Stat 123 Homework Assignment 2

Due Friday 20h by 8 pm

Using R Markdown, please complete the following assignment. If an answer does not require any R code, you can type the answer to the question outside of a chunk. Make sure that your assignment is well labelled so that it is clear where each question's answer begins. Your assignment should be submitted as a pdf (whether you knit directly to PDF, or knit to HTML or Word and then convert the file to a pdf).

1. The built-in UCBAdmissions data set is a 3-dimensional array that contains the following information:

   - Dimension 1: Admit Admitted, Rejected
   - Dimension 2: Gender Male, Female
   - Dimension 3: Dept A, B, C, D, E, F

   Note: You can find the UCBAdmissions data set using data().

   If you wanted to access the data, you could type in UCBAdmissions[1,1,2] to get this value. If you wanted to create a table with the number of rejected students in department A, you could type UCBAdmissions [2, ,1].

   (a) Create (and print out) a table that contains all students in department D

   (b) Create (and print out) a vector called *department.D* that contains the admitted and rejected students.
   Hint: You may need to use rowSums() on your answer from part (a).

   (c) Create a bar plot displaying the admitted and rejected students in Department D. Make sure to include a main title and label your x-axis. Also, make sure that each bar is a different color.

   (d) Create (and print out) a vector called *admitted.females* which contains the admitted females in Department D.

   (e) Create (and print out) a vector called *pct.admitted.females* which contains the percentage of admitted females in department D.

   (f) Create a pie chart that displays the *pct.admitted.femalesdata*. Be sure to include a main title for your pie chart.

   (g) What does the pie chart imply about the number of admitted females in department D?

2. The following question deals with the data set *Government_expenditure_per_student*.csv, which you will need to download from the assignment page.

   (a) Create (but do not print) a vector called *expenditure* containing the last column of the dataset named *Value*.

   (b) Create a histogram displaying the distribution of this variable. Be sure to have both a main title and a title on your x-axis.

   (c) Describe the shape of the distribution (symmetric, left-skewed, right-skewed).

   (d) Compute the appropriate center value and the corresponding measures of variability.

   (e) Remove decimals from the vector named *expenditure* by using round() and create a stem plot.