Clustering in Heart Disease Data

Important things that I have learnt in this project are:-

1. Scale()- Use this function to scale data or standardize data(z-scores). More modifications can also be made, for eg Sample Variance for each observation can also be computed using this function
2. Before using a clustering, it is crucial to decide if data should be scaled or not. If two different variables are in different units then data must be checked carefully.
3. While using kmeans algorithm, we can start by using only 1 iteration and run it twice. If in two different iteration there is a significant difference in cluster sizes then we can infer that kmeans algorithm is not clustering the data appropriately
4. We can assign clusters to data using cbind() and then plot the data to see how data is clustered.
5. To see summary statistic for observations in each cluster made using hierarchical clustering algorithm, we used do.call() which is similar to lapply(). This function created a dataframe having mean and sd for each of the variables grouped by cluster.