Data wrangling in R

This lab is preconfigured to include all dependencies (libraries, packages, and datasets) you'll need to complete your work in RStudio. You can practice, run test cases, and work on assignments from your browser.

Assignment Overview

The purpose of this lab is to assess your proficiency in utilizing R for fundamental data wrangling tasks, which serve as a crucial skillset for upcoming data visualization tasks. The data for this lab can be found in Module2 "Salaries.csv".

The basic steps should be included in this lab:

1. Install and load required packages.
2. Load the dataset using read_csv().
3. Select columns of "rank", "discipline", "sex", "salary".
4. Filter the rows where "salary" is not missing value.
5. Group based on the required columns ("rank","discipline", "sex") and have the tally count of the rows.

   Use select(), filter(), group_by(), count() in this part and combine them with pipe operator.

6. Reshape the table from long to wide: using spread() to split the column "sex" in the data frame you obtained in Steps 3-5 to many new columns. Provide values for the new columns based on "n" (tally count).
7. Reshape the table from wide to long: using gather() to make the two sex columns (from Step 6) into one column and name the column as "sex". Please have the values in the two columns into one column and name the column as "n" (tally count).

Grading Criteria

This week, your .R code will be autograded on the following elements:

1. Your code should match the sequential operations required by the instructor:

- Install and load required packages
- Load the dataset using read_csv().
- Select columns of "rank", "discipline", "sex", "salary".
- Filter the rows where "salary" is not missing value.
- Group based on the required columns ("rank","discipline", "sex") and have the tally count of the rows.

  Use select(), filter(), group_by(), count() in this part and combine them with pipe operator.

- Reshape the table from long to wide: using spread() to split the column "sex" in the data frame you obtained in Steps 3-5 to many new columns. Provide values for the new columns based on "n" (tally count).
- Reshape the table from wide to long: using gather() to make the two sex columns (from Step 6) into one column and name the column as "sex". Please have the values in the two columns into one column and name the column as "n" (tally count).

2. Your code should be run successfully.
3. You should provide comments for each step.
4. You should submit a .R script file in your lab for grading.

How to Submit Your Work for a Grade

- **If you're working in the In-Browser RStudio:** When you've completed your lab, please be sure that your submit your .R file is in the Module 2 folder in your lab. From there use the "Submit Assignment" button in your lab's upper toolbar to submit your code for autograding.
- **If you completed your work in a local Desktop verison of RStudio:** Delete the existing "Data Wrangling in R.R" starter file from your lab and upload your final Data Wrangling in R.R File to the Module 2 folder for autograding using the "Upload -> Choose File" button in your RStudio "Files" tab. Your target directory should be "~/Module2".

For both options, you'll see a final grade and feedback for your work in the Programming Assignment item you launched this lab from. Once you've completed your submission and received a passing score, you can close this RStudio Lab.