Peter David Smits

Seattle, Washington; 609-933-7042
peterdavidsmits@gmail.com

psmits.github.io/
github.com/psmits
linkedin.com/in/pdsmits/

Skills

Languages/Tools: R, Stan, Python, SQL, bash, LaTeX, git
Packages:

- **R:** brms, rstan, rstanarm, ggplot2, tidyverse, knitr, caret, shiny
- **Python:** pandas, numpy, beautifulsoup, keras, flask

Statistics: Bayesian data analysis, multilevel/hierarchical modeling, survival analysis, time series analysis
Machine Learning: regression, decision trees/CART, random forests, k-means, hierarchical clustering, regularization, feature engineering/selection, dimensionality reduction/PCA, NLP, neural networks

Experience

- **Data Science Fellow** Seattle, WA
  Insight Data Science *Sept 2019 – present*
  - Designed Copyprism, a web application for generating draft product descriptions from an image.
  - Implemented web scraper for collecting IKEA product descriptions in **python** using **beautifulsoup**.
  - Fine-tuned GPT-2 natural language deep learning model on IKEA catalog in **python** using **Google Colab**, whose generated text was indistinguishable from human text 26% of the time.
  - Deployed Copyprism web application on AWS using **flask**.
- **Postdoctoral Scholar** Berkeley, CA
  University of California – Berkeley *Sept 2017 – July 2019*
  - Designed multilevel discrete-time survival model using **Stan** to predict species extinction which identified species at risk of extinction within 1-million years with an AUC of 0.78.
  - Created multievel Bayesian time series model in **Stan** for predicting when rare extinction events were likely to be clustered in time based on geological information sourced from multiple databases.
  - Wrote nine chapters of online book on the analysis of paleontological and macroecological data using **R**, **tidyverse**, and **brms**.
- **Graduate Researcher** Chicago, IL
  University of Chicago *Sept 2012 – June 2017*
  - Identified how differences in mammal species ecologies affected their survival rates over the last 65 million years using a multilevel Bayesian survival model implemented in **Stan** and **R** applied to a database of fossil occurrences (approx. 40k rows) accessed via web API.
  - Created hidden Markov birth-death model for incompletely observed longitudinal data in **Stan** where transition probabilities were modeled as multilevel regressions in order to identify how speciation and extinction rates varied over time by species ecology and environmental state.
  - Developed ensemble framework for identifying species based on 2D shape information using **R** and **caret** where average of multinomial logistic regression, neural network, and random forest models correctly differentiated seven closely related turtle species with an AUC over 0.98.
  - Mentored and taught graduate and undergraduate students in R, statistics, Stan, and pedagogy.

Projects

- **mathhammr: R** package and **shiny** web application for generating dice rolls for Warhammer 40k and comparing those rolls to a simulated distribution of results.

Education

- **University of Chicago** Ph.D. Evolutionary Biology June 2017
- **Monash University** M.Sc. Biological Sciences Aug 2012
  - Vice-Chancellor's Commendation for Master's Thesis Excellence
- **University of Washington** B.S. Biology June 2010