Brian P. Abelson

New York City / brianabelson@gmail.com / http://abelson.live

Profile

Experienced Data Scientist and Engineer, Open Source Software Developer, Digital Journalist, and Civic Hacker with a track record of using data to solve problems, tell stories, and make people dance.

Skills

- **Data Science**
  - Machine Learning, Natural Language Processing, Computer Vision, Neural Networks, GIS, Visualization, Web Scraping.
- **Software Development**
  - Applications, APIs, Database Management, Automation, DevOps, Testing.
- **Digital Journalism**
  - Creative and inventive storyteller skilled in carrying out acts of journalism online.

Technologies

- **Software**: Python / Ruby / Node / Java / Go
- **Stats**: R / Hadoop / Hive / Spark / Stata / Excel
- **Frontend**: HTML / CSS / Javascript
- **DBs**: Redshift / SQL / ElasticSearch / MongoDB / Redis / DBT
- **GIS**: PostGIS / CartoDB / ArcGIS / QGIS
- **Visualization**: D3.js / ggplot2 / Looker
- **Automation**: Docker / Serverless / Chef / Ansible / Vagrant
- **Cloud**: AWS / Google Cloud / Heroku
- **OS**: Mac / Linux / Ubuntu

Experience

Kickstarter (Data Engineer, 2017-2018)

- Rearchitected Kickstarter's aging data infrastructure in two months, moving to a combination of Airflow, DBT, Redshift Spectrum, and Fivetran, increasing cluster performance. availability, and analytic insight.

- In less than a month, singlehandedly designed, prototyped, and deployed a spam classifier capable of greater than 98% accuracy and close to a 0% false positive rate, ameliorating one of the worst user experiences on the site.

- Developed a timeseries algorithm for predicting the likelihood that a Kickstarter project will succeed or fail and how much it will raise (>90% accuracy one day after launch). By constructing the model as an HTTP service, its output has been incorporated in domains as diverse as revenue forecasting, search ranking, customer support, and social media marketing.

- Provide daily support, guidance, and mentorship to three Data Analysts.

Vox Media (Chief Data Scientist, 2016)

- Responsible for the design, development, and ongoing maintenance of a centralized Hive warehouse for all of the company's key data sources, including Google Analytics Premium (BigQuery), Chartbeat, DoubleClick for

Publishers (DFP), Chorus (our proprietary content management system), Facebook Insights, Twitter Analytics, Youtube Analytics, Snapchat Analytics, and several others.

- Wrote Hive queries and custom Map-Reduce jobs to join and analyze these data sources. These analyses were presented directly to the CEO and were fundamental in steering company-wide revenue strategy and identifying opportunities for future growth.

Enigma (Chief Data Scientist, 2013-2015)

- Designed and developed a global database of government procurement contracts within ten non-english speaking countries for a Fortune 100 company. This involved extracting structured data from messy websites, PDFs, Word Documents, PowerPoints, and poorly-formatted Excel spreadsheets. To accomplish this task, I hired and trained a team of four people, designed custom, reusable software to expedite the data acquisition process, and interfaced directly with the clients. The successful delivery of this application was fundamental in securing a multi-year, multi-million dollar contract which increased Enigma's revenue ten-fold and aided in raising $27 million in venture capital.

- Worked with the City of New Orleans to develop machine learning-based tool to assist in addressing the city's backlog of blight abatement cases. Within three months of the tool's implementation, the backlog of cases had been fully eliminated. This project recieved an award from the Bureau of Governmental Research.

- In collaboration with the Red Cross, DataKind, and local fire departments, I developed a novel machine learning model capable of predicting the likelihood that residents in a given census block lacked a smoke alarm. This project was powered by multiple open-source libraries and resulted in a free web application which enabled local governments to upload their own fire incident data and return a list of addresses to target for inspections. The tool is actively used by multiple fire departments throughout the country, as well as by the Red Cross to optimize their Home Fire Preparedness Outreach Campaign.

- Analyzed and visualized 50 years of daily temperature readings from NOAA to identify daily temperature anomalies. The resulting map and analysis helped communicate how climate change results in not only in warmer weather, but more *anomalous* weather. This visualization generated significant traffic to Enigma's corporate website, received multiple awards, was presented to the UN's Intergovernmental Panel on Climate Change, was tweeted by Bill Gates, and was also deemed the "best visualization of climate change" by Jer Thorpe (a prominent data artist).

NewsLynx (Co-founder and CTO, 2013-2015)

- Conceived of, secured funding for, and led development of a platform for measuring the quantitative and qualitative impact of nonprofit journalism. The project resulted in a white paper and numerous open source libraries.

DataKind (DataCorps Ambassador, 2013-2014)

- Worked with GiveDirectly, a non-profit which addresses extreme poverty via direct cash transfers, to create an algorithm that uses satellite images to help staff more efficiently identify villages in need. By optimizing the organization's outreach efforts, the algorithm was estimated to save almost 400 hours of volunteer time. The project was published in *Big Data* and was awarded "Best Social Good Paper" at KDD, the world's leading Data Mining conference.

New York Times / Knight-Mozilla OpenNews (Data Scientist / OpenNews Fellow, 2013)

- Working with multiple terabytes of raw event logs, as well as novel data sources, I developed Pageviews Above Replacement – a metric for better assessing the performance of online content. In addition to this principal project, I assisted in internal analytics efforts and collaborated with the Computer Assisted Reporting team on data journalism projects.

Treasury.IO (Technical Lead, 2013)

- With a group of other journalists, developers and data scientists, I built and open-sourced a first-of-its-kind data feed for how the U.S. government spends and takes in money on a daily basis. The project has been used in stories by Al Jazeera America, Time and others. The project was funded through a Knight-Mozilla Code OpenNews Sprint grant.

Harmony Institute (Data Scientist, 2012)

- Utilized data from social media, the web, and EEG scans to develop new methodologies for measuring the impact of narrative media. This work resulted in a peer-reviewed publication in *Nature Communications*.

Columbia University (Research Assistant, 2011)

- Conducted literature reviews, translated primary interviews in French, and provided detailed edits of Dr. Severine Autesserre's book *Peaceland*, published by Cambridge University Press.

Harry Frank Guggenheim Foundation (Assistant Program Officer, 2009-2011)

- Reviewed over 400 applications for annual Research Grants and Dissertation Fellowships, providing detailed assessments of each project's academic merit.

Grassroots Campaigns (Assistant Office Manager, 2004)

- Recruited, trained, and managed a team of 10+ door-to-door canvassers, raising money for the Kerry Campaign.

Education

Columbia University, New York, NY

- Masters in Applied Statistics, 2011-2012
- Master's Thesis: "The Neural and Emotional Correlates of Social Media."
- Coursework in Data Mining, Timeseries Analysis, Bayesian Statistics, Survey Design, Python, and GIS.

Whitman College, Walla Walla, WA

- Bachelors in Politics with Honors, 2005-2008
- Honors Thesis: "Another end of the world is possible: The apocalyptic narratives of American Imperialism"

Awards / Grants

- Excellence in Innovation, Bureau of Governmental Research, 2016
- Best Data Visualization Projects, Flowing Data, 2014
- Short List - Interactive Visualization, Information is Beautiful Awards, 2014
- Best Social Good Paper, KDD, 2014
- Fellow, Tow Center for Digital Journalism, Columbia University, 2013-2015
- Fellow, Knight-Mozilla OpenNews / New York Times, 2013-2014
- Code Sprint Grant, Knight-Mozilla OpenNews, 2013
- Best in Show, New York Times Open Source Science Fair, 2013
- Best in Innovation and Best in Show, Bicoastal Datafest - Columbia / Stanford University, 2013

Publications {#publications)

Keller, Michael and Abelson, Brian (2015). Newslynx: A tool for newsroom impact measurement. *Tow Center for Digital Journalism, Columbia University*.

Varshney, Kush R., Abelson, Brian, *et. al.* (2014). Targeting villages for rural development using satellite image analysis. *Big Data* 3:1.

Dmochowski, J.P., Abelson, Brian, *et. al.* (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications* 5:4567.