

Dr. Inna Novianty, S.Si, M.Si., M.Si,  
 Gema Parasti Mindara, S.Si, M.Kom,  
 Bayu Widodo, S.T., M.T.  
 2023/1/15 (updated: 2023-01-26)

## Daftar Isi

<b>1</b>	<b>Deskripsi Data</b>	<b>1</b>
1.1	Dataset . . . . .	2
1.2	Ukuran Pemusatan . . . . .	4
1.2.1	Nilai Maksimum & Minimum . . . . .	4
1.2.2	Mean & Median . . . . .	5
1.2.3	Modus . . . . .	5
1.3	Ukuran Penyebaran (Dispersi) . . . . .	6
1.3.1	Quartile Pertama dan Ketiga . . . . .	6
1.3.2	Rentang Interquartile . . . . .	7
1.3.3	Standard Deviasi dan Variance . . . . .	7
1.3.4	Skewness & Kurtosis . . . . .	9
1.4	Paket pastecs . . . . .	10
1.4.1	Koefisien Variasi . . . . .	11
1.5	Korelasi & Koefisien Korelasi . . . . .	12
1.5.1	Matrix Korelasi . . . . .	13
1.6	Distribusi Frekuensi . . . . .	14
1.7	state.x77 . . . . .	18
1.8	Soal Latihan . . . . .	21
	Daftar Pustaka . . . . .	22

## 1 Deskripsi Data

*Statistik Deskriptif merupakan cabang dari ilmu statistik yang bertujuan untuk meringkas, mendeskripsikan, dan menyajikan serangkaian nilai atau kumpulan data. Statistik deskriptif seringkali merupakan langkah pertama dan merupakan bagian penting dalam setiap analisis statistik. Statistik deskriptif memungkinkan untuk memeriksa kualitas data dan membantu untuk "memahami" data secara lengkap, jelas dan mendalam.*

Jika disajikan dengan baik, statistik deskriptif akan menjadi titik awal yang baik untuk analisis data lebih lanjut. Ada banyak ukuran yang mampu mewakili dan menjelaskan kondisi umum atau ringkasan dari suatu dataset. Secara umum ukuran-ukuran untuk meringkas dataset dikelompokkan dalam 2 (dua) jenis, yakni:

1. Ukuran kecenderungan memusat (ukuran pemusatan) atau ukuran lokasi.  
 Ukuran yang menyatakan dimana distriusi data terpusat tepatnya adalah suatu nilai yang digunakan untuk menjelaskan pusat dari sekelompok data yang dapat mewakili data secara keseluruhan. Terdapat tiga ukuran pemusatan data yang sering digunakan, yaitu:

- **Mean** (Rata-rata hitung/rata-rata aritmetika),
  - **Median** (Nilai Tengah), dan
  - **Mode/Modus** (Nilai yang paling sering muncul).
2. Ukuran penyebaran (dispersi)
- Ukuran yang menyatakan seberapa besar nilai-nilai data berbeda atau bervariasi dengan nilai ukuran pusatnya atau seberapa besar penyimpangan nilai-nilai data dengan nilai pusatnya.

Ukuran pemusatan memberikan pemahaman tentang kecenderungan sentral dari data, sedangkan pengukuran dispersi memberikan pemahaman tentang penyebaran data.

Pada bagian ini, kami hanya fokus pada implementasi di R dari statistik deskriptif paling umum dan visualisasinya (bila dianggap tepat).

## 1.1 Dataset

R sudah menyediakan banyak dataset siap pakai (built-in) salah satunya yang populer adalah dataset *iris*. Dataset Iris merupakan dataset multivariate yang diperkenalkan oleh ahli statistika dan biologi Inggris bernama *Ronald Fisher* pada tahun 1936.<sup>1</sup>

R menyediakan kurang lebih 100 dataset yang dapat digunakan oleh user (**built-in-dataset**). Fungsi `data()` digunakan untuk melihat daftar dataset tersebut.

Beberapa diantara paket tersebut masih memerlukan pemanggilan secara eksplisit dengan menggunakan fungsi `data()`. Fungsi `data()` ini memuat satu data frame dengan menspesifikasikan di argumen.

Dataset ini terdiri dari 3 spesies Iris yaitu Iris Setosa, Iris Virginica, dan Iris Versicolor dan tiap spesiesnya memiliki 50 sampel. Dalam data Iris terdapat 4 atribut yang dapat mempengaruhi klasifikasi yaitu, sepal length, sepal width, petal length, dan petal width dalam centimeter yang berbeda-beda.

Untuk melihat seperti apa dataset iris, jalankan R atau Rstudio lalu ketikkan:

```
1 # load dataset iris
2 iris
```

atau

```
1 # load dataset iris dan
2 # menampilkan 10 baris pertama
3 head(iris,10)
```

<sup>1</sup>Untuk menampilkan semua dataset bawaan di R, dapat mengetikkan perintah `data()` di prompt R

Seperti kebanyakan bahasa pemrograman, R memungkinkan penggunaanya mengimport dan menyimpan data tertentu dalam sebuah obyek (variable). Perhatikan contoh berikut:

```
1 # menyimpan dataset iris ke dalam objek atau varaibel dat
2 iris
3 dat <- iris
4 print(dat)
```

Pada contoh di atas dataset iris disimpan ke dalam variabel dat. Anda dapat mengetikkan perintah **dat** atau **print(dat)** untuk melihat isi dari objek (variabel) dat.

Mengenal struktur dataset merupakan hal yang perlu diketahui agar dapat lebih jelas akan diapakan dataset tersebut, atau membantu dalam pemisahan data numerik dan non-numerik. Berikut perintahnya:

```
1 # menampilkan struktur dataset
2 str(iris)

'data.frame':  150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Jika Anda ingin tahu lebih banyak mengenai dataset iris, ketikkan perintah berikut:

```
?iris
```

Anda juga dapat memisahkan atribut numerik dan non numerik pada data iris menggunakan perintah:

```
1 # memisah atribut numerik dengan atribut non numerik
2 # atriut numerik disipan ke dalam variabel iris.num
3 iris.num <- iris[,1:4]
4 str(iris.num)

'data.frame':  150 obs. of  4 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
```

Dataset iris berisi 150 observasi dan 5 variabel yang mewakili panjang dan lebar sepal, panjang dan lebar petal serta spesies dari 150 bunga. Panjang dan lebar sepal dan petal merupakan variabel numerik dan spesies merupakan faktor dengan 3 level (ditunjukkan dengan num dan Faktor w/ 3 level setelah nama variabel).

Pada data yang ada juga dapat diketahui statistik dari data tersebut, misal: mean, median, quantile, variansi, kovariansi, korelasi. Dari dataset ini Anda dapat mengetahui nilai ringkasan dari dataset iris dengan menggunakan perintah *summary()*.

```
1 ## menampilkan min, mean, median, quantile dan frekuensi untuk data kategori
2 summary(iris)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500

Species  
 setosa :50  
 versicolor:50  
 virginica :50

Untuk jenis data numerik, untuk melihat ikhtisar data setiap variabelnya dapat menggunakan perintah `fivenum()`. Berikut contohnya:

```
1 fivenum(iris$Sepal.Length)
[1] 4.3 5.1 5.8 6.4 7.9
```

**Latihan 1.1** Jalankan RStudio dan di R Console atau Code Editor. Ketik dan jalankan perintah berikut. Catat hasilnya dan bila perlu berilah penjelasan singkat dari hasil yang diperoleh.

```
fivenum(iris$Petal.Width)
fivenum(iris$Petal.Length)
```

## 1.2 Ukuran Pemusatan

Salah satu aspek yang paling penting untuk menggambarkan distribusi data adalah nilai pusat data pengamatan (Ukuran Pemusatan, Tendensi Sentral atau Ukuran Lokasi, Ukuran Kecenderungan Memusat). Ukuran Pemusatan Data adalah Nilai yang digunakan untuk menjelaskan pusat dari sekelompok data yang dapat mewakili data secara keseluruhan (Nilai tunggal yang dapat mewakili seluruh data)

Setiap pengukuran aritmatika yang ditujukan untuk menggambarkan suatu nilai yang mewakili nilai pusat atau nilai sentral dari suatu dataset (gugus data, himpunan pengamatan) dikenal sebagai ukuran *tendensi sentral*.

### 1.2.1 Nilai Maksimum & Minimum

Minimum dan maksimum dapat ditemukan menggunakan fungsi `min()` dan `max()` atau menggunakan fungsi `range()`.

```
1 nilaiMin <- min(dat$Sepal.Length)
2 print(paste("Nilai Min = ",nilaiMin))
```

```

3 nilaiMax <- max(dat$Sepal.Length)
4 print(paste("Nilai Max = ",nilaiMax))
5 m <- min(dat$Sepal.Length)
6 print(paste("m = ",m))
7 M <- max(dat$Sepal.Length)
8 print(paste("M = ",M))
9 range(dat$Sepal.Length)
10 rng <- range(dat$Sepal.Length)
11 rng[1]
12 rng[2]
13 max(dat$Sepal.Length) - min(dat$Sepal.Length)
14 rng[2]-rng[1]
15 M - m

```

### 1.2.2 Mean & Median

Mean dapat dihitung dengan menggunakan fungsi `mean()` dan Median dapat dihitung dengan menggunakan fungsi `median()` atau fungsi `quantile()`.

```

1 mean(dat$Sepal.Length)
2 median(dat$Sepal.Length)
3 quantile(dat$Sepal.Length)

```

```

[1] 5.843333
[1] 5.8
 0%  25%  50%  75% 100%
4.3  5.1  5.8  6.4  7.9

```

Dalam mencari atau menghitung nilai mean, jika ada setidaknya satu nilai yang hilang dalam dataset, gunakan fungsi `mean(dat$Sepal.Length, na.rm = TRUE)` untuk menghitung rata-rata dengan NA (*Not Available*) yang dikecualikan. Argumen ini dapat digunakan untuk sebagian besar fungsidi R, tidak hanya untuk rata-rata.

Catatan: secara matematis jika kita memiliki nilai observasi  $x_1, x_2, \dots, x_n$ , maka rata-ratanya (mean) adalah:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

### 1.2.3 Modus

R tidak memiliki fungsi bawaan standar untuk menghitung mode. Fungsi `table()` dapat digunakan untuk mencari nilai mode. Berikut contoh mencari nilai modus untuk variabel `Sepal.Length` pada dataset `iris`.

```

1 iris
2 str(iris)
3 var1<-iris$Sepal.Length

```

```

4 var1
5 sort(var1)
6 table(var1)
7 sort(table(var1))
8 names(table(var1))

1 sort(table(var1))
2 names(table(var1))[table(var1)==max(table(var1))]

var1
4.3 4.5 5.3 7 7.1 7.3 7.4 7.6 7.9 4.7 6.6 4.4 5.9 6.8 7.2 4.6 5.2 6.2 6.9 7.7
  1  1  1  1  1  1  1  1  1  2  2  3  3  3  3  4  4  4  4  4
4.8 6.5 4.9 5.4 5.6 6 6.1 5.5 5.8 6.4 5.7 6.7 5.1 6.3 5
  5  5  6  6  6  6  6  7  7  7  8  8  9  9 10
[1] "5"

```

**Latihan 1.2** Jalankan RStudio dan di R Console atau Code Editor. Ketik dan jalankan perintah berikut. Catat hasilnya dan bila perlu berilah penjelasan singkat dari hasil yang diperoleh.

```

data <- c(2,2,8,2,7,1,2,9,8,2,2,10,9,2)
sum(data)/length(data)
mean(data)
median(data)
sort(data)
table(data)
names(table(data))[table(data)==max(table(data))]

```

### 1.3 Ukuran Penyebaran (Dispersi)

Ukuran penyebaran data adalah suatu ukuran yang menyatakan seberapa besar nilai-nilai data berbeda atau bervariasi dengan nilai ukuran pusatnya atau seberapa besar penyimpangan nilai data dengan nilai pusatnya.

#### 1.3.1 Quartile Pertama dan Ketiga

Sebagai median, kuartil pertama dan ketiga dapat dihitung dengan fungsi `quantile()` dan dengan mengatur argumen kedua menjadi 0,25 atau 0,75 akan diperoleh hasil yang sama.

```

1 quantile(dat$Sepal.Length, 0.25) # first quartile
2 quantile(dat$Sepal.Length, 0.75) # third quartile

25%
5.1

```

75%  
6.4

**Latihan 1.3** Jalankan RStudio dan di R Console atau Code Editor. Ketik dan jalankan perintah berikut. Catat hasilnya dan berilah penjelasan singkat dari hasil yang diperoleh.

```
quantile(dat$Sepal.Length, 0.4)
quantile(dat$Sepal.Length, 0.90)
```

### 1.3.2 Rentang Interquartile

Rentang interkuartil (yaitu perbedaan antara kuartil pertama dan ketiga) dapat dihitung dengan fungsi IQR() atau menggunakan fungsi quantile().

```
1 IQR(dat$Sepal.Length)
2 iqr <- as.vector(quantile(dat$Sepal.Length))
3 iqr[4]-iqr[2]
```

### 1.3.3 Standard Deviasi dan Variance

Standar deviasi dan varian dihitung dengan fungsi sd() dan var().

```
1 sd(dat$Sepal.Length) # standard deviation
2 var(dat$Sepal.Length) # variance

[1] 0.8280661
[1] 0.6856935
```

Perlu diingat bahwa standar deviasi dan varian berbeda untuk sampel atau populasi (lihat perbedaan antara sampel dan populasi). Dalam R, standar deviasi dan varians dihitung seolah-olah data mewakili sampel (sehingga penyebutnya adalah n1, di mana n adalah jumlah pengamatan). R secara default tidak menyediakan fungsi untuk menghitung standar deviasi atau varians suatu populasi.

Catatan: Ukuran varians menggunakan prinsip pencarian jarak antara setiap data dengan rata-ratanya. Secara matematis, hal tersebut dirumuskan:

$$s^2 = \frac{\sum (x_i - \bar{X})^2}{n - 1} \quad (2)$$

Untuk menghitung standar deviasi (atau varians) dari beberapa variabel sekaligus, gunakan fungsi lapply() dengan statistik yang sesuai sebagai argumen kedua.

```
1 lapply(dat[,1:4], sd)
2 lapply(dat[,1:4], var)
```

Perintah dat[, 1:4] memilih variabel 1 sampai 4 karena variabel kelima adalah variabel kualitatif dan standar deviasi tidak dapat dihitung pada jenis variabel tersebut.

Anda dapat menghitung minimum, 1<sup>st</sup> quartile, median, mean, 3<sup>st</sup> quartile dan nilai maksimum untuk seluruh variabel numerik dari suatu dataset sekaligus dengan menggunakan fungsi `summary()`.

```
1 summary(dat)

  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Min.   :4.300  Min.   :2.000  Min.   :1.000  Min.   :0.100
1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st Qu.:0.300
Median :5.800  Median :3.000  Median :4.350  Median :1.300
Mean   :5.843  Mean   :3.057  Mean   :3.758  Mean   :1.199
3rd Qu.:6.400  3rd Qu.:3.300  3rd Qu.:5.100  3rd Qu.:1.800
Max.   :7.900  Max.   :4.400  Max.   :6.900  Max.   :2.500

  Species
setosa   :50
versicolor:50
virginica :50
```

Gunakan fungsi `by()` jika Anda memerlukan statistik deskriptif ini berdasarkan grup. Di mana argumennya adalah nama dataset, variabel pengelompokan (factor level), dan fungsi yang akan diterapkan. Syntax dari fungsi `by()` adalah: `by(X, data$var, FUN)`, dimana parameter X adalah dataset berbentuk data frame, `data$var` adalah factor (list factor) dimana fungsi akan diterapkan, FUN adalah fungsi yang digunakan, dapat berupa fungsi dasar matematika, statistika maupun user defined function. Pada contoh diatas, `dat$var` adalah factor level dari variabel species dan FUN berupa fungsi `summary`. Berikut contohnya:

```
1 by(dat, dat$Species, summary)

dat$Species: setosa
  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Min.   :4.300  Min.   :2.300  Min.   :1.000  Min.   :0.100
1st Qu.:4.800  1st Qu.:3.200  1st Qu.:1.400  1st Qu.:0.200
Median :5.000  Median :3.400  Median :1.500  Median :0.200
Mean   :5.006  Mean   :3.428  Mean   :1.462  Mean   :0.246
3rd Qu.:5.200  3rd Qu.:3.675  3rd Qu.:1.575  3rd Qu.:0.300
Max.   :5.800  Max.   :4.400  Max.   :1.900  Max.   :0.600

  Species
setosa   :50
versicolor: 0
virginica : 0
```

```
-----
dat$Species: versicolor
```



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min. :4.900	Min. :2.000	Min. :3.00	Min. :1.000	setosa : 0
1st Qu.:5.600	1st Qu.:2.525	1st Qu.:4.00	1st Qu.:1.200	versicolor:50
Median :5.900	Median :2.800	Median :4.35	Median :1.300	virginica : 0
Mean :5.936	Mean :2.770	Mean :4.26	Mean :1.326	
3rd Qu.:6.300	3rd Qu.:3.000	3rd Qu.:4.60	3rd Qu.:1.500	
Max. :7.000	Max. :3.400	Max. :5.10	Max. :1.800	

---

```

dat$Species: virginica
  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
Min.   :4.900   Min.   :2.200   Min.   :4.500   Min.   :1.400
1st Qu.:6.225   1st Qu.:2.800   1st Qu.:5.100   1st Qu.:1.800
Median :6.500   Median :3.000   Median :5.550   Median :2.000
Mean   :6.588   Mean   :2.974   Mean   :5.552   Mean   :2.026
3rd Qu.:6.900   3rd Qu.:3.175   3rd Qu.:5.875   3rd Qu.:2.300
Max.   :7.900   Max.   :3.800   Max.   :6.900   Max.   :2.500

  Species
setosa    : 0
versicolor: 0
virginica :50

```

### 1.3.4 Skewness & Kurtosis

Ukuran kemiringan (Koefisien Kemiringan) atau skewness merupakan suatu nilai yang mengukur ketidaksimetrisan distribusi data. Suatu data dikatakan berdistribusi simetris sempurna bila nilai rata-rata, median, dan modus dalam data adalah sama. (Distribusi normal merupakan distribusi yang simetris dan nilai skewness adalah 0).

1. Skewness yang bernilai positif menunjukkan ujung dari kecondongan menjulur ke arah nilai positif (ekor kurva sebelah kanan lebih panjang, nilai kemiringan  $> 0$ ).
2. Skewness yang bernilai negatif menunjukkan ujung dari kecondongan menjulur ke arah nilai negatif (ekor kurva sebelah kiri lebih panjang, nilai kemiringan  $< 0$ ).
3. Jika Skewness sama dengan nol ( $= 0$ ), model distribusinya simetris

Kurtosis menggambarkan keruncingan (peakedness atau kerataan flatness) suatu distribusi data dibandingkan dengan distribusi normal.

1. Pada distribusi normal, nilai kurtosis sama dengan 0.
2. Nilai kurtosis yang positif menunjukkan distribusi yang relatif runcing, sedangkan nilai kurtosis yang negatif menunjukkan distribusi yang relatif rata.

Skewness dan kurtosis merupakan dua alat ukur dalam menelusuri distribusi data yang diperbandingkan dengan distribusi normal.

**Latihan 1.4** Jalankan RStudio dan di R Console atau Code Editor. Hitung statistik data berikut dan berikan interpretasi terhadap output yang dihasilkan.

Diberikan dua buah vector X dan vector Y yang masing-masing mewakili variabel Luas (Ukuran) Rumah dalam feet<sup>2</sup> dan Konsumsi Listrik perbulan dalam Kwh.

```
X <- c(1290, 1350, 1470, 1600, 1710, 1840, 1980, 2230, 2400, 2930)
```

```
Y <- c(1182, 1172, 1264, 1493, 1671, 1711, 1804, 1840, 1956, 1954)
```

Sumber: Irawan and Astuti (2006)

## 1.4 Paket pastecs

Paket (library)<sup>2</sup> pastecs digunakan untuk mendapatkan deskriptif statik dari dataset yang lebih lengkap. Library pastecs perlu di instal terlebih dahulu dengan perintah:

```
# install pacakage
install.packages("pastecs")
```

```
1 # mengaktifkan library
2 library(pastecs)
3 # menjalankan fungsi stat.desc()
4 stat.desc(dat[,1:4])
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
nbr.val	150.00000000	150.00000000	150.00000000	150.00000000
nbr.null	0.00000000	0.00000000	0.00000000	0.00000000
nbr.na	0.00000000	0.00000000	0.00000000	0.00000000
min	4.30000000	2.00000000	1.00000000	0.10000000
max	7.90000000	4.40000000	6.90000000	2.50000000
range	3.60000000	2.40000000	5.90000000	2.40000000
sum	876.50000000	458.60000000	563.70000000	179.90000000
median	5.80000000	3.00000000	4.35000000	1.30000000
mean	5.84333333	3.05733333	3.75800000	1.19933333
SE.mean	0.06761132	0.03558833	0.1441360	0.06223645
CI.mean.0.95	0.13360085	0.07032302	0.2848146	0.12298004
var	0.68569351	0.18997942	3.1162779	0.58100626
std.dev	0.82806613	0.43586628	1.7652982	0.76223767
coef.var	0.14171126	0.14256420	0.4697441	0.63555114

Anda dapat memiliki lebih banyak statistik (seperti uji kemiringan, kurtosis, dan normalitas) dengan menambahkan argumen `norm = TRUE` di fungsi sebelumnya. Perhatikan bahwa variabel Spesies bukan numerik, sehingga statistik deskriptif tidak dapat dihitung untuk variabel ini dan NA ditampilkan.

<sup>2</sup>library yang sudah di-install bisa diaktifkan dengan menggunakan perintah `library(nama_packages)` tanpa menggunakan tanda `" "` atau `' '`

```
1 options(digits = 3)
2 stat.desc(dat, norm=TRUE)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
nbr.val	150.0000	150.0000	1.50e+02	1.50e+02	NA
nbr.null	0.0000	0.0000	0.00e+00	0.00e+00	NA
nbr.na	0.0000	0.0000	0.00e+00	0.00e+00	NA
min	4.3000	2.0000	1.00e+00	1.00e-01	NA
max	7.9000	4.4000	6.90e+00	2.50e+00	NA
range	3.6000	2.4000	5.90e+00	2.40e+00	NA
sum	876.5000	458.6000	5.64e+02	1.80e+02	NA
median	5.8000	3.0000	4.35e+00	1.30e+00	NA
mean	5.8433	3.0573	3.76e+00	1.20e+00	NA
SE.mean	0.0676	0.0356	1.44e-01	6.22e-02	NA
CI.mean.0.95	0.1336	0.0703	2.85e-01	1.23e-01	NA
var	0.6857	0.1900	3.12e+00	5.81e-01	NA
std.dev	0.8281	0.4359	1.77e+00	7.62e-01	NA
coef.var	0.1417	0.1426	4.70e-01	6.36e-01	NA
skewness	0.3086	0.3126	-2.69e-01	-1.01e-01	NA
skew.2SE	0.7792	0.7893	-6.80e-01	-2.55e-01	NA
kurtosis	-0.6058	0.1387	-1.42e+00	-1.36e+00	NA
kurt.2SE	-0.7696	0.1762	-1.80e+00	-1.73e+00	NA
normtest.W	0.9761	0.9849	8.76e-01	9.02e-01	NA
normtest.p	0.0102	0.1012	7.41e-10	1.68e-08	NA

### 1.4.1 Koefisien Variasi

Koefisien variasi (coefficient of variation) merupakan perbandingan (rasio) antara standar deviasi dengan nilai rata-rata. Koefisien variasi biasa dinyatakan dengan persentase. Kegunaan koefisien variasi adalah untuk melihat sebaran/distribusi data dari rata-rata hitungannya.

Koefisien variasi merupakan statistik yang berguna untuk membandingkan variasi data dari satu dataset dengan dataset lainnya apabila satuan pengukuran dari variabel-variabel yang diukur berbeda satu sama lain (misalnya tinggi badan dalam cm dan berat badan dalam kg).

Koefisien variasi dapat diperoleh dengan menggunakan fungsi `stat.desc()` (lihat baris `coef.var` pada tabel di atas) atau dengan menghitung secara manual. Berikut langkah menghitungnya:<sup>3</sup>

```
1 options(digits = 3)
2 sd(dat$Sepal.Length) / mean(dat$Sepal.Length)

[1] 0.142
```

<sup>3</sup>Koefisien variasi adalah simpangan baku dibagi rata-rata dan hasilnya dikalikan 100%

Analisis statistik deskriptif merupakan cara atau alat analisis data statistik yang paling sederhana. Statistik deskriptif akan menghasilkan informasi dan insight yang berguna dalam membantu menjawab permasalahan. Statistik deskriptif adalah cara atau metode untuk mengumpulkan, meringkas dan menyajikan data dalam bentuk yang mudah dibaca.

Beberapa statistik sederhana menunjukkan ukuran kecenderungan memusat seperti mean, median kuartil dan ukuran penyebaran seperti standar deviasi, rentang interkuartil, rentang.

## 1.5 Korelasi & Koefisien Korelasi

Korelasi merupakan istilah yang biasa digunakan untuk menggambarkan ada tidaknya hubungan suatu hal dengan hal lain (variabel yang satu dengan variabel yang lain) untuk menyatakan derajat hubungan linier antar dua variabel.

Sedangkan Koefisien Korelasi merupakan angka yang menunjukkan arah dan kuatnya hubungan antar variabel dan **hubungan linier** antara 2 variabel.<sup>4</sup>

Koefisien Korelasi dinyatakan dengan simbol “  $r$  ” yang dapat mempunyai nilai antara -1 s.d +1. Dikatakan hubungan -1 jika memiliki hubungan yang kuat negatif, sebaliknya jika nilai  $r=1$  berarti mempunyai hubungan kuat positif. Jika nilai  $r$  semakin mendekati 0 (nol) artinya hubungan variabel tersebut semakin lemah. Berikut *Rule of thumb* penentuan derajat kuat tidaknya hubungan antara 2 variabel:

- Nilai  $r$  yang mendekati -1 atau +1 menunjukkan hubungan kuat antara dua variabel.
- Nilai  $r$  yang mendekati 0 menunjukkan lemahnya hubungan antara dua variabel.

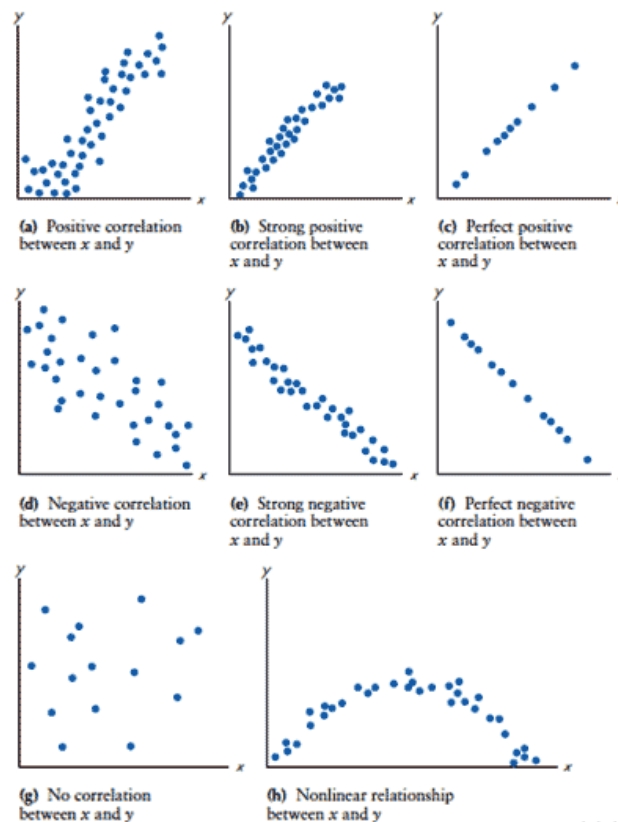
Untuk menghitung koefisien korelasi dapat menggunakan fungsi `cor()`. Perhatikan contoh menghitung koefisien korelasi berikut:

```
1 names(dat)
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
1 cor(dat$Sepal.Length, dat$Petal.Width)
[1] 0.818
```

Pada contoh diatas korelasi dihitung secara default dengan metode Korelasi Pearson menggunakan fungsi `cor()`. Jika Anda ingin menghitung korelasi Spearman, tambahkan metode argumen = “spearman” ke fungsi `cor()`.

```
1 # Spearman correlation antar 2 variables
2 cor(dat$Sepal.Length, dat$Petal.Width, method = "spearman")
[1] 0.834
```

<sup>4</sup>Jika ada hubungan nonlinier, maka koefisien korelasi bernilai 0



Gambar 1: Jenis Korelasi antara 2 variabel

### 1.5.1 Matrix Korelasi

Misalkan sekarang Anda ingin menghitung korelasi untuk beberapa pasang variabel. Anda dapat dengan mudah melakukannya untuk semua kemungkinan pasangan variabel dalam kumpulan data. Berikut langkahnya:

```
1 round(cor(iris.num), digits=2)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.00	-0.12	0.87	0.82
Sepal.Width	-0.12	1.00	-0.43	-0.37
Petal.Length	0.87	-0.43	1.00	0.96
Petal.Width	0.82	-0.37	0.96	1.00

Cara lain menghitung Matrix Korelasi adalah dengan menggunakan **library corrplot**. Untuk menggunakannya, Anda perlu menginstall library corrplot terlebih dahulu. Berikut langkahnya:

```
install.packages("corrplot")
```

```
1 library(corrplot)
```

```
corrplot 0.92 loaded
```

```
1 corrplot(cor(iris.num),method="number",type="upper")
```



## 1.6 Distribusi Frekuensi

Distribusi frekuensi adalah suatu daftar atau tabel yang membagi data dalam beberapa kelas. Distribusi frekuensi terdiri dari 2 macam, yaitu distribusi *frekuensi kategorikal* dan *distribusi frekuensi numerical*.

Tabel distribusi frekuensi dibuat untuk menyederhanakan bentuk dan jumlah data sehingga ketika disajikan akan dapat dengan mudah dipahami atau dinilai. Fungsi `table()` dapat digunakan untuk membuat tabel distribusi frekuensi. Hal-hal yang perlu dipahami tentang distribusi frekuensi antara lain:

1. Distribusi Frekuensi merupakan “cara” untuk menyajikan data dalam bentuk yang ringkas.
2. Distribusi Frekuensi berarti mengelompokkan “data mentah” ke dalam kelas-kelas kategori tertentu atau interval tertentu yang disertai dengan nilai frekuensi yang sesuai. Sehingga membentuk suatu tabel frekuensi yang berisikan kategori-kategori tersebut.
3. Daftar frekuensi ini akan memberikan gambaran yang khas tentang bagaimana keragaman data.

Bagaimana cara membuat Distribusi Frekuensi?

1. Gunakan fungsi `cut()` untuk membuat suatu interval dengan menggunakan argumen `break` yang digunakan untuk menentukan batas-batas interval.

2. Syntax fungsi cut(): cut(x,breaks,labels=), dimana:

- x: vector
- breaks: vector dari titik potong cutpoint
- labels: label dari setiap kategori

3. Gunakan fungsi table() untuk melihat distribusi frekuensinya.

Untuk memudahkan pemahaman, kita gunakan data dummy hasil ujian Basis Data berikut:

Scores on Test #2 - Males						
42 Scores: Average = 73.5						
84	88	76	44	80	83	51
93	69	78	49	55	78	93
64	84	54	92	96	72	97
37	97	67	83	93	95	67
72	67	86	76	80	58	62
69	64	82	48	54	80	69

Gambar 2: Hasil Ujian Basis Data

Data mentah pada Gambar 2 di atas kemudian dikelompokkan menjadi beberapa kelas (kelompok) dengan rentang nilai: 30 - 39; 40 - 49; . . .; 90 - 99.

Pertama-tama kita pahami perilaku fungsi cut() secara default. perjakan langkah berikut:

1. Memasukkan data mentah ke dalam variabel, misalnya nama variabelnya adalah: nilai.

```
1 baris1<-c(84,88,76,44,80,83,51)
2 baris2<-c(93,69,78,49,55,78,93)
3 baris3<-c(64,84,54,92,96,72,97)
4 baris4<-c(37,97,67,83,93,85,67)
5 baris5<-c(72,67,86,76,80,58,62)
6 baris6<-c(69,64,82,48,54,80,69)
7 nilai<-c(baris1,baris2,baris3,baris4,baris5,baris6)
8 sort(nilai)

[1] 37 44 48 49 51 54 54 55 58 62 64 64 67 67 67 69 69 69 72 72 76 76 78 78 80
[26] 80 80 82 83 83 84 84 85 86 88 92 93 93 93 96 97 97
```

```
1 length(nilai)
```

```
[1] 42
```

2. Mengelompokkan data mentah dalam 7 kelompok (kategori) menggunakan fungsi cut.

```
1 cut(nilai,breaks=7)

[1] (79.9,88.4] (79.9,88.4] (71.3,79.9] (36.9,45.6] (79.9,88.4] (79.9,88.4]
[7] (45.6,54.1] (88.4,97.1] (62.7,71.3] (71.3,79.9] (45.6,54.1] (54.1,62.7]
```

```
[13] (71.3,79.9] (88.4,97.1] (62.7,71.3] (79.9,88.4] (45.6,54.1] (88.4,97.1]
[19] (88.4,97.1] (71.3,79.9] (88.4,97.1] (36.9,45.6] (88.4,97.1] (62.7,71.3]
[25] (79.9,88.4] (88.4,97.1] (79.9,88.4] (62.7,71.3] (71.3,79.9] (62.7,71.3]
[31] (79.9,88.4] (71.3,79.9] (79.9,88.4] (54.1,62.7] (54.1,62.7] (62.7,71.3]
[37] (62.7,71.3] (79.9,88.4] (45.6,54.1] (45.6,54.1] (79.9,88.4] (62.7,71.3]
7 Levels: (36.9,45.6] (45.6,54.1] (54.1,62.7] (62.7,71.3] ... (88.4,97.1]
```

### 3. Melihat distribusi frekuensi dengan menggunakan fungsi table()

```
1 table(cut(nilai,breaks=7))
      (36.9,45.6] (45.6,54.1] (54.1,62.7] (62.7,71.3] (71.3,79.9] (79.9,88.4]
              2           5           3           8           6          11
      (88.4,97.1]
              7

1 cbind(table(cut(nilai,breaks=7)))
      [,1]
(36.9,45.6] 2
(45.6,54.1] 5
(54.1,62.7] 3
(62.7,71.3] 8
(71.3,79.9] 6
(79.9,88.4] 11
(88.4,97.1] 7
```

Terlihat bahwa dengan hanya menggunakan fungsi `cut()` dan `table()`, kita sudah mendapatkan tabel distribusi frekuensi dengan sangat mudah. Fungsi `cut()` membagi suatu variabel numeric menjadi interval kelas, jumlah interval kelas diberikan dalam argumen “`breaks=...`”. R akan menentukan batas bawah dan batas atas setiap kelas untuk ke sepuluh kelas. Notasi  $(x,y]$  berarti interval kelas dari  $x$  ke  $y$ , dengan  $x$  tidak termasuk tapi  $y$  termasuk.

Bagaimana cara membuat kategori sesuai dengan kasus di atas, yakni dengan mengelompokkan ke dalam kelas: 30 - 39; 40 - 49; . . . ; 90 - 99 ? Tiap-tiap kelompok ini disebut *interval kelas* dan angka 30, 40, . . . , 90 disebut sebagai batas bawah kelas dan angka 39, 49, . . . , 99 disebut sebagai batas atas kelas. Kelas didefinisikan dengan membuat daftar batas kelas.

Tepi kelas atas dan tepi kelas bawah dicari dengan rumus:

1. Tepi Atas Kelas = batas atas kelas + 0.5
2. Tepi Bawah Kelas = batas bawah kelas - 0.5

Sehingga dari rumus di atas diperoleh angka 29.5 untuk tepi bawah kelas ke-1 dan 99.5 untuk tepi atas kelas ke-7. Kemudian, untuk membagi kelas interval sesuai keinginan, dapat dilakukan dengan menetapkan titik-titik potong (break) dalam sebuah vector. Berikut langkahnya:



```
1 cut(nilai,breaks=seq(29.5,99.5,10))
```

[1]	(79.5,89.5]	(79.5,89.5]	(69.5,79.5]	(39.5,49.5]	(79.5,89.5]	(79.5,89.5]
[7]	(49.5,59.5]	(89.5,99.5]	(59.5,69.5]	(69.5,79.5]	(39.5,49.5]	(49.5,59.5]
[13]	(69.5,79.5]	(89.5,99.5]	(59.5,69.5]	(79.5,89.5]	(49.5,59.5]	(89.5,99.5]
[19]	(89.5,99.5]	(69.5,79.5]	(89.5,99.5]	(29.5,39.5]	(89.5,99.5]	(59.5,69.5]
[25]	(79.5,89.5]	(89.5,99.5]	(79.5,89.5]	(59.5,69.5]	(69.5,79.5]	(59.5,69.5]
[31]	(79.5,89.5]	(69.5,79.5]	(79.5,89.5]	(49.5,59.5]	(59.5,69.5]	(59.5,69.5]
[37]	(59.5,69.5]	(79.5,89.5]	(39.5,49.5]	(49.5,59.5]	(79.5,89.5]	(59.5,69.5]

7 Levels: (29.5,39.5] (39.5,49.5] (49.5,59.5] (59.5,69.5] ... (89.5,99.5]

```
1 table(cut(nilai,breaks=seq(29.5,99.5,10)))
```

(29.5,39.5]	(39.5,49.5]	(49.5,59.5]	(59.5,69.5]	(69.5,79.5]	(79.5,89.5]
1	3	5	9	6	11
(89.5,99.5]					
7					

```
1 cbind(table(cut(nilai,breaks=seq(29.5,99.5,10))))
```

	[,1]
(29.5,39.5]	1
(39.5,49.5]	3
(49.5,59.5]	5
(59.5,69.5]	9
(69.5,79.5]	6
(79.5,89.5]	11
(89.5,99.5]	7

```
1 # dapat juga menggunakan fungsi transform()
2 transform(table(cut(nilai,breaks=seq(29.5,99.5,10))))
```

	Var1	Freq
1	(29.5,39.5]	1
2	(39.5,49.5]	3
3	(49.5,59.5]	5
4	(59.5,69.5]	9
5	(69.5,79.5]	6
6	(79.5,89.5]	11
7	(89.5,99.5]	7

Anda dapat juga menggunakan langkah berikut:

```
1 # memasukkan nilai cut ke dalam variabel nilaiDF
2 nilaiDF<-cut(nilai,breaks=seq(29.5,99.5,10))
3 table(nilaiDF)
```

	Var1	Freq
1	(29.5,39.5]	1
2	(39.5,49.5]	3
3	(49.5,59.5]	5
4	(59.5,69.5]	9
5	(69.5,79.5]	6
6	(79.5,89.5]	11
7	(89.5,99.5]	7

7

```
1 transform(table(nilaiDF))
```

```
      nilaiDF Freq
1 (29.5,39.5]    1
2 (39.5,49.5]    3
3 (49.5,59.5]    5
4 (59.5,69.5]    9
5 (69.5,79.5]    6
6 (79.5,89.5]   11
7 (89.5,99.5]    7
```

Untuk melihat bentuk distribusi data dalam grafik, dapat digunakan diagram dahandaun (stem and leaf diagram) dapat dilakukan dengan fungsi `stem()` sebagai berikut:

```
1 stem(nilai)
```

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

```
3 | 7
4 | 489
5 | 14458
6 | 244777999
7 | 226688
8 | 00023344568
9 | 2333677
```

## 1.7 state.x77

```
1 state.x77
2 class(state.x77)
3 dataState <- data.frame(state.x77)
4 names(dataState)
5 colnames(dataState)
6 rownames(dataState)
7 dim(dataState)
```

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost
Alabama	3615	3624	2.1	69.0	15.1	41.3	20
Alaska	365	6315	1.5	69.3	11.3	66.7	152
Arizona	2212	4530	1.8	70.5	7.8	58.1	15
Arkansas	2110	3378	1.9	70.7	10.1	39.9	65
California	21198	5114	1.1	71.7	10.3	62.6	20
Colorado	2541	4884	0.7	72.1	6.8	63.9	166
Connecticut	3100	5348	1.1	72.5	3.1	56.0	139
Delaware	579	4809	0.9	70.1	6.2	54.6	103
Florida	8277	4815	1.3	70.7	10.7	52.6	11
Georgia	4931	4091	2.0	68.5	13.9	40.6	60

## Teknik Analisis Data

(Deskripsi Data)

Hawaii	868	4963	1.9	73.6	6.2	61.9	0
Idaho	813	4119	0.6	71.9	5.3	59.5	126
Illinois	11197	5107	0.9	70.1	10.3	52.6	127
Indiana	5313	4458	0.7	70.9	7.1	52.9	122
Iowa	2861	4628	0.5	72.6	2.3	59.0	140
Kansas	2280	4669	0.6	72.6	4.5	59.9	114
Kentucky	3387	3712	1.6	70.1	10.6	38.5	95
Louisiana	3806	3545	2.8	68.8	13.2	42.2	12
Maine	1058	3694	0.7	70.4	2.7	54.7	161
Maryland	4122	5299	0.9	70.2	8.5	52.3	101
Massachusetts	5814	4755	1.1	71.8	3.3	58.5	103
Michigan	9111	4751	0.9	70.6	11.1	52.8	125
Minnesota	3921	4675	0.6	73.0	2.3	57.6	160
Mississippi	2341	3098	2.4	68.1	12.5	41.0	50
Missouri	4767	4254	0.8	70.7	9.3	48.8	108
Montana	746	4347	0.6	70.6	5.0	59.2	155
Nebraska	1544	4508	0.6	72.6	2.9	59.3	139
Nevada	590	5149	0.5	69.0	11.5	65.2	188
New Hampshire	812	4281	0.7	71.2	3.3	57.6	174
New Jersey	7333	5237	1.1	70.9	5.2	52.5	115
New Mexico	1144	3601	2.2	70.3	9.7	55.2	120
New York	18076	4903	1.4	70.5	10.9	52.7	82
North Carolina	5441	3875	1.8	69.2	11.1	38.5	80
North Dakota	637	5087	0.8	72.8	1.4	50.3	186
Ohio	10735	4561	0.8	70.8	7.4	53.2	124
Oklahoma	2715	3983	1.1	71.4	6.4	51.6	82
Oregon	2284	4660	0.6	72.1	4.2	60.0	44
Pennsylvania	11860	4449	1.0	70.4	6.1	50.2	126
Rhode Island	931	4558	1.3	71.9	2.4	46.4	127
South Carolina	2816	3635	2.3	68.0	11.6	37.8	65
South Dakota	681	4167	0.5	72.1	1.7	53.3	172
Tennessee	4173	3821	1.7	70.1	11.0	41.8	70
Texas	12237	4188	2.2	70.9	12.2	47.4	35
Utah	1203	4022	0.6	72.9	4.5	67.3	137
Vermont	472	3907	0.6	71.6	5.5	57.1	168
Virginia	4981	4701	1.4	70.1	9.5	47.8	85
Washington	3559	4864	0.6	71.7	4.3	63.5	32
West Virginia	1799	3617	1.4	69.5	6.7	41.6	100
Wisconsin	4589	4468	0.7	72.5	3.0	54.5	149
Wyoming	376	4566	0.6	70.3	6.9	62.9	173

### Area

Alabama	50708
Alaska	566432
Arizona	113417
Arkansas	51945

---

California	156361
Colorado	103766
Connecticut	4862
Delaware	1982
Florida	54090
Georgia	58073
Hawaii	6425
Idaho	82677
Illinois	55748
Indiana	36097
Iowa	55941
Kansas	81787
Kentucky	39650
Louisiana	44930
Maine	30920
Maryland	9891
Massachusetts	7826
Michigan	56817
Minnesota	79289
Mississippi	47296
Missouri	68995
Montana	145587
Nebraska	76483
Nevada	109889
New Hampshire	9027
New Jersey	7521
New Mexico	121412
New York	47831
North Carolina	48798
North Dakota	69273
Ohio	40975
Oklahoma	68782
Oregon	96184
Pennsylvania	44966
Rhode Island	1049
South Carolina	30225
South Dakota	75955
Tennessee	41328
Texas	262134
Utah	82096
Vermont	9267
Virginia	39780
Washington	66570
West Virginia	24070
Wisconsin	54464

```

Wyoming          97203
[1] "matrix" "array"
[1] "Population" "Income"      "Illiteracy" "Life.Exp"    "Murder"
[6] "HS.Grad"    "Frost"      "Area"
[1] "Population" "Income"      "Illiteracy" "Life.Exp"    "Murder"
[6] "HS.Grad"    "Frost"      "Area"
[1] "Alabama"      "Alaska"      "Arizona"      "Arkansas"
[5] "California"   "Colorado"    "Connecticut"  "Delaware"
[9] "Florida"     "Georgia"     "Hawaii"       "Idaho"
[13] "Illinois"    "Indiana"     "Iowa"         "Kansas"
[17] "Kentucky"    "Louisiana"   "Maine"        "Maryland"
[21] "Massachusetts" "Michigan"    "Minnesota"    "Mississippi"
[25] "Missouri"    "Montana"     "Nebraska"     "Nevada"
[29] "New Hampshire" "New Jersey" "New Mexico"   "New York"
[33] "North Carolina" "North Dakota" "Ohio"         "Oklahoma"
[37] "Oregon"      "Pennsylvania" "Rhode Island" "South Carolina"
[41] "South Dakota" "Tennessee"   "Texas"        "Utah"
[45] "Vermont"     "Virginia"    "Washington"   "West Virginia"
[49] "Wisconsin"   "Wyoming"
[1] 50 8

```

## 1.8 Soal Latihan

1. Hitung statistik deskriptif dari dataset iris dan interpretasikan outputnya untuk variabel:
  1. Petal.Length
  2. Petal.Width
2. Hitung nilai ukuran dispersi untuk variabel Petal.Length pada dataset iris
3. Hitung kurtosis dan skewness dari variabel Petal.Width pada dataset iris.
4. Hitung matrix korelasi pada data set iris, interpretasikan outputnya dan visualisasikan hasilnya.
5. Berikut data tentang jumlah pengunjung XXI hari ke-20 s.d hari ke-49. Buat tabel distribusi frekuensinya dan bagi kelas interval sesuai dengan yang Anda inginkan.
6. Jalankan R atau RStudio kemudian ketikkan skrip berikut:

```

pairs(~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width,
data = iris,
lower.panel = NULL
main="Simple Scatter Plot Matrix")

```

Interpretasikan hasilnya!

7. Buat diagram dahan daun (*Steam and Leaf Plot*) untuk data di atas.

	No	Hari	Jml Pengunjung		No	Hari	Jml Pengunjung	
<input type="checkbox"/>	1	20	67		16	35	77	<input type="checkbox"/>
<input type="checkbox"/>	2	21	87		17	36	86	<input type="checkbox"/>
<input type="checkbox"/>	3	22	67		18	37	45	<input type="checkbox"/>
<input type="checkbox"/>	4	23	54		19	38	67	<input type="checkbox"/>
<input type="checkbox"/>	5	24	67		20	39	77	<input type="checkbox"/>
<input type="checkbox"/>	6	25	86		21	40	99	<input type="checkbox"/>
<input type="checkbox"/>	7	26	76		22	41	60	<input type="checkbox"/>
<input type="checkbox"/>	8	27	100		23	42	87	<input type="checkbox"/>
<input type="checkbox"/>	9	28	56		24	43	87	<input type="checkbox"/>
<input type="checkbox"/>	10	29	45		25	44	112	<input type="checkbox"/>
<input type="checkbox"/>	11	30	78		26	45	45	<input type="checkbox"/>
<input type="checkbox"/>	12	31	44		27	46	56	<input type="checkbox"/>
<input type="checkbox"/>	13	32	90		28	47	78	<input type="checkbox"/>
<input type="checkbox"/>	14	33	89		29	48	65	<input type="checkbox"/>
<input type="checkbox"/>	15	34	78		30	49	67	<input type="checkbox"/>

Gambar 3: Banyaknya Pengunjung XXI Bulan Juni 2023

### Daftar Pustaka

10 Irawan, Nur, and Septin Puji Astuti. 2006. *Mengolah Data Statistik Dengan Mudah Menggunakan Minitab 14*. Yogyakarta: Penerbit Andi.