Data and Programming for Public Policy I

Course Syllabus

Course number: 30525
Professor: Nate Barker
Course email: harris.dppp1@gmail.com

Lectures:
Section 1: Monday and Wednesday, 3:00pm-4:20pm
Section 2: Monday and Wednesday, 4:30pm-5:50pm

Labs (to be assigned once class starts):
Lab Section A: Monday and Wednesday, 6:00pm-7:20pm
Lab Section B: Tuesday and Thursday, 8:00am-9:20am

Note: On the first week of class, we will meet on Friday, September 30th, at the same times (Section 1 at 3:00pm, Section 2 at 4:30pm)

Description

This course offers an introduction to programming and data analysis for students with no prior coding experience.[1] This course is targeted towards MPP students; undergraduates are welcome to enroll as well.

The course will be taught in R. We will learn how to write and share code, and how to clean, wrangle, and visualize data. This course is the first of a three-quarter sequence in the Harris Data Analytics certificate. There are two main objectives associated with the class:

- For those of you who go on to work in policy or policy-adjacent jobs, we want you to feel comfortable using programming (more generally) and R (specifically) as a way of answering policy questions.
- For those of you who discover after taking this course that you enjoy programming (or at the very least, find it valuable and tolerable), we want to provide a strong coding foundation for subsequent courses in programming and analysis at Harris.

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.

[1] Or experience with a proprietary platform like Stata, SASS, SPSS, etc.

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 labs will offer a chance for you to work on your problem sets, with teaching assistants available.

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 by September 30th.

Materials:

Textbook: We will closely follow Hadley Wickham and Garret Grolemund's R for Data Science (R4DS). The online textbook is free.[2]

Software and platforms: we will use R, the graphical user interface (gui) RStudio, and the version control platform Git (hosted on the website GitHub). Each of these services is free (which has the added benefit that you can freely download any of these products and use them in your future jobs to take advantage of the skills learned in this course).

Grading:

Problem Sets: Grades for the course will be determined by 9 problem sets, roughly one per week. They will consist of two types:

1) 5 Skills Problem Sets, worth 60% of grade. They will be done individually, and will be similar to the exercises in the books.
2) 4 Applied Problem Sets, worth 40% of grade. They will be done in pairs, will be less structured, and will practice using, cleaning and visualizing policy-related datasets.

We will drop the problem set that most negatively affects your grade.

Late coins: You will 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.

Passing: grade of 60% or more

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.

Integrity:

[2] You are welcome to buy the book if you would like. I find the online version substantially easier to use though, with e.g. sample code you can paste directly into your RStudio.

All submitted code should be your own

What you CAN'T do:

- Share your code
- Ask for someone else's code
- Copy and paste online solutions for book exercises
- Copy and paste (or read and rewrite) code from online sources (stackoverflow, github, medium, etc)

What you CAN do:

- Via Zoom/Screen sharing or in person:
  - Clarify questions
  - Discuss conceptual aspects (pseudo code)
  - Show output and error messages
  - Always list all collaborators on top of your pset
- Canvas:
  - Ask questions
  - Share error messages
  - Share generic or pseudo code
- Online resources: always cite the source by leaving the link commented on your code

Course Outline:

The topics we will cover in this class will (provisionally) include

- Intro to R and GitHub
- Data Visualization
- Data Transformation
- Exploratory Data Analysis
- Wrangling Data
- Tidying Data
- Joins
- Factor Data
- Working with Dates and Times
- Pipes and Functions
- Iteration
- Graphs for Communication
- Intro to Maps

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 8:30am-5pm) 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.