Data Skills for Public Policy - 30531

Professor: Peter Ganong
TAs: Ari Anisfeld (Head), Rose Gao, Dan Snow, and Joseph Lovins
Course email: uchicagodatasci@gmail.com
Section 1: Lecture Mon and Wed 8:00-9:20 am in Room 289B,
Required Lab 2:00-3:20pm in Room 224
Section 2: Lecture Mon and Wed 9:30-10:50 am in Room 289B,
Required Lab 3:30-4:50pm in Room 289B

Course Description

This course is the second 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. 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 eight problem sets. Through repeated analysis, you will gain knowledge of up to five public sector datasets:

- Flights data
- Medicare home health claims data
- Parking ticket data
- Human resources data on payroll and absences from a large public agency.
- Traffic data for Chicago captured at 5-minute intervals from Waze

The last two datasets are proprietary. To use these two datasets, you will need to agree to abide by the confidentiality rules from the data providers.

This course will differ in three 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 Monday and Wednesday morning lecture, the course will have a mandatory lab to meet where you will work on your problem sets on Monday and Wednesday afternoon.

3. Different students have different styles for learning how to code. My lectures will largely be a substitute for the textbook and you should learn in whatever way is most effective for you. If you attend lecture, please arrive on time. Arriving late disrupts your peers' learning. In addition, attendance at guest lectures is mandatory. Lecture is interactive. Bring your name card every day so Professor Ganong can learn your name and call on you.

Prerequisite: As a part of the Harris Data Science certificate, you must have taken 30550 "Introduction to Programming for Public Policy". The course is also open to students with significant prior programming experience. If you have not taken 30550 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 the first lecture.

Topics & Dates

We will have approximately 15 lectures by Professor Ganong and 6 guest lectures. We have five confirmed guest lectures:

- Jascha Franklin-Hodge, Former Chief Information Officer of the City of Boston
- Adam Freeman, U.S. Department of Health and Human Services, Harris MPP 2010
- Melissa Sanchez, ProPublica
- Rob Rose, CEO, Cook County Land Bank
- Rebekah Scheinfeld, Commissioner of Transportation, City of Chicago

There will probably be one more guest lecture, but that person is not confirmed yet.

10/1	**Lecture canceled**, Lab optional
10/3	Lecture + Lab. **Bring your name card to lecture.**
10/4	PS1 due
10/8	Lecture + Lab
10/10	Lecture + Lab
10/14	PS2 due
10/15	Lecture + Lab
10/17	Lecture + Lab
10/21	PS3 due
10/22	Lecture (Jascha Franklin-Hodge) + make-up lecture

10/24	Lecture + Lab
10/28	PS4 due
10/29	Lecture (Adam Freeman) + Lab
10/31	Lecture + Lab
11/4	PS5 due
11/5	Lecture + Lab
11/7	Lecture (Melissa Sanchez) + Lab
11/9	PS6 due
11/12	Lecture + Lab
11/14	Lecture (Rob Rose) + Lab
11/19	Lecture + Lab
11/20	PS7 due
11/21	AM Lecture is on. Lab canceled for Thanksgiving.
11/26	Lecture + Lab
11/28	Lecture + Lab
12/3	Lecture (Rebekah Scheinfeld) + Lab
12/5	Lecture (content not on PS8) + Lab. PS8 due.

Grades

Problem sets (90% of grade) will be submitted using github. Register here. There are 8 problem sets. I will drop your lowest problem set grade. Most will have two components:

- highly structured exercises from the textbook as well as
- less structured prompts where you will create data products for decision-makers using the public sector datasets

Quizzes (10% of grade) at lab to work on problem sets. Quizzes occur at the start of lab. You must be physically present to take the quiz. I will drop the two lowest quiz grades.

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

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

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. In-person collaboration at lab
        i. clarify ambiguities in p-set questions
        ii. discuss conceptual aspects of psets (e.g. at the whiteboard)
        iii. show output on screen (e.g. a graph or table)
    b. Electronic collaboration outside lab
        i. Piazza message board
            1. ask questions
            2. share error messages (but not code)
        ii. Code online
            1. cite all code you use, even a one-line snippet
4. How do these rules change for problem sets working in pairs?
    a. You and your partner will submit a single problem set.
    b. You and your partner can talk as much as you want.
5. Unsure about some aspect of this policy? Please ask!

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

Is this class for me?

This class is designed for people who

1. Have no prior knowledge of programming other than 30550 (introduction to programming for public policy
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. My experience from last year was that the students who were happiest with the course were those who wanted to make a significant time investment in the topic.