F. William (Will) Townes — Updated March 11, 2019

CONTACT INFORMATION

Department of Biostatistics
Harvard University
Cambridge, MA USA

website: willtownes.github.io
email: ftownes@g.harvard.edu

SUMMARY

I am a fifth-year Biostatistics Ph.D. candidate specializing in high-dimensional, unsupervised data analysis. My interests include machine learning, genomics, and infectious disease. My goal is to produce statistical methods and open source software implementations to help scientists achieve their research goals more effectively.

EDUCATION

2019[1] **Ph.D. Biostatistics**, Harvard University, Cambridge, MA USA
Doctoral Committee: Rafael Irizarry, Ph.D. (Dana-Farber Cancer Institute, Biostatistics)
Jeff Miller, Ph.D. (Harvard University, Biostatistics)
Martin Aryee, Ph.D. (Massachusetts General Hospital, Pathology)

2016 **A.M. Biostatistics**, Harvard University, Cambridge, MA USA
2013 **M.S. Mathematics & Statistics**, Georgetown University, Washington, DC USA
2007 **B.S. Biology** *magna cum laude*, Washington & Lee University, Lexington, VA USA

PROFESSIONAL EXPERIENCE

2011 - 2014 **Software Test Engineer**, SRA International, Arlington, VA
Developed automated testing of web services and ETL processes for the Energy Star Qualified Product Exchange. Trained other testers on the use of APIs and version control.

2008 - 2010 **Software Analyst**, Perrin Quarles Associates, Charlottesville, VA
Supported development of the Environmental Protection Agency's National Emissions Inventory System.

2007 - 2008 **Fulbright Scholar**, University of the Philippines
Conducted independent forest ecology research while living with an indigenous *Aeta* community in Bataan Province.

RESEARCH

Preprints

[1] **Townes FW**, Hicks SC, Aryee MJ, Irizarry RA. Feature Selection and Dimension Reduction for Single Cell RNA-Seq based on a Multinomial Model. *bioRxiv* (2019). doi: https://doi.org/10.1101/574574.

In Press

[1] Marsh DM, **Townes FW**, Cotter K, Farroni K, McCreary K, Petry R, Tilghman J. Temperature Tolerance and Species Range in Mountaintop Salamanders and Their Widespread Competitors. *Journal of Herpetology* (2019).

Peer-reviewed Journal Articles

[1] Hicks SC, **Townes FW**, Teng M, Irizarry RA. Missing Data and Technical Variability in Single-Cell RNA-Sequencing Experiments. *Biostatistics* (2018). doi: http://dx.doi.org/10.1101/025528.

[2] Hecker, J, Xu X, **Townes FW**, Fier HL, Corcoran C, Laird N, Lange C. Family-Based Tests for Associating Haplotypes With General Phenotype Data: Application to Asthma Genetics. *Genetic Epidemiology* (2017). doi: https://doi.org/10.1002/gepi.22094.

[1] anticipated

[3] Valeri L, Patterson-Lomba O, Gurmu Y, Ablorh A, Bobb J, **Townes FW**, Harling G. Predicting Subnational Ebola Virus Disease Epidemic Dynamics from Sociodemographic Indicators. PLoS ONE (2016). doi: https://doi.org/10.1371/journal.pone.0163544.

[4] **Townes, W**. Seed dispersal of the genus Leea in forest patches of Bataan, Philippines. *Ecotropica* (2010).

Dissertation Topics In Preparation

[1] **Townes FW**, Hicks SC, Aryee MJ, Irizarry RA. Feature Selection, Normalization, and Dimension Reduction of High Dimensional, Sparse Count Matrices with Application to Single Cell Gene Expression.

[2] **Townes FW**, Miller J. Identifying Pro-Longevity Genes by Integrating Gene Expression and Curated Annotations.

[3] **Townes FW**. The Compound Poisson Extreme Stable Family and a Generalization of the Hermite Distribution.

Class Projects and Other Unpublished Work

[1] Townes FW, Comment L. Bayesian Methods for Dependent Data. Harvard Biostatistics 245 (Multivariate and Longitudinal Analysis). April 2016.

[2] Townes FW. Variational Inference for Mixtures of Dirichlet Network Distributions. Harvard Computer Science 282R (Bayesian Nonparametrics Seminar).

[3] Townes FW, Liu JZ. Bayesian Nonparametric Time Series Modeling. MIT 6.882 (Bayesian Modeling and Inference).

[4] Townes FW, Marquez-Luna C, Onnela JP. Network Connectivity and Infectious Disease Modeling.

[5] **Townes FW**, Hicks SC, Aryee MJ, Irizarry RA. Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. *bioRxiv* (2017). doi: https://doi.org/10.1101/166736.

Talks

Seminars

[1] Townes FW. Feature Selection and Dimension Reduction for Single Cell RNA-Seq. *Englehardt Lab, Computer Science Department, Princeton University.* February 2019.

[2] Townes FW. Feature Selection and Dimension Reduction for Single Cell RNA-Seq. *Greene Lab, Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania.* February 2019.

[3] Townes FW. Unsupervised Learning for Single Cell Gene Expression. *Harvard Biostatistics Cancer Working Group.* September 2018.

[4] Townes FW. Single Cell Housekeeping Genes and Normalization. *Dana-Farber Cancer Institute cBio Seminar.* April 2018.

[5] Townes FW. Informative Missing Data in Single Cell RNA-Seq. *Dana-Farber Cancer Institute Genomics Seminar.* October 2016.

[6] Townes FW, Lange C. Family Based Association Tests for Rare Variants. *Brigham & Women's Hospital Channing Network Medicine Seminar.* September 2015.

Contributed Talks

[1] Townes FW. Varying-Censoring Aware Matrix Factorization for Single Cell RNA-Sequencing. *Bioconductor Annual Meeting.* July 2017.

Posters

[1] Townes FW, Shukla C. Gene Expression Autoencoders. Harvard Biomedical Informatics 707 (Deep Learning in Healthcare). April 2018.

[2] Townes FW, Marquez-Luna C. Mixture of Experts Analysis of Infectious Disease Outbreak Characteristics. Harvard Computer Science 281 (Advanced Machine Learning). December 2015.

[3] Townes FW, Karaayvaz M, Gillespie S, Bernstein B, Ellisen L, Aryee M. Single Cell RNA-Seq Technical and Biological Confounders. Program in Quantitative Genomics Conference, Harvard Medical School. November 2015.

Funding

2018 - 2019	NIH T32 Training Grant: Cancer
2017	Chan-Zuckerberg Foundation Travel Grant
	Human Cell Atlas Jamboree. European Bioinformatics Institute, Hinxton, UK
2016 - 2018	NIH T32 Training Grant: Big Data to Knowledge
2016	NSF-CBMS Travel Award: Topology, Geometry, and Statistics
2014 - 2016	NIH T32 Training Grant: Clinical Epidemiology of Lung Diseases

Teaching Experience

2018 Fall	**Teaching Assistant**, Applied Regression (Harvard - BST 210)
	Grade homeworks, hold office hours, and teach a weekly lab section.
2018 Summer	**Co-Instructor**, StatStart (Harvard)
	Statistics program for under-represented high school students.
	Gave two interactive lectures on graphing data and regression.
2017 Fall	**Teaching Assistant**, Applied Bayesian Analysis (Harvard - BST 228)
	Grade homeworks, provide solutions, and hold office hours.
2017 Summer	**Lead Instructor**, Introduction to Data Science (PARSE Ltd.)
	Nonprofit, week-long statistics program for high school students.
	Developed course material and gave interactive lectures using R.
2017 Spring	**Teaching Assistant**, Applied Longitudinal Analysis (Harvard - BST 226)
	Grade homeworks and hold office hours.
2016 Fall	**Teaching Assistant**, Intro to Statistical Methods (Harvard - BST 201)
	Grade homeworks, provide solutions, hold office hours, and teach a weekly lab section.
2016 Summer	**Project Mentor**, Pipelines into Biostatistics (Harvard)
	Biostatistics program for undergraduates.
	Guided three students on a pharmacogenomics data analysis.
2016 Spring	**Teaching Assistant**, Rates and Proportions (Harvard - BST 210)
	Grade homeworks, provide solutions, hold office hours, and teach a weekly lab section.

Technical Skills

Programming: R, Python, SQL, Git, LaTeX, Stata, Julia, Matlab, Shell
Operating Systems: Mac OS X, Linux, Windows
Machine Learning Frameworks: Caret, Sklearn, Keras, Tensorflow, Stan, JAGS