

GSS 2010 Sample Panel Wave 2, Release 1

August 27, 2013

I. Overview

This GSS panel dataset has two waves of interviews: originally sampled and interviewed in 2010 and for the second wave in 2012. Among the 2,044 cases newly interviewed in 2010, we ended up re-interviewing 1,551 cases in 2012 (see Table 1). This data file contains those 2,044 respondents and those variables that were asked in either of the two waves.

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

	GSS Year	
	2010	2012
1 st wave	2044	1974
2 nd wave	1581	1551
3 rd wave	1276	1295
Combined N	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. Thus, each row (case) is unique and some cases are not re-interviewed.
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 (2010), and EDUC_2 is education in the second wave (2012).
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 2010 data and ID_2 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 are labeled in the data).

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, GENDER_{3_1} and GENDER_{3_2} do not necessarily show the gender of the same household member.
9. Interviewers’ ID numbers (INTID) were newly assigned in each wave. Thus, INTID₁=56 and INTID₂=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 AGE₁) would have COHORT₁ 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 AGE₂ and AGE₃ and his COHORT₂ and COHORT₃ would be 1917 in the second and third waves.

III. Weights

The panel data include two different weights: WTPAN₁₂ and WTPANNR₁₂. WTPAN₁₂ is a regular weight variable that does not make a special adjustment to non-response whereas WTPANNR₁₂ is a weight variable that makes a nonresponse adjustment. Below is the description of two 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 2010. It accounts for all four stages of selection for the 2010 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 2010-12 panel data, you need to use this weight variable
Weight Variable Name with "NR"	This weight has the same case base as WEIGHT _{panel2010} and also includes all stages of selection, <i>but also includes a nonresponse adjustment.</i>
wtpannr12	If you want to analyze only 2010-12 panel data, you need to use this weight variable

Selection of respondents from the 2010 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 2010 design weights WEIGHT_{2010XSec} to account for these additional stages of selection. Then, WEIGHT_{panel2010} is simply:

$$WEIGHT_{panel2010} = \frac{WEIGHT_{2010XSec}}{(\pi_{segment}^{panel2010} \cdot \pi_{case}^{panel2010})}$$

The final weight WTPAN12 is just this variable, rescaled to sum to the number of completed 2010-sample cases in 2012.

To adjust WTPAN12 for nonresponse in 2012 (WTPANNR12), we use logistic regressions to predict a set of response propensity (RP) scores for the 2010 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.

$$WEIGHTNR_{panel} = \frac{WA_{2010}}{(\pi_{segment}^{panel} \cdot \pi_{case}^{panel})} \cdot \frac{1}{mean_h(\widehat{RP}_{2010 \text{ panel in 2012}})}$$

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