Cavan Donohoe

[cavandonohoe@gmail.com](mailto:cavandonohoe@gmail.com) | [github.com/cavandonohoe](https://github.com/cavandonohoe) | [linkedin.com/in/cavan-donohoe](https://www.linkedin.com/in/cavan-donohoe) | (805) 404-3321

### Looking to improve your company’s data workflow

I have a level of expertise in both R and Excel that helps me both manipulate and analyze data as well as automate monthly tedious tasks. With the mind of a tutor, I make sure to document everything and answer questions quickly and efficiently when I explain my methodology to coworkers or clients. Let’s have a chat so I can help your data from beginning to end.

## Professional Experience

### Senior Data Scientist

#### GRAIL

#### Menlo Park, CA

February 2021 - August 2024

* Authored and maintained new R Packages used on cross production within entire company
* Worked extensively with R (tidyverse, dplyr, ggplot2) and SQL (PostgreSQL) for data wrangling, validation, and integration across large-scale clinical datasets
* Created new R Markdowns to display hundreds of TFLs (Tables, Figures, and Listings) in order to assess clinical trial accuracy
* Created and documented new Derived Variables for ongoing studies using git/Gitlab for collaboration, R for coding, and AWS S3 for cloud storage and data management using CDISC standards (SDTM, ADaM, CDASH) for clinical trial data management (raw EDC data)
* Created a timeline visualizer for significant cancer screenings and events leading to an eventual diagnosis using R Shiny Output
* Maintained complete and auditable programming documentation for analysis of clinical studies
* Provided statistical programming support to data cleaning and locking activities, pre-defined and exploratory analysis, formal reports, publications, presentations, and new statistical methodologies
* Worked closely with biostatisticians to create data and analysis program specifications based on the statistical analysis plan (SAP) and Statistical Process Control
* Contributed to the development, documentation and maintenance of a reusable programming code library while devoloping reproducible R pipelines integrating statistical modeling, simulation, and visualization for trial-level analyses
* Wrote unit tests to ensure programmatic updates will be compatible to previous versions
* Utilized R programming to analyze survival data and assess risk factors
* Managed the installation and configuration of R packages across multiple environments, ensuring compatibility and stability
* Performed jackknife and bootstrap resampling simulations in R to assess model robustness, estimate uncertainty, and support clinical trial design evaluations.

### Reporting Analyst

#### 24 Hour Home Care

#### El Segundo, CA

April 2020 - January 2021

* Automated a weekly report that calculates the Accounts Receivable aging for our company. This weekly report is now being calculated in R and automatically sends an email formatted in HTML with CSS to show summary tables and a highlighted list of achievements; saved 5 hours of time every week
* Imported, Cleaned, Manipulated, and Visualized data for ad hoc dashboard requests
* Assumed role of Analytics team liaison to other teams within the company
* Automated batch invoice emails via R for use between teams to run
* Connected R to Azure Web App to allow for script driven data exports now allowing for automated exports saving teams hundreds of hours per year

### Actuarial Analyst

#### HMSA

#### Honolulu, HI

January 2019 - March 2020

* Automated the CRG (Community Rated Groups) Band Renewal Process in R (8 hour process that now takes 1 minute)
* Developed the demographic and plan risk factors for AMS and Underwriting Department (now automated in R)
* Worked on mandatory ACA rate filing
* Automated the Mental Health Parity process for hundreds of plans
* Automated the Forecasting Model with user friendly Excel and R integration for obtaining data and easing the process
* Automated a tedious CRG proposal process that previously took an hour and now takes 4 seconds
* Explained processes with easily read documentation and hosted meetings to explain modeling changes and methodologies
* Fit empirical distributions to theoretical distributions using Kolmogorov-Smirnov tests and Maximum Likelihood Estimation in R while splitting every single partition of a sample in a time efficient manor

## Education

### Bachelor of Science

#### University of California, Irvine

#### Irvine, CA

2018

* Major in Mathematics
* Minor in Statistics

## Awards

### BAS UCLA Case Competition Finalist ([Project Presentation](http://www.math.ucla.edu/~actuary/caseCompetition/Case%20Comp%202018%20-%20Team%20UCI%20Presentation.pdf) and [Project Memorandum](http://www.math.ucla.edu/~actuary/caseCompetition/Case%20Comp%202018%20-%20Team%20UCI%20Executive%20Summary.pdf))

## Credentials

### Passed 3 Actuarial Exams

## Skills

* Expert in Microsoft Excel and Wide Knowledge of VBA (pivot tables, vlookups, sumifts, etc.)
* Expert in R (tidyverse, RODBC, blastula (email automation), openxslx, APIs, automation, data manipulation, R Markdown, bootstrap/jackknife methods, etc.)
* Expert in SQL (PostgreSQL)
* Literally made this [resume](https://cavandonohoe.github.io/cv.html) and [website](https://cavandonohoe.github.io/) using R