Applying Racial Equity Awareness in Data Visualization

**Keywords:** Diversity, equity, inclusion, racial equity, accessibility.

**Abstract:** A data visualization style guide does for graphs what the *Chicago Manual of Style* does for English grammar: it defines the components of a graph and their proper, consistent use. In the process of revising and expanding the style guide for the Urban Institute, a nonprofit research institution based in Washington, DC, we are taking a more diverse, equitable, and inclusive (DEI) perspective to our research, data, and visualizations. Here, we discuss eight techniques that data visualization producers should consider when creating visuals. We consider this a first step in our process of creating graphs through a DEI lens.

# Introduction

The Urban Institute is a nonprofit research institution based in Washington, DC. With more than 500 researchers, staff members, and communications experts, Urban’s mission is to open minds, shape decisions, and offer solutions. To advance that mission, we must think intentionally about how we can learn from and speak to audiences that reflect the rich diversity of America’s communities, and we must foster a culture where employees from different backgrounds and perspectives enjoy mutual respect, inclusivity, and collegiality.

For the past few months, we have been working on revising and updating Urban’s Data Visualization Style Guide [1]. In addition to reorganizing the guide and updating various style decisions, we have also been thinking critically about how we communicate data and information about the groups we study. Just as Urban has carefully considered the words we use in our written reports and platforms, we should be equally careful in how we visually present data to our readers, users, and audiences and what words we should use in and around those visuals. (We are also working on a separate section about accessibility in data visualization, which we will share at a later date.)

Although more people are thinking and writing about these issues, there hasn’t been much agreement around best practices for taking an equity lens to data visualization, especially as it applies to setting standards for entire organizations. As best we can, we have been reading a variety of posts and papers (a short list can be found below) and discussing ways we can develop a more diverse, equitable, and inclusive approach to presenting and visualizing data. We view this effort as just the beginning of our process and anticipate growing and expanding our work as we learn more and receive feedback from colleagues, partners, and readers.

To that end, we have identified eight areas in which researchers, analysts, and anyone working with data can be more inclusive in how they present their data.

# Using language with a racial equity awareness

Titles, text, and labels are among the first things readers scan [2] when encountering a chart, so they present an important opportunity to apply racial equity awareness thinking. Urban researchers often rely on our organization’s guides for accessible and inclusive language for labeling and annotating graphs. Consistent with Urban style, we want our researchers to label their data using people-first language, such as “people with disabilities” rather than “disabled people.” Terms should also refer to people and not strictly to their skin color; for example, “Black people” not “Blacks.”

The terms and phrases we use continue to change. In writing about terminology around and about people with disabilities, Nicholas Steenhout [3] writes, “Disability language is never straightforward. It’s always nuanced. It always evolves.” That sentiment can extend to any underrepresented group, so as researchers, developers, and designers, we need to monitor the current lexicon and reflect the experiences of the people we study and the people we communicate with. Graph creators may want to consider including a footnote or endnote explaining why a particular term was used. We also encourage researchers to talk to their target audiences to give them the option to self-identify their preferred terminology to determine what language is most appropriate for them.

# Ordering data labels in a purposeful way

Many of our graphs and tables include different demographic groups, ordered simply as they appear in the raw data. There is likely little thought given to *how* estimates in tables or bars in graphs or small multiples are ordered, such as “White,” “Black,” “Hispanic,” “Other.” As graph producers, however, we should take a more active role in choosing how to order and present data values for different groups. Which group we choose to show as the first row in a table or the first bar in a graph can affect how readers perceive the relationship or hierarchy between groups; always starting with “White” or “Men” can make these groups appear as the default against which other groups should be compared, suggesting they’re the most important populations. How we choose to order may also reflect who we view as the intended audience for our visualizations. Starting with “White” or “Men” can make it seem as though those are the most important groups we are trying to communicate with. Urban does not have a universal rule that applies to all visuals, but a few issues are worth considering:

* Does your study focus on a particular community? If it does, that group should be presented first.
* Is there a particular argument or story you are trying to tell? If so, the order or presentation of results should reflect that argument.
* Is there a quantitative relationship that can guide how the groups are ordered? Can they be sorted alphabetically or by population size, sample size (weighted or unweighted), or magnitude or effect of the results?

# Considering the missing groups

It is also important to acknowledge who is and is not included in our data and charts. Many charts on race and ethnicity only show white people, Black people, and Latinx people but not smaller racial or ethnic groups. Often this is because of data limitations, in particular, small sample sizes. But even in those cases, how can researchers be more proactive to help organizations conducting surveys be more inclusive? How can we communicate to those organizations to help them conduct better surveys? Just because it may be harder to obtain data about certain groups doesn’t mean we shouldn’t still try to better understand their lives.

# Using colors with a racial equity awareness

Urban’s color palette is consistent for people with certain color vision deficiencies, and the contrast between those colors and white/black text meet basic accessibility guidelines [4]. Urban does not use color palettes that reinforce gender or racial stereotypes, such as baby pink and baby blue to represent women and men. We have not set a specific standard for which Urban color refers to which gender group or racial group, although we certainly recommend that researchers avoid using colors associated with skin tones (or worse, racial stereotypes).

In general, as data visualization producers, we need to be aware of how our use of colors, words, and categorizations can perpetuate or exacerbate inequities.

# Using icons and shapes with a racial equity awareness

We want to be careful and thoughtful when using icons in any data visualization. When showing groups of people, we should consider a mix of genders, races, and ethnicities. We don’t often use icons at Urban, but when we do, we should consider to whom we are presenting our results and how our icons might be perceived. We need to be conscious of how certain icons may not correspond to the content, such as an icon of a baby in a chart about child mortality.

Mis- or underrepresentation of certain groups in imagery and iconography also fails to take a racial or gender equity awareness perspective toward our data visualization work. A 2018 study by the Pew Research Center found that “men are overrepresented in online image search results across a majority of jobs examined” and that “women appear lower than men in such search results for many jobs.” Data visualization creators should ensure a variety of races and genders are depicted [5] when using icons and avoid icons that make inappropriate depictions of people or communities or reinforce stereotypes.

# Demonstrating empathy

One of the big challenges in visualizing data, and quantitative research in general, is the ability to help readers connect with the content. Standard graphs like bar charts, line charts, and pie charts, while informative, can abstract from the content and people being represented. Taking an empathetic view of the reader’s needs as they read or perceive information is one step to better data communication [6]. This kind of empathy is often couched in terms of producing specific graphs that meet the needs and expertise of our readers. But viewing empathy through a DEI lens would mean considering how the specific lived experiences and perspectives of our readers (not to mention the actual study populations) will perceive the information. As journalist Kim Bui [7] wrote in 2019, “approaching stories—and people—with more empathy creates better relationships with marginalised communities, builds trust and increases diverse coverage.”

# Engaging or reflecting lived experience

Ideally, chart makers would not just implement our guidelines above but would also reach out to members of the communities being visualized and ask for their feedback and advice. Are we using language consistent with how they refer to themselves and others? What have we missed in our visuals that are inconsistent with a DEI framework? How can we take a more empathetic approach to creating data visualizations that accurately and respectfully account for other people’s lived experiences? It’s important to remember that data are a reflection of the lives of real people, not just a sterile abstraction.

# Looking forward

As we continue to revise our data visualization style guide, we’ve noticed few, if any, style guides from other organizations [8] mention accessibility, inclusion, or diversity. Our approach has been to create a set of recommendations and issues to consider rather than a set of rules that researchers must follow. Ultimately, we hope researchers will be thoughtful and deliberate about their design choices [9], not just relying on software defaults or the status quo. If you or your organization has gone about setting guidelines for DEI in your visualizations, we would love to hear about them.

Acknowledgements

References

[1] Urban Institute. “Urban Institute Data Visualization Style Guide.” Accessed July 23, 2020. <http://urbaninstitute.github.io/graphics-styleguide/>.

[2] Borkin, Michelle A., Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. “Beyond Memorability: Visualization Recognition and Recall.” *IEEE Transactions on Visualization and Computer Graphics* 22, no. 1 (August 2015): 519–28.

[3] Steenhout, Nicholas. “Disability Language Is a Nuanced Thing,” *Part of a Whole* (blog). July 9, 2020. https://incl.ca/disability-language-is-a-nuanced-thing/.

[4] WebAIM. “Contrast Checker.” Accessed July 23, 2020. <https://webaim.org/resources/contrastchecker/>.

[5] Kim, Erika. “Visible Women: Gender in Iconography.” Noun Project, October 29, 2019. <https://blog.thenounproject.com/visible-women-gender-in-iconography-c6518ec1ddb8>.

[6] Schwabish, Jonathan. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks.* (New York: Columbia University Press, forthcoming).

[7] Bui, Kim. “Designing Data Visualisations with Empathy.” Datajournalism.com, June 11, 2019. <https://datajournalism.com/read/longreads/designing-data-visualisations-with-empathy>

[8] PolicyViz. “Style Guides.” November 30, 2016. <https://policyviz.com/2016/11/30/style-guides/>.

[9] Costanza-Chock, Sasha. “Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice.” *Proceedings of the Design Research Society* (2018).