

Package ‘rsinaica’

February 8, 2024

Type Package

Title Download Data from Mexico's Air Quality Information System

Version 1.0.0

Description Easy-to-use functions for downloading air quality data from the Mexican National Air Quality Information System (SINAICA). Allows you to query pollution and meteorological parameters from more than a hundred monitoring stations located throughout Mexico. See <https://sinaica.inecc.gob.mx> for more information.

License MIT + file LICENSE

URL <https://hoyodesmog.diegovalle.net/rsinaica/>,
<https://github.com/diegovalle/rsinaica>

BugReports <https://github.com/diegovalle/rsinaica/issues>

Depends R (>= 3.2)

Imports dplyr, httr, jsonlite, lubridate, stats, stringr, utils

Suggests knitr, rmarkdown, testthat

Encoding UTF-8

LazyData TRUE

RoxygenNote 7.3.1

NeedsCompilation no

Author Diego Valle-Jones [aut, cre]

Maintainer Diego Valle-Jones <diego@diegovalle.net>

Repository CRAN

Date/Publication 2024-02-08 08:10:07 UTC

R topics documented:

params_sinaica	2
sinaica_param_data	2
sinaica_station_data	4

sinaica_station_dates	6
sinaica_station_params	7
stations_sinaica	7

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

params_sinaica	<i>Valid air quality parameters</i>
----------------	-------------------------------------

Description

Valid air quality parameters

Format

A data frame with 55 rows and 2 variables:

param_code Abbreviation of the air quality parameter

param_name Name of the air quality parameter

Source

SINAICA

Examples

```
head(params_sinaica)
```

sinaica_param_data	<i>Get air quality data from all stations by parameter</i>
--------------------	--

Description

Download data from all stations for a single parameter by specifying a date range

Usage

```
sinaica_param_data(
  parameter,
  start_date,
  end_date,
  type = "Crude",
  remove_extremes = FALSE
)
```

Arguments

parameter	<p>type of parameter to download</p> <ul style="list-style-type: none"> • BEN - Benceno • CH4 - Metano • CN - Carbono negro • CO - Monóxido de carbono • CO2 - Dióxido de carbono • DV - Dirección del viento • H2S - Acido Sulfhídrico • HCNM - Hidrocarburos no metánicos • HCT - Hidrocarburos Totales • HR - Humedad relativa • HRI - Humedad relativa interior • IUV - Índice de radiación ultravioleta • NO - Óxido nítrico • NO2 - Dióxido de nitrógeno • NOx - Óxidos de nitrógeno • O3 - Ozono • PB - Presión Barométrica • PM10 - Partículas menores a 10 micras • PM2.5 - Partículas menores a 2.5 micras • PP - Precipitación pluvial • PST - Partículas Suspendidas totales • RS - Radiación solar • SO2 - Dióxido de azufre • TMP - Temperatura • TMPI - Temperatura interior • UVA - Radiación ultravioleta A • VV - Radiación ultravioleta B • XIL - Xileno
start_date	start of range in YYYY-MM-DD format
end_date	end of range from which to download data in YYYY-MM-DD format
type	<p>The type of data to download. One of the following:</p> <ul style="list-style-type: none"> • Crude - Crude data that has not been validated • Manual - Manually collected data that is sent to an external lab for analysis (may no be collected daily). Mostly used for suspend particles collected by pushing air through a filter which is later sent to a lab to be weighted
remove_extremes	<p>whether to remove extreme values. For O3 all values above .2 are set to NA, for PM10 those above 600, for PM2.5 above 175, for NO2 above .21, for SO2 above .2, and for CO above 15. This is done so that the values match exactly those of the SINAICA website, but it is recommended that you use a more complicated statistical procedure to remove outliers.</p>

Value

data.frame with a column named *value* containing the air quality parameter values. If the data was validated the column named *date_validated* will contain the validation date. Care should be taken when working with hourly data since each station has their own timezone (available in the [stations_sinaica](#) data.frame) and some stations reported the timezone in which they are located erroneously.

Examples

```
## Not run:  
## May take several seconds  
df <- sinaica_param_data("03", "2015-10-14", "2015-10-14")  
head(df)  
  
## End(Not run)
```

sinaica_station_data *Get air quality data from a single measuring station*

Description

Download data from a single station by specifying a parameter and a date range

Usage

```
sinaica_station_data(  
  station_id,  
  parameter,  
  start_date,  
  end_date,  
  type = "Crude",  
  remove_extremes = FALSE  
)
```

Arguments

station_id	the numeric code corresponding to each station. See stations_sinaica for a list of stations and their ids.
parameter	type of parameter to download <ul style="list-style-type: none">• BEN - Benceno• CH4" - Metano• CN - Carbono negro• CO - Monóxido de carbono• CO2 - Dióxido de carbono• DV - Dirección del viento• H2S - Acido Sulfhídrico

	<ul style="list-style-type: none"> • HCNM - Hidrocarburos no metánicos • HCT - Hidrocarburos Totales • HR - Humedad relativa • HRI - Humedad relativa interior • IUV - Índice de radiación ultravioleta • NO - Óxido nítrico • NO2 - Dióxido de nitrógeno • NOx - Óxidos de nitrógeno • O3 - Ozono • PB - Presión Barométrica • PM10 - Partículas menores a 10 micras • PM2.5 - Partículas menores a 2.5 micras • PP - Precipitación pluvial • PST - Partículas Suspendidas totales • RS - Radiación solar • SO2 - Dióxido de azufre • TMP - Temperatura • TMPI - Temperatura interior • UVA - Radiación ultravioleta A • VV - Radiación ultravioleta B • XIL - Xileno
start_date	start of range in YYYY-MM-DD format
end_date	end of range from which to download data in YYYY-MM-DD format
type	The type of data to download. One of the following: <ul style="list-style-type: none"> • Crude - Crude data that has not been validated • Validated - data which has undergone a validation process during which it was cleaned, verified, and validated • Manual - Manually collected data that is sent to an external lab for analysis (may not be collected daily). Mostly used for suspended particles collected by pushing air through a filter which is later sent to a lab to be weighted
remove_extremes	whether to remove extreme values. For O3 all values above .2 are set to NA, for PM10 those above 600, for PM2.5 above 175, for NO2 above .21, for SO2 above .2, and for CO above 15. This is done so that the values match exactly those of the SINAICA website, but it is recommended that you use a more complicated statistical procedure to remove outliers.

Value

data.frame with air quality data. Care should be taken when working with hourly data since each station has their own timezone (available in the [stations_sinaica](#) data.frame) and some stations reported the timezone in which they are located erroneously.

See Also

Crude data comes from <https://sinaica.inecc.gob.mx/data.php>, validated data from <https://sinaica.inecc.gob.mx/data.php?tipo=V>, and manual data from <https://sinaica.inecc.gob.mx/data.php?tipo=M>

Examples

```
stations_sinaica[which(stations_sinaica$station_name == "Xalostoc"), 1:5]
df <- sinaica_station_data(271, "03", "2015-09-11", "2015-09-11", "Crude")
head(df)
```

sinaica_station_dates *Dates supported by a station*

Description

Start date and end date of the range for which SINAICA has data for an air quality station

Usage

```
sinaica_station_dates(station_id, type = "Crude")
```

Arguments

station_id	the numeric code corresponding to each station. See stations_sinaica for a list of stations and their ids.
type	The type of data to download. One of the following: <ul style="list-style-type: none">• Crude - Crude data that has not been validated• Validated - Validated data (may not be the most up-to-date)• Manual - Manual data

Value

a vector containing the date the station started reporting and end reporting date

Examples

```
## id 271 is Xalostoc. See `stations_sinaica`
df <- sinaica_station_dates(271, "Manual")
head(df)
```

`sinaica_station_params`*Parameters supported by a station*

Description

List of air quality parameters of a measuring station for which SINAICA has data

Usage

```
sinaica_station_params(station_id, type = "Crude")
```

Arguments

<code>station_id</code>	the numeric code corresponding to each station. See stations_sinaica for a list of stations and their ids.
<code>type</code>	The type of data to download. One of the following: <ul style="list-style-type: none">• Crude - Crude data that has not been validated• Validated - Validated data (may not be the most up-to-date)• Manual - Manual data

Value

a data.frame with the parameters supported by the station

Examples

```
## id 271 is Xalostoc. See `stations_sinaica`  
df <- sinaica_station_params(271, "Crude")  
head(df)
```

`stations_sinaica`*Air quality measuring stations in Mexico*

Description

This data set contains all the stations that report to the National Air Quality Information System **SINAICA**.

Usage

```
data(stations_sinaica)
```

Format

A data frame with 341 rows and 26 variables:

station_id Numeric code of the station
station_name Name of the station
station_code Abbreviation of the station
network_id Numeric code for the network
network_name Name of the network
network_code Abbreviation of the network
street street
ext exterior number
interior interior number
colonia colonia
zip zip code
state_code state code
municipio_code municipio code
year_started date the station started operations
altitude altitude in meters
address address
date_validated last date the station was validated
date_validated2 second to last date the station was validated
passed_validation did the station pass validation
video link to video of the station
lat latitude
lon longitude
date_started date the station started operations
timezone time zone in which the station is located (may contain errors)
street_view link to Google Street View
video_interior link to video of the interior of the station

Source

[SINAICA](#) ans Solicitud de Información 1612100005118

Examples

```
head(stations_sinaica)
```


Index

`params_sinaica`, [2](#)

`sinaica_param_data`, [2](#)

`sinaica_station_data`, [4](#)

`sinaica_station_dates`, [6](#)

`sinaica_station_params`, [7](#)

`stations_sinaica`, [4-7](#), [7](#)