

# GSS Panel Data Release Notes

Released in July 2011

---

## I. Overview

This GSS panel dataset has three waves of interviews: originally sampled and interviewed in 2006, interviewed for the second time in 2008, and for the third wave in 2010. Among the 4,510 cases interviewed in 2006, we randomly selected 2,000 cases for re-interviews. We re-interviewed 1,536 cases in 2008, and 1,276 in 2010 (see Table 1).

This data file contains those 2,000 respondents who were pre-selected among the 2006 samples and those variables that were asked at least twice in three waves. The only exceptions are listed in section II.3 below.

	2006	2008	2010
1 <sup>st</sup> wave	4510 <sup>a</sup>	2023	2044
2 <sup>nd</sup> wave		1536	1581
3 <sup>rd</sup> wave			1276
Combined N	4510	3559	4901

Note:<sup>a</sup> Of the 4,510 respondents, 2,000 respondents were selected for the 2008 panel (2nd wave).

## 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”, “\_2”, or “\_3” to the existing GSS variable names. For example, EDUC\_1 is the years of education in the first wave (2006), EDUC\_2 is education in the second wave (2008), and EDUC\_3 is education in the third wave (2010).
3. The values of the following variables do not change over waves so they are included as single variables (without \_1, \_2, or \_3): BALLOT, FORM, FORMWT, OVERSAMP, SAMCODE, and SAMPLE.
4. YEAR\_1 is GSS year of the first wave while YEAR\_2 and YEAR\_3 are GSS year of the second wave and the third wave.
5. ID\_1 is the identification number used in the GSS 2006 data, ID\_2 in 2008, and ID\_3 in 2010. ID generally differs across years.

6. PANSTAT\_2 and PANSTAT\_3 indicate 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.
7. For those cases that were not re-interviewed in the second and/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. OLD1 to OLD14, GENDER1 to GENDER14) do not necessarily indicate the same persons over waves. For example, GENDER13\_1 and GENDER13\_2 do not necessarily show the gender of the same household member.
9. Interviewers’ ID numbers (INTID) were newly assigned in each wave. Thus, INTID\_1=56 and INTID\_2=56 do not indicate they are the same interviewer.

### **III. Weights**

Three wave panel data include four different weights, such as WTPAN12, WTPAN123, WTPANNR12, and WTPANNR123. 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 (06-08), weight variables with 123 indicate the three-wave panel, which are 06-08-10 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 all 1,276 cases that completed the panel interview in 2010 that originated from 2006. It accounts for all four stages of selection for the 2006 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 06-08 panel data, you need to use this weight variable (This weight variable is same as the weight variable (wtpan_2) in panel wave 2 data set, which was previously released in 2010)
wtpan123	With the exception of the 06-08 panel analysis, you need to use this weight variable for analyzing three-wave panel data.
Weight Variable Name with NR	This weight has the same case base as WEIGHTpanel2006 and also includes all stages of selection, <b>but also includes a nonresponse adjustment.</b>
wtpannr12	If you want to analyze only 06-08 panel data, you need to use this weight variable (This weight variable is same as the weight variable (wtpan_2) in panel wave 2 data set, which was previously released in 2010)
wtpannr123	With the exception of the 06-08 panel analysis, you need to use this weight variable for analyzing three-wave panel data.

Selection of respondents from the 2006 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 2006 design weights  $W_{3NR_{2006XSec}}^1$  to account for these additional stages of selection. Only completes from 2008 were fielded in 2010. So, with the extra step of nonresponse adjustment,  $WA_{panel2006}$  (WTPAN123) is simply:

$$WA_{panel2006} = \frac{W_{3NR_{2006XSec}}}{\left( \pi_{segment}^{panel2006} \cdot \pi_{case}^{panel2006} \right)} \cdot \frac{1}{mean_h \left( \widehat{RP}_{2006 \text{ Panel in 2008}} \right)}$$

<sup>1</sup> It refers to WTSSNR of 2006 in the cumulative GSS data file. This weight variable is not included in this panel file.

The final weight WTPAN123 is just these WA variables, rescaled to sum to the number of completed 2006 cases in 2010. Cases originated from 2006 would sum to 2006 totals.

To adjust WTPAN123 for nonresponse in 2010 (WTPANNR123), we use logistic regressions to predict a set of response propensity scores for the 2006 panel cases. The independent variables in this regression are factual (not attitudinal) responses collected in 2008<sup>2</sup>. The variables used were: born outside the U.S.<sup>3</sup>, 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.

$$WA_{panel} = \frac{WA_{2006}}{\left(\pi_{segment}^{panel} \cdot \pi_{case}^{panel}\right)} \cdot \frac{1}{mean_h\left(\widehat{RP}_{2006 \text{ Panel in 2010}}\right)}$$

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

#### IV. Miscellaneous

1. The question wordings of the below variables in Wave 2 had references to year 2007, but the Wave 3 questionnaire changed their references to year 2008. In addition, the phrase “last 12 months” in Wave 2 was changed to “last 24 months” in Wave 3 for these variables.
  - HLTH10
  - HRDSHP6
  - HLTH11
  - WORK10
  - FINAN4
  - HRDSHP1
  - LAW5
2. A new category, “Two or more race”, was added to INTETHN\_2 and INTETHN\_3 (interviewers’ race) after the first wave was conducted, because NORC changed the way it keeps information on interviewers in 2007. Thus, this category was not used in INTETHN\_1.

---

<sup>2</sup> In 2008, missing data on the 40 ineligible cases was permitted. In 2010, we used the same independent variables as in 2008 regardless of missing data.

<sup>3</sup> Of all variables studied, this one most strongly predicted response in 2008.

## V. Variables Not Found in the *GSS Cumulative Codebook*

Most of the variables in the panel data are documented in *Codebook* for the GSS cross-sectional cumulative data. The following variables are not included in the *Codebook* because they have never been asked in the cross-section:

- CONTERRR\_1 to CONTERRR\_3
- IMMAFFUS\_1 to IMMAFFUS\_3
- PAYENVIR\_1 to PAYENVIR\_3
- PANSTAT\_2, PANSTAT\_3
- WTPAN<sub>12</sub>, WTPAN<sub>123</sub>
- WTPANNR<sub>12</sub>, WTPANNR<sub>123</sub>

PANSTAT, WTPAN, and WTPANNR were described earlier in this document. This section lists questions and response categories for CONTERRR, IMMAFFUS, and PAYENVIR.

### CONTERRR

How concerned are you about the possibility there will be more major terrorist attacks in the United States? Is that something that worries you a great deal, somewhat, not too much or not at all?

<b>HANDCARD I<sub>2</sub></b>
-------------------------------

A great deal.....	1
Somewhat .....	2
Not too much.....	3
Not at all .....	4
DON'T KNOW.....	DK
REFUSED .....	REF

**IMMAFFUS**

In the long run, do you think that people who are immigrating to the United States today will make American society better, will make American society worse, or do you think that today's immigrants won't affect American society one way or another?

Better.....	1
Worse.....	2
Won't affect American society one way or another.....	3
DON'T KNOW.....	DK
REFUSED.....	REF

**PAYENVIR**

How willing would you be to pay much higher prices in order to protect the environment? Would you say . . .

**HANDCARD I1**

Very willing.....	1
Fairly willing.....	2
Neither willing nor unwilling.....	3
Not very willing, or.....	4
Not willing at all?.....	5
DON'T KNOW.....	DK
REFUSED.....	REF