> wine1 <- read.csv(file.choose())

> wine <- wine1[,-1]

> View(wine)

> summary(wine)

Alcohol Malic Ash Alcalinity Magnesium

Min. :11.03 Min. :0.740 Min. :1.360 Min. :10.60 Min. : 70.00

1st Qu.:12.36 1st Qu.:1.603 1st Qu.:2.210 1st Qu.:17.20 1st Qu.: 88.00

Median :13.05 Median :1.865 Median :2.360 Median :19.50 Median : 98.00

Mean :13.00 Mean :2.336 Mean :2.367 Mean :19.49 Mean : 99.74

3rd Qu.:13.68 3rd Qu.:3.083 3rd Qu.:2.558 3rd Qu.:21.50 3rd Qu.:107.00

Max. :14.83 Max. :5.800 Max. :3.230 Max. :30.00 Max. :162.00

Phenols Flavanoids Nonflavanoids Proanthocyanins Color

Min. :0.980 Min. :0.340 Min. :0.1300 Min. :0.410 Min. : 1.280

1st Qu.:1.742 1st Qu.:1.205 1st Qu.:0.2700 1st Qu.:1.250 1st Qu.: 3.220

Median :2.355 Median :2.135 Median :0.3400 Median :1.555 Median : 4.690

Mean :2.295 Mean :2.029 Mean :0.3619 Mean :1.591 Mean : 5.058

3rd Qu.:2.800 3rd Qu.:2.875 3rd Qu.:0.4375 3rd Qu.:1.950 3rd Qu.: 6.200

Max. :3.880 Max. :5.080 Max. :0.6600 Max. :3.580 Max. :13.000

Hue Dilution Proline

Min. :0.4800 Min. :1.270 Min. : 278.0

1st Qu.:0.7825 1st Qu.:1.938 1st Qu.: 500.5

Median :0.9650 Median :2.780 Median : 673.5

Mean :0.9574 Mean :2.612 Mean : 746.9

3rd Qu.:1.1200 3rd Qu.:3.170 3rd Qu.: 985.0

Max. :1.7100 Max. :4.000 Max. :1680.0

> describe(wine)

vars n mean sd median trimmed mad min max range

Alcohol 1 178 13.00 0.81 13.05 13.01 1.01 11.03 14.83 3.80

Malic 2 178 2.34 1.12 1.87 2.21 0.77 0.74 5.80 5.06

Ash 3 178 2.37 0.27 2.36 2.37 0.24 1.36 3.23 1.87

Alcalinity 4 178 19.49 3.34 19.50 19.42 3.04 10.60 30.00 19.40

Magnesium 5 178 99.74 14.28 98.00 98.44 14.83 70.00 162.00 92.00

Phenols 6 178 2.30 0.63 2.36 2.29 0.75 0.98 3.88 2.90

Flavanoids 7 178 2.03 1.00 2.13 2.02 1.24 0.34 5.08 4.74

Nonflavanoids 8 178 0.36 0.12 0.34 0.36 0.13 0.13 0.66 0.53

Proanthocyanins 9 178 1.59 0.57 1.56 1.56 0.56 0.41 3.58 3.17

Color 10 178 5.06 2.32 4.69 4.83 2.24 1.28 13.00 11.72

Hue 11 178 0.96 0.23 0.96 0.96 0.24 0.48 1.71 1.23

Dilution 12 178 2.61 0.71 2.78 2.63 0.77 1.27 4.00 2.73

Proline 13 178 746.89 314.91 673.50 719.30 300.23 278.00 1680.00 1402.00

skew kurtosis se

Alcohol -0.05 -0.89 0.06

Malic 1.02 0.22 0.08

Ash -0.17 1.03 0.02

Alcalinity 0.21 0.40 0.25

Magnesium 1.08 1.96 1.07

Phenols 0.09 -0.87 0.05

Flavanoids 0.02 -0.91 0.07

Nonflavanoids 0.44 -0.68 0.01

Proanthocyanins 0.51 0.47 0.04

Color 0.85 0.30 0.17

Hue 0.02 -0.40 0.02

Dilution -0.30 -1.11 0.05

Proline 0.75 -0.31 23.60

> str(wine)

'data.frame': 178 obs. of 13 variables:

$ Alcohol : num 14.2 13.2 13.2 14.4 13.2 ...

$ Malic : num 1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ...

$ Ash : num 2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 2.17 2.27 ...

$ Alcalinity : num 15.6 11.2 18.6 16.8 21 15.2 14.6 17.6 14 16 ...

$ Magnesium : int 127 100 101 113 118 112 96 121 97 98 ...

$ Phenols : num 2.8 2.65 2.8 3.85 2.8 3.27 2.5 2.6 2.8 2.98 ...

$ Flavanoids : num 3.06 2.76 3.24 3.49 2.69 3.39 2.52 2.51 2.98 3.15 ...

$ Nonflavanoids : num 0.28 0.26 0.3 0.24 0.39 0.34 0.3 0.31 0.29 0.22 ...

$ Proanthocyanins: num 2.29 1.28 2.81 2.18 1.82 1.97 1.98 1.25 1.98 1.85 ...

$ Color : num 5.64 4.38 5.68 7.8 4.32 6.75 5.25 5.05 5.2 7.22 ...

$ Hue : num 1.04 1.05 1.03 0.86 1.04 1.05 1.02 1.06 1.08 1.01 ...

$ Dilution : num 3.92 3.4 3.17 3.45 2.93 2.85 3.58 3.58 2.85 3.55 ...

$ Proline : int 1065 1050 1185 1480 735 1450 1290 1295 1045 1045 ...

> attach(wine)

> a <- cor(wine)

> a

Alcohol Malic Ash Alcalinity Magnesium

Alcohol 1.00000000 0.09439694 0.211544596 -0.31023514 0.27079823

Malic 0.09439694 1.00000000 0.164045470 0.28850040 -0.05457510

Ash 0.21154460 0.16404547 1.000000000 0.44336719 0.28658669

Alcalinity -0.31023514 0.28850040 0.443367187 1.00000000 -0.08333309

Magnesium 0.27079823 -0.05457510 0.286586691 -0.08333309 1.00000000

Phenols 0.28910112 -0.33516700 0.128979538 -0.32111332 0.21440123

Flavanoids 0.23681493 -0.41100659 0.115077279 -0.35136986 0.19578377

Nonflavanoids -0.15592947 0.29297713 0.186230446 0.36192172 -0.25629405

Proanthocyanins 0.13669791 -0.22074619 0.009651935 -0.19732684 0.23644061

Color 0.54636420 0.24898534 0.258887259 0.01873198 0.19995001

Hue -0.07174720 -0.56129569 -0.074666889 -0.27395522 0.05539820

Dilution 0.07234319 -0.36871043 0.003911231 -0.27676855 0.06600394

Proline 0.64372004 -0.19201056 0.223626264 -0.44059693 0.39335085

Phenols Flavanoids Nonflavanoids Proanthocyanins Color

Alcohol 0.28910112 0.2368149 -0.1559295 0.136697912 0.54636420

Malic -0.33516700 -0.4110066 0.2929771 -0.220746187 0.24898534

Ash 0.12897954 0.1150773 0.1862304 0.009651935 0.25888726

Alcalinity -0.32111332 -0.3513699 0.3619217 -0.197326836 0.01873198

Magnesium 0.21440123 0.1957838 -0.2562940 0.236440610 0.19995001

Phenols 1.00000000 0.8645635 -0.4499353 0.612413084 -0.05513642

Flavanoids 0.86456350 1.0000000 -0.5378996 0.652691769 -0.17237940

Nonflavanoids -0.44993530 -0.5378996 1.0000000 -0.365845099 0.13905701

Proanthocyanins 0.61241308 0.6526918 -0.3658451 1.000000000 -0.02524993

Color -0.05513642 -0.1723794 0.1390570 -0.025249931 1.00000000

Hue 0.43368134 0.5434786 -0.2626396 0.295544253 -0.52181319

Dilution 0.69994936 0.7871939 -0.5032696 0.519067096 -0.42881494

Proline 0.49811488 0.4941931 -0.3113852 0.330416700 0.31610011

Hue Dilution Proline

Alcohol -0.07174720 0.072343187 0.6437200

Malic -0.56129569 -0.368710428 -0.1920106

Ash -0.07466689 0.003911231 0.2236263

Alcalinity -0.27395522 -0.276768549 -0.4405969

Magnesium 0.05539820 0.066003936 0.3933508

Phenols 0.43368134 0.699949365 0.4981149

Flavanoids 0.54347857 0.787193902 0.4941931

Nonflavanoids -0.26263963 -0.503269596 -0.3113852

Proanthocyanins 0.29554425 0.519067096 0.3304167

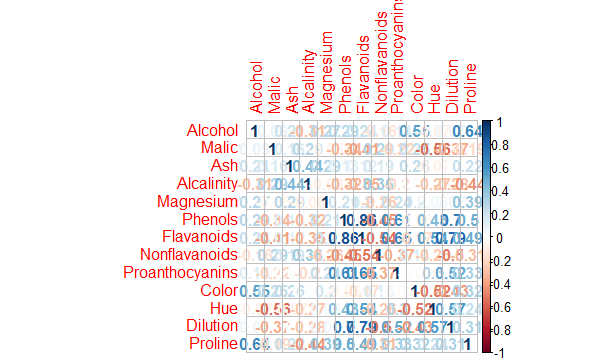
Color -0.52181319 -0.428814942 0.3161001

Hue 1.00000000 0.565468293 0.2361834

Dilution 0.56546829 1.000000000 0.3127611

Proline 0.23618345 0.312761075 1.0000000

> corrplot(a,method = "number")



> pcawine <- princomp(wine,cor = TRUE,scores = TRUE,covmat = NULL) #PCA

> summary(pcawine)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5

Standard deviation 2.1692972 1.5801816 1.2025273 0.9586313 0.92370351

Proportion of Variance 0.3619885 0.1920749 0.1112363 0.0706903 0.06563294

Cumulative Proportion 0.3619885 0.5540634 0.6652997 0.7359900 0.80162293

Comp.6 Comp.7 Comp.8 Comp.9 Comp.10

Standard deviation 0.80103498 0.74231281 0.59033665 0.53747553 0.50090167

Proportion of Variance 0.04935823 0.04238679 0.02680749 0.02222153 0.01930019

Cumulative Proportion 0.85098116 0.89336795 0.92017544 0.94239698 0.96169717

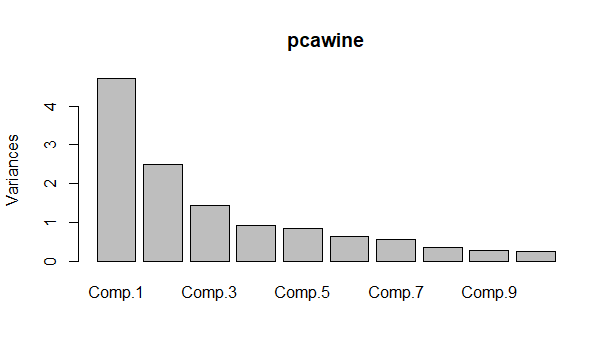
Comp.11 Comp.12 Comp.13

Standard deviation 0.47517222 0.41081655 0.321524394

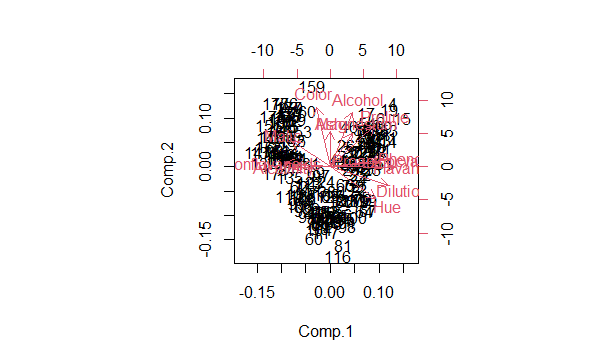
Proportion of Variance 0.01736836 0.01298233 0.007952149

Cumulative Proportion 0.97906553 0.99204785 1.000000000

>plot(pcawine)



>biplot(pcawine)



>pcawine$scores

>wine\_score <- cbind(wine,pcawine$scores[,1:3]) #[,1:3]considering 66.5% pca components scores as per problem statement

>View(wine\_score)

#cluster analysis

############### hierarchial clustering for principle components scores ###############

>library(factoextra)

>library(NbClust)

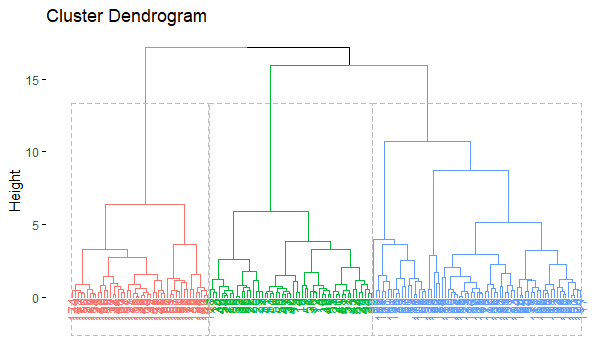
>library(dendextend)

>clust\_data <-wine\_score[,14:16]

>norm\_clust <- scale(clust\_data)

>clust <- eclust(norm\_clust,"hclust",k=3,graph = FALSE)

>fviz\_dend(clust,rect = TRUE)



> groups <- cutree(clust,k=3)

> finalhclust <- data.frame(groups,wine)

> aggregate(finalhclust,by=list(clust$cluster),FUN=mean)

Group.1 groups Alcohol Malic Ash Alcalinity Magnesium Phenols

1 1 1 13.75421 2.015439 2.449474 16.95965 107.36842 2.853333

2 2 2 12.30890 1.896712 2.255342 20.14932 94.57534 2.262603

3 3 3 13.15771 3.386042 2.437083 21.51042 98.54167 1.681667

Flavanoids Nonflavanoids Proanthocyanins Color Hue Dilution Proline

1 2.9950877 0.2878947 1.942105 5.570877 1.061228 3.162807 1116.2982

2 2.0909589 0.3578082 1.594932 3.110685 1.055973 2.781781 536.9589

3 0.7885417 0.4558333 1.167708 7.410833 0.684375 1.698542 627.5000

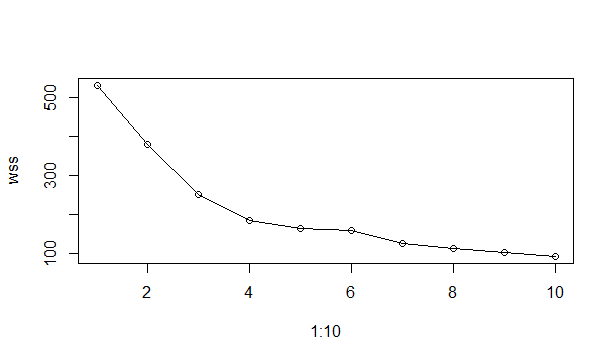
> wss <- (nrow(norm\_clust)-1)\*sum(apply(norm\_clust,2,var))

> for(i in 1:10){

+ wss[i]=sum(kmeans(norm\_clust,centers = i)$withinss)

+ }

> plot(1:10,wss, type = "o") #from scree plot no of clusters is 3



> #alternative method

> noofclust <- NbClust(clust\_data,distance = "euclidean",method = "kmeans",min.nc = 2,max.nc = 10,index = "all")

\*\*\* : The Hubert index is a graphical method of determining the number of clusters.

In the plot of Hubert index, we seek a significant knee that corresponds to a

significant increase of the value of the measure i.e the significant peak in Hubert

index second differences plot.

\*\*\* : The D index is a graphical method of determining the number of clusters.

In the plot of D index, we seek a significant knee (the significant peak in Dindex

second differences plot) that corresponds to a significant increase of the value of

the measure.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Among all indices:

\* 4 proposed 2 as the best number of clusters

\* 15 proposed 3 as the best number of clusters

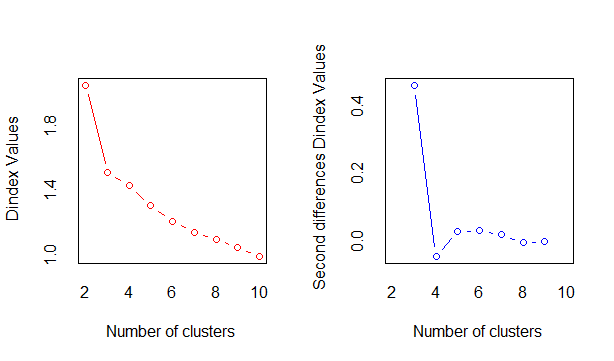
\* 1 proposed 4 as the best number of clusters

\* 1 proposed 7 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 3



> fviz\_nbclust(noofclust) #optimal number of clusters

Among all indices:

===================

\* 2 proposed 0 as the best number of clusters

\* 1 proposed 1 as the best number of clusters

\* 4 proposed 2 as the best number of clusters

\* 15 proposed 3 as the best number of clusters

\* 1 proposed 4 as the best number of clusters

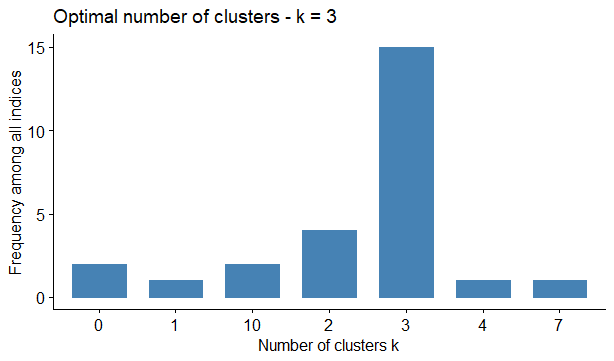
\* 1 proposed 7 as the best number of clusters

\* 2 proposed 10 as the best number of clusters

Conclusion

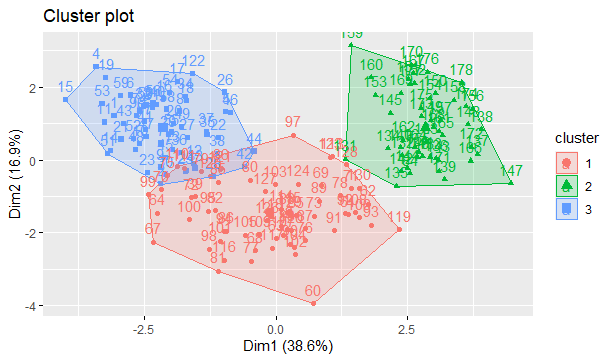
=========================

\* According to the majority rule, the best number of clusters is 3 .



> km <- kmeans(norm\_clust,3)

> fviz\_cluster(km,data = wine[-1])



> final <- data.frame(km$cluster,wine)

> aggregate(final,by=list(km$cluster),FUN = mean)

Group.1 km.cluster Alcohol Malic Ash Alcalinity Magnesium Phenols

1 1 1 12.27632 1.906176 2.222500 19.88824 92.73529 2.237500

2 2 2 13.15163 3.344490 2.434694 21.43878 99.02041 1.678163

3 3 3 13.68672 2.006066 2.472295 17.49508 108.13115 2.854918

Flavanoids Nonflavanoids Proanthocyanins Color Hue Dilution Proline

1 2.0375000 0.3601471 1.593971 3.030882 1.0578824 2.773824 514.750

2 0.7979592 0.4508163 1.163061 7.343265 0.6859184 1.690204 627.551

3 3.0091803 0.2922951 1.931148 5.482295 1.0636066 3.171148 1101.541

no of clusters of principle components scores is similar to the wine data (i.e. problem statement)