Report for Plots for KMeans + Hierarchical Clustering

Dataset : mtcars

Clustering Type : KMeans+ Hierarchical Clustering

Motive: To check the clustering plots and dendograms of different linkages in KMeans Clustering

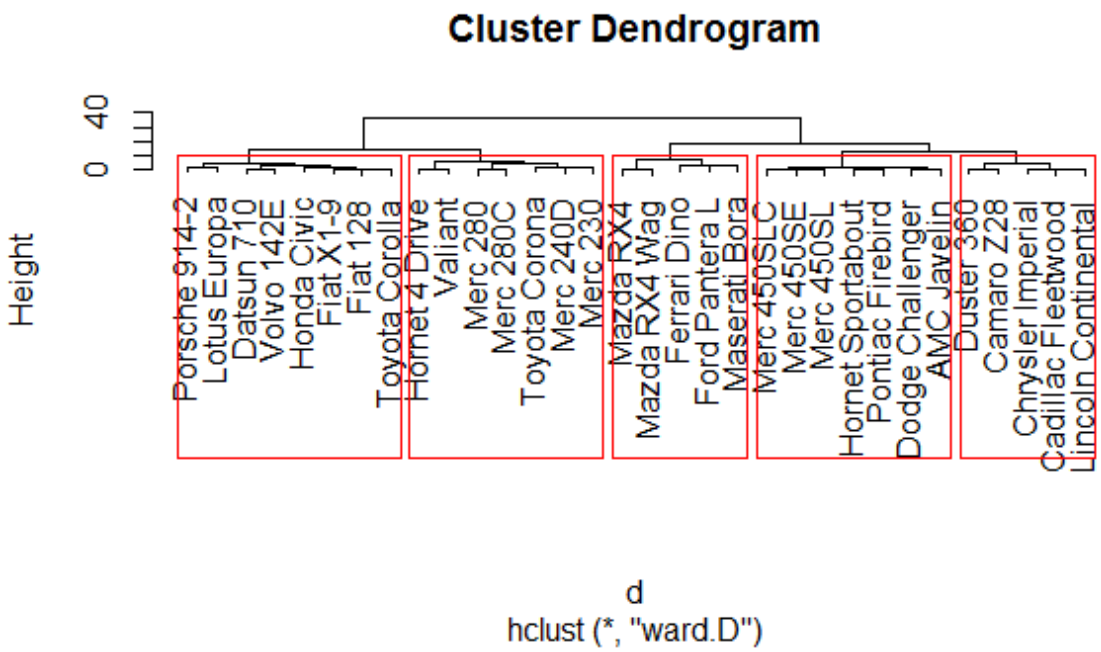
Observation :

R-Studio

The given mtcars K-Means + Hierarchical clustering shows the below result in R :

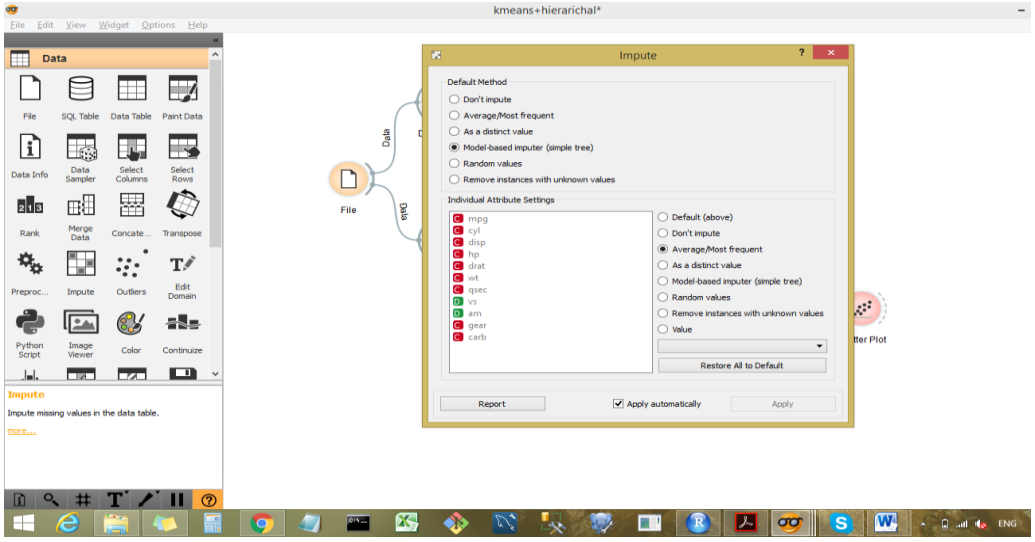
Plot for Mapping Sum of Squares within the cluster against Number of clusters

Dendogram for Hierarichical Clustering of the data with output showing car names for linkage type Ward



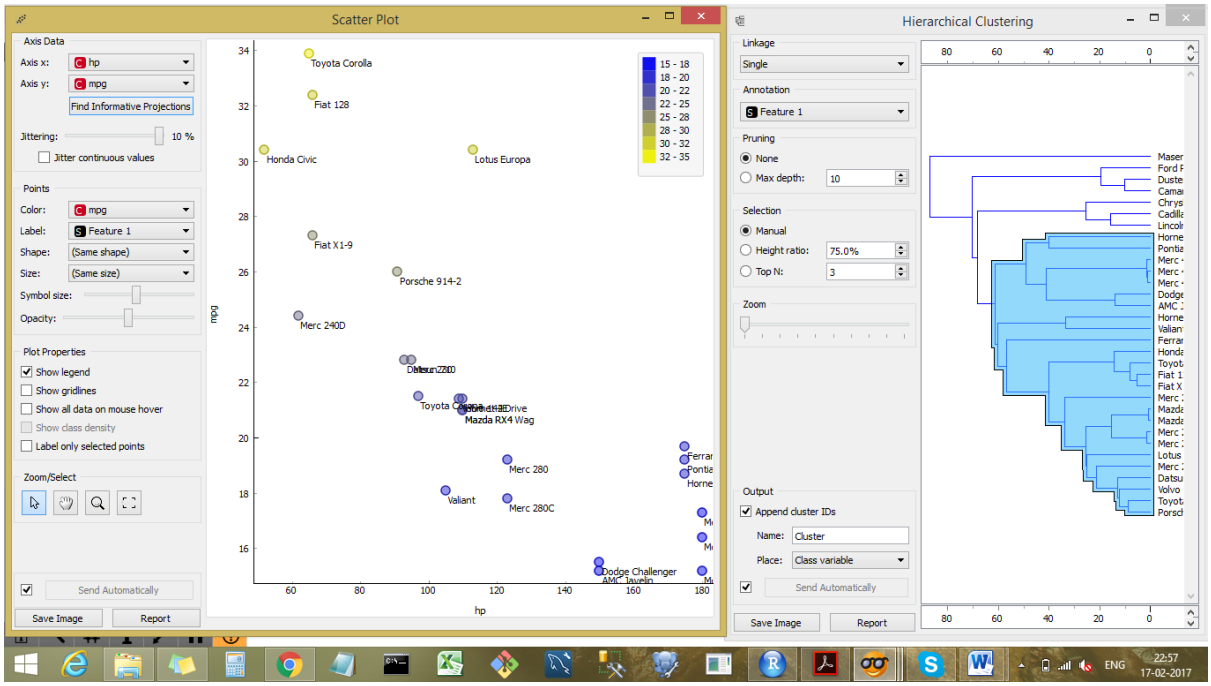
Orange

The given example shows data mapping for Hierarchical and K-Means in one work flow. The file contains missing values for some independent variables. Inorder to extract a true analysis of the clustering of data, we compute the missing values by using the widget ‘Impute’. This widget provides us with a choice of method to input values these missing variables.

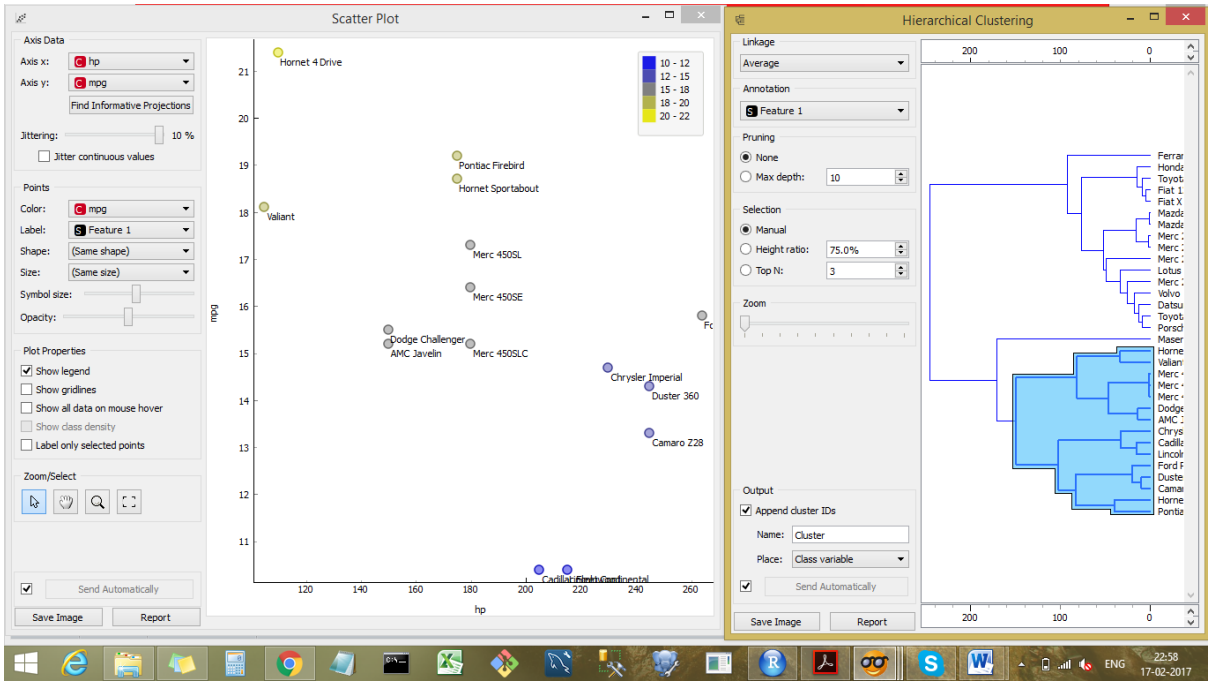


The mtcars dataset consists of categorical values which need to be converted here into continuous values by using the widget "Continuize" in order to display the data in Hierarchal cluster and its types of linkages :

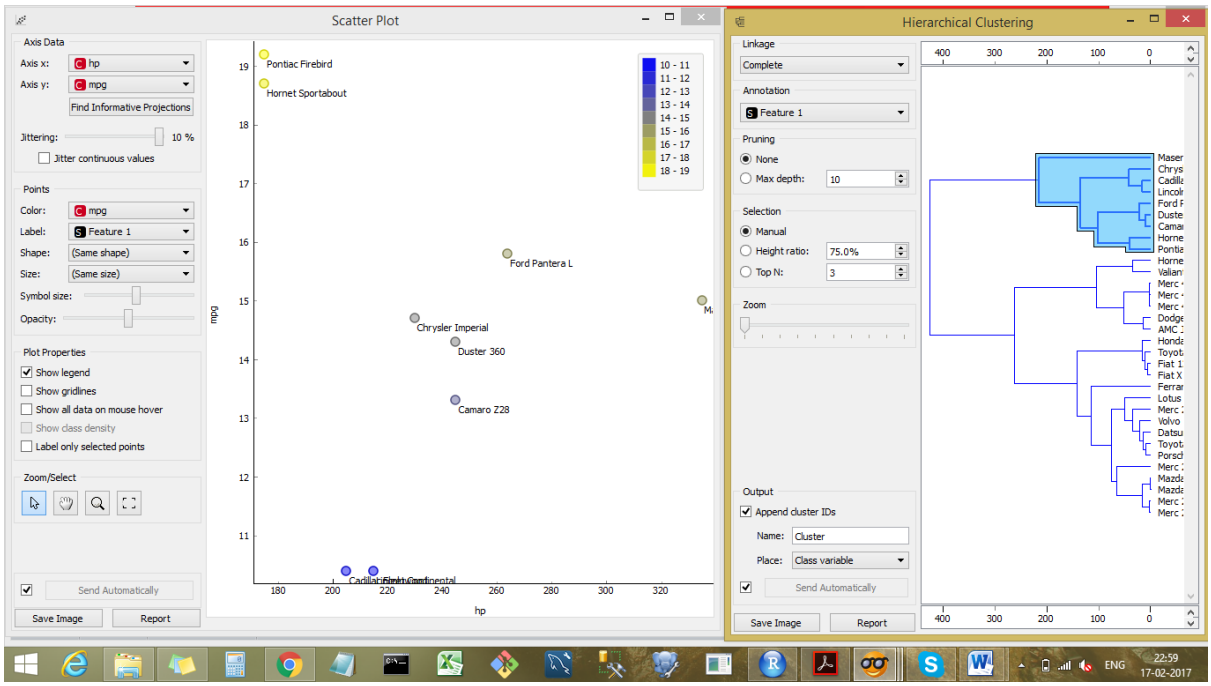
Dendogram Single Linkage Hierarchal Cluster



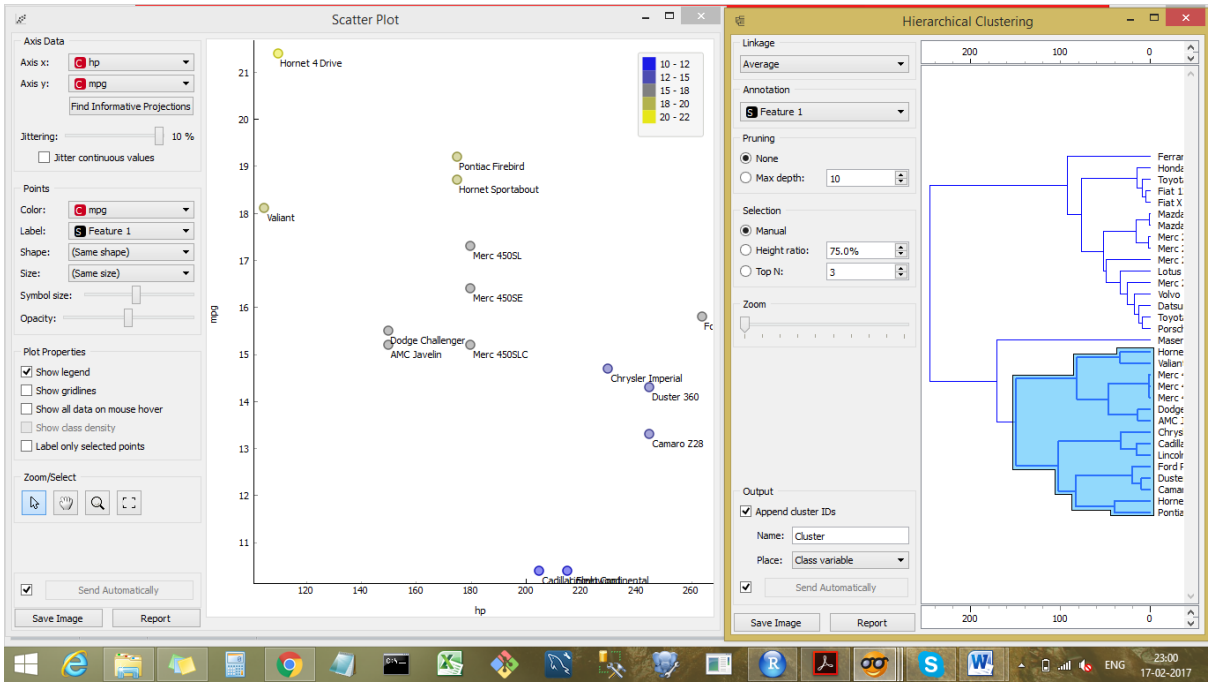
Dendogram Weighted Linkage Hierarchal Cluster



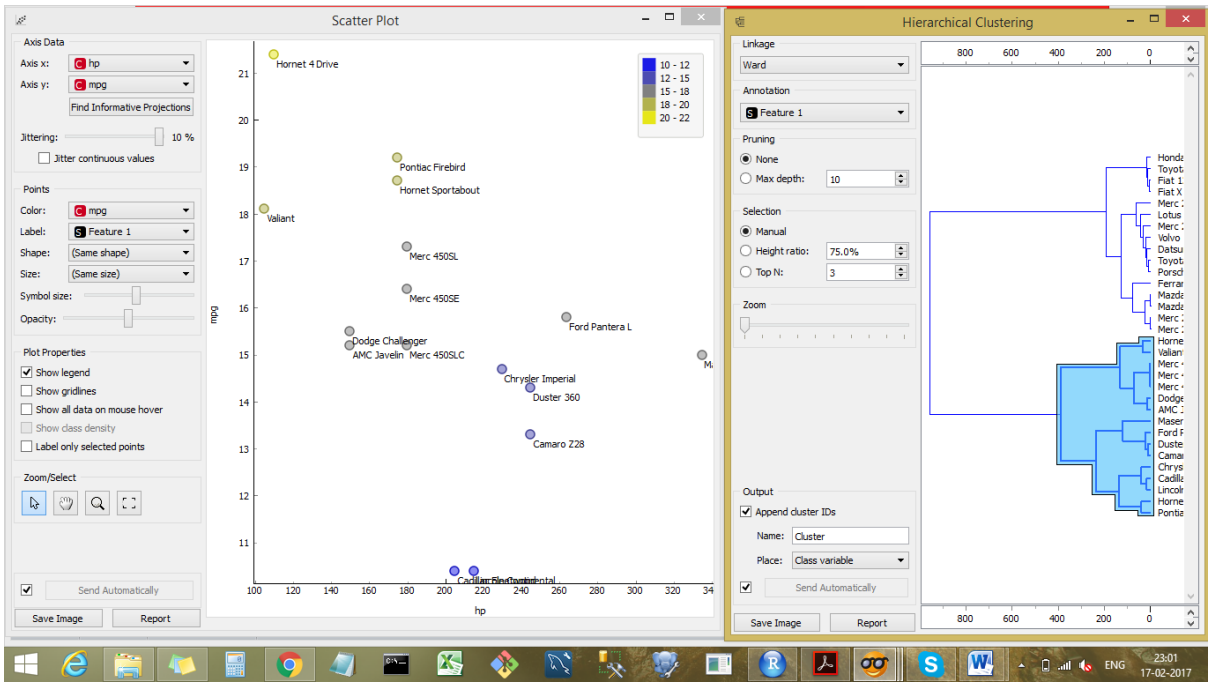
Dendogram Complete Linkage Hierarchal Cluster



Dendogram Average Linkage Hierarchal Cluster



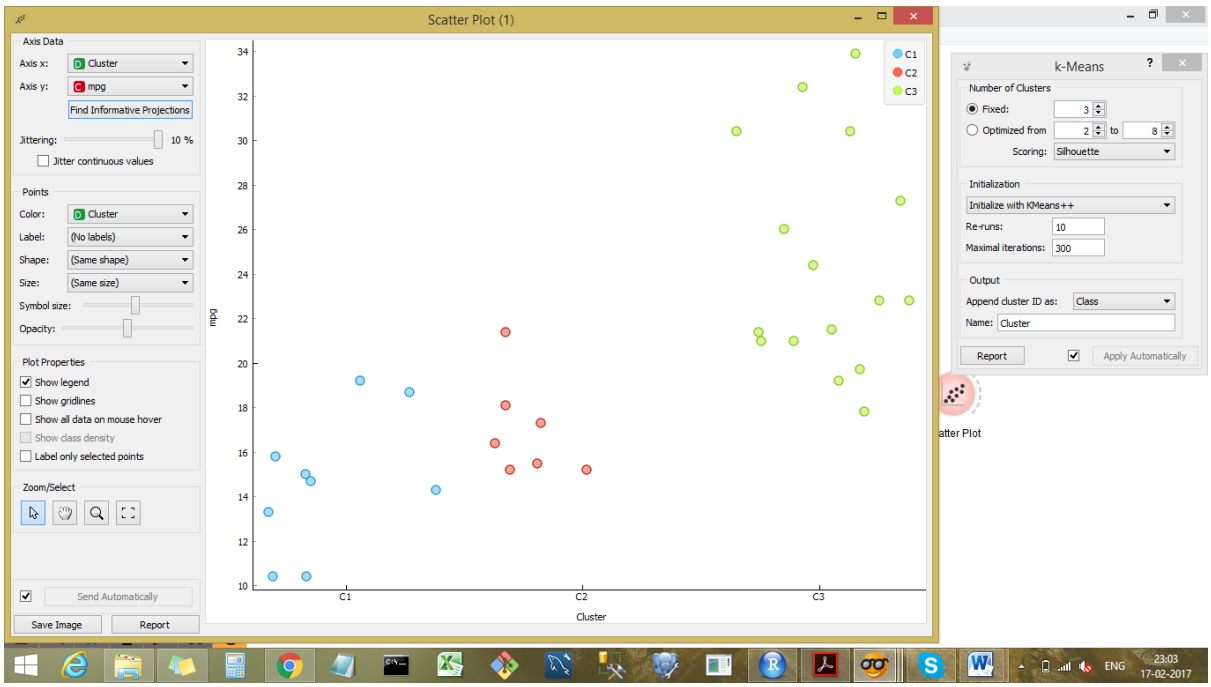
Dendogram Ward Linkage Hierarchal Cluster



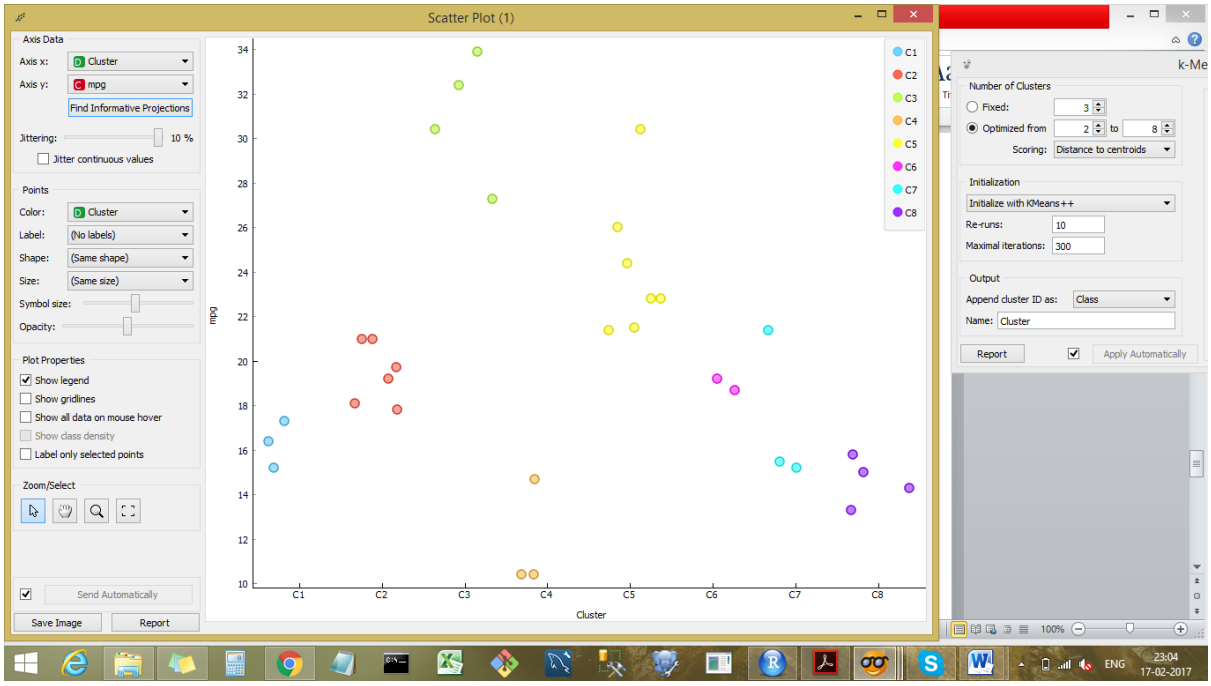
For KMeans Clustering of the data -

We are able to control the plot of the independent variable by choosing between fixed number of clusters or a range in the K-Means cluster widget. We are able to see a view a variety of plots depending for different factors that we choose in the K-Means widget. Below are some of the examples :

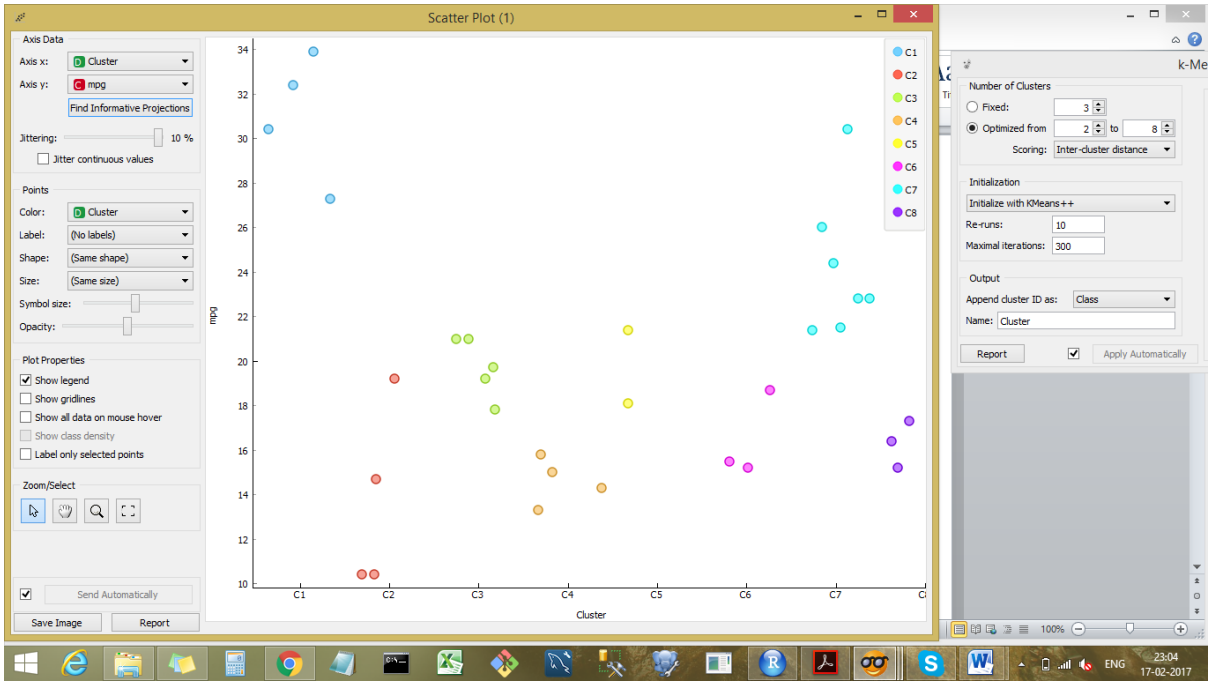
Output Plot for KMeans when the number of cluster is fixed=3



Output Plot for KMeans for the Iris when the number of cluster is ranging from 2 to 8 and within the cluster



Output Plot for KMeans for the Iris when the number of cluster is ranging from 2 to 8 and between the cluster



Conclusion :

The plots viewed on each platforms (R-studio and Orange) show that behaviour of the different linkage types in Hierarchal Clustering against the selected data from their respective dendograms remain same. These plots viewed on these platforms (R-studio and Orange) show that there are factors which affect the behaviour of the KMeans factor. Some of them are whether we choose to keep the clusters fixed or keep a range. Keeping a range allows us the flexibility of even further get into details by choosing between inter-cluster graph or within the cluster (as shown above).