

Data Visualization with Python

20 Mar 2024



Agenda

1. Intro to Data Visualization
2. Matplotlib and Seaborn
3. Plot with Pandas

Data Visualization

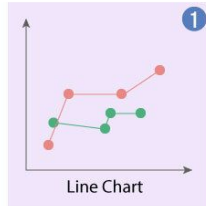
Data Visualization

- Visualisasi data merupakan proses **representasi informasi atau data secara grafis** atau visual.
- Tujuan utamanya adalah untuk **menyampaikan informasi secara jelas dan efektif** melalui penggunaan elemen visual seperti grafik, diagram, peta, dan berbagai alat grafis lainnya.
- Visualisasi data memungkinkan pengguna untuk **melihat dan memahami tren, outlier, dan pola dalam data** dengan cara yang lebih intuitif dibandingkan dengan melihat melalui tabel data atau laporan tertulis.

Pentingnya Visualisasi Data

- **Mempercepat Pemahaman:** Otak manusia dapat memproses informasi visual lebih cepat daripada teks. Visualisasi membantu memahami kompleksitas data dengan cepat.
- **Mengungkapkan Wawasan:** Dengan visualisasi, kita bisa mengidentifikasi tren, korelasi, dan outlier yang mungkin tidak terlihat dari data mentah.
- **Memfasilitasi Komunikasi:** Grafik atau diagram dapat membantu menyampaikan pesan atau temuan penting dari data kepada orang lain, memudahkan komunikasi dan pembuatan keputusan.
- **Mendorong Eksplorasi Data:** Visualisasi memungkinkan pengguna untuk memanipulasi dan mengeksplorasi data, membuka peluang untuk menemukan insight baru.

TYPES OF DATA VISUALIZATION CHARTS



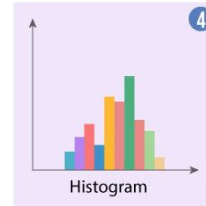
Display trends over time



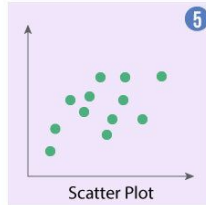
A line chart with areas below the lines filled with colors



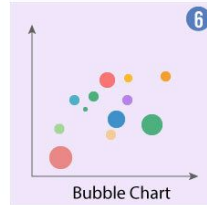
Display trends with multiple variables



Display the shape and spread of continuous dataset samples



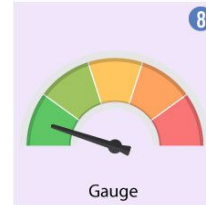
Show correlation in a dataset



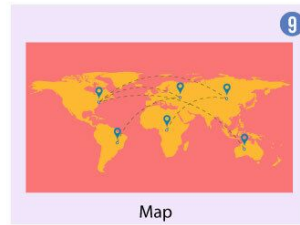
Show and compare the relationship between the labelled circles



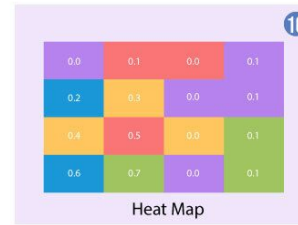
Show the contribution of data point inside a whole dataset



Visualize the distance between intervals



Show data with location as a variable



Show magnitude of a phenomenon

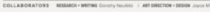
Proses Visualisasi Data

- **Pemahaman Tujuan:** Mengidentifikasi tujuan visualisasi, siapa audiensnya, dan pesan apa yang ingin disampaikan.
- **Pengumpulan dan Persiapan Data:** Memilih data yang relevan dan melakukan pembersihan data jika diperlukan.
- **Pemilihan Jenis Visualisasi:** Memilih jenis grafik atau diagram yang paling efektif untuk menyampaikan informasi.

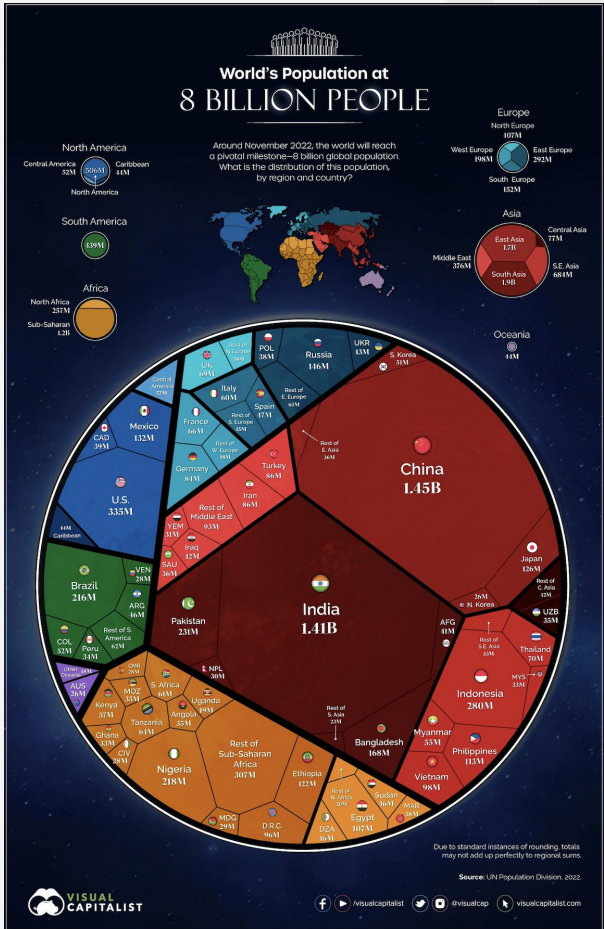
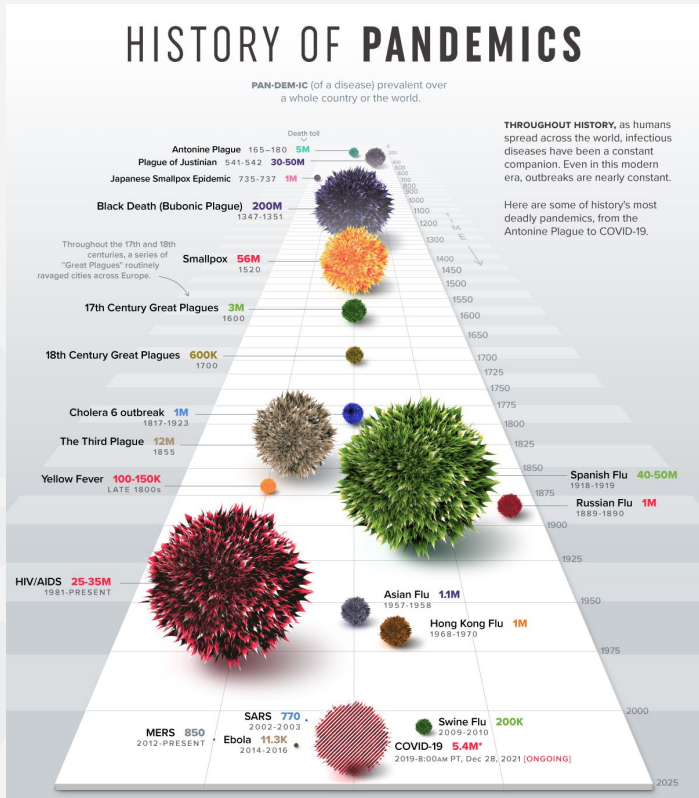
Proses Visualisasi Data

- **Desain Visualisasi:** Mengatur elemen visual seperti warna, font, dan layout untuk meningkatkan kejelasan dan estetika.
- **Implementasi dengan Alat Visualisasi:** Menggunakan perangkat lunak atau alat visualisasi data seperti Excel, Looker, Tableau, Microsoft Power BI, Matplotlib, atau Seaborn untuk membuat visualisasi.
- **Evaluasi dan Iterasi:** Menilai efektivitas visualisasi dan melakukan iterasi berdasarkan umpan balik untuk perbaikan.

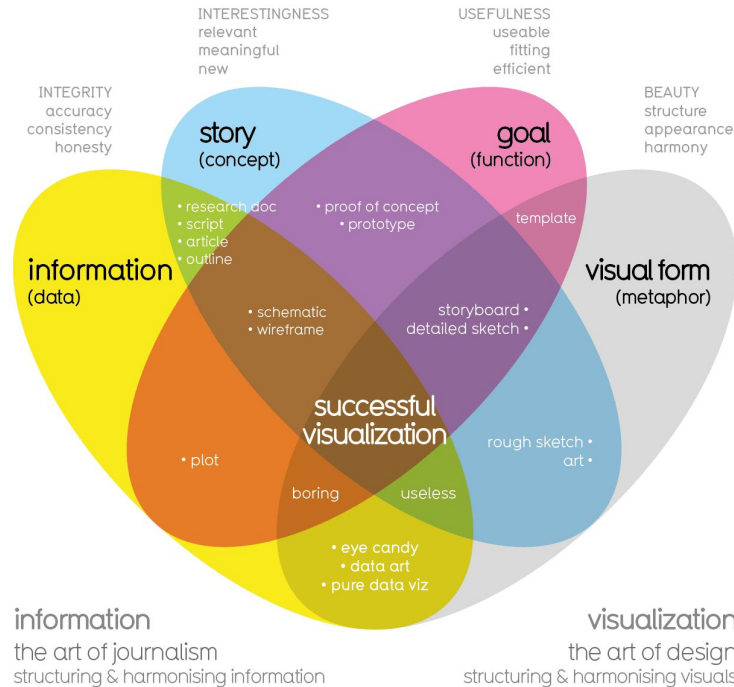
Below, we show the key players—from Google to Twitter—that currently dominate the Internet.



Contoh Visualisasi Data



What Makes a Good Visualization?



David McCandless
InformationIsBeautiful.net

find out more
bit.ly/KIB_Books



Matplotlib

- library plotting yang sangat populer untuk Python yang memungkinkan pengguna untuk membuat berbagai jenis visualisasi data 2D dengan cara yang relatif mudah dan intuitif.
- Dikembangkan oleh John D. Hunter pada tahun 2003
- Dapat menghasilkan kualitas tinggi figure dalam berbagai format file, seperti PNG, PDF, SVG, dan lain-lain.

Instalasi & Import Matplotlib

bash

 Copy code

```
pip install matplotlib
```

python

 Copy code

```
import matplotlib.pyplot as plt
```

Matplotlib

- library plotting yang sangat populer untuk Python yang memungkinkan pengguna untuk membuat berbagai jenis visualisasi data 2D dengan cara yang relatif mudah dan intuitif.
- Dikembangkan oleh John D. Hunter pada tahun 2003
- Dapat menghasilkan kualitas tinggi figure dalam berbagai format file, seperti PNG, PDF, SVG, dan lain-lain.




Seaborn

- Seaborn adalah library visualisasi data Python yang berbasis pada Matplotlib.
- Seaborn sangat bagus untuk dengan cepat membuat plot yang menarik secara visual dengan kode minimal, sementara Matplotlib menawarkan lebih banyak opsi kustomisasi dan kontrol yang detail atas setiap aspek dari sebuah plot.
- Seaborn menyediakan beberapa dataset bawaan yang tersimpan dalam dataframe pandas, memudahkan penggunaannya untuk visualisasi data dan analisis statistik.

Instalasi & Import Seaborn

bash

 Copy code

```
pip install seaborn
```

python

 Copy code

```
import seaborn as sns
```

Plot with Pandas

Pandas Plot

- Pandas menyediakan kemampuan untuk melakukan visualisasi data langsung dari DataFrame atau Series, memanfaatkan Matplotlib di balik layar untuk menghasilkan grafik.
- Fitur visualisasi ini membuatnya mudah untuk menjelajahi data tanpa perlu menulis banyak kode Matplotlib secara manual.

Pandas Plot

- Visualisasi data dengan Pandas memungkinkan Anda untuk:
 - Membuat berbagai jenis plot seperti line plot, bar chart, histogram, dan scatter plot dengan sintaks yang sederhana.
 - Menyesuaikan plot dengan berbagai opsi styling dan formatting.
 - Menampilkan data yang terstruktur dalam DataFrame atau Series secara grafis, yang sangat berguna untuk analisis eksploratif data.



Terima Kasih

