数据可视化

Table of Contents

官方使用教程是非常重要的学习来源。## 图形的构成

我们用一个例子来说明绘图的过程。下面的全球电影票房数据来自维基百科全球最高電影票房收入列表

```
import pandas as pd
import re
data = pd.read_excel("datasets/highest_gross_films.xlsx")
data['全球票房'] = data['全球票房'].apply(lambda ser: pd.to_numeric(re.sub(r'\D','', ser)))
df = data[:10]
df
```

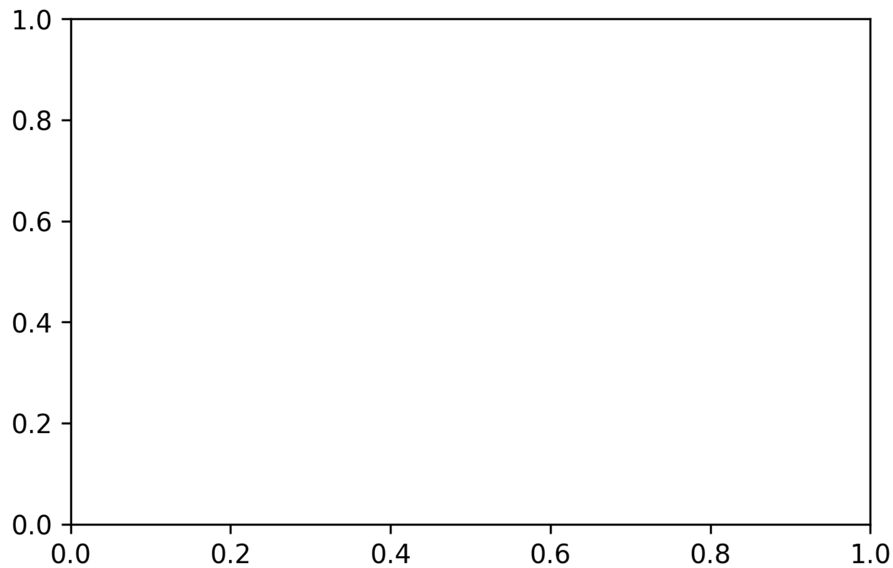
	排名	峰值	影片名称	全球票房	年份
0	1	1	阿凡达	2923706026	2009
1	2	1	复仇者联盟：终局之战	2797501328	2019
2	3	3	阿凡达：水之道	2320250281	2022
3	4	1	泰坦尼克号	2257844554	1997
4	5	5	哪吒 2	2217080000	2025
5	6	3	星球大战：原力觉醒	2068223624	2015
6	7	4	复仇者联盟：无限战争	2048359754	2018
7	8	6	蜘蛛侠：英雄无归	1922598800	2021
8	9	8	头脑特工队 2	1698863816	2024
9	10	3	侏罗纪世界	1671537444	2015

0.0.1 图形的解剖图

0.0.2 步骤

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots()
```



现在有了图形轴（Axes）的实例，就可以在上面绘制图形了：

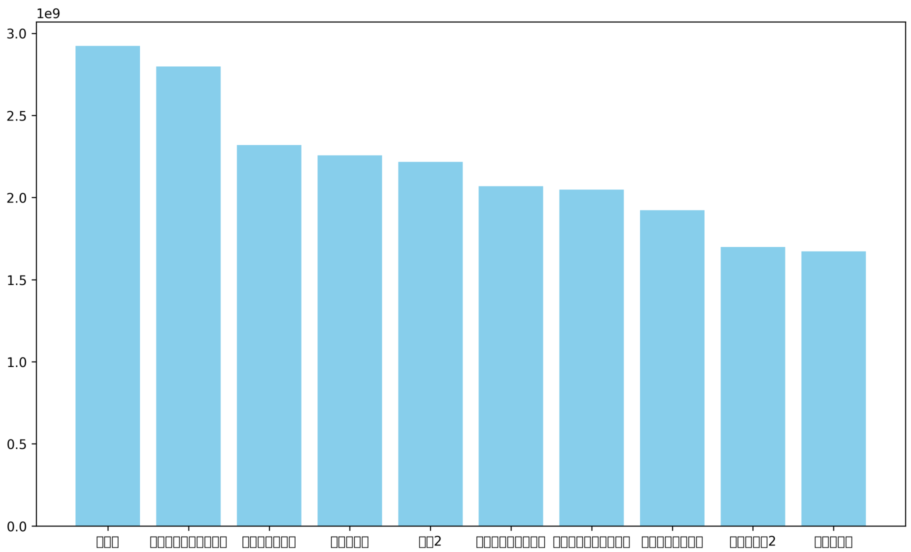
```
fig, ax = plt.subplots(figsize=(12, 7))
ax.bar(x=df['影片名称'], height=df['全球票房'], color='skyblue')
plt.show()
```

```
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
```

```
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
```

```
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabt
  fig.canvas.print_figure(bytes_io, **kw)
```

```
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
D:\Project\Quarto Books\EconomicAnalysis\eco\Lib\site-packages\IPython\core\pylabtools.py:17
  fig.canvas.print_figure(bytes_io, **kw)
```



0.1 常用的图形

0.1.1 直方图

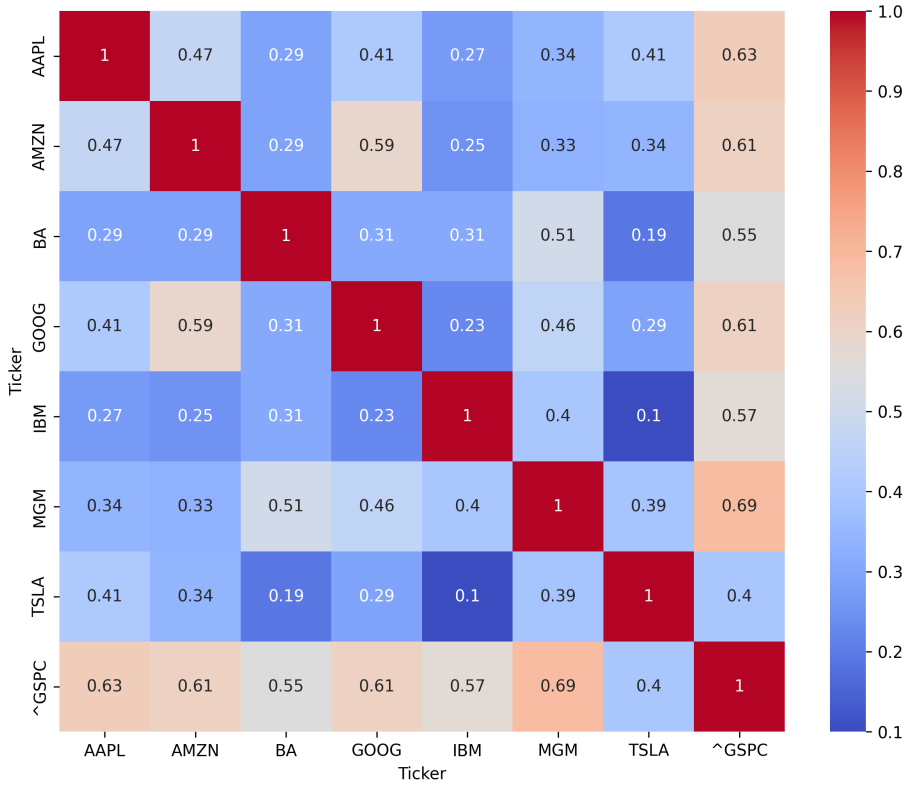
0.1.2 核密度图

0.1.3 热图

```
import yfinance as yf
import seaborn as sns

stocks_list = ['AAPL','BA','MGM','AMZN','IBM','TSLA','GOOG','^GSPC']
start_date = "2012-01-01"
end_date = "2024-12-31"
stocks_price_us = yf.download(tickers=stocks_list,
         start=start_date, end=end_date,
         interval="1mo",
         auto_adjust=True,
         progress=False)['Close']
returns = stocks_price_us.pct_change()
corr = returns.corr().round(2)

fig, ax = plt.subplots(figsize=(10, 8))
sns.heatmap(corr, annot=True,
            ax=ax,
            cmap='coolwarm')
plt.show()
```



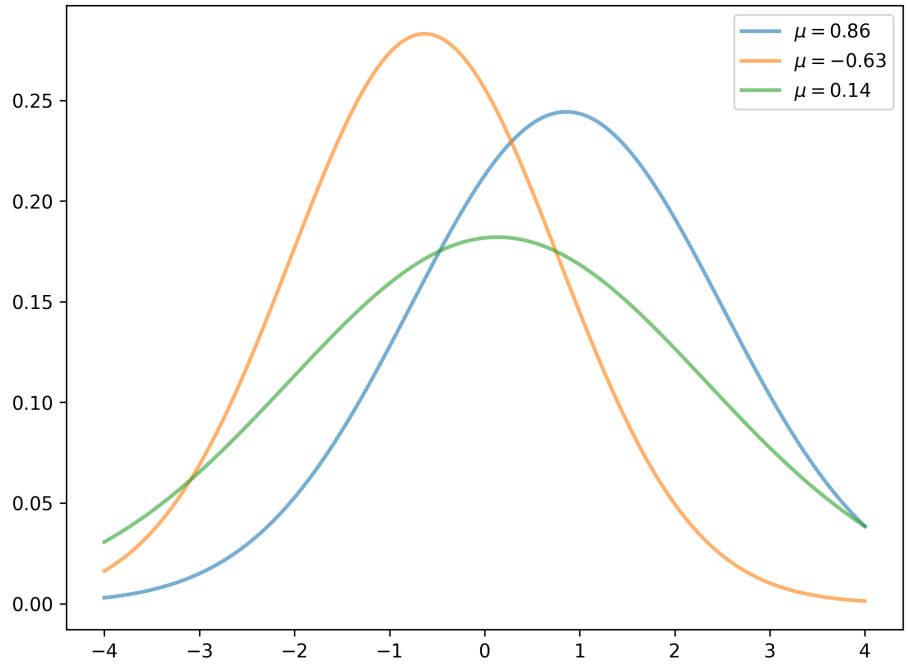
0.2 多图和子图

0.3 多图

```
# 正态分布
import numpy as np
from scipy.stats import norm
np.random.seed(12345)

fig, ax = plt.subplots(figsize=(8, 6))
x = np.linspace(-4, 4, 500)
for i in range(3):
```

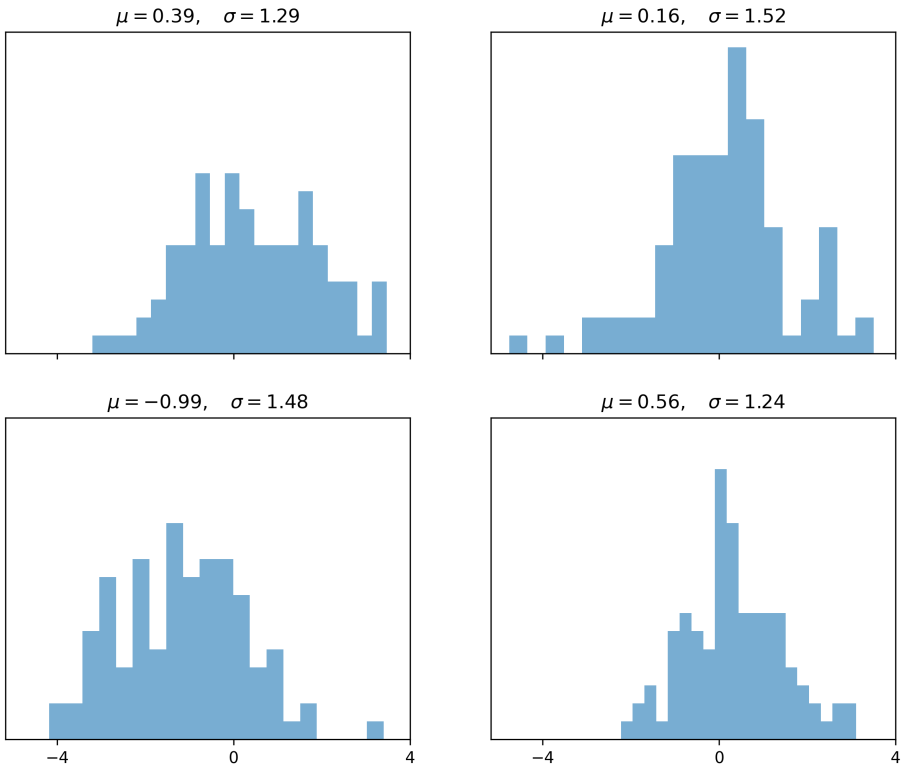
```
    mu, std = np.random.uniform(-1,1), np.random.uniform(1, 3)
    y = norm.pdf(x, loc = mu, scale = std)
    current_label = rf"$\mu = {mu:.2f}$"
    ax.plot(x, y, linewidth = 2, alpha = 0.6, label = current_label)
ax.legend()
plt.show()
```



0.3.1 子图

```
np.random.seed(123)
fig, axes = plt.subplots(nrows=2, ncols=2,figsize=(10,8),sharex=True,sharey=True)
for i in range(2):
    for j in range(2):
        m, s = np.random.uniform(-1, 1), np.random.uniform(1, 2)
        x = np.random.normal(m,s,100)
        axes[i, j].hist(x, alpha = 0.6, bins=20)
```

```
        title = rf"$\mu = {m:.2f},\quad \sigma = {s:.2f}$"
        axes[i, j].set(title = title, xticks = [-4, 0, 4],yticks = [])
plt.show()
```



0.3.2 图形风格

```
plt.style.available
```

```
['Solarize_Light2',
 '_classic_test_patch',
 '_mpl-gallery',
 '_mpl-gallery-nogrid',
 'bmh',
 'classic',
```

```
 'dark_background',
 'fast',
 'fivethirtyeight',
 'ggplot',
 'grayscale',
 'petroff10',
 'seaborn-v0_8',
 'seaborn-v0_8-bright',
 'seaborn-v0_8-colorblind',
 'seaborn-v0_8-dark',
 'seaborn-v0_8-dark-palette',
 'seaborn-v0_8-darkgrid',
 'seaborn-v0_8-deep',
 'seaborn-v0_8-muted',
 'seaborn-v0_8-notebook',
 'seaborn-v0_8-paper',
 'seaborn-v0_8-pastel',
 'seaborn-v0_8-poster',
 'seaborn-v0_8-talk',
 'seaborn-v0_8-ticks',
 'seaborn-v0_8-white',
 'seaborn-v0_8-whitegrid',
 'tableau-colorblind10']
```

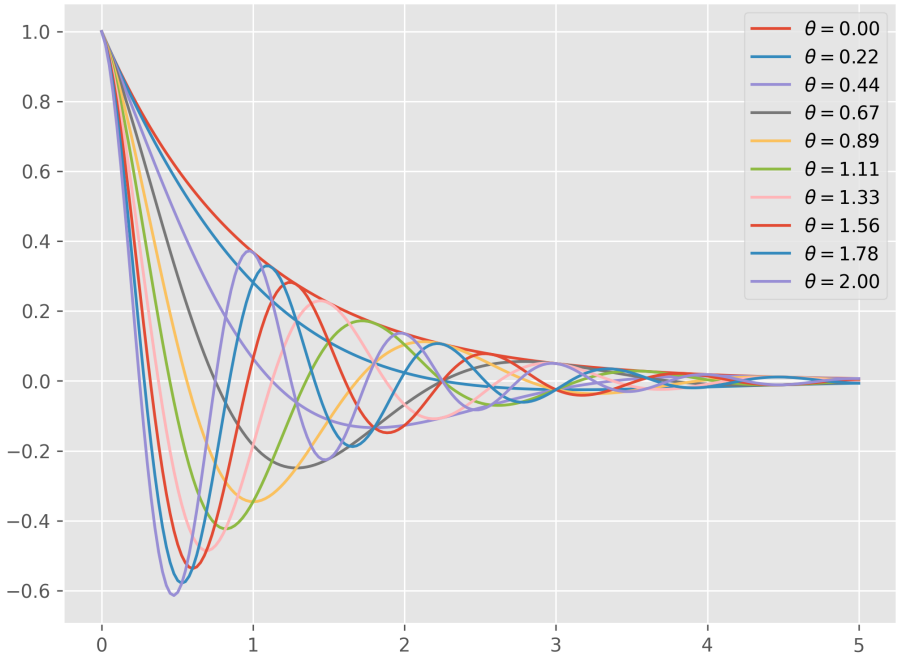
```
import numpy as np
import matplotlib.pyplot as plt
plt.style.use("ggplot")

def f(x, theta):
    return np.cos(np.pi * theta * x ) * np.exp(- x)

 _vals = np.linspace(0, 2, 10)
x = np.linspace(0, 5, 200)
fig, ax = plt.subplots(figsize=(8, 6))
```

```
for theta in  _vals:
    ax.plot(x, f(x, theta), label = rf"$\theta = {theta:.2f}$")

ax.legend()
plt.show()
```



0.4 应用：股票价格及收益率