Stat 311 R Assignment 2

This assignment uses the same patient data set (Patient_Data.csv) that was used for the first R assignment. As with assignment 1, use `rmarkdown` to create HTML file output that you then convert to pdf. Your final product will be a pdf file uploaded to Canvas by 11:30 PM on Wednesday, January 27th.

The R2 Tutorial introduces all the functions you will need to complete this assignment. As mentioned in the R2 Tutorial videos, the tutorial shows multiple ways to do the same thing. For this assignment, you pick the functions you want to use to answer the questions. Do not put in any redundant, unnecessary code in your final product. Create a new .Rmd file for this assignment. I recommend copying over the title block and setup chunk into your new file. Then, copy, paste and edit code from the tutorial as needed for the assignment. Remember to include section and subsection headers as needed to make your assignment readable.

Your assignment includes the following problems.

1. Do a simple univariate EDA of the total cholesterol (TotChol). Your response must include a 7-number summary, a histogram and a boxplot. Summarize the distribution of total cholesterol, addressing shape, spread, and outliers.

2. Make comparative boxplots of total cholesterol (TotChol) by sex (Sex). Compare and contrast the distribution of total cholesterol by sex.

3. Create a contingency table for marital status (MaritalStat) by sex (Sex), where Sex are rows and MaritalStat are columns. Show a table of counts and a table of joint percentages. No interpretation needed for this question.

4. Find the marginal percentages for marital status (MaritalStat). No interpretation needed for this question.

5. Find the conditional percentages for marital status (MaritalStat) for females (Sex = F). No interpretation needed for this question.

6. Create a side-by-side bar graph for marital status (MaritalStat) within levels of sex (Sex categories should be on the *x*-axis). Do marital status and sex appear to be statistically independent or dependent? Make a qualitative assessment by looking at the pattern within each level of sex. If the patterns are similar then that suggests that the two variables are independent.

7. Explore the relationship between TotChol (*y*) and Weight (*x*).
   a. Make a scatterplot and describe the relationship.
   b. Calculate and interpret the correlation coefficient.
   c. Regardless of the relationship you observe in part (a), estimate the regression parameters for total cholesterol on weight.
   d. Write out the regression equation.
   e. Interpret the estimated regression slope parameter in the context of the problem.
   f. Does using simple linear regression make sense for modeling the relationship between total cholesterol and weight?