PSYC 259: Final Project

Goals of final project

- Take an existing project and:
  1. Improve its efficiency
  2. Improve its fidelity
  3. Improve its reproducibility and capacity for sharing
- You *do not* need to cover each of these equally
  - e.g., if improvements to efficiency are most important to your project, spend more time on that part
- It is possible (and likely) that changes will cover more than one of these goals
  - e.g., automating reading data is efficient (#1) and reduces error (#2)

Goals of final project

- Your final project should contain *at least one* of the following:
  1. Custom function to reduce repetitive code and/or split long scripts into more manageable files
  2. Automation (e.g., map, for loop, access, read_csv) to replace repetitive code or manual data entry/copy & paste
  3. Explore the data through plotted graphs or created data checks
  4. Documentation of project, organization of files, or readability of code improved

Goals of final project

- We do not expect you to make every single change that you might like to make to the workflow
  - You have only a few weeks for this project
  - Point is to apply what you have learned in class
  - Perfectly ok (and encouraged) to focus on improving one section of your workflow and then explain how you might use similar techniques in other sections of the workflow
  - Prioritize changes that will improve your workflow the most and help you learn skills that you want to learn

Two parts to project

1. Github repository to show changes you made to workflow
2. Written report to explain changes you made to workflow

Part 1: Github repository

- Set this up similar to problem sets
  - Private repo
  - Share with OST + Madison by adding us as collaborators

Part 1: Github repository

- In addition:
  - Repo needs to show “before and after”
  - So, upload **raw** data files
  - And scripts that import and clean raw data files
  - And scripts that explore the data and check for errors
  - Caveat: If your raw data contain sensitive information, please de-identify the data before uploading OR simulate a data file that has the same format/structure so that we can understand the project

Part 1: Github repository

- Finally:
  - For us to see the changes you make, you will need to commit and push changes to your repository as you work on your final project
    - One way to do this is to simply commit and push all changes you make during each work session you devote to the project
  - If you already started making changes and did not track them on Github, please reach out to us, and we will work on a way to try to see the “before” version
  - You are allowed to use a coding language other than R for this part
    - Please let us know in advance so that we can make sure to have a way to understand what your project does

Part 2: Written report

- Whereas the Github repo will let us see *exactly* what changes you made, the written report will explain to us *why* you made them
- Thus, the written report should:
  1. Describe the changes you made
  2. Explain why you made those changes
  3. Explain how each change should benefit your workflow/project

Part 2: Written report

- We anticipate that everyone's changes will be different, so we expect that you will structure the written report in a way that makes sense for your project
  - For each change that you make, please explain how and why you did it
  - Code chunks in Quarto will be useful for illustrating changes you made to code (before and after)
    - But no need to show *every* code chunk you changed
    - Just show changes that are representative of the types of changes you made
  - We also encourage you to embed figures that illustrate the results of your data checking and exploration

Part 2: Written report

- Formatting:
  - The report should be written in Quarto and, thus, in R (regardless of what language you use for Part 1)
  - The report should:
    1. Illustrate changes you made to code (e.g., with screenshots)
    2. Examples of error checking
    3. Examples of exploratory data analysis (EDA)
    4. Figures made from your data
  - The report should also be uploaded to your Github repository from Part 1 of the final project as a PDF or HTML file (so that we can read it more easily)

Questions?