HW 3, STAT 450
Due: Thursday, October 28

Reading: Section 11.1, 11.2, 12.1–12.4 from *R for Data Science*

```
library(tidyverse)
```

Exercise 1. Use `read_csv()` to read the data set `hate_crimes.csv` into R (lecture 13). This data set was used for the FiveThirtyEight article *Higher Rates Of Hate Crimes Are Tied To Income Inequality*. A description of the variables can be found at this link:

https://github.com/fivethirtyeight/data/tree/master/hate-crimes

(a) The Gini Index is a measure of income inequality.[1] The Gini Index is between 0 and 1, where values closer to 1 indicate greater income inequality. Which states have the highest Gini Index? Which states have the lowest Gini Index? [Hint: use `arrange()`]

(b) Use `ggplot()` to make a scatter plot with `gini_index` on the x-axis and `avg_hatecrimes_per_100k_fbi` on the y-axis. Use `geom_smooth()` to add a smooth trend line to the scatter plot. Label the x-axis "Gini Index" and the y-axis "Average hate crimes per 100,000 residents". Describe the association between the two variables in the scatter plot, and identify any potential outliers.

Exercise 2

(a) What function would you use to read a file where fields were separated with a semicolon `";"`?

(b) What function would you use to read a file where fields were separated with a vertical bar `"|"`?

Exercise 3. Identify what is wrong with the following inline CSV file. What happens when you run the code?

```
read_csv("a,b,c\n1,2\n1,2,3,4")
```

[1] https://en.wikipedia.org/wiki/Gini_coefficient

Exercise 4. Consider the following data from a Pew religion and income survey.

```
relig_income
```

```
## # A tibble: 18 x 11
##    religion `<$10k` `$10-20k` `$20-30k` `$30-40k` `$40-50k` `$50-75k` `$75-100k`
##    <chr>      <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>      <dbl>
##  1 Agnostic      27        34        60        81        76       137        122
##  2 Atheist       12        27        37        52        35        70         73
##  3 Buddhist      27        21        30        34        33        58         62
##  4 Catholic     418       617       732       670       638      1116        949
##  5 Don't k~      15        14        15        11        10        35         21
##  6 Evangel~     575       869      1064       982       881      1486        949
##  7 Hindu          1         9         7         9        11        34         47
##  8 Histori~     228       244       236       238       197       223        131
##  9 Jehovah~      20        27        24        24        21        30         15
## 10 Jewish        19        19        25        25        30        95         69
## 11 Mainlin~     289       495       619       655       651      1107        939
## 12 Mormon        29        40        48        51        56       112         85
## 13 Muslim         6         7         9        10         9        23         16
## 14 Orthodox      13        17        23        32        32        47         38
## 15 Other C~       9         7        11        13        13        14         18
## 16 Other F~      20        33        40        46        49        63         46
## 17 Other W~       5         2         3         4         2         7          3
## 18 Unaffil~     217       299       374       365       341       528        407
## # ... with 3 more variables: $100-150k <dbl>, >150k <dbl>,
## #   Don't know/refused <dbl>
```

Use the `pivot_longer()` function to make the `relig_income` data set into a tidy data set, with the variables along the columns and observations along the rows. Your code should produce the following output:

```
## # A tibble: 180 x 3
##    religion income             count
##    <chr>    <chr>              <dbl>
##  1 Agnostic <$10k                 27
##  2 Agnostic $10-20k               34
##  3 Agnostic $20-30k               60
##  4 Agnostic $30-40k               81
##  5 Agnostic $40-50k               76
##  6 Agnostic $50-75k              137
##  7 Agnostic $75-100k             122
##  8 Agnostic $100-150k            109
##  9 Agnostic >150k                 84
## 10 Agnostic Don't know/refused    96
## # ... with 170 more rows
```