Ziyue (Patrick) Zheng

ziyuezheng@ucsc.edu • ziyuezheng.github.io • github.com/ZiyueZHENG

Summary

PhD student in Statistics at UC Santa Cruz working on **semi-supervised learning**, **uncertainty quantification**, and **interpretable statistical/ML models** with applications in oceanography and biological sciences. Build end-to-end ML pipelines in **Python/R**, including modeling, evaluation, and deployment. Prior industry experience as an **ML Engineer** and **Data Analyst**.

Education

University of California, Santa Cruz — *Aug 2023 – Present* Ph.D. in Statistics
University of Wisconsin–Madison — *2021 – 2022* M.S. in Data Science
Zhongnan University of Economics and Law — *Sep 2016 – Jun 2020* B.S. in Statistics

Technical Skills

Languages	Python, R, SQL, JavaScript, C++, Shell/Linux
ML / Stats	PyTorch, TensorFlow; scikit-learn, tindyverse, ggplot2
Data	pandas, NumPy; ETL; MySQL; MongoDB
Big Data / Dev	Spark, Hadoop, Hive; Git; REST APIs; Docker
Web / Serving	ShinyR, Django, Flask; basic API/service deployment on AWS/GCP

Research Projects

Subcellular proteome niche discovery using semi-supervised functional clustering *2025*

Ziyue Zheng, Loay J. Jabre, Matthew McIlvin, Mak A. Saito, and Sangwon Hyun.
arXiv preprint arXiv:2512.08087, 2025. PDF

- Implemented a semi-supervised functional clustering workflow designed to leverage a small labeled subset alongside abundant unlabeled data.
- Developed novel model assessment and evaluation criteria for choosing hyper parameters. Compared model result with different methods on different datasets.
- Built a reusable package FSPmix pipeline (data preparation → model fitting → interpretation) and documented usage with vignettes/tutorial material.
- Supported analysis for a spatial proteomics application motivating the method; linked paper experiments and code organization for reproducibility.

Debiased inference for semi-supervised MoE model for flow cytometry data *2026 (in progress)*

- Developed methodology for semi-supervised mixture-of-experts models tailored to high-dimensional flow cytometry data.
- Built an evaluation pipeline (simulation + real data) to assess estimation quality and uncertainty; packaging code for repeatable experiments.
- Contributed to an R package flowmix implementing sparse mixture of multivariate regressions for continuous-time flow cytometry data collected over space and time.
- Reproduced and extended analysis workflows using provided paper code and data structure (model fitting, diagnostics, and reporting).

GRUMP project: relative abundance clustering for plankton metabarcoding data *2026 (in progress)*

- Working on statistical analysis and methodology development for Global rRNA Universal Metabarcoding of Plankton (GRUMP) data.
- Prototyping clustering/abundance modeling workflows and organizing code for collaborative research.

Industry Experience

Machine Learning Engineer, Tencent Inc. *May – Sep 2021*

- Built and trained a machine learning model for face recognition in Python.
- Developed a Django-based web application to serve model predictions end-to-end; exposed REST APIs for online inference.
- Deployed the service on a cloud platform; tuned inference pipeline and improved code efficiency.

- Evaluated model performance using statistical methods and iterative experiments.

Data Analyst, Skyworth Inc. *Apr – Jul 2020*

- Maintained and managed financial databases using MySQL and MongoDB; executed ETL to integrate data across sources.
- Performed analysis using pandas/NumPy; produced reports with tables and visualizations to support business decisions.
- Conducted risk identification analyses for financial products using statistical models.

Teaching

- **STAT 131: Intro to Probability** (UCSC) Winter / Spring / Fall 2025
- **STAT 5: Statistics** (UCSC) Spring / Fall 2024

Last updated: January 30, 2026