**实验七 数据可视化实验**

**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、实验仪器及实验环境**

（1）硬件：PC机；

（2）软件：Anaconda Jupyter Notebook，Spyder，Pandas

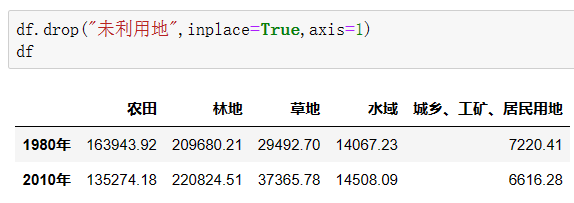
**三、实验内容：**

（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。

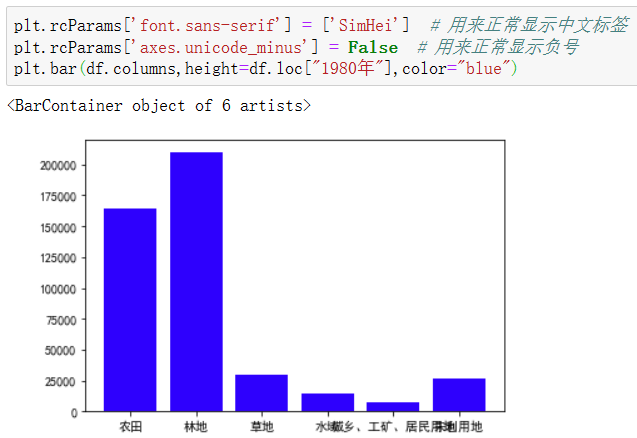


3、利用柱状图显示1980年各类数据的面积，并将柱状图中柱体的颜色设置为你喜欢的颜色。

提示：为了正确的显示中文和负号，需要在脚本文件中导入相关库后加入如下两行代码：

plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签

plt.rcParams['axes.unicode\_minus'] = False # 用来正常显示负号

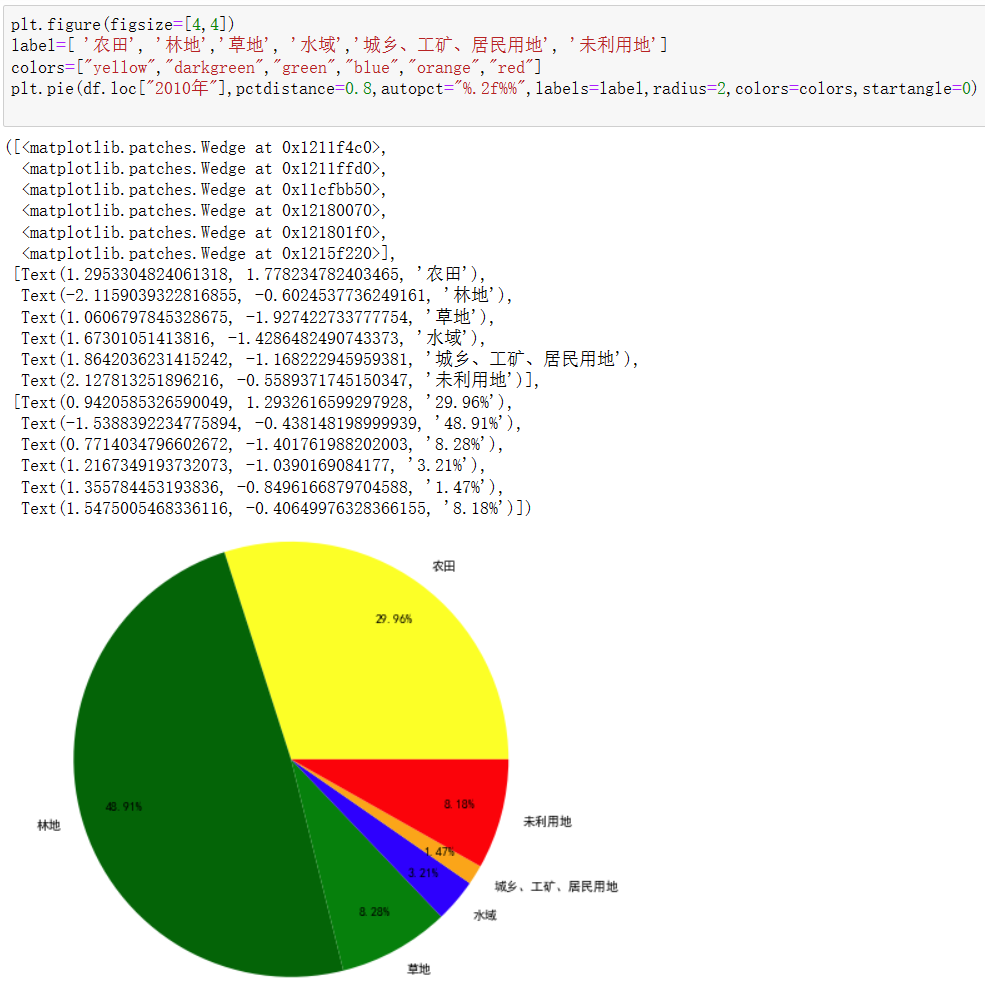


4、利用并列柱状图对1980年和2010年的各类用地面积进行对比分析。

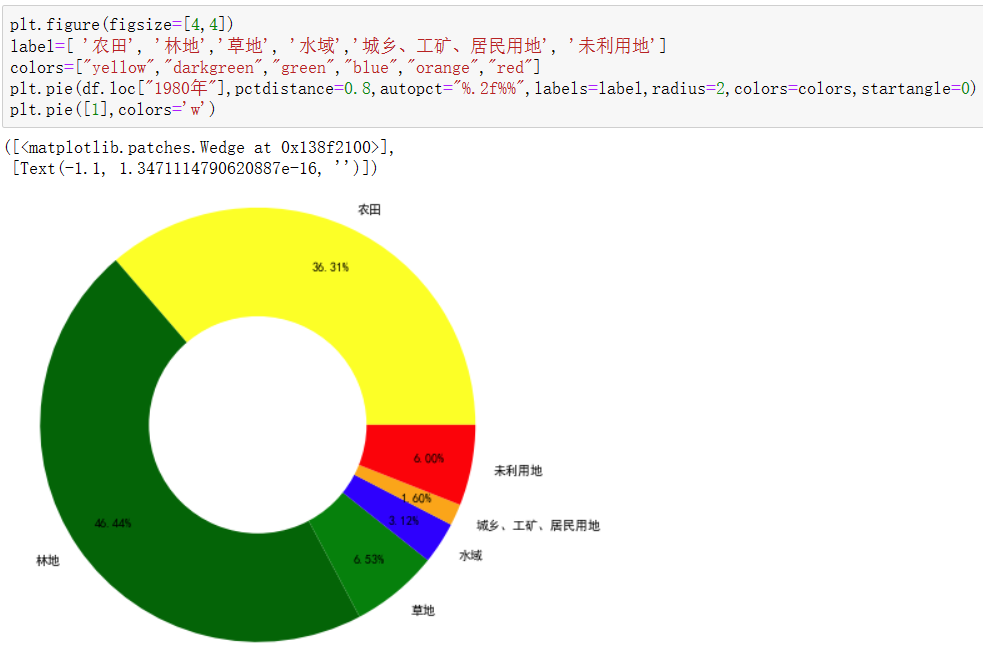
提示：并列柱状图中要计算各个主体的位置，利用宽度进行控制，默认柱体宽度为0.8，显示两个属性时，可以将宽度设置为0.4，两个样本之间有一定分割。1980年数据的6个柱体的位置用1-6来进行表示，那么2010年数据的柱体位置就应该是1.4,2.4…，也就是在1980年柱体的右侧。X轴的刻度要显示为不同利用类型，利用xticks进行设置。



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。

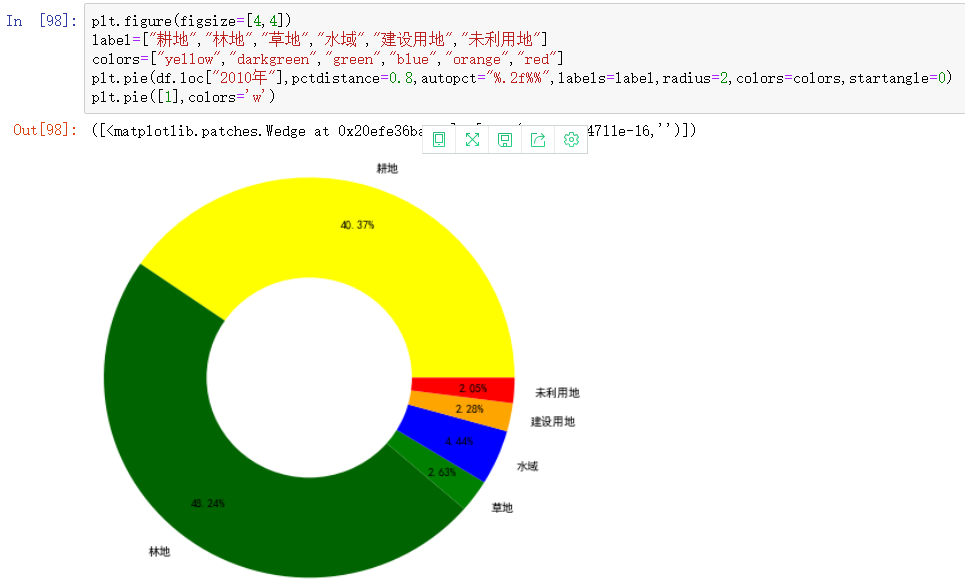
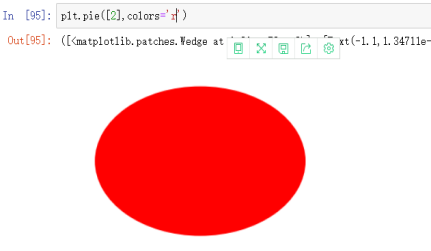


6、利用环状图显示1980年各类用地的比例，要求标注面积比例和各类用地名称，面积比例保留两位小数。

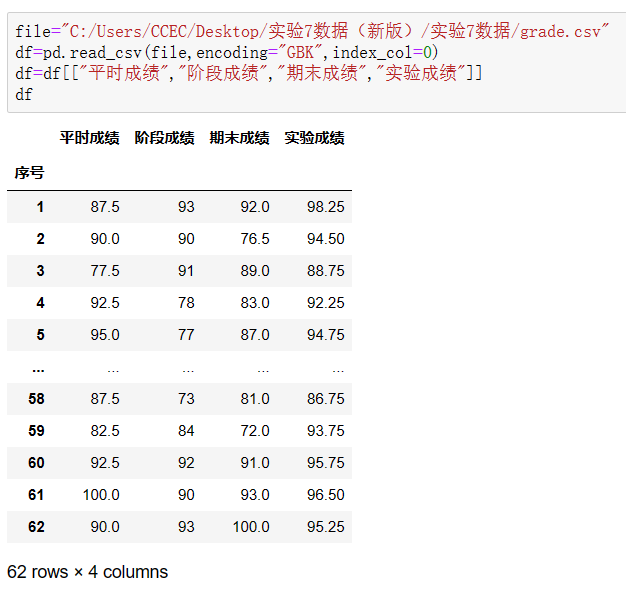


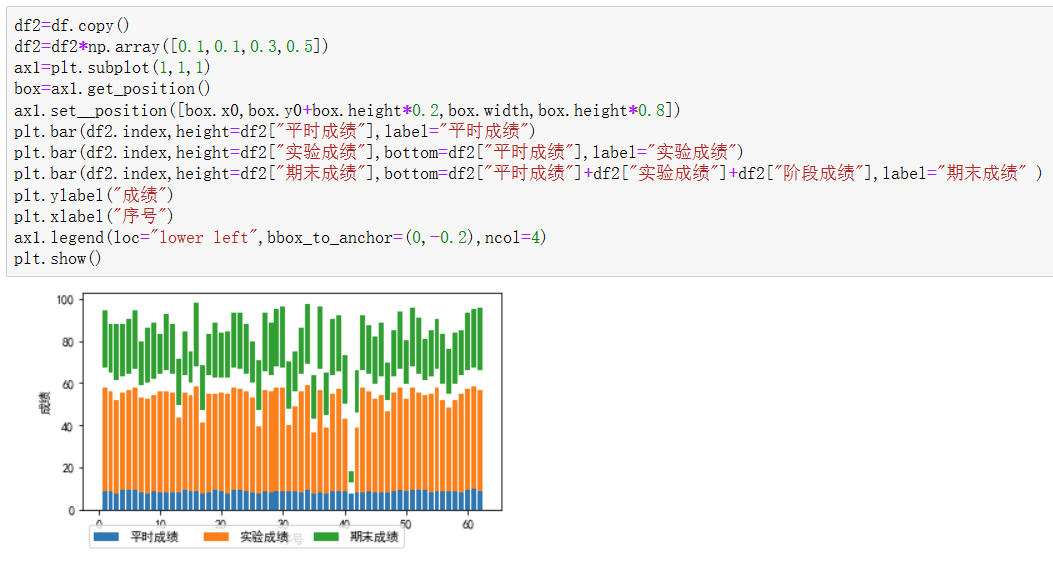
提示：在饼状图的中心绘制一个小的白色饼填充中间部分，即可形成一个环状图。此时，只用一个>0的数据进行饼状图的可视化即可。如：

plt.pie([1],colors=‘r’)就可以产生一个红色的饼图了。将其的半径调小一点，颜色设置为白色，与题4中的饼图叠加就可以生成环状图了。

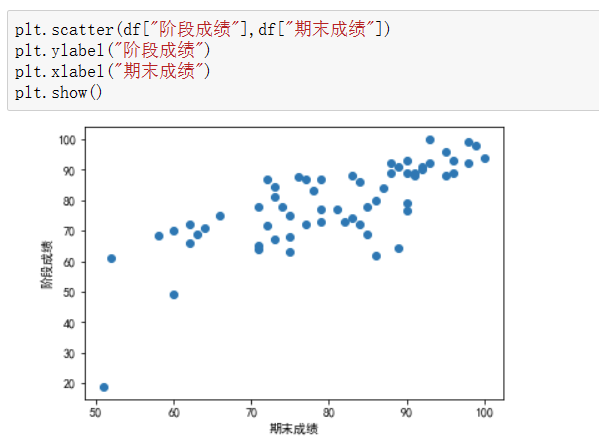


7、打开学生成绩数据集，对各列进行重新排列，保留平时成绩、实验成绩、阶段成绩、期末成绩。总成绩中各部分的比例分别为10%,10%,30%,50%



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。

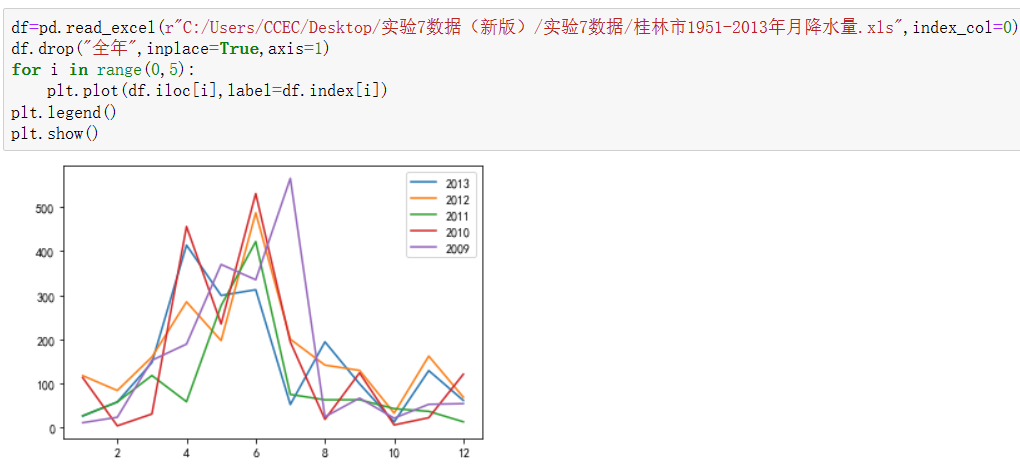
9、利用散点图查看学生阶段成绩和期末成绩，简单分析一下其相关性。



1. 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。

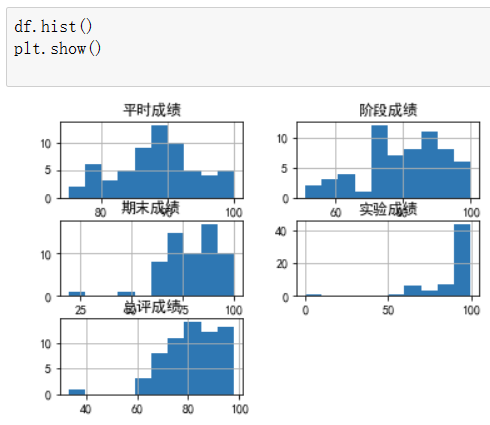


（二）基于Pandas库的可视化

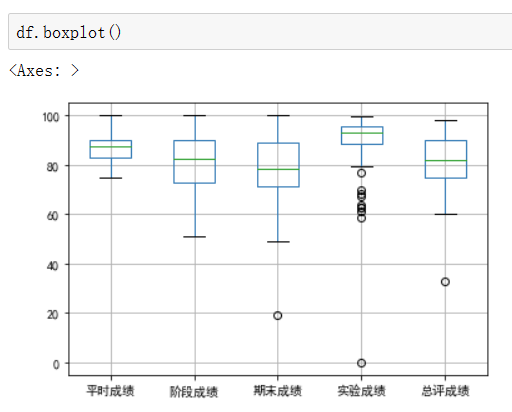
1、打开学生成绩数据集，检查其中数值型数据的分布。



2、利用直方图显示学生各类成绩的分布情况。

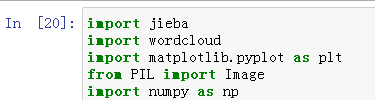


3、利用箱线图查看学生成绩中的异常点。

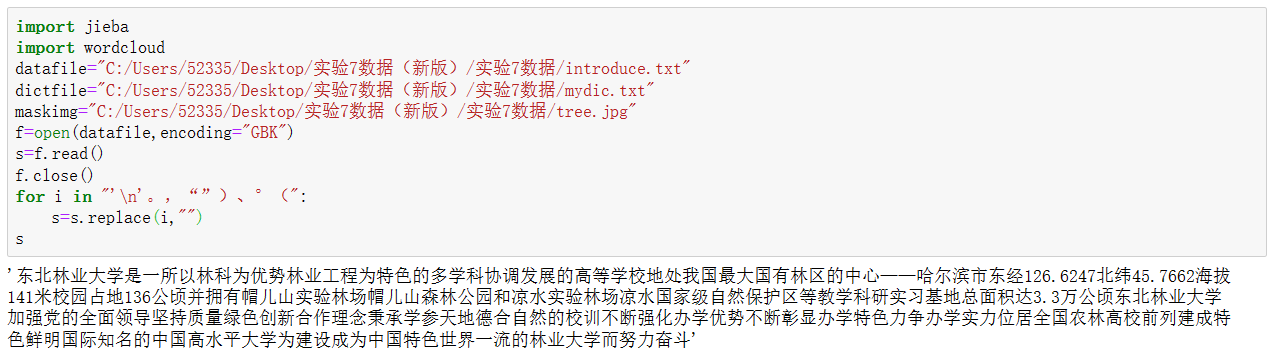


（三）词云图的制作

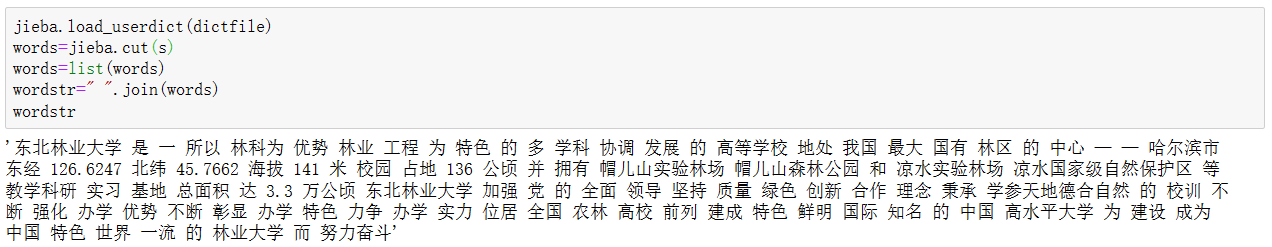
1、导入相关库



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、实验结果与分析：**

（1）验证实现的程序功能与题目要求是否一致。

（2）查找程序出错的语句代码，调试分析出错的原因并改进语句。