library(cluster) #conduct cluster analysis

library(compareGroups) #build descriptive statistic tables

library(HDclassif) #contains the dataset

library(NbClust) #cluster validity measures

library(sparcl) #colored dendrogram

data(wine)

str(wine)

names(wine) = c("Class","Alcohol","MalicAcid","Ash","Alk\_ash","magnesium","T\_phenols","Flavanoids","Non\_flav","Proantho","C\_Intensity","Hue","OD280\_315","Proline")

names(wine)

df = as.data.frame(scale(wine[,-1]))

str(df)

table(wine$Class)

numComplete = NbClust(df, distance="euclidean", min.nc=2, max.nc=6, method="complete", index="all")

numComplete$Best.nc

dis = dist(df, method="euclidean")

hc = hclust(dis, method="complete")

plot(hc, hang=-1,labels=FALSE, main="Complete-Linkage")

comp3 = cutree(hc, 3)

ColorDendrogram(hc, y = comp3, main="Complete", branchlength=50)

table(comp3)

table(comp3,wine$Class)

(51+50+48)/178

NbClust(df, diss=NULL, distance="euclidean", min.nc=2, max.nc=6, method="ward.D2", index="all")

hcWard = hclust(dis, method="ward.D2")

plot(hcWard, hang=-1, labels=FALSE, main="Ward's-Linkage")

ward3 = cutree(hcWard, 3)

table(ward3,wine$Class)

table(comp3, ward3)

aggregate(wine[,-1],list(comp3),mean)

aggregate(wine[,-1],list(ward3),mean)

par(mfrow=c(1,2))

wine$comp\_cluster = comp3

wine$ward\_cluster = ward3

boxplot(Proline~comp\_cluster, data=wine, main="Proline by Complete Linkage")

boxplot(Proline~ward\_cluster, data=wine, main="Proline by Ward's Linkage")

NbClust(df, min.nc=2, max.nc=15, method="kmeans")

set.seed(1234)

km=kmeans(df,3,nstart=25)

table(km$cluster)

#km$tot.withinss

km$centers

wine$km\_cluster = km$cluster

boxplot(Alcohol~km\_cluster, data=wine, main="Alcohol Content, K-Means")

boxplot(Alcohol~ward\_cluster, data=wine, main="Alcohol Content, Ward's")

table(km$cluster, wine$Class)

df$class = as.factor(wine$Class)

df$Alcohol = as.factor(ifelse(df$Alcohol>0,"High","Low"))

str(df)

disMat = daisy(df, metric="gower")

set.seed(123)

pamFit = pam(disMat, k=3)

table(pamFit$clustering)

table(pamFit$clustering, wine$Class)

df$cluster = pamFit$clustering

group = compareGroups(cluster~., data=df)

clustab = createTable(group)

clustab

export2csv(clustab,file="wine\_clusters.csv")