YUTANG XIONG

Parker, CO 80134 ◇ 312-874-0451 ◇ yutang.xiong@du.edu ◇ LinkedIn ◇ GitHub

EDUCATION

University of Denver — Denver, CO
Ph.D. candidate, Research Methods & Statistics, GPA: 4.0/4.0 — Expected Spring 2025

University of Missouri-Columbia — Columbia, MO
M.A., Statistics, GPA: 3.7/4.0 — May 2017

Hunan University — Changsha, China
B.S., Mathematics, GPA: 80.9/100 — Jul 2014

SKILLS

Relevant Coursework: Statistical Inference, Advanced Probability, Statistical Learning, Bayesian Statistics, Algorithms & Data Structures, Database Organization & Management, Complex Data Analytics, Text Mining, Data Mining

Programming & Software: Python, R, SQL, SAS (basic), Stata (basic), Bash (basic)

PROFESSIONAL EXPERIENCE

Pardee Institute for International Futures — Denver, CO
Lead Data Operations Manager — Oct 2019 - Present

- Managed and updated relational databases (SQLite, Microsoft Access) for an integrated assessment model. Automated data processing and merging using Python libraries (Pandas, NumPy, pyodbc, sqlite3), ensuring data integrity and reducing manual effort by 50%
- Developed Python web scrapers using Selenium, Requests, BeautifulSoup to collect structured and unstructured data from APIs and web sources. Extracted and processed large datasets, such as commodity-level trade data, enhancing data availability for analysis
- Conducted ad-hoc statistical analyses to support research projects. Employed and optimized predictive models ranging from traditional regression to machine learning techniques like PCA, clustering, and network analysis. Utilized Scikit-Learn for model development and evaluation, enhancing model performance through feature selection and tuning
- Performed text analysis, including frequency and sentiment analysis, on speech data extracted from PDF files. Utilized regular expressions, NLTK, wordcloud, and SpaCy in Python to process and analyze unstructured data
- Created a streamlined process to enable interaction between custom-built Python modules and two .NET applications. Automated database modifications and model output analysis to establish a feedback loop for an integrated assessment model
- Maintained and updated technical documentation in MediaWiki to streamline data update and vetting processes.

National Bureau of Statistics, Jiangxi Province — Nanchang, China
Intern, Division of Comprehensive Statistics — Oct 2013 – Dec 2013

- Assisted in workflow of the Survey Office, including raw data collection, data processing and data analysis
- Reviewed research papers and internal work reports to ensure publication quality

Chinese Academy of Sciences — Beijing, China
Remote Research Intern — Jul 2013 - Sep 2013

- Reviewed literature and provided some numerical examples for National Survey on Enterprise Salary in China
- Derived the estimator of the finite population quantile under the stratified cluster sampling with probability proportional to size under a professor's supervision

RESEARCH EXPERIENCE

Identifying Patterns in the Structural Drivers of Intrastate Conflict
University of Denver, research funded by the Minerva Initiative — Nov 2021 - Jun 2022

- Automated the process of compiling and selecting intrastate conflict observations and drivers using Python
- Conducted hierarchical clustering analysis, compared model behaviors, and tuned model parameters to identify 3 major patterns of intrastate conflict

- Visualized research findings and outputs of sensitivity analysis through R

From Insights to Action: Gender Equality in the Wake of COVID-19
Pardee Center, work commissioned by the UN Women-UNDP collaboration May 2020 - Sep 2020

- Compiled poverty rate data at different poverty lines from multiple groups
- Verified the structure of 5 targeted issues areas driving poverty and gender-poverty gap by using principal component analysis with over 90 variables
- Produced a final predictive model based on iterative regression analyses and diagnosis of model outputs

Conditions for Success in the Implementation of the African Continental Free Trade Agreement
Pardee Center, work commissioned by the African Union Development Agency Aug 2019 - Feb 2020

- Compiled large scale commodity-level trade and tariff data (over 40GB) from multiple resources
- Calculated time-series trade complementarity indices and weighted tariff values between countries
- Built a SQL database in Microsoft Access for dyadic trade and tariff information

Multiple Imputation with Simulated Non-Normal Missing Variable at Level 2
University of Denver, paper presented at the 2019 AERA Annual Meeting conference Oct 2017 - Apr 2019

- Simulated nested structure data sets with a non-normal missing variable at the 2nd level
- Compared the performance between three multi-level imputation methods and the regular single-level imputation
- Identified the negative effects brought by non-normality to the imputed data, and discussed each method's change of performance from data size shift

Leaf Classification, Competitions on Kaggle.com
University of Missouri-Columbia, ranked top 10% over 1000 teams Nov 2016 - Dec 2016

- Implemented statistical learning algorithms through h2o package in R, including Support Vector Machine, weighted Nearest Neighbor method, Decision Tree Based methods, and Deep Learning Methods
- Improved each model's performance by tuning optimal parameters using boosting and 5-fold cross validation
- Compared models and achieved miss classification error less than 1% with log-loss less than 0.05

ADDITIONAL EXPERIENCE

Python Workshop for Graduate Students at Josef Korbel School of International Studies
University of Denver Feb 2020

Stats & R Tutor for Undergraduate Students in the 2017 ASA DataFest Competition
University of Missouri-Columbia Jan 2017 - Apr 2017

Vice President of Student Union, School of Mathematics
Hunan University Apr 2011 - Jun 2012