

Updated: November 16th, 2022

Log of Suspected Issues under Investigation in the GSS Cumulative File, Panel Datafiles, and Data Explorer

Curation of GSS survey data is an ongoing activity, and we rely on the user community to help us report possible issues with the data. Users are encouraged to report potential data issues to gss@norc.org. Making the data available to the community in a timely manner is a foundational principle for our work. For example, the GSS team holds no data embargos.

The GSS team corrects suspected errors as soon as possible but we do not have the resources to send out a notification each time one is confirmed and corrected. Therefore, we plan to release updated data about twice a year (Spring and Summer). However, if you have a question about a specific question and you believe you may have discovered an error or want to check in on one you notified us about, please get in touch with us directly so we can discuss it.

Current known errors that are scheduled to be fixed in the cumulative datafile, panel datafiles, and single-year datafiles are listed below, along with notes about anticipated corrections for the data.

This list will be constantly updated. Once the issues have been addressed and updated data has been released, the item will be removed from this list, and corrections will be included in release notes. Details on the actual updates will be provided in the GSS data documentation.

SAS formats

Years affected: All years

Datasets affected: All data files

Due to the format limitation in SAS, the Release 3 of the GSS 1972-2021 data uses *two* format catalogs (format.sas7bcat) instead of one (please refer to Appendix B in the [2021 GSS Codebook](#) for more information). This two-format setup is causing inconvenience to SAS users when they use the GSS cumulative data. The GSS team is working on a solution to simplify the data use for SAS users.

Expected release: Spring 2023

Household enumeration form variables, XNORCSIZ, SRCBELT, SIZE

Years affected: 2021 GSS

Datasets affected: All data files

Due to issues raised by the shift from F2F to web data collection in 2021, as well as changing data cleaning processes, variables from the household enumeration form (OLD1-14, GENDER1-14, MAR1-14, and RELATE1-14, as well as its derivatives) and geocoded variables (XNORCSIZ, SRCBELT, SIZE) are not yet available for 2021. Please note that due to the nature of the web-only data collection, AWAY1-14 and WHERE1-14 will never be available for 2021, as they were not asked.

Expected release: Spring 2023

HIVTEST1

Years affected: 2006-2021

Datasets affected: All data files

Some value labels are missing in HIVTEST1.

999802: Feb, don't know year

999806: June, don't know year

999898: Don't know year & month

999899: Don't know year & no answer year

Expected release: Spring 2023

ETHWORLD1-ETHWORLD16, ETHREGION1-ETHREGION96

Years affected: 2021

Datasets affected: All data files

The label for value 0 is missing; it should be "No".

Expected release: Spring 2023

DECAUTO, DECBED, DECGIFT

Years affected: 1996

Datasets affected: All data files

The label for reserve code .p is labeled as "Not imputable"; it should be "Not applicable (I have not faced this decision)".

Expected release: Spring 2023

Exports using the GSSDE-generated .dat file

Years affected: All years

Datasets affected: GSS Data Explorer

GSS Data Explorer utilizes negative values to denote missing values (e.g., don't know, inapplicable), but many statistical software packages (e.g., SAS, Stata, SPSS) do not recognize these values as missing. Users may need to recode these negative values to missing, according to their chosen package's specifications, to exclude these cases from analysis.

In addition, the associated syntax files generated for Stata try to assign decimal values a label, even when the value is just a number. Stata does not allow decimal values to be assigned labels and the associated syntax will produce an error.

The GSS team is actively working to find a fix for these issues.

Expected release: Spring 2023

Visualization on variables on GSS Data Explorer

Years affected: All years

Datasets affected: GSS Data Explorer

Some variables appear displayed twice. This does not have an impact on underlying data. This is a visualization issue. The GSS is working on correcting this issue.

The screenshot shows the GSS Data Explorer search results page. At the top, it says "GSS CELEBRATING 50 YEARS | 1972-2022". Below that is a "Search Data" section with a search bar and filters. The search results show 6404 results matching criteria. The results are displayed in a table with columns for variable name, description, and a visualization bar. The variable "year" is listed with the description "GSS year for this respondent". Below it, "wrkstat" is listed with the description "Labor force status". This "wrkstat" entry is circled in red. Below that, "wrkstat" is listed again with the description "labor force status". This second "wrkstat" entry is also circled in red. The page is on "Page 1 of 257" and shows navigation controls.

Expected release: Spring 2023

Incorrect occupation codes

Years affected: 2018

Datasets affected: All data files

OCC10, PAOCC10, MAOCC10, SPOCC10, and COOCC10 contain 20 incorrect values (170, 590, 1070, 1480, 3890, 4270, 5670, 6080, 6870, 6990, 7270, 7490, 7570, 7590, 7780, 7870, 8270, 8690, 9170, 9370) affecting a total of 22 cases across the five variables. These values do not correspond to any values in the Census Occupation Code 2010 codeframe. Initial investigation shows that all of these codes are incorrect, and do not seem to match verbatims supplied by respondents.

Expected release: Spring 2023

EVSTRAY 'Never married' Error

Years affected: 2021

Datasets affected: All data files, GSS Data Explorer

Respondents in Ballot A have punch 3, 'never married' assigned to them programmatically, but are not ever asked EVSTRAY, as it is an item that does not appear on Ballot A. As a result, EVSTRAY appears to be available on all 3 ballots, but was only asked of respondents on Ballots B and C. All respondents on Ballot A will be recoded to .i.

Expected release: Spring 2023

WANT MORE INFORMATION?

Call Us
(888) 837-8988

Email Us
GSS@NORC.org

Visit Us Online
<http://gss.norc.org>