

Package ‘GSODR’

June 13, 2017

Type Package

Title Global Summary Daily Weather Data in R

Version 1.0.3

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

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

Description Provides automated downloading, parsing, cleaning, unit conversion and formatting of Global Surface Summary of the Day (GSOD) weather data from the from the USA National Centers for Environmental Information (NCEI) for use in R. Units are converted from from United States Customary System (USCS) units to International System of Units (SI). Stations may be individually checked for number of missing days defined by the user, where stations with too many missing observations are omitted. Only stations with valid reported latitude and longitude values are permitted in the final data. Additional useful elements, saturation vapour pressure (es), actual vapour pressure (ea) and relative humidity are calculated from the original data and included in the final data set. The resulting data include station identification information, state, country, latitude, longitude, elevation, weather observations and associated flags. Data may be automatically saved to disk. File output may be returned as a comma-separated values (CSV) or GeoPackage (GPKG) file. Additional data are included with this R package: a list of elevation values for stations between -60 and 60 degrees latitude derived from the Shuttle Radar Topography Measuring Mission (SRTM). For information on the GSOD data from NCEI, please see the GSOD readme.txt file available from, <<http://www1.ncdc.noaa.gov/pub/data/gsod/readme.txt>>.

Depends R (>= 3.2.0)

License MIT + file LICENSE

Imports curl, data.table, dplyr, fields, magrittr, purrr (>= 0.2.0),
R.utils, readr, rgdal (>= 1.1-9), sp, stats, utils

Suggests ggplot2, knitr, lubridate, plotKML, raster, reshape2, rgeos,
rmarkdown, roxygen2 (>= 6.0.1), spacetime, testthat, tibble,
tidyr, covr

RoxygenNote 6.0.1

Encoding UTF-8
NeedsCompilation no
Repository CRAN
LazyData FALSE
LazyLoad FALSE
ByteCompile TRUE
VignetteBuilder knitr
Author Adam Sparks [aut, cre] (<http://orcid.org/0000-0002-0061-8359>),
Tomislav Hengl [aut] (<http://orcid.org/0000-0002-9921-5129>),
Andrew Nelson [aut] (<http://orcid.org/0000-0002-7249-3778>)
Maintainer Adam Sparks <adamhsparks@gmail.com>
Date/Publication 2017-06-13 07:06:58 UTC

R topics documented:

country_list	2
get_GSOD	3
GSODR	6
GSODR-deprecated	7
isd_history	7
nearest_stations	9
reformat_GSOD	10
update_station_list	12
Index	13

country_list	<i>country_list</i>
--------------	---------------------

Description

FIPS Federal Information Processing Standards (FIPS) code
COUNTRY_NAME English language name
iso2c ISO 3166-1 alpha-2 – two-letter country codes
iso3c ISO 3166-1 alpha-3 – three-letter country codes

Usage

country_list

Format

An object of class `data.table` (inherits from `data.frame`) with 293 rows and 4 columns.

Note

Users of these data should take into account the following (from the NCEI website): “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

The country_list data are automatically loaded with the GSODR package.

Source

<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/country-list.txt>

get_GSOD	<i>Download, clean, reformat generate new elements and return a tidy data.frame of GSOD weather data</i>
----------	--

Description

This function automates downloading, cleaning, reformatting of data from the Global Surface Summary of the Day (GSOD) data provided by the US National Centers for Environmental Information (NCEI), <https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>, and elements three new variables; saturation vapour pressure (es) – Actual vapour pressure (ea) and relative humidity (RH). Stations reporting a latitude of < -90 or > 90 or longitude of < -180 or > 180 are removed. Stations may be individually checked for number of missing days to assure data quality and omitting stations with too many missing observations. All units are converted to International System of Units (SI), e.g., Fahrenheit to Celsius and inches to millimetres. Alternative elevation measurements are supplied for missing values or values found to be questionable based on the Consultative Group for International Agricultural Research’s Consortium for Spatial Information group’s (CGIAR-CSI) Shuttle Radar Topography Mission 90 metre (SRTM 90m) digital elevation data based on NASA’s original SRTM 90m data. Further information on these data and methods can be found on GSODR’s GitHub repository here: https://github.com/ropensci/GSODR/blob/master/data-raw/fetch_isd-history.md.

Usage

```
get_GSOD(years = NULL, station = NULL, country = NULL, dsn = NULL,
  filename = NULL, max_missing = NULL, agroclimatology = FALSE,
  CSV = FALSE, GPKG = FALSE)
```

Arguments

years	Year(s) of weather data to download.
station	Optional. Specify a station or multiple stations for which to retrieve, check and clean weather data using STNID. The NCEI reports years for which the data are available. This function checks against these years. However, not all cases are properly documented and in some cases files may not exist on the ftp server

even though it is indicated that data was recorded for the station for a particular year. If a station is specified that does not have an existing file on the server, this function will silently fail and move on to existing files for download and cleaning from the FTP server.

country	Optional. Specify a country for which to retrieve weather data; full name or ISO codes can be used. See country_list for a full list of country names and ISO codes available.
dsn	Optional. Local file path to write file out to. Must be specified if CSV or GPKG parameters are selected. If unspecified and CSV or GPKG are set to TRUE, dsn will default to the current working directory.
filename	Optional. The filename for resulting file(s) to be written with no file extension. File extension will be automatically appended to file outputs. If unspecified by the user it will default to "GSOD" followed by the file extension(s) set using CSV or GPKG.
max_missing	Optional. The maximum number of days allowed to be missing from a station's data before it is excluded from final file output.
agroclimatology	Optional. Logical. Only clean data for stations between latitudes 60 and -60 for agroclimatology work, defaults to FALSE. Set to TRUE to include only stations within the confines of these latitudes.
CSV	Optional. Logical. If set to TRUE, create a comma separated value (CSV) file and save it locally in a user specified location, if dsn is not specified by the user, defaults to the current working directory.
GPKG	Optional. Logical. If set to TRUE, create a GeoPackage file and save it locally in a user specified location, if dsn is not specified by the user, defaults to the current working directory.

Details

Data summarise each year by station, which include vapour pressure and relative humidity elements calculated from existing data in GSOD.

If the option to save locally is selected. Output may be saved as comma- separated value (CSV) or GeoPackage (GPKG) files in a directory specified by the user, defaulting to the current working directory.

When querying selected stations and electing to write files to disk, all years queried and stations queried will be merged into one final output file.

All missing values in resulting files are represented as NA regardless of which field they occur in.

For a complete list of the fields and description of the contents and units, please refer to the vignette("GSODR_output_fields").

For more information see the description of the data provided by NCEI, http://www7.ncdc.noaa.gov/CDO/GSOD_DESC.txt.

Value

A [data.frame](#) object of weather data or a comma-separated value (CSV) or GeoPackage (GPKG) file saved to local disk.

Note

Some of these data are redistributed with this R package. Originally from these data come from the US NCEI which states that users of these data should take into account the following: “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

Author(s)

Adam H Sparks, <adamhsparks@gmail.com>

References

Jarvis, A., Reuter, H. I, Nelson, A., Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database <http://srtm.csi.cgiar.org>

See Also

[reformat_GSOD](#)

Examples

```
## Not run:
# Download weather station for Toowoomba, Queensland for 2010
t <- get_GSOD(years = 2010, station = "955510-99999")

# Download data for Philippines for year 2010 and generate a yearly
# summary GeoPackage file, Philippines_GSOD-2010.gpkg, file in the user's
# home directory with a maximum of five missing days per station allowed.

get_GSOD(years = 2010, country = "Philippines", dsn = "~/",
filename = "Philippines_GSOD", GPKG = TRUE, max_missing = 5)

# Download global GSOD data for agroclimatology work for years 2009 and 2010
# and generate yearly summary files, GSOD-agroclimatology-2010.csv and
# GSOD-agroclimatology-2011.csv in the user's home directory.

get_GSOD(years = 2010:2011, dsn = "~/",
filename = "GSOD_agroclimatology_2010-2011", agroclimatology = TRUE,
CSV = TRUE)

## End(Not run)
```

Description

Provides automated downloading, parsing, cleaning, unit conversion and formatting of Global Surface Summary of the Day (GSOD) weather data from the from the USA's National Oceanic and Atmospheric Administration's (NOAA) National Centre for Environmental Information (NCEI) for use in R. All units are converted from from United States Customary System (USCS) units to International System of Units (SI). Stations may be individually checked for number of missing days defined by the user, where stations with too many missing observations are omitted. Only stations with valid reported latitude and longitude values are permitted in the final data. Additional useful elements, saturation vapour pressure (es), actual vapour pressure (ea) and relative humidity are calculated from the original data and included in the final data set. The resulting data include station identification information, state, country, latitude, longitude, elevation, weather observations and associated flags. Data may be automatically saved to disk. File output may be returned as a comma-separated values (CSV) or GeoPackage (GPKG) file. Additional data are included with this R package: a list of elevation values for stations between -60 and 60 degrees latitude derived from the Shuttle Radar Topography Measuring Mission (SRTM). For information on the GSOD data from NCEI, please see the GSOD readme.txt file available from, <http://www1.ncdc.noaa.gov/pub/data/g sod/readme.txt>. For climate data that have been formatted specifically for use with the GSODR package, please see the GSODRdata package (Sparks *et al.*) available on GitHub: <https://adamhsparks.github.io/GSODRdata/>. Four data frames of climate data are provided from various sources for GSOD station locations.

Author(s)

Adam Sparks, Tomislav Hengle and Andrew Nelson

Source

<https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>

References

- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., *et al.* (2016) Climatologies at high resolution for the Earth land surface areas. *arXiv preprint arXiv:1607.00217*
- New, M., Lister, D., Hulme, M., Makin, I., (2002) A high-resolution data set of surface climate over global land areas. *Climate Research* **21**:1–25
- Sparks A., Hengl T., Nelson A. (2017) GSODRdata: Extra Climate Data for the GSODR Package. R package version 0.1.0, <https://adamhsparks.github.io/GSODRdata/index.html>.
- Wilson A. M., Jetz W. (2016) Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLoS Biol* **14**(3): e1002415. doi:10.1371/journal.pbio.1002415

See Also**GSODR functions:**

[get_GSOD](#) Fetch, clean and reformat data from NCEI GSOD database

[reformat_GSOD](#) Clean and reformat local files from NCEI GSOD database

[nearest_stations](#) Find stations within a given radius expressed in kilometres for a given point using Latitude and Longitude

[update_station_list](#) Download the latest station list from the NCEI FTP server updates the package's internal database of stations and their metadata.

[isd_history](#) GSODR provides a cleaned, local copy of the "isd_history.csv" file to help make some processes faster with elevation data derived from SRTM90m data also included in these data

[country_list](#) GSODR provides a cleaned list of the countries which are represented in the GSOD data set

Useful links:

- <https://github.com/ropensci/GSODR>
- Report bugs at <https://github.com/ropensci/GSODR/issues>

 GSODR-deprecated

Deprecated function(s) in the GSODR package

Description

These functions are provided for compatibility with older version of the GSODR package. They may eventually be completely removed.

Usage

```
get_station_list()
```

Details

`get_station_list` now superceded by [update_station_list](#)

`get_station_list` was used to download the list of weather stations and the corresponding meta-data. The station list is no longer fetched on-the-fly. Instead a version is supplied with the GSODR package upon installation. To update the internal database, please use [update_station_list](#).

 isd_history

isd_history

Description

USAF Air Force Datsav3 station number

WBAN Weather Bureau Army Navy (5 digit identifier)

STN_NAME Unique station name

CTRY FIPS country ID

STATE If applicable, US states only (2 letter code)

CALL ICAO Identifier, identifiers approved for use under the International Civil Aviation Administration plan of identifiers (4 letter identifier)

LAT Latitude in thousandths of decimal degrees

LON Longitude in thousandths of decimal degrees

ELEV_M Elevation to tenths in metres

BEGIN First available date of data for station, YYYYMMDD format

END Last available date of data for station, YYYYMMDD format

STNID Unique station ID, a concatenation of USAF and WBAN number, used for merging with station data weather files

ELEV_M_SRTM_90m Elevation in metres extracted from SRTM data (Jarvis *et al.* 2008)

Usage

isd_history

Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 28330 rows and 13 columns.

Details

Station elevation information for the US National Centers for Environmental Information (NCEI) Global Surface Summary of the Day (GSOD) weather data. The original file has missing and incorrect geographic data including location (LAT/LON) and elevation. These data provide an alternative set of elevation values with the following changes to the original list of stations from the NCEI:

1. Stations with both a latitude and longitude of 0 degrees were removed.
2. Stations with longitude values that are beyond -180/180 degrees were removed.
3. Stations with latitude values that are beyond -90/90 degrees were removed.
4. A new field for elevation is included, `ELEV_M_SRTM_90m`. This was created using mean values of a 200m buffer around the reported LAT/LON station location within the CGIAR-CSI hole-filled 90m SRTM digital elevation model (Jarvis *et al.* 2008).

For more on this, please consult the document available from the GSODR GitHub repository detailing the process used to generate these data, https://github.com/ropensci/GSODR/blob/master/data-raw/fetch_isd-history.md

Note

Users of these data should take into account the following (from the NCEI website): “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

The `isd_history` data are automatically loaded with the [GSODR](#) package and merged with the latest available data from the NCEI in the "isd-history.csv" file.

To update these data with the latest available, use [update_station_list](#).

Source

<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-history.csv>

References

Jarvis, A., Reuter, H. I, Nelson, A., Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database <http://srtm.csi.cgiar.org>

nearest_stations	<i>Find nearest GSOD stations to specified latitude and longitude</i>
------------------	---

Description

Given a latitude and longitude value entered as decimal degrees (DD), this function returns a list of STNID values, which can be used in [get_GSOD](#) to query for specific stations as an argument in the station parameter of that function.

Usage

```
nearest_stations(LAT, LON, distance)
```

Arguments

LAT	Latitude expressed as decimal degrees (DD) [WGS84]
LON	Longitude expressed as decimal degrees (DD) [WGS84]
distance	Distance in kilometres from point for which stations are to be returned.

Value

[vector](#) object of station identification numbers

Note

The GSOD data, which are downloaded and manipulated by this R package, stipulate that the following notice should be given. “The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification.”

Author(s)

Adam H Sparks, <adamhsparks@gmail.com>

Examples

```
## Not run:
# Find stations within a 100km radius of Toowoomba, QLD, AUS

n <- nearest_stations(LAT = -27.5598, LON = 151.9507, distance = 100)

## End(Not run)
```

reformat_GSOD

Clean, reformat and generate new variables from GSOD weather data

Description

This function automates cleaning and reformatting of GSOD, <https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>, station files in "WMO-WBAN-YYYY.op.gz" format that have been downloaded from the United States National Center for Environmental Information's (NCEI) FTP server.

Usage

```
reformat_GSOD(dsn = NULL, file_list = NULL)
```

Arguments

dsn	User supplied file path to location of station file data on local disk for reformatting.
file_list	User supplied list of files of station data on local disk for reformatting.

Details

For automated downloading and processing see the [get_GSOD](#) function which provides expanded functionality for automatically downloading and expanding annual GSOD files and cleaning station files.

This function reformats the data into a more usable form and calculates three new elements; saturation vapour pressure (es), actual vapour pressure (ea) and relative humidity (RH). All units are

converted to International System of Units (SI), e.g., Fahrenheit to Celsius and inches to millimetres. Alternative elevation measurements are supplied for missing values or values found to be questionable based on the Consultative Group for International Agricultural Research's Consortium for Spatial Information group's (CGIAR-CSI) Shuttle Radar Topography Mission 90 metre (SRTM 90m) digital elevation data based on NASA's original SRTM 90m data.

Data summarise each year by station, which include vapour pressure and relative humidity elements calculated from existing data in GSOD.

All missing values in resulting files are represented as NA regardless of which field they occur in.

Only station files in ".op.gz" file format are supported by this function. If you have downloaded the full annual "gsod_YYYY.tar" file you will need to extract the individual station files first to use this function.

For a complete list of the fields and description of the contents and units, please refer to the vignette("GSODR_output_fields")

Value

A `data.frame` object of weather data or a comma-separated value (CSV) or GeoPackage (GPKG) file saved to local disk.

Note

Some of these data are redistributed with this R package. Originally from these data come from the US NCEI which states that users of these data should take into account the following: "The following data and products may have conditions placed on their international commercial use. They can be used within the U.S. or for non-commercial international activities without restriction. The non-U.S. data cannot be redistributed for commercial purposes. Re-distribution of these data by others must provide this same notification."

Author(s)

Adam H Sparks, <adamhsparks@gmail.com>

References

Jarvis, A., Reuter, H.I, Nelson, A., Guevara, E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database <http://srtm.csi.cgiar.org>

See Also

[get_GSOD](#)

Examples

```
## Not run:

# Reformat station data files in local directory
x <- reformat_GSOD(dsn = "~/tmp")

# Reformat a list of data files
y <- c("~/GSOD/gsod_1960/200490-99999-1960.op.gz",
```

```
      "~/GSOD/gsod_1961/200490-99999-1961.op.gz")
x <- reformat_GSOD(file_list = y)

## End(Not run)
```

update_station_list	<i>Download the latest station list from the NCEI server and update internal database</i>
---------------------	---

Description

This function downloads the latest station list (isd-history.csv) from the NCEI FTP server and updates the data distributed with GSODR so that you have the latest list of stations available. These data provide unique identifiers, country, state (if in US), latitude, longitude, elevation and when weather observations begin and end. Stations with invalid latitude and longitude values will not be included.

Usage

```
update_station_list()
```

Details

There is no need to use this unless you know that a station exists in the GSODR data that is not available in the database distributed with [GSODR](#) in the [isd_history](#) data distributed with [GSODR](#).

Value

[data.table](#) object of station metadata.

Author(s)

Adam H Sparks, <adamhsparks@gmail.com>

Examples

```
## Not run:
update_station_list()

## End(Not run)
```

Index

*Topic **datasets**

country_list, [2](#)

isd_history, [7](#)

country_list, [2](#), [4](#), [7](#)

data.frame, [4](#), [11](#)

data.table, [12](#)

get_GSOD, [3](#), [7](#), [9–11](#)

get_station_list (GSODR-deprecated), [7](#)

GSODR, [6](#), [9](#), [12](#)

GSODR-deprecated, [7](#)

GSODR-deprecated-package
(GSODR-deprecated), [7](#)

GSODR-package (GSODR), [6](#)

isd_history, [7](#), [7](#), [12](#)

nearest_stations, [7](#), [9](#)

reformat_GSOD, [5](#), [7](#), [10](#)

update_station_list, [7](#), [9](#), [12](#)

vector, [9](#)