PSY 390 Pre-Exam 1

"Pre-exams" are assignments that are designed to be very similar to the exams. My mindset is that the pre-exam will cover the exact same data analysis steps and skills.

Pre-exam 1 covers the following skills/concepts/etc

- Create a R notebook and be able to knit it
- Import data as data frames from a .csv file or from Google Sheets.
- Be able to calculate summary statistics, summary statistics by a grouping variable, t-tests, and correlations
- Be able to interpret t-test and correlation output
- Be able to use the tidyverse to select a subset of a dataframe, select some columns from a larger data frame, sort the data, and be able to make new variables
- Be able to use ggplot to generate a histogram and scatterplot. Also be able to format the plot

In this pre-exam, we will cover step data from a fitbit with number of steps per day as the main variable of analysis. The data frame is available as "preexam1.csv".

Make sure you create a R notebook to answer these questions and when you are completed with this assignment, knit this to create an output that you will then upload.

Step 1: Load the psych and tidyverse packages into R.

Step 2: Load the data in the preexam1.csv file as the dataframe `step`.

Step 3: View the data in the viewer to make sure it load correctly. There should be 402 observations of 5 variables. The variables are as follows:

- time - the date of each observation
- steps - number of steps in that day
- day.week - the day of the week
- month - the month for each observation
- day.month - the day of the month of each observation

Step 4: Write code chunks to complete the following questions. Write annotations to answer the questions.

1. Use a summarizing function to output the questions below

- What is the mean number of steps?
- What is the standard deviation for the number of steps

2. Use a summarizing function to answer the following questions. Include your answers as annotations.

- Which day of the weeek has the highest average number of steps? How many are there?
- Which day of the week has the lowest average number of steps? How many are there?
- Which month has the highest and lowest average number of steps? How many?

2. Sort the data to answer the following questions:

- What are the three dates with the most steps?
- What are the three dates with the least steps? (not counting zero)

3. How would you create a data frame which removed all the rows where there are zero steps? What reason might there be for there to be rows with zero steps?

4. Create a dataframe called `steps.clean` which removes all observations with less than 1000 steps. Use this data frame for all the other questions in this pre-exam.

5. Create a new data frame called "steps.weekend". Use this data frame to conduct a t-test comparing Saturday to Sunday. Which day has a higher number of steps? Is the difference significant? Report your answer using APA formatting

6. Create a histogram for `steps.clean`. Give the histogram an informative x-axis label and a title. Make the bars blue with a red outline. Write down what code you would use here.

7. What is the correlation between steps and day.month? What would this correlation mean?