**Data Science for Social Scientists**

PSYC 546, Spring 2023

Homework Assignment 6

**Due Date**: March 15th (by 8:15 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 Homework 6 – 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. Using the lubridate package, create a date object of your choosing. This could be a birthday, an anniversary, a graduation date, or any other special date of your choosing. You do not need to explain what this date refers to, but just save it as an object. Then, using this date object, create code to provide answers for the following items. [2 points overall]
   1. The duration (in seconds) between today (i.e., the day you are working on this item) and the date object you created: 24883200s
   2. The difference (in days) between today and the date object you created: 288d
   3. The day of the week that the date occurred: Saturday
   4. The day of the year that the date occurred: 141
2. Import the **credit\_card.sav** data file from Canvas. The variable “dob” is the date of birth of the credit card user. The variable “card\_date” is the date the primary credit card was issued to the credit card user. Create a new variable in the data frame called “age\_at\_issue”. This variable should consist of the difference in years between the person’s date of birth and when they were issued their primary credit card. Make sure to round the age values in the age\_at\_issue variable to an integer (i.e., there should be no decimals in the age values in age\_at\_issue). [2 points]
3. Using the **credit\_card.sav** data frame, Question 3 will involve the “card” variable. This variable categorizes the type of credit card that was issued. First, make sure that the card variable is being classified as a factor variable in the data frame. Then, recode the factor card variable so that the text of the credit card is stored/displayed as the value instead of the number. The correspondence of numbers and text for the card variable is: 1 = “American Express”, 2 = “Visa”, 3 = “Mastercard”, 4 = “Discover”, and 5 = “Other”. [2 points]
4. Use the recoded card variable for Question 4. Grouping by card, calculate the median amount of money spent (derived from the variable “spent” in the data frame). Then create a geom\_point() ggplot figure that has money spent on the x-axis and card type on the y-axis. Make sure that the card type is reordered based on the median money spent. Finally, make the x-axis label be “Money Spent Per Transaction” and the y-axis label be “Credit Card Type”. Paste your final figure below and state which credit card was associated with the highest average expenditures in the data set. [2 points]

* American Express was associated with the highest median expenditures.

Chart, scatter chart

Description automatically generated

1. I believe that Questions 1-4 comprehensively cover the new material learned this week. However, to keep the HW assignments standardized at 10 points, this question is still included as a gimme. Hope you enjoy your spring break and I will see you on 3/15! [2 points]

* Pleasantly surprised! Thank you.