

AVA HOFFMAN

Data Scientist

avamariehoffman@gmail.com

(804) 687-7476

/in/avahoffman

avahoffman

www.avahoffman.com



EXPERIENCE



Fred Hutch Cancer Center / Johns Hopkins University · Senior Staff Scientist

May 2021 - present · Seattle, WA (remote)

- Built outreach software for publishing content, previewing data, and helping scientists comply with NIH policies using Shiny, GitHub Actions, and other tools
- Co-lead, analyzed, and curated data for **BioDIGS**, a soil microbe genetics research consortium of over 100 students and faculty
- Co-lead outreach and user research for the NIH **AnVIL Platform**, enabling genomic data science via cloud computing for 500+ users
- Led and managed \$579k NIH-funded program ("**GEMs**") that designs open-source genomic data science training modules with under-resourced colleges
- Led, developed, and managed \$200k/yr **Data Science for Environmental Health program**, educating 40 trainees and public health government professionals annually
- Taught graduate-level courses in R programming and community data science consulting



Johns Hopkins University · Postdoctoral Fellow

Mar 2020 - May 2021 · Baltimore, MD

- Used bioinformatic tools to quantify genomic patterns among 1000+ plants adapted to 6 city landscapes



Boston Consulting Group · Data Scientist

Mar 2019 - Mar 2020 · Boston, MA

- Productionalized PySpark pipeline describing and modeling client's 20K+ commercial banking customers for better product recommendations & risk intervention
- Designed & hosted fully customized RShiny dashboard for reducing internal carbon emissions by 30% in 5 years



Colorado State University · Researcher

Aug 2013 - Mar 2019 · Fort Collins, CO

- Led the Blue Grama Diversity Project, implementing hierarchical Bayesian models and analyzing 9,000+ genomic features to understand plant traits from population genetics

EDUCATION



PhD · Ecology

Colorado State University · 05/2019



BS · Biology

University of Virginia · 05/2012

TECHNICAL SKILLS

Programming languages · Python, R

Python packages · scikit-learn, pandas, NumPy, SciPy, pytest, statsmodels, seaborn, matplotlib, Jupyter

R packages · tidyverse, RStan, shiny, bioconductor, ggplot2 + add ons, rMarkdown/bookdown, RStudio

Version control · Git, GitHub, Bitbucket

Statistics · Bayesian methods, multivariate statistics, permutational methods, dimensionality reduction, repeated measures

Supervised learning · time series, random forest, linear/nonlinear/logistic regression, LDA, SEM

Unsupervised learning · PCA, k-means/hierarchical clustering, feature engineering

Cloud technologies · GCP, GitHub Actions

Other technologies · Docker, PySpark, SQL, APIs, Tableau, Alteryx, SAS

ACHIEVEMENTS



Excellence in Teaching Award

2022 - 2025 (7x) · Johns Hopkins University



29 peer-reviewed publications

2010-2025 · [Google Scholar](#)