Pertemuan 6 - Tidyverse dan Importing Data

YOHANES FEBRYAN KANA NYOLA_123220198

2024-10-23

import library

```
library(dslabs)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

tibble

perbedaan pertama

```
# subsetting dataset murders kolom total
use_df = murders[,'total']
use_df
```

```
##  [1]  135   19  232   93 1257   65   97   38   99  669  376    7   12  364  142
## [16]   21   63  116  351   11  293  118  413   53  120  321   12   32   84    5
## [31]  246   67  517  286    4  310  111   36  457   16  207    8  219  805   22
## [46]    2  250   93   27   97    5
```

```
# dengan tibble
use_tbl = as_tibble(murders)[,'total']
use_tbl
```

total <dbl>
135
19

total <dbl>
232
93
1257
65
97
38
99
669

1-10 of 51 rows Previous **1** 2 3 4 5 6 Next

```
# fungsi class
class(use_df)
```

```
## [1] "numeric"
```

```
class(use_tbl)
```

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

perbedaan kedua

```
# memanggil kolom yang tidak tersedia
murders$Population
```

```
## NULL
```

```
as_tibble(murders)$Population
```

```
## Warning: Unknown or uninitialised column: `Population`.
```

```
## NULL
```

```
tibble(
  nama = c('tes', 'tes1', 'tes2'),
  fungsi = c(sum, mean, sd)
)
```

nama <chr>	fungsi <list>
tes	<builtin>
tes1	<fun>
tes2	<fun>

3 rows

```
grades <- data.frame(names = c("John", "Juan", "Jean", "Yao"),
exam_1 = c(95, 80, 90, 85),
exam_2 = c(90, 85, 85, 90))
class(grades$names)
```

```
## [1] "character"
```

dot operator

```
# tanpa pipeline
filtered_data = filter(murders,region == 'South')
mutated_data = mutate(filtered_data, rate = total/population * 10^6)
mutated_data$rate
```

```
##  [1]  28.24424  31.89390  42.31937 164.52753  33.98069  37.90323  26.73201
##  [8]  77.42581  50.74866  40.44085  29.99324  29.58934  44.75323  34.50936
## [15]  32.01360  31.24600  14.57101
```

```
# tanpa dot operator
filtered_data_pipeline = murders %>%
  filter(region == 'South') %>%
  mutate(rate = total/population * 10^6)
filtered_data_pipeline$rate
```

```
##  [1]  28.24424  31.89390  42.31937 164.52753  33.98069  37.90323  26.73201
##  [8]  77.42581  50.74866  40.44085  29.99324  29.58934  44.75323  34.50936
## [15]  32.01360  31.24600  14.57101
```

```
# dengan dot operator
murders %>%
  filter(region == 'South') %>%
  mutate(rate = total/population * 10^6) %>%
  .$rate
```

```
##  [1]  28.24424  31.89390  42.31937 164.52753  33.98069  37.90323  26.73201
##  [8]  77.42581  50.74866  40.44085  29.99324  29.58934  44.75323  34.50936
## [15]  32.01360  31.24600  14.57101
```

```
# fungsi pull
murders %>%
  filter(region == 'South') %>%
  mutate(rate = total/population * 10^6) %>%
  pull(rate)
```

```
##  [1]  28.24424  31.89390  42.31937 164.52753  33.98069  37.90323  26.73201
##  [8]  77.42581  50.74866  40.44085  29.99324  29.58934  44.75323  34.50936
## [15]  32.01360  31.24600  14.57101
```

purrr package (map)

```
compute_s_n = function(n){
  x = 1:n
  sum(x)
  # as_character(sum(x)) -> sebagai karakter
  # tibble(sum=sum(x)) -> sebagai dataframe
}
n = 1:25
s_n = sapply(n, compute_s_n)
class(s_n)
```

```
## [1] "integer"
```

```
s_n
```

```
##  [1]   1   3   6  10  15  21  28  36  45  55  66  78  91 105 120 136 153 171 190
## [20] 210 231 253 276 300 325
```

```
# library(purrr)
# s_n = map_int(n, compute_s_n) -> integer
# s_n = map_dbl(n, compute_s_n) -> double
# s_n = map_chr(n, compute_s_n) -> char
# s_n = map_lgl(n, compute_s_n) -> logika
# s_n = map_df(n, compute_s_n) -> dataframe
s_n = map(n, compute_s_n) # -> list
class(s_n)
```

```
## [1] "list"
```

```
s_n
```

```
## [[1]]
## [1] 1
##
## [[2]]
## [1] 3
##
## [[3]]
## [1] 6
##
## [[4]]
## [1] 10
##
## [[5]]
## [1] 15
##
## [[6]]
## [1] 21
##
## [[7]]
## [1] 28
##
## [[8]]
## [1] 36
##
## [[9]]
## [1] 45
##
## [[10]]
## [1] 55
##
## [[11]]
## [1] 66
##
## [[12]]
## [1] 78
##
## [[13]]
## [1] 91
##
## [[14]]
## [1] 105
##
## [[15]]
## [1] 120
##
## [[16]]
## [1] 136
##
## [[17]]
## [1] 153
##
## [[18]]
## [1] 171
##
## [[19]]
```

```
## [1] 190
##
## [[20]]
## [1] 210
##
## [[21]]
## [1] 231
##
## [[22]]
## [1] 253
##
## [[23]]
## [1] 276
##
## [[24]]
## [1] 300
##
## [[25]]
## [1] 325
```

tidyverse conditional

```
x = c(-2, -1, 0, 1, 2)
case_when(x < 0 ~ "Negatif Njir",
          x > 0 ~ "Positif Bang",
          TRUE ~ "Dasar Kosongg")
```

```
## [1] "Negatif Njir"  "Negatif Njir"  "Dasar Kosongg" "Positif Bang"
## [5] "Positif Bang"
```

```
# pada dataset murders
case_when(murders$population > 1e7 ~ "Banyak banget",
          between(murders$population, 1e6, 1e7) ~ "Banyak",
          TRUE ~ "Dikit")
```

```
##  [1] "Banyak"        "Dikit"         "Banyak"        "Banyak"
##  [5] "Banyak banget" "Banyak"        "Banyak"        "Dikit"
##  [9] "Dikit"         "Banyak banget" "Banyak"        "Banyak"
## [13] "Banyak"        "Banyak banget" "Banyak"        "Banyak"
## [17] "Banyak"        "Banyak"        "Banyak"        "Banyak"
## [21] "Banyak"        "Banyak"        "Banyak"        "Banyak"
## [25] "Banyak"        "Banyak"        "Dikit"         "Banyak"
## [29] "Banyak"        "Banyak"        "Banyak"        "Banyak"
## [33] "Banyak banget" "Banyak"        "Dikit"         "Banyak banget"
## [37] "Banyak"        "Banyak"        "Banyak banget" "Banyak"
## [41] "Banyak"        "Dikit"         "Banyak"        "Banyak banget"
## [45] "Banyak"        "Dikit"         "Banyak"        "Banyak"
## [49] "Banyak"        "Banyak"        "Dikit"
```

importing data

```
library(here)
```

```
## here() starts at D:/KULIAH IF/SEMESTER 5/PRAK DATA SCIENCE/Praktikum (Practice)/Pertemuan
6
```

```
here()
```

```
## [1] "D:/KULIAH IF/SEMESTER 5/PRAK DATA SCIENCE/Praktikum (Practice)/Pertemuan 6"
```

```
# dr_here()
here("data raw", "Iris1.csv")
```

```
## [1] "D:/KULIAH IF/SEMESTER 5/PRAK DATA SCIENCE/Praktikum (Practice)/Pertemuan 6/data raw/I
ris1.csv"
```

import dataset excel

```
library(readxl)
path = here("data raw", "JKT48_frame_2020.xlsx")
jkt48 = read_xlsx(path)
# View(jkt48)
```

import dataset delimited

```
library(readr)
path = here("data raw", "Iris1.csv")
iris1 = read_csv(path)
```

```
## Rows: 74 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Species
## dbl (5): Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm
##
## 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.
```

```
# View(iris1)

library(vroom)
```

```
##
## Attaching package: 'vroom'
##
## The following objects are masked from 'package:readr':
##
##     as.col_spec, col_character, col_date, col_datetime, col_double,
##     col_factor, col_guess, col_integer, col_logical, col_number,
##     col_skip, col_time, cols, cols_condense, cols_only, date_names,
##     date_names_lang, date_names_langs, default_locale, fwf_cols,
##     fwf_empty, fwf_positions, fwf_widths, locale, output_column,
##     problems, spec
```

```
iris1 = vroom(path)
```

```
## Rows: 74 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Species
## dbl (5): Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm
##
## 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.
```

import dataset delimited

```
list_iris <- list.files(path = here("data raw"), pattern =
"Iris", full.names = TRUE)
list_iris
```

```
## [1] "D:/KULIAH IF/SEMESTER 5/PRAK DATA SCIENCE/Praktikum (Practice)/Pertemuan 6/data raw/I
ris1.csv"
## [2] "D:/KULIAH IF/SEMESTER 5/PRAK DATA SCIENCE/Praktikum (Practice)/Pertemuan 6/data raw/I
ris2.csv"
```

functional iteration

```
iris_lapply = lapply(list_iris, read_csv)
```

```
## Rows: 74 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Species
## dbl (5): Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm
##
## 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: 76 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Species
## dbl (5): Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm
##
## 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.
```

```
iris_lapply = Reduce(rbind, iris_lapply)
# View(iris_lapply)
```

```
iris_vroom = vroom(list_iris)
```

```
## Rows: 150 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Species
## dbl (5): Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm
##
## 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.
```

```
identical(iris_lapply, iris_vroom)
```

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