

Syllabus

SOC352 Growing Up Poor Across America: An Application of Sociological Reasoning to Data

Spring 2018

Instructor: John Kuk

Business Address

Class: T, TH 10:00–11:30AM

Room: Seigel 210

Course website: Piazza

Home Address

John Kuk

Room: Seigel 228

Email: jskuk@wustl.edu

Office Hours: Wednesday 1:00-3:00 PM

Course Description

The American Dream is built around the idea that children born into poor or disadvantaged families can, through education or hard work, eventually break into the middle and upper classes. This process, known as intergenerational mobility, is fundamental to sociological research and to social policies aiming to reduce income inequality. Yet recent research shows children who are raised in poor families are much more likely to enter the middle or upper classes in some cities than others—for instance, the probability that a child reaches the top quintile of the national income distribution starting from a family in the bottom quintile is more than three times higher in San Jose than it is in Charlotte. What might be driving these geographic differences in opportunities for low-income youth? Across the semester, we will work together as a class to try to answer this question.

Through our focus on intergenerational mobility, we will learn core skills for making an argument with quantitative data. By working throughout the semester with the Fragile Families dataset, students will learn the fundamentals of statistical thinking. Students will also develop the computational skills to put these ideas into practice by learning to manipulate data and make graphs that are effective at communicating their statistical knowledge to others.

Prerequisites

A willingness to work hard on possibly unfamiliar material is key.

About the Class and the Dataset

Class Structure: This is a class about a particular empirical puzzle in the study of intergenerational mobility. Recent research has shown some surprisingly large differences among communities

in the propensity of children born into poor families to move into the middle or upper class. Together we will investigate the factors that might drive these differences. Because this is a real puzzle, we don't have an answer for you. What we do have is some interesting data that might allow us to explore this question a bit more thoroughly. Our goal as instructors is to help you acquire the set of skills that will allow you to not only explore the data yourself, but also communicate the results of that analysis to others. In the course schedule below we outline the core skills we think will be most relevant but the actual course will be responsive to what we discover together as a class as well as the pace at which participants master the relevant data analysis techniques.

Class Expectations: This problem-driven structure means that the success of the course depends heavily on how seriously all of you take each session. We will occasionally assign light readings to provide background on the day's topic and we expect you to come to class having read and reflected on them. To help with learning the programming skills necessary to analyze data, we will be giving assignments in an online interactive learning system called Datacamp (www.datacamp.com). Learning to program is a lot like learning a language, it is best to have consistent, daily practice. Because we will be working hands-on with the data in class, we expect you to keep up with the Datacamp programming exercises so you can actively participate. Finally, we want every student to bring a laptop to class every day so we can all explore together. If you don't own a laptop please email the instructors and we will make arrangements for you.

Fragile Families: In order to learn more about intergenerational mobility we need access to some data. For this class we will make use of the Fragile Families & Child Wellbeing Study which follows a cohort of nearly 5,000 children born in large U.S. cities between 1998 and 2000. This study is ideal for our purposes because it includes information not only on the children themselves but also on their families. The sample includes a number of individuals who are growing up in poor households in a wide variety of cities around the U.S. From the study's website:

The core Study consists of interviews with both mothers and fathers at birth and again when children are ages one, three, five, and nine. The parent interviews collect information on attitudes, relationships, parenting behavior, demographic characteristics, health (mental and physical), economic and employment status, neighborhood characteristics, and program participation. Additionally, in-home assessments of children and their home environments were conducted at ages three, five, and nine. The in-home interview collects information on children's cognitive and emotional development, health, and home environment. Several collaborative studies provide additional information on parents' medical, employment and incarceration histories, religion, child care and early childhood education.

Requirements and Evaluation

Class participation: 20%
Weekly Homework Assignments: 30%
Midterm Exam: 15%
Final Project: 35%

Programming Exercises: Throughout the semester we will learn the statistical programming language R. We will use the open-source statistical software environment *R Studio*, which makes it

much easier and more intuitive to work with data using R. There is a steep learning curve with R, and you will discover that learning to program is fun and exciting, but it can also be frustrating at times. To facilitate learning of R, we will be using DataCamp. DataCamp will enable you to work through the programming exercises at your own pace, while accessing various types of support, both within DataCamp and the broader class community. The system will teach you all you need to know to use R for your own analyses, and you will have access to several supplementary courses that you can use to extend your knowledge beyond what is covered in the course. With the resources and exercises provided by DataCamp, activities and instruction during class and experience working through the problem sets, we are confident that all of you will learn the language of R during the semester, but students should expect to spend additional time learning and practicing.

Midterm (15%): The midterm exam will be designed to evaluate whether you have sufficiently developed the basic skills required to explore relationships between variables in R and also ask you to reflect on the problem of intergenerational mobility. The midterm will be a take-home format. We will distribute it on Tuesday of exam week after class and it will be due on Friday at 3:00 PM.

Final Paper (35%): Our goal in this class is to work together to learn something new about intergenerational mobility. In the process we will learn how to make convincing arguments using quantitative data. As a capstone project, each student will be required to choose a research topic you are interested in, pull data from the Fragile Families data, and use it to make an argument. The student will then create a short memo consisting of carefully-designed figures and short paragraphs making the argument. In week 6, students will present their choice of research topic.

Logistics

Emails and Piazza: Email should be used for personal issues, such as to schedule an appointment outside of office hours, to request an excused absence, or for feedback about grades. All other questions should be asked on *Piazza*. If you have a question about course content assignments, or logistics, please check *Piazza* first to see whether it has been asked already. If you email us with a question that is relevant to other members of the class, we will respond by directing you to post your comment to *Piazza*. Note that you can ask questions anonymously on *Piazza*. *Piazza* is designed so that students can answer each other's questions? we encourage you to use this feature. *Piazza* will be checked at least every 48 hours, and more frequently around the end of the term when papers are due. The link to *Piazza* will be announced later.

Late Submissions: For the seminar to work, all students must keep up with the course load throughout the semester. To encourage students to keep up with the material and to allow the instructors to provide timely feedback, assignments must be turned in on time. All assignments turned in after the deadline will be docked 10%, and an additional 10% for every 24 hours the assignment is delayed.

Course Schedule

We have provided a general overview of the four units and the weekly topics. Weekly topics are largely framed in terms of the data analysis skills we will be teaching but will also be driven by

topics in the domain area. However, because we are working through addressing a real problem, we won't quite know what aspects of the intergenerational mobility we will cover when.

Date	Topic	Coding and Projects
Unit 1: Beginnings		
Jan 16, 18	Introduction to Problem and Data	Intro to R Research ethics
Jan 23, 25	Understanding the Intergenerational Mobility Mystery	R programming building blocks Research design
Jan 30, Feb 1	Exploring the Data	R programming building blocks
Feb 6, 8	Measurement and Sampling	Subset and loop
Unit 2: Visualizing the Fragile Families Data		
Feb 13, 15	Plotting the Data	Tidy data and base plot
Feb 20, 22	Labeling and Annotating Graphs	ggplot2 and dplyr
Feb 27, Mar 1	Principles of Graphical Communication Correlation and regression	ggplot2, merge, and recode Creating initial project dataset
Unit 3: Relationships Between Variables		
Mar 5, 8	Measuring Associations Regression and prediction	Visualize project variables
Mar 13, 15		Spring break
Mar 20, 22	Regression	Presentation: literature review
Unit 4: Measuring Uncertainty		
Mar 27, 29	Measuring Uncertainty	Simulation and bootstrap Flash talk
Apr 3, 5	Visualizing Uncertainty	Bootstrap Flash talk 2
Unit 5: Predicting Poverty		
Apr 10, 12	Fragile Families Challenge	
Apr 17, 19	Text as data	
Apr 24, 26	Wrapup and Discussion of Findings	Final presentation

Readings

The majority of course readings will be articles. There is one required book for the class: Imai, Kosuke, 2017, *Quantitative Social Science*. Princeton University Press (QSS hereafter).

- Week 1
 - Chetty, Raj and Nathaniel Hendren. “The Effects of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects.” *Quarterly Journal of Economics*.
- Week 2
 - Reichman, Nancy, Julien Teitler, Irwin Garfinkel, and Sara McLanahan. 2001. “Fragile Families: Sample and Design.” *Children and Youth Services Review*. Vol. 23, Nos. 4/5: 303-326.
- Week 3
 - Cherlin, Andrew J. 2005. “American Marriage in the Early Twenty-First Century.” *The Future of Children*, Vol. 15, No. 2: 33-55
 - DeParle, Jason. 2012. “Two Classes, Divided by ‘I Do’.” *The New York Times*. July 14, 2012.
 - QSS Ch. 1
- Week 4
 - QSS Ch. 2
- Week 5
 - Desmond, Matthew. 2016. *Evicted*. New York NY: Crown Publishers.
- Week 6
 - Goffman, Alice. 2009. “On the Run: Wanted Men in a Philadelphia Ghetto.” *American Sociological Review* 74: 339–357.
- Week 7
 - Wildeman, Christopher. 2010. “Paternal Incarceration and Children’s Physically Aggressive Behaviors: Evidence from the Fragile Families and Child Wellbeing Study.” *Social Forces* 89(1): 285–310.
 - Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Graphics Press. 2nd ed.
- Week 8
 - QSS Ch. 3
- Week 9: Spring break
- Week 10

- Phillips, Meredith and Tiffani Chin. 2004. “School Inequality: What Do We Know?” Pp. 467-510 in *Social Inequality* edited by Kathryn Neckerman. New York: Russell Sage Foundation Press.
- Week 11
 - Heckman, James. 2006. “Skill Formation and the Economics of Investing in Disadvantaged Children.” *Science* 312: 1900-1902.
- Week 12
 - QSS Ch. 4
- Week 13
 - Boyd, Danah, and Kate Crawford. 2012. “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.” *Information, Communication & Society* 15: 662–679.
- Week 14
 - Pebly, Anne and Narayan Sastry. 2004. “Neighborhoods, Poverty, and Children’s Well-Being.” Pp. 119-145 in *Social Inequality* edited by Kathryn Neckerman. New York: Russell Sage Foundation Press.

Homework Assignments

1. Comparing Perceptions of Neighborhood Violence/Parents’ Concerns across Groups
 - Practice `subset`, `table`, `prop.table`, `for loop`, and summary statistics
2. Eviction
 - Practice for `loop`, `ggplot2` and learn survey non-response.
3. Simulation, Plotting, Tidy Data
 - Learn simulation, practice `ggplot2`, `tidyr`, `dplyr`
4. Regression
 - Learn t-test, regression and practice `lm`
5. Associations Between 2 Variables
 - Practice more `ggplot2` and learn bootstrap, LOESS

Acknowledgment

I learned tremulously from my teachers and friends. This course is built upon, or borrows from, course materials prepared by Brandon Stewart and Margaret Frye (Princeton Sociology 320), and Molly Roberts (UCSD POLI170A).

February, 2018