waiyuamandang@gmail.com
amanda-ng518.github.io/
amanda-ng518
amanda-ng-2323591a6

Wai Yu Amanda Ng

Skills

- 3+ years of experience in data analysis across interdisciplinary projects
- proficient in data analysis, visualizations, stakeholder focused reporting
- Experience working within a Linux environment and HPC environment (supercomputer).
- Experiences working in collaborative teams, strong problem-solving and critical thinking skills, attentive to details.
- Languages: English, Cantonese, Mandarin
- Tools: SQL, R, Python (Pandas, NumPy, Matplotlib, PyMC, Scikit-learn, Plotly), Excel, Tableau, Power BI, Git Version Control, QGIS

Education

Sep 2025 – Apr 2026 (expected)
MSc in Statistics, *University of Toronto*, Toronto, ON, Canada
Courses: Applied Multivariate Analysis, Methods of Applied Statistics, Special Topics in Data Science, Survival Analysis

Sep 2021 – Apr 2025
H.B.Sc in Statistical Science (Specialist), Mathematics (Major), Health Studies (Focus), *University of Toronto*, Toronto, ON, Canada **cGPA: 3.94/4.00**
Courses: Applied Statistical Methods, Applied Spatial Analysis, Time Series Analysis, Applied Bayesian Statistics, Surveys, Sampling and Observational Data, Statistical Consultation
Awards: University of Toronto Excellence Award (2024), Eleventh Annual Canadian Statistics Student Conference Undergraduate Level First Prize (2023)

Professional Experiences

Sep 2025 – Current
Statistical Learning Designer, *University of Toronto Human Biology Program*
- Built 4 auto-testable R assignments for a cohort of 400 students, cutting manual grading workload by 90% and accelerating feedback turnaround.
- Designed interactive LearnR and LearnSQL modules with RShiny, reducing student debugging requests and enhancing self-directed learning efficiency.
- Standardized version-controlled releases on Git, enabling multi-developer collaboration and achieving zero student-reported issues across all instructional materials.

Sep 2024 – Apr 2025
Data Visualization Analyst, *University of Toronto Human Biology Program*
- Analyzed learning resource engagement data for 1,000+ course enrollments in Excel, identifying behavioral patterns linked to final grades and informing assessment redesign.
- Developed 100+ Tableau dashboards visualizing key performance metrics (discussion board activity, lecture videos completion, early-term quiz grades), enabling instructors to identify at-risk students by Week 3.
- Communicated analytical findings into data-driven recommendations that improved assessment design and feedback practices, enabling instructors to address early-term performance gaps and refine course delivery pacing.

Jun 2024 – Apr 2025
Clinical Education Research Analyst, *Centre for Addiction and Mental Health*
- Analyzed 90+ VR cybersickness surveys in R to evaluate how physical movement affects user discomfort in medical training environments, contributing to a peer-reviewed publication and two conference posters.
- Applied descriptive and non-parametric statistics analyses to identify a positive movement–nausea relationship, revealing key factors, such as frequent turning contributing to increased cybersickness in clinical training.
- Collaborated with clinical educators to translate findings into evidence-based movement thresholds, informing VR and desktop module designs to balance instructional effectiveness with user comfort.

May 2023 – Aug 2023
Data Analyst Intern, *Canadian Urban Institute*
- Developed regression models in R to analyze nationwide household carbon-emission patterns, identifying key drivers of high-emission regions and guiding targeted interventions for a national sustainable city-planning project.
- Designed an interactive scenario-modeling tool illustrating how urbanization and housing composition influence greenhouse-gas emissions, enabling stakeholders to test policy options on a dashboard.
- Produced 300+ geospatial maps in QGIS and integrated them with statistical findings, uncovering location-specific emission drivers and delivering actionable planning recommendations.

Oct 2022 – May 2023
Computational Social Science Researcher, *University of Toronto Population Well-being Lab*
- Applied unsupervised machine learning methods, Gaussian mixture model, k-means, and hierarchical clustering to 150K+ survey observations, uncovering subpopulations based on life satisfaction and demographics.
- Created four accessible statistical method tip-sheets, improving onboarding efficiency and strengthening methodological literacy across the lab.
- Collaborated with faculty and graduate researchers to design a poster on clustering workflows and result interpretation, earning first prize at Statistic Society of Canada conference.

Statistics Research/ Course Projects

Oct 2025 – Dec 2025 **Uncovering Calendar-Based Structures Through Clustering of Daily Traffic Data**
- Applied PCA and UMAP in R to a year-long traffic dataset across 26 locations, uncovering recurring traffic pattern clusters tied to school calendars, holidays, and seasonal routines.
- Created visualizations and prepared a report informing transportation planning and infrastructure maintenance decisions.

Oct 2025 – Dec 2025 **Fatal US School Shootings (1999-2025)**
- Modeled fatality risk in 400+ US school shootings using logistic regression in R, highlighting actionable risk indicators including shooting intent, absence of non-shooter injuries, and shooter connection to the school community.
- Built a reproducible GitHub repository documenting all data cleaning, modeling, and analysis steps to support transparency and collaboration.
- Synthesized findings into a report informing safety and emergency preparedness strategies.

Sep 2025 – Dec 2025 **Statistical Inference for Conditional Spatialtemporal Correlation in Brain MRI Data**, *Graduate Research Project*
- Pioneered analysis framework for assessing genetic heritability of intermodal coupling using multi-modal brain MRI data, generating methodological insights for advanced neuroimaging inference.
- Conducted 50+ HPC simulation experiments and applied adaptive clustering techniques with parallel computing to validate proposed model's Type I error control and statistical power.
- Summarized findings in a report advancing robust methodological neuroimaging statistical approaches.

Feb 2025 – Apr 2025 **Modeling Cocoa Futures Prices: A Statistical Analysis of Ghana's Cocoa Production**
- Integrated 3,700+ daily observations across four datasets, performing extensive cleaning to produce a consistent, analysis-ready dataset, enabling reliable comparison across climate, economic, and price variables.
- Developed ARIMA and Random Forest forecasting models in R (astsa, forecast, randomForest), engineering lagged features and covariates that improved predictive accuracy by capturing both linear and nonlinear drivers of cocoa price volatility.
- Conducted full diagnostics and authored a report translating model findings into actionable insights, highlighting need for additional economic and production data in future forecasting systems.

Oct 2024 – Dec 2024 **Transit and Crime: Investigating Assault Clusters Around Toronto's TTC Subway Lines**
- Analyzed 200k+ Toronto assault incidents using kernel density estimation, point process models, and HDBSCAN clustering, identifying assault hotspots within 100–500m of TTC subway lines, especially around key transfer stations.
- Developed an interactive geospatial report translating complex spatial patterns into map-based insights, providing location recommendations for targeted patrol allocation and station lighting improvement.

Sep 2024 – Apr 2025 **Quercus Engagement as a Predictor of First-Year University Student Dropout Risk: A Logistic Regression Analysis**, *Statistical Consulting Course*
- Preprocessed 3,000+ student engagement records by resolving missingness, standardizing features across heterogeneous time windows, and designing multi-week activity metrics, enabling clear identification of early disengagement patterns.
- Built and validated logistic regression models to assess the relationship between Quercus engagement and dropout risk, identifying disengagement in first 4 weeks as a critical retention indicator.
- Delivered client-focused reports and presentations with clear visualizations and accessible interpretations, directly informing revisions to early-term outreach and academic support strategies.

Sep 2024 – Apr 2025 **Fair Machine Learning in Healthcare**, *Undergraduate Reading Course*
- Developed R/Python data pipeline (dplyr, Pandas, PyTorch) on 40,000+ EHR records from MIMIC-III database, producing a clean, standardized dataset that improved downstream model stability and interpretability.
- Developed embeddings for 6,000+ medical concepts and applied feature selection algorithms, yielding clinically relevant features that drove over 85% accuracy in subsequent prediction tasks.
- Trained logistic regression and random forest models for depression diagnosis, identifying key demographic disparities, such as race and insurance type, linked to label bias through fairness analysis.
- Built a reproducible workflow and curated a GitHub repository, enhancing research transparency and replicability.

May 2024 – Current **Aspect of Robust Regression Analysis**, *Undergraduate Research Project*
- Developed and evaluated 4 parameter-estimation approaches for high-dimensional, heavy-tailed data, producing methodological insights that improved estimation accuracy over traditional regression methods.
- Implemented and optimized algorithms in R/Python, resolving non-convergence issues and increasing computational efficiency for large-scale simulations.
- Delivered a departmental research talk and released an R package with 20+ reusable functions, strengthening reproducibility and statistical communication.

Feb 2024 – Apr 2024 **Investigating The Factors Affecting Cycling Accident Severity: A Bayesian Hierarchical Mixture Model Approach**
- Conducted EDA on 800K+ accident records in Python (Pandas, NumPy, Matplotlib, Seaborn), identifying key temporal and environmental predictors, such as evening and dim environment, of severe injuries.
- Restoring 60K+ incomplete records by engineering Bayesian missing-data imputation pipeline (Scipy) using posterior predictive sampling, reducing model bias by 18% in downstream severity estimates.
- Built a Bayesian hierarchical mixture model in PyMC incorporating time-of-day structure, spike-and-slab selection, and demographic mixture components, achieving predictive accuracy gains over baseline logistic regression.

Sep 2023 – Apr 2024 **An Agent-Based Simulation Approach to Investigate the Effect of Decreasing Birth Rates and the Efficacy of Potential Solutions**, *Undergraduate Research Opportunity Program*

- Partnered with environmental researchers to design 20+ agent-based simulations modeling population dynamics under declining birth rates, identifying two policy levers that improved long-term population stability.
- Synthesized 30+ multidisciplinary sources into an annotated bibliography, strengthening model assumptions and improving simulation realism.
- Conducted time-series and regression analyses in Python (SciPy) to quantify intervention effects, revealing short- vs long-term trade-offs that informed policy recommendations, resulting in a poster presentation.

Health Studies Course Projects

Feb 2024 – Apr 2024 **Effectiveness of Usage of Virtual Reality in Raising Awareness about Cannabis Misuse Issues among Teenages**, *HMB342*

- Designed a randomized controlled trial study comparing virtual reality–based education with traditional seminars to evaluate effectiveness in raising awareness of cannabis misuse among Canadian high school students.
- Developed evaluation metrics and statistical analysis plan (pre/post surveys, T-tests, knowledge retention, behavioral intentions) to measure intervention impact on awareness and future cannabis use.
- Applied health promotion and program design frameworks to integrate emerging technologies into adolescent substance misuse education, highlighting cost-effectiveness and policy implications for public health strategies.

Feb 2023 – Apr 2023 **Food Insecurity Intervention Program Planning**, *HST330*

- Conducted root cause analysis and developed a program proposal to address food insecurity among Toronto low-income families, modeling intervention design on the Supplemental Nutrition Assistance Program with a digital e-coupon system.
- Designed a quasi-experimental evaluation plan, including baseline and post-intervention measures (Food Insecurity Experience Scale, BMI, Healthy Eating Index, stress levels) to assess program effectiveness.
- Created a logic model and theory of change framework to outline expected outcomes, health implications, and long-term policy applications, demonstrating strong skills in program evaluation and health policy analysis.

Oct 2022 – Dec 2022 **Proposal to Address Teenager Substance Misuse Problem**, *HST209*

- Conducted comprehensive analysis of risk factors (family influence, peer pressure, media exposure, mental health) contributing to teenage substance misuse, synthesizing evidence-based research into actionable insights.
- Designed and evaluated multi-level interventions including individual counseling, family therapy, in-school workshops, and peer-led support communities to promote adolescent health and reduce substance misuse.
- Developed health promotion strategies grounded in empowerment and community engagement, aligning interventions with public health models to improve long-term outcomes for at-risk youth.

Extracurricular/ Other Experiences

Sep 2023 – Current **Teaching Assistant and Learning Support**, *University of Toronto Department of Statistical Sciences and Data Sciences Institute*

- Led tutorials and office hours across multiple statistics courses in person and online, strengthening undergraduate students and industry professionals understanding of R, Python, SQL, and statistical concepts.
- Supported high-school outreach events, Florence Nightingale Day, translating advanced statistical ideas into engaging, accessible explanations.

Sep 2022 – Dec 2023 **Statistical Consultant**, *University of Toronto Data Sciences Café*

- Provided on-demand statistical guidance to 20+ interdisciplinary researchers, transforming broad research questions into statistically rigorous analysis plans.
- Diagnosed design limitations and recommended tailored analytical approaches, helping researchers avoid common biases and strengthen study validity.
- Authored 5 reports on advanced statistical topics, translating complex concepts and coding methods into clear guidance that enhanced statistical literacy among non-statistician researchers.

Sep 2022 – Apr 2023 **First-year Learning Community Assistant Peer Mentor**, *University of Toronto*

- Facilitated 12 sessions supporting 20+ first-year students in academic planning, study strategies, and transition management.
- Built an inclusive learning community and coordinated discussions with faculty and library staff to enhance student engagement.
- Managed the online Quercus course page to facilitate remote communication and resource access.

May 2022 – Sep 2022 **Bilingual Client Care Representative**, *Telus Health*

- Served as first point of contact for 100+ daily client inquiries, delivering bilingual (English/Chinese) support that improved satisfaction and reduced escalations.
- Assessed client needs, identified the purpose of each inquiry, and provided tailored service recommendations, ensuring clients felt informed, supported, and guided toward the right solutions.
- Coordinated service delivery by booking appointments and routing callers to the appropriate service-specific teams, contributing to faster resolution times and smoother operations.