Stat 311 R Assignment 1

This assignment will introduce you to `rmarkdown`, a package that allows you to embed R code and output into a document. We will use `rmarkdown` to create HTML file output that you will then convert to pdf. Your final product will be a pdf file uploaded to Canvas.

Make a copy of the R1 Tutorial .Rmd file and save it named something like, R1_FirstNameLastName.Rmd. Modify/add to the code in the .Rmd file to address the problems below. Make sure to add a ##### Header for each problem. Upload your completed assignment to Canvas by 11:30 PM on the due date listed in Canvas.

For questions that have multiple parts, you can put the code in one chunk, but you should add a comments with the chucks indicating Part (a), Part (b), etc. Remember to use a single # for a comment within a chunk of R code.

1. Install `rmarkdown` if you did not do it when watching the R1 Tutorial. Also install the `knitr`, `ggplot2` and `dplyr` packages. The template already attaches the `knitr` package. Add code to attach the `ggplot2` and `dplyr` packages.

2. Using R as a calculator: write code to calculate the area of triangles $\left(A = \frac{bh}{2}\right)$, with bases that vary from 0.5 to 2 by 0.5 increments and height are 1 and 3. To get a sequence of base values that vary by 0.5 use the code `bases <- seq(from=0.5, to=2, by=0.5)`. Thus, you are calculating the area for $4 \text{ x } 2 = 8$ triangles. Write your code using the vectors as opposed to calculating eight individual areas. Note: you may have to do this in two steps (4 areas and 4 areas). Make sure your 8 areas show up in the knitted output. Describe the combinations of base and height that yield the minimum area, maximum area and areas of 0.75.

3. Reading the data into R.

   Read the data patient data (Patient_Data.csv) into R, creating a data frame. The command you need is already in the template.

   a) Add code to turn all categorical variables into factors.

   b) What is the value of the total cholesterol (TotChol) for the 50th observation?

   c) Using one line of code, display the values of total cholesterol (TotChol) and systolic blood pressure (SystolicP) for observations in rows 15, 25 and 99.

4. Subsetting Data

   a) Create a new object for the subset of patients whose marital status (MaritalStat) is divorced ("D"). Do not display the data but do report how many observations are in this subset.

   b) Report the mean total cholesterol value for divorced patients.

   c) Report the standard deviation (SD) of total cholesterol for divorced patients.

   d) How do the mean and SD of total cholesterol for divorced patients compare to married (MaritalStat of "M") patients?

After you finish coding and adding narrative to your .Rmd file, knit the file to HTML. Read it over to make sure you are happy with the content and formatting. Then open the HTML file in a browser and save as a .pdf file. Upload the .pdf file to Canvas under R Assignment 1 before the due date.