



[www.esaunggul.ac.id](http://www.esaunggul.ac.id)

**EKSPLORASI DATA & DATA WAREHOUSE**  
**PERTEMUAN - 2**  
**NOVIANDI**  
**PRODI MIK | FAKULTAS ILMU-ILMU KESEHATAN**

# KEMAMPUAN AKHIR YANG DIHARAPKAN

Mahasiswa mampu mengeksplorasi data, mengetahui adanya *missing* data values, dan data noisy serta dapat mengatasinya

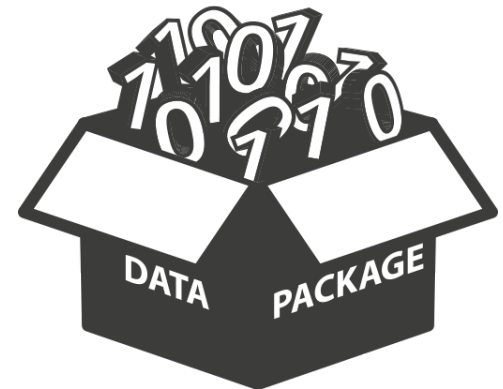


# EKSPLORASI DATA

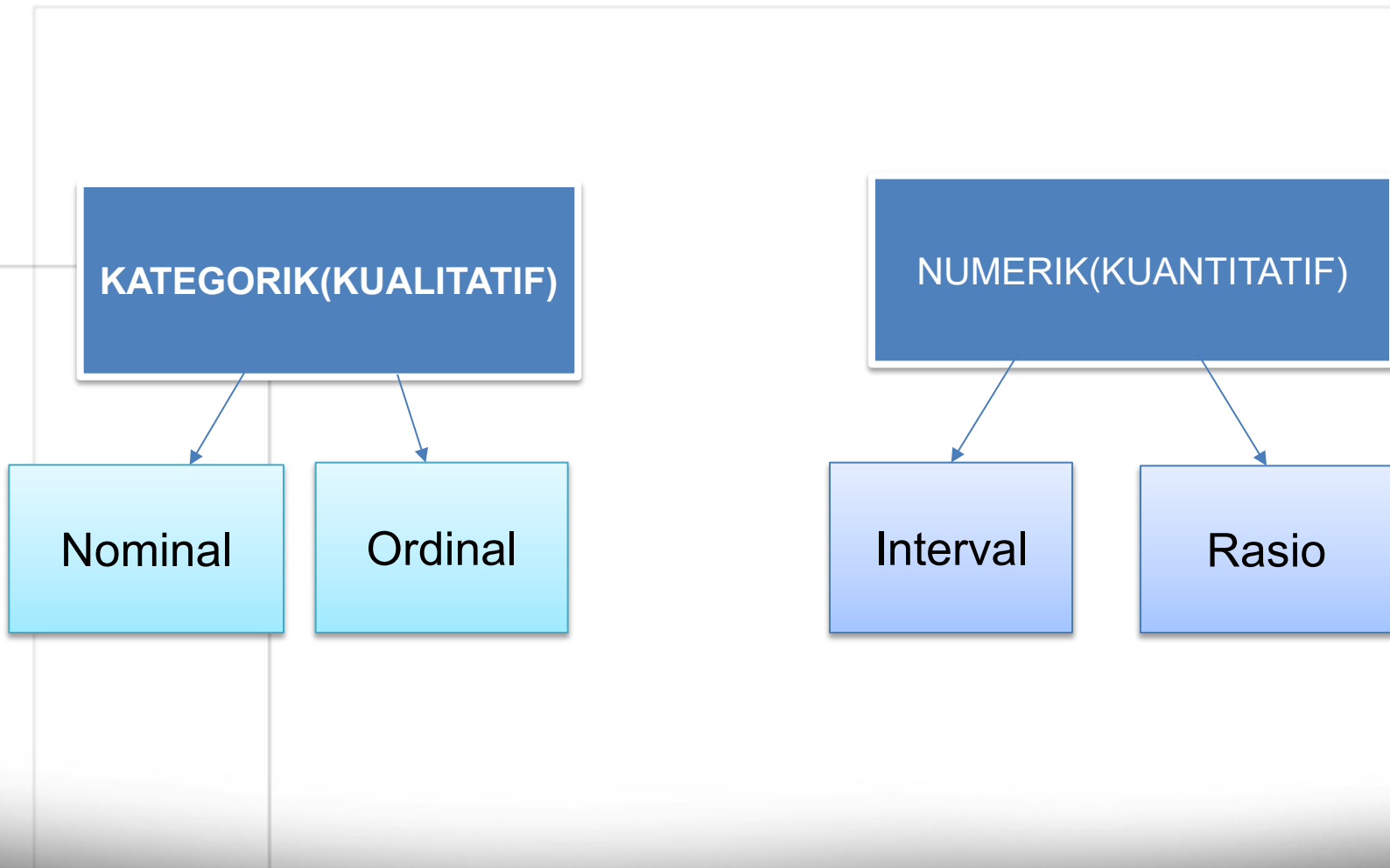
Eksplorasi data merupakan langkah untuk memahami data sebelum dilakukan praproses. Tujuan dari ekplorasi data adalah menyeleksi teknik pemrosesan dan analisis data yang sesuai dengan dataset yang dimiliki.

Hal-hal yang harus diperhatikan:

- Tipe data
- Kualitas data
- Statistika ringkasan
- Visualisasi



# TIPE DATA



# TIPE DATA

Tipe data		Deskripsi	Contoh
Kategorik (kualitatif)	Nominal	<p>Nilai dari atribut nominal adalah nama-nama sebagai pembeda antara satu dengan yang lain. Nilai nominal menyediakan informasi yang cukup untuk membedakan satu objek dengan objek yang lain.</p> <p>Operator aritmatik yang dapat digunakan ialah sama dengan (=) atau tidak sama dengan (<math>\neq</math>).</p>	PIN kartu ATM , NIM Mahasiswa, warna RGB , kode biner.
	Ordinal	<p>Nilai dari atribut ordinal adalah nama-nama yang selain sebagai pembeda juga dapat menjadi ukuran perbandingan satu dengan yang lain. Nilai ordinal menyediakan informasi yang cukup untuk mengurutkan objek.</p> <p>Operator aritmatik yang dapat digunakan ialah sama dengan (=), tidak sama dengan (<math>\neq</math>) lebih besar (<math>&gt;</math>) , lebih kecil (<math>&lt;</math>).</p>	Nomor antrian, <i>Grade</i> , kecepatan prosesor {lambat, cepat, sangat cepat}

# TIPE DATA

Tipe data		Deskripsi	Contoh
Numerik (kuantitatif)	Interval	<p>Dalam atribut interval perbedaan antarnilai merupakan sesuatu yang berarti, pada atribut interval terdapat unit pengukuran.</p> <p>Operator aritmatik yang dapat digunakan ialah sama dengan (=), tidak sama dengan (<math>\neq</math>) lebih besar (<math>&gt;</math>) , lebih kecil (<math>&lt;</math>), penambahan (+), pengurangan (-).</p>	Tanggal pada kalender, Temperatur.
	Rasio	<p>Dalam atribut rasio, perbedaan rasio merupakan hal yang berarti.</p> <p>Operator aritmatik yang dapat digunakan ialah sama dengan (=), tidak sama dengan (<math>\neq</math>) lebih besar (<math>&gt;</math>) , lebih kecil (<math>&lt;</math>), penambahan (+), pengurangan (-), perkalian (*), pembagian (/).</p>	Umur, berat badan, tinggi badan.

# KUALITAS DATA

*Noise*

*Outliers*

*Missing Value*

Duplicate

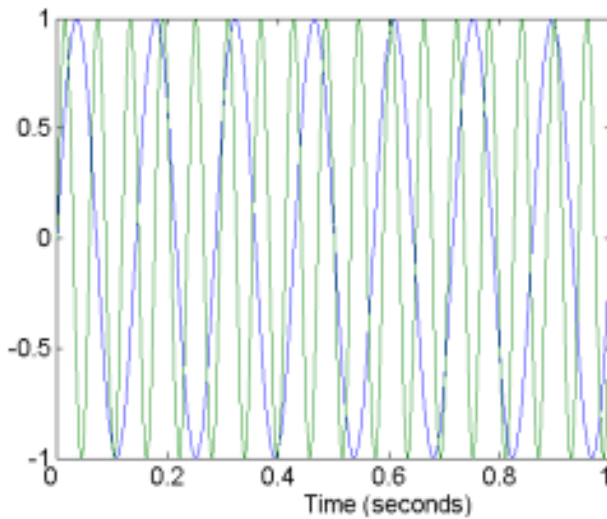


# KUALITAS DATA

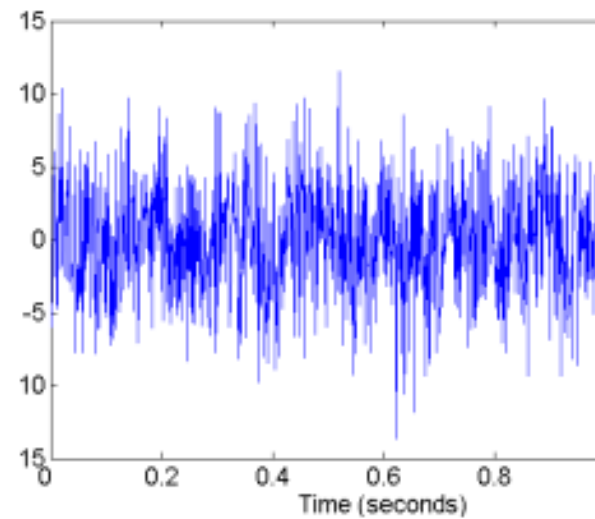
**NOISE**

Modifikasi dari  
nilai sebenarnya

Ex: 'Snow'  
pada tv



Two Sine Waves



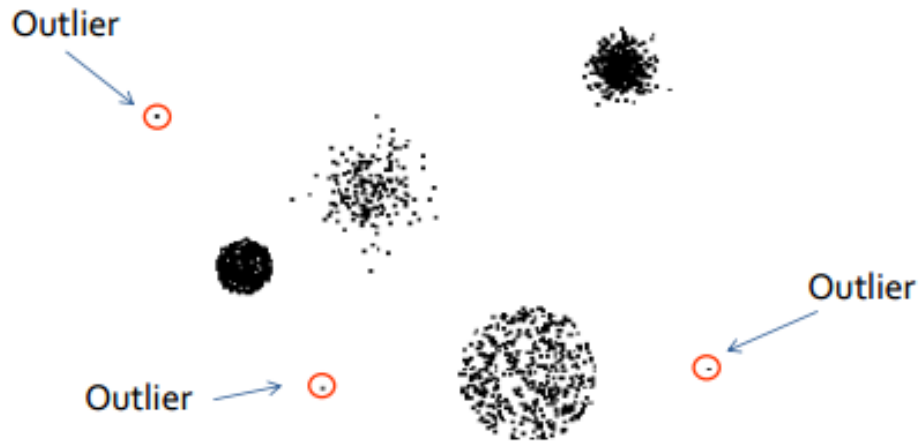
Two Sine Waves + Noise



# KUALITAS DATA

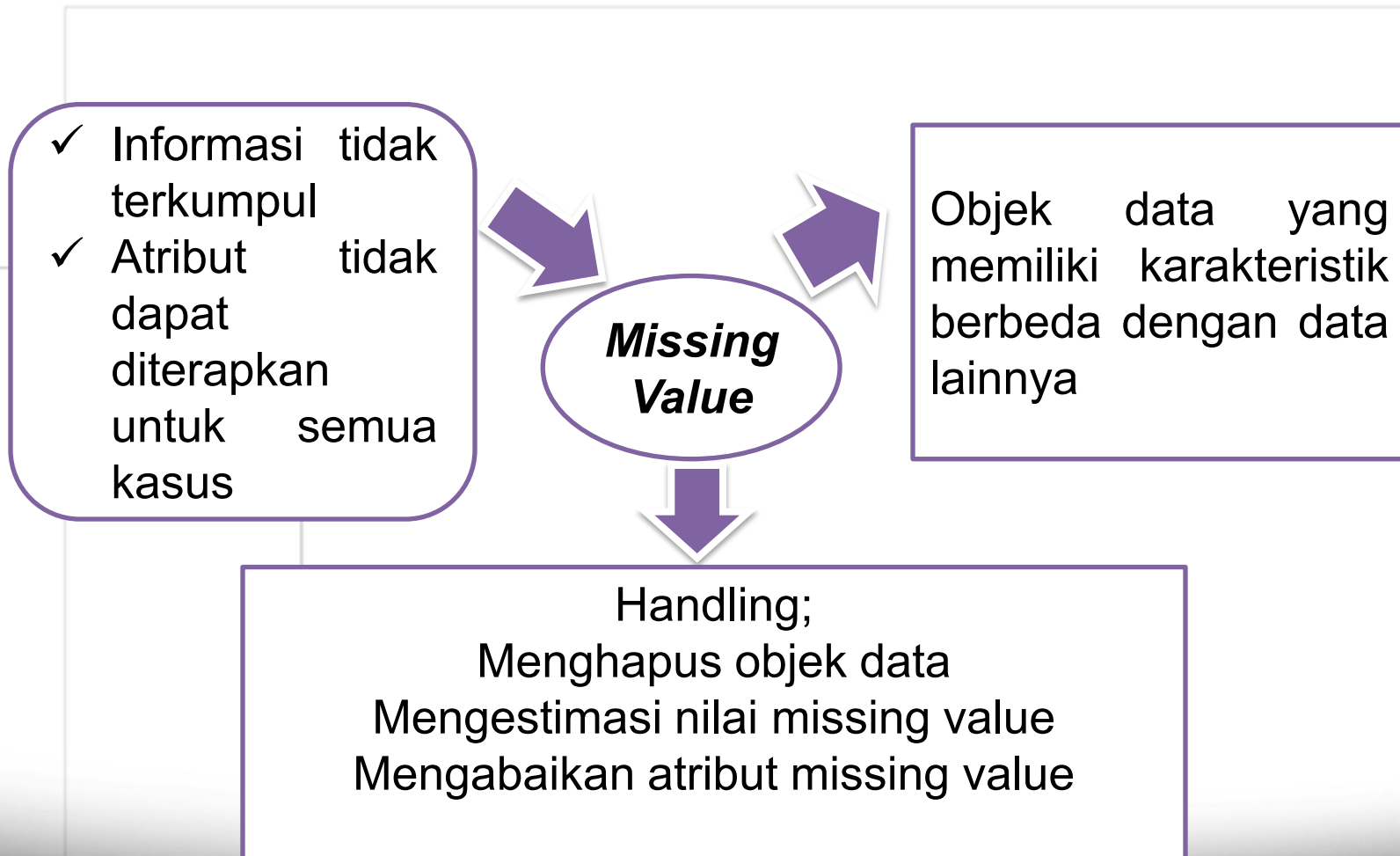
**OUTLIER**

Objek data yang memiliki karakteristik berbeda dengan data lainnya



Outlier dapat dipandang sebagai *noise* tetapi berguna dalam *fraud detection*, *rare event analysis*

# KUALITAS DATA



# KUALITAS DATA

***Duplicate  
data***

Masalah utama ketika  
menggabungkan data dari  
berbagai sumber



Data Cleaning  
Menghilangkan noise dan  
data yang tidak konsisten



# STATISTIKA RINGKASAN

Statistika ringkasan adalah sejumlah ringkasan property dari data.

## FREKUENSI

Presentase nilai atribut tersebut muncul pada *dataset*

## MODUS

Nilai atribut yang paling sering muncul

## MEAN

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$$

## RANGE

Jarak antara nilai maks dan min

## MEDIAN

Untuk  $n$  ganjil:  $Me = x_{\frac{1}{2}(n+1)}$

Untuk  $n$  genap:  $Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

## VARIANCE

$$s_x^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2$$

# VISUALISASI DATA

Visualisasi data adalah salah satu teknik dalam eksplorasi data.

Manfaat visualisasi data:

- Dapat mendeteksi general pola dan trends
- Dapat mendeteksi *outlier* atau *unusual* trends



Boxplots



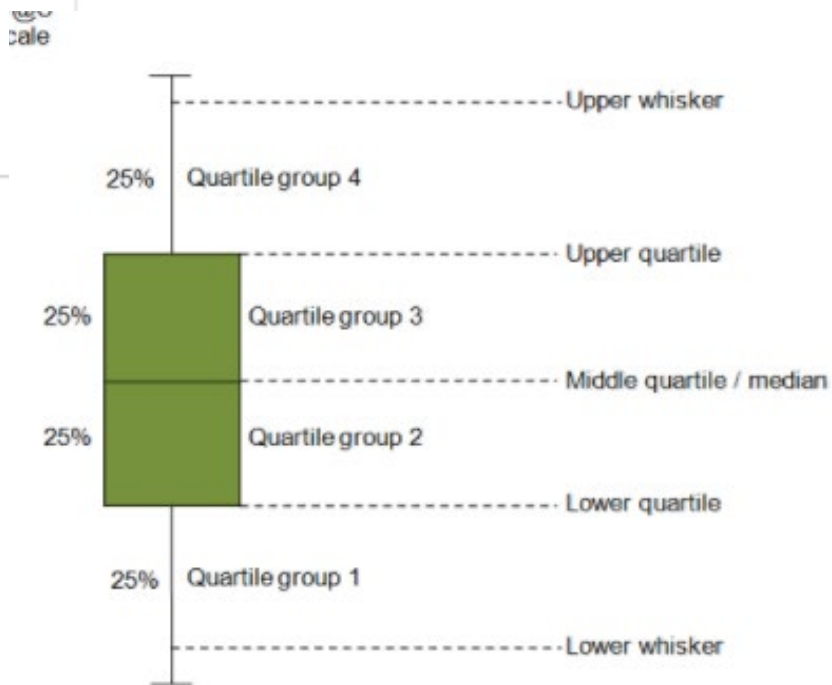
Histogram



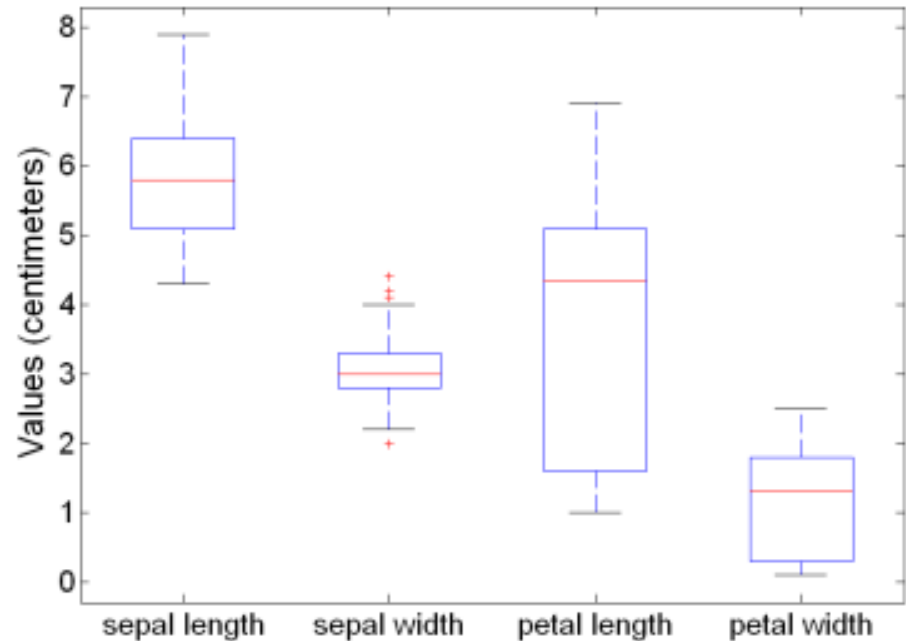
Scatter  
plot

# VISUALISASI DATA

- BOXPLOTS



Bagian-bagian dari boxplots

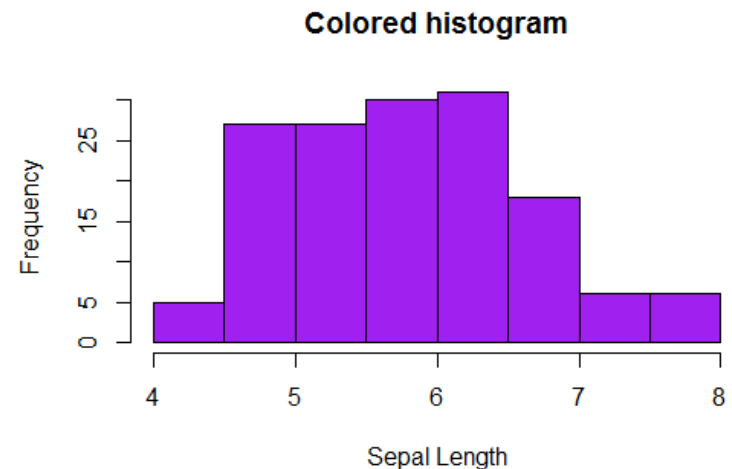


Boxplots dapat digunakan untuk membandingkan atribut

# VISUALISASI DATA

- HISTOGRAM

- ☐ Histogram mendistribusikan nilai-nilai suatu atribut
- ☐ Membagi nilai menjadi bin-bin dan barplot menunjukkan jumlah objek pada setiap bin
- ☐ Tinggi dari setiap bar mengidentifikasikan jumlah objek

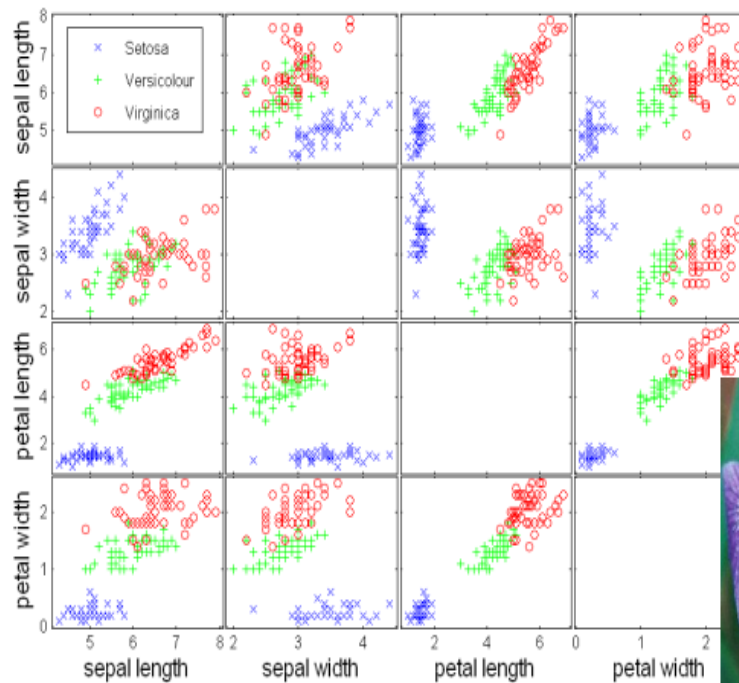




# VISUALISASI DATA

- SCATTER PLOT

- ❖ Nilai atribut menentukan posisi
- ❖ Atribut ditunjukkan dengan warna maupun bentuk yang berbeda dengan atribut lainnya
- ❖ Dapat melihat hubungan beberapa pasangan atribut



Source: <http://miralab.org>

**TERIMA KASIH 😊**