

(626) 993-0786  
Irvine, CA  
xiwenj@ics.uci.edu

# Xiwen (Josie) Jiang

Data Scientist

Portfolio: [xiwenjiang.github.io](https://xiwenjiang.github.io)  
[github.com/xiwenjiang](https://github.com/xiwenjiang)  
[linkedin.com/in/xiwen-jiang-115202224](https://linkedin.com/in/xiwen-jiang-115202224)

## EDUCATION

**PhD in Statistics**, *University of California, Irvine (UCI)* (Expected) Dec 2024  
**Master of Science in Statistics**, *University of California, Irvine (UCI)* Jul 2019 — Jun 2021  
**Bachelor of Science in Statistics**, *University of California, Los Angeles (UCLA)* Jul 2016 — Jun 2019  
*Dean's List, Department of Statistics at UCLA* 2016 — 2019  
*Outstanding Graduate Teaching Assistant, Department of Statistics at UCI* 2021

## WORK EXPERIENCE

**Data Scientist Intern/ Sequential AB Testing** Jun 2023 — Aug 2023  
*Nike Digital Commerce Analytics & Research (Nike.com and Nike app)* Beaverton, OR

- Optimized the existing A/B testing methodology for the Nike Direct organization, resulting in over 50% reduction in time and resource costs for A/B testing compared to traditional methods.
- Developed a comprehensive suite of automated scripts for efficient data computation and analysis within the newly sequential AB Testing method.
- Designed and implemented an innovative R package and interactive ShinyApp, revolutionizing daily tasks for the data science team and enhancing operational efficiency.
- Proactively projected a potential yearly revenue increase of \$6 million for Nike Direct organization through the implementation of the optimized sequential A/B testing method.

**Research Data Scientist Intern / COVID-19 and cardiovascular disease research** Jun 2022 — Jun 2023  
*Children's Health Orange County* Orange, CA

- Utilized data analysis tools, including Python and R, to efficiently clean, organize, and conduct in-depth statistical analyses on a complex and imbalanced longitudinal dataset.
- Explored the correlation between cardiovascular conditions and COVID-19 severity, contributing to a collaborative research paper set for submission to JAMA Network Open.
- Developed a robust model for estimating the impact of COVID-19 severity on cardiovascular conditions and drew meaningful statistical inferences from the data.

**College Instructor/ Summer Class of Introduction to Probability and Statistics** Jun 2022 — Sep 2022  
*University of California, Irvine* Irvine, CA

- Instructed college-level statistics courses to undergraduate students, providing comprehensive coverage of statistical principles, methods, and applications.
- Developed and delivered engaging and interactive lesson plans, ensuring a clear understanding of complex statistical concepts.
- Offered guidance and support to students through one-on-one consultations, aiding in their comprehension of statistical topics and improving their performance in assessments.
- Demonstrated strong communication and presentation skills to foster a positive and productive learning environment for students.

## RESEARCH EXPERIENCE

**Innovative Approach for Modeling Time Series Network Data** 2021 — 2023  
*Dynamic Network Data Estimation and Recovery Structure* Irvine, CA

- Developed and implemented an advanced approach for modeling time series network data, addressing key limitations in existing models.
- Proposed a groundbreaking solution by introducing a new family of models called "block-structured dynamic exponential-family random graph models" (ERGMs), which divided time into blocks, allowing smoother parameter transitions, resulting in more accurate trend estimation.
- Utilized ERGMs to estimate dynamic network trends, and developed a semi-parametric model to extract trends from hidden block structures, enhancing trend accuracy.
- Validated the approach through extensive simulation studies and applied the method to analyze real-world data, including interbank transaction networks and political blog citations networks.
- Published the scientific paper in the journal of *Stat* as the first author, demonstrating the significance and recognition of the work within the scientific community.

(626) 993-0786  
Irvine, CA  
xiwenj@ics.uci.edu

# Xiwen (Josie) Jiang

Data Scientist

Portfolio: [xiwenjiang.github.io](https://xiwenjiang.github.io)  
[github.com/xiwenjiang](https://github.com/xiwenjiang)  
[linkedin.com/in/xiwen-jiang-115202224](https://linkedin.com/in/xiwen-jiang-115202224)

## Time Series Clustering for Biomedical Data Analysis

2022 — 2023

*Simultaneous denoising and heterogeneity learning for time series data*

Irvine, CA

- Developed a novel time series clustering method that emphasizes smoothing over time and capturing subject-level variations, improving data heterogeneity understanding.
- Extended the method to integrate with various distance-based clustering techniques, enhancing its versatility and applicability.
- Demonstrated the method's efficacy through numerical experiments, showcasing its ability to recover cluster structures and enhance time series data quality.
- Applied the method to real-world biomedical data, including an U.S. lung cancer mortality study and a suicide rate study, yielding valuable insights.
- Published the scientific paper in the journal of *Statistics in Biosciences* as the first author, highlighting the recognition of the work in the field.

## Amazon Customer Sentiment Analysis and Sales Prediction

2021 — 2022

*Analyze Amazon customer sentiment and sales data to make data-driven predictions*

Irvine, CA

- Collaborated seamlessly with a diverse team of economists and data analysts to manage and evaluate a substantial textual dataset comprising over 2 million questions and answers on the Amazon platform.
- Organized and meticulously evaluated the large-scale textual dataset, ensuring data integrity and accessibility.
- Utilized advanced machine learning techniques in Natural Language Processing (NLP) for sentiment analysis on the textual data. This involved extracting nuanced sentiments and insights from customer interactions, leveraging collective expertise.
- Employed Vector Autoregressive (VAR) modeling as a part of the team effort to assess the correlation between product sales ratings and sentiment scores. Utilized these models to make data-driven predictions for future sales ratings, contributing to informed decision-making.

## SKILLS

Tools and Languages	Python (Pandas, NumPy, ScikitLearn, Matplotlib), R, SQL, Spark, SAS, STATA, SPSS, $\text{\LaTeX}$ , CSS, Bash, Markdown
Quantitative Research	AB Testing, Time Series Analysis, Spatial Analysis, Social Network/Graph Analysis, Mathematical optimization, Statistical Modeling, Machine Learning, Natural Language Processing, Survival Analysis
Communication	English, Mandarin, Cantonese

## PUBLICATIONS

- Jiang, X., Margarita Livas, S., Yin, F., Banerjee, S., Butts, C. T., & Shen, W. (2023). Structure recovery and trend estimation for dynamic network analysis. *Stat*, 12(1), e593. ([link](#))
- Jiang, X., & Shen, W. (n.d.). Simultaneous denoising and heterogeneity learning for time series data. *Statistics in Biosciences*. ([link](#))
- Garó, H., Jiang, X., Grant, C., & et al. (2023). Survival impact of post-operative immunotherapy in resected stage iii cutaneous melanomas in the checkpoint era. *ESMO Open*. (Under Review)
- Lewis, S. A., Sureshchandra, S., Zulu, M. Z., Doratt, B., Jankeel, A., Ibraim, I. C., & Messaoudi, I. (2021). Differential dynamics of peripheral immune responses to acute SARS-CoV-2 infection in older adults. *Nature Aging*, 1(11), 1038-1052. ([link](#))
- Sureshchandra, S., Chan, C., Robino, J., & et al., L. (2022). Maternal western-style diet remodels hematopoietic stem and progenitor cells in the bone marrow of fetal rhesus macaques. *Stem Cell Reports*. ([link](#))