

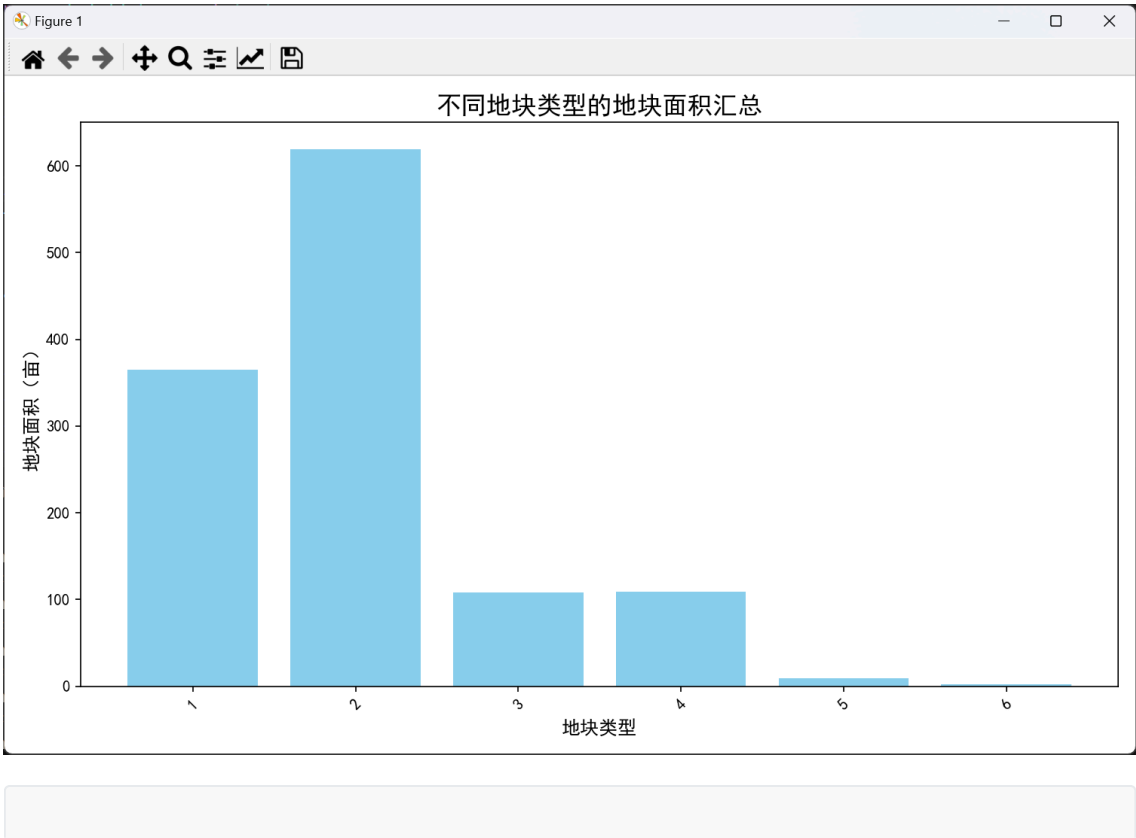
2024.9.30作业

复习题

- T1
数据采集，数据存储，数据管理，数据计算，数据分析，数据展示
- T2
数据采集是指从真实世界对象中获得原始数据的过程。获取途径包括传感器，日志文件，web爬虫
- T3
数据管理是利用计算机硬件和软件技术对数据进行有效的收集#存储#处理和应用的过程。传统数据管理和大数据管理技术都会处理存储分析数据，都会定期备份数据。但是传统数据管理存储量更低，存储效率更低。大数据技术会采用分布式并行化的方式来处理数据，而传统技术只会使用数据仓库。
- T4
数据批量计算模式，数据流计算模式，数据交互式计算模式，数据图计算模式
- T5
数据分析处理来自对某一兴趣现象的观察#测量或者实验的信息。分析方法有：描述性分析，预测性分析，规则性分析
- T6
数据可视化可以吸引使用者，可以使大家获取更多的信息量，还可以在小空间中展示大规模数据，让大家对数据有更加全面的认识。

练习题

- T7



```

import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import font_manager

# 1. 设置中文字体
plt.rcParams['font.sans-serif'] = ['SimHei'] # 设置中文字体为 SimHei（黑体）
plt.rcParams['axes.unicode_minus'] = False

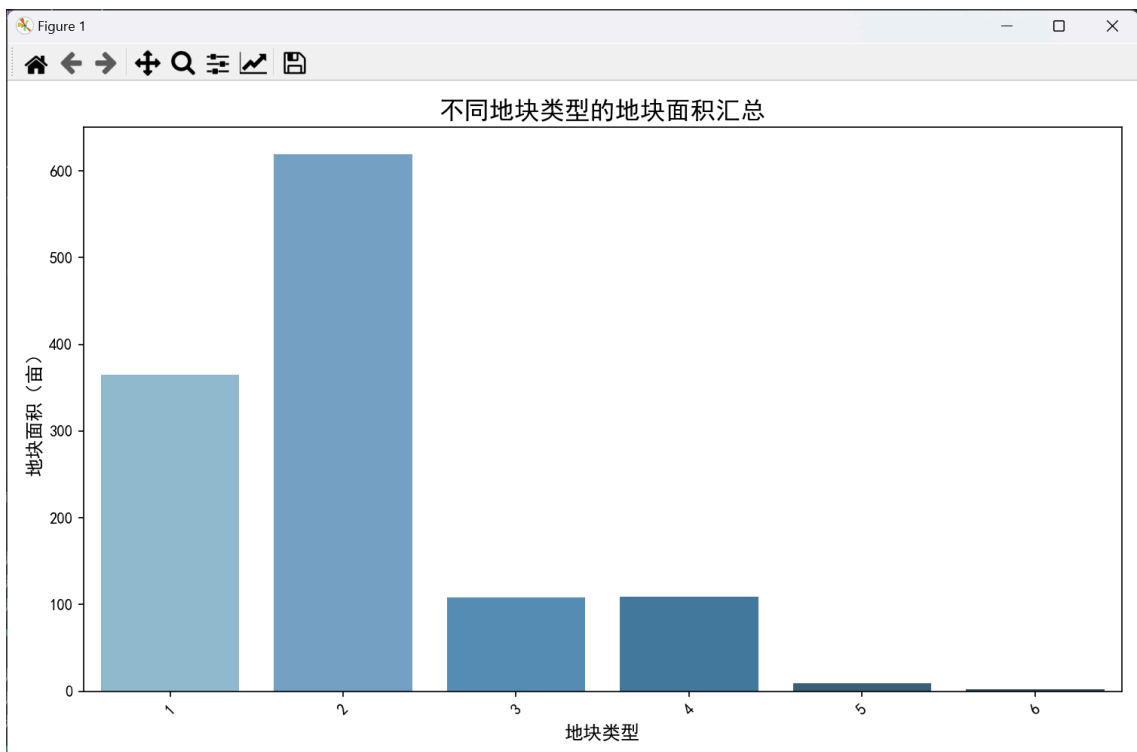
# 2. 读取 Excel 文件中的特定工作表
file_path = '附件1.xlsx'
df = pd.read_excel(file_path, sheet_name='乡村的现有耕地')

# 3. 按地块类型汇总地块面积
land_summary = df.groupby('地块类型')['地块面积/亩'].sum().reset_index()

# 4. 绘制柱状图
plt.figure(figsize=(10, 6))
plt.bar(land_summary['地块类型'], land_summary['地块面积/亩'], color='skyblue')
plt.title('不同地块类型的地块面积汇总', fontsize=16)
plt.xlabel('地块类型', fontsize=12)
plt.ylabel('地块面积（亩）', fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```

- T8



```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

```

```
file_path = '附件1.xlsx'
df = pd.read_excel(file_path, sheet_name='乡村的现有耕地')
land_summary = df.groupby('地块类型')['地块面积/亩'].sum().reset_index()

plt.figure(figsize=(10, 6))
sns.barplot(x='地块类型', y='地块面积/亩', data=land_summary, palette='Blues_d')

plt.title('不同地块类型的地块面积汇总', fontsize=16)
plt.xlabel('地块类型', fontsize=12)
plt.ylabel('地块面积（亩）', fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```