

Xinyue Wu

📍 Pullman, WA ✉ xinyue.wu@wsu.edu 🌐 xinyue-wu

Education

Washington State University

Aug 2022 – May 2027

- Ph.D. in Sociology: Qualifying exams passed in April 2024
- M.S. in Computer Science
- M.S. in Statistics achieved in May 2025

The University of Tokyo

Sep 2020 – Aug 2022

- M.A., Interdisciplinary Studies of Social Sciences
- Completed Program: Data Science School

Wuhan University

Sep 2016 – May 2020

- B.A., History

Research Interests

Population Health, Demography, Causal Inference, LLMs

Publications

- Amorim, Mariana, Benjamin Brunjes, Gregg Colburn, Crystal Hall, Julia Karon, Sharon Kioko, Kayla Lendy, and **Xinyue Wu**. 2025. *Home Equity Sharing Agreements in Washington State (Policy Report)* [🔗](#). Washington State Department of Financial Institutions, Olympia, WA.
- Kioko, Sharon, **Xinyue Wu**, and Mariana Amorim. 2025. *Earned Wage Access Financial Services (Policy Report)* [🔗](#). Washington State Department of Financial Institutions, Olympia, WA.
- Wang, Zeqiang, Yuqi Wang, **Xinyue Wu**, Chenxi Li, Yiran Liu, Linghan Ge, Zhan Yu, Jiaxin Shi, and Suparna De. 2025. *PMPO: A Self-Optimizing Framework for Creating High-Fidelity Measurement Tools for Social Bias in Large Language Models* [🔗](#). In *Proceedings of IJCNLP–AAACL 2025 (Main Track)*. Association for Computational Linguistics.

Manuscripts in Progress

- **Wu, Xinyue**, and Mariana Amorim. “Return Migration and Educational Investment in Children in China: Household Educational Expenditures, Time Inputs, and Parental Coresidence.” *Revise & Resubmit, Demographic Research*.
- **Wu, Xinyue**. “Disparities in the Education-Health Gradient Across Race and Nativity in the United States.” *Under Review, Demography*.

Research Experience

Research Assistant, Dr. Mariana Amorim

Pullman, WA

Washington State Department of Financial Institutions HESAs Project

Aug 2024 – Aug 2025

- Built automated web-scraping pipelines in **Python** (Selenium, BeautifulSoup) to collect and clean data on $N = 2,150$ properties engaged in Home Equity Sharing Agreements (HESAs).
- Automated geocoding via the U.S. Census Geocoder API; implemented name parsing and entity resolution using an LLM API; mapped spatial distributions in **R**.
- Integrated and crosswalked data from the Home Mortgage Disclosure Act (HMDA), IPUMS National Historical Geographic Information System (NHGIS), and American Community Survey (ACS) in **Stata** to develop tract-level socioeconomic and demographic indicators.
- Designed a sampling frame of 25,000 single-family residences and condominiums across six Washington counties to append individual-level socioeconomic characteristics from Experian for profiling HESA users.
- Estimated Bayesian conditional autoregressive (CAR) models in **R** to identify spatial and temporal patterns

in HESA usage and inform tract-level policy analysis.

- Translated findings into accessible insights for policymakers and stakeholders.

Research Assistant, Dr. Mariana Amorim

Pullman, WA

Changes Over Time in Investments in Non-Resident Children

Jan 2024 – Nov 2024

- Integrated 1980–2019 Consumer Expenditure Survey microdata with contextual datasets (BLS unemployment rates, ACS/Census demographics, IRS Gini coefficients) to build a longitudinal state-year panel on household spending for non-resident children.
- Engineered individualized Gini coefficients in **R** to quantify within-household expenditure inequality and flag outliers for further study.
- Conducted **Blinder–Oaxaca decompositions** to isolate drivers of financial-transfer disparities and created demographic trend visualizations in **Stata**.
- Performed **age–period–cohort (APC) analyses** to surface generational shifts in financial investments, informing evidence-based narratives on changing family dynamics.

Data Science Program Practice

Tokyo, JP

The University of Tokyo & GMO Payment Gateway

May 2022 – Aug 2022

- Applied machine learning (e.g., neural networks) and statistical models (e.g., logistic regression) in **Python** to detect fraudulent transactions, achieving 70.8% accuracy on held-out data.
- Engineered features using one-hot encoding and visualized relationships via heatmaps to identify patterns in high-dimensional categorical data.
- Prepared formal analytic reports and presented findings to stakeholders at GMO Payment Gateway, supporting data-driven fraud prevention strategies.

Work Experience

Data Analyst

Pullman, WA

Institutional Research, Washington State University

Aug 2025 – Present

- Built automated, reusable ETL pipelines for university-wide datasets using **SAS macros** and **SQL**, implementing validation checks (schema, row counts, key uniqueness, value ranges) to improve data integrity.
- Built and maintained longitudinal crosswalks for academic plans/majors and awarded degrees across terms, including consistent cohort/retention tracking.
- Collaborated with cross-functional stakeholders to translate incomplete or shifting reporting requirements into reproducible metrics; delivered parameterized, version-controlled code, documentation, and Excel-ready exports aligned with internal surveys and annual reporting.

Teaching Experience

Teaching Assistant, *Social Psychology of Communication, WSU*

Spring 2024

Lab Instructor, *Quantitative Techniques in Sociology, WSU*

Fall 2023

Teaching Assistant, *Social Problems, WSU*

Fall 2023

Instructor, *Introduction to Sociology, WSU*

Summer 2023

Teaching Assistant, *Quantitative Techniques in Sociology, WSU*

Spring 2023

Teaching Assistant, *Introduction to Sociology, WSU*

Fall 2022

Teaching Assistant, *Society and Technology, WSU*

Fall 2022

Teaching Assistant, *Introduction to Python Programming, The University of Tokyo*

Spring 2022

Conference Presentations & Workshop Participation

International Advanced Studies in Demography (IDEM) – Max Planck

Nov 2025

Institute for Demographic Research

- Acquired advanced training in collecting and analyzing unstructured web data, applying methods for bias correction in non-representative samples, and nowcasting demographic indicators using computational techniques.

Summer Institute in Computational Social Science – NYU Shanghai

Jun 2025

- Participated in lectures and hands-on workshops on text analysis, web scraping, digital field experiments,

machine learning, and research ethics.

- Collaborated on and presented the group project “From Style to Substance: Linguistic Features and Quality Evaluation of Human vs. LLM Concepts in Sociology.”

Panel Study of Income Dynamics (PSID) Data User Workshop *Jun 2025*

- Gained hands-on experience in accessing, cleaning, and analyzing longitudinal household survey data.

Population Association of America, Annual Meeting *Apr 2025*

- Education–Health Gradient Disparity Across Race and Nativity in the United States

Population Association of America, Annual Meeting *Apr 2024*

- Return Migrant Children in Contemporary China: Estimating Impacts of Return Migration on Children’s Educational Aspirations

American Sociological Association, Annual Meeting *Aug 2023*

- Return Migrant Children in Contemporary China: Estimating Impacts of Return Migration on Children’s Educational Aspirations

Honors & Awards

Karen J. MacDonald and David A. MacDonald Scholarship	<i>2025 – 2026</i>
PSID Travel Stipend	<i>2025 Summer</i>
Foundry10 Bridge Fellowship	<i>2025 Summer</i>
Don Dillman Fellowship	<i>2025 Summer</i>
James F. Short Jr. Research Award	<i>2025 Summer</i>
Fred Yoder Graduate Fellowship Endowment in Sociology	<i>2024 – 2025</i>
Sociology Excellence Fund – Lois DeFleur	<i>2024 Summer</i>
Third Position at the GPSA Research Exposition	<i>2024</i>
Craft Family Scholarship	<i>2023 – 2024</i>
Fred Yoder Graduate Fellowship Endowment in Sociology	<i>2023 – 2024</i>
CAS Graduate Research Award	<i>2023 – 2024</i>
Half Exemptions of Tuition Fees of The University of Tokyo	<i>2020, 2021, 2022</i>
Scholarship for an Outstanding Student of Wuhan University	<i>2017, 2018, 2019</i>
Outstanding Youth Volunteer of Wuhan University	<i>2017, 2018</i>
Scholarship for Freshmen in Basic Subjects of Wuhan University	<i>2016</i>

Technical Skills

Programming & Tools: Stata, SQL, R (tidyverse, ggplot2, dplyr), Python (pandas, numpy), SAS, ArcGIS Pro

Statistics & Experiments: Causal inference (propensity score methods, difference-in-differences, regression discontinuity), experimental design (A/B testing), generalized linear models, survival analysis, Bayesian inference (MCMC), time series

Data & Visualization: Reproducible pipelines, data cleaning and validation, static visualization (e.g. map); web scraping (Selenium, BeautifulSoup)

Data Domains: Unstructured data, survey, panel, and administrative data (NHIS, PSID, CEPS, ACS, HMDA, CEX)