# Understanding Screen-Reader Users' Experiences with Online Data Visualizations

Ather Sharif
asharif@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering | DUB Group, University of Washington
Seattle, Washington, USA

Sanjana S. Chintalapati
sanjanac@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering, University of Washington
Seattle, Washington, USA

Jacob O. Wobbrock
wobbrock@uw.edu
The Information School | DUB Group, University of Washington
Seattle, Washington, USA

Katharina Reinecke
reinecke@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering | DUB Group, University of Washington
Seattle, Washington, USA

## ABSTRACT

Online data visualizations are widely used to communicate information from simple statistics to complex phenomena, supporting people in gaining important insights from data. However, due to the defining visual nature of data visualizations, extracting information from visualizations can be difficult or impossible for screen-reader users. To assess screen-reader users' challenges with online data visualizations, we conducted two empirical studies: (1) A qualitative study with nine screen-reader users, and (2) a quantitative study with 36 screen-reader and 36 non-screen-reader users. Our results show that due to the inaccessibility of online data visualizations, screen-reader users extract information 61.48% less accurately and spend 210.96% more time interacting with online data visualizations compared to non-screen-reader users. Additionally, our findings show that online data visualizations are commonly indiscoverable to screen readers. In visualizations that are discoverable and comprehensible, screen-reader users suggested tabular and textual representation of data as techniques to improve the accessibility of online visualizations. Taken together, our results provide empirical evidence of the inequalities screen-readers users face in their interaction with online data visualizations.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in visualization**; **Visualization design and evaluation methods**; **Empirical studies in accessibility**.

## KEYWORDS

visualizations, screen readers, data, challenges, techniques


*ASSETS '21, October 18–22, 2021, Virtual Event, USA*

ACM ISBN 978-1-4503-8306-6/21/10…$15.00
https://doi.org/10.1145/3441852.3471202

**ACM Reference Format:**
Ather Sharif, Sanjana S. Chintalapati, Jacob O. Wobbrock, and Katharina Reinecke. 2021. Understanding Screen-Reader Users' Experiences with Online Data Visualizations. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3441852.3471202

## 1 INTRODUCTION

Online data visualizations are being increasingly utilized to communicate essential insights, assisting users to explore, interact with, and extract meaningful information from complex data [28]. These insights help people make critical and informed decisions for themselves and their families concerning health (*e.g.*, COVID-19), finances (*e.g.*, stock trends), and the current events (*e.g.*, polling data), among other important life domains [25].

The wide adoption of online data visualizations for information uptake, learning, and decision-making means that those who are unable to access the data presented in these visualizations may be at a disadvantage. In particular, screen-reader users—over 7.6 million people in the United States [1]—must extract the information contained within data visualizations in alternative ways. However, no prior work has provided empirical studies of whether and to what degree screen-reader users are in fact disadvantaged with respect to online data visualizations.

We define "screen-reader users" as users who utilize a screen reader (*e.g.*, JAWS[1], NVDA[2], or VoiceOver[3]) to read the contents of their computer screen. They might have conditions including complete or partial blindness, low vision, learning disabilities (such as alexia), motion sensitivity, and/or vestibular hypersensitivity.

Prior work has explored automatically generating alternative text from data visualizations summarizing common statistics [32, 38] as well as providing screen-reader users with alternative mediums for interacting with digital visualizations (such as sonification [17, 31, 47], haptic graphs [42, 46], and 3-D printing [13, 26, 39]). However, such alternative mediums require auxiliary resources and are not practical for daily web browsing by screen-reader users. To

[1]https://www.freedomscientific.com/products/software/jaws/
[2]https://www.nvaccess.org/
[3]https://www.apple.com/accessibility/vision/

our knowledge, no work has evaluated the needs and performance of screen-reader users when interacting with online visualizations.

To understand and assess current challenges with online data visualizations for screen-reader users compared to their sighted counterparts, we conducted two empirical studies. First, through semi-structured contextual interviews [23] with nine screen-reader users, we found that 33% of the visualizations in our study, from a sample of 27 visualizations, were undiscoverable to screen readers. For those that the screen readers were able to detect, screen-reader users seemingly endured an excessive workload burden in extracting information from visualizations. Screen-reader users looked for a holistic overview of the information contained in data visualizations before deciding to delve further and looking at individual data points. However, exploring the data both holistically and in a drill-down manner was challenging and time-consuming for them.

Second, we conducted a controlled experiment to quantify the difference in interaction times and the accuracy of extracted information from data visualizations (generated using D3, Google Charts, and ChartJS) between screen-reader users ($N$=36) and non-screen-reader users ($N$=36). It is important to note that our work is an evaluation of the online data visualizations (the technology) and not of the abilities of people who use screen readers to interpret data. Our results show that the inaccessibility of online visualizations causes screen-reader users to extract information 61% less accurately and to spend 211% more time interacting with online data visualizations compared to non-screen-reader users. Google Charts had the best performance in terms of accuracy of extracted information, as it provides an alternate tabular representation of data that is only visible to screen readers.

While our studies showed that screen-reader users have a significant disadvantage when interacting with online data visualizations compared to non-screen-reader users, our findings also suggest ways to reduce this gap. In particular, screen-reader users identified data representation through tables and text, overall trends, and multi-modality as techniques that can improve the accessibility of online visualizations.

The main contributions of this work are as follows:

- We provide a detailed account of the challenges screen-reader users face with data visualizations, showing that: (1) Many visualizations are entirely indetectable to screen readers, (2) those that are detectable, are identified as "blank," "graphic," "frame," or "object," and are therefore, incomprehensible, and (3) those that are at least passably comprehensible, are time-consuming and can cause screen-reader users an excessive workload in extracting information—both holistically and in a drilled-down manner.
- We present the empirical evaluation of the performance of screen-reader users when interacting with online data visualizations. Our findings show that due to the inaccessibility of online visualizations, screen-reader users spend 211% more time interacting with online data visualizations, and are 61% less accurate in extracting information compared to non-screen-reader users.
- We provide guidelines for the design of accessible online data visualizations and suggest the need for a new approach for creating accessible data visualizations.

## 2 RELATED WORK

We review prior work related to the experiences of screen-reader users with technology, including comparisons to non-screen-reader users. Additionally, we review the previous research on tools and systems designed to improve the accessibility of data visualizations for screen-reader users.

### 2.1 Interaction Experiences of Screen-Reader Users with Technology

Several research efforts have explored the interaction of screen-reader users with technology via interviews [3, 4, 7, 21, 27], showing that screen-reader users encounter a number of challenges. For example, Kane *et al.* [27] conducted interviews with eight screen-reader users, identifying usability issues with mobile devices and especially touch screen smartphones, which were new at the time. Billah *et al.* [7] conducted a study with 21 screen-reader users, reporting on the usage of screen-readers in remote access scenarios. They utilized various screen-readers on different types of computers to measure screen-reader users' experiences with using computers at home, in the workplace, and at school. Most recently, Schaadhardt *et al.* [37] studied screen-readers' experiences of 2-D digital artboards, such as those appearing in Microsoft PowerPoint and Adobe Illustrator. Their findings detail the challenges of using screen-readers in 2-D environments, rather than for 1-D text streams. Challenges emerged that are reinforced by our findings, such as high cognitive loads and a need for better feedback. In our work, we conduct contextual interviews with screen-reader users to understand the common information they seek in online visualizations, the pain points in their interactions, and the wide variety of techniques and strategies that they prefer.

Perhaps most similar to our work is prior work by Yu and Brewster [45], who compared a multi-modal data visualization system and traditional tactile diagrams, measuring the accuracy of information extracted and the interaction time. They found that the multi-modal approach improved the extraction of accurate information from graphs. Similarly, Brewster [12] employed the same dependent variables to compare the performance of speech and pitch sound graphs for screen-reader users, finding that non-speech sound and haptics can significantly improve interaction with visualizations. Both of these studies only explored the performance of screen-reader users. In our work, we employ the same dependent variables but explore the performance difference between screen-reader and non-screen-reader users. Additionally, we also analyze the main and interaction effects of different online visualization libraries, data complexity, difficulty level, and age.

### 2.2 Accessibility of Data Visualizations

Several prior research efforts have attempted to improve upon the accessibility of data visualizations by using a variety of techniques, including automatically generating alternative text for visualization elements [32, 38], sonification [17, 31, 47], haptic graphs [42, 46], and 3-D printing [13, 26, 39]. For example, Sharif *et al.* [38] developed a jQuery plugin to create accessible graphs, automatically generating a summary of the data and utilizing it as alternative text for the graph. Flowers *et al.* [17] examined the equivalence of visual and auditory scatterplots, finding sonification of the data to be similarly efficient compared to visual graphs. Yu *et al.* [46] developed a system to generate haptic graphs and evaluated their system via an experiment employing both blind and sighted people, reporting that haptic interfaces are useful to provide the information contained in a graph to blind computer users. Brown *et al.* [13] worked with six individuals with low or limited vision to understand the usefulness of 3-D printed custom tactile visualizations, building software that automatically generates tactile visualizations leveraging 3-D printing technology. However, while all of these approaches are a plausible step forward in making visualizations accessible to screen-reader users, they only target a singular approach, and do not explore the varying preferences of visualization interaction among screen-reader users.

## 2.3 Online Data Visualization Libraries

Visualizations are usually embedded on web pages either as images or using visualization libraries. In our work, we focus on the visualizations embedded on web pages using JavaScript visualization libraries (*e.g.*, ChartJS, D3, and Google Charts.) Visualization libraries differ in their underlying implementation; some utilize SVG elements and some make use of HTML Canvas—causing varying screen reader outputs from one visualization to another. They also differ in their application of accessibility measures. For example, Google Charts, by default, appends a tabular representation of the data for screen-reader users whereas D3 and ChartJS rely on the developers to add the appropriate alternative text and Accessible Rich Internet Applications (ARIA) attributes [2]. ARIA attributes are a set of attributes to make web content more accessible to people with disabilities. Both alternative text and ARIA attributes rely on the developer to be added appropriately.

# 3 STUDY 1: CONTEXTUAL INTERVIEWS ABOUT DATA VISUALIZATIONS

To understand screen-reader users' interaction with online data visualizations, we conducted contextual interviews with nine screen-reader users. We observed the participants interacting with real-world online data visualizations embedded in websites, and subsequently interviewed them about their experiences. Our research questions were: (1) What challenges do screen-reader users face when interacting with online data visualizations?; (2) What information do they commonly seek in online data visualizations?; and (3) What techniques and strategies could improve their interaction experience with online data visualizations?

## 3.1 Participants

Our participants (Table 1) were nine screen-reader users, recruited using word-of-mouth, snowball sampling, and advertisements through social media channels (Facebook and Twitter) and email distribution lists for people with disabilities. Four participants identified as women, and five as men. Their average age was 50.2 (*SD*=18.4). Two participants were blind since birth, and seven lost vision gradually. The highest level of education attained or in pursuit was graduate level for seven participants; for the remaining two participants, it was undergraduate and high school level, respectively. The daily computer usage was greater than 5 hours for five participants, 3-5 hours for two participants, and 1-2 hours for the remaining two participants. The average visualization interaction frequency was over two visualizations per day. All participants were located in the United States.

We ceased recruitment of participants once we reached saturation of insights, as per prior work [44]. Participants were compensated with a $15 Amazon gift card for 45-60 minutes of their time.

## 3.2 Materials

We curated a set of 50 web pages that contained a visualization created using one of three common visualization libraries (Google Charts, ChartJS, or D3) and one of three common chart types (Bar, Scatter, or Line) [36], by searching for terms including "d3 visualizations," "google charts visualizations," and "chartjs visualizations" on Google. We incorporated different chart types to present participants with a diverse set of real-world visualizations. To avoid inaccessible web pages from affecting participants' experiences with the visualizations contained therein, we screened the web pages to ensure that the 50 web pages themselves (but not the visualizations) were accessible to screen-reader software. Out of the 50 web pages, we randomly selected 27, nine for each of the three chart types. We used stratified random sampling to assign each participant three unique web pages, each with a different chart type.

## 3.3 Procedure

We conducted the study online using Zoom video conferencing in August 2020. Interview sessions lasted 45-60 minutes. We first asked participants about their self-identified gender, age, screen-reader software, vision-loss level, and diagnosis (Table 1). They were also asked about their education level, daily computer usage, and their interaction frequency with online data visualizations.

To prepare for the observational part of our contextual interviews, participants were asked to share their screens and make their screen-readers' audio outputs audible to the researchers. All interviews were conducted by two authors, recorded using Zoom's built-in recording feature and via a smartphone, and were later transcribed.

Participants were asked to interact with the web pages containing a visualization as they normally would in their daily life, following a "think-aloud" protocol. Figure 1 shows a subset of three visualizations, one for each visualization library, that participants interacted with during the study session. The order of the chart type was randomized across participants. Participants spent 10-15 minutes interacting with each web page. We took detailed notes during their interactions. Participants were interviewed in a semi-structured manner with open prompts at the end of each of their interactions. Specifically, participants were asked about the information they sought, the difficulties they faced during their interactions, and the improvements that could optimize their performance in extracting information from visualizations. Additionally, participants were asked about whether and how their experiences with online data visualizations from this study differed as compared to those that they interact with in their daily life.

**Table 1: Screen-Reader Participants for Study 1, their gender identification, age, screen-reader, vision-loss level, and diagnosis. Under the *Gender* column, $M$ = *Male*, and $F$ = *Female*.**

| Participant | Gender | Age | Screen-Reader | Vision Loss Level | Diagnosis |
|---|---|---|---|---|---|
| P1 | M | 26 | NVDA | Blind since birth | Optic Nerve Hypoplaxia |
| P2 | M | 55 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| P3 | F | 30 | NVDA | Blind since birth, Partial vision | Rhetonopy Prematurity |
| P4 | F | 67 | Fusion | Lost vision gradually | Juvenile Macular Degeneration |
| P5 | F | 72 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| P6 | M | 51 | JAWS | Lost vision gradually | Demacular Degeneration |
| P7 | F | 75 | JAWS | Lost vision gradually | Rhegonitis Stignitosa |
| P8 | M | 35 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| P9 | M | 41 | JAWS | Lost vision gradually | Angle Glaucoma |

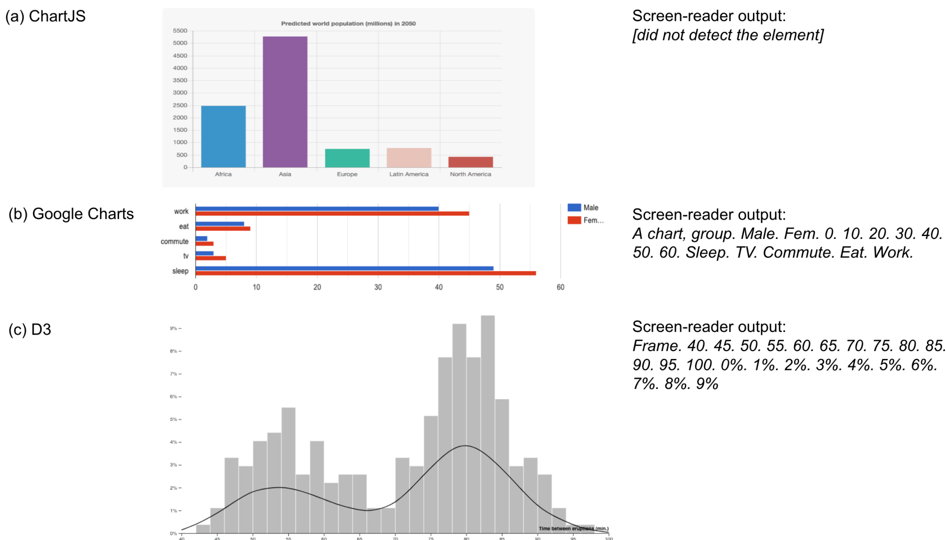


**Figure 1: Examples of three visualizations that participants interacted with during the contextual interviews (Study 1), one for each visualization library. Transcribed text from VoiceOver (MacOS built-in screen reader) for each visualization is shown under each example. (a) was implemented using ChartJS and was not detected by the screen reader; (b) was implemented using Google Charts and was identified as a "chart"; and (c) was implemented using D3 and was identified as a "frame". Axes labels were also read for (b) and (c).**

## 3.4 Data Analysis

We used thematic analysis, following a semantic approach, in which themes are identified within the explicit or surface meanings of the data [33], guided by an essentialist paradigm [35, 43]. The essentialist paradigm focuses on reporting the experiences, meanings, and the reality of the participants [9]. We developed an initial set of codes based on two interviews, before coding all nine interview transcripts, adding new codes as necessary. This resulted in 29 codes. Each interview transcript was coded by two researchers independently, and disagreements were resolved through mutual discussions until consensus was reached. We calculated inter-rater reliability (IRR), expressed as percentage agreement among raters before resolving disagreements, dividing the total number of codes agreed upon by the total number of identified codes across nine transcripts [22]. IRR was therefore calculated as $88 \div 106 \times 100 = 83\%$, demonstrating an acceptable level of agreement [20, 22].

We combined our 29 codes into 10 axial codes using affinity diagramming. Axial codes are codes generated after combining the initial codes into broader, over-arching categories. We followed the thematic analysis approach by Braun and Clarke [9] for analysis. Our final analysis, after searching and reviewing themes (Phases 3 and 4 in [9]), revealed a total of nine themes across our three research questions. Additionally, our analysis was conducted independent of the chart types, as chart types were only used to diversify our visualization dataset.

## 4 STUDY 1: RESULTS

We identified nine main themes across the three main research questions.

## 4.1 RQ1: Challenges and Pain Points

Our analysis revealed three themes relevant to challenges that screen-reader users experience with online data visualizations: (1) Visualizations are often completely indiscoverable by the screen-reader software; (2) information read out by the screen-reader software often lacks context, making the information difficult to comprehend; and (3) data contained in the visualizations is not available to the screen-reader users, restricting screen-reader users to explore the data.

*4.1.1 The First Problem: Invisible Visualizations.* The first theme that emerged in our study was that in 33% of the web pages presented to the participants, their screen-reader software did not detect the visualization at all. As a consequence, participants interacted with the web page, but were unaware that it contained a visualization, as P9 mentioned:

> `It didn't seem like there's anything.` (P9)

Participants expressed similar experiences when interacting with visualizations in their daily lives. P3, who has only had partial vision since birth, said:

> `So the actual visualizations and the graphs, I can't access at all. Unless they have an image description with them, which they usually don't.` (P3)

This finding shows that screen-reader users commonly interact with websites without knowing that a visualization may be present.

*4.1.2 The Second Problem: Incomprehensible Visualizations.* Our second theme showed that even when the visualization was accurately detected, the information the screen-reader software read out to the participants was insufficient to fully comprehend the information that the visualization conveyed. The screen-reader software often only identified the visualization as "blank," "graphic," "frame," or "object". Noticeably frustrated, P6 commented:

> `It says ''graphic, graphic, graphic...'' It means nothing to a blind person. Have to be more descriptive, saying ''graphic'' means nothing.` (P6)

Similarly, P2 was confused when their screen-reader software identified the visualization as an "object"—a term that usually identifies the HTML object tag:

> `I guess a pain point would be when [screen reader] said ''object.'' I wouldn't normally expect to hear that in any kind of visualization, so that was a little confusing for me.` (P2)

Screen-reader users were also often unable to infer what the data was referring to. P3, who has been blind since birth, reported being confused whether the screen reader was reading out axes labels or other data:

> `When it was saying like ''12k,'' ''1k,'' I had no idea what that was referring to.` (P3)

When screen readers read out the data from a data table (as provided in Google Charts), screen-reader users must go through each data point, which can be tedious and cognitively challenging, and is further exacerbated with larger datasets. P9, who interacts with several visualizations on a weekly basis, found complex visualizations in our study very time-consuming:

> `It seemed like I went through a lot of [data] points.` (P9)

*4.1.3 Where's the Data: Lack of Access to Data Points.* Our third theme showed that screen-reader users were frustrated due to the lack of direct access to the data points enclosed in the visualizations. For example, P3 and P8 considered access to the data points as a key piece to explore visualizations, and shared their frustrations:

> `The information that was actually graphed, so again, like the points---I wasn't able to access that. Right now I have zero access to the data point, which is the whole point of having a graph.` (P3)

> `I mean, the data's not accessible, so that's why I got stuck.` (P8)

Overall, the answer to our first research question is that visualizations are oftentimes entirely invisible to screen readers, taking away the opportunity from screen-reader users to explore the data and extract meaningful information. In cases where the visualization element is detectable, it is recognized meaninglessly as "blank," "graphic," "frame," or "object." Additionally, screen-reader users trying to access with visualizations produced by ChartJS or D3 do not have access to the data contained in the visualizations. Google Chart provides access to the data, but our participants found going through each data point tedious and time-consuming.

## 4.2 RQ2: Commonly Sought Information

Two main themes emerged from our analysis that are relevant to answering our second research question on which information screen-reader users commonly seek in visualizations: (1) Screen-reader users first explore the data holistically; and (2) After an initial holistic exploration, they look for and compare individual data points.

*4.2.1 The Flyover: Holistic Exploration and Trend Assessment.* Our first theme for RQ2 showed that screen-reader users first sought to obtain a holistic sense of the data in the online visualizations, getting the "feel" for the information contained in the visualization, before deciding to explore the data further, similar to non-screen-reader users. For example, P6 described how screen-reader users approach online visualizations:

> `When we first interact with something new to us, first we want to try to read everything at a quick glance and then the second time, the third time, the fourth time we really want to understand.` (P6)

To get holistic overview of the data, our participants specifically looked for the *overall trend*, *extrema* (data points representing minimum and maximum values), and *axes information* (labels and ranges for each axis) in the visualizations presented to them during the study session.

When looking for the overall trend, we found that our participants developed a mental image of the data. They did so by interacting with one data point at a time, by navigating through either the SVG elements or the items in the data table, depending on the visualization library. For example, P2 and P4, both of whom lost their vision gradually, described their thought process:

> `In my mind, I try to move along the graph, and visualize [the trend] as it goes up and down.` (P2)

> `I can just listen to data and get an idea what that represents over time.` (P4)

Additionally, P5, in their interaction was interested in the maximum and minimum values (extrema):

> `I would gather the bulk of the info, such as the minimum and maximum.` (P5)

Similarly, P1 was able to get the holistic overview of the data from the axes labels and ranges:

> `X-axis is the time and Y-axis is basically giving the popularity score where 100 is the most popular.` (P1)

*4.2.2 The Drill-Down: Investigating Specific Data Points.* Our second theme showed that participants were interested in the *individual data points*. It is worth noting that the participants were only able to explore the data in a detailed manner after gaining a holistic view of the information contained in the data visualization. Once participants were able to obtain that view, then they explored the values of particular data points. Additionally, participants explored how particular data points compared to other data points (greater, lesser, or the same). For example, P9 and P2 shared their experiences of comparing data points:

> `[I look for] the opening price and the closing price of one minute worth of data all the way up to days, or maybe for months, for data.` (P9)

> `What I would think is that the male is higher, and the female is a lower number right there.` (P2)

Overall, we found that screen-readers first explore the data holistically, seeking the overall trend, extrema, and axes information. For visualizations that pique the interests of screen-reader users after an initial holistic exploration, screen-reader users explore the individual data points both exclusively and in comparison to the other data points.

## 4.3 RQ3: Techniques and Strategies for Accessible Visualizations

To understand future possibilities of information visualizations for screen-reader users, we explored strategies that would improve the accessibility of the visualizations. We identified four strategies as themes for RQ3: (1) Tabular representation; (2) Textual representation; (3) Overall trends in non-visual formats; and (4) Multi-modality.

*4.3.1 Tabular Representation.* Our first theme that emerged for RQ3 was that our participants highlighted the importance of having a tabular representation of the data in lieu of a visualization. For example, P3 and P4 described their positive experiences from everyday life with tabular representations of data:

> `One of my friends developed a website for blind and visually impaired people with COVID-19 data. It's actually based in a [table] instead of a bar graph, so he created this tool that pulls data every day from sources and puts it in a chart form.` (P3)

> `The daily newspapers that I listen to, they list all the municipalities in that county and it's like a three column table, and so it's the name of the municipality, the number of cases to date, and then the number of deaths to date, and some of them have what increase or decrease that represents since March 1st, so that's wonderful for me. I don't need to look at a chart that way I can just listen to data and get an idea of which municipalities have the greatest number of cases, which of them have had the most fatalities, and what that represents over time.` (P4)

*4.3.2 Textual Representation.* Our second theme showed that screen-reader users value the importance and benefits of using textual representations of data to aid in the visualization interaction experience, within and outside of the study session. For example, P1 highlighted the benefits of including alternative text in visualizations:

> `Most visualizations are, of course, inaccessible. If somebody cares to put alternative text, yes, that helps a bit.` (P1)

In another instance during the study session, P1 found the alternative text to be helpful in extracting the holistic overview of the data:

> `This help blob really helped distill the information down more clearly. The alternative text gave me enough information about what the visualization was about.` (P1)

P6 considered alternative text as the best solution:

> `Like I said, alternative text is the best solution to solve this [inaccessible visualizations] problem. So we need to have alternative text to describe the graphics.` (P6)

*4.3.3 Overall Trend In Non-Visual Formats.* Our third theme showed that screen-reader users emphasized the importance of presenting the trend in an alternate format compared to the visual format, reducing the burden of deducing the overall trend on screen-reader users, and increasing the holistic exploration of the data. As P1, who is blind since birth, said:

> It would be helpful to know how the visualization would look in a visual aspect. For example, when people talked about COVID-19, they said it's an exponential growth and flatten the curve---but what does that mean? It means one thing numerically, and it means another thing visually. And for you to participate in a conversation, you should know what the visual aspect means for you to sensibly contribute.

Our participants also identified *sonification*, *summarization*, and *braille printouts* as three techniques to present online visualizations involving data trends to screen-reader users. For example, P3 and P8 shared their thoughts:

> *Sonification*: Having an auditory graph is also super helpful; I just learned how to do that in the health app on my phone yesterday, and it was really helpful to get that audio feedback about what this graph looked like. (P3)

> *Summarization*: Showing the highest point, lowest point, maybe mean or average, and a trend, is it generally going up, down, up and down, is it stagnant? Just a brief description that you would be able to see by just looking at it for one or two seconds. Oh, just a quick paragraph before the data just explaining it, you know, this is the dataset from whatever organization ''X'' that is showing, you know, ''X'' trend over whatever amount of time or something, you know what I mean? Just something like preferably in plain language as well. I'm a big fan of just kind of explaining, you know. (P8)

> *Braille Printouts*: My preferred method is still, because I am a visual learner, so I do appreciate when I have access to a hard copy braille of a graph. Just because I am more visual. So it depends on which information you're trying to convey on whether or not a table is more useful than an actual graph. (P3)

*4.3.4 The More Options, The Better: Multi-Modality.* Our last theme showed that screen-reader users consider a multi-modal solution beneficial in exploring the data contained in the visualizations. For example, P8 and P1 expressed the importance multi-modality and having various options in accessing information contained in a visualization:

> The most ideal way to access [information] is like having options, multi-modal approach. Where like, I can view it as a table or a list or I can download it as an Excel spreadsheet, you know? (P8)

> Like overview kind of a thing. Like to look at the general trend. And I want to be able to control how I consume the data. And I want to know the visual attributes of the plot or the graph. (P1)

Overall, we found that screen-reader users prefer the following techniques and strategies to improve the accessibility of online data visualizations: (1) Tabular representation; (2) Textual representation; (3) Overall trends in non-visual formats; and (4) Multi-modality.

# 5 STUDY 2: TASK-BASED USABILITY STUDY OF DATA VISUALIZATIONS

To evaluate the accessibility of online data visualizations and its effects on participants' ability to extract information accurately and efficiently, we conducted a mixed factorial experiment with screen-reader and non-screen-reader users. The experiment was conducted online, without supervision.

## 5.1 Participants

We recruited 72 participants for our study, half of whom (36) were screen-reader users. Participants were recruited using word-of-mouth, snowball sampling, and advertisements through social media channels (Facebook and Twitter). We also advertised on email distribution lists for people with disabilities. The sample size was calculated at 0.8 power to detect a large effect size at the standard .05 alpha significance threshold. Among screen-reader users, 17 identified as women, 18 as men, and 1 as genderfluid. Their average age was 44.1 ($SD$=14.1) years. In the group of non-screen-reader users, 21 identified as women, and 15 as men, with their average age being 43.4 ($SD$=12.8). The age was not significantly different between the two subject groups ($t(70)=0.23$, $p=.823$).

Screen-reader users were compensated with a $15 Amazon gift card for about 30-45 minutes of their time, whereas non-screen-reader user were compensated with $10 Amazon gift card for about 10-20 minutes of their time. The difference in the compensation amount was calculated based on the average time of study completion. No participant was allowed to partake in the experiment more than once.

## 5.2 Apparatus

We implemented an online experiment using the React framework, ensuring maximum and proper accessibility measures by testing it both with and without a screen reader ourselves and with screen-reader users. The experiment was deployed as a website on our server and the link to the website was shared with the participants.

*5.2.1 Visualization Dataset.* Following prior work [34, 40], we collected 27 different datasets from Kaggle[4]—one of the largest online resources for open datasets. In choosing the visualization datasets, our goal was to ensure that the datasets were both topically diverse and realistic. The topics were mutually agreed upon by the authors and were selected such that: (a) a dataset for a given topic only existed at most once in our pool of datasets, and (b) datasets represented different fields of interest, filtered using "tags" on Kaggle. For line charts, nine out of the 27 datasets represented temporal data, which showed a clear trend, to avoid misinterpretation.

[4]https://www.kaggle.com/datasets

**Table 2: Cardinality summary from Borkin *et al.*'s[8] dataset comprising 2,068 two-dimensional single-panel visualizations used to determine the range for the different complexity levels.**

| | N | Minimum | Maximum | Median | Mode | Mean | $25^{th}$ Percentile | $75^{th}$ Percentile |
|---|---|---|---|---|---|---|---|---|
| **Line** | 50 | 5 | 168 | 29 | 22 | 41 | 17 | 56 |
| **Bar** | 50 | 6 | 42 | 20 | 20 | 21 | 15 | 27 |
| **Scatter** | 50 | 7 | 170 | 39 | 25 | 52 | 25 | 72 |

To present participants with visualizations that have a range of data points, we further subdivided our dataset into different levels of complexity. To do so, we randomly sampled 50 visualizations for each chart type, totaling to 50×3=150 visualizations, from Borkin *et al.*'s[8] dataset comprising 2,068 two-dimensional single-panel visualizations. We used this sample to find the minimum, maximum, $25^{th}$ percentile, and $75^{th}$ percentile of their cardinalities (Table 2), which we used to determine the range for the different complexity levels. Specifically, the datasets had a random cardinality between the minimum and $25^{th}$ percentile, the $25^{th}$ and $75^{th}$ percentile, and the $75^{th}$ percentile and the maximum, for low, medium, and high complexity levels, respectively.

We used the datasets acquired from Kaggle to implement the visualizations for our online experiment, following the WCAG 2.0 Guidelines [14] in combination with the official accessibility recommendations from the visualization libraries. All visualizations were interactive by default, and were generated using all three visualization libraries. All the visualizations used in our experiment are provided in supplementary materials.

*5.2.2 Question Categories.* The study utilized four question categories to measure the accuracy of extracted information from the data visualizations. The categories were derived based on Brehmer and Munzner's task topology [11]. Specifically, we considered one Search action (*lookup and locate*) and two Query actions (*identify and compare*), similar to prior work [10]. Each category was assigned a difficulty level, determined by a discussion and mutual agreement between at least two authors based on their knowledge and familiarity of the subject matter. The categories, in an ascending order of difficulty, were:

(1) *Order Statistics*: Participants were asked either about the maximum data point or the minimum data point, chosen randomly (for example, "what is the minimum data point in the visualization?" or "what is the maximum data point in the visualization?")
(2) *Symmetry Comparison*: Participants were asked to identify the relationship between two data points (for example, "how is [random data point 1]'s value in comparison to that of [random data point 2]")
(3) *Chart Type-Specific Questions*:
   - *Value Retrieval*: Participants were asked to extract the information from a given individual data point (for example, "what is the corresponding value for [random data point]"). This question was only asked for bar charts.
   - *Trend Summary*: Participants were asked about the overall data trend (for example, "what was the overall trend of the visualization?"). We curated the dataset to ensure no ambiguity in the answer. This question was only asked for line charts.
   - *Correlation*: Participants were asked about the correlation between the dependent and independent variables in the visualization (for example, "what was the correlation between [dependent variable] and [independent variable]?"). We curated the dataset to ensure no ambiguity in the answer while keeping the data points scattered along the axes. This question was only asked for scatter plots.

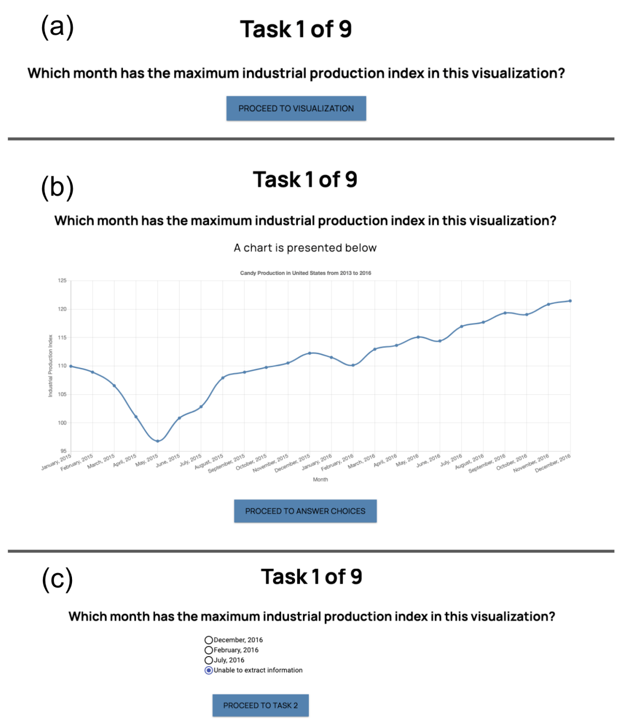


**Figure 2: Participants in Study 2 were shown three pages in a single task. (a): Page 1 presented the question to explore. (b): Page 2 displayed the same question and a visualization. (c): Page 3 showed the question again with a set of four multiple choice responses.**

All questions were multiple-choice questions with four choices: the correct answer, two incorrect answers, and the option for "Unable to extract information." The order of the choices was randomized per trial.

### 5.3 Procedure

The study was conducted online, without supervision. The participants were shown the study purpose, eligibility criteria, and the statement of IRB approval on the first page of the study. On the next page, the participants were asked to fill out a pre-study questionnaire to record their demographic information, screen-reader software, vision-loss level, and diagnosis (see Appendix C, Table 7). They were also asked about their education level, daily computer usage, and their interaction frequency with visualizations. Then, participants were shown instructions for completing the study tasks.

Each participant was shown three visualizations, created using one of three commonly utilized visualization libraries (Google Charts, ChartJS, D3). Figure 2(b) shows an example visualization. For each visualization, the participants were asked to answer three questions. Each of the three questions represented a different difficulty level (Low, Medium, High), assigned by mutual agreement from at least two authors based on the ease of extracting the answers from the visualizations. The complexity of the visualization (Low, Medium, High) and the chart type (Bar, Line, Scatter) for the visualizations were counterbalanced for each order of visualization libraries across participants.

For each of the three *Visualization Library* × *Complexity* conditions, participants were shown three pages: Page 1 contained the question to explore; page 2 displayed the question and a visualization; and page 3 presented the question with a set of four multiple choice responses from which participants chose the answer, as shown in Figure 2. The order of the questions was randomized, per visualization, and the question was shown at the top of the page. Participants were asked to interact with the visualization as they normally would in their daily lives. For screen-reader users, a study session took 30-45 minutes from start to finish, whereas for non-screen-reader users, the total time for the study session ranged between 10-20 minutes.

### 5.4 Design & Analysis

The experiment was a mixed factorial design with the following factors and levels:

- *Screen-Reader User*, between-Ss.: {yes, no}
- *Visualization Library*, within-Ss.: {ChartJS, D3, Google Charts}
- *Data Complexity*, within-Ss.: {Low, Medium, High}
- *Question Difficulty*, within-Ss.: {Low, Medium, High}

Our dependent variables were *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT). For tractability, we treated *AEI* as binary, classifying *AEI* as "inaccurate" if the user incorrectly answered the question or was unable to extract the information, and as "accurate" otherwise. As for *IT*, for screen-reader users, *IT* was calculated as the total time of focus on the root visualization element to accurately represent a screen-reader user's interaction experience. For non-screen-reader users, *IT* was calculated simply as the total time of focus on the webpage containing the visualization element.

To analyze *AEI*, we used mixed logistic regression [19] with the above factors, their interactions, a covariate of *Age*, and a random effect for *Subject* to account for repeated measures. The statistical model was *AEI* = *SRU* × *VL* + *SRU* × *CMP* + *SRU* × *DF* + *Age* + *Subject*. To analyze *IT*, we used a linear mixed model analysis of variance [18, 30], with the same model as for *AEI*.

Participants were tested over three *Visualization Library* × *Complexity* conditions, resulting in a total of 3×3 = 9 trials per participant. With 72 participants, a total of 72×9 = 648 trials were produced and analyzed in this study.

## 6 STUDY 2: RESULTS

In this section, we present the results of the experiment focusing on the *Accuracy of Extracted Information* (AEI) and *Interaction Time* (IT) for screen-reader and non-Screen-Reader Users, when interacting with online data visualizations. It is worth re-emphasizing that our work does not assess the cognitive and/or intellectual abilities of our participants, especially screen-reader users; rather, our work focuses on *AEI* and *IT* as a function of the *accessibility* of online data visualizations.

As a preliminary matter, we checked our regression models for multicollinearity by calculating the variance inflation factor (VIF) and found that multicollinearity between *AEI* and *IT* was not a concern ($VIF$=1.49).

### 6.1 Accuracy of Extracted Information (AEI)

Our results show a significant main effect of *Screen-Reader User* (SRU) on *AEI* overall ($\chi^2(1, N{=}72){=}67.22$, $p<.001$, Cramer's $V{=}0.18$), indicating that *AEI* differs significantly between the two *Screen-Reader User* groups. In fact, *AEI* is considerably lower for screen-reader users (34%) compared to non-screen-reader users (87%)—a percentage difference of 61.48%.

There was also a significant main effect of *Visualization Library* (VL) on *AEI* overall ($\chi^2(2, N{=}72){=}40.45$, $p<.001$, Cramer's $V{=}0.14$). This result indicates that *AEI* differs significantly between different visualization libraries. Figure 4 and Table 4 show the *AEI* percentages across different visualization libraries. *Google Charts* had the best performance (73%) for screen-reader users, followed by *D3* (17%) and *ChartJS* (11%). For non-screen-reader users, all three visualization libraries performed almost identical.

We also examined whether changes in *AEI* were proportionally similar or different between the visualization libraries for participants in each *Screen-Reader User* group. To do so, we examined the *SRU* × *VL* interaction, and found it to have a significant effect on *AEI* overall ($\chi^2(2, N{=}72){=}50.35$, $p<.001$, Cramer's $V{=}0.15$). This result indicates that *AEI* not only significantly differs between visualization libraries overall but also between screen-reader and non-screen-reader users. Table 4 shows *AEI* percentages across different visualization libraries for each user group.

Additionally, we also investigated the effects of *Complexity, Difficulty, Age*, and interactions between *SRU* × *Complexity*, and *SRU* × *Difficulty*, but did not find a significant effect on *AEI*. We report our findings in Table 3.

### 6.2 Interaction Time (IT)

Anderson-Darling [5] goodness-of-fit tests of normality showed that the interaction times were conditionally non-normal. Further inspection revealed these values to be conditionally lognormal, and so a logarithmic transformation was applied prior to analysis, as

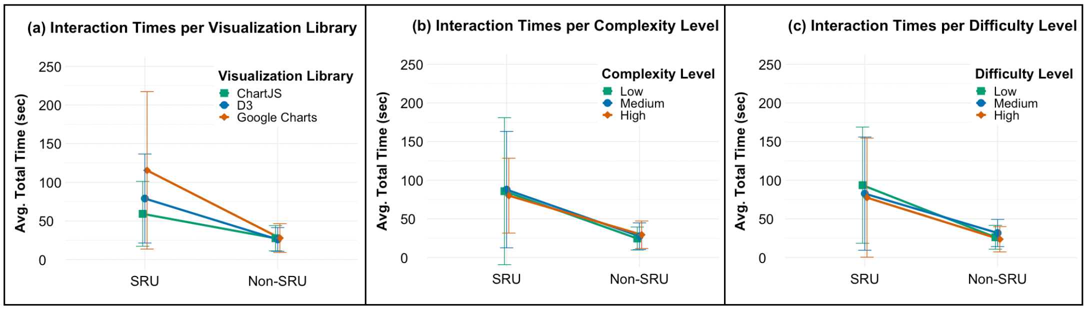


**Figure 3: Interaction Times ($IT$), in seconds, for screen-reader and non-screen-reader users, per (a) Visualization Library ($VL$), (b) Complexity Level ($CMP$), and (c) Difficult Level ($DF$).**

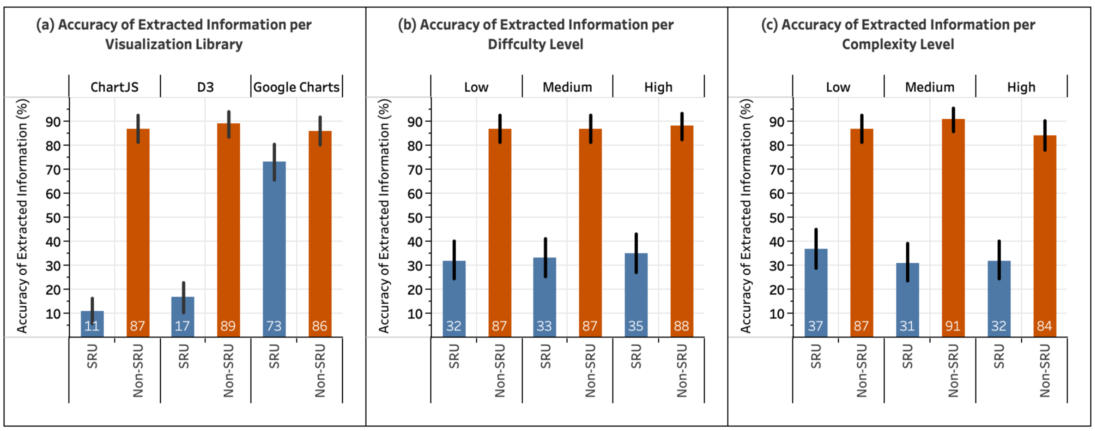


**Figure 4: Accuracy of Extracted Information (*AEI*), as a percentage, for screen-reader and non-screen-reader users, per (a) Visualization Library, (b) Difficulty Level, and (c) Complexity Level. *AEI* was classified as "inaccurate" if the user incorrectly answered the question or was unable to extract the information, and as "accurate" otherwise. The percentage represents the "accurate" answers. Therefore, higher is better.**

**Table 3: Summary results from 72 screen-reader and non-screen-reader participants using mixed logistic regression [19]. The statistical model was *AEI = SRU × VL + SRU × CMP + SRU × DF + Age + Subject*, where *Subject* was modeled with a random intercept to account for repeated measures. *AEI* is "inaccurate" if the user incorrectly answered the question or was unable to extract the information, and as "accurate" otherwise. Cramer's $V$ is a measure of effect size [16].**

| | $N$ | $\chi^2$ | $p$ | Cramer's $V$ |
|---|---|---|---|---|
| Screen Reader Usage (SRU) | 72 | 67.22 | < .001 | 0.18 |
| Visualization Library (VL) | 72 | 40.45 | < .001 | 0.14 |
| SRU × VL | 72 | 50.35 | < .001 | 0.15 |
| Complexity (CMP) | 72 | 1.94 | .380 | 0.03 |
| SRU × CMP | 72 | 2.20 | .332 | 0.03 |
| Difficulty (DF) | 72 | 0.35 | .838 | 0.01 |
| SRU × DF | 72 | 0.06 | .972 | 0.00 |
| Age | 72 | 2.70 | .100 | 0.04 |

is common practice for time measures [6, 24, 29]. Examination of this log-transformed response indicated that it was indeed normal ($p \approx .234$). For ease of communication, plots of this dependent variable are shown using the original non-transformed values.

The factor *Screen-Reader User* (SRU) had a significant main effect on *Interaction Time* (IT) overall ($F(1,69)$=115.33, $p$<.001, $\eta_p^2$=0.63). Specifically, the average IT for screen-reader users was 84.6 seconds ($SD$=75.2). For non-screen-reader users, it was 27.2 seconds

**Table 4: Numerical results for the $N = 648$ questions asked of screen reader users and non-screen reader users for each level of *Visualization Library*, *Difficulty Level*, and *Complexity Level*. $N$ is the total number of questions asked, $AA$ is the number of "accurate answers," and $AA(\%)$ is the percentage of "accurate answers."**

| | *Both Groups* | | | *Screen Reader Users* | | | *Non-Screen Reader Users* | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | AA | AA (%) | N | AA | AA (%) | N | AA | AA (%) |
| **Overall** | 648 | 392 | 60% | 324 | 109 | 34% | 324 | 283 | 87% |
| **Visualization Library** | | | | | | | | | |
| - ChartJS | 216 | 106 | 49% | 108 | 12 | 11% | 108 | 94 | 87% |
| - D3 | 216 | 114 | 53% | 108 | 18 | 17% | 108 | 96 | 89% |
| - Google Charts | 216 | 172 | 80% | 108 | 79 | 73% | 108 | 93 | 86% |
| **Difficulty Level** | | | | | | | | | |
| - Low | 216 | 129 | 60% | 108 | 35 | 32% | 108 | 94 | 87% |
| - Medium | 216 | 130 | 60% | 108 | 36 | 33% | 108 | 94 | 87% |
| - High | 216 | 133 | 62% | 108 | 38 | 35% | 108 | 95 | 88% |
| **Complexity Level** | | | | | | | | | |
| - Low | 216 | 134 | 62% | 108 | 40 | 37% | 108 | 94 | 87% |
| - Medium | 216 | 132 | 61% | 108 | 34 | 31% | 108 | 98 | 91% |
| - High | 216 | 126 | 58% | 108 | 35 | 32% | 108 | 91 | 84% |

**Table 5: Summary results from 72 screen-reader and non-screen-reader participants using a mixed-effects model analysis of variance [18, 30]. The statistical model was *IT = SRU× VL + SRU× CMP + SRU× DF + Age + Subject*, where *Subject* was modeled with a random intercept. Partial eta-squared ($\eta_p^2$) is a measure of effect size [15].**

| | $df_n$ | $df_d$ | $F$ | $p$ | $\eta_p^2$ |
|---|---|---|---|---|---|
| Screen Reader Usage (SRU) | 1 | 69 | 115.33 | < .001 | 0.63 |
| Visualization Library (VL) | 2 | 564 | 25.10 | < .001 | 0.08 |
| SRU:VL | 2 | 564 | 31.58 | < .001 | 0.10 |
| Complexity (CMP) | 2 | 564 | 6.00 | .003 | 0.02 |
| SRU:CMP | 2 | 564 | 2.01 | .136 | 0.01 |
| Difficulty (DF) | 2 | 564 | 24.07 | < .001 | 0.08 |
| SRU:DF | 2 | 564 | 12.75 | < .001 | 0.04 |
| Age | 1 | 69 | 7.03 | .010 | 0.09 |

(*SD*=16.8). The average *IT* for participants who used screen readers was 210.96% higher than for participants who did not.

There was a significant main effect of *Visualization Library (VL)* ($F$(2,564)=25.10, $p$<.001, $\eta_p^2$=0.08) on *IT*. Furthermore, the interaction between *SRU*× *VL* was also significant ($F$(2,564)=31.58, $p$<.001, $\eta_p^2$=0.10). These results indicate that *IT* not only significantly differed between visualization libraries overall, but also differentially between visualization libraries for participants in each group. Figure 3 and Table 5 show interaction times across different visualization libraries for each user group. For screen-reader users, *ChartJS* had the minimum interaction time, followed by *D3* and *Google Charts*. For non-screen-reader users, all three visualization libraries performed almost identical.

We found a significant main effect of *Difficulty (DF)* ($F$(2,564)=24.07, $p$<.001, $\eta_p^2$=0.08) on *IT*. Furthermore, the interaction between *SRU* × *DF* was also significant ($F$(2,564)=12.75, $p$<.001, $\eta_p^2$=0.04). These results indicate that *IT* not only significantly differed between question difficulty levels overall, but also differentially between difficulty levels for participants in each screen-reader user group. Figure 3 and Table 5 show interaction times across different difficulty levels for each user group.

Additionally, *Complexity (CMP)* had a significant main effect on *IT* overall ($F$(2,564)=6.00, $p \approx$.003, $\eta_p^2$=0.02), indicating that *IT* differed significantly between different complexity levels. Figure 3 and Table 5 show interaction times across different complexity levels. We also examined the interaction between *SRU*× *CMP*, but did not find a significant effect.

We investigated the effects of *Age* on *IT*. *Age* had a significant effect on *IT* ($F$(1,69)=7.03, $p$<.05, $\eta_p^2$ =0.09), indicating that *IT* differed significantly across the ages of our participants, with participants over the age of 50 showing higher interaction times by about 12.18%, compared to participants under the age of 50. Table 6 (Appendix A) shows the average *IT* for each age range, for both screen-reader and non-screen-reader users.

In addition to the above analyses, we also examined the interactions between *VL* × *CMP*, *VL* × *DF*, and *CMP*× *DF* to explore the relationship between our independent variables, but did not find any significant effects.

# 7 DISCUSSION

Our work is the first empirical work to show the inequalities screen-reader users face, fighting an uphill battle when attempting to interact with online data visualizations. To understand their challenges, we conducted contextual interviews with nine screen-reader users and found that online data visualizations are often indetectable to screen readers. And when visualizations are detected, they are predominantly identified as "blank," "frame," "graphic," or "object," leaving screen-reader users in a state of ambiguity—whether the underlying element is a visualization or something else.

We also found that screen-reader users seek a holistic view of the data first, for example, by examining overall trends. One way to gain such a holistic view is through adequate alternative text, which is rarely provided at all. Data tables made available by default in Google Charts provide more details about the data, but do not adequately support gaining a holistic overview first. For answering specific questions about the data, such as what the maximum value is, or how a specific value compares to another, screen-reader users have to remember and compute more data than is humanly possible to keep in working memory [41].

Based on these observations, we explored the gap in access to information between screen-reader and non-screen-reader users. Our controlled experiment showed that due to the inaccessibility of online data visualizations, screen-reader users have to interact 210% as long with an online data visualization as compared to non-screen reader users. Additionally, non-screen-reader users were able to accurately answer 87% of questions in our experiment after viewing online data visualizations, screen-reader users only answered 34% correctly—a performance gap of 61% in accurately extracting information, attributed the online visualizations being inaccessible.

Interestingly, different visualization libraries mean different trade-offs for screen-reader users. For example, Google Charts appends a non-visual data table to the embedded visualization element, attributing to screen-reader users being able to answer more questions accurately (73%) as compared to D3 (17%) and ChartJS (11%). But the trade-off is that navigating through the data table consequently increased their interaction time with the visualizations, attributing to screen-reader users spending almost two minutes on each visualization—1.5 and 1.9 times more as compared to D3 (1.3 minutes on average) and ChartJS (1 minute on average), respectively.

Providing adequate alternative text is a great way to quickly give screen-reader users an overview of what the visualization shows. For answering specific questions, however, a data table is currently the best option and allows screen-reader users to accurately answer a great percentage of questions (73%)—even though this is still vastly different from the performance of non-screen-reader users (87.3% averaged across three visualization libraries).

Screen-reader users indicated in our contextual interviews that a tabular format to represent data is indeed a good option to allow access to data. However, a tabular representation of the data, in most cases, only allows for sequential exploration, thereby increasing the time it takes and working memory load. This is especially the case when a visualization contains several data points. Other participants preferred techniques such as summarization (as sometimes provided with alternative text) or data sonification, where data is transformed into auditory signals. Given the different preferences of different screen-reader users, who may differ further based on their interest in the topic of the visualization, our findings emphasize that a "one-size-fits-all" approach is not necessarily the best solution, especially when the goal is to make visualizations as meaningful and as accessible for screen-reader users as they are for non-screen-reader users.

## 7.1 Design Recommendations

Our results have several important implications for the design of future online data visualizations that equalize access for screen-reader users.

**Recommendation 1: Online data visualizations need to be discoverable and comprehensible.** The biggest disadvantage screen-reader users face with respect to online visualizations is that many visualizations are indetectable to screen readers. Therefore, the information present within such visualizations is kept completely hidden from screen-reader users. We recommend developers and visualization creators to adequately provide alternative text and use ARIA attributes to improve the discoverability of the visualization elements.

**Recommendation 2: Online data visualization libraries need to offer screen-reader users both a holistic view of the data and support for a drilled-down exploration.** As our studies showed, current popular online visualization libraries either encourage the use of alternative text (ChartJS and D3) or they automatically include a data table (Google Charts). However, even though alternative text can support gaining a holistic overview, it is usually not comprehensive enough to allow understanding the data in detail. In contrast, data tables allow exploration of the data in detail, but adds a mental burden for screen-reader users to process all data and lacks efficient support for gaining a holistic overview. We recommend providing support for both to better support screen-reader users' exploration preferences, as found in this study.

**Recommendation 3: Alternative text could be auto-generated based on the underlying data.** Our findings showed that alternative text is rarely available, and when it is, it is mostly inadequate. Additionally, the quality of the alternative text primarily depends on the developer, which can produce inconsistencies in visualization interaction experiences for screen-reader users from one visualization to another. Therefore, we recommend dynamically generating alternative text, similar to how it has been proposed in prior work [32, 38]. Additionally, given the fact that every user is unique and may prefer different information in the alternative text, we recommend generating personalized alternative text to cater to the individual preferences of screen-reader users.

**Recommendation 4: Online data visualization libraries should offer different modes for exploring the data.** Our study found that while data tables are beneficial, many screen-reader users preferred additional approaches—such as data sonification—to explore data. Presenting screen-reader users with multiple modes for exploring visualizations would improve their overall experience with online data visualizations. This is especially true in the case of complex visualizations, which would likely require more exploration time.

Altogether, we found that current approaches to make online data visualizations accessible insufficiently support screen-reader users in fully extracting the data and require them to spend significantly more time on visualizations as compared to non-screen-reader users.

## 8 LIMITATIONS & FUTURE WORK

Our experiments included three different visualization libraries and two-dimensional data to evaluate and compare the visualization interaction experiences of screen-reader users with that of non-screen-reader users. While we chose the most commonly used visualization libraries and all of our participants were evaluated using two-dimensional data, different visualization libraries and data dimensionality could extend this work. To address this limitation, future work could examine the visualization interaction experiences of screen-reader users in comparison to that of non-screen-reader users, using different visualization libraries, such as Highcharts and Recharts, and a different data dimensionality, such as three-dimensional data.

The empirical findings from our work indicate that due to the inaccessibility of online data visualizations, the visualization interaction experiences of screen-reader users are significantly worse than that of non-screen-reader users. We found that visualizations are either undetectable by screen readers or are identified as "blank," "frame," "graphic," or "object." Future work could use this finding to build visualization tools that allow for the visualizations to be meaningfully recognized by the screen readers. We also found that screen-reader users have individual preferences of techniques and strategies to make visualizations more accessible. Using this finding, future work can cater to the individual needs of the screen-reader users in a personalized manner, and develop tools that enable the screen-reader users to explore the visualizations and extract information both holistically and in a drill-down manner, whichever way they prefer.

## 9 CONCLUSION

Online data visualizations are an increasingly popular mode of communicating information online. We empirically evaluated the experiences of screen-reader users interacting with online data visualizations as compared to those of non-screen-reader users. We used a mixed-methods approach, employing both contextual interviews and a quantitative task-based experiment, with 45 screen-reader and 36 non-screen-reader users. Our findings indicate that screen readers often do not even "see" data visualizations, and when they do, these visualizations are commonly identified as "object" or similar, without conveying the existence of a data visualization. Our results also show that those visualizations that are detected by screen readers still inadequately support screen-reader users. In fact, extracting information from online data visualizations, in their current state, using a screen reader is both inaccurate and time-consuming, creating significant disparities between screen-reader and non-screen-reader users. Thus, this work emphasizes the need for accessible online data visualizations. It is our hope that our findings will motivate and guide designers, developers, and researchers in the creation of more accessible online data visualizations.

## ACKNOWLEDGMENTS

This work was supported in part by The Mani Charitable Foundation and in part by NSF award 2006104. We extend our gratitude to the AccessComputing staff for their support and assistance in recruiting participants. We would also like to thank the anonymous reviewers for their helpful comments and suggestions. Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect those of any supporter.

## REFERENCES


[1] [n.d.]. Blindness Statistics | National Federation of the Blind. https://nfb.org/resources/blindness-statistics?q=resources%2Fblindness-statistics. (Accessed on 04/10/2021).

[2] [n.d.]. WAI-ARIA Overview | Web Accessibility Initiative (WAI) | W3C. https://www.w3.org/WAI/standards-guidelines/aria/. (Accessed on 04/11/2021).

[3] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[4] Dustin Adams, Tory Gallagher, Alexander Ambard, and Sri Kurniawan. 2013. Interviewing blind photographers: design insights for a smartphone application. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–2.

[5] Theodore W Anderson and Donald A Darling. 1954. A test of goodness of fit. *Journal of the American statistical association* 49, 268 (1954), 765–769.

[6] Donald A Berry. 1987. Logarithmic transformations in ANOVA. *Biometrics* (1987), 439–456.

[7] Syed Masum Billah, V. Ashok, D. Porter, and I. Ramakrishnan. 2017. Ubiquitous Accessibility for People with Visual Impairments: Are We There Yet? *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017).

[8] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. 2018. Visualizing ranges over time on mobile phones: a task-based crowdsourced evaluation. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 619–629.

[11] Matthew Brehmer and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2376–2385.

[12] Stephen Brewster. 2002. Visualization tools for blind people using multiple modalities. *Disability and rehabilitation* 24, 11-12 (2002), 613–621.

[13] Craig Brown and Amy Hurst. 2012. VizTouch: automatically generated tactile visualizations of coordinate spaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 131–138.

[14] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)* (2008).

[15] Jacob Cohen. 1973. Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and psychological measurement* 33, 1 (1973), 107–112.

[16] Christopher J Ferguson. 2016. An effect size primer: A guide for clinicians and researchers. (2016).

[17] John H Flowers, Dion C Buhman, and Kimberly D Turnage. 1997. Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors* 39, 3 (1997), 341–351.

[18] Brigitte N Frederick. 1999. Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results. (1999).

[19] AR Gilmour, RD Anderson, and AL Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 3 (1985), 593–599.

[20] Matthew Graham, Anthony Milanowski, and Jackson Miller. 2012. Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance

Ratings. *Online Submission* (2012).

[21] William Grussenmeyer, Jesel Garcia, Eelke Folmer, and Fang Jiang. 2017. Evaluating the accessibility of the job search and interview process for people who are blind and visually impaired. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*. 1–4.

[22] Donald P Hartmann. 1977. Considerations in the choice of interobserver reliability estimates. *Journal of applied behavior analysis* 10, 1 (1977), 103–116.

[23] Karen Holtzblatt and Sandra Jones. 1995. Conducting and analyzing a contextual interview (excerpt). In *Readings in Human–Computer Interaction*. Elsevier, 241–253.

[24] MH Hoyle. 1973. Transformations: An introduction and a bibliography. *International Statistical Review/Revue Internationale de Statistique* (1973), 203–223.

[25] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2014. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (2014), 420–433.

[26] Amy Hurst and Shaun Kane. 2013. Making" making" accessible. In *Proceedings of the 12th international conference on interaction design and children*. 635–638.

[27] Shaun K Kane, Jeffrey P Bigham, and Jacob O Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 73–80.

[28] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications* 35, 5 (2015), 84–90.

[29] Eckhard Limpert, Werner A Stahel, and Markus Abbt. 2001. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience* 51, 5 (2001), 341–352.

[30] Ramon C Littell, PR Henry, and Clarence B Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *Journal of animal science* 76, 4 (1998), 1216–1231.

[31] David K McGookin and Stephen A Brewster. 2006. Soundbar: exploiting multiple views in multimodal graph browsing. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. 145–154.

[32] Silvia Mirri, Silvio Peroni, Paola Salomoni, Fabio Vitali, and Vincenzo Rubano. 2017. Towards accessible graphs in HTML-based scientific articles. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1067–1072.

[33] Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE Publications, inc.

[34] Azzurra Pini, Jer Hayes, Connor Upton, and Medb Corcoran. 2019. AI Inspired Recipes: Designing Computationally Creative Food Combos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[35] Jonathan Potter and Margaret Wetherell. 1987. *Discourse and social psychology: Beyond attitudes and behaviour*. Sage.

[36] Bahador Saket, Alex Endert, and Çağatay Demiralp. 2018. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics* 25, 7 (2018), 2505–2512.

[37] Anastasia Schaadhardt, Alexis Hiniker, and Jacob O Wobbrock. 2021. Understanding blind screen reader users' experiences of digital artboards. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM New York, NY, USA, To appear.

[38] Ather Sharif and Babak Forouraghi. 2018. evoGraphs—A jQuery plugin to create web accessible graphs. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 1–4.

[39] Lei Shi, Idan Zelzer, Catherine Feng, and Shiri Azenkot. 2016. Tickers and talker: An accessible labeling toolkit for 3D printed models. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 4896–4907.

[40] Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system. In *23rd International Conference on Intelligent User Interfaces*. 293–304.

[41] Jonathan E Thiele, Michael S Pratte, and Jeffrey N Rouder. 2011. On perfect working-memory performance with large numbers of items. *Psychonomic Bulletin & Review* 18, 5 (2011), 958–963.

[42] Frances Van Scoy, Don McLaughlin, and Angela Fullmer. 2005. Auditory augmentation of haptic graphs: Developing a graphic tool for teaching precalculus skill to blind students. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, Vol. 5.

[43] Sue Widdicombe and Robin Wooffitt. 1995. *The language of youth subcultures: Social identity in action*. Harvester/Wheatsheaf.

[44] Susan P Wyche and Rebecca E Grinter. 2009. Extraordinary computing: religion as a lens for reconsidering the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 749–758.

[45] Wai Yu and Stephen Brewster. 2002. Multimodal virtual reality versus printed medium in visualization for blind people. In *Proceedings of the fifth international ACM conference on Assistive technologies*. 57–64.

[46] Wai Yu, Ramesh Ramloll, and Stephen Brewster. 2000. Haptic graphs for blind computer users. In *International Workshop on Haptic Human-Computer Interaction*. Springer, 41–51.

[47] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment: a case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 1 (2008), 1–28.

## A INTERACTION TIME PER AGE RANGE

**Table 6: Summary results from 72 screen-reader and non-screen-reader participants showing the numerical results for *Interaction Time* (IT), for each age range.** $N$ **is the total number of participants for the given age range, *Mean* is the average IT in seconds, and** $SD$ **represents the standard deviation.**

| | *Both Groups* | | | *Screen Reader Users* | | | *Non-Screen Reader Users* | | |
|---|---|---|---|---|---|---|---|---|---|
| Age Range | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| 18-20 | 2 | 65.8 | 43.9 | 1 | 96.8 | - | 1 | 34.8 | - |
| 20-30 | 12 | 40.6 | 28.0 | 6 | 63.7 | 20.3 | 6 | 17.4 | 4.9 |
| 30-40 | 14 | 44.2 | 23.1 | 7 | 60.6 | 16.5 | 7 | 27.7 | 16.1 |
| 40-50 | 19 | 67.5 | 78.7 | 10 | 106.7 | 93.1 | 9 | 24.0 | 10.7 |
| 50-60 | 14 | 47.6 | 27.3 | 5 | 79.1 | 19.2 | 9 | 30.1 | 8.1 |
| 60-70 | 9 | 64.8 | 33.6 | 6 | 76.8 | 33.7 | 3 | 40.6 | 19.0 |
| > 70 | 2 | 127.0 | 127.4 | 1 | 217.1 | - | 1 | 36.9 | - |

# B SCREEN-READER PARTICIPANTS FOR STUDY 2

**Table 7: Screen-Reader Participants for Study 2, their gender identification, age, screen-reader, vision-loss level, and diagnosis. Under the *Gender* column, *M* = *Male*, *F* = *Female*, and *GF* = *Genderfluid*.**

| Subject | Gender | Age | Screen-Reader | Vision-Loss Level | Diagnosis |
|---|---|---|---|---|---|
| S3 | F | 67 | Fusion | Partial vision, Lost vision gradually | Juvenile Macular Degeneration |
| S4 | M | 55 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| S5 | F | 30 | NVDA | Lost vision gradually, Partial vision | Retinopathy of Prematurity |
| S6 | F | 63 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| S12 | F | 35 | JAWS | Blind since birth | Leber's Congenital Amaurosis |
| S13 | M | 41 | JAWS | Lost vision gradually | Juvenile Onset Open Angle |
| S15 | M | 40 | JAWS | Partial vision, Lost vision gradually | Retinitis Pigmentosa |
| S16 | M | 47 | JAWS | Lost vision gradually | Leber's Congenital Amaurosis |
| S17 | M | 35 | JAWS | Blind since birth | Leber's Congenital Amaurosis |
| S18 | F | 51 | JAWS | Blind since birth | Blind |
| S19 | M | 51 | JAWS | Blind since birth | Blind |
| S20 | M | 31 | NVDA | Blind since birth, Lost vision gradually | Peter's Anomaly |
| S21 | M | 48 | NVDA | Lost vision gradually | Retinitis Pigmentosa |
| S22 | GF | 24 | VoiceOver | Partial vision | Partial Sight Impairment |
| S23 | F | 27 | NVDA | Blind since birth, Lost vision gradually | Retinal Detachment |
| S25 | F | 64 | JAWS | Partial vision | Did not disclose |
| S26 | F | 39 | Fusion | Lost vision gradually | Did not disclose |
| S28 | F | 53 | JAWS | Lost vision gradually | Optic Neuropathy |
| S29 | F | 22 | NVDA | Blind since birth | Retinitis Pigmentosa |
| S30 | M | 60 | JAWS | Partial vision | Optic Neuropathy |
| S31 | M | 46 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| S32 | M | 29 | NVDA | Blind since birth | Leber's Congenital Amaurosis |
| S33 | F | 46 | JAWS | Lost vision gradually | Optic Neuritis/Atrophy, Diabetic Retinopathy |
| S34 | M | 57 | JAWS | Lost vision gradually | Glaucoma |
| S35 | M | 18 | NVDA | Blind since birth | Retinopathy of Prematurity |
| S36 | F | 63 | JAWS | Lost vision gradually | Cataracts |
| S40 | F | 28 | NVDA | Blind since birth, Lost vision gradually | Optic Nerve Hypoplasia and Glaucoma |
| S41 | M | 27 | NVDA | Blind since birth | Optic Nerve Hypoplasia |
| S42 | F | 68 | JAWS | Blind since birth | Retinopathy of Prematurity |
| S44 | F | 34 | JAWS | Blind since birth | Renal Retinal Dysplaysia |
| S46 | F | 72 | JAWS | Lost vision gradually | Retinitis Pigmentosa |
| S47 | M | 39 | JAWS | Lost vision gradually | Retinitis Pigmentosa, Maculae Degeneration |
| S48 | M | 47 | JAWS | Lost vision gradually, Partial vision | Leber's Congenital Amaurosis |
| S49 | M | 41 | JAWS | Blind since birth | Micropthalmia |
| S50 | M | 43 | Other | Blind since birth | Retinopathy of Prematurity |
| S51 | F | 46 | JAWS | Lost vision gradually | Optical Nerve Damage |