

Tugas Modul 3 dan 4

Hazlan Muhammad Qodri

2022-09-14

R Markdown

Nama : Hazlan Muhammad Qodri
NIM : 123190080
Kelas : Praktikum Data Science - C

Tugas Modul 3

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 ...
```

- Terdiri dari 51 negara.

```
murders$state
```

## [1] "Alabama"	"Alaska"	"Arizona"
## [4] "Arkansas"	"California"	"Colorado"
## [7] "Connecticut"	"Delaware"	"District of Columbia"
## [10] "Florida"	"Georgia"	"Hawaii"
## [13] "Idaho"	"Illinois"	"Indiana"
## [16] "Iowa"	"Kansas"	"Kentucky"
## [19] "Louisiana"	"Maine"	"Maryland"
## [22] "Massachusetts"	"Michigan"	"Minnesota"
## [25] "Mississippi"	"Missouri"	"Montana"
## [28] "Nebraska"	"Nevada"	"New Hampshire"
## [31] "New Jersey"	"New Mexico"	"New York"

```
## [34] "North Carolina"      "North Dakota"      "Ohio"
## [37] "Oklahoma"            "Oregon"             "Pennsylvania"
## [40] "Rhode Island"        "South Carolina"     "South Dakota"
## [43] "Tennessee"          "Texas"              "Utah"
## [46] "Vermont"             "Virginia"           "Washington"
## [49] "West Virginia"       "Wisconsin"          "Wyoming"
```

- Data berisi tingkat pembunuhan pada 50 negara bagian dan DC.

```
murders[1:50,]
```

```
##           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
## 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
## 11     Georgia  GA      South   9920000    376
## 12     Hawaii  HI       West    1360301      7
## 13     Idaho   ID       West    1567582     12
## 14     Illinois IL North Central 12830632   364
## 15     Indiana IN North Central 6483802    142
## 16     Iowa    IA North Central 3046355     21
## 17     Kansas  KS North Central 2853118     63
## 18     Kentucky KY      South    4339367    116
## 19     Louisiana LA     South    4533372    351
## 20     Maine   ME      Northeast 1328361     11
## 21     Maryland MD     South    5773552    293
## 22     Massachusetts MA    Northeast 6547629    118
## 23     Michigan MI North Central 9883640    413
## 24     Minnesota MN North Central 5303925     53
## 25     Mississippi MS     South    2967297    120
## 26     Missouri MO North Central 5988927    321
## 27     Montana MT       West     989415     12
## 28     Nebraska NE North Central 1826341     32
## 29     Nevada  NV       West    2700551     84
## 30     New Hampshire NH    Northeast 1316470      5
## 31     New Jersey NJ     Northeast 8791894    246
## 32     New Mexico NM       West    2059179     67
## 33     New York NY      Northeast 19378102   517
## 34     North Carolina NC     South    9535483   286
## 35     North Dakota ND North Central 672591      4
## 36     Ohio    OH North Central 11536504   310
## 37     Oklahoma OK      South    3751351    111
## 38     Oregon  OR       West    3831074     36
## 39     Pennsylvania PA    Northeast 12702379   457
## 40     Rhode Island RI    Northeast 1052567     16
## 41     South Carolina SC     South    4625364    207
## 42     South Dakota SD North Central 814180      8
```

```
## 43      Tennessee TN      South  6346105  219
## 44      Texas TX      South  25145561  805
## 45      Utah UT      West  2763885  22
## 46      Vermont VT      Northeast  625741  2
## 47      Virginia VA      South  8001024  250
## 48      Washington WA      West  6724540  93
## 49      West Virginia WV      South  1852994  27
## 50      Wisconsin WI North Central  5686986  97
```

```
murders[murders$abb == 'DC',]
```

```
##      state abb region population total
## 9 District of Columbia DC South  601723  99
```

Referensi: <https://sparkbyexamples.com/r-programming/select-rows-in-r/>

- Data berisi Nama negara bagian, singkatan dari nama negara bagian, wilayah negara bagian, dan populasi negara bagian serta jumlah total pembunuhan pada tahun 2010. Kalau Menampilkan Data Murders

```
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
## 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
## 11     Georgia GA      South  9920000  376
## 12     Hawaii HI      West   1360301    7
## 13     Idaho ID      West   1567582   12
## 14     Illinois IL North Central 12830632  364
## 15     Indiana IN North Central  6483802  142
## 16     Iowa IA North Central  3046355   21
## 17     Kansas KS North Central  2853118   63
## 18     Kentucky KY      South  4339367  116
## 19     Louisiana LA      South  4533372  351
## 20     Maine ME      Northeast 1328361   11
## 21     Maryland MD      South  5773552  293
## 22     Massachusetts MA      Northeast 6547629  118
## 23     Michigan MI North Central 9883640  413
## 24     Minnesota MN North Central 5303925   53
## 25     Mississippi MS      South  2967297  120
## 26     Missouri MO North Central 5988927  321
## 27     Montana MT      West   989415   12
## 28     Nebraska NE North Central 1826341   32
## 29     Nevada NV      West  2700551   84
## 30     New Hampshire NH      Northeast 1316470    5
```

## 31	New Jersey	NJ	Northeast	8791894	246
## 32	New Mexico	NM	West	2059179	67
## 33	New York	NY	Northeast	19378102	517
## 34	North Carolina	NC	South	9535483	286
## 35	North Dakota	ND	North Central	672591	4
## 36	Ohio	OH	North Central	11536504	310
## 37	Oklahoma	OK	South	3751351	111
## 38	Oregon	OR	West	3831074	36
## 39	Pennsylvania	PA	Northeast	12702379	457
## 40	Rhode Island	RI	Northeast	1052567	16
## 41	South Carolina	SC	South	4625364	207
## 42	South Dakota	SD	North Central	814180	8
## 43	Tennessee	TN	South	6346105	219
## 44	Texas	TX	South	25145561	805
## 45	Utah	UT	West	2763885	22
## 46	Vermont	VT	Northeast	625741	2
## 47	Virginia	VA	South	8001024	250
## 48	Washington	WA	West	6724540	93
## 49	West Virginia	WV	South	1852994	27
## 50	Wisconsin	WI	North Central	5686986	97
## 51	Wyoming	WY	West	563626	5

Kalau Menampilkan Jumlah Total Pembunuhan pada Tahun 2010

```
sum(murders$total)
```

```
## [1] 9403
```

- str tidak menunjukkan informasi yang relevan.

```
str(murders) # gak ngerti maksud soalnya apa :(
```

```
## '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 ...
```

2. Sebutkan apa saja nama kolom yang digunakan pada data frame

```
names(murders)
```

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

3. 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"
```

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

```
b = murders$abb[]  
class(b)
```

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

```
print(paste("Apakah variabel 'a' dan 'b' bernilai sama? ", a == b))
```

```
## [1] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [2] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [3] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [4] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [5] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [6] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [7] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [8] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [9] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [10] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [11] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [12] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [13] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [14] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [15] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [16] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [17] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [18] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [19] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [20] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [21] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [22] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [23] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [24] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [25] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [26] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [27] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [28] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [29] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [30] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [31] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [32] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [33] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [34] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [35] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [36] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [37] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [38] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [39] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [40] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [41] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [42] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"  
## [43] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
```

```
## [44] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [45] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [46] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [47] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [48] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [49] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [50] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
## [51] "Apakah variabel 'a' dan 'b' bernilai sama? TRUE"
```

Jawaban: Ya, bernilai sama

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

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

```
## [1] 4
```

6. 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(murders$region)
```

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

Tugas Modul 4

Pada latihan ini seluruh soal akan menggunakan dataset: AS murders

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

1. Gunakan operator aksesor (\$) 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.

```
pop <- murders$population
sorted_pop <- sort(pop)
sorted_pop[1]
```

```
## [1] 563626
```

2. Tampilkan indeks dari data yang memiliki nilai populasi terkecil. Petunjuk: gunakan fungsi order.

```
index <- order(murders$population)
index[1]
```

```
## [1] 51
```

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

```
which.min(murders$population)
```

```
## [1] 51
```

4. Tampilkan nama negara yang memiliki populasi terkecil.

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

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

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

```
##           state rank
## 1      Alabama   29
## 2       Alaska    5
## 3      Arizona   36
## 4    Arkansas   20
## 5    California  51
## 6     Colorado   30
## 7   Connecticut  23
## 8     Delaware    7
## 9 District of Columbia  2
## 10     Florida   49
## 11     Georgia   44
## 12      Hawaii   12
## 13      Idaho    13
## 14    Illinois   47
## 15     Indiana   37
```

```
## 16          Iowa 22
## 17          Kansas 19
## 18          Kentucky 26
## 19          Louisiana 27
## 20          Maine 11
## 21          Maryland 33
## 22          Massachusetts 38
## 23          Michigan 43
## 24          Minnesota 31
## 25          Mississippi 21
## 26          Missouri 34
## 27          Montana 8
## 28          Nebraska 14
## 29          Nevada 17
## 30          New Hampshire 10
## 31          New Jersey 41
## 32          New Mexico 16
## 33          New York 48
## 34          North Carolina 42
## 35          North Dakota 4
## 36          Ohio 45
## 37          Oklahoma 24
## 38          Oregon 25
## 39          Pennsylvania 46
## 40          Rhode Island 9
## 41          South Carolina 28
## 42          South Dakota 6
## 43          Tennessee 35
## 44          Texas 50
## 45          Utah 18
## 46          Vermont 3
## 47          Virginia 40
## 48          Washington 39
## 49          West Virginia 15
## 50          Wisconsin 32
## 51          Wyoming 1
```

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. Petunjuk: buat objek “ind” yang akan menyimpan indeks yang diperlukan dalam mengurutkan data populasi

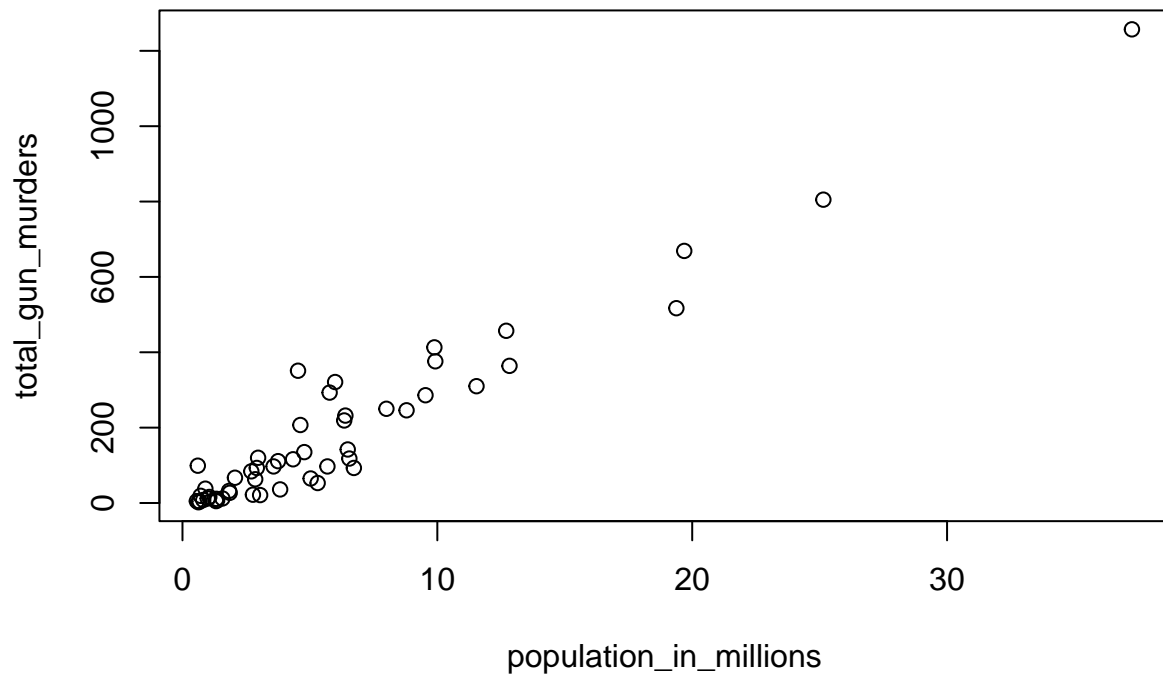
```
ranks <- rank(murders$population)
ind <- order(ranks)
my_df <- data.frame(state = murders$state[ind], rank = ranks[ind])
my_df
```

```
##          state rank
## 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

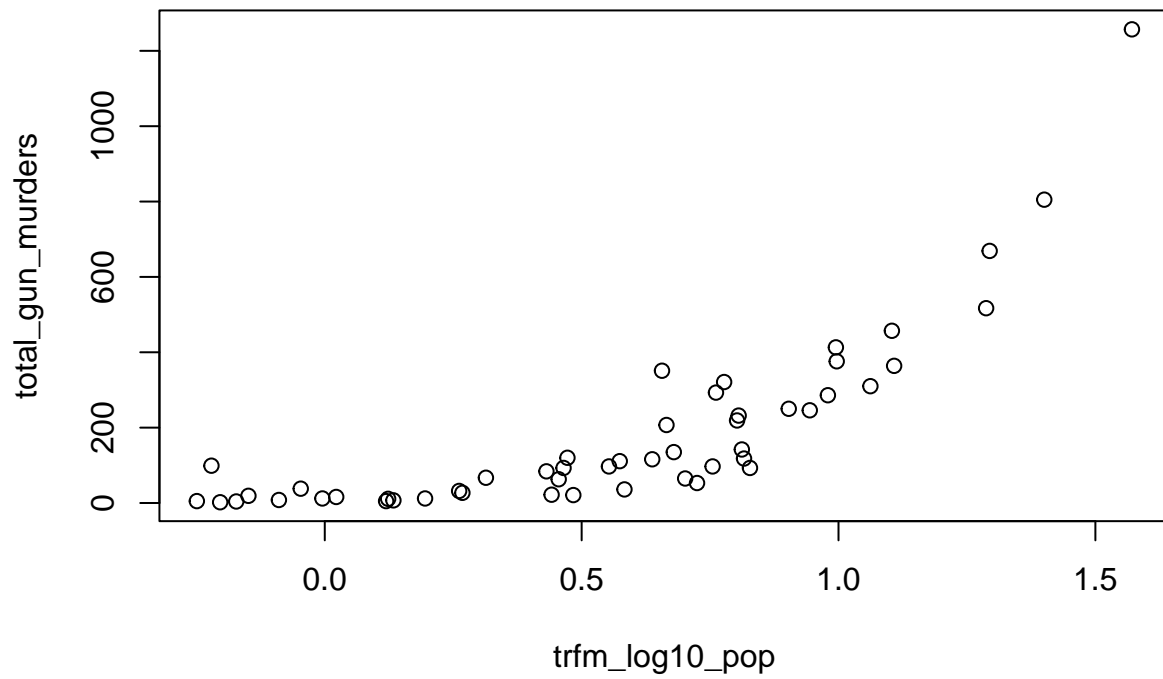
7. Untuk keperluan analisis data, akan dibuat plot yang memvisualisasikan total pembunuhan terhadap populasi dan mengidentifikasi hubungan antara keduanya. Script yang digunakan:

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



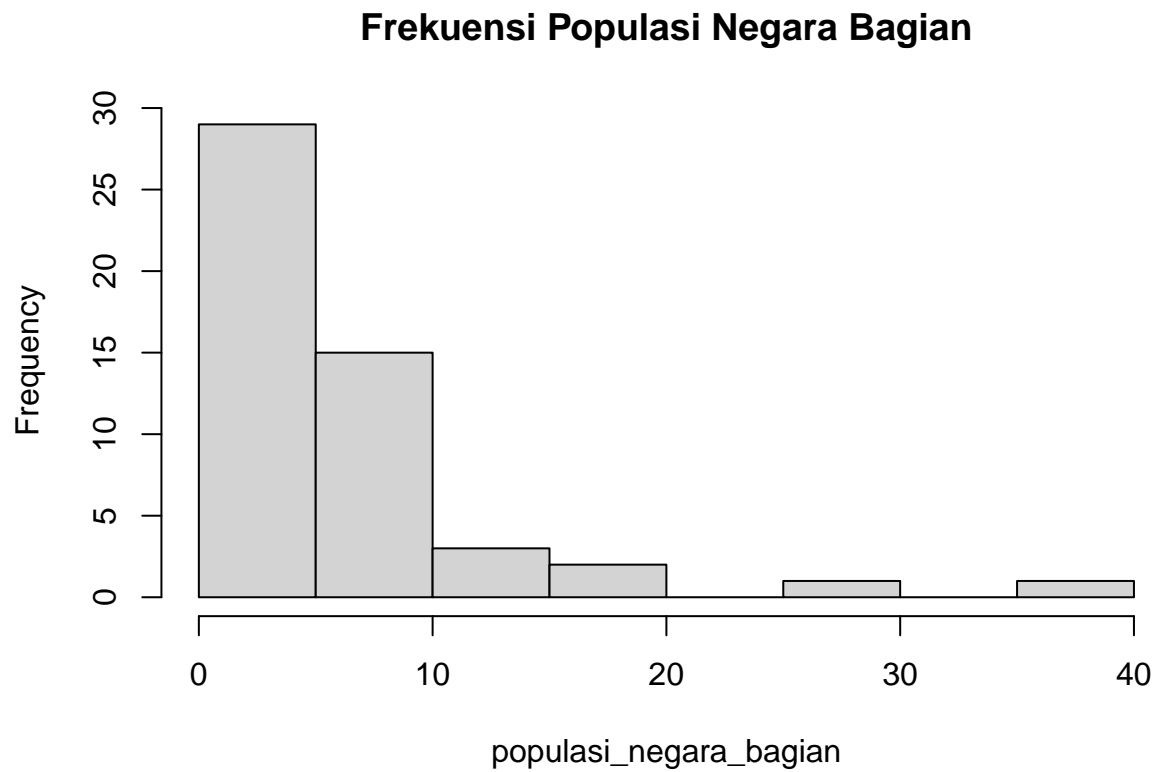
Perlu diingat bahwa beberapa negara bagian memiliki populasi di bawah 5 juta, sehingga untuk mempermudah analisis, buat plot dalam skala log. Transformasi nilai variabel menggunakan transformasi log10, kemudian tampilkan plot-nya.

```
population_in_millions <- murders$population/10^6  
trfm_log10_pop <- log10(population_in_millions)  
total_gun_murders <- murders$total  
plot(trfm_log10_pop, total_gun_murders)
```



8. Buat histogram dari populasi negara bagian.

```
populasi_negara_bagian <- with(murders, population/10^6)
hist(populasi_negara_bagian, main = "Frekuensi Populasi Negara Bagian")
```



9. Hasilkan boxplot dari populasi negara bagian berdasarkan wilayahnya.

```
populasi_negara_bagian <- with(murders, population/106)  
boxplot(populasi_negara_bagian~region, data = murders)
```

