Zhijian Liu

subaiwen.github.io
zl1409a@american.edu

SKILLS

Statistical techniques:

- **Data science**: querying, web-scraping, cleaning, manipulating, modeling, visualizing, reporting
- **Regression**: regression, GLM, multilevel models, multivariate analysis (PCA/LDA)
- **Machine learning**: KNN, spline, cluster, tree, kernel machine, model selection and regularization
- **Visualization**: R-Shiny (shiny app), Tableau

Tools:

- **Statistical software:** SAS, SPSS, STATA, StatCrunch, EXCEL, Tableau
- **Programming Language:** SQL, R, Python, Java
- **Linguistic Language:** English, Cantonese, Madarin

EDUCATION

Master of Statistics | American University – DC 05/2019

- ML, GLM, Data science, Experimental design, Bayesian, Time series, Stochastic process, Survey design
- GPA: 3.87

Bachelor of Business Administration | South China Normal University – China 07/2017

EXPERIENCES

Statistical Software Support Staff | Part-time – American University 09/2018 – 05/2019

- Provided software support for students with *R, SAS, STATA, SPSS* and *StatCrunch* daily in the library
- Guided students to learn and apply statistical software to their research, projects and dissertations

Teaching Assitant | Part-time – American University 09/2018 – 05/2019

- Graded 50 copies of assignments weekly and summarized students' common mistakes for professors
- Recorded class notes with ***LaTeX*** and ***Rsweave***, and output them as handouts for students
- Introduced tools, such as *ggplot*, to students to optimize codes and graphical output of assignments

Data Analyst | Internship – Igola, China 05/2018 – 08/2018

- Web-scraped the information of hotels in 20 major Asian cities in apps, such as *Expedia* and *Booking*, in order to locate tags, distance, price and comments to provide references for UI optimization and algorithm optimization of a hotel ranking system
- Tested the quality of geographic data provided by different web-service regarding hotels and city centers in over 800 cities in China to rank and filter the content suppliers
- Cleaned and wrangled the data, making it more user-friendly for a more diverse audience
- Analyzed, visualized, and reported the data to support marketing and operating decisions

H1B Data Analysis | Course project – American University 03/2018 – 05/2018

- Used *R* to analyze over 200,000 H1B Visa Petitions from 2011 to 2016 all over the US
- Detected, corrected and rematched miscoded data using general expression and *stringr* package
- Queried geographical data from *Google API* and visualized data with *ggplot* package
- Built shiny app for users to look up geographical information regarding the number of H1B applications and salaries on an interactive map

Energy Competition Exit Survey | Course project – American University 10/2017 – 12/2017

- Designed and implemented an exit poll survey of a 159 sample with questionnaires for the purpose of detecting the effect of a Campus-wide Energy Competition
- Imputed the missing data using KNN and committed preliminary analysis with descriptive plots in *R*
- Utilized *STATA* to build multiple regression to determine both the efficacy of outreach efforts, as well as the impact of the competition on students' actions