

# Mengapa perlu memahai Visualisasi Data ?

- Visualiasi data bukan hanya sekedar memperlihatkan data , tetapi **bercerita tentang Data**.
- We (human ) aren't naturally good at storytelling with data [5]



# Contoh visualisasi data (showing vs story telling)

## Survey Results

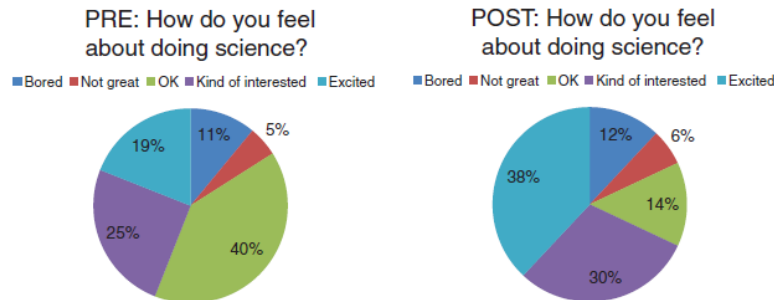
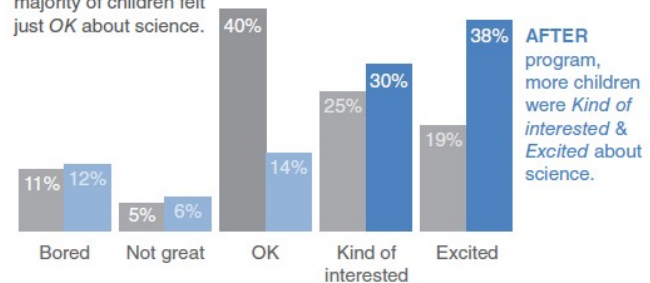


FIGURE 0.4 Example 2 (before): showing data

## How do you feel about science?

BEFORE program, the majority of children felt just OK about science.



Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 0.5 Example 2 (after): storytelling with data

# Contoh visualisasi data (showing vs story telling)

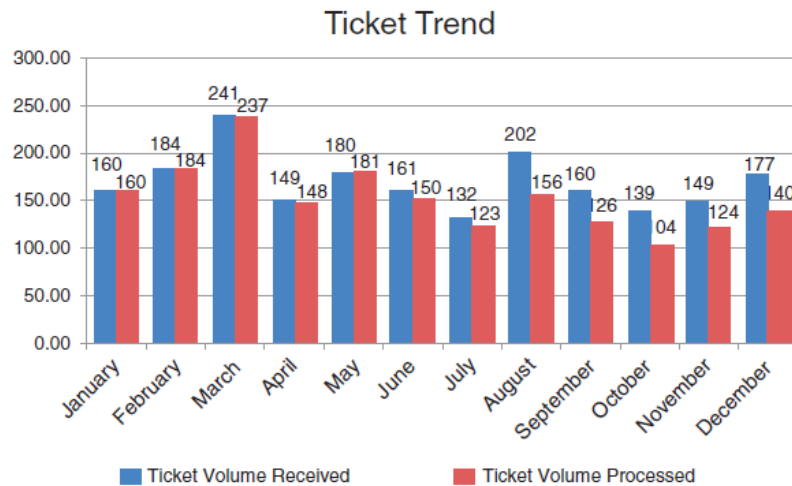
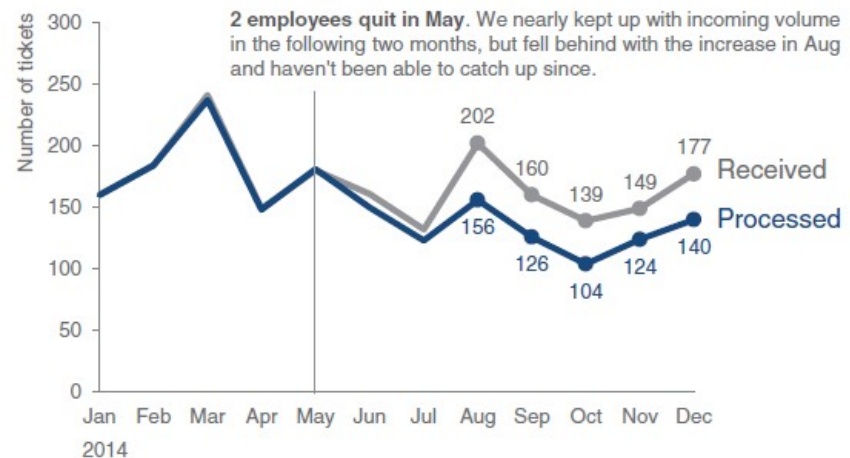


FIGURE 0.2 Example 1 (before): showing data



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

FIGURE 0.3 Example 1 (after): storytelling with data

# Tujuan dari Visualisasi

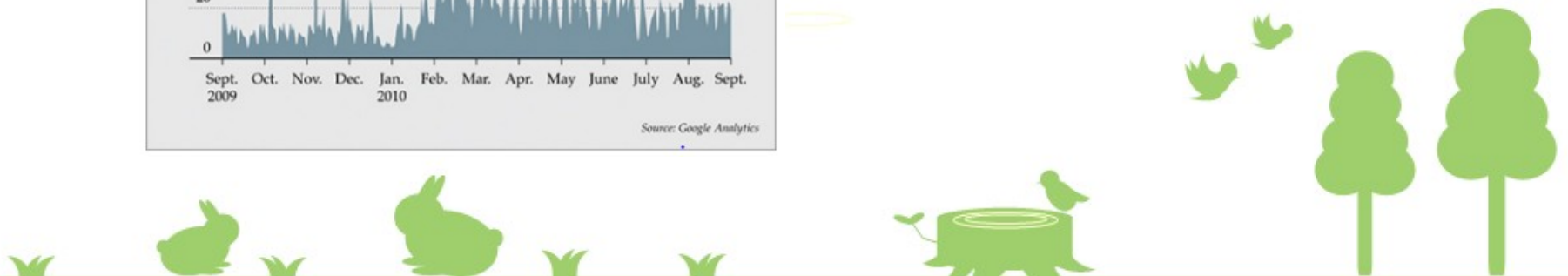
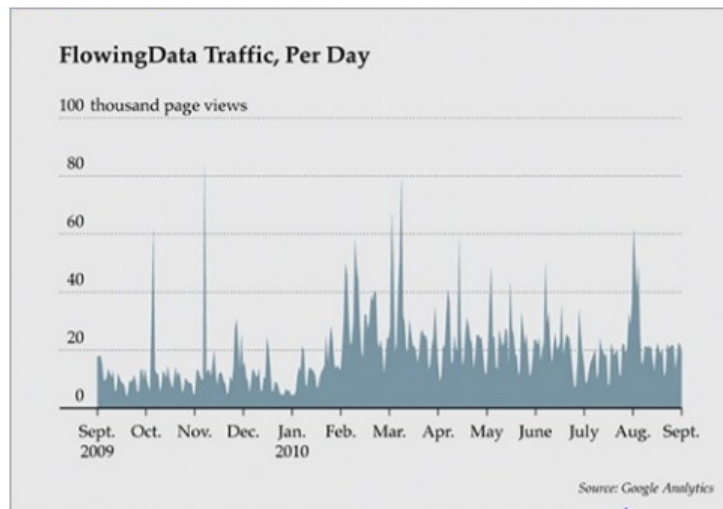
Secara umum, visualisasi data mempunyai tujuan untuk mencari:

1. Pattern (Pola)
2. Relations ( Keterkaitan)



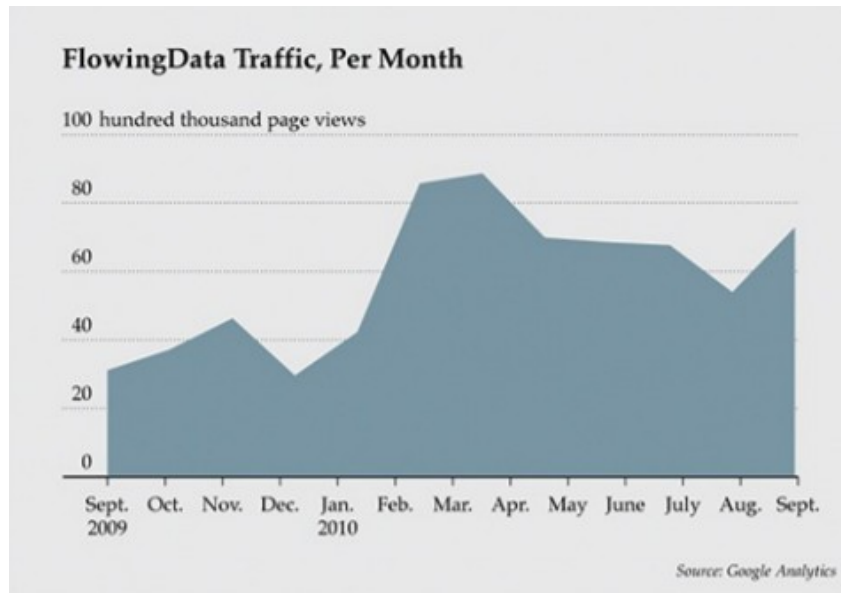
# Tujuan dari Visualisasi Data : Pattern

Pola dapat muncul pada sebuah data dan dapat di analisis untuk mendeteksi pemicu perubahan data.



# Tujuan dari Visualisasi Data : Relation

Banyak variable mungkin membentuk sebuah keterkaitan antara satu dengan yang lainnya.



# Cara membangun sebuah data visual

Terdapat 6 tahapan dalam membentuk data visual:

- 1. Understand the context*
- 2. Choose an appropriate visual display*
- 3. Eliminate clutter*
- 4. Focus attention where you want it*
- 5. Think like a designer*
- 6. Tell a story*



# Understand the context

- Untuk memahami konteks, dapat dimulai dari menjawab beberapa pertanyaan berikut:
  - 1. Who is your audience?*
  - 2. What do you need your audience to know or do?*
  - 3. (how) What data is available that will help make mypoint?*





# Memilih visualisasi yang tepat

Ada banyak cara untuk memvisualisasikan data, antara lain adalah menggunakan chart (paling umum).

91%

Simple text



Scatterplot

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Table



Line



# Eliminate Clutter

- Merujuk pada Gestalt Principles of Visual Perception. Pada Gestalt Principles of Visual Perception mempunyai 6 prinsip, yaitu:
  - *Proximity*
  - *Similarity*
  - *Enclosure*
  - *Closure*,
  - *Continuity*
  - *Connection*



# Focus attention where you want it

- Pada pembuatan visualisasi data perlu memerhatikan beberapa hal berikut:

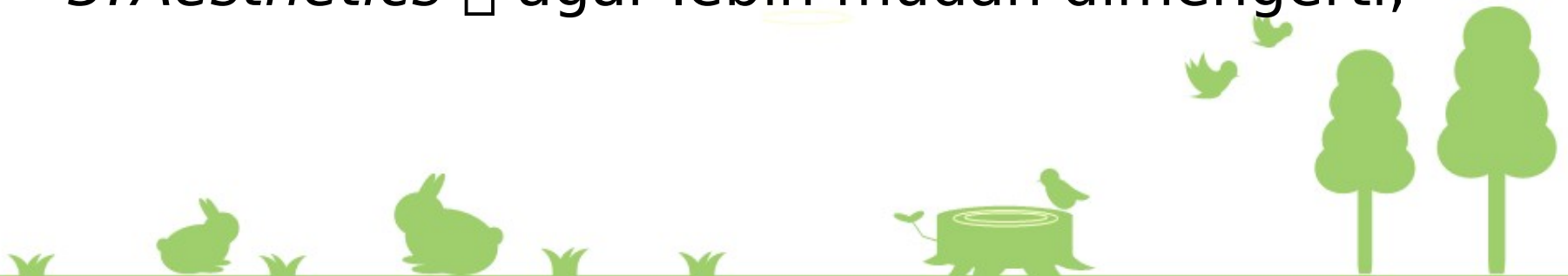
1. Ukuran
2. Warna
3. Tata Letak



# Think like a Designer

Terdapat beberapa aspek yang perlu diperhatikan , yaitu:

1. *Affordances* □ Kejelasan tentang aksi yang mungkin di antara objek dan sebuah individu
2. *Accessibility* □ dapat dimengerti oleh banyak kalangan
3. *Aesthetics* □ agar lebih mudah dimengerti,



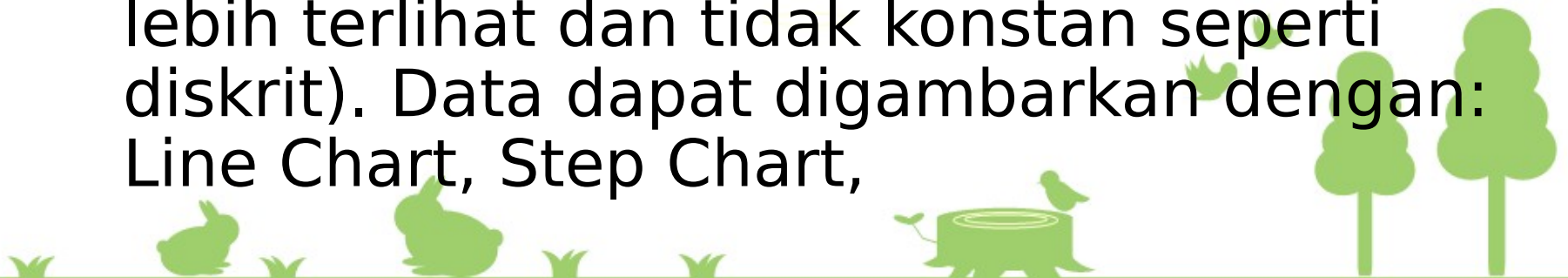
# Tell a Story

Membuat sebuah data untuk memiliki informasi yang lebih mudah untuk dibaca dan juga memiliki informasi yang lebih humanis.



# Visualizing Pattern over Time

- Discrete Temporal Data = Statis, tidak berkelanjutan dan dalam satu kurun waktu saja. Data dapat digambarkan dengan: Bar Chart, dan Point Chart
- Continuous Temporal Data = Data berkelanjutan, dapat dilihat progress data dalam jangka waktu tertentu (perubahan lebih terlihat dan tidak konstan seperti diskrit). Data dapat digambarkan dengan: Line Chart, Step Chart,

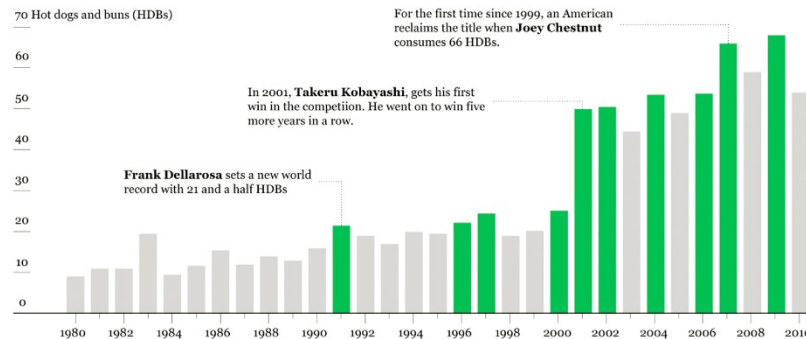


# Bar Chart

- Untuk menunjukkan perbandingan dari dua aspek (X dan Y)
- Umumnya digunakan untuk perbandingan waktu atau besaran konstan lainnya

## HOT DOG EATING

Nathan's hot dog eating contest every July 4th has been going on since the early 1900s, but it wasn't until 2001 when things got serious. Takeru Kobayashi from Japan raised the bar, more than doubling the previous world record. Highlighted bars indicate new records.



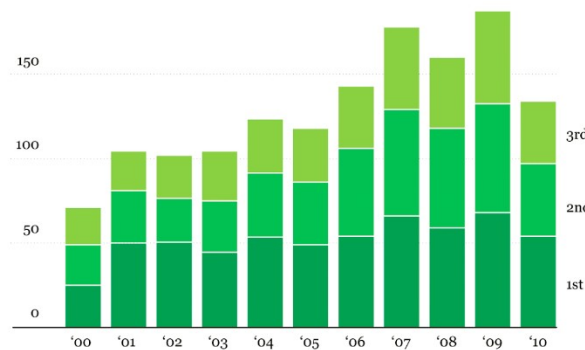
# Stacked Bar Chart

- Untuk menunjukkan perbandingan dari dua aspek (X dan Y)
- Serupa dengan bar chart, namun bar memiliki lapisan yang menjelaskan sub kategori

## TOP THREE HOT DOG EATERS

The year before Takeru Kobayashi started to compete in Nathan's Hot Dog Eating Contest, the top three eaters were close in skills. However, from 2001 to 2005, Kobayashi always had a substantial lead. That changed in 2006 when Joey Chestnut started competing.

200 Hot dogs and buns (HDBs)



Source: Wikipedia | Nathan Yau

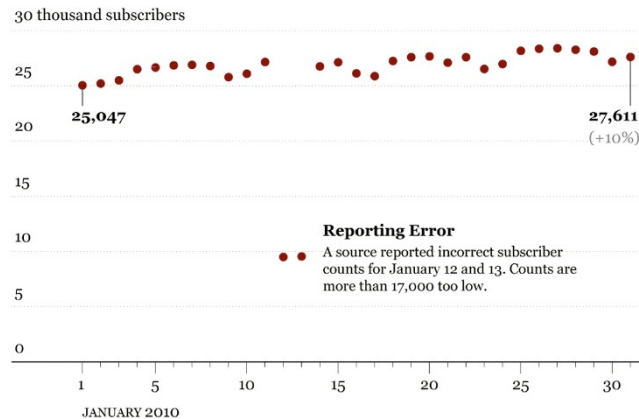


# Titik (Points)

- Untuk menunjukkan perbandingan dari dua aspek (X dan Y)
- Data bersifat padat, sehingga bar digantikan dengan titik

## INCREASE IN SUBSCRIBERS

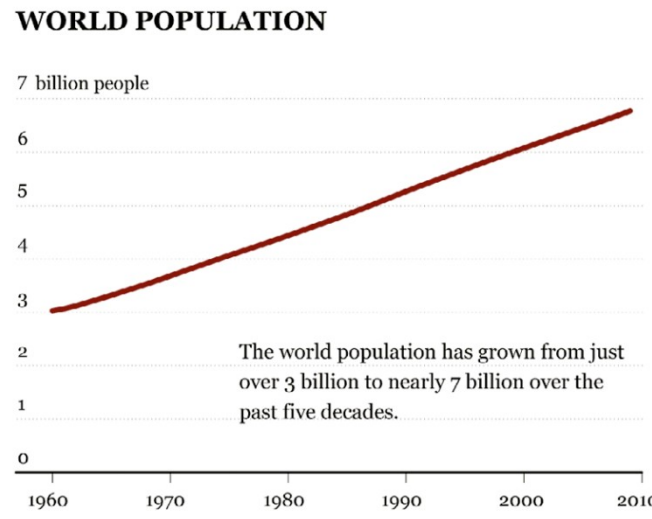
In January 2010, the number of subscribers via RSS and email increase to 27,611, making it the tenth month in a row with at least a ten percent increase.



Source: Feedburner | Nathan Yau

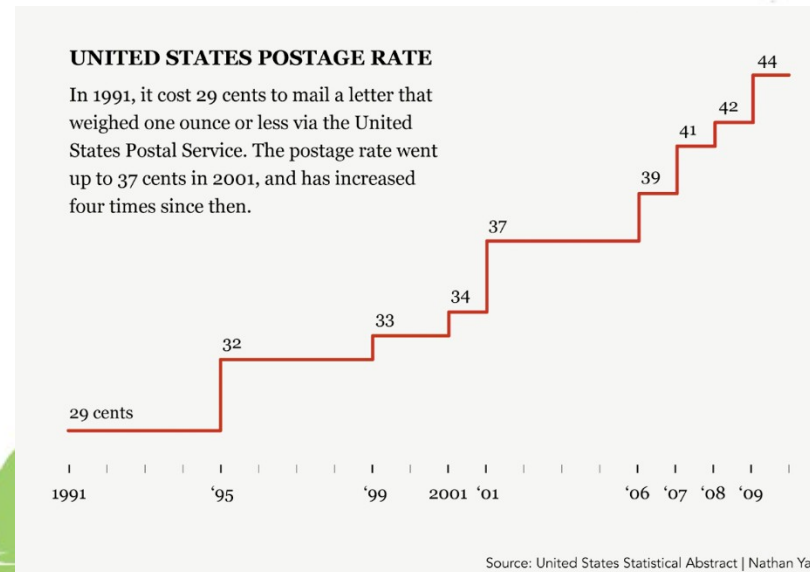
# Line Chart

- Pengembangan dari point chart yang dapat menunjukkan perubahan secara jelas
- Setiap titik dihubungkan dengan garis untuk menjelaskan proses perubahan



# Step Chart

- Digunakan untuk data yang statis naik/turun perkembangannya.
- Modifikasi dari line chart



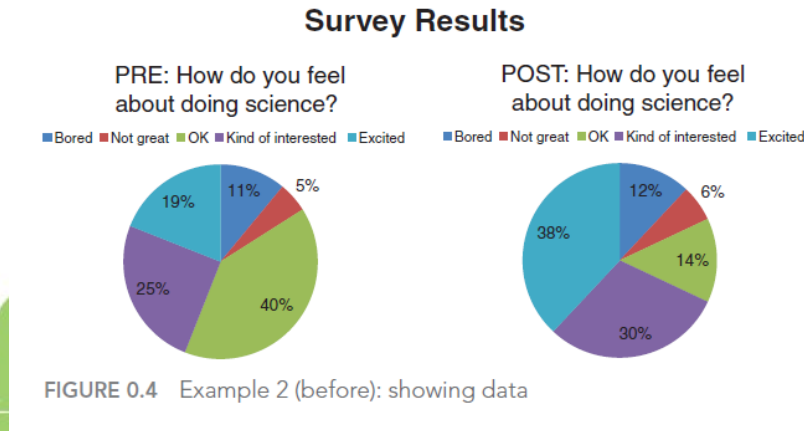
# Visualizing Proportion

- Untuk menjelaskan besaran porsi dari satu hal mutlak
- Contoh chart yang dapat digunakan adalah = Pie Chart dan Area Chart



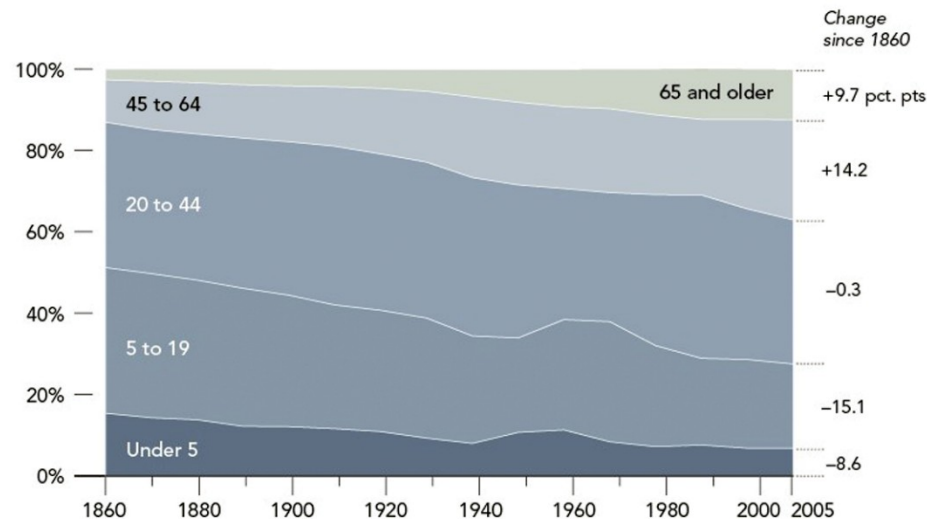
# Pie Chart

- Untuk data yang menunjukkan persentase atau besaran yang dimiliki suatu hal
- Cocok digunakan untuk menggambarkan perbandingan dari suatu hal yang memiliki proporsi numerik



# Area Chart

- Menunjukkan proporsi dalam satuan waktu
- Modifikasi dari line chart yang menggambarkan proporsi data



# Visualizing Relationship

- Pola yang menggambarkan relasi antar data
- Menunjukkan pengaruh suatu data terhadap data lainnya
- Contoh chart yang dapat digunakan:  
Scatterplot, Bubble Chart



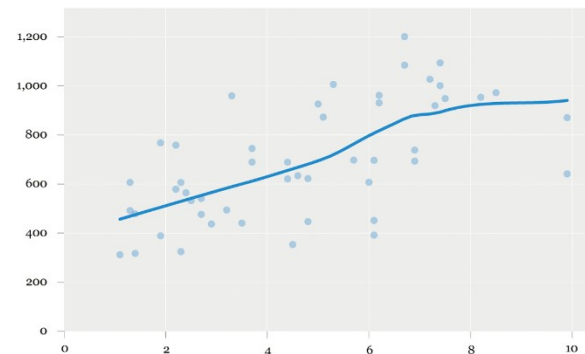
# Scatterplot

- Implementasi dari dua data pada line chart dan point chart
- Umumnya kedua jenis data saling berkaitan untuk mengetahui dampak data satu sama lain

**MURDERS VERSUS BURGLARIES IN THE UNITED STATES**

States with higher murder rates tend to have higher burglary rates.

**Burglaries**  
per 100,000 population



**Murders**  
per 100,000 population

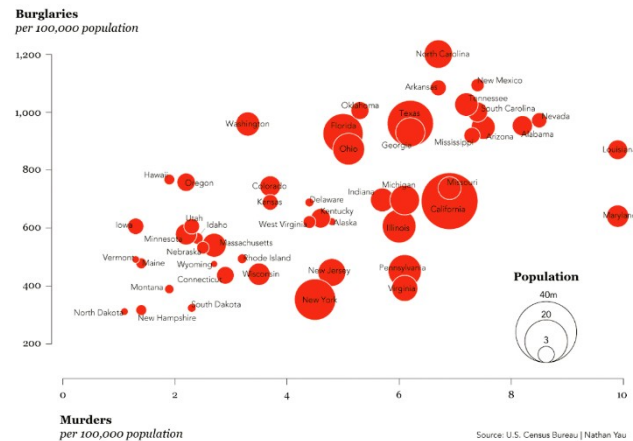
Source: U.S. Census Bureau | Nathan Yau



# Bubble Chart

- Modifikasi dari point chart
- Dapat digunakan untuk komparasi 3 hal sekaligus, dengan sumbu X, sumbu Y, dan besarnya titik

MURDERS VERSUS BURGLARIES IN THE UNITED STATES



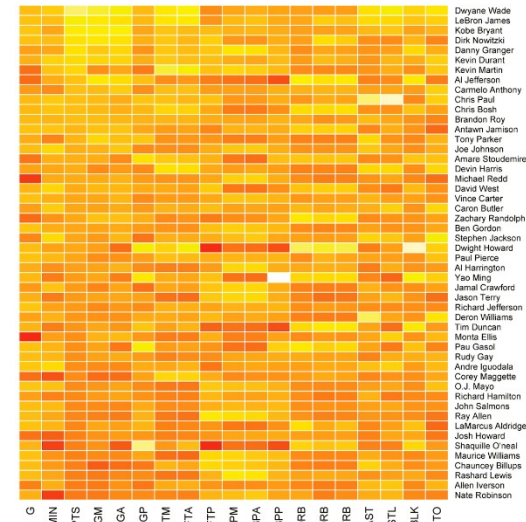
# Spotting Differences

- Untuk memperlihatkan perbedaan kondisi antar masing-masing data
- Contoh chart yang dapat digunakan: Heatmap, Chernoff Faces, Star Chart



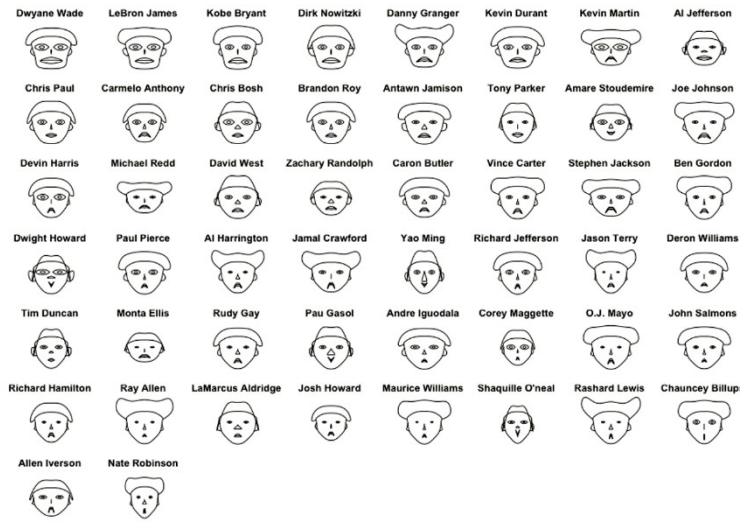
# Heatmap

- Menggambarkan kondisi perbedaan antar masing-masing data dengan warna
- Data digambarkan dari warna, dimana warna dapat ditinjau dari kecerahan maupun kegelapan



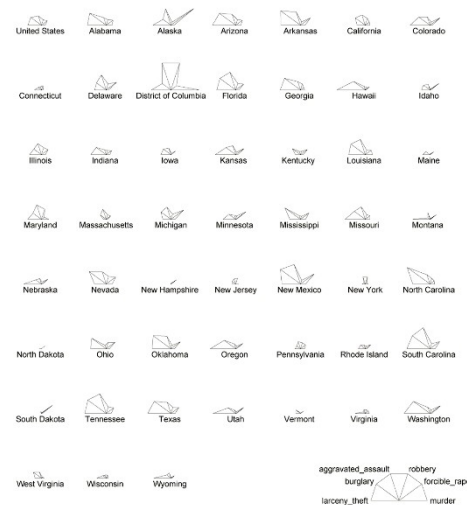
# Chernoff Faces

- Menggambarkan variasi data dari bentuk wajah yang dihasilkan
- Beberapa faktor pembeda dapat dilihat dari atribut wajah, seperti mata, hidung, mulut, dsb



# Star Chart

- Serupa dengan Chernoff Faces, namun menggunakan bentuk bangun datar yang menggambarkan data
- Bangun datar dapat ditinjau dari ukuran, sudut, bentuk, dsb



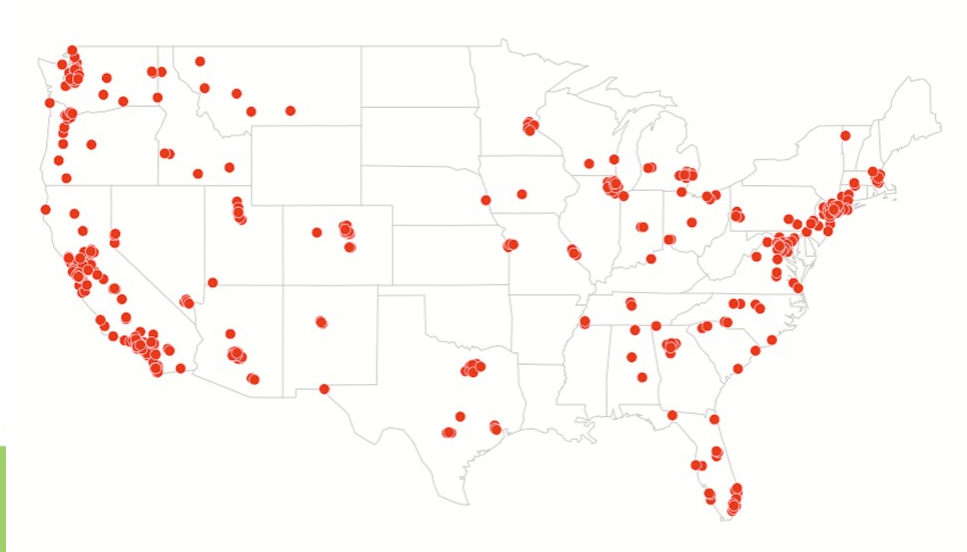
# Visualizing Spatial Relationship

- Menggambarkan besaran spasial yang berkaitan antar data
- Umumnya menggunakan peta sebagai acuan data



# Map with Dots

- Penggunaan titik pada peta yang dapat digunakan untuk menjelaskan lokasi atau keberadaan suatu data pada peta
- Dapat menunjukkan keberadaan serta kepadatan berdasarkan titik



# Map with Lines

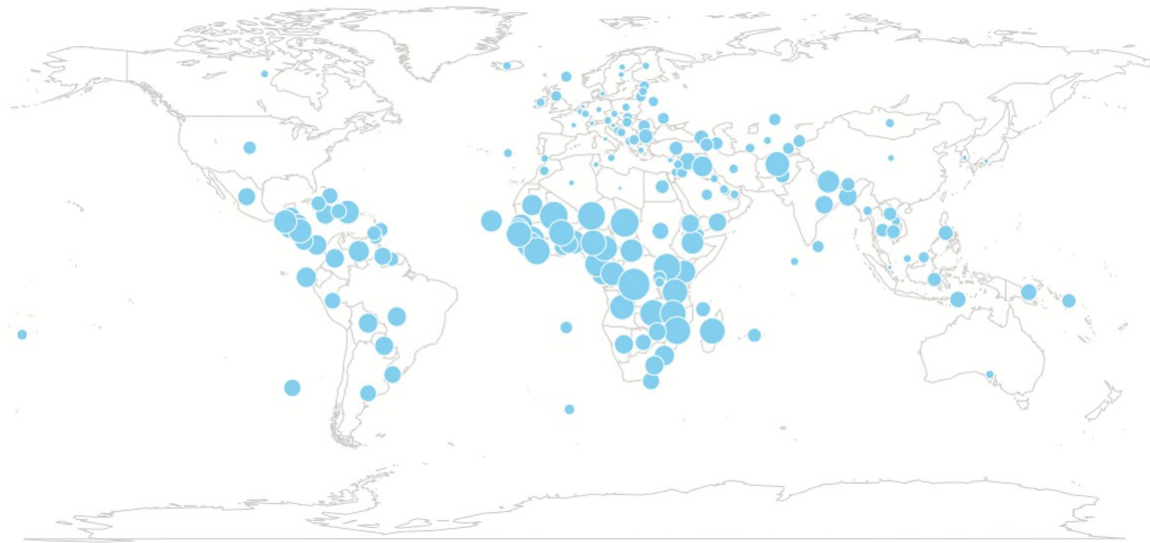
- Dapat menggambarkan jarak antar titik dari Map with Dots
- Digunakan untuk identifikasi jarak, rute, koneksi, dsb





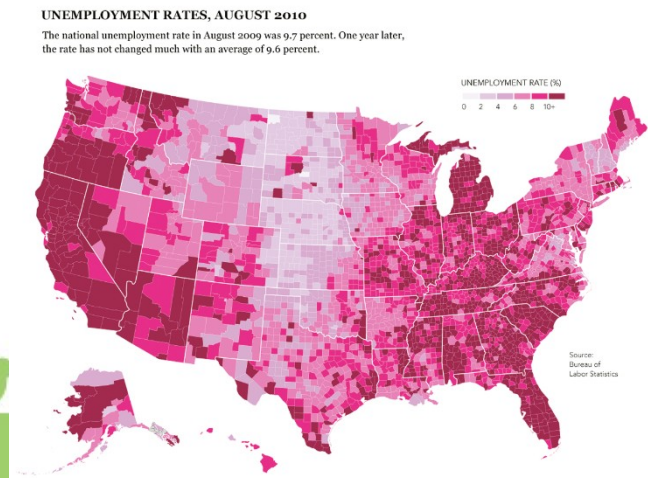
# Map with Bubbles

- Sama seperti Map with Dots, namun titik menggambarkan besaran data pada lokasi tersebut



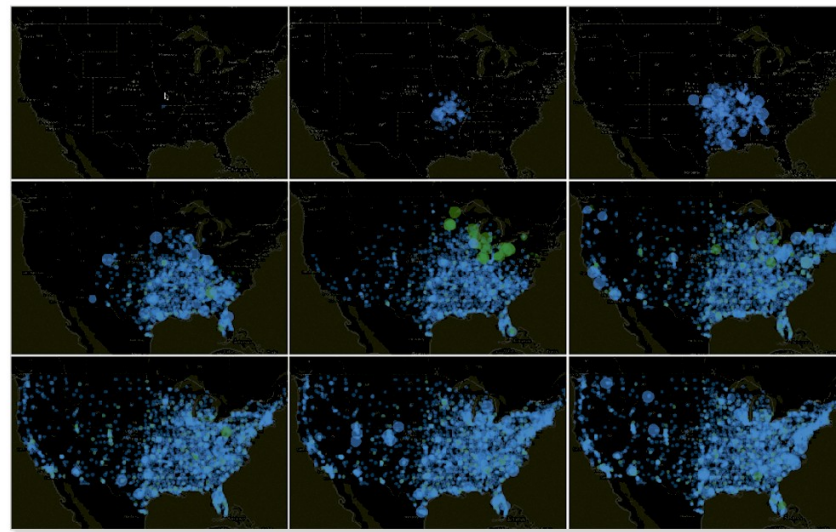
# Map with Regions

- Menggambarkan data sebagai bagian dari peta
- Data digambarkan dari pemecahan peta menggunakan garis/warna yang menggambarkan data



# Map over Space and Time

- Menunjukkan perkembangan data dalam kurun waktu tertentu
- Umumnya menggunakan Map with Region dengan jumlah yang lebih dari satu



# Designing with a Purpose

- Prepare yourself
- Prepare for readers
- Visual Cues
- Good Visualization



# Tugas 1: Aneka Visualisasi

Ikuti tutorial di:

<https://www.geeksforgeeks.org/data-visualization-on-different-charts-python/>



# Sample Code To Visualize Data

- Membuat histogram

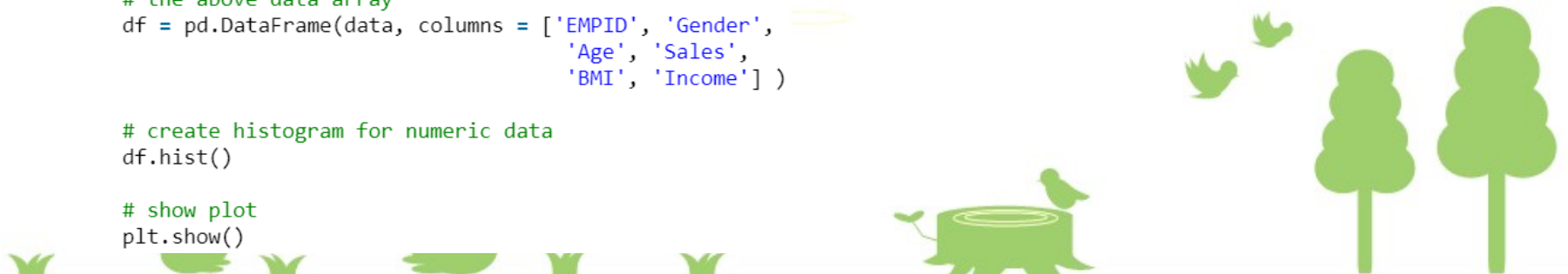
```
# import pandas and matplotlib
import pandas as pd
import matplotlib.pyplot as plt

# create 2D array of table given above
data = [['E001', 'M', 34, 123, 'Normal', 350],
        ['E002', 'F', 40, 114, 'Overweight', 450],
        ['E003', 'F', 37, 135, 'Obesity', 169],
        ['E004', 'M', 30, 139, 'Underweight', 189],
        ['E005', 'F', 44, 117, 'Underweight', 183],
        ['E006', 'M', 36, 121, 'Normal', 80],
        ['E007', 'M', 32, 133, 'Obesity', 166],
        ['E008', 'F', 26, 140, 'Normal', 120],
        ['E009', 'M', 32, 133, 'Normal', 75],
        ['E010', 'M', 36, 133, 'Underweight', 40] ]

# dataframe created with
# the above data array
df = pd.DataFrame(data, columns = ['EMPID', 'Gender', 'Age', 'Sales', 'BMI', 'Income' ])

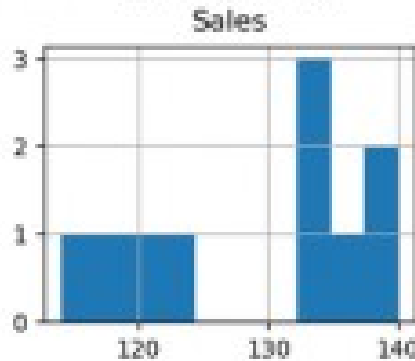
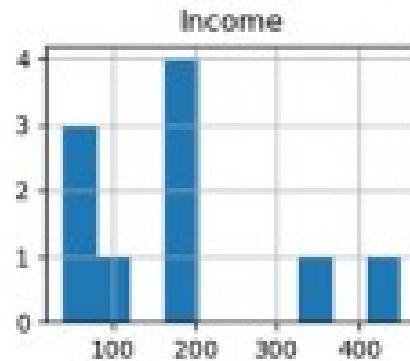
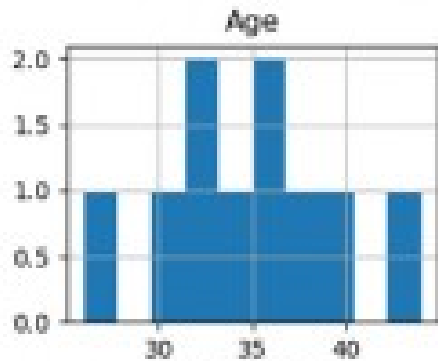
# create histogram for numeric data
df.hist()

# show plot
plt.show()
```



# Sample Code To Visualize Data

- Membuat histogram



# Sample Code To Visualize Data

- Membuat column chart

```
# Dataframe of previous code is used here
```

```
# Plot the bar chart for numeric values  
# a comparison will be shown between  
# all 3 age, income, sales  
df.plot.bar()
```

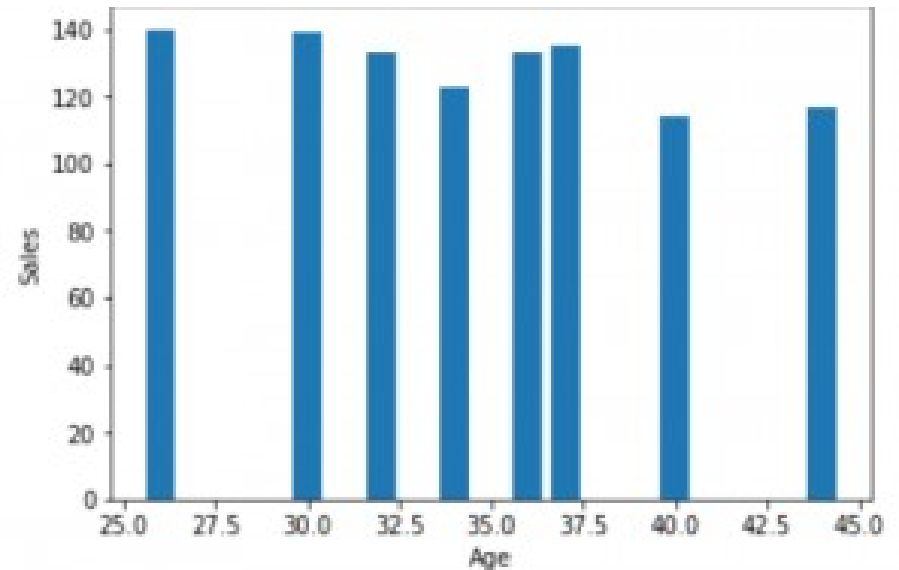
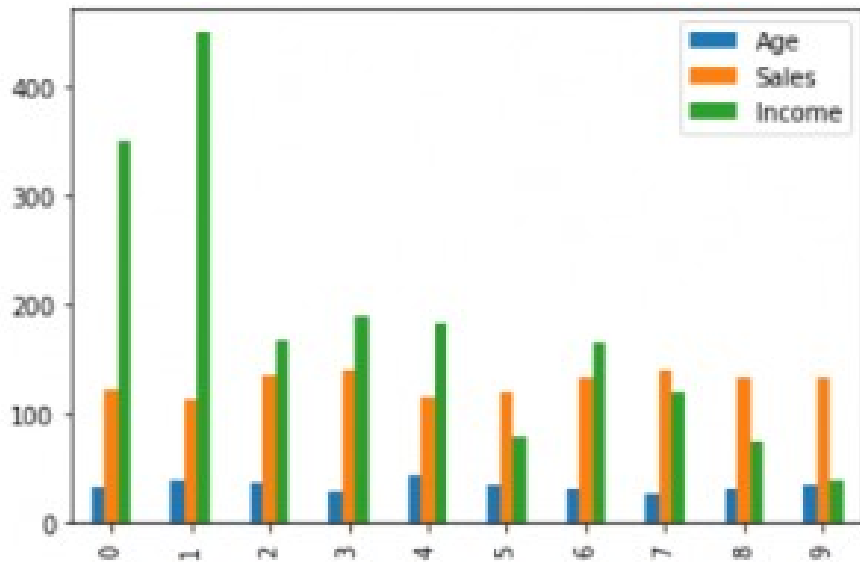
```
# plot between 2 attributes  
plt.bar(df['Age'], df['Sales'])  
plt.xlabel("Age")  
plt.ylabel("Sales")  
plt.show()
```





# Sample Code To Visualize Data

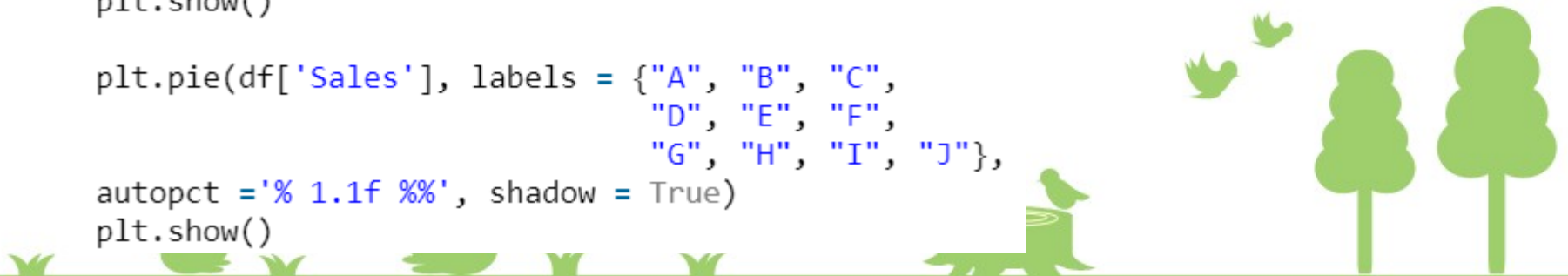
- Membuat column chart



# Sample Code To Visualize Data

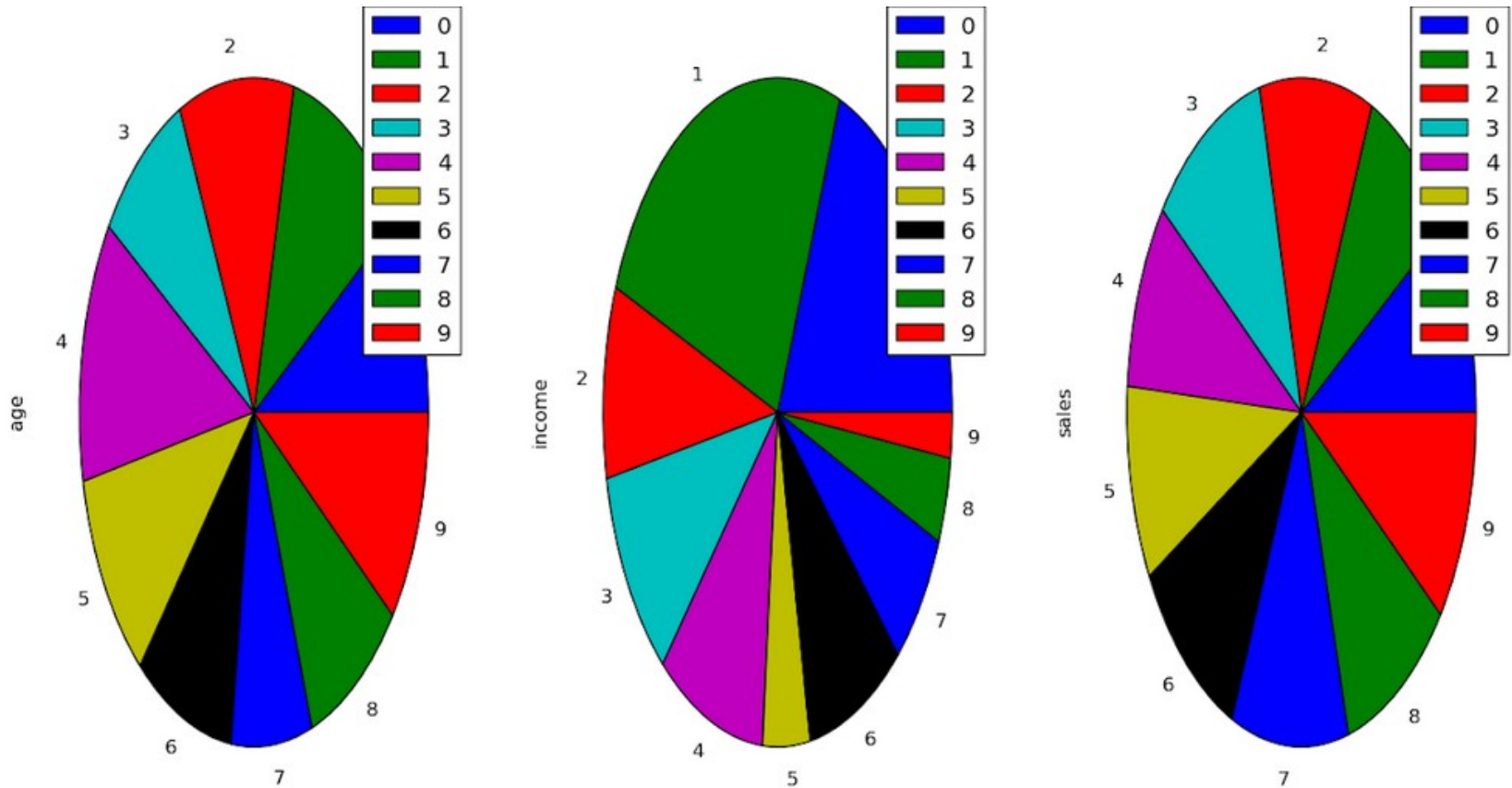
- Membuat pie chart

```
plt.pie(df['Age'], labels = {"A", "B", "C",  
                             "D", "E", "F",  
                             "G", "H", "I", "J"},  
  
autopct = '% 1.1f %%', shadow = True)  
plt.show()  
  
plt.pie(df['Income'], labels = {"A", "B", "C",  
                                "D", "E", "F",  
                                "G", "H", "I", "J"},  
  
autopct = '% 1.1f %%', shadow = True)  
plt.show()  
  
plt.pie(df['Sales'], labels = {"A", "B", "C",  
                               "D", "E", "F",  
                               "G", "H", "I", "J"},  
  
autopct = '% 1.1f %%', shadow = True)  
plt.show()
```



# Sample Code To Visualize Data

- Membuat pie chart



# Sample Code To Visualize Data

- Membuat scatter plot

```
# scatter plot between income and age  
plt.scatter(df['income'], df['age'])  
plt.show()
```

```
# scatter plot between income and sales  
plt.scatter(df['income'], df['sales'])  
plt.show()
```

```
# scatter plot between sales and age  
plt.scatter(df['sales'], df['age'])  
plt.show()
```



# Sample Code To Visualize Data

- Membuat scatter plot

