Taylor & Francis
Taylor & Francis Group

Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujse21

Inclusivity in Statistics and Data Science Education

Jeff Witmer

To cite this article: Jeff Witmer (2021) Inclusivity in Statistics and Data Science Education, Journal of Statistics and Data Science Education, 29:1, 2-3, DOI: 10.1080/26939169.2021.1906555

To link to this article: https://doi.org/10.1080/26939169.2021.1906555

Published online: 17 May 2021.

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION
2021, VOL. 29, NO. 1, 2–3
https://doi.org/10.1080/26939169.2021.1906555

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

EDITORIAL

Inclusivity in Statistics and Data Science Education

The real-world nature of data science and statistics makes it at once interesting and potentially uncomfortable. As educators we want all students to feel welcome. In statistics and data science, context matters. We get to apply statistical reasoning in a wide array of applications, including issues of how society functions.

Many harsh realities affect our lives, and the lives of our students, both in and outside of the classroom. Many face income inequality, structural racism, differential health outcomes during a pandemic, and sexism on a daily basis. In particular, tensions around race are running high these days in America, and elsewhere in the world. Social and political issues can motivate our teaching, as they are central to our collective experiences of being human, but such examples are challenging to address without contributing to increased polarization. They have to be used with care, as students can be alienated if they think their professor is promoting a particular agenda or is insensitive or unaware of potential biases lurking unexamined in those examples. As teachers, many of us try to develop awareness of our own ideological biases so that we can help our students to learn *how* to think (and as statistics educators, how to reason from data) but not *what* to think. This can be difficult to navigate, but given that statistics are applied to real-world problems we have the opportunity to demonstrate problems and how to find solutions.

The history of statistics includes many regrettable elements related to systemic racism and eugenics. A recent article in *Nautilus*, entitled "How Eugenics Shaped Statistics," is sobering, to put it mildly. As I'm sure you know, last year the American Statistical Association looked at some of this history and took an initial step in recognition of these inequities by recommending the renaming of the Fisher Lecture.

As the history of eugenics comes into clearer focus and society continues to wrestle with how racism in the past shapes lives today, how should I think about race? More to the point, how do my thoughts about race affect my teaching? I have lived in a world awash with race all of my life, but I was stopped cold by a sentence Ta-Nehisi Coates wrote in his book *Between the World and Me*: "Race is the child of racism, not the father." I had presumed, without ever thinking about it, that the existence of race led to the development of racist behaviors. The idea that we judge others who are different from us and then justify our treatment of them by creating labels, such as race, was not part of my education. What else am I not thinking about but merely taking for granted?

Yet race is by no means the only social issue of the day. When teaching data science and statistics we routinely use gender as an example of a binary predictor variable–which can be great for illustrating the idea of parallel regression lines, or for setting up a two-sample t-test, but not so great for making everyone feel welcome. As you know, sex is a biological variable with more than two levels. Gender is a social construct with more than two levels. Some of our students may think of sex and gender as interchangeable — and may benefit from thinking about the difference between them; but whichever of these we use, if we present it as binary we send a subtle, or maybe not so subtle, message. One action we can take is to acknowledge that there are more than two levels. But we can also replace the use of gender with a substitute variable that is binary.

Aside from using gender (or sex) as a binary variable, we may do other things, while working with real data, that may inadvertently work against the environment we want to create. In addition to my choices of examples to present[1] and datasets to analyze, there are many other questions I could ask myself about how I structure my classes. Does my course structure or assessment system favor white men, for example? Is my classroom a welcoming place in which full potential is cultivated, rather than blocked by thoughtless habits and careless words? Do I present examples of successful statisticians from underrepresented groups who might be inspiring models for my students? Am I sensitive to how the data I present to my students depict aspects of race, gender, socioeconomic status, etc.?

Many people are thinking about these issues and no one claims to have all of the answers, but here are some resources that you might find helpful:

- A recent Public Health Post points out, among other things, that differences in health outcomes that are primarily due to social factors are often falsely attributed to race, something that we have seen during the COVID-19 pandemic; see *https://www.publichealthpost.org/research/why-are-we-still-using-race-as-a-variable-in-health-research/*.
- One of the five goals of the strategic plan of the Council on Undergraduate Research concerns diversity, equity, and inclusion; see *https://www.cur.org/who/organization/plan/*.

[1]My attention was drawn to a paper we have published with an example in it that made some students uncomfortable because of how it dealt with body size.

- The Mathematical Association of America conducted a webinar series last year called Conversations for the Math Community about diversity, equity, and inclusion; see *http://info.maa.org/pages/1780913/23513?utm_source=newsletter&utm_medium=email&utm_content=DEI%20webinar%20series&utm_campaign=ProgramsCampaign_Email%206*.
- An editorial recently published in *JAMA* talks through definitions of race and ethnicity and recommends guidelines for reporting race, ethnicity, and other demographic information; see *https://jamanetwork.com/journals/jama/fullarticle/2776936*. This could provide a starting point for a conversation with students on labels in data collection and in usage.[2]
- Bryan Martin has created *https://github.com/bryandmartin/gibboda*, where gibboda is short for Gender Isn't Binary But Other Data Are. Martin invites contributions to this collection of datasets with binary variables that might be used in place of gender.
- Jo Hardin and Jack Miller wrote an article on "The Evolution of Variables and the Existence of Trans People" for the March 2019 *Amstat News*: *https://magazine.amstat.org/blog/2019/03/01/evolutionofvariables/*
- The June 2019 issue of *SIGNIFICANCE* included a column on "Friends and allies: LGBT+ inclusion in statistics and data science." (*https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2019.01280.x*)

One more point is that your students will make mistakes and when they do their classmates might call them out. At times this is appropriate, but often it is more constructive for all involved if they can be extended the grace that I have often received of being called in, rather than being called out; see *https://www.nytimes.com/2020/11/19/style/loretta-ross-smith-college-cancel-culture.html*.

The world of statistics may have taken small steps in recent years to promote inclusivity, but there is more that we can do. I encourage you to think about your pedagogical practice, including unexamined assumptions and behaviors. I also encourage you to submit articles to *JSDSE* that will help us grow as a community that supports an ever-more-diverse population of students.

Disclaimer: I am grateful for the input of many people who have helped me draft this editorial, particularly Scarlett Bellamy, William Cipolli, Jennifer Green, Matt Hayat, Nick Horton, Debra Hydorn, Jack Miller, and Bill Notz, but the opinions expressed here are mine and I am not speaking on behalf of the American Statistical Association.

Jeff Witmer
Editor-in-Chief

[2] Although this editorial is valuable, JAMA does not have a clean bill of health. In February they aired a podcast that questioned the existence of racist doctors and of structural racism. They are now reviewing editorial processes and they acknowledge that "structural racism is real, pernicious and pervasive in health care;" see *https://www.medscape.com/viewarticle/947256?src=WNL_trdalrt_210315_MSCPEDIT&uac=104660MZ&impID=3250064&faf=1*.