

Package ‘edbuildr’

June 2, 2021

Type Package

Title Automated School District Data Download and Processing

Version 0.3.0

Description Import the 'EdBuild' master dataset of school district finance, student demographics, and community economic indicators for every school district in the United States.

The master dataset is built from the US Census, Annual Survey of School System Finances (F33) and joins data from the National Center for Education Statistics, Common Core of Data; the US Census, Small Area Income and Poverty Estimates; and the US Census, Education Demographic and Geographic Estimates. We apply 'EdBuild' standard processing to the dataset and provide the option to select from four different exclusion criteria - see the masterpull() help file for more details.

The master dataset is available for any school year from 2013 to 2019 or longitudinally for all years 2013-2019.

School year is identified by the end year. For example, the 2018-19 school year is 2019.

Additional functions in the package use 'EdBuild' master data to analyze the difference between neighboring school districts and create formatted excel tables of school district data. For full details about

'EdBuild' data processing please see 'EdBuild' (2020) <<http://data.edbuild.org>>.

License CC0

URL <https://github.com/EdBuild/edbuildr>,
<http://viz.edbuild.org/workshops/edbuildr/>,
<http://viz.edbuild.org/workshops/data-overview/>

BugReports <https://github.com/EdBuild/edbuildr/issues>

Imports dplyr, magrittr, openxlsx, scales, stringr, tidysselect (>= 1.0.0)

Encoding UTF-8

RoxygenNote 7.0.2

Suggests knitr, rmarkdown

NeedsCompilation no

Author Megan Brodzik [cre, aut],
 Cecilia Depman [aut],
 Sara Hodges [aut],
 Kailey Spencer [ctb]

Maintainer Megan Brodzik <meganbrodzik@gmail.com>

Repository CRAN

Date/Publication 2021-06-02 13:40:02 UTC

R topics documented:

big_borders	2
edbuildr	3
f33pull	4
f33pull_raw	6
long_masterpull	7
masterpull	7
master_codebook	9
neigh_diff	10
round2	11
sd_neighbor	12
sd_neighbor_xlsx	13
sd_table_xlsx	14
Index	15

big_borders	<i>A function to model local revenue pooling</i>
-------------	--

Description

This function allows you to pool local revenue for school districts at either the county or state level to see the effects on individual districts.

Usage

```
big_borders(data_year= "2019", pooling_level = "county")
```

Arguments

data_year	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
pooling_level	Character string indicating the level of local revenue pooling. Either county or state. Defaults to county.

Value

A dataframe with 17 variables where each observation is a school district.

See Also

[masterpull](#), [master_codebook](#)

Examples

```
state_pool <- big_borders(  
  data_year = "2019",  
  pooling_level = "state"  
)
```

edbuildr

edbuildr: A package for automated downloading and processing of school district data

Description

This package allows users to import EdBuild's master dataset of school district finance, student demographics, and community economic indicators for every school district in the United States. The master dataset is built from the US Census, Annual Survey of School System Finances (F33) and joins data from the National Center for Education Statistics, Common Core of Data (CCD); the US Census, Small Area Income and Poverty Estimates (SAIPE); and the US Census, Education Demographic and Geographic Estimates (EDGE). We apply EdBuild's standard processing to the dataset and provide the option to select from four different exclusion criteria - see the masterpull help file for full details. The master dataset is available for any school year from 2013 to 2019 or longitudinally for all years 2013-2019. School year is identified by the end year. For example, the 2018-19 school year is 2019. Additional functions in the package use EdBuild's master data to analyze the difference between neighboring school districts and create formatted excel tables of school district data. For full details about EdBuild's data processing please see: EdBuild (2020) <<http://data.edbuild.org/>>.

edbuildr functions

The edbuildr functions are:

big_borders Pools local revenue for school districts at either the county or state level to see the effects on individual districts.

f33pull Pulls in the US Census's Annual Survey of School System Finances (F33) and processes according to EdBuild's adjustments. Available for 2006-2019.

f33pull_raw Pulls in the the full US Census's Annual Survey of School System Finances (F33) with no EdBuild adjustments. Available for 1990-2019.

long_masterpull Pulls in EdBuild's master dataset longitudinally for years 2013-2019.

master_codebook Reads in a codebook for EdBuild's master data. The codebook tells the user what each variable name represents and the source of each variable.

masterpull Pulls in EdBuild's master dataset which is a compilation of national level school district data from the Census's Annual Survey of School System Finances (F33) and Small Area Income and Poverty Estimates (SAIPE) and NCES's Common Core of Data (CCD) and Education Demographic and Geographic Estimates (EDGE). Cost adjustments were calculated using C2ER. Available for 2013-2019.

neigh_diff Calculates the difference and national rank for a selected variable between all school district neighbors.

round2 Rounds 0.5 up.

sd_neighbor Creates a dataframe of any school district and their neighbors with selected variables.

sd_neighbor_xlsx Creates a formatted table (.xlsx) of any school district and their neighbors with selected variables.

sd_table_xlsx Creates a formatted excel table (.xlsx) of school districts in a state or county with selected variables.

Author(s)

- Megan Brodzik (meganbrodzik@gmail.com), maintainer
- Cecilia Depman (cecilia.depman@gmail.com), author
- Sara Hodges (sara.hodges@gmail.com), author
- Kailey Spencer (kailey.spencer.depoe@gmail.com), contributor

f33pull

A function to import and process F33 data

Description

This function imports the US Census, Annual Survey of School System Finances (F33) data with standard EdBuild processing steps applied to revenues.

Usage

```
f33pull(data_year = "2019", path=NULL, additional_var=NULL, keep_calcs=FALSE)
```

Arguments

data_year	Four digit year of F33 data you would like to pull. Available for 2006-2019.
path	Path name of F33 file if data is stored on local computer. Defaults to NULL to pull in F33 data from EdBuild's server.
additional_var	List any additional F33 variables you would like to pull.
keep_calcs	Do you want to keep all of the variables used to calculate adjusted revenues? Defaults to FALSE.

Format

A data frame with 14 variables:

year data year

State State name

STATE_FIPS State code

NCEISD NCES school district ID

NAME School district name

CONUM County number

ENROLL School district enrollment

TFEDREV Total federal revenue- no EdBuild adjustments

LOCREV_adj Local revenue - with EdBuild adjustments

LOCREV_adj_PP Local revenue per pupil- with EdBuild adjustments

STREV_adj State revenue - with EdBuild adjustments

STREV_adj_PP State revenue per pupil - with EdBuild adjustments

STLOCREV_adj State and local revenue - with EdBuild adjustments

STLOCREV_adj_PP State and local revenue per pupil - with EdBuild adjustments

Value

A dataframe where each observation is a school district.

Note

The following processing was made to state and local revenues for each school district:

1. Revenues are multiplied by 1,000.
2. Sale of property (U11) is removed from local revenue.
3. Capital outlay and debt service programs (C11) is removed from state revenue.
4. In Arkansas, the Census state NCES local revenue (C24) is subtracted from state revenues and added to local revenues for all districts in the state.
5. In Texas, recapture, reported as payments to state governments (L12), is subtracted from local revenue.
6. Charter passthrough dollars, V92, is subtracted proportionately from state and local revenues since students attending charter schools are not included in F33 enrollment.

For full details about F33 processing please visit [EdBuild's dividing lines map site](#).

Source

<https://s3.amazonaws.com/data.edbuild.org/public/Raw+Data/F33/csv/2019.csv>

See Also

[f33pull_raw](#)

Examples

```
f33_2019 <- f33pull(data_year="2019",  
                  additional_var=c("V40", "TCAPOUT"),  
                  keep_calcs=FALSE)
```

f33pull_raw

A function to pull raw F33 data

Description

This function imports raw data from the US Census, Annual Survey of School System Finances (F33).

Usage

```
f33pull_raw(data_year = "2019")
```

Arguments

`data_year` Four digit year of F33 data you would like to pull. Available for 1990-2019.

Format

A data frame with 40 variables. Definitions of each variable name can be found in the [US Census's Annual Survey of School System Finances](#) technical documentation.

Value

A dataframe where each observation is a school district.

Source

<https://s3.amazonaws.com/data.edbuild.org/public/Raw+Data/F33/csv/2019.csv>

See Also

[f33pull](#)

Examples

```
f33_2019 <- f33pull_raw('2019')
```

long_masterpull	<i>A function to import a longitudinal version of EdBuild's master dataset</i>
-----------------	--

Description

This function imports a longitudinal dataset of EdBuild's master data for the years 2013- 2019. The master dataset is a compilation of national level school district data from the US Census Annual Survey of School System Finances (F33); US Census Small Area Income and Poverty Estimates (SAIPE); National Center for Education Statistics (NCES) Common Core of Data (CCD); and Education Demographic and Geographic Estimates (EDGE). Cost adjustments were calculated using C2ER.

Usage

```
long_masterpull()
```

Format

A data frame with 130,419 observations and 44 variables. To view descriptions of variable names and sources for each use `master_codebook()`

Value

A dataframe where each observation is a school district.

See Also

[master_codebook](#), [masterpull](#)

Examples

```
long_master <- long_masterpull()
```

masterpull	<i>A function to import EdBuild's master dataset</i>
------------	--

Description

This function allows you to read in EdBuild's master dataset for the years 2013- 2019. The master dataset is a compilation of national level school district data from the US Census Annual Survey of School System Finances (F33); US Census Small Area Income and Poverty Estimates (SAIPE); National Center for Education Statistics (NCES) Common Core of Data (CCD); and Education Demographic and Geographic Estimates (EDGE). Cost adjustments were calculated using C2ER.

Usage

```
masterpull(data_year = "2019", data_type = "gen", disaggregated =
  FALSE)
```

Arguments

data_year	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
data_type	Type of master data to pull in <ul style="list-style-type: none"> • geo pulls in all school districts that have physical school district boundaries. To be used for map-based analysis and other analyses that pertain to school districts with geographic boundaries. For instance, an analysis using median owner-occupied property value would use the geography exclusion. • fin pulls in all school districts that meet EdBuild's criteria for financial analysis. To be used for finance analysis. • gen pulls in all school districts that meet enrollment and district type requirements. To be used for all other, non-finance analysis. • full pulls in all school districts in the given year. To be used with great care, or to find a district that does not appear in any other exclusion, for example charter school districts. <p>Defaults to gen</p>
disaggregated	For the full, general, and finance exclusions in 2019, users have the option to view disaggregated Vermont school district data without EdBuild's aggregation processing.

Format

A data frame with 42 variables. To view descriptions of variable names and sources for each use `master_codebook()`

Value

A dataframe where each observation is a school district.

Note

There are three types of exclusions that can be applied to the master dataset:

1. Geography based:

- Excludes districts that do not have physical school district boundaries and thus are not included in the US Census, EDGE shapefile.

2. Finance based:

- Excludes districts that are of types 5 (vocational or special education), 6 (non-operating) or 7 (educational service agency) in the F33 data.

- If F33 school type is missing, excludes districts that are of types 4 (regional education service agency), 5 (state agency), 6 (federal agency), 7 (charter agency) or 8 (other education agency) based on CCD.
- Excludes districts that zero missing or total enrollments.
- Excludes districts that have missing or zero operational schools.
- Excludes districts that have missing revenues.
- Excludes districts that have very low revenues (<\$500).
- Excludes districts that have very high revenues (>\$100,000 in inflation-adjusted 2017 dollars).
- Excludes districts from the US territories.

3. General:

- Excludes districts that are of types 5 (vocational or special education), 6 (nonoperating) or 7 (educational service agency) in the F33 data.
- If F33 school type is missing, excludes districts that are of types 4 (regional education service agency), 5 (state agency), 6 (federal agency), 7 (charter agency) or 8 (other education agency) based on CCD.
- Excludes districts that have missing or zero total enrollments.
- Excludes districts from the US territories.

Source

https://s3.amazonaws.com/data.edbuild.org/public/Processed+Data/Master/2019/full_data_19_type_exc.csv

See Also

[master_codebook](#), [long_masterpull](#)

Examples

```
master19_geo <- masterpull("2019", data_type = "geo")
```

master_codebook

A function to view EdBuild's data sources and variable names

Description

This function imports a codebook for EdBuild's master data. The codebook tells the user what each variable name represents and the source of each variable.

Usage

```
master_codebook()
```

Value

A dataframe where each observation is a variable from the master dataset.

See Also

[long_masterpull](#), [masterpull](#)

Examples

```
codebook <- master_codebook()
```

<code>neigh_diff</code>	<i>A function to find the difference between school district neighbors</i>
-------------------------	--

Description

This function allows you to find the difference between each pair of school district neighbors and calculate the national rank from largest to smallest.

Usage

```
neigh_diff(data_year= "2019",
  diff_var="Percentage Point Difference in Poverty Rate", type= "like")
```

Arguments

<code>data_year</code>	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
<code>diff_var</code>	Character string on which to rank the difference between school district neighbors. Use <code>diff_var = "options"</code> to print a list of the variables. Defaults to Percentage Point Difference in Poverty Rate.
<code>type</code>	Character string to indicate which types of neighbors to return. Defaults to "like" which returns a list of neighbors that are the same district type (that is, unified to unified, elementary to elementary and secondary to secondary). To view all neighbors use "all". This selection becomes important for districts like Chicago which have upwards of 50 neighboring school districts, but only 1 type-like neighbor. Chicago is a unified district and it has 1 neighbor that is also unified, 16 neighbors that are secondary districts, and 32 neighbors that are elementary districts.

Value

A dataframe where each observation is a pair of neighboring school districts.

See Also

[masterpull](#), [master_codebook](#), [sd_neighbor_xlsx](#)

Examples

```
tr_diff <- neigh_diff(  
  data_year = "2019",  
  diff_var = "Difference in Total Revenue Per Pupil"  
)
```

round2	<i>A function to round 0.5 down</i>
--------	-------------------------------------

Description

This function rounds values based on mathematical rules. That is, numbers ending below 0.5 round down and numbers ending with 0.5 and greater round up.

Usage

```
round2(x, n)
```

Arguments

x	a numeric vector
n	number of decimal places

Value

The value returned is a numeric vector, x, rounded to n decimals places.

References

<https://stackoverflow.com/questions/12688717/round-up-from-5>

Examples

```
round2(143.05, 1)  
round2(143.048, 2)
```

sd_neighbor	<i>A function to create a dataframe of a school district and all its neighbors</i>
-------------	--

Description

This function allows you to create a table to view any school district and its neighbors with selected data from EdBuild's master datafile.

Usage

```
sd_neighbor(data_year = "2019", school_district = NULL, table_vars =
  c('Name', 'Enrollment', 'Poverty Rate', 'Percent Nonwhite', 'Local Revenue PP',
    'State Revenue PP', 'Type'))
```

Arguments

data_year	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
school_district	Seven digit NCESID of the school district. Default is NULL. To find the NCESID for any school district, use masterpull to search for your district.
table_vars	Variable or list of variables to include in the table. Use tables_vars = "options" to print a list of the variables. Defaults to: Name; Enrollment; Poverty Rate; Percent Nonwhite; Local Revenue, per Pupil; State Revenue, per Pupil; Type

Value

A dataframe where each observation is a school district.

See Also

[sd_neighbor_xlsx](#), [sd_table_xlsx](#), [master_codebook](#), [masterpull](#) [neigh_diff](#)

Examples

```
table <- sd_neighbor(
  data_year = "2019",
  school_district = "0622710",
  table_vars = c("Name",
    "Percent Nonwhite",
    "Median Household Income",
    "State Revenue PP")
)
```

sd_neighbor_xlsx *A function to create a table of a school district and all its neighbors*

Description

This function allows you to write out a table of any school district and its neighbors with selected data from EdBuild's master datafile, ready to export as a formatted excel file.

Usage

```
sd_neighbor_xlsx(data_year = "2019", school_district = NULL,
  table_vars = c('Name', 'Enrollment', 'Poverty Rate', 'Percent Nonwhite',
    'Local Revenue PP', 'State Revenue PP', 'Type'))
```

Arguments

data_year	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
school_district	Seven digit NCESID of the school district. Default is NULL. To find the NCESID for any school district, use masterpull to search for your district.
table_vars	Variable or list of variables to include in the table. Use tables_vars = "options" to print a list of the variables. Defaults to: Name; Enrollment; Poverty Rate; Percent Nonwhite; Local Revenue, per Pupil; State Revenue, per Pupil; Type

Value

An excel workbook which can be written out with `openxlsx::saveWorkbook(my_table, file = '~/Documents/neighbor_table.xlsx', overwrite = TRUE)`

See Also

[sd_table_xlsx](#), [master_codebook](#), [masterpull](#) [neigh_diff](#)

Examples

```
table <- sd_neighbor_xlsx(
  data_year = "2019",
  school_district = "0622710",
  table_vars = c("Name",
    "Percent Nonwhite",
    "Median Household Income",
    "State Revenue PP")
)
```

sd_table_xlsx	<i>A function to create a table of school districts in a selected state or county</i>
---------------	---

Description

This function allows you to write out a table of any school districts with selected data as a formatted excel file.

Usage

```
sd_table_xlsx(data_year = '2019', state = "New Jersey", county = NULL,
  table_vars = c('Name', 'Enrollment', 'Poverty Rate', 'Percent Nonwhite',
    'Local Revenue PP', 'State Revenue PP', 'Total Revenue PP'))
```

Arguments

data_year	Four digit year of master data to pull in. Options include 2013- 2019. Defaults to 2019.
state	State name. Defaults to New Jersey.
county	County name. Defaults to NULL.
table_vars	Variable or list of variables to include in the table. Use tables_vars = "options" to print a list of the variables. Defaults to: Name; Enrollment; Poverty Rate; Percent Nonwhite; Local Revenue, Per Pupil; State Revenue, Per Pupil; Total Revenue, Per Pupil.

Value

An excel workbook which can be written out with `openxlsx::saveWorkbook(my_table, file = '~/Documents/state_year.xlsx', overwrite = TRUE)`

See Also

[sd_neighbor_xlsx](#), [master_codebook](#), [masterpull](#)

Examples

```
my_table <- sd_table_xlsx(data_year = "2019",
  state = "Maryland",
  county = c("Baltimore County", "Baltimore City", "Howard County", "Carroll County"),
  table_vars = c("Name", "Poverty Rate")
)
```

Index

- * **CCD**
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **EDGE**
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **EdBuild**
 - big_borders, 2
 - f33pull, 4
 - f33pull_raw, 6
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - neigh_diff, 10
 - sd_neighbor, 12
 - sd_neighbor_xlsx, 13
 - sd_table_xlsx, 14
 - * **F33**
 - f33pull, 4
 - f33pull_raw, 6
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **SAIPE**
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **borders**
 - big_borders, 2
 - * **codebook**
 - master_codebook, 9
 - * **data**
 - f33pull, 4
 - f33pull_raw, 6
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **difference**
 - neigh_diff, 10
 - sd_neighbor, 12
 - sd_neighbor_xlsx, 13
 - sd_table_xlsx, 14
 - * **master**
 - long_masterpull, 7
 - master_codebook, 9
 - masterpull, 7
 - * **neighbors**
 - neigh_diff, 10
 - sd_neighbor, 12
 - sd_neighbor_xlsx, 13
 - sd_table_xlsx, 14
 - * **pooling**
 - big_borders, 2
 - * **rank**
 - neigh_diff, 10
 - * **revenue**
 - big_borders, 2
 - * **round**
 - round2, 11
 - * **table**
 - sd_neighbor, 12
 - sd_neighbor_xlsx, 13
 - sd_table_xlsx, 14
- big_borders, 2, 3
- edbuildr, 3
- f33pull, 3, 4, 6
- f33pull_raw, 3, 5, 6
- long_masterpull, 3, 7, 9, 10
- master_codebook, 3, 7, 9, 9, 10, 12–14
- masterpull, 3, 4, 7, 7, 10, 12–14
- neigh_diff, 4, 10, 12, 13

round2, [4](#), [11](#)

sd_neighbor, [4](#), [12](#)

sd_neighbor_xlsx, [4](#), [10](#), [12](#), [13](#), [14](#)

sd_table_xlsx, [4](#), [12](#), [13](#), [14](#)