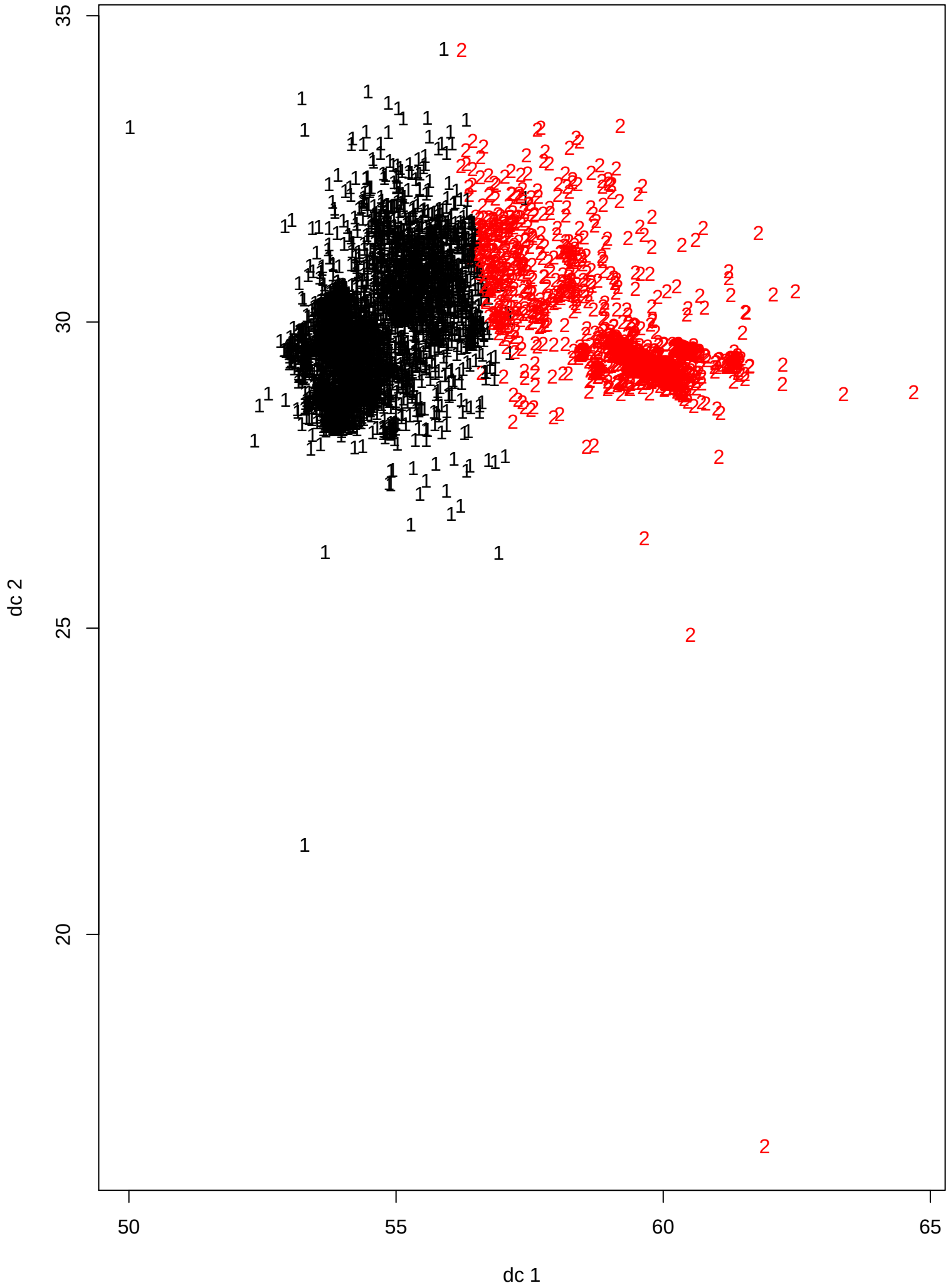
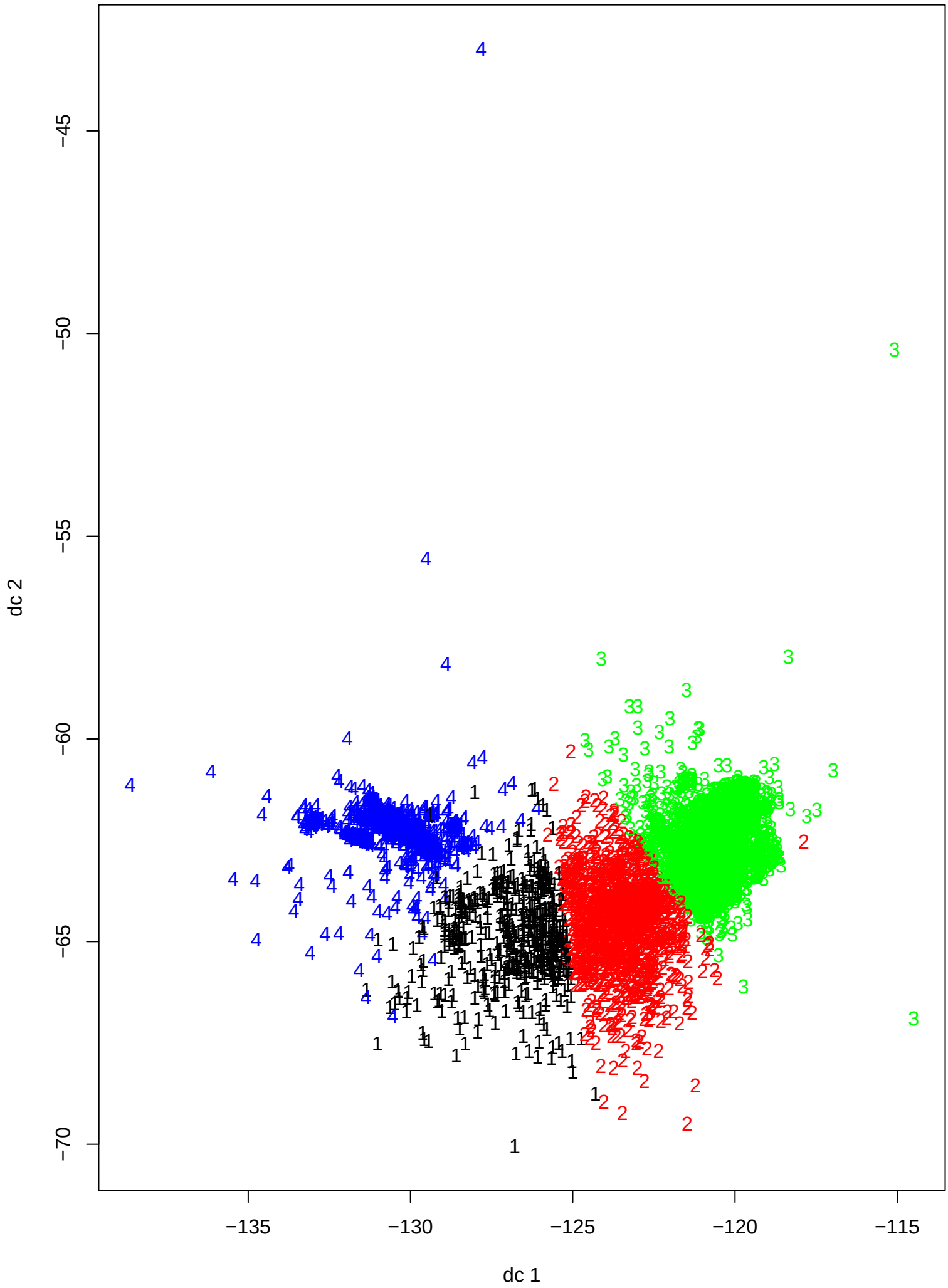
```
Exercise 4
Philipp Schlieker - RA191996


#--------------RESULT--------------#
[1] "Optimal Number of Clusters using Silhouette  2"
[1] "Optimal Number of Clusters using Normalized/Adjusted Rand  4"

#---------------CODE----------------#
library("fpc")
library("cluster")

inputdata <- read.csv("http://www.ic.unicamp.br/~wainer/cursos/2s2016/ml/cluster-data.csv")
classes <- read.csv("http://www.ic.unicamp.br/~wainer/cursos/2s2016/ml/cluster-data-class.csv")

d <- dist(inputdata)

#Storing information for Internal Eval
silhouette <- 0
kint <- 0

#Storing Information for External Eval
rand <- -1
kext <- 0

for(k in 2:10){
  clust <- kmeans(x = inputdata, centers = k, nstart = 5)

  #Evaluate Clusters using Silhouette
  eval <- cluster.stats(d = d, clust$cluster, silhouette = TRUE)
  if(eval$avg.silwidth > silhouette){
    silhouette <- eval$avg.silwidth
    kint <- k
  }

  #Evaluate Clusters using Normalized/Adjusted Rand
  eval <- cluster.stats(d = d, clustering = clust$cluster, alt.clustering=classes$x,
                        compareonly = TRUE)
  if(eval$corrected.rand > rand){
    rand <- eval$corrected.rand
    kext <- k
  }
}
print(paste("Optimal Number of Clusters using Silhouette ", kint))
print(paste("Optimal Number of Clusters using Normalized/Adjusted Rand ", kext))

#Create Plots
clust <- kmeans(x = inputdata, centers = 2, nstart = 5)
plotcluster(inputdata, clust$cluster)
clust <- kmeans(x = inputdata, centers = 4, nstart = 5)
plotcluster(inputdata, clust$cluster)
```


35
30
25
20
dc 2
50
55
60
65
dc 1


dc 2
−45
−50
−55
−60
−65
−70
−135
−130
−125
−120
−115
dc 1