

Package ‘getTBinR’

September 3, 2019

Type Package

Title Access and Summarise World Health Organisation Tuberculosis data

Version 0.7.0

Description Quickly and easily import analysis ready

Tuberculosis (TB) burden data, from the World Health Organisation (WHO), into R. The aim of getTBinR is to allow researchers, and other interested individuals, to quickly and easily gain access to a detailed TB data set and to start using it to derive key insights. It provides a consistent set of tools that can be used to rapidly evaluate hypotheses on a widely used data set before they are explored further using more complex methods or more detailed data. These tools include: generic plotting and mapping functions; a data dictionary search tool; an interactive shiny dashboard; and an automated, country level, TB report. For newer R users, this package reduces the barrier to entry by handling data import, munging, and visualisation. All plotting and mapping functions are built with ggplot2 so can be readily extended.

License GPL-3 | file LICENSE

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

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

Depends R (>= 3.3.0)

Imports data.table (>= 1.12.2), dplyr (>= 0.8.3), ggplot2 (>= 3.2.1), ggthemes (>= 4.2.0), magrittr (>= 1.5), plotly (>= 4.9.0), purrr (>= 0.3.2), scales (>= 1.0.0), tibble (>= 2.1.3), tidyr (>= 0.8.3), viridis (>= 0.5.1), rlang (>= 0.4.0)

Suggests spelling (>= 2.1), covr (>= 3.3.1), devtools (>= 2.1.0), usethis (>= 1.5.1), knitr (>= 1.24), pkgnet (>= 0.4.0), rmarkdown (>= 1.15), shiny (>= 1.3.2), shinycssloaders (>= 0.2.0), shinydashboard (>= 0.7.1), shinyWidgets (>= 0.4.8), testthat (>= 2.2.1), tidyverse (>= 1.2.1), vdiff (>= 0.3.1)

VignetteBuilder knitr

Encoding UTF-8**LazyData** true**RoxygenNote** 6.1.1**Language** en-GB**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** 2019-09-03 13:50:06 UTC

R topics documented:

available_datasets	2
get_data	3
get_data_dict	4
get_tb_burden	5
map_tb_burden	7
plot_tb_burden	9
plot_tb_burden_overview	11
plot_tb_burden_summary	13
prepare_df_plot	17
render_country_report	18
run_tb_dashboard	19
search_data_dict	20
summarise_metric	21
summarise_tb_burden	22
who_shapefile	25
Index	26

available_datasets	<i>Available Datasets</i>
--------------------	---------------------------

Description

Datasets that `getTBinR` supports importing into R, see [here](#) additional dataset details. This dataset is also used by `get_tb_burden`. Use `search_data_dict` in order to get details of the variables in each dataset.

Usage

```
available_datasets
```

Format

A data frame with 7 rows and 5 variables.

dataset Dataset name used by the WHO data dictionary.

description Either the data description supplied by the WHO or a user generated description.

timespan The timespan of the data

default Whether the dataset is downloaded by default or not.

url The URL for downloading the data - used by [get_tb_burden](#)

 get_data

Generic Get Data Function

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 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 tempdir .
<code>save_name</code>	Character string, name to save the data under. Defaults to NULL.
<code>return</code>	Logical, should the data be returned as a dataframe. Defaults to TRUE.
<code>verbose</code>	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
<code>use_utils</code>	Logical, defaults to FALSE. Used for testing alternative data download function. When TRUE data is downloaded using <code>read.csv</code> .
<code>use_direct_download</code>	Logical, defaults to FALSE. Used for testing alternative data download function. When TRUE data is downloaded using <code>download.file</code> .
<code>retry_download</code>	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 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 = NULL, download_data = TRUE, save = TRUE,
  dict_save_name = NULL, return = TRUE, verbose = TRUE,
  use_utils = FALSE, retry_download = TRUE)
```

Arguments

url	Character string, indicating the url of the data dictionary. This argument is deprecated and will be removed from future releases.
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. This argument is deprecated and will be removed from future releases. Dataset naming is now handled internally.
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	<i>Get the WHO TB Burden Data</i>
---------------	-----------------------------------

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 = NULL, download_data = TRUE, save = TRUE,
  burden_save_name = NULL, add_mdr_data = TRUE,
  additional_datasets = NULL, mdr_save_name = NULL, mdr_url = NULL,
  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. This argument is depreciated and will be removed from future releases. The TB burden URL is now supplied internally - see available_datasets for details.
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. This argument is depreciated and will be removed from future releases. Dataset naming is now handled internally.
add_mdr_data	Logical, defaults to TRUE. Should MDR TB burden data be downloaded and joined to the TB burden data.
additional_datasets	A character vector specifying the names of the additional datasets to import. See available_datasets for available datasets. Use "all" to download all available datasets (experimental datasets such as incidence by age and sex are excluded from this list).
mdr_save_name	Character string, name to save the MDR data under. This argument is depreciated and will be removed from future releases. Dataset naming is now handled internally.
mdr_url	Character string, indicating the url of the MDR TB data. This argument is depreciated and will be removed from future releases. The MDR-TB burden URL is now supplied internally - see available_datasets for details.
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

```
## Default datasets
```

```

tb_burden <- get_tb_burden(additional_datasets = available_datasets$dataset[3])

head(tb_burden)

## Add in the latent TB dataset as an additional dataset (see getTBinR::available_datasets)
tb_with_latents <- get_tb_burden(additional_datasets = available_datasets$dataset[3])

head(tb_with_latents)

```

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, legend = "bottom",
  year = NULL, annual_change = FALSE, trans = "identity",
  interactive = FALSE, download_data = TRUE, save = TRUE,
  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 target.
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.
legend	Character string, defaults to "bottom". Position of the legend see <code>?ggplot2::theme</code> for defaults but known options are: "none", "top", "right" and "bottom".
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 .
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 arguments 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()
## Not run:
#' ## Map raw incidence rates
map_tb_burden(year = 2014:2017, facet = "year")
```

```

## 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, smooth = FALSE, conf = c("_lo", "_hi"),
  countries = NULL, years = NULL, compare_to_region = FALSE,
  facet = NULL, legend = "none", annual_change = FALSE,
  trans = "identity", scales = "fixed", interactive = FALSE,
  download_data = TRUE, save = TRUE, 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.

smooth	Logical, defaults to FALSE. Should the data be smoothed (using <code>ggplot2::geom_smooth</code>) prior to plotting. If set to TRUE then the confidence intervals shown are derived from the smooth and do not represent the underlying uncertainty in the data.
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 NULL 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 target.
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.
legend	Character string, defaults to "none". Position of the legend see <code>?ggplot2::theme</code> for defaults but known options are: "none", "top", "right" and "bottom".
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> .
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 arguments to pass to <code>get_tb_burden</code> .

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)

## Not run:
## Plot smoothed incidence rates in these countries
plot_tb_burden(df = tb_burden, dict = dict, facet = "country", smooth = TRUE,
               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)

## End(Not run)
```

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", legend = "right",
  scales = "free_y", interactive = FALSE, download_data = TRUE,
  save = TRUE, 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 target.
<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>legend</code>	Character string, defaults to "right". Position of the legend see <code>?ggplot2::theme</code> for defaults but known options are: "none", "top", "right" and "bottom".
<code>scales</code>	Character string, see <code>?ggplot2::facet_wrap</code> for details. Defaults to "fixed", alternatives are "free_y", "free_x", or "free".
<code>interactive</code>	Logical, defaults to FALSE. If TRUE then an interactive plot is returned.
<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 tempdir .
<code>viridis_palette</code>	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 arguments to pass to get_tb_burden .

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)
```

plot_tb_burden_summary

Plot Summaries of TB Burden - By Region, Globally and for Custom Groups

Description

Plot summaries of 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 four statistics are supported; the mean (with 95% confidence intervals) and the median (with 95% interquartile range), rates and proportions.

Usage

```
plot_tb_burden_summary(df = NULL, dict = NULL, metric = "e_inc_num",
  metric_label = NULL, conf = c("_lo", "_hi"), years = NULL,
  samples = 1000, countries = NULL, compare_to_region = FALSE,
  compare_to_world = TRUE, custom_compare = NULL,
  compare_all_regions = TRUE, stat = "rate", denom = "e_pop_num",
  rate_scale = 1e+05, truncate_at_zero = TRUE, annual_change = FALSE,
  smooth = FALSE, facet = NULL, legend = "top", trans = "identity",
  scales = "fixed", interactive = FALSE, viridis_palette = "viridis",
  viridis_direction = -1, viridis_end = 0.9, download_data = TRUE,
  save = TRUE, 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 defaulting to NULL. If supplied this metric will be looked up using the WHO data dictionary to provide a label. A use case would be when calculating incidence rates using <code>e_inc_100k</code> to get the WHO TB incidence rate label.
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 NULL 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.
years	Numeric vector of years. Defaults to NULL which includes all years in the data.
samples	Numeric, the number of samples to use to generate confidence intervals (only used when <code>conf</code> are present)
countries	A character string specifying the countries to target.
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".

compare_to_world	Logical, defaults to TRUE. Should a comparison be made to the metric of interests global value.
custom_compare	Logical, defaults to NULL. A named list of custom countries.
compare_all_regions	Logical, defaults to TRUE. Should all regions be compared.
stat	Character string, defaults to "rate". The statistic to use to summarise the metric, currently "mean", "median", "rate" and "prop" are supported. Note "mean" and "median" do not recompute the supplied country levels values but can be used to summarise the distribution of region or global metrics. "prop" and "rate" compute the overall incidence rate for a given grouping (i.e the sum of the metric divided by the sum of the denominator).
denom	Character string defaulting to e_pop_num (country level population). If stat is set to rate or prop then this is the parameter to use as the denominator.
rate_scale	Numeric defaults to 100,000. The scaling to use for rates. If stat is to set to prop then this defaults to 1.
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.
smooth	Logical, defaults to FALSE. Should the data be smoothed (using ggplot2::geom_smooth) prior to plotting. If set to TRUE then the confidence intervals shown are derived from the smooth and do not represent the underlying uncertainty in the data.
facet	Character string, the name of the variable to facet by.
legend	Character string, defaults to "top". Position of the legend see ?ggplot2::theme for defaults but known options are: "none", "top", "right" and "bottom".
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 ggplot2::continuous_scale.
scales	Character string, see ?ggplot2::facet_wrap 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.
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.
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> .
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional arguments to pass to <code>get_tb_burden</code> .

Value

A plot of TB Incidence Rates by Country

See Also

`search_data_dict` `plot_tb_burden` `summarise_tb_burden`
`summarise_tb_burden` `get_tb_burden` `search_data_dict`

Examples

```
## Get an overview of incidence rates regionally and globally compared to the UK
plot_tb_burden_summary(metric = "e_inc_num",
                       metric_label = "e_inc_100k",
                       stat = "rate",
                       countries = "United Kingdom",
                       compare_to_world = TRUE,
                       compare_all_regions = TRUE,
                       verbose = FALSE,
                       scales = "free_y",
                       facet = "Area")
```

```
## Not run:
```

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

```
## End(Not run)
```

```
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,
  verbose = TRUE, ...)
```

Arguments

<code>df</code>	Dataframe of TB burden data, as sourced by <code>get_tb_burden</code> . 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 <code>get_data_dict</code> 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>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>metric_label</code>	Character string specifying the metric label to use.
<code>countries</code>	A character string specifying the countries to target.
<code>years</code>	Numeric vector of years. Defaults to <code>NULL</code> which includes all years in the data.
<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>facet</code>	Character string, the name of the variable to facet by.
<code>annual_change</code>	Logical, defaults to <code>FALSE</code> . If <code>TRUE</code> 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 <code>TRUE</code> . 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> .
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional arguments 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")
```

`render_country_report` *Render a Country Level TB Report*

Description

This function renders a country level TB report.

Usage

```
render_country_report(country = "United Kingdom",
  format = "html_document", interactive = FALSE, save_dir = NULL,
  filename = NULL)
```

Arguments

country	Character string, defaults to "United Kingdom". Specify the country to report on.
format	Character string, defaults to "html_document". The format to render the report to. See <code>?rmarkdown::render</code> for details.
interactive	Logical, defaults to FALSE. When the format allows should graphs be interactive.
save_dir	Character string, defaults to NULL. If not given then the report is rendered to a temporary directory (although only if filename is also not given).
filename	Character string defaults NULL. Name to save the report under, defaults to "country_report".

Value

Renders a country level TB report

Examples

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

## Run the TB dashboard
render_country_report()

## End(Not run)
```

run_tb_dashboard	<i>Run a TB Shiny Dashboard</i>
------------------	---------------------------------

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, dataset = NULL,
  dict = NULL, download_data = TRUE, save = TRUE,
  dict_save_name = NULL, verbose = TRUE, ...)
```

Arguments

<code>var</code>	A character vector of variable names.
<code>def</code>	A character vector of terms to use to search the variable definitions for partial matches.
<code>dataset</code>	A character vector of terms to use to search the dataset names for partial matches.
<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>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 tempdir .
<code>dict_save_name</code>	Character string, name to save dictionary under. This argument is depreciated and will be removed from future releases. Dataset naming is now handled internally.
<code>verbose</code>	Logical, defaults to TRUE. Should search information be returned.
<code>...</code>	Additional parameters to pass to get_data_dict .

Value

A tibble containing the information in the data dictionary matching the variables searched for. If nothing is found then NULL is returned.

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 all variables in the Estimates dataset.
search_data_dict(dataset = "Estimates")

## 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")
```

summarise_metric *Summarise a TB Metric - A Quick Look Summary*

Description

This function provides a curated list of summary measures for a given TB metric in countries of interest. It can be used to facilitate reporting and is used extensively in the TB report included in the package (see [render_country_report](#)). It outputs the most recent year of data in the target country for a given metric, along with the year this data was recorded, the regional and global rank and the average change in the last decade. For a more customisable metric summary see [summarise_tb_burden](#) as a starting point.

Usage

```
summarise_metric(df = NULL, metric = NULL, countries = NULL,
  conf = c("_lo", "_hi"), download_data = TRUE, save = TRUE,
  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>).
metric	Character string specifying the metric to plot
countries	A character string specifying the countries to target.
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.
download_data	Logical, defaults to <code>TRUE</code> . 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> .
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional arguments to pass to <code>get_tb_burden</code> .

Value

A tibble containing the name of the target country, the year of the most recent data available, the most recent value for the metric, the regional rank, the global rank and the average change in the previous decade.

Examples

```
## Get a summary of TB incidence rates for the united kingdom and germany
summarise_metric(metric = "e_inc_100k", countries = c("United Kingdom", "Germany"))

## Get a summary of case detection rates in France
summarise_metric(metric = "c_cdr", countries = "France")

## Get a summary of case detection rates in France - without confidence intervals
summarise_metric(metric = "c_cdr", countries = "France", conf = NULL)

## Provide a dataset and get summary measures from it.
tb <- get_tb_burden()
summarise_metric(df = tb, metric = "c_cdr", countries = "France")
```

summarise_tb_burden *Summarise TB Burden - By Region, Globally and for Custom Groups*

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), rates and proportions.

Usage

```
summarise_tb_burden(df = NULL, dict = NULL, metric = "e_inc_num",
  metric_label = NULL, conf = c("_lo", "_hi"), years = NULL,
  samples = 1000, countries = NULL, compare_to_region = FALSE,
  compare_to_world = TRUE, custom_compare = NULL,
  compare_all_regions = TRUE, stat = "rate", denom = "e_pop_num",
  rate_scale = 1e+05, truncate_at_zero = TRUE, annual_change = FALSE,
  download_data = TRUE, save = TRUE, 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 NULL for which no confidence intervals are used. Used by <code>annual_change</code> .
years	Numeric vector of years. Defaults to NULL which includes all years in the data.
samples	Numeric, the number of samples to use to generate confidence intervals (only used when <code>conf</code> are present)
countries	A character string specifying the countries to target.
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".
compare_to_world	Logical, defaults to TRUE. Should a comparison be made to the metric of interests global value.
custom_compare	Logical, defaults to NULL. A named list of custom countries.
compare_all_regions	Logical, defaults to TRUE. Should all regions be compared.
stat	Character string, defaults to "rate". The statistic to use to summarise the metric, currently "mean", "median", "rate" and "prop" are supported. Note "mean" and "median" do not recompute the supplied country levels values but can be used to summarise the distribution of region or global metrics. "prop" and "rate" compute the overall incidence rate for a given grouping (i.e the sum of the metric divided by the sum of the denominator).
denom	Character string defaulting to <code>e_pop_num</code> (country level population). If <code>stat</code> is set to <code>rate</code> or <code>prop</code> then this is the parameter to use as the denominator.
rate_scale	Numeric defaults to 100,000. The scaling to use for rates. If <code>stat</code> is set to <code>prop</code> then this defaults to 1.
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 <code>tempdir</code> .
verbose	Logical, defaults to TRUE. Should additional status and progress messages be displayed.
...	Additional arguments to pass to <code>get_tb_burden</code> .

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,
  stat = "mean",
  samples = 100,
  compare_all_regions = TRUE,
  compare_to_world = TRUE,
  verbose = TRUE)

## Not run:
## 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)

## Get an overview of incidence rates regionally and globally compared to the UK
summarise_tb_burden(metric = "e_inc_num",
  stat = "rate",
  countries = "United Kingdom",
```

```
compare_to_world = TRUE,  
compare_to_region = TRUE,  
verbose = FALSE)  
  
## End(Not run)
```

who_shapefile	<i>WHO shapefile</i>
---------------	----------------------

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**

available_datasets, 2
who_shapefile, 25

available_datasets, 2, 6

get_data, 3
get_data_dict, 4, 7, 9, 12, 14, 17, 20, 23
get_tb_burden, 2, 3, 5, 7–10, 12–14, 16–18,
21–24

map_tb_burden, 7, 25

plot_tb_burden, 9, 17
plot_tb_burden_overview, 11, 17
plot_tb_burden_summary, 13
prepare_df_plot, 17

render_country_report, 18, 21
run_tb_dashboard, 19

scale_color_viridis, 8, 10, 13, 15
search_data_dict, 2, 20
summarise_metric, 21
summarise_tb_burden, 21, 22

tempdir, 3–6, 8, 10, 12, 16, 18, 20, 22, 24

viridis_pal, 8, 10, 12, 15

who_shapefile, 25