

JAE YEON KIM

Overview

I study political learning, organizing, and mobilization among marginalized populations using big data and data science. I also build tools that make social science research more efficient and reproducible. I have extensive experience analyzing survey, experimental, administrative, and text data using statistical and machine learning methods



EDUCATION

- 2016**
|
2021
(expected)
University of California, Berkeley
PhD Candidate in Political Science - Concentrations: computational political science and political behavior (passed exam with distinction)
📍 Berkeley, California, USA
- Summer 2019**
Summer Institute in Computational Social Science
Participant (10% acceptance rate)
📍 Princeton University, Princeton, USA
- 2014**
|
2016
University of California, Berkeley
MA in Political Science
📍 Berkeley, California, USA
- 2012**
Korea University
BA in English & Political Science (graduated with Great Honors; study abroad: Hong Kong, Taiwan)
📍 Seoul, South Korea



PROFESSIONAL EXPERIENCE

- September 2020**
|
Present
Visiting Fellow
[P3 Lab, SNF Agora Institute](#) 📍 Johns Hopkins University,
 - Collaborated with Milan de Vries (former Director of Analytics at MoveOn.Org)
 - Parsed more than 3 million tax reports ([XML files stored in Amazon AWS](#)) filed by nonprofits in the United States and linked this data source with these organizations' websites and social media handles
 - Designed and built a data infrastructure that tracks civic organizations in the US and their relationships with food security, polarization, and the 2020 racial justice movement
- May 2019**
|
Present
Senior Data Science Fellow, Instructor, and Statistical Consultant
[Data-intensive Social Sciences Lab](#) 📍 UC Berkeley
 - Created a Shiny dashboard for the [D-Lab's Online Hate Speech Index Project](#)
 - Consulted 70+ Berkeley faculty, students, and staff on applied statistics, machine learning, and database management
 - Developed and taught [seven original workshops](#) on machine learning, functional programming, package development, advanced data wrangling, and reproducible project management

Last updated on 2020-12-24.

CONTACT INFO

✉ jaeyeonkim@berkeley.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

SKILLS

📊 **Quantitative:** Statistical modeling (OLS, GLM, Multilevel Modeling), Causal inference (a.k.a. impact evaluations or A/B testing), Experimental and survey design and analysis, Measurement (constructing and testing measures)

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

Fall 2019
|
Spring
2020

Data Science Project Manager

Data Science Discovery Program

📍 UC Berkeley

- Hired and managed four undergraduate RAs, set goals, plans, and deadlines, and helped these RAs' professional development
- Published [one paper](#) in *Proceedings of the Fourteenth International Conference on Web and Social Media (ICWSM)*, *Data Challenge Workshop* and [one preprint](#), which is under review at the *American Political Science Review*

Spring
2020

Data Science Education Program Fellow

Data Science Education Program

📍 UC Berkeley

- Served as research lead for the undergraduate students and project partners involved in 40+ [data science discovery projects](#)
- Published a [blog post](#) on project management published 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 an original graduate-level course on [computational tools for social science research](#) as an instructor and an undergraduate-level course on applied statistics as a teaching assistant
- Created an [open-access textbook](#) titled "Computational Thinking for Social Scientists"
- Received [the Outstanding Graduate Student Instructor Award](#), which is given to less than 10% of Berkeley TAs



PEER-REVIEWED PUBLICATIONS

Kim J. (2020). [Integrating Human and Machine Coding to Measure Political Issues in Ethnic Newspaper Articles](#). Forthcoming at the *Journal of Computational Social Science*. [[GitHub](#)]

- Winner of the **Don T. Nakanishi Award for Distinguished Scholarship and Service in Asian Pacific American Politics**, Western Political Science Association (2020)

Kim J, Carlos Ortiz, Sarah Nam, Sarah Santiago, and Vivek Datta. (2020). [Intersectional Bias in Hate Speech and Abusive Language Datasets](#). 2020. *Proceedings of the Fourteenth International Conference on Web and Social Media, Data Challenge Workshop* [[GitHub](#)]

Kim J. (2020). [How Other Minorities Gained Access: The War on Poverty and Asian American and Latino Community Organizing](#). Forthcoming at *Political Research Quarterly*. [[GitHub](#)]



PREPRINTS

Chan N, Jae Yeon Kim, and Vivien Leung. (2020). [COVID-19 and Asian Americans: How Social Exclusion Shapes Asian American Partisanship](#). *SocArXiv*. [[GitHub](#)]

Kim J and Andrew Thompson. (2020). [How Threats Shape the Politics of Marginalized: Evidence from a Natural Experiment and Machine Learning](#). *SocArXiv*. [[GitHub](#)]

Kim J and Alan Yan. (2020). [The Two Faces of Linked Fate: How Survey Respondents Interpret Linked Fate Question](#). *SocArXiv*. [[GitHub](#)]



SOFTWARE

tidytweetjson: R package for turning Tweet JSON files into a tidy dataframe. The package takes 12.34 seconds to wrangle 5,685 articles.

tidyethnicnews: R package for turning search results from one of the largest databases on ethnic newspapers and magazines published in the U.S. into a tidy dataframe. The package takes 4 minutes to wrangle 2 million tweets.

TidyChaseBankStatements: R package for turning Chase bank Statements into a tidy dataframe.