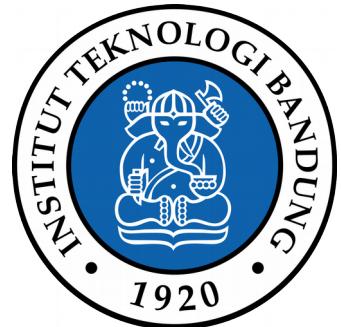
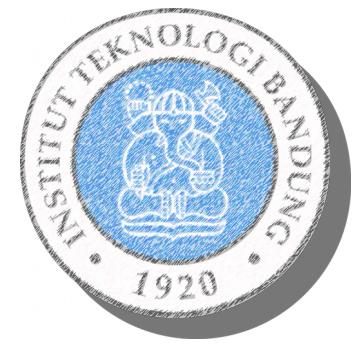


# Visualisasi Data (tool: Python)

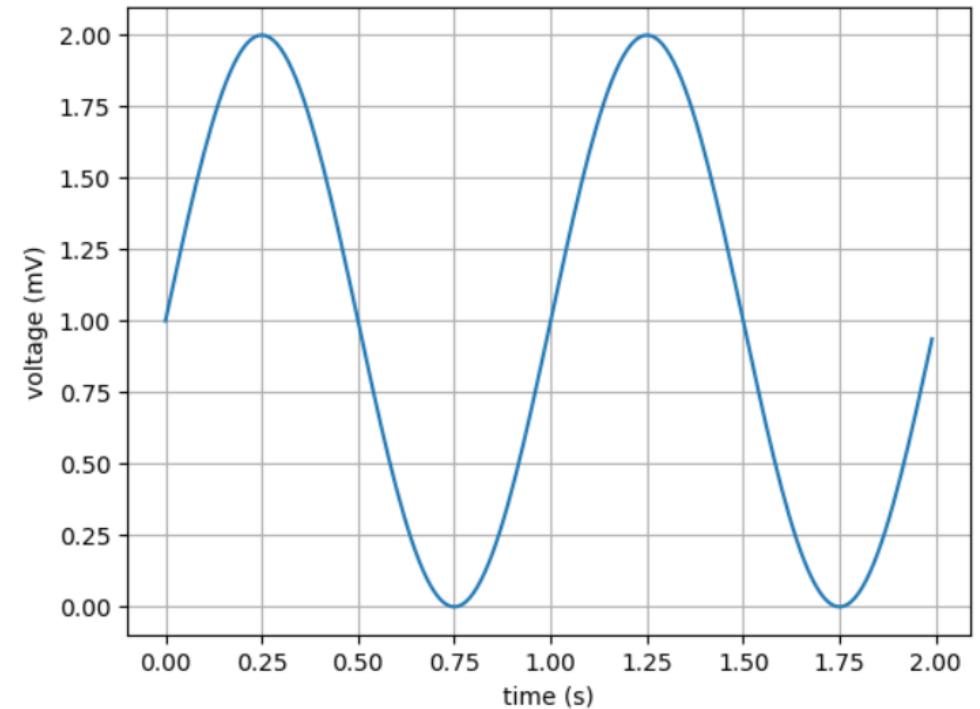
Tim Penyusun Materi Pengenalan Teknologi Informasi  
Institut Teknologi Bandung © 2018





# Matplotlib

- **Matplotlib** adalah library di Python yang digunakan untuk plotting grafik 2 dimensi dalam berbagai format
  - **pyplot** adalah modul untuk plotting sederhana dengan interface yang mirip seperti MATLAB
- Dikembangkan oleh John Hunter (1968-2012) dan rekan-rekan, merupakan salah satu library yang banyak dimanfaatkan untuk



# Contoh data-1

- data.csv
- Load pandas dan dataframe dari file data.csv

```
import pandas as pd  
df = pd.read_csv("D:/data.csv")
```

- Load library matplotlib.pyplot
- Load library matplotlib.pyplot

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

	name	age	gender	state	num_children	num_pets
0	john	23	M	CA	2	5
1	marry	78	F	DC	0	1
2	peter	22	M	CA	0	0
3	jeff	19	M	DC	3	5
4	bill	45	M	CA	2	2
5	lisa	33	F	TX	1	2
6	jose	20	M	TX	4	3

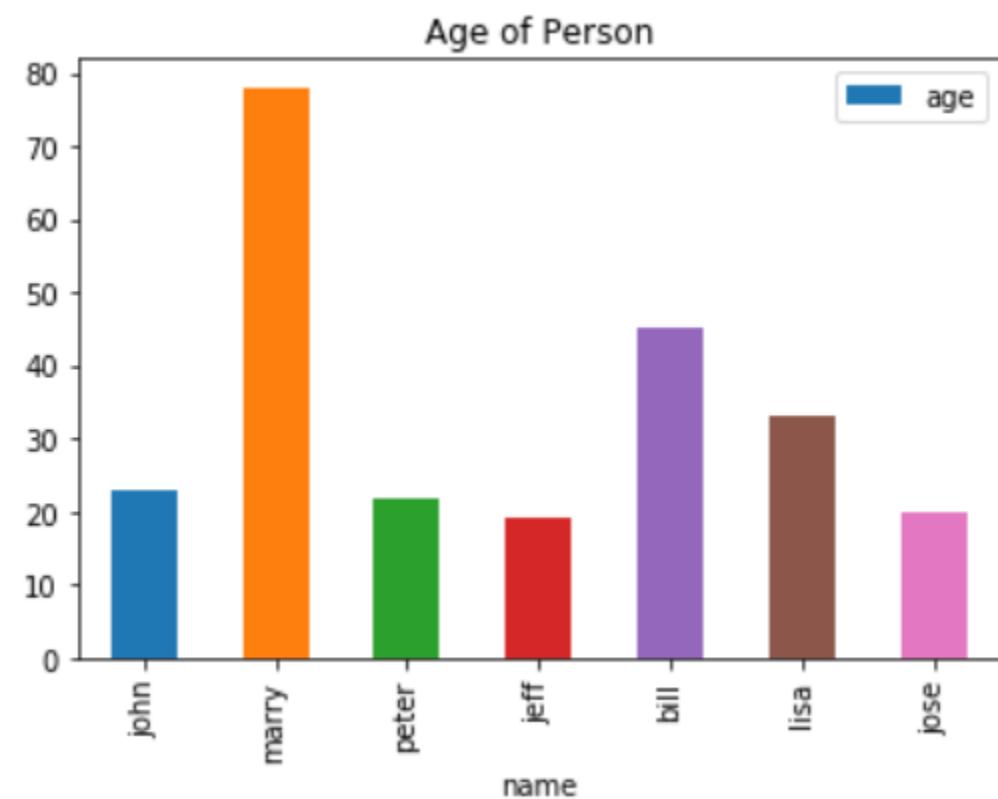
Lebih lanjut fungsi plot pada dataframe:  
<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.plot.html>

# Bar chart (1)

- Buat vertical bar chart untuk menampilkan umur dari setiap orang

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

```
df = pd.read_csv("D:/data.csv")  
df.plot(kind="bar", x="name", y="age", title="Age of Person")  
plt.show()
```

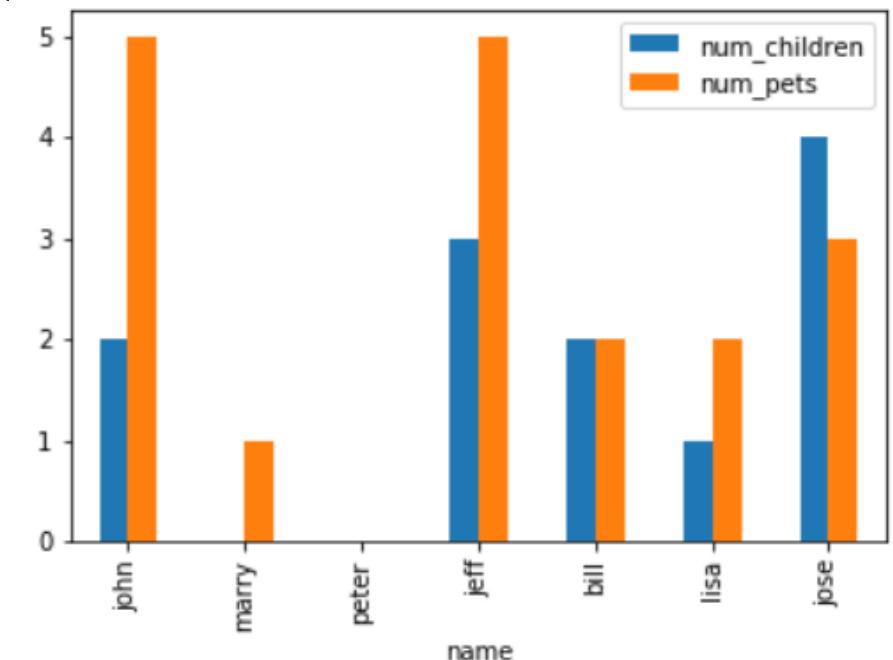




# Bar Chart (2)

- Tampilkan banyaknya anak (num\_children) dan banyaknya piaraan (num\_pets) dalam 1 grafik vertical bar chart

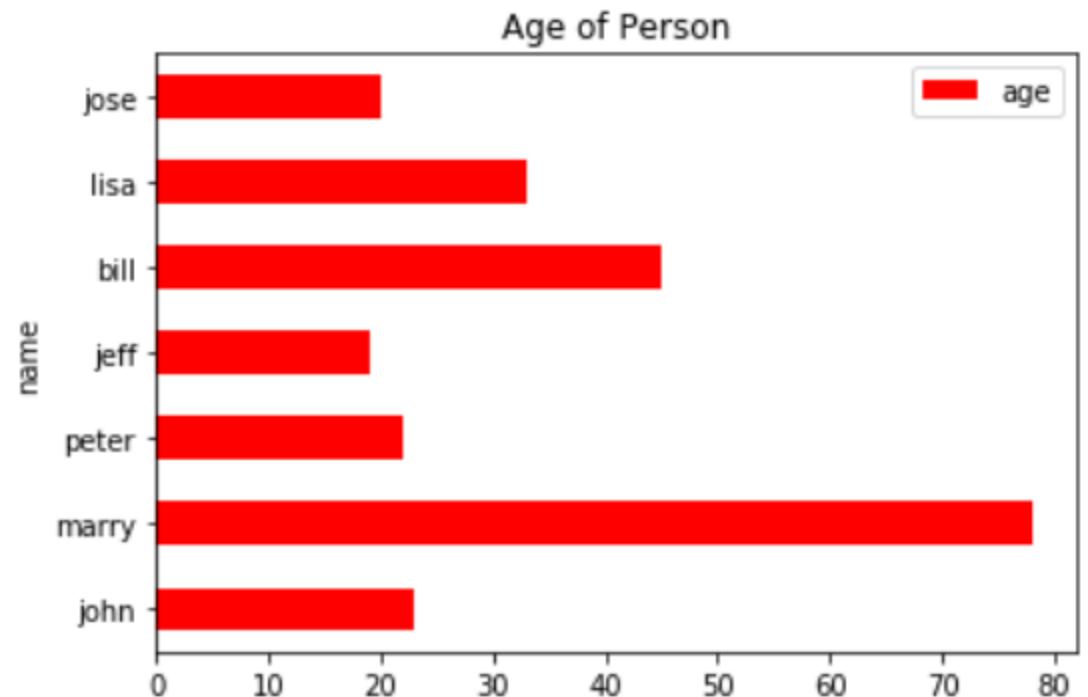
```
df.plot(kind="bar", x="name", y=["num_children", "num_pets"])
plt.show()
```

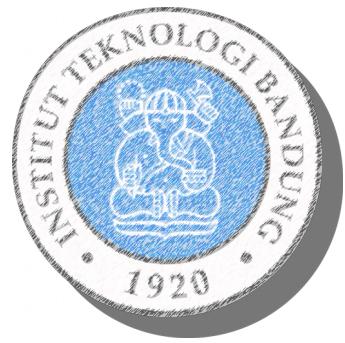




# Horizontal Bar Chart

```
df.plot(kind="barh", x="name", y="age", title="Age of Person",  
color="red")  
plt.show()
```

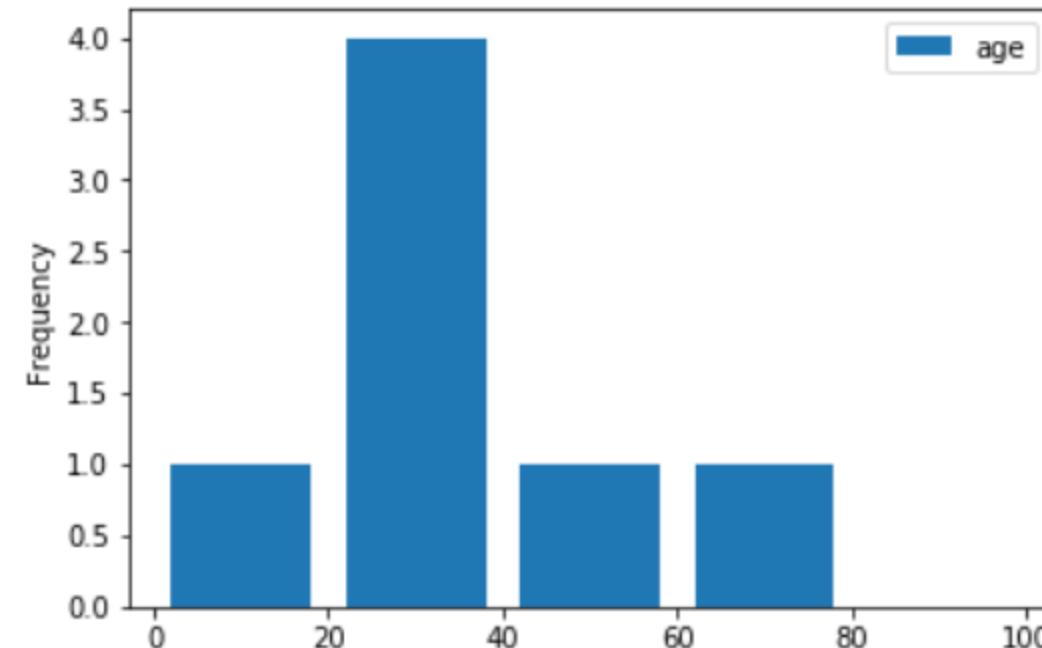




# Histogram

- Tampilkan distribusi orang berdasarkan kelompok umur:  
0-20; 21-40; 41-60; 61-80; 81-100

```
df[["age"]].plot(kind="hist", bins=[0, 20, 40, 60, 80, 100], rwidth=0.8)  
plt.show()
```

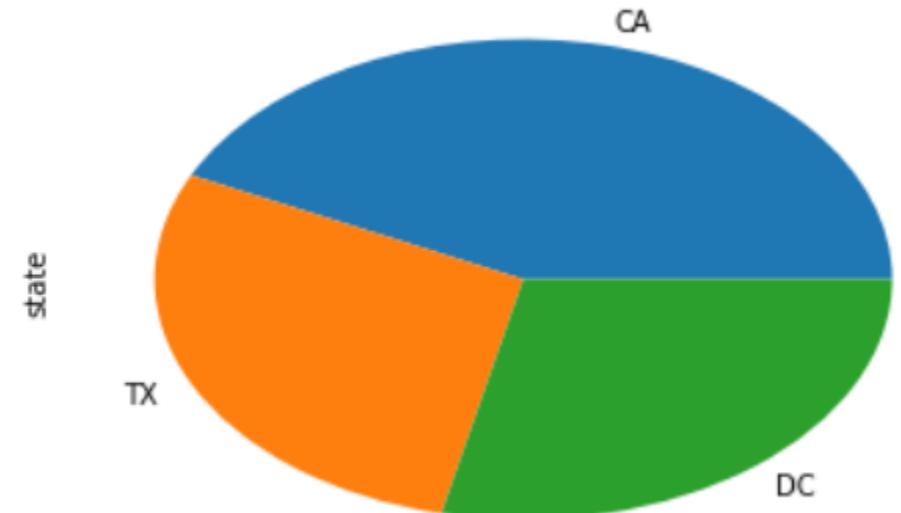


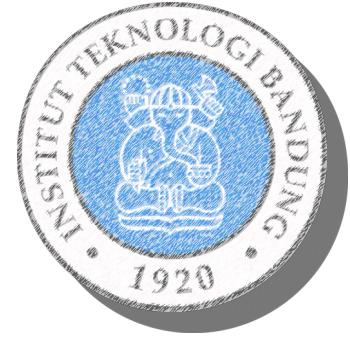


# Pie Chart

- Tampilkan komposisi banyaknya orang berdasarkan negara
- Langkah:
  - Hitung distribusi frekuensi (lihat bahan kuliah sebelumnya)
  - Plot ke pie chart

```
df2 = df["state"].value_counts()  
df2.plot(kind = "pie")  
plt.show()
```





# Stacked Bar Chart (1)

- Tampilkan data banyaknya data per jenis kelamin (gender) per negara bagian (state)
- Langkah:
  1. Membuat **tabel pivot**: menggunakan perintah group by, kelompokkan data terlebih dahulu

Misalnya: group by berdasarkan kolom gender, state, lalu count banyaknya data (misal berdasarkan kolom name)

```
df3 = df.groupby(["gender", "state"])["name"].size().unstack()  
df3
```

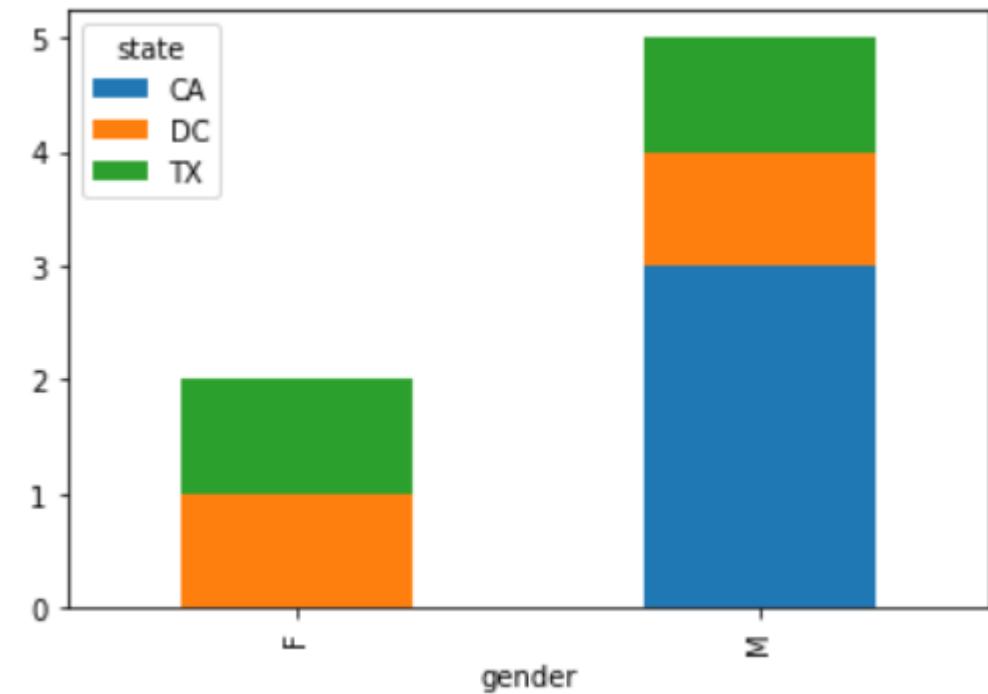
gender	CA	DC	TX
F	NaN	1.0	1.0
M	3.0	1.0	1.0



# Stacked Bar Chart (2)

- Langkah:  
2. Plot df3 ke bar chart dan stacked = True

```
df3.plot(kind="bar", stacked=True)  
plt.show()
```

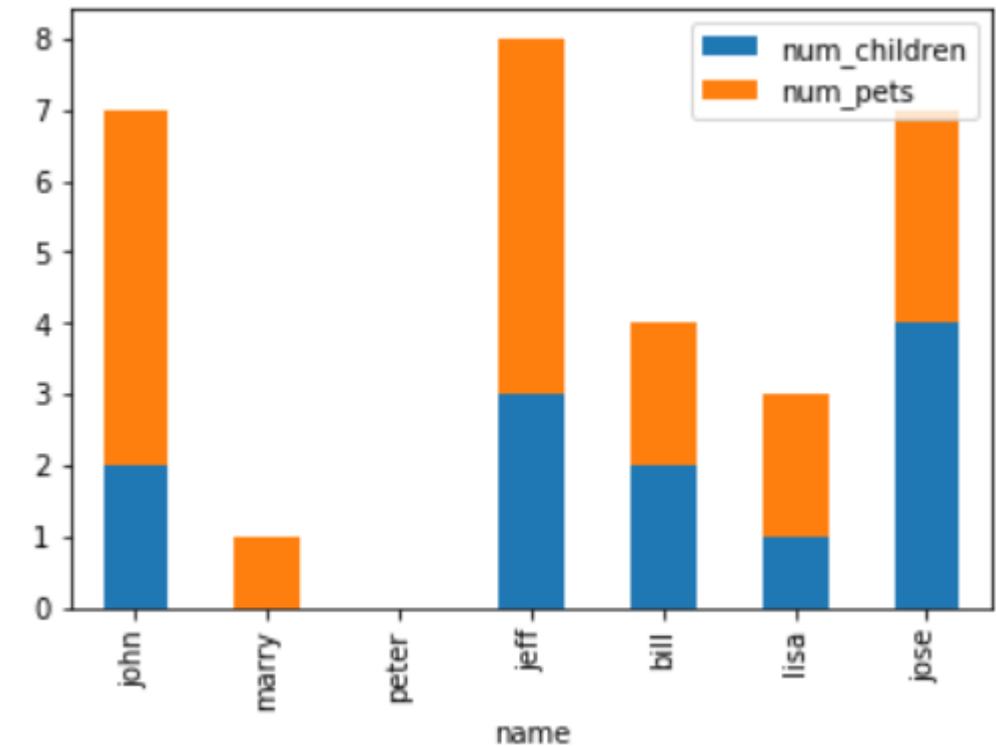




# Stacked Bar Chart (3)

- Tampilkan banyaknya anak (num\_children) dan banyaknya piaraan (num\_pets) dalam 1 grafik stacked bar chart

```
df.plot(kind = "bar", x = "name",
y=["num_children", "num_pets"],
stacked = True)
plt.show()
```





# Contoh Data-2: Time-series

- animal.csv
- Load data:

```
df1 = pd.read_csv("D:/animal.csv")
```

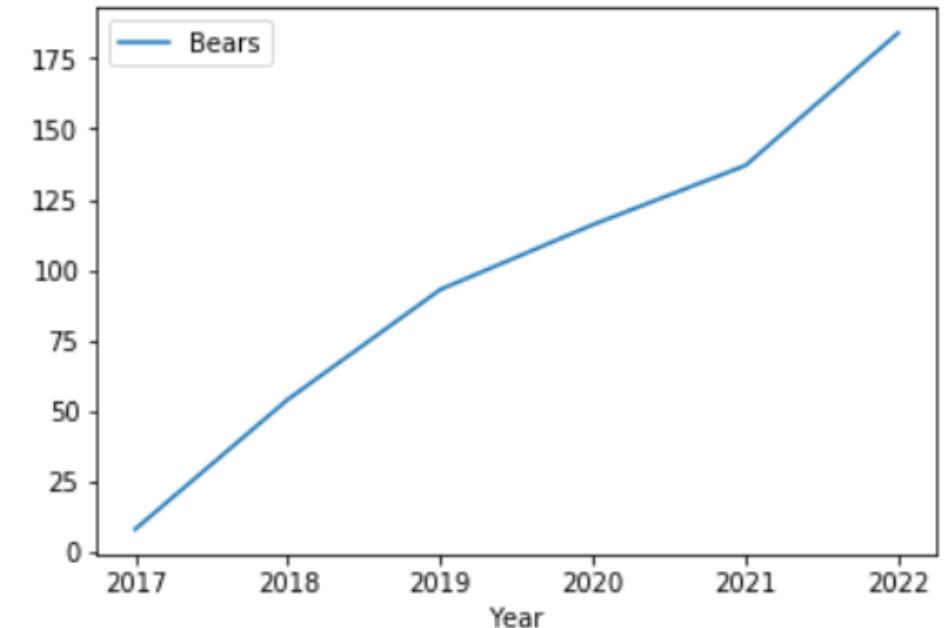
	Year	Bears	Dolphins	Whales
0	2017	8	150	80
1	2018	54	77	54
2	2019	93	32	100
3	2020	116	11	76
4	2021	137	6	93
5	2022	184	1	72

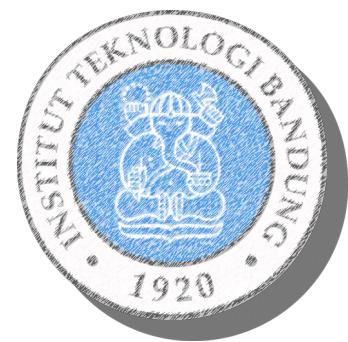


# Line Chart (1)

- Tampilkan pertumbuhan populasi beruang (Bears) dari tahun ke tahun dalam line chart

```
df1.plot(kind="line", x="Year", y="Bears")  
plt.show()
```

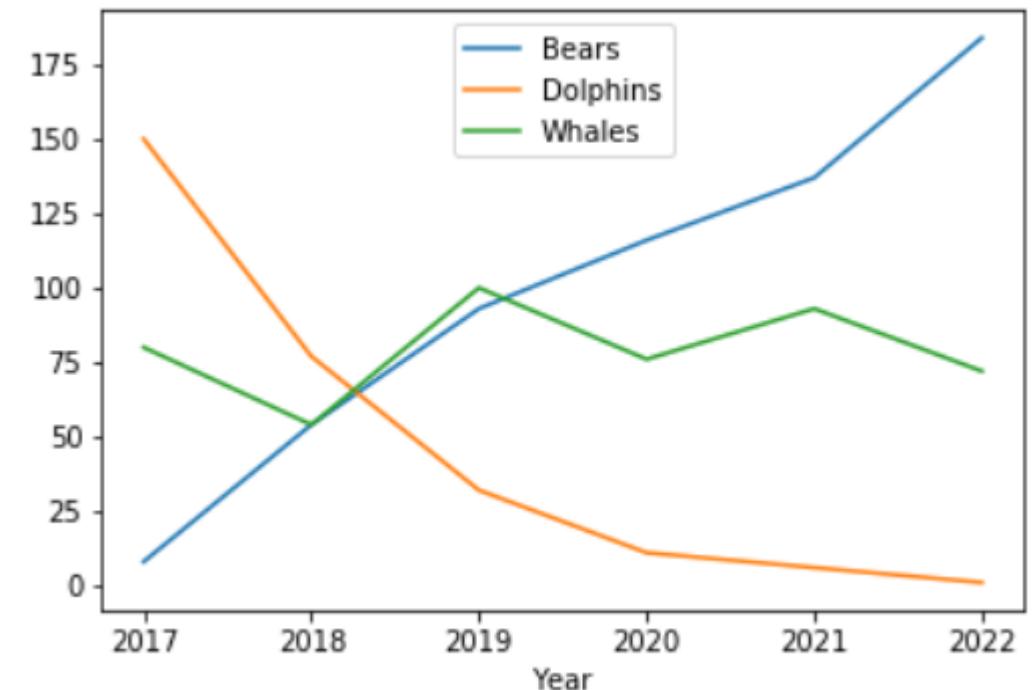




# Line Chart (2)

- Tampilkan pertumbuhan populasi beruang (Bears), lumba-lumba (Dolphins), dan ikan paus (Whales) dari tahun ke tahun dalam 1 line chart

```
df1.plot(kind="line", x="Year",  
y=["Bears", "Dolphins", "Whales"])  
plt.show()
```

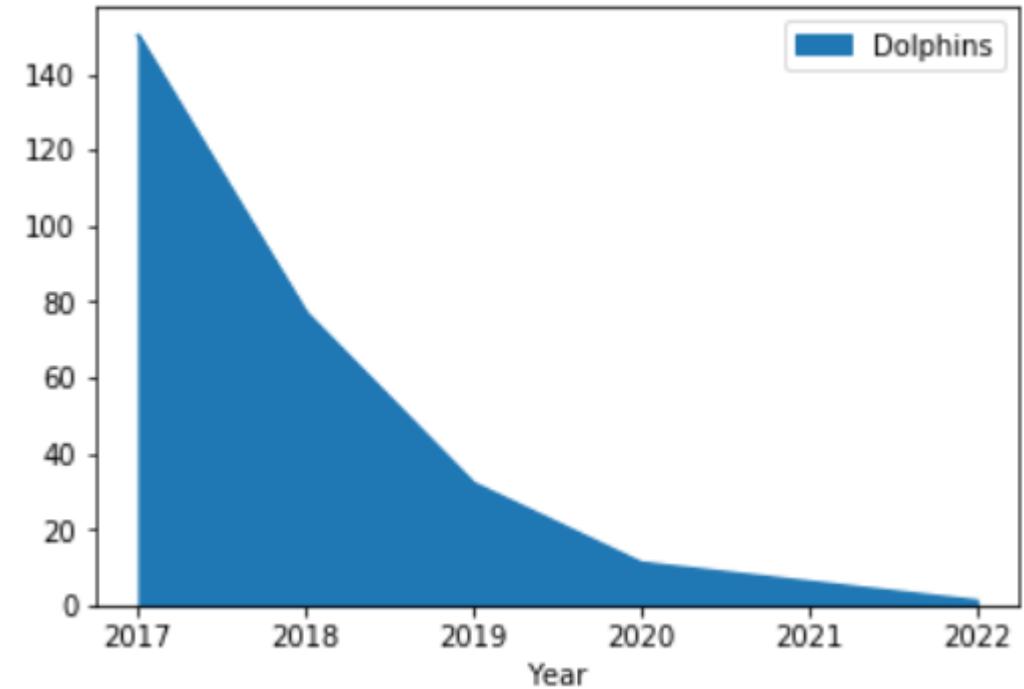




# Area Chart

- Tampilkan pertumbuhan populasi lumba-lumba (Dolphins) dari tahun ke tahun dalam area chart

```
df1.plot(kind="area", x="Year",  
y="Dolphins")  
plt.show()
```

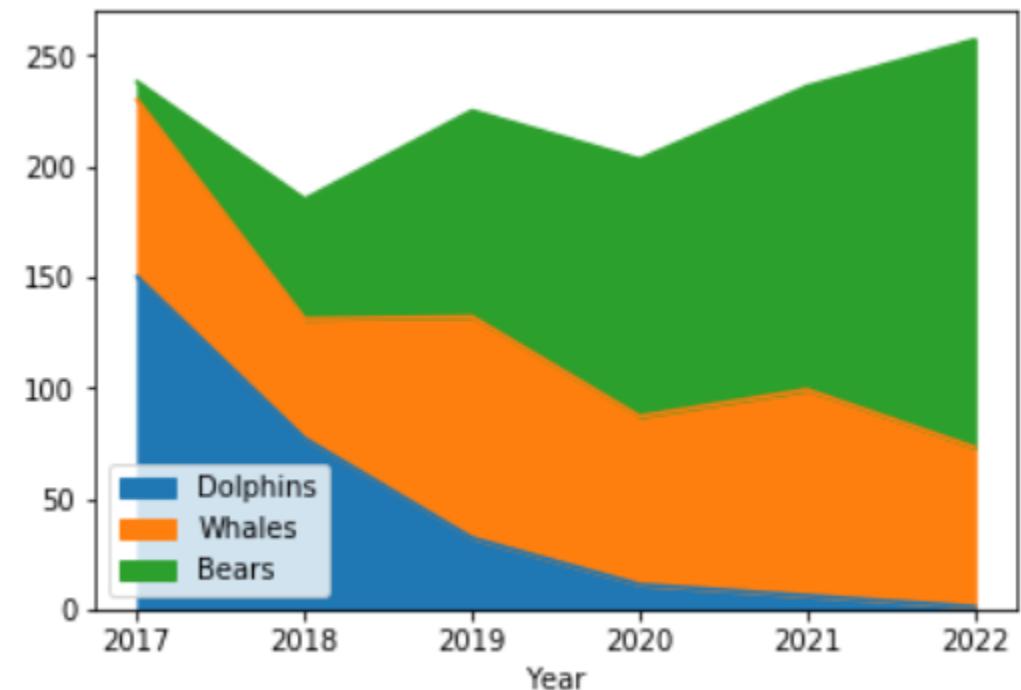


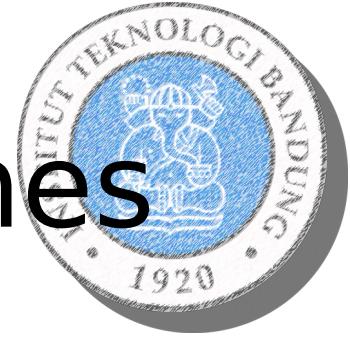


# Stacked Area Chart

- Tampilkan pertumbuhan populasi lumba-lumba (Dolphins), ikan paus (Whales), dan beruang (Bears), dari tahun ke tahun dalam stacked area chart

```
df1.plot(kind="area", x="Year",  
y=["Dolphins", "Whales", "Bears"])  
plt.show()
```





# Contoh Data-3: medali Asian Games

- Kembali ke contoh pada materi kuliah sebelumnya: file medali.csv
- Load data:

```
df4 = pd.read_csv("D:/medali.csv")
```



# Scatter Plot

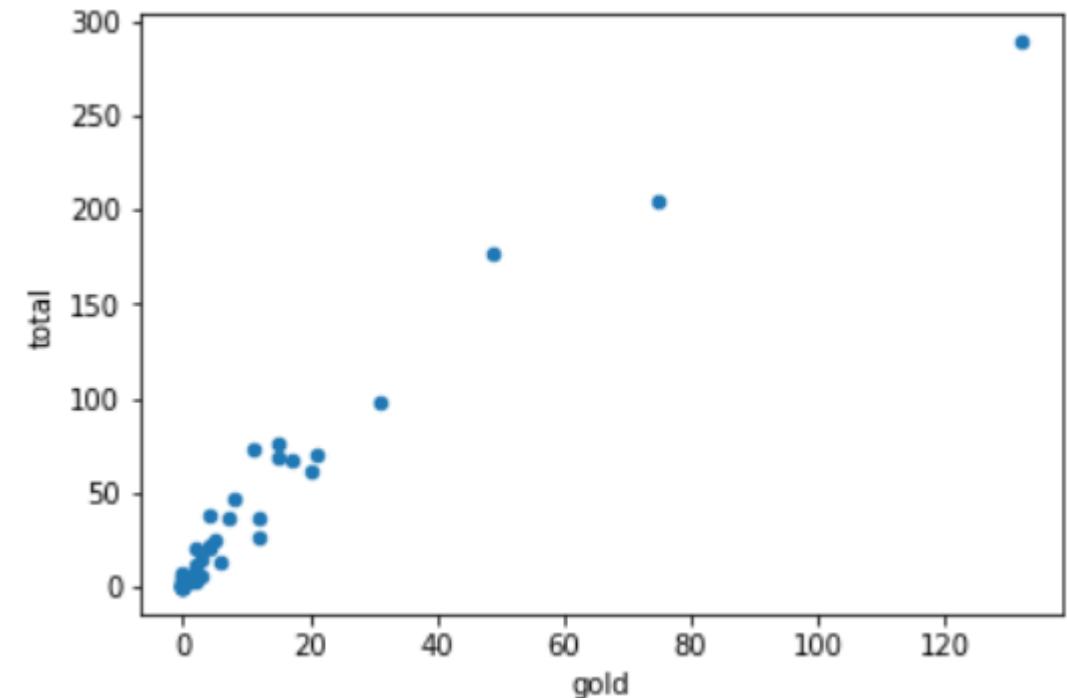
- Tampilkan relationship antara variable gold dan total dalam grafik scatter plot dan tunjukkan adanya korelasi positif

```
df4.plot(kind="scatter",  
x="gold", y="total")
```

Alternatif:

```
df4.plot.scatter(x="gold",  
y="total")
```

```
plt.show
```

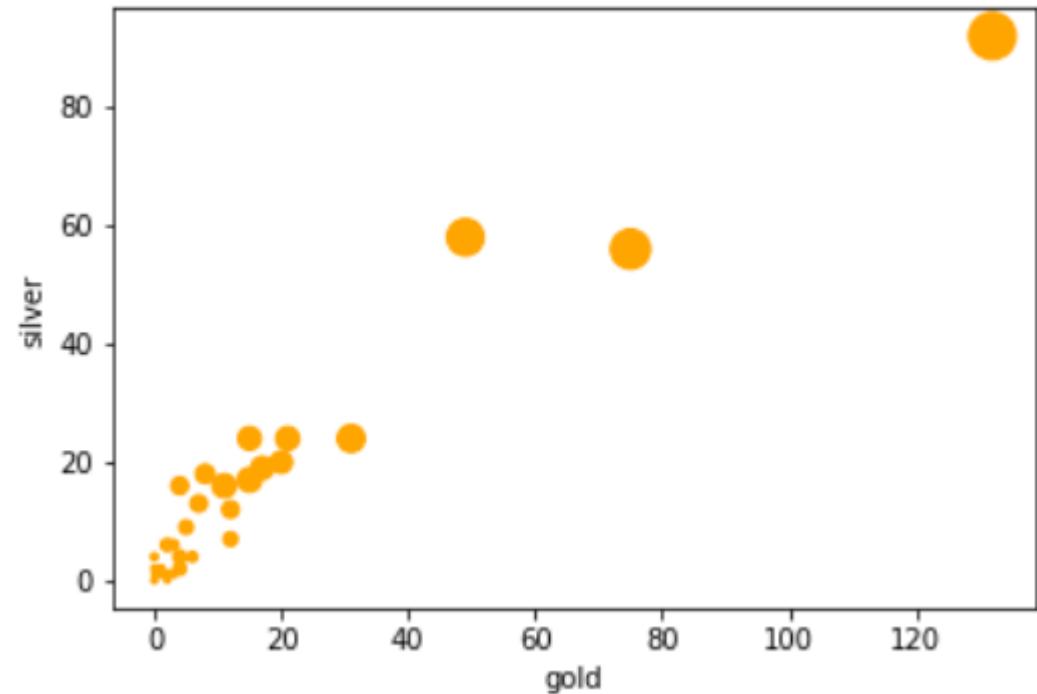


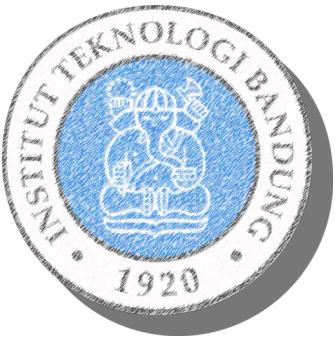


# Bubble Plot

- Tampilkan banyaknya total medali dikaitkan dengan perolehan nilai medali emas (gold) pada sumbu x dan perolehan medali perak (silver) pada sumbu y dalam grafik bubble plot
- Bubble plot di Python dibuat berdasarkan scatter plot

```
df4.plot(kind="scatter", x="gold",
y="silver", sizes=df4["total"],
color="orange")
plt.show()
```





# Menyimpan grafik ke file

```
df.plot(kind="bar", x="name", y="age", title="Age of Person")
plt.savefig("D:/agebarchart.png")
```