

Package ‘rscorecard’

March 29, 2018

Title A Method to Download Department of Education College Scorecard Data

Version 0.7.1

Maintainer Benjamin Skinner <btskinner@virginia.edu>

URL <https://github.com/btskinner/rscorecard>

BugReports <https://github.com/btskinner/rscorecard/issues>

Description A method to download Department of Education College Scorecard data using the public API <<https://collegescorecard.ed.gov/data/documentation/>>. It is based on the 'dplyr' model of piped commands to select and filter data in a single chained function call. An API key from the U.S. Department of Education is required.

Depends R (>= 3.2.3)

License MIT + file LICENSE

LazyData true

Imports dplyr, httr, jsonlite, lazyeval, magrittr

RoxygenNote 6.0.1

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

NeedsCompilation no

Author Benjamin Skinner [aut, cre] (<<https://orcid.org/0000-0002-0337-7415>>)

Repository CRAN

Date/Publication 2018-03-29 21:43:18 UTC

R topics documented:

rscorecard	2
sc_dict	2
sc_filter	3
sc_get	3

sc_init	4
sc_key	5
sc_select	6
sc_year	6
sc_zip	7

Index	9
--------------	----------

rscorecard	<i>rscorecard: A Method to Download College Scorecard Data.</i>
------------	---

Description

The `rscorecard` package provides a series of piped functions (a la `dplyr`) to facilitate downloading Department of Education College Scorecard data. In reality it is simply a method for converting idiomatic R code into a properly formatted URL string that is then queried. This package requires an API key, which can be requested at <https://api.data.gov/signup/>.

Details

All command pipes must start with `sc_init()`, end with `sc_get()`, and be linked with the `magrittr` pipe function, `%>%`. Internal commands, `sc_select`, `sc_filter`, `sc_year`, `sc_zip`, can come in any order in the pipe chain. Only `sc_select` is required.

sc_dict	<i>Search data dictionary.</i>
---------	--------------------------------

Description

This function is used to search the College Scorecard data dictionary.

Usage

```
sc_dict(search_string, search_col = c("all", "description", "varname",
  "dev_friendly_name", "dev_category", "label"), ignore_case = TRUE,
  limit = 10, confirm = FALSE)
```

Arguments

<code>search_string</code>	Character string for search. Can use regular expression for search. Must escape special characters, <code>.</code> <code>\</code> <code> </code> <code>(</code> <code>)</code> <code>[</code> <code>{</code> <code>^</code> <code>\$</code> <code>*</code> <code>+</code> <code>?</code> , with a doublebackslash <code>\\</code> .
<code>search_col</code>	Column to search. The default is to search all columns. Other options include: "varname", "dev_friendly_name", "dev_category", "label".
<code>ignore_case</code>	Search is case insensitive by default. Change to <code>FALSE</code> to restrict search to exact case matches.
<code>limit</code>	Only the first 10 dictionary items are returned by default. Increase to return more values. Set to <code>Inf</code> to return all items matched in search'
<code>confirm</code>	Use to confirm status of variable name in dictionary. Returns <code>TRUE</code> or <code>FALSE</code> .

Examples

```
sc_dict('state')
sc_dict('^st', search_col = 'varname') # variable names starting with 'st'
sc_dict('.', limit = Inf) # return full dictionary (not recommended)
```

sc_filter	<i>Filter scorecard data by variable values.</i>
-----------	--

Description

This function is used to filter the downloaded scorecard data. It converts idiomatic R into the format required by the API call.

Usage

```
sc_filter(sccall, ...)
```

Arguments

sccall	Current list of parameters carried forward from prior functions in the chain (ignore)
...	Expressions to evaluate

Examples

```
## Not run:
sc_filter(region == 1) # New England institutions
sc_filter(stabbr == c('TN','KY')) # institutions in Tennessee and Kentucky
sc_filter(control != 3) # exclude private, for-profit institutions
sc_filter(control == c(1,2)) # same as above
sc_filter(control == 1:2) # same as above
sc_filter(stabbr == 'TN', control == 1, locale == 41:43) # TN rural publics

## End(Not run)
```

sc_get	<i>Get scorecard data.</i>
--------	----------------------------

Description

This function gets the College Scorecard data by compiling and converting all the previous piped output into a single URL string that is used to get the data.

Usage

```
sc_get(sccall, api_key, debug = FALSE, print_key_debug = FALSE)
```

Arguments

sccall	Current list of parameters carried forward from prior functions in the chain (ignore)
api_key	Personal API key requested from https://api.data.gov/signup stored in a string. If you first set your key using sc_key, then you may omit this parameter. A key set here will take precedence over any set in the environment (DATAGOV_API_KEY).
debug	Set to true to print and return API call (URL string) rather than make actual request. Should only be used when debugging calls.
print_key_debug	Only used when debug == TRUE. Default masks the api_key value. Set to TRUE to print the full API call string with the api_key unmasked.

Obtain a key

To obtain an API key, visit <https://api.data.gov/signup>

Examples

```
## Not run:
sc_get('<API KEY IN STRING>')
key <- '<API KEY IN STRING>'
sc_get(key)

## End(Not run)
```

sc_init	<i>Initialize chained request.</i>
---------	------------------------------------

Description

This function initializes the data request. It should always be the first in the series of piped functions.

Usage

```
sc_init(dfvars = FALSE)
```

Arguments

dfvars	Set to TRUE if you would rather use the developer-friendly variable names used in actual API call.
--------	--

Examples

```
## Not run:  
sc_init()  
sc_init(dfvars = TRUE)  
  
## End(Not run)
```

sc_key	<i>Store Data.gov API key in system environment.</i>
--------	--

Description

This function stores your data.gov API key in the system environment so that you only have to load it once at the start of the session. If you set your key using `sc_key`, then you may omit `api_key` parameter in the `sc_get` function.

Usage

```
sc_key(api_key)
```

Arguments

api_key	Personal API key requested from https://api.data.gov/signup stored in a string.
---------	---

Obtain a key

To obtain an API key, visit <https://api.data.gov/signup>.

Examples

```
## Not run:  
sc_key('<API KEY IN STRING>')  
  
## End(Not run)
```

sc_select	<i>Select scorecard data variables.</i>
-----------	---

Description

This function is used to select the variables returned in the final dataset.

Usage

```
sc_select(sccall, ...)
```

Arguments

sccall	Current list of parameters carried forward from prior functions in the chain (ignore)
...	Desired variable names separated by commas (not case sensitive)

Examples

```
## Not run:
sc_select(UNITID)
sc_select(UNITID, INSTNM)
sc_select(unitid, instnm)

## End(Not run)
```

sc_year	<i>Select scorecard data year.</i>
---------	------------------------------------

Description

This function is used to select the year of the data.

Usage

```
sc_year(sccall, year)
```

Arguments

sccall	Current list of parameters carried forward from prior functions in the chain (ignore)
year	Four-digit year (default is 2013)

Important notes

1. Not all variables have a year option.
2. At this time, only one year at a time is allowed.
3. The year selected is not necessarily the year the data were produced. It may be the year the data were collected. For data collected over split years (fall to spring), it is likely the year represents the fall data (*e.g.*, 2011 for 2011/2012 data).

Be sure to check with the College Scorecard [data documentation report](#) when choosing the year.

Examples

```
## Not run:  
sc_year()  
sc_year(2012)  
  
## End(Not run)
```

sc_zip

Subset results to those within specified area around zip code.

Description

Subset results to those within specified area around zip code.

Usage

```
sc_zip(sccall, zip, distance = 25, km = FALSE)
```

Arguments

sccall	Current list of parameters carried forward from prior functions in the chain (ignore)
zip	A 5-digit zipcode
distance	An integer distance in miles or kilometers
km	A boolean value set to TRUE if distance should be in kilometers (default is FALSE for miles)

Note

Zip codes with leading zeros (Northeast) can be called either using a string ('02111') or as a numeric (02111). R will drop the leading zero from the second version, but `sc_zip()` will add it back before the call. The shortened version without the leading zero may also be used (2111 and '2111' both become '02111'), but is not recommended for clarity.

Examples

```
## Not run:
sc_zip(37203)
sc_zip(37203, 50)
sc_zip(37203, 50, km = TRUE)
sc_zip('02111')      # 1. Using string
sc_zip(02111)        # 2. Dropped leading zero will be added
sc_zip(2111)         # 3. Will become '02111' (not recommended)

## End(Not run)
```


Index

rscorecard, [2](#)
rscorecard-package (rscorecard), [2](#)

sc_dict, [2](#)
sc_filter, [2, 3](#)
sc_get, [3](#)
sc_init, [4](#)
sc_key, [5](#)
sc_select, [2, 6](#)
sc_year, [2, 6](#)
sc_zip, [2, 7](#)