

Package ‘weathercan’

October 8, 2018

Type Package

Title Download Weather Data from the Environment and Climate Change Canada Website

Version 0.2.8

Description Provides means for downloading historical weather data from the Environment and Climate Change Canada website (<http://climate.weather.gc.ca/historical_data/search_historic_data_e.html>). Data can be downloaded from multiple stations and over large date ranges and automatically processed into a single dataset. Tools are also provided to identify stations either by name or proximity to a location.

License GPL-3

Language en-CA

BugReports <https://github.com/ropensci/weathercan/issues>

LazyData TRUE

URL <https://github.com/ropensci/weathercan>

Depends R (>= 3.1.2)

Imports dplyr (>= 0.7.0), httr (>= 1.1.0), lubridate (>= 1.7.1), methods (>= 3.2.2), purrr (>= 0.2.4), rlang (>= 0.1.4), stringi (>= 1.1.2), tidyr (>= 0.4.1), xml2 (>= 0.1.2)

RoxygenNote 6.1.0

Suggests devtools, ggplot2, ggmap, knitr, lutz, rmarkdown, sf, sp, stringr, testthat

VignetteBuilder knitr

Encoding UTF-8

NeedsCompilation no

Author Steffi LaZerte [aut, cre] (<<https://orcid.org/0000-0002-7690-8360>>), Sam Albers [ctb] (<<https://orcid.org/0000-0002-9270-7884>>)

Maintainer Steffi LaZerte <steffi@steffi.ca>

Repository CRAN

Date/Publication 2018-10-08 20:30:02 UTC

R topics documented:

weathercan-package	2
finches	3
flags	4
get_tz	4
glossary	4
kamloops	5
kamloops_day	6
pg	8
stations	9
stations_dl	10
stations_search	11
tz_calc	12
weather_dl	12
weather_interp	14
Index	16

weathercan-package	<i>Easy downloading of weather data from Environment and Climate Change Canada</i>
--------------------	--

Description

weathercan is an R package for simplifying the downloading of Historical Climate Data from the Environment and Climate Change Canada (ECCC) website (http://climate.weather.gc.ca/historical_data/search_historic_data)

Details

Bear in mind that these downloads can be fairly large and performing multiple, downloads may use up Environment Canada's bandwidth unnecessarily. Try to stick to what you need.

There are three main aspects of this package:

1. Access **stations** lists
 - `stations` (a data frame listing stations)
 - `stations_search()` identify stations by name or proximity to a location
 - `stations_dl()` re-download stations data
2. Download **weather** data
 - `weather_dl()`
3. Merge **weather** data into other data sets through interpolation over time
 - `weather_interp()`

We also include several practice data sets:

- `finches`

- [kamloops](#)
- [kamloops_day](#)
- [pg](#)

As well as several vignettes:

- General Usage: `vignette("usage")`
- Merging and Interpolating: `vignette("interpolation")`
- Data Flags: `vignette("flags")`
- Data Glossary: `vignette("glossary")`

Online we also have an advanced article:

- Using weathercan with [tidyverse](#) ([here](#))

References

Environment and Climate Change Canada: <https://www.ec.gc.ca/>
Glossary of terms http://climate.weather.gc.ca/glossary_e.html
ECCC Historical Climate Data: <http://climate.weather.gc.ca/>

finches

RFID Data on finch visits to feeders

Description

RFID Data on finch visits to feeders

Usage

finches

Format

An example dataset of finch RFID data for interpolation:

bird_id Bird ID number

time Time

feeder_id feeder ID

species Species

lat Latitude of station location in degree decimal format

lon Longitude of station location in degree decimal format

flags *Meaning of coded 'flags'*

Description

A reference dataset containing 'flags' matched to their meaning. Data downloaded using the `weather_dl()` function contains columns indicating 'flags' these codes are presented here for interpretation.

Usage

flags

Format

A data frame with 16 rows and 2 variables:

code Flag code

meaning Explanation of the code

get_tz *DEFUNCT: Get timezone from lat/lon*

Description

DEFUNCT: Get timezone from lat/lon

Usage

get_tz()

glossary *Glossary of units and terms*

Description

A reference dataset matching information on columns in data downloaded using the `weather_dl()` function. Indicates the units of the data, and contains a link to the ECCC glossary page explaining the measurement.

Usage

glossary

Format

A data frame with 77 rows and 5 variables:

interval Data interval type, 'hour', 'day', or 'month'.

ECCC_name Original column name when downloaded directly from ECCC

weathercan_name R-compatible name given when downloaded with the `weather_dl()` function using the default argument `format = TRUE`.

units Units of the measurement.

ECCC_ref Link to the glossary or reference page on the ECCC website.

kamloops

Hourly weather data for Kamloops

Description

Downloaded with `weather()`. Terms are more thoroughly defined here http://climate.weather.gc.ca/glossary_e.html

Usage

kamloops

Format

An example dataset of hourly weather data for Kamloops:

station_name Station name

station_id Environment Canada's station ID number. Required for downloading station data.

prov Province

lat Latitude of station location in degree decimal format

lon Longitude of station location in degree decimal format

date Date

time Time

year Year

month Month

day Day

hour Hour

qual Data quality

weather The state of the atmosphere at a specific time.

hmdx Humidex

hmdx_flag Humidex data flag

pressure Pressure (kPa)
pressure_flag Pressure data flag
rel_hum Relative humidity
rel_hum_flag Relative humidity data flag
temp Temperature
temp_dew Dew Point Temperature
temp_dew_flag Dew Point Temperature flag
visib Visibility (km)
visib_flag Visibility data flag
wind_chill Wind Chill
wind_chill_flag Wind Chill flag
wind_dir Wind Direction (10's of degrees)
wind_dir_flag wind Direction Flag
wind_spd Wind speed km/hr
wind_spd_flag Wind speed flag
elev Elevation (m)
climate_id Climate identifier
WMO_id World Meteorological Organization Identifier
TC_id Transport Canada Identifier

Source

http://climate.weather.gc.ca/index_e.html

kamloops_day

Daily weather data for Kamloops

Description

Downloaded with `weather()`. Terms are more thoroughly defined here http://climate.weather.gc.ca/glossary_e.html

Usage

kamloops_day

Format

An example dataset of daily weather data for Kamloops:

station_name Station name

station_id Environment Canada's station ID number. Required for downloading station data.

prov Province

lat Latitude of station location in degree decimal format

lon Longitude of station location in degree decimal format

date Date

year Year

month Month

day Day

cool_deg_days Cool degree days

cool_deg_days_flag Cool degree days flag

dir_max_gust Direction of max wind gust

dir_max_gust_flag Direction of max wind gust flag

heat_deg_days Heat degree days

heat_deg_days_flag Heat degree days flag

max_temp Maximum temperature

max_temp_flag Maximum temperature flag

mean_temp Mean temperature

mean_temp_flag Mean temperature flag

min_temp Minimum temperature

min_temp_flag Minimum temperature flag

snow_grnd Snow on the ground (cm)

snow_grnd_flag Snow on the ground flag

spd_max_gust Speed of the max gust km/h

spd_max_gust_flag Speed of the max gust flag

total_precip Total precipitation (any form)

total_precip_flag Total precipitation flag

total_rain Total rain (any form)

total_rain_flag Total rain flag

total_snow Total snow (any form)

total_snow_flag Total snow flag

elev Elevation (m)

climate_id Climate identifier

WMO_id World Meteorological Organization Identifier

TC_id Transport Canada Identifier

Source

http://climate.weather.gc.ca/index_e.html

pg

Hourly weather data for Prince George

Description

Downloaded with `weather()`. Terms are more thoroughly defined here http://climate.weather.gc.ca/glossary_e.html

Usage

pg

Format

An example dataset of hourly weather data for Prince George:

station_name Station name

station_id Environment Canada's station ID number. Required for downloading station data.

prov Province

lat Latitude of station location in degree decimal format

lon Longitude of station location in degree decimal format

date Date

time Time

year Year

month Month

day Day

hour Hour

qual Data quality

weather The state of the atmosphere at a specific time.

hmdx Humidex

hmdx_flag Humidex data flag

pressure Pressure (kPa)

pressure_flag Pressure data flag

rel_hum Relative humidity

rel_hum_flag Relative humidity data flag

temp Temperature

temp_dew Dew Point Temperature

temp_dew_flag Dew Point Temperature flag

visib Visibility (km)

visib_flag Visibility data flag

wind_chill Wind Chill
wind_chill_flag Wind Chill flag
wind_dir Wind Direction (10's of degrees)
wind_dir_flag wind Direction Flag
wind_spd Wind speed km/hr
wind_spd_flag Wind speed flag
elev Elevation (m)
climate_id Climate identifier
WMO_id World Meteorological Organization Identifier
TC_id Transport Canada Identifier

Source

http://climate.weather.gc.ca/index_e.html

stations	<i>Station data downloaded from Environment and Climate Change Canada</i>
----------	---

Description

A dataset containing station information downloaded from Environment and Climate Change Canada. Note that a station may have several station IDs, depending on how the data collection has changed over the years. Station information can be updated by running `stations_new <- stations_dl()` and then by specifying `stn = stations_new` in most functions.

Usage

stations

Format

A data frame with 26211 rows and 12 variables:

prov Province
station_name Station name
station_id Environment Canada's station ID number. Required for downloading station data.
climate_id Climate ID number
WMO_id Climate ID number
TC_id Climate ID number
lat Latitude of station location in degree decimal format
lon Longitude of station location in degree decimal format
elev Elevation of station location in metres

tz Local timezone excluding any Daylight Savings
interval Interval of the data measurements ('hour', 'day', 'month')
start Starting year of data record
end Ending year of data record

Source

http://climate.weather.gc.ca/index_e.html

stations_dl	<i>Get available stations</i>
-------------	-------------------------------

Description

This function can be used to download a Station Inventory CSV file from Environment and Climate Change Canada. This is only necessary if the station you're interested was only recently added. The 'stations' data set included in this package contains station data downloaded when the package was last compiled. This function may take a few minutes to run.

Usage

```
stations_dl(url = NULL, skip = NULL, verbose = FALSE,
            quiet = FALSE)
```

Arguments

url	Character. Url from which to grab the station information (see details)
skip	Numeric. Number of lines to skip at the beginning of the csv. If NULL, automatically derived.
verbose	Logical. Include progress messages
quiet	Logical. Suppress all messages (including messages regarding missing data, etc.)

Details

URL defaults to `ftp://client_climate@ftp.tor.ec.gc.ca/Pub/Get_More_Data_Plus_de_donnees/ Station`

Value

A tibble containing station names, station ID codes and dates of operation

Examples

```
# Update stations data frame
s <- stations_dl()

# Use new data frame to search for stations
stations_search("Winnipeg", stn = s)
```

stations_search	<i>Search for stations by name or location</i>
-----------------	--

Description

Returns stations that match the name provided OR which are within `dist` km of the location provided. This is designed to provide the user with information with which to decide which station to then get weather data from.

Usage

```
stations_search(name = NULL, coords = NULL, dist = 10,
  interval = c("hour", "day", "month"), stn = weathercan::stations,
  verbose = FALSE, quiet = FALSE)
```

Arguments

name	Character. A vector of length 1 or more with text against which to match. Will match station names that contain all components of name, but they can be in different orders and separated by other text.
coords	Numeric. A vector of length 2 with latitude and longitude of a place to match against. Overrides <code>lat</code> and <code>lon</code> if also provided.
dist	Numeric. Match all stations within this many kilometres of the coords.
interval	Character. Return only stations with data at these intervals. Must be any of "hour", "day", "month".
stn	Data frame. The stations data frame to use. Will use the one included in the package unless otherwise specified.
verbose	Logical. Include progress messages
quiet	Logical. Suppress all messages (including messages regarding missing data, etc.)

Details

To search by coordinates, users must make sure they have the [sp](#) package installed.

Value

Returns a subset of the stations data frame which match the search parameters. If the search was by location, an extra column 'distance' shows the distance in kilometres from the location to the station. If no stations are found withing 'dist', the closest 10 stations are returned.

Examples

```
stations_search(name = "Kamloops")
stations_search(name = "Kamloops", interval = "hour")

stations_search(coords = c(53.915495, -122.739379))

loc <- gmap::geocode("Prince George, BC")
stations_search(coords = loc[c("lat", "lon")])
```

 tz_calc

DEFUNCT: Get timezone from lat/lon

Description

Accessed Google API to determine local timezone from coordinates. Defunct as the API is no longer accessible without a Key.

Usage

```
tz_calc()
```

 weather_dl

Download weather data from Environment Canada

Description

Downloads data from Environment Canada for one or more stations. For details and units, see the glossary vignette (vignette("glossary", package = "weathercan")) or the glossary online http://climate.weather.gc.ca/glossary_e.html.

Usage

```
weather_dl(station_ids, start = NULL, end = NULL, interval = "hour",
  trim = TRUE, format = TRUE, string_as = NA, tz_disp = NULL,
  stn = weathercan::stations, url = NULL, encoding = "UTF-8",
  list_col = FALSE, verbose = FALSE, quiet = FALSE)
```

Arguments

station_ids	Numeric/Character. A vector containing the ID(s) of the station(s) you wish to download data from. See the stations data frame or the stations_search function to find IDs.
start	Date/Character. The start date of the data in YYYY-MM-DD format (applies to all station_ids). Defaults to start of range.
end	Date/Character. The end date of the data in YYYY-MM-DD format (applies to all station_ids). Defaults to end of range.
interval	Character. Interval of the data, one of "hour", "day", "month".
trim	Logical. Trim missing values from the start and end of the weather dataframe. Only applies if 'format = TRUE'
format	Logical. If TRUE, formats data for immediate use. If FALSE, returns data exactly as downloaded from Environment and Climate Change Canada. Useful for dealing with changes by Environment Canada to the format of data downloads.
string_as	Character. What value to replace character strings in a numeric measurement with. See Details.
tz_disp	Character. What timezone to display times in (must be one of OlsonNames()).
stn	Data frame. The stations data frame to use. Will use the one included in the package unless otherwise specified.
url	Character. Url from which to grab the weather data. If NULL uses default url (see details)
encoding	Character. Text encoding for download.
list_col	Logical. Return data as nested data set? Defaults to FALSE. Only applies if 'format = TRUE'
verbose	Logical. Include progress messages
quiet	Logical. Suppress all messages (including messages regarding missing data, etc.)

Details

Data can be returned 'raw' (format = FALSE) or can be formatted. Formatting transforms dates/times to date/time class, renames columns, and converts data to numeric where possible. If character strings are contained in traditionally numeric fields (e.g., weather speed may have values such as "< 30"), they can be replaced with a character specified by 'string_as'. The default is NA. Formatting also replaces data associated with certain flags with NA (M = Missing).

Start and end date can be specified, but if not, it will default to the start and end date of the range (this could result in downloading a lot of data!).

Times are returned as the Etc/GMT offset timezone corresponding to the location. This does not include daylight savings. However, for compatibility with other data sets, timezones can be converted by specifying the desired timezone in 'tz_disp'.

By default, downloads from "http://climate.weather.gc.ca/climate_data/bulk_data_e.html"

Data is downloaded from ECCC as a series of files which are then bound together. Each file corresponds to a different month, or year, depending on the interval. Metadata (station name, lat, lon,

elevation, etc.) is extracted from the start of the most recent file (i.e. most recent dates) for a given station. Note that important data (i.e. station name, lat, lon) is unlikely to change between files (i.e. dates), but some data may or may not be available depending on the date of the file (e.g., station operator was added as of April 1st 2018, so will be in all data which includes dates on or after April 2018).

Value

A tibble with station ID, name and weather data.

Examples

```
kam <- weather_dl(station_ids = 51423,
                 start = "2016-01-01", end = "2016-02-15")

stations_search("Kamloops A$", interval = "hour")
stations_search("Prince George Airport", interval = "hour")

kam.pg <- weather_dl(station_ids = c(48248, 51423),
                   start = "2016-01-01", end = "2016-02-15")

library(ggplot2)

ggplot(data = kam.pg, aes(x = time, y = temp,
                        group = station_name,
                        colour = station_name)) +
  geom_line()
```

weather_interp

Interpolate and add weather data to a dataframe

Description

When data and the weather measurements do not perfectly line up, perform a linear interpolation between two weather measurements and merge the results into the provided dataset. Only applies to numerical weather columns (see `weather` for more details).

Usage

```
weather_interp(data, weather, cols = "all", interval = "hour",
              na_gap = 2, quiet = FALSE)
```

Arguments

data	Dataframe. Data with dates or times to which weather data should be added.
weather	Dataframe. Weather data downloaded with <code>weather</code> which should be interpolated and added to data.
cols	Character. Vector containing the weather columns to add or 'all' for all relevant columns. Note that some measure are omitted because they cannot be linearly interpolated (e.g., wind direction).
interval	What interval is the weather data recorded at? "hour" or "day".
na_gap	How many hours or days (depending on the interval) is it acceptable to skip over when interpolating over NAs (see details).
quiet	Logical. Suppress all messages (including messages regarding missing data, etc.)

Details

Dealing with NA values If there are NAs in the weather data, `na_gap` can be used to specify a tolerance. For example, a tolerance of 2 with an interval of "hour", means that a two hour gap in data can be interpolated over (i.e. if you have data for 9AM and 11AM, but not 10AM, the data between 9AM and 11AM will be interpolated. If, however, you have 9AM and 12PM, but not 10AM or 11AM, no interpolation will happen and data between 9AM and 12PM will be returned as NA.)

Examples

```
# Weather data only
head(kamloops)

# Data about finch observations at RFID feeders in Kamloops, BC
head(finches)

# Match weather to finches

## Not run
finch_weather <- weather_interp(data = finches, weather = kamloops)
```

Index

*Topic **datasets**

- finches, [3](#)
- flags, [4](#)
- glossary, [4](#)
- kamloops, [5](#)
- kamloops_day, [6](#)
- pg, [8](#)
- stations, [9](#)

`add_weather (weather_interp)`, [14](#)

`finches`, [2](#), [3](#)

`flags`, [4](#)

`get_tz`, [4](#)

`glossary`, [4](#)

`kamloops`, [3](#), [5](#)

`kamloops_day`, [3](#), [6](#)

`pg`, [3](#), [8](#)

`sp`, [11](#)

`stations`, [2](#), [9](#), [13](#)

`stations_all (stations_dl)`, [10](#)

`stations_dl`, [2](#), [10](#)

`stations_search`, [2](#), [11](#), [13](#)

`tz_calc`, [12](#)

`weather`, [5](#), [6](#), [8](#), [15](#)

`weather (weather_dl)`, [12](#)

`weather_dl`, [2](#), [12](#)

`weather_interp`, [2](#), [14](#)

`weathercan (weathercan-package)`, [2](#)

`weathercan-package`, [2](#)