

Jaren R. Haber, PhD

✉ jhaber@berkeley.edu 🌐 jarenhaber.com 🐙 github.com/jhaber-zz 🆔 [0000-0002-5093-8895](https://orcid.org/0000-0002-5093-8895)

Education

- 2020 PhD, Sociology, University of California (UC) Berkeley
- 2013 MA, Sociology, UC Berkeley
- 2010 BA (with honors), Socio-Cultural Anthropology, UC Davis

Research Interests

- quantitative methods • social stratification • social policy • sociology of education • misinformation
- organizational sociology • natural language processing • web-scraping

Publications

Haber, Jaren, Lisa Singh, Ceren Budak, Josh Pasek, Meenakshi Balan, Ryan Callahan, Rob Churchill, Brandon Herren, and Kornraphop Kawintiranon. 2021. Research Note: Lies and Presidential Debates: How Political Misinformation Spread Across Media Streams during the 2020 Election. *Harvard Kennedy School (HKS) Misinformation Review* 2(6). doi: [10.37016/mr-2020-84](https://doi.org/10.37016/mr-2020-84).

Haber, Jaren. 2021. Sorting Schools: A Computational Analysis of Charter School Identities and Stratification. *Sociology of Education* 94(1):43–64. doi: [10.1177/0038040720953218](https://doi.org/10.1177/0038040720953218).

- Replication data and code: https://bit.ly/sorting_schools
- Post-print (Open Science Foundation): <http://bit.ly/OSF-postprint-SocEd>
- Media coverage: <https://bit.ly/charters-highlight>, <https://bit.ly/BIDS-post-2020>

Haber, Jaren. 2016. Institutionalized Involvement: Teams and Stress in 1990s U.S. Steel. *Industrial Relations: A Journal of Economy and Society* 55(4):632–61. doi: [10.1111/irel.12154](https://doi.org/10.1111/irel.12154).

Papers Under Review

Haber, Jaren, Lisa Singh, Alex Chen, Kornraphop Kawintiranon, Aidan Pizzo, Anna Pogrebivsky, and Joyce Yang. Identifying High Quality Training Data for Misinformation Detection. Under review at *Journal of Computational Social Science*.

Gresenz, Carole Roan, Lisa Singh, Yanchen Wang, **Jaren Haber**, and Yaguang Liu. Advancing Gun Violence Research with Social Media Data: A Computational Analysis of Gun Ownership. Under review at *American Journal of Preventive Medicine*.

Work in Progress

Haber, Jaren, and Heather A. Haveman. Toward Computational Literature Reviews: Applying Expert-Built Dictionaries for Automated Analysis of Academic Texts. [Draft available](#).

Haber, Jaren, Nick Camp, and Jae Yeon Kim. Digitizing Segregation: How Charter School Websites Signal Race- and Class-Specific Educational Ideologies. [Pre-registration available](#).

Singh, Lisa, Rob Churchill, **Jaren Haber**, Kornraphop Kawintiranon, Yaguang Liu, Agoritsa Polyzou, and Yanchen Wang. Fact Context Framer: Illuminating the Relationship Between Facts and Misinformation. Draft available on request.

Fellowships, Grants, and Awards

Allocations received from *Extreme Science and Engineering Discovery Environment (XSEDE)*:

- Research Allocation, Jetstream2 (\$136,575 value), with Heather Haveman. Jan. 2022–Dec. 2022
 - Research Allocation, Jetstream (\$15,000 value), with Heather Haveman. Jan. 2021–Dec. 2021
- Post-SICSS Grant. September 27, 2019 (\$6,463)

Selected participant. *Summer Institute in Computational Social Science (SICSS)*. Princeton, June 2019

Leo Lowenthal Fellowship. UC Berkeley, Fall 2018 (\$10,000)

Fellowship Award. *HathiTrust Research Center UnCamp*, UC Libraries, Jan. 25–26, 2018

Dissertation Completion Fellowship. UC Berkeley, 2016–2017 (\$18,000)

Outstanding Graduate Student Instructor award. UC Berkeley, 2016 (\$250)

Cota Robles Fellowship. Sociology Dept., UC Berkeley, 2010–2012 (\$50,000)

Citation of Outstanding Performance. Socio-Cultural Anthropology Dept., UC Davis, 2007 (\$1,000)

First Year Scholar and Dean’s List. UC Davis, 2006–2007

President’s List. California Polytechnic State University, San Luis Obispo, 2003–2006

Research Experience

Postdoctoral Fellow, Quantitative Social Science , Dartmouth College	Sept. 2022–present
Postdoctoral Fellow, Massive Data Institute (MDI) , Georgetown University	Aug. 2020–Aug. 2022
Collaborator, “Ideology and Institutions in the Diffusion of Disinformation in the Digital Era”, Sciences Po/McCourt Institute grant with Jen Schradie, PI .	Apr. 2022–present
Co-Investigator, “The use of online obituaries as a tool for public health surveillance,” NIH R03 grant with Maria Alva, PI, MDI .	Jan. 2022–present
Graduate Student Researcher with Heather Haveman, Sociology, UC Berkeley	May–Dec. 2019
Graduate Student Researcher with Annette Bernhardt, Institute for Research on Labor and Employment, UC Berkeley	May–July 2014
Graduate Student Researcher with Andreea Gorbatai, Management of Organizations Program, Haas School of Business, UC Berkeley	May–Aug. 2013
Honors Research Project at UC Davis and University of Ghana, Accra	Aug.–Dec. 2007

Teaching Experience

Course Designer and Lecturer

Modern Statistical Computing (Dartmouth)	Fall 2022, Winter 2023
Sociology of School Choice (UC Berkeley)	Spring 2020, Fall 2017

Graduate Student Assistant (UC Berkeley)

Sociology of Education (American Cultures Engaged Scholarship)	Spring 2016, Fall 2014
Sociological Methods: Evaluation of Evidence	Spring 2018, Spring 2013
Introduction to Sociology	4x Fall 2012–Spring 2015
Politics & Social Change	Fall 2015

Workshop Leader (selected)

“Web-Crawling: Essential Tools for Wrangling Web Data”. International Conference on Computational Social Science (IC2S2).	July 19, 2022
“Topic Models: A Practical Introduction in Python”. MDI	Mar. 15–16, 2022
“Text Classification: A Practical Introduction in Python”. MDI	Oct. 12–13, 2021
“Natural Language Processing: An Introduction in Python”. MDI	Oct. 19–20, 2020
“Separating ‘signal’ from ‘noise’ in text with Structural Topic Models”. Computational Text Analysis Working Group (CTAWG) , D-Lab, UC Berkeley	Apr. 29, 2020
“Introduction to text as data”. Training for Text Across Domains symposium (TextXD) , Berkeley Institute for Data Science (BIDS) , UC Berkeley	Dec. 3, 2019
“Publishing computational research in social sciences journals”. CTAWG	Nov. 13, 2019
“Web-scraping at scale”. Guest lecture for <i>Introduction to Computational Tools for Social Science</i> , graduate course, UC Berkeley	Mar. 20, 2019
“Introduction to word embedding models”. Training for <i>TextXD</i> , BIDS	Dec. 4, 2018
“Validating themes in text with word embedding models”. CTAWG	Sept. 12, 2018
“Capturing a corpus of inconsistent websites”. <i>Making Text Research-Ready Symposium</i> , Social Science Matrix, UC Berkeley	Apr. 30, 2018
“Scraping the unscrapable with Python”. <i>Data Science Faire</i> , BIDS	May 2, 2017
“Introduction to a thorough, practical CTA training”. CTAWG	Oct. 4, 2017

Presentations (selected; [see my talks](#))

“Lies and presidential debates: How political misinformation spread across media streams during the 2020 election”. IC2S2 .	July 20, 2022
“Toward computational literature reviews: Combining supervised and unsupervised text-analysis”. Pacific Sociological Association Conference.	April 9, 2022
“Sorting schools: A computational analysis of charter school identities and stratification”. Sociology of Education Association (SEA) annual meeting	Feb. 25, 2022
“Data-driven school choice policy?: How school web data can inform solutions to educational segregation”. MDI Scholars meeting, Georgetown U.	Apr. 16, 2021

“Digitizing segregation: How charter school websites signal race- and class-specific educational ideologies”. Colloquium, Sociology Dept., Georgetown U.	Mar. 17, 2021
“Lessons from text analysis on the Wild Web”. Guest lecture for <i>Text as Data in the Digital Era</i> , undergraduate course, Mt. Holyoke College	Feb. 12, 2021
“Applied NLP and organizational analysis: Research on charter schools and organizational theories”. <i>Georgetown U. Computational Linguistics</i> lab meeting.	Nov. 13, 2020
“Coding for reproducibility”. MDI Scholars meeting, Georgetown U.	Nov. 13, 2020
“A supercomputer reviews the literature: Combining supervised and unsupervised text analysis”. <i>Computational Social Science Forum</i> , BIDS	Sept. 28, 2020
“Marketing drives charter school segregation”. Poster/talk at <i>Conference on Educational Data Science</i> , Institute for Research in Social Sciences, Stanford U.	Sept. 18, 2020
“Sorting schools: A computational analysis of charter school identities and stratification”. Roundtable, American Sociological Association (ASA) meeting	Aug. 13, 2020
“Do charter schools sell segregation?: Pilot results from text analysis and online experiments”. Roundtable, ASA	Aug. 12, 2019
“Inductive dictionary creation with word embedding models”. <i>TextXD</i> , BIDS	Dec. 5, 2018
“Stratification through organizational identity: Do charter schools today reinforce U.S. social inequalities?” Roundtable, ASA	Aug. 12, 2018
“Charter schools as big data: Early results of computational analyses”. <i>Education Research Day</i> , Graduate School of Education (GSE), UC Berkeley	Mar. 16, 2018
“Moral education and state regulation”. <i>Education Research Day</i> , GSE	Apr. 3, 2015

Professional Activities and Service

Service at Georgetown University

Project Supervisor, “Web-Crawling for All: Toward a Universal, Accessible Framework for Collecting Digital Data”. MDI Scholars program	2020–2022
Founder and Organizer. <i>Georgetown University Interdisciplinary Text Analysis Research (GUITAR)</i> working group	2020–2022
Workshop Co-Lead. MDI Scholars program	2020–2022

Service at UC Berkeley

Project Supervisor of 40+ undergraduates. Undergraduate Research Apprentice and BIDS Data Science Discovery Programs	2017–present
Allocation Manager. Startup, Renewal, and Research Allocations, XSEDE	2017–present
Graduate Affiliate. Oski Lab data science initiative	2017–present
Topic Area Support, Text Analysis. D-Lab	Spring 2019

Organizer. Job Market Practice Talks, Sociology Dept.	2014, 2016
Graduate Student Scholar. Institute for the Study of Societal Issues	2010–2020
Editorial Board. <i>Berkeley Journal of Sociology</i> , Sociology Dept.	2010–2012

Conferences, Symposia, and Working Groups

Chair. <i>Belonging & Education</i> panel, Pacific Sociological Association Conference	July 22, 2022
Co-Organizer. First <i>Summer Institute in Computational Social Science</i> in San Francisco Bay Area (BAY-SICSS). http://bit.ly/BAYSICSS-2020-results	2019–2020
Co-Organizer. CTAWG, D-Lab, UC Berkeley	2017–2020
Co-Organizer, Data Lead, and Cloud Support. <i>TextXD</i> , BIDS, UC Berkeley	2018, 2019
Co-Organizer. <i>Making Text Research-Ready</i> , Social Science Matrix, UC Berkeley	Apr. 30, 2018
Graduate Student Evaluator. <i>Sociological Research Symposium</i> , Sociology Dept., UC Berkeley	Apr. 20, 2018

Certificates

Certificate in Educational Finance. <i>Edunomics</i> , Georgetown University	2020
Waldorf Teaching Certificate. Bay Area Center for Waldorf Teacher Training	2017
Certificate of Teaching and Learning in Higher Education. UC Berkeley	2015

Professional Associations and Memberships

Board member, Sociology of Education Association (SEA)	2022–2024
Affiliate, Sociology Dept., Georgetown University	2021–present
<i>American Sociological Association</i>	2011–present
<i>Golden Key Honor Society</i>	2007–present
<i>National Society of Collegiate Scholars</i>	2004–present
<i>Mathematics, Science, and Engineering Achievement</i>	2003–present

Peer Review

<i>Sociology of Education</i>	2022
<i>Sociological Methods and Research</i>	2021
<i>Social Currents</i>	2021
<i>Journal of Research on Educational Effectiveness</i>	2020

Mentoring

Tong Xiang. Research apprentice, GU, 2021. Now PhD student, Computer Science, VA Tech.

Krutika Ingale. Research apprentice, UCB, 2018–19. Finished MS, Info. & Data Science, UCB

Jun Luo. Research apprentice, UCB, 2017. Finished MA, Public Policy, UCLA

Jin Kweon. Research apprentice, UCB, 2017. Now PhD student, Quant. Life Science, McGill U.

Emily Qian. Research apprentice, UCB, 2017. Finished MA, Finance, MIT Sloan School Mgt.

Giovanni Román-Torres. Honors mentee, UCB, 2015. Now PhD student in Sociology, U. Michigan

Ji Shi. Research apprentice, UCB, 2017–2019. Finished MS, Operations Research & Info.

Engineering, Cornell U.

Rudy Venguswamy. Research apprentice, UCB, 2019. Finished MS, Info. & Data Science, UCB

Jiahua Zou. Apprentice, UCB, 2017–2019. Finished MS, Bus. Analytics, College William & Mary

Programming Skills

Python: • Pandas • gensim • NLTK • spacy • scikit-learn • BeautifulSoup • scrapy • keras • selenium • matplotlib • geopandas • hdbscan • numpy • scipy

R: • stm • tidytext • dplyr

Other: • Stata • Git • Docker • Zotero • Tableau • bash • UNIX • XSEDE/Jetstream • MS Office

References

Michael A. Bailey

Director of the Massive Data Institute and the Data Science for Public Policy Program

Colonel William J. Walsh Professor of American Government

Department of Government and McCourt School of Public Policy

Georgetown University

ICC - Edward B. Bunn S.J. Intercultural Center 590

Washington, DC 20057

Email: Michael.Bailey@georgetown.edu

Lisa Singh

Professor

Department of Computer Science

Georgetown University

332 St. Mary's Building

Washington, DC 20057

Email: Lisa.Singh@georgetown.edu

Heather A. Haveman

Professor

Department of Sociology and Haas School of Business

University of California, Berkeley

410 Barrows Hall

Berkeley, CA 94720-1980

Email: haveman@berkeley.edu

Samuel R. Lucas

Professor

Department of Sociology

University of California, Berkeley

410 Barrows Hall

Email: Lucas@berkeley.edu

Calvin Morrill

Stefan A. Riesenfeld Professor of Law and Professor of Sociology

Jurisprudence and Social Policy and School of Law

University of California, Berkeley

2240 Piedmont Avenue

Berkeley, CA 94720-2150

Email: cmorrill@law.berkeley.edu