CMPT 459 Programming Assignment 1 Report
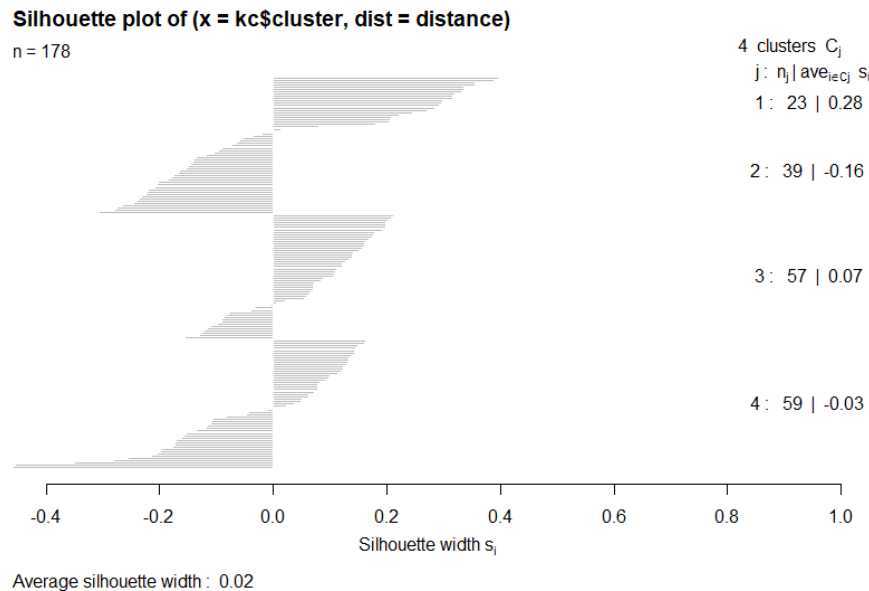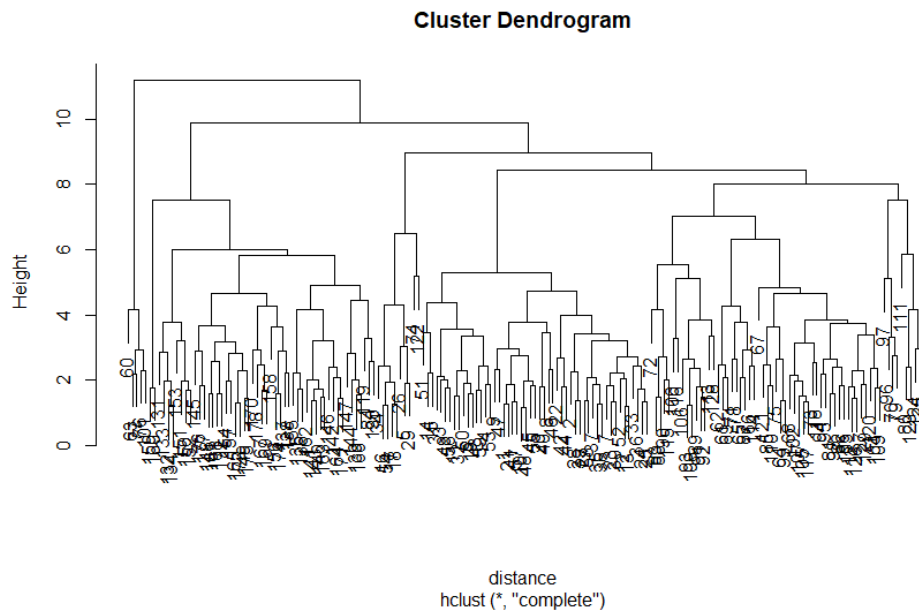
Predefined function *apply* were used to find the mean and standard deviation in R, where 2 was used in the parameter which corresponds to columns. Normalization was done using another predefined function *scale*. This can be seen in the source code. After performing K-means clustering with k = 4, the average silhouette score is 0.02 which can be seen in the plot below.

Silhouette plot of (x = kc$cluster, dist = distance)

n = 178

4 clusters C_j

$j : n_j | ave_{i \in Cj} \; s_i$

1 : 23 | 0.28

2 : 39 | -0.16

3 : 57 | 0.07

4 : 59 | -0.03
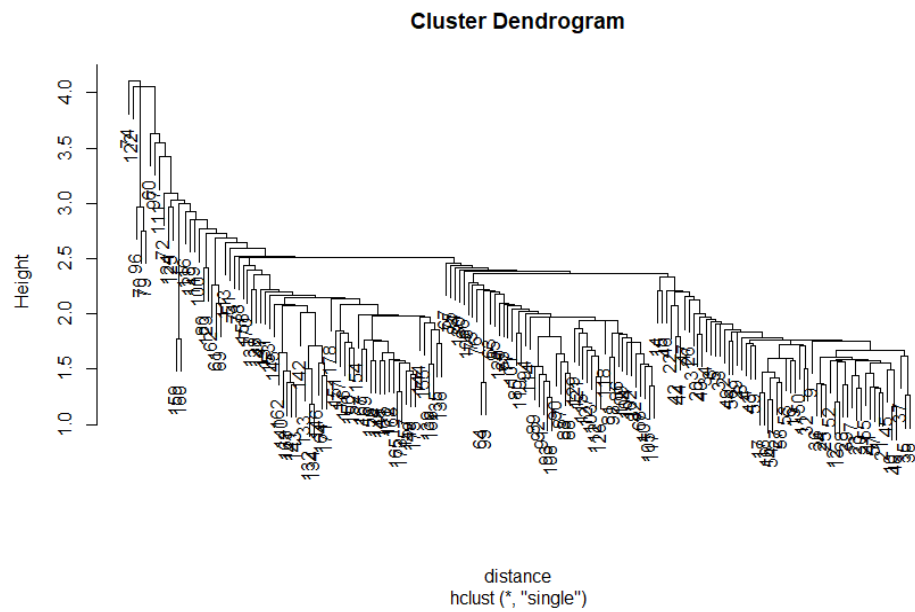
Silhouette width s_i

Average silhouette width : 0.02

The different types of linkage leads to the following cluster dendrograms below. Clusters/shape of the dendrograms are similar with increasing amount of linkages from complete to single, respectively.

Cluster Dendrogram with Complete Linkage

Cluster Dendrogram

Height

distance
hclust (*, "complete")

Cluster Dendrogram with Average Linkage

Cluster Dendrogram

distance
hclust (*, "single")

Cluster Dendrogram with Single Linkage

Cluster Dendrogram

distance
hclust (*, "average")