HW2: Programming Assignment
Due date: 12th October, 11.55pm

Objective: Implement bisecting k-means

Notes:

(1) Bisecting k-means internally calls k-means, and that k-means should be your own implementation
(2) Several variations of bisecting k-means is possible, so write down your assumptions clearly
(3) Bisecting k-means can stop on "K" or given threshold on SSE
(4) Play with existing k-means implementation(s) in R
    a. To compare results with your own implementation of k-means
    b. To get a good idea on what would be appropriate SSE
(5) Use example R code, to generate clustering plots
(6) Any programming language is fine

Dataset (filename = d-c4hw2.csv)
Dataset contains two variables, named "length", and "width" and 400 objects (or records)

Question:

(1) Plot the given data (single color – as you haven't clustered yet!), example is shown below.
(2) Apply your own K-means clustering, with K=4 and K=8. Generate clustering plots (see examples below), use random colors or symbols to distinguish clusters. Comment on which clustering result is better (just by visual comparison).
(3) Apply your own bisecting k-means on the given data for K=2 to 8, and answer the following sub-questions
    a. Plot clustering SEE (y-axis) against number of clusters (x-axis). (e.g., is shown below)
    b. From the plot identify optimal K
    c. For identified optimal K, show clustering plot

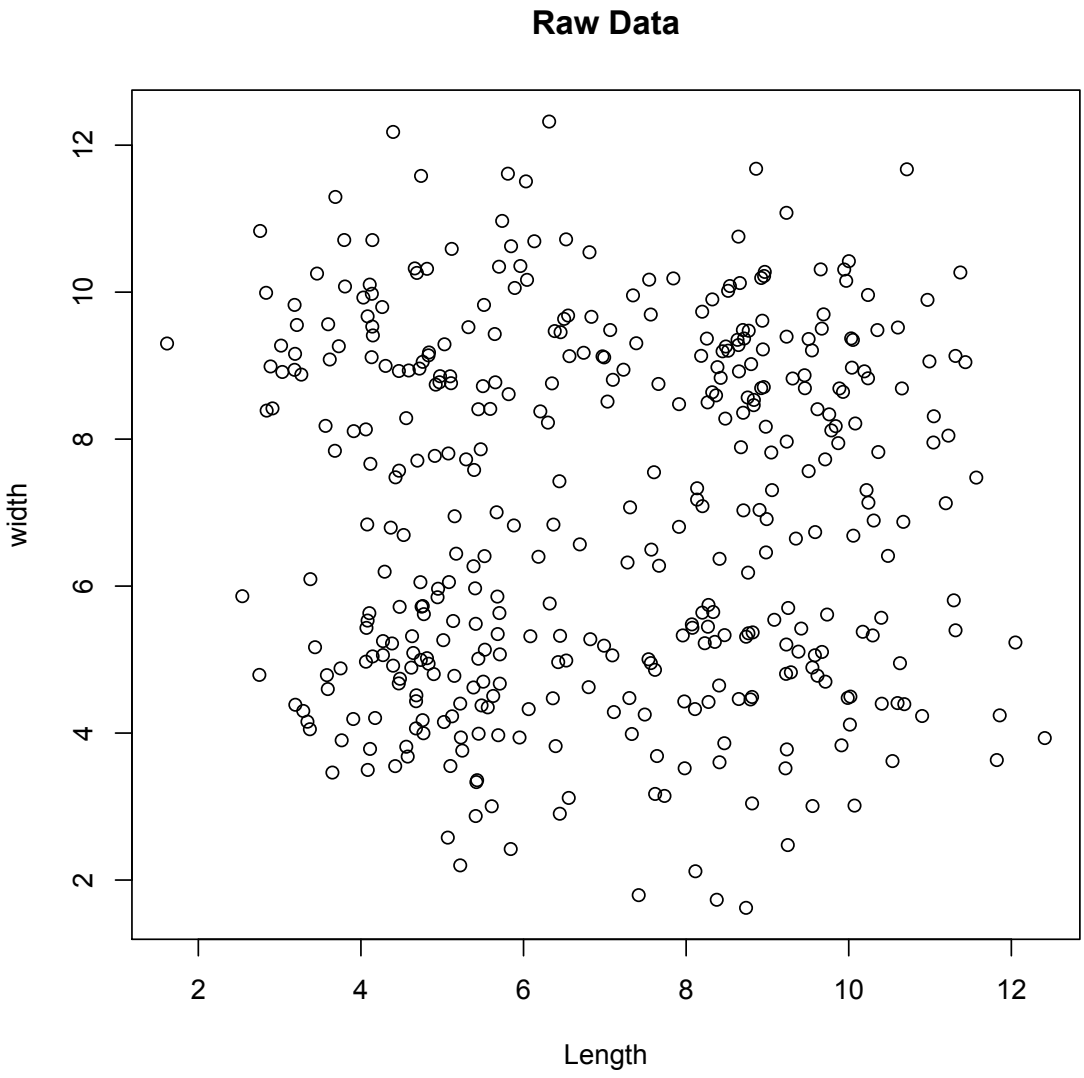
What needs to be submitted?
Single zip file containing:

(1) Your code (with suitable commenting)
(2) Answers to all 3 questions above, with plots embedded (doc or pdf).

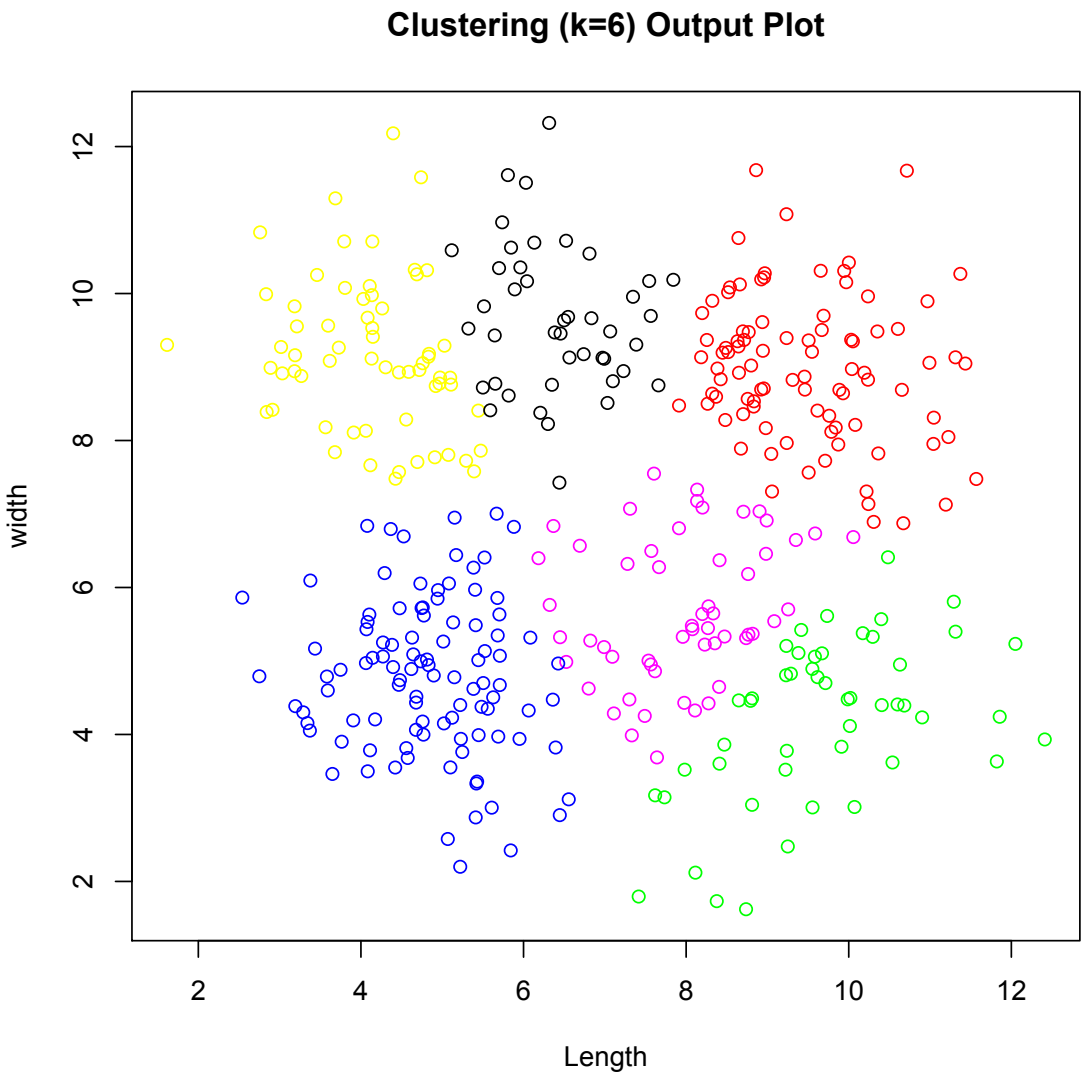
Make sure that the system shows "**Submitted**" status (not "draft"). I changed system configuration, probably this time it should not ask twice, but any case please make sure that your submission is complete.

Example plots:
Note: These are sample plots, your plots may or may not match these and your objective should not be to match with these figures. These are provided to help you debug your outputs, as they should look similar to these sample plots.

1) Raw data plot



2) Clustering Plot (with K=6)



3) SSE vs. K plot (or elbow plot).