**Data Science for Social Scientists**

PSYC 546, Spring 2023

Week 6 – In-Class Assignment

**Due Date**: February 23rd (by 11:59 PM)

**Reminder**: See the assigned readings and the Week 6 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 6 – 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. Import the **class\_roster.csv** data file from Canvas. The file contains a single column that contains the last name, first name for each student in class. Separate this single column into two columns (named LastName and FirstName). [2 points]
2. Now, unite the data frame created in Question 1 into a single column called FullName. The column should consist of the first name followed by the last name, with a space separating the two names. [2 points]
3. Import the **wide\_data\_example.xlsx** data file from Canvas. Gather the data into long format with the key column being named “time” and the value column being named “test\_score”. Assign this to an object called “long\_data”. How many rows are in the long\_data object? Does this make sense? Why or why not? [2 points]

* 300 rows in the object long\_data. This makes sense because the wide data has a total of 100 observations/rows (id), and each row has 3 times (time1, time2, time3). And when we transform it into long data, it becomes 100\*3 = 300 rows in total.

1. For the next two questions, you will need to import both the **time1.sav** and **time2.sav** data files on Canvas. Perform a left join merge on these two data sets. Report the number of rows contained in the merged data frame below. [2 points]

* The total number of rows is 150

1. Using the same two data sets, perform a full join merge. Report the number of rows contained in this merged data frame below. Based on your answers in the past two problems, how many new participants must have joined the study in the second wave of data collection? [2 points]

* The total number of rows is 250
* So a total of (250-150) or 100 new participants joined the study in the second wave