**Lab 2 – Requirements**

**A.** Given data in the **grades\_km\_input\_samle.csv**, please complete the following tasks.

1. Import data in .csv file to **grade\_input** data frame in R
2. Because the student ID is not used in the clustering analysis, pleas exclude that column from the k-means input matrix, kmdata
3. Use k-means algorithm identify clusters for k = 1, 2, …, 15 . For each value of k, calculate the WSS
4. Use R to loops through several k-means analyses for the number of centroids, k, varying from 1 to 15. For each k, the option nstart=25, specifies that the k-means algorithm will be repeated 25 times, each starting with k random initial centroids. The corresponding value of WSS for each k-mean analysis is stored in the ***wss***vector
5. Using the basic R plot function, each WSS is plotted against the respective number of centroids, 1 through 15
6. Summarize information with k=3
7. Verifies that the results of wss
8. Visualize the data and assigned clusters by ggplot2 package in R
9. Do diagnostics with n= 2

**B. The result report must be included in a document which has:**

- All necessary commands used to do task

- And output screen of command result