R - Practice 02 - v1.1

Sangkon Han(sangkon@pusan.ac.kr)

2023-09-10

Contents

Data import (tibble & readr)

Table called tibble

```
# some build in tibbles (inside ggplot2)
ggplot2::economics
```

```
## # A tibble: 574 x 6
##    date         pce    pop psavert uempmed unemploy
##    <date>     <dbl>  <dbl>   <dbl>   <dbl>    <dbl>
##  1 1967-07-01  507. 198712    12.6     4.5     2944
##  2 1967-08-01  510. 198911    12.6     4.7     2945
##  3 1967-09-01  516. 199113    11.9     4.6     2958
##  4 1967-10-01  512. 199311    12.9     4.9     3143
##  5 1967-11-01  517. 199498    12.8     4.7     3066
##  6 1967-12-01  525. 199657    11.8     4.8     3018
##  7 1968-01-01  531. 199808    11.7     5.1     2878
##  8 1968-02-01  534. 199920    12.3     4.5     3001
##  9 1968-03-01  544. 200056    11.7     4.1     2877
## 10 1968-04-01  544  200208    12.3     4.6     2709
## # i 564 more rows
```

```
class(ggplot2::economics)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

```
ggplot2::diamonds
```

```
## # A tibble: 53,940 x 10
##    carat cut       color clarity depth table price     x     y     z
##    <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
##  1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
##  2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
##  3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
##  4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
##  5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
##  6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
##  7  0.24 Very Good I     VVS1     62.3    57   336  3.95  3.98  2.47
##  8  0.26 Very Good H     SI1      61.9    55   337  4.07  4.11  2.53
##  9  0.22 Fair      E     VS2      65.1    61   337  3.87  3.78  2.49
## 10  0.23 Very Good H     VS1      59.4    61   338  4     4.05  2.39
## # i 53,930 more rows
```

```
class(ggplot2::diamonds)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
ggplot2::faithfuld %>% head()
```

```
## # A tibble: 6 x 3
##   eruptions waiting density
##       <dbl>   <dbl>   <dbl>
## 1      1.6       43 0.00322
## 2      1.65      43 0.00384
## 3      1.69      43 0.00444
## 4      1.74      43 0.00498
## 5      1.79      43 0.00542
## 6      1.84      43 0.00574
```

```
class(ggplot2::faithfuld)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
# hflights data set  (not a tibble - can be converted, shown later)
hflights::hflights %>% head()
```

```
##      Year Month DayofMonth DayOfWeek DepTime ArrTime UniqueCarrier FlightNum
## 5424 2011     1          1         6    1400    1500            AA       428
## 5425 2011     1          2         7    1401    1501            AA       428
## 5426 2011     1          3         1    1352    1502            AA       428
## 5427 2011     1          4         2    1403    1513            AA       428
## 5428 2011     1          5         3    1405    1507            AA       428
## 5429 2011     1          6         4    1359    1503            AA       428
##      TailNum ActualElapsedTime AirTime ArrDelay DepDelay Origin Dest Distance
## 5424  N576AA                60      40      -10        0    IAH  DFW      224
## 5425  N557AA                60      45       -9        1    IAH  DFW      224
## 5426  N541AA                70      48       -8       -8    IAH  DFW      224
## 5427  N403AA                70      39        3        3    IAH  DFW      224
## 5428  N492AA                62      44       -3        5    IAH  DFW      224
## 5429  N262AA                64      45       -7       -1    IAH  DFW      224
##      TaxiIn TaxiOut Cancelled CancellationCode Diverted
## 5424      7      13         0                         0
## 5425      6       9         0                         0
## 5426      5      17         0                         0
## 5427      9      22         0                         0
```

```
## 5428      9       9         0                         0
## 5429      6      13         0                         0
```

```
class(hflights::hflights)
```

```
## [1] "data.frame"
```

Create a tibble

as_tibble() - From data frame (conversion)

```
# Convert hflights data frame
class(hflights)
```

```
## [1] "data.frame"
```

```
dft <- as_tibble(hflights)
class(dft)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
# Convert custom data frame
df <- data.frame(x = 1:10,
                 y = seq.Date(from = as.Date("2021-01-01"),
                              to = as.Date("2021-01-10"),
                              by = "day"))
class(df)
```

```
## [1] "data.frame"
```

```
dft <- as_tibble(df)
class(dft)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
# tibble() - Create custom tibble
tibble(v1 = seq(from = 1, to = 100, by = 1),
       v2 = pi,
       v3 = sqrt(v1),
       v4 = seq.Date(from = as.Date("2021-01-01"), length.out = 100, by = "day")) %>%
  head()
```

```
## # A tibble: 6 x 4
##      v1    v2    v3 v4
##   <dbl> <dbl> <dbl> <date>
## 1     1  3.14  1    2021-01-01
## 2     2  3.14  1.41 2021-01-02
## 3     3  3.14  1.73 2021-01-03
## 4     4  3.14  2    2021-01-04
## 5     5  3.14  2.24 2021-01-05
## 6     6  3.14  2.45 2021-01-06
```

```
# Use strange non-syntactic column names
tibble(`123` = 123, `.` = "period", `,` = "comma", `,*/-+?!` = "strange name")
```

```
## # A tibble: 1 x 4
##   `123` .      `,`   `,*/-+?!`
##   <dbl> <chr>  <chr> <chr>
## 1   123 period comma strange name
```

```r
# tribble() - Create transposed tibble
tribble(
  ~name,  ~surname,  ~male,  ~age, # header
  #--------------------------------#
  "Max",  "Smith",   T,      35,
  "Lily", "Brown",   F,      27
  )
```

```
## # A tibble: 2 x 4
##   name  surname male    age
##   <chr> <chr>   <lgl> <dbl>
## 1 Max   Smith   TRUE     35
## 2 Lily  Brown   FALSE    27
```

data.frame VS tibbles

```r
# Print output
hflights %>% as_tibble()
```

```
## # A tibble: 227,496 x 21
##     Year Month DayofMonth DayOfWeek DepTime ArrTime UniqueCarrier FlightNum
##    <int> <int>      <int>     <int>   <int>   <int> <chr>             <int>
##  1  2011     1          1         6    1400    1500 AA                  428
##  2  2011     1          2         7    1401    1501 AA                  428
##  3  2011     1          3         1    1352    1502 AA                  428
##  4  2011     1          4         2    1403    1513 AA                  428
##  5  2011     1          5         3    1405    1507 AA                  428
##  6  2011     1          6         4    1359    1503 AA                  428
##  7  2011     1          7         5    1359    1509 AA                  428
##  8  2011     1          8         6    1355    1454 AA                  428
##  9  2011     1          9         7    1443    1554 AA                  428
## 10  2011     1         10         1    1443    1553 AA                  428
## # i 227,486 more rows
## # i 13 more variables: TailNum <chr>, ActualElapsedTime <int>, AirTime <int>,
## #   ArrDelay <int>, DepDelay <int>, Origin <chr>, Dest <chr>, Distance <int>,
## #   TaxiIn <int>, TaxiOut <int>, Cancelled <int>, CancellationCode <chr>,
## #   Diverted <int>
```

```r
hflights %>% head()
```

```
##      Year Month DayofMonth DayOfWeek DepTime ArrTime UniqueCarrier FlightNum
## 5424 2011     1          1         6    1400    1500            AA       428
## 5425 2011     1          2         7    1401    1501            AA       428
## 5426 2011     1          3         1    1352    1502            AA       428
## 5427 2011     1          4         2    1403    1513            AA       428
## 5428 2011     1          5         3    1405    1507            AA       428
## 5429 2011     1          6         4    1359    1503            AA       428
##      TailNum ActualElapsedTime AirTime ArrDelay DepDelay Origin Dest Distance
## 5424  N576AA                60      40      -10        0    IAH  DFW      224
## 5425  N557AA                60      45       -9        1    IAH  DFW      224
## 5426  N541AA                70      48       -8       -8    IAH  DFW      224
## 5427  N403AA                70      39        3        3    IAH  DFW      224
## 5428  N492AA                62      44       -3        5    IAH  DFW      224
## 5429  N262AA                64      45       -7       -1    IAH  DFW      224
```

```
##      TaxiIn TaxiOut Cancelled CancellationCode Diverted
## 5424      7      13         0                        0
## 5425      6       9         0                        0
## 5426      5      17         0                        0
## 5427      9      22         0                        0
## 5428      9       9         0                        0
## 5429      6      13         0                        0
```

Subsetting

```
# Extract by name - $
mpg$manufacturer %>% head()
```

```
## [1] "audi" "audi" "audi" "audi" "audi" "audi"
```

```
# Extract by name - [[]]
mpg[["manufacturer"]] %>% head()
```

```
## [1] "audi" "audi" "audi" "audi" "audi" "audi"
```

```
# Extract by position - [[]]
mpg[[1]] %>% head()
```

```
## [1] "audi" "audi" "audi" "audi" "audi" "audi"
```

```
# Using with a pipe operator
mpg %>% .$manufacturer %>% head()
```

```
## [1] "audi" "audi" "audi" "audi" "audi" "audi"
```

```
mpg %>% .[["manufacturer"]] %>% head()
```

```
## [1] "audi" "audi" "audi" "audi" "audi" "audi"
```

Read files

```
# Read inline csv file
read_csv("c1,c2,c3
          1,a,T,
          2,b,T,
          3,c,F")
```

```
## # A tibble: 3 x 3
##      c1 c2    c3
##   <dbl> <chr> <chr>
## 1     1 a     T,
## 2     2 b     T,
## 3     3 c     F
```

```
#   Inline files with a meta header lines
read_csv("First meta line
          Second meta line
          c1,c2,c3
          1,a,T,
          2,b,T,
          3,c,F", skip = 2)
```

```
## # A tibble: 3 x 3
```

```
##      c1 c2    c3
##   <dbl> <chr> <chr>
## 1     1 a     T,
## 2     2 b     T,
## 3     3 c     F
```

```
#   Inline files with comments
read_csv("c1,c2,c3 # comment
          1,a,T,   # comment
          2,b,T,
          3,c,F", comment = "#")
```

```
## # A tibble: 3 x 3
##      c1 c2    c3
##   <dbl> <chr> <chr>
## 1     1 a     T,
## 2     2 b     T,
## 3     3 c     F
```

Read comma separated files - .csv from your disk

```
# Small mpg table
df <- read_csv(file = "./data/mpg_mini.csv")
head(df)
```

```
## # A tibble: 6 x 4
##   manufacturer model             displ  year
##   <chr>        <chr>             <dbl> <dbl>
## 1 pontiac      grand prix          5.3  2008
## 2 toyota       toyota tacoma 4wd   4    2008
## 3 toyota       4runner 4wd         4.7  2008
## 4 audi         a4 quattro          3.1  2008
## 5 toyota       corolla             1.8  1999
## 6 subaru       impreza awd         2.5  2008
```

```
# Small mpg table (column separator = ";)
df <- read_csv2(file = "./data/mpg_mini2.csv")
head(df)
```

```
## # A tibble: 6 x 4
##   manufacturer model             displ  year
##   <chr>        <chr>             <dbl> <dbl>
## 1 pontiac      grand prix           53  2008
## 2 toyota       toyota tacoma 4wd     4  2008
## 3 toyota       4runner 4wd          47  2008
## 4 audi         a4 quattro           31  2008
## 5 toyota       corolla              18  1999
## 6 subaru       impreza awd          25  2008
```

```
# Read tab separated files - .tsv from your disk
df <- read_tsv(file = "./data/mpg.tsv")
head(df)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <dbl> <dbl> <chr>      <chr> <dbl> <dbl> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(15)   f        18    29 p     compa~
```

```
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# Read files with selected delimiter
df <- read_delim(file = "./data/mpg_delim.txt", delim = "~")
head(df)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <dbl> <dbl> <chr>      <chr> <dbl> <dbl> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# Read text file:
#   - " " as separator
#   - decimal separator ","
#   - quotations around strings
#   - meta lines in header
#   - empty lines
#   - missing values
df <- read_delim(file = "./data/mpg.txt",
                 col_names = T,
                 skip = 3,
                 skip_empty_rows = T,
                 delim = " ",
                 quote = "\"",
                 na = "")
head(df)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <dbl> <dbl> <chr>      <chr> <dbl> <dbl> <chr> <chr>
## 1 audi         a4       18  1999     4 auto(l5)   f        18    29 p     <NA>
## 2 audi         a4       18  1999     4 manual(m5) f        21    29 p     compa~
## 3 <NA>         a4        2  2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4        2  2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4       28  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4       28  1999     6 manual(m5) f        18    26 p     compa~
```

```
# Read log based file:
read_log(file = "./data/example.log") %>% head()
```

```
## # A tibble: 2 x 7
##   X1           X2    X3                   X4                    X5       X6    X7
##   <chr>        <lgl> <chr>                <chr>                 <chr> <dbl> <dbl>
## 1 172.21.13.45 NA    "Microsoft\\JohnDoe" 08/Apr/2001:17:39:0~ GET ~   200  3401
## 2 127.0.0.1    NA    "frank"              10/Oct/2000:13:55:3~ GET ~   200  2326
```

```
# Read large .csv file
#  - read.csv   VS   read_csv
```

```
#  - execution times
system.time(df <- read.csv(file = "./data/mpg_maxi.csv")) %>% head()
```

```
##  user.self   sys.self    elapsed user.child  sys.child
##       2.10       0.15       2.27         NA         NA
```

```
system.time(df <- read_csv(file = "./data/mpg_maxi.csv")) %>% head()
```

```
##  user.self   sys.self    elapsed user.child  sys.child
##       0.86       0.21       0.50         NA         NA
```

Vector parsing

```
# Parse character vector
parse_character(c("one", "two", "three"))
```

```
## [1] "one"   "two"   "three"
```

```
parse_character(c("one", "two", 3))
```

```
## [1] "one" "two" "3"
```

```
# Other type encoding
konnichiwa <- "\x82\xb1\x82\xf1\x82\xc9\x82\xbf\x82\xcd" # Japanese word
parse_character(konnichiwa) # UTF-8 encoding by default
```

```
## [1] "\x82\xb1\x82\xf1\x82\xbf\x82\xcd"
```

```
parse_character(konnichiwa, locale = locale(encoding = "Shift-JIS")) # switch encoding
```

```
## [1] "こんにちは"
```

```
# Parse logical vector
parse_logical(c("TRUE", "FALSE", "T", "F"))
```

```
## [1]  TRUE FALSE  TRUE FALSE
```

```
parse_logical(c("TRUE", "FALSE", "T", "F", "NA"))
```

```
## [1]  TRUE FALSE  TRUE FALSE    NA
```

```
# check parsing problems
x <- parse_logical(c("TRUE", "FALSE", "T", "F", "NA", "string"))
problems(x)
```

```
## # A tibble: 1 x 4
##     row   col expected           actual
##   <int> <int> <chr>              <chr>
## 1     6    NA 1/0/T/F/TRUE/FALSE string
```

```
# Parse integer vector
parse_integer(c("10", "20", "30", "40"))
```

```
## [1] 10 20 30 40
```

```
parse_integer(c("10", "20", "30", "40.5"))
```

```
## [1] 10 20 30 NA
## attr(,"problems")
## # A tibble: 1 x 4
##     row   col expected               actual
```

```
##   <int> <int> <chr>                  <chr>
## 1     4    NA no trailing characters 40.5
```

```
# Parse factor
parse_factor(c("one", "two", "one"))
```

```
## [1] one two one
## Levels: one two
```

```
parse_factor(c("one", "two", "one"), levels = c("two", "one"))
```

```
## [1] one two one
## Levels: two one
```

```
# Parse double vector
parse_double(c("11.7", "4.13"))
```

```
## [1] 11.70  4.13
```

```
# Different decimal mark
parse_double(c("11,7", "4,13"))
```

```
## [1] NA NA
## attr(,"problems")
## # A tibble: 2 x 4
##     row   col expected               actual
##   <int> <int> <chr>                  <chr>
## 1     1    NA no trailing characters 11,7
## 2     2    NA no trailing characters 4,13
```

```
parse_double(c("11,7", "4,13"), locale = locale(decimal_mark = ","))
```

```
## [1] 11.70  4.13
```

```
# Parse number
parse_number(c("1", "2.2", "$1000", "20%", "1,000"))
```

```
## [1]    1.0    2.2 1000.0   20.0 1000.0
```

```
# Grouping mark specified
parse_number(c("100,000.2"), locale = locale(grouping_mark = ","))
```

```
## [1] 100000.2
```

```
# Parse date
parse_date("2021-12-31") %>% head()
```

```
## [1] "2021-12-31"
```

```
# Specify date format
parse_date("20211231", "%Y%m%d")
```

```
## [1] "2021-12-31"
```

```
parse_date("21/12/31", "%y/%m/%d")
```

```
## [1] "2021-12-31"
```

```
# Parse time
parse_time("00:01")
```

```
## 00:01:00
```

```
parse_time("00:01 am")
```

```
## 00:01:00
```

```
parse_time("00:01:00")
```

```
## 00:01:00
```

```
# Parse datetime
parse_datetime("2021-12-31 00:01")
```

```
## [1] "2021-12-31 00:01:00 UTC"
```

File parsing

```
# Guess parser heuristic
guess_parser(c("T", "F"))
```

```
## [1] "logical"
```

```
guess_parser("2021-12-31")
```

```
## [1] "date"
```

```
guess_parser("2021-12-31 00:01")
```

```
## [1] "datetime"
```

```
guess_parser(c("5", "10"))
```

```
## [1] "double"
```

```
# Parse each column mpg table
read_tsv(file = "./data/mpg.tsv",
         col_types = cols(
           manufacturer = readr::col_factor(),
           model = readr::col_factor(),
           displ = col_double(),
           year = col_integer(),
           cyl = col_integer(),
           trans = col_character(),
           drv = col_character(),
           cty = col_number(),
           hwy = col_number(),
           fl = col_character(),
           class = col_character())) %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <fct>        <fct> <dbl> <int> <int> <chr>      <chr> <dbl> <dbl> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(15)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(15)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# Import table
#  - do not specify column types at import
```

```
#  - change column types inside R
read_tsv(file = "./data/mpg.tsv") %>%
  mutate_at(.vars = c("year", "cyl"), .funs = as.integer) %>%       # integer conversion
  mutate_at(.vars = c("manufacturer", "model"), .funs = as.factor) %>%   # factor conversion
  head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <fct>        <fct> <dbl> <int> <int> <chr>      <chr> <dbl> <dbl> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

Other useful import libraries

```
# readxl
read_excel(path = "./data/mpg.xlsx")
```

```
## # A tibble: 234 x 11
##    manufacturer model      displ  year   cyl trans drv     cty   hwy fl    class
##    <chr>        <chr>      <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr>
##  1 audi         a4           1.8  1999     4 auto~ f        18    29 p     comp~
##  2 audi         a4           1.8  1999     4 manu~ f        21    29 p     comp~
##  3 audi         a4           2    2008     4 manu~ f        20    31 p     comp~
##  4 audi         a4           2    2008     4 auto~ f        21    30 p     comp~
##  5 audi         a4           2.8  1999     6 auto~ f        16    26 p     comp~
##  6 audi         a4           2.8  1999     6 manu~ f        18    26 p     comp~
##  7 audi         a4           3.1  2008     6 auto~ f        18    27 p     comp~
##  8 audi         a4 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
##  9 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
## 10 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
## # i 224 more rows
```

```
read_excel(path = "./data/mpg.xlsx", sheet = "Sheet 1") # specify sheet
```

```
## # A tibble: 234 x 11
##    manufacturer model      displ  year   cyl trans drv     cty   hwy fl    class
##    <chr>        <chr>      <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr>
##  1 audi         a4           1.8  1999     4 auto~ f        18    29 p     comp~
##  2 audi         a4           1.8  1999     4 manu~ f        21    29 p     comp~
##  3 audi         a4           2    2008     4 manu~ f        20    31 p     comp~
##  4 audi         a4           2    2008     4 auto~ f        21    30 p     comp~
##  5 audi         a4           2.8  1999     6 auto~ f        16    26 p     comp~
##  6 audi         a4           2.8  1999     6 manu~ f        18    26 p     comp~
##  7 audi         a4           3.1  2008     6 auto~ f        18    27 p     comp~
##  8 audi         a4 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
##  9 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
## 10 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
## # i 224 more rows
```

```
read_excel(path = "./data/mpg.xlsx", range = "A1:C10")  # specify range
```

```
## # A tibble: 9 x 3
##   manufacturer model      displ
##   <chr>        <chr>      <dbl>
## 1 audi         a4           1.8
## 2 audi         a4           1.8
## 3 audi         a4           2
## 4 audi         a4           2
## 5 audi         a4           2.8
## 6 audi         a4           2.8
## 7 audi         a4           3.1
## 8 audi         a4 quattro   1.8
## 9 audi         a4 quattro   1.8
```

```
read_excel(path = "./data/mpg.xlsx") %>% class()
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
# rio
rio::import(file = "./data/mpg.xlsx") %>% head()
```

```
##   manufacturer model displ year cyl      trans drv cty hwy fl   class
## 1         audi    a4   1.8 1999   4   auto(l5)   f  18  29  p compact
## 2         audi    a4   1.8 1999   4 manual(m5)   f  21  29  p compact
## 3         audi    a4   2.0 2008   4 manual(m6)   f  20  31  p compact
## 4         audi    a4   2.0 2008   4   auto(av)   f  21  30  p compact
## 5         audi    a4   2.8 1999   6   auto(l5)   f  16  26  p compact
## 6         audi    a4   2.8 1999   6 manual(m5)   f  18  26  p compact
```

```
rio::import(file = "./data/mpg.xlsx") %>% class() %>% head()
```

```
## [1] "data.frame"
```

```
rio::import(file = "./data/mpg.xlsx", sheet = "Sheet 1") %>% head() # specify sheet
```

```
##   manufacturer model displ year cyl      trans drv cty hwy fl   class
## 1         audi    a4   1.8 1999   4   auto(l5)   f  18  29  p compact
## 2         audi    a4   1.8 1999   4 manual(m5)   f  21  29  p compact
## 3         audi    a4   2.0 2008   4 manual(m6)   f  20  31  p compact
## 4         audi    a4   2.0 2008   4   auto(av)   f  21  30  p compact
## 5         audi    a4   2.8 1999   6   auto(l5)   f  16  26  p compact
## 6         audi    a4   2.8 1999   6 manual(m5)   f  18  26  p compact
```

```
rio::import(file = "./data/mpg.xlsx", range = "A1:C10") %>% head() # specify range
```

```
##   manufacturer model displ
## 1         audi    a4   1.8
## 2         audi    a4   1.8
## 3         audi    a4   2.0
## 4         audi    a4   2.0
## 5         audi    a4   2.8
## 6         audi    a4   2.8
```

```
# Import large flat file with fread
df.f <- fread(file = "./data/mpg_maxi.csv", sep = ",")
head(df.f)
```

```
##    manufacturer      model displ year cyl    trans drv cty hwy fl   class
## 1:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
## 2:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
```

```
## 3:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
## 4:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
## 5:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
## 6:      pontiac grand prix   5.3 2008   8 auto(s4)   f  16  25  p midsize
```

```
# Read large .csv file
#  - compare execution times
#  - read.csv VS read_csv VS fread
print("Execution time read.csv():")
```

```
## [1] "Execution time read.csv():"
```

```
system.time(df1 <- read.csv(file = "./data/mpg_maxi.csv")) %>% head()
```

```
##  user.self   sys.self    elapsed user.child  sys.child
##       1.67       0.07       1.75         NA         NA
```

```
print("Execution time read_csv():")
```

```
## [1] "Execution time read_csv():"
```

```
system.time(df2 <- read_csv(file = "./data/mpg_maxi.csv")) %>% head()
```

```
##  user.self   sys.self    elapsed user.child  sys.child
##       0.93       0.10       0.48         NA         NA
```

```
print("Execution time fread():")
```

```
## [1] "Execution time fread():"
```

```
system.time(df3 <- fread(file = "./data/mpg_maxi.csv")) %>% head()
```

```
##  user.self   sys.self    elapsed user.child  sys.child
##       0.00       0.00       0.15         NA         NA
```

Write files

```
# Comma separated
write_csv(x = mpg, file = "./output/mpg_w.csv", col_names = T) %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# Semicolon separated
write_csv2(x = mpg, file = "./output/mpg_w2.csv", col_names = T) %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
```

```
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# write a xlsx file
rio::export(x = mpg, file = "./output/mpg.xlsx") %>% head()
```

```
## [1] "./output/mpg.xlsx"
```

```
# write/read to/from a .rds file
write_rds(x = mpg, file = "./output/mpg.rds") %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
read_rds(file = "./output/mpg.rds") %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
# feather
write_feather(x = mpg, path = "./output/mpg.feather") %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
read_feather(path = "./output/mpg.feather") %>% head()
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```