

Tugas 4

Sayyid Quthub

9/30/2021

Latihan

Pada latihan ini seluruh soal akan menggunakan dataset: AS murders

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

Nomor 1

Gunakan operator aksesori (\$) untuk mengakses variabel populasi dan menyimpannya pada objek baru "pop". Kemudian gunakan fungsi sort untuk mengurutkan variabel "pop". Pada langkah terakhir, gunakan operator [] untuk menampilkan nilai populasi terkecil.

```
library(dslabs)
data("murders")
pop <- murders$population
pop <- sort(pop)
pop[1]

## [1] 563626
```

Nomor 2

Tampilkan indeks dari data yang memiliki nilai populasi terkecil.

```
order(murders$population)[1]

## [1] 51
```

Nomor 3

Dengan fungsi which.min, Tulis satu baris kode yang dapat menampilkan hasil yang sama dengan langkah diatas

```
which.min(murders$population)

## [1] 51
```

Nomor 4

Tampilkan nama negara yang memiliki populasi terkecil.

```
i_min <- which.min(murders$population)
murders$state[i_min]
```

```
## [1] "Wyoming"
```

Nomor 5

Untuk membuat data frame baru, contoh script yang dapat digunakan adalah sebagai berikut:

```
temp <- c(35, 88, 42, 84, 81, 30)
city <- c("Beijing", "Lagos", "Paris", "Rio de Janeiro",
"San Juan", "Toronto")
city_temps <- data.frame(name = city, temperature = temp)
```

Gunakan fungsi rank untuk menentukan peringkat populasi dari tiap negara bagian, dimulai dari nilai terkecil hingga terbesar. Simpan hasil pemeringkatan di objek baru "ranks", lalu buat data frame baru yang berisi nama negara bagian dan peringkatnya dengan nama "my_df".

```
ranks <- (murders$population)
my_df <- data.frame(state = murders$state, ranks = ranks)
my_df
```

##	state	ranks
## 1	Alabama	4779736
## 2	Alaska	710231
## 3	Arizona	6392017
## 4	Arkansas	2915918
## 5	California	37253956
## 6	Colorado	5029196
## 7	Connecticut	3574097
## 8	Delaware	897934
## 9	District of Columbia	601723
## 10	Florida	19687653
## 11	Georgia	9920000
## 12	Hawaii	1360301
## 13	Idaho	1567582
## 14	Illinois	12830632
## 15	Indiana	6483802
## 16	Iowa	3046355
## 17	Kansas	2853118
## 18	Kentucky	4339367
## 19	Louisiana	4533372
## 20	Maine	1328361
## 21	Maryland	5773552
## 22	Massachusetts	6547629
## 23	Michigan	9883640
## 24	Minnesota	5303925
## 25	Mississippi	2967297
## 26	Missouri	5988927
## 27	Montana	989415
## 28	Nebraska	1826341
## 29	Nevada	2700551

```
## 30      New Hampshire 1316470
## 31      New Jersey   8791894
## 32      New Mexico   2059179
## 33      New York    19378102
## 34      North Carolina 9535483
## 35      North Dakota  672591
## 36      Ohio        11536504
## 37      Oklahoma     3751351
## 38      Oregon       3831074
## 39      Pennsylvania 12702379
## 40      Rhode Island 1052567
## 41      South Carolina 4625364
## 42      South Dakota  814180
## 43      Tennessee    6346105
## 44      Texas        25145561
## 45      Utah         2763885
## 46      Vermont      625741
## 47      Virginia     8001024
## 48      Washington   6724540
## 49      West Virginia 1852994
## 50      Wisconsin    5686986
## 51      Wyoming      563626
```

Nomor 6

Ulangi langkah sebelumnya, namun kali ini urutkan `my_df` dengan fungsi `order` agar data yang ditampilkan merupakan data yang telah diurutkan dari populasi yang paling tidak padat hingga ke yang terpadat.

```
ranks <- (murders$population)
my_df <- data.frame(state = murders$state, ranks = ranks)
ind <- order(my_df$ranks)
my_df <- data.frame(state = murders$state[ind], ranks = 1:nrow(my_df))
my_df
```

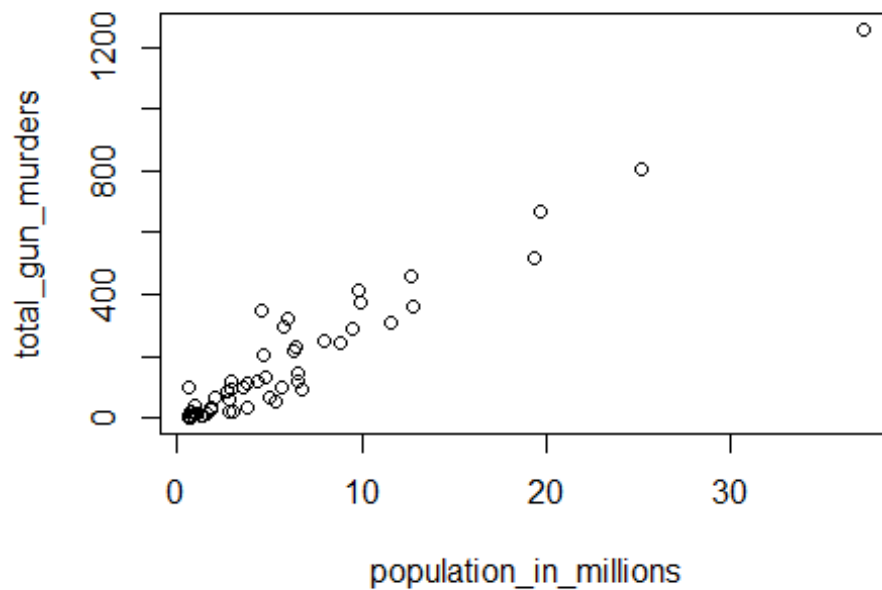
	state	ranks
## 1	Wyoming	1
## 2	District of Columbia	2
## 3	Vermont	3
## 4	North Dakota	4
## 5	Alaska	5
## 6	South Dakota	6
## 7	Delaware	7
## 8	Montana	8
## 9	Rhode Island	9
## 10	New Hampshire	10
## 11	Maine	11
## 12	Hawaii	12
## 13	Idaho	13
## 14	Nebraska	14

## 15	West Virginia	15
## 16	New Mexico	16
## 17	Nevada	17
## 18	Utah	18
## 19	Kansas	19
## 20	Arkansas	20
## 21	Mississippi	21
## 22	Iowa	22
## 23	Connecticut	23
## 24	Oklahoma	24
## 25	Oregon	25
## 26	Kentucky	26
## 27	Louisiana	27
## 28	South Carolina	28
## 29	Alabama	29
## 30	Colorado	30
## 31	Minnesota	31
## 32	Wisconsin	32
## 33	Maryland	33
## 34	Missouri	34
## 35	Tennessee	35
## 36	Arizona	36
## 37	Indiana	37
## 38	Massachusetts	38
## 39	Washington	39
## 40	Virginia	40
## 41	New Jersey	41
## 42	North Carolina	42
## 43	Michigan	43
## 44	Georgia	44
## 45	Ohio	45
## 46	Pennsylvania	46
## 47	Illinois	47
## 48	New York	48
## 49	Florida	49
## 50	Texas	50
## 51	California	51

Nomor 7

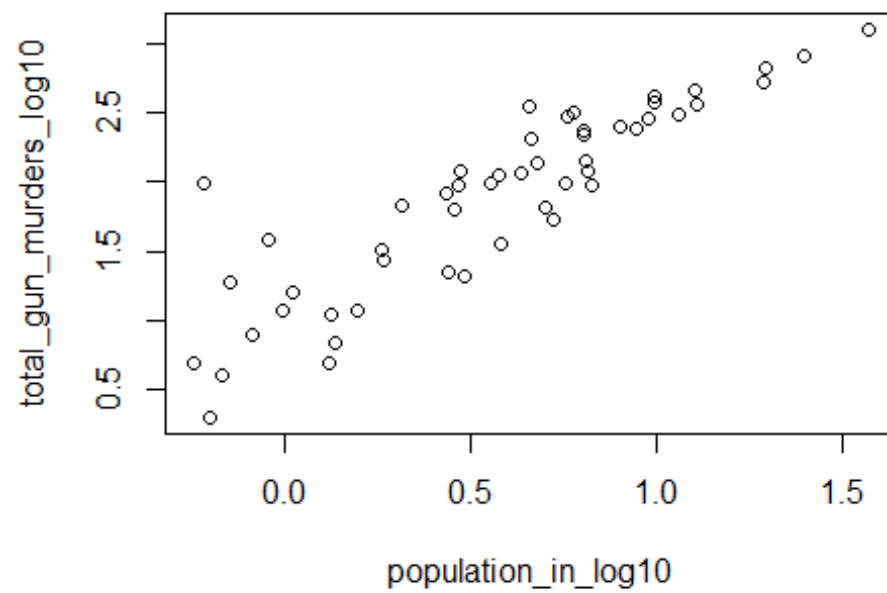
Untuk keperluan analisis data, akan dibuat plot yang memvisualisasikan total pembunuhan terhadap populasi dan mengidentifikasi hubungan antara keduanya.

```
population_in_millions <- murders$population/10^6
total_gun_murders <- murders$total
plot(population_in_millions, total_gun_murders)
```



Perlu diingat bahwa beberapa negara bagian memiliki populasi di bawah 5 juta, sehingga untuk mempermudah analisis, buat plot dalam skala log. Transformasi nilai variabel menggunakan transformasi log10, kemudian tampilkan plot-nya.

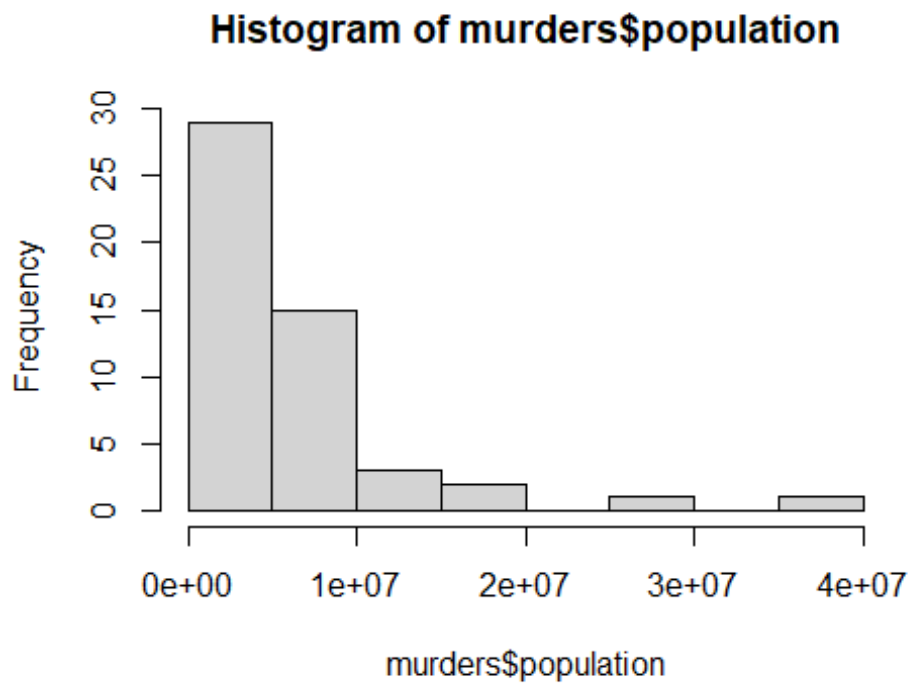
```
population_in_log10 <- log10(population_in_millions)
total_gun_murders_log10 <- log10(total_gun_murders)
plot(population_in_log10, total_gun_murders_log10)
```



Nomor 8

Buat histogram dari populasi negara bagian

```
hist(murders$population)
```



Nomor 9

Hasilkan boxplot dari populasi negara bagian berdasarkan wilayahnya.

```
boxplot(population~region, data = murders)
```

