

Jaren Haber

POSTDOCTORAL FELLOW @ QUANTITATIVE SOCIAL SCIENCE, DARTMOUTH COLLEGE

jhaber@berkeley.edu | jarenhaber.com | [jhaber-zz](https://twitter.com/jhaber-zz) | [jaren-haber-2b17a85](https://github.com/jaren-haber-2b17a85) | [jhaber43](https://twitter.com/jhaber43)

Summary

Experienced computational social scientist specializing in educational data science and reproducible, open-access workflows. Diverse data science background, from project conception & capacity-building to delivery. Skilled in natural language processing, machine learning, and statistical techniques. Committed to positive social impact through collaboration and research on inequality and policy.

Skills

Researcher	Conceptualize, analyze, & publish research; Clearly communicate findings to broad audiences; Collaborate on grants
Education Analyst	Common Core of Data , EdFacts , ACS (school districts), CMO directory , school finances , web-scraped charter schools corpus
Project Manager	Recruit, mentor, & lead data science teams (60+ apprentices); Manage codebases; Manage compute infrastructure
Programmer	Python (pandas, nltk, gensim, scikit-learn, beautifulsoup, keras), R (stm, tidytext), Stata, Git, Docker, UNIX, Tableau, Office
Content Creator	Natural Language Processing , Topic Modeling , Classification , Web-Crawling , Word Embeddings , Sociology of School Choice
Lead Organizer	Fund-raise & lead data science institutes (BAY-SICSS , TextXD) and NLP trainings (CTAWG , GUITAR)

Experience

Quantitative Social Science

Dartmouth College

POST-DOCTORAL FELLOW

2022–2023

- Taught [QSS20: Modern Statistical Computing](#), a data science course & major requirement w/ [community-engaged capstone project](#)
- Designed & conducted digital experiments with ~600 subjects to analyze perceptions of charter school performance ([methods registration](#))
- Designed & led [web-scraping and text parsing of obituary aggregator Legacy.com](#) to provide mortality estimates faster than official sources, yielding ~84K obituaries in Washington, D.C., from 2015–2021; build regression models to predict demographic-specific mortality

Massive Data Institute

Georgetown University, D.C.

POST-DOCTORAL FELLOW

2020–2022

- [Wrote & published analyses](#) of the diffusion of misinformation topics from political elites through mainstream & social media into public memory during 2020 presidential election, using ~187K news articles, ~1.2K TV channels' transcripts, ~170m tweets, and ~18K surveys
- Developed [computational workflow](#) for rapid visualization of massive data on per-pupil expenditures & demographics for [Edonomics Lab](#)
- Implemented web-scraping for Fact Context Framer, an interactive tool to visualize misinformation in newspapers and social media
- Innovated blended supervised classification and keyword sampling workflow for tracking COVID-related myths in millions of Twitter posts
- Wrote & published analyses [to inform gun policy](#), estimating gun-related mortality and gun ownership using social media data
- Led development of Crawl4All, a universal, accessible web-crawling platform for massive web-based data collection ([server code](#))

Doctoral Program, Sociology

University of California, Berkeley, CA

GRADUATE STUDENT RESEARCHER / PH.D. CANDIDATE / PROJECT MANAGER

2015–2020

- Web-crawled ~6,000 charter school websites ([crawling code](#)), linked federal data to custom web corpus, built & implemented word embeddings and mixed regression models to develop linguistic measures of school race & class signals ([replication code](#)) ([major journal publication](#))
- Implemented transformers to classify articles for computational literature review, with average F_1 score of 0.856 ([Towards Data Science post](#))
- Developed [open-source codebase](#) for supervised and unsupervised machine learning to process & analyze complex texts ([preprint](#))
- Developed & implemented text analysis workshops and curriculum, hackathon workflows, and cloud infrastructure for coding sessions

Recent Awards

2023	Research Allocation, Jetstream2 (\$136,575 value) , ACCESS network (Co-PI w/ Prof. H. Haveman)	IU & TACC
2022	Sciences Po & McCourt grant (\$500,000) , <i>Diffusion of Digital Disinformation</i> (Collab. w/ Prof. J. Schradie)	Paris & D.C.
2022	NIH R03 grant (\$100,000) , <i>Digital Obituaries for Public Health Surveillance</i> (Co-Investigator w/ Prof. M. Alva)	Washington, D.C.

Education

University of California, Berkeley

Berkeley, CA

PH.D. IN SOCIOLOGY, EXAM AREAS: ORGANIZATIONS, SOCIAL PSYCHOLOGY, THEORY

May 2020

University of California, Berkeley

Berkeley, CA

M.A. IN SOCIOLOGY

University of California, Davis

Davis, CA

B.A. IN SOCIO-CULTURAL ANTHROPOLOGY (WITH HONORS)