

# Ben Ewing

DATA SCIENTIST

Oakland, California

☎ 209-535-2125 | ✉ [ben.ewing@duke.edu](mailto:ben.ewing@duke.edu) | 🏠 [improperprior.com](http://improperprior.com) | 📷 [ben-e](#) | 🌐 [ben-e](#)

## Education

---

### Duke University

MS IN ECONOMICS AND COMPUTATION

Relevant Coursework: Econometrics, Machine Learning, Bayesian Statistics, Artificial Intelligence

Durham, NC

2018 - 2020

### University of California, Santa Cruz

BA ECONOMICS AND MATHEMATICS

Santa Cruz, CA

2011 - 2015

## Skills

---

**Software:** Python (NumPy, Pandas, PyTorch, scikit-learn), R (tidyverse), Spark, Stata, SQL

**Methods:** Bayesian Statistics, Causal Inference, Econometrics, Machine Learning, Optimization

## Experience

---

### Duke University, Computer Science Department

GRADUATE TEACHING ASSISTANT

- Lead labs, office hours, and mini-lectures introducing common data science tools including **SQL**, **NumPy**, **scikit-learn**, and **matplotlib**.
- Advise students on data science and machine learning projects.

Durham, NC

2020 - Current

### Duke University, DevLab

GRADUATE RESEARCH ASSISTANT

- Authored an **R** package for programming XLSForm-compliant surveys.
- Built dashboards and automated reports to flag fraudulent data entry and monitor data quality using **R** and **Shiny**.
- Analyzed panel data for causal effects using difference-in-differences and machine-learning based matching methods using **Stata** and **Python**.

Durham, NC

2019 - Current

### Wells Fargo, Enterprise Analytics and Data Science

DATA SCIENCE INTERN

- Developed a model to optimally structure complex loans with multiple lenders.
- Explored and consolidated disparate datasets describing previous loans using **PySpark**.
- Implemented machine learning and optimization algorithms for each step of the loan optimization process using **pandas** and **scikit-learn**.
- Built a web application to provide banking analysts access to loan optimization tools using **Flask**.

Charlotte, NC

2019 - 2019

### The Cloudburst Group

DATA SCIENTIST

- Collected and analyzed data for five U.S. Agency for International Development land rights impact evaluations.
- Implemented regression based approaches and matching methods for causal inference using **R** and **Stata**.
- Used supervised and unsupervised learning techniques (among others: (random forest, PCA, clustering) for data exploration using **H2O**, **Python**, and **R**.
- Visualized results and wrote reports and presentations for technical and non-technical audiences using **R Markdown** and **ggplot2**.
- Programmed and tested surveys using ODK, SurveyCTO and R.
- Trained and managed data collection teams of up to 30 people in Liberia and Zambia.
- Monitored data quality by building dashboards and automated reports using **R** and **Shiny**.

Oakland, CA

2015 - 2018

## Projects

---

### Software: XLSFormTools

(GITHUB: [BEN-E/XLSFORMTOOLS](#))

2019 - Current

- Authored an R package which implements an embedded domain specific language for writing surveys compliant with the XLSForm standard.
- This package provides a plain-text alternative to survey programming in Excel, which allows users to take advantage of **Git** for collaboration and version control.
- Read my full blog post at [improperprior.com/post/a-language-for-survey-programming/](http://improperprior.com/post/a-language-for-survey-programming/).

### Project: Adversarial Knapsacks

(GITHUB: [NEILPRUTHI/AD\\_KNAP](#))

2018

- Explored a competitive extension of the knapsack problem where players attempt to form high value knapsacks full of random variables.
- Developed optimization and machine learning based solutions using **Python**, **R**, and **Julia**.