

Data kualitatif dan kuantitatif

Pendahuluan

- Di R, data umumnya disimpan dalam bentuk vektor atau data frame.
- Data kualitatif, di dalam statistika dikenal sebagai data kategorikal.
- Data kualitatif dapat disimpan dalam bentuk *Factors*.
- Data kuantitatif, di dalam statistika dikenal sebagai data kontinyu atau data numerik.
- Data kuantitatif dapat disimpan dalam bentuk *Numeric*.

Data kualitatif

Data kualitatif merupakan data non-statistik yang umumnya bersifat tidak terstruktur atau semi-terstruktur.

- Data kualitatif tidak melulu berasal dari pengukuran.
- Data kualitatif dikategorikan berdasarkan sifat - sifat, atribut, label, dll.
- Data ini digunakan untuk interpretasi dan pembuatan hipotesis.
- Data ini tidak dapat dikumpulkan dan dianalisa menggunakan metode - metode konvensional.

Contoh - contoh data kualitatif:

- Jenis kelamin,
- Ukuran sepatu,
- *Rating*.

```
ukuranBaju <- c('S', 'M', 'L', 'XL',  
                'XXL', 'M', 'L', 'XL',  
                'XXL', 'S', 'M')
```

ukuranBaju

1. 'S'
2. 'M'
3. 'L'
4. 'XL'
5. 'XXL'
6. 'M'
7. 'L'
8. 'XL'
9. 'XXL'
10. 'S'
11. 'M'

```
ukuran_baju <- factor(ukuranBaju) # dijadikan dalam bentuk Factor
ukuran_baju
```

```
1. S
2. M
3. L
4. XL
5. XXL
6. M
7. L
8. XL
9. XXL
10. S
11. M
```

► **Levels:**

```
str(ukuran_baju)
```

```
Factor w/ 5 levels "L","M","S","XL",...: 3 2 1 4 5 2 1 4 5 3 ...
```

```
summary(ukuran_baju)
```

```
L
  2
M
  3
S
  2
XL
  2
XXL
  2
```

```
levels(ukuranBaju)
```

```
NULL
```

```
levels(ukuran_baju)
```

```
1. 'L'
2. 'M'
```

3. 'S'
4. 'XL'
5. 'XXL'

Visualisasi data kualitatif

Gunakan:

- Diagram batang,
- Diagram lingkaran.

```
ukuran_baju
```

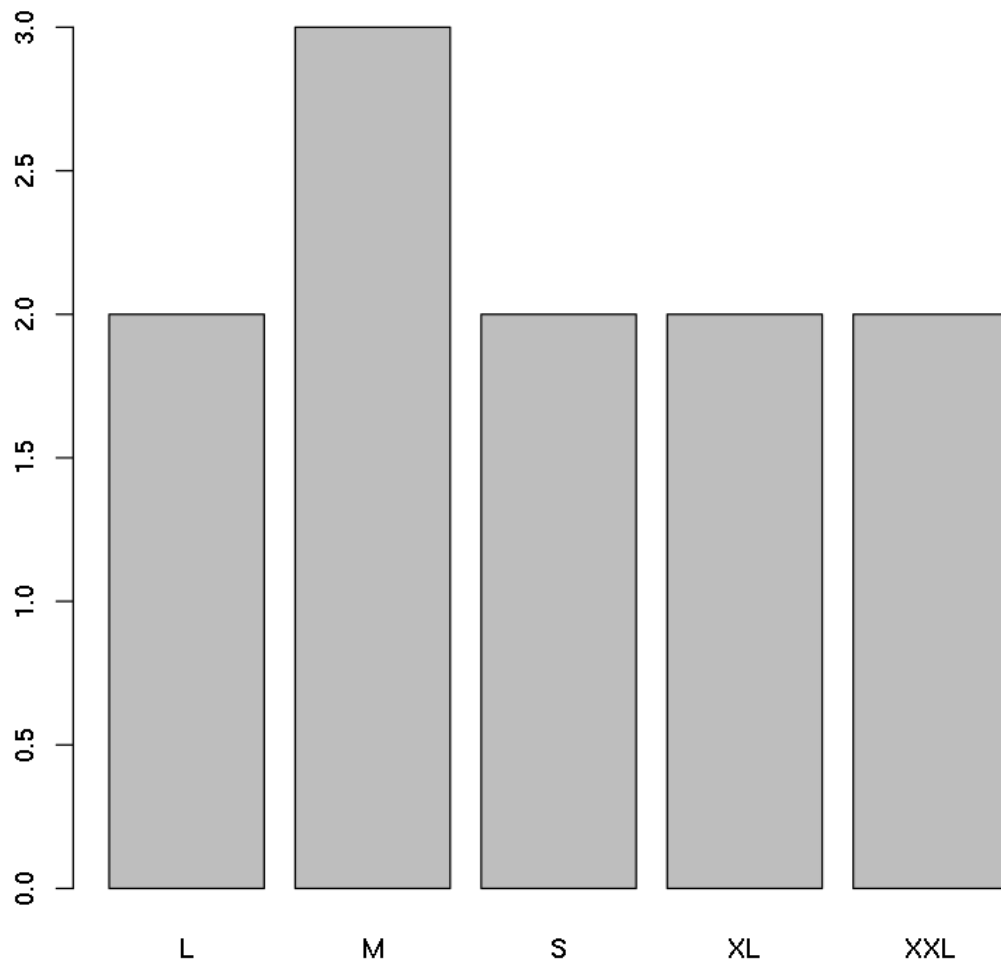
1. S
2. M
3. L
4. XL
5. XXL
6. M
7. L
8. XL
9. XXL
10. S
11. M

► **Levels:**

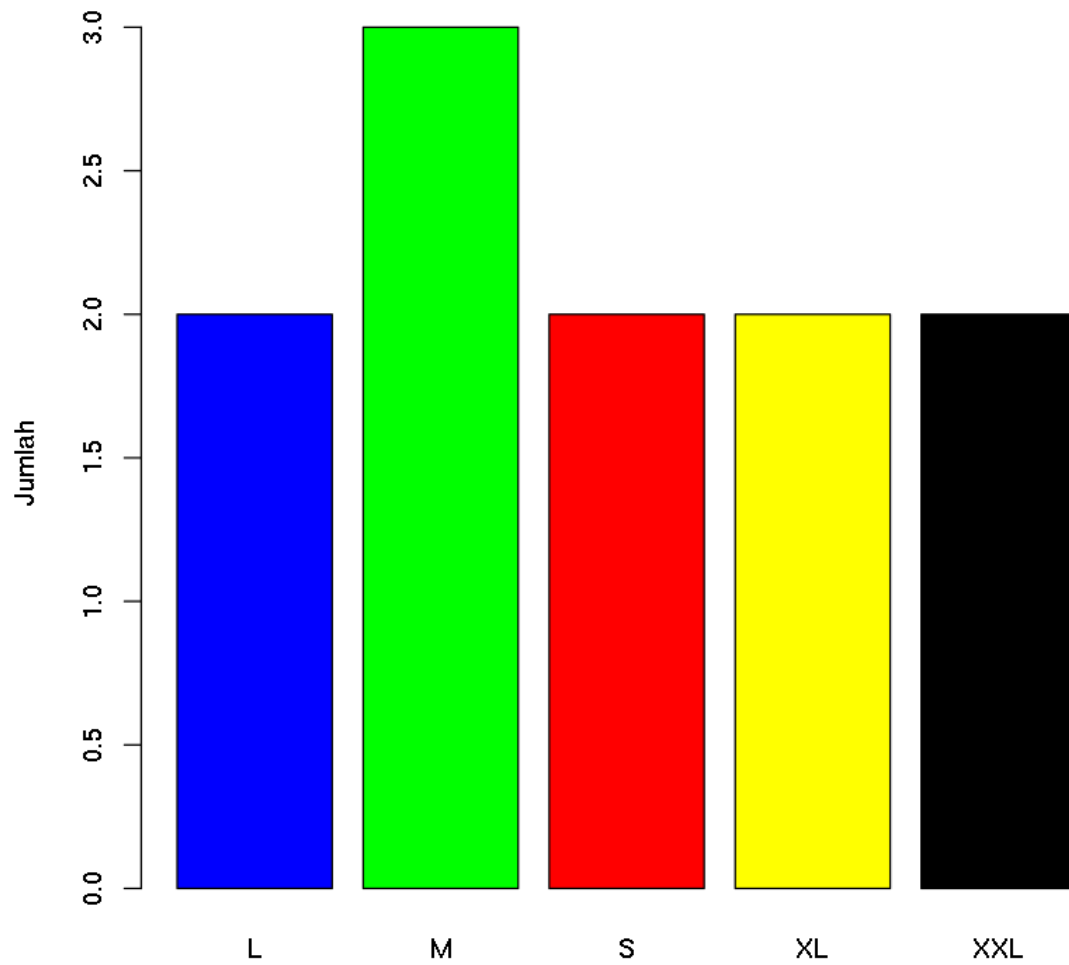
```
tabelUkuranBaju <- table(ukuran_baju)
tabelUkuranBaju
```

```
ukuran_baju
  L  M  S  XL XXL
2  3  2  2  2
```

```
barplot(tabelUkuranBaju)
```



```
# kostumisasi diagram batang
barplot(tabelUkuranBaju,
        col = c('blue', 'green', 'red', 'yellow', 'black'),
        ylab = 'Jumlah')
```



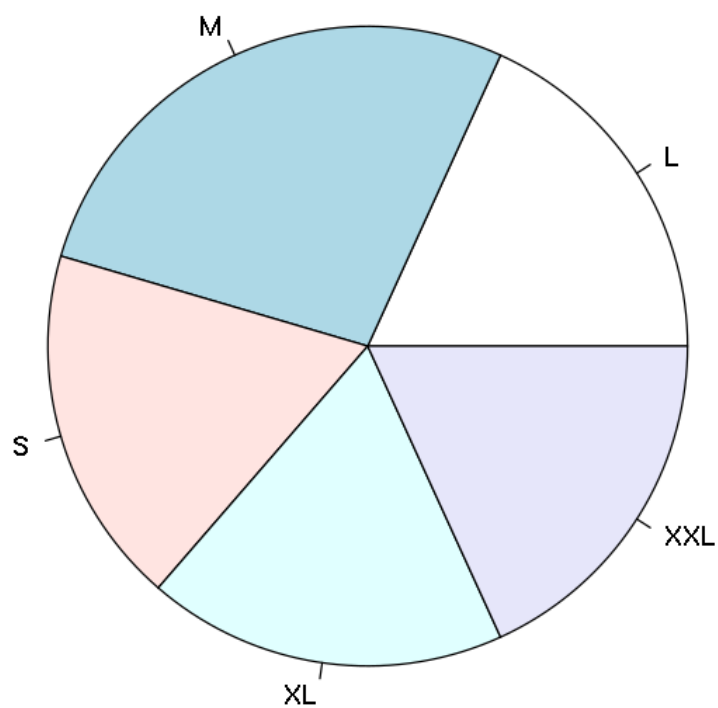
```
# misalkan kita hanya ingin ukuran M  
ukuran_baju == 'M'
```

1. FALSE
2. TRUE
3. FALSE
4. FALSE
5. FALSE
6. TRUE
7. FALSE
8. FALSE
9. FALSE
10. FALSE
11. TRUE

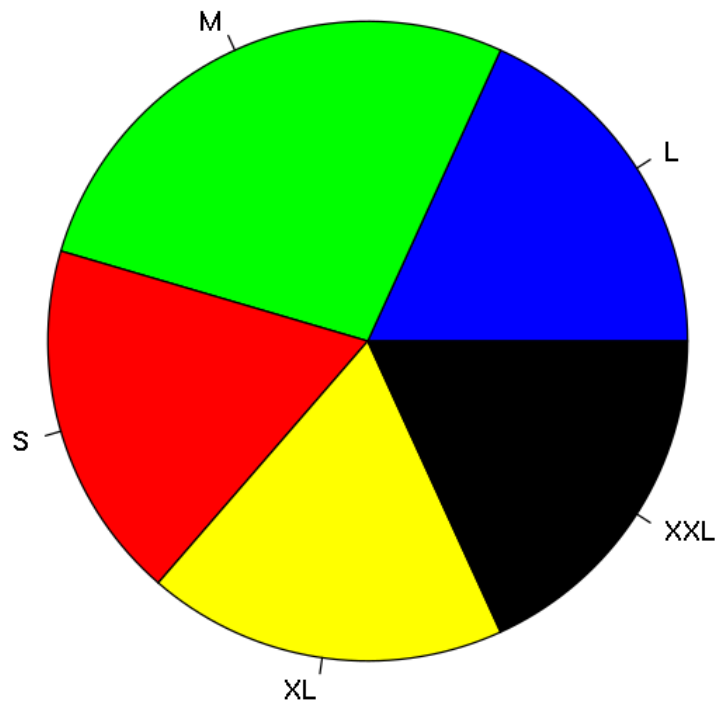
```
jumlah_ukuranM = sum(ukuran_baju == 'M')  
jumlah_ukuranM
```

3

```
# Penggunaan diagram lingkaran  
pie(tabelUkuranBaju)
```



```
# kostumisasi diagram lingkaran  
pie(tabelUkuranBaju, col = c('blue', 'green', 'red', 'yellow', 'black'))
```



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

```
levels(kategori_usia)
```

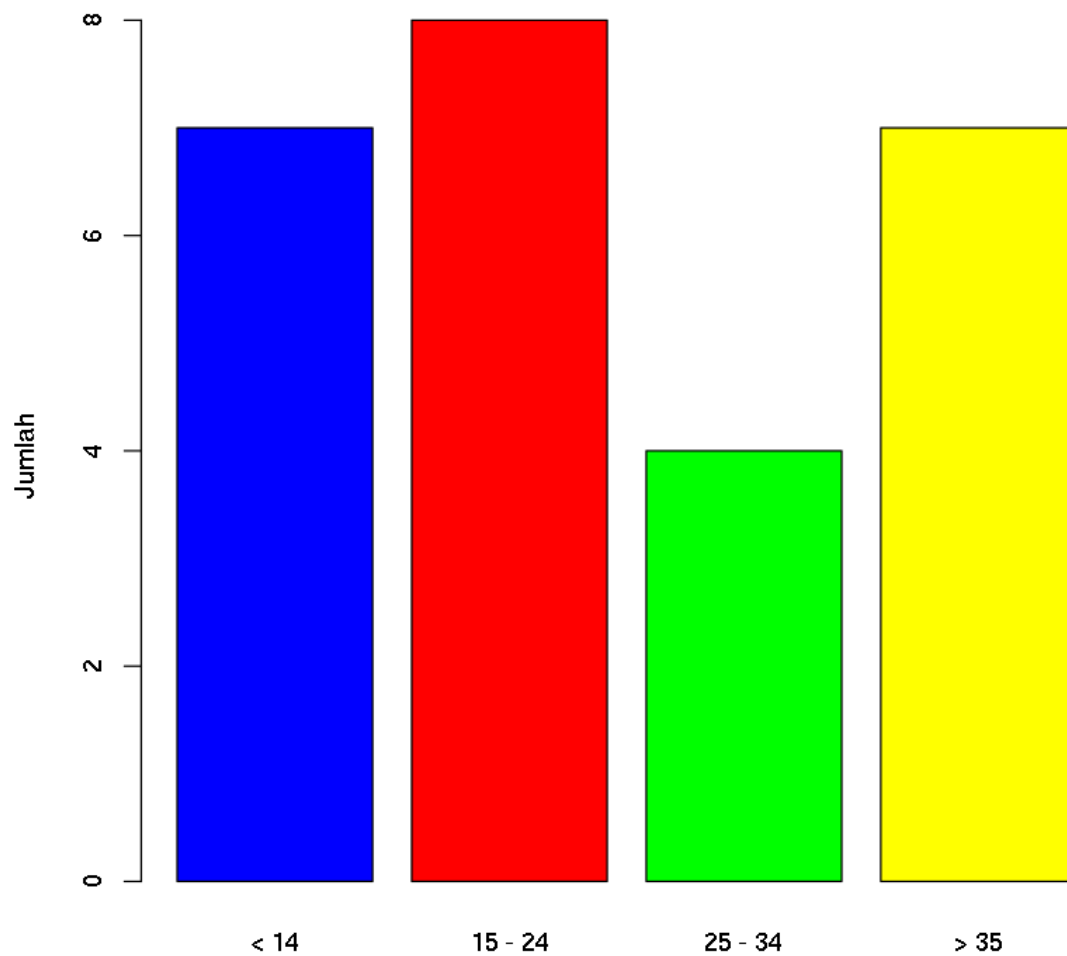
1. '1'
2. '2'
3. '3'
4. '4'

```
# mengubah level kategori usia  
levels(kategori_usia) <- c('< 14', '15 - 24', '25 - 34', '> 35')
```

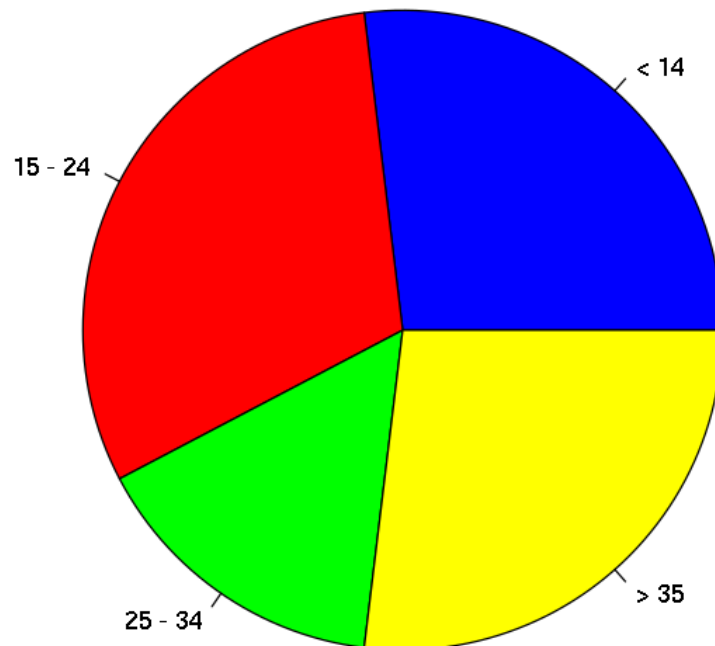
```
tabelKategoriUsia <- table(kategori_usia)  
tabelKategoriUsia
```

```
kategori_usia  
< 14 15 - 24 25 - 34 > 35  
7      8      4      7
```

```
barplot(tabelKategoriUsia,  
        col=c('blue', 'red', 'green', 'yellow'),  
        ylab = 'Jumlah')
```



```
pie(tabelKategoriUsia, col=c('blue', 'red', 'green', 'yellow'))
```

Data kuantitatif

- Data kuantitatif dapat dihitung, diukur, dan diekspresikan secara numerik.
- Data kualitatif sendiri bersifat deskriptif dan konseptual.
- Data kuantitatif bersifat terstruktur.

Contoh:

- Pengukuran temperatur udara,
- Harga saham, dll.

```
# panjang lagu (dalam menit)
lagu <- c(5.3, 3.6, 5.5, 4.7, 6.7, 4.3, 6.2, 4.3, 4.9, 5.1, 5.8, 4.4)
lagu
```

1. 5.3
2. 3.6
3. 5.5
4. 4.7
5. 6.7

- 6. 4.3
- 7. 6.2
- 8. 4.3
- 9. 4.9
- 10. 5.1
- 11. 5.8
- 12. 4.4

Visualisasi data kuantitatif

- Histogram
- *Boxplot*
- *strip-chart* \rightarrow alternatif *boxplot* ketika ukuran sampel kecil.

Histogram

```
length(lagu) # jumlah elemen di dalam vektor lagu
```

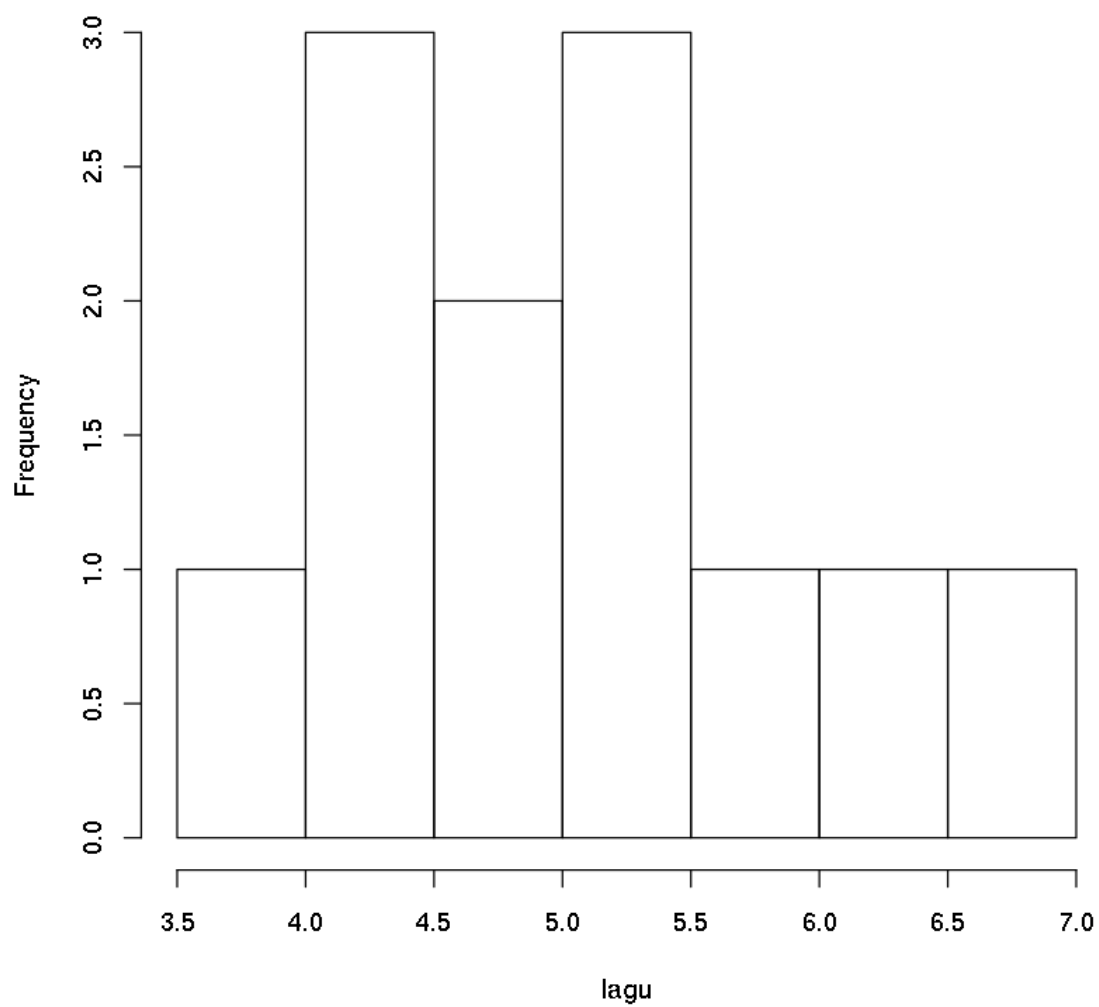
12

```
summary(lagu)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.600	4.375	5.000	5.067	5.575	6.700

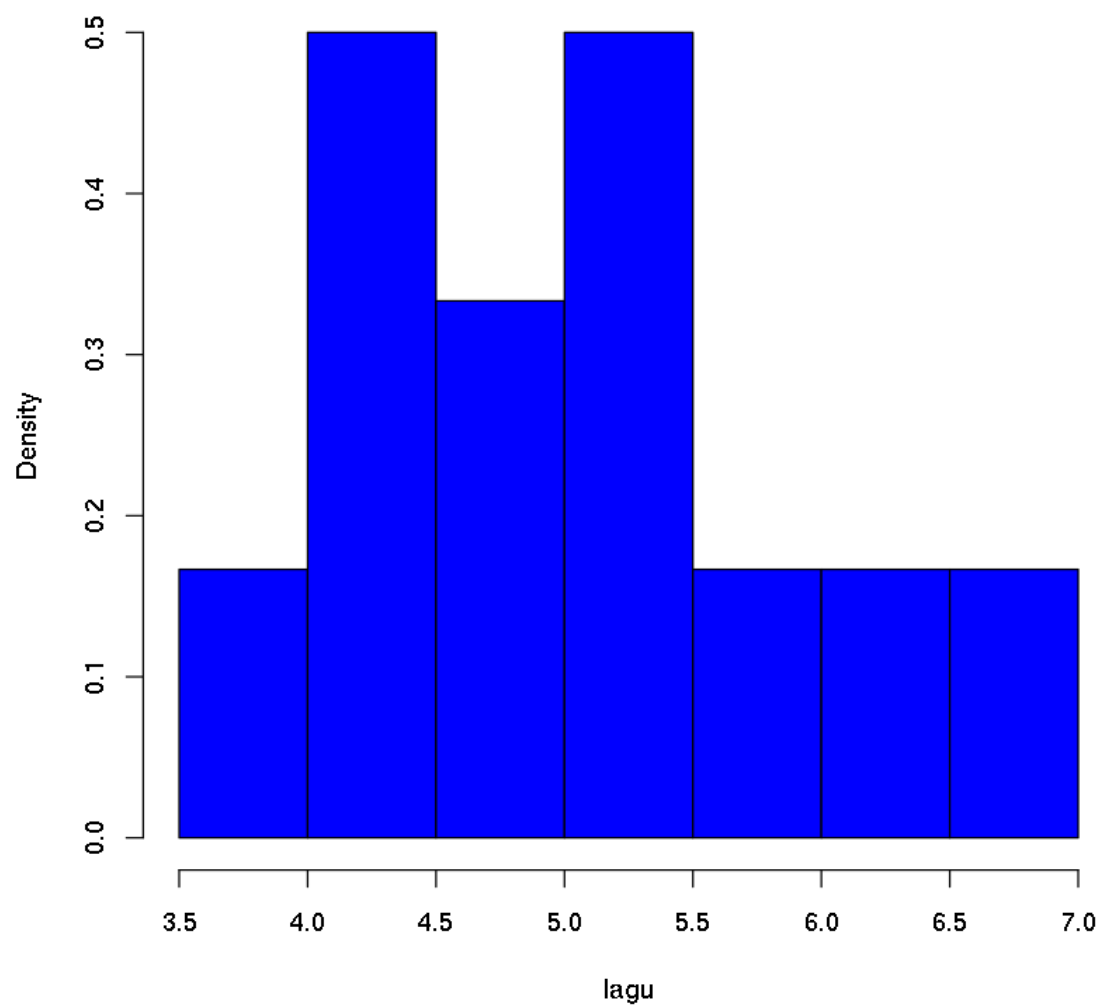
```
hist(lagu)
```

Histogram of lagu



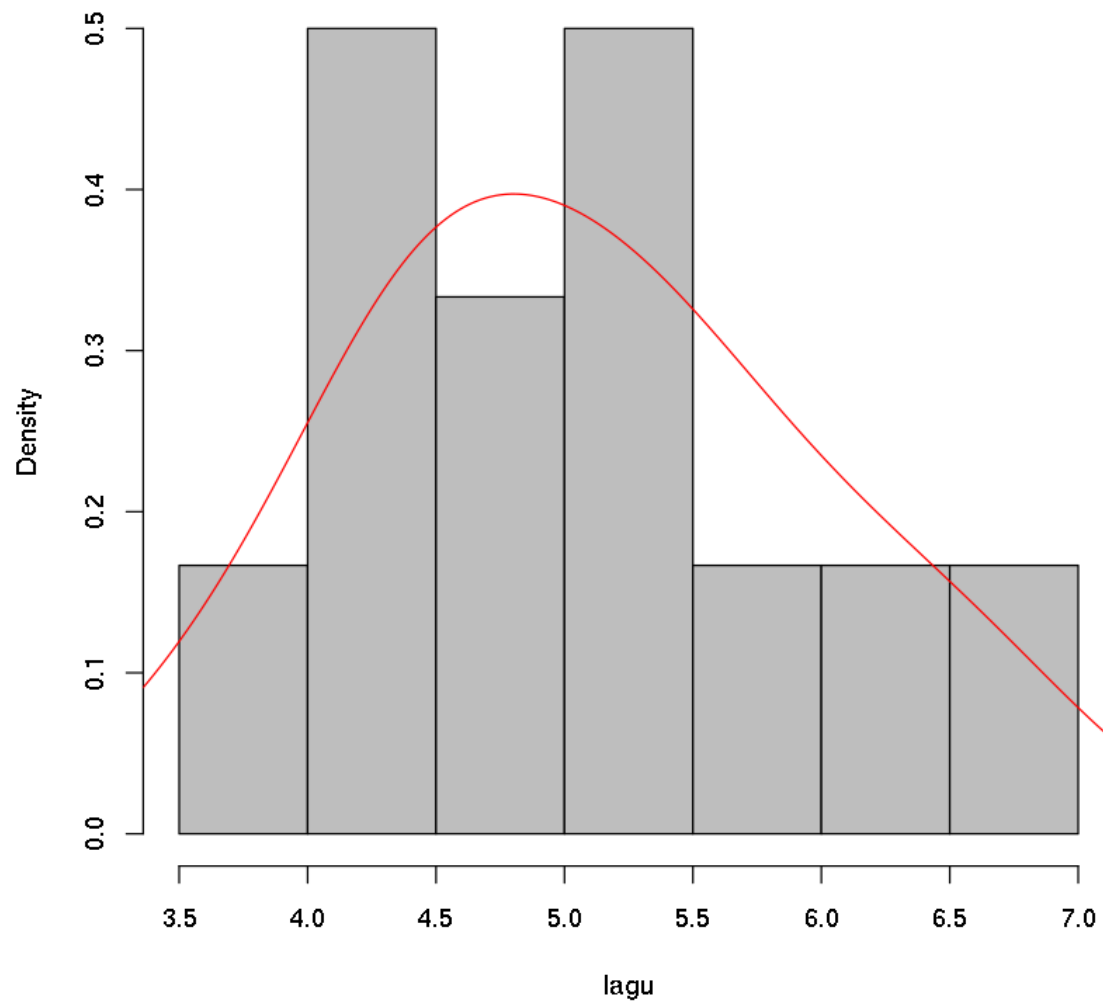
```
hist(lagu, col='blue', prob=T) # pdf: probability density function
```

Histogram of lagu



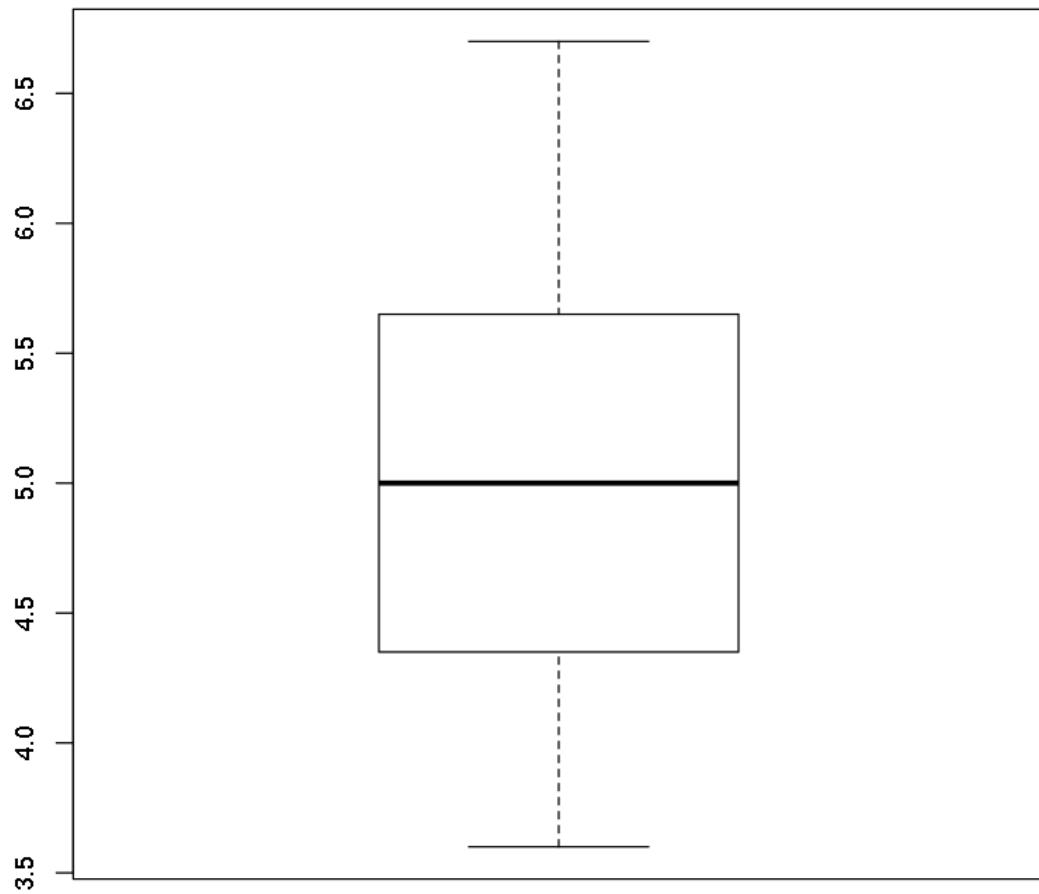
```
hist(lagu, col='grey', prob=T)
lines(density(lagu), col='red')
```

Histogram of lagu



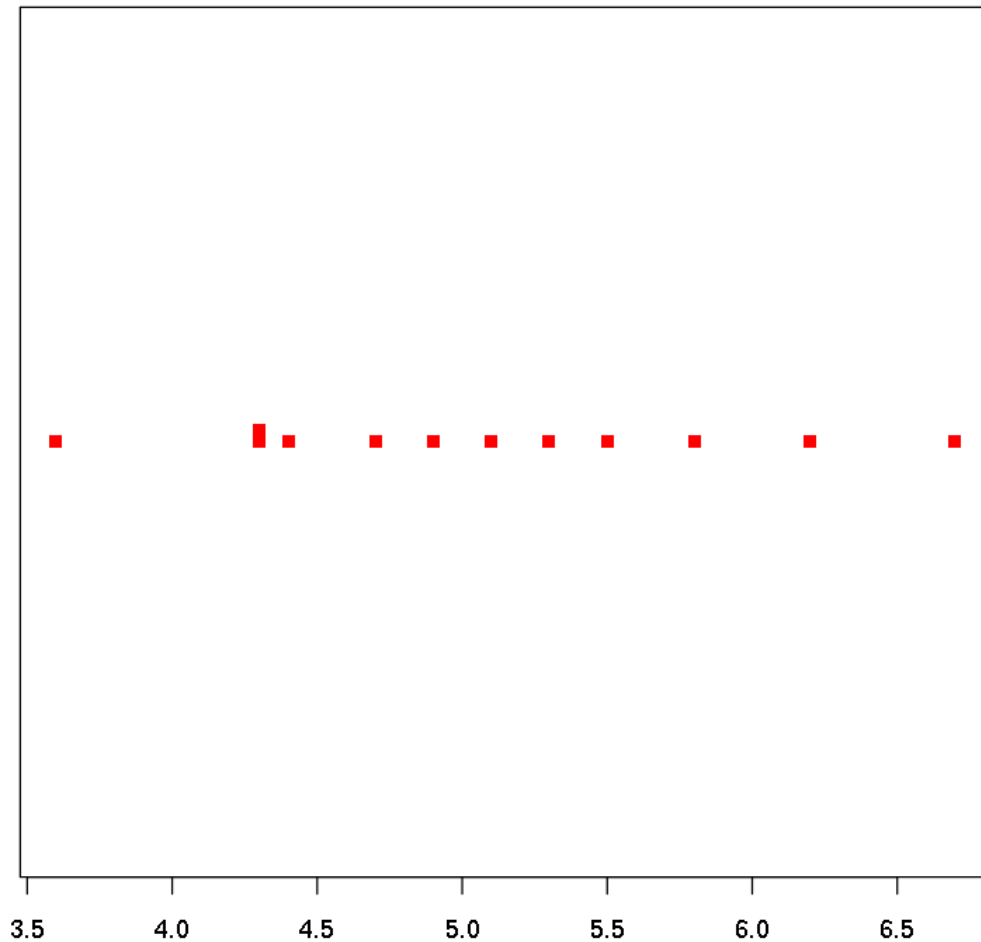
boxplot

```
boxplot(lagu)
```



strip-chart

```
stripchart(lagu,col='red', pch=15, method='stack')
```



Visualisasi data saham

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
gedata <- read.csv("../data/GESStock.csv")
```

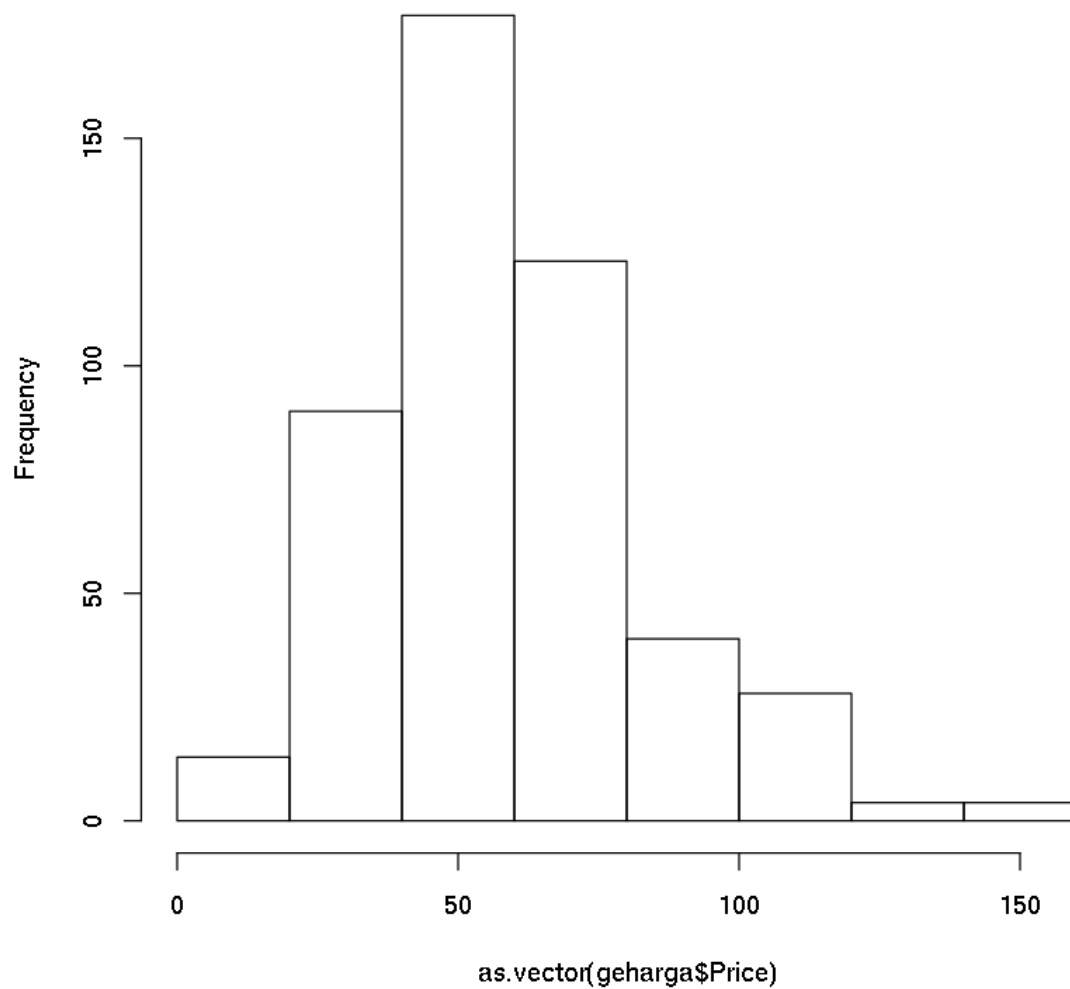
```
geharga <- select(gedata, Price)
```

```
summary(geharga)
```

```
Price
Min.   : 9.294
1st Qu.: 44.214
Median : 55.812
Mean   : 59.303
3rd Qu.: 72.226
Max.   :156.844
```

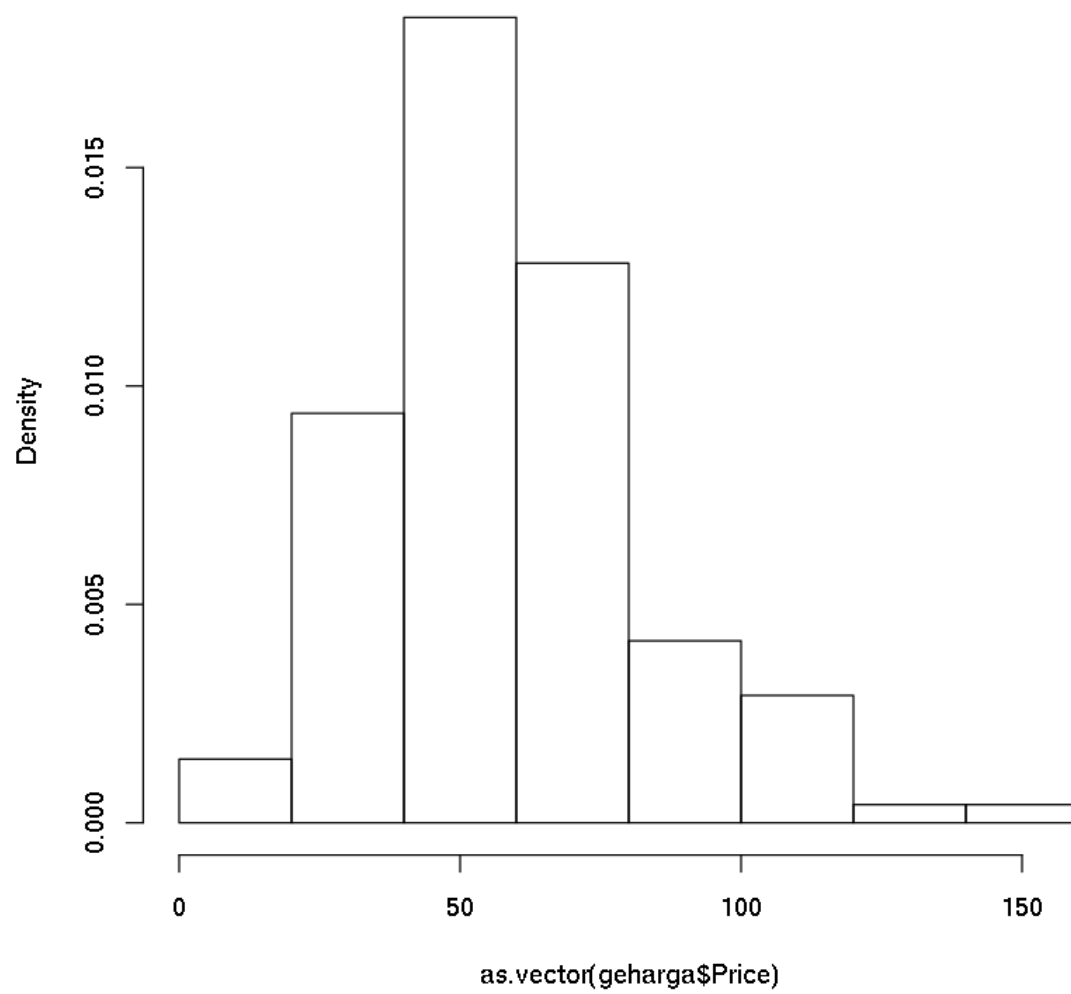
```
hist(as.vector(geharga$Price))
```

Histogram of as.vector(geharga\$Price)



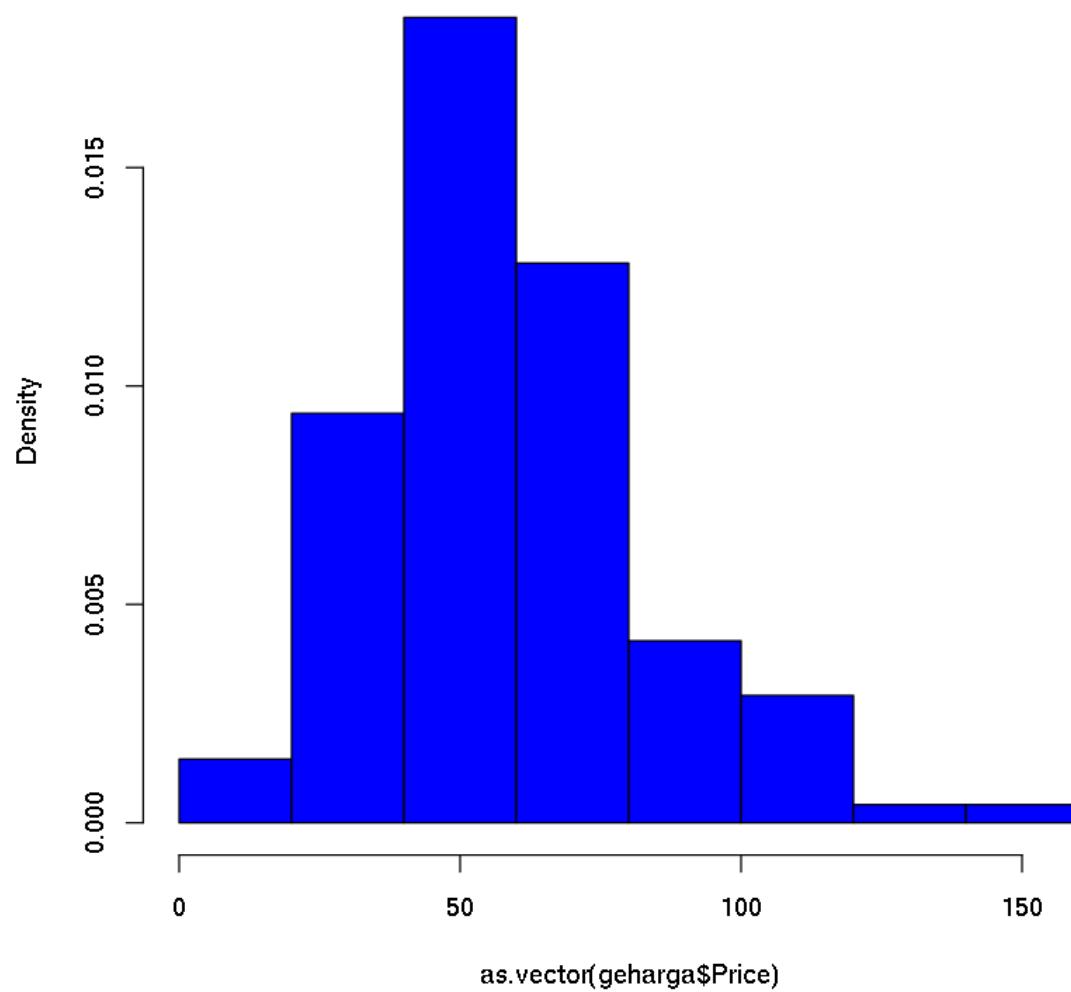
```
hist(as.vector(geharga$Price), prob=T)
```


Histogram of as.vector(geharga\$Price)



```
hist(as.vector(geharga$Price), prob=T, col='blue')
```

Histogram of as.vector(geharga\$Price)



```
hist(as.vector(geharga$Price), prob=T, col='blue', ylim=c(0,0.025))  
lines(density(geharga$Price), col='red')
```

Histogram of as.vector(geharga\$Price)

