A flow cytometry data set is provided: facs.csv

Perform k-means clustering using `kmeans()` function of R.

Perform clustering for different values of k from 2 to 6.

Calculate and plot WCSS vs k.

Decide the optimum number of clusters or k. Note that increasing k does not always help. Optimum k is decided based on WCSS, and cluster overlapping detected visually on the cluster diagram.

Plot two scatter plots: FL2.H vs FL3. H and FL3.H vs FL4.H. Color the data points based on the cluster. These are cluster diagrams.

Submit:

a) The R script

b) A report in MS Word. The report should have the following:

1. Name and roll number
2. WCSS vs k plot
3. The number of clusters or k you chose for the final clustering. Give the reasoning behind your choice.
4. Cluster diagrams with the k decided by you: FL2.H vs FL3. H and FL3.H vs FL4.H Color the data points based on the cluster.

   Note: In all plots, axes must be suitably labelled.