Jacob Kaplan, PhD

Email: jk7785@princeton.edu
Websites: jacobdkaplan.com & crimedatatool.com
GitHub: github.com/jacobkap

Professional Profile

Data Scientist with a Ph.D. in Criminology, possessing extensive experience in managing and analyzing large, complex data. I am an expert in data cleaning, analysis, and visualization, and have developed user-friendly tools, such as the Crime Data Tool, and built several R packages, demonstrating my ability to make data analysis accessible to various audiences. My academic portfolio includes 14 peer-reviewed publications and a textbook that introduces the programming language R. My unique blend of technical expertise, research acumen, and communication skills underscores my commitment to transforming complex data into valuable insights.

Education

PhD in Criminology	University of Pennsylvania	2016-2020
MS in Criminology	University of Pennsylvania	2015-2016
BS in Criminal Justice	California State University, Sacramento	2011-2015

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

Experience

Professional Specialist **Princeton University** **2021-Present**

- Developing a toolkit to measure racial bias in law enforcement among traffic stops.
- Designed, tested, and deployed automated workflows for categorizing and clustering thousands of unique strings using large language models.
- Collaborated on a cross-disciplinary project comparing sociodemographic and political attributes of police to civilians in corresponding jurisdictions.
- Managed and cleaned data from 99 police agencies, performing probabilistic matching against a massive-scale (200+ million rows) voting database.
- 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 Data Science Projects

Crime Data Tool | crimedatatool.com

- Built an interactive platform enabling non-programmers to access government crime data.
- Simplified data processing workflows and provided customizable visualizations and downloadable data for users.
- Developed the tool using JavaScript and HTML, with data processing in R.

Uniform Crime Reporting (UCR) Program Data | Harvard Dataverse

- Cleaned, standardized, and published FBI UCR data as R and Stata files, making the dataset accessible to researchers worldwide.
- Has over 250 citations in peer-reviewed journal articles.
- Standardizes column names and values, resolved data errors, and accounts for naming convention changes over time.

fastDummies | **CRAN** (2.1 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.

Selected Publications

A Criminologist's Guide to R: Crime by the Numbers | Chapman & Hall/CRC The R Series

- Authored an introductory textbook on the R programming language, combining crime data examples with practical programming techniques.
- Covers data gathering, cleaning, visualization (through graphs and maps), project management, and introduces the concept of *mise en place* in the context of project planning.
- 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. "**Political diversity in U.S. police agencies."** (Accepted at *American Journal of Political Science*.)

- 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.
- Compared officer behavior on these traits in Houston and Chicago to measure if they behave differently.

Block, K., & **Kaplan, J.** (2022). **An analysis of National Hockey League playoff games and city level crime counts.** *Crime & Delinquency.* 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