

## PAM 2070: Big Data for Big Policy Problems

Summer 2020, July 13-30

Professors Maria Fitzpatrick and Matt Hall

This class is all online and students are spread throughout the world in different time zones. We have structured the course to allow you to take it wherever you are and receive full credit. We will have live interactive sessions to go over coding and answers to the problem sets. We have scheduled those for the times listed below, which are meant to be accessible across time zones. They will be recorded and posted online so that all students can view them, even if they are not able to attend live. We are also available to help you through the Canvas discussion board, email, and by appointment.

### Assignment Due Dates (all times Eastern Standard Time; see below for more info on the assignments):

Problem set 1 – Thursday July 16, 12:00pm

Problem set 2 – Tuesday July 21, 12:00pm

Problem set 3 – Friday July 24, 12:00pm

Problem set 4 – Wednesday July 29, 12:00pm

Policy Brief Project – Monday August 3, 11:59pm

### Live interactive coding sessions (all times Eastern Standard Time):

Tuesday July 14 at 9:00am – Live discussion: Getting Moving (to Opportunity)

Friday July 17 at 9:00am – Problem set 1 answer review

Monday July 20 at 9:00am – Live discussion: The Rent is too High

Wednesday July 22 at 9:00am – Problem set 2 answer review

Thursday July 23 at 7:00pm – Live discussion: Explore COVID-19 data

Monday July 27 at 7:00pm – Problem set 3 answer review

Tuesday July 28 at 9:00am – Live discussion: 8 Can't Wait

Thursday July 30 at 9:00am – Problem set 4 answer review

### Live discussion sections (all times Eastern Standard Time):

Thursday July 16 at 7:00pm – Inequality

Friday July 24 at 9:00am – COVID

Friday July 31 at 7pm – Criminal Justice

### Relevant email addresses:

Fitzpatrick: [maria.d.fitzpatrick@cornell.edu](mailto:maria.d.fitzpatrick@cornell.edu)

Hall: [mhall@cornell.edu](mailto:mhall@cornell.edu)

Postdoc: Chris Hess, [clh285@cornell.edu](mailto:clh285@cornell.edu)

TAs: Katharine Sadowski ([ks2373@cornell.edu](mailto:ks2373@cornell.edu)) Rene Crespin ([rac437@cornell.edu](mailto:rac437@cornell.edu)) Daphne Blakey ([dab487@cornell.edu](mailto:dab487@cornell.edu))

**Summary:** This course is an introductory overview of new innovations in data science, management, and programming, with accessible applications to a dynamic set of the most pressing social policy issues. Each semester, we will cover rotating topical modules ranging from income inequality to education to climate change to diversity and nativism to health. This semester, we will be covering inequality,

COVID-19, and criminal justice. Students will learn about important policy issues through both the economics and sociological lenses, how big data are being used to address policy problems, and technical skills for how to begin using data in similar ways to address policy problems.

**Grading:** Your grade in the course will be based on problem sets and a policy brief. The four problem sets will count for 15 percent of your grade each and the policy brief will count for 40 percent.

**Problem sets:** There will be 4 problem set assignments. Each problem set is designed to promote your development of data skills, as well as provide you with hands-on experience analyzing policy issues related to the three topics covered in this course. You can work on these in groups, though each person must hand in their own work. The problem sets will involve hands-on programming problems using R. These problem sets are designed without any prior knowledge of programming.

**Policy briefs:** As the culmination of the course you will be asked to write a policy brief memo to a mayor of a city. You will have the option to choose from one of the three different policy topics covered in the course. The questions will be posted during the first week of the course. Grades will be determined by the quality and clarity of the ideas and writing on the assignment, including how well the ideas are supported by existing data and how well you discuss the merits and pitfalls of use of data in the given context. Also, you should bring in data from the relevant section of the course, and will receive full credit for using the data appropriately.

**Texts:** Much of the reading for this class will be taken from popular summaries of research work and media discussion of important policy issues. In addition, we may use the following texts as useful materials for introducing ideas of how data science is incorporated into social science research. We also recommend the online textbook R for Data Science, found at this link: <https://r4ds.had.co.nz/>

**Big Data and Social Science: A practical guide to models and tools.** Taylor Francis 2016, Ian Foster, Rayid Ghani, Ron Jarmin, Frauke Kreuter and Julia Lane

**Bit by Bit: Social Research in the Digital Age.** Matthew J. Salganik. (2017). Princeton University Press.

**Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.** Cathy O’Neil. (2016). Penguin Books.

### **Income equality**

Leonhardt, David. 2013. “In Climbing Income Ladder, Location Matters.” *The New York Times*.

Leonhardt, David, Amanda Cox, and Claire Cain Miller. 2015. “An Atlas of Upward Mobility Shows Paths Out of Poverty.” *The New York Times*.

Aisch, Gregor, Eric Buth, Matthew Block, Amanda Cox, and Kevin Quealy. 2015. “The Best and Worst Places to Grow Up: How Your Area Compares.” *The New York Times*.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *Quarterly Journal of Economics* 29 (4): 1553–1623. Non-technical summary

Non-technical summary “The Impacts of Neighborhoods on Intergenerational Mobility ”

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.” *American Economic Review* 106 (4): 855–902. Non-technical summary

Davis, Bob. 2016. “Barely Half of 30-Year-Olds Earn More Than Their Parents.” *The Wall Street Journal*.

Non-technical summary “The Fading American Dream: Trends in Absolute Income Mobility Since 1940.”

Glaeser, Edward L., Scott Duke Kominers, Michael Luca, Nikhil Naik, 2018. “Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life,” *Economic Inquiry* 56(1): 114–137, January 2018

Cimpian, Andrei and Sarah-Jane Leslie. 2017. “Why Young Girls Don’t Think They Are Smart Enough.” *The New York Times*.

Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian. 2017. “Gender Stereotypes about Intellectual Ability Emerge Early and Influence Children’s Interests.” *Science* 391 (6323): 389–91.

#### *Optional Reading*

Alex Bell, Raj Chetty, Xavier Jaravel, Neviana Petkova, John Van Reenen. 2016. “The Lifecycle of Inventors.” Working Paper.

#### **Crime and punishment:**

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, Harold A. Pollack. 2015. “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago.” NBER Working Paper No. 21178.

Hvistendahl, Mara. 2016. Can ‘Predictive Policing’ Prevent Crime Before It Happens? *Science News*.

James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani, “Tree-Based Methods,” Chapter 8 in *An Introduction to Statistical Learning*.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. “Human Decisions and Machine Predictions.” NBER Working Paper No. 23180.

Kleinberg, John, Jens Ludwig, and Sendhil Mullainathan. 2016. *A Guide to Solving Social Problems with Machine Learning*. Harvard Business Review.

Mohler, George, Martin Short, P. Jeffrey Brantingham, Frederick Schoenberg, and George Tita. 2011. “Self-Exciting Point Process Modeling of Crime.” *Journal of the American Statistical Association* 106 (493): 100–108.

Algorithmic Risk Assessment Tools in the Hands of Humans. 2019. Doleac, Jen and Megan Stevenson.

Keep the Kids Inside? Juvenile Curfews and Urban Gun Violence (with Jillian B. Carr). *Review of Economics and Statistics*, 100(4): 609–618.