**个人报告(CSE579)**

王毅Arizona State University[ywan1369@asu.edu](mailto:ywan1369@asu.edu)（  
1224641314）

1. **思考**

通过课程CSE598的学习，使我了解到，

对于数据的展示和分析，除了传统的数据表格，文本形式之外，还存在着可视化这样的的更加直观形象的展示方式。

通过这种可视化的展示方式，除了将数据本身进行展示之外，还能够通过，图形上的变化趋势对未来的变化进行趋势性的预测和判断。

感受比较深刻的主要有以下几个方面：

首先，通过数据可视化，能够比较清晰的发现数据的变化趋势，比如，对于项目中，职业和个人收入的可视化，就可以很容易的发现高收入人群的职业分布。

不同的可视化，对于不同的数据类型就有着不同的展示力度，只有选取最恰当的可视化图形，才能拿恰如其分的优雅的展示数据并预测趋势，比如，对于连续性数据年龄就很难通过饼图的形式进行展示。

当然，可视化的展示，需要首先对原始的数据集进行适当的数据选择和清洗之后，才能够更好的进行展示。整体来说，这种可视化的方式展示数据是直观的，高效的。

1. **分析**

尽管这已经是我学习的倒数第2门课程了，但参与这样的团队项目还是第一次。团队项目需要小组各成员之间相互配合，合理的分工给合作之后才能够较为合理的一起完成好项目。通过大家的不懈努力，最后完成项目，即完成了项目，学习到了可视化的课程知识，同时也加深了和同学之间的认识，了解。

1. **实际贡献**
2. 技术贡献

本项目中，本人的贡献，主要在于完成了3个可视化的实现部分，分别是

1. 使用词云，展示职业（occupation）和收入（income）的关系。采用以上数据的可视化，可以直观地看到收入50K以上和以下各种职业的分布，从而能够快速的确定哪些职业为高收入人群，这样的人群一般更愿意且有能力支持投入学习中。和中期的报告相比，将此项中的教育替换成了职业
2. 使用堆叠柱状图，展示sex,education,hours\_ per week 之间的关系，采用以上几种数据类型，从而揭示2性不同学历和工作时间之间的关系，从而大致明确2性中各自什么样的学历人群具有较长的工作时间，什么样的学历人群具有较短的工作时间，从而便于设计针对性的教学内容，预期2性时间充裕的工作不一样。和中期的报告相比，将此项中的职业替换成了教育。
3. 使用气泡图（一种变形的散点图），展示婚姻状况（marital\_status），工作类型（workclass）,年龄（age）之间的关系。如图示，红色代表较为年轻的群体，蓝色为中间群体，绿色群体年龄较大，气泡的大小，代表这个群体数量，越大说明人数越多，越小人数越少，从而可以观测不同年龄段的人群所属的工作类型和婚姻状态。
4. 非技术贡献

本人编写了项目报告中与本人完成的3个可视化相关部分的报告内容的编写，以及个人报告的编写部分。