

# Eric Li

Cell: 410-369-6751 E-mail: [xiaodalee@gmail.com](mailto:xiaodalee@gmail.com)

[Personal Website](#) [LinkedIn](#)

## EDUCATION

University of Maryland - College Park (Expected to graduate in May 2024)

Bachelor of Science: Statistics and Computer Science Dual Degree

Unweighted Cumulative GPA: 3.90 (out of 4.0)

## AWARDS/HONORS

UMD Presidential Scholarship

UMD University Honors Living Learning Program

Dean's List Fall 2020, Spring 2021, Fall 2021, Spring 2022, Fall 2022, Spring 2023

## TECHNICAL EXPERIENCE

Software: Python (pandas, numpy, matplotlib), R (dplyr, ggplot), SAS, MySQL, MATLAB, Java, C, Ruby

Skills: Data extraction, cleaning, visualization, web-scraping, linear regression, hypothesis testing, machine learning

Relevant Coursework: Data Science, Probability Theory, Statistical Inference, Real Analysis, Multivariable Calculus, Differential Equations, Linear Algebra, Algorithms, Object-Oriented Programming, Biology and Chemistry w/ Labs

Languages: English, Mandarin, Spanish

## WORK/RESEARCH EXPERIENCE

Big Data Summer Institute Participant, Ann Arbor, MI — June - July 2022

- Conducted and presented a [genomics research project](#) on differential DNA Methylation between ancestry groups with three other students. Developed quantitative analysis skills through an intensive, interdisciplinary training program in biostatistics, data science, and human health. Heavy use of R programming language and packages.

Undergraduate Research Assistant, College Park, MD — September 2021 - May 2022

- Performed data analyses related to natural disasters in order to conduct probabilistic risk assessments. Projects include analysis of Twitter data to investigate impact of hurricanes on public health centers, and [geospatial modeling](#) of predicted hurricane tracks in correspondence with nuclear power plant operating statuses.

## PROJECTS (FOUND ON GITHUB)

Men's Professional Tennis Data Analysis

- Created a [data science tutorial](#) using data from the ATP Tour. Used linear regression and hypothesis testing to investigate whether certain physical attributes such as height and dominant hand are significant advantages.

Gapminder Data Visualization/Machine Learning

- Using data from [gapminder.org](#), fitted a generalized linear model factored by continent to predict future life expectancies. Also experimented with classification algorithms such as decision trees and kNN. [Link](#)

MoneyBall Analysis

- Investigated MLB team salaries from 1990-2014 to determine if the Oakland Athletics really performed better than other teams while spending less money. Gained experience extracting data using SQL queries and creating visualizations with the matplotlib package in Python. [Link](#)

NASA Solar Flare Data Analysis

- Analyzed NASA solar flare data in order to gain experience working with and cleaning datasets. Web-scraped the data directly from the NASA website using BeautifulSoup. Used Pandas to clean and organize the data. [Link](#)