Guide to the Asexual Community Survey Data

Tristan Miller | asexualcensus@gmail.com | asexualcensus.wordpress.com
February 6, 2019

Table of Contents

Recommended citation: Miller, T., (2019). Guide to the Asexual Community Survey Data. Retrieved from https://asexualcensus.files.wordpress.com/2019/2/dataguide.pdf.

I. About this document

The Ace Community Survey is an annual online survey of major asexual communities, run by a team of volunteers, the Ace Community Survey Team. The purpose of the survey is to track the makeup and well-being of asexual communities. The survey has been administered every year from 2014 to 2018, timed to coincide with Asexual Awareness Week in October.

The Ace Community Survey Team publishes reports on their website (https://asexualcensus.wordpress.com/) in order to educate asexual activists, researchers, and the general public. Data is also made available to researchers on request.

The purpose of this document is to explain the methodology of the survey in sufficient detail such that researchers who are working directly with the data can better understand it. We include details about the survey administration, the preparation and formatting of data, as well as a list of notable year-by-year issues.

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, Tumblr, Twitter, Reddit, Livejournal, etc.). 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 English-speaking asexual communities in order to have encountered the survey. Furthermore, some ace communities are over- or under-represented because of differences in recruiting effectiveness. Lastly, the non-ace respondents are not representative of the general population.

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 Approval

Since the Ace Community Survey is not affiliated with any university or formal research organization, we do not currently undergo any IRB approval process. We do, however, remove any participants below 13 years of age, and inform participants upfront about potentially sensitive questions.

III. Data Formatting

A. General Format

Researchers are provided with two csv files from each year. The first file is the "data" file, containing the response data output by Google Forms (with additional processing as described in section IV). Each row in the data file corresponds to a survey response, and each column corresponds to a variable. The header contains the names of all the variables.

The second file, the "questions" file, contains information about each variable. The questions file has two columns, the first containing the list of variable names, and the second column contains a description of the variable.

Most variables correspond to survey questions, thus the "description" of the variable is simply the question text.   This question text should be cross-referenced with the original survey text, which will contain further information such as the context of the question, explanatory text below the question, and a list of available options.

B. Checkbox Questions (or, Why are Some Columns Mostly Empty?)

A checkbox question is one that allows respondents to check off any number of boxes from a provided list of responses. All checkbox data is processed in a special way to make further analysis easier, but the format can be counterintuitive at first glance.

For each of the options in a checkbox question, we append a new column to the data file, and we refer to these additional columns as dummy variables. In each dummy variable column, there are three possible values: "y", "n", and "NA". "y" means that the respondent checked the corresponding option. "n" means that the respondent did not check the option. "NA" means

that the respondent did not check any of the options in the question, and is presumed to have skipped the question.

For some of the checkbox questions, the last available option says "Other:" and provides a space for respondents to write in anything they want. This option is given a dummy variable in the same manner as the other options. Additionally, if the respondent wrote anything, their response remains in the original column. The original column is otherwise left empty.

The naming conventions of the dummy variables have changed over time. In 2014-2015, the variable names were generated automatically. For example, a checkbox question about race in the US is given the variable name "USRace", and the dummy variables are named "USRace1", "USRace2", "USRace3", etc. "USRace1" corresponds to the first available option, "USRace2" corresponds to the second available option, and so on. The original survey text must be consulted to determine the order of the options.

From 2016 onwards, the dummy variables are given descriptive names based on the options they correspond to. For example, for the "CanadaRace" question, "CanRaceFilip" is the name of the dummy variable that corresponds to the option "Filipino". The questions csv contains information about these in case the variable name cannot be understood.

In some years, we append an additional column to the data file containing the "raw" checkbox data, as it was originally output by Google Forms. These columns contain a string with all checked boxes, separated by commas.

IV. Data processing

Here we describe the details of how the data is processed before it is given to researchers.

- Each survey response is assigned a unique ID. The responses are then sorted by timestamp.
- A common error in these surveys is for people to write down a year when we asked for an age. In these cases, we infer the age using information about their birth year. We also remove ages that are inconsistent with their current age, and any zeroes (because respondents sometimes write zero as a way of skipping the question). For certain questions (such as identification) we remove any ages below 5.
- Many likert variables are initially strings, because the response options contain additional text explaining the meaning of the numbers. The additional text is removed so they are integers.
- Some responses are removed for various reasons:
  - We remove duplicate responses, sometimes caused by software glitches.
  - We remove respondents who may have been under age 13. Most such respondents are blocked by Google Form's automatic answer validation, but there are a few age/month combinations that must be removed manually.
  - While most questions are optional and do not have a 100% response rate, we remove individual responses that left the survey mostly blank. We count up the

number of skipped questions on each survey (looking at a selection of questions that few people have good reason to skip), make a table, and then cut off people who are on the tail.

- Some questions allow long-form responses. All these responses are removed from the data given to researchers, because of privacy concerns.

V. Additional data

Additional data can be made available to researchers, if they make a special request, and have sufficient justification. Specifically, researchers may request data with write-in interpretation, or they may request long-form response data.

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 all the interpretations backcoded into it. The backcoded data can be made available to researchers who need it.

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 (this is very rare) then one of the interpretations is selected at random.

Write-in responses for checkboxes require special consideration. As explained in Section IIIB, the checkbox columns contain only the write-in responses, and the rest of the information is extracted to the dummy variables. Interpreters were only given access to the write-in responses, and not the dummy variables. From 2014-2015, the interpreters were only allowed to categorize write-in responses, and were not allowed to change dummy variables.

From 2016 onwards, a more sophisticated procedure was developed. Interpreters are still only given access to the write-in responses, but are empowered to place respondents into any of the existing dummy variables. Additionally, they are empowered to create new dummy variables, if enough people fall into a newly created category. Any subthreshold interpretations are in a new "interp" column.

B. Long-form responses

There are some questions that allow respondents to provide long-form answers (such as the "feedback" question). This data is normally withheld for privacy reasons, but can be made available if there is good justification.

VI. 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.

A. General Issues

- **Suicidality**: We are aware that the rates of suicidality in our surveys are quite high, higher than the rates found in the literature.[1] However, they might be more comparable to the suicidality rates seen in the National Transgender Discrimination Survey.[2] Please exercise caution when interpreting these data.
- **Aces with non-ace orientations**: In most years, we asked respondents if they identify with any orientations besides those on the asexual spectrum. We find that there are many ace respondents who also identify as straight, gay, lesbian, bisexual, or pansexual. These labels have multiple interpretations. They may be sexual orientations, especially if demisexual or gray-asexual respondents find they are sexually attracted towards certain genders. They could also be romantic orientations, for example a homoromantic asexual might identify as gay or lesbian. The labels may also communicate some other aspect of a person's identity, such as other kinds of attraction.

B. 2014 Survey

- **Character encoding**: We are aware that there are problems with the character encoding in this data.
- **Religion**: Due to an error, question 5 (what is your religious preference) was missing an option for "other religious". So be aware that the answers to the religion question may be slightly off because of that omission.
- **Sexual history**: There was some confusion from some subjects about whether the sexual history section was asking for consensual sex only, or any sexual encounters. Help text was added a few days in indicating that we are specifically interested in consensual sex only, but it should be noted that some participants (especially those who took the survey before the help text was added) may have included data from non-consensual encounters or encounters of dubious consensuality.

---

[1] For example, see a study of New Zealand secondary school students, Lucassen et al. Aust N Z J Psychiatry. 2011 March;45(5):376-83

[2] https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf

C. 2015 Survey

No significant issues.

D. 2016 Survey

- **Race**: There was a typographical error in question 8 ("Do you identify with any of the following racial/ethnic categories?"). The checkbox "Black and/or of the African Diaspora" was replaced with the word "No". This may have led some African Diaspora respondents to instead check off "Black: African". The typo was corrected during the survey on Monday October 24th around 22:45 PDT. We created a dummy variable called "raceTypo" which is "y" for all respondents that we believe were affected.

E. 2017 Survey

- **Age of identification**: In questions 41 and 42 (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 656 responses to question 41, and 365 responses to question 42.
- **Age validation**: Due to answer validation that was out of date, the minimum allowed age was 14 instead of 13.

F. 2018 Survey

- **Corrected questions**: Question 71 (regarding sexual coercion) was not allowing responses of "0" in the first few hours of release. This was corrected around October 23rd, 9:00 AM PDT.