Reading & Writing Data with R

Evans_G02504972

2023-10-20

```r
library("readr")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```r
library(bench)
library("ggplot2")
```

```r
table <- SwiftSongs %>%
  slice_tail(n = 13) %>%
  select(30:34)
```

```
## Error in eval(expr, envir, enclos): object 'SwiftSongs' not found
```

```r
table
```

```
## function (..., exclude = if (useNA == "no") c(NA, NaN), useNA = c("no",
##     "ifany", "always"), dnn = list.names(...), deparse.level = 1)
## {
##     list.names <- function(...) {
```

```
##         l <- as.list(substitute(list(...)))[-1L]
##         if (length(l) == 1L && is.list(..1) && !is.null(nm <- names(..1)))
##             return(nm)
##         nm <- names(l)
##         fixup <- if (is.null(nm))
##             seq_along(l)
##         else nm == ""
##         dep <- vapply(l[fixup], function(x) switch(deparse.level +
##             1, "", if (is.symbol(x)) as.character(x) else "",
##             deparse(x, nlines = 1)[1L]), "")
##         if (is.null(nm))
##             dep
##         else {
##             nm[fixup] <- dep
##             nm
##         }
##     }
##     miss.use <- missing(useNA)
##     miss.exc <- missing(exclude)
##     useNA <- if (miss.use && !miss.exc && !match(NA, exclude,
##         nomatch = 0L))
##         "ifany"
##     else match.arg(useNA)
##     doNA <- useNA != "no"
##     if (!miss.use && !miss.exc && doNA && match(NA, exclude,
##         nomatch = 0L))
##         warning("’exclude’ containing NA and ’useNA’ != \"no\"’ are a bit contradicting")
##     args <- list(...)
##     if (length(args) == 1L && is.list(args[[1L]])) {
##         args <- args[[1L]]
##         if (length(dnn) != length(args))
##             dnn <- paste(dnn[1L], seq_along(args), sep = ".")
##     }
##     if (!length(args))
##         stop("nothing to tabulate")
##     bin <- 0L
##     lens <- NULL
##     dims <- integer()
##     pd <- 1L
##     dn <- NULL
##     for (a in args) {
##         if (is.null(lens))
##             lens <- length(a)
##         else if (length(a) != lens)
##             stop("all arguments must have the same length")
##         fact.a <- is.factor(a)
##         if (doNA)
##             aNA <- anyNA(a)
##         if (!fact.a) {
##             a0 <- a
##             op <- options(warn = 2)
##             on.exit(options(op))
##             a <- factor(a, exclude = exclude)
##             options(op)
##         }
##         add.na <- doNA
##         if (add.na) {
##             ifany <- (useNA == "ifany")
##             anNAc <- anyNA(a)
##             add.na <- if (!ifany || anNAc) {
##                 ll <- levels(a)
##                 if (add.ll <- !anyNA(ll)) {
##                   ll <- c(ll, NA)
##                   TRUE
##                 }
##                 else if (!ifany && !anNAc)
##                   FALSE
##                 else TRUE
##             }
##             else FALSE
##         }
##         if (add.na)
##             a <- factor(a, levels = ll, exclude = NULL)
##         else ll <- levels(a)
##         a <- as.integer(a)
##         if (fact.a && !miss.exc) {
##             ll <- ll[keep <- which(match(ll, exclude, nomatch = 0L) ==
##                 0L)]
##             a <- match(a, keep)
##         }
##         else if (!fact.a && add.na) {
##             if (ifany && !aNA && add.ll) {
##                 ll <- ll[!is.na(ll)]
##                 is.na(a) <- match(a0, c(exclude, NA), nomatch = 0L) >
##                   0L
##             }
##             else {
##                 is.na(a) <- match(a0, exclude, nomatch = 0L) >
##                   0L
##             }
##         }
##         nl <- length(ll)
##         dims <- c(dims, nl)
##         if (prod(dims) > .Machine$integer.max)
##             stop("attempt to make a table with >= 2^31 elements")
##         dn <- c(dn, list(ll))
##         bin <- bin + pd * (a - 1L)
##         pd <- pd * nl
##     }
##     names(dn) <- dnn
##     bin <- bin[!is.na(bin)]
##     if (length(bin))
##         bin <- bin + 1L
##     y <- array(tabulate(bin, pd), dims, dimnames = dn)
##     class(y) <- "table"
##     y
## }
## <bytecode: 0x13eb833d8>
## <environment: namespace:base>
```

2(c)Import data on Taylor Swift songs directly from the URL using the read.csv(), read_csv(), and the fread() functions, comparing the read times using the mark function from the bench package, storing the results of the mark function in an object called readTimes. Specify a minimum of 5 iterations in the mark function.

```
url <- "https://raw.githubusercontent.com/dilernia/STA418-518/main/Data/swiftSongs.csv"
swift_songs <- read.csv(url)

readTimes <- mark(
  read.csv = {
    bench::mark(
      read.csv(url),
      min_iterations = 5
    )
  },
  read_csv = {
    bench::mark(
      read_csv(url),
      min_iterations = 5
    )
  },
  fread = {
    bench::mark(
      fread(url),
      min_iterations = 5
    )
  },
  check = FALSE
)
```

```
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
```

```
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
## Rows: 151 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): youtube_title, youtube_description, youtube_duration, youtube_url...
## dbl  (22): youtube_view_count, youtube_like_count, youtube_favorite_count, y...
## lgl   (1): explicit
## dttm  (1): youtube_publish_date
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
## Warning: Some expressions had a GC in every iteration; so filtering is
## disabled.
```

```
#  results
readTimes
```

```
## # A tibble: 3 x 6
##   expression      min   median ‘itr/sec‘ mem_alloc ‘gc/sec‘
##   <bch:expr> <bch:tm> <bch:tm>     <dbl> <bch:byt>    <dbl>
## 1 read.csv      1.18s    1.18s     0.846      400B     0
## 2 read_csv   685.16ms 685.16ms     1.46       400B     1.46
## 3 fread      559.03ms 559.03ms     1.79       400B     1.79
```