

附录：散点图

2023 年 9 月 18 日

5.3 散点图

散点图（Scatter Diagram）又称为散点分布图，是以利用坐标点（散点）的分布形态反映特征间的相关关系的一种图形。实际中一般使用二维散点图，通过散点的疏密程度和变化趋势表示两个特征间关系。散点图有以下 3 个特点。1. 表现特征之间是否存在数值或者数量的关联趋势，关联趋势是线性的还是非线性的。2. 凸显出离群点（异常点），及对整体的影响。3. 数据量越大，能发挥的作用越好。pyplot 模块中使用 scatter 函数绘制散点图，其使用语法如下：

```
matplotlib.pyplot.scatter(x, y, s=None, c=None, marker=None,  
                           cmap=None, norm=None, vmin=None,  
                           vmax=None, alpha=None, linewidths=None,  
                           verts=None, edgecolors=None, hold=None,  
                           data=None, **kwargs)
```

参数名称	说明
x, y	接收 array。表示 x 轴和 y 轴对应的数据。无默认值
s	接收数值或者一维的 array。指定点的大小，若传入一维 array 则表示每个点的大小。默认为 None
c	接收颜色或者一维的 array。指定点的颜色，若传入一维 array 则表示每个点的颜色。默认为 None
marker	接收特定 str。表示绘制的点的类型。默认为 None
alpha	接收 0~1 的小数。表示点的透明度。默认为 None

以datasets/ .csv为例子:

```
[1]: import pandas as pd
df = pd.read_csv('datasets/国民经济核算季度数据.csv', index_col=0)
df.head()
```

[1]: 国内生产总值 _ 当季值 (亿元) 第一产业增加值 _ 当季值 (亿元) 第二产业增加值 _ 当季值 (亿元) 第三产业增加值 _ 当季值 (亿元) \

2000 年第一季度	21329.9	1908.3	9548.0	↪
↪9873.6				
2000 年第二季度	24043.4	3158.2	11127.5	↪
↪9757.7				
2000 年第三季度	25712.5	4140.6	11887.0	↪
↪9684.9				
2000 年第四季度	29194.3	5510.2	13102.3	↪
↪10581.7				
2001 年第一季度	24086.4	2015.3	10641.7	↪
↪11429.4				

 农林牧渔业增加值 _ 当季值 (亿元) 工业增加值 _ 当季值 (亿元) 建筑业增加值 _ 当季值 (亿元) 批发和零售业增加值 _ 当季值 (亿元) \

2000 年第一季度	1947.5	8798.7	777.1	↪
↪2100.9				
2000 年第二季度	3209.7	9799.9	1359.0	↪
↪2073.0				
2000 年第三季度	4196.1	10503.1	1417.4	↪
↪1943.2				
2000 年第四季度	5590.2	11158.0	1980.5	↪
↪2041.5				
2001 年第一季度	2061.6	9838.3	834.2	↪
↪2351.6				

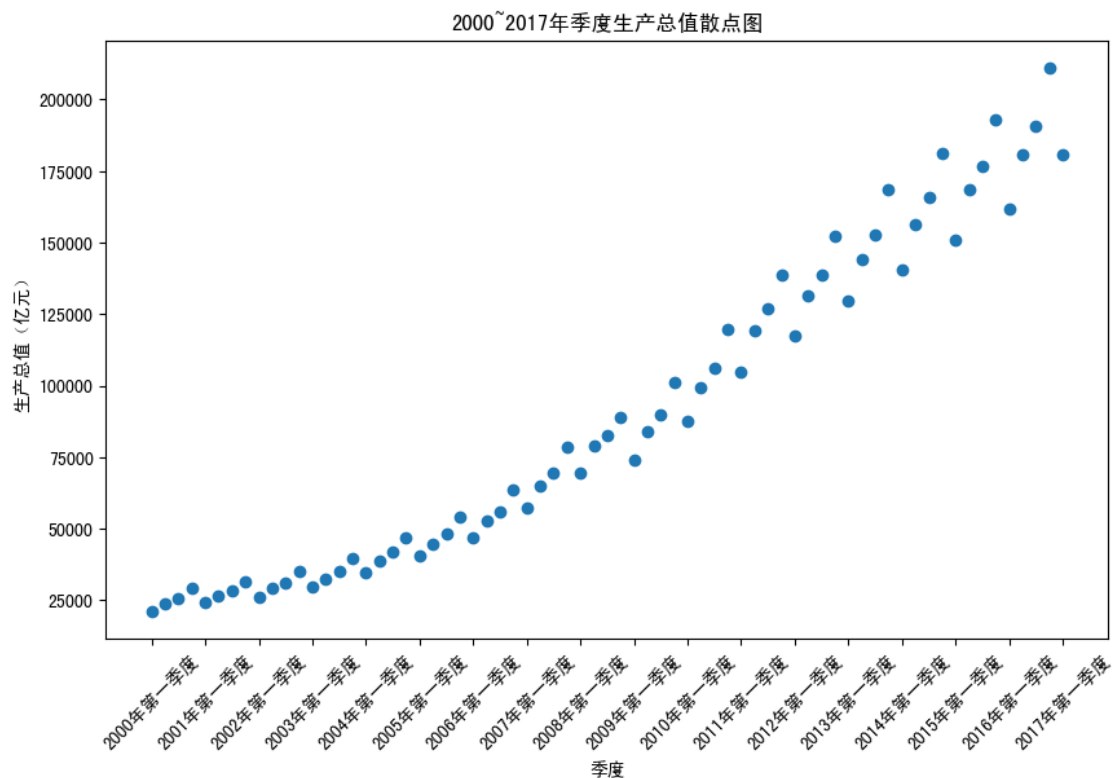
 交通运输、仓储和邮政业增加值 _ 当季值 (亿元) 住宿和餐饮业增加值 _ 当季值 (亿元) 金融业增加值 _ 当季值 (亿元) \

2000 年第一季度	1379.4	570.5	1235.9
2000 年第二季度	1571.7	536.5	1124.0

2000 年第三季度	1370.0	523.2	1170.4
2000 年第四季度	1840.8	516.1	1306.0
2001 年第一季度	1604.1	638.6	1341.5

	房地产业增加值 _ 当季值 (亿元)	其他行业增加值 _ 当季值 (亿元)
2000 年第一季度	933.7	3586.1
2000 年第二季度	904.7	3464.9
2000 年第三季度	1070.9	3518.2
2000 年第四季度	1239.7	3521.5
2001 年第一季度	1074.4	4342.2

```
[2]: # step1 -----
import matplotlib.pyplot as plt
# step2 -----
X = df.index.values
Y = df['国内生产总值 _ 当季值 (亿元)']
# step3 -----
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
fig, ax = plt.subplots(figsize = (10, 6))
ax.scatter(X, Y, marker = 'o') # 散点图的语句
ax.set_xticks(range(0, len(X), 4), df.index[range(0, len(X), 4)], rotation = 45)
ax.set_xlabel('季度')
ax.set_ylabel('生产总值 (亿元)')
ax.set_title('2000~2017 年季度生产总值散点图')
plt.show()
```



```
[3]: range(0, len(X),4), df.index[range(0, len(X), 4)])
```

```
[3]: (range(0, 69, 4),
      Index(['2000 年第一季度', '2001 年第一季度', '2002 年第一季度', '2003 年第一季度',
            '2004 年第一季度',
            '2005 年第一季度', '2006 年第一季度', '2007 年第一季度', '2008 年第一季度',
            '2009 年第一季度',
            '2010 年第一季度', '2011 年第一季度', '2012 年第一季度', '2013 年第一季度',
            '2014 年第一季度',
            '2015 年第一季度', '2016 年第一季度', '2017 年第一季度'],
            dtype='object'))
```

绘制 2000~2017 年第一产业、第二产业、第三产业各季度的国民生产总值散点图，如下代码所示。

```
[4]: # step1 -----
import matplotlib.pyplot as plt
# step2 -----
```

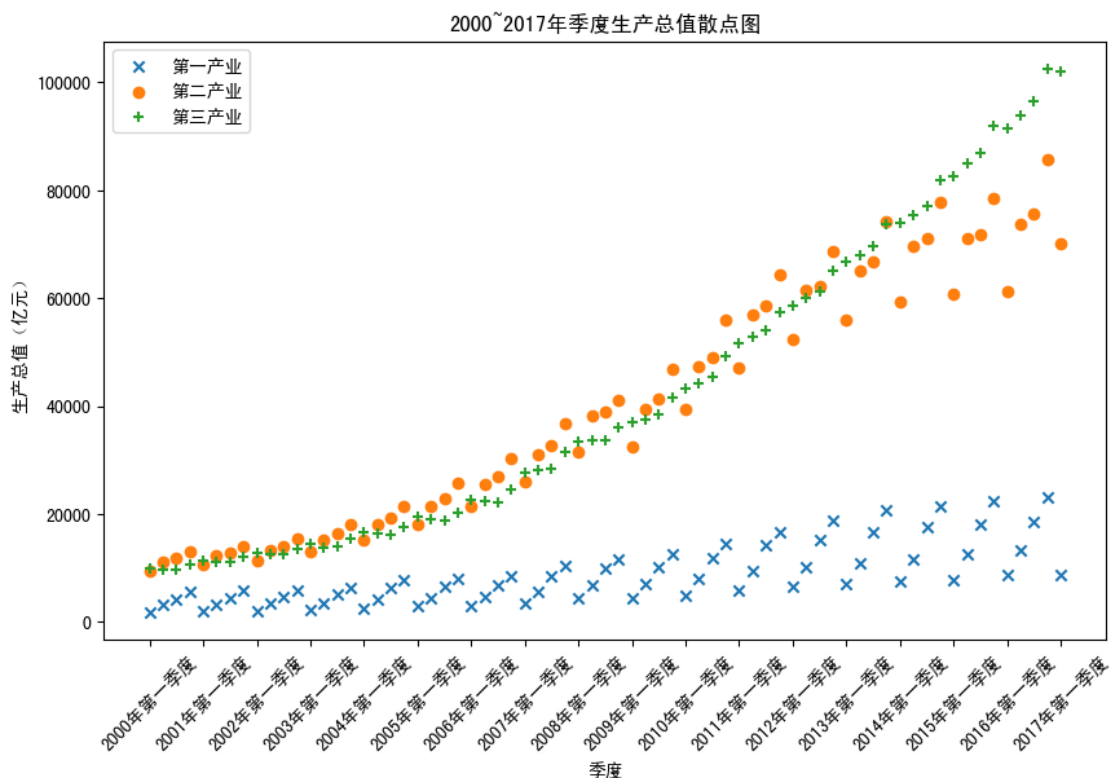
```

X = df.index.values
Y1 = df['第一产业增加值 _ 当季值 (亿元)']
Y2 = df['第二产业增加值 _ 当季值 (亿元)']
Y3 = df['第三产业增加值 _ 当季值 (亿元)']

# step3 -----
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
fig, ax = plt.subplots(figsize = (10, 6))
ax.scatter(X, Y1, marker = 'x')           # 散点图的语句
ax.scatter(X, Y2, marker = 'o')         # 散点图的语句
ax.scatter(X, Y3, marker = '+')         # 散点图的语句
plt.xlabel('季度')
plt.ylabel('生产总值 (亿元)')
plt.xticks(range(0, len(X), 4), df.index[range(0, len(X), 4)], rotation = 45)
plt.legend(['第一产业', '第二产业', '第三产业'])
plt.title('2000~2017 年季度生产总值散点图')
# plt.savefig('三种产业散点图.png')

```

[4]: Text(0.5, 1.0, '2000~2017 年季度生产总值散点图')



[]: