Problem Set 1

Intro to `R`

[YOUR NAME]

Due Date: 2024-01-19

Getting Set Up

Open `RStudio` and create a new RMarkDown file (`.Rmd`) by going to `File -> New File -> R Markdown...`. Accept defaults and save this file as `[LAST NAME]_ps1.Rmd` to your `code` folder.

Copy and paste the contents of this `.Rmd` file into your `[LAST NAME]_ps1.Rmd` file. Then change the `author: [Your Name]` to your name.

We will be using the `sc_debt.Rds` file from the course github page (https://github.com/jbisbee1/DS1000_F2023/blob/main/data/sc_debt.Rds).

All of the following questions should be answered in this `.Rmd` file. There are code chunks with incomplete code that need to be filled in.

This problem set is worth 8 total points, plus two extra credit points. The point values for each question are indicated in brackets below. To receive full credit, you must have the correct code. In addition, some questions ask you to provide a written response in addition to the code.

You are free to rely on whatever resources you need to complete this problem set, including lecture notes, lecture presentations, Google, your classmates…you name it. However, the final submission must be complete by you. There are no group assignments. To submit, compiled the completed problem set and upload the PDF file to Brightspace on Friday by midnight. Also note that the TAs and professors will not respond to Campuswire posts after 5PM on Friday, so don't wait until the last minute to get started!

Good luck!

*Copy the link to ChatGPT you used here: _____

Question 0 [0 points]

***Require** `tidyverse` **and load the** `sc_debt.Rds` **data by assigning it to an object named** `df`.*

```
require() # Load tidyverse
```

```
## Loading required package:
```

```
df <- read_rds() # Load the dataset directly from github
```

```
## Error in read_rds(): could not find function "read_rds"
```

Question 1 [2 points]

Which school has the lowest admission rate (`adm_rate`) and which state is it in (`stabbr`)?

```
df %>%
  arrange() %>% # Arrange by the admission rate
  select() # Select the school name, the admission rate, and the state
```

```
## Error in df %>% arrange() %>% select(): could not find function "%>%"
```

Write answer here

Question 2 [2 points]

Which are the top 10 schools by average SAT score (`sat_avg`)?

```
df %>%
  arrange() %>% # arrange by SAT scores in descending order
  select() %>% # Select the school name and SAT score
  print() # Print the first 12 rows (hint: there is a tie)
```

```
## Error in df %>% arrange() %>% select() %>% print(): could not find function "%>%"
```

Write answer here

Question 3 [2 points]

Create a new variable called `adm_rate_pct` which is the admissions rate multiplied by 100 to convert from a 0-to-1 decminal to a 0-to-100 percentage point.

```
df <- df %>% # Use the object assignment operator to overwrite the df object
  mutate() # Create the new variable adm_rate_pct
```

```
## Error in df %>% mutate(): could not find function "%>%"
```

Question 4 [2 points]

Calculate the average SAT score and median earnings of recent graduates by state.

```
df %>%
  group_by() %>% # Calculate state-by-state with group_by()
  summarise(sat_avg = , # Summarise the average SAT
            earn_avg = ) # Summarise the average earnings
```

```
## Error in df %>% group_by() %>% summarise(sat_avg = , earn_avg = ): could not find fun
ction "%>%"
```

Extra Credit [2 points]

Plot the average SAT score (x-axis) against the median earnings of recent graduates (y-axis) by school, and add the line of best fit. What relationship do you observe? Why do you think this relationship exists?

```
# INSERT CODE HERE
```

Write answer here