Daniel Ryan Furman | Programming Competency Statement

GitHub Profile

Table of contents: Python, R, C++, OCaml, Zsh, MATLAB, ArcGIS

1. Python. The language I have most applied experience with. **Applied Work:** Mudd Data Science REU ('19-present), Ice Physics Lab ('18-20), and OGWR statistical report ('19). **Explanation:** ML cross-validation tuning and validation error pipelines, feature engineering, feature selection, raster-based analytics, data mining and pre-processing, data visualization, unsupervised learning, and constitutive physics modeling. Some of these programs are centrally based on recursion (e.g. feature selection), while others develop object-oriented classes (e.g. MCMC parameter space exploration). I am comfortable in writing style guide enforcement (e.g. flake8) for Python programs, and with leveraging well-developed third party libraries (see resume for examples). I am proudly an open-source contributor to PyImpute (a high-level geo-classification library), updating Sci-kit Learn pipelines and geo-data processing in several of its functions. For more detail on my Python programming, see my GitHub repos "Furman-DS-programs", "ensemble-climate-projections", and "Furman-and-Goldsby".

- **Courses:** *Stats, Data Mining, Machine Learning*; University of Pennsylvania; A | *Data Analysis in Earth Science*; University of Pennsylvania; A | *Modeling Geographic Space*; University of Pennsylvania; B+. *See notes in appendix for more information on these courses.
- **More Python Background:** Python was my first language (followed second by MATLAB and R). In my ice physics research, I collaborated on Python programs with geophysicists including David Goldsby, Douglas Jerolmack, and Andrew J. Cross. At Harvey Mudd, I collaborated on Python programs with statistician Tanja Srebotnjak and ecologist Stephen Adolph. Python is also my go-to language for work, as was the case in my OGWR analyses.

2. R. The language I have second-to-most applied experience with. **Applied Work:** Mudd Data Science REU ('19-present), Australian Wet Tropics bird abundance project ('18) (applied statistics research during my semester abroad). **Explanation:** I implement R scripts for statistical modeling, data mining and pre-processing, geo-spatial statistics, random aerial sampling, alpha hull analyses in three dimensional Principal Component space, and similarity clustering (ie. with Jaccard metric). In the Harvey-Mudd Data Science REU, I worked heavily with primarily R statistician Dr. Tanja Srebotnjak.

- **Courses:** *Statistical Data Mining for Big Data*; University of Pennsylvania; A | *Statistics for Biologists*; University of Pennsylvania; A-.

- **R Course Note:** *Statistical Data Mining for Big Data* was (~4/5) oriented to R; yet, I was also introduced to coding in SQL, NoSQL and Apache Hadoop, including parallel processing and hierarchical schema assignments.

3. C++ (as well as OCaml). Introductory exposure to topics such as object-oriented data structures, data structures, program design, value-oriented programming, lists and recursion, nested patterns, abstraction/modularity, and the transform/fold higher-order functions. Languages employed for self-study and an array of audited courses as an undergraduate at the University of Pennsylvania. These experiences inform the programming practices I use in Python and R, as well as my curiosity to learn more about software engineering and production-level code as a graduate student.

4. Zsh & Shell scripting. I rely on the terminal for a multitude of tasks, to set up my Python programming environments, to manage my Mac's files, and to compile code (such as C++ in the AWS Cloud9 IDE). I also rely on the terminal for management of my GitHub repositories and for pull-request collaborations.

5. MATLAB. Employed for topics in mathematics. **Courses:** *Numerical Mathematics*; University of Pennsylvania; A | *Computational Linear Algebra*; University of Pennsylvania; A. **Explanation:** Proficiency to write, from scratch, numerical methods in errors and floating point numbers, solutions of equations, polynomial interpolation, calculus, approximation of functions, solutions of differential equations, solving linear systems equations, vector spaces, orthogonality and the Gram–Schmidt algorithm, determinants, eigenvalues and eigenvectors, and linear transforms, as well as run time analysis across these topics.

6. ArcGIS. Cartographic modeling for a wide variety of geo-spatial analytics. **Courses:** *Modeling Geographic Space*; University of Pennsylvania; B+ (double-listed with Python). **Explanation:** With coursework from the inventor of Map algebra (Dana Tomlin), I am versed in the nature and use of raster-based information. Fluency with Python source-code for ArcGIS functions, as well as with the writing code with the ArcGIS API for Python. Concepts from the course were critical in my geo-modeling research, in many of the raster-based analyses I have implemented in Python and R.

Appendix. *Python course explanations: *Stats, Data Mining, Machine Learning* instilled fluency in coding object-oriented classes. My assignments often employed linear algebra computation with Numpy (ie. coding without relying on external libraries) for MCMC parameter space exploration, maximum-likelihood estimation, bootstrap resampling, forward model selection, ChiSquare error ellipses for linear models, as well as writing functions for pipelines in machine learning, PCA, and deep learning. In the morphological galaxy classification final, my cross-validation tuned Random Forest classifier tied for the most predictive among ~35 projects (for more detail see my GitHub "Furman-DS-programs" folder "machine-learning-statistics"). In contrast, *Data Analysis in Earth Science* first focused on functions and program decomposition, time-series analysis, style-enforcement, and engineering programming practices, secondly on data analytics, e.g. writing numerical methods in geoscience from scratch: random walks, pattern recognition, covariance and correlation, modeling flow, and image analyses.