STAT 201: Problem Set 2 (R)

Your name

2/25/25

0. Change your name in the YAML. Then load in the `tidyverse` and `openintro` packages in the code chunk below. We will work with the `starbucks` data from the `openintro` package.

```
# load packages here
```

Get a feel for the data by looking at its Help file and View()-ing it.

1. Write code that creates two variables: one for the average number of calories in a Starbucks food item and another for the standard deviation.

2. Using ggplot, create a histogram of the calories in Starbucks food items. Change the binwidth or number of bins to make a more "pleasing" plot. Add an informative title.

3. Modify your code chunk above so your plot includes:

- A vertical line that displays the mean number of calories. Do this by adding a layer to your plot and using the `geom_vline()` (which stands for *v*ertical line). You can look at the Help file to figure out exactly how to use this function. You should use one of the variables you created in number 1.
- Make your line a color of your choice by specifying the `col` argument within the `geom_vline()` function.
- A caption that provides a brief description of what the line represents.

4. Using the plots you created, along with some summary statistics, describe the distribution of calories in Starbucks food items. Make sure you discuss shape, center, spread, and potential outliers.

Answer:

5. Now, create a scatterplot in ggplot with calories on the y-axis. For the x-axis variable, choose a variable that displays a strong association/pattern with calories. You may have to play around a bit! Add an informative title.

6. In the code chunk above, color the points by another variable in the data set. Then, add a layer of code using the function `scale_color_viridis_c()`. When you run this code, you'll notice that this function changes the color palette to something called the Viridis color palette. Play around with different palettes by looking to the Help file for this function and looking at the *option* parameter. Choose one of the eight options!

7. Briefly interpret this last plot you created. Your interpretation may include discussing associations/trends/patterns (or lack thereof)!

Answer:

Once you're done, render this file one more time, and submit this file to Gradescope alongside the other homework problems.