Modul 3

Ahmad Abdullah Azzam

9/28/2021

## D.Latihan

*Import* dataset *“murders”*. Menggunakan langkah berikut :

library(dslabs)  
data(murders)

1. Gunakan fungsi str untuk memeriksa struktur objek *“murders”*. Manakah dari pernyataan berikut ini yang paling menggambarkan karakter dari tiap variabel pada *data frame*?

str(murders)

## 'data.frame': 51 obs. of 5 variables:  
## $ state : chr "Alabama" "Alaska" "Arizona" "Arkansas" ...  
## $ abb : chr "AL" "AK" "AZ" "AR" ...  
## $ region : Factor w/ 4 levels "Northeast","South",..: 2 4 4 2 4 4 1 2 2 2 ...  
## $ population: num 4779736 710231 6392017 2915918 37253956 ...  
## $ total : num 135 19 232 93 1257 ...

a. Terdiri dari 51 Negara.  
b. Data berisi tingkat pembunuhan pada 50 negara bagian dan DC.  
c. Data berisi Nama negara bagian, singkatan dari nama negara bagian, wilayah negara bagian, dan populasi negara bagian serta jumlah total pembunuhan pada tahun 2010.  
d. `str` tidak menunjukkan informasi yang relevan.

Dari pernyataan diatas dapat disimpulkan bahwa ***Opsi C*** merupakan pernyataan yang paling menggambarkan karakter tiap variabel pada *data frame*.

1. Sebutkan apa saja nama kolom yang digunakan pada *data frame*. Terdapat 5 nama kolom yaitu :
   * state
   * abb
   * region
   * population
   * total
2. Gunakan operator aksesor ($) untuk mengekstrak informasi singkatan negara dan menyimpannya pada objek “*a*”. Sebutkan jenis class dari objek tersebut.

a<-murders$abb  
class(a)

## [1] "character"

1. Gunakan tanda kurung siku untuk mengekstrak singkatan negara dan menyimpannya pada objek “*b*”. Tentukan apakah variabel “*a*” dan “*b*” bernilai sama ?

b<-murders['abb']  
print(a)

## [1] "AL" "AK" "AZ" "AR" "CA" "CO" "CT" "DE" "DC" "FL" "GA" "HI" "ID" "IL" "IN"  
## [16] "IA" "KS" "KY" "LA" "ME" "MD" "MA" "MI" "MN" "MS" "MO" "MT" "NE" "NV" "NH"  
## [31] "NJ" "NM" "NY" "NC" "ND" "OH" "OK" "OR" "PA" "RI" "SC" "SD" "TN" "TX" "UT"  
## [46] "VT" "VA" "WA" "WV" "WI" "WY"

print(b)

## abb  
## 1 AL  
## 2 AK  
## 3 AZ  
## 4 AR  
## 5 CA  
## 6 CO  
## 7 CT  
## 8 DE  
## 9 DC  
## 10 FL  
## 11 GA  
## 12 HI  
## 13 ID  
## 14 IL  
## 15 IN  
## 16 IA  
## 17 KS  
## 18 KY  
## 19 LA  
## 20 ME  
## 21 MD  
## 22 MA  
## 23 MI  
## 24 MN  
## 25 MS  
## 26 MO  
## 27 MT  
## 28 NE  
## 29 NV  
## 30 NH  
## 31 NJ  
## 32 NM  
## 33 NY  
## 34 NC  
## 35 ND  
## 36 OH  
## 37 OK  
## 38 OR  
## 39 PA  
## 40 RI  
## 41 SC  
## 42 SD  
## 43 TN  
## 44 TX  
## 45 UT  
## 46 VT  
## 47 VA  
## 48 WA  
## 49 WV  
## 50 WI  
## 51 WY

1. Variabel *region* memiliki tipe data: *factor*. Dengan satu baris kode, gunakan fungsi level dan length untuk menentukan jumlah *region* yang dimiliki *dataset*.

region<-murders$region  
levels(region)

## [1] "Northeast" "South" "North Central" "West"

length(region)

## [1] 51

1. Fungsi table dapat digunakan untuk ekstraksi data pada tipe vektor dan menampilkan frekuensi dari setiap elemen. Dengan menerapkan fungsi tersebut, dapat diketahui jumlah *state* pada tiap *region*. Gunakan fungsi table dalam satu baris kode untuk menampilkan tabel baru yang berisi jumlah *state* pada tiap *region*.

table(region)

## region  
## Northeast South North Central West   
## 9 17 12 13