

# Guide to the Ace Community Survey Data, 2021

*Tristan Miller, Caroline Bauer | [tristanmiller@gmail.com](mailto:tristanmiller@gmail.com) |  
[asexualcensus.wordpress.com](http://asexualcensus.wordpress.com)  
December 4th, 2021*

## Table of Contents

<b>I. About this document</b>	<b>1</b>
<b>II. Survey Design</b>	<b>2</b>
A. Sampling Method	2
B. Survey text	3
C. IRB Status	3
<b>II. Data cleaning</b>	<b>3</b>
<b>III. Metadata</b>	<b>4</b>
<b>IV. Additional data</b>	<b>5</b>
A. Write-in interpretation	5
B. Additional metadata	6
C. Long responses	6
<b>V. Known issues with the surveys</b>	<b>6</b>
E. 2019 Survey	6
G. 2020 Survey	6
G. 2021 Survey	6

# I. About this document

The Ace Community Survey is an annual online survey of asexual communities, run by a team of activist volunteers known as the Ace Community Survey Team. The Ace Community Survey tracks the makeup and well-being of asexual communities, providing an evidence-based assessment of community needs. The survey has been administered every year from 2014 to 2020 so far, timed to begin on Asexual Awareness Week in October.

Although the primary purpose of the Ace Community Survey is to serve the needs of activists and community members, rather than serve the needs of research, survey data is also made available to researchers on request. Researchers are provided with preprocessed data, and metadata. This document seeks to explain the preprocessing steps and metadata format, as well as some other background to understanding the survey.

In 2019, another Data Guide was published (<https://asexualcensus.files.wordpress.com/2019/2/dataguide.pdf>), and that guide remains applicable to data from the 2014-2019 surveys. However, beginning with the 2020 Ace Community Survey, there were improvements and changes in the preprocessing and metadata. Therefore, this document supersedes the previous Data Guide, and is applicable to data from the 2020 Ace Community Survey onwards.

## II. Survey Design

### A. Sampling Method

The survey represents a convenience sample recruited via snowballing sampling techniques. Announcements containing a link to the survey were posted on several major asexual websites (AVEN, The Asexual Agenda, etc.), as well as in asexuality-themed groups on various popular social networking sites (Facebook, Instagram, Tumblr, Twitter, Reddit, etc.). In some cases, the survey may be shared on non-English asexual websites, and translation guides may be provided to assist people to take the survey in English. Respondents were encouraged to share the link with any other asexual communities or individuals they knew. Although recruitment was focused on asexual-spectrum (ace) respondents, non-ace respondents who encountered the survey were encouraged to take it as well.

It is crucial to note that the asexual spectrum respondents do not represent asexual spectrum people in general, but rather, the people who have sufficient contact with asexual communities in order to have encountered the survey. Likewise, the non-ace respondents do not represent the general population, but rather those who have enough contact with asexual communities to

encounter the survey. Finally, some ace communities are over- or under-represented because of differences in recruiting effectiveness.

## **B. Survey text**

The survey is made up of a core set of questions that are repeated year after year, and topical questions that are rotated from year to year. Each year we revise the survey, choosing what sections to rotate in or out, and iteratively improving upon the core questions.

The surveys were administered in Google Forms. The text of the surveys can be accessed at <https://asexualcensus.wordpress.com/past-censuses/>.

## **C. IRB Status**

According to the US Department of Health and Human Services, "research" is "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge," and "human subjects research" "obtains, uses, studies, analyzes, or generates identifiable private information." Under this definition, the Ace Community Survey does not qualify as research, but rather as needs assessment for ace communities. Furthermore, if the Ace Community Survey did qualify as research, it would not qualify as human subjects research, because it does not collect identifiable private information. Therefore, the Ace Community Survey does not require IRB review to operate.

However, any research based on the Ace Community Survey would need to seek IRB review on its own. With that in mind, the Ace Community Survey prevents people under the age of 13 from participating, and follows informed consent practices. Specifically, participants are informed up front and throughout the survey that participation is voluntary. Participants also informed about potentially triggering sections, and offered the option to skip them entirely.

# **II. Data cleaning**

In order to prepare the Ace Community Survey data for analysis, certain data cleaning steps are performed. For the convenience of researchers, we provide data that has already been cleaned. Data cleaning consists of the following steps:

- A unique variable name is assigned to each column, replacing the question text. For example, "In which year were you born?" is renamed as "year"
- Where appropriate, column values are converted to integers. For example, a likert question may have responses ranging from "0 - Strongly disagree" to "4 - Strongly agree". These are converted to integers from 0 to 4.

- In questions that ask for the age at time of a certain event, we find that some respondents instead provide the year of the event. We correct this error, and remove any ages that are not between 0 and 99. For example, if a person of age 40 is responding to the 2020 survey, and they enter "1999", this gets converted to "19".
- Certain questions have options such as "Prefer not to answer". These responses are converted to blanks.
- We drop duplicate responses, which are occasionally generated by technical glitches in Google Forms.
- Although most questions are optional, if a respondent skips too many questions, their responses are dropped. This is determined automatically, by detecting which questions have a 99% response rate (to eliminate questions that are commonly skipped), and dropping responses that skip at least 20% of these questions.
- Checkbox questions (where respondents are instructed to "check all that apply") are dummy encoded. This generates a collection of new columns, one for each of the possible options, with a 1 if the option was selected, 0 if it was not selected, and blank if none of the options were selected. Additionally, if there was a write-in option, then the write-ins are extracted to a new column.
- The responses to long response questions are removed. Although we do not believe these fields contain personally identifiable information, as a precaution we do not provide this data to researchers without special cause.

### III. Metadata

In addition to the survey data, researchers are also provided with a metadata file, titled `AceSurvey20XX_columns.csv`. Each row in the metadata file describes a column in the survey data file.

In most cases, each question is associated with a single column. However, due to the dummy encoding, each checkbox question is associated with multiple columns. One column, containing the raw checkbox response, is designated the main column of that question.

The metadata file contains the following information:

- "question" - The full text of the associated question.
- "abbreviation" - The variable name, used as the header of the associated column.
- "category" - The thematic category to which the associated question belongs.
- "type" - The data type of the associated column. All columns are string types unless otherwise specified.
- "age" - This is marked 1 if the associated column contains ages.
- "long response" - This is marked 1 if the associated question asks for a long response. Long responses are removed, and these columns are left blank.

- "checkbox" - This is marked 1 if the associated column is the main column of a checkbox question.
- "writein" - This is marked 1 if the associated question had a write-in response option. For checkbox questions, this is only marked for the main column.
- "mask" - This metadata is currently not in use.
- "group" - If multiple columns belong together, they will all be designated as part of the same group. This occurs in two cases: if all the columns are associated with a single checkbox question, or if multiple questions were contained in the same grid question.
- "group info" - This column contains info on the column's relation to its own group.
  - For checkbox groups: "Raw response" means that it contains the raw responses to a checkbox question. "Writein response" means that it contains the raw writein responses. "Option dummy" means that it contains the dummy variable for the specified option.
  - For grid questions, "group info" contains the text belonging to the associated row in the grid.

## IV. Additional data

In addition to the data that we normally provide to researchers, we also have other data files that may be provided on a case by case basis.

### A. Write-in interpretation

Many questions offer a space for respondents to write anything they like. These are called write-in responses. In the data that we provide to most researchers, the write-in responses are preserved in full, without interpretation.

However, we generate write-in interpretations as part of our own data analysis. We then generate a copy of the data, with interpretations backcoded into data. This data may be made available to researchers upon special request.

To generate the interpretations, one or more interpreters go through the list of write-in responses, and writes out an interpretation for each one. If there are multiple interpreters, then they look at discrepancies between their interpretations, discuss them, and come to an agreement.

Then, we generate a copy of the data file with each write-in response replaced by the interpretation. In some cases, the interpretation may match one of the options already provided. In other cases, interpreters may create new categories. If multiple equally valid interpretations are provided then one of the interpretations is selected at random.

Write-in responses for checkboxes are handled differently. Interpreters are only given access to the write-in responses, and not the dummy variables. Interpreters may place each respondent into one or more of the existing categories, and may create new categories. Some new categories, if they're sufficiently prevalent, will be dummy encoded. Any other new categories that are not sufficiently prevalent are collected into a separate column.

## **B. Additional metadata**

If needed, there is additional metadata available providing specifics on some of the data cleaning steps. For example, there is a complete list of which questions were coerced to integers, and a complete list of which "Prefer not to answer" options were converted to blanks.

## **C. Long responses**

As described in the Data Cleaning section, certain long responses are removed from the data. This data is available separately on special request.

# **V. Known issues with the surveys**

We maintain extensive documentation on issues with each survey. Below we list the biggest issues that we are aware of. These issues cover 2019-2021. For issues regarding earlier years, please consult with the older guide to the Ace Community Survey data (<https://asexualcensus.files.wordpress.com/2019/2/dataguide.pdf>).

## **E. 2019 Survey**

- **Age of identification:** In questions 39-41 (regarding when respondents privately identified as ace, and came out as ace to another ace), we mistakenly allowed any written response, rather than restricting to integers. All non-integer responses have been removed. This resulted in the loss of 764, 494, and 432 responses, respectively, to each question.
- **Age validation:** Due to answer validation that was out of date, the minimum allowed age was 14 instead of 13.

## **G. 2020 Survey**

- **Translation guides:** In some previous years, we've granted individuals permission to produce translations of the survey, allowing people to take the survey when they otherwise couldn't understand it. In 2020 we scaled this process up, allowing Spanish, German, Portuguese, and Italian translations, which were also announced on the website. Unfortunately, it was too late to add a question to indicate how many people used these guides. Not everything in the survey necessarily translates well, however; for

instance the question regarding education might be difficult for people outside the U.S. educational system. The effect on written responses is unknown.

## **G. 2021 Survey**

- **Translation guides:** We further scaled up the translation guides, producing translations into nine languages: Danish, Dutch, French, German, Japanese, Polish, Portuguese, Russian, and Spanish. There is now a question for people to indicate whether they used any of these guides.