Harvard Data Science Review • Issue 1.2, Fall 2019

Data Science as a Foundation for Inclusive Learning

Robert A. Lue^{1,2}

¹Department of Molecular and Cellular Biology, Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America,

²Derek Bok Center for Teaching and Learning, Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Published on: Nov 19, 2019

DOI: <https://doi.org/10.1162/99608f92.c9267215>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

STEM (science, technology, engineering, and mathematics) fields such as computer science and statistics have traditionally struggled to increase participation from underrepresented groups such as women and minorities. And while progress has been made in developing more inclusive classroom pedagogies and programs, rates of representation have remained low over the past 20 years. The emerging field of data science with its roots in both computer science and statistics faces the same challenges, but also presents a unique opportunity to build a more inclusive framework for the teaching of STEM. The widening application of data science methods to nearly every field imaginable in the natural sciences, social sciences, and humanities opens up avenues for engagement based on what students care about and the challenges they are most interested in tackling. Data science therefore provides an opportunity to build an inclusive STEM curriculum from the ground up that connects with multiple disciplines as well as the personal passions of students.

Keywords: education, inclusive teaching, pedagogy, STEM, curriculum, statistics, computer science

---

Data: a word that brings daily excitement to researchers and equally as often mystifies laypersons everywhere. It encompasses the results of observation, serendipity, experiments, and research that can be shared with the wider society, but can also be adulterated to support fabricated worldviews. This latter risk is the impetus for one commonly held vision of an open-sharing world where the ability to properly gather, interpret, and use data is an essential part of a basic education that everyone deserves. But if this is true, how close are we to making data science inclusive enough to be accessible to all? And given the emergence of data science from statistics and computer science, have we made sufficient progress in those fields when it comes to representation so that data science can build on those successes? While these questions may seem reasonable, they are in fact placing the cart in front of the horse. It is not progress in statistics and computer science that will make data science more inclusive and equitable, rather, it is the other way around. Data science, being a nascent field, represents not just an early opportunity to address the historical structures in STEM that often exclude women and underrepresented minorities, its very nature opens up new ways of engaging a far broader audience of students.

Nevertheless, the statistical and computational lineage of data science allows us to take a hard look at what has or hasn't been achieved in those fields and to learn from those continuing efforts. The most complete datasets focus on women and minority groups (specifically Black, Hispanic, and Native American) where current employment figures indicate that they continue to be underrepresented in the fields of statistics and computer science ([Bureau of Labor Statistics, 2018](#)). For example, women accounted for a mere 18% of computer science degrees in 2015 ([National Science Foundation, 2018](#)), and in 2014, Black and Native American students comprised only 4.6% of degrees in mathematics and statistics; a 20-year low. The overall trend is clear: women and minorities have experienced decreasing participation in these fields over the past 20 years.

There are multiple issues that contribute to the disparity in the number of women and minorities studying and continuing in computer science and statistics. One is the well-documented phenomenon of stereotype threat (reviewed in [Spencer, Logel, & Davies, 2016](#)) where both women and minorities internalize the stereotype that they are not adept at quantitative work, which results in decreasing performance and a higher rate of exit from quantitative fields. The problem of underrepresentation is further exacerbated by the overrepresentation of minority students in underresourced secondary schools. Thus, many minority students come from high schools with limited instructional resources, contributing to a first experience with quantitative course work in college that is demoralizing due to a lack of preparation. Young women can be similarly disadvantaged in their first computer science and statistics courses due to inappropriate messaging in high school that mathematically inclined course work is not for them. So even if they go against these negative cues in college, they are not as well prepared having avoided relevant coursework in high school.

So, what has been done to remedy this situation? Progress has been made in developing classroom practices designed to mitigate stereotype threat and to establish a culture of active engagement and belonging (reviewed in [Lawrie et al., 2017](#)). The issue of less background preparation is often addressed by programs designed to level-set student backgrounds in introductory classes. Prematriculation summer bridge is one common response that combines remedial training in mathematics with workshops designed to acclimatize students to the culture of college ([Murphy, Gaughan, Hume, & Moore, 2010](#)). All of these programs and pedagogical innovations are worthwhile, but instead of simply reapplying them to the field of data science, which by extension suffers from the same disparities as the parental fields of computer science and statistics, why not take advantage of the very essence of data science to foster the first STEM-related discipline designed to be taught inclusively from the beginning?

Data science is uniquely well suited to inclusive education because of its emphasis on data; how research is designed around it and how it is gathered and analyzed. Since the kind and source of data is entirely open, it makes data science relevant to any field. This provides an unmatched breadth of disciplinary scope, from the analysis of textual patterns in poetry to the large-scale mapping of neuronal connections in the brain; from the prediction of audience responses to a new TV series to the dynamic remodeling of supply chains in the fashion industry. There is virtually no discipline that has not been affected by the approaches and tools of data science, which means that for the first time it is highly likely that every student will be able to connect what they care about with an approach or set of analytical tools from this emergent field.

In most STEM courses we expend enormous effort on trying to convince students that, say, biology or physics is the field for them. We passionately relay stories of how we and others found connection with the field that we teach and hope that it will strike a chord in as many students as possible. We frame our discipline in the service of challenges that humanity must face, thereby grasping for a profound relevance that is so compelling as to be universal. But it is exactly this hope that our perspective and these stories are somehow universally compelling that is the fallacy. If there is one thing that the growing diversity of students in the college

classroom has revealed, it is the equally diverse range of motivations and personal histories that inform what our students find inspiring. To ignore this is to ignore our duties as instructors to foster genuine agency in our students. So instead of selling them on our stories of what gets us out of bed in the morning, how can we help them discover their own passions? For starters, we can move beyond the simplistic notion of having our students fill a toolbox with abstract concepts and analytical tools and instead use data science to provide them with a set of skills and habits of mind relevant to virtually every field. Better yet, we can also foster the learning of said skills directly in the context of a problem or field of their choice. Such a learner-centered approach is enabled by data science, providing a breadth of opportunity to apply what they learn to a field they care about.

The university context also provides a rich test bed for exploring whether data science can indeed catalyze greater personal agency. Take the University of California at Berkeley Data Science Connector Course model, for example, where students beginning data science coursework can concurrently pursue their interests in numerous other fields through allied connector courses—in fact, they are strongly encouraged to do so. Since the fall of 2015, there have been connector courses offered at the intersections of data science and civil engineering, cognitive science, computer science, demography, earth science, environmental science, ecology, geography, history, legal studies, molecular/cellular biology, statistics, and general liberal arts. It is noteworthy that some of these cross-connecting courses are traditionally characterized by overrepresentations of women and minorities. This signals that connectors can be used by women and minorities to gain access to the approaches of statistics and computer science under the auspices of application to issues drawn from another field.

The approach of using data science as a common foundation also expands the horizons of those students traditionally overrepresented in statistics and computer science as they are faced with the myriad applications beyond those that first drew them to quantitative fields in the first place. Simply put, data science could provide a context for a transdisciplinary diffusion of students that fosters both exploration and agency; the two pillars that support personal motivation. Furthermore, it can connect fields that have become increasingly gender segregated, sharing skills and perspectives from one to the other. Crucially, this might also allow women and minorities to enter data science through any one of many avenues informed by what motivates them, thereby bypassing the hurdles often erected by situational and social cues. Equally as important, the experience of data science in context can deepen enthusiasm for allied disciplines and contribute to students continuing in, say, computer science or statistics.

The Berkeley Connector model represents one curricular instance of how data science might open up more inclusive opportunities in quantitative and computational fields. This claim is worthy of careful evaluation, especially in the context of an approach that is both systemic and wide-ranging in terms of the students that it engages. At its core, it is also about presenting students with opportunities for agency, which lies at the heart of outcomes-oriented pedagogies that emphasize problem-solving and addressing challenges. It is true that despite

decades of awareness that STEM fields tend to exclude women and minorities, limited progress, if any, has been made when it comes to shifting the demographics of these fields. This humbling reality further underscores that there are no simple fixes for such a systemic problem. Nevertheless, it is not every day that a new field emerges that has the potential to connect disciplines across the natural sciences, social sciences, and humanities. So, we would be remiss to not leverage this opportunity to throw open the doors of STEM, not simply through evangelism, but also through emphasizing intellectual openness and an authentic relevance to fields and ideas that students care about.

Disclosure Statement

Robert Lue has no financial or non-financial disclosures to share for this article.

References

Bureau of Labor Statistics. (2018). Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity. Retrieved June 27, 2019, from <https://www.bls.gov/cps/cpsaat11.htm>

Lawrie, G., Marquis, E., Fuller, E., Newman, T., Qiu, M., Nomikoudis, M., . . . van Dam, L. (2017). Moving towards inclusive learning and teaching: A synthesis of recent literature. *Teaching & Learning Inquiry: The ISSOTL Journal*, 5(1), 10. <http://doi.org/10.20343/teachlearningqu.5.1.3>

Murphy, T. E., Gaughan, M., Hume, R., & Moore, S. G., Jr. (2010). College graduation rates for minority students in a selective technical university: Will participation in a summer bridge program contribute to success? *Educational Evaluation and Policy Analysis*, 32(1), 70–83. <https://doi.org/10.3102/0162373709360064>

National Science Foundation. (2018). *S&E Indicators 2018*. Retrieved June 27, 2019, from <https://www.nsf.gov/statistics/2018/nsb20181/report/sections/higher-education-in-science-and-engineering/undergraduate-education-enrollment-and-degrees-in-the-united-states>

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>

©2019 Robert Lue. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.