Data Science & Diversity

Seeing the People Behind Diversity-Focused DataJam Projects:
The Urban Institute's Do No Harm Guide Quick Reference
By Srija Konduru

Diverse population data requires an empathetic lense and approach.

Diverse Populations (n):

"A population with one or more of the following possible characteristics that differs from others: socioeconomic status, race, ethnicity, language, disabilities, and gender."

-Cultivating Teacher Leadership in Rural School Districts

- Diverse populations include underserved communities that have access to fewer financial and educational resources

-FEMA.gov

Empathy

"understanding, being aware of, being sensitive to, and vicariously experiencing the feelings, thoughts, and experience of another"

-Merriam-Webster Dictionary

Applying an Empathic Approach

- Applying empathy to research is employed in multiple aspects of the process
  - Generating Project Questions
  - Validating Data sources
  - Presenting Findings
  - Communicating with communities about possible solutions

Putting it all together

When creating a DataJam project focused on Data that affects real people, there are a few questions to ask yourself as you work on the project:

- Generating the Project Question
  - Is this question helpful? Harmful? Offensive?
  - Will answers to this question benefit the communities involved?
- Validating the Data Sources
  - Who generated the data? Who benefits?
- Presenting the Findings
  - Visualizations should avoid biased icons/colors, but Icons are great! They help audiences humanize the data
- Communicating solutions to a broader audience
  - Carefully frame the question to communities affected by the question
  - Use plain language
    - Know your audience
  - Interact with audience

Further Reading

- Report by Jonathan Schwabish and Alice Feng
  - Experts in data analysis and visualization
- "In this guide and its associated toolkits, we focus on how data practitioners can approach their work through a lens of diversity, equity, and inclusion."

DO NO HARM GUIDE

Applying Equity Awareness in Data Visualization

JONATHAN SCHWABISH AND ALICE FENG

JUNE 9, 2021
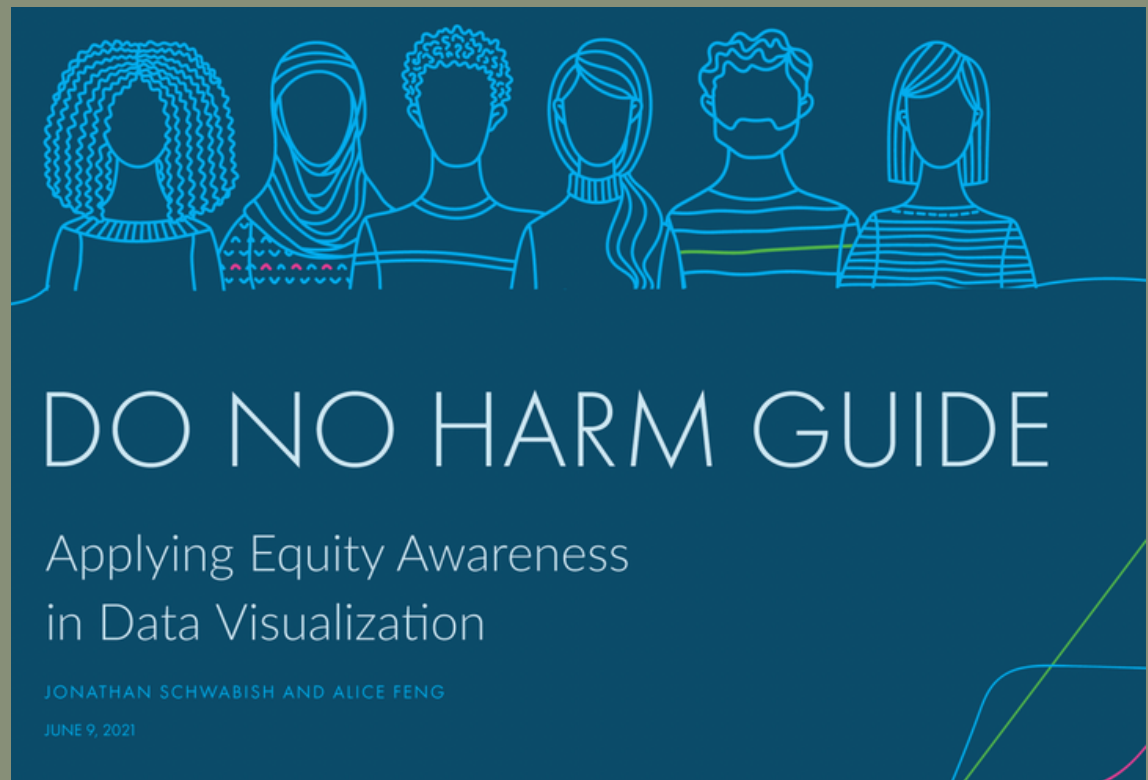
In other words, although human beings have a natural tendency to make categorical distinctions, the categories themselves and meanings associated with those categories are socially constructed rather than natural.

- Article by John A. Powell and Stephen Menendian
  - Experts in civil rights, civil liberties, and ethnic studies
- Higher-level piece on the idea of "othering" groups
- How data represents some groups, while othering (leaving out or not representing) groups is something to keep in mind when presenting how your findings apply to real people
- "This article argued that othering...serves as a conceptual framework featuring a generalizable set of processes that engender group-based marginality."

- Article by Lena V. Groeger
  - Graphics Director at ProPublica
- Article with some examples of how Data visualization may have inadvertently misrepresented diverse populations
  - And how we can learn from them
- One example of a Boston study, "Street Bump" is pictured to the right

A few years ago, the City of Boston released <u>Street Bump</u>, a smartphone app that uses GPS and accelerometer data to detect potholes as Bostonians drive through the city. Every bump gets submitted to the city, and three or more bumps at the same location trigger an inspection and, in theory, a quick repair.
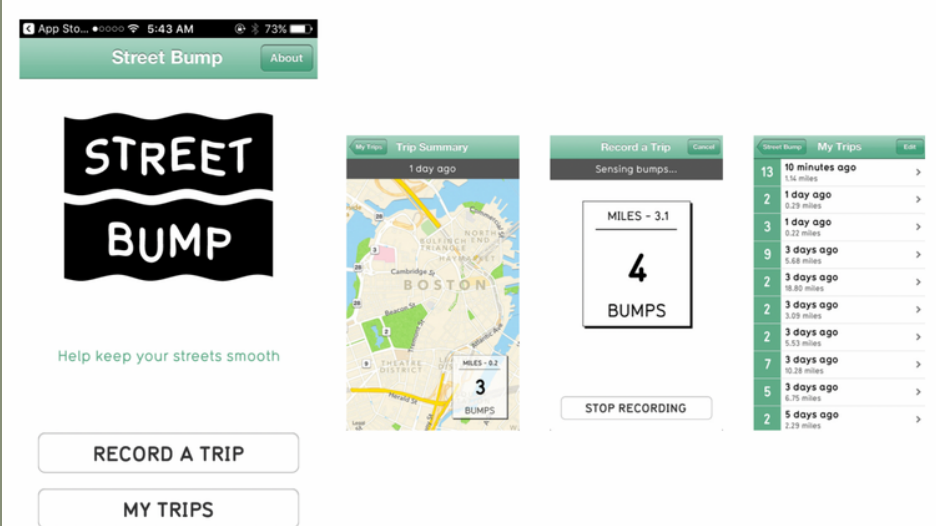
Street Bump, a project to improve neighborhood streets (StreetBump)

Sounds like a great idea: A massive stream of data that reveals the status of street surfaces at low cost to the city! And it *is* a great idea — but there's a problem. Not everyone in Boston has a smartphone. The experiences of people without smartphones (people who tend to be poorer and older) <u>don't show up in the data</u>.

Luckily, Boston was aware of this issue and addressed it from the beginning, by giving the app to city workers to use as they drove around the city. But the implication is clear. Data doesn't speak for itself — it echoes its collectors.

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*