

Package ‘cansim’

March 13, 2020

Type Package

Title Accessing Statistics Canada Data Table and Vectors

Version 0.3.5

Maintainer Jens von Bergmann <jens@mountainmath.ca>

Description Searches for, accesses, and retrieves new-format and old-format Statistics Canada data tables, as well as individual vectors, as tidy data frames. This package deals with encoding issues, allows for bilingual English or French language data retrieval, and bundles convenience functions to make it easier to work with retrieved table data. Optional caching features are provided.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

Imports dplyr (>= 0.7), httr (>= 1.0.0), readr, rlang, xml2, rvest, stringr, purrr, tibble, utils, jsonlite

RoxygenNote 7.1.0

Suggests knitr, rmarkdown, scales, tidyr, ggplot2

URL <https://github.com/mountainMath/cansim>,
<https://mountainmath.github.io/cansim/>

BugReports <https://github.com/mountainMath/cansim/issues>

VignetteBuilder knitr

Language en-CA

NeedsCompilation no

Author Jens von Bergmann [cre],
Dmitry Shkolnik [aut]

Repository CRAN

Date/Publication 2020-03-13 10:00:20 UTC

R topics documented:

add_provincial_abbreviations	2
adjust_cansim_values_by_variable	3
cansim_old_to_new	3
categories_for_level	4
correspondence	5
get_cansim	5
get_cansim_changed_tables	6
get_cansim_code_set	7
get_cansim_column_categories	7
get_cansim_column_list	8
get_cansim_cube_metadata	9
get_cansim_data_for_table_coord_periods	10
get_cansim_ndm	10
get_cansim_table_info	11
get_cansim_table_notes	12
get_cansim_table_overview	13
get_cansim_table_short_notes	13
get_cansim_table_subject	14
get_cansim_table_survey	15
get_cansim_table_url	16
get_cansim_vector	16
get_cansim_vector_for_latest_periods	17
get_cansim_vector_info	18
list_cansim_cubes	18
list_cansim_tables	19
normalize_cansim_values	20
search_cansim_cubes	21
search_cansim_tables	22
view_cansim_webpage	23
Index	24

add_provincial_abbreviations

Add provincial abbreviations as factor

Description

Add provincial abbreviations as factor

Usage

add_provincial_abbreviations(data)

Arguments

data codecansim package data frame with provincial level data

Value

a codecansim package data frame with additional factor GEO.abb that contains language-specific provincial abbreviations

Examples

```
df <- get_cansim("17-10-0005")
df <- add_provincial_abbreviations(df)
```

adjust_cansim_values_by_variable

(Deprecated) Adjust values in a retrieved Statistics Canada data table using a scaling variable

Description

(Deprecated) Adjust retrieved data table values by a scaled amount; however French does not work, probably due to encoding issues. This function is now deprecated and should not be used.

Usage

```
adjust_cansim_values_by_variable(data, var)
```

Arguments

data	A downloaded CANSIM data table
var	A now deprecated input

cansim_old_to_new	<i>Translate deprecated CANSIM table number into new NDM-format table catalogue number</i>
-------------------	--

Description

Returns NDM table catalogue equivalent given a standard old-format CANSIM table number

Usage

```
cansim_old_to_new(oldCansimTableNumber)
```

Arguments

oldCansimTableNumber	deprecated style CANSIM table number (e.g. "427-0001")
----------------------	--

Value

new-format NDM table number

Examples

```
cansim_old_to_new("026-0018")
```

categories_for_level *Use metadata to extract categories for column of specific level*

Description

For tables with data with hierarchical categories, metadata containing hierarchy level descriptions is used to extract categories at a specified level of hierarchy only.

Usage

```
categories_for_level(
  data,
  column_name,
  level = NA,
  strict = FALSE,
  remove_duplicates = TRUE
)
```

Arguments

data	data table object as returned from <code>get_cansim()</code>
column_name	the quoted name of the column to extract categories from
level	the hierarchy level depth to which to extract categories, where 0 is top category
strict	(default FALSE) when TRUE will only extract that specific hierarchy level
remove_duplicates	(default TRUE) When set to TRUE higher level grouping categories already captured by lower level hierarchy data will be removed

Value

A vector of categories

Examples

```
data <- get_cansim("16-10-0117")
categories_for_level(data, "North American Industry Classification System (NAICS)", level=2)
```

correspondence	<i>The correspondence file for old to new StatCan table numbers is included in the package</i>
----------------	--

Description

The correspondence file for old to new StatCan table numbers is included in the package

Author(s)

Statistics Canada

References

https://www.statcan.gc.ca/eng/developers-developpeurs/cansim_id-product_id-concordance.csv

get_cansim	<i>Retrieves a Statistics Canada data table using CANSIM code or NDM table number</i>
------------	---

Description

Retrieves a data table as a tidy dataframe using either an old-style CANSIM code or a new-format NDM table number. This function will automatically convert old-style CANSIM codes into their new equivalents. Retrieved table data is cached for the duration of the current R session only by default.

Usage

```
get_cansim(
  cansimTableNumber,
  language = "english",
  refresh = FALSE,
  timeout = 200
)
```

Arguments

cansimTableNumber	the table number to load, accepts old CANSIM or new NDM table numbers
language	"en" or "english" for English and "fr" or "french" for French language versions (default is set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Value

tibble format data table output

Examples

```
# Retrieve a table with an NDM code
get_cansim("34-10-0013")
# Retrieve a table with an old-style CANSIM code
get_cansim("026-0018")
```

get_cansim_changed_tables

Retrieve a list of modified tables since a given date

Description

Retrieve a list of tables that have been modified or updated since the specified date.

Usage

```
get_cansim_changed_tables(start_date)
```

Arguments

start_date	Starting date in YYYY-MM-DD format to look for changes that changed on or after that date
------------	---

Value

A tibble with Statistics Canada data table product ids and their release times

Examples

```
get_cansim_changed_tables("2018-08-01")
```

get_cansim_code_set *Get NDM code sets*

Description

Useful to get a list of surveys or subjects and used internally

Usage

```
get_cansim_code_set(  
  code_set = c("scalar", "frequency", "symbol", "status", "uom", "survey", "subject",  
              "wdsResponseStatus"),  
  refresh = FALSE  
)
```

Arguments

code_set	the code set to retrieve.
refresh	Default is FALSE, repeated calls during the same session will hit the cached data. To refresh the code list during a running R session set to TRUE

Value

a tibble with english and french labels for the given code set

Examples

```
get_cansim_code_set("survey")
```

get_cansim_column_categories
Retrieve Statistics Canada data table categories for a specific column

Description

Returns table column details given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_column_categories(  
  cansimTableNumber,  
  column,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
column	the specified column for which to retrieve category information for
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_column_categories("34-10-0013", "Geography")
```

```
get_cansim_column_list
```

Retrieve Statistics Canada data table column list

Description

Returns table column details given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_column_list(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```


Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_column_list("34-10-0013")
```

```
get_cansim_cube_metadata
```

Retrieve table metadata from Statistics Canada API

Description

Retrieves table metadata given an input table number or vector of table numbers using either the new or old table number format. Patience is suggested as the Statistics Canada API can be very slow. The 'list_cansim_tables()' function can be used as an alternative to retrieve a (cached) list of CANSIM tables with (more limited) metadata.

Usage

```
get_cansim_cube_metadata(cansimTableNumber)
```

Arguments

cansimTableNumber	A new or old CANSIM/NDM table number or a vector of table numbers
-------------------	---

Value

a tibble containing the table metadata

Examples

```
get_cansim_cube_metadata("34-10-0013")
```

```
get_cansim_data_for_table_coord_periods
```

Retrieve data for specified Statistics Canada data product for last N periods for specific coordinates

Description

Allows for the retrieval of data for a Statistics Canada data table with specific coordinates for the N most-recently released periods. Caution: coordinates are concatenations of table member ID values and require familiarity with the TableMetadata data structure. Coordinates have a maximum of ten dimensions.

Usage

```
get_cansim_data_for_table_coord_periods(
  cansimTableNumber,
  coordinate,
  periods = 1
)
```

Arguments

cansimTableNumber	Statistics Canada data table number
coordinate	A string of table coordinates in the form "1.1.1.36.1.0.0.0.0.0"
periods	Numeric value for number of latest periods to retrieve data for

Value

A tibble with data matching specified coordinate and period input arguments

Examples

```
get_cansim_data_for_table_coord_periods("35-10-0003",coordinate="1.12.0.0.0.0.0.0.0",periods=3)
```

```
get_cansim_ndm
```

Retrieve a Statistics Canada data table using NDM catalogue number

Description

Retrieves a data table using an NDM catalogue number as a tidy data frame. Retrieved table data is cached for the duration of the current R session only by default.

Usage

```
get_cansim_ndm(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (defaults to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Value

tibble format data table output

Examples

```
get_cansim("34-10-0013")
```

get_cansim_table_info *Retrieve Statistics Canada data table information*

Description

Returns table information given an NDM table catalogue number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_table_info(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_table_info("34-10-0013")
```

```
get_cansim_table_notes
```

Retrieve Statistics Canada data table notes and column categories

Description

Returns table notes given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_table_notes(
  cansimTableNumber,
  language = "en",
  refresh = FALSE,
  timeout = 200
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_table_notes("34-10-0013")
```

`get_cansim_table_overview`*Retrieve Statistics Canada data table overview text*

Description

Prints table overview information as console output. In order to display table overview information, the selected CANSIM table must be loaded entirely to display overview information. Overview information is printed in console an in English or French, as specified.

Usage

```
get_cansim_table_overview(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)

Value

none

Examples

```
get_cansim_table_overview("34-10-0013")
```

`get_cansim_table_short_notes`*Retrieve Statistics Canada data table short notes*

Description

Returns table notes given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_table_short_notes(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_table_short_notes("34-10-0013")
```

```
get_cansim_table_subject
```

Retrieve Statistics Canada data table subject detail

Description

Returns table subject detail given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_table_subject(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_table_subject("34-10-0013")
```

```
get_cansim_table_survey
```

Retrieve Statistics Canada data table survey detail

Description

Returns table survey detail given an NDM table number in English or French. Retrieved table information data is cached for the duration of the R session only.

Usage

```
get_cansim_table_survey(  
  cansimTableNumber,  
  language = "english",  
  refresh = FALSE,  
  timeout = 200  
)
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (default set to English)
refresh	(Optional) When set to TRUE, forces a reload of data table (default is FALSE)
timeout	(Optional) Timeout in seconds for downloading cansim table to work around scenarios where StatCan servers drop the network connection.

Examples

```
get_cansim_table_survey("34-10-0013")
```

get_cansim_table_url *Retrieve a Statistics Canada data table URL given a table number*

Description

Retrieve URL of a table from the API given a table number. Offers a more stable approach than manually guessing the URL of the table.

Usage

```
get_cansim_table_url(cansimTableNumber, language = "en")
```

Arguments

cansimTableNumber	the NDM table number to load
language	"en" or "english" for English and "fr" or "french" for French language versions (defaults to English)

Value

String object containing URL for specified table number

Examples

```
get_cansim_table_url("34-10-0013")  
get_cansim_table_url("34-10-0013", language = "fr")
```

get_cansim_vector *Retrieve data for a Statistics Canada data vector released within a given time frame*

Description

Retrieve data for a Statistics Canada data vector released within a given time frame

Usage

```
get_cansim_vector(  
  vectors,  
  start_time,  
  end_time = Sys.time(),  
  use_ref_date = TRUE  
)
```


Arguments

vectors	The list of vectors to retrieve
start_time	Starting date in YYYY-MM-DD format, applies to REF_DATE or releaseTime, depending on use_ref_date parameter
end_time	Set an optional end time filter in YYYY-MM-DD format (defaults to current system time)
use_ref_date	Optional, TRUE by default. When set to TRUE, uses REF_DATE of vector data to filter, otherwise it uses Statistics Canada's releaseDate value for filtering the specified vectors

Value

A tibble with data for vectors released between start and end time

Examples

```
get_cansim_vector("v41690973", "2015-01-01")
```

get_cansim_vector_for_latest_periods
Retrieve data for specified Statistics Canada data vector(s) for last N periods

Description

Allows for the retrieval of data for specified vector series for the N most-recently released periods.

Usage

```
get_cansim_vector_for_latest_periods(vectors, periods = 1)
```

Arguments

vectors	The list of vectors to retrieve
periods	Numeric value for number of latest periods to retrieve data for

Value

A tibble with data for specified vector(s) for the last N periods

Examples

```
get_cansim_vector_for_latest_periods("v41690973", 10)
```

`get_cansim_vector_info`*Retrieve metadata for specified Statistics Canada data vectors*

Description

Allows for the retrieval of metadata for Statistics Canada data vectors

Usage

```
get_cansim_vector_info(vectors)
```

Arguments

`vectors` a vector of cansim vectors

Value

A tibble with metadata for selected vectors

Examples

```
get_cansim_vector_info("v41690973")
```

`list_cansim_cubes`*Get overview list for all Statistics Canada data cubes*

Description

Generates an overview table containing metadata of available Statistics Canada data cubes.

Usage

```
list_cansim_cubes(lite = FALSE, refresh = FALSE)
```

Arguments

`lite` Get the version without cube dimensions and comments for faster retrieval, default is FALSE.

`refresh` Default is FALSE, repeated calls during the same session will hit the cached data. To refresh the code list during a running R session set to TRUE

Value

A tibble with available Statistics Canada data cubes, including NDM table number, cube title, start and end dates, achieve status, subject and survey codes, frequency codes and a list of cube dimensions.

Examples

```
list_cansim_cubes()
```

```
list_cansim_tables
```

Get overview list for all Statistics Canada data tables

Description

Generates an overview table containing metadata of available Statistics Canada data tables. A new and updated table will be generated if this table does not already exist in cached form or if the force refresh option is selected (set to FALSE by default). This can take some time as this process involves scraping through hundreds of Statistics Canada web pages to gather the required metadata. If option `cansim.cache_path` is set it will look for and store the overview table in that directory.

Usage

```
list_cansim_tables(refresh = FALSE)
```

Arguments

`refresh` Default is FALSE, and will regenerate the table if set to TRUE

Value

A tibble with available Statistics Canada data tables, listing title, Statistics Canada data table catalogue number, deprecated CANSIM table number, description, and geography

Examples

```
list_cansim_tables()
```

```
normalize_cansim_values
```

Normalize retrieved data table values to appropriate scales

Description

Facilitates working with Statistics Canada data table values retrieved using the package by setting all units to counts/dollars instead of millions, etc. If "replacement_value" is not set, it will replace the VALUE field with normalized values and drop the scale column. Otherwise it will keep the scale columns and create a new column named replacement_value with the normalized value. It will attempt to parse the REF_DATE field and create an R date variable. This is currently experimental.

Usage

```
normalize_cansim_values(  
  data,  
  replacement_value = NA,  
  normalize_percent = TRUE,  
  default_month = "01",  
  default_day = "01",  
  factors = FALSE,  
  strip_classification_code = FALSE  
)
```

Arguments

data	A retrieved data table as returned from <code>get_cansim()</code> or <code>get_cansim_ndm()</code>
replacement_value	(Optional) the name of the column the manipulated value should be returned in. Defaults to replacing the current value field
normalize_percent	(Optional) When true (the default) normalizes percentages by changing them to rates
default_month	The default month that should be used when creating Date objects for annual data (default set to "01")
default_day	The default day of the month that should be used when creating Date objects for monthly data (default set to "01")
factors	(Optional) Logical value indicating if dimensions should be converted to factors. (Default set to false).
strip_classification_code	(strip_classification_code) Logical value indicating if classification code should be stripped from names. (Default set to false).

Value

Returns the input tibble with with adjusted values

Examples

```
cansim_table <- get_cansim("34-10-0013")
normalize_cansim_values(cansim_table)
```

search_cansim_cubes *Search through Statistics Canada data cubes*

Description

Searches through Statistics Canada data cubes using a search term.

Usage

```
search_cansim_cubes(search_term, refresh = FALSE)
```

Arguments

search_term	User-supplied search term used to find Statistics Canada data cubes with matching titles, table numbers, subject and survey codes.
refresh	Default is FALSE. The underlying cube list is cached for the duration of the R sessions and will regenerate the cube list if set to TRUE

Value

A tibble with available Statistics Canada data cubes, listing title, Statistics Canada data cube catalogue number, deprecated CANSIM table number, survey and subject.

Examples

```
search_cansim_cubes("Labour force")
```

search_cansim_tables *Search through Statistics Canada data tables*

Description

Searches through Statistics Canada data tables using a search term. A new table is generated if it already does not exist or if refresh option is set to TRUE. Search-terms are case insensitive, but will accept regular expressions for more advanced searching. The search function can search either through table titles or through table descriptions, depending on the whether or not search_description is set to TRUE or not. If refresh = TRUE, the table will be updated and regenerated using Statistics Canada's latest data. This can take some time since this process involves scraping through several hundred web pages to gather the required metadata. If option cache_path is set it will look for and store the overview table in that directory.

Usage

```
search_cansim_tables(search_term, search_fields = "both", refresh = FALSE)
```

Arguments

search_term	User-supplied search term used to find Statistics Canada data tables with matching titles
search_fields	By default, this function will search through table titles and keywords. Setting this parameter to "title" will only search through the title, setting it to "keyword" will only search through keywords
refresh	Default is FALSE, and will regenerate the table if set to TRUE

Value

A tibble with available Statistics Canada data tables, listing title, Statistics Canada data table catalogue number, deprecated CANSIM table number, description and geography that match the search term.

Examples

```
search_cansim_tables("Labour force")
```

view_cansim_webpage *View CANSIM table information in browser*

Description

Opens CANSIM table on Statistics Canada's website using default browser. This may be useful for getting further info on CANSIM table and survey methods.

Usage

```
view_cansim_webpage(cansimTableNumber)
```

Arguments

cansimTableNumber
CANSIM or NDM table number

Value

none

Examples

```
view_cansim_webpage("34-10-0013")
```

Index

*Topic **data**

- correspondence, [5](#)

- [add_provincial_abbreviations, 2](#)
- [adjust_cansim_values_by_variable, 3](#)

- [cansim_old_to_new, 3](#)
- [categories_for_level, 4](#)
- [correspondence, 5](#)

- [get_cansim, 5](#)
- [get_cansim_changed_tables, 6](#)
- [get_cansim_code_set, 7](#)
- [get_cansim_column_categories, 7](#)
- [get_cansim_column_list, 8](#)
- [get_cansim_cube_metadata, 9](#)
- [get_cansim_data_for_table_coord_periods, 10](#)
- [get_cansim_ndm, 10](#)
- [get_cansim_table_info, 11](#)
- [get_cansim_table_notes, 12](#)
- [get_cansim_table_overview, 13](#)
- [get_cansim_table_short_notes, 13](#)
- [get_cansim_table_subject, 14](#)
- [get_cansim_table_survey, 15](#)
- [get_cansim_table_url, 16](#)
- [get_cansim_vector, 16](#)
- [get_cansim_vector_for_latest_periods, 17](#)
- [get_cansim_vector_info, 18](#)

- [list_cansim_cubes, 18](#)
- [list_cansim_tables, 19](#)

- [normalize_cansim_values, 20](#)

- [search_cansim_cubes, 21](#)
- [search_cansim_tables, 22](#)

- [view_cansim_webpage, 23](#)