

*revision 1 | 26 aug*

# Introduction to Urban Data and Informatics

## Data for Urban Planning and Design, Fall 2019

Tuesdays, 11:00a-1:00p

*Wednesday Hands-On Recitations, 5:00-5:50p*

#GSAPPUrbanData

Anthony Vanky, PhD

a.vanky@columbia.edu

324M Fayerweather Hall

**Office Hours:** Tuesday Afternoons 3:00-5:00pm, or by Request

book ahead at <http://vanky.co/officehours>

## I. Summary

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, civic technologies—that seeks to understand the intersection of digital technologies and the human environment. These forces have created new opportunities for planners to make data actionable through analysis and visualization, as well as avenues for new types of practice including startups and data advocacy. Furthermore, many of the most urgent problems facing cities—including those as a result of climate change—are problems known, presently, through their representations in the output of predictive models. These forces and challenges have placed an imperative for planners to develop the facilities to collect, analyze, communicate (visually and textually) using these large and often-messy data sets.

This **Introduction to Urban Data and Informatics** is intended to provide students an entrée into the technical, theoretical and practice-based dimensions of data analysis in cities. The course is centered around technical lectures interspersed with guest presentations and class debates grouped into five topical categories—data acquisition, numerical analysis, spatialization, visualization and interaction, and civic technologies. Students will also have an opportunity to develop their project—based on their research question—that combines these technical aspects in a final analysis and visualization. Within the seminar and lecture sessions, we will discuss the policy and design questions around the creation of, and use of urban data within the language of planning. Seminar and lecture sessions cover topics related to the context and practice models associated with urban technologies, including civic technology, indicators, smart cities, and performance management.

## Learning Objectives

Students taking this class will spend most of their time learning tools essential for working with big data, including the Python programming language. We will begin with the technical skills of data acquisition skills including querying APIs and data scraping, create and use databases, utilize common libraries to analyze data (Pandas), and design compelling visualizations (D3, Leaflet, Carto). Assignments will also challenge you to think about the communicability of these findings to the communities in which we engage through our professional practice. The intent of this class is not to produce experts, but to give you the foundation to begin to formulate your own questions and to confidently find the mechanisms by which to answer them.

In parallel, we will engage in the critical debates around data and planning to become conversant in these vital conversations and learn about how they are being considered through emerging planning practices. We will analyze 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.

## Prerequisites

This course assumes no prior programming experience, but this course is focused on coding and analysis. Due to the nature of the course, an introductory understanding of programming or statistics or quantitative methods 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. **The best way to learn is by doing, but doing != easy.**

## II. Assignments

### Problem Sets

This class will involve completing several problem sets that will have you learn how to code through practice. These may seem somewhat unrelenting, particularly with reading responses and the final project, but this is by design. The best way to learn is practice, practice, practice. These assignments are essential for developing your skills. In general, you are given two weeks to complete a PSet, yet we may be introducing topics concurrent to when you are completing the assignment. Generally, you will have been introduced to the concepts required to complete the assignments at least one week prior to the due date. However, with this said, you should not wait to look and think about the assignment. When they are distributed, you should already think through how you might approach the assignment and offer questions.

Each problem set includes some type of translation toward policymakers, community groups or non-technical audiences; you are asked to distill your findings in a communicable fashion. These documents should comprehensively present the findings and objectives of the analysis, but accurately and appropriately framed for a non-technical audience.

### Participation

It is very difficult to pass this course without participation. Your reading responses are intended to ensure you are engaging with the reading material and to give you time to prepare for in-class participation. You can also participate online via Twitter using #GSAPPUrbanData by sharing current events, your raw thoughts, policy questions, etc. I'll use Twitter to share current events that may not fit within the very tight timing of our course. This will also be considered as part of your participation!

## Responses

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. The selected readings provide a background and frame a foundation for debate. Each day, you're asked to submit a few observations and/or questions based on the readings. In some cases, you will get a specific prompt to consider (such as in preparation for in-class debates or case studies). These are not meant to be exhaustive essays but a few bullet points that distill your consideration.

## Final Project (and Precedent Analysis) ... aka Assignment 5

*"Demo or Die."* – Nicholas Negroponte (but the class will be a little less dramatic about it.)

In the second half of the term, you will have the chance to apply your newfound technical savviness to develop your own project that analyzes and synthesizes urban data around a question of your choosing. You are welcome to utilize any tool or skill you learned in the course, or you may have picked up along the way, to build something that interests you, that solves an actual problem, or impacts the community in a meaningful way. You are allowed to collaborate with one or two classmates, although the scope and effort should reflect the collective energies of the whole team.

The final project will have two major milestones en route to the development and implementation of a working demo. First, you are required to meet with the teaching team/instructor to review your pre-proposal. This is to help you scope appropriately and to make sure the data you are interested in is either feasible or available. The second is the project pitch, where you will present your concept. You will also analyze an analysis, visualization or interactive tool that you find compelling that informs your thinking about how you will approach your project as part of this presentation.

Projects that are visualizations and utilize NYC Open Data are also highly encouraged to submit them to the [Data Through Design](#) competition.

## Case Studies

In order to connect the technical learning in the class to innovative classes, you will be asked to present on a contemporary case study of how topics in class are being used in practice or applied to policy-making. Students, as individuals or as groups, are asked to analyze the data and processes being utilized as well as engage in a critical debate on what the potential pitfalls and gaps may be. You are asked to sign up on a voluntary basis, and credit will be applied through participation. International, in-progress, or projects in which you engaged are welcome.

## Grading

Grading will be performed through a numerical assessment of your submitted work. Your final "score" will be translated into GSAPP's grading system. The breakdown is as such:

Attendance and Participation .....	15
(Quick Case Study .....	5)
Problem Sets (PSets, Assignments) .....	40
Responses .....	15
Final Project .....	30

Proposal Pitch and Precedent Analysis ...	5
Final Presentation .....	5
Implemented Deliverable .....	20

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

### III. Schedule

	Contextual	Practical	Due
<b>Getting Set Up</b>			
Tue, 03 Sep	Introduction & Syllabus	Data and Coding Environment I	
Wed, 04 Sep		Getting Set Up - iPython and its Variants	
Mon, 09 Sep			<b>Problem Set 0 - Getting Into It</b>
<b>Fundamentals in Data Analysis</b>			
Tue, 10 Sep	Introduction to Data & Computational Thinking	Data and Coding Environment II	
Wed, 11 Sep		Data Cleaning and Tables (Pandas)	
Tue, 17 Sep	The promise of quantification... Of cities?	Statistics Primer	
Wed, 18 Sep		Table Manipulation + Matplotlib	
Mon, 23 Sep			<b>Problem Set 1 - Racial Disparities in Policing</b>
<b>Accessing Online Resources</b>			
Tue, 24 Sep	The Politics of Data	Accessing Online Data	
Wed, 25 Sep		APIs and Open Data	
Tue, 01 Oct	Data Opportunism and Data Exhaust		<i>Case Study - Quick Presentation</i>
Wed, 02 Oct		Scraping	
Tue, 08 Oct	Algos and Ethics		<i>Case Study - Quick Presentation</i>
Wed, 09 Oct		Machine Learning + the Cloud (Sentiments)	
Mon, 14 Oct			<b>Problem Set 2 - "Yum" as a Planning Dataset</b>
<b>Visualizing and Communicating</b>			
Tue, 15 Oct	Visuaization and Communication	Introduction to Tools	
Wed, 16 Oct		d3 and d3 plus / Python Viz (brief)	
Tue, 22 Oct	Agency and Reciprocity	more d3 and d3 plus	
Wed, 23 Oct		(none)	
Tue, 29 Oct	People and Participation / The Crowd		<i>Case Study - Quick Presentation</i>
Wed, 30 Oct		Publishing for the Web	

Mon, 04 Nov			<b>Problem Set 3 - Geography and Upward Mobility</b>
<b>Spatial Data</b>			
Tue, 05 Nov	<b>A5 (Final) - Pitch Presentations</b>		<b>Problem Set 5 - Proposal Presentations!</b>
Wed, 06 Nov		Geopandas	
Tue, 12 Nov	Spatial Analysis in Context		
Wed, 13 Nov		Ushahidi (extra!)	
Tue, 19 Nov	Big Data and Small Spaces	Sensors and Machine Data	<i>Case Study - Quick Presentation</i>
Wed, 20 Nov		SQL and PostGIS (extra!)	
Mon, 25 Nov			<b>Problem Set 4 - Mobility Differences in New York</b>
<b>Hacking Planning</b>			
Tue, 26 Nov	New Regulations and Regimes	Technical Meetings	<i>Case Study - Quick Presentation</i>
Tue, 03 Dec	A5 (Final) - Presentations!		<b>Problem Set 5 - Final Presentations!</b>

## IV. Recitations and Outside of Class

### Technical Recitation

A session is available Wednesday evenings as a resource. In some cases, we will go in depth on tools and techniques or do hands-on workshops that may assist with the PSets. Although these are optional, and there is no attendance requirement, they are highly recommended where we will go through coding exercises that allow you to practice. Materials from this will be made available as a resource after the session.

### Conversations over Coffee

Part of the introductory nature of this class is to expose you to practitioners who are applying the course concepts in practice, and we are fortunate to have some of the world's best researchers as guest lectures. These will almost exclusively happen outside the timing of the course, and most likely online via a Hangout. Come in your PJs, and have a chat with interesting professionals and researchers.

## V. Unsolicited Advice

### “Why won’t this work?!!”

It is guaranteed that you will be frustrated in this class—your code will not work and you won’t know why, or you’ve tried several solutions and they all fail. Technical learning is that it is best accomplished by doing, failing, failing to the point of frustration and doing some more. Expect frustrating moments to come, and be prepared to persevere past them. Mastering a technology always involves learning how to seek help with problems that exceed your understanding, even when Google or your chosen search engine fails you. It is worth noting that much of your learning will come through searching for answers, figuring out how to adapt what you found, and iterating until it works. **Use each other as resources and share resources amongst yourselves.** (Anecdotally, I once spent 8 hours frustratingly reviewing a modest piece of code that would not compile due to a single errant semi-colon.)

The best advice I can offer is to start early, and ask for help. Should you feel that despite your perseverance and resourcefulness that you are still finding problems, you should not hesitate to contact the instructor.

PS... You may want to remember this statement exists for later in the semester! <3

### **Writing 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, the writing center is a great resource that you should feel welcome (and excited!) to take advantage of:

<https://www.college.columbia.edu/core/uwp/writing-center>.

## **VI. Policies**

### **Late Submissions**

Late submissions for responses and PSets will incur an automatic penalty of 25% for each day late (compounding), up to 2 days, after which assignments are no longer accepted. But, things\* come up... You are granted a penalty-free one-day late submission that will be applied at the end of the course—no questions asked. However, no late submissions will be granted for any assignment related to the Final Project.

An impetus for this policy is that the technical material builds upon itself in this course, and falling behind makes it difficult to understand the new material. This is only to aid you in maintaining pace through the semester through what is sometimes difficult material.

\*things = conferences, field work, studio deadlines, etc.

### **Collaboration & Quoting Policy**

Coding has unique challenges when it comes to collaboration and plagiarism, so please familiarize yourself 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 PSets 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 PSets.

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.

### **Academic Integrity**

This course will strictly enforce the GSAPP honor code, which can be viewed at <https://www.arch.columbia.edu/honor-system>. Plagiarism, including the use of another's work in the class, is automatic grounds—at minimum—for failing. For more information, please refer to <https://www.arch.columbia.edu/plagiarism-policy>.

P.S.--Your code may be run through a standard program to check for plagiarism and quoting. It's been in use for about a decade across many universities, so please don't copy. Don't cheat.

### **Accommodations**

The College will make reasonable accommodations for persons with documented disabilities. Services are available only to students who have registered and submit appropriate documentation. As your instructor, I am happy to discuss specific needs with you as well. Please report any access related concerns about instructional material to Office of Disability Services and to me as your instructor.

### **Sexual Harassment and Violence Reporting**

The University is committed to maintaining a safe environment for students. Because of this commitment and because of federal and state regulations, we must advise you that if you tell any of your instructors about sexual harassment or gender-based misconduct involving a member of the campus community, your instructor is required to report this information to the Title IX Coordinator, Margorie Fisher. While this information is considered private, it may lead to follow up. For more information on these policies, see <https://www.arch.columbia.edu/discrimination-policy>.

## **VII. Course Resources**

Students are not required to purchase any books for the course, although you will need to obtain a few case studies (as noted in the readings). The assigned readings for this course will be uploaded to the website, and a range of books will be placed on reserve for your perusal. (References are provided in the appendix of this syllabus.)

Additionally, participants may find the following resources useful for additional programming practice or reference. Many before have faced similar stumbling blocks in one's one analysis and projects, and your favorite search engine will certainly pull up possible solutions for your problem. Among many others, I recommend the following:

<http://stackoverflow.com>

Codecademy

Learn Code the Hard Way

<http://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. These services should only be used for assistance with your PSets, and should not be used past an advisory or consultative capacity.

And in addition, the following (e-)books from O'Reilly Media may be of great use, and worth a purchase for your own library. Many books from that publishing house may be available in print or electronic copy from the library.

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, Scott. 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.

## **VIII. Readings**

*Refer to supplemental document.*