Isai Garcia-Baza

https://isaigb.github.io/ github.com/isaigb linkedin.com/in/isaigarciabaza

Education

University of North Carolina at Chapel Hill **Chapel Hill, NC**
Ph.D. Education Policy, graduate minor in Computer Science *Expected May 2026*

- **Selected Coursework:** Causal Inference, Machine Learning, Natural Language Processing, Linear Regression, Panel Methods

University of North Carolina at Chapel Hill **Chapel Hill, NC**
B.A. Psychology with Honors *May 2017*

Data and Programming Skills

STATA: Advanced. 6+ years of experience. Develop internal tools for data cleaning, management, and quality monitoring.

Python: Beginner. 1 year of experience. Utilize scikit-learn, imbalanced-learn, statsmodels, pandas, NumPy, TensorFlow.

R: Beginner. 2 years of experience, mainly for coursework and RMarkdown.

Reporting and Data Visualization: Jupyter Notebook, Matplotlib , R Markdown, ggplot2, Quarto.

Version Control: Git, GitHub.

Data Collection: Survey (RedCap, SurveyMonkey, SurveyGizmo), focus groups, interviews in English, Spanish.

Statistical Expertise: Causal Inference, Statistical Modelling, Machine Learning, Econometrics.

Research Experience

UNC-CH School of Education **Chapel Hill, NC**
Graduate Research Assistant *Aug. 2021 - Present*

- NSF STEM (Github) – Python: `scikit-learn` , `imbalanced-learn` ; STATA
  - Analyze large-scale administrative data by first combining administrative data sourced from 5 databases across 16 institutions resulting in 13 million observation data set
  - Trained Random Forest and alternative classification models to predict student grades
  - Iterated over model prototyping, stakeholder input, and feature engineering cycles
  - Balanced data using SMOTE, Random Over Sampling, Random Under Sampling
  - Tuned models using K-fold cross validation and assessed using balanced accuracy to select best model
- SES Proxy (Github) – STATA, R: `tidycensus` ; Python: `census` , `statsmodels`
  - Analyzed American Community Survey (ACS) zip code-level data from the United States Census API
  - Conducted Principal Component Analysis (PCA) to identify proxy variables for socioeconomic status
  - Trained LASSO regression model for comparison against PCA regression
  - Compared models utilizing proxy variables against ground-truth administrative data to evaluate gaps in measures
- Misc.
  - Developed internal STATA tool for automatically generating data quality reports with variable labels
  - Strengthened causal inference claims by implementing propensity score matching

Child Trends **Bethesda, MD**
Senior Research Assistant *Aug. 2017 - Jan. 2021*

- Developed and analyzed survey for a randomized control trial evaluation of a positive parenting informational intervention
- Drafted internal and external reports of research findings
- Supported project with web scraping and human review of Twitter text data analysis
- Conducted usability testing of data visualizations
- Managed quantitative data and monitored national reach of news service using Meltwater data
- Served as coordinator for the Child Trends Institutional Review Board