Sarah Kathryn Stein

345 Franklin St. #104
Cambridge, MA 02139

skstein2@gmail.com
(303) 817-7196
LinkedIn

Skills

- Statistical Packages: R (advanced), Stata (advanced), SAS (basic)
- Programming Languages: Python (intermediate), Java (intermediate), SQL (basic)
- GIS: ArcGIS (advanced), R (advanced)
- Big Data Tools: web scraping, natural language processing (NLP), MapReduce, AWS
- Other Competencies: data visualization, experimental design, survey design (Qualtrics), MTurk

Education

Stanford Graduate School of Business — Stanford, CA
Ph.D. in Organizational Behavior — *Sept 2013 – June 2018*

- Dissertation - *Distinguishing Round From Square Pegs: Understanding Hiring Based on Pre-hire Language Use*
- Merged job applications to HR data in Python using fuzzy name matching
- Predicted job candidates' hiring outcomes and cultural fit using their pre-hire language
- Quantified the cultural differences between hired and not hired job candidates using NLP
- Programmed the R-package RCA (Relational Class Analysis)
- Designed, ran, and analyzed experiments using Qualtrics and MTurk
- Scraped and cleaned data from public websites using Python

Brown University — Providence, RI
B.A. with Honors in Economics; GPA: 4.00 — *Sept 2006 – May 2010*

- Honors Thesis - *Salt and the Timing of the Neolithic Revolution*
- Combined raster elevation data & coastline shapefiles to determine average coastal gradients
- Samuel C. Lamport Prize - awarded for the best thesis on international understanding
- Phi Beta Kappa

Experience

Data Analyst — San Francisco, CA
Structural Engineers Association of Northern California — *May 2016 - Sep 2016*

- Analyzed a large nationwide survey on gender dynamics among structural engineers
- Wrote and presented explanations of the results that were accessible to non-experts

Research Assistant — Stanford, CA
Stanford University Department of Political Science — *May 2015 - May 2016*

- Created boundary shapefiles from scanned historical maps of Africa using ArcGIS
- Geocoded school locations & calculated the distance between survey clusters and schools in R
- Created an area-weighting methodology to translate contemporary data to historical districts
- Visualized educational, religious, and ethnographic data using graduated color maps

Research Assistant Princeton, NJ
Princeton University Department of Politics *Feb 2012 - May 2013*

- Collected and cleaned Census data (>80GB) using MS Access and Stata
- Merged Census data to ArcGIS shapefiles to create demographic visualizations
- Added demographic shapefiles to Google Maps to facilitate door-to-door data collection

Research Associate/Assistant Economist New York, NY
Federal Reserve Bank of New York *Jul 2010 – Jun 2013*

- Built a spatial algorithm to match land parcels across borders using Stata, Python & ArcGIS
- Geocoded real estate sales & created heat map visualizations of local markets in R
- Cleaned a quarterly panel of U.S. credit reports (>1TB) using MapReduce
- Estimated diff-in-diff and regression discontinuity (RD) models in SAS and Stata
- Created a spreadsheet model of FX bid and ask spreads
- Aggregated, analyzed, and visualized data on the FX reserves of industrial countries
- Wrote documentation and created process diagrams for all on-going projects

Ph.D. Research

Distinguishing Round from Square Pegs

Predicting Hiring, Performance, and Cultural Fit from Pre-Hire Free-Response Text

- Co-authors: Amir Goldberg, Sameer Srivastava
- This article examines how cultural matching relates to a job applicant's likelihood of getting hired into an organization and identifies the components of cultural similarity that matter most for hiring success. Cultural compatibility at the hiring stage can forecast an individual's post-hire productivity but is difficult to reliably measure in the selection process. As a consequence, cultural matching is often subject to various informational and identity-based biases. We develop a language-based model that provides a means for directly assessing job candidates' cultural similarity. Based on variegated data from a mid-sized technology firm—including job applicants' free text responses at the pre-hire stage, applicant characteristics, applicant-interviewer assignments, and hiring outcomes—we find that linguistic similarity with previously hired employees increases a job candidate's chances of being hired, even after controlling for the applicant's human and social capital. We further find that, although all three forms of cultural fit that we assess—fit based on work preferences, lifestyles, and ideology—predict hiring in between-interviewer models, only work preferences fit predicts hiring in within-interviewer models. Supplemental analyses indicate that pre-hire cultural fit is also predictive of successful enculturation in the firm over the first six months of employment. Together, these results indicate that cultural matching leads to sorting on attributes that are both relevant and potentially irrelevant for job success.

Separating the Wheat from the Chaff

Examining Pre-hire Language Use by Hiring Outcome

- Co-authors: Amir Goldberg, Sameer Srivastava
- Although it is generally acknowledged that cultural matching plays an important role in the hiring process, the relevant cultural characteristics are often nuanced and can vary widely from organization to organization. Because of these intricacies, researchers have typically examined cultural matching using qualitative interviews, surveys, and participant observation. However, with the introduction of linguistic techniques to the study of organizations and culture, natural language processing has become an increasingly viable method for examining cultural preferences and experiences. As such, this study leverages these new techniques to

provide a detailed accounting of the linguistic differences separating hired and not hired job candidates at a mid-sized technology firm. Specifically, we develop a set of diverse analyses across four linguistic slices aimed at elucidating the cultural elements relevant to this firm's hiring process. Moreover, we further distinguish between three dimensions of cultural fit—work preferences, lifestyles, and ideologies—when quantifying differences between the hired and not hired groups. Ultimately, we find evidence that hired candidates differ meaningfully from their not hired counterparts on a number of cultural dimensions including: (1) their focus on job characteristics vs. cooperation/teamwork, (2) their interest in outdoors activities—particularly camping, hiking, and biking, and (3) their attention to broad, abstract ideological concerns vs. narrow, concrete ones.