

# PERTEMUAN 11

## Visualisasi Data

# Visualisasi

Visualisasi berperan peran penting dalam bidang machine learning dan data science. Seringkali kita perlu menyaring informasi kunci yang ditemukan dalam sejumlah data data menjadi bentuk yang bermakna dan mudah dicerna.

Visualisasi yang baik dapat menceritakan sebuah cerita tentang data Anda dengan cara yang tidak dapat dilakukan oleh sebuah kalimat

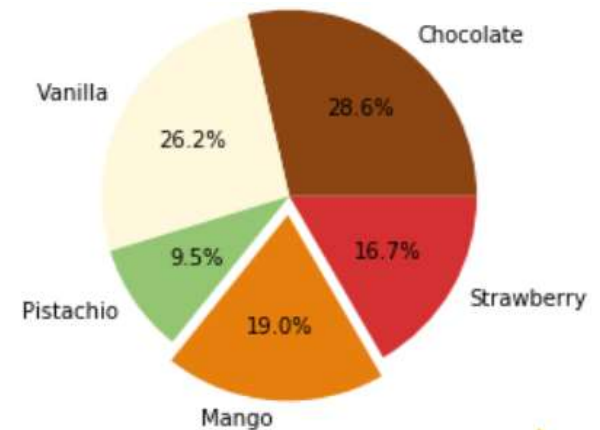
# Pie Chart

Pie chart digunakan untuk menunjukkan seberapa banyak dari setiap jenis kategori dalam dataset berbanding dengan keseluruhan.

Variabel label berisi tupel rasa es krim

Variabel voting berisi tupel voting.

Data tersebut mewakili jumlah voting rasa es krim favorit

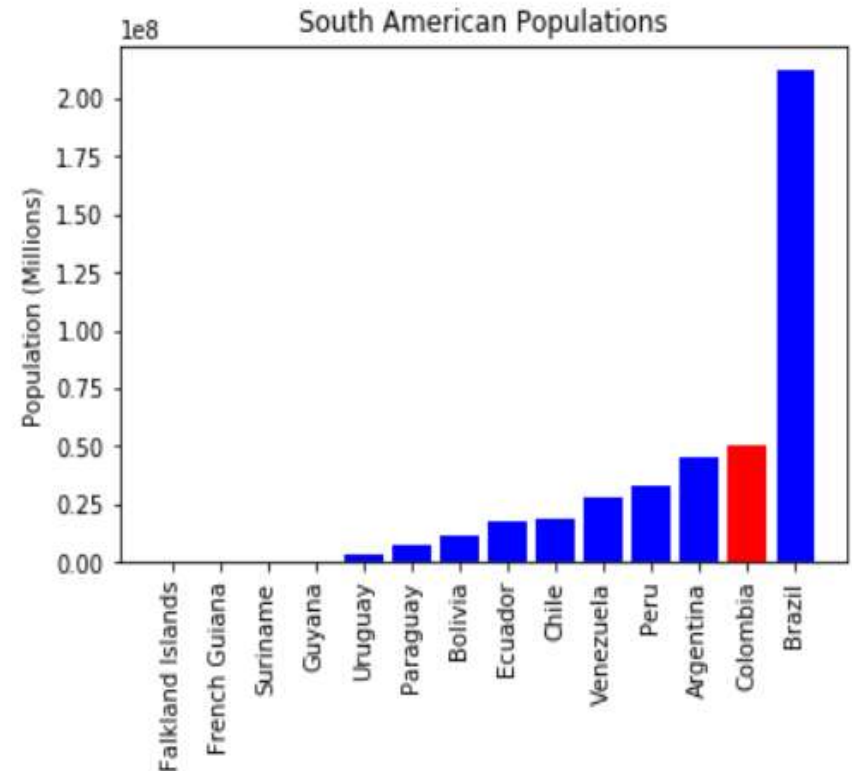


# Bar Chart

Bar Chart adalah merupakan tools visualisassi yang dapat digunakan untuk membandingkan data kategorikal.

Mirip dengan diagram lingkaran, diagram ini dapat digunakan untuk membandingkan kategori data satu sama lain.

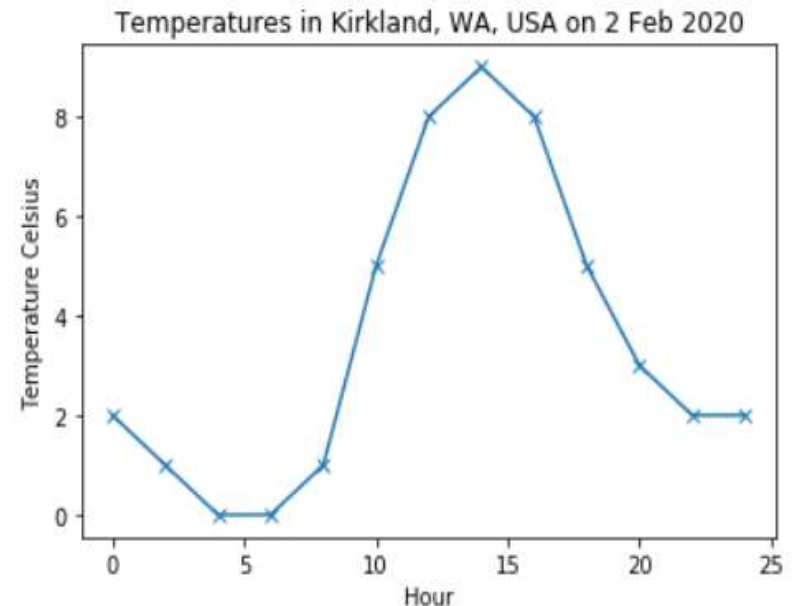
Diagram batang dapat menampilkan lebih banyak kategori data daripada diagram lingkaran.



# Line Graph

Line Graph adalah bentuk visualisasi lainya selain diagram lingkaran dan diagram batang.

- Diagram garis lebih berguna untuk menunjukkan bagaimana kemajuan data selama beberapa periode.
- Misalnya, grafik garis dapat berguna dalam membuat grafik temperatur dari waktu ke waktu, harga saham dari waktu ke waktu, berat menurut hari, atau metrik berkelanjutan lainnya.



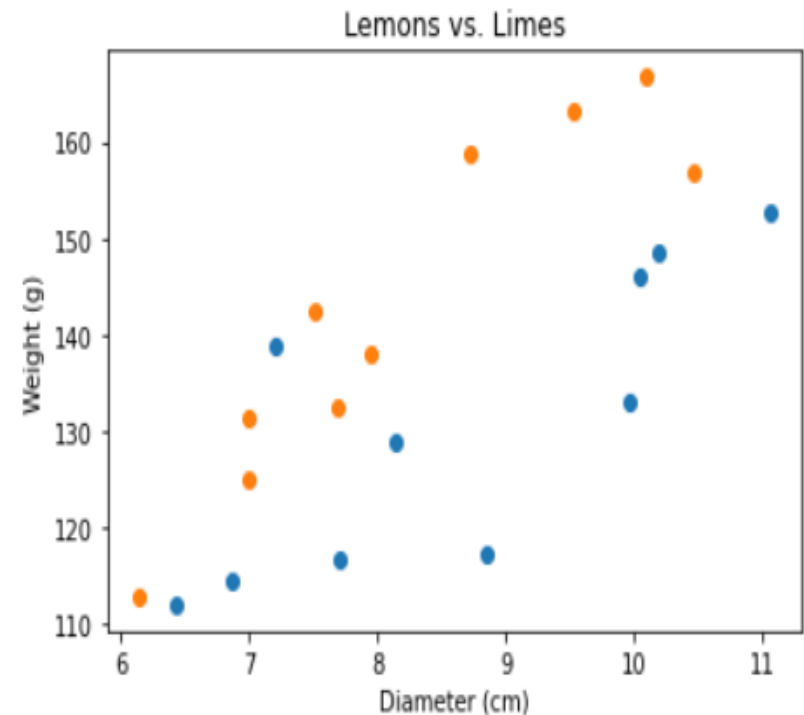
# Scatter Plot

Scatter plot berfungsi baik untuk data dengan dua komponen numerik.

Scatter plot dapat memberikan informasi yang berguna terutama mengenai pola atau pencilan.

Pada contoh di bawah ini, kita memiliki data yang terkait dengan perbedaan lemon dan lime berdasarkan karakteristik fisiologis.

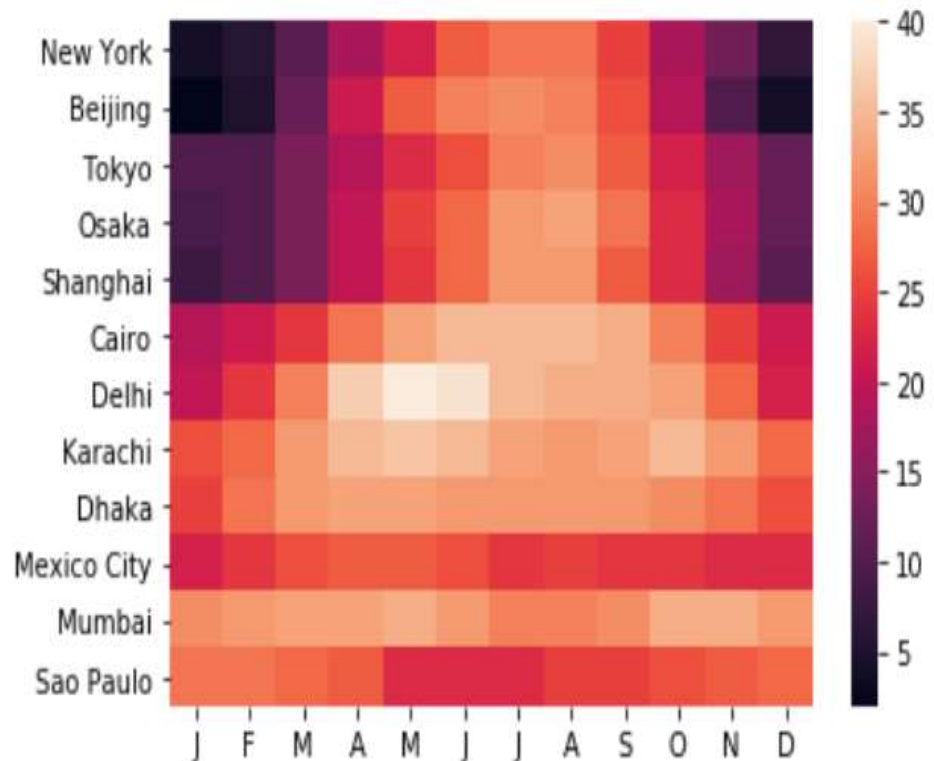
- Berat (g)
- Diameter (cm)



# Heatmap

Heatmap adalah jenis visualisasi yang menggunakan kode warna untuk mewakili nilai / kepadatan relatif data di seluruh permukaan.

Warna-warna ini kemudian dapat digunakan untuk memeriksa data secara visual guna menemukan kelompok dengan nilai serupa dan mendeteksi tren dalam data



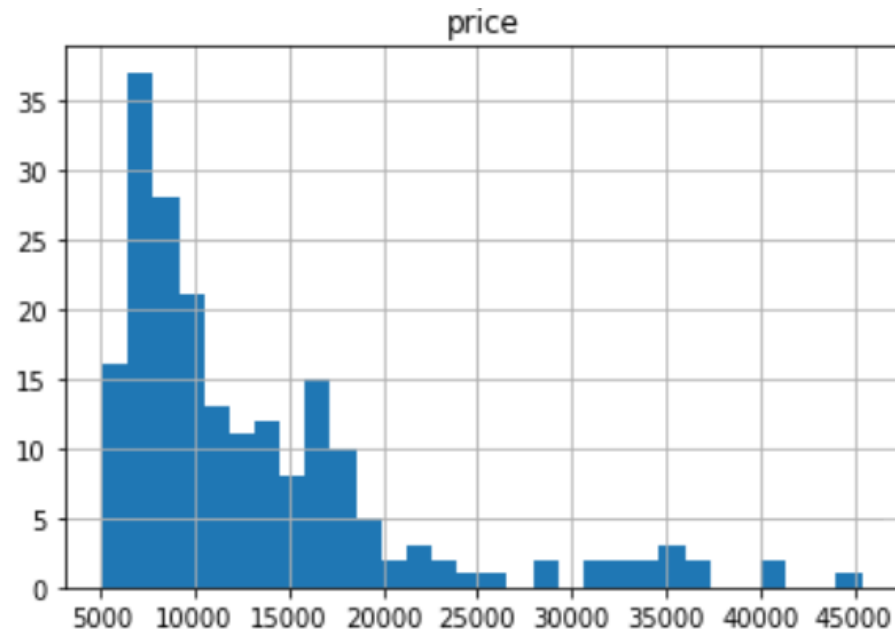
# Histogram

Histogram adalah salah satu visualisasi yang cukup penting dalam memahami distribusi pada data kita.

Histogram menyediakan method yang memudahkan kita untuk membuat histogram.

Plot histogram secara tradisional hanya membutuhkan satu dimensi data.

Ini dimaksudkan untuk menunjukkan jumlah nilai atau kumpulan nilai secara serial.



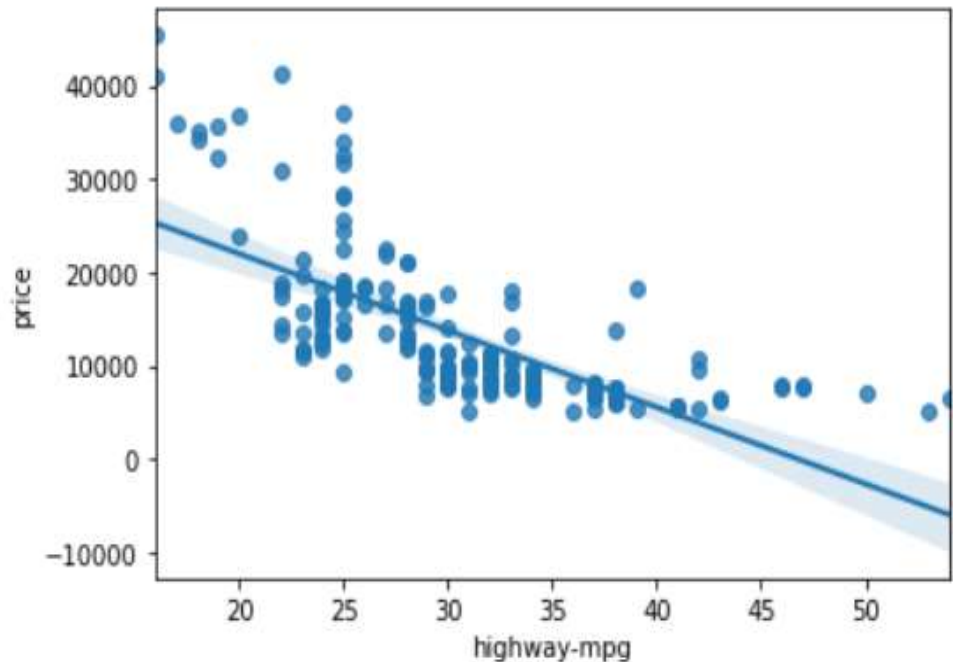


# Correlation & Causation

Korelasi merupakan suatu pengukuran sejauh mana nilai saling ketergantungan antar variabel.

Causation merupakan hubungan antara sebab dan akibat antara dua variable. Penting untuk mengetahui perbedaan antara keduanya dan bahwa korelasi tidak mendeskripsikan sebab-akibat.

Menentukan korelasi jauh lebih sederhana menentukan sebab memerlukan analisis lebih lanjut



# Correlation & Causation

## P-Value:

Berapa nilai P ini? Nilai P adalah nilai probabilitas bahwa korelasi antara kedua variabel ini signifikan secara statistik. Biasanya, kita memilih tingkat signifikansi 0,05, yang berarti bahwa kami yakin bahwa 95% korelasi antar variabel signifikan.

Dengan konvensi, Ketika

- nilai p adalah  $\leq 0,001$ : kami katakan ada bukti kuat bahwa korelasinya signifikan. nilai p adalah  $\leq 0,05$ : terdapat bukti moderat bahwa korelasi tersebut signifikan.
- nilai p adalah  $\leq 0,1$ : ada bukti lemah bahwa korelasinya signifikan.
- nilai p adalah  $> 0,1$ : tidak ada bukti bahwa korelasi tersebut signifikan.