Gonzalo Rivero

This version: July 4, 2025

Contact

Home:	me@grivero.net	(917) 446-7449	Washington, DC
Work:	grivero@wmata.com	(202) 962-1051	Washington, DC

Employment

WMATA, Data Lab — *Washington, DC (USA)*
Director — Jan. 2024–Present

- Established and led a new research-to-practice organization at the intersection of data, research, and engineering. Collaborated with senior leadership to develop strategic plan for data innovation and long-term planning challenges beyond immediate needs
- Supervised a $1.5M budget and 35+ staff across managers, researchers, engineers, and contractors. Oversaw key initiatives including: methods to assess bus priority interventions using high-resolution AVL data, tools to forecast platform crowding using computer vision and real-time faregate data, analyses of long-term impact of remote work using LBS data, or demand modeling tools to estimate ridership impacts of fare and service changes.
- Restructured an existing partnership with the Transit Lab at MIT by centralizing project management, expanding researcher autonomy, and focusing on strategic projects. Developed a "last mile" engineering program to ensure that academic research met authority operational needs. Established a Research Council to coordinate cross-departmental initiatives and create structured channels for executive leadership input.
- Incubated a modern cloud analytics platform (`Trino`, `Iceberg`, `Dagster`, `dbt`, `OpenMetadata` on Azure) to replace legacy Oracle systems in collaboration with the IT department that could address critical issues: lack of data governance, fragmented pipeline ownership, slow performance, and poor quality control practices. Established reproducibility standards and fostered an Authority-wide data science community of practice.
- Continued individual contributor work alongside leadership responsibilities, personally building a Bayesian state-space model to identify changes in trends in rail ridership (`Stan`), a custom dashboard to monitor and document "ghost buses" (`Flask` and `Vue.js`), a pipeline to calculate transit competitiveness (`R`), and a tool to estimate bus stop times from coarse GTFS-RT data (`PyMC3`).

Pew Research Center, Data Labs — *Washington, DC (USA)*
Associate Director — Feb. 2021–Jan. 2024

- Supervised a team of 8 researchers and engineers delivering computational social science projects with major national media coverage and academic impact. Rewrote the team's strategic plan, mission and vision statements, and key performance metrics in collaboration with executive leadership for board approval.
- Transformed internal operations into a professional research team by implementing structured processes for hiring, onboarding, career progression, and project management. Established methodological standards and data governance processes that improved research trustworthiness and reliability.
- Directed a broad research portfolio and provided methodological and editorial support for projects using new data sources and methods including research on social media behaviors of political elites and the general population, online sermons, school mission statements, digital content decay, and object prevalence in street-level imagery.
- Led an organization-wide generative AI policy development to assess impacts of LLM on research practices and organizational outreach strategy. Standardized open-source policies with Legal and Communications teams and published 8 internal research tools for broader use.
- Managed the modernization of the team's infrastructure by consolidating a fragmented toolset and standardizing workflows (`Prefect` and `JupyterHub` running on `AWS`) with comprehensive documentation. Eliminated project setup barriers that previously required specialized expertise, democratizing research infrastructure access.

Westat, Statistics and Evaluation Sciences Unit — *Rockville, MD (USA)*
Principal Data Scientist and Manager — Dec. 2020–Feb. 2021
Senior Data Scientist and Manager — Jul. 2018–Dec. 2020
Senior Data Scientist — Oct. 2015–Jul. 2018

- Built and led the first data science team integrating machine learning and novel data sources (text, images, audio, video) into existing operations for a large traditional statistical organization.
- Delivered hands-on technical solutions on core strategic projects like NLP models for drug identification in medical records (`TensorFlow`, `spaCy`, `Flask`), a COVID fieldwork decision dashboard (`Python`, `Angular`, `PostgreSQL`, `Kubernetes`), tools for paradata extraction and preparation from CASI logs, an SPC toolkit for quality monitoring of coding teams, or an extensible framework for responsive and adaptive survey design.

- Modernized organization-wide research practices by leading workshops introducing a reproducibility toolkit to a SAS-heavy organization. Resulted in white paper and the adoption of modern software engineering best practices for statistical programmers and software developers.
- Served as lead consultant for statistical modeling and machine learning projects across the company. Represented researchers and statisticians in the technical working group on emerging data technologies that advised senior leadership on planning for data infrastructure investments. Led a data science working group tasked with preparing a vision for statistical modernization.
- Contributed technical expertise and designed and wrote responses to procurements from government agencies at local, state, and federal levels.

YouGov, Scientific Research Group — *Palo Alto, CA (USA)*
Statistical Scientist — Nov. 2014–Oct. 2015
Senior Analyst — Apr. 2013–Nov. 2014

- Performed matching and weighting for custom survey studies on a opt-in panel and maintained and improved an `R` toolkit for internal research operations. Evaluated and incorporated new methods for survey data processing.
- Redesigned `R` data pipelines for client-facing interactive dashboards, reducing runtime by 10x and automating brittle manual processes. Extended capabilities to accommodate analysis of multilingual open-ended responses.
- Developed a Bayesian conjoint analysis tool (`Stan`) to bring statistical modeling of choice experiments in-house. Served as statistical consultant for clients on statistical modeling, providing expertise in Item Response Theory, linear state space models, and survey experiments.
- Designed and led a C-suite commissioned project for observational and experimental evaluation of survey design practices. Developed an automated tool for topic classification of survey instruments.

Technical Skills

Statistical Computing & ML: R, Python, Stan/Jags, Julia
Analysis of social and behavioral data (experimental and observational) using frequentist and Bayesian approaches. Expertise in statistical learning, choice modeling, longitudinal and survival analysis, and causal inference. Experience developing and deploying custom models in computer vision and NLP applications. Professional experience in `R` package development for scientific applications.

Software & Data Engineering: `Python`, `JavaScript`, `Go`, `R`
Full-stack design and implementation using synchronous and asynchronous patterns in `Flask` with `Vue.js` frontends. Database design and engineering in `SQL` (`PostgreSQL`, `Oracle`) and NoSQL (`MongoDB`, `ElasticSearch`). Data engineering using `SQLAlchemy`, `dbt` and `Airflow`. Implementation of data quality monitoring and validation with `Great Expectations`. Software development in Scrum/Agile teams with strong focus on computational reproducibility.

Cloud & DevOps: `Amazon AWS`, `Docker/Podman`, `Kubernetes`, `Prefect/Airflow`
Experience working in cloud-native environments with containerization (`Docker`) and orchestration (`Kubernetes`). GitOps and DataOps practices using `GitHub/GitLab` workflows. Hands-on systems administration and operational work in legacy environments.

Education

New York University — *New York City, NY (USA)*
Ph.D. in Political Science — 2013
M.A. in Political Science — 2009

Instituto Juan March de Estudios e Investigaciones — *Madrid, Spain*
M.A. in Social Sciences

Universidade of Vigo — *Ourense, Spain*
D.E.A. in History (Prehistory & Archaeology)
Licenciado in History

Books

1. Jungherr, A.; **Rivero, G.** and Gayo-Avello, D. (2020) *Retooling Politics: How Digital Media are Shaping Democracy*. Cambridge University Press.
   Reviews: International Journal of Press/Politics, Medien & Kommunikationswissenschaft
2. **Rivero, G.** (2011) *Análisis de datos incompletos en Ciencias Sociales*. CIS, Madrid.
   Reviews: Metodología de Encuestas

Peer-reviewed Articles

1. Asensio, O.I.; Lan, T.; Moore, C.; Simsekler, M.C.E.; Ulibarri, N. and **Rivero, G.** (2024): "Data Technologies and Analytics for Policy and Governance: A Landscape Review." *Data & Policy*, 7: e25.
2. Campione, J.R.; Nooney, J.G.; Kirkman, M.S.; Pfaff, E.; Mardon, S.; Benoit, S.R.; McKeever-Bullard, K.; Yang, D.H., **Rivero, G.**; Rolka, D. and Saydah, S. (2023): "Validated Models Using EHRs or Claims Data to Distinguish Diabetes Type among Adults." *Advances in Diabetes & Endocrinology*, 7 (1): 1-8.
3. Billington, C.; **Rivero, G.**; Jannett, A. and Chen, K. (2022): "A Machine Learning Model Helps Process Interviewer Comments in CAPI Instruments: A Case Study." *Field Methods*, 34 (4): 275-287.
4. Gallego, J.; **Rivero, G.** and Martínez-Gordillo, J.D. (2021): "Preventing Rather than Punishing: An Early Warning Model for Corruption in Public Procurement." *International Journal of Forecasting*, 37 (1): 360-377.
5. Suhay, E.; Klasnja, M. and **Rivero, G.** (2021): "Ideology of Affluence: Rich American's Explanations for Inequality and Redistributive Attitudes." *Journal of Politics*, 83 (1): 367-380.
   Awards: APSA Best Paper Prize 2018. Class and Inequality Section.
6. Nooney, J.; Kirkman, M. S.; Bullard, K. M.; White, Z.; Meadows, K.; Campione, J.; Mardon, R.; **Rivero, G.**; Benoit, S. R.; Pfaff, E.; Rolka, D. and Saydah, S. (2020): "Identifying Optimal Survey-Based Algorithms to Distinguish Diabetes Type among Adults with Diabetes." *Journal of Clinical and Translational Endocrinology*, 21: 100231.
7. Tourangeau, R.; Sun, H.; Yan, T.; Maitland, A.; **Rivero, G.** and Williams, D. (2018): "Web Surveys by Smartphones and Tablets: Effects on Data Quality." *Social Science Computer Review*. 36 (5): 542-556.
8. Tourangeau, R.; Maitland, A.; **Rivero, G.**; Sun, H.; Williams, D. and Yan, T. (2017): "Web Surveys by Smartphones and Tablets: Effects on Survey Responses." *Public Opinion Quarterly*, 81 (4): 896-929.
9. **Rivero, G.** (2017): "Preaching to the Choir. Ideology and Following Behavior in Social Media." *Contemporary Social Science*, 14 (1): 54-70.
10. Fernández-Vázquez, P.; Barberá, P. and **Rivero, G.** (2015): "Rooting Out Corruption or Rooting For Corruption? The Heterogeneous Electoral Consequences of Scandals." *Political Science Research and Methods*, 4 (2): 379-397
    Media coverage: Pacific Standard, El País
11. **Rivero, G.** (2015): "Heterogeneous Preferences in Multidimensional Voting Models: Ideology and Nationalism in Spain." *Electoral Studies*, 40: 136-145.
12. Przeworski, A.; **Rivero, G.** and Xi, T. (2015): "Elections as a Conflict Processing Mechanism." *European Journal of Political Economy*, 39: 235-248.
13. Barberá, P. and **Rivero, G.** (2015): "Understanding the Political Representativeness of Twitter Users," *Social Science Computer Review*, 33 (6): 712-729.
    Media coverage: The Atlantic

Chapters in Books

1. Gayo-Avello, D.; Jungherr, A. and **Rivero, G.** (2023): "Social Media and Electoral Prediction: Ten Years After." In T. Yasseri (ed.): *Handbook of Computational Social Sciences*. Edward Elgar Publishing.
2. Riddles, J.; Burke, J.; **Rivero, G.** and Rust, K. (2017): Adjustments to Student Weights to Account for Student Nonresponse in the National Assessment of Educational Progress. *Proceedings of the Joint Statistical Meetings*. American Statistical Association: 561–570.
3. Barberá, P. and **Rivero, G.** (2012): ¿Un tweet, un voto? Desigualdad en el debate político en Twitter. In R. Cotarelo and I. Crespo (Ed.): *La comunicación política y las nuevas tecnologías*. Ed. Libros de la Catarata: 200–220.
4. Lapuente, V.; Fernández-Vázquez, P. and **Rivero, G.** (2011): Carencias en el control vertical y horizontal de la corrupción. In J. Estefanía (Ed.): *Informe sobre la Democracia en España, 2011*. Fundación Alternativas: 199–222.

Reports

1. Chapekis, A.; Bestvater, S.; Remy, E. and **Rivero, G.** (2024): When Online Content Disappears. *Pew Research Center*.
   Media coverage: EuroNews, The Independent, Axios, The Week, The Verge, The Register, Library Journal, El País
2. **Rivero, G.** and Muñoz, J. (2022): La metodologia de les estimacions electorals del CEO. *Papers de Treball del Centre d'Estudis d'Opinió*.
3. Bestvater, S.; Shah, S.; **Rivero, G.** and Smith, A. (2022): Politics on Twitter: One-Third of Tweets From U.S. Adults Are Political. *Pew Research Center*.
   Media coverage: AP, Washington Post, ABC News, CNN, NBC, The Hill, Washington Times, Tech Crunch, Engadget, Adweek

4. McClain, C.; Widjaya, R.; **Rivero, G.** and Smith, A. (2021): The Behaviors and Attitudes of U.S. Adults on Twitter. *Pew Research Center*.
   Media coverage: New York Times, CNN, Fox News, Axios, Politico, The Hill, The Star, Chicago Tribune, Protocol, Adweek, Daily Mail, Fast Company, Mashable
5. Shah, S.; Widjaya, R.; Smith, A.; **Rivero, G.** and Chapekis, A. (2021): Charting Congress on Social Media in the 2016 and 2020 Elections. *Pew Research Center*.
   Media coverage: Protocol, Adweek
6. **Rivero, G.** and Fernández-Vázquez, P. (2011): Las consecuencias electorales de los escándalos de corrupción municipal, 2003–2007. *Estudios de Progreso*, 59.
   Media coverage: El País, Expansión

Other Publications

1. Remy, E. and **Rivero, G.** (2024): Reproducibility as part of code quality control. *Decoded, Pew Research Center*.
2. Broderick, B. and **Rivero, G.** (2023): How we adopted Kubernetes for our data science infrastructure. *Decoded, Pew Research Center*.
3. Castner, H. and **Rivero, G.** (2023): What Twitter users say versus what they do: Comparing survey responses with observed behaviors. *Decoded, Pew Research Center*.
4. Remy, E. and **Rivero, G.** (2023): How we review code at Pew Research Center. *Decoded, Pew Research Center*.
5. Guirola, L.M. and **Rivero, G.** (2022): Political polarization contaminates the link with both partisan and independent institutions: Evidence from 138 cabinet shifts. *Documentos de trabajo del Banco de España*, 2237.
6. **Rivero, G.** (2021): Democracia, evaluación y participación: los laboratorios de políticas públicas. *Strategic Metropolitan Plan of Barcelona*.
7. **Rivero, G.** and Chen, J. (2020): Good Coding Practices to Ensure Reproducibility. *Westat Issue Brief Series*.
8. **Rivero, G.** (2011): Integrality and separability in multidimensional voting models: Ideology and nationalism in Spanish regional elections. *Estudios/Working Papers of the CEACS*, 265.

Published Software

1. `escons`: An R package to estimate seat allocations (Commissioned by the Centre d'Estudis d'Opinió)
2. `dshare`: An R package to estimate district shares (Commissioned by the Centre d'Estudis d'Opinió)

Additional Training

Short courses

R and Spark: Tools for Data Science Workflows. National Institute of Statistical Sciences	Sep. 2017
Statistical Learning and Data Mining. Trevor Hastie & Rob Tibshirani	Oct. 2016

Long courses

Quantitative Methods for Social Research. ICPSR, University of Michigan	June–Aug. 2005
P.G.C. in Advanced Methods of Applied Statistics, UNED	2005
P.G.C. in Celtic Studies, Universidad Internacional Menéndez Pelayo	2003
P.G.C. in Management of Cultural Heritage, Laboratory of Archaeology, CSIC	2003

Research Affiliations

Doctor Miembro (Research Fellow). Juan March Institute, Madrid (Spain)	2013–Present
Visiting Scholar. The Institute for Economic Analysis, CSIC, Barcelona (Spain)	2010–2011

Teaching Experience

Writing for Survey Researchers (2 hours). *Universidad Complutense*, Madrid (Spain)	Apr. 2023
Survey Statistics for Public Opinion Research (4 hours). *40db*, Madrid (Spain)	Jan. 2020
Reproducibility in Quantitative Research (8 hours). *Westat*, Rockville, MD	Sept.–Dec. 2019
Big Data for Social Research (20 hours)	
Universidad Católica, Montevideo (Uruguay)	July 2019
Universidad del Rosario, Bogotá (Colombia)	June 2017
Advanced R (6 hours). *Westat*, Rockville, MD	June 2016
Introduction to Statistical Programming in R (6 hours). *Westat*, Rockville, MD	Feb. 2016
Incomplete Data Analysis in Surveys (8 hours). *Universitat Autònoma*, Barcelona (Spain)	Nov. 2010

Invited Talks

Universidad del Rosario (Colombia) 2018
University of Durham, Universidad del Rosario (Colombia) 2017
Trinity College of Dublin 2016
TIGER (Toulouse School of Economics) 2014
Instituto Tecnológico Autónomo de México, Spanish National Research Council 2012
Universitat Autònoma de Barcelona 2011

Outreach

Mentions and interviews in media

El País. "Políticos, periodistas y tuitstars: ¿qué grupo está más polarizado?" Dec. 25, 2020.
El Universal. "2020: ¿año de la redención de las encuestas en EU?" Sept. 6, 2020.
El Periódico. "'Patrullas ciudadanas virtuales': la fina línea entre la colaboración vecinal y la alarma social." Sept. 4, 2019.
La Voz de Galicia. "Lo que trasciende de los datos." Dec. 30, 2016.
El Mundo. "El voto oculto de Donald Trump." Oct. 27, 2016.
Univisión. "Por qué es muy improbable que fallen las encuestas como dice Trump." Oct. 25, 2016.
Xataka. "¿Predice el big data sobre redes sociales mejor que las encuestas quiénes ganan las elecciones?" Dec. 15, 2015.
El País. "Rajoy y Rubalcaba empatan en la valoración de los usuarios de Twitter". Sep. 26, 2011.

Writings in media

The Washington Post. "Sanders and Bloomberg want to redistribute wealth. Most millionaires and billionaires don't," with Liz Suhay and Marko Klasnja. Feb. 17. 2020
Ahora. "Cuando PP y PSOE perdieron a los jóvenes," with María Ramos. Jul. 8 2016.
El País. "El éxito de las encuestas," with Kiko Llaneras. Mar. 30, 2015.
European Politics and Policy Blog, London School of Economics. "Political discussions on Twitter during elections are dominated by those with extreme views," with Pablo Barberá. Dec. 13, 2014.

Service & Membership

Editorial Service

Associate Editor, *Data & Policy* 2020–2022

Referee

Journals: Public Opinion Quarterly; British Journal of Political Science; Journal of Conflict Resolution; Political Communication; EPJ Data Science; Social Science Computer Review; International Journal of Press/Politics; International Journal of Public Opinion Research; Social Media & Society; Government and Opposition; European Political Science Review; Communication Research; Congress & the Presidency; Journal of Survey Statistics and Methodology; Data & Policy; Armed Forces and Society; War in History; Dados; Spanish Journal of Political Science; Southern European Politics and Society; El Profesional de la Información; Revista Internacional de Sociología

Book publishers: Cambridge University Press

Funding organizations: Spanish National Agency for Evaluation and Prospective, Institute of Education Sciences

Service in Organizations

Data for Policy (Area Committee for Data Technologies & Analytics) 2021–2022
Washington Statistical Society (Board of Directors) 2018–2020
AAPOR (Communications Committee) 2019–2020

Expert panels and advisory roles

National Office of Foresight and Strategy, *Ministerio de la Presidencia* (Spain) 2022
Centre d'Estudis d'Opinió, *Generalitat de Catalunya* (Spain) 2022
MA in Computational Social Sciences, *Universidad Carlos III* (Spain) 2021
Bureau of Justice Statistics, *U.S. Department of Justice* 2018