Zachary M. Jones, Ph.D.

Seattle, WA // zmj@zmjones.com // zmjones.com // github.com/zmjones // (912) 401-8711

Summary

- a social scientist with a strong background in machine learning, statistics and associated software development with 7+ years experience doing research, communication skills from numerous lectures given, and the ability to learn quickly as needed.
- high-quality reusable software development experience with R, and daily use for 7 years with a focus on machine learning, data manipulation, and visualization.
- software development and data analysis using (7 years) Git/GitHub, GNU Make, Unix Shell, LaTeX, Emacs/Elisp, and Python (data processing and analysis as well as web applications).
- experience presenting technical and non-technical material to students, academics, and government/industy professionals (7 years).
- working with an international team on a large codebase (3 years) and managing a research team (3 years).

Experience

- Moore/Sloan Data Science Postdoctoral Fellowship, eScience Insitute, *University of Washington* (2017-2019).
  - automated collection and digitization and cleanup of images of treaty text from the United Nations' database in Python, coding of the content of these treaties by extracting features from the text and prediction of multiple human-coded labels using Python's `scikit-learn` library.
  - predicted political violence using interpretable machine learning software I developed in R resulting in several peer-reviewed publications.
  - implemented machine learning models to code archeological artifacts by period in R suggest new groupings to archeologists as well as evaluate standard groupings.
  - developed a web application to predict cyclist's power/duration dynamics using a semi-parametric anti-isotonic regression in Python using the Strava API.
- Google Summer of Code Fellowship, mlr: **M**achine **L**earning in **R** (2015)
  - implemented generic decomposition methods to make black-box machine learning output interpretable, and visualization methods for interpretation methods, performance and hyperparameter tuning. Participated in evaluation of hyperparameter tuning methods.
  - reviewed and approved code contributions and wrote technical documentation and user tutorials.
- Graduate Research/Teaching, *Pennsylvania State University*, 2014-2017
  - conducted dissertation work on intereptable machine learning methods which involved the development of several R packages implementing methods for decomposing black-box prediction functions and visualizing them resulting in several peer-reviewed publications which include applications of these methods as well as theory, and multiple talks.

  - developed large-scale statistical simulations for analysis of the latent space model for networks in `R` run on a `SLURM` cluster, and wrote and submitted papers for peer-review based on the simulation and analysis.
  - taught a graduate course on data science that covered applications of machine learning in a variety of disciplines.
  - developed teaching material for machine learning which were peer-reviewed and published. Code examples and simulations were written in `R`. Developed Software Carpentry style teaching materials for reproducible quantitative research.
  - developed and implemented hierarchical gradient boosting algorithm for grouped data in `R`.
- Graduate Research/Teaching, *University of Georgia* (2011-2013)
  - wrote web scraping software for analysis of United Nations Treaty database in Python later used in peer-reviewed publications.
  - performed large-scale analysis of data on human rights abuses written in `R`, run on Amazon's Elastic Compute Cloud and later published in the discpline's flagship journal. Managed software development using Git/Github. Presented research at several large research conferences.
  - implemented system for tracking drug trafficking patterns using open source data on retail prices using `R` and Python.

Education

- Ph.D. Political Science (statistical and computational methods), *Pennsylvania State University* (2017).
- M.A. Political Science, *University of Georgia* (2013).
- B.A. Political Science and Philosophy, *Georgia Southern University* (2010).

Selected Publications, Software, and Talks

1. Interpretable Statistical Learning Methods (under review)
   - talks at New Faces in Political Methodology (2018), the University of Colorado: Boulder (2016), and the International Methods Colloquium (2015)
2. mmpf: Monte-Carlo Methods for Prediction Functions, *The R Journal* (2018), Zachary M. Jones.
3. Is There More Violence in the Middle?, *American Journal of Political Science* (2018), Zachary M. Jones and Yonatan Lupu.
4. edarf: Exploratory Data Analysis using Random Forests, *Journal of Open Source Software* (2016), Zachary M. Jones and Fridolin Linder.
5. mlr: Machine Learning in R, *Journal of Machine Learning Research* (2016), Bernd Bischl, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio and Zachary M. Jones.
6. Enhancing Validity in Observational Settings When Replication is Not Possible, *Political Science Research and Methods* (2016), Christopher J. Fariss and Zachary M. Jones.
7. An Empirical Evaluation of Explanations for State Repression, *American Political Science Review* (2014), Daniel W. Hill Jr. and Zachary M. Jones.
8. Git/GitHub, Transparency, and Legitimacy in Quantitative Research, *The Political Methodologist* (2013).