

## 复习题

### 1.数据的全生命周期管理包括哪些阶段？

数据的全生命周期管理包括数据采集、数据存储、数据管理、数据计算、数据分析和数据展示等阶段。

### 2.数据采集的概念是什么？都有哪些方法？

数据采集是指从真实世界对象中获得原始数据的过程。不准确的数据采集将影响后续的数据处理并最终得到无效的结果。数据采集方法的选择不但要依赖于数据源的物理性质，还要考虑数据分析的目标。

课堂上讲的三种数据采集方法是：传感器、日志文件（例如视频网站记录你的浏览记录和时间等等）和Web爬虫。

### 3.什么是数据管理？比较传统的数据管理和大数据管理技术有什么异同？

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。其目的在于充分有效地发挥数据的作用。大数据管理技术需要处理的数据类型更多、体量更大，且对实时性有更高的要求。还有就是大数据更注重数据本身，而没有那么在意服务的端

### 4.大数据的计算模式可以分为哪几类？

(书上原话)

数据的计算模式大致分为批量计算模式#流式计算模式#交互式计算模式和图计算模式四类

### 5.什么是数据分析？有哪些数据分析的方法或者模型？

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

数据分析的方法可以分为：描述性分析、预测性分析和规则性分析

数据分析技术主要有：统计分析、数据挖掘、机器学习和可视化分析

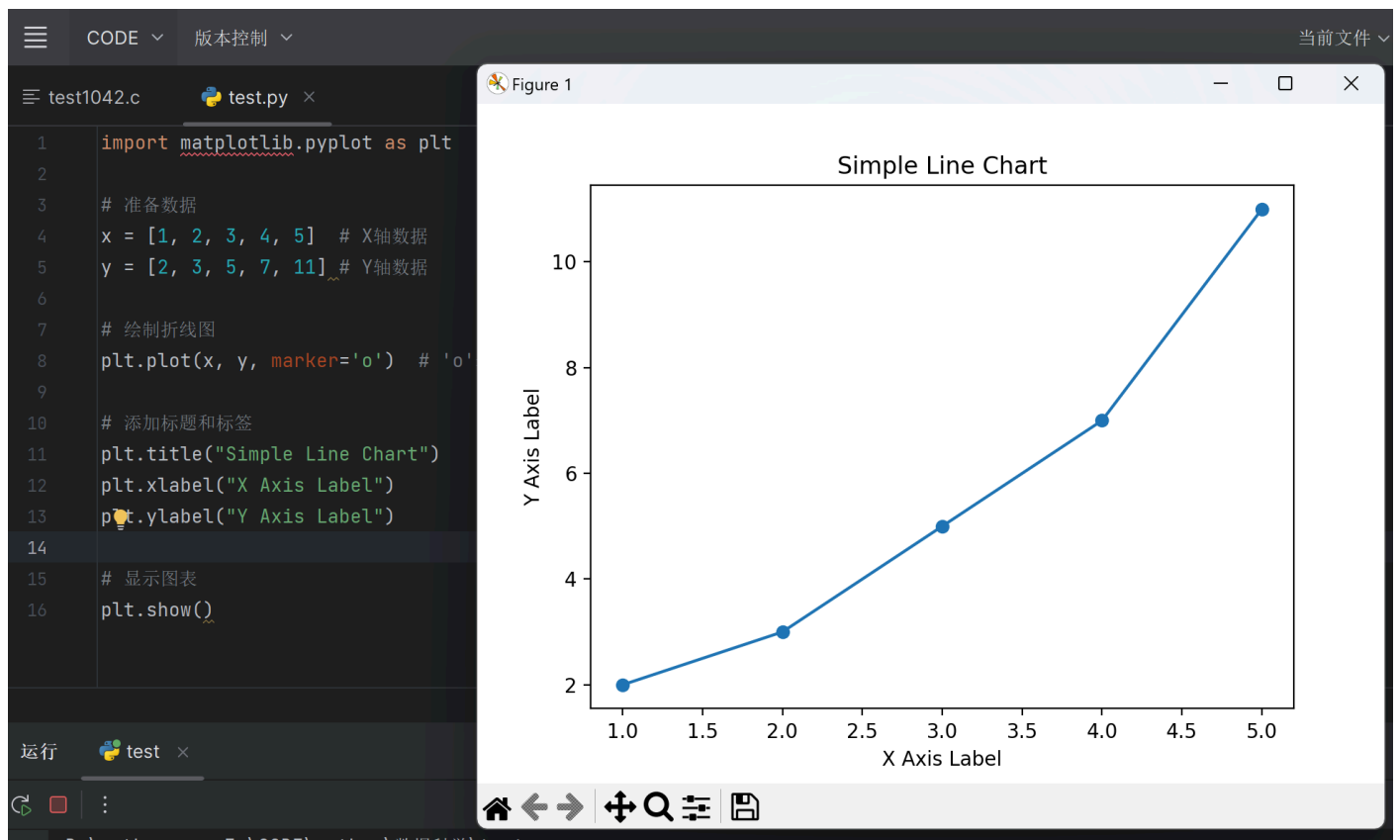
### 6.数据可视化的原因有哪些？

(书上原话。。无聊咩。。)

数据可视化的原因包括：我们利用视觉获取的信息量远远比其他感官要多，数据可视化能够在小空间中展示大规模数据，数据可视化能够帮助我们对数据有更加全面的认识，受人类大脑记忆能力的限制

## 练习题

### 1.matplotlib



### 2.Seaborne

```

import pandas as pd # 导入Pandas库, 并使用pd作为别名
import seaborn as sns
import matplotlib.pyplot as plt

# 准备数据
data = {
    "Date": ["2021-01-01", "2021-01-02", "2021-01-03", "2021-01-04", "2021-01-05"],
    "Value": [10, 15, 7, 10, 9]
}

df = pd.DataFrame(data) # 使用pd别名创建DataFrame

# 确保日期列是日期类型
df['Date'] = pd.to_datetime(df['Date'])

# 绘制折线图
sns.lineplot(x='Date', y='Value', data=df)

# 添加标题

test
:
python.exe E:\CODE\python\数据科学\test.py

```

