Problem Set 1

YOUR NAME HERE

due 9/8/21

For this assignment, you'll be working with the College Scorecard (debt) dataset to predict the debt of college graduates using conditional means. You'll need to select the college-level characteristics that you think might be related to eventual debt.

Structural stuff:
1. Be sure to change the "author" above to your name.
2. Save your .Rmd file as LastName_FirstName.Rmd (do this before you knit).
3. You need to submit your .Rmd code file AND a knit file (upload both simultaneously to the course webpage; you can't upload them one-by-one). You will only receive full credit if you upload both files.
4. Below I have set up the file for you with the library you'll need and I have included code to read in the dataset (note, I won't do this every time).
5. I expect that the .Rmd file you submit will run cleanly, and that the knit file won't contain any errors (LOOK at the knit file after you create it - if questions/text are running into each other, if you see error messages, etc., you're not done).
6. You can use comments to tell me what you are doing either in text or in code chunks, but remove "old" code that didn't run/work.

```
knitr::opts_chunk$set(echo = TRUE)

library(yardstick)
```

```
## For binary classification, the first factor level is assumed to be the event.
## Use the argument `event_level = "second"` to alter this as needed.
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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

Load the data here (I provide the code this time, but I won't always).

```
collegedebt<-readRDS("sc_debt.Rdata")
```

1. Calculate the mean of the outcome: `grad_debt_mdn`

2. Use your mean as a prediction: Create a new variable that consists of the mean of the outcome.

3. Calculate a summary measure of the errors for each observation—the difference between your prediction and the outcome.

4. Calculate the mean of the outcome at levels of a predictor variable of your choosing.

5. Use these conditional means as a prediction: for every college, use the conditional mean to provide a "best guess" as to that college's level of the outcome.

6. Calculate a summary measure of the error in your predictions.

7. Repeat the above process using the tool of conditional means, try to find 3-4 combined variables that predict the outcome with better (closer to 0) summary measures of error. Report the summary measures of error and the variables as text in your `.Rmd` file.