

JORGE ALBERTO ROA CONTRERAS

Software Developer · Data Scientist

 jorgeroa@stanford.edu ·  jurjoroa ·  jorge-roac ·  0000-0002-3972-9793 ·  Jorge Roa
 Google Scholar

EDUCATION

Hertie School

Master of Science: Data Science for Public Policy

Berlin, Germany

09/2022 - 06/2024

- Held a 75% tuition scholarship.
- Dissertation focused on the effect of the judicial reform on the homicides in Mexico: A quasi-experimental design employing a Generalized Additive Mixed Model.

Center for Research and Teaching in Economics

Bachelor of Arts: Public Policy

Aguascalientes, Mexico

08/2016 - 07/2020

- Held an 80% tuition scholarship.
- GPA: 3.5/4
- Dissertation focused on the relationship between violence, political actors and criminal groups in Mexico employing Instrumental Variables.

Complutense University of Madrid

Faculty of Political Science and Sociology

Madrid, Spain

01/2019 - 07/2019

- Exchange semester: Took five courses, including Introduction to Demographics and International Security Studies.

EXPERIENCE

Stanford University School of Medicine

Software Developer Associate

Palo Alto, California, US

10/2024-Present

- Software development of a colorectal cancer simulation model for the Cancer Intervention and Surveillance Modeling Network (CISNET) program, focusing on cost-effectiveness analysis and decision modeling.
- Collaboration with the National Cancer Institute (NIH), CISNET and multidisciplinary team of researchers and clinicians to develop and implement simulation models for colorectal cancer (CRC) screening and treatment in the U.S. population.
- Colorectal cancer simulation model (SimCRC) model adapted for Chile to optimize FIT screening (threshold, age, frequency) and estimate long-term impact on cancer outcomes and resource use.
- Developing a cost-effective colorectal cancer screening program for high-risk individuals in Mexico, with plans for nationwide implementation.
- Designed and maintained reproducible simulation models and R packages for health policy research, emphasizing software accessibility, transparency, and open science principles.
- Built dynamic Shiny applications and Quarto reports for interactive data visualization, stakeholder communication, and technical documentation.
- Led data wrangling, statistical analysis, and visualization workflows using R (data.table, ggplot2, etc.), while ensuring software quality through unit testing, Git-based version control, and collaborative development across research teams.
- Reduced calibration execution time by 60% by replacing full 0–100 age loops with a difference-array/cumsum algorithm that computes tumor bucket counts only at change points, using rolling joins to broadcast results across reporting ages.

- Built automated web scrapers and crawlers to collect real-time data from Twitter, enabling early detection of flu and allergy-related trends.
- Extracted and structured political speeches and online content to support analysis of narrative framing and public discourse.
- Constructed a comprehensive database of international organizations by scraping official websites, speeches, and event listings.
- Tagged causal relationships in political and health-related texts to support the development of training datasets for large language models (LLMs).

- Working on optimization of an algorithm for cost-effectiveness analysis for colorectal cancer.
- Migration of natural history, screening and surveillance colorectal cancer models from C++ to R.
- Estimated the infection force of *Helicobacter pylori* among the U.S. population using catalytic models based on data from the NHANES Survey.
- Designed a Shiny application and a website to show interactively the force of infection and the age to infection based on the research of the *Helicobacter pylori* bacteria.
- Determined the cumulative mortality risk associated with COVID-19 across population at both state and national levels in Mexico building survival models.
- Conducted a twenty-year trend analysis of homicides in Mexico utilizing Bayesian methods to understand changes over time.
- Evaluated the impact of the COVID-19 pandemic on homicide trends in Mexico using multiphase mixed-effects models to isolate pandemic-related variations.

- Created a dataset locating and characterizing 148 criminal groups in Mexico, computing web scraping techniques from the information reported by the press and social media implementing code written in Python and text analysis written in R.
- Designed and programmed an electronic survey to study harm reductions associated with psychoactive substances and involvement in violent sociability practices.
- Coordinated and supervised a team for implementing the geocode of more than 10,000 addresses related to executions, confrontations, attacks, arrests, and different crimes in Mexico manually and automatically developing code written in Python.

FELLOWSHIPS

- Data Science for Social Good (DSSGx) program.
- International Organization for Migration Program: built a model to predict migration and an Early Warning System.

SKILLS

Programming Languages and Tools - proficient: R, Python, Quarto, Shiny, Selenium, HTML, CSS, GitHub, git, LaTeX, Stata, ArcGis, SurveyToGo, **intermediate:** Tableau, SQL, Tensorflow/Keras, Bash, GeoDa, Qgis, Javascript

Languages - Spanish (native), English (TOEFL iBT: 100)

PEER-REVIEWED PUBLICATIONS

- Kriit, H.K., Chen-Xu, J., Semenza, J.C., ... **Roa-Contreras, J.A.**, ... Rocklöv, J. (2026). [The 2026 Europe report of the Lancet Countdown on health and climate change: narrowing window for decisive health action](#). *The Lancet Public Health*. DOI: [https://doi.org/10.1016/S2468-2667\(26\)00025-3](https://doi.org/10.1016/S2468-2667(26)00025-3)
- van den Berg, D.M.N., Brück, C.C., Nascimento de Lima, P., Alarid-Escudero, F., Hahn, A.I., Lansdorp-Vogelaar, I., & Colorectal Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET) [incl. **Roa, J.**]. (2026). [The best screening test is the one that gets followed up on](#). *Gastroenterology*. DOI: <https://doi.org/10.1053/j.gastro.2026.02.013>
- Romanello, M., Walawender, M., Hsu, S.-C., ... **Roa, J.**, ... Costello, A. (2025). [The 2025 report of the Lancet Countdown on health and climate change](#). *The Lancet*. DOI: [https://doi.org/10.1016/S0140-6736\(25\)01919-1](https://doi.org/10.1016/S0140-6736(25)01919-1)
- Taylor, C.S., McMahon, M.V., Ward, Z.J., Alarid-Escudero, F., Camargo, M.C., Laszkowska, M., **Roa, J.**, & Yeh, J. M. (2024). [Birth cohort and age-specific trends in global Helicobacter pylori seroprevalence: A scoping review](#). *The Lancet Regional Health - Americas*, 100877. DOI: <https://doi.org/10.1016/j.lana.2024.100877>
- McMahon, M. V., Taylor, C. S., Ward, Z. J., Alarid-Escudero, F., Camargo, M. C., Laszkowska, M., **Roa, J.**, & Yeh, J. M. (2024). [Helicobacter pylori infection in the United States beyond NHANES: a scoping review of seroprevalence estimates by racial and ethnic groups](#). *Lancet regional health. Americas*, 41, 100890. DOI: <https://doi.org/10.1016/j.lana.2024.100890>
- Alarid-Escudero, F., Gracia, V., Luviano, A., **Roa, J.**, Peralta, Y., Reitsma, M. B., Claypool, A. L., Salomon, J. A., Studdert, D. M., Andrews, J. R., & Goldhaber-Fiebert, J. D. (2021). [Dependence of COVID-19 Policies on End-of-Year Holiday Contacts in Mexico City: A Modeling Study](#). *MDM Policy & Practice*. DOI: [10.1177/23814683211049249](https://doi.org/10.1177/23814683211049249)

FORTHCOMING PUBLICATIONS

- Sun, D., Han, W., Alarid-Escudero, F., **Roa-Contreras, J.A.**, Wang, S., Müller, D.T., Huang, R.J., Ladabaum, U., Spaander, M.C.W., Camargo, M.C., Park, J.Y., Chen, W., Lansdorp-Vogelaar, I., & Hontelez, J. (Forthcoming, 2026). [Trends in Helicobacter pylori Prevalence, Force of Infection, and Attributable Disease Burden in China from 1990 to 2023: A Systematic Review, Meta-Analysis, and Modelling Study](#).
- Kaufmann, M. B., **Roa, J.**, Yeh, J. M., Goldhaber-Fiebert, J. D., & Alarid-Escudero, F. (The Lancet Regional Health – Americas). [Modeling the impact of mass testing and eradication strategies of Helicobacter pylori in the United States](#).
- **Roa, J.**, Peralta, Y., McMahon, M., Taylor, C., Ward, Z. J., Camargo, M. C., Yeh, J., & Alarid-Escudero, F. (in preparation). Race-specific estimates of the force of infection of Helicobacter pylori using two national seroprevalence surveys in the US: A hierarchical Bayesian catalytic epidemic modeling study.
- Manohara, K., **Roa, J.**, Jankin, S., Garcia-Corral, P., et al. (Forthcoming, 2024). From searches to sneezes: Evaluating digital indicators for allergic rhinitis surveillance in the United Kingdom. In *Proceedings of the 2nd International Workshop on Big Data Analytics for Health and Medicine (IEEE BigData 2024)*, Washington, D.C., USA.

CONFERENCE ABSTRACTS

- **Roa, J.**, Gracia, V., Goldhaber-Fiebert, J.D., & Alarid-Escudero, F.(2025). Accurate uncertainty quantification in Bayesian calibration accounting for correlated targets, Society for Medical Decision Making (SMDM) 47th Annual Meeting, Michigan Union, University of Michigan, Ann Arbor, USA, accepted.

CONFERENCE PRESENTATIONS

A Hierarchical Bayesian Meta-Analytic Helicobacter Pylori (*H. pylori*) Infection Generator for the U.S. Population.

- *Poster*: Society for Medical Decision Making (SMDM) 18th Biennial European Conference, 2023. **Berlin, Germany**

SOFTWARE

ggpop: Visualizing Population Data — R package extending ggplot2 for icon-based population visualizations using `geom_pop()` and `geom_icon_point()`, with 2,000+ Font Awesome icons and built-in themes.

- **Roa-Contreras, J. A.**, Soultanova, R., Alarid-Escudero, F., & Pineda-Antunez, C. (2024). ggpop: Visualizing Population Data (Version 1.7.0) [R package]. <https://jurjoroa.github.io/ggpop/>

DEVELOPED WEBSITES

Jorge Roa, MSc — Personal Site — Multi-page Quarto site aggregating peer-reviewed publications, media appearances, tutorials, software, and a curated portfolio. Custom dark design system, animated SVG identity, and a case-study format for past web/data projects.

DARTH Workgroup — Official website for the Decision Analysis in R for Technologies in Health collaborative, featuring open-source tools, tutorials, and resources for health technology assessment and decision analytic modeling.

Ethics × Data Science — Teaching resource platform providing tutorials, case studies, and instructional materials on data science ethics, designed for professional schools and educators across disciplines.

Hertie Coding Club — Website for the student coding club founded at the Hertie School, offering learning resources, workshops, and event information for a community of 200+ members across R, Python, and other languages.

ggpop documentation site — Package documentation and tutorial site for the ggpop R package, built with pkgdown and hosted via GitHub Pages.

Stanford CEA Chile 2025 — Course website for a Stanford University simulation modeling program in collaboration with Chilean institutions, covering cost-effectiveness analysis, decision modeling, and R programming for health policy.

SHINY APPS

Hp Generator — Interactive Shiny application for generating Helicobacter pylori infection scenarios. Produces parameter sets and synthetic cohorts for downstream modelling, with a reactive UI for exploring inputs and exporting results.

TUTORIALS & TECHNICAL WRITING

Making the Transition: A Guide for Switching from Stata to R — Comprehensive migration guide for researchers and analysts moving from Stata to R, covering syntax, workflows, and equivalent

functions. *jorgeroac.com, 2024.*

How We Met Quanteda: Text Analysis with R – Introduction to corpus construction, tokenization, and document-feature matrices using the quanteda ecosystem. *jorgeroac.com, 2022.*

Bias in AI: Detection and Mitigation – Applied tutorial on identifying and mitigating algorithmic bias in machine learning pipelines using aif360 and aequitas. *jorgeroac.com, 2023.*

Enable Multithreading with data.table on Mac (Intel + Apple Silicon) – Step-by-step guide to unlocking parallel processing in R's data.table on macOS. *jorgeroac.com, 2022.*

PDF to Text: Extract Text and Tables with Python – Practical walkthrough of PDF parsing pipelines using pdfplumber and Camelot for research data extraction. *jorgeroac.com, 2023.*

Counting Without Loops: Difference Arrays and Cumulative Sums in R – Algorithmic tutorial on vectorized range-update techniques and their applications in epidemiological modeling. *jorgeroac.com, 2025.*

AI STACK

cv-to-website – Claude Code plugin and portable agentic skill that turns a CV (PDF, DOCX, image, or plain text) into a deployed Quarto personal website. Runs a five-phase pipeline (`detect` → `install` → `auth` → `render` → `deploy`), producing a profession-aware multi-page site with listing-card publications, optional Works/Gallery widgets for visual portfolios, and a live URL on GitHub Pages.

AWARDS

2022 CISNET's best Junior Investigator poster award based on the research of *Helicobacter pylori* (*H. pylori*) Infection Generator.

TEACHING ASSISTANT EXPERIENCE

Universidad de Chile, School of Public Health **Santiago, Chile**
Cost-Effectiveness Analysis and Decision Modeling in R 11/2025–12/2025

- Teaching assistant for the Stanford CEA Chile 2025 intensive course, held at the Universidad de Chile School of Public Health over one week (November 24 – December 5, 2025).
- Led a session on setting up R for cost-effectiveness analysis, covering environment configuration, package installation, and introductory CEA workflows.

Center for Research and Teaching in Economics **Aguascalientes, Mexico**
Quantitative Methods I 01/2021–07/2021

- Instructed a weekly two-hour class for undergraduate students where it was covered the theory of statistical methods and practical applications in R.

Catholic University of Uruguay **Remote**
Methods for Public Policy Evaluation 08/2019–07/2020

- Assisted a course in methods for public policy evaluation for the Winter School in Methods 2020. We checked Difference-in-Difference, Matching, Instrumental Variables, etc.

INTERNSHIPS

Department of Health Policy, Stanford University **Palo Alto, California, US**
Research Assistant 06/2023–09/2023

- Implemented and learned techniques of Discrete Event Simulation, Markov models, value of information (VOI), and cost-effectiveness analysis.

- Learned the introduction to implement natural history, surveillance, and screening methods to simulate colorectal cancer in U.S. populations.
- Learned to run cancer models at a grander scale using bash and Sherlock, Stanford's supercomputer.

Criminal Analysis Unit: Security Secretary State of Mexico

Toluca, Mexico

Data Analyst

08/2019–07/2020

- Analyzed databases related to crimes, as well as a review of characteristics of criminal groups in the State of Mexico. These analyzes served as inputs for making security policy decisions within the secretariat.

ACADEMIC MEMBERSHIPS

Decision Analysis in R for Technologies in Health (DARTH)

Member

2024-Present

Society for Medical Decision Making (SMDM)

Member

2023-Present

Stanford

Stanford Data Science x Decision Science [DS]² Research Hub

2023-Present

Stanford Impact Labs

2025-Present

Cancer Intervention and Surveillance Modeling Network (CISNET)

Gastric Cancer site

2021-Present

Colorectal Cancer site

2023-Present

CERTIFICATIONS

Cost-effectiveness and decision modeling using R, 2023

Decision Analysis in R for Technologies in Health (DARTH)

Fundamentals of Accelerated Computing with CUDA Python, 2023

NVIDIA: Deep Learning Institute

Data Science Summer School, 2022

Hertie School

Drug Policy, Health and Human Rights, 2021

Center for Research and Teaching in Economics

Introduction to R and Spatial Analysis with GIS, 2020

Winter School in Methods. Catholic University of Uruguay

LEADERHSIP/INITIATIVE

Founder of the Hertie Coding Club

Berlin, Germany

Hertie School

11/2022

- Student club initiative for founding a coding club with the objective to teach and practice code skills among students in different programming languages for data analysis and data science for public policy.
- Designed and instructed multiple lessons of R and other languages as well as coordinate different teams between students and school authorities.
- The Club has a base of more than 200 members including students and members of the Hertie School.

TALKS

2023 CIVICA Student Association Forum. Hertie Coding Club: A student initiative to promote coding skills. **Warsaw, Poland**

2021 Seminario de Intersecciones de lo Religioso. “Malverde no me ha fallado”: iglesias y narcotráfico. **Mexico**

HACKATHONS

DataFest Germany, 2023

Ludwig-Maximilians-Universität (LMU)

Argonne Leadership Computing Facility (ALCF) Incite Hackathon, 2023

ALCF, NVIDIA and OpenACC

POLICY BRIEF

Peralta, Y., **Roa, J.**, Miranda, A., Alarid-Escudero, A. (2021). [Homicidios en México: Su evolución en las últimas dos décadas](#). Programa de Política de Drogas.

Peralta, Y., **Roa, J.**, Alarid-Escudero, A. (2021). [La tendencia de los homicidios en México durante la pandemia por COVID-19](#). Programa de Política de Drogas.

MEDIA PUBLICATIONS

Velázquez, S., **Roa, J.**, Padilla, A., and Alarid-Escudero, F. (2021, June 23). [¿Qué tan abiertos son los “datos abiertos” de vacunación en México?](#) Animal Político.

Roa, J. (2020, June 3). [La geografía de la violencia](#). Animal Político.

Roa, J. (2020, December 10). [Analizando el indicador estatal covid de Aguascalientes](#). LJA.mx.

Roa, J. (2020, December 14). [De la corte a las calles](#). Animal Político.

INTERVIEWS ON TELEVISION, RADIO, NEWSPAPERS AND INTERNATIONAL MEDIA

Mexico: *emeequis*. (2022, January 29). [Grupos criminales fuera del radar: los que no salen en informes de FGR](#). *emeequis — Al Día*.

emeequis. (2022, January 27). [La Fiscalía no habla de estos grupos criminales en las sombras: base de datos revela nombres](#). *emeequis — Investigaciones*.

El Herald de México. Alatorre, J. (2022, January 27). [Las Noticias con Javier Alatorre — Entrevista sobre seguridad en México](#).

La Silla Rota. (2021, October 19). [Estos 12 nuevos grupos criminales surgieron durante el gobierno de AMLO](#).

Debate. (2021, July 12). [Durante administración de AMLO se han creado 12 nuevos grupos criminales: CIDE](#).

Meganoticias Mx. Olvera, A. (2021, April 23). [Crimen organizado, actor principal en el proceso electoral](#). *Meganoticias Mx*.

Mexican Institute of the Radio. Iglesias, A. (2021, April 27). [Análisis de los grupos delictivos en México](#).

El Siglo de Torreón. (2021, April 18). [Operan en México hasta 148 cárteles](#).

Reforma. Martínez, C. (2021, April 16). [Operan en México hasta 148 cárteles](#). *Reforma* (Print Edition).

Tercera Vía. (2021, April 15). [El mapa del crimen: 148 grupos delincuenciales se disputan México](#).

El Financiero-Bloomberg. Jiménez, H., Salazar, A. & Cepeda, O. (2020, November 15). [En México hay 148 grupos criminales: CIDE](#). *El Financiero Bloomberg TV*.

Spain: *El País*. (2021, July 18). [Los 12 grupos criminales de la era López Obrador](#).

France: *Radio France International*. (2021, May 17). [México: A tres semanas de las elecciones, los candidatos son víctimas de la violencia](#).

United States: *Univision Noticias — América Latina.* (2021, April 20). Estos son los estados de México en los que 148 cárteles operan con más fuerza.

Argentina: *Infobae México.* (2021, April 17). Radiografía del narco: así están distribuidos los 148 grupos criminales que operan en el país.

Honduras: *La Prensa.* (2021, April 20). En México operaron 148 cárteles, 34 son brazos armados.