

# RUI QIU

RQ47@GEORGETOWN.EDU | (571)363-5917 | [REXARSKI.COM](https://rexarski.com) | [GITHUB/REXARSKI](https://github.com/rexarski) | [LINKEDIN/RQIU](https://www.linkedin.com/in/rqiu)

## EDUCATION

- **Georgetown University** Washington D.C., United States  
*Master of Science in Data Science and Analytics. cGPA: 4.0* 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 – June 2016

## SKILLS

- **Programming:** SQL, Python, R, JavaScript, HTML, CSS, Java
- **Machine Learning:** tidyverse, tidymodels, caret, jupyter-notebook, scikit-learn, numpy, pandas
- **Visualization:** ggplot2, Shiny, d3.js, matplotlib, seaborn, plotly, leaflet, bokeh
- **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 (EC2, EMR, S3, SageMaker), Azure, Databricks, Spark, Hadoop, MongoDB, Docker, Linux, Shell, Hugo, Figma

## EXPERIENCE

- **International Finance Corporation (World Bank Group)** Washington, D.C.  
*Data Scientist Intern* May 2022 – Present
  - Develop metrics (such as fairness, explainability, and adversarial robustness score) to evaluate the performance of MALENA (Machine Learning ESG Analyst)
  - Train and fine-tune transformer-based NLP and NLU models (including RoBERTa and DistilBERT) to enhance the robustness and stability of MALENA
  - Collaborate with CEG and IFC stakeholders to understand their requirements and designed innovative dashboards for machine learning inference data visualization
  - Improve efficiency by cleaning up and scheduling batch jobs in the Databricks cluster, which provided periodic summary statistics for the dashboard
- **Georgetown University** Washington, D.C.  
*Teaching Assistant / Bootcamp Manager* Aug. 2021 – Aug. 2022
  - Served as a graduate teaching assistant for ANLY-503 (Advanced Data Visualization) and ANLY-560 (Time Series), providing weekly office hours and grading assignments for a group of over 30 students
  - Facilitated course materials, managed the course Slack channel, and organized the course GitHub organization
- **K2L Canberra** Canberra, Australia  
*Product Data Analyst* July 2019 – June 2021
  - Utilized Python and SQL to track and analyze product data, resulting in a 15% increase in recommendation accuracy
  - Created data visualizations with MySQL and R for product key performance indicators (KPIs), reducing manual reporting time by 3 hours per week
  - Improved user experience (UX) measures by 20% through increased user feedback collection and frequent iteration based on customer needs
  - Worked with the engineering team to launch an internal kanban tool to improve the work-from-home experience
  - Successfully launched 6 products (including websites, mobile apps, and web apps) for clients in 2 years, resulting in a 65% revenue increase
- **K2L** Canberra, Australia  
*Co-founder/Data Analyst* Feb. 2019 – June 2019
  - Successfully led a team of 12 in developing an online Q&A community, which merged the concepts of StackOverflow and TikTok. The community grew to 1,000 users in just 3 months
  - Conducted market competition analysis using scraped data to inform the initial product planning process
  - Analyzed business goals, schedules, and budgeting during the early stages of development and created strategies to align with these objectives
  - Developed functional prototypes that improved communication efficiency by 30% for clients

## PROJECTS

- **What Ingredients Are You Tasting In Authentic Japanese Flavors?** A comprehensive interactive visualization project on Japanese cuisine with D3, plotly and leaflet. [Website](#)
- **Basketball Analytics and Beyond** A project on various basketball related topics, both on-court and off-court. [Website](#)
- **No Pie Newsletter** A newsletter curating the latest articles, packages, tools and podcasts about data visualization. [Archive](#)
- **xG Visualization Pipeline** An automation R script to parse and clean raw data from soccer games. [Tweet](#), [archive](#), [repository](#)
- **tarantino** An R package that generates color palettes inspired by Quentin Tarantino's movies for visualization. [R-Weekly](#), [repository](#)
- **@canberramapbot** A Twitter bot utilizes Mapbox and Twitter APIs to post bird's-eye view over Canberra. [Repository](#), [Twitter](#)
- **sentRy** A DevOps monitoring tool that sends Django error logs to subscribers' Telegram account. Reduced the total amount of notifications by 40% via aggregation and trimming redundant information. [Blog](#), [repository](#)