ECON 1152 and HKS SUP 135: Using Big Data to Solve Economic and Social Problems

Instructor: Professor Raj Chetty
Email: chetty@fas.harvard.edu
For appointments, contact Madeleine Marino: mjmarino@g.harvard.edu

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

Teaching Fellows:

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gregory Davis</td>
<td><a href="mailto:gdavis@g.harvard.edu">gdavis@g.harvard.edu</a></td>
</tr>
<tr>
<td>Pedro Gabriel Degiovanni</td>
<td><a href="mailto:pedrodegiovanni@g.harvard.edu">pedrodegiovanni@g.harvard.edu</a></td>
</tr>
<tr>
<td>Michael Droste</td>
<td><a href="mailto:mdroste@fas.harvard.edu">mdroste@fas.harvard.edu</a></td>
</tr>
<tr>
<td>Diana Goldemberg, Ph.D.</td>
<td><a href="mailto:diana_goldemberg@g.harvard.edu">diana_goldemberg@g.harvard.edu</a></td>
</tr>
<tr>
<td>Caroline Geiger</td>
<td><a href="mailto:caroline_kelley@g.harvard.edu">caroline_kelley@g.harvard.edu</a></td>
</tr>
<tr>
<td>Gru Han</td>
<td><a href="mailto:gruhan@fas.harvard.edu">gruhan@fas.harvard.edu</a></td>
</tr>
<tr>
<td>Hanno Hilbig</td>
<td><a href="mailto:hhilbig@g.harvard.edu">hhilbig@g.harvard.edu</a></td>
</tr>
<tr>
<td>Alejandro Lagomarsino</td>
<td><a href="mailto:alagomarsino@g.harvard.edu">alagomarsino@g.harvard.edu</a></td>
</tr>
<tr>
<td>John Macke</td>
<td><a href="mailto:jmacke@g.harvard.edu">jmacke@g.harvard.edu</a></td>
</tr>
<tr>
<td>Ian Nason</td>
<td><a href="mailto:nason@g.harvard.edu">nason@g.harvard.edu</a></td>
</tr>
<tr>
<td>Lena Shi</td>
<td><a href="mailto:yulenashi@g.harvard.edu">yulenashi@g.harvard.edu</a></td>
</tr>
<tr>
<td>Mikko Silliman</td>
<td><a href="mailto:silliman@g.harvard.edu">silliman@g.harvard.edu</a></td>
</tr>
<tr>
<td>Niharika Singh</td>
<td><a href="mailto:niharikasingh@g.harvard.edu">niharikasingh@g.harvard.edu</a></td>
</tr>
<tr>
<td>Tanggang Yuan, Ph.D.</td>
<td><a href="mailto:tanggangyuan@g.harvard.edu">tanggangyuan@g.harvard.edu</a></td>
</tr>
</tbody>
</table>

Lectures: Tuesday & Thursdays, 12-1:15 p.m in Sanders Theater, Memorial Hall, 45 Quincy Street

Sections: 1 per week at times to be arranged. Sections will be divided into two groups: one for students with no prior coursework in statistics/econometrics and another intended for those who have taken courses in statistics/econometrics. You may choose which type of section you would like to attend depending upon your background. Kennedy School students must enroll in the more advanced section.

Course Description: This course will show how "big data" can be used to understand and solve some of the most important social and economic problems of our time. The course will give students an introduction to frontier research in applied economics and social science that does not require prior coursework in Economics. Topics include equality of opportunity, education, innovation and entrepreneurship, health care, climate change, and crime. In the context of these topics, the course will also provide an introduction to basic statistical methods and data analysis techniques, including regression analysis, causal inference, quasi-experimental methods, and machine learning. The course will include discussions with leading practitioners who will discuss how they use big data in real-world applications.

Note: 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. It satisfies the Empirical and Mathematical Reasoning general education requirement.

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 to how practitioners are using data to analyze social problems.

**Academic Accommodations:** Students needing academic accommodations because of a documented disability must present their Faculty Letter from the Accessible Education Office (AEO) by Thursday, February 7.

**Grading:** Grades will be based on a midterm exam (30 percent), final exam (40 percent), four empirical projects (25 percent), and lecture attendance (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 (70 percent), empirical projects (25 percent), and lecture attendance (5 percent). See below for tentative dates for each of these assignments and exams.

<table>
<thead>
<tr>
<th>Schedule for Exams and Empirical Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assigned</strong></td>
</tr>
<tr>
<td>Empirical Project 1</td>
</tr>
<tr>
<td>Empirical Project 2</td>
</tr>
<tr>
<td>Midterm</td>
</tr>
<tr>
<td>Empirical Project 3</td>
</tr>
<tr>
<td>Empirical Project 4</td>
</tr>
<tr>
<td>Final exam</td>
</tr>
</tbody>
</table>

**Empirical Projects:** A key element of the course will be four empirical projects, which will give students hands-on experience in working with data. We recommend and will support using the statistical software program Stata for these projects, but students are welcome to use other programs (e.g., SPSS, R, Python), provided that their code and work is clearly documented. The empirical projects are designed to be more substantial than traditional problem sets and will include significant coding, reading, and writing elements that will give students a sense of how social scientists approach research.

**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 for a documented reason (e.g., an illness), you may request a recording of the lecture by emailing Gregory Bruich with documentation of the reason for your absence. Since guest discussants 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 above).

**Required readings.** Students are responsible for reading a small number of required papers, which appear in bold on the course reading list (starting on page 5 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 for this purpose.

**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 or final exam.
### Lectures

<table>
<thead>
<tr>
<th>Num.</th>
<th>Topic</th>
<th>Selected Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Part I: Equality of Opportunity</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>The Geography of Upward Mobility in America</td>
<td>correlation, regression</td>
</tr>
<tr>
<td>2</td>
<td>Causal Effects of Neighborhoods</td>
<td>quasi-experiments</td>
</tr>
<tr>
<td>3</td>
<td>Moving to Opportunity vs. Place-Based Approaches</td>
<td>experiments</td>
</tr>
<tr>
<td>4</td>
<td>The American Dream in Historical Perspective</td>
<td>distributional analysis</td>
</tr>
<tr>
<td>5</td>
<td>Upward Mobility, Innovation, and Economic Growth</td>
<td>propensity score reweighting</td>
</tr>
<tr>
<td></td>
<td><strong>Part II: Education</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Higher Education and Upward Mobility</td>
<td>Bayes rule</td>
</tr>
<tr>
<td>7</td>
<td>The Causal Effect of Colleges</td>
<td>regression discontinuity</td>
</tr>
<tr>
<td>8</td>
<td>Primary Education</td>
<td>experiments</td>
</tr>
<tr>
<td>9</td>
<td>Teachers and Charter Schools</td>
<td>event study designs, competitive equilibrium</td>
</tr>
<tr>
<td></td>
<td><strong>Part III: Racial Disparities</strong></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Racial Disparities in Economic Opportunity</td>
<td>dynamic models and steady states</td>
</tr>
<tr>
<td></td>
<td><strong>Part IV: Health</strong></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Improving Health Outcomes</td>
<td>hazard models</td>
</tr>
<tr>
<td>12</td>
<td>The Economics of Health Care and Insurance</td>
<td>adverse selection</td>
</tr>
<tr>
<td></td>
<td><strong>Part V: Criminal Justice</strong></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Improving Judicial Decisions</td>
<td>machine learning</td>
</tr>
<tr>
<td></td>
<td><strong>Part VI: Climate Change</strong></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Effects of Air and Water Pollution</td>
<td>diff-in-differences, externalities</td>
</tr>
<tr>
<td>15</td>
<td>Policies to Mitigate Climate Change</td>
<td>discount rates, external validity</td>
</tr>
<tr>
<td></td>
<td><strong>Part VII: Tax Policy</strong></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Income Taxation</td>
<td>supply and demand models,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>synthetic controls</td>
</tr>
<tr>
<td>17</td>
<td>Behavioral Public Economics</td>
<td>behavioral economics</td>
</tr>
<tr>
<td></td>
<td><strong>Part VIII: Economic Development and Institutional Change</strong></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Institutions and Economic Development</td>
<td>historical data analysis</td>
</tr>
</tbody>
</table>

*Note: the Spring 2019 course at Harvard included 5 guest lectures by external speakers, which are not included in the list above.*
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


Historical Trends


Innovation, Mobility, and Growth


Part II: Education

Higher Education


Primary Education


**Charter Schools**


**Part III: Racial Disparities**

**Racial Disparities and Segregation**


**Discrimination and Bias**


Part IV: Health

Improving Health Outcomes


The Economics of Health Care and Insurance


Part V: Criminal Justice


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


Part VI: Climate Change

Effects of Air and Water Pollution


Policies to Mitigate Climate Change


Part VII: Tax Policy

*Income Taxation*


*Part VIII: Economic Development and Institutional Change*


Statistics References


Stata Resources

Stata is available for download from FAS IT.

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


The Stata Blog: https://blog.stata.com/