CS5542 Big Data Apps and Analytics

LAB ASSIGNMENT #2

REPORT and SCREEN SHOTS

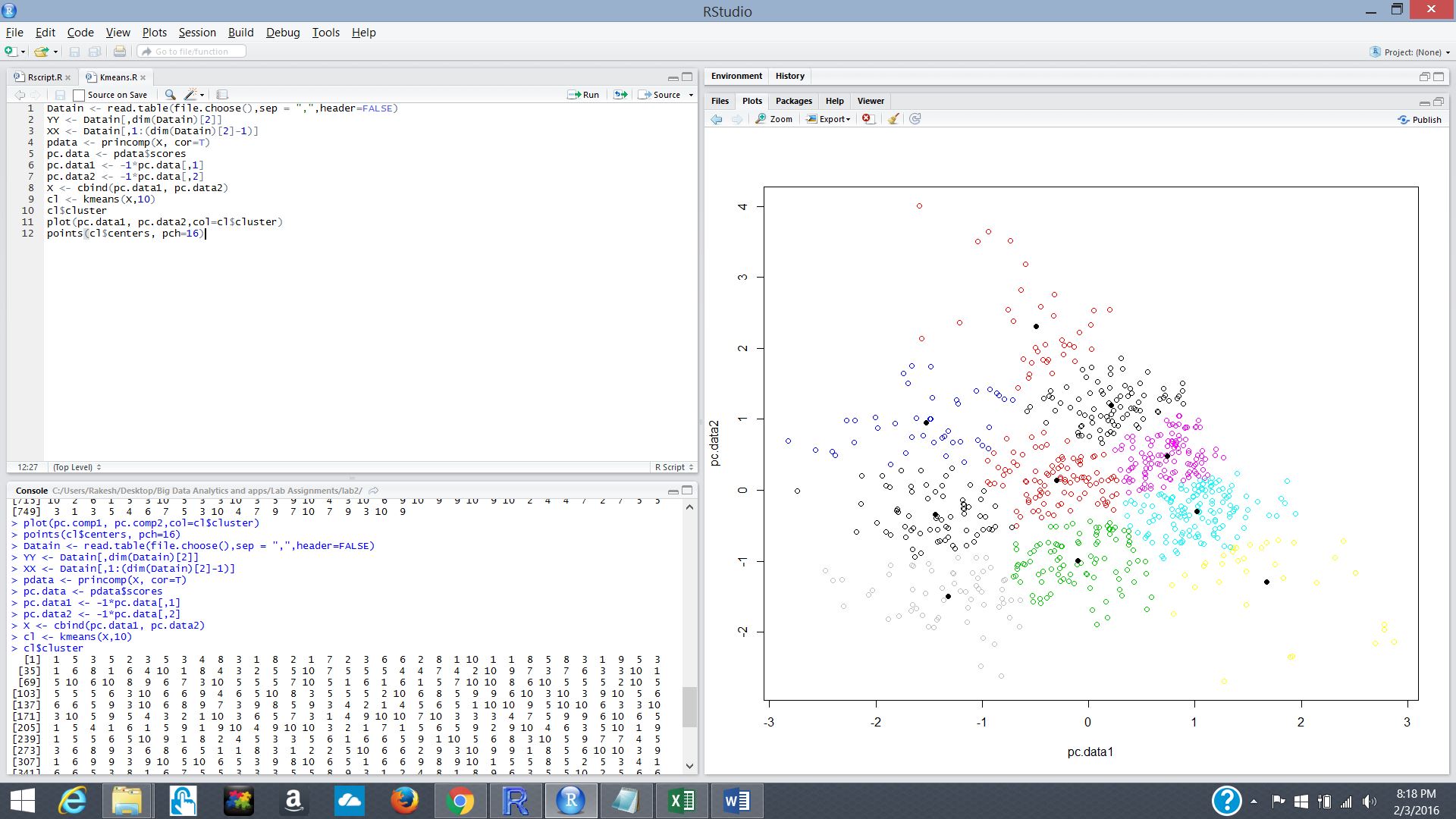
**1. R Project**

**Prepare a dataset and perform k-means clustering.**

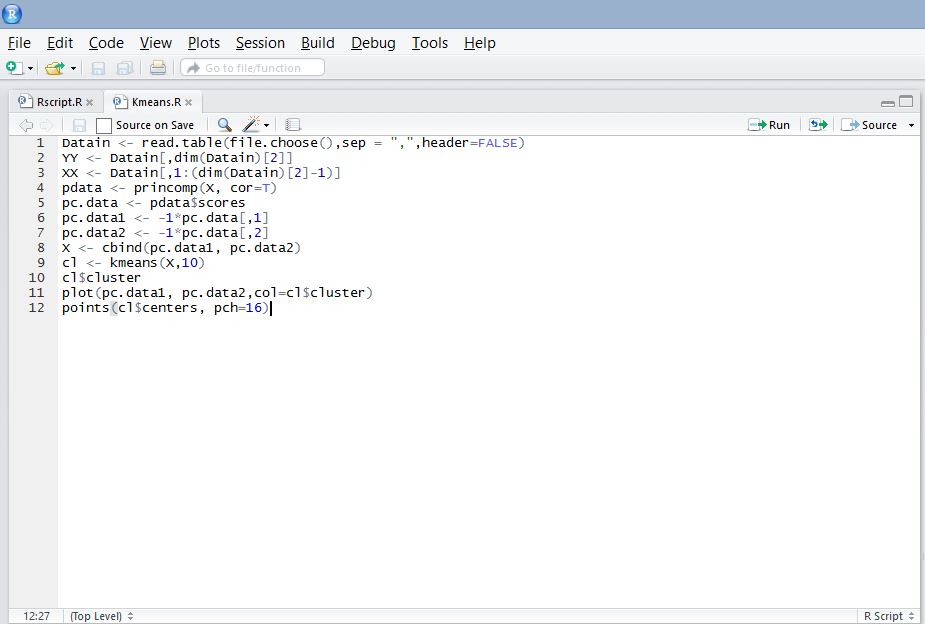
The diabetes data set is taken from the UCI machine learning database repository at:

<http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>

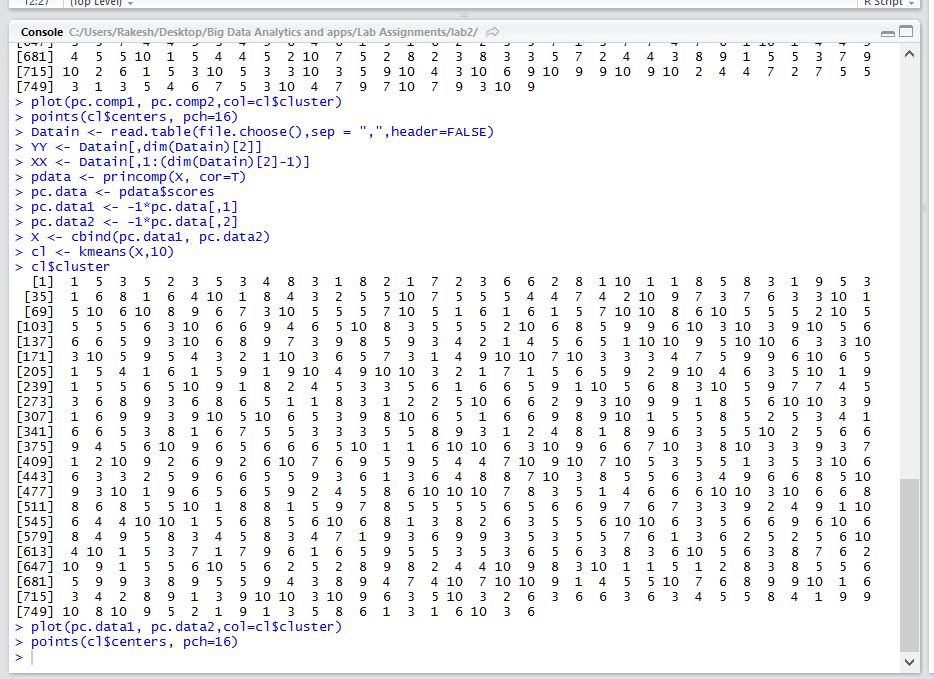
1. We LOAD the data into R using the file.choose() method and read csv file into Dataframe in R.

****

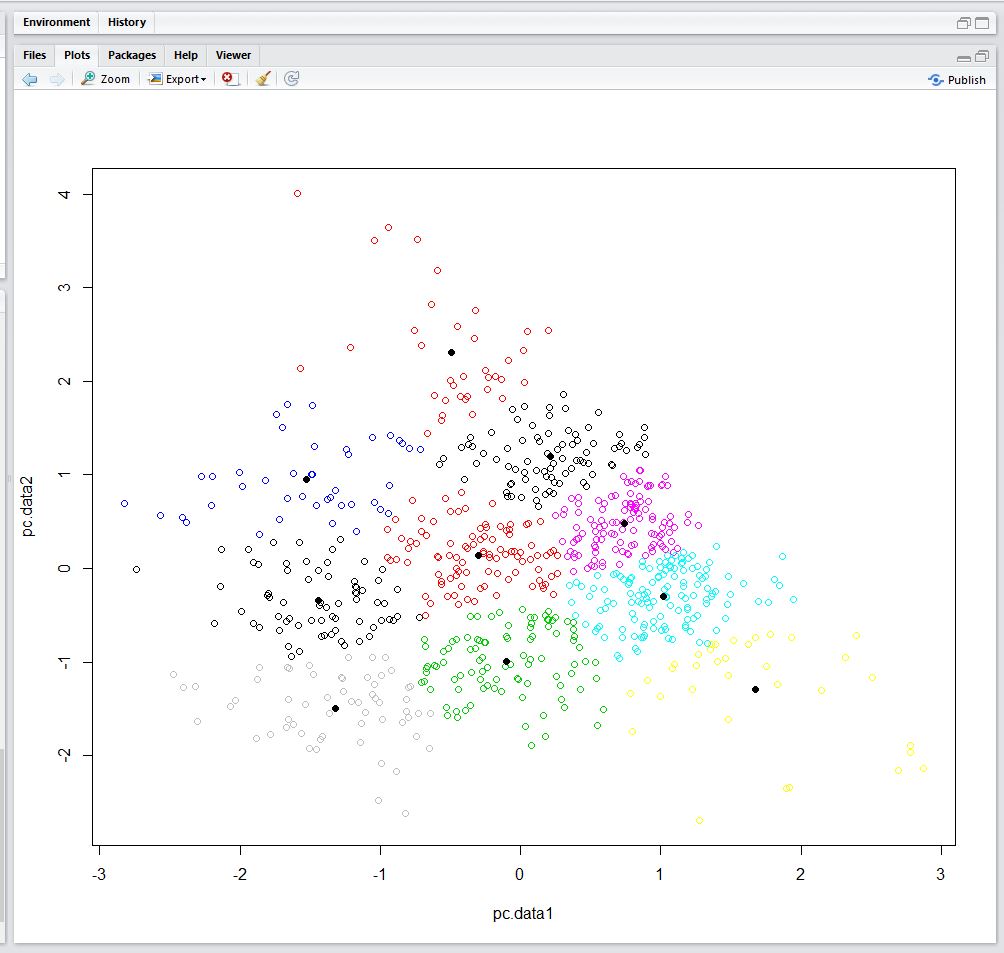
2. Write the R script in R studio.

****

3. See the cluster output in R console using command Cluster

****

4. Graph of cluster output with 10 clusters and Dark black dots indicate the centroid of means of cluster.

****