

Overview

Course Information

Instructor: Professor Jack Reilly

Office: Eggers 225F

Office Hours: Tuesday and Thursday, 11 AM - Noon

Phone: 315-443-2687 (office)

e-mail: jlreilly@syr.edu

Overview

This course introduces students to data management, wrangling, and visualization as well as the technical tools necessary to do such analysis in an open and reproducible fashion.

Description

Data preprocessing, wrangling, and management often consumes a large fraction of the time spent doing quantitative data analysis in public administration, public policy, and the broader social sciences. This class introduces students to the technical tools necessary to do these tasks in an open and reproducible fashion suitable for modern computational data workflows. Throughout the course of the semester, students will learn the principles and practice of conducting reproducible quantitative research, including readable programming and coding, version control, methods of documentation, data storage, workflow management, and exploratory data visualization. A variety of relevant open technical software tools will be introduced and used, including but not limited to R (and RStudio), git (and github), markdown, LaTeX, and a variety of helper programs to tie things together.

Prerequisites

No formal pre-requisites. It is assumed you have either previously taken or are currently enrolled in an “Introduction to Statistics” or “Quantitative Methods” class (ie, PAI 721 or MAX 201).

Materials

Books

- Required:
 - Weidmann, Nils. *Data Management for Social Scientists*. Open access: <https://doi.org/10.1017/9781108990424>
 - Healy, Kieran. *Data Visualization: A Practical Introduction*. Open access: <https://socviz.co>
- Recommended: A book on R programming or data wrangling
 - *Recommended*: Braun & Murdoch, *A First Course in Statistical Programming*, 3rd Edition
 - *Other options*:
 - * Freeman & Ross, *Programming Skills for Data Science*
 - * Hadley Wickham, Garrett Golemund, and Mine Çetinkaya-Rundel, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, 2nd ed. <https://r4ds.hadley.nz>.

i A Programming Book

There are, in essence, three kinds of books that are useful for the class - a book on data management, a book on data visualization, and a book on data programming. For the first two, I've required the open access books by Weidmann and Healy (**DMSS** and **DV**, respectively). For the third, you have options. I recommend Braun and Murdoch (**FCSP**), which is a good general overview of the R language from a statistical programming perspective. Freeman & Ross (**PSDS**) is a more general introduction to the overall data science technical environment, and Wickham et al (**RDS**) has the advantage of being open access (always useful). **PSDS**, however, doesn't have as detailed information on actual programming in it, and **RDW** but is primarily focused on the tidyverse, which our course is not exclusively focused on. We will cover all programming and scripting content in class, so the book you choose for background reference is up to you, but you will find it valuable to have one.

Computing

You will need access to a personal computer for this class. It will need to run a full operating system, where you have the ability to install local applications outside of app stores and have access to the command line. MacOS, Windows, and Linux are all fine. Tablet or web-book OSes - like Chromebooks or iPads - won't be sufficient. Aside from the computer, all significant software we use will be free/open-source, and we'll cover usage and installation in class.

Online Course Resources

Blackboard is our internet-based course platform: <http://blackboard.syr.edu>. In it, you will find submission portals for assignments and a link to this course webpage, where you can find the course syllabus, problem sets, and links to readings.

You can also link to our [course drive here](#), which contains lecture slides, data sets, and some other useful things for the class.

Please note that class attendance is the primary source of course-related announcements and material.

Course Requirements

Overview

Satisfactory completion of the course requires completion of the following:

- Regular course participation and attendance (10%)
- Weekly Assignments (30%)
- Core Exam (30%)
- Final Project (30%)

Attendance

One of the guiding principles of my class is that you are adults, and thus, capable of managing your own time. I have little interest in policing your lives. Attendance is kept for each day of class, but you will lose no points on attendance if you happen to miss a couple days: everyone has things that occasionally come up in life that need to be dealt with, and I fully realize that some of those things are things you - very understandably - may not want to discuss with your professor. *That's OK!*

That said, attendance in class is an important element to doing well in the course. If you must miss more than a couple days, it's a good idea to check in with me so that I don't mark you

off for chronic absenteeism. The easiest way to do this is just email me with a brief reason when something comes up and you have to miss class (which will also allow me to tell you if you're missing anything particularly important).

If you must miss class, the way to make up what you've missed is straightforward: make sure to look over the posted material, do the reading, get notes from a friend, and still complete the assignment if you are able (and make sure to look over any assignment solutions). If you do these things and still feel like you're missing something, please feel free to come into my office hours and we can talk it through.

i Participation

There is no formal grade for "participation". However, I reserve the right to dock a couple points here if you do ridiculous/unprofessional things in class (like answering your cell phone, always coming in late and regularly distracting others, spontaneously breaking out into ribald song in the middle of class, etc).

Assignments

There is an assignment each week in class, **due Thursday by class time**.

Assignments will vary in nature: some will be one-off problem sets, some may build on problem sets from a prior week. All material needed for an assignment will be covered by the Tuesday before the assignment is due (usually much earlier), and the assignment itself will be given a week ahead of time. **No assignment work is accepted after class**, as we will go over answers for assignments in class.

Core Exam

The core exam will have in-class and out-of-class components. More information will be given as the exam gets closer.

Final Project

A project utilizing data of your own choice. Graduate students will have higher expectations than undergraduate students.

Course Expectations & Guidelines

Etiquette & Decorum

A university course is fundamentally a learning community. Be courteous to fellow students and the professor. Don't let yourself be distracted by your cell phone in class, and don't let what is on your computer screen distract fellow students in the class, either.

This is a graduate course: I take it for granted that you have a basic interest in the material, an enthusiastic attitude toward participation, and a respectful attitude to everyone in the room.

Office & Consultation Hours, Appointments

I encourage you to chat with me at any point if you have questions about the course. You can schedule a meeting with me by going to my website here: <http://jacklreilly.github.io> and sign up for time at your convenience. You can also always just drop in during my regularly scheduled drop-in office hours without appointment.

E-mail

Email is the best way to contact me. I'm usually pretty responsive, but as a baseline, I always aim to get back to you in a modified 24-hour fashion: by the end of the business day the day after you email. So if you email me at 2 PM Tuesday, I'll get back to you by 6 PM Wednesday; if 10 PM Thursday, by 6 PM Friday; if you email me at 3 PM on Friday, by 6 PM Monday, etc.¹

Note

If your email requires a long response, expect me to encourage you to schedule an appointment with me so that we can more effectively discuss the matter.

¹Again: usually I'm much faster! But if you don't hear from me by this baseline, feel free to bump a reminder.