计算机 22A4 2022015232 孙春辉
2025 年 4 月 2 日 星期三 第八次作业

Report 8

Part 1: Course exercises from reading material Ch6.1

1.1: Line2D

1.2: 堆积图

1.3: 10

2.1: F

2.2: T

2.3: F

3.1: D

3.2: C

3.3: A

3.4: D

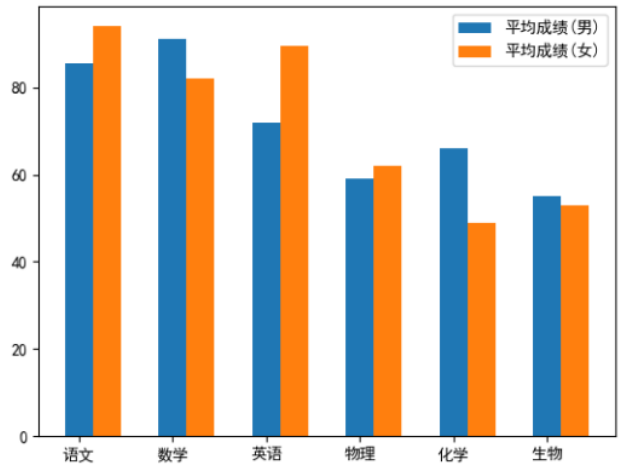
3.5: C

4.1:

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

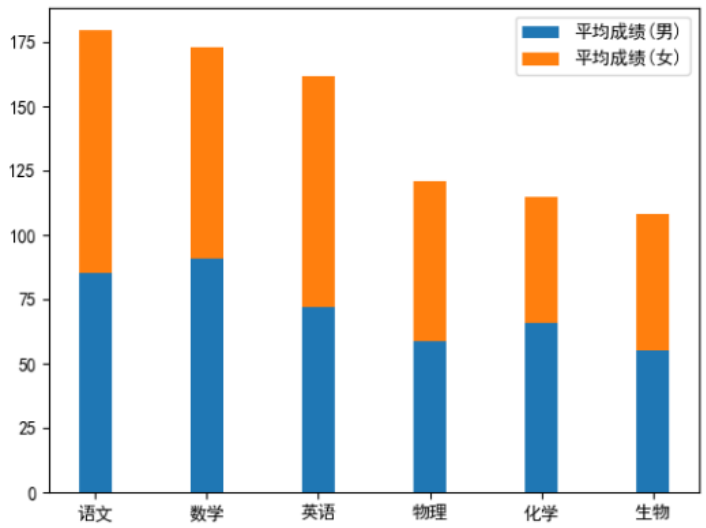
计算机 22A4 2022015232 孙春辉
2025 年 4 月 2 日 星期三 第八次作业

```
df = pd.DataFrame({'学科':['语文','数学','英语','物理','化学','生物'],
                   '平均成绩(男)':[85.5,91,72,59,66,55],
                   '平均成绩(女)':[94,82,89.5,62,49,53]})
print(df)

x = np.arange(6)
x1 = df['学科']
y1 = np.array(df['平均成绩(男)'])
y2 = np.array(df['平均成绩(女)'])

bar_width = 0.3

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.bar(x,y1,tick_label=x1,width=bar_width)
plt.bar(x+bar_width,y2,width=bar_width)

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



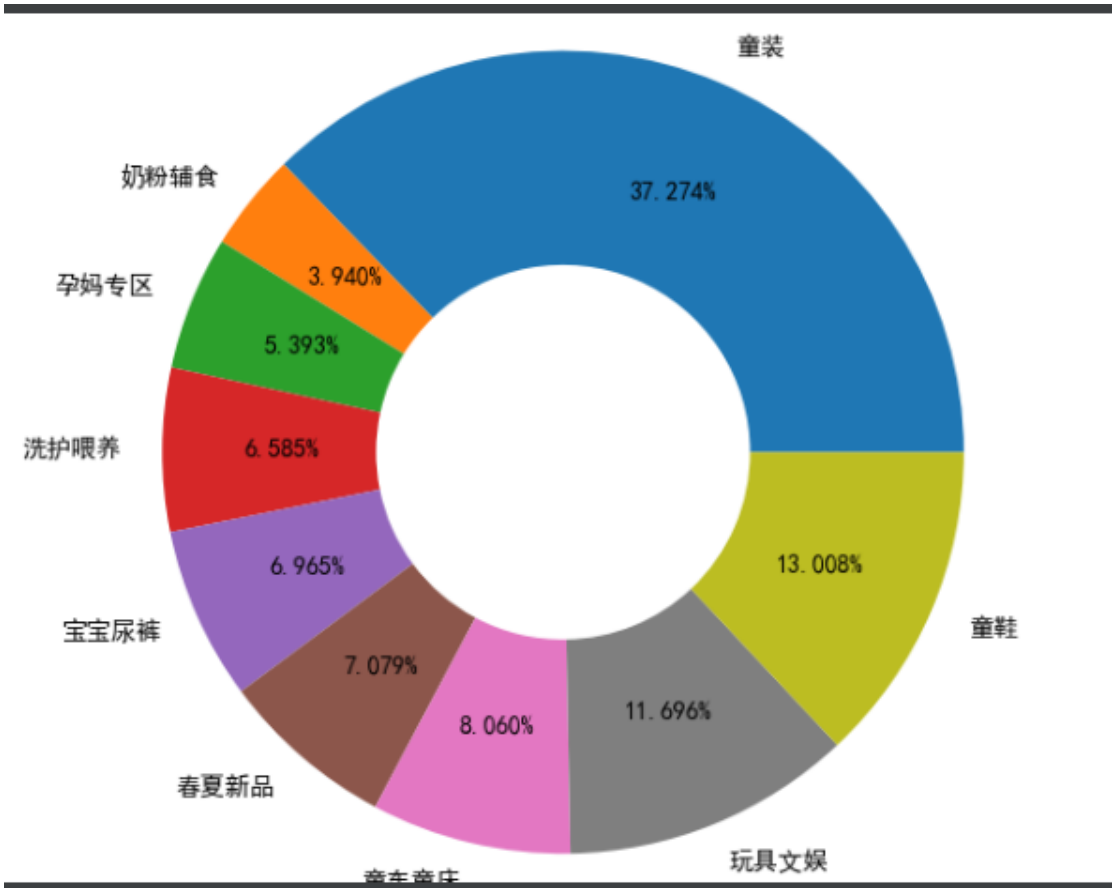

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


df = pd.DataFrame({'学科':['语文','数学','英语','物理','化学','生物'],
                   '平均成绩(男)':[85.5,91,72,59,66,55],
```

```
                    '平均成绩(女)':[94,82,89.5,62,49,53]})
print(df)

x = np.arange(6)
x1 = df['学科']
y1 = np.array(df['平均成绩(男)'])
y2 = np.array(df['平均成绩(女)'])

bar_width = 0.3

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.bar(x,y1,tick_label=x1,width=bar_width,label='平均成绩(男)')
# 通过修改第二个柱状图的起始 y 坐标为上一个数据的 值，最终就能实现柱状图的堆叠
plt.bar(x,y2,bottom=y1,width=bar_width,label='平均成绩(女)')

# plt.bar(x,y1,tick_label=x1,width=bar_width,label='平均成绩
(男)',color='blue')
# plt.bar(x+bar_width,y2,width=bar_width,label='平均成绩
(女)',color='red')
# plt.legend(['平均成绩(男)'])
plt.legend()

plt.show()
```



4.2:

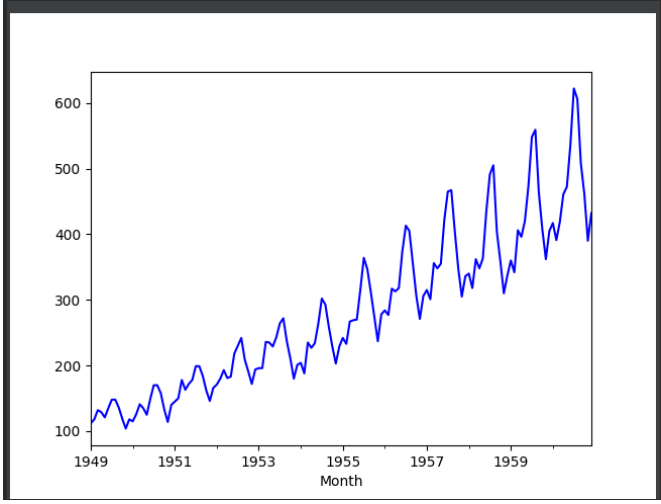
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt


df = pd.DataFrame({'子类目':['童装','奶粉辅食','孕妈专区','洗护喂养','宝宝
尿裤','春夏新品','童车童床','玩具文娱','童鞋'],
                   '销售额
':[29665,3135.4,4292.4,5240.9,5543.4,5633.8,6414.5,9308.1,10353]})

name = np.array(df['子类目'])
data = np.array(df['销售额'])
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.pie(data,radius=1.5,wedgeprops={'width' : 0.8},labels=name,
        autopct='%0.3lf%%',pctdistance=0.7)
plt.show()
```



Part 2:

Use matplotlib to implement data visualization tasks in the Assignment7.xlsx on Moodle. You need to choose 1 method to plot the data, explain why you choose this method and analyze what you see from the charts in the report.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv(r'D:\schoolzwu\AI\machine-study\data_analyse\3-30\Report7-AirPassengers.csv', encoding='utf-8', index_col='Month')
df.index = pd.to_datetime(df.index)  # 添加频率信息
ts = df['#Passengers']  # 生成 pd.Series 对象
f = plt.figure(facecolor='white')
ts.plot(color='blue')
plt.show()
```



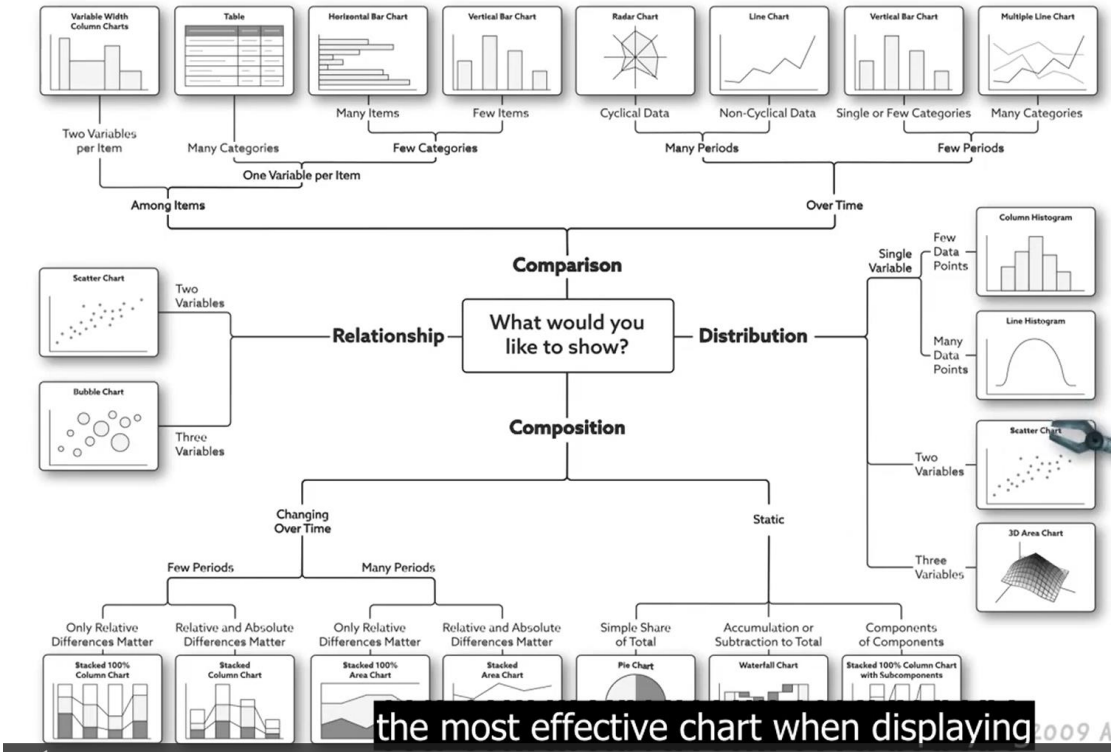
Explain:

We choose the line chart. Because we want to show the relationships between two variables , timeseries and quantity. The line chart can show the change in quantity over time.

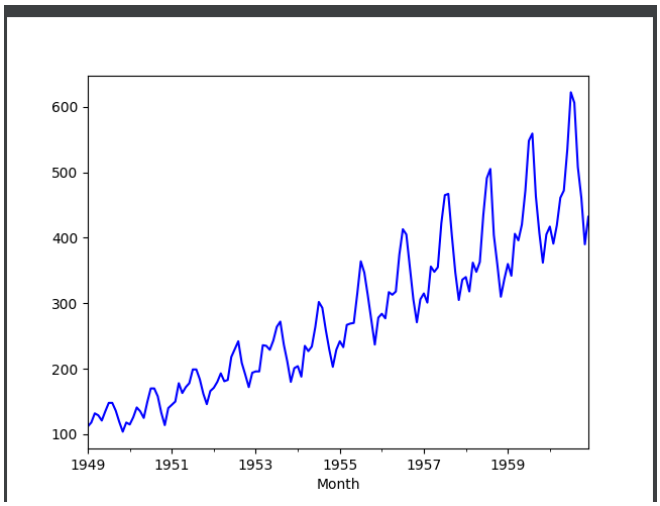
Part 3:

Watch the video“How to choose the right chart for your data visualization” on Moodle and in this video, how to choose the right chart for your data

visualization is introduced. Please first summarize what you learn from this video and then re-draw the graph that created by Andrew Avila by using a propreate software and make sure all contents are in Chinese. Note: export your graph into an image format (e.g. jpeg/png/bmp etc.) and attach it in the end of the Report 7.

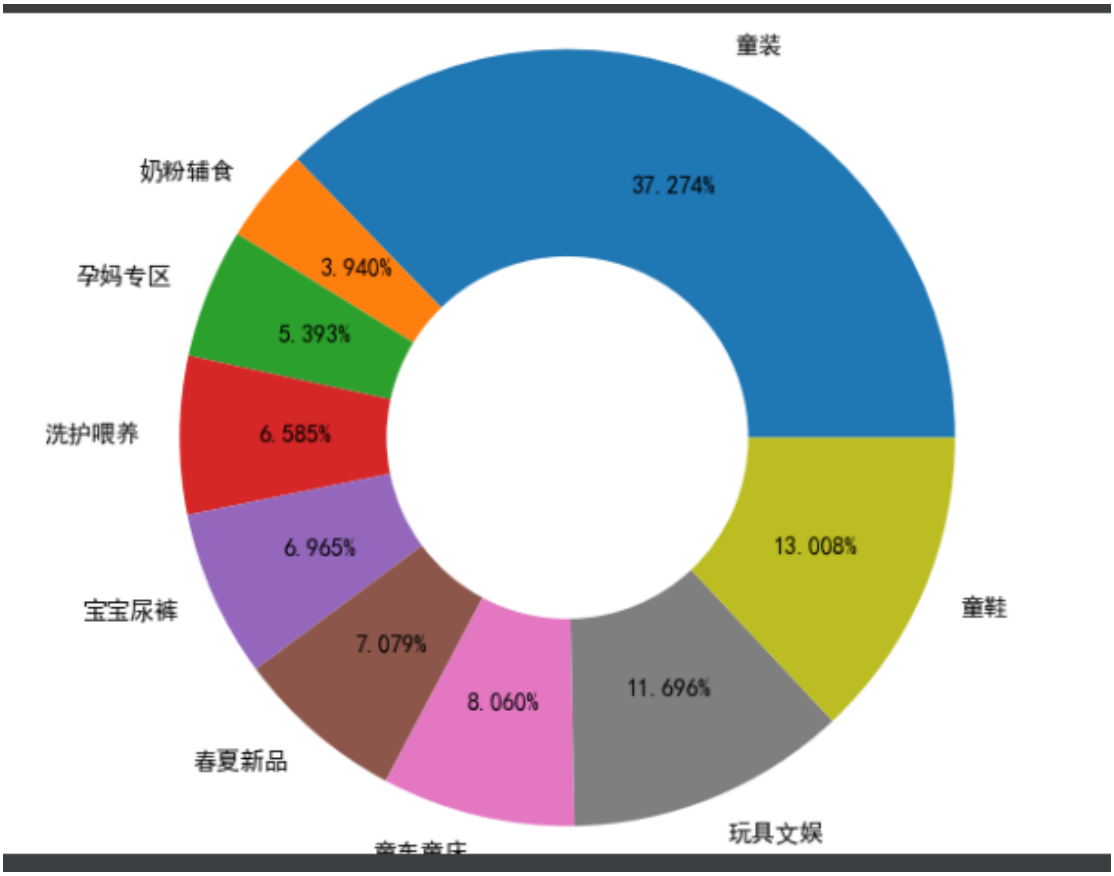


The point is what would you like to show with your data distribution relationships. We can divide four categories according to this standard , including relationship、composition、distribution、comparision. Then we can choose different kinds chart according to the numbers of variables.



Part 4: research task

Find a data visualization chart from any source (internet, your Alipay report, etc.) and try to use a different method to represent the same data, summarize your results in the report.




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


df = pd.DataFrame({'子类目':['童装','奶粉辅食','孕妈专区','洗护喂养','宝宝
尿裤','春夏新品','童车童床','玩具文娱','童鞋'],
                   '销售额
':[29665,3135.4,4292.4,5240.9,5543.4,5633.8,6414.5,9308.1,10353]})

name = np.array(df['子类目'])
data = np.array(df['销售额'])
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.pie(data,radius=1.5,labels=name,
        autopct='%0.3lf%%',pctdistance=0.5)
plt.show()
```

计算机 22A4 2022015232 孙春辉
2025 年 4 月 2 日 星期三 第八次作业