

Data and Programming for Public Policy I

Course number: 30535

Professor: Peter Ganong

Course email: uchicagodatasci@gmail.com

Section 1: Lecture Mon and Wed 8:00-9:20 am, Required Lab Mon and Wed 3:30-4:50pm

Section 2: Lecture Mon and Wed 9:30-10:50 am, Required Lab Tue and Thu 3:30-4:50pm

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 Golemund'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
- Home value assessing 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 in the 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.

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 the first lecture.

Guest lectures

Please bring your namecards and a notebook. Please do not bring laptops or tablets.

We will have 4-6 guest lectures this year. They will all occur 9:30am-10:50am. If you are enrolled in the 8:00am-9:20am section, you are still required to attend the guest lectures. If you are enrolled in a course which conflicts, we will have a procedure by which you can document the conflict and receive an exemption from this requirement.

Confirmed speakers this year

- Melissa Sanchez, ProPublica
- Rob Ross, Chief Data Officer, Chicago Assessor's Office, Harris MPP 2017

To give you a sense of the type of people who might come, last year's speakers were

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

Course-related requests from professor

Sometimes, students will request an exception on some sort of work (e.g. to be exempted from a quiz). Such requests are accepted only via email to uchicagodatasci@gmail.com and not via an in-person meeting.

Grades

Problem sets (90% of grade) will be submitted using github. Register [here](#). There are 8 problem sets. We 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. We 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 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 and quizzes. Time spent answering questions on Piazza 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. 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 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. My experience is that the students who are happiest with the course are those who want to make a significant time investment in the topic.