

# PSYC 259: Principles of Data Science

Prof. John M. Franchak

Winter 2021

E-mail: [franchak@ucr.edu](mailto:franchak@ucr.edu)

<https://github.com/PSYC-259-Data-Science>

Office Hours: By Appointment

Class Hours: W 1:30-4:20 PM

<https://ucr.zoom.us/my/johnfranchak>

<https://ucr.zoom.us/j/97985633776>

---

## Course Description

Most quantitative courses (importantly) focus on the final steps of data analysis—conducting and understanding statistical tests. However, much of the work in data science is taking raw data, often from multiple, incompatible sources, and processing those data into a usable form. This course will emphasize the importance of robust, documented, and automated workflows for processing data to save time, reduce errors, improve reproducibility, and facilitate collaboration among multiple researchers. We will also spend time on data visualization and communication—an important part of creating, checking, and collaborating on data workflows. We will use the R programming language, Github, and Rmarkdown to work through examples, but the focus is on concepts/best practices that can be applied to any software or programming language. The course is open to students who have little programming experience or experience with R. The goal is for students at all levels of programming experience to set goals to improve their data science skills.

## Course Objectives

The goals of this course are for you to critically analyze and improve your data analysis workflows. Implementing robust, automated procedures for handling data will allow you to:

- Foster open science through increased transparency, reproducibility, and easier data sharing
- Increase the fidelity of your data and analyses by detecting and preventing errors
- Better understand and communicate about your data
- Save time by preventing errors, automating tasks, and reusing code
- Facilitate collaboration with organized and documented workflows

## Course Materials

- The course [Github page](#) has links to project files, readings, and the syllabus
- We will work through examples and complete assignments through a course workspace on [RStudio Online](#). To sign up, follow this [link](#). You will need to create a free user account. We will go over how to access materials on the first day of class.
- Readings from *R for Data Science* (Wickham & Golemund, 2017) are available [online](#).
- PDFs of other course readings referenced below are available on Github.

## **Course Policies**

TBD

## **Assignments and Grading**

## **Remote Instruction**

## Class Schedule

Readings should be completed prior to each class.

### Week 01, 01/06: Course Goals

- Goals of the class
- Logistics (RStudio/Github)
- Readings and assignments

### Week 02, 01/13: Data Workflow

- File organization
- Version control
- Workflow examples [John's data "Show and Tell"]
- *SKILLS*: Basic I/O functions, using Github
- **READING**: Minimizing Mistakes (Rouder, Haaf, & Snyder, 2019)

### Week 03, 01/20: Automation

- Copy/paste is evil (automating data pipeline)
- Drop-down menus are evil (automating your analyses)
- Machine-readable formats
- *Skills*: Basic scripting, iteration, merging data

### Week 04, 01/27: Data "Show and Tell"

- Present your current workflow
- Identify places for potential errors
- Identify places to increase efficiency/automation

### Week 05, 02/03: Data Structure

- Data types
- Data organization (variables/observations)
- **READING**: Tidy Data (Wickham, 2014)

### Week 06, 02/10: Error Checking

### Week 07, 02/17: Encapsulation: Reusing Functions

### Week 08, 02/24: Reproducible Reports and Documentation

(Klein et al., 2018)

### Week 09, 03/03: Visualization

### Week 10, 03/10: Project Presentations

## References

- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., ... Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4(1), 1–15.
- Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science*, 2(1), 3–11. doi:[10.1177/2515245918801915](https://doi.org/10.1177/2515245918801915)
- Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23.
- Wickham, H., & Grolemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc."