

**Nianyao (Amber) Du**  
dnyuianao@berkeley.edu | (510) 693-1132

## EDUCATION

**BA, University of California, Berkeley**

Expected May 2022

Double Major in Applied Math and Statistics | GPA:3.2

**Courses:** Introduction to Machine Learning, Numerical Analysis, Introduction to Analysis, Linear Algebra, Discrete Mathematics, Introduction to Abstract Algebra, Calculus, Data Structures, Concepts of Statistics, Game Theory, Concepts of Probability, Economic Analysis – Micro, The Structure and Interpretation of Computer Programs

## SKILLS

**Tools:** Python, SQL, R, Stata, MATLAB, Java, JavaScript, Git, QGIS, CSS, Scheme, Linux Environment (Commands, Shell Scripting), LaTeX, Excel, Power Point, Word, Jupyter Lab, Tableau, Google G Suite, Zotero, Mailchimp, Squarespace, CITI Training Certificate

**Key Skills:** Machine Learning (scikit-learn, sciPy, TensorFlow, PyTorch), Data Analysis (Regression, Classification), Time Series Analysis, Natural Language Processing (spaCy, Spark NLP), Data Science (NumPy, pandas), Data Visualization (Matplotlib, ggplot2), Web Scraping (xml2, rvest, Beautiful Soup), Data Cleaning, Geospatial Analysis, App Development (Shinyapps), Web Development, Software Engineering, Computational Social Science, Privacy Experiment Design, Presentation

## RELEVANT EXPERIENCE

**Research Assistant, UC Berkeley School of Law, Berkeley, CA**

Feb 2022 – Present

- Research the language ESG (Environmental, social, and governance) companies used in their earning calls and 10-k report with NLP techniques such as name–entity recognition, dependency parsing, and text classification
- Create SQL database to store, access, analyze, and manage research data

**Member of Data Consulting and Insights & Analytics, Student Association for Applied Statistics, Berkeley, CA**

Feb 2021 – Present

- Worked on a geospatial analysis client project to determine number and location of aircrafts that are hidden by cloud covers using spatial covariance model with Kriging on satellite images of airports, which reduced error rate by 20 percent
- Worked on a text analysis client project, used NLP models to tag, identify, and categorize customer feedback on their products
- Used NumPy, Pandas, and data visualization libraries to analyze world-wide coffee bean production and establish the relationship between country and coffee bean properties such as flavor and aroma
- Classified furniture product data such as color, price, and shape based on the idea of decision tree using libraries such as scikit-learn

**Co-Lab Manager & Project Lead & Budget Lead, Coleman Lab, Institute for the Study of Societal Issues, Berkeley, CA**

Jan 2019 – Present

- Lead a group of 12 to clean a data breach database provided by Privacy Rights Clearinghouse, conducted data analysis, built a procedure for collecting missing information, and co-author a book chapter on data breaches in higher education institutions
- Use sentiment analysis via Python to research company's privacy and cybersecurity considerations and establish a relationship between their privacy evaluations and stock market performances
- Collaborated with Annotation Database Team to construct a database using Zotero that students and clients can use to quickly access more than 500+ annotations of peer - reviewed academic articles written by lab
- Managed fund, kept track of budget, submitted reimbursement, and completed multiple grant applications
- Researched publications related to privacy in user experience and design experiments
- Completed Wikipedia training, researched spatial cloaking, wrote 20+ annotations, a blog and a Wikipedia article, and presented project to representatives from the Wiki-Ed Foundation and at a privacy symposium

**Research Associate, American Energy Society, Palo Alto, CA**

Sep – Dec 2021

- Built a pipeline with Python for analyzing 40,000+ rows of data in the reader reports of a biweekly publication with 10,000+ recipients, including time series analysis, visualizations, and recommendations for refining current structure of sending publications
- Implemented a standardized process for industry comparison based on comparing user clicks on energy related topics with general industry trends and developed a report on the monthly changes of energy related topic

**Research Assistant, Department of Politics, Princeton University**

Jul – Aug 2021

- Researched Varieties of Democracy dataset and built a new dataset by merging it with environmental indicators such as tree cover loss, biomass loss, and CO2 emission, and social development indicators such as GDP and population
- Conducted exploratory data analysis on merged dataset, including time series analysis and regression analysis on 4000+ lines of data
- Identified correlations between how environmental and societal indicators influence democracy levels

**Event Assistant, Summer Institute in Computational Social Science (SICSS) – Howard / Mathematica**

Oct 2020 – Jun 2021

- Reviewed instructional materials, prepared examples for group exercises, took notes on potential questions, developed resource lists, and assisted the development of daily action items for 30+ participants
- Worked on a blog post for the plenary speaker of the event which was published by SAGE Ocean as one of nine in a series called “The Future of Computational Social Science is Black”