David G. Robinson

✉ admiral.david@gmail.com · 646-620-6630

varianceexplained.org

Work

Period	**September 2019 — Present**	
Position	**Data Insights Engineering Manager**	
Company	**Flatiron Health**	New York, NY

- Develops data products of cancer data on the Structured Data Platforms team

- Manages three data scientists

Period	**January 2018 — September 2019**	
Position	**Chief Data Scientist**	
Company	**DataCamp**	New York, NY

- Led and mentored a team of nine data scientists, analysts and engineers

- Managed the development of a data pipeline for ingesting and cleaning data in a Redshift data lake

- Built technical platforms and packages using R and Shiny Server to enable internal analytics and support six departmental dashboards

- Represented the needs of data science to the leadership team

Period	**June 2015 — January 2018**	
Position	**Data Scientist**	
Company	**Stack Overflow**	New York, NY

- Built internal data platforms and R tools to enable effective analysis and experimentation

- Trained and tested machine learning models for ad targeting and job matching

- Published posts on the company blog with insights from our data, leading to 1.8 million unique visits

Education

Period	**September 2010 — May 2015**	
Degree	**PhD in Quantitative and Computational Biology**	
University	**Princeton University**	Princeton, NJ

Period	**September 2006 — May 2010**	
Major	**A.B in Statistics, *cum laude***	
University	**Harvard College**	Cambridge, MA

Books

- Silge, Julia and Robinson, David. (2017) Text Mining with R: a Tidy Approach. O'Reilly.

- Robinson, David. (2017) Introduction to Empirical Bayes: Examples from Baseball Statistics.

Selected Open-Source Packages

- broom (Creator): R package for converting statistical models to tidy data frames

- tidytext (Co-creator): R package for text mining using dplyr, ggplot2, and other tidy tools

- fuzzyjoin (Creator and maintainer): R package for merging tables based on inexact matching

- widyr (Creator and maintainer): R package for widening, processing, and then re-tidying a dataset

- gutenbergr (Creator and maintainer): R package for searching and downloading public domain texts from Project Gutenberg

Selected Talks

- **The unreasonable effectiveness of public work**. rstudio::conf 2019 (Keynote)

- **Teach the Tidyverse to Beginners**. rstudio::conf 2018

- **We R What We Ask: The Landscape of R Users on Stack Overflow**. EARL Boston 2017 (Keynote)

- **broom: Converting statistical models to tidy data frames**. useR 2016

Selected Articles

- Robinson, D.G. (2015) broom: an R package for converting statistical analysis objects into tidy data frames. arXiv preprint. arXiv:1412.3565 [stat.CO].

- Robinson, D.G., Wang, J., and Storey, J.D. (2015) A nested parallel experiment demonstrates differences in intensity-dependence between RNA-seq and microarrays. Nucleic Acids Research, 43(20):e131. doi: 10.1093/nar/gkv636.

- Robinson, D.G. and Storey, J.D. (2014) subSeq: Determining appropriate sequencing depth through efficient read subsampling. Bioinformatics; doi: 10.1093/bioinformatics/btu552.

- Robinson, D.G., Chen, W., Storey, J.D., and Gresham, D. (2014) Design and Analysis of Bar-seq Experiments. G3: Genes/Genomes/Genetics, 4(1), 11-18

- Robinson, D.G., Lee, M.C. and Marx, C.J. (2012) OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. Nucleic Acids Research, 10.1093/nar/gks778.

Skills

Programming Languages	R, Python
Data	PostgreSQL, MySQL, AWS S3, Redshift, Airflow
Tools	Git, Markdown, UNIX