GSS 2008 Sample Panel Wave 2

Released in January 2012

I. Overview

This GSS panel dataset has two waves of interviews: originally sampled and interviewed in 2008 and for the second wave in 2010. Among the 2,023 cases newly interviewed in 2008, we ended up re-interviewing 1,581 cases in 2010 (see Table 1). This data file contains those 2,023 respondents and those variables that were asked in both waves (the only exceptions are listed in section II.3 below).

<Table 1> GSS design features: Cross-sectional and Panel Components

	GSS Year	
	2008	**2010**
1st wave	2023	2044
2nd wave	1536	1581
3rd wave		1276
Combined N	3559	4901

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 is GSS year of the second wave.

5. ID_1 is the identification number used in the GSS 2008 data and ID_2 in 2010. ID generally differs 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 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.

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_1) would have COHORT_1 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_2 and AGE_3 and his COHORT_2 and COHORT_3 would be 1917 in the second and third waves.

III.   Weights

Three-wave panel data include four different weights, such as WTPAN12 and WTPANNR12. The two different weight variables are differentiated by whether they include NR in the variable name or not 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, weight variables with 12 indicate two-wave panel (08-10) data. Below is the description of two weight variables and methodology used to calculate these weights.

IV.   <Table 2> GSS panel data three wave weight variables

Variable name	Description
Weight Variable Name Without NR	This weight is assigned to all 1,581 cases that completed the panel interview in 2010 that 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 08-10 panel data, 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, **but also includes a nonresponse adjustment.**
wtpannr12	If you want to analyze only 08-10 panel data, you need to use this weight variable.

Selection of respondents from the 2008 round for the panel was also done in two phases: first we selected segments and second we selected cases within those segments. To calculate weights for the panel cases, we simply adjusted the 2008 design weights $W3NR_{2008XSec}$ to account for these additional stages of selection. Then $WA_{panel2008}$ is simply:

$$WA_{panel2008} = \frac{W3NR_{2008XSec}}{\pi_{segment}^{panel2008} \cdot \pi_{case}^{panel2008}}$$

Since the 2008 cross section was not subsampled, we have

$\pi_{segment}^{panel2008} = \pi_{case}^{panel2008} = 1$.

The final weight WTPAN12 is just these WA variables, rescaled to sum to the number of completed 2008 cases in 2010.

        To adjust WTPAN12 for nonresponse in 2010 (WTPANNR12), 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 2008[1]. The variables used were: born outside the U.S.[2], living alone, gender, race (white / nonwhite) and Census division. We used the predicted response propensities from this model to divide the responding and non-

---

[1] 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.
[2] Of all variables studied, this one most strongly predicted response in 2008.

responding 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.

$$WA_{panel} = \frac{WA_{2008}}{\left(\pi_{segment}^{panel} \cdot \pi_{case}^{panel}\right)} \cdot \frac{1}{mean_h\left(\widehat{RP}_{2008 \text{ Panel in } 2010}\right)}$$

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

GSS 2008-sample Wave 2, Release 2 (September 2012)

Variable	Wave affected	Issue	Fix
SOCREL, SOCOMMUN, SOCFREND, SOCBAR	2	These questions did not appear on Ballot C but they were coded to NA (9) anyway.	Recoded to IAP (0)
NUMEMPS	2	The response category "0 employee" was mixed with IAP	"0 employee" and IAP are separated; IAP is now coded to -1 in SPSS
OTHER, OTH16, SPOTHER	All	No value labels	Value labels added
OCC80, SPOCC80, PAOCC80, MAOCC80	All	No value labels	Value labels added
INDUS80, SPIND80, PAIND80, MAIND80	All	No value labels	Value labels added
VSTRATA, VPSU	1	No issue	Newly added. For more information, please refer to a documentation about strata in GSS.