



Data and Programming for Public Policy I

PPHA 30535

Instructors _____



Fernanda Sobrino



datasciharris@gmail.com

Logistics _____

Lectures T and Th 2-3:20pm and 3:30-4:50pm

Labs: TBA

Office hours: TBA

Resources _____

Coding: R and R Studio

Upload Psets: Github

Public Communication: Ed Discussion

Private Communication: Email

Description

This course introduces programming and data analysis for students with no prior coding experience ¹. By the end of this quarter, you should be able to write and share code, retrieve, clean, wrangle, and visualize structured data sets to help you start answering some explanatory policy questions.

****Learning to code is like learning a new language, so you will need to code a lot and not be afraid to ask questions. The class will have several problem sets and different lab times to exercise your new coding skills.**

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 TBH

Materials:

- Hadley Wickham and Garret Grolemund's R for Data Science (R4DS). The online textbook is free [here](#).

Grading

Problem Sets: There will be 9 problem sets ~ 1 per week. Two different kinds of problem sets:

- 6 skills (60%): solo authored book/similar to the book exercises
- 3 applied (40%): done in pairs, less structured using different data sets

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

9 Late Coins: a late coin allows you (and your partner) to turn psets in one day late. In a partner problem set, you just need one per day (have fun bargaining). You can use a maximum of 2 per assignment.

Passing: grade of 60% or more.

Curve: for passing grades, 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 ALL CODE SUBMITTED 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 an 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

¹Or experience in proprietary platforms like Stata, SASS, SPSS, etc

Course outline

Tentative outline it might change later

Week 1

- Introduction to R and Github
- Visualization I

Week 2

- Visualization II
- Data Transformation I
- Problem set skills I

Week 3

- Data Transformation II
- Exploratory Analysis
- Problem set skills II

Week 4

- Tidy I
- Tidy II
- Problem set applied I

Week 5

- Data Types I
- Data Types II
- Problem set Skills III

Week 6

- Joins
- Pipes
- Problem set Skills IV

Week 7

- Functions
- Vectors
- Problem set Applied II

Week 8

- Logic controls and Loops
- Graphics for Communication I
- Problem set skills V

Week 9

- Graphics for Communication II
 - Maps
 - Problem set skills VI and Applied III
-