SOCIOL 90Z: Inequality

Xiang Zhou

03/20/2020

E-mail: xiang_zhou@fas.harvard.edu
Office Hours: Monday 3-5pm
Office: 406 William James Hall

Web: scholar.harvard.edu/xzhou
Class Hours: Monday 9:45am-11:45am
Class Room: Robinson 105 (FAS)

Course Description

This course introduces students to both classical and frontier research on social and economic inequalities. By analyzing data from the Current Population Survey (CPS) and National Longitudinal Surveys of Youth (NLSY), we investigate educational inequality, labor market inequality, and intergenerational mobility in the United States. In the context of these topics, students will build skills in data analysis and data visualization in R and learn a variety of quantitative methods including regression, causal inference, and machine learning.

This is a small, hands-on class. I will likely adapt the syllabus to better fit students' needs and interests.

Prerequisites

A basic course in statistics such as Sociology 156, Stat 100/102/104, or an equivalent. Prior experience with R is encouraged but not required.

Course Details

Data sets

In this course, we are going to analyze data from two large-scale nationally representative social surveys: the Current Population Survey (CPS) and the National Longitudinal Survey of Youth 1979 (NLSY79). They can be accessed through IPUMS-CPS and NLS investigator, respectively.

- IPUMS-CPS: https://cps.ipums.org/cps/
- NLS investigator: https://www.nlsinfo.org/investigator/pages/login

Readings

To learn R, we will use DataCamp, an intuitive and interactive online platform, as well as *R for Data Science* by Grolemund and Wickham, available at https://r4ds.had.co.nz/. If you'd like a physical copy of the book, you can purchase it from Amazon.

For each substantive topic, we will read one or two research articles/book chapters. These readings will be made available on Canvas.

DataCamp

We will make extensive use of lessons from DataCamp. All DataCamp lessons will be graded pass/fail. Given the nature of the questions and the hints provided, it is not hard to get full credit as long as you make an honest effort.

All DataCamp lessons are due at **Sunday 10pm before class**.

Presentations

Each student is expected to make two presentations during the semester, once on the assigned readings and once on the final project (in the last week).

Final Project

The ultimate goal of this course is for you to apply the data analysis techniques learned over the semester to answer a substantively interesting question about social and economic inequality. You are encouraged to use the CPS/NLSY data provided in the class, but you are also free to find and use other data sets. The final project should be 5-15 pages long and focus on the description of research question, data, methodology, and results. Literature reviews or background material should be relatively brief. The final project write-up will be letter-graded.

Final project milestones

- **March 27**: A one-page description of your proposed project with at least two references and one regression table. Schedule a meeting with me to discuss your project.
- **April 17**: A 2-3 page description of your progress, including a description of your research question, analyses, preliminary results in tables and figures.
- **May 15**: A 5-15 page writeup of your final project.

Grading

- Attendance and participation (25%)
- DataCamp (15%)
- Presentation (10%)
- Final project milestones (10%)
- Final project presentation (10%)
- Final project write-up (30%)

Piazza

In this course, we will use an online discussion board called Piazza. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. We encourage you to use the Piazza Q&A board when asking questions about class materials and assignments outside of office hours. You can sign up to the Piazza course page using the link below (there are also free Piazza apps for Android and iOS devices):

http://piazza.com/harvard/spring2020/sociol90z

Using Piazza will allow you to see and learn from questions others have. I will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. Your respectful and constructive participation on the forum will count toward your class participation grade.

Computing

We will use R in this class, which you can download for free at https://cloud.r-project.org. R is open source and available on all operating systems. You can find a virtually endless set of resources for R on the internet, including this Getting Started With R page.

In this course, we will use DataCamp, an intuitive and interactive online platform, to learn R. We will also use RStudio, an editor and development environment for R.

Late Assignments

Late DataCamp assignments will receive half the credit if submitted within 48 hours of the deadline. Assignments received more than 48 hours after the deadline will not receive credit.

The final project will be marked down by one grade step for each 24 hour period past the deadline the paper is submitted (i.e. an A paper submitted within 24 hours past the deadline will receive an A-; an A paper submitted between 24 and 48 hours after the deadline would receive a B+).

Preliminary Schedule

The following is an anticipated schedule of course topics. The plan is to cover one topic per week, but we will go as fast as needed to make sure that everyone is understanding the material. Check the Canvas site to know what we will be covering in an upcoming class.

Week 01, 01/27: Introducing the Course & the Data sets

Week 02, 02/03: Data Manipulation & Visualization in R

DataCamp

- Install R and RStudio
- Introduction to the Tidyverse
- (Optional) R for Data Science, chapters 1-4

In class

- Working with dplyr and ggplot2

Week 03, 02/10: Describing Inequality

Readings

- Grusky and MacLean. 2016. "The Social Fallout of a High-inequality Regime"
- Autor 2014. "Skills, Education, and the Rise of Earnings Inequality among the 'Other 99 Percent'"

Presentation guideline

- Briefly summarize Grusky and MacLean (2016) and Autor (2014).
- What research questions have come to your mind while reading these articles? Identify a research question that you might pursue with the CPS data. Present this question and how you would pursue it with the CPS data.

DataCamp

- Introduction to Data Visualization with ggplot2
- (Optional) R for Data Science, chapters 5-8
- Browse and play with the CPS data

In class

- Describing wage/income inequality with the CPS data.

Week 04, 02/17: No Class (President's Day).

Week 05, 02/24: Categorical Inequality: Gender, Race, and Class

Readings

- Snipp and Cheung. 2016. "Changes in Racial and Gender Inequality since 1970."
- Xie, Killewald, and Near. 2016. "Between- and Within-Occupation Inequality: The Case of High-Status Professions"

Presentation guideline

- Briefly summarize Snipp and Cheung (2016) and Xie et al. (2016).
- What research questions have come to your mind while reading these articles? Identify a research question that you might pursue with the CPS data. Present this question and how you would pursue it with the CPS data.

DataCamp

- Working with Data in the Tidyverse, chapters 1-2
- (Optional) R for Data Science, chapters 9-11, 15
- Browse and play with the CPS data

In class

- Describing categorical inequality with the CPS data.

Week 06, 03/02: Educational Inequality

Readings

- Bailey and Dynarski. 2011. "Gains and Gaps: Changing Inequality in U.S. College Entry and Completion."

Presentation guideline

- Briefly summarize Bailey and Dynarski (2011).
- What research questions have come to your mind while reading this article? Identify a research question that you might pursue with the NLSY data. Present this question and how you would pursue it with the NLSY data.

DataCamp

- Working with Data in the Tidyverse, chapters 3-4.
- (Optional) R for Data Science, chapters 12-14
- Browse and play with the NLSY data

In class

- Analyzing educational inequality with the NLSY data.

Week 07, 03/09: Intergenerational Mobility

Readings

- Torche 2015. "Analyses of Intergenerational Mobility: An Interdisciplinary Review."
- Chetty, Hendren, Kline, and Saez. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility."

Presentation guideline

- Briefly summarize Torche (2015) and Chetty et al. (2014).
- What research questions have come to your mind while reading these articles? Identify a research question that you might pursue with the NLSY data. Present this question and how you would pursue it with the NLSY data.

DataCamp

- Modeling with Data in the Tidyverse, chapters 1-2.
- Browse and play with the NLSY data

In class

- Analyzing intergenerational mobility with the NLSY data.

Week 08, 03/16: No Class (Spring Break)

Week 09, 03/23: Income Attainment

Readings

- Witteveen and Attewell. 2017. "Family Background and Earnings Inequality among College Graduates."

Presentation guideline

- Briefly summarize Witteveen and Attewell (2017).
- What research questions have come to your mind while reading this article? Identify a research question that you might pursue with the NLSY data. Present this question and how you would pursue it with the NLSY data.

DataCamp

- Modeling with Data in the Tidyverse, chapter 3.
- Multiple and Logistic Regression in R, chapter 1
- Browse and play with the NLSY data

In class

- Modeling income attainment with the NLSY data; making regression tables.

Week 10, 03/30: Returns to Education.

Readings

- Hout 2012. "Social and Economic Returns to College Education in the United States."

Presentation guideline

- Briefly summarize Hout (2012).
- What research questions have come to your mind while reading this article? Identify a research question that you might pursue with the NLSY data. Present this question and how you would pursue it with the NLSY data.

DataCamp

- Multiple and Logistic Regression in R, chapters 2-3.

In class

- Estimating returns with education using multiple regression
- Interpreting and visualizing interaction effects.

Week 11, 04/06: Regression and Causality I

Readings

- Morgan and Winship, 2015. Counterfactuals and Causal Inference: Methods and Principles for Social Research. Chapters 1.1-1.5, 3.1-3.3

Presentation guideline

- Summarize, in your own words, of the rules of conditioning in regressions, as described in the reading, along with questions about things you find confusing in Morgan and Winship (2015)

In class

- Drawing causal diagrams; improving regression models.
- Heterogeneous causal effects.

Week 12, 04/13: Regression and Causality II

Readings

- Zhou. 2019. "Equalization or Selection? Reassessing the 'Meritocratic Power' of a College Degree in Intergenerational Income Mobility."

Presentation guideline

- Briefly summarize Zhou (2019).
- What are the things that have confused/inspired you? Share your confusions, thoughts, and inspirations.

In class

- Drawing causal diagrams; improving regression models.
- Mediation analysis.

Week 13, 04/20: Communicating Research Findings with RMarkdown

DataCamp

- Communicating with Data in the Tidyverse. chapter 3.
- (Optional) R for data science, chapters 26-30

In class

- Working with RMarkdown

Week 14, 04/27: Student Presentations