Peter David Smits

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

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

Skills

Languages: R, Stan, Python, SQL

Packages:

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

Statistics: Bayesian data analysis, hierarchical modeling, survival analysis, time series analysis, machine learning

Other: git, bash, LaTeX, cmdstan

Experience

- **Data Science Fellow** — Seattle, WA
  Insight Data Science — *Sept 2019 – present*
  - Designed Copyprisim, a web application for generating rough draft product descriptions from an image of the product.
  - Wrote web scraper for collecting Ikea product descriptions in python using beautifulsoup.
  - Trained text generation deep learning model in python using keras with TensorFlow backend.
  - Deployed web application using flask.
- **Postdoctoral Scholar** — Berkeley, CA
  University of California – Berkeley — *Sept 2017 – July 2019*
  - Designed a discrete-time survival model using Stan to predict probability of species extinction which generalized to correctly ranking which of two species is more likely to go extinct in 1 million years with a 79% probability.
  - Created a hierarchical Bayesian time series model in Stan for predicting when rare events (species extinctions) 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 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 from a database of fossil occurrences in space and time using a hierarchical Bayesian survival model implemented in Stan and R.
  - Modeled how the strength of factors influencing extinction risk change in response to average extinction rate increasing or decreasing estimated from incompletely observed longitudinal data using Stan and R.
  - Developed ensemble machine learning framework for differentiating between similar species based on 2d shape information in R which capable of correctly identifying seven closely related turtle species with an AUC over 0.98.
  - Mentored and taught graduate and undergraduate students in R, statistics, Stan, and pedagogy.

Education

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