



南通师范高等专科学校  
Nantong Normal College



# Matplotlib数据可视化之 图表绘制

朱亚林





# 散点图

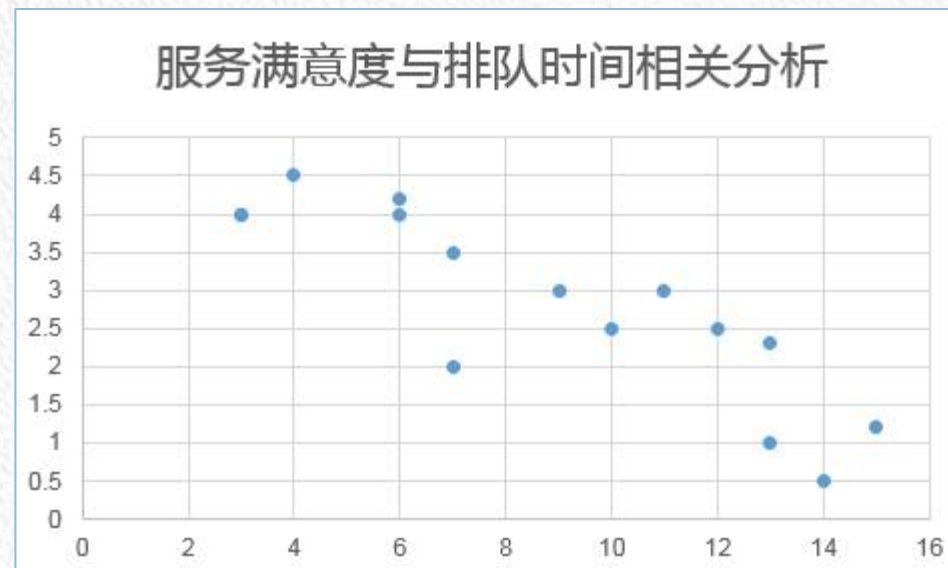
scatter的用法

■ 散点图是一种利用坐标点（散点）的分页形态反映特征间的统计关系的一种图形。值由点在图表中的位置表示，类别由图表中的不同标记表示，通常用于比较跨类别的数据。

■ 散点图一般可以提供以下两类关键信息。

（1）特征之间是否存在数值或数量的关联趋势，关联趋势是线性还是非线性的

（2）如果某一个点或某几个点偏离大多数点，则这些点是离群值。通过散点图可以一目了然。

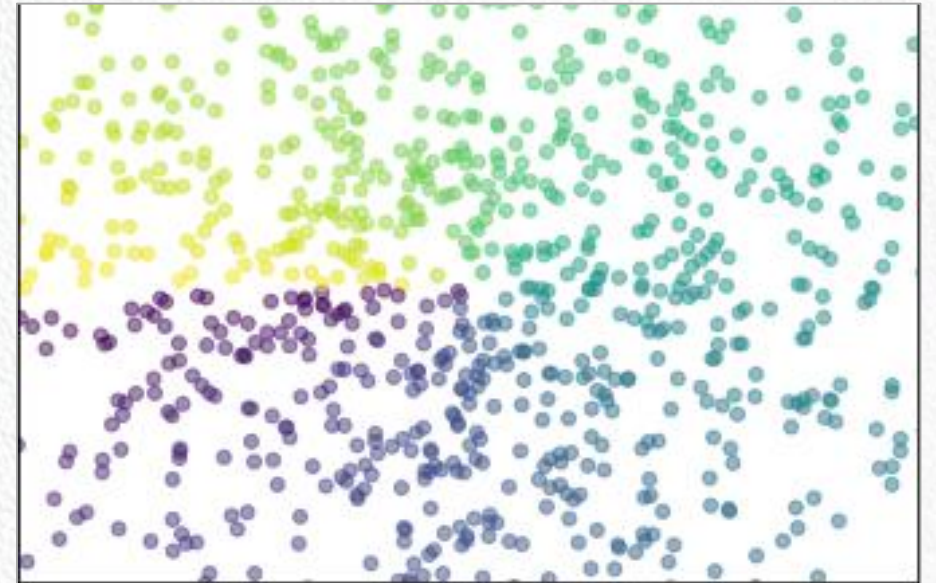


- 使用scatter()函数，可以绘制散点图像。
- scatter()函数常用的参数有以下几个。

参数	说明
x,y	接收要描绘的点的值
s	设置要描绘的点的大小，可以是一个数组
c	设置要描绘点的颜色，可以是一个数组
marker	接收一个string值，表示绘制的点的类型
alpha	接收一个0~1之间的小数，表示点的透明度

# 一个散点图的实例

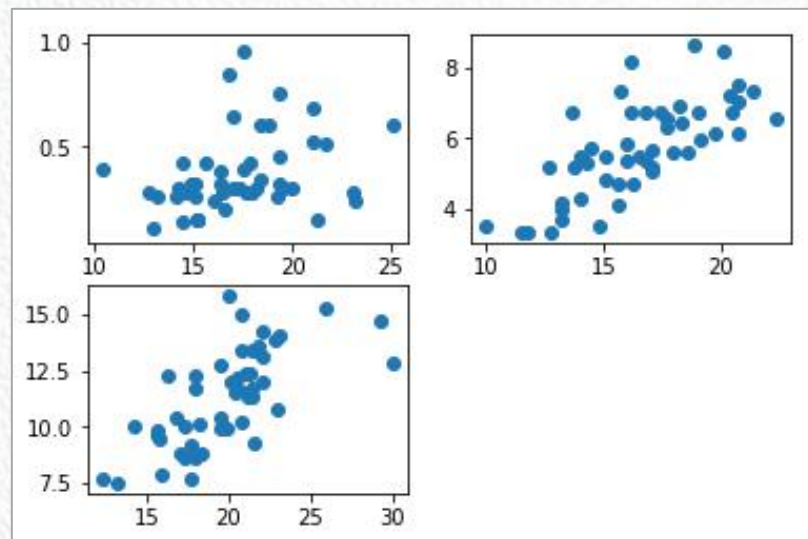
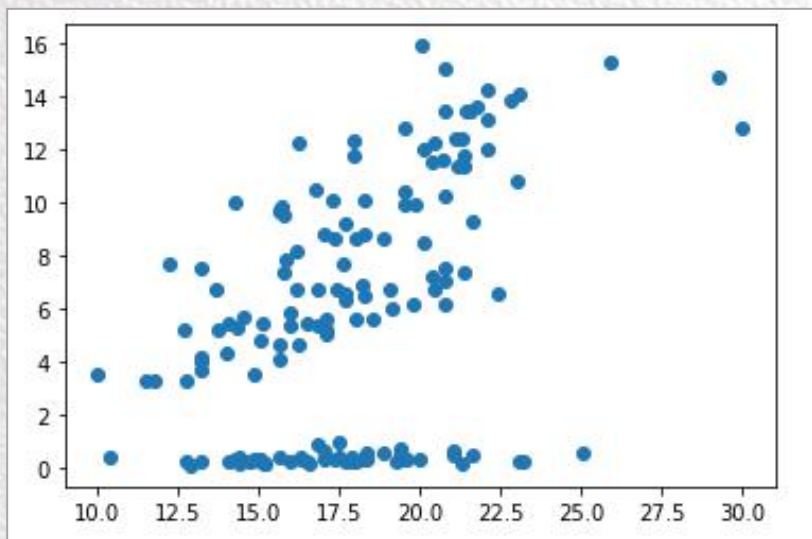
```
import matplotlib.pyplot as plt
import numpy as np
n = 1024
X = np.random.normal(0, 1, n)
Y = np.random.normal(0, 1, n)
T = np.arctan2(Y,X)
plt.scatter(X, Y, s=25, c=T, alpha=.5)
plt.xlim(-1.5, 1.5)
plt.xticks(())
plt.ylim(-1.5, 1.5)
plt.yticks(())
plt.show()
```



# 一个散点图练习

目标：通过对鸢尾花数据集中萼片和花瓣数据的分析以散点图的形式完成下列关系的呈现

- 萼片（sepal）和花瓣（petal）的大小关系
- 不同种类（species）鸢尾花萼片和花瓣的大小关系



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('file')
df['萼片尺寸']=df['萼片宽度(cm)']* df['萼片长度(cm)']
df['花瓣尺寸']=df['花瓣宽度(cm)']*df['花瓣长度(cm)']
plt.scatter(df['萼片尺寸'],df['花瓣尺寸'])
```

```
flag = 1
for name in df.groupby(['分类']).size().index:
    sepal_size = df[df['分类'].values == name]['萼片尺寸']
    petal_size = df[df['分类'].values == name]['花瓣尺寸']
    plt.subplot(2,2,flag)
    plt.scatter(sepal_size.values,petal_size.values)
    flag += 1
plt.show()
```



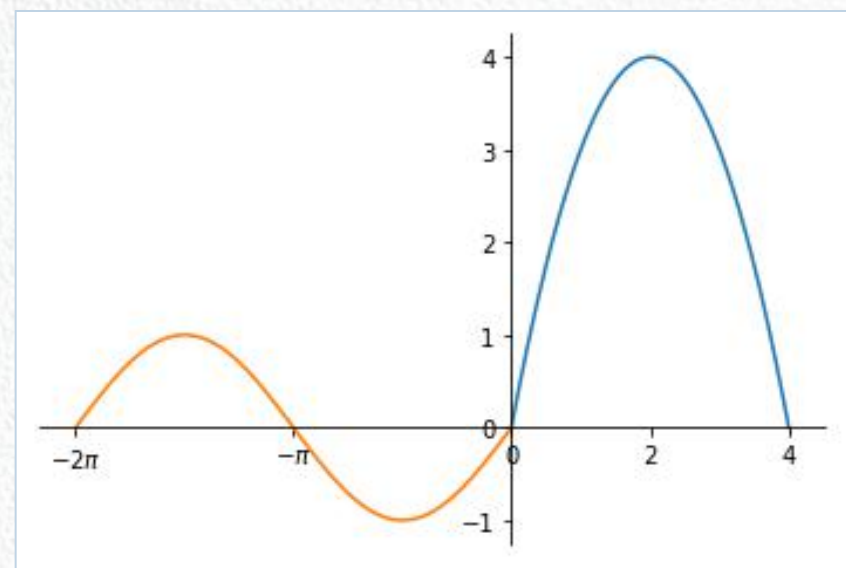
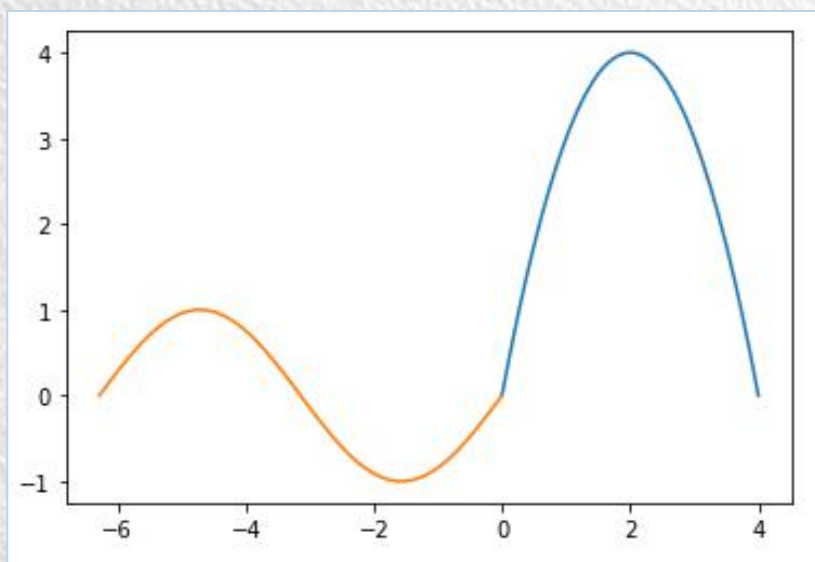
# 线形图

Still plot



# 回忆线形图，做个小练习

左边跟我一起画个龙，在你右边画一道彩虹



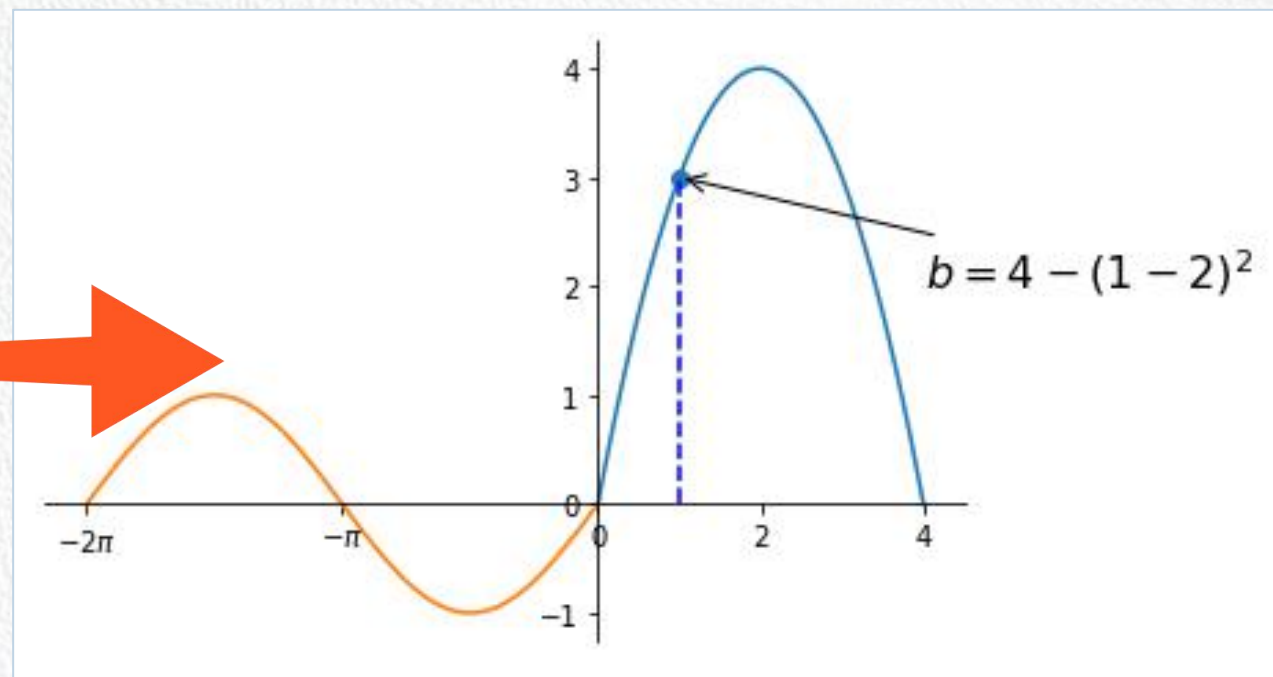
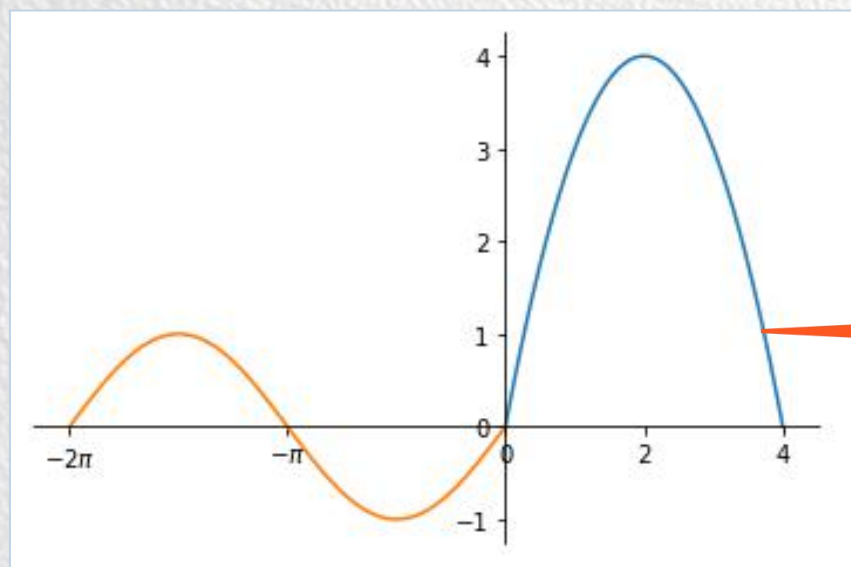
先画左边一幅图，对比思考右图有什么不同？

## 不同点一：坐标轴移动了

```
ax = plt.gca() #gca是get current axis
ax.spines['right'].set_color('none') #设置右边颜色无
ax.spines['top'].set_color('none') #设置顶边颜色无
ax.xaxis.set_ticks_position('bottom')
ax.spines['bottom'].set_position(('data', 0)) #设置底边在0上
ax.yaxis.set_ticks_position('left')
ax.spines['left'].set_position(('data',0))
```

不同点二：坐标轴上的数值是自定义的

```
plt.xticks([-2*np.pi,-np.pi,0,2,4],[' $-2\pi$ ',' $-\pi$ ',0,2,4])
```



有哪些变化？

- 在图像上指定了一点 (`scatter`)
- 过该点作出了一条垂直的线 (`plot`)
- 指向该点的函数表达式注释

```
plt.annotate(s="$b=4-(%s-2)^2 $"% a1, fontsize=16, xytext=(+4, +2),xy=(a1, b1),  
arrowprops=dict(arrowstyle="->"))
```

- 丨 **s**: 注释的内容，一段文字；
- 丨 **xytext**: 这段文字所处的位置；
- 丨 **xy**: 箭头指向的位置；
- 丨 **arrowprops**: 通过**arrowstyle**表明箭头的风格或种类。

学习链接：[公式的表达方式](#)

