ECON 50 / HKS SUP 135 / GSE A218
Using Big Data to Solve Economic and Social Problems

Professor Raj Chetty
Email: chetty@opportunityinsights.org

Course Head: Gregory Bruich, Ph.D.
Email: gbruich@fas.harvard.edu
Office: Littauer Center 113

LECTURES Mondays and Wednesdays, 1:30-2:45 p.m. in Sanders Theater, 45 Quincy Street

COURSE DESCRIPTION

This course will show how "big data" can be used to understand and address some of the most important social and economic problems of our time. The course will give students an introduction to frontier research and policy applications in economics and social science in a non-technical manner that does not require prior coursework in economics or statistics, making it suitable both for students exploring economics for the first time, as well as for more advanced students. Topics include equality of opportunity, education, racial disparities, innovation and entrepreneurship, health care, climate change, criminal justice, and tax policy. In the context of these topics, the course will also provide an introduction to basic methods in data science, including regression, causal inference, and machine learning. The course will include discussions with leading researchers and practitioners who use big data in real-world applications.

INSTRUCTORS

Raj Chetty is the William A. Ackman Professor of Economics at Harvard. Raj’s current research, in collaboration with his Opportunity Insights research team, focuses on equality of opportunity: how can we give children from disadvantaged backgrounds the best chances of succeeding? Raj got his AB from Harvard in 2000 (where he lived in Hurlbut Hall as a first-year and then Pforzheimer House) and Ph.D. in 2003.

Gregory Bruich is a Lecturer in the Department of Economics at Harvard. He received his Ph.D. from the Department of Economics at Harvard University and his BA and BS from UC-Berkeley (where he took courses and worked with Raj). His research has focused on topics such as the impact of food stamp benefits on consumption and disability insurance on labor supply. In addition to Economics 50, he also teaches undergraduate and Ph.D.-level econometrics classes at Harvard. He is the Department of Economics’ Concentration Adviser for Dunster House and Mather House.
LABS

Labs (sections), held once a week for 75 minutes, are designed to give you hands-on experience in *doing* economics yourself by working through empirical problems with your teaching fellow. Please bring your computer. There will be two types of labs:

1. **No Prerequisite labs**: no specific background or prior knowledge required – we’re confident we’ll be able to bring you up to speed!

2. **Prerequisite labs**: prerequisite of previous completion of one of the following: a statistics class (Stat 100, 104, 110, or equivalent) OR an econometrics class (previous or concurrent enrollment in Economics 1123/1126) OR a computer science class.

If you are unsure about which lab is better for you, please discuss with Gregory Bruich or the teaching fellows. Both labs cover the same core content and course grades *did not differ* across students in the two types of labs in last year’s course; therefore, please simply enroll in the lab that matches your background without concern about its potential implications for your performance. Kennedy School students must enroll in the Prerequisite Lab.

You will be able to rank your preferences over time slots. The lab preference form will go live on Monday January 27. Please submit your preferences by Friday 1/31 at 11:59 p.m.

STUDENT ENGAGEMENT

Although this is a large course, we are eager to provide as personalized and hands-on a learning experience to each student as possible, through several avenues:

**OFFICE HOURS**: For office hours with the teaching fellows, there is a Google calendar on the course website where all the regular office hours and extra office hours times and locations will be posted.

Office hours with Dr. Bruich are 12-1 p.m. on Wednesdays in Littauer 113 and by appointment by writing to gbruich@fas.harvard.edu.

For office hours with Professor Chetty, please email Maddie Marino <mjmarino@g.harvard.edu> with a brief description of what you’d like to discuss. Appointments will typically be scheduled between 3-4 p.m. on Wednesdays.

**CANVAS DISCUSSION BOARD**: We encourage you to post questions on the Canvas website and teaching fellows will respond on the site. This is valuable because other students often have the same question that you do and will benefit from seeing your interaction with your teaching fellow. Of course, feel free to email TFs directly as well if you prefer not to post your question publicly.

**INFORMAL LUNCHES/DINNERS**: Professor Chetty and his research team at Opportunity Insights will host a small number of lunches/dinners with 12 students each to facilitate more informal interactions. If you are interested, please sign up using a Google form available on Canvas.
COURSE GOALS

The course has three principal learning objectives: 1) to introduce students to frontier social science research on key social and economic issues, 2) to teach students how to analyze data using modern quantitative methods and basic programming techniques, and 3) to show students how practitioners are using data to analyze social problems.

PEDAGOGICAL APPROACH

We seek to teach economics like a laboratory science, showing students how to do economics rather than presenting lectures about long-established results. Our approach draws inspiration from research on best practices in teaching in other settings. Mehta and Fine (2019) compared teaching practices across high school courses and found that in the most effective classes, “rather than touring students through the textbook, teachers invited students to participate in the authentic work of the field….For example, a skillful science teacher in a high-poverty-district high school offered a course in which her students designed, researched, carried out and wrote up original experiments.” We seek to apply this approach to teaching introductory empirical economics by discussing frontier research in lectures and having students engage in research themselves in labs and empirical projects. This is a work in progress, and we welcome your feedback on how we can improve this class as we teach the course at Harvard and also seek to support this approach at other colleges and high schools.

GENERAL EDUCATION AND OTHER CREDIT

This class satisfies the Quantitative Reasoning with Data (QRD) general education requirement. This class (when taken for a letter grade) meets the writing elective requirement for the Economics concentration. It also is an approved economics elective for the Applied Math-Economics concentration and the Economics Secondary Field. It counts towards the Technology and Governance Requirement for the Government Department’s Tech Science program. This class is a course connector with Statistics 10/Computer Science 10/Data Science 10. Students who enroll in both classes concurrently may be able to have overlap between their final projects for both classes, although the project must be approved by both courses’ faculties and expectations are higher for a project that is being used for both courses since it reflects effort and skills for two courses. This class is intended to complement Econ 10a/b by focusing on statistical methods and showing students how to apply the tools of economics using modern data science techniques.

EMPIRICAL PROJECTS

A key learning element of the course will be four empirical projects, which will give students hands-on experience in doing economics and working with data. We will teach and support the statistical software program Stata 16 for these projects, but students are welcome to use other programs (e.g., SAS, SPSS, R, Python), provided that their code and work is clearly documented. The empirical projects are more substantial than traditional problem sets and will include significant coding, reading, and writing elements that will put students in the shoes of social scientists doing research. Labs will be structured to provide the tools necessary to solve the empirical projects, and support will be provided so that coding skill is not a hindrance to achieving success on the projects.
LECTURE ATTENDANCE

In the interest of fostering interaction and discussion, students are required to attend all lectures. If you are unable to make it to a lecture because of illness (with doctor’s note) or Harvard Varsity athletics travel, you may email your documentation to Dr. Bruich to request access to a recording of the lecture for a limited time (7 days from when the lecture took place). Since guest lecturers have generously offered their time to our class, student attendance will be taken when we have a guest. Students’ grades will be partly based on attendance at those lectures (see below).

READINGS

There is no textbook for the course because the material is based entirely on recent research papers, mostly written within the past few years. Students are responsible for reading a small number of research papers, which appear in bold on the course reading list below. As we go along, we will let you know when each of the required readings should be done. The first reading should be done in the first week. Please focus on understanding the main ideas, rather than technical details. We recommend starting with non-technical summaries and introductions of research papers for this purpose.

GRADING

Grades will be based on a midterm exam (25 percent), final exam (25 percent), four empirical projects (40 percent), lecture attendance (5 percent), and class and lab activities (5 percent). Graduating seniors who are finishing theses this semester may choose to omit the midterm and have their grade based on the final exam (50 percent), empirical projects (40 percent), lecture attendance (5 percent), and class and lab activities (5 percent). All others are required to take the midterm.

The distribution of grades in this course will be similar to other large General Education courses at Harvard.

SCHEDULE FOR EXAMS AND EMPIRICAL PROJECTS

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project 1 part 1</td>
<td>11:59 p.m. on Wednesday, 2/12</td>
</tr>
<tr>
<td>Project 1 part 2</td>
<td>11:59 p.m. on Wednesday, 2/19</td>
</tr>
<tr>
<td>Project 2</td>
<td>11:59 p.m. on Wednesday, 3/4</td>
</tr>
<tr>
<td>Midterm exam</td>
<td>1:30-2:45 p.m. on Wednesday, 3/11</td>
</tr>
<tr>
<td>Project 3</td>
<td>11:59 p.m. on Wednesday, 4/8</td>
</tr>
<tr>
<td>Project 4 part 1</td>
<td>11:59 p.m. on Wednesday, 4/22</td>
</tr>
<tr>
<td>Project 4 part 2</td>
<td>11:59 p.m. on Wednesday, 4/29</td>
</tr>
<tr>
<td>Final exam</td>
<td>Scheduled by the FAS registrar for Tuesday, May 12 at 9:00 a.m.</td>
</tr>
</tbody>
</table>

PASS/FAIL: If you want to take the course pass/fail, please bring your pass/fail form to Gregory Bruich in Littauer 113.
ACADEMIC ACCOMMODATIONS: Students needing academic accommodations because of a documented disability must present their Faculty Letter from the Accessible Education Office (AEO) by Wednesday, February 5.

COLLABORATION POLICY: Discussion and the exchange of ideas are essential to academic work. You are encouraged to consult with your classmates on the empirical projects and to share sources. However, you should ensure that any work you submit for evaluation is the result of your own research and that it reflects your own approach to the topic. You must also adhere to standard citation practices and properly cite any books, articles, websites, lectures, etc. that have helped you with your work. If you received any help with your work (e.g., feedback on drafts, help with code or programming), you must also acknowledge this assistance. No collaboration of any kind is allowed during the midterm exam or the final exam.

ACADEMIC INTEGRITY: You are expected to uphold the Harvard College honor code and abide by the other University policies on academic honesty and integrity as given in the Harvard College Handbook for Students. As required by the College, all instances of suspected cheating will be referred to the Administrative Board.

ACKNOWLEDGMENTS

We are grateful to many people who have contributed to the development of this course and help support it, including Amanda Bayer, Syon Bhanot, Andy Housiaux, and Rebecca Toseland as well as Abigail Hiller, Madeleine Marino, Kate Musen, Trina Ott, Federico Gonzalez-Perez, Shannon Spence, and other members of the Opportunity Insights team.

TEACHING FELLOWS AND COURSE ASSISTANTS

Michael Droste………………… mdroste@fas.harvard.edu
John Macke………………… jmacke@g.harvard.edu
Lena Shi…………………. yulenashi@g.harvard.edu
Ian Nason …………………. nason@g.harvard.edu
Mikko Silliman…………… silliman@g.harvard.edu
Eric Andersen…………… ericandersen@g.harvard.edu
Harris Eppsteiner………….. heppsteiner@g.harvard.edu
José Ramón Enríquez…… jrenriquez@g.harvard.edu
Jiacheng Feng………………. fengji@g.harvard.edu
Maclaine Fields…………… mfields@college.harvard.edu
Sofia Garcia……………….. sofia_garcia@college.harvard.edu
Ariel Gomez……………….. arielgomez@g.harvard.edu
Chris Kuang……………….. ckuang@college.harvard.edu
Sophia Lee………………….. sophia_lee@college.harvard.edu
Dhruv Mohnot…………….. dhruvmohnot@college.harvard.edu
Jose Ramon Morales-Arilla.. jrm488@g.harvard.edu
Eddie Richardson…………. erichardson@college.harvard.edu
LJ Ristovska………………. ristovska@g.harvard.edu
Jay Sastry………………….. jsastry@college.harvard.edu
Naomi Vickers…………….. naomivickers@college.harvard.edu
Anthony Yu…………………. ayu@g.harvard.edu
## CLASS SCHEDULE (as of February 23, 2020; subject to change)

<table>
<thead>
<tr>
<th>Date</th>
<th>Num.</th>
<th>Topic</th>
<th>Selected Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Part I: Inequality and Social Mobility</strong></td>
<td></td>
</tr>
<tr>
<td>1/27</td>
<td>1</td>
<td>The Geography of Upward Mobility in America</td>
<td>correlation, regression</td>
</tr>
<tr>
<td>1/29</td>
<td>2</td>
<td>Causal Effects of Neighborhoods</td>
<td>experiments</td>
</tr>
<tr>
<td>1/31</td>
<td>3</td>
<td>Introductory Statistical Concepts</td>
<td>experiments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[optional lecture, taught by Gregory Bruich and Michael Droste]</td>
<td></td>
</tr>
<tr>
<td>2/3</td>
<td>4</td>
<td>Causal Effects of Neighborhoods and Characteristics of High-Mobility Areas</td>
<td>quasi-experiments</td>
</tr>
<tr>
<td>2/5</td>
<td>5</td>
<td>Policies to Increase Upward Mobility</td>
<td></td>
</tr>
<tr>
<td>2/10</td>
<td>6</td>
<td>Historical and International Evidence on the Drivers of Inequality and Mobility</td>
<td>cost-benefit analysis</td>
</tr>
<tr>
<td>2/12</td>
<td>7</td>
<td>Discussion with Edward Glaeser, Professor of Economics, Harvard University</td>
<td></td>
</tr>
<tr>
<td>2/17</td>
<td></td>
<td><em>No class - Presidents' Day</em></td>
<td></td>
</tr>
<tr>
<td>2/19</td>
<td>8</td>
<td>Upward Mobility, Innovation, and Growth</td>
<td>propensity score reweighting</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Part II: Education</strong></td>
<td></td>
</tr>
<tr>
<td>2/24</td>
<td>9</td>
<td>Higher Education and Upward Mobility</td>
<td>regression discontinuity</td>
</tr>
<tr>
<td>2/26</td>
<td>10</td>
<td>Primary Education</td>
<td>experiments</td>
</tr>
<tr>
<td>3/2</td>
<td>11</td>
<td>Teachers and Charter Schools</td>
<td>event study designs, competitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>equilibrium</td>
<td></td>
</tr>
<tr>
<td>3/4</td>
<td>12</td>
<td>Discussion with Geoffrey Canada, Founder, Harlem Children's Zone</td>
<td></td>
</tr>
<tr>
<td>3/9</td>
<td>13</td>
<td>Review [Bruich]</td>
<td></td>
</tr>
<tr>
<td>3/11</td>
<td></td>
<td><em>Midterm Exam</em></td>
<td></td>
</tr>
<tr>
<td>3/14-3/22</td>
<td></td>
<td><em>Spring Break - no class</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Part III: Racial Disparities &amp; Criminal Justice</strong></td>
<td></td>
</tr>
<tr>
<td>3/23</td>
<td>14</td>
<td>Racial Disparities in Economic Opportunity</td>
<td>dynamic models and steady states</td>
</tr>
<tr>
<td>3/25</td>
<td>15</td>
<td>Improving Judicial Decisions</td>
<td>machine learning, implicit bias</td>
</tr>
<tr>
<td>3/30</td>
<td>16</td>
<td>Discussion with Mahzarin Banaji, Professor of Social Ethics, Harvard University</td>
<td></td>
</tr>
<tr>
<td>4/1</td>
<td>17</td>
<td>Discussion with Jens Ludwig, Professor at the Univ. of Chicago Harris School of Public Policy</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Part</td>
<td>Session</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>4/6</td>
<td>Part IV: Health</td>
<td>Improving Health Outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>hazard models</td>
<td></td>
</tr>
<tr>
<td>4/8</td>
<td></td>
<td>The Economics of Health Care and Insurance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>adverse selection</td>
<td></td>
</tr>
<tr>
<td>4/13</td>
<td>Part V: Climate Change</td>
<td>Effects of Air and Water Pollution and Policies to Mitigate Climate Change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>diff-in-differences, externalities, discount rates, external validity</td>
<td></td>
</tr>
<tr>
<td>4/15</td>
<td></td>
<td>Discussion with Frances Moore, Professor of Environmental Science and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Policy at Univ. of California, Davis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>supply &amp; demand, synthetic control</td>
<td></td>
</tr>
<tr>
<td>4/22</td>
<td></td>
<td>Tax Policy 2: Savings and Wealth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>behavioral economics</td>
<td></td>
</tr>
<tr>
<td>4/27</td>
<td>Part VII: Economic Development and Institutional Change</td>
<td>Discussion with Melissa Dell, Professor of Economics, Harvard University</td>
<td></td>
</tr>
<tr>
<td>4/29</td>
<td></td>
<td>Institutions and Economic Development</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>historical data analysis</td>
<td></td>
</tr>
<tr>
<td>5/4</td>
<td></td>
<td>No class - Reading period</td>
<td></td>
</tr>
<tr>
<td>5/6</td>
<td></td>
<td>Review [Bruich]</td>
<td></td>
</tr>
</tbody>
</table>
COURSE READINGS

Students are responsible for reading a small number of required papers (in bold below). Please focus on understanding the main ideas, rather than technical details. We recommend starting with non-technical summaries and introductions for this purpose. The other papers will be discussed in lecture, in section, or in the empirical projects, and may be useful references in those contexts.

Part I: Equality of Opportunity

Geography of Economic Mobility


Policies to Improve Upward Mobility


Historical Trends


Using Big Data to Measure and Understand Cities


Innovation, Mobility, and Growth


Part II: Education

Higher Education


Primary Education


Charter Schools


Part III: Racial Disparities

Racial Disparities and Segregation


**Discrimination and Bias**


**Criminal Justice**


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


**Part IV: Health**

**Improving Health Outcomes**


**The Economics of Health Care and Insurance**


**Part V: Climate Change**

**Effects of Air and Water Pollution**


**Policies to Mitigate Climate Change**


**Part VI: Tax Policy**

*Income Taxation*


*Savings and Wealth*


**Part VII: Economic Development and Institutional Change**


**Statistics References**


**Stata Resources**

Stata 16 is available for [download from FAS IT](#).

Introduction to Stata and R for Economists: [https://canvas.harvard.edu/courses/19323](https://canvas.harvard.edu/courses/19323)


The Stata Blog: [https://blog.stata.com/](https://blog.stata.com/)