Teaching Visual Accessibility in Introductory Data Science Classes with Multi-Modal Data Representations

JOOYOUNG SEO,∗ AND MINE DOGUCU

1School of Information Sciences, University of Illinois Urbana-Champaign, 614 E. Daniel St., Champaign, IL 61820, USA
2Department of Statistical Science, University College London, UK
3Department of Statistics, University of California Irvine, USA

Abstract
Although there are various ways to represent data patterns and models, visualization has been primarily taught in many data science courses for its efficiency. Such vision-dependent output may cause critical barriers against those who are blind and visually impaired and people with learning disabilities. We argue that instructors need to teach multiple data representation methods so that all students can produce data products that are more accessible. In this paper, we argue that accessibility should be taught as early as the introductory course as part of the data science curriculum so that regardless of whether learners major in data science or not, they can have foundational exposure to accessibility. As data science educators who teach accessibility as part of our lower-division courses in two different institutions, we share specific examples that can be utilized by other data science instructors.

Keywords curriculum; data representations; R

1 Introduction
According to LinkedIn’s U.S. Emerging Jobs Report, data scientists rank among the top emerging jobs (2020). With the urgent need for training a higher number of skilled data scientists, many institutes of higher education are developing their own data science curricula. However, a consensus on what should be included in such curricula has not yet been reached. One of the few guidelines on the topic is written by the American Computing Machinery Data Science Task Force and titled Computing Competencies for Undergraduate Data Science Curricula. In this report, accessibility is a “foundational consideration” at the knowledge level when it comes to displaying data (Danyluk et al. 2021).

In fact, training on accessibility has been increasingly becoming an important job-readiness aspect in computing and information sciences (Shinohara et al. 2018; Kawas, Vonessen, and Ko 2019). For instance, some notable big tech companies, such as Yahoo!, Facebook, Microsoft, Google, and Adobe, have been leading a TeachAccess (teachaccess.org) Consortia to represent this increased demand for accessibility knowledge and to increase the number of computer and information science faculty teaching accessibility (2018). The Partnership on Employment & Accessible Technology (PEAT), in their Accessible Technology Skills Gap Report (2018), also highlights this trend—84% of industry correspondents that they work with say “it is important or

∗Corresponding author. Email: jseo1005@illinois.edu.

© 2023 The Author(s). Published by the School of Statistics and the Center for Applied Statistics, Renmin University of China. Open access article under the CC BY license.
Received July 31, 2022; Accepted February 27, 2023
very important for them to hire developers and designers with accessible technology skills.” The growing interest in and importance of teaching accessibility can also be found in professional education societies. For example, accessibility has been recently added to engineering design accreditation requirements by Accreditation Board for Engineering and Technology [ABET; abet.org] (Shinohara et al. 2018).

Even though this statistic may be different specifically for data scientists, data science as a highly interdisciplinary field across computing and information sciences also shares the rising demand for accessibility preparedness. Furthermore, data science is broad with the intersection in careers such as developers (e.g., data dashboard and package developers) and designers (e.g., data visualization experts). Therefore, the impact of teaching and learning accessibility is as far-reaching as its wide scope.

Data science consists of an iterative process of importing, tidying, transforming, visualizing, and modeling data (Wickham and Grolemund 2016). While each stage has unique challenges, data visualization may pose fundamental barriers against people with permanent, situational, and temporary disabilities (e.g., visual, cognitive, and motor impairments, to name a few) in accessing information (Marriott et al. 2021; Lee et al. 2020). For example, people with cognitive and learning disabilities often have difficulties understanding the abstractions and symbolic conventions used in visualizations, and people lacking fine motor coordination or having tremors may be unable to operate interactive visualizations, not to mention people who are blind and have low vision facing extra challenges in visualization access (Marriott et al. 2021; Lee et al. 2020). Marriott et al. (2021), in their call to action paper, highlights the serious impact of inaccessible visualizations: “This is an equity issue with severe consequences because a lack of access to visualizations and underlying data impacts health, educational and work opportunities, and lifestyle” (p. 48).

Despite the increasing demand for accessibility knowledge in both industry and academia and the critical impact of accessible visualizations, accessibility has not yet found its place in the data science course and teaching materials. One main reason for this could be that accessibility is often incorporated into human-computer interaction (HCI) and design courses (Shinohara, Bennett, and Wobbrock 2016; Shinohara et al. 2017), introductory web development (Ko and Ladner 2016; Rosmaita 2006), software engineering courses (Martin-Escalona, Barcelo-Arroyo, and Zola 2013; Stephanie Ludi 2007; S. Ludi 2002) or the digital arts (Barata et al. 2019), if at all covered. However, data science is an interdisciplinary field that is taught in various disciplines, including not only computer and information sciences but also statistics, business schools, political science, biological sciences, and more (Schwab-McCoy, Baker, and Gasper 2021). Therefore, we posit that the resources and conversations around accessibility have not yet extended to the broader data science education community at the scale that they should. For instance, many introductory data science courses (Yan and Davis 2019; Baumer 2015), textbooks (Adhikari, DeNero, and Wagner 2019; Wickham and Grolemund 2016), or curricula (Schwab-McCoy, Baker, and Gasper 2021) either fail to mention accessibility at all or only focus on accessing the data.

In response to the need to develop accessible teaching materials within the data science discipline, this paper demonstrates and shares some teaching strategies that we have employed in our teaching introductory data science courses. In this manuscript, we argue that accessibility should be taught early in the data science curriculum so that regardless of whether learners major in data science or not, they can have foundational exposure to accessibility.

One may wonder why we need to teach accessibility as part of an introductory data science course as opposed to covering accessibility in a separate and general course module. According
to the recent faculty survey on teaching accessibility (Shinohara et al. 2018), the most critical barriers to teaching accessibility reported by 1,857 faculty from 318 institutions were “the absence of clear and discipline-specific accessibility [emphasis added] learning objectives and the lack of faculty knowledge about accessibility” (p. 197). This reveals faculty often have trouble making sense of how they can integrate accessibility into their specific teaching domains, which requires applied knowledge of accessibility beyond general concepts.

We share our experiences and strategies in the hope that our discipline-specific (i.e., data science and data visualization) accessibility teaching cases can be used as scaffolding guides for other data science instructors to build upon.

Another important aspect of curriculum design includes the timing of when to introduce accessibility concepts in data science courses. For example, some instructors may prefer to put teaching accessibility on the back burner so that they can save more time for their students to learn data science fundamentals first and adopt accessibility principles in an upper-level course later. While the decision is subject to each institution’s and individual instructor’s discussion, we believe the sooner, the better. This is because: (1) some students may take an introductory level only, which means they will lose a chance to get exposed to this demanding knowledge unless they elect the upper-level course covering accessibility; (2) this kind of separation may unintentionally contribute to misconceptions, such as accessibility is difficult and/or accessibility is optional, not essential. When properly incorporated, accessibility can be taught along with other data science fundamentals (e.g., wrangling, merging, tidying, visualizing) while not necessarily putting more cognitive burden on students’ current learning. Rather, students can take a chance to become creative design thinkers by making their data science work more meaningful and accessible through “empathy learning” (Putnam et al. 2015). Thus, we suggest teaching accessibility as part of an introductory data science course for students to practice coupling their data science knowledge with accessibility principles as early as possible.

As data science educators who teach accessibility as part of our lower-division courses in two different institutions, we share specific examples that can be utilized by other data science instructors. We teach R (R Core Team 2022) as the main language in our courses; thus, our examples are shaped around using R. However, instructors using a different set of language(s) or even teaching language-agnostic courses may still find the content applicable to their own courses.

2 Learning Goals and Objectives

Although there are various ways to represent data patterns and models, visualization has been primarily taught in many data science courses for its efficiency (Kim et al. 2021). Such vision-dependent output may cause critical barriers against not only blind and visually impaired people but also people with cognitive and learning impairments and people having motor impairments (Kim et al. 2021; Marriott et al. 2021; Lee et al. 2020). We argue that instructors need to teach multiple data representation methods so that all students can produce data products that are more accessible. To this end, we have considered multi-data representations and their relation to accessibility in our courses, and that’s what we will demonstrate throughout this paper.

In incorporating accessibility to our course contents, we also wanted our students to be familiar with assistive technologies, and since our focus has been on visual accessibility, it was a natural fit to include screen readers. Screen reader is an assistive technology that supports blind or visually impaired people in using their computer by reading aloud the contents. Recently, screen reader has been provided as one of the built-in accessibility features on operating systems, such as Microsoft Windows Narrator, Apple iOS VoiceOver, Android TalkBak and Linux Gnome Orca.
Seo, J. and Dogucu, M.

However, we also want to make our teaching as inclusive as possible to other types of dis/abilities so that we are not just confined to issues of visual impairments. Therefore, our curricular goal is to broadly ensure that the current and the next generation of data scientists provide public-facing outputs, including websites and analysis reports, in accessible forms and representations for people with diverse abilities. To achieve this goal, we set three specific and minimal learning objectives like below.

Students should:
1. get familiar with Americans with Disabilities Act;
2. use at least one assistive technology;
3. consider different representations of data.

The first objective provides our students with a chance to learn legislative concepts and empathy building concerning diversity issues. As both authors teach data science in the United States, it was a natural choice for us to make students aware of the Americans with Disabilities Act (ADA). However, many countries these days may have their own laws similar to ADA, such as Canada’s Accessibility for Ontarians with Disabilities Act, European Accessibility Act, United Kingdom’s Equality Act, Ireland’s Equal Status Acts, Australia’s Disability Discrimination Act, Israel’s Equal Rights of Persons with Disabilities Act, South Korea’s Act on Welfare of Persons with Disabilities, and Japan’s Elimination of Discrimination against Persons with Disabilities Act to name a few. The ADA is a civil rights law that prohibits discrimination against individuals with disabilities. Regardless of their own disability status, as future data scientists, we expect our students to produce data science products that are accessible to all, not only because it is the right thing to do but also because they may have future responsibilities as data scientists to comply with ADA.

The second learning objective echoes our argument on accessible data representation beyond one-type modality. People with disabilities may rely on different types of assistive technology in their day-to-day lives. Assistive technology (AT) is defined as “products, equipment, and systems that enhance learning, working, and daily living for persons with disabilities” (Assistive Technology Industry Association; https://bit.ly/3fJN2aJ). Examples of assistive technology include, but are not limited to, walking sticks, wheelchairs, screen readers, foot-activated switches, joy-sticks, sip-and-puff controllers, eye-tracking devices, and speech recognition input. As we noted above, we introduced screen reader as one of these examples when our students were working on a data visualization project; however, we acknowledge that various assistive technologies, such as speech recognition and video captioning, could be integrated into data science courses in diverse contexts (e.g., data collection and data presentation contexts).

It is important to note that the first two objectives are not data science specific but are related to accessibility on a broader scale. The third objective, however, is data science specific. In the remaining section of this paper, we will focus further on the third objective by illustrating how we have taught multiple data representations.

2.1 Data Visualization

Data visualization is the most common representation of data. Despite their use by many, data visualizations are not accessible to all. Colors are an important part of data visualizations and an aspect that can limit accessibility for color-blind people. When teaching data visualizations, we include three basic rules for incorporating colors into data visualization. Students should be able to
1. simulate color blindness (if they are sighted and not color blind);
Figure 1: Sample scatterplot showing the relationship between flipper length (mm) and bill length (mm) of penguins for three different species of penguins.

2. pick color blind friendly colors;
3. not only rely on color for differentiating data.

We show an example of incorporating these three basic rules by using the Palmer Penguins data (Horst, Hill, and Gorman 2022). We visualize the relationship between flipper length and bill length of penguins and color the data points differently for each species in Figure 1.

The colorblindr package (McWhite and Wilke 2022) provides a set of functions that can be useful in teaching and practicing color blindness accessibility. For instance, `colorblindr::cvd_grid()` function provides color-deficiency simulations of a given plot. In the code chunk below, the plot is named as `fig`. The code creates a grid of four plots where each plot shows a different color vision deficiency, including deuteranomaly, protanomaly, protanomaly, and desaturated. This grid allows those without any color vision deficiencies to experience what data visualizations would look like to a sighted person that has a color deficiency.

```
colorblindr::cvd_grid(fig)
```

For their data visualizations, instead of relying on the default color selection of R packages, students are introduced to the Okabe-Ito color palette, which is known to be accessible to people with color vision deficiencies (Okabe and Ito 2008). One can find specific colors in the palette by the R code `palette.colors(palette = "Okabe-Ito")`. In addition, students can be introduced to `colorblindr::scale_color_OkabeIto()` and `colorblindr::scale_fill_OkabeIto()` to automatically switch to using this color palette by adding a layer to a ggplot presentation. For instance, the plot in Figure 1 can be switched to the Okabe-Ito color palette as shown in Figure 3 by using the code below.

```
fig + colorblindr::scale_color_OkabeIto()
```

Last but not least, students should learn not to rely only on color to differentiate data points, in this case, Species. For instance, they can be introduced to shapes and have data
points that are circles, triangles, and squares for each species. This can also be extended to line graphs as having continuous or dashed lines. Facetting can also be utilized. For instance, each of the species could have its own scatterplot side-by-side.

In this manuscript and in our introductory courses, we focus on static data visualizations. Those who teach interactive data visualizations may also consider covering cognitive and motor accessibility. Our focus here is on visual accessibility, but it is worth noting that all subdomains of accessibility are relevant to data science learners and can be incorporated into the curriculum as seen fit.

2.2 Data Verbalization (i.e., Alt Text)

Our definition of “data verbalization” refers to a way of representing data patterns using verbal description with alternative text (i.e., alt text). Images without alt text markup are not readily accessible to assistive technologies (i.e., screen readers; refreshable braille displays; text-based voice commands) and web search parsers. However, appropriately designed alt text can not only give assistive technology users minimum access to visualized data but also provide all students with a much richer context (Lundgard and Satyanarayan 2022).

There are two ways to create alt text: (1) auto-generated alt text; and (2) manual alt text. Currently, `BrailleR::VI()` (Godfrey et al. 2021) function has a capability to auto-generate alt text for basic R graphics and ggplot objects (see Figure 4). Some of the supported graphic types include: `graphics::hist()`, `graphics::boxplot()`, `ggplot2::geom_bar()`, `ggplot2::geom_histogram()`, and `ggplot2::geom_boxplot()`.

```r
library(ggplot2)
library(BrailleR)
library(palmerpenguins)

# Create graph
g_species <- ggplot(data = penguins, mapping = aes(x = species)) + geom_bar()

g_species
```

```
## This is an untitled chart with no subtitle or caption.
## It has x-axis 'species' with labels Adelie, Chinstrap and Gentoo.
## It has y-axis 'count' with labels 0, 50, 100 and 150.
## The chart is a bar chart with 3 vertical bars.
## Bar 1 is centered horizontally at Adelie, and spans vertically from 0 to 152.
## Bar 2 is centered horizontally at Chinstrap, and spans vertically from 0 to 68.
## Bar 3 is centered horizontally at Gentoo, and spans vertically from 0 to 124.
```

```r
# Obtain auto verbalization
# (Added feature in BrailleR >= 0.32.1): `BrailleR::VI()` function is automatically called against supported graphics if BrailleR package is currently loaded in R session
#BrailleR::VI(g_species)
```

Even though the automated alternate text has some use with its limitations, it does not often directly convey the message a visualization is displaying. Thus, it is important to write alternate texts manually and teach students how to write one. We rely on Cesal’s “simple formula
for writing alt text for data visualization” (2020), which is as follows: “alt= Chart type of type of data where reason for including chart.” She also recommends the inclusion of a link to data somewhere within the text. Similarly, for a meaningful alternate text, Canelón and Hare (2021) suggest that the description should convey the meaning of the data, variables on the axes, scale, and type of plot should be included.

Given these recommendations, we provide the alternate text we wrote for Figure 1. “Sample scatterplot showing the relationship between flipper length in mm on the x-axis and bill length in mm on the y-axis. Flipper lengths vary from about 170 to 230, and bill lengths vary from about 35 to 60. Overall there is a moderate positive relationship. Each data point is colored differently for three species as Adelie, Chinstrap, and Gentoo. Adelie has a short flipper length and short bill length in comparison. Chinstrap has low flipper length but high bill length. Gentoo has a high flipper length and high bill length. Data points are in red, green, and blue for Adelie, Chinstrap, and Gentoo penguins, respectively.” Note that we specify colors as they would be relevant as readers switch from Figure 1 to Figure 3.

Once alt text is prepared either programmatically or manually, it can be embedded in static and dynamic plot objects in R, such as `ggplot2::labs(..., alt = "[alt text]")`, `shiny::renderPlot(..., alt = "[alt text]")` (Chang et al. 2022). We also ask our students to provide alt text within their R Markdown chunks by using `fig.alt = "[alt text]"` knitr chunk option (Xie 2014), and within Pandoc-flavor/Common Markdown with `![alt text here.](path/to/file)` syntax (MacFarlane [2010] 2022).
2.3 Data Sonification Using Sound

Our definition of sonification refers to a data representation method using stereo sound and various audible patterns. For example, Figure 5 can be made audible by using `sonify` package (Siegert and Williams 2017) like below. The output is audio which can also be accessed as an audio file at https://jooyoungseo.github.io/teaching-accessibility-manuscript/audio/sonify-example.wav.

# Representing the relationship between the two numerical variables via stereo sound
library(sonify())

# Prepare variables
x <- 1:5
y <- 1:5

sonify(x, y)

In this example, a scatter plot is represented by sound. Position on the X axis is communicated using the left-to-right stereo effect; the pitch of the sound indicates the position on the Y axis. Data sonification is very useful when representing linear patterns; however, there is much room for improvement in terms of complex data sonification as of this writing. Thus, we suggest using sonification for simple linear regression models.

2.4 Data Tactualization Using Swell Form Machine

Data tactualization refers to making data visualization in a tactile form so that it can be touchable by fingers. Although this requires some expensive embossing hardware (e.g., braille embosser or Swell Form Heating Machine), this may be one of the most effective and accessible data representations for those who are blind or visually impaired. The tactileR (Seo 2022) package in R can generate a ready-to-emboss pdf file from R graphics and ggplot objects. Students and instructors can produce their own tactile graphics by using the following functions: tactileR::brl_begin(); [their graphic object]; tactileR::brl_end() (see Figure 6). For those who do not have access to embossing hardware, the learning process on tactualization can be supported with a video that shows the process of printing a tactile graph available at https://www.youtube.com/watch?v=CIi5554Z1M&ab_channel=JooYoungSeo.
Seo, J. and Dogucu, M.

Figure 6: Ready-to-print tactile graph in braille.

```r
library(palmerpenguins)
library(tactileR)

brl_begin("figure/tactile-boxplot.pdf")
with(penguins, boxplot(body_mass_g ~ species))
brl_end()
```

3 Delivery of Content

We believe that a natural connection between data science and accessibility can be achieved in the data science classroom while presenting data visualization and then introducing the other aforementioned forms of data representation. While delivering the content, instructors should try to provide as many opportunities as possible for students to experience using these tools and accessibility principles themselves. For instance, each student could use screen reader at least once regardless of their visual acuity and hear the alternate text that they have written.

Ample opportunities in learning accessibility should also include assessments. For instance, instructors should modify assessment instructions and rubrics to include accessibility. This will prevent students from just hearing about accessibility in the lecture and then forgetting about it. Ideally, the assessments should not treat accessibility as a learning objective at a single point in the academic term. For instance, if accessibility is covered in Week 3 of the term, students
can (and should) still be expected to write alternate texts for visualizations until the end of the semester.

In this manuscript, we have focused on domain-specific (i.e., data science) knowledge of accessibility using a particular language (R). Depending on the courses taught, instructors may want to extend the accessibility content to non-data contexts. In addition, they may adopt the content in their courses with other languages such as Python and SAS or even in their language-agnostic courses. Overall, if taught well, a student learning accessibility content, for instance, should be able to include alternate text in their PowerPoint presentation in their Anthropology class. In other words, accessibility knowledge should extend beyond data science and beyond any tool.

Although we have mainly focused on tools and different representations of data to have students develop accessible content, a more crucial accessibility feature is reproducible workflows. Scholars often focus on the reproducibility and replication of scientific findings. We believe reproducibility can often serve as an accessibility feature in a data science context. A reproducible open-source workflow, for example, allows users with and without dis/abilities to examine data in their preferred form of data representation. Thus, in addition to teaching different representations of data, instructors may also consider teaching reproducibility.

4 Closing Remarks

We believe that an accessible future in data science requires educators to teach accessibility as part of the curriculum deliberately. This includes both teaching the topic as well as assessing it so that students are held accountable for their accessibility practices. We also acknowledge that what we have presented here is only a small portion of the intersection of visual accessibility and data science.

Thus far, we have just demonstrated some methods to teach accessibility in introductory data science courses with a special focus on accessible visualization. However, as Shinohara et al. (2018) highlight, teaching accessibility at scale requires large-scale adoption and implementation of these methods by faculty, who are organically contributing to, and constructing, instructors’ knowledge (Barnett, Parry, and Coate 2001; Kirk and MacDonald 2001). We hope that this manuscript has just opened the discussion on teaching accessibility in the data science field, and we encourage other data science faculty to help us move towards large-scale curricular change beyond our examples.

Supplementary Material

The reproducible R source code of this manuscript can be found on a GitHub repository at https://github.com/jooyoungseo/teaching-accessibility-manuscript.

Funding

The second co-author was supported through a Teach Access grant to develop curricular materials on teaching accessibility as part of the data science curriculum. Educators interested in teaching accessibility can follow the organization’s work at https://teachaccess.org/.
References


Canelón S, Hare E (2021). Revealing room for improvement in accessibility within a social media data visualization community. https://www.youtube.com/watch?v=DxLkv2iRdf8&ab_channel=csvconf


Okabe M, Ito K (2008). Color Universal Design (CUD) – How to make figures and presentations that are friendly to colorblind people –. https://jfly.uni-koeln.de/color/


