Rebecca Nugent

Department of Statistics & Data Science, Baker Hall 232G   412-268-7830
Carnegie Mellon University   rnugent@stat.cmu.edu
Pittsburgh, PA 15213   http://www.stat.cmu.edu/~rnugent

Positions

Department of Statistics & Data Science, Carnegie Mellon University

Department Head	June 2021 - present
Stephen E. and Joyce Fienberg Professor of Statistics & Data Science	Fall 2019 - present
Associate Department Head	March 2017 - June 2021
Co-Director of Undergraduate Studies	August 2019 - June 2021
Director of Undergraduate Studies	July 2016 - August 2019
Teaching Professor	July 2015
Co-Director of Undergraduate Studies	July 2014 - July 2016
Associate Teaching Professor	July 2012
Assistant Teaching Professor	July 2009
NSF VIGRE Postdoctoral Fellow/Visiting Assistant Professor	Fall 2006 - Summer 2009

Education

PhD in Statistics, University of Washington   Dec 2006
Dissertation Title: "Algorithms for Estimating the Cluster Tree of a Density"
Advisor: Werner Stuetzle; *Committee:* Adrian Raftery, Marina Meila, Peter Hoff, Jim Hermanson (GSR)

M.S. in Statistics, Stanford University   June 2001

B.A. with Majors in Mathematics, Statistics, and Spanish, Rice University   May 1999

Research Interests

Behavioral Data Science, Science of Data Science
characterizing and analyzing the data analysis/science pipeline; distributions/populations of statistical data analyses;
how do subjective decisions propagate? impact of human decisions on reproducibility/replicability

Integrated Statistics Learning Environment (ISLE)
building interactive statistics & data science platforms that support the entire data science pipeline:
research question hypothesis, data cleaning/manipulating, visualization, modeling, dissemination/interpretation;
platform tracks student actions and provides instructor summaries and student characterization
focused on researching how people write about data; http://www.stat.cmu.edu/isle

Clustering
nonparametric multivariate analysis methods; cluster trees; spanning trees; generalized single linkage clustering
assessing cluster significance; pruning and merging; high-dimensional density estimation and visualization
longitudinal clustering possibly in the presence of informative missingness; agreement indices
clustering with distributions of estimated distances or dissimilarities

Record Linkage, Classification, Text Mining
classification/clustering algorithms for disambiguating and linking text records (and other related structures);
determining networks of colleagues, historical family structures; focusing on classifier aggregation and clustering
distributions using representative records; applications include networks in 15th/16th century England, early 20th century
Ireland Census, large-scale U.S. Census administrative data, quantifying human rights violations, and characterizing
intellectual innovation via patents Collaborators include CMU Digital Humanities, UC Dublin Mathematics & Statistics,
U.S Census Bureau, Human Rights Data Analysis Group, CMU Engineering & Public Policy

Online Learning: How to Design to Optimize Learning and Student Interaction
clustering methodologies as proxies for cognitive diagnosis models; psychometrics
modeling student use of and interaction with different online learning media
Joint work with Department of Modern Languages, Human Computer Interaction Institute, Carnegie Mellon University

Public Health
example projects include characterizing sleep duration, quality, consistency and its association with health
and assessing medical residency education programs
Joint work with Department of Internal Medicine, Texas Tech Health Sciences Center

Honors and Awards

President, International Federation of Classification Societies 2023
https://ifcs.boku.ac.at/site/doku.php
President-Elect, American Statistical Association Section on Statistics & Data Science Education 2020
(elected but declined due to new position as Dept Head)
Stephen E. and Joyce Fienberg Professorship in Statistics & Data Science Fall 2019
Carnegie Mellon University
American Statistical Association Waller Education Award 2015
national award for innovation in statistics education
The William H. and Frances S. Ryan Award for Meritorious Teaching 2015
Carnegie Mellon University (top teaching honor given by the university)
Statistician of the Year 2015
American Statistical Association Pittsburgh Chapter
Elliott Dunlap Smith Award for Distinguished Teaching and Educational Service 2012-2013
Dietrich College of Humanities & Social Sciences, Carnegie Mellon University
Distinguished Visiting Professor, Bucknell University October 2012, April 2014
Honorarium Recipient for invited seminars and research trips
City University of New York Graduate School June 2011
Honorarium Recipient for invited talk
Chikio Hayashi Award 2009
Young Promising Researcher award (under 35 years old); monetary prize and travel grant
presented biennially by the International Federation of Classification Societies (IFCS)
Carnegie Mellon University Panhellenic Role Model and Mentor Award Fall 2007-2013
undergraduate students nominated award
Winner, American Statistical Association Statistical Computing 2006
and Graphics Student Paper Competition
University of Washington Excellence in Teaching Award 2003
Dorothy M. Gilford Award: Excellence in Teaching by a Graduate Student 2001 - 2002
Department of Statistics, University of Washington, Seattle, WA
Rice Engineering Alumni Senior Engineering Award/Scholarship May 1999
Impact Award - Rice Women's Resource Center May 1999
Who's Who in American Universities 1998-1999
Rice University Vice Presidential Appreciation Award May 1998
Commendation from the Senate of the State of Texas 1992

Advisory Boards/Management Teams

NSF AI Institute for Societal Decision Making, Senior Management Team 2023 - present

West Point Applied Math and Statistics, Data Science Advisory Board 2022 - present
West Point Academy, NY

Learning the Earth and Artificial Intelligence and Physics (LEAP), Advisory Board 2021 - present
NSF Science and Technology Center, Columbia University

Testimony

Pennsylvania Senate Majority Policy Committee, September 2023
Pennsylvania as an Innovation Leader: Artificial Intelligence Breakthroughs
Pittsburgh Technology Council, Pittsburgh, PA

National Academies of Science, Engineering, and Medicine

CNSTAT: AI Day for Federal Statistics May 2024
Moderator, Representative for the NSF AI Institute for Societal Decision Making

Board on Atmospheric Sciences and Climate Feb 2022
Invited Expert, *Broadening Data Science Education for the Future Biomanufacturing Workforce*

Division of Engineering and Physical Sciences Nov 2021
Invited Expert/Speaker, *Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenge*

Idea Competition, Symposium on Imagining the Future of Undergraduate STEM Education　Nov 2020
Winner, *Personalized Learning Environments for Student-Centric STEM*
`https://www.youtube.com/watch?v=Yv8mTG_MUlM&feature=youtu.be`

Division on Eng. and Physical Sciences, Board on Mathematical Sciences and Analytics　2019-2020
Co-Chair, Study Committee on *Improving Defense Acquisition Workforce Capability in Data Use*

Policy and Global Affairs, Board on Research Data and Information　March 2020
Panelist on *Workshop on Accelerating Scientific Discovery through Advanced and Automated Workflows*

Committee on Applied and Theoretical Statistics　2016-2018
Study Committee member on *Envisioning the Data Science Discipline: The Undergraduate Perspective*

Division on Eng. and Physical Sciences, Board on Mathematical Sciences and Analytics　March 2019
Speaker/Panelist on *Roundtable on Data Science Postsecondary Education: Mechanisms for Engaging and Fostering Industry Partnerships*

Board on Chemical Sciences and Technology　Winter 2018
Speaker/Panelist on *Data Science: Opportunities to Transform Chemical Sciences and Engineering*

External Reviews

George Washington University Data Science Program	Winter 2023
Willams College Dept of Mathematics and Statistics	Fall 2021
Bucknell University Mathematical Economics Program, Chair	Winter 2020
Minerva College of Computational Sciences Program	Winter 2020

External Leadership/Board Positions

Mahindra & Mahindra Financial Services Limited, Independent Director, Executive Board　2021 - present

United Nations Task Team on Training, Competencies, and Capacity Development　Winter, Spring 2021
External Member

Grants & Research Gifts

NSF AI Institute for Societal Decision Making　June 2023 - May 2028
NSF-funded institute focusing on bridging human and autonomous decision-making
in public health and disaster management use cases; `https://www.cmu.edu/ai-sdm/`
Consortium with CMU (lead), Harvard, Texas A&M, Howard, U Washington, and others
Senior Management Team; Education and Workforce Lead
$20,000,000

IES: DS4EDU: Methods Training in Data Science for Education Researchers　Sept 2023-Aug 2026
Three year training program to develop network of educational technology practitioners
using data-driven methods for learning innovation
`https://www.cmu.edu/simon/projects/flagship-projects/ds4edu/index.html`
Co-PI w/ John Stamper (PI),Carolyn Rose (Co-PI),Ken Koedinger (Co-PI),Norman Bier (Co-PI)
$799,523; three years

NSF: Improving Undergraduate STEM Education　June 2022 - May 2026
SCORE with Data: Building a sustainable national network for developing and disseminating
Sports Content for Outreach, Research, and Education in data science
PI with N Clark (co-PI), K Pelechrinis (co-PI), M Schuckers (co-PI), R Sturdivant (co-PI)
$1,100,000; four years

Optum (UnitedHealth Group)　Summer 2021
PHIGHT COVID: research on the impact of non-pharmaceutical interventions and school modalities
on COVID-19 case rates
PI with Seema Lakdawala (PI, U Pittsburgh Microbiology); $50,000 unrestricted gift

Carnegie Mellon ProSeed/Simon Initiative　Summer/Fall 2020
Integrating a Statistical Learning Environment into the Writing Classroom
David West Brown (PI), Rebecca Nugent (co-PI), Philipp Burckhardt (co-PI); $15,000

NSF: Division of Mathematical Sciences   October 2017 - September 2019
Improving Probabilistic Record Linkage
$92,348; Sub-grant from Jared Murray, University of Texas

Carnegie Mellon ProSEED/Simon Initiative   May 2018
Data Analysis Think-Alouds: Student Engagement and Workflow on
an Online Interactive Statistical Analysis Tool; $15,000
PI with Philipp Burckhardt (Co-Investigator)

NIH: National Institute on Aging   January 2018 - May 2018
How Does Automated Record Linkage Affect Inferences about Population Health?
$27,049; Sub-grant from Martha Bailey (LIFE-M), University of Michigan

Carnegie Mellon ProSEED/Crosswalk   February 2018
Women in Statistics at CMU: Fostering collaboration
through formal mentorship; one year, $1685
Faculty advisor; K Frisoli (PI), S Gallagher (PI), A Luby (PI)

NSF: The NSF-Census Research Network Supplement   September 2016 - August 2017
Census Research Node: Data Integration, Online Data Collection,
and Privacy Protection for Census 2020; $\approx \$650,000$
co-PI with S Fienberg (PI), W Eddy (PI), A Acquisti (co-PI)

Carnegie Mellon ProSEED/Crosswalk   Summer 2015
What is Statistics? An Interactive Platform that Engages and Educates the Non-Statistician
co-PI with Paige Houser (PI) and Howard Seltman (co-PI); $2500

NSF: The NSF-Census Research Network   Sept 2011 - Sept 2016
Census Research Node: Data Integration, Online Data Collection,
and Privacy Protection for Census 2020; $\approx \$3,000,000$
co-PI with S Fienberg (PI), W Eddy (PI), A Acquisti (co-PI)

NSF: Research Training Groups in the Mathematical Sciences   May 2011 - June 2016
Statistics and Machine Learning for Scientific Inference; $2,250,979
co-PI with R Kass (PI), W Eddy (PI)

CMU Berkman Faculty Development Fund   Summer 2014
Determining the Cluster Structure in High-Dimensional Data: A Visualization Tool for Merging Clusters
graduate student summer support; $2975

Association for Women in Mathematics   Summer 2011
Self-Tuning Diffusion Maps: Finding Local Cluster Structure
while Reducing Dimensionality; research/travel grant: $2000
(declined due to last minute unavailability to travel)

The Royal Society of Edinburgh   Summer 2008
Merging Clustering Methodologies for Visualization and Estimation of Group Structure
with Nema Dean, Department of Statistics, University of Glasgow
research/travel grant: $\approx \$5500$ (based on then conversion rate)

Carnegie Mellon University Summer Undergraduate Research Fellowship   Summer 2008
Using Statistical Techniques to Improve Disease Classification
with Ryan Sieberg, Departments of Mathematical Sciences, Carnegie Mellon University
research grant: $3500; (declined due to another student support opportunity)

Corporate-Sponsored Research and Educational Projects

Corporate Capstone Program: Department of Statistics & Data Science, Carnegie Mellon University

Founding Director, experiential learning data science program for undergraduate, master's, PhD   2018 - present

Partners have included: Dexcom, Pittsburgh Pirates, Pittsburgh Penguins, Allegheny County Health Department,
Pittsburgh Park Conservancy, United States Olympic Committee, Optum/UnitedHealth Group, PNC,
ThermoFisher Scientific, The NPD Group, Giant Eagle, Koppers, Chain of Demand, Penguin Random House,
Principal Financial, IKOS, Pack Up + Go, Steady (app), TruMedia, CivicScience, Black & Veatch,
C.H. Robinson Worldwide, Inc

Executive Education/Professional Development Programs

PNC Rising Data Scientists   2022-2024
Dept of Statistics & Data Science, Tepper School of Business, Carnegie Mellon
four month program professional development program for rising data scientists
Co-Faculty Director/Developer, instructor, online modules, custom ISLE analytics platform

Mahindra Group Tech Vision,   Spring 2023
Dept of Statistics & Data Science, Tepper School of Business, Carnegie Mellon
AI and data science-focused content for the Mahindra & Mahindra C-level executives

Moderna AI Academy   2021-2022
Partnership between Carnegie Mellon Statistics & Data Science/Dietrich College, Tepper School
of Business & Moderna to create customized AI/data science academy for entire enterprise
Co-Faculty Director/Developer, instructor, online modules, custom ISLE analytics platform

Optum/United Health Group Data Science for Business Leaders (three cohorts/yr)   Spring 2021 - present
Dept of Statistics & Data Science, Tepper School of Business, Carnegie Mellon
Co-Faculty Director, instructor, online modules, custom ISLE analytics platform

Optum AI for Business Leaders, Tepper School of Business, Carnegie Mellon   October 2019 - present
The Role of Data Science, Data Life Cycle (three cohorts/yr)
instructor, online modules, custom ISLE analytics platform

PNC Strategy and Innovation,   2022-present
Data Visualization workshops, instructor, online modules, custom ISLE analytics platform

Chief Digital Officer Certificate, (open enrollment)   2021-present
Heinz College of Information Systems and Public Policy, Carnegie Mellon
instructor, online modules, custom ISLE analytics platform

Vapotherm AI for Business, Tepper School of Business, Carnegie Mellon   Spring 2022
The Role of Data Science, Data Life Cycle and LInkage
instructor, online modules, custom ISLE analytics platform

Mahindra Group, Carnegie Mellon Silicon Valley Campus   January 2019
Innovating with Data Science: Combining Human Wisdom with Data Analytics

Online Certificate Programs

Department of Statistics & Data Science, Carnegie Mellon University:

Foundations of Data Science   full launch Fall 2024
online graduate-level certificate designed for workforce professionals and students
wanting to increase their practical data science skills; part-time/full-time available
https://www.cmu.edu/online/fds/

Summer Research Programs

Department of Statistics & Data Science, Carnegie Mellon University:

Bridges to Healthcare Technology   Summer 2022-present
summer data science/engineering undergrad research program sponsored by UnitedHealth Group
targeting under-resourced populations; includes real-world health care projects, professional
development, and internship opportunities https://www.stat.cmu.edu/summerbridge

Carnegie Mellon Sports Analytics Camp   Summer 2019-present
summer data science undergraduate research program
includes real-world data science problems with professional sports teams
https://www.stat.cmu.edu/cmsac

Statistics & Data Science Degrees

Department of Statistics & Data Science, Carnegie Mellon University:

Co-developer of:
Data Science in Society undergraduate minor   Slated for Fall 2024 launch
Statistics & Machine Learning undergraduate degree   Spring 2015
Statistics & Neuroscience undergraduate degree track   Spring 2015
Statistics & Mathematical Sciences undergraduate degree track   Spring 2015

Teaching & Course Development

Department of Statistics & Data Science, Carnegie Mellon University: (ordered by level)

36-764 - Teaching Statistics   Fall 2017, 2019; Spring 2017, 2018
PhD pedagogy journal club/training course
36-792/692/492 - Topic Detection and Document Clustering   Spring 2014
PhD/master's/undergraduate methodology/application course
36-721 - Statistical Graphics and Visualization   Fall 2010
PhD course for graduate students in statistics, computer science, machine learning
36-729 - Unsupervised Learning   Fall 2009
PhD course for graduate students in statistics, computer science, machine learning
36-691/491 - Data Matching Methods and Their Uses   Fall 2013
master's/undergraduate record linkage course
36-497 - Corporate Capstone   Spring 2018 - present
industry data science research projects; invitation-only course
36-493 - Sports Analytics Capstone   Spring 2020
sports analytics data science research projects; invitation-only course
36-490 - Undergraduate Research   Spring 2011, 2015; Fall 2020
undergraduate research projects; invitation-only course
36-462 - Topics in Statistics: Statistical Learning   Spring 2010
master's/senior undergraduate course in unsupervised, supervised learning
36-401 - Modern Regression   Fall 2006-2012, 2014, 2016
undergraduate/master's linear regression/data analysis for Stat, Math, CS majors
36-375 - Data Ethics & Responsibility   Spring 2018
undergraduate level seminar on data ethics, responsibility, integrity, etc
36-315 - Statistical Graphics and Visualization   Spring 2008-2012, 2014
undergraduate graphics/programming course for majors in statistics, math, CS, etc.
36-303 - Sampling, Survey, and Society   Spring 2007
undergraduate course examining role of sample surveys in U.S. society
36-226 - Introduction to Statistical Inference   Spring 2012, 2015
undergraduate mathematical statistics course for majors in statistics, math, CS, etc.
36-202 - Methods for Statistics & Data Science   Spring 2019
intro level 2nd course for quantitative majors on modeling and statistical learning
36-200 - Reasoning with Data   Fall 2017, 2018; Spring 2017, 2018
introductory level course for all majors focusing on data science and its application
36-149 - Freshmen Statistics Seminar   Fall 2012, Spring 2016
Networks: Where do they come from? What do they tell us?

Tepper School of Business, Master's in Computational Finance Program, Carnegie Mellon University:

46-921 - Probability   Fall 2013
46-923 - Statistical Inference   Fall 2013

China Education Association for International Exchange, Summer China Program:

Calculus II, Intro to Statistics   Summer 2012
Hosted at University of Science & Technology Beijing; http://en.ceaie.edu.cn/
CEAIE established by the Chinese Ministry of Education and the Ministry of Foreign Affairs

Center for Statistics & Social Sciences, University of Washington:

Instructor/Developer - CS&SS Math Camp   Sept 2004, Sept 2005
Developer of one week intensive introduction to fundamental concepts of
mathematics, probability, and statistics designed for graduate students in the social sciences
Author of all lecture and homework materials; http://www.csss.washington.edu/MathCamp

Workshops and Tutorials

Villanova Center for Statistics Education Workshop   May 2017
Classification and Clustering: The Basics, The Next Level

Park City Math Institute Undergraduate Summer School (PCMI 2016)   July 2016
Visualizing and Learning the Structure in Data
lecturer and author of materials of month-long program; `http://www.stat.cmu.edu/~rnugent/PCMI2016/`

ASA Conference on Statistical Practice (CSP 2015)   February 2015
An Overview of Clustering: Finding and Extracting Group Structure in High-Dimensional Data
presenter and author of lecture materials; `http://www.stat.cmu.edu/~rnugent/CSP2015/`
supplemental materials and code provided by Sam Ventura

7th International Conference on Educational Data Mining (EDM 2014)   June 2014
An Overview of Clustering: Finding Group Structure in Educational Research Data
presenter and author of materials; `http://www.stat.cmu.edu/~rnugent/EDM2014/`

Refereed Publications

36) Pane J, Murray J, Nugent R, Yang S, Nugent K. "Electronic cigarette use by and perceptions of middle and high school students in the United States". *Journal of Investigative Medicine*, 2023 Mar; 71(3): 212-222. `https://doi.org/10.1177/10815589221140588`

35) Yang S, Nugent R, Nugent K. "The likelihood of adolescents trying conventional or electronic cigarettes varies with their use and their perception of harm of other tobacco products, including cigars". *Southern Medical Journal*, 2022 Sep; 115(9): 665-673. `https://doi.org/10.14423/SMJ.0000000000001445`

34) Reinhart, A, Evans C, Luby A, Orellana J, Meyer M, Wieczorek J, Elliott P, Burckhardt P, Nugent R. "Think-Aloud Interviews: A Tool for Exploring Student Statistical Reasoning". *Journal of Statistics and Data Science Education*, 30:2, 100-113, `https://doi.org/10.1080/26939169.2022.2063209`

33) Burckhardt P, Nugent R, Genovese C. "Teaching Statistical Concepts and Modern Data Analysis with a Computing-Integrated Learning Environment". *Journal of Statistics Education*, 29:sup1, S61-S73, `https://doi.org/10.1080/10691898.2020.1854637`

32) Nugent E, Nugent A, Nugent R, Nugent K, Nugent C "The management of women's health care by internists with a focus on the utility of ultrasound". *The American Journal of the Medical Sciences*, Volume 360, Issue 5, November 2020, p. 435-446. `https://doi.org/10.1016/j.amjms.2020.05.016`

31) Nugent K, Raj R, and Nugent, R. "Sleep Patterns and Health Behaviors in Health Care Students". *Southern Medical Journal*, Vol 113, No. 3 March 2020, p.104-110.

30) Frisoli K, LeRoy, B, and Nugent, R. "A novel record linkage interface that incorporates group structure to rapidly collect richer labels". *6th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, September 2019, pp. 580-589. DOI: `10.1109/DSAA.2019.00073`

29) Flynt A, Dean N, Nugent R. "A *soft* agreement measure for class partitions incorporating assignment probabilities". *Advances in Data Analysis and Classification*, March 2019, Vol 13, Number 1, p.303-323.

28) Frisoli, K and Nugent, R. "Exploring the effect of household structure in historical record linkage of early 1900s Ireland census records". *IEEE International Conference on Data Mining Workshops (ICDMW)*, November 2018, pp.502-509. DOI: `10.1109/ICDMW.2018`

27) Youngs B, Prakash A, Nugent R. "Statistically-driven Visualizations of Student Interactions in an French Online Course Video". *Journal of Computer-Assisted Language Learning, Special Edition on Learning Analytics*. Published online September 2017, `http://www.tandfonline.com/doi/full/10.1080/09588221.2017.1367311`.

26) Nugent E, Nugent A, Nugent R, Nugent K. "Zika virus: epidemiology, pathogenesis, and human disease". *The American Journal of the Medical Sciences*, Volume 353, No. 5, May 2017, p. 466-473.

25) Unger L, Fisher A, Nugent R, Ventura S, and MacLellan C. "Development Changes in the Semantic Organization of Living Kinds", *Journal of Experimental Child Psychology*, Vol 146, June 2016, p.202-222.

24) Yang C, Nugent R, Fuchs E. "Gains from Others' Losses: Technology Trajectories and the Global Division of Firms." *Research Policy*, Vol 45, Issue 3, April 2016, p.724-745.

23) Ventura S, Nugent R, Fuchs E. "Seeing the non-stars: (Some) sources of bias in past disambiguation approaches and a new public tool leveraging labeled records". *Research Policy (Special Issue on Big Data)*, Vol 44, Issue 9, Nov 2015, p.1672-1701.

22) Narayanan R, Nugent R, Nugent K. "An Investigation of the Variety and Complexity of Statistical Methods Used in Current Internal Medicine Literature", *Southern Medical Journal*, Vol 108, No. 10, Oct 2015.
Selected for Invited Commentary

21) Ventura S, Nugent R, and Fuchs E. "Hierarchical Linkage Clustering with Distributions of Distances for Large-Scale Record Linkage". *Privacy in Statistical Databases (Lecture Notes in Computer Science 8744)*, ed. J. Domingo-Ferrer, Springer, p.283-298 (2014).

20) Nugent R, Althouse, A, Yaqub Y, Nugent K, Raj, R "Modeling the relationship between obesity and sleep parameters in children referred for dietary weight reduction intervention". *Southern Medical Journal, Special Series: Obesity*, Vol 107, Issue 8, p. 473-480 (2014).

19) Dean, N and Nugent, R. "Clustering student skill set profiles in a unit hypercube using mixtures of multivariate betas". *Advances in Data Analysis and Classification*, Vol 7, No, 3, p.339-357 (2013).

18) Ayers E, Rabe-Hesketh S, Nugent R. "Incorporating Student Covariates in Cognitive Diagnosis Models". *Journal of Classification*, 30: 195-224 (2013).

17) Rupp, A, Nugent R, Nelson B. "Evidence-centered Design for Diagnostic Assessment within Digital Learning Environments: Integrating Modern Psychometrics and Educational Data Mining". *Journal of Educational Data Mining*, Volume 4, Issue 1, October 2012. Pages 1-10.

16) Nelson B, Nugent R, Rupp A. "On Instructional Utility, Statistical Methodology, and the Added Value of ECD: Lessons Learned from the Special Issue". *Journal of Educational Data Mining*, Volume 4, Issue 1, October 2012. Pages 224-230.

15) Nourbaksh E, Nugent R, Wang H, Cevik C, and Nugent K. "Medical Literature Searches: PubMed Central or Google Scholar". *Health Information and Libraries Journal*. 2012; 29:214-22.
Additionally selected to be part of a special issue on *The Role of the Health Information Professional* marking the CILIP Health Libraries Group Conference, Oxford, 2014.

14) Rinaldo A, Singh A, Nugent R, Wasserman L. "Stability of Density-Based Clustering". *Journal of Machine Learning Research* 13(Apr):905-948, 2012.

13) Friedenberg D, Nugent R. "Exploration of the Use of a Self-Tuning Diffusion Map Framework". *Int. Statistical Institute: Proceedings 58th World Statistical Congress, 2011, Dublin (Session IPS040)*, Dec 2012.

12) Dean N, Nugent R. "Comparing Different Clustering Methods on the Unit Hypercube". *Int. Statistical Institute: Proceedings 58th World Statistical Congress, 2011, Dublin (Session IPS040)*, Dec 2012.

11) Nugent R, Dean N, Ayers E. "Skill Set Profile Clustering: The Empty K-Means Algorithm with Automatic Specification of Starting Cluster Centers". *Educational Data Mining 2010: 3rd International Conference on Educational Data Mining, Proceedings* (refereed). Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.), p.151-160.

10) Nugent R, Stuetzle W. "Clustering with Confidence: A Low-Dimensional Binning Approach". *"Classification as a Tool for Research". Proceedings of the 11th International Federation of Classification Societies Conference* (refereed), Herman Locarek-Junge, Claus Weihs (editors), University of Dresden, Germany, March 13-18, 2009. Springer-Verlag, Heidelberg-Berlin, 2010, p.117-126.

9) Stuetzle W, Nugent R. "A Generalized Single Linkage Method for Estimating the Cluster Tree of a Density". *The Journal of Computational and Graphical Statistics*, 2010, Vol. 19, 2, p.397-418.

8) Wang H, Nugent R, Nugent C, Nugent K, Phy M. "A Commentary of the Use of the Internal Medicine In-Training Examination". *The American Journal of Medicine*. Vol 122, No 9, September 2009, p.879-883.

7) Nugent R, Ayers E, Dean N. "Conditional Subspace Clustering with Skill Mastery Information: Identifying Skills that Separate Students". *Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings* (refereed). Barnes, T., Desmarais, M., Romero, C., and Ventura, S. (Eds), Cordoba, Spain, July 1-3, 2009, p.101-110.

6) Ayers E, Nugent R, Dean N. "A Comparison of Student Skill Knowledge Estimates". *Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings* (refereed). Barnes, T., Desmarais, M., Romero, C., and Ventura, S. (Eds), Cordoba, Spain, July 1-3, 2009, p.1-10.

5) Ayers E, Nugent R, Dean N. "Skill Set Profile Clustering Based on Student Capability Vectors Computed from Online Tutoring Data". *Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings* (refereed). R.S.J.d. Baker, T. Barnes, and J.E. Beck (Eds), Montreal, Quebec, Canada, June 20-21, 2008. p.210-217.

4) Buscemi D, Kumar A, Nugent R, Nugent K. "Short Sleep Times Predict Obesity in Internal Medicine Clinic Patients". *Journal of Clinical Sleep Medicine*, Vol 3, No. 7. Dec 2007. p. 661-688.

3) Glaser SL, Clarke CA, Keegan THM, Gomez SL, Nugent RA, Topol B, Stearns CB, Stewart SL. "Attenuation of social class and reproductive risk factors for Hodgkin lymphoma due to selection bias in controls". *Cancer Causes Control*: 2004; 15:731-9.

2) Glaser SL, Clarke CA, Nugent RA, Stearns CB, Dorfman RF. "Reproductive Factors in Hodgkin's disease in women". *American Journal of Epidemiology*: 2003; 158(6):553-563.

1) Glaser SL, Clarke CA, Nugent RA, Dorfman RF, Stearns CB. "Social class and risk of Hodgkin's disease in young adult women in 1988-94". *International Journal of Cancer*: 2002; 98(1):110-17.

In Revision/Submitted Manuscripts

Yang S, Nugent R, Nugent K. "When should clinicians use the term syndrome?" **In revision**, Fall 2022.

Avery A, Wang J, Ma X, Pan A, McGrady E, Yuan Z, Liang E, Nugent R, Lakdawala S. "Variations in Non-Pharmaceutical Interventions by State Correlate with COVID-19 Disease Outcomes". **Submitted**, July 2021.

Ehman C, Luo Y, Yang Z, Zhu Z, Donovan S, Avery A, Wang J, Lawler J, Nugent R, Ventura V, Lakdawala S. "K-12 School Teaching Posture Correlates with COVID-19 Disease Outcomes in Ohio". **Submitted**, July 2021.

Yurko, R and Nugent, R. "MMA: Maximum Model Agreement for Model-Based Clustering with Variable Selection". **In revision**

Invited Commentaries & Discussions

10) Yurko R and Nugent R. "Discussion of *Flexible marked spatio-temporal point processes with applications to event sequences from association football* by Narayanan, Kosmidis, and Dellaportas". *Journal of the Royal Statistical Society, Series C* (2023), Vol 72, Issue 5, p.1134-1135.

9) Frisoli K and Nugent R. "Discussion of *Statistical challenges of administrative and transaction data* by Hand". *Journal of the Royal Statistical Society, Series A* (2018), Vol 181, Issue 3, p.590.

8) Alvarez, Espanol, Faridani, Flores, Marr, McNulty, Newman, Nugent, Seneres, Shott, Velez, Walker (alphabetical order). "The PCMI workshop for mentors: A weeklong workshop on diversity?". *Notices of the American Mathematical Society* (May 2018), Vol 65, Issue 5, p. 586-591.

7) Nugent R, Lorenzi E, and Frisoli K. "Discussion of *A Bayesian Information Criterion for Singular Models* by Drton and Plummer". *Journal of the Royal Statistical Society, Series B* (2017), Vol 79, Issue 2, p.371.

6) Ventura S and Nugent R. "Discussion of *Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings* by Ehm, Gneiting, Jordan, and Kruger". *Journal of the Royal Statistical Society, Series B* (2016), Vol 78, Issue 3, p.555.

5) Flynt A and Nugent R. "Discussion of *Statistical Modelling of Citation Exchange Among Statistics Journals* by Varin, Cattelan, and Firth". *Journal of the Royal Statistical Society, Series A* (2016), Vol 179, Issue 1, p.47-49.

4) Nugent R and Lorenzi E. "Discussion of *Analysis of forensic DNA mixtures with artefacts* by Cowell, Graversen, Lauritzen, and Mortera". *Journal of the Royal Statistical Society C* (2015), Vol 64, Issue 1, p.43.

3) Nugent R and Flynt A. "Discussion of *How to find an appropriate clustering for mixed type variables with application to socio-economic stratification* by Hennig and Liao". *Journal of the Royal Statistical Society C* (2013), Vol 62, Part 3, p.47-48.

2) Nugent R. "Maintaining Quality in the Face of Rapid Program Expansion". *AMSTATNEWS: The Membership Magazine of the American Statistical Association*. August 2012, Issue #422, p 14-15.

1) Nugent R, Rinaldo A, Singh A, Wasserman L. "Discussion on *Stability Selection* by Meinshausen and Buhlmann". *Journal of the Royal Statistical Society B* (2010), Vol 72, Part 4, p.465. (authorship in alphabetical order).

Book Chapters

Ngamruengphong S, Nugent A, Nugent K, Nugent R. (authorship in alphabetical order) "Case 49: Prostate-Ca-Survival". *Case Files: Geriatrics (LANGE Case Files)*, Andrew Dentino, MD (editor). McGraw-Hill Medical, 2014.

Nugent R and Meila M. "An Overview of Clustering Applied to Molecular Biology". *Statistical Methods in Molecular Biology*. Springer/Humana Press, 2010.

Invited Keynotes, Plenaries, Addresses, and Talks *(slated/upcoming in italics)*

AI24 by Lamarr Institute,   Sept 2024
Keynote, TBD

Koppers Tech Talk, March 2024
Keynote, *Data Science: Then, Now, and Forever*

Corporate Startup Lab Forum, Swartz Center for Entrepreneurship, Carnegie Mellon November 2024
Keynote/Interview, *The State of AI*

Digital Week Dortmund, September 2023
Keynote/Transatlantic Fireside Chat on AI (traveled as part of Pennsylvania delegation)

Joint Statistical Meetings (JSM 2023), August 2023
Talk, *Data Science Upskilling in Industry: What to Do when No One Has Any Time or Knows How to Code*

Joint Statistical Meetings (JSM 2023), August 2023
Invited Poster, *SCORE with Data: Sports Content for Outreach, Research, and Education*

International Association for Statistical Education (IASE 2023), July 2023
Talk/Paper, *SCORE with Data: Sports Content for Outreach, Research, and Education*

Corporate Startup Lab Forum, Swartz Center for Entrepreneurship, Carnegie Mellon November 2022
Keynote, *It's All About the Data (Pipelines)*

XXXI Scientific Conference of the Classification and Data Analysis Section September 2022
Polish Statistical Society, Keynote, *Optimizing and Clustering Data Science Workflows*

Joint Statistical Meetings (JSM 2022) August 2022
Invited Panel, *Promoting Diversity in Sports Analytics*

International Federation of Classification Societies (IFCS 2022) July 2022
Presidential Address, *Embracing, Optimizing, and Empowering with Data Science*

Earth Science Information Partners (ESIP 2022) July 2022
Keynote, *Democratizing Data: "Everyone is a Data Scientist"*

Classification Society (CS2022) June 2022
Invited Talk, *Demystifying, Optimizing, and Clustering Data Science Workflow Patterns*

Industry Federation of the State of Rio de Janeiro (FIRJAN 2022) May 2022
Keynote, *Data as a Business Strategy*

6th Conference of the Deutsche Arbeitsgemeinschaft Statistik (DAGStat 2022) March 2022
Invited Talk, *Demystifying and Optimizing Data Science*

James Madison SUMS Conference November 2021
Closing Keynote, *Demystifying Data Science*

Working Group on Model-Based Clustering, Greece (virtual) October 2021
Invited Long Talk, *Tackling how to Cluster Student Writing and Learning*

Joint Statistical Meetings (JSM 2021) August 2021
Opening Invited Poster Session, *Supporting and Studying Collaborative Data Analysis and Writing
in Statistics and Data Science using the Integrated Statistics Learning Environment*
Invited Talk/Panel, *Using Sports Analytics to Inspire Student Interest in Statistics
and Storytelling with Data*

U.S. Conference on Teaching Statistics June 2021
Opening Keynote, *Democratizing Data (Science): Empowering and Expanding Opportunities
for Both Students and Educators*

Uber Data Science December 2020
Data Science: Starts with People, Ends with People...and They're in the Middle too

New England Statistical Society (NESS), NextGen 2020 November 2020
Keynote, *Demystifying Data Science: Starts with People, Ends with People*

Academic Data Science Alliance, Annual Meeting/Leadership Summit October 2020
Teaching and Researching Data Science with ISLE

JazzHR, Leadership Retreat October 2020
Demystifying Data Science: Leveraging Data as an Asset but also...Starts with People, Ends with People

Joint Statistical Meetings (JSM2020), Virtual   August 2020
Invited Poster: *ISLE: An Integrated Learning (and Research) Environment for Statistics & Data Science*
Invited Panel: *Teaching-Focused Careers in Colleges, Universities, and Industry*

NSF/Berkeley 2020 National Workshop on Data Science Education, Invited Panel, Virtual   June 2020
Institutional Transformations

ASA Symposium on Data Science and Statistics, Opening Keynote, Pittsburgh   June 2020
Instead of Just Teaching Data Science, Let's Understand How and Why People Do it

The Fort AI and Data Summit, Fortive, Keynote   June 2020
Demystifying Data Science: Transportation, Pandemics, and People

goto Chicago 2020, Keynote   April 2020
Data Science for Everyone with ISLE: Leveraging Web Technologies to Increase Data Acumen

United Nations, Statistics Division, NYC   *March 2020, postponed (COVID-19)*

*Keynote, **NIH Friday Before Pi Day**, Baltimore*   *March 2020, postponed (COVID-19)*

Moore-Sloan Data Science Leadership Summit, Santa Fe   November 2019
Experiential Learning with Corporate Capstones

Machine Learning Workshop Galicia (WGML 2019 Keynote), A Coruna, Spain   October 2019
Before Teaching Data Science, Let's Understand How and Why People Do it

International Federation of Classification Societies, Thessaloniki, Greece   August 2019
President's Invited Session: Data Science Education
Before Teaching Data Science, Let's Understand How People Do it

National Academies on Sciences, Engineering, and Medicine, Washington D.C.   August 2019
Board Meeting on Chemical Sciences and Technology (Data Science testimony)

Google, Data Conference Keynote, Bay Area   August 2019
Data Science: Everyone is Doing It, But What Are They Actually Doing?

Joint Statistical Meetings (JSM 2019) Denver, CO (two talks)   August 2019
Making an Impact in Statistics Education: Waller Award Winner Perspectives
Experiential Learning (Chairs' Workshop)

Model-Based Clustering Working Group Vienna Austria (Short Talk)   July 2019
MMA: Maximum Model Agreement for Model-Based Clustering with Variable Selection

Statistics in the Liberal Arts Workshop, Amherst, MA   July 2019
Overview of ISLE and Behavioral Data Science

Classification Society Annual Meetings, Edmonton, Canada   June 2019
Exploring the impact of household structure and user-driven labels on the linkage
of early 1900s Irish census records

Institute for Mathematics and its Applications/Math Alliance, Minneapolis, MA   June 2019
Facilitated Graduate Admissions Process Workshop: Plenary on Data Science

U.S. Conference on Teaching Statistics, Penn State, PA   May 2019
'Many Students, One Dataset': Using ISLE to Teach Reproducibility
and the Impact of Data Analysis Decisions on Conclusions

Conference Board of the Mathematical Sciences, Alexandria, VA   May 2019
Overview of the National Academies Report on Envisioning the Data Science Discipline:
The Undergraduate Perspective

Eastern North American Region International Biometric Society (ENAR)   March 2019
Before Teaching Data Science, Let's First Understand How People Do It

National Academies of Science, Engineering, and Medicine, Irvine, CA   March 2019
Roundtable on Data Science Postsecondary Education: Mechanisms for Engaging and Fostering
Industry Partnerships

Consortium for the Advancement of Undergraduate Statistics Education (CAUSE)   December 2018
national webinar on ISLE, a new data science analytics platform

Joint Statistical Meetings (JSM 2018) Vancouver, Canada   August 2018
Data Science for Everybody: Building an Interactive, Adaptive Software Platform (ISLE) &
Analyzing Student Data Analysis Pipelines

Joint Statistical Meetings (JSM 2018) Vancouver, Canada   August 2018
Introductory Overview Lecture: *Evolution of the Undergraduate Statistics & Data Science Program*

ASA Symposium on Data Science and Statistics Reston, Virginia   May 2018
Data Science for Everybody: Building and Characterizing Student-Driven Pathways
in Introductory Statistics Courses

Bryant University Research Engagement Day *Smithfield, RI*   April 2018
Keynote Address: *Why Data Science? Why Not? You're Already Doing It*

Joint Statistical Meetings (JSM 2017) Baltimore, MD   August 2017
Lessons Learned in Transitioning from "Intro to Statistics" to "Reasoning with Data"

U.S. Conference on Teaching Statistics State College, PA   May 2017
Opening Address speaker

TEDx CMU: Pivot Pittsburgh, PA   April 2017
Embracing Your Inner Data Scientist

National Academy of Science, Engineering, and Medicine Washington, D.C.   December 2016
Workshop on "Envisioning the Data Science Discipline: The Undergraduate Perspective"
If You Build It, They Will Come: Perspectives on an Undergraduate Statistics/Data Science Program

NSF-Census Research Network/National Academy of Sciences Washington, D.C.   May 2015
Building and Training the Next Generation of Survey Methodologists and Researchers

Model-Based Clustering Working Group Dublin, Ireland (Short Talk)   July 2014
To Merge or Not to Merge: Interactive Visualization Tool for Local Merges of Mixture Model Components

European Conference on Data Analysis/GfKl Annual Meetings Bremen, Germany   July 2014
Keynote Address: *Solving the Identity Crisis: Large-Scale Clustering with Distributions of Distances*
and Applications in Record Linkage

ENAR: International Biometric Society Baltimore, MD   March 2014
Massive Online Open Statistics: Should We Be Teaching Statistics to 100s of Thousands of Students?

International Federation of Classification Societies Meeting (IFCS 2013) Tilburg, Netherlands   July 2013
Deduplicating Text Records in "Big Data": Scaling Up with Aggregated Classifiers and
Blocked Hierarchical Clustering

Classification Society Annual Meetings (CS 2013) Milwaukee, WI   June 2013
Presidential Address: *Mixture Model Component Trees: A Tool for Merging Clusters (and much more!)*

Bucknell University, Department of Mathematics Lewisburg, PA   October 2012
Distinguished Visiting Professor: *Solving the Identity Crisis: Inventor Disambiguation*
using a Forest of Random Forests

Bucknell University, Department of Mathematics, Distinguished Visiting Professor, Lewisburg, PA October 2012
How/What/When Do We Learn? Providing Real-Time Cognitive Diagnostic Feedback
in the Assistment Project Using Clustering Methods

Joint Statistical Meetings - JSM 2012 San Diego, CA   August 2012
Maintaining Quality in the Face of Rapid Program Expansion

Classification Society Annual Meetings (CS 2012) Pittsburgh, PA   June 2012
Presidential Address: *A Self-Tuning Diffusion Map Framework for Use in Document Clustering*

International Federation of Classification Societies 2011 Meetings Frankfurt, Germany   August 2011
Estimating and Visualizing Cluster Structure in a Constrained Hypercube as a Proxy
for Cognitive Diagnosis Models

International Statistical Institute 2011 Meetings Dublin, Ireland   August 2011
Using Self-Tuning Diffusion Maps to Reduce Dimensionality and Find Cluster Structure

Model-based Clustering Working Group (Short Talk), Glasgow, Scotland   July 2011
Model-Based Clustering in a Diffusion Map Framework

International Classification Conference 2011 St Andrews, Scotland   July 2011
Clustering Students by Their Skill Set Profiles on the Unit Hypercube

City University of New York Graduate School (Honorarium Recipient)   June 2011
An Overview of Assistments:Providing Diagnostic Feedback using Modern Clustering Methods

Model-based Clustering Working Group (Short Talk) Grenoble, France   July 2010
Estimating a Mixture of Latent Class Trajectories in the Presence of Missing/Censored Observations

International Conference on Educational Data Mining (EDM 2010) Pittsburgh, PA   June 2010
Skill Set Profile Clustering: The Empty K-Means Algorithm with Automatic Specification Starting Cluster Centers

American Educational Research Association Meetings (AERA 2010) Denver, CO   May 2010
Providing Diagnostic Feedback using Modern Clustering Methods in the Assistment Project

International Meeting of the Psychometric Society - IMPS09 Cambridge, UK   July 2009
Finding and Visualizing Hierarchical Cluster Structure as an Alternative to Cognitive Diagnosis Models

Working Group on Model-Based Clustering (Long Talk) Paris, France   July 2009
Mixture Model Component Trees: Visualizing the Hierarchical Component Structure of Complex Groups

SAMSI Cognitive Diagnosis Modeling Working Group Research Triangle Park, NC   July 2009
Clustering in Cognitive Diagnosis: Some Recent Work and Open Questions

International Symposium on Business and Industrial Statistics Prague, Czech Republic   July 2008
Finding and Visualizing the Hierarchical Cluster Structure in Data

Classification Society of North America 2008 Meetings St Louis, MO   June 2008
Visualizing Hierarchical Cluster Structure via Linkage Algorithms

Southern Society of General Internal Medicine Meetings New Orleans, LA   February 2008
Symposium: *A Review of Clustering, Classification, and Regression Analysis in Biomedical Research*

Working Group on Model-Based Clustering, Trinity College (Short Talk) Dublin, Ireland   July 2007
Visualization of Clusters Via a Density-Based Similarity Measure

Joint Statistical Meetings - ASA Section on Statistical Computing Seattle, WA   August 2006
Generalized Single Linkage Clustering; Session: Density-based Clustering
*Note: ASA rules prohibited giving this talk (also won a Student Award);
Werner Stuetzle presented this work in my place*

Texas Tech Health Sciences Center, Department of Internal Medicine Lubbock, TX   March 2005
*A Statistical Overview of "Combination of Isosorbide Dinitrate and
Hydralazine in Blacks with Heart Failure"*

Texas Tech Health Sciences Center, Department of Internal Medicine Lubbock, TX   June 2004
The Metabolic Syndrome and Chronic Kidney Disease in U.S. Adults: Statistical Analysis

Department of Biostatistics, University of Washington, Student Seminar, Seattle   January 2004
Clustering With Confidence: Overview of Current Clustering Methods - How Well Do They Do?

Invited Departmental Seminars and Lectures *(slated/upcoming in italics)*

Texas Tech Health Sciences Center Grand Rounds,   Sept 2023
Talk, *Use of Non-Traditional Data Sources to Improve Population Health with Applications to COVID-19*

Central Michigan University Data Science   April 2023
Data, huh, Yeah. What is it good for? Absolutely Everything!

Williams College, Department of Mathematics & Statistics   September 2020
Demystifying Data Science

University of Virginia, School of Data Science   July 2020
Instead of Just Teaching Data Science, Let's Also Understand How and Why People Do It

University of Washington Center for Statistics & Social Sciences, Virtual   April 2020
Instead of Just Teaching Data Science, Let's Understand How and Why People Do It

University of Texas Department of Statistics Austin, TX   September 2019
Instead of Just Teaching Data Science, Let's Understand How and Why People Do It

Michigan State University, Department of Statistics and Probability East Lansing, Michigan   March 2019
Before Teaching Data Science, Let's First Understand How People Do It

Binghamton University Department of Mathematical Sciences New York   March 2019
Before Teaching Data Science, Let's First Understand How People Do It

The Ohio State University Department of Statistics Columbus, Ohio   November 2018
Before Teaching Data Science, Let's First Understand How People Do It

Rice University Department of Statistics Houston, TX   April 2017
Seminar: *Harnessing the Data Science (R)evolution*

Rice University Department of Statistics Houston, TX — March 2015
Seminar: *Solving the Identity Crisis: Large-Scale Clustering with Distributions of Distances and Applications in Record Linkage*

Indiana University Department of Statistics Bloomington, IN — September 2014
Seminar: *Solving the Identity Crisis: Large-Scale Clustering with Distributions of Distances and Applications in Record Linkage*

Indiana University Department of Statistics Bloomington, IN — August 2014
Honorarium Recipient (undergraduate program consultation): *Choose-Your-Own Capstone Adventure: Providing Flexible Paths for Undergraduate Majors*

Bucknell University, Department of Mathematics, Undergraduate Seminar Series, Lewisburg, PA — April 2014
The Next Big Sound? Modeling the Path from Relative Obscurity to the Top of the Charts using Social Metrics

Johns Hopkins University, Department of Biostatistics Baltimore, MD — October 2012
GaPMM: Modeling Mixtures of Trajectory Classes in the Presence of Informative Missingness

U of Guelph Dept of Mathematics & Statistics Ontario, Canada — November 2011
A Self-Tuning Diffusion Map Framework for Use in Document Clustering

U of Chicago Booth School of Business Chicago, IL — September 2011
A Self-Tuning Diffusion Map Framework for Use in Document Clustering

Center for Statistics & Social Sciences, University of Washington Seattle, WA — January 2010
Mixture Model Component Trees as an Alternative to High-Dimensional Cognitive Diagnosis Models

Department of Statistics, Iowa State University Ames, IA — October 2009
Mixture Model Component Trees: Finding and Visualizing Hierarchical Component Structure (with Applications in Cognitive Diagnosis)

RAND Statistics Seminar Series, Pittsburgh, PA — November 2008
Visualizing Hierarchical Cluster Structure via Density-Based Linkage Methods and Component Trees

Department of Statistics, University of Glasgow, Scotland — August 2008
Cluster Tree Estimation using a Generalized Single Linkage Method (and Extensions)

Le Colloque de Statistique de Montreal - joint seminar of Montreal statistics departments — November 2007
Invited by Department of Mathematics and Statistics, University of Montreal
Visualization of Clusters Via a Density-Based Similarity Measure

Department of Mathematics, Grinnell College Grinnell, IA — March 2007
Lecture: *Intro to Clustering*

Department of Statistics, University of Pittsburgh — November 2006
Lecture: *Overview of Clustering Gene Expression Data*

Department of Statistics, Carnegie Mellon University Pittsburgh, PA — March 2006
Clustering with Confidence

Department of Statistics, University of Pittsburgh, PA — February 2006
Clustering with Confidence

Interviews/Panels

Pittsburgh Bar Institute: Employment Law Institute West 2022, *Discrimination by Algorithm* Panel, November 2022

BW meets Pittsburgh: Market and cooperation opportunities - a joint perspective on the application of AI in the vehicle and health care sectors, *AI in Health Care* Panel, October 2022

Pittsburgh Digital Therapeutics Forum: (Echo Strategies), May 2021

goto; book club: Interview of Phil Winder, author of *Reinforcement Learning: Industrial Applications of Intelligent Agents*, https://gotopia.tech/bookclub; Winter 2021

Corporate Startup Lab 2020 Forum, Carnegie Mellon Swartz Center: Panel on Data-based Corporate Startups, November 2020

Carnegie Mellon Women's Association: Leadership Panel, November 2020

Carnegie Mellon Futures Summit/Make Possible Campaign: Speaker, Panel; November 2019 (Pittsburgh), February 2020 (Silicon Valley/San Francisco), March 2020 (NYC)

Carnegie Mellon Mortar Board series: *Life Advice and Wisdom...Don't Do What I Did*, February 2014

ASA Webinar: *The Role and Variety of Undergraduate Statistics Capstones*, with C Shalizi, December 2013

Simply Statistics (Authors: Jeff Leek, Roger Peng, Rafa Irizarry; Johns Hopkins Biostatistics). *http://simplystatistics.org/post/33892683645/interview-with-rebecca-nugent-of-carnegie-mellon.* October 2012.

Advising/Supervising

Postdoctoral Advisor

Philipp Burckhardt, Department of Statistics & Data Science, Carnegie Mellon University   2019-2022
ISLE: Integrated Statistics Learning Environment

Ph.D. Advisor

Xiaoyi Yang, Department of Statistics & Data Science, Carnegie Mellon University   August 2021
Learning social networks from text data
applications in social sciences, digital humanities; co-advised with Nynke Niezink

Kayla Frisoli, Department of Statistics & Data Science, Carnegie Mellon University   December 2020
Exploring the impact of household structure and user-driven labels on the linkage
of early 1900s Irish census records

Brendan McVeigh, Department of Statistics & Data Science, Carnegie Mellon University   May 2020
A Sequential Algorithm for Bayesian Inference of Large-Scale Record Linkage Structure
primary advisor: Jared Murray, Dept of Information, Risk & Operations Mgmt, UT Austin

Samuel Ventura, Department of Statistics, Carnegie Mellon University   Summer 2015
Large Scale Classification and Clustering Methods with Applications in Record Linkage

Ph.D. External Examiner

Katherine Morris, Department of Mathematics and Statistics, University of Guelph   January 2014
Dimension Reduction and Clustering using Non-Elliptical Mixtures

Arthur White, School of Mathematical Sciences (Statistics), University College Dublin   December 2013
Clustering Multivariate and Network Data Using Mixture and Mixed Membership Models

Ph.D. Thesis Committees

Purvasha Chakravarti, Department of Statistics & Data Science, Carnegie Mellon University   May 2020
Inference for Clustering and Anomaly Detection

Yotam Hechtlinger, Department of Statistics & Data Science, Carnegie Mellon University   December 2019
Topics in Prediction

Philipp Burkhardt, Department of Statistics & Data Science, Carnegie Mellon University   May 2019
Learning Data Science

Amanda Luby, Department of Statistics & Data Science, Carnegie Mellon University   July 2019
Accounting for individual differences among decision-makers with applications
to the evaluation of forensic evidence

Xiao Hui Tai, Department of Statistics & Data Science, Carnegie Mellon University   June 2019
Forensic Data Matching Problems

David Luke Oates, Department of Engineering & Public Policy, Carnegie Mellon University   February 2015
Low Carbon Electricity and the Fossil Fleet

Mauricio Sadinle-Garcia, Department of Statistics, Carnegie Mellon University   January 2015
A Bayesian Framework for Duplicate Detection and Record Linkage

Chia-Hsuan Yang, Department of Engineering and Public Policy, Carnegie Mellon University   May 2014
Challenges for Emerging Innovation: Offshoring Manufacturing, Inventor Mobility
and Technology Trajectories in U.S. Optoelectronic Components for Telecommunications

Saeed Amizadeh, Intelligent Systems Program, University of Pittsburgh   Sept 2013
Graph-Based Methods for Large Scale Problems

Casey Studer, Department of Statistics, Carnegie Mellon University *Incorporating Learning Over Time into the Cognitive Assessment Framework*	August 2012
Gabrielle Flynt, Department of Statistics, Carnegie Mellon University *Clustering Longitudinal Trajectories in the Presence of Informative Monotone Missingness*	May 2012
Joseph Richards, Department of Statistics, Carnegie Mellon University *Fast and Accurate Estimation for Astrophysical Problems in Large Databases*	July 2010
David Friedenberg, Department of Statistics, Carnegie Mellon University *Adaptive Cluster Detection*	May 2010
Elizabeth Ayers, Department of Statistics, Carnegie Mellon University *Predicting Performance and Scaling Up Estimates of Student Skill Knowledge*	May 2010

PhD Advanced Data Analysis Project Advisor

Ron Yurko, Department of Statistics & Data Science, Carnegie Mellon University *Modeling Understanding of the Data Analysis Workflow through Written Work*, affiliated with the revamp of the Dietrich College General Education Program	2018-2019
Alan Mishler, Department of Statistics, Carnegie Mellon University *Online Learning for Introductory French*, co-advised by Bonnie Youngs; Modern Languages, CMU	2017-2018
Xiaoyi Yang, Department of Statistics, Carnegie Mellon University *Historical Record Linkage*, co-advised by Jared Murray; Information, Risk & Oper Mgmt, UT Austin	2017-2018
Sam Ventura, Department of Statistics, Carnegie Mellon University *Disambiguating USPTO Inventors with Forest of Random Forests*, co-advised by Erica Fuchs, EPP, CMU	2013

Undergraduate Honors Examiner

Swarthmore College, Department of Mathematics & Statistics Honors examiner in Statistics	May 2018, 2019

Undergraduate Honors Thesis Advisor

Carlo Duffy, Department of Statistics & Data Science, Carnegie Mellon University *Association between Prescription Propensity and Medicare Patient Panels' Mean HCC Risk Scores* co-advised, Mark Patterson, Social & Decision Sciences/QSSS	2020-2021
Yeuk Yu (Tiffany) Lee, Department of Statistics, Carnegie Mellon University *User-Driven Visualization Tools* (co-advised, Robin Mejia, Center for Human Rights Science)	2017-2018
Josh Ragen, Department of Statistics, Carnegie Mellon University *Risk Assessment Tools for Domestic Violence Recidivism* (co-advised, Mark Patterson, QSSS)	2017-2018
Theo Peterson, Department of Statistics, Carnegie Mellon University *Modeling the Threat of Domestic Violence via Hotline Calls* (co-advised, Mark Patterson, QSSS)	2016-2017
Joseph Pane, Department of Statistics, Carnegie Mellon University *Hitting the Wall: Mixture Models of Long Distance Running Trajectories*	2014-2015
Emily Wright, Department of Statistics, Carnegie Mellon University *Music Artist Discovery: The Digital Road to the Top of Radio*	2013-2014
Yi Xiang Chong, Department of Mathematics and Statistics, Carnegie Mellon University *Incorporating Flexibility into the Normalized-Cut Image Segmentation Algorithm*	2011-2012
Christopher Peter Makris, Department of Statistics, Carnegie Mellon University *Exploration of Imputation Methods for Missingness in Image Segmentation*	2010-2011
Andrew Klein, Departments of Information Systems, Statistics, Carnegie Mellon University *Semi-automated Collection of Pitch Location and Intent in Baseballs* (co-advised, Andrew Thomas)	2010-2011

University and Departmental Service

Carnegie Mellon University:

Presidential Advisory Board for Enterprise Creation, Entrepreneurship, and Industry Engagement, President-appointed	2023-present

Co-Chair, Academic Calendar Advisory Board, Provost-appointed	2023-present
Co-Chair, Academic Calendar Innovation Committee, Provost-appointed	2020-2023
Co-Chair, Data & Information Literacy Core Competency, Vice Provost-appointed	2021-2022
Carnegie Mellon NCAA Faculty Athletics Representative	2021 - present
Search Committee, Vice President for Communications & Marketing	2021
Dietrich College of Humanities and Social Sciences, Promotion & Tenure Committee	2020-2021
Chair, University Libraries Search for Associate Dean for Innovation, Preservation, and Access	2018
Faculty Senate University Libraries Advisory Committee	2016-2019
Chair	Summer 2019
Advisory Board - digital Sciences, Humanities, and Arts: Research and Publication	2018 - 2019
dSHARP center co-sponsored by University Libraries and Dietrich College	
Carnegie Mellon 2018 Middle States Accreditation Self-Study Committee	2016 - 2018
Dietrich College General Education Requirements Redesign Committee	2015 - 2018
Dietrich College Assistant Dean/Head of Academic Advisory Center Search Committee	Spring 2015
Undergraduate Research Office Grant/Fellowship Committee	Fall 2010 - present
Responsible for the yearly disbursement of undergraduate research funding	
Carnegie Mellon Ryan Award for Meritorious Teaching Selection Committee	2015-2018
Dietrich College Elliott Dunlap Smith Award Selection Committee	2013-2015
Andrew Carnegie Society Scholar Selection Committee	2011-2016
Senior Capstone Seminar - facilitator; discussant	Fall 2008

Department of Statistics (& Data Science):

Department Head	June 2021 - present
Associate Department Head	Spring 2017 - Summer 2021
Founding Director, Corporate Capstone Program	Spring 2018 - present
(co-)Director of Undergraduate Studies	2014 - 2021
Education Programs Administration	2012-2021
oversee scheduling, faculty teaching assignments, TA hiring and training, etc	
Chair, Teaching Track Faculty Search Committee	2013-2014, 2018-2019
(Interim) Director of Undergraduate Studies	Fall 2010 - Summer 2011
Undergraduate Curriculum Committee	Fall 2009 - 2021
(co-)Chair	2010-2011, 2014 - 2021
Undergraduate Research	Fall 2009 - 2021
Master's Applied Data Analysis Qualifier Committee	2011, 2012
CMART Management Committee	2010 - 2014
Carnegie Mellon & RAND Traineeships in Methodology and Interdisciplinary Research program	
Master's in Statistical Practice Graduate Program Committee	Fall 2008 - Summer 2011
program development, admissions, etc	
Teaching Teas - chair; organizer of colloquia focusing on departmental teaching	Fall 2007 - Summer 2008
Seminar Committee; co-organizer of weekly seminars	Fall 2006 - Fall 2007

University of Washington:

Unsupervised Learning Working Group	August 2003 - June 2006
Co-organizer of weekly research working group	
http://www.stat.cmu.edu/~rnugent (under Research)	
Statistics Tutor & Study Center	Spring 2003 - Winter 2005
Department of Statistics, University of Washington, Seattle, WA	
Co-founder and student director of pilot program providing free statistics tutoring	
and study space; employed graduate students to tutor introductory statistics courses	
Center is still in operation. http://www.stat.washington.edu/tutorcenter	

Professional Service

Associations:

American Statistical Association
Section on Statistics & Data Science Education

President-Elect (elected but declined due to new Dept Head role)	2020
Executive Committee	2017-2019

Section on Statistical Graphics

Council of Sections Representative	2014-2016
Judge for John M Chambers Statistical Software Award	2014
Publications Officer	2012-2014

Pittsburgh ASA Chapter

Past-President/Board Member	2012-2013
President	2011-2012

International Federation of Classification Societies

President-Elect	2020

Classification Society

President	Jan 2012-Jan 2014
Co-organizer, Classification Society 2011, 2012 Annual Meetings	2010-2012
Representative to the International Federation of Classification Societies	2010-2012
Board of Directors	Jan 2008-Jan 2010
Dissertation Award Committee	2009 - 2011

Journals & Proceedings:

Co-editor, **Springer Texts in Statistics** with Genevera Allen and Richard DeVeaux	2019-present
Co-editor, **International Federation of Classification Societies 2019 Proceedings**, Springer	2019-2020

Journal of Classification

Chair, Editor Search Committee	2014
Journal of Computational & Graphical Statistics, Associate Editor	Fall 2011- Fall 2016
The Southwest Respiratory and Critical Care Chronicles, Editorial Board	Fall 2012- present
Journal of Educational Data Mining, Guest Co-editor, Special Issue	2012
AISTAT 2009: 12th International Conference on Artificial Intelligence and Statistics, Area Chair	2009

Reviewer
Panelist: National Science Foundation Division of Graduate Education, Division of Mathematical Sciences

Grants: National Science Foundation, Engineering and Physical Sciences Research Council (EPSRC - UK),
National Sciences and Engineering Research Council of Canada (NSERC)

Journals: Statistics in Medicine, JRSS-B, Psychometrika, World Scientific Publishing, Journal of Computational
and Graphical Statistics, Statistics and Computing, Journal of Machine Learning Research, AISTATS (2010, 2012, 2014),
EURASIP IVP, Journal of Classification, Computational Statistics and Data Analysis, Statistical Analysis
and Data Mining, International Federation of Classification Societies Proceedings (2011), Pattern Recognition,
Journal of Educational and Behavioral Statistics, Teaching and Learning in Medicine, Annals of Applied Statistics,
Nature: Methods, Transactions on Knowledge and Data Engineering, Journal of Educational Data Mining,
Journal of Statistical Software, Journal of the American Statistical Association, Annals of Statistics, Management Science
Journal of Statistics Education, The American Statistician, Demographic Research, NIPS (2014, 2016), WIRES

Professional Societies:

Institute of Mathematical Statistics, American Statistical Association, Classification Society (of North America),
Interface Foundation of North America, Psychometric Society, International Educational Data Mining Society

Conferences:

Women in Data Science Conference, Chair, Program Committee; Conference Organizer	2018-present
http://www.stat.cmu.edu/wids	
ASA Symposium on Data Science and Statistics, Program Committee	2020

electronic Conference on Teaching Statistics (eCOTS 2018), Program Chair   May 2018

Carnegie Mellon Sports Analytics Conference, Chair, Program Committee   2017-present

ICDM, the 13th International Conference on Data Mining   2013
Program Committee for Workshop IClaNov

Young Statistician's Meeting (YSI) - Satellite of the 58th ISI World Congress Meeting   2011
Scientific Program Committee

International Federation of Classification Societies Meetings: Scientific Committee   2010-2011, 2012-2013

University of Washington Working Group on Model-Based Clustering:Organizing Committee Summer 2011

Interface Foundation of North America 2010 Meetings: Organization Committee   2009-2010

Languages

English (native language), Spanish (proficient, both verbal and written)

Community Service

Carnegie Mellon Athletics Department   2007-2017
Head Coach, Women's Ultimate Frisbee
National Certification Level I
Nominated Ohio Valley Region Coach of the Year   2012-2013, 2013-2014
Nominated CMU Outstanding Student Organization Advisor   2015

Seattle Girls' School   (Fall only) 2002 - 2005
Head Cross Country Coach: 5th-8th grades
undefeated for four years; Service Award from the SGS Board (2005)

Girls on the Run   Spring 2003 - Summer 2006
Race Coordinator and Volunteer; Coach for the Spring 2003 Program
Mentor/Running partner for girls age 9-13