

# Randi H Griffin

I'm a Data Scientist with a background in the social/biological sciences and marketing. I'm passionate about crafting creative yet practical solutions to problems and deeply committed to fostering psychological safety in the workplace.

[rgriff23@gmail.com](mailto:rgriff23@gmail.com) | <https://rgriff23.github.io> | <https://github.com/rgriff23> | [www.linkedin.com/in/randigriffin](https://www.linkedin.com/in/randigriffin)

## SKILLS

**Programming:** Python (pandas, scikit-learn, matplotlib); R (ggplot2, tidyverse, caret), SQL

**Statistics:** Generalized linear models, survival analysis, time series analysis, network analysis, meta-analysis

**Machine learning:** Classification, regression, clustering, forecasting, feature engineering, NLP

**Other tools & techniques:** git, bash, Airflow, R Shiny, dash, Periscope, Google Data Studio

## EDUCATION

**Ph.D. in Biological Anthropology**, Duke University

May 2018

**B.A. in Human Evolutionary Biology**, Harvard University

May 2010

## PROFESSIONAL DATA SCIENCE EXPERIENCE

**Stand Up America, Data Consultant (contract role)**

Remote, Oct 2020 – Present

- Built Periscope dashboards to track Get Out The Vote initiatives for the 2020 presidential election.
- Conducted deep analysis of Google and Facebook marketing campaigns which contributed directly to a report which earned an Expy Award from the Analyst Institute: [https://members.analystinstitute.org/research/stand-up-america-electoral-advertising-program-11988?mc\\_cid=ad56434e15&mc\\_eid=74b6c8c8cb](https://members.analystinstitute.org/research/stand-up-america-electoral-advertising-program-11988?mc_cid=ad56434e15&mc_eid=74b6c8c8cb)

**KAYAK Software Corporation, Data Scientist (Performance Marketing)**

Cambridge, MA, Nov 2018 – Present

- Developed the company's first predictive model of keyword revenue-per-click to support our searchads algorithms, leveraging historical data and NLP techniques to provide smart estimates for long-tail keywords. I also designed a dash app with an interactive 3D TSNE plot to facilitate visual exploration of keyword predictions semantic space.
- Developed a rigorous counterfactual experimentation framework using Bayesian structural time series models for 1. Measuring lift in marketing experiments, 2. Identifying the best markets to use as treatment and control groups in future experiments, and 3. Estimating the statistical power of future experiments.
- Regularly contributed production Python and SQL code for ETLs and ad bidding algorithms.

**Insight Data Science, Data Science Fellow**

Boston, MA, Sep 2018 – Nov 2018

- Built a dash app for a babysitting service that automatically geocodes user addresses, links them with census data, and uses logistic regression to estimate the probability that new users will subscribe to the babysitting app.

**Duke University, NSF Graduate Research Fellow**

Durham, NC, Sep 2013 – 2018

- Used multivariate GLMs to identify ecological predictors of primate skull morphology based on CT scans.
- Conducted simulation studies to evaluate statistical methods for estimating evolutionary ancestral states.
- Wrangled 10 years of data and used survival models to quantify parasite-mediated 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

- 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.

## OPEN SOURCE CONTRIBUTIONS

**Parsons:** This Movement Cooperative project provides ETL connectors to integrate NGPVAN with other data sources commonly used by progressive organizations. I have contributed 2 new connectors, over 20 documentation PRs, and ~3000 lines of code: <https://github.com/move-coop/parsons/commits?author=rgriff23>

**'btw' R package:** R wrapper for the BayesTrait phylogenetic modeling software. <https://github.com/rgriff23/btw>

## INDEPENDENT DATA SCIENCE PROJECTS

**Web scraping Olympic history data:** Scraped data on 135k Olympians from [www.sports-reference.com](http://www.sports-reference.com) and shared the dataset and my analysis on Kaggle. The dataset has been downloaded >56K times on Kaggle as of Oct, 2020. <https://www.kaggle.com/heesoo37/olympic-history-data-a-thorough-analysis>

**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](https://github.com/rgriff23/Katie_Hinde_Twitter_storm_text_analysis)

**StackOverflow developer survey analysis:** Won a \$1000 Kaggle Kernel award for identifying relationships between age, gender, and sexual orientation that influence feelings of membership in the programming community. <https://www.kaggle.com/heesoo37/stack-overflow-2018-survey-age-gender-sexuality>

## UNIVERSITY TEACHING

**Northeastern University, Lecturer (Masters in Analytics)**

Boston, MA, Feb 2019 – Present

- Taught 2 semesters of the Analytics Capstone course, coaching 5 teams of students as they carry out a real data science project for a sponsoring company. Projects focused on marketing and learning analytics, respectively.
- Developed original materials and taught three quarters of Data Mining in R with excellent teacher ratings.

**AnthroTree Workshop, Organizer and Lecturer**

Multiple cities, 2012 – 2015

- Participated in every aspect of organizing an annual workshop on specialized statistical methods for PhD students, post-docs, and faculty in evolutionary biology
- Developed tutorial materials and lead a 3 hour workshop introducing the application of Generalized Linear Models to evolutionary datasets using R

## PEER-REVIEWED PUBLICATIONS

Fox, S.D., **Griffin, R.H.**, Pachankis, J.E. 2020. Minority Stress, Social Integration, and the Mental Health Needs of LGBTQ Asylum Seekers in North America. *Social Science & Medicine*, 246, 112727.

Schneider-Crease, I.A., **Griffin, R.H.**, Gomery, M.A., Bergman, T.J., and J.C. Beehner. 2017. High mortality associated with parasitism in geladas (*Theropithecus gelada*) in the Simien Mountains National Park, Ethiopia. *American Journal of Primatology*, 79(9).

Schneider-Crease, I.A., **Griffin, R.H.**, Dorny, P., Noh, J.C., Handali, S., Chastain, H.M., Wilkins, P.P., Nunn, C.L., Snyder-Mackler, N., Beehner, J.C., and T.J. Bergman. 2017. Identifying wildlife reservoirs of neglected taeniid tapeworms: non-invasive diagnosis of endemic *Taenia serialis* infection in wild primates. *PLOS Neglected Tropical Diseases*, 11(7): p.e0005709.

**Griffin, R.H.**, and G.S. Yapunchich. 2017. A critical comment on the ‘multiple variance Brownian motion’ model of Smaers et al. (2016). *Biological Journal of the Linnean Society*, 121(1): 223-228.

Reiskind, M., **Griffin, R.H.**, Janairo, M.S., and K.A. Hopperstad. 2016. Mosquitoes of Field and Forest: The Scale of Habitat Segregation in a Diverse Mosquito Assemblage. *Medical & Veterinary Entomology*, 31(1): 44-54.

**Griffin, R.H.**, and G.S. Yapunchich. 2015. The Independent Evolution method is not a viable phylogenetic comparative method. *PLoS ONE* 10(12): e0144147.

Coburn, R.A., **Griffin, R.H.**, & S.D. Smith. 2015. Genetic basis for a rare floral mutant in an Andean species of Solanaceae. *American Journal of Botany* 102(2): 264-272.

Young, H., **Griffin, R.**, Wood, C.L., and Nunn, C.L. 2013. Does habitat disturbance increase infectious disease risk for primates? *Ecology Letters*, 16(5): 656-663.

Cooper, N., **Griffin, R.**, Franz, M., Omotayo, M., and Nunn, C.L. 2012. Phylogenetic host specificity and understanding parasite sharing in primates. *Ecology Letters* 15(12): 1370-77. [Science Daily press release](#)

**Griffin, R.H.**, Matthews, L.J., and Nunn, C.L. 2012. Evolutionary Disequilibrium and Activity Period in Primates: A Bayesian Phylogenetic Approach. *American Journal of Physical Anthropology* 147:409-416.

**Griffin, R.H.** and Nunn, C.L. 2011. Community structure and the spread of infectious disease in primate social networks. *Evolutionary Ecology* 26(4): 779-800.

## ATHLETIC ACHIEVEMENT & COACHING

**South Korean 2018 Olympic Team and Women’s Ice Hockey National Team Player, 2015-2018.**

<https://today.duke.edu/2018/03/duke-olympian-will-soon-defend-her-phd>

**USA Hockey Certified Coach.** Completed 4 years of training and have coached youth teams from ages 12 to 19.