JONATHAN KROPKO

Phone: (440) 227-0386
jkropko@virginia.edu
ORC-ID: 0000-0003-3403-7648
Google Scholar: https://scholar.google.com/citations?user=wFgLdqUAAAAJ

1366 Stoney Ridge Rd
Charlottesville, VA 22902

ACADEMIC POSITIONS

Associate Professor, General Faculty Teaching Track 2024-present
School of Data Science
University of Virginia

Director, Master's of Data Science, Online 2022-present
School of Data Science
University of Virginia

Assistant Professor, General Faculty Teaching Track 2019-2024
School of Data Science
University of Virginia

Assistant Professor 2013-2019
Department of Politics
University of Virginia

Postdoctoral Fellow 2011-2013
Applied Statistics Center
Columbia University
Supervisors: Andrew Gelman, Jennifer Hill

EDUCATION

PhD University of North Carolina at Chapel Hill, Political Science May 2011
Dissertation: "New Approaches to Discrete Choice and Time Series
    Cross Section Methodology for Political Research"
Committee: George Rabinowitz (chair), John Aldrich (Duke University),
    Thomas Carsey, Skyler Cranmer, Justin Gross,
    and Michael Mackuen

MA University of North Carolina at Chapel Hill, Political Science May 2009

BS Ohio State University, Mathematics, *with honors* June 2005

BA Ohio State University, Political Science, *with distinction* June 2005

HONORS AND AWARDS

Quantitative Foundation Endowed Chair   2024-present
School of Data Science
University of Virginia

PhD Program Teaching Award   2024
School of Data Science
University of Virginia

Partnership of the Year Award   2022
Awarded to Code for Charlottesville
Charlottesville Business Innovation Council

All-University Teaching Award   2018
University of Virginia

Excellence in Graduate Student Mentorship Award   2015
Society for Political Methodology

TEACHING AND MENTORING EXPERIENCE

University of Virginia, Charlottesville, VA, USA   August 2019 to present
Assistant Professor, School of Data Science/Data Science Institute
- DS 6600: Data Engineering 1 (Ph.D.)
    - Developed for the launch of the Ph.D. in data science program in Fall 2022
    - Python-based
    - Covers:
        - Virtual environments, containers, virtual machines
        - Workflow and remote code repositories
        - Data ingest using flat files, APIs, and scraping
        - Data wrangling
        - Database design using relational, graph, and document store schemas
        - Data visualization and dashboards
- DS 6001: Practice and Applications of Data Science 1 (residential and online)
    - Wrote a companion textbook: "Surfing the Data Pipeline with Python"
    - Master's level and Python-based
    - Covers:
        - Data ingestion: from CSVs and flat files, JSONs, APIs, web scraping
        - Databases: MySQL, PostgreSQL, SQlite, MongoDB
        - Data wrangling: SQL, Mongo DB's query language, and pandas
        - Data exploration and visualization: matplotlib, seaborn, plotly, and dashboards
    - Generated labs, recorded lectures, and quizzes in an active learning framework

- DS 6011/6013: Masters of Data Science Capstone Projects (residential and online)
    - "Covering the Housing Affordability Crisis in Charlottesville, VA: A Case Study in Open Data and Data Journalism". Spencer Bozsik, Xinlun Cheng, Malvika Kuncham and Evan Mitchell. April 2022.
        - Winner, Best Paper Award, Policy Track, IEEE Systems and Information Engineering Design Symposium, 2022
    - "Longitudinal Classification and Predictive Modeling for Historical CPS Data Using Random Forests". Cecile Johnson and Hannah Schmuckler. April 2022.
    - "Investigating the Illicit Trade of Cultural Property with an Automated Data Pipeline Architecture". Felipe Barraza, Nicholas Landi, Elizabeth Lee, and Karolina Naranjo-Velasco. April 2022.
    - "noderank: An R package for differential gene expression analysis and method benchmarking". Taylor Derby, Brad Howlett, Monish Dadlani, Patrick Chatfield. August 2021.
    - "Incorporation of Publicly Available Data for Improvement of Insurance Loss Prediction". Brittany Durkin, Aaron Oliver, Ali Moradkhany. August 2021.
    - "Fair Location-Based Clustering for General Liability Rate Making Models". Jon Baynes, Luke Haushalter, Chas Smith, Natalie Zimmer. August 2021.
    - "Measuring and Predicting Digital Force Applied Towards the Public Through Police Surveillance Technologies". Sarah Adams, Claire Setser. August 2021.
    - "Detecting Infrastructure Damage using Satellite Imagery and Neural Networks". Maurizio Boano, Sam Tyree, Travis Moore, Anvar Sarbanov. August 2021.
    - "Towards Automating Search and Classification of Protostellar Images". Pavan Kumar ; Pengwei Hu; Shannon E ; John Zhang. April 2021.
    - "Multi-Output Random Forest Regression to Emulate the Earliest Stages of Planet Formation". Kevin Hoffman; Jae Yoon Sung; André Zazzera. April 2021.
    - "Assessing Infrastructure Damage with Satellite Images". Will Carruthers; Jordan Bales. December 2020.
    - "An Analysis of COVID-19 Spread in the U.S. and Potential Vaccination Effects" Brooke Williams; Kevin Lennon; Cory Yemen. December 2020.
    - "Forecasting the 2020 Presidential Election: a Comparison of Methods" Matt Thomas; Chad Sopata; Ben Rogers; Spencer Marcuso. December 2020.
    - "A Text Analysis of the 2020 U.S. Presidential Election Campaign Speeches". Kevin Finity; Ramit Garg; Max McGaw. December 2020.
    - "Understanding Public AttitudesToward COVID-19 with Twitter". Jae Hyun Lee; Jason Lwin; Cullen Baker. December 2020.
    - "Lessons Learned: A Case Study in Creating a Data Pipeline using Twitter's API" Jason Tiezzi; Rice Tyler; Suchetha Sharma. April 2020.
    - "The Deployment of a LoRaWAN-Based IoT Air Quality Sensor Network for Public Good". James Montgomery Howerton; Benjamin Leo Schenck. April 2020.

o "A Digital Green Thumb: Neural Networks to Monitor Hydroponic Plant Growth" Mark L Tenzer; Nicholas C Clifford. April 2020.

Assistant Professor, Department of Political Science   August 2013 – August 2019
- PLAD 7100/7500: Mathematics for Social Scientists (Doctoral level)
  - Precalculus, probability, differential, integral, and multivariate calculus, and linear algebra
- PLAD 8310: Linear Regression and Data Management (Doctoral level)
  - Basic programming in R, data wrangling and visualization with the tidyverse packages, linear models, interactions and transformed models, robust estimation methods
- PLAD 8320: Generalized Linear Models (Doctoral level)
  - Foundations of likelihood-based statistics, optimizing custom-built likelihood functions in R, models for nonnormal, binary, ordinal, nominal, count, and survival dependent variables
- PLAD 7500: Time Series and Panel Statistics (Doctoral level)
  - ARIMA models, Granger causality and vector autoregressions, error correction models, fixed and random effects models, and programming and visualizing time series statistics in R
- PLAD 8500: Statistical Measurement (Doctoral level)
  - Exploratory and confirmatory factor analysis, item response theory and scaling methods, cluster analysis, and methods for analyzing social media
- PLAD 4500: Analysis of Political Data Using R (Undergraduate level)
  - Basic programming in R, data wrangling and visualization with the tidyverse packages, linear models, logistic regression, and text analysis

Instructor, U.S. Department of State Foreign Service Institute   2020
District Data Labs
- Programming in R (with Brian Wright)
  - Basic programming in R, data wrangling and visualization with the tidyverse packages, linear models, logistic regression, and text analysis

Instructor, University of Virginia Launchpad Summer Program   2020
- Python Programming
  - Data ingestion from CSVs and flat files, JSONs, APIs, web scraping, databases I ncludingMySQL, PostgreSQL, SQlite, and MongoDB, data wrangling using SQL, Mongo DB's query language, and pandas, and data exploration and visualization with matplotlib, seaborn, plotly, and dashboards

Instructor, Essex Summer School in Social Science Data Analysis   Summer 2015-2019
University of Essex, Colchester, UK
- Designing Your Own Maximum Likelihood Models
  - Foundations of likelihood-based statistics, optimizing custom-built likelihood functions in R, models for nonnormal, binary, ordinal, nominal, count, and survival dependent variables

Doctoral Committees
- Kimberly Ganczak, "Broadening the Political Methodologist's Toolkit:A Population Dynamics Model of Political Science Time Series Data", May 2021.
- Danilo Medeiros. "Extremism and Polarization: How Income Inequality Affects Legislative Behavior in Brazil". August 2019.
- Robert Kubinec, "Crony Capitalist, Democracy and the Arab Uprisings in North Africa", May 2018.
- Min-Gyu Paik, "Going Above and Beyond: Dependence and Military Coalition Participation", May 2016

PUBLICATIONS

Books

Kropko, Jonathan. 2015. *Mathematics for Social Scientists*. New York: Sage.

Journal Publications

Ali, Christopher, Hilde Van den Bulck, and Jonathan Kropko. 2025. "An Island of Trust: Public Broadcasting in the United States." *Journal of Communication*. Forthcoming. https://doi.org/10.1093/joc/jqaf009

Christ, Bryan R., Zachary Gottesman, Jonathan Kropko, Thomas Hartvigsen. 2025. "Math Neurosurgery: Isolating Language Models' Math Reasoning Abilities Using Only Forward Passes." ACL: Annual Meeting of the Association for Computational Linguistics. https://arxiv.org/pdf/2410.16930

Christ, Bryan R., Jonathan Kropko, and Thomas Hartvigsen. 2024. "MATHWELL: Generating Educational Math Word Problems Using Teacher Annotations." In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the Association for Computational Linguistics*: EMNLP 2024 (pp. 11914–11938). Association for Computational Linguistics. https://aclanthology.org/2024.findings-emnlp.696

Kropko, Jonathan, and Robert Kubinec. 2020. "Interpretation and Identification of Within-unit and Cross-sectional Variation in Panel Data Models." *PloS one.* 15(4). https://doi.org/10.1371/journal.pone.0231349

Kropko, Jonathan and Jeffrey J. Harden. 2020. "Beyond the Hazard Ratio: Generating Expected Durations from the Cox Proportional Hazards Model." *British Journal of Political Science*. 50(1): 303-320. https://doi.org/10.1017/S000712341700045X

Kropko, Jonathan and Jeffrey J. Harden. 2019. "coxed: An R Package for Computing Duration-Based Quantities of Interest from the Cox Proportional Hazards Model" *R Journal*. 11(2): 38-45. https://doi.org/10.32614/RJ-2019-042

Harden, Jeffrey J. and Jonathan Kropko. 2019. "Simulating Duration Data for the Cox Model." *Political Science Research and Methods*. 7(4): 921-928. https://doi.org/10.1017/psrm.2018.19

Kropko, Jonathan and Kevin Banda. 2018. "Issue Scales, Information Cues, and the Proximity and Directional Models of Voter Choice." *Political Research Quarterly*. 71(4): 772-787. https://doi.org/10.1177/1065912918760729

Kropko, Jonathan, Ben Goodrich, Andrew Gelman, and Jennifer Hill. 2014. "Multiple Imputation for Continuous and Categorical Data: Comparing Joint Multivariate Normal and Conditional Approaches." *Political Analysis*. 22(4): 497-519. https://doi.org/10.1093/pan/mpu007

Liu, Jingchen, Andrew Gelman, Jennifer Hill, Yu-Sung Su, and Jonathan Kropko. 2013. "On the Stationary Distribution of Iterative Imputations." *Biometrika*. 101(1): 155-173. https://doi.org/10.1093/biomet/ast044

Clark, Trenette T., Anh B. Nguyen, and Jonathan Kropko. 2013. "Epidemiology of Drug Use Among Biracial/Ethnic Youth and Young Adults: Results from a US Population-Based Survey." *Journal of Psychoactive Drugs*. 45(2): 99-111. https://doi.org/10.1080/02791072.2013.785804

Working papers

Christ, B. R.**,** Gottesman, Z., Kropko, J., & Hartvigsen, T. 2024. "*Math Neurosurgery: Isolating Language Models' Math Reasoning Abilities Using Only Forward Passes*" (arXiv:2410.16930). arXiv. https://doi.org/10.48550/arXiv.2410.16930

Ali, Christopher, Van den Bulk, Hilde, Kropko, Jonathan. 2025. "An Island of Trust: Public Broadcasting in the United States". Revise and Resubmit at *Journal of Communications*.

Additional Writings (Not peer reviewed)

Kropko, Jonathan. 2020. *Surfing the Data Pipeline with Python.* https://jkropko.github.io/surfing-the-data-pipeline/

SCHOOL AND UNIVERSITY SERVICE

- Public Interest Technology University Network steering committee
- Aperio (open access academic publishing) steering committee
- UVA SDS Ph.D., Academic Affairs, and Math Curriculum committees
- Public Service Pathways Steering Committee

EXTERNAL SERVICE

- Reviewer for Public Interest Technology Challenge, New America Foundation

GRANT ACTIVITY

- "The Interconnected Cosmos Initiative: Accelerating Discovery through Interdisciplinary Collaboration". Strategic Planning for Research, Scholarship, and Creative Activity in Arts & Sciences at the University of Virginia. Member, steering committee.

SOFTWARE

- CoxED: methods for calculating expected duration and marginal change in duration from the Cox proportional hazard model. Available for Stata and R.
- MI: multiple imputation through iterative equations, with support for continuous, binary, ordinal, nominal, censored, and multilevel data, and Bayesian models. With Ben Goodrich and Andrew Gelman. Available for R.

COMMUNITY SERVICE

Code for Charlottesville
Co-founder and Captain

- Organizing volunteers and developing community service tech project with local government, legal aid, and area nonprofits
- Founded September 2019. Currently more than 300 volunteer members and 18 community partner organizations