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 & Mark D. Ward, Purdue University, Center for Science of Information

Broader Impacts Training Goal

Train the next generation of science of information scholars:

- Develop annual summer workshop introducing data science techniques to a broad spectrum of students
- Foster interdisciplinary team science across multiple institutions and domains

Student Population

- 149 Students (six annual workshops)
- 25 Universities, 22 distinct departments
- Approaching gender balance
- Mix of grad, undergrad, postdoc
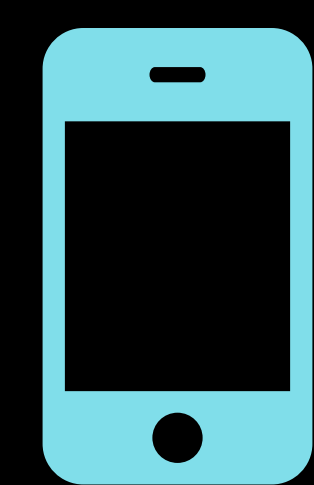- No pre-reqs required

Data Science Training

- Introduction to R (4-wk online course)
- Hands-on with R (1-wk in person workshop)
- First principles and concepts from data analysis
- Data scraping, parsing, cleaning, analysis
- Big data
- SQL databases
- Data visualization
- LaTeX

Team Science Training

- Interdisciplinary
- Multi-institutional
- Mix of grad, undergrad, postdoc
- Best practices from science of team science
- Grant writing – Professional development
- Post workshop funding – NSF style grants
- 1-2 Year team projects
- Teams co-present results at conferences and in journal publications
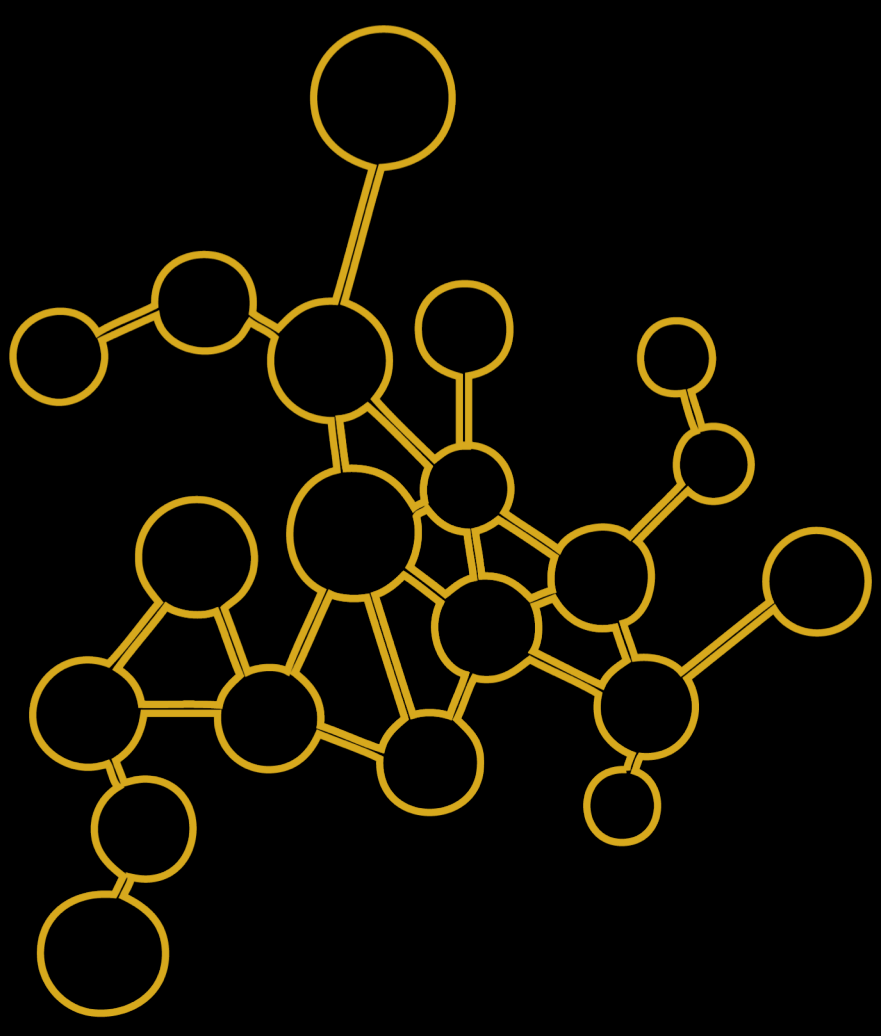
Our **engaged learning** model of training provides diverse students with immediately useful **data science skills**, while learning to work in interdisciplinary, multi-institutional research **teams** that quickly progress to co-producing **conference** and **journal publications**.
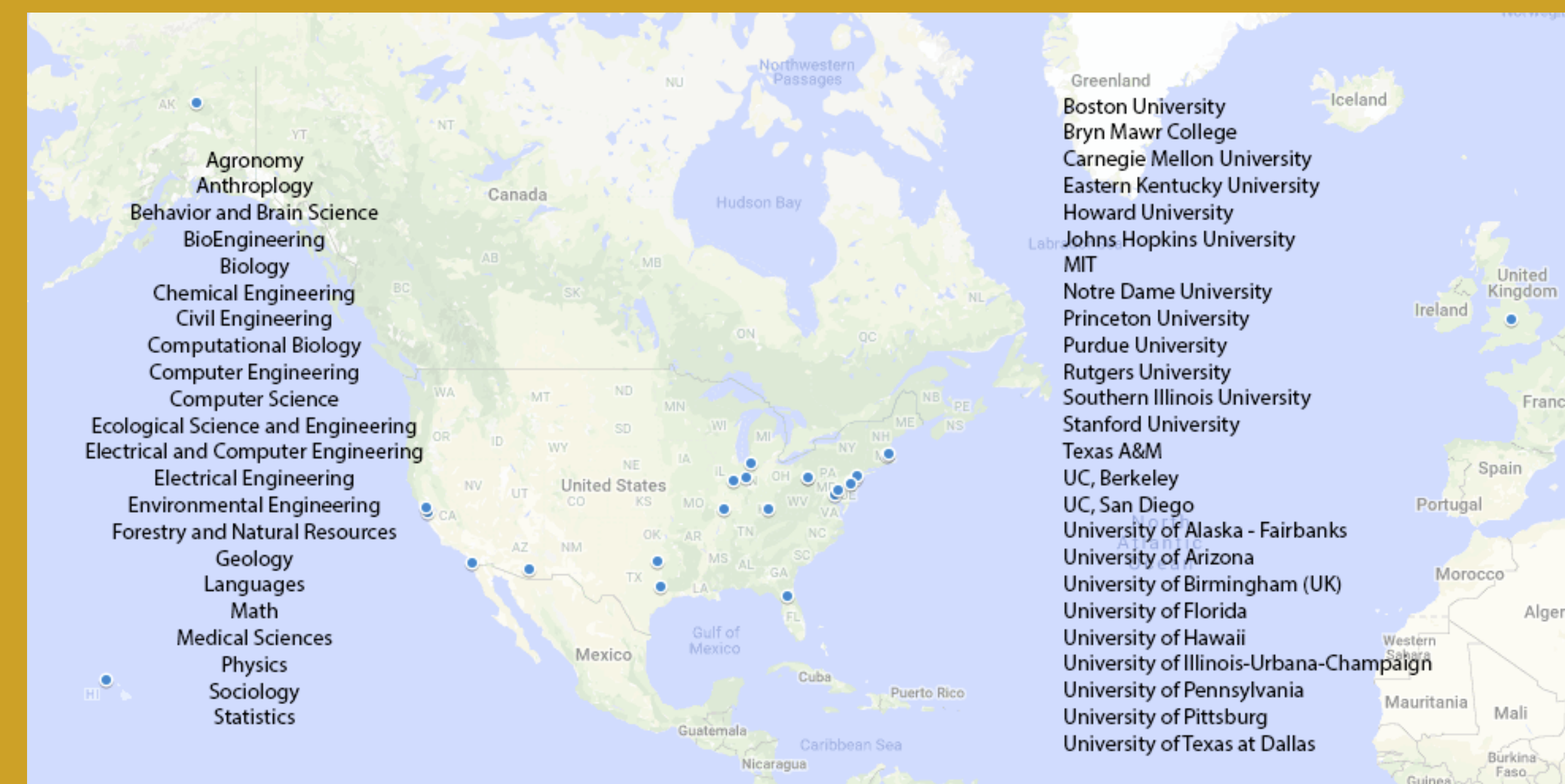
Take a picture to download the working paper

Interdisciplinary Student Team Results

- 18 multi-institutional teams funded for year-round collaborative research
- 25 co-authored papers
- 44 co-presented conference posters

Student teams co-present their results at conferences and in journal publications

Ripple Effect in our Community

GLMM analysis comparing our students who collaborated with others in our community vs. those that did not collaborate* reveals that our collaborating graduate students are significantly more productive in publishing journal papers (2.81 vs. 2.04, p< .001, *n=256, F=11.89,) as well as producing more conference posters/presentations (3.06 vs. 2.59, p= .07), with these results due primarily to the act of collaboration.

*only students with at least one publication in a sample year are included for that year in the analysis

Professional Development and Learning Outcomes from the Annual Data & Team Science Workshop (2012-2018)*

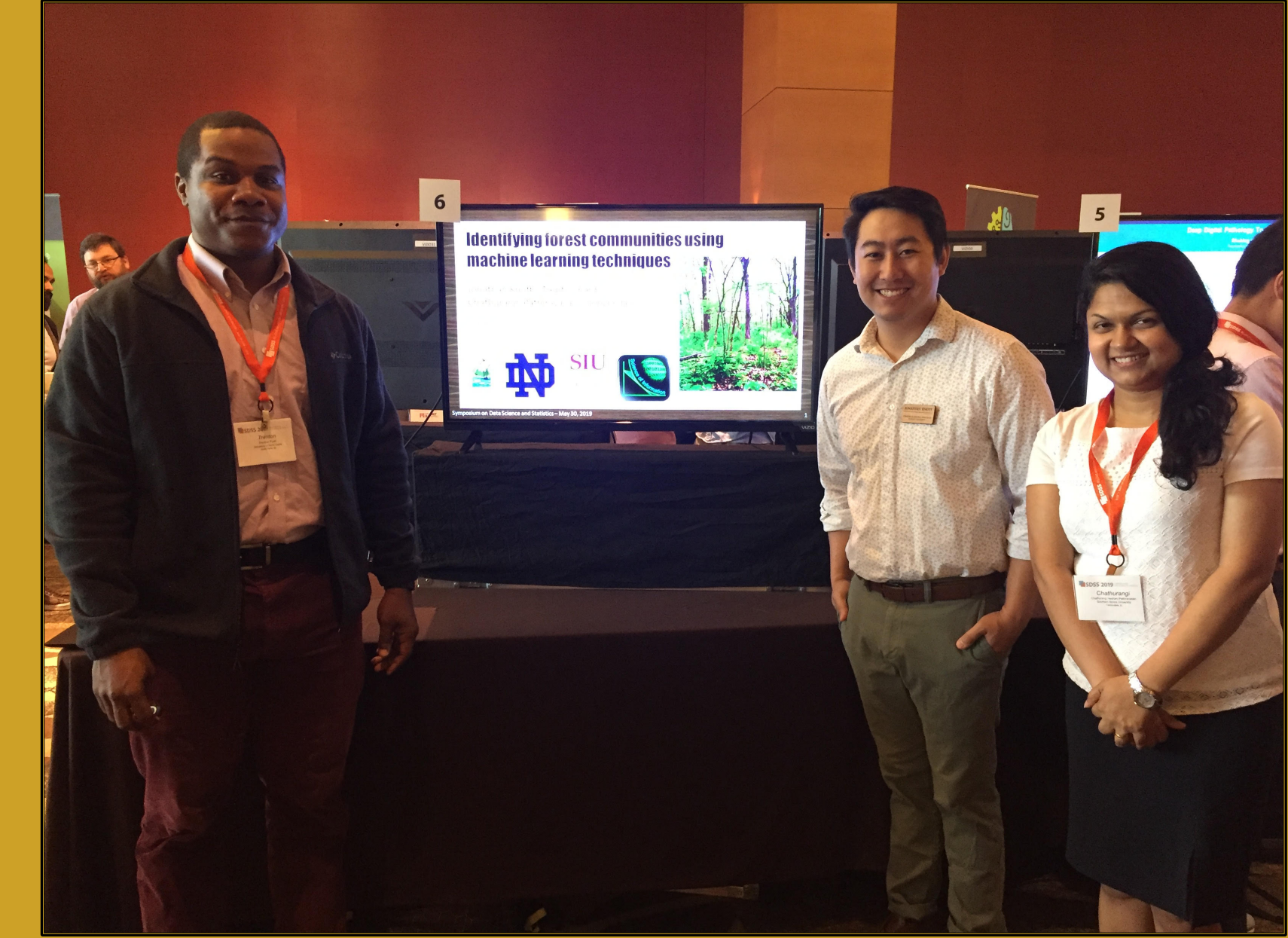| Student Evaluations Reflected Against Learning Objectives:	Mean/4.0
I received useful feedback to my own research by my interactions with peers and faculty in the workshop	3.81
I gained an improved interdisciplinary understanding to approaching a research problem	3.66
My overall experience of working in a multi-institutional interdisciplinary team during the workshop	3.66
Overall, I learned specific skills I can put to use in my own research/courses	3.63
I started some level of professional connections with peers through the workshop	3.60
I improved my ability to explain my research to others as a result of interactions during the workshop.	3.53

n=85 anonymous responses. 4 point Likert scale 1=poor, 2=fair, 3=good, 4=excellent
the workshop was not offered in 2013 due to hosting the NASIT summer school

Disciplinary and Institutional Breadth of the Student Research Team Members

Agronomy
Anthropology
Behavior and Brain Science
BioEngineering
Biology
Chemical Engineering
Civil Engineering
Computational Biology
Computer Engineering
Computer Science
Ecological Science and Engineering
Electrical and Computer Engineering
Electrical Engineering
Environmental Engineering
Forestry and Natural Resources
Geology
Languages
Math
Medical Sciences
Physics
Sociology
Statistics

Boston University
Bryn Mawr College
Carnegie Mellon University
Eastern Kentucky University
Howard University
Johns Hopkins University
MIT
Notre Dame University
Princeton University
Purdue University
Rutgers University
Southern Illinois University
Stanford University
Texas A&M
UC, Berkeley
UC, San Diego
University of Alaska - Fairbanks
University of Arizona
University of Birmingham (UK)
University of Florida
University of Hawaii
University of Illinois-Urbana-Champaign
University of Pennsylvania
University of Pittsburg
University of Texas at Dallas

Collaboration on Research Publications?