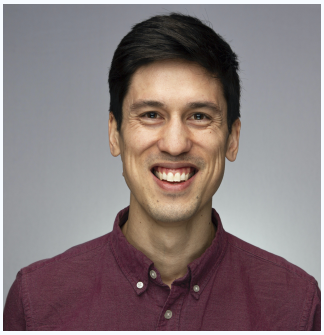
RICH PAULOO, PHD

I spend most of my days writing code (mostly `R`, `Python`, `SQL`) to clean, visualize, and model data. I have a PhD in computational hydrogeology, where I simulated and visualized 3D contaminant transport in aquifers.

I'm an expert-level #rstats user. A few projects I'm proud of include `R` packages to query water quality data 📦 [1] and text yourself from `R` 📱 [2], `R` data science curriculum 📚 [3], a dashboard that makes millions of water quality observations understandable 📈 [4], and a model that predicts the risk of wells going dry 💧 [5] funded by Microsoft's AI for Earth Grant.



View this CV online with links at *richpauloo.com/cv*

CONTACT

- richpauloo@gmail.com
- RichPauloo
- github.com/richpauloo
- richpauloo.com
- (415) 275-4981

FOREIGN LANGUAGE SKILLS

Conversational Spanish

EDUCATION

2020 | 2015

PhD, Computational Hydrogeology
University of California Davis — Davis, CA

- Published 6 scientific papers (3 first-author).
- Tools used: R, Python, SQL, git/Github, bash, AWS, cron, dplyr, ggplot2, shiny, flexdashboard, leaflet, sf, MODFLOW, RW3D, Paraview, Illustrator, ArcGIS, Envi, LaTeX
- Won ~$153,000 in national, compeitive grants and awards from NASA, Microsoft AI for Earth, AGU, and others.

2011 | 2006

B.S., Integrative Biology (minor in Conflict Resolution)
University of California Berkeley — Berkeley, CA

PROFESSIONAL & RESEARCH EXPERIENCE

present | 2022

Data Science Consultant
Accenture — San Francisco, CA

- Aligned with Applied Intelligence West

present | 2020

Co-Founder & Principal Data Scientist
Water Data Lab — Remote

- Managed $105k in annual contracts for specialized data science consulting. Managed technical teams.
- Developed reproducible pipelines for the largest water quality and spatial database covering 306.88 million people served by 44,919 water systems (97.2% of the US population).
- Co-developed r4wrds.com

The source code is available at github.com/richpauloo/cv.

Last updated on 2022-08-24.

2022 | 2020

Data Scientist

Larry Walker Associates — Berkeley, CA

- Built and automated ETL pipelines for ~180 real-time sensor networks and dashboards that process > 100,000 daily observations.
- Turned messy data into actionable information and automated reports.
- Tools used: R, Python, SQL, git/Github, bash, AWS, cron, dplyr, ggplot2, shiny, flexdashboard, leaflet, sf
- Managed multiple six-figure contracts, scoped work, contributed to strategic marketing, and trained staff.

2020 | 2018

Data Engineer

UC Water — Davis, CA

- Developed a monitoring dashboard with interactive data visualization using AWS with R, SQL, Shiny, and Shiny Server. I also built an automated ETL pipeline that pulled data from an IoT sensor network to feed the dashboard. REsults were peer-reviewed and published[6].

2019 | 2018

Data Lab Researcher

Computational Institute for Geodynamics (CIG) — UC Davis

- NLP, text mining, and network analysis in R on a corpus of ~600 PDFs.
- Developed a R Shiny dashboard[7] to understand the corpus.
- Results were peer-reviewed and published[8].

SELECTED DATA SCIENCE WRITING

2020 — **Automating R scripts on Linux with cron**[9]

2019 — **Race to the Bottom**[10]

2019 — **Text Analysis of the Mueller Report**[11]

2018 — **Tidy Chi Squared stats in infer**[12]

View all of my blog posts here.

LINKS

1: https://caopenwater.github.io/sdwisard/
2: https://github.com/richpauloo/textme

3: https://r4wrds.com/
4: http://calwaterquality.com/
5: https://www.gspdrywells.com/
6: https://doi.org/10.3390/w12041066
7: https://richpauloo.shinyapps.io/cig_nlp
8: https://ieeexplore.ieee.org/document/8827910
9: https://www.richpauloo.com/post/crontabs/
10: https://www.richpauloo.com/post/race-to-the-bottom/
11: https://www.richpauloo.com/post/mueller/
12: https://www.richpauloo.com/post/infer/