

Package ‘raustats’

March 13, 2019

Type Package

Title Access Data and Statistics from the ABS and RBA Websites

Description Functions for downloading Australian economic statistics from the Australian Bureau of Statistics (ABS) (see <<https://www.abs.gov.au/>>) and Reserve Bank of Australia (RBA) (see <<https://www.rba.gov.au/>>) websites.

Version 0.1.0

Date 2019-03-03

Maintainer David Mitchell <david.pk.mitchell@gmail.com>

URL <https://github.com/mitcda/raustats>

License GPL-3

Depends R (>= 3.3.0), readxl

Imports stats, dplyr, httr, jsonlite, rvest, tidyr, xml2, lubridate

Suggests repo, knitr, rmarkdown, testthat, ggplot2

LazyData TRUE

RoxygenNote 6.1.1

VignetteBuilder knitr

NeedsCompilation no

Encoding UTF-8

Author David Mitchell [aut, cre]

Repository CRAN

Date/Publication 2019-03-13 15:20:03 UTC

R topics documented:

abs_cachelist	2
abs_cat_cachelist	3
abs_cat_download	3
abs_cat_stats	4
abs_cat_tables	5

abs_cat_unzip	6
abs_datasets	6
abs_dimensions	7
abs_metadata	8
abs_read_tss	9
abs_search	9
abs_stats	10
aus_state_codes	12
raustats	13
rba_cachelist	13
rba_file_download	14
rba_read_tss	14
rba_search	15
rba_stats	16
rba_table_cache	17

Index	18
--------------	-----------

abs_cachelist	<i>Datasets available through the ABS API</i>
---------------	---

Description

This data set provides a list of all datasets, and the associated metadata, available through the ABS API.

Usage

abs_cachelist

Format

A data frame containing three columns:

- id ABS dataset identifier.
- agencyID Source agency identifier (ABS).
- name ABS dataset name.

abs_cat_cachelist	<i>List of ABS catalogue tables</i>
-------------------	-------------------------------------

Description

This data set provides a list of the most common ABS catalogue tables.

Usage

```
abs_cat_cachelist
```

Format

A data frame containing five columns:

- `publication_title` ABS publication title.
- `catalogue_no` ABS catalogue number.
- `abs_url` ABS URL.
- `last_updated` Publication last updated.
- `type` Publication type – one of either 'time series', 'panel' or 'summary'.RBA URL

abs_cat_download	<i>Function to download files from the ABS website and store locally</i>
------------------	--

Description

Gets files from ABS website specified by URL

Usage

```
abs_cat_download(data_url, exdir = tempdir())
```

Arguments

<code>data_url</code>	Character vector specifying an ABS data URLs.
<code>exdir</code>	Target directory for downloaded files (defaults to <code>tempdir()</code>). Directory is created if it doesn't exist.

Value

Downloads data from the ABS website and returns a character vector listing the location where files are saved.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

abs_cat_stats *Get ABS catalogue series data*

Description

TBC

Usage

```
abs_cat_stats(cat_no, tables = "All", releases = "Latest",
              types = "tss")
```

Arguments

cat_no	Character vector specifying one or more ABS collections or catalogue numbers to download.
tables	A character vector of regular expressions denoting tables to download. The default ('All') downloads all time series spreadsheet tables for each specified catalogue. Use a list to specify different table sets for each specified ABS catalogue number.
releases	Date or character string object specifying the month and year denoting which release to download. Default is "Latest", which downloads the latest available data. See examples for further details.
types	One of either 'tss' – ABS time series spreadsheet (the default) or 'css' – ABS data cube (cross-section spreadsheet).

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## Download quarterly Australian National Accounts, Tables 1 & 2
ana_q <- abs_cat_stats("5206.0", tables=c("Table 1\\W+", "Table 2\\W+"));

## Download December 2017 Australian National Accounts, Table 1
ana_q_2017q4 <- abs_cat_stats("5206.0", tables="Table 1\\W+", release="Dec 2017");
```

abs_cat_tables	<i>Return ABS catalogue tables</i>
----------------	------------------------------------

Description

Return list of tables from specified ABS catalogue number

Usage

```
abs_cat_tables(cat_no, releases = "Latest", types = c("tss", "css"),
  include_urls = FALSE)
```

Arguments

cat_no	ABS catalogue numbers.
releases	Date or character string object specifying the month and year denoting which release to download. Default is "Latest", which downloads the latest available data. See examples for further details.
types	ABS publication types to return. Permissible options include one or more of: 'tss' – ABS Time Series Spreadsheets, 'css' - ABS Data Cubes and 'pub' – ABS Publications. The default returns all Time Series Spreadsheets and Data Cubes.
include_urls	Include full URLs to returned ABS data files. Default (FALSE) does not include data file URLs.

Value

Returns a data frame listing the data collection tables and links.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## List latest available quarterly National Accounts tables
ana_tables <- abs_cat_tables("5206.0", releases="Latest");
ana_tables_url <- abs_cat_tables("5206.0", releases="Latest", include_urls=TRUE);

## List latest available CPI Time Series Spreadsheet tables only
cpi_tables <- abs_cat_tables("6401.0", releases="Latest", types="tss");
cpi_tables_url <- abs_cat_tables("5206.0", releases="Latest", types="tss", include_urls=TRUE);

## List latest available ASGS Volume 3 Data Cubes
asgs_vol3_tables <- abs_cat_tables("1270.0.55.003", releases="Latest", types="css");
asgs_vol3_tables_url <- abs_cat_tables("1270.0.55.003", releases="Latest",
  types="css", include_urls=TRUE);
```

```
## List latest available ASGS ANZSIC publications (PDF) files
anzsic_2006 <- abs_cat_tables("1292.0", releases="Latest", types="pub", include_urls=TRUE);
```

abs_cat_unzip *Uncompress locally-stored ABS Catalogue data file archives*

Description

Function to uncompress locally-stored ABS Catalogue data file archives

Usage

```
abs_cat_unzip(files, exdir)
```

Arguments

files	One or more local zip files.
exdir	Target directory for extracted archive files. Directory is created if it doesn't exist. If missing, creates a new subdirectory in tempdir() using the respective zip files (specified in files).

Value

Returns a character vector listing the names of all files extracted.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

abs_datasets *Download updated data series information from the ABS API*

Description

TBC

Usage

```
abs_datasets(lang = "en", include_notes = FALSE)
```

Arguments

lang	Preferred language (default 'en' - English).
include_notes	Include ABS annotation information for each series.

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
datasets <- abs_datasets()
datasets <- abs_datasets(include_notes=TRUE)
```

abs_dimensions	<i>Return available dimensions of ABS series</i>
----------------	--

Description

This function returns the available dimensions for a specified ABS API dataset.

Usage

```
abs_dimensions(dataset, update_cache = FALSE)
```

Arguments

dataset	Character vector of dataset codes. These codes correspond to the indicatorID column from the indicator data frame of abs_cache or abs_cachelist, or the result of abs_indicators.
update_cache	Logical expression, if FALSE (default), use the cached list of available ABS.Stat datasets, if TRUE, update the list of available datasets.

Value

a data frame with available dataset dimensions.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## CPI - Consumer Price Index
x <- abs_dimensions("CPI");
str(x)
## LF - Labour Force
x <- abs_dimensions("LF");
str(x)
```

abs_metadata

Download updated data series information from the ABS API

Description

TBC

Usage

```
abs_metadata(id, lang = "en")
```

Arguments

id	ABS dataset ID.
lang	Preferred language (default 'en' - English).

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
datasets <- abs_datasets();
x <- abs_metadata("CPI");
x <- abs_metadata(grep("cpi", datasets$id, ignore.case=TRUE, value=TRUE));
names(x)
y <- abs_metadata(datasets$id[1]);
names(y)
```

abs_read_tss	<i>Read ABS time series data file(s)</i>
--------------	--

Description

This function extracts time series data from ABS data files.

Usage

```
abs_read_tss(files, type = "tss")
```

Arguments

files	Names of one or more ABS data files
type	One of either 'tss' – ABS Time Series Spreadsheet (the Default) or 'css' – Data Cube.R

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## Read specified ABS Excel time series files
tables <- abs_cat_tables("5206.0", releases="Latest", include_urls=TRUE);
downloaded_tables <- abs_cat_download(tables$path_2[1], exdir=tempdir())
extracted_files <- abs_cat_unzip(downloaded_tables)
x <- abs_read_tss(extracted_files);
```

abs_search	<i>Search dataset information from the ABS.Stat API</i>
------------	---

Description

This function finds datasets or dimensions within a specific that match a specified regular expression and returns matching results.

Usage

```
abs_search(pattern, dataset = NULL, ignore.case = TRUE,
           code_only = FALSE, update_cache = FALSE)
```

Arguments

pattern	Character string or regular expression to be matched.
dataset	Character vector of ABS.Stat dataset codes. These codes correspond to the indicatorID column from the indicator data frame of abs_cache or abs_cachelist, or the result of abs_indicators. If NULL (default), then function undertakes a dataset mode search. If not NULL, function searches all dimensions of specified dataset.
ignore.case	Case sensitive pattern match or not.
code_only	If FALSE (default), all column/fields are returned. If TRUE, only the dataset identifier or indicator code are returned.
update_cache	Logical expression, if FALSE (default), use the cached list of available ABS.Stat datasets, if TRUE, update the list of available datasets.

Value

A data frame with datasets and data items that match the search pattern.

Note

With acknowledgements to wb_search function.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## ABS dataset search
x <- abs_search(pattern = "consumer price index")
x <- abs_search(pattern = "census")
x <- abs_search(pattern = "labour force")

## ABS indicator search
x <- abs_search(pattern = "all groups", dataset="CPI")
x <- abs_search(pattern = c("all groups", "capital cities"), dataset="CPI")
```

abs_stats

Download data from the ABS API

Description

This function downloads the specified ABS data series from the ABS API.

Usage

```
abs_stats(dataset, filter, start_date, end_date, lang = c("en", "fr"),
  dimensionAtObservation = c("AllDimensions", "TimeDimension",
    "MeasureDimension"), detail = c("Full", "DataOnly", "SeriesKeysOnly",
    "NoData"), enforce_api_limits = TRUE, return_url = FALSE,
  update_cache = FALSE)
```

Arguments

dataset	Character vector of ABS.Stat dataset codes. These codes correspond to the indicatorID column from the indicator data frame of abs_cache or abs_cachelist, or the result of abs_indicators.
filter	A list that contains filter of dimensions available in the specified series to use in the API call. If NULL, no filter is set and the query tries to return all dimensions of the dataset. Valid dimensions to include in the list supplied to filter include: MEASURE, REGION, INDEX, TSEST and FREQUENCY.
start_date	Numeric or character. If numeric it must be in year). For data at the sub-annual granularity the API supports a format as follows: Monthly data – '2016-M01', Quarterly data – '2016-Q1', Semi-annual data – '2016-B2', Financial year data – '2016-17'.
end_date	Numeric or character (refer to startdate).
lang	Language in which to return the results. If lang is unspecified, english is the default.
dimensionAtObservation	The identifier of the dimension to be attached at the observation level. The default order is: 'AllDimensions', 'TimeDimension' and 'MeasureDimension'. AllDimensions results in a flat list of observations without any grouping.
detail	This argument specifies the desired amount of information to be returned. Possible values are: <ul style="list-style-type: none"> • Full: all data and documentation, including annotations (default) • DataOnly: attributes – and therefore groups – will be excluded • SeriesKeysOnly: only the series elements and the dimensions that make up the series keys • NoData: returns the groups and series, including attributes and annotations, without observations (all values = NA)
enforce_api_limits	If TRUE (the default), the function enforces the ABS.Stat RESTful API limits and will not submit the query if the URL string length exceeds 1000 characters or the query would return more than 1 million records. If FALSE, the function submits the API call regardless and attempts to return the results.
return_url	Default is FALSE. If TRUE, the function returns the generated request URL and does not submit the request.
update_cache	Logical expression, if FALSE (default), use the cached list of available ABS.Stat datasets, if TRUE, update the list of available datasets.

Value

Returns a data frame of the selected series from the specified ABS dataset.

Note

The data query submitted by this function uses the ABS RESTful API based on the SDMX-JSON standard. It has a maximum allowable character limit of 1000 characters allowed in the data URL.

Further limitations known at this time include:

- Only anonymous queries are supported, there is no authentication
- Each response is limited to no more than 1 million observations
- Errors are not returned in the JSON format but HTTP status codes and messages are set according to the Web Services Guidelines
- The lastNObservations parameter is not supported
- Observations follow the time series (or import-specific) order even if dimensionAtObservation=AllDimensions is used.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
x <- abs_stats(dataset="CPI", filter="all", return_url=TRUE);
x <- abs_stats(dataset="CPI", filter=list(MEASURE=1, REGION=c(1:8,50),
                                         INDEX=10001, TSEST=10, FREQUENCY="Q"));
x <- abs_stats(dataset="CPI", filter=list(MEASURE="all", REGION=50,
                                         INDEX=10001, TSEST=10, FREQUENCY="Q"));
x <- abs_stats(dataset="CPI", filter=list(MEASURE="all", REGION=50, INDEX=10001,
                                         TSEST=10, FREQUENCY="Q"), return_url=TRUE);
```

aus_state_codes

Table of Australian state and territory codes

Description

A list of Australian state and territory codes.

Usage

```
aus_state_codes
```

Format

A data frame containing three columns:

- `state_code` One-digit state code.
- `state_abb` State/territory abbreviation.
- `state_name` State/territory name.

raustats	<i>raustats: An R package for accessing data and statistics from the ABS and RBA websites</i>
----------	---

Description

The raustats package provides structured access to all data and statistics available from the Australian Bureau of Statistics and Reserve Bank of Australia website, as well as draft access to the ABS.Stat - Beta data catalogue API.

Details

To learn more about the raustats package, start with the vignettes: `browseVignettes(package = "raustats")`

rba_cachelist	<i>Cached list of statistical tables provided by the RBA</i>
---------------	--

Description

This data is a cached result of the `rba_table_cache` function. By default functions `rba_search` and `rba_stats` use this data if the `update_cache` parameter is TRUE.

Usage

```
rba_cachelist
```

Format

A data frame containing three columns:

- `table_code` RBA table code.
- `table_name` RBA table name.
- `table_type` One of either current statistical tables, historical data or discontinued data
- `url` RBA URL

rba_file_download	<i>Function to download statistics files from the RBA website and store locally</i>
-------------------	---

Description

This function downloads one or more RBA data files at the specified by URLs and saves a local copy.

Usage

```
rba_file_download(data_url, exdir = tempdir(), update_cache = TRUE)
```

Arguments

data_url	Character vector specifying an RBA data set URL.
exdir	Target directory for downloaded files (defaults to tempdir()). Directory is created if it doesn't exist.
update_cache	Logical expression, if FALSE (default), use the cached list of available RBA datasets, if TRUE, update the list of available datasets.

Value

Downloads data from the ABS website and returns a character vector listing the location where files are saved.

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

rba_read_tss	<i>Read RBA statistical time series spreadsheet</i>
--------------	---

Description

Functions to extract data from a specified RBA time series spreadsheet.

Usage

```
rba_read_tss(files)
```

Arguments

files	Names of one or more ABS data file
-------	------------------------------------

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
rba_urls <- rba_search(pattern = "Liabilities and Assets")$url
rba_files <- sapply(rba_urls, rba_file_download)
data <- rba_read_tss(rba_files);
```

rba_search

Return list of data tables from RBA website

Description

Function to return a list of all RBA data tables.

Usage

```
rba_search(pattern, fields = c("table_no", "table_name"),
  ignore.case = TRUE, update_cache = FALSE)
```

Arguments

pattern	Character string or regular expression to be matched
fields	Character vector of column names through which to search. By default, the function searches 'table_no' and 'table_name'.
ignore.case	Case sensitive pattern match or not.
update_cache	Logical expression, if FALSE (default), use the cached list of available RBA tables (rba_cachelist), if TRUE, update the list of available datasets.

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
rba_datasets <- rba_search(pattern = "Liabilities and Assets");
```

rba_stats	<i>Return data for a specified RBA time series</i>
-----------	--

Description

Function to download and return specified RBA time series data.

Usage

```
rba_stats(table_no, pattern, url, update_cache = FALSE, ...)
```

Arguments

table_no	Character vector specifying one or more RBA table numbers to download.
pattern	Character string or regular expression to be matched.
url	Valid URL for RBA dataset (Excel format only).
update_cache	Logical expression, if FALSE (default), use the cached list of available RBA datasets, if TRUE, update the list of available datasets.
...	Other arguments to rba_search

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
## Example - Selecting by table_no  
x <- rba_stats("A1");  
  
## Example - Selecting by pattern  
x <- rba_stats(pattern="Liabilities and Assets");
```

rba_table_cache	<i>Return list of RBA tables</i>
-----------------	----------------------------------

Description

Function to return an updated list of data tables available from the RBA website.

Usage

```
rba_table_cache()
```

Value

data frame in long format

Author(s)

David Mitchell <david.pk.mitchell@gmail.com>

Examples

```
rba_cachelist <- rba_table_cache();
```

Index

*Topic **datasets**

- [abs_cachelist](#), [2](#)
- [abs_cat_cachelist](#), [3](#)
- [aus_state_codes](#), [12](#)
- [rba_cachelist](#), [13](#)

- [abs_cachelist](#), [2](#)
- [abs_cat_cachelist](#), [3](#)
- [abs_cat_download](#), [3](#)
- [abs_cat_stats](#), [4](#)
- [abs_cat_tables](#), [5](#)
- [abs_cat_unzip](#), [6](#)
- [abs_datasets](#), [6](#)
- [abs_dimensions](#), [7](#)
- [abs_metadata](#), [8](#)
- [abs_read_tss](#), [9](#)
- [abs_search](#), [9](#)
- [abs_stats](#), [10](#)
- [aus_state_codes](#), [12](#)

- [raustats](#), [13](#)
- [raustats-package \(raustats\)](#), [13](#)
- [rba_cachelist](#), [13](#)
- [rba_file_download](#), [14](#)
- [rba_read_tss](#), [14](#)
- [rba_search](#), [13](#), [15](#), [16](#)
- [rba_stats](#), [13](#), [16](#)
- [rba_table_cache](#), [13](#), [17](#)