

Intermediate Statistics

Advanced Statistics

Introduction to Data
Visualization

Learning Progress Review Week 10

By:
MARVEL TEAM

Fikrie | Natalia | Satria

1. *Intermediate Statistics*

KORELASI

Korelasi adalah kesamaan pola kemunculan. Pola hubungan antara dua hal atau variable yang muncul bersamaan dalam satu waktu. Variabel adalah sesuatu yang bisa berubah dan bisa kita ukur.

Contoh kasus korelasi:

- ❖ Hubungan antara kenaikan harga BBM (X) **dengan** harga kebutuhan pokok (Y).
- ❖ Hubungan tingkat pendidikan (X) **dengan** tingkat pendapatan (Y).

KAUSALITAS

Kausalitas adalah hubungan antara dua hal sebab akibat.

Kausalitas di tinjau dari segi knowledge domain, apakah suatu variable benar menyebabkan terjadinya suatu variable lainnya.

Contoh kasus kausalitas:

- ❖ Kamu akan menjadi juara pertama di kelas *jika* kamu belajar dengan rajin.
- ❖ Lingkungan rumah kami kebanjiran *karena* banyak warga yang membuang sampah sembarangan ke sungai

Probabilitas

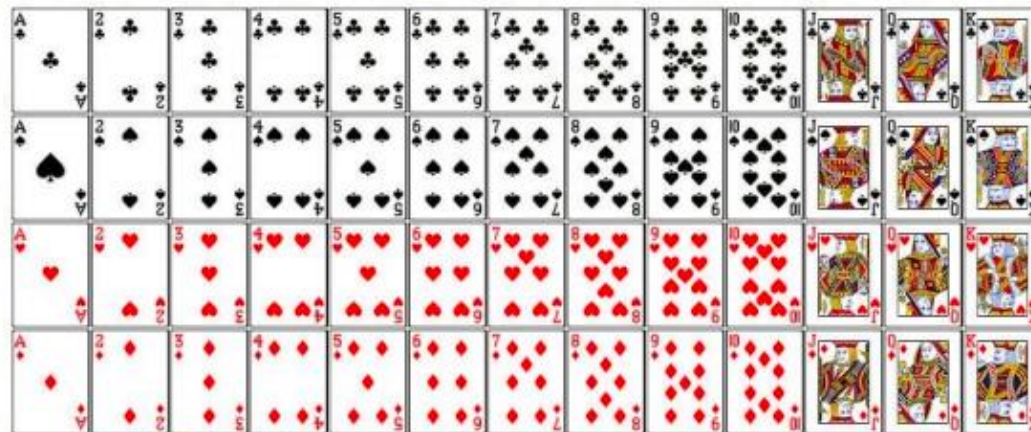
Probabilitas adalah suatu nilai yang digunakan untuk mengukur tingkat terjadinya suatu kejadian yang acak. Kata probabilitas itu sendiri sering disebut dengan peluang atau kemungkinan

Probabilitas

Suatu kejadian dinyatakan memiliki nilai probabilitas 0(nol), jika status peristiwa atau kejadian tidak memiliki peluang sama sekali untuk terjadi (Tidak akan Terjadi).

Sebaliknya, justru kejadian juga dinyatakan memiliki nilai probabilitas 1(satu), jika suatu peristiwa atau kejadian tersebut pasti terjadi, dan tidak ada kemungkinan selain itu.

Contoh kasus Probabilitas:

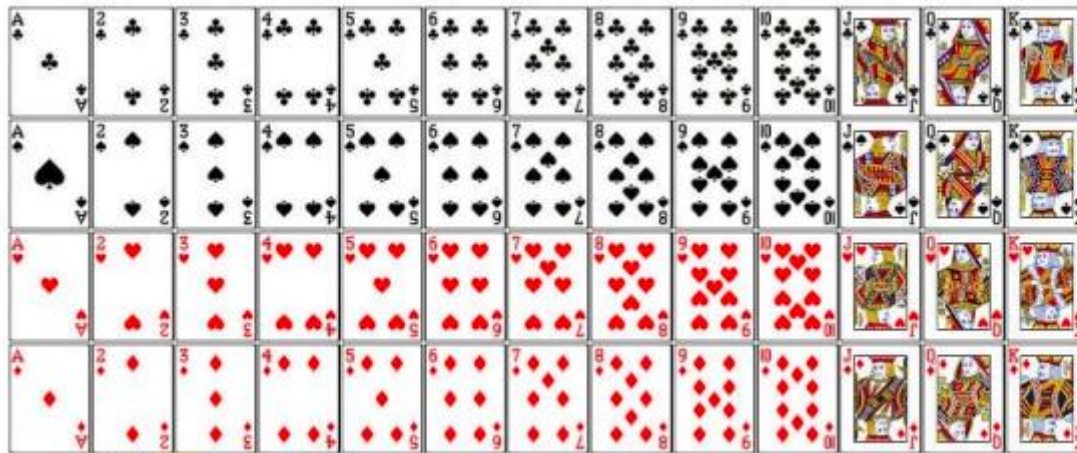


$P(\text{Queen}) = ?$

$P(\text{Red}) = ?$

$P(3) = ?$

Probabilitas



$$P(\text{Queen}) = 4/52$$

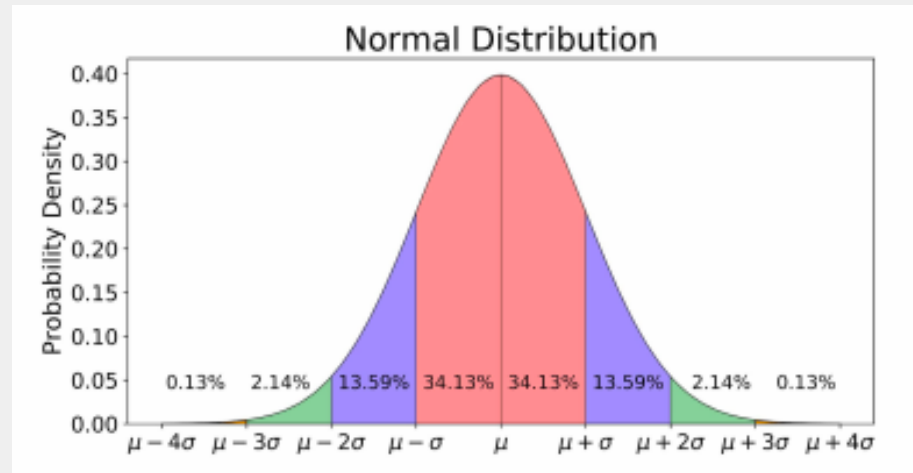
$$P(\text{Red}) = 26/52$$

$$P(3) = 4/52$$

Distribusi

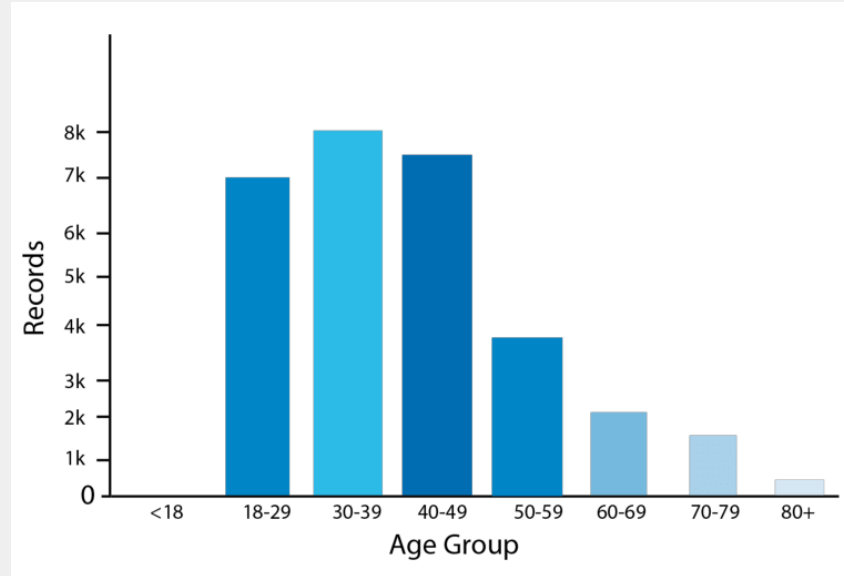
Distribusi Data adalah suatu fungsi yang menunjukkan semua nilai dari sebuah data dan seberapa sering nilai tersebut terjadi.

Contoh kasus Distribusi: **Distribusi normal** adalah distribusi peluang kontinu yang paling sering penting dalam statistika



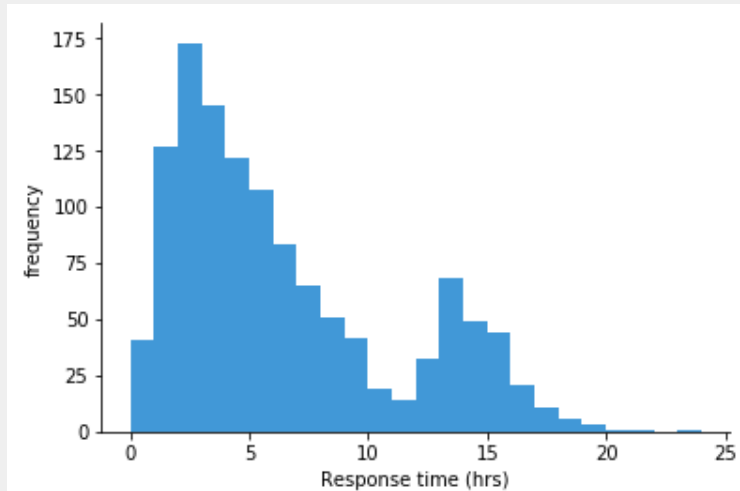
Jenis – jenis Statistics Plot

Barchart

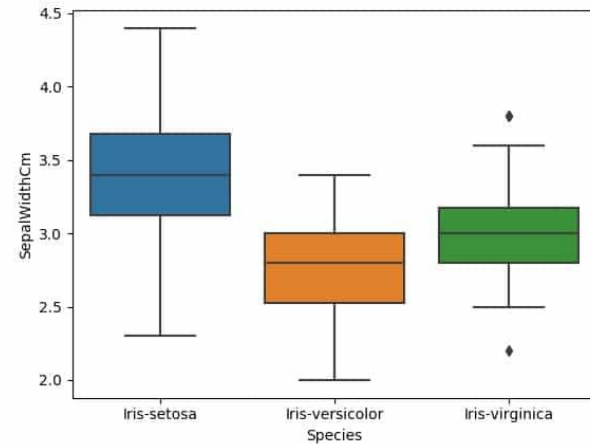


Statistics Plot

Histogram

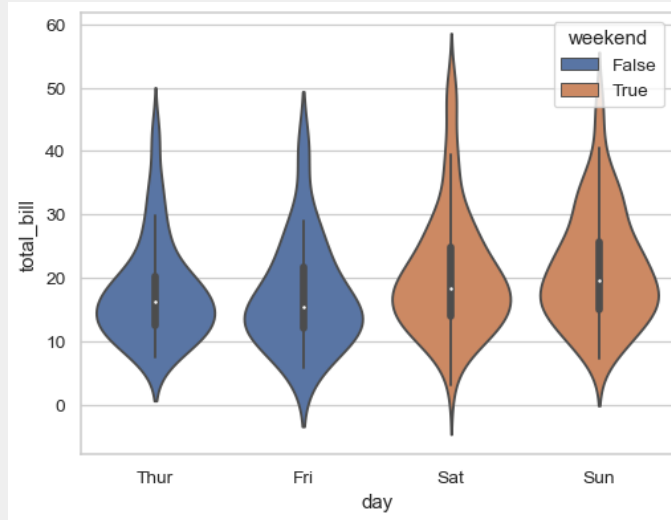


Box Plot

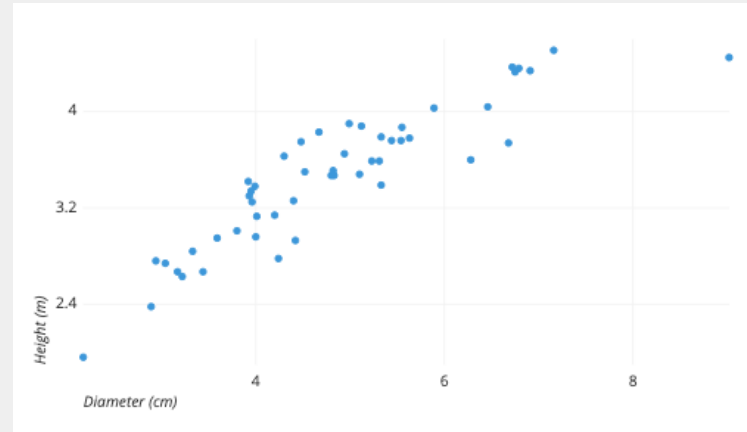


Statistics Plot

Violin Plot

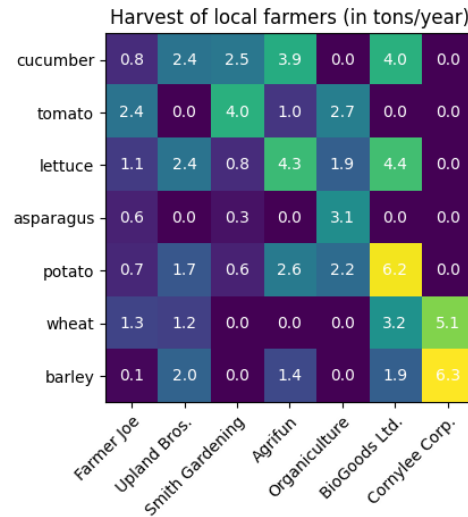


Scatter Plot

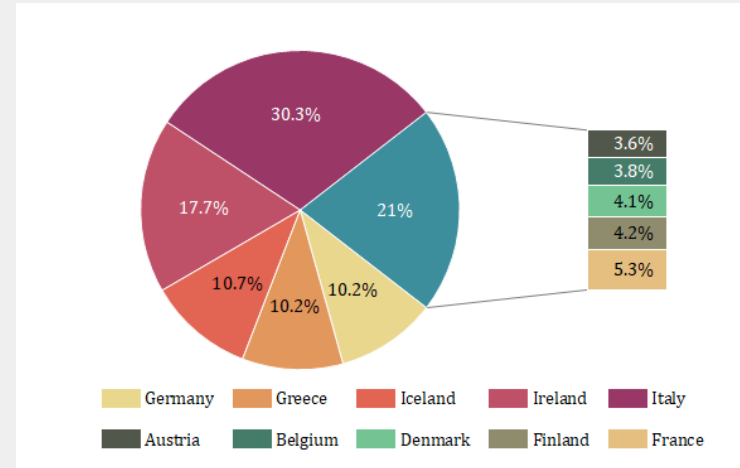


Statistics Plot

Heatmap Plot



Pie Chart



2. *Advanced Statistics*

Metode *Sampling*

Probability Sampling

Teknik pengambilan sampel yang memungkinkan seluruh anggota populasi memiliki peluang yang sama untuk terpilih sebagai sampel.

Non-Probability Sampling

Teknik pengambilan sampel yang tidak mempertimbangkan peluang anggota populasi untuk terpilih sebagai sampel.

Mengapa *Sampling?*

- ❖ Biaya sangat besar & waktu yang sangat lama jika mengambil data keseluruhan populasi.
- ❖ Biaya komputasi sangat besar jika mengolah data yang mencakup seluruh populasi.
- ❖ Data yang didapatkan efektif untuk diolah dan dianalisis.

Kesalahan *Sampling*

Sampling Error

Kesalahan yang merupakan konsekuensi dari sampling karena hanya mengambil data dari sebagian populasi.

Non-Sampling Error

Kesalahan-kesalahan teknis yang merupakan konsekuensi dari berjalannya survei.

Ukuran Sampel

Slovin :

$$n = \frac{N}{1 + Ne^2}$$

N : Number of Population

e : Margin of Error

Pengujian Hipotesis

- ❖ Bertujuan membuktikan suatu pernyataan umum dengan menggunakan data-data sampel.
- ❖ Berprinsip praduga tak bersalah.
- ❖ Hipotesis Nol/*Null Hypothesis* (H_0) adalah hipotesis yang berkontradiksi dari pernyataan yang diuji.
- ❖ Hipotesis Alternatif/*Alternative Hypothesis* (H_1/H_a) adalah hipotesis yang ingin diuji dan bersifat spesifik.

Metriks Pengukur Resiko Kesalahan

- ❖ Metriks untuk mengukur resiko kesalahan dinamakan *P-value* (*Probability Value*).
- ❖ *P-value* berbentuk persentase/peluang yang akan menentukan apakah akan menolak H_0 atau menerima H_1/H_a .
- ❖ Pengambilan keputusan didasarkan :
 1. *P-value* $< \alpha$, maka H_1/H_a .
 2. *P-value* $> \alpha$, maka H_0 .
- ❖ Penentuan α berdasarkan analisa kita.

Pengujian Statistika

Pengujian statistika sangat banyak, namun yang jamak digunakan seperti,

- **T-Test**
- **ANOVA**
- **Pearson Correlation Test**
- **Chi-Square**
- **Wilcoxon Rank-Sum Test**
- **Sign Test**

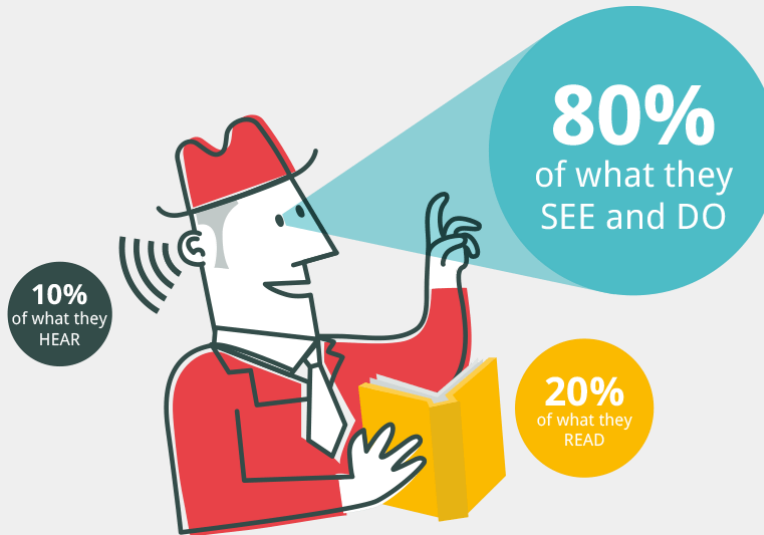
3. *Introduction to Data Visualization*

Apa itu visualisasi data?



Visualisasi data atau data visualization adalah tampilan berupa grafis atau visual dari informasi dan data.

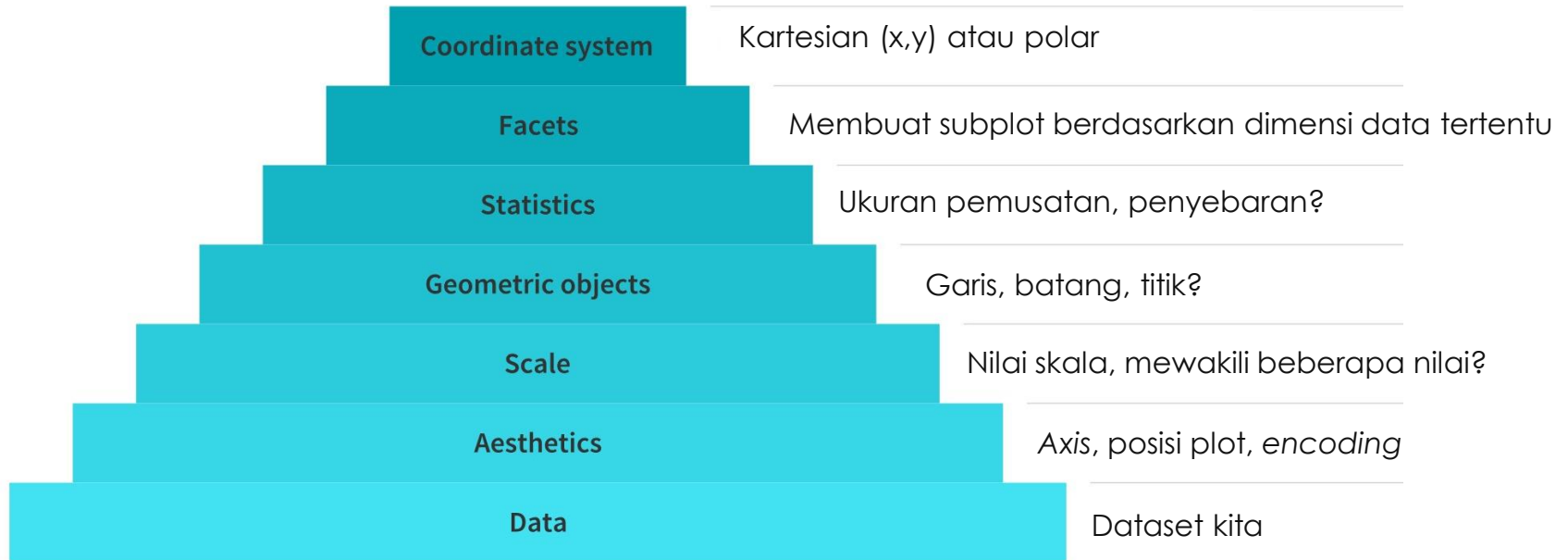
Mengapa visualisasi data itu penting?



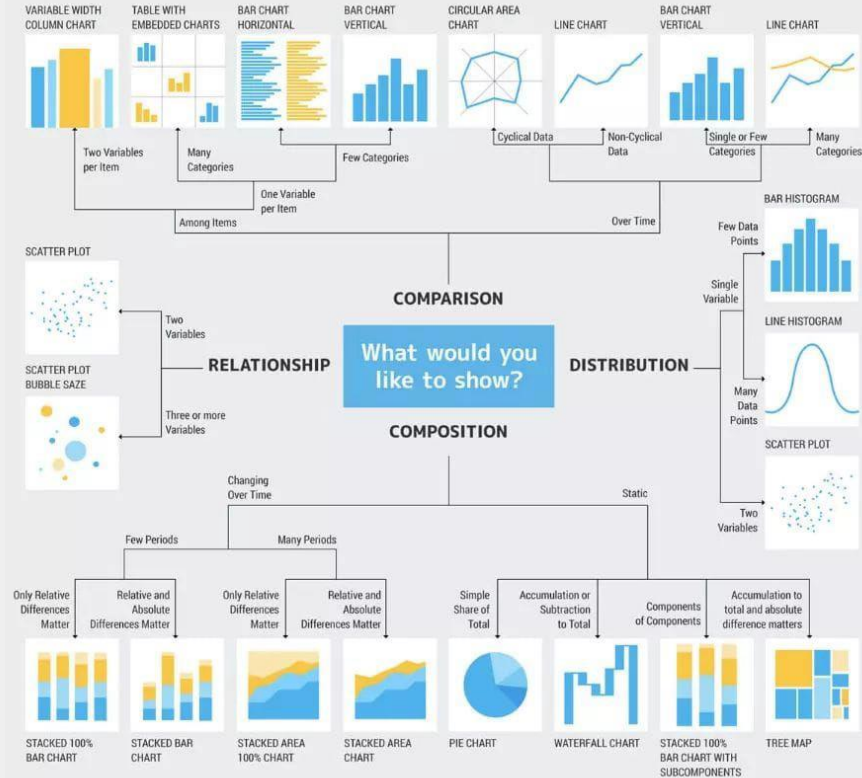
Data yang ditampilkan secara visual memungkinkan banyak orang untuk memahami data dengan lebih cepat.

Menurut penelitian, orang mengingat 80% dari apa yang mereka lihat, tapi cuma 20% dari apa yang mereka baca. Jadi, otak kita dapat mengingat gambar jutaan kali lebih cepat daripada kata.

Komponen dalam menggambarkan visualisasi data



Framework untuk memilih visualisasi data



Kesalahan umum ketika visualisasi data

1. Grafik tidak efektif

2. Nilai axis memakai cara yang tidak intuitif

3. Dalam format yang sulit dipahami

4. Mengabaikan bagaimana data disimpan

Special Thanks to :



Slide template by SlideCarnival