

Jonathan Barnes | Data Scientist

Minneapolis, Minnesota | <https://jonathanmbarnes.github.io> | Jonathanmbarnes@outlook.com

PROFESSIONAL EXPERIENCE

University of Minnesota, Minneapolis, MN

Analytics Engineer | Center for Public Health Systems

(May 2025 - Present)

- Established and operate an in-house analytics and data warehousing platform for federal health funding analysis.
- Integrated and standardized 3M+ HHS award records into a unified analytical model supporting evaluation.
- Designed analytical data models to reconcile complex federal funding and organizational structures.
- Engineered modular, logged data pipelines supporting recurring analytical workflows.
- Authored complex Python and SQL queries over large-scale funding data for analytical workflows.
- Delivered analysis-ready datasets and Tableau dashboards communicating HHS funding trends, allocations, and flows.
- Enabled program-level evaluation of PHIG funding allocations using standardized federal award data.
- Developed internal tooling to support controlled execution and refresh of analytical data assets.

Graduate Teaching Assistant | Division of Biostatistics and Health Data Science

(Aug 2024 - May 2025)

- Taught and evaluated applied statistical analysis in R, improving student code quality and statistical reasoning.

Graduate Research Assistant | Medical School

(Oct 2023 - May 2024)

- Designed custom plasma biomarker phenotypes and gene expression features to enable survival modeling.
- Applied Cox proportional hazards regression to model time to cognitive decline using biomarker phenotypes.

University of St. Thomas, St. Paul and Minneapolis, MN

ITS Support Technician Intern & Senior Student Technician | ITS Support Services

(Aug 2021 - Aug 2023)

- Resolved or consulted on 4,500+ IT support tickets across hardware, software, networking, and security.
- Developed a Python application that enabled core ITS service desk workflows and improved operational efficiency.
- Mentored junior student technicians to independently handle complex service desk operations and consultations.
- Imaged and administered software on 1,500+ summer replacement computers, doubling the typical annual volume.
- Assembled an in-house virtual lab to replace AWS EC2 instances, reducing cloud footprint and infrastructure costs.
- Implemented digital signage across service desks and campuses to improve access to real-time IT service information.
- ITS Outstanding Student Award.

EDUCATION

University of Minnesota School of Public Health, Minneapolis, MN

Master of Public Health - Public Health Data Science, GPA 3.73/4.00

(Expected Jan 2026)

- APEx — Pre-Statistics Modules for MPH Students
- ILE – Evaluation of US Federal Spending and Systems.
- President's Student Leadership & Service Award Nominee (2025)

University of St. Thomas, St. Paul, MN

Bachelor of Science - Data Analytics (Biology Domain), Completed while employed full-time

- Minors - Biology (Biochemistry & Computational Biology courses), Applied Statistics.
- Capstone - Development of Obesity and Nutritional Trends Over Time in the United States Using Combined NHANES.

SKILLS

Languages: R, Python, SQL, Bash.

Tools: Shiny, APIs, Git, ArcGIS, Tableau, CI/CD, Microsoft Azure, Docker.

Competencies: Data Visualization, Data Communication, ETL Pipelines, Geospatial Visualization, Longitudinal Data Analysis, Machine Learning, Artificial Intelligence, Relational Databases, Columnar Data Architectures, System Design, Experimental Design, Data Acquisition, Consulting, Health Informatics, Electronic Health Records, HIPAA, PHI, SNOMED, ICD-10.

Statistical Methods: Dimensionality Reduction, Supervised Learning, Unsupervised Learning, Regression, Regularization, Tree Models, Mixed Effects Models, Survival Analysis, Structural Equation Modeling, Natural Language Processing.

PROJECTS

Public Health Federal Funding Evaluation Platform & Dashboard (Grant Supported)

- Developed a modular ingestion and modeling pipeline to handle millions of U.S. federal public health funding records, spanning prime awards, sub-awards, and derived transactions across inconsistent reporting structures.
- Designed reconciliation logic to accurately link awards, recipients, and health departments across divergent reporting hierarchies, identifier schemes, and funding relationships.
- Implemented rules-based normalization to standardize agencies, recipients, locations, and award classifications across fragmented HHS source systems with inconsistent metadata.
- Resolved duplicated entities, missing identifiers, and conflicting award relationships in large-scale federal funding data.
- Optimized analytical storage using DuckDB and Parquet, reducing data size by 75%+ while preserving query performance.
- Materialized analysis-ready datasets and Tableau Hyper extracts for reproducible visualization workflows.
- Built a GUI-based orchestration tool to manage pipeline execution, refresh cycles, and validation.
- Designed dashboards to analyze funding flows, geographic distribution, and program-level spending.
Methods & Tools: Python, Polars, SQL, DuckDB, Apache Parquet, Tableau (Hyper), Data Modeling, Data Pipelines, Data Warehousing, Data Validation, Data Engineering, Data Visualization, Public Health Informatics.

Mapping Misery – Predicting Poor Physical and Mental Health Days in the U.S.

- Developed predictive models for poor physical and mental health days in 3,000+ U.S. counties using R.
- Engineered clusters based on 10 years of trends across 15 metrics using Gaussian Mixture Models (GMM) by transforming variables into 8 level categories with NAs for robust Multiple Correspondence Analysis (MCA).
- Achieved an R^2 of 0.844 (RMSE: 0.263) for physical health days with Elastic Net Regression.
Methods & Tools: Data Pipelines, PCA, MCA, GMM Clustering, Feature Engineering, Longitudinal Data Analysis, Elastic Net Regression, Predictive Modeling, R, Geospatial Visualization, GIS, Data Visualization.

Tech Desk Tools Python

- Built a GUI-based desktop application for the University of St. Thomas Tech Desk to streamline common support tasks such as ticket creation, password resets, system diagnostics, and resource access in use continuously since February 2023.
- Redesigned and modernized legacy AutoHotKey application in Python, improving functionality and future scalability.
- Integrated role-specific admin features for senior student technicians based on assigned Azure roles.
- Deployed the application to all staff and student workers at the University of St. Thomas's Tech Desk.
Methods & Tools: Python, PowerShell, Bash, Dashboards, Data Communication, GUI Development, Active Directory, Azure, Technical Writing, Packaging and Deployment.

Applied Biostatistics Introduction Modules (Grant Supported)

- Designed modular teaching materials introducing probability, inference, and regression through applied public health examples at the request of Public Health Administration & Policy program director
- Developed intuition-focused explanations to support students with apprehension toward statistics.
- Addressed recurring conceptual gaps identified by faculty and me from teaching.
- Emphasized ethical interpretation and justification of statistical choices in applied public health contexts.
Methods & Tools: Biostatistics, Statistical Reasoning, Probability, Curriculum Design, Technical Writing, R, Data Communication, Public Health Education.

Data Inspection Dashboard | <https://jonathanbarnes.shinyapps.io/DescriptiveApp/>

- Developed an interactive tool that allows users to conduct baseline Exploratory Data Analyses (EDAs) without requiring R knowledge and accepts common dataset formats in public health.
- Automates and facilitates data cleaning, NA handling, imputation, and produces a data table with inline distribution plots and summary statistics.
Methods & Tools: R, Shiny, Data Pipelines, Dashboards, Data Visualization, Data Communication, JavaScript.

Biostatistics Reference Guide

- Provides step-by-step instructions and R code for linear regression, logistic regression, Poisson regression, Cox proportional hazards models, survival analysis, ordinal regression, and log-binomial regression.
Methods & Tools: R, Regression Analysis, Cox Proportional Hazard, Logistic Regression, Poisson Regression, Ordinal Regression, Log-binomial Regression, Tidyverse, Technical Writing.

Predicting Child Mortality and Cesarean Delivery Outcomes with Generalized Linear Mixed Models.

- Built Generalized Linear Mixed Models (GLMM) to investigate factors affecting child mortality and cesarean delivery.
- Determined a 1,124.5% increase in the odds of cesarean delivery for hospital births (95% CI: 618.7% - 1,986.3%, $p < 0.001$) — a result consistent with logical expectations.
Skills & Tools: GLMM, GEE, R, Data Cleaning, Data Pipelines, Data Visualization, Feature Engineering, Collaboration.