Instruction_of_downloading_GEO_dataset

hank yao

2023-11-27

Installation of GEOquery

```r
# Install GEOquery with fix of timeout bug
# by using function(install_github()) from 'remotes' package.
library(remotes)
install_github("seandavi/GEOquery", force = TRUE)
```

Approach 1: From github of GEOquery's author

```
## Downloading GitHub repo seandavi/GEOquery@HEAD

##
## -- R CMD build -----------------------------------------------------------------
##      checking for file '/private/var/folders/tg/l8s3lr394p99bs171s50q_n80000gn/T/Rtmp7Ugloh/remotes1
##    - preparing 'GEOquery':
##      checking DESCRIPTION meta-information ...  v  checking DESCRIPTION meta-information
##    - checking for LF line-endings in source and make files and shell scripts
##    - checking for empty or unneeded directories
##    - building 'GEOquery_2.69.1.tar.gz'
##
##
```

```r
# Download modified GEOquery package from my github
# by using function(install_github()) from 'remotes' package.
library(remotes)
install_github("curryhank08/GEOquery_with_modifiable_timeout_seconds", force = TRUE)
```

Approach 2: From my github

```
## Downloading GitHub repo curryhank08/GEOquery_with_modifiable_timeout_seconds@HEAD

##
## -- R CMD build -----------------------------------------------------------------
##      checking for file '/private/var/folders/tg/l8s3lr394p99bs171s50q_n80000gn/T/Rtmp7Ugloh/remotes1
##    - preparing 'GEOquery':
##      checking DESCRIPTION meta-information ...  v  checking DESCRIPTION meta-information
##    - checking for LF line-endings in source and make files and shell scripts
##    - checking for empty or unneeded directories
##    - building 'GEOquery_2.65.2.tar.gz'
##
##
```

Load GEOquery

```r
# Load modified GEOquery or GEOquery 2.69.1/2.70.0up version
library(GEOquery)
```

```
## Loading required package: Biobase
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Setting options('download.file.method.GEOquery'='auto')
```

```
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```r
# Setting the max timeout_seconds (Whatever you want to set)
options(timeout=100000)
# Check the input timeout_seconds
getOption("timeout")
```

```
## [1] 1e+05
```

Download GEO dataset through accession id

```r
# Download GSE40279 by a fuction getGEO() from GEOquery package.
gse30870 <- getGEO("GSE30870", GSEMatrix = TRUE, AnnotGPL = TRUE)
```

```
## Found 1 file(s)
```

```
## GSE30870_series_matrix.txt.gz
```

```
## Annotation GPL not available, so will use submitter GPL instead
```

```r
gse30870_matrix <- gse30870[[1]]
expression_data <- exprs(gse30870_matrix)
#head(expression_data)
```