

Michael J. Kane

CONTACT INFORMATION	22 Mill Pond Drive Guilford, CT 06437	<i>Phone:</i> (585) 613-1640 <i>E-mail:</i> michael.kane@yale.edu
ACADEMIC APPOINTMENTS	Assistant Professor, 7/2016 – Present Associate Research Scientist, 9/2010 – 7/2015 Biostatistics Department Yale University New Haven, CT, USA	
RESEARCH INTERESTS	Scalable Statistical/Machine Learning Methodology, Early Phase Clinical Trials	
EDUCATION	Yale University , New Haven, CT Ph.D., Statistics 2010 • Dissertation Title: Scalable Strategies for Computing with Massive Sets of Data • Dissertation Committee: John Emerson, Associate Professor of Statistics (Primary Advisor) David Pollard, Professor of Statistics Martin H Schultz, Arthur K. Watson Professor of Computer Science Emeritus M.A., Statistics 2006 Rochester Institute of Technology , Rochester, New York M.S., Electrical Engineering 2003 • Concentration: Signal and Image Processing • Thesis Title: A Two Step Approach to Bayesian Network Model Selection • Advisor: Professor Andreas Savakis B.S., Computer Engineering, 2000	
AWARDS AND HONORS	Grants In Execution The Patient-Centered Outcomes Research Institute , 7/2016 – 12/2019. Study Registration Information: HSRP20164106. Expansion of Methods for Two-Stage Trial Designs for Testing Treatment, Self-Selection and Treatment Preference Effects. Position: Co-Investigator. The National Institutes of Health , 08/2013 – 08/2018. Award: P30 CA 016359-33S3. Yale Comprehensive Cancer Center Grant. Position: Biostatistician. Grants Completed The Bill and Melinda Gates Foundation , Grand Challenge in Global Health. 11/2013–4/2015. <i>Semantic clustering of human-animal medical corpuses</i> . Position: PI. Defense Advanced Research Projects , XDATA Initiative. 10/2012–12/2016. <i>ECON: Elastic Computational Numerics</i> . Award Number: FA8750-12-C-0324. Position: Co-PI.	

Academic Awards

ASA's Chambers Award for the Bigmemory Project Software, 2010

Yale University Dissertation Fellowship, 2009-2010

Yale University Teaching Fellowship, 2007-2009

Yale University Fellowship, 2005-2009

Fred Emerson Foundation Scholarship, 1997-2000

REFEREED JOURNAL ARTICLES

T. Arnold, **M. J. Kane**, S. Urbanek "iotools: High-Performance I/O Tools for R," *The R Journal*, 2017.

M. J. Kane, B. Lewis, S. Tatikonda, and S. Urbanek "Scatter Matrix Concordance: A Diagnostic for Regressions on Subsets of Data," *Statistical Analy Data Mining*, 2015.

F. Costa, J. E. Hagan, J. Calcagno, **M. Kane**, P. Torgerson, M. S. Martinez-Silveira MS, et al. "Global Morbidity and Mortality of Leptospirosis: A Systematic Review," *PLoS Negl Trop Dis.*, 2015.

P. R. Torgerson, J. E. Hagan, F. Costa, J. Calcagno, **M. Kane**, M. S. Martinez-Silveira, M. G. Goris, C. Stein, A. I. Ko, B. Abela-Ridder. Global Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. *PLoS Negl Trop Dis*, 2015.

P. M. Rabinowitz, I. B. Slizovskiy, V. Lamers, S. J. Trufan, T. R. Holford, J. D. Dziura, P. N. Peduzzi, **M. J. Kane**, J. S. Reif, T. R. Weiss, M. H. Stowe. Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania. *Environ Health Perspect*, 2015.

A. Wrzesniewski, B. Schwartz, X. Cong, **M. J. Kane**, A. Omar, and T. Kolditz "Multiple types of motives don't multiply the motivation of West Point cadets," *PNAS*, 2014.

M. J. Kane and J. Emerson "Scalable Strategies for Computing with Massive Data," *The Journal of Statistical Software*, 2013.

M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz "Comparison of ARIMA and Random Forest Time Series Models for Prediction of Avian Influenza H5N1 Outbreaks," *BMC Bioinformatics*, 2013.

M. Scotch, C. Mei, Y. J. Makonnen, J. Pinto, A. Ali, S. Vegso, **M. J. Kane**, I. N. Sarkar, P. Rabinowitz "Phylogeography of Influenza A H5N1 Clade 2.2.1.1 in Egypt." *BMC Genomics (Section: Comparative and evolutionary genomics)*, 2013.

P. M. Rabinowitz, D. Galusha, S. Vegso, J. Michalove, S. Rinne, M. Scotch, **M. J. Kane** "Comparison of Human and Animal Surveillance Data for H5N1 Influenza A in Egypt 2006-2011," *PLoS ONE*, 2012.

M. E. K. Evans, R. Ferrier, **M. J. Kane**, D. L. Venable, "Bet Hedging via Seed Banking in Desert Evening Primrose (*Oenothera*, *Onagraceae*): Demographic Evidence from Natural Populations," *The American Naturalist*, 2007.

MANUSCRIPTS IN
SUBMISSION

M. DeVeaux, **M. J. Kane**, and D. Zelterman “A Two-Stage, Phase II Clinical Trial Design with Nested Criteria for Early Stopping and Efficacy.”

MANUSCRIPTS IN
PREPARATION

A. Athreya, M. Tang, V. Lyzinski, Y. Park, B. Lewis, **M. Kane**, C. Priebe “Numerical tolerance for spectral decompositions of random dot product graphs.”

J. Baglama, **M. J. Kane**, B. Lewis, and A. Poliakov “Efficient Thresholded Correlation using Truncated Singular Value Decomposition.”

M. DeVeaux, **M. J. Kane**, and D. Zelterman “A Stopped Negative Binomial Distribution,” *The arXiv*.

M. DeVeaux, **M. J. Kane**, and D. Zelterman “A Bayesian Sequential Probability Ratio Test.”

M. J. Kane “An Reproducing Kernel Hilbert Space Approach to Fitting Models from Sampled Sets of Data.”

BOOKS

M. van der Laan, **M. Kane**, P. Bühlmann, P. Drineas, *The Handbook of Big Data*.

OTHER
PUBLICATIONS

M. Kane “Review of Data-Intensive Computing,” *The Journal of The American Statistical Association*, 2015.

M. Kane “Cleveland’s Action Plan and the Development of Data Science over the Last 12 Years,” *Statistical Analysis and Data Mining*, 2015.

M. Kane, “Strategies for Exploring a 12 Gigabyte Data Set: Airline Flight Delays,” Invited book chapter in *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*, 2015.

M. J. Kane and B. Lewis “cnidaria: A Generative Communication Approach to Scalable, Distributed Learning,” *Proceedings of the NIPS Big Learning Workshop*, 2013.

J. Baglama, **M. J. Kane**, B. Lewis, L. Reichel “IRLBA, a fast partial SVD,” *Proceedings of the 12th annual Scientific Computing with Python Conference*, 2013.

S. A. Kazmi, **M. J. Kane**, M. Krauthammer “Benchmarking Technology Infrastructures for Embarrassingly and Non-embarrassingly Parallel Problems in the Biomedical Domain,” *Proceedings of the Annual ORNL Biomedical Sciences and Engineering Conference*, 2013.

M.J. Kane “Using bigmemory with Rcpp,” *The Rcpp Gallery*, 2013.

J. Emerson and **M. Kane**, “Don’t drown in the data,” *Significance*, 9(4): 38–39. Aug. 2012.

A. Savakis, J. Luo, and **M. Kane**, “Bayesian Networks for Image Understanding,” Invited book chapter in *Bayesian Network Technologies: Applications and Graphical Models*, A. Mittal and A. Kasim Eds., IGI Global, Hershey, PA, 2007.

M. J. Kane and A. Savakis, “Bayesian Network Structure Learning and Inference in Indoor vs. Outdoor Image Classification,” *Proceedings of the International Conference on Pattern Recognition*, 2004.

M. J. Kane, F. Sahin, and A. Savakis “A Two-Phase Approach to Bayesian Network Model Selection and Comparison between the MDL and DGM Scoring Heuristics,” *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, 2003.

OTHER MEDIA

“Practical Principles for Scalable Statistical Analysis.” *Presentation for the New York R Conference*. 2015.

“Distributed Data Structures in R for General, Large-Scale Computing.” *Presentation for R user group*. June 3, 2014.

“Podcast interview with Michael Kane,” Oracle R Enterprise Blog. Sept. 17, 2012. .

B. Alpert and L. Stryjewski, “Hitting the Switch on New Circuit Breakers,” *Barron’s Magazine*. Aug. 13, 2011.

OPEN SOURCE PROJECTS

bigalgebra: An R package providing high-performance linear algebra functions for use with R matrices along with matrices from **bigmemory**. (<http://www.bigmemory.org/>)

biganalytics: An R package which providing extended functionality for working with matrices provided by **bigmemory**. (<http://www.bigmemory.org/>)

bigmemory: An R package for handling massive data sets. (<http://www.bigmemory.org/>)

bigtabulate: An R package providing fast tabulate functions for use with R matrices along with matrices from **bigmemory**. (<http://www.bigmemory.org/>)

irlbpy: Truncated SVD by implicitly restarted Lanczos bidiagonalization for Python Numpy. (<https://github.com/bwlewis/irlbpy>)

Ratlab: An Octave/Matlab project which allows tight-coupling of data between Octave/Matlab and R. It also exposes R functionality to Octave/Matlab. (<http://github.com/bwlewis/ratlab>)

synchronicity: An R package providing interprocess communication and synchronization. (<http://www.bigmemory.org/>)

COURSES TAUGHT

BIS 678, Statistical Consulting, Fall 2011, 2013, 2014, 2015

STAT 625a, Case Studies, Fall 2010

COMMUNITY ACTIVITIES

Editor for the Journal of Statistical Software, 2014-Present.

Treasurer of the Connecticut Chapter of the American Statistical Association, 2017-2020.

Past-Chair of Statistical Graphics Section for the American Statistical Association, 2017.

Chair of Statistical Graphics Section for the American Statistical Association, 2016.

Secretary of the Connecticut Chapter of the American Statistical Association, 2016.

Chair of the Statistical Graphics Section for the Joint Statistical Meetings, 2015.

Chair Elect of the Statistical Graphics Section for the Joint Statistical Meetings, 2014.

Program Chair Elect of the Statistical Graphics Section for the Joint Statistical Meetings, 2013.

Reviewer: *Annals of Applied Statistics*, *The American Statistician*, *Journal of Computation and Graphical Statistics*, *Journal of Statistical Software*, *BMC Bioinformatics*, *Open-access Journal*, *Statistics and its Interface*, *PLOS ONE*, *Bioinformatics*