

Package ‘KNMI’

July 14, 2017

Title functions for retrieving climate data from the Dutch meteorological institute KNMI.

Version 0.2.0

Author person("Bart", "van Hest", email = "hestbv@gmail.com", role = c("aut", "cre"))

Maintainer Bart van Hest <hestbv@gmail.com>

Description This package provides functions to retrieve raw and processed climate data from both the official Dutch meteorological institute (KNMI). For some stations, data goes back to 1900. The data can be subsetted based on year and specified variables. The returned data is converted to SI-units, when necessary. Several processing functions are available, such as a function that calculates the Hugin- or VE-index. The package also contains two KNMI dataset with all measurement stations and with long term averages from 2000-2015.

Depends R (>= 3.2.1), data.table, lubridate, plyr, RCurl

Imports data.table, lubridate, plyr, RCurl

License GPL-3

LazyData true

Encoding UTF-8

RoxygenNote 6.0.1

R topics documented:

berekenHuglinIndex	1
berekenVEIndex	2
date2DayOfYear	3
dayOfYear2Date	4
delta.ljgem	4
find_nearest_KNMI_station	5
getClimateDataInBulk	5
getClimateDataSet	7
knmi.langJarigGem	10
plot_KNMI_stations	10

printStations	11
renameColumnsKNMIdata	11
stations	12
subsetKNMIdata	12
voegToe	13

berekenHuglinIndex *calculate the Huglin index*

Description

This function calculates the Huglin index, which is one of the statistics that represents the total amount of heat in a season that is bennefical to the growth of plants. More specific, its used to measure the amount of warmth that is required by grapes to start growing, blossom and ripen.

Usage

```
berekenHuglinIndex(dgg, startDate = paste(format(Sys.Date(), format = "%Y"),
      "01-01", sep = "-"), endDate = paste(format(Sys.Date(), format = "%Y"),
      "12-31", sep = "-"))
```

Arguments

dgg	dotaframe containing three mandatory columns; the day-of-the-year (dagVjaar), the daily mean temperature (gemTemp), the daily maximum temperature (maxTemp) and an optional column containing the year (jaar). Temperatures in degrees Celcius. The data frame can contain the ranges for multiple years.
startDate	the start of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the start of the year.
endDate	the end of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the end of the year.

Details

It is calculated as the cumulative summation over the days of the year of the mean of the daily average temperature (in degrees Celcius) plus the daily maximum temperature, subtracted with 10 (the temperature at which grapes start growing). This value is only added to the total when its positive. The value is multiplied by a correction for the latitude of the measurement location (1.06 for The Netherlands). The default summation interval, from the first of April until the end of September, has been altered to be able to deal with cool-climate winegrowing that requires a prolonged growing season.

Documentation: <https://de.wikipedia.org/wiki/Huglin-Index>

Value

dataframe containing the year (jaar), the day-of-the-year (dagVjaar) and the Huglin-index at the dagVjaar.

berekenVEIndex	<i>calculate the VE index</i>
----------------	-------------------------------

Description

This function calculates the VE index, which is one of the statistics that represents the total amount of heat in a season that is beneficial to the growth of plants. More specific, its used to measure the amount of warmth that is required by grapes to start growing, blossom and ripen.

Usage

```
berekenVEIndex(dgg, startDate = paste(format(Sys.Date(), format = "%Y"),  
  "01-01", sep = "-"), endDate = paste(format(Sys.Date(), format = "%Y"),  
  "12-31", sep = "-"))
```

Arguments

dgg	dotaframe containing two mandatory columns; the day-of-the-year (dagVjaar), the daily mean temperature (gemTemp) and an optional column containing the year (jaar). Temperatures in degrees Celcius. The data frame can contain the ranges for multiple years.
startDate	the start of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the start of the year.
endDate	the end of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the end of the year.

Details

It is calculated as the cumulative summation over the days of the year of the mean of the daily minimum temperature (in degrees Celcius) subtracted with 10 (the temperature at which grapes start growing). This value is only added to the total when its positive. The default summation interval, from the first of April until the end of September, has been altered to be able to deal with cool-climate winegrowing that requires a prolonged growing season.

Documentation: <https://de.wikipedia.org/wiki/Wachstumsgradtag>

Value

dataframe containing the year (jaar), the day-of-the-year (dagVjaar) and the VE-index at the dagV-jaar.

date2DayOfYear	<i>convert a date into an integer that represents the day of the year.</i>
----------------	--

Description

convert a date into an integer that represents the day of the year.

Usage

```
date2DayOfYear(datum)
```

Arguments

datum	the date that needs to be converted. Date class.
-------	--

Value

an integer representing the day-of-the-year.

Examples

```
date2DayOfYear(as.Date("2015-01-01")) # expected value: 1
date2DayOfYear(as.Date("2015-12-31")) # expected value: 365
```

dayOfYear2Date	<i>convert a 'day of the year'-value into a date.</i>
----------------	---

Description

convert a 'day of the year'-value into a date.

Usage

```
dayOfYear2Date(doy, year = format(Sys.Date(), format = "%Y"))
```

Arguments

doy	the day of the year that needs to be converted. Integer.
year	the year for which the day of the year is applicable. Integer. Default is the current year.

Value

the date corresponding to the day-of-the-year.

Examples

```

dayOfYear2Date(1) # '01-01-201x'
dayOfYear2Date(1, year=2010) # '01-01-2010'
dayOfYear2Date(0) # 'NA'

```

delta.ljgem	<i>calculate the difference in days between the two ranges with daily values</i>
-------------	--

Description

This function determines the two dates in two ranges of values that have the most similar value. Next, the difference between the two days is calculated.

Usage

```
delta.ljgem(ac, ljg)
```

Arguments

dataframe	containing the columns doy, value
dataframe	containing the columns doy, value

Value

difference in days between the two ranges with daily values.

find_nearest_KNMI_station	<i>find_nearest_KNMI_station</i>
---------------------------	----------------------------------

Description

This function returns the KNMI measurement station that is closed to the provided location. One can select from the active stations (active = TRUE, the default) or *all* stations (active = FALSE).

Usage

```
find_nearest_KNMI_station(location, active = TRUE)
```

Arguments

location	data-frame with a lat/lon-column (using the WGS84 coordinate system).
active	boolean to select only currently active stations. Default = TRUE.

Value

data-frame with the id, name, url to station information and the lat/lon of the nearest KNMI-station.

```
getClimateDataInBulk
```

```
get Climate Data In Bulk
```

Description

This function retrieves raw climate data collected by the official KNMI weather stations. It is optimised for retrieving large sets of data that have been prepared by the KNMI for download. If a year in the past is selected, the call is forwarded to the function 'getClimateDataSet()'. The function 'getClimateDataSet()' in this package is better suited to retrieve data for very specific date-ranges.

Usage

```
getClimateDataInBulk(station = "ALL", from = paste(format(today(), format = "%Y"), "0101", sep = ""), to = format(today() - 1, format = "%Y%m%d"))
```

Arguments

station	ID for the KNMI station. The available stations can be retrieved with the function 'listStations()'. Defaults to "ALL". . Note: a string of characters in the format 'iii'.
from	startdate for the time-window. Defaults to the start of the current year. If the returned data is from a later date, no prior data is available for the selected station. Note: a string of characters in the format 'yyyymmdd'.
to	enddate for the time-window. Defaults to yesterday. If the returned data is from an earlier date, no recent data is available for the selected station. Note: a string of characters in the format 'yyyymmdd'.

Format

The returned data frame contains the following columns:

- STN = ID of measurementstation;
- YYYYMMDD = Datum (YYYY=jaar MM=maand DD=dag);
- DDVEC Vectorgemiddelde windrichting in graden (360=noord, 90=oost, 180=zuid, 270=west, 0=windstil/variabel). Zie <http://www.knmi.nl/kennis-en-datacentrum/achtergrond/klimatologische-brochures-en-boeken>;
- FHVEC Vectorgemiddelde windsnelheid (in 0.1 m/s). Zie <http://www.knmi.nl/kennis-en-datacentrum/achtergrond/klimatologische-brochures-en-boeken>;
- FG Etmaalgemiddelde windsnelheid (in 0.1 m/s);
- FHX Hoogste uurgemiddelde windsnelheid (in 0.1 m/s);
- FHXH Uurvak waarin FHX is gemeten;
- FHN Laagste uurgemiddelde windsnelheid (in 0.1 m/s);
- FHNH Uurvak waarin FHN is gemeten;

- FXX Hoogste windstoot (in 0.1 m/s);
- FXXH Uurvak waarin FXX is gemeten;
- TG Etmaalgemiddelde temperatuur (in 0.1 graden Celsius);
- TN Minimum temperatuur (in 0.1 graden Celsius);
- TNH Uurvak waarin TN is gemeten;
- TX Maximum temperatuur (in 0.1 graden Celsius);
- TXH Uurvak waarin TX is gemeten;
- T10N Minimum temperatuur op 10 cm hoogte (in 0.1 graden Celsius);
- T10NH 6-uurs tijdvak waarin T10N is gemeten;
- SQ Zonneschijnduur (in 0.1 uur) berekend uit de globale straling (-1 voor <0.05 uur);
- SP Percentage van de langst mogelijke zonneschijnduur;
- Q Globale straling (in J/cm²);
- DR Duur van de neerslag (in 0.1 uur);
- RH Etmaalsom van de neerslag (in 0.1 mm) (-1 voor <0.05 mm);
- RHX Hoogste uursom van de neerslag (in 0.1 mm) (-1 voor <0.05 mm);
- RHXH Uurvak waarin RHX is gemeten;
- PG Etmaalgemiddelde luchtdruk herleid tot zeeniveau (in 0.1 hPa) berekend uit 24 uurwaarden;
- PX Hoogste uurwaarde van de luchtdruk herleid tot zeeniveau (in 0.1 hPa);
- PXH Uurvak waarin PX is gemeten;
- PN Laagste uurwaarde van de luchtdruk herleid tot zeeniveau (in 0.1 hPa);
- PNH Uurvak waarin PN is gemeten;
- VVN Minimum opgetreden zicht;
- VVNH Uurvak waarin VVN is gemeten;
- VVX Maximum opgetreden zicht;
- VVXH Uurvak waarin VVX is gemeten;
- NG Etmaalgemiddelde bewolking (bedekkingsgraad van de bovenlucht in achtsten, 9=bovenlucht onzichtbaar);
- UG Etmaalgemiddelde relatieve vochtigheid (in procenten);
- UX Maximale relatieve vochtigheid (in procenten);
- UXH Uurvak waarin UX is gemeten;
- UN Minimale relatieve vochtigheid (in procenten);
- UNH Uurvak waarin UN is gemeten;
- EV24 Referentiegewasverdamping (Makkink) (in 0.1 mm);

Details

You can specify a specific station or get data from all the stations at once (the default). When the from and to date parameters are not provided, all measurements for the current year are returned. Otherwise the data is subsetted to the given interval.

Value

a data frame.

```
getClimateDataSet
```

get Climate Data Set

Description

This function retrieves raw climate data collected by the official KNMI measurement stations for a specific station and/or date-range. It uses the, somewhat slower, KNMI-API to collect the data. The function 'retrieveHistoricData()' in this package is optimized to collect data for larger date-ranges (e.g. for > 10 years), but is less flexible with combinations of all or specific stations and date ranges.

Usage

```
getClimateDataSet(stationID = "ALL", from = paste(format(Sys.Date(), format
= "%Y"), "0101", sep = ""), to = format(Sys.Date() - 1, format =
"%Y%m%d"))
```

Arguments

stationID	ID for the KNMI measurement station. The available stations can be retrieved with the function 'getStations()'. Defaults to "all". . Note: a string of characters in the format 'iii'.
from	startdate for the time-window. Defaults to the start of the current year. Note: a string of characters in the format 'yyyymmdd'.
to	enddate for the time-window. Defaults to yesterday. Note: a string of characters in the format 'yyyymmdd'.

Format

The returned data frame contains the following columns:

- STN = ID of measurementstation;
- YYYYMMDD = Datum (YYYY=jaar MM=maand DD=dag);
- DDVEC Vectorgemiddelde windrichting in graden (360=noord, 90=oost, 180=zuid, 270=west, 0=windstil/variabel). Zie <http://www.knmi.nl/kennis-en-datacentrum/achtergrond/klimatologische-brochures-en-boeken>;
- FHVEC Vectorgemiddelde windsnelheid (in 0.1 m/s). Zie <http://www.knmi.nl/kennis-en-datacentrum/achtergrond/klimatologische-brochures-en-boeken>;
- FG Etmaalgemiddelde windsnelheid (in 0.1 m/s);
- FHX Hoogste uurgemiddelde windsnelheid (in 0.1 m/s);
- FHXH Uurvak waarin FHX is gemeten;
- FHN Laagste uurgemiddelde windsnelheid (in 0.1 m/s);

- FHNH Uurvak waarin FHN is gemeten;
- FXX Hoogste windstoot (in 0.1 m/s);
- FXXH Uurvak waarin FXX is gemeten;
- TG Etmaalgemiddelde temperatuur (in 0.1 graden Celsius);
- TN Minimum temperatuur (in 0.1 graden Celsius);
- TNH Uurvak waarin TN is gemeten;
- TX Maximum temperatuur (in 0.1 graden Celsius);
- TXH Uurvak waarin TX is gemeten;
- T10N Minimum temperatuur op 10 cm hoogte (in 0.1 graden Celsius);
- T10NH 6-uurs tijdvak waarin T10N is gemeten;
- SQ Zonneschijnduur (in 0.1 uur) berekend uit de globale straling (-1 voor <0.05 uur);
- SP Percentage van de langst mogelijke zonneschijnduur;
- Q Globale straling (in J/cm²);
- DR Duur van de neerslag (in 0.1 uur);
- RH Etmaalsom van de neerslag (in 0.1 mm) (-1 voor <0.05 mm);
- RHX Hoogste uursom van de neerslag (in 0.1 mm) (-1 voor <0.05 mm);
- RHHX Uurvak waarin RHX is gemeten;
- EV24 Referentiegewasverdamping (Makkink) (in 0.1 mm);
- PG Etmaalgemiddelde luchtdruk herleid tot zeeniveau (in 0.1 hPa) berekend uit 24 uurwaarden;
- PX Hoogste uurwaarde van de luchtdruk herleid tot zeeniveau (in 0.1 hPa);
- PXH Uurvak waarin PX is gemeten;
- PN Laagste uurwaarde van de luchtdruk herleid tot zeeniveau (in 0.1 hPa);
- PNH Uurvak waarin PN is gemeten;
- VVN Minimum opgetreden zicht;
- VVNH Uurvak waarin VVN is gemeten;
- VVX Maximum opgetreden zicht;
- VVXH Uurvak waarin VVX is gemeten;
- NG Etmaalgemiddelde bewolking (bedekkingsgraad van de bovenlucht in achtsten, 9=bovenlucht onzichtbaar);
- UG Etmaalgemiddelde relatieve vochtigheid (in procenten);
- UX Maximale relatieve vochtigheid (in procenten);
- UXH Uurvak waarin UX is gemeten;
- UN Minimale relatieve vochtigheid (in procenten);
- UNH Uurvak waarin UN is gemeten;

Details

You can specify a specific station or get data from all the stations at once (the default). When the from and to date parameters are not provided, all measurements for the current year are returned. Otherwise the data is subsetted to the given interval.

Value

a data frame.

knmi.langJarigGem	<i>KNMI longterm averages</i>
-------------------	-------------------------------

Description

A dataset containing the longterm averages (from 01-01-1999 to 31-12-2014) of measurements.

Usage

```
data(knmi.langJarigGem)
```

Format

A data frame with 13346 rows and 11 variables:

- station: ID of the station (210–391)
- dagVjaar: day of the year (1-366)
- gemTemp:
- gemMaxTemp:
- absMaxTemp:
- gemMinTemp:
- absMinTemp:
- gemNeerslag:
- gemZon:
- gemPercZon:
- gemStraling:

plot_KNMI_stations	<i>plot_KNMI_stations</i>
--------------------	---------------------------

Description

This function plots a map of The Netherlands and shows the locations of the KNMI measurement station and their id and name. One can show the active stations (active = TRUE, the default) or *all* stations (active = FALSE).

Usage

```
plot_KNMI_stations(active = TRUE)
```

Arguments

`active` boolean to select only currently active stations. Default = TRUE.

Value

data-frame with the id, name, url to station information and the lat/lon of the nearest KNMI-station.

<code>printStations</code>	<i>print KNMI measurement stations</i>
----------------------------	--

Description

Simple helper function to print the id's and names of the KNMI measurement stations and if they are active or not to the console.

Usage

```
printStations(active = TRUE)
```

Arguments

`active` boolean to select only currently active stations. Default = TRUE.

Format

The returned data-frame contains the following columns:

- `station` = ID of measurement station;
- `plaats` = city closest to the measurement station;
- `active` = indicates if the station is still active;

Value

a data-frame.

<code>renameColumnsKNMIdata</code>	<i>rename Columns KNMI data</i>
------------------------------------	---------------------------------

Description

This function provides meaningfull names for the different variables.

Usage

```
renameColumnsKNMIdata(data)
```

Arguments

<code>data</code>	data-frame with KNMI-data that has been obtained with the function <code>'getClimateDateSet()'</code> , <code>'getClimateDateInBulk()'</code> or <code>'subsetKNMIdata()'</code> .
-------------------	--

Value

data-frame met subset van de KNMI-data.

<code>stations</code>	<i>KNMI stations</i>
-----------------------	----------------------

Description

A dataset containing the ID's and meta-data on the official KNMI measurement stations.

Usage

```
data(stations)
```

Format

A data frame with 35 rows and 4 variables:

- station: ID of the station (210–391)
- plaats: City where the station is located
- lon: geographical longitude (format: Decimal Degrees DDD.DDDDD°)
- lat: geographical latitude (format: Decimal Degrees DDD.DDDDD°)

subsetKNMIdata	<i>subset KNMI data-set</i>
----------------	-----------------------------

Description

This function returns a filtered subset of the KNMI data-set. The filtering is two-fold; select the most-used variables from the complete set of variables and include the data from the provided start-year upward to the most current date. More importantly, it modifies the column values to SI-units and it adds some helper-columns, like day-of-the-year, weeknumber, daynumner, year.

Usage

```
subsetKNMIdata(data, startyear = 2006, variabeles = c("FG", "TG", "TN",
  "TX", "SQ", "SP", "Q", "RH", "NG"))
```

Arguments

data	data-frame with KNMI-data that has been obtained with the function 'getClimateDateSet' or 'getClimateDateInBulk'.
startyear	start-year for the selection. Default is 2006. Note that the end-year is always the most current year in the data-set.
variabeles	list with variables that should be returned. Default is ("FG","TG","TN","TX","SQ","SP","Q","RH","NG").

Format

The default data frame contains the following columns:

- stationID = ID of measurementstation;
- datum = Datum (YYYY=jaar MM=maand DD=dag);
- FG : gemWind = Etmaalgemiddelde windsnelheid (in m/s);
- TG : gemTemp = Etmaalgemiddelde temperatuur (in graden Celsius);
- TN : minTemp = Minimum temperatuur (in graden Celsius);
- TX : maxTemp = Maximum temperatuur (in graden Celsius);
- SQ : zon = Zonneschijnduur (in uur) berekend uit de globale straling (-1 voor <0.05 uur);
- SP : percZon = Percentage van de langst mogelijke zonneschijnduur;
- Q : straling = Globale straling (in J/cm2);
- RH : dagTotaalNeerslag = Etmaalsom van de neerslag (in mm) (-1 voor <0.05 mm);
- NG : gemBewolking = Etmaalgemiddelde bewolking (bedekkingsgraad van de bovenlucht in achtsten, 9=bovenlucht onzichtbaar);

Value

data-frame met subset van de KNMI-data.

voegToe	<i>this function adds the VE-index to the original data-frame. See berekenVEIndex for more details.</i>
---------	---

Description

this function adds the VE-index to the original data-frame. See berekenVEIndex for more details.

Usage

```
voegToe(dgg, indexType, startDate = paste(format(Sys.Date(), format = "%Y"),
      "01-01", sep = "-"), endDate = paste(format(Sys.Date(), format = "%Y"),
      "12-31", sep = "-"))
```

Arguments

dgg	dataframe containing two mandatory columns; the day-of-the-year (dagVjaar), the daily mean temperature (gemTemp) and an optional column containing the year (jaar). Temperatures in degrees Celcius. The data frame can contain the ranges for multiple years.
startDate	the start of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the start of the year.
endDate	the end of the summation interval, formatted as a string 'yyyy-mm-dd'. Defaults to the end of the year.

Value

dataframe containing the year (jaar), the day-of-the-year (dagVjaar) and the VE-index at the dagVjaar.