

山西大学计算机与信息技术学院

实 验 报 告

姓 名	付 天 豪	学 号	201702904036	专业班级	2017 级大数据 2 班
课程名称	数据挖掘与分析				
指导教师	段菲	实验日期	2019. 10. 11	成 绩	
实 验 名 称	实验一 鸢尾花数据集的处理				
<p>一、 实验目的：</p> <p>学会利用 python 进行简单的数据分析</p> <p>二、实验内容：</p> <p>1、实验原理：</p> <p>Jupyter notebook 是一个进行数据分析的强大利器</p> <p>Python 包 numpy 提供了丰富的函数 例如进行求解百分位点等等</p> <p>Python 包 pandas 提供了丰富的函数 例如进行绘制箱图、直方图、散点图、散点图矩阵等等</p> <p>2、实验步骤：</p> <p>（一） 求解百分位点</p> <p>查阅相关资料，发现 python np.percentile 可进行求解数据集的百分位点，因此选用此函数进行求解。</p> <p>同时看了此函数的源码，对求解百分位点有了新的体会。</p> <p>核心代码：</p> <pre>x1 = np.percentile(iris_1,(25,50,75)) x2 = np.percentile(iris_2,(25,50,75)) x3 = np.percentile(iris_3,(25,50,75)) x4 = np.percentile(iris_4,(25,50,75)) print(x1) print(x2) print(x3) print(x4)</pre> <p>（二） 绘图</p> <p>查阅相关资料，发现 python 包 pandas 自带一些绘图函数极为方便，因此采用 pandas 包中带的函数 .boxplot() ， .hist() ， .plot() ， .plotting.scatter_matrix()，完成本次实验内容。</p> <p>同时看了以上四个函数实现的源码，对绘图有了一些新的体会。</p> <p>核心代码：</p> <pre>df.boxplot() plt.show() df1.hist() df.plot() pd.plotting.scatter_matrix(df)</pre>					

3、实验结果：

(1) 鸢尾花各属性百分位点：

示例：属性名：[25%分位点 50%分位点 75%分位点]

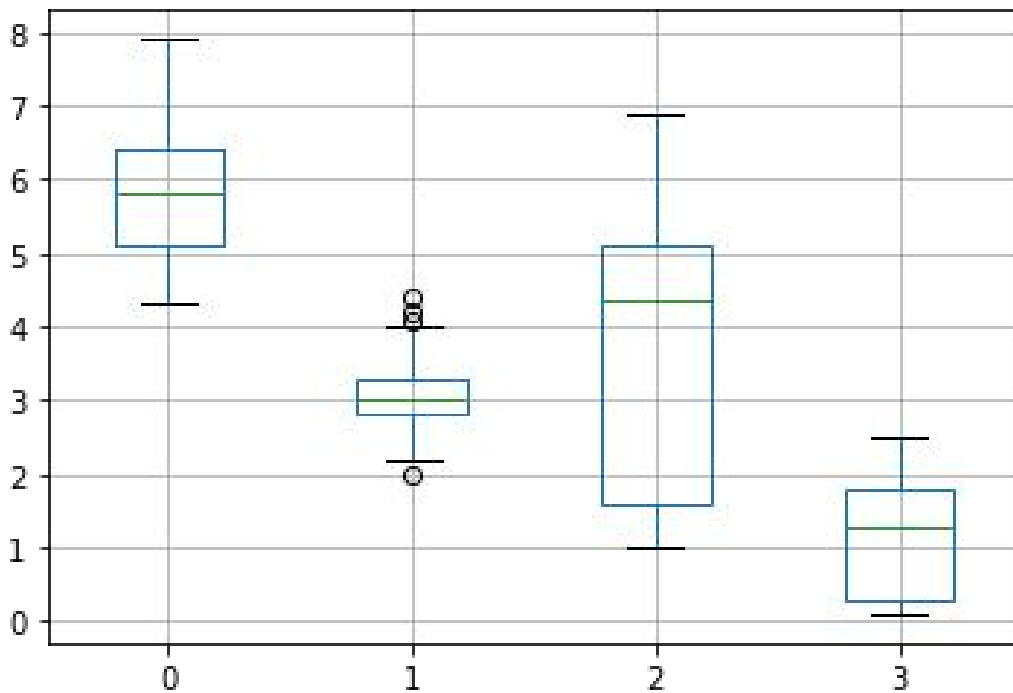
萼片长度：[5.1 5.8 6.4]

萼片宽度：[2.8 3. 3.3]

花瓣长度[1.6 4.35 5.1]

花瓣宽度[0.3 1.3 1.8]

(2) 箱图：



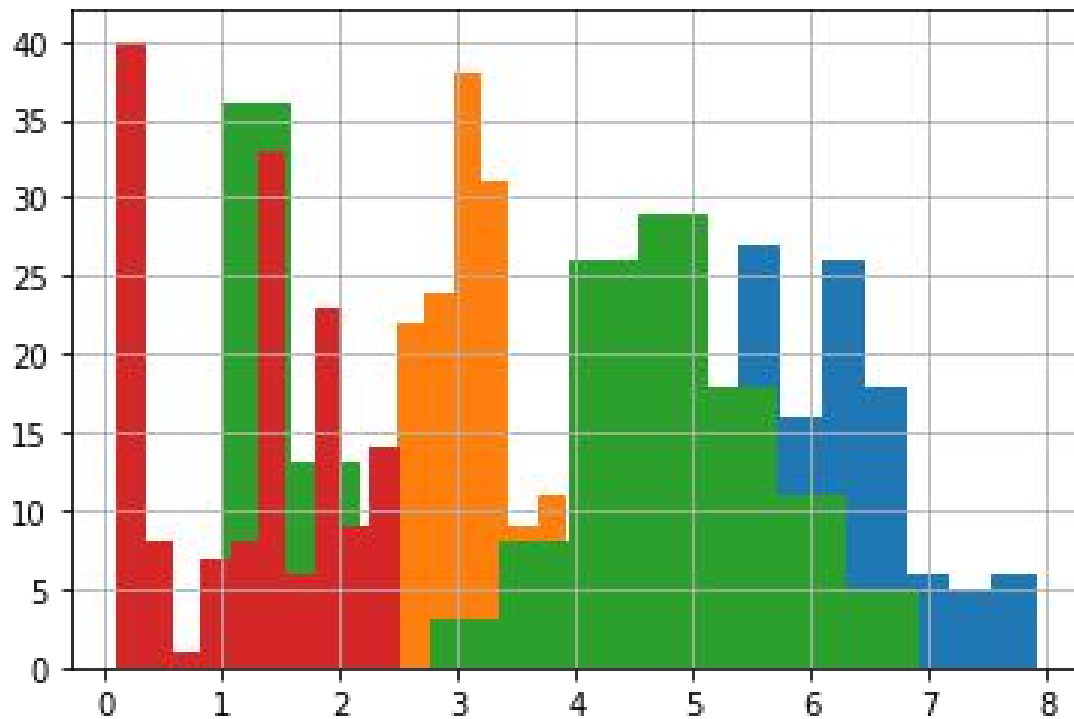
0: 萼片长度

1: 萼片宽度

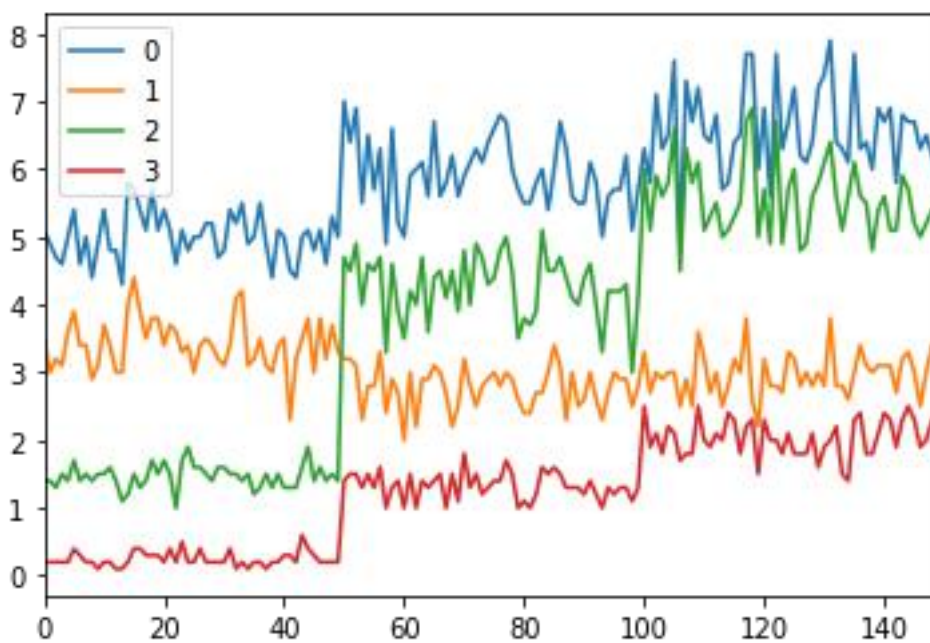
2: 花瓣长度

3: 花瓣宽度

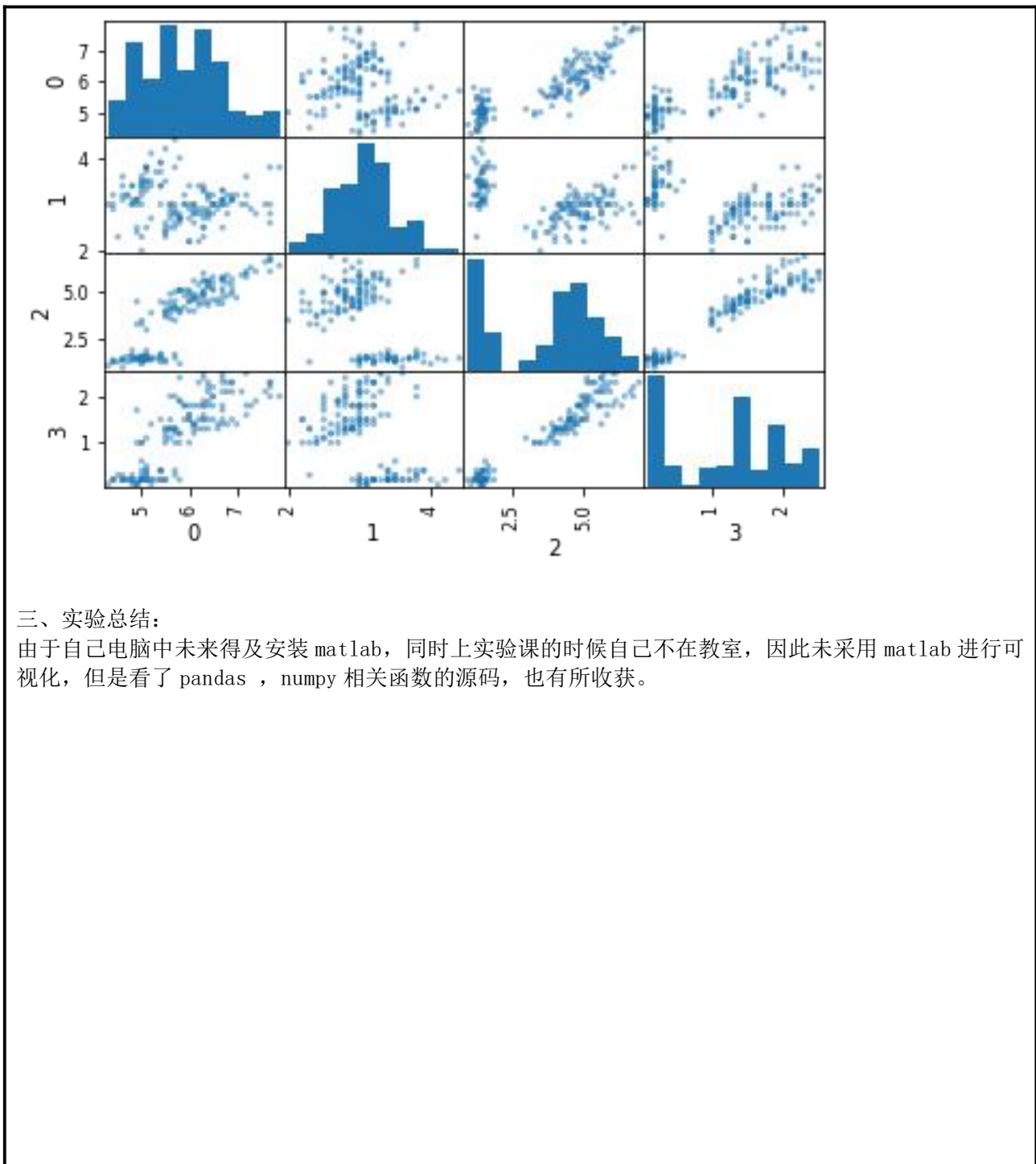
(3) 直方图



(4) 条形图



(5) 散点图矩阵



三、实验总结：

由于自己电脑中未来得及安装 matlab，同时上实验课的时候自己不在教室，因此未采用 matlab 进行可视化，但是看了 pandas ， numpy 相关函数的源码，也有所收获。