JAE YEON KIM

Overview

- Computational social scientist: using statistics and data science to solve public problems and strengthen civic engagement
- Research software developer: building tools that help data curation
- Domain expertise: political psychology, public opinion and election, race, inequality, immigration, social policy, and urban and local politics
- Data experience: survey, experimental, administrative, and text data

🎓 EDUCATION

2021

University of California, Berkeley
PhD in Political Science (**Co-Winner of the 2022 Best Dissertation Award, Urban and Local Politics Section, American Political Science Association**)

📍 Berkeley, California, USA

🏢 PROFESSIONAL EXPERIENCE

January 2023 | Present

Senior Data Scientist
Product and Technology Team   📍 Code for America
- Embedded in a team working with partner states to improve safety net experiences and outcomes

January 2022 | December 2022

Assistant Professor of Data Science
Core faculty, data science concentration
📍 KDI School of Public Policy Management
- Co-developed South Korea's first public policy-focused data science program from scratch
- Co-designed the data science concentration curriculum and set the standards for faculty hiring
- Designed and instructed R fundamentals for public policy and data communication and visualization

Fall 2020 | Present

Pre- and Postdoctoral Fellow (Currently: Affiliated Researcher)
P3 Lab, SNF Agora Institute   📍 Johns Hopkins University,
- Created the most comprehensive dataset on U.S. civil society (Mapping the Modern Agora project) with Milan de Vries (former Director of Analytics at MoveOn.Org)
- Developed open-source research software (MapAgora) that supports the data curation with Milan de Vries
- Analyzed the statistical relationships between civic opportunities and policy outcomes (e.g., vaccination)

Last updated on 2022-11-12.

CONTACT INFO

✉ jkim638@jhu.edu

🔗 jaeyk.github.io

🐙 github.com/jaeyk

in linkedin.com/jae-yeon-kim

📞 +1 510-646-5183

For more information, please contact me via email.

COURSEWORK

Statistical and Causal Inference, Experimental Design, Survey Methods, Game Theory, Computational Social Science

Passed Political Behavior (social and cognitive psychology, survey, and experimental design) field exam with distinction

SKILLS

🔢 **Quantitative**: Statistical and causal inference, Experimental and survey design

💻 **Computational**: Natural language processing, Machine learning, R (tidyverse, tidymodels, statistical packages), Python (pandas, scikit-learn), Git/GitHub, SQL (PostgreSQL), Linux Command Line

May 2019 | May 2021

Senior Data Science Fellow, Instructor, and Consultant

Data-intensive Social Sciences Lab   📍 UC Berkeley

- Consulted 60+ Berkeley faculty, students, and staff on applied data science and statistics
- Developed and taught six original data science workshops, ranging from programming fundamentals to machine learning
- Organized and streamlined consulting and workshop production

Spring 2020

Data Science Education Program Fellow

Data Science Education Program   📍 UC Berkeley

- Advised the undergraduate students and project partners involved in 40+ data science discovery projects
- Published an article on project management in SAGE Ocean, an initiative from SAGE Publishing focusing on computational social science

Fall 2016 | Present

Graduate Student Instructor

Department of Political Science   📍 UC Berkeley

- Taught graduate-level courses on computational tools for social science research as lead instructor
- Wrote an open-access online textbook on computational thinking for social scientists
- Taught an undergraduate-level applied statistics as a teaching assistant and received the Outstanding Graduate Student Instructor Award, which is given to less than 10% of Berkeley TAs

🖥 SOFTWARE

MapAgora: R package for getting tax reports, websites, and social media handles related to nonprofit organizations in the United States (with Milan de Vries)

tidytweetjson: R package for turning Tweet JSON files into a cleaned and wrangled dataset.

tidyethnicnews: R package for turning search results from one of the largest databases on ethnic newspapers and magazines published in the United States into a cleaned and wrangled dataset.

⚗ SELECTED PUBLICATIONS

COVID-19 and Asian Americans: How Elite Messaging and Social Exclusion Shape Partisan Attitudes [GitHub]

Co-author   📍 **Perspectives on Politics (2021)**

- Developed an R package that automates parsing a large Tweet JSON file (>5 GB) into a cleaned and wrangled dataset
- Applied keyword-based topic modeling to 1.4 million tweets and traced the rise of anti-Asian sentiment in the post-pandemic U.S.

- **Misinformation and Hate Speech: The Case of Anti-Asian Hate Speech During the COVID-19 Pandemic [GitHub]**
  Lead author   **📍 Journal of Online Trust and Safety (2021)**
  - Invited for the inaugural issue of the Journal of Online Trust and Safety hosted by Stanford's Internet Observatory
  - Applied embedding regression methods to 50k+ U.S. tweets related to COVID-19 and anti-Asian hate and found the tie between hate speech and misinformation

- **Intersectional Bias in Hate Speech and Abusive Language Detection Datasets [GitHub]**
  Lead author
  📍 The International AAAI Conference on Web and Social Media Data Workshop (2020)
  - Classified gender, racial, and party identities of the 100k tweets
  - Demonstrated African American tweets were up to 3.7 times more likely to be labeled as abusive, and African American male tweets were up to 77% more likely to be labeled as hateful compared to the others

👥 COMMUNITY BUILDING

- **The First Summer Institute in Computational Social Science in South Korea (co-hosted by KAIST and KDI School)**
  Co-organizer   📍 November 2021 - June 2022
  - Reviewed 60+ applicants and selected 18 participants
  - Designed the curriculum and guided the project development

- **The First Summer Institute in Computational Social Science in the San Francisco Bay Area (co-hosted by UC Berkeley and Stanford) [Berkeley Institute of Data Science blog post]**
  Co-organizer   📍 August 2019 - July 2020
  - Raised 50k+, reviewed 100+ applicants, and selected 20 participants
  - Developed close partnerships with Bay Area nonprofits (e.g., Code for America, DonorsChoose, Hopelab)
  - Designed the curriculum and guided the project development