Optimizing Data Visualization for Social Good: A Thematic Analysis

Arleen Aksay

Iris Yusupov

Dr. Michael Friendly

June 25, 2021

Introduction

Presenting data with compelling visualizations can be an effective tool to communicate information to a target audience (Cleveland, 1985). Graphical representation of data does not have to be confined to the research lab; journalists, educators, and health practitioners may need to communicate data clearly and effectively to the public, students, and patients, respectively. Often, infographics and images can be a quick and digestible method to disseminate information on social media to a wider audience.

Data for Good is a movement that has been gaining momentum within the last three years. Data is utilized to engender a social change and a key element of the movement is, "harnessing the power of data visualization" (Avidon, 2020). One organization that employs data visualization to highlight and promote an organization working to solve a social issue is Viz for Social Good (Viz for Social Good, 2020). This organization consists of volunteers partnering with non-profit groups to help transform their data into visually impactful graphics meant to promote their work and raise awareness. This is an especially valuable collaboration considering the scarcity of resources within non-profit organizations limiting their employment of graphic designers and data visualization specialists. An example of such a partnership is with the non-profit Crowd2Map, which helps rescue girls at risk of female genital mutilation in Tanzania and move them to safe houses (Dash & Gabor, 2021). See Figure 1 below as one of the maps created by a volunteer to depict the work of Crowd2Map to date, raise awareness, and share the importance of the cause. Organizations working toward social good can use data visualizations to communicate their work quickly and effectively, which is an asset in the age of social media.

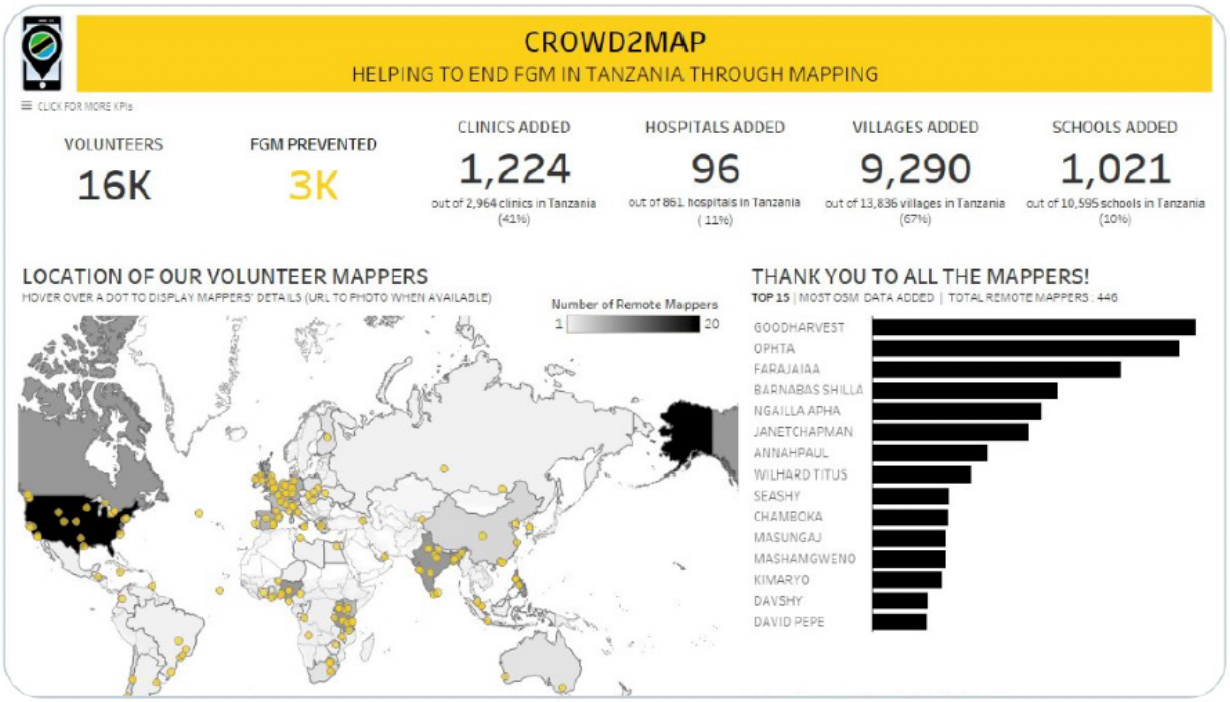

Figure 1. A map created by volunteer, Delphine Rouvillois, shared on social media as her submission to Viz for Social Good. The goal of the visualization was to highlight the work of

Crowd2Map volunteer mappers, who dedicate their time in mapping out remote rural areas to facilitate in the rescue of girls at risk of female genital mutilation.

Research on Elements of Data Visualization

Graphical presentation can impact judgements and interpretations of the information. In fact, the term "graphical perception" refers to the ability of observers to interpret and encode information presented in graphs (Cleveland & McGill, 1984). The literature has shown that visual variables such as shape, length, colour, and position are key factors in visualizing data. For example, Cleveland and McGill (1984) demonstrated that viewers are more accurate at judging the length of visual data rather than its area or angles. Due to this finding, it is preferable to present of data with bars decreasing sequentially in length than circles decreasing sequentially in area. Further, viewers are less accurate in judging angles of objects with small radii (Cleveland & McGill, 1987). For this reason, bar graphs may be preferable to pie charts to display proportional data, though other researchers have found no difference in accuracy of judgements between bar graphs and pie charts (Spence & Lewandowsky, 1991; Simkin & Hastie, 1987). In addition, when plotting data points in a scatterplot, it is important to avoid overplotting, so viewers can make accurate conclusions about the layout of points (Wickham, 2010). A solution would be to saturate the colour of points according to density of the distribution (Unwin, Hawkins, Hofmann & Siegl, 1996). Further, it has also been shown that colour is an effective property for visual discrimination of scatterplot data (Lewandowsky & Spence, 1989). Organizing the presentation of categorical data according to a meaningful property such as a measure of central tendency is preferable to an arbitrary characteristic such as alphabetical order (Becker, Cleveland & Shyu, 1996; Friendly & Kwan, 2003). Overall, the literature on data visualization has shown that visual properties, such as point distribution and length discussed above, impact the perception and interpretation of data.

Although this is an emerging field, data visualization techniques have been used to investigate perceptions of social issues, specifically, climate change data in a high school student focus group in Sweden (Ballantyne, Wibeck & Neset, 2016). Considering the complexity of the social cause, the authors presented climate change information through salient images and videos in film in an immersive theatre environment. Through mind maps and interviews, the authors found that this presentation encouraged reflection about the issue among the focus group, and that students' prior knowledge about climate change impacted the interpretation of the data presented (Ballantyne, Wibeck & Neset, 2016). In addition, this study demonstrated the importance of engaging the public in discovering opinions about social issues after presenting information in visual and audiovisual forms. Lastly, the study revealed that focus groups can be a useful tool in investigating the interpretation of data presented visually and through different modalities.

Current Study

A purpose of visually displaying quantitative information is to effectively tell a story (Gelman & Unwin, 2013). Presenting data regarding a social issue involves telling the stories of communities and individuals dealing with difficult circumstances, of which often others are not aware. There is an important opportunity in which to engage with the public and communicate quantitative information visually, thereby telling the stories of often marginalized individuals. It is possible to implement best practices of data visualization found in the literature for the purposes of communicating the emotional stories presented by social organizations. However, to

our knowledge, no study has investigated which aspects of visualizations are most effective in drawing attention toward and communicating the data found by a social organization from the perspective of a wide range of audience members. We propose to systematically investigate visual presentations of quantitative data in improving awareness, knowledge, and willingness to get involved in social cause through thematic analyses of a focus group in order to capture the complexity of the human experience when perceiving, processing, and forming an emotional connection to a cause. We will manipulate the method of data visualization by presenting information in uniform text or through various infographics (i.e., tables, bar graphs, pie charts, and heat maps) to investigate credibility, memorability, and general importance of the particular social issue to the viewer. Our study will provide valuable information for various organizations seeking to optimize the impact of their data visualizations, thereby maximizing opportunities for raising awareness and inspiring participation in the social cause.

Methods

Participants

Thirty participants will be recruited through online advertisements across Canada in order to recruit participants from a wide range of demographics (i.e., age, gender, socioeconomic status, education, ethnicity etc.). Participants will be screened for inclusion criteria through an online survey: English speaking, access to a computer and internet connection, and general availability of a 2-hour time slot. Informed consent will also be obtained through an online questionnaire as per Ethical Guidelines from the York University Research Ethics Board.

Procedure

Based on participants' availability, groups of 4-5 will be created. Research on focus groups demonstrates the benefits of the methodology for understanding the process of how participants perceive, learn, and make meaning of new information (Wibeck et al., 2007). Focus groups were also utilized as a part of a semiotic framework to explore the effectiveness of Information and Communication Technologies (ICT) of communicating the complexities of climate changes through data visualization (Ballantyne et al., 2016). Based on this work, we will be conducting focus groups of semi-structured interviews led by a moderator who will provide structure to the sessions while simultaneously inviting open-ended discussion and follow-up to key ideas. As Ballantyne et al. (2016) discovered that preconceptions of climate change influenced participant interpretation of the data visualization, we decided to choose the less popularized social cause of food deserts in North America.

Each focus group will be conducted on a video conference platform in consideration of the COVID-19 pandemic and the wide adoptability of this platform in research since. Issues related to data collection through virtual means were carefully considered. In order to mitigate scheduling and communication difficulties, participants will be asked for their preferred method of communication (e.g., email, text message, phone) (Tuttas, 2015). The moderator will verbally ask permission from all participants prior to video recording the session. Lastly, the researchers will test the technology prior to the session to avoid any technological difficulties that could lead to delays and subsequently to participant attrition (Tuttas, 2015).

The focus group will begin by gauging any preconceptions or prior knowledge related to food deserts. Next, the moderator will present factual information on food deserts through text (e.g., definitions, statistics) and lead a discussion related to four key areas of in-the-moment reactions: 1) thoughts and cognitions, 2) emotions, 3) personal relevance and meaning, and 4) knowledge of communicated information. Based on the literature of principles of graphical excellence (Gordon & Finch, 2015), and effectiveness of various data visualization elements such as colour (Zeileis et al., 2020), pictographs (Haroz et al., 2015), infographics (Lazard & Atkinson, 2015), accessibility (Ferres et al., 2013), eliciting emotions (Kennedy, & Hill, 2018; Kostelnick, 2016) and interactivity (Young et al., 2018), a variety of graphs, infographics, maps, and interactive images will be presented to the focus group to investigate which elements are particularly effective in communicating the social causes' message and building a sense of concern for the cause. These materials will be gathered from various sources including social media, non-profit organizations, educational databases, and created by the authors in collaboration with a data visualization expert in the field. The moderator will engage the group in a series of questions related to the aforementioned four key areas after the presentations of each new form of data visualization. Each session will be videorecorded and will be approximately 2 hours in length with a 10-minute break at the midway point.

Data Analysis

Each session's audio recording will be transcribed, and videos will be coded for participant reactions to stimuli. A thematic analysis will be conducted according to guidelines outlined by Braun and Clarke (2006). Namely, this will involve data familiarisation, initial coding of information, theme searching, and theme reviewing and defining (Braun & Clarke, 2006). The free and accessible software, *Taguette*, will be used for data familiarisation, coding, and theme organization.

Lastly, it is important to discuss author positionality when conducting a thematic analysis (Braun & Clarke, 2019). Authors A.A. and I.Y. are both doctoral students in psychology at York University with areas of research focus in visual perception and clinical neuropsychology, respectively. The project will be completed in collaboration with an expert within the field of data visualization, Dr. Michael Friendly, a faculty member in the Psychology Department at York University (see here for publication history within the area: http://www.datavis.ca/papers/vita/). All research members do not have an invested interest or stake within any organization related to food deserts.

Conclusion

Visualizing data creatively and effectively is a valuable tool to communicate data to various audiences. The literature on data visualization has demonstrated certain qualities of a graphic that impact accuracy in judging the information presented. Using these techniques can be a valuable opportunity to engage the public and inform others about a social issue. The aim of this proposal is to investigate the visualizations which are most effective in drawing attention toward and communicating the data found by a social organization from the perspective of a viewer from a wide range of demographics. We will assess these perspectives using focus groups, and we will use thematic analyses to gain insight into the processes of emotions and

cognitions elicited when viewing information of a social cause. This way, we aim to elucidate the optimal data visualization methods for use by social organizations to promote awareness of their cause, with the ultimate goal of making a difference in the world.

References

Avidon, E. (2020). *'Data for good' movement spurs action in fight for causes*. SearchBusinessAnalytics. Retrieved 20 June 2021, from https://searchbusinessanalytics.techtarget.com/news/252487703/Data-for-good-movement-spurs-action-in-fight-for-causes.

Ballantyne, A. G., Wibeck, V., & Neset, T. (2016). Images of climate change – a pilot study of young people's perceptions of ICT-based climate visualization. *Climatic Change, 134*(1), 73-85. https://doi.org/10.1007/s10584-015-1533-9

Becker, R. A., Cleveland, W. S., & Shyu, M.-J. (1996). The Visual Design and Control of Trellis Display. Journal of Computational and Graphical Statistics, 5(2), 123–155. https://doi.org/10.1080/10618600.1996.10474701

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., Clarke, V. (2019). Novel insights into patients' life-worlds: The value of qualitative research. *The Lancet Psychiatry, 6*(9), 720–72. https://doi.org/10.1016/S2215-0366(19)30296-2

Cleveland, W. S. (1985). The elements of graphing data . Wadsworth Advanced Books and Software.

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Cleveland, W. S., & McGill, R. (1987). Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. Journal of the Royal Statistical Society. Series A. General, 150(3), 192–229. https://doi.org/10.2307/2981473

Dash, B., & Gabor. (2021, May 25). *30% – 60% of girls and women aged 15-49 still go through FGM in Tanzania's rural communities*. Crowd2Map Tanzania. https://crowd2map.org/.

Ferres, L., Lindgaard, G., Sumegi, L., & Tsuji, B. (2013). Evaluating a tool for improving accessibility to charts and graphs. *ACM Transactions on Computer-Human Interaction, 20*(5), 1-32. https://doi.org/10.1145/2533682.2533683

Friendly, M., & Kwan, E. (2003). Effect ordering for data displays. Computational Statistics & Data Analysis, 43(4), 509–539. https://doi.org/10.1016/S0167-9473(02)00290-6

Gelman, A., & Unwin, A. (2013). Infovis and Statistical Graphics: Different Goals, Different Looks. Journal of Computational and Graphical Statistics, 22(1), 2–28. https://doi.org/10.1080/10618600.2012.761137

Gordon, I., & Finch, S. (2015). Statistician heal thyself: Have we lost the plot? *Journal of Computational and Graphical Statistics, 24*(4), 1210-1229. https://doi.org/10.1080/10618600.2014.989324

Haroz, S., Kosara, R., & Franconeri, S. L. (2015). ISOTYPE Visualization: Working Memory, Performance, and Engagement with Pictographs. *Association for Computing Machinery 1191-1200*. https://doi.org/10.1145/2702123.2702275

Kennedy, H., & Hill, R. L. (2018). The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology (Oxford), 52*(4), 830-848. https://doi.org/10.1177/0038038516674675

Kostelnick, C. (2016). The re-emergence of emotional appeals in interactive data visualization. *Technical Communication (Washington), 63*(2), 116.

Lazard, A., & Atkinson, L. (2015). Putting environmental infographics center stage: The role of visuals at the elaboration likelihood Model's critical point of persuasion. *Science Communication, 37*(1), 6-33. https://doi.org/10.1177/1075547014555997

Lewandowsky, S., & Spence, I. (1989). Discriminating Strata in Scatterplots. Journal of the American Statistical Association, 84(407), 682–688. https://doi.org/10.1080/01621459.1989.10478821

Simkin, D., & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. Journal of the American Statistical Association, 82(398), 454–465. https://doi.org/10.1080/01621459.1987.10478448

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. Applied Cognitive Psychology, 5(1), 61–77. https://doi.org/10.1002/acp.2350050106

Tuttas, C. A. (2015). Lessons learned using web conference technology for online focus group interviews. *Qualitative Health Research, 25*(1), 122-133. https://doi.org/10.1177/1049732314549602

Unwin, A., Hawkins, G., Hofmann, H., & Siegl, B. (1996). Interactive Graphics for Data Sets with Missing Values-MANET. Journal of Computational and Graphical Statistics, 5(2), 113–122. https://doi.org/10.1080/10618600.1996.10474700

Viz for Social Good. (2020). https://www.vizforsocialgood.com/.

Wibeck, V., Dahlgren, M. A., & Öberg, G. (2007). Learning in focus groups: An analytical dimension for enhancing focus group research. *Qualitative Research : QR, 7*(2), 249-267. https://doi.org/10.1177/1468794107076023

Wickham, H. (2010). A Layered Grammar of Graphics. Journal of Computational and Graphical Statistics, 19(1), 3–28. https://doi.org/10.1198/jcgs.2009.07098

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism?: A content analysis of data journalism awards finalists 2012-2015. *Journalism Practice, 12*(1), 115-135. https://doi.org/10.1080/17512786.2016.1270171

Zeileis, A., Fisher, J. C., Hornik, K., Ihaka, R., McWhite, C. D., Murrell, P., Stauffer, R., & Wilke, C. O. (2020). colorspace: A toolbox for manipulating and assessing colors and palettes. *Journal of Statistical Software, 96*(1), 1-49. https://doi.org/10.18637/jss.v096.i01