

Hilary Parker

77 Van Ness Ave
Ste 101 #1722
San Francisco, CA 94102
331-472-7537
me@hilaryparker.com

Employment

Data Science Product Manager -- Biden for President (remote)

AUGUST 2020 - NOVEMBER 2020

Managed a complex deployment cycle with multiple stakeholders and QA checkpoints for all the voter-level scores produced by the central Data Science team (e.g. support probability, turnout probability used at national and state levels by all analytics teams). Successfully completed the cycle with zero incidents of score rollbacks. Advised on statistical modeling choices and provided code review for voter-level modeling as well as overall polling.

Designed and implemented voter-survey research to estimate the number of “mailed-by-deadline-but-not-yet-received” ballots in key swing states such as NV and PA. Provided estimates on election night (and beyond) which were used by the Chief Analytics Officer to shape public messaging about the election outcome.

Data Scientist -- Stitch Fix (San Francisco, CA)

MAY 2016 - AUGUST 2020

Drove company-wide initiative to use outfits as a core client-facing algorithmic product by establishing foundational system framework, building data-collection platform to power multiple algorithms, and partnering with product, creative, strategy, and merchandising to leverage creative IP in the form of data, QA algorithmic output and prototype client-facing experiences. ([Wired](#))

Provided decision and strategic support for the evolution of the Stitch Fix experience into an interactive client-facing experience, such as the introduction of client-facing quizzes, and addressing technical debt associated with over-reliance on direct client answers to specific onboarding questions.

Senior Data Analyst -- Etsy, Inc. (Brooklyn, NY)

JUNE 2013 - MARCH 2016

Developed statistical methodology (hypothesis testing, generalized linear modeling, survival analysis, multiple-testing correction) and process for the design and analysis of online experiments. Influenced buyer-facing product development with experimentation, exploratory data analysis, and opportunity sizing with deliverables ranging from ad-hoc analyses to internal white papers

Teaching Experience

Visiting Lecturer -- American University of Armenia (Yerevan, Armenia)

NOVEMBER 2011

Taught "Statistical Modeling and Sampling" in an intensive 3-week session as part of the core curriculum for the Masters of Public Health Program.

Lead Teaching Assistant -- Johns Hopkins School of Public Health (Baltimore, MD)

FALL 2012

Taught lab sessions for the 500+ person "Introduction to Biostatistics" course as part of core curriculum for the Masters of Public Health Program.

Education

Ph.D., Biostatistics -- Johns Hopkins Bloomberg School of Public Health (Baltimore, MD)

AUGUST 2008 - MAY 2013

Thesis: Practical Statistical Issues in Translational Genomics

M.S., Biostatistics -- Johns Hopkins Bloomberg School of Public Health (Baltimore, MD)

AUGUST 2008 - MAY 2011

B.A., Mathematics; B.A., Molecular Biology -- Pomona College (Claremont, CA)

AUGUST 2004 - MAY 2008

Honors and Awards

Hopkins Sommer Scholar (Johns Hopkins Bloomberg School of Public Health)

AUGUST 2008 - MAY 2013

Helen Abbey Award for Excellence in Teaching (Biostatistics Department, Johns Hopkins Bloomberg School of Public Health)

SPRING 2012

Gertrude M. Cox Scholarship, Honorable Mention (American Statistical Association)

FALL 2008

Bruce J. Levy Prize in Mathematics (Pomona College)

SPRING 2007

Howard Hughes Medical Institute Undergraduate Summer Research Grant (Pomona College)

SUMMER 2007

Publications

Refereed

Jeffrey T. Leek, W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, John D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics*, Volume 28, Issue 6, 15 March 2012, Pages 882–883, <https://doi.org/10.1093/bioinformatics/bts034>

Hilary S. Parker, Jeffrey T. Leek, Alexander V. Favorov, Michael Considine, Xiaoxin Xia, Sameer Chavan, Christine H. Chung, Elana J. Fertig, Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction, *Bioinformatics*, Volume 30, Issue 19, October 2014, Pages 2757–2763, <https://doi.org/10.1093/bioinformatics/btu375>

Parker HS, Corrada Bravo H, Leek JT. 2014. Removing batch effects for prediction problems with frozen surrogate variable analysis. *PeerJ* 2:e561 <https://doi.org/10.7717/peerj.561>
Leek, Jeffrey T., W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey. "sva: Surrogate Variable Analysis R package version 3.10. 0." DOI 10 (2014): B9.

Parker, Hilary S, and Jeffrey T Leek. "The practical effect of batch on genomic prediction." *Statistical applications in genetics and molecular biology* vol. 11,3 (2012): Article 10. doi:10.1515/1544-6115.1766

Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Elana J. Fertig, Andrew E. Jaffe, John D. Storey, Yuqing Zhang, and Leonardo Collado Torres. "sva: Surrogate variable analysis, 2018." R package version 3, no. 0: 25-27.

Considine, M., Parker, H., Wei, Y., Xia, X., Cope, L., Ochs, M., & Fertig, E. (2015). AGA: Interactive pipeline for reproducible genomics analyses. *F1000Research*, 4, 28. <https://doi.org/10.12688/f1000research.6030.1>

Not Refereed

Parker H. Opinionated analysis development. *Peer J Preprints*. 2017;5(e3210v):1.

Public Presentations

*denotes keynote or plenary talk

2020

***Electronic Conference on Teaching Statistics (eCOTS) 2020** (remote): [Using Data Effectively: Beyond Art & Science](#)

***rstudio::conf(2020L)** (San Francisco, CA): [Not So Standard Deviations Episode 100](#)

2019

Directions in Statistical Computing 2019 (Palo Alto, CA): Enabling Data Scientists at Stitch Fix

***APRA Data Analytics Symposium 2019** (Phoenix, AZ): [Using Data Effectively: Beyond Art and Science](#)

***Plant Health 2019** (Cleveland, OH): [A Design Mindset for Working with Data](#)

***Connect@IPSDS Conference** (Mannheim, Germany): [Design Thinking for Data Science](#)

***Women in Analytics Conference** (Menlo Park, CA): [Design Thinking for Data Science](#)

***Women in Data Science Conference** (Palo Alto, CA): [Using Data Effectively: Beyond Art and Science](#)

rstudio::conf(2019L) (Austin, TX): [Cultivating creativity in data work](#)

2018

***QCon** (San Francisco, CA): [Using Data Effectively: Beyond Art and Science](#)

***Electronic Undergraduate Statistics Research Conference (eUSR)** (remote): [Cultivating Creativity in Data Work](#)

***Big Data Ignite** (Grand Rapids, MI): Cultivating Creativity in Data Work

***Fields Institute Data Science in Industry Day** (Toronto, Canada): [Cultivating Creativity in Data Work](#)

***International Conference on Teaching Statistics (ICOTS) 2018** (Kyoto, Japan): [Cultivating Creativity in Data Work](#)

***Rose Hulman Undergraduate Math Conference** (Terra Haute, IN): [The Statistics Behind a Personalized Look](#)

Women in Analytics Conference (Menlo Park, CA): [Cultivating Creativity in Data Work](#)

Data Day Texas (Austin, TX): Opinionated Analysis Development

2017

***EARL London 2017** (London, England): Opinionated Analysis Development

Pomona College (Claremont, CA): visiting panelist

JupyterCon 2017 (New York, NY): [Opinionated Analysis Development](#)

***EARL San Francisco 2017** (San Francisco, CA): Opinionated Analysis Development

csv,conf,3 (Portland, OR): [Opinionated Analysis Development](#)

McGill University (Montreal, Canada): visiting speaker

rstudio::conf(2017L) (Orlando, FL): [Opinionated Analysis Development](#)

2016

***Joy of Coding Conference** (Rotterdam, Netherlands): [The Joy of Analysis Development](#)

NYR Conference (New York, NY): Scaling Analysis Responsibly

Joint Statistical Meetings (Chicago, IL): [Data Science at Stitch Fix \(poster\)](#)

2015

NYR Conference (New York, NY): [Reproducible Analysis in Production: Lessons from Etsy](#)

Macalester College (Saint Paul, MN): Mathematics & Computer Science Senior Capstone speaker

2014

Trinity College (Hartford, CT): visiting speaker

***True North Conference** (Stockholm, Sweden): [speaker](#)

Other Professional Activities

Co-host of [Not So Standard Deviations](#), a data science podcast (10k+ listens per episode)
2015 - PRESENT