

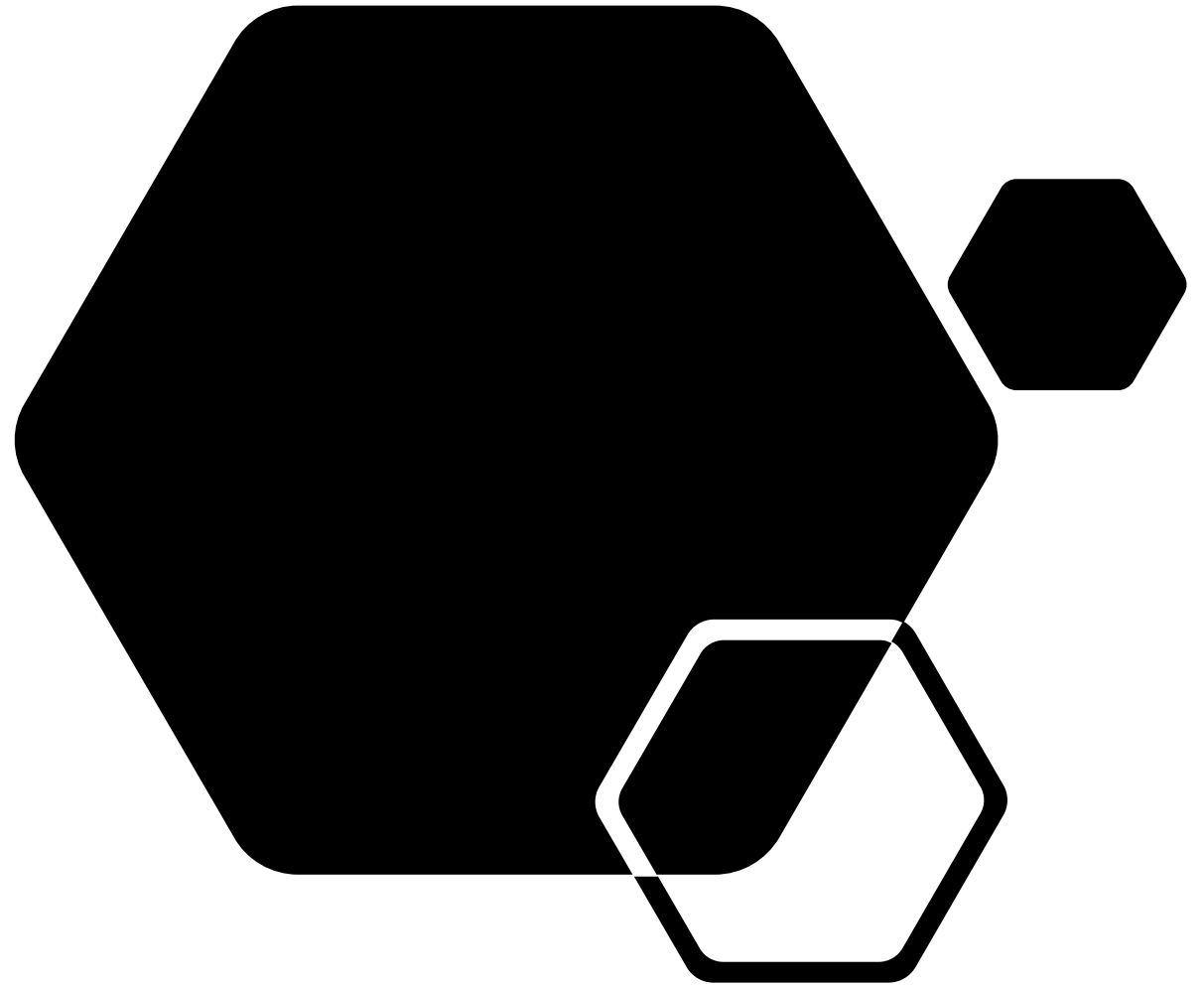


# *Analisis Data Eksploratif*

## *Pertemuan 4*

### *Visualisasi Data*

# *Visualisasi Data*



# Diagram Batang

- Pada umumnya digunakan untuk menggambarkan perkembangan nilai-nilai suatu objek penelitian dalam kurun waktu tertentu.
- Diagram batang menunjukkan berbagai keterangan dengan batang-batang tegak ataupun mendatar dan sama lebar dengan batang-batang terpisah.

# Contoh

Menggunakan data VADeaths pada Rstudio diperoleh data sbb:

```
> VADeaths
```

	Rural	Male	Rural	Female	Urban	Male	Urban	Female
50-54		11.7		8.7		15.4		8.4
55-59		18.1		11.7		24.3		13.6
60-64		26.9		20.3		37.0		19.3
65-69		41.0		30.9		54.6		35.1
70-74		66.0		54.3		71.1		50.0

- Diagram Batang Untuk Rural Male

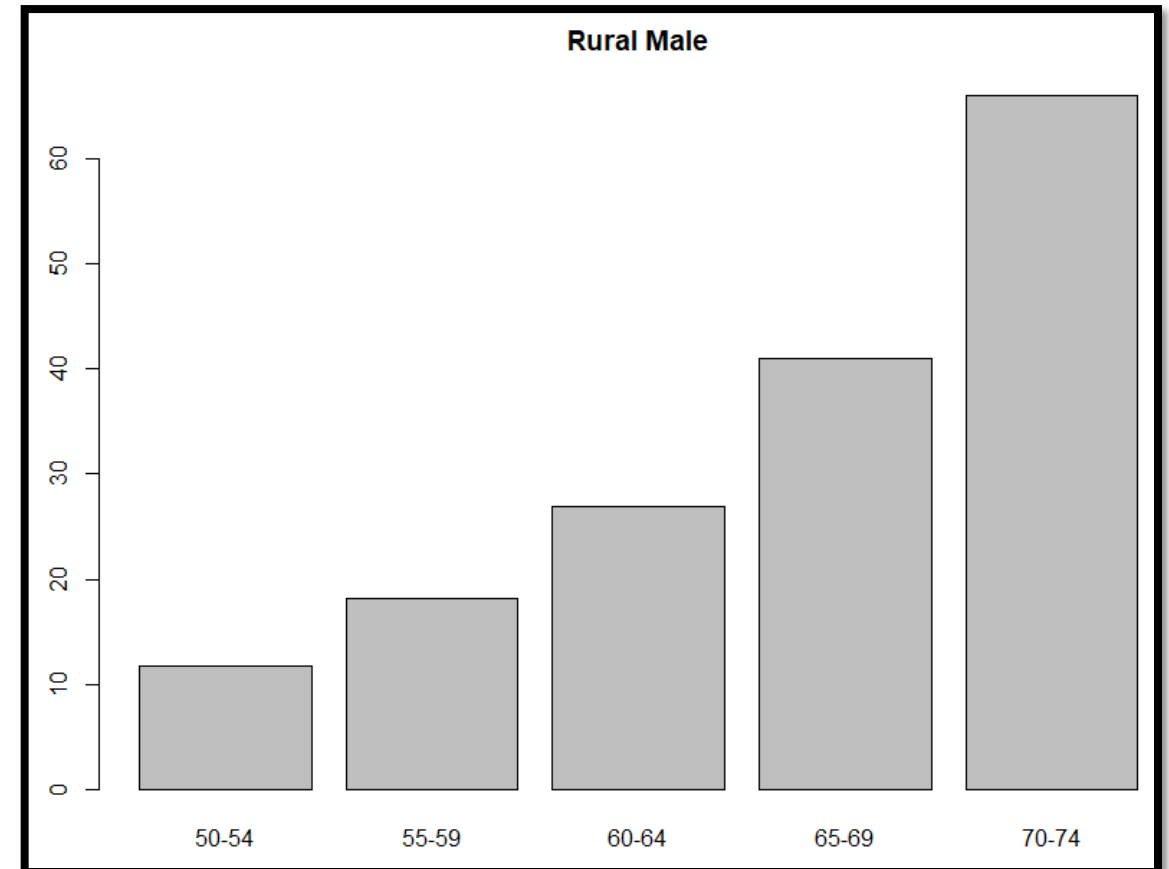
```
#Membuat Barplot
```

```
>
```

```
> barplot(VADeaths[, "Rural Male"], main="Rural Male")
```

Informasi yang dapat diambil dari diagram batang disamping:

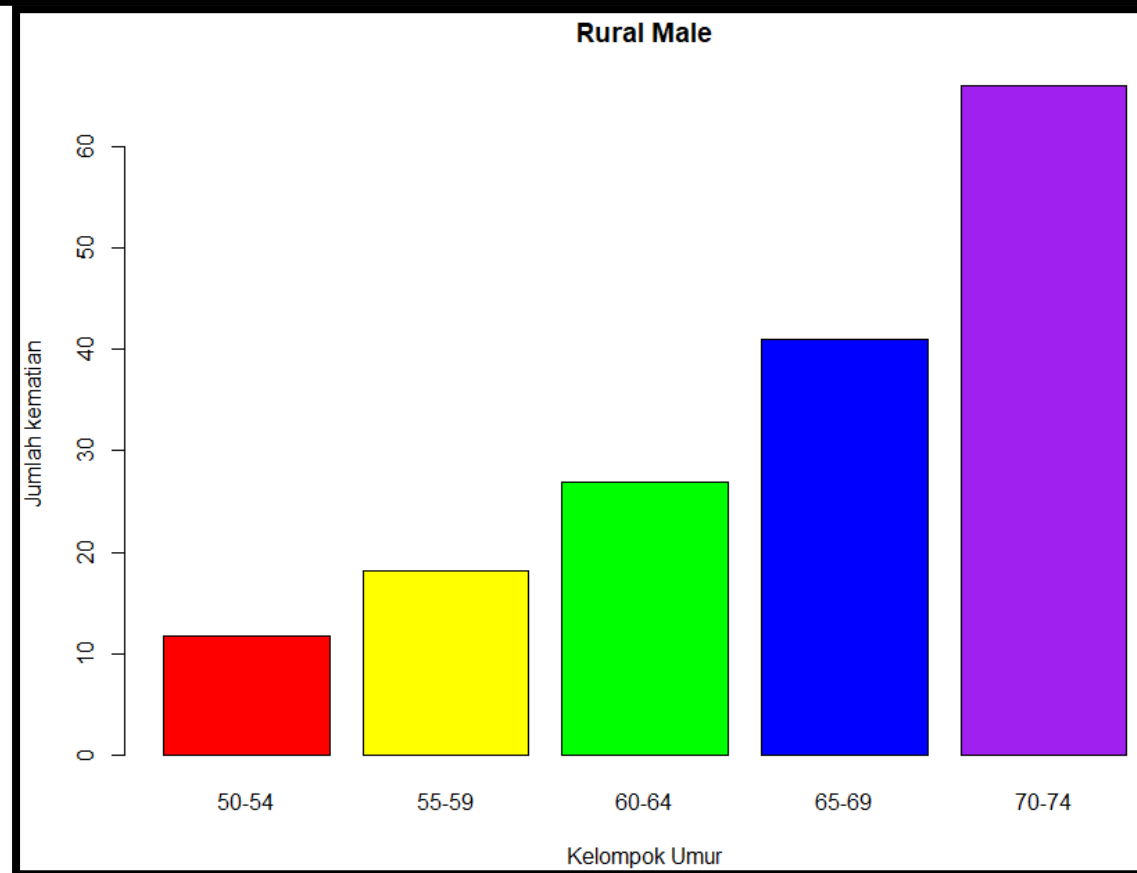
1. Kelompok usia 70-74 pada kategori Rural Male memiliki kerentanan untuk meninggal paling tinggi
2. Kelompok usia 50 – 54 pada kategori Rural Male memiliki kerentanan terendah dibandingkan kelompok usia lainnya



Atau dapat juga diubah warnanya dengan syntax

```
barplot(VADeaths[, "Rural Male"], main="Rural Male", xlab="Kelompok Umur", ylab="Jumlah kematian", border = "black", col=c("red","yellow","green", "blue", "purple"))
```

**Hasil**



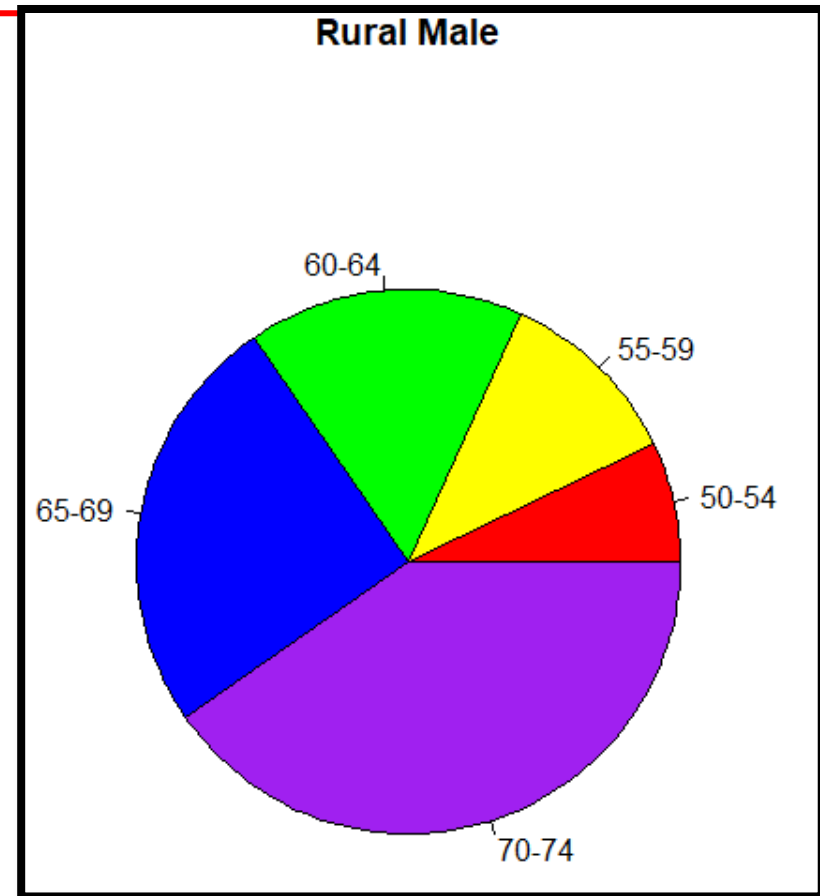
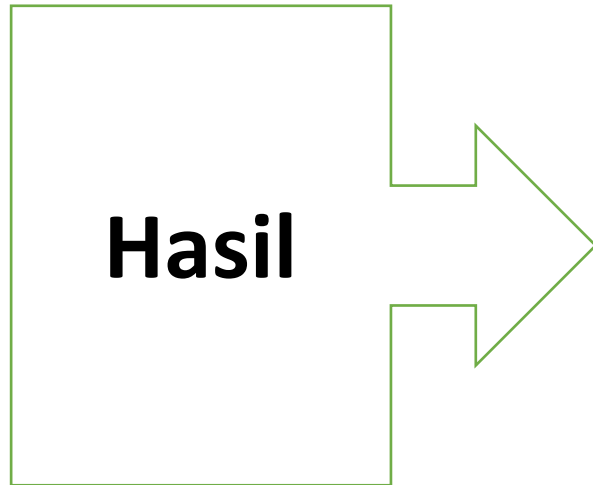
# Diagram Lingkaran

- Penyajian data statistik dengan memakai gambar yang berbentuk lingkaran. Lalu bagian-bagian dari daerah lingkaran, menunjukkan persen data.
- Untuk membuat diagram lingkaran, pertama-tama terlebih dahulu ditentukannya besar persentase tiap objek terhadap keseluruhan data dan besarnya sudut pusat sektor lingkaran.



# Contoh

```
pie(VADeaths[, "Rural Male"], main="Rural Male", border = "black",  
col=c("red","yellow","green", "blue", "purple"))
```



# Stem & Leaf

- Dapat digunakan untuk melihat distribusi data

```
> stem(VADeaths)
```

The decimal point is 1 digit(s) to the right of the |

```
0 | 89224589
```

```
2 | 047157
```

```
4 | 1045
```

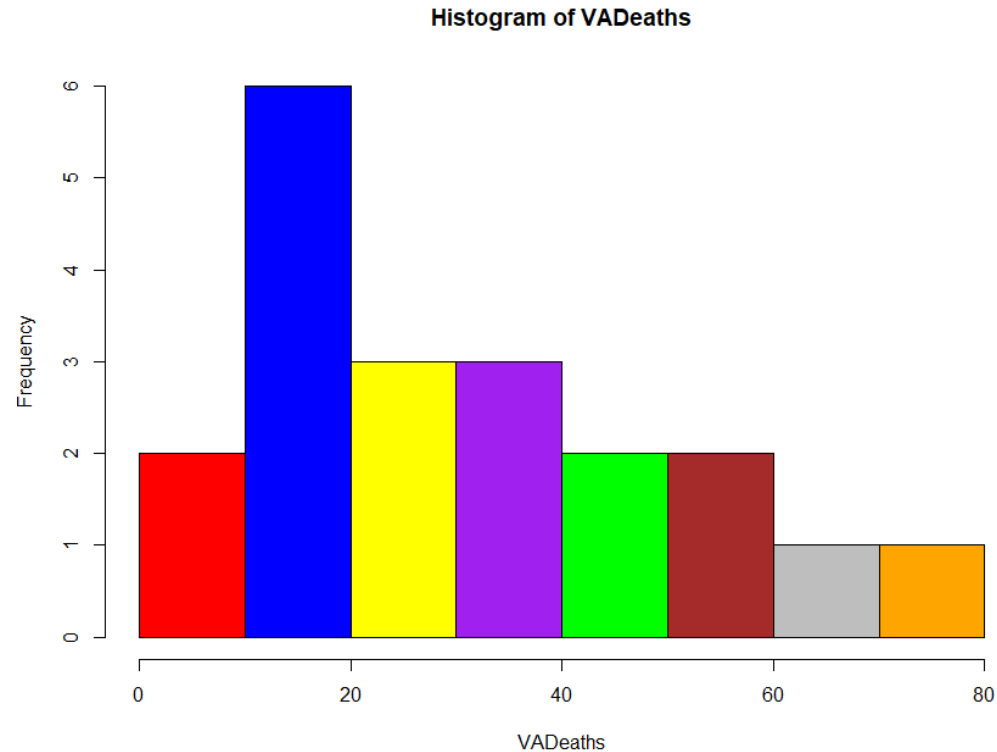
```
6 | 61
```

# Histogram

1. Membuat Garis Horizontal dengan menggunakan skala berdasarkan pada unit pengukuran data
2. Membuat Garis Vertikal dengan menggunakan skala frekuensi
3. Menggambarkan Grafik Batang, tingginya sesuai dengan Frekuensi setiap Kelas Interval
4. Jika terdapat batasan Spesifikasi yang ditentukan oleh Customer (Pelanggan) maka tariklah garis vertikal sesuai dengan spesifikasi tersebut.

# Contoh

```
> hist(VADeaths, col=c("red", "blue", "yellow",  
"purple", "green", "brown", "grey", "orange"))
```



# Ringkasan 5 Angka

- Ringkasan Numerik yang diperlukan adalah Median,  $Q_A$ ,  $Q_B$ , dan Nilai Ekstrim ( $x_{max}$  dan  $x_{min}$ ).
- Susunan Tabel Ringkasan 5 Angka:

Median	
$Q_B$	$Q_A$
$X_{min}$	$X_{max}$

atau

TRIRATA	
$Q_B$	$Q_A$
$X_{min}$	$X_{max}$

# Contoh

Diberikan data Jumlah pengunjung pameran computer selama 6 hari (dalam ratusan) sbb:

Hari Ke-	1	2	3	4	5	6
Jumlah	36	29	21	37	34	45

Diperoleh data terurut:

Urutan	1	2	3	4	5	6
Jumlah	21	29	34	36	37	45

Diperoleh ringkasan 5 angka sbb:

$$Me = \frac{34+36}{2} = 35$$

$$Q_B = \text{data ke } 2 = 29$$

$$Q_A = \text{data ke } 5 = 37$$

Sehingga didapat

<b>35</b>	
<b>29</b>	<b>37</b>
<b>21</b>	<b>45</b>

# Boxplot

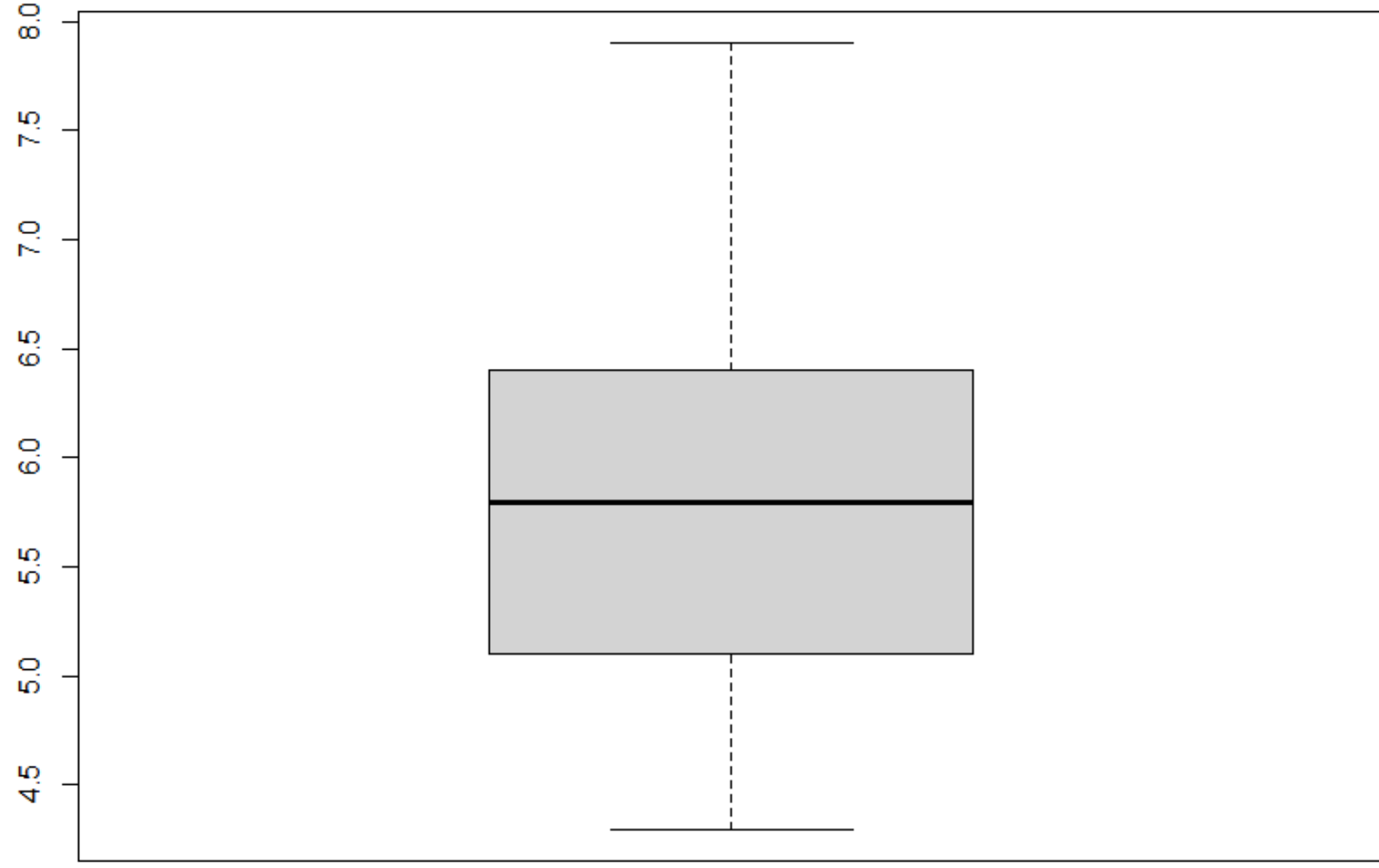
- Data statistik yang dipakai untuk menggambarkan diagram kotak garis yaitu statistik Lima Serangkai, yang dimana terdiri dari data ekstrim (data yang terkecil dan data yang terbesar),  $Q_B$ ,  $Med$  dan  $Q_A$ .
- **Keuntungan Boxplot:**
  1. Sangat baik digunakan untuk Angkatan dengan jumlah observasi besar
  2. Dapat dibuat dalam waktu relative singkat
  3. Gambaran yang jelas dapat dilihat dari data, karena tidak terganggu oleh titik-titik lainnya yang tidak penting



# Contoh

Menggunakan data iris yang ada di Rstudio yang berisikan tentang data mahkota dan kelopak bunga iris. Bunga Iris diklasifikasikan dalam 3 jenis yaitu iris setosa, iris virginica dan iris versicolor.

```
> boxplot(iris$Sepal.Length)
```



**Keterangan:**

IQR = Inner Quartile Range

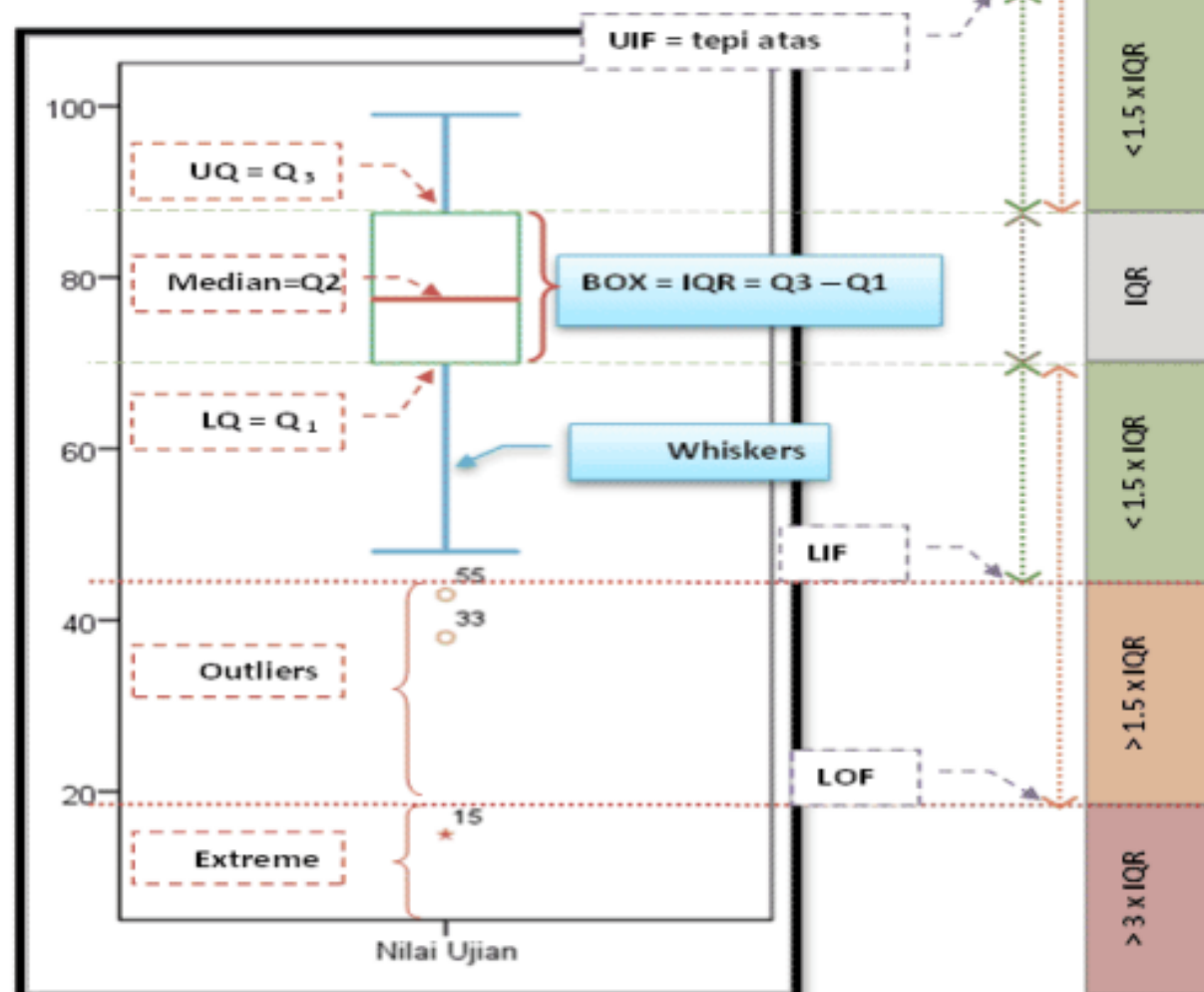
UQ = Upper Quartile

LQ = Lower Quartile

UIF = Upper Inner Fence

LIF = Lower Inner Fence

UOF = Upper Outer Fence



- Garis atas kotak (UQ) = Q3 (Kuartil ketiga) dimana 75% data pengamatan lebih kecil atau sama dengan nilai Q1
- Garis yang merupakan perpanjangan dari box (baik ke arah atas ataupun ke arah bawah) dinamakan dengan whiskers.
- Whiskers bawah menunjukkan nilai yang lebih rendah dari kumpulan data yang berada dalam IQR.
- Whiskers atas menunjukkan nilai yang lebih tinggi dari kumpulan data yang berada dalam IQR.
- Panjang whisker  $\leq 1.5 \times \text{IQR}$ . Masing-masing garis whisker dimulai dari ujung kotak IQR, dan berakhir pada nilai data yang bukan dikategorikan sebagai outlier (*Pada gambar, batasnya adalah garis UIF dan LIF*). Dengan demikian, nilai terbesar dan terkecil dari data pengamatan (tanpa termasuk outlier) masih merupakan bagian dari Boxplot yang terletak tepat di ujung garis tepi whiskers.

- Bagian utama boxplot adalah kotak berbentuk persegi (**Box**) yang merupakan bidang yang menyajikan interquartile range (**IQR**), dimana 50 % dari nilai data pengamatan terletak di sana.
- Panjang kotak sesuai dengan jangkauan kuartil dalam (inner Quartile Range, IQR) yang merupakan selisih antara Kuartil ketiga (Q3) dengan Kuartil pertama (Q1). IQR menggambarkan u... lata. Semakin panjang bidang IQR menunjukkan data semakin menyebar. *Pada Gambar,  $IQR = UQ - LQ = Q3 - Q1$*
- Garis bawah kotak (LQ) = Q1 (Kuartil pertama), dimana 25% data pengamatan lebih kecil atau sama dengan nilai Q1
- Garis tengah kotak = Q2 (median), dimana 50% data pengamatan lebih kecil atau sama dengan nilai ini

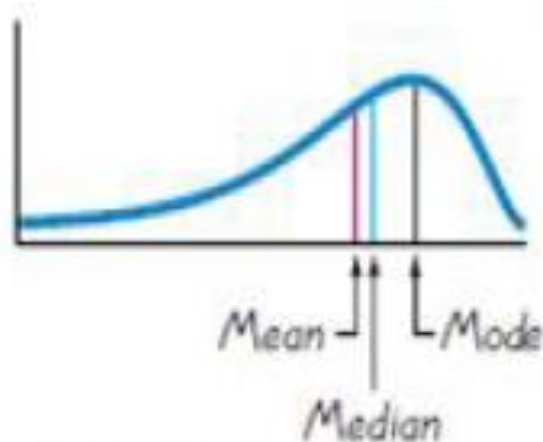
- Nilai yang berada di atas atau dibawah whisker dinamakan nilai outlier atau ekstrim.
- Nilai outlier adalah nilai data yang letaknya lebih dari 1.5 x panjang kotak (IQR), diukur dari UQ (atas kotak) atau LQ (bawah kotak). *Pada Gambar di atas, ada 2 data pengamatan yang merupakan outlier, yaitu data pada case 33 dan case 55 (ada pada baris ke 33 dan baris 35)*
- $Q3 + (1.5 \times IQR) < \text{outlier atas} \leq Q3 + (3 \times IQR)$
- $Q1 - (1.5 \times IQR) > \text{outlier bawah} \geq Q1 - (3 \times IQR)$
- Nilai ekstrim adalah nilai-nilai yang letaknya lebih dari 3 x panjang kotak (IQR), diukur dari UQ (atas kotak) atau LQ (bawah kotak). *Pada gambar di atas, ada 1 data yang merupakan nilai ekstern, yaitu data pada case 15.*
- Ekstrim bagian atas apabila nilainya berada di atas  $Q3 + (3 \times IQR)$  dan
- Ekstrim bagian bawah apabila nilainya lebih rendah dari  $Q1 - (3 \times IQR)$

Boxplots dapat membantu kita dalam memahami karakteristik dari distribusi data. Selain untuk melihat derajat penyebaran data (yang dapat dilihat dari tinggi/panjang boxplot) juga dapat digunakan untuk menilai kesimetrisan sebaran data. Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang whisker menggambarkan tingkat kesimetrisannya.

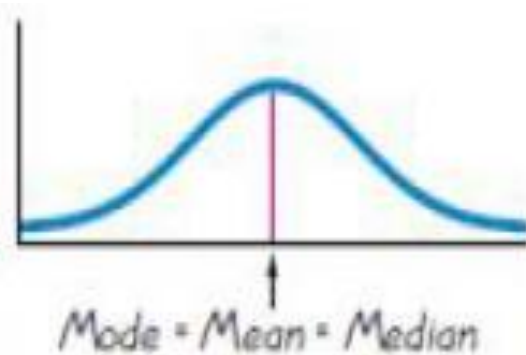
Jika data simetris (berasal dari distribusi normal):

- garis median akan berada di tengah box dan whisker bagian atas dan bawah akan memiliki panjang yang sama serta tidak terdapat nilai outlier ataupun nilai ekstrim.
- diharapkan nilai-nilai pengamatan yang berada di luar whiskers tidak lebih dari 1%.
- Jika data tidak simetris (miring), median tidak akan berada di tengah box dan salah satu dari whisker lebih panjang dari yang lainnya.
- Adanya outlier di bagian atas boxplot yang disertai dengan whisker bagian atas yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kanan (positive skewness).
- Sebaliknya, adanya outlier di bagian bawah boxplot yang disertai dengan whisker bagian bawah yang lebih panjang, menunjukkan bahwa distribusi data cenderung menjulur ke arah kiri (negatif skewness).

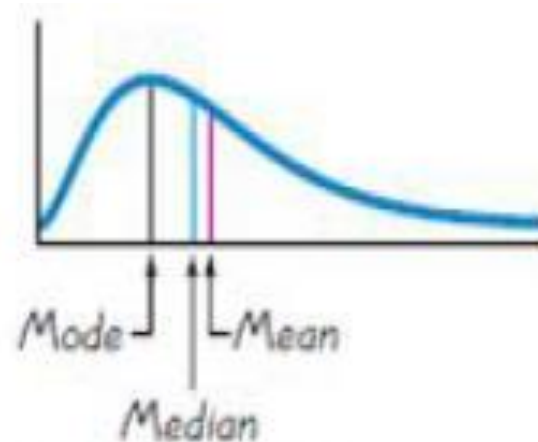




**(a) Skewed to the Left  
(Negatively Skewed):** The mean and median are to the *left* of the mode.



**(b) Symmetric  
(Zero Skewness):** The mean, median, and mode are the same.



**(c) Skewed to the Right  
(Positively Skewed):** The mean and median are to the *right* of the mode.

# Tugas 1

- Carilah dataset dari sumber yang valid (Kaggle.com atau bps.go.id atau satudata.id atau dari sumber lain)
- Buatlah visualisasi datanya menggunakan Rstudio meliputi Boxplot, histogram dan stem & leaf
- Interpretasikan informasi apa saja yang dapat diperoleh dari visualisasi yang kalian lakukan
- Kemudian, lakukan peringkasan numerik menggunakan Rstudio saja
- Masing-masing mahasiswa harus berbeda datasetnya
- Kumpulkan dalam bentuk pdf di elearning paling lambat hari Minggu 12/3/2023