实验二 《数据统计和可视化》

题目

基于实验一中清洗后的数据练习统计和视化操作，100个同学（样本），每个同学有11门课程的成绩（11维的向量）；那么构成了一个100x11的数据矩阵。以你擅长的语言C/C++/Java/Python/Matlab，编程计算：

1. 请以课程1成绩为x轴，体能成绩为y轴，画出散点图。

2. 以5分为间隔，画出课程1的成绩直方图。

3. 对每门成绩进行z-score归一化，得到归一化的数据矩阵。

4. 计算出100x100的相关矩阵，并可视化出混淆矩阵。（为避免歧义，这里“协相关矩阵”进一步细化更正为100x100的相关矩阵，100为学生样本数目，视实际情况而定）

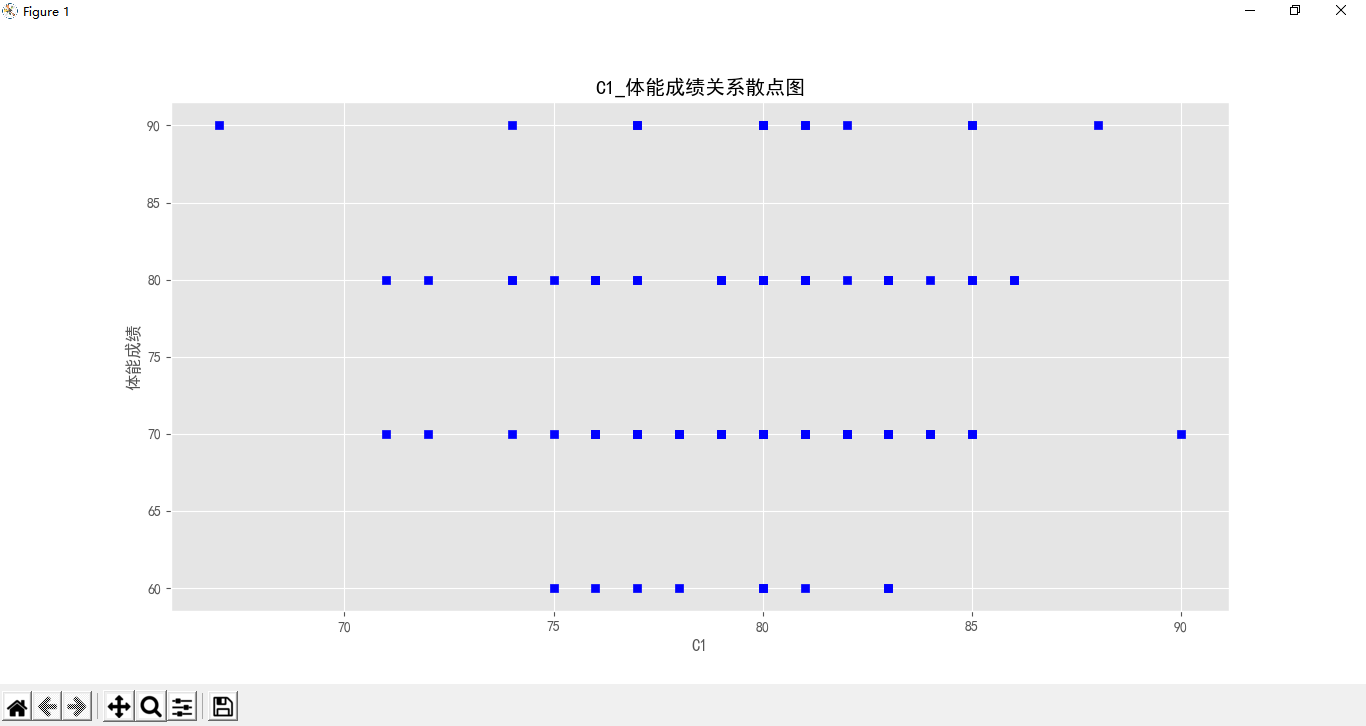
5. 根据相关矩阵，找到距离每个样本最近的三个样本，得到100x3的矩阵（每一行为对应三个样本的ID）输出到txt文件中，以\t,\n间隔。

提示：

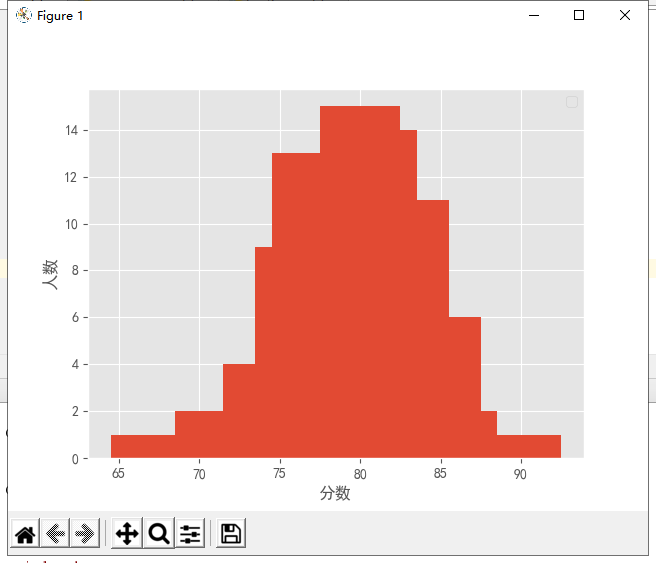
计算部分不能调用库函数；画图/可视化显示可可视化工具或API实现。

实验结果截图：

问题1散点图

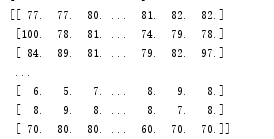


问题2直方图

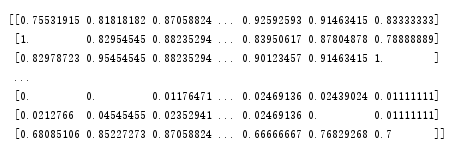


问题3

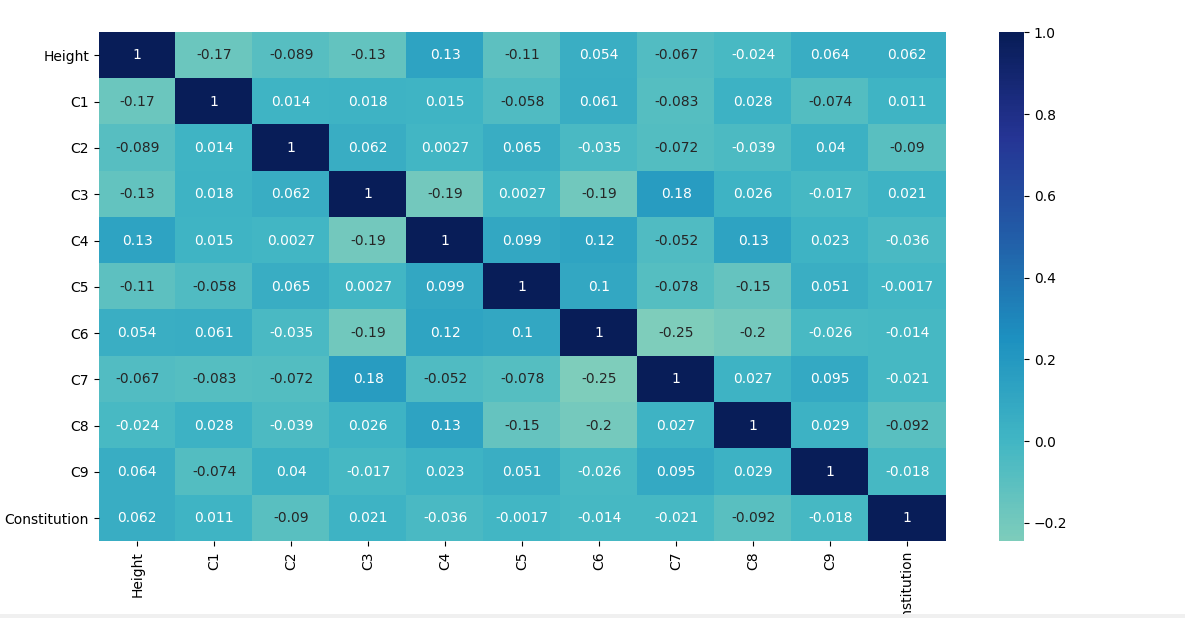
验证输出得矩阵



对应相关系数矩阵



问题四：



实验心得：

问题四和五其实弄得不是很明白这个相关系数到底要求的是哪个相关系数，因此没有做出来。不过画图部分的实验难度不大，多去网上搜索资料学习就可以了解要怎么做。