Syllabus - URP 6224 Intermediate Urban Analytics
2024 Spring

Instructor

Shenhao Wang, Ph.D. (Google Scholar)
Director of Urban AI Lab @UFL
Email: shenhaowang@ufl.edu
Thursday morning: 8:30am – 11:30am
Location: Arch 0439
Office hour: TBD

Course description

This course introduces the primary modeling paradigms to analyze cities with an emphasis on analytical perspectives and urban applications. The course consists of five modules with Module 1 focused on urban analytical basics, Modules 2-4 on statistical analysis, network science, and machine learning, and Module 5 on the final project. Specifically, Module 1 introduces Python syntax, urban data sources, static data description, and other urban analytical basics. Module 2 presents urban statistical analysis, mainly focused on linear regressions from the statistics tradition with applications to urban economy and mobility. Module 3 introduces the spatial visualization and regressions for urban network analysis. Module 4 discusses machine learning, including supervised learning, unsupervised learning, and deep learning with applications to mobility and socioeconomic predictions. It also discusses algorithmic fairness and analytical perspectives for computational justice in cities. Students will learn Python packages, such as Pandas, GeoPandas, and Scikit-learn to analyze urban mobility, economic development, energy consumption, and housing through the in-class Python Labs. This course focuses on analytical intuition and urban applications, rather than theory or math foundations. It provides future urban planners, designers, and engineers the critical analytical capacity to understand cities and address upcoming urban challenges.

Course prerequisites

No strict prerequisite course is needed. However, students have prior coding experiences in Python or need to take Practicum AI at UFL as a concurrent requirement. Prior knowledge in probability, statistics, and linear algebra can also facilitate your learning experiences. If you have questions, please contact the instructor to discuss your qualification.

Coding

This course uses Python as the programing language. Students are expected to use Google Colab in the lab sessions. You don't need to install Python and other libraries on your laptop because Colab provides a standardized coding platform. The course will teach Python coding modules including Numpy, Pandas, Matplotlib, Statsmodels, GeoPandas, and Scikit-Learn.

Textbook

No textbook is required. The four textbooks are recommended as references for Modules 1-4.

1. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc. GitHub link here.
2. Stock, J. H., & Watson, M. W. (2011), Introduction to Econometrics, 3rd Edition.
3. Newman, M. (2018). Networks. Oxford University Press, second edition.
4. Bishop, C. M. (2006). Pattern recognition and machine learning. GitHub link here.

Course schedule

Week	Dates	Lectures	Lab sessions	Work sessions	Psets	Project
1	Jan 11	Class overview				
Module 1: Urban analytical basics						
2	Jan 18	Urban data sources and Python basics	Lab 02. Python in Colab and GitHub basics			
3	Jan 25	Data description and visualization	Lab 03. ACS and spatial data			
4	Feb 1	Correlational analysis and urban application	Lab 04. Gasoline consumption and cities		pset 1 out	
Module 2: Urban statistical analysis						
5	Feb 8	Univariate linear regression	Lab 05. Car ownership regression part 1			idea guideline out
6	Feb 15	Multivariate linear regression	Lab 06. Car ownership regression part 2		pset 1 due	
7	Feb 22	Urban application in linear regressions: built environment and travel behaviors	Lab 07. Built environment and travel behavior		pset 2 out	idea due
Module 3: Urban network analysis						
8	Feb 29	Network representation and spatial visualization	Lab 08. Visualizing property values			
9	Mar 7	Spatial autocorrelation and regression for property values	Lab 09. Spatial analysis of property values		pset 2 due pset 3 out	proposal guideline out
NA	Mar 14	No class (spring break)	NA			
10	Mar 21	Review, mid-term survey, and proposal presentation	NA			proposal presentation; proposal due
Module 4: Machine learning in cities						
11	Mar 28	Supervised learning: classifications and logistic regression	Lab 11. Predicting sustainable travel behavior part 1		pset 3 due pset 4 out	
12	Apr 4	Supervised learning: diverse classifiers for mobility predictions	Lab 12. Predicting sustainable travel behavior part 2			
13	Apr 11	Unsupervised learning and analytical fairness	Lab 13: Clustering to redefine spatial boundaries	Y	pset 4 due	
Module 5: Final Project						
14	Apr 18	Work session	NA	Y		
15	Apr 25	Final presentation	NA			final presentation; final report due

Note: This schedule is subject to changes.

Course communication
The instructor can be reached through Canvas inbox or by email. Expect a response within 48 hours, excluding holidays and weekends. General questions can be posted to the Canvas class website discussion board.

Details in Course Schedule

Practicum AI provides a wonderful overview for using Python, Colab, GitHub, and other tools in computation. It is highly recommended for the students to read the materials.

- Link to Practicum AI: https://practicumai.org/
- GitHub repository: https://github.com/PracticumAI

Module 1. Urban analytical basics

- Google Colab: Google Tutorial about Colab features
  - https://colab.research.google.com/notebooks/basic_features_overview.ipynb
- GitHub: Getting started with GitHub
  - https://docs.github.com/en/get-started/quickstart/hello-world
- Python basics.
  - 30-min crash course: https://github.com/srebalaji/python-crash-course
  - 4-hour crash course: https://github.com/Python-Crash-Course/Python101
  - MIT Open Course for Python Programming: https://ocw.mit.edu/courses/6-0001-introduction-to-computer-science-and-programming-in-python-fall-2016/
- Python Pandas and Matplotlib for data description and visualization
  - Pandas and Matplotlib: Chapters 3 and 4 in VanderPlas, J. (2016) Python data science handbook: Essential tools for working with data.
  - Python tutorials: Chapters 3 and 4 in the GitHub repository: https://github.com/jakevdp/PythonDataScienceHandbook/tree/master/notebooks
- Data Sources.
  - American Community Survey: https://www.census.gov/programs-surveys/acs
  - Census Geography and TIGER Lines: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
- Readings
  - (Required) Newman, P. W., & Kenworthy, J. R. (1989). Gasoline consumption and cities: a comparison of US cities with a global survey. Journal of the American Planning Association, 55(1), 24-37.

Module 2: Urban statistical analysis

- Python statsmodels for regressions
  - Getting started: https://www.statsmodels.org/stable/gettingstarted.html
  - Linear regression example: https://www.statsmodels.org/stable/regression.html
- Readings
  - (Required) Cervero, R. and K. Kockelman (1997). "Travel demand and the 3Ds: density, diversity, and design." Transportation Research Part D: Transport and Environment.
  - (Optional) Baum-Snow & Kahn, (2000) "The effects of new public projects to expand urban rail transit." Journal of public economics 77(2): 241-263.
  - (Optional) Glaeser, Scheinkman, & Shleifer, (1995) "Economic growth in a cross-section of cities." Journal of monetary Economics 36(1): 117-143.
  - (Optional) Chapters 4, 6, and 9 in James Stock and Mark Watson (2011), Introduction to Econometrics, 3rd Edition.

- (Optional) Chapter 3 in Bishop, C. M. (2006). Pattern recognition and machine learning.

Module 3: Urban network analysis

- Python GeoPandas, Pysal, and NetworkX for spatial analysis.
  - GeoPandas overview: https://geopandas.org/en/stable/index.html
  - GeoPandas GitHub tutorial: https://github.com/jorisvandenbossche/geopandas-tutorial
  - Spatial visualization with GeoPandas: https://geopandas.org/en/stable/docs/user_guide/mapping.html
  - Spatial regressions with Pysal: https://sustainability-gis.readthedocs.io/en/latest/lessons/L4/spatial_regression.html
  - NetworkX: https://networkx.org/documentation/stable/tutorial.html
  - Spatial Data Science Python tutorials: https://sustainability-gis.readthedocs.io/en/latest/index.html
- Readings
  - (Required) Hui, E. C., Chau, C. K., Pun, L., & Law, M. (2007). Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix. Building and environment, 42(6), 2333-2343.
  - (Optional) Chapters 3, 7 and 8 in M. Newman (2018). Networks. Oxford University Press, second edition.
  - (Optional) Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782.
  - (Optional) Eagle, N., et al. (2010). "Network diversity and economic development." science 328(5981): 1029-1031.
  - (Optional) LeSage, J. P. and R. K. Pace (2008). "Spatial econometric modeling of origin-destination flows." Journal of Regional Science 48(5): 941-967.

Module 4: Machine learning in cities

- Python Scikit-learn for machine learning.
  - Scikit learn tutorial: https://scikit-learn.org/stable/
  - Supervised learning tutorial in Scikit-learn: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
  - Unsupervised learning tutorial in Scikit-learn: https://scikit-learn.org/stable/unsupervised_learning.html
  - GitHub tutorials for Chapters 4, 5, and 7 for Bishop, C. M. (2006): https://github.com/gerdm/prml
- Readings
  - (Required) Wang, S., Mo, B., Hess, S., & Zhao, J. (2021). Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. ArXiv:2102.01130.
  - (Optional) Blumenstock, J., et al. (2015). "Predicting poverty and wealth from mobile phone metadata." science 350(6264): 1073-1076.
  - (Optional) Jiang, S., Ferreira, J., & González, M. C. (2012). Clustering daily patterns of human activities in the city. Data Mining and Knowledge Discovery, 25(3), 478-510.
  - Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation research part C: emerging technologies, 19(3), 387-399.
  - (Optional) Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29, 3315-3323.
  - (Optional) Chapters 4, 5, and 7 in Bishop, C. M. (2006). Pattern recognition and machine learning.

Grading

Your grade consists of the follow three components.

Components	Total points	Percentage of final grade
Course participation	100	10%
Problem sets (4)	100 each	40%
Project	100	50%

Course participation (10 pts)

Attendance and participation in the class are required. Attendance and participation grade will be computed in proportion to the number of presences. Students are also highly encouraged to get engaged in the class discussions. The university policy can be found here.

Problem sets (40 pts)

Problem sets are designed to help you learn how to apply the analytical tools to cities. Students are allowed to work in groups, as long as each group is comprised of no more than three people and each member submits their own written answers.

1. Pset 1 (10 pts) – Urban analytical basics.
2. Pset 2 (10 pts) – Urban statistical analysis.
3. Pset 3 (10 pts) – Urban network analysis.
4. Pset 4 (10 pts) – Machine learning in cities.

Project: three stages (50 pts)

1. Idea (5 pt). Limit to 1 page.
2. Proposal (10 pt). Limit to 3 pages.
3. Final paper (30 pt). Limit to 8 pages.

Grading scale

The following table is used as **an example only**.

	Percent Grade	4.0 Scale		Percent Grade	4.0 Scale
A	94-100	4.0	C	73-76	2.0
A-	90-93	3.67	C-	70-72	1.67
B+	87-89	3.33	D+	67-69	1.33
B	83-86	3.0	D	65-66	1.0
B-	80-82	2.67	E/F	Below 65	0.0
C+	77-79	2.33			

The grading follows the university policy here.

Late Submission

It is important to meet deadlines. All work must be completed and submitted by the designated date and time on Canvas. However, life is sometimes uncertain. Therefore, you are allowed to submit your assignments late, but with **2 points deducted for every 24 hours**. For example, if an assignment is submitted 1 hour after the deadline, its full grade will automatically drop from 10 points to 8 points. This policy applies to both Psets and the milestones of the project.

Other UF policies and resources

Accommodating Students with Disabilities

Students requesting accommodation for disabilities must first register with the Dean of Students Office (https://disability.ufl.edu/). The Dean of Students Office will provide documentation to the student who must then provide this documentation to the instructor when requesting accommodation. You must submit this documentation prior to submitting assignments or taking the quizzes or exams. Accommodations are not retroactive, therefore, students should contact the office as soon as possible in the term for which they are seeking accommodations.

Academic integrity and UF honor code

Academic honesty and integrity are fundamental values of the University community. UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Course evaluation

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available here. Students will be notified when the evaluation period opens and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via this link. Summaries of course evaluation results are available here.

Other academic resources

- For On-Campus URP Students: Graduate Coordinator contact information: Laura Dedenbach, laurajd@ufl.edu, 352-294-1493.
- Career Connections Center: Reitz Union Suite 1300, 352-392-1601. Career assistance and counseling services career.ufl.edu/.
- Library Support: various ways to receive assistance with respect to using the libraries or finding resources. cms.uflib.ufl.edu/ask
- Teaching Center: Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420. General study skills and tutoring. teachingcenter.ufl.edu/
- Writing Studio: 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers. writing.ufl.edu/writing-studio/
- For issues with technical difficulties for E-learning in Canvas, please contact the UF Computing Help Desk at 352-392-4357 or via e-mail at helpdesk@ufl.edu
- Student Complaints: sccr.dso.ufl.edu/policies/student-honor- code- student-conduct-code/

Health and wellness

- U Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit umatter.ufl.edu/ to refer or report a concern and a team member will reach out to the student in distress.
- Counseling and Wellness Center: Visit counseling.ufl.edu/ or call 352-392-1575 for information on crisis services as well as non-crisis services.

- Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need or visit shcc.ufl.edu/.
- University Police Department: Visit police.ufl.edu/ or call 352-392-1111 (or 9-1-1 for emergencies).
- UF Health Shands Emergency Room / Trauma Center: For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; ufhealth.org/emergency-room-trauma-center