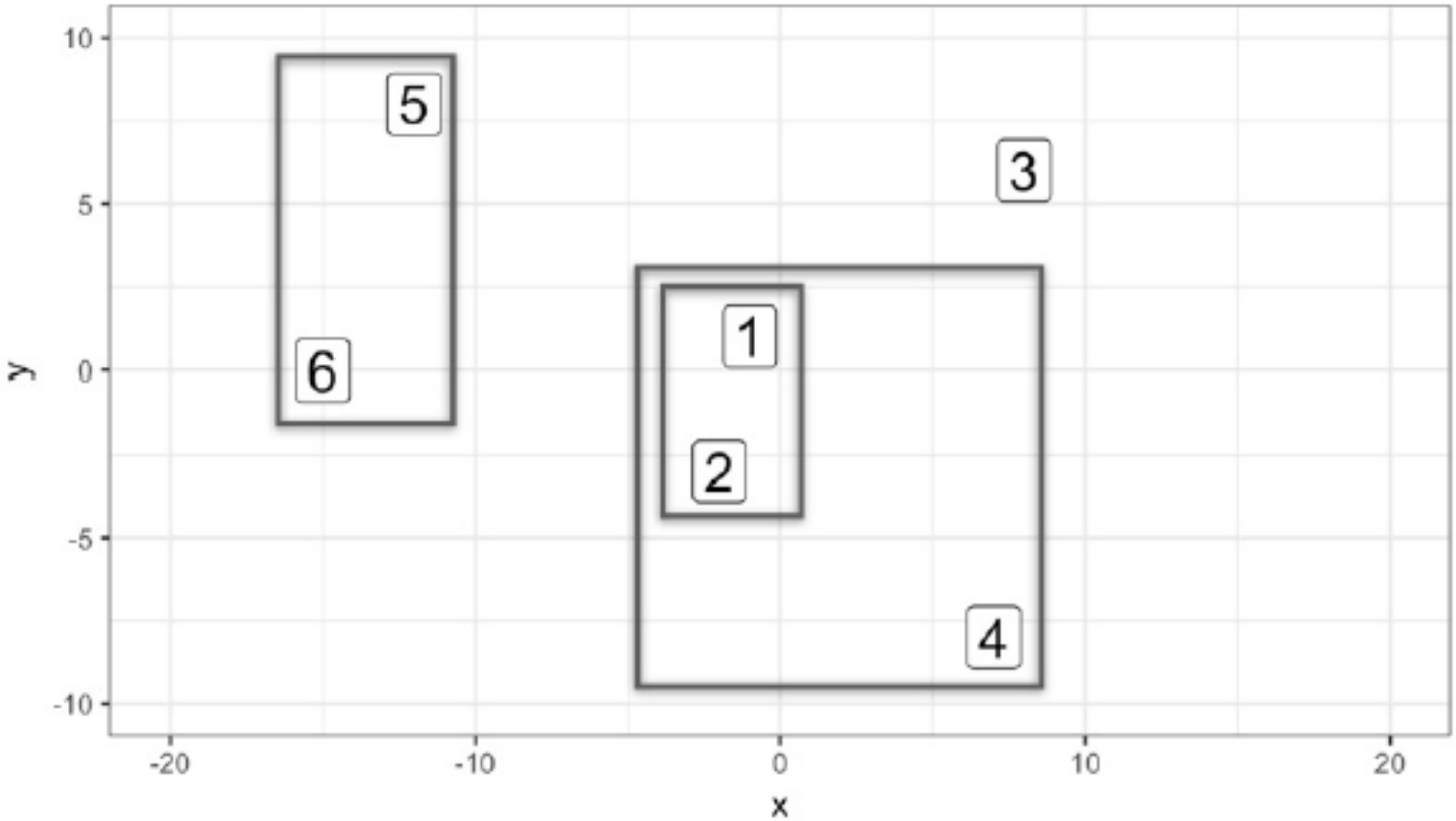


Figure 16:

Defining the number of clusters

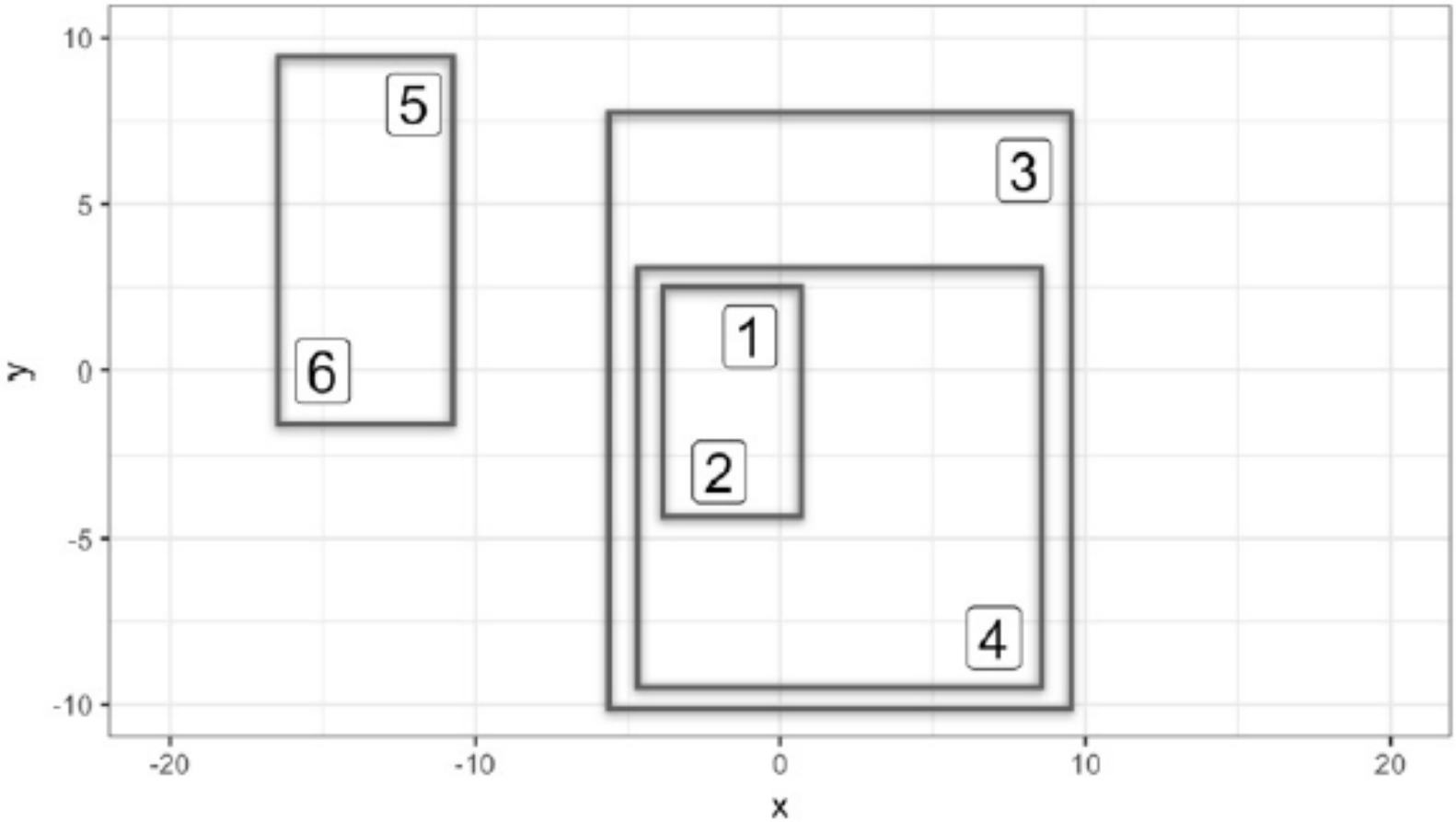


Figure 17:

Defining the number of clusters

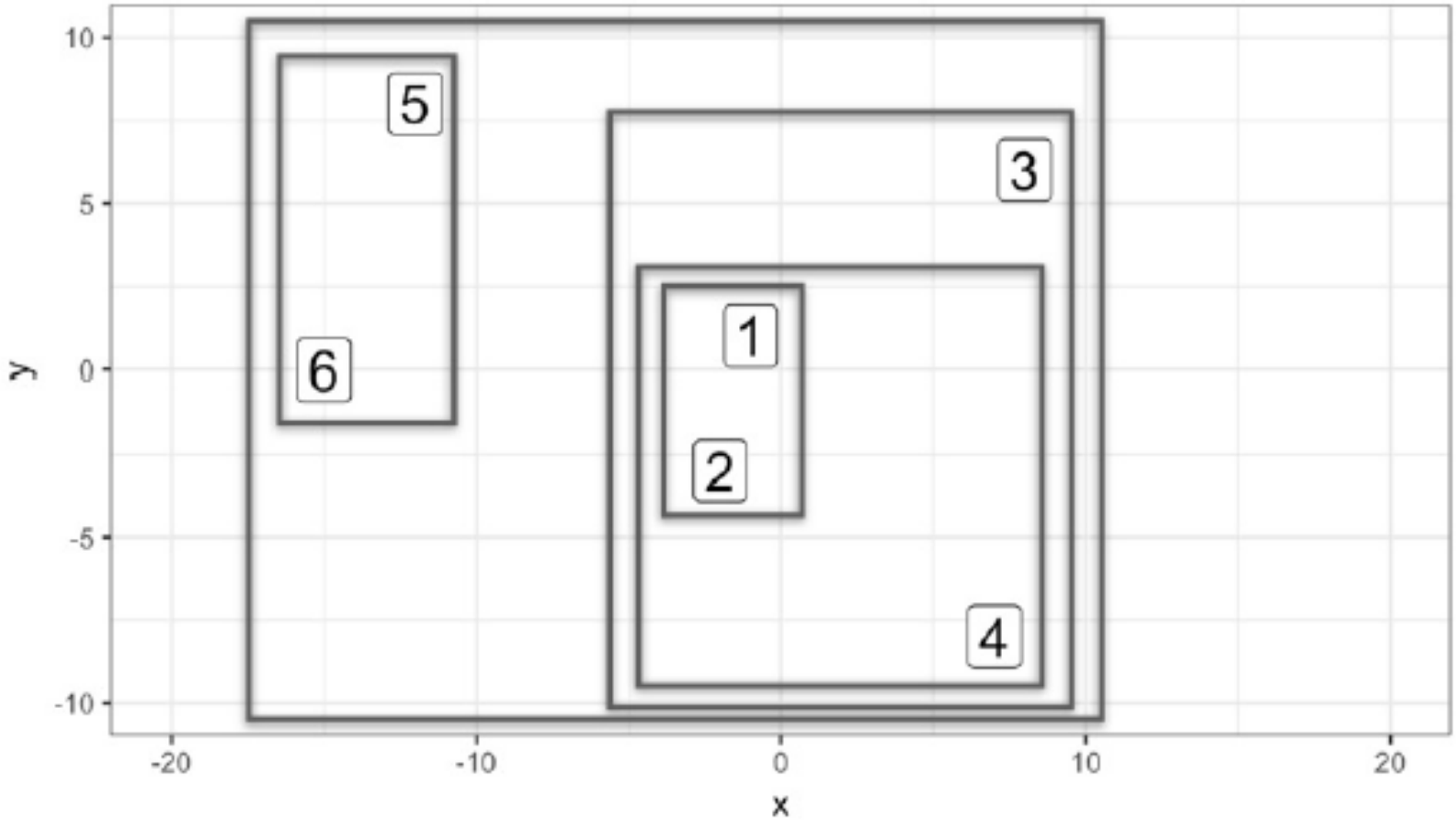


Figure 18:

Defining the number of clusters

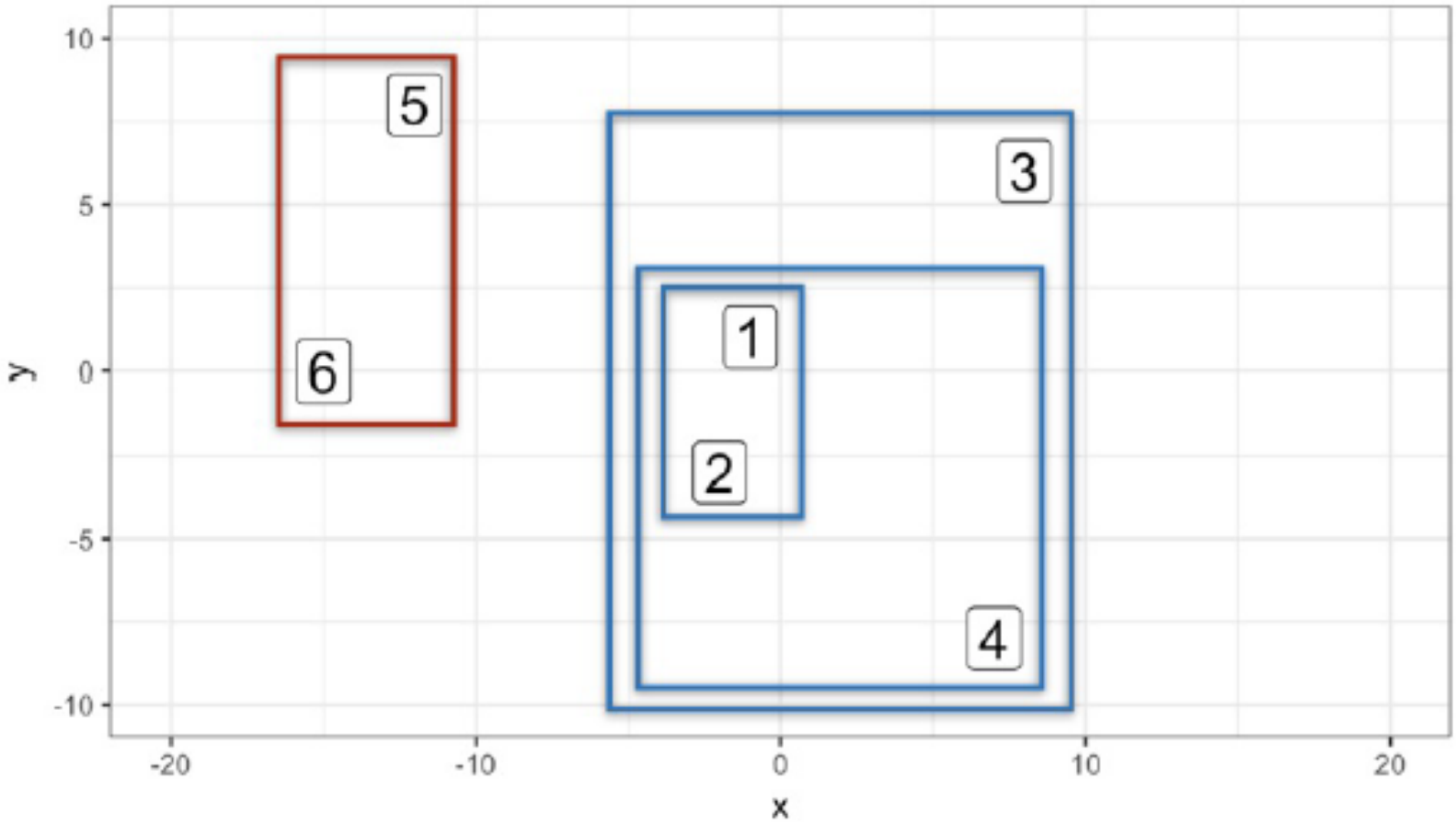


Figure 19:

Defining the number of clusters

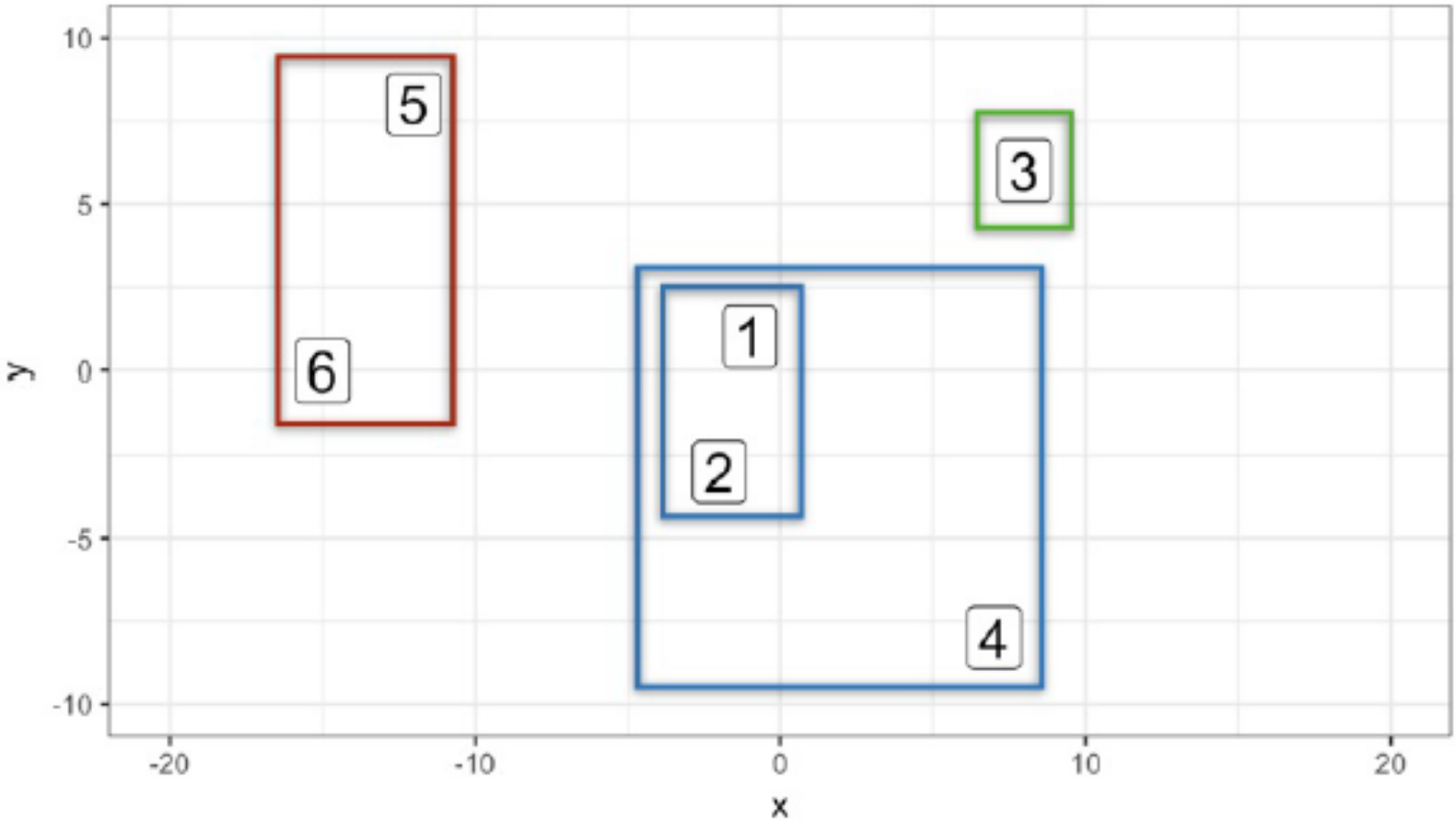


Figure 20:

Hierarchical Clustering in R

```
dist_players <- dist(players, method = "euclidean")
hc_players <- hclust(dist_players, method = 'complete')
```

Extracting K Clusters

```
cluster_assignments <- cutree(hc_players, k = 2)
print(cluster_assignments)
```

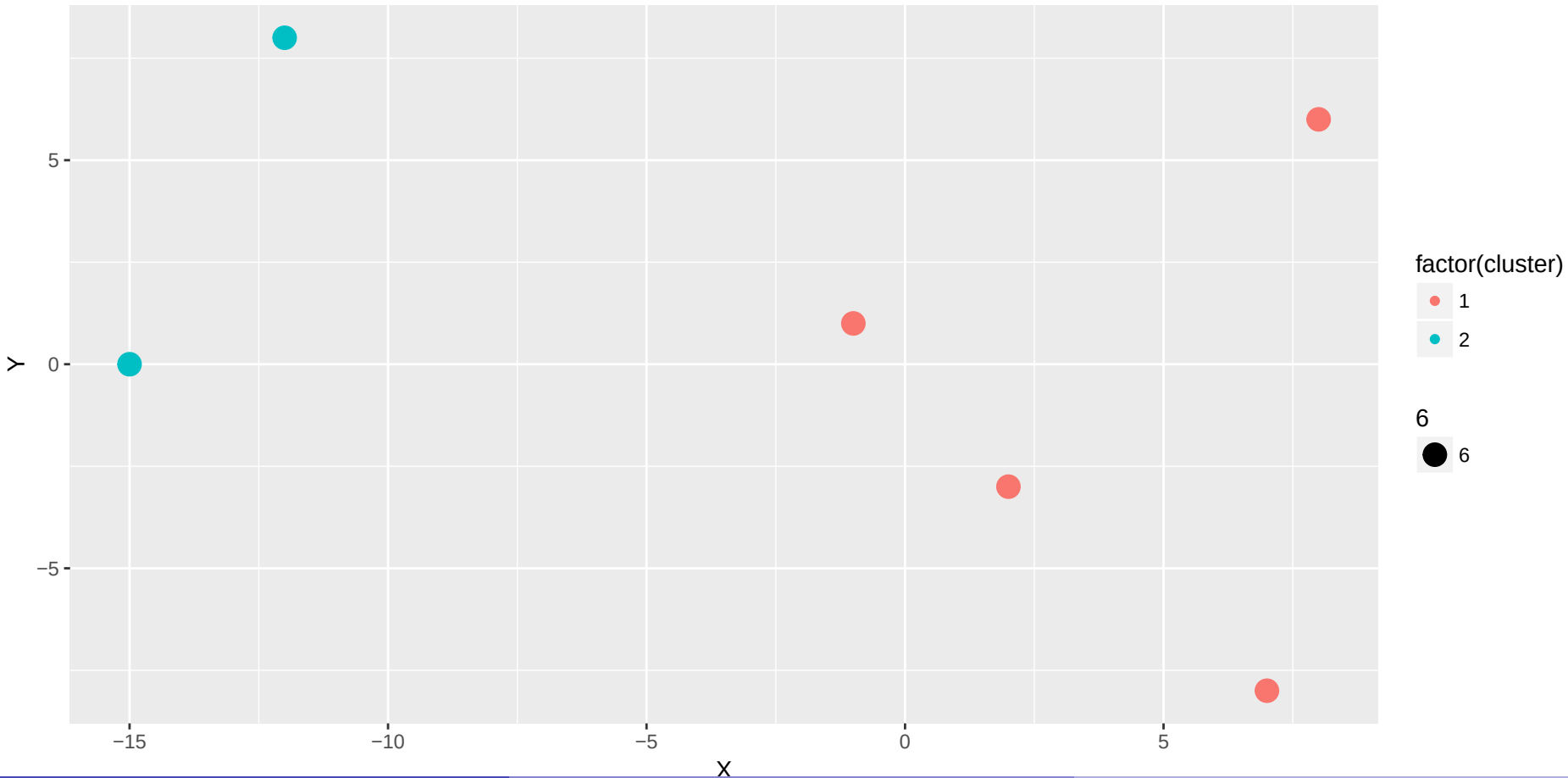
```
## [1] 1 1 1 1 2 2
```

```
library(dplyr)
players_clustered <- mutate(players, cluster =
                            cluster_assignments)
players_clustered
```

```
##      X  Y cluster
## 1   -1  1       1
## 2    2 -3       1
## 3    8  6       1
## 4    7 -8       1
## 5  -12  8       2
## 6  -15  0       2
```

Visualizing K-Clusters (ggplot)

```
library(ggplot2)
ggplot(players_clustered, aes(x = X, y = Y, color =
                                  factor(cluster), size = 6)) +
geom_point()
```



Building the Dendogram

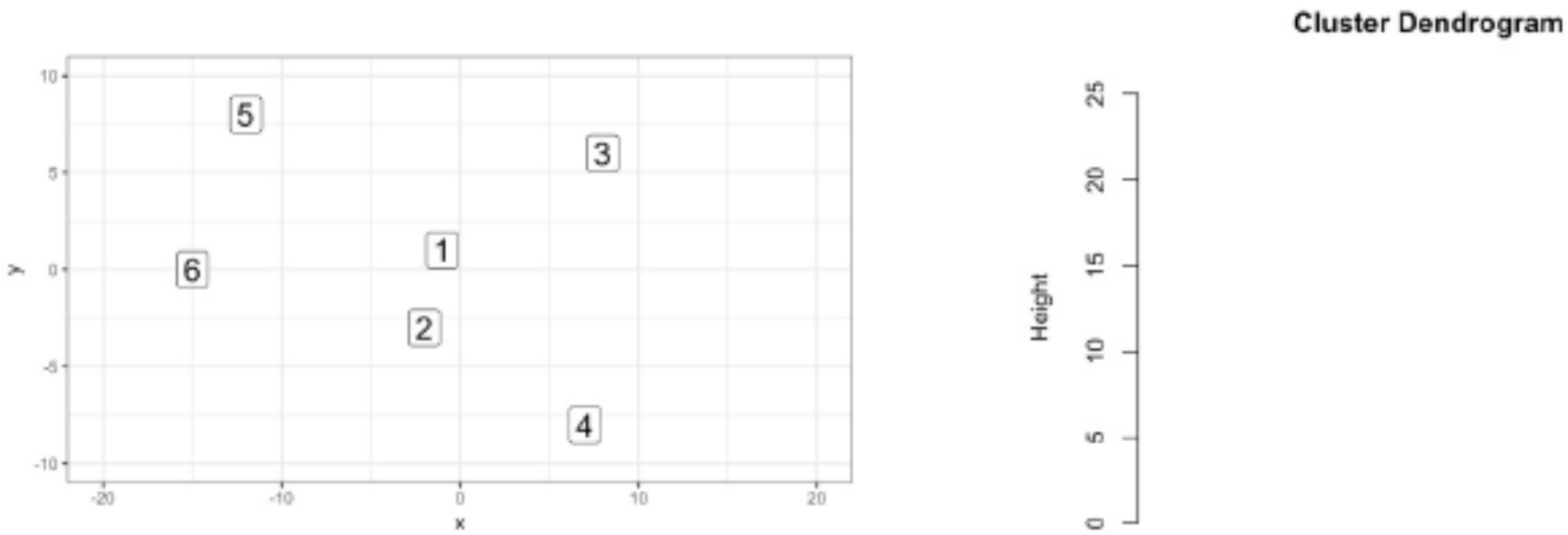


Figure 21:

Building the Dendogram

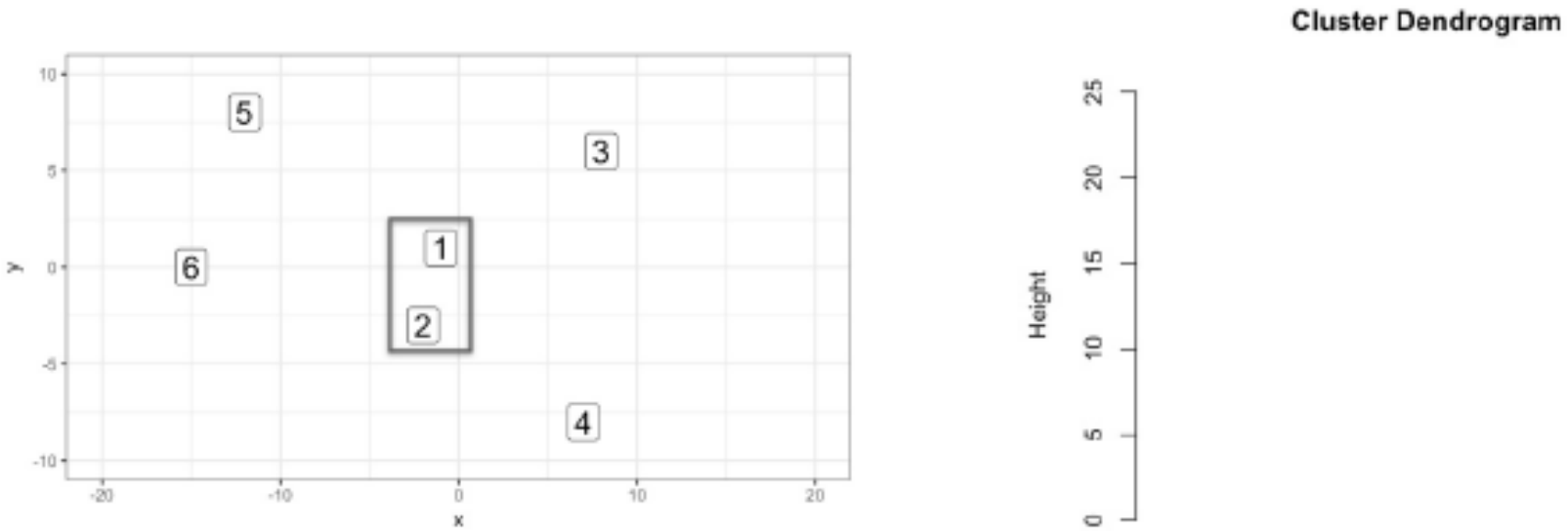


Figure 22:

Building the Dendogram

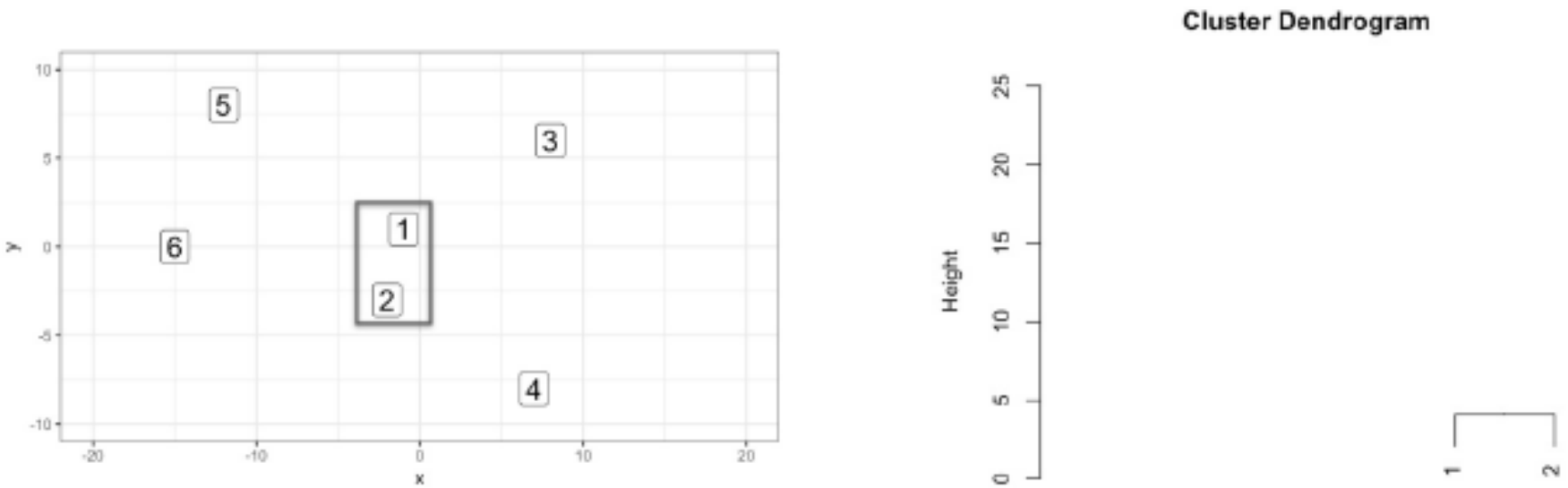


Figure 23:

Building the Dendogram

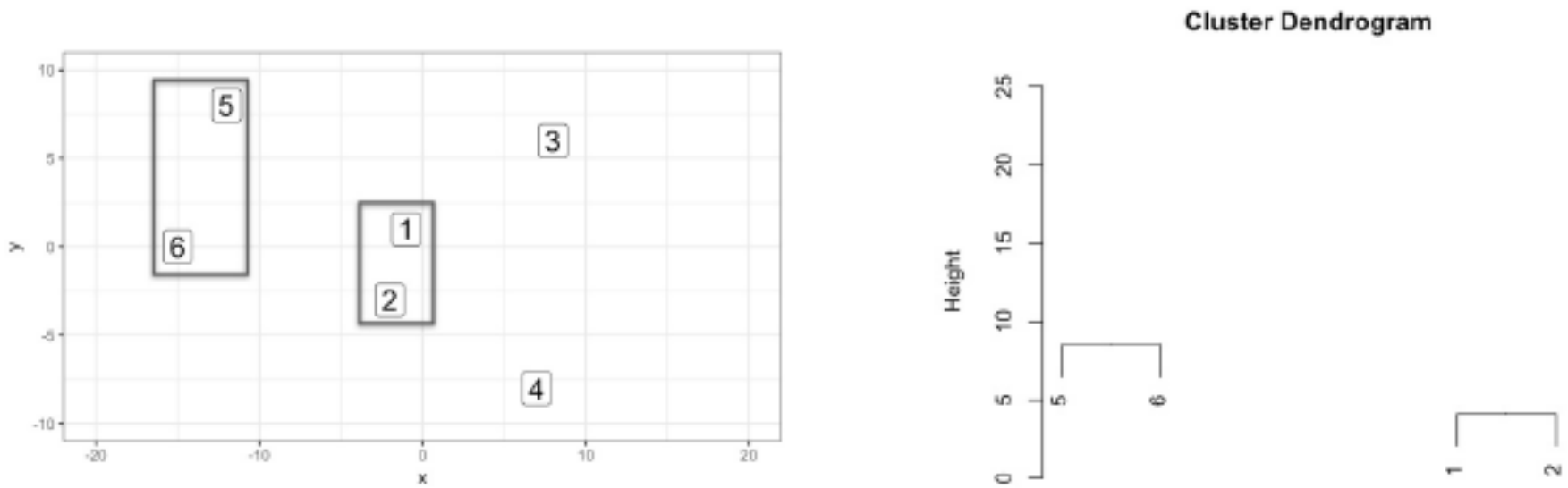


Figure 24:

Building the Dendogram

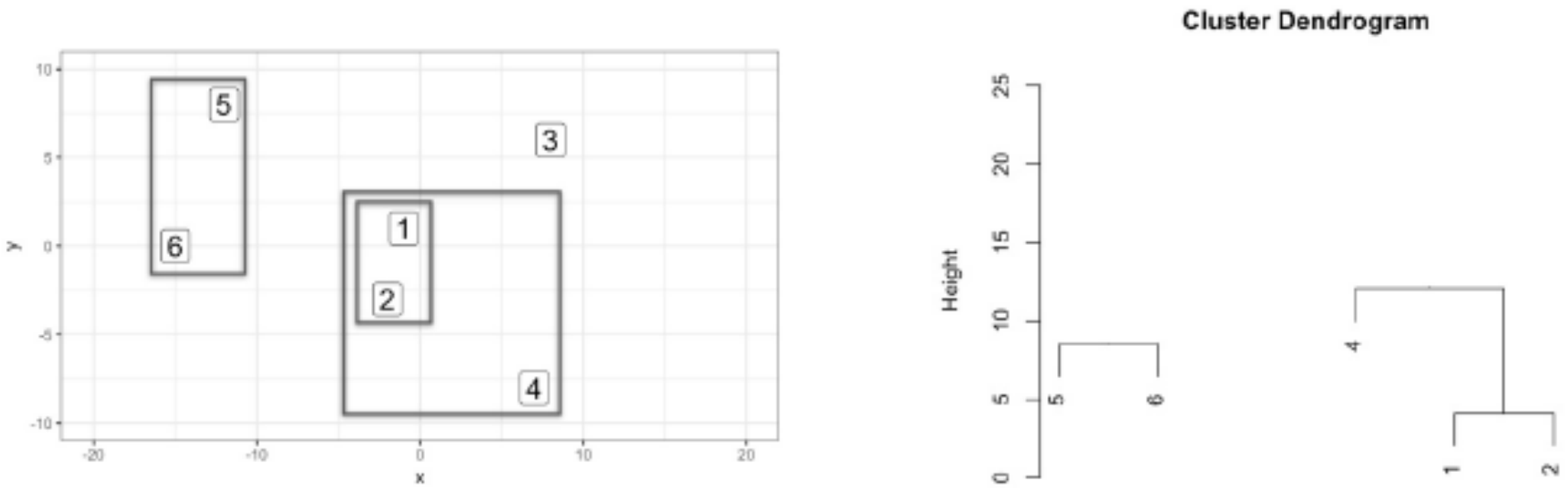
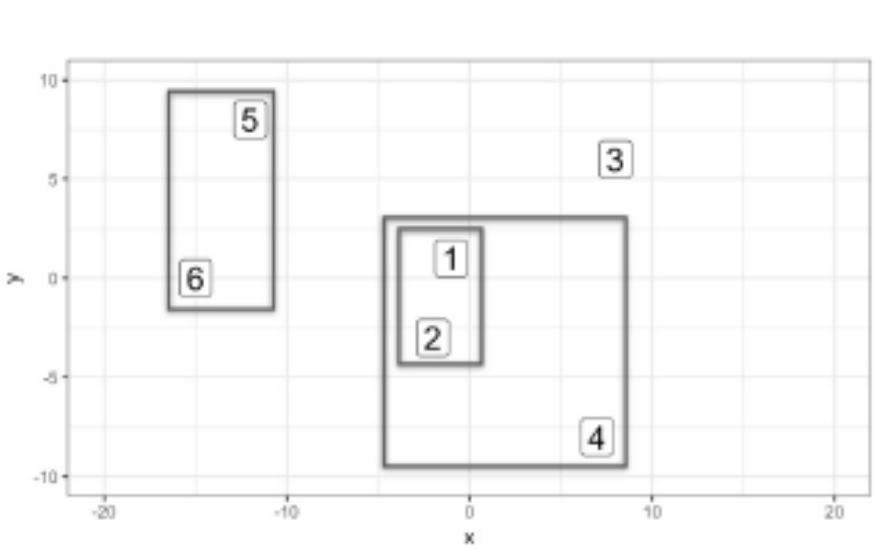


Figure 25:

Building the Dendogram



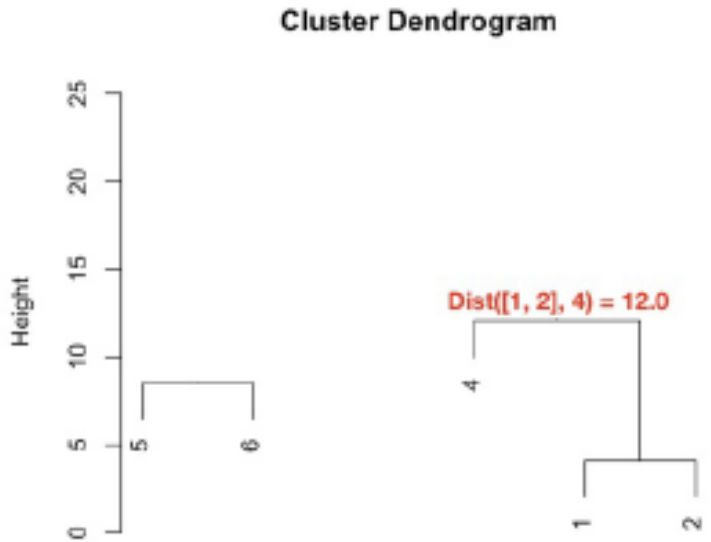


Figure 26:

Building the Dendogram

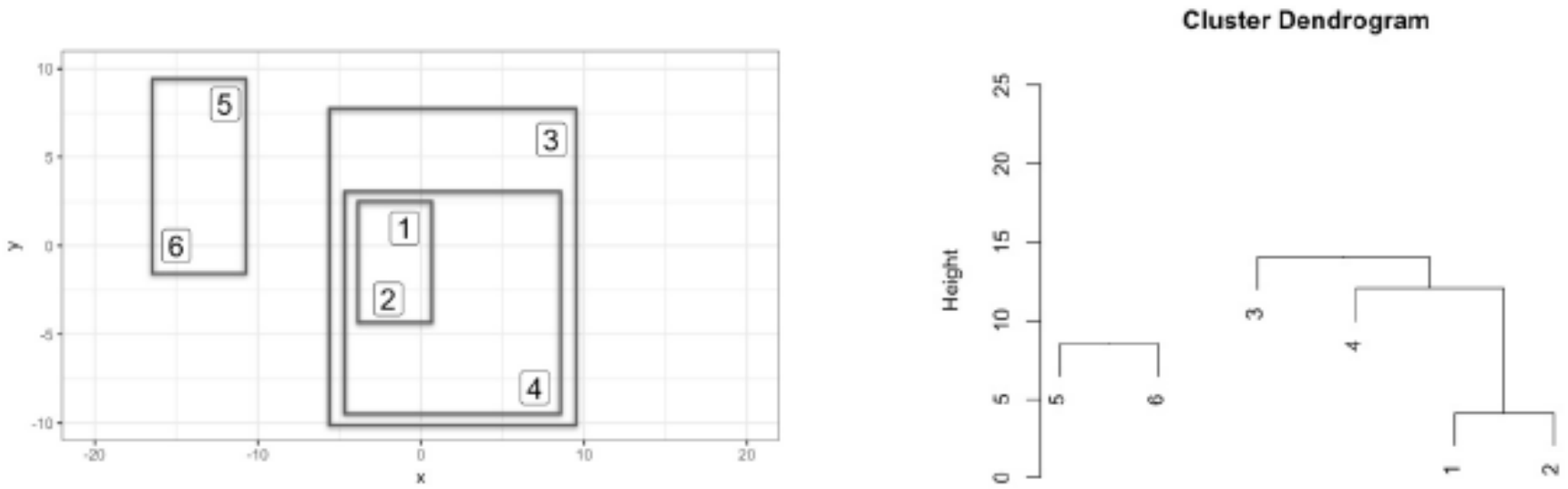
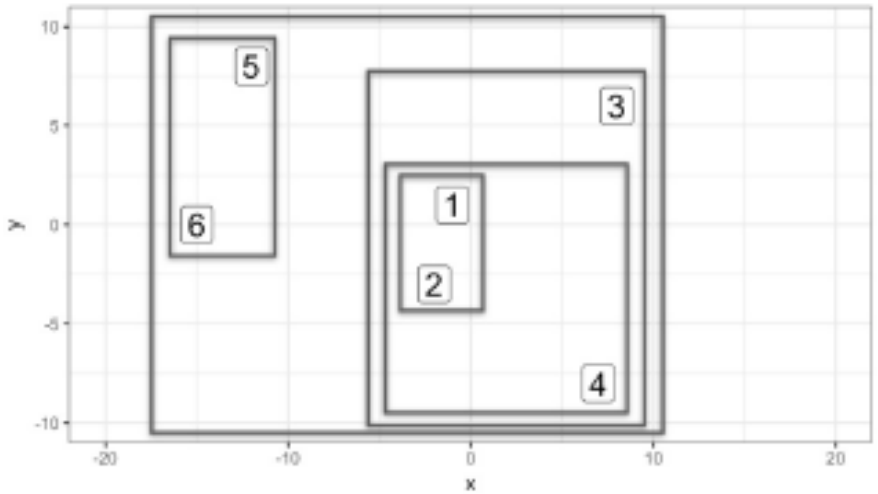


Figure 27:

Building the Dendogram



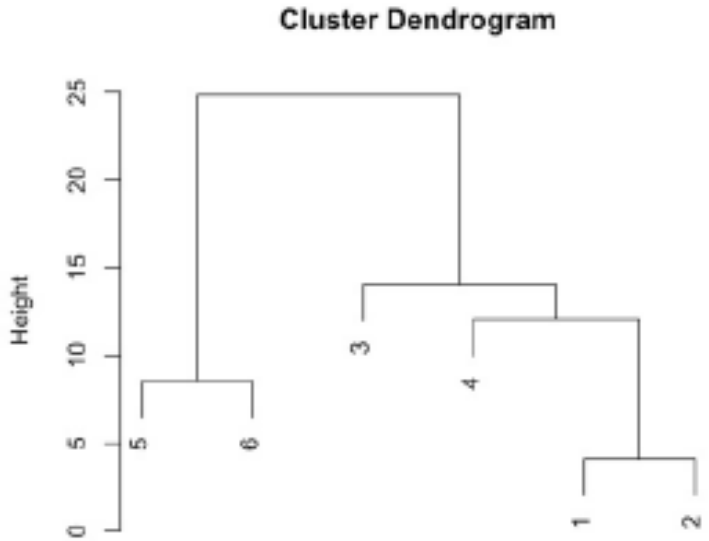
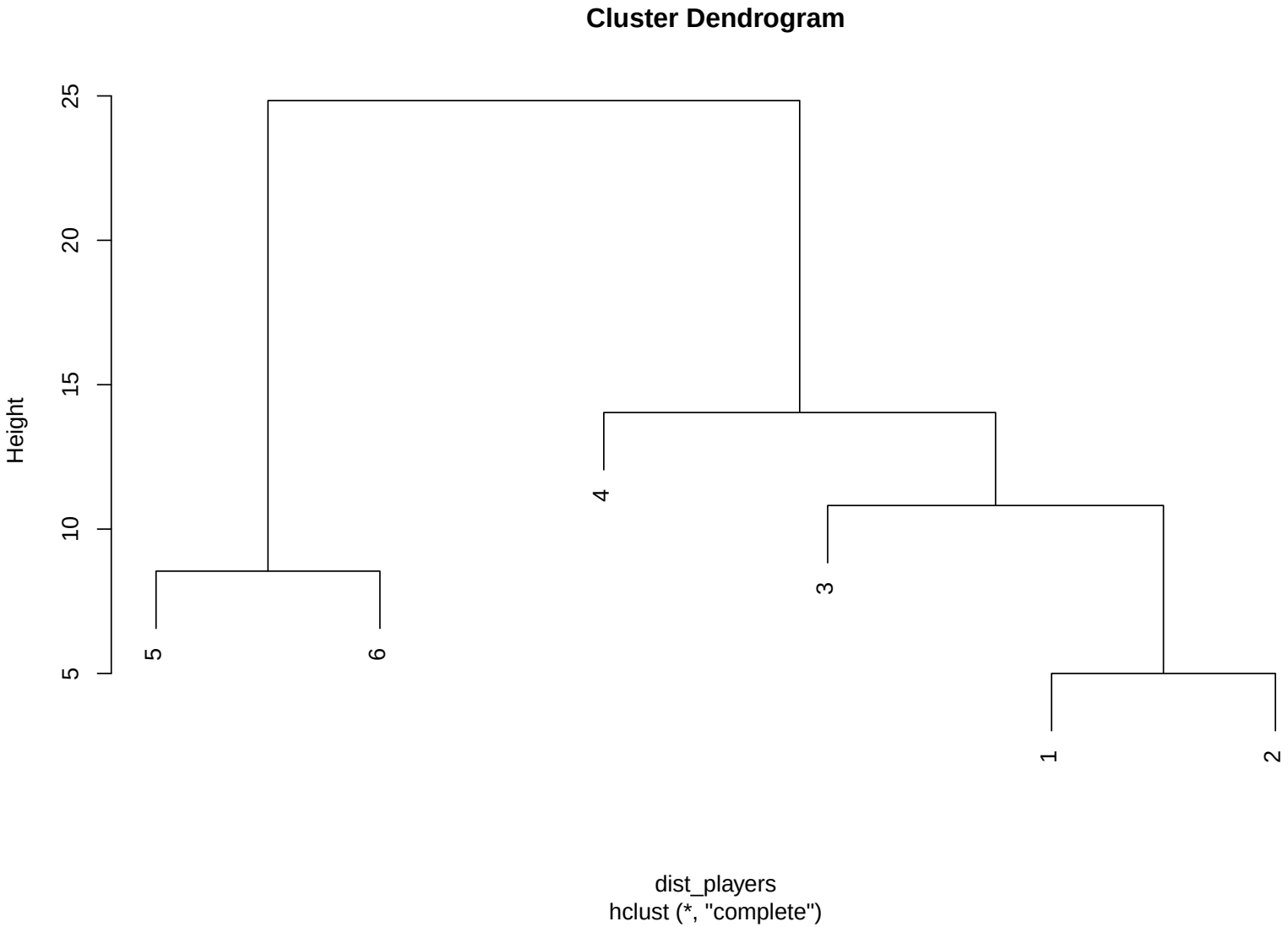


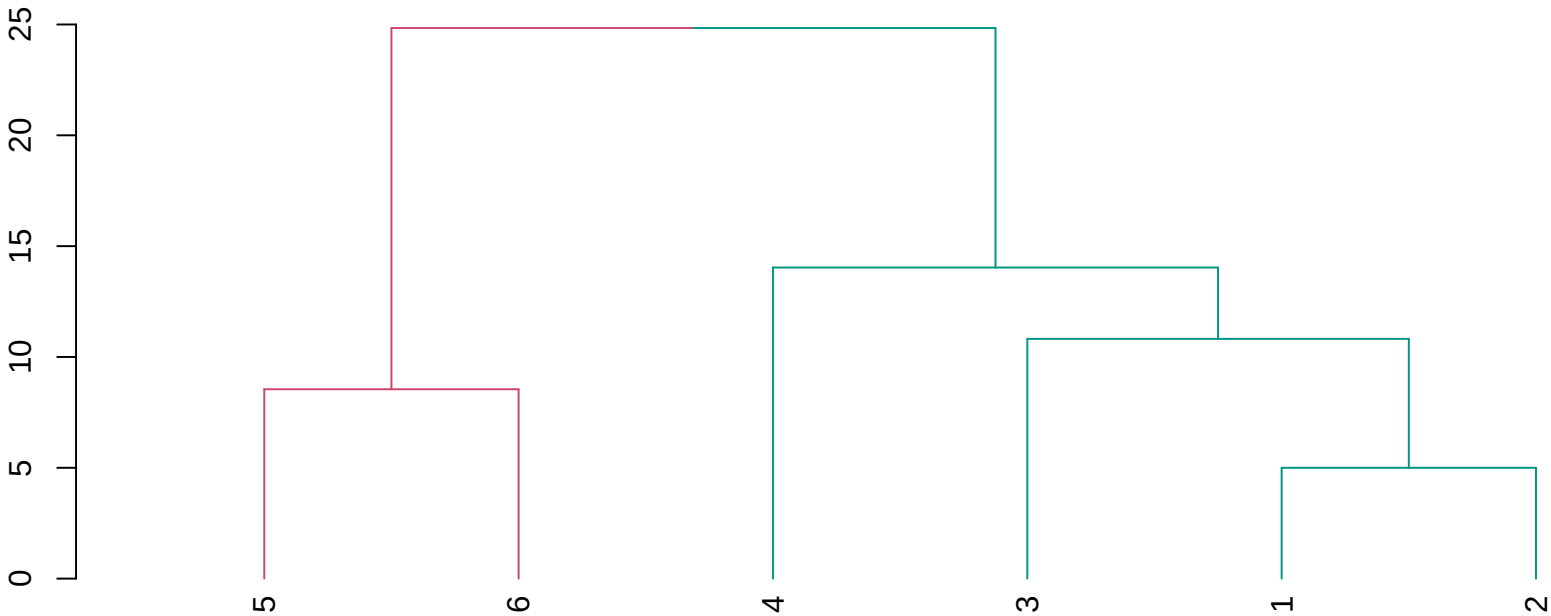
Figure 28:

Plotting the Dendogram



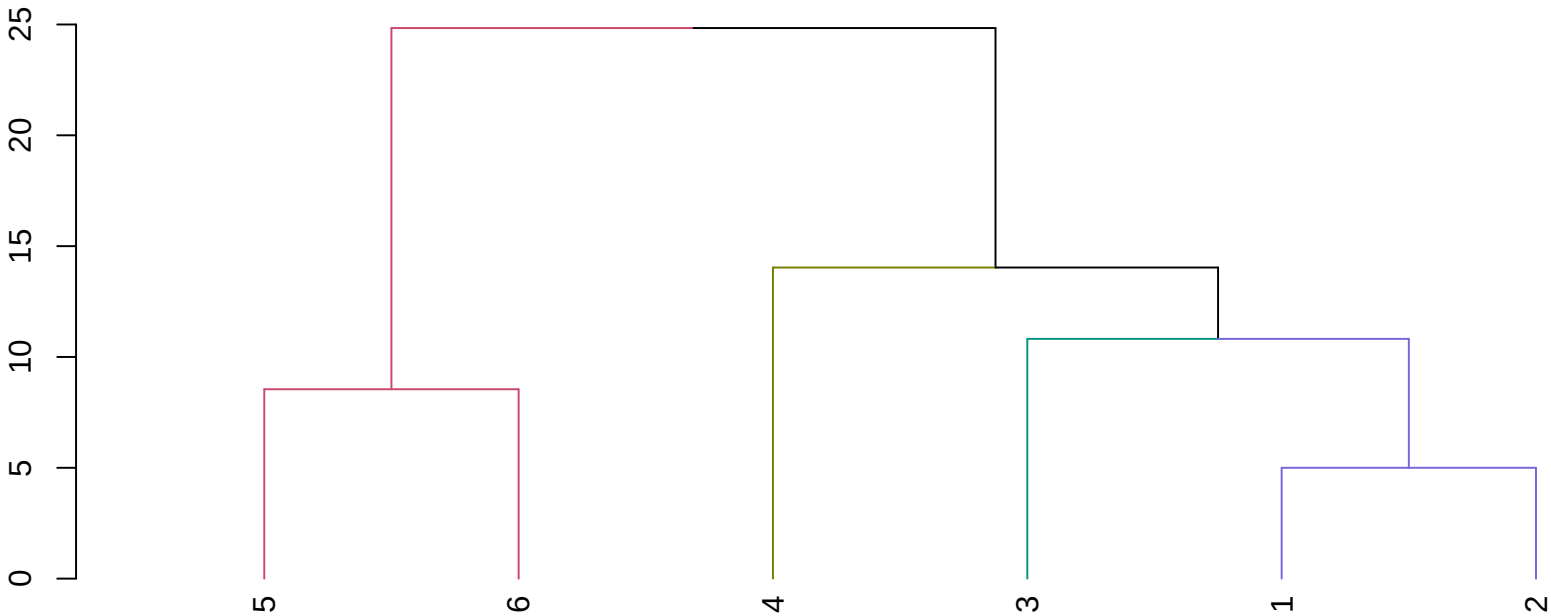
A better looking Dendogram

```
library(dendextend)
dend_players <- as.dendrogram(hc_players)
dend_colored <- color_branches(dend_players, h = 15)
plot(dend_colored)
```



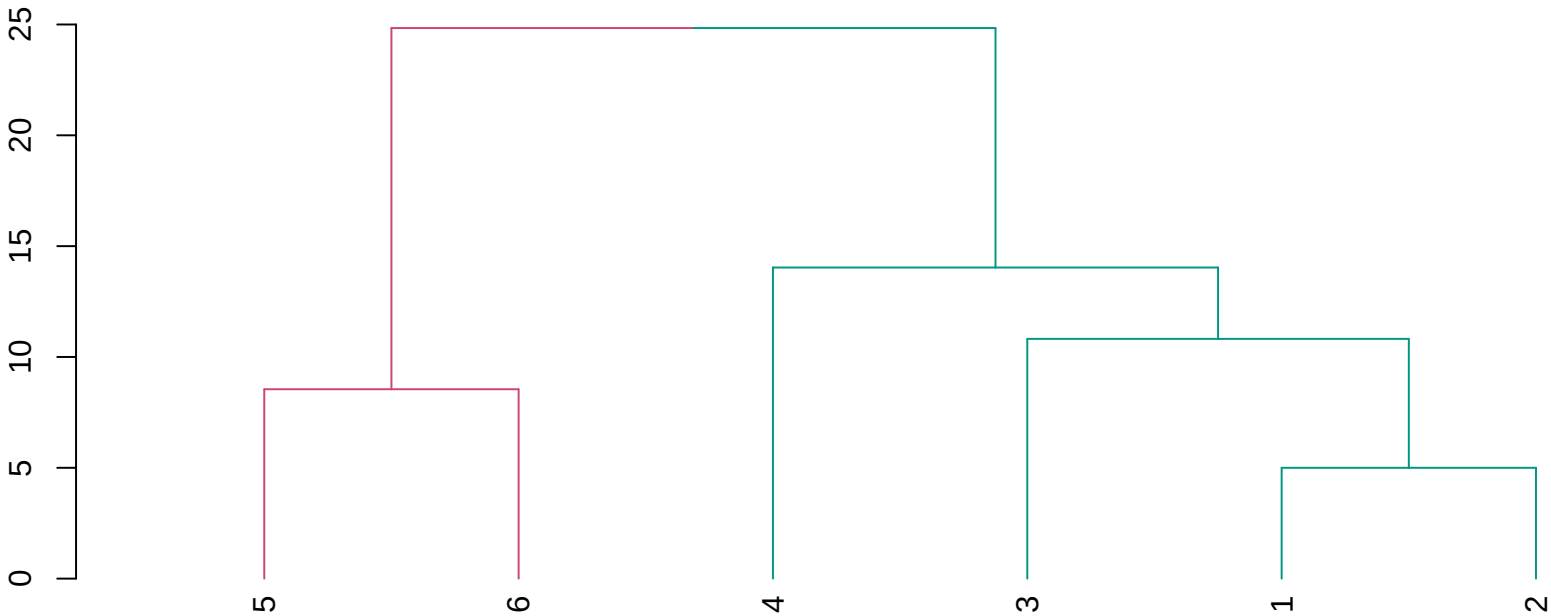
A better looking Dendogram (cut = 10)

```
library(dendextend)
dend_players <- as.dendrogram(hc_players)
dend_colored <- color_branches(dend_players, h = 10)
plot(dend_colored)
```



A better looking Dendogram (k = 2)

```
library(dendextend)
dend_players <- as.dendrogram(hc_players)
dend_colored <- color_branches(dend_players, k=2)
plot(dend_colored)
```



cutree using height

```
cluster_assignments <- cutree(hc_players, h = 15)
print(cluster_assignments)
```

```
## [1] 1 1 1 1 2 2
```

```
players_clustered <- mutate(players, cluster =
                            cluster_assignments)
```