学业预警数据分析：

一：目的和思路：

|  |
| --- |
| 1. 各个学院出现学业预警的人数：   水平条形图barh()：x轴是各个学院，y轴是每个学院有学院预警的人数   1. 计算机学院学业预警（蓝色，黄色，橙色，红色）的人数占比：   饼图：pie()   1. 计算机学院每个年级出现学业预警的人数   折线图：plot()   1. 各个学院，不及格门次   散点图:scatter()  5，计算机学院2016级，每个班级出现学业预警的人数  直方图bar()  6, 计算机学院2016级每个班级出现的学业预警人数  堆叠直方图 |

二：数据收集准备

学业预警级别认定汇总表（2019-2020-1学期）.csv

三：数据处理分析

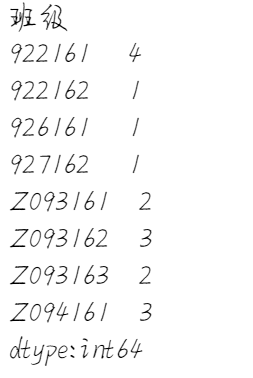
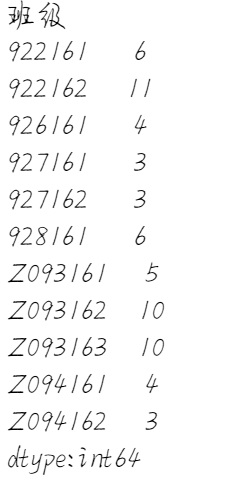
1. 数据清洗：

空值，缺失值，重复值

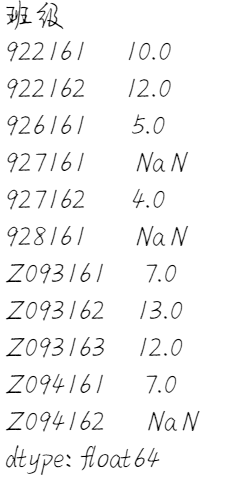
四：数据展现

遇到的问题：主要是问题六，堆叠直方图遇到的，在筛选出计算机学院的数据后，按照预警级别进行分组，再对每个预警级别按照班级进行分组统计人数，每个预警级别对应的人数作为y，预警级别作为x

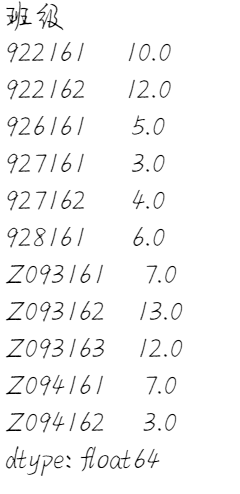
但分析数据的时候发现，有的班级并没有相应的预警级别，比如说，预警级别为橙色的班级只有八个，预警级别为黄色的班级有11个，这样在数据进行堆叠的时候就不能直接将数据进行相加，在这里就涉及到如何将两个len不同的series数据进行相加

比如和进行加法操作，索引相同的数值会进行

得到的结果如下：



而预期的数据为：



我解决这个问题的方法是，先创建一个索引值为所有班级的值为0 的数组init，然后将索引值为部分班级的数据加上init这样把len统一后，再进行加法操作，就能得到想要的数据了。