THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL OF PUBLIC POLICY

Data and Programming for Public Policy I

Course number: 30535
Professor: Peter Ganong
Head TA: Sushant Banjara

Lecture: Mon and Wed 9:00am-10:20am or 10:30am-11:50am in Keller 0001[1]

Labs: Mon and Wed 1:30pm-2:50pm or 3:00pm-4:20pm in Keller 1002

Last updated: 2023-03-19. Latest version here.

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).

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 four public sector datasets:

- Flights data
- Medicare billing 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 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 is mandatory.

[1] There are five guest lectures which will always occur 9:00AM-10:20AM. Details below.

THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL OF PUBLIC POLICY

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 14 lectures by Professor Ganong.

Guest lectures

- 3/29 9AM -- Matt Yglesias
- 4/17 9AM -- Vedika Ahuja, City of San Rafael & Harris MSCAPP 2020
- 4/24 9AM -- Adam Freeman, Department of Health and Human Services, Harris MPP 2010
- 5/1 10:30AM -- Melissa Sanchez, ProPublica
- 5/17 9AM -- Dan Snow, Cook County & Harris MPP 2019

Is this class for me and how can I get help?

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

Getting help

- Come to lab
- Post in Ed
    - The teaching staff will respond to questions 9AM-5PM Monday-Friday and 9AM-12PM Saturday.
- Send us a private message in Ed
    - We discourage the use of email since it's easy for messages to get lost in our inboxes. Most emails will generate a response "please send a message in Ed". However, if you have a sensitive matter

which cannot be shared with the teaching staff, then please email ganong@uchicago.edu w "30535 confidential".

- Tutoring (more on this below)

Grades

The course will have a weekly problem set due Saturday at 5PM. (80% of grade)

- *4 applied problem sets* -- less structured prompts where you will create data products for decision-makers using the public sector datasets.
- *5 skills problem sets* -- highly structured exercises from the textbook. due every week there is not an applied problem set.

We will drop your lowest problem set grade.

Class and Lab Participation (10% of grade). Participation grades will be based on attendance, bringing name tent for first 3 weeks, and asking questions in class/Ed/MUD cards.

Quizzes (10% of grade) occur at the start of each lab. We will drop the two lowest quiz grades.

Helping peers (extra credit, up to 5% of grade) for helpful answers to classmate's questions on our discussion board, Ed, or being acknowledged by many classmates in their problem set submissions for helping them.

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

Grade cutoffs A: 100 – 95%, A-: 90% – 94.9%, B+: 85% – 89.9%, B: 80% – 84.9%, B-: 60% - 79.9%

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 or on zoom
        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 for method (a) at top of pset
    b. Via Ed message board

        i. Ask questions
        ii. Share error messages or reprex code
    c. Code you find online
        i. Cite all code you use with a URL, even a one-line snippet.
        ii. For ChatGPT provide the query string you used.
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 with both names at the top.
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.

Course policies

- *Late problem sets* You have four late coins. A late coin enables you (and your partner, if appropriate) to turn a problem set in one day late. The maximum number of late coins per assignment is one. After 24 hours if your problem set is not turned in you will receive a zero for the assignment.
- *Regrades* You may request a regrade of a problem set by emailing Sushant Banjara (sbanjara@uchicago.edu) within a week of the problem set being returned. If this occurs, we will regrade the entire problem set and your grade may go up or down.
- *Absence from class*
    - Class will be recorded via Zoom. Send a private message in Ed if you are out sick and we will give you access to the recording. If you send a private message in advance requesting a live Zoom stream, we can try to set this up.
    - Every parent has had the experience of child care falling through at the last minute. Children are welcome in a pinch although please bring some quiet toys for them to play with.
- *Absence from lab* -- Lab will not be recorded. If for some reason you are unable to attend lab, please post any questions you have to Ed.

Schedule

- Guest lectures
    - Although there are two sections of the course, there will be only a single guest lecture. It will usually be at 9:00AM-10:20AM.
    - If you are enrolled in a course that conflicts with the time of the guest lecture, we will provide you with a recording of the guest lecture which you can watch prior to the quiz.

- Idiosyncrasies relative to the regular schedule
    - Week of April 3:
        - Monday April 3:
            - Regular lecture in AM slot in 0001
            - Additional lecture in PM slot in 1002
        - Wednesday April 5 AM
            - No lecture in AM, optional lab 10:30-12PM in 0001
            - Regular lab in PM slot in 1002
    - Week of April 10:
        - Monday April 5
            - Regular lecture in AM slot in 0001
            - Additional lecture in PM slot in 1002
        - Wednesday April 12
            - No lecture in AM, optional lab 10:30-12PM in 0001
            - Regular lab in PM slot in 1002

Textbook details

Lectures

1. Ggplot (Ch2)
2. Dplyr (Ch4)
3. Tidyr (Ch6)
4. Style (Ch5) & Import (Ch8)
5. EDA (Ch11)
6. Layers (Ch10)
7. Logical vectors (Ch13) & numbers (Ch14)
8. Factors (Ch17) & dates (Ch18)
9. Missing values (Ch19)
10. Joins (Ch20)
11. Functions (Ch26)
12. Iteration (Ch27)

Support

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.

Tutoring

Harris offers 10 hours of free tutoring support to Harris students for coding in Stata, R, and Python. Tutoring will be available starting March 30th for this Spring Quarter. You can read more about the program on the Harris Student Handbook Canvas site. Any questions should be directed to your academic advisor or harrisdeanofstudents@uchicago.edu.