Jacob Kaplan, PhD (Current as of May 2023)

Websites: jacobdkaplan.com crimedatatool.com
GitHub: https://github.com/jacobkap
Email: jkkaplan6@gmail.com
LinkedIn: https://www.linkedin.com/in/jacob-kaplan1/

Professional Profile

Data Scientist with a PhD in Criminology, possessing extensive experience in managing and analyzing large, complex data. I am proficient in data analysis, cleaning, 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 12 peer-reviewed publications as well as practical guides for data usage and programming. My unique blend of technical expertise, research acumen, and effective communication skills, combined with my experience in handling big data and building R packages, underscores my commitment to transforming complex data into valuable insights.

Technical Skills: R (8+ years), Git, JavaScript, HTML, CSS, Stata, LaTeX, Microsoft Office

Experience

2021-Present Professional Specialist Princeton University
Chief Data Scientist Research on Policing Reform and Accountability (RoPRA)

- Collaborated on a cross-disciplinary project comparing sociodemographic and political attributes of police to civilians in corresponding jurisdictions.
- Managed and cleaned data from 98 police agencies, performing probabilistic matching against the massive-scale (200+ million rows) L2 voting database.
- Conducted code reviews for team members to ensure coding accuracy and adherence to style guidelines.
- Developing a toolkit to measure racial bias in law enforcement at various stages of police-citizen encounters, including stops, arrests, and use of force. Cleaning and standardizing data from several different police agencies.

2020-2021 Postdoctoral Fellow University of Pennsylvania

- Embedded researcher at the Philadelphia District Attorney's Office.
- Spearheaded a privacy-focused project determining optimal anonymity levels for safely releasing datasets containing Personal Identifiable Information (PII).
- Established robust data cleaning and preprocessing pipelines to manage and analyze large-scale criminal justice datasets.
- Analyzed the life-cycle of gun cases from arrest to final disposition, studying if and how changes in policy affected these cases.

Selected Data Science Projects

Crime Data Tool | https://crimedatatool.com

- Developed an interactive tool enabling non-programmers to access government crime data, built entirely in JavaScript and HTML, with data gathered and processed in R.
- Facilitated selection of specific datasets, agencies, variables, and presentation preferences to

generate downloadable tables and time-series graphs.
- Made the entire data freely downloadable on openICPSR for users with programming skills (link to a list of datasets is here; data has been used in over 100 peer-reviewed papers).

A Criminologist's Guide to R: Crime by the Numbers | https://crimebythenumbers.com/
- Authored a comprehensive guide to the R programming language published by Chapman & Hall/CRC The R Series, with examples from criminology to cater to all readers.
- Covered data gathering, cleaning, data visualization (through graphs and maps), project management in R and RStudio, and introduced the concept of *mise en place* in the context of project planning.

Uniform Crime Reporting (UCR) Program Data: A Practitioner's Guide
National Incident-Based Reporting System (NIBRS) Data: A Practitioner's Guide
https://ucrbook.com/ | https://nibrsbook.com/
- Created user-friendly guides to major FBI crime and arrest datasets, detailing their proper usage and potential pitfalls.
- Targeted at a layman's audience, with an emphasis on enabling users without extensive data or programming skills to effectively use the data.

fastDummies | https://cran.r-project.org/web/packages/fastDummies/index.html
- 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.
- To date, it has achieved 1.3 million downloads and is used in 31 other R packages.

asciiSetupReader | https://cran.r-project.org/web/packages/asciiSetupReader/index.html
- Developed an R package for reading fixed-width ASCII files with .sps or .sas setup files.
- The package has received 40,000 downloads.

Education

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