PAI 400/600: Data Wrangling and Visualization

Course Syllabus

F2025

Course Information

Class

Time: Tuesdays and Thursday, 2:00 PM - 3:20 PM

Location: Eggers 225B

Instructor

Professor Jack Reilly

Office: Eggers 225F

Office Hours: Tuesday and Thursday, 11 AM - Noon (drop in) and by appointment.[1]

Zoom Hours: Friday, 1-3 (schedule online)

Phone: 315-443-2687 (office)

e-mail: jlreilly@syr.edu

---

[1]In addition to formal office hours, I have an open door policy - feel free to stop by if you see the door open. In particular, I'll be around for the half an hour after class each Tuesday and Thursday until 4, and my office is right next door to the classroom. Feel free to stay after and meet with me. I'll also be around after office hours until at least 12:30 each Tuesday and Thursday according to student need.

Course Description

Every step in policymaking relies on data. This course introduces students to data management, wrangling, communication, and visualization in the context of public policy, public administration, and behavioral science as well as the technical tools necessary to do such work in an open and reproducible fashion.

Expanded 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 behavioral science research. Yet these topics frequently do not receive regular attention in methodological courses that focus on statistical inference. 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 in the public sector. 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, and a variety of helper programs to tie things together. Special attention will be paid to data frequently used in public policy, public administration, and behavioral science.

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), and are conversant enough in statistics to be able to work with concepts like "mean" and "standard deviation".

While this course has no *formal* pre-requisites, it does have a substantial *informal* prerequisite: **motivation**. Learning a programming langauge is challenging work, and students must be prepared to invest the appropriate time, energy, effort, and - *above all* - patience.

Learning Objectives

1. Demonstrate capability in open science and contemporary reproducible data processing, wrangling, and management tools
2. Apply appropriate principles of data and file management to data projects
3. Evaluate the credibility and clarity of data visualizations

4. Create effective, reproducible, and well designed data visualizations with appropriate tools
5. Analyze large-N datasets commonly used in public policy and behavioral science

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. *Purchase links:* cambridge amazon
  - *Other options:*
    * Freeman & Ross, *Programming Skills for Data Science*
    * Hadley Wickham, Garrett Grolemund, 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 perspecitve. 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 the 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%)
- Practicum (Take Home) (15%)
- Core Exam (In Class) (15%)
- 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.

Students may miss up to two assignments with no penalty. Students may also work together on assignments - in fact, I encourage you to do so - although each student is ultimately responsible for their own learning and work.

> **ℹ Effort-Based Evaluation**
>
> Assignments are evaluated based upon effort and a check completion system. Students who answer every problem will earn a check, with each check worth one point toward their final assignment grade.

Practicum

A practicum is a large assignment that is worth more and graded on a scale.[2] It is untimed, take-home, cumulative, and will be completed on your own time (and computer). Unlike the weekly assignments, you are not allowed to work together on the practicum. Essentially, think of it as take home test that complements the in-class core exam.

Core Exam

The core exam will be in class. 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. More information will be given as the exam gets closer.

Course Expectations & Guidelines

Etiquette & Decorum

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. A university classroom 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 or stop by to see if my door is open: if it's open, come on in. (Don't feel like you're intruding! I'll tell you if it's not a good time.)

---

[2]While regular weekly assignments are evaluated on an effort based, check/no-check system, on the practicum, you'll receive a certain number of points out of the total, like a test.

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 at the latest; if 10 PM Thursday, by 6 PM Friday; if you email me at 3 PM on Friday, by 6 PM Monday, etc.[3]

> **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.

Course Schedule

subject to change

Week	Topic	Tools
	Preliminaries	
1	Course Introduction	Scripts; R and RStudio
2	Reproducible Data Analyses	Markup Languages; Quarto
3	File Management & Version Control	Filesystems; git; GitHub
4	A Field Guide to Data	Data Formats; readr; tidyr
5	Structural Data Manipulation	dplyr; srvyr
6	Data Visualization I	Grammar of Graphics; ggplot2
7	Data Visualization II	ggplot2
8	**Core Exam (Thursday, October 16)**	
9	Workflow & Data Retrieval	database access, APIs, SQL overview
10	Social Networks & Network Data	iGraph; statnet
11	Census Data	tidycensus
12	Maps & GIS	sf; tigris; mapgl; mapbox; osm
13	TBD	TBD
14	Project Work	
15	Project Presentations	
F	**Finals Week (Project Due)**	

---

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

Advanced Topics (if we have time)

Week	Topic	Tools
16	Text Data & Data Scraping	
17	Web Apps & Visualization	quarto, shiny
18	AI Pair Programming & "Vibe Coding"	github copilot
19	Local LLMs	

Campus Academic Resources & Policies

The following reflects the approved campus-wide academic policies of Syracuse University. For more information and detail, please consult the full documentation of University policies here.

Academic Drop Deadline

As part of our efforts to track satisfactory academic progress, the Academic Drop Deadline and the Financial Drop deadline will both occur on September 15, 2025, for the fall semester and February 2, 2026, for the spring semester. Students may still withdraw from courses after these deadlines; this would place a 'WD' grade on their transcripts. Students enrolled in "flex" classes (Flexibly formatted classes) have different deadlines and will need to check MySlice for the Academic and Financial Drop deadlines that pertains to their class.

Academic Integrity

As a pre-eminent and inclusive student-focused research institution, Syracuse University considers academic integrity at the forefront of learning, serving as a core value and guiding pillar of education. Syracuse University's Academic Integrity Policy provides students with the necessary guidelines to complete academic work with integrity throughout their studies. Students are required to uphold both course-specific and university-wide academic integrity expectations such as crediting your sources, doing your own work, communicating honestly, and supporting academic integrity. The full Syracuse University Academic Integrity Policy can be viewed by visiting the Syracuse University Policies website.

Upholding Academic Integrity includes the protection of faculty's intellectual property. Students should not upload, distribute, or share instructors' course materials, including presentations, assignments, exams, or other evaluative materials without permission. Using websites that charge fees or require uploading of course material (e.g., Chegg, Course Hero) to obtain exam solutions or assignments completed by others, which are then presented as your own violates academic integrity expectations in this course and may be classified as a Level 3 violation. All

academic integrity expectations that apply to in-person assignments, quizzes, and exams also apply online.

Students found in violation of the policy are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered. Students may not drop or withdraw from courses in which they face a suspected violation. Any established violation in this course may result in course failure regardless of violation level.

Artificial Intelligence

Based on the specific learning outcomes and assignments in this course, artificial intelligence is permitted on the following: all weekly assignments. Artificial intelligence is *not* permitted in tests and practicums. See each assignment, quiz, or exam instructions for more information about what artificial intelligence tools are permitted and to what extent, as well as citation requirements. If no instructions are provided for a specific assignment, then no use of any artificial intelligence tool is permitted. Any AI use beyond that which is detailed in course assignments is explicitly prohibited except when documented permission is granted.

Please also see Syracuse University's general campus AI policies here.

Academic Integrity Online

All academic integrity expectations that apply to in-person quizzes and exams also apply to online quizzes and exams. In this course, all work submitted for quizzes and exams must be yours alone. Discussing quiz or exam questions with anyone during the quiz or exam period violates academic integrity expectations for this course.

Using websites that charge fees or require uploading of course material (e.g., Chegg, Course Hero) to obtain exam solutions or assignments completed by others and present the work as your own violates academic integrity expectations in this course and may be classified as a Level 3 violation, resulting in suspension or expulsion from Syracuse University.

Attendance Policy

Attendance in classes is expected in all courses at Syracuse University. It is a federal requirement that faculty promptly notify the university of students who do not attend or cease to attend any class. Faculty will use Early-Semester Progress Reports and Mid-Semester Progress Reports in Orange SUccess to alert the Registrar and Financial Aid Office on non-attendance. For more information visit:

Faculty: Information for Faculty: Non-attendance or Stopped Attending Students: Information for Students: Non-attendance or Stopped Attending

If a student is unable to participate in-person or virtually for an extended period of time (48 hours or more), the student may request an absence notification from their home school/college Dean's Office or through Student Outreach and Support office. Instructors will be notified via the "Absence Notification" flag in Orange SUccess.

Barnes Center at the Arch (Health, Counseling, etc.) staff will not provide medical excuse notes for students. When Barnes Center staff determine it is medically necessary to remove a student from classes, they will coordinate with Student Outreach and Support case management staff to provide appropriate notification to faculty through Orange Success. For absences lasting less than 48 hours, students are encouraged to discuss academic arrangements directly with their faculty.

Additional information may be found at Student Outreach and Support: Absence Notifications.

Blackboard

This class will use the Blackboard Learning Management to house the syllabus, course content, links to external course materials, assignments, quizzes, exams, feedback, and grades. Due dates and times in Blackboard are stored in Coordinated Universal Time (UTC) and displayed for each user based on the time zone setting of their computer and data from their internet browser. The system will always display the time zone being used. If an instructor sets a due date of 11pm Eastern time, a student in the Pacific time zone will see a due date of 8pm."

Information about Blackboard is available on Answers Blackboard; alternatively, you can contact Information Technology Services by sending an email to help@syr.edu, calling 315.443.2677, or in-person at the ITS Service Center, located at 1-227 CST in the Life Sciences Complex. Business hours for the Service Center can be found on the ITS website.

Discrimination and Harassment

The University does not discriminate and prohibits harassment or discrimination related to any protected category including creed, ethnicity, citizenship, sexual orientation, national origin, sex, gender, pregnancy, reproductive health decisions, disability, marital status, political or social affiliation, age, race, color, veteran status, military status, religion, sexual orientation, domestic violence status, genetic information, gender identity, gender expression or perceived gender.

Any complaint of discrimination or harassment related to any of these protected bases should be reported to Sheila Johnson-Willis, the University's Chief Equal Opportunity & Title IX Officer for Faculty and Staff. She is responsible for coordinating compliance efforts under the various laws including Titles VI, VII, IX and Section 504 of the Rehabilitation Act. She can be contacted at Equal Opportunity, Inclusion, and Resolution Services, 621 Skytop Road, Suite

1001, Syracuse University, Syracuse, NY 13244-1120; or by email: equalopp@syr.edu; or by telephone: 315-443-4018.

Diversity

It is the intent of this course for students from all diverse backgrounds and perspectives to be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength, and benefit. It is also critical to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let your instructor know ways to improve the effectiveness of the course for you personally or for other students or student groups.

Inclusion

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. There may be aspects of the instruction or design of this course that result in barriers to your inclusion and full participation in this course. I invite any student to contact me to discuss strategies and/or accommodations (academic adjustments) that may be essential to your success and to collaborate with the **Center for Disability Resources (CDR)** in this process.

If you would like to discuss disability accommodations or register with CDR, please visit the Center for Disability Resources. Call (315) 443-4498 or email disabilityresources@syr.edu for more information.

The CDR is responsible for coordinating disability-related academic accommodations and will work with the student to develop an access plan. Since academic accommodations may require early planning and generally are not provided retroactively, please contact CDR as soon as possible.

Faith and Religious Accommodations

Syracuse University's Religious Observances Policy recognizes the diversity of faiths represented in the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their traditions. Under the policy, students are given an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance, provided they notify their instructors no later than the academic drop deadline. For observances occurring before the drop deadline, notification is required at least two academic days in advance. Students may enter their observances in MySlice under Student Services/Enrollment/My Religious Observances/Add a Notification.

Health and Wellness

Well-being and mental health are significant predictors of academic success. It is critical to take care of yourself physically and emotionally and to effectively navigate stress, anxiety, and depression. Please familiarize yourself with the range of resources the Barnes Center provides https://ese.syr.edu/bewell/ and seek out support for mental health concerns as needed. Counseling services are available 24/7, 365 days, at 315-443-8000.