Kasey Zapatka, PhD

San Francisco, CA | kaseyzapatka@berkeley.edu | www.kaseyzapatka.com | kaseyzapatka
kaseyzapatka | 0000-0002-7278-3660

Education

PhD **City University of New York, The Graduate Center**, Sociology — New York City, NY, Aug 2016 – Feb 2023

MA **Fordham University**, Sociology — Bronx, NY, Aug 2014 – May 2016

BA **Point Loma Nazarene University**, Spanish Literature and Language — San Diego, CA, Aug 2006 – May 2010

Experience

Clean Air Task Force, Consultant – NEPA Clean Energy Review Analysis — Remote, Jan 2026 – present, 5 months

- Built a BERT+LLM pipeline to extract and classify records from 120,000+ NEPA documents, enabling CATF to analyze clean energy review timelines at scale
- Integrated Federal Register API data, targeted web scraping, and ML/regex classifiers to categorize 20,000+ projects by technology, review process, and capacity
- Scoped research questions collaboratively with CATF and delivered phased findings across six structured deliverables mapped to specific policy priorities
- Published findings via public project website and HuggingFace document browser

University of California, Berkeley, Postdoctoral Researcher and Lecturer — Berkeley, CA, July 2023 – present, 2 years 11 months

- Lead cross-functional research team to develop, maintain, and analyze eviction database (150M+ records) using SQL
- Applied R-based machine learning, multilevel regression, and synthetic control methods to study impact of rental assistance on eviction filings
- Coordinated with UPenn partners, managed research process, and led report writing
- Presented findings and strategic recommendations to federal housing policy stakeholders
- Taught machine learning, natural language processing, and causal inference to PhD students
- Led 32 weekly labs in R and Python libraries (tidyverse, sklearn, TensorFlow)
- Supervised 8 student projects on predictive modeling and decision-making

NYC Department of Housing Preservation & Development, Data Scientist — New York City, NY, Sept 2022 – June 2023, 10 months

- Collaborated with US Census research team to establish sample survey research frame for Housing Vacancy Survey
- Addressed Housing Vacancy Survey measurement error by creating post-stratification weights
- Developed methodology with survey weights to determine where to build affordable housing, presented to non-technical stakeholders, and persuaded Commissioner to adopt methodology

City University of New York, The Graduate Center, Research Scientist, Professor Van Tran — New York City, NY, Jan 2021 – Dec 2022, 2 years

- Built US Census dataset to study ethno-racial neighborhood integration in metro New York
- Fitted multinomial logistic regression models and interpreted marginal effects plots in R
- Wrote methodological sections, created descriptive plots in R, and managed project development on GitHub

City University of New York, The Graduate Center, Research Scientist, Professor Paul Attewell — New York City, NY — Oct 2018 – May 2021 — 2 years 8 months

- Developed research projects to study divergence of benefits for groups with different educational attainment
- Managed 4 large national, longitudinal, and cross-sectional datasets used in analyses
- Used data mining, HLM, OLS, and logistic regression techniques to understand the impact of educational attainment on mid-life labor market earnings

CUNY, Center for Urban Research, UX Researcher — New York City, NY — Sept 2016 – Jan 2019 — 2 years 5 months

- Designed, executed, and led focus group research to identify key labor market trends across 10 different projects
- Distilled findings into actionable insights and presented recommendations to clients

Empirical Creative, UX Researcher — New York City, NY — Nov 2014 – Aug 2016 — 1 year 10 months

- Designed and led mock jury trials (focus groups) to identify favorable juror characteristics and attitudes
- Conducted strategic research to understand user characteristics that shape case opinions and what facts change decisions
- Translated insights into actionable recommendations — trial strategy, user profiles, and visual graphics

Max Planck Institute, Data Scientist – Superdiversity in Metro New York Project — Remote — Jan 2020 – Dec 2021 — 2 years

- Managed end-to-end project development of interactive Superdiversity Website and Teaching Tool
- Built and analyzed 6 cross-sectional, longitudinal databases exploring changing diversity in Metro New York
- Supervised design team and ensured project aligned with stakeholder vision

Selected Projects

NEPA Clean Energy Environmental Review Analysis — 2026–present

- Trained BERT classifiers and LLM adjudication layer to extract review dates and process types from 120,000+ unstructured federal documents
- Integrated Federal Register API, web scraping, and ML/regex classifiers to categorize 20,000+ projects by technology, capacity, and review type
- Published findings via a public project website and an interactive HuggingFace document browser, making results accessible to policymakers and researchers

Voter Turnout in New York City — 2025–2026

- Used Bayesian Improved Surname Geocoding (BISG) to impute race/ethnicity from surnames in 4.6 million NYC voter registration records
- Applied multilevel regression and poststratification (MrP) to impute educational attainment and build survey post-stratification weights
- Applied RAG techniques and fine-tuned open-source LLMs to extract insights from voter file records

SF Residential Inspection Risk — Jan 2025

- Linked fire incidents, building violations, inspection records, and parcel tax data at the parcel level using APN identifiers
- Built composite risk scores and k-means spatial clusters to prioritize inspections for SFFD and the Department of Building Inspection
- Deployed two interactive Shiny dashboards: SFFD, DBI

Superdiversity in Metro New York — 2020–2021

- Built 6 cross-sectional longitudinal databases and led end-to-end development of interactive website and teaching tool

Housing Literacy 2017–2019

- Developed online tool annotating NYC rent regulation legal documents for rent-stabilized tenants and housing advocates

Publications

New Frontiers of Integration: Convergent Pathways of Neighborhood Diversification in Metropolitan New York Feb 2023
Kasey Zapatka, Van C. Tran
10.7758/RSF.2023.9.1.03 (RSF: The Russell Sage Foundation Journal of the Social Sciences)

Affordable Regulation: New York City Rent Stabilization as Housing Affordability Policy Oct 2022
Kasey Zapatka, Juliana de Castro Galvao
10.1177/15356841221123762 (City & Community)

Superdiversity in Metropolitan New York: Technical Report June 2021
Kasey Zapatka
www.superdiv-newyork.mmg.mpg.de (Max Planck Institute)

Reordering Occupation, Race, and Place in Metropolitan New York Jan 2021
Kasey Zapatka
(Springer International Publishing)

Does Demand Lead Supply? Gentrifiers and Developers in the Sequence of Gentrification, New York City 2009–2016 Aug 2020
Kasey Zapatka
(Urban Studies)

Skills

Programming: R, Python, SQL, Stata, Git & GitHub, Linux

Quantitative Methods: Spatial econometrics, Machine learning, Generalized linear regression, Natural language processing, Survey analysis, LLMs and GenAI, A/B testing

Qualitative Methods: Focus groups, Interviews, Stakeholder analysis, Survey design, User perception studies

Languages: English (native), Spanish (fluent)