Dana Paige Seidel

Data Scientist with 10 years experience in data wrangling and statistical modeling applications in biology and manufacturing.

CONTACT

 @dpseidel dpseidel dana-seidel

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY

PhD in Environmental Science, Policy, & Management
July 2019 | Berkeley, CA

UNIVERSITY OF ALBERTA

MSc in Biological Sciences
Aug 2014 | Edmonton, AB, Canada
Concentration in Ecology

CORNELL UNIVERSITY

BSc in Natural Resources
May 2011 | Ithaca, NY
Concentration in applied ecology
College of Agriculture and Life Sciences
Graduated *cum laude*

SKILLS

- Data Analysis • R • Tidyverse
- Geospatial Analysis (R, Python, ArcGIS)
- Biostatistics • Regression Analysis
- Experimental Design
- Stitch • Sisense (PeriscopeData)
- Data Visualization • R Shiny
- DBT (data-build-tool)
- Python (numpy, pandas, pymc3)
- Open Source Development in R
- Version Control, Git/Github
- Testing & Continuous Integration
- Teaching • R Markdown • LaTeX

AWARDS

- Data Science for the 21st Century: National Science Foundation Research Traineeship
- Outstanding Graduate Student Instructor Award, UC Berkeley 2018
- Letter of Commendation for Excellence in Teaching, Univ. of Alberta 2012

PUBLICATIONS

For full list of my publications, please see Google Scholar

WORK EXPERIENCE

RIVIAN | Senior Data Scientist • Decision Science - Factory Data
Oct 2021 - Present | Palo Alto, CA
Support factory ramp and data-driven decision making within Rivian's manufacturing engineering org.

PLENTY | Staff Data Scientist • Software & Data
July 2019 - Oct 2021 | South San Francisco, CA

- Aggregate multiple data streams from production farms (e.g. environmental signals, operations, imagery) to develop complex derived features and metrics for downstream models, alerting, and business users
- Fit canopy height and harvest yield models in Python (pymc3) to continuously forecast growth curves and expected harvest yield from plant cohorts
- Develop and maintain deployed model parsing canopy temperature and detecting hotspots in production images

RSTUDIO | Software Development Intern • Tidyverse
June 2018 – Aug 2018 | Remote

- Full-time developer of widely-used open source R packages for data visualization, ggplot2 (486K downloads/month) and scales
- Worked within large existing codebases to resolve issues, submit bug fixes, add new features, write unit tests, and update documentation to provide increased functionality for user-defined scales, themes, and aesthetic manipulation for data visualization in R

GOOGLE | Quantitative Analyst Intern • GeoData Analytics
May 2016 – Aug 2016 | Mountain View, CA

- Queried, manipulated, and analysed mobile and web generated user-impressions data for tens of millions of Google Maps features
- Built and evaluated predictive models for average time to "maturity" of user-interactions with novel business features
- Fit spatially-implicit mixed effect regression models and evaluated model capacity for predicting user-interaction with novel business features

PROJECT EXPERIENCE

Graduate Student Researcher | UC Berkeley

- Developed a R package to streamline best practices and automated report building for exploratory data analysis of telemetry data
- Analyzed multi-banded MODIS and LandSat imagery to understand environmental context of animal movement data

Modeling Chronic Wasting Disease | Alberta Fish & Wildlife

- Built an R Shiny application with infrastructure to load raw data, extract environmental covariates from rasters, model and map spatial risk
- Used Python (e.g. GeoPandas, Fiona, Shapely, ArcPy) & R (e.g. sf, raster, velox, mapview) libraries to manipulate spatial raster and vector data for analysis
- Built spatiotemporal models estimating disease prevalence and spread using a hierarchical Bayesian framework in WinBUGs and R

Southwest Alberta Montane Research Project | U Alberta

- Designed relational database and analysis flow for field-collected foraging & movement data from 182 radio-collared elk
- Implemented multivariate regression analyses in R, including mixed negative binomial regression and paired conditional logistic regression