Package ‘wcde’

October 12, 2022

Title  Download Data from the Wittgenstein Centre Human Capital Data Explorer

Version  0.0.5

URL  https://guyabel.github.io/wcde/

BugReports  https://github.com/guyabel/wcde/issues/

Description  Download and plot education specific demographic data from the Wittgenstein Centre for Demography and Human Capital Data Explorer <http://dataexplorer.wittgensteincentre.org/>.

License  GPL-3

Encoding  UTF-8

LazyData  true

RoxygenNote  7.1.2

Imports  dplyr, magrittr, tidyr, progress, countrycode, tibble, purrr, stringr, readr,forcats

Depends  R (>= 2.10)

Suggests  spelling, knitr, rmarkdown, tidyverse, lemon

VignetteBuilder  knitr

Language  en-US

NeedsCompilation  no

Author  Guy J. Abel [aut, cre, ctb] (<https://orcid.org/0000-0002-4893-5687>),
        Samir K.C. [ctb] (<https://orcid.org/0000-0002-5213-9181>),
        Michaela Potancokova [ctb],
        Claudia Reiter [ctb] (<https://orcid.org/0000-0002-1485-3851>),
        Andrea Tamburini [ctb],
        Dilek Yildiz [ctb] (<https://orcid.org/0000-0001-6192-0634>)

Maintainer  Guy J. Abel <g.j.abel@gmail.com>

Repository  CRAN

Date/Publication  2022-06-06 17:30:02 UTC
**edu_group_sum**

**R topics documented:**

- `edu_group_sum` .................................................. 2
- `every_other` ..................................................... 3
- `find_indicator` .................................................. 4
- `get_wcde` .......................................................... 4
- `get_wcde_single` .................................................. 7
- `past_epop` ........................................................... 7
- `wic_col4` ........................................................... 8
- `wic_col6` ............................................................ 8
- `wic_col8` ............................................................ 9
- `wic_indicators` ...................................................... 9
- `wic_locations` ..................................................... 10
- `wic_scenarios` .................................................... 11

**Index** 12

---

**edu_group_sum** | **Education group sums**

---

**Description**

Cleans epop data, downloaded using the `wcde()` function, for summations of population by 4, 6 or 8 education groups.

**Usage**

```r
edu_group_sum(d = NULL, n = 4, strip_totals = TRUE, factor_convert = TRUE)
```

**Arguments**

- `d` Data frame downloaded from the
- `n` Number of education groups (from 4, 6 or 8)
- `strip_totals` Remove total sums in epop column. Will not strip education totals if year < 2015 and n = 8 as past data on population size by 8 education groups is unavailable.
- `factor_convert` Convert columns that are character strings to factors, with levels based on order of appearance.

**Details**

Strips the epop data set to relevant rows for the n education groups.

**Value**

A tibble with the data selected.
every_other

Examples

library(tidyverse)
past_epop %>%
  filter(year == 2020) %>%
  edu_group_sum()

every_other(x = letters)
every_other(LETTERS, n = 3, start = 6)
every_other(x = letters, fill = "")

description

Select every other (nth) element from a vector

Usage

every_other(x, n = 2, start = 1, fill = NULL)

Arguments

x Vector to select (remove) elements from
n Numeric value for the number of elements to skip. Default is 2, i.e. skips every second element
start Numeric value to indicate which element of the vector to commence from.
fill Character string to be used in place of skipped element. By default is NULL and hence skipped elements are removed rather than replaced.

Value

Vector with elements removed

Examples

every_other(x = letters)
every_other(LETTERS, n = 3, start = 6)
every_other(x = letters, fill = "")
find_indicator  
*Find available indicator code names in the Wittgenstein Centre Human Capital Data Explorer*

**Description**
Find available indicator code names in the Wittgenstein Centre Human Capital Data Explorer

**Usage**
`find_indicator(x)`

**Arguments**
- **x** Character string on key word or name related to indicator of potential interest.

**Value**
A subset of the `wic_indicators` data frame with one or more of the `indicator`, `description` or `definition` columns matching the keyword given to `x`. Use the result in the `indicator` column to input to the get_wcde function for downloading data.

**Examples**
```r
find_indicator("education")
find_indicator("migr")
find_indicator("fert")
```

get_wcde  
*Download data from the Wittgenstein Centre Human Capital Data Explorer*

**Description**
Downloads data from the Wittgenstein Centre Human Capital Data Explorer. Requires a working internet connection.

**Usage**
```r
get_wcde(
  indicator = "pop",
  scenario = 2,
  country_code = NULL,
  country_name = NULL,
  pop_age = c("total", "all"),
  pop_sex = c("total", "both", "all"),
)```
get_wcde

pop_edu = c("total", "four", "six", "eight"),
include_scenario_names = FALSE
)

Arguments

indicator One character string based on the indicator column in the wic_indicators data frame, representing the variable to be downloaded.
scenario Vector of length one or more with numbers corresponding the scenarios. See details for more information. Defaults to 2 for the SSP2 Medium scenario.
country_code Vector of length one or more of country numeric codes based on ISO 3 digit numeric values.
country_name Vector of length one or more of country names. The corresponding country code will be guessed using the countrycodes package.
pop_age Character string for population age groups if indicator is set to pop. Defaults to no age groups total, but can be set to all.
pop_sex Character string for population sexes if indicator is set to pop. Defaults to no sex total, but can be set to both or all.
pop_edu Character string for population educational attainment if indicator is set to pop. Defaults to total, but can be set to four, six or eight.
include_scenario_names Logical vector of length one to indicate if to include additional columns for scenario names and short names. FALSE by default.

Details

If not country_name or country_code is provided data for all countries and regions are downloaded. A full list of available countries and regions can be found in the wic_locations data frame.

indicator must be set to a value in the first column in the table below of available demographic indicators:

<table>
<thead>
<tr>
<th>indicator</th>
<th>Indicator Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pop</td>
<td>Population Size (000’s)</td>
</tr>
<tr>
<td>bpop</td>
<td>Population Size by Broad Age (000’s)</td>
</tr>
<tr>
<td>epop</td>
<td>Population Size by Education (000’s)</td>
</tr>
<tr>
<td>prop</td>
<td>Educational Attainment Distribution</td>
</tr>
<tr>
<td>bprop</td>
<td>Educational Attainment Distribution by Broad Age</td>
</tr>
<tr>
<td>growth</td>
<td>Average Annual Growth Rate</td>
</tr>
<tr>
<td>nirate</td>
<td>Average Annual Rate of Natural Increase</td>
</tr>
<tr>
<td>sexratio</td>
<td>Sex Ratio</td>
</tr>
<tr>
<td>mage</td>
<td>Population Median Age</td>
</tr>
<tr>
<td>tdr</td>
<td>Total Dependency Ratio</td>
</tr>
<tr>
<td>ydr</td>
<td>Youth Dependency Ratio</td>
</tr>
<tr>
<td>odr</td>
<td>Old-age Dependency Ratio</td>
</tr>
<tr>
<td>ryl15</td>
<td>Age When Remaining Life Expectancy is Below 15 years</td>
</tr>
<tr>
<td>pryl15</td>
<td>Proportion of Population with a Remaining Life Expectancy below 15 Years</td>
</tr>
</tbody>
</table>
get_wcde

mys  Mean Years of Schooling by Age
bmys Mean Years of Schooling by Broad Age
ggapmys15  Gender Gap in Mean Years Schooling (15+)
ggapmys25  Gender Gap in Mean Years Schooling (25+)
ggapedu15  Gender Gap in Educational Attainment (15+)
ggapedu25  Gender Gap in Educational Attainment (25+)
tfr  Total Fertility Rate
etfr Total Fertility Rate by Education
asfr Age-Specific Fertility Rate
easfr Age-Specific Fertility Rate by Education
cbr  Crude Birth Rate
macb Mean Age at Childbearing
e0 Life Expectancy at Birth
cdr  Crude Death Rate
assr Age-Specific Survival Ratio
eassr Age-Specific Survival Ratio by Education
net  Net Migration

See wic_indicators data frame for more details.

scenario must be set to one or values in the first column table below of the available future scenarios:

<table>
<thead>
<tr>
<th>scenario</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rapid Development (SSP1)</td>
</tr>
<tr>
<td>2</td>
<td>Medium (SSP2)</td>
</tr>
<tr>
<td>3</td>
<td>Stalled Development (SSP3)</td>
</tr>
<tr>
<td>21</td>
<td>Medium - Zero Migration (SSP2 - ZM)</td>
</tr>
<tr>
<td>22</td>
<td>Medium - Double Migration (SSP2 - DM)</td>
</tr>
</tbody>
</table>

See wic_scenarios data frame for more details.

Value

A tibble with the data selected.

Examples

# SSP2 tfr for Austria and Bulgaria
get_wcde(indicator = "tfr", country_code = c(40, 100))

# SSP1 and SSP2 life expectancy for Vietnam and United Kingdom (guessing the country codes)
get_wcde(scenario = c(1, 2), indicator = "e0", country_name = c("Vietnam", "UK"))

# SSP1 and SSP3 population by education for all countries
get_wcde(scenario = c(1, 3), indicator = "tfr")
# population totals (aggregated over age, sex and education)
get_wcde(indicator = "pop", country_name = "Austria")

# population totals by education group
get_wcde(indicator = "pop", country_name = "Austria", pop_edu = "four")

# population totals by age-sex group
get_wcde(indicator = "pop", country_name = "Austria", pop_age = "all", pop_sex = "both")

---

**get_wcde_single**

Pull multiple vectors for a given indicator, scenarios and .Rdata file names

---

**Description**

Requires a working internet connection. Intended for internal use.

**Usage**

```r
get_wcde_single(indicator = NULL, scenario = 2, country_code = NULL)
```

**Arguments**

- **indicator**: One character string based on the name column in the wic_indicators data frame, representing the variable to be interested.
- **scenario**: Vector with a numbers corresponding the scenarios. See details in wcde for more information.
- **country_code**: Vector of length one or more of country numeric codes based on ISO 3 digit numeric values.

**Value**

A tibble with multiple columns.

---

**past_epop**

Past population sizes for all countries by education

---

**Description**

A data set containing population sizes for all countries by education between 1950 and 2020

**Usage**

```r
past_epop
```
Format

A data frame with 840,126 rows and 7 variables, including:

- **name**  Area name
- **country_code**  ISO 3 digit country code
- **year**  Year of observation from 1950 to 2020 in five-year steps
- **age**  Five-year age groups
- **education**  Education group
- **sex**  Sex
- **epop**  Population size in thousands for each age, sex and education group

Source

http://dataexplorer.wittgensteincentre.org/wcde-v2/

---

wic_col4  
*Colours used in Wittgenstein Centre for Demography and Human Capital Data Explorer*

Description

Three sets of colours used for filling education based plots based on the different availability of detailed education categories (four, six or eight groups)

Usage

wic_col4

Format

A named vector

---

wic_col6  
*Colours used in Wittgenstein Centre for Demography and Human Capital Data Explorer*

Description

Three sets of colours used for filling education based plots based on the different availability of detailed education categories (four, six or eight groups)

Usage

wic_col6
**wic_col8**

**Format**

A named vector

**Description**

Three sets of colours used for filling education based plots based on the different availability of detailed education categories (four, six or eight groups)

**Usage**

wic_col8

**Format**

A named vector

**wic_indicators**

**Indicators used in the Wittgenstein Centre Human Capital Data Explorer**

**Description**

A data set containing the indicator codes, names and further details used in the Wittgenstein Centre Human Capital Data Explorer

**Usage**

wic_indicators

**Format**

A data frame with 31 rows and 8 variables, including:

- **indicator** Short name of indicator to be used in the `indicator` argument of the `get_wcde()` function
- **description** Brief description of indicator
- **age** Availability of indicator by five-year age groups
- **bage** Availability of indicator by broad age groups
- **sage** Availability of indicator with a new born age group
- **sex** Availability of indicator by sex
edua Availability of indicator by education
period Indicator is a period (flow)
past Availability of past data for indicator
definition Full definition for indicator

Source

http://dataexplorer.wittgensteincentre.org/wcde-v2/

| wic_locations | Locations used in the Wittgenstein Centre Human Capital Data Explorer |

Description

A dataset containing the location codes, names and further details used in the Wittgenstein Centre Human Capital Data Explorer

Usage

wic_locations

Format

A data frame with 230 rows and 7 variables, including:

name Area name
isono ISO 3 digit country code
continent Continent of country
region UN region of country
dim Category or country/region/area

Source

http://dataexplorer.wittgensteincentre.org/wcde-v2/
### Description

A data set containing the scenario codes, names short names used in the Wittgenstein Centre Human Capital Data Explorer

### Usage

wic_scenarios

### Format

A data frame with 5 rows and 3 variables, including:

- **scenario_name**  Full scenario name
- **scenario**  Code to match help file of get_wcde function
- **scenario_abb**  Short scenario name

### Source

http://dataexplorer.wittgensteincentre.org/wcde-v2/
Index

* datasets
  past_epop, 7
  wic_col4, 8
  wic_col6, 8
  wic_col8, 9
  wic_indicators, 9
  wic_locations, 10
  wic_scenarios, 11

edu_group_sum, 2
every_other, 3
find_indicator, 4
get_wcde, 4
get_wcde_single, 7
past_epop, 7
tibble, 6
wic_col4, 8
wic_col6, 8
wic_col8, 9
wic_indicators, 9
wic_locations, 10
wic_scenarios, 11