Analisis data bivariat dan multivariat

Data bivariat

Data bivariat mendeskripsikan hubungan antar dua buah variabel. Misalnya:

- Hubungan antara berat dan tinggi badan.
- Hubungan antara risiko penyakit jantung dengan jenis kelamin.
- dll.

Data bivariat, terdiri dari:

- 2 variabel kualitatif
- 1 variabel kualitatif dan 1 variabel kuantitatif.
- 2 variabel kuantitatif.

Data bivariat kualitatif

```
ratings <-
factor(c(2,4,3,3,2,1,1,2,3,4,2,3,3,4,1,3,2,4,3,2,1))
ratings
```

1. 2
2. 4
3. 3
4. 3
5. 2
6. 1
7. 1
8. 2
9. 3
10. 4
11. 2
12. 3
13. 3
14. 4
15. 1
16. 3
17. 2
18. 4
19. 3
20. 2
21. 1

▶ **Levels**:

```
kursus <-
factor(c(1,1,0,0,1,1,0,0,1,0,0,0,1,0,1,0,1,1,1,0,1))
kursus
```

1. 1
2. 1
3. 0
4. 0
5. 1

6. 1
7. 0
8. 0
9. 1
10. 0
11. 0
12. 0
13. 1
14. 0
15. 1
16. 0
17. 1
18. 1
19. 1
20. 0
21. 1

▶ **Levels**:

```
levels(kursus) <- c('R', 'Python')
```

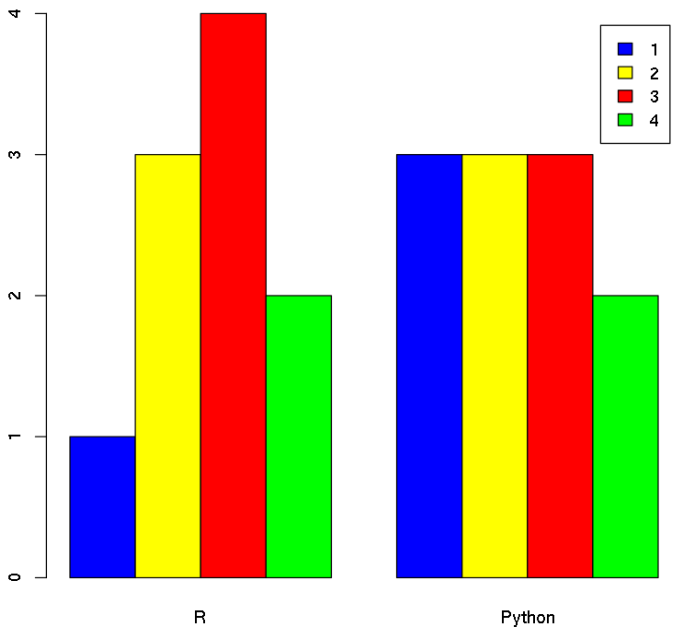
```
table(ratings, kursus)
```

```
       kursus
ratings R Python
      1 1      3
      2 3      3
      3 4      3
      4 2      2
```

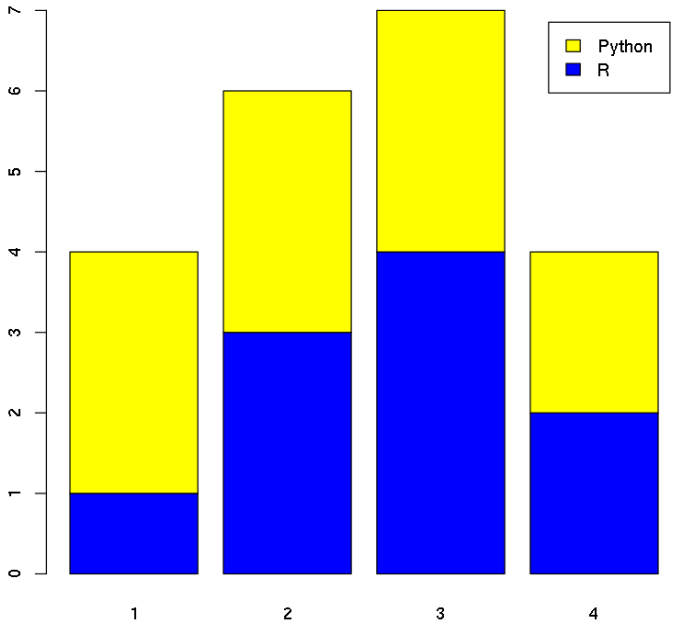
```
table(kursus, ratings)
```

```
        ratings
kursus   1 2 3 4
  R      1 3 4 2
  Python 3 3 3 2
```

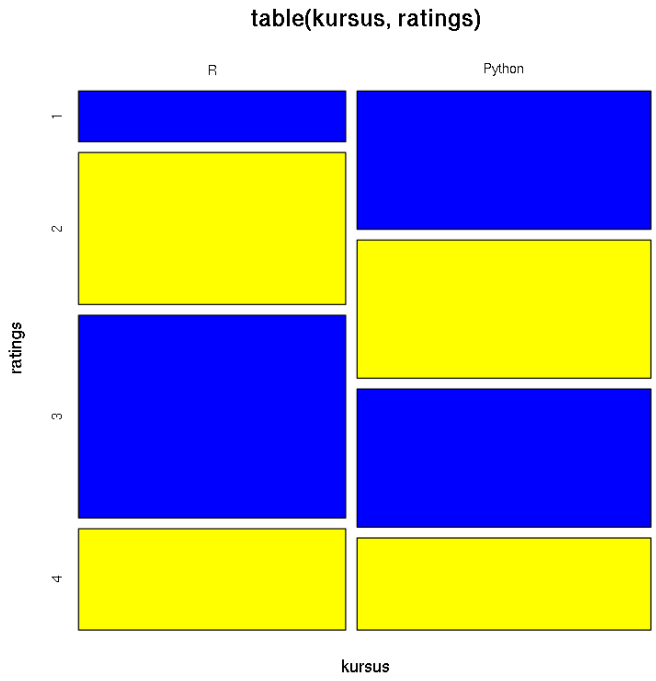
```
barplot(table(ratings, kursus),
       col=c('blue', 'yellow', 'red', 'green'), legend=T,
beside=T)
```




```
barplot(table(kursus, ratings),
        col=c('blue', 'yellow', 'red', 'green'),
        legend.text=T)
```



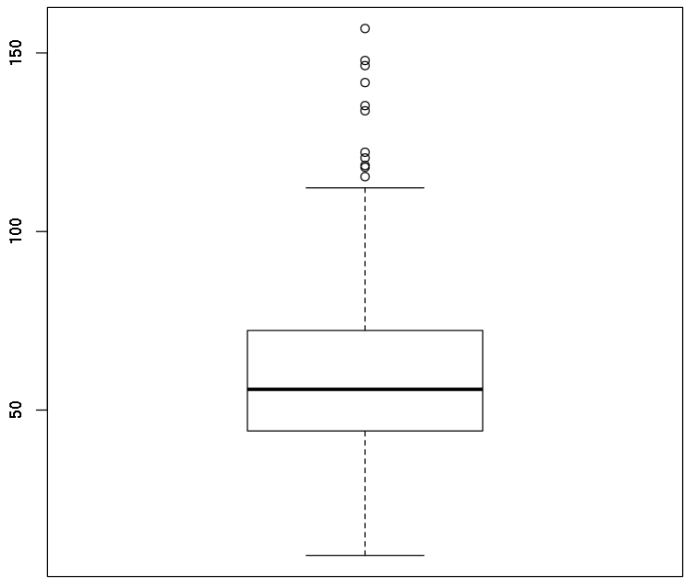

```
mosaicplot(table(kursus,ratings),
           col=c('blue', 'yellow'))
```



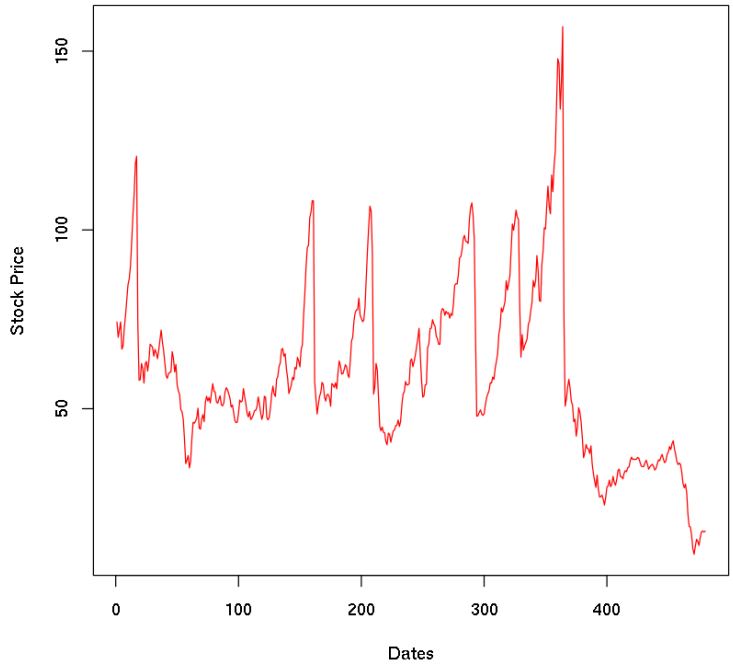
Data bivariat kuantitatif

```
library(dplyr)
df <- read.csv("../data/GEStock.csv")
dates <- select(df, Date)
price <- select(df, Price)
```

```
boxplot(price) # hanya untuk univariat
```



```
plot(df$Price,
    xlab='Dates',
    ylab='Stock Price',
    col='red',
    type='l')
```




```
max(df$Price)
```

156.8436842

```
which(df$Price == max(df$Price)) # indeks maksimum
```

364

```
df[which(df$Price == max(df$Price)),]
```

	DATE	PRICE
364	4/1/00	156.8437

Data multivariat

```
df <- read.csv("../data/murders.csv")
head(df)
```

STATE	ABB	REGION	POPULATION	POPULATIONDENSITY	MURDERS	GUNMURDERS	GUNOW

STATE	ABB	REGION	POPULATION	POPULATIONDENSITY	MURDERS	GUNMURDERS	GUNOW
Alabama	AL	South	4779736	94.65	199	135	0.517
Arizona	AZ	West	6392017	57.05	352	232	0.311
California	CA	West	37253956	244.20	1811	1257	0.213
Colorado	CO	West	5029196	49.33	117	65	0.347
Connecticut	CT	Northeast	3574097	741.40	131	97	0.167
Florida	FL	South	19687653	360.20	987	669	0.245

```
str(df)
```

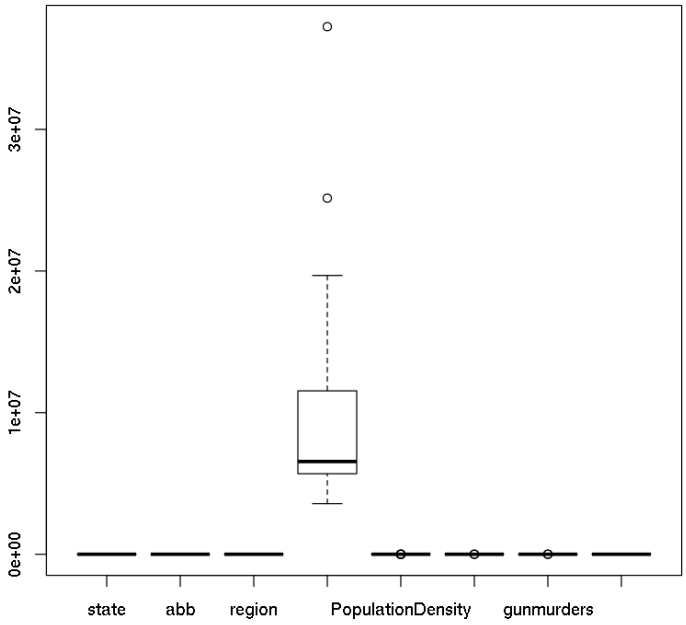
```
'data.frame':   25 obs. of  8 variables:
 $ state            : Factor w/ 25 levels
"Alabama","Arizona",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ abb              : Factor w/ 25 levels
"AL","AZ","CA",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ region           : Factor w/ 4 levels "North
Central",..: 3 4 4 4 2 3 3 1 1 3 ...
 $ population       : int  4779736 6392017 37253956
5029196 3574097 19687653 9920000 12830632 6483802 4339367
...
 $ PopulationDensity: num  94.7 57 244.2 49.3 741.4 ...
 $ murders          : int  199 352 1811 117 131 987 527
453 198 180 ...
 $ gunmurders       : int  135 232 1257 65 97 669 376 364
142 116 ...
 $ gunownership     : num  0.517 0.311 0.213 0.347 0.167
0.245 0.403 0.202 0.391 0.477 ...
```

```
summary(df)
```

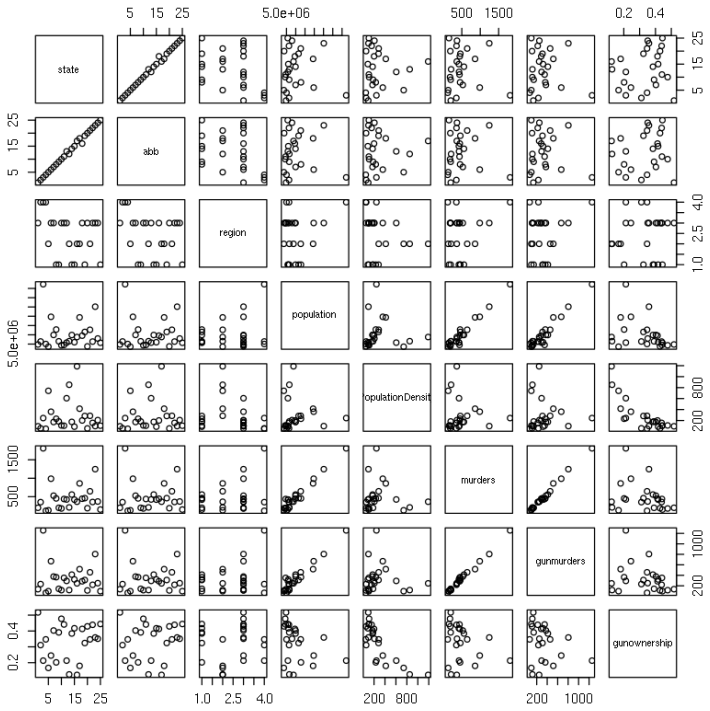
```
         state         abb               region
population
 Alabama    : 1   AL     : 1   North Central: 6   Min.   :
3574097
 Arizona    : 1   AZ     : 1   Northeast    : 5   1st Qu.:
5686986
 California : 1   CA     : 1   South        :11   Median :
6547629
 Colorado   : 1   CO     : 1   West         : 3   Mean
:10155719
 Connecticut: 1   CT     : 1                      3rd
Qu.:11536504
 Florida    : 1   FL     : 1                      Max.
:37253956
 (Other)    :19   (Other):19

 PopulationDensity    murders         gunmurders
gunownership
 Min.   :  49.33   Min.   : 117.0   Min.   :  65.0   Min.
  :0.1230
 1st Qu.: 105.00   1st Qu.: 199.0   1st Qu.: 135.0   1st
Qu.:0.2130
```

```
 Median : 182.50   Median : 419.0   Median : 286.0
Median :0.3510
 Mean   : 282.57   Mean   : 483.4   Mean   : 329.9   Mean
  :0.3305
 3rd Qu.: 285.30   3rd Qu.: 527.0   3rd Qu.: 376.0   3rd
Qu.:0.4170
 Max.   :1189.00   Max.   :1811.0   Max.   :1257.0   Max.
  :0.5170
```

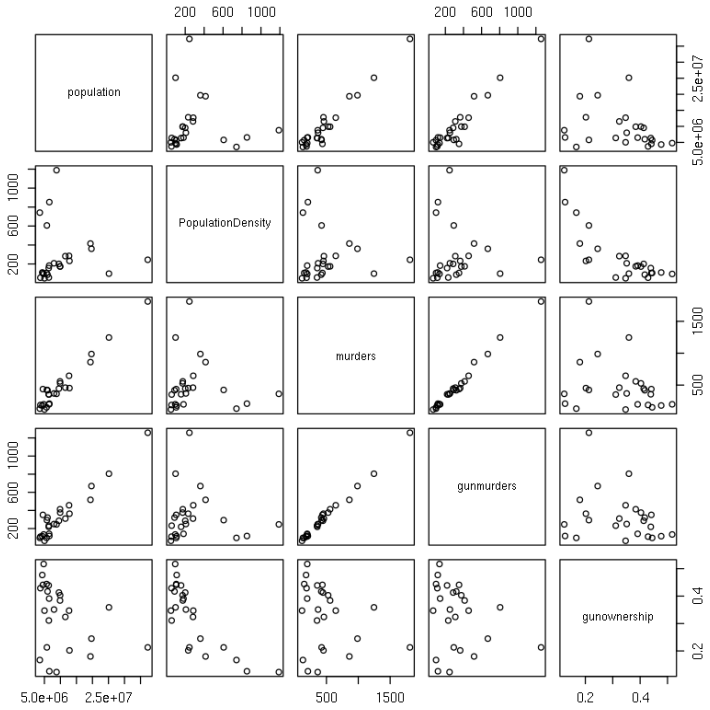
```
boxplot(df)
```




```
plot(df)
```




```
pairs(df[,-c(1,2,3)]) # tidak memasukan kolom no 1,2,3
```




```
dfsel <- df[,-c(1,2,3,4,5)] # hanya memasukan data
kuantitatif
```

```
mat <- data.matrix(dfsel) # konversi data terseleksi ke
matriks
mat <- t(mat)
```

```
barplot(mat,
        col=c('red', 'yellow',
              'green'),
        beside=T,
        names.arg=dfsel$state,
        legend.text=T)
```