CSC 522: AUTOMATED LEARNING AND DATA ANALYSIS

HW 23R – PLOTS AND EXPLANATION

Student Name: Akshay Nalwaya
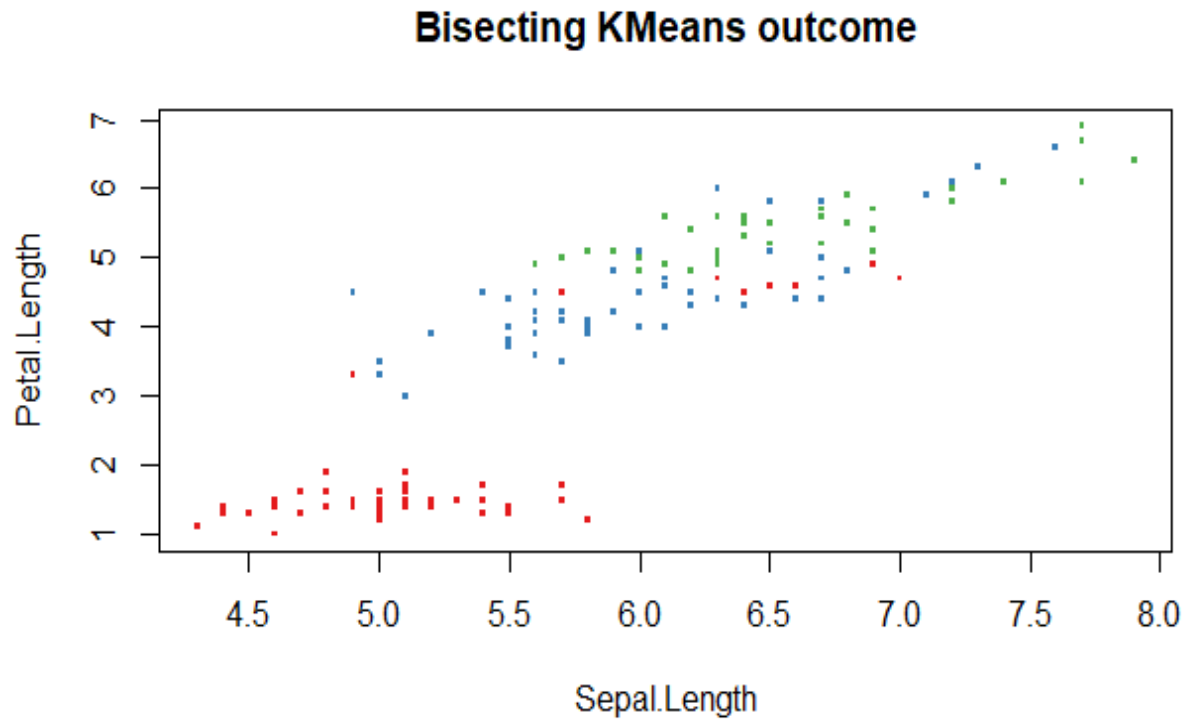**Student ID:** 200159155
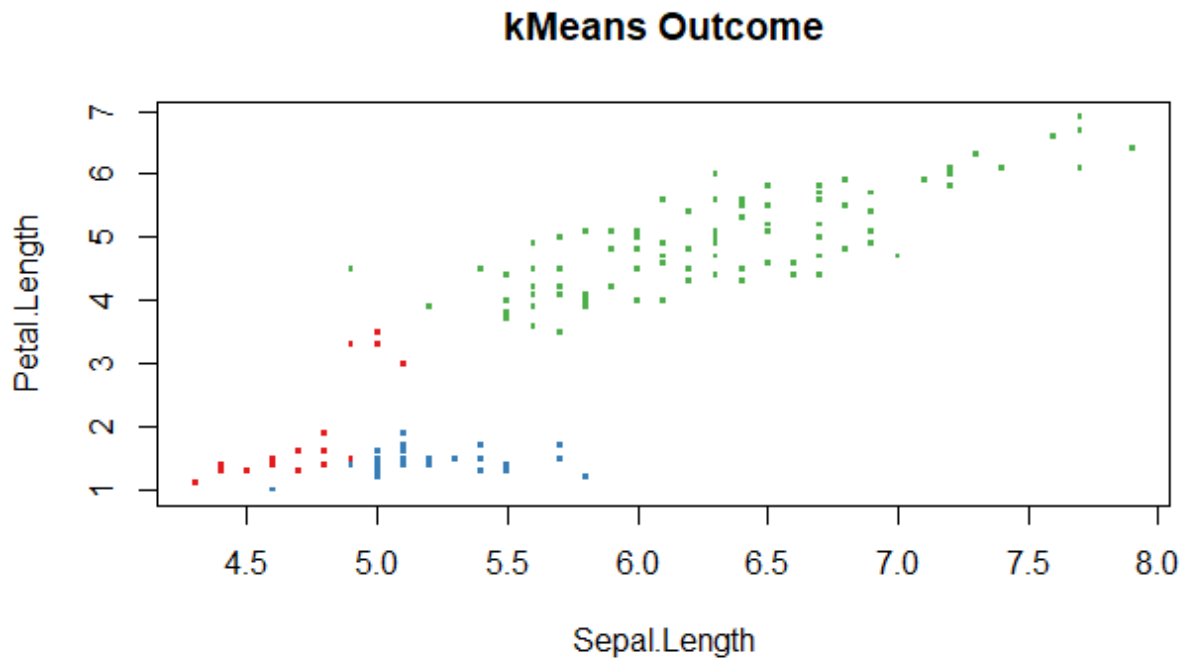**Unity ID:** analway

Q1. BISECTING KMEANS

COMPARISON OF BISECTING KMEANS WITH KMEANS

The plots obtained after running bisecting kMeans and kMeans on the data *(data.df)* are shown below.

Bisecting KMeans outcome

kMeans Outcome

The following finding/inferences can be made based on the results and plots obtained:

- Since, kMeans was initialized with 3 centers, it partitions the data cluster present in the lower part of graph into two sub-clusters and makes the upper data cluster into a single cluster. This is not the case in bisecting kMeans which classifies all the lower data cluster into a single cluster and divides the upper data cluster further into 2 clusters.
- The result of bisecting kMeans is more logical than the result obtained from kMeans algorithm as it correctly splits the big upper data cluster into two clusters which will increase the prediction accuracy.
- ***Comparing Overall SSE:***
  - Bisecting kMeans = 76.31102
  - kMeans = 142.75352

```
[1] "Total SSE (Bisecting kMeans): 76.3110228401192"
[1] "Total SSE (kMeans with default parameters): 142.753520021645"
```