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 files are provided in `dta` Stata format on an "as-is" basis: please use them for teaching purposes only, and do not redistribute them for any 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

Coursework:

You are expected to use the `lookfor` and `lookfor_all` commands to explore the course datasets, in order to select a set of variables for analysis in your research project.

Bibliographic information and additional documentation is available from the online data sources. Once you have identified a dataset, retrieve that information to describe and reference the data in your paper.

Indications on weighting the data with the `syvset` command are provided for point estimation if you need to estimate a particular statistic with high precision.

The Stata `use` commands mentioned in this document require that the `SRQM` folder has been set as the working directory, as explained in its [README](#) file.

---

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

Survey weights:

```
use datasets/ess2008, clear
svyset [aw=dweight]
```

See the [ESS weighting guide](#) for instructions.

Production notes:

The dataset was created by subsetting the [ESS cumulative dataset](#) to Round 4, and then by removing variables with no observations. The codebook was downloaded from the [ESS data](#) website.

---

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/

Survey weights:

```
use datasets/gss2010, clear
svyset vpsu [pw=wtssall], strata(vstrata)
```

See "Calculating Design-Corrected Standard Errors for the General Social Survey, 1988-2010" by Steven Pedlow and Appendix A of the GSS codebook for other options.

Production notes:

The dataset is a trimmed-down version of the GSS 2010 cross-sectional dataset (Release 2, Feb. 2012). A mock codebook is bundled with the data.

---

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/

Production notes:

The data and codebook come from the QOG Standard download page. 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 Wave 4 (1999-2004) 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/

Survey weights:

```
use datasets/wvs2000, clear
svyset [aw=s017]
```

See the WVS weighting guide for other options.

Production notes:

The data come from the WVS 2000 official file found at the WVS website. This version had encoding issues that are used as examples in class for recode commands. The cumulative dataset has different variable names and proper variable encoding. More recent data is also available up to Wave 6 (2010-2013). A mock codebook is bundled with the data.