Designing Born-Accessible Courses in Data Science and Visualization: Challenges and Opportunities of a Remote Curriculum Taught by Blind Instructors to Blind Students

JooYoung Seo[1] , Sile O'Modhrain[2] , Yilin Xia[1] , Sanchita Kamath[1] , Bongshin Lee[3] , and James M. Coughlan[4]

[1]University of Illinois at Urbana-Champaign, School of Information Sciences, USA
[2]University of Michigan, School of Information, USA
[3]Yonsei University, Republic of Korea
[4]Smith-Kettlewell Eye Research Institute, USA

Abstract
While recent years have seen a growing interest in accessible visualization tools and techniques for blind people, little attention is paid to the learning opportunities and teaching strategies of data science and visualization tailored for blind individuals. Whereas the former focuses on the accessibility and usability issues of data visualization tools, the latter is concerned with the learnability of concepts and skills for data science and visualization. In this paper, we present novel approaches to teaching data science and visualization to blind students in an online setting. Taught by blind instructors, nine blind learners having a wide range of professional backgrounds participated in a two-week summer course. We describe the course design, teaching strategies, and learning outcomes. We also discuss the challenges and opportunities of teaching data science and visualization to blind students. Our work contributes to the growing body of knowledge on accessible data science and visualization education, and provides insights into the design of online courses for blind students.

CCS Concepts
• Applied computing → Education;

1. Introduction

Over the past decade, there has been a growing body of research on making data visualization accessible to a broader range of people [KJRK21, LCI*20, MLB*21, SD23, LMSW24]. The data visualization and accessibility communities, in particular, have actively investigated non-visual data representation and interaction methods for blind and low-vision (BLV) individuals who face extra challenges in directly accessing data visualizations. Techniques that have been explored include data sonification using spatial sound [SLS*19, HEEB23], natural language data descriptions [LS22, CJP*19], tactile graphics with embossed braille or haptic feedback [BH12, DGMB21], and multimodal data representations that combine various methods to supplant visual elements [TMS*23, SWM*22, SXYM23].

However, the focus has been predominantly on addressing accessibility and usability issues in data visualization systems and techniques. In addition, as Fitzpatrick and colleagues have highlighted, there exists a tendency to inadvertently categorize blind individuals as mere consumers rather than creators of data visualizations [FGS17].

In contrast, this paper explores the learnability of data science

and visualization by, with, and for blind people. We introduce a novel approach to teaching data science and visualization to blind students through an online course, facilitated by blind instructors. We present a two-week summer course, which took place in August 2023, involving nine blind learners with diverse professional backgrounds from banking, to managers of databases for state and federal administration, and to neuroscience research.

This paper addresses the following research questions:

1. How can we scaffold the learning of data science and visualization for blind students in an online setting?
2. How much can blind students learn about data science and visualization in a two-week online course?
3. What are the challenges and opportunities of teaching data science and visualization to blind students, especially by blind instructors?

We detail the course design, teaching strategies, and learning outcomes, and discuss the challenges and opportunities encountered in teaching data science and visualization to blind students. Our work contributes to the expanding knowledge base on accessible data science and visualization education, and offers insights into designing online courses for blind students.

2. Related Work

2.1. Teaching Statistics and Visualizations to Blind Learners

Teaching statistics and data visualization to blind students poses significant challenges due to the heavy reliance on visual representations of data and concepts. Previous research has explored various methods and tools to make statistics and data science education more accessible and inclusive for blind learners.

Gibson and Darron [GD99], for example, conducted an exploratory case study with one blind student to identify effective adaptations for making statistical concepts accessible without visual aids. They provided the student with tactile graphics and braille handouts created using a Perkins Brailler and swell paper. The student achieved above average scores on exams and assignments, suggesting that blind students can learn foundational statistics with appropriate accommodations.

Marson et al. [MHW13] reviewed the literature on teaching statistics to blind students and offered practical advice based on their own experiences. They recommended useful classroom aids such as Nemeth Code for math braille, tactile models of distributions, and kinesthetic explanations. They also discussed assessment strategies and potential pitfalls for instructors, such as being overly sympathetic or lowering expectations.

Godfrey and Loots [GL15] provided a unique perspective as blind PhD statisticians and university lecturers reflecting on their own experiences learning and teaching statistics. They emphasized the importance of communication and direct questions to elicit blind students' needs. They also suggested specific accommodations such as accessible texts, recorded lectures, quality embossed images, and accessible software like R or SAS.

Our work contributes to this knowledge base on accessible data science education with the insights into designing online courses for blind students.

2.2. Accessible Tools for Data Science Education

The most comprehensive work on making statistical software accessible to blind students and professionals has been carried out by Jonathan Godfrey, a senior faculty member at Massey University in New Zealand who is himself blind. Godfrey outlined the accessibility issues faced by blind users when accessing statistical software and explained how R has proven to be the most accessible option [God13]. He introduced a new R package called BrailleR [God18] aimed at converting graphical output to text descriptions for blind users. He also proposed some minor improvements in R to further enhance its accessibility, such as alternate formats for documentation and functions that provide text representations of graphs. Godfrey has also collaborated with other BLV scholars to extend this work. For example, Fitzpatrick, Godfrey, and Sorge presented a method for producing accessible statistical diagrams in R for blind users [FGS17]. They used the gridSVG [MP14] package in R to generate detailed SVG graphics files encoded with semantic information [FGS17]. They then used the DIAGcess JavaScript library [GMS18a, SLW15] to enable interactive web-based exploration of the diagrams through screen readers, synchronized highlighting, and magnification. They demonstrated the method for various types of graphs, such as bar charts, histograms, box plots, and time series graphs.

In addition, Godfrey and his colleagues extended the previous work by proposing a hierarchical navigation model to support exploring the accessible diagrams [GMS18b]. They provided summary descriptions of the full chart first, then allowed users to drill down into specific chart components. They also discussed the limitations of screen readers and the need for additional modalities such as braille and sonification.

More recently, Seo et al. have developed a multimodal access and interactive data representation (MAIDR) system to enhance the accessibility and inclusivity of data science and visualization education for BLV learners [SXYM23, SXL*24]. The MAIDR system provides customizable multimodality modes (i.e., braille, text, and sonification) to learners with varying degrees of visual disabilities. It works with four types of graphs (bar plot, heatmap, box plot, and scatter plot).

Taken together, the synthesis of prior research underscores that the application of embossed tactile graphs for scaffolding visual schema concepts, combined with the provision of accessible statistical software and packages (e.g., R, BrailleR). This is accomplished by leveraging a multimodal data representation system such as MAIDR [SXYM23, SXL*24], which offers a viable pathway for blind individuals to achieve competencies in statistics, data science, and visualization comparable to those of their sighted counterparts.

3. Learning Design

The course presented in this paper was specifically designed for blind individuals to explore data science and visualization using R. It was developed as part of an initiative funded through the Rehabilitation and Engineering Research Center (RERC) on Blindness and Low Vision at the Smith-Kettlewell Eye Research Institute. It included four interactive sessions, each three-hour long, delivered via Zoom and led by a blind professor who's an expert in using R for data science work. We allocated three hours per session to allow for a balance between lecture, hands-on practice, and Q&A time. Since both instructors and students used screen readers, we had to be mindful of the time needed to navigate through the code and the screen reader's speech output. While a course of this duration cannot cover everything in the vast field of data science, our main focus was on data visualization and exploration in non-visual ways. The concepts taught, while centered around R, can be easily applied to other programming languages such as Python.

Our target learners were those who require data visualization skills in their academic or professional journey. This included faculty, undergrad/graduate students, programmers, analysts, engineers, and scientists. Due to the personalized nature of this course, we could only accommodate up to 10 participants. Selection was based on the application and immediate needs of the participants. Participation in pre- and post-course interviews and surveys, as well as completing a short assignment after each session, was requested for research purposes. The study was conducted in accordance with the ethical guidelines and regulations set forth by
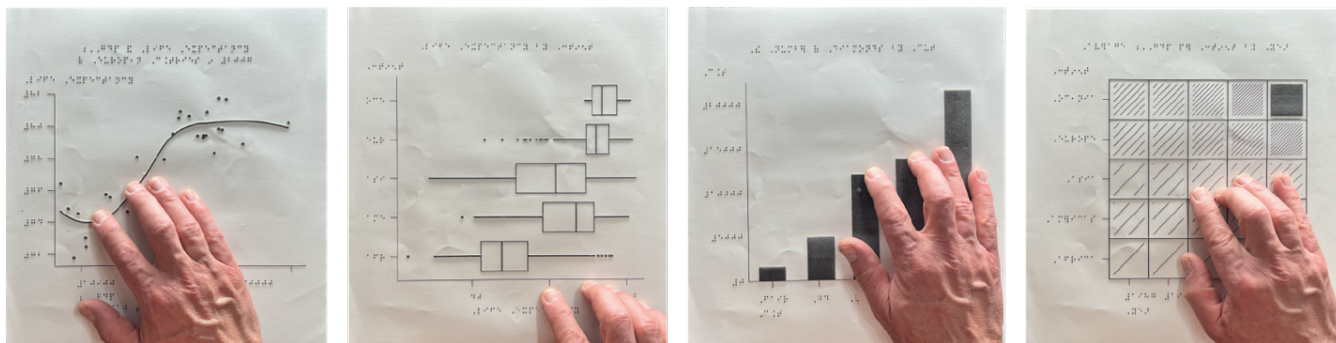
Figure 1: *Tactile graphics in swell paper shared with participants: (A) scatter plot, (B) box plot, (C) bar plot, and (D) heatmap.*

the approved Institutional Review Board (IRB) at both the Smith-Kettlewell Eye Research Institute and the University of Illinois at Urbana-Champaign.

3.1. Objectives

Our primary objective in designing and teaching this course was to equip blind individuals with the tools and knowledge necessary to independently create and interpret data visualizations, contributing to a more inclusive data literacy movement. By doing so, we aimed to increase participants' effectiveness in their professional careers and thereby to bridge the representation gap in the data science field. Specifically, the course addressed:

- How to set up the data science environment, including installing R and Visual Studio Code (VSCode), and configuring them accessible to blind learners.
- Basic concepts of data science workflows.
- How to create and explore multimodal data visualization in an accessible, non-visual fashion using ggplot2 [Wic16] and MAIDR packages.
- How to effectively share and communicate with sighted people through their own data and visualization.

3.2. Sessions and Materials

The course consisted of four sessions across two weeks (Aug 7 and 9 for the first week and Aug 14 and 16 for the second week, with each session lasting three hours (from 8-11am in US Pacific Time), resulting in a total of 12 hours. Besides, office hours will be made available upon request from participants.

We provided the participants with the course materials, including the technical setup guide, lecture slides, and R scripts, in advance via our GitHub website at https://jooyoungseo.github.io/a11y_ds/. The technical setup was prepared by the primary instructor and the course content slide decks were adapted from the open-source "Data Science in a Box" curriculum [Tid24]. We also provided tactile graphics in swell paper to participants in advance to help them understand the basic concepts of data visualization (Figure 1), which is available on the course website for download.

Session 1: Setting Up The Environment. We used VSCode as our primary programming integrated development environment (IDE) because it has been proven one of the most accessible programming code editors for screen reader users [SR23]. Participants learned how to set up their environment, including installing R, VSCode, and configuring assistive devices for accessibility.

Session 2: Data Science Fundamentals. In this session, participants were introduced to the basic concepts of data science, including data importing, tidying, transforming, visualizing, and modeling. This workflow was inspired by Hadley Wickham's R for Data Science [WG16].

Session 3: Data Visualization and Exploration. Participants explored data visualization and how to interpret data non-visually. In advance of the course, tactile graphics illustrating four common data visualizations (Figure 1) – Bar Plots, Scatter Plots, Box Plots and Heatmaps – were prepared and shipped to participants. Behind the selection of these four Cartesian-based visualizations was the consideration of their widespread use in various fields, their distinct statistical properties (e.g., frequency distribution; correlation and regression; interquartile range and outliers; cross-table and chi-square), and their compatibility with MAIDR system [SXL*24]. During the course, we guided learners to get familiar with the basic concepts of these four visualizations and then participants learned how to create some of these data visualizations using the ggplot2 package. Later, participants were taught to explore the visualized data using dynamic sonification and a refreshable braille display using MAIDR package [SXYM23, SXL*24].

Session 4: Data Visualization Do It Yourself. In the final session, our goal was to have participants brought their own data and to create their own visualizations. Although this turned out to be an overambitious goal, we were able to guide participants through the process of importing an external dataset and creating a visualization of their choice following the tidyverse principles [WAB*19]. Due to the wide range of participants' technological background and required assistance (Section 5.2), we were not able to cover all the participants' personalized data visualization needs within the time frame of the course. However, participants were encouraged to continue practicing and exploring data visualization on their own after the course.

Table 1: *Overview of participant demographics, and refreshable braille displays and screen readers employed in the study. The numerical values in the names of the braille devices typically indicate the number of braille cells they contain. An exception is the HIMS Inc BrailleSense 6, which has 32 cells.*

PID	Braille Device Name	Gender	Age	Education	Major	Screen Reader
P01	Freedom Scientific Focus 40	Female	30	Doctorate	Cognitive Psychology	JAWS
P02	VarioUltra 40	Male	38	Bachelor	Computer Animation and Media Arts	VoiceOver
P03	VarioUltra 20	Male	30	Master	Political Science	JAWS
P04	HIMIS inc BrailleSense 6	Male	28	Master	Computer Science	JAWS, NVDA, VoiceOver
P05	Humanware Brailliant BI 40X	Cisgender Female	37	Doctorate	Special Education	NVDA
P06	Freedom Scientific Focus 40	Male	48	Doctorate	Political Science	JAWS
P07	Humanware Brailliant BI 20X	Cisgender Female	29	Master	Psychology	JAWS
P08	Freedom Scientific Focus 40 Blue	Cisgender Female	23	Bachelor	Flip Performance	JAWS
P09	Freedom Scientific Focus 40	Female	50	Doctorate	Communication	JAWS, NVDA

3.3. Course Team and Students

The course team consists of one primary instructor, one assistant instructor, one teaching assistant (TA), and one course supervisor. Two instructors are blind while the TA and course supervisor are sighted. All team members are located in the USA.

JooYoung Seo was the primary Instructor for this course. Dr. Seo is an assistant professor in the School of Information Sciences, University of Illinois at Urbana-Champaign. He is a certified RStudio data science educator, a blind scientist, and has contributed to various open-source data science projects to improve their accessibility. He is also closely collaborating with industry partners, including Posit (formerly RStudio) and Microsoft VSCode team, to enhance the accessibility of data science tools and technologies.

Sile O'Modhrain was the co-instructor. Dr. O'Modhrain is also a blind scientist. She is an associate professor of music in the School of Music, Theatre & Dance, and associate professor in the School of Information at the University of Michigan.

Teaching Assistant Yilin Xia is a doctoral student from the School of Information Sciences at the University of Illinois at Urbana-Champaign, with a focus on accessible data science. He was responsible for hosting office hours and helping resolve technical problem encountered by the students.

Course Supervisor Dr. Coughlan, director of the Smith-Kettlewell Rehabilitation Engineering Research Center (RERC) on Blindness and Low Vision, provided overall guidance in developing this course.

Participants for this course were recruited primarily from the Blind Academics list, a private email forum for BLV professionals and graduate students who actively engaged in research. Participants were required to:

- Be 18 years or older.
- Be a US citizen or current resident.
- Be an everyday screen reader user (e.g., JAWS, NVDA, VoiceOver, ORCA) using speech output and/or braille display.
- Have access to a laptop or desktop that can run Zoom, VSCode, and R.
- Have access to a refreshable braille display or braille notetaker that can be connected to a computer and screen reader.
- Have a basic understanding of computer programming, such as variables, functions, loops, and package and data loading.
- Have an immediate need for data visualization skills in their academic or professional careers.

The nine participants in the course had an average age of 34.78 years (*SD* = 8.73), with ages spanning from 23 to 50 years old (Table 1). Gender distribution among the group were two females, four males, and three individuals identifying as cisgender females. Their educational backgrounds varied, with qualifications ranging from bachelor's degrees to master's and PhDs, in disciplines including Cognitive Psychology, Political Science, and Computer Science. The participants came from a range of work environments including banking, education administration, and industrial and academic research labs.

In terms of their visual impairments, the majority of participants had been blind since infancy. Two participants experienced visual impairments after turning five years old, and P05 was born with low vision and became entirely blind at 25. In terms of assistive technology, the average starting age for using a screen reader was 16.33 (*SD* = 6.65) with a duration of use ranging from just 10 to over 30 years. Most have been using braille for about five years, although P03 has only recently begun to use braille displays. To optimize their experience with the course, we gathered information on their preferred screen readers and operating systems in advance. Seven participants preferred using Job Access With Speech (JAWS) on Windows, two of whom also utilized Non-Visual Desktop Access (NVDA), while P02 exclusively used VoiceOver on a Mac.

Overall, this course demonstrated the presence of an active community of blind professionals with a declared need to visualize their own data for analysis purposes, but also a need to confidently and independently generate visual representations of their data to communicate with their sighted peers.

3.4. Assessments

To determine how participants' confidence level in data visualization changed after the course, we conducted a pre- and post-self-report survey and interviews (refer to Appendix). For the four visualizations covered in the course (i.e., bar plot, box plot, heatmap, and scatter plot), pre- and post-assessments of confidence in creating and interpreting them were measured using Likert scales [Ber07]. Semi-structured interviews were also conducted along with the pre- and post-surveys for deeper interpretation.

Table 2: *The changes of confidence levels. * indicates a statistically significant difference.*

Chart Type	Tasks	Median Scores	p	Z	IQR
Bar Plot	Creation	increased from 2 to 5*	0.016	0.0	remained at 1
Heatmap	Creation	increased from 1 to 3	0.088	6.0	decreased from 2 to 1
Box Plot	Creation	increased from 1 to 4*	0.016	0.0	increased from 1 to 2
Scatter Plot	Creation	increased from 2 to 4*	0.017	0.0	remained at 1
Bar Plot	Interpretation	remained at 5	0.317	0.0	remained at 0
Heatmap	Interpretation	increased from 2 to 3	0.221	3.0	increased from 1 to 2
Box Plot	Interpretation	increased from 3 to 5*	0.017	0.0	increased from 0 to 1
Scatter Plot	Interpretation	remained at 4	0.414	1.5	remained at 2

4. Results

4.1. Quantitative Analysis

We used the Wilcoxon Signed-Rank Test [Woo07] to compare pre- and post-test scores for each metric, suitable due to the paired nature of our data and the non-normal distribution of score changes.

Significant improvements in participants' confidence levels were observed in several areas of graph creation and interpretation as a result of attending this data science course (see Table 2). These results indicate statistically significant improvements in participants' confidence in creating bar plots, box plots, and scatter plots, and interpreting box plots as a result of attending the data science summer course. While some metrics showed non-significant changes, the overall trend suggests a positive impact of the course on enhancing data science skills among participants. Future work will involve recalculating effect sizes accurately to fully understand the magnitude of these changes.

4.2. Qualitative Analysis

We collected qualitative feedback from participants to understand their experiences and identify areas for improvement. The feedback was analyzed using thematic analysis and the following three common themes and patterns emerged: the challenges of audio-based learning, the difficulties of setting up and using the software environment, and the benefits of the course for enhancing their data literacy and confidence. We present each theme below with illustrative quotes from the participants.

4.2.1. Challenges of Audio-based Learning

Students reported that one of the major challenges of the course was to divide their auditory attention between multiple sources of speech, such as their own screen reader, the instructor's screen reader, and the instructor's spoken presentation. This made it hard for them to follow the code examples and the data visualizations.

You're juggling the verbal information of the code and like understanding it. Plus the spatial information and the. And then you're just making sense of like the visual representation of the graph. What is like what does this line actually mean? Yeah. So it's like two to three things going on at once. – P01

I think the whole set of thing was a mess ... everybody was trying to ask things at the same time and JooYoung was, you know, like doing his best and still like people were really lost and we have like, sometimes like two or like screen readers at the same time and people. Yeah. That was not ideal. – P03

I think when I was following with the recording, so second session, I followed it with recording completely. And sometimes I felt like I wish there was kind of instruction given and then give like, I don't know, like one minute to two minutes time so that people can try it on their own and then kind of catch up to it. So that, so like what I had to do then was like pause the recording to do, do follow the instruction and then see if that works out and then on pause the recording go on. And then I realized that when I joined real time, it wasn't actually super hard to follow real time. And I think in large parts, I was kind of stop put my jaws on speech on demand and kind of follow it with real display while listening to the jobs that was being shared. So I think I was able to follow through that way. – P07

4.2.2. Difficulties of Setting Up and Using the Software Environment

Another challenge that participants faced was the setup and use of the software environment, which consisted of VSCode as the code editor, R as the programming language, and various R packages for data analysis and visualization. Participants had different levels of proficiency and familiarity with their access technologies (AT), such as screen readers and braille displays, and each AT had its own configuration issues with VSCode and R. Some participants expressed frustration and confusion with the VSCode interface and the R syntax, while others appreciated learning new skills and tools.

The most challenging was definitely VSCode. Having to deal with the interface. I don't know. Just think that there were a bunch of gaps between like getting it set up and understanding in the way that it worked. It didn't click until maybe the third or fourth class that the way our works, the way VSCode works, and the way that we were being taught is it's not like a functional language in the term that it starts at the top of the bottom. It's like a scratch pad where you just sort of write where something were any executable code, wherever you wanted in the source file, and then just executed it from there. It was odd to me. And then ultimately, like, it was just, yeah, I just did not like the VScode experience at all. – P02

The most challenging aspect of the course was to get the environment set up. That was the most challenging for me. That's my environment was set up. It was easy to follow the course content as well as the instruction. – P08

I wish I could have done everything ahead though and even though you might not have explained it, maybe going ahead and doing all of the packages, doing everything that we would need so that the instruction could flow. Maybe even if that had been because there was Mac, there were NVDA users and jaws. – P09

4.2.3. Data Literacy and Confidence

Despite the challenges, participants also reported that the course was a valuable experience for them, as it enhanced their data literacy and confidence. Participants were able to create and interact with various data visualizations, such as bar plots, scatter plots, and box plots, using non-visual techniques, such as sonification and braille representations. Participants also gained a better understanding of the underlying data and the concepts of data analysis and manipulation. Participants expressed a desire for more advanced-level courses on data science and visualization in the future.

The most rewarding part was the fact that I could not only run and create these various artifacts of data representation myself, but also understand the underlying data. To begin with, there are two parts to what I just said. One is to interact with the software and create these graphs and charts. Then the other part, equally important or maybe more important in some ways, was to understand the underlying data. That had not happened until this. I had interacted with maps and graphs and charts here and there, but then I could never get to the underlying data that was feeding into those charts and graphs. – P08

I think there is a difference between just manually using the keyboard to produce it versus being asking somebody else to do it. So I think that was really like a meaningful experience. – P07

academically, getting first-hand experience with some of the R packages that I wouldn't otherwise be exposed to. such as BrailleR, R, MAIDR and Tidyverse. . . . Great. Thank you. That's a really rewarding. – P04

The combination of sonification is very cool. Oh, we really appreciate it. . . . it's really cool to have a tool like that as a blind person where visually everyone has that tool everyone sighted looks at their outputs . . . That's the whole point of visual data analysis. And this does make that visualization come into a position of being almost as efficient for me as it is for sighted people. – P05

5. Discussion

Teaching data science to blind students in a remote setting, especially delivered by blind instructors, presents unique challenges. This exploration has not only highlighted these challenges but also prompted thoughtful considerations on how to better design accessible learning experiences in the future.

5.1. Challenges of Real-Time Auditory Processing

As reported in Section 4.2.1, a primary challenge was the difficulty of processing multiple, overlapping sources of auditory information

in real time. In a typical coding class, an instructor will likely guide students through code by explaining how particular lines of code work while either typing or highlighting the corresponding code expression. A blind student sitting in a classroom and trying to follow or implement example code in real time has to listen not only to their instructor, but also to the spoken output of their screen reader (often via a single earbud) if they do not have access to a refreshable braille device. They have two competing sources of complex spoken information to attend to. In an online environment where the instructor, too, is using a screen reader to demonstrate code in an accessible way, the problem becomes even more intractable.

In the classroom environment, a blind student can at least rely on the fact that the two auditory sources are spatially separated to selectively attend to one or the other as the need arises, i.e., to rely on the perceptual phenomenon known as 'auditory stream segregation,' whereby distinguishing attributes of auditory sources such as timbre, pitch or spatial location can function to afford each source a unique identity within the auditory environment, an identity that a listener can use to selectively attend to or ignore individual sound sources as when choosing who to listen to at a cocktail party [McD09]. But if the instructor, their screen reader, and the student's own screen reader are all arriving through headphones and are not spatially separated or sufficiently differentiated in timbre, the cognitive load can become extremely high.

A more accessible solution would be to use stereo streaming over headphones, or better to take advantage of the higher spatial fidelity of binaural audio [HGH16]. This would allow the instructor's voice, their screen reader and the student's screen reader to be rendered in different locations within a virtual soundscape. Further, as Zekveld and colleagues have noted, source separation by means of timbre is even more effective at reducing cognitive load for a listener than separation by spatial location [ZRK*14]. For the students in our online class, this could be accomplished by employing different synthesizer voices for the screen reader used by the instructor versus those used by the students. Recent advancements in screen readers, such as JAWS and NVDA, facilitate this through their ability to split system sounds and screen reader sounds into separate audio channels, thereby improving the clarity and spatial distinction of auditory information. In our questionnaire, we did not ask participants whether they had chosen to change the voice of their synthesizer to separate it timbrally from that used by the instructor's screen reader. Even if students solved the source segregation problem, there remains a temporal processing challenge as code must be listened to sequentially, without visual scanning abilities. Sufficient gaps must be provided for cognitive processing. As it is impossible to address the individual pacing needs of every student in a real-time classroom environment, we believe that a better approach would be to implement a flipped classroom [RDM16] to support self-paced learning.

5.2. Variability in Assistive Technology Skills

Another major barrier was variability in assistive technology skills as noted in Section 4.2.2. Although learners were professionals, their experience with screen readers, coding tools, and advanced keyboard shortcuts varied enormously. Pre-teaching exercises or modules focused specifically on onboarding this technical knowl-

edge could more evenly prepare students for advanced data science content. The sheer multitude of interacting mental models is intensely demanding for blind users. Ultimately, successfully teaching data science to blind students in an online environment demands a curriculum intentionally designed for nonvisual learning, as well as attention to the particular needs of a speech-based learning environment [WB15]. Simply making visual content accessible reactively is insufficient. A proactive, "born accessible" approach considers blind learning needs throughout. Findings suggest that quality learning experiences for blind students require additional scaffolds like flipped classrooms, differentiated streams of information, and extensive onboarding to the complex technical environment [RDM16].

5.3. Limitations and Future Directions

Limitations of this initial course are primarily due to its small sample size and brief duration. Furthermore, the course was not structured as a controlled experiment, which means our findings rely on self-reported experiences. To build on this groundwork, future studies should aim for a larger participant base and employ a more rigorous, controlled experimental design to improve the current data science courses. The pilot nature of the course restricted our ability to proactively tackle the two principal challenges highlighted in this study: real-time auditory processing and the wide range in proficiency with assistive technologies. Consequently, future research efforts can be dedicated to the development and empirical evaluation of flipped classroom strategies specifically designed to overcome these barriers. Additionally, future research may warrant a more in-depth exploration of how the data science and visualization education for blind students have unique characteristics compared to the general challenges of non-visual approaches in computer science and programming education identified by the prior literature [SLAM19, BBL19].

6. Conclusion

Overall, the 2-week long online course presented in this paper provides an important first step in exploring the barriers and solutions to making data science and visualization education inclusive for blind students. Participants in the course appreciated the opportunity to learn about data science and visualization and applied their new knowledge to create accessible visualizations (Section 4.2.3). They also provided valuable feedback on the course content and structure, which will be used to improve future iterations of the course. Despite the challenges in remote setting and the short duration, our quantitative and qualitative findings suggest that the course was successful in increasing participants' confidence in their data literacy skills, and in their ability to create accessible visualizations. With a conscious design, data science and visualization can produce new career opportunities rather than obstacles for the blind community. Accessible pedagogy paired with accessible tools can enable their full participation.

7. Acknowledgement

We wish to thank our nine participants who gave so generously of their knowledge and professional experience as data scientists.

References

[BBL19] BAKER C. M., BENNETT C. L., LADNER R. E.: Educational Experiences of Blind Programmers. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis MN USA, Feb. 2019), ACM, pp. 759–765. doi:10.1145/3287324.3287410. 7

[Ber07] BERTRAM D.: Likert scales. *Retrieved November 2*, 10 (2007), 1–10. 4

[BH12] BROWN C., HURST A.: VizTouch: Automatically generated tactile visualizations of coordinate spaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (Kingston Ontario Canada, Feb. 2012), ACM, pp. 131–138. doi:10.1145/2148131.2148160. 1

[CJP*19] CHOI J., JUNG S., PARK D. G., CHOO J., ELMQVIST N.: Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum 38*, 3 (June 2019), 249–260. doi:10.1111/cgf.13686. 1

[DGMB21] DE GREEF L., MORITZ D., BENNETT C.: Interdependent Variables: Remotely Designing Tactile Graphics for an Accessible Workflow. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event USA, Oct. 2021), ACM, pp. 1–6. doi:10.1145/3441852.3476468. 1

[FGS17] FITZPATRICK D., GODFREY A. J. R., SORGE V.: Producing Accessible Statistics Diagrams in R. In *Proceedings of the 14th International Web for All Conference* (New York, NY, USA, Apr. 2017), W4A '17, Association for Computing Machinery, pp. 1–4. doi:10.1145/3058555.3058564. 1, 2

[GD99] GIBSON W., DARRON C.: Teaching Statistics to a Student who is Blind. *Teaching of Psychology 26*, 2 (Jan. 1999), 130–131. doi:10.1207/s15328023top2602_13. 2

[GL15] GODFREY A. J. R., LOOTS M. T.: Advice From Blind Teachers on How to Teach Statistics to Blind Students. *Journal of Statistics Education 23*, 3 (Nov. 2015), null. doi:10.1080/10691898.2015.11889746. 2

[GMS18a] GODFREY A. J. R., MURRELL P., SORGE V.: An accessible interaction model for data visualisation in statistics. In *Computers Helping People with Special Needs: 16th International Conference, ICCHP 2018, Linz, Austria, July 11-13, 2018, Proceedings, Part I 16* (2018), Springer, pp. 590–597. 2

[GMS18b] GODFREY A. J. R., MURRELL P., SORGE V.: An Accessible Interaction Model for Data Visualisation in Statistics. In *Computers Helping People with Special Needs* (Cham, 2018), Miesenberger K., Kouroupetroglou G., (Eds.), Lecture Notes in Computer Science, Springer International Publishing, pp. 590–597. doi:10.1007/978-3-319-94277-3_92. 2

[God13] GODFREY A. J. R.: Statistical Software from a Blind Person's Perspective. *The R Journal 5*, 1 (2013), 73. doi:10.32614/RJ-2013-007. 2

[God18] GODFREY A. J. R.: *BrailleR: Improved access for blind users*. Massey University, 2018. R package version 0.29.1. URL: https://cran.r-project.org/package=BrailleR. 2

[HEEB23] HOQUE M. N., EHTESHAM-UL-HAQUE M., ELMQVIST N., BILLAH S. M.: Accessible Data Representation with Natural Sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2023), CHI '23, Association for Computing Machinery, pp. 1–19. doi:10.1145/3544548.3581087. 1

[HGH16] HEEREN J., GRIMM G., HOHMANN V.: The Influence of Dynamic Binaural Cues on Speech Intelligibility at Low and High Frequencies. 6

[KJRK21] KIM N. W., JOYNER S. C., RIEGELHUTH A., KIM Y.: Accessible Visualization: Design Space, Opportunities, and Challenges. *Computer Graphics Forum 40*, 3 (June 2021), 173–188. doi:10.1111/cgf.14298. 1

[LCI*20] LEE B., CHOE E. K., ISENBERG P., MARRIOTT K., STASKO J.: Reaching Broader Audiences With Data Visualization. *IEEE Computer Graphics and Applications 40*, 2 (Mar. 2020), 82–90. doi:10.1109/MCG.2020.2968244. 1

[LMSW24] LEE B., MARRIOTT K., SZAFIR D., WEBER G.: Inclusive Data Visualization (Dagstuhl Seminar 23252). *Dagstuhl Reports 13*, 6 (2024), 81–105. URL: https://drops.dagstuhl.de/entities/document/10.4230/DagRep.13.6.81, doi:10.4230/DagRep.13.6.81. 1

[LS22] LUNDGARD A., SATYANARAYAN A.: Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (Jan. 2022), 1073–1083. doi:10.1109/TVCG.2021.3114770. 1

[McD09] MCDERMOTT J. H.: The cocktail party problem. *Current biology: CB 19*, 22 (Dec. 2009), R1024–1027. doi:10.1016/j.cub.2009.09.005. 6

[MHW13] MARSON S. M., HARRINGTON C. F., WALLS A.: Teaching introductory statistics to blind students. *Teaching Statistics 35*, 1 (2013), 21–25. doi:10.1111/j.1467-9639.2012.00510.x. 2

[MLB*21] MARRIOTT K., LEE B., BUTLER M., CUTRELL E., ELLIS K., GONCU C., HEARST M., MCCOY K., SZAFIR D. A.: Inclusive data visualization for people with disabilities: A call to action. *Interactions 28*, 3 (Apr. 2021), 47–51. doi:10.1145/3457875. 1

[MP14] MURRELL P., POTTER S.: The gridsvg package. *R J. 6*, 1 (2014), 133. 2

[RDM16] REYNA J., DAVILA Y. C., MEIER P.: Enhancing the Flipped Classroom Experience with the Aid of Inclusive Design. In *EdMedia + Innovate Learning* (June 2016), Association for the Advancement of Computing in Education (AACE), pp. 1795–1807. 6, 7

[SD23] SEO J., DOGUCU M.: Teaching Visual Accessibility in Introductory Data Science Classes with Multi-Modal Data Representations. *Journal of Data Science* (Mar. 2023), 1–14. doi:10.6339/23-JDS1095. 1

[SLAM19] STEFIK A., LADNER R. E., ALLEE W., MEALIN S.: BEST PAPER AT SIGCSE 2019 IN THE EXPERIENCE REPORTS AND TOOLS TRACK: Computer science principles for teachers of blind and visually impaired students. *ACM Inroads 10*, 2 (Apr. 2019), 50–57. doi:10.1145/3324894. 7

[SLS*19] SUMMERS I. C. E., LANGSTON J. E., SOOKNE J. D., OLLEY J. B., TROUT K. L. S., IV C. D. H., KALAT S. E., LAYNE P. W.: Converting graphical data-visualizations into sonified output, Jan. 2019. 1

[SLW15] SORGE V., LEE M., WILKINSON S.: End-to-end solution for accessible chemical diagrams. In *Proceedings of the 12th International Web for All Conference* (2015), pp. 1–10. 2

[SR23] SEO J., ROGGE M.: Coding Non-Visually in Visual Studio Code: Collaboration Towards Accessible Development Environment for Blind Programmers. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA, Oct. 2023), ASSETS '23, Association for Computing Machinery, pp. 1–9. doi:10.1145/3597638.3614550. 3

[SWM*22] SHARIF A., WANG O. H., MUONGCHAN A. T., REINECKE K., WOBBROCK J. O.: VoxLens: Making Online Data Visualizations Accessible with an Interactive JavaScript Plug-In. In *CHI Conference on Human Factors in Computing Systems* (New Orleans LA USA, Apr. 2022), ACM, pp. 1–19. doi:10.1145/3491102.3517431. 1

[SXL*24] SEO J., XIA Y., LEE B., MCCURRY S., YAM Y. J.: MAIDR: Making statistical visualizations accessible with multimodal data representation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2024). doi:10.1145/3613904.3642730. 2, 3

[SXYM23] SEO J., XIA Y., YAM Y. J., MCCURRY S.: MAIDR: Multimodal Access and Interactive Data Representation System for Inclusive Data Science Education. In *Proceedings of the 3rd Annual Meeting of the International Society of the Learning Sciences* (2023), pp. 51–54. 1, 2, 3

[Tid24] Tidyverse/datascience-box. tidyverse, Apr. 2024. 3

[TMS*23] THOMPSON J. R., MARTINEZ J. J., SARIKAYA A., CUTRELL E., LEE B.: Chart Reader: Accessible Visualization Experiences Designed with Screen Reader Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2023), CHI '23, Association for Computing Machinery, pp. 1–18. doi:10.1145/3544548.3581186. 1

[WAB*19] WICKHAM H., AVERICK M., BRYAN J., CHANG W., MCGOWAN L. D., FRANÇOIS R., GROLEMUND G., HAYES A., HENRY L., HESTER J., KUHN M., PEDERSEN T. L., MILLER E., BACHE S. M., MÜLLER K., OOMS J., ROBINSON D., SEIDEL D. P., SPINU V., TAKAHASHI K., VAUGHAN D., WILKE C., WOO K., YUTANI H.: Welcome to the Tidyverse. *Journal of Open Source Software 4*, 43 (Nov. 2019), 1686. doi:10.21105/joss.01686. 3

[WB15] WESTERMANN A., BUCHHOLZ J.: The effect of spatial separation in distance on the intelligibility of speech in rooms. *Journal of the Acoustical Society of America 137*, 2 (2015), 757–767. doi:10.1121/1.4906581. 7

[WG16] WICKHAM H., GROLEMUND G.: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. "O'Reilly Media, Inc.", Dec. 2016. 3

[Wic16] WICKHAM H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. URL: https://ggplot2.tidyverse.org. 3

[Woo07] WOOLSON R. F.: Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007), 1–3. 5

[ZRK*14] ZEKVELD A. A., RUDNER M., KRAMER S. E., LYZENGA J., RÖNNBERG J.: Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience 8* (Apr. 2014), 88. doi:10.3389/fnins.2014.00088. 6