# KATHERINE KEMPFERT

kempfert@berkeley.edu ⋄ https://kkempfert.github.io

## EDUCATION

**University of California Berkeley** Aug. 2019 - Present
PhD in Statistics
GPA: 3.87
Advisor: Dr. Christopher Paciorek

**University of Florida** Aug. 2015 - May 2019
Bachelor of Science in Statistics Summa Cum Laude
Bachelor of Science in Mathematics Cum Laude
Cumulative GPA: 3.96, Major GPA: 4.0
Advisor: Dr. Samuel Wong

## PUBLICATIONS

**Kempfert, K**. & Wong, S. (2020). Where Does Haydn End and Mozart Begin? Composer Classification of String Quartets. *Journal of New Music Research*, 1-20.

**Kempfert, K.**, Wang, Y., Chen, C., & Wong, S. (2020). A comparison study on nonlinear dimension reduction methods with kernel variations: Visualization, optimization and classification. *Intelligent Data Analysis*, 24(2), 267-290.

Yip, B., Bingham, G.*, **Kempfert, K.***, Fabish, J., Kling, T., Chen, C., & Wang, Y (2018). Preliminary Studies on a Large Face Database. *2018 IEEE International Conference on Big Data.* ***These authors contributed equally**

## AWARDS, SCHOLARSHIPS, & FELLOWSHIPS

**National Science Foundation Graduate Research Fellowship** 2020

**University of California Berkeley Chancellor's Fellowship** 2019

**Outstanding Four-Year Scholar Award** 2019
Awarded to less than 12 undergraduate students out of thousands graduating in May 2019 at the University of Florida

**Best Oral Presentation** 2017
Awarded for presentation *Nonlinear dimension reduction with kernel functions for computer vision tasks* at the 13th Annual Regional Mathematics and Statistics Conference at the University of North Carolina Greensboro

**Florida Academic Scholarship** 2015

## RESEARCH PROJECTS

**Generative Deep Learning for Rainfall Modeling** Oct. 2019 - Present
*Advisor: Dr. Christopher Paciorek*
*University of California Berkeley*

- Fit generative long short-term memory (LSTM) neural networks on rainfall time series from weather stations across the United States; generated thousands of synthetic time series from these fits; and analyzed the results to detect and characterize climate change

- Performed simulation studies to draw connections between state-space models and recurrent neural networks

**Predicting Classical Composers with Musical Scores** Aug. 2017 - May 2019
*Advisor: Dr. Samuel Wong*
*University of Florida*

- Developed novel features based on the sonata form that can be automatically computed from musical scores; classified the composer of Haydn and Mozart string quartets with logistic regression; and set benchmark results that exceed 84% leave-one-out classification accuracy
- Provided model-based interpretations about Haydn and Mozart that could be relevant to musicologists

**Nonlinear Dimension Reduction for Gender Classification** May 2017 - Feb. 2019
*Advisors: Dr. Cuixian Chen, Dr. Yishi Wang, & Dr. Samuel Wong*
*University of North Carolina Wilmington & University of Florida*

- Developed a novel machine learning pipeline for the large face database Morph-II; classified over 55,000 photographs in Morph-II as picturing either a male or a female; and reached over 95% cross-validated accuracy (competitive with benchmark)
- Compared the performance of kernel principal component analysis (KPCA), supervised KPCA, and kernel linear discriminant analysis via simulation studies and results on Morph-II

**Forecasting Dengue Fever in Brazil with Diverse Data Streams** June 2018 - Aug. 2019
*Advisors: Dr. Carrie Manore, Dr. Geoffrey Fairchild, Dr. Nidhi Parikh, & Dr. David Osthus*
*Los Alamos National Laboratory*
*Parallel Computing Summer School and Information Systems & Modeling (A-1) research group*

- Forecasted dengue fever with high accuracy and confidence for all 27 states of Brazil using time series variables from heterogeneous data streams (doctors' offices, weather stations, satellites, and Google Health Trends)
- Systematically compared predictive performance among variants of seasonal autoregressive integrated moving average (SARIMA), vector autoregression, seasonal trend decomposition, and ensembles combining these methods

## TEACHING

**Teaching Assistant for Introduction to Statistics (STA 2023)** Aug. 2016 - May 2019
*Professors: Maria Ripol, Megan Mocko, & Stephanie Stine*
*University of Florida*

- Led two to three lab sections of 40 to 126 students every week and introduced the main topics of statistics; engaged students in relevant activities, such as simulation of sampling methods, data collection, and use of statistical software Minitab; assigned worksheets for students to complete in lab; graded and recorded labs
- Held office hours every week, providing walk-in tutoring to students; proofread and proctored three exams each semester

## ORAL PRESENTATIONS

*Heterogenous Data Fusion of Time Series to Nowcast Dengue at the State Level of Brazil.* Presented at the Information Systems & Modeling (A-1) research group at Los Alamos National Laboratory, Los Alamos, New Mexico, 2019

*Parallel Forecasting of Dengue Fever in Brazil.* Presented at the Information Science & Technology Institute's Summer Schools Presentation Day at Los Alamos National Laboratory, Los Alamos, New Mexico, 2018

*Kernel-Based Nonlinear Dimension Reduction for Automatic Gender Classification.* Presented a contributed paper to Methods for Analysis of High-Dimensional Data session at Joint Statistical Meetings (JSM), Vancouver, British Columbia, 2018

*Forecasting Dengue in Brazil with Time Series Modeling in Parallel.* Presented at the Theoretical-Division Lightning Talk Series at Los Alamos National Laboratory, Los Alamos, New Mexico, 2018

*Nonlinear dimension reduction with kernel functions for computer vision tasks.* Won best oral presentation at the 13th Annual Regional Mathematics and Statistics Conference at the University of North Carolina Greensboro, Greensboro, North Carolina, 2017

## POSTER PRESENTATIONS

*Kernel Variants of Component Analysis and Discriminant Analysis for Gender Classification via Faces.* Presented poster at University of Florida Research Computing HiPerGator Symposium, Gainesville, Florida, 2018

*Forecasting Dengue in Brazil with Time Series Modeling in Parallel.* Presented poster at Los Alamos National Laboratory Student Symposium poster session, Los Alamos, New Mexico, 2018

*Kernel-Based Nonlinear Dimension Reduction for Face Analysis.* Presented poster at the Mathematical Association of America (MAA) Undergraduate Poster Session at the Joint Mathematical Meetings (JMM), San Diego, California, 2018

*Nonlinear Dimension Reduction Using Kernel Representations.* Presented poster at University of North Carolina Wilmington Summer Research Showcase, Wilmington, North Carolina, 2017

## SERVICE

**University of California Berkeley Statistics Graduate Student Association**

- Member of Social Committee — Aug. 2020 - Present
- Co-lead of Fellowships Committee — Aug. 2020 - May 2022
- Member of Service Committee — Aug. 2019 - May 2020

**Mentor for Berkeley Artificial Intelligence Lab** — Aug. 2019 - May 2020

**Mentor for Mentor UF** — Aug. 2016 - May 2017

## PROGRAMMING & SOFTWARE SKILLS

R, Python, Java, GIS, LaTeX