Neal Grantham

Full Stack Statistician in San Francisco, CA. Email: neal@nsgrantham.com
nsgrantham.com | github.com/nsgrantham | twitter.com/nsgrantham | linkedin.com/in/nsgrantham

Education

Ph.D. in Statistics, North Carolina State University, Raleigh, NC	Aug 2017
Master of Statistics, North Carolina State University, Raleigh, NC	Aug 2014
B.S. in Mathematics, B.S. in Statistics, California Polytechnic State University, San Luis Obispo, CA	Jun 2012

Experience

Data Scientist, Phylagen, San Francisco, CA   Oct 2017 — present
- **Machine Learning**: Led research and development on custom machine learning platform for forensic geoanalysis of microbiome data from environmental sources.
- **Data Engineering**: Built a production-level, end-to-end data analysis pipeline with steps for data acquisition, ETL, large DNA sequence file processing, REST API queries, machine learning analysis, and report generation.
- **Startup Generalist**: "Wearer of many hats" at <20 employee biotech company, contributing to work with government agencies, industry partners, and VC investors, leading to a successful Series A fundraising round.

Graduate Research and Teaching Assistant, North Carolina State University, Raleigh, NC   Aug 2012 — Aug 2017
- **Spatial Modeling**: Constructed a spatial algorithm to identify the geographic origins of dust samples using microbial DNA sequences — a "biogeographical fingerprint," funded by the US Department of Defense.
- **Bayesian Inference**: Developed a Bayesian hierarchical mixed-effects model for the analysis of high-dimensional microbiome data as a response variable from designed experiments.
- **Statistics Education**: Adapted core statistics curriculum to 90+ students with backgrounds in business, biology, and the social sciences. Trained students in statistical thinking via "flipped classroom" approach.

Intern, NASA Langley Research Center, Hampton, VA   Jun 2015 — Aug 2015
- **Time Series**: Classified cognitive state of airline pilots using eye-tracking, biophysical, and electrical brain signals during critical flight scenarios using hidden Markov models.
- **Collaboration**: Worked with aerospace, mechanical, and biomedical engineers on open problems in aviation safety.

Undergraduate Research Assistant, California Polytechnic State University, San Luis Obispo, CA   Jun 2011 — Jun 2012
- **Self-Learning**: Characterized the clustering patterns of colonial marine invertebrates in capstone Senior Project, requiring independent study of spatial statistics.

Skills

R: tidyverse, ggplot2, Shiny, proficient writing R packages and contributing to open source
Python: pandas, numpy, scipy, keras, scikit-learn, Flask, Beautiful Soup, proficient writing Python modules
SQL: PostgreSQL, SQLite, and database front-ends SQLAlchemy and dbplyr
Unix: Bash scripts, cron jobs, tmux, vim, zsh, dotfiles
Statistics: Bayesian hierarchical models with MCMC, generalized linear models, experimental design
Machine Learning: Deep neural networks, random forests, hidden Markov models
Containers and Scale: Docker, AWS EC2, RDS, S3
Web Dev: Knowledgeable in HTML, CSS, and JavaScript including D3 and Leaflet
Reproducible Research: RMarkdown, Jupyter Notebook, LaTeX
Other Software: GSuite, Airtable, Notion

Projects

tidytuesday.rocks: Tweet catalogue made with shiny | nsgrantham.shinyapps.io/tidytuesdayrocks
ggdark: Dark mode extension for ggplot2 | github.com/nsgrantham/ggdark
uspops: R package with US state population data since 1900 | github.com/nsgrantham/uspops