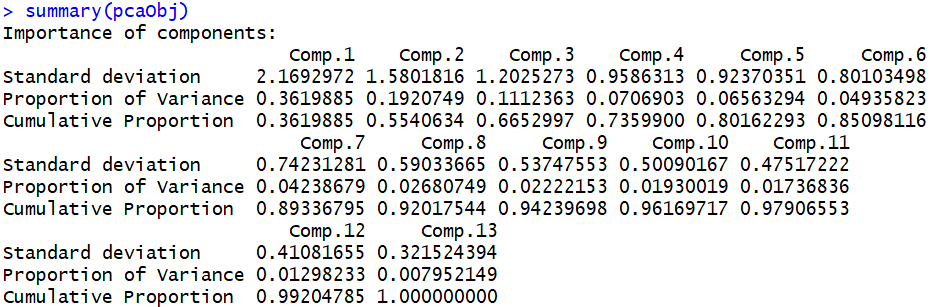
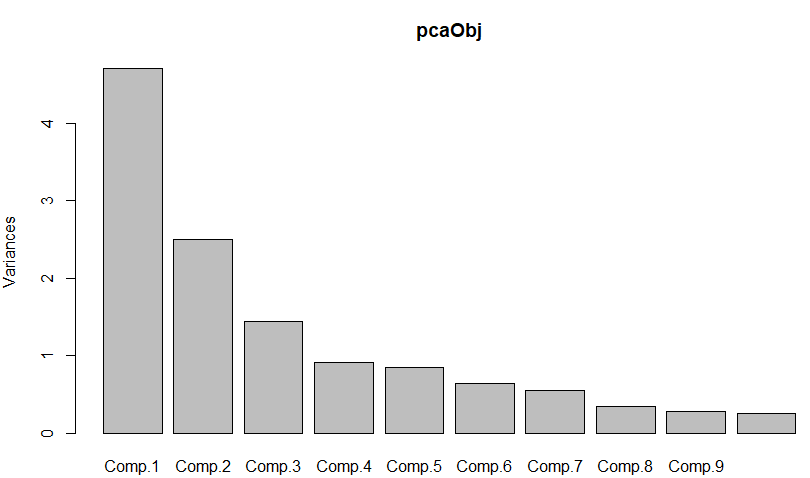
**Wine data from the UCI machine learning repository**

**Do a Principal Components Analysis (PCA) on the data.**



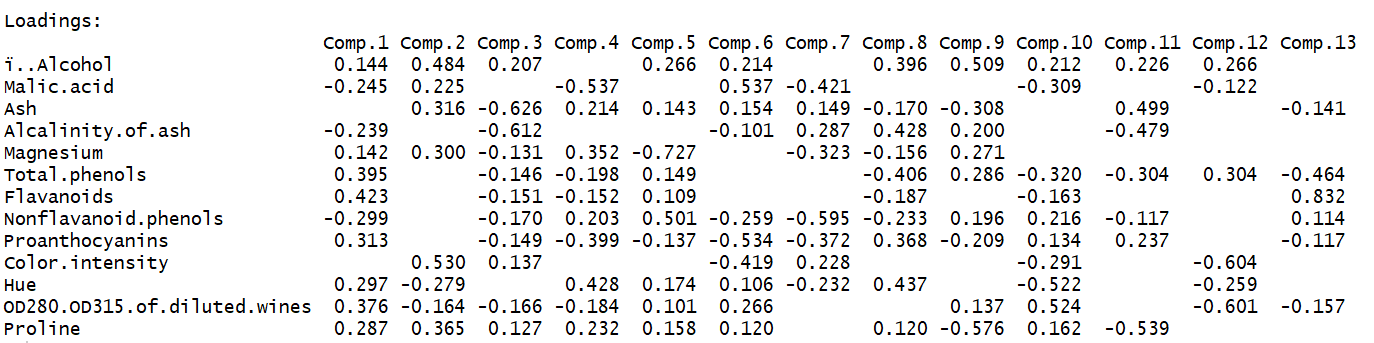
Component1 i.e. captures high variance as per PCA design, subsequently variance reduces for further components

1. **Enumerate the insights you gathered during your PCA exercise. Please do not clutter your report with too many insignificant insights as it will dilute the value of your other significant findings.**

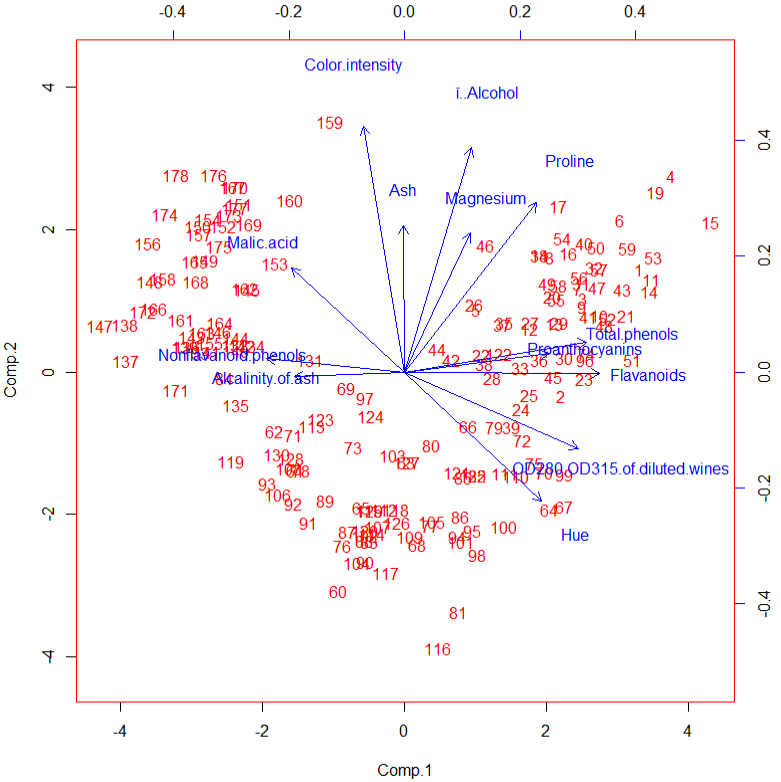


Above scree plot confirms Principal Component 1 captures most of the variance in the data and subsequently other columns variance in decrease in order.

Almost 80 – 85% of data can be recovered from first 5-6 columns, which is also confirmed from summary table.



**Biplot:**

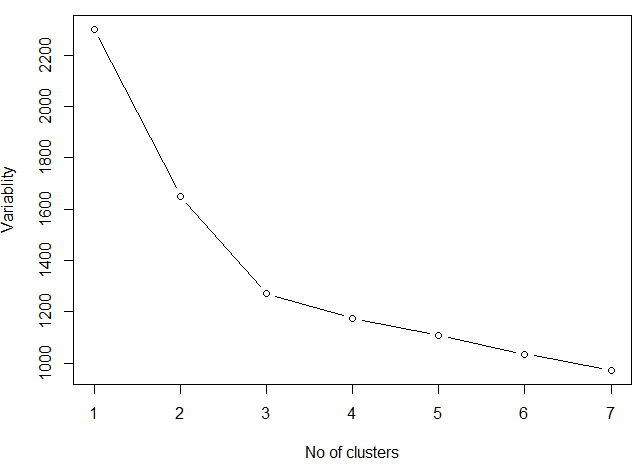


1. **What are the social and/or business values of those insights, and how the value of those insights can be harnessed―enumerate actionable recommendations for the identified stakeholder in this analysis?**

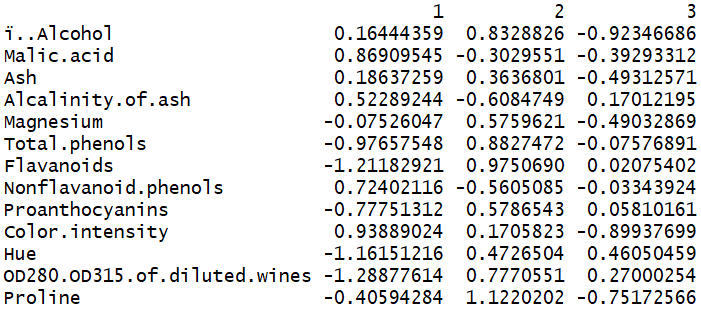
Loading table which have correlation >=0.5 will be considered as important. In such way, we can conclude that

1. Alcohol, Magnesium & Proline are positive correlated between comp1 & comp2.
2. Malic Acid, Nonflavanoid phenols are negatively correlated.
3. Hue & Malic Acid are dissimilar contents
4. **Do a cluster analysis―you may try different algorithms or approaches and go with the one that you find most appropriate**

**Using All Chemical Measurements**



From Elbow plot, we can see k=3 will be the best fit. Hence will proceed K-Means clustering with K=3.



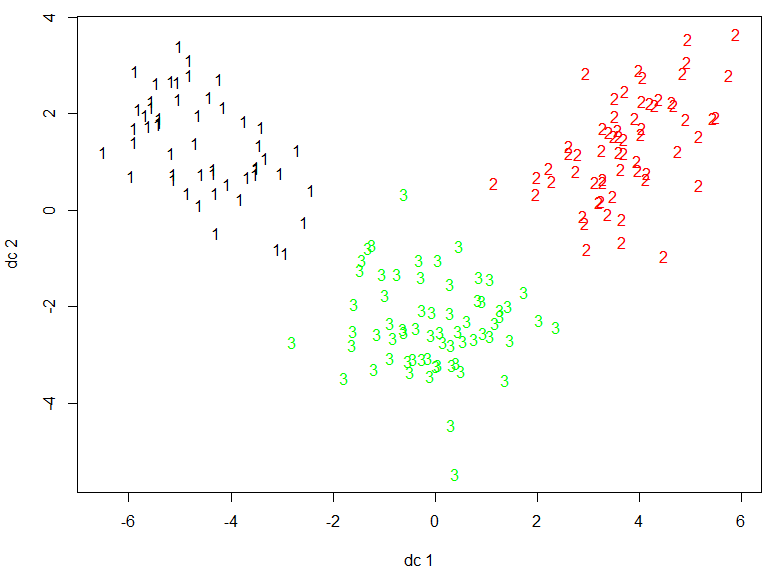
Cluster2 have very high Alcohol, Magnesium, Phenol contents were very high. Hence this type of Beverage is bad for health

Cluster3 have vert less Alcohol content and OD280.OD315.of.diluted.wines is positive, hence this type of beverages are little safe to consume.

Cluster1 will fall into intermediate category with medium Alcoholic contents in addition with good malic acid contents.

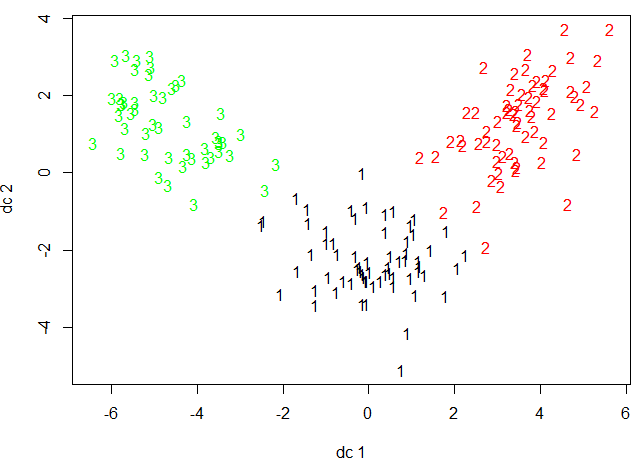
**Are there clearly separable clusters of wines? How many clusters did you go with? How the clusters obtained in part (i) are different from or similar to clusters obtained in part (ii), qualitatively**?

From the plot, it is very clear that wine data is separable with 3 clusters as best fit.



**using two most significant PC scores**

We can similar 3 clusters by using only 2 most significant PC scores from K-Means algorithm



**Could you suggest a subset of the chemical measurements that can separate wines more distinctly? How did you go about choosing that subset? How do the rest of the measurements that were not included while clustering, vary across those clusters?**

From Biplot & loadings table, we can identify that Alcohol, OD280/OD315 of diluted wines mainly differentiates clusters groups whether high alcoholic content beverages or low alcoholic content beverages. Also Malic acid, Phenols & magnesium plays its part to determine medium category cluster.