123200028\_PraktikumDataScience\_Modul7

2022-11-01

1. Gunakan as\_tibble untuk mengkonversi tabel dataset “US murders” dalam bentuk tibble dan simpan dalam objek baru bernama ‘murders\_tibble’

library(dslabs)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

data(murders)  
as\_tibble(murders) %>% class()

## [1] "tbl\_df" "tbl" "data.frame"

murders\_tibble <- as\_tibble(murders) %>% class()

1. Gunakan fungsi group\_by untuk mengkonversi dataset “US murders” menjadi sebuah tibble yang dikelompokkan berdasarkan ‘region’

as\_tibble(murders) %>% group\_by(region)

## # A tibble: 51 × 5  
## # Groups: region [4]  
## state abb region population total  
## <chr> <chr> <fct> <dbl> <dbl>  
## 1 Alabama AL South 4779736 135  
## 2 Alaska AK West 710231 19  
## 3 Arizona AZ West 6392017 232  
## 4 Arkansas AR South 2915918 93  
## 5 California CA West 37253956 1257  
## 6 Colorado CO West 5029196 65  
## 7 Connecticut CT Northeast 3574097 97  
## 8 Delaware DE South 897934 38  
## 9 District of Columbia DC South 601723 99  
## 10 Florida FL South 19687653 669  
## # … with 41 more rows

1. Tulis script tidyverse yang menghasilkan output yang sama dengan perintah berikut: exp(mean(log(murders$population))) Gunakan operator pipe sehingga setiap fungsi dapat dipanggil tanpa menambahkan argumen. Gunakan dot operator untuk mengakses populasi.

library(dslabs)  
library(dplyr)  
data(murders)  
murders %>%   
 pull(population) %>%   
 log %>%  
 mean %>%   
 exp

## [1] 3675209

exp(mean(log(murders$population)))

## [1] 3675209

1. Gunakan map\_df untuk membuat data frame yang terdiri dari tiga kolom: ‘n’, ‘s\_n’, dan ‘s\_n\_2’. Kolom pertama harus berisi angka 1 hingga 100. Kolom kedua dan ketiga masingmasing harus berisi penjumlahan 1 hingga n, dimana n menyatakan jumlah baris.

library(purrr)  
compute\_s\_n <- function(n){   
 x <- 1:n   
 sum(x)   
}   
n <- 1:100   
s\_n <- sapply(n, compute\_s\_n)  
compute\_s\_n <- function(n){   
 x <- 1:n   
 tibble(sum = sum(x))   
}   
s\_n <- map\_df(n, compute\_s\_n)  
as\_tibble(s\_n)

## # A tibble: 100 × 1  
## sum  
## <int>  
## 1 1  
## 2 3  
## 3 6  
## 4 10  
## 5 15  
## 6 21  
## 7 28  
## 8 36  
## 9 45  
## 10 55  
## # … with 90 more rows