**Data Science for Social Scientists**

Psyc 546, Spring 2023

Week 8 – In-Class Assignment

**Due Date**: March 16th (by 11:59 PM)

**Reminder**: See the assigned readings, resources 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 8 – 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. First, create the following two character strings in R exactly as written:



Then, combine these two strings together into a single string named “combined\_string”. Make sure that a comma and a space separates the two clauses in the combined string. [2 points]

1. Using the combined string object created in Q1, perform the following functions utilizing appropriate stringr functions: [2 points overall]
   1. Replace the single exclamation point at the end of the sentence with three explanation points.
   2. Make all the letters upper case.
   3. Count the number of characters in the string and provide the answer: 72
   4. Count the number of vowels in the string and provide the answer: 22
2. Using the janeaustenr package, assign the text of the book “Pride & Prejudice” to an object named pride\_and\_pred. This requires filtering the austen\_books for Pride & Prejudice. Then, make it in tidytext format by unnesting it based on words. Next, remove stop words from pride\_and\_pred. Finally, perform a frequency descriptive analysis so that the top ten most frequent words are printed out to the console. [2 points]
3. Use the gutenbergr package to download “The Time Machine” by H.G. Wells. For reference, it has an ID of 35. Assign it to an object called time\_machine. Then, make it in tidytext format by unnesting it based on words. Finally, remove stop words from time\_machine. [2 points]
4. Using the “bing” sentiment lexicon, find the five most frequent positive words in time\_machine and the five most frequent negative words in time\_machine. Report these words in the table below. [2 points]

|  |  |  |
| --- | --- | --- |
| **Frequency Rank** | **Positive** | **Negative** |
| 1 | soft | strange |
| 2 | bright | darkness |
| 3 | silent | dark |
| 4 | beautiful | fear |
| 5 | comfort | fell |