

JSM 2019 Abstract – #304581

Training Students Concurrently in Data Science and Team Science: Results and Lessons Learned from Multi-institutional Interdisciplinary Student-led Research Teams 2012-2018.

Brent T. Ladd* and Mark Daniel Ward, Purdue University

Keywords: Collaboration, Data Science, Diversity, Engaged Learning, Interdisciplinary, R

Our student training was designed and offered annually, to introduce diverse cohorts of students and early-career scientists to first principles and concepts from data analysis. Participants completed a pre-workshop online four-week Introduction to R course. Our week-long workshop emphasized hands-on tutorials with techniques for data wrangling and visualization including data scraping, parsing, cleaning, and analysis while also fostering interdisciplinary team science. Diverse backgrounds and experience were prioritized during the selection of participants, along with disciplinary interests from the full spectrum of STEM disciplines and beyond. Teams were organized around real-world, data-driven research projects. Students from statistics, math, and computer science domains were matched with students from engineering, life sciences, and liberal arts. Multi-institutional interdisciplinary teams received funds for continuing collaborative research with the goal of co-publishing results. Outcomes demonstrate successful long-term student collaborations across institutions and topic domains at the nexus of data science. This work is supported by NSF grant CCF-0939370.

*presenting author

Detailed version:

Training Students Concurrently in Data Science and Team Science: Empirical Results and Lessons Learned from Multi-institutional Interdisciplinary Student-led Research Teams 2012-2018.

With a goal of training the next generation of science of information scholars, an annual active learning summer workshop was designed to introduce diverse cohorts of students in data science processes while fostering team science best practices. Active graduate student-led research projects were the focus of team collaborations. Diverse backgrounds, experience, gender equality, and institutional breadth were a priority in the training model. A gender balance of 1:1 female: male ratio was maintained. 166 students participated in trainings from 2012-2018 including advanced undergraduate, graduate, and post-doc levels with 24 universities and 22 distinct departments represented. The large majority of participants did not have previous data science training. Participants completed pre-workshop tutorials, and an intensive series of hands-on exercises using R with tools and techniques for data scraping, parsing, cleaning, and analysis during the workshop. Teams were organized and facilitated using best practices in team science. Professional development in grant writing was available following the workshop, with funding for teams to continue active collaborations throughout the year. To date, 18 multi-institutional interdisciplinary teams have been funded for year-round collaborative research. These teams have collectively produced 25 published papers, and 44 conference posters. Results demonstrate that providing focused training, facilitation, and small amounts of funding can lead to highly successful student collaborations when including participants across institutions and topic domains at the nexus of data science. This work is supported by NSF grant CCF-0939370.