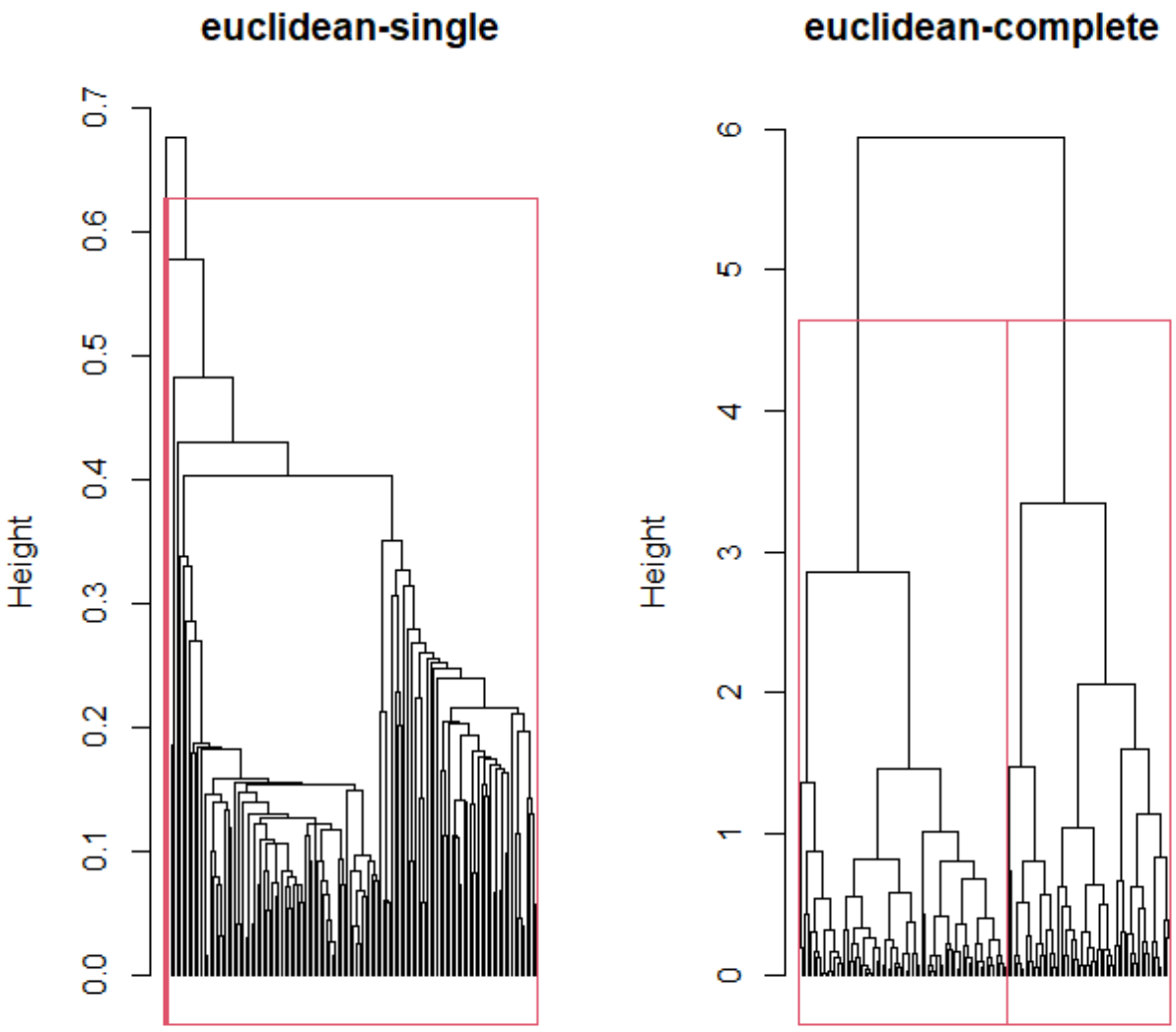
APPLIED STATISTICS EXAM **18/01/2023**

MANUEL PERACCI 10824742

EXERCISE 2

A)



After scaling the variables and permute the rows, I performed a cluster analysis of the beers by using a hierarchical clustering method (Euclidean distance) with single linkage and complete linkage.

The two dendrograms are very different:

The one on the left comes from the single linkage and we can clearly see that it is affected by the chain-effect. In this case I chose two clusters, but also three were fine. While the right one comes from the complete linkage and I chose two clusters.

B)

Between the two dendrograms I would choose definitely the second one (complete linkage), because the single linkage is affected by the chain effect. So, the two groups are not correctly separated.

With the complete linkage, I would use two clusters because looking at the height of dendrogram, they seem quite appropriate.

C)

The centroids are:

M_clus_1

alcohol ibu

5.901857 40.059429

M_clus_2

alcohol ibu

9.415273 100.624545

The sizes are: 70 for the first cluster and 55 for the second one.

The cophenetic coefficient is 0.7919285 for the complete linkage.