



第7章 数据可视化

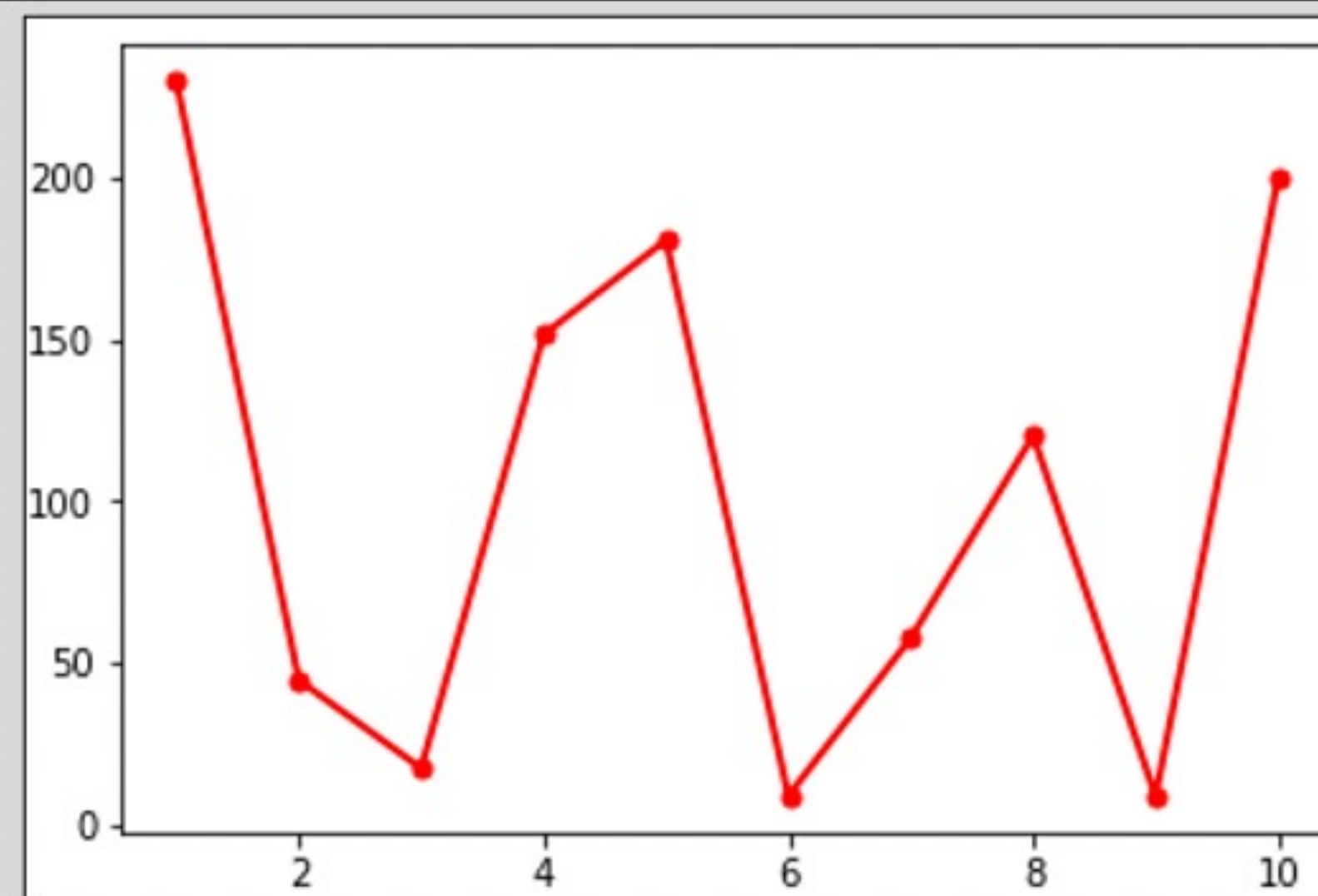
Jupyter金融应用从入门到实践

示例中plot()函数使用了两遍：第一遍是用于绘制折线，参数依次是x轴的数值、y轴的值数值、线的颜色（其中“r”代表红色，“-”代表实线）、线的粗细；第二遍是用于绘制坐标点，参数依次是x轴的数值、y轴的数值、坐标点的颜色（其中“r”代表红色，“o”代表实心点）、点的大小。

[In]

```
1. import os
2. import pandas as pd
3. import matplotlib.pyplot as plt
4. path = "dataset/advertising.csv"
5. data = pd.read_csv(path, sep=',')
6. df = data.head(10)
7. l1 = list(df['TV'])
8. x = [1,2,3,4,5,6,7,8,9,10]
9. plt.figure()
10. plt.plot(x, l1, "r-", linewidth=2)
11. plt.plot(x, l1, "ro", markersize=5)
12. plt.show()
```

[Out]

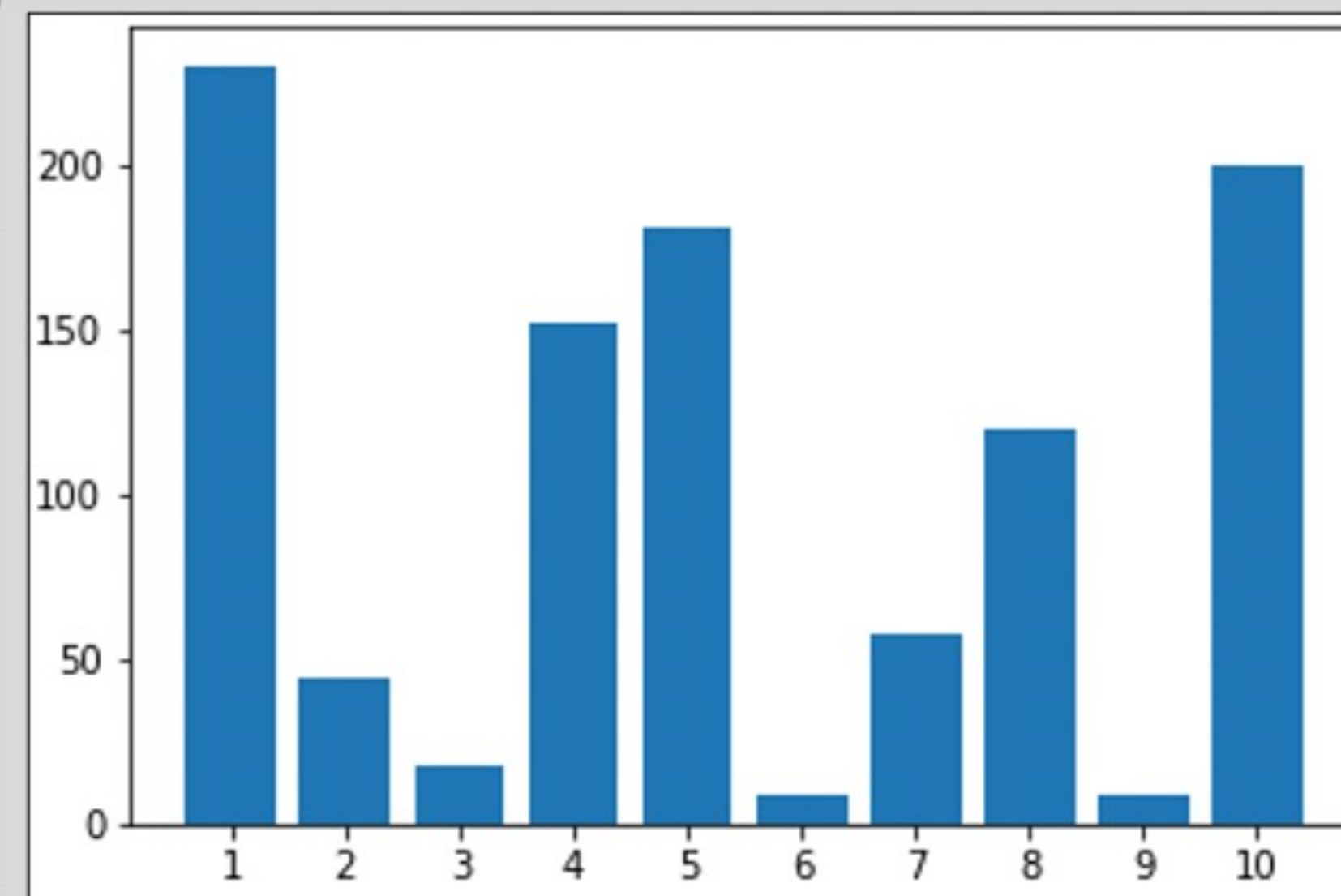


柱状图通过`bar()`函数绘制，同样设有x轴和y轴参数，也有额外的参数（例如`tick_label`）用于定义标签。输入的数据类型与折线图相同。

[In]

```
1. plt.figure()  
2. plt.bar(x, l1, tick_label=x)  
3. plt.show()
```

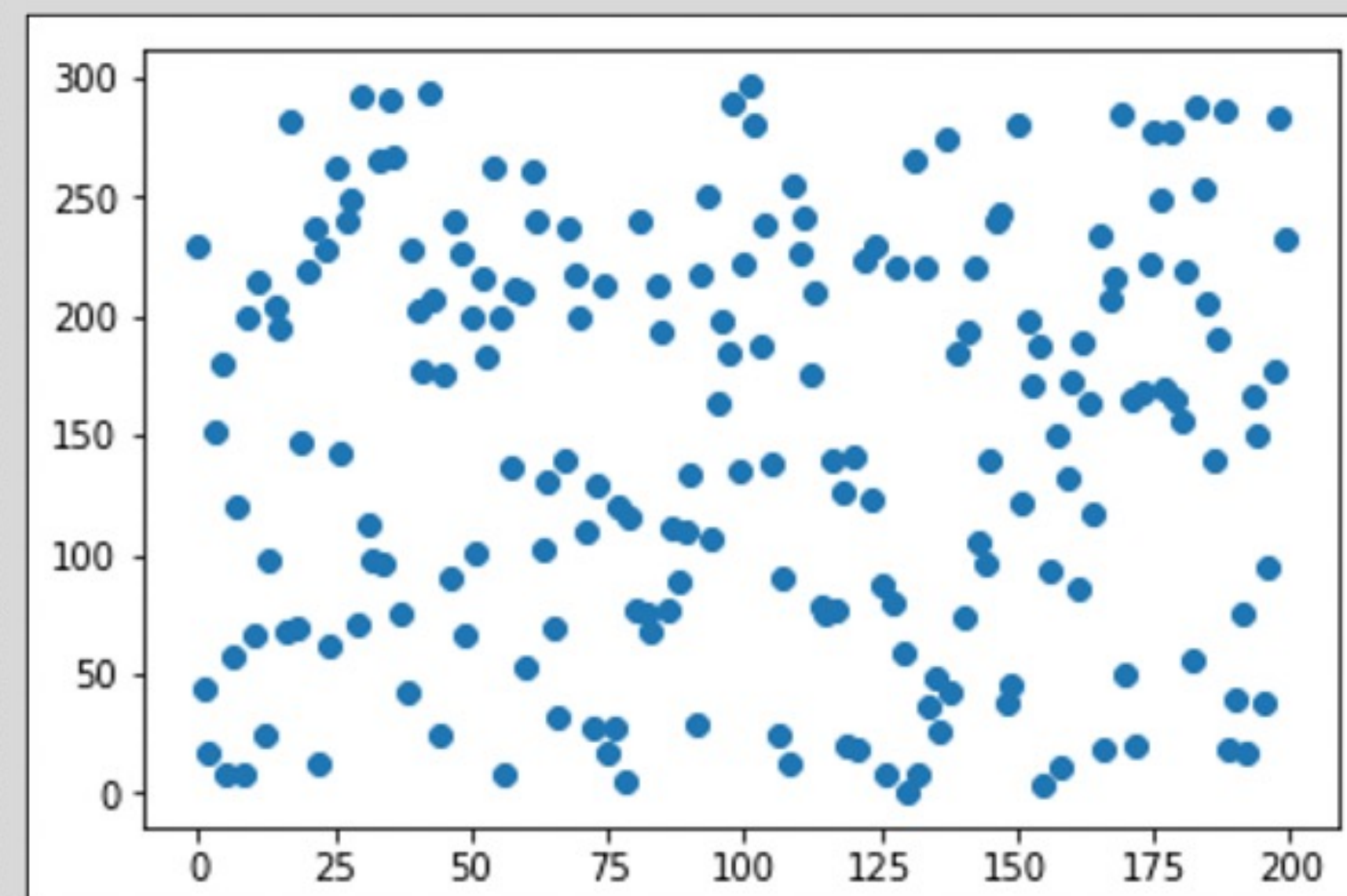
[Out]



散点图一般包括大量数据，用于观察数据的分布以及数据与x轴、y轴（甚至z轴）的关联，Matplotlib通过scatter()函数来绘制散点图。

```
[In] 1. import os
      2. import pandas as pd
      3. import matplotlib.pyplot as plt
      4. path = "dataset/advertising.csv"
      5. data = pd.read_csv(path, sep=',')
      6. l1 = list(data['TV'].index)
      7. l2 = list(data['TV'])
      8. plt.figure()
      9. plt.scatter(l1,l2)
     10. plt.show()
```

[Out]

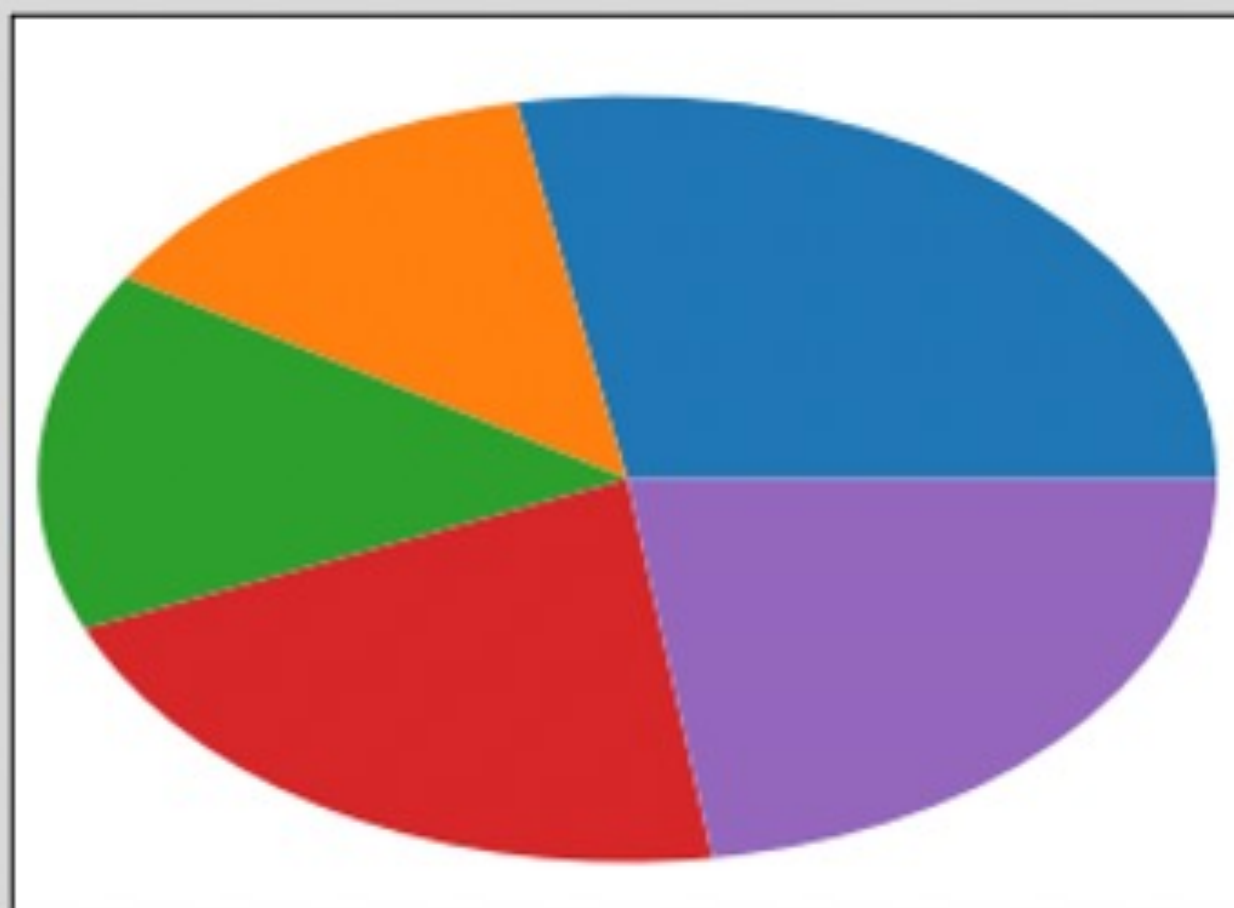


饼状图的共同特点是由多个扇形组成，通常用于展示各个类别所占的比例。饼状图的一般展示形式有饼图和环形图。饼图可以使用Matplotlib中的pie()函数来绘制饼图。

[In]

```
1. import os
2. import pandas as pd
3. import matplotlib.pyplot as plt
4. path = "dataset/advertising.csv"
5. data = pd.read_csv(path, sep=',')
6. data = data.head()
7. l1 = list(data['Sales'])
8. plt.figure()
9. plt.pie(l1)
10. plt.show()
```

[Out]

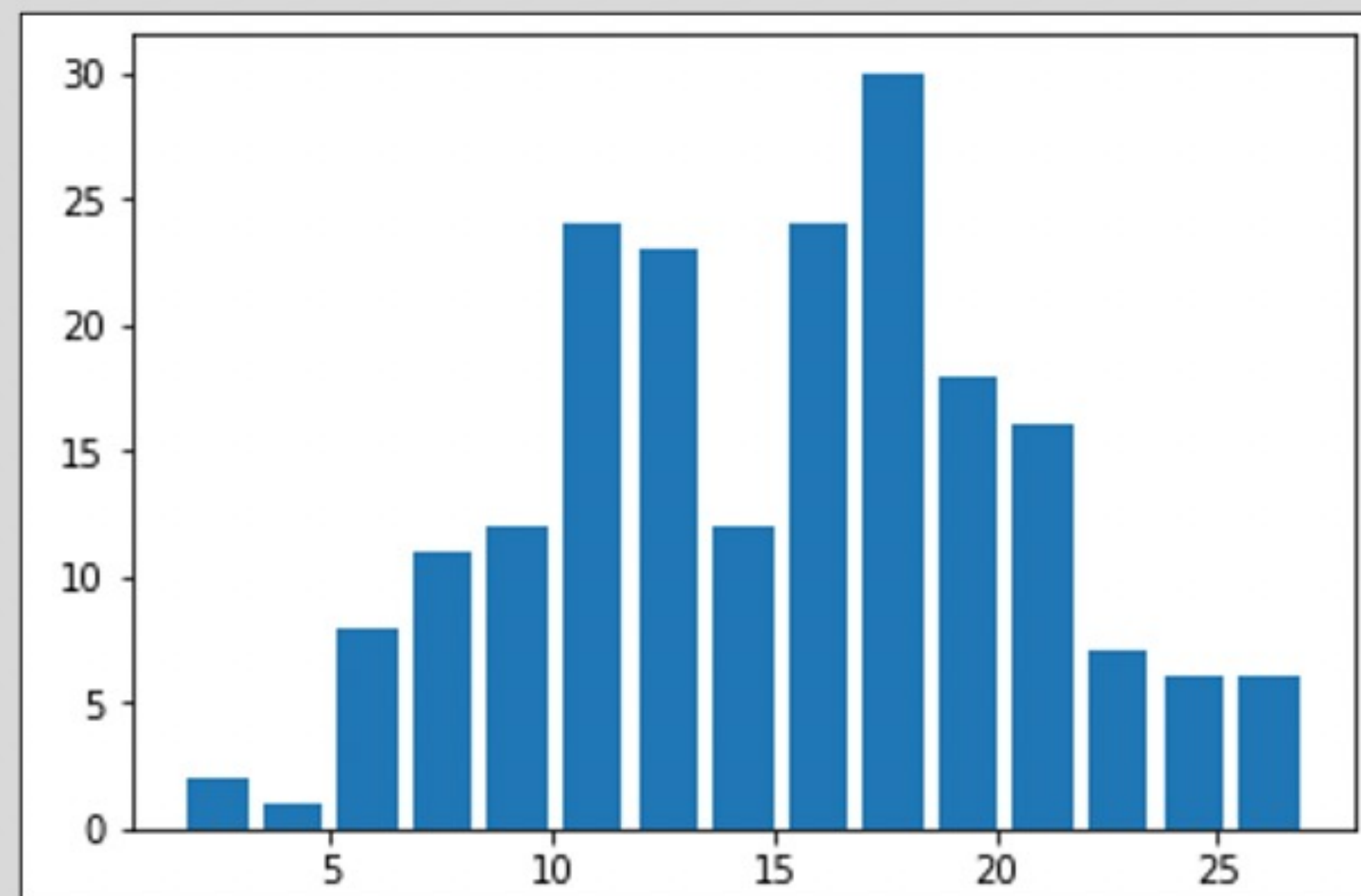


直方图用于展示数据分布的情况，横轴是该变量的取值范围分段，纵轴是范围内存在的变量数目（频率），在统计中有重要意义。Matplotlib中通过hist()函数来绘制直方图。

[In]

```
1. import os
2. import pandas as pd
3. import matplotlib.pyplot as plt
4. path = "dataset/advertising.csv"
5. data = pd.read_csv(path, sep=',')
6. l1 = list(data['Sales'])
7. plt.figure()
8. plt.hist(l1, rwidth=0.8, bins=15)
9. plt.show()
```

[Out]

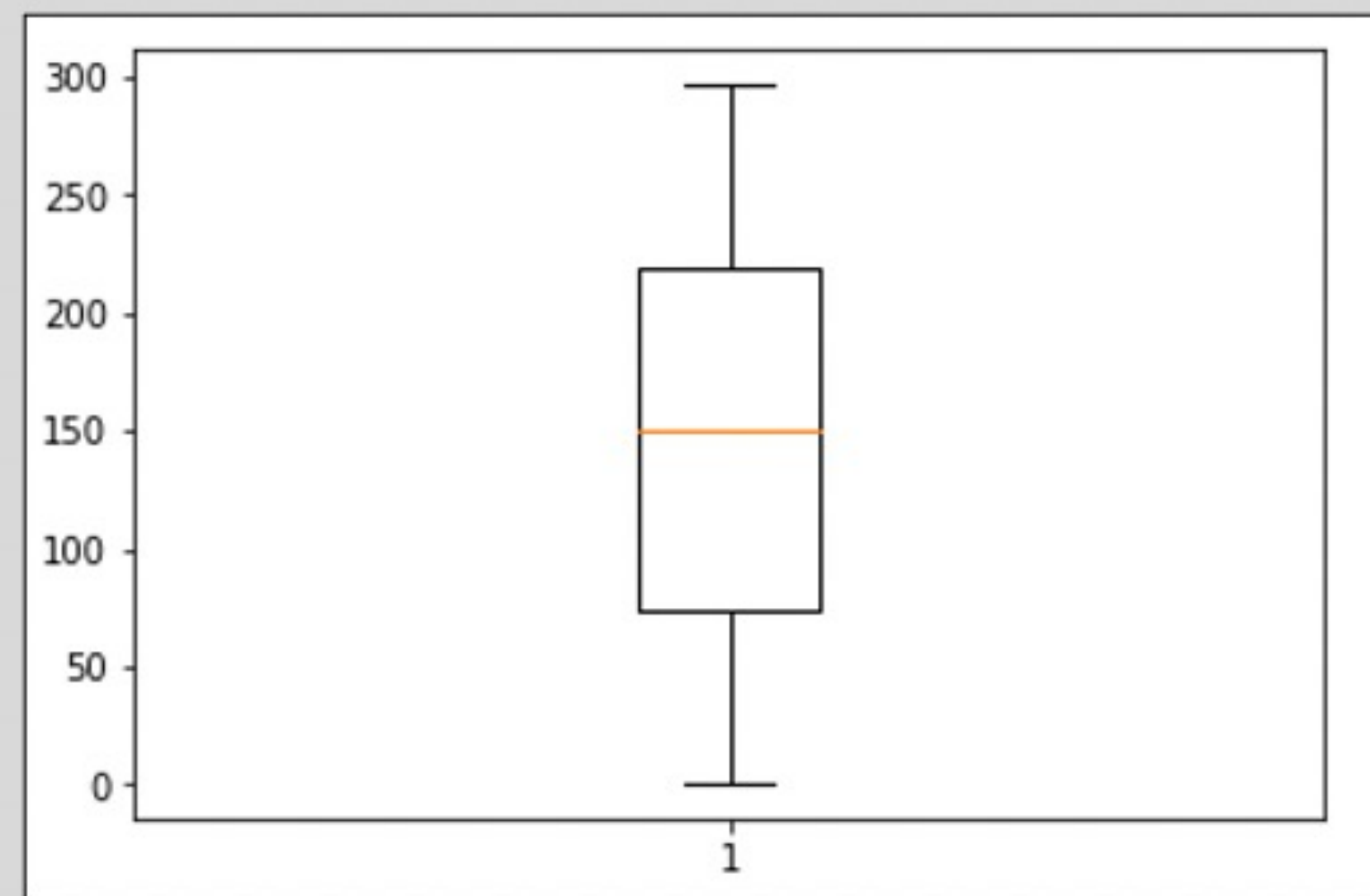


箱型图用于观察数据的分散情况，可以很直观地观察异常值、分位数、中位数等，Matplotlib中使用boxplot()函数来绘制箱型图。

[In]

```
1. import os
2. import pandas as pd
3. import matplotlib.pyplot as plt
4. path = "dataset/advertising.csv"
5. data = pd.read_csv(path, sep=',')
6. l1 = list(data['TV'])
7. plt.figure()
8. plt.boxplot(l1)
9. plt.show()
```

[Out]

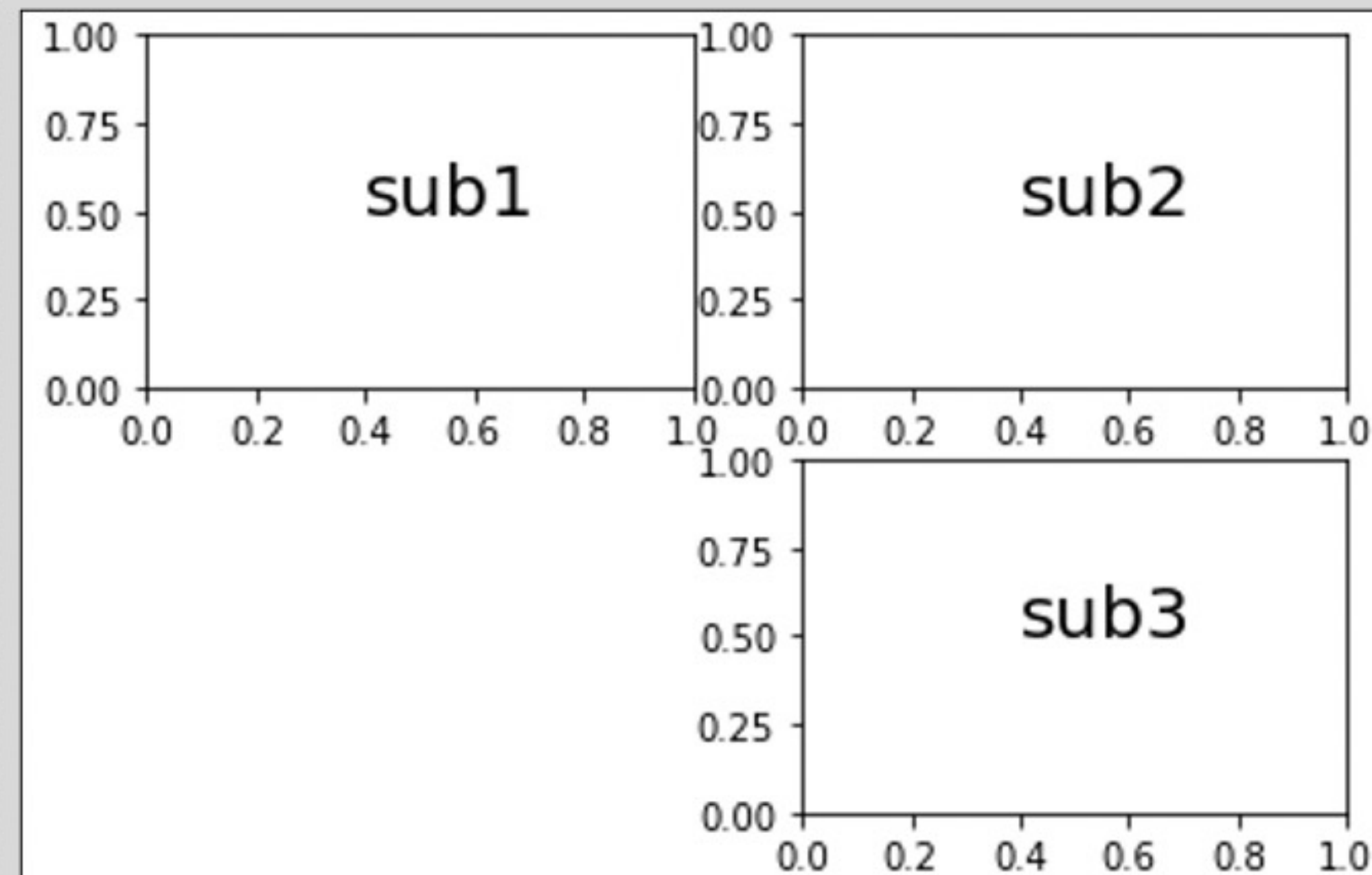


- 想要同时在画布上并列绘制多个图形，则需要使用subplot()函数。两者之间的关系是包含关系：figure是绘制subplot的区域，subplot是绘制图形的区域。
- 想要在同一figure或subplot中同时绘制多个图形，则只需要进行叠加即可

[In]

```
1. fig = plt.figure()
2. sub1 = fig.add_subplot(2,2,1)
3. sub1.text(0.4,0.5,'sub1',fontsize=20)
4. sub2 = fig.add_subplot(2,2,2)
5. sub2.text(0.4,0.5,'sub2',fontsize=20)
6. sub3 = fig.add_subplot(2,2,4)
7. sub3.text(0.4,0.5,'sub3',fontsize=20)
8. plt.show()
```

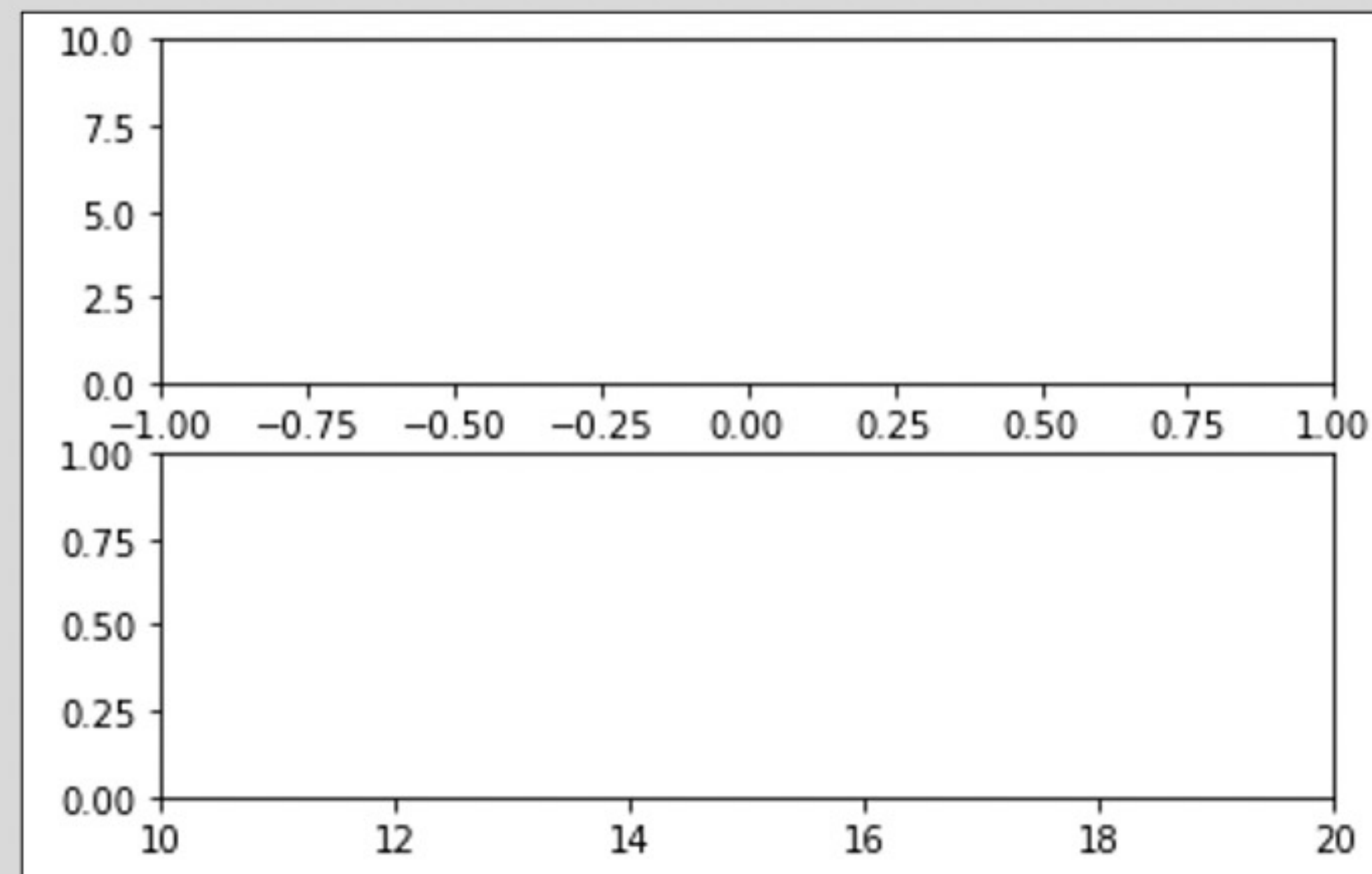
[Out]



- 对于横纵坐标的刻度范围，如果使用 `add_subplot()` 函数逐个建立子图，则可以使用 `set_xlim()` 和 `set_ylim()` 设置坐标范围，内部参数为下界和上界。

```
[In] 1. import os
      2. import matplotlib.pyplot as plt
      3. fig = plt.figure()
      4. sub1 = fig.add_subplot(2,1,1)
      5. sub1.set_xlim(-1,1)
      6. sub1.set_ylim(0,10)
      7. sub2 = fig.add_subplot(2,1,2)
      8. sub2.set_xlim(2,5)
      9. sub2.set_xlim(10,20)
     10. plt.show()
```

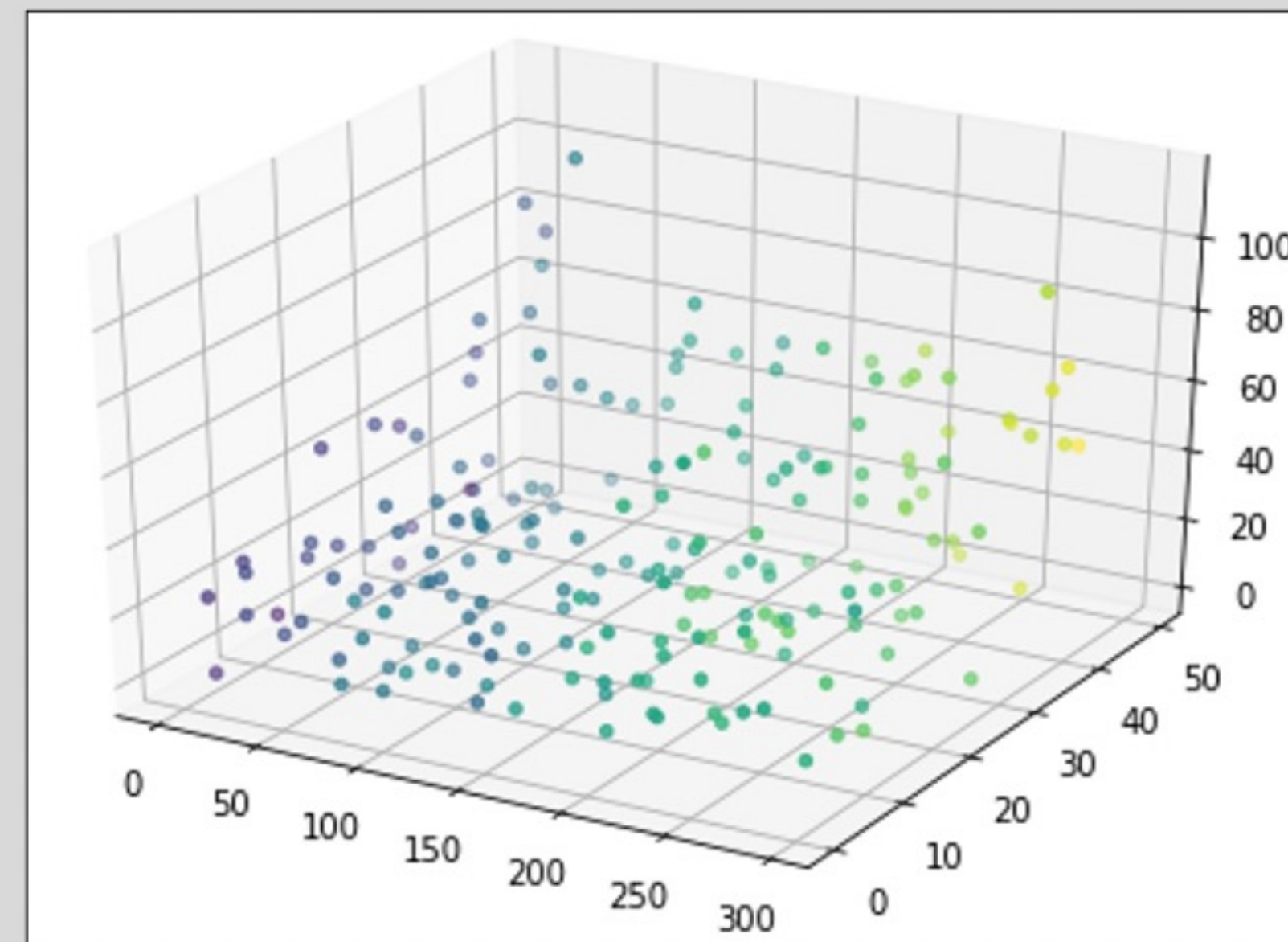
[Out]



- 在进行完包的引入和3D画布的设置后，直接使用前文介绍过的绘图函数进行绘制即可。

```
[In] 1. import pandas as pd
      2. import matplotlib.pyplot as plt
      3. from mpl_toolkits.mplot3d import axes3d
      4. path = "dataset/advertising.csv"
      5. df = pd.read_csv(path, sep=',')
      6. l1 = list(df['TV'])
      7. l2 = list(df['Radio'])
      8. l3 = list(df['Newspaper'])
      9. l4 = list(df['Sales'])
     10. fig = plt.figure()
     11. ax = Axes3D(fig)
     12. ax.scatter(l1, l2, l3, c=l4, marker='.', s=50)
     13. plt.show()
```

[Out]





谢谢！