Package ‘wid’

February 20, 2026

Type Package

Title Download Data from the World Inequality Database

Version 0.0.1

Author Thomas Blanchet [aut],
Ignacio Flores [cre]

Maintainer Ignacio Flores <stats@wid.world>

Description Tools to download data from the online World Inequality Database directly into R. The World Inequality
Database is an extensive source on the historical evolution of the
distribution of income and wealth both within and between countries.
It relies on the combined effort of an international network of over a
hundred researchers covering more than seventy countries from all continents.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

Depends R (>= 2.10)

Imports httr (>= 1.2.1), base64enc (>= 0.1), plyr (>= 1.8.4), jsonlite (>= 1.6.1)

Suggests testthat (>= 1.0.2), knitr (>= 1.16), rmarkdown (>= 1.6), dplyr (>= 1.0.0), ggplot2 (>= 2.2.1), scales (>= 0.4.1), tidyverse (>= 1.1.1)

NeedsCompilation no

Repository CRAN

Date/Publication 2026-02-20 11:20:02 UTC

Contents

`check_ages` *Check list of age codes*

Description

Check that the list of age codes submitted by the user is valid.

Usage

```
check_ages(ages)
```

Arguments

`ages`	List of age codes

Author(s)

Thomas Blanchet

`check_areas` *Check list of area codes*

Description

Check that the list of area codes submitted by the user is valid.

Usage

```
check_areas(areas)
```

Arguments

`areas`	List of area codes

Author(s)

Thomas Blanchet

`check_indicators` *Check list of indicator codes*

Description

Check that the list of indicator codes submitted by the user is valid.

Usage

```
check_indicators(indicators)
```

Arguments

`indicators`	List of indicators.

Author(s)

Thomas Blanchet

`check_perc` *Check list of percentiles*

Description

Check that the list of percentiles submitted by the user is valid

Usage

```
check_perc(perc)
```

Arguments

`perc`	List of percentiles

Author(s)

Thomas Blanchet

`check_pop` *Check list of population codes*

Description

Check that the list of population codes submitted by the user is valid.

Usage

```
check_pop(pop)
```

Arguments

`pop`	List of population codes

Author(s)

Thomas Blanchet

`check_years` *Check list of years*

Description

Check that the list of years submitted by the user is valid

Usage

```
check_years(years)
```

Arguments

`years`	List of years

Author(s)

Thomas Blanchet

download_wid *Download data from WID.world*

Description

Downloads data from the World Wealth and Income Database (https://wid.world) into a `data.frame`. Type `vignette("wid-demo")` for a detailed presentation.

Usage

```
download_wid(
  indicators = "all",
  areas = "all",
  years = "all",
  perc = "all",
  ages = "all",
  pop = "all",
  metadata = FALSE,
  include_extrapolations = TRUE,
  verbose = FALSE
)
```

Arguments

`indicators` List of six-letter strings, or `"all"`: code names of the indicators in the database. Default is `"all"` for all indicators. See 'Details' for more.

`areas` List of strings, or `"all"`: area code names of the database. `"XX"` for countries/regions, `"XX-YY"` for infra-national regions. `"XX-YYY"` for supra-national regions. Default is `"all"` for all areas. See 'Details' for more.

`years` Numerical vector, or `"all"`: years to retrieve. Default is `"all"` for all years.

`perc` List of strings, or `"all"`: percentiles take the form `"pXX"` or `"pXXpYY"`. Default is `"all"` for all percentiles. See 'Details' for more.

`ages` Numerical vector, or `"all"`: age category codes in the database. 999 for all ages, 992 for adults. Default is `"all"` for all age categories. See 'Details' for more.

`pop` List of characters, or `"all"`: type of population. `"t"` for tax units, `"i"` for individuals. Default is `"all"` for all population types. See 'Details' for more.

`metadata` Should the function fetch metadata too (ie. variable descriptions, sources, methodological notes, etc.)? Default is `FALSE`.

`include_extrapolations`
Should the function return estimates that are the results of extrapolations and interpolations based on limited data? Default is `TRUE`.

`verbose` Should the function indicate the progress of the request? Default is `FALSE`.

Details

Although all arguments default to "all", you cannot download the entire database by typing download_wid(). The command requires you to specify either some indicators or some areas. **To download the entire database, please visit https://wid.world/data/ and choose "download full dataset".**

If there is no data matching you selection on WID.world (maybe because you specified an indicator or an area that doesn't exist), the command will return NULL with a warning.

All monetary amounts for countries and country subregions are in constant local currency of the reference year (i.e. the previous year, the database being updated every year around July). Monetary amounts for world regions are in EUR PPP of the reference year. You can access the price index using the indicator inyixx, the PPP exchange rates using xlcusp (USD), xlceup (EUR), xlcyup (CNY), and the market exchange rates using xlcusx (USD), xlceux (EUR), xlcyux (CNY). To check the current reference year, you can look at when the price index is equal to 1.

Shares and wealth/income ratios are given as a fraction of 1. That is, a top 1% share of 20% is given as 0.2. A wealth/income ratio of 300% is given as 3.

The arguments of the command follow a nomenclature specific to WID.world. We provide more details with a few examples below. **For the complete up-to-date documentation of the structure of the database, please visit https://wid.world/codes-dictionary/.**

Indicators: The argument indicators is a vector of 6-letter codes that corresponds to a given series type for a given income or wealth concept. The first letter correspond to the type of series. Some of the most common possibilities include:

one-letter code	description
a	average (local currency unit, last year's prices)
b	inverted Pareto-Lorenz coefficient
f	female population (fraction between 0 and 1)
g	Gini coefficient (between 0 and 1)
i	index
n	population
s	share (fraction between 0 and 1)
t	threshold (local currency unit, last year's prices)
m	total (local currency unit, last year's prices)
p	proportion of women (fraction between 0 and 1)
w	wealth-to-income ratio or labor/capital share (fraction of national income)
r	Top 10/Bottom 50 ratio
x	exchange rate (market or PPP)
e	Total emissions (tons of CO2 equivalent emissions)
k	Per capita emissions (tons of CO2 equivalent emissions)
l	Average per capita group emissions (tons of CO2 equivalent per capita emissions)

The next five letters correspond a concept (usually of income and wealth). Some of the most common possibilities include:

five-letter code	description
ptinc	pre-tax national income

pllin	pre-tax labor income
pkkin	pre-tax capital income
fiinc	fiscal income
hweal	net personal wealth

For example, sfiinc corresponds to the share of fiscal income, ahweal corresponds to average personal wealth. If you don't specify any indicator, it defaults to "all" and downloads all available indicators. Check https://wid.world/codes-dictionary/ for a full list of codes.

Area codes: All data in WID.world is associated to a given area, which can be a country, a region within a country, an aggregation of countries (eg. a continent), or even the whole world. The argument areas is a vector of codes that specify the areas for which to retrieve data. Countries and world regions are coded using 2-letter ISO codes. Country subregions are coded as XX-YY where XX is the country 2-letter code. If you don't specify any area, it defaults to "all" and downloads data for all available areas.

Years: All data in WID.world correspond to a year. Some series go as far back as the 1800s. The argument years is a vector of integer that specify those years. If you don't specify any year, it defaults to "all" and downloads data for all available years.

Percentiles: The key feature of WID.world is that it provides data on the whole distribution, not just totals and averages. The argument perc is a vector of strings that indicate for which part of the distribution the data should be retrieved. For share and average variables, percentiles correspond to percentile ranges and take the form pXXpYY. For example the top 1% share correspond to p99p100. The top 10% share excluding the top 1% is p90p99. Thresholds associated to the percentile group pXXpYY correspond to the minimal income or wealth level that gets you into the group. For example, the threshold of the percentile group p90p100 or p90p91 correspond to the 90% quantile. Variables with no distributional meaning use the percentile p0p100. If you don't specify any percentile, it defaults to "all" and downloads data for all available parts of the distribution.

Age groups: Data may only concern the population in a certain age group. The argument ages is a vector of age codes that specify which age categories to retrieve. Ages are coded using 3-digit codes. Some of the most common possibilities include:

three-digit code	**description**
999	all ages
014	ages 0 to 14
156	ages 15 to 64
997	ages 65 and older
991	ages below 20
992	ages 20 and older

If you don't specify any age, it defaults to "all" and downloads data for all available age groups. Visit https://wid.world/codes-dictionary/ for a comprehensive list of options.

Population types: The data in WID.world can refer to different types of population (i.e. different statistical units). The argument pop is a vector of population codes. They are coded using one-letter codes. Some of the most common possibilities include:

one-letter code	description
i	individuals
j	equal-split adults (i.e., income or wealth divided equally among spouses)
m	male
f	female
t	tax unit
e	employed

If you don't specify any code, it defaults to `"all"` and downloads data for all available populations.

Extrapolations/interpolations: Some of the data on WID.world is the result of interpolations (when data is only available for a few years) or extrapolations (when data is not available for the most recent years) that are based on much more limited information that other data points. We include these interpolations/extrapolation by default as a convenience, and also because these values are used to perform regional aggregations. Yet we stress that these estimates, especially at the level of individual countries, can be fragile.

For many purposes, it can be preferable to exclude these data points. For that, use the option `include_extrapolations = FALSE`.

Value

A `data.frame` with the following columns:

`country` The country or area code.

`variable` The variable name, which combine the indicator, the age code and the population code.

`percentile` The part of the distribution the value relates to.

`year` The year the value relates to.

`value` The value of the indicator.

If you specify `metadata = TRUE`, the `data.frame` also has the following columns:

`countryname` The full name of the country/region.

`shortname` A short version of the variable full name in plain english.

`shortdes` A description of the type of series.

`pop` The population type, in plain english.

`age` The age group, in plain english.

`source` The source for the data.

`method` Methodological notes, if any.

`imputation` Type of estimate (when applicable). The `imputation` field is a short qualitative description of the type of estimate provided, which is strongly related to data quality. For technical details, see the `method` field and papers cited in `source`.

`quality` Data quality (when applicable). The `quality` field is a score from 0 to 5 indicating the quality of the data.

Author(s)

Thomas Blanchet, with updates by Ignacio Flores

`environment` *Get variables associated to a list of area codes*

Description

Package API environment selector.

Usage

```
environment
```

Format

An object of class `character` of length 1.

Author(s)

Thomas Blanchet

`get_data_variables` *Get data associated to a list of variables*

Description

Perform GET request to the server to retrieve data associated to a list of variables.

Usage

```
get_data_variables(areas, variables, no_extrapolation = FALSE)
```

Arguments

`areas` List of area codes.

`variables` List of variables, of the form: `"xxxxxx_pXXpYY_999_i"`

`no_extrapolation` Logical: should interpolated/extrapolated years be included or not?

Author(s)

Thomas Blanchet

`get_metadata_variables` *Get metadata associated to a list of variables*

Description

Perform GET request to the server to retrieve metadata associated to a list of variables.

Usage

```
get_metadata_variables(
  areas,
  variables,
  report_missing = TRUE,
  collected_metadata = NULL
)
```

Arguments

`areas` List of area codes.

`variables` List of variables, of the form: `"xxxxxx_pXXpYY_999_i"`

`report_missing` Logical: report any missing metadata when set to TRUE.

`collected_metadata`
List used to accumulate missing metadata across calls.

Author(s)

Thomas Blanchet

Index