CP255 Urban Informatics and Visualization

Scheduling: Mon/Wed, 9:30am-11:00am
Location: 214 Wurster Hall computer lab

Instructors:

Paul Waddell, waddell@berkeley.edu Office Hours: M 11-12 in 326A

Geoff Boeing, gboeing@berkeley.edu Office Hours: M/W 11-12 in 214 lab

This is a hands-on course that trains students to analyze urban data, develop indicators, and create visualizations and maps using the Python programming language, open source tools, and public data. The course will first introduce the fundamentals of programming in Python before moving on to a survey of data analysis/visualization tools and technologies. Classroom sessions will include lectures and workshops. A series of exercises will reinforce the skills and topics being presented, and a final project will provide an opportunity for students to develop a more complete project from harvesting data from Open Data portals to synthesizing and analyzing those data to explore a question or problem, to communicating their results in a web map and blog, as well as a final presentation.

COURSE SYNOPSIS AND OBJECTIVES

This course is designed to provide future city planners with a toolkit of technical skills for quantitative problem solving. It requires some tolerance for experimentation, self-directed trial and error, and an interest in learning to write code. If you are willing to roll up your sleeves and embrace some uncertainty, you'll learn the fundamentals of urban data analysis and visualization, and might discover an entirely new lens through which to study, plan, and design neighborhoods, cities, and regions.

Topics to be covered include:

- Fundamentals of programming with Python and IPython notebooks
- Cleaning, manipulating, and analyzing urban data with Python's pandas library
- Using Git for source code version control and collaboration
- Visualizing data in Python with charts, graphs, and tables
- Accessing public data from the web with scraping and APIs (including Twitter, Google, Census data, and the Open Data portals of cities)
- Developing spatial indicators and mapping urban data with open source GIS tools, CartoDB, and Mapbox

COURSE PREREQUISITES

Prior coursework (such as CP 204c) and experience using a GIS is required. Students are not required to have prior programming experience, although it will be beneficial. Python is an accessible language and the course will emphasize learning by doing. Prior or concurrent course work in statistics and data analysis (e.g. CP 204A or 204D) is encouraged as this course will not provide the theoretical foundations of statistical analysis.

This course is open to students from across campus, but priority enrollment will be given to students in the Master of City Planning program.

ASSIGNMENTS AND ASSESSMENT

Students will develop skills gradually through exercises paced over the semester. These will typically involve writing some code and documenting it, using IPython Notebooks that can be shared and interactively run inside a web browser, and a writeup discussing the assignment and its results. Exercises will generally be due Sunday night at 11:59pm.

Each exercise will be worth 20 points and the final project/presentation will be worth 100. Exercises will be tasks that demonstrate a degree of mastery of skills, and will be used as a means of ensuring that students are keeping up with the material and not falling behind. Final projects will require harnessing the skills practiced in the exercises and developing a more independent work plan to accomplish an analysis of data.

Please make sure you arrive on time to class (we will start at 9:40 - please be in class by then). Due to the nature of the computer lab, it can be very distracting to have students trickling in after class has begun. Make sure you have a keycard with access to room 214 and please do not knock on the door to be let in after class has begun. If you are taking this class and are not a Master of City Planning student, you *must* sign up for a computer account and access to the 214 lab. This may require paying a fee. You can sign up in 477 Wurster Hall.

READINGS AND RESOURCES

There will be a reading or set of short readings required for most class sessions. These are essential for coming prepared as the lectures will proceed under the assumption that everyone has already read the assigned reading materials. Most class sessions will begin with a quick discussion of the reading material, and you will have the opportunity to ask us questions about anything you didn't understand.

For the most part, the lecture notes, tutorials and examples provided on this website will provide a good foundation for the skills students will gain in thew course. Some of the key software packages we will be using have free documentation available, and you should make extensive use of these to help you climb the learning curve.

1. Downey, Allen. 2013. Think Python: How to Think Like a Computer Scientist. Green Tea Press. Available under Files > Reading Materials. This text provides a fairly comprehensive overview of Python as a programming language.
2. IPython Documentation. 2013. This online documentation explains what IPython is and how to use it.
3. McKinney, Wes. 2012. Python for Data Analysis. O'reilly. Available under Files > Reading Materials. You may also want to access the online Documentation available in pdf.

This course requires a lot of experimentation and trial-and-error. Google is your best friend! Google your questions, Google any error messages, and if you can't find an answer, email Paul and Geoff. The key to success in this course has been repetition: keeping up on the readings, writing lots of code over and over, and stretching to figure out "what next".

SCHEDULE

Week	Date	Topic	Homework
1	Wed 26 Aug	Introduction and Setup	Exercise 1: Anaconda and OCF
2	Mon 31 Aug	Python fundamental data types: numerics, strings, lists, dicts	
2	Wed 2 Sep	Wordpress	
3	Mon 7 Sep	No Class	
3	Wed 9 Sep	CartoDB	Exercise 2: blog post with CartoDB
4	Mon 14 Sep	Python control: conditions, loops, functions	Exercise 3: Python script
4	Wed 16 Sep	Guest speaker: Eddie Tejeda	
5	Mon 21 Sep	QGIS	
5	Wed 23 Sep	Python control: conditions, loops, functions	Exercise 4: Python script
6	Mon 28 Sep	Pandas basics	
6	Wed 30 Sep	Guest speaker: Matt Davis, using Git	
7	Mon 5 Oct	Pandas basics	Exercise 5: pandas operations
7	Wed 7 Oct	Working with APIs: Sam Maurer	

8	Mon 12 Oct	Guest speaker: Donna Lasala Berkeley Open Data Portal	Exercise 6: API data
8	Wed 14 Oct	Pandas data wrangling	
9	Mon 19 Oct	Guest speaker: Michael Migurski	
9	Wed 21 Oct	Pandas data visualization	Exercise 7: data visualization
10	Mon 26 Oct	Pandas data science	
10	Wed 28 Oct	Pandas data science	Exercise 8: data analysis
11	Mon 2 Nov	Web scraping	
11	Wed 4 Nov	Guest speaker: Sam Maurer - Web Mapping	Exercise 9: web mapping
12	Mon 9 Nov	Reverse geocoding	
12	Wed 11 Nov	No Class	
13	Mon 16 Nov	Pandas spatial analysis	Exercise 10: spatial analysis
13	Wed 18 Nov	Guest Speaker: Prashant Singh and Alicia Roualt: Local Data	
14	Mon 23 Nov	Urban Data Science Toolkit	
14	Wed 25 Nov	No Class	
15	Mon 30 Nov	Workshop	
15	Wed 2 Dec	Workshop	
16	Mon 7 Dec	Final presentations	RRR week: final projects
16	Wed 9 Dec	Final presentations	RRR week: final projects