SYLLABUS

Data and the State: How Governments "See" People and Places

PUBPOL 2130 / INFO 3130
Moon Duchin, Spring 2025

1   Course Basics

Instructor & Lecture.   This course is taught by Moon Duchin (mduchin@cornell.edu), Professor of Mathematics and Public Policy. Lectures are Tuesday/Thursday 1:25-2:40. Moon's office is in Malott 505, office hours will be posted at a calendar link.

Teaching Assistants & Sections.   Jennah Gosciak (jrg377) and Tanya Popli (tp399) will run required discussion/lab sections on Friday mornings. Two sections run 9:05-9:55 and two run 10:10-11:00.

Description.   In order to regulate and govern, states must marshall data: the whos, whats, and wheres of a nation get rendered in records and statistics, at least approximately. One of the most fundamental instruments for the datafication of the population is a census – and the very first Article of the U.S. Constitution calls for one to be conducted every ten years. That same sentence is also the source of the notorious "Three-Fifths Compromise" establishing an enumeration rule where enslaved people count fractionally as much as free people: arithmetic in the service of politics. In this course we will use Census data as a jumping-off point for an investigation of data practices of governance, with a focus on race, geography, and a host of frameworks and mechanisms to sort, measure, and locate. Students will learn to use a suite of modern packages for geospatial data science in Python, including Pandas, GeoPandas, Shapely, and Matplotlib, and readings will include texts from geography, anthropology, science and technology studies, and urban sociology. No prior knowledge assumed.

2   Assignments and Assessment

Each week we will be working through a Python notebook together (posted to the course GitHub) and you will be expected to do ≤ 30 pages of reading.

On a weekly basis, students will submit two items. First is a small data product (such as a plot or table) created during the Friday lab section and submitted at the end. This doubles as taking attendance for labs. Second is a weekly assignment in which students will be expected to briefly address the reading, and to creatively modify the code in the notebook, or write independent code, to produce a novel data product.

There are two short in-class exams ($\approx$ Feb 14, Mar 21) and one final project. Projects will be short reports and visual data products created in teams of 4-5. Students can find a novel data source and curate it for use, or can use the datasets introduced in the class in a novel way. Projects must articulate an argument about the topic of choice, supported by data. Project teams and topics must be approved by April 8; first draft due May 1; final submission date (roughly May 12) is pending, to be assigned by the university.

Academic integrity. Attribution statements are expected on homework assignments, documenting human or non-human assistance. We ask that you do not rely on ChatGPT or similar tools on a regular basis, and that you note any use. Assignments can be worked on in groups, but precise contents must be individual to the submitter.

Grading. For each prompt, work is typically graded on a four-point scale, where 3 out of 4 is good/solid work, and the score of 4/4 is reserved for exceptionally strong or creative work. At the end of the term, a score of 75% will earn an A-. Class participation is largely measured in lab sections, though participation in lecture is highly welcome. The grade breakdown is 10% participation, 30% weekly assignments, 30% exams, and 30% final project.

Lab attendance is required. In the case of an absence, students are expected to fill out this Absence Form, typically by the following day, or as soon as possible in extraordinary circumstances.

The university maintains the following grading scale:

`A+=4.3 A=4.0 A-=3.7 B+=3.3 B=3.0 B-=2.7 C+=2.3 C=2.0 C-=1.7 D+=1.3 D=1.0 D-=0.7 F=0.0`

3 Learning Outcomes

The goal of this course is to build up tools from data science and computing and tackle questions of social structure and culture, enabling robust analysis and design of public policy. Skills to be developed include coding, statistical analysis, visualization/narrative-building, and critique. The core learning objectives are:

1. Write code to curate, clean, analyze, and visualize spatial data;
2. Gain an understanding of historical processes of classification and regulation;
3. Implement a descriptive or predictive analysis using appropriate data and statistical and/or computational methods;
4. Clearly communicate your process and results as a data narrative through visualizations, context, textual description, and oral presentation;
5. Identify the limitations and potential biases in the data, data-generating processes, and tools and methods in addressing a research topic

4 Accommodations and Course Norms

I am committed to the accessibility of my courses for students from all backgrounds. For years I have worked to build courses in keeping with what is now called UDL, or Universal Design for Learning. This means that I will make the course meetings, assignments, and assessments as flexible and inclusive as possible. Anyone can access extra time on exams, and alternative formats can be proposed for assignments within the confines of administrability for a large course. Official accommodations can be registered through the office of Student Disability Services. Agreement about appropriate accommodations must be reached at least two weeks in advance of assignments or assessments for which an alternative structure is requested.

Laptop policy. If you want to use a laptop in class, I ask that you sit in the front row. For everyone else, please consider taking notes on paper! It is a different learning experience and you might like it. The reason for this policy is just to encourage more engagement and eye contact during class, though I understand if a laptop or other tech is best suited to your learning needs.

Inclusivity. Cornell seeks to create an inclusive learning environment where diversity and individual differences are acknowledged and respected. I expect that students in this class will honor our mutual differences in perspective, background, and lived experience. Be kind and treat each other well—I will do the same!

5   Topics and Schedule

This schedule is subject to change, especially because I am working to secure guest speakers. Here is the currently planned pacing. As speaker scheduling is finalized, I will post updated pacing plans in Canvas.

Week 1 Course Intro and Census I – why we count, who counts
Week 2 Census II – kinds of people
Week 3 Geography I – units and hierarchies
Week 4 Geography II – the Modifiable Areal Unit Problem (MAUP)
Week 5 War, Borders, Migration
Week 6 Health and Disease
Week 7 Sex, Family, Wealth
Week 8 Policing and Incarceration
Week 9 Cities and Transit
Week 10 Housing and Segregation
Week 11 Neighborhoods and Communities
Week 12 Elections
Week 13 Redistricting
Week 14 Privacy
Week 15 Wrap-up