Empowering Students to Assess the State of Diversity, Equity, and Inclusion on Campus

Jonathan Auerbach[1] and Christi Wilcox[2]

[1] Department of Statistics

[2] CEC Diversity, Outreach & Inclusive Learning

Volgenau School of Engineering

George Mason University

Bios

Jonathan Auerbach is an Assistant Professor in the Department of Statistics at the Volgenau School of Engineering at George Mason University.

Christi Wilcox is Associate Director and Deputy Title IX Coordinator of CEC Diversity, Outreach & Inclusive Learning at the Volgenau School of Engineering at George Mason University.

Abstract

Universities and colleges have strengthened their commitment to diversity, equity, and inclusion (DEI) over the past decade. Many have established DEI offices tasked with identifying and addressing barriers that limit the participation of students, faculty, and staff. The effects of these barriers are often reflected in the volumes of data on enrollment, graduation, and other facets of university life reported to the U.S. Department of Education each year. But it is unrealistic to expect even a generously resourced office to monitor the vast amount of data available.

In this paper, we propose that instructors of project-based statistics and data science courses work with DEI offices to have students use this data to investigate the state of DEI on campus. We describe a class project in which undergraduate students conducted their own independent assessments and submitted a short report summarizing their findings. The lesson proved rewarding to both the students and the DEI office: Students reported a variety of new and interesting trends. Students also gained a different perspective of campus diversity and a deeper appreciation for how diversity is calculated in practice. We conclude that students are an untapped resource that can help improve campus life as they develop their statistics and data science skills.

Keywords

Diversity, Equity, and Inclusion; DEI; Capstone Design Courses or Projects; Diversity in STEM; Campus as a Laboratory; Statistics; Data Science; Teaching; Statistics Education

Introduction

Universities and colleges have strengthened their commitment to diversity, equity, and inclusion (DEI) over the past few years. Many have established DEI offices tasked with identifying and addressing barriers that limit the ability of students, faculty, and staff to fully participate in university activities. For example, DEI offices can provide undergraduate students with additional resources to complete their coursework, such as peer mentoring and study spaces. They can also assist a wide range of faculty—from K-12 educators who facilitate the transition to higher education to college instructors who develop lesson plans that better support students of all backgrounds.

The barriers that limit the ability of students, faculty, and staff to participate in university activities are formidable. Many originate from inequities deeply ingrained in society, and when these inequities manifest as barriers, they limit participation in complicated ways. Data can provide much needed insight. Indeed, the effects of the barriers that limit participation are often reflected in the volumes of data on enrollment, graduation, and other facets of university life reported to the U.S. Department of Education each year. But it is unrealistic to expect even a generously resourced office to monitor the vast amount of data available. Assistance is needed.

At the same time, instructors seek meaningful data that motivate students to practice coding and analytical skills. Ideally, meaningful data are both consequential, so that students feel invested in the outcome of their analysis, and familiar, so that students understand what the numbers represent. But the search for meaningful data can be difficult in diverse classrooms because data meaningful to students of one background may not be meaningful to students of another.

In this paper, we describe a project aimed at addressing these challenges. We propose that instructors work with DEI offices to have students assess the state of DEI on campus. We argue that such projects can benefit both DEI offices and instructors. DEI data are easily obtained, analyzed, and meaningful to all students. In addition, DEI offices gain the perspective of the students that attend their institution. Though student assessments may not be exhaustive or conclusive, new findings encourage dialogue and can inform future programs and policies.

We provide the details of our proposal in five sections. The first two sections provide background information. In the first section, we describe the upper-level statistics communication course for which the proposed project was designed. This information is important for understanding our motivation and the open-ended nature of the assignment. However, we argue that the project can be adapted for other settings, such as an introductory statistics course. In the second section, we describe the mission of a DEI office. This information provides important context for students, who we have found are greatly motivated by aiding that mission. We also contrast our proposal with other activities proposed to address DEI on campus.

In the next two sections, we describe the details of the project as it was implemented at the author's institution. In the third section, we discuss the data available to students to assess DEI as well as several measures of diversity students might consider for their investigation, from which the barriers that limit participation might be inferred. In the fourth section, we describe the students' findings, including an example analysis based on a student submission. We also review the results of a post-project survey the students completed. We find that awareness of DEI increased following the project. Only 6% of students were aware of how diversity and inclusion are measured before the project, compared to 100% after. In addition, roughly one third of the students indicated that the project challenged beliefs they held prior to their analysis.

We conclude by discussing some minor challenges we experienced conducting the proposed project. For example, we found that the students in the class had a limited perspective of the barriers facing faculty and staff. Yet overall, we conclude that students are an untapped resource that can help improve campus life while developing their data analysis and coding skills. We argue this resource is becoming only more important as some states have moved to defund or otherwise restrict DEI offices in recent years.

Section 1. The proposed project was part of an undergraduate statistics communication course

The proposed project was assigned to undergraduate students as part of a three-credit-hour statistics communication course offered during the 2022 Fall semester. The course is a core course at the authors' institution, and one of two final courses required to complete the statistics major. Students typically complete the course the semester before graduation, and the coursework assumes students have completed one semester each of undergraduate probability, statistics, and regression, the first two of which are calculus-based. The coursework also assumes students are familiar with the programming language `R`—in particular, the set of packages that make up the `tidyverse`. Students first become familiar with this software while completing two additional data science courses required for the statistics major: a coding course and an exploratory data analysis course.

The goal of the course is to round out the statistics curriculum as recommended in the Curriculum Guidelines for Undergraduate Programs in Statistical Science (American Statistical Association 2014). Students apply the statistics and data science knowledge they have gained from previous courses to solve a series of real problems facing campus stakeholders and communicate their findings to non-statisticians. The benefits associated with community-oriented classes have been documented, for example, by Phelps and Dostilio (2008) and Tijsma et al. (2020).

The project described in this paper is the first of five projects assigned to students over the semester and the only project that involved DEI. Each project culminated in an approximately 1,500-word report. We followed the Guidelines for Assessment and Instruction in Statistics Education (Carver et al. 2016) in choosing projects that attained the characteristics described in Appendix C and in grading student submissions using the example rubric and numeric scoring approach described in Appendix E.

Classes were conducted in person and met two days a week over fifteen weeks. The first meeting of the week was reserved for lectures and the second meeting of the week was reserved for labs. Lectures both summarized important background information regarding each project and reviewed core concepts from probability, statistics, and regression, spending approximately the first month on probability, the second on statistics, and the third on regression.

The project described in this paper was assigned during the second week of the course, while the lectures reviewed probability concepts such as combinatorics and the laws of probability. The reason this project was assigned first was because diversity is often assessed using combinatorial principles, some of which we describe in Section 3. While students were not limited to these methods, we found it helpful that the review of core concepts reinforce the background information for the project. Indeed, we have found that when we do not review probability (i.e., when we teach diversity measures without reviewing combinatorial principles), students default to mindlessly applying linear regression. (Of course, we encourage the mindful application of linear regression.)

Although the proposed project was designed for an upper-level statistics communications class, we believe it is easily adapted to a wide range of settings. Indeed, we argue it is particularly relevant for an introductory statistics course. This is because students begin their professional journey in statistics and data science when they take their first class and that early interactions between students are formative. The DEI data we describe in the following sections provide a reasonably objective basis for students to reflect on their colleagues and think about how their experiences may differ from others. Moreover, we believe devoting class time to DEI tells students that DEI is important for their personal and professional development.

Note that the course is taught at one of the most diverse public schools in the United States. The sixteen students who participated in the course reflect that diversity: Eight students identified as non-White, five students identified as Black or Hispanic, and four students identified as female. Two students chose not to share any demographic information.

Section 2. The proposed project advances the mission of DEI offices

Among their many responsibilities, DEI offices are tasked with an ostensibly simple mission—to ensure all members of the community can participate regardless of background, culture, and identity. Overcoming barriers to participation at post-secondary institutions is particularly important. A comprehensive review by the U.S. Department of Education (DOE) found that "higher education is a key pathway for social mobility in the United States" and "gaps in college opportunity contributed to diminished social mobility" (U.S. Department of Education 2016).

The DOE recommends institutions build "their capacity to collect and analyze the data required to set and track their diversity and inclusion goals." But this recommendation is no small task. Barriers exist at multiple points across the higher education pipeline, including at application, admission, enrollment, persistence, and completion. A large data collection and analysis effort is necessary to examine all such points. Thus, by participating in the proposed project, students can join their peers to assist the DEI office in identifying and addressing the barriers responsible for decreased participation, potentially increasing the social mobility of their peers.

The benefits of participation extend beyond the immediate institution in which the students are enrolled. As mentioned in Section 1, we take the position that students begin their professional journeys in statistics and data science the moment they begin their first statistics course, and thus coursework should prepare future leaders that are both technically competent and socially responsible as outlined by the codes of conduct set by their prospective professional associations. In the case of the American Statistical Association, for example, the codes include the Justice, Equity, Diversity, and Inclusion (JEDI) objectives:

1. Learn from our members and others how to identify and overcome systemic racism and hindering biases of any kind.

2. Critically reappraise and improve the effectiveness of our JEDI efforts.

3. Identify and develop resources for individuals and organizations in our professional community to enable growth and appreciation for cultural humility.

4. Share openly our diversity and inclusion efforts and the solutions we have implemented.

We end this section by comparing the proposed project to others proposed to promote DEI. Our literature search revealed a wide variety of such activities over the past few years, and we divide these projects into two categories for comparison purposes.

The first category—and the most common by far—is the creation of courses, panels, and workshops that raise awareness of DEI barriers, such as microaggressions and implicit biases. These efforts are often interdisciplinary and may also include staff and members of the community. See Lesser and Nordenhaug (2004), Lesser (2007), Rawat et al. (2017), Herrera et al. (2021), and Asgarpoor et al. (2021) for examples. The project we propose is similar in that we raise awareness of DEI barriers by having students study DEI on campus.

The second category is the use of technology, outreach, or other strategies to make courses and other programs at universities and colleges more accessible. These projects are particularly impactful when course materials are difficult for some students to access. See Gray et al. (2016), Angrave et al. (2020), Dickens (2021), and Liao (2023) for examples. The project we propose is similar in spirit in that students use technology (i.e., statistical software) to provide the DEI office with insight that can help make campus more accessible. However, it is different in that the technology does not itself eliminate DEI barriers—the DEI office must act on any insight to make campus more accessible.

After conducting our literature search, we concluded that though a wide variety of projects have been proposed to promote DEI on campus, we could find no project that affords students the opportunity to use their data analysis and coding skills to further the mission of DEI offices on campus. In the remainder of this paper, we discuss how the proposed project can be conducted in practice. As mentioned in Section 1, our discussion is based on a version of the project we recently conducted in an upper level, undergraduate statistical communications course offered at the authors' institution. However, we also include additional information that may help instructors adapt the project for an introductory undergraduate statistics course or even a statistics course at the graduate level.

Section 3. Students at the authors' institution assessed DEI on campus in a month-long project

The proposed project was assigned to the sixteen students enrolled in the statistical communications course described in Section 1. The project utilized two and a half hours of class time over two days. One class was taught by a member of the DEI office, in which the mission of the office was discussed. The other class was taught by the statistics instructor, in which official DEI statistics were reviewed and three example DEI measures were demonstrated. By official DEI statistics, we mean race, ethnicity, gender, and related data collected by the U.S. Department of Education from postsecondary institutions as authorized under the Civil Rights and Higher Education Acts. See Subsection 3.a for details.

Students were given two weeks to conduct their individual investigations and draft a report of approximately 1,500 words. The reports were reviewed for accuracy and clarity, and the students were then provided another two weeks to revise their investigation and submit their final reports. As mentioned in Section 1, the course reviewed core concepts from probability theory while students worked on their projects. Time was also provided at the end of each class and during labs and office hours to answer questions related to the project.

The project instructions were intentionally simple to recreate the types of open-ended problems that typically arise in the real world. (See Carver et al. (2016) for a discussion of real-world and open-ended problems.) Students were asked to use official DEI statistics to answer the question: Is the authors' institution diverse? Students could examine the entire university or a portion of it, for example the engineering school or students in STEM programs. Students were encouraged to let their personal experience guide their analysis; however, they were instructed to answer the question (yes, no, somewhat, etc.), support their answer with the official statistics as the primary evidence, and use their findings to suggest policies or recommendations that would make the authors' institution more equitable and inclusive.

Students were told that the DEI office at the authors' institution was the target audience, that both the DEI office and instructor would review the projects, and that grades would be assigned using the rubric and numeric scoring approach mentioned in Section 1. There was no required format for the report.

3.a. Students were required to support their reasoning using official DEI statistics

The primary data for the project were the official DEI statistics published by the Department of Education's National Center for Education Statistics through the Integrated Postsecondary Education Data System (IPEDS). IPEDS data are collected using a system of interrelated surveys conducted annually by the National Center for Education Statistics. The data contain detailed enrollment, funding, and graduation information from roughly 6,400 colleges, universities, and technical and vocational institutions. Hundreds of variables are available for analysis, including the number of students attending each institution by race, ethnicity, gender, area of study, degree sought, and whether full or part-time.

The data are of extremely high quality since all institutions that participate in the federal student aid programs are required to complete the IPEDS surveys by law as authorized under Title VI of the Civil Rights Act and Title IV of the Higher Education Act. The data are also well documented. Students were encouraged to complete the five module-based tutorials designed to introduce them to the IPEDS data (U.S. Department of Education 2022).

Students were taught several ways to access the data. The preferred approach was for students to use the Urban Institute Education Data Portal. The Urban Institute supports an Application Programming Interface (API) as well as “wrapper” packages for the `R`, `Python`, `Javascript`, and `Stata` programming languages. (“Wrapper” packages refer to third-party software built on top of the API that provide a more user-friendly interface.) Most of the students in the class were sufficiently familiar with `R` that the `R` package was demonstrated in class (Urban Institute 2022). We provide a short example using this approach in Subsection 4.a. The `R` code to reproduce this example is contained in the Appendix.

Alternative approaches taught to the students include a demonstration of how to download data directly from the IPEDS website by constructing “statistical tables,” a demonstration of how to access the complete data files, and a demonstration of a comprehensive dataset created by Urban Institute from IPEDS and other sources (Urban Institute 2020). We found it was important to provide multiple ways to access the data to accommodate students less comfortable with coding. We also believe this approach is appropriate for introductory statistics classes, where familiarity with coding is not required or expected.

3.b. Students were provided with three examples of how to measure diversity

Students were taught three indices that are commonly used to measure the diversity of a group. The lack of diversity according to any of these measures may suggest the existence of barriers that prevent inclusion and the equitable distribution of campus resources. All three measures assume that the population has been divided into K distinct groups, $K > 1.$ The three indices are:

Disparity Index — calculate the percentage of the population from historically marginalized communities. Often the percentage of Black or Hispanic students is used. However, students might also consider the percentage of women in STEM for example.

Diversity Index — randomly pair individuals and calculate the percentage of pairs in which more than one group is represented. Note that this index is closely related to the Simpson's Diversity Index and the Herfindahl–Hirschman Index.

Diffusion Index — calculate the percentage of students not in the k largest groups, where $k < K$ is determined by the student. For example, the percentage of non-White students might be considered, provided students who identify as White are the most prevalent group.

Students were told they could compare one or more universities over time or one or more universities to the state or county population that contained the institution. For further reading, students were assigned to read the Urban Institute report "How Racially Representative Is Your College?" (Urban Institute 2020), which uses the Disparity Index, and the Census Bureau blog post "Measuring Racial and Ethnic Diversity for the 2020 Census" (Jensen et al. 2021), which uses the Diversity and Diffusion Indices. Additional readings that demonstrate diversity analyses are given by the Chronicle of Higher Education (2017), De Brey et al. (2019), and Auerbach and DeLazzero (2022).

We note that one population may be considered more diverse than another if it has a higher index according to one or more of these measures. However, the three measures may disagree on which of two populations is more diverse. Furthermore, the measures themselves may change depending on which groups are chosen for the analysis. For these reasons, we stressed that it is incumbent upon the student to choose meaningful groups and an appropriate measure that makes sense within the context of the students' investigation. Students were not required to use one of these measures, and several students chose alternative measures using concepts such as correlation, regression, and entropy.

We stress to the students that the goal of the project is to offer actionable insight that could help make the authors' institution more equitable and inclusive. Thus, while students are asked to use the data to measure diversity, a complete report must use those measurements to characterize the state of equity and inclusion on campus. Such characterizations can borrow from students' own experiences. For example, a student might first establish that a specific major is not as diverse as the rest of the school using an appropriate index. They then might recommend that students in that major be interviewed and suggest some possible questions to ask based on their own personal experience of the barriers facing specific groups of students pursuing that major. However, personal experiences must be noted as such and should not be stated as fact. See Subsection 4.a for an example analysis.

3.c. Students were encouraged to create their own diversity measures

As described in Section 1, the purpose of the proposed project is to have students address a real, open-ended statistics problem facing a campus stakeholder and communicate their findings to non-statisticians. A key learning outcome was that students demonstrate proficiency addressing an open-ended problem using an appropriate statistical tool of their choice. As mentioned in Subsection 3.b, we encouraged students to derive a new measure that best addresses the problem as they understand it. We conclude this section with some additional resources we shared with individual students, who sought direction beyond what we reviewed in class and summarized in Subsection 3.b.

Some students were interested in comparing the diversity at the author's institution to the diversity of a specific reference population. For example, one student was interested in whether the diversity of veterans enrolled in the author's institution matched the diversity of the military. While there is rarely a single "correct" reference population, students may benefit from understanding how reference populations are used to establish statistical evidence of discrimination in court. For example, the statistical rule used for employment discrimination follows the Supreme Court's decision in *Hazelwood School District v. United States*, in which a binomial significance test was used to assess the difference between the proportion of black teachers employed by the Hazelwood School and the proportion of black teachers in the relevant labor market. See Meier et al. (1984) and the more student friendly summary by Zabell (1989).

Other students were interested in creating a more complicated measure of diversity. For example, one student was interested in the probability a class of thirty students would have at least two members from each of K predefined groups, assuming the students were randomly chosen. (The number two was chosen so that each group would be represented, and no student would be the sole individual representing their group.) This measure is similar to the indices described in Subsection 3.b in that students are "collected" and their differences assessed. We would refer students to the so-called "coupon collector's problem" (Ross 2014). Note that these problems can be tricky to solve analytically, but solutions are often easily obtained by simulation.

Section 4. Students were split over whether the authors' institution was diverse.

All sixteen reports investigated the state of diversity among undergraduate students. This was not surprising since the students authoring the reports were themselves undergraduates. Most students examined diversity in enrollment, however many also examined diversity among the graduating class or by the degrees they received.

Roughly half the class reported positively about diversity at the authors' institution, while the other half found the authors' institution lacked diversity in some crucial respect. The split over whether the author's institution was diverse depended on what variables the student used and the comparison the student made: Out of sixteen reports total, thirteen (81%) considered racial or ethnic diversity, while six (38%) focused on gender diversity. Nine (56%) found the authors' institution had much greater racial diversity than comparable universities. Seven (44%) reported on engineering schools specifically—most of those finding that the gender gap at the authors' institution was larger than comparable engineering schools. Other surprising or notable trends reflected barriers disproportionately facing students with disabilities, students with a specific immigration status, and older students.

The most common recommendation was to conduct further study. Other recommendations included recruitment practices and raising awareness of diversity among campus students, faculty, and staff. Several reports suggested that the lack of diversity was itself a barrier to inclusion. That is, students may be more comfortable participating with others who share similar identities and experiences.

4.a. An example student analysis that argued the authors' institution is diverse

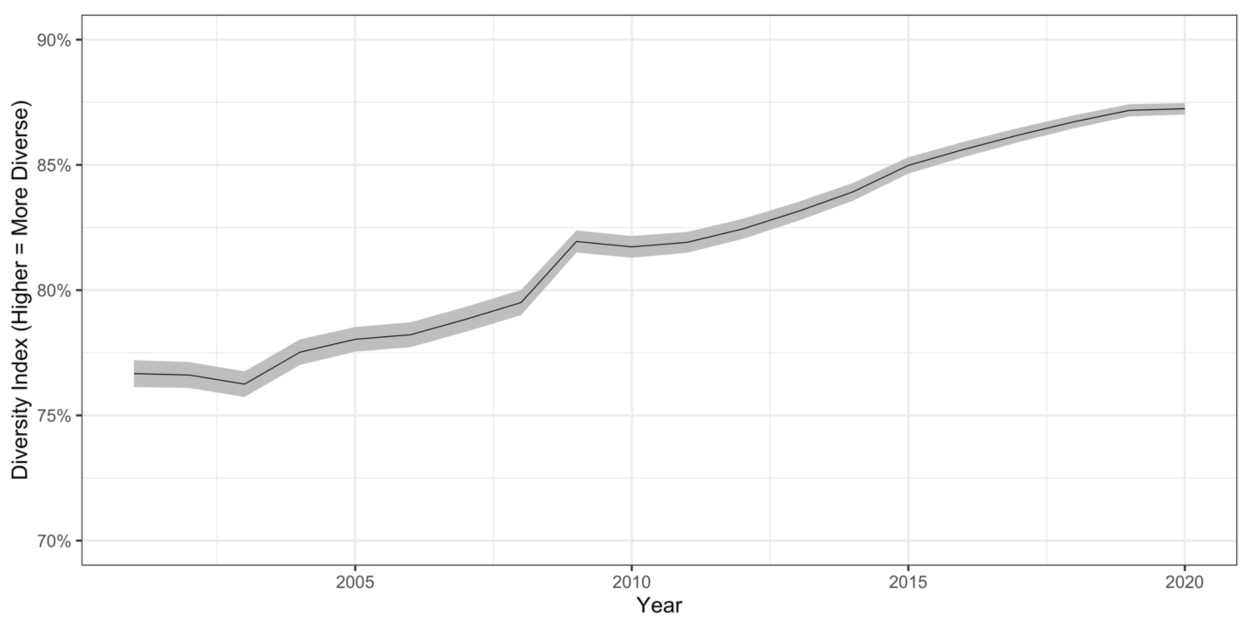
We outline a statistical analysis based on a project submitted by a student with their permission. The purpose is to provide an example of the sort of analysis that can be conducted. We note that the student had personal experiences with the barriers that arise by being classified under a specific immigration status. We do not include any additional nonstatistical details, such as references to the student's background and motivations for conducting their analysis. The `R` code necessary to reproduce the analysis and figures in this paper is included in the Appendix.

The analysis used the Diversity Index described in Subsection 3.b to monitor the diversity of the authors' institution over time. As mentioned in Section 3, the primary question the student sought to answer was whether the authors' institution is diverse. Secondary questions included whether diversity differed among undergraduates and graduates, whether diversity differed between part-time and full-time students, and whether diversity has been increasing over time. The student suspected that immigration status prevents students from enrolling in graduate school part time.

To answer these questions, students enrolled at the authors' institution between 2001 and 2020 were divided into 16 groups based on their gender (male or female) and race/ethnicity (whether they identified as White, Black, Hispanic, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Two or more races, or Nonresident alien). Students with Unknown race or ethnicity were not included.

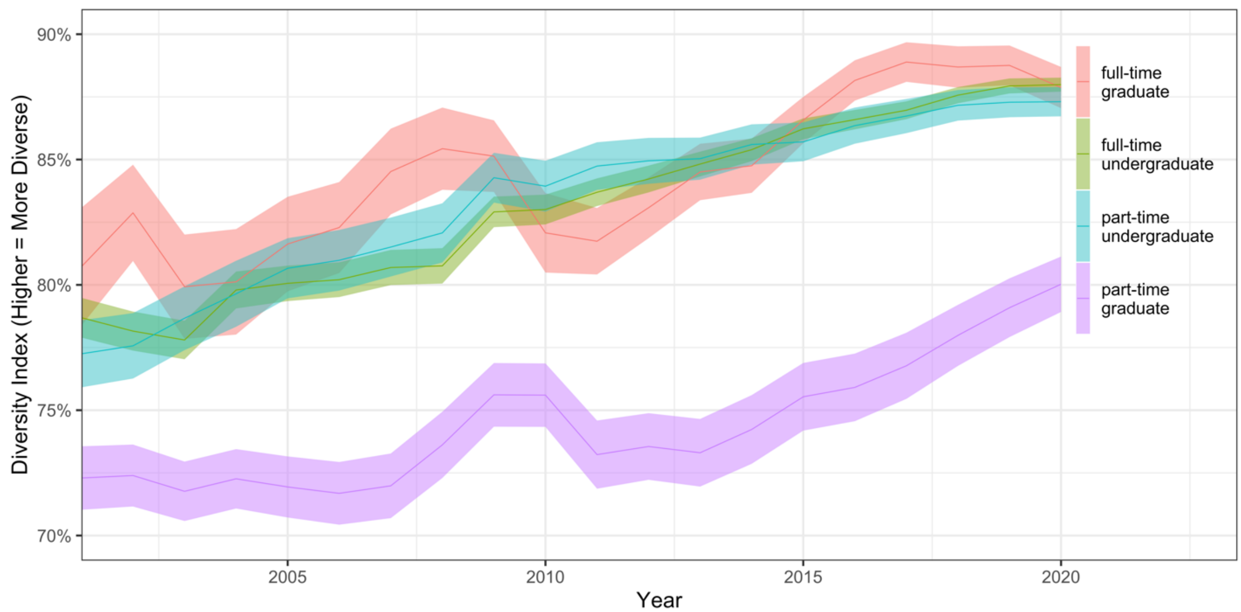
The diversity index was calculated for each year between 2001 and 2020, and a trend line was plotted as in Figure 1 below. The grey region represents a Bonferroni-corrected 95 percent confidence interval, calculated from standard errors described at the end of this subsection. The plot shows that the diversity index has increased markedly at the authors' institution from 2001 to 2020. If students were randomly paired in 2001, about one in four would identify as belonging to the same group. In 2020, the rate was one in nine. The student concluded the author's institution was diverse.

Figure 1: Diversity has increased markedly at authors' institution over the past two decades



The analysis was repeated separately for full-time undergraduates, part-time undergraduates, full-time graduates, and part-time graduates. The results are displayed in Figure 2. The main finding is that the Diversity Index has increased at approximately the same rate among all student segments. However, diversity among part-time graduates lags twenty years behind other groups. The student recommended that the DEI office investigate the source of this disparity.

Figure 2: Diversity among part-time graduate students lags twenty years behind others



The student described specific barriers to students interested in pursuing a graduate degree part-time that disproportionately affected students depending on their immigration status. These barriers did not affect part-time undergraduate students. The student suggested that the DEI office consider these barriers as the main source of the gap.

The purpose of the standard errors is to account for chance variation in the diversity of the student population that may arise between years, particularly in small populations. Students were modeled as being drawn at random from a superpopulation of possible students, and of interest is whether superpopulation parameters are changing over time.

Specifically, standard errors were calculated as follows. Let K denote the number of groups, and let p_i denote the probability a student identifies as a member of group i, $i = 1, \ldots, K$. Suppose two students had a chance encounter on campus. The probability those students identify as part of the same group is $\sum_{i=1}^{K} p_i^2$. The diversity index is $1 - \sum_{i=1}^{K} p_i^2$, the probability two students identify differently.

Let $[X_1, \ldots, X_K] \sim Multinomial([p_1, \ldots, p_K], n)$ denote the number of students enrolled in a given year who self-identify as a member of groups $i = 1, \ldots, K$, and let $\hat{p}_i = X_i / n$ denote the observed proportion of students that identify as belonging to group i.

It follows from the delta method that when n is large, the sample index, $1 - \sum_{i=1}^{K} \hat{p}_i^2$, is approximately normal with mean $1 - \sum_{i=1}^{K} p_i^2$ and variance

$$\frac{4}{n}\left(\sum_{i=1}^{K-1} p_i(p_i - p_K)^2 \; - \left[\sum_{i=1}^{K-1} p_i(p_i - p_K)\right]^2\right)$$

Substituting the sampling frequencies, $\hat{p}_i$ for the probabilities, yields the plug-in estimate of the variance, which was used to make Figures 1 and 2.

4.b. Post-project survey of participating students

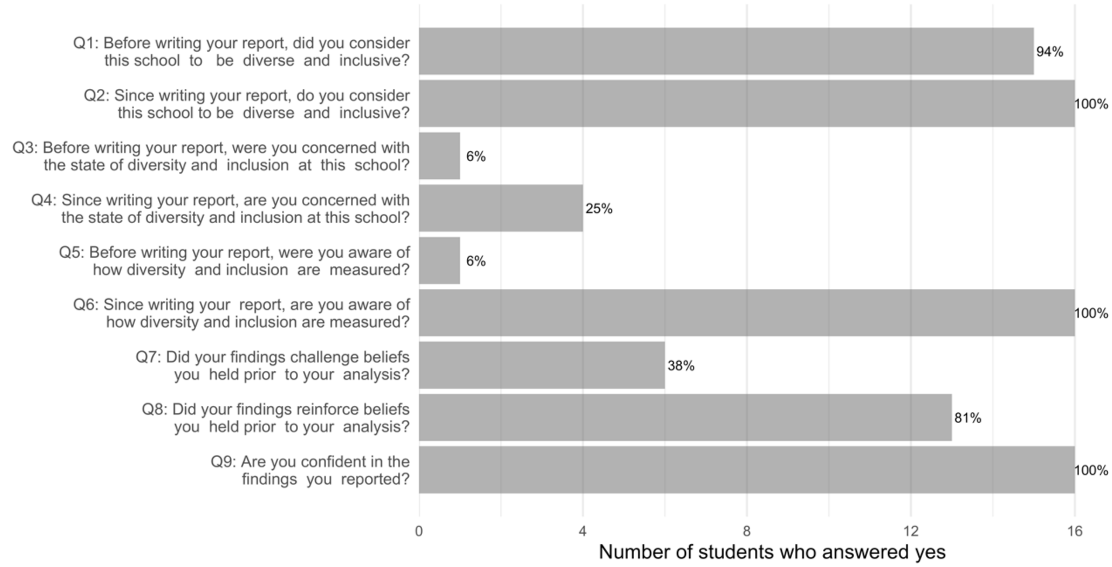
Following the project, we conducted a short follow-up survey to determine whether the assignment affected students' views. The survey was administered online using Microsoft Forms and took roughly two minutes for students to complete. To encourage truthful answers, participation was voluntary and all responses to the survey were anonymous. The survey was reviewed and approved by the authors' Institutional Review Board (IRBNet reference number 1972134-1). Note that the survey was administered to all students after completion of the project. No information was collected before the project was assigned, and there was no control group.

All students who completed the project chose to participate in the follow-up survey. The survey results are displayed in Figure 3 below. The results indicate that students' awareness of DEI increased following the project, as measured by several questions. Only 6% of students were aware of how diversity and inclusion are measured before the project (Question 5), compared to 100% after (Question 6). In addition, roughly one third of the students indicated that the project challenged beliefs they held prior to the analysis (Question 7).

That said, the authors' institution has more diverse enrollment than the typical university, and most students were already aware of that diversity. The project only slightly increased the percentage of students who viewed the authors' institution as diverse and inclusive from 94% (Question 1) to 100% (Questions 2). Most students (81%) indicated that the project reinforced beliefs they held prior to the analysis (Question 8).

At the same time, however, the authors' institution still lacks equity, and the assignment increased the percentage of students concerned about diversity and inclusion on campus—indicating more work is needed (Questions 3 and 4).

Figure 3: Students who participated in the project reported an increased awareness of diversity.



Although few students were aware of how diversity was measured before the project was assigned, all students who participated in the project were confident in their findings (Question 9). Confidence is a well-documented factor that determines whether students will use their skills in future coursework and in their chosen occupations (Han et al. 2021).

Discussion

In this paper, we proposed a project in which students use their data analysis and coding skills to investigate the state of diversity, equity, and inclusion (DEI) on campus and report their findings to the DEI office. We first provided some background information about the course for which this project was designed and the mission of the DEI office. We then described a version of the proposed project we taught at the authors' institution. Finally, we reviewed an example based on a student submission and a post-report survey in which the students reflected on the project.

The lesson proved rewarding to both the students and the DEI office. In light of the students' findings, the DEI office is considering the following actions

1. Reaching out to institutions with a smaller gender gap in engineering to discuss best practices and initiatives,

2. Working with offices on campus that facilitate adult learners and online learning, such as the offices of Continuing Studies and Teaching and Learning, and
3. Performing an audit on the current efforts to increase the retention of engineering students, identifying opportunities to retain traditional and nontraditional students.

Overall, the project ran smoothly. Students were well motivated by aiding the DEI office in fulfilling their mission. Nevertheless, two minor challenges did arise, which we believe should be considered when future versions of this project are conducted.

We found that students are limited by their own experiences. For example, undergraduate students largely experience the university through their interactions with other students—not through their interactions with faculty and staff. It is perhaps for this reason that no student chose to investigate the diversity of faculty or staff, even though such data were easily accessible. Future versions of the project might explicitly ask students to study faculty and staff diversity along with student diversity.

Another challenge was creating a uniform standard for evaluating and grading the wide variety of investigations the students conducted. In particular, it was difficult to set student expectations for how the assignment would be graded. Much of this difficulty arose from the fact that we intentionally placed little structure on the assignment to obtain the widest set of perspectives possible. While students appreciated the importance of this flexibility, several still expressed concern over whether their answer was “correct.” Future versions of this project might address this concern by showing past projects as an example. However, doing so might limit the creativity of the students’ investigations as students may feel pressured to emulate the example provided.

Despite these challenges, we believe the proposed project resonated with every student in the class. We conclude that students are an untapped resource that can help improve campus life as they develop their statistics and data science skills.

This resource may be more important now than ever. Representatives from twenty-eight states have introduced “anti-DEI” legislation since 2023 according to The Chronicle of Higher Education. Six states have restricted DEI offices: Alabama, Florida, Iowa, Texas, Utah, and Wyoming. For example, in 2024, Florida prohibited public colleges from spending state or federal funds on DEI unless required by federal law. For another example, the 2024 Wyoming budget eliminated state spending for the DEI office at the University of Wyoming (Chronicle of Higher Education 2024).

These DEI restrictions are for the most part aimed at college and university programs, not classroom activities. No legislation appears to materially limit the activity proposed in this paper: having students investigate the state of DEI on behalf of campus stakeholders. Students at institutions without a DEI office can still work with other offices concerned with student enrollment and success, such as the admissions office, the office of student life, and the office of career services.

We have argued in this paper that the proposed project is particularly important at institutions where DEI resources are limited. Indeed, the collection and reporting of race, ethnicity, and gender data are mandatory for all institutions that receive Federal financial assistance under the Civil Rights and Higher Education Acts—and these data constitute official statistics which, under Statistical Policy Directive 4 of the U.S. Office of Budget and Management, must be made accessible to the public equitably, reflecting the right of all U.S. citizens, students, and other stakeholders to public information. By participating in the proposed project, students provide campus stakeholders access to the information contained within the data, which limited resources would otherwise make inaccessible.

Disclosure Statement

The authors have no competing interests to disclose.

Data Availability Statement

Deidentified responses are available in the Appendix. The IRBNet reference number is 1972134-1.

References

American Statistical Association. (2014). Curriculum guidelines for undergraduate programs in statistical science. https://www.amstat.org/docs/default-source/amstat-documents/edu-guidelines2014-11-15.pdf

Angrave, L., Jensen, K., Zhang, Z., Mahipal, C., Mussulman, D., Schmitz, C.D., Baird, R.T., Liu, H., Sui, R., Wu, M.S., and Kooper, R. (2020). Improving Student Accessibility, Equity, Course Performance, and Lab Skills: How Introduction of ClassTranscribe is Changing Engineering Education at the University of Illinois. American Society for Engineering Education Annual Conference. https://peer.asee.org/34796

Asgarpoor, Jena Shafai, Meg Handley, Alisha L. Sarang-Sieminski, John Brooks Slaughter, Meagan C Pollock, Homero Murzi, and Monica Farmer Cox. (2021). Embracing Diversity, Equity, and Inclusion in Our Classroom and Teaching. American Society for Engineering Education Annual Conference. https://peer.asee.org/37024

Auerbach, J., & DeLazzero, C. E. (2022). Linked Data Detail a Gender Gap in STEM That Persists Across Time and Place. Harvard Data Science Review, 4(2). https://doi.org/10.1162/99608f92.49954afb

Carver, Robert, Michelle Everson, John Gabrosek, Nicholas Horton, Robin Lock, Megan Mocko, Allan Rossman, Ginger Holmes Roswell, Paul Velleman, Jeffrey Witmer and Beverly Wood. (2016). Guidelines for assessment and instruction in statistics education (GAISE) college report 2016. https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-(gaise)-reports

Chronicle of Higher Education. (2017). Diversity Index. Retrieved January, 15, 2023. https://www.chronicle.com/package/diversity-indexes

Chronicle of Higher Education. (2024). DEI Legislation Tracker. Retrieved August, 1, 2024. https://chronicle.com/article/here-are-the-states-where-lawmakers-are-seeking-to-ban-colleges-dei-efforts

De Brey, C., Musu L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., and Wang, X. (2019). Status and Trends in the Education of Racial and Ethnic Groups 2018. NCES 2019-038. National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019038

Dickens, J. C. (2021). The Data Science Public Schools Outreach Project: An Initiative for Diversity and Inclusion in Statistics. CHANCE, 34(1), 25-27. https://doi.org/10.1080/09332480.2021.1885930

Gray, C., de Cresce El Debs, L., Exter M., and Krause, T.S. (2016). Instructional Strategies for Incorporating Empathy in Transdisciplinary Technology Education. American Society for Engineering Education Annual Conference. https://peer.asee.org/25746

Han, J., Kelly T., and Knowles, J.G. (2021). Factors Influencing Student STEM Learning: Self-Efficacy and Outcome Expectancy, 21st Century Skills and Career Awareness. Journal for STEM Education Research, 4(2), 117-137. https://doi.org/10.1007/s41979-021-00053-3

Herrera, D., Leader C.M., Patel S., and Behrouzi, A. (2021). Student-Led Summer Diversity Workshops for Built-Environment Majors. American Society for Engineering Education Annual Conference. https://peer.asee.org/37766

Jensen, E., Jones, N., Orozco, K., Medina, L., Perry, M., Bolender, B. and Battle, K. (2021), "Measuring Racial and Ethnic Diversity for the 2020 Census." U.S. Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html

Lesser, L. M., & Nordenhaug, E. (2004). Ethical Statistics and Statistical Ethics: Making an Interdisciplinary Module. Journal of Statistics Education, 12(3). https://doi.org/10.1080/10691898.2004.11910630

Lesser, L. M. (2007). Critical Values and Transforming Data: Teaching Statistics with Social Justice. Journal of Statistics Education, 15(1). https://doi.org/10.1080/10691898.2007.11889454

Liao, S. M. (2023). SCRATCH to R: Toward an Inclusive Pedagogy in Teaching Coding. Journal of Statistics and Data Science Education, 31(1), 45–56. https://doi.org/10.1080/26939169.2022.2090467

Meier, P., Sacks, J., & Zabell, S. L. (1984). What happened in Hazelwood: Statistics, employment discrimination, and the 80% rule. American Bar Foundation Research Journal, 9(1), 139-186. https://doi.org/10.1111/j.1747-4469.1984.tb00900.x

Phelps, A. L., & Dostilio, L. (2008). Studying Student Benefits of Assigning a Service-Learning Project Compared to a Traditional Final Project in a Business Statistics Class. Journal of Statistics Education, 16(3). https://doi.org/10.1080/10691898.2008.11889574

Rawat, K., Lawrence E.E., and Gooden O.D. (2017). Mobile Aerospace Education Lab (m-AEL): A NASA Supported K-12 "Roadshow-In-A-Box" Initiative to Advance Aviation/Aerospace Education in Underserved Counties. American Society for Engineering Education Annual Conference. https://peer.asee.org/28675

Ross, Sheldon M. (2014). Introduction to probability models. Academic press. https://doi.org/10.1016/C2012-0-03564-8

Tijsma, G., Hilverda, F., Scheffelaar, A., Alders, S., Schoonmade, L., Blignaut, N., & Zweekhorst, M. (2020). Becoming productive 21st century citizens: A systematic review uncovering design principles for integrating community service learning into higher education courses. Educational Research, 62(4), 390–413. https://doi.org/10.1080/00131881.2020.1836987

Urban Institute. (2020). How Racially Representative Is Your College? Retrieved January, 15, 2023. https://apps.urban.org/features/college-racial-representation/

Urban Institute. (2020). Racial and Ethnic Representativeness of U.S. Postsecondary Education Institutions. Retrieved January, 15, 2023. https://datacatalog.urban.org/dataset/racial-and-ethnic-representativeness-us-postsecondary-education-institutions

Urban Institute. (2022). Education Data Explorer. Retrieved January, 15, 2023.
https://educationdata.urban.org/data-explorer/explorer
https://educationdata.urban.org/documentation/schools.html

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2016), Advancing diversity and inclusion in higher education: key data highlights focusing on race and ethnicity and promising practices. https://www2.ed.gov/rschstat/research/pubs/advancing-diversity-inclusion.pdf

U.S. Department of Education, National Center for Education Statistics. (2022). Integrated Postsecondary Education Data System (IPEDS). Retrieved January, 15, 2023. https://nces.ed.gov/ipeds/use-the-data

U.S. Department of Education, National Center for Education Statistics. (2022). IPEDS tutorials. Retrieved January, 15, 2023. https://nces.ed.gov/ipeds/use-the-data/overview-of-ipeds-data

Zabell, Sandy L. (1989). Statistical proof of employment discrimination. Statistics: A Guide to the Unknown. https://dl.acm.org/doi/book/10.5555/21791

Appendix

Student Responses to Survey

Question \ Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Q1: Before writing your report, did you consider this school to be diverse and inclusive?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q2: Since writing your report, do you consider this school to be diverse and inclusive?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q3: Before writing your report, were you concerned with the state of diversity and inclusion at this school?	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Q4: Since writing your report, are you concerned with the state of diversity and inclusion at this school?	Yes	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No	No	Yes
Q5: Before writing your report, were you aware of how diversity and inclusion are measured?	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No
Q6: Since writing your report, are you aware of how diversity and inclusion are measured?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q7: Did your findings challenge any beliefs you held prior to your analysis?	Yes	No	Yes	No	No	No	No	Yes	No	No	No	No	Yes	Yes	No	Yes
Q8: Did your findings reinforce any beliefs you held prior to your analysis?	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q9: Are you confident in the findings you reported?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

R Code to Recreate Example Analysis

```
library("educationdata")
library("tidyverse")

# IPEDS data read into R using Urban Institute educationdata package
# Source: https://educationdata.urban.org/documentation/colleges.html
# <college id> denotes the six-digit code that identifies the authors' institution

data <- get_education_data(level = "college-university",
                           source = "ipeds",
                           topic = "fall-enrollment",
                           filters = list(year = 2001:2020,
                                          unitid = <college id>),
                           subtopic = list("race", "sex"))

# Figure 1

 data_summary <-
  data %>%
  filter(level_of_study == 99,
         degree_seeking == 99,
         ftpt == 99,
         class_level == 99,
         sex != 99,
         race != 99,
         race != 9) %>%
  replace_na(list(enrollment_fall = 0)) %>%
  add_count(year,
            wt = enrollment_fall,
            name = "enrollment_total") %>%
  transmute(year,
            race,
            sex,
            n = enrollment_total,
            p = enrollment_fall / enrollment_total)

data_summary %>%
  filter(!(race == 8 & sex == 2)) %>%
  left_join(data_summary %>%
              filter(race == 8, sex == 2) %>%
              transmute(year, p_k = p)) %>%
  group_by(year) %>%
  summarize(index = 1 - sum(p^2) + first(p_k)^2,
            var = 4 / first(n) * (sum(p * (p - p_k)^2) - sum(p * (p - p_k))^2),
            se = sqrt(var)) %>%
  ggplot() +
  theme_bw(base_size = 20) +
  aes(year, index) +
  geom_ribbon(aes(ymin = index - qnorm(1 - .025 / 20) * se,
                  ymax = index + qnorm(1 - .025 / 20) * se),
              fill = "grey") +
  geom_line() +
  scale_y_continuous(name = "Diversity Index (higher = more diverse)",
                     labels = scales::percent,
                     limits = c(.7, .9)) +
  scale_x_continuous(name = "Year")

# Figure 2

data_summary <-
  data %>%
  filter(level_of_study != 99,
         degree_seeking == 99,
         ftpt != 99,
```

```
         class_level == 99,
         sex != 99,
         race != 99,
         race != 9) %>%
  replace_na(list(enrollment_fall = 0)) %>%
  add_count(year,
            level_of_study,
            ftpt,
            wt = enrollment_fall,
            name = "enrollment_total") %>%
  transmute(year,
            level_of_study,
            ftpt,
            race,
            sex,
            n = enrollment_total,
            p = enrollment_fall / enrollment_total)

data_summary %>%
  filter(level_of_study %in% 1:2) %>%
  filter(!(race == 8 & sex == 2)) %>%
  left_join(data_summary %>%
              filter(race == 8, sex == 2) %>%
              transmute(year, level_of_study, ftpt, p_k = p),
            by = c("year", "level_of_study", "ftpt")) %>%
  group_by(year, level_of_study, ftpt) %>%
  summarize(index = 1 - sum(p^2) + first(p_k)^2,
            var = 4 / first(n) * (sum(p * (p - p_k)^2) - sum(p * (p - p_k))^2),
            se = sqrt(var)) %>%
  ungroup() %>%
  mutate(label = case_when(
    level_of_study == 1 & ftpt == 1 ~ "\nfull-time\nundergraduate\n",
    level_of_study == 1 & ftpt == 2 ~ "\npart-time\nundergraduate\n",
    level_of_study == 2 & ftpt == 1 ~ "\nfull-time\ngraduate\n",
    level_of_study == 2 & ftpt == 2 ~ "\npart-time\ngraduate\n")) %>%
  mutate(label = fct_relevel(factor(label), "\npart-time\ngraduate\n", after = Inf)) %>%
  ggplot() +
  theme_bw(base_size = 20) +
  aes(year, index, label = label, color = label, fill = label) +
  geom_ribbon(aes(ymin = index - qnorm(1 - .025 / 60) * se,
                  ymax = index + qnorm(1 - .025 / 60) * se),
              alpha = .5, color = NA) +
  geom_line() +
  scale_y_continuous(name = "Diversity Index (Higher = More Diverse)",
                     labels = scales::percent,
                     limits = c(.7, .9)) +
  scale_x_continuous(name = "Year",
                     expand = c(0, 0, .18, 0)) +
  theme(legend.position = c(.92, .7),
        legend.background = element_rect(fill = NA)) +
  labs(fill = "", color = "")
```