

Package ‘getTBinR’

September 28, 2018

Type Package

Version 0.5.5

Title Get WHO Tuberculosis Data

Description Facilitates easy import of analysis ready World Health Organisation Tuberculosis data and provides plotting functions for exploratory data analysis.

License GPL-3 | file LICENSE

VignetteBuilder knitr

URL <https://www.samabbott.co.uk/getTBinR>,
<https://github.com/seabbs/getTBinR>

BugReports <https://github.com/seabbs/getTBinR/issues>

Encoding UTF-8

RoxygenNote 6.1.0

LazyData true

Depends R (>= 2.10)

Suggests devtools, testthat, vdiff, covr, pkgnet, rmarkdown, knitr,
tidyverse, shiny, shinydashboard, shinyWidgets, shinycssloaders

Imports data.table, tibble, dplyr, tidyr, purrr, ggplot2, magrittr,
plotly, ggthemes, viridis, scales

NeedsCompilation no

Author Sam Abbott [aut, cre] (<<https://orcid.org/0000-0001-8057-8037>>)

Maintainer Sam Abbott <contact@samabbott.co.uk>

Repository CRAN

Date/Publication 2018-09-28 12:00:09 UTC

R topics documented:

<code>get_data</code>	2
<code>get_data_dict</code>	3
<code>get_tb_burden</code>	4

map_tb_burden	5
plot_tb_burden	8
plot_tb_burden_overview	10
prepare_df_plot	12
run_tb_dashboard	14
search_data_dict	14
summarise_tb_burden	15
who_shapefile	18

Index 19

get_data	<i>Generic Get Data Function</i>
----------	----------------------------------

Description

If the data is found locally in the temporary directory then this function will load the data into R. Otherwise if `download_data = TRUE` then the data will be retrieved from the specified URL. Data can then be saved to the temporary directory by specifying `save = TRUE`.

Usage

```
get_data(url = NULL, data_trans_fn = NULL, download_data = TRUE,
  save = TRUE, save_name = NULL, return = TRUE, verbose = TRUE,
  use_utils = FALSE, use_direct_download = FALSE,
  retry_download = TRUE)
```

Arguments

<code>url</code>	Character string, indicating the url of the data to download.
<code>data_trans_fn</code>	Function that takes a <code>data.table</code> as input and returns a single dataframe of any type. If not specified defaults to transforming the data into a tibble.
<code>download_data</code>	Logical, defaults to <code>TRUE</code> . If not found locally should the data be downloaded from the specified URL?
<code>save</code>	Logical, should the data be saved for reuse during the current R session. Defaults to <code>TRUE</code> . If <code>TRUE</code> then the data is saved to the temporary directory specified by tempdir .
<code>save_name</code>	Character string, name to save the data under. Defaults to <code>NULL</code> .
<code>return</code>	Logical, should the data be returned as a dataframe. Defaults to <code>TRUE</code> .
<code>verbose</code>	Logical, defaults to <code>TRUE</code> . Should additional status and progress messages be displayed.
<code>use_utils</code>	Logical, defaults to <code>FALSE</code> . Used for testing alternative data download function. When <code>TRUE</code> data is downloaded using <code>read.csv</code> .
<code>use_direct_download</code>	Logical, defaults to <code>FALSE</code> . Used for testing alternative data download function. When <code>TRUE</code> data is downloaded using <code>download.file</code> .
<code>retry_download</code>	Logical defaults to <code>TRUE</code> . When <code>TRUE</code> , if downloading fails, the function will try repeatedly to download the data within 3 seconds, up to 3 times.

Value

The data loaded from a local copy or downloaded from the given url as a dataframe, exact format specified by data_trans_fn

See Also

get_tb_burden get_data_dict

Examples

```
tb_burden <- get_data(url = "https://extranet.who.int/tme/generateCSV.asp?ds=estimates",
  save_name = "TB_burden")

head(tb_burden)
```

get_data_dict

Get the WHO Data Dictionary for TB Data

Description

Get the data dictionary for TB burden data from the WHO, see [here](#) for details. This function will first attempt to load the data from the temporary directory (`tempdir`). If that fails, and `download_data = TRUE`, it will instead download the data.

Usage

```
get_data_dict(url = "https://extranet.who.int/tme/generateCSV.asp?ds=dictionary",
  download_data = TRUE, save = TRUE, dict_save_name = "TB_data_dict",
  return = TRUE, verbose = TRUE, use_utils = FALSE,
  retry_download = TRUE)
```

Arguments

url	Character string, indicating the url of the data dictionary. Default is current url.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by <code>tempdir</code> .
dict_save_name	Character string, name to save dictionary under. Defaults to TB_data_dict.
return	Logical, should the data dictionary be returned as a dataframe. Defaults to TRUE.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.

use_utils	Logical, defaults to FALSE. Used for testing alternative data download function. When TRUE data is downloaded using read.csv.
retry_download	Logical defaults to TRUE. When TRUE, if downloading fails, the function will try repeatedly to download the data within 3 seconds, up to 3 times.

Value

The WHO TB data dictionary as a tibble with 4 variables: variable_name, dataset, code_list, definition.

See Also

get_data search_data_dict

Examples

```
dict <- get_data_dict()
head(dict)
```

get_tb_burden *Get the WHO TB Burden Data*

Description

Get the TB burden data, and multi-drug resistant TB (MDR-TB) data from the WHO, see [here](#) for details. This function will first attempt to load the data from the temporary directory (`tempdir`). If that fails, and `download_data = TRUE`, it will instead download the data. The MDR TB data is only available for the latest year of data.

Usage

```
get_tb_burden(url = "https://extranet.who.int/tme/generateCSV.asp?ds=estimates",
  download_data = TRUE, save = TRUE, burden_save_name = "TB_burden",
  add_mdr_data = TRUE, mdr_save_name = "MDR_TB",
  mdr_url = "https://extranet.who.int/tme/generateCSV.asp?ds=mdr_rr_estimates",
  return = TRUE, verbose = TRUE, use_utils = FALSE,
  retry_download = TRUE)
```

Arguments

url	Character string, indicating the url of the TB burden data. Default is current url.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?

save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by <code>tempdir</code> .
burden_save_name	Character string, name to save the data under. Defaults to "TB_burden".
add_mdr_data	Logical, defaults to TRUE. Should MDR TB burden data be downloaded and joined to the TB burden data.
mdr_save_name	Character string, name to save the MDR data under. Defaults to "MDR_TB"
mdr_url	Character string, indicating the url of the MDR TB data.
return	Logical, should the data be returned as a dataframe. Defaults to TRUE.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
use_utils	Logical, defaults to FALSE. Used for testing alternative data download function. When TRUE data is downloaded using <code>read.csv</code> .
retry_download	Logical defaults to TRUE. When TRUE, if downloading fails, the function will try repeatedly to download the data within 3 seconds, up to 3 times.

Value

The WHO TB burden data as a tibble.

See Also

`get_data` `search_data_dict`

Examples

```
tb_burden <- get_tb_burden()

head(tb_burden)
```

map_tb_burden

Map TB Burden

Description

Map measures of TB burden by country by specifying a metric from the TB burden data. Specify a country or vector of countries in order to map them (the default is to map all countries). Various other options are available for tuning the plot further.

Usage

```
map_tb_burden(df = NULL, dict = NULL, metric = "e_inc_100k",
  metric_label = NULL, fill_var_type = NULL, countries = NULL,
  compare_to_region = FALSE, facet = NULL, year = NULL,
  annual_change = FALSE, trans = "identity", interactive = FALSE,
  download_data = TRUE, save = TRUE, burden_save_name = "TB_burden",
  dict_save_name = "TB_data_dict", viridis_palette = "viridis",
  viridis_direction = -1, viridis_end = 0.9, verbose = TRUE, ...)
```

Arguments

df	Dataframe of TB burden data, as sourced by get_tb_burden . If not specified then will source the WHO TB burden data, either locally if available or directly from the WHO (if <code>download_data = TRUE</code>).
dict	A tibble of the data dictionary. See get_data_dict for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and <code>download_data = TRUE</code> then the dictionary will be downloaded.
metric	Character string specifying the metric to plot
metric_label	Character string specifying the metric label to use.
fill_var_type	A character string, defaults to NULL. To set the fill variable type to be discrete use "discrete" and to be continuous use "continuous".
countries	A character string specifying the countries to plot.
compare_to_region	Logical, defaults to FALSE. If TRUE all countries that share a region with those listed in <code>countries</code> will be plotted. Note that this will override settings for <code>facet</code> , unless it is set to "country".
facet	Character string, the name of the variable to facet by.
year	Numeric, indicating the year of data to map. Defaults to the latest year in the data. If <code>interactive = TRUE</code> then multiple years may be passed as a vector, the result will then be animated over years.
annual_change	Logical, defaults to FALSE. If TRUE then the percentage annual change is computed for the specified metric.
trans	A character string specifying the transform to use on the specified metric. Defaults to no transform ("identity"). Other options include log scaling ("log") and log base 10 scaling ("log10"). For a complete list of options see <code>ggplot2::continuous_scale</code> .
interactive	Logical, defaults to FALSE. If TRUE then an interactive plot is returned.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by tempdir .
burden_save_name	Character string, name to save the data under. Defaults to "TB_burden".
dict_save_name	Character string, name to save dictionary under. Defaults to TB_data_dict.

viridis_palette	Character string indicating the viridis colour palette to use. Defaults to "viridis". Options include "cividis", "magma", "inferno", "plasma", and "viridis". For additional details see viridis_pal for additional details.
viridis_direction	Numeric, indicating the direction for the colour palette (1 or -1), defaults to -1. See scale_color_viridis for additional details.
viridis_end	Numeric between 0 and 1, defaults to 0.9. The end point of the viridis scale to use. #' See scale_color_viridis for additional details.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional parameters to pass to get_tb_burden .

Value

A plot of TB Incidence Rates by Country

See Also

[plot_tb_burden](#) [plot_tb_burden_overview](#) [get_tb_burden](#) [search_data_dict](#)

Examples

```
## Map raw incidence rates
map_tb_burden()

#' ## Map raw incidence rates
map_tb_burden(year = 2014:2017, facet = "year")
## Not run:
## Map log10 scaled incidence rates
map_tb_burden(trans = "log10")

## Map percentage annual change in incidence rates
map_tb_burden(annual_change = TRUE)

## Find variables relating to mortality in the WHO dataset
search_data_dict(def = "mortality")

## Map mortality rates (exc HIV) - without progress messages
map_tb_burden(metric = "e_mort_exc_tbhiv_100k", verbose = FALSE)

## Can also use a discrete metric if one is available
map_tb_burden(metric = "g_whoregion", metric_label = "WHO world region")

## End(Not run)
```

plot_tb_burden *Plot TB Burden by Country*

Description

Plot measures of TB burden by country by specifying a metric from the TB burden data. Specify a country or vector of countries in order to plot them (by default plots all countries). Various other options are available for tuning the plot further.

Usage

```
plot_tb_burden(df = NULL, dict = NULL, metric = "e_inc_100k",
  metric_label = NULL, conf = c("_lo", "_hi"), countries = NULL,
  years = NULL, compare_to_region = FALSE, facet = NULL,
  annual_change = FALSE, trans = "identity", scales = "fixed",
  interactive = FALSE, download_data = TRUE, save = TRUE,
  burden_save_name = "TB_burden", dict_save_name = "TB_data_dict",
  viridis_palette = "viridis", viridis_direction = -1,
  viridis_end = 0.9, verbose = TRUE, ...)
```

Arguments

df	Dataframe of TB burden data, as sourced by get_tb_burden . If not specified then will source the WHO TB burden data, either locally if available or directly from the WHO (if <code>download_data = TRUE</code>).
dict	A tibble of the data dictionary. See get_data_dict for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and <code>download_data = TRUE</code> then the dictionary will be downloaded.
metric	Character string specifying the metric to plot
metric_label	Character string specifying the metric label to use.
conf	Character vector specifying the name variations to use to specify the upper and lower confidence intervals. Defaults to <code>c("_lo", "_hi")</code> , if set to <code>NULL</code> then no confidence intervals are shown. When <code>annual_change = TRUE</code> the confidence intervals represent the annual percentage change in the metrics confidence intervals.
countries	A character string specifying the countries to plot.
years	Numeric vector of years. Defaults to <code>NULL</code> which includes all years in the data.
compare_to_region	Logical, defaults to <code>FALSE</code> . If <code>TRUE</code> all countries that share a region with those listed in <code>countries</code> will be plotted. Note that this will override settings for <code>facet</code> , unless it is set to "country".
facet	Character string, the name of the variable to facet by.
annual_change	Logical, defaults to <code>FALSE</code> . If <code>TRUE</code> then the percentage annual change is computed for the specified metric.

trans	A character string specifying the transform to use on the specified metric. Defaults to no transform ("identity"). Other options include log scaling ("log") and log base 10 scaling ("log10"). For a complete list of options see <code>ggplot2::continuous_scale</code> .
scales	Character string, see <code>?ggplot2::facet_wrap</code> for details. Defaults to "fixed", alternatives are "free_y", "free_x", or "free".
interactive	Logical, defaults to FALSE. If TRUE then an interactive plot is returned.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by tempdir .
burden_save_name	Character string, name to save the data under. Defaults to "TB_burden".
dict_save_name	Character string, name to save dictionary under. Defaults to TB_data_dict.
viridis_palette	Character string indicating the viridis colour palette to use. Defaults to "viridis". Options include "cividis", "magma", "inferno", "plasma", and "viridis". For additional details see viridis_pal for additional details.
viridis_direction	Numeric, indicating the direction for the colour palette (1 or -1), defaults to -1. See scale_color_viridis for additional details.
viridis_end	Numeric between 0 and 1, defaults to 0.9. The end point of the viridis scale to use. # See scale_color_viridis for additional details.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional parameters to pass to get_tb_burden .

Value

A plot of TB Incidence Rates by Country

See Also

[get_tb_burden](#) [search_data_dict](#)

Examples

```
## Get the WHO TB burden data and the data dictionary
tb_burden <- get_tb_burden()
dict <- get_data_dict()

## Get a random sample of 9 countries
sample_countries <- sample(unique(tb_burden$country), 9)

## Plot incidence rates in these countries
plot_tb_burden(df = tb_burden, dict = dict, facet = "country", countries = sample_countries)
```

```

## Use data caching to plot incidence rates with free y scales
plot_tb_burden(facet = "country", countries = sample_countries, scales = "free_y")

## Plot annual percentage change in incidence rates in selected countries
plot_tb_burden(df = tb_burden, dict = dict, facet = "country", scales = "free_y",
               countries = sample_countries, annual_change = TRUE, conf = NULL)

## Find variables relating to mortality in the WHO dataset
search_data_dict(def = "mortality")
## Plot mortality rates (exc HIV) - without progress messages
plot_tb_burden(metric = "e_mort_exc_tbhiv_100k", facet = "country",
               countries = sample_countries, scales = "free_y", verbose = FALSE)

```

```
plot_tb_burden_overview
```

Plot an overview of TB Burden for Multiple Countries

Description

This functions returns a dot plot for a given metric over a specified list of countries. If `compare_to_region` is specified then a given country will be compared to others in its region. This enables the user to rapidly understand trends in Tuberculosis over time and the progress towards global elimination.

Usage

```

plot_tb_burden_overview(df = NULL, dict = NULL,
                        metric = "e_inc_100k", metric_label = NULL, countries = NULL,
                        years = NULL, compare_to_region = FALSE, facet = NULL,
                        annual_change = FALSE, trans = "identity", scales = "free_y",
                        interactive = FALSE, download_data = TRUE, save = TRUE,
                        burden_save_name = "TB_burden", dict_save_name = "TB_data_dict",
                        viridis_palette = "viridis", viridis_direction = -1,
                        viridis_end = 0.9, verbose = TRUE, ...)

```

Arguments

<code>df</code>	Dataframe of TB burden data, as sourced by get_tb_burden . If not specified then will source the WHO TB burden data, either locally if available or directly from the WHO (if <code>download_data = TRUE</code>).
<code>dict</code>	A tibble of the data dictionary. See get_data_dict for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and <code>download_data = TRUE</code> then the dictionary will be downloaded.
<code>metric</code>	Character string specifying the metric to plot
<code>metric_label</code>	Character string specifying the metric label to use.
<code>countries</code>	A character string specifying the countries to plot.

years	Numeric vector of years. Defaults to NULL which includes all years in the data.
compare_to_region	Logical, defaults to FALSE. If TRUE all countries that share a region with those listed in <code>countries</code> will be plotted. Note that this will override settings for <code>facet</code> , unless it is set to "country".
facet	Character string, the name of the variable to facet by.
annual_change	Logical, defaults to FALSE. If TRUE then the percentage annual change is computed for the specified metric.
trans	A character string specifying the transform to use on the specified metric. Defaults to no transform ("identity"). Other options include log scaling ("log") and log base 10 scaling ("log10"). For a complete list of options see <code>ggplot2::continuous_scale</code> .
scales	Character string, see <code>?ggplot2::facet_wrap</code> for details. Defaults to "fixed", alternatives are "free_y", "free_x", or "free".
interactive	Logical, defaults to FALSE. If TRUE then an interactive plot is returned.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by <code>tempdir</code> .
burden_save_name	Character string, name to save the data under. Defaults to "TB_burden".
dict_save_name	Character string, name to save dictionary under. Defaults to <code>TB_data_dict</code> .
viridis_palette	Character string indicating the <code>viridis</code> colour palette to use. Defaults to "viridis". Options include "cividis", "magma", "inferno", "plasma", and "viridis". For additional details see <code>viridis_pal</code> for additional details.
viridis_direction	Numeric, indicating the direction for the colour palette (1 or -1), defaults to -1. See <code>scale_color_viridis</code> for additional details.
viridis_end	Numeric between 0 and 1, defaults to 0.9. The end point of the <code>viridis</code> scale to use. #? See <code>scale_color_viridis</code> for additional details.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional parameters to pass to <code>get_tb_burden</code> .

Value

A dot plot of any numeric metric by country.

See Also

`get_tb_burden` `search_data_dict`

Examples

```
## Plot incidence rates over time for both the United Kingdom and Botswana
plot_tb_burden_overview(countries = c("United Kingdom", "Botswana"),
                        compare_to_region = FALSE)

## Plot percentage annual change in incidence rates.
plot_tb_burden_overview(countries = c("United Kingdom", "Botswana"),
                        compare_to_region = FALSE, annual_change = TRUE)

## Compare incidence rates in the UK and Botswana to incidence rates in their regions
plot_tb_burden_overview(countries = c("United Kingdom", "Botswana"),
                        compare_to_region = TRUE)

## Find variables relating to mortality in the WHO dataset
search_data_dict(def = "mortality")

## Compare mortality rates (exc HIV) in the UK and Botswana to mortality rates in their regions
## Do not show progress messages
plot_tb_burden_overview(metric = "e_mort_exc_tbhiv_100k",
                        countries = c("United Kingdom", "Botswana"),
                        compare_to_region = TRUE, verbose = FALSE)
```

```
prepare_df_plot
```

```
Generic Function to Prepare TB Burden Data for Plotting
```

Description

This function is used internally by [plot_tb_burden](#) and [plot_tb_burden_overview](#) to prepare data for plotting.

Usage

```
prepare_df_plot(df = NULL, dict = NULL, metric = "e_inc_100k",
               conf = NULL, metric_label = NULL, countries = NULL, years = NULL,
               compare_to_region = FALSE, facet = NULL, annual_change = FALSE,
               trans = "identity", download_data = TRUE, save = TRUE,
               burden_save_name = "TB_burden", dict_save_name = "TB_data_dict",
               verbose = TRUE, ...)
```

Arguments

df Dataframe of TB burden data, as sourced by [get_tb_burden](#). If not specified then will source the WHO TB burden data, either locally if available or directly from the WHO (if `download_data = TRUE`).

dict A tibble of the data dictionary. See [get_data_dict](#) for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and `download_data = TRUE` then the dictionary will be downloaded.

<code>metric</code>	Character string specifying the metric to plot
<code>conf</code>	Character vector specifying the name variations to use to specify the upper and lower confidence intervals. Defaults to NULL for which no confidence intervals are used. Used by <code>annual_change</code> .
<code>metric_label</code>	Character string specifying the metric label to use.
<code>countries</code>	A character string specifying the countries to plot.
<code>years</code>	Numeric vector of years. Defaults to NULL which includes all years in the data.
<code>compare_to_region</code>	Logical, defaults to FALSE. If TRUE all countries that share a region with those listed in <code>countries</code> will be plotted. Note that this will override settings for <code>facet</code> , unless it is set to "country".
<code>facet</code>	Character string, the name of the variable to facet by.
<code>annual_change</code>	Logical, defaults to FALSE. If TRUE then the percentage annual change is computed for the specified metric.
<code>trans</code>	A character string specifying the transform to use on the specified metric. Defaults to no transform ("identity"). Other options include log scaling ("log") and log base 10 scaling ("log10"). For a complete list of options see <code>ggplot2::continuous_scale</code> .
<code>download_data</code>	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
<code>save</code>	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by <code>tempdir</code> .
<code>burden_save_name</code>	Character string, name to save the data under. Defaults to "TB_burden".
<code>dict_save_name</code>	Character string, name to save dictionary under. Defaults to TB_data_dict.
<code>verbose</code>	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
<code>...</code>	Additional parameters to pass to <code>get_tb_burden</code> .

Value

A list containing 3 elements, the dataframe to plot, the facet to use and the label to assign to the metric axis.

See Also

`plot_tb_burden` `plot_tb_burden_overview`

Examples

```
prepare_df_plot(countries = "Guinea")
```

run_tb_dashboard *Run a TB Shiny Dashboard*

Description

This functions runs a TB dashboard that has been built using other package functionality. The dashboard can be used to explore the global burden of TB interactively. A hosted version of this dashboard is available [here](#).

Usage

```
run_tb_dashboard()
```

Value

Starts a shiny Tuberculosis dashboard

Examples

```
## Only run the example if in an interactive session
## Not run:

## Run the TB dashboard
run_tb_dashboard()

## End(Not run)
```

search_data_dict *Search the WHO TB Data Dictionary by Variable Name.*

Description

Searches the WHO data dictionary for TB burden data. When run for the first time it will download the data dictionary, if `download_data = TRUE`, and save it into the temporary

Usage

```
search_data_dict(var = NULL, def = NULL, dict = NULL,
  download_data = TRUE, save = TRUE, dict_save_name = "TB_data_dict",
  verbose = TRUE, ...)
```

Arguments

var	A character vector of variable names.
def	A character vector of terms to use to search the variable definitions for partial matches
dict	A tibble of the data dictionary. See get_data_dict for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and <code>download_data = TRUE</code> then the dictionary will be downloaded.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by tempdir .
dict_save_name	Character string, name to save dictionary under. Defaults to <code>TB_data_dict</code> .
verbose	Logical, defaults to TRUE. Should search information be returned.
...	Additional parameters to pass to get_data_dict .

Value

A tibble containing the information in the data dictionary matching the variables searched for.

See Also

[get_data_dict](#)

Examples

```
## Search for a known variable
## Download and save the dictionary if it is not available locally
search_data_dict(var = "country")

## Search for all variables mentioning mortality in their definition
search_data_dict(def = "mortality")

## Search for both a known variable and for mortality being mentioned in there definition
## Duplicate entries will be omitted.
search_data_dict(var = "e_mort_exc_tbhiv_100k", def = "mortality")
```

Description

Summarise TB burden metrics by region, globally, and for custom groupings. For variables with uncertainty represented by confidence intervals bootstrapping can be used (assuming a normal distribution) to include this in any estimated summary measures. Currently two statistics are supported; the mean (with 95% confidence intervals) and the median (with 95% interquartile range). An incidence rate and weighted proportion statistic are in development.

Usage

```
summarise_tb_burden(df = NULL, dict = NULL, metric = "e_inc_100k",
  metric_label = NULL, conf = c("_lo", "_hi"), years = NULL,
  samples = 1000, countries = NULL, compare_to_region = FALSE,
  compare_to_world = FALSE, custom_compare = NULL,
  compare_all_regions = FALSE, stat = "mean", denominator = NULL,
  truncate_at_zero = TRUE, annual_change = FALSE,
  download_data = TRUE, save = TRUE, burden_save_name = "TB_burden",
  dict_save_name = "TB_data_dict", verbose = TRUE)
```

Arguments

<code>df</code>	Dataframe of TB burden data, as sourced by get_tb_burden . If not specified then will source the WHO TB burden data, either locally if available or directly from the WHO (if <code>download_data = TRUE</code>).
<code>dict</code>	A tibble of the data dictionary. See get_data_dict for details. If not supplied the function will attempt to load a saved version of the dictionary. If this fails and <code>download_data = TRUE</code> then the dictionary will be downloaded.
<code>metric</code>	Character string specifying the metric to plot
<code>metric_label</code>	Character string specifying the metric label to use.
<code>conf</code>	Character vector specifying the name variations to use to specify the upper and lower confidence intervals. Defaults to <code>NULL</code> for which no confidence intervals are used. Used by <code>annual_change</code> .
<code>years</code>	Numeric vector of years. Defaults to <code>NULL</code> which includes all years in the data.
<code>samples</code>	Numeric, the number of samples to use to generate confidence intervals (only used when <code>conf</code> are present)
<code>countries</code>	A character string specifying the countries to plot.
<code>compare_to_region</code>	Logical, defaults to <code>FALSE</code> . If <code>TRUE</code> all countries that share a region with those listed in <code>countries</code> will be plotted. Note that this will override settings for <code>facet</code> , unless it is set to "country".
<code>compare_to_world</code>	Logical, defaults to <code>FALSE</code> . Should a comparison be made to the metric of interests global value.
<code>custom_compare</code>	Logical, defaults to <code>NULL</code> . A named list of custom countries.
<code>compare_all_regions</code>	Logical, defaults to <code>FALSE</code> . Should all regions be compared.

stat	Character string, defaults to "mean". The statistic to use to summarise the metric, currently "mean" and "median" are supported. Note "mean" and "median" do not recompute the supplied country levels values. Future options will include the weighted proportion and the summarised incidence rate.
denominator	Character string identifying the variable to use as a denominator in any proportion or rate calculation.
truncate_at_zero	Logical, defaults to TRUE. Should lower bounds be truncated at zero?
annual_change	Logical, defaults to FALSE. If TRUE then the percentage annual change is computed for the specified metric.
download_data	Logical, defaults to TRUE. If not found locally should the data be downloaded from the specified URL?
save	Logical, should the data be saved for reuse during the current R session. Defaults to TRUE. If TRUE then the data is saved to the temporary directory specified by tempdir .
burden_save_name	Character string, name to save the data under. Defaults to "TB_burden".
dict_save_name	Character string, name to save dictionary under. Defaults to TB_data_dict.
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.

Value

A tibble containing summarised values (with 95 stratified by area and year).

Examples

```
## Get the most recent year of data
tb_burden <- get_tb_burden()
most_recent_year <- max(tb_burden$year)

## Get summary of the e_mdr_pct_rr_new cases
summarise_tb_burden(metric = "e_mdr_pct_rr_new",
                    years = most_recent_year,
                    samples = 100,
                    compare_all_regions = TRUE,
                    compare_to_world = TRUE,
                    verbose = TRUE)

## Get median (with 95% IQR) of the case fatality rate for regions and the world
## Bootstrapping uncertainty in country measures
summarise_tb_burden(metric = "cfr",
                    years = most_recent_year,
                    samples = 100,
                    stat = "median",
                    compare_all_regions = TRUE,
                    compare_to_world = TRUE,
                    verbose = FALSE)
```

```
## Get summary data for the UK, Europe and the world
## Bootstrapping CI's
summarise_tb_burden(metric = "e_inc_num",
                    samples = 100,
                    stat = "median",
                    countries = "United Kingdom",
                    compare_to_world = TRUE,
                    compare_to_region = TRUE,
                    verbose = FALSE)
```

who_shapefile

WHO shapefile

Description

The shapefile used in the WHO TB report, see [here](#) for the original source. This shapefile is used in [map_tb_burden](#).

Usage

who_shapefile

Format

A data frame with 15243 rows and 7 variables.

long Longitude

lat Latitude

order The shape order.

hole

piece

group The country group

id The country acronym

Index

*Topic **datasets**

who_shapefile, 18

get_data, 2

get_data_dict, 3, 6, 8, 10, 12, 15, 16

get_tb_burden, 4, 6–13, 16

map_tb_burden, 5, 18

plot_tb_burden, 8, 12

plot_tb_burden_overview, 10, 12

prepare_df_plot, 12

run_tb_dashboard, 14

scale_color_viridis, 7, 9, 11

search_data_dict, 14

summarise_tb_burden, 15

tempdir, 2–6, 9, 11, 13, 15, 17

viridis_pal, 7, 9, 11

who_shapefile, 18