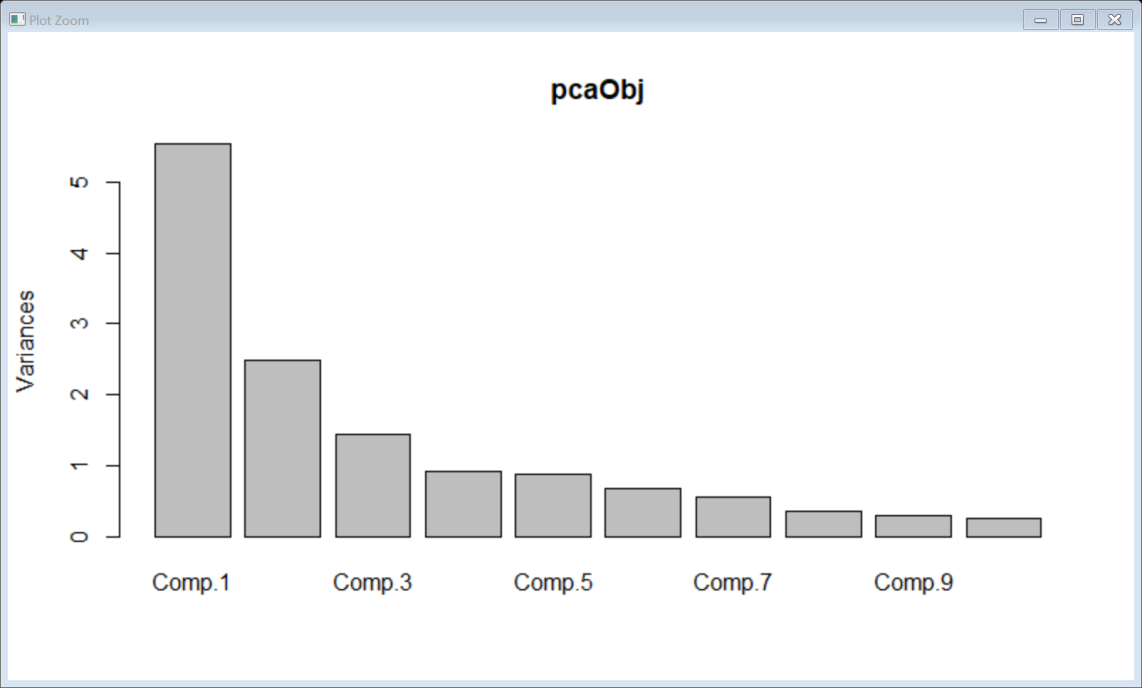
# Principle Component Analysis: PCA

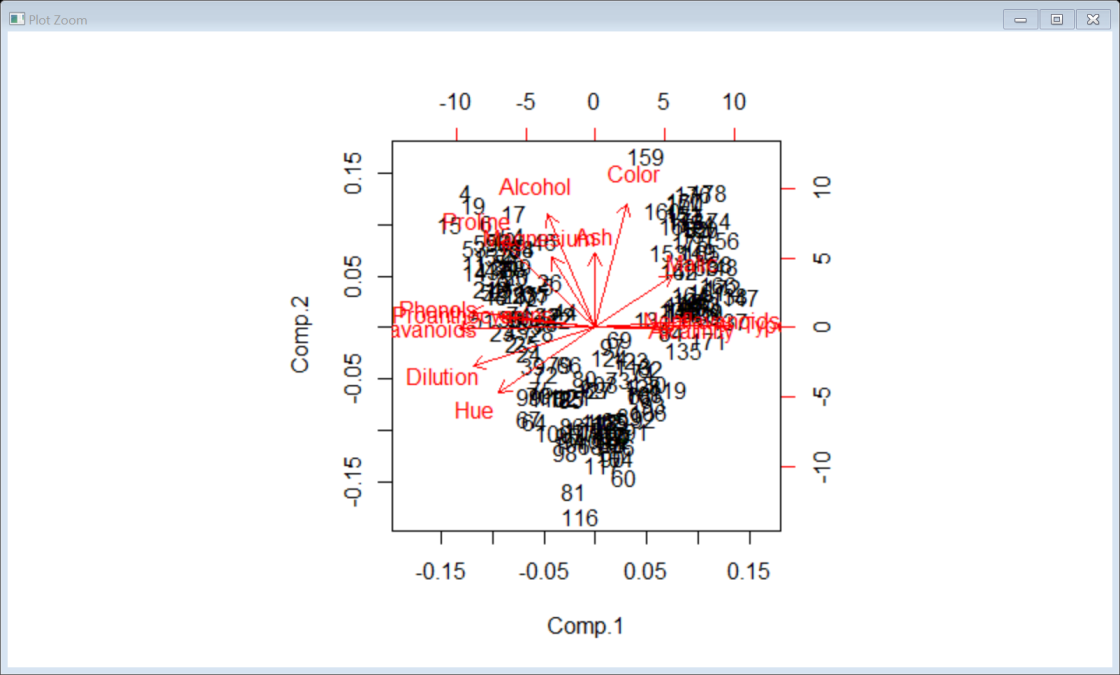
**Problem Statement**: Perform Principal component analysis and perform clustering using first 3 principal component scores

Step 1: princomp performs a principal components analysis on the given numeric data matrix and returns the results as an object of class princomp.

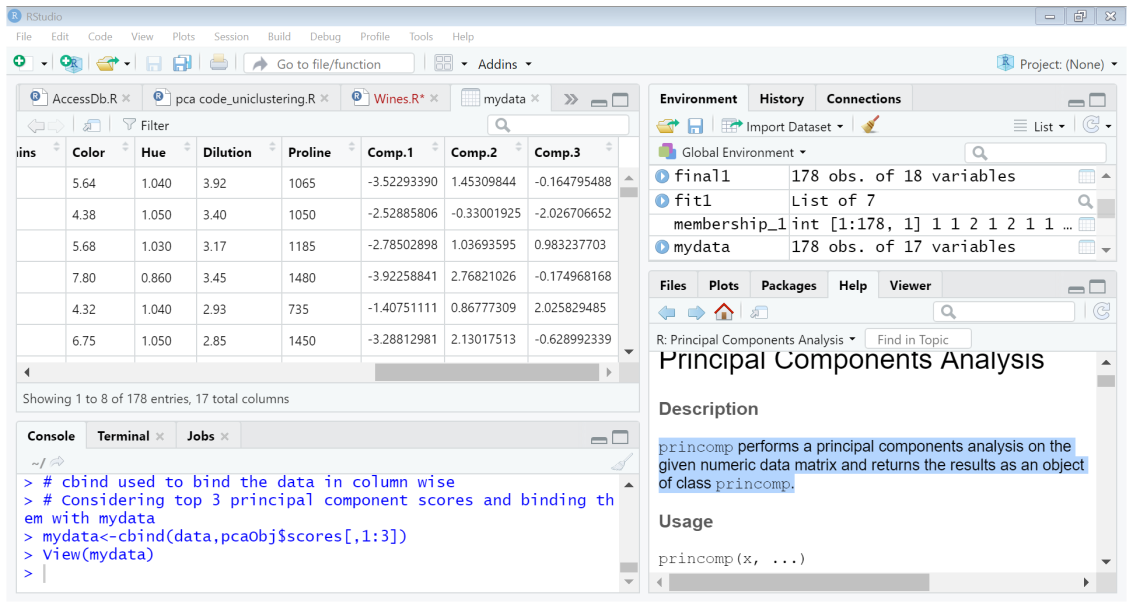
Step 2: Below histogram graph showing importance of principal components . Comp.1 having highest importance (highest variance)



Step 3:Create a Biplot on the same



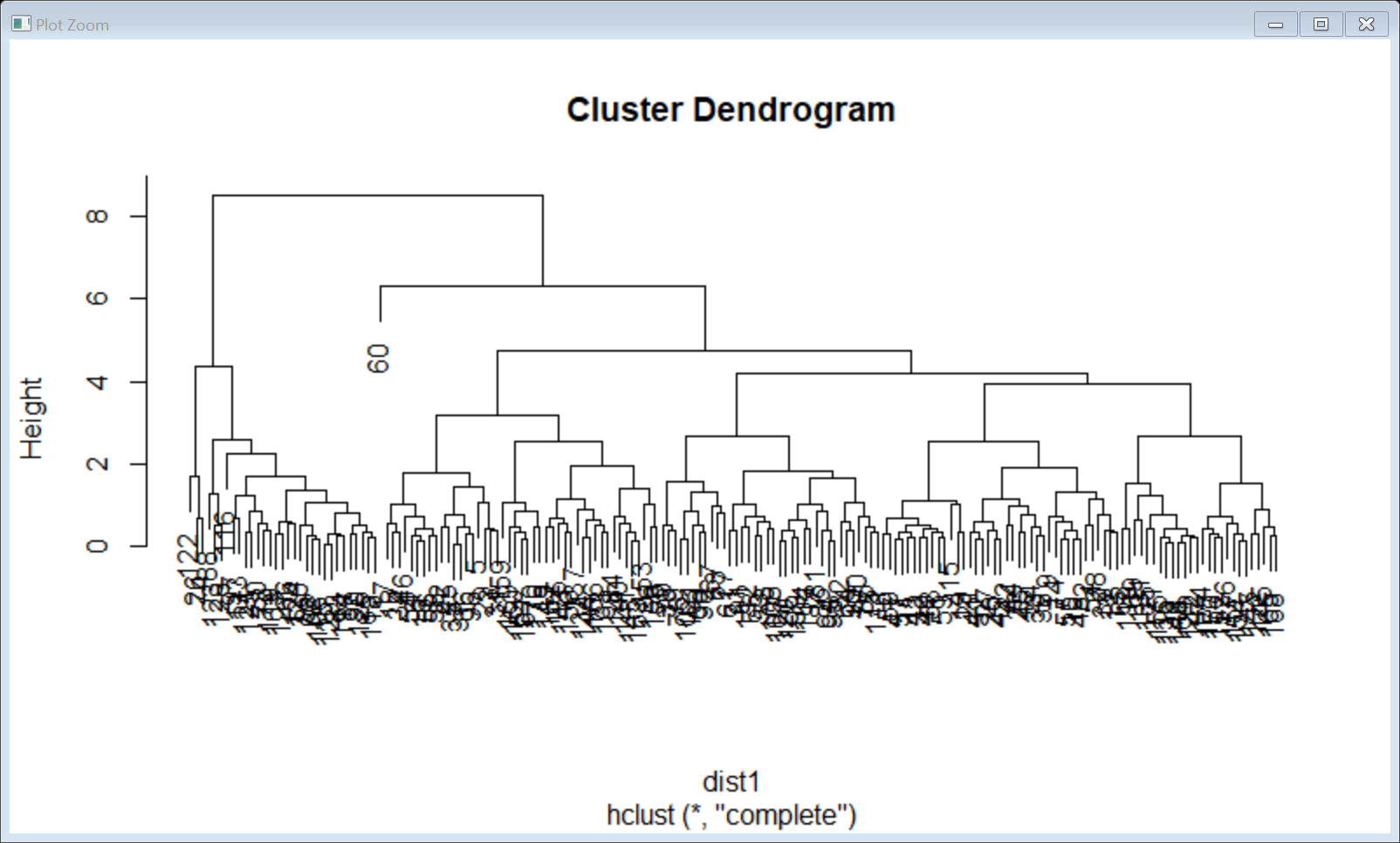
Step 4: Top 3 PCA Scores which represents the whole data .Considering top 3 principal component scores and binding them with mydata



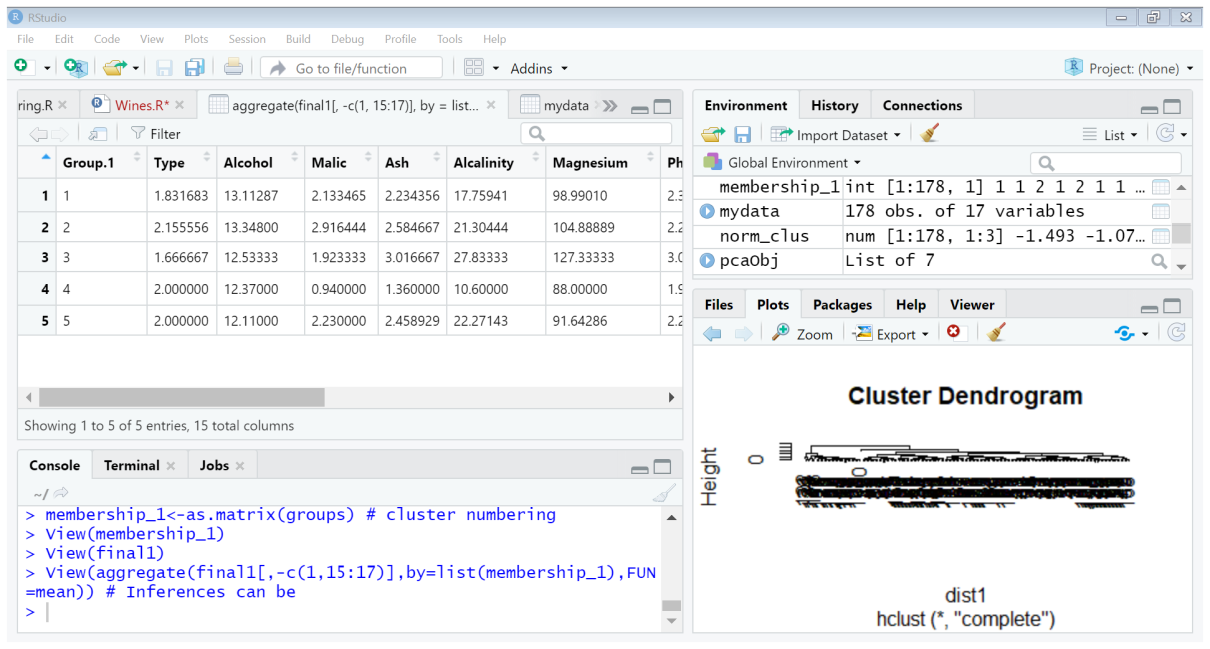
Step 5: Preparing data for clustering considering only pca scores as they represent the entire data.

Step 6: Use Scale function to normalize data use Euclidean distance method.Use complete linkage function for clustering.

Step 7: Display using Dendrogram below.

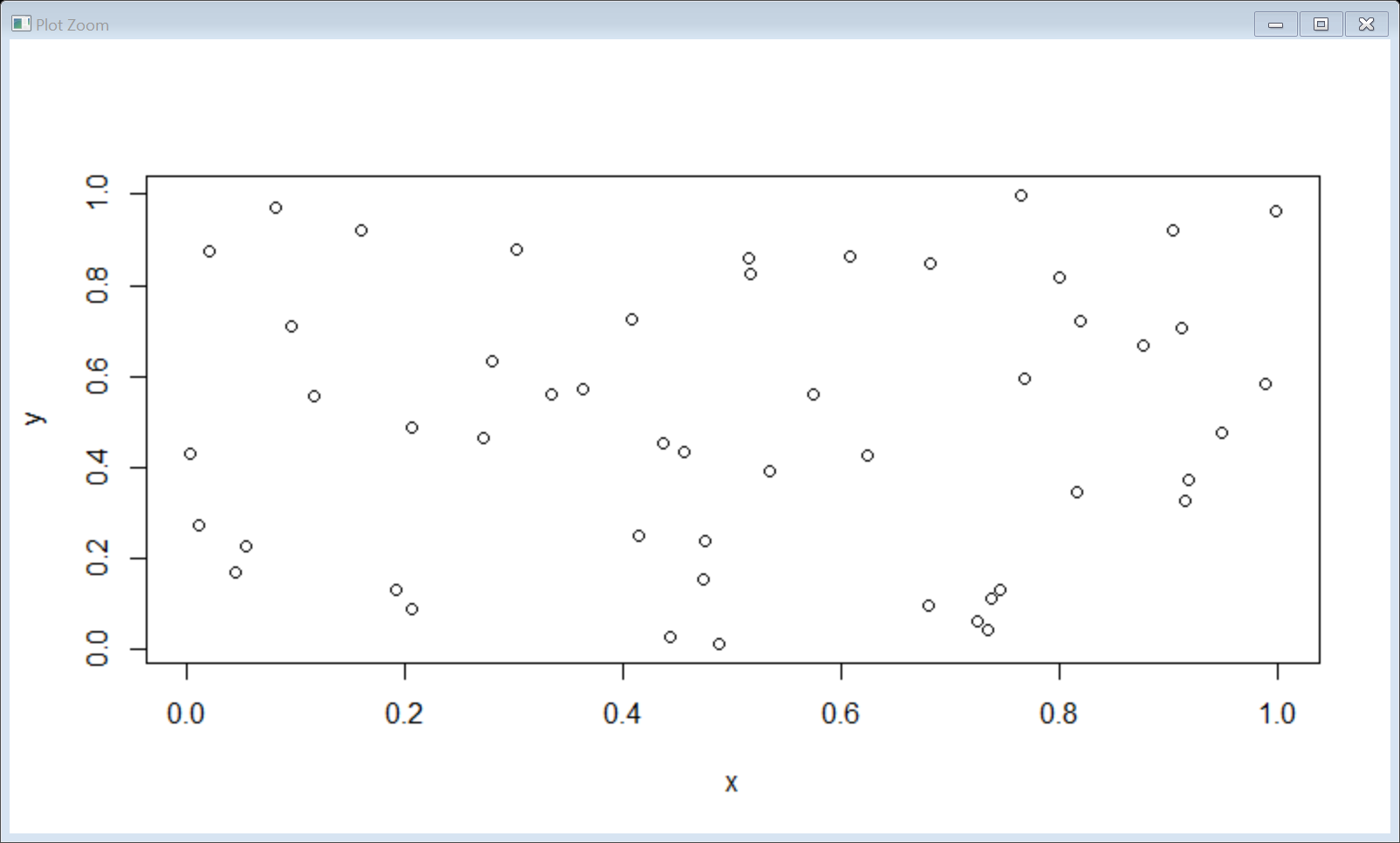


Step 8: drawn from the aggregate of the wine data on membership\_1

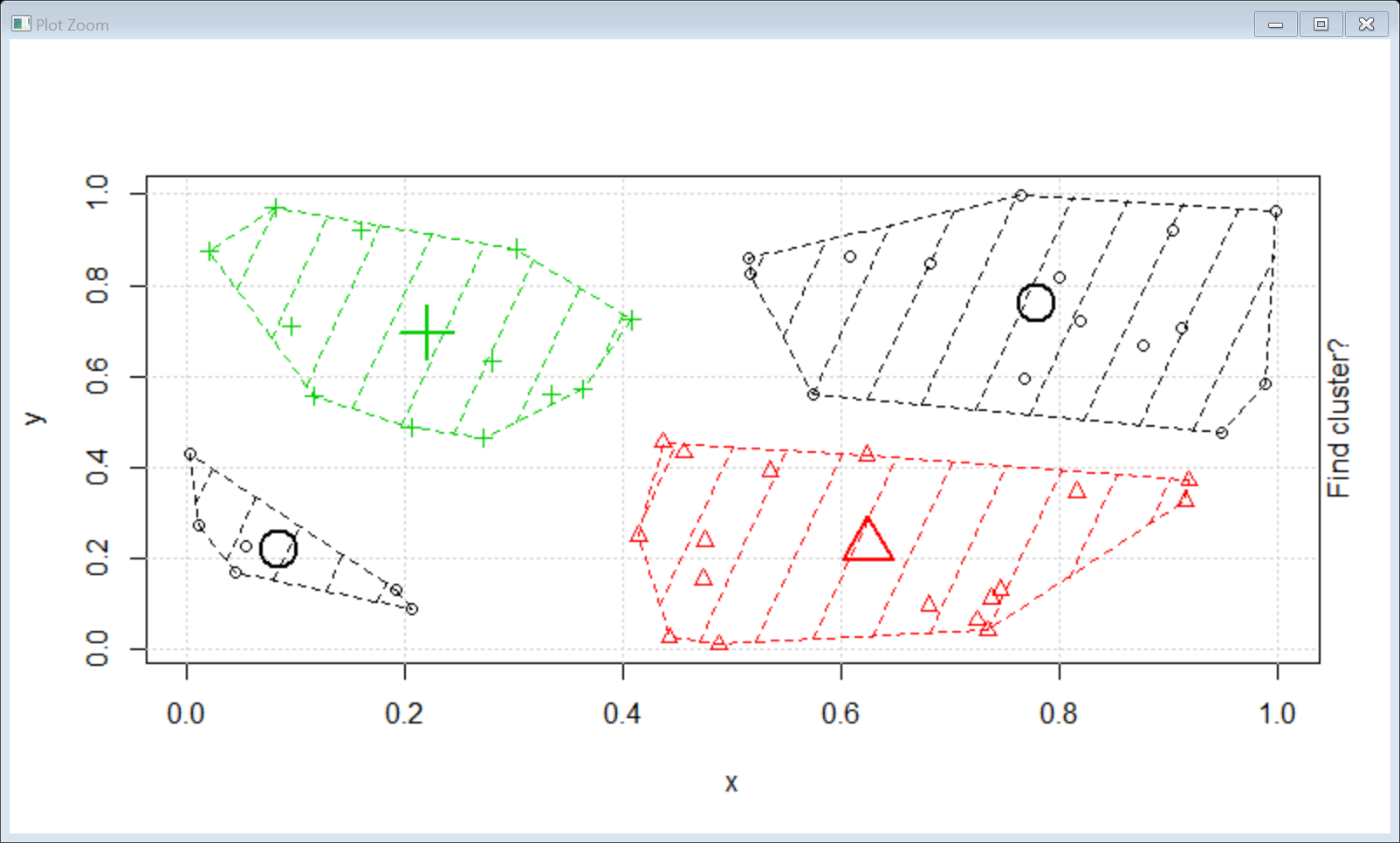


Step 9: Create the finale write.csc file showing the result.

Step 10: Kmeans Clustering : generating 50 random numbers of x and y and combining them and then scatter plot



Step 11: Animated representation of the data which forms the centroid and checks each data point with that centroid and keeps calculating until the most appropriate clusters are formed since 4 clusters are mentioned so it calculated with 4 centroids accordingly.



Step 12: determine the centers of clusters

x y

[1,] 0.08581447 0.2195191

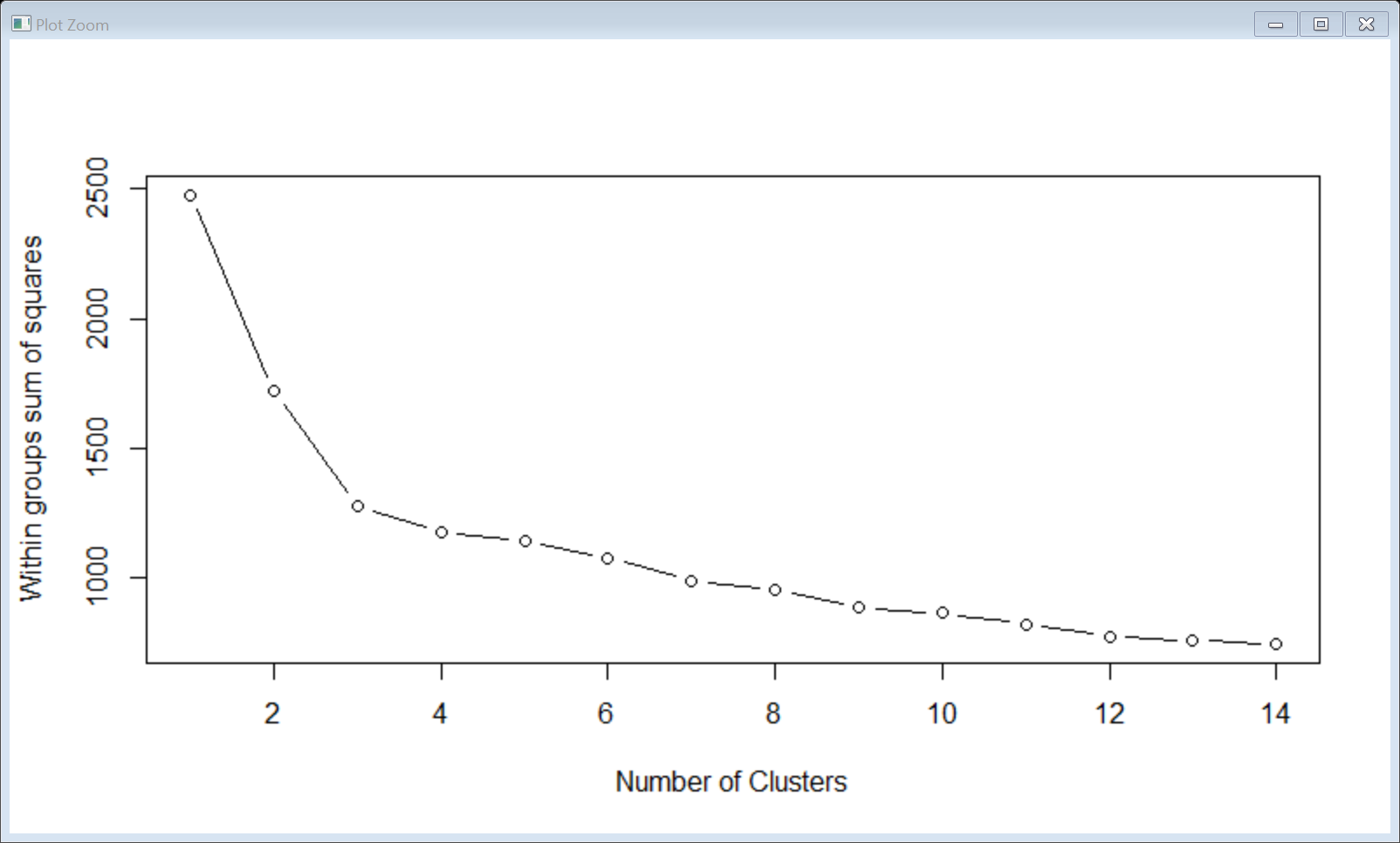
[2,] 0.62534619 0.2271623

[3,] 0.22050637 0.6956555

[4,] 0.77919568 0.7606356

Step 13: Normalize the data and calculate the K means with 4 clusters

Step 14: Determine number of clusters by scree-plot and determining the elbow curve



Conclusion: 4 clusters are formed