# Package ‘edbuildr’

March 19, 2020

**Type** Package  
**Title** Automated School District Data Download and Processing  
**Version** 0.1.1  
**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.  
**License** CC0  
**Imports** dplyr, magrittr, openxlsx, scales, stringr, tidyselect (>= 1.0.0)  
**Encoding** UTF-8  
**LazyData** true  
**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 <megan@edbuild.org>  
**Repository** CRAN  
**Date/Publication** 2020-03-19 12:50:05 UTC
**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 2013-2017 or longitudinally for 2013-2017. 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 (2019) <https://edbuild.org/content/dividing-lines/main/methodology>.

**edbuildr functions**

The edbuildr functions are:

- **f33pull** Pulls in the US Census’s Annual Survey of School System Finances (F33) and processes according to EdBuild’s adjustments. Available for 2006-2017.


- **long_masterpull** Pulls in EdBuild’s master dataset longitudinally for years 2013-2017.

- **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.
f33pull  

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 2013-2017.

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

round2 Rounds 0.5 up.

sd_neighbor_xlsx Creates a formatted table (.xlsx) of school districts 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 (megan@edbuild.org), maintainer
- Cecilia Depman (cecilia@edbuild.org), author
- Sara Hodges (sara@edbuild.org), author
- Kailey Spencer (kailey.spencer.depoe@gmail.com), contributor

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, path=NULL, additional_var=NULL, keep_calcs=FALSE)

Arguments


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


See Also

- f33pull_raw
**Examples**

```r
f33_2017 <- f33pull(data_year="2017",
                     additional_var=c("V40", "TCAPOUT"),
                     keep_calcs=FALSE)
```

### Description

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

### Usage

```r
f33pull_raw(data_year = "2017")
```

### Arguments

- `data_year` 

### 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


### See Also

- `f33pull`

### Examples

```r
f33_2017 <- f33pull_raw('2017')
```
A function to import a longitudinal version of EdBuild’s master dataset

Description

This function imports a longitudinal dataset of EdBuild’s master data for the years 2013-2017. 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 92,636 observations and 41 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()

A function to import EdBuild’s master dataset

Description

This function allows you to read in EdBuild’s master dataset for the years 2013-2017. 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 = "2017", data_type = "gen")

Arguments

data_type: Type of master data to pull in
  • 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.

Defaults to gen

Format

A data frame with 40 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/2017/full_data_17_type_exc.csv

See Also
master_codebook, long_masterpull

Examples
master17_geo <- masterpull("2017", 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
neigh_diff  

Examples

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

---

**neigh_diff**  
*A function to find the difference between school district neighbors*

**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**

```r
neigh_diff(data_year = "2017",
           diff_var = "Percentage Point Difference in Poverty Rate",
           type = "like")
```

**Arguments**

- `data_year`  

- `diff_var`  
  Character string on which to rank the difference between school district neighbors. Use `diff_var = "options"` to print a list of the variables. Defaults to Percentage Point Difference in Poverty Rate.

- `type`  
  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**

```r
tr_diff <- neigh_diff(
  data_year = "2017",
  diff_var = "Difference in Total Revenue Per Pupil"
)
```

**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**

```r
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](https://stackoverflow.com/questions/12688717/round-up-from-5)

**Examples**

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

**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**

```r
ds_neighbor_xlsx(data_year = "2017", school_district = NULL,
                 table_vars = c('Name', 'Enrollment', 'Poverty Rate', 'Percent Nonwhite',
                                'Local Revenue PP', 'State Revenue PP', 'Type'))
```
Arguments

- **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 `table_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

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

```r
table <- sd_neighbor_xlsx(
  data_year = "2017",
  school_district = "3402640",
  table_vars = c("Name",
                 "Percent FRL",
                 "Median Household Income",
                 "State Revenue PP")
)
```

Description

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

Usage

```r
sd_table_xlsx(data_year = '2017', state = "New Jersey", county = NULL,
              table_vars = c('Name', 'Enrollment', 'Poverty Rate', 'Percent Nonwhite',
                             'Local Revenue PP', 'State Revenue PP', 'Total Revenue PP'))
```
Arguments

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 = "2017",
state = "Maryland",
county = c("Baltimore County", "Baltimore City", "Howard County", "Carroll County"),
table_vars = c("Name", "Poverty Rate")
)
Index

*Topic CCD
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic EDGE
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic EdBuild
  f33pull, 3
  f33pull_raw, 5
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6
  neigh_diff, 9
  sd_neighbor_xlsx, 10
  sd_table_xlsx, 11

*Topic F33
  f33pull, 3
  f33pull_raw, 5
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic SAIPER
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic codebook
  master_codebook, 8

*Topic data
  f33pull, 3
  f33pull_raw, 5
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic difference
  neigh_diff, 9
  sd_neighbor_xlsx, 10
  sd_table_xlsx, 11

*Topic master
  long_masterpull, 6
  master_codebook, 8
  masterpull, 6

*Topic neighbors
  neigh_diff, 9
  sd_neighbor_xlsx, 10
  sd_table_xlsx, 11

*Topic rank
  neigh_diff, 9

*Topic round
  round2, 10

*Topic table
  sd_neighbor_xlsx, 10
  sd_table_xlsx, 11

edbuildr, 2
f33pull, 2, 3, 5
f33pull_raw, 2, 4, 5
long_masterpull, 2, 6, 8
master_codebook, 6, 8, 9, 11, 12
masterpull, 3, 6, 6, 8, 9, 11, 12
neigh_diff, 3, 9, 11
round2, 3, 10
sd_neighbor_xlsx, 3, 9, 10, 12
sd_table_xlsx, 3, 11, 11