CP101- Introduction to Urban Data Analytics [4 credits]

Lead Instructor: Manuel Santana Palacios [manuel.santana@berkeley.edu]

UGSIs: Rivka Batlan [rivkabatlan@berkeley.edu], Lauren Hom [laurenhom@berkeley.edu]

When: Mon/Wed/Fri 9:00 am – 12:00 pm

Where: Zoom Virtual Classroom:

Office Hours:

-Manuel: Wednesdays from 2 to 4 pm (sign-up via http://calendly.com/mansanp)

-Rivka: Wednesdays from 12 to 1 pm

- Lauren: Weeks 1 - 2 on Friday from 1 to 2 pm; Weeks 3-6 on Mondays from 1 to 2 pm

Notes

[1] Please notify UGSIs ahead by email that you are attending their office hours

[2] Instructors will for holding virtual office using the same Zoom link provided for lectures

Course Description

CP 101 introduces students to the systematic analysis of urban data in its institutional context. Recognizing that defining this context relies on critical thinking concerning economic, social, and environmental outcomes, the course explores how stakeholders conceptualize smart and inclusive urbanity. Accordingly, this course teaches students systematic approaches to collecting, analyzing, modeling, and interpreting data to inform research, and, ultimately, urban planning practice and policymaking. Beyond instruction in urban data science and analytics, students will be introduced to theory and critical discourses on topics such as big data, open data, and e-governance. Instructors will expect students to engage with technical and theoretical - with a particular focus placed on ethical - considerations associated with these subjects in lecture and laboratory sections. The course will introduce students to data analysis in Excel and coding in Python, using open source software, accessing open and scraped data, and other tools and techniques for urban analysis.

Prerequisites

CP 101 reserves seats for CED majors and others from around campus (e.g., Data Science majors); others can enroll with the instructor's permission. No prior statistics coursework is assumed. This class provides a foundation to pursue further undergraduate data science courses at UC-Berkeley. For Urban Studies, CP 101 satisfies one of the four additional City Planning courses for Upper Division Urban Studies Core. For the City Planning minor, CP 101 satisfies one of the four additional City Planning courses for Upper Division courses under List 1. For SED, CP 101 can count as an Upper Division outside SED major for Fall 2016 admits and later.

Course Requirements

Assignments will involve the use of data and software available online, through campus IT. You need to use your personal computer for labs, assignments, and exams. Labs sessions will be offered to introduce students to Excel, Census Data & Social Explorer, CARTO, WordPress, and Python programming using Jupyter Hub. Students are welcome to explore other programming languages and tools; however, please check with the graduate student instructor before using any alternative software.

Course Structure

Module 1: Introduction to Data Science for Planners

During this module, students will be introduced to the smart cities construct and associated topics and will explore fundamental data applications in urban planning. In Module 1, students will gain skills in working with Census and economic data, statistical testing, and static data visualization.

Module 2: Mapping the City

In the course's second module, students will learn different tools to make maps. Students will gain an understanding of the essential elements of maps, how to map with online programs and geographic information systems open tools such as CARTO, and how to construct story maps.

Module 3: Big Data and Analytics

In the course's final module, students will use the knowledge acquired in earlier modules to explore urban data science in the context of smart cities. Sessions will cover topics such as big data and smart cities, and students will gain skills in real-time and crowd-sourced data collection and use, as well as in interactive data visualization.

Course Materials

CP 101 has one required book: Singleton, Spielman, and Folch. 2018. Urban Analytics. Thousand Oaks, CA: Sage. There is no course reader; all readings that are not part of the required book will be posted to the CP 101 bCourses except for the ones from which we provide a link to. Lecture slides and other course materials will be posted on the CP 101 bCourses site. Lectures will be recorded and posted on bCourses. **PLEASE DO NOT DISTRIBUTE THESE VIDEO RECORDING WITH ANYONE**. Visit the course website, http://www.cp101.org, for a variety of additional (optional) resources related to the course.

Grading

Students will complete four group assignments and take one quiz at the end of the course. Students will be expected to read multiple articles, book chapters, and online posts before lecture sessions and attend and actively participate in class and lab. Students are asked to submit reading response questions online on our course bCourse site. Grades will be assigned as follows:

- Assignment 1: Neighborhood profile (20%)
- Assignment 2: Class project first deliverable - Research question, data, and methods (5%)
- Assignment 3: Class project second deliverable - Brief Story Map (15%)
- Assignment 4: Class project final deliverable (30%)
- Reading Response Questions (10%)
- Quiz (20%)

Class Participation and Reading Responses

Students are expected to attend every lecture and actively contribute to class discussions. We encourage you to **use your video camera** when engaging in discussions, but understand some of you may not want to so. Attendance to computer lab sessions is optional but highly encouraged. **Students are expected to respond to different sets of readings by submitting at least one question per session marked with an asterisk [*].** Questions must demonstrate a high level of awareness and react critically to concepts and analyses presented by the readings and could be used to prompt an in-class discussion. Instructors may randomly select one or more of the questions to discuss during lecture or lab sessions. Questions should be posted by 11:59 pm the day before class to the bCourses discussion thread for the day.

Assignments

In the first assignment, students will explore neighborhood change through an in-depth analysis of census data. The reports will contain compelling visualizations and text describing neighborhood change. Assignment 1 will be conducted by groups of two students, randomly selected by instructors.

Students will also work on a class project that explores at least two datasets of their choice to answer one research question. At least one dataset must fit into the definition of big data or open data. The class project can be conducted in groups of up to three students and is divided into the following three assignments:

1. A one-page description of their research question and approach (Assignment 2)
2. A brief story map describing some preliminary findings with an interactive online map(s) and other visualizations (Assignment 3).
3. a post answering the research question(s) with a fully developed narrative supported by data analyses (Assignment 4)

All deliverables must be published on WordPress (or a similar online content management system), including references in proper format and follow data visualization and ethical principles covered in the course material.

Quiz

There will be one take-home open-book quiz. Students will choose one of four questions provided and will write a short essay in response. The quiz will be held during week five and will assess students learning of the applications of urban data analytics, potential ethical dilemmas, and critical thinking.

Statement on Academic Integrity

Any test, visualization, blog post, paper, or report submitted by you and that bears your name is presumed to be your original work that has not previously been submitted for credit in another course. You may use words or ideas written by other individuals in publications, websites, or other sources, but only with proper attribution. If you are not clear about the expectations for completing an assignment or taking an exam, be sure to ask a course instructor.

You should also keep in mind that as a member of the campus community, you are expected to demonstrate integrity in all of your academic work and be evaluated on your own merits. The consequences of cheating and academic misconduct — including a formal discipline record and possible loss of future opportunities — are not worth the risks.

Statement on Accommodations for Students with Disabilities

If you have been issued a letter of accommodation from the Disabled Students Program (DSP), please see the course lead instructor as soon as possible to work out the necessary arrangements. If you need an accommodation and have not yet seen a Disability Specialist at the DSP, please do so as soon as possible. If you would need any assistance in the event of an emergency evacuation of the building, the DSP recommends that you make a plan for this in advance. (Contact the DSP access specialist at 643-6456.)

Statement on Scheduling Conflicts

Please notify course instructors by the second week of the term about any known or potential extracurricular conflicts (such as religious observances, graduate or medical school interviews, or team activities). We will try our best to help you with making accommodations but cannot promise them in all cases.

Module 1: Introduction to Urban Data

Week 1 - Wednesday, May 27

Lecture Session 1: Introduction to Urban Analytics
- Singleton, Spielman, and Folch, 2018. Chapter 1, "Questioning the city through urban analytics."
- [Optional] Kim, Annette (Jun 5, 2018). Satellite Images Can Harm the Poorest Citizens. The Atlantic.

Lecture Session 2 - Data Fundamentals for Planners
- Singleton, Spielman, and Folch, 2018. Chapter 2, "Sensing the city."
- Boyd and Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." Information, Communication & Society 15 (5)

Week 1 - Friday, May 29

*Lecture Session 3: US Census Data [Part 1]
- US Census Bureau (Feb 27, 2020). The Importance of the American Community Survey and the 2020 Census. YouTube
- Alba, Richard (Jun 11, 2015). The Myth of a White Minority. The New York Times
- Singleton, Spielman, and Folch (2018) Chapter 6 "Explaining the city," **pg. 97-113** [Optional: pg. 114-118]

Lab Session 1: Excel Basics

Week 2 - Monday, Jun 1

Lecture Session 4: US Census Data [Part 2]
- No readings for this lecture

*Lecture Session 5: Statistics for the American Community Survey
- Jurjevich et al., 2018. Navigating Statistical Uncertainty: How Urban and Regional Planners Understand and Work with the American Community Survey (ACS) Data for Guiding Policy. Journal of the American Planning Association, 84(2), 112-126
- Cochran, Abby. 2020. Stats for CP 101 (Video available on bCourses)
- [Optional] Wheelan (2013) Chapters 2, 3, & 4 "Descriptive Statistics," "Descriptive Deception," "The Central Limit Theorem."

Lab Session 2: WordPress + Census Data [Part 1]

Week 2 - Wednesday, Jun 3

*Lecture Session 6: Static Data Visualization
- Tufte, 1983. The Visual Display of Quantitative Information. Graphics Press. Chapter 2, "Graphical Integrity," pg. 53-77
- Few, 2012. Show Me the Numbers: Designing Tables and Graphs to Enlighten. 2nd ed. USA: Analytics Press. Appendix A "Table and Graph Design at a Glance" pg. 309-310

- [Optional] Few (2012) Chapter 3 pg. 39-60 "Differing Roles of Tables and Graphs," Chapter 4 pp. 53-60 "Fundamental Variations of Tables" Chapter 5, pg. 67-79 "Attributes of Pre-attentive Processing & "Applying Visual Attributes to Design," Chapter 6, pg. 101-135 "Graph Design Solutions," Chapter 11, pg. 257-270 "Displaying Many Variables at Once," Chapter 13, pg. 295-306 "Telling Compelling Stories with Numbers."

Lab Session 3: Census Data [Part 2]

Week 2 - Friday, Jun 5

Lecture Session 7: Intro to the Longitudinal Employment-Household Dynamics Data
- Abowd, Haltiwanger, & Lane, 2004. Integrated longitudinal employer-employee data for the United States. American Economic Review, 94(2), 224-229

Lab Session 4: Margins of Error, Social Explorer, and LEHD data

>>> Assignment 1 [Neighborhood Profile] due Sunday, June 7 <<<

Module 2: Mapping the City

Week 3 - Monday, Jun 8

*Lecture Session 8: Neighborhood Indicators: The Urban Displacement Project
- Singleton, Spielman, & Folch, 2018. Chapter 5, "Differences Within Cities"
- Chapple & Zuk, 2016. "Forewarned: The Use of Neighborhood Warning Systems for Gentrification and Displacement" Cityscape: A Journal of Policy Development and Research, 18(3), 109-130
- Urban Displacement Project [Skim]

Lecture Session 9: Spatial Data & GIS Fundamentals
- Singleton, Spielman, and Folch, 2018. Chapter 4, "Visualizing the city."
- [Optional] Monmonier, 1996. How to Lie with Maps. University of Chicago Press. Chapters 2, "Elements of the Map."

Lab Session 5: Mapping with CARTO [Part 1]

Week 3 - Wednesday, Jun 10

*Lecture Session 10 – Introduction to Story Mapping
- Tufte, 1997. "Visual Explanations: Images and Quantities, Evidence and Narrative." Chapter 2 'Visual and Statistical Thinking: Display of Evidence for Making Decisions' only pg. 27-37
- Review the following examples:
    o Mapping Segregation in DC
    o Creating a neighborhood change zoning plan for Spruce Hill
    o Katrina +10: A Decade of Change in New Orleans
    o Displacement in the Bay Area

- [Optional] Monmonier, 1996 How to Lie with Maps. University of Chicago Press. Chapters 3, 4, and 10.

Lab Session 6: Mapping with CARTO [Part 2] + WordPress [Part 2]

Week 3 - Friday, Jun 12

*Lecture Session 11: Power, Place, and Mapping
- Parker, Brenda. "Constructing Community through Maps? Power and Praxis in Community Mapping." Professional Geographer, 58:4, (2006): 470-484
- Norwood, Carla, and Gabriel Cumming. "Making maps that matter: Situating GIS within community conversations about changing landscapes." Cartographica: The International Journal for Geographic Information and Geovisualization 47.1 (2012): 2-17
- Harley, J. Brian. "Maps, knowledge, and power" (Chapter 8). In Henderson, George and Waterstone, Marvin. Geographic thought: a praxis perspective, 1988. 129-148

Lab Session 7: Python: Intro to Pandas

>>> Assignment 2 [Class Project Research Question, Data, and Methods] due Friday, Jun 12 <<<

Module 3: Big Data and Analytics

Week 4 - Monday, Jun 15

*Lecture Session 12: Introduction to Big Data
- Foster et al., 2017. "Introduction." Pp. 1-19 in Big Data and Social Science: A Practical Guide to Methods and Tools. Boca Raton, FL: Taylor & Francis Group.
- Schweitzer, 2014. "Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter." Journal of the American Planning Association 80 (3): 218–38.
- [Optional] B. Strasser and P. Edwards, "Big Data is the Answer… But What is the Question?" Osiris 32, 2017: pp. 328-345

*Lecture Session 13: Big Data: Ethical Considerations in Urban Studies/Planning
- Zook et al., 2017. "Ten simple rules for responsible big data research." PLoS Comput Biol 13(3)
- Crawford, 2013. "The hidden biases in big data." Harvard Business Review 1
- [Optional] Gitelman and Jackson. 2013. Introduction. Raw data is an oxymoron. MIT Press
- [Optional] Crawford. "The Trouble with Bias," NIPS conference keynote, December 2017 (minutes 14:00 - 38:00)

Lab Session 8: Python: Web Scraping and Data Cleaning

Week 4 - Wednesday, Jun 17

*Lecture Session 14: Volunteered Geographic Information [VGI]
- Jiang, Bin, and Jean-Claude Thill. 2015. "Volunteered Geographic Information: Towards the Establishment of a New Paradigm." Computers, Environment and Urban Systems, Special Issue on Volunteered Geographic Information, 53 (September): 1–3.

- Boeing, Geoff, and Paul Waddell. 2016. "New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings." Journal of Planning Education and Research.

Lab Session 9: Python: Mapping and Geopandas [Part 1]

Week 4 - Friday, Jun 19

Lecture Session 15: Street Networks, Open Street Maps, and Accessibility
- Singleton, Spielman, and Folch (2018) Chapter 8, "Cities as Networks and Flows"
- Jones. "Measuring Pedestrian Accessibility." Medium. Jun 27

Lab Session 10: Python: Mapping and Geopandas [Part 2]

>>> Assignment 3 [Story Map] due Sunday, June 21 <<<

Week 5 - Monday, Jun 22

*Lecture Session 16: Complex Urban Modeling: Machine Learning
- Foster et al., 2017. "Machine Learning." Pp. 147-186 in Big Data and Social Science: A Practical Guide to Methods and Tools. Boca Raton, FL: Taylor & Francis Group.
- Domingos, 2012. "A few useful things to know about machine learning." Communications of the ACM, 55(10), 78-87

Lecture Session 17: Interactive Visualizations
- Hemmersam et al., 2015. "Exploring Urban Data Visualization and Public Participation in Planning." Journal of Urban Technology 22 (4): 45–64
- Anderson, 2016. "12 Complex Concepts Made Easier Through Great Data Visualization — ReadThink (by HubSpot)." Medium. Jun 27
- Explore additional interactive visualizations here:
    o http://polygraph.cool/history/
    o http://goodcitylife.org/chattymaps/index.html
    o http://hubcab.org/#13.00/40.7219/-73.9484
    o http://218consultants.com/interactive-suitability-map/
    o https://ourworldindata.org/a-history-of-global-living-conditions-in-5-charts/ http://www.urban.org/features/vision-equitable-dc
    o http://www.urbandisplacement.org
- [Optional] Foster et al. 2017. "Working with Web Data and APIs." Pp. 23-70 and "Information Visualization." Pp. 243-263.

Lab Session 11: Python: Mapping and Geopandas [Part 3]

Week 5 - Wednesday, Jun 24

*Lecture Session 18: Smart Cities [Part 1]
- Pierce and Shoup, 2013. SFpark: Pricing Parking by Demand
- Explore Side|Walk|Labs

- [Plans to Develop High-Tech 'Smart City' in Toronto Met with Resistance](#). NPR, Feb 16, 2020.

Lab Session 12: Working with Big Data

Week 5 - Friday, Jun 26, 2020

*Lecture Session 19: Smart Cities [Part 2]
- Batty, 2016. "How Disruptive Is the Smart Cities Movement?" Environment and Planning B: Planning and Design 43 (3): 441–43.
- T. Misra, "[The New Digital Sanctuaries,](#)" Citylab, Nov 14, 2017.

Lab Session 13: Open Help Session

>>> Take-home Open-book Quiz due Saturday, June 27 <<<

Week 6 - Monday, Jun 29, 2020

*Lecture Session 20: Smart Institutions & e-Governance
- Noveck, Beth Simone. 2015. Smart Citizens, Smarter State: The Technologies of Expertise and the Future of Governing. Harvard University Press.; Chapter 1 & Conclusion, "From Open Government to Smarter Governance," pg. 1 - 43; "Conclusion: The Daedalus Project," pg. 267 – 275
- V. Eubanks, "Want to Predict the Future of Surveillance? Ask Poor Communities," The American Prospect, Jan 15, 2014.

*Lecture Session 21: Open Data and Access
- Johnson, Jeffrey Alan. 2014. "From Open Data to Information Justice." Ethics and Information Technology 16 (4): 263–74
- Explore [The Digital Matatus Project](#)

Lab Session 14: Open Help Session

>>> Assignment 4 [Final Project] due Tuesday, Jun 30 <<<

Week 6 - Wednesday, Jul 1, 2020

Lecture Session 22: Deploying Urban Data Analytics in Research and the Future of the Field
- Singleton, Spielman, and Folch, 2018. Chapter 9, "The Future of Urban Data Analytics."

Closing remarks by Karen Chapple, Professor of City & Regional Planning and Data Science