

Tugas3-123170036

Muhammad Rizal

2/20/2020

Praktikum Data Science

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

No 5. Membuat Data Frame baru kemudian mengurutkan menggunakan rank

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

```
##           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
```

Penjelasan No 5

Menampilkan data populasi serta ranking urutan dalam populasi, menggunakan perintah **rank** kemudian setelah itu disimpan kedalam object/variabel **ranks** kemudian buat object data.frame baru yang disimpan di variabel **my_df**

No 6. Membuat data.frame yang terurut

```
sort_rank <- order(my_df$rank)
sort_my_df <- data.frame(state = my_df$state[sort_rank], rank = my_df$rank[sort_rank])
sort_my_df # order by population
```

```
##      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

Penjelasan No 6

Mirip dengan data sebelumnya tetapi disini telah diorder berdasarkan rank populasi terendah ke tertinggi, menggunakan indeks dari data.frame tersebut. Mengurutkan menggunakan perintah **order**

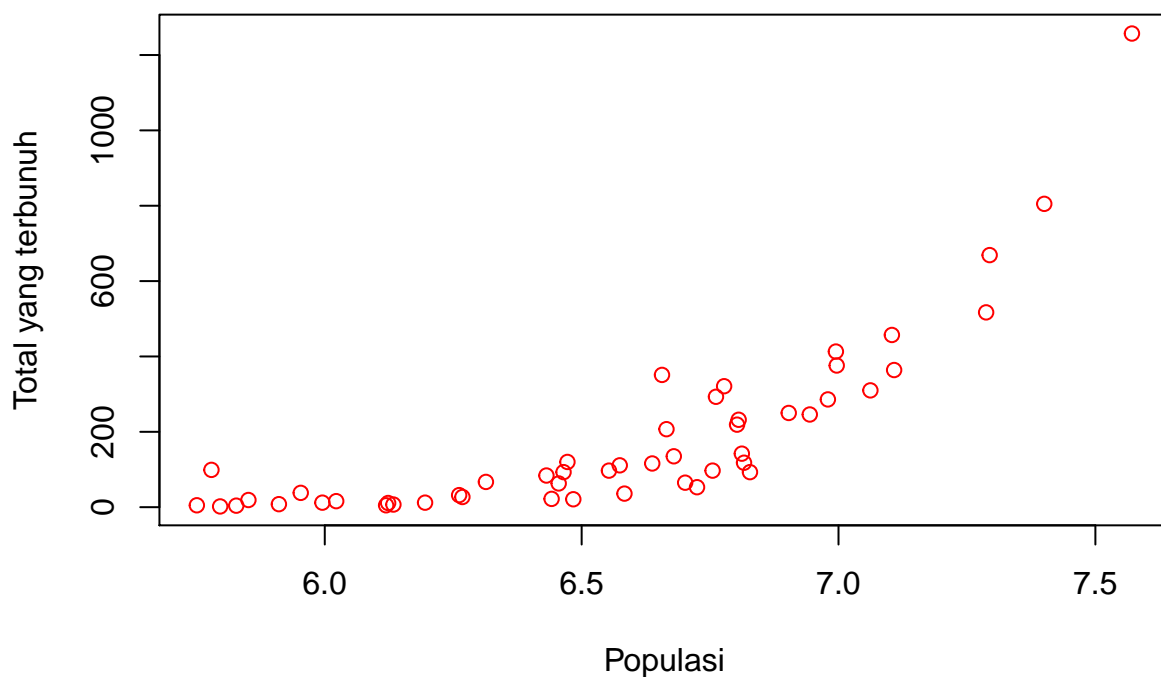
No 7. membuat plot dengan basis log10

```

population_in_millions <- log10(murders$population)
total_gun_murders <- murders$total
plot(population_in_millions,
      total_gun_murders,
      main = "Plot Data Tingkat pembunuhan berdasarkan populasi",
      ylab = "Total yang terbunuh",
      xlab = "Populasi",
      col = "red")

```

Plot Data Tingkat pembunuhan berdasarkan populasi



Penjelasan No 7 Menampilkan plot populasi dengan total orang yang terbunuh dalam setiap populasi tersebut, karena datanya ada yang paling banyak kurang dari 10 juta maka digunakan perintah **log10** untuk menyederhanakan statistik plot tersebut. Untuk menampilkan plot menggunakan perintah **plot** dengan parameter pertama dan kedua data yg ingin diplot, data yang digunakan disini adalah data integer. parameter **main** untuk judul dari plot tersebut, **ylab** untuk keterangan dari data plot vertical, sedangkan **xlab** untuk keterangan data horizontal, memberi warna dengan **col**.

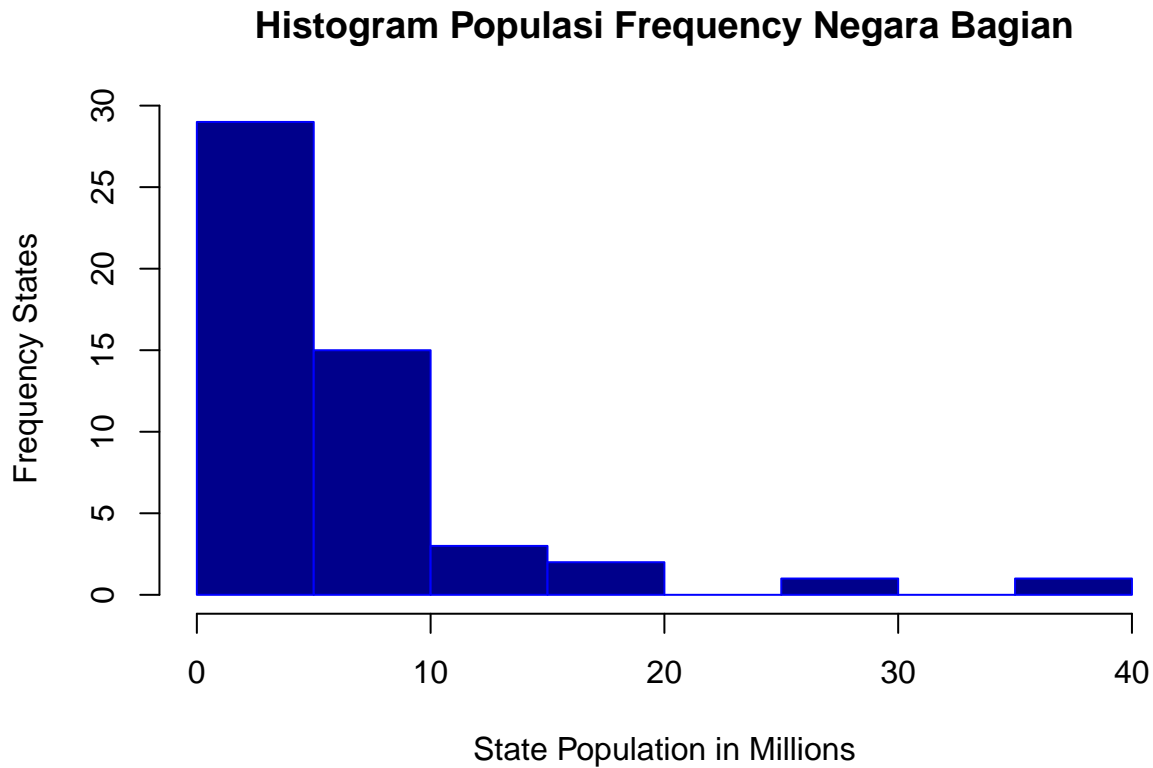
No 8. Histogram negara bagian

```

state_population <- with(murders, population/10^6)
hist(state_population,
      main = "Histogram Populasi Frequency Negara Bagian",
      breaks = 10,

```

```
xlab = "State Population in Millions",
xlim = c(0, 40),
ylab = "Frequency States",
border="blue",
col="darkblue")
```



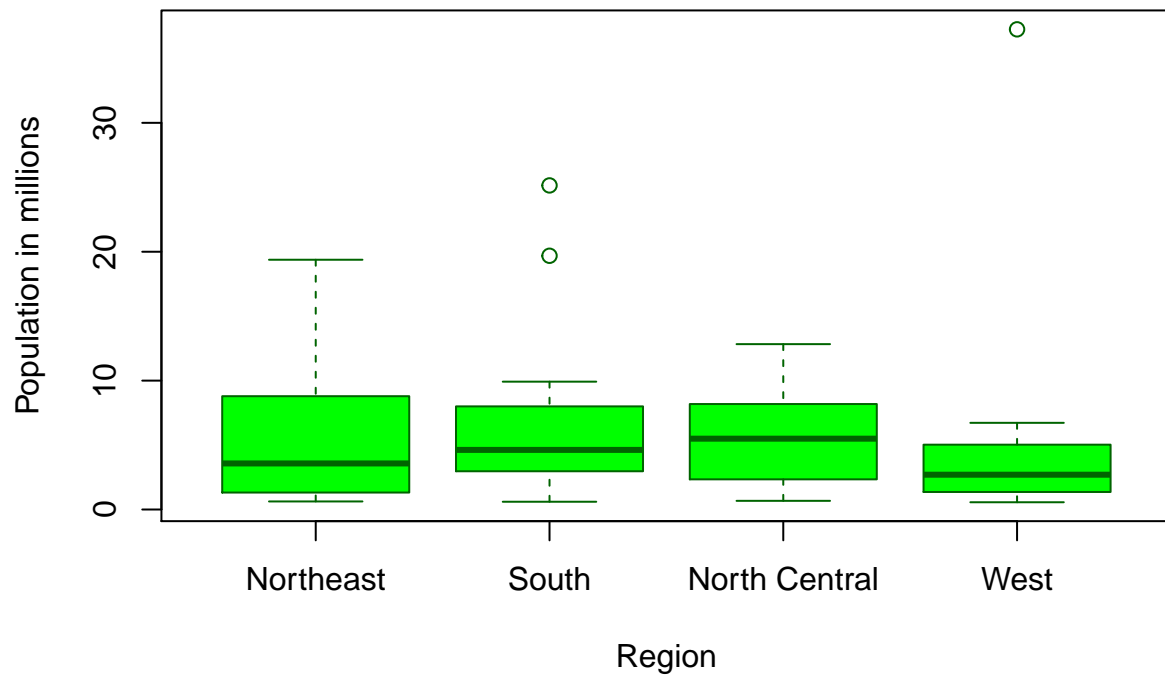
Penjelasan No 8

Menampilkan histogram dengan perintah **hist** data yang digunakan adalah integer, penjelasan dari parameternya mirip dengan yang sebelumnya, sedangkan disini **xlim** untuk range dari absis histogram tersebut.

No 9. Menampilkan Boxplot dari populasi bagian.

```
region_population <- with(murders, population/10^6)
boxplot(region_population~region,
        main = "Boxplot populasi berdasarkan wilayah",
        data = murders,
        xlab = "Region",
        ylab = "Population in millions",
        col = "green",
        border = "darkgreen")
```

Boxplot populasi berdasarkan wilayah



Penjelasan No 9

Menampilkan data populasi berdasarkan region dengan tipe boxplot menggunakan **boxplot** data populasinya disini dinormalisasikan terlebih dahulu dengan membaginya dengan 10^6 , penjelasannya mirip dengan yang sebelumnya, parameter pertama, menggunakan `~` untuk pemisah antara data yang dikelompokkan dengan data yang ditampilkan.