Final Workflow Assignment

PSYC 259

Winter 2021

Assignment Overview

The goal of this assignment is to take an existing workflow improve its 1) efficiency, 2) fidelity, and 3) sharing/reproducibility. You will write a report that describes the changes you made and their expected effects. This report should use R Markdown to demonstrate changes you made to code, the results of error checking and EDA, and figures made from your data. Your final project should be shared in a well-organized GitHub repository that contains data files, scripts, figures, and the R Markdown report.

The final project is due on March 17 at 1:30pm. You should make your Github repository public and turn in the URL of the repository on Canvas.

Github repository

Before you start to make changes to your project, set up a Github repository with your current project files so that we can see the "before". Because the focus of the course has been on data wrangling and exploratory analysis, your project should contain raw data files, script(s) that import and clean the data, and scripts that explore the data and/or check for errors. If you have raw data that contains sensitive information, you can deidentify the data or make a few "fake" data files that have the correct format/structure so that we can understand the project. As you work on your project and make changes, commit and push them to Github so that we can see what things were changed. If you've already started making changes and they're not tracked in Github, let us know and we can try to figure out a way to see the "before" version of the project. If your project uses a coding language other than R for data import/cleaning/checking, that's OK. Let us know in advance so that we can make sure we have a way to see/understand what your project does. No matter what language(s) you use in the project, everyone will need to use R to create the R Markdown report with figures.

What changes should you make?

Based on your self-critique assignment and our feedback, you should make changes to improve the efficiency, fidelity, and sharing/reproducibility of the project. You don't need to cover these three categories equally. If improving the efficiency is most important to your project, it makes sense to focus on that. You can describe the reasoning behind the changes you chose to make in the report. Of course, some changes might fall into multiple categories (e.g., automating reading data from a file is more efficient and is less likely to lead to mistakes). Your final project should contain at least one example of the following items:

- Wrote a custom function to reduce code repetition, or split long scripts into more manageable files

- Used some form of automation (map, for loop, across, read_csv) to replace repetitive code or manual data entry/copy/paste
- Plotted graphs or created data checks to explore the data
- Improved documentation of the project, file organization, and/or readability of the code

R Markdown report

Use R Markdown to prepare a report that summarizes the work you did to improve your project. Because everyone's changes will be different, you should structure the report in a way that makes sense to you. For each change that you made, please be sure to explain *how* you did it and *why* you did it. Render your R Markdown file as a PDF or HTML so that we can view it easily.