Stat 21 Homework 1

Your name here Collaborators: [list any collaborators here]

Due: Sept 15, by noon ET

This assignment is due on to be submitted on Gradescope on **September 15** by **12:00pm ET**. Please use the `homework-q-and-a` channel on Slack to post any related questions or error messages.

General instructions for all assignments:

You must submit your completed assignment as a single **PDF** or **HTML** document uploaded to **Gradescope**. For instructions on how to do this, please watch this 2 minute video: https://youtu.be/KMPoby5g_nE. You must use R markdown to write up your solutions. For any homework problems that involve coding in R, you must provide **both** the code and the requested output. You can find a R markdown homework template here: http://www.swarthmore.edu/NatSci/sthornt1/Stat021/Stat21.html. Please make sure each problem is **clearly labeled** and that any handwritten components (such as pictures or equations) are easily readable as pictures within the R markdown document. You may want to use a service like CamScanner (https://www.camscanner.com/) to help you upload handwritten pages.

You are allowed to work with your classmates on this homework assignment but you must disclose the names of anyone you collaborate with at the top of your solutions. Each homework assignment is worth 20 points. One problem will be chosen at random to be graded for correctness and the other problems will be graded for completion. At the end of the semester, your lowest homework grade will be dropped. No homework solutions will be provided.

- Use this file as the template for your submission. You can delete unnecessary text (e.g. these instructions) but make sure to keep the nicely formatted "Problem 1", "Problem 2", "a.", "b.", etc
- Upload your knitted HTML or PDF file to the Homework 1 submission section on Gradescope. Name this file as: [SwatID]_stat11_hw01.html or [SwatID]_stat21_hw01.pdf (e.g. and "sthornt1_stat21_hw01.html"). You only need to upload one file, but please make sure that your graphs, code, and answers to each question appear in the appropriate place. If we cannot see your code/graphs/answers, we cant give you credit for your work!
- Your file should contain the code to answer each question in its own code block. Your code should produce plots/output that will be automatically embedded in the output (.html or.pdf) file.
- Each answer must be supported by written statements (unless otherwise specified).
- Include the name of anyone you collaborated with at the top of the assignment.
- In order to knit this document, make sure you have installed the following packages in your version of RStudio: `ggplot2`, `tidyverse`, `gridExtra`, `gcookbook`, `knitr`

Note: This assignment is shorter than most of your future assignments becasue I want to make sure you have time to get set up with RStudio and getting used to writing up your homework in R Markdown documents.

Problem 1

Install the R package called *swirl* in RStudio by navigating to 'Tools -> Install Packages' and setting "Install From" to "Repository (CRAN)", typing in "swirl" under "Packages", and checking the box "install all dependencies". Next, call this package into your working library by typing `library("swirl")` in the R

console window. Follow the prompts that appear in the console. Select the course option "1: R Programming: The basics of programming in R" and then type in the course option "1". Complete the following lessons:

- 1: Basic Building Blocks
- 2: Workspace and Files
- 3: Logic

Once your have completed the above lessons you can exit the tutorial by typing `bye()` into the R console.

Solution Problem 1:

[Confirm you have run through the specified parts of the tutorial here.]

Problem 2

Respond to the prompt in the "weekly-checkin" channel of our Slack group.

Solution Problem 2:

[Confirm your response here.]

Problem 3

Meet with your assigned group members sometime outside of class using Zoom. Take at least 10 minutes to get to know one another. Answer the following questions about each of your group members and in the space below, include a PNG image of a screenshot of everyone in the Zoom meeting.

Questions:

1. What year are you?
2. Where are you working from?
3. Why are you taking this class?
4. What's your favorite color and why?

Solution Problem 3:

[Write your solution here. You can use the R chunk of code below to incorporate a .jpeg or .png image in this file. Alternatively, you can save the image as a PDF file and then use something like https://smallpdf.com/merge-pdf to merge your homework PDF and your image PDF.]

```
# knitr::include_graphics("file-path/name-of-image.png")
# Un-comment the line above if you would like to include an image here. You may want to play around wit
```

Problem 4

In a large class of introductory Statistics students, the professor has each peron toss a coin 16 time and calculate the proportion of each person's tosses that were heads. The students then report their results, and the professor plots a histogram of these several proportions.

(a) What shape would you expect this histogram to be? Why?

(b) Where do you expect the histogram to be centered?

(c) How much variabilitiy would you expect among these proportions?

(d) Explain why a Normal model should **not** be used here.

Solution Problem 4:

[Write your solution here.]

Problem 5

Census data for a certain country shows that 19% of the adult residents are Latinx. Suppose 72 people are called for jury duty and only 9 of them are Latinx. Does this apparent underrepresentation of Latinx jurors call into question the fairness of the jury selection system. Explain your answer with statistical reasoning.

Solution Problem 5:

[Write your solution here.]

```
## Uncomment this line and put any r-code you used for your solution here
```

Problem 6

A company with a fleet of 150 cars found that the emissions systems of 7 out of the 22 they tested failed to meet pollution control guidelines. Is this strong evidence that more than 20% of the fleet might be out of compliance? Test an appropriate hypothesis and state your conclusion. Be sure the appropriate assumptions and conditions are satisfied before you proceed.

Solution Problem 6:

[Write your solution here.]

```
## Uncomment this line and put any r-code you used for your solution here
```

Problem 7

It is widely believed that regular mammogram screening may detect breast cancer early, resulting in fewer deaths from that disease. One study that investigated this issue over a period of 18 years was published during the 1970s. Among 30,565 people with breast tissue who had never had mammograms, 196 died of breast cancer, while only 153 of 30,131 who had undergone screening died of breast cancer.

Do these results suggest that mammograms may be an effective screening tool to reduce breast cancer deaths? Use appropriate statistical methods to support your answer.

Solution Problem 7:

[Write your solution here.]

```
## Uncomment this line and put any r-code you used for your solution here
```

Problem 8

In July of 2004, the Gallup Poll asked 1005 US adults if they actively try to avoid carbohydrates in their diet. That number increased to 27% from 20% in a similar 2002 poll. Is this what statisticians would call a "statistically significant" increase? Use a difference in proportions test or CI to justify your answer.

Solution Problem 8:

[Write your solution here.]

```
## Uncomment this line and put any r-code you used for your solution here
```