Wangqian Ju

LinkedIn: linkedin.com/in/wangqian-will-ju/
Github: github.com/willju-wangqian

Email: wju@iastate.edu
Mobile: +1(515)708-2829

Education

Iowa State University — Ames, IA
- *Ph.D. Candidate in Statistics; Advised by Dr. Heike Hofmann* — *Aug. 2019 - Expected May. 2024*
  3.94 GPA; Kempthorne Award in Statistics;
- *Master of Science in Statistics* — *Aug. 2019 - June. 2021*

University of California - Davis — Davis, CA
- *Bachelor of Science in Statistics; Minor in Computer Science;* — *Aug. 2015 - June. 2019*
  3.92 GPA; UC Davis Statistics Outstanding Performance Citation; Graduation with Honors

Skills

- **Statistical Analysis:** Generalized Linear Model, ANOVA, Bayesian, Time Series, Statistical Inference
- **Coding:** R, Python, Shiny App, C, Matlab, SQL, Bash, Git, JavaScript
- **Frameworks:** Scikit, TensorFlow, Keras

Research Experience

Open-source Implementation of the CMPS Algorithm - R and C — May. 2020 - Present
Center for Statistics and Applications in Forensic Evidence - Research Assistant

- Implemented the Congruent Matching Profile Segment (CMPS) algorithm in **R** and built an **R** package called `cmpsR`. The CMPS algorithm is conceptually described by the National Institute of Standards and Technology for objective comparison of striated tool marks. As an open-source implementation, the `cmpsR` package facilitates field experts to cross-validate the parameters, estimate its accuracy, compare with other algorithms, and move further towards the framework of open science
- Optimized running speed with **C** code, applied the `cmpsR` implementation to the processed bullet signature data, achieved a 100% classification rate, and obtained comparable and reproducible results. These results validate the `cmpsR` implementation and suggest further research on algorithm improvements and case studies
- Developed a principled evaluation framework of algorithmic results. The framework can be used for comparing algorithmic results from different parameter settings, different models, or different datasets. The results of the framework generated by different researchers can be incorporated together and contribute to the construction of a comprehensive database of algorithms for algorithmic assessment and comparison
- **Highlights:** R package development, algorithm implementation and assessment, pattern matching

Data Processing and Visualization tool - R and Shinyapp — Jan. 2021 - Present
Center for Statistics and Applications in Forensic Evidence - Research Assistant

- Built a visualization tool called `bulletInspectR` using **R** shiny. It is designed for forensic field experts who are not familiar with R. Without digging into the code, users can use the interactive `bulletInspectR` to perform data processing, identify and mark damaged data, visualize possible results, and download results in their desired format
- Applied the idea of modularization to the design of `bulletInspectR`. Users can implement, insert, and remove modules freely to have their desired functionality
- **Highlights:** design and development of interactive tools, data visualization and production

Data Analysis Projects

Green Beans Daily Sales Prediction - R — Jan. 2021 - Apr. 2021

- Conducted exploratory data analysis on data of individual stores and fitted primary models for further analysis
- Fitted a statistical model with a two-stage prior that allows individual stores to investigate the relationship between sales and prices through the posterior estimations of the model parameters
- Used **JAGS** (Just Another Gibbs Sampler) to obtain credible intervals, posterior distributions, posterior predictive distributions, and other estimations of the model parameters

New York City Real Estate Price Prediction - R and Python — Jan. 2020 - Apr. 2020

- Investigated and processed 50,000+ real estate data in New York City and performed exploratory data analysis
- Built a random forest model to predict the real estate price using model prediction results (Ordinary Least Squares, K-Nearest Neighbors, Principal Component Regression, Elastic Net, etc) as features and achieved a loss score of 0.35
- Built an artificial neural network model in **Python** with Keras and compared its result with that of the random forest model

Publications

- **"An Open-Source Implementation of the CMPS Algorithm for Assessing Similarity of Bullets,"** The R Journal, In Preparation