Principal Component Analysis(PCA)

Problem statement 1

Perform Principal component analysis and perform clustering using first

3 principal component scores (both heirarchial and k mean clustering(scree plot or elbow curve) and obtain

optimum number of clusters and check whether we have obtained same number of clusters with the original data

Answer:

Rcode:

```
wine <- read.csv(file.choose())

View(wine)

str(wine)

names(wine)

class(wine)

summary(wine)

attach(wine)


sd(Type)

sd(Alcohol)

sd(Malic)

sd(Alcalinity)

sd(Magnesium)

sd(Phenols)

sd(Flavanoids)

sd(Nonflavanoids)

sd(Proanthocyanins)

sd(Color)
```

```
sd(Hue)
sd(Dilution)
sd(Proline)

var(Type)
var(Alcohol)
var(Malic)
var(Alcalinity)
var(Magnesium)
var(Phenols)
var(Flavanoids)
var(Nonflavanoids)
var(Proanthocyanins)
var(Color)
var(Hue)
var(Dilution)
var(Proline)

library(moments)
skewness(Type)
skewness(Alcohol)
skewness(Malic)
skewness(Alcalinity)
skewness(Magnesium)
skewness(Phenols)
skewness(Flavanoids)
skewness(Nonflavanoids)
```

```
skewness(Proanthocyanins)
skewness(Color)
skewness(Hue)
skewness(Dilution)
skewness(Proline)

kurtosis(Type)
kurtosis(Alcohol)
kurtosis(Malic)
kurtosis(Alcalinity)
kurtosis(Magnesium)
kurtosis(Phenols)
kurtosis(Flavanoids)
kurtosis(Nonflavanoids)
kurtosis(Proanthocyanins)
kurtosis(Color)
kurtosis(Hue)
kurtosis(Dilution)
kurtosis(Proline)
names(wine)
plot(Alcohol)
plot(Malic)
plot(Ash)
plot(Alcalinity)
plot(Magnesium)
plot(Phenols)
plot(Flavanoids)
```

```
plot(Nonflavanoids)
plot(Proanthocyanins)
plot(Color)
plot(Hue)
plot(Dilution)
plot(Proline)

hist(Alcohol)
hist(Malic)
hist(Ash)
hist(Alcalinity)
hist(Magnesium)
hist(Phenols)
hist(Flavanoids)
hist(Nonflavanoids)
hist(Proanthocyanins)
hist(Color)
hist(Hue)
hist(Dilution)
hist(Proline)

boxplot(Alcohol,horizontal = T,xlab="Alcohol")
boxplot(Malic,horizontal = T,xlab="Malic")
boxplot(Ash,horizontal = T,xlab="Ash")
boxplot(Alcalinity,horizontal = T,xlab="Alcalinity")
boxplot(Magnesium,horizontal = T,xlab="Magnesium")
boxplot(Phenols,horizontal = T,xlab="Phenols")
```

```
boxplot(Flavanoids,horizontal = T,xlab="Flavanoids")
boxplot(Nonflavanoids,horizontal = T,xlab="Nonflavanoids")
boxplot(Proanthocyanins,horizontal = T,xlab="Proanthocyanins")
boxplot(Color,horizontal = T,xlab="Color")
boxplot(Hue,horizontal = T,xlab="Hue")
boxplot(Dilution,horizontal = T,xlab="Dilution")
boxplot(Proline,horizontal = T,xlab="Proline")

mydata <- wine[,-1]
cor(mydata)

pcaobj <- princomp(mydata,cor = T,scores = T,covmat = NULL)
str(pcaobj)
loadings(pcaobj)
plot(pcaobj)
biplot(pcaobj)
plot(cumsum(pcaobj$sdev*pcaobj$sdev)*100/(sum(pcaobj$sdev*pcaobj$sdev)),type="b")
pcaobj$scores[,1:3]
mydata1<-cbind(wine,pcaobj$scores[,1:3])
View(mydata1)

# Hierarchial Clustering

clus_data<-wine[,8:10]

# Normalizing the data
```

```
norm_clus<-scale(clus_data)
distance<-dist(norm_clus,method = "euclidean")

fit<-hclust(distance,method="complete")

plot(fit)

rect.hclust(fit, k=7, border="red")
groups <- cutree(fit,5)
clust_1 <- as.matrix(groups)
View(clust_1)
final <- cbind(clust_1,mydata1)
View(final)
aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = mean)
aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = max)
aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = min)
write.csv(final,file = "data_clust.txt",row.names = F,col.names = F)
write.csv(final,file = "data_clust.csv",row.names = F,col.names = F)
getwd()

# K-Means Clustering :
library(plyr)
mydata2 <- final
View(mydata)

normalized_data<-scale(mydata2[,15:17])
kmeans_clust <- kmeans(normalized_data,7)
```

str(kmeans_clust)

final1<- cbind(kmeans_clust$cluster,mydata2)

View(final1)

wss = (nrow(normalized_data)-1)*sum(apply(normalized_data, 2, var))   # Determine number of clusters by scree-plot

for (i in 1:7) wss[i] = sum(kmeans(normalized_data, centers=i)$withinss)

plot(1:7, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of squares")   # Look for an "elbow" in the scree plot #

title(sub = "K-Means Clustering Scree-Plot")

aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = mean)

aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = min)

aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = max)

kmeans_clust$centers

table(kmeans_clust$cluster)

write.csv(final1,file = "data_clust1.csv",row.names = F,col.names = F)

getwd()

Console:

```
> wine <- read.csv(file.choose())
> View(wine)
```

PCA.R × wine ×

Filter

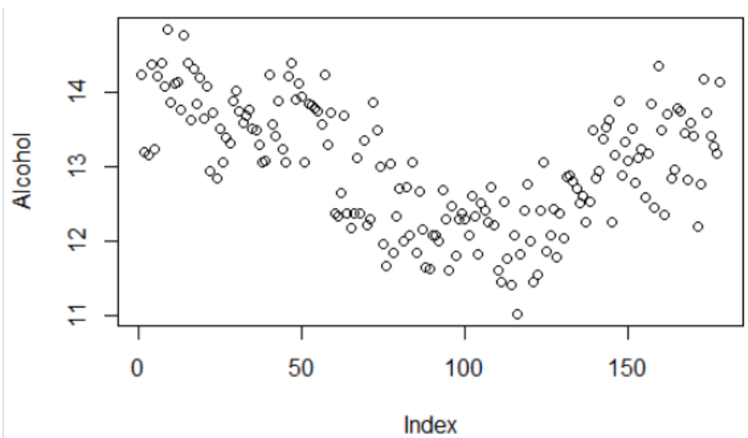
	Type	Alcohol	Malic	Ash	Alcalinity	Magnesium	Phenols	Flavanoids	Nonflavanoi
1	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28
2	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26
3	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30
4	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24
5	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39
6	1	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34
7	1	14.39	1.87	2.45	14.6	96	2.50	2.52	0.30
8	1	14.06	2.15	2.61	17.6	121	2.60	2.51	0.31
9	1	14.83	1.64	2.17	14.0	97	2.80	2.98	0.29
10	1	13.86	1.35	2.27	16.0	98	2.98	3.15	0.22

Showing 1 to 10 of 178 entries, 14 total columns

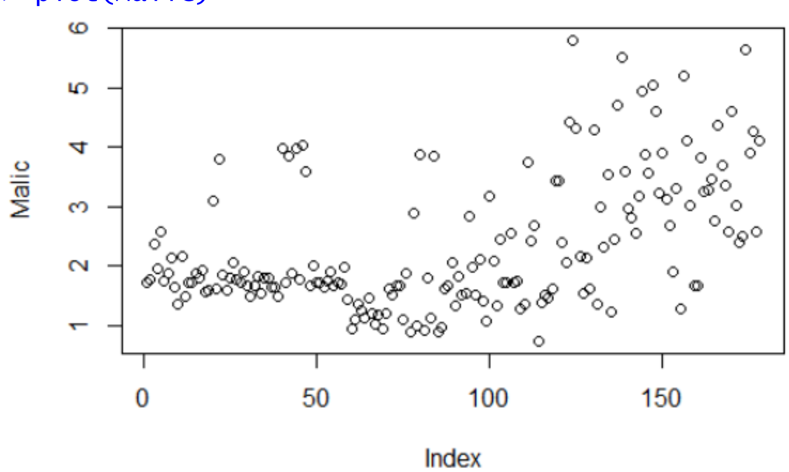
```
> str(wine)
'data.frame':	178 obs. of  14 variables:
 $ Type           : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Alcohol        : num  14.2 13.2 13.2 14.4 13.2 ...
 $ Malic          : num  1.71 1.78 2.36 1.95 2.59 1.76 1.87 2.15 1.64 1.35 ..
.
 $ Ash            : num  2.43 2.14 2.67 2.5 2.87 2.45 2.45 2.61 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 ...
> names(wine)
 [1] "Type"            "Alcohol"         "Malic"           "Ash"
 [5] "Alcalinity"      "Magnesium"       "Phenols"         "Flavanoids"
 [9] "Nonflavanoids"   "Proanthocyanins" "Color"           "Hue"
[13] "Dilution"        "Proline"
> class(wine)
[1] "data.frame"
> summary(wine)
      Type          Alcohol          Malic            Ash
 Min.   :1.000   Min.   :11.03   Min.   :0.740   Min.   :1.360
 1st Qu.:1.000   1st Qu.:12.36   1st Qu.:1.603   1st Qu.:2.210
 Median :2.000   Median :13.05   Median :1.865   Median :2.360
 Mean   :1.938   Mean   :13.00   Mean   :2.336   Mean   :2.367
 3rd Qu.:3.000   3rd Qu.:13.68   3rd Qu.:3.083   3rd Qu.:2.558
 Max.   :3.000   Max.   :14.83   Max.   :5.800   Max.   :3.230
   Alcalinity      Magnesium         Phenols        Flavanoids
 Min.   :10.60   Min.   : 70.00   Min.   :0.980   Min.   :0.340
 1st Qu.:17.20   1st Qu.: 88.00   1st Qu.:1.742   1st Qu.:1.205
 Median :19.50   Median : 98.00   Median :2.355   Median :2.135
 Mean   :19.49   Mean   : 99.74   Mean   :2.295   Mean   :2.029
 3rd Qu.:21.50   3rd Qu.:107.00   3rd Qu.:2.800   3rd Qu.:2.875
 Max.   :30.00   Max.   :162.00   Max.   :3.880   Max.   :5.080
 Nonflavanoids    Proanthocyanins     Color             Hue
 Min.   :0.1300   Min.   :0.410   Min.   : 1.280   Min.   :0.4800
 1st Qu.:0.2700   1st Qu.:1.250   1st Qu.: 3.220   1st Qu.:0.7825
 Median :0.3400   Median :1.555   Median : 4.690   Median :0.9650
 Mean   :0.3619   Mean   :1.591   Mean   : 5.058   Mean   :0.9574
 3rd Qu.:0.4375   3rd Qu.:1.950   3rd Qu.: 6.200   3rd Qu.:1.1200
 Max.   :0.6600   Max.   :3.580   Max.   :13.000   Max.   :1.7100
    Dilution        Proline
 Min.   :1.270   Min.   : 278.0
 1st Qu.:1.938   1st Qu.: 500.5
 Median :2.780   Median : 673.5
 Mean   :2.612   Mean   : 746.9
 3rd Qu.:3.170   3rd Qu.: 985.0
 Max.   :4.000   Max.   :1680.0
> attach(wine)
> sd(Type)
[1] 0.775035
> sd(Alcohol)
[1] 0.8118265
> sd(Malic)
[1] 1.117146
> sd(Alcalinity)
[1] 3.339564
> sd(Magnesium)
[1] 14.28248
> sd(Phenols)
```

```
[1] 0.625851
> sd(Flavanoids)
[1] 0.9988587
> sd(Nonflavanoids)
[1] 0.1244533
> sd(Proanthocyanins)
[1] 0.5723589
> sd(Color)
[1] 2.318286
> sd(Hue)
[1] 0.2285716
> sd(Dilution)
[1] 0.7099904
> sd(Proline)
[1] 314.9075
>
> var(Type)
[1] 0.6006792
> var(Alcohol)
[1] 0.6590623
> var(Malic)
[1] 1.248015
> var(Alcalinity)
[1] 11.15269
> var(Magnesium)
[1] 203.9893
> var(Phenols)
[1] 0.3916895
> var(Flavanoids)
[1] 0.9977187
> var(Nonflavanoids)
[1] 0.01548863
> var(Proanthocyanins)
[1] 0.3275947
> var(Color)
[1] 5.374449
> var(Hue)
[1] 0.05224496
> var(Dilution)
[1] 0.5040864
> var(Proline)
[1] 99166.72
>
> library(moments)
> skewness(Type)
[1] 0.1065237
> skewness(Alcohol)
[1] -0.05104747
> skewness(Malic)
[1] 1.030869
> skewness(Alcalinity)
[1] 0.2112473
> skewness(Magnesium)
[1] 1.088915
> skewness(Phenols)
[1] 0.08590677
> skewness(Flavanoids)
[1] 0.02512948
> skewness(Nonflavanoids)
[1] 0.446349
> skewness(Proanthocyanins)
[1] 0.512769
> skewness(Color)
[1] 0.8612481
> skewness(Hue)
[1] 0.02091312
> skewness(Dilution)
[1] -0.3046899
```

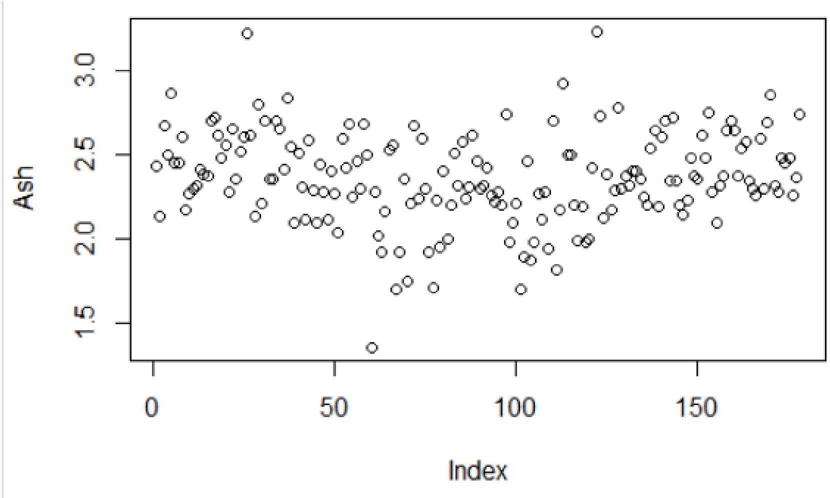
```
> skewness(Proline)
[1] 0.7613362
> kurtosis(Type)
[1] 1.68056
> kurtosis(Alcohol)
[1] 2.13774
> kurtosis(Malic)
[1] 3.257348
> kurtosis(Alcalinity)
[1] 3.440823
> kurtosis(Magnesium)
[1] 5.012806
> kurtosis(Phenols)
[1] 2.154143
> kurtosis(Flavanoids)
[1] 2.110635
> kurtosis(Nonflavanoids)
[1] 2.347048
> kurtosis(Proanthocyanins)
[1] 3.505671
> kurtosis(Color)
[1] 3.33737
> kurtosis(Hue)
[1] 2.631975
> kurtosis(Dilution)
[1] 1.910325
> kurtosis(Proline)
[1] 2.725
> plot(Alcohol)
```



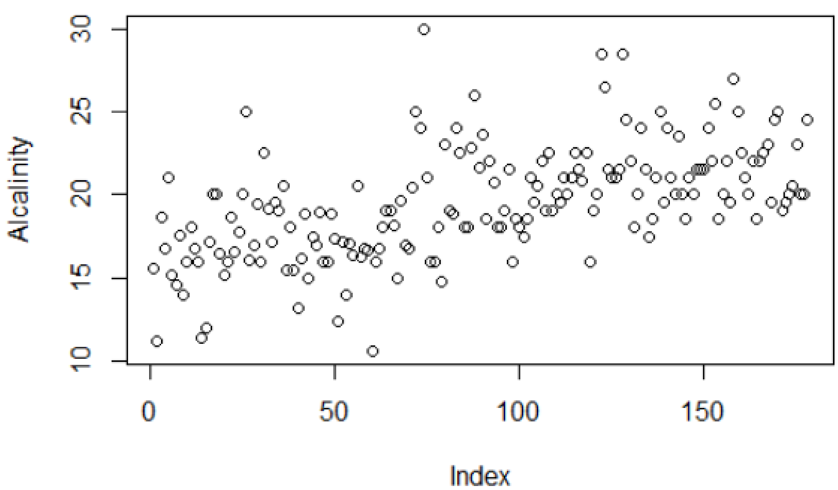

```
> plot(Malic)
```



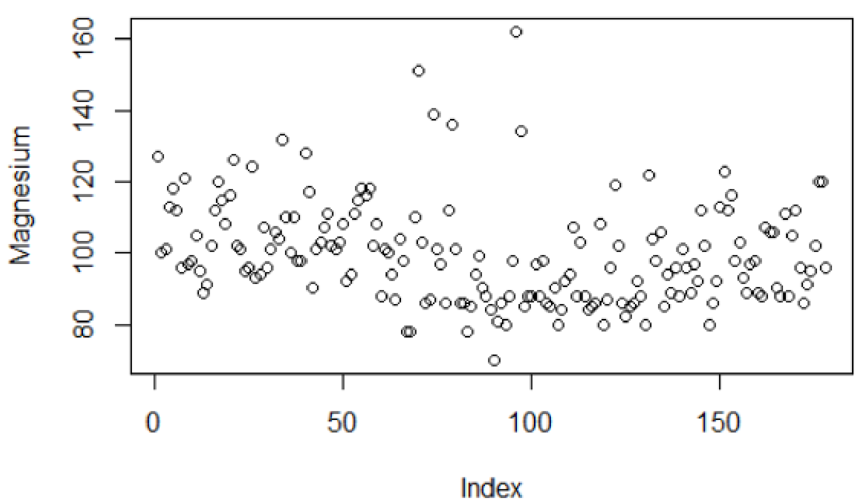
```
> plot(Ash)
```



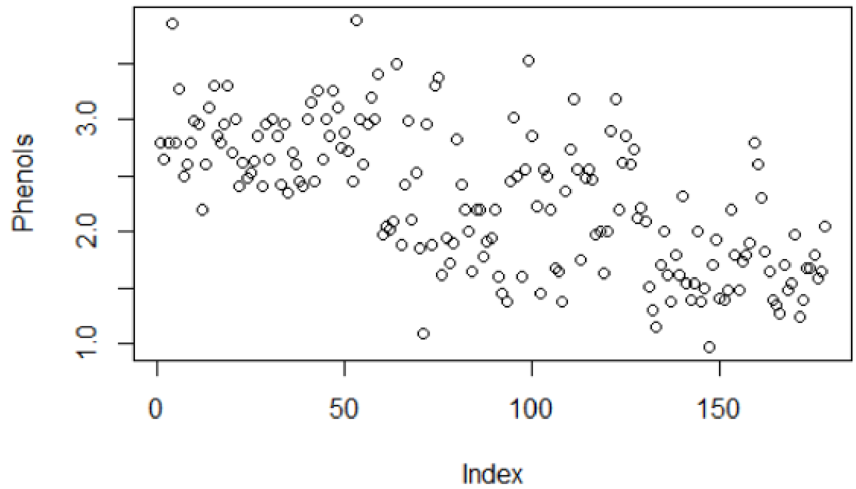

```
> plot(Alcalinity)
```



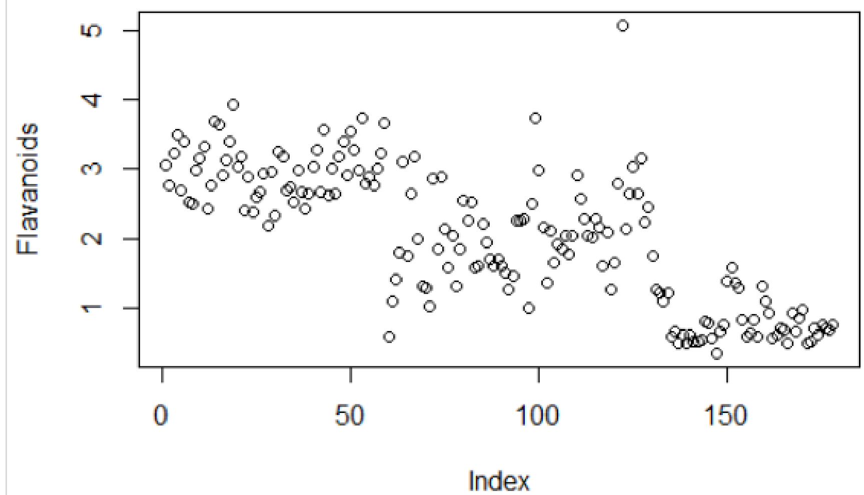

```
> plot(Magnesium)
```



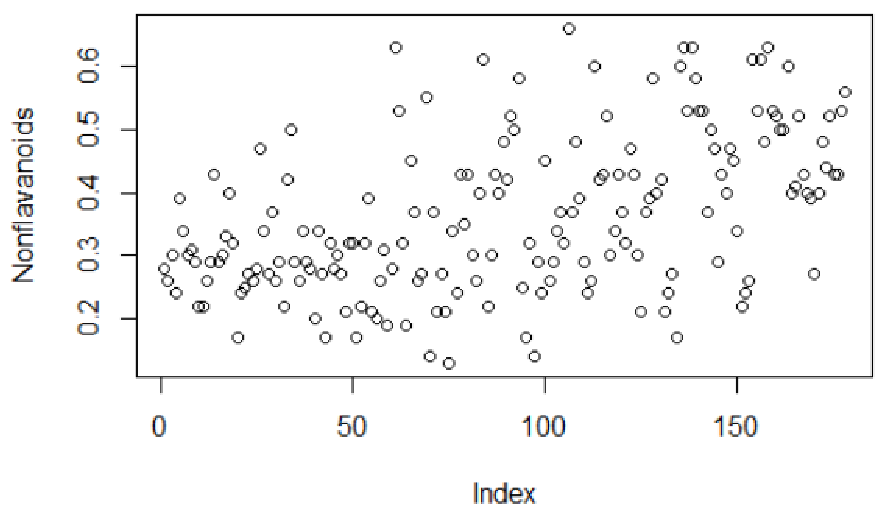
```
> plot(Phenols)
```



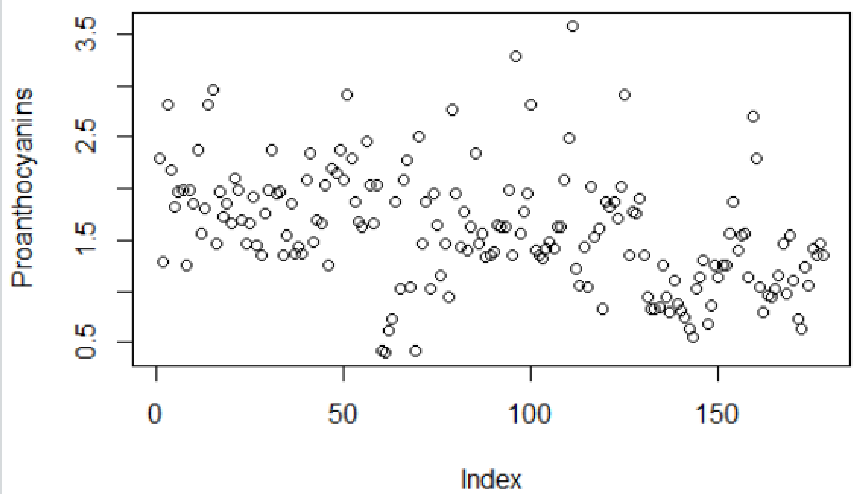

```
> plot(Flavanoids)
```



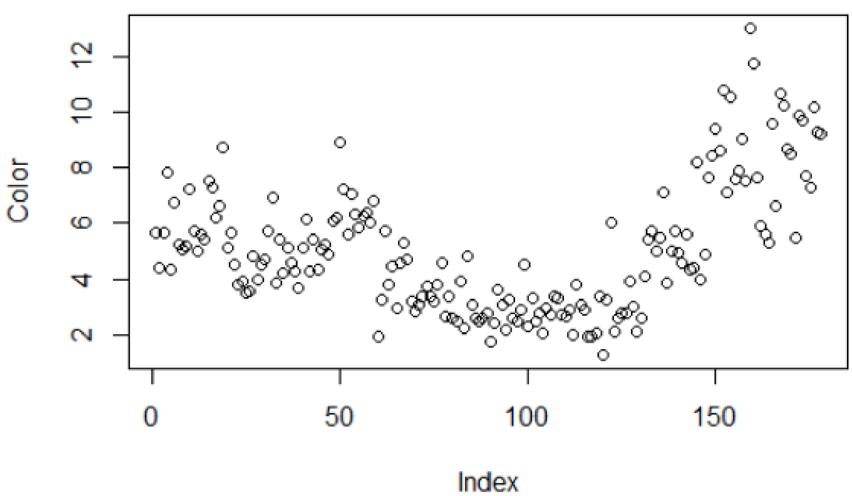

```
> plot(Nonflavanoids)
```



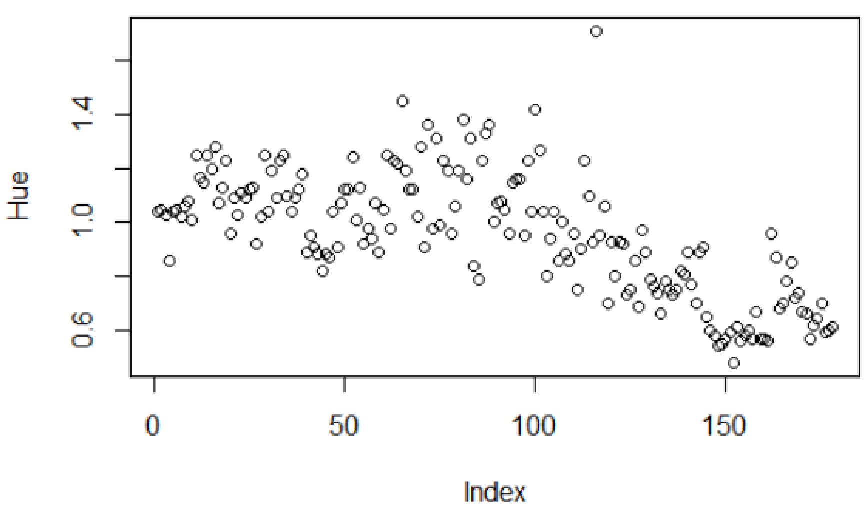
```
> plot(Proanthocyanins)
```



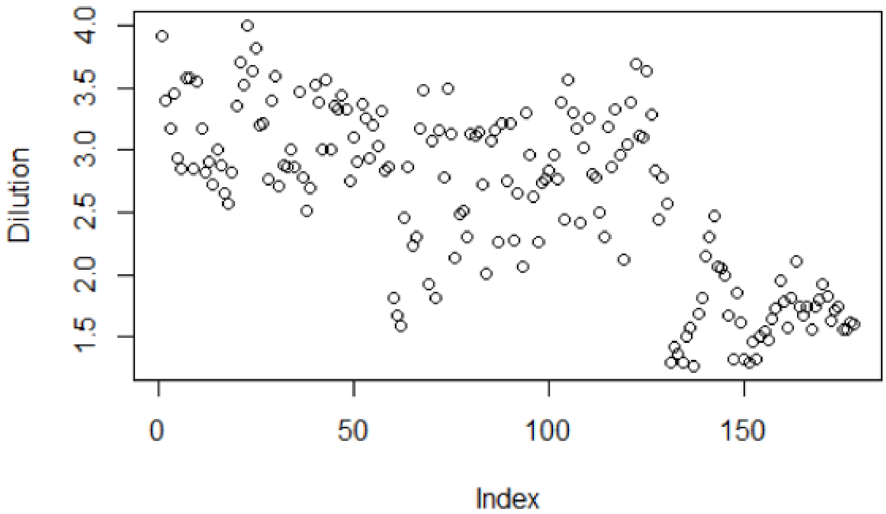

```
> plot(Color)
```



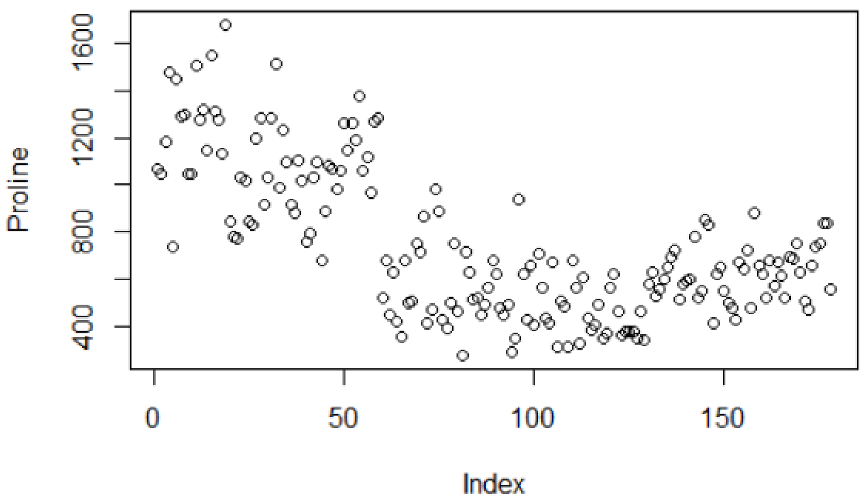

```
> plot(Hue)
```



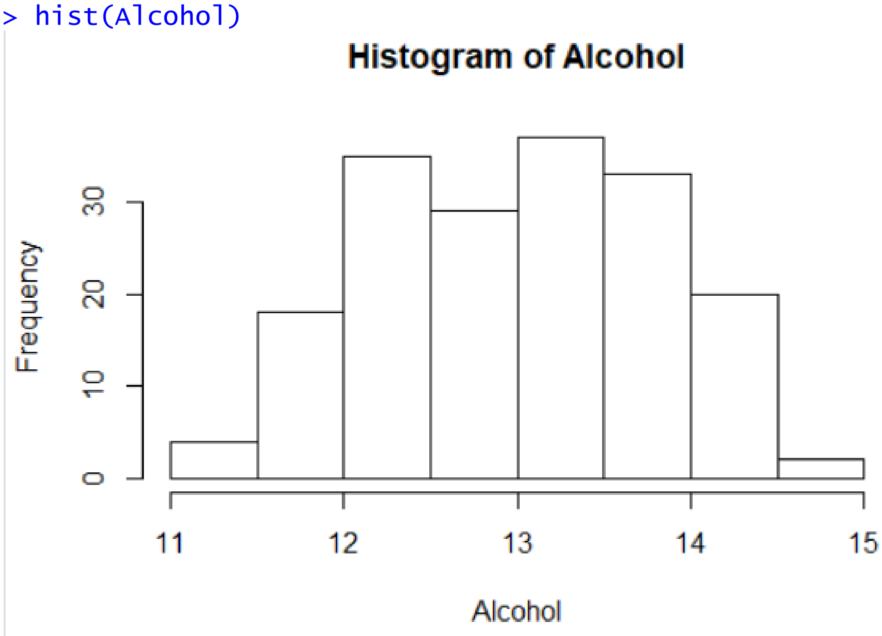
```
> plot(Dilution)
```



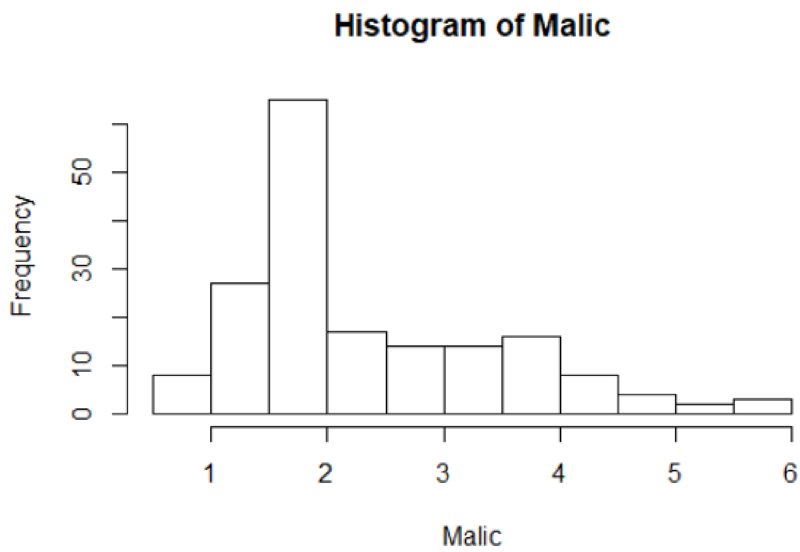

```
> plot(Proline)
```



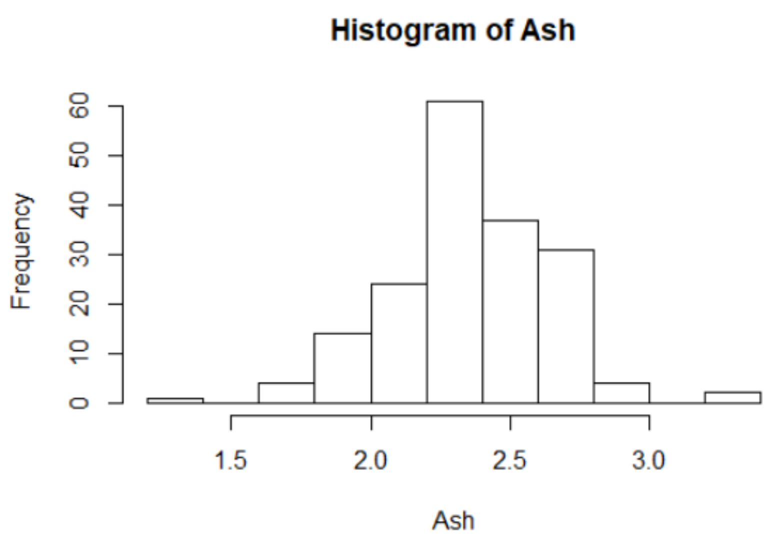

```
> hist(Alcohol)
```



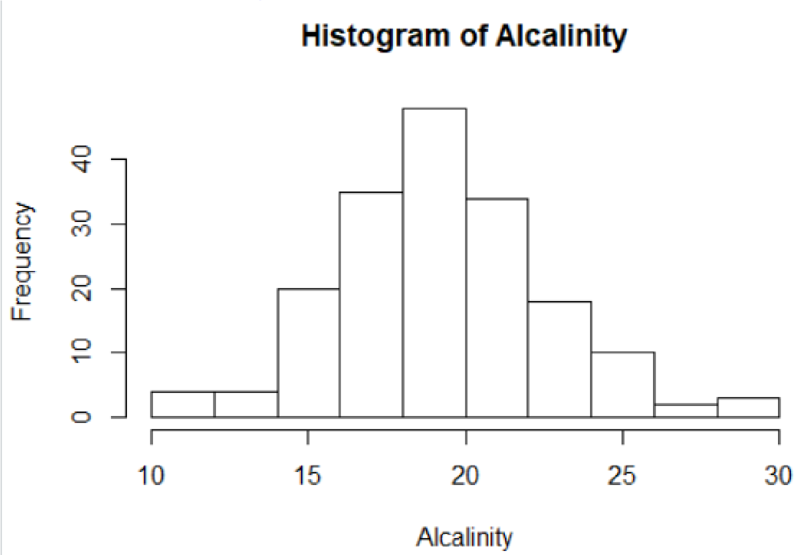
```
> hist(Malic)
```



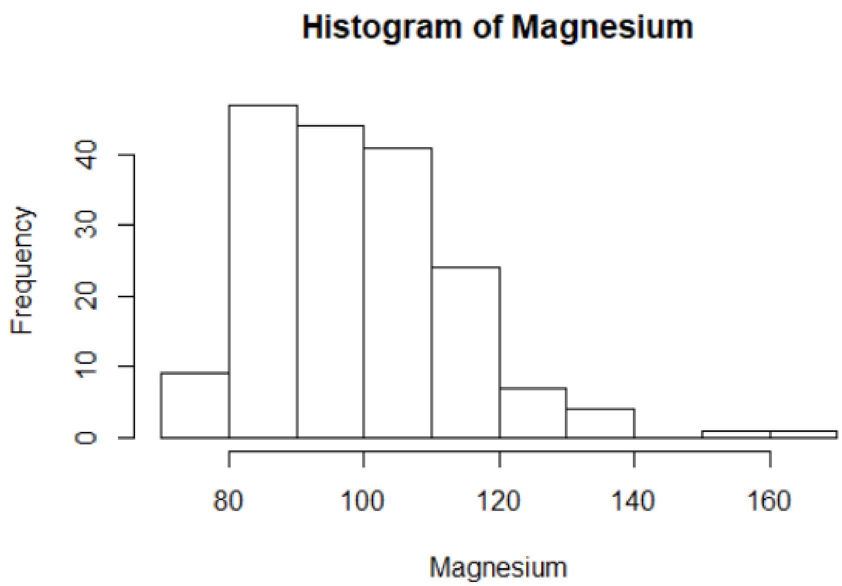

```
> hist(Ash)
```



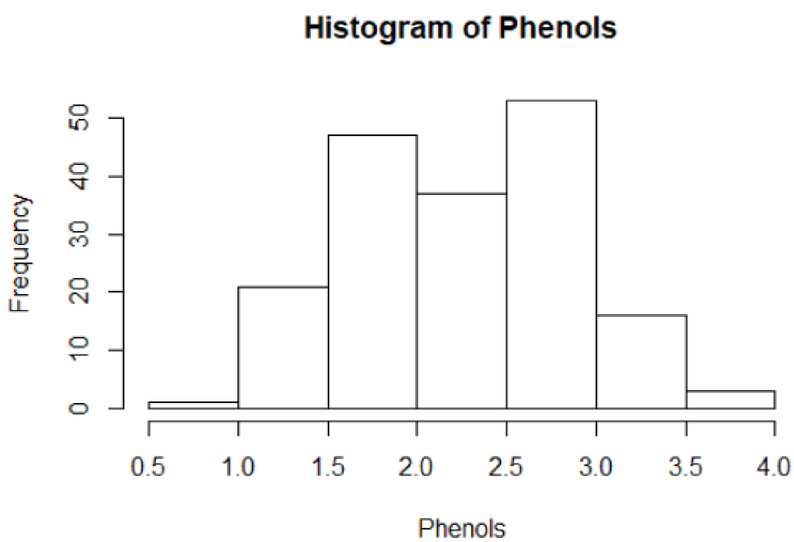

```
> hist(Alcalinity)
```



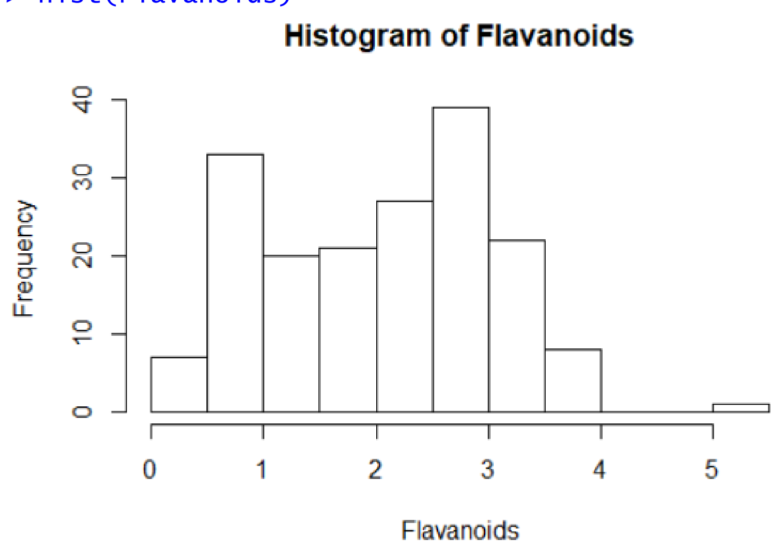

```
> hist(Magnesium)
```



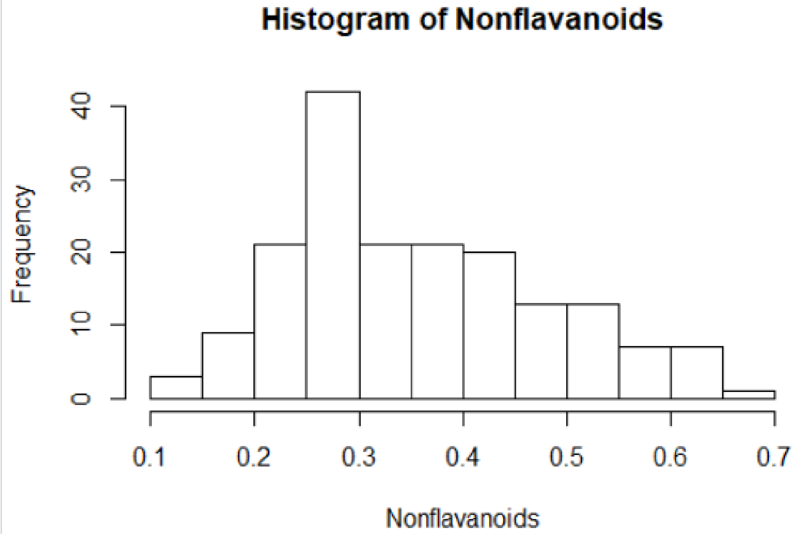

```
> hist(Phenols)
```



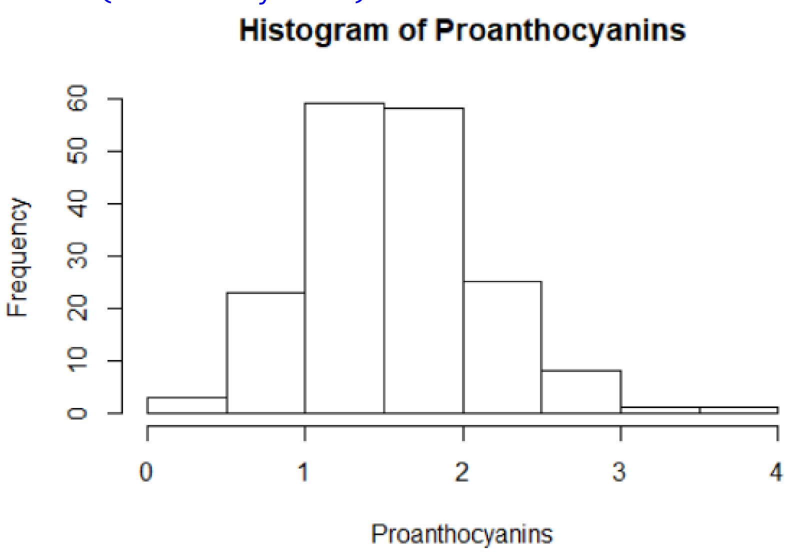

```
> hist(Flavanoids)
```



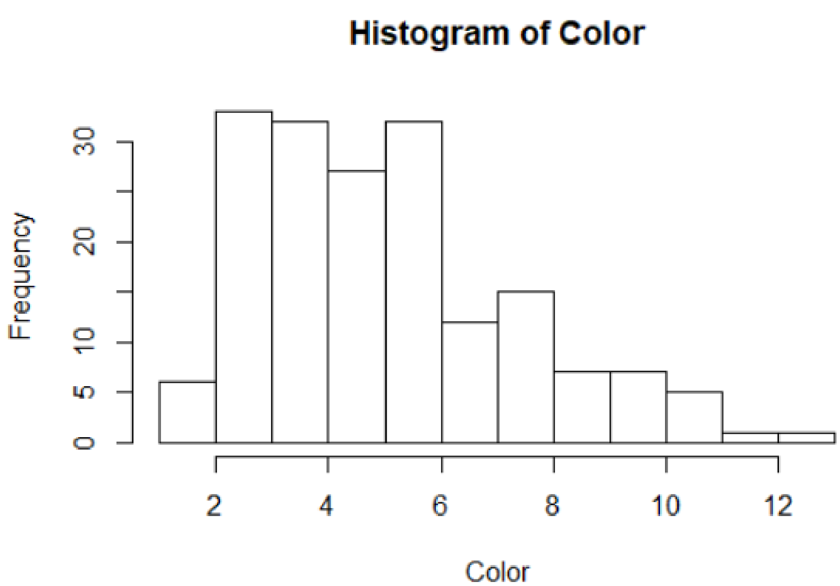
```
> hist(Nonflavanoids)
```



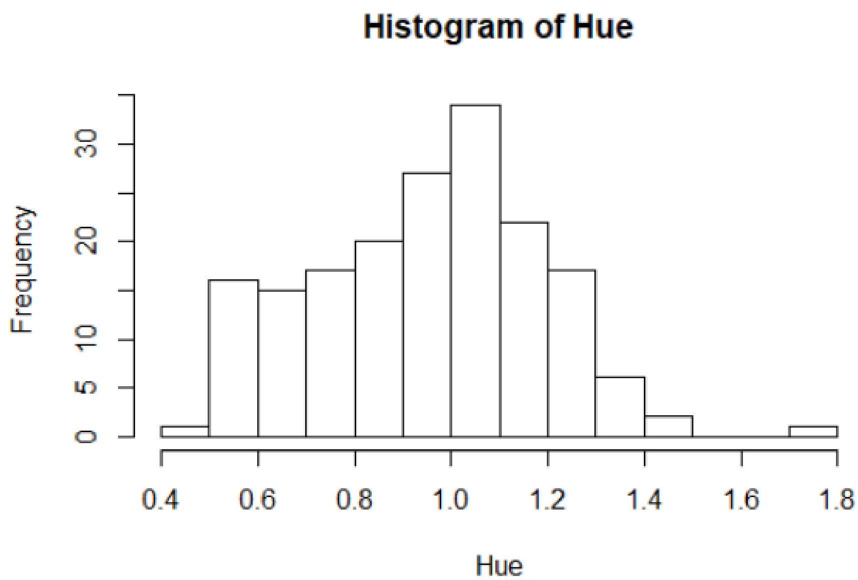

```
> hist(Proanthocyanins)
```



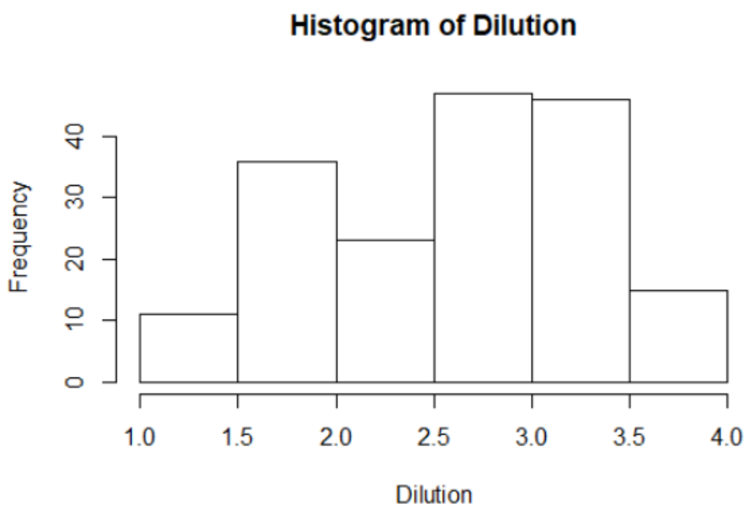

```
> hist(Color)
```



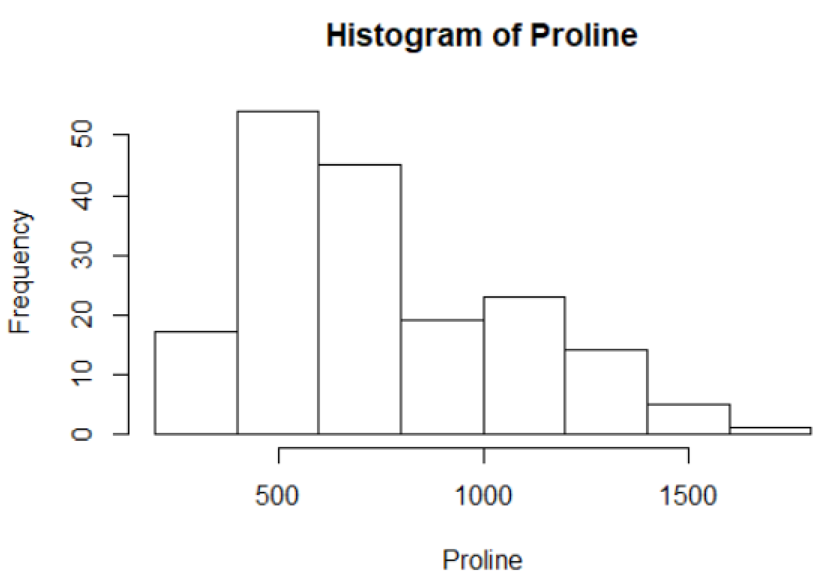
```
> hist(Hue)
```



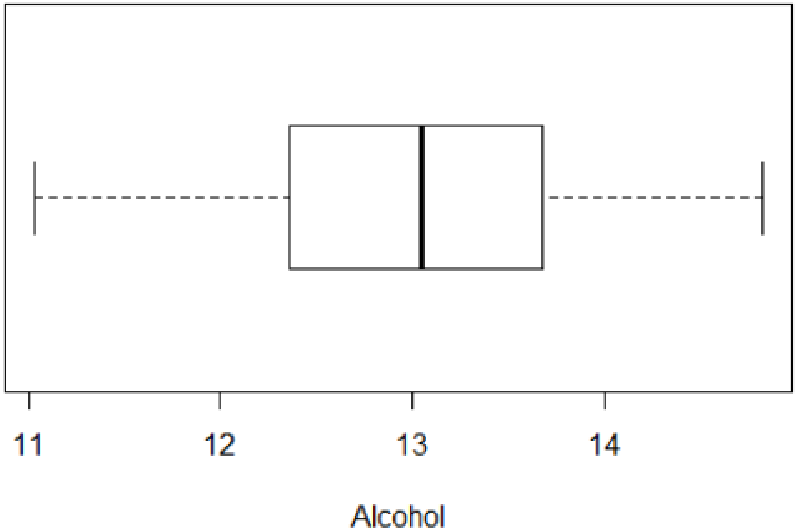

```
> hist(Dilution)
```



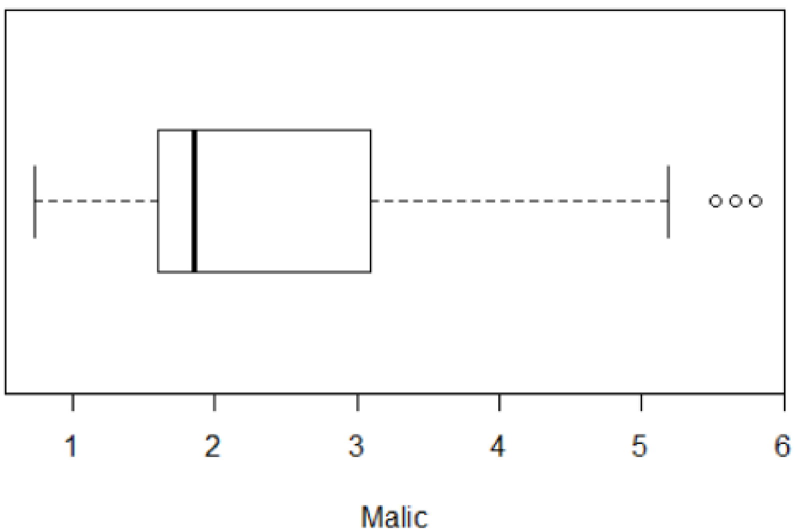

```
> hist(Proline)
```



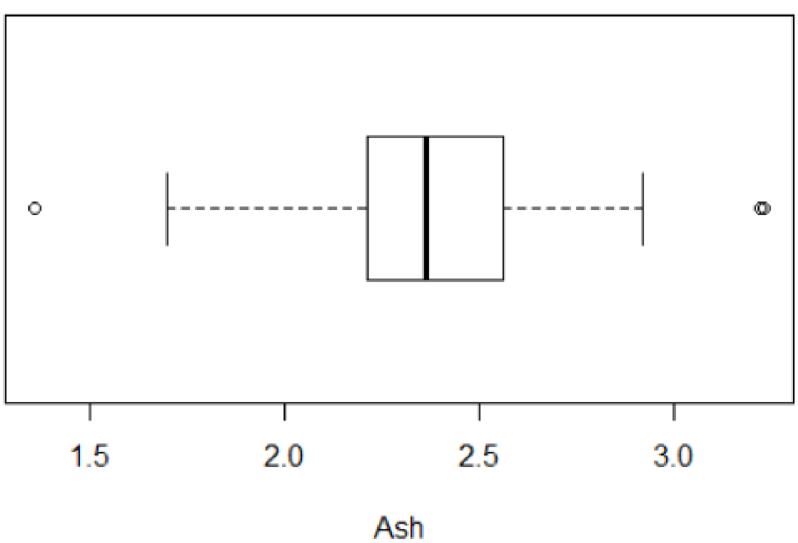
```
> boxplot(Alcohol,horizontal = T,xlab="Alcohol")
```



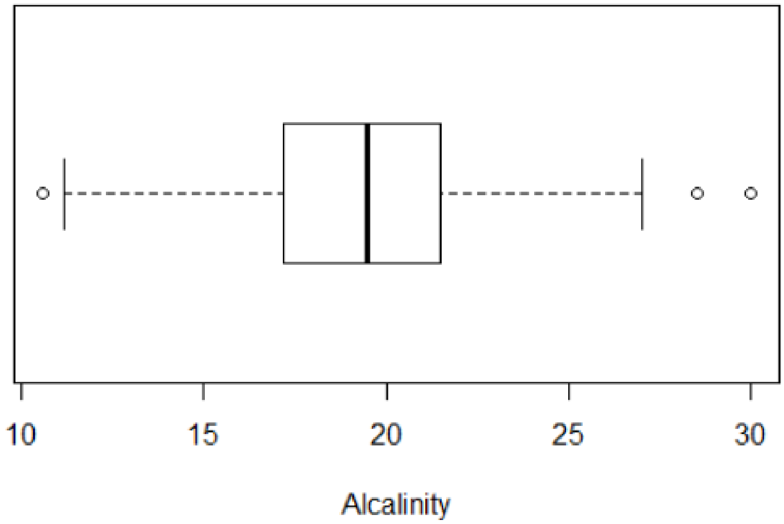

```
> boxplot(Malic,horizontal = T,xlab="Malic")
```



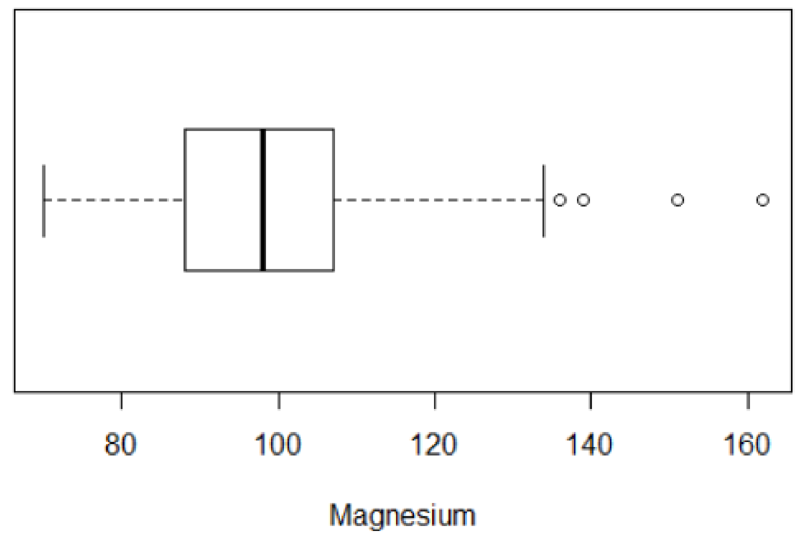

```
> boxplot(Ash,horizontal = T,xlab="Ash")
```



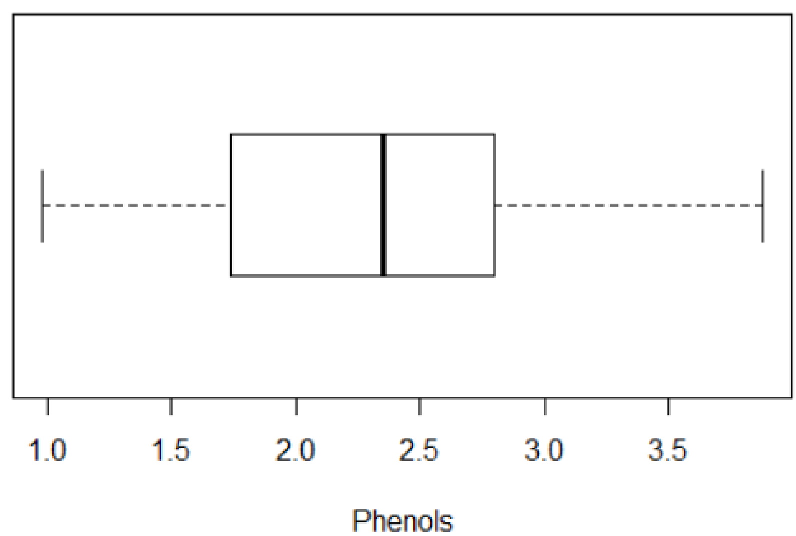
```
> boxplot(Alcalinity,horizontal = T,xlab="Alcalinity")
```



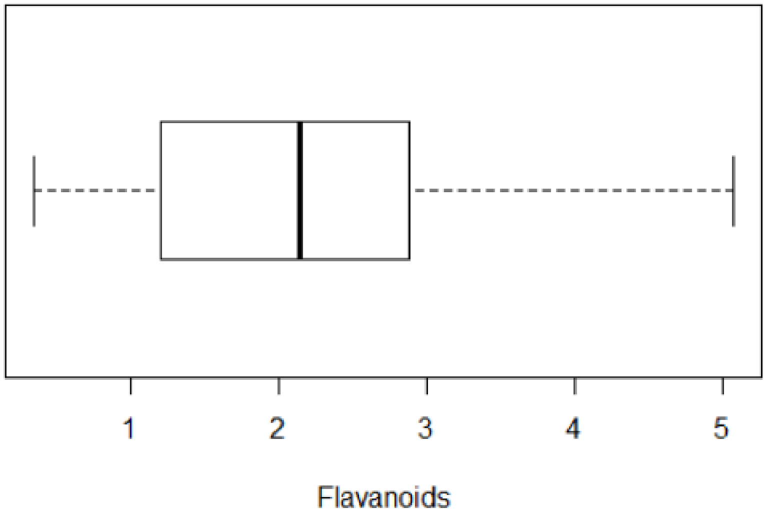

```
> boxplot(Magnesium,horizontal = T,xlab="Magnesium")
```



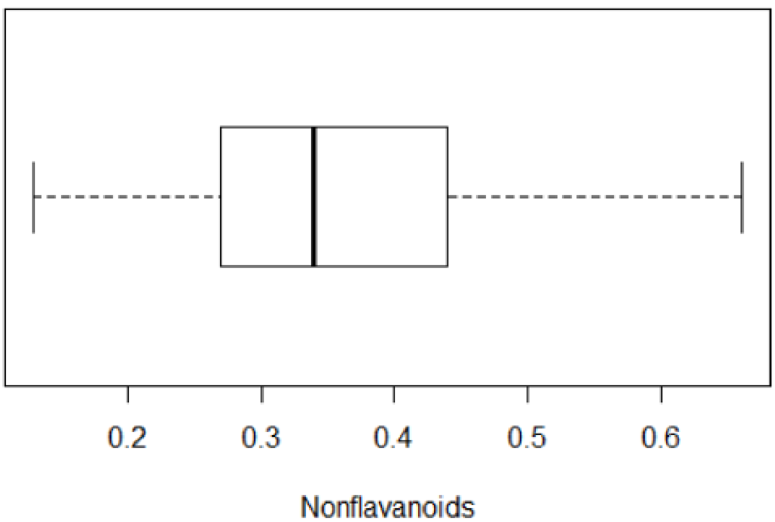

```
> boxplot(Phenols,horizontal = T,xlab="Phenols")
```



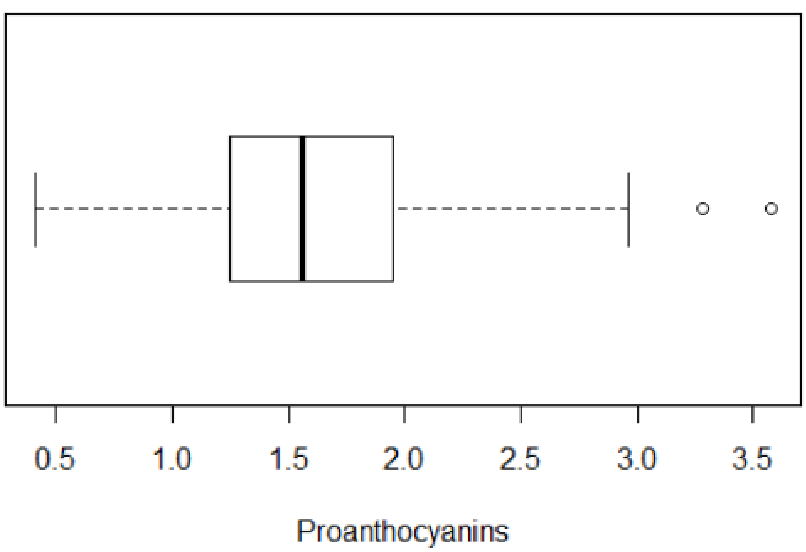
```
> boxplot(Flavanoids,horizontal = T,xlab="Flavanoids")
```



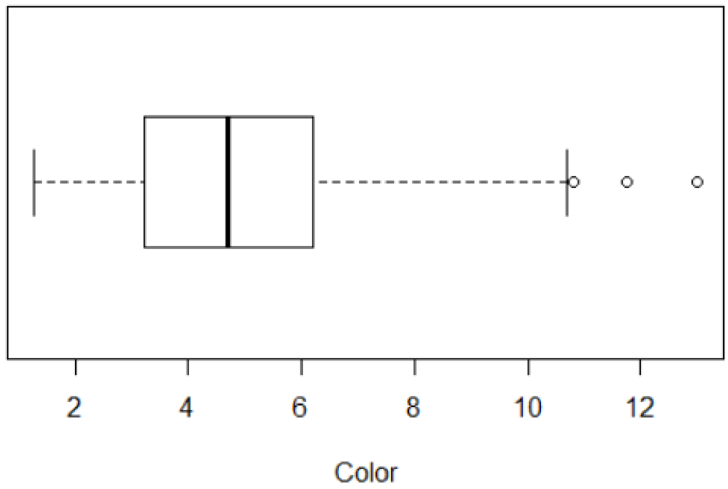

```
> boxplot(Nonflavanoids,horizontal = T,xlab="Nonflavanoids")
```



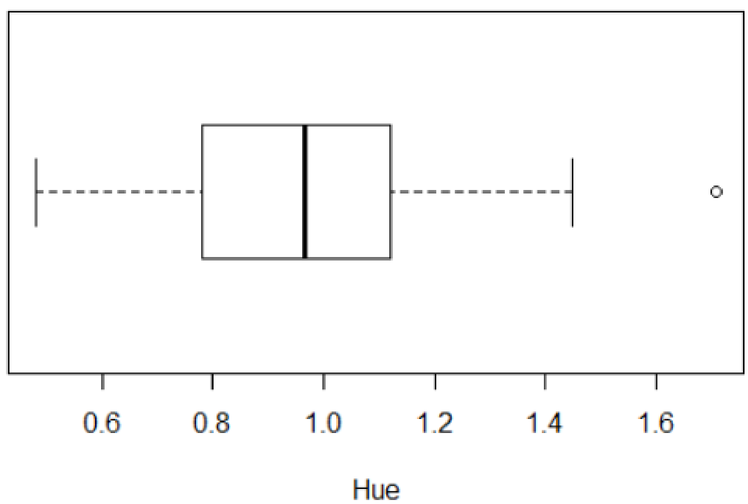

```
> boxplot(Proanthocyanins,horizontal = T,xlab="Proanthocyanins")
```



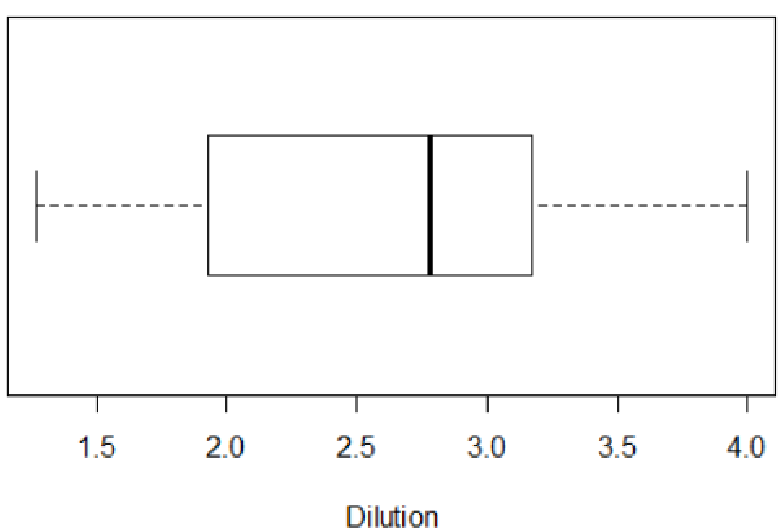
```
> boxplot(Color,horizontal = T,xlab="Color")
```



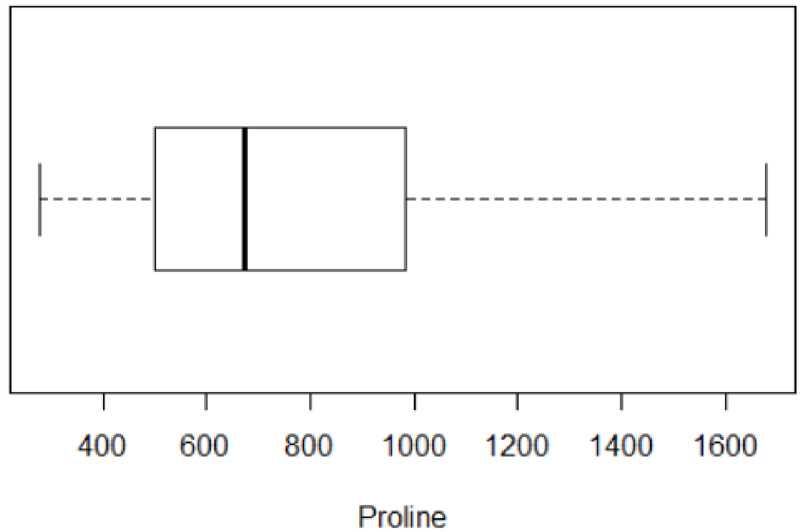

```
> boxplot(Hue,horizontal = T,xlab="Hue")
```




```
> boxplot(Dilution,horizontal = T,xlab="Dilution")
```



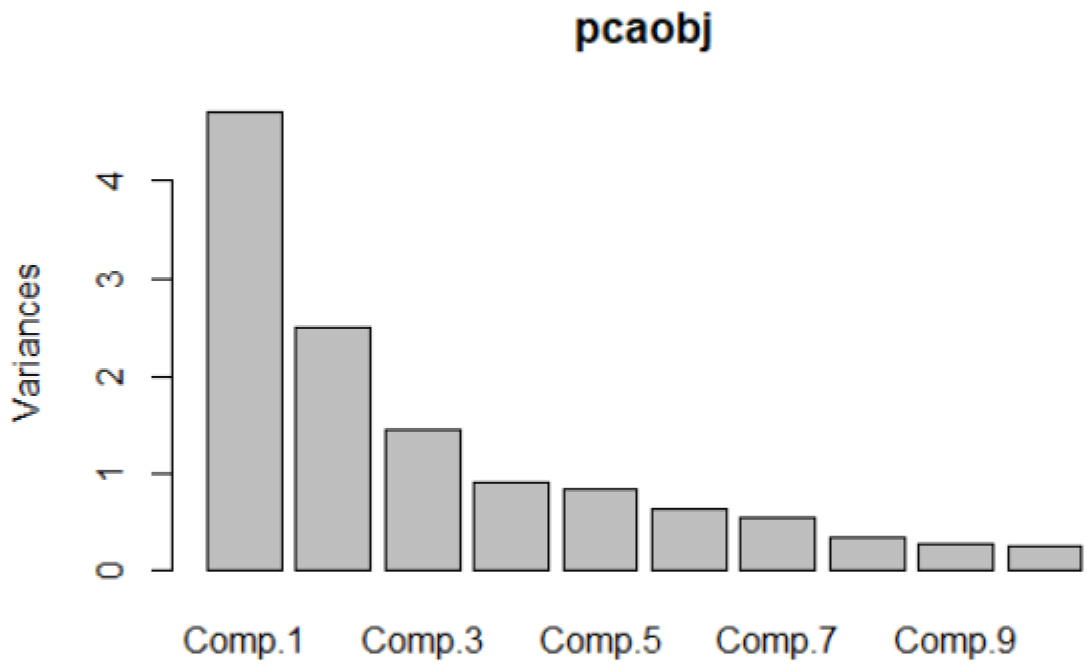
```
> boxplot(Proline,horizontal = T,xlab="Proline")
```



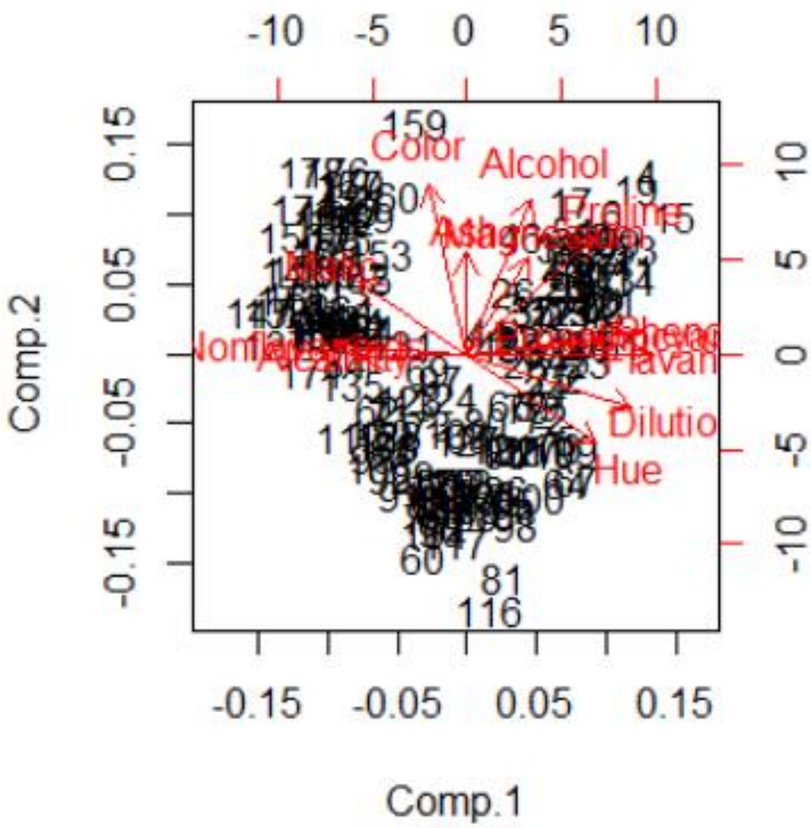

```
> mydata <- wine[,-1]
> cor(mydata)
                    Alcohol       Malic          Ash  Alcalinity   Magnesium
Alcohol          1.00000000  0.09439694  0.211544596 -0.31023514  0.27079823
Malic            0.09439694  1.00000000  0.164045470  0.28850040 -0.05457510
Ash              0.21154460  0.16404547  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
Alcohol          0.28910112  0.2368149    -0.1559295     0.136697912
Malic           -0.33516700 -0.4110066     0.2929771    -0.220746187
Ash              0.12897954  0.1150773     0.1862304     0.009651935
Alcalinity      -0.32111332 -0.3513699     0.3619217    -0.197326836
Magnesium        0.21440123  0.1957838    -0.2562940     0.236440610
Phenols          1.00000000  0.8645635    -0.4499353     0.612413084
Flavanoids       0.86456350  1.0000000    -0.5378996     0.652691769
Nonflavanoids   -0.44993530 -0.5378996     1.0000000    -0.365845099
Proanthocyanins  0.61241308  0.6526918    -0.3658451     1.000000000
Color           -0.05513642 -0.1723794     0.1390570    -0.025249931
Hue              0.43368134  0.5434786    -0.2626396     0.295544253
Dilution         0.69994936  0.7871939    -0.5032696     0.519067096
Proline          0.49811488  0.4941931    -0.3113852     0.330416700
                      Color         Hue     Dilution    Proline
Alcohol          0.54636420 -0.07174720  0.072343187  0.6437200
Malic            0.24898534 -0.56129569 -0.368710428 -0.1920106
Ash              0.25888726 -0.07466689  0.003911231  0.2236263
Alcalinity       0.01873198 -0.27395522 -0.276768549 -0.4405969
Magnesium        0.19995001  0.05539820  0.066003936  0.3933508
Phenols         -0.05513642  0.43368134  0.699949365  0.4981149
Flavanoids      -0.17237940  0.54347857  0.787193902  0.4941931
Nonflavanoids    0.13905701 -0.26263963 -0.503269596 -0.3113852
Proanthocyanins -0.02524993  0.29554425  0.519067096  0.3304167
Color            1.00000000 -0.52181319 -0.428814942  0.3161001
```

```
Hue              -0.52181319  1.00000000  0.565468293  0.2361834
Dilution         -0.42881494  0.56546829  1.000000000  0.3127611
Proline           0.31610011  0.23618345  0.312761075  1.0000000
> pcaobj <- princomp(mydata,cor = T,scores = T,covmat = NULL)
> str(pcaobj)
List of 7
 $ sdev    : Named num [1:13] 2.169 1.58 1.203 0.959 0.924 ...
  ..- attr(*, "names")= chr [1:13] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...
 $ loadings: 'loadings' num [1:13, 1:13] 0.14433 -0.24519 -0.00205 -0.23932 0
.14199 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:13] "Alcohol" "Malic" "Ash" "Alcalinity" ...
  .. ..$ : chr [1:13] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...
 $ center  : Named num [1:13] 13 2.34 2.37 19.49 99.74 ...
  ..- attr(*, "names")= chr [1:13] "Alcohol" "Malic" "Ash" "Alcalinity" ...
 $ scale   : Named num [1:13] 0.81 1.114 0.274 3.33 14.242 ...
  ..- attr(*, "names")= chr [1:13] "Alcohol" "Malic" "Ash" "Alcalinity" ...
 $ n.obs   : int 178
 $ scores  : num [1:178, 1:13] 3.32 2.21 2.52 3.76 1.01 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : NULL
  .. ..$ : chr [1:13] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...
 $ call    : language princomp(x = mydata, cor = T, scores = T, covmat = NULL
)
 - attr(*, "class")= chr "princomp"
> loadings(pcaobj)

Loadings:
                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
Alcohol          0.144  0.484  0.207         0.266  0.214         0.396
Malic           -0.245  0.225        -0.537         0.537 -0.421       
Ash                     0.316 -0.626  0.214  0.143  0.154  0.149 -0.170
Alcalinity      -0.239        -0.612               -0.101  0.287  0.428
Magnesium        0.142  0.300 -0.131  0.352 -0.727        -0.323 -0.156
Phenols          0.395        -0.146 -0.198  0.149               -0.406
Flavanoids       0.423        -0.151 -0.152  0.109               -0.187
Nonflavanoids   -0.299        -0.170  0.203  0.501 -0.259 -0.595 -0.233
Proanthocyanins  0.313        -0.149 -0.399 -0.137 -0.534 -0.372  0.368
Color                   0.530  0.137               -0.419  0.228       
Hue              0.297 -0.279         0.428  0.174  0.106 -0.232  0.437
Dilution         0.376 -0.164 -0.166 -0.184  0.101  0.266              
Proline          0.287  0.365  0.127  0.232  0.158  0.120         0.120
                Comp.9 Comp.10 Comp.11 Comp.12 Comp.13
Alcohol          0.509  0.212   0.226   0.266         
Malic                  -0.309          -0.122         
Ash             -0.308          0.499          -0.141 
Alcalinity       0.200         -0.479                 
Magnesium        0.271                                
Phenols          0.286 -0.320  -0.304   0.304  -0.464 
Flavanoids             -0.163                   0.832 
Nonflavanoids    0.196  0.216  -0.117           0.114 
Proanthocyanins -0.209  0.134   0.237          -0.117 
Color                  -0.291          -0.604         
Hue                    -0.522          -0.259         
Dilution         0.137  0.524          -0.601  -0.157 
Proline         -0.576  0.162  -0.539                 

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.077  0.077  0.077  0.077  0.077  0.077  0.077  0.077
Cumulative Var  0.077  0.154  0.231  0.308  0.385  0.462  0.538  0.615
               Comp.9 Comp.10 Comp.11 Comp.12 Comp.13
SS loadings     1.000   1.000   1.000   1.000   1.000
Proportion Var  0.077   0.077   0.077   0.077   0.077
Cumulative Var  0.692   0.769   0.846   0.923   1.000
```

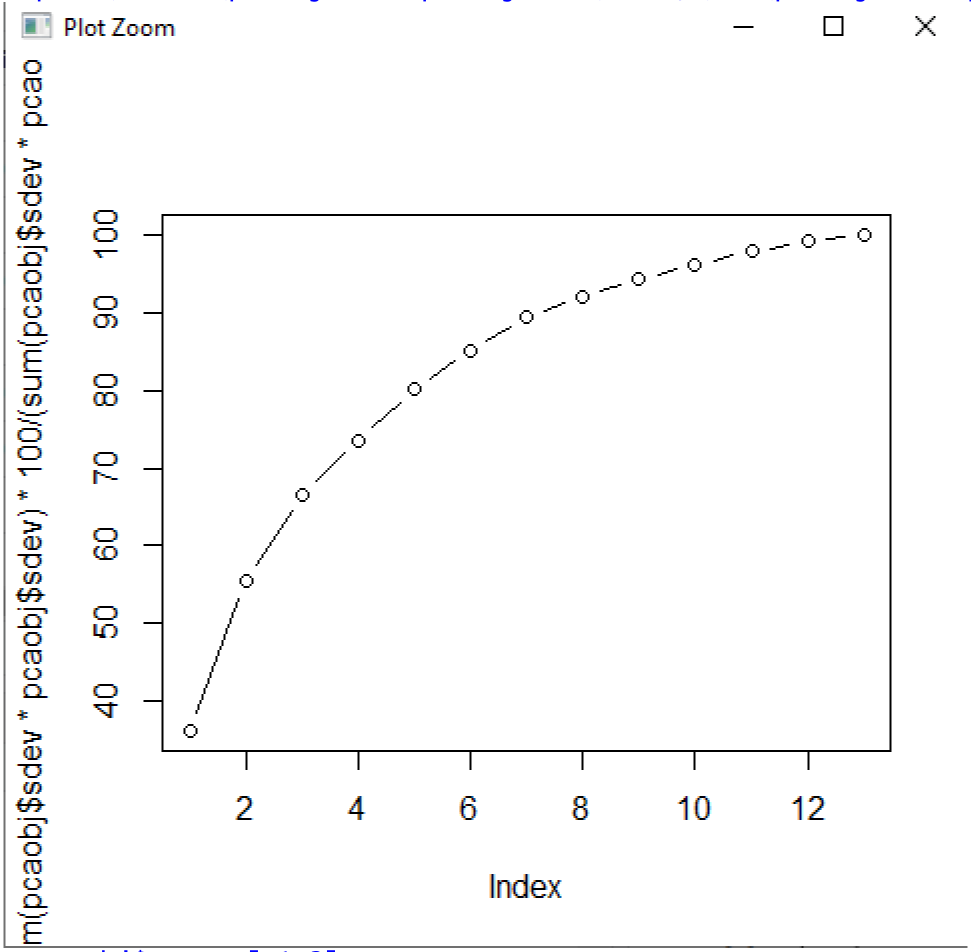
```
> plot(pcaobj)
```




```
> biplot(pcaobj)
```



```
> plot(cumsum(pcaobj$sdev*pcaobj$sdev)*100/(sum(pcaobj$sdev*pcaobj$sdev)),type="b")
```




```
> pcaobj$scores[,1:3]
          Comp.1      Comp.2       Comp.3
 [1,] 3.31675081  1.44346263  0.165739045
 [2,] 2.20946492 -0.33339289  2.026457374
 [3,] 2.51674015  1.03115130 -0.982818670
 [4,] 3.75706561  2.75637191  0.176191842
 [5,] 1.00890849  0.86983082 -2.026688219
 [6,] 3.05025392  2.12240111  0.629395827
 [7,] 2.44908967  1.17485013  0.977094891
 [8,] 2.05943687  1.60896307 -0.146281883
 [9,] 2.51087430  0.91807096  1.770969027
[10,] 2.75362819  0.78943767  0.984247490
[11,] 3.47973668  1.30233324  0.422735217
[12,] 1.75475290  0.61197723  1.190878320
[13,] 2.11346234  0.67570634  0.865086426
[14,] 3.45815682  1.13062988  1.204276353
[15,] 4.31278391  2.09597558  1.263912752
[16,] 2.30518820  1.66255173 -0.217902616
[17,] 2.17195527  2.32730534 -0.831729866
[18,] 1.89897118  1.63136888 -0.794913792
[19,] 3.54198508  2.51834367  0.485458508
[20,] 2.08452220  1.06113799  0.164746678
[21,] 3.12440254  0.78689711  0.364887083
[22,] 1.08657007  0.24174355 -0.936961600
[23,] 2.53522408 -0.09184062  0.311932659
[24,] 1.64498834 -0.51627893 -0.143885095
[25,] 1.76157587 -0.31714893 -0.890285647
[26,] 0.99007910  0.94066734 -3.820908008
[27,] 1.77527763  0.68617513  0.086700406
[28,] 1.23542396 -0.08980704  1.386896545
[29,] 2.18840633  0.68956962 -1.394566881
```

```
[30,]  2.25610898  0.19146194  1.092657258
[31,]  2.50022003  1.24083383 -1.386017855
[32,]  2.67741105  1.47187365  0.332261728
[33,]  1.62857912  0.05270445  0.167128706
[34,]  1.90269086  1.63306043 -1.172082119
[35,]  1.41038853  0.69793432 -0.479743025
[36,]  1.90382623  0.17671095 -0.450835040
[37,]  1.38486223  0.65863985 -0.458438581
[38,]  1.12220741  0.11410976  0.039107277
[39,]  1.50219450 -0.76943201  1.426177346
[40,]  2.52980109  1.80300198  0.343152389
[41,]  2.58809543  0.77961630  0.118477466
[42,]  0.66848199  0.16996094  0.783362548
[43,]  3.07080699  1.15591896  0.312758084
[44,]  0.46220914  0.33074213  0.201476496
[45,]  2.10135193 -0.07100892  0.655849415
[46,]  1.13616618  1.77710739 -0.028705736
[47,]  2.72660096  1.19133469  0.539773261
[48,]  2.82133927  0.64625860  1.155552411
[49,]  2.00985085  1.24702946  0.057293988
[50,]  2.70749130  1.75196741  0.643113612
[51,]  3.21491747  0.16699199  1.973571680
[52,]  2.85895983  0.74527880 -0.004719502
[53,]  3.50560436  1.61273386  0.520774530
[54,]  2.22479138  1.87516800 -0.339549850
[55,]  2.14698782  1.01675154  0.957762762
[56,]  2.46932948  1.32900831 -0.513437453
[57,]  2.74151791  1.43654878  0.612473396
[58,]  2.17374092  1.21219984 -0.261779593
[59,]  3.13938015  1.73157912  0.285661413
[60,] -0.92858197 -3.07348616  4.585064007
[61,] -1.54248014 -1.38144351  0.874683112
[62,] -1.83624976 -0.82998412  1.605702186
[63,]  0.03060683 -1.26278614  1.784408010
[64,]  2.05026161 -1.92503260  0.007368777
[65,] -0.60968083 -1.90805881 -0.679357938
[66,]  0.90022784 -0.76391147 -0.573361302
[67,]  2.24850719 -1.88459248  2.031840193
[68,]  0.18338403 -2.42714611  1.069745560
[69,] -0.81280503 -0.22051399  0.707005396
[70,]  1.97562050 -1.40328323  1.238276220
[71,] -1.57221622 -0.88498314  0.628997950
[72,]  1.65768181 -0.95671220 -1.952584217
[73,] -0.72537239 -1.06364540 -0.080332229
[74,]  2.56222717  0.26019855 -3.374393962
[75,]  1.83256757 -1.28787820 -0.458280027
[76,] -0.86799290 -2.44410119  1.563333179
[77,]  0.37001440 -2.15390698  2.449386348
[78,] -1.45737704 -1.38335177  0.227306902
[79,]  1.26293085 -0.77084953  1.184224517
[80,]  0.37615037 -1.02704340 -1.794466295
[81,]  0.76206390 -3.37505381  0.357470056
[82,]  1.03457797 -1.45070974  0.363011773
[83,] -0.49487676 -2.38124353 -1.335743176
[84,] -2.53897708 -0.08744336 -0.474251393
[85,]  0.83532015 -1.47367055 -0.610093576
[86,]  0.78790461 -2.02662652  0.254723404
[87,] -0.80683216 -2.23383039 -0.772855797
[88,] -0.55804262 -2.37298543 -2.307611404
[89,] -1.11511104 -1.80224719 -0.959253308
[90,] -0.55572283 -2.65754004 -0.849126898
[91,] -1.34928528 -2.11800147  0.047652321
[92,] -1.56448261 -1.85221452 -0.781067031
[93,] -1.93255561 -1.55949546  0.089274676
[94,]  0.74666594 -2.31293171 -0.114679769
[95,]  0.95745536 -2.22352843 -0.142444774
[96,]  2.54386518  0.16927402 -0.788696991
[97,] -0.54395259 -0.36892655 -1.308895932
```

[98,]	1.03104975	-2.56556935	1.086390174
[99,]	2.25190942	-1.43274138	0.230208244
[100,]	1.41021602	-2.16619177	-0.748896411
[101,]	0.79771979	-2.37694880	1.568112531
[102,]	-0.54953173	-2.29312864	1.498935323
[103,]	-0.16117374	-1.16448332	-1.003713103
[104,]	-0.65979494	-2.67996119	0.764920868
[105,]	0.39235441	-2.09873171	0.471850008
[106,]	-1.77249908	-1.71728847	-0.947033174
[107,]	-0.36626736	-2.16935330	0.481324235
[108,]	-1.62067257	-1.35558339	-0.287159001
[109,]	0.08253578	-2.30623459	0.463574989
[110,]	1.57827507	-1.46203429	-1.779645955
[111,]	1.42056925	-1.41820664	-0.139275829
[112,]	-0.27870275	-1.93056809	-0.078670553
[113,]	-1.30314497	-0.76317231	-1.999596510
[114,]	-0.45707187	-2.26941561	-1.061338968
[115,]	-0.49418585	-1.93904505	-1.323938072
[116,]	0.48207441	-3.87178385	-1.344271223
[117,]	-0.25288888	-2.82149237	0.302639785
[118,]	-0.10722764	-1.92892204	-0.690148243
[119,]	-2.43301260	-1.25714104	1.903027404
[120,]	-0.55108954	-2.22216155	0.356228830
[121,]	0.73962193	-1.40895667	-1.125345492
[122,]	1.33632173	0.25333693	-5.345388179
[123,]	-1.17708700	-0.66396684	-3.010221888
[124,]	-0.46233501	-0.61828818	-0.483442366
[125,]	0.97847408	-1.44557050	-1.481236975
[126,]	-0.09680973	-2.10999799	-0.434826116
[127,]	0.03848715	-1.26676211	-0.687577913
[128,]	-1.59715850	-1.20814357	-3.361175555
[129,]	-0.47956492	-1.93884066	-1.296507519
[130,]	-1.79283347	-1.15028810	-0.782800173
[131,]	-1.32710166	0.17038923	1.180013355
[132,]	-2.38450083	0.37458261	0.723822595
[133,]	-2.93694010	0.26386183	0.167639816
[134,]	-2.14681113	0.36825495	0.453301301
[135,]	-2.36986949	-0.45963481	1.101399789
[136,]	-3.06384157	0.35341284	1.099124104
[137,]	-3.91575378	0.15458252	-0.221827800
[138,]	-3.93646339	0.65968723	-1.712215419
[139,]	-3.09427612	0.34884276	1.026831413
[140,]	-2.37447163	0.29198035	-1.241914333
[141,]	-2.77881295	0.28680487	-0.609670124
[142,]	-2.28656128	0.37250784	0.971643032
[143,]	-2.98563349	0.48921791	-0.946952932
[144,]	-2.37519470	0.48233372	0.252883994
[145,]	-2.20986553	1.16005250	1.245125226
[146,]	-2.62562100	0.56316076	0.855961082
[147,]	-4.28063878	0.64967096	1.458196962
[148,]	-3.58264137	1.27270275	0.110784038
[149,]	-2.80706372	1.57053379	0.472527935
[150,]	-2.89965933	2.04105701	0.495959810
[151,]	-2.32073698	2.35636608	-0.437681744
[152,]	-2.54983095	2.04528309	0.312267999
[153,]	-1.81254128	1.52764595	-1.362589782
[154,]	-2.76014464	2.13893235	0.964628688
[155,]	-2.73715050	0.40988627	1.190404684
[156,]	-3.60486887	1.80238422	0.094036861
[157,]	-2.88982600	1.92521861	0.782322556
[158,]	-3.39215608	1.31187639	-1.602025969
[159,]	-1.04818190	3.51508969	-1.160038566
[160,]	-1.60991228	2.40663816	-0.548559697
[161,]	-3.14313097	0.73816104	0.090998724
[162,]	-2.24015690	1.17546529	0.101376932
[163,]	-2.84767378	0.55604397	-0.804215218
[164,]	-2.59749706	0.69796554	0.884939521
[165,]	-2.94929937	1.55530896	0.983400727

```
[166,] -3.53003227  0.88252680  0.466029128
[167,] -2.40611054  2.59235618 -0.428226211
[168,] -2.92908473  1.27444695  1.213358272
[169,] -2.18141278  2.07753731 -0.763782552
[170,] -2.38092779  2.58866743 -1.418044029
[171,] -3.21161722 -0.25124910  0.847129152
[172,] -3.67791872  0.84774784  1.339420231
[173,] -2.46555580  2.19379830  0.918780960
[174,] -3.37052415  2.21628914  0.342569512
[175,] -2.60195585  1.75722935 -0.207581355
[176,] -2.67783946  2.76089913  0.940941877
[177,] -2.38701709  2.29734668  0.550696197
[178,] -3.20875816  2.76891957 -1.013913664
> mydata1<-cbind(wine,pcaobj$scores[,1:3])
> View(mydata1)
```

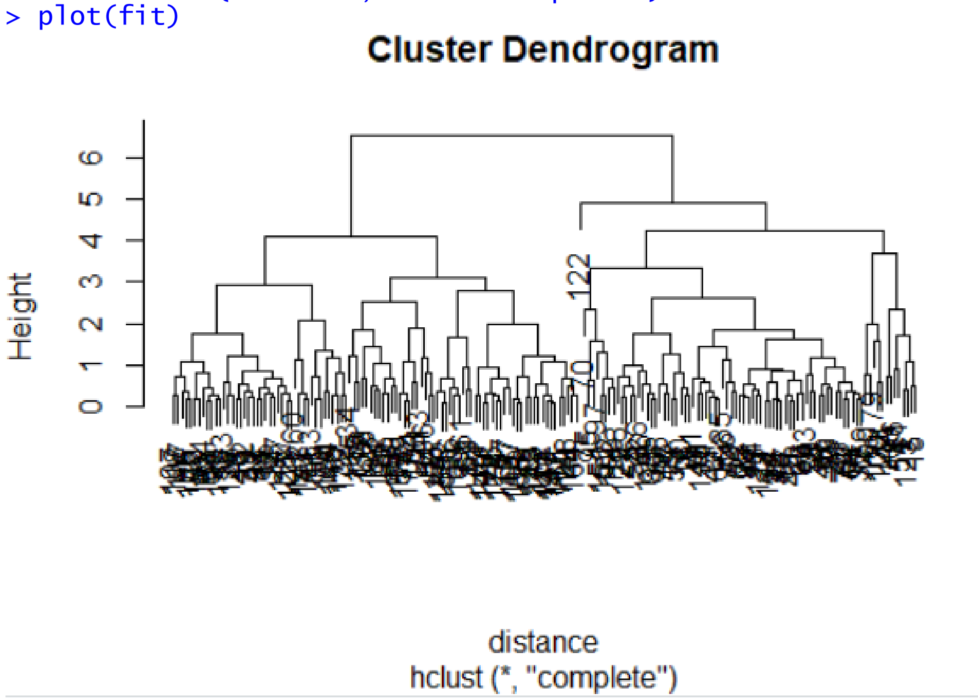
PCA.R × | mydata1 × | wine ×

Filter

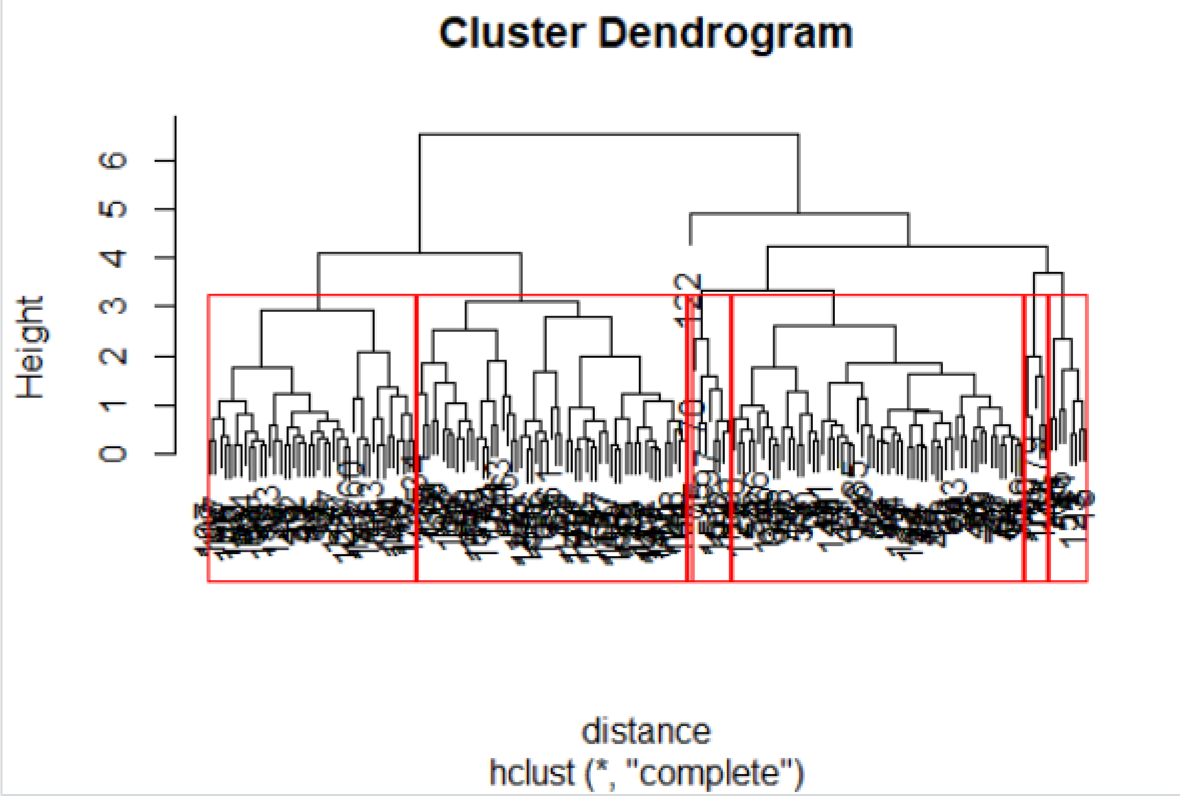
anoids	Proanthocyanins	Color	Hue	Dilution	Proline	Comp.1	Comp.2	Comp.3
	2.29	5.64	1.040	3.92	1065	3.31675081	1.44346263	0.165739045
	1.28	4.38	1.050	3.40	1050	2.20946492	-0.33339289	2.026457374
	2.81	5.68	1.030	3.17	1185	2.51674015	1.03115130	-0.982818670
	2.18	7.80	0.860	3.45	1480	3.75706561	2.75637191	0.176191842
	1.82	4.32	1.040	2.93	735	1.00890849	0.86983082	-2.026688219
	1.97	6.75	1.050	2.85	1450	3.05025392	2.12240111	0.629395827
	1.98	5.25	1.020	3.58	1290	2.44908967	1.17485013	0.977094891
	1.25	5.05	1.060	3.58	1295	2.05943687	1.60896307	-0.146281883
	1.98	5.20	1.080	2.85	1045	2.51087430	0.91807096	1.770969027

Showing 1 to 10 of 178 entries, 17 total columns

```
> clus_data<-wine[,8:10]
> # Normalizing the data
> norm_clus<-scale(clus_data)
> distance<-dist(norm_clus,method = "euclidean")
> fit<-hclust(distance,method="complete")
> plot(fit)
```



```
> rect.hclust(fit, k=7, border="red")
```




```
> groups <- cutree(fit,5)
> clust_1 <- as.matrix(groups)
> View(clust_1)
```

PCA.R | clust_1 | final

Filter

	V1
1	1
2	2
3	3
4	1
5	1
6	1
7	1
8	2
9	1
10	1
11	1

Showing 1 to 11 of 178 entries, 1 total columns

```
> final <- cbind(clust_1,mydata1)
```

```
> View(final)
```

PCA.R | final | clust_1 | mydata1 | wine

Filter

	clust_1	Type	Alcohol	Malic	Ash	Alcalinity	Magnesium	Phenols	Flavanoids
1	1	1	14.23	1.71	2.43	15.6	127	2.80	3.06
2	2	1	13.20	1.78	2.14	11.2	100	2.65	2.76
3	3	1	13.16	2.36	2.67	18.6	101	2.80	3.24
4	1	1	14.37	1.95	2.50	16.8	113	3.85	3.49
5	1	1	13.24	2.59	2.87	21.0	118	2.80	2.69
6	1	1	14.20	1.76	2.45	15.2	112	3.27	3.39
7	1	1	14.39	1.87	2.45	14.6	96	2.50	2.52
8	2	1	14.06	2.15	2.61	17.6	121	2.60	2.51
9	1	1	14.83	1.64	2.17	14.0	97	2.80	2.98

Showing 1 to 10 of 178 entries, 18 total columns

```
> aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = mean)
  Group.1 clust_1  Alcohol    Malic      Ash Alcalinity Magnesium  Phenols
1       1       1 13.33134 1.981343 2.419403   18.49851 105.02985 2.772687
2       2       2 12.77190 2.188571 2.238333   19.36190  95.19048 2.053095
3       3       3 12.90769 2.268462 2.256154   17.86154 102.07692 2.690000
4       4       4 12.82055 2.902909 2.410364   21.03273  95.87273 1.788727
5       5       5 11.56000 2.050000 3.230000   28.50000 119.00000 3.180000
  Flavanoids Nonflavanoids Proanthocyanins    Color       Hue Dilution
1   2.808657     0.2762687        1.913284 5.102687 1.0310448 3.058955
2   1.811905     0.3090476        1.223095 4.071190 0.9807143 2.634286
3   2.511538     0.3515385        2.751538 5.270000 0.9746154 2.756923
4   1.076364     0.5069091        1.199636 5.690182 0.8464727 1.995636
5   5.080000     0.4700000        1.870000 6.000000 0.9300000 3.690000
   Proline
1 901.6418
2 680.1190
3 786.3846
4 605.1636
5 465.0000
> aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = max)
  Group.1 clust_1 Alcohol Malic  Ash Alcalinity Magnesium Phenols Flavanoids
1       1       1   14.83  5.80 3.22       30.0       151    3.88       3.93
2       2       2   14.21  5.04 2.86       26.0       122    2.85       2.76
3       3       3   14.75  4.31 2.70       25.0       162    3.30       3.69
4       4       4   14.16  5.65 2.92       28.5       132    2.95       2.74
5       5       5   11.56  2.05 3.23       28.5       119    3.18       5.08
  Nonflavanoids Proanthocyanins Color  Hue Dilution Proline
1          0.47            2.50 10.80 1.36     4.00    1680
2          0.42            1.63  9.40 1.38     3.63    1295
3          0.53            3.58 13.00 1.42     3.64    1547
4          0.66            2.01 10.68 1.71     3.30    1235
5          0.47            1.87  6.00 0.93     3.69     465
> aggregate(final[,-c(2,16:18)],by=list(clust_1),FUN = min)
  Group.1 clust_1 Alcohol Malic  Ash Alcalinity Magnesium Phenols Flavanoids
1       1       1   11.45  0.89 1.70       13.2        78    1.40       0.99
2       2       2   11.65  0.90 1.36       10.6        70    0.98       0.34
3       3       3   11.46  0.99 1.82       11.4        82    1.90       1.10
4       4       4   11.03  0.74 1.98       16.0        80    1.10       0.47
5       5       5   11.56  2.05 3.23       28.5       119    3.18       5.08
  Nonflavanoids Proanthocyanins Color  Hue Dilution Proline
1          0.13            1.25  2.12 0.48     1.30     290
2          0.17            0.42  1.74 0.57     1.29     278
3          0.17            1.87  1.28 0.57     1.78     312
```

```
4           0.37            0.41  1.90 0.54      1.27      315
5           0.47            1.87  6.00 0.93      3.69      465
> write.csv(final,file = "data_clust.txt",row.names = F,col.names = F)
> write.csv(final,file = "data_clust.csv",row.names = F,col.names = F)
> getwd()
[1] "C:/Users/Arti Patel/Documents"
> # K-Means Clustering :
> library(plyr)
> mydata2 <- final
> normalized_data<-scale(mydata2[,15:17])
> kmeans_clust <- kmeans(normalized_data,7)
> str(kmeans_clust)
List of 9
 $ cluster     : int [1:178] 7 1 7 7 1 7 7 7 7 7 ...
 $ centers     : num [1:7, 1:3] 0.4252 -0.5863 0.0416 -0.3266 -0.8588 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:7] "1" "2" "3" "4" ...
  .. ..$ : chr [1:3] "Proline" "Comp.1" "Comp.2"
 $ totss       : num 531
 $ withinss    : num [1:7] 15.17 3.29 3.84 6.18 19.14 ...
 $ tot.withinss: num 70.6
 $ betweenss   : num 460
 $ size        : int [1:7] 30 20 10 21 43 18 36
 $ iter        : int 4
 $ ifault      : int 0
 - attr(*, "class")= chr "kmeans"
> final1<- cbind(kmeans_clust$cluster,mydata2)
> View(final1)
```

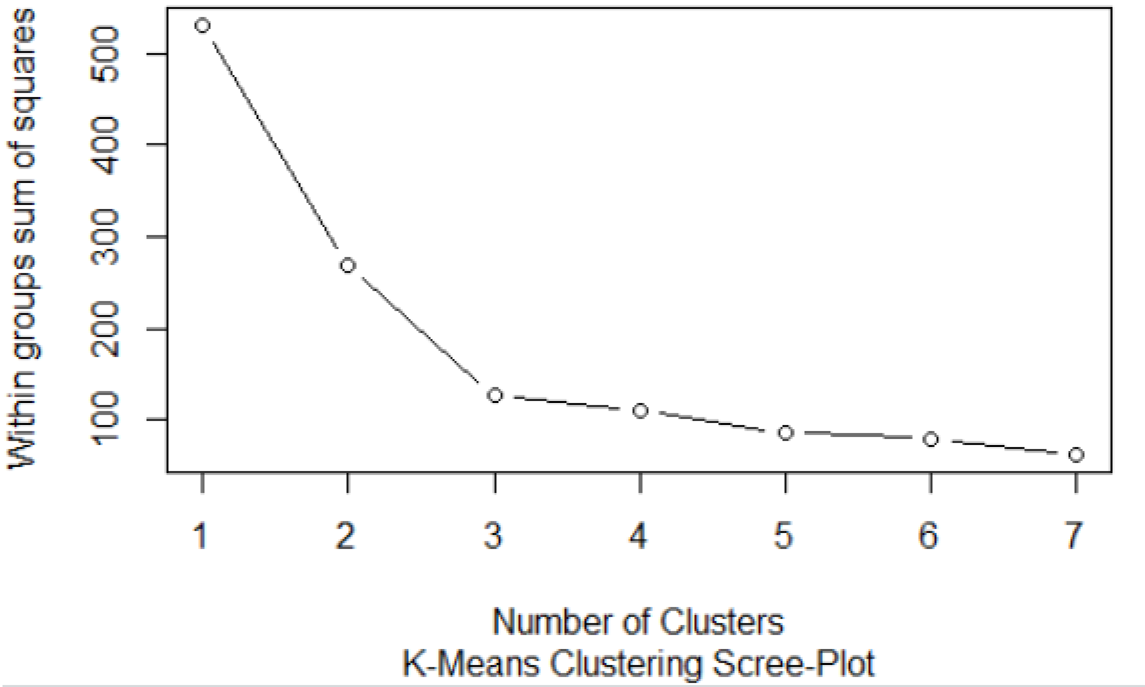
PCA.R | final1 | mydata | final | clust_1 | mydata1 | wine

Filter

	kmeans_clust$cluster	clust_1	Type	Alcohol	Malic	Ash	Alcalinity	Magnesium
1	7	1	1	14.23	1.71	2.43	15.6	127
2	1	2	1	13.20	1.78	2.14	11.2	100
3	7	3	1	13.16	2.36	2.67	18.6	101
4	7	1	1	14.37	1.95	2.50	16.8	113
5	1	1	1	13.24	2.59	2.87	21.0	118
6	7	1	1	14.20	1.76	2.45	15.2	112
7	7	1	1	14.39	1.87	2.45	14.6	96
8	7	2	1	14.06	2.15	2.61	17.6	121
9	7	1	1	14.83	1.64	2.17	14.0	97

Showing 1 to 10 of 178 entries, 19 total columns

```
> wss = (nrow(normalized_data)-1)*sum(apply(normalized_data, 2, var))      # D
etermine number of clusters by scree-plot

> for (i in 1:7) wss[i] = sum(kmeans(normalized_data, centers=i)$withinss)
> plot(1:7, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum
of squares")   # Look for an "elbow" in the scree plot #

> title(sub = "K-Means Clustering Scree-Plot")
```




```
> aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = mean)
  Group.1     Type  Alcohol    Malic      Ash Alcalinity Magnesium  Phenols
1       1 1.233333 13.18033 2.024000 2.439000   18.31333 111.76667 2.673333
2       2 2.950000 12.92300 3.535000 2.403000   21.30000  93.95000 1.582500
3       3 2.800000 12.88700 2.489000 2.337000   19.79000 103.30000 1.654000
4       4 3.000000 13.46333 3.373333 2.497143   22.09524 102.52381 1.793810
5       5 2.000000 12.21837 1.756047 2.162791   19.76047  89.04651 2.392326
6       6 2.000000 12.32389 2.532222 2.383333   21.06111  93.33333 1.872222
7       7 1.000000 13.92833 1.878611 2.452778   16.77778 106.30556 2.941667
  Flavanoids Nonflavanoids Proanthocyanins    Color
1  2.7346667     0.2903333        1.889333 4.356333
2  0.7145000     0.4795000        0.921500 5.811000
3  0.7820000     0.4350000        1.003000 5.224000
4  0.9195238     0.4319048        1.448095 9.408571
5  2.2176744     0.3251163        1.680698 2.932093
6  1.6838889     0.4622222        1.326667 3.113333
7  3.1133333     0.2886111        1.985556 6.152500
> aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = min)
  Group.1 Type Alcohol Malic  Ash Alcalinity Magnesium Phenols Flavanoids
1       1    1   11.56  0.99 1.75       11.2        90    1.85       1.28
2       2    2   12.20  1.29 2.10       18.5        80    0.98       0.34
3       3    2   12.25  0.94 2.15       17.0        85    1.10       0.50
4       4    3   12.79  1.67 2.26       18.5        88    1.35       0.61
5       5    2   11.03  0.74 1.36       10.6        70    1.45       0.57
6       6    2   11.64  1.10 1.98       16.0        80    1.38       0.99
7       7    1   13.05  1.35 2.04       11.4        89    2.20       2.19
  Nonflavanoids Proanthocyanins Color
1          0.13            1.28  2.60
2          0.17            0.55  3.85
3          0.21            0.42  3.05
4          0.22            0.97  7.10
5          0.17            0.42  1.28
6          0.14            0.41  2.08
7          0.17            1.25  3.95
> aggregate(mydata2[,2:12],by=list(kmeans_clust$cluster),FUN = max)
  Group.1 Type Alcohol Malic  Ash Alcalinity Magnesium Phenols Flavanoids
1       1    2   14.22  3.99 3.23       30.0       162    3.38       5.08
2       2    3   13.88  5.51 2.72       25.0       106    2.32       1.59
3       3    3   13.69  3.88 2.64       27.0       122    2.53       1.30
4       4    3   14.34  5.65 2.86       25.5       123    2.80       1.57
```

```
5       5    2   13.86  4.31 2.70       26.0       108    3.52        3.75
6       6    2   13.49  5.80 2.92       28.5       134    2.62        2.65
7       7    1   14.83  4.04 2.72       22.5       132    3.88        3.93
  Nonflavanoids Proanthocyanins     Color
1          0.47            3.28  6.130000
2          0.63            1.62  9.899999
3          0.63            1.46  8.210000
4          0.61            2.70 13.000000
5          0.52            3.58  5.300000
6          0.66            2.01  5.750000
7          0.50            2.96  8.900000
> kmeans_clust$centers
      Proline     Comp.1      Comp.2
1  0.42522567  0.8186532  0.09610543
2 -0.58634130 -1.4144988  0.28686687
3  0.04162093 -0.9852757  0.24524737
4 -0.32659002 -1.1790025  1.36464749
5 -0.85875428  0.1510432 -1.31632360
6 -0.81171614 -0.6227871 -0.78032720
7  1.58193227  1.1960417  0.85881203
> table(kmeans_clust$cluster)

 1  2  3  4  5  6  7 
30 20 10 21 43 18 36 
> write.csv(final1,file = "data_clust1.csv",row.names = F,col.names = F)
> getwd()
[1] "C:/Users/Arti Patel/Documents"
```