

Dustin Stansbury, PhD

Staff Data Scientist with over 15 years of technical experience spanning world class research, machine learning, analytics, and software development. Seeking to contribute to the world through technology, innovation, education, and art.

[Email](#)
[LinkedIn](#)
[Personal Website](#)
[Blog](#)
[Github](#)

PROFESSIONAL EXPERIENCE

Gap Year, Berkeley, CA — *Personal Development*

JANUARY 2023 - PRESENT

Took a sabbatical away from industry to continue my education, pursue personal projects and hobbies, and recalibrate life and career goals.

- Developed [spearmint](#), a python package that simplifies AB Test analysis.
- Wrote a free [AB Testing application](#) powered by spearmint.
- Audited the popular [Statistical Rethinking](#) course on causality, Bayesian inference, and scientific philosophy. Rewrote all the course's R code in Python/PMC5; [those codes](#) are now part of the official course resources.
- Consulted for a sustainable machine learning [startup](#).
- Wrote many blog posts (I even [published](#) some of them).
- Learned Spanish (intermediate, conversational level).
- Traveled to Europe (my first time! Great way to practice Spanish).
- Improved my music production chops.
- Currently working on an interactive book teaching causal inference.

Patreon, San Francisco, CA — *Staff Data Scientist, Growth*

JUNE 2022 - JANUARY 2023

- Technical and thought leader, drove the development of processes and technology for running quality online experiments (AB testing) at large scale.
- Improved the number of successful experiments run by over 400%.
- Worked closely with cross-functional leadership to define and refine business strategy and metrics for driving and monitoring growth across the business.
- Developed forecasting system linking improvements in various funnels to future revenue growth; system now used for opportunity sizing and planning across the org.
- Hired and mentored numerous [L2-L5](#) Data Scientists.

Numerai, San Francisco, CA — *Staff Data Scientist*

SEPTEMBER 2021 - APRIL 2022

- Expanded both R&D efficacy and trading capabilities by designing and implementing a unified trading system that supports backtesting, paper trading, and live trading.
- Improved the organization's R&D flexibility and efficiency by designing and implementing a generalized portfolio optimization system that supports multiple trading strategies and optimizer backends.
- Contributed to the fund's alpha by improving upon existing, and developing novel machine learning models, as well as exploring candidate feature sets used in the public Numerai tournament.

SKILLS

Statistics, Machine Learning, MLOps, Experiment Design, Data Visualization, Software Design & Engineering, NLP, Data Engineering & ETL, Recommendation Systems, Predictive Analytics, Business Intelligence, Product Development, Quantitative Finance

TECHNOLOGIES

Programming Languages:

Expert: Python, SQL, Julia, MATLAB

Proficient: R, Go, C++, C, Java, Groovy, Scala, Ruby, Javascript, PHP

Analytics & Data Visualization:

pandas, PyMC, Stan, statsmodels, seaborn, Streamlit, Tableau, Looker, Periscope, Mode

Machine Learning:

PyTorch, TensorFlow, Spark, Databricks, sklearn, xgboost, cvxpy, gensim, Prophet

Data & MLOps:

AWS stack—S3, EC2, RDS, Aurora, Redshift, Sagemaker, Batch

GCP stack—BigQuery, GCS, Data Store, Big Table, GKE, Kubernetes

Relational Databases—MySQL, Postgres, Redshift, BigQuery

NoSQL—Redis, MongoDB, Hadoop, Elasticsearch, Cassandra

Graph Databases—Tinkerpop/Gremlin, JanusGraph, Neo4j/Cypher

Orchestration & Deployment—Apache Airflow, Docker, Terraform, K8s, GKE, DBT

Miscellaneous & Frameworks:

bash, Git, poetry, pipenv, conda, HTML/CSS, REST, CRUD, Plotly.js, D3.js, Node.js, Vue.js, FastAPI, Flask, Rails, django, Jekyll, Domino DataLab

Quizlet, San Francisco, CA — *Staff Data Scientist*

SEPTEMBER 2016 - SEPTEMBER 2021

- Designed and implemented a large-scale knowledge graph that now powers many features in the application.
- Designed and implemented internal framework used to manage the training, comparison, and deployment of the organization's machine learning models.
- Designed and implemented the company's site optimization and AB Testing system, including the statistical analysis backend, which is now released as the public python package, [abracadabra](#).
- Designed and implemented a scalable computational framework for classifying billions of text records into hierarchical academic taxonomies.
- Wrote and maintained ETL systems in Apache Airflow to perform regular analytics, data management, and machine learning tasks.
- Wrote multiple [blog posts](#) for the company that publicly share approaches to solving novel research and engineering projects.
- Hired and mentored multiple data scientists and product analysts.

8tracks, San Francisco, CA — *Principal Data Scientist*

OCTOBER 2014 - SEPTEMBER 2016

- Enabled real-time, personalized programming by designing, implementing, and maintaining key recommendation systems.
- Enabled the application to be systematically and incrementally improved by designing and implementing a site optimization experiment framework.
- Enabled the analytics team to be more data-driven by designing and implementing a predictive analytics framework used by multiple teams to automatically segment users.
- Designed, implemented, and maintained multiple ETLs responsible for data processing, machine learning, royalty reporting, and business intelligence metrics.

University of California, Berkeley, CA — *Postdoctoral Researcher*

MAY 2014 - SEPTEMBER 2014

Used Neuroimaging (fMRI) and Machine Learning techniques to [study](#) how humans visually process the environment.

University of California, Berkeley, CA — *Graduate Student Researcher*

AUGUST 2007 - MAY 2014 - [DISSERTATION](#)

Studied functional analogs between artificial neural networks (i.e. "Deep Learning") and biological neural networks in the primate visual system.

EDUCATION

University of California, Berkeley, CA — *PhD, Visual Neuroscience*

AUGUST 2007 - MAY 2014

Used machine learning, computer vision, and signal processing methods to devise models of neural function in the visual system. Validated models on large scale biological data sets recorded from neurophysiological and neuroimaging experiments.

PROJECTS

- [theclevermachine Blog](#)
- [spearmint](#)
- [refreshing-ab-testing](#)
- [statistical-rethinking](#)

INTERESTS

Open source software, running, rock climbing, cycling, travel, reading, music (listening & creating)

University of North Carolina, Asheville, NC — *Supplemental Education in Statistics and Computer Science*

AUGUST 2006 - MAY 2007

Coursework: Multivariate Statistics & Probability Theory, Discrete Mathematics, and Algorithms and Data Structures

Appalachian State University, Boone, NC — *BS, Physics & Psychology*

AUGUST 2000 - MAY 2005

Minor in Applied Mathematics

Coursework: statistics, linear algebra, calculus I-III, differential equations, computational physics (C programming), optics, visual perception, aesthetics, experimental methods and design