CRP and DESIGN 4580/5680:
Introduction to Urban Data Science: Data, Interpretation, and Presentation
Spring 2025

Instructor: Wenzheng Li (wl563@cornell.edu)
Class time: 8:40 – 9:55am Tuesday and Thursday, Sibley Hall 101
Lab session: 4:45pm – 5:35pm Thursday, Sibley Hall 305
Office Hours: Monday 2:30 – 4:30pm and Wednesday 11:00am-1:00pm in Sibley Hall 214. Book a time here.

TA and GTRS
Yujin Hazel Lee (TA) yl3276@cornell.edu
Office hours: Thursday 5:30-6:30 pm in Sibley Hall 305
Xi Guan (GTRS) xg298@cornell.edu
Office hours: Tuesday 4:30-5:30pm in Sibley Hall 305

Course Description
Urban data science is an emergent practice in geography and urban planning that combines: 1) the set of data analysis tools and methods used to understand a wide array of big data and big spatial data sources and, 2) questions of urban development, structure, complexity, theory, policy, dynamics, and outcomes. These approaches enable more spatiotemporally dynamic and granular analyses of cities and allows researchers new insight into urban dynamics.

This course will provide a toolkit to speak through data, code, statistics, and visualization. Using open-source data and computational tools in Python and the Jupyter Notebook environment, we will learn how to design testable research questions, collect and prepare data, apply relevant analytical techniques, present our process and results in an engaging and informative way, and identify the limitations of quantitative analysis. A personal laptop will be required.

Learning Objectives and Outcomes
The goal of this course is to provide an introduction to a wide range of tools and concepts that will enable future, deeper exploration of urban dynamics. In other words, this course provides a "sampler" of the foundations of urban data science through coding, statistical analysis, visualization/narrative-building, and critique.

The core learning objectives are:
1) Use code to clean, analyze, and visualize spatial data
2) Implement a descriptive or predictive analysis using appropriate data and statistical and/or computational methods
3) Clearly communicate your process and results as a data narrative through visualizations, context, textual description, and presentation
4) Identify the limitations and potential biases in the data, data-generating processes, and tools and methods in addressing your research topic

Class Structure
Weeks 1-13: Every class will include a brief lecture and a tutorial at the end of the class to practice the concept you just learned. There are 4-5 coding homeworks during this period.

Weeks 14-16: The last three weeks of the class will be devoted to a final project of your choosing that addresses an urban development question. The aim is to synthesize and further develop the skills you have learned throughout the course of the semester. The proposals for the project will be due Week 10. In

class for Weeks 14 and 15, each project group or individual will meet with me to discuss your progress, technical or conceptual roadblocks, and next steps. Our last two classes of the semester will be devoted to final presentations. Your final projects will be due on **May 16 at 11:59 pm.**

Course Prerequisites
This course is designed for masters students and upper class undergraduate students.

- CRP 4080/5080 (Intro to GIS) or an equivalent GIS course is a prerequisite for this course.
- Additionally, I assume you have some basic statistics knowledge, such as hypothesis testing, basic regression and some familiarity using spreadsheet software (Excel, Google spreadsheets).
- Prior or concurrent coursework in quantitative methods, visualization, and programming is recommended.

Technology
A personal laptop with permissions to install software is required for this course. Your laptop can be Mac, Windows, or Linux. We will be using entirely Free and Open Source Software (FOSS). We will set up the Python coding environment during the first week of this semester.

Assignments and Grading
The assignments in this course will consist of completion of in-class exercises, four or five homework assignments, a project proposal, and a final project including a presentation on your project. Grading breakdown as follows:
In-class exercises: 15%
Homework: 35%
Final Project Proposal: 5%
Final Project: 35% (5% presentation + 30% paper)
Class attendance and Participation: 10%

In-class exercises: All in-class exercises are meant to be an opportunity for you to practice the concepts taught in class. These will be 100% if submitted with completed answers without errors and 0% if not submitted or completed. There is in-class exercise in each class. Submit both by Thursday per week at **11:59pm** as a single file (e.g., a published html or PDF). TA will assist you in finishing these in-class exercises during the Thursday lab session.

Homework assignments: The homework assignments will be released at least a week before they are due. They should be submitted as a published html or PDF file with your code, answers, and results. In addition, a zip file of all the relevant notebooks and datasets for the assignment are required. Students are responsible for ensuring that their submissions go through in time. *Submit early to avoid tech issues and please check that the notebooks run without error*.

Late submission will automatically be downgraded unless there is a medical or family emergency:
2%: if turned in the following day;
8%: if turned in within 1 week;
20%: after 1 week.

Final Project: The final research project will address an urban question using the tools and concepts from the course. It will consist of presenting the issue, its context and background, relevant data analyses and visualizations, conclusions, and limitations to the analysis. In addition to the project proposal due on **March 28 at 11:59pm**, the final deliverables will consist of a presentation, a well-documented Jupyter Notebook and associated datasets and an in-class presentation. The specific project prompt and grading criteria will be announced at a later date. The final project deliverables are due **May 16 at 11:59pm.** Late

submission of the final projects will not be accepted. The final projects can either be individual submissions or submissions of groups of two.

Attendance: If you need to miss class, please let me or the PhD TA know beforehand. Arriving at class 20 minutes later counts as an absence. **Four** unexcused absences from either day of class will result in lowering one level grade (for example: A+ to A). **Six or more** unexcused absences will result in lowering one letter grade (for example: A to B) in the course. If you have to miss class for any reason, please get the notes from class from a classmate. If you miss class, I still expect you to complete the in-class exercise.

Lab
The labs are an opportunity to practice and review the concepts from the week. Additionally, our TAs will discuss the weekly in-class exercises and homework assignments. Attendance of labs is required for the course.

Generative AI
Tools such as ChatGPT can often facilitate the generation of functions, processes, and frameworks in coding. They can be a useful aid in our analytical process. As such, we will learn how to work with ChatGPT in our homework assignments. *The use of generative artificial intelligence (AI) tools is permitted for coding with proper attribution*. Additionally, as I mentioned in the "Academic Integrity" section, be prepared to verbally explain what your code is doing. There are some very fundamental concepts in Python, machine learning, and regression that I would like you to actually learn.

Undergrad vs. Grad:
There is an undergrad and graduate version of this course. My expectations for graduate students enrolled in this course is for the final projects to be more research-oriented. I will expect you to complete a literature review for your proposal in order to justify the research question you investigate.

A note about learning to code: Coding is an iterative process involving (a lot of) trial and error, patience, self-direction, and clever Googling. It can seem daunting and intractable at first. Here are some guidelines and resources to help you through this process:

1) Look for typos in the code.
2) Search for the issue on Google. This will often lead you to sites such as Stack Overflow or Medium, which provides code snippets and sometimes step-by-step instructions on how to resolve your question. Try to be specific in your search. Do not be afraid to sound silly. My search generally involves the following keywords:
    a. [language or tool] ex: "Python", "Pandas", "Matplotlib"
    b. [function or action] ex: "plt.subplots", "plotting multiple plots in one figure"
    c. [error or issue] ex: "plots are tiny", "not showing all plots", etc.
3) If trying to implement a fairly standard process, look through our class notebooks or the readings. There are often code snippets for reference there.
4) Ask classmates.
5) If none of the above is fruitful, a) come to our TAs and my office hours. b) you may want to message Yujin (Hazel), Xi Guan, and me on the discussion session via Canvas with the specific task you are trying to implement and the relevant code snippet either as a screenshot or a Github Gist. Do not send code in the body of an email as rich text editors often add hidden formatting that can introduce new code errors.

Academic Integrity
By its nature, coding involves sharing and replication, especially given the availability of online resources. However, when using significant chunks (around five lines is a good rule of thumb) of

repurposed code, make sure to indicate the source in a comment, tailor the code for your specific needs, and be prepared in class to explain how the code works.

Students are expected to follow Cornell University's Code of Academic Integrity. Violations of the Code such as plagiarism (from any source, including fellow classmates) can result in failure or even expulsion from Cornell. Group work should summarize each student's contribution.

Disabilities and Health
If you need a disability-related adjustment in the course, please meet with Student Disability Services (SDS) and provide me an accommodation letter. We can also meet in private to discuss adjustments to the course you may need. Also, know that Cornell has resources for mental health for anyone who may need it.

If you test positive for COVID, please submit your status to Daily Check. This will trigger an accommodation period. If students are still struggling with the impacts of long COVID or health issues, you are again encouraged to get an accommodation letter from SDS, send this, and discuss accommodations with me.

Inclusivity Statement

We understand that our members represent a rich variety of backgrounds and perspectives. The Department of City and Regional Planning is committed to providing an atmosphere for learning that respects diversity. While working together to build this community we ask all members to:

- share their unique experiences, values, and beliefs.
- be open to the views of others.
- honor the uniqueness of their colleagues.
- appreciate the opportunity that we have to learn from each other in this community.
- value each other's opinions and communicate in a respectful manner.
- keep confidential discussions that the community has of a personal (or professional) nature.
- use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the Cornell community.

Bias Reporting
If you have seen, heard, or experienced bias in any form you should make a report. This helps to track our culture, progress, and flag problematic behavior and can help in identifying patterns long term. A bias report is not a punitive action, and can be made anonymously.

Since 2000, Cornell University has had a program to track bias that is occurring on all campuses in an effort to be proactive in creating an inclusive climate for all. You can access resources or report incidents that are experienced or witnessed, in campus, university event or virtually using the Bias Reporting Portal at: https://diversity.cornell.edu/our-commitments/bias-reporting-cornell

In addition to university-wide civil rights compliance, the Office of Institutional Equity and Title IX (OIETIX) is responsible for collecting and tracking all reported bias activity that occurs at Cornell University that could potentially impact our commitment to diversity and inclusion, including all reports made by faculty, staff, students, and visitors to Ithaca, Weill Cornell Medicine, and Cornell Tech campuses.

Office of Institutional Equity & Title IX
Students who experience sexual violence, sexual harassment, or discrimination based on gender or sexual

identity are encouraged to report their experience to the Office of Institutional Equity and Title IX titleix@cornell.edu to explore formal and informal reporting options, and explore the support and resources available. Information shared in class assignments, class discussions, and at public events do not constitute an official disclosure, and faculty and staff do not have to report these to the Title IX Coordinator. Any other disclosure to faculty and staff needs to be reported to the Title IX Coordinator.

Nearly all faculty and staff members are legally required to disclose any reports of sexual misconduct to the Title IX Coordinator and cannot guarantee confidentiality. Confidential resources are not required to disclose reports of sexual misconduct to the

Title IX Office. All students have the right to report to local law enforcement (Cornell Police, Tompkins County Sheriff, or New York State Police.) All students have the right to be protected by the University from retaliation for reporting an incident and to receive assistance and resources from the University. More information at: https://titleix.cornell.edu

Religious Observances

Cornell University is committed to supporting students who wish to practice their religious beliefs. Students are urged to discuss religious absences with their instructors well in advance of the religious holiday so that arrangements for make up work can be resolved before the absence. Cornell's faculty are governed by the Faculty Handbook, which requires them to provide reasonable accommodations to students when their religious observance conflicts with exam-taking, class attendance, and other course-related requirements in compliance with New York State law. The Office of Spirituality and Meaning-Making (OSMM) maintains a religious accommodation website that is an extremely valuable resource for both students and faculty. If you have questions or concerns, you may call the OSMM office at 607.255.6002 or email Oliver Goodrich, Associate Dean of Students for Spirituality and Meaning-Making. You may also contact the Office of the Dean of the Faculty by email or by phone at 607.255.4843.

Syllabus Change Policy

Except for changes that substantially affect implementation of the evaluation (grading), the activities and assignments described below will be subject to change at any time during the semester to allow class time and activities to maximize the course's benefit to the students, depending on their ongoing development and interests.

Documentation

Course work and activities may be captured and recorded to document course activities and/or progress. The documentation and display may include identifiable student images, voices, and likenesses.

Course Schedule

Week	Tuesday	Thursday
	Section 1: Introduction to Python and Data Techniques	
1	**Jan 21** **Introductions and Course Overview** • Read over the syllabus together • Introducing ourselves	**Jan 23** **Coding environment setup** • Setting up your Python and Anaconda coding environment

	• Open science and the modern urban data science software stack	**Lab Session:** • Setting up the coding environment. • Organize notebooks through markdown
2	**Jan 28** **Basics of Python:** • Basic syntax; • Variables and flows; • List, tuple, dictionary, set; • If-statement and for-loop.	**Jan 30** **Data management using *Pandas* 1** • Python Packages • Basics of Pandas: DataFrame, import and export datasets, built-in functions
3	**Feb 4** **Data management using *Pandas* 2** • filtering a DataFrame: indexing and slicing • Data cleaning	**Feb 6** **Data management using *Pandas* 3** • Linking datasets, • overlaying and aggregating data, • re-classifying data with pandas
4	**Feb 11** **Data visualization:** • Basic plots using *Pandas*, *Matplotlib*, and S*eaborn* • Customizing your plots • Interactive visualization using *Folium* and *Bokeh*	**Feb 13**
	Section 2: Exploratory Spatial Data Analysis (ESDA) and Spatial Econometrics	
5	**Feb 18** **NO CLASS – FEBRUARY BREAK**	**Feb 20** **Geospatial operations 1:** • Basics of GeoPandas • Geometry and Projection • Spatial join
6	**Feb 25** **Geospatial operations 2:** • Spatial data visualization • Choropleth maps using *GeoPandas*	**Feb 27** **Spatial data analysis 1:** Spatial weights with *pysal*
7	**Mar 4** **Spatial data analysis 2:** Spatial autocorrelation with *pysal*	**Mar 6** **Spatial data analysis 3:** Point pattern analysis
8	**Mar 11** **Regression 1:** Linear regression with *statsmodels* and *scikit-learn*	**Mar 13** **Regression 2:** Spatial regression with *pysal*
9	**Mar 18** **Data Collection 1:** • Google Map APIs for geocoding and distance calculation	**Mar 20** **Data Collection 2:** • Web-scraping or using OSMnx package to obtain OpenStreetMap dataset
	Section 3: Machine Learning	
10	**Mar 25** **Unsupervised learning 1**: Dimensionality reduction and K-means clustering with *scikit-learn*	**Mar 27** **Unsupervised learning 2:** Spatial clustering through DBSCAN
11	**Apr 1** **NO CLASS – SPRING BREAK**	**Apr 3** **NO CLASS – SPRING BREAK**

12	**Apr 8** **Supervised learning 1:** Ensemble learning with decision trees and random forest models with scikit-learn	**Apr 10** **Supervised learning 2:** Regression vs classification, model selection, bias-variance tradeoff, and cross-validation with scikit-learn
13	**Apr 15** **Special topics**: • Spatial networks or • Web-scraping or • Guest speaker: Interactive mapping or • Guest speaker: Natural language processing	**Apr 17** **Guest Speaker:** Prof.Waishan Qiu from the University of Hong Kong
14	**Apr 22** In-class work and one-on-one final projects meetings with Wenzheng	**Apr 24** In-class work and one-on-one final projects meetings with Wenzheng
15	**April 29** In-class work and one-on-one final projects meetings with Wenzheng	**May 1** **Final Project Presentations 1**
16	**May 6** **Final Project Presentations 2**	**Final project materials due May 16 at 11:59pm**