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)

- Led user research for NIH AnVIL Platform through survey design and interview analysis, identifying key feature priorities to increase platform adoption
- Developed interactive data planning dashboards with R Shiny, improving NIH policy compliance for NIH scientists
- Co-developed and deployed automated content publishing software using GitHub Actions and R that reduced manual content publishing time processing time by 50%
- Designed and implemented statistical analyses for BioDIGS soil microbiome project, processing 5+ TB of genomic sequencing and soil property data across 300+ samples and 100+ students and faculty
- Secured and managed $579k NIH grant through data-driven proposal development, creating genomic data science curriculum among 6 faculty at 4 institutions
- Established the $200k/yr Data Science for Environmental Health program, teaching and evaluating data science pedagogy using repeated measures ML approaches
- Taught R programming courses focusing on data wrangling, reproducibility, and machine learning applications in public health

Johns Hopkins University · Postdoctoral Fellow

Mar 2020 - May 2021 · Baltimore, MD

- Developed and implemented custom bioinformatic pipelines in R and Python to analyze 1TB+ of genomic data from 1,000+ plant samples
- Implemented evolutionary isolation models to understand genetic differences among distinct city environments
- Visualized complex genomic relationships using dimensional reduction techniques and genetic structure plots, leading to a publication in preparation

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, using logistic regression and random forest methods
- Created interactive RShiny dashboard integrating multiple data sources for carbon emissions tracking, enabling leadership to explore strategies for reducing internal carbon emissions by 30% in 5 years
- Optimized and clustered client's vehicle inventory with Python and k-means clustering approach, improving confidence in leadership's nationwide product delivery decisions

Colorado State University · Researcher

Aug 2013 - Mar 2019 · Fort Collins, CO

- Led statistical analysis for Blue Grama Diversity Project, implementing hierarchical Bayesian models in R to analyze 9,000+ genomic markers across 300+ geospatially distributed plant samples
- Developed the first custom computational pipelines for next-gen sequencing analysis in several species, improving reproducibility and future research capabilities
- Created reproducible analysis workflows using R Markdown and version control

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

Statisics · 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, QGIS, Tableau, Alteryx, SAS

ACHIEVEMENTS

Excellence in Teaching Award
2022 - 2025 (7x) · Johns Hopkins University

29 peer-reviewed publications
2010-2025 · Google Scholar