All icons are hyperlinks

Last Updated on 17th January 2021

Dana Paige Seidel

Senior Data Scientist with 10 years experience in data wrangling and statistical modeling specifically for spatial and time series data.

CONTACT

dana.seidel@gmail.com
612.232.6708 @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 • SQL
- 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

PLENTY | Senior Data Scientist • Data Science & Metrology

July 2019 - Present | South San Francisco, CA

- Build, maintain, test, and surface Plenty's core reporting data, leveraging a modern data science pipeline including Stitch, Snowflake data warehouseing, DBT, & Sisense
- Aggregate multiple data streams from a production farm (e.g. environmental signals, material traceability, equipment performance, farm imagery) to develop complex derived features and metrics for downstream models, alerting, and business users
- Build and deploy statistical models in R and Python to create novel data streams, characterize system performance and health, and predict harvest yield

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 regression 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