GEORGETOWN UNIVERSITY
McCourt School of Public Policy

Data Visualization for Data Science
PPOL 5202: Fall 2024

Professor

Rebecca A. Johnson

Office Hrs: Wednesdays 10:00-12:00. In person or Zoom. Sign up on Calendly

125 E St NW, Room 768

http://www.rebeccajohnson.io

rj545@georgetown.edu

Course Info

Mondays & Wednesdays

14:00-15:15

Room 490

TA Info

Jiaqin Wu

Office Hrs: Wednesdays 17:00-18:00 + Sundays 10:00-11:00

Calender

Virtual: Zoom

jw2104@georgetown.edu

Overview

This course introduces students to the tools, methods, and skills necessary for making compelling quantitative graphics for the analysis and communication of public policy research. Students will be trained in programming and software applications useful for data visualization including R, Python, and Tableau. Students will have the opportunity to develop and profile their skills in data manipulation and visualization design through an online portfolio. Thanks to Professors Taylor Corbett, Alex Lundry, and Wesley Joe, whose curricula provided inspiration for this course and content structure, and especially to Professor Maddie Pickens whose Fall 2023 iteration of the course heavily shaped the current course structure.

Course Materials

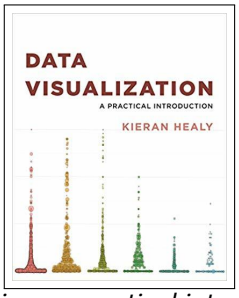
Software

The course will use the following software. We will note setup instructions on Canvas in the weeks leading up to when each software is used. For `R` and `Python`, please make sure to have the software updated to the latest versions compatible with your OS.

1. R + R Studio
2. Python + Jupyter Notebooks
3. Tableau Desktop
    - Free license for students available here: `https://www.tableau.com/university-students`
4. Quarto
5. Git/GitHub

Readings/Textbook

There is no assigned textbook for the course. We will occasionally have readings from this textbook, but all reading material will be posted on Canvas:



Healy, Kieran. *Data visualization: a practical introduction.* Princeton University Press, 2018.

Learning Objectives

- A foundation in the theoretical, practical, and aesthetic elements of data visualization.
- Skills in multiple software tools for making both static and dynamic visualizations.
- An understanding of different data visualization techniques, trends, and challenges they are likely to encounter.

Grading

Item	%	Details
Problem Sets	45	There will be three problem sets that are graded.
Final Project/ Presentation	35	There will be a group final project that is a guided interactive visualization of a dataset your group has found or created, along with a GitHub repo to produce the visualization and technical documentation. There will be a 10-minute final presentation on the project.
Data Viz. Tutorial	10	In different groups, you will create a data visualization tutorial covering a software or concept we have not covered in the class; you will present this during one of the lab sessions
Code Review	5	You will conduct a code review of another group's final project for clarity of code, correctness, and efficiency and write up the results.
Team Player/ Participation	5	This will be an individual grade assessing your contributions to group work, participation in in-class discussions and lab. It is not a measure of attendance but of engagement.
Total	100	

There is no midterm or final examination for the course.

The grade cutoffs are: $\geq$ 95% A; 90-94 A-; 87-89 B+; 84-86 B; 80-83 B-; and so on.

Details on Assignments

As each assignment approaches, we will provide a detailed handout on Canvas outlining milestones, deadlines, and grading criteria. In this syllabus, we provide a preview of each type of assignment:

Problem Sets

There will be three problem sets. These will be completed either individually or in pairs, depending on the problem set. They will be graded for accuracy and efficiency of coding. You will submit the following for each problem set:

- A raw source file for the code (e.g., an .Rmd, .ipynb, or .qmd file; scripts if necessary)
- A compiled HTML version

Final Project / Presentation

This will be completed in groups of 3-4 students, and we will either assign these based on common interests or let you choose your own group. The broad goal of the project, consistent with Fall 2023's course iteration, is to: "create a data visualization project that tells a story with data." The story and underlying dataset—whether using a pre-existing dataset developed by a reputable policy organization or creating a pipeline that generates a new dataset—should be related to public policy. The Final Project will have the following components:

- A viewable interactive data visualization (the final project guide will specify platforms but these include things like a Shiny app, Plotly, or other ways of displaying the data interactively)
- A 4-5 page description of the data sources, key conclusions revealed by the visualization, and caveats/limitations
- A GitHub repo with code + a README that describes all steps to transform the raw data into the visualization
- A 10-minute in-class presentation

Code Review of Other Projects

Groups will be assigned to do a peer code review of another group's final project, with a focus on the GitHub repo. The group will fill out an evaluation form assessing the clarity of the code + thoroughness of the documentation. During the last session of class, we will discuss these evaluations.

Data Visualization Tutorial

In different groups than the final presentation groups, you will create a 10-15 minute data visualization tutorial in the form of a compiled notebook you walk through in class. This tutorial can cover a software we do not cover in the course (e.g., Power BI), a different module or package within `R` or `Python`, or a data visualization concept we do not cover (e.g., visualizing audio or video data). The presentation should be in the form of a tutorial workbook. We will provide examples on Canvas.

Team Player/Participation

This grade will reflect:

- Participation in class: I will not take formal attendance records but hope to have you all participate in Lectures and lab sessions actively.
- Equitable contributions to the group projects: for the data viz. tutorial and final project, we will conduct peer feedback; this will be focused on a yes or no question of whether each group member made a reasonable effort to contribute.

Course-Specific Policies

How to Get Help

We will use two tools that we prefer over e-mail for course-related questions:

- GitHub issues: for each problem set, we will create a GitHub issue for each problem set question; you can post questions on that issue
- Slack: for course logistics + broader, non-pset assignments, we will use Slack for communications

Please direct the following types of questions to each of us:

- Prof. Johnson: lecture concepts; course management; personal challenges/issues (no need to reach out about absences unless it's a significant amount of classtime and you want to discuss makeup work)
- Jiaqin Wu: code/debugging issues; he will escalate these to me as needed

Late Days

In the course, you will have three, no questions asked late days. These can only be used on the problem sets, since the other assignments involve time-sensitive group work. If you're using your late days for an assignment, no need to notify us; we will track late submissions on Canvas.

ChatGPT/AI Use

You are free to use ChatGPT and other forms of AI to help with troubleshooting code. The following guidelines apply (taken from Professor Pickens' Fall 2023 course):

1. Only use it to produce and/or refine initial drafts. The text it generates often lacks a "human" touch and it will certainly not capture your voice. You should make any language it produces your own.
2. Do not rely on it for factual information. It is frequently confidently wrong about things.
3. If you do use it (including for code generation), you must include it in your bibliography or otherwise cite it.
4. When citing usage, you must include each of the specific prompts you used to generate the text or code.

Class Schedule

The class will generally be structured as:

- Mondays: less interactive Lecture session covering that week's topic
- Wednesdays: any residual Lecture + a lab session where you work on a structured coding activity with help from me + your classmates that will preview the problem set content

Below is the current schedule. This will be adapted based on the pace at which we proceed + your feedback on important-to-cover topics. The current schedule will always be available at this Google Sheet.

Last updated: August 28, 2024

Day of Week	Date	Topic	Language
Wednesday	2024-08-28	Introductions	
Tuesday	2024-09-03	Lecture: Tidyverse refresher + ggplot2 for data exploration	R
Wednesday	2024-09-04	Lab: Tidyverse refresher + ggplot2 for data exploration	R
Monday	2024-09-09	Lecture: Automating ggplot2 graphics generation	R
Wednesday	2024-09-11	Lab: Automating ggplot2 graphics generation	R
Monday	2024-09-16	Lecture: ggplot2 for statistics/ML model exploration	R
Wednesday	2024-09-18	Lab: ggplot2 for statistics/ML model exploration	R
Monday	2024-09-23	Lecture: geospatial data in R	R
Wednesday	2024-09-25	Lab: geospatial data in R	R
Monday	2024-09-30	Lecture: interactive visualizations with Shiny; week one	R
Wednesday	2024-10-02	Lab: interactive visualizations with Shiny; week one	R
Monday	2024-10-07	Lecture: interactive visualizations with Shiny; week two	R
Wednesday	2024-10-09	Lab: interactive visualizations with Shiny; week two	R
Monday	2024-10-14	University Holiday: No Class	
Wednesday	2024-10-16	Lab: interactive visualizations with Shiny; week two continued	R
Monday	2024-10-21	Lecture: geospatial data in Python	Python
Wednesday	2024-10-23	Lab: geospatial data in Python	Python
Monday	2024-10-28	Lecture: interactive visualizations with Vega-Altair	Python
Wednesday	2024-10-30	Lab: interactive visualizations with Vega-Altair	Python
Monday	2024-11-04	Lecture: interactive visualizations with Bokeh	Python
Wednesday	2024-11-06	Lab: interactive visualizations with Bokeh	Python
Monday	2024-11-11	Lecture: Storytelling with data (Tableau)	Tableau
Wednesday	2024-11-13	Lab: Storytelling with data (Tableau)	Tableau
Monday	2024-11-18	Lecture: Storytelling with data (TBD software)	
Wednesday	2024-11-20	Lab: Storytelling with data (TBD software)	
Monday	2024-11-25	In-class time for final project work/troubleshooting	
Wednesday	2024-11-27	Fall Recess: No class	
Monday	2024-12-02	Final presentations	Misc
Wednesday	2024-12-04	Final presentations	Misc
Monday	2024-12-09	Peer review of presentations + wrap-up	

Georgetown University Policies

Academic Resource Center/Disability Support

If you believe you have a disability, contact the Academic Resource Center (arc@georgetown.edu) for further information. The Center is located in the Leavey Center, Suite 335 (202- 687-8354). The Academic Resource Center is the campus office responsible for reviewing documentation provided by students with disabilities and for determining reasonable accommodations in accordance with the Americans with Disabilities Act (ASA) and University policies. For more information, go to http://academicsupport.georgetown.edu/disability/.

Important Academic Policies and Academic Integrity

McCourt School students are expected to uphold the academic policies set forth by Georgetown University and the Graduate School of Arts and Sciences. Students should therefore familiarize themselves with all the rules, regulations, and procedures relevant to their pursuit of a Graduate School degree. The policies are located at: https://sites.google.com/a/georgetown.edu/gsas-graduate-bulletin/vi-academic-integrity-policies-procedures.

Provost's Policy Accommodating Students' Religious Observances

Georgetown University promotes respect for all religions. Any student who is unable to attend classes or to participate in any examination, presentation, or assignment on a given day because of the observance of a major religious holiday or related travel shall be excused and provided with the opportunity to make up, without unreasonable burden, any work that has been missed for this reason and shall not in any other way be penalized for the absence or rescheduled work. Students will remain responsible for all assigned work. Students should notify professors in writing at the beginning of the semester of religious observances that conflict with their classes. The Office of the Provost, in consultation with Campus Ministry and the Registrar, will publish, before classes begin for a given term, a list of major religious holidays likely to affect Georgetown students. The Provost and the Main Campus Executive Faculty encourage faculty to accommodate students whose bona fide religious observances in other ways impede normal participation in a course. Students who cannot be accommodated should discuss the matter with an advising dean.

Title IX/Sexual Misconduct

Georgetown University and its faculty are committed to supporting survivors and those impacted by sexual misconduct, which includes sexual assault, sexual harassment, relationship violence, and stalking. Georgetown requires faculty members, unless otherwise designated as confidential, to report all disclosures of sexual misconduct to the University Title IX Coordinator or a Deputy Title IX Coordinator. If you disclose an incident of sexual misconduct to a professor in or outside of the classroom (with the exception of disclosures in papers), that faculty member must report the incident to the Title IX Coordinator, or Deputy Title IX Coordinator. The coordinator will, in turn, reach out to the student to provide support, resources, and the option to meet. [Please note that the student is not required to meet with the Title IX coordinator.]. More information about reporting options and resources can be found on the Sexual Misconduct Website: https://sexualassault.georgetown.edu/resourcecenter. If you would prefer to speak to someone confidentially, Georgetown has a number of fully confidential professional resources that can provide support and assistance. These resources include: Health Education Services for Sexual Assault Response and Prevention: confidential email sarp@georgetown.edu Counseling and Psychiatric Services (CAPS): 202.687.6985 or after hours, call (833) 960-3006 to reach Fonemed, a telehealth service; individuals may ask for the on-call CAPS clinician More information about reporting options and resources can be found on the Sexual Misconduct Website.

———————————————————————