**-1**

**Data Science for Social Scientists**

PSYC 546, Spring 2023

Week 7 – In-Class Assignment

**Due Date**: March 2nd (by 11:59 PM)

**Reminder**: See the assigned readings, cheat sheets on Canvas, and the lecture slides for a tutorial on how to use R to perform the various functions included in the in-class assignment below. **Once completed, you should submit a completed version of this document and your final R script file to the Week 7 – In-Class Assignment – Submission Portal on Canvas**.

Your submitted R script file should contain code to answer all of the questions below. Please use comments (e.g., #Question 1) to label the code for each question.

1. Questions 1 to 3 use the **gss­\_cat** data frame from the tidyverse. Collapse the marital variable into a new variable named marital\_recoded. The new variable should be binary with two levels (“married” and “not married”). All levels on the original marital variable not “Married” should be collapsed into the “not married” level on the recoded variable. Make sure this recoded variable is appended to the gss\_cat data frame. Report the frequencies of these two groups below. [2 points]
   1. married: 11366
   2. not married: 10117
2. The variable “year” in gss\_cat is an integer class type. First, coerce it to be a factor. Then recode/collapse the factor into a new variable named year\_recoded so that all years less than 2010 have a level name of “2000s” and all years 2010 or after have a level name of “2010s”. Make sure this recoded variable is appended to the gss\_cat data frame. Report the frequencies of these two year-levels below. [2 points]
   1. 2000s: 14927
   2. 2010s: 6556
3. Use the original “marital” variable for this question (i.e., not the recoded one you created for Q1). Grouping by marital status, calculate the mean age. Then create a geom\_point() ggplot figure that has age on the x-axis and marital status on the y-axis. Make sure that the marital factor is reordered based on the average age. Paste your final figure below. [2 points]

Chart, scatter chart

Description automatically generated

1. Using the lubridate package, create a date-time object that represents the exact start time of our first class this semester (i.e., when you officially embarked on your data science journey this semester). As a reminder, our first class was on January 18th at 8:15 PM. For this question, imagine class began at the exact second of 8:15 PM. That was probably not the case, but let’s pretend for this question. [2 points]
2. Using the date-time object created in Question 4, calculate the duration in seconds from the current moment to the beginning of class. Report this value below. [2 points]

* 3655594.7181201s