

Envisioning the Data Science Discipline: The Undergraduate Perspective

Committee

Laura M. Haas

Co-Chair

Laura Haas (NAE) is an IBM Fellow and Director of IBM Research's Accelerated Discovery Lab. She was the director of computer science at IBM Almaden Research Center from 2005-2011, and had worldwide responsibility for IBM Research's exploratory science program from 2009 through 2013. Previously, she was senior manager for Information Integration Solutions Architecture and Development in IBM's Software Group. Dr. Haas was one of the founders of Information Integration Solutions, which today includes InfoSphere Information Server and a suite of products that can integrate both structured and unstructured data via federation and materialization. Dr. Haas managed the development team in its first two years, moving to lead the architecture team for the product suite with the acquisition of Ascential Software in 2005. Before joining Information Integration, she managed the DB2 UDB Query Compiler development, including key technologies for information integration and the Life Sciences industry, such as federated database and XML Query. Dr. Haas spent the first twenty years of her career as a research staff member and manager at IBM's Almaden Research Center in San Jose. Dr. Haas joined IBM in 1981 as a research staff member. Before joining IBM, Dr. Haas studied Applied Mathematics and Computer Science at Harvard University, and Computer Science at the University of Texas at Austin, where she received her PhD in 1981. Dr. Haas was vice-chair of the Association of Computing Machinery (ACM) Special Interest Group on the Management of Data (SIGMOD) from 1989-1997. She has served as an Associate Editor of the ACM journal Transactions on Database Systems, as Program Chair of the 1998 ACM SIGMOD technical conference and the IIS track of the 2005 VLDB conference, as general co-chair of VLDB 2008, and as Vice President of the VLDB Endowment Board of Trustees (from 2004-2009). She received an IBM Corporate Award for her work on federated database technology, IBM awards for Outstanding Technical Achievement for her work on R* and on DiscoveryLink, an IBM Outstanding Contribution award for Starburst, an Outstanding Innovation Award for Clio, a YWCA Tribute to Women in INdustry (TWIN) award, the Anita Borg Institute's Technical Leadership Award, and the ACM SIGMOD Codd Innovation Award. Dr. Haas is an ACM Fellow and a Fellow of the American Academy of Arts and Sciences, a member of the National Academy of Engineering, on the NRC's Computer Science and Telecommunications Board (CSTB) and past Vice-Chair of the board of the Computing Research Association.

Alfred O. Hero, III

Co-Chair

Alfred O. Hero III is the R. Jamison and Betty Williams Professor of Engineering at the University of Michigan. He received the BS (summa cum laude) from Boston University (1980) and PhD from Princeton University (1984), both in electrical engineering. His primary appointment is in the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. In 2008 he was awarded the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, located at the Ecole Supérieure d'Electricité, Gif-sur-Yvette, France. He is an Institute of Electrical and Electronics Engineers (IEEE) Fellow and several of his research articles have received best paper awards. Professor Hero was awarded the University of Michigan Distinguished Faculty Achievement Award (2011). He received the IEEE Signal Processing Society Meritorious Service Award (1998) and the IEEE Third Millennium Medal (2000). He was president of the IEEE Signal Processing Society (2006-2008) and was on the Board of Directors of the IEEE (2009-2011) where he served as Director of Division IX (Signals and Applications). Dr. Hero's recent research interests have been in detection, classification, pattern analysis, and adaptive sampling for spatio-temporal data. Of particular interest are applications to network security, multimodal sensing and tracking, biomedical imaging, and genomic signal processing.

Ani Adhikari

Member

Ani Adhikari is a Senior Lecturer in Statistics at UC Berkeley, has received the Distinguished Teaching Award at Berkeley and the Dean's Award for Distinguished Teaching at Stanford University. While her research interests are centered on applications of statistics in the natural sciences, her primary focus has always been on teaching and mentoring students. She teaches courses at all levels and has a particular affinity for teaching statistics to students who have little mathematical preparation. She received her undergraduate degree from the Indian Statistical Institute, and her Ph.D. in Statistics from Berkeley.

David E. Culler

Member

David Culler (NAE) received his B.A. from UC Berkeley in 1980, and an M.S. and Ph.D. from MIT in 1985 and 1989, respectively. He joined the EECS faculty in 1989 and is the founding Director of Intel Research, UC Berkeley and was Associate Chair of the EECS Department, 2010-2012 and Chair from 2012 through June 30, 2014.

He won the Okawa Prize in 2013. He is a member of the National Academy of Engineering, an ACM Fellow, and an IEEE Fellow. He has been named one of Scientific American's Top 50 Researchers and the creator of one of MIT's Technology Review's 10 Technologies that Will Change the World. He was awarded the NSF Presidential Young Investigator and the Presidential Faculty Fellowship.

His research addresses networks of small, embedded wireless devices, planetary-scale internet services, parallel computer architecture, parallel programming languages, and high performance communication. It includes TinyOS, Berkeley Motes, PlanetLab, Networks of Workstations (NOW), Internet services, Active Messages, Split-C, and the Threaded Abstract Machine (TAM).

David L. Donoho

Member

David Donoho (NAS) is an Anne T. and Robert M. Bass Professor of Humanities and Sciences and Professor of Statistics at Stanford University. Dr. Donoho is a mathematician who has made fundamental contributions to theoretical and computational statistics, as well as to signal processing and harmonic analysis. His algorithms have contributed significantly to the understanding of the maximum entropy principle, of the structure of robust procedures, and of sparse data description. His theoretical research interests have focused on the mathematics of statistical inference and on theoretical questions arising in applying harmonic analysis to various applied problems. His applied research interests have ranged from data visualization to various problems in scientific signal processing, image processing, and inverse problems. Donoho received an A.B. from Princeton University and a Ph.D. from Harvard University.

E. Thomas Ewing

Member

E. Thomas Ewing is an Associate Dean for Graduate Studies, Research, and Diversity in the College of Liberal Arts and Human Sciences and a professor in the Department of History. His education included a BA from Williams College and a PhD in history from the University of Michigan. He teaches courses in Russian, European, Middle Eastern, and world history, gender / women's history, and historical methods. His publications include, as author, *Separate Schools: Gender, Policy, and Practice in the Postwar Soviet Union* (2010) and *The Teachers of Stalinism. Policy, Practice, and Power in Soviet Schools in the 1930s* (2002); as editor, *Revolution and Pedagogy. Transnational Perspectives on the Social Foundations of Education* (2005); and as co-editor, with David Hicks, *Education and the Great Depression. Lessons from a Global History* (2006). His articles on Stalinist education have been published in *Gender & History*, *American Educational Research Journal*, *Women's History Review*, *History of Education Quarterly*, *Russian Review*, and *The Journal of Women's History*. He has received funding from the National Endowment for the Humanities, the Spencer Foundation, and the National Council for Eurasian and East European Research.

Louis J. Gross

Member

Louis J. Gross is a James R. Cox and Alvin and Sally Beaman Distinguished Professor of Ecology and Evolutionary Biology and Mathematics and Director of The Institute for Environmental Modeling at The University of Tennessee, Knoxville. He is also Director Emeritus of the National Institute for Mathematical and Biological Synthesis, a National Science Foundation-funded center to foster research and education at the interface between math and biology. He completed a B.S. degree in Mathematics at Drexel University and a Ph.D. in Applied Mathematics at Cornell University, and has been a faculty member at UTK since 1979. His research focuses on applications of mathematics and computational methods in many areas of ecology, including disease ecology, landscape ecology, spatial control for natural resource management, photosynthetic dynamics, and the development of quantitative curricula for life science undergraduates. He led the effort at UT to develop an across trophic level modeling framework to assess the biotic impacts of alternative water planning for the Everglades of Florida. He has co-directed several Courses and Workshops in Mathematical Ecology at the International Centre for Theoretical Physics in Trieste, Italy, served as Program Chair of the Ecological Society of America, as President of the Society for Mathematical Biology, President of the UTK Faculty Senate, Treasurer for the American Institute of Biological Sciences and as Chair of the National Research Council Committee on Education in Biocomplexity Research. He is the 2006 Distinguished Scientist awardee of the American Institute of Biological Sciences and is a Fellow of the American Association for the Advancement of Science. He has served on the National Research Council Board on Life Sciences and was liaison to the NRC Standing Committee on Emerging Science for Environmental Health Decisions. He is a long-time volunteer for Jubilee Community Arts and Community Shares, hosted and produced folk music programs for WUOT-FM, performs with the Lark in the Morn English Country Dancers and serves as House Sound Engineer for concerts at the Laurel Theatre in Knoxville.

Nicholas Horton

Member

Nicholas Horton is a Professor of Statistics at Amherst College. He has taught a variety of courses in statistics and related fields, including probability, mathematical statistics, regression and design of experiments and is passionate about improving quantitative literacy for students with a variety of backgrounds as well as engagement and mastery of higher-level concepts and capacities to undertake research. He is the Chair of the Committee of Presidents of Statistical Societies and has published more than 150 papers in statistics and biomedical research and four books on statistical computing and data science. He has been the recipient of a number of teaching awards. As an applied biostatistician, Dr. Horton's work is based squarely within the mathematical sciences, but spans other fields in order to ensure that research is conducted on a sound footing. The real-world research problems that these investigators face often require the use of novel solutions and approaches, since existing methodology is sometimes inadequate. Bridging the gap between theory and practice in interdisciplinary settings is often a challenge, and has been a particular focus of Dr. Horton's work. Dr. Horton earned his A.B. from Harvard College and my Sc.D. in biostatistics from the Harvard School of Public Health.

Julia I. Lane

Member

Dr. Julia Lane is a Professor at CUSP, and at NYU's Wagner Graduate School of Public Service. She also serves as a Provostial Fellow for Innovation Analytics and Senior Fellow at NYU's GovLab. As part of the CUSP team, Dr. Lane works with the research team to build the CUSP Data User Facility.

Dr. Lane is an economist who is the co-founder of the Longitudinal Employer-Household Dynamic (LEHD) partnership with the Census Bureau. LEHD data has been used to analyze commuting patterns for transportation planning, and the study of workforce turnover, pensions, and low-wage work. Dr. Lane has authored over 65 refereed articles and edited or authored seven books. She has been working with a number of national governments to document the results of their science investments. Her work has been featured in several publications including Science and Nature. Work Dr. Lane started at the National Science Foundation (as Senior Program Director of the Science of Science and Innovation Policy Program) to quantify the results of federal stimulus spending is the basis of the new Institute for Research on Innovation and Science at the University of Michigan. The data will be used to describe the structure of the research workforce, the nature and evolution of research collaborations, and the diffusion of sponsored research results.

Dr. Lane has had leadership positions in a number of policy and data science initiatives at her other previous appointments, which include Senior Managing Economist at the American Institutes for Research; Senior Vice President and Director, Economics Department at NORC/University of Chicago; various consultancy roles at The World Bank; and Assistant, Associate and Full Professor at American University. Please [click here](#) for additional information on her professional achievements.

Dr. Lane received her PhD in Economics and Master's in Statistics from the University of Missouri

Andrew McCallum

Member

Andrew McCallum is a Professor and Director of Center for Data Science, as well as the Information Extraction and Synthesis Laboratory, in the College of Information and Computer Science at University of Massachusetts Amherst. He has published over 250 papers in many areas of AI, including natural language processing, machine learning and reinforcement learning; his work has received over 45,000 citations. He obtained his PhD from University of Rochester in 1995 with Dana Ballard and a postdoctoral fellowship from CMU with Tom Mitchell and Sebastian Thrun. In the early 2000's he was Vice President of Research and Development at at WhizBang Labs, a 170-person start-up company that used machine learning for information extraction from the Web. He is a AAAI Fellow, the recipient of the UMass Chancellor's Award for Research and Creative Activity, the UMass NSM Distinguished Research Award, the UMass Lilly Teaching Fellowship, and research awards from Google, IBM, Microsoft, and Yahoo. He was the General Chair for the International Conference on Machine Learning (ICML) 2012, and is the current President of the International Machine Learning Society, as well as member of the editorial board of the Journal of Machine Learning Research. For the past ten years, McCallum has been active in research on statistical machine learning applied to text, especially information extraction, entity resolution, social network analysis, structured prediction, semi-supervised learning, and deep neural networks for knowledge representation.

Richard D. McCullough

Member

Richard McCullough has a B.S. in chemistry from the University of Texas, Dallas, earned his M.A. and Ph.D. in chemistry at Johns Hopkins University. He did his postdoctoral fellowship at Columbia University. Since 2012, Dr. McCullough has been the Vice Provost for Research, working with the President and Provost to encourage, cultivate, and coordinate high impact academic research across all of Harvard's schools and affiliated institutions. The Office of the Vice Provost for Research (VPR) has broad responsibility and oversight for the development, review, and implementation of strategies, planning, and policies related to the organization and execution of academic research across the entire university. Dr. McCullough leads a new office of Foundation and Corporate Development. He also assists in oversight of many of the interdisciplinary institutes, centers and initiatives across Harvard. Under Vice Provost McCullough's leadership, the Office of the VPR is particularly focused on removing barriers to collaboration, whether in University policies, or financial or administrative systems. Additionally, the Vice Provost for Research works with the President and Provost to foster and encourage entrepreneurship among undergraduates, graduate students, and faculty members. He also helps to lead the development of the new innovation campus. Richard McCullough is also a Professor of Materials Science and Engineering at Harvard and is a member of numerous professional societies and boards. Prior to being named Vice Provost for Research at Harvard, Dr. McCullough was the Vice President for Research at Carnegie Mellon University in Pittsburgh, where he previously served as the Dean of the Mellon College of Science, and Professor and head of the Department of Chemistry. Dr. McCullough has founded two companies: Plextronics Inc, and Liquid X Printed Metals.

Rebecca Nugent

Member

Rebecca Nugent is a teaching professor in the Department of Statistics at Carnegie Mellon University and has been teaching at CMU since she completed her Ph.D. in Statistics from University of Washington in 2006. Prior to that, she received her B.A with Majors in Mathematics, Statistics and Spanish at Rice University and her M.S. in Statistics at Stanford University. She recently was awarded top teaching honors with the American Statistical Association Waller Education Award; The William H. and Frances S. Ryan Award for Meritorious Teaching; and Statistician of the Year by the American Statistical Association Pittsburgh Chapter. Rebecca's research interests lie in clustering, record linkage, educational data mining/psychometrics, public health, tech/innovation/entrepreneurship and semantic organization.

Lee Rainie

Member

Lee Rainie is the director of internet, science and technology research at Pew Research Center. Under his leadership, the Center has issued more than 500 reports based on its surveys that examine people's online activities and the internet's role in their lives. He also directs the Center's new initiative on the intersection of science and society. The American Sociological Association gave Dr. Rainie its award for "excellence in the reporting on social issues" in 2014 and described his work as the "most authoritative source of reliable data on the use and impact of the internet and mobile connectivity." Dr. Rainie is a co-author of *Networked: The new social operating system* and five books about the future of the internet that are drawn from the Center's research. He gives several dozen speeches a year to government officials, media leaders, scholars and students, technology executives, librarians, and nonprofit groups about the changing media ecosystem. He is also regularly interviewed by major news organizations about technology trends. Prior to launching Pew Research Center's technology research, Dr. Rainie was managing editor of U.S. News & World Report. He is a graduate of Harvard University and has a master's degree in political science from Long Island University.

Rob A. Rutenbar

Member

Rob Rutenbar received his PhD from the University of Michigan in 1984, and then joined the faculty at Carnegie Mellon University. He spent 25 years in Electrical and Computer Engineering at CMU, ultimately holding the Stephen J. Jastras (E'47) Chair. He was the founding Director of the Center for Circuit & System Solutions (called "C2S2"), a large consortium of US schools (e.g., CMU, MIT, Stanford, Berkeley, Caltech, Cornell, Columbia, GaTech, UCLA, etc.) supported by DARPA and the US semiconductor industry, focused on design problems at the end of Moore's Law scaling. In 2010 he moved to the University of Illinois at Urbana-Champaign, where he is Abel Bliss Professor and Head of Computer Science. At Illinois, he pioneered the novel "CS+X" program, which combines a core Computer Science curriculum with a disciplinary "X" curriculum, leading to a Bachelor's degree in "X"; student pipelines for CS+Anthropology, Astronomy, Chemistry, Linguistics, are now underway, with several more CS+X degrees under design. His research has focused in three broad areas: tools and for integrated circuit design; statistics of nanoscale chip designs; and custom architectures for machine learning and perception. In 1998 he founded Neolinear, Inc., to commercialize the first practical synthesis tools for non-digital ICs, and served as Neolinear's Chief Scientist until its acquisition by Cadence in 2004. In 2006 he founded Voci Technologies Inc., to commercialize enterprise-scale voice analytics. He has won numerous awards, including the IEEE CASS Industrial Pioneer Award and the Semiconductor Research Corporation Aristotle Award. His work has been featured in venues ranging from Slashdot to the Economist magazine.

Kristin M. Tolle

Member

Kristin M. Tolle is the Director of the Data Science Initiative in Microsoft Research Outreach, Redmond, WA. Since joining Microsoft in 2000, Dr. Tolle has acquired numerous patents and worked for several product teams including the Natural Language Group, Visual Studio, and the Microsoft Office Excel Team. Since joining Microsoft Research's outreach program in 2006, she has run several major initiatives from Biomedical computing and environmental science to more traditional computer and information science programs around natural user interactions and data curation. She was also directed the development of the Microsoft Translator Hub and the Environmental Science Services Toolkit. Dr. Tolle is an editor, along with Tony Hey and Stewart Tansley, of one of the earliest books on data science, *The Fourth Paradigm: Data Intensive Scientific Discovery*. Her current focus is developing an outreach program to engage with academics on data science in general and more specifically around using data to create meaningful and useful user experiences across devices platforms. Prior to joining Microsoft, Tolle was an Oak Ridge Science and Engineering Research Fellow for the National Library of Medicine and a Research Associate at the University of Arizona Artificial Intelligence Lab managing the group on medical information retrieval and natural language processing. She earned her Ph.D. in Management of Information Systems with a minor in Computational Linguistics. Dr. Tolle's present research interests include global public health as related to climate change, mobile computing to enable field scientists and inform the public, sensors used to gather ecological and environmental data, and integration and interoperability of large heterogeneous environmental data sources. She collaborates with several major research groups in Microsoft Research including eScience, computational science laboratory, computational ecology and environmental science, and the sensing and energy research group.

Talithia Williams

Member

Talithia Williams takes sophisticated numerical concepts and makes them understandable and relatable to everyone. As illustrated in her popular TedTalk, "Own Your Body's Data," she demystifies the mathematical process in amusing and insightful ways, using statistics as a way of seeing the world in a new light and transforming our future through the bold new possibilities inherent in the STEM fields. As an Associate Professor of Mathematics at Harvey Mudd College, she has made it her life's work to get people—students, parents, educators and community members—more excited about the possibilities inherent in a STEM education. In her present capacity as a faculty member, she exemplifies the role of teacher and scholar through outstanding research, with a passion for integrating and motivating the educational process with real world statistical applications. Her educational background includes a Bachelor's degree in Mathematics from Spelman College, Masters' degrees in both Mathematics from Howard University and Statistics from Rice University, and a Ph.D. in Statistics from Rice University. Her professional experiences include research appointments at the Jet Propulsion Laboratory (JPL), the National Security Agency (NSA), and NASA. Dr. Williams develops statistical models which emphasize the spatial and temporal structure of data and has partnered with the World Health Organization in developing a cataract model used to predict the cataract surgical rate for countries in Africa. Through her research and work in the community at large, she is helping change the collective mindset regarding STEM in general and math in particular—rebranding the field of mathematics as anything but dry, technical or male-dominated but instead a logical, productive career path that is crucial to the future of the country.

Andrew Zieffler

Member

Andrew Zieffler is a Senior Lecturer and researcher in the Quantitative Methods in Education program within the Department of Educational Psychology at the University of Minnesota. He currently teaches undergraduate and graduate-level courses in statistics, and trains and supervises graduate students in statistics education. In a previous life Dr. Zieffler taught mathematics and A.P. Statistics at ROCOR High School in Cold Spring, MN. Dr. Zieffler scholarship focuses on statistics education. He has also authored/co-authored several papers and book chapters related to statistics education, and has been a co-PI on many NSF-funded statistics education research projects. Additionally, he has co-authored two textbooks that serve as an introduction to modern statistical and computational methods for students in the educational and behavioral sciences. Dr. Zieffler currently serves as co-editor of the journal Technology Innovations in Statistics Education, and as a member of the Research Advisory Board for the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE). Dr. Zieffler received his Ph.D. in Quantitative Methods in Education from the University of Minnesota.

Jon Eisenberg

Staff Officer