



Sociology 320

Growing Up Poor Across America: An Introduction to Reasoning With Data

Fall 2016

Margaret Frye
mfrye@princeton.edu
Office: Wallace Hall Rm. 147
Office Hours: TBD

Brandon Stewart
bms4@princeton.edu
Office: Wallace Hall Rm. 149
Office Hours: TBD

Course Website: Blackboard

Lectures: Monday and Wednesday, 11am-12:20pm

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: This course has no prerequisites but is designed to target students majoring in sociology, to be taken in advance of SOC 305: Quantitative Social Science. Juniors enrolling in the course would take SOC 305 in their senior year.

Grade Distribution:

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

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).¹ 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 childrens 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.

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 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

¹Datacamp has a free trial that gives you access to some of the basic lessons if you want to see what its like before committing to the class

with the software in order to become comfortable enough to use it in their own analyses. We will periodically assign additional exercises specific to the Fragile Families data.

Midterm: 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 Monday of exam week after class and it will be due on Thursday at 5:00 PM.

Final Paper: 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 variable from the Fragile Families data and use it to make an argument about the relationship between that variable and the cognitive skills outcome variable discussed in class. The student will then create a short memo consisting of a single carefully-designed figure and a short paragraph making the argument.

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. *Piazza* can be accessed directly through Blackboard, and you will already be signed in if you click the link through blackboard.

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.

Recognizing that unexpected delays may occur, we will allow each of you two automatic, no-questions-asked, 24-hour extension on any of the homework assignments. Extensions of more than 24 hours, or extensions on more than two assignments, are not automatic and must be requested to us in writing. If you find yourself in such a situation, please send us an email as soon as possible. No extensions will be permitted on the midterm exam or the final paper without written note from a residential dean or director of studies.

Readings: The majority of course readings will be articles. We have two required purchases:

- Tufte, Edward R. *The Visual Display of Quantitative Information*. 2001
- Subscription to Datacamp²

For those who would like a textbook based treatment of R, we also recommend *A Beginner's Guide to R* by Zuur, Ieno and Meesters which is available for free online through the Princeton library.

²Not really a book, but a required purchase! We will explain how to purchase this in class.

Academic Integrity All Princeton students pledge to adhere to the Honor Code in the conduct of all written examinations, tests, and quizzes that take place in class. However, much of your work at Princeton – from papers to problem sets to the senior thesis – will be produced outside of class, whether in the library, your dorm room, or elsewhere. Such work is governed by the University’s academic regulations, which are designed to ensure the integrity of your academic work. Intellectual honesty is the cornerstone of our academic community. You must always distinguish your own words and ideas from the words and ideas of others – including the authors of print or electronic sources, faculty members, classmates, and friends.

Failure to acknowledge the sources – textual, personal, electronic – upon which you’ve relied is a serious breach of academic integrity. Such a failure can lead to the accusation of plagiarism – defined as the use of any source, published or unpublished, without proper acknowledgment. Plagiarism is a very serious charge at Princeton, which can result in disciplinary probation, suspension, or expulsion. As per university regulations, ignorance of the university policy is not an acceptable defense against the charge of plagiarism. We will be sure to clearly communicate which aspects of the course you are allowed to work together on, but it is your responsibility to ensure that you fully understand the parameters of each assignment.

Course Schedule: The course schedule below is subject to change in order to adapt to participants’ interests. 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.

- Unit 1: Beginnings
 - Week 1: Introduction to Problem and Data (September 14)
 - Week 2: Understanding the Intergenerational Mobility Mystery (September 19, 21)
 - Week 3: Exploring the Data (September 26, 28)
- Unit 2: Visualizing the Fragile Families Data
 - Week 4: Plotting the Data (October 3, 5)
 - Week 5: Labeling and Annotating Graphs (October 10, 12)
 - Week 6: Principles of Graphical Communication (October 17, 19)
- Unit 3: Relationships Between Variables
 - Week 7: Visualizing Relationships (October 24, 26)
 - Week 8: Measuring Associations (November 7, 9)
 - Week 9: Communicating Associations (November 14, 16)
- Unit 4: Measuring Uncertainty
 - Week 10: Measuring Uncertainty (November 21)
 - Week 11: Visualizing Uncertainty (November 28, 30)
 - Week 12: Communicating Uncertainty (December 5, 7)
 - Week 13: Wrapup and Discussion of Findings (December 12, 14)