

# BIG DATA AND PUBLIC POLICY

PPHA 3059 / SSA 4059

University of Chicago

Winter, 2021 (updated 1.27.21)

**Instructors:** Chris Berry  
crberry@uchicago.edu

**Office Hours:**  
Fridays 2 to 3pm

Nicole Marwell  
nmarwell@uchicago.edu

**Office Hours:**  
Wednesdays 2 to 3pm

## **COURSE SUMMARY**

This course examines the conceptual underpinnings of data science and social science approaches to policy analysis. We discuss epistemologies of quantification, data production and the phenomenon of “datafication,” predictive versus causal analytic paradigms, algorithmic fairness, and issues of data ethics, regulation and governance. The course is open to graduate and undergraduate students who have taken at least one course in quantitative methods.

## **REMOTE LEARNING TEACHING METHODS AND EXPECTATIONS**

The University of Chicago is offering most Winter 2021 classes remotely, due to the COVID-19 pandemic. This syllabus has been developed in response to the online format, with best practices for online instruction in mind.

It is very important that you activate your University of Chicago Zoom account before the first day of class. You will only be able to access our course meetings through your university Zoom account. You cannot use a Zoom account that you have set up outside the university. You should also make sure to claim your University of Chicago G-suite account; that will allow you to use google tools (such as docs, forms, slides, etc.) from within the secure university system. If you have any questions regarding how to set up either of these accounts, please contact the University’s technical support.

This course will be taught using both synchronous and asynchronous activity. We will have a total of 2 hours and 50 minutes of contact time each week; the ratio of synchronous and asynchronous activity may vary over the weeks. The weekly schedule below indicates what you should prepare prior to each week’s class. Course materials and instructions are available on the course Canvas site; you are expected to have access to and regularly check Canvas.

We expect that you will be actively engaged and on camera during our class Zoom meetings. If your situation requires your camera to be regularly turned off, please contact us before the start of Week 2 of the quarter to discuss.

## **REQUIRED READINGS**

There is one required book for the course, Rob Kitchin's *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. You should locate it from your favorite online bookseller. All other required readings are listed in the Course Schedule and posted on the Canvas site, in the Module for the indicated week.

## **ASSIGNMENTS AND EVALUATION**

1. *Pre-recorded video presentation. 25% of final grade. Sign up to make one presentation in Weeks 3-8. Due by Monday of your selected week at noon; further instructions to follow.*
  - a. 5 minute limit. **DO NOT GO LONGER THAN FIVE MINUTES.** Apply ideas from the readings for one week to a substantive topic, e.g., high school dropout, bail, mortgage approval, child protection, etc. You may choose to present on the same topic as your final paper, or on a different topic. We will provide questions to guide each week's presentation; use these as a jumping-off point. Choose the questions that are most interesting and relevant to your topic. Do not try to cover all of the questions.
  
2. *Class participation. 25% of final grade.*
  - a. Come to class having read all the assigned readings, and watched the pre-recorded lectures and student presentations.
  - b. Actively participate in full-class discussions and breakout rooms.
  
3. *Final paper. 50% of final grade.*

Papers should apply concepts from the course to a substantive topic of interest (e.g., high school dropout, bail, mortgage approval, child protection, etc.). We are open to a range of paper topics on any of the three main areas covered in the course: data production (weeks 3-4), epistemologies (weeks 5-6), and policy implications (weeks 7-8). You should choose **ONE** of these areas as the focus of your paper. Papers may: undertake original data analysis, critique prior work by others, reflect on your own work to date, offer a plan for future analysis or data collection, etc. These are only examples to give you a flavor of possible options. We are available to discuss your ideas prior to the due date of the paper proposal, and will require each of you to meet with one of us after we have read your proposal.

- a. One-paragraph summary of final paper topic: due by 9:00 am of Week 2 (January 20), via Canvas.
- b. Final paper proposal. Due NO LATER than March 3 (but you can turn it in earlier if you like, so as to be able to start working on your paper sooner), via Canvas.
  - i. Provide a preliminary list of readings, data sources, etc. that you will draw on for writing your paper.
  - ii. Your paper should focus on ONE of the course's three main substantive areas: data production (weeks 3-4), epistemologies (weeks 5-6), and policy implications (weeks 7-8).
  - iii. Provide a preliminary description of the key ethical issue(s) at stake in your topic.
- c. Final paper feedback meeting: Meetings to be held on Friday, March 5. Sign up for a meeting time on our google spreadsheet. If you submit your paper proposal earlier than March 3 and would like to meet earlier than March 5, sign up for regular office hours with one of us.
- d. Final paper: 10 pages, double-spaced, due March 19, via Canvas.

## CLASS SCHEDULE

### PART I: SETTING THE STAGE

#### *Week 1 – January 13: Introduction*

##### Required Readings:

1. Mayer-Schoenberger, Viktor, and Kenneth Neil Cukier. 2013. "The Rise of Big Data." *Foreign Affairs*.
2. Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Chapter 7.
3. Kingdon, John W. 2003. *Agendas, Alternatives, And Public Policies*. Chapters 4 & 6.

#### *Week 2 – January 20: Developing Ethical Competencies for Data Science*

##### Required Readings:

1. O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Introduction.
2. Salganik, Matthew. *Bit by Bit*, Chapter 6.
3. Hasselbalch, Gry. 2019. "Making Sense of Data Ethics. The Powers behind the Data Ethics Debate in European Policymaking"
4. Brayne, Sarah. 2017. "Big Data Surveillance: The Case of Policing." *American Sociological Review* 82: 977-1008.

##### Recommended Supplemental Readings:

1. Bueno, Ethan de Mesquita. 2019. *Political Economy for Public Policy*, Chapter 1.
2. Kahn, Falaah Arif & Julia Stoyanovich. "Mirror, Mirror (Data Responsibly, #1)." Comic book.
3. Kitchin, Chapter 10

### PART II: DATA PRODUCTION

#### *Week 3 – January 27: How to Think About Data*

##### Required Readings

1. Kitchin. Chapters 1-4
2. Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *The American Political Science Review* 95: 529-546.

*Week 4 – February 3: Does Data Tell Us What We Think it Does?*

Required Readings

1. Kitchin. Chapters 5 and 8
2. Pink, Sarah, Minna Ruckenstein, Robert Willim and Melisa Duque. 2018. "Broken Data: Conceptualising Data in an Emerging World." *Big Data & Society* January-June: 1-13.
3. Richardson, Rashida, Jason M. Schultz and Kate Crawford. 2021. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York University Law Review* 94: 15-55.

PART III: EPISTEMOLOGIES

*Week 5 – February 10: Two Cultures*

1. Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*
2. Meyer, Bruce D. "Natural and Quasi Experiments in Economics." 1995. *Journal of Business Economics and Statistics*
3. Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review Papers and Proceedings*
4. Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures." *Statistical Science*

*Week 6 – February 17: Using the Right Tool for the Job*

1. Deaton, Angus. 2020. "Randomization in the Tropics Revisited: A Theme and Eleven Variations." *National Bureau of Economic Research*.
2. Titiunik, Rocío. 2015. "Can Big Data Solve the Fundamental Problem of Causal Inference?" *PS: Political Science & Politics*.
3. Hofman, Jake M., Amit Sharma, and Duncan J. Watts. 2017. "Prediction and Explanation in Social Systems." *Science*
4. Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science*

## PART IV: THINKING ABOUT POLICY IMPLICATIONS

### *Week 7 – February 24: External Validity: Understanding Implications of Evidence for Policy*

#### Required Readings

1. Czibor, Eszeter, et al. 2019. "The Dozen Things Experimental Economists Should Do (More of)." *Southern Economic Journal*.
2. Dehija, Rajeev. 2015. "Experimental and Non-Experimental Methods in Development Economics: A Porous Dialectic." *Journal of Globalization and Development*.
3. Lucas, Jeffrey. 2003. "Theory-Testing, Generalization, and the Problem of External Validity." *Sociological Theory*.
4. Muller, Sean. 2015. "Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations." *World Bank Economic Review*.
5. Pritchett, Lant, and Justin Sandefur. 2014. "Context Matters for Size: Why External Validity Claims and Development Practice do not Mix." *Journal of Globalization and Development*.

### *Week 8 – March 3: Fairness*

#### Required Readings

1. Mitchell, Shira, et al. 2021. "Prediction-Based Decisions and Fairness: A Catalogue of Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application*.
2. Paulus, Jessica, and David Kent. 2020. "Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities." *Nature Digital Medicine*.
3. Kravitz, Richard, et al. 2004. "Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages." *The Milbank Quarterly*.

#### Also Required: Read ONE of the following:

1. Less technical option: Barocas, Solon, and Andrew Selbst. 2016. "Big Data's Disparate Impact." *California Law Review*. (Read part I and skim the rest.)
2. More technical option: Kleinberg, Jon, et al. 2017. "Inherent Trade-Offs in the Fair Determination of Risk Scores." *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*

## PART V: BRINGING IT ALL TOGETHER

### *Week 9 – March 10: Bail Decisions and Recidivism Prediction*

1. Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 2016. "Machine Bias." *ProPublica*. May 23, 2016.
2. Rudin, Cynthia, Caroline Wang, and Beau Coker. 2020. "The Age of Secrecy and Unfairness in Recidivism Prediction." *Harvard Data Science Review*.
3. Kleinberg, Jon, et al. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*.
4. Alexandra Chouldechova. 2018. "The accuracy, fairness, and limits of predicting recidivism." *Science Advances*
5. Feller, Avi, et al. 2016. "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear." *Washington Post*. 10/17/2016.
6. Jackson, David, et al. 2020. "Bail reform analysis by Cook County chief judge based on flawed data, undercounts new murder charges." *Chicago Tribune*. 2/13/2020.

## FINAL PAPERS DUE ON MARCH 19

### THE FINE PRINT

1. Zoom Policy: It is easy to get distracted while on Zoom. Please do your best to stay engaged in class and avoid conducting other activities (email, text, etc.) during our synchronous time.
2. Participation Policy: As this is a class where we do a great deal of learning from our discussions with each other, your participation is required. Attending our class Zoom meetings is important, but it is not sufficient for full participation credit. For full credit, we expect that you will attend regularly *and* demonstrate that you are fully prepared, by asking questions, making comments, and otherwise working to help make this an engaging, interesting, and productive class. We will also expect you to be respectful and encouraging of your colleagues. We will do our best to live up to these standards as well.
3. Attendance Policy: Sometimes life events (illness, travel, family problems, etc.) cause students to miss class. If you miss class once, we understand. If you miss class twice, we understand, but it will be reflected in your participation grade. If you miss class three times your final grade for the course will not be higher than a B-minus. If you miss class

four times without a medical excuse from a doctor, you will not be eligible to pass the class and should consider dropping it.

4. Late Assignment Policy: You are responsible for finishing your work on time. All assignments are due via Canvas at the due dates and times indicated—afterwards is considered late. Late papers are not eligible for full credit. Specifically, they will be marked down by a third of a grade the first day (e.g., an A paper would now be an A-minus), another third of a grade down the second day (e.g., an A paper would now be a B-plus), etc.
5. Formatting of assignments: For all assignments, we expect you to write clearly, proofread carefully, and adhere to page limits. Please format your papers in 12-point font, with 1-inch margins, double-spaced, and reference all sources in either APA or ASA style.
6. Getting help: If you are worried about meeting a class requirement or are having trouble with the material, please contact us *immediately* by email. We want to help you succeed in this course, but that becomes difficult at the last minute. “Incomplete” grades will be given only under extraordinary circumstances (as in, almost never).
7. Accommodations for Students with Different Abilities: The University of Chicago is committed to ensuring equitable access to our academic programs and services. Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please meet with one of us to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations. To contact SDS: website: [disabilities.uchicago.edu](http://disabilities.uchicago.edu) phone: (773) 702-6000 email: [disabilities@uchicago.edu](mailto:disabilities@uchicago.edu)
8. Academic Integrity: All University of Chicago students are expected to uphold the highest standards of academic integrity and honesty. Among other things, this means that students shall not represent another’s work as their own, use un-allowed materials during exams, or otherwise gain unfair academic advantage. The University’s policies regarding academic integrity and dishonesty are described [here](#). In summary, “It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another’s statements or ideas as one’s own work. To do so is plagiarism or cheating, offenses punishable under the University’s disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously.”

## **ZOOM RECORDING AND DELETION POLICIES FOR ACADEMIC YEAR 2020-21**

The Recording and Deletion Policies for the current academic year can be found in the Student Manual under Petitions, Audio, and Video Recording on Campus.

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be automatically deleted 90 days after the completion of the recording.

### **COVID-19 RELATED INFORMATION**

All students on campus are required to adhere to the guidelines in the UChicago Health Pact in order to promote a safe environment on campus:

- Secure face coverings must be worn appropriately at all times at all times while in University buildings.
- Maintain a distance of 6 feet from others.
- Do not come to any campus facility if you feel unwell or are experiencing COVID-19 related symptoms.

The complete text of the UChicago Health Pact along with additional information about COVID-19 protocols can be found [here](#).

If you were potentially exposed to COVID-19 or are experiencing symptoms of COVID-19, contact UChicago Student Wellness immediately to be tested. If your COVID-19 test results come back positive, reach out immediately to [C19HealthReport@uchicago.edu](mailto:C19HealthReport@uchicago.edu).

Any concerns over inappropriate PPE usage (e.g., people on campus not wearing masks), physical distancing, cleaning/disinfection, or other COVID-19 related public health concerns should be directed toUCAIR. If there is an emergency, call 773-702-8181 or dial 123 on any campus phone.