

Package: cranology (via r-universe)

June 28, 2024

Title The CRAN Chronology

Version 0.1.0

Description Scraping routines and datasets to monitor the evolution of the number of packages on CRAN.

License MIT + file LICENSE

Depends R (>= 2.10)

Imports cli, devtools, dplyr, furr, ggplot2, lubridate, magrittr, purrr, readr, rvest, stringi, stringr, utils

Suggests checkhelper (>= 0.0.0.9000), knitr, testthat, usethis, withr

VignetteBuilder knitr

Remotes thinkr-open/checkhelper

Encoding UTF-8

LazyData true

RoxygenNote 7.2.0

Repository <https://thinkr-open.r-universe.dev>

RemoteUrl <https://github.com/ThinkR-open/cranology>

RemoteRef HEAD

RemoteSha c221bc4e5981cac33cedbf6e0f6b15c48d9237d8

Contents

cran_monthly_package_number	2
cran_packages_history	2
get_package_first_release	3
mran	3
plot_cran_monthly_package_number	4
scrape_cran_history	4
update_monthly_package_number	5

Index	6
--------------	----------

```
cran_monthly_package_number
      cran_monthly_package_number
```

Description

The evolution of the number of packages on CRAN since its beginning. Last update: 2022-06-29.

Usage

```
cran_monthly_package_number
```

Format

A data frame with 297 rows and 2 variables:

date [Date] Month of release.
number_packages [numeric] Number of packages available on CRAN at that given date.

Source

<https://cran.rstudio.com/src/contrib/> and <https://mran.microsoft.com/>

```
cran_packages_history  cran_packages_history
```

Description

All packages ever available on CRAN. Last update: 2022-06-29.

Usage

```
cran_packages_history
```

Format

A data frame with 312 rows and 10 variables:

file_name [character] Either the name of the .tar.gz or the name of the archive folder holding the .tar.gzs of all versions ever released of a given package.
date [POSIXct,POSIXt] The date of upload on CRAN.
time [character] The time of upload on CRAN.
size [character] The size of the .tar.gzs. '-' in case of archive folder.
package_name [character] The name of the package.
last_archived [POSIXct,POSIXt] The date when one version was last archived.

archive [logical] Was a version ever archived ?
first_date [POSIXct,POSIXt] The date of the first release.
n_versions [integer] The number of versions released.
last_modified [POSIXct,POSIXt] The date of last release.

Source

<https://cran.rstudio.com/src/contrib/>

```
get_package_first_release
  Get first release of package
```

Description

Scrape every folder of the CRAN archive to retrieve both the date of the first release and the number of versions released for all archived packaged.

Usage

```
get_package_first_release(package_name)
```

Arguments

`package_name` A character string. The package name.

Value

A tibble with three columns: `_package_name_`, `_first_date_` and `_n_versions_`.

```
mran
  Get number of package on CRAN on a given date using MRAN
```

Description

This function queries MRAN to retrieve the number of package on CRAN on a given date.

Usage

```
get_package_number_mran(dates, parallelize = FALSE)
```

Arguments

`dates` A vector of dates. Either a character vector of the form "yyyy-mm-dd" or a vector of class "Date". All dates must be posterior to "2014-09-17", the day of MRAN first CRAN snapshot.

`parallelize` A logical. If TRUE furr is used to asynchronously scrap mran.

Value

A data.frame with two columns 'date' and 'n' the number of packages on CRAN at that given 'date'.

Examples

```
get_package_number_mran(c("2018-04-10", "2020-03-19"))
```

```
plot_cran_monthly_package_number
```

Plot monthly evolution of package number on CRAN

Description

This function is a convenience tool to quickly draw a line showing the evolution of packages number on CRAN since its beginning. It uses the 'cran_monthly_package_number' dataset.

Usage

```
plot_cran_monthly_package_number()
```

Value

A ggplot object

Examples

```
plot_cran_monthly_package_number()
```

```
scrape_cran_history
```

Scrape CRAN packages history

Description

This function is the workhorse of cranology. It scrapes <https://cran.rstudio.com> and generates two datasets:

Usage

```
scrape_cran_history()
```

Details

* 'cran_packages_history': A data.frame gathering information about every package that has ever been on CRAN including the first release date the number of versions released so far... * 'cran_monthly_package_number': A data.frame holding the number of packages available on CRAN since its beginning. Data is provided on a montly basis.

Value

A list of two data.frames: 'cran_packages_history' and 'cran_monthly_package_number'.

Examples

```
## Not run:
scrape_cran_history()

## End(Not run)
```

```
update_monthly_package_number
```

Update 'cran_monthly_package_number' dataset

Description

The creation 'cran_monthly_package_number' using 'scrape_cran()' is a long process as the underlying scrapping operations are time consuming. To more rapidly update 'cran_monthly_package_number' it is easier to rely on data from MRAN. This what this function does. It uses 'get_package_number_mran()' to quickly update the dataset.

Usage

```
update_monthly_package_number(
  cran_monthly_package_number_df,
  parallelize = FALSE
)
```

Arguments

```
cran_monthly_package_number_df
```

A data.frame similar to the 'cran_monthly_package_number' dataset included within cranology.

```
parallelize
```

A logical. If TRUE furr is used to asynchronously scrap mran.

Examples

```
# Simulate `cran_monthly_package_number` update
date_lag <- 3
df <- cran_monthly_package_number[
  1:(nrow(cran_monthly_package_number) - date_lag),
]
update_monthly_package_number(
  cran_monthly_package_number_df = df
)
```

Index

* datasets

- cran_monthly_package_number, [2](#)
- cran_packages_history, [2](#)

cran_monthly_package_number, [2](#)
cran_packages_history, [2](#)

get_package_first_release, [3](#)
get_package_number_mran (mran), [3](#)

mran, [3](#)

plot_cran_monthly_package_number, [4](#)

scrape_cran_history, [4](#)

update_monthly_package_number, [5](#)