PSC 103B

Homework 1

**Instructions**

The data for this homework is in the ﬁle ‘HW1Data.csv’ (on canvas).

Please use R/RStudio to complete the following questions. **Include the code you used to generate your answer for each question.**

I encourage you to do every question, but only those marked with an asterisk will be submitted (via canvas) and graded.

You may get help from your classmates while working on the assignment, but you must do all the work yourself. You may work beside each other, and consult each other, but everything you turn in must be your own code and words. Please make good use of the Canvas discussion forum if you get stuck.

Please submit your completed assignment as a PDF to Canvas by the start of class (10 AM) on the assigned due date.

1. **Loading and Exploring the Data**
2. \*(1 pt) Read the data in using R (HW1Data.csv). Use the head() function to make sure the data loaded in correctly as well as to see what variables are in the dataset. Include the output from the head() function in your answer.
3. What are the dimensions of the data set?
4. \*(1 pt) What data type (e.g., numeric, character, etc) is in each column?
5. How many values are missing for each variable?
6. \*(1pt) Subset the height and weight variables and save them as a new object under the label heightWeightData. Include both the code you used to subset and a screenshot of what your new object--heightWeightData--looks like.
7. **Describing the Data**
8. How many participants were female? How many were male?
9. \*(1pt) What is the frequency (tally) of education completed (e.g., HS and MA) in the data set (ignoring missing values)? That is, for each level of educational attainment, how many participants were in that level?
10. \*(1pt) Use the group\_by() and summarize() functions to calculate the average (i.e., mean) weight for men and the average weight for women.
11. \*(1pt) What is the average age of participants based on their level of educational attainment? That is, what is the mean age of participants who had completed HS, 2 years of college, etc.?
12. **Analyzing the Data**
13. Calculate the 95% confidence interval around the mean of one of the variables. Assume the population standard deviation is 1.
14. \*(2pt) Report and interpret the confidence interval from Question 10.
15. \*(1 pt) Calculate the z-score for a value of your choice. Use the same variable as Question 11.
16. \*(1 pt) Interpret the z-score from Question 12