

# Homework 07: Instructions

## Overview

This homework set will work with the current state of your **project's joined data frame**. At this point in the project, you need to start drafting and finalizing some major part so coding for data summaries and visualizations. Your liaison may ask you to clean up the data used for the summaries and visualizations but if they do, you can simply add the cleaning elements later. At very least, however, you will want to start building some code for basic plots types that will visualize the distribution of data for single variables as well as visualize variable relationships. Fine-tuning of both data and plots can come later but such fine-tuning will require working code nevertheless. This homework moves you closer to understanding your data in both tabular and graphical forms.

Some of the code for your visualization will likely be used later to incorporate into your report, so discuss among your team members which plots might be best to include and add them to your working notes and consider which plot code might be appropriate to add to your project visualization script(s).

## Thinking about Variables Models

1. (1 pt) Based on your knowledge about statistics covered in previous classes or in your discussions with your liaison, identify a statistical model that might be relevant to test. Provide that model below by listing the outcome variable and the predictor variable(s).
2. (1 pts) Examine your project data. Identify any variables that are qualitative in nature that you may need to convert into factors. List them.

## Summarizing Data

3. (4 pts) Identify **two** numeric and **one** character variables that are relevant to your project. Create the following summary tables using your **joined data frame**.
  - a. Create a summary table of the *mean*, *standard error*, and *confidence interval*, and *sample size* for one of the numeric variables.
  - b. Do the same for the other numeric variable.
  - c. Create a summary table of the *mean*, *standard error*, and *confidence interval*, and *sample size* for one of the numeric variables but now ensure that your table contains the data summary for subsets of data based on a character/factor type variable for grouping.
  - d. Do the same for the other numeric variable.
  - *Tip:* Consider making the tables more presentable by using `{gt}`.

## Visualizing Data

4. (2 pts) Select *one* of the numeric variables used earlier.
  - a. Following concepts addressed in the visualization module slide materials create a **histogram** to represent the distribution one of the numeric variables.
  - b. In a second plot, create the same histogram but now use one of the character/factor variables to facet the plot so that you have a histogram for each subgroup.
    - *Tip:* Consider using `facet_wrap()` and wrap them all in a column so that you can easily compare the distributions.
5. (1 pt) Following concepts addressed in the visualization module slide materials create a **scatterplot** to represent the relationship between the two numeric variables.
  - *Tip:* Map your predictor variable to the `x` axis parameter of the function and the outcome variable to the `y` axis and parameter.
6. (1 pts) In a second plot, create the same scatter plot but now use one of the character/factor variables to facet the plot so that you have a scatter plot for each subgroup.

### BONUS:

Consider examining some Pearson's  $r$  correlations for variables of interest.