HOMEWORK 3

Use
library(ggplot2)
library(dplyr)
library(factoextra)
library(cluster)

1. Call iris dataset

2. Plot "sepal width" vs "sepal length", coloring the points according to their species.

3. Apply principle component anaysis to see potential number of clusters.

4. Write your comment about the importance of components, and another sentence about how many components will you use.

5. Use fviz_pca_ind function to plot 2-D PCA plot with your pre-defined number of components.

Use
library(stats)

set.seed (102)

6. Do k-means clustering in R using the function kmeans(). The primary argument to this function is centers, which tells the function how many cluster centers you want to use.

7. Compose a table for predicted vs true valies of species

8. Perform agglomerative HC with hclust:

   First compute the dissimilarity values with "dist" function (using euclidian distance) and then feed these values into "hclust" and specify the agglomeration method to be used as "average".

9. Plot the dendrogram.

10. Take the cutree function and cut the dendogram from height 3.

11. Compose a table for predicted vs true valies of species.


After the analysis in each part, your comments are also extra points.