Maya Krishnamoorthy

Email • 919-594-3840 • LinkedIn • New York, NY • GitHub

EDUCATION

Columbia University, New York, NY Expected May 2026
Master of Science, Biostatistics (Public Health Data Science Track) *GPA: 4.0/4.0*

University of North Carolina at Chapel Hill, Chapel Hill, NC Aug 2020 – May 2024
Bachelor of Science, Computer Science; Bachelor of Science, Public Health (Biostatistics Concentration) *GPA: 3.9/4.0*

Relevant Coursework: Data Science & Visualization (R), Statistical Learning (Python), Causal Inference, Biostatistical Methods (R), Graphical Models for Health Data (R), Relational Databases & SQL, Machine Learning Theory, Statistical Genetics

SKILLS

Languages: R (tidymodels, caret, shiny), SQL (MySQL), Python (pandas, numpy, PyTorch), SAS
Statistical & Analytical Methods: Predictive modeling, regression methods, causal inference, longitudinal and survival analysis

RESEARCH & WORK EXPERIENCE

Aiello Research Group, Columbia University New York, NY
Graduate Research Assistant Aug 2024 – Present

- Apply survey-weighted regression and structural equation modeling on longitudinal Add Health data (N ≈ 9,720) using R to quantify associations between dual caregiving responsibilities, chronic stress, and biological aging biomarkers in midlife adults
- Design a proof-of-concept for an internal HIPAA-compliant RAG system indexing 1,000+ research articles related to the Add Health study to support secure literature synthesis and evidence-based analysis workflows
- Develop course materials and serve on the instructional team for "Generative AI for Epidemiologists," focusing on best practices, prompt engineering, learning guidelines, and responsible use of generative AI in epidemiological research and academic settings
- Deliver weekly technical briefings translating high-dimensional statistical results to interdisciplinary teams of epidemiologists, clinicians, and project managers

Fidelity Investments Durham, NC
Data Engineering Intern Jun – Aug 2022, Jun – Aug 2023

- Built a high-throughput Dataverse-Kafka-Oracle data pipeline (500K+ transactions/day) to automate compliance operations
- Developed Kafka-based ingestion tools that reduced manual data processing by ~10 hours/week
- Integrated AI-driven modeling tools into cloud infrastructure (AWS) to support financial trend forecasting

ACADEMIC ANALYTICAL PROJECTS

Relational Database Design & Analytics on Yelp Public Data Dec 2025

- Designed and implemented a normalized MySQL database modeling user–restaurant–review interactions across 300K+ records, enforcing referential integrity through primary and foreign key constraints
- Built Python ETL pipelines to clean, validate, and load raw JSON data into structured tables
- Wrote advanced SQL queries to identify user engagement cohorts, restaurant performance segments, and temporal trends

Causal Inference Analysis of Medicaid Expansion and Mortality Dec 2025

- Implemented DiD and event-study estimators on multi-state panel data to estimate Medicaid expansion effects on mortality
- Ran 1,000 Monte Carlo bootstrap simulations to evaluate estimator bias, variance, and sensitivity under alternative data-generating processes, stress-testing parallel trends and identification assumptions
- Conducted robustness checks across multiple model specifications to assess causal validity and quantify policy-relevant effects

Utilizing Graphical Models for fMRI-Based Classification of Autism Spectrum Disorder (ASD) Dec 2025

- Estimated subject-specific functional connectivity networks from resting-state fMRI data using both the Peter-Clarke (PC) algorithm and graphical lasso in R across 40 subjects and 60 ROIs, revealing method-dependent topology with graphical lasso producing denser, globally integrated graphs and PC yielding sparser, fragmented structures
- Predicted ASD status using $\ell 1$-regularized logistic regression with LOOCV on network summary features with limited stability due to small sample size (AUC ≈ 0.76), identifying global connectedness as a key but highly variable feature under graphical lasso

Epidemiological Approach to Type II Diabetes (T2D) in the U.S Nov– Dec 2024

- Analyzed 60K+ BRFSS records using regression methods to identify demographic and behavioral risk factors for Type II diabetes
- Deployed a user-friendly interactive RShiny dashboard visualizing geographic and population-level risk patterns across states, enabling exploratory cohort analysis

EXTRACURRICULARS

Teaching Assistant, Data Science I, Columbia University Aug – Dec 2025

- Support instruction for a 190-student graduate-level data science course, assisting with statistical computing, model interpretation, and reproducible analysis in R
- Lead weekly office hours providing technical guidance on regression modeling, data wrangling, and visualization workflows

Biostatistics Computing Club Member Aug 2024 – Present