

SRQM: Data sources

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 `dta` Stata format on an “as-is” basis. Please use these datasets for replication purposes only and do not redistribute them for any other reason.

Data sources:

Filename	Data	Year of release
ess2008	European Social Survey	2008
gss2010	General Social Survey	2010
nhis2009	National Health Interview Survey	2009
qog2011	Quality of Government	2011
wvs2000	World Values Survey	2000

Remember to **systematically cite the data source and authors** after retrieving the bibliographic information from the source and formatting it into an academic citation.

Usage commands:

The Stata `use` commands included at the end of each dataset description require that the `SRQM` folder has been set as the working directory, as explained in its [README](#) file.

Students are encouraged to use the `lookfor` and `lookfor_all` commands to explore the course datasets, and to weight the data by using the `svyset` command with the appropriate weight.

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>

Additional documentation:

<http://ess.nsd.uib.no/>

Variables that appear in the documentation but not in the dataset are part of the overall ESS rotating module (thanks to the ESS data team for the precision).

Usage in Stata:

```
use "Datasets/ess2008.dta", clear
svyset [aw=dweight]
```

First multiply `dweight` by `pweight` if you are working on combined countries or averages. See the [ESS weighting guide](#) for more details.

Technical note:

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/>

Additional documentation:

To extract a quick codebook out of the dataset, run the following:

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

Usage in Stata:

```
use "Datasets/gss2010.dta", clear
svyset sampcode [pw=wtssnr]
```

There is no stratum variable in the GSS dataset, but the sampling design does include stratification. See [Appendix A](#) of the GSS documentation for more details.

Technical note:

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 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
```

nhis2009

The `nhis2009` dataset holds sample adult 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>

Usage in Stata:

```
use "Datasets/nhis2009.dta", clear  
svyset psu [pweight=perweight], strata(strata)
```

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/>

Usage in Stata:

```
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.

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/>

Usage in Stata:

```
use "Datasets/wvs2000.dta", clear  
svyset [aw=s017]
```

Use `s017a` if you are using country split-ups. See the [WVS weighting guide](#) for more details.

Technical note:

The data were extracted from the WVS cumulative dataset. The code to capitalize country names is available as the `proper_labels` command (see the [README](#) file of the `Programs` folder).
