Package 'rclimateca'

March 25, 2018

Type Package

Title Fetch Climate Data from Environment Canada

Version 1.0.1

Author Dewey Dunnington <dewey@fishandwhistle.net>

Maintainer Dewey Dunnington <dewey@fishandwhistle.net>

Description The Environment Canada climate archives <http://climate.weather.gc.ca/> are an important source of data for climate researchers in Canada and world wide. The repository contains temperature, precipitation, and wind data for more than 8,000 locations. The functions in this package simplify the process of downloading, subsetting, and manipulating these data for the purposes of more efficient workflows in climate research.

License GPL-3

Depends R (>= 2.10)

Imports digest, httr, prettymapr, reshape2, lubridate, mudata2, dplyr(>= 0.7.0), stringr, readr, tibble, purrr, tidyr(>= 0.7.0), magrittr, rlang

URL https://github.com/paleolimbot/rclimateca

BugReports https://github.com/paleolimbot/rclimateca/issues

Suggests ggplot2, testthat, covr, knitr, rmarkdown

LazyData TRUE

RoxygenNote 6.0.1

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2018-03-25 18:16:53 UTC

R topics documented:

`as_ec_climate_location` *Environment Canada Historical Climate Locations*

Description

Environment Canada Historical Climate Locations

Usage

```
as_ec_climate_location(x, ...)

## S3 method for class 'ec_climate_location'
as_ec_climate_location(x, ...)

## S3 method for class 'double'
as_ec_climate_location(x, ...)

## S3 method for class 'integer'
as_ec_climate_location(x, ...)

## S3 method for class 'character'
as_ec_climate_location(x, ...)

is_ec_climate_location(x)

new_ec_climate_location(x)
```

```
validate_ec_climate_location(x)

## S3 method for class 'ec_climate_location'
as_tibble(x, ...)

## S3 method for class 'ec_climate_location'
as.data.frame(x, ...)

## S3 method for class 'ec_climate_location'
print(x, ...)

## S3 method for class 'ec_climate_location'
as.character(x, ...)

## S3 method for class 'ec_climate_location'
as.integer(x, ...)

## S3 method for class 'ec_climate_location'
as.numeric(x, ...)

## S3 method for class 'ec_climate_location'
x[i, ...]
```

Arguments

x	A vector to convert to EC historical climate locations
...	Not used in these functions
i	Used to subset an EC historical climate location vector

Value

An object of type ec_climate_location

`clear_cache` *Cache options*

Description

Clears the local cache of downloaded files (by default in the ec.cache folder in the working directory).

Usage

```
clear_cache(cache = "ec.cache")

set_default_cache(cache = "ec.cache")

get_default_cache()
```

Arguments

cache	A folder name where cached files should be stored.

Examples

```
clear_cache()
```

`climatelong` *DEPRECATED: Transform EC data to long format*

Description

DEPRECATED: Transform EC data to long format

Usage

```
climatelong(df, rm.na = FALSE)
```

Arguments

df	A wide data frame (obtained from getClimateData or getClimateDataRaw)
rm.na	Flag to remove rows with an empty value. This may help compress large datasets.

Value

A melted `data.frame` (see reshape2::melt)

`ecclimatelocs` *DEPRECATED: Deprecated climate locations (February 2017)*

Description

Climate locations for Environment Canada, as of February 2017. This object is available for historical reasons, and will be removed in future versions. Instead, use ec_climate_locations_all.

Usage

```
ecclimatelocs
```

Format

A data frame with 8735 rows and 19 variables. There are many columns, only several of which are used within this package.

Name the name of the location (in all caps)

Province the province containing the location (in all caps)

Climate ID IDs that may be used outside of EC

Station ID the ID to be used in getClimateData and getClimateDataRaw

WMO ID IDs that may be used outside of EC

TC ID IDs that may be used outside of EC

Latitude (Decimal Degrees) the latitude of the site

Longitude (Decimal Degrees) the longitude of the site

Latitude integer representation of the latitude

Longitude integer representation of the longitude

Elevation (m) The elevation of the site (in metres)

First Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

Last Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

MLY First Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

MLY Last Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

DLY First Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

DLY Last Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

HLY First Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

HLY Last Year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

Source

ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/

`ec_climate_data` *Load Environment Canada Historical Climate Data*

Description

Load Environment Canada Historical Climate Data

Usage

```
ec_climate_data(location, timeframe = c("monthly", "daily", "hourly"),
  start = NA, end = NA, value_parser = readr::parse_double,
  cache = get_default_cache(), quiet = TRUE)

ec_climate_mudata(location, timeframe = c("monthly", "daily", "hourly"),
  start = NA, end = NA, cache = get_default_cache(), quiet = TRUE)
```

Arguments

`location`	A vector of unambiguous name identifiers or station IDs (resolved using as_ec_climate_location).
`timeframe`	One of monthly, daily, or hourly.
`start`	The first date to be included in the output as a Date object or in YYYY-MM-DD format (passed through as.Date)
`end`	The last date to be included in the output as a Date object or in YYYY-MM-DD format (passed through as.Date)
`value_parser`	A readr parse function (like parse_double or parse_character) to apply to value columns. The default is to use parse_double, but occasionally values are in the form ">30", or "<30", especially for wind speed. When this happens a warning will occur, and problems() can be used to see which values were dropped. Use parse_character) to skip parsing and extract the values yourself.
`cache`	A directory in which to cache downloaded files
`quiet`	Use FALSE for verbose output

Value

A data.frame (tibble) with an attribute "flag_info", containing the flag information. `ec_climate_mudata()` returns a mudata object.

References

http://climate.weather.gc.ca/historical_data/search_historic_data_e.html ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt http://climate.weather.gc.ca/glossary_e.html

Examples

```
# station 27141 is Kentville CDA CS
monthly <- ec_climate_data(27141, timeframe = "monthly")
daily <- ec_climate_data(27141, timeframe = "daily", start = "1999-01-01", end = "1999-12-31")
hourly <- ec_climate_data(27141, timeframe = "hourly", start = "1999-07-01", end = "1999-07-31")

# get climate data in mudata format
library(mudata2)
monthly_md <- ec_climate_mudata(27141, timeframe = "monthly")
daily_md <- ec_climate_mudata(27141, timeframe = "daily",
                              start = "1999-01-01", end = "1999-12-31")
hourly_md <- ec_climate_mudata(27141, timeframe = "hourly",
                               start = "1999-07-01", end = "1999-07-31")

# mudata objects can easily be plotted
autoplot(monthly_md)
autoplot(daily_md)
autoplot(hourly_md)
```

---

ec_climate_data_base *Low-level access to the EC Climate Bulk Data Service*

Description

Low-level access to the EC Climate Bulk Data Service

Usage

```
ec_climate_data_base(location, timeframe = c("monthly", "daily", "hourly"),
  year = NULL, month = NULL, cache = NULL, quiet = FALSE,
  check_dates = FALSE,
  endpoint = "http://climate.weather.gc.ca/climate_data/bulk_data_e.html")
```

Arguments

location	A vector of unambiguous name identifiers or station IDs (resolved using as_ec_climate_location).
timeframe	One of monthly, daily, or hourly.
year	The year for which to get data (required for daily requests)
month	The month for which to get data (required for hourly requests)
cache	A directory in which to cache downloaded files
quiet	Use FALSE for verbose output
check_dates	Check the request data against ec_climate_locations_all to avoid downloading data that is known not to exist. Pass FALSE to circumvent this check.
endpoint	The web address for the EC data service

Value

A data.frame (tibble) of the downloaded data frame, with all columns as character vectors

References

http://climate.weather.gc.ca/historical_data/search_historic_data_e.html ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt

Examples

```
# station 27141 is Kentville CDA CS
monthly <- ec_climate_data_base(27141, timeframe = "monthly")
daily <- ec_climate_data_base(27141, timeframe = "daily", year = 1999)
hourly <- ec_climate_data_base(27141, timeframe = "hourly", year = 1999, month = 7)
```

`ec_climate_data_read_csv`

Read a historical climate data CSV file

Description

Read a historical climate data CSV file

Usage

```
ec_climate_data_read_csv(path)
```

Arguments

`path`	A path to an (unmodified) CSV file downloaded from the EC service

Value

A tibble with an attribute "flag_info", containing the flag information.

`ec_climate_locations_all`
Climate locations (January 2018)

Description

Climate locations for Environment Canada, as of January 2018, with column names as nice_names and a unique string identifier added.

Usage

```
ec_climate_locations_all
```

Format

A data frame with 8735 rows and 19 variables.

location A unique string identifier for each location, consisting of the name, province, and station_id

latitude latitude of the site (decimal degrees)

longitude longitude of the site (decimal degrees)

timezone_id The timezone ID of the site

lst_utc_offset The offset of local standard time to UTC (estimated)

station_id The primary identifier of the site used by Environment Canada

name the name of the location (in all caps)

province the province containing the location (in all caps)

climate_id IDs that may be used outside of EC

wmo_id IDs that may be used outside of EC

tc_id IDs that may be used outside of EC

elevation_m The elevation of the site (in metres)

first_year The first year where data exists for this location (for any resolution)

last_year The last year where data exists for this location (for any resolution)

mly_first_year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

mly_last_year The last year where data exists for this location (for MLY, DLY, or HLY resolution)

dly_first_year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

dly_last_year The last year where data exists for this location (for MLY, DLY, or HLY resolution)

hly_first_year The first year where data exists for this location (for MLY, DLY, or HLY resolution)

hly_last_year The last year where data exists for this location (for MLY, DLY, or HLY resolution)

Source

ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/

ec_climate_long *Transform EC Climate Data to Parameter-Long Form*

Description

Transform EC Climate Data to Parameter-Long Form

Usage

```
ec_climate_long(climate_df, na.rm = FALSE)
```

Arguments

climate_df	The outuput of ec_climate_data
na.rm	TRUE to remove measurements for which there is no information

Value

A data.frame (tibble) with one row per measurement

Examples

```
# station 27141 is Kentville CDA CS
monthly <- ec_climate_data(27141, timeframe = "monthly")
ec_climate_long(monthly)

# or use the pipe
ec_climate_data(27141, timeframe = "monthly") %>%
  ec_climate_long()
```

ec_climate_params_all *Climate parameters (January 2018)*

Description

Climate parameters for Environment Canada datasets, as of January 2018.

Usage

```
ec_climate_params_all
```

Format

A data frame with 32 rows and 6 variables.

dataset The dataset name (ec_climate)

param The parameter identifier

nice_label The column name from the output of ec_climate_data

label The label for the parameter (with unit)

flag_label The raw column name used for the Flag parameter

unit The unit of measurement for the parameter

`ec_climate_search_locations`
Search climate locations

Description

Search climate locations

Usage

```
ec_climate_search_locations(query = NULL, ..., timeframe = c("NA",
  "monthly", "daily", "hourly"), year = NULL, limit = NULL)

ec_climate_geosearch_locations(query = NULL, ..., timeframe = c("NA",
  "monthly", "daily", "hourly"), year = NULL, limit = NULL)
```

Arguments

`query`	A query in several forms
`...`	Additional arguments are used to filter ec_climate_locations_all
`timeframe`	The target timeframe for the query
`year`	An optional year when the location must have data
`limit`	The maximum number of locations to return (or NULL for no limit). Lon/lat queries are automatically capped at 30 locations.

Examples

```
# character searches match the location column of ec_climate_locations_all
# (case-insensitive)
ec_climate_search_locations("ottawa")

# multiple values use OR logic
ec_climate_search_locations(c("ottawa on", "halifax"))
```

```
# you can use a year and a timeframe to find locations that are known to have some
# data for that year/timeframe
ec_climate_search_locations("ottawa", year = 2016)
ec_climate_search_locations("ottawa", timeframe = "daily", year = 2016)

# you can also use a vector of years
ec_climate_search_locations("ottawa", timeframe = "daily", year = 2000:2016)

# if you need to search geographically, you can pass a numeric vector in the form
# c(lon, lat)
ec_climate_search_locations(c(-75.69031, 45.42111))


# to use a human readable geocoded location, use ec_climate_geosearch_locations()
ec_climate_geosearch_locations("ottawa on")
```

`getClimateData` *DEPRECATED: Get Parsed and aggregated Environment Canada data*

Description

Use this function to get Environment Canada climate data in bulk over multiple years and/or stations.

Usage

```
getClimateData(stationID, timeframe = c("monthly", "daily", "hourly"),
  year = NULL, month = NULL, day = NULL, cache = "ec.cache",
  quiet = TRUE, progress = c("text", "none", "tk"), format = c("wide",
  "long"), rm.na = FALSE, parsedates = TRUE, checkdates = TRUE,
  nicenames = FALSE, ply = plyr::adply)

getClimateMUData(stationID, timeframe = c("monthly", "daily", "hourly"),
  year = NULL, month = NULL, day = NULL, cache = "ec.cache",
  quiet = TRUE, progress = c("text", "none", "tk"), rm.na = FALSE,
  dataset.id = "ecclimate")
```

Arguments

`stationID`	The station ID, possibly found by getClimateSites
`timeframe`	One of "monthly", "hourly", or "daily"
`year`	A vector of years for which to fetch data
`month`	A vector of months for which to fetch data
`day`	A vector of day numbers for which to fetch data

cache	A folder in which to cache downloaded files
quiet	Suppress update messages
progress	A plyr progress bar, one of "text", "tk", or "none"
format	One of "wide" or "long". Use long for easy integration with ggplot.
rm.na	Pass rm.na=TRUE to remove rows with no values. This may help compress large datasets
parsedates	Flag to parse date/time information (useful for plotting).
checkdates	If checkdates is TRUE, the loop will not attempt to download if a year is marked as missing in ecclimatelocs. Note that this information may be out of date, but this flag is useful to minimize the amount of downloading that needs to occur. This will also subset the resulting data frame to only contain the years/months requested.
nicenames	Use lower-case, unit-free names for columns.
ply	The plyr-like function that executes the loop and returns the result. Pass your own function accepting named arguments .data, .margins=1, .fun=function(row), .progress. This may be useful if all you need to do is extract information out of a large amount of climate data without the need to store it to disk.
dataset.id	The dataset identifier to use for mudata creation.

Value

A data.frame (or the results of ply if passed).

References

http://climate.weather.gc.ca/historical_data/search_historic_data_e.html ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt

---

getClimateDataRaw *DEPRECATED: Get parsed CSV data from Environment Canada*

Description

This function just downloads a .csv file from the bulk data service from Environment Canda. It follows as closely as possible the EC specifications, and does not modify the result except to remove the header information. To apply this function over multiple months/stations/years/months, use getClimateData.

Usage

```
getClimateDataRaw(stationID, timeframe = c("monthly", "daily", "hourly"),
  Year = NA, Month = NA,
  endpoint = "http://climate.weather.gc.ca/climate_data/bulk_data_e.html",
  flag.info = FALSE, ...)
```

Arguments

stationID	A stationID (you could find this using getClimateSites)
timeframe	One of "montly" "daily" or "hourly"
Year	The year for which to fetch the data
Month	The month for which to fetch the data
endpoint	The url from which to fetch data (in case this changes in the future)
flag.info	Pass TRUE to get a list with elements $data and $flags
...	further arguments passed on to the downloading function

Value

A data.frame of results, or a list if flag.info=TRUE

References

http://climate.weather.gc.ca/historical_data/search_historic_data_e.html ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt

---

getClimateSites *Get Environment Canada locations.*

Description

Get Environment Canada locations.

Usage

```
getClimateSites(location, year = NULL, n = 5, locs = NULL,
  nicenames = FALSE, cols = c("Name", "Province", "Station ID", "distance",
  "Latitude (Decimal Degrees)", "Longitude (Decimal Degrees)", "First Year",
  "Last Year"))
```

Arguments

location	A human-readable location that will be geocoded
year	A vector of years that the location must have data for
n	The number of rows to return
locs	The data.frame of locations. Use NULL to for ecclimatelocs.
nicenames	Sanitize names to type-able form
cols	The columns to return (use NULL for all columns)

Value

A subset of ecclimatelocs

References

ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/Readme.txt
ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/

`print.ec_climate_location_search` *Print climate location search results*

Description

Print climate location search results

Usage

```
## S3 method for class 'ec_climate_location_search'
print(x, limit = 20, ...)
```

Arguments

`x`	A climate location search result from ec_climate_search_locations or ec_climate_geosearch_locations.
`limit`	The number of results to show (use NULL for no limit)
`...`	Not used.

Value

The input, invisibly.

`rclimateca` *Environment Canada data in R*

Description

Environment Canada data in R

Index