STAC33

Assignment 7

Due Tuesday March 24 at 11:59pm

To begin:

```
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Optionally, install and load package `broom`. (You don't need this: you can do the assignment without it.)

1. Work through at least some of Chapter 14 of PASIAS. There are lots of problems there. Problems 14.4–14.9 are good practice for the problem you'll be handing in.

   Hand in the next one.

2. The SAT is a standardized test used in the US as part of the college admissions process. Two of the sections of the test are Math and Verbal. Students receive a score on each. Are the two scores related? The data in `http://ritsokiguess.site/STAC33/sat.csv` are Math and Verbal SAT scores for a number of students. The data file also contains the sex of each student, which we will ignore in this question. Our aim is to predict math SAT score from verbal SAT score.

   (a) (2 marks) Read in and display (some of) the data. How many students are there in the data set?

   (b) (2 marks) Make a suitable plot of the two SAT scores for each student. Add a smooth trend.

   (c) (3 marks) What do you see in your plot? Explain briefly. Hint: think about (i) form: linear or curved, (ii) direction: up or down or a mixture of both, (iii) strength: strong relationship, moderate, or weak.

   (d) (2 marks) Fit a linear regression predicting math SAT score from verbal SAT score. Display the output.

   (e) (2 marks) Is there a relationship between SAT verbal and math scores for all students (of whom the students in this data set are a sample)? Explain briefly.

   (f) (3 marks) Obtain a plot of residuals against fitted values. What do you conclude from it? Explain briefly.

3. Work through (at least some of) the remaining problems in Chapter 14 of PASIAS. 14.1–14.4 are multiple regressions; 14.10 has a categorical variable in it, and the remaining problems are a mixture.