**多视图可视化 作业报告**

信息科学技术学院熊伟民

1900011621

1. **数据描述和分析**

本次可视化设计选择的数据内容为明代进士户籍数据。对于每个数据项共有年份、姓名、排名、籍贯、户籍、科目、家族信息、官职等属性。因为这次可视化需要自选分析的数据的维度，并且要实现一个多视图可交互的可视化，需要充分考虑可视化数据的内容，例如两个维度之间是否有直接的关系。经过权衡，我最终选择了户籍、籍贯、年份、科目四个维度的信息。选择籍贯这一维度可以帮助利用之前等方树图的可视化代码来观察不同地区进士的分布关系；选择时间这一维度可以帮助我们画出时间轴，并沿时间轴方向添加缩放组件，研究进士数据随时间的变化；选择户籍和科目维度一方面因为这两个维度的种类比较少，毕竟容易对原始数据清洗，另一方面可以从其他维度帮助我们分析明代进士分布情况。例如我们可以通过刷选和过滤，选择指定科目、户籍、时间或者籍贯，控制变量地研究进士分布与某一维度之间的关系。

在可视化工具的选择方面，考虑到这次要绘制的图表比较多，使用d3和jquery逐个绘制会比较麻烦，因此我选择具有良好封装的echarts作为每个图表绘制的数据可视化图表库。一方面echarts省去了排布每个可视化元件的麻烦，可以直接在setOption()的参数中指定需要选择的图表、需要加入的功能以及要可视化的数据。另一方面echarts具有默认的设计良好的css，不需要再专门设计图表的样式，具有良好的视觉显示效果。不过，因为echarts的封装比较彻底，在加入图表交互的时候比较困难，需要其他javascript库的帮助。

在本次数据的清洗过滤方面，我选择了类似于上次等方树图可视化的数据清洗方法，对于籍贯数据，我复用了上次数据清洗的代码，对于户籍数据，我经过在维基百科上的查找，将所有进士分为民籍、军籍、官籍、匠籍、灶籍、监籍、站籍几类。对于科目数据，我将所有进士分为选考诗经、书经、易经、礼记、春秋几类。最后生成 data.js，huji.js，kemu.js三个数据文件，保存在js文件夹中。相关处理数据的文件calculate\_time.py，filter.py，manage\_json.py，process\_jiguan.py，xlsx\_to\_json.py保存在script文件夹中。

1. **设计的宗旨和目标**

我在这次可视化中总共采用了四种图表，折线图、柱状图、等方树图、饼图。设计的目标是经过刷选和过滤，定性或者定量分析户籍、科目、地区等属性与进士人数的关系。设计的宗旨是保持数据的准确性和完整性；提供清晰的上下文帮助用户浏览数据，构建功能时，优先考虑数据的检索和比较功能；减少认知负担，专注于真正重要的东西。使用视觉设计来传达层次结构，并提高一致性。

选择折线图，优势在于折线图是读者都很熟悉的形式，非常适合于一目了然地表现数据趋势；缺点是只能关注到轨迹线，比较困难看到和探讨离散的数据点。选择柱状图，优势在于比较适合类别间的简单比较，缺点是多组条形可能变得难以解析。选择饼图，优势在于可以很好地显示主导份额和非主导份额；缺点是人们对于扇形楔形快的面积估计得不是很好；如果楔形块过多，将使得值难以区分和量化。选择等方树图，优势在于可以很清晰的看到数据的层次关系，缺点在于需要较多的计算，同时值很小的数据项难以被观察到，增加读者的视觉负担。

1. **可视化结果描述**

户籍图表