PLA6619 Introduction to Urban Data and Informatics I

Data for Urban Planning and Design, Fall 2021

Instructor: Boyeong Hong (b.hong@columbia.edu)

Lecture and hands-on lab: Tuesday 11:00am – 1:00pm
Room: UP Planning Lab
E-Office hour: Wednesday after 4:30pm. Please check the calendar and sign up here or by appointment via email.

Course description and objectives

Data analytics and data-driven processes have been used to make urban planning decisions and to improve related city service operations. With the proliferation of digital data, new opportunities are being availed to measure, understand and propose changes to the communities in which we live, work, and play. This has led to a host of new terms and disciplines – urban science, big data, smarter cities, urban informatics, civic analytics – that seeks to understand the intersection of digital technologies and the human environment. The most benefit of those disciplines is not only an in-depth understanding of urban phenomena but also predicting and preparing for future scenarios in cities composed of complex systems. There are immense opportunities with big data and analytic capacities to support responsive and effective urban systems, ultimately aiming at sustainable and livable cities through a problem-driven analytic approach.

This course is intended to provide an introduction to the technical, theoretical and practice-based dimensions of urban analytics. The course is centered around data acquisition, numerical analysis, spatialization, visualization and interaction, and civic technologies. Students will learn major concepts, software tools, and analytical techniques to extract meaningful information from various data sources and analytical practices. Students also will have an opportunity to develop their project that combines the technical aspects in a final analysis and visualization.

This course will engage the role of technologies and computational methods in the planning process. The main objective of this course is to familiarize students with modern computational techniques and demonstrate how they can be applied to real-world problems alongside the planning perspectives (e.g. climate change or public health crisis like the COVID-19 pandemic). Students taking this class will spend most of their time to learn

- Tools essential for working with big urban data, including the Python programming language
- Data acquisition skills including querying APIs and data scraping
- To create and use databases
- To utilize common libraries to analyze data
- To design compelling visualization to communicate with stakeholders and deliver insights

The intent of this class is not to produce experts, but to provide the foundation to begin to formulate urban planning questions and to confidently find the mechanisms by which to answer them. In parallel, the course will discuss the ethical implications of new urban technologies and develop responses to professional ethical situations in urban informatics with the overall ambition of informing future practice.

Course structure

The course is a practice-oriented class, learning concepts and techniques are motivated and illustrated by applications to urban problems and datasets. The class will include a mix of lectures and interactive coding lab sessions. The relevant theoretical background is provided and students will discuss the policy and design questions around the creation of, and use of urban data within the language of planning.

Note: the course will not go into every detail of each technique. That being said, students who wish to engage more with specific techniques / skills / concepts behind urban informatics are encouraged and supported through discussions and further resources.

Prerequisites and software

This course assumes no prior programming experience, but this course is focused on coding and analysis using programming languages and computational thinking. Students who don't have programming experience are welcome, but those should have a strong willingness to learn and build up related skills. Due to the nature of the course, an introductory understanding of programming / quantitative methods / data science / data visualization is undoubtedly helpful. Due to the vast variation in skills present, the general mantra for the class is that course participants are required, at a minimum, to approach the activities and lectures with enthusiasm and/or perseverance. This course will use a variety of software tools and packages. Python (usually through Jupyter notebooks including packages like Pandas, Numpy, and Matplotlib) will be the primary programming language, although R and other tools may also be used. There will be a significant programming component and basic exploratory and visualization abilities will be assumed. A computer is required (either Mac or Windows).

Textbooks and course resources

This course will use a combination of articles, book chapters, and instructor notes. There is no required textbook and the assigned readings will be uploaded. The followings are recommended reference for programming practice and self-learning resources:

- Python Data Science Handbook: Essential Tools for Working with Data by Jake VanderPlas (https://jakevdp.github.io/PythonDataScienceHandbook/)
- Wheelan, Charles. 2013. Naked Statistics: Stripping the Dread from the Data. W. W. Norton & Company.
- McKinney, Wes. 2017. Python for Data Analysis. 2nd ed. Boston: O'Reilly Media.
- Downey, Allen. 2013. Think Python: How to Think Like a Computer Scientist. Green Tea Press. (http://greenteapress.com/wp/think-python-2e/)
- IPython Documentation. (http://ipython.readthedocs.io/en/stable/)
- Murray, S. 2017. Interactive Data Visualization for the Web: An Introduction to Designing with D3. 2nd ed. Boston: O'Reilly Media.
- Tufte, E., & Graves-Morris, P. (2014). The visual display of quantitative information.; 1983.
- Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.
- Munzner, T. (2014). Visualization analysis and design. CRC press.
- https://www.google.com (this is not a joke. Your favorite search engine will certainly pull up possible solutions for coding problems)
- https://stackoverflow.com
- Codeacademy
- https:/www.data-analysis-in-python.org/
- Harvard CS50 (Introduction to Computer Science) and CS109 (Data Science)

On campus, the Department of Statistics offers free consultations to students by appointment (http://stat.columbia.edu/consulting-information/), and the Columbia Libraries with CUIT (https://library.columbia.edu/services/research-data-services.html) offer support for research data preparation and analysis.

Participation

Participation is a key of this course and it is very difficult to pass this course without participation. Firstly, students not attending a full session of class without legitimate reasons will be recorded as absent. For each unexcused absence, one point will be deducted from the attendance grade. In-class participation through asking questions, a group discussion, and a hands-on lab session is the most important piece to engage with this course.

Assignments

There will be four assignments, consisting of problem sets and/or that reinforce and propel topics covered in class. The assignments will be an extension of a lab session. In addition to regular assignments, students will be asked to participate in a contemporary case study once per semester. This assignment is called “urban innovation”, dealing with how topics in class are being used in practice or applied to policy-making. Details will be announced during the first lecture.

Late assignments

Assignments will be deducted 10% for each day a submission is late unless there is a legitimate reason that the instructor is informed of in advance. Assignments later than a week will not be accepted. No late submissions will be granted for the midterm and the final deadlines.

Readings

Imperative to the use of data within the context of planning is understanding the biases and debates with regard to the use of such information. There will be weekly readings assigned, providing a background and frame a foundation for debate. Students will participate in a weekly group discussion.

Midterm and final deliverables

There will be an in-class midterm datathon challenge and a final project.

The datathon typically means a time-limited online competition where people are challenged to work on real-world problems from different areas of data science, computer science, or machine learning. Students will be provided data and problem sets and be asked to analyze data and answer the questions within two hours (in-class online challenge).

Students, as groups, are asked to work on a final project to apply newfound technical savviness to analyze and synthesize urban data around a research question to deliver meaningful planning insights. It requires i) project proposal, ii) project report, iii) presentation, and iv) technical documentation. Details will be announced in due course.

Grading

Grading will be performed through a numerical assessment of students' submitted work. A final score will be translated into the GSAPP grading system. The breakdown is as such:

- Attendance and participation (10%)
- Weekly urban innovation memo and 3 minutes pitch – 2 to 3 students per week (10%)
- Four (4) Assignments; each assignment is 10% point (40%)
- Midterm datathon challenge (20%)
- Final packet (20%)

"High Pass" will be offered to the top 20% of students based on their numerical score and level of participation. "Pass" will be given to all final scores above 75. "Low Pass" will be 60-74, or automatically offered as a maximum if any assignment is missing.

GSAPP honor system and plagiarism

Students must adhere to the principles of academic honesty (https://www.arch.columbia.edu/honor-system) and ensure that all work submitted is fully theirs and adhere to the GSAPP Plagiarism Policy (https://www.arch.columbia.edu/plagiarism-policy) set forth. Students found guilty of plagiarism or academic dishonesty will be subject to appropriate disciplinary action. Academic honesty does seriously matter in this course.

Collaboration and quoting policy*

Coding has unique challenges when it comes to collaboration and plagiarism, so please familiarize with this section.

Firstly, all the work you turn in must be your own (as an individual or as teams, as appropriate). However, you are welcome to discuss course materials, ideas, and assignments with others. When working through code with others, you must not discuss specific code—what you are going to implement within the computer that will be compiled—but you may discuss resources logic, structure and/or pseudo code with others. Nor may you provide or make available solutions to assignments to individuals who take or may take this course in the future. You may not directly use code found on the internet (cut-copy'ing) for assignments.

For the project, you may "quote" from resources online. You must acknowledge any source code that was not written by you by mentioning the original author(s) directly in your source code (comment or header). You can also acknowledge sources in a README.txt file if you used whole classes or libraries. Do not remove any original copyright notices and headers. However, you are encouraged to use libraries, unless explicitly stated otherwise by copyright, the code author or the teaching team! Although you may be using code found elsewhere, it is expected that your final projects are of substantive originality in concept and implementation.

** Developed by Prof. Anthony Vanky*

Writing and Technical Assistance

The strength of GSAPP and the urban planning program is the diversity of experiences among its community members. However, with the diversity of languages, academic writing in English is a difficult art to master. While you will gain practice in communicating to diverse audiences in this class, 1) the writing center is a great resource that you should feel welcome to take advantage of: https://www.college.columbia.edu/core/uwp/writing-center and 2) a doctoral student will be a mentor providing support for academic writing specialized in the urban planning context. Also, a second-year student (Mauricio Enrique Rada Orellana) will help general data and coding related problem solving.

Lecture schedule

Week	Date	Topics	Assignment (release)
01	09/14/2021	Preview of the course Data for Urban Planning and Design Lab 00 - Environment set up	
02	09/21/2021	Fundamentals in data analysis 1 Computational thinking / data and coding environment Lab 01 - Intro to Python	
03	09/28/2021	Fundamentals in data analysis 2 Quantifying cities with statistics primer Lab 02 - Data cleaning and table manipulation	Assignment 01*
04	10/05/2021	Fundamentals in data analysis 3 Data visualization and communication Lab 03 - Exploratory data analysis and visual representation	
05	10/12/2021	Accessing online resources 1 Urban data taxonomy / accessing data Lab 04 - APIs and open data	Assignment 02*
06	10/19/2021	Accessing online resources 2 Data opportunism and data exhaust Lab 05 - Scraping and natural language processing	
07	10/26/2021	Spatial data 1 Geolocation information with urban context Lab 06 - Geopandas	Assignment 03*
08	11/02/2021	Election Day, University Holiday (No class)	
09	11/09/2021[+]	Midterm *In-class datathon challenge**	
10	11/16/2021[+]	Guest lecture - Urban informatics and urban sensing Final project pitch presentation	
11	11/23/2021	Spatial data 2 Origin-destination movements in cities Lab 07 - NetworkX	Assignment 04*
12	11/30/2021	Web visualizing and communicating 1 Introduction to web visualization tools Lab 08 – Plotly and D3	
13	12/07/2021	Web visualizing and communicating 2 People and participation / the crowd Lab 09 - More D3 and web viz publishing	
14	12/14/2021	Final presentation	
15	12/21/2021	Final packet due*	

** Assignment/midterm/final deadlines will be specified*

+ Online session