Reimagined-Vis: Project Proposal

Kathryn (Katy) Koenig **Lucy Li** **Belen Michel Torino**

OVERVIEW

Throughout daily life, we are often confronted with data visualizations for diverse purposes. This includes from a simple graphs at newspapers till sophisticated visualizations that represent stocks values time series. This research is focused on people who are blind or legally-blind, specifically adults because they encounter data visualizations more often and they are expected to know how to read and interpret them. Often, those who are creating graphs do not take into account our target population. Although there is braille for reading, there is no system for reading data. An interface or system that can quickly adapt to any visualization, can help those who are blind or legally-blind to understand what is being represented. An interface or system would also be much quicker than trying to create a different representation for every data visualization. There are many different possibilities for solving this issue, and we hope to find one that is easy to access and effective. Throughout our research we hope to answer the following questions:

1. What is the best way to help represent visual data for those who are blind?

2. What type of data is the most important/common in daily life to translate?

RESEARCH METHODS

We will conduct semi-structured interviews with a length of 30-45 minutes. During these interviews, we will use an experience-centered approach, focused on their current use of technology and digital consumption of data and visualizations.

Due to time constraints (<10 week period), we believe semi-structured interviews will allows us to get valuable and accurate information and, at the same time, ameliorate scheduling difficulties that would arise to coordinate a focus group.

We expect to have at least five participants, but, understanding that connecting with people from the blind and low vision community might be unfeasible in the short term, we might only be able to interview three low-vision participants and three colorblind people. We expect to have a snowball effect to find participants.

We will first ask some "ice breaker" questions where we intend to get to know them better: understand their level of vision and their daily challenges in relation to reading/consuming data visualizations in professional or academic settings. We will then ask questions about their digital experiences when consuming data and visualizations. Finally, we will design an activity where they will select the ways in which they believe some concepts (i.e. trends, size, bounciness, etc) can be better represented for them. We have attached our interview questions as well as our consent forms to the appendix of this proposal. Note, we plan to provide consent forms digitally so that our participants can use text-to-speech technologies to understand them. We will also orally confirm consent at the onset of our interview.

We will record the interviews and at the same time use Google's Voice to Text technology to transcribe it while we interview. After the meeting, we will improve Google transcriptions manually. Later, we will codify the interviews and conduct affinity diagramming. Our interviews will be conducted both in English and Spanish depending on the language that the participants feel most comfortable with.

In order to analyze the data, we will create a code book that will help us to structure the text of the interviews in a way that allow us to understand better, the challenges that the participants face, the technologies that they are already using, and the new technologies that they will feel most comfortable using in the feature.

DESCRIPTION OF CO-DESIGNERS

In particular, we are studying those are blind or near-blind. In particular, blind or near-blind adults as this is the demographic most likely to need to understand data visualizations. Of course we hope what we develop can help all age groups, but we will focus on designing for blind or near-blind adults.

Data visualizations are common in daily life, and not having access to them can lead to missing lots of information. We are focusing on adults specifically, because they are the most likely to encounter data visualizations that are important to them. Common examples are stock graphs, data graphs, sports stats, political polls, etc. When you are unable to see certain data visualizations it is very difficult to understand how the data has changed and progressed over time. This is why we hope to help blind or near-blind people in order to help them better understand the world and how it is changing around them.

PROBLEM AND OUR SOLUTION

People with visual impairments find several challenges on their daily life. In recent years, many of those have been mitigated with the help of technology. However, there are some challenges that were not solved yet, such as being able to consume data visualizations. One participant mentioned that when he reads a text that has visualizations, charts or

other type of images, he skips the visualizations, accepting loss of the information presented. He also mentioned that his university allowed him access to a limited but useful book-sharing program, in which he can find texts with paragraph descriptions of where the visualizations. While this option is better than the complete loss of information due to entirely skipping the visualization, those descriptions are sometimes biased due to the writer's interpretations.

Data visualizations are now present everywhere on our daily life. We encounter them in the newspaper, magazines, books, at work or at school. People with low vision that try to access this information are currently very limited to either finding a digital book with good descriptions of those visualizations, or must request assistance from professors, classmates, co-workers, friends or family to describe the data shown. One participant that is currently pursuing an MS degree told us that he likes to explore data and to program, however, he sometimes designs visualizations that he is not able to fully read by his own.

People with low vision usually support their daily activities either with their phones or their laptops, where they have access to read-a-loud applications. Also, their limited access to data visualizations is often achieved through these two devices, too.

We will develop a system to change the status quo in data visualizations: instead of placing the burden of translating data visualizations to a consumable product for people with visual impairments, we seek to create a program to allow data visualization designers to easily create a counter part that can be understood by people with visual impairments and allow this product to be easily integrated into digital version of text documents, just as a data visualization would be.

LITERATURE REVIEW

Exploration of alternative renderings of data visualizations can be divided into data physicalization and data sonification.

While previous research attempts to increase the speed and ease of the creation of direct translations of plots[6], the printing of one-to-one representations of data visualizations onto embossed paper or as 3D objects fails convey the overall information in its visual counterpart[7]. Specifically, while visual learners can understand a plot in a whole-to-part format, using direct translation, tactile learners must interpret plots part-to-whole[7], introducing difficulty in understanding the plot in its entirety.

Other data physicalizations[5] presents data through novel encoding of its attributes such as temperature, haptic feedback, friction, weight[5] and taste[9]. Much of this research seeks to deepened our understanding of data through exploration with multiple senses and is not targeted at ameliorating issues faced by the low vision and blind community when encountering data visualizations regularly. This presents issues of scaling due to time and cost needed to create data physicalization. Additionally, while people with visual impairments prefer 3D printed graphics over embossed paper graphics, the low resolution of affordable 3D printers lack the precision to print legibly [7].

Other approaches to the translation of data visualization for people with visual impairments have sought solutions in sonification. Motivated through the desire to teach statistics to people with visual impairments, previous research has found that individuals could understand distributional properties and the first four moments of data, i.e. mean, variances, skew and kurtosis, as well as slope and linearity through sonified data[4],[2]. Furthermore, the successful translation of a box-and-whisker plot to an "arpeggio plot" shows that the mapping of a specific data visualization to a particular data sonification is possible[2].

Unfortunately, much of the data sonification research lacks participants experiencing visual impairments[4], [2] due to lack of access to the community, so much of this validation has been completed using sighted test subjects. Additionally, successful use of data sonification requires a small amount of training as there is a learning curve to adapt to the way in which sounds are presented[2], but previous research lacks guidelines not only on the grammar of sonification[3] but also on methods of training for widespread use of data sonification.

Previous literature in both data physicalization and data sonification notes the lack of standards and democratization to allow experimentation of these methods[5],[2]. We seek to address these issues in order to foster widespread use of accessible explorations of data.

EXISTING SYSTEMS

Descriptions of data visualizations can be easily understood by people with visual impairments using text-to-speech (TTS) technologies and screen readers. TTS technologies though are only useful when a visual designer prepared accompanying descriptions of their data visualization, which poses two issues regarding accessibility: (1) it is not standard practice to include a description with each data visualization; (2) descriptions are subject to the creator's bias in that the user only hears the statistics which are described. For example, if outliers are visibly present in a scatter plot, but a description only includes information regarding the correlation between x and y variables, TTS technology will fail to provide all information conveyed in the plot.

Currently, software exists to explore data visualizations aurally. Specifically, SAS, the statistical software suite, offers SAS Graphics Accelerator[8], a free Google Chrome extension allows creation of visualizations and accompanying sonification of data. It claims to provide sonification of varying plots, including bar charts, box plots, and histograms. For visualizations which the program cannot translate into sound, e.g. contour plots, mosaic plots and needle plots, the software offers creation of descriptions which can be heard using screen readers. Additionally, SAS Graphs Accelerator allows users to create data tables, import data and extract data tables on webpages to create their own accessible graphs.

The sonified plots in SAS Graphics Accelerator do not always convey data clearly. For example, data that becomes a familiar box-and-whisker plot with one categorical type becomes a single sound: all information regarding outliers and spread are lost. Furthermore, every sonification is initialized with a

dataframe, whether from a csv file or a table from a webpage, as opposed to translating an existing data visualization into sound. [8]

Recent computer science research aims to translate data visualizations into data tables on websites using machine learning. A prototype extension for Google Chrome interprets a plot through a deep neural network and outputs raw data for a user's inspection[1].

While these technologies certainly seek to make data sonification more accessible, both place the burden of sonification design on the consumer: the visualization designer is not responsible for creating a sonification. This workflow upholds the norm of inaccessible visualization instead of allowing data analysts to create as rich a sonification of data as they do a visualization.

EVALUATION CRITERIA

To evaluation our system, we plan to conduct semi-structured follow-up interviews with our participants using the criteria described below. Broadly, we will evaluate our system based on both the statistical understanding gained from use and the general user experience. Below, we enumerate our specific criteria:

Statistical Understanding

1. Does the user have a better understanding of the data after using our project?

   To ensure that our product adds value to the user, we will randomly select participants to use our output while other participants will use their current methods of evaluating visualizations. We will then ask questions regarding the underlying data to see if participants that used our product can convey more information regarding the data than those that did not.

2. Does our product's output convey the same information as a data visualization?

   To test that our product is comparable to a visualization in relaying data, we will translate a visualization using our product. Then, our participants will visual impairments will consume the output of our product and be asked a series of questions regarding the underlying data. We will then have people without visual impairments view the original data visualization and ask them the same questions regarding the underlying data. We will then compare the results to judge the accuracy with which our product conveys information.

User Experience

1. Is our output easy to use and access?

   We want to drastically lower the burden translating data visualizations to a product that can be consumed by a person with visual impairments. To do so, our data physicalization or sonification must integrate as seamlessly into text analysis as a data visualization. Therefore, we will ask our participants to think aloud when using the output of our program as well as give feedback regarding its integration.

2. How large is our learning curve?

   We assume that there will exist a learning curve as our data must be encoded in a novel format. We want to ensure that the training period to consume the product is quite small. Therefore, we will not only time our training process but also ask our participants to give their opinions regarding the training materials and time period.

3. How confident in the user in their interpretation of data?

   We want our users to not only interpret data but also feel confident in the insights gleaned from our product. While direct questions regarding the users ideas of the underlying data and the strength of their idea will be used, we also hope to evaluate participants' convictions by introducing doubt. For example, after asking for the information they grasped from a sonified bar chart, we plan to then suggest that some of this information is incorrect and record whether participants persist in their opinions of the data or acquiesce to different understanding of the data. In our debrief, we plan to acknowledge this introduction of doubt to ensure that our participants are fully aware of our entire user study design.

4. Does the user feel empowered and more knowledgeable regarding the data after using our product?

   Most importantly, we want our product to bring additional knowledge to our users that text analysis cannot and for our users to feel good about using our product. As empowerment is difficult to address directly, we plan to ask users for overall evaluations of the product and for their feelings towards future use of the product.

REFERENCES

[1] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum* (2019). DOI: http://dx.doi.org/10.1111/cgf.13686

[2] John Flowers, Kimberly Turnage, and Dion Buhman. 2005b. Desktop Data Sonification: Comments on Flowers et al., ICAD 1996. *TAP* 2 (10 2005), 473–476. DOI:http://dx.doi.org/10.1145/1101530.1101545

[3] John H Flowers, Dion C Buhman, and Kimberly D Turnage. 2005a. Data sonification from the desktop: Should sound be part of standard data analysis software? *ACM Transactions on Applied Perception (TAP)* 2, 4 (2005), 467–472.

[4] John H. Flowers and Terry A. Hauer. 1993. "Sound" alternatives to visual graphics for exploratory data analysis. *Behavior Research Methods, Instruments, & Computers* 25, 2 (01 Jun 1993), 242–249. DOI: http://dx.doi.org/10.3758/BF03204505

[5] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3227–3236.

[6] Chandrika Jayant, Matt Renzelmann, Dana Wen, Satria Krisnandi, Richard Ladner, and Dan Comden. 2007. Automated tactile graphics translation: in the field. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*. 75–82.

[7] Alan Lundgard, Crystal Lee, and Arvind Satyanarayan. 2019. Sociotechnical Considerations for Accessible Visualization Design. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 16–20.

[8] SAS. 2019. SAS Products SAS Graphics Accelerator. `http://support.sas.com/software/products/graphics-accelerator/index.html`. (2019). Accessed: 2020-01-27.

[9] Yun Wang, Xiaojuan Ma, Qiong Luo, and Huamin Qu. 2016. Data Edibilization: Representing data with food. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 409–422.

APPENDIX

University of Chicago Online Consent Form for Research Participation

Study Title: Reimagines-Vis

Researcher(s): Kathryn (Katy) Koenig, Lucy Li, Belen Michel Torino

Description: We are researchers at the University of Chicago doing a research study about how people with visual impairments access to data visualizations. This research will allow us to better understand the challenges they face, the tools that they use now a days and the possible tools that could be developed to help them explore and consume data visualizations. To participate better understand participants and their every day challenges we will ask them their level of vision.The procedure will involve a semi-structured interview of an estimated length of 30/45mins.

Risks and Benefits: Your participation in this study does not involve any risk to you beyond that of everyday life. Taking part in this research study may not benefit you personally, but we may learn new things that could help others.

Rights as a participant: Your participation in this interview is voluntary. If there are any questions in the survey that you do not wish to answer, you may skip them. You may stop participating at any time.

Confidentiality: The researchers will take all possible steps to keep information about you confidential and anonymous, and to protect any information collected as part of this project from unauthorized disclosure, tampering, or damage.

- Your name will not be viewed by anyone except the three researchers in this team.
- All names will be deleted from our data files before saving and preparing the data for analysis.All data will be stored in password-protected files on password-protected computers or on secure cloud storage only accessible by members of this research.
- After the project is completed, any data that is shared for the purpose of scientific inquiry will remain completely anonymous and all identifying information will have been removed. When results from this study are shared, your name or the names of the people you mention during the interviews will never be mentioned.
- Audio recordings and data collected as part of this research will be stored de-identified. Please note that if you decide to withdraw from the study and de-identified materials or data have already been submitted to an archive, it is possible that your data will not be able to be removed. When possible, however, your data will be withdrawn upon your decision to leave the study.

Contacts & Questions: If you have questions or concerns about the study, you can contact the researchers at:

Belen Michel Torino

belenmichel@uchicago.edu

+1(312)493-7726

Consent: Participation is voluntary. Refusal to participate or withdrawing from the research will involve no penalty or loss of benefits to which you might otherwise be entitled.

By clicking “Agree” below, you confirm that you have read the consent form, are at least 18 years old, and agree to participate in the research. Please print or save a copy of this page for your records.

☐ **I agree to participate in the research**

☐ **I do NOT agree to participate in the research. In this case we thank you for your consideration and we will not conduct the interview.**

Interview guide

The following guide will be used to conduct semi-structured interviews among people with visual impairments.

Getting to know the participant:

1. How would you describe your visual impairment?
2. Which is your level of vision?
3. Do you know the name of your diagnosis?
4. Are you currently working? Are you currently studying?
   (we will use this response to ask the following questions in the setting mentioned)
5. Is there any sense in which you rely the most to compensate your view?

Encountering data Visualizations:

1. Which tasks are the most challenging for you at school/work? why?

   [Describe what exactly we mean by data visualization]
2. When do you encounter data visualizations?
3. What do you do when you encounter data visualizations?
4. What tools do you use to interpret data visualization?
5. What do you like about this tool?
6. What do you dislike about this tool?
7. Has there ever been a moment in time where not being able to analyze
   data visualizations prevented you from doing well in a class/ job?

Creating data Visualizations:

1. Have you created data visualizations?
   [Maybe remind that this could be simple]

   If yes:
2. What are the tools you have used to create visualizations?
3. How have you evaluated your creations?

Potential Solutions:

1. Do you have any idea of what can help you to access visualizations?

   We will tell them about the literature review we have done, mentioning the different options explored (3d printings, music, deep learning + voice, firework experience).
   If the interview is conducted in person, this last part will be done in a dynamic way with cards explaining each of the 4 options. Otherwise, if the interview is conducted virtually, this section will be guided by the following questions.
2. Do you think any of these options could be useful to you?
3. Could you please rank them from the best solution to the worst in terms of:
   - Getting all the information?
   - Easy to use?
   - Solving all your challenger?
   - Global experience?