

## Release Notes (July 27, 2016)

This release has updated the cross-section data up to date. Most notably, the new occupation and industry codes as well as new prestige and SEI scores are added. For more information, please refer to the previous release notes for the cumulative data for a history of fixes.

# GSS 2008 Sample Panel Wave 3, Release 1

August 6, 2013

---

## I. Overview

This GSS panel dataset has three waves of interviews: originally sampled and interviewed in 2008, for the second wave in 2010, and the third interview in 2012. Among the 2,023 cases newly interviewed in 2008, we ended up re-interviewing 1,581 cases in 2010, and 1,295 in 2012 (see Table 1). This data file contains those 2,023 respondents and those variables that were asked in any of the three waves.

<Table 1> GSS Design Features: Cross-Sectional and Panel Components

	GSS Year		
	2008	2010	2012
1 <sup>st</sup> wave	2023	2044	1974
2 <sup>nd</sup> wave	1536	1581	1551
3 <sup>rd</sup> wave		1276	1295
Combined N	3559	4901	4820

## II. Data File Organization

1. The released data file is in the wide format: cases in rows and variables of each wave in columns.
2. To denote waves, we have added a suffix “\_1” or “\_2” to the existing GSS variable names. For example, EDUC\_1 is the years of education in the first wave (2008), and EDUC\_2 is education in the second wave (2010).
3. The values of the following variables do not change over waves so they are included as single variables (without \_1 or \_2): BALLOT, FORM, FORMWT, OVERSAMP, SAMCODE, and SAMPLE.
4. YEAR\_1 is the GSS year of the first wave while YEAR\_2 and YEAR\_3 are GSS year of the second and third wave.
5. ID\_1 is the identification number used in the GSS 2008 data, ID\_2 in 2010, and ID\_3 in 2012. ID generally changes across years.
6. PANSTAT\_2 indicates panel selection status. Users can identify those cases that were: (1) selected, eligible for re-interviews, and actually re-interviewed; (2) selected, eligible, but

not re-interviewed; and (3) selected, but not eligible and not re-interviewed. If we have more information about why the selected cases were not eligible, we used codes 31 through 33 instead of 3 in the data set (codes labeled).

7. For those cases that were not re-interviewed in the second or the third wave, values in all variables are coded to “Inapplicable (IAP)” (actual codes vary by variables).
8. The variables related to respondents’ household members (e.g. OLD<sub>1</sub> to OLD<sub>14</sub>, GENDER<sub>1</sub> to GENDER<sub>14</sub>) do not necessarily indicate the same persons over waves. For example, GENDER<sub>3\_1</sub> and GENDER<sub>3\_2</sub> do not necessarily show the gender of the same household member.
9. Interviewers’ ID numbers (INTID) were newly assigned in each wave. Thus, INTID<sub>1</sub>=56 and INTID<sub>2</sub>=56 do not indicate they are the same interviewer.
10. COHORT reflects year of birth for respondents age 18-89 on AGE. Respondents older than 89 are coded as 89 on AGE and for them COHORT does not match year of birth, but a somewhat more recent year due to the top coding of AGE at 89. Re-interview cases that are older of than 89 are coded to reflect their COHORT at Wave 1 since the top coding of age at 89 prevents their aging from showing up in the data. For example, a respondent who was 90 in the first wave (and top coded as 89 on AGE<sub>1</sub>) would have COHORT<sub>1</sub> as 1917 in GSS 2006. If he was re-interviewed in the second and the third wave and reported ages as respective 92 and 94, he would have been top coded as 89 on AGE<sub>2</sub> and AGE<sub>3</sub> and his COHORT<sub>2</sub> and COHORT<sub>3</sub> would be 1917 in the second and third waves.

### **III. Weights**

Three wave panel data include four different weights: WTPAN<sub>12</sub>, WTPAN<sub>123</sub>, WTPANNR<sub>12</sub>, and WTPANNR<sub>123</sub>. The four different weight variables are differentiated by whether they include NR in the variable name or not, or if they use 12 or 123 at the end of variable name. First, weight variables with NR indicate if the weight variables considered a nonresponse adjustment in addition to selection. Second, while weight variables with 12 indicate two-wave panel (2008-10), weight variables with 123 indicate the three-wave panel, which are 2008-10-12 panel data. Below is the description of four weight variables and methodology used to calculate these weights.

<Table 2> GSS panel data three wave weight variables

Variable name			Description
Weight Variable	Name		
without “NR”			This weight is assigned to the cases originated from 2008. It accounts for all four stages of selection for the 2008 samples (NFA, segment, HU, and respondent) and also for the selection of the segment and the case into the panel sample.
	wtpan12		If you want to analyze only 2008-10 panel data, you need to use this weight variable
	wtpan123		For an analysis of all three waves, you need to use this weight variable.
Weight Variable	Name		
with “NR”			This weight has the same case base as WEIGHTpanel2008 and also includes all stages of selection, <i>but also includes a nonresponse adjustment.</i>
	wtpannr12		If you want to analyze only 2008-10 panel data, you need to use this weight variable
	wtpannr123		For an analysis of all three waves, you need to use this weight variable.

Selection of respondents from the 2008 round for the panel was done in three phases: first, we selected segments; second, we selected cases within those segments; and third, we selected the completes. To calculate weights for the panel cases, we simply adjusted the 2008 design weights  $W_{3NR_{2008XSec}}$  to account for these additional stages of selection. Only completes from 2010 were fielded in 2012. So, with the extra step of nonresponse adjustment,  $WA_{panel2008}$  (WTPAN123) is simply:

$$WA_{panel2008} = \frac{W_{3NR_{2008XSec}}}{(\pi_{segment}^{panel2008} \cdot \pi_{case}^{panel2008})} \cdot \frac{1}{mean_h(\widehat{RP}_{2008 \text{ panel in 2012}})}$$

The final weight WTPAN123 is just these WA variables, rescaled to sum to the number of completed 2008-sample cases in 2012. Cases originated from 2008 would sum to 2008 totals.

To adjust WTPAN123 for nonresponse in 2012 (WTPANNR123), we use logistic regressions to predict a set of response propensity scores for the 2008 panel cases. The independent variables in this regression are factual (not attitudinal) responses collected in 2010. The variables used were: born outside the U.S., living alone, gender, race (white / nonwhite) and Census division. We used

the predicted response propensities from this model to divide the responding and nonresponding cases into five equal size adjustment cells; and, within each cell, we inflated the weights of the responding cases by the inverse of the mean response propensity in that cell.

$$WA_{panel} = \frac{WA_{2008}}{(\pi_{segment}^{panel} \cdot \pi_{case}^{panel})} \cdot \frac{1}{mean_h(\widehat{RP}_{2008 \text{ panel in } 2012})}$$

where h is the nonresponse adjustment cell. This weight was then, again, scaled to sum to the number of completed panel cases.