

# CAROLYN LIU

[linkedin.com/in/carolynliu4](https://www.linkedin.com/in/carolynliu4) | [crlu4.github.io](https://crlu4.github.io)

## EDUCATION

---

### University of Chicago, Harris School of Public Policy

Master of Science in Computational Analysis and Public Policy

Chicago, IL

June 2023

- Relevant coursework: Machine Learning, Databases, Data Visualization, Program Evaluation, Large-Scale Computing

### Wellesley College

Bachelor of Arts in Economics, Minor in Mathematics, *Magna Cum Laude*

Wellesley, MA

June 2018

- Awarded the Natalie Bolton Faculty Prize in Econometrics (2017): Prize for the best econometrics group paper

## EXPERIENCE

---

### NYC Taxi & Limousine Commission (TLC)

*Data Analyst*

New York, NY

August 2023 – Present

- Produce and present to senior leadership high-quality, organized, and thorough data analyses, contributing to significant projects such as adjusting wait time thresholds for for-hire vehicle accessibility rules
- Utilize PowerBI and Python to develop, launch, and maintain TLC's public data dashboard: [TLC Factbook](#) (November 2023)
- Developed new methodology and ensured reproducibility for TLC's annual [Safety Honor Roll](#), coordinating with multiple teams to gather necessary data and feedback; reduced project time from one year to three months with minimal prior documentation
- Refactored and automated monthly data update processes for industry indicators and Vision Zero metrics, converting manual tasks in R and Excel to Python scripts

### Coding it Forward

*Civic Digital Fellow (Data Science) at US Census Bureau*

Remote

June 2022 - August 2022

- Conducted exploratory record linkage and analysis in Python between the Business Formation Statistics data product and other demographic datasets (Annual Business Survey & Nonemployer Statistics by Demographics) to develop a process (matching by unique identifier; resolving duplicates to select primary record) for an improved data product
- Presented findings to external audiences and senior internal leadership on the process for assembling a more equitable & improved data product to help businesses, policymakers, regional planners, and researchers assess the current state of early entrepreneurship at the national and state levels

### Harvard Business School

*Research Associate*

Boston, MA

July 2018 - July 2021

- Utilized Stata and SAS to collect, clean, and apply econometric models using administrative and financial data; example projects included the effect of internet and information inequality on job flows, impact of a salary disclosure law on the gender pay gap, & creating an adjusted living wage benchmark
- Collaborated in the development of and co-authored HBS cases used in the MBA curriculum by conducting company due diligence, interviewing company staff, and creating requisite data representations

## TECHNICAL SKILLS

---

- Programming: Python (numpy, pandas, scikit-learn, matplotlib, plotly), SQL
- Tools: Docker, Dash, Git, PowerBI, Databricks, ArcGIS
- Statistics/econometrics: OLS, randomized controlled trials, instrumental variables, regression discontinuity

## PROJECTS

---

### Data Science Clinic: Electronic Monitoring Device Shielding Detection (Python, Docker, SQL); Available upon request

- Worked on a team of 3 with an external stakeholder; designed & conducted field tests to build sufficient ground-truth on device GPS signal loss
- Built logistic and XGBoost models to identify whether device shielding was occurring; models had accuracy of ~80% (device vendor's built-in detection had accuracy of only ~50%)

### [Data Visualization - The Tate Collection through the Years](#) (D3.js, HTML/CSS, Python)

- Coded primarily in D3.js to conduct visual storytelling (used stacked bar, interactive treemap & modified beeswarm charts)