Framework for Accessible and Inclusive Teaching Materials for Statistics and Data Science Courses

Mine Dogucu, Alicia A. Johnson & Miles Ott

To cite this article: Mine Dogucu, Alicia A. Johnson & Miles Ott (2023): Framework for Accessible and Inclusive Teaching Materials for Statistics and Data Science Courses, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2023.2165988

To link to this article: https://doi.org/10.1080/26939169.2023.2165988

© 2023 The Author(s). Published with license by Taylor and Francis Group, LLC.

View supplementary material

Published online: 22 Feb 2023.

Submit your article to this journal

Article views: 737

View related articles
Framework for Accessible and Inclusive Teaching Materials for Statistics and Data Science Courses

Mine Dogucua,b, Alicia A. Johnson1, and Miles Ott4

1Department of Statistical Science, University College London, London, UK; bDepartment of Statistics, University of California Irvine, Irvine, CA; cDepartment of Mathematics, Statistics, Computer Science, Macalester College, Saint Paul, MN; dTubi, San Francisco, CA

ABSTRACT
Despite rapid growth in the data science workforce, people of color, women, those with disabilities, and others remain underrepresented in, underserved by, and sometimes excluded from the field. This pattern prevents equal opportunities for individuals, while also creating products and policies that perpetuate inequality. Thus, it is critical that, as statistics and data science educators of the next generation, we center accessibility and inclusion throughout our curriculum, classroom environment, modes of assessment, course materials, and more. Though some common strategies apply across these areas, this article focuses on providing a framework for developing accessible and inclusive course materials (e.g., in-class activities, course manuals, lecture slides, etc.), with examples drawn from our experience co-writing a statistics textbook. In turn, this framework establishes a structure for holding ourselves accountable to these principles.

1. Introduction
According to LinkedIn’s U.S. Emerging Jobs Report, data scientists rank among the top emerging jobs with a 37% annual growth rate (2020). Yet the current STEM workforce, unfortunately, lacks gender and racial diversity. Though specifics vary by sector and company, statistics from two large employers are especially striking. Facebook’s Diversity Report notes that only 4.3% of technical roles are held by Hispanic employees, 1.7% by Black employees and 24.1% by women (2020). The figures are similarly low for Google’s technical roles where 0.7% are held by Black women, 1.1% by Latina women, and 0.2% by Native American women (2020). People with disabilities are also underrepresented in, and sometimes excluded from, the STEM workforce. For instance, whereas the Centers for Disease Control and Prevention estimate that 26% of U.S. adults have some type of disability (2020), only 6.1% of Google’s employees (technical and nontechnical) report having a disability.

Beyond the lack of equal opportunities for individuals in underserved or excluded groups, such a stunning lack of diversity in the STEM workforce can lead to products and policies that perpetuate bias and inequality (Noble 2018). For instance, statistical translation tools (e.g., Google Translate) yield male pronoun defaults more frequently (Prates et al. 2019). Facial recognition algorithms are much less accurate in identifying women in men, and darker-skinned people than lighter-skinned people (Raji and Buolamwini 2019). Autonomous vehicles fail to recognize behaviors of people with disabilities (Disability Rights UK 2021). After the death of a pedestrian by an autonomous vehicle while pushing her bicycle, scholars also raise concerns about the identification of people in wheelchairs (Whittaker et al. 2019).

The lack of diversity in the broader STEM workforce is due in part to a “leaky pipeline” (Preston 2004), and a good portion of this leak occurs throughout higher education. Students of color disproportionately attend community colleges and for-profit institutions as opposed to four-year institutions and thus they are not exposed to the same breadth of STEM curriculum and opportunities (Brenner et al. 2022). According to the American Statistical Association (ASA), only 42.9% of Bachelor’s degrees and 37.4% of Ph.D.’s in 2020, the most recent year of data collection, were earned by students that identify as women (2020). To this end, creating a healthy learning environment and departmental climate can factor into the persistence of students who identify with groups that have been underserved and excluded from STEM (Packard 2015). Yet, historically, the typical statistics graduate program has not provided significant training on the accessibility and inclusion principles that are critical to this goal. Further, related training provided by institutions of higher education typically (and necessarily) focus on general best practices. Thus, centering accessibility and inclusion in our statistics programs in particular requires the proactive alignment of these general practices with our training in statistics education. The goal of this article is to provide resources to that end.

A complete treatment of accessibility and inclusion in statistics education will require an ongoing series of articles and ongoing discussions—these principles must be woven
throughout our departments, curricula, classrooms, assessment, and course materials. Though some of the strategies herein apply more broadly, the current article focuses on the last of these: developing a framework for accessible and inclusive course materials (e.g., in-class activities, course manuals, lecture slides, etc). We share the specific goals behind this framework, as well as the processes and tools for reaching these goals. Throughout, we draw examples from our recent collaboration and original inspiration behind this work, a Bayesian statistics textbook for undergraduates and comparably trained practitioners (Johnson et al. 2022). Finally, it's important to note that the current article belongs to a growing body of work on similar topics. Among other resources, the Being Human in STEM (2022) initiative provides “a framework to understand and navigate diverse identities in the classroom, lab, and beyond.” The Data Science Course Framework outlines strategies to “build, scale, and sustain policies, practices, and structures that ensure that each student has equal access to, and successfully engages in” Data Science / STEM (Launch Years Initiative 2021).

2. Accessibility and Inclusion Goals

As with our textbook project, we believe that the development and delivery of statistics course materials should be driven by two broad goals.

Goal 1. Course materials should be physically accessible. In general, students cannot engage with materials that they cannot access. Just as we offer online, open access to our book, we recommend using open access materials in the classroom whenever possible. The reasons are many. First and foremost, many students can’t afford physical textbooks. It is estimated that 65% of students opt out of buying a college textbook and, among those students, 94% state that they suffer academically as a consequence (U.S. PIRG 2014). Second, open-access materials help engage a more global audience that might be outside the range of a publisher’s physical distribution, thereby increasing the range of perspectives in statistics and data science. Yet open-access delivery is not an exhaustive accessibility wand. It should also be a priority to make course materials accessible to students with color blindness or other visual impairments.

Goal 2. The development and delivery of course materials should be inclusive of a diverse body of learners. Our students don’t all share the same gender, race, cultural, socioeconomic class, immigration status, age, sexuality, and religious identities. They have varying interests, academic experiences, learning styles, and goals. We want all students to be seen, supported, and encouraged to engage in our classes. More broadly, we want to create a sense of belonging, not only in our work together, but in the broader statistics and data science community. We follow several strategies to this end:

• showcase the diversity of the field through a broad group of scholars;
• use inclusive language, assumptions, and examples;
• use active learning approaches that encourage students to learn by doing;
• embrace the challenges and failures which are critical to learning; and
• build rapport.

3. Accessibility and Inclusion Approaches

Creating more inclusive and accessible course materials simply does not happen without intention and accountability. Throughout our book writing and editing process, we developed and returned to a set of specific strategies. This same framework can be used in developing course materials for the statistics and data science classroom. We introduce these strategies here, along with tools for their implementation. Though some of the tools are based in R, the broader themes should resonate no matter your preferred software.

3.1. Increasing Accessibility

3.1.1. Open-Access Materials

For instructors developing their own open-access course materials, textbooks, or other educational materials, a few (free!) tools can help streamline the publishing process. We recommend three tools below along with helpful resources for their implementation. Though there’s definitely a learning curve, the benefits of investing time to learn these tools extend beyond open-access educational materials—the same workflow can be used for personal websites and blogs.

• For writing the materials: R Markdown allows users to seamlessly weave text, figures, R code, R output, and more into the same reproducible document. In turn, the R bookdown package (Xie 2021a) allows users to build longer course materials from R Markdown documents. For those who already use R Markdown, the documentation for bookdown is quite extensive (Xie 2016). For those who are not familiar with R Markdown, RStudio’s online tutorials provide a starting point (R Studio n.d.).

• For collaboration with colleagues / coauthors: Bryan et al. (2021) is an excellent resource for learning GitHub and version control.

• For deploying materials online: Hvitfeldt’s blog post (2020) provides step-by-step instructions for deploying online materials with Netlify.

3.1.2. Color Blind Accessibility

Figures and data visualizations are critical components of data analysis and communication and hence of the statistics and data science classroom. Interpreting and learning from these figures can require the distinction between multiple colors that, for example, represent different groups or models. We recommend two strategies for making such figures more accessible to students with color blindness. First, use more color blind friendly color palettes. For example, for those that use the ggplot2 package in R to create figures for their course materials, it is important to know that the default color palette is not accessible to the most common types of color blindness (e.g., red-green color blindness). Instead, we recommend the color palette suggested
by Okabe and Ito (2008). The `palette.colors()` function lists the Okabe-Ito color palette in both hex code and R form.

```r
# Check out the color palette
palette.colors(palette = 'Okabe-Ito')
## black orange skyblue bluishgreen yellow
## '#000000' '#E69F00' '#56B4E9' '#009E73' '#F0E442'
## blue vermillion reddishpurple gray
## '#0072B2' '#D55E00' '#CC79A7' '#999999'
```

When using R Markdown or bookdown to create course materials, placing the following code at the front of the document sets this as the default color palette throughout:

```r
# Set the ggplot2 color palette
palette('Okabe-Ito')
scale_colour_discrete <- function(...)
  scale_colour_manual(values = palette())
scale_fill_discrete <- function(...)
  scale_fill_manual(values = palette())
```

A second recommendation in making figures more color blind friendly is to test your color choices with a color blindness simulator (Colblindor 2021). These simulators help those that don’t have any color vision deficiencies learn how an image looks to people that do. For instance, Figure 1 displays an original plot that uses our color palette of choice (left) alongside a simulated version (right) of how this plot would look to someone who has green-blind color blindness, that is, deuteranopia. In this case, we learn that some people will not distinguish the colors here as “yellow,” “green,” and “blue” (thus, we shouldn’t rely on those descriptions in our discussion), yet they will view these colors as distinct.

### 3.1.3. Screen Reader Accessibility

Teaching statistics to blind and visually impaired students comes with a set of considerations (Godfrey and Loots 2015). The development of course material should be mindful of students with visual impairments that rely on screen reader software that reads aloud the content on a computer screen. We highly encourage educators without vision impairments to test out a screen reader to experience how inaccessible course materials can be. In general, screen readers are effective at reading basic text and code, yet cannot automatically process another major component of the typical statistics and data science resource: figures. Unless a figure has accompanying alt (or alternative) text, it is read simply as an “image” by a screen reader. As such, alt text which describes the contents of an image for screen readers is an important supplement to figure captions and discussions which appear within the text.

As of version 1.32, the `knitr` (Xie 2021b) package behind R Markdown documents supports alt text (Hill et al. 2021). Consider the bar plot in Figure 2 which displays the number of artists in the Museum of Modern Art (MoMA) collection that are still living (TRUE) versus those that aren’t (FALSE). We created this figure, along with alt text, using the code below. Whereas the figure caption provides a short description of the image which appears in the text and is read by screen readers, the alt text is only read by screen readers. Our understanding of writing effective alt text continues to evolve. To this end, we have found the talk by Canelón and Hare (2021b) and the supplementary resources in the GitHub repo (2021a) to be particularly helpful.

---

1The authors requested this feature through a GitHub issue.
As an alternative to writing your own alt text, resources such as the `BrailleR` (Godfrey et al. 2021) package provide automated alt text. For example, the `VI()` function in `BrailleR` automates the following alt text for Figure 2.

```
BrailleR::VI(example_barplot)
```

```
## This is an untitled chart with no subtitle or caption.
## It has x-axis 'alive' with labels FALSE and TRUE.
## It has y-axis 'count' with labels 0, 2000 and 4000.
## The chart is a bar chart with 2 vertical bars.
## Bar 1 is centered horizontally at FALSE, and spans vertically from 0 to 5584.
## Bar 2 is centered horizontally at TRUE, and spans vertically from 0 to 5380.
```

![Figure 2](image.png)

Figure 2. The number of living (TRUE) and nonliving (FALSE) artists in the MoMA collection.

Resources like `BrailleR` do provide an efficient approach to including alt text, and are especially appealing when developing course materials with lots of figures. Yet the current automated alt text toolbox has drawbacks. First, though it can be effective for simple plots like that in Figure 2, automated alt text might not adequately summarize the nuances and themes within a figure. Second, automated alt text tools are not yet compatible with more advanced plotting options (e.g., the Markov chain trace plots within our book).

Finally, in addition to providing alt text, we recommend improving the screen reader experience by removing the hashtags in R output. By default, R displays output with double hashtags:

```
mean(c(2, 3, 4))
```

```
## [1] 3
```

These hashtags are then read by screen readers, unnecessarily distracting from the code. When creating documents within R Markdown or bookdown, setting the global `comment` option to blank (`""`) at the start of the document removes the hashtags throughout:

```
knitr::opts_chunk$set(comment = "")
```

The result applied to our mean calculation is illustrated below:

```
mean(c(2, 3, 4))
```

```
[1] 3
```

### 3.2. Increasing Inclusivity

#### 3.2.1. Showcasing the Diversity within Statistics and Data Science

One simple strategy for developing and delivering course materials that are more inclusive of students with different identities
and experiences is to make them more inclusive of scholars in the fields of statistics and data science. The Data Feminism book (D’Ignazio and Klein 2020a) provides inspiration here. At the onset of their project, the authors outlined a set of values and aspirational metrics for holding themselves accountable to these values (D’Ignazio and Klein 2020b). For example, to address structural racism, the authors aspired to have at least 75% of their citations of feminist scholarship to be from people of color.

With or without numerical diversity metrics, being intentional about the scholars we cite and discuss in our course materials prevents the perpetual visibility of a narrow, homogeneous, and extensively cited group at the expense of others who are also doing groundbreaking work. For example, throughout the writing of our book, we committed to citing scholars that represent diverse fields and identities. To this end, several organizations support and highlight the work, both present and historical, of diverse communities within statistics and data science. We find the following organizations to provide particularly invaluable resources (e.g., articles, talks, and profiles) for learning more about this work ourselves and incorporating it into our classrooms: Mathematically Gifted and Black,2 the Society for Advancement of Chicanos/Hispanics and Native Americans in Science,3 the ASA Committee on Minorities in Statistics,4 the ASA Committee on Women in Statistics,5 and the ASAs new Justice, Equity, Diversity, and Inclusion Outreach group (JEDI).6

Not only is it critical to be mindful of who we cite and highlight in our course materials, but how we cite. For example, since many publishers make it difficult for authors to correct their names after publication, trans scholars might have historical publications under their deadnames (Tanenbaum 2020). Further, we cannot infer a scholar’s gender identity by their name alone. As such, we recommend taking extra steps to verify that you are using the correct name and pronouns in any citation. Though this process is difficult for deceased scholars, we rely on personal websites, ORCID IDs, and social media accounts to help verify this information for living scholars. When in doubt, we use they/them pronouns.

3.2.2. Using Inclusive Language, Assumptions, and Examples

In developing and delivering course materials, our language, assumptions, and examples should be mindful of two facts: not all students share the same specialized knowledge; and, not all students share the same cultural or personal identities and lived experiences. Consider the first of these. Statistical applications often require deep, specialized knowledge (e.g., about biology, physics, sports, etc.). Assuming that all students share the same specialized knowledge is problematic for a couple of reasons. First, trying to teach a technical statistical concept through a specialized context is like trying to teach kids to read using passages about retirement plans—it’s challenging if not impossible to learn a concept when it’s delivered within a context that you don’t understand. Second, specialized settings risk alienating students which, in turn, can detract from the learning process.

To this end, we recommend showcasing applications that are either fairly universal or at least have a well explained context. For example, in writing our textbook, we were considerate of the fact that we teach many international students and that we have readers outside the United States. As such, we eliminated some of our original examples which required niche knowledge about the United States (e.g., about the unique primary and caucus phase of U.S. election cycles). Though we do bring a variety of perspectives and experiences to our book and classrooms, we also acknowledge that evaluating our own work along these lines, and recognizing our own niche perspectives, is challenging.

Intentionality is also critical in developing course materials that are inclusive with respect to cultural or personal identities and lived experiences. Consider a few representative but non-exhaustive recommendations based on strategies we used throughout our book. First, in developing examples and exercises, avoid cultural stereotypes and use names that are associated with different cultures. For example, in addition to Matt and Taylor, our book’s discussions include Muhammad, Zuofu, Kimya, and Fernando. Second, be mindful of the evolving and more inclusive understanding of sexual orientation and gender identity (Phillips et al. 2019). For example, we employed three strategies to avoid heteronormative defaults in our book: (a) our examples span topics from same-sex marriage to LGBTQ+ anti-discrimination laws; (b) we use three singular person pronouns, he/she/they, when referring to people in our examples; and (c) as recommended by Thornton et al. (2019), we avoid examples and datasets that treat gender as binary.

3.2.3. Embracing Challenges and Failures

In the classroom, assumptions and statements about what should be “obvious” to our students can be both wrong and demoralizing. The same goes for the materials we use and develop for our classes. To this end, we recommend two important strategies. First, don’t just normalize, but actively embrace the mistakes and questions that are critical to learning. For example, our book explicitly states up front that: “As you read the book and put Bayesian methodology into practice, you will make mistakes. Many mistakes. Making and learning from mistakes is simply part of learning.” As another example, we note in the second chapter of the book that, no matter their level of previous experience, readers will be introduced to many new vocabulary terms and concepts, and so should expect to take this chapter slowly. When used in classroom conversations, policies, and materials, such language can help counter any impostor syndrome by normalizing the fact that humans don’t magically ingest and process information—learning takes time and iteration. Second, use language that acknowledges and welcomes students with a variety of technical or academic backgrounds. For example, in our book as in our course materials, we don’t assume that all students have seen the Greek alphabet, thus, give students a heads up on Greek letters and how to pronounce them. Similarly, in our course materials, we weave in reminders of definitions and notation from the course prerequisites.

---

2 https://mathematicallygiftedandblack.com/
3 https://www.sacnas.org/
4 https://community.amstat.org/cmis/home
5 https://community.amstat.org/cowis/home
6 http://www.datascijedi.org/
students feel more seen themselves. For example, our book logo is a somewhat silly disco ball, a loose analogy of the Bayesian philosophy. We use casual imagery approach here is to just be ourselves, not statistics robots.

3.2.5. Building Rapport

Our final strategy for creating inclusive course materials is to build rapport. In the classroom, rapport can be fostered through one-on-one conversations, small group activities, and other opportunities for conversation and sharing. Though building rapport through written course materials might be tougher to imagine, it’s equally important and possible. Our simple approach here is to just to be ourselves, not statistics robots. For example, our book logo is a somewhat silly disco ball, a loose analogy of the Bayesian philosophy. We use casual imagery and analogies to build intuition for technical concepts. Similarly, our discussions tend to be conversational in tone. In general, and by sharing some of our own personal identities, we are almost certainly missing goals without realizing it. As educators, we also expect our understanding of accessibility and inclusion to evolve as we continually seek to improve our teaching practices. This checklist will become outdated or insufficient quite soon. With this understanding, we consider the framework we’ve provided here to be a starting point for continued collaboration and conversation with other statistics and data science educators. As our understanding and practice around accessibility and inclusion evolve, we welcome any feedback on these topics through a Google Form.\footnote{The Google Form can be found at \url{https://bit.ly/bayesrules-feedback}.}

### Supplementary Materials

We provide the R session information used writing this manuscript.

### References


Bryan, J., STAT 545 TAs and Hester, J. (2021), Happy Git and GitHub for the useR. Available at \url{https://happygitwithr.com}


Canelon, S., and Hare, E. (2021b), “Revealing Room for Improvement in Accessibility within a Social Media Data Visualization Community,” Available at \url{https://www.youtube.com/watch?v=DxLkv2Rdj8e&ab_channel=csvconf}.csvconf.


### Table 1.

<table>
<thead>
<tr>
<th>Accessibility and inclusion criteria</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>Is the cost affordable for learners from diverse socioeconomic backgrounds?</td>
</tr>
<tr>
<td></td>
<td>Are plots distinguishable to color blind learners?</td>
</tr>
<tr>
<td></td>
<td>Is alt text provided for images?</td>
</tr>
<tr>
<td>Inclusivity of scholars</td>
<td>Do the cited scholars represent diversity across identities, experiences, and expertise?</td>
</tr>
<tr>
<td></td>
<td>Are scholars cited using the correct names and pronouns?</td>
</tr>
<tr>
<td>Inclusivity of students</td>
<td>Do examples avoid the necessity of specialized knowledge?</td>
</tr>
<tr>
<td></td>
<td>Do names and pronouns reflect diverse cultural and personal identities?</td>
</tr>
<tr>
<td></td>
<td>Are there examples that could potentially speak to younger as well as older students?</td>
</tr>
<tr>
<td></td>
<td>Does the delivery embrace mistakes and critical thinking?</td>
</tr>
<tr>
<td></td>
<td>Are efforts made to accommodate different academic experiences and create a shared foundation?</td>
</tr>
</tbody>
</table>

4. Concluding Remarks

Table 1 summarizes our strategies to be intentional about, and hold ourselves accountable to, our accessibility and inclusion goals. Though we originally developed these strategies for use in the writing and revision of our book, we have since used them in developing course materials for our classrooms. It is possible that someone reading our book might disagree that we met all of these criteria. With the subjective measures as well as our own personal identities and experiences, we are almost certainly missing goals without realizing it. As educators, we also expect our understanding of accessibility and inclusion to evolve as we continually seek to improve our teaching practices. This checklist will become outdated or insufficient quite soon. With this understanding, we consider the framework we’ve provided here to be a starting point for continued collaboration and conversation with other statistics and data science educators. As our understanding and practice around accessibility and inclusion evolve, we welcome any feedback on these topics through a Google Form.\footnote{The Google Form can be found at \url{https://bit.ly/bayesrules-feedback}.}