GSS Missing-Data Codes (or Reserve Codes)

The 2022 GSS Cross-section now uses a standardized set of missing values for all variables:

.d – Don't know (DK). This reserve code is used when respondents indicate that they do not know the answer to a question. For some variables (mostly ISSP variables), this code is used for "Can't choose" which was a response option offered to respondents unlike "Don't know" which was not offered.

.i – Inapplicable (IAP). This reserve code is used when a respondent does not see a question, either due to structural factors (e.g., for a respondent on ballot C, they will be marked inapplicable for a question exclusive to ballots A and B) or to personal factors (e.g., a respondent who is not married will be marked inapplicable for a spouse-related item).

.n – No Answer (NA). This reserve code is used when a respondent refuses to answer a question.. If a respondent gives a "No Answer" response to a screener question, all follow-up questions are marked as "No Answer."

.s – Skipped on Web. This reserve code is used when a respondent skips a question in the web mode. If a respondent skips a screener question, follow-up questions are marked as "Skipped on Web."

.x – Not available in this release. Variables that were asked in the 2022 GSS Cross-section but that have proven difficult to clean have been omitted from the first data release. Instead of dropping the variable entirely, all cases are marked with .x, indicating that the data will be made available in a subsequent release.

.y – Not available in this year. Variables that are included in the cumulative GSS datafile but that were not asked in 2022 are coded .y for all cases in 2022. The same is true for variables that are unavailable in in 2021, 2018, 2016, etc. Our hope is that .y makes it easier to tell if two variables that are not part of the replicating core have ever overlapped in a given year.

.z – Variable-specific reserve code. Several variables also include a .z reserve code, which documents circumstances or changes that might lead to odd missing data. For example, in 2021, some respondents in the Religion topical module have .z codes because their assigned religion was recoded based on open-ended responses, and as a result, their movement through the module was different than what their responses to RELIG and DENOM imply.

.u – Uncodeable. Some open-ended responses were not codeable to numeric codes, and .u was used instead in such cases.

.m – DK, NA, IAP. Variables created in the 1970s (e.g., OCC, INDUSTRY) used a single missing value that had multiple meanings. .m can mean "Don't know", "No answer", or "Inapplicable".

.b – Blank. This reserve code indicates 'blank' in open-ended responses.

.p – Not imputable. This missing code is used in variables POVLINE and INCDEF. For more information about these variables, please refer to Methodological Report 57 available on the GSS main website.

.f – Birthdate not available. This missing code is used in the variable VIETDRAFT. Please see the 2021 GSS codebook on the GSS main website for more information.