



DO NO HARM GUIDE

Applying Equity Awareness
in Data Visualization

JONATHAN SCHWABISH AND ALICE FENG

JUNE 9, 2021

TABLE OF CONTENTS

01 INTRODUCTION	3
02 RACE AS A CONSTRUCTION	4
03 DEMONSTRATING EMPATHY	5
04 ENGAGING OR REFLECTING LIVED EXPERIENCES	9
Methods	10
05 THE DATA BEHIND THE VIZ	11
Data Confidentiality and Security	12
06 USING LANGUAGE WITH A RACIAL EQUITY AWARENESS	13
07 ORDERING DATA PURPOSEFULLY	15
08 CONSIDERING THE MISSING GROUPS	17
Lumpers and Splitters	17
Male, Female, In Between, and Beyond	19
Othering the Other	20
Choosing to Not Include All Groups	21
09 USING COLORS WITH A RACIAL EQUITY AWARENESS	22
10 USING ICONS AND SHAPES WITH A RACIAL EQUITY AWARENESS	23
11 EMBRACING CONTEXT AND COMPLEXITY	25
12 BUILDING DIVERSE AND INCLUSIVE DATA COMMUNICATION TEAMS AND ORGANIZATIONS	26
13 THE FEEDBACK LOOP	27
14 THE GATEKEEPERS	28
Editors and Peer Reviewers	28
Government Agencies, Regulations, and Directives	29
Funders	30
15 CASE STUDY FROM THE URBAN INSTITUTE	30
16 LOOKING FORWARD	33
17 ACKNOWLEDGMENTS	34
18 REFERENCES	35
19 INTERVIEW GUIDE: DO NO HARM GUIDE AGENDA	39
20 OTHER DATA AND DATA VISUALIZATION RACIAL EQUITY RESOURCES	40
21 DIVERSITY, EQUITY, AND INCLUSION IN DATA VISUALIZATION: GENERAL RECOMMENDATIONS	41
22 THE RACIAL EQUITY IN DATA VISUALIZATION CHECKLIST	42

01 INTRODUCTION

People who work with data every day seek to make discoveries, elicit insights, and offer solutions. To do so in an equitable way, data analysts should think intentionally about how we can learn from and speak to audiences that reflect the rich diversity of the people and communities of focus. Failing to recognize the diversity in these groups will further exacerbate—and potentially contribute to—the inequities that have shaped the world. Today, in the era of big data, data visualization, machine learning, and artificial intelligence, we need to be more purposeful about where data are coming from and how research and the communication of that research can affect people, their communities, and the policies that touch their lives.

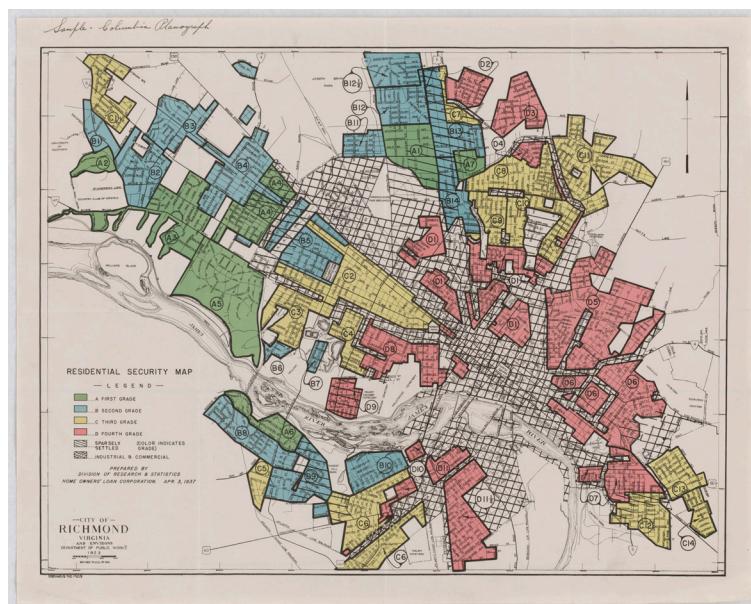
Personal and organizational efforts to foster diversity, equity, and inclusion (DEI) need to extend not only to our internal processes of hiring, promotion, and professional development but also to our external communication efforts to represent the people and communities we focus on. To that end, we need to consider how to apply a DEI lens not just to the words, colors, icons, and other elements in our writing, graphs, charts, and diagrams,

but also to the process of crafting these communication products.

This guide and the associated checklists and toolkits focus on the often hidden or subtle ways that data analysts and communicators fail to incorporate equitable awareness in the data they use and the products they create. Some data and data visualizations, however, are overtly racist and discriminatory. For example, the map shown in figure 1 was produced in 1937 by the Home Owners' Loan Corporation (HOLC), a federal agency tasked to appraise home values and neighborhoods across the United States. Richard Rothstein described the HOLC and these maps in his book, *The Color of Law*: “The HOLC created color-coded maps of every metropolitan area in the nation, with the safest neighborhoods colored green and the riskiest colored red. A neighborhood earned a red color if African Americans lived in it, even if it was a solid middle-class neighborhood of single-family homes” (Rothstein 2017). The bulk of this guide does not focus on data and visuals that are overtly racist, sexist, or discriminatory, but as data practitioners, we still need to

FIGURE 01

This redlining map of Richmond, Virginia, from the Home Owners' Loan Corporation shows how data and data visualization can be used to further systematic discrimination.



Source: National Archives.

be aware of how such work often creates irreparable and long-term harm. Systematic discrimination is and can be generated by how we use and misuse our data.

In this guide, we explore ways to help data scientists, researchers, and data communicators take a more purposeful DEI approach to their work. To do so, we conducted more than a dozen interviews with nearly 20 people about their experiences and approaches to being more inclusive with their data exploration, analysis, and communication. Our interviewees included data journalists in major media outlets, researchers at universities and colleges, and people working in the public and private sectors. Our long (and growing) list of people we would have liked to speak with for this project indicates how the work of creating diverse and inclusive products in the data and data visualization areas will continue to grow. We also take a broad view of the “data community” to include anyone working with and communicating data: researchers, scholars, data analysts and scientists, journalists, web developers, data visualization specialists, and more.

Our recommendations are not one-size-fits-all and should be adapted to the particular needs of a project and organization. A large research project with a team of researchers that spans multiple months or years might embrace some of these outreach efforts differently than a single data visualization intended for a 600-word blog post or a standalone dashboard. Through honest conversation and evaluation of different projects, data project creators can be better prepared to incorporate the strategies described here in future work.

02 RACE AS A CONSTRUCTION

In this guide, we focus primarily on race and ethnicity, but similar issues of inclusivity around gender, sexuality, class, ability, and other characteristics (as well as their intersectionality) should always be at the forefront for every data analyst and communicator. The principles outlined here are also applicable to addressing these other aspects of identity.

With respect to race specifically, it is important to remember that race is an artificial human construct. Racial and ethnic categories such as “Asian,” “Black,” “Hispanic,” “white,” and others are classifications and labels society

The motivation for this entire project can be summarized in one word: **empathy**. Applying a DEI lens to how we analyze, visualize, and communicate data requires empathizing with both the communities whose data we are visualizing as well as the readers and target audiences for our work. This means considering how the lived experiences and perspectives of our study populations and readers affect how they will receive and perceive the information. This quote from journalist Kim Bui succinctly captures the concept: “If I were one of the data points on this visualization, would I feel offended?” As we consider the use of words, colors, icons, and more in our data visualizations, asking whether we would be offended makes for a good checkpoint.

If I were one of the data points on this visualization, would I feel offended?

- KIM BUI

This guide offers a set of guidelines rather than ironclad rules for presenting data through a DEI lens. Our goal is not to tell data analysts what to do or not to do but to instead ask them to be thoughtful in the ways they work with and communicate their data and to be aware of the decisions they have made and why they have made them. As we will show, the choices we make about the colors, shapes, words, and representations in our data analysis and visualizations can affect how people perceive the final results, how change might be implemented, and how that change will impact different people and communities.

places on people, but they are not rooted in biology or genetics. “Race,” writes scholar Ibram X. Kendi (2019) in *How to Be an Antiracist*, is “a power construct of collected or merged difference that lives socially” (35). Later in the book, Kendi writes:

...race is a mirage, which doesn’t lessen its force ... What people see in themselves and others has meaning and manifests itself in ideas and actions and policies, even if what they are seeing is an illusion. (37)

Thus, as communicators consider how to analyze and visualize issues of race, ethnicity, and other characteristics, they should be cognizant of how these categorizations are used and perceived in different fields and areas of study. In the United States, as well as many other countries, the social construct of race has been, and continues to be, used to create and maintain systems that establish the power and privilege of certain groups (i.e., white people) while discriminating against and oppressing other groups (i.e., Black, Hispanic, Native American, Asian, and other people). Such racism does not just operate at the individual and interpersonal levels but is baked into institutions, laws, and norms, and thus it manifests in the form of institutional and structural racism that perpetuates these benefits and harms across generations. This has resulted in lasting impacts among people and communities of color, such as the lower levels of wealth held among Black and Hispanic families (McKernan et al. 2017) and higher rates of chronic health conditions among people of color (Carratala and Maxwell 2020). As Isabel Wilkerson (2020) writes in her book *Caste*, "The

human impulse to create hierarchies runs across societies and cultures, predates the idea of race, and thus is farther reaching, deeper, and older than raw racism and the comparatively new division of humans by skin color" (67).

When thinking about how to analyze and communicate data with a racially equitable lens, we should ask ourselves whether the issue we are analyzing and visualizing is worse or exacerbated for people of color and what factors contribute to or compound racial inequities for this issue. These issues have taken even greater urgency and focus with the announcement from the Biden administration that the federal government "should pursue a comprehensive approach to advancing equity for all, including people of color and others who have been historically underserved, marginalized, and adversely affected by persistent poverty and inequality" (White House 2021). It is important to be informed and aware of the policies, institutions, and actors that have shaped such inequities and disparities and to acknowledge this history and context in our communication products.

03 DEMONSTRATING EMPATHY

You can't feel empathy for a point on a map, but you can feel empathy for a mom who's crying.

— KIM BUI

One of the big challenges in visualizing data, and of quantitative research in general, is the ability to help readers connect with the content. Merriam-Webster defines empathy as "the action of understanding, being aware of, being sensitive to, and vicariously experiencing the feelings, thoughts, and experience of another of either the past or present without having the feelings, thoughts, and experience fully communicated in an objectively explicit manner." Connecting directly with people and communities and trying to better understand their lived experiences can help content producers create visualizations and tell stories that better reflect those people's true experiences.

This is a key dimension for people working with data to explore to help put their work into the hands of policymakers, stakeholders, and community members who can use it to effect change. Inclusive and thoughtful data visualizations that respectfully reflect the people we are studying can help us build trust with those communities. As Kim Bui (2019) wrote, "approaching stories—and people—with more empathy creates better relationships with marginalized communities, builds trust and increases diverse coverage."

We think of empathy as it applies to communicating data across six main themes:

1. Put people first.

First and foremost, we need to remember and communicate that the data shown reflect the lives and experiences of real people. Data communicators must help readers better understand and recognize the people behind the data. As Jacob Harris (2015) wrote, "If your data is about people, make it extremely clear who they are or were."

2. Use personal connections to help readers and users better connect with the material.

As Kim Bui told us, "You can't feel empathy for a point on a map, but you can feel empathy for a mom who's crying." Pairing data-driven charts with personal stories centered on individual experiences can help readers understand and identify with the people represented in the research and data visualizations. Techniques that can be used in tandem with data visualizations to help lift up personal stories include photography, illustrations, pull quotes, and oral histories. Helping readers understand the "near and far" (Harris 2015)—a wide lens of overall metrics or data paired with individual- or group-level data—may help readers connect with content (Klein, n.d.).

Making charts relatable can also be important for audience engagement and impact. In their study of what factors influenced attitudes and perceptions about data visualizations in rural populations, Evan Peck and coauthors found that visualizing topics that the audience had personal connections to trumped design factors when rating which charts they liked the most. They write, "Regardless of style, clarity, or ease-of-understanding, our interviews served as a reminder that data can be intimate and personal, and that those ties may supersede many other dimensions of design" (Peck, Ayuso, and El-Etr 2019). Similarly, allowing users to find themselves in the visualization by, for example, being able to search for their city or county in a map, can also strengthen audience engagement with the data. There is an important balance here, of course, between being able to search and filter for a specific geographic area or subpopulation and data privacy and confidentiality, which we address in section 5.

3. Use a mix of quantitative and qualitative approaches to tell a story.

Most charts and graphs are built on top of spreadsheets or databases of quantitative data. However, focusing on numbers alone without any context can overlook important aspects of a story including the "why" and the "how." We discuss other research methods that analysts can use in section 4.

4. Create a platform for engagement.

This can take the form of interactivity in which users are able to manipulate buttons, sliders, tooltips, and other elements to make selections, filter the dataset, or create customized views of a chart. Such engagement can be leveraged as a way to allow users to find

themselves in the data or discover the stories that most interest them. Another form of engagement is offering audiences a means of providing feedback about a data tool or visualization. Feedback forms offer users a way to share their own experiences with the topic being visualized. We discuss this further in section 12.

5. Consider how your framing of an issue can create a biased emotional response.

In our discussion with Kim Bui, she pointed to coverage and discussion about racial justice protests in the spring and summer of 2020. Much of that coverage, she argued, tended to focus on the number of riots that occurred or how much property damage took place, which portrayed the protests and protesters in a negative light while missing the cause of those protests: "people trying to get more justice for Black and brown lives that have been taken from us." Similarly, in her article "When the Designer Shows Up in the Design," Lena Groeger discusses two different ways to visualize the impact of crime on local communities: maps showing the locations where crimes occurred versus maps showing the percentage of residents in a neighborhood who were in prisons. The former focus on victims of crimes while the latter "show the urban costs of incarceration and suggest how those dollars might be better spent on investing in communities" (Groeger 2017). What data we choose to focus on and what we choose to ignore can bias our audiences' perceptions of the issues we are communicating about.

6. Recognize the needs of your audience.

Taking an empathetic view of the readers' needs as they read or perceive information is also an important step to better data communication (Schwabish 2021). This kind of empathy is often couched in terms of producing specific graphs that meet the needs and expertise of our readers. For example, nearly 28 million people in the United States alone have challenges accessing content on the web, ranging from issues of sight to physical or intellectual impairments (Brault 2012). Making sure our visualizations are accessible is important so people with disabilities can view and use online content (see also Lundgard, Lee, and Satyanarayan 2019). Similarly, overly technical or jargon-laden language can make reports and articles inaccessible to broader audiences and less likely to be cited by other researchers in the field (Martínez and Mammola 2021). Another consideration is translating materials into languages most used by your audiences.

Being empathetic to the people and communities of focus does not imply sacrificing the data and methods in responsible, in-depth, sophisticated research. In fact, the opposite is true: research and analysis must be responsible and as objective as possible to create effective policy solutions and recommendations. We are not arguing that empathy replaces high-quality research or data work; rather, such research and empathy for people and communities can be complementary. Effective research necessarily means understanding someone else's point of view nonjudgmentally and recording that perspective as accurately and truthfully as possible. Engaging in empathy allows researchers and data analysts to better understand the source and full context of the data they are working with. (Given the considerations around bias in data and in human decisionmaking generally, we hesitate to use the term "objective" as it applies to research, but that is essentially what we are trying to imply here—that empathy and objectivity in research are not opposites but complements for generating more effective research and policy solutions.)

In thinking about embedding empathy into the design of data visualizations, it is worth considering whether certain chart types better lend themselves to connecting with the human element of the data being portrayed. Commonly used charts and graphs like bar charts, line charts, and pie charts, while informative, can abstract the content by collapsing all the people represented into one shape.

These are not just data points, and so when you add them as dots on a map, or lines on a chart, I think you have to remember that these are true people.

— TIM MEKO

Unit charts and dot plots, which use multiple repeated shapes to depict the data, might offer more opportunity to connect with the subject by reminding readers of the number of people represented, particularly if each dot represents one person.

Taking this thinking one step further, the use of icons instead of abstract shapes such as circles and rectangles may also improve the ability of readers to empathize by reinforcing that they are looking at people and not just numbers or statistics. As Tim Meko from the *Washington Post* told us, “These are not just data points, and so when you add them as dots on a map, or lines on a chart, I think you have to remember that these are true people.” Graphics that specifically represent people—the anthropomorphizing of data graphics, or, as Jeremy Boy and colleagues refer to it, anthropographics (Boy et al. 2014)—is sometimes seen as a way to evoke empathy (though Boy and his coauthors do not find this to be the case; see also Groeger 2014). An example of a stacked bar chart compared with a unit chart that uses icons is shown in figure 2.

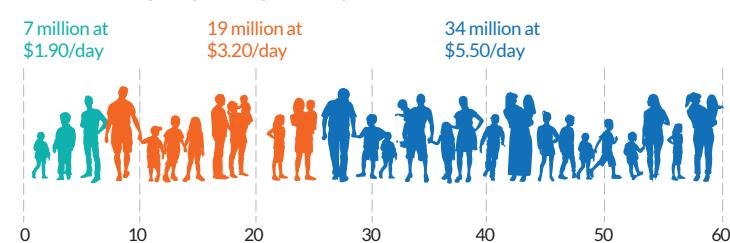
FIGURE 02

A comparison of a standard stacked bar chart and a unit chart with icons.

Millions of people in poverty



Millions of people in poverty



Source: Recreated based on Schwabish (2021).

In a paper examining the design space of anthropographics, Luis Morais and authors articulate seven design dimensions that may affect the ability of a data visualization to elicit empathy or compassion from the audience (Morais 2020). These include *granularity* (the number of people represented by one shape or graphical mark), *specificity* (how completely each item in a chart can be distinguished from another), *coverage* (the number of people visualized), and *authenticity* (whether the chart shows any additional attributes that are not derived from the source dataset). In addition to what and how much information is shown, how that information is represented is also worth considering. This includes the *realism* of the visualization, with abstract shapes such as bars and lines being the least realistic and photographs being the most; *physicality*, meaning the use of physical versus virtual objects; and *situatedness*, or how physically close or far a visualization is from the people or setting it describes. Although current research does not yet offer guidance on the effectiveness of these dimensions, they may be aspects worth considering when designing your data visualizations.

Using human icons to visualize data does not necessarily guarantee empathy, however. The chart in figure 3 has been ridiculed as an example of how not to use icons. The icons are scaled inappropriately by both height and width, the y-axis does not start at zero, and the icons use stereotypical pink colors. The scaling distorts the representation of the data, and the resulting chart has been perceived as offensive to many women, particularly those in the six countries portrayed in the chart. Sabah Ibrahim, an educator and environmental campaigner, shared this tweet in August 2020: "As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me." While the tweet is somewhat tongue-in-cheek, it hearkens to the initial motivation in this paper: "If I were one of the data points, would I feel offended?"

The issue of empathy and how data communicators should take an empathetic view to the data they are communicating and the people they are communicating

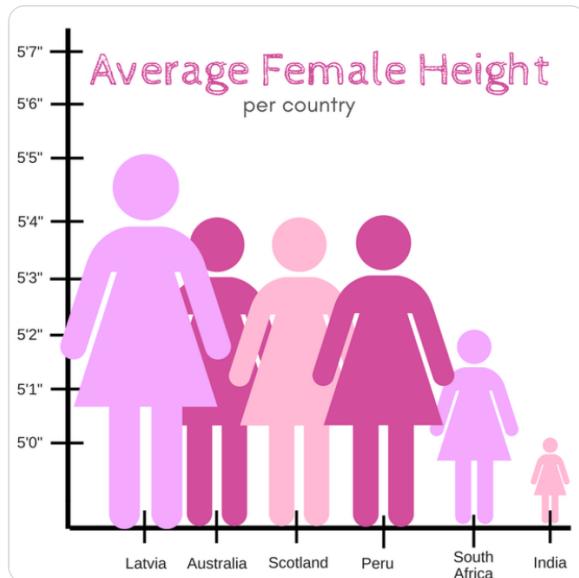
about underlies the concept of conducting research and making data visualizations that uphold DEI. In the case of racial equity, we can take that concept of empathy a step further and recognize the histories of racism, injustice, abuse, and discrimination that serve as the background for many of the current issues we analyze, visualize, and debate.

FIGURE 03

A tweet from Sarah Ibrahim demonstrates how people can see themselves in the data.



As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me



6:58 PM · Aug 6, 2020 · Twitter Web App

Source: Sarah Ibrahim (@reina_sabah), "As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me," Twitter, August 6, 2020, 6:58 p.m., https://twitter.com/reina_sabah/status/1291509085855260672?lang=en.

04 ENGAGING OR REFLECTING LIVED EXPERIENCES

Don't assume, ask people.

— KIMBERLYN LEARY

The need for data communicators to understand and form relationships with the people and communities that are the focus of their research emerged as a common theme in all of our discussions. Research and analysis should be done with communities, not on behalf of them, except in the rare cases where doing so is not feasible or appropriate. Such communities should derive value from the research and data and have some level of input into the design of the project itself and the final deliverables. This kind of buy-in—from members of the community, policymakers, and other stakeholders—can help the research be more impactful, relevant, and embraced by a wider audience.

We do not, however, intend to suggest that research objectivity and independence be sacrificed to engage with communities. We also recognize that interacting with people and communities does not work for every research project or data visualization: national, aggregated data from the federal government, for example, might be less appropriate for these approaches than projects involving first-hand data collection. But for projects with opportunity for such engagement, there is likely a middle ground where community members can weigh in and feel heard and recognized.

In light of these considerations, researchers and analysts should reach out to and engage communities from the start of the project. This is an important step for making sure a DEI framework and empathy are baked into the project and kept in mind from the beginning rather than “retrofitted” at the end. Reaching out to local organizations and nonprofits, service providers, social justice groups, and advocacy groups, as well as to your personal network for connections or searching the internet for people talking about the topic you are interested in, can help identify community partners. Some universities and newsrooms have engagement teams or offices that specialize in working with local organizations; such teams can also be a source for forming connections with communities.

A critical aspect to keep in mind is understanding a community's culture, history, and values. Different communities have different preferences for how they wish to be approached, and it is important to respect those preferences. For example, members of the University of California Los Angeles (UCLA) Center for Health Policy Research told us that Native Hawaiian Pacific Islander communities prefer to meet face to face and have the researchers go to them rather than asking community members to come to the researchers. Relationships, they argued, need to be built before you can ask for a favor (see also Costanza-Chock 2018).

After establishing such relationships, researchers and analysts should seek to understand the needs of the community and how the work they are doing would benefit them, engaging them as partners early on in the process. Kimberlyn Leary, senior vice president at the Urban Institute, recommended that researchers “mak[e] sure that you put in the time to understand not just the disparities, but engage with people to find out and offer some solutions... don't assume, ask people.” Kim Bui echoed that sentiment, saying it is important to tell stories with a community rather than on behalf of them and to seek out what they would want to learn and what would be useful for them along with their concerns. Asking community members for their perspectives can also help researchers better understand the context surrounding the issue they wish to study as well as how the community perceives that issue. Similarly, community members can help interpret the data, but which community members (or groups) researchers should speak to and how many people they should speak with are important considerations, especially when research funds are limited or there is significant variation within the community. Those issues underscore the importance of engaging with communities at the beginning of the research project.

While engaging with communities, researchers and analysts should be clear and transparent up front on what their intentions are and what they will be doing with the data they gather. Where possible and appropriate, the project's goals should align with the community's goals, and benefits should ideally be tangible early on in the

project cycle rather than occurring at the end. Research questions should reflect the community's values and account for cultural and historical context. Community members' input should be sought to ensure that the design of the research upholds these goals (but again, the data used and methods applied should be rooted in the science objectively and honestly). Above all, researchers should not exploit people and their experiences by taking data from the community and not giving back. Community members should not only benefit from the results of the work but also have a vested interest in the research, data, and final products. However, because not all community members will agree on certain findings or policy recommendations, the ultimate responsibility for determining the final conclusions still lies in the hands of the researchers.

As the project nears the end, results, data, and final products should be shared with the community first before being released to the public. Community members should be given a chance to review and provide feedback before publication. Questions to ask about communications products include the following: Are we using language consistent with how community members refer to themselves and others? What in our visuals is inconsistent with a DEI framework? And how does the community prefer to have the research results presented? The feedback received should then be critically reviewed by the research team and incorporated into final visualizations and products as appropriate, and community members should be credited for their contributions.

Finally, once built, these relationships should be sustained. This is important for fostering and maintaining trust so the community does not feel like it is being used or taken advantage of and to ensure that the community is able to act on the recommendations that result from the research. Such relationships can also spur collaborations and partnerships for future projects.

METHODS

Quantitative researchers and analysts especially should consider how best to incorporate qualitative methods when conducting research that engages lived experiences. Long-form surveys, interviews, and focus groups can provide an important opportunity for community members to share their experiences

and lift up their voices. Such questions can also help answer "why" questions and surface themes in the responses. Evan Peck from Bucknell University, for example, interviewed 42 people in rural Pennsylvania to better understand their relationships to data and data visualizations (Peck et al. 2019). Peck and coauthors found that people's reactions to data visualizations were often driven or framed by their personal experience (e.g., where they live or work, their level of educational attainment, or their political identity), suggesting that our existing knowledge of how data visualizations facilitate understanding is much broader than whether and how people can discern quantities from the graph (Cleveland and McGill 1984; Heer, Bostock, and Ogievetsky 2010). Their conversations with study participants enabled them to better understand the detailed explanations behind their reasoning about how they read and understand graphs that more standard quantitative methods may not have permitted. In our interview with Peck, he elaborated further and noted that he suspects his team had higher participation in their interviews because participants could see how the data were being collected (rather than being collected on an impersonal form), so they had "some sort of understanding of how the data is collected."

More generally, our interviews suggest that embracing a "reciprocal research" strategy—in which research participants see concrete, actionable benefits—makes research stronger or more accessible and its recommendations more likely to succeed (Myers and MacDonald 2021). With data visualization specifically, some have argued that information visualization should empower users to allow them to "question visual representations, utilize them to tell their own story, and shift from awareness to action" (Dörk et al. 2013, 7). Dr. Sasha Costanza-Chock at the Massachusetts Institute of Technology has helped popularize the Design Justice model that also recognizes that community participation is not singular nor necessarily consistent during the project. "The point" Costanza-Chock writes, "is to encourage a more complex understanding of participation and to emphasize that very few design processes are characterized by community control throughout" (Costanza-Chock 2018).

Again, these approaches will not work for every research project and every data visualization. Many projects will not interact with people directly or will address

aggregate concepts affecting numerous groups, and many visualizations will be geared toward a specific purpose for a specific audience. But research and data communities should at least consider these issues as they progress through the project.

We conclude this section with a final caveat about these kinds of qualitative methods and research: we are not suggesting that pursuing qualitative research is as simple as conducting a few interviews. Entire fields are devoted to qualitative methods, which require study, training, and practice to master. In our experience, we have found that

many primarily qualitative researchers have at least some quantitative training: they know how to clean quantitative data, estimate summary statistics, and run regressions. By comparison, many quantitative researchers have little concept of what it means to conduct qualitative research—the methods, skills, and practices to effectively collect and understand such data using these kinds of techniques. We hope that this guide can help move qualitative and quantitative research fields closer to one another.

05 THE DATA BEHIND THE VIZ

Before we jump into discussing ways to apply racial equity awareness to charts and maps, we should first step back and critically examine the data underlying them. Just as trying to analyze flawed data generates flawed results, trying to visualize flawed data will also generate misleading and deceptive charts. When thinking about DEI issues specifically, we should check our data for ways they may be biased or the product of racist or oppressive data collection systems. As Evan Peck noted, data visualizations amplify data by making them more accessible. Thus, if the data underlying the visualization are biased, the chart will amplify that bias and the harm that bias generates.

A full discussion of how to identify and handle biased data is beyond the scope of this guide, but the questions below offer a good starting point for what data analysts and communicators should be asking when they receive a dataset (for more ideas, see our list of additional resources provided at the end of this guide):

- How were these data generated?
- Are these data demographically representative?
- Who is included and who is excluded from these data?
- Whose voices, lives, and experiences are missing?
- How much can this data be disaggregated by race, gender, ethnicity, etc.?
- Why were these data collected?
- Who stands to benefit from these data?
- Who might be harmed by the collection or publication of these data?

Data, particularly data collected about people, are not neutral or objective: they reflect a human act that was done by someone for a specific purpose and intent (Correll 2019; Lupi 2017). Data visualizers and communicators must first thoroughly understand the data they are using, how those data were collected, why they were collected, and who is and is not reflected in those data before they begin creating charts and graphics. We strongly emphasize this point because applying DEI thinking to visualizations alone will not fix the problems of data or analyses that are biased or racist. Researchers and analysts should also consider the root causes of disparities that are revealed in the data and frame their work and interpretations in light of that knowledge.

Taking a racially equitable perspective to data should be baked into a project from the beginning and continue throughout the process. This entails critically examining how the study is designed and how the data are collected as well as how the analysis of that data is being conducted and the results communicated. If the data or analysis are inherently racist or biased, applying the principles in this guide to the way you visualize these data will not magically make your visualization “anti-racist” (Kendi 2019). It can also be valuable to think about the impact of racism and other forms of oppression on the data you are visualizing even if your charts do not explicitly show racial or ethnic breakdowns. DEI thinking cannot be a Band-Aid slapped on at the end to fix a project that is inherently racist or perpetuates oppression.

DATA CONFIDENTIALITY AND SECURITY

Two areas of working with data do not always lend themselves to full and transparent data communication: uncertainty, and privacy and security. Uncertainty in data adds an important challenge to all data analysis and is one that affects not only how data analysis is performed but also when and how uncertainty is presented in data communication (Hullman 2019). Certain audiences are familiar with measures of uncertainty and dispersion, such as confidence intervals, standard errors, percentiles, and variances; other audiences are less fluent in these concepts. Thus, recognizing your audience and providing appropriate labels and annotations to help them better understand these concepts can be central to conveying issues of uncertainty and distribution. Certain conclusions from uncertain evidence can be just as bad as incorrect conclusions from certain evidence. For this reason, good analysis should try to include estimates and measures of uncertainty. Failure to properly communicate uncertainty is a problem with many data projects and can become an even bigger problem when using disaggregated data, especially owing to small sample sizes. Data communicators should explore new and different ways

to communicate uncertainty to their audiences, such as different plot types, better annotation, and even different tools to help users interact with data (Hullman 2019; Nguyen et al. 2020).

With respect to data confidentiality and security, we recognize that many recommendations in this report—such as talking to members of the community or enabling users to see themselves in the data—do not necessarily align with data privacy and security considerations, especially when data are disaggregated. Publishing detailed, disaggregated data can help policymakers and decisionmakers better target policy solutions, but we must carefully consider the unintended harms such detailed data can cause to certain people and groups. A recent report from the Urban Institute notes that “people of color with low incomes are more susceptible to privacy attacks because of their higher reliance on smartphones for internet access and how much personal information they give up for free cell phone app services. This information collection makes them more easily identifiable, especially if they are outliers in small geographies” (Bowen, Williams, and Narayanan 2021; see also Madden 2017 and Bowen 2020).

06 USING LANGUAGE WITH A RACIAL EQUITY AWARENESS

Titles do such important work to frame the visual part.

— CATHERINE D’IGNAZIO

Titles, text, and labels are among the first things readers scan when encountering a chart (Borkin et al. 2015), thus presenting an important opportunity to apply racial equity awareness thinking. When possible, forces of oppression such as racism, how those forces are operating, and their historical context should be named directly in the chart through titles, annotations, labels, and notes rather than buried in surrounding text.

The importance of using concise, active titles in data visualizations should not be underestimated. “Titles,” Catherine D’Ignazio mentioned in our conversation, “do such important work to frame the visual part.” Thus, they present an important opportunity to address the

role racism and other forms of oppression play in the topic the chart is visualizing. In their book *Data Feminism* (2020), D’Ignazio and coauthor Lauren Klein offer an example of a bar chart showing the rate of mental health diagnosis of incarcerated people by race (figure 4). The initial title for the chart is “Mental Health in Jail: Rate of mental health diagnosis of inmates.” Although this title seems neutral and objective, it ignores the role racism and discrimination play in how likely incarcerated people are to receive a mental illness diagnosis; it also uses the term “inmate,” which some have argued is dehumanizing and references people by their crimes and punishments (Bartley 2021). The authors then show a different title for the chart: “Racism in Jail: People of color less likely to get mental health diagnosis.” They argue that this alternative title more accurately reflects the main findings of the research (which focused on racial disparities in the jail system), names the forces of oppression at work (racism in prison), and references people, not inmates.

Another alternative takes the process a step further. In figure 5, the subtitle is changed to “White people get more mental health diagnoses,” which focuses not on what people of color lack, but as D’Ignazio and Klein (2020) describe, on the “unfair advantages that are given to the dominant group.” “Placing numbers in context,” they write, “and naming racism or sexism when it is present in those numbers should be a requirement—not only for feminist data communication, but for data communication full stop.”

In addition to the way language can be used to frame the larger picture and context of a chart, it is important to be aware of how language is used in the more detailed parts of a chart, such as labels. Labels should use people-first language, such as “people with disabilities” rather than “disabled people,” or “people in prison” instead of “inmate,” (Bartley 2021) and they should refer to people, not to their skin color (for example, “Black people” rather than “Blacks”).

As an example, in a June 2020 project we examined, a series of bivariate choropleth maps showed the relationship between race and poverty. In the original visualization, the labels along the legend were “More Poverty” and “More Black” (figure 6). That language is

not inclusive of different groups: poverty refers to an experience, not a static description; and “More Black” references skin color, not people. A more inclusive way to label the legend might be “Larger proportion of people experiencing poverty” and “Larger Black population” (the author of the visualization later changed “More Black” to “Larger Black Population,” and we are grateful the author provided us with permission to include this example here).

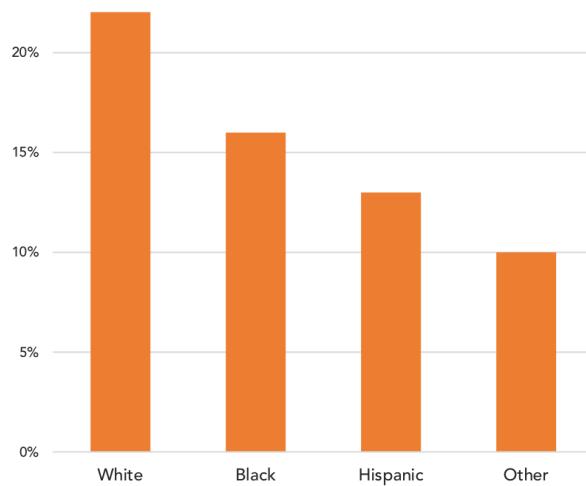
Another challenge when it comes to language is how the terms and phrases used to describe people and communities are constantly changing. In writing about terminology concerning people with disabilities, Nicholas Steenhout (2020) writes, “Disability language is never straightforward. It’s always nuanced. It always evolves.” That sentiment can extend to many underrepresented groups; thus, data communicators should monitor the current lexicon and reflect the experiences of the people we study and the people we communicate with. For example, in an August 2020 report, the Pew Research Center (Noe-Bustamante, Mora, and Lopez 2020) found that in their sample of more than 3,000 people who self-identified as Hispanic or Latino, only 23 percent had heard of the term “Latinx” (a gender-neutral alternative that has emerged in the past few years), and

FIGURE 04

Example of moving from a generic descriptive title to one that reflects the main findings of the research.

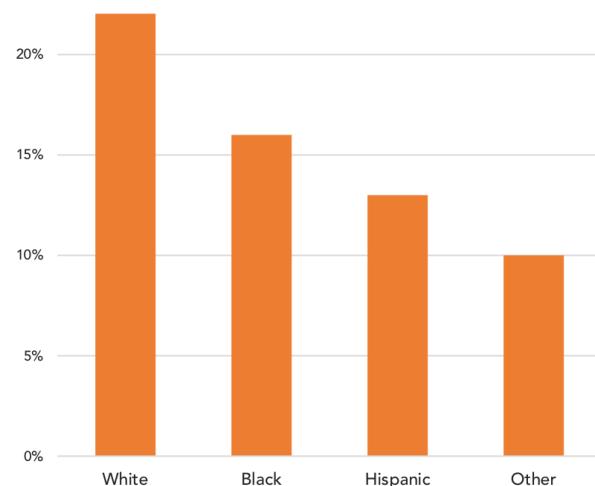
Mental Health in Jail

Rate of mental health diagnosis of inmates



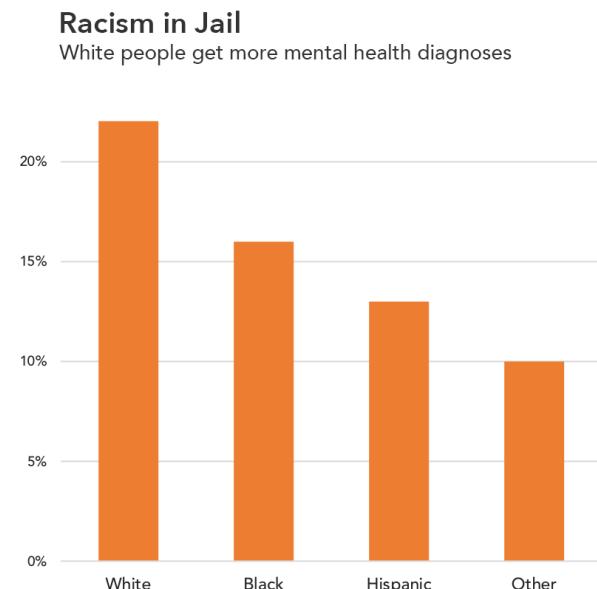
Racism in Jail

People of color less likely to get mental health diagnosis



Source: D’Ignazio and Klein (2020); Data from Fatos Kaba et al., “Disparities in Mental Health Referral and Diagnosis in the New York City Jail Mental Health Service,” *American Journal of Public Health* 105, no. 9 (2015): 1911–16.

FIGURE 05
Example of a chart title that focuses on the dominant group.



only 3 percent use the term. Of those who have heard the term "Latinx," 65 percent say it should not be used to describe the Hispanic or Latino population.

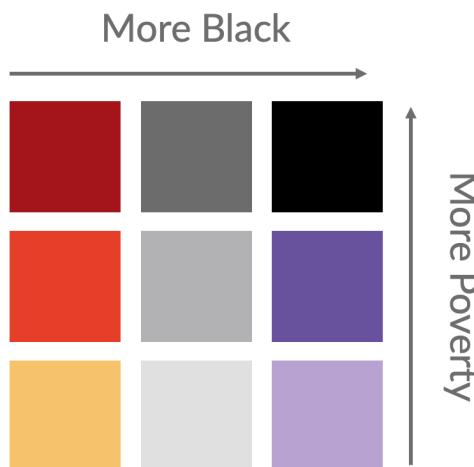
There may also be generational differences in preferred terms. The same Pew survey found that younger people were more likely to have heard of the term "Latinx" than older generations. G. Cristina Mora captured this well in our discussion by asking whether the data communicator wants to "land future oriented" or "land now." In that sense, she argued, using the term "Latina/o/x" might be "the safest best or the most inclusive bet" (although there is also a growing movement to use the term "Latine" because it can be easier to pronounce and "-e" can be used with other Spanish words more easily than "-x"; Blas 2019; Del Real 2020.) Given that younger people who identified as Hispanic or Latino might also be more politically progressive, this might also reflect the broader intersectionality of politics, age, and race or ethnicity.

Whichever term the data communicator chooses, one theme that was consistently raised in our discussions was the need for data communicators to talk with their study population and readers to understand and identify

their preferred terminology. Doing so also offers the community whose data were used an opportunity to gain more ownership of the data, research, and policy recommendations.

Finally, we also recognize that surveys or datasets that use outdated or less preferred terminology can present a tricky situation in terms of language. Data communicators may be understandably reluctant to change the word or phrase that was used in the original survey in their final report, graph, or dashboard. If, for example, survey respondents choose the "Hispanic or Latino" option in the survey, is it appropriate for the final product to use a different term, such as "Latinx"? Several of our interviewees signaled that this kind of change is acceptable as long as it is noted in the text or a note. Speaking from the perspective of a data journalist, Lena Groeger suggested that this strategy is similar to paraphrasing an interview and that the creator should explain the decision and the process used to arrive at it. Others noted that such change would not be acceptable because survey participants responded to these terms, which may carry particular connotations. In such cases, the original survey language should be used in tables or charts, while an explanation of why this is not the preferred way of describing such populations can be woven into the text.

FIGURE 06
Legend where the labels use language that is not inclusive of different groups.



Source: Recreated from Tableau dashboard.

07 ORDERING DATA PURPOSEFULLY

Often, little thought is given to how estimates in tables or bars in graphs are ordered other than how they appear in the raw data. As graph producers, however, we should take a more active role in choosing how to order and present data values for different groups.

There is a historical legacy from which this sort of ordering and implied racial hierarchy derives. Many of the largest demographic surveys conducted in the United States order race starting with "white" and "Black" as the first two options. The images in figure 7 are the questions about race from the decennial census conducted by

the US Bureau of the Census, the Survey of Consumer Finances from the Federal Reserve Board, the Panel Study of Income Dynamics from the University of Michigan, and the Census Bureau's Current Population Survey. Each lists "white" as the first option and codes that group with a "1" in the data. No wonder data analysts use it as the default option and thus the norm.

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

FIGURE 07

Examples of survey questions about race in four major US federal government surveys illustrating how "white" is presented as the first response option.

<p>9. What is Person 1's race? Mark <input checked="" type="checkbox"/> one or more boxes AND print origins.</p> <p><input type="checkbox"/> White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc. ↵ <input type="checkbox"/> Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc. ↵ <input type="checkbox"/> American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc. ↵ <input type="checkbox"/> Chinese <input type="checkbox"/> Vietnamese <input type="checkbox"/> Native Hawaiian <input type="checkbox"/> Filipino <input type="checkbox"/> Korean <input type="checkbox"/> Samoan <input type="checkbox"/> Asian Indian <input type="checkbox"/> Japanese <input type="checkbox"/> Chamorro <input type="checkbox"/> Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc. ↵ <input type="checkbox"/> Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc. ↵ <input type="checkbox"/> Some other race – Print race or origin. ↵</p>	<p>IN PERSON VERSION: (SHOW CARD 20) Please look at this card. Which of these categories do you feel best describe you: (white, black or African-American, Hispanic or Latino, Asian, American Indian or Alaska Native, Hawaiian Native or other Pacific Islander, or another race?)</p> <p>TELEPHONE VERSION: Which of these categories do you feel best describe you: white, black or African-American, Hispanic or Latino, Asian, American Indian or Alaska Native, Hawaiian Native or other Pacific Islander, or another race?</p> <p>CODE ALL THAT APPLY: CODE RESPONSES IN THE ORDER THEY ARE GIVEN</p> <p>1. *WHITE (INCLUDE MIDDLE EASTERN/ARAB WITH WHITE); Caucasian 2. *BLACK/AFRICAN-AMERICAN 3. *HISPANIC/LATINO 4. *ASIAN 5. *AMERICAN INDIAN/ALASKA NATIVE 6. *NATIVE HAWAIIAN/PACIFIC ISLANDER -7. *OTHER 0. Inap. (/no further responses)</p> <p>***** FOR THE PUBLIC DATA SET: ONLY X6809 AND X6810 ARE INCLUDED. FOR X6809, CODES 4, 5, AND 6 ARE COMBINED WITH CODE -7. IF AN ADDITIONAL RESPONSE WAS GIVEN IN X6810-X6814, X6810 IS SET TO 1; OTHERWISE X6810 IS SET TO 5. *****</p> <p>RACE</p> <p>I am going to read you a list of five race categories. You may choose one or more races. For this survey, Hispanic origin is not a race. (Are/Is) (NAME/you) White; Black or African American; American Indian or Alaska Native; Asian; OR Native Hawaiian or Other Pacific Islander?</p> <p>*Do not probe unless response is Hispanic or a Hispanic origin Enter all that apply, separate with commas</p> <p>1. White 2. Black or African American 3. American Indian or Alaska Native 4. Asian 5. Native Hawaiian or Other Pacific Islander 6. Other - DO NOT READ</p>						
<p>L7. What is your race? Are you white, black, American Indian, Alaska Native, Asian, Native Hawaiian or Other Pacific Islander? • ENTER all that apply • PROBE: Any others? • If R says something not listed in response options, record under [7. Other] and list response verbatim at "specify screen"</p> <table border="1"><tr><td>1. White</td><td>2. Black, African-American, or Negro</td></tr><tr><td>3. American Indian or Alaska Native</td><td>4. Asian</td></tr><tr><td>5. Native Hawaiian or Pacific Islander</td><td>7. Some other race-specify → L7OS. Specify (String 500)</td></tr></table>	1. White	2. Black, African-American, or Negro	3. American Indian or Alaska Native	4. Asian	5. Native Hawaiian or Pacific Islander	7. Some other race-specify → L7OS. Specify (String 500)	
1. White	2. Black, African-American, or Negro						
3. American Indian or Alaska Native	4. Asian						
5. Native Hawaiian or Pacific Islander	7. Some other race-specify → L7OS. Specify (String 500)						

Sources: Decennial census from the US Census Bureau (top-left); Survey of Consumer Finances from the Board of Governors of the Federal Reserve System (top-right); Panel Study of Income Dynamics from the University of Michigan (bottom-right); and Current Population Survey from the US Census Bureau (bottom-left).

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. Again, starting with "white" or "men" can make it seem as though those are the most important groups we are trying to communicate with. A few issues are worth considering to help guide the decision of how to order different racial and ethnic groups:

- 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?

Interactivity is another potential way to address the problematic aspects of choosing an order for your charts or tables. Enabling users to select which groups they want to see in a chart by, for example, toggling the lines that appear in a line chart or selecting which race appears as the baseline in a stacked bar chart, can empower users to focus on the groups they are most interested in or customize which comparisons they want to make. Instead of imposing a static view of a data visualization, allowing interactivity can help users find and explore the stories most relevant to them.

08 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 show white people, Black people, and Hispanic/Latino people but not other racial or ethnic groups. Many charts also present data at the level of these broader racial groupings rather than at a more disaggregated level. Often this is because of data limitations, usually small sample sizes or no data at all. But even in such cases, how can researchers 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.

Groups whose data are and are not collected or shown often reflects who society deems as most important or valuable. An example of this can be found in cartography, where maps of the United States often do not display US territories, erasing the populations that live there (Dougherty and Ilyankou 2021). Not only do the citizens of those areas not have any political representation in Congress (or the right to vote for president), these areas also have majority-nonwhite populations. Is it a coincidence that roughly 3.5 million people are not represented in these visualizations of the US? As mentioned in section 5, omitting certain populations

from the data can have important ramifications for what products or public policies are created and who benefits from those products and policies.

We can think about missing data even more broadly than data we have but do not include in our charts, graphs, and diagrams. Artist and researcher Mimi Onuoha's *The Library of Missing Datasets* project identifies "missing datasets," or entire categories of data and information that do not exist but perhaps should. Onuoha (n.d.) writes, "That which we ignore reveals more than what we give our attention to ... Spots that we've left blank reveal our hidden social biases and indifferences."

There are instances, however, in which collecting data on marginalized groups can be harmful, especially when done without their consent. Doing so can risk exposing people with adverse consequences for their livelihoods and safety (see our earlier note about data confidentiality and security). Some groups can also be overrepresented in datasets. In some instances, this is a purposeful (and useful) factor in conducting a survey that seeks to better represent small groups and the variation within those groups. In other cases, however, such overrepresentation can be harmful, such as when parents with low incomes were overrepresented in a dataset that was used as an input to predict the risk of child abuse in Allegheny County, Pennsylvania (Eubanks 2018).

In thinking about missing groups, we refer to four specific situations: lumping or splitting groups; using nonbinary gender categories; using catch-all groups often labeled as “other”; and choosing to not include all groups.

LUMPERS AND SPLITTERS

For researchers and analysts, a big issue when deciding which groups to include in an analysis or show in a data visualization is the underlying sample size. When there are “too few” observations, populations may be lumped together to make analysis more convenient.

Doing so, however, can have harmful effects on the communities that are lumped together. The UCLA Center for Health Policy Research conducts outreach and works with Asian American, Native Hawaiian, and Pacific Islander (NHPI) communities. Ninez Ponce told us that “It’s a disservice to NHPIs to aggregate to an [Asian American and Pacific Islander] group because the generally better stats of Asians hide the vulnerabilities of NHPIs. That actually could be harmful for the Native Hawaiian Pacific Islander community.”

As a relatively simple example of how aggregating racial groups can mask important variation, we looked at the 2019 poverty rate across 139 detailed race categories recorded in the Census Bureau’s American Community Survey, an annual survey that helps determine how more than \$675 billion in government funds are distributed each year (US Bureau of the Census, n.d.).

The official poverty rate in the United States stood at 12.3 percent in 2019 (Benson 2020). Within that overall estimate, different racial and ethnic groups vary significantly. The primary question in the American Community Survey concerning race consists of 15 separate checkboxes; for some of these, interviewees are able to print additional names or races (a separate question refers to Hispanic or Latino origin). Digging slightly deeper into the data presents a more nuanced picture of poverty in the United States.

The dots in figure 8 show our estimated poverty rates for all 139 detailed racial groups from the American Community Survey, along with the overall poverty rate for major racial groupings frequently used in analyses (Hispanic/Latino is not shown in our chart because it is asked in a separate question about ethnicity). Within

the “American Indian or Alaska Native” category, for example, poverty rates vary from 5.9 percent for the Aleut (an indigenous community primarily living in Alaska) to 36.9 percent for those who identify as part of the Sioux Native American tribe (living primarily in Nebraska, North Dakota, and South Dakota). Within the “Asian or Pacific Islander” category, which is no longer a specific category in the American Community Survey but is often used in data analysis, we find a range of poverty rates, from 4.5 percent for those who identify as both Chinese and Japanese to 27.8 percent for those who identify as Mongolian. Even within the Pacific Islander category, we find a range of poverty rates, from 6.3 percent for Fijians to 24.6 percent for people who identify as more than one Micronesian race. We could tell similar stories about the more than 30 American Indian and Alaska Native tribes identified in the data, an issue that has raised calls for groups to take greater ownership of their own data (Rodriguez-Lonebear 2016). (It is also worth noting that interviewees from the NHPI Data Policy Lab stated that there is broad consensus among Asian American and NHPI organizations to retire the phrase “Asian and Pacific Islander” to help support more inclusive and accurate assessments of people in the United States; see Ishisaka 2020.)

A main reason more analyses don’t include more racial groups is likely sample size. In the data we used here, the number of people in poverty ranged from around 100 people for some groups to tens of thousands for others. Sample size limitations are a function of multiple factors, including the actual size of the group, staffing, and funding as well as the active choices made by the surveying organization, survey funders, and society. Heather Krause, founder of We All Count, noted in our interview that small sample sizes for groups that are already underrepresented is a “choice on the part of the privileged, not limitations that are inherent in small populations.” Thus, the burden to do a better job collecting data that reflects the lives of all people should not fall on already-marginalized communities but on the surveying and research institutions, which should strive for better representation of such groups in the full dataset.

MALE, FEMALE, IN BETWEEN, AND BEYOND

Sometimes data are not collected at all about certain populations. Many major federal surveys, for example, do not offer “nonbinary” or “transgender” as response options when asking about gender. In such situations, data consumers should make it clear to their reader what information they have and do not have in their data. Lauren Klein said she includes a note in her work when certain groups are not included in the data, such as nonbinary people: “Just by calling out the absence and by doing it explicitly, it gives a prompt to the reader who may be the next survey designer to be like, ‘Maybe I should think about nonbinary gender.’” The entire ecosystem of data producers, consumers, and communicators should work in tandem to be more inclusive and remind each other of how these different aspects can be improved.

In January 2021, Billy Jacobson (also known as Anna Lytical, @theannalytical on Twitter) asked followers to submit “the worst gender selection forms.” A selection of those submissions are shown in figure 9 and demonstrates the challenges (and sometimes absurd results) websites and groups conducting surveys have in asking about missing groups. Although we are primarily

focusing on racial equity in this paper, we raise this issue around identifying gender to highlight both the challenges of missing data and the intersectionality between different characteristics that can raise even greater challenges to people collecting, analyzing, and communicating data.

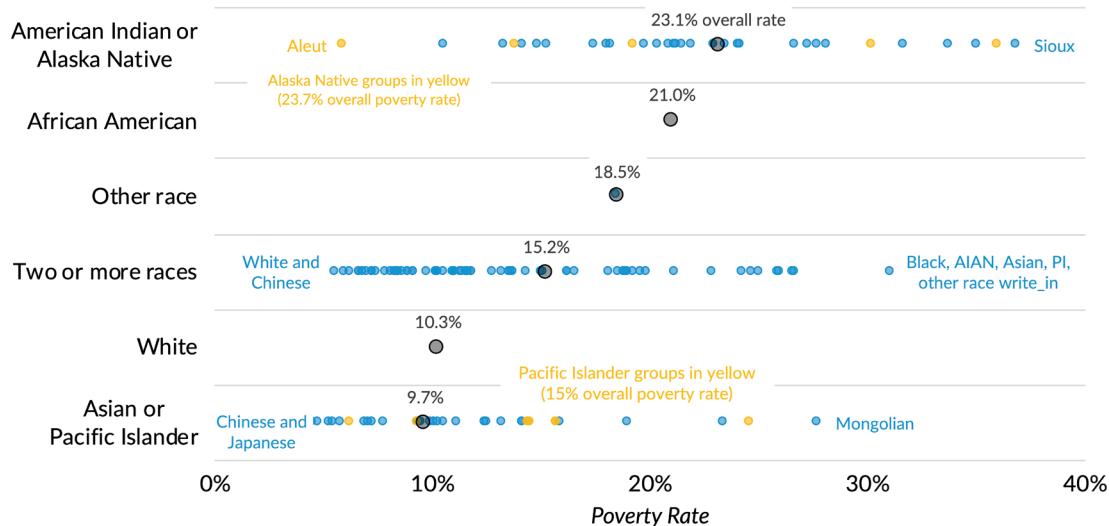
This concept of missing data raises two further questions: First, do data collectors always need to ask for every piece of information? There seems to be a perspective that it is better to collect all possible pieces of information just in case additional research will be conducted later on. But asking for all this information may not be necessary. In a July 2020 blog post in *Health Affairs*, Rhea Boyd and coauthors recommend that health researchers “define race during the experimental design, and specify the reason for its use in the study” (Boyd et al. 2020). It may not be necessary to include race, gender, or other demographic variables in every data collection effort, project, or visualization.

As a personal example, the children of one of this guide’s authors (Schwabish) were participating in a virtual movie event with other kids during the COVID-19 pandemic.

FIGURE 08

Poverty rates across different racial groups when disaggregating data from the American Community Survey reveal variation that is missing when metrics are only presented for overall groups.

Variation in Poverty Rates



Source: Schwabish and Feng (2021).

Just by calling out the absence and by doing it explicitly, it gives a prompt to the reader who may be the next survey designer to be like, "Maybe I should think about nonbinary gender."

- LAUREN KLEIN

Kids would register online and go to a curbside pickup location for bags of popcorn and hot chocolate. The registration form included basic information, such as the student's grade and parent's names, but it also provided a space for gender, which had no bearing on the event (figure 10). Website developers and data collectors should ask themselves whether collecting this information is necessary.

Second, products and research authored by more diverse teams can contribute to creating better and more inclusive products and projects because of the variety of perspectives and lived experiences team members bring with them. Such diversity can help survey and product designers identify gaps in the way they are collecting data. Perhaps the most famous example of how designing for a specific group can help everyone are the small ramps built into the curbs of sidewalks. The ramps

were initially an effort by disability advocates to assist people in wheelchairs, but they also serve parents with strollers, travelers with luggage, and workers pushing carts. This "curb-cut" effect shows how designing for one group can help many others (Blackwell 2017). Similarly, collecting inclusive data and building inclusive tools and visualizations can make the experience better for all users.

OTHERING THE OTHER

A topic of particular interest for many people collecting and working with data is how to treat the "other" category. There are two primary issues here: First, what is the right approach to analyzing and communicating about the "other" category? Second, can we use more inclusive language to talk about this group?

On the first issue, we need to first understand what the "other" category is actually measuring. Depending on the data, the "other" category may include people with a wide array of races, ethnicities, religions, and other characteristics. (This is similar to the data lumping issue mentioned earlier, except here we do not have data that captures specifics about the individual and cannot disaggregate this category further.) Although "other" can be an important option for survey respondents who do not see themselves reflected in the usual racial

FIGURE 09

Selection of gender drop-down menus with questionable response options.

The figure displays four examples of gender selection interfaces:

- * Gender:** A dropdown menu with options: Female, Male, N/A, Unknown, and Tax Entity. The "Male" option is highlighted.
- Does the Beneficiary reside a** [redacted] *** Contact Phone:** A dropdown menu with options: Female, Male, N/A, Unknown, and Tax Entity. The "Male" option is highlighted.
- GENDER** [redacted] **BIRTHDATE** [redacted] **LANGUAGE** [redacted]
A table with three rows:
 - Row 1: GENDER [redacted] Female
 - Row 2: BIRTHDATE [redacted] Male
 - Row 3: LANGUAGE [redacted] Female
HumanThe "Male" option under GENDER is highlighted.
- Spouse's gender.***
Three radio button options:
 - Female
 - Male
 - IndeterminateThe "Male" option is selected.

Source: Coding Drag Queen Anna Lytical (@theannalytical). 2021. "send me the worst gender selection forms you've seen, I'll start." Twitter, January 13, 2021, 11:25 a.m., <https://twitter.com/theannalytical/status/1349392166716657664>.

categories, we should question whether grouping people with such varied characteristics and life experiences together is actually meaningful. Although including all observations in the data sample is important for the math and statistical analysis, it may not be particularly useful in terms of policy solutions, as Calvin Chang from the UCLA team noted:

In terms of whether that would be useful at all, in terms of the organizations that you would expect to use a data report. It won't be useful for them to know what that coefficient is, or being able to dive any deeper into that "other" category, if their community is in the "other" category. So I can see it being part of the calculations, but in terms of the data being presented, it's just simply not going to be of much use. I could never go to a policymaker as a member of the NHPI community and say, "We need resources because the 'other' category doesn't have resources."

Regarding the issue of more inclusive language, it is important to recognize that the term "other" literally others individuals, emphasizing how they differ from the norm. The word "other" itself can be perceived to have negative

FIGURE 10
Example form that asks for gender even though it is not relevant.

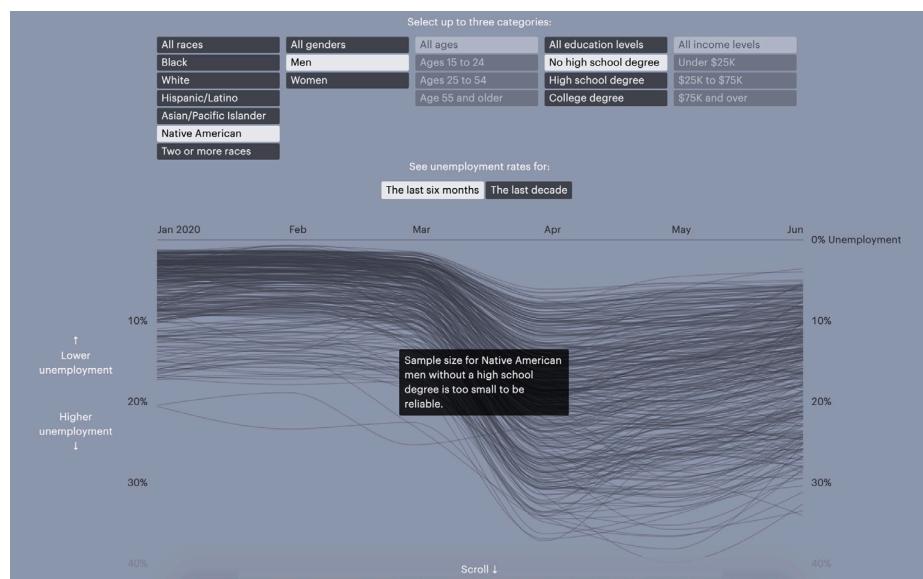
Source: Original creation by authors in likeness of actual web form.

connotations. One meaning of "other" in the Merriam-Webster dictionary is "disturbingly or threateningly different," clearly evoking a negative or dangerous association.

Moving away from the word "other" is one step data communicators can take. Through our various discussions, we have identified six possible alternatives to "other":

1. Another race
2. Additional groups
3. All other self-descriptions

FIGURE 11
A message box explaining how sample size limitations did not allow for reliable estimates.



Source: "What Coronavirus Job Losses Reveal about Racism in America" (Groeger 2020).

4. People identifying as other or multiple races
5. Identity not listed
6. Identity not listed in the survey

Some of these terms are more verbose and may not fit as nicely in a table or under a bar in a bar chart, but they are more inclusive and avoid some of the issues noted.

As discussed in section 6, some may worry that these labels are not the exact words or phrases available to respondents in the survey. That is, the survey participant checked the box “other” because that was what was available. But we believe these alternatives are, in the general sense, still true and accurate; a separate note or footnote could be added to the graph, table, or text to make the change clear to readers.

CHOOSING TO NOT INCLUDE ALL GROUPS

In cases where data were collected about a specific group but that group was not presented in the chart or included in the analysis, it may be worth listing in the notes section

all racial and ethnic groups included in the original dataset. This can both inform readers that data exist for these smaller groups (even though they may be limited) and make the chart-maker’s decision about groups they included in their visualization transparent.

ProPublica took another approach in its piece “What Coronavirus Job Losses Reveal about Racism in America” (Groeger 2020), a data tool that allows users to explore disparities in unemployment rates by race, gender, age, education, and income (figure 11). Some user selections would have produced sample sizes that are too small (such as Native American men without a high school degree). Rather than excluding those options from the tool and thus the population this combination of selections represents, ProPublica displayed a message over the chart explaining how the lack of sufficient data would result in estimates that were too imprecise to be reliable. This approach allowed ProPublica to be inclusive while not misinforming its readers.

09 USING COLORS WITH A RACIAL EQUITY AWARENESS

Good color palettes for data visualization should, at minimum, meet basic accessibility guidelines (WebAIM 2020) and offer sufficient contrast between colors for readers with vision difficulties. Going beyond accessibility, color choices should also avoid reinforcing gender or racial stereotypes, such as baby pink and baby blue to represent women and men or colors associated with skin tones or racial stereotypes (e.g., black to represent Black people, yellow to represent Asian people).

A data visualization that does not take a racially equitable approach to using colors is the Massachusetts Institute of Technology Office of the Provost’s “Diversity Dashboard” (MIT, n.d.), which enables users to explore the demographic characteristics of the school’s students, faculty, and staff. The June 2020 version of the dashboard used three distinct hues to represent nine racial and ethnic groups (figure 12). Six groups were represented with shades of red; the “international” and “unknown” groups were two shades of gray; and the “white” group was the only one in blue.

This design presents many problems. First, the shades of red that are used to represent the six groups of students of color create a visual divide that seems to pit students of color against white students. Second, a graduated color palette should not be used for categorical data. A graduated color palette shows greater or higher values in darker colors and smaller or lower values in lighter colors. Thus, the graduated palette here appears to suggest that “Black or African American” students are somehow

FIGURE 12

Legend showing a problematic color scheme applied to data on race and ethnicity.



Source: Recreated based on the June 2020 version of the *Diversity Dashboard* from the Massachusetts Institute of Technology, Office of the Provost.

"more" or "higher" than students who identify as "two or more races." Instead, separate colors should be used for each of the six categories. Third, because gray colors tend to fade to the background (Schwabish 2021; Shirley 2020), this color choice diminishes international students and students whose race or ethnicity is unknown. All these design decisions create an effect where the "white" category moves to the foreground and is highlighted, as if that is the most important group and norm with which all other groups should be compared.

About a month after it was launched, the dashboard was updated but not dramatically improved (figure 13). Here, we see a blue color palette used for four student groups, a dark green representing Black or African American students, and a light purple for the "two or more races" category. White and international students now have similar orange/red colors, and the "unknown" category sits by itself to the side in grey.

Many of the same issues persist in this redesign. The use of graduated color palettes still suggests a hierarchy among races, and the dark green and light purple are close enough to the blue palette to still be perceived as being part of a singular color "ramp." The similar orange/red colors for white and international students also seems to imply a relationship between these two groups.

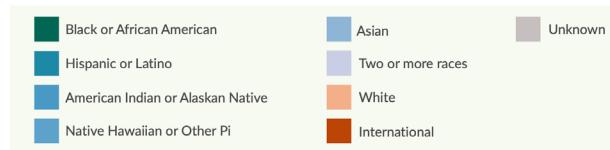
It is not necessarily easy to pick out nine separate colors, but many free color tools are available for designers and developers. The palette in figure 14, for example, was chosen from the free ColorBrewer tool in less than a minute. Another possible approach would be to use a series of different graphs (the small multiples or panel chart approach) instead of putting all the data in a single view, which we discuss in more detail in section 14.

Another aspect of color to be aware of is the emotional connotations associated with certain hues. In Western

cultures, colors such as red can be perceived as threatening or aggressive and therefore can paint the population we're visualizing in a negative light. In a map of participation in employment services for those receiving welfare benefits in Minnesota, participants were represented as small red dots. When the participants who were being visualized in the map were shown that map, they felt that the design presented them as a "threat" and made them look like an "infestation" (Evergreen, n.d.). The implication of this visual approach is that the problem lies with the people rather than the legacy of racist institutions and policies. Again, it is important to ask, "If I were one of the data points on this visualization, would I feel offended?"

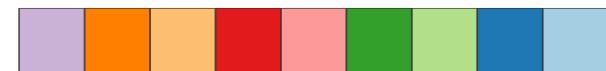
In sum, as data visualization producers, we need to be aware of how our use of colors, words, and categorizations can perpetuate or exacerbate inequities and stereotypes.

FIGURE 13
Legend with the color scheme modified (though still problematic).



Source: Recreated based on the July 2020 version of the *Diversity Dashboard* from the Massachusetts Institute of Technology, Office of the Provost.

FIGURE 14
Example color palette that avoids gradients and hierarchies.



Source: Created with the ColorBrewer online tool.

10 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, ethnicities, ages, and other characteristics. When using icons, we should consider to whom we are presenting

our results, anticipate how our icons might be perceived, and ensure they do not perpetuate harmful and offensive stereotypes. 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.

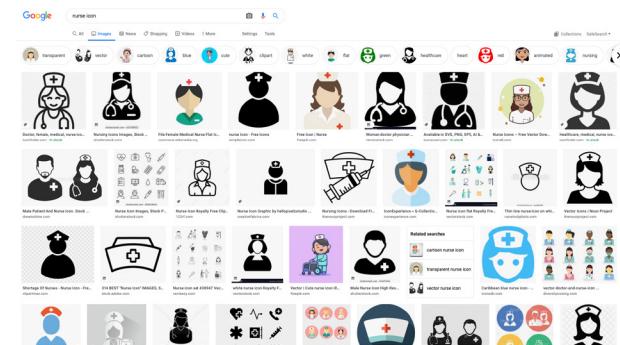
Mis- or underrepresentation of certain groups in imagery and iconography can fail to take a racial or gender equity awareness perspective toward how we visualize our data. A 2018 study by the Pew Research Center found that “men are overrepresented in online image search results across a majority of jobs examined” and that “women appear lower than men in such search results for many jobs.” Similarly, scholar Safiya Noble (2018), in her book *Algorithms of Oppression*, records a racist and sexist culture in our online searches and other actions on the internet. These disparities continue to this day; the top image in figure 15 shows Google search results for the phrase “nurse icon,” and the image on the bottom shows results for the phrase “boss icon.” Notice how nearly all the images on the top we might imagine as feminine and the images on the bottom we might traditionally consider masculine.

When choosing icons or imagery to represent racial or ethnic groups, avoid images that depict or reinforce stereotypes associated with those groups, such as African Americans in poverty or American Indians in traditional headdresses on horseback (Kim 2019). Pictures also should not show people in situations where they appear as helpless victims (so-called “poverty porn”; Dortonne 2016) or reinforce power hierarchies (e.g., a white male supervisor with a person of color as a subordinate). Instead, images should represent people as active and empowered and reinforce their dignity, agency, and humanity. (A separate but related issue is how to use emojis with different skin tones, which has spurred all sorts of discussions and debates about whether the original, default yellow skin tone represents a white default; see McGill 2016.) And when possible, try to solicit feedback about the visuals used to make sure they are not offensive and to identify problematic cultural associations that were missed.

Another factor to consider when using icons or shapes that resemble people and communities is when they might imply something untrue about the data. The 2019 ProPublica story “What Happened to All the Jobs Trump Promised?” (Porat, Groeger, and Arnsdorf 2019) used

icons of people to represent the number of jobs President Trump claimed to have created (figure 16). In designing this story, Lena Groeger and her colleagues debated whether it was appropriate to include an icon with a person in a wheelchair. Including such an icon could be seen as representing people with disabilities and making them more visible. But readers might also interpret the icon to mean that the people in those jobs were disabled when that might not be true. In the end, Groeger and her colleagues decided not to include the icon of the person in a wheelchair. Ultimately, one should exercise caution when using highly specific icons.

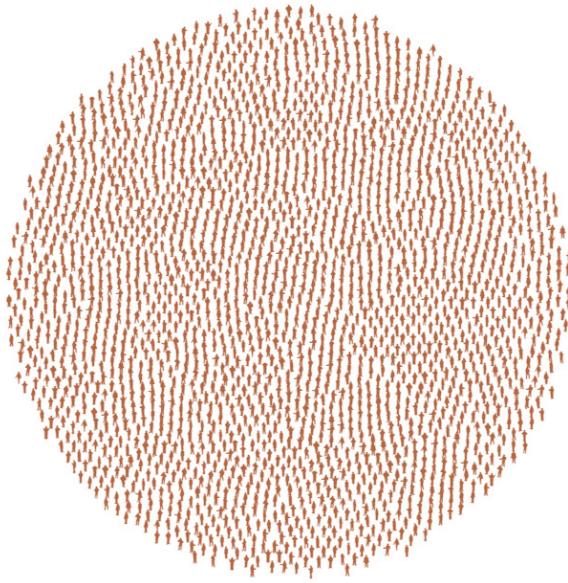
FIGURE 15
Image search results showing problematic stereotypes and biases.



Source: 2021 Google image search of “nurse icon” and “boss icon.”

FIGURE 16

A chart that uses icons representing people.



(Each person icon represents about 4,000 jobs)

Source: "What Happened to All the Jobs Trump Promised?" (Porat, Groeger, and Arnsdorf 2019).

Since the election, Trump has made 35 claims that companies would create 8.9 million jobs in the U.S. thanks to his policies and actions. Some of the new jobs he touted failed to materialize. Many of the new jobs he took credit for were, in fact, planned before he took office.

11 EMBRACING CONTEXT AND COMPLEXITY

**If I only saw this chart on Twitter,
would I draw the correct conclusion?**

- KIM BUI

As mentioned in section 6, it is important that data visualizations acknowledge the context and history surrounding the topic being shown. Structural racism, historic discrimination, other barriers and inequities, and the mechanisms by which they may be operating should be woven into the framing of the research and the chart (Schwabish and Kijakazi 2021). Visualizations should not let the data "speak for themselves": data are not neutral or objective, so data visualizations are not neutral either (as we saw with the example of the HOLC redlining map in the introduction). When data visualizations are not critically examined and thoughtfully designed within a DEI framework, they tend to express the viewpoint of the dominant group in society while, as Catherine

D'Ignazio and Lauren Klein (2020, 76) write in their book *Data Feminism*, "mask[ing] the people, the methods, the questions, and the messiness that lies behind clean lines and geometric shapes." When weaving in context, data communicators should have diverse references and citations, particularly from scholars of color or scholars who are members of the community of focus, and they should lift up the voices of the communities they worked with, highlighting lived experience where possible.

Data communicators and visualizers also should not shy away from embracing complexity and nuance in their visuals. Even though simple charts and the idea of having a high "data-ink ratio" (Tufte 1985) are often lauded as goals in data visualization, too much simplicity can strip away necessary context and the inherent complexity of many social issues that cannot be boiled down to a simple bar chart. As Kim Bui put it, "If I only saw this chart on Twitter, would I draw the correct conclusion?"

Data visualizations should use more complex designs if that would more accurately reflect and promote a better understanding of the topic being shown.

Complex designs can also foster better audience connection and engagement with the data. In her articulation of "Data Humanism," information designer Giorgia Lupi argues that dense, custom-designed data visualizations presenting multiple layers of information can encourage more careful reading, better personal connection, and a deeper understanding as readers are offered various paths to explore the data. Using a more personal approach to visualizing data, one that incorporates qualitative information, expresses subjectivity, and acknowledges uncertainty can "unlock [data's] profound nature and shed light on its real meaning for ourselves and others" (Lupi 2017). Data visualizations should be clear and informative while not oversimplifying the complex world we live in.

It is worth noting that not all data needs to be visualized; not every graph needs to be created. Data communicators should consider whether their instinct to create a graph, chart, or diagram is warranted. Sometimes a simple number or set of numbers can be more effective than a graph. And sometimes—perhaps many times—data visualizations are unable to convey the complexity of the

underlying data or the bias inherent in the data or the visuals. As Sarah Williams (2020, 184) writes in her book, *Data Action*, "Sharing data through visual interpretation might make underlying policies more transparent, but all data visualizations hold the bias of the creator." It is important to remember that data are a reflection of the lives of real people, not just a sterile abstraction.

Further, charts simply may not be as effective for forming an empathetic connection to the subject as other visual or storytelling methods (Correll 2019). In a story on families with children who have a fatal disease, Sarah Slobin and her colleagues at the *Wall Street Journal* at first tried to visualize data on an index that tracked the progression of the disease (Marcus 2013; Slobin 2014). After struggling to find a compelling way to visualize these data and the children affected by it, Slobin and her team realized that charts were not necessary after all and that simply showing photos of children diagnosed with these fatal diseases was sufficient, perhaps even more compelling: "When you looked at the kids—really looked at them, even though it was hard to look at them—you understood what the data was capturing, how the disease progressed, what it wrought. And you saw these beautiful little people and you understood what the Severity Index was all about without a chart or a visualization or an explainer" (Slobin 2014).

12 BUILDING DIVERSE AND INCLUSIVE DATA COMMUNICATION TEAMS AND ORGANIZATIONS

Across most of our interviews, we heard that diverse teams, in terms of demographics, life experiences, skills, and subject matter expertise, are critical to creating communication products that apply a racial equity lens and fight other forms of oppression. Diverse teams can help identify biases and make connections between different fields of study whose relevance may not be evident at first glance. They can also better reflect the demographics of the populations they wish to study. However, having diverse teams is not in itself sufficient. The internal culture of the organization should also value, promote, and embody DEI so that the communication products that are created uphold such values. The concept of building teams rather than leaving the

responsibility in the hands of individuals is a primary theme that emerged from our conversations.

We can distill our conversations around building diverse and inclusive organizations into four main areas:

1. A ground-up approach and buy-in from leadership go hand in hand.

In many organizations, change first occurs because an individual or small group of people see an opportunity to improve or evolve. In his work at the *Washington Post*, Aaron Williams saw this firsthand: "I think that there were reporters there first who were like, 'I'm going to start doing this in my work,' and then editors who got behind their reporters and then it started bubbling up." Kim Leary echoed this point, saying

that individuals can ask questions such as "Why not?" and "How come?" challenging existing institutional norms and bringing creativity and innovation needed for fresh thinking. Though efforts may start at the individual level, buy-in and support are eventually needed from leaders and management to ensure that change is implemented throughout the institution.

2. Processes for incorporating equity and inclusiveness need to be codified for change to happen.

Working toward a goal of equity and inclusiveness is fine at the conceptual level, but concrete, actionable guidelines need to come out of such discussions for organizations to successfully implement changes to their culture and the way they work. Guidelines must be codified (through documentation, checklists, guidelines, requirements, and testing) and made available to all members of the organization so everyone is following a standardized process.

3. Guidelines should be actionable and instill accountability.

Ayesha McAdams-Mahmoud from Salesforce told us about her experiences building toolkits and resources for product developers to implement in order to build accessible and equitable tools. Those guidelines needed to include concrete steps that could be taken and "baked into the product development process" for them to be useful to developers. Those toolkits included measures of accountability, which are not necessarily punitive actions but rather guidelines and metrics that imbued a sense of responsibility. It is important at this step for organizations to also examine the official and unofficial practices and norms that incentivize or reward people for doing (or not doing) certain things.

4. Updating and reevaluating norms should have a regular cadence.

Ongoing checks, team conversations, and retrospective analyses can all be ways to ensure products and projects are taking a racial equity lens and instituting organizational guidelines and norms. Guidelines and norms themselves should be updated and reevaluated periodically as well (Chartoff 2018).

Cultural shifts of this magnitude need to happen at both the organizational level and the individual and interpersonal levels. McAdams-Mahmoud and others talked about the importance of working with members

of the community when it comes to external data communication, but that theme also emerged when we discussed creating cultural change within an organization. From McAdams-Mahmoud's perspective, involving individuals at all levels helps drive home a sense of ownership among the entire organization: "[We want to] make sure we're bringing people along before we institute these processes. Our hope is that it's not a surprise to anybody once these checklists come down the barrel because we've already said, 'Hey this is important. We've had conversations about it.'"

The question of making cultural change then becomes, "How do we persuade people to do something, especially something that could be viewed as burdensome or extra work?" How to effect cultural change, change habits, and help businesses evolve is beyond the scope of this work, but three examples of major public campaigns from the past 50 years show how it can be done:

- Antismoking campaigns
- "Don't Mess with Texas," a campaign to reduce litter in the state
- "Click It or Ticket" and other efforts to encourage Americans to wear seatbelts

Ongoing antismoking campaigns (Durkin, Biener, and Wakefield 2009), along with changes in smoke-free laws and changing prices for cigarettes (American Heart Association 2018), have contributed to significant declines in the percentage of people who smoke. The original "Don't Mess with Texas" slogan was created by the Texas Department of Transportation to reduce litter and used TV commercials featuring famous Texans as well as bumper stickers and other merchandise to spread the message. And it worked: between 1986 and 1990, road litter declined 72 percent (Nodjimbadem 2017). Finally, the "Click It or Ticket" campaign, which started in 1993 in North Carolina and has since expanded nationwide, is credited with increasing seat-belt usage and saving thousands of lives (Solomon, Compto, and Preusser 2004; Tison et al. 2010). When it comes to persuading individuals and organizations to change, varied approaches are needed: both "carrots" that help individuals see the benefits of such change and "sticks" that increase the cost of not adopting these changes.

13 THE FEEDBACK LOOP

In section 4, we noted the importance of working closely with and empowering the people and communities our research focuses on. This includes using a variety of research methods—such as interviews, focus groups, and long-form surveys—for a given body of work. Another approach is to solicit feedback from communities and audiences about our work. This can allow researchers and communicators to engage with audiences, understand how their work is being used and what they may have missed, and continue to develop and build relationships with communities. Across our discussions, we saw various approaches that different organizations took to soliciting feedback.

ProPublica. The ProPublica newsroom has an engagement team that works with reporters to communicate with readers, commentors, survey respondents, and more. These team members will sometimes be responsible for fielding calls and comments and will work directly with reporters on specific stories. ProPublica also often includes a form at the end of certain stories that provides readers with an opportunity to share their experiences or provide more information on how an event or policy affected them, as well as whether they are involved in the issue in some way.

UCLA Center for Health Policy Research. The UCLA team has several internal and external ways to receive feedback on data projects. Internally, they have a significant review process: team members who identify as different races and ethnicities review each product they produce, including infographics and other standalone visual content. Externally, the UCLA team has invested time in building relationships with the people and communities that are in their content focus areas. For example, when developing a survey for the Native Hawaiian and Pacific Islander community, the team created posters and flyers to generate awareness of the survey and asked members of the community to review the products. They also highlighted that the diversity of their own team helps them build connections and develop relationships with community groups who can then see themselves reflected in the research team. In our interview, they noted they have found it essential

to implement community-based participatory research principles (discussed also in section 9), which emphasizes incorporating both input and feedback loops from community organizations and leaders from the initial planning and design phase through implementation and dissemination.

Evan Peck. Evan Peck and his team at Bucknell University applied the concept of “reciprocal research” to how they conduct their projects. Reciprocal research (Hanrahan et al 2020; also discussed in section 9) is the idea of trying to create immediate, concrete benefits project participants can receive while they are participating in the study rather than waiting until the project is over before seeing how such work helps them. As Peck states, “You aren’t just getting data for your sake and something that might benefit the community back in 15 years, but you start thinking a little bit more short-term or how can you provide benefit back to the community as you’re collecting this data ... it’s how to be more deliberate in making sure we’re not unintentionally exploiting these people’s experiences in ways that don’t benefit them.” In applying such an approach, researchers may need to be flexible and responsive with their plans, pivoting to a new approach when they realize their original ideas do not meet the needs of the study’s participants and community.

You start thinking a little bit more short-term or how can you provide benefit back to the community as you’re collecting this data ... it’s how to be more deliberate in making sure we’re not unintentionally exploiting these people’s experiences in ways that don’t benefit them.

- EVAN PECK

Institute of Governmental Studies at the University of California, Berkeley. G. Cristina Mora described what we perceive to be a more common approach in the academic fields of having a faculty advisory board who is responsible for reviewing certain projects for equity

and inclusion. As Mora described, the Berkeley advisory board consists of about 20 faculty across campus and is diverse in race, ethnicity, and gender. Mora noted that participating in these kinds of advisory positions can be an extra burden on faculty members' time, so compensation and consideration of that time is central: "You have to compensate them. Faculty of color are so taxed already ... We get most of the graduate students of color. We sit on almost every diversity committee that

exists on campus, and we still have to do our own work ... so you [have to] have creative ways of making this not be an additional tax on them and their time."

There are many ways to solicit and receive feedback, but the overarching themes in these and other conversations is to (1) build relationships with people and communities over time and (2) value people's time and compensate them accordingly.

14 THE GATEKEEPERS

Most data communicators, especially researchers and analysts, rely on funders for financial support to perform their work and on peer-reviewed publications to communicate the findings of their research. Thus, gatekeepers—funders, project officers, editors, peer reviewers, federal agencies, and others—can play an important role in enforcing a racial equity lens in the work they fund and publish because of the amount of influence they have on researchers' careers. Here, we discuss how three of these gatekeepers can help promote racial equity across the data ecosystem.

EDITORS AND PEER REVIEWERS

Attributing findings to structural racism is not yet common in scholarly writing for many fields. Regarding economics, Howard University professor William Spriggs wrote, "The fact that far too many economists blindly agree that negative attributes correlate to being African American and cannot see that relationship to police officers assuming all Black men are criminals is stupefying. The fact that a discipline that prides itself on being objective and looking for data to test hypotheses fails to see how negative attributes do not correlate with being African American is a constant irritant for Black economists" (Spriggs 2020). In a July 2020 blog post in *Health Affairs*, Rhea Boyd and coauthors wrote that "despite racism's alarming impact on health and the wealth of scholarship that outlines its ill effects, preeminent scholars and the journals that publish them, including *Health Affairs*, routinely fail to interrogate racism as a critical driver of racial health inequities" (Boyd et al. 2020). Researchers and scholars need to do a better job of examining how racism impacts their findings, and

journal editors and reviewers can play an important role in demanding this happens (Schwabish and Kijakazi 2021).

Between February and May 2020, over 807,000 manuscripts were submitted to more than 2,300 journals on the Elsevier publishing platform (Squazzoni et al. 2020). Each of these manuscripts was presumably reviewed by an editor and sent to experts in the field to review and comment. If editors and reviewers asked authors specific questions about racial equity for each manuscript they review, we would likely see a shift in how race is analyzed, visualized, and written about in academic publishing.

Journals could also start to require "positionality statements" as part of any manuscript submission or publication. A positionality statement describes an individual's social and political context that helps shape their identity in terms of race, class, gender, sexuality, and ability status. Such positionality statements—in short, a summary of the author's lived experiences—may not only highlight existing gaps in the composition of research teams (e.g., an all-white, all-male research team writing about Black women's fertility), but it might also help the research team become more attuned to its own possible biases and perspectives as it relates to their work. (Obviously, journals will need to carefully consider how such statements would work with blind reviews and author privacy.) As Heather Krause noted, "You're not going to get actual empathy if there isn't first some examination of privilege." Positionality statements should examine not only individual team members' identities but also the identities of their team as a whole and that

of the institution they work in, as well as the implications of those identities. Such identities necessarily affect the biases we hold, which in turn influence the decisions made during a research project, such as the questions that are asked, how data are gathered, how findings are interpreted, and who the main audience is for the work. Journals are increasingly requiring financial and data disclosure documents as part of the publication process; positionality statements may be an appropriate next step.

GOVERNMENT AGENCIES, REGULATIONS, AND DIRECTIVES

In addition to being funders of research, federal governments have significant power and influence on how surveys are designed and thus the data that are collected, which in turn affects how research is conducted. In 1977, the Office of Management and Budget adopted Directive Number 15 (OMB 1977), which directed federal statistical agencies to collect and report data on race and ethnicity across four different race categories (American Indian or Alaska Native, Asian or Pacific Islander, Black, and white) and one ethnic category (Hispanic origin). In 1997, OMB revised that directive to define five minimum categories for race (American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Black, and white) and changed the ethnic label designation from "Hispanic" to "Hispanic or Latino" (OMB 1997). Federal agencies follow these directives, and this directly influences how other firms and companies collect their own data. If the OMB were to revisit these categories that were created more than 20 years ago, other agencies, firms, and organizations would follow.

Another agency with significant influence in how racial and ethnic data are reported in the United States is the Bureau of the Census. Charged with collecting and publishing some of the most widely used demographic and economic data in the country (including the constitutionally mandated decennial census), the Bureau has what G. Cristina Mora described as "symbolic power" around defining race and ethnicity:

It's what we call symbolic power. It's like the cultural, symbolic power of the state. And to define who we are lies often really largely in the Census Bureau

in particular....The whole idea that even South Asians right now would see themselves, however tenuously, as part of the Asian-American category is through the power of the Census Bureau. Nobody else ever called them part of having anything to do with East Asians. And now we see it 30 years later, it's pretty strong and a tangible identity that's connected to it. It's not perfect, but that's the power of the Census.

Were the Census Bureau to, for example, revise how it asks about race and ethnicity—that is, how questions about race (e.g., white or Black) are separate from questions about ethnicity (e.g., Hispanic or Latino)—and how it publishes and codes those responses, researchers and analysts would change their approach to discussing these groups as well.

FUNDERS

Many organizations and individual researchers are responsible for raising funds to support their work. Just as researchers should be held responsible for taking a racial equity lens to their work, funders should be held to similar account in the funds they disburse and their requests for analysis. In some cases, funders may request data in formats and categories that are short sighted or that do not align with best practices. As an example, in summer 2020, we (the authors) were asked to create a series of charts for a membership organization to highlight the challenges their readers were facing during the COVID-19 pandemic. We insisted that data around race and ethnicity play a central part in the data-driven project even though it was not explicitly asked for. Without that focus, we argued, the analysis would fail to acknowledge the adverse impacts people and communities of color were facing during the pandemic.

In other cases, funders may not recognize the need for overarching changes to their funding models or project processes. In the early part of summer 2020, the Rockefeller, W.K. Kellogg, and Walmart Foundations issued a wide-ranging request for proposals (RFP) to assess the impacts of COVID-19 pandemic responses across the food system. In response to the quick-

turnaround RFP, 17 Black, Indigenous, and other people of color (BIPOC)-led groups working in the food and agriculture sectors called for the RFP to be canceled and to “move in concert with BIPOC-led orgs doing work in BIPOC communities” (HEAL Food Alliance 2020). They argued the call reinforced existing inequalities in the food and agricultural justice sector by offering funding to “white-led organizations to do service work in BIPOC communities, or to fund a white-led organization with an established funder relationship to subgrant to an under-resourced BIPOC-led organization.” This, combined with the short time to submit project proposals (the RFP was issued on June 8, 2020, and proposals were due on July

1, 2020) that would require multiple teams and hundreds of thousands of dollars (if not more than a million) to put together, effectively cut smaller, BIPOC-led organizations from participation. Other organizations elevated these critiques, and the funders ultimately pulled the RFP. The organizations suggested several steps the funders could take in the future, including cultivating real partnerships with BIPOC communities and BIPOC-led organizations; targeting funding to smaller, BIPOC-led organizations; and investing in multiyear grants using a participatory grantmaking model.

15 CASE STUDY FROM THE URBAN INSTITUTE

We now walk through an example when the Urban Institute data visualization team sought to more consciously and critically apply a racial equity lens to a data visualization project. “Tracking COVID-19’s Effects by Race and Ethnicity” (Urban Institute, n.d.), which was published in summer 2020, is a data tool that enables users to visualize the disparate impact of the COVID-19 pandemic by race and ethnicity for various metrics. Initially, the central graphic was a series of line charts that showed how specific metrics had changed over time for each of five racial or ethnic groups. These charts displayed all the races together, with each racial group represented by its own line, a common approach for visualizing this type of data. The image in figure 17 shows, by racial and ethnic groups, the national share of adults in households with children enrolled in public or private school where at least one child had class canceled because of the pandemic between April and May 2020.

Given this project’s focus on racial and ethnic disparities, however, we questioned whether our initial design for visualizing this data truly reflected a racial-equity-aware approach. Though it is common to present data disaggregated by race or ethnicity with all groups on the same chart, this can have negative effects. As consultant Pieta Blakely writes, showing all groups together “encourages [the reader] to compare each of the groups to the highest performing groups,” which can lead to a “deficit-based perspective” that focuses attention

on what low-performing groups are lacking compared with the high performers (Blakely 2019). In addition to portraying members of groups that have been harder hit by the pandemic in a negative light, readers may neglect the needs of individuals within groups that appear to be, on average, doing “well,” even though members from all races and ethnicities have been negatively affected by the pandemic.

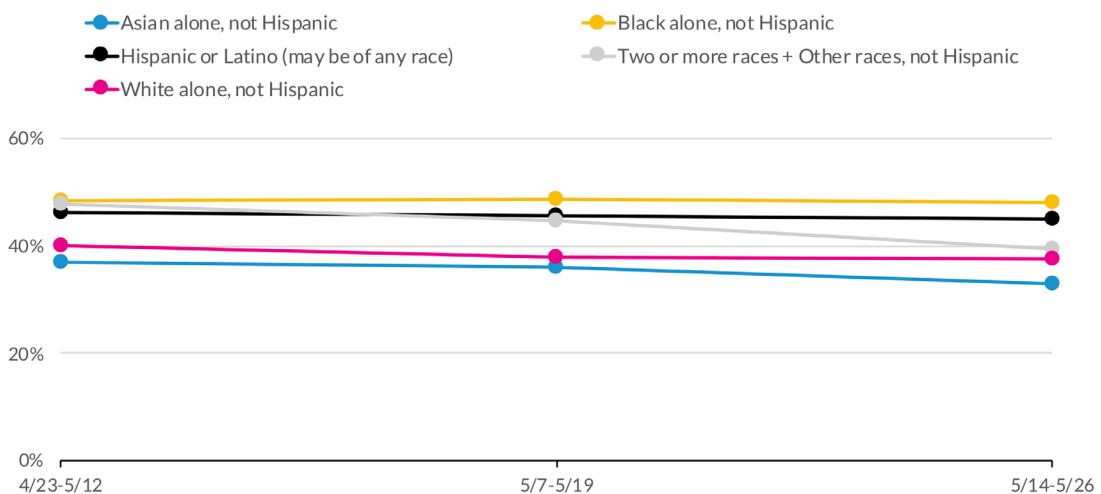
Some counter that this deficit framing perspective is not always applicable depending on how much context regarding the issue being visualized is offered. The UCLA team raised an important point that showing all racial and ethnic groups on the same chart might not be interpreted with a deficit-based lens if sufficient context on the historical inequities and discrimination that have led to these differences between groups is presented. Members of those groups may not necessarily view the chart as deficit based because they understand the gaps are the result of historical discrimination. In the case of our project, however, we felt we did not go into such historical factors in sufficient detail to mitigate the possibility of a deficit-based interpretation.

Another drawback of visualizing data in this manner is that such charts can also make it seem that the racial group least impacted by the pandemic is what the other racial groups should seek to match. But we should ask ourselves why, for example, is 12 percent of households experiencing food insufficiency acceptable just because that’s the share of white households in that situation at a particular point

FIGURE 17

Image from an early draft of the Urban Institute's "Tracking COVID-19's Effects by Race and Ethnicity" project.

Share of adults in households with children enrolled in public or private school where at least one child had class cancelled because of COVID-19



Source: Urban Institute (n.d.).

during the pandemic? The real goal to strive toward is eliminating food insufficiency altogether, not just bringing all racial groups to the same level as the population that has been least negatively impacted. As such, it is worth thinking about how the way we visualize data might reinforce or challenge the way we frame these types of social challenges.

In the end, we shifted our approach to use multiple smaller charts (known as panel charts, trellis charts, or small multiples) rather than plotting all groups on the same visual (figure 18). We felt that by showing each race and ethnicity individually, this design might better encourage readers to think about the specific needs and challenges facing each group. Having six charts to work with (one for each demographic group and one showing all groups) also allowed us to add the relevant location-specific average as a consistent benchmark.

In ordering the six charts, we consciously sorted the groups alphabetically rather than by data value. Because this is an interactive data tool where users can select different metrics, sorting the charts alphabetically

maintains the same order across all views rather than having the graphs shift around, which can be disruptive to the user experience.

Finally, rather than showing estimates as lines, we displayed each data point as a dot surrounded by confidence interval blocks to reflect the uncertainty around our estimates caused by statistical and data collection issues. We used opacity to signal when differences were or were not statistically significant (i.e., more transparent blocks and dots indicate that the difference between the group-specific and the statewide estimates was not statistically significant; more opaque blocks and dots indicate the difference was statistically significant). This was important to show because often the data collected did not allow us to make precise estimates, and we wanted to be transparent about these limitations so readers would not incorrectly interpret what these charts are showing or falsely conclude there was or was not a racial or ethnic disparity when that may not have been the case.

FIGURE 18

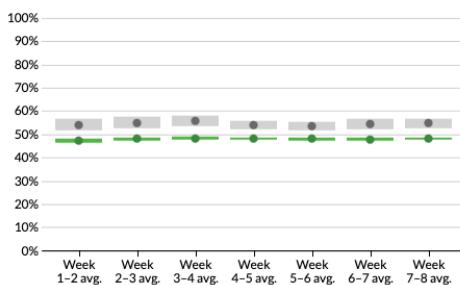
Image of the final design of the Urban Institute's "Tracking COVID-19's Effects by Race and Ethnicity" project.

Share of adults in households where at least one person has lost employment income since March 13

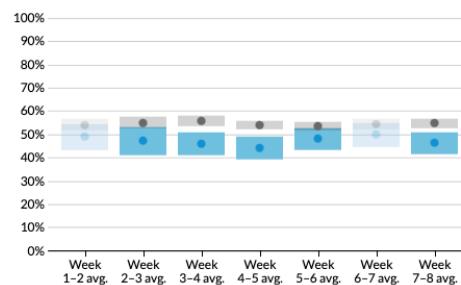
● National average ● State or metro area average ● Race/ethnicity average

■ Statistically significant difference ■ No statistically significant difference

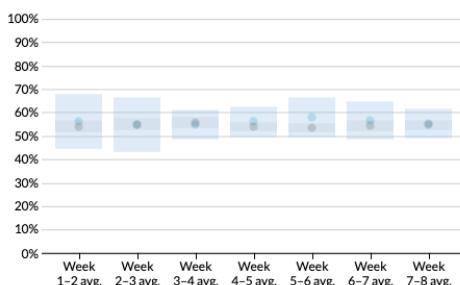
All



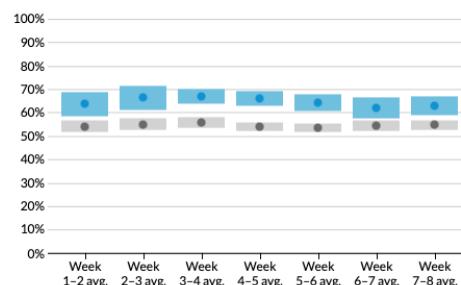
Asian



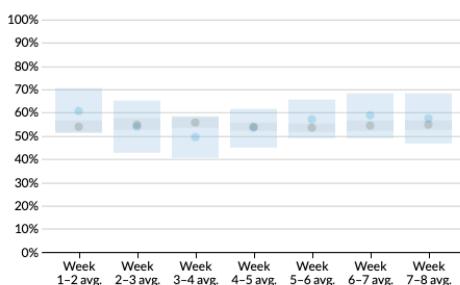
Black



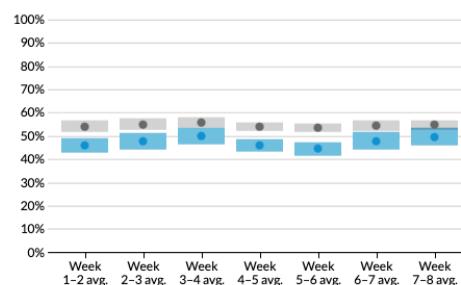
Hispanic or Latino



Two or more/other



White



Source: Urban Institute (n.d.).

16 LOOKING FORWARD

We have laid out many recommendations and guidelines that people and organizations should consider when taking a racially equitable view of their data and data visualizations. The issues we have highlighted here are by no means exhaustive—we have not explored issues relating to, for example, data visualization interactivity, accessibility, database structure and management, cultural differences, data privacy and security, and myriad other issues people working with data should consider daily. Nor are our recommendations ironclad rules. We view the principles outlined here as a starting point for people to think more carefully and critically about how to embrace equity and inclusion throughout the data pipeline of creation, analysis, and communication. Rather than dictating the right or wrong approach, we urge analysts and data communicators to be aware of the decisions they have made and know why they made those decisions.

Many issues we have highlighted here and that our interviewees brought to our attention will change and evolve with society and technology. And experts in data science, data communication, and similar fields are in positions to improve how people view and understand race, ethnicity, and other aspects of identity. Moreover, those views will in turn affect how different policies are created and implemented or products designed and built. Doing so through an equitable and inclusive lens will ultimately create a better society for everyone.

17 ACKNOWLEDGMENTS

The authors thank Greg Acs, Shena Ashley, Kreg Steven Brown, Michael Correll, Heather Hahn, Allen R. Hillery, Billy Jacobson (a.k.a. Anna Lytical), Kilolo Kijakazi, Marla McDaniel, Alexandra Tammaro, and Aaron Williams for their thoughtful comments and feedback. We also thank the people we interviewed during this project for their invaluable insights and experiences: Kim Bui, Kim Campbell, Richard Calvin Chang, Catherine D'Ignazio, Arthur Gailes, Samir Gambhir, Lena Groeger, Lauren Klein, Heather Krause, Kimberlyn Leary, Tiffany Lopes, Ayesha McAdams-Mahmoud, Tim Meko, G. Cristina Mora, Evan Peck, Ninez Ponce, Peter Mattingly, Jacob Rosalez, and Aaron Williams.

THE PEOPLE WE INTERVIEWED

We are especially thankful to the people we interviewed during this project for their invaluable insights and experiences.



KIM CAMPBELL



RICHARD
CALVIN CHANG



CATHERINE
D'IGNAZIO



SAMIR GAMBHIR



LENA GROEGER



LAUREN KLEIN



HEATHER KRAUSE



KIMBERLYN LEARY



TIM MEKO



G. CRISTINA MORA



EVAN PECK



NINEZ PONCE



AARON WILLIAMS

NOT PICTURED ABOVE

Arthur Gailes

Kim Bui

Ayesha McAdams-Mahmoud

Peter Mattingly

Jacob Rosalez

Tiffany Lopes

18 REFERENCES

- American Heart Association. 2018. "Smoking in America: Why More Americans Are Kicking the Habit." American Heart Association, August 30, 2018. <https://www.heart.org/en/news/2018/08/29/smoking-in-america-why-more-americans-are-kicking-the-habit>.
- Bartley, Lawrence, Lisette Bamenga, Adria Watson, Rahsaan Thomas, and Wilbert L. Cooper. 2021. "The Language Project." New York: The Marshall Project. <https://www.themarshallproject.org/2021/04/12/the-language-project>.
- Benson, Craig. "Poverty: 2018 and 2019, American Community Survey Briefs." Suitland–Silver Hill, MD: US Bureau of the Census. <https://www.census.gov/content/dam/Census/library/publications/2020/acs/acsbr20-04.pdf>.
- Betzendahl, Lindsay. 2020. "Racial Inequality in the United States." Tableau public data visualization, June 24. <https://public.tableau.com/en-us/gallery/racial-inequality-united-states?tab=viz-of-the-day>.
- Blackwell, Angela Glover. 2017. "The Curb-Cut Effect." *Stanford Social Innovation Review*, Winter. https://ssir.org/articles/entry/the_curb_cut_effect.
- Blakely, Pieta. 2019. "Presenting Data for a Targeted Universalist Approach." *PietaBlakely.com* (blog), March 6, 2019. <http://blog.pietablakely.com/presenting-data-for-a-targeted-universalist-approach>.
- Blas, Terry. 2019. "'Latinx' Is Growing in Popularity. I Made a Comic to Help You Understand Why." Vox, October 23, 2019. <https://www.vox.com/the-highlight/2019/10/15/20914347/latin-latina-latino-latinx-means>.
- Borkin, Michelle A., Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. "Beyond Memorability: Visualization Recognition and Recall." *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 519–28.
- Bowen, Claire, Aaron R. Williams, and Ajjit Narayanan. 2021. "To Advance Racial Equity, Releasing Disaggregated Data while Protecting Privacy Will Be Key." *Urban Wire*, March 2, 2021. <https://www.urban.org/urban-wire/advance-racial-equity-releasing-disaggregated-data-while-protecting-privacy-will-be-key>.
- Bowen, Claire. 2020. "Will the Census's Data Privacy Efforts Erase Rural America?" *Urban Wire*, March 4, 2020. <https://www.urban.org/urban-wire/will-censuss-data-privacy-efforts-erase-rural-america>.
- Boy, Jeremy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico Bertini. 2017. "Showing People behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?" In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, by Association for Computing Machinery, 5462–74. New York: Association for Computing Machinery.
- Boyd, Rhea W., Edwin G. Lindo, Lachelle D. Weeks, and Monica R. McLemore. 2020. "On Racism: A New Standard for Publishing On Racial Health Inequities." *Health Affairs Blog*, July 2 2020. <https://www.healthaffairs.org/do/10.1377/hblog20200630.939347/full/>.
- Brault, Matthew W. 2012. "Americans With Disabilities." Report P70-131. Suitland–Silver Hill, MD: US Bureau of the Census. <https://www.census.gov/library/publications/2012/demo/p70-131.html>.
- Bui, Kim. 2019. "Designing Data Visualisations with Empathy." DataJournalism.com, June 11, 2019. <https://datajournalism.com/read/longreads/designing-data-visualisations-with-empathy>.
- Carratala, Sofia, and Connor Maxwell. 2020. "Health Disparities by Race and Ethnicity." Washington, DC: Center for American Progress. <https://www.americanprogress.org/issues/race/reports/2020/05/07/484742/health-disparities-race-ethnicity/>.
- Chartoff, Ben. 2018. "Urban Institute's Data Visualization Style Guide: A Living Document." Urban Institute Medium (blog), March 18, 2018. <https://urban-institute.medium.com/urban-institutes-data-visualization-style-guide-a-living-document-b90710702a9f>.
- Cleveland, William S., and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79 (387): 531–54.
- Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: MIT Press.

- Correll, Michael. 2019. "Ethical Dimensions of Visualization Research." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, by Association for Computing Machinery. New York: Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300418>
- Del Real, Jose A. 2020. "'Latinx' Hasn't Even Caught On among Latinos. It Never Will." *Washington Post*, December 18, 2020. https://www.washingtonpost.com/outlook/latinx-latinos-unpopular-gender-term/2020/12/18/bf177c5c-3b41-11eb-9276-ae0ca72729be_story.html
- D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.
- Dörk, Marian, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. 2013. "Critical InfoVis: Exploring the Politics of Visualization." In *CHI 2013: Extended Abstracts on Human Factors in Computing Systems*, edited by Patrick Baudisch, Michel Beaudouin-Lafon, and Wendy E. Mackay, 2189–98. New York: Association for Computing Machinery.
- Dortonne, Nathalie. 2016. "The Dangers of Poverty Porn." CNN, December 8 2016. <https://www.cnn.com/2016/12/08/health/poverty-porn-danger-feat>.
- Dougherty, Jack, and Ilya Ilyankou. 2021. *Hands-On Data Visualization*. Newton, MA: O'Reilly Media. <https://handsondataviz.org/>.
- Durkin, Sarah J., Lois Biener, and Melanie A. Wakefield. 2009. "Effects of Different Types of Antismoking Ads on Reducing Disparities in Smoking Cessation among Socioeconomic Subgroups." *American Journal of Public Health* 99 (12): 2217–23.
- Eubanks, Virginia. 2018. *Automating Inequality*. New York: St. Martin's Press.
- Evergreen, Stephanie. n.d. "How Dataviz Can Unintentionally Perpetuate Inequality: The Bleeding Infestation Example." EvergreenData. Accessed February 25, 2021. http://stephanieevergreen.com/dataviz-inequality_pt1/.
- Groeger, Lena V. 2014. "A Big Article about Wee Things." ProPublica, September 25, 2014. <https://www.propublica.org/nerds/a-big-article-about-wee-things>.
- Groeger, Lena V. "When the Designer Shows Up in the Design." ProPublica, April 4, 2017. <https://www.propublica.org/article/when-the-designer-shows-up-in-the-design>
- Groeger, Lena. 2020. "What Coronavirus Job Losses Reveal About Racism in America." ProPublica, July 20, 2020. <https://projects.propublica.org/coronavirus-unemployment/>.
- Hanrahan, Benjamin V., Ning F. Ma, Eber Betanzos, and Saiph Savage. 2020. "Reciprocal Research: Providing Value in Design Research from the Outset in the Rural United States." *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, by the Association for Computing Machinery. New York: Association for Computing Machinery.
- Harris, Jacob. 2015. "Connecting with the Dots." Source, January 15, 2015. <https://source.opennews.org/articles/connecting-dots/>.
- Kim, Erika. 2019. "Visible Women: Gender in Iconography." Noun Project blog, October 29, 2019. <https://blog.thenounproject.com/visible-women-gender-in-iconography-c6518ec1ddb8>.
- HEAL Food Alliance. 2020. "An Open Letter from BIPOC Leaders in Food and Agriculture to Food Systems Funders." HEAL Food Alliance Medium (blog), July 10, 2020. https://medium.com/@HEAL_Food/an-open-letter-from-bipoc-leaders-in-food-agriculture-to-food-systems-funders-9e00c1f4cb60.
- Heer, Jeffrey, Michael Bostock, and Vadim Ogievetsky. 2010. "A Tour through the Visualization Zoo." *Communications of the ACM* 53 (6): 59–67.
- Hullman, Jessica. 2019. "Why Authors Don't Visualize Uncertainty." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 130–39.
- Hullman, Jessica. 2019. "How to Get Better at Embracing Unknowns." *Scientific American*, September 2019. <https://www.scientificamerican.com/article/how-to-get-better-at-embracing-unknowns/>
- Ishisaka, Naomi. 2020. "Why It's Time to Retire the Term 'Asian Pacific Islander'." *Seattle Times*, November 30, 2020. <https://www.seattletimes.com/seattle-news/why-its-time-to-retire-the-term-asian-pacific-islander/>.
- Kaba, Fatos, Angela Solimo, Jasmine Graves, Sarah Glowa-Kollisch, Allison Vise, Ross MacDonald, Anthony Waters, Zachary Rosner, Nathaniel Dickey, Sonia Angell, and Homer Venters. 2015. "Disparities in Mental Health Referral and Diagnosis in the New York City Jail Mental Health Service." *American Journal of Public Health* 105 (9): 1911–16.
- Kendi, Ibram X. 2019. *How to Be an Antiracist*. New York: One World.

- Klein, Scott. n.d. "Intro: The Design and Structure of a News Application." ProPublica GitHub repository, last updated March 4, 2013. <https://github.com/propublica/guides/blob/master/design-structure.md>.
- Lundgard, Alan, Crystal Lee, and Arvind Satyanarayan. 2019. "Sociotechnical Considerations for Accessible Visualization Design." Presentation given at the 2019 IEEE Visualization Conference, October 22–25.
- Lupi, Giorgia. 2017. "Data Humanism: The Revolutionary Future of Data Visualization." *PrintMag*, January 30, 2017. <https://www.printmag.com/post/data-humanism-future-of-data-visualization>.
- Madden, Mary. 2017. "Privacy, Security, and Digital Inequality," New York: Data & Society Research Institute. https://datasociety.net/pubs/prv/DataAndSociety_PrivacySecurityandDigitalInequality.pdf.
- Marcus, Amy Dockser. 2013. "Trials." *Wall Street Journal*, November 14, 2013. <https://graphics.wsj.com/trials/#chapter=1>.
- Martínez, Alejandro, and Stefano Mammola. 2021. "Specialized Terminology Reduces the number of Citations of Scientific Papers." *Proceedings of the Royal Society B: Biological Sciences* 288 (1948): 20202581.
- MIT (Massachusetts Institute of Technology). n.d. "Diversity Dashboard." MIT Institutional Research, Office of the Provost. Accessed July 2020. <https://ir.mit.edu/diversity-dashboard>.
- McGill, Andrew. 2016. "Why White People Don't Use White Emoji." *Atlantic*, May 9, 2016. <https://www.theatlantic.com/politics/archive/2016/05/white-people-dont-use-white-emoji/481695/>.
- McKernan, Signe-Mary, Caroline Ratcliffe, C. Eugene Steuerle, Caleb Quakenbush, and Emma Kalish. 2015. "Nine Charts about Wealth Inequality (Updated)." Urban Institute, last updated October 5, 2017. <https://apps.urban.org/features/wealth-inequality-charts/>.
- Morais, Luiz, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic. 2020. "Showing Data about People: A Design Space of Anthropographics." *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2020.3023013>.
- Myers, Jan, and Martha MacDonald. 2014. "Reciprocal Relationships: The Role of Government and the Social Economy in the Co-Construction of Social Policy in Atlantic Canada." *Canadian Public Policy / Analyse De Politiques* 40 (1S): S17–25. www.jstor.org/stable/24365086.
- Nadjimbadem, Katie. 2017. "The Trashy Beginnings of 'Don't Mess With Texas.'" *Smithsonian Magazine*, March 10, 2017. <https://www.smithsonianmag.com/history/trashy-beginnings-dont-mess-texas-180962490/>.
- Nguyen, Francis, X. Qiao, J. Heer, and Jessica Hullman. 2020. "Exploring the Effects of Aggregation Choices on Untrained Visualization Users' Generalizations from Data." *Computer Graphics Forum* 39 (6).
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- Noe-Bustamante, Luis, Lauren Mora, and Mark Hugo Lopez. 2020. "About One-in-Four U.S. Hispanics Have Heard of Latinx, but Just 3% Use It." Pew Research Center, August 11, 2020. <https://www.pewresearch.org/hispanic/2020/08/11/about-one-in-four-u-s-hispanics-have-heard-of-latinx-but-just-3-use-it/#fn-29384-5>.
- OMB (Office of Management and Budget). 1977. "Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting." Washington, DC: Office of Management and Budget. https://transition.fcc.gov/Bureaus/OSEC/library/legislative_histories/1195.pdf.
- . 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." *Federal Register* 62 (210). <https://www.federalregister.gov/documents/1997/10/30/97-28653/revisions-to-the-standards-for-the-classification-of-federal-data-on-race-and-ethnicity>.
- Ọnụoha, Mimi. 2016. "The Library of Missing Datasets." <https://mimionuoha.com/the-library-of-missing-datasets>.
- Peck, Evan M., Sofia E. Ayuso, and Omar El-Etr. 2019. "Data Is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, by the Association for Computing Machinery. New York: Association for Computing Machinery.

- Porat, Daniela, Lena Groeger, and Isaac Arnsdorf. 2019. "What Happened to All the Jobs Trump Promised?" ProPublica, May 7, 2019. <https://projects.propublica.org/graphics/trump-job-promises>.
- Rodriguez-Lonebear, Desi. 2016. "Building a Data Revolution in Indian Country." In *Indigenous Data Sovereignty: Toward an Agenda*, edited by Tahu Kukutai and John Taylor, 253–72. Canberra: Australian National University, College of Arts and Social Sciences, Centre for Aboriginal Economic Policy Research.
- Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: Liveright.
- Schwabish, Jonathan. 2021. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. New York: Columbia University Press.
- Schwabish, Jonathan, and Kilolo Kijakazi. 2021. "Rethinking Economics in the Black Lives Matter Era." Washington, DC: Urban Institute. <https://www.urban.org/research/publication/rethinking-economics-black-lives-matter-era>.
- Shirley, Rachel. 2020. "How Colors Advance and Recede in Art." Science of Color. Accessed July 23, 2020. <https://sites.google.com/site/scienceofcolour/how-colors-advance-and-recede-in-art>.
- Slobin, Sarah. 2014. "What if the Data Visualization Is Actually People?" Source, April 24, 2014. <https://source.opennews.org/articles/what-if-data-visualization-actually-people/>.
- Solomon, Mark G., Richard P. Compton, and David F. Preusser. 2004. "Taking the Click It or Ticket model nationwide." *Journal of Safety Research* 35 (2): 197–201.
- Spriggs, William. 2020. "A Teachable Moment? Will George Floyd's Death Spur Change in Economics?" Federal Reserve Bank of Minneapolis, June 9, 2020. <https://www.minneapolisfed.org/article/2020/a-teachable-moment-will-george-floyds-death-spur-change-in-economics>.
- Squazzoni, Flaminio, Giangiacomo Bravo, Francisco Grimaldo, Daniel Garcia-Costa, Mike Farjam, and Bahar Mehmani. 2020. "Only Second-Class Tickets for Women in the COVID-19 Race. A Study on Manuscript Submissions and Reviews in 2329 Elsevier Journals." Amsterdam: Elsevier SSRN. <https://ssrn.com/abstract=3712813> or <http://dx.doi.org/10.2139/ssrn.3712813>.
- Steenhout, Nicholas. 2020. "Disability Language Is a Nuanced Thing." *Part of a Whole* (blog), July 9, 2020. <https://incl.ca/disability-language-is-a-nuanced-thing/>.
- Tison, Julie, Allan F. Williams, and Preusser Research Group. 2010. *Analyzing the First Years of the Click It or Ticket Mobilizations*. DOT HS 811 232. Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.
- Tufte, Edward R. 1985. "The Visual Display of Quantitative Information." *Journal for Healthcare Quality* 7 (3): 15. https://journals.lww.com/jhqonline/citation/1985/07000/the_visual_display_of_quantitative_information.12.aspx
- Urban Institute. n.d. "Tracking COVID-19's Effects by Race and Ethnicity." Urban Institute. Last updated July 23, 2020. <https://www.urban.org/features/tracking-covid-19s-effects-race-and-ethnicity>.
- US Bureau of the Census. n.d. "About the American Community Survey." US Bureau of the Census. Last revised January 4, 2021. <https://www.census.gov/programs-surveys/acs/about.html>.
- White House. 2021. "Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government." January 20. <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>.
- Wilkerson, Isabel. 2020. *Caste: The Origins of Our Discontents*. New York: Random House.
- Williams, Sarah. 2020. *Data Action: Using Data for Public Good*. Cambridge, MA: MIT Press.
- WebAIM. n.d. "Contrast Checker." WebAIM. Accessed May 12, 2021. <https://webaim.org/resources/contrastchecker/>.

19 INTERVIEW GUIDE: DO NO HARM GUIDE AGENDA

Our interviews were conducted between December 2020 and February 2021. Each lasted for about an hour and were conducted via Zoom and transcribed for internal purposes.

1. Comments or feedback about our existing paper.
 - a. Is there anything we missed? What other topics do we need to discuss and include in the guide?
2. How do you and your organization, or organizations you have worked with, approach taking a racial equity lens to communicating data? This can encompass anything from language to colors to icons to different data visualization types.
 - a. If you've worked on any projects that explicitly applied a racial equity lens to the way you communicated data, we'd love to hear more about that experience. What was the process like? What issues did you consider? What decisions were made? Did you encounter any pushback?
 - b. Can you think of any examples of data projects or data visualizations that did not take a racial equity lens?
3. How do we as data visualizers—and people working with data generally—build relationships and bridges with the people and communities we are visualizing/studying?
 - a. How can we solicit feedback from members of these communities on the design of our visualizations? How should we engage with them as fellow participants in the design process?
 - i. We'd love to hear more about any projects that you have worked on that directly involved the participation of the communities that you were visualizing or studying.
 - b. How can we sustain these relationships so that they last beyond the duration of any single project?
4. How do internal organizational culture and analysis overlap?
 - a. How can data visualization and data science teams (and the larger organizations they're housed in) do a better job of reflecting the diverse experiences of the communities we are visualizing/studying?
 - b. Are there aspects of the process of designing and creating a data visualization that could be improved on to be more diverse, equitable, and inclusive?
5. Is there anything else you'd like to add? Anything we missed? Do you have any recommendations for other people we should speak with?

20 OTHER DATA AND DATA VISUALIZATION RACIAL EQUITY RESOURCES

Below we have listed a sampling of books, organizations, and other resources (in addition to those found in the references section) that may be useful to help you and your organization build an equitable approach to data and data visualization. This list will be published on the Racial Equity Data Hub from the Tableau Foundation.

Selected Bibliography

- Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020).
- Sasha Costanza-Chock, *Design Justice: Community-Led Practices to Build the Worlds We Need* (Cambridge, MA: MIT Press, 2020).
- Sarah Williams, *Data Action: Using Data for Public Good* (Cambridge, MA: MIT Press, 2020).
- G. Cristina Mora, *Making Hispanics: How Activists, Bureaucrats, and Media Constructed a New American* (Chicago: University of Chicago Press, 2014).
- Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018).
- Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity Press, 2019).
- Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018).
- Witney Battle-Baptiste and Britt Rusert, eds., *WEB Du Bois's Data Portraits: Visualizing Black America* (Hudson, NY: Princeton Architectural Press, 2018).
- Tactical Technology Collective, *Visualizing Information for Advocacy* (Berlin: Tactical Technology Collective, 2013), <https://visualisingadvocacy.org/get.html>.

Organizations

- Allied Media Projects, <https://alliedmedia.org/>
- Design Justice Network, <https://designjustice.org/>
- Detroit Digital Justice Coalition, <http://detroitdjc.org/>
- The A11Y Project, <https://www.a11yproject.com/>
- We All Count, <https://weallcount.com/>
- Civic Data Lab (MIT), <https://civicdatadesignlab.mit.edu/>
- The Communications Network, <https://comnetworkdei.org/>
- Black Futures Lab, <https://blackfutureslab.org/>
- Data for Black Lives, <https://d4bl.org/>
- The Marshall Project, <https://www.themarshallproject.org/>
- Teachers 4 Social Justice, <https://t4sj.org/>
- Chartability, <https://chartability.fizz.studio/>

Other Resources

- Child Trends' *How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process*, https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective_ChildTrends_October2019.pdf
- Actionable Intelligence for Social Policy's *A Toolkit for Centering Racial Equity Throughout Data Integration*, <https://www.aecf.org/resources/a-toolkit-for-centering-racial-equity-within-data-integration/>
- Chicago Beyond's *Why Am I Always Being Researched*, <https://chicagobeyond.org/researchequity/>
- Black Design in America, <https://bipocdesignhistory.com/>
- *Scratching the Surface* podcast, <https://scratchingthesurface.fm/>
- Urban Institute
 - "Taking an Equity Lens to our Data Practice," <https://medium.com/local-data-for-equitable-recovery/taking-an-equity-lens-to-our-data-practice-8a980fe14d0>
 - "Principles for Advancing Equitable Data Practices," <https://www.urban.org/research-publication/principles-advancing-equitable-data-practice>
 - "Equitable Data Practice," <https://www.urban.org/elevate-data-equity/resources-elevate-data-equity/equitable-data-practice>

21 DIVERSITY, EQUITY, AND INCLUSION IN DATA VISUALIZATION: GENERAL RECOMMENDATIONS

- **Critically examine your data.**
 - Understand where your data come from, who is included and excluded from these data, how these data were collected, why they were collected, and who benefits or is harmed by them.
- **Use people-first language.**
 - Start with the person, not the characteristic such as “people with disabilities,” “a person with asthma,” or “communities of color.”
- **Label people, not skin color.**
 - Use full labels such as “Black people” rather than “Black.”
 - Remember that language continues to evolve. Certain labels that may have been acceptable years ago may no longer be.
- **Order labels purposefully.**
 - Don’t simply order data in tables, graphs, and charts as they are ordered in the data, which may reflect historical biases. Consider alternative sorting parameters such as study focus, specific story or argument, quantitative relationship (i.e., magnitude of the results), alphabetical, or sample size (weighted or unweighted).
- **Consider missing groups.**
 - What groups are not included in your data? Consider adding notes to highlight how the data are not inclusive or representative.
 - Consider alternatives to labeling the “other” catch-all category:
 - *Another*
 - *Another race*
 - *Additional groups*
 - *All other self-descriptions*
 - *People identifying as other or multiple races*
 - *Identity not listed*
 - *Identity not listed in the survey*
- **Carefully consider colors.**
 - Avoid reinforcing gender or racial stereotypes, such as by using baby pink and baby blue to represent women and men or colors associated with skin tones or racial stereotypes.
 - Avoid using incremental color palettes (e.g., light to dark) to represent different demographic groups.
- **Consider icons and shapes.**
 - Recognize how readers might be better able to connect with the data by using small shapes or icons, but use them carefully.
 - Avoid stereotypical, discriminatory, and racist imagery.
 - Use images that show people as empowered and dignified, and avoid images that depict people as helpless victims.
- **Communicate with people and communities of focus.**
 - Reach out to the people and communities you are focusing on in your work and hoping to connect with through the final product.
 - Build teams and connections with outside groups to build and maintain these relationships. This kind of work takes time and effort.
- **Reflect lived experiences.**
 - Not everyone has the same experiences, especially when it comes to characteristics such as race, ethnicity, and gender. Consider what your work may be missing and seek out colleagues and communities to help identify them.
- **Consider the needs of your audiences.**
 - Make sure results are presented in a format that is useful to the audience.
 - Make sure the language used is written in a way that is easily understandable by your readers.
 - Consider translating your products into the languages used by your audiences.

22 THE RACIAL EQUITY IN DATA VISUALIZATION CHECKLIST

- Does the communicator understand the data they worked with, including how it was sourced, who was or was not represented in it, why it was collected, and who benefits or is harmed by having these data collected?
 - Has the communicator carefully considered words, phrases, and labels that are used to describe people, groups, and communities?
 - Has the communicator considered colors that are inclusive of different groups and that are accessible for people with different abilities?
 - Has the communicator considered the order of numbers or estimates in tables, charts, and diagrams? Some options include sorting alphabetically or by magnitude of estimate/number, population size, and sample size (weighted or unweighted).
 - Has the communicator considered alternative words or phrases for groups that may be classified as “other” in the original data? Some alternatives include another race; additional groups; all other self-descriptions; people identifying as other or multiple races; identity not listed; and identity not listed in the survey.
 - Have all icons and images been reviewed with a racial equity lens?
 - Would alternative graph types do a better job presenting the data? Do all groups need to be positioned within the same graph?
- Has the research team communicated with the people or communities that they are focusing on or wishing to communicate with? If not, what people, groups, or organizations can the team contact?
 - Is the research team—and the organization more generally—diverse in its composition and work practices to be able to facilitate better understanding of different groups?
 - Does the final communication product meet the needs of the audience or user?

Teams should consider these issues throughout the research and communication process. Use these checkboxes as a reminder:

- Proposal development stage*
- Strategic planning stage*
- Data collection phase*
- Analysis phase*
- Data visualization phase*
- Writing phase*
- Editing phase*
- Outreach and final publication stage*

ABOUT THE PROJECT

This guide was funded by the Tableau Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Copyright © June 2021. Urban Institute.

Permission is granted for reproduction of this file, with attribution to the Urban Institute.



500 L'Enfant Plaza SW | Washington, DC 20024