# Package 'getTBinR'

June 11, 2020

**Type** Package

**Title** Access and Summarise World Health Organization Tuberculosis Data

**Version** 0.7.1

**Description** Quickly and easily import analysis ready Tuberculosis (TB) burden data, from the World Health Organization (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 (>= 1.12.0), dplyr (>= 1.0.0), ggplot2 (>= 3.3.1), ggthemes (>= 4.2.0), magrittr (>= 1.5), plotly (>= 4.9.2.1), purrr (>= 0.3.4), scales (>= 1.1.1), tibble (>= 3.0.1), tidyr (>= 1.1.0), viridis (>= 0.5.1), rlang (>= 0.4.6), grDevices

**Suggests** spelling (>= 2.1), covr (>= 3.5.0), devtools (>= 2.3.0), usethis (>= 1.6.1), knitr (>= 1.28), pkgnet (>= 0.4.1), rmarkdown (>= 2.2), shiny (>= 1.4.0.2), shinycssloaders (>= 0.3), shinydashboard (>= 0.7.1), shinyWidgets (>= 0.5.3), testthat (>= 2.3.2), vdiffr (>= 0.3.2), styler (>= 1.3.2), pkgdown (>= 1.5.1)

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.0

**Language** en-GB

**NeedsCompilation** no

**Author** Sam Abbott [aut, cre] (<https://orcid.org/0000-0001-8057-8037>),
Maria Bekker-Nielsen Dunbar [ctb]
(<https://orcid.org/0000-0002-7249-3524>)

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

**Repository** CRAN

**Date/Publication** 2020-06-11 05:00:02 UTC

# R topics documented:

## available_datasets *Available Datasets*

### 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

## for_presentation *Presentation style plots*

### Description

This function adjusts the aspect ratio and default font size of a plot to improve readability when including figures in presentations.

### Usage

```
for_presentation(plot, aspect_ratio = 0.5, font_increase = 1.5)
```

### Arguments

| | |
|---|---|
| plot | The ggplot2 object to be adjusted. |
| aspect_ratio | The desired ratio between the height and width of the plot, defaults to 0.5. |
| font_increase | The amount to increase the font size by, default to 1.5. |

### Value

A plot adjusted for presentation.

## Author(s)

Maria Bekker-Nielsen Dunbar

Sam Abbott

## Examples

```
plot <- plot_tb_burden_summary(
  countries = "United Kingdom",
  compare_all_regions = FALSE, compare_to_region = TRUE
)
plot # Original
for_presentation(plot) # After adjustments
```

---

# 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 = FALSE,
  use_utils = FALSE,
  use_direct_download = FALSE,
  retry_download = TRUE
)
```

## Arguments

| | |
|---|---|
| `url` | Character string, indicating the url of the data to download. |
| `data_trans_fn` | Function that takes a data.table as input and returns a single dataframe of any type. If not specified defaults to transforming the data into a tibble. |
| `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. |
| save_name | Character string, name to save the data under. Defaults to NULL. |
| return | Logical, should the data be returned as a dataframe. Defaults to TRUE. |
| verbose | Logical, defaults to FALSE. 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. |
| use_direct_download | Logical, defaults to FALSE. Used for testing alternative data download function. When TRUE data is downloaded using download.file. |
| 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 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 = FALSE,
  use_utils = FALSE,
  retry_download = TRUE
)
```

## Arguments

| | |
|---|---|
| url | Character string, indicating the url of the data dictionary. This argument is depreciated 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 tempdir. |
| dict_save_name | Character string, name to save dictionary under. This argument is depreciated 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 FALSE. 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 = 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 = FALSE,
  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 FALSE. 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 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::avaiable_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 = FALSE,
  ...
)
```

## 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. |
| 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 countries will be plotted. Note that this will override settings for facet, 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 ?ggplot2::theme 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 interactive = TRUE 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 ggplot2::continous_scale. |
| 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 FALSE. 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 = FALSE,
  ...
)
```

## 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. |
| 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 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. |
| conf | Character vector specifying the name variations to use to specify the upper and lower confidence intervals. Defaults to c("_lo", "_hi"), if set to NULL then no confidence intervals are shown. When annual_change = TRUE 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 countries will be plotted. Note that this will override settings for facet, 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 ?ggplot2::theme 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 ggplot2::continous_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. |
| 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 FALSE. 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

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"
)
```

```
## 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 = "bottom",
  scales = "free_y",
  interactive = FALSE,
  download_data = TRUE,
  save = TRUE,
  viridis_palette = "viridis",
  viridis_direction = -1,
  viridis_end = 0.9,
  verbose = FALSE,
  ...
)
```

## 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. |
| `metric` | Character string specifying the metric to plot |

| | |
|---|---|
| metric_label | Character string specifying the metric label to use. |
| 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 countries will be plotted. Note that this will override settings for facet, 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 ggplot2::continous_scale. |
| legend | Character string, defaults to "right". Position of the legend see ?ggplot2::theme for defaults but known options are: "none", "top", "right" and "bottom". |
| 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. |
| 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 FALSE. 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\

## 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 = "bottom",
  trans = "identity",
  scales = "fixed",
  interactive = FALSE,
  viridis_palette = "viridis",
  viridis_direction = -1,
  viridis_end = 0.9,
  download_data = TRUE,
  save = TRUE,
  verbose = FALSE,
  ...
)
```

## 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. |
| 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 e_inc_100k 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 c("_lo", "_hi"), if set to NULL then no confidence intervals are shown. When annual_change = TRUE 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 `conf` 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 `countries` will be plotted. Note that this will override settings for `facet`, 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::continous_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 tempdir.

verbose
: Logical, defaults to FALSE. 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

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 = FALSE,
  ...
)
```

## 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. |
| metric | Character string specifying the metric to plot |
| 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 annual_change. |

| | |
|---|---|
| metric_label | Character string specifying the metric label to use. |
| 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 countries will be plotted. Note that this will override settings for facet, 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 ggplot2::continous_scale. |
| 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. |
| verbose | Logical, defaults to FALSE. Should additional status and progress messages be displayed. |
| ... | Additional arguments to pass to get_tb_burden. |

## 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 ?rmarkdown::render 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 interative session
## Not run:

## Run the TB dashboard
render_country_report()

## End(Not run)
```

---

# 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 interative session
## Not run:

## Run the TB dashboard
run_tb_dashboard()

## End(Not run)
```

---

`scale_colour_who` *Adds World Health Organization-inspired colours to plots*

---

## Description

Applies WHO inspired colours for `ggplot2` plots. Currently a continuous palette has not been implemented so the `viridis` palettes will be used as a fallback in this situation.

## Usage

```
scale_colour_who(
  palette = "main",
  discrete = TRUE,
  reverse = FALSE,
  n = NULL,
  add_missings = FALSE,
  ...
)
```

## Arguments

| | |
|---|---|
| `palette` | A character string. the colours of choice, current options are "main", "light" and "misc" for mixed colours as well as "purple", "turquoise", "blue", "magenta", "brown", "misc", "green", and "red". |
| `discrete` | Logical, defaults to `TRUE`. Should the palette be discrete or continuous. |
| `reverse` | Logical, defaults to `FALSE`. Should the palette be reversed. |
| `n` | Number of colours desired. If a specific value is given which corresponds to the length of a palette used in the 2019 WHO TB report, this palette is returned, else the number of entries in the palette with the most colour options (up to its full length) is returned. Defaults to `NULL` which returns the longest palette. |
| `add_missings` | Add a grey and white colour to the palette, defaults to `FALSE`. |

... Pass additional arguments to ggplot2::discrete_scale or ggplot2::scale_colour_viridis_c depending on the discrete setting. @param ... Pass additional arguments to ggplot2::discrete_scale or ggplot2::scale_colour_viridis_c depending on the discrete setting.

## Author(s)

Maria Bekker-Nielsen Dunbar

Sam Abbott

## See Also

scale_colour_who scale_fill_who who_palettes

## Examples

```
plot_tb_burden_summary(
  countries = "United Kingdom", compare_all_regions = FALSE,
  compare_to_region = TRUE, conf = NULL, verbose = FALSE
) +
  theme_who() +
  scale_colour_who(reverse = TRUE)
```

---

# scale_fill_who *Adds World Health Organization-inspired colours as fills to plots*

---

## Description

Applies WHO inspired colours as a fill for ggplot2 plots. Currently a continuous palette has not been implemented so the viridis palettes will be used as a fallback in this situation.

## Usage

```
scale_fill_who(
  palette = "light",
  discrete = TRUE,
  reverse = FALSE,
  n = NULL,
  add_missings = FALSE,
  ...
)
```

### Arguments

| | |
|---|---|
| palette | A character string. the colours of choice, current options are "main", "light" and "misc" for mixed colours as well as "purple", "turquoise", "blue", "magenta", "brown", "misc", "green", and "red". |
| discrete | Logical, defaults to TRUE. Should the palette be discrete or continuous. |
| reverse | Logical, defaults to FALSE. Should the palette be reversed. |
| n | Number of colours desired. If a specific value is given which corresponds to the length of a palette used in the 2019 WHO TB report, this palette is returned, else the number of entries in the palette with the most colour options (up to its full length) is returned. Defaults to NULL which returns the longest palette. |
| add_missings | Add a grey and white colour to the palette, defaults to FALSE. |
| ... | Pass additional arguments to ggplot2::discrete_scale or ggplot2::scale_colour_viridis_c depending on the discrete setting. @param ... Pass additional arguments to ggplot2::discrete_scale or ggplot2::scale_colour_viridis_c depending on the discrete setting. |

### Author(s)

Maria Bekker-Nielsen Dunbar

Sam Abbott

### See Also

scale_colour_who who_palettes

### Examples

```
plot_tb_burden_summary(
  countries = "United Kingdom", compare_all_regions = FALSE,
  compare_to_region = TRUE, verbose = FALSE
) +
  theme_who() +
  scale_colour_who(reverse = TRUE) +
  scale_fill_who(reverse = TRUE)
```

---

## 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 = FALSE,
  ...
)
```

## Arguments

| | |
|---|---|
| var | A character vector of variable names. |
| def | A character vector of terms to use to search the variable definitions for partial matches. |
| dataset | A character vector of terms to use to search the dataset names 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 download_data = TRUE 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. This argument is depreciated and will be removed from future releases. Dataset naming is now handled internally. |
| 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. 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 been 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 = FALSE,
  ...
)
```

## 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`). |
| `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 c("_lo", "_hi"), if set to `NULL` then no confidence intervals are shown. |

| | |
|---|---|
| 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. |
| verbose | Logical, defaults to FALSE. Should additional status and progress messages be displayed. |
| ... | Additional arguments to pass to get_tb_burden. |

## 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\

## 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 = FALSE,
  ...
)
```

## 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. |
| 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 annual_change. |
| 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 conf 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 countries will be plotted. Note that this will override settings for facet, 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.

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.

verbose
: Logical, defaults to FALSE. Should additional status and progress messages be displayed.

...
: Additional arguments to pass to get_tb_burden.

## Value

A tibble containing summarised values (with 95% confidence intervals) for the metric of choice 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
## Boostrapping 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)
```

---

theme_who *World Health Organization-inspired ggplot2 theme*

---

## Description

This theme is inspired by that used in the annual global Tuberculosis report. See the report here.

## Usage

```
theme_who()
```

## Author(s)

Maria Bekker-Nielsen Dunbar

Sam Abbott

## See Also

bbplot::bbc_style()

## Examples

```
plot_tb_burden_summary(conf = NULL, verbose = FALSE) +
  theme_who()
```

---

`who_palettes` *World Health Organization-inspired palettes*

---

## Description

World Health Organization-inspired palettes

## Usage

```
who_palettes(
  palette = "main",
  reverse = FALSE,
  n = NULL,
  add_missings = FALSE,
  ...
)
```

## Arguments

| | |
|---|---|
| `palette` | A character string. the colours of choice, current options are "main", "light" and "misc" for mixed colours as well as "purple", "turquoise", "blue", "magenta", "brown", "misc", "green", and "red". |
| `reverse` | Logical, defaults to `FALSE`. Should the palette be reversed. |
| `n` | Number of colours desired. If a specific value is given which corresponds to the length of a palette used in the 2019 WHO TB report, this palette is returned, else the number of entries in the palette with the most colour options (up to its full length) is returned. Defaults to `NULL` which returns the longest palette. |
| `add_missings` | Add a grey and white colour to the palette, defaults to `FALSE`. |
| `...` | Pass additional arguments to `colorRampPalette`. |

## Author(s)

Maria Bekker-Nielsen Dunbar

Sam Abbott

## See Also

scale_colour_who scale_fill_who

## Examples

```
# Set up the main palette but reversed.
who_palettes(palette = "main", reverse = TRUE)
```

---

`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