Dav J. W. Clark, Ph.D.

Data and Learning Scientist

917-544-8408 | davclark@gmail.com | 1201 W Mount Royal Ave #512, Baltimore, Maryland
https://www.linkedin.com/in/davclark | https://github.com/davclark

Summary

I have a strong track record of getting things done and executing on collaborative data-intensive projects across academic, enterprise, consulting and start-up settings. My experience includes computational linguistics, statistics, machine learning, static and interactive visualization, and physical computing. I'm a fast learner looking to join a great team.

Experience

*Open source projects indicated with * are linked at the end of the document.*

Research Scientist, Kennedy Krieger Institute, Baltimore, MD 2016-2017

Program implementation and evaluation spanning survey, clinical, physiologic and brain-imaging data; hiring and supervision of instructional staff; partnerships; grant and academic writing; Data and Mental Health Working Groups in OSI-funded School Climate Collaborative; developed Taichi Site for web-based video curriculum in Elm*

UC Berkeley, Berkeley, CA

Fellow, Berkeley Institute for Data Science (BIDS) 2014-2016

*Co-founded BIDS collaborative incubator for data-intensive projects with real-world impact; Discussant in numerous panels on data collection, sharing, and ethics; Resident at ManyLabs (a Moore-funded science hackerspace in San Francisco); contributions to rpy2 project**

Data Scientist, D-Lab 2013-2016

Training and empowering data-driven social scientists, grant-writing, specified and administered "fat" compute node (ZFS, GPU), ran corporate training, managed undergraduate staff, trained staff and organized hosting for projects and curricula on GitHub, EdX (MOOC) Data Czar, Presenter at Bloomberg Data for Good Exchange, Training Assistant at U Chicago Data Science for Social Good, big financial data including marketflow library, developed BCE* data science VM for instructional and research needs in collaboration with Research IT and Statistics department, UC Berkeley SPOT Award for outstanding service*

Instructor, Machine Learning (Masters in Data Science); Hacking Measurement 2014, 2015

Developed missing "catch-up" curriculum for scientific python, automated GitHub course workflows, organized Hacking Measurement as an open project site*

Chief Scientist, Oroeco / Startup Chile, San Francisco, CA / Santiago, Chile 2012-2015

Analysis of user behavior, Rails development, implemented multiple testing frameworks to encourage developer usage

Partner, KeepOpen Web Design, Euless, TX (remote) 2004-2013

Concierge-style web design and hosting using Silva (a Zope-based CMS)

Principal Scientist, Entrieva (now LucidMedia), Reston, VA 2006-2007

Analyzed and reported on the core business algorithm that no employees understood, wrote new algorithm that was comprehensible to taxonomy team, Tomcat/Spring

Education

Ph.D., Psychology, UC Berkeley 2007-2013

Thesis: Climate change and conceptual change
Research in Cognition and Math Education (RCME) Fellow, helped start and architect the NiPype project, developed MTurk Admin* for longitudinal online experiments*

MS, Cognitive Neuroscience, MIT 1999-2002

Thesis: Neurocognitive circuitry supporting neoword learning
Harvard Business Plan Competition runner up ($4000 in-kind services), MIT Sloan Business Plan Competition semi-finalist, President of MIT Grad Student Volunteer Corps

BA, Linguistics; BS, Computer Science; BS, Mathematics, *U of MD, College Park* 1995-1999

Thesis: Modeling language change with markov models*; Magna Cum Laude*
Founder of Juggling Club - extensive youth outreach and education

Skills

Programming **- Data Science:** Python, R, Spark, SQL, MongoDB, HDF5; **Visualization:** ggplot2, Matplotlib, Bokeh, SVG (D3); **Front End:** Elm, HTML, CSS, JavaScript / React, Jekyll; **Back End:** Ruby on Rails, Express, Python, Tomcat

Ops **-** Amazon EC2 & MTurk, Docker, Ansible, Packer, Vagrant, Linux admin, system building, GPU

Science **- Data Collection:** brain imaging, physiology, computerized testing (desktop / web / mobile), surveys / crowd-sourcing, experiment design, Arduino & Raspberry Pi; **Statistics:** classical, Bayesian, non-parametric, time-series & spatial, machine learning

Writing **-** technical, grants, documentation, policy / MOU

Sample Open Source Projects

marketflow - efficient ingest of large financial datasets http://marketflow.readthedocs.io/en/latest/

rpy2 - interoperate with R from Python http://rpy2.readthedocs.io/

NiPype - data processing graphs for brain imaging analysis http://nipype.readthedocs.io/

BCE – the Berkeley Common Environment, a VM for data science http://bce.berkeley.edu/

Hacking Measurement - incubating grad student projects http://hackingmeasurement.berkeley.edu/

MTurk Admin - longitudinal experiments on Amazon MTurk https://github.com/davclark/mturk_admin

Taichi Site - static site using Elm for video-based practice https://github.com/davclark/taichi-site

Hotaru - controlling LPD8806 LED strip using React.js (hobby) https://github.com/davclark/hotaru-material-ui