

Katharina Roesler

katroesler@gmail.com — kroesler.com — GitHub

EDUCATION Ph.D. Stanford University, Department of Sociology June 2017
M.A. Stanford University, Department of Sociology June 2016
B.A. Princeton University, Sociology, *Summa Cum Laude* June 2012

SKILLS *Programming:* Python, SQL (incl. Redshift), Javascript, Flask, AngularJS, NetLogo
Statistics: Pandas, R, Matlab, Stata
Other: Linux/Unix, Vim, AWS, Google Cloud, MTurk, Qualtrics, L^AT_EX,
Languages: English(native), German(native), French(proficient), Spanish(basic),
Hebrew(basic)

PROFESSIONAL EXPERIENCE *Data Scientist at Quora* July 2017 - Present
I am currently a data scientist at Quora and love my job. I spend most of my days writing SQL and Python to analyze data and maintain our internal dashboards.

Consulting Data Scientist June 2016 - September 2016
I did a data science internship at Applied Research Works, Inc., which is building a healthcare analytics tool for doctors and insurance companies. Together with Dr. Rupinder Singh, I wrote [a research paper](#) on fee-for-service Medicare costs, which we are currently submitting to medical policy journals.

DATA SCIENCE PROJECTS *Machine Learning Project* Spring 2016
I took a Machine Learning class (CS 229), for which I wrote [this paper](#) and made [this poster](#). I used boosting to identify key features and then predicted political Independence with 89 percent test set accuracy.

Yelp Dataset Challenge Autumn 2016
I submitted a [paper](#) to Round 8 of Yelp's Dataset Challenge. I classified 1.6 million reviews' topics using logistic regression and Naive Bayes classifiers, with test set accuracies ranging from 84 to 98 percent.

NetLogo Model Winter 2017
I built a [NetLogo model](#) of workplace friendships and job satisfaction.

COMPUTER SCIENCE COURSES Programming Methodology (CS 106A) Dr. Mehran Sahami
Programming Abstractions (CS 106B) Dr. Marty Stepp
Introduction to Scientific Python (CME 193) Daniel Frank
Social Information and Network Analysis (CS 224W) Dr. Lada Adamic
Machine Learning (CS 229) Dr. John Duchi
Design and Analysis of Algorithms (CS 161) Dr. Gregory Valiant
Modern Algorithmic Toolbox (CS 168) Drs. Roughgarden and Valiant

STATISTICS COURSES Applied Matrix Theory (MATH 104) Dr. Angela Hicks
Introduction to Probability (STATS 116) Dr. David Donohu
Sociological Methodology I: Introduction Dr. Michael Rosenfeld
Sociological Methodology II: Multivariate Regression Dr. Cristobal Young
Sociological Methodology III: Models for Discrete Outcomes Dr. Xueguang Zhou

**RESEARCH
EXPERIENCE**

Dissertation *Autumn 2014 - Present*
My dissertation examines policing and racial disparities in arrest rates. The primary objective is to understand the causes and extent of police discrimination against black men. I ask whether police departments with more discretionary policing and with more white officers are more likely to discriminate against black men. I scraped [Mugshots.com](#) and then used [NIBRS data](#) on criminal offenders, [LEMAS data](#) on police agencies, and [ACS data](#) on county demographics. I used Python to clean and merge the data and R to fit multilevel and robust logistic regression models predicting offenders' likelihood of arrest.

Research Assistant *Spring 2014 - Present*
I am conducting research on marital dissolution rates with Dr. Michael Rosenfeld and Taylor Orth, using data from the [National Survey of Family Growth \(NSFG\)](#). I merged nine waves of the NSFG and transformed them into longitudinal format. We then estimated individuals' likelihood of divorce and/or separation using discrete time logistic regression and Cox proportional hazards models.

Research Assistant *Summer 2015 - Present*
Dr. Tomás Jiménez, Anna Boch, and I are examining Mexicans' assimilation into the US as viewed through the lens of food. We examine whether Mexican food is more "mainstream" (ethnicity is less salient) in areas with larger Mexican populations. To obtain our data, I scraped 191,055 Yelp.com reviews of Mexican restaurants across the US and cleaned Round 8 of Yelp's Academic Dataset.

Research Assistant *Autumn 2015 - Summer 2016*
Dr. Cristobal Young and I examined "billionaire migration," billionaires' choice of residence in response to tax rates. I scraped Forbes.com and Fortune.com to compile a dataset of billionaires and estimated discrete-time event history models to predict their migration, as well as linear and Poisson regression models to predict states' and countries' billionaire populations.

Research Assistant *Spring 2015 - Summer 2015*
I cleaned a 100 GB file with all California tax records from 1987 to 2012 using only my own laptop (due to privacy concerns). I transformed the data into longitudinal format and created maps of zip codes' average incomes.

Research Assistant *Winter 2015 - Spring 2015*
I administered a Mechanical Turk survey experiment and analyzed the results in Stata for Dr. Corey Fields.

**WEB
DEVELOPMENT**

Transcribr *Present*
I am making [transcribr.io](#), a website that converts audio files to text. The goal is to enable people with audio from interviews to easily extract the text.

Yelp Map *Autumn 2016*
I made a [Flask app](#) showing Yelp restaurant reviews.

FotoMap *Autumn 2016*
I made [foto-map.com](#) (currently deactivated), a website that visualizes your photos on a map. I used AngularJS with a MySQL database and Flask backend.

TEACHING

Instructor *Spring, Autumn 2016*
Introductory Sessions to R and [Python](#)