Training for Trainer (TOT)
Microcredential Certification
Bidang
Data Science dan Artificial Intelligence
2021

Program Fresh Graduate Academy Digital Talent Scholarship 2019 | Machine Learning

Pengantar Visualisasi Data dengan Python



Achmad Benny Mutiara



Bagian 1

Pendahuluan Visualisasi Data

Pendahuluan: Visualisasi Data

Alasan membuat visualisasi data:

1. Untuk melakukan eksplorasi pada analisis data
2. Mengkomunikasikan data dengan jelas
3. Representasi data agar tidak ada data yang bias
4. Digunakan untuk mendukung rekomendasi kepada para pemangku kepentingan (stakeholders) yang berbeda-beda

Pendahuluan: Visualisasi Data

Best Practice

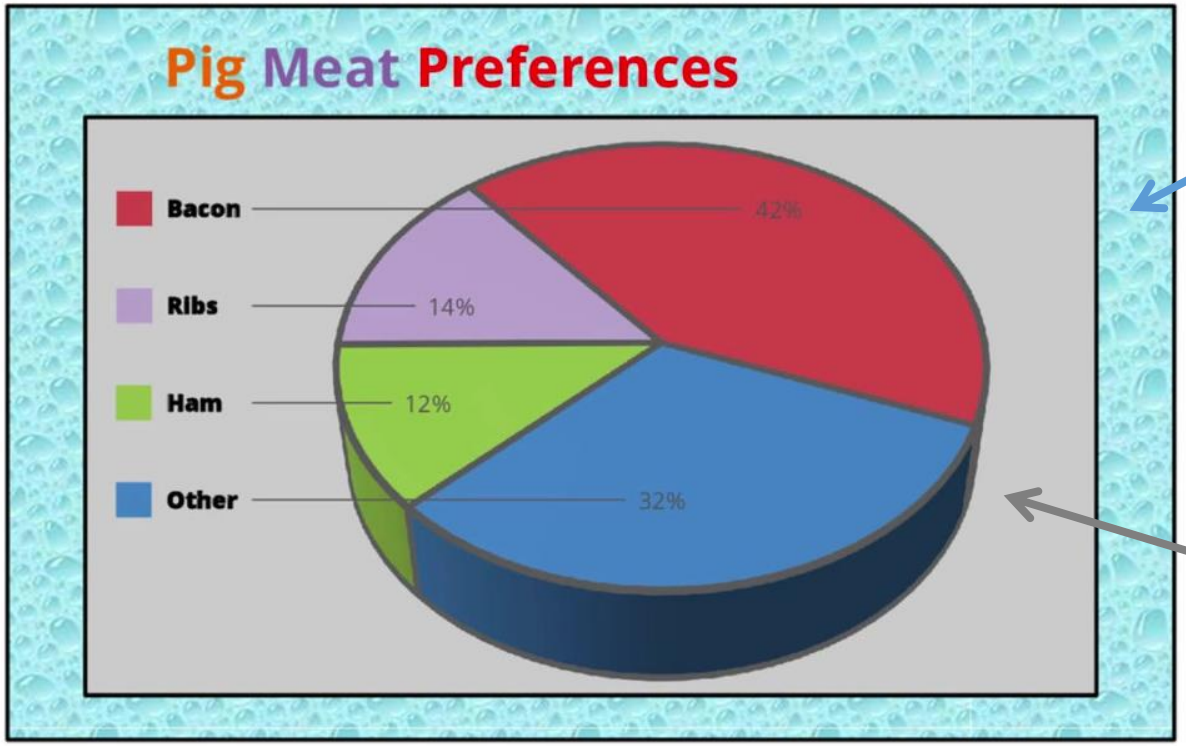
Beberapa hal yang perlu diingat pada saat membuat visualisasi

1. Visualisasi yang sederhana lebih efektif
2. Visualisasi yang sederhana lebih menarik
3. Visualisasi yang sederhana lebih memiliki dampak

Pendahuluan: Visualisasi Data

Contoh ke-1:

Diagram Pie 3D menggunakan 2 warna background pada gambar (warna biru dan abu-abu)



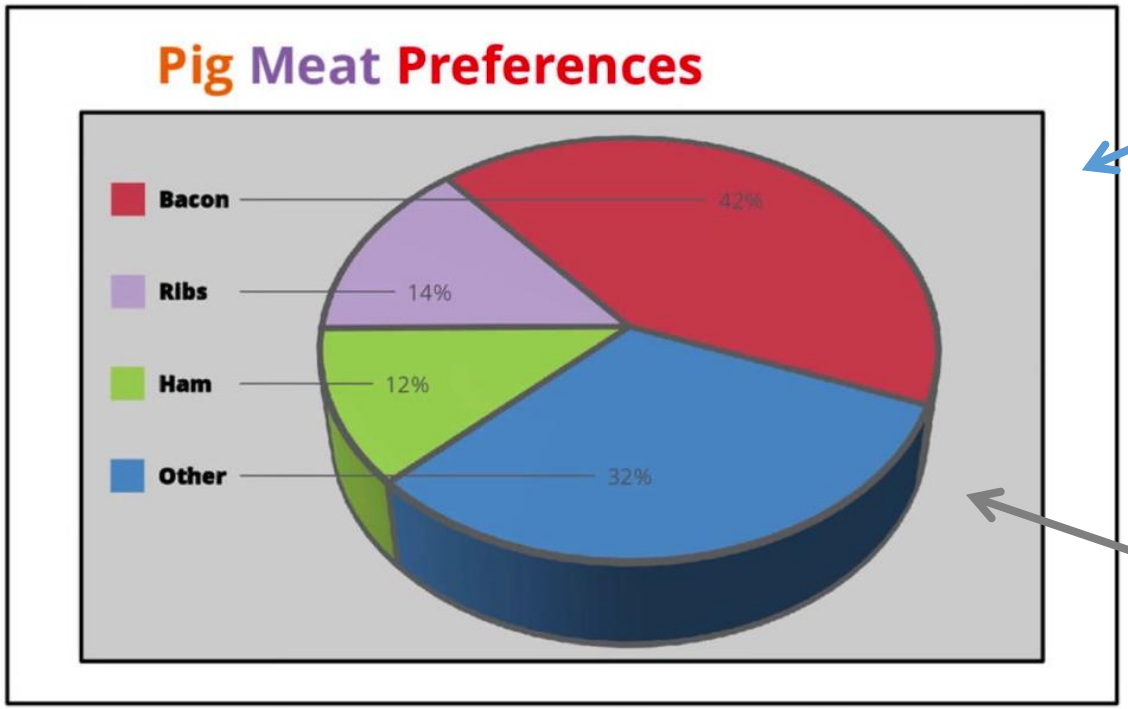
Warna background kotak luar

Warna background kotak dalam

Pendahuluan: Visualisasi Data

Contoh ke-2:

Diagram Pie 3D hanya 1 warna background pada gambar (warna abu-abu)



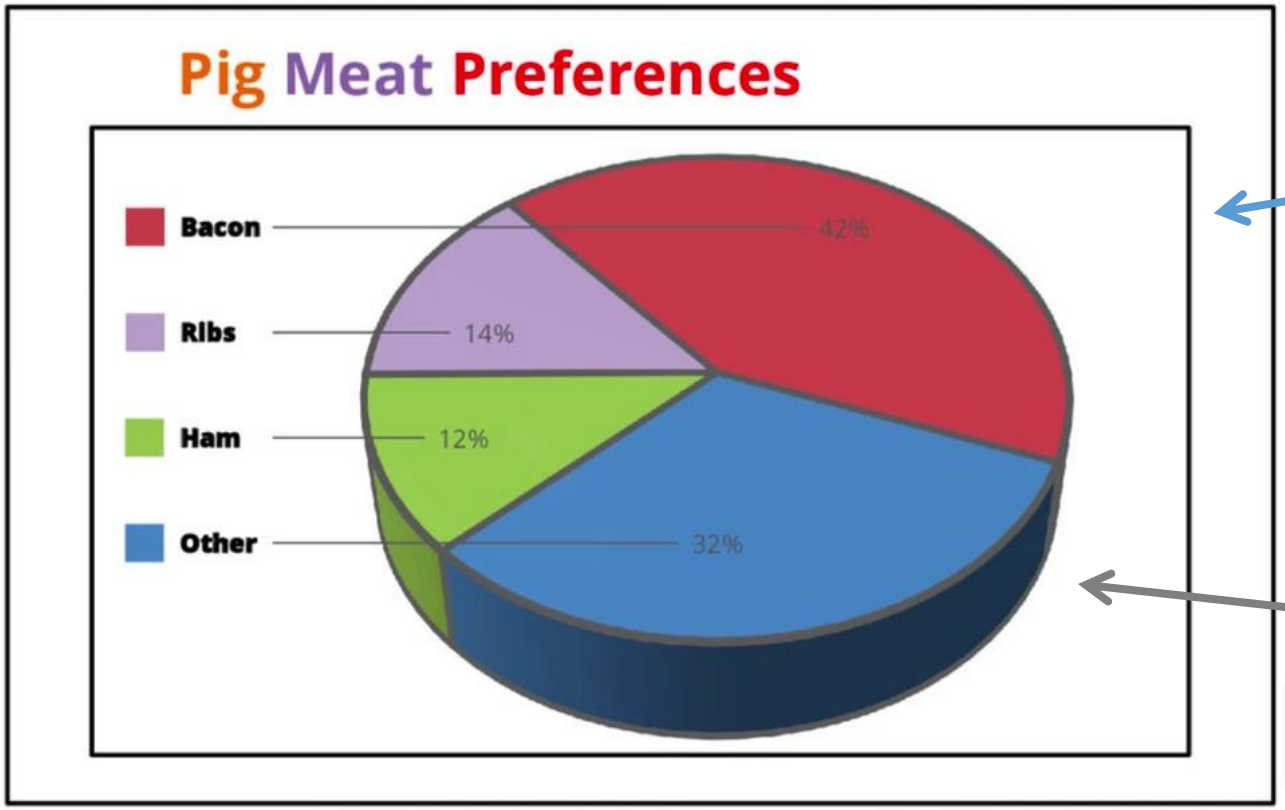
Warna background kotak luar DIHAPUS

Warna background kotak dalam

Pendahuluan: Visualisasi Data

Contoh ke-3:

Diagram Pie 3D TANPA warna background pada gambar



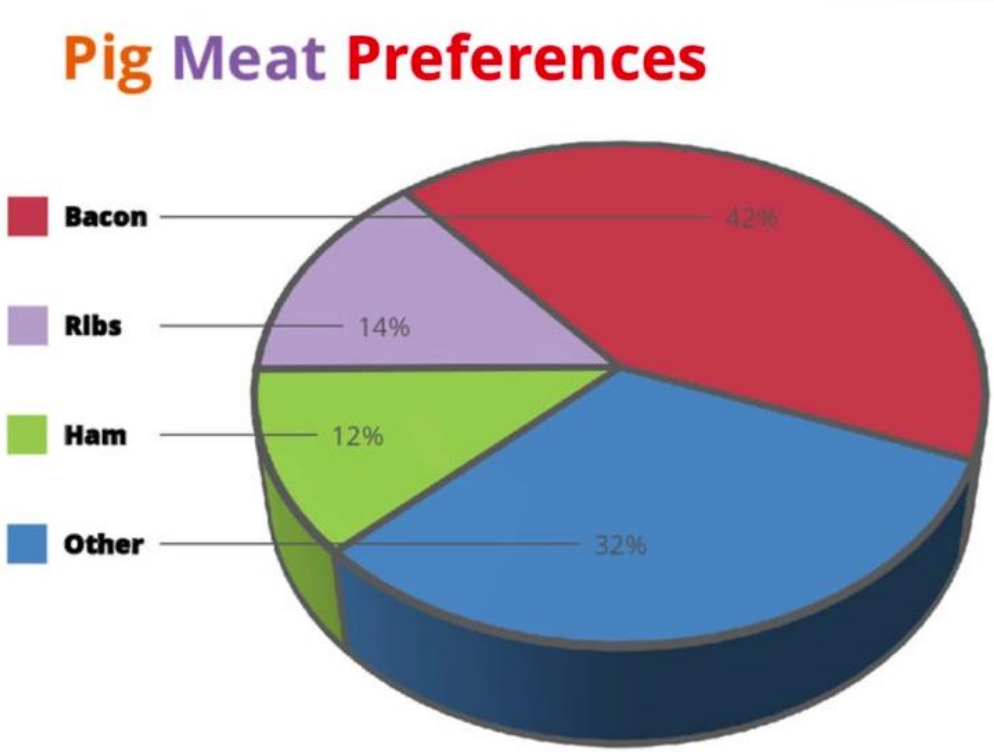
Warna background kotak luar DIHAPUS

Warna background kotak dalam

Pendahuluan: Visualisasi Data

Contoh ke-4:

Diagram Pie 3D TANPA garis bingkai

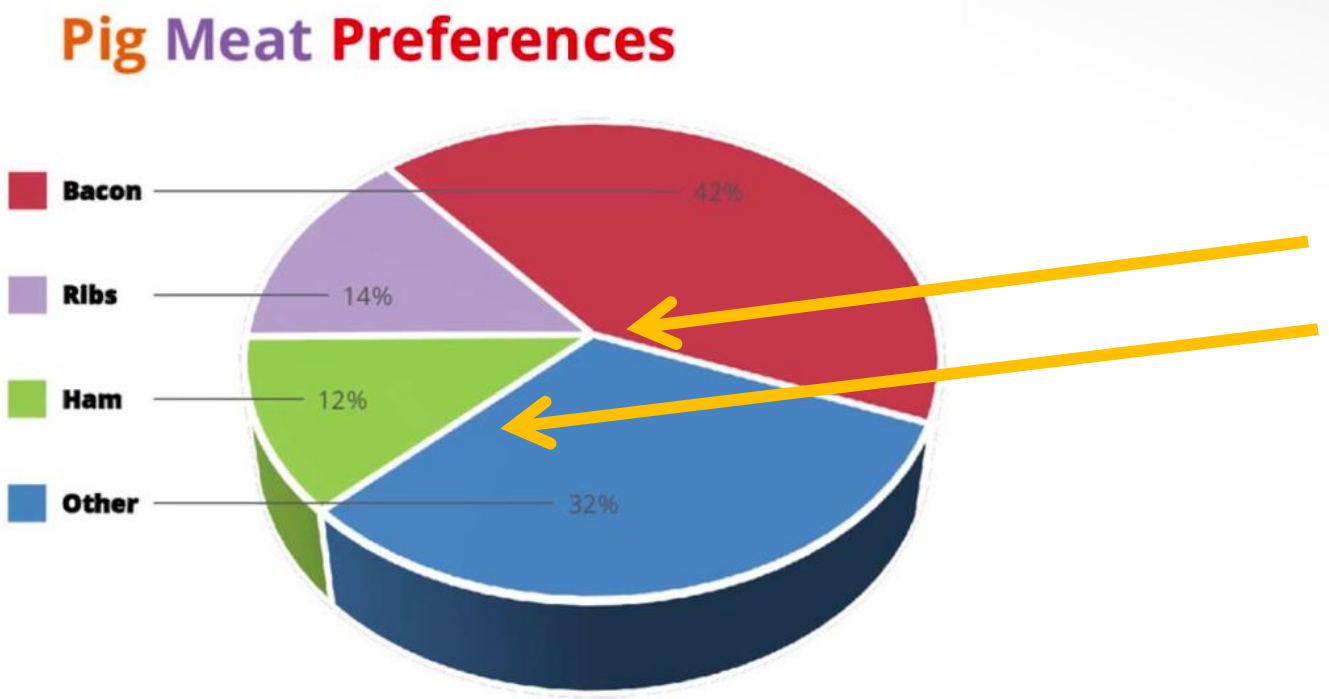


TANPA Garis bingkai

Pendahuluan: Visualisasi Data

Contoh ke-5:

Diagram Pie 3D TANPA garis pembatas antar visualiasi data

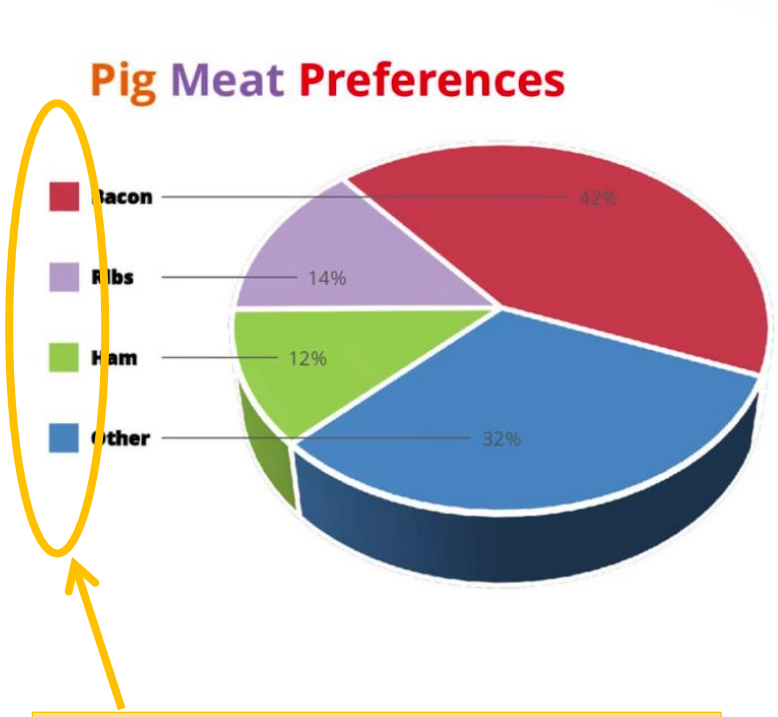


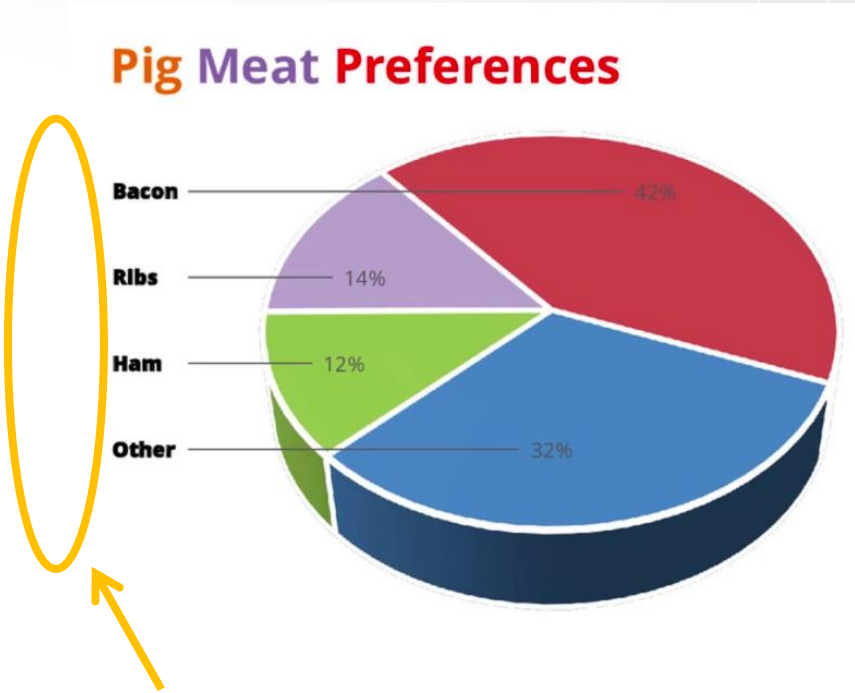
TANPA Garis pembatas antar visualiasi data

Pendahuluan: Visualisasi Data

Contoh ke-6:

Diagram Pie 3D TANPA keterangan warna dari visualisasi data (legend)





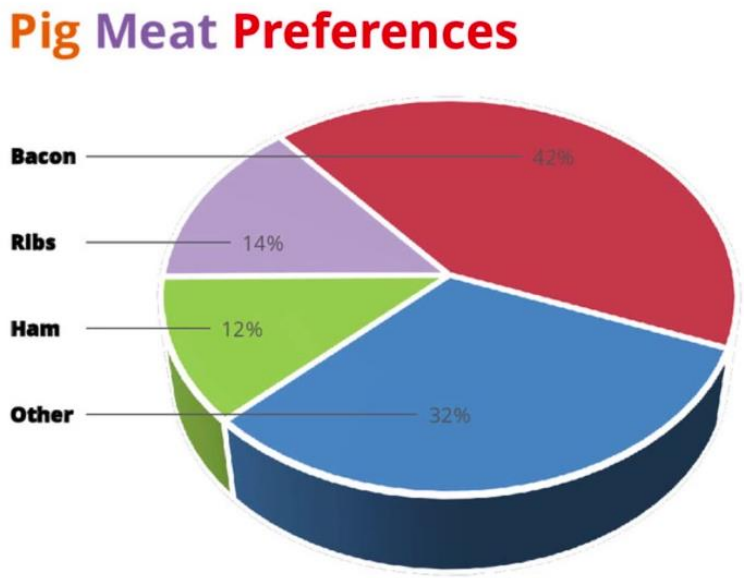
Dengan keterangan warna

TANPA keterangan warna

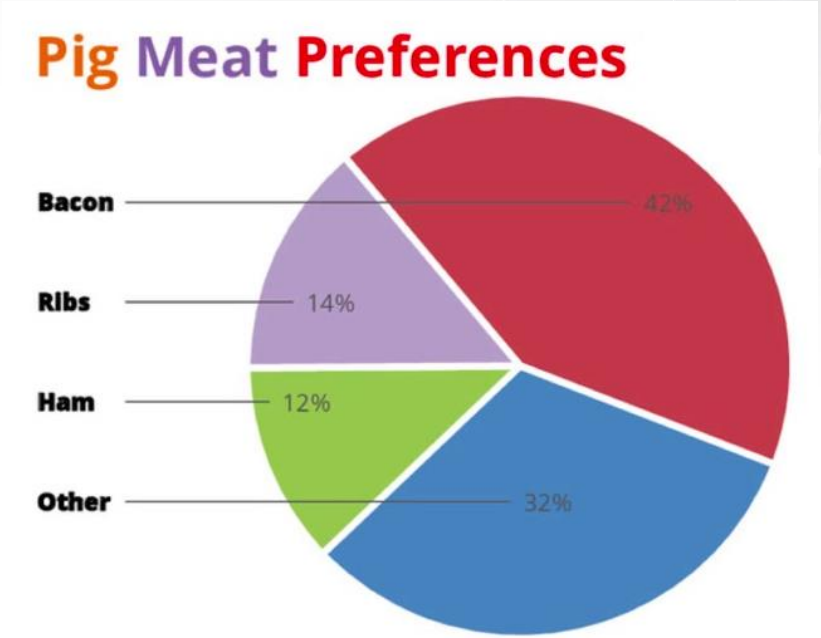
Pendahuluan: Visualisasi Data

Contoh ke-7:

Diagram Pie dapat diisualisasikan dalam 3D dan 2D



3D

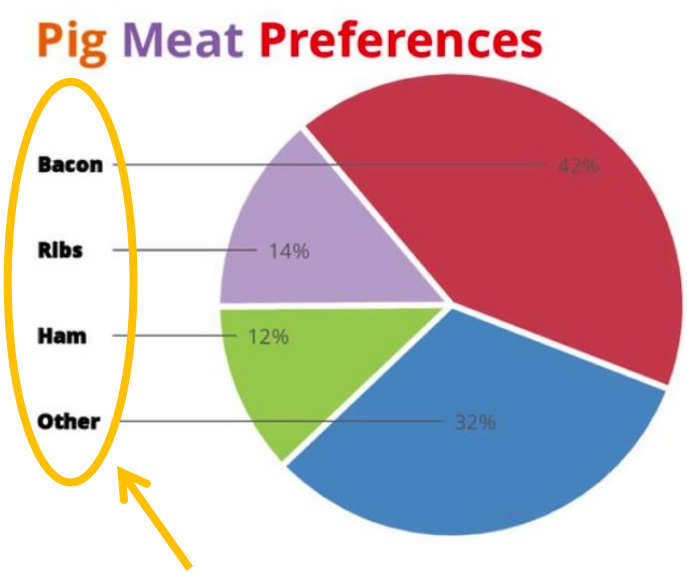


2D

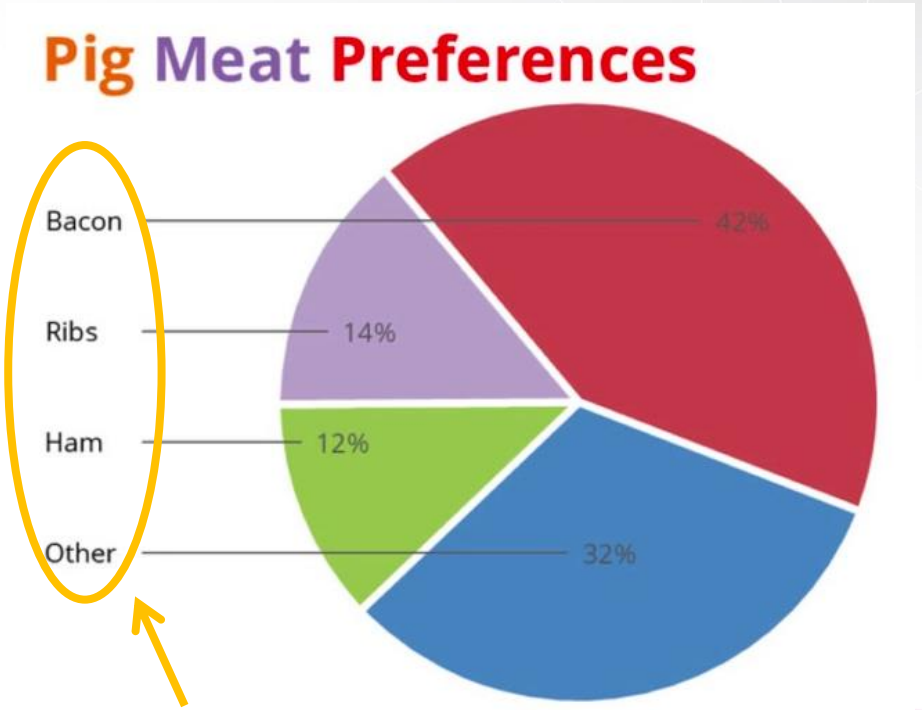
Pendahuluan: Visualisasi Data

Contoh ke-8:

Diagram Pie 2D dengan huruf TANPA cetak tebal pada tulisan keterangan yang ada di gambar



Huruf cetak tebal

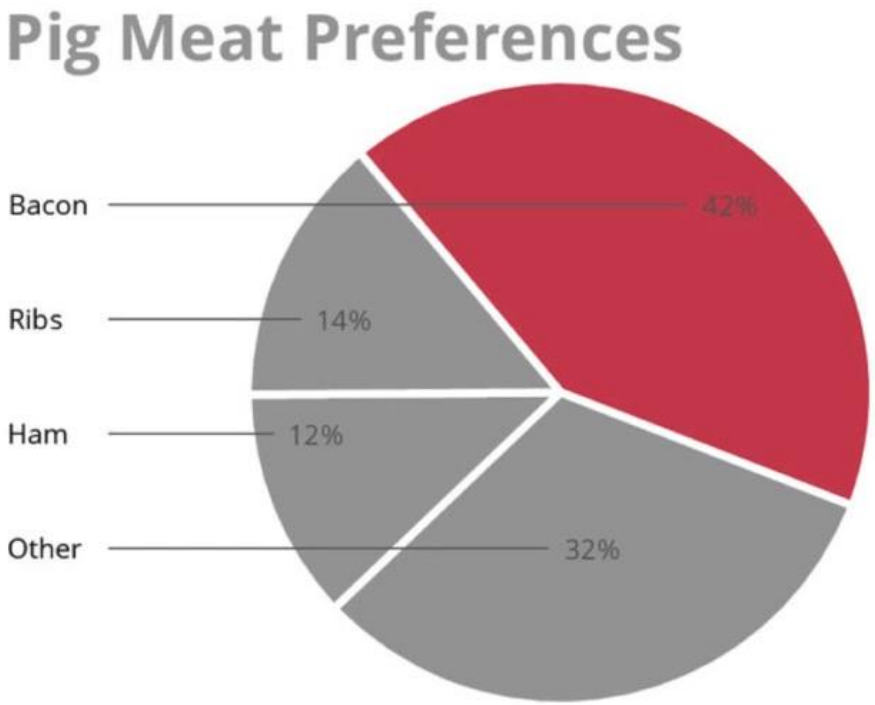


TANPA cetak tebal

Pendahuluan: Visualisasi Data

Contoh ke-9:

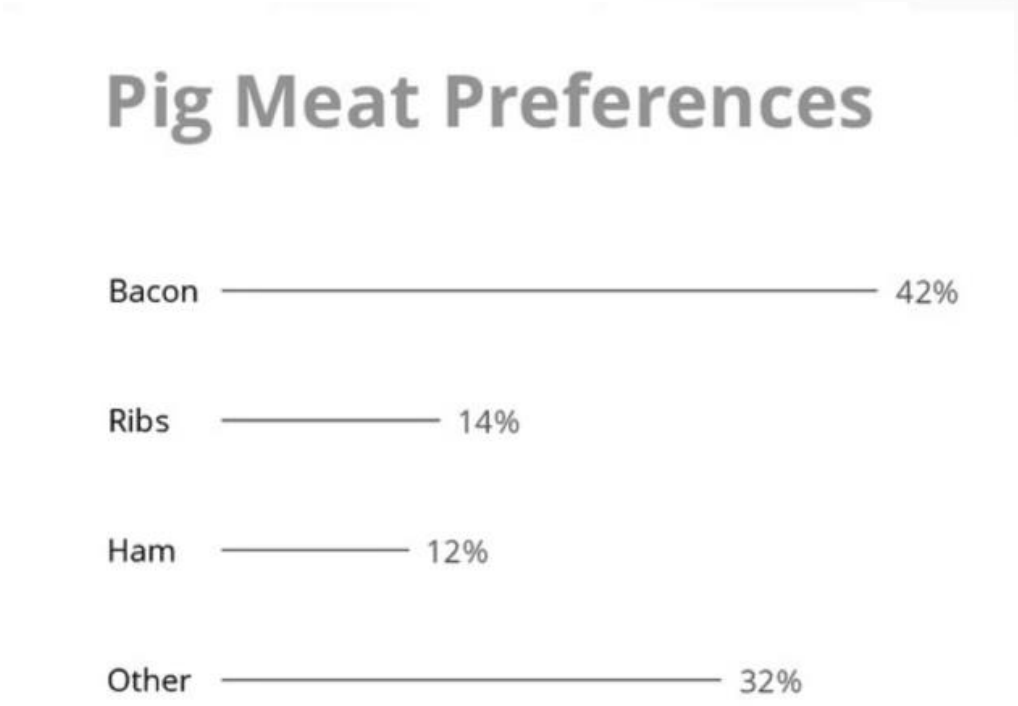
Diagam pie 2D dapat dimeminimalkan warna yang ada di gambar.



Pendahuluan: Visualisasi Data

Contoh ke-10:

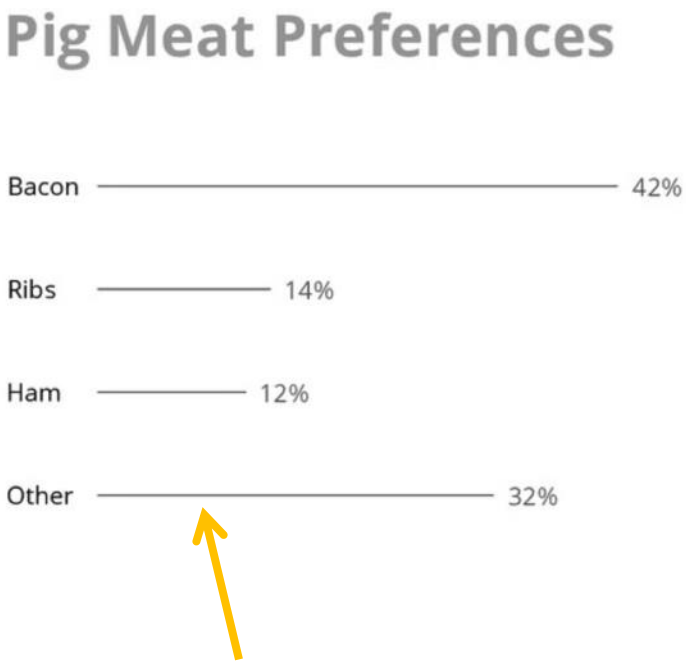
Diagram TANPA irisan (diagram dalam bentuk lines/garis)



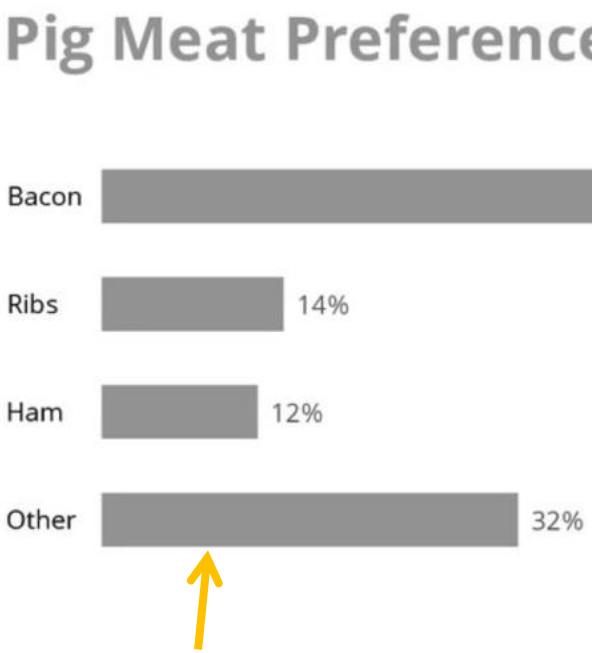
Pendahuluan: Visualisasi Data

Contoh ke-11:

Diagram lines/garis dapat dibuat dengan garis tipis maupun garis tebal.



Garis Tipis

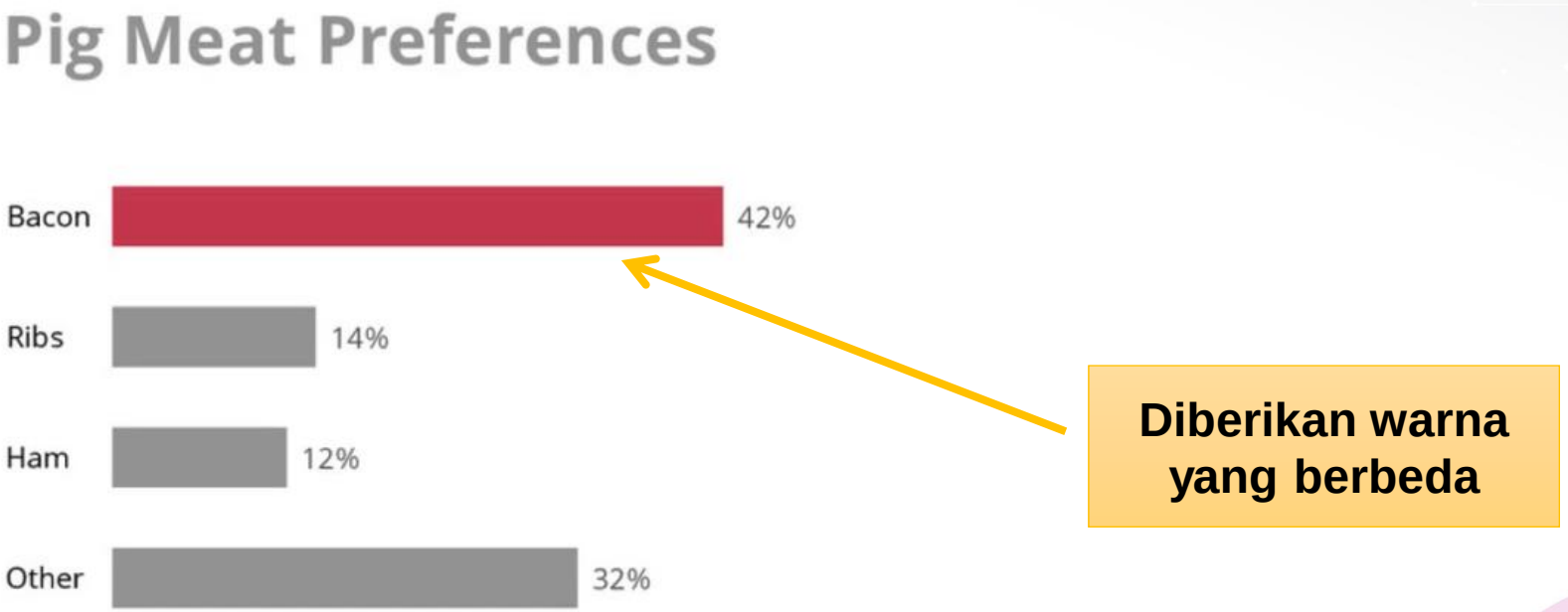


Garis Tebal

Pendahuluan: Visualisasi Data

Contoh ke-12:

Diagram lines/garis dapat menekankan salah satu/lebih data yang diinginkan dengan memberikan warna berbeda.

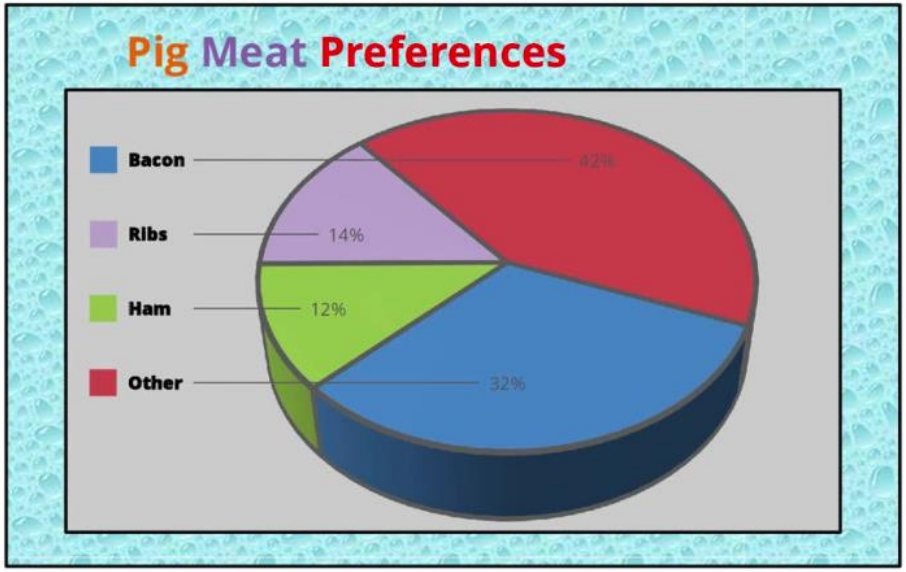


Pendahuluan: Visualisasi Data

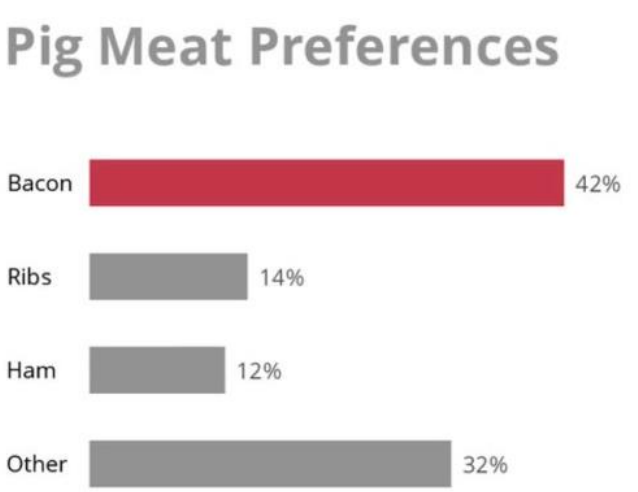
Contoh ke-13:

Perbandingan antara diagram pie dengan diagram garis

Before



After





Bagian 2

Pendahuluan Matplotlib

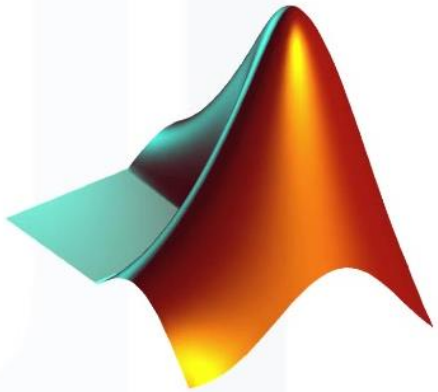
Sejarah Matplotlib



John Hunter (1968 – 2012)

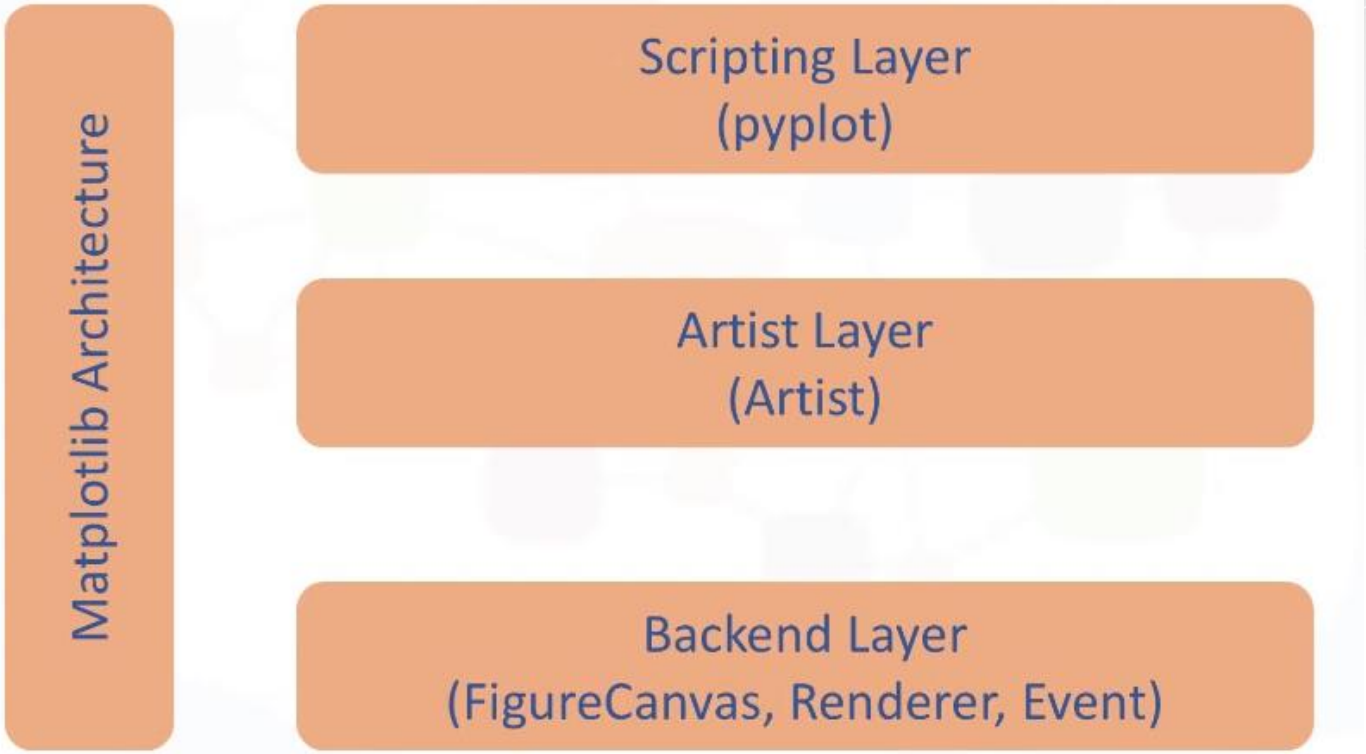


EEG/ECoG Visualization Tool



Analogous to Matlab scripting interface

Arsitektur Matplotlib



Arsitektur Matplotlib

1. Backend Layer

Memiliki 3 (tiga) class antarmuka:

1. FigureCanvas: **matplotlib.backend_bases.FigureCnvas**

   Mencakup area tempat untuk menggambar

2. Renderer: **matplotlib_backend_bases.Renderer**

   Bagaimana menggambar pada FigureCanvas

3. Event: **matplotlib.backend_bases.Event**

   Menangani input pengguna, seperti klik mouse dan tekanan/gerakan pada keyboard

Arsitektur Matplotlib

2. Artist Layer

- Terdiri dari satu objek utama – **Artist**:

  Dapat mengetahui cara menggunakan Renderer untuk menggambar pada canvas
- Title, lines, tick tables, and images, dan sebagainya sesuai pada **Artist** instances.
- Dua tipe dari **Artist** objects:

  - Primitive: line2D, rectangle, circle, and text.

  - Composite: Axis, tick, axes, and figure
- Masing-masing composite artist dapat berisi composite artists yang lainnya seperti primitive artists.

Arsitektur Matplotlib

Menempatkan Artist Layer untuk Digunakan

Membuat histogram dari contoh data dengan **Artist** layer:

```
from matplotlib.backends.backend_agg import FigureCanvasAgg as FigureCanvas # import FigureCanvas
from matplotlib.figure import Figure # import Figure artist
fig = Figure()
canvas = FigureCanvas(fig)

# create 10000 random numbers using numpy
import numpy as np
x = np.random.randn(10000)

ax = fig.add_subplot(111) # create an axes artist

ax.hist(x, 100) # generate a histogram of the 10000 numbers

# add a title to the figure and save it
ax.set_title('Normal distribution with $\mu=0, \sigma=1$')
fig.savefig('matplotlib_histogram.png')
```

Perintah figure artist

Perintah random 10000 dengan numpy

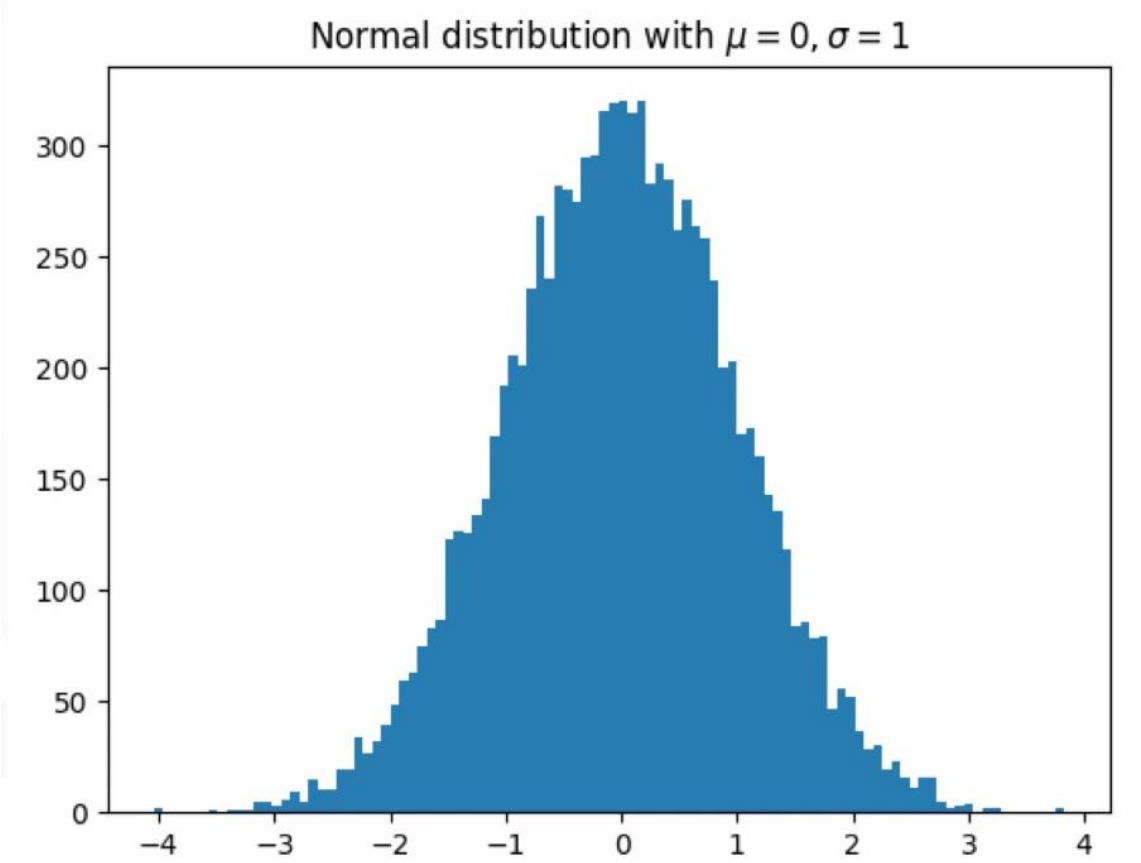
Perintah membuat sumbu artist

Perintah membuat histogram

Arsitektur Matplotlib

Menempatkan Artist Layer untuk Digunakan

Membuat histogram dari contoh data dengan **Artist** layer:



Arsitektur Matplotlib

3. Scripting Layer

1. Terdiri dari yang paling utama adalah **pyplot**, antar muka dengan **artist** layer.
2. Perintah yang digunakan untuk membuat histogram dari angka random 10000 menggunakan **pyplot** interface.

```
import matplotlib.pyplot as plt
import numpy as np

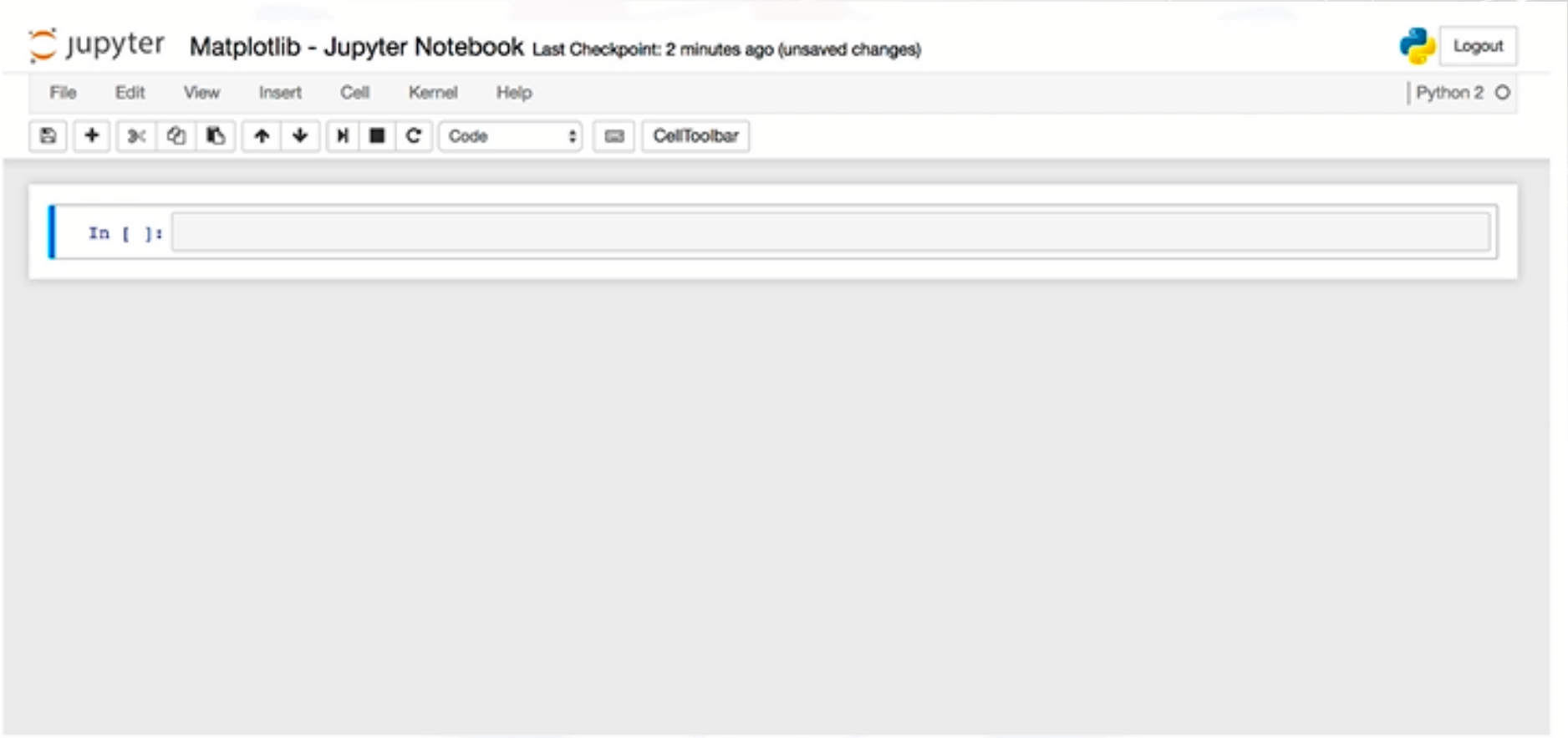
x = np.random.randn(10000)
plt.hist(x, 100)
plt.title(r'Normal distribution with $\mu=0, \sigma=1$')
plt.savefig('matplotlib_histogram.png')
plt.show()
```



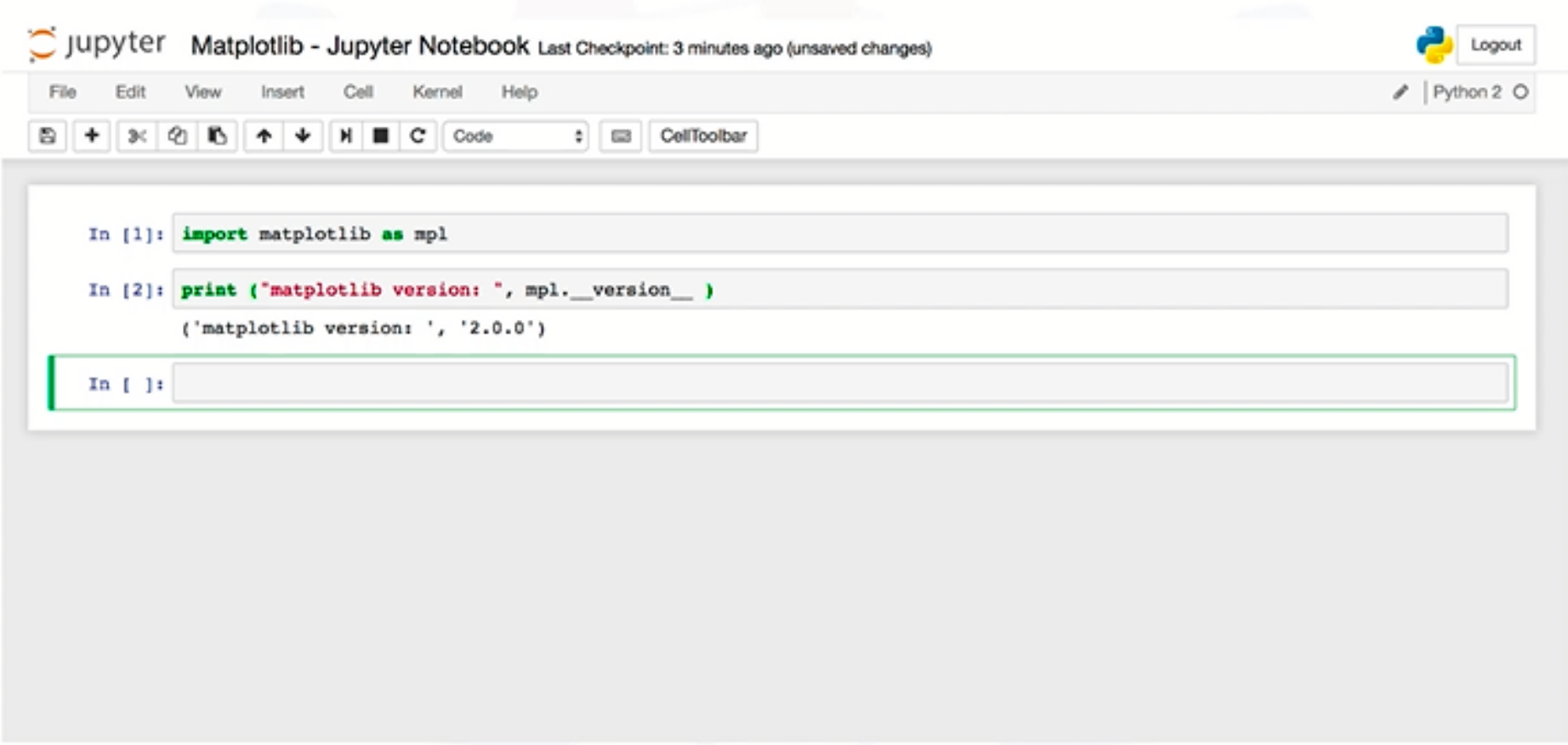
Bagian 3

Menggambar dengan Matplotlib

Matplotlib – Jupiter Notebook



Matplotlib – Jupiter Notebook



Matplotlib – Plot Function

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 39 minutes ago (unsaved changes)

Logout

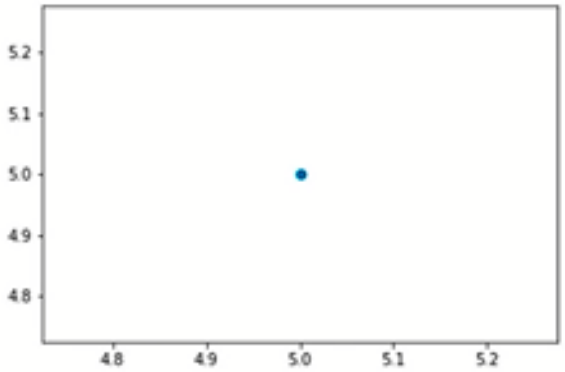
File Edit View Insert Cell Kernel Help

Python 2

Code CellToolbar

```
In [1]: import matplotlib.pyplot as plt

In [2]: plt.plot(5, 5, 'o')
        plt.show()
```




```
In [ ]:
```

Matplotlib Backends – Inline

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 2 hours ago (autosaved)

Logout

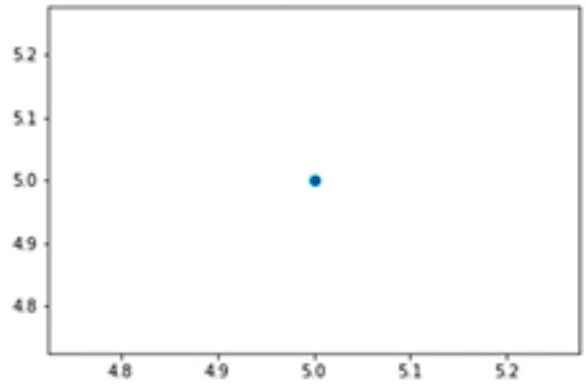
File Edit View Insert Cell Kernel Help

Python 2

Code CellToolbar

In [1]:
```
%matplotlib inline
import matplotlib.pyplot as plt
```

In [2]:
```
plt.plot(5, 5, 'o')
plt.show()
```



In [ ]:

Matplotlib Backends – Inline

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 2 hours ago (autosaved)

Logout

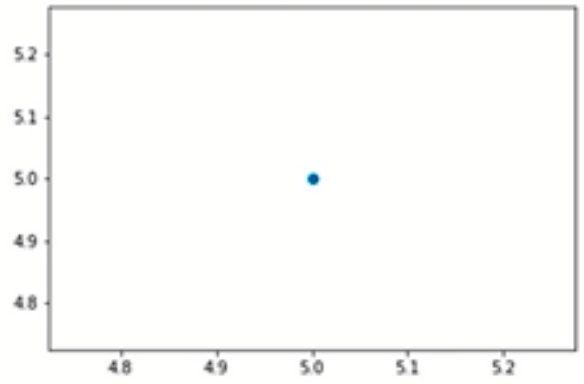
File Edit View Insert Cell Kernel Help

Python 2

Code CellToolbar

```
In [1]: %matplotlib inline
        import matplotlib.pyplot as plt

In [2]: plt.plot(5, 5, 'o')
        plt.show()
```



Matplotlib Backends – Inline

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 2 hours ago (autosaved)

Logout

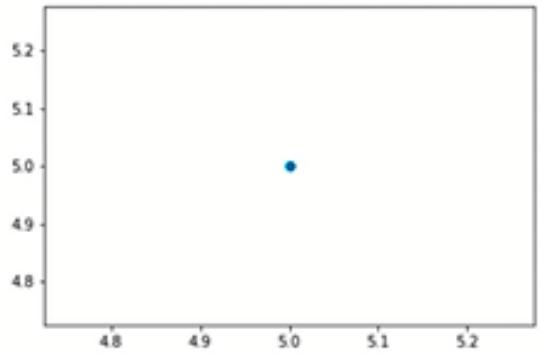
File Edit View Insert Cell Kernel Help

Python 2

Code CellToolbar

```
In [1]: %matplotlib inline
        import matplotlib.pyplot as plt

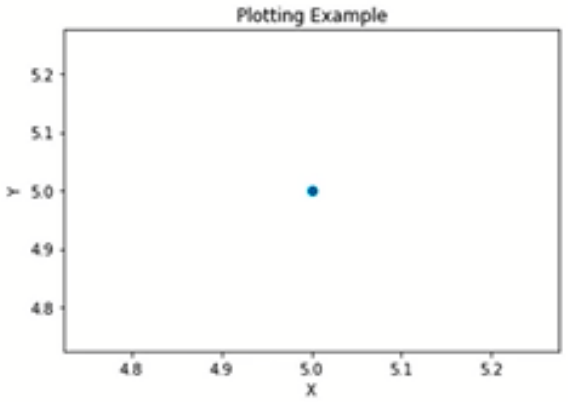
In [2]: plt.plot(5, 5, 'o')
        plt.show()
```




```
In [3]: plt.plot(5, 5, 'o')

Out[3]: plt.ylabel("Y")
        plt.xlabel("X")
        plt.title("Plotting Example")

        plt.show()
```



Matplotlib Backends – Notebook

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 16 minutes ago (autosaved)

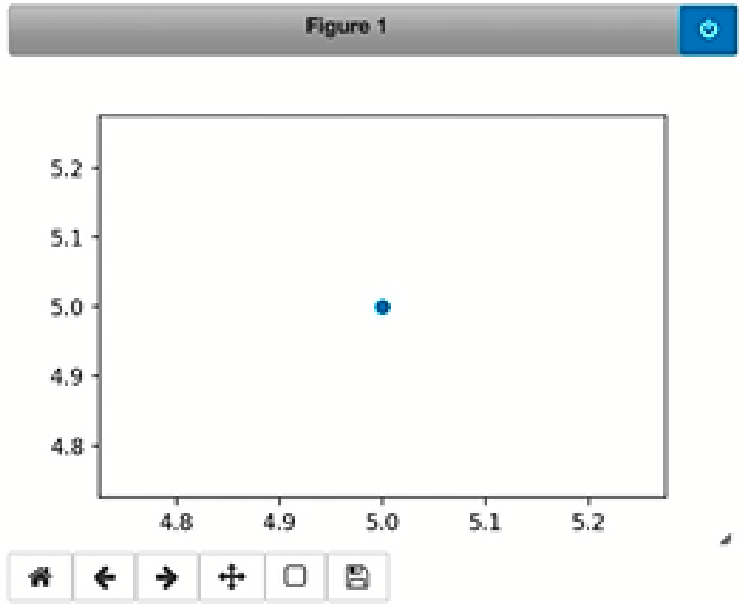
Logout

File Edit View Insert Cell Kernel Help Python 2

Code CellToolbar

```
In [1]: %matplotlib notebook
        import matplotlib.pyplot as plt

In [2]: plt.plot(5, 5, 'o')
```




```
Out[2]: [<matplotlib.lines.Line2D at 0x10784c790>]
```

Matplotlib Backends – Notebook

jupyter Matplotlib - Jupyter Notebook Last Checkpoint: 19 minutes ago (unsaved changes)

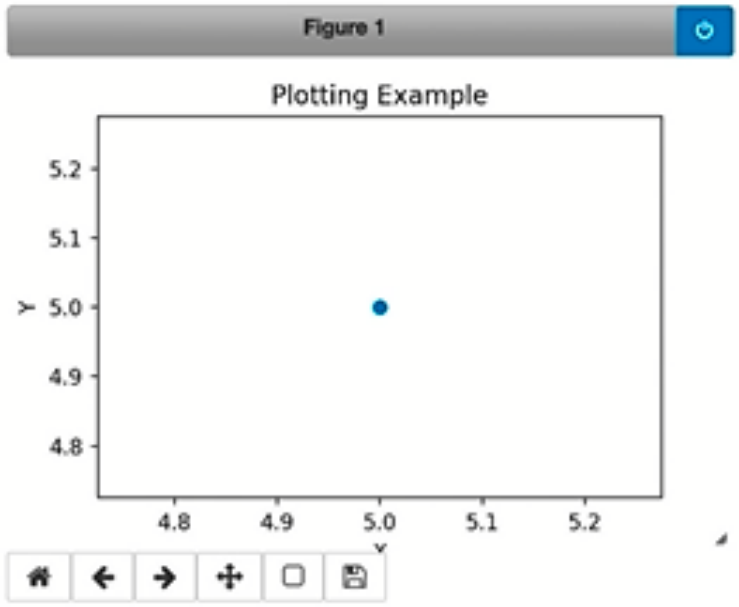
Logout

File Edit View Insert Cell Kernel Help Python 2

Code CellToolbar

```
In [1]: %matplotlib notebook
        import matplotlib.pyplot as plt

In [2]: plt.plot(5, 5, 'o')
```




```
Out[2]: [<matplotlib.lines.Line2D at 0x10784c790>]

In [3]: plt.ylabel("Y")
        plt.xlabel("X")
        plt.title("Plotting Example")

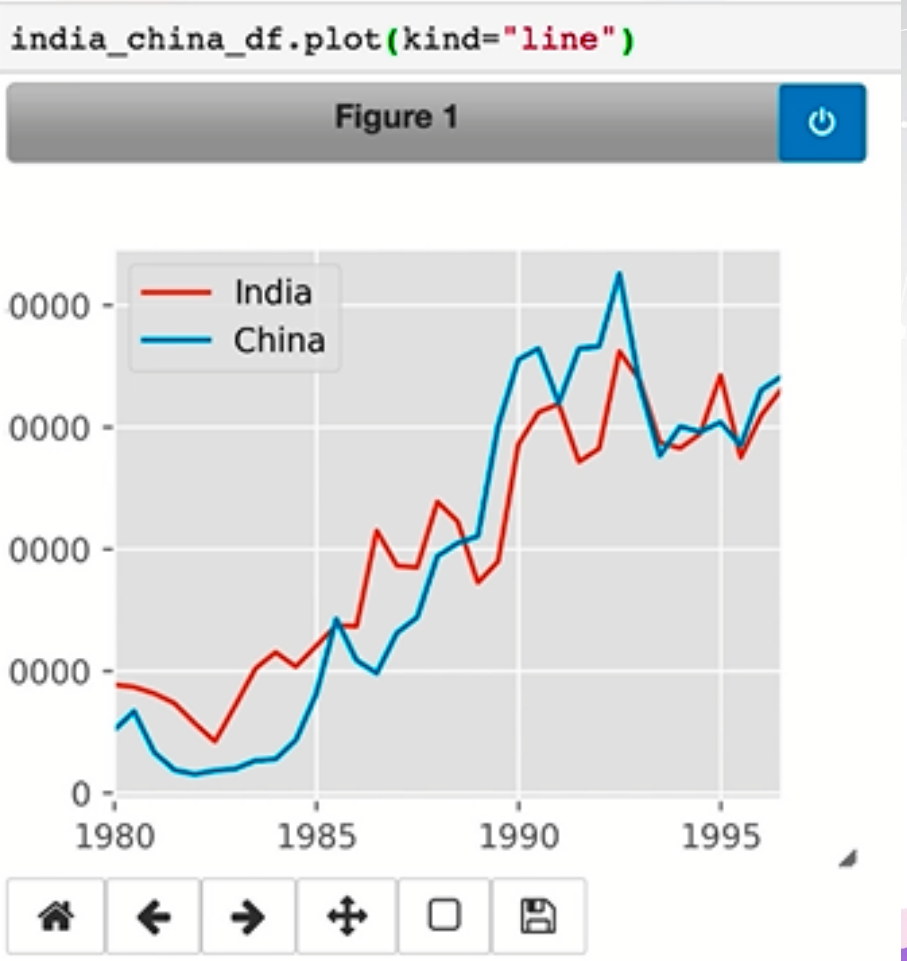
Out[3]: <matplotlib.text.Text at 0x1077f2910>
```

Matplotlib – Pandas

india_china_df

	India	China
1980	8880	5123
1981	8670	6682
1982	8147	3308
1983	7338	1863
1984	5704	1527

```
india_china_df.plot(kind="line")
```

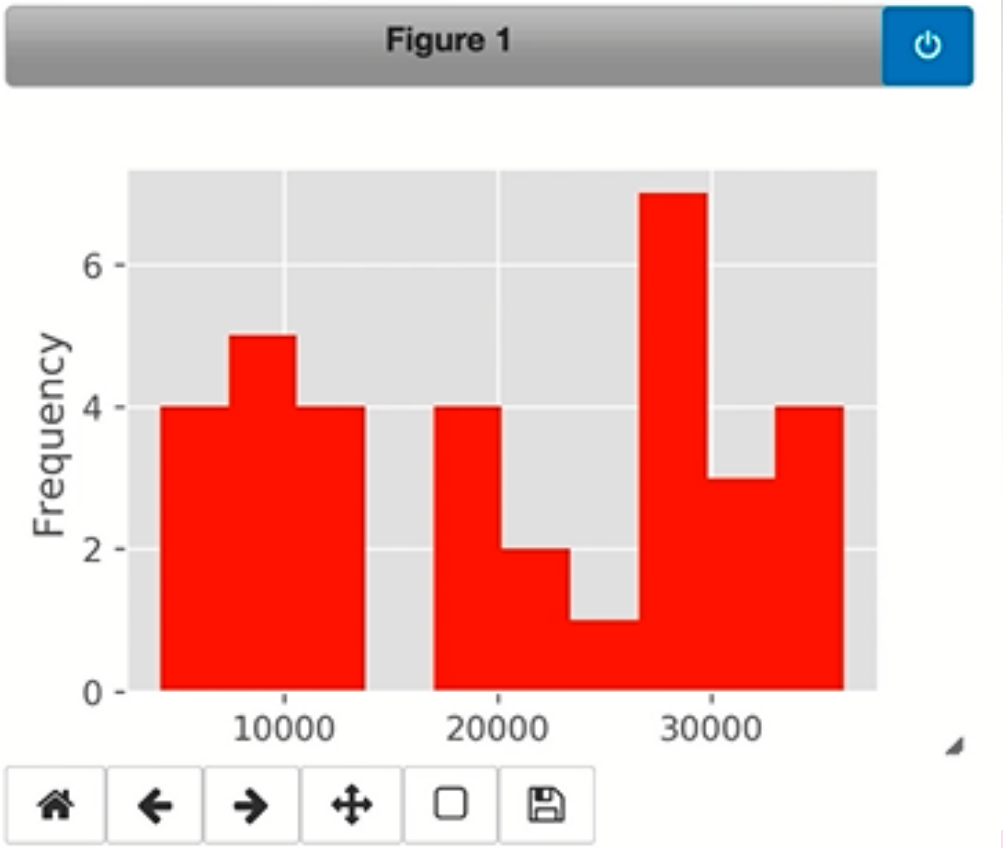


Matplotlib – Pandas

india_china_df

	India	China
1980	8880	5123
1981	8670	6682
1982	8147	3308
1983	7338	1863
1984	5704	1527

```
india_china_df["India"].plot(kind="hist")
```





Bagian 3

Pandas dan Matplotlib

Dataset

1. Divisi populasi negara-negara Amerika mengumpulkan data yang berkaitan dengan 45 negara.
2. Untuk setiap negara, data tahunan tentang arus migran internasional dilaporkan selain metadata lainnya.
3. Data yang digunakan sebagai latihan pada paket metplotlib adalah data PBB tentang imigrasi ke Kanada.

Data Imigrasi di Canada

United Nations
Population Division
Department of Economic and Social Affairs

International Migration Flows to and from Selected Countries: The 2015 Revision

POP/DB/MIG/Flow/Rev.2015

December 2015 - Copyright © 2015 by United Nations. All rights reserved

Suggested citation: United Nations, Department of Economic and Social Affairs, Population Division (2015).
International Migration Flows to and from Selected Countries: The 2015 Revision. (United Nations database,

Reporting country: Canada
Criterion: Citizenship

Classification		Origin/Destination	Major area		Region		Development region							
Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	1981	1982	1983	1984	1985
Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	39	39	47	71	340
Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	0	0	0	0	0
Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	67	71	69	63	44
Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	1	0	0	0	0
Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	0	0	0	0	0

Data Imigrasi di Canada

United Nations
Population Division
Department of Economic and Social Affairs

International Migration Flows to and from Selected Countries: The 2015 Revision

POP/DB/MIG/Flow/Rev.2015

December 2015 - Copyright © 2015 by United Nations. All rights reserved

Suggested citation: United Nations, Department of Economic and Social Affairs, Population Division (2015).
International Migration Flows to and from Selected Countries: The 2015 Revision. (United Nations database,

Reporting country: Canada
Criterion: Citizenship

Classification		Origin/Destination	Major area		Region		Development region						
Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	1981	1982	1983	1984	1985
Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	39	39	47	71	340
Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	0	0	0	0	0
Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	67	71	69	63	44
Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	1	0	0	0	0
Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	0	0	0	0	0

Membaca data pada Dataframe Pandas

```
import numpy as np  # useful for many scientific computing in Python
import pandas as pd # primary data structure library
from __future__ import print_function # adds compatibility to python 2
```

```
# install xlrd
!pip install xlrd

print('xlrd installed!')
```

Membaca data pada Dataframe Pandas

```
import numpy as np  # useful for many scientific computing in Python
import pandas as pd # primary data structure library
from __future__ import print_function # adds compatibility to python 2
```

```
# install xlrd
!pip install xlrd

print('xlrd installed!')
```

```
df_can = pd.read_excel(
    'https://ibm.box.com/shared/static/lw190pt9zpy5bd1ptyg2aw15awomz9pu.xlsx',
    sheetname="Canada by Citizenship",
    skiprows=range(20),
    skip_footer = 2)
```

Tampilan Dataframe

```
df_can.head()
```

	Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	...	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	...	2978	3436	3009	2652	2111	1746	1758	2203	2635	2004
1	Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	...	1450	1223	856	702	560	716	561	539	620	603
2	Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	...	3616	3626	4807	3623	4005	5393	4752	4325	3774	4331
3	Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	...	0	0	1	0	0	0	0	0	0	0
4	Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	...	0	0	1	1	0	0	0	0	1	1

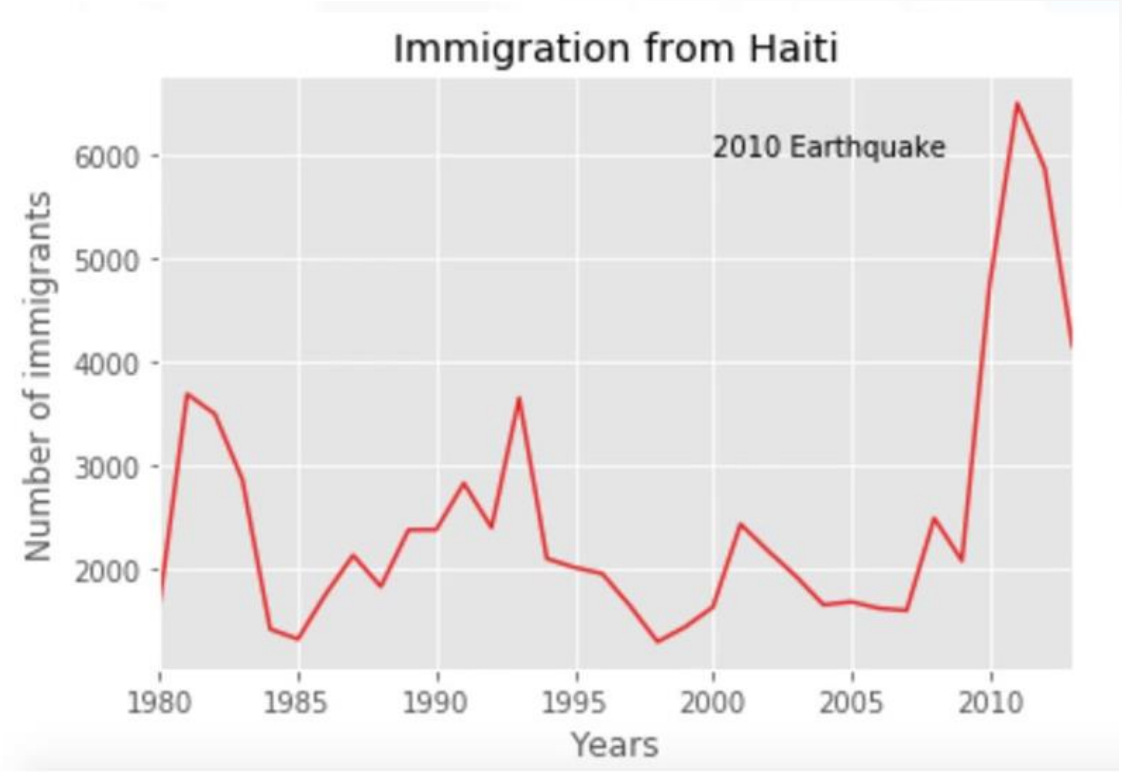


Bagian 4

Line Plot

Line Plots

Line plots adalah jenis plot yang menampilkan informasi sebagai serangkaian titik data yang disebut "penanda" yang dihubungkan oleh segmen garis lurus.



Dataset - Recap

	Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	...	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	...	2978	3436	3009	2652	2111	1746	1758	2203	2635	2004
1	Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	...	1450	1223	856	702	560	716	561	539	620	603
2	Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	...	3616	3626	4807	3623	4005	5393	4752	4325	3774	4331
3	Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	...	0	0	1	0	0	0	0	0	0	0
4	Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	...	0	0	1	1	0	0	0	0	1	1

Dataset - Processed

	Continent	Region	DevName	1980	1981	1982	1983	1984	1985	1986	...	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Country																					
Afghanistan	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	496	...	3436	3009	2652	2111	1746	1758	2203	2635	2004	58639
Albania	Europe	Southern Europe	Developed regions	1	0	0	0	0	0	1	...	1223	856	702	560	716	561	539	620	603	15699
Algeria	Africa	Northern Africa	Developing regions	80	67	71	69	63	44	69	...	3626	4807	3623	4005	5393	4752	4325	3774	4331	69439
American Samoa	Oceania	Polynesia	Developing regions	0	1	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	6
Andorra	Europe	Southern Europe	Developed regions	0	0	0	0	0	0	2	...	0	1	1	0	0	0	0	1	1	15

df_canada

Membuat Line Plots

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

```
years = list(map(str, range(1980, 2014)))

df_canada.loc['Haiti', years].plot(kind = 'line')
plt.title('Immigration from Haiti')
plt.ylabel('Number of immigrants')
plt.xlabel('Years')

plt.show()
```

Membuat Line Plots

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

```
years = list(map(str, range(1980, 2014)))

df_canada.loc['Haiti', years].plot(kind = 'line')
plt.title('Immigration from Haiti')
plt.ylabel('Number of immigrants')
plt.xlabel('Years')

plt.show()
```

