

## 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
- **Linguistic Language:** English, Cantonese, Madarin

## EDUCATION

**Master of Statistics** | American University – DC 05/2019

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

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

## EXPERIENCES

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

- Established procedures for students to transform raw data into consistent data that can be analyzed for GLMs, such as binomial/multinomial logistic model, survival model, log-linear model and so on
- Designed more than 40 visualization templates, including graphical correlation matrixes, distribution plots and diagnostic plots by using visualization packages, such as *ggplot2* and *GGally*, in R
- Utilized *LaTeX* and *Rsweave* to generate handouts and exemplary homework solutions weekly for students

**Data Analyst** | Internship – [lgola](#), 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 *ggplot2* 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