

Modul3

Iadhty

9/17/2021

```# Dasar Teori

Variasi tipe data pada R memfasilitasi keberagaman jenis variabel data. Sebagai contoh, terdapat data yang terdiri dari sekumpulan angka dan data lain yang berisi sekumpulan karakter. Pada contoh lain, ada pula data yang berbentuk tabel maupun kumpulan (list) angka sederhana.

Dengan bantuan fungsi class, kita akan mendapatkan kemudahan dalam mendefinisikan tipe data yang kita miliki:

```
a <- 2
class(a)
```

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

## Data Frames

Cara paling umum yang dapat digunakan untuk menyimpan dataset dalam R adalah dalam tipe data frame. Secara konseptual, kita dapat menganggap data frame sebagai tabel yang terdiri dari baris yang memiliki nilai pengamatan dan berbagai variabel yang didefinisikan dalam bentuk kolom. Tipe data ini sangat umum digunakan untuk dataset, karena data frame dapat menggabungkan berbagai jenis tipe data dalam satu objek.

Contoh dataset pada library(dslabs) dan pilih dataset "murders" menggunakan fungsi data:

```
library(dslabs)
data(murders)
```

Untuk memastikan bahwa dataset tersebut tipenya adalah data frame, dapat digunakan perintah berikut:

```
class(murders)
```

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

Untuk memeriksa lebih lanjut isi dataset, dapat pula digunakan fungsi str untuk mencari tahu lebih rinci mengenai struktur suatu objek:

```
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 ...
```

Dengan menggunakan fungsi `str`, dapat diketahui bahwa dataset “murders” terdiri dari 51 baris dan lima variabel: `state`, `abb`, `region`, `population`, dan `total`. Selanjutnya, untuk melihat contoh enam baris pertama pada dataset, dapat digunakan fungsi `head`:

```
head(murders)
```

```
state abb region population total
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
```

Untuk analisis awal tiap variabel yang diwakili dalam bentuk kolom pada tipe data frame, dapat digunakan operator aksesori (`$`) dengan cara berikut:

```
murders$population
```

```
[1] 4779736 710231 6392017 2915918 37253956 5029196 3574097 897934
[9] 601723 19687653 9920000 1360301 1567582 12830632 6483802 3046355
[17] 2853118 4339367 4533372 1328361 5773552 6547629 9883640 5303925
[25] 2967297 5988927 989415 1826341 2700551 1316470 8791894 2059179
[33] 19378102 9535483 672591 11536504 3751351 3831074 12702379 1052567
[41] 4625364 814180 6346105 25145561 2763885 625741 8001024 6724540
[49] 1852994 5686986 563626
```

Untuk mengetahui nama-nama dari lima variabel yang dapat dievaluasi menggunakan operator aksesori, sebelumnya, melalui fungsi `str`, telah kita ketahui bahwa variabel yang dimiliki dataset adalah: `state`, `abb`, `region`, `population`, dan `total`. Sebagai alternatif, terdapat pula fungsi `names`, yang dapat digunakan seperti contoh dibawah ini:

```
names(murders)
```

```
[1] "state" "abb" "region" "population" "total"
```

## Vector : Numeric, character, dan logical

Untuk mengidentifikasi banyaknya entri dalam suatu vector dapat digunakan fungsi `length` seperti contoh berikut:

```
length(murders$population)
```

```
[1] 51
```

Vector khusus ini bertipe numeric karena populasi terdiri dari data-data angka:

```
class(murders$population)
```

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

Vector juga dapat digunakan untuk menyimpan string dengan tipe character, Sebagai contoh nama negara pada dataset “murders”:

```
class(murders$state)
```

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

Jenis vector penting lainnya adalah logical yang nilainya berupa TRUE atau FALSE

```
z <- 3 == 2
z
```

```
[1] FALSE
```

```
class(z)
```

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

## Factors

Dalam dataset “murders”, variabel state yang berisi data karakter bukan bertipe vector character, namun, tipe datanya adalah factor:

```
class(murders$region)
```

```
[1] "factor"
```

Faktor berguna untuk menyimpan data kategorikal. Dapat dilihat, bahwa hanya terdapat 4 wilayah pada variabel state. Untuk melihat jumlah kategori yang dimiliki oleh variabel dengan tipe data factor dapat digunakan fungsi level:

```
levels(murders$region)
```

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

## List

Data frame merupakan sekumpulan list yang memiliki kelas yang berbeda-beda. Sama halnya dengan data frame, analisis list dapat dilakukan dengan menggunakan operator aksesor (\$) dan dua kurung siku ([]).

## Matriks

Matriks merupakan tipe data yang mirip dengan data frame karena keduanya memiliki dua dimensi, yaitu: baris dan kolom. Namun, sama halnya dengan tipe data vector numerik, karakter dan logis, entri dalam matriks harus terdiri dari jenis vector yang sama. Dalam hal ini, data frame dapat dikatakan sebagai tipe data

yang paling cocok untuk menyimpan data, karena kita dapat memiliki karakter, faktor, dan angka sekaligus dalam satu data frame. Namun matriks memiliki satu keunggulan yang tidak dimiliki oleh tipe data frame: pada matriks dapat dilakukan operasi aljabar.

Untuk mendefinisikan matriks, dapat digunakan fungsi `matrix` dengan mendefinisikan pula argumen berupa jumlah baris dan kolom yang diinginkan.

```
mat <- matrix(1:12, 4, 3)
mat
```

```
[,1] [,2] [,3]
[1,] 1 5 9
[2,] 2 6 10
[3,] 3 7 11
[4,] 4 8 12
```

Untuk mengakses entri tertentu dalam matriks, dapat digunakan tanda kurung siku (`[]`). Sebagai contoh, kita akan menampilkan data pada baris kedua, kolom ketiga, menggunakan:

```
mat[2,3]
```

```
[1] 10
```