

Package ‘wcde’

July 28, 2021

Title Download Data from the Wittgenstein Human Capital Data Explorer

Version 0.0.2

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

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

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

Depends R (>= 2.10)

Suggests spelling, knitr, rmarkdown

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 2021-07-28 06:40:02 UTC

R topics documented:

edu_group_sum	2
every_other	3
find_indicator	3

get_wcde	4
past_epop	6
wcde_pull	7
wic_col4	7
wic_col6	8
wic_col8	8
wic_indicators	8
wic_locations	9
wic_scenarios	10

Index 11

edu_group_sum	<i>Education Group Sums</i>
---------------	-----------------------------

Description

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

Usage

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

Arguments

<code>d</code>	Data frame downloaded from the
<code>n</code>	Number of education groups (from 4, 6 or 8)
<code>strip_totals</code>	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.
<code>factor_convert</code>	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.

Examples

```
library(tidyverse)
past_epop %>%
  filter(year == 2020) %>%
  mutate(scenario = 2) %>%
  edu_group_sum()
```

every_other	<i>Select every other (nth) element from a vector</i>
-------------	---

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	<i>Find available indicator code names in the Wittgenstein Centre Human Capital Data Explorer</i>
----------------	---

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

```
find_indicator("education")
find_indicator("migr")
find_indicator("fert")
```

<code>get_wcde</code>	<i>Download data from the Wittgenstein Centre Human Capital Data Explorer Data Explorer</i>
-----------------------	---

Description

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

Usage

```
get_wcde(
  indicator = "pop",
  scenario = 2,
  country_code = NULL,
  country_name = NULL,
  include_scenario_names = FALSE
)
```

Arguments

<code>indicator</code>	One character string based on the <code>indicator</code> column in the <code>wic_indicators</code> data frame, representing the variable to be downloaded.
<code>scenario</code>	Vector of length one or more with numbers corresponding the scenarios. See details for more information. Defaults to 2 for the SSP2 Medium scenario.
<code>country_code</code>	Vector of length one or more of country numeric codes based on ISO 3 digit numeric values.
<code>country_name</code>	Vector of length one or more of country names. The corresponding country code will be guessed using the <code>countrycodes</code> package.
<code>include_scenario_names</code>	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:

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

scenario	description
1	Rapid Development (SSP1)
2	Medium (SSP2)

3	Stalled Development (SSP3)
21	Medium - Zero Migration (SSP2 - ZM)
22	Medium - Double Migration (SSP2 - DM)

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")
```

past_epop	<i>Past population sizes for all countries by education</i>
-----------	---

Description

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

Usage

```
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/>

wcde_pull	<i>Pull multiple vectors for a given indicator, scenarios and .Rdata file names</i>
-----------	---

Description

Requires a working internet connection. Intended for internal use.

Usage

```
wcde_pull(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.

wic_col4	<i>Colours used in Wittgenstein Human Capital Data Explorer</i>
----------	---

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 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`**Format**

A named vector

`wic_col8`*Colours used in Wittgenstein 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_col8`**Format**

A named vector

`wic_indicators`*Indicators used in the Wittgenstein Data Explorer*

Description

A data set containing the indicator codes, names and further details used in the Wittgenstein 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

edu 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 Data Explorer

Description

A dataset containing the location codes, names and further details used in the Wittgenstein 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/>

wic_scenarios	<i>Scenarios used in the Wittgenstein Data Explorer</i>
---------------	---

Description

A data set containing the scenario codes, names short names used in the Wittgenstein 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, 6

wic_col4, 7

wic_col6, 8

wic_col8, 8

wic_indicators, 8

wic_locations, 9

wic_scenarios, 10

edu_group_sum, 2

every_other, 3

find_indicator, 3

get_wcde, 4

past_epop, 6

tibble, 6

wcde_pull, 7

wic_col4, 7

wic_col6, 8

wic_col8, 8

wic_indicators, 8

wic_locations, 9

wic_scenarios, 10