Laboratory in Justice Data Science 4 Credits - Spring 2022

Prof. Niall Bolger nb2229@columbia.edu

Prof. Geraldine Downey, gd20@columbia.edu

TA: Jacob Edwards, je2527@columbia.edu

Syllabus is subject to change: Revisions will be posted on Courseworks

Course description. The course will provide the rigorous data science training and core content knowledge students need to use data science to effect policy changes that promote a more just society. The course will leverage the academic expertise of psychologists, lawyers and data scientists, the perspectives and experiences of community members and students affiliated with the Center for Justice, and policymakers from government agencies and community organizations. The focus will be on collaborating with community and government organizations to propose data-informed solutions that center on those most impacted by failures of the justice system.

Students will learn how to promote a more just society through combining data, disciplinary knowledge, and fine-grained, and on-the-ground experience. They will learn how to approach policy relevant data with an explicit justice mindset such that they consider the implications of specific policies for achieving a more just, racially equitable outcome.

Learning goals:

1. Using specific examples from criminal, gender, racial and environmental justice, students will learn through "doing" how to identify relevant administrative and other datasets; how to clean, integrate, analyze and learn from data; how to make and communicate the inferences and/or predictions; how to integrate close-to-the-problem expertise into problem solving; and how to approach data and problem solving in ways that are consistent with justice values, where assumptions about data and statistical models are transparent.
2. Students will learn how relevant data are identified and results are effectively communicated; how data can inform policy changes and bring about meaningful and just societal change; how the impact of policies (and other changes) can be assessed; and how quantitative and qualitative approaches can be integrated.
3. Students will learn how data may be misused in ways that perpetuate racial inequalities and biases, how policies that rely on gut-instinct and opinion can perpetuate injustice and structural racism (e.g., the policies that generated mass incarceration). Thus, students will learn to adopt a critical approach to data and policy formulation and to be vigilant for unintended consequences of well-intentioned efforts when data are used without an understanding of context and history.

Integration into curriculum

Students are interested in connecting academics with data-informed action. The work of the Center for Justice with communities and city and government agencies makes clear the value of educating students in how to use data to inform the transformation of law, policy and institutional practices. Psychology is becoming more focused on connecting to other disciplines, addressing questions of societal relevance, and preparing students to grapple with issues of social justice. The need for psychology students to be well-versed in new approaches to data (that go beyond traditional lab-based experiment methods) is something in which the Psychology department is building expertise as we prepare students to be future leaders in using data to understand and promote positive change.

This course is viewed as a second course in the Frontiers of Justice sequence (although it is also a stand-alone course) and an essential step in broadening our curricular offering to prepare undergraduates for the burgeoning interest in connecting psychology and neuroscience with public policy and law, and providing undergraduates the data science training they need to go on to graduate study, careers in public policy, etc. Recognizing the growing interdisciplinarity of the field and the need to provide undergraduates with more applied, integrative training, last year, the Psychology department created a new interdisciplinary requirement for majors. In addition to fulfilling core area and seminar requirements for neuroscience and psychology students, the Frontiers of Justice sequence will be an integral part of this new applied/integrative requirement.

Many students are now combining majors in neuroscience and psychology with majors and concentrations in political science, African American and Diaspora Studies, public health, and human rights. There is an expanding need for rigorous, interdisciplinary training in neuroscience/psychology, data science, and public policy. This Frontiers of Justice sequence represents an important step in expanding our undergraduate curriculum to meet the challenges of our time.

Prerequisites: *Frontiers of Justice: An introduction* and *an intro to statistics course* preferred. Permission of instructor is required.

Role in the Curriculum: This is a lab course with lectures, open to undergraduate students, post-baccalaureate students, master's students and graduate students.

Course Expectations and Grading:

1. Class preparation and participation: The readings and data assignments will be selected from the list below and are designed to expand your knowledge of data science and justice and

to hone your critical thinking and hands-on data skills. Strong preparation and participation will enable us to have high-level, thought-provoking discussion, both in the lecture and lab components of the class. To ensure that everyone is accountable for thoroughly engaging with the material during class discussions, your active participation in these discussions and completion of lab assignments will contribute to your final grade. For some people participating regularly in class discussions can be difficult. Those students who might be concerned about their ability to contribute to class discussions should see the course instructors or TA.

After each Wednesday lecture, students will have the remainder of the week to prepare the related lab. The lab assignments are due in the lab on Monday. There will be 8 assignments worth 60 points in total. Students will receive 5 points for on-time assignments and 4 points for late assignments and no points for incomplete assignments.

At the end of the semester, students will summarize the analyses of their own data in a 10-15 minute class presentation and written paper. The paper should (loosely) follow an APA format with a brief introduction to the topic, a detailed methods section, a thorough results section, and a concise discussion.

Presentations will be on April 27th and May 2nd and are worth 20 points (all students are expected to attend all presentation days).

Final papers are due via upload to Canvas on Friday, May 6th, at 5 pm and are worth 20 points.

Course materials: Readings will comprise empirical papers, review papers, and approaches identifying relevant administrative and other data sets; how to clean, integrate, analyze and learn from data; how to make and communicate the impacts of inferences and/or predictions; how to integrate close-to-the-problem expertise into problem solving; and how to approach data and problem solving in ways that are consistent with justice values, and where assumptions about the meaning and modelling of data are transparent. Students will also be introduced to open access Python and R. PDFs of readings and relevant links will be posted on Courseworks.

Schedule: The schedule below details themes and key topics for each week. Required readings/assignments will be selected from these lists; other readings/assignments will be posted as supplemental but not required. ***This list remains subject to revision and guest speakers' availability.***

Course Structure

Students will attend a lecture (10:10 -12 Wednesday and a lab (10:10 -12 Monday). The lecture will be led by the instructors and will include guest speakers some weeks. In the Lab Section the students will gain hands-on experience with data science approaches to justice-related questions.

LAB Mondays	LECTURE Wednesdays
1/17 NO CLASS	**1/19** Evidence in Support of Justice
1/24 Software choices: R, Python	**1/26** Case Studies in Data Use and Misuse
1/31 Online Data Sources	**2/2** Multilevel Thinking in Justice Policy
2/7 Data Wrangling	**2/9** Thinking About Causal Processes
2/14 Effective Visualization	**2/16** Probability Thinking: Frequentist & Bayesian
2/21 NO CLASS: President's Day	**2/23** Social Interventions and Experiments
2/28 Using Probability Statements	**3/2** Natural Experiments and How to Use Them
3/7 Nonexperimental Data	**3/9** How to Study Change as it Relates to Justice
3/14 SPRING BREAK	**3/16** SPRING BREAK
3/21 Analyzing Change Over Time	**3/23** Narrative Evidence
3/28 Integrating Data Analysis and Narrative	**3/30** Socio-Historical Changes and the Life Course
4/4 Predictive Analytics	**4/6** Social Contagion in Networks
4/11 Social Network Analysis	**4/14** Intersectional Processes
4/18 Community-Engaged Research	**4/20** Using Data to Influence People, Practice, and Policy
4/25 Meta-Analysis	**4/27** Policy Presentation
5/2 Policy Presentation	

Readings will be selected from the following list (subject to change)

Dodge, Kenneth (2008), Framing public policy and prevention of chronic
    violence in American youth. *American Psychologist. 63,* 573-590*.*

DU Patton, S Mathiyazhagan, AY Landau. Meet Them Where They Are: Social Work Informed Considerations for Youth Inclusion in AI Violence Prevention Systems. Algorithmic Rights and Protections for Children.

Gone, Joseph (2021). The (Post)Colonial Predicament in Community Mental Health Services for American Indians: Explorations in Alter-Native Psy-Ence. *American Psychologist.*

Gone, Joseph (2019). "The thing happened as he wished." Recovering an American Indian Cuulrural psychology*, Journal of Community Psychology.*

ML Hatzenbuehler, S McKetta, R Kim, S Leung, SJ Prins, ST Russell. Evaluating litigation as a structural strategy for addressing bias-based bullying among youth. JAMA pediatrics.

MR Lattanner, J Ford, N Bo, W Tu, JE Pachankis… (2021). A Contextual Approach to the Psychological Study of Identity Concealment: Examining Direct, Interactive, and Indirect Effects of Structural Stigma on Concealment Motivation Across Proximal and Distal Geographic Levels. Psychological Science.

MA Price, JR Weisz, S McKetta, NL Hollinsaid, MR Lattanner, AE Reid, …Meta-analysis: Are Psychotherapies Less Effective for Black Youth in Communities With Higher Levels of Anti-Black Racism? Journal of the American Academy of Child & Adolescent Psychiatry.

Mark L. Hatzenbuehler (2020). Multilevel Approaches to Understanding LGBTQ Health Disparities (2020). LGBTQ Health Research: Theory, Methods, Practice,

D Rao, A Elshafei, M Nguyen, ML Hatzenbuehler… A systematic review of multi-level stigma interventions: state of the science and future directions
- BMC medicine, 2019

David Gillborn, Paul Warmington & Sean Demack (2018) QuantCrit: education, policy, 'Big Data' and principles for a critical race theory of statistics, Race Ethnicity and Education, 21:2, 158-179, DOI: 10.1080/13613324.2017.1377417

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). Boston: O'Reilly.

Grus, Joel. (2018) *Data science from scratch: First principles with Python*. Sebastopol, CA:

O'Reilly.

Zumel, Nina, & Mount, John. (2020) *Practical data science with R* (2nd ed.) Shelter Island, NY: Manning.

Hacking, I. (2001). *An introduction to probability and inductive logic*. New York: Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (1st ed.). New York: Basic Books.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Carlsson, C., & Sarnecki, J. (2015). *An introduction to life-course criminology*. Thousand Oaks, CA: Sage.

Prell, C. (2012). *Social network analysis: History, theory & methodology*. Los Angeles: SAGE.

Luke, D. A. (2015). *A user's guide to network analysis in R*. New York: Springer.

Taris, T. W. (2000). *A primer in longitudinal data analysis*. London: Sage.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-Analysis with R: A hands-on guide* (1st ed.). New York: Chapman and Hall/CRC.

Siegel, E. (2016). *Predictive analytics: The power to predict who will click, buy, lie, or die* (Revised ed.). New York: Wiley.

Legewie, J., & Fagan, J. (2019). Aggressive policing and the educational performance of minority youth. American Sociological Review, 84, 220-247. doi: 10.1177/0003122419826020

 Desmond, M., & Valdez, N. (2013). Unpolicing the urban poor: Consequences of third-party policing for inner-city women. American Sociological Review, 78, 117-141. doi: 10.1177/0003122412470829

Patrick Sharkey "The gaps between White and Black America, in charts." Link to article here: https://www.nytimes.com/interactive/2020/06/19/opinion/politics/opportunity -gaps-race-inequality.html

Work by Prof. Laura Kurgan and her students on geospatial analysis of justice

https://urbanomnibus.net/series/location-of-justice/

https://c4sr.columbia.edu/news/human-face-big-data-features-million-dollar-blocks

https://c4sr.columbia.edu/news/how-mass-incarceration-creates-'million-dollar-blocks'-poor-neighborhoods-washington-post

https://www.moma.org/interactives/exhibitions/2013/designandviolence/million-dollar-blocks-by-the-spatial-information-design-lab/

https://www.moma.org/interactives/exhibitions/2013/designandviolence/design-and-violence-the-book/

https://www.propublica.org/article/when-the-designer-shows-up-in-the-design (Picking a good question)

Work by Prof. Mark Hansen and students on integrating narrative and data

https://www.themarshallproject.org/2016/02/03/policing-the-future

https://www.themarshallproject.org/2016/02/21/what-you-need-to-know-about-predictive-policing

Data sources from the Marshall project

https://www.themarshallproject.org/records/2093-data-access

https://www.themarshallproject.org/records/2491-big-data

https://www.themarshallproject.org/2016/08/18/crime-in-context

https://www.themarshallproject.org/2020/03/11/how-we-pulled-off-a-groundbreaking-political-survey-behind-bars

Data science isn't "apolitical"

Green, B. (2020). Data science as political action: grounding data science in a politics of justice. Available at SSRN 3658431. https://arxiv.org/abs/1811.03435

Nost, E., & Goldstein, J. E. (2021). A political ecology of data. Environment and Planning E: Nature and Space, 25148486211043503. https://journals.sagepub.com/doi/abs/10.1177/25148486211043503

Wright, R. F., Chavis, K., & Parks, G. S. (2018). The Jury Sunshine Project: Jury selection data as a political issue. U. Ill. L. Rev., 1407. https://heinonline.org/HOL/LandingPage?handle=hein.journals/unilllr2018&div=42&id=&page=

History of statistics as a tool for perpetuating injustice

https://nautil.us/issue/92/frontiers/how-eugenics-shaped-statistics

Prediction Gone Awry / "Automatization" of Injustice

https://www.theverge.com/22444020/chicago-pd-predictive-policing-heat-list

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://www.science.org/doi/abs/10.1126/science.aax2342

https://www.wired.com/story/excerpt-from-automating-inequality/ (from https://us.macmillan.com/books/9781250074317/automatinginequality)

Physiognomy & 'Gaydar'

https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html

https://www.glaad.org/blog/glaad-and-hrc-call-stanford-university-responsible-media-debunk-dangerous-flawed-report

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of personality and social psychology, 114(2), 246.

Course Policies:

Fostering an Inclusive Classroom: Our aim is to foster a learning environment that supports a diversity of perspectives and experiences and honors your identities. Please reach out to us with any concerns or suggestions you may have to better address your learning needs and to improve the effectiveness of this course. We look forward to working together to create a classroom community built on mutual respect and inclusivity.

 Students with special needs who may require classroom/test accommodations should make an appointment with me before or during the first week of class. You should also contact the Office of Disability Services (ODS) in Lerner Hall before the start of the course to register for these accommodations. The procedures for registering with ODS can be found at https://health.columbia.edu/content/disability-services or by calling (212) 854-2388.

Promoting Wellness: Many of us have periods in which our mental health and well-being suffer. Please take care of yourselves – and of each other. Please prioritize your mental health and wellbeing and know that there are many resources available to you both within our classroom community and throughout the university:

https://health.columbia.edu/content/counseling-and-psychological-services

http://blogs.cuit.columbia.edu/nightline/

We are in this together. Please reach out for help if you need it, and, if you see others who are struggling, please make sure they know how to find the support they need.

Ensuring Academic Integrity: As members of this academic community, we are responsible for maintaining the highest level of personal and academic integrity, which includes presenting only our own work on assignments and exams. You can find detailed definitions and examples in Columbia University's Guide to Academic Integrity (http://www.college.columbia.edu/academics/academicintegrity). Any questions of academic integrity will be automatically referred to Columbia's office of Student Conduct and Community Standards. The semester progresses very quickly, and there is a lot of material to learn. If you find yourself in a situation – e.g., starting an assignment too late – in which it seems like the best option may be to violate your academic integrity, please see me. Together, we can work out a solution. Plagiarism—whether intentional or inadvertent—is a serious violation of academic integrity. If you have any questions about what constitutes plagiarism and/or how to properly cite sources, please come to me. I am more than happy to help.

Syllabus is subject to change: Revisions will be posted on Courseworks