Tiffany (Yifan) Hu

Department of Statistics and Data Science, UCLA, Los Angeles, California

eve9031bruin@g.ucla.edu 310-905-1233 github.com/Tiff-Hu

Research Interests

- Statistical and Machine Learning for social, health, and behavioral systems
- Optimization
- Natural Language Processing for scientific and policy applications
- Causal Inference
- Bayesian Statistics
- Computational Social Science
- Scalable Probabilistic Modeling
- High-dimensional and structured data modeling
- Network and Relational Data Analysis

Related Coursework

- **Mathematics:** Multivariable Calculus; Linear Algebra; Differential Equations; Optimization.
- **Statistical Theory & Methods:** Probability and Statistics; Linear Models; Design and Analysis of Experiment.
- **Computation & Simulation:** Computational Statistics; Monte Carlo Methods.
- **Machine Learning & Data Science:** Statistical Learning and Data Analytics; Data Mining; Text Mining.
- **Networks, Bayesian & Causal Inference:** Statistical Analysis of Networks; Applied Bayesian Social Statistics; Causality.
- **Professional Practice:** Statistical Consulting.

Education

University of California, Los Angeles (UCLA) **Sep. 2021 – Sep. 2025**

B.S. in Statistics and Data Science *Los Angeles, CA*

GPA: 3.98/4.00

Honors & Awards

- **UCLA Dean's List**: The Dean's Honors List recognizes high scholastic achievement in any one term. **2021 – 2022 & Spring 2025**
- **Magna Cum Laude**: The College and schools award Latin honors according to overall grade-point average (GPA) at graduation. **Class of 2025**

Publications

Journal Papers

3. **Contributing author.** *Global health and climate benefits from walking and cycling infrastructure. PNAS* (2025). https://doi.org/10.1073/pnas.2422334122.
2. **Co-author.** *Developing an LLM-driven phenotyping and treatment-outcome Bayesian model for children with epilepsy in Uganda.* (Manuscript in preparation; targeted for submission to a leading epilepsy journal/conference).
1. **Co-author.** *Reinforcement learning for decision-making under uncertainty.* (Manuscript in preparation; targeted for submission to IEEE).

Patents & Inventions

- **Co-inventor.** *Bridge Icing Monitoring, Early Warning, and Mitigation System.* Chinese Patent Application: CN202310558553.8. Developed a real-time hazard detection system integrating distributed meteorological and infrared sensors with a multivariate signal-processing pipeline. Designed probabilistic early-warning thresholds to infer icing risk and contributed to automated actuation workflows for de-icing deployment.
- **Co-inventor.** *Self-cleaning Multilayer Radiative Cooling Film.* Chinese Patent Application: 202511190946.3. Engineered a multilayer hybrid nanomaterial structure for passive radiative cooling, incorporating PVDF-HFP fiber networks, fluorinated SiO_2 nanoparticles, and PDA-reinforced interfaces. Evaluated performance using statistical characterization of emissivity (8–13 μm), durability, and hydrophobicity, and contributed to data-driven optimization of material layering.

Labs

Practical Causal Inference (PCI) Lab **Sep. 2024 – Present**

Department of Statistics and Data Science, UCLA

- Worked on methodological and applied causal inference projects, including treatment effect estimation, policy evaluation, and scalable implementations of double machine learning and Bayesian hierarchical models.
- Participated in group discussions on transportability, identifiability, and inference under limited overlap.

Relational Data Learning Group **Sep. 2024 – Present**

Department of Statistics and Data Science, UCLA

- Conducted research on social networks and relational data, focusing on robust Bayesian inference, network sampling, presence-only data, and differential privacy for network models.
- Contributed to projects on:
  - Non-probability surveys and presence-only data
  - Balance testing in social networks
  - Robust Bayesian inference for intractable likelihoods
  - Differential privacy in network models
  - Applications of Graph Neural Networks in social statistics
  - Algorithmic bias, model validity, and inference under non-probability sampling
- Participated in presentations and collaborative work on latent class models, synthetic network data, and causal inference in networked settings.

Other Research Experience

Research Assistant **Apr. 2025 – Jun. 2025**

UCLA Anderson School of Management

- Developed automated matching algorithms in Python and structured tabular data to process 40,000+ financial records, achieving 73% accuracy using text normalization, Jaccard similarity, and Levenshtein distance.
- Supported financial research analysis by developing record-linkage algorithms for noisy, high-dimensional tabular data.
- Built multi-dataset integration workflows by combining master databases and reference lists through programmatic lookup tables and cross-validation algorithms to ensure data integrity.
- Implemented a tiered data validation system with certainty classification (Level 1/2) and automated error-detection protocols to verify accuracy.

Researcher **Feb. 2025 – Apr. 2025**

UCLA Luskin School of Public Affairs

- Designed and implemented Bayesian hierarchical models in Stan and Python to analyze the impact of instrumental variables on population density trends.
- Integrated GIS and demographic data into a reproducible spatial analysis pipeline.
- Collaborated with policy researchers to translate results into actionable insights.
- Optimized model convergence and diagnostics, improving scalability by 40%.

Software Engineer **Dec. 2024 – Mar. 2025**

UCLA Anderson School of Management

- Developed a full-stack application integrating React, Flask, and AWS for chronic disease data analysis, incorporating machine learning models.
- Partnered with MBA researchers to translate complex healthcare datasets into strategic business insights through interactive dashboards.
- Applied iterative prototyping and scalable deployment of data-driven systems to refine pipelines and improve real-time data processing efficiency.
- Integrated predictive modeling workflows into the application, supporting exploratory analysis of chronic disease risk.

Data Research Team Lead **Oct. 2024 – Sep. 2025**
Department of Economics, UCLA

- Led a team analyzing cognitive impairment trends across multiple countries for the International Long-Term Care Project, contributing to aging-related policy research.
- Oversaw collection and cleaning of multi-source datasets, applying advanced EDA in Stata to uncover key insights.
- Integrated R and Python into statistical workflows, reducing manual processing time by 30% and improving model performance.
- Produced publication-ready tables and figures, translating raw data into impactful outputs for researchers and stakeholders.

Data/AI Researcher **Sep. 2023 – Nov. 2023**
City of Palo Alto

- Developed predictive models to optimize city services, applying advanced statistical methods and machine learning to infrastructure datasets.
- Automated extraction, transformation, and storage workflows, reducing reporting time by 37%.
- Built interactive dashboards in Tableau and Power BI to make urban analytics accessible to policymakers and department heads.
- Partnered with city officials and cross-functional teams to ensure technical findings informed actionable municipal policies.

Data Assistant **Sep. 2022 – Dec. 2023**
Department of Mathematics, UCLA

- Automated grading workflows in Python, reducing manual workload by 26% and improving efficiency for 50+ students.
- Analyzed and visualized student performance data, generating insights that informed curriculum adjustments.
- Built scalable pipelines with Python, R, and Excel, ensuring accuracy through robust quality checks.

Industry / Applied Experience

Data Analyst **Aug. 2025 – Oct. 2025**
UCLA Alumni Affairs

- Developed data integration between UCLA One (Graduway) and Salesforce Education Cloud by leveraging the Graduway REST API and Salesforce sandbox environment.
- Performed field mapping of Graduway datasets, including user profiles, mentorship records, and feed activity, into Salesforce Contacts and related objects, ensuring schema compatibility.
- Validated integration in Salesforce sandbox through test data loads, API call debugging, and iterative refinement of mapping rules.

Project Manager **Jan. 2025 – Mar. 2025**
UCLA Alumni Affairs

- Led a cross-functional team in designing an interactive, self-service alumni geographic visualization tool using ArcGIS and Python.
- Developed a prototype for dot-density mapping with filtering and drill-down capabilities to enhance usability for non-technical users.
- Created a comprehensive implementation plan addressing data privacy, system reliability, and user experience.
- Delivered a final pitch deck highlighting key project deliverables and long-term scalability strategies for Alumni Affairs staff.

Data Analyst Intern **Apr. 2024 – Jul. 2024**
Ciba Health

- Developed data/metadata models using statistical and data mining techniques to improve treatment methodologies and decision-making. Utilized SQL, Python, and Looker Studio to generate business-critical reports.
- Collaborated with patient care teams, healthcare providers, and engineers to ensure insights were actionable and integrated into clinical strategies.
- Translated client requirements into technical ETL specifications, implementing data cleansing to maintain integrity and reliability.

Software Engineer Intern **Oct. 2023 – Dec. 2023**
NeuroLeap

- Designed and implemented scalable software solutions aligned with business needs, collaborating with cross-functional teams to define and deliver new features.
- Applied diverse frameworks and languages (C++, Python, JavaScript) to create efficient systems.
- Maintained and optimized existing software, conducting trial runs and quality checks to ensure reliable output.

Data Analyst Intern **Jul. 2023 – Oct. 2023**
PricewaterhouseCoopers (PwC)

- Performed exploratory and predictive statistical modeling for client datasets using Python and Tableau.
- Automated ETL workflows, reducing data preparation time and enabling faster quantitative analysis.
- Built forecasting models to evaluate KPIs and inform strategic recommendations.

Other Projects

Text-based Financial Market Prediction

Constructed a pipeline to scrape and preprocess financial news and social media with NLP (TF-IDF, BERT embeddings). Combined sentiment features with stock price data and trained ML models (LSTM, Random Forest) for trend prediction. Emphasized interpretable modeling by analyzing feature importance and uncertainty in predictions. Evaluated models with accuracy, F1, and AUC-ROC; improved performance via hyperparameter tuning, PCA, and feature engineering.

Bayesian Analysis on Household Dietary Diversity

Applied Bayesian Poisson and Negative Binomial regression to study socio-economic effects on dietary diversity in Tanzania. Improved model robustness using Bayesian feature selection and shrinkage priors (e.g., LASSO priors), generating insights for nutritional epidemiology and policy.

U.S. Presidential Election Voting Predictors

Analyzed demographic influences on voting behavior using ensemble models (XGBoost, LightGBM) and Support Vector Machines. Leveraged model stacking to improve prediction accuracy and minimize overfitting, emphasizing advantages of diverse base model utilization.

Technical Skills

- **Statistics & Machine Learning**
  - Bayesian Inference: Stan, rstan, rstanarm, CmdStanPy, coda, loo
  - Causal Inference: MatchIt, dagitty, SuperLearner, ivreg
  - ML/DL: Pandas, NumPy, scikit-learn, Matplotlib, Seaborn, TensorFlow, PyTorch, Keras, Transformers, Hugging Face, JAX, VLLM
  - Spatial Modeling: GIS-based data integration (ArcGIS, QGIS)

- **Programming:**
  - Python, R, JavaScript, MATLAB, C++
- **Network Analysis:**
  - sna, NetworkX, igraph
- **Data & Systems:**
  - SQL, Git, Docker
  - Tableau, Looker Studio, PowerBI