

July 14, 2025

## Kenneth E. Shirley

website: <http://www.kennyshirley.com>

email: [kennetheshirley@yahoo.com](mailto:kennetheshirley@yahoo.com)

---

### EDUCATION

2007 The Wharton School, University of Pennsylvania, Ph.D., Statistics (advisor Dylan Small)

2003 Harvard College, A.B. with honors, Mathematics.

### POSITIONS

2020- Principal Research Scientist, Amazon

2016-2020 Senior Research Scientist, Amazon

2013-2016 Principal Member, Technical Staff - Research, AT&T Labs, Inc.

2009-2013 Senior Member, Technical Staff - Research, AT&T Labs, Inc.

2007-2009 Postdoctoral Fellow, Earth Institute, Columbia University (advisor Andrew Gelman)

### PUBLICATIONS

2016 Flynn, C., Shirley, K., and Wang, W. "Deconstructing Domain Names to Reveal Latent Topics", Proc. IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, Canada. (\*Honorable Mention Best Paper Award)

2016 Foster, D., Karloff, H., and Shirley, K. "How well does the standard body mass index or variations with a different exponent predict human lifespan?", *Obesity*, in press.

2015 Shirley, K. and Wang, W. "Breaking Bad: Detecting malicious domains using word segmentation", *IEEE 2015 workshop on Web 2.0 Security and Privacy (W2SP)*, San Jose, CA.

2015 Shirley, K., and Gelman, A., "Hierarchical Models for Estimating State and Demographic Trends in U.S. Death Penalty Public Opinion", *Journal of the Royal Statistical Society, Series A. (JRSS-A)*, Vol. 178, No. 1, pp.1-28.

2014 Sievert, C., and Shirley, K. "LDAvis: A method for visualizing and interpreting topics", *Associations for Computational Linguistics (ACL) 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, MD.

2013 Reibman, A., Shirley, K., and Tian, C., "A Probabilistic Pairwise Preference Predictor for Image Quality", *Proc. of IEEE Intl. Conference on Image Processing (ICIP)*, Melbourne, Australia, Sept. 15-18, 2013. pp. 413 - 417.

- 2013 Karloff, H., Shirley, K. "Maximum Entropy Summary Trees", *Computer Graphics Forum (Proc. EuroVis)*, Volume 32, Issue 3, Part 1, pp. 71-80. (\*Honorable Mention Best Paper Award)
- 2010 Shirley, K., Small, D., Lynch, K., Maisto, S., Oslin, D., "Hidden Markov Models for Alcoholism Treatment Trial Data", *Annals of Applied Statistics*, Vol. 4, No. 1, 366-395.
- 2009 Jensen, S.T., Shirley, K., Wyner, A., "Bayesball: A Bayesian Hierarchical Model for Evaluating Fielding in Major League Baseball", *Annals of Applied Statistics*, Vol. 3, No. 2, 491-520.

## BOOK CHAPTERS

- 2012 Osgood, D. and Shirley, K., "The Value of Information in Index Insurance for Farmers in Africa". In Ramanan Laxminarayan and Molly K. Macauley (eds.), "The Value of Information: Methodological Frontiers and New Applications in Environment and Health", pp. 1-18, Springer, New York, USA.
- 2011 Gelman, A., and Shirley, K., "Inference and Monitoring Convergence". In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng (eds.), "Handbook of Markov Chain Monte Carlo", pp. 163-174, Chapman & Hall/CRC, New York, USA.

## SOFTWARE

- 2015 **summarytrees** R package: <https://github.com/kshirley/summarytrees> (with H. Karloff)  
A method for summarizing (via maximum entropy) and visualizing large, node-weighted trees. Components written in R, C and d3.js.
- 2014 **LDAvis** R package: <https://github.com/cpsievert/LDAvis> (with C. Sievert)  
A d3-based interactive visualization to help users interpret the topics learned from Latent Dirichlet Allocation. As of January 2016, the repository had about 170 stars on Github.

## PATENTS

- 2021 "Data Analytics System and Methods for Text Data", with P. Bogdan, G. Gressel, G. Reser, A. Rubarkh, USPTO Patent #20210232764.
- 2019 "Method and apparatus for generating quality estimators", with A. Reibman and C. Tian, USPTO Patent #10432985.

## TALKS

- 2022 Aug "Using Subsampling to Speed up Training in Attention-Based NNs: A Case Study at Doing Statistics at Scale in the Amazon Supply Chain", JSM, Washington, DC
- 2017 Aug "Discussion: The Essential Power of Interactive Graphics", JSM, Baltimore,

	MD
2015 Aug	"Interactive Visualization of Trees Using R and D3", JSM, Seattle, WA
2015 June	"Text Mining on Domain Names", Frontiers in Applied and Computational Mathematics (FACM), New Jersey Institute of Technology, Newark, NJ
2014 Sep	Panelist, Data Science Career Panel, Columbia University Postdoc Research and Career Symposium, New York, NY
2013 Nov	"Maximum Entropy Summary Trees", Columbia University Statistics Department Seminar, New York, NY
2013 Aug	"Maximum Entropy Summary Trees", JSM, Montreal, QC.
2013 June	"Maximum Entropy Summary Trees", EuroVis, Leipzig, Germany.
2013 Feb	"Maximum Entropy Summary Trees", University of Kentucky Statistics Department Seminar, Lexington, KY
2011 Oct	"Bayesian Models for Weekly Customer Volume at AT&T Stores", Case Studies in Bayesian Statistics and Machine Learning, CMU, Pittsburgh, PA.
2011 Aug	"The ABCs of Xqjkz: A New Scrabble Rating System Based on a Statistical Model for Tile-by-Tile Play", JSM, Miami, FL
2010 Aug	"Hierarchical Bayes models for rainfall at multiple sites", JSM, Vancouver, BC
2010 Mar	"Hierarchical Bayes time series models for death penalty public opinion data", Virginia Tech Statistics Department, Blacksburg, VA
2009 Aug	"Index Insurance and Statistics", JSM, Washington, D.C.
2009 Feb	"Hierarchical Bayes time series models for death penalty public opinion data", QuaSSI, Department of Political Science, Penn State University, State College, PA
2008 Oct	"Modeling and Simulating Rainfall", IRI Workshop on Technical Issues in Index Insurance, Columbia University, New York, NY
2008 Aug	"Modeling time series of death penalty public opinion", JSM, Denver, CO
2008 May	"Bayesian hidden Markov models for alcoholism treatment trial data", Statistics in Psychiatry Symposium, New York, NY
2007 Aug	"A Bayesian hidden Markov model for alcoholism treatment trial data", JSM, Salt Lake City, UT
2007 Ma	"Hidden Markov models for alcoholism treatment trial data", ENAR Annual Meeting, Atlanta, GA
2006 Aug	"Hidden Markov models for ordinal longitudinal data", JSM, Seattle, WA
2006 Apr	"The effects of missing data on Value-Added Model (VAM) estimates", AERA Annual Meeting, San Francisco, CA (Joint with Henry Braun).

## AWARDS

- 2016 Honorable Mention, Best Paper Award, DSAA 2016, for "Deconstructing Domain Names to Reveal Latent Topics"
- 2014 AT&T Labs Excellence Team Award (LDA Topic Model Web-based Software Tool)
- 2013 Honorable Mention, Best Paper Award, EuroVis 2013, for "Maximum Entropy Summary Trees"
- 2012 AT&T Labs Vice President's Excellence Award (Recruiting Committee Leadership)
- 2011 AT&T Labs Vice President's Excellence Award (Forecasting model for store visits)
- 2007 Statistics in Epidemiology (Section of the ASA) Travel Award for JSM, Salt Lake City,

UT.

2004 Murray Award for Excellence in Teaching, Department of Statistics, Wharton School

## **MENTORSHIP**

2024 Amazon Summer Intern, Abhinandan Dalal, University of Pennsylvania

2017 Amazon Summer Intern, Xiang Xu, University of Houston

2014 AT&T Labs Research Summer Intern, Jingjing Zou, Columbia University

2012 AT&T Labs Research Summer Intern, Carson Sievert, Iowa State University

2011 AT&T Labs Research Summer Intern, Jie Shen, UC-Irvine

## **SERVICE**

2017 Program Chair, ASA Section on Statistical Graphics (for JSM 2017)

2014-2018 Associate Editor, Journal of Quantitative Analysis in Sports (JQAS)

2015 Organizing Committee, AT&T Labs Machine Learning Summit, New York, NY

2013-2014 Secretary/Treasurer, ASA Section on Statistical Graphics

2010-2011 Co-coordinator, AT&T Labs Research Summer Intern Program

Referee: Journal of the American Statistical Association (JASA)

Bayesian Analysis

The American Statistician

Statistics in Medicine (SIM)

MIT Sloan Sports Analytics Conference

Journal of Quantitative Analysis in Sports (JQAS)

Psychometrika

Journal of Statistical Computation and Simulation (JSCS)

ASA Section on Statistical Learning and Data Mining (SLDM) Student Paper  
Competition