

Package 'GSODR'

August 29, 2016

Type Package

Title Download, Clean and Generate New Variables from GSOD Data

Version 0.1.9

Date 2016-07-15

Maintainer Adam Sparks <adamhsparks@gmail.com>

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

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

Description Download, clean, reformat and create new variables from the USA National Climatic Data Center (NCDC) Global Surface Summary of the Day (GSOD) weather stations data, <<https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>>. The function, `get_GSOD()`, retrieves data from the GSOD ftp site and reformats it from United States Customary System (USCS) units to International System of Units (SI). Stations are individually checked for number of missing days, as defined by the user, to assure data quality. Stations with too many missing observations are omitted from final file. Stations with missing latitude or longitude or values for both of 0 are omitted. Also omitted are stations with a latitude of < -90 or > 90 or longitude of < - 180 or > 180. Output is returned as a comma-separated values (CSV) file written to disk in a location selected by the user, which summarises each year by station and includes new variables: actual vapor pressure, saturation vapor pressure and relative humidity, calculated from the original GSOD data. The resulting files can be as large as 500mb depending on the user's stringency for missing data and geographic area of interest. Be sure to have sufficient RAM and disk space as well as a reasonably fast internet connection to use this package to perform this operation. However, for much smaller and more manageable data sets, an individual country of interest may be selected as well as only stations falling between -60/60 degrees latitude for agroclimatology work or individual stations if the station ID is known. The resulting files include station data (e.g., station name, country, latitude, longitude, elevation) for use in a geographic information system (GIS). The function was largely based on T. Hengli's 'getGSOD.R' script, available from

<<http://spatial-analyst.net/book/system/files/getGSOD.R>> with enhancements to be cross-platform, faster and more flexible. For information on the original data from NCDC, please see the GSOD readme.txt file available from, <<http://www1.ncdc.noaa.gov/pub/data/g sod/readme.txt>>.

Depends R (>= 3.0.0)

License GPL (>= 3)

Imports curl, data.table, dplyr, doParallel, foreach, lubridate, iotools, iterators, parallel, raster, readr, rgdal, settings, sp, stats, stringr, utils

RoxygenNote 5.0.1

Encoding UTF-8

NeedsCompilation no

Repository CRAN

LazyData TRUE

ByteCompile TRUE

Author Adam Sparks [aut, cre],
Tomislav Hengl [ctb],
Andy Nelson [ctb],
Kay Sumfleth [ctb]

Date/Publication 2016-07-15 10:14:31

R topics documented:

country_list	2
get_GSOD	3
stations	7

Index	9
--------------	----------

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

Description

country_list

Usage

country_list

Format

A data frame with 293 observations of 4 variables:

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

Note

Users of these data should take into account the following (from the NCDC 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

Download, Clean, Reformat and Generate New Variables From GSOD Weather Data

Description

This function automates downloading, cleaning, reformatting of data from the Global Surface Summary of the Day (GSOD) data provided by the US National Climatic Data Center (NCDC), <https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>, and calculates three new variables; Saturation Vapor Pressure (ES) – Actual Vapor Pressure (EA) and relative humidity (RH). Stations are individually checked for number of missing days to assure data quality, stations with too many missing observations are omitted, stations with a latitude of < -90 or > 90 or longitude of < -180 or > 180 are removed. 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/adamhsparks/GSODR/blob/master/data-raw/fetch_isd-history.md

Usage

```
get_GSOD(years = NULL, station = NULL, country = NULL, path = "",
          max_missing = 5, agroclimatology = FALSE, shapefile = FALSE,
          CSV = TRUE, merge_station_years = FALSE)
```

Arguments

years	Year(s) of weather data to download.
station	Specify single station for which to retrieve, check and clean weather data.
country	Specify a country of interest for which to retrieve weather data; full name. For stations located in locales having an ISO code 2 or 3 letter ISO code can also be used if known. See country_list for a full list of country names and ISO codes available.
path	Path entered by user indicating where to store resulting output file. Defaults to the current working directory.
max_missing	The maximum number of days allowed to be missing from a station's data before it is excluded from final file output. Defaults to five days. If a single station is specified, this option is ignored and any data available, even an empty file, from NCDC will be returned.
agroclimatology	Only clean data for stations between latitudes 60 and -60 for agroclimatology work, defaults to FALSE. Set to FALSE to override and include only stations within the confines of these latitudes.
shapefile	If set to TRUE, create an ESRI shapefile of vector type, points, of the data for use in a GIS. Defaults to FALSE, no shapefile created.
CSV	If set to TRUE, create a comma separated value (CSV) file of data, defaults to TRUE, a CSV file is created.
merge_station_years	If set to TRUE, merge output files into one output file for all years when selecting a single station, defaults to FALSE.

Details

Due to the size of the resulting data, output is saved as a comma-separated, csv, file (default) or ESRI shapefile in a directory specified by the user or defaults to the current working directory. The files summarize each year by station, which includes vapour pressure and relative humidity variables calculated from existing data in GSOD. Optionally, because the file sizes are much smaller, when selecting a single station, all years queried may be merged into one final output file (CSV or shapefile) using the `merge_station_years` option.

All missing values in resulting files are represented as -9999 regardless of which field they occur in.

Be sure to have disk space free and allocate the proper time for this to run. This is a time, processor and disk input/output/space intensive process. This function was largely based on T. Hengl's "getGSOD.R" script, available from <http://spatial-analyst.net/book/system/files/getGSOD.R> with enhancements to be cross-platform, faster and more flexible. For more information see the description of the data provided by NCDC, http://www7.ncdc.noaa.gov/CD0/GSOD_DESC.txt.

The CSV or ESRI format shapefile in the respective year-directory will contain the following fields/values:

STNID Station number (WMO/DATSAV3 number) for the location

WBAN Number where applicable—this is the historical "Weather Bureau Air Force Navy" number
- with WBAN being the acronym

STN.NAME Unique text string identifier

CTRY Country

LAT Latitude

LON Longitude

ELEV.M Station reported elevation (metres to tenths)

ELEV.M.SRTM.90m Corrected elevation data in whole metres for stations derived from Jarvis et al. (2008), extracted from DEM using reported LAT/LON values in metres

YEARMODA Date in YYYY-MM-DD format

YEAR The year

MONTH The month

DAY The day

YDAY Sequential day of year (not in original GSOD)

TEMP Mean daily temperature converted to degrees C to tenths. Missing = -9999

TEMP.CNT Number of observations used in calculating mean daily temperature

DEWP Mean daily dew point converted to degrees C to tenths. Missing = -9999

DEWP.CNT Number of observations used in calculating mean daily dew point

SLP Mean sea level pressure in millibars to tenths. Missing = -9999

SLP.CNT Number of observations used in calculating mean sea level pressure

STP Mean station pressure for the day in millibars to tenths Missing = -9999

STP.CNT Number of observations used in calculating mean station pressure

VISIB Mean visibility for the day converted to kilometers to tenths Missing = -9999

VISIB.CNT Number of observations used in calculating mean daily visibility

WDSP Mean daily wind speed value converted to metres/second to tenths Missing = -9999

WDSP.CNT Number of observations used in calculating mean daily windspeed

MXSPD Maximum sustained wind speed reported for the day converted to metres/second to tenths. Missing = -9999

GUST Maximum wind gust reported for the day converted to metres/second to tenths. Missing = -9999

MAX Maximum temperature reported during the day converted to Celsius to tenths—time of maximum temperature report varies by country and region, so this will sometimes not be the maximum for the calendar day. Missing = -9999

MAX.FLAG Blank indicates maximum temperature was taken from the explicit maximum temperature report and not from the 'hourly' data. " * " indicates maximum temperature was derived from the hourly data (i.e., highest hourly or synoptic-reported temperature)

MIN Minimum temperature reported during the day converted to Celsius to tenths—time of minimum temperature report varies by country and region, so this will sometimes not be the minimum for the calendar day. Missing = -9999

MIN.FLAG Blank indicates minimum temperature was taken from the explicit minimum temperature report and not from the 'hourly' data. " * " indicates minimum temperature was derived from the hourly data (i.e., lowest hourly or synoptic-reported temperature)

PRCP Total precipitation (rain and/or melted snow) reported during the day converted to millimetres to hundredths will usually not end with the midnight observation—i.e., may include latter part of previous day. ".00" indicates no measurable precipitation (includes a trace). Missing = -9999. *Note:* Many stations do not report '0' on days with no precipitation— therefore, '-9999' will often appear on these days. For example, a station may only report a 6-hour amount for the period during which rain fell. See PRCP.FLAG column for source of data

PRCP.FLAG A = 1 report of 6-hour precipitation amount

B = Summation of 2 reports of 6-hour precipitation amount

C = Summation of 3 reports of 6-hour precipitation amount

D = Summation of 4 reports of 6-hour precipitation amount

E = 1 report of 12-hour precipitation amount

F = Summation of 2 reports of 12-hour precipitation amount

G = 1 report of 24-hour precipitation amount

H = Station reported '0' as the amount for the day (eg, from 6-hour reports), but also reported at least one occurrence of precipitation in hourly observations—this could indicate a trace occurred, but should be considered as incomplete data for the day

I = Station did not report any precipitation data for the day and did not report any occurrences of precipitation in its hourly observations. It's still possible that precipitation occurred but was not reported

SNDP Snow depth in millimetres to tenths. Missing = -9999

I.FOG Fog, (1 = yes, 0 = no/not reported) for the occurrence during the day

I.RAIN_DZL Rain or drizzle, (1 = yes, 0 = no/not reported) for the occurrence during the day

I.SNW_ICE Snow or ice pellets, (1 = yes, 0 = no/not reported) for the occurrence during the day

I.HAIL Hail, (1 = yes, 0 = no/not reported) for the occurrence during the day

I.THUNDER Thunder, (1 = yes, 0 = no/not reported) for the occurrence during the #' day

I.TDO_FNL Tornado or funnel cloud, (1 = yes, 0 = no/not reported) for the occurrence during the day

Values calculated by this package and included in final output:

ea Mean daily actual vapour pressure

es Mean daily saturation vapour pressure

RH Mean daily relative humidity

Note

Users of these data should take into account the following (from the NCDC 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.”

References

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

Examples

```
## Not run:
# Download weather station for Toowoomba, Queensland for 2010, save resulting
# file, GSOD-955510-99999-2010.csv, in the user's home directory.

get_GSOD(years = 2010, station = "955510-99999", path = "~/")

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

get_GSOD(years = 2010, country = "Philippines", path = "~/")

# 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 with a maximum
# of five missing days per weather station allowed.

get_GSOD(years = 2010:2011, path = "~/", agroclimatology = TRUE)

## End(Not run)
```

stations

*Stations***Description**

Stations

Usage

stations

Format

A data frame with 24878 observations of 13 variables:

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)

Station information for the US National Climatic Data Centre (NCDC) Global Surface Summary of the Day (GSOD) weather data. The original file has missing and incorrect information. This is a clean version of this dataset, provided by this package. The following changes were made.

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. For convenience elevation is converted from decimetres to metres.
5. STNID is added as a new field, a concatenation of USAF and WBAN.
6. A new field for elevation is included, ELEV.M.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)

Note

Users of these data should take into account the following (from the NCDC 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 stations data are automatically loaded with the GSODR package.

Source

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

References

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

Index

*Topic **datasets**

country_list, [2](#)

stations, [7](#)

country_list, [2, 4](#)

get_GSOD, [3](#)

stations, [7](#)