Week 7 Tutorial Worksheet

AY22/23 Semester 2

Question 1. College majors and earnings

In this question, we will work on `college.csv` to explore the relationship between college majors and earnings. This is a *modified* version of a data set from TidyTuesday (https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-03-10/readme.md).

Download and use the data file from Canvas. The link above is only for your reference. Here are the definition of the variables:

Variable	Definition
`rank`	Rank by median early-career pay.
`major`	Name of the college major.
`major_category`	Category of the major.
`major_stem`	Whether the major is considered a STEM field.
`total`	Total number of graduates of this major.
`sharewomen`	Women as a share of total graduates.
`unemployment_rate`	Number of unemployed graduates/total graduates.
`median`	Median early-career pay of employed graduates.

1. Read the CSV file into `R` as `college`. Remove the rows with missing values (if any). Convert the following columns to factor: `major_category` and `major_stem`. Create new column, `num_women`, as the number of female graduates of each major.

2. Create a data frame `qn1_2` that stores the top three highest paying major categories.

3. Create a data frame `qn1_3` that stores the top 10 majors with the highest early-career pay.

4. Create a data frame `qn1_4` that contains information on the ranking, the name of the major, and the unemployment rate the top 10 majors with the highest unemployment rate. Round the variable `unemployment_rate` to two decimal places.

5. Explore the data. **Create a graph** to answer one question you find interesting about the data. Include the code you use. After that, summarize (in 50 – 100 words) what you found in an `Rmd` text section entitled "Question 1.5".

Question 2. An `Rmd` file to fix

A classmate has sent you the file `rmd-to-fix.Rmd` (click to download), which is an `R` Markdown document analyzing the data in Question 1. They could not knit the file to HTML, and are providing you with a raw `Rmd` file instead.

1. Go through the file and fix things that are preventing it from knitting.
2. Change the lines of bold-font text to `Rmd` text sections entitled "Question 1.1", ..., "Question 1.5" to organize the document.
3. Does the file generate **all** required objects? Make necessary changes if this is not the case.
4. Is the plot explanation written in an `Rmd` text section of the required title? Make necessary changes if this is not the case.

Before sending the modified `R` Markdown file back to the classmate, make sure that you can render it into HTML. To do so, clear your current RStudio Environment and then hit the "Knit" button again.

Requirements

- Data frames (tibbles): `college`, `qn1_2`, `qn1_3`, `qn1_4`.
- A graph that answers the question you find interesting about the data.
- A brief summary of the your findings under a text section entitled "Question 1.5".
- Able to fix the problems in the `Rmd` file.

Note: Make sure you understand and follow all requirements. In an exam, we would deduct marks if we are not able to run/knit your `Rmd` file. Among other reasons for marks deduction:

- Data set not read in.
- Syntax errors in the code.
- Required objects not found within the code.
- Explanations/discussions not written under `Rmd` text sections of the required title.