



中國人民大學
RENMIN UNIVERSITY OF CHINA

第2编 Python数据分析

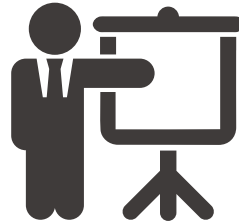
第2讲 Matplotlib可视化

余力

buaayuli@ruc.edu.cn



中國人民大學
RENMIN UNIVERSITY OF CHINA



01. pyplot与pylab

pyplot的方式

```
import numpy as np
```

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

```
x=np.arange(0,10,1)
```

```
y=np.random.randn( len(x) )
```

```
plt.title("pyplot")
```

```
plt.plot(x,y)
```

```
plt.show()
```

类封装的方式

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

x = np.arange(0, 10, 1)
y = np.random.randn(len(x))

# 生成一个figure对象
fig = plt.figure()
ax = fig.add_subplot(111)
ax.set_title("object oriented")
plt.plot(x, y)

plt.show()
```

pylab的方式

```
#coding: utf-8
```

```
from pylab import *
```

```
x = arange(0, 10, 1)
```

```
y = randn(len(x))
```

```
title("pylab")
```

```
plot(x, y)
```

```
show()
```

典型图形

```
import pylab as pl
import numpy as np
pl.figure(figsize=(8, 6), dpi=80)
pl.subplot(1, 1, 1)
X = np.linspace(-np.pi, np.pi, 256, endpoint=True)
C, S = np.cos(X), np.sin(X)
pl.plot(X, C, color="blue", linewidth=1.0, linestyle="-")
pl.plot(X, S, color="green", linewidth=1.0, linestyle="-")
pl.xlim(-4.0, 4.0)
pl.xticks(np.linspace(-4, 4, 9, endpoint=True))
pl.ylim(-1.0, 1.0)
pl.yticks(np.linspace(-1, 1, 5, endpoint=True))
# savefig("exercice_2.png", dpi=72)
pl.show()
```

美化图形线条

- 1) 控制颜色

- 红色：把`pl.plot(x, y, 'o')`改成`pl.plot(x, y, 'or')`

- 颜色之间的对应关系为

b---blue **c**---cyan (青色)

g---green **k**----black

m---magenta (紫红) **r**---red

w---white **y**----yellow

美化图形线条

■ 2) 控制标记风格***

- 蓝色星型markers: `plot(x, y, ' b*')`

'o'	圆圈	':'	点
'D'	菱形	's'	正方形
'h'	六边形1	'*'	星号
'H'	六边形2	'd'	小菱形
'_'	水平线	'v'	一角朝下的三角形
'8'	八边形	'<'	一角朝左的三角形
'p'	五边形	'>'	一角朝右的三角形
'^'	像素	'^'	一角朝上的三角形
'+'	加号	'\'	'
'None', ',', ''	无	'x'	X

美化图形线条

■ 3)控制线型

- 实线
- 短线
- . 短点相间线
- : 虚点线

美化图形线条-例

```
import matplotlib.pyplot as plt  
import numpy as np  
y=np.arange(1, 3, 0.3)  
plt.plot(y,'cx--',y+1, 'mo:',y+2,'kp-.')  
plt.show()
```

坐标刻度

- **设置极值**
- `pl.xlim(X.min() * 1.1, X.max() * 1.1)`
- `pl.ylim(C.min() * 1.1, C.max() * 1.1)`
- **设置坐标轴刻度值**
- `pl.xticks([-np.pi, -np.pi/2, 0, np.pi/2, np.pi])`
- `pl.yticks([-1, 0, +1])`
- **设置刻度标签**
- `pl.xticks([-np.pi, -np.pi/2, 0, np.pi/2, np.pi], [r'$-\pi$', r'$-\pi/2$', r'0', r'$+\pi/2$', r'$+\pi$'])`
- `pl.yticks([-1, 0, +1], [r'-1', r'0', r'$+1$'])`

添加图例

- **pl.plot**(X, C, color="blue", linewidth=2.5, linestyle="-", label="cosine")
- **pl.plot**(X, S, color="red", linewidth=2.5, linestyle="-", label="sine")
- **pl.legend**(loc='upper left')

Legend位置与两个图

```
import numpy as np
import matplotlib.pyplot as plt
fig = plt.figure()
ax1 = fig.add_axes([0.1, 0.1, 0.4, 0.7]) # 固定坐标画, [left, bottom, width, height]
ax2 = fig.add_axes([0.55, 0.1, 0.4, 0.7])
x = np.arange(0.0, 2.0, 0.02)
y1 = np.sin(2*np.pi*x)
y2 = np.exp(-x)
l1,l2 = ax1.plot(x, y1, 'rs-', x, y2, 'go')
y3 = np.sin(4*np.pi*x)
y4 = np.exp(-2*x)
l3,l4 = ax2.plot(x, y3, 'yd-', x, y4, 'k^')
fig.legend((l1, l2), ('Line 1', 'Line 2'), 'upper center',bbox_to_anchor=(0.5,
0.90),ncol=2,fontsize=10,shadow=True)
fig.legend((l3, l4), ('Line 3', 'Line 4'), 'upper right')
```

图形

参数	默认值	描述
num	1	图形编号
figsize	figure.figsize	以英寸表示的图形大小 (宽、高)
dpi	figure.dpi	分辨率以每英寸点数表示
facecolor	figure.facecolor	背景色
edgecolor	figure.edgecolor	背景边缘色
frameon	True	是否绘制框架

子图

subplot(2,1,1)

subplot(2,1,2)

Axes 1

Axes 2

Axes 4

Axes 5

Axes 3

subplot(1,2,1)

subplot(1,2,2)

subplot(2,2,1)

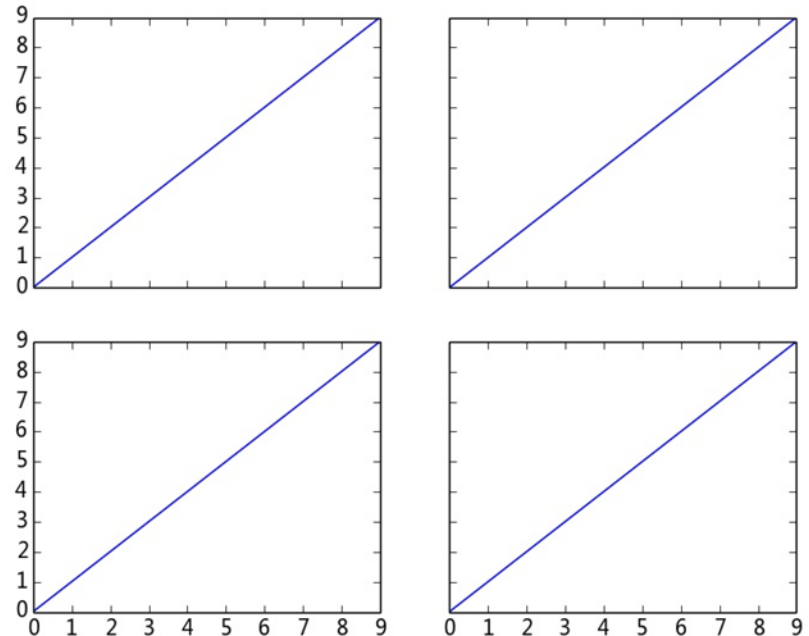
subplot(2,2,2)

subplot(2,2,3)

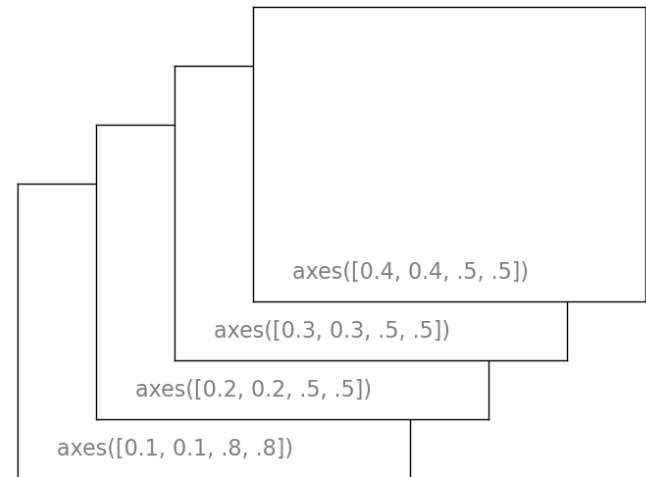
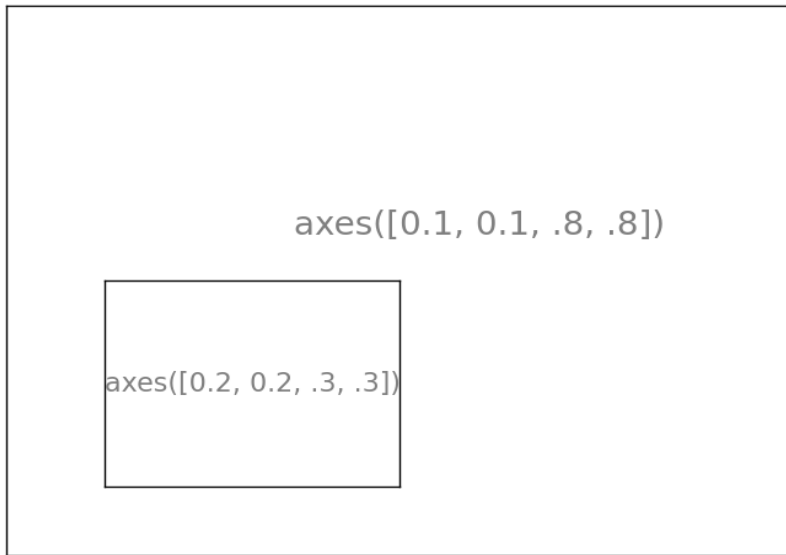
subplot(2,2,4)

共享x、y轴

```
import matplotlib.pyplot as plt
x = range(10)
y = range(10)
fig, ax=plt.subplots(nrows=2,ncols=2, sharex=True, sharey=True)
for row in ax:
    for col in row:
        col.plot(x, y)
plt.show()
```



轴





中國人民大學
RENMIN UNIVERSITY OF CHINA



02. 主要统计图

常用图

EMPID	Gender	Age	Sales	BMI	Income
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

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np
df = pd.read_excel('example.xlsx', 'Sheet1')
```

1、直方图

```
fig = plt.figure()  
ax = fig.add_subplot(111)  
ax.hist(df['Age'], bins=7)  
plt.title('Age distribution')  
plt.xlabel('Age')  
plt.ylabel('Employee')  
plt.show()
```

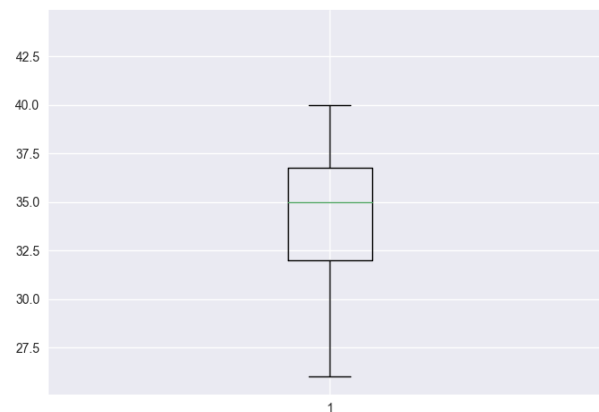
2、箱线图

```
fig = plt.figure()

ax = fig.add_subplot(111)

ax.boxplot(df['Age'])

plt.show()
```

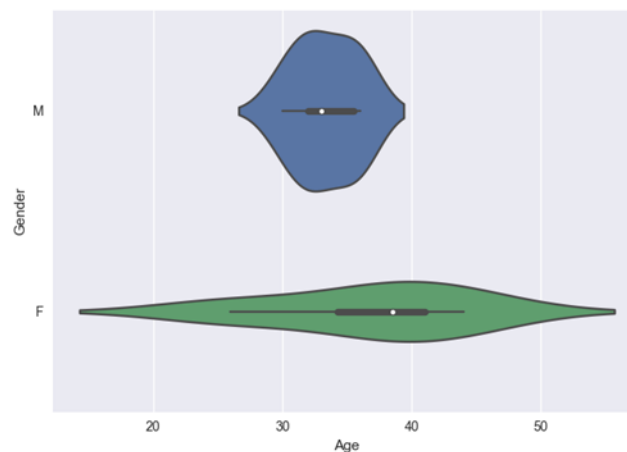


3、小提琴图

```
sns.violinplot(df['Age'], df['Gender'])
```

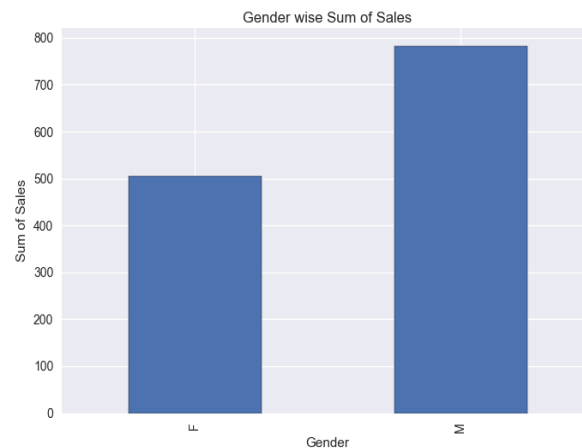
```
sns.despine()
```

```
plt.show()
```



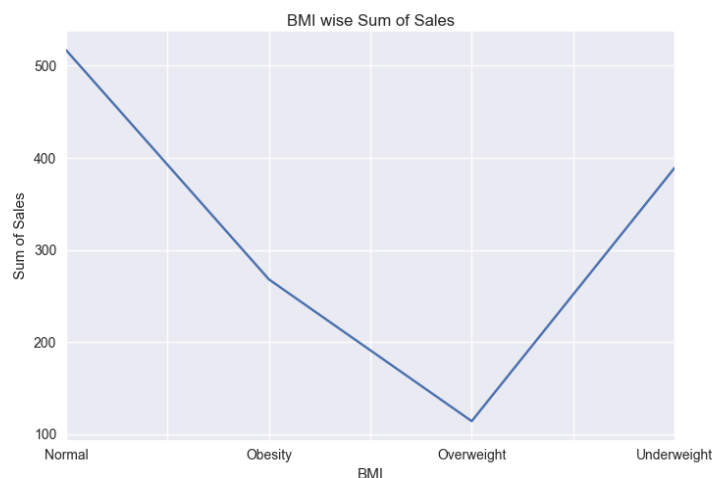
4、条形图

```
var = df.groupby('Gender').Sales.sum()
fig = plt.figure()
ax1 = fig.add_subplot(111)
ax1.set_xlabel('Gender')
ax1.set_ylabel('Sum of Sales')
ax1.set_title('Gender wise Sum of Sales')
var.plot(kind='bar')
plt.show()
```



5、折线图

```
var = df.groupby('BMI').Sales.sum()
fig = plt.figure()
ax = fig.add_subplot(111)
ax.set_xlabel('BMI')
ax.set_ylabel('Sum of Sales')
ax.set_title('BMI wise Sum of Sales')
var.plot(kind='line')
plt.show()
```

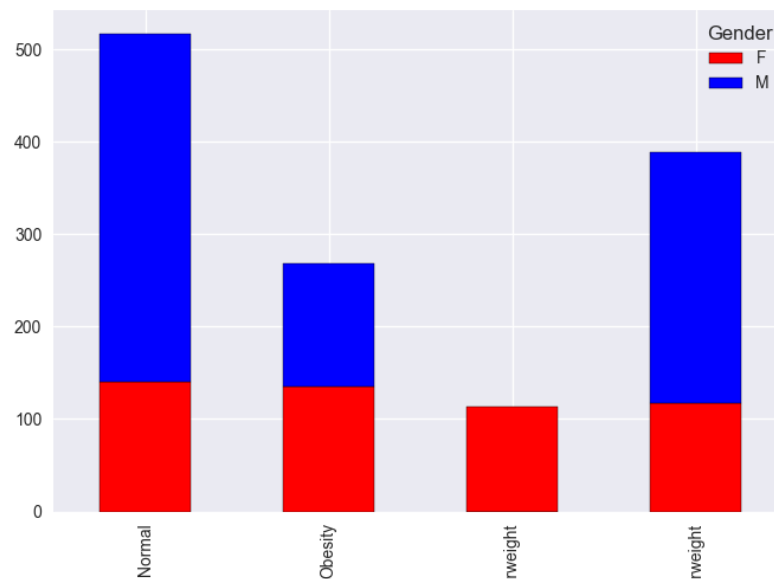


6、堆积柱形图

```
var = df.groupby(['BMI', 'Gender']).Sales.sum()
```

```
var.unstack().plot(kind='bar', stacked=True, color=['red', 'blue'])
```

```
plt.show()
```



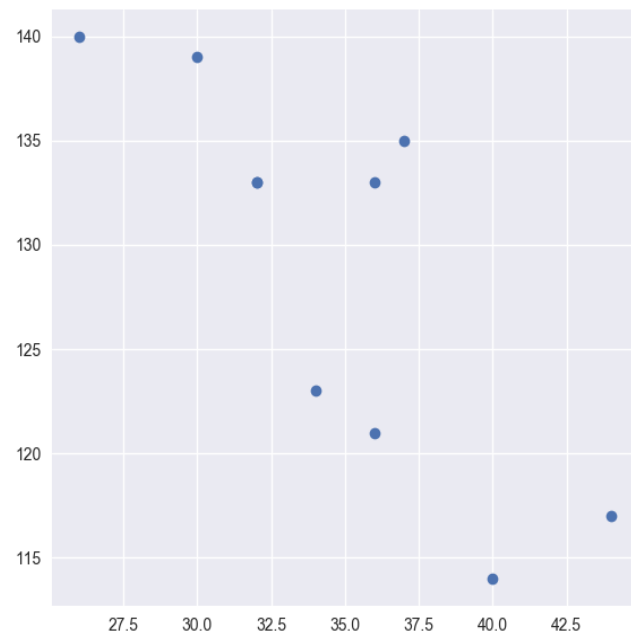
7、散点图

```
fig = plt.figure()

ax = fig.add_subplot(111)

ax.scatter(df['Age'], df['Sales'])

plt.show()
```



8、气泡图

```
fig = plt.figure()
```

```
ax = fig.add_subplot(111)
```

```
ax.scatter(df['Age'], df['Sales'], s=df['Income'])
```

```
# 第三个变量表明根据收入气泡的大小
```

```
plt.show()
```

9、饼图

```
var = df.groupby(['Gender']).sum().stack()
```

```
temp = var.unstack()
```

```
type(temp)
```

```
x_list = temp['Sales']
```

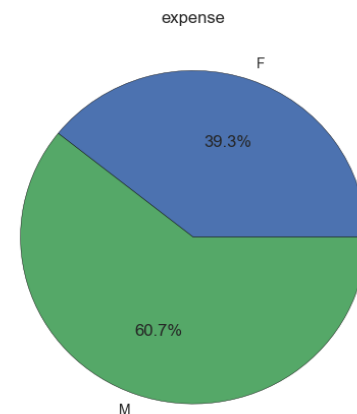
```
label_list = temp.index
```

```
plt.axis('equal')
```

```
plt.pie(x_list, labels=label_list, autopct='%')
```

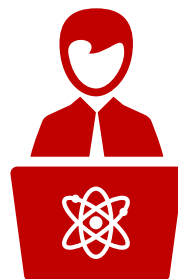
```
plt.title('expense')
```

```
plt.show()
```





中國人民大學
RENMIN UNIVERSITY OF CHINA



谢谢大家!

