

# Siyu Wang

5002 Sheboygan Ave, Madison, WI, 53705 | (608) 381-5100 | [iamwangsiyu@gmail.com](mailto:iamwangsiyu@gmail.com) | [linkedin.com/in/siyu79/](https://www.linkedin.com/in/siyu79/)

## PROFESSIONAL SUMMARY

---

- Independently developed and implemented a decision tree package in both R and Python.
- Led the development and implementation of a new survival analysis framework during an internship at a medical device company, providing critical insights to guide future clinical trials.
- Experienced in end-to-end A/B testing, from planning and execution to analysis, with proficiency in tools like Optimizely, Tableau, and Snowflake.

## SKILLS

---

- **Programming & Tools:** R, Python, C++, SAS, SQL, Bash, GitHub, Jupyter Notebook, Optimizely, AWS
- **Statistical Modeling & Machine Learning:** Decision Trees, Survival Analysis, Monte Carlo Simulations, Linear Regression, Experimental Design, Causal Inference, Nonparametric Statistics, Multivariate Analysis
- **Data Visualization & Analysis:** Matplotlib, ggplot2, Excel, Tableau

## EDUCATION

---

### UNIVERSITY OF WISCONSIN-MADISON

*Ph.D. Candidate in Statistics*

*Master of Science in Data Science*

*Madison, WI*

*Expected Dec 2024*

*May 2019*

### UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

*Bachelor of Science in Statistics*

*Hefei, China*

*May 2018*

## PROFESSIONAL EXPERIENCE

---

### Inari Medical

#### Biostatistics Intern

*Irvine, CA*

*May - Aug 2022*

- Participated in quality control for a data preprocessing pipeline, transforming complex datasets into analysis-ready formats using R and SAS. Instrumental in detecting and rectifying data issues during the integration of health records in CRF from real-world registries.
- Collaborated with the Clinical Insights team to evaluate the influence of biomarkers on forecasting patient severity. Implemented innovative survival models for ad-hoc counterfactual analyses.
- Engaged in methodological research focusing on the nuances between SAS and R in handling repeated measurements. Conducted extensive literature review on marginal homogeneity. Presented these research findings at a monthly department meeting, contributing to knowledge sharing and departmental learning.

### Python and R Package for an Enhanced Decision Tree Method ([LDATree](#))

*Sep 2021 - Now*

- Developed and maintained the LDATree package in R (CRAN) and Python (PyPI), which handles missing values, outperforms traditional LDA, and matches Random Forest's accuracy with improved efficiency.
- Enhanced performance by integrating a novel stepwise LDA variant, utilizing C++ for optimization and advanced statistical methods. Developed visualization tools to improve clarity and accessibility for users from diverse backgrounds.

### Bootstrap-Based Confidence Interval Methods for Mediation Effects

*Sep - Dec 2019*

- Simulated data for time-varying causal models to evaluate the mediation effect between treatment and outcomes using advanced statistical techniques.
- Ran thousands of simulations on a high-performance computing (HPC) server with Bash and R scripts to accurately estimate type-I error rates and compute calibrated confidence intervals via bootstrap methods.