



## Measuring Diversity

"Measuring diversity" specifically refers here to measuring differences in **self-identifying factors** that people use to **confirm themselves** in relation to others.

Using principles from biodiversity + ecosystem diversity, we focus on more than just individuals or flattened profiles.

- "Nature abhors a vacuum" or "Nature ...umm...finds a way."

We are not measuring in a vacuum, designing for monocultures, or assuming spherical cows.



## Self-Identifying Factors include..

Diversity is often measured in open source communities through surveys by factors such as:

- **Geographic region** (i.e. city, state, country, continent)
- **Role in an organization** (i.e. developers, managers, executives)
- **Types of problems someone works on** (i.e. design, front-end engineering)
- **Experience level** (i.e. number of months or years working with a project)
- **Project community** (i.e. Gophers are members of the Go community)

Diversity is also been measured in open source communities by factors such as: **Gender, Age, Sexual orientation, Race, Ethnicity**

## Known Measurement Problems

- Survey Design<sup>1</sup>
- Biased sampling
- Oversimplification
- Overstating results

Representation and counting people matters, but for whom do these kinds of statistic serve? Who does this data belong to? **We must also consistently ask of our demographic collection and goals:**

- Who does this serve?
- Who does this leave out?
- Who does this harm?
- Who does this expose?

## Pattern: *Understand Your Intent + Design with Purpose*

1. Diversity, equity, and inclusion programs may be centralized or decentralized, but DEI work is integrated into all Google human-centered programs
2. If your mission statement includes words like "everyone", "all", or "community", you will want to know how your work is meeting those goals

## Pattern: *Collecting Data for Accountability*

1. Be clear + transparent with your communities on your diversity goals
2. Explain why you ask specific questions + what you find in plain language
  - a. "These questions allow us to measure diversity in the community and highlight opportunities for outreach and growth." (Golang Developer Survey 2021)

## Pattern: *Ask Well-Designed Questions (build with!)*

1. Designing survey questions which reflect a community is a community effort.
2. Research newest, best practices
3. Ask for reviews from a wide range of expertise and experience
4. Context is important – why are you asking about these identifying factors?
  - a. What do these factors represent? Are they proxies for assumed lived experience?
  - b. Are you curious about barriers certain subpopulations face when joining a space or staying in a space?
  - c. When do you ask about assumed experiences versus when do you ask about the gates which keep them out?

## Anti-Pattern: *Tacking demographic questions onto any available mechanism*

1. Relying on demographic questions from ad-hoc surveys can complicate survey design
2. Regularly scheduled data collection and transparent analysis demonstrate commitment to community goals

## Anti-Pattern: *Every Single Inferential Method assigning identity*

At Google, we don't use, recommend, or release products which algorithmically infer gender, race, age, sexual orientation, ethnicity ...

These inferential algorithmic methods continue to be debunked as pseudoscience, even when dressed in technology:

1. Physiognomy's New Clothes, 2017<sup>4</sup>
2. Name-based demographic inference and the unequal distribution of misrecognition, 2023<sup>5</sup>

## Anti-Pattern: *Perpetuating Harm through Careless Design (build for)*

Identity is not protected everywhere - digital world is laid on top of geographic laws and social systems. We have to be aware of who in our community we could be putting into harm's path by asking identifying questions.

1. How we ask a question is as important as the questions we choose to ask.
2. Building inclusive spaces starts with the words you give people to describe themselves.
3. Expect your data to represent real people - complex, intersectional, nuanced.

## Open Source Programs Office

[opensource.google](https://opensource.google)

## About this Project

Google Open Source's mission is to bring the value of open source to Google and all of the resources of Google to open source. We believe that open source solves real-work problems for everyone.

Because when we say everyone, we mean everyone.  
So, how do we know we are - enabling everyone?  
encouraging everyone?  
creating space for everyone?

So...how do you measure "everyone"?

We set specific goals.  
We measure against them.  
We aim to do even better.

## References

1. "Writing Survey Questions", Pew Research Center, <https://www.pewresearch.org/our-methods/u-s-surveys/writing-survey-questions/>
2. *Data Feminism*, Catherine D'Ignazio and Lauren F. Klein, (2020)
3. Golang Developer Survey, Go Community (2021).
4. Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov, Physiognomy's New Clothes, (2017).
5. Lockhart, J.W., King, M.M. & Munsch, C. Name-based demographic inference and the unequal distribution of misrecognition. *Nat Hum Behav* (2023).