

Python程序设计实验报告

数据可视化实验

| | |
|------|------------|
| 姓 名： | 王仕和 |
| 学 号： | 2019213681 |
| 日 期： | 2022.01.07 |

- 1 作业1：依据图表信息绘制历次人口普查全国人口数量柱状图
- 2 作业2：依据图信息绘制某人2018年支付宝年支出情况饼图
- 3 作业3：使用IRIS数据集，在一个figure中绘制出右侧的16个子图
- 4 作业4：绘制八年级期末考试成绩直方图
- 5 作业5：绘制北京PM2.5指数月平均折线图

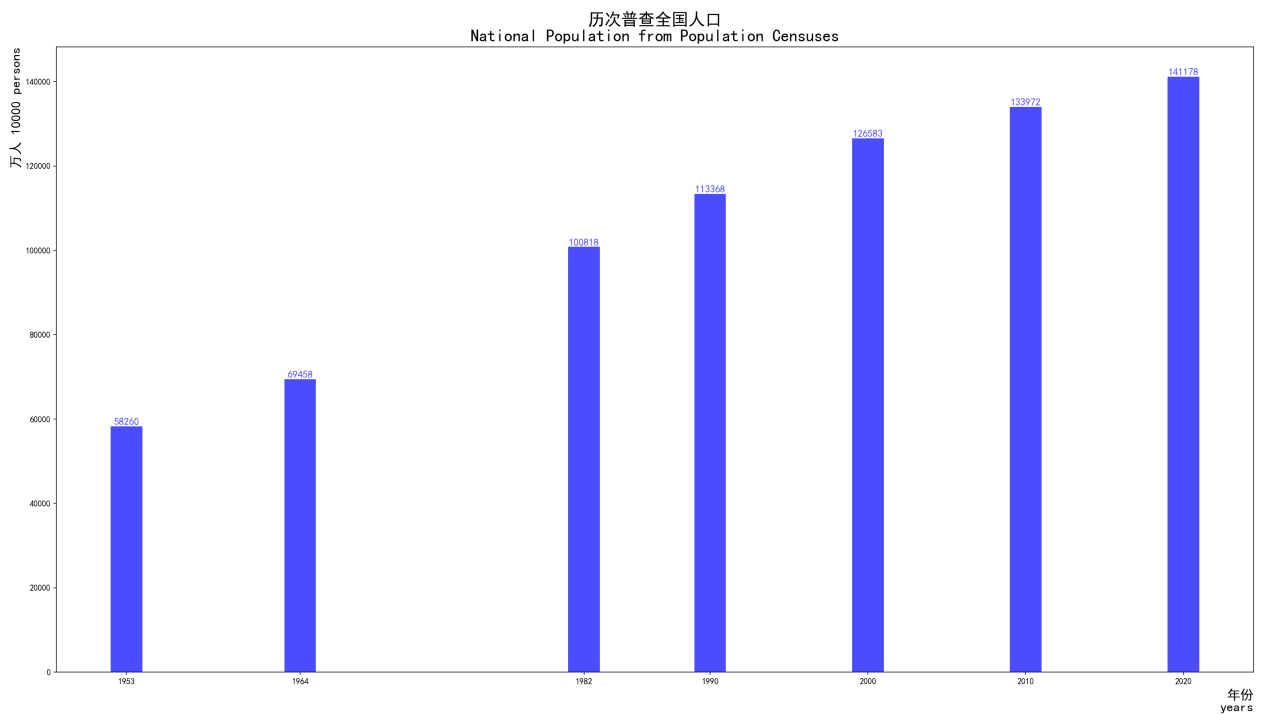
Python程序设计实验报告

数据可视化实验

1 作业1：依据图表信息绘制历次人口普查全国人口数量柱状图

依据图表信息绘制历次人口普查全国人口数量柱状图。

```
1 import matplotlib.pyplot as plt
2
3 year = [1953, 1964, 1982, 1990, 2000, 2010, 2020]
4 height = [58260, 69458, 100818, 113368, 126583, 133972, 141178]
5
6 plt.rcParams['font.sans-serif'] = ['SimHei']
7 plt.title('历次普查全国人口\nNational Population from Population Censuses',
8           fontsize=20)
9 plt.xlabel('年份\nyears', loc='right', fontsize=16)
10 plt.ylabel('万人 10000 persons', loc='top', fontsize=16)
11 plt.bar(year, height, width=2.0, color='b', alpha=0.7, align='center')
12 plt.xticks(ticks=year)
13 for i, j in zip(year, height):
14     plt.text(i, j, format(j, ''), ha='center', va='bottom', fontsize=12,
15             color='b', alpha=0.7)
16 plt.show()
```

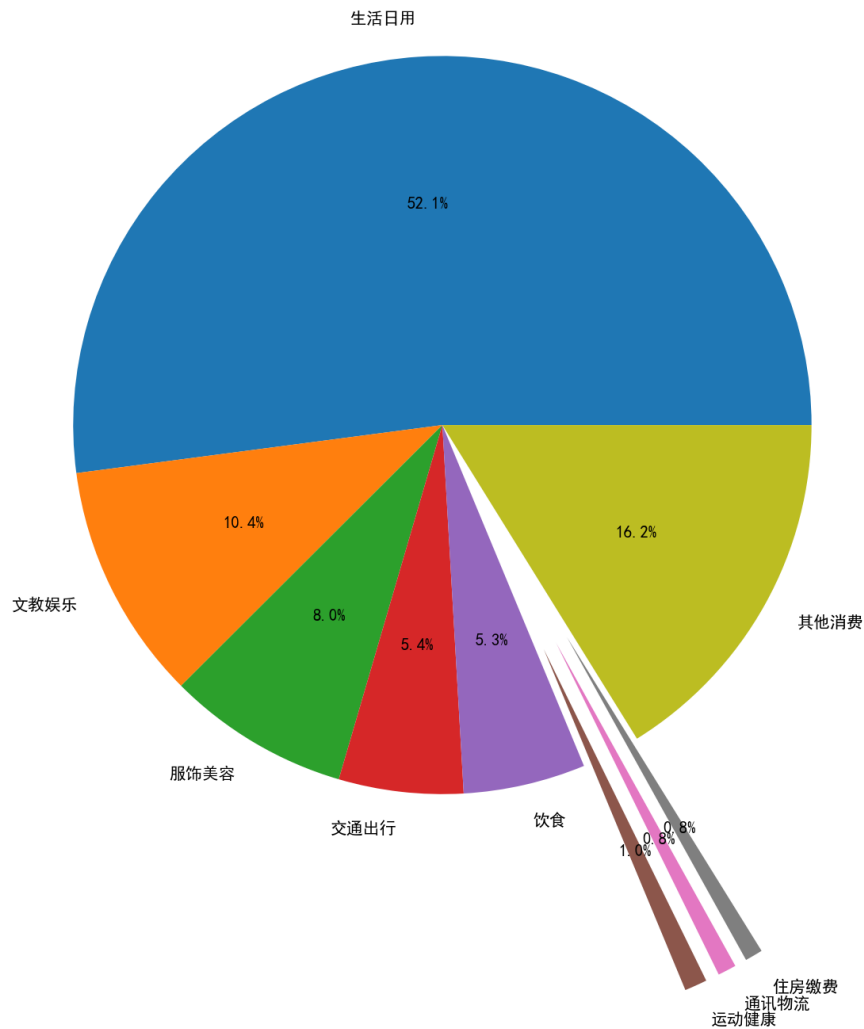


2 作业2：依据图信息绘制某人2018年支付宝年支出情况饼图

依据图信息绘制某人2018年支付宝年支出情况饼图，图中应展示各类型支出占总支出的比例。

```
1 import matplotlib.pyplot as plt
2
3 label = ['生活日用', '文教娱乐', '服饰美容', '交通出行', '饮食', '运动健康', '通讯物
   流', '住房缴费', '其他消费']
4 money = [51686.39, 10325.60, 7927.91, 5407.22, 5304.16, 980.11, 801.55,
   770.00, 16041.03]
5
6 plt.rcParams['font.sans-serif'] = ['SimHei']
7 plt.title('2018年支付宝年支出情况', fontsize=20)
8 explode = [0] * len(money)
9 for i in range(len(money)):
10     if money[i] < 1000:
11         explode[i] = 0.5
12     elif money[i] < 5000:
13         explode[i] = 0.1
14     else:
15         explode[i] = 0
16 plt.pie(money, labels=label, autopct='%.1f%%', explode=explode, radius=0.75)
17 plt.show()
```

2018年支付宝年支出情况



3 作业3：使用IRIS数据集，在一个figure中绘制出右侧的16个子图

使用IRIS数据集，在一个figure中绘制出右侧的16个子图。

分别使用花瓣长度、花瓣宽度、花萼长度和花萼宽度这四种数据，两两组合，绘制散点图。

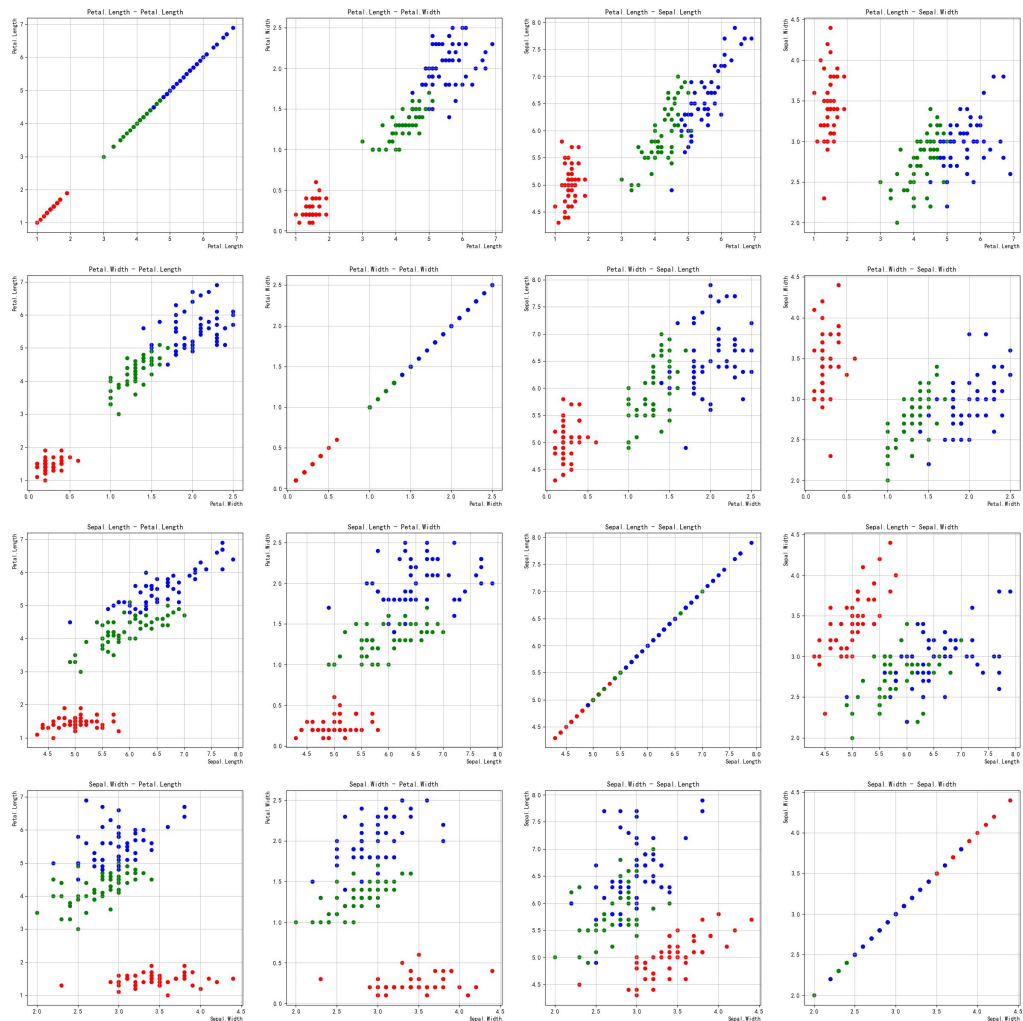
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 plt.rcParams['font.sans-serif'] = ['SimHei']
5 df = pd.read_csv("iris.csv")
6 attribute = ['Petal.Length', 'Petal.Width', 'Sepal.Length', 'Sepal.Width']
7 species = df.Species.unique()
8
9 color = ['r', 'g', 'b']
10 plt.figure(figsize=(30, 30))
11 plt.legend()
```

```

12 plt.suptitle('IRIS散点图')
13 for xid, x in enumerate(attribute):
14     for yid, y in enumerate(attribute):
15         plt.subplot(4, 4, 4 * xid + yid + 1)
16         for i, specie in enumerate(species):
17             plt.scatter(x=df.loc[df.Species == specie, x],
18 y=df.loc[df.Species == specie, y], s=35, c=color[i], label=specie)
19             plt.title(f'{x} - {y}')
20             plt.xlabel(x, loc='right', fontsize=10)
21             plt.ylabel(y, loc='top', fontsize=10)
22             plt.grid(True, alpha=0.7)
23 plt.savefig('3.jpg')

```

IRIS散点图

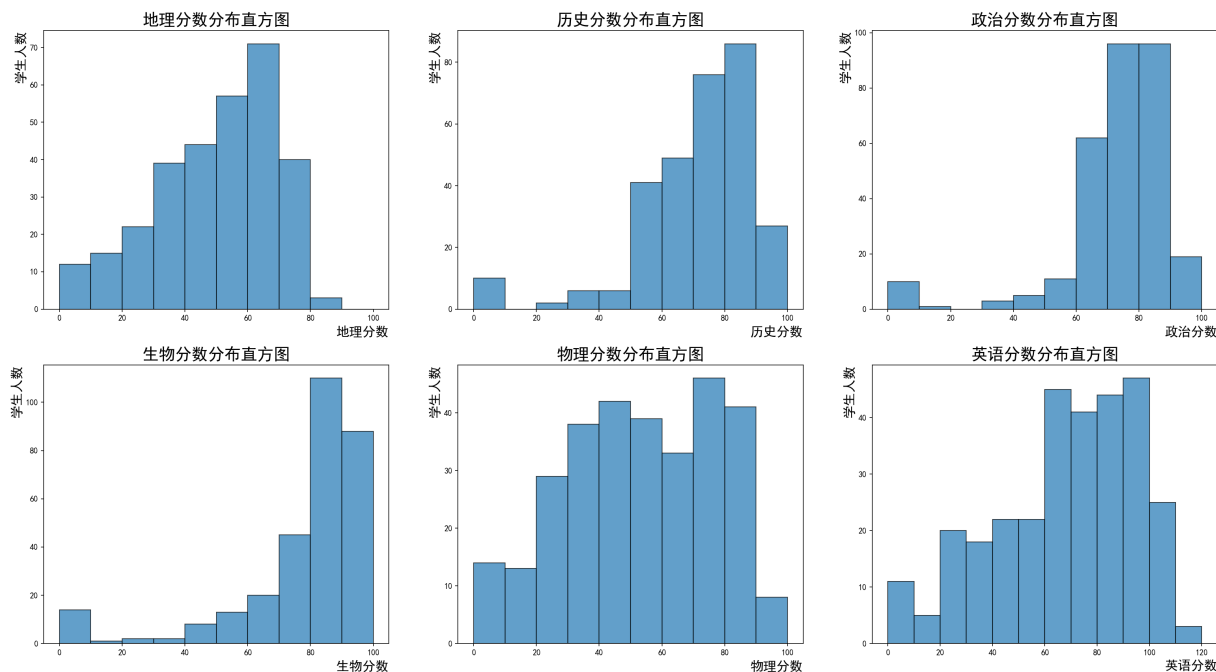


4 作业4：绘制八年级期末考试成绩直方图

使用给出的“八年级期末考试成绩表.xlsx”，在一个figure中绘制六个子图，分别绘制六门课程的成绩分段统计情况直方图，每10分一个分段。

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 plt.rcParams['font.sans-serif'] = ['SimHei']
5 df = pd.read_excel('八年级期末考试成绩表.xlsx')
6 courses = ['地理分数', '历史分数', '政治分数', '生物分数', '物理分数', '英语分数']
7
8 plt.suptitle('八年级期末考试成绩分布直方图', fontsize=24)
9 for i, course in enumerate(courses):
10     plt.subplot(2, 3, i + 1)
11     plt.title(f'{course}分布直方图', fontsize=20)
12     plt.xlabel(course, loc='right', fontsize=16)
13     plt.ylabel('学生人数', loc='top', fontsize=16)
14     score = df[course]
15     limit = 120 if course == '英语分数' else 100
16     bin = [i for i in range(0, limit + 10, 10)]
17     plt.hist(score, bins=bin, alpha=0.7, edgecolor='black')
18
19 plt.show()
```

八年级期末考试成绩分布直方图



5 作业5：绘制北京PM2.5指数月平均折线图

使用“BeijingPM20100101_20151231.csv”数据集，展示北京市2010~2015年PM2.5指数月平均数据的变化情况，在同一幅图中绘制六条折线，每年一条折线。

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4
5 plt.rcParams['font.sans-serif'] = ['SimHei']
6 df = pd.read_csv("BeijingPM20100101_20151231.csv")
7
8 columns = ['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US Post']
9 # 计算每一行四个pm2.5监测点的平均值
10 meanpm = round(df[columns].mean(axis=1), 1)
11 # 创建一个四个监测点名称到平均值的映射
12 fill = {}.fromkeys(columns, meanpm)
13 # 使用映射，将四个监测点的平均值替换某些监测点的空值
14 df.fillna(value=fill, inplace=True)
15 df.to_csv('./after.csv', index=False, encoding='utf-8')
16
17 avg = df.groupby(['year', 'month']).agg({'PM_Dongsi': np.mean})
18
19 plt.legend()
20 plt.title('北京2010-2015各年份PM2.5指数月均值折线图', fontsize=20)
21 plt.xlabel('月', loc='right', fontsize=16)
22 plt.ylabel('PM2.5指数月均值', loc='top', fontsize=16)
23 for i, year in enumerate(range(2010, 2016)):
24     plt.plot(list(range(1,13)), avg.loc[year]['PM_Dongsi'].to_list(),
25             label=year)
26
27 plt.legend()
28 plt.show()
```

北京2010-2015各年份PM2.5指数月均值折线图

