# Documentation for the National Survey of Nonprofit Trends and Impacts, Spring 2021, Public Use File (PUF)

The National Survey of Nonprofit Trends and Impacts, Spring 2021 was fielded between January and April 2021. During that time, a nationally representative sample of 24,598 operating 501(c)(3) nonprofit organizations were invited to participate in the survey. The Urban Institute worked in collaboration with researchers at American University, and George Mason University, and the nonprofit sector association Independent Sector to conduct the survey.

The survey and subsequent research report were funded by the Generosity Commission, an independent project of the Giving Institute and Giving USA Foundation, which provided support for this first national survey of nonprofits. The survey was also supported by the National Science Foundation Human Networks and Data Science – Infrastructure Program collaborative award numbers 2024310, 2024307, 2024320, and 2024330, which support a collaboration between the Urban Institute, American University, George Mason University, and the Georgia Institute of Technology.

This survey is the first wave of a panel dataset to collect repeat-observations that will help policymakers, nonprofits, and researchers better understand the long-term experiences of U.S. nonprofits. For more information about the project, released reports and analyses, and information about accessing restricted use data, please visit our project page: https://www.urban.org/partnering-understand-long-term-trends-nonprofit-organization-activities-and-needs

We are releasing de-identified survey responses as a Public Use File (PUF) that others can use to conduct their own analyses. This is in keeping with the disclosure that was made to survey respondents that results would only be released without identifying information. Removing these key pieces of identifiable information from the PUF means that full reproducibility of results from published reports is not possible. However, new or additional analyses beyond what was conducted for the research report is still possible.

This document describes the process used to determine which variables to include and exclude from the PUF; as well as manipulations in order to preserve the privacy of the individuals and organizations that responded. Please refer to the full survey codebook for a comprehensive list of included variables, as well as the type of variable (numeric or character) and values for categorical variables.

## Consent and choice to opt-out

Survey respondents were asked to review an informed consent document before beginning the survey. The document stated that they could choose to opt-out of the survey entirely or opt-out of individual questions if they chose.

Because the survey included several skip patterns, it was important for the research team to differentiate between situations where the respondent skipped over a question as part of a skip pattern

vs. when respondents saw the question but chose not to answer. In the PUF the value is replaced with minus 99 (-99) if the respondent saw the question but chose not to respond. If the question was part of a skip pattern, the value is missing (or empty depending on the statistical software, you use).

## Inclusion in the final dataset

Of those invited to complete the survey, 1,548 progressed all the way to the end and submitted their responses via the final survey screen. So long as they did not choose to skip more than 50% of the questions, we call these responses "fully complete".

Others started the survey but did not make it all the way to the end. We chose to keep some of these responses in the dataset. To be included in the dataset the response had to meet the following conditions:

1. The respondent progressed to at least section six (out of nine survey sections)
2. The variables volimp (importance of volunteering) and donimp (importance of donations from individuals) were answered

There are an additional 758 responses in the dataset that meet these conditions. We call these responses "partially complete". PUF users can choose to analyze either all responses (**Group A**, n=2,306) or only fully complete responses (**Group B**, n=1,548). However, the correct weight must be applied depending on which set of responses you wish to analyze.

## Survey weights

There are two survey weight variables included in the PUF. The variable names are:

1. weight_complete_only – this weight should be used when working with data from **Group B** (the 1,548 respondents who completed the full survey)
2. weight_complete_partials – this weight should be used when working with data from **Group A** (the 2,306 total responses)

The survey weights adjust the estimates to account for nonresponse. The weights also adjust for the oversampling of small nonprofit organizations below $100,000 in annual expenses and the oversampling of nonprofit organizations in rural and low-income areas. These survey weights reduce potential non-response bias by adjusting the sample so that the respondents and non-respondents end up with the same distribution of characteristics (organization size, location, and nonprofit classification) that we have information on for the full nonprofit population. The correct weight should be applied depending on whether you are analyzing Group A or Group B. The data should always be analyzed using the appropriate sample weights described above and should not be analyzed unweighted.

## Removing Employer ID numbers

We promised not to disclose the name of the organization or name of the person responding to the survey on behalf of the organization. Employer Identification Numbers (EINs) are issued by the IRS and can be used to search for publicly available tax forms like the IRS form 990. We used EINs to identify a sample of organizations across size, geography and nonprofit classification. These IDs are removed from the Public Use File (PUF), which prevents users from identifying organizations that completed the

survey. Therefore, the PUF does not provide the information needed to compare or link survey responses directly to publicly available information on the IRS form 990.

According to the National Center for Charitable Statistics (NCCS), more than 1.5 million nonprofit organizations are registered in the United States. Our sampling methodology reduced the sample frame to 117,714 organizations, from which random samples were drawn for the survey invitations. We sent invitations to 24,598 organizations to participate in the survey. 2,306 of those (or approximately 2% of the full sample frame) responded either fully or fully enough for the research team to include them in the restricted use dataset.

Only a fraction of a percent (0.16%) of all nonprofit organizations in the US were invited to participate. And only a fraction of those actually responded to the invitation. Since we are not releasing the sample to the public, it would be exceedingly difficult for a user to identify the organization without the EIN, because they would have to compare survey responses against the complete 1.5 million nonprofit organizations rather than just those in our sample or sample frame.

## Removing Other Directly Identifiable Information

In this case directly identifiable information refers to both information that identifies the organization (for example organization name or address) or the individual person responding to the survey.

This information is necessary to remove because with no other source of information, someone could identify exactly which organization the responses are for and then with a simple Google search also obtain the EIN another other linking information. For example, if a bad actor had the website URL variable, they would easily be able to identify the organization and therefor their EIN.

Variables / questions removed because they contain identifying information are:

- Organization name
- Organization mailing address (including city and zip code)
- Organization physical address (including city and zip code)
- Year the organization was founded
- Name of person responding
- Job title of person responding
- Email of person responding
- Organization website
- Addresses and other geographic identifiers of where the organization is located or provides services (including state, county, city, and zip code)

## Removing Comments and Open Text Variables

Several multiple-choice questions in the survey include an "other, specify" option. Users can still see when the respondent selected "other" but the variables containing text in the specify box have all been removed from the PUF as there is a risk that the information included could disclose the respondent or the organization that they represent. Open text questions that also asked respondents for general comments are removed for the same reason.

## Adding Noise to Financial Data

Financial data is an important topic in the survey. We asked respondents to estimate their financial values, but it is possible that some respondents used exact numbers from a recent 990 report. If they used data from a publicly available 990 report, it could potentially be used to link the two sources together and identify responses. Since many financial variables are continuous data (rather than categories) we "added noise" to them. This creates a small amount of imprecision but makes it impossible to merge with another data source like the IRS form 990.

To create noise, we take the original dollar amount, then calculate a new dollar amount, at random, that differs by a maximum of three percentage points above and below the original value. For example, if an organization reports revenue of $100,000, then we randomly assign a new value for this organization between $97,000 and $103,000. When calculating statistics like mean or median, this additional noise has very little impact on the statistic since the values are assigned at random and very by a maximum of 3% from the original value. However, it significantly reduces the risk of identification since none of the values are the actual reported value.

In addition to adding noise, another technique we use is top coding financial data. This means that we calculate the percentile for every value of a variable, then any value that is greater than the 95 percentile is recoded as exactly the 95th percentile. This reduces the chance that large nonprofits can be identified. Since most nonprofits are on the small end of the scale and the distribution skews in that direction, even adding noise to the data may not obscure some of the very largest organizations.

## Citation and license

These data are published under an ODC-BY 1.0 license. You are free to share these data, produce works from these data, and adapt the files as long as you attribute any public use of the PUF (the database) or works produced from the database. The citation should be:

Urban Institute. 2021. Documentation for the National Survey of Nonprofit Trends and Impacts, Spring 2021 Public Use File (PUF). Accessible from https://datacatalog.urban.org/dataset/national-survey-nonprofit-trends-and-impactspublic-use-files