

Kelly Zhu

Atlanta, GA | meng.khz@gmail.com | LinkedIn: [kelly-zhu-h41423](#) | Github: [khz123](#)
U.S. Citizen

SUMMARY

Master-level Biostatistician with experience in end-to-end analyses of large, real-world datasets using regression, survival, longitudinal, and machine learning methods. Proficient in R, SAS, and Python; highly motivated team-player praised for exceptional communication of complex results to non-statistical audiences and delivering high-quality work.

SKILLS & CERTIFICATIONS

Programming Languages: R, SAS, Python, SQL/NoSQL, SPSS, MATLAB

Statistical Analysis: Linear, Logistic, Multivariate Regression; ANOVA; Longitudinal and Survival Analysis

Machine Learning: Supervised (regression, k-NN, random forest, XGBoost); Unsupervised (clustering); Deep Learning (neural networks)

Data Visualization & Reporting: Tableau, PowerBI, R Shiny, Excel, Word, PowerPoint, PCA, t-SNE, UMAP

Certifications: SAS Advanced Programming Using SAS 9.4, CITI Human Subjects Protection

EDUCATION

Emory University, Rollins School of Public Health - Atlanta, GA Expected May 2026

Master of Public Health in Biostatistics and Bioinformatics, Data Science Certificate (GPA: 3.96/4.00)

Coursework: Regression Analysis, Machine Learning, Survival Analysis, Clinical Trials, Geographic Information Systems, Longitudinal Data Analysis, Probability Theory

New York University, College of Arts and Science - New York, NY May 2024

Bachelor of Arts in Psychology and Economics, Minor in Data Science

Coursework: Causal Inference, Analytical Statistics, Econometrics, Microeconomics Analysis

PROFESSIONAL EXPERIENCE

Emory School of Medicine, Graduate Research Assistant - Atlanta, GA Jan 2025 - Present

- Conducted end-to-end statistical analyses across 6 ophthalmology studies in R; co-authored 2 manuscripts and contributed to a third, delivering reports and visualizations to clinical teams and hospital leadership
- Fit GLMMs to model retinopathy progression across placebo and L-DOPA dose groups (n=45) over 12 months; applied correlation heatmaps to quantify structure-function relationships
- Analyzed 15,000+ consult records using hypothesis testing, regression, and geospatial analysis to surface decade-long rise in trauma-driven consult volume and regional care demand
- Applied regression modeling to a pediatric HSV keratitis cohort (n=85), identifying corneal scarring and age at presentation as key predictors of visual outcomes

Z&H Technology Consulting, Contract Data Analyst - New York, NY

July 2024 - Present

- Cleaned multi-line cancer therapy datasets, evaluated prediction powers of AI-based predictive models using Karas and XGBoost, and generated recommendations to support client decision-making
- Conducted gene-protein activity analysis using R and estimated information content in large-scale omics data using linear, non-linear tree-based, and deep-learning models

RESEARCH & PROJECTS

Capstone - Modeling Renal Function Outcomes: Precision and Clinical Interpretation | SAS Spring 2026

- Implemented cLDA, LDA, and ANCOVA to model eGFR trajectories across nine time points in a two-arm randomized non-inferiority trial comparing Abatacept and Belatacept
- Evaluated model fit, assumptions, and precision; produced manuscript-ready figures and RTF tables per clinical reporting standards
- Developed an interactive R Shiny app to visualize covariance structure selection in longitudinal models
- Summarized results in manuscript and presented results as a poster at a departmental symposium

Omics Project on Epigenetic Age & Accelerated Aging | R, Bioconductor (methylock)

Spring 2026

- Analyzed DNA methylation data across 482,000+ CpG sites from a [GEO dataset](#) using multi-clock epigenetic age estimation to quantify biological age acceleration
- Modeled relationships between age acceleration and smoking status via multivariate linear regression, controlling for sex; visualized patterns using PCA, heatmaps, and volcano plots

Data Science Project on Cross-Country Caffeine, Alcohol, and Sleep | R

Fall 2025

- Conducted reproducible regression analysis of caffeine, alcohol, and sleep patterns in 10,000 people across 20 countries
- Built containerized workflow with Docker and managed project dependencies using Make and renv, ensuring reproducibility and scalability