Final Workflow Assignment

PSYC 259

Winter 2021

Assignment Overview

The goal of this assignment is to take an existing workflow and improve its 1) efficiency, 2) fidelity, and 3) sharing/reproducibility. There are two parts of the project. The first part is a Github repository that tracks changes you made to your project and shows how your new workflow is set up. The second part is a written report that describes the changes you made, why you made them, and what benefits each one should create. This report should use R Markdown to illustrate changes you made to code, the examples of error checking and EDA, and figures made from your data. All of the files, including data files, scripts, figures, and the R Markdown report, should be stored on the Github repository.

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. There is no longer a presentation associated with this assignment during Week 10. Instead, we will spend Week 10 helping you figure out how to implement changes to your project and creating the Github repository.

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. *Please do not share anything publicly on Github that you are not allowed share*. Another option is to tell Github to ignore the folder containing raw/sensitive data and share a screenshot with us showing how the raw data are formatted. Or, you can create a private repository and share it directly with the instructors. If in doubt, ask, and we can figure out something that works for your project.

As you work on your project and make changes, commit and push the changes to Github so that we can see what you 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, focus on that. 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). Please describe the reasoning behind the changes you chose to make. Your final project should contain at least one example of the following items:

- Wrote a custom function to reduce code repetition and/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

We don't expect the final project to be perfect (whatever that means!). It would be impossible to make every change you might want to make in a few weeks. You can, instead, demonstrate in one section of the code how you used automation to improve the workflow, and describe how you might apply a similar technique in other sections of your project. Prioritize making changes that will help your project the most and help you learn skills you want to learn.

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. Make use of R Markdown code chunks to illustrate the before and after of a change you made to code (not every change, but changes that are representative of the type of change). You are encouraged to embed figures to illustrate the results of your data checking and exploration. Render your R Markdown file as a PDF or HTML so that we can view it easily.