Open Questions about the Visualization of Sociodemographic Data

Florent Cabric [ID] *
LISN, Université
Paris-Saclay, CNRS,
Inria, Saclay, France

Margrét Vilborg Bjarnadóttir [ID] †
DOIT, Robert H. Smith
School of Business,
University of Maryland,
United States

Anne-Flore Cabouat [ID] ‡
LISN, Université
Paris-Saclay, CNRS,
Inria, Saclay, France

Petra Isenberg [ID] §
LISN, Université
Paris-Saclay, CNRS,
Inria, Saclay, France

ABSTRACT

This paper collects a set of open research questions on how to visualize sociodemographic data. Sociodemographic data is a common part of datasets related to people, including institutional censuses, health data systems, and human-resources files. This data is sensitive, and its collection, sharing, and analysis require careful consideration. For instance, the European Union, through the General Data Protection Regulation (GDPR), protects the collection and processing of any personal data, including sexual orientation, ethnicity, and religion. Data visualization of sociodemographic data can reinforce stereotypes, marginalize groups, and lead to biased decision-making. It is, therefore, critical that these visualizations are created based on good, equitable design principles. In this paper, we discuss and provide a set of open research questions around the visualization of sociodemographic data. Our work contributes to an ongoing reflection on representing data about people and highlights some important future research directions for the VIS community. A version of this paper and its figures are available online at osf.io/a2u9c.

Index Terms: Human-centered computing—Visualization;

1 INTRODUCTION

Data visualizations are commonly based on an abstraction of data according to some criteria, such as an average or a clustering. These abstractions enable us to make comparisons,3 for instance, between multiple groups or to identify outliers that do not match the characteristics of the group. When we visualize data about people, this kind of abstraction may lead to the promotion of stereotypes and biases. Stereotypes are built on grouping together people with presumed shared characteristics as a means to describe group membership, contrast that group with others, and identify outsiders [3, 34, 41]. Abstraction can also lead to invisibilizing marginalized groups and their issues [22, 37, 38]. Thus, the process of abstraction of data in visualizations can be very harmful — for example, by invisibilizing marginalized groups, leading to biased decisions, or conveying a feeling of not being included. The design of nonharmful visualization raises many design [38] and ethical [12] questions.

Our work is in line with previous work on how to visualize data about people [12, 14, 16, 38]. In particular, we focus on expanding on the "Do No Harm Guide" [38] that discusses a set of considerations for designing visualizations considering diversity, equity, and inclusion. We elaborate on this work through a discussion of open research questions. Our work is a reflection on the challenges that researchers in the community need to tackle to effectively and equitably visualize sociodemographic data in a nonharmful way and an opportunity to discuss our current practices.

---

*e-mail: florent.cabric@inria.fr
†e-mail: mbjarnad@umd.edu
‡e-mail: anne-flore.cabouat@inria.fr
§e-mail: petra.isenberg@inria.fr

2 PROTECTION AND VISUALIZATION OF SOCIODEMOGRAPHIC DATA

The protection of private attributes has a long history but has seen rapid changes in recent decades. The number of policies regarding the protection of sociodemographic data is growing across companies, states, and institutions. One such policy, the EU's General Data Protection Regulation (GDPR) [20], strongly regulates the processing of five categories of sensitive data: *personal, trade-union membership, genetic or biometric, health-related,* and *sex life or sexual orientation.* The UK [44] defines nine attributes as protected: *age, disability, sex, sexual orientation, gender reassignment, marriage and civil partnerships, pregnancy/maternity, race,* and *religion/belief.* Discrimination against people based on protected attributes is forbidden by law. Similarly, Canada differentiates 16 characteristics about people [11] that require protection against discrimination: *race, national or ethnic origin, color, religion, age, sex, sexual orientation, gender identity and expression, marital status, family status, genetic characteristics, disability, creed, irrational fears* and *source of income*; and the US defines 11 [45]: *race, religion, national origin, age, sex, sexual orientation, pregnancy, family status, disability status, veteran status,* and *genetic information.*

Several papers have discussed broader issues around the visualization of data about people. The belief that data visualizations might be neutral has been disparaged by many: Visualizations are political [16], can promote empathy [7], often do not surface inherent uncertainty [24], and have even been called inhuman [17]. Beyond the visualization, data and its collection and processing are also regularly prone to biases. Criado Perez [13] has summarized how unrepresentative data collection about people (for example, only collecting data from men) has led to overgeneralizations to the general population and thus has caused discrimination and other problems such as discomfort in women's daily lives [13, Chapter 8] or medical complications [13, Chapter 10].

Out of the above-mentioned personal attributes, a few have been the subject of dedicated papers on how to visualize them in a nonharmful way. Disability has gained attention from visualization researchers, with most work focused on making data visualization accessible to people with disabilities [18,19,30,47]. Only a few studies have explored how to represent data about people with disabilities. For example, Barstow and colleagues [1] investigated the design of inclusive visual symbols beyond the "traditional" wheelchair icon, which fails to represent many people with a physical disability who need specific access or consideration. Race or ethnicity has received attention regarding the design of nonharmful visualization. Schwabish and Feng [38] discussed a set of considerations for data collection, use of language, visibility of data, and use of colors, icons, and shapes. The authors emphasize talking to people to ensure they are represented correctly and harmlessly in data visualizations. Dhawka and colleagues [14] explored the challenges and opportunities of representing diverse people grouped by race using anthropographics. Gender has also been the subject of studies related to the diversity of the community itself [43]. Papers on the visual representation of gender as a demographic variable are rare. In a blog post, Muth [33] collected multiple examples of professional visualizations of gender by institutions such as Bloomberg

and the New York Times. The visualizations used a variety of colors, including pink/blue, but also many other combinations, such as orange/green or green/purple, without a clear consistency. Finding the right visual variables, layouts, and marks has been shown to be challenging for these sociodemographic attributes (disability, ethnicity, and gender), and many others may surface as more of the protected attributes are explored. However, multiple personal attributes raise common open research questions that will be discussed next.

3 Some Open Questions about the Visualization of Sociodemographic Data

As pointed out in the introduction, visualizations are often inherently built on a process of abstraction. When data about people have been abstracted into groups, these groups need to be assigned marks, visual variables, and a layout. This visualization design process involves choices that may reflect a certain way of thinking or "not thinking" [13, Preface], built-in default encodings, or the use of over-general guidelines. Visualization challenges in this process are related to the potential amplification of stereotypes and the marginalization of groups of people. In addition, designers have the difficult task of balancing how the visualized populations will see themselves reflected in the visualization and the ability of decision-makers to use the visualization effectively. We next outline open questions we would like to discuss as part of the Vis4Good workshop.

3.1 Balancing Efficiency

Visual decision-making tasks are cognitively complex: The reader must understand the visualization, define the underlying decision problem, compare different options, and then make one choice [15]. Such complexity may require balancing inclusive visual design with the need for time-efficient task completion. Air traffic control, crowd management, or emergency medicine regularly require decision-makers to complete critical tasks quickly and accurately. To reduce the time taken to make a decision, people can rely on stereotypes to activate heuristics [5, 6], sometimes at the cost of accuracy [36]. This Speed-Accuracy Trade-off (SAT) [26] has been discussed for a long time in psychology [21]. In the context of visualization, we are interested in discussing the impact of stereotyped visual representations on the decision-making processes.

Creating visualization with stereotyped attributes such as red and blue colors for gender may speed up cognitive tasks and decisions.
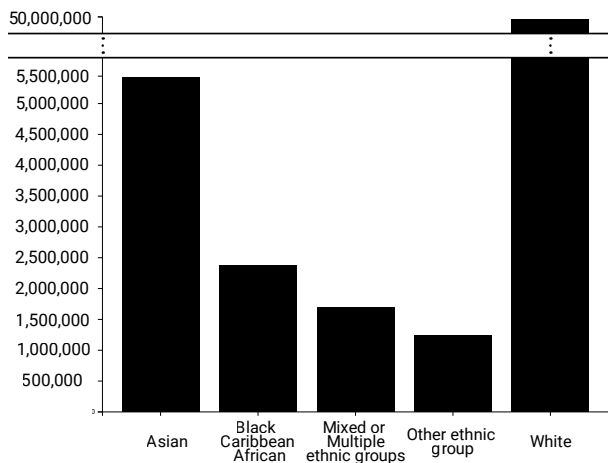

Figure 1: England's and Wales's residents according to five different ethnic groups. Contrast this figure with Fig. 2 and consider the balance between information and visualization simplicity and how included different types of viewers may feel. The data is from the UK census and uses its data labels [35]. *Note: the y-axis is broken between the 5.5M and 50M ticks.*

Studies have assessed the effectiveness of using semantically resonant colors [28]. For example, using a color associated with the data represented, such as yellow for a bar representing bananas. In contrast, there is some evidence that using a reversed mapping — associating a visual mapping that is typically associated with a specific group (brighter colors = women) to another group (brighter colors = men) — slows down classification and association tasks [39].

We are not aware of studies that evaluated the efficiency of stereotyped and unstereotyped visualizations for common tasks [8]. Suppose previous results on semantically resonant colors are confirmed with stereotyped visualizations. In that case, a new dilemma will arise for designers of visualizations used in situations that require quick and accurate task completion: Is the efficiency of my visualizations worth the use of stereotyped data encodings? One may argue that this question is not even worth asking because the well-being of the people represented should trump concerns about efficiency. Yet, people's well-being might also depend on quick and correct decisions. What has not been studied to date is: To what extent do the potential benefits of stereotyped visualizations matter? Perhaps the efficiency benefits are so small that we should never consider using stereotyped visualizations, and research should focus on raising awareness of stereotypes. Or, perhaps the benefits are significant enough to motivate future works on how to avoid stereotyped visualizations while still achieving the same efficiency. For example, training people to associate groups with non-stereotyped visual variables such as colors may solve the problem; but how and to what extent to invest in such training in different scenarios is an open question. To be clear, we do not call for the use of stereotyped visual encodings. Instead, we call for more research to fully understand the effects of using undesired stereotyped visualizations.

3.2 Balancing Simplicity

In their "Do No Harm Guide," Schwabish and Feng [38] recommend that "data visualizations should use more complex designs if that would more accurately reflect and promote a better understanding of the topic being shown." In addition, Baumer and colleagues [2] argue that complexity can lead to inclusivity and help data visualization readers to gain deeper insights. We relate complexity here to the number of data categories being displayed. To facilitate data exploration, designers must disaggregate and split specific data dimensions into different factors and subfactors. For instance, the 2021 UK Census dataset [35] proposed three levels of disaggregation for ethnicity data: 5 (see Fig. 1), 20 (see Fig. 2), and 288 categories.
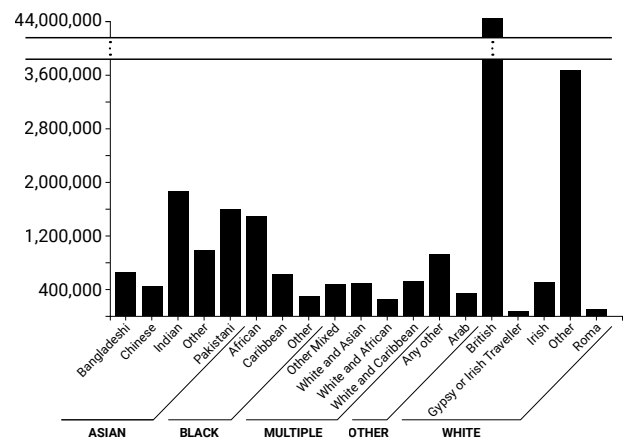

Figure 2: England's and Wales's residents grouped into 20 different ethnic groups. Contrast this figure with Fig. 1 and consider the balance between information and visualization simplicity and how included different types of viewers may feel. The data is from the UK census and uses its data labels [35]. *Note: the y-axis is broken between the 3.6M and 40M ticks.*

On the one hand, the representation of a small number of groups can result in the invisibilization of small and sometimes marginalized groups. Invisibilizing groups can reinforce the feeling of being outside society for the people represented [22]. On the other hand, *hypervisibilizing* groups [40] can be interpreted as *tokenization* of the people represented [46] and lead to increased surveillance from the main group. Overrepresentation in the data or visualization can reinforce feelings of marginalization.

The continuum of disaggregation is vast, and visualization designers must choose between simplicity, which can lead to overgeneralization and invisibilization, or better representativity of the population, which can result in hypervisibilization and visual clutter. Minimizing the number of categories can considerably reduce the number of items to process and, thus, facilitate data exploration and decision-making. Representing the 288 ethnic groups of the UK census [35] can give a sense of representativeness of the population but make the visualization less time-efficient for tasks that require a lot of attention and carefulness. This trade-off between simplicity and better representation raises an important question: To what extent can better representation of people from marginalized groups reduce efficiency or lead to biased decisions that may harm the people represented?

To be clear again, we are not advocating the invisibilization of groups in data visualizations. On the contrary, we want visualization researchers to invest time in understanding the effects of this trade-off on tasks as complex as decision-making. For example, it would be interesting to study tools that help decision-makers explore different disaggregation levels while remaining simple. One notable example is VisPilot [27], which supports the exploration of subgroups while avoiding drill-down fallacies that lead to incomplete insights. It should be noted that many visualizations, in particular those designed for the general public, do not require the same type of efficiency as discussed above. Instead, designers should focus on other objectives, such as the representativity of the people visualized, understandability, or message reception.

Few previous works in the visualization community have discussed the effect of grouping people [2, 38]. Future studies can take inspiration from other research domains, such as the social sciences [22, 40], to develop new methods to investigate how people feel harmed or represented when looking at visualizations under different levels of aggregation.

3.3 Inclusiveness of Different Representation Types

Some visualization researchers have raised the question of how best to represent the individuals behind sociodemographic data. Recently, researchers have studied visual representations to show the diversity [14] or disparity [23] of people inside a group. Morais and colleagues defined three levels of *granularity* to represent individuals [32]:

Low: The marks aggregate all people from the same group to hide variability (see Fig. 1 and Fig. 2).
Intermediate: The marks visualized represent a fixed number of people (greater than 1; see Fig. 3)
Maximum: Each mark is associated with one individual.

This continuum poses interesting challenges for the visualization community. First, it is unclear how people represented in the data are affected by specific types of visualization or how they prefer to be represented. Second, it is unclear how a specific representation affects the stakeholders' understanding of the underlying data.

Holder and Xiong [23], for example, showed that the type of representation mattered. The authors compared visualizations of low granularity (bar charts, dot plots) with those of higher levels of granularity (jitter plots, prediction intervals). The authors found low granularity visualizations increased stereotyped conclusions.

The representation of intermediate or maximum levels of granularity seems promising. However, visualizing individuals has multiple
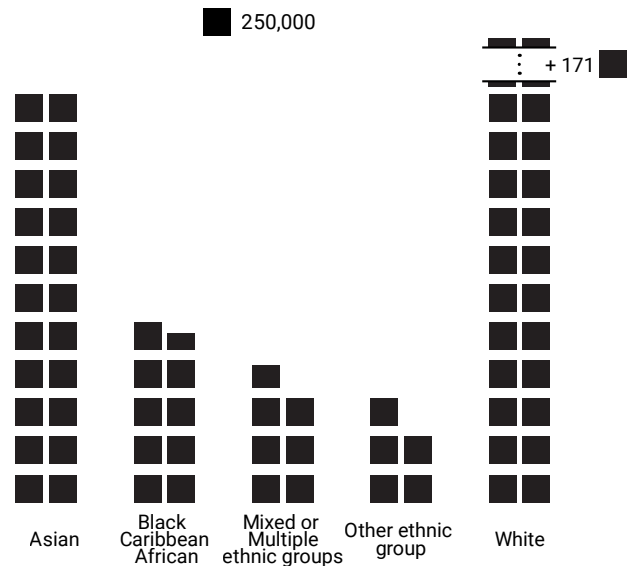

Figure 3: Unit-based visualization from the UK census [35], where each full square represents 250,000 UK residents by ethnic group. While a unit-based visualization may be beneficial to show that each group is made up of individuals, our choice of unit size doesn't allow full representation of the White group.

challenges, including the scalability of maximum granularity to a large dataset [32], and finding an adequate unit size, in particular when representing groups with large differences in population size. In Fig. 3, we decided on a unit size that allowed a more refined representation of smaller groups but made it difficult to fully represent the largest group within a suitable layout. Perhaps early work on focus+context visualization should be revisited in this context [42].

A self-perception of the individual in data visualization can also be achieved with *anthropographics* — data visualizations that use human-shaped graphical representation. Researchers studied such visualizations to elicit more empathy for human rights [7] and promote prosocial behavior [31]. Despite initial results that did not show clear benefits of anthropographics, this stream of research is growing. Notably, Dhawka and colleagues [14] recently investigated the challenges of representing the diversity of people's races. In particular, they identified the mapping between sociodemographic groups and visual channels as one that can lead to problematic choices, depending on whether or not the designer is aware of the social construction of groups and the stereotypes that can exist between the group and the visual channel. In the context of Virtual Reality, Ivanov and colleagues [25] investigated individual representations of 3D human figures who died in mass shootings. The authors rendered human figures from six unique models as flat-grey silhouettes according to age and gender. Anthropographics, as a new paradigm in visualization , requires further studies to understand its power, in particular, to promote diversity, equity, and inclusion.

Anthropographics per se requires grouping and visualizing people based on a physical characteristic shared by an entire group. In the "Do No Harm Guide" [38], Schwabish and Xiong advised choosing unstereotyped icons to represent people. However, grouping people based on a physical characteristic requires that the entire group shares the physical characteristic and that this characteristic is not a topic of discrimination. Anthropographics raises many questions, including: To what extent is it even possible to represent people based on a physical characteristic that is not stereotyped? Also, to what extent does showing certain types of human-shaped icons lead to feelings of inclusion or exclusion?

3.4 Summary

As mentioned in Section 2, visualizations are not neutral. They can lead to biased decisions, the perpetuation of stereotypes, the reinforcement of discrimination, or the amplification of groups' marginalization. Visualizing demographic data has the potential to cause great harm, thus requiring careful and responsible design choices. Visualization designers can select what data they want to show, the visual marks mapped to each category, and what extra information to disclose. Such decisions place the designer in charge of how the visualization can be used or interpreted [12] and, consequently, influencing the people visually represented. Designers must be aware that any design choice involves trade-offs between reducing potential harm and, for example, time efficiency, visual simplicity, or efforts to elicit empathy.

We know little about the potential side effects of designing non-harmful visualizations. The set of open questions we raised focused on the trade-off between the use of visual stereotypes and efficiency, the trade-off between the inclusiveness of the marginalized group represented and simplicity, and the right level of individuation and perception to convey a sense of inclusion. But many other important questions remain open, including the efficiency of anthropographics or the scalability of unit-based visualizations.

4 DISCUSSION

At the beginning of this paper, we introduced protected attributes as important data dimensions to focus on in order to accurately represent individuals and groups of people and their characteristics. Overall, there is very limited visualization research on this topic, and until now, it has been mainly focused on three of these attributes: gender [33], disability [1], and ethnicity/race [14]. We have not come across any dedicated discussion on the representation of religion, age, marital status, or any of the other attributes. The question is, why? Are some of these attributes less associated with visual stereotypes than others? For example, stereotyped colors come to mind immediately when we think of gender; conversely, fewer and weaker visual stereotypes may come to mind when considering age.

As researchers from both the US and Europe, we also are aware of the continental and national differences in calls for social changes broadly linked to diversity, equity, and inclusion. For example, a person's veteran status is a protected attribute in the US, while it is not in the UK or Canada. As such, studies on how visualizations of sociodemographic data are perceived should be a multi-cultural effort to study how and if perceptions are local.

Including people from marginalized populations throughout the design process is an important step in mindful representations of such groups — as suggested in recent works on demographic and nonharmful visualization [1, 2, 14, 16, 38]. For example, Mack and colleagues [29] interviewed 18 people with disabilities or related identities across two months to understand their perception, design choices, and use of avatars on online platforms. Respondents expressed the need to modify the avatar's disclosure of their disability according to context or environment. The choice of disclosing (or not) or aggregating (or not) personal data or attributes in visualizations should be made in consultation with the persons represented.

Visualization designers should also consider disclosing design rationales and information about the people involved in the design process as an integral part of a visualization [10]. Dörk and colleagues [16] highlight the importance of empowering the reader by providing background information about the data shown, the data hidden, and the intention of the visualization designer. Recently, Burns and colleagues [9] showed that disclosing metadata, such as design rationales (for instance, why the designer mapped a specific visual variable to a category) or data sources (where the data comes from) can improve transparency and viewer's understanding.

Many topics related to sociodemographic data visualization require a cross-disciplinary approach. For instance, inclusiveness of

data collection requires collaboration between scientists and institutions that collect data. Visualization researchers also have a role to play in designing innovative data collection methods. For example, Beischel and colleagues [4] designed a visual tool to let participants express their sex and gender based on multiple dimensions.

5 CONCLUSION

Visualizing sociodemographic data requires designers to address many design challenges, from data collection to visualization. The design of visualization is not neutral and can convey different messages with the same underlying data. Sometimes, these messages can harm people, in particular, people from marginalized groups, and we need to understand these potential harms better. Through a set of open questions, this paper aims to advance the discussion on how to design the least harmful visualization.

Many of the questions we raise represent current considerations in the visualization community and in society. It should be noted that all the constructs that can harm people (invisibilization, hypervisibilization, stereotypes in representation, bias in decision-making) are socially and contextually dependent. Therefore, inclusive visualization practices may always be a moving target. When we base these visualizations on key principles, progress can be achieved, a point made well in the "Do No Harm Guide" [38]. We have an obligation to be careful and inclusive, but, more critically, additional research is needed to understand better how our design choices affect both the people represented and the people using our visualizations.

REFERENCES

[1] B. A. Barstow, J. Vice, S. Bowman, T. Mehta, S. Kringen, P. Axelson, and S. Padalabalanarayanan. Examining perceptions of existing and newly created accessibility symbols. *Disabil. Health J.*, 12(2):180–186, Apr. 2019. doi: 10.1016/j.dhjo.2018.11.012

[2] E. P. S. Baumer, M. Jasim, A. Sarvghad, and N. Mahyar. Of course it's political! a critical inquiry into underemphasized dimensions in civic text visualization. *Comput. Graph. Forum*, 41(3):1–14, June 2022. doi: 10.1111/cgf.14518

[3] H. S. Becker. *Outsiders: Studies in the Sociology of Deviance*. Free Press, New York, 1963.

[4] W. J. Beischel, Z. C. Schudson, and S. M. van Anders. Visualizing gender/sex diversity via sexual configurations theory. *Psychol. Sex. Orientat. Gend. Divers.*, 8(1):1–13, 2021. doi: 10.1037/sgd0000449

[5] G. V. Bodenhausen. Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychol. Sci.*, 1(5):319–322, Sept. 1990.

[6] G. V. Bodenhausen and R. S. Wyer Jr. Effects of stereotypes on decision making and information-processing strategies. *J. pers. soc. psychol.*, 48(2):267–282, Feb. 1985. doi: 10.1037/0022-3514.48.2.267

[7] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, pp. 5462–5474. ACM, 2017. doi: 10.1145/3025453.3025512

[8] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Trans. Vis. Comput. Graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124

[9] A. Burns, C. Lee, T. On, C. Xiong, E. Peck, and N. Mahyar. From invisible to visible: Impacts of metadata in communicative data visualization. *IEEE Trans. Vis. Comput. Graphics*. To appear. doi: 10.1109/TVCG.2022.3231716

[10] A. Burns, T. On, C. Lee, R. Shapiro, C. Xiong, and N. Mahyar. Making the invisible visible: Risks and benefits of disclosing metadata in visualization. In *IEEE Workshop Vis. Soc. Good (VIS4Good)*, pp. 11–15, 2021. doi: 10.1109/VIS4Good54225.2021.00008

[11] Canada. Canadian Human Rights Act R.S.C., 1985, c. h-6. `https://laws-lois.justice.gc.ca/eng/acts/H-6/`, 2023.

[12] M. Correll. Ethical dimensions of visualization research. In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, pp. 2019:1—-2019:13. ACM, 2019. doi: 10.1145/3290605.3300418

[13] C. Criado Perez. *Invisible Women: Data Bias in a World Designed for Men*. Abrams Press, New York, 2019. doi: 10.1145/3290605.3300418

[14] P. Dhawka, H. A. He, and W. Willett. We are the data: Challenges and opportunities for creating demographically diverse anthropographics. In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, pp. 807:1–807:14. ACM, 2023. doi: 10.1145/3544548.3581086

[15] E. Dimara and J. Stasko. A critical reflection on visualization research: Where do decision making tasks hide? *IEEE Trans. Vis. Comput. Graphics*, 28(1):1128–1138, 2022. doi: 10.1109/TVCG.2021.3114813

[16] M. Dörk, P. Feng, C. Collins, and S. Carpendale. Critical infovis: Exploring the politics of visualization. In *Extended Abstracts on Human Factors in Computing Systems (CHI)*, pp. 2189—2198. ACM, 2013. doi: 10.1145/2468356.2468739

[17] S. Dragga and D. Voss. Cruel pies: The inhumanity of technical illustrations. *Technical Communication*, 48:265–274, Aug. 2001.

[18] F. Elavsky, C. Bennett, and D. Moritz. How accessible is my visualization? evaluating visualization accessibility with chartability. *Comput. Graph. Forum*, 41(3):57–70, 2022. doi: 10.1111/cgf.14522

[19] N. Elmqvist. Visualization for the blind. *Interactions*, 30(1):52—-56, Jan. 2023. doi: 10.1145/3571737

[20] European Union. Article 9 – regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Off. J. Eur. Union*, (L 119):38, 2016.

[21] G. Gigerenzer and W. Gaissmaier. Heuristic decision making. *Annu. Rev. Psychol.*, 62(1):451–482, 2011. doi: 10.1146/annurev-psych-120709-145346

[22] B. Herzog. Invisibilization and silencing as an ethical and sociological challenge. *Soc. Epistemol.*, 32(1):13–23, 2018. doi: 10.1080/02691728.2017.1383529

[23] E. Holder and C. Xiong. Dispersion vs disparity: Hiding variability can encourage stereotyping when visualizing social outcomes. *IEEE Trans. Vis. Comput. Graphics*, 29(1):624–634, Jan. 2023. doi: 10.1109/TVCG.2022.3209377

[24] J. Hullman. Why authors don't visualize uncertainty. *IEEE Trans. Vis. Comput. Graphics*, 26(1):130–139, Jan. 2020. doi: 10.1109/TVCG.2019.2934287

[25] A. Ivanov, K. Danyluk, C. Jacob, and W. Willett. A walk among the data. *IEEE Comput. Graph. Appl.*, 39(3):19–28, 2019. doi: 10.1109/MCG.2019.2898941

[26] J. S. Larson and G. E. Hawkins. Speed-accuracy tradeoffs in decision making: Perception shifts and goal activation bias decision thresholds. *J. Exp. Psychol.: Learn. Mem. Cogn.*, 49(1):1–32, Jan. 2023. doi: 10.1037/xlm0000913

[27] D. J.-L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. Parameswaran. Avoiding drill-down fallacies with VisPilot: Assisted exploration of data subsets. In *Proc. Int. Conf. Intell. User Interfaces (IUI)*, pp. 186–196. ACM, 2019. doi: 10.1145/3301275.3302307

[28] S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer. Selecting semantically-resonant colors for data visualization. *Comput. Graph. Forum*, 32(3pt4):401–410, 2013. doi: 10.1111/cgf.12127

[29] K. Mack, R. C. L. Hsu, A. Monroy-Hernández, B. A. Smith, and F. Liu. Towards inclusive avatars: Disability representation in avatar platforms. In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, pp. 607:1–607:13. ACM, 2023. doi: 10.1145/3544548.3581481

[30] K. Marriott, B. Lee, M. Butler, E. Cutrell, K. Ellis, C. Goncu, M. Hearst, K. McCoy, and D. A. Szafir. Inclusive data visualization for people with disabilities: A call to action. *Interactions*, 28(3):47—-51, May 2021. doi: 10.1145/3457875

[31] L. Morais, Y. Jansen, N. Andrade, and P. Dragicevic. Can anthropographics promote prosociality? A review and large-sample study. In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*, pp. 611:1–611:18. ACM, 2021. doi: 10.1145/3411764.3445637

[32] L. Morais, Y. Jansen, N. Andrade, and P. Dragicevic. Showing data about people: A design space of anthropographics. *IEEE Trans. Vis. Comput. Graphics*, 28(3):1661–1679, 2022. doi: 10.1109/TVCG.2020.3023013

[33] L. C. Muth. An alternative to pink & blue: Colors for gender data. `https://blog.datawrapper.de/gendercolor/`, Mar. 2018. Accessed: July 2023.

[34] T. N. Nelson. *Handbook of prejudice, stereotyping, and discrimination*. Psychology Press, New York, NY, 2009. doi: 10.4324/9781841697772

[35] Office for National Statistics - UK. Census 2021 — ethnic group. `https://www.ons.gov.uk/datasets/TS021/editions/2021/versions/3`, Nov. 2022. Accessed: July 2023.

[36] J. W. Payne, J. R. Bettman, and E. J. Johnson. Adaptive strategy selection in decision making. *J. Exp. Psychol.: Learn. Mem. Cogn.*, 14(3):534–552, July 1988. doi: 10.1037/0278-7393.14.3.534

[37] K. Rall, M. L. Satterthwaite, A. V. Pandey, J. Emerson, J. Boy, O. Nov, and E. Bertini. Data visualization for human rights advocacy. *J. Hum. Rights Pract.*, 8(2):171–197, jul 2016. doi: 10.1093/jhuman/huw011

[38] J. Schwabish and A. Feng. Do No Harm Guide: Applying equity awareness in data visualization. Published Online `https://www.urban.org/research/publication/do-no-harm-guide-applying-equity-awareness-data-visualization`, 2021.

[39] G. R. Semin and T. A. Palma. Why the bride wears white: Grounding gender with brightness. *J. Consum. Psychol.*, 24(2):217–225, Apr. 2014. doi: 10.1016/j.jcps.2013.09.003

[40] I. H. Settles, N. T. Buchanan, and K. Dotson. Scrutinized but not recognized: (In)visibility and hypervisibility experiences of faculty of color. *J. Vocat. Behav.*, 113:62–74, Aug. 2019. Managing visibility and invisibility in the workplace. doi: 10.1016/j.jvb.2018.06.003

[41] B. A. Susie Heywood. *Challenging Gender Stereotypes in the Early Years*. Routledge, London, 2022. doi: 10.4324/9781841697772

[42] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann. Interactive lenses for visualization: An extended survey. *Comput. Graph. Forum*, 36(6):173–200, Sept. 2017. doi: 10.1111/cgf.12871

[43] N. Tovanich, P. Dragicevic, and P. Isenberg. Gender in 30 years of IEEE visualization. *IEEE Trans. Vis. Comput. Graphics*, 28(1):497–507, Jan. 2022. doi: 10.1109/TVCG.2021.3114787

[44] UK. Types of discrimination ('protected characteristics'). Published Online `https://www.gov.uk/discrimination-your-rights`, 2010. Last Accessed: 06-20-2023.

[45] USA. Types of discrimination ('protected characteristics'). `https://www.ftc.gov/policy-notices/no-fear-act/protections-against-discrimination`. Last Accessed: 06-20-2023.

[46] A. H. Wingfield and J. H. Wingfield. When visibility hurts and helps: How intersections of race and gender shape black professional men's experiences with tokenization. *Cult. Divers. Ethn. Minor. Psychol.*, 20(4):483–490, Oct. 2014. doi: 10.1037/a0035761

[47] K. Wu, E. Petersen, T. Ahmad, D. Burlinson, S. Tanis, and D. A. Szafir. Understanding data accessibility for people with intellectual and developmental disabilities. In *Proc. Conf. Hum. Factors Comput. Syst. (CHI)*. ACM, 2021. doi: 10.1145/3411764.3445743