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.

Please also remember to **systematically cite the data source and authors**, after retrieving the bibliographic information from the source and formatting it as required by academic standards.

Data sources:

Filename	Data	Author/Year of release
ebm2009	Eurobarometer Special Report	2009
ess2008	European Social Survey	2008
gss2010	General Social Survey	2010
lobbying2010	'Lobbying and Policy Change'	(Baumgartner *et al.* 2010)
nhis2009	National Health Interview Survey	2009
qog2011	Quality of Government	2011
trust2012	'Bread and Peace'	(Hibbs 2012)
wvs2000	World Values Survey	2000

A selection of other sources is shown at the end of this document. Please remember that other sources are allowed for research projects *if and only if you can quickly show skills in data management.*

Usage commands:

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

---

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 thematical 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

Additional documentation:

http://www.gesis.org/en/eurobarometer/home/

Usage in Stata:

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

The sample can be weighted with the `v8` and `v38` variables.

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/

Usage in Stata:

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

The sample can be weighted with the `dweight` and `pweight` variables.

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
```

Variables that appear in the documentation but not in the dataset are part of the overall ESS rotating module.

---

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.org/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
```

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
```

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).

Additional documentation:

http://lobby.la.psu.edu/

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

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
```

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/

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.

---

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

Additional 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 in Stata:

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

The data were kindly provided by John Sides, following his blog post "What Will Make People Trust Goverment 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/

Usage in Stata:

```
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:

```
* Capitalize country names.
local varname v2
local sLabelName: value label `varname'
di "`sLabelName'"
levelsof `varname', local(xValues)
foreach x of local xValues {
    local sLabel: label (`varname') `x', strict
    local sLabelNew = proper("`sLabel'")
    if "`sLabelNew'"=="Usa" {
        local sLabelNew = upper("`sLabel'")
    }
    noi di "`x': `sLabel' ==> `sLabelNew'"
    label define `sLabelName' `x' "`sLabelNew'", modify
}
```

Sources

The following list was inspired by Robert J. Franzese, Jr., "Empirical Strategies for Various Manifestations of Multilevel Data," *Political Analysis* 13:430–446, 2005; Paul Pennings *et al.*, *Doing Research in Political Science: An Introduction to Comparative Methods and Statistics*, Sage, 2005; Berenica Vejvoda, "Selected Data Resources for Political Science Research," UC San Diego Social Science Data Services; Emiliano Grossman and Nicolas Sauger; friends, students, colleagues, etc.

More sources can be identified using websites like CKAN, the Council of European Social Science Data Archives, DataMarket, EUROLAB and GESIS, IPSAportal, the Macro Data Guide and data libraries from your home universities (see e.g. the Data Libary at the University of Edinburgh).

Intergovernmental organizations

- Eurostat
- International Monetary Fund (IMF)
- World Bank (World Development Indicators)
- OECD (OECD Health, OECD Social Expenditure)
- ILO (labour force surveys)
- United Nations (Statistics Division, Millennium Development Goals)
- World Health Organization (WHO) (Global Health Observatory, WHO Health for All Database); Gapminder (excellent graphing)

Nongovermental organizations

- Freedom House (freedom of the press)
- Reporters Without Borders (Press Freedom Index)
- Transparency International (corruption perception indexes)

National providers

- United Kingdom:
    - Office of National Statistics (ONS)
    - UK Data Archive, browsable at ESDS; see e.g. British Social Attitudes (BSA) or British Household Panel Survey (BHPS)
    - Survey Question Bank (ESRC Survey Resources Network)
- United States:
    - Census Bureau
    - General Social Survey (GSS)
    - American National Election Studies (ANES)
    - Federal Reserve (U.S. economic time series)

- - [Pew Research Center for the People & the Press](#)
  - [Pew Research Center's Social & Demographic Trends](#)
  - [Cultural Policy & the Arts National Data Archive](#) (CPANDA)
- France:
  - [INSEE](#) (very large and detailed data)
  - [Réseau Quételet](#) (strict access conditions)

Comparative politics

- [Parties and elections](#) (elections and political parties in Europe since 1945)
- [Comparative political data sets](#) (politics and expenditures in OECD and European countries)
- [ICPSR](#) (includes comparative datasets used by American political scientists)
- [IPU](#) (parliaments and electoral systems)
- [Worldwide Elections](#) (electoral competition and outcomes)
- [Comparative Study of Electoral Systems](#)
- [Comparative Manifestos](#) (pools parties' election manifestos by election across several democracies)
- [Quality of Government](#) (used in course)
- [Polity IV](#) (regime characteristics and transitions)
- [Comparative Policy Agendas](#) (initially an [American project](#))

Comparative public opinion

- [Afrobarometers](#)
- [Eurobarometers](#) (EB)
- [European Social Survey](#) (ESS) (used in course)
- [European Values Surveys](#) (EVS) — see [download guide](#)
- [International Social Survey Program](#) (ISSP)
- [World Values Surveys](#) (WVS)

International relations

- [SIPRI](#) (Stockholm International Peace Research Institute; world armament and disarmament)
- [Correlates of War](#) (includes Militarized Interstate Disputes)