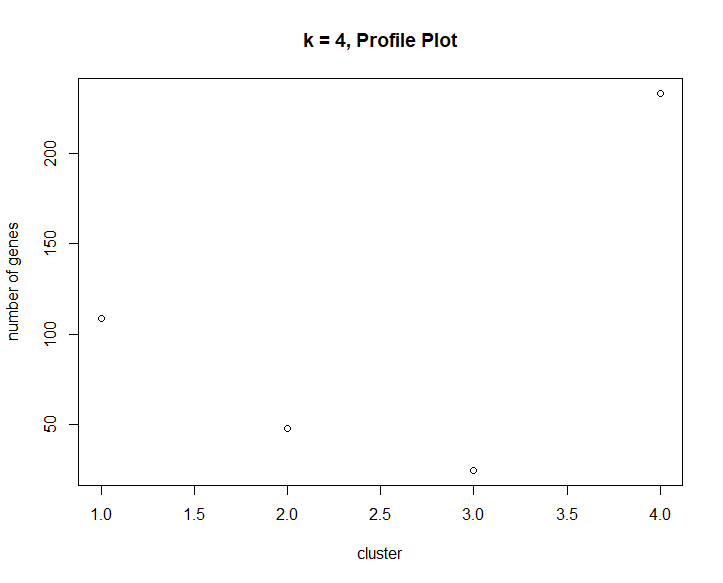


2 b)

According to the silhouette plots in the zip file and the conclusion above, the most appropriate k value should be 16 because the average silhouette width of the entirety is greatest when k = 16, which is quite different from the judgement made in Task 3.

3 c)



3 d)

Because, in this case, the distance comparation between every single data is no longer necessary. By contrast, with a proper k value, the algorithm of renewing the least intravariance sum will adjust the position of the cluster centroids until they do not change. The times needed for calculation are significantly reduced.