THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL OF PUBLIC POLICY

Data and Programming for Public Policy I

Course number: 30535
Professors: Peter Ganong, Fernanda Sobrino Macias
Teaching Staff: Miguel Valenzuela
Course email: uchicagodatasci@gmail.com

Lecture Mon and Wed 9:00am-10:20am or 10:30am-11:50am
You will be assigned to one of two labs. Each lab will meet for 80 minutes twice per week in the afternoon. Times will be announced soon.
Last updated: 2022-01-21. Latest version here.

Course Description

This course is the first of a three-quarter sequence for the Harris Data Analytics certificate. Although the course is designed for MPPs, undergraduates are welcome to enroll as well.

The goal of this course is to teach you to *quickly engage a policy question with a data visualization.* Doing this requires two new skills.

First, we will teach you to be proficient in R. We will closely follow Hadley Wickham and Garret Grolemund's R for Data Science (R4DS). The online textbook is free.

Second, we will teach you to use data to improve the performance of public sector organizations. The course material draws on Professor Ganong's experience helping to start the Citywide Analytics Team in Boston. The certificate description contains more examples of how teams like this are transforming government. During the course, you will complete eleven problem sets. Through repeated analysis, you will gain knowledge of three public sector datasets:

- Flights data
- Parking ticket data
- Traffic data for Chicago captured at 5-minute intervals from Waze

The last dataset is proprietary. To use this dataset, you will need to agree to abide by the confidentiality rules from the data provider.

This course will differ in two ways from the typical Harris course. Learning R, just like learning a foreign language, is hard and requires lots of repetition.

1. The best way to learn to write good code is to write lots of code. As a result, this course will not have any exams and will have approximately one problem set per week.
2. It is easiest to learn to write code if you set aside time to work on just this and have help available. In addition to lecture, the course will have dedicated lab periods where you will

work on your problem sets. In-person attendance has been mandatory in past years. This year, because of the switch to remote learning, attendance is strongly encouraged.

Prerequisites: Harris Statistics for Data Analysis I & II. If you are a non-Harris student and therefore have not taken these two courses and would like to enroll in this course, you may petition to join by sending an email to the course address with what languages you know and examples of code you have written. These petitions must be submitted before April 1.

Lectures

R4DS lectures (Sobrino Macias & Ganong)
Professor Sobrino Macias will teach on March 28, March 30, and April 4
Professor Ganong will teach the remaining 7.5 weeks (April 6-May 25)

Public sector applications (guests)
We usually have 3-5 guest lectures. Many of the guest lectures are connected to the datasets that we are working with. Here are the guest lectures that we had in 2020 and Professor Ganong is working to line up the guest lectures for 2022:

- Chicago Assessor's Office: Rob Ross, Chief Data Officer, Harris MACRM 2017 & Dan Snow, Data Analyst, Harris MPP 2019
- Melissa Sanchez, ProPublica
- Quoctrong Bui, NYTimes Upshot
- Brenna Berman, CityTech Collaborative, Harris MPP 2000

Code and communication platforms

Code

- Github -- upload psets, collaborate on code
- RStudio and R -- write code

Public communication

- Piazza -- message board for pset questions
- Canvas -- post assignments, calendar, and real-time chat room

The teaching staff will monitor Piazza and Canvas chat 9AM-5PM Monday-Friday.

Grades

The course will have two types of problem sets

- *3 applied problem sets* (40% of grade) -- done in pairs, due once every three weeks. Less structured prompts where you will create data products for decision-makers using the public sector datasets.
- *6 skills problem sets* (60% of grade) -- done solo, due every week. Highly structured exercises from the textbook.

We will drop the problem set grade that most negatively affects your total grade. You will also begin the quarter with nine late coins. A late coin enables you (and your partner, if appropriate) to turn a problem set in one day late. In a partner problem set, only one late coin total is required per day. The maximum number of late coins for one assignment is two.

Helping peers (extra credit, up to 5% of grade) for helpful answers to classmate's questions on our discussion board, Piazza.

Passing You need a grade of 60% to pass this course.

Curve Among students who pass, the curve is one-third A, one-fourth A minus, one-fourth B plus, one-twelfth B and one-twelfth lower grades. The curve is applied to problem sets. Time spent on "helping peers" activities improves your grade and does not affect your classmates' grades.

Integrity

1. Academic dishonesty will not be tolerated. If you commit plagiarism, you may receive an F.
2. All work must be your own. Do **not**
    a. Share your problem set code
    b. Ask for someone else's problem set code
    c. Use online solutions which you might find to the R4DS questions
3. So how can I collaborate?
    a. Via Zoom or other screen share
        i. Clarify ambiguities in p-set questions
        ii. Discuss conceptual aspects of psets (e.g. at the whiteboard)
        iii. Show output and error messages
        iv. List names of all collaborators at top of pset
    b. Via Piazza message board
        i. Ask questions
        ii. Share error messages (but not code)
    c. Code you find online
        i. cite all code you use, even a one-line snippet
4. Can I screen share code? This is allowed but be careful not to inadvertently copy someone else's code, thereby violating rule # 2.
5. How do these rules change for problem sets working in pairs? You and your partner will submit a single problem set.
6. Unsure about some aspect of this policy? Please ask!

Source: This policy draws heavily on the CS12100 academic honesty policy and CMSCC 23300 policy.

THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL OF PUBLIC POLICY

Is this class for me?

This class is designed for people who

1. Have no prior knowledge of programming other than Harris Stats 1 and 2.
2. Want to invest significant time in learning data skills
3. Are working on an independent data project or plan to start one

Should I take this class even if I don't meet the three criteria? Maybe. Our experience is that the students who are happiest with the course are those who want to make a significant time investment in the topic.

Schedule

- Lecture and lab
    - Lectures are in the morning, lab is in the afternoon
    - Guest lectures will occur once at 10:30AM-11:50AM on Mondays and Wednesdays. We will attempt to record them for students who are enrolled in the 9AM section and have a course scheduling conflict.
    - Exceptions -- there will be no lecture on Mon April 18. Make-up lecture will occur on Wed April 20 during the usually scheduled lab time.
- Problem set due dates
    - Every week at 5PM Friday: Skills problem set is due
    - Every other week at 5PM Friday: Applied problem set is due
    - Exceptions -- Final skills and applied psets due Wednesday 5/25

Counseling services

If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, remote counseling services are available. Student Counseling Service (SCS) urges you to attend to your mental wellbeing and to reach out to them for support during these challenging times. All SCS services are covered by the Student Life Fee, and there is no additional cost for students to access their services. Students seeking new services/resources can call 773.702.9800 during business hours (Monday–Friday 8:30 a.m.–5 p.m.) and ask to speak with a clinician**.** Students needing urgent mental health care can speak with clinicians over the phone 24/7 by calling the SCS at 773.702.3625.