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

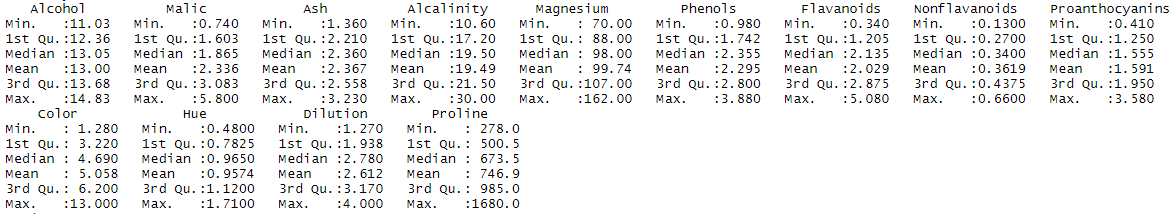
(class column we have ignored at the begining who shows it has 3 clusters)df

1.) Principal component Analysis

Datasets :- wine.csv

After removing the ”type variable”, there are 13 variables

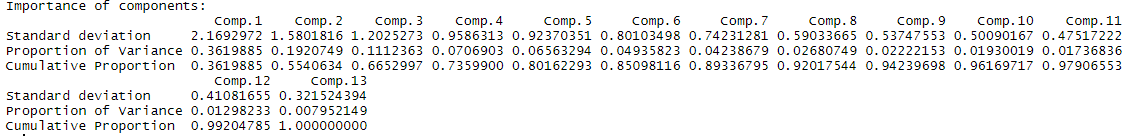
Summary



Applying PCA

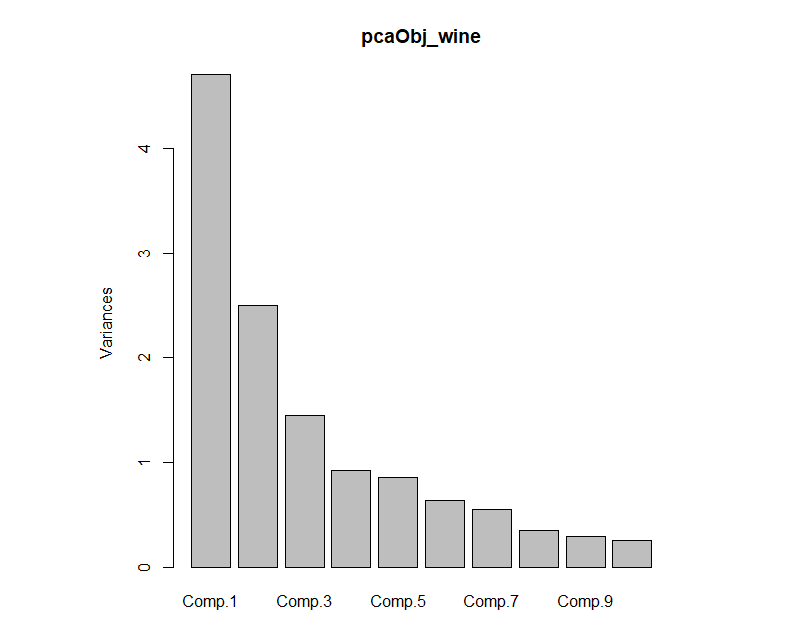
pcaObj\_wine<-princomp(wine\_data,cor = TRUE,scores = TRUE, covmat=NULL)

**Summary**



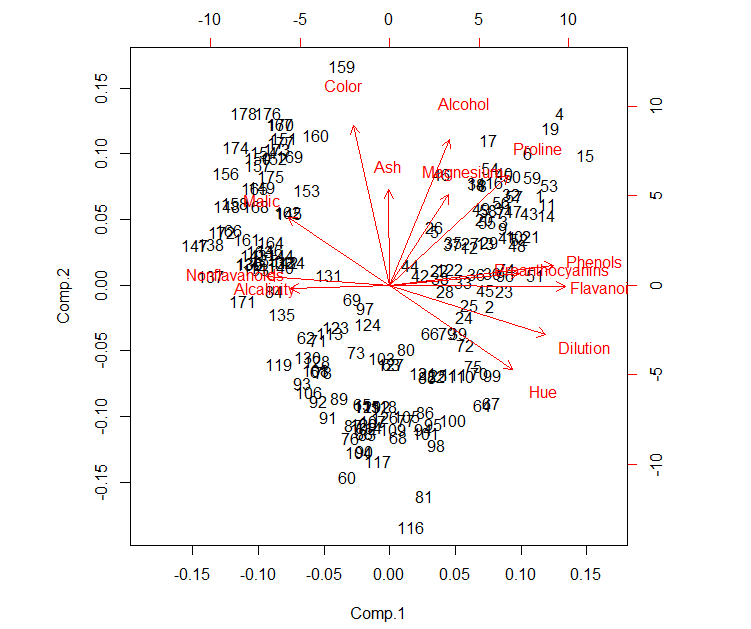
Now we can see that from summary, 95 percent data is captured by 10 components from the cumulative proportion value. Variance is highest in the 1st component and reduces in descending order as we move from 1st component to 13th component.

**Graphical representation**



From the above plot, it can be said that the Comp 1 covers maximum information.

Biplot



Considering first three components and bind with the original wine data

Clustering

1.) Hierarchical

Considering first three components as it represents the entire data

Applied Scale function for Normalization

**Summary of normalized data**

Comp.1 Comp.2 Comp.3

Min. :-1.96773 Min. :-2.4433 Min. :-4.4326

1st Qu.:-0.99878 1st Qu.:-0.8824 1st Qu.:-0.6303

Median : 0.02782 Median : 0.1654 Median : 0.1174

Mean : 0.00000 Mean : 0.0000 Mean : 0.0000

3rd Qu.: 0.91996 3rd Qu.: 0.7785 3rd Qu.: 0.6893

Max. : 1.98251 Max. : 2.2182 Max. : 3.8021

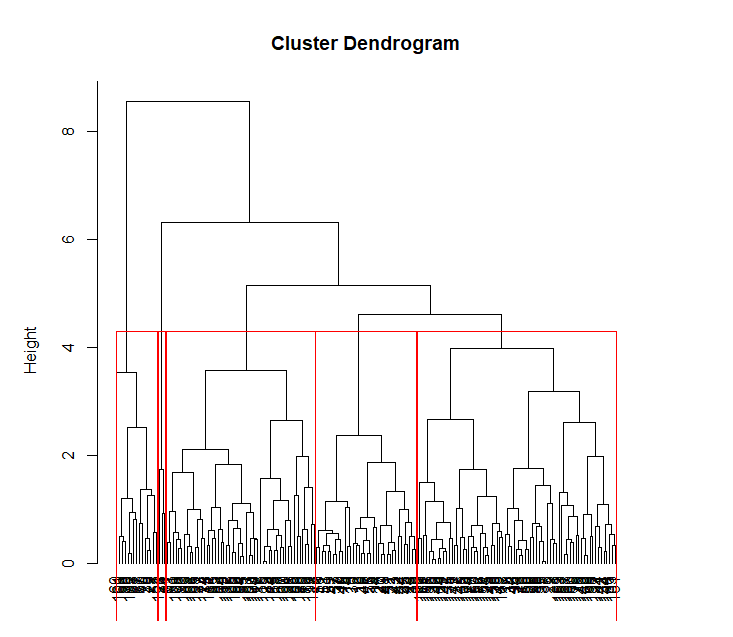
**Finding Distance**

Euclidean Distance is calculated

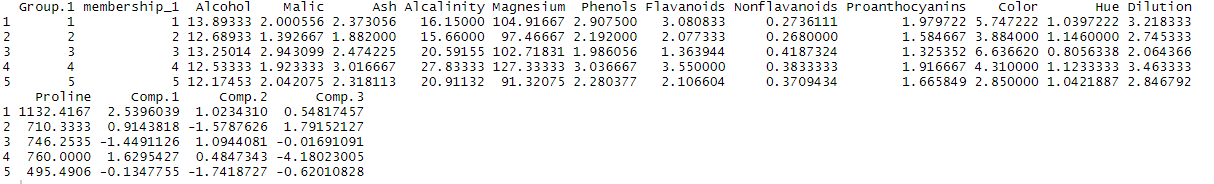
For clustering complete linkage method is used.

**Dendogram**

Cutting into five clusters



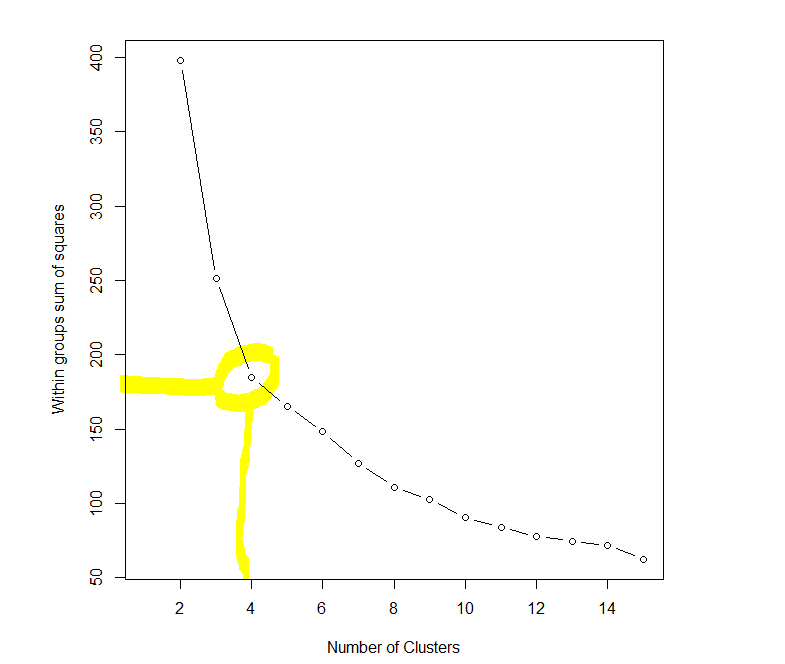
Aggregation: this will give the mean value of the variables as per cluster.



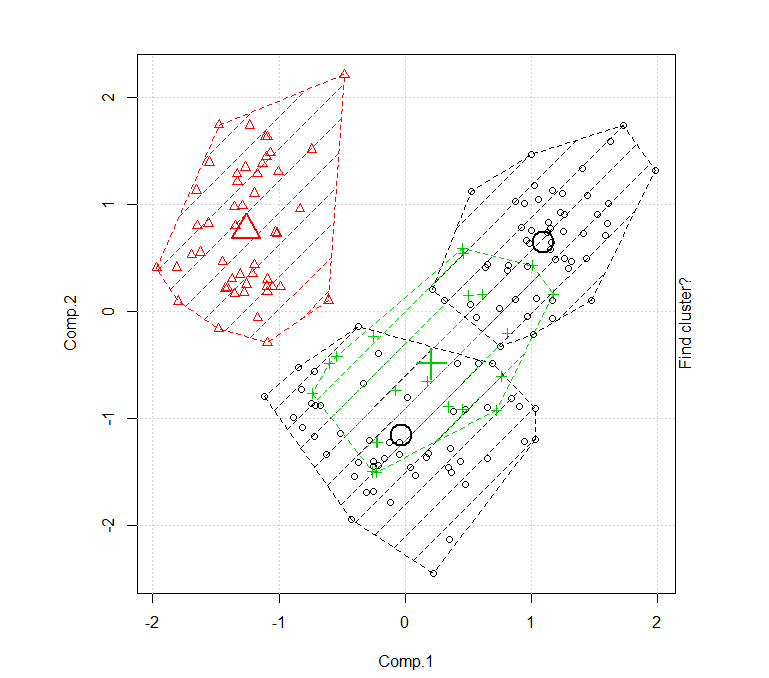
From the above aggregation, wine belong to group 1 are strong alcohol and wine belong to group 5 have mild in nature

**K-means**

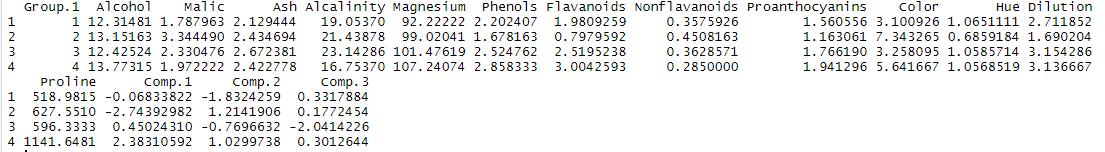
Scree Plot/Elbow curve



From the above plot, optimal number of cluster should be 4

Now applying k-means at 4 cluster 

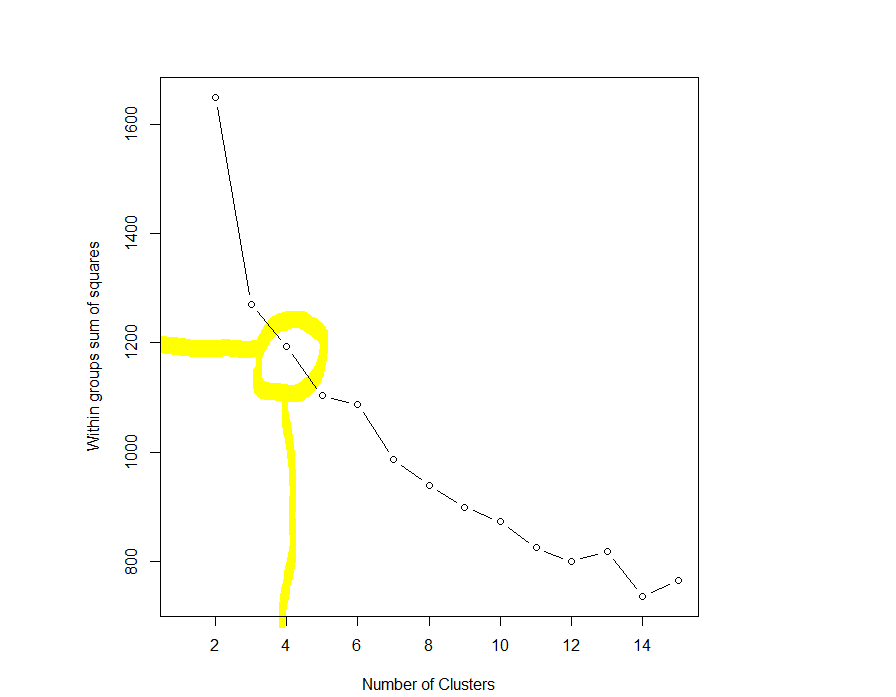
Aggregation of 4 clusters



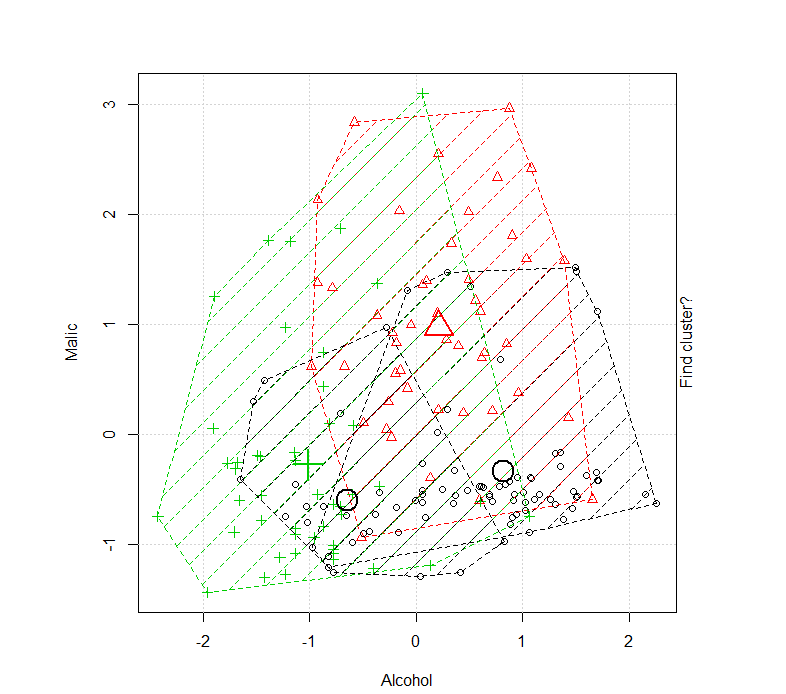
In this case wine belong to group 4 are strong in nature and wine belong to group 2 are mild in nature.

**Clustering with original data**

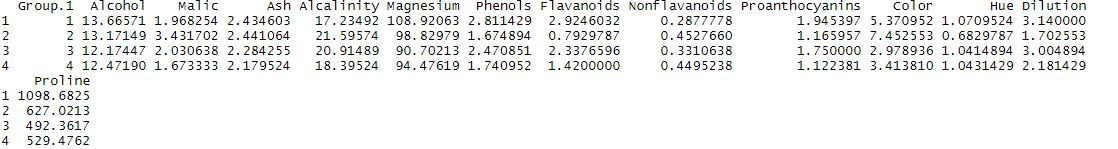
**Scree plot and k-means**



**Here the optimal clusters should be 4**

**Applying k-means**

In the original data sets, clusters are different



Here the wine belongs to group 1 are strong and wine belongs to group 4 are mild.

We have attained the same number of clusters through scree plot of Principal components and also through original data.