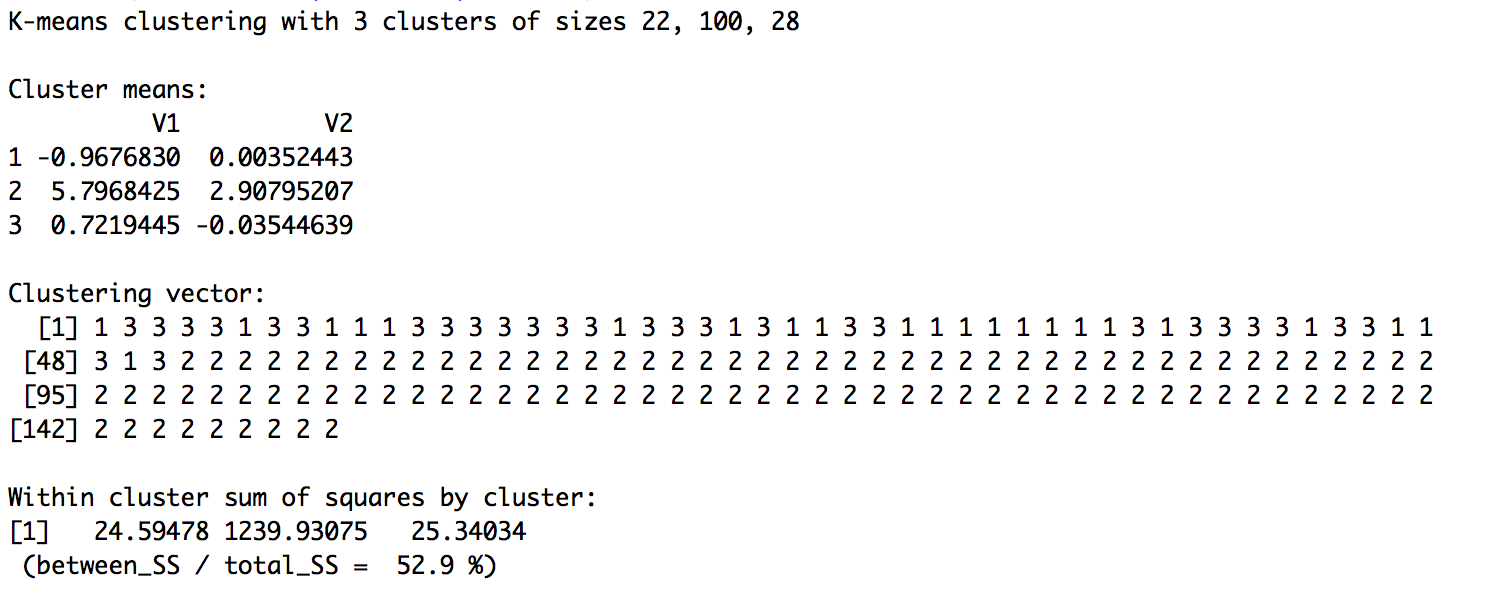
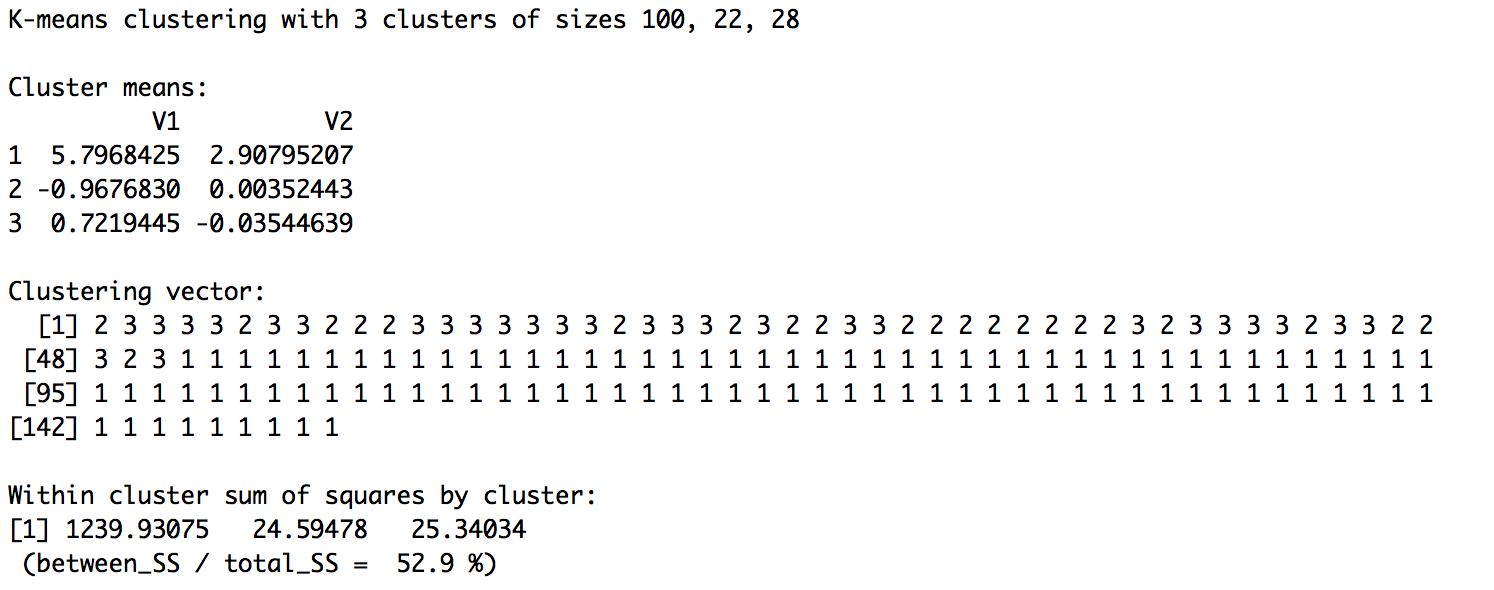
Q1: For nstart being 1

*>R kmeans(clusterdata1, centers=3, nstart=1)*

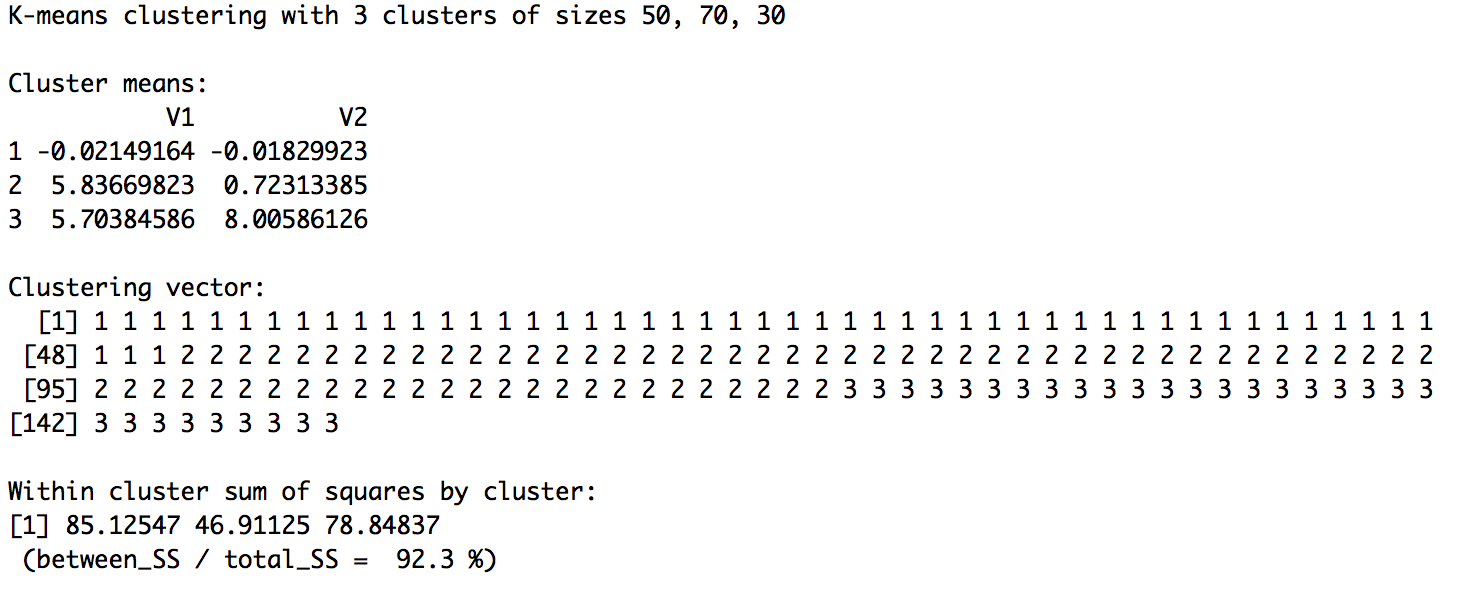
**1st:**



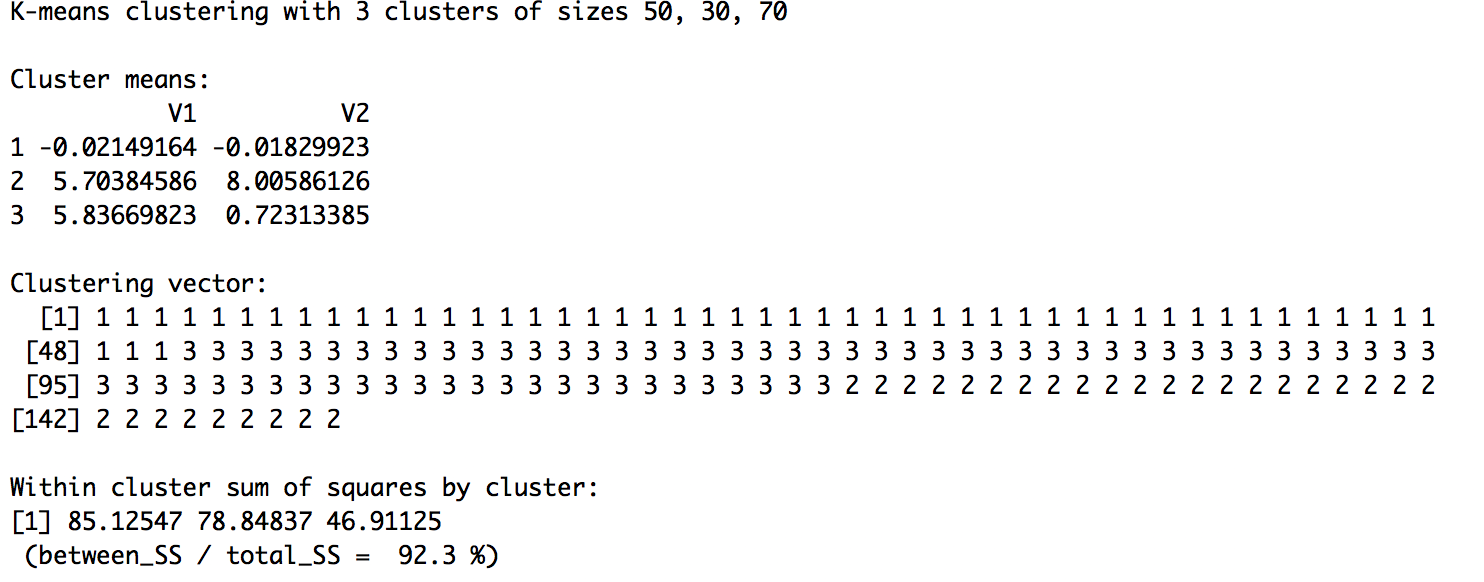
**2nd:**



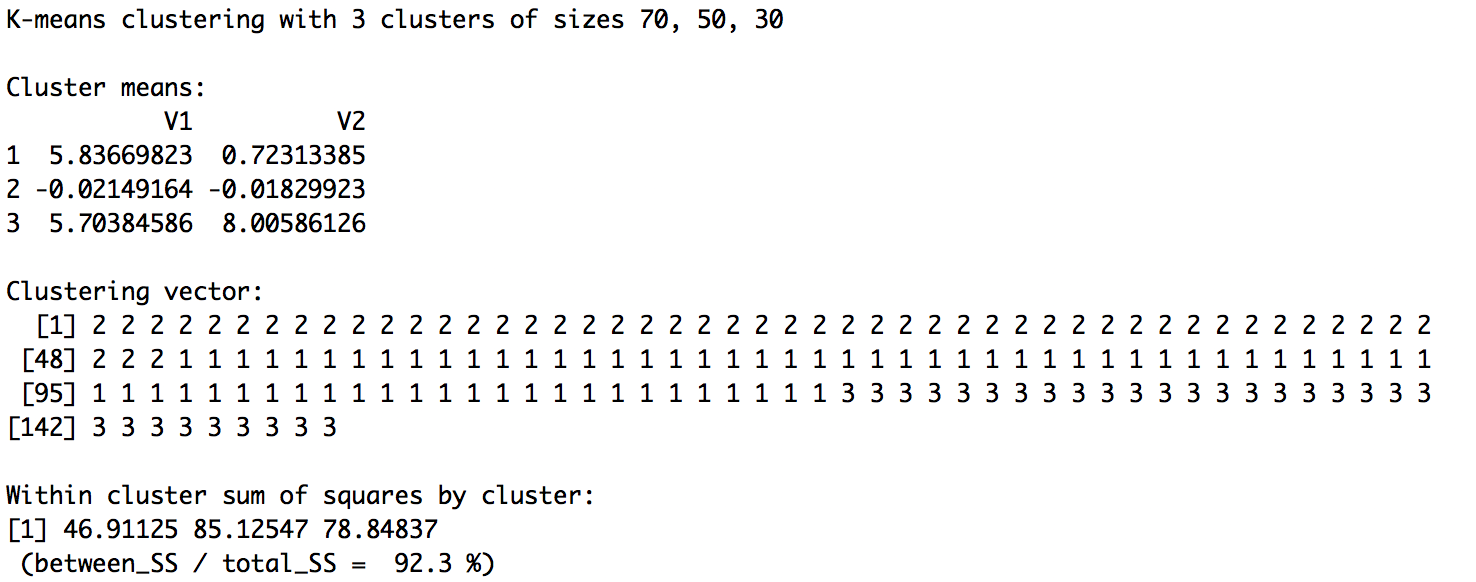
**3rd:**



**4th:**



**5th:**



For nstart = 1:

The results for ‘K-means clustering with 3 clusters of sizes’ & ‘within cluster sum of squares by cluster’ and ‘between\_SS / total\_SS’ have changed over times. This may be due to the changed of the centroids. So that the stability builds up as we runs more times of trying.

For nstart=100:

Within cluster sum of squares are unchanged for 10 times and the between\_SS / total\_SS remain stable at 92.3%

Q2:

> R:

k3<-kmeans(olive, centers = 3, nstart = 100)

plot(olive, col=k3$cluster,pch=clusym[k3$cluster])

### After scale

solive<- scale(olive)

pairs(solive, cex=0.3)

sk3<- kmeans(solive, 3, 100)

plot(olive, col=sk3$cluster,pch=clusym[sk3$cluster])

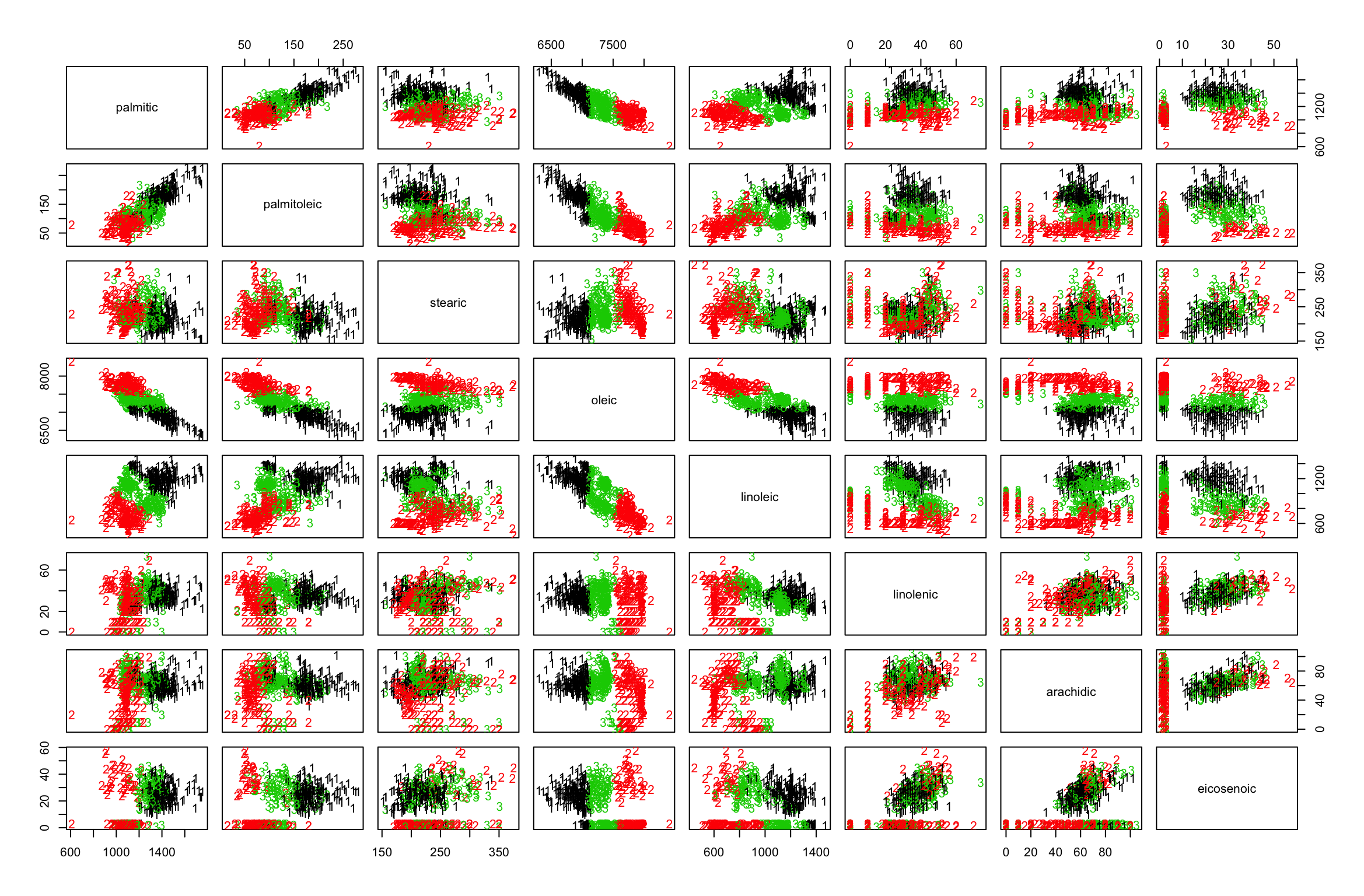
table(sk3$cluster, oliveoil$macro.area)

sk9 <- kmeans(solive, 9, 100)

plot(solive, col=sk9$cluster,pch=clusym[sk9$cluster])

table(sk9$cluster, oliveoil$macro.area)

With K=3: before scale



After scale:

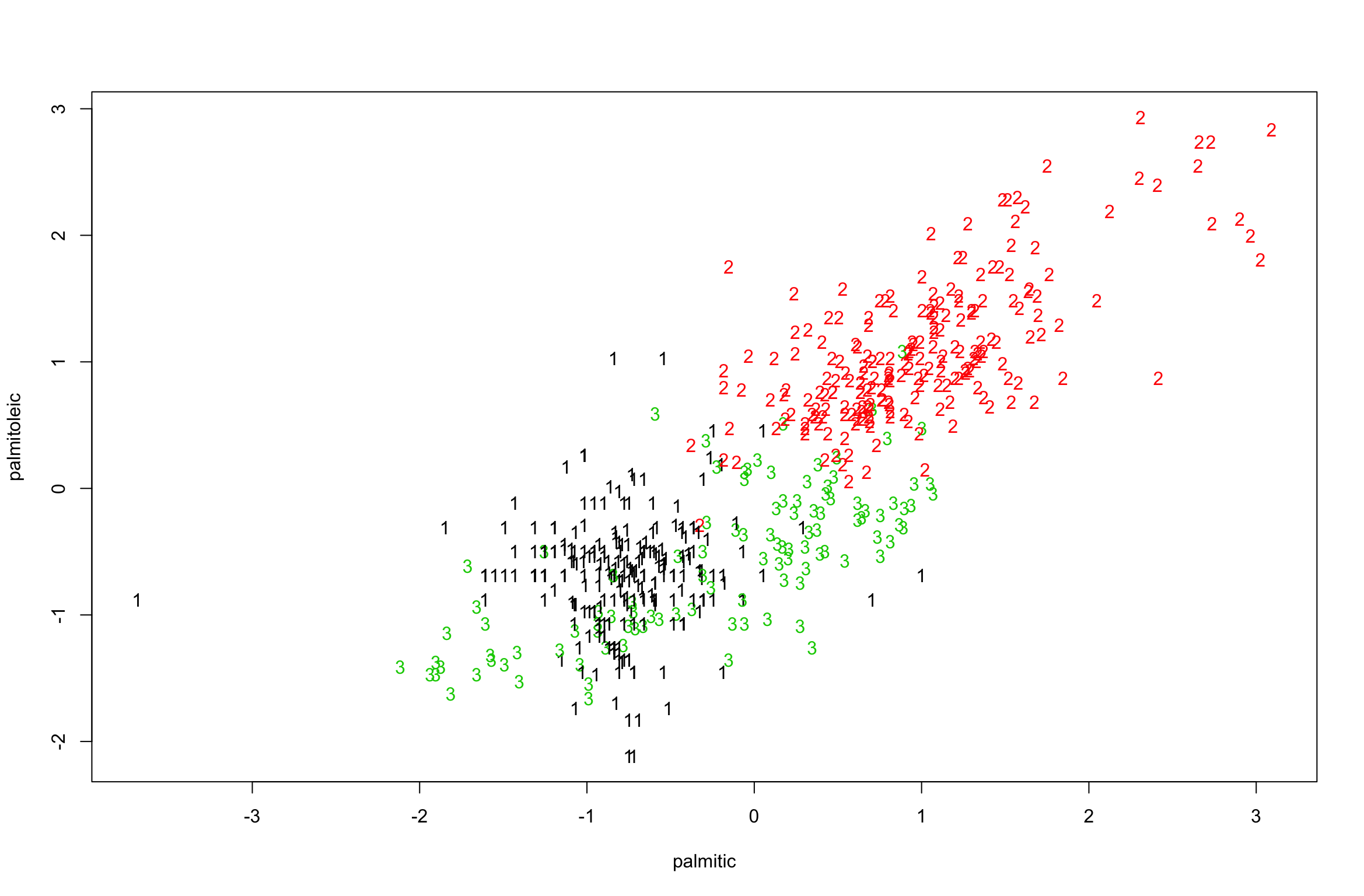
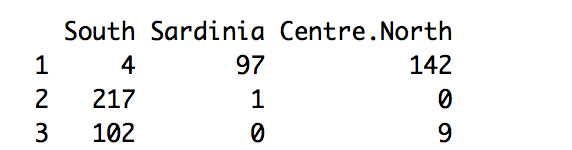


table:



Looks not bad…

I don’t see quiet much how the criteria of how well this clustering is … I could see that South has been allocated to two major parts of clusters 2 and 3, and Centre.North and Sardinia are allocated to 1.

k=9

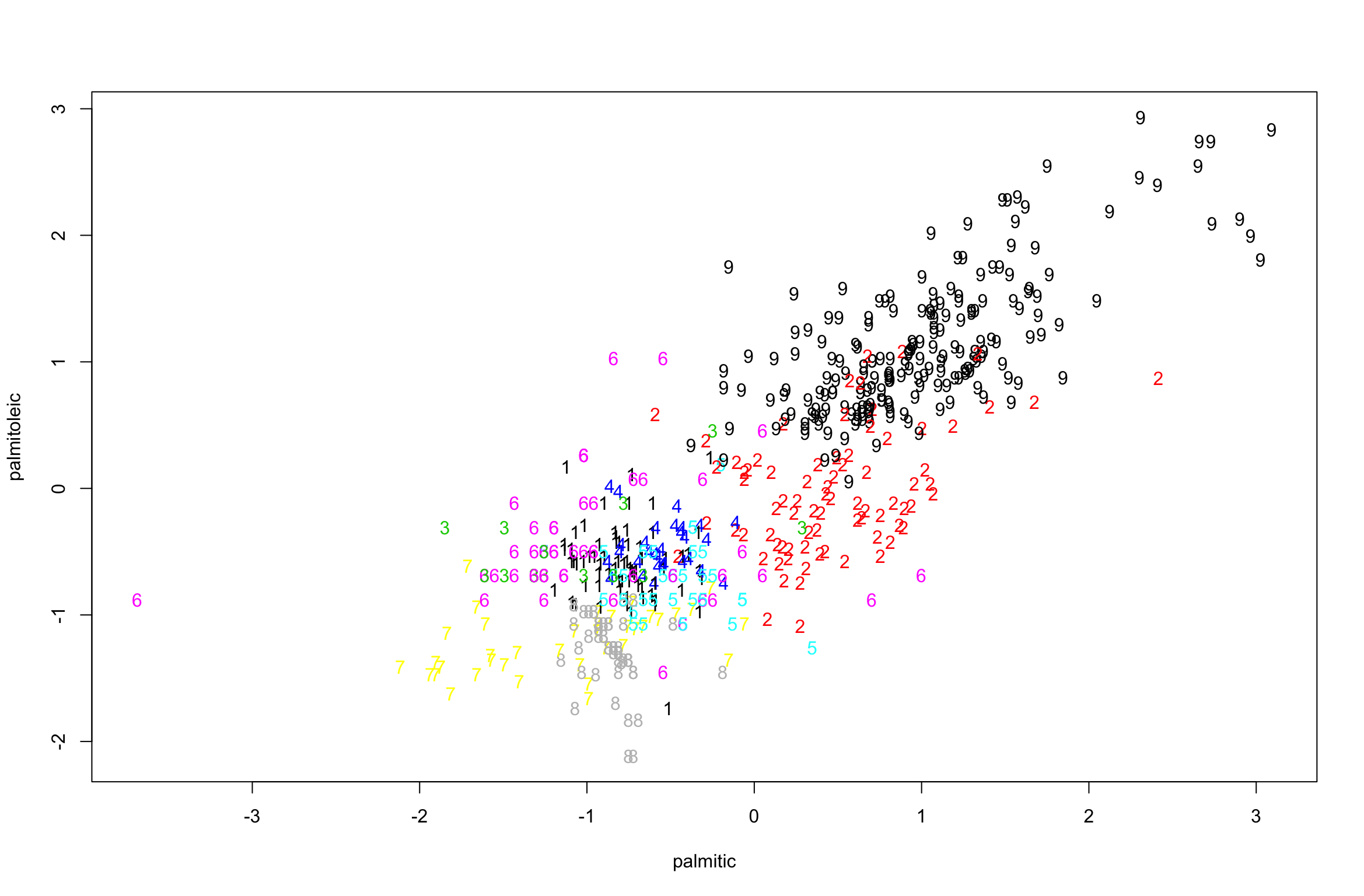
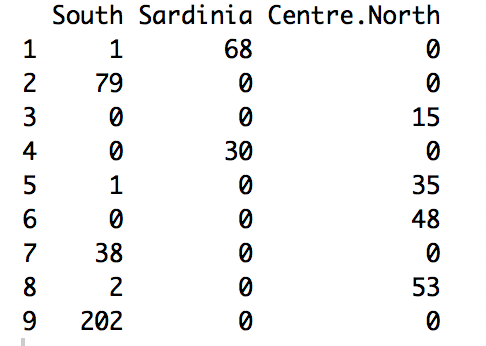


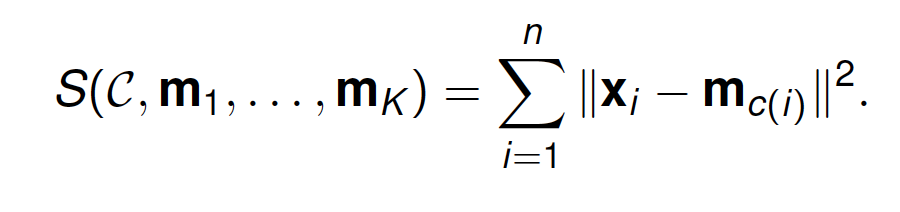
table:



I think this is just an expand of K=3. But at least this trial makes Sardinia and Centre.North in different clusters. So that they are separated when k=9.

Q3:

Multiplying the variables by the same constant q won’t change the variation of the data. Therefore, the Euclidean distances between each xi and the centroids mi…kkm will not change. Thus, the K-means clustering of D which is defined as choosing the ms and cs to minimise:



will not change.

However, multiplying the variables by a same constant does not affect the variation but the value of the data. The new centroids are affected and are multiplied by the q.