Syllabus

Course Information

Instructor: Professor Jack Reilly

Office: Eggers 225F

Office Hours: TBD

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 techinal 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:
    - Braun & Murdoch, *A First Course in Statistical Programming*, 3rd Edition

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 course announcements, a submission portal for (some) of your assignments, and a link to this coruse webpage, where you can find the course syllabus, problem sets, and links to readings. In addition, during the semester, solutions to the problem sets and lecture slides will be posted. Please note, however, 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 slides, do the reading for the day, get notes from a friend, and still complete the assignment if you are able (or, if it is a day we are going over the assignment, make sure to look over the 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.

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 loudly, 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.

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]

> **i** 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.

Artificial Intelligence, LLMs, etc

You can find the formal AI policy for this course in the "Policies" section of the syllabus/website. Because of the nature of this course, however, a deeper discussion of AI is warranted.

Artificial intelligence – in particular, large language models (LLMs) like ChatGPT – are incredibly useful tools. However, in the very beginning phases of learning how to write code and conduct data analysis, using them can hamper your ability to understand fundamental elements of programming. Accordingly, I **strongly encourage** you to **avoid** using generative-AI tools in this course (except where I introduce them in the content of the class).

> **💡** Tip
>
> LLMs are, in a formal sense, **bullshit** of the first order. And yes: "**bullshit**" here is a technical term, meaning "language produced without regard for truth."[2] An LLM, at its core, is a probabilistic model whose primary function is to produce the next token – usually a word – that is most likely to make sense, based upon some prompt. It **does not care if this token represents truth, reality, or a functional line of code** – it only cares if the next token *seems* to make sense. Frequently, what it produces will incidentally be true, functional code, or accurate representation. That is what makes it useful! But this *incidental* relationship with truth is also what makes it dangerous. You can't reliably evaluate AI outputs until you already know enough about the subject.

My job, goal, and interest in this class is to help **you, a human, learn how to code, wrangle, and visualize data**. My interest is not in correcting AI slop; nor does my interest lie

---

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

[2] Frankfurt, *On Bullshit* (2005); Hicks, Humphries, and Slater, *ChatGPT is Bullshit* (2024).

in trying to figure out if you are attempting to pass AI slop off as your own work.[3] Accordingly, I will not spend time trying to guess if your assignments are AI generated.[4] As always, you are responsible for your own learning and work - as your professor, I simply ask that you not try to outsource the important process of your own learning with a literal philosophical bullshit machine.

---

[3]AIs have their own reinforcement learning processes, with humans employed to correct their foibles and mistakes – an unfortunately unending and unenviable task, like Sisyphus with his rock.

[4]The technical tools that try to do this are imperfect, anyways, resulting in a lot of uncertainty and false positives. Furthermore, experience has taught me that typically, unedited-AI produced content will earn poor marks, regardless, and often do so with some pretty clear "tells".