

Package: gssr (via r-universe)

November 11, 2024

Title US General Social Survey (GSS) Data for R

Version 0.6

Description The General Social Survey (GSS) is a long-running, mostly annual survey of US households. It is administered by the National Opinion Research Center (NORC). This package contains the GSS Cumulative Data and GSS Panel Data files packaged for R. Its companion package, gssrdoc, provides the codebook integrated into R's help system For more information on the GSS see [\url{http://gss.norc.org}](http://gss.norc.org).

License MIT + file LICENSE

Depends R (>= 4.0.0)

Imports curl, dplyr, fs, haven, here, magrittr, tibble

Suggests drat, forcats, ggplot2, scales, stringr, survey, srvyr, tidy, knitr, rmarkdown

Encoding UTF-8

LazyData false

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

URL <https://github.com/kjhealy/gssr>

BugReports <https://github.com/kjhealy/gssr/issues>

VignetteBuilder knitr

Config/pak/sysreqs make libssl-dev libx11-dev zlib1g-dev

Repository <https://kjhealy.r-universe.dev>

RemoteUrl <https://github.com/kjhealy/gssr>

RemoteRef HEAD

RemoteSha 541cd79c069ba3a074922b695dad26bf2e6751ed

Contents

gss_all	2
gss_get_yr	2
gss_panel06_long	3
gss_panel08_long	4
gss_panel10_long	5
gss_panel20	5
gss_sub	6
gss_which_years	7
Index	8

gss_all	<i>General Social Survey Survey Cumulative Data File 1972-2022 R2a</i>
---------	--

Description

A tibble containing Release 2a of the GSS Cumulative Data (1972-2022) file.

Usage

```
data(gss_all)
```

Format

An object of class `tibble` with 72,390 rows and 6,694 columns. Variables are encoded as labelled vectors. The GSS Codebook is the authoritative source for the variables in this dataset. It is available at <http://gss.norc.org/Get-Documentation>. Summary information is available in `gss_doc`, a tibble supplied with this package.

Source

National Opinion Research Center, <http://gss.norc.org>.

gss_get_yr	<i>Download GSS data file for a single year from NORC</i>
------------	---

Description

Use `gss_get_yr()` to get GSS data for a single year from NORC's GSS website (where it is available as a zipped Stata file) and put it directly into a tibble.

Usage

```
gss_get_yr(  
  year = 2022,  
  url = "https://gss.norc.org/documents/stata/",  
  fname = "_stata",  
  ext = "zip",  
  dest = "data-raw/",  
  save_file = c("n", "y")  
)
```

Arguments

year	The desired GSS survey year, as a number (i.e., not in quotes). Defaults to 2022.
url	Location of the file. Defaults to the current NORC URL for Stata files.
fname	Non-year filename component. Defaults to '_stata'. Usually should not be changed.
ext	File name extension. Defaults to 'zip'. Usually should not be changed.
dest	If save_file is "y", the directory to put the file in. Defaults to data-raw in current directory.
save_file	Save the data file as well as loading it as an object. Defaults to 'n'.

Value

A tibble with the requested year's GSS data.

Examples

```
gss80 <- gss_get_yr(1980)
```

gss_panel06_long

General Social Survey Survey 2006 Three Wave Panel Data

Description

A tibble containing the General Social Survey 2006 Three Wave Panel Data File, in long format.

Usage

```
data(gss_panel06_long)
```

Format

A tibble with 6,000 rows and 1,572 columns. Variables are encoded as numerics or factors. The GSS Codebook is the authoritative source for the variables in this dataset. It is available at <http://gss.norc.org/Get-Documentation>. Summary information is available in `gss_panel_doc`, a tibble supplied with this package. Respondent ids are contained in the variable `firstid` (from the GSS `id_1` variable). Survey waves (years 2006, 2008, 2010) are contained in the `wave` variable as 1, 2, and 3. See also the `gss_panel_doc` object in this package.

Source

National Opinion Research Center, <http://gss.norc.org>.

`gss_panel08_long`

General Social Survey Survey 2008 Three Wave Panel Data

Description

A tibble containing the General Social Survey 2008 Three Wave Panel Data File, in long format.

Usage

```
data(gss_panel08_long)
```

Format

A tibble with 6,069 rows and 1,243 columns. Variables are encoded as as numerics or factors. The GSS Codebook is the authoritative source for the variables in this dataset. It is available at <http://gss.norc.org/Get-Documentation>. Summary information is available in `gss_panel_doc`, a tibble supplied with this package. Respondent ids are contained in the variable `firstid` (from the GSS `id_1` variable). Survey waves (years 2008, 2010, 2012) are indicated by the `wave` variable as 1, 2, and 3. See also the `gss_panel_doc` object in this package.

Source

National Opinion Research Center, <http://gss.norc.org>.

`gss_panel10_long`*General Social Survey Survey 2010 Three Wave Panel Data*

Description

A tibble containing the General Social Survey 2010 Three Wave Panel Data File, in long format.

Usage

```
data(gss_panel10_long)
```

Format

A tibble with 6,132 rows and 1,191 columns. Variables are encoded as numerics or factors. The GSS Codebook is the authoritative source for the variables in this dataset. It is available at <http://gss.norc.org/Get-Documentation>. Summary information is available in `gss_panel_doc`, a tibble supplied with this package. Respondent ids are contained in the variable `firstid` (from the GSS `id_1` variable). Survey waves (years 2010, 2012, 2014) are indicated by the wave variable as 1, 2, and 3. See also the `gss_panel_doc` object in this package.

Source

National Opinion Research Center, <http://gss.norc.org>.

`gss_panel20`*General Social Survey Survey 2020 Panel Data*

Description

A tibble containing the General Social Survey 2020 Panel Data File, in wide format.

Usage

```
data(gss_panel20)
```

Format

A tibble with 5,215 rows and 4,296 columns. Variables are encoded as labelled vectors. The GSS Codebook is the authoritative source for the variables in this dataset. It is available at <http://gss.norc.org/Get-Documentation>. Due to the COVID-19 pandemic, in 2020 the conducted the GSS was conducted as two studies: (1) a panel re-interview of past respondents from the 2016 and 2018 cross sectional GSS studies (referred to as the 2016-2020 GSS Panel), and (2) an independent fresh cross-sectional address-based sampling push to web study (referred to as 2020 cross-sectional survey). This data object is for the first study; namely, the study empaneling former 2016 and 2018 GSS respondents to answer a GSS questionnaire in 2020 (i.e., the 2016-2020 GSS panel).

This data focuses on Wave 2 of the 2016-2020 GSS Panel – i.e. the panel reinterviews with 2018 GSS respondents and a randomly selected subset of 2016 GSS respondents. The GSS has used a panel format previously, as parts of the 2006-2014 GSS. In the 2016-2020 GSS Panel, variables only contain data from one of the three years. To differentiate between versions of each variable, they have been appended with suffixes. Variables from 2016 (Wave 1a) have `_1a` appended, variables from 2018 (Wave 1b) have `_1b` appended, and variables from 2020 (Wave 2) have `_2` appended. Users can also track cases from 2016 and 2018, and reinterviews from 2020 with the variable `~samptype``.

Because of its relatively complex nature, users are strongly encouraged to consult the official [GSS documentation for this dataset](<https://gss.norc.org/Documents/codebook/2016-2020%20GSS>).

Source

National Opinion Research Center, <http://gss.norc.org>.

gss_sub

Example subset of the GSS Cumulative Data File 1972-2022

Description

A tibble containing just a few variables from the GSS Cumulative Data File. See <http://gss.norc.org/Get-Documentation> for full documentation of the variables.

Usage

```
data(gss_sub)
```

Format

A tibble with 72,390 rows and 19 columns.

`year` Year of the survey.

`id` Respondent id.

`ballot` Survey ballot

`age` Age of respondent

`race` Race of respondent

`sex` Sex of respondent

`degree` Highest level of education obtained

`padeg` Father's education

`padeg` Mother's education

`relig` Religion (simple coding)

`Polviews` Political views

`fefam` Response to a statement that it is better for man to go out to work, and for a woman to tend the home

vpsu Variance primary sampling unit
 vstrat Variance stratum
 oversamp Weights for black oversamples
 formwt Survey weight for experimental randomization
 wtssall Survey weight (1972-2018)
 wtssps Poststratification survey weight (1972-2022)
 sampcode Sampling error code
 sample Sampling frame and method

Source

National Opinion Research Center, <http://gss.norc.org>.

gss_which_years	<i>gss_which_years</i>
-----------------	------------------------

Description

See which years a particular question was asked in the GSS.

Usage

```
gss_which_years(data, variable)
```

Arguments

data	A tibble of data, usually gss_all
variable	The variable or variables we want to check. Provide variables in tidyselect style, i.e. unquoted, and for multiple variables enclose unquoted in c()

Details

What years was a particular question asked in the GSS?

Value

A tibble showing whether the question or questions were asked in each of the GSS years

Examples

```
## Not run:
data(gss_all)
gss_all %>%
  gss_which_years(fefam)

gss_all %>%
  gss_which_years(c(industry, indus80, wrkgovt, commute))

## End(Not run)
```

Index

* datasets

- gss_all, [2](#)
- gss_panel06_long, [3](#)
- gss_panel08_long, [4](#)
- gss_panel10_long, [5](#)
- gss_panel20, [5](#)
- gss_sub, [6](#)

- gss_all, [2](#)
- gss_get_yr, [2](#)
- gss_panel06_long, [3](#)
- gss_panel08_long, [4](#)
- gss_panel10_long, [5](#)
- gss_panel20, [5](#)
- gss_sub, [6](#)
- gss_which_years, [7](#)