Exam 1

Thanks to Drs. James Dickens, Richard Ressler, and David Gerard

Problem:

Read the help page of the `gss_cat` data frame from the forcast package.

   i. What are the variables?

   ii. What is the class of `gss_cat`?

   iii. What is the class of each variables?

   iv. Reorder the levels of the 'relig" variable so that the levels are in alphabetical order. write a code that shows the order has been changed. Change order to descending.

   v. Find the frequency of each categories.

   vi. Put levels in descending order of how frequently each level occurs in the data

   vii. Modify the factor levels of `marital` to be abbreviations of their long-names. For example, "Divorced" can just be "D"

Problem:

The first two numbers of the Fibonacci Sequence are 0 and 1. Each succeeding number is the sum of the previous two numbers in the sequence. For example, the third element is $1 = 0 + 1$, while the fourth elements is $2 = 1 + 1$, and the fifth element is $3 = 2 + 1$.

   i. Use a for loop to calculate the first 100 Fibonacci Numbers.

   ii. Return the first 15 Fibonacci Numbers.

   iii. Write a code that finds the nth Fibonacci Number. What is the 30th Fibonacci Number?

   iv. Sanity Check: The $\log_2$ of the 100th Fibonacci Number is about 67.57.

Problem:

Load the `wmata_ridership` data frame into R from https://dcgerard.github.io/stat_412_612/data/wmata_ridership.csv.
i. Save the data in your local machine in your working directory (use write_csv()). ii. Upload it into R (use read_csv() and relative path) and name it `wmata`. iii. what are the variables?
iv. Separate variable `Date` to `year`, `month` and `day`.
v. For each month, calculate the proportion of rides made on a given day of the month.
vi. Then make box plots of the proportions of ridership vs day of the weak. But exclude any days from 2004.

Problem:

i. Create a new repository in Github. Name it `repositoryexam_1`

ii. Drag and Drop 3 files from your desktop to your new repository (any files that you think is appropriate)

iii. Take a screenshot of the created repository showing evidence of the three files uploaded.

iv. Now go to the bash terminal and clone the repository back to your Descktop

Problem:

i. Type your `PAT` token

ii. Push the `exam_1` file (without solution) to your `repositoryexam_1`
iii. Take screenshot and post the url of your Github page that shows the file being pushed along with the commit message "Add exam 1 problems set"

Problem: You must use data.table package to solve this problem. If you solve it otherwise for instance by tidyverse you get a zero.

Use the `nycflights13` package

i. Add the full airline names to the `flights` data.table.
ii. Use data.table to calculate the median air time for each month.
iii. Use data.table to calculate the number of trips from each airport for the carrier code DL.

iv. Calculate the mean departure delay for each origin in the months of January and February.

Problem:

(from OpenIntro): The 2010 General Social Survey asked 1,500 US residents: "Do you think the use of marijuana should be made legal, or not?" 35% of the respondents said it should be made legal. a. Is 35% a sample statistic or a population parameter? Explain. b. Construct a 95% confidence interval for the proportion of US residents who think marijuana should be made legal, and interpret it in the context of the data.

Problem:

Read up on the `ex0330` dataset from the Sleuth3 R package. Determine if education level is associated with income. Interpret any estimates and confidence intervals you derive.