

UP206A: Introduction to GIS and Spatial Data Science

Fall 2020

Mondays 2pm - 5pm

Instructor:

Yoh Kawano (yohman@gmail.com)

Teaching Assistants:

Bo Liu (blu17@g.ucla.edu)

Chris Giamarino (cgiamarino@g.ucla.edu)

Office Hours:

Please email in advance for appointments.

Yoh: Mondays 5PM - 6PM and by appointment

Bo and Chris: TBD and by appointment

Course Description

Welcome to the world of planning and spatial thinking. In recent years, our relationship with maps and map-making has changed dramatically. No longer are we limited to a mode of map-making that is dominated by industry giants like Google and ESRI. Instead, a suite of data-science tools have matured to a point where they can produce similar, if not, more powerful and creative ways that advance spatial exploration.

Our understanding of social phenomena through spatial constructs in urban data allows us to address questions on social justice, the environment, transportation, community development and design, amongst many other themes, and how these factors may affect different population groups in different ways. This course prepares you for challenges in urban data beyond off-the-shelf cartographic approaches. It looks at the various components of data's journey—acquisition, exploration, modeling, and communication through visualization—and how it enables and advances your research from a data science perspective.

The goal for this course is to expose you to the foundations of spatial data science. Where once there was a dearth of available digital information, we now live in a world of too much data. How can these data be transformed to human expressions and narratives that are utilized in planning? We begin with an introduction to various data science tools, and review the basics of programming with Python. Once a foundation of Python programming and data wrangling is achieved, spatial analysis through Python Libraries, and subsequently, through advanced geoprocessing will be introduced. All lessons will be based on "real" data with analytical methods addressing relevant and contemporary urban problems. At the conclusion of this course, students will be able to critically describe, analyze, and visualize spatial data for planning practices and research.

While there are no prerequisites for taking this course, people who are approaching programming for the first time will admittedly find the course to be intense and challenging.

Learning Objectives for UP206A

At the conclusion of this course, students will be able to critically describe, analyze, and visualize spatial data for planning practices and research.

Specifically, students will learn to:

- Demonstrate the value of data for planning purposes through discovery, exploration, and analysis
- Critically evaluate data and create frameworks to prepare data for researchApply data science programming techniques to produce relevant visualizations that inform urban policy
- Learn to visualize data spatially to communicate the importance of place-based informatics
- Learn to produce publication-ready scholarly materials in the form of tables, charts, and maps

Inspirations

I would be remiss if I do not mention various inspirations that have led to the creation of this course. First and foremost, the late and great [Leo Estrada](#), my mentor and confidant over the years, who taught GIS at our department for many decades. Leo and I spoke extensively about modifying this course to a more “modern” approach, and I am delighted and honored to uphold his legacy moving forward.

Second, I have taken the liberty (with permission) to borrow ideas and materials from other courses. I specifically took inspiration from [Paul Waddell’s “Urban Informatics and Visualization”](#) course at UC Berkeley, and [Geoff Boeing’s “Data, Evidence, and Communication for the Public Good”](#) at the Department of Urban Planning and Spatial Analysis at USC’s Sol Price School of Public Policy.

Course materials

The course will almost entirely be conducted on Jupyter Notebooks. A [JupyterHub](#), a web-based Jupyter Notebook environment, has been set up specifically for this class, and is available at the following [URL](#). Note that you will need multi-factored authentication to login:

- [JupyterHub](#) (choose UCLA)

Weekly course materials, including presentations, interactive notebooks (.ipynb), and data will be made available through a course github repository (TBD) that you will interact with through your JupyterHub account.

Assignments and Evaluation

All assignments, unless otherwise specified, must be posted on your individual GitHub accounts by midnight of the Sunday prior to our class on Monday.

TBD

- Participation, individual progress 10%
- Group assignments 40%
- Mid-term 25%
- Finals 25%

Remote zoom learning expectations

Welcome to the new normal. It is expected that the 2020 Fall quarter will be conducted remotely and via zoom in its entirety. The good news is that this course—coding, data, maps!—is situated perfectly for remote learning. Nevertheless, there are some expectations and guidelines I would like to enforce in order to make this experience as fruitful and productive for all of us.

- Lectures will be recorded and made available one day after instruction
- While lectures are recorded, you are expected to be in attendance during live sessions
- Unless prior reasoning is provided by email to the instructor, students are expected to turn on their video’s during zoom lectures and discussions

- Students and instructors must mute their mics unless they are speaking
- Virtual backgrounds are allowed, but keep them as distraction-free as possible
- Chatrooms will be open and monitored by the T.A.'s

How to ask a technical question

Given the nature of the course, you will have many, many questions along the way. However, we ask that you adhere to the following guidelines in order to make consultations as productive as possible. Students who do not follow these guidelines will be asked to reschedule.

Before asking a question:

1. Close all open programs, restart your computer, then try your task again
2. Search google and stackoverflow for the topic/problem (for example, the name of the function you're struggling with or the error message you are seeing)
3. Go back through the relevant lecture materials to look for any insights
4. Go back through the assigned reading materials to look for any insights

If the above steps haven't solved your problem, send an email (or attend office hours) and include the following information:

1. A detailed description of what you're trying to do, why, and how
2. A complete [minimal reproducible example](#) of your code so far (never send screenshots of code)
3. What you've already tried to do to solve your problem and what you have learned from it (specifically, explain the results of steps 1-4 above, including relevant links from stackoverflow etc)

Readings and Resources

- Think Python 2nd Edition by Allen B. Downey
 - <https://greenteapress.com/wp/think-python-2e/>
- Jupyter Notebooks
 - <https://jupyter.readthedocs.io/en/latest/index.html>

Weekly Schedule

Weekly content is subject to change, and will be modified as needed based on class discussions, needs, and progress.

Preparation

- If you do not have a GitHub account, create one for the class
 - <https://github.com/>
- Fill out the [pre-class survey](#)

Week 1 (10/5): Introduction to spatial data science

- Lecture
 - Overview of the course
 - Expectations
 - Where did ArcMap go?
 - What is Data Science?
 - How is it relevant to planning?
 - Mid-term and final projects will be based on groups (pairings)
- Hands on
 - Introducing GitHub

- Introducing JupyterHub
- Intro to Python
- Assignments
 - Create a GitHub account, and create a project repo named "UP206A"
 - Create a Readme.md file in your GitHub class repo, and introduce your Data Science project space (hint: use this [markdown guide](#))
 - Search for and document up to three potential data sources for your own research inquiry; explain how it can be used to answer a potential research question, and how it can be visualized spatially

Week 2 (10/12): Data in Urban Studies: The challenge in data acquisition

- Lecture
 - The digital globe: where to get data, how to use it, what can it do for you?
- Data resources
 - Census data: <https://data.census.gov/>
 - Census reporter: <https://censusreporter.org/>
 - Social Explorer (requires UCLA VPN for full access): <https://www.socialexplorer.com/explore-maps>
 - LA city data portal: <https://data.lacity.org/>
 - LA county data portal: <https://data.lacounty.gov/>
 - LA Metro: <https://developer.metro.net/>
 - LA Times: <http://boundaries.latimes.com/sets/>
 - Census TIGER/Line: <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>
 - NHGIS: <https://www.nhgis.org/>
 - NHTS: <https://nhts.ornl.gov/>
 - LATCH: <https://www.bts.gov/statistical-products/surveys/local-area-transportation-characteristics-households-latch-survey>
- Hands on
 - Python Boot Camp
 - Intro to the Pandas Library
- Assignments
 - Group Assignment #1
 - Write one paragraph for your midterm and final project proposal. Include:
 - A paragraph that explains an issue of interest
 - A research question
 - Data sources
 - Create a final project repo in one of your group member's github account, then assign other group members as collaborators
 - Individual assignment:
 - Launch JupyterHub, import a dataset of your choice, and conduct data exploration, making sure to document your steps and your preliminary findings
 - Add your first data exploration Jupyter Notebook to your GitHub repository

Week 3 (10/19): Understanding communities: Census data profiles

- Lecture
 - Census Data
 - Python for urban data analysis
- Hands on
 - Census data exploration in python using pandas
 - Plotting with matplotlib
 - Quick interactive mapping with Folium
- Assignments

- Group Assignment #2
 - Download census variables relevant to your research question
 - Create a notebook and use pandas to explore and visualize the data
 - Produce several charts, including a map, make sure to document each process

Week 4 (10/26): Open data

- Lecture
 - Open data, API's and why that matters
 - Learning SQL and why it is relevant to understanding data filtering
 - 4th Annual Battle of the Maps: A Humanitarian Mapathon
- Case Study: LA Open Data Portal
 - [Example tutorial](#)
 - LA Times Data Desk
- Hands on
 - SQL in Python
 - Sodapy library to access socrata data
- Assignments
 - Group Assignment #3
 - Create a Jupyter Notebook that takes in, queries, and explores a dataset acquired using a Socrata API

Week 5 (11/2): Data visualization: Open Street Maps

- Mapping: OSM
- Overview of mid-term requirements next week
- Hands on
 - What is OSM?
 - Open Street Map with osmnx
- Assignments
 - Choose a Python mapping library of your choice (folium, geopandas, Plotly Express) and provide three map-based visualizations
 - Mid-term prep!

Week 6 (11/9): Mid-terms

- Midterm lightning presentations
- Hands on
 - Introducing geopandas
 - How to load different data types
 - Managing projections

Week 7 (11/16): Geo-processing with geopandas

- Introducing spatial analysis with geopandas
- Hands on
 - Joining data
 - Spatial analysis with geopandas
- Assignments
 - Group Assignment #4
 - Create a new jupyter notebook, and use geopandas to load your past project data or newly acquired data
 - Join two datasets, visualize the results
 - Create 3 or more maps

Week 8(11/23): Advanced methods in spatial analysis

- Content TBD and subject to student demand. Proposed topics:
 - Network analysis
- Hands on
 - TBD

Week 9: Spatial Data Science: R to the "r"escue

- Alternatives to Spatial Data Science: Using R
- Special topics: Spatial narratives through social media
- Hands on
 - R Studio goes spatial

Week 10: Final Project Preparation

- Publishing your work
- Final project guidelines review

Resources:

- [gboeing/ppd534: USC PPD534: Data, Evidence, and Communication for the Public Good](#)
 - University of Southern California
 - Professor: Geoff Boeing
- [CP255: Urban Informatics and Visualization](#)
 - University of California, Berkeley Department of City and Regional Planning
 - Professor: Paul Waddell
- [Automating GIS-processes 2019](#)
 - University of Helsinki, Finland
- [Computing for the Social Sciences](#) (R focused)
 - University of Chicago