Zeyu (Zoey) Li

Tel: (812) 361-6967 **Email:** lizeyu@berkeley.edu **Website:** lizeyuyuz.github.io/lizeyu/

EDUCATION

Master of Science, Applied Statistics Sep 2016 - Apr 2018
University of Michigan, Ann Arbor

Bachelor of Arts, Economics and Statistics (Honors in Economics) Aug 2011 - Dec 2015
University of California, Berkeley

- Thesis - Earnings Inequality in Urban Labor Market: Discrimination and Decomposition of Wage Differentials.

WORK EXPERIENCE

Research Specialist Aug 2018 - Present
University of California, Berkeley

- Investigate the issue of racial bias generated by commercial risk prediction algorithms in the healthcare sector.
- Write R and SQL code on remote server to integrate biomarkers and laboratory test results to the master dataset. Develop an R shiny app to interactively presents summary statistics.

RELEVANT PROJECTS

Hierarchical Bayesian Analysis on Housing Prices Oct 2017 - Dec 2017
Bayesian Computation Project, University of Michigan

- Incorporated Bayesian simulation techniques in modeling housing priced with hierarchical models. Wrote STAN (efficient Hamiltonian Monte Carlos) scripts.
- Merged and cleaned large datasets of housing sales transactions using SQL, Python pandas, and R `data.table`. Imputed missing observations with random forest.

Clustering Analysis on NBA draftees Feb 2017 - Apr 2017
Machine Learning Project, University of Michigan

- Scraped NBA draftees' statistics with `XML`, `rvest`, `stringr` and `stringdist`.
- Performed analysis using PCA and MDS. Investigated groupings of NBA draftees using clustering methods such as K-means, K-means++, and mixture models.
- Experimented with labeling and classifying players using classification methods such as LDA, QDA, logistic regression, KNN, SVM and decision tree.
- See `https://github.com/lizeyuyuz/NBA_Clustering`

TECHNICAL SKILLS

Proficient: **R** (4+yr, `data.table`, `sqldf`, `ggplot`, `shiny`), **Python** (2+yr, `pandas`, `numpy`, `scikit-learn`, `scipy`, `matplotlib`), **SQL** (2+yr).
Intermediate: C++, Stata, SAS, MapReduce, Hadoop, Spark, HTML, CSS, XML (1-2 semesters).
Software: Excel, Git, LaTeX, RStudio, Tableau, vim, Unix.