实验报告

东北林业大学

信息与计算机科学技术实验中心

|  |
| --- |
| 一、实验目的  （1）理解Matplotlib库中画板、画布、子图及图形元素的基本概念，掌握利用Matplotlib库中相关方法实现散点图、折线图、直方图、柱状图、饼图、热力图等图形与图像的可视化操作，并能够根据数据的内容选择适当的方法进行数据的可视化展示；  （2）掌握利用pandas库中的hist（）、plot（）、boxplot( )等方法实现数据的可视化；  （3）掌握利用jieba库进行中文分词的方法；  （4）掌握利用wordcloud库实现词云图的绘制。 |
| 二、实验环境  （1）硬件：PC机；  （2）软件：Anaconda Jupyter Notebook，Spyder，Pandas |
| 三、实验内容及结果  （一）基于Maplotlib库的数据可视化  1、读取土地利用数据集统计结果：各县市土地利用面积统计表.xls，并显示数据。  2、去除最后一列。    3、利用柱状图显示1980年各类数据的面积，并将柱状图中柱体的颜色设置为你喜欢的颜色。  4、利用并列柱状图对1980年和2010年的各类用地面积进行对比分析。  5、利用饼状图显示2010年各类用地的比例，要求标注面积比例和各类用地名称，面积比例保留两位小数。    6、利用环状图显示1980年各类用地的比例，要求标注面积比例和各类用地名称，面积比例保留两位小数。  7、打开学生成绩数据集，对各列进行重新排列，保留平时成绩、实验成绩、阶段成绩、期末成绩。总成绩中各部分的比例分别为10%,10%,30%,50%    8、利用堆叠柱状图显示各同学的平时成绩、实验成绩、阶段成绩和期末成绩。  9、利用散点图查看学生阶段成绩和期末成绩，简单分析一下其相关性。    10、利用PIL的Image对象打开数据集中的im1.tif和im2.tif，文件，并对其进行对比显示，查看区域前后的变化情况。      11、打开数据集中桂林市1951-2013年月降水量.xls，利用折线图显示其中的数据，分析区域的降水量特征。    （二）基于Pandas库的可视化  1、打开学生成绩数据集，检查其中数值型数据的分布。    2、利用直方图显示学生各类成绩的分布情况。    3、利用箱线图查看学生成绩中的异常点。    （三）词云图的制作  1、导入相关库    2、选择你感兴趣的资料，存入txt文件，然后利用open（）打开文件，读取数据并剔除标点符号。  3、首先利用jieba进行中文分词，如果需要针对特有名词进行处理，则制作用户词典，然后再分词。    4、选择背景图片，打开背景图片进行显示。    3、利用wordcloud制作词云图。 |

|  |
| --- |
| 四、实验过程分析与讨论  1． 实验过程分析与讨论：  （1）项目1：使用Pandas的read\_excel函数读取Excel文件，并将第一列作为索引列，将数据存储在名为df的DataFrame中。df.drop("总计", inplace=True, axis)：删除DataFrame中的名为"总计"的行。labels = df.columns：将DataFrame的列名赋值给labels变量。x = np.arange(len(labels))：创建一个等差数组，用于设置x轴刻度标签的位置。plt.xticks(x, labels=labels)：设置x轴刻度标签的位置和内容。plt.legend()显示图例。  （2）项目2：df = pd.read\_csv(file, encoding="GBK", index\_col=0)：使用Pandas的read\_csv函数读取CSV文件，并将第一列作为索引列，将数据存储在名为df的DataFrame中。df.copy()：复制df的数据。df2 = df2 \* np.array([0.1, 0.1, 0.3, 0.5])：将df2中的每列数据与给定的权重数组相乘，得到加权后的数据。ax1 = plt.subplot(1, 1, 1)：创建一个子图对象。box = ax1.get\_position()：获取子图对象的位置和尺寸。  柱状图展示了不同成绩类型在不同序号下的加权成绩情况，而散点图则展示了阶段成绩和期末成绩之间的关系。通过这些可视化图表，可以更直观地分析和讨论成绩数据的分布和趋势。  （3）项目3：是用Python中的PIL库来处理图像，并使用matplotlib库来进行图像的显示和数据的可视化。  首先通过PIL库读取了两个图像文件，并使用matplotlib库将它们显示在一个图像窗口中。然后通过pandas库读取了一个Excel文件中的数据，并使用matplotlib库将数据绘制成折线图进行可视化。通过这段代码，可以直观地了解图像的内容，并对数据进行分析和比较。  （4）SyntaxError: positional argument follows keyword argument  出现这个bug的原因在于参数位置不正确，关键字参数必须跟随在位置参数后面! 因为python函数在解析参数时, 是按照顺序来的, 位置参数是必须先满足, 才能考虑其他可变参数. |
| 五、指导教师意见    指导教师签字：  年 月 日 |