

Yi Luo

Luo18125@163.com | +86 17600583506

<https://ymo26.github.io/>

Xinyao South Road, Changsha City, Hunan, 410114, China

Commented [A1]: 邮箱之后需要换成我们申请用的 outlook 邮箱

✧ Education

Lafayette College

Easton, PA

Bachelor of Arts in Mathematics and Economics (Minor in Data Science)

09/2022 – 12/2025

Total GPA: 3.5/4.0 Major GPA: 3.54/4.0 Minor GPA: 3.77/4.0

✧ Research

Environmental Determinants of Health: Spatial Analysis of Air Quality, Weather, and Mortality (2000–2023)

Research Assistant (Advisor: Sayorn Chin from Lafayette College)

01/2025 – 05/2025

- Developed a harmonized, longitudinal dataset integrating air quality (EPA), weather (NOAA), and health outcome (CDC WONDER) data across U.S. counties (2000–2023) using automated data retrieval, cleaning, and aggregation workflows in R
- Created comprehensive standardized mortality rate maps and indexes for air quality (PM2.5, ozone) and weather conditions (temperature, humidity, pressure), revealing significant urban-rural, demographic, and temporal disparities
- Conducted robust regression analyses to identify key environmental, demographic, and socioeconomic determinants of mortality, highlighting actionable insights for policymakers to mitigate environmental health disparities and economic burdens, particularly in rural and vulnerable populations

Feature Engineering, AutoML, Metrics Design and Smart Data Visualization

12/2024 – 07/2025

Research Assistant (Advisor: J. Yang from Columbia University)

- Performed empirical data analysis, including data cleaning, missing data imputation, and data visualization
- Generated new features based on existing variables, trained different machine learning models using Scikit-learn and TensorFlow from Python
- Evaluated models using metrics such as their accuracy, precision, and F1 scores; interpreted the results, formed a research paper on them, and published

✧ Project

A Comparative Study on Model Theory and Dataset Performance

01/2025 – 05/2025

- Examined fundamental mechanics of decision trees and random forests, highlighting interpretability, the Gini impurity criterion, and ensemble techniques through illustrative examples of student data.
- Conducted comprehensive evaluations using medium-scale skiing resort data and large-scale income data, analyzing accuracy, precision, recall, and F1 scores to benchmark model performance; concluded that random forests consistently provided greater accuracy and stability in complex and larger datasets

Election Campaign Projections Using Python

09/2024 – 12/2024

- Utilized Python to analyze historical U.S. election data, identifying critical swing states by assessing voter behavior, demographic changes, and margins of victory from previous election cycles
- Developed comprehensive visualizations with Pandas and Matplotlib to highlight key battleground states, employing data cleaning, state indexing, and margin-of-victory computations for accuracy and relevance

Analyzing Urban Heat Islands Using Spatial and Temporal Data in R

09/2024 – 12/2024

- Employed R and spatial analysis packages to visualize building distributions within defined urban buffers, examining urban structures' role in heat retention
- Utilized ggplot2 to create comparative visualizations of day and night temperature differences across different months, highlighting temperature variation patterns in urban areas (Chicago and Berlin)

Lion and Hyena Population Dynamics Modeling and Stability Analysis

01/2024 – 05/2024

- Developed R scripts to visualize the long-term behaviors of single-species and dual-species population dynamics, comparing scenarios of coexistence, competitive exclusion, and extinction events
- Interpreted model outputs to explain species interactions and competition in Maasai Mara National Reserve

Regression Analysis: Exponential and Quadratic Model Comparisons *01/2024 – 05/2024*

- Performed detailed regression analyses (Exponential vs. Power models; General vs. Ordinary Quadratic models), evaluating fit using R^2 and residual diagnostics
- Determined optimal model selection based on comparative graphical analyses, residual distribution, and quantitative metrics, highlighting scenarios where each model type excels

Tumor Growth Analysis: Logistic vs. Gompertz Model Evaluation *01/2024 – 05/2024*

- Investigated tumor growth dynamics using Logistic and Gompertz differential equation models, identifying equilibrium states, stability conditions, and analytical solutions
- Compared model performance through phase portraits, growth curves, and Residual Sum of Squares, demonstrating the Gompertz model's superior fit to observed data

Assessing Long-Term Cancer Risks Associated with Early Smoking Initiation *01/2024 – 05/2024*

- Evaluated the relationship between smoking intensity, age at smoking initiation, education, and risk of cancer diagnosis by conducting logistic regression analysis; incorporated sociodemographic factors for robustness
- Emphasized early intervention as critical to lowering long-term cancer risks and proposed underage smoking policy recommendations

Economic Indicator Analysis, Predictive Modeling, and Statistical Evaluation *09/2023 – 12/2023*

- Performed detailed exploratory data analysis including data loading, missing value imputation, and preliminary statistical examination of Iowa economic indicators
- Identified key variables through stepwise feature selection; created polynomial features and standardized data for enhanced model accuracy; constructed a linear regression model in R
- Assessed model performance using cross-validation and statistical diagnostics (RMSE, R^2 , ANOVA F-test)

✧ **Skills & Interests**

Languages: Mandarin (Native), English (Proficient)

Computer skills: R, Stata, Python

Hobbies: 4 years of Sanda (participated in small-scale competitions), 6 years of skiing (expected to obtain a coaching certificate)