Package 'wdiexplorer'

April 21, 2026

Type Package

Title Explore World Development Indicators Data

Version 0.1.2

Description Provides a workflow for exploring World Development Indicators (WDI) country-level panel data. It downloads WDI data using the 'WDI' package and computes diagnostic indices that capture the temporal behaviour of the data by incorporating the grouping structure of the data. The set of diagnostic indices implemented includes variation features, trend and shape features, and sequential temporal features. This method is described in Akinfenwa, Cahill, and Hurley (2025) ``'wdiexplorer': An R package Designed for Exploratory Analysis of World Development Indicators (WDI) Data" <doi:10.48550/arXiv.2511.07027>. We adapt the clustering diagnostics and visualisation methodology described in Rousseeuw (1987) <doi:10.1016/0377-0427(87)90125-7> and selected time series features from Hyndman and Athanasopoulos (2021) ``Forecasting: Principles and Practice" <https://otexts.com/fpp3/>.

License MIT + file LICENSE

Encoding UTF-8

Imports dplyr, tidyr, tidyselect, tibble, tsibble, rlang, WDI, cluster, fabletools, feasts, forcats, ggplot2, ggiraph, ggtext, ggdist, scales, patchwork, ggnewscale

Suggests knitr, rmarkdown, naniar, testthat

Depends R (>= 4.1.0)

RoxygenNote 7.3.2

LazyData true

URL https://github.com/Oluwayomi-Olaitan/wdiexplorer

BugReports https://github.com/Oluwayomi-Olaitan/wdiexplorer/issues

VignetteBuilder knitr

NeedsCompilation no

Author Oluwayomi Akinfenwa [aut, cre],
Niamh Cahill [aut, ths],
Catherine Hurley [aut, ths]

Maintainer Oluwayomi Akinfenwa <oluwayomiakinfenwa@gmail.com>

Repository CRAN

Date/Publication 2026-04-21 20:40:02 UTC

Contents

`add_group_info` *Add grouping information of the WDI data to a metric summary*

Description

Add grouping information of the WDI data to a metric summary

Usage

```
add_group_info(metric_summary, wdi_data)
```

Arguments

`metric_summary`	A data frame containing the calculated diagnostic indices generated by any of the following functions: `compute_variation`, `compute_trend_shape_features`, `compute_temporal_features`, or `compute_diagnostic_indices`
`wdi_data`	A data frame of the indicator data generated by `get_wdi_data`

Value

A data frame containing the calculated diagnostic indices and the grouping variables in the WDI data set.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
```

`compute_diagnostic_indices` *Compute the set of diagnostic indices*

Description

Calculates the collection of diagnostic indices at once

Usage

```
compute_diagnostic_indices(wdi_data, index = NULL, group_var)
```

Arguments

`wdi_data`	A data frame of the indicator data generated by `get_wdi_data`
`index`	An optional character string specifying the indicator code Defaults to `NULL`
`group_var`	A grouping variable in the WDI data set (e.g., "region" or "income")

Value

A data frame with columns `country`, `country_avg_dist`, `within_group_dist`, `sil_width`, `trend_strength`, `linearity`, `curvature`, `smoothness`, `crossing_points`, `flat_spot`, and `acf`.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
```

`compute_dissimilarity` *Compute dissimilarity between pair of countries Calculate pairwise dissimilarities and convert the output to matrix.*

Description

Compute dissimilarity between pair of countries Calculate pairwise dissimilarities and convert the output to matrix.

Usage

```
compute_dissimilarity(wdi_data, index = NULL, metric = "euclidean")
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	An optional character string specifying the indicator code Defaults to NULL
metric	A character string specifying the dissimilarity metric to use Defaults to "euclidean" and uses the daisy() function which handles missing values.

Value

A matrix of pairwise dissimilarities between countries.

Examples

```
pm_diss_mat <- compute_dissimilarity(pm_data)
```

---

compute_temporal_features

Compute sequential temporal features

Description

Calculates number of crossing points, longest flat spot using the feasts package functionality and an additional time series feature - autocorrelation.

Usage

```
compute_temporal_features(wdi_data, index = NULL)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	An optional character string specifying the indicator code Defaults to NULL

Value

A data frame with columns country, crossing_points, flat_spot, and acf.

Examples

```
pm_temporal <- compute_temporal_features(pm_data)
```

compute_trend_shape_features *Compute trend and shape features*

Description

Calculates trend strength, linearity, and curvature using the `feasts` and `fabletools` packages functionality.

Usage

```
compute_trend_shape_features(wdi_data, index = NULL, verbose = TRUE)
```

Arguments

`wdi_data`	A data frame of the indicator data generated by `get_wdi_data`
`index`	An optional character string specifying the indicator code Defaults to `NULL`
`verbose`	Logical, if TRUE, the message about the data download is printed. If FALSE, it is silenced.

Value

A data frame with columns `country`, `trend_strength`, `linearity`, `curvature`, and `smoothness`.

Examples

```
pm_trend_shape <- compute_trend_shape_features(pm_data, verbose = TRUE)
```

compute_variation *Compute variation features*

Description

Calculates average dissimilarities between countries, group-wise country dissimilarities, and silhouette widths.

Usage

```
compute_variation(
  wdi_data,
  index = NULL,
  diss_matrix = compute_dissimilarity(wdi_data, index = index),
  group_var
)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	An optional character string specifying the indicator code Defaults to NULL
diss_matrix	An optional dissimilarity matrix generated by compute_dissimilarity
group_var	A grouping variable in the WDI data set (e.g., "region" or "income")

Value

A data frame with columns country, group, country_avg_dist, within_group_dist, and sil_width.

Examples

```
pm_variation <- compute_variation(pm_data, group_var = "region")
```

---

get_valid_data *Extract valid data from the WDI data Reports countries with no data point, countries with one data point, as well as years for which no data are available.*

Description

Extract valid data from the WDI data Reports countries with no data point, countries with one data point, as well as years for which no data are available.

Usage

```
get_valid_data(wdi_data, index = NULL, verbose = TRUE)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	An optional character string specifying the indicator code Defaults to NULL
verbose	Logical, if TRUE, the message about countries and years will one or no data point is printed. If FALSE, it is silenced. Default to TRUE

Value

A tibble with the valid data for the provided WDI indicator data set and a detailed report of missing entries.

Examples

```
get_valid_data(pm_data, verbose = TRUE)
```

get_wdi_data *Download WDI data using the* WDI *R package*

Description

Create and store the data for the specified indicator code in a folder called wdi_data.

Usage

```
get_wdi_data(indicator, verbose = TRUE)
```

Arguments

indicator	A valid WDI indicator code
verbose	Logical, if TRUE, the message about the data download is printed. If FALSE, it is silenced. Default to TRUE

Value

An .rds file containing the data set for the specified indicator code.

Examples

```
pm_data <- get_wdi_data(indicator = "EN.ATM.PM25.MC.M3", verbose = TRUE)
```

pisa_data *PISA mathematics average scores*

Description

The Programme for International Student Assessment (PISA) is a study conducted by the Organisation for Economic Co-operation and Development (OECD) that evaluates education systems by measuring 15-year-old students' performance in reading, mathematics, and science every three years.

Usage

```
pisa_data
```

Format

A data frame with 15,407 observations with 13 variables

country Country name (character)

iso2c 2-letter ISO country code (character)

iso3c 3-letter ISO country code (character)

year Calendar year representing the time index of the observation (integer)

LO.PISA.MAT Observational values for the specified indicator code (numeric)

status An empty variable meant to indicate the operational status of variables (character)

lastupdated Timestamp that indicates the most recent update of the indicator date (character)

region Geographical region variable (character)

capital Name of the capital city of each country (character)

longitude Geographic coordinate that measures the longitude of the city (character)

latitude Geographic coordinate that measures the latitude of the city (character)

income World Bank income classification variable (character)

lending World Bank income classification variable (character)

Source

World Development Indicator, using the WDI R package

Examples

```
data(pisa_data)

head(pisa_data)
```

---

plot_data_trajectories

Plot of data trajectories

Description

Generates the trajectory of each country data series and supports two plot modes: The display of all series uniformly or a mode that highlight countries with metric values within a specified percentile. Each mode can be rendered in two versions: ungrouped and grouped. Hovering over each highlighted line displays the corresponding country name and metric value

Usage

```
plot_data_trajectories(
  wdi_data,
  index = NULL,
  group_var = NULL,
  metric_summary = NULL,
  metric_var = NULL,
  percentile = 0.95
)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	A character string specifying the indicator code Defaults to NULL
group_var	A grouping variable in the WDI data set (e.g., "region" or "income") Default to NULL If NULL, trajectories are ungrouped and if specified, trajectories are grouped by the levels of the variable
metric_summary	A data frame containing computed diagnostic metrics and the pre-defined grouping information, generated by passing the output of any diagnostic metrics function to add_group_info Defaults to NULL. If NULL, data trajectories are plotted per country series If specified, it highlight countries using a colour palette based on a metric threshold
metric_var	Character string specifying metric variable name in metric_summary to plot
percentile	A percentile threshold (between 0 and 1) for highlighting countries based on their metric values Defaults to 0.95, meaning countries that fall within the top 5% of metric_var values

Value

An ungrouped or grouped interactive plot object displaying the trajectory of country-level data series. It supports both the display of all series uniformly, and also a mode that highlight countries that fall within a specified percentile of any chosen diagnostic metric values.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
plot_data_trajectories(pm_data, group_var = "region",
metric_summary = pm_diagnostic_metrics_group, metric_var = "within_group_avg_dist")
```

plot_metric_distribution *Plot distribution(s) of diagnostic metric(s)*

Description

Generates faceted ggplot displaying the distribution of either selected metric(s) or all the set of diagnostic indices. By default, distribution(s) are ungrouped; if a group_var is specified, distributions are grouped by its levels within each panel. If only one metric is specified in metric_var, a single panel is displayed.

Usage

```
plot_metric_distribution(
  metric_summary,
  colour_var,
  metric_var = NULL,
  group_var = NULL
)
```

Arguments

metric_summary	A data frame containing computed diagnostic metrics and the pre-defined grouping information, generated by passing the output of any diagnostic metrics function to add_group_info
colour_var	A variable in metric_summary data frame whose levels are mapped to distinct colours in the resulting plot
metric_var	Character string or vector of character strings specifying metric variable name(s) in metric_summary to plot If NULL (default), distributions are plotted for all metric variables in metric_summary Is specified, only the distribution for the specified metric(s) will be plotted
group_var	A grouping variable in the WDI data set (e.g., "region" or "income") Default to NULL If NULL, distributions are ungrouped and if specified, distributions are grouped by the levels of the variable

Value

A ggplot object displaying either the ungrouped or grouped distribution of metric(s) in metric_summary. Each metric is displayed in a separate facet panel; if one metric is specified, a single panel is shown.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
plot_metric_distribution(pm_diagnostic_metrics_group, colour_var = "region", group_var = "region")
```

plot_metric_linkview *Plot of diagnostic metrics linked to data trajectories*

Description

Creates an interactive plot linking the scatterplot of two selected metrics with data trajectories. The scatterplot showing the relationship between specified metrics are presented in one panel, and the data trajectories are presented in another panel. Hovering over a point in the scatterplot highlights the corresponding trajectory with the country name, and vice versa.

Usage

```
plot_metric_linkview(
  wdi_data,
  index = NULL,
  metric_summary,
  metric_var,
  group_var = NULL
)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	A character string specifying the indicator code Defaults to NULL
metric_summary	A data frame containing computed diagnostic metrics and the pre-defined grouping information, generated by passing the output of any diagnostic metrics function to add_group_info
metric_var	A vector of character strings specifying metric variable names in metric_summary
group_var	A grouping variable in the WDI data set (e.g., "region" or "income") Default to NULL If NULL, both plots are ungrouped and if specified, they are grouped by the levels of the specified grouping variable

Value

An ungrouped or grouped interactive girafe object displaying the two panels, one with the scatterplot of two specified metrics and the other with the data trajectories.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
plot_metric_linkview(pm_data, metric_summary = pm_diagnostic_metrics,
metric_var = c("linearity", "curvature"))
```

plot_metric_partition *Plot of metric values partitioned by grouping variable*

Description

Generates bars representing the metric value of each country, countries are partitioned by the levels of a specified variable. The partition plot is restricted to group levels containing more than one country, because meaningful comparisons are not possible for single-country levels. The metric value of each country is represented by a coloured bar ordered in descending order, while a lighter-shaded rectangular bar beneath indicates the group-level average for the metric. Countries in each group-level are represented by the same colour.

Usage

```
plot_metric_partition(metric_summary, metric_var, group_var, x_breaks = NULL)
```

Arguments

metric_summary	A data frame containing computed diagnostic metrics and the pre-defined grouping information, generated by passing the output of any diagnostic metrics function to add_group_info
metric_var	Character string specifying metric variable name in metric_summary to plot
group_var	A grouping variable in the WDI data set (e.g., "region" or "income")
x_breaks	Numeric vector specifying the limits and breaks, default to NULL which automatically breaks the x_axis

Value

A ggplot object displaying the metric value of each country by a coloured bar ordered in descending order. A lighter-shaded rectangular bar is displayed beneath the bars indicating their respective group-level average.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
plot_metric_partition(metric_summary = pm_diagnostic_metrics_group,
metric_var = "sil_width", group_var = "region")
```

plot_missing *Missingness plot of the indicator data*

Description

Missingness plot of the indicator data

Usage

```
plot_missing(wdi_data, index = NULL, group_var)
```

Arguments

wdi_data	A data frame of the indicator data generated by get_wdi_data
index	An optional character string specifying the indicator code Defaults to NULL
group_var	A grouping variable in the WDI data set (e.g., "region" or "income")

Value

A plot that provides a structured overview of missing data and shows its distribution over time, across countries, and by the specified grouping variable.

Examples

```
plot_missing(pm_data, group_var = "region")
```

plot_parallel_coords *Plot of diagnostic metrics parallel coordinate plot*

Description

Generates interactive parallel coordinate plots of all diagnostic indices. Hovering over a line across x-axis displays the country name, corresponding metric and its metric value.

Usage

```
plot_parallel_coords(diagnostic_summary, colour_var, group_var = NULL)
```

Arguments

diagnostic_summary	A data frame containing the computed set of diagnostic indices generated by compute_diagnostic_indices
colour_var	A variable in metric_summary data frame whose levels are mapped to distinct colours in the resulting plot
group_var	A grouping variable in the WDI data set (e.g., "region" or "income") Default to NULL If NULL, parallel coordinates are ungrouped and if specified, parallel coordinates are grouped by the levels of the specified grouping variable

Value

An ungrouped or grouped interactive parallel coordinate plot of all diagnostic metrics, with each metric represented as a vertical axis. Each country is shown as an interactive line that intersects all axes, with the position along the x-axis corresponding to the diagnostic indices.

Examples

```
pm_diagnostic_metrics <- compute_diagnostic_indices(pm_data, group_var = "region")
pm_diagnostic_metrics_group <- add_group_info(metric_summary = pm_diagnostic_metrics,pm_data)
plot_parallel_coords(pm_diagnostic_metrics_group, colour_var = "region", group_var = "region")
```

---

pm_data *PM2.5 air pollution data*

Description

This data set contains the mean annual exposure levels to ambient PM2.5 air pollution across various countries, measured in micrograms per cubic meter.

Usage

```
pm_data
```

Format

A data frame with 13,910 observations with 13 variables

country Country name (character)

iso2c 2-letter ISO country code (character)

iso3c 3-letter ISO country code (character)

year Calendar year representing the time index of the observation (integer)

EN.ATM.PM25.MC.M3 Observational values for the specified indicator code (numeric)

status An empty variable meant to indicate the operational status of variables (character)

lastupdated Timestamp that indicates the most recent update of the indicator date (character)

region Geographical region variable (character)

capital Name of the capital city of each country (character)

longitude Geographic coordinate that measures the longitude of the city (character)

latitude Geographic coordinate that measures the latitude of the city (character)

income World Bank income classification variable (character)

lending World Bank income classification variable (character)

Source

World Development Indicator, using the WDI R package

Examples

```
data(pm_data)

head(pm_data)
```

Index