

Jacob Kaplan, Ph.D.

Email: jk7785@princeton.edu
Website: jacobdkaplan.com
GitHub: github.com/jacobkap

Professional Profile

I'm a Professional Specialist at Princeton University with a Ph.D. in criminology from the University of Pennsylvania. My work sits at the intersection of data science and public policy, focusing on automating the cleaning and processing of large administrative police datasets using large language models (LLMs). I am also leading a project that uses a large-scale geospatial mobility dataset (~15TB) to study how homelessness impacts voting behavior. I primarily work in R and have developed several open-source R packages. I'm the author of *A Criminologist's Guide to R: Crime by the Numbers* (CRC Press), a practical programming guide for working with crime data. To support broader data accessibility, I built the [Crime Data Tool](#)—an interactive platform for visualizing trends in crime, arrests, and police employment across jurisdictions. My work emphasizes reproducibility, transparency, and making complex data usable for both researchers and non-technical users.

Technical Skills: R (10+ years), Git, JavaScript, Stata, LaTeX

Selected Data Science Projects

Crime Data Tool | CrimeDataTool.com

- Developed an interactive web-based dashboard for visualizing public safety data across time and agencies, enabling non-technical users to explore trends in crime, arrests, and police employment using open data.
- Developed the website using JavaScript and HTML, with data processing in R.

Uniform Crime Reporting (UCR) Program Data | [Harvard Dataverse](#) (~250 academic citations)

- Cleaned, standardized, and published FBI Uniform Crime Reporting (UCR) data as R and Stata files, improving accessibility, reproducibility, and data transparency for researchers worldwide.

fastDummies | [CRAN](#) (2.2 million downloads)

- Developed an R package for quick and efficient creation of binary columns from categorical columns, and dummy rows based on all combinations of categorical and date columns.

Education

Ph.D. in Criminology	University of Pennsylvania	2016-2020
M.S. in Criminology	University of Pennsylvania	2015-2016
B.S. in Criminal Justice	California State University, Sacramento	2011-2015

Experience

Professional Specialist Princeton University 2021-Present

- Used geospatial mobility data (~15TB of data) to measure how homelessness in a city affects voting behavior.
- Built automated NLP pipelines using large language models (LLMs) to classify and cluster thousands of unique strings, enabling scalable and reproducible workflows for large-scale police data.

- Cleaned and harmonized multi-agency administrative data from 99 police departments; performed probabilistic record linkage with a voter registration database (~200M rows) using custom ETL workflows.
- Conducted code reviews for team members to ensure coding accuracy and adherence to style guidelines.

Committee Member **Federal Bureau of Investigation –
Uniform Crime Reporting Subcommittee** **2022**

- Served as one of 17 voting members, providing subject matter expertise to inform policy.
- Reviewed, advised, and voted on potential changes to the FBI’s crime data collection, focusing on a balance between the usefulness of the data and the burden on agencies to generate it.

Postdoctoral Fellow **University of Pennsylvania** **2020-2021**

- Embedded researcher at the Philadelphia District Attorney’s Office, analyzing the life cycle of gun cases from arrest to final disposition.
- Spearheaded a privacy-focused project determining optimal anonymity levels for safely releasing datasets containing Personal Identifiable Information (PII).

Selected Publications

A Criminologist’s Guide to R: Crime by the Numbers | [Chapman & Hall/CRC The R Series](#)

- Covers full data science workflow: data ingestion, cleaning, exploratory analysis, mapping, and automated reporting, with real-world public safety examples. Teaches tools for reproducible research, tidyverse, web scraping, and geospatial analysis.
- Includes lessons on ggplot2, dplyr, tidyr, readr, stringr, regular expressions through grep and gsub, sf, leaflet, rvest, pdftools, and tidygeocoder.

Ba, B., Ge, H., **Kaplan, J.**, Knox, D, Komisarchik, M., Lanzalotto, G., Mariman, R., Mummolo, J., Rivera, R., & Torres, M. (2025). **Political diversity in U.S. police agencies**. *American Journal of Political Science*. Online First. <https://doi.org/10.7910/DVN/CZOPH3>

- Compared whether police officers in 99 of the largest 100 agencies are similar to civilians in their jurisdiction in terms of demographics, socioeconomic factors, and political affiliation. Measured whether officer demographics affect behavior.
- We found that most police agencies are different than their community on a number of factors, and that officer outcomes differ by officer race, but not by political party.

Block, K., & **Kaplan, J.** (2023). **An analysis of National Hockey League playoff games and city level crime counts**. *Crime & Delinquency*, 69(11), 2194-2217. <https://doi.org/10.1177/00111287221118879>

- Analyzed crime data from 15 cities to study whether crime changed when an NHL playoff game occurred at home compared to away.
- Found that home games are associated with 7% more disorder and 4% more property crimes.
- Selected coverage: [New York Times](#)