

Anthony Cu

San Jose, CA | (925) 854-8627 | anthonycu@ucsb.edu | [linkedin.com/in/anthonycu](https://www.linkedin.com/in/anthonycu) | [anthonycu10.github.io](https://github.com/anthonycu10)

EDUCATION

University of California, Santa Barbara (UCSB)

JUNE 2025

Bachelor of Science (B.S.) in Statistics & Data Science and Mathematics (Double Major).

GPA: 3.93/4.0

Minor: Science and Mathematics Education. Awards: **Phi Beta Kappa** (2024), **Dean's Honors**.

SKILLS

Languages/Tools: R (tidyverse, ggplot2), Python (Scikit-learn, Pandas, Matplotlib), SQL, GitHub, C++, CLI, LaTeX

Domains: Big Data (Visualization, Wrangling), Machine Learning, Design and Analysis of Experiments, API

Integration, Technical Documentation

RELEVANT EXPERIENCE

Chan Zuckerberg Biohub

MAY 2024 - JULY 2024

Data Science Intern

San Francisco, CA

- Analyzed a dataset of 30k genes across 60k experiments, compiled from a wide variety of sources, discovering underlying trends about library size and gene detection.
- Implemented and evaluated 5 data normalization methods to determine the most suitable approach.
- Visualized UMAPs to facilitate biological interpretations and synthesis with metadata.
- Identified 16 potential housekeeping genes in zebrafish, based on variability and detection rates.

Lean On Me National

JUNE 2022 - AUGUST 2022

Data Visualization Intern

Remote

- Created 100+ data visualizations, using a dataset of 1k user conversation histories, for display on the [Lean On Me National](#) and 12 chapter websites.
- Visualized stress levels before vs. after conversations to illustrate mental health impact, supporting each chapter in raising \$2k annually.
- Analyzed trends in conversation topics and user activity, providing actionable insights on how to improve supporter training protocols.

HIGHLIGHTED PROJECTS

Understanding Classroom Connections

SEPTEMBER 2024 - PRESENT

Education Research Study (Research Assistant)

- Designed a survey for 300+ students to analyze interactions while ensuring data privacy.
- Formulated a binary encoding and scoring system to quantify 9 different interaction types.
- Created sociograms to reveal social networks and patterns in study group formation based on 8 demographic factors and levels of nomination status.

[Spotify x Musixmatch \(Taylor's Version\)](#)

JANUARY 2024 - MARCH 2024

Text Mining Music Analysis (Independent Project)

- Utilized Spotify API and Musixmatch API to extract and build a dataset of 200+ song titles and lyrics.
- Developed and implemented a NLP pipeline using a bag-of-words model to identify 4 thematic clusters through unsupervised machine learning methods (e.g. K-means, hierarchical clustering).

Diseases of Affluence

AUGUST 2023 - MARCH 2024

Public Health Nutrition Research Study (Research Assistant)

- Cleaned and aggregated 2 large datasets: nutritional data (600k observations) and health data (800k observations), to analyze the correlation between wealth, diet, and disease.
- Enhanced dataset granularity by applying AI-driven techniques to enrich food descriptions, enabling more precise hypothesis testing.

LEADERSHIP EXPERIENCE

UCSB Statistics and Applied Probability Department

SEPTEMBER 2022 - JUNE 2025

Undergraduate Learning Assistant

Santa Barbara, CA

- Led sections of 30+ students and facilitated class lectures for 150+, adapting communication methods to accommodate diverse skill levels in data science principles (e.g. R, probability theory, SQL).
- Curated and organized a database of 700+ past course quiz questions in Google Sheets.