

## EDUCATION

---

- **Georgetown University** Washington DC, United States  
Master of Science in Data Science Science  
Teaching Assistant for ANLY-503 Advanced Data Visualization  
Aug. 2021 – May. 2023 (Est.)
- **The Australian National University** Canberra, Australia  
Master of Science in Statistics  
Feb. 2017 – Dec. 2018
- **University of Toronto** Toronto, Canada  
Honors Bachelor of Science with Distinction (Maths, Statistics and Computer Science)  
Sept. 2011 – Jun. 2016

## SKILLS

---

- **Programming:** SQL, Python, R, JavaScript, HTML, CSS, Java
- **Machine Learning:** scikit-learn, numpy, pandas, caret, tidymodels, tidyverse
- **Visualization:** ggplot2, R Shiny, leaflet, d3.js, matplotlib, seaborn, bokeh, plotly
- **Data Science:** A/B testing, ETL, pipeline (cleansing, wrangling, visualization, modeling, interpretation), scraping, feature engineering, NLP, cluster analysis, time series analysis, PCA, stochastic process, experimental design, hypothesis testing
- **Big Data & Miscellaneous Technologies:** AWS, Spark, MongoDB, Docker, Linux, Shell, Hugo, Figma

## EXPERIENCE

---

- **K2L Canberra** Canberra, Australia  
Product Data Analyst  
Jul. 2019 – Jun. 2021
  - Utilized Python and SQL to track and analyze product data, resulting in a 15% increase in recommendation accuracy
  - Built data visualization with MySQL and R for product KPIs that reduced manual reporting time by 3 hours per week
  - Improved UX measures by 20% through increased user feedback collection and frequent iteration to customer needs
  - Collaborated with the engineering team to launch an internal kanban tool for a better work-from-home experience
  - Launched 6 products (websites, mobile apps, and web apps) for clients in 2 years for a total 65% revenue gain
- **K2L** Canberra, Australia  
Co-founder/Data Analyst  
Feb. 2019 – Jun. 2019
  - Led a team of 12 to develop an online Q&A community combining the idea of StackOverflow and TikTok which grew to 1,000 users in 3 months
  - Performed market competition analysis based on scraped data to direct the initial product planning
  - Analyzed and strategized regarding business goals, schedules, and budgeting in the early stage of the development
  - Created functional prototypes that increased communication efficiency by 30% for clients
- **Zmate Network Technology** Shenzhen, China  
Data Analyst Intern  
Aug. 2016 – Dec. 2016
  - Cleaned 10TB of unstructured data of company registration information with SQL and Python, which reduced the storage space by 50% and increased the data accessibility
  - Utilized Python to implement supervised machine learning techniques to generate an evaluation metric of companies' intellectual property awareness, which highlighted 3% of potential customers for the marketing team
  - Implemented an automatic domain name generating system based on Chinese naming conventions, which also cross-checked with the existing data, leading to full automation in the routine of the marketing team

## PROJECTS

---

- **No Pie Newsletter:** A newsletter curating latest articles, packages, tools and podcasts about data visualization; [Website](#), [Telegram](#)
- **xG Visualization Pipeline (R, sports analytics, visualization):** An automation script to parse and clean raw data from soccer games and visualize it in 3 different aspects; [Blog](#), [repository](#)
- **tarantino (R package development):** An R package that generates color palettes inspired by Quentin Tarantino's movies for visualization; Featured in R-Weekly newsletter; [Repository](#)
- **@canberramapbot (R, API):** A Twitter bot posts bird's-eye view over Canberra periodically; [Repository](#), [Twitter](#)
- **sentRy (R Server, AWS, API):** A DevOps monitoring tool that automatically sends Django error logs to subscribers' Telegram account; Reduced the total amount of notifications by 40% via aggregation and trimming redundant information; [Blog](#), [repository](#)
- **aussie (R, Selenium, Python, crawler, data cleaning):** A collection of Python and R web crawlers that scraped over 76,000 courses offered by 25 Australian universities in 2020; Cleansed and wrangled the data into a tidied structure for further analysis; [Repository](#)