

ANDREW G. REECE

+1.617.395.6841

andrew.reece@gmail.com

andrewgarrettreece.com

[github:andrewreece](https://github.com/andrewreece)

[stackoverflow:andrew_reece](https://stackoverflow.com/users/andrew_reece)

OVERVIEW

- Full-stack data scientist: **6 years experience** in analytics, research, and product development
- Skill set combines **statistical analysis**, **software engineering**, and **scientific research**
- Proven experience in both **client-facing** and **engineering roles** on solo and team projects
- Skill highlights: **Machine learning** (Python, R), **big data analytics** (Spark), **front-end** (JS, D3)

TECH SKILLS

- **Strong** Python, R, JavaScript (D3, jQuery), AWS (EC2, EMR, S3)
- **Proficient** Golang, Docker, Spark, Hive, Kafka, Flask, SQL, PHP, Linux
- **Basic** Cython, OpenCL, Perl, Matlab, Angular

EXPERIENCE

- | | | |
|-------------|--|--|
| 2017 | MEMBER OF TECHNICAL STAFF | StackRox, Inc |
| | <ul style="list-style-type: none">• Led project to develop automated analysis of machine learning performance and threat detection efficacy• Contributed to engineering efforts across the full stack: back end, UI/UX, API development, machine learning• Helped launch company's first enterprise-ready product | |
| 2016 | DATA SCIENCE & STRATEGY CONSULTANT | Shortlist, LLC |
| | <ul style="list-style-type: none">• Created analytics pipeline for early-stage talent acquisition startup• Defined strategic initiatives for near-term data/analytics goals | |
| 2015 - 2016 | DATA SCIENCE CONSULTANT | DrivenData, Inc |
| | <ul style="list-style-type: none">• Generated predictive modeling solutions for multi-sector client base• Lead data science trainings for clients | |
| 2015 - 2016 | CYBER SECURITY DATA SCIENCE INTERN | Rapid7, Inc. |
| | <ul style="list-style-type: none">• Applies machine learning methods to identify email phishing attacks• Generated 3 patent applications in first 3 months on the job• Lead project placed on Q4 corporate roadmap for product integration | |
| 2014 | DATA SCIENCE FELLOW | Chicago Department of Public Health |
| | <ul style="list-style-type: none">• Created predictive model to reduce residential lead-poisoning hazards• Helped City of Chicago secure \$3.9M federal grant from US Dept. of Housing & Urban Development for lead prevention data science project• Published findings in peer-reviewed 2015 KDD Conference Proceedings | |

DATA SCIENCE PROJECTS

- GaugingDebate** **Real-time sentiment tracking of 2016 US Presidential debates**
- End-to-end analytics pipeline: **streaming Twitter data**, parallelized **sentiment analysis**, dynamic front-end web app
 - Tech: **Spark**, Kafka, Flask, Plotly, Twitter API, AWS
- HelloMarc.us** **Hollywood box office prediction with Natural Language Processing**
- **Latent Dirichlet Allocation** (collapsed Gibbs) on 40k+ movie reviews provided features for machine learning algorithm (Random Forests) to predict box office success, **coded algorithm from scratch**
 - Tech: **Python**, Pandas, Scikit-Learn, Flask, Bootstrap, D3
- SEE:Net Graph** **Social network graph with continuous-tracking sensor data**
- **Graph analysis** and visualization web app used MIT Reality Mining experiment dataset; demonstrated **emergence of social networks** with interactive time series dashboard
 - Tech: **D3**, jQuery, R, Python, jQuery

AWARDS & RECOGNITION

2016 Kaggle Master

Ranked in **Top 1% of all 475,000+ worldwide** Kaggle.com data science competitors

2014 Eric & Wendy Schmidt Data Science for Social Good Summer Fellowship

Awarded for excellence in data science and **demonstrated commitment to social causes**

PATENTS

- **Identifying Malicious Identifiers** Pending
Algorithm for identifying malicious links in emails PCT: 15/177,555
- **Classifying Locator Generation Kits** Pending
URL parsing engine for discovering malware attack campaigns PCT: 15/200,530
- **Neutralizing Malicious Locators** Pending
Cyber counter-attack for triggering shutdown of malicious websites PCT: 15/196,072

HARVARD COURSEWORK

- **Machine Learning**: Mathematical foundations in Bayesian modeling and optimization
- **Monte-Carlo Methods & Stochastic Optimization**: Range of MCMC and optima methods
- **Data Science**: Full-stack predictive analytics (scraping, munging, modeling, visualization)
- **Data Visualization**: Design principles and interactive visualization with D3
- **Computing Foundations for Computational Science**: Parallel, GPU optimized computing
- **Learning From Big Data**: Team Kaggle competitions in big data, machine learning
- **Introduction to Probability**: Random variables and distributions, basic combinatorics

EDUCATION

2017 **PhD** **Psychology | Computational Science** **Harvard University**

Dissertation: "A method for identifying predictive markers of mental illness in social media data"
GPA: 3.89

