# Yuxuan Chen

Email: cyxnbnb@bu.edu · cyxnbnbnb@gmail.com
Phone: +1 8572564713 · +86 13957992893
[Github](Github)

## EDUCATION

**University of Virginia**                                                                               Charlottesville, VA
Master of Science: Data Science

**Brigham Young University**                                                                                      Provo, UT
Bachelor of Arts: Political Science

## SKILLS

**Languages and Tools** - *proficient:* Python, R, Bash *intermediate:* Spark, SQL, Tensorflow/Keras, Git, Docker
**Selected Coursework** - Deep Learning, Bayesian Machine Learning, Big Data Systems with Spark, Linear Models for Data Science, Statistical Learning, Data Engineering, Linear Algebra, Econometrics I & II, Data Visualization, Programming in Python

## EXPERIENCE

**Dynata**                                                                                              Aug 2022-Present
*Data Scientist, Marketing*                                                                                   Herndon, VA

- Delivered 30+ advanced analytical studies (MTA, drivers analysis, diff-in-diff) to high profile clients including 15+ to marketing teams at **Google and Meta**
- Built diff-in-diff and multi-touch attribution model codebase in Python(OOP) for Ad Data Science Team. This codebase has **saved 100+ hours and brought over in $2 million in revenue to date**

**Internet Archive**                                                                                   Nov 2022-May 2023
*Intern, Data Science*                                                                                   Washington, D.C.

- **Solved entity resolution problem with Gradient Boosted Trees** [here] linking sparse wikipedia references to Internet Archive Database References
- Implemented model in Python [here] which, given a wikipedia reference, will return a match in the Internet Archive Database along with a probability of matching.

**Morning Consult**                                                                                    Nov 2021-Aug 2022
*Senior Data Analyst*                                                                                    Washington, D.C.

- **Developed and automated wave-over-wave statistical tests on time series data** in 17 tracker surveys of 5 countries
- Led project to build a Python Web Bot (Selenium) to automate generation of test cases in surveys, contributed this to data science code base (used by 60+ data scientists)
- Over 300+ requests, pulled data from API or large database into R, wrangled data using R, and output figures and tables

**Echelon Insights**                                                                                 April 2020-Nov 2021
*Research Analyst*                                                                                           Alexandria, VA

- Led in modeling projects **predicting election turnout for entire U.S. in 2022**, phone response rates, etc.
- Wrangled, cleaned, weighted, or made presentations for 60+ survey datasets with R, SQL, and AWS
- Using R Shiny, built a codeless-crosstab tool for company's research team

**Center for Elections and Democracy**                                                                Dec 2018-April 2020
*Research Fellow*                                                                                                 Provo, UT

- Designed and executed multiple survey experiments in original research projects
- Mentored 60+ students in solving econometrics problems in weekly office hours
- Visualized data using R creating 50+ informative figures for AFS official report [here], news outlets, and professor's projects

## SELECTED PROJECTS

*Bayesian County-Level School Shooting Analysis*                                                            August 2022

- Compared and estimated several Bayesian Regression models with PYMC3 in Python. Ultimately, used **hierarchal negative binomial model** to predict shootings and make inferences about gun laws

- recieved award from Political Science department to fund a panel survey (n=1000).
- estimated linear regression and multinomial logit in observational study context and wrote paper [here] on findings.