

Candid.

Analyzing Candid's demographic data



About Candid Insights

This manual is written by Candid Insights. Candid's Insights team aims to make meaning out of Candid's data and share these insights with the field. These insights are often reflected in Candid's blogs, research reports, infographics, and manuals. The recommendations in this manual on how to understand and analyze Candid data for research reflect decades of evolving institutional knowledge and standards.

Contributors

Anna Koob
Senior lead researcher

Mantin Diomande
Senior research analyst

Cathleen Clerkin
Senior director of Insights

Guy Mika
Research operations associate

Correspondence

Please contact Candid Insights at researchteam@candid.org with questions or to learn more about this data.

Citation

Candid (2023). Candid research manual: Analyzing Candid's demographic data.
doi.org/10.15868/socialsector.42925

DOI

doi.org/10.15868/socialsector.42925

Copyright

© 2023, Candid. This work is licensed under a Creative Commons AttributionNoDerivatives 4.0 International License, creativecommons.org/licenses/by-nd/4.0.

This work was supported by the Robert Wood Johnson Foundation

Date last updated: November 27, 2023

Contents

4 Overview

6 Part I: Demographic data collection and categories

7 Data collection process and procedures

8 Data recentness

9 Survey categories and response options

12 Timeline of changes to Candid's demographic data

14 Part II: Demographic data files and formats

14 Candid nonprofit profile

14 Demographics via Candid dashboard

15 Downloadable flat file

16 Demographics API

17 Part III: Understanding the demographic data set

17 Dynamic, cross-sectional data

18 Data coverage across the sector

18 Data representativeness

21 Data completeness and response attrition

23 Changes to the racial/ethnic categories (2023)

25 Data limitations

27 Part IV: Working with Candid's demographic data

27 Archival data and longitudinal analyses

28 Comparing and joining Candid's demographic data
with other data sets

30 Removing "unuseful" responses

31 Cleaning and recoding write-in responses

33 Defining BIPOC and BIPOC-led

Overview

To better understand the demographic identities of those working in the nonprofit sector and to reduce the reporting burden on nonprofits, Candid collects and shares demographic data about staff and board members from a growing number of U.S. nonprofits and foundations. This manual provides background and best practices for those looking to understand and use this data set for research and analysis.

The document is divided into four parts:

Part 1: Demographic data collection and categories describes what type of demographic data Candid collects from nonprofits. This section outlines the questions that are asked in Candid's demographic form and describes how the data is collected. It also includes a timeline detailing the history of Candid's demographic data collection since 2014, as well as the various changes that have been made since then.

Part 2: Demographic data files and formats outlines four ways to access the demographic data that an organization has shared with Candid: 1) By looking at an organization's nonprofit profile (formerly known as GuideStar profile); 2) By visiting Candid's publicly available Tableau dashboard, which provides users with aggregate demographics across organizations sharing data; 3) By requesting a flat file of the data set, which will include the latest demographic information provided by each participating organization; and 4) by using Candid's Demographics API.

Part 3: Understanding the demographic data set shares some features of the data that will affect how researchers interpret it. This section highlights the dynamic, cross-sectional nature of the data set. It describes the degree to which U.S. nonprofits across the sector are sharing demographic data with Candid and references detailed analyses that Candid's insights team has done to evaluate representativeness in the set. Additionally,

the section describes response attrition in the data set, details changes to the race/ethnicity categories in the survey and presents some key limitations of the data.

For a shorter version of this guide, featuring key highlights see [FAQ: Candid's Nonprofit Demographic data](#).

Part 4: Working with Candid's demographic data offers tips to researchers looking to conduct their own analyses using the data set. Topics include opportunities (and challenges) around longitudinal analysis and joining the demographic data with other commonly used data sets. The section also describes cleaning processes that can be used to remove unusable responses and recode write-in responses. Finally, the manual wraps up with a discussion about defining and operationalizing 'BIPOC' (Black, Indigenous, and People of Color) and 'BIPOC-led' within Candid's demographic data set.

Part I: Demographic data collection and categories

Resources for nonprofits

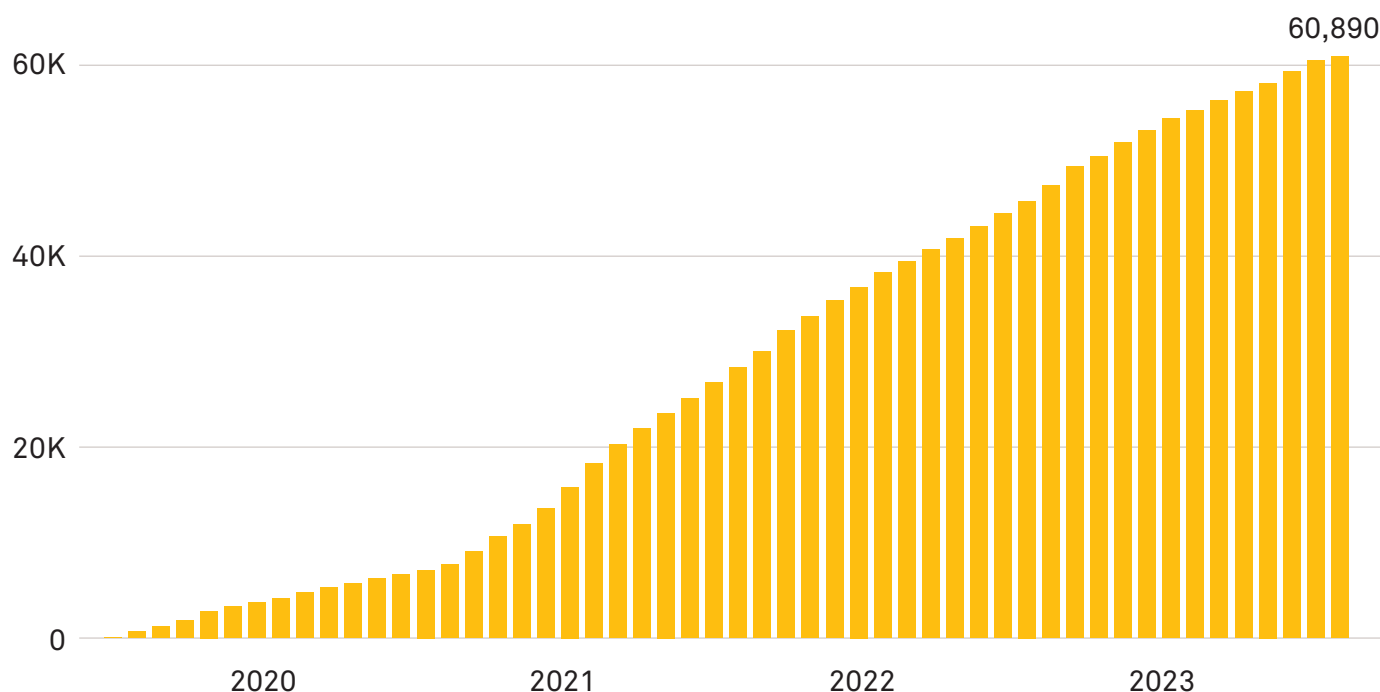
Candid provides a number of resources for nonprofits interested in collecting and sharing demographic data. [This overview](#) describes Candid's demographic data initiative. [This learning module](#) describes how to get started in collecting demographic data and [this guide](#) details what demographic questions to ask and how to collect the results. Finally, see [this](#) for guidance on how to share data with Candid through your organization's Candid profile.

“Demographics” refers to data about groups of people based on shared social identities or characteristics (e.g., race, gender). Candid began collecting organizational demographic data in 2014, inviting U.S. nonprofits to share demographic information about their leaders, staff, and board on their [Candid nonprofit profiles](#) (formerly known as GuideStar profiles) as part of a larger effort to increase transparency in the social sector. In mid-2019, the demographic section of the nonprofit profiles was revised with updated questions to reflect changing language around social identities.¹

Candid's current demographic data set dates back to 2019. Between July 2019 and October 2023, over 60,000 nonprofits had shared at least some demographic information about their organizations with Candid. In 2023, on average, about 1,000 new organizations shared demographic data with Candid per month (see **Figure 1**).

1. These revisions were made in partnership with and under the advisement of [CHANGE Philanthropy](#).

Figure 1: Number of nonprofits sharing demographic data with Candid over time
Ending October 10, 2023



Note: This total includes all nonprofits that responded to Candid’s demographic survey including 1,141 organizations that skipped all questions or responded “unknown” or “decline to state” in all cases. In subsequent charts and figures, these organizations are excluded.

Data collection process and procedures

Candid’s demographic data set is based on nonprofits’ self-reported information.² All U.S. nonprofits are eligible to submit data about the demographic make up of their staff and board via their [Candid profile](#). Sharing this information is completely voluntary, and therefore, the resulting data set should be considered a convenience sample. However, preliminary analyses suggest that the sample is fairly representative of the U.S. nonprofit sector as a whole (see page 18 for more details on data representativeness).

To share demographic data with Candid, nonprofits are asked to first collect this information about their organization independently. Specifically, nonprofits are asked to gather information about respondents’ racial/ethnic identity, gender

2. This includes both public charities and private foundations, as well as other 501c classified organizations such as advocacy groups, labor unions, etc.

identities, sexual orientation, and disability status, as well as information about each respondent's position within the organization—whether they are an organizational leader, board member, senior staff member, or staff member (see *Survey categories and response options* on page 9 for more information). Next, nonprofits are asked to calculate totals for each identity category across each staffing level (e.g. number of male senior staff = 4) and input the totals directly into their Candid nonprofit profile.

Candid requests that nonprofits only share demographic information that has been directly gathered from and approved by respondents. It is never appropriate to fill out demographic information for another person without their consent, or to guess how another person may identify. Moreover, Candid instructs nonprofits to ask how individuals in their organization publicly identify. Given that some individuals may hold different private and public identities (e.g., someone may privately identify as disabled but not want that information shared publicly), organizations should not share any information with Candid where it should not be made public. In cases where individuals do not want to share information publicly, or nonprofits do not know specific answers to demographic questions, there is an option to note that identities are unknown or that individuals have declined to state their identity.

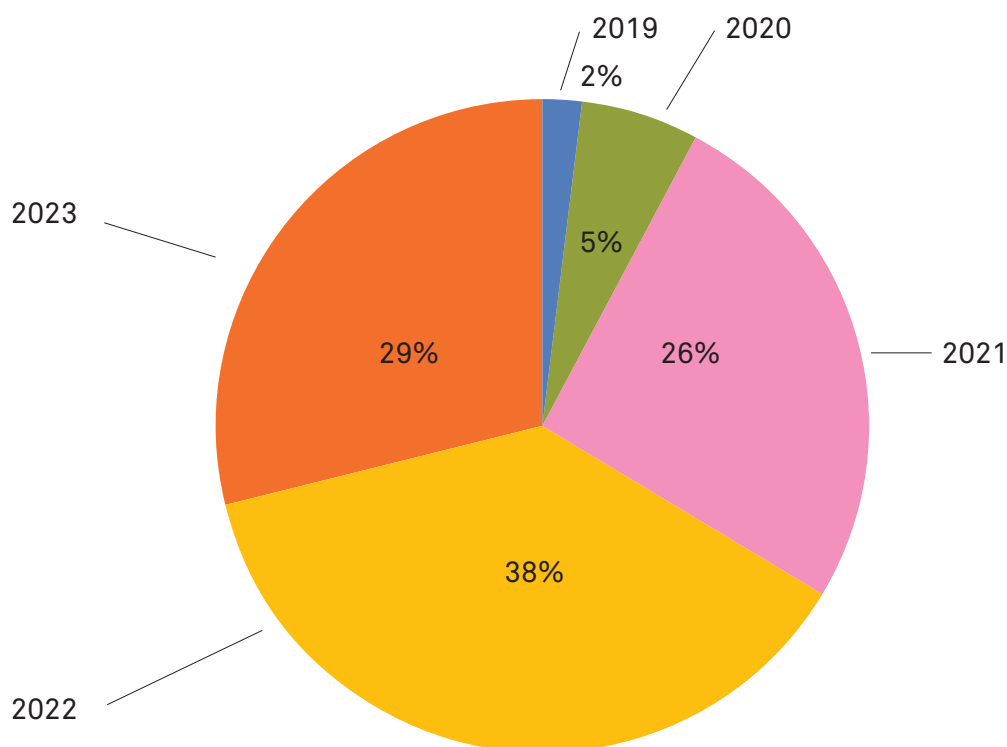
Organizations are not required to share demographic data, however, nonprofits who want to earn a Gold or Platinum Seal of Transparency must complete some demographic information about their leader.

Data recentness

Responding organizations are encouraged to update their demographic data on an annual basis, and many—but not all—do so. The current demographic data set includes organizations' most recently available demographic data, supplied between 2019 and 2023. A nonprofit's most recent demographic information is also listed on their Candid nonprofit profile and remains there until updated. As of October 2023, over 90% of organizations sharing demographic data provided their latest update since 2021 (see **Figure 2**).

Figure 2: Nonprofits sharing demographic data by year of most recent update

As of October 2023



Survey categories and response options

All demographic information is based on self-reported data provided by a given organization. Organizations are asked to collect and share how their staff members publicly self-identify across five demographic identities and four leadership or staffing levels.

Demographic categories

Within each demographic category, participants are asked to select the single best response. Participants may also decline to state if they are uncomfortable disclosing information.

Race & Ethnicity Identity. There is one question about race and ethnicity, and 11 possible responses:

Why not check all that apply?

When surveying individuals about their demographic identities, a common best practice is to allow respondents to “check all that apply” (e.g., a single person may identify as Asian, White, and Multiracial). However, because Candid’s demographic data set aims to capture *aggregate demographic counts at the organizational level*, checking multiple responses introduces complications (e.g., if an organization reported a single person as 1 white board member, 1 Asian board member, and 1 multiracial board member, the same person would be counted multiple times resulting in inaccurate data). Individual organizations may still format questions as “check all that apply” when gathering data. However, the person submitting aggregate data would still have to select the best single option to represent identities selected (multiracial, prefer to identify with another racial identity, etc.)

Race & ethnicity. (Select a single response)

- a. Asian/Asian American
- b. Black/African American
- c. Hispanic/Latino/Latina/Latinx
- d. Middle Eastern/North African
- e. Native American/American Indian/Alaska Native/Indigenous
- f. Native Hawaiian/Pacific Islander
- g. White/Caucasian/European
- h. Multi-racial/Multi-ethnic (two or more races or ethnicities)
- i. Prefer to identify with another race or ethnicity
- j. Unknown or decline to state

Gender identity 1. This question allows individuals to share the extent to which they identify with a binary gender. The terms “nonbinary”, “genderqueer”, and “gender non-conforming” signal that an individual does not identify with traditional binary gender (e.g., female/woman or male/man).

Gender identity. (Select a single response)

- a. Female
- b. Male
- c. Gender nonbinary/Genderqueer/Gender non-conforming
- d. Unknown or decline to state

Gender identity 2. The second question about gender allows individuals to share whether they identify as transgender, not transgender, or another related identity. The term “transgender” indicates that an individual identifies with a different gender than they were assigned at birth. “Not transgender” indicates that an individual identifies with the same gender that they were assigned at birth. This blog post, [Using data to make transgender people visible](#), explains why transgender identity is included as a separate question. The specific question and response options are:

Transgender identity. (Select a single response)

- a. Transgender
- b. Not transgender
- c. Unknown or decline to state

Sexual orientation. There is one question about sexual orientation with three response options. LGBTQIA+ represents gay, lesbian, bisexual, trans, queer/questioning, intersex, and asexual/aromantic/agender identities as well as other sexual orientations within the LGBTQIA+ community. Exact question and response options are:

Sexual orientation. (Select a single response)

- a. Heterosexual or straight
- b. LGBTQIA+ community
- c. Unknown or decline to state

Disability identity. There is one question about disability identity with three response options. According to RespectAbility (respectability.org), disabilities can be physical, learning, cognitive, sensory, mental or chronic health, or another condition that is a barrier to everyday living. Here is the exact question and response options:

Disability status. (Select a single response)

- a. Person with a disability
- b. Person without a disability
- c. Unknown or decline to state

Staffing levels

Demographic data is collected for four staffing or leadership levels within the organization. These groupings are not always mutually exclusive—i.e., ‘staff’ includes ‘leaders’, ‘co-leaders’, and ‘senior staff’, whose data is also collected separately. The levels and their definitions are as follows:

Leaders: This is the person who leads the organization; typically an Executive Director, CEO, or President.

Co-leaders: Nonprofits may list up to one co-leader, in addition to their leader. The majority of organizations indicate that they do not have a co-leader.

Board members: Board members include all members of an organization’s board of governors.

Senior staff: Senior staff includes the leader and co-leader and is defined as people with the authority over budget. Typically this includes C-suite staff like VPs and, depending on the structure of the organization, may include directors.

Staff: Staff includes all staff at all levels, including senior staff and the leader and co-leader (where applicable). Beyond these guidelines, it is left up to the responding organization to determine who they include as staff.

Timeline of changes to Candid’s demographic data

Candid has collected demographic data in some form since 2014 and has periodically updated the demographic data questions and data submission platform to respond to feedback, align with evolving standards, and improve the structure, completeness and cleanliness of the data. The details and implications of these changes are outlined in the timeline below. For more on the principles that Candid uses to guide decision-making around changes, see the following blog post: [*We’re listening: Updates to the race and ethnicity options on the Candid profile.*](#)

Table 1: Timeline of demographic data changes

Year	Change
2014	Candid (then GuideStar) begins collecting data from nonprofits about staff and leadership demographics through the nonprofit profile on GuideStar. Developed in partnership with <u>D5</u> , a five-year coalition to advance philanthropy’s diversity, equity, and inclusion, it was the first demographic data collection effort of its kind in the nonprofit sector.
July 2019	Candid introduces improvements to the data submission portal and revises demographic questions in partnership with <u>CHANGE Philanthropy</u> (see related <u>blog post</u>). Specific changes included: <ul style="list-style-type: none">— Leaders and co-leaders were added as a distinct staffing category— Options for “decline to state” and “unknown” were added for each question— The data submission portal was modified to validate that the number of staff or board members identified within a demographic category totaled the organization’s overall board and staff count as reported on the survey <p>Because of the changes to the resulting data set, it is not practical to compare demographic data collected from 2014-2018 to that collected from July 2019 to the present. Thus, Candid’s current data set is limited to demographic information collected since 2019.</p>

Year

Change

Late 2020

Candid makes it a requirement for organizations to share demographic information about their leader and co-leader (where applicable) in order to receive a Gold Seal of Transparency. Nonprofits earn Candid's [Seals of Transparency](#) by sharing data about their organization on their Candid profile. The new requirements for earning the Gold Seal coincide with a significant increase in organizations sharing demographic data with Candid. In the first year after the introduction of this requirement, the number of nonprofits sharing some demographic data with Candid nearly quadrupled – from 6,000 to 22,000.

2023

Candid announces [Demographics via Candid](#), a campaign to collect, centralize, and share out data on nonprofit demographics. The campaign partners with organizations across the social sector to encourage demographic data sharing one time (Data1x) through nonprofits' Candid profile. The main goal of the initiative is to reduce the burden associated with collecting and sharing this information for nonprofits, while simultaneously providing the sector with data it needs to meaningfully advance equity.

Candid launches a freely available Tableau dashboard that allows users to analyze demographics across nonprofit sharing this data with Candid. For more on this tool, see [relevant section].

Candid introduces updates to the race and ethnicity section of the demographic survey. The changes include two new additional categories: Middle Eastern/North African and Native Hawaiian/Pacific Islander, and updated labels for several existing categories. For more on the details, background, and implications of these changes, see the related blog post: [We're listening: Updates to the race and ethnicity options on the Candid profile](#).

Part II: Demographic data files and formats

Candid offers four different formats for retrieving this data for use in research or evaluation. As of 2023, all data formats are free. Researchers and other users are advised to understand the differences between these formats and determine which one best suits their purposes. For additional information about accessing this data, see [*4 free ways to access Candid's demographic data*](#).

1. Candid nonprofit profile

Best for: Looking up demographics for a single organization.

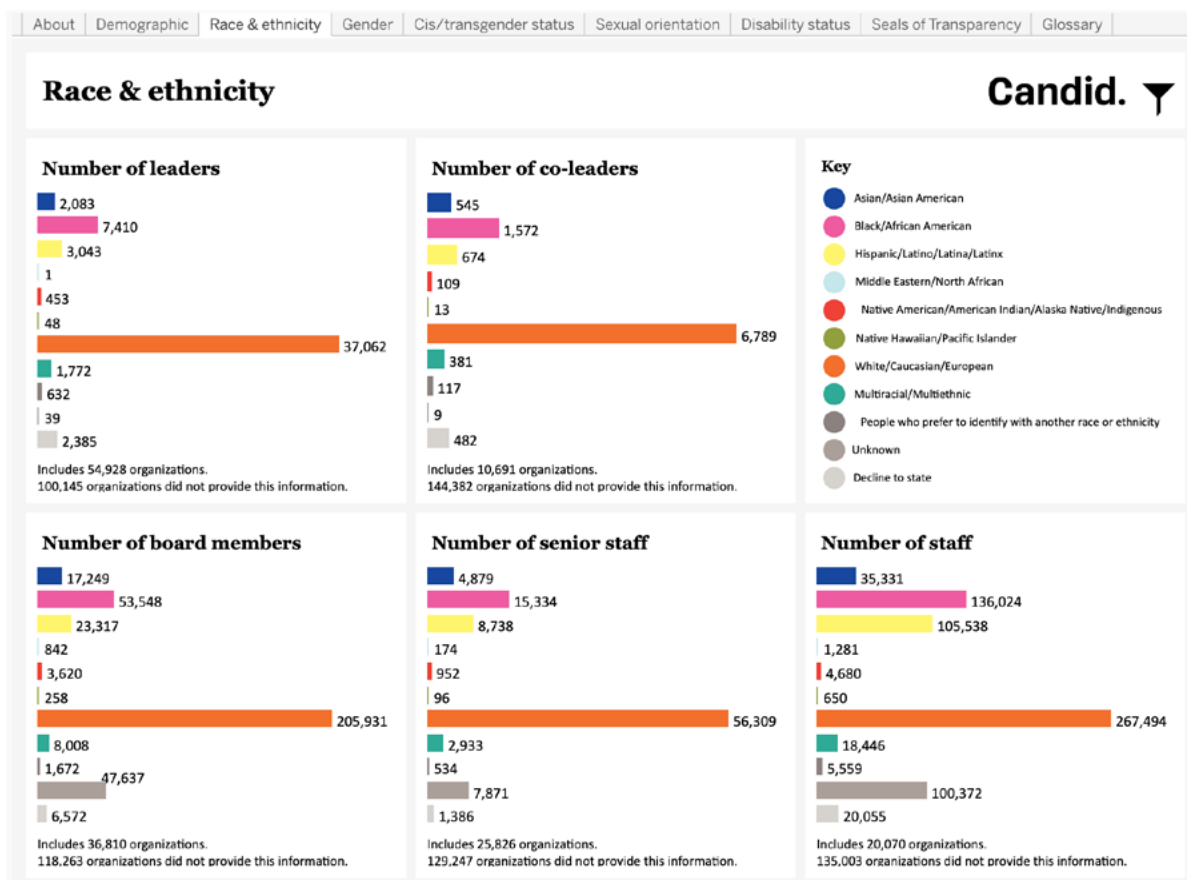
Individuals interested in viewing a particular organization's demographic make-up can go to that organization's [profile on Candid's GuideStar](#) and scroll down to the *Organizational demographics* section. This section will only be available where an organization has shared their demographic data through their Candid profile. The data featured here represents the latest available data that the organization has provided, and updates immediately after an organization shares it.

2. Demographics via Candid dashboard

Best for: Generating charts of demographic data in aggregate, with filtering capabilities. [Demographics via Candid](#) is a publicly available dashboard built in Tableau that allows users to see Candid's demographic data aggregated across nonprofits that share this data (see **Figure 3**). Released in Spring 2023, the dashboard is primarily useful for viewing a quick snapshot of how demographic data breaks down across all nonprofits sharing data or among a particular subset of nonprofits (for example, a funder's cohort of grantees). The dashboard allows for organizational filtering by state, subject and population code

(based on the [Philanthropy Classification System](#)), expense range, and region. It also allows users to build a view based on a customized list of organizations (such as a foundation's cohort of grantees) by entering organizational EINs. The data represents the latest available demographic information for a given organization and refreshes on a weekly basis.

Figure 3: Screenshot of the *Demographics via Candid dashboard*, Race & ethnicity page.



3. Downloadable flat file

Best for: Basic research. Researchers and others interested in raw organizational-level data will likely find the flat .csv file of demographic data most useful. This data set is organized such that each row of data represents one organization's demographic information. It is updated on a monthly basis and reflects the most recent demographic information available for each organization. To receive access to the latest version of the data, complete [this data access form](#). Those requesting access will be required to agree to Candid's [terms of use](#) and are asked to share how they plan on using the data.

To view sample data and the data dictionary associated with this file, visit [this page](#).

4. Demographics API

Best for: Developers, data scientists, and users looking to integrate data into platforms and portals. An API (Application Programming Interface) is a software intermediary that allows two applications to share data and information. As of May, 2023, Candid's demographic data is available as a [free API](#). Demographic data is also included in Candid's [Premier API](#) (a paid API that includes much more data). The demographic data in the API reflects each organization's most recent demographic information and is updated instantly after an organization submits new demographic data.

Table 2: Comparison of data files and formats

	Candid Profile	Demographics via Candid Dashboard	Flat file	Demographics API
Format	Website	Dashboard on Tableau Public; dashboard views downloadable as PPT or PDF file	CSV file	API
Cost	Free	Free	Free	Free
Update cadence	Instantly	Weekly	Monthly	Instantly
Oldest data included	2019	2019	2019	2019

Part III: Understanding the demographic data set

There are several aspects of Candid’s demographic data that are important to understand and account for in order to clean, analyze, and interpret the data accurately. Key considerations include: understanding Candid’s “live” data collection processes; how the data has grown and changed over time; missing data and response attrition; and data limitations.

Dynamic, cross-sectional data

As with most of Candid’s data, the demographic data set is “dynamic”—i.e., it is constantly being updated. Any and every day, a nonprofit might share demographic data with Candid for the first time or might make changes to demographic data they had previously shared via their Candid profile. As of 2023, around 1,000 organizations share new demographic data with Candid every month. This means that demographic counts are never static, especially when examining the entire data set. It also is important to note that different demographic data formats get updated at different cadences (see Table 2 on page 16), so there may be discrepancies when examining data points across sets even when accessed on the same date.

The standard format for this data always includes the latest demographic information for a given organization, and currently includes data shared between 2019-2023. The reason for this approach is that the primary objective of the data is to increase transparency around organizations’ current leadership and staff. In other words, the data is collected in an on-going, cross-sectional manner. It should not be considered an annual panel, nor does it include changes to demographics within a single organization over time.

Those who are interested in studying changes to demographics over time are welcome to create their own longitudinal data sets. Additionally, Candid is working on building more complex historical views of the demographic data so these data sets may become available in the coming years. For more information, see *Archival data and longitudinal analyses* on page 27.

Data coverage across the sector

All U.S. nonprofits that have a Candid profile—including foundations, small volunteer-run organizations, and organizations outside of 501(c)(3) public charities (e.g. 501(c)(4)s, 501(c)(6)s)—have the option of sharing demographic data with Candid. Because Candid profiles encompass such a broad range of organizations, only about 3% of Candid profiles include demographic data. However, among a subset of more typical, active nonprofit organizations, about 11% of organizations share demographic data as of mid-2023.³

Data representativeness

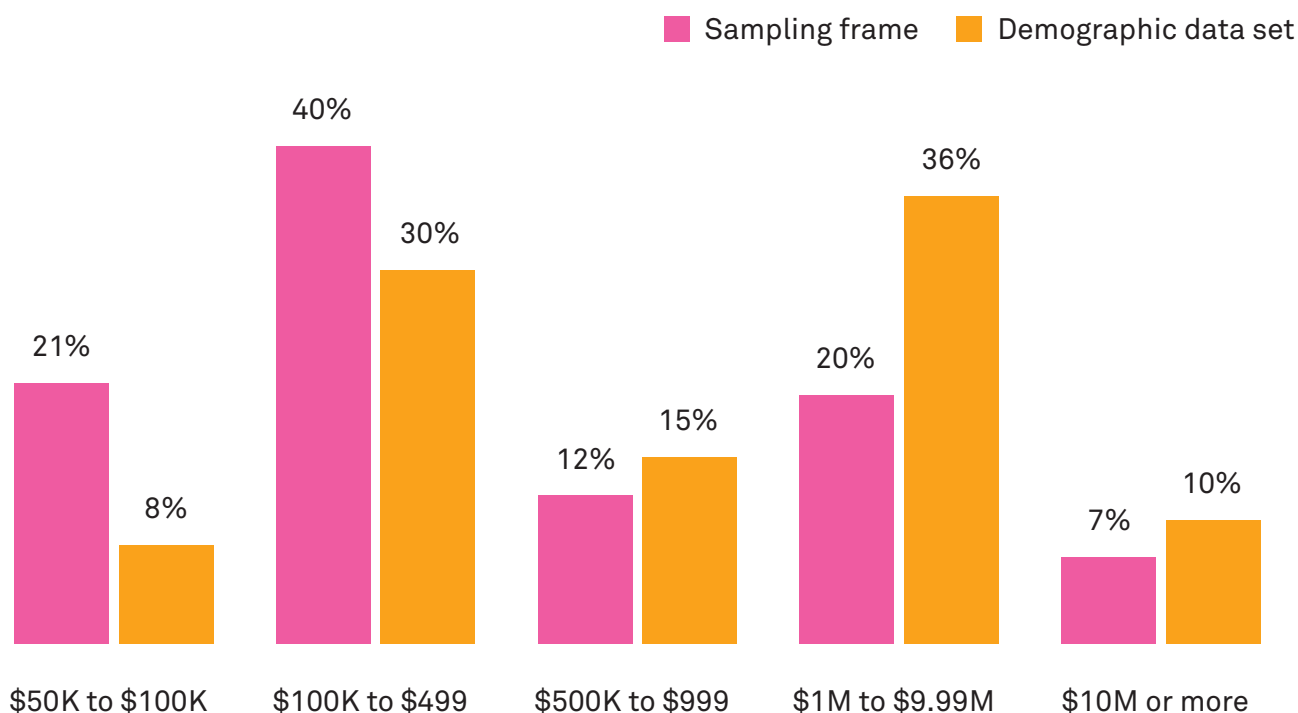
Because U.S. nonprofits and foundations are invited—but not required—to share demographic data with Candid, those who do are a self-selected group. However, preliminary analysis by Candid suggests that the organizations that do share demographic data are fairly diverse and representative of the overall sector in terms of organization size, location, and subsector.

In 2023, Candid’s Insights team updated a previous analysis comparing Candid’s sample to a larger population of active nonprofits (i.e. the “sampling frame”) by organization size, subsector and US region. 312,102 501(c)(3) public charities were included in the larger sampling frame; of these, 42,352 (or about 14%) shared some form of demographic data with Candid.

3. Here, ‘typical, active nonprofit organizations’ are defined as any 501(c)(3) public charity that filed a 990 or 990-EZ within the past five years and had at least \$50,000 in annual expenses per their latest IRS filing.

The analysis found that, overall, the demographic sample largely reflects the field, with nonprofits represented across expense tiers, subsectors, and U.S. regions. However, there were also slight differences. For example, organizations in the demographic sample skewed larger in terms of expenses (see **Figure 4**). When it comes to subsectors, education and religious nonprofits are underrepresented, while organizations focused on the environment are slightly overrepresented (see **Figure 5**). In terms of region, organizations from the Western United States are slightly overrepresented and those based in the Midwest and Northeast are slightly underrepresented (see **Figure 6**). Ultimately, this data set should be considered a large, national, sector-wide convenience sample.

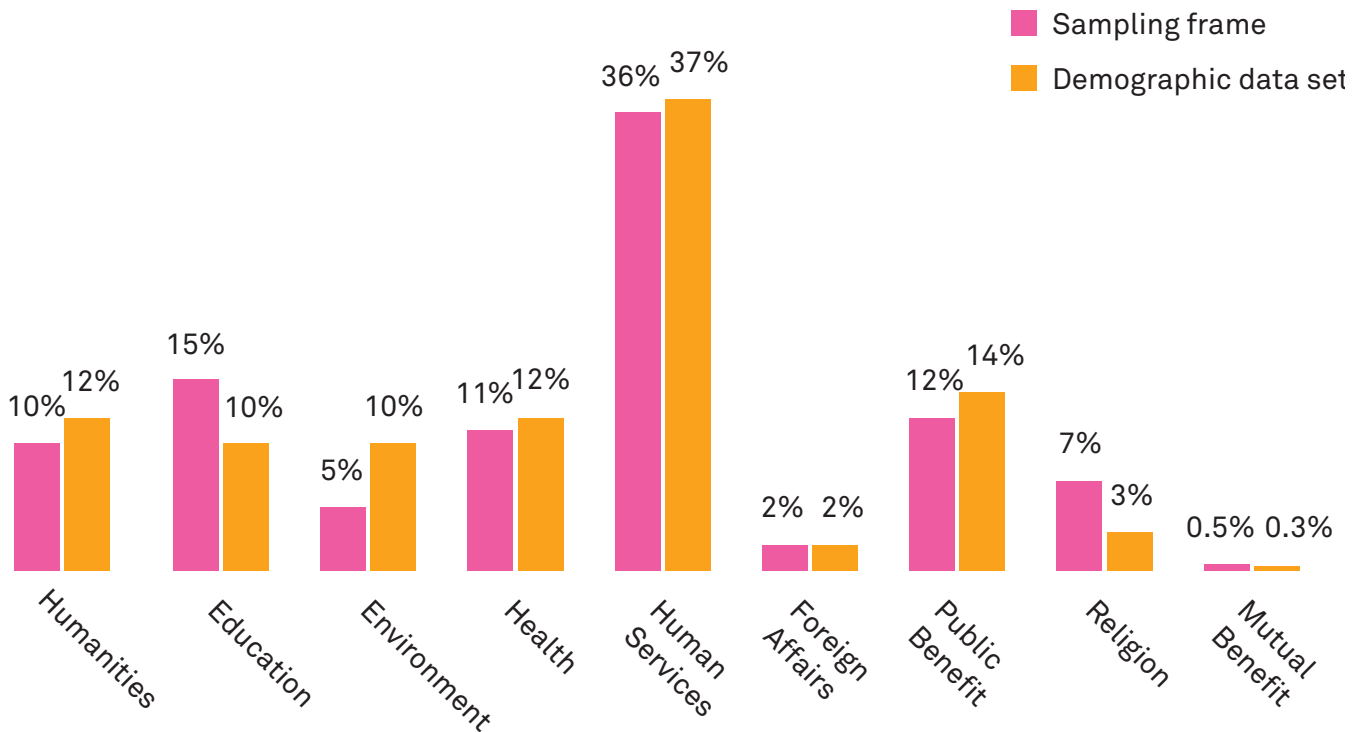
Figure 4: Comparison by expense tiers
As of November 2023



Note: Figures for expenses are based on the organization's latest 990/EZ filing. Sampling frame includes all 501(c)(3) public charities that filed a 990 or 990-EZ within the past five years and have at least \$50,000 in annual expenses. The demographic data set is the subset of organizations in the sampling frame that have shared demographic data with Candid.

Figure 5: Comparison by sub-sector type

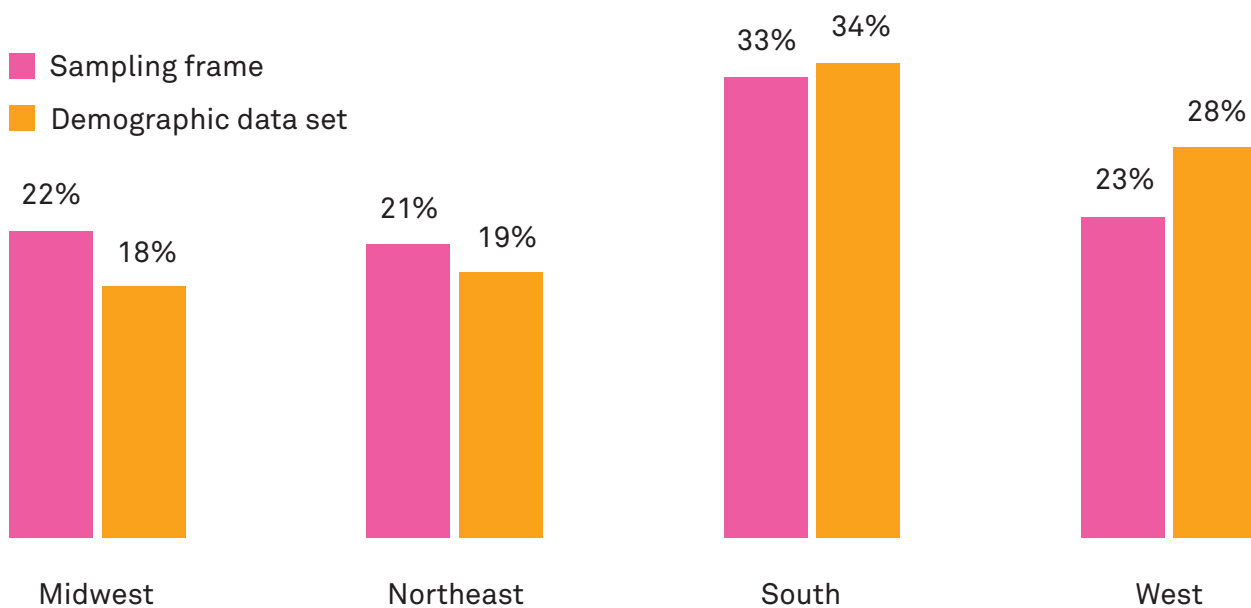
As of November 2023



Note: Based on NTEE codes. Sampling frame includes all 501(c)(3) public charities that filed a 990 or 990-EZ within the past five years and have at least \$50,000 in annual expenses. The demographic data set is the subset of organizations in the sampling frame that have shared demographic data with Candid.

Figure 6: Comparison by U.S. region

As of November 2023



Note: Region based on U.S. Census designations. Sampling frame includes all 501(c)(3) public charities that filed a 990 or 990-EZ within the past five years and have at least \$50,000 in annual expenses. The demographic data set is the subset of organizations in the sampling frame that have shared demographic data with Candid.

Additional analysis conducted by Candid's Insights team suggests that the best predictors of whether an organization will share demographic data with Candid include (in order of importance): the organization's dependence on contributions as a source of revenue; organizational size (as measured by assets, expenses, and revenue); and whether the organization serves "ethnic and racial groups" (those who do are more likely to share demographic data).⁴ These preliminary findings also indicate that organizations serving racial and ethnic groups tend to share more demographic data than other organizations providing this data. More information about this analysis can be found in this blog post: [*Who is sharing nonprofit demographic data with Candid?*](#)

Researchers looking to account for selection bias and/or create a representative sample may want to consider using weights or selecting a stratified random sample of the demographic data based on a selection of the variables mentioned above. A good example of similar techniques as applied to nonprofit data can be found in the [*Nonprofit Trends and Impacts 2021*](#) report produced by the Urban Institute.

Data completeness and response attrition

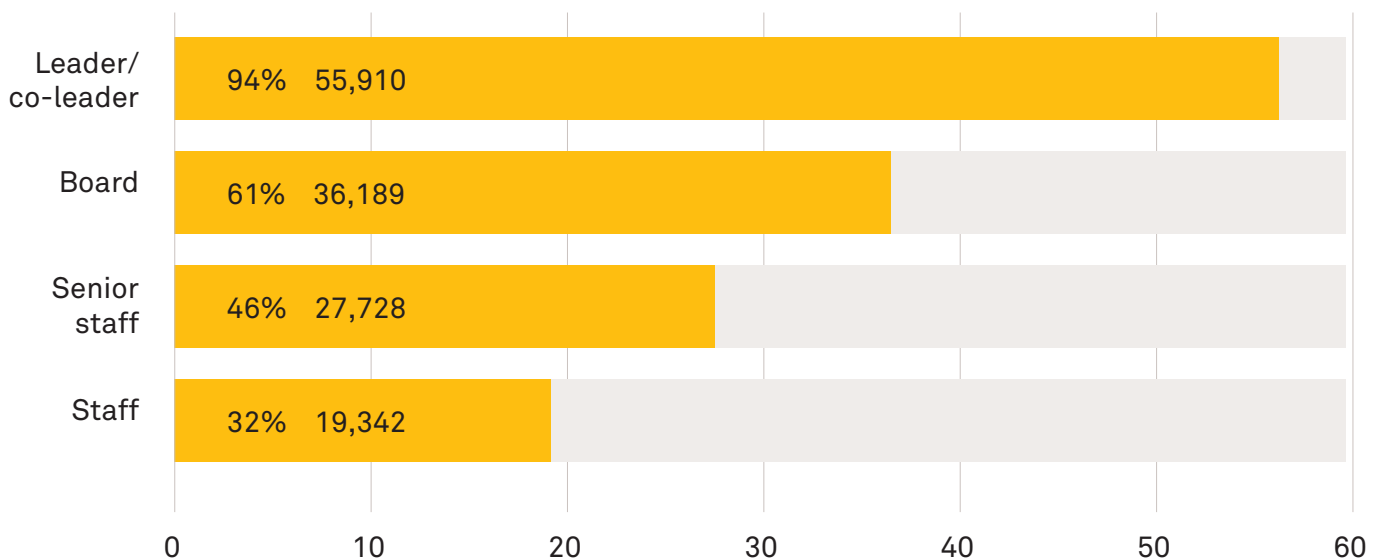
Candid asks nonprofits to share data in response to five demographic questions across four staffing levels, for a grand total of twenty questions. Organizations can selectively respond to questions. They can also select 'unknown' or 'decline to state' for some or all of their staff. About 2% of responding organizations either skip all questions or select 'unknown' or 'decline to state' in response to every question. Candid's Insights team considers these responses "not usable" and excludes them when analyzing the data.

Among organizations providing some usable data, there may be substantial missing data and response attrition. Overall, data is much less complete for broader staff levels. Based on an

4. The term "[ethnic and racial groups](#)" is a code under the taxonomy that Candid uses to classify nonprofits and philanthropic grantmaking. Here, the term is used to identify organizations that focus on non-white populations as the beneficiaries of their services and programming.

analysis updated in October 2023, 94 percent of organizations answer at least one demographic question about their CEO/leader, compared with only 32 percent sharing any demographic data about overall staff (see **Figure 7**). The increased response rate for CEO/leader demographics is most likely due to at least two factors: 1) It is a significantly simpler task to collect and share demographics about one individual as opposed to an entire organization, and 2) in late 2020, Candid began requiring that nonprofits share CEO/leader demographic data in order to earn a Gold Seal of Transparency. This change provides an incentive for organizations that might not otherwise have volunteered this information.

Figure 7: Nonprofits sharing demographic data by staffing level
As of October 2023

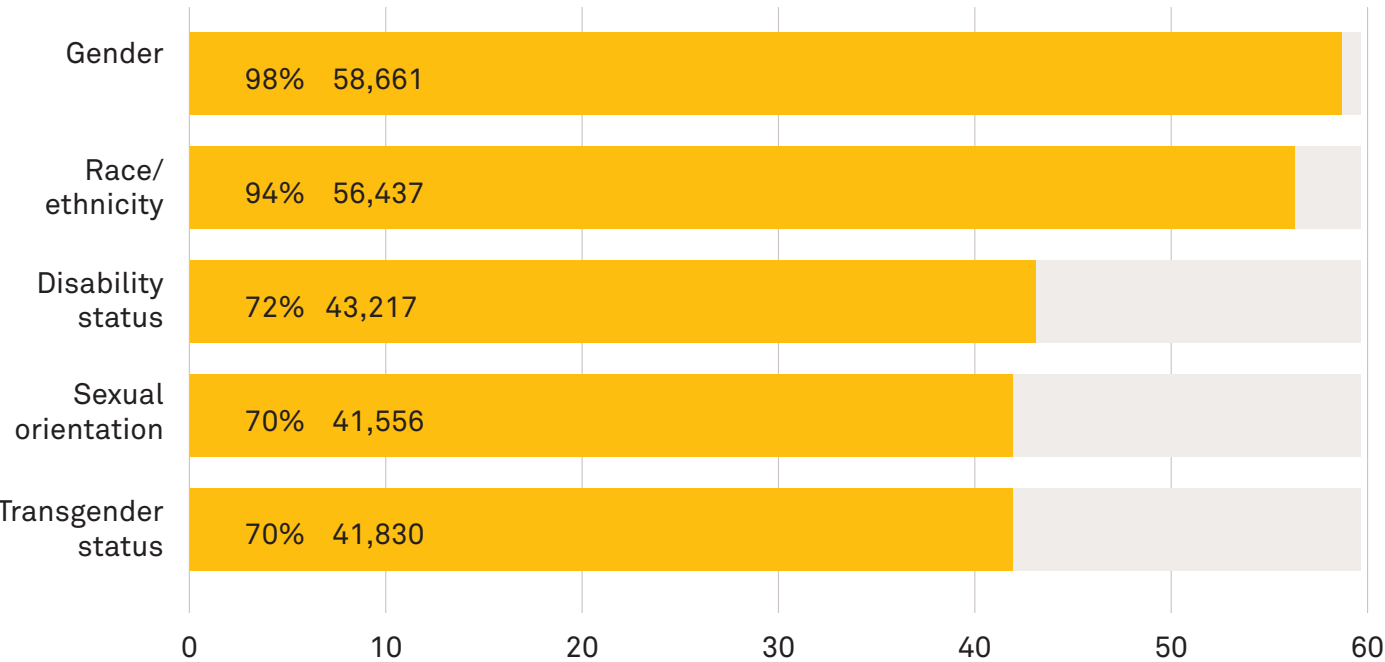


Note: The total number of all responding organizations is 59,749. This total excludes organizations that skipped all questions or responded “unknown” or “decline to state” in all cases.

There is also response attrition for some of the more sensitive demographic questions. In particular, there is more missing data for questions around disability status, sexual orientation, and transgender identity compared to race and gender (see **Figure 8**). Overall, gender was the most commonly answered question, with at least one person from any given organization sharing information about gender in 98 percent of the sample.

Figure 8: Nonprofits sharing data by demographic category

As of October 2023



Note: The total number of all responding organizations is 59,749. This total excludes organizations that skipped all questions or responded “unknown” or “decline to state” in all cases.

Changes to the racial/ethnic categories (2023)

In February 2023, Candid made several small changes to the racial categories included in the demographic data. These changes reflect feedback Candid received from those sharing this data as well as from experts in the field (most notably, [Viewpoint Consulting](#)). They also bring Candid’s efforts into much closer alignment with existing and emerging federal standards on collecting race and ethnicity data. For example, since 1997 federal standards have called for the collection of data on Native Hawaiian/Pacific Islander identity separate from Asian identity and included “Alaska Native” in the category for American Indians. The government’s recent proposal for updating the 1997 standards also recommends collecting race and ethnicity as a single question, as Candid does, and adding Middle Eastern/North African (MENA) as a new category, separate from White. The latter has been a long-felt gap; the MENA community has advocated for this addition for more than 30 years. See **Table 3** for a side-by-side comparison of the previous and current racial categories.

Table 3: Changes to the race/ethnicity options in Candid's demographic survey, February 2023

Previous categories	New categories
Asian American/ Pacific Islander /Asian	Asian/Asian American (<i>label change</i>)
Black/African American/ African	Black/African American (<i>label change</i>)
Hispanic/Latino/Latina/Latinx	Hispanic/Latino/Latina/Latinx (<i>no change</i>)
Native American/American Indian/Indigenous	Native American/American Indian/ Alaska Native /Indigenous (<i>label change</i>)
N/A	Middle Eastern/North African (NEW)
N/A	Native Hawaiian/Pacific Islander (NEW)
Multi-racial/Multi-ethnic (two or more races or ethnicities)	Multi-racial/Multi-ethnic (two or more races or ethnicities) (<i>no change</i>)
White/Caucasian/European	White/Caucasian/European (<i>no change</i>)
Different identity (please specify)	Prefer to identify with another race or ethnicity (please specify) (<i>label change</i>)
Decline to state	Decline to state (<i>no change</i>)

Candid's hope is that these refinements give individuals across the sector more room to select the race or ethnicity with which they truly identify. Of course, the updates have implications for how the data is represented, analyzed, and understood. In particular:

1. New categories of Native Hawaiian/Pacific Islander and Middle Eastern/North African will initially be underreported. As organizations begin to share their demographic data using these new options, the numbers will increase.
2. In some cases, former respondents will be misclassified until they update their information. For example, if someone identifying as 'Pacific Islander' selected Asian/Asian American/Pacific Islander prior to these changes, they will be included under the newly defined Asian/Asian American category until they update their information. Candid is actively reaching out to organizations to encourage these updates.

3. Over time, these changes will likely cause fluctuations in the distribution of responses across categories. For example, if people are newly selecting ‘Middle Eastern/North African’ but had previously selected ‘White’ or ‘Different identity’ (or some other category), those aggregate figures may go down as a result. In analyzing the data, researchers will need to be careful about how to represent these data fluctuations so as not to conflate them with actual demographic changes among nonprofit staff.

Researchers analyzing demographic data in the near term (e.g. in 2023, when there is likely a mix of responses using both old and new categories), may want to look closely at categories affected and determine whether additional cleaning or coding should be done to increase accuracy. For example, in the short-term researchers might want to sum the two new categories of “Pacific Islander” and “Asian/Asian American” back into “Asian/Asian American/Pacific Islander”, to prevent inaccurate under representation of Pacific Islanders. Researchers may also want to categorize new “Middle Eastern/North African” responses as “Prefer to identify with another race or ethnicity”, temporarily, until more people have had an opportunity to utilize the new category.

Data limitations

Like all data sets, Candid’s demographic data set has limitations that impact what research questions can be answered with the data. Researchers should understand these limitations to determine whether this data set is appropriate for their purposes.

Limited intersectional data. As mentioned in Part I, nonprofits are asked to report *individual* level data for leaders and co-leaders (e.g., CEO’s race, gender, etc.), and *aggregate* demographic information for other staffing levels (e.g., 4 Black board members, 5 white board members, 2 Asian board members). The result is that leaders and co-leaders are the only individuals within an organization for whom Candid has intersectional data. In other words, one can analyze combinations of identities (e.g., race, gender, sexual orientation, etc.) for leaders or co-leaders because they are single individuals. However, it is not possible to identify

combinations of identities at the individual level for other organizational roles.

Limited validation of data collection. While Candid's explicit instructions and guidelines state that it is never appropriate to guess the demographic identities of others, there is no way to verify whether those providing information follow these instructions. Candid's nonprofit profile interface requires that one person upload aggregate demographic numbers on behalf of their organization.

Limited historical data for longitudinal analyses. Changes to Candid's demographic data collection process in 2019 make it extremely difficult to compare data collected before and after 2019. Furthermore, because the data is dynamic and nonprofits can update their profiles at any time, the standard demographic data set always only include the *latest* demographic data an organization shared with Candid. See page 27 for further discussion.

Convenience sample and selection bias. As previously mentioned Candid's demographic data set is effectively a convenience sample, as organizations voluntarily elect to share demographic data on their Candid profiles. As such, the data set is subject to selection bias. For a detailed analysis and assessment of selection bias in the data set, see page 18.

Part IV: Working with Candid's demographic data

Archival data and longitudinal analyses

As noted in the previous section, the demographic data set is a cross-sectional data set based on the latest available information from any given organization. As of 2023, the standard data is not formatted to examine change over time within organizations. However, Candid's Insights team has experimented with creating a historical data set that draws on the limited archived demographic data that is available in Candid's systems. This preliminary set captures an organization's demographic data as of December 31 of a given year, going back to 2019. The data set presents several limitations in terms of accuracy and comparability:

1. The archival set won't necessarily capture all of an organization's demographic data updates. If an organization updated their data multiple times in a given year (due to multiple staff changes, for example), this view only captures the latest instance of the demographic data they shared (as of December 31 of that year).
2. Very few organizations have shared data for all available years. Candid's current demographic data collection efforts began partway through the year in July 2019. Only 161 organizations have shared demographic data in 2019, 2020, 2021, *and* 2022. The small number of organizations sharing data in each of these four years limits generalizability about changes in demographics in the nonprofit sector over this period.
3. There may be variations in the type of data that a nonprofit shared year-over-year. For the small set of organizations that have shared demographic data with Candid on an annual basis, there may still be year-to-year fluctuations in what

they share. For example, one year an organization may only share their leader's race/ethnicity, the next year they might share additional leader-level data about gender and disability status, and then the following year, they might start sharing some demographic data about their board. In other words, even when organizations consistently share data, they may be inconsistent in what they're sharing, which would affect a researcher's ability to do time series analysis.

Given these limitations, it is unlikely that sufficient archival data exists to conduct within-subjects longitudinal analyses at this time. However, given the steady increase in organizations sharing demographic data, such analyses may be more plausible in future years.

Comparing and joining Candid's demographic data with other data sets

Researchers often use demographic data as an independent variable (e.g., examining how the demographic make-up of nonprofit boards predicts outcomes such as total donations received, organization size, etc.). Because of this, researchers will likely want to merge or join this data with other data sets to draw conclusions about outcomes of interest. Below are two examples of additional data sources that researchers may want to combine with Candid's demographic data, along with a few tips on how to do so.

Forms 990 data

Many U.S. tax-exempt organizations are required to file Forms 990 with the IRS annually, sharing information about their organization's size, location, revenue, grantmaking, etc. In a sector where transparency is not always a given, 990s offer a window into what is happening within and across organizations. As such, it is one of the most common data sources for research in the social sector. Because Candid's demographic data set includes EINs, merging or joining the demographic data with 990 data is very straightforward. However, there are two aspects of merging these data sources that warrant caution and consideration:

1. **Aligning data years.** As mentioned throughout this manual, the demographic data set focuses on the latest available information from an organization and the majority of the data tends to be from the *current and previous calendar year* (e.g. 2023 and 2022 at the time of this manual's publication). In contrast, 990 data tends to be released on a 2-3 year delay, and provides information about a given organization's *fiscal year*. Because of this, it can take some extra effort to align timelines and ensure that appropriate conclusions are drawn when merging these data sources (for example, demographic data collected and shared in 2023 is not an ideal independent variable for an analysis using 990 data from 2017).
2. **Differentiating between employees and staff.** It is also worth noting that the number of staff members reported in the demographic data set might not always align with the number of employees reported on Forms 990. There are two reasons for this. Firstly, only full 990s require that nonprofits share a count of employees, thus organizations who file 990-EZs, 990-Ns, or 990-PFs will not have the number of employees listed on their IRS forms. Secondly, Candid offers a more open-ended definition of staff, simply asking organizations to report a count of individuals that they consider to be staff; in contrast, the Form 990 has a narrower definition of who is considered an employee. Because of this, it is not recommended that demographic data collected through Candid be assumed to match the organizations' employees as reported to the IRS.

U.S. Census demographics

The U.S. Census Bureau serves as the most prominent source of demographic data about the U.S. population. As such, researchers may want to compare Candid demographic data with Census data. While the structure and design of the decennial Census served as a guide in developing Candid's demographic questions, the two differ in several ways that are important to note. For one, the U.S. Census collects demographic data at the individual level, whereas Candid's demographic data is aggregated and reported at the organization level (except for leader/co-leader data). Additional key differences include:

- The Census survey asks respondents to provide their age, while Candid does not.

- The Census separates its question about Hispanic origin from its question on race. Candid’s format combines race and ethnicity options under a single question.
- When asking about race, the Census uses a “check all that apply” format while Candid uses a “select one response” question format.
- Candid’s demographic questions ask respondents to provide data on staff and leader transgender identity, sexual orientation, and disability status while the Census questionnaire does not.
- Candid’s race and ethnicity question allows individuals to identify as Middle-Eastern and North African, while the Census does not.

For these reasons, researchers should exercise caution when comparing the data sets and make sure they familiarize themselves with Census data methodology. Different Census data sets use different methodologies to calculate percentages. In particular, researchers may want to seek out Census data that has been aggregated to add up to 100%; and to make sure to use “__race alone” aggregates (e.g. “white alone” numbers and percentages reflect individuals who identify as white only, rather than both white and other identities).

Removing “unuseful” responses

Candid follows the best practice of always allowing respondents to skip questions or decline to state demographic information. Because of this, it is possible for organizations to “complete” the demographic section of their profile but skip nearly all questions or solely respond with ‘Unknown’ or ‘Decline to state’ across all questions. These responses are not useful for many types of analyses (e.g., if all board members’ identities are marked as ‘unknown’, it is impossible to draw conclusions about that board’s demographics). As of October 2023, there were 1,141 organizations (2% of the total 53,936) that had provided fully unusable responses—either by skipping all questions or by exclusively responding with ‘Unknown’ or ‘Decline to state’ in all cases. The Candid Insights team typically flags and removes these organizations from the data set before conducting analyses.

Additionally, responding organizations may skip or select ‘unknown’ or ‘decline to state’ for certain questions only. Researchers looking at specific subsets of the data (e.g., disability status for all staff) may want to further remove organizations that haven’t shared that data. The chart below (see **Figure 9**) provides the number of organizations sharing usable data for each demographic category x staffing level combination as of October 2023. It illustrates that the most usable data is available at the leader level and for questions about gender and racial/ethnic identity and that data is more likely to be missing at the staff level for questions of transgender identity and sexual orientation.

Figure 9: Nonprofits sharing data by demographic category and staffing level

As of October 2023

		Staffing level							
		Leader/co-leader		Board		Senior staff		Staff	
		#	%	#	%	#	%	#	%
Demographic Category	Gender identity	54,873	92%	33,288	56%	25,858	43%	17,393	29%
	Race/ethnicity	52,738	88%	31,828	53%	25,778	43%	17,809	30%
	Disability Status	39,758	67%	15,899	27%	15,104	25%	9,510	16%
	Transgender identity	38,889	65%	16,042	27%	15,036	25%	9,017	15%
	Sexual Orientation	38,230	64%	15,613	26%	15,037	25%	9,013	15%

Note: The total number of all responding organizations is 59,749. This total excludes organizations that skipped all questions or responded “unknown” or “decline to state” in all cases.

Cleaning and recoding write-in responses

Candid’s demographic survey allows individual leaders to write in their own responses to the race/ethnicity question. The write-in option is limited to the leader level because at all other levels, data is collected across staff in aggregate. However, Candid does recommend that nonprofits allow for write-in responses at all

staffing levels when collecting data within their organization. Write-in responses make up less than 1% of total responses at the leadership level. Whether to exclude these responses, create a new “write-in” variable, or re-code the write-in responses is at the discretion of the researcher.

To see how Candid has matched write-in responses to existing racial/ethnic categories, visit [this mapping](#).

In 2022, the Candid Insights team conducted a preliminary analysis of the write in responses under the racial category. Out of 274 responses, the five most common *verbatim* write in responses were: “Middle Eastern” (n = 25), “Jewish” (n=17), “Human” (n = 14), “American” (n =6) and “Iranian” (n=6). Content coding further revealed 6 themes as to why people may have chosen to write in their response:

- 1. A more specific identity** (n =~62). Many responses listed an identity (often an ethnicity) that falls within the existing categories but that is more specific than the broad categories offered. Examples include “Irish”, or “White/European/Jewish” (which would typically fall under White/Caucasian/European). Or “Hmong” or “Filipino” (which would typically fall under Asian/Asian American).
- 2. A broader cultural identity** (n =~47). Conversely, a number of responses focused on sharing broader identities including national identity (e.g. “American”, “Australian”), vague racial information (e.g. “minority”, “non-ethnic”, “non-white”), or other cultural identities (e.g. Jewish, Sikh, Nepali, Slavic). It is also worth noting that there were a number of responses that blurred the line between these first two themes (for example, are “Kenyan” and “Jamaican” more specific responses to Black/African American/African? Or broader responses, focusing on nationality over race?).
- 3. Middle Eastern/North African** (MENA; n = ~99). One notable pattern in the write in responses was that over one third of them indicated an affiliation with the Middle East or North Africa. Note that this analysis was conducted before “Middle Eastern/North African” was added to the demographic survey as a separate racial/ethnic category in February 2023. In total, 51 unique responses were identified as MENA including: Arab, Arab American, Arab/Middle Eastern, Egyptian, Iranian,

Iraqi American, Kurdish - Middle Eastern, Lebanese/Middle Eastern, MENA, Middle East/Arab, Middle Eastern, Middle Eastern North African, Middle Eastern-American, Middle-Eastern (Middle East/North Africa – MENA), North African, Palestine Arab, Persian, Persian/Iranian American, and Syrian.

4. Multiple identities (n = ~17). Some respondents opted to write in their response in order to share the specifics of their multiple identities. For example, “White/Hispanic”, “White/ Native American”, “Latino & African American”.

5. Irrelevant identities and information (n = ~33). Other respondents shared information that was unrelated to race. Examples include “woman”, “A female in STEM...”, “Human”, “Veteran”, “Disabled Veteran”, “Dual citizenship...”, “lgbt”, “Child of the Earth”, “Spiritual being”, “not an issue”. (It should be noted that there are questions pertaining to gender, sexual orientation, and disability status elsewhere on the demographic form).

6. Race unknown or confidential (n = ~4). A few respondents also used the write-in section as an opportunity to further explain that racial information was unavailable. For example: “pending hire”, “We do not disclose information about any one individual.”

Defining BIPOC and BIPOC-led

Creating BIPOC variables. In some cases, researchers may find it useful to distinguish between white and non-white racial/ethnic identities among respondents at any staffing level—leader, board, senior staff, or staff. The term BIPOC (Black, Indigenous, and People of Color) is generally defined as individuals who are not white. The term is primarily used in the United States as a way to center the unique and historic injustices that Black and Indigenous people have experienced. Other groups (like the Data Equity Workgroup described below) may use different terms, such as PoC (People of Color) to describe non-white people.

The Candid Insights team generally considers BIPOC to be individuals who report that they publicly identify as one of the following:

- Asian/Asian American
- Black/African American
- Hispanic/Latino/Latina/Latinx
- Middle Eastern/North African
- Native American/American Indian/Alaska Native/Indigenous
- Native Hawaiian/Pacific Islander
- Multi-racial/Multi-ethnic

Additionally, Candid’s demographic format allows for respondents at the leader and co-leader level to provide write-in responses to the race/ethnicity question. Determining whether write-in responses should be considered BIPOC requires some decision-making on the part of the researcher. When in doubt, the Insights team errs on marking write-ins as “unknown”. See above section, *Cleaning and recoding write-in responses*, for more discussion on the range of write-in responses Candid receives.

While the standard data set does not include a specific BIPOC variable, it is fairly straightforward to create a new variable by summing the results of different racial categories (e.g. summing the number of board members of a given organization that identify with one of the BIPOC categories above to create a new BIPOC board members count).

Creating “BIPOC-led” variables. Some researchers are interested in going one step further and examining not only the number of BIPOC individuals at a given organization, but determining which organizations are “BIPOC-led.” Creating such variables is slightly more complicated and requires researchers to make numerous decisions. However, there are several things all researchers should consider in defining and creating such variables:

1. Who counts as a leader? Candid’s data includes three staffing levels that may be considered part of “leadership”: 1) leaders or co-leaders; 2) board; 3) senior staff. Deciding what level of leadership one is considering will determine which variables one needs to be aggregate. Researchers may also want to

Data Equity Workgroup definition of “PoC-led”

In 2023, a group of foundations and diversity experts—led by the Kellogg Foundation, the California Endowment, and the Duke Endowment—came together to align on a common, measurable definition of “People of Color-led nonprofits”. The definition that came out of this workgroup is: *“At least 1 CEO/ Executive director identifies as a person of color; AND, more than 50% of board members identify as people of color.”* Given this workgroup, this definition may become a standard in the years to come. At the same time, we recognize that other researchers may have justifiable reasons for defining “PoC-led” or “BIPOC-led” in different ways depending on their research questions, methodologies, and data sets.

consider combining different leader levels in their definition—for example, requiring that both the leader/co-leader and a majority of board members or senior staff identify as BIPOC in order for an organization to qualify as “BIPOC-led”.

- 2. What percentage of staff need to be Black, Indigenous, and People of Color to consider an organization as being “BIPOC-led”?** For example, Candid’s demographic data allows for information about two organizational leaders (leader and co-leader). Researchers should consider whether one or both leaders need to be BIPOC or in order for an organization to be considered BIPOC-led at this leadership level. Additionally, when considering staff levels with multiple people (e.g., board or senior staff), researchers should determine whether BIPOC individuals need to make up a plurality or a majority of the people in those positions.
- 3. Data limitations.** Defining BIPOC-led and operationalizing that definition in the data can be a difficult exercise. Data limitations can play a major role in what definitions researchers may decide to use. For one, a definition that’s based on a percentage can be less meaningful for small organizations with two or three senior staff and/or board members. Moreover, missing data (particularly at the staff and board level) may mean that many organizations cannot be categorized as either BIPOC-led or white-led. For example, if half of a board’s racial/ethnic data is missing then that organization won’t ‘count’ as having either a majority BIPOC or white board. For further discussion and specific examples of these challenges, see *So you’ve defined “BIPOC-led”...but how will you measure it?*