Michael Kim

michaelkkim@ufl.edu | LinkedIn | Website | GitHub

Education

University of Florida, Gainesville, FL — *Expected: 2022*
Ph.D. Statistics

University of Florida, Gainesville, FL — *August 2017 - December 2019*
M.S. Statistics

University of North Carolina at Chapel Hill, Chapel Hill, NC — *August 2013 - May 2017*
B.S. Statistics **B.A.** Mathematics

Coursework

MCMC | Probability Theory | Inference | GLM | Linear Models | Regression | Design of Experiments | Time Series | Optimization | Measure Theory | Real Analysis | PDE | Matrix Algebra | OOP | Data Structures | Machine Learning

Skills

Languages: R | Python | Java | LaTeX
Libraries: scikit-learn | NumPy | pandas | dplyr | ggplot | shiny
Statistics: Missing Data | Latent Variable Models | Bayesian Inference | Network Science | NLP | Clustering | Dimensionality Reduction

Research Experience

Graduate Researcher, University of Florida — *August 2019 - Present*
- Advised under Michael Daniels
- Missing data in latent variable models

Undergraduate Researcher, University of North Carolina at Chapel Hill — *January 2016 - May 2017*
- Advised under Professor Shankar Bhamidi
- Network visualization, network centrality analysis, PageRank algorithm implementation, logistic regression, and rank-score model on the citation network of the Supreme Court of the United States (SCOTUS)
  * Exploratory Analysis of the SCOTUS Citation Network – Shiny Apps | GIFs | Code
  * Centrality Analysis of the SCOTUS Citation Network – Visualizations | Code #1 | Code #2
- Natural Language Processing, clustering, and dimensionality reduction of the SCOTUS citation network and text data to return words that appear most frequently within clusters of the network and text data
  * Word Contextualization of Various Clusters of the SCOTUS Citation Network and Text Data – Shiny App | Code

Teaching Experience

Instructor, University of Florida
- STA 3032 (Engineering Statistics) – Fall 2019
- STA 2023 (Intro to Statistics I) – Summer 2018

Teaching Assistant, University of Florida
- STA 3024 (Intro to Statistics II) – Summer 2019
- STA 2023 (Intro to Statistics I) – Fall 2017, Spring 2018, Fall 2018, Spring 2019

Presentations

ASA DataFest, Duke University — *April, 2017*
- Presented on how Expedia can better recommend travel locations and vacation package deals based on a user's previous booking information and others' endorsements, through the following:
  * Data cleaning on 11 million observations from a dataset provided by Expedia
  * K-means clustering of 30,000+ travel locations based on endorsement likelihoods
  * Cross validation and logistic regression to accurately predict types of travel from booking information

PyData Carolinas, Research Triangle, NC — *September 2016*
- Presented on open-source legal data, graph packages, network structures, centrality measures, and community detection for exploring legal precedent and important SCOTUS cases
- Presentation with Iain Carmichael: "Open Data, Networks and the Law"

Publications

Carmichael, I., Wudel, J., **Kim, M.**, Jushchuk, J. (2017). Examining the Evolution of Legal Precedent through Citation Network Analysis *N.C. L. Rev., 96: 227-269*. Code

Awards

Grinter Fellowship — *August 2017 - Spring 2020*