Howell Lu

 Howell Lu | in Howell Lu | 🌐 Personal Website | ✉ hl4631@nyu.edu | 📱 917-605-9173

Summary

Interested in statistical computing, sequential data analysis, and applying the former to the latter.

Research and Professional Experience

Research Assistant | New York University   Aug 2025 - Present

– Researched under Dr. Hill to investigate and evaluate the feasibility, implications and limitations of GPU-accelerated BART (Bayesian Additive Regression Trees) and other parallel BART engines.

– Researched under Dr. Scott to analyze and investigate the applicability of CFDA (Categorical Functional Data Analysis) on messy sequential data.

Data Analyst | Mega ICBC   Sept 2023 - August 2025

– Architected terabyte-scale ETL pipelines and automated data interrogation frameworks using Python and MS SQL Server, engineering a parallel compliance workflow to execute end-to-end testing for real-time SWIFT and Fedwire integrations.

– Developed statistical validation frameworks utilizing Chi-Squared, Mann-Whitney U, and Kolmogorov-Smirnov tests to quantify distributional shifts between legacy SWIFT MT and converted MX formats, successfully identifying critical data anomalies.

– Served as Technical Lead for the ISO 20022 migration, translating complex data discrepancies into actionable insights for senior management and external vendors to drive workflow optimization.

Teaching Assistant / Grader | New York University   May 2022 - August 2022

Designed assessments, solutions, rubrics and lab materials using LaTeX, conducted lab lectures and held office hours for two weekly lab sections

Research Assistant | New York University/Ripple   May 2022 - August 2022

A collaborative initiative pairing NYU graduate students with industry partners to solve pressing, real-world research challenges. Researched outlier behavior and arbitrage opportunities between 21 different crypto-markets.

Education

2025 - Present   **Post-Master's Graduate Coursework** at **New York University**
Objective: Advanced theoretical preparation & methodological research.

Fall 2025:
- **Graduate Research / Independent Study** (Advisor: Prof. Jennifer Hill)
  Topic: Parallelizing Bayesian Trees: Algorithmic Insights from GPU BART.
- **Multi-Level Modeling** (APSTA-GE 2042) & **Consulting Practicum** (Advisor: Prof. Marc Scott)

Spring 2026 (Expected):
- **Essentials of Probability** (Measure-Theoretic) – *NYU Courant MATH-GA 2901.*
- **Applied Real Analysis** (Athabasca Univ. MATH 370) – *Completion: February 2026.*
- **Advanced Causal Inference**
- **Inference**

2021 - 2023   **M.S. in Data Science** at **New York University**   (GPA: 3.73/4.0)

Selected Coursework: Advanced Python for Data Science, Causal Inference, Machine Learning, Optimization and Computational Linear Algebra, Introduction to Computer Vision, Text as Data, Big Data, Digital Currency.

2019 - 2021   **B.A. in Applied Mathematics** at **York University**   (GPA: 3.79/4.0)

Selected Coursework: Differential Equations, Regression Analysis, Mathematical Statistics, Stochastic Processes.

2013 - 2017   Bachelor's Degree (BBA) at **York University**   (GPA: 3.53/4.0)

PROJECTS

My Relationship Future   Link

Uses Finite Mixture Models to classify an individual's past relationship into one of 11 components, which are used to predict future relationship history, 47% higher accuracy compared to XGBoost

Is My Degree Worth It?   Link

A ML model created from data from the DOE's College Scorecard, the Census ACS, and the Bureau of Labor's NLSY dataset to create income, ROI and NPV projections from college, major and debt taken

My College Voyage   Link

LLM enhanced college browsing and data visualization site. Mostly designed to see how well text could be converted into SQL Queries.

SimplyIncluded   Link

Created a particle swarm optimization algorithm to generate optimal exclusion criteria for multidimensional problems.

Crypto Anomaly Detection with Ripple   Link

A joint research project between NYU's Center for Data Science and Ripple that utilized deep reinforcement learning to label anomalous cryptocurrency data using a multitude of features.

Lexicon based Similarity Analysis between Population Groups   Link

Scraped 396,041 reddit comments from the multiple demographic groups on the popular social media website Reddit, and applied NLP algorithms such as WordScore to determine the distance between different demographic groups

Metropolis Algorithm for Expectations   Link

Proved and extended the Metropolis algorithm with a probability distribution $\mu = 1$ to the Travelling Salesperson Problems

SKILLS

Statistical Methodology	High-Dimensional Inference ($p \gg n$), Causal Inference, Bayesian Statistics, Convex Optimization, Survival Analysis, Longitudinal Data, Multi-level Modeling.
Machine Learning	Deep Learning (Transformers, CNNs), Reinforcement Learning, Computer Vision, Large Language Models (NLP). Unsupervised Learning
Languages & Computing	Python, R (Advanced), SQL, Linux/Bash, High-Performance Computing (HPC). PyTorch, TensorFlow, JAX, CUDA