Randi H Griffin

Boston, MA

617-548-2608 | rgriff23@gmail.com | https://rgriff23.github.io | https://github.com/rgriff23 | www.linkedin.com/in/randigriffin

SKILLS

Languages: Python (pandas, scikit-learn, matplotlib, seaborn); R (ggplot2, tidyverse, caret); SQL; HTML/CSS
Statistics: Generalized linear models, survival analysis, multivariate statistics, meta-analysis, hypothesis testing
Machine learning: Classification, regression, clustering, feature engineering, NLP, dimension reduction (e.g., PCA)
Tools & techniques: git, Bash, R Markdown, Jupyter, web scraping, data visualization, simulation, survey data

EXPERIENCE

KAYAK Software Corporation, *Data Scientist (Performance Marketing)*   Cambridge, MA, Nov 2018 – Present

Insight Data Science, *Data Science Fellow*   Boston, MA, Sep 2018 – Nov 2018
- Developed a logistic regression pipeline in Python to predict which users will subscribe to a babysitting app.
- Generated novel features by geocoding 4K user addresses and linking them with geospatial census data.
- Built a dash app that allows the company to estimate the probability that new users will subscribe.

Duke University, *NSF Graduate Research Fellow*   Durham, NC, Sep 2013 – 2018
- Implemented multivariate GLMs in R to model 3D point clouds derived from CT scans to identify ecological predictors of skull morphology.
- Designed computer simulation studies to evaluate two new statistical methods for estimating ancestral states in evolutionary biology, demonstrating that the methods performed far less effectively than several alternatives.
- Wrangled 10 years of longitudinal data and used Cox survival models to provide the first demonstration that tapeworms can reduce mortality in wild primates.
- Demonstrated fine-scale habitat segregation in mosquito communities using GLMMs and PCA, providing a recommendation of <20 meters for the minimum resolution of spatial data in mosquito-borne disease models.

Harvard University, *Research assistant*   Cambridge, MA, Sep 2011 – 2013
- Maintained and queried a MySQL database with ~20K records of parasites reported in wild mammals.
- Performed formal meta-analysis of 14 published studies and 164 effect sizes, with results contradicting the popular claim that elevated parasite loads in wild animals are driven by human-caused habitat disturbance.
- Simulated pathogen transmission on social networks and identified network characteristics (clustering, centrality) that increase susceptibility to epidemic and endemic pathogens.

PROJECTS

'btw' R package: R wrapper for BayesTraits, an executable C++ program for fitting Bayesian phylogenetic models.
https://github.com/rgriff23/btw

Web scraping Olympic history data: Scraped and wrangled data on 135k Olympians from www.sports-reference.com. This dataset has been downloaded >10k times (top 0.3%) on Kaggle as of Sep, 2018.
https://github.com/rgriff23/Olympic_history

Twitterstorm analysis: Compiled data on 4.5k users and 5k tweets in a politically-charged Twitterstorm, then used social network and sentiment analysis to identify liberal and conservative clusters.
https://github.com/rgriff23/Katie_Hinde_Twitter_storm_text_analysis

StackOverflow developer survey analysis: Identified relationships between age, gender, and sexual orientation that influence feelings of membership in the programming community. This analysis won a $1000 Kaggle Kernel Award in June 2018. https://www.kaggle.com/heesoo37/stack-overflow-2018-survey-age-gender-sexuality

EDUCATION

Ph.D. in Evolutionary Anthropology, Duke University   May 2018
B.A. in Human Evolutionary Biology, Harvard University   May 2010