Data Skills for Public Policy - 30531

Professor: Peter Ganong
TAs: Yuxi Wu and Emma Peterson
Course email: uchicagodatasci@gmail.com
Lecture: Monday and Wednesday 9:00 AM - 10:20 AM, Harris Room 289B
Lab (Required): Monday and Wednesday, 1:30 PM - 2:50 PM, Harris Room 140C

Course Description

This course is the second of a three-quarter sequence for the Harris Data Science certificate. This certificate is designed to train you to work in the rapidly-expanding field of data analytics in the public sector after graduation. 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 Peter'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 nine problem sets. Through repeated analysis, you will gain an in-depth knowledge of three public sector datasets:
- Medicare data
- traffic data for Chicago captured at 5-minute intervals from Waze
- shift-level human resources data from a large transit agency.

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. Different students have different styles for learning how to code. As a result, we will use a flipped classroom. 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.
2. 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 one problem set per week.
3. You are encouraged to discuss challenges with your classmates, but what you submit must be your own work. There is a strict academic honesty policy described below.

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 cover *R for Data Science* in approximately the first eight weeks of the course. In the last two weeks, we will briefly cover mapping and dashboards.

Date	What's special
1/4	Extra optional lab in 289A, 1:30-2:50
1/10	Guest: Jeff Liebman, Government Performance Lab, Founder and Director
1/15	No class -- MLK Day
1/16	Lab in 289A, 1:30-2:50
1/24	Guest: Adam Freeman, U.S. Department of Health and Human Services, Harris MPP 2010

Grades

Problem sets (89% of grade) will be submitted using github. Register here. There are 10 problem sets and they are due each Thursday by 11:59PM. Most problem sets 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

I will assign you to groups, which will rotate every week. Late problem sets will not be accepted. However, I will drop your lowest problem set grade.

Quizzes (11% 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. I will also drop your quiz grade on 1/16 if you have a scheduling conflict.

Academic Honesty

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. What resources can I use?
   a. with group members
      i. discuss conceptual aspects of psets (e.g. at the white board)
   b. with anyone in the class
      i. ask questions on Piazza
      ii. share error messages (but not code) on Piazza
   c. with code online
      i. cite all code you use, even a one-line snippet
4. Unsure about some aspect of this policy? Please ask!

Source: This policy draws heavily on the CS12100 academic honesty [policy](#) and CMSCC 23300 [policy](#).