

## **Michael A. Babb, PhD**

[babb\\_mike@outlook.com](mailto:babb_mike@outlook.com) | (206) 295 - 3612

### **Data Scientist**

**University of Washington**

**January 2016 – December 2021**

Design and implementation, in python and R, of data engineering pipelines using linear optimization and linear, logistic, and Poisson regression modelling resulting in a 30-percent increase in the study population (136M to 178M) and an 8-fold increase in complete cases.

Trained and deployed 155K regression models in a parallelized R computing environment.

Design and implementation in R of parallelized data visualization pipelines.

Data engineering, explanatory regression modelling, and visualization of historical and contemporary data using python, R, and SQLite.

Consultation and instruction in machine learning, data visualization, feature engineering, geospatial analysis, and cartographic visualization using python, R, and ArcGIS.

### **GIS Analyst**

**Zillow.com**

**August 2013 – January 2016**

Authorship of geospatial feature engineering and machine learning algorithms in python and R.

Feature engineering in support of live-site operations and maintenance using python and R.

Data-driven analysis of the technical feasibility and business-need suitability of dataset acquisition and incorporation into production, at-scale, machine learning pipelines.

Authorship of new geospatial feature engineering algorithms in python and R that led to a 10-fold increase over then-existing algorithm efficiency and coverage.

Data-driven business-needs assessment of the strengths and efficiencies of existing geospatial feature engineering algorithms.

Use of R to perform model diagnostics leading to the identification of biases in production machine learning models.

### **Geographer**

**U.S. Census Bureau**

**April 2012 – October 2014**

Proposal development and project management of federal-academic research projects.

Site operation and administration of a secure-site Federal Statistical Research Data Center.

Completed build out and received certification to operate a federally compliant Federal Statistical Research Data Center (1 of 16 at the time of opening).

### **Software proficiency**

Python – 10 years

Pandas, NumPy, sci-kit learn, multiprocessing

R – 10 years

data.table, tidyverse, baseR, doParallel

SQL – 10 years

MS SQL Server, PostgreSQL, SQLite

ArcGIS – 15 years

### **Data Science**

Supervised learning – 10 years

Unsupervised learning – 10 years

Data visualization – 10 years

Feature engineering – 10 years

Unstructured data – 10 years

Text analysis – 5 years

Geospatial data engineering – 15 years

### **Education**

**Ph.D., Geography, University of Washington, 2021**