

Package ‘DSIR’

May 23, 2026

Title Data Science Infrastructure for Global Health

Version 0.7.1

Description Supports global health data analysis, including a publication-ready 'ggplot2' theme, a 'flextable' defaults helper, a thin pie chart wrapper, built-in regional country-code datasets with a WHO region lookup helper, a geometric mean function for indicator aggregation, and convenience clients for the World Health Organization Global Health Observatory (GHO) OData API <<https://ghoapi.azureedge.net/api/>> and the United Nations Sustainable Development Goals (SDG) API <<https://unstats.un.org/SDGAPI/swagger/>>.

License MIT + file LICENSE

Language en-US

URL <https://github.com/shanlong-who/DSIR>,
<https://shanlong-who.github.io/DSIR/>

BugReports <https://github.com/shanlong-who/DSIR/issues>

Encoding UTF-8

Depends R (>= 4.1.0)

Imports cli, flextable (>= 0.9.0), ggplot2 (>= 3.4.0), httr2, rlang,
tibble (>= 3.0.0)

Suggests patchwork, officer, testthat (>= 3.0.0), httptest2, knitr,
rmarkdown, dplyr

Config/testthat/edition 3

LazyData true

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Shanlong Ding [aut, cre] (ORCID:
<<https://orcid.org/0000-0001-9048-6670>>)

Maintainer Shanlong Ding <dings@who.int>

Repository CRAN

Date/Publication 2026-05-23 11:22:45 UTC

Contents

bind_indicators	2
dsi_flextable_defaults	3
geomean	4
ggpie	5
gho_clean	6
gho_count	8
gho_coverage	9
gho_data	10
gho_dimensions	11
gho_has_data	12
gho_indicators	13
iso3_to_m49	14
iso3_to_region	15
m49_to_iso3	16
pic_cty	17
scale_dsi_col	17
sdg_areas	18
sdg_clean	19
sdg_coverage	20
sdg_data	21
sdg_goals	22
sdg_indicators	23
sdg_targets	24
theme_dsi	25
theme_dsi_facet	26
who_countries	27
who_region_vectors	29

Index **31**

bind_indicators	<i>Bind Cleaned Indicator Tibbles</i>
-----------------	---------------------------------------

Description

Combines two or more tibbles produced by `gho_clean()` or `sdg_clean()` into a single tibble. Because both cleaners output the same 15-column schema, the result is a uniform table that can be filtered, joined, or visualised without source-specific code paths; use the `source` column to tell GHO rows apart from SDG rows.

Usage

```
bind_indicators(...)
```

Arguments

... Two or more tibbles returned by `gho_clean()` or `sdg_clean()` (or any data frame with the same column set). NULL arguments are dropped. Calling with no inputs — or only NULL inputs — returns the empty 15-column tibble.

Details

Inputs do not need to be in any particular order. NULL inputs are silently dropped, which makes it ergonomic to write code like `bind_indicators(maybe_gho, maybe_sdg)` where some sources may not have been fetched.

Value

A single [tibble](#) with the unified cleaned- indicator schema (15 columns). Row order is `c(input_1, input_2, ...)`, preserving within-input order.

See Also

[gho_clean\(\)](#), [sdg_clean\(\)](#).

Examples

```
gho <- gho_data("NCDMORT3070", area = wpro_pty) |> gho_clean()
sdg <- sdg_data("3.4.1", area = wpro_pty) |> sdg_clean()
bind_indicators(gho, sdg)
```

dsi_flexable_defaults

Set DSIR Flextable Defaults

Description

Applies a consistent set of flextable formatting defaults for publication-ready tables (booktabs theme, bold headers, modest padding). Pick any font you like — the default "" leaves the flextable default in place so the call is safe on systems where Cambria is not installed.

Usage

```
dsi_flextable_defaults(  
  font_size = 12,  
  font_family = "",  
  font_color = "#333333",  
  border_color = "black",  
  padding = c(3, 3, 4, 4)  
)
```

Arguments

font_size	Font size in points. Default 12.
font_family	Font family name. Default "" keeps the existing flextable default; try "Cambria" for the original DSIR look on Windows.
font_color	Font color. Default "#333333".
border_color	Border color. Default "black".
padding	Numeric vector of length 1 (applied to all sides) or length 4 (top, bottom, left, right). Default c(3, 3, 4, 4).

Value

Invisibly returns NULL. Called for its side effect of mutating the flextable global defaults via [flextable::set_flextable_defaults\(\)](#).

Examples

```
dsi_flextable_defaults()
```

geomean

Geometric Mean

Description

Computes the geometric mean of a numeric vector, with optional weights. Useful for aggregating multiplicative quantities such as ratio-based health indicators (e.g. UHC service-coverage tracers, where the composite index is the geometric mean of component coverage values).

Usage

```
geomean(x, w = NULL, na.rm = TRUE)
```

Arguments

x	A numeric vector. Zeros produce a result of 0. Negative values produce NaN with a warning, since the geometric mean is undefined for negative numbers.
w	Optional numeric vector of weights, the same length as x. Must be non-negative. If NULL (default), the unweighted geometric mean is returned.
na.rm	Logical. Should missing values in x (and w, if provided) be removed before computation? Default TRUE.

Details

Pass w to compute a weighted geometric mean, defined as $\exp(\text{weighted.mean}(\log(x), w))$.

Value

A numeric scalar. Returns NA_real_ when the input is empty, when it is entirely NA, or when na.rm = FALSE and any element is NA. Returns NaN with a warning when x contains negative values, or when all weights are zero.

Examples

```
# Unweighted
geomean(c(1, 4, 16))           # 4
geomean(c(0.6, 0.8, 0.95))    # ~0.772 - typical UHC tracer aggregation
geomean(c(1, NA, 4))          # 2
geomean(c(1, NA, 4), na.rm = FALSE) # NA_real_
geomean(c(1, 0, 4))           # 0

# Weighted
geomean(c(1, 4, 16), w = c(1, 1, 1)) # 4 (equal weights = unweighted)
geomean(c(1, 4, 16), w = c(1, 2, 1)) # weighted toward 4
geomean(c(0.6, 0.8, 0.95), w = c(2, 1, 1))
```

ggpie

Create a Pie Chart with ggplot2

Description

Builds a pie chart from a data frame using one categorical column and one numeric column. Slices are labeled with the category name and percentage share.

Usage

```
ggpie(
  df,
  .x,
  .y,
  .offset = 1,
```

```

    .color = "white",
    .legend = FALSE,
    .label = TRUE,
    .label_size = 3.5
  )

```

Arguments

<code>df</code>	A data frame.
<code>.x</code>	Column name (string) of the categorical variable used for the slices.
<code>.y</code>	Column name (string) of the numeric variable used for the slice values.
<code>.offset</code>	Numeric scalar (> 0). Controls label position along the slice radius. Default 1 places the label at the middle of the slice. Smaller values (e.g. 0.5) move the label inward toward the centre; larger values (e.g. 2) move it outward toward the edge or beyond, useful for donut-style layouts.
<code>.color</code>	Border color between slices. Default "white".
<code>.legend</code>	Logical. Show the legend? Default FALSE.
<code>.label</code>	Logical. Draw "name\n(pct%)" labels on the slices? Default TRUE.
<code>.label_size</code>	Label text size in mm. Default 3.5.

Value

A ggplot object.

Examples

```

df <- data.frame(
  category = c("A", "B", "C"),
  value = c(40, 35, 25)
)
ggpie(df, "category", "value")

```

gho_clean

Tidy a GHO Data Frame

Description

Selects, renames, and type-casts the most useful columns from a GHO observation table returned by `gho_data()`, producing a compact tibble in the **unified DSIR cleaned-indicator schema** — the same schema produced by `sdg_clean()`, so the two outputs can be combined directly with `bind_indicators()`.

Usage

```
gho_clean(df)
```

Arguments

`df` A data frame returned by `gho_data()`.

Details

The mapping (GHO source → unified column) is:

- `IndicatorCode` → `id`
- `IndicatorCode` resolved against the GHO indicator catalog → `indicator` (the human-readable name; cached at session level after the first call)
- `SpatialDim` → `location`; also `iso3` when it matches a WHO Member State, otherwise `iso3 = NA`
- `TimeDim` → `year` (integer)
- `Value` → `value` (character; raw)
- `NumericValue` → `value_num` (numeric)
- `Low, High` → `low, high` (numeric)
- `Dim1, Dim2, Dim3` → `dim1, dim2, dim3` (character)

The `series` column is always NA for GHO output (it is an SDG-only concept). The `location_name` column is populated by looking up `location` (an ISO3 code or a WHO region code) against the `who_countries` dataset and a hardcoded set of WHO regional names; locations that match neither (e.g. non-Member State areas) are left as NA.

Source columns absent from `df` (e.g. `Low / High` for indicators without confidence intervals) are filled with typed NA, so the output always has the same 15 columns with the same column types.

The GHO data endpoint (`/api/{IndicatorCode}`) does not return `IndicatorName`; that field lives on the catalog endpoint queried by `gho_indicators()`. On the first call within an R session, `gho_clean()` fetches the catalog once and caches it for the rest of the session, so the `indicator` column carries the full human-readable indicator name. If the catalog cannot be fetched (e.g. no network), `gho_indicators()` emits a warning and the `indicator` column falls back to NA.

Value

A `tibble` with 15 columns: `source` (always "gho"), `id`, `indicator`, `location`, `iso3`, `location_name`, `year`, `value`, `value_num`, `low`, `high`, `series` (NA), `dim1`, `dim2`, `dim3`. Sorted by `location` then `year`. Empty input returns an empty tibble with the same columns and types.

See Also

`gho_data()`, `sdg_clean()`, `bind_indicators()`.

Examples

```
gho_data("NCDMORT3070", spatial_type = "country") |>
  gho_clean()
```

gho_count

*Count Observations for a GHO Indicator Filter***Description**

Sends a `$top=0&&$count=true` request to the WHO GHO OData API, which returns the matching row count without transferring any observations. Useful for sizing a download before issuing it.

Usage

```
gho_count(
  indicator,
  spatial_type = NULL,
  area = NULL,
  year_from = NULL,
  year_to = NULL
)
```

Arguments

<code>indicator</code>	Character scalar. The indicator code (e.g. "NCDMORT3070"). Use gho_indicators() to find codes.
<code>spatial_type</code>	Character. Spatial dimension to filter on: one of "country", "region", "global", or NULL (all levels, the default).
<code>area</code>	Character vector of country or region codes (e.g. <code>c("FRA", "DEU")</code>). Default NULL returns all areas.
<code>year_from</code>	Numeric. Start year filter (inclusive). Default NULL.
<code>year_to</code>	Numeric. End year filter (inclusive). Default NULL.

Value

An integer scalar — the number of observations the server would return for the same filter via [gho_data\(\)](#). Returns `NA_integer_` (with a warning) if the request fails.

See Also

[gho_data\(\)](#), [gho_has_data\(\)](#), [gho_coverage\(\)](#).

Examples

```
# How many rows would gho_data() pull for France?
gho_count("WHOSIS_000001", area = "FRA")

# Compare coverage across regions
gho_count("NCDMORT3070", spatial_type = "country")
gho_count("NCDMORT3070", spatial_type = "region")
```

`gho_coverage`*Summarise Per-Location Data Coverage of a GHO Indicator*

Description

Fetches only the `SpatialDim` and `TimeDim` columns for a GHO indicator (much lighter than `gho_data()`) and summarises the year range and observation count per location. Useful for answering "which countries have data, and for what years?" before committing to a full download.

Usage

```
gho_coverage(  
  indicator,  
  spatial_type = "country",  
  area = NULL,  
  year_from = NULL,  
  year_to = NULL  
)
```

Arguments

<code>indicator</code>	Character scalar. The indicator code (e.g. "WHOSIS_000001").
<code>spatial_type</code>	Character. Spatial dimension to filter on: one of "country", "region", "global". Defaults to "country" since per-country coverage is the typical use case. Pass NULL for all spatial levels.
<code>area</code>	Character vector of country or region codes (e.g. <code>c("FRA", "DEU")</code>). Default NULL returns all areas for the chosen <code>spatial_type</code> .
<code>year_from</code>	Numeric. Start year filter (inclusive). Default NULL.
<code>year_to</code>	Numeric. End year filter (inclusive). Default NULL.

Value

A [tibble](#) with one row per location and columns:

- `location` (chr) — the `SpatialDim` value (typically ISO3).
- `year_min` (int) — earliest year with data.
- `year_max` (int) — latest year with data.
- `n_obs` (int) — number of observations.

Sorted by location. Empty input or service failure returns an empty tibble with the same four columns.

See Also

[gho_data\(\)](#), [gho_has_data\(\)](#), [gho_count\(\)](#).

Examples

```
# Year coverage of life expectancy for three countries
gho_coverage("WHOSIS_000001", area = c("FRA", "DEU", "JPN"))

# All countries with any life-expectancy data, since 2010
gho_coverage("WHOSIS_000001", year_from = 2010)
```

gho_data

*Fetch GHO Data***Description**

Retrieves observations for a specific indicator from the WHO GHO OData API, with optional filters by spatial level, country / region and year range.

Usage

```
gho_data(
  indicator,
  spatial_type = NULL,
  area = NULL,
  year_from = NULL,
  year_to = NULL
)
```

Arguments

indicator	Character scalar. The indicator code (e.g. "NCDMORT3070"). Use gho_indicators() to find codes.
spatial_type	Character. Spatial dimension to filter on: one of "country", "region", "global", or NULL (all levels, the default).
area	Character vector of country or region codes (e.g. c("FRA", "DEU")). Default NULL returns all areas.
year_from	Numeric. Start year filter (inclusive). Default NULL.
year_to	Numeric. End year filter (inclusive). Default NULL.

Value

A [tibble](#) of indicator observations, or an empty tibble when the service is unreachable.

See Also

[gho_indicators\(\)](#), [gho_dimensions\(\)](#).

Examples

```
# Country-level data for one indicator
gho_data("NCDMORT3070", spatial_type = "country")

# Specific countries and years
gho_data("WHOSIS_000001", area = c("FRA", "DEU"), year_from = 2015)
```

gho_dimensions	<i>List Dimensions of a GHO Indicator</i>
----------------	---

Description

Returns the unique values of a given dimension across all observations of a GHO indicator. Useful for discovering which ages, sexes, regions, or other breakdowns are available before calling [gho_data\(\)](#).

Usage

```
gho_dimensions(indicator, dimension = "SpatialDimType")
```

Arguments

indicator	Character scalar. The indicator code (e.g. "NCDMORT3070").
dimension	Character. Name of the dimension column in the indicator data. Common values include "SpatialDim", "SpatialDimType", "TimeDim", "Dim1", "Dim2", and "Dim3". Default "SpatialDimType".

Value

A character vector of unique, sorted dimension values, or an empty character vector when the service is unreachable or the dimension is missing.

See Also

[gho_data\(\)](#), [gho_indicators\(\)](#).

Examples

```
gho_dimensions("NCDMORT3070")
gho_dimensions("NCDMORT3070", dimension = "Dim1")
```

`gho_has_data`*Check Whether a GHO Indicator Has Data for a Filter*

Description

Sends a minimal request (`$top=1&$select=Id`) to the WHO GHO OData API to find out whether any observations exist for the given indicator and filter combination, without downloading the full result set. Useful as a quick precheck before `gho_data()`.

Usage

```
gho_has_data(  
  indicator,  
  spatial_type = NULL,  
  area = NULL,  
  year_from = NULL,  
  year_to = NULL  
)
```

Arguments

<code>indicator</code>	Character scalar. The indicator code (e.g. "NCDMORT3070"). Use <code>gho_indicators()</code> to find codes.
<code>spatial_type</code>	Character. Spatial dimension to filter on: one of "country", "region", "global", or NULL (all levels, the default).
<code>area</code>	Character vector of country or region codes (e.g. <code>c("FRA", "DEU")</code>). Default NULL returns all areas.
<code>year_from</code>	Numeric. Start year filter (inclusive). Default NULL.
<code>year_to</code>	Numeric. End year filter (inclusive). Default NULL.

Value

A logical scalar:

- TRUE if at least one observation exists for the filter.
- FALSE if the server returns an empty result.
- NA if the request fails (network failure, unreachable host, or the indicator code does not exist and the server returns an HTTP error). A warning is emitted in the failure case.

See Also

`gho_data()`, `gho_count()`, `gho_coverage()`.

Examples

```
# Does WHO have life-expectancy data for France?
gho_has_data("WHOSIS_000001", area = "FRA")

# Quickly screen a list of indicators before downloading any data
inds <- c("WHOSIS_000001", "NCDMORT3070")
vapply(inds, gho_has_data, logical(1), area = "FRA")
```

gho_indicators	<i>List GHO Indicators</i>
----------------	----------------------------

Description

Fetches the catalog of indicators from the WHO Global Health Observatory (GHO) OData API.

Usage

```
gho_indicators(search = NULL)
```

Arguments

search	Optional character. Search keywords matched against IndicatorName (case-insensitive). All terms must match (AND semantics). Accepts either: <ul style="list-style-type: none">• a single string, which is split on whitespace into terms (e.g. "child mortality" matches indicators containing both "child" and "mortality"), or• a character vector, whose elements are used as terms verbatim (whitespace inside an element is treated as part of the term). Single quotes in any term are escaped for the OData filter.
--------	---

Value

A [tibble](#) with columns IndicatorCode, IndicatorName and Language. Returns an empty tibble (with a message) when the service is unreachable.

See Also

[gho_data\(\)](#), [gho_dimensions\(\)](#).

Examples

```
# All indicators
inds <- gho_indicators()

# Single keyword
gho_indicators("mortality")

# Multiple keywords from one string (AND): both terms must appear
```

```
gho_indicators("child mortality")

# Or pass terms as a vector
gho_indicators(c("child", "mortality"))
```

 iso3_to_m49

Convert ISO3 Codes to UN M49 Numeric Codes

Description

Maps ISO 3166-1 alpha-3 country codes to UN M49 numeric area codes using the [who_countries](#) dataset shipped with DSIR. Useful when moving from data sources keyed by ISO3 (e.g. the WHO GHO API) to sources keyed by M49 (e.g. the UN SDG API).

Usage

```
iso3_to_m49(iso3)
```

Arguments

`iso3` Character vector of ISO3 codes. Case-insensitive; values are upper-cased before lookup.

Details

Codes that do not correspond to a WHO Member State return NA. This includes Associate Members (e.g. Puerto Rico) and other non-Member areas that some indicator data sets cover.

Most users will not need to call this function directly: [sdg_data\(\)](#) and [sdg_coverage\(\)](#) accept ISO3 codes for their area argument and convert internally. This helper is exported for cases where you want to inspect or manipulate the conversion yourself.

Value

A character vector the same length as `iso3`, with M49 codes in the same format as `who_countries$m49_code` (three-character zero-padded strings, e.g. "076"). Non-Member areas return NA.

See Also

[who_countries](#), [iso3_to_region\(\)](#), [sdg_data\(\)](#).

Examples

```
iso3_to_m49(c("PHL", "FRA", "JPN"))
# "608" "250" "392"

# Case-insensitive
iso3_to_m49("phl")
# "608"
```

```
# Non-Member areas return NA
iso3_to_m49(c("PRI", "PHL"))
# NA "608"
```

iso3_to_region *Look Up the WHO Region for ISO3 Codes*

Description

Maps ISO 3166-1 alpha-3 country codes to WHO region codes using the [who_countries](#) dataset shipped with DSIR. Stays in sync with WHO governance changes reflected in DSIR — for example, Indonesia's reassignment from SEAR to WPR following EB156 (May 2025).

Usage

```
iso3_to_region(iso3, long = FALSE)
```

Arguments

iso3	Character vector of ISO3 codes. Case-sensitive (uppercase, as in who_countries).
long	Logical. If TRUE, return long-form region names (e.g. "Western Pacific"). If FALSE (default), return the short codes used elsewhere in DSIR: "AFR", "AMR", "SEAR", "EUR", "EMR", "WPR".

Details

Codes that do not correspond to a WHO Member State return NA. This includes Associate Members (Puerto Rico, Tokelau) and other non-Member areas that some indicator data sets cover.

Value

A character vector the same length as iso3.

See Also

[who_countries](#), [wpro_cty](#).

Examples

```
iso3_to_region(c("PHL", "FRA", "USA", "COK"))
# "WPR" "EUR" "AMR" "WPR"

iso3_to_region(c("IDN", "JPN"), long = TRUE)
# "Western Pacific" "Western Pacific" (Indonesia in WPR since May 2025)

# Non-Member areas return NA
iso3_to_region(c("PRI", "TKL", "PHL"))
# NA NA "WPR"
```

m49_to_iso3

Convert UN M49 Numeric Codes to ISO3 Codes

Description

Maps UN M49 numeric area codes to ISO 3166-1 alpha-3 country codes using the [who_countries](#) dataset shipped with DSIR. Counterpart to [iso3_to_m49\(\)](#) and used internally by [sdg_clean\(\)](#) to populate the `iso3` column on SDG output.

Usage

```
m49_to_iso3(m49)
```

Arguments

`m49` Character vector of M49 codes.

Details

M49 codes that do not correspond to a WHO Member State return NA. This includes region / world aggregates (e.g. "900" for World, "001" for World, "419" for Latin America and the Caribbean) and codes for non-Member areas (e.g. Puerto Rico, Tokelau).

Input accepts either the zero-padded form ("076") or the bare form ("76"); both are normalised before lookup. Non-numeric input returns NA (with a single warning from the underlying [as.integer\(\)](#) coercion).

Value

A character vector the same length as `m49`. Non-Member codes (region aggregates, non-Member areas) return NA.

See Also

[who_countries](#), [iso3_to_m49\(\)](#), [sdg_clean\(\)](#).

Examples

```
m49_to_iso3(c("608", "250", "392"))
# "PHL" "FRA" "JPN"

# Zero-padded and bare forms both accepted
m49_to_iso3(c("076", "76"))
# "BRA" "BRA"

# Non-Member areas / aggregates return NA
m49_to_iso3(c("900", "608"))
# NA "PHL"
```

pic_cty *Pacific Island Country ISO3 codes*

Description

Character vector of ISO3 codes for the 14 WHO Member States classified as Pacific Island Countries (PICs). Sorted alphabetically.

Usage

```
pic_cty
```

Format

A character vector of 14 ISO 3166-1 alpha-3 codes.

Details

All 14 PICs are within the Western Pacific Region, so `all(pic_cty %in% wpro_cty)` is TRUE. Non-Member Pacific areas (e.g. New Caledonia, French Polynesia, American Samoa, Tokelau) are not included.

The 14 PIC Member States are: Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, and Vanuatu.

See Also

[who_countries](#), [wpro_cty](#).

Examples

```
length(pic_cty)      # 14
all(pic_cty %in% wpro_cty) # TRUE
```

scale_dsi_col *Continuous Scales for DSIR Bar / Column Charts*

Description

Thin wrappers around `ggplot2::scale_y_continuous()` and `ggplot2::scale_x_continuous()` that remove the default lower expansion so that columns sit flush with the axis — the convention for WHO and most publication-style bar charts. The upper expansion is preserved at 5% so the tallest column has breathing room above (or to the right of) it.

Usage

```
scale_y_dsi_col(...)
```

```
scale_x_dsi_col(...)
```

Arguments

... Arguments forwarded to the underlying `ggplot2::scale_y_continuous()` / `ggplot2::scale_x_continuous()`.

Details

Use `scale_y_dsi_col()` for vertical bars (`geom_col()` / `geom_bar()`) and `scale_x_dsi_col()` when bars are horizontal (via `coord_flip()` or `geom_col(orientation = "y")`).

Pass any other `scale_*_continuous()` argument (labels, breaks, limits, ...) through

Value

A ggplot2 Scale object, to be added to a plot with +.

Examples

```
library(ggplot2)

# Vertical bars
ggplot(mtcars, aes(factor(cyl))) +
  geom_bar(fill = "#0093D5") +
  scale_y_dsi_col() +
  theme_dsi()
```

sdg_areas

List SDG Geographic Areas

Description

Fetches the list of geographic areas available from the UN SDG database.

Usage

```
sdg_areas()
```

Value

A `tibble` with area codes and names, or NULL when the service is unreachable.

See Also

[sdg_data\(\)](#).

Examples

```
sdg_areas()
```

```
sdg_clean
```

Tidy an SDG Data Frame

Description

Selects, renames, and type-casts the most useful columns from an SDG observation table returned by `sdg_data()`, producing a compact tibble in the **unified DSIR cleaned-indicator schema** — the same schema produced by `gho_clean()`, so the two outputs can be combined directly with `bind_indicators()`.

Usage

```
sdg_clean(df)
```

Arguments

`df` A data frame returned by `sdg_data()`.

Details

The mapping (SDG source → unified column) is:

- `indicator` (list-column, flattened) → `id` (e.g. "3.4.1")
- `seriesDescription` → `indicator` (human-readable label; NA if the API response does not include it)
- `geoAreaCode` → `location` (UN M49 numeric, as character); also `iso3` via `m49_to_iso3()` for WHO Member States — region / world aggregates and non-Member areas get `iso3 = NA`
- `location_name` is resolved by looking up `iso3` against `who_countries` (so a WHO Member State has the same `location_name` here and in `gho_clean()` output), with a fallback to the SDG API's raw `geoAreaName` for non-Member-State rows (e.g. regional / world aggregates)
- `timePeriodStart` → `year` (integer)
- `value` → `value` (character; raw) and `value_num` (numeric; NA for non-numeric entries like "<0.1" or aggregate notes)
- `lowerBound`, `upperBound` → `low`, `high` (numeric)
- `series` → `series`

Three columns are always present but never populated for SDG output: `dim1`, `dim2`, `dim3` (GHO-only concepts).

Value

A [tibble](#) with 15 columns: source (always "sdg"), id, indicator, location, iso3, location_name, year, value, value_num, low, high, series, dim1 (NA), dim2 (NA), dim3 (NA). Sorted by location then year. Empty input returns an empty tibble with the same columns and types.

See Also

[sdg_data\(\)](#), [gho_clean\(\)](#), [bind_indicators\(\)](#), [m49_to_iso3\(\)](#).

Examples

```
sdg_data("3.2.1", area = "156", year_from = 2015) |>
  sdg_clean()
```

sdg_coverage

Explore Series Coverage of an SDG Indicator

Description

A single SDG indicator (for example "3.4.1", NCD mortality) is typically published as several **series** stratified by sex, age, or cause. Different series may have different country and year coverage. `sdg_coverage()` summarises year range and observation count per (location, series) combination, so you can see which series exist for an indicator and how each one is covered before committing to a downstream analysis.

Usage

```
sdg_coverage(indicator, area = NULL, year_from = NULL, year_to = NULL)
```

Arguments

indicator	Character vector of SDG indicator codes (e.g. "3.4.1").
area	Character vector of country/area codes. Accepts either ISO3 codes (e.g. <code>c("PHL", "FRA")</code>) — converted automatically via iso3_to_m49() — or UN M49 numeric codes (e.g. <code>c("608", "250")</code>) as returned by sdg_areas() . Do not mix the two formats in a single call. Default NULL returns all areas. Unknown ISO3 codes are dropped with a warning before the network call.
year_from	Numeric. Start year filter (inclusive). Default NULL.
year_to	Numeric. End year filter (inclusive). Default NULL.

Details

Unlike the GHO availability helpers, this function is a series-exploration tool rather than a payload-saving precheck: SDG data is generally complete enough that GHO-style `has_data()` / `count()` helpers add little value, so they are intentionally not provided. The SDG API also offers no payload-reduction option (no `$select` equivalent), so `sdg_coverage()` calls [sdg_data\(\)](#) internally and aggregates the result client-side.

Value

A [tibble](#) with one row per (location, series) and columns:

- location (chr) — area code (geoAreaCode).
- series (chr) — SDG series code.
- year_min (int) — earliest year with data.
- year_max (int) — latest year with data.
- n_obs (int) — number of observations.

Sorted by location then series. Empty input or service failure returns an empty tibble with the same five columns.

See Also

[sdg_data\(\)](#), [sdg_indicators\(\)](#), [gho_coverage\(\)](#).

Examples

```
# Series available for NCD mortality in China and Brazil
sdg_coverage("3.4.1", area = c("156", "076"))

# Filter to a year range
sdg_coverage("3.4.1", area = "156", year_from = 2015)
```

sdg_data

Fetch SDG Data

Description

Retrieves data for one or more SDG indicators from the UN SDG API, with optional filters by area and year.

Usage

```
sdg_data(  
  indicator,  
  area = NULL,  
  year_from = NULL,  
  year_to = NULL,  
  page_size = 1000L  
)
```

Arguments

indicator	Character vector of indicator codes (e.g. "1.1.1"). Use <code>sdg_indicators()</code> to find codes.
area	Character vector of country/area codes. Accepts either ISO3 codes (e.g. <code>c("PHL", "FRA")</code>) — converted automatically via <code>iso3_to_m49()</code> — or UN M49 numeric codes (e.g. <code>c("608", "250")</code>) as returned by <code>sdg_areas()</code> . Do not mix the two formats in a single call. Default NULL returns all areas.
year_from	Numeric. Start year filter (inclusive). Default NULL.
year_to	Numeric. End year filter (inclusive). Default NULL.
page_size	Integer. Number of records per page. Default 1000, maximum 10000.

Value

A [tibble](#) of indicator observations, or an empty tibble when the service is unreachable or there are no matching rows.

See Also

[sdg_indicators\(\)](#), [sdg_areas\(\)](#), [iso3_to_m49\(\)](#).

Examples

```
# One indicator, one country – the typical entry point
sdg_data("1.1.1", area = "PHL")

# Specific area and year range (M49 code)
sdg_data("3.2.1", area = "156", year_from = 2015, year_to = 2023)

# ISO3 codes work directly – DSIR's regional vectors can be passed in
sdg_data("3.4.1", area = c("PHL", "FRA", "JPN"))
```

sdg_goals

List SDG Goals

Description

Fetches the list of Sustainable Development Goals from the UN SDG API.

Usage

```
sdg_goals(include_children = FALSE)
```

Arguments

include_children	Logical. Include targets and indicators nested under each goal? Default FALSE.
------------------	--

Value

A list (or [tibble](#)) of SDG goals, or NULL when the service is unreachable.

See Also

[sdg_targets\(\)](#), [sdg_indicators\(\)](#), [sdg_data\(\)](#).

Examples

```
sdg_goals()
sdg_goals(include_children = TRUE)
```

sdg_indicators

List SDG Indicators

Description

Fetches the list of SDG indicators from the UN SDG API, with optional keyword filtering on the indicator description.

Usage

```
sdg_indicators(search = NULL)
```

Arguments

search Optional character. Search keywords matched against the description column (case-insensitive). All terms must match (AND semantics). Accepts either:

- a single string, which is split on whitespace into terms (e.g. "mortality cancer" keeps rows whose description contains both "mortality" and "cancer"), or
- a character vector, whose elements are used as terms verbatim (so a term may itself contain whitespace, e.g. c("mortality rate", "attributed")).

The filter is applied client-side using [grepl\(\)](#) with `fixed = TRUE` because the UN SDG /Indicator/List endpoint is not OData and exposes no server-side search parameter; the full list is small (~250 rows) so this is cheap.

Value

A list (or [tibble](#)) of SDG indicators, or NULL when the service is unreachable. When search matches no rows, an empty tibble with the same columns as the unfiltered response is returned.

See Also

[sdg_targets\(\)](#), [sdg_data\(\)](#).

Examples

```
# Full list
sdg_indicators()

# Single keyword
sdg_indicators("mortality")

# Multi-keyword – AND semantics
sdg_indicators("mortality cancer")
sdg_indicators(c("maternal", "mortality"))
```

sdg_targets

List SDG Targets

Description

Fetches the list of SDG targets from the UN SDG API.

Usage

```
sdg_targets(include_children = FALSE)
```

Arguments

`include_children`
Logical. Include indicators nested under each target? Default FALSE.

Value

A list (or [tibble](#)) of SDG targets, or NULL when the service is unreachable.

See Also

[sdg_goals\(\)](#), [sdg_indicators\(\)](#).

Examples

```
sdg_targets()
```

 theme_dsi

DSIR ggplot2 theme — WHO publication-style

Description

A clean, modern theme tuned for WHO and global-health publications. Removes the panel border, draws light grid lines, and uses a muted text colour so that the data — not the chart chrome — is the visual focus. The `grid` argument controls which direction(s) the grid lines run, so the theme works equally well for vertical bars, horizontal bars (via `coord_flip()`), scatter plots, and line charts.

Usage

```
theme_dsi(
  base_size = 12,
  base_family = "",
  accent = "#0093D5",
  grid_color = "grey92",
  grid = c("both", "x", "y", "none"),
  legend_position = "bottom"
)
```

Arguments

<code>base_size</code>	Base font size in points. Default 12.
<code>base_family</code>	Base font family. Default "" (system default). Set to "sans", "Arial", "Helvetica", or "Calibri" for a specific look. Empty default keeps the package portable on CRAN's Linux check machines, where Calibri is unavailable.
<code>accent</code>	Accent colour used for axis lines and as a default for highlight elements. Default "#0093D5" (WHO blue). Pass any colour string.
<code>grid_color</code>	Colour of the major grid. Default "grey92" — a very light grey, close to <code>bbplot</code> and OECD style.
<code>grid</code>	Which direction(s) to draw major grid lines. One of "both" (default — draws both horizontal and vertical, works correctly under <code>coord_flip()</code>), "y" (only horizontal grid lines, the look used in DSIR <= 0.5.x), "x" (only vertical grid lines), or "none".
<code>legend_position</code>	Position of the legend. Default "bottom". Pass "none", "top", "right", or a numeric vector <code>c(x, y)</code> .

Value

A `ggplot2` theme object.

See Also

[theme_dsi_facet\(\)](#) for a sibling theme tuned for faceted plots.

Examples

```
library(ggplot2)

# Default - grid in both directions, works under coord_flip()
ggplot(mtcars, aes(wt, mpg, color = factor(cyl))) +
  geom_point(size = 3) +
  theme_dsi() +
  labs(title = "Fuel efficiency by weight",
       x = "Weight (1000 lbs)", y = "Miles per gallon",
       color = "Cylinders")

# Minimal look - only horizontal grid lines
ggplot(mtcars, aes(wt, mpg)) +
  geom_point(size = 3, color = "#0093D5") +
  theme_dsi(grid = "y") +
  labs(x = "Weight (1000 lbs)", y = "Miles per gallon")
```

 theme_dsi_facet

DSIR ggplot2 theme for faceted plots

Description

A sibling of `theme_dsi()` tuned for faceted plots. Where `theme_dsi()` uses half-frame axis lines and only horizontal grid lines — a look that suits a single panel — repeating those across every facet looks heavy and the panels run together. `theme_dsi_facet()` replaces the axis lines with a light panel border, draws grid lines on both axes, gives the strip a soft background, and inserts whitespace between panels.

Usage

```
theme_dsi_facet(
  base_size = 12,
  base_family = "",
  accent = "#0093D5",
  grid_color = "grey92",
  strip_fill = "grey95",
  strip_color = "grey20",
  grid = c("both", "x", "y", "none"),
  legend_position = "bottom"
)
```

Arguments

<code>base_size</code>	Base font size in points. Default 12.
<code>base_family</code>	Base font family. Default "" (system default). Set to "sans", "Arial", "Helvetica", or "Calibri" for a specific look. Empty default keeps the package portable on CRAN's Linux check machines, where Calibri is unavailable.

accent	Accent colour, kept for parity with <code>theme_dsi()</code> so that the argument set is interchangeable. Not currently used in the facet variant — the panel border replaces the accent-coloured axis lines. Default "#0093D5" (WHO blue).
grid_color	Colour of the major grid (both axes). Default "grey92".
strip_fill	Background fill colour for facet strips. Default "grey95" — a light neutral grey. Avoid blues, which clash with the WHO-blue accent.
strip_color	Text colour inside facet strips. Default "grey20".
grid	Which direction(s) to draw major grid lines. One of "both" (default — both horizontal and vertical), "y" (only horizontal), "x" (only vertical), or "none". Matches the identical argument on <code>theme_dsi()</code> so the two themes can be swapped without changing other settings.
legend_position	Position of the legend. Default "bottom". Pass "none", "top", "right", or a numeric vector <code>c(x, y)</code> .

Details

Use `theme_dsi()` for single-panel plots and `theme_dsi_facet()` for plots with `facet_wrap()` or `facet_grid()`. Shared elements (text styles, title block, legend, plot margins) match `theme_dsi()` exactly, so the two themes feel like the same family.

Value

A `ggplot2` theme object.

See Also

[theme_dsi\(\)](#) for the single-panel sibling.

Examples

```
library(ggplot2)
ggplot(mtcars, aes(wt, mpg)) +
  geom_point(size = 2, color = "#0093D5") +
  facet_wrap(~ cyl, labeller = label_both) +
  theme_dsi_facet() +
  labs(title = "Fuel efficiency by cylinder count",
       x = "Weight (1000 lbs)", y = "Miles per gallon")
```

who_countries

WHO Member States with regional and Pacific classifications

Description

A tibble of the 194 World Health Organization (WHO) Member States, with standard country identifiers and WHO/Pacific groupings used across DSIR analytical workflows.

Usage

who_countries

Format

A tibble with 194 rows and 7 columns:

iso3 ISO 3166-1 alpha-3 code (3-letter, e.g. "PHL").

iso2 ISO 3166-1 alpha-2 code (2-letter, e.g. "PH").

m49_code UN M49 numeric code, stored as a 3-character string with leading zeros where present (e.g. "008" for Albania). Stored as character because some downstream APIs (notably the UN SDG API) expect the leading-zero form.

name_official WHO official English name (e.g. "Iran (Islamic Republic of)"). Use for formal documents and reports.

name_short A shorter form suitable for charts and tables (e.g. "Iran"). Equal to name_official for countries whose official name is already concise. See Details.

who_region WHO region code: one of "AFR", "AMR", "SEAR", "EUR", "EMR", "WPR".

is_pic Logical. TRUE for the 14 Pacific Island Country (PIC) Member States in WPR, FALSE otherwise.

Details

Scope. WHO has 194 Member States plus 2 Associate Members (Puerto Rico, Tokelau) and reports on additional non-Member areas (e.g. West Bank and Gaza Strip). This dataset includes only the 194 Member States. Cook Islands and Niue are full WHO Member States and are included even though they are not UN member states.

Region coverage (as of May 2025, after WHO EB156 reassignment of Indonesia from SEAR to WPR):

- AFR (African Region): 47
- AMR (Region of the Americas): 35
- SEAR (South-East Asia Region): 10
- EUR (European Region): 53
- EMR (Eastern Mediterranean Region): 21
- WPR (Western Pacific Region): 28

Short names. name_short differs from name_official for 13 countries where the official name is a parenthetical descriptor or is otherwise long. Examples: "DPR Korea", "DR Congo", "Lao PDR", "United Kingdom". These short forms follow conventions used in WHO regional reports and OECD Health at a Glance: Asia/Pacific.

PICs. The Pacific Island Countries flag (is_pic) marks the 14 WHO Member States in the Pacific sub-region: Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, and Vanuatu. Non-Member Pacific areas (e.g. New Caledonia, French Polynesia, American Samoa) are not included in this dataset.

Source

- WHO official region listing: <https://www.who.int/countries>
- ISO 3166-1 codes and UN M49 numeric codes: UN Statistics Division <https://unstats.un.org/unsd/methodology/m49/>

Examples

```
# All WPR Member States, sorted alphabetically by short name
wpr <- who_countries[who_countries$who_region == "WPR", ]
wpr <- wpr[order(wpr$name_short), c("iso3", "name_short")]
head(wpr)

# Filter a data frame to PIC member states
# df_pic <- subset(my_data, country_iso3 %in% who_countries$iso3[who_countries$is_pic])
```

who_region_vectors *WHO regional Member State ISO3 vectors*

Description

Convenience character vectors of ISO3 codes for the WHO Member States in each region. Each vector is the regional subset of `who_countries`'s `iso3` column, sorted alphabetically.

Usage

```
afro_cty
amro_cty
searo_cty
euro_cty
emro_cty
wpro_cty
```

Format

Character vectors of ISO 3166-1 alpha-3 codes.

- An object of class character of length 47.
- An object of class character of length 35.
- An object of class character of length 10.
- An object of class character of length 53.
- An object of class character of length 21.
- An object of class character of length 28.

Details

These are derived directly from [who_countries](#) and exist for ergonomic filtering, e.g. `dplyr::filter(data, iso3 %in% wpro_cty)`. If you need to update group membership, edit `data-raw/who_countries.R` and re-run that script — the vectors regenerate from the master table.

`afro_cty` 47 Member States in the WHO African Region.

`amro_cty` 35 Member States in the WHO Region of the Americas.

`searo_cty` 10 Member States in the WHO South-East Asia Region.

`euro_cty` 53 Member States in the WHO European Region.

`emro_cty` 21 Member States in the WHO Eastern Mediterranean Region.

`wpro_cty` 28 Member States in the WHO Western Pacific Region. Includes Indonesia from May 2025 (per WHO EB156).

See Also

[who_countries](#), [pic_cty](#).

Examples

```
length(wpro_cty)      # 28
"IDN" %in% wpro_cty   # TRUE - Indonesia in WPR since May 2025
```

Index

* datasets

- pic_cty, [17](#)
 - who_countries, [27](#)
 - who_region_vectors, [29](#)
- afro_cty (who_region_vectors), [29](#)
amro_cty (who_region_vectors), [29](#)
as.integer(), [16](#)
- bind_indicators, [2](#)
bind_indicators(), [6, 7, 19, 20](#)
- dsi_flextable_defaults, [3](#)
- emro_cty (who_region_vectors), [29](#)
euro_cty (who_region_vectors), [29](#)
- flextable::set_flextable_defaults(), [4](#)
- geomean, [4](#)
ggpie, [5](#)
ggplot2::scale_x_continuous(), [17, 18](#)
ggplot2::scale_y_continuous(), [17, 18](#)
gho_clean, [6](#)
gho_clean(), [2, 3, 19, 20](#)
gho_count, [8](#)
gho_count(), [9, 12](#)
gho_coverage, [9](#)
gho_coverage(), [8, 12, 21](#)
gho_data, [10](#)
gho_data(), [6–9, 11–13](#)
gho_dimensions, [11](#)
gho_dimensions(), [10, 13](#)
gho_has_data, [12](#)
gho_has_data(), [8, 9](#)
gho_indicators, [13](#)
gho_indicators(), [7, 8, 10–12](#)
grepl(), [23](#)
- iso3_to_m49, [14](#)
iso3_to_m49(), [16, 20, 22](#)
- iso3_to_region, [15](#)
iso3_to_region(), [14](#)
- m49_to_iso3, [16](#)
m49_to_iso3(), [19, 20](#)
- pic_cty, [17, 30](#)
- scale_dsi_col, [17](#)
scale_x_dsi_col (scale_dsi_col), [17](#)
scale_y_dsi_col (scale_dsi_col), [17](#)
sdg_areas, [18](#)
sdg_areas(), [20, 22](#)
sdg_clean, [19](#)
sdg_clean(), [2, 3, 6, 7, 16](#)
sdg_coverage, [20](#)
sdg_coverage(), [14](#)
sdg_data, [21](#)
sdg_data(), [14, 18–21, 23](#)
sdg_goals, [22](#)
sdg_goals(), [24](#)
sdg_indicators, [23](#)
sdg_indicators(), [21–24](#)
sdg_targets, [24](#)
sdg_targets(), [23](#)
searo_cty (who_region_vectors), [29](#)
- theme_dsi, [25](#)
theme_dsi(), [26, 27](#)
theme_dsi_facet, [26](#)
theme_dsi_facet(), [25](#)
tibble, [3, 7, 9, 10, 13, 18, 20–24](#)
- who_countries, [7, 14–17, 19, 27, 29, 30](#)
who_region_vectors, [29](#)
wpro_cty, [15, 17](#)
wpro_cty (who_region_vectors), [29](#)