**Data Science for Social Scientists**

PSYC 546, Spring 2023

Week 4 – In-Class Assignment

**Due Date**: February 10th (by 5:00 PM)

**Reminder**: See the assigned Week 3 readings and the Week 4 Lecture Slides for a tutorial on how to use Excel and R to perform the various functions included in the in-class assignment below. **Once completed, you should submit a completed version of this document, an Excel file, and your final R script file to the Week 4 – In-Class Assignment – Submission Portal on Canvas**.

**Excel**

For problems 1-3, use the week4\_class Excel file on Canvas. You should submit a copy of this file with the data figures included. The point of these exercises is to get training in creating APA formatted charts that would be acceptable to submit in peer-reviewed journals.

1. Use the **scatter** sheet for Problem 1. Create a scatter plot with variable X on the horizontal axis and variable Y on the vertical axis. The axis titles should be “Variable X” and “Variable Y”, respectively; the chart title should be “Correlation between Variables X and Y”; all text should be in Times New Roman; the dots should have a black fill with no border; there should be no gridlines or figure border; both the horizontal and vertical axes limits should be 0 to 100 with intervals of 20; the horizontal and vertical axes should be black and not grey; a solid black linear trendline should be added to the scatter plot. Once finished, paste the scatter plot below as an image. [2 points]



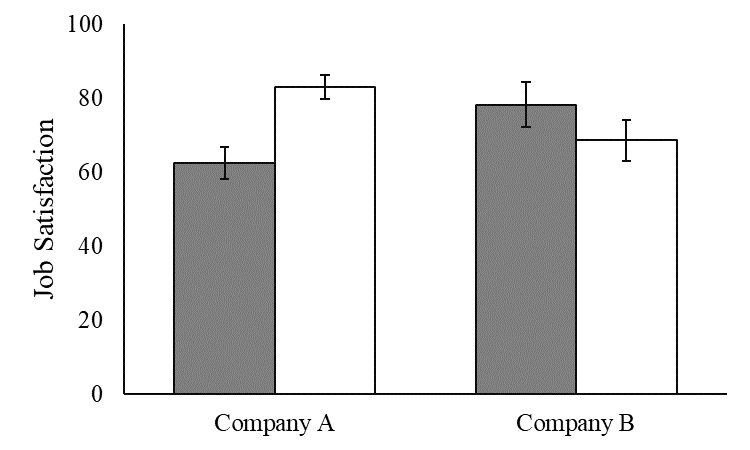
1. Use the **pie** sheet for Problem 2. Use the data in that sheet to re-create the pie chart below (created in Excel) the best that you can. Feel free to also paste your image next to the one below to compare. [1 point]



My pie-chart:



1. Use the **bar** sheet for Problem 3. Use the data in that sheet to re-create the bar chart below (created in Excel) the best that you can. First you will need to use the standard deviations (SD) and sample sizes (N) to calculate the standard errors (SE), so look online for that formula if you have forgotten it. When you are done, feel free to also paste your image next to the one below to compare. [2 points]



My bar chart:



**R/RStudio**

Your submitted R script file should contain code to answer all of the questions below. Please use comments (e.g., #Question 4) to label the code for each question. For all the below questions, use the traditional plotting functions in R covered in class. Next week we will spend the entire class on ggplot2 to get experience with tidyverse graphing functions.

1. Use the **nyc\_regents\_scores** data set in the dslabs library. Make a histogram of the us\_history scores. Have the horizontal x axis labelled as “US History Scores” and the chart tile as “Histogram of US History Scores”. Paste the figure below. [1 point]

Chart, histogram

Description automatically generated

1. Using the same data set as Question 4, make a scatter plot between integrated\_algebra scores on the horizontal x axis and english on the vertical y axis. Make the x axis and y axis labels “Integrated Algebra Scores” and “English Scores”, respectively. Finally, make the markers on the figure have a point shape of a diamond (5). Paste the figure below. [1 point]

Chart, scatter chart

Description automatically generated

1. Using the same data set, make a boxplot of the global\_history variable. Paste it below and report how many potential outliers are being identified by the visualization. [1 point]

Chart, box and whisker chart

Description automatically generated

* 3 outliers

1. Using the **olive** data set from the dslabs library, create a bar plot based on the region of Italy variable. Paste the figure below. [1 point]

Chart, bar chart

Description automatically generated

1. Using the same data set as Question 7, create a pie chart based on the area of Italy variable. Paste the figure below. [1 point]

Chart, pie chart

Description automatically generated