Zeyu (Zoey) Li

536 South Forest Avenue
Ann Arbor, MI 48104
(812) 361-6967
lizeyu@umich.edu

SUMMARY

Graduate student in Applied Statistics with knowledge of statistical models, machine learning, bayesian statistics, web scraping, text processing and data visualization. Familiar with regression, logistic regression, decision trees, neural networks, clustering etc. Strong programming and software skills, including proficiency in R, Python, SQL, Git (command line), vim, Hadoop. More about me on my website: `https://lizeyuyuz.github.io/lizeyu`.

EDUCATION

Master of Arts, Applied Statistics, GPA 3.7 Sep 2016 - Expected Apr 2018
University of Michigan, Ann Arbor

- Coursework: Machine Learning, Bayesian Modeling and Computation, Big Data Manipulation and Analysis (Python, MapReduce, Hadoop, Spark), Computational Methods (R, Stata, SAS and Spark), Statistical Computing (C++), Data Science Computing (R packages), Probability Theory.
- Proposed thesis: "Bayesian Analysis of Schizophrenia fMRI Data""

Graduate Data Science Certificate Sep 2017 - Expected Apr 2018
Michigan Institute For Data Science, University of Michigan

Bachelor of Arts, Double Major in Economics and Statistics (with honors in Economics) Aug 2011 - Dec 2015
University of California, Berkeley

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

RELEVANT PROJECTS

MLB Web Scraping R Package Apr 2017
R Package Project, University of Michigan

- Created a fully functioning R package with documentation and test suite.
- Wrote functions that efficiently scrape baseball team and player statistics from websites using R packages such as `XML`, `rvest`, and `stringr`.
- Refined for fuzzy and incomplete search using `stringdist`.
- See MLB Web Scraping R Package on `https://lizeyuyuz.github.io/lizeyu/menu/projects.html`

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 exploratory analysis using PCA and MDS.
- Investigated groupings of NBA draftees using clustering methods such as K-means, K-means++, 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`

Handwritten Digit Recognition using Multivariate Logistic Regression in R and C++ Oct 2016 - Dec 2016

Rcpp Project, University of Michigan

- Wrote a rudimentary R Package with functions written in C++ through the use of Rcpp and RcppArmadillo.
- Used Multivariate Logistic Regression method as way of classification in handwritten digit recognition.

WORK EXPERIENCE

Research Assistant Jun 2015 - Aug 2015

Department of Statistics, UC Berkeley

- Developed dynamic web documents using R, HTML, CSS, and Shiny package by RStudio.
- Created an `R` package that covers a brief case study on factors that influence SAT scores as well as a tutorial that demonstrates Simpson's paradox. See `https://github.com/debnolan/DynDocs/tree/master/SAT`.

Research Assistant Jan 2014 - Dec 2014

Department of Economics, UC Berkeley

- Collected, cleaned and coded economics performance data of more than 1500 Chinese Special Economic Zones (SEZs). The results are used in the research paper, "Place-based policies in a development context—evidence from China," by Yiwen Cheng.
- Collected, cleaned and coded more than 20 sets of qualitative economic performance data. The results are used in the research paper, "Cities leading counties and patterns of urbanization in China," by Yiwen Cheng.

Research Assistant Sep 2014 - Dec 2014

Berkeley Law, UC Berkeley

- Assembled and coded more than 100 U.S. law school professor profiles in Excel.
- The results are used to augment a dataset used in the research paper, "Ideological Diversity and Law School Hiring," by James Philips.

SKILLS

Programming: R, Python, C++, Stata, SAS, SQLite, MapReduce, Hadoop, Spark, HTML, CSS, XML.

Software: Excel, Git, LaTeX, RStudio, RShiny, Vim, VirualBox, Unix.

Language Skills: English (Fluent), Chinese (Native), French (Limited Working Proficiency), Korean (Basic).