

# Yiliu Cao

✉ yiliu.cao@mail.utoronto.ca   ① yiliuc   in Yiliu Cao   ② yiliuc.github.io

## Education

**University of Toronto**  
*Honours Bachelor of Science*

Toronto, ON  
Sep 2020 - Apr 2024

- Double Major in Economics & Statistics, Mathematics Minor
- cGPA: 3.81/4
- Awards: Dean's List Scholar (2021 Summer, 2022 Winter, 2023 Winter, 2023 Summer)
- Core Courses: Statistical Machine Learning, Methods of Multivariate Data, Design and Analysis of Experiments, Time Series, Data Visualization

**University of Oxford**  
*Summer Abroad Program*

Oxford, UK  
Aug 2023 - Sep 2023

- Grade: 94/100
- Core Course: Big Data Tools and Applied Machine Learning for Economists

## Professional Experience

**United Nations Global Platform for Big Data China Hub**  
*Big Data Research Intern*

Hangzhou, China  
Jun 2023 - Aug 2023

- Used R and Python to process and analyze about 60,000 industries and examined patterns of industry flows as well as the potential reasons.
- Understood the data collection process from "State Grid" and learnt the practical application of statistical analysis in expected energy consumption by type in 2030.
- Engaged in three 'Beyond GDP Sprint' meetings organized by the UN Statistics Division to understand international statistical frameworks and their impacts.

**Boston Consulting Group Data Science Virtual Experience**  
*Data Science Intern*

Remote  
May 2023 - Jul 2023

- Engaged in multiple projects at BCG X, learning how to identify the clients' goals and perform the data analysis using Python and real data
- Assisted PowerCo in determining the price sensitivity for different clients by performing tasks such as Hypothesis Framing, Exploratory Data Analysis etc.

## Independent Research

**Causal inference between COVID-19 and 2020 US Presidential Election - Used R**

- Investigated the causal inference between COVID-19 and Trump's loss in 2020. Data comes from MIT EDSC and CSSE at JHU, roughly 40,000 observations.
- Employed propensity score matching coupled with an exploratory investigation of the best treatment, and also conducted a counterfactual analysis.
- Discovered that counties with extra death per case rate have a lower voting preference for Trump and Trump could re-elect if the inequality was addressed.
- Delivered key findings and insights through a presentation and received feedback from faculty members and fellow students.

**An Electoral Analysis for the 2020 US Election - Used Python**

- Preliminary study on the relationship between COVID-19 and economic factors in relation to Donald Trump's 2020 voting patterns.

- Used the election data in 2016 and 2020, and ACS 5-year estimates in 2016 and 2020 with above 30,000 observations
- Performed machine learning techniques, such as Regression Trees and Random Forest, and summarized with the best model
- Concluded that COVID significantly influenced Trump's 2020 votes, and high-income and counties that previously supported Trump in 2016 were less likely to vote for him

#### **Inequality of Infants Mortality Rate Between Black and White in US - Used R**

- Employed data from American Community Survey 2016 and CDC WONDER
- Analyzed the disparity in infant mortality rates between White and Black infants across various income-level counties, using OLS regression
- Concluded that White infants consistently have less mortality rate than Black infants but the inequality is less pronounced in high-income counties

#### **Assessing Racial Mortality Rate Changes During and Before Pandemic - Used R**

- Further research on the prior topic, examined the change in mortality inequality between Whites and Blacks during and before the pandemic by states and counties.
- Used data from ACS 2019, 2019 CDC WONDER, and CDC "Health Disparities" database 2023, with around 100,000 observations.
- Employed OLS regression to assess the variability in mortality among two races in terms of income and education levels.
- Concluded, although White people generally have less mortality rate change than the Black, factors such as age and health insurance are more crucial contributing to racial disparities in mortality rates during the pandemic.

### **Papers & Projects**

#### **STA304: Analysis of Popular Vote for the Liberal and the Conservative Party in the 2025 Canadian Federal Election - Used R**

- Introduced data from General Social Survey Data 2017 and Canada Election Survey Data 2019
- Used post-stratification to divide the population into cells based on age, sex and university completion
- Predicted that the Conservative Party will obtain a higher popular votes than the Liberal Party.

#### **STA313: Advocating for Sustainable Development within Canada - Used Tableau**

- Visualized Canada's energy consumption and its relationship with people's health, including the incidence rate of asthma etc.
- Created four interactive dashboards displaying energy usage, impacts on people's health, and future predictions using linear regression
- Collaborated with a team of six people and served as the leader for the Tableau session.

### **Extracurricular Activities**

#### **Henan Normal University Statistics Tutor**

*Jun 2023 - Jul 2023*

- Developed and delivered statistics tutor sessions to a group of Masters and PhD Physics students
- Systemically explained a wide range of statistical subjects from basic to more advanced theories

#### **University of Toronto Volunteer Notetaker Program**

*Jan 2023 - Apr 2023*

- Provided academic support to students with disabilities by taking detailed notes during lectures for two courses (STA305 and STA437)
- Received co-curricular credit for participation in the program

## Skills

---

- Language Skills: English (Proficient); Mandarin (Native)
- Technical skills: R, Tableau, Python, LaTeX, SQL, HTML/CSS, C++
- Packages
  - Python: Pandas, Numpy, Matplotlib, Sklearn (Sklearn.Ensemble, Sklearn.Metrics), Beautifulsoup, Geopandas, Seaborn, Request, Linearmodels
  - R: Tidyverse, ggplot2, Shiny, Openintro, Knitr, Kable, GridExtra, lsmens, MASS, boot, glmnet