**Data Science for Social Scientists**

PSYC 546, Spring 2023

Week 5 – In-Class Assignment

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

**Reminder**: See the assigned readings and the Week 5 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 5 – In-Class Assignment – Submission Portal on Canvas**.

**ggplot2 in R/RStudio**

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 **death\_prob** data set from the dslabs package, create a scatter plot with age on the x-axis and probability of death within 1 year on the y-axis. Copy/paste your final figure below. Make sure the figure includes all the following elements [5 points overall]:
   1. Colors of the points based on the variable sex
   2. The size of the points equaling .75
   3. The label of the x-axis being “Age (in years)”
   4. The label of the y-axis being “Probability of Death”
   5. The limits of the y-axis being 0 and 1

Chart, line chart

Description automatically generated

1. Using the **diamonds** data set from ggplot2, create a histogram with diamond price on the x-axis. Copy/paste your final figure below. Make sure the figure includes all the following elements [5 points overall]:
   1. Only diamonds with a cut of “Premium” are included in the figure
   2. The bin width equaling 200
   3. The borders of the bars being black (using the col option)
   4. The label of the x-axis being “Price (in dollars)”
   5. The limits of the x-axis being 0 and 15000
   6. Faceting the figure based on the color of the diamond (include 2 columns in the faceted figure)

A picture containing text, sky, screenshot

Description automatically generated