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, Docker
Packages:

- **R:** brms, rstan, rstanarm, lme4, ggplot2, tidyverse, tidymodels, rmarkdown/knitr, shiny
- **Python:** pandas, numpy, beautifulsoup, flask

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

Experience

- **Data Scientist II** Seattle, WA
  Amazon – Advertising *January 2020 – present*
  - Developed Recommendation Impact, a data processing and modeling pipeline for measuring the effect of advertising campaign suggestions on daily advertising performance **R**, **Docker**, **ECR** and other AWS tools.
  - Designed multilevel/mixed-effects models for describing effectiveness of automated advertising campaign setup suggestions on display advertisement performance, with Maximum Likelihood and Bayesian implementations, written in **R** using **lme4** and **Stan**.
  - Coordinated with international team to measure and report the quality and value of 24 automated display advertising performance recommendations.
- **Data Science Fellow** Seattle, WA
  Insight Data Science *Sept 2019 – December 2019*
  - 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** whose generated text 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 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 multilevel Bayesian time series model in **Stan** for predicting when rare extinction events were likely to be clustered in time based on geological information across multiple databases.
  - Wrote nine lesson short course on analyzing paleontological and macroecological data using **R**, **tidyverse**, and **brms** which was used as curriculum for graduate-level paleontology course.
- **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 accessed via web API.
  - Created hidden Markov birth-death model in **Stan** for estimating fossil species observation, origination, and extinction rates over time, where each rate was modeled as its own multilevel regression which incorporated species ecology and environmental context.
  - 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 8th Edition 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