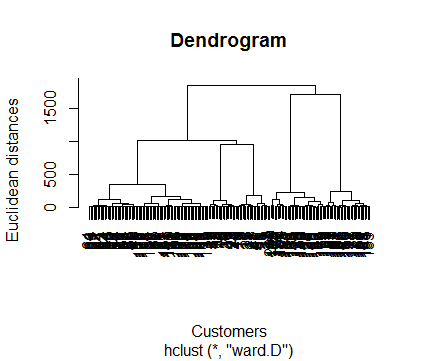
**Business problem**

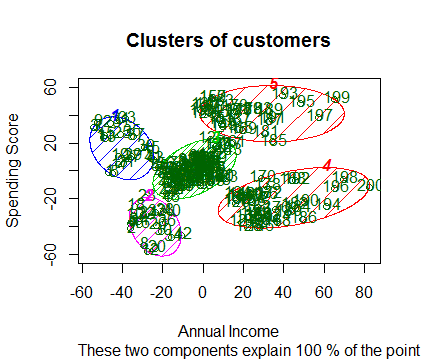
Create a model that groups the customers of a specific mall by their annual income and their spending on that mall.

**Explaining the model**

1. Import the dataset, which is in a csv format.
2. We use the columns of interest, which are the annual income and spending.
3. After that, we need to know the optimal number of clusters. To do so, we use the function hclust, which takes a first argument called dist (takes our dataset and the ‘euclidean’ method), the second argument is the “ward.D” method and then we plot the graph.
4. To find the number of clusters, we need to find the largest vertical distance we can make without crossing any horizontal lines and we count the number of vertical lines we can find at that level.



**Plotting the results**



As we can see, we got our 5 clusters. Cluster number 2 is the cluster with the customers with low income and low spending. Cluster number 3 is the cluster with the customers with average income and average spending. Cluster number 1 is the cluster with the customers with low income and high spending. Cluster number 4 is the cluster with the customers with high income and low spending. Cluster number 5 is the cluster with the customers with high income and high spending.