# MODULE 22

1. **WINE Data**

**Package used in python** :

Pandas : Used for data manipulation

Numpy : Scientific Calculating

Matplotlib : Used to create plot for dendrogram

Sklearn

PCA: To Run the PCA algorithm

Scale: Use the scale function to normalize the data

KMeans : Used to run K Mean algorithm

AgglomerativeClustering : Used to run hierarchical Cluster

Scipy

cluster.hierarchy : Used to create cluster on dendrogram

Linkage : Used as a linkage function for the dendrogram

**Loading the data**

Loading the Wine dataset in R and Python.

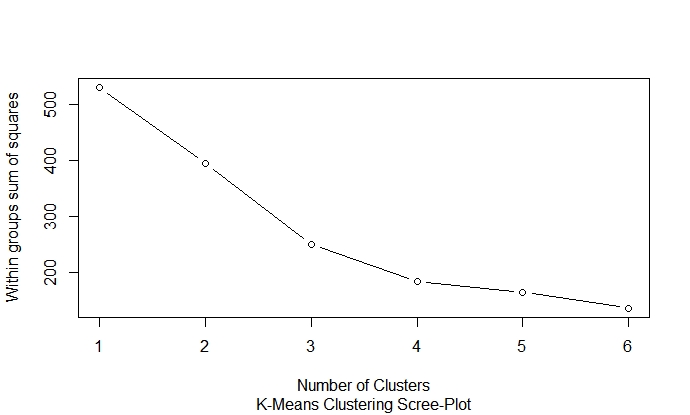
**PCA**

Ran the algorithm on the dataset and created component of data.

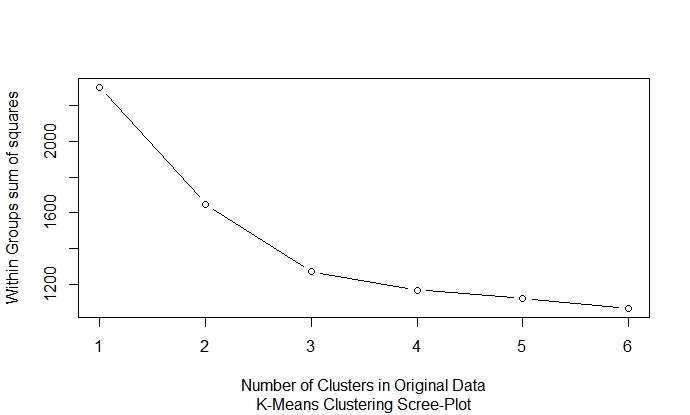
Selected First 3 scores which contains 66% of the dataset and ran K Mean and Hierarchical Cluster to check whether the top 3 contains same amount of data as the Original Data

**K Mean**

In PCA Component Data K Mean Scree Plot Elbow Like Curve is found at 3rd point of the plot



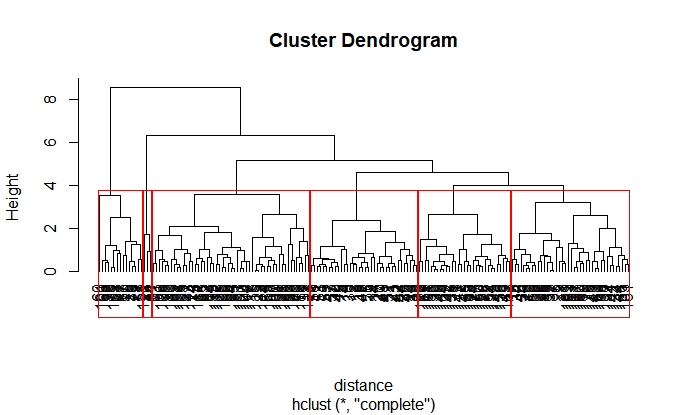
In Original Data K Mean Scree Plot Elbow like Curve is found in the 3rd point of the curve



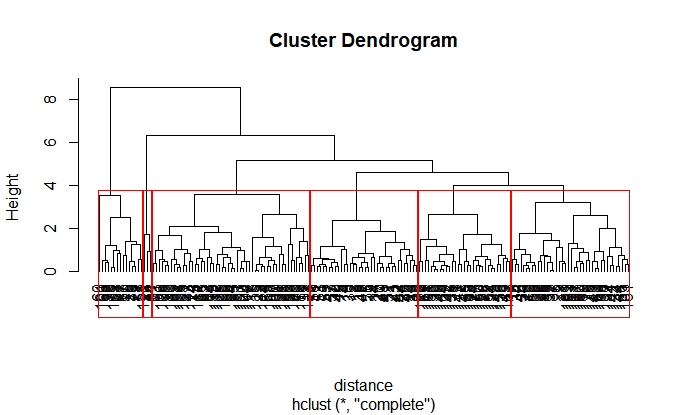
**Both the PCA Component Data and Original Wine Data Elbow Curve found at 3rd Point of the curve.**

**Hierarchical Cluster**

**Created Dendrogram on the PCA Component data with 6 Clusters with euclidean Distance method**

****

Created Dendrogram on the Original Wine data with 6 Clusters with euclidean Distance method



**Both the PCA Component Data and Original Wine Data has same type of clusters.**