

SRQM: Datasets

This file listing covers the datasets distributed as part of the **Statistical Reasoning and Quantitative Methods** course taught by François Briatte and Ivaylo Petev.

All datasets are provided in Stata format on an “as-is” basis, with selected documents enclosed in the [docs](#) folder. Please use these datasets for replication purposes only and do not redistribute them for any other reason. Please also remember to systematically cite the source and authors for each dataset.

The usage commands assume that the [SRQM](#) folder has been set as the working directory.

Contents:

- **ebm2009**: Eurobarometer Special Report (2009)
- **ess2008**: European Social Survey (2008)
- **gss2010**: General Social Survey (2010)
- **lobbying2010**: ‘Lobbying and Policy Change’ data (Baumgartner *et al.* 2010)
- **nhis2009**: National Health Interview Survey (2009)
- **qog2011**: Quality of Government (2012)
- **trust2012**: ‘Bread and Peace’ data (Hibbs 2012)
- **wvs2000**: World Values Survey (2000)

ebm2009

The [ebm2009](#) dataset holds data from the Special Eurobarometer, **Ref. 311**, Wave 71.1 (2009), on “Europeans and the Economic crisis”:

Special Eurobarometer reports are based on in-depth thematic studies carried out for various services of the European Commission or other EU Institutions and integrated in Standard Eurobarometer’s polling waves. Reproduction is authorized, except for commercial purposes, provided the source is acknowledged.

http://ec.europa.eu/public_opinion/archives/eb_special_en.htm

Documentation:

- **Survey analysis** by the European Commission (2009)
- **Sample factsheet (descriptive results) for Greece** (2009)

Usage:

```
use "Datasets/ebm2009.dta", clear
```

ess2008

The [ess2008](#) dataset holds Round 4 (2008) of the European Social Survey (ESS):

The European Social Survey (the ESS) is an academically-driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations.

<http://www.europeansocialsurvey.org>

Documentation:

- [ESS description brochure](#)
- [ESS4 questionnaire](#) (2008)
- [ESS4 documentation \(survey design\) report](#) (2011)
- [ESS1 – 4 variable list](#)
- [ESS guide to survey weights](#)

Usage:

```
use "Datasets/ess2008.dta", clear
```

The dataset was created by subsetting the ESS cumulative dataset to Round 4, and then by removing variables with no observations. The code follows:

```
drop if essround != 4

foreach v of varlist * {
    qui count if !mi(`v')
    if r(N)==0 {
        di "Dropping: " "`v'"
        drop `v'
    }
}

save ess2008.dta, replace
```

gss2010

The [gss2010](#) dataset holds data from the U.S. General Social Survey (GSS) for year 2010:

The GSS contains a standard 'core' of demographic, behavioral, and attitudinal questions, plus topics of special interest. Many of the core questions have remained unchanged since 1972 to facilitate time-trend studies as well as replication of earlier findings.

<http://www3.norc.umd.edu/GSS+Website/>

Documentation:

- **GSS variable lists** (online)
- **Guide to weighting GSS data** (GSS Codebook, Appendix A)
- Stata codebook (for quicker access to variable codings; see “Usage” below)

Usage:

```
use "Datasets/gss2010.dta", clear
```

The dataset is a subset from the **GSS 1972-2010 cross-sectional cumulative dataset** (Release 1.1, Feb. 2011). The code used to trim the data and to export the Stata codebook follows:

```
// Subset.
keep if year==2010

// Remove empty variables.
foreach v of varlist * {
    qui su `v'
    if r(N)==0 {
        di "Removing " "`v'"
        drop `v'
    }
}
save gss2010.dta, replace

// Export codebook.
log using gss2010_codebook.log, name(gss2010) replace
d
codebook
log close gss2010
```

lobbying2010

The `lobbying2010` dataset contains the variables required to replicate part of the analysis published in Frank R. Baumgartner, Jeffrey M. Berry, Marie Hojnacki, David C. Kimball and Beth L. Leech, ***Lobbying and Policy Change. Who Wins, Who Loses, and Why*** (University of Chicago Press, 20009).

Documentation:

- **Advocacy and Public Policymaking** (online)
- **Issue-level codebook** (2010)

The issue-level dataset is used in Chapter 5 of the book.

nhis2009

The [nhis2009](#) dataset holds data for years 2000—2009 of the U.S. National Health Interview Survey (NHIS):

The National Health Interview Survey (NHIS) has monitored the health of the nation since 1957. NHIS data on a broad range of health topics are collected through personal household interviews. For over 50 years, the U.S. Census Bureau has been the data collection agent for the National Health Interview Survey. Survey results have been instrumental in providing data to track health status, health care access, and progress toward achieving national health objectives.

<http://www.cdc.gov/nchs/nhis.htm>

Documentation:

- **Description brochure** (2010)
- **Description of NHIS 2009** (2010)

Usage:

```
use "Datasets/nhis2009.dta", clear
```

If you want to use the data with individual weights, the code should probably look like this:

```
svyset psu [pweight=perweight], strata(strata) vce(linearized)  
singleunit(missing)
```

qog2011

The [qog2011](#) dataset holds the Quality of Government (QOG) Standard dataset in its most recent revision of April 6, 2011:

Our research addresses the questions of how to create and maintain high quality government institutions and how the quality of such institutions influences public policy in a broader sense.

<http://www.qog.pol.gu.se/>

Documentation:

- **QOG Annual report** (2011)
- **QOG Standard codebook** (2011)

Usage:

```
use "Datasets/qog2011.dta", clear
```

A simpler version of the dataset, **QOG Basic**, is also available for students who prefer to work on a more accessible and less extensive version of the data.

trust2012

The [trust2012](#) dataset contains the variables required to replicate Douglas Hibbs' 'Bread and Peace' model of the U.S. presidential election:

Aggregate votes for president in postwar elections are well explained by just two fundamental determinants: (1) weighted-average growth of per capita real disposable personal income over the term, and (2) cumulative US military fatalities owing to unprovoked, hostile deployments of American armed forces in foreign conflicts. No other objectively measured, exogenous factor systematically affects postwar aggregate votes for president.

<http://www.douglas-hibbs.com/Election2012/2012Election-MainPage.htm>

Documentation:

- Douglas Hibbs, "**The 2012 US Presidential Election: Implications of the 'Bread and Peace' Model For Obama's Re-election Prospects as of 2011:q4**", 29 February 2012 (with **slides** material).

Usage:

```
use "Datasets/trust2012.dta", clear
```

The data were kindly provided by John Sides, following his blog post "**What Will Make People Trust Government Again?**" published at *The Monkey Cage* on February 14, 2010.

wvs2000

The [wvs2000](#) dataset holds data from the 1999-2004 wave of the World Values Survey (WVS):

The World Values Survey (WVS) is a worldwide network of social scientists studying changing values and their impact on social and political life. The WVS in collaboration with EVS (European Values Study) carried out representative national surveys in 97 societies containing almost 90 percent of the world's population. These surveys show pervasive changes in what people want out of life and what they believe. In order to monitor these changes, the EVS/WVS has executed five waves of surveys, from 1981 to 2007.

<http://www.worldvaluessurvey.org/>

Documentation:

- **Full questionnaire of all WVS waves** (1981-2008)
- **Questionnaire for Egypt** (2000)

Usage:

```
use "Datasets/wvs2000.dta", clear
```

The data were extracted from the WVS cumulative dataset. The code to capitalize country names was kindly provided at **StackOverflow** by William A. Huber.