课程设计日报

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 | | 张胤民 | 分组编号 |  |
| 学号 | | 201694069 | 班级 | 软国1601 |
| 日期 | | 2017-7-5 星期四 | | |
| 当日工作总结 | 1. 为了更好的掌握数据分析的知识，我们小组今天共同进行了一个不大不小的挑战 2. 首先，我们基于各自昨天独立的学习，交流了各自对于包的掌握情况，交换了学习经验， 3. 然后，我们小组一致决定，挑战一个kaggle上面的入门级数据分析项目，我们拿到了近五十年来美国各个州不同枪支的持有数量以及五十年来美国包含经济领土住房等多方面的数据。   我们按照基本的数据分析流程，首先由高若宇对数据进行了初步的清理，使得数据更加易于分析，然后张胤民同学对数据进行了基本程度的分析，寻找不同因素对于枪支持有情况的影响，举个有趣的例子，我们发现在白人较多的州，随着黑人数量的上升，枪支持有率会上升，但是在黑人较多的州，随着黑人数量的上升，枪支的持有率却会下降。最后，刘东亮同学对我们分析出的结论进行了可视化分析和展示 | | | |
| 问题汇报 | 1. 对于数据清理我们遇到的第一个问题是数据的缺失，对于缺失的数据我们主要用两种方法进行数据填充，补0或者取均值，对于不合适填补的数据我们选择了丢弃 2. 数据分析困难的是如何找到两个有关系的量，我们利用包内置的函数可以一次性统览所有变量之间可能存在的大致关系，然后从中进行选择 3. 对于可视化来说，困难的地方在于选择一个适合这些数据的图表类型，解决的方法我们选择的是大量的对比和尝试 | | | |