# CSP 518 - Exercises: 1b - Plotting

*Due: 2019-02-01 @ 11:59pm*

## Instructions

1. Make a .Rmd file to complete these exercises.

- Title it "CSP 518 - Exercises: Intro to R" (in quotes)
- Add your name(s) in the author slot (in quotes)
- The output can be either be `pdf_document` or `html_document` (no quotes)

2. Turn in just your .Rmd file on Blackboard. Note, I will need to be able to `knit` it myself, so make sure everything runs the way you intend, before you turn it in.

3. For clarity, write "## Question " befor each question to signal which one you are answering.

## Exercises

1. Load ggplot2 (which also automatically loads the `mpg` dataset).

2. Create a histogram of the `hwy` variable using base R's `hist()` function.

3. Using `ggplot()`, show how you would build up a plot, going from blank canvas to detailed visualization (i.e., depict each of the steps below, as you add to the plot).

   1. Show me a blank plot.
   2. Add points to that plot, with any two variables you want (dont forget the `mapping = aes()` part of the process).
   3. Choose another numeric variable (not a categorical one) and have the points change size, according to that variable. Hint: this is the same as what we did with color, but call `size = <my-variable>` in the `aes()` function, instead of color).
   4. Change the labels to something more descriptive and publication ready.

— stop here, wait for block 2 of lecture

4. Show me how you would save the output of your plot from #4 as a .png file. Make the file 8 inches wide and 6 inches high. Note, you don't need to actually save it and turn it in, just show me the code you would use.

5. Run the code below to import some new data. All this code does is change the formatting of the data a little to make it easier for you to plot. In future classes, we'll learn why this matters, but for now just trust that I've taken care of some of the more tedious work for you.

```
data('Seatbelts')

seatbelts_df <- as.data.frame(Seatbelts)
seatbelts_df$time <- time(Seatbelts)
seatbelts_df$law <- factor(seatbelts_df$law)
```

6. Make a blank plot with *time* on the x-axis and *DriversKilled* on the y-axis.

1. Fill the plot with points and color the points with the *law* variable. Then draw a vertical line at time = 1983, when the seatbelt law was passed. Does it appear that the law made a difference?
2. Connect the dots with a line. Is there a cyclical trend to traffic fatalities? What do you think accounts for that?

7. Investigate the relationship between the price of petrol and driver fatalities.

   1. Make a plot using `seatbelts_df`. Place *PetrolPrice* on the x-axis and *DriversKilled* on the y-axis. Add some points to the plot.
   2. Add a smoothed line to the data with `geom_smooth()`
   3. Change that smoothed line to a regression line with `geom_smooth(method = 'lm')`. Note "lm" is short for "linear model".
   4. Add a stat_function() to your figure. Try to make it approximate your regression line from above as much as possible. HINT: you will need a function of the form $ax + b$.