

# 可交互多视图可视化作业报告

张启哲 1900011638

2021 年 11 月 29 日



## 目录 (Contents)

<b>1 数据描述与维度分析 (Data Description &amp; Dimension Analysis)</b>	<b>1</b>
1.1 数据描述 .....	1
1.2 维度分析 .....	1
<b>2 设计目标与宗旨 (Design Objective)</b>	<b>1</b>
2.1 堆叠面积图 .....	2
2.2 堆叠柱状图 .....	2
2.3 矩形树图 .....	3
2.4 交互设计 .....	3
<b>3 可视化结果描述 (Visualization Result Description)</b>	<b>4</b>
3.1 结果展示 .....	4
3.2 总结与分析 .....	4

# 1 数据描述与维度分析 (Data Description & Dimension Analysis)

## 1.1 数据描述

本次可视化作业使用的数据是来自于China Biographical Database Project(CBDB)中 Ming Jinshi List 的明进士数据。数据集包括明代 52 年科举考试的所有进士原始资料，共 14116 人。

数据集以 excel 数据 xlsx 的格式呈现，共有 3 个工作表。第 1 个工作表 Ming Jinshi Lists 中包括了所有进士的详细信息，第 2 个工作表 Release Notes 中描述了 3 个工作表中的内容并列出了各次科举考试的年份和资料出处，第 3 个工作表 dazi rule 描述了进士数据项的格式。

## 1.2 维度分析

数据集中的数据项包含了出处、年份、姓名、甲次、名次、籍贯、户籍、科目、年龄等諸多属性，而部分属性在某些年份的数据存在缺失，故本次多视图可视化项目主要采用了数据完整的 5 个属性 (年份、户籍、科目、甲次与籍贯) 作为分析与展示的维度。

年份维度：1592 年的进士户籍记录有明显异常，可视化结果中不予展示，其余年份的数据无误，但与数据集给出的 52 年这一数字有偏差，除去 1592 年剩余共 50 个年份；户籍维度：根据《明史》、《后湖志》等古籍记载，明代的户籍主要分为军、民、匠、灶四类，考虑到数据集本身的记录，本项目将进士户籍共分为民、军、官、匠、灶 5 类；科目维度：按照科举考试的常规科目五经分为诗、书、礼、易、春秋 5 类；甲次维度：按照科举考试结果分为一、二、三甲，其中一甲共 3 人，第一名为状元，第二名为榜眼，第三名为探花，二甲每年为十几人到一百多人不等，三甲则为一百人到三百多人不等，最多不超过三百五十人；籍贯维度：按照行政区划分为司、府两个行政级别，其中司 17 个、府 146 个。

数据分析时将年份与户籍类型组织在一起，可以了解到考生户籍类型随时间的分布情况，有助于分析明代各社会阶层在不同年代的生存与发展状况；将考试科目与甲次组织在一起，可以了解到各考试科目的热门程度以及对不同分段考生的偏差；籍贯维度单独呈现，分为司、府两个级别，展示社会精英的地域分布特点，也反映出各地区的教育水平差异。

# 2 设计目标与宗旨 (Design Objective)

本项目的设计目标是通过多视图的形式将不同维度有机结合，并通过不同视图之间的交互动态地分析高维数据，宗旨是结构清晰、布局合理、操作简洁。

具体设计方面，采用了 3 个视图结合的方式布局。整个画布的左上角通过堆叠面积图 (StackedAreachart) 的形式将年份与户籍类型两个维度结合查看，右上角通过堆叠柱状图 (StackedBarchart) 的形式将考试科目与中举甲次两个维度结合查看，下方则通过矩形树图 (SquarifiedTreemap) 的形式呈现进士的籍贯分布，有司、府两个行政级别可以切换。

上方的 2 个视图可以通过点击交互互相控制，而下方的树图则可以同时控制上方的视图。下面具体介绍 3 个视图以及交互设计。

## 2.1 堆叠面积图

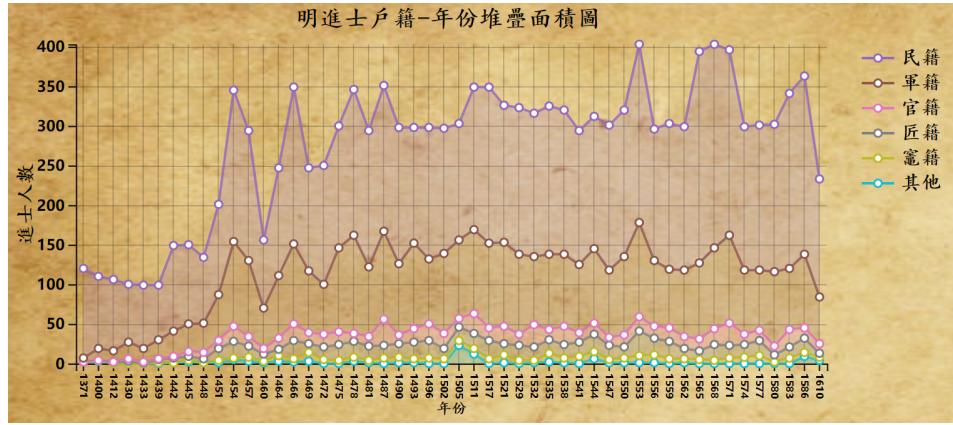


Figure 1: 户籍 -年份堆叠面积图

堆叠面积图中，横轴为年份，从 1371-1610 共划分为 50 次，纵轴为进士人数，按照其他 - 灶籍 - 匠籍 - 官籍 - 军籍 - 民籍的顺序从下到上堆叠，不同颜色代表了不同的户籍种类，每个小圆圈代表一个数据点，线与线之间的距离代表了该类户籍的进士人数，面积表示该户籍的总人数，而最高点则表示了当年考取的进士总人数。

## 2.2 堆叠柱状图

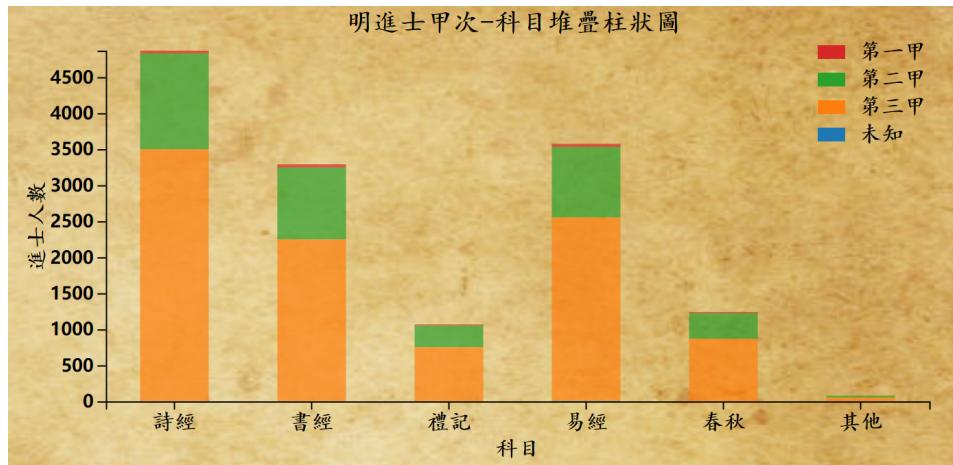


Figure 2: 甲次 -科目堆叠柱状图

堆叠柱状图中，横轴为考试科目，共有诗经、书经、礼记、易经、春秋与其他 6 类，纵轴为进士人数，按照未知 - 第三甲 - 第二甲 - 第一甲的顺序从下到上堆叠，不同颜色代表了不同的甲次，柱子的总高度则代表了该考试科目下考取的进士总人数。

## 2.3 矩形树图



Figure 3: 籍贯矩形树图

矩形树图中，通过 SquarifiedTreemap 可视化了按照行政区划层次结构组织的进士数据，分为司、府两个行政等级，可以点击切换。

## 2.4 交互设计



Figure 4: 动态数据项显示



Figure 5: 刷选帮助显示

**动态数据项：**当鼠标移至数据项上时，画布左侧会显示对应数据项内容：堆叠面积图(年份、户籍、进士人数)、堆叠柱状图(科目、甲次、进士人数)、矩形树图(行政级别、行政区划、进士人数)。

**动态刷选交互：**鼠标点击堆叠面积图上的小圆圈时，右侧堆叠柱状图会按照对应年份筛选；点击不同颜色面积区域时，右侧堆叠柱状图会按照对应户籍类型筛选；点击堆叠柱状图中的立柱时，左侧堆叠面积图会按照对应考试科目筛选；点击下方矩形树图中的小矩形时，上方 2 个视图均会按照对应行政级别与地区进行筛选。同时，树图上方的司、府级切换按钮可切换不同行政级别；上方的 4 个与右侧的 1 个重置按钮均可重置筛选条件。

### 3 可视化结果描述 (Visualization Result Description)

#### 3.1 结果展示

代码库已上传至 github: [Theia-4869/Interactive-Multiple-Views](https://github.com/Theia-4869/Interactive-Multiple-Views)。

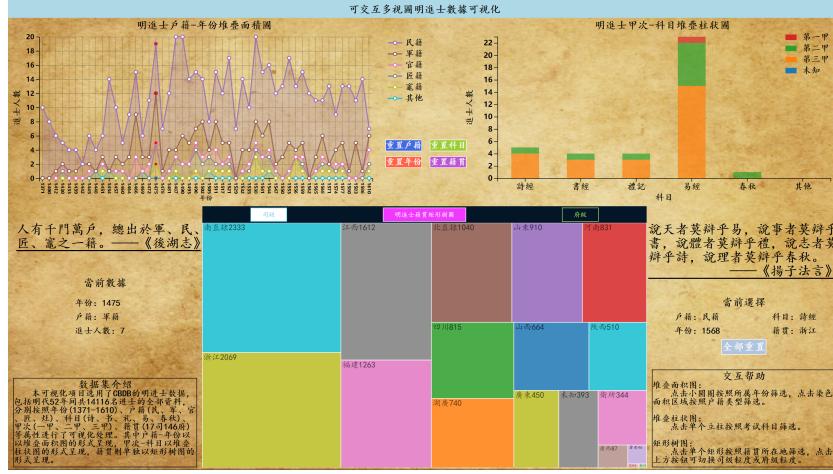


Figure 6: 可交互多视图明进士数据可视化

#### 3.2 总结与分析

从数据集的统计结果来看，1451 年后进士数量迅速增长，从之前的每年 100 人作业增长到 300 人以上并在以后基本保持稳定，进士人数最多的年份为 1568 年，共 403 人，最少的则为 1433 和 1439 年并列，仅有 99 人。户籍为民籍的考生有 8404 人，占全部考生的 60%，军籍次之，占到 28%，其余户籍的考生就较少，这可能是由于科举考试是改变民籍百姓的几乎唯一出路，参军的青年大多身体素质较好但成绩和思维能力就相对差些，官家的小孩无论成绩如何在看重阶级身份与血统的社会都可以衣食无忧，而匠籍与灶籍的百姓平日则要承受繁重的劳役，加之经济情况不好，很难拥有参加科举考试的机会和条件。从考试科目来看，选择诗经的考生最多，达到了 4867 人，占 34%，易经与书经次之，占 23%~25%，春秋与礼记最少，只占到 8% 左右，值得一提的是，一甲率最高的科目不是选择人数最多的诗经，而是书经 (1.3%) 与易经 (1.2%)，这大概是因为诗经科目较为常规，难有新意，书经和易经有时更能给考官带来惊喜，而选择人数过少的春秋和礼记则较难。而考生的地域分布依然与今日的全国教育格局类似，南方沿海地区教育水平高，中举人数多，西北部和偏远地区则相对落后，进士人数较少。

从多视图的交互来看，不同年份与户籍的考生在考试科目的选择上基本一致，都是诗经 > 书经、易经 > 礼记、春秋，这可能与不同典籍的内容及考试的题目类型有关，是长期的考试经验决定的，数据也证明了这一点，在 1460 年之前选择各科目考生的数量比例还时有波动，但 1460 年后比例基本稳定。从单个科目来看，春秋在早间年更受考生们的偏爱，而易经则成为了后来考生的坚定选择。地域分布方面，同样是南方考生的程度由于北方，进士人数多，一甲人數也多，而各地区考生在户籍类型的分布上也与全国平均水平相近 (除卫所这一特殊的军籍管理行政机构外)，又从另一个侧面证明了上文中解释的合理性。

从整体设计来看，整体布局与操作较为简洁，可以考虑采用更多的视图与不同的图表类型展示不同维度数据之间的关系，也可以添加更多的交互机制为用户分析数据提供便捷与选择。