

123190075_Tugas1

Yohanes Ageng Hendrawan Zhamudra Putra

9/19/2022

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

##MODUL 3

```
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?
 - 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.

```
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
## $ population: num  4779736 710231 6392017 2915918 37253956 ...
## $ total     : num  135 19 232 93 1257 ...
```

Pilihan C merupakan pernyataan paling menggambarkan tiap variabel

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']  
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
```

Sama, variabel a dan b mempunyai nilai yang sama

5. 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(levels(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(region)

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

##MODUL 4

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.

```

pop = murders$population
asc = sort(pop)
asc [1]

## [1] 563626

min(pop)

## [1] 563626

```

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

```

x = murders$population
index = order(x)
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.

```

q = murders$population
p = order(q)
m = p[1]
s = murders$state
s[m]

## [1] "Wyoming"

```

5. 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".

```

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)
rank(city_temps$temperature)

## [1] 2 6 3 5 4 1

my_df = city_temps[order(rank(city_temps$temperature)),]
my_df

##           name temperature
## 6      Toronto          30
## 1      Beijing          35

```

```
## 3      Paris      42
## 5      San Juan   81
## 4 Rio de Janeiro 84
## 2      Lagos     88
```

- 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

```
ind = rank(murders$populations)
my_Df = murders[order(rank(murders$population)),]
my_Df
```

##	state	abb	region	population	total
## 51	Wyoming	WY	West	563626	5
## 9	District of Columbia	DC	South	601723	99
## 46	Vermont	VT	Northeast	625741	2
## 35	North Dakota	ND	North Central	672591	4
## 2	Alaska	AK	West	710231	19
## 42	South Dakota	SD	North Central	814180	8
## 8	Delaware	DE	South	897934	38
## 27	Montana	MT	West	989415	12
## 40	Rhode Island	RI	Northeast	1052567	16
## 30	New Hampshire	NH	Northeast	1316470	5
## 20	Maine	ME	Northeast	1328361	11
## 12	Hawaii	HI	West	1360301	7
## 13	Idaho	ID	West	1567582	12
## 28	Nebraska	NE	North Central	1826341	32
## 49	West Virginia	WV	South	1852994	27
## 32	New Mexico	NM	West	2059179	67
## 29	Nevada	NV	West	2700551	84
## 45	Utah	UT	West	2763885	22
## 17	Kansas	KS	North Central	2853118	63
## 4	Arkansas	AR	South	2915918	93
## 25	Mississippi	MS	South	2967297	120
## 16	Iowa	IA	North Central	3046355	21
## 7	Connecticut	CT	Northeast	3574097	97
## 37	Oklahoma	OK	South	3751351	111
## 38	Oregon	OR	West	3831074	36
## 18	Kentucky	KY	South	4339367	116
## 19	Louisiana	LA	South	4533372	351
## 41	South Carolina	SC	South	4625364	207
## 1	Alabama	AL	South	4779736	135
## 6	Colorado	CO	West	5029196	65
## 24	Minnesota	MN	North Central	5303925	53
## 50	Wisconsin	WI	North Central	5686986	97
## 21	Maryland	MD	South	5773552	293
## 26	Missouri	MO	North Central	5988927	321
## 43	Tennessee	TN	South	6346105	219
## 3	Arizona	AZ	West	6392017	232

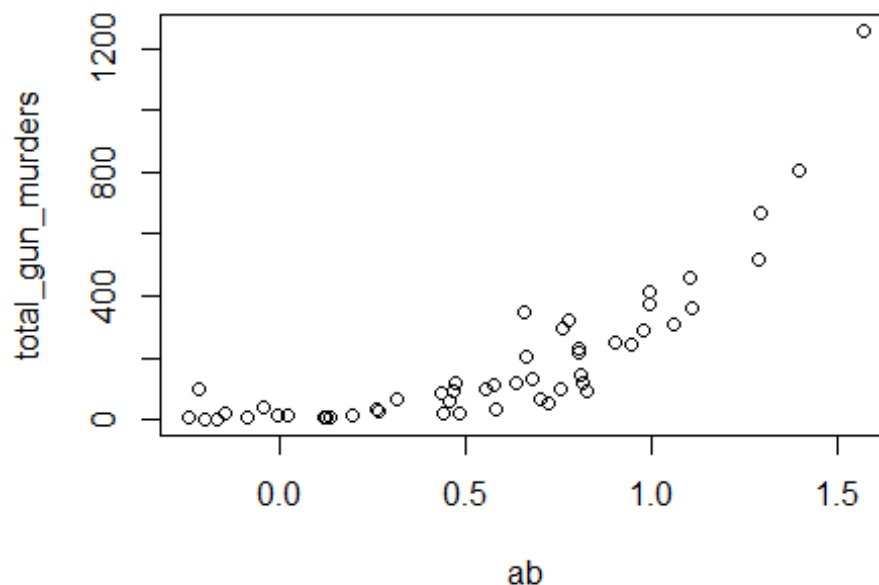
## 15	Indiana	IN	North Central	6483802	142
## 22	Massachusetts	MA	Northeast	6547629	118
## 48	Washington	WA	West	6724540	93
## 47	Virginia	VA	South	8001024	250
## 31	New Jersey	NJ	Northeast	8791894	246
## 34	North Carolina	NC	South	9535483	286
## 23	Michigan	MI	North Central	9883640	413
## 11	Georgia	GA	South	9920000	376
## 36	Ohio	OH	North Central	11536504	310
## 39	Pennsylvania	PA	Northeast	12702379	457
## 14	Illinois	IL	North Central	12830632	364
## 33	New York	NY	Northeast	19378102	517
## 10	Florida	FL	South	19687653	669
## 44	Texas	TX	South	25145561	805
## 5	California	CA	West	37253956	1257

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
ab = log10(population_in_millions)
plot(ab, total_gun_murders)

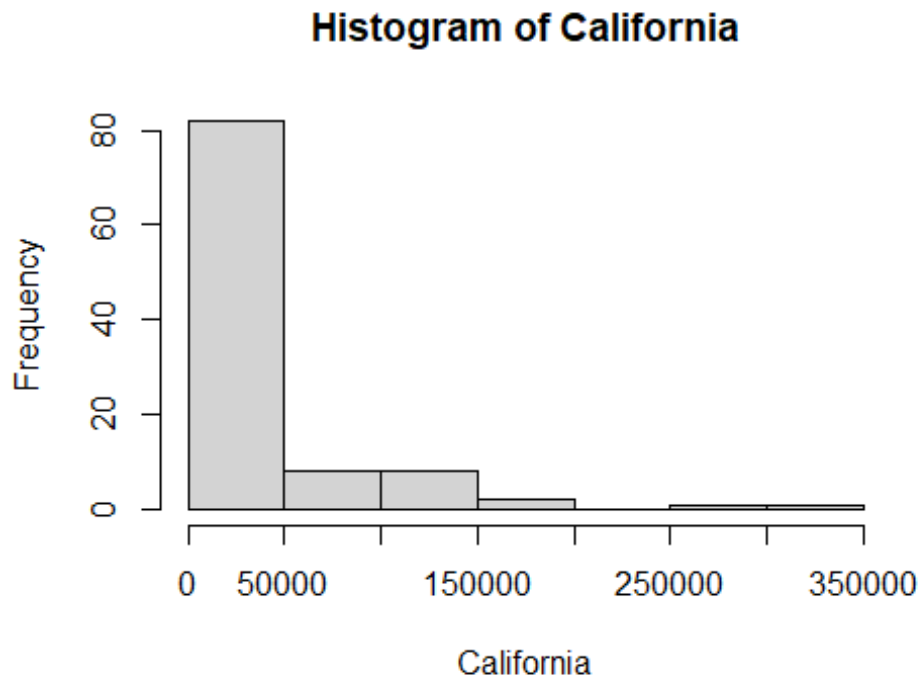
```



Buat histogram

dari populasi negara bagian.

```
Populasi = log10(murders$population)
California = c(Populasi, murders$population/murders$total)
hist(California)
```



9. Hasilkan boxplot

dari populasi negara bagian berdasarkan wilayahnya.

```
murders$rate = with(murders, total / population * 100000)
boxplot(rate~region, data = murders)
```

