

Additional items

abstract

The world is full of text data, yet text analytics has not traditionally played a large part in statistics education. We consider four different ways to provide students with opportunities to explore whether email messages are unwanted correspondence (spam). Text from subject lines are used to identify features that can be used in classification. The approaches include use of a Model Eliciting Activity, exploration with CODAP, modeling with a specially designed Shiny app, and coding more sophisticated analyses using R. The approaches vary in their use of technology and code but all share the common goal of using data to make better decisions and assessment of the accuracy of those decisions.

bios

Nicholas J. Horton is Beitzel Professor of Technology and Society (Statistics and Data Science) at Amherst College. He received his doctorate in biostatistics from the Harvard School of Public Health in 1999, and has co-authored a series of books on data science and statistical computing. He is a member of the ASA board of directors, and co-chair of the National Academies Committee on Applied and Theoretical Statistics. This work is part of a larger project with this team that stemmed from Nick's work as a Tinker Fellow with the Concord Consortium.

Jie Chao is a learning scientist at the Concord Consortium. She received her Ph.D. in Instructional Technology and STEM education from the University of Virginia in 2012. Jie is the PI of multiple NSF-funded projects on innovative approaches to STEM teaching and learning. Her research is focused on designing learning environments that help students develop computational thinking skills, mathematical modeling competencies, and understanding of artificial intelligence.

William Finzer is a Senior Scientist at the Concord Consortium where he leads the development of the Common Online Data Analysis Platform (CODAP). Bill serves as co-PI on the NSF-funded StoryQ, M2Studio and Boosting Data Fluency projects. His work centers on bringing data science into the K12 curriculum integrated across subject areas primarily through the creation of data exploration software designed for learning and to be accessible and usable in the classroom.

Phebe Palmer is a recent graduate from Amherst College, having received a B.A. in Statistics in 2021. Her research centers largely around STEM education, having assisted with projects focused on approaches to statistics pedagogy, as well as equitable access to STEM curriculum. She works as a research assistant at SageFox Consulting Group based in Amherst, MA.