

Yuxuan(Susan) Wang

MASTERS STUDENT · DATA SCIENCE

✉ susanwyx@umich.edu | 🏠 wang264y.github.io | 📧 wang264y

Education

University of Michigan

MASTERS OF DATA SCIENCE

Ann Arbor, MI

Aug. 2023 - Present

- GPA: 3.94/4.00
- Research Advisor: Dr. Ivo Dinov and Dr. Simeone Marino
- Coursework: Databases Management Systems, Data Mining, Multivariate Analysis, Statistics Theory, Regression Analysis(IP), Bayesian Inference(IP)

Mount Holyoke College

BACHELARS OF ART IN MATHEMATICS AND DATA SCIENCE

South Hadley, MA

Aug. 2019 - May. 2023

- GPA: 3.98/4.00
- Advisor: Dr. Timothy Chumley
- Coursework: Mathematical Statistics, Probability Theory, Complex Analysis, Abstract Algebra, Applied Regression Methods, Stochastic Processes, Real Analysis

Publications

PUBLISHED

Bradt, S. A., Elder, J., Harris, P. E., Kirby, G. R., Reuterer, E., **Wang, Y.**, Whidden, J. (2024). Unit Interval Parking Functions and the r -Fubini Numbers. *La Matematica*, 1-15. (The authors equally contributed to this paper and are listed alphabetically.)

Xu, M., **Wang, Y.**, Huang, Y., Xu, M. (2021). Leveraging Financial News Analysis to Predict Stock Price Movement. *Frontiers in Economics and Management*, 2(7), 265-276.

IN PREPARATION

Marino, S., **Wang, Y.**, Liu, Y., Tang, M., Dinov, I., Burns, M. et al. (2024) Rationale and Strategies for Medical Data Sharing and Synthetic Clinical Data Generation – Maximizing Data Value Utilization and Minimizing Risks. *In preparation*.

Research Experience

Privacy and Utility Metrics Generation and Data Analytics

Ann Arbor, MI

UNIVERSITY OF MICHIGAN, ANN ARBOR Advisors: Dr. Ivo Dinov & Dr. Simeone Marino

Jan. 2024 - Present

- Utilizing Python SDV and SDMetrics packages to evaluate the privacy and utility aspects of the DataSifter and compare the DataSifter obfuscation methods with existing data synthesizers.
- Assessing and developing the alternative version of the DataSifter that addresses longitudinal data.
- Preparing for an academic paper titled "Rationale and Strategies for Medical Data Sharing and Synthetic Clinical Data Generation – Maximizing Data Value Utilization and Minimizing Risks".

News Representations of Chinese International Students before and after COVID-19

South Hadley, MA

MOUNT HOLYOKE COLLEGE Advisor: Dr. Kenneth Mulder

Jan. 2023 - May. 2023

- Built the structural topic model, conducted LIWC sentiment analysis and 2 sample t-tests to compare public portrayal before and after the COVID-19 outbreak.
- Interpreted models and results to investigate sentiment differences in news media about Chinese international students before and after the COVID-19 outbreak.
- Delivered presentation and report as the Undergraduate Data Science Capstone Project.

Data Science Infrastructure Construction using AWS and Docker Container

South Hadley, MA

MOUNT HOLYOKE COLLEGE Advisors: Dr. Ben Gebre-Medhin

Mar. 2022 - May. 2023

- Developed data science docker container with R working environment based on existing images.
- Assisted in setting up the Littlest JupyterHub on AWS for MHC campus to serve 5+ courses 3+ semesters.
- Tested out the JupyterHub functions of user login, folder permissions, backup processes, SSN authorizations.

Parking Functions with Fixed Ascents, Descents, and Ties Sets

Providence, RI

ICERM *Advisors: Dr. Gordon Rojas Kirby, Dr. Pamela E. Harris & Dr. Jennifer Elder*

Jun. 2022 - Aug. 2022

- Conducted literature reviews on various combinatoric patterns of parking functions with different statistics subsets.
- Enumerated parking functions with fixed statistics sets by coding in Python and Sagemath.
- Extended the studies in the field of combinatorics by publishing a journal paper and delivering poster and conference presentations discussing the association between Fubini, R-Fubini numbers and unit interval parking functions with certain tie subsets.

Using Data to Inform ECEP's CS Education Policy & Advocacy Efforts

South Hadley, MA

DSC-WAV *Advisor: Dr. Valerie Barr; Client: ECEP Organization*

Jun. 2021 - Dec. 2021

- Utilized data mining, data cleaning, data visualization skills through R studio to help a national organization called Expanding Computing Education Pathways (ECEP) to analyze the trends of CS education in the past 5 years.
- Built relevant Github repository, and communicate with fellow data scientists through git.
- Provided strategies on how to facilitate broad participation in computing across states, especially focusing on minority populations and non-ECEP states.

Billiard, Probability, and their Interplay: Determination of Average Time that A

South Hadley, MA

Particle Escapes from a Channel with Diverse Parameters

MOUNT HOLYOKE COLLEGE *Advisor: Dr. Timothy Chumley*

Jun. 2021 - Dec. 2021

- Conducted literature reviews on probability distribution laws and random walks derived from billiards escaping a channel.
- Performed simulations in R and proposed conjectures about the influence mechanism of geometric microstructures in a channel on the mean escape time of a billiard following the specular-diffuse random collision law with various parameters.
- Designed an R-shiny app, presentation, and report to present the findings, allowing users to observe the changes on the mean escape time by altering the parameters, such as the portions of specular or diffuse collisions and the incoming angles of the billiards.

Leveraging Financial News Analysis to Predict Stock Price Movement

Online

CATHAYPATH INTERNATIONAL SUMMER (CIS) PROGRAM

May. 2020 - Jul. 2020

Advisor: Dr. Patrick Houlihan

- Collected and cleaned stock market data for 3538 stock tickers over the period of 2017-2020 using NLTK tools.
- Performed sentiment analysis using VADER and fitted the models to predict stock market price data involving Logistic Regression, Naive Bayes model and Random Forest.

Presentations

CONTRIBUTED PRESENTATIONS

Reuteracrona, E., **Wang, Y.**, Whidden, J., 2023. Parking Functions with Fixed Ascent and Descent Sets.

Oral Presentation: Joint Mathematics Meetings 2023 - Pi Mu Epsilon Contributed Session on Research by Undergraduates, Boston, MA

Reuteracrona, E., **Wang, Y.**, Whidden, J., 2023. Subsets of Unit Interval Parking Functions Enumerated by Fubini Numbers.

Poster Presentation: Joint Mathematics Meetings 2023 - AMS - PME Undergraduate Student Poster Session, Boston, MA

Teaching/Tutoring Experience

Spring 2023 **Real Analysis**, Teaching Assistant

Fall 2022 **Real Analysis**, Teaching Assistant

Fall 2021 **Introduction to Proof Through Analysis**, Teaching Assistant

Spring 2021 **AP Calculus**, Undergraduate Student Instructor