

# 多视图的、可交互的可视化

---

谢悦 1900013055

## 一、选择数据维度

来自 China Biographical Database Project (CBDB) 的明代进士录记录了明代 52 年科考进士的相关信息，本作业选择进士年代、籍贯、户籍、科目、年龄五个比较重要的数据维度对数据进行可视化。

### (一) 数据各维度含义：

年代是一个进士科举中进士的年份，进士信息和明代历史息息相关；

籍贯是一个进士所属地区的司府县信息，不同地域的进士数会反应地区发展水平、文化水平；

户籍是一个进士居民身份（有民籍、军籍、官籍、匠籍、灶籍、监籍、站籍），代表着进士的出生，稀有身份的进士涌现某种程度可以反应明代底层（除官籍）民众的进步；

科目是一个进士科考命题范围的科目（有诗经、易经、书经、春秋、礼记），不同科目某种程度上反应国家文化思想状态；

年龄是一个进士中进士时的年纪，年龄某种程度上可以反应特定情况中进士的天资和为国效力精力，虽然整体大概率会呈一个高斯分布。

### (二) 数据维度之间的关系：

就这五个数据维度而言，有很多可以探讨的维度之间关系，例如：

1. 年代与籍贯：籍贯可视化可以看出明代那些地区盛产人才，而跟随着年代的变化，各个地区发展不断变化，其为朝廷提供人才的占比人数也会有不一样的改变。
2. 年代与科目：明代学风先受程朱理学影响，然后又受陆王心学影响，且明代“八股取士”，科举考试在中期一度繁荣，年代变化，历史因素会给进士的科举科目分布会带来什么变化？
3. 年代与户籍：不同户籍，随着年代变化，数量是否会有呈某种趋势改变，明代科举制度不断壮大，各户籍进士人数是否都会呈增长趋势？
4. 籍贯与科目：同一地区中进士考试科目的分布，是否存在某一个地区某一门科目发达，大多喜欢考同一门科目？
5. 籍贯和户籍：是否某一个地区的进士大多是同一个户籍，或者某一个户籍的进士大多来自于某一个地区？
6. 户籍和科目：户籍大概率代表着一个进士的出生，以及其受教育经历，特定的户籍是否存在不同比例的科目分布？
7. 年龄与籍贯：来自明代发达地区的进士是否普遍中进士入仕年龄较早？不发达地区是否普遍入仕年龄较迟？年龄还可以与其他维度联系，反应某种特定情况进士水平。

8. 多个维度下的一个维度上的数据：比如某一段年间，同一个地区，进士所选科目是否有特殊；或者某一段年间，同一个地区，进士年龄分布是否普遍偏早。这样在确定某几个维度情况后，进士在其他维度下的分布水平。

## 二、设计宗旨及目标

本作业设计宗旨是希望从年代、籍贯、户籍、科目、年龄这五个维度下描绘明代进士数据，做到数据间最大自由度联动，方便明代进士数据分析者，更好更自由的在这张可视化上验证自己对明代进士规模和趋势的认知；且可以对数据有效筛选，分析特定情况下某一维度明代进士数据的规模或趋势。

故本作业采用 **Squarified Treemap** 展示籍贯信息，多颜色分类折线图展示户籍信息，饼图展示科目信息，折线图展示年龄分布。利用时间框选组件让使用者自选年代范围，然后可视化上的所有展示的数据都是所选年代范围之间的。通过选择特定的户籍，所有数据在当前基础上滤去其他户籍的数据。通过选择特定的科目，所有数据在当前基础上滤去其他科目的数据。选择进士所属司府县，所有数据在当前基础上限定到所选择的地区信息中。

用年龄范围做筛选对其他数据意义不大，故年龄数据只作为一个数据展示，展示特定情况下进士年龄分布，折线图较为直观的展现了特定情况下进士年龄分布，同时突出了很年轻就高中进士的“年轻学霸”，或者暮年入仕的科举“忠实爱好者”这类个例，为了方便使用者查到具体某些情况下进士的姓名及详细信息，点击特定年龄，本可视化可以列出符合该情况所有进士的详细信息。所有的交互都可以叠加且可以通过重复操作回退撤销，由于所用 **echarts** 库限制，**teemap** 对数据的筛选只能放在最后，优先级是最低的。

### 优势：

- (1) 这样的可视化有很高的交互自由度，功能较为完善灵活，让使用者更加自由主观的分析不同维度之间的关系，明代进士录是很多维度的数据，通过选择和科举考试非常相关的几个维度并进行高自由度的交互，可以满足对该数据感兴趣的研究者多维度的猜想和认识。
- (2) 树图合适的表达了籍贯这样一个多层数据不同层级（司府县）中进士人员分布，较好地展现了进士按地理维度的分布，使用者通过交互选择，可以实现在国家选拔人才的考试中，对优秀生源地的探索研究。
- (3) 折线图较为清晰的展现了一特定情况下进士的年龄分布，一定程度上反应了该特定情况下进士水平，数据样本越多，该分布呈高斯分布越明显，同时可以让年龄过小和年龄过大这类特殊样本，世人感兴趣的样本浮现出来。
- (4) 对于户籍和科目数据，分别采用了以年代为 **x** 轴的分类折线图和饼图，饼图可以很好地展示同一情况下科目的分布，折线图展示了随年代变化下各个户籍人数的分布及变动趋势，某种程度上反映了科举的发展趋势。

### 不足：

- (1) 较高的交互自由度容易给使用者带来不必要的观看干扰和较高的理解难度，本可视化视图排版还不够优化，信息传达还不够直观，在这方面需要深入分析。存在使用者只想探索某几维度下数据信息，但收到其他维度的干扰的情况。几个维度在含义上很难有逻辑顺序，让使用者找对应维度的图表时存在一定困难。这些还没有仔细分析。

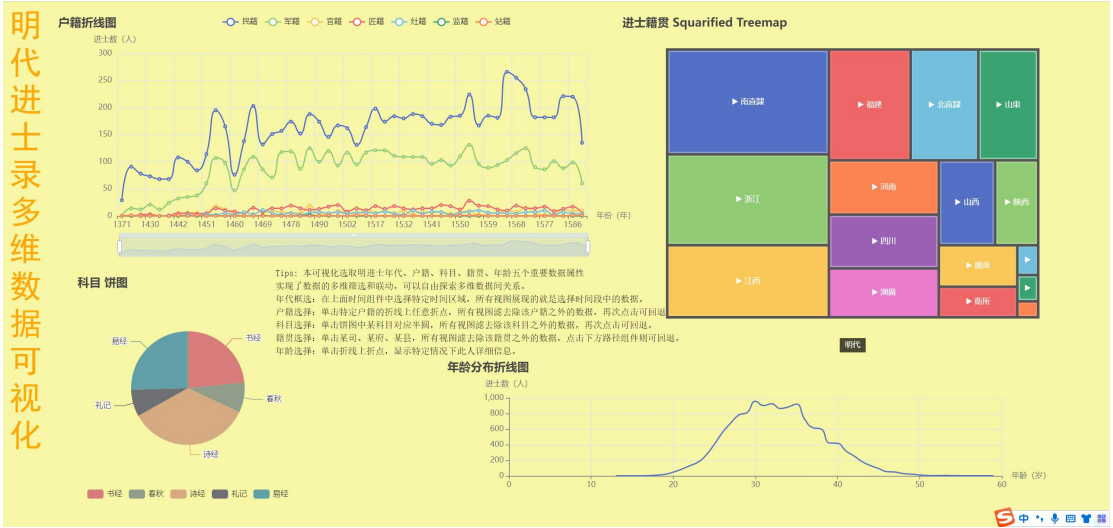
(2) 户籍之所以选择折线图，一方面是想将其和年代联动，展现随年代变化，进士人数分布趋势及不同户籍进士人数的占比变化，另一方面因为存在很多类人数不多，占比极小，饼图难以展示，但折线图未加上户籍在 y 轴上累加的功能，达到累加和非累加可相互切换的效果，展现户籍占比方面稍欠缺

(3) 由于通过饼图展现科目的分布较为简单，不容易和年代维度联系起来展现出关系，但由于总体可视化折线图过多，看上去容易视觉疲劳和造成混乱，故科目数据直接选择了饼图，这方面可视化方式选取还有待改进。

(4) 树图在显示层级和通过交互点击选择数据范围之间还有待修正，目前是只显示一个层级，在树图上向下点击则代表选择了数据范围。这可能需要使用者一定程度的适应和理解。

### 三、可视化结果

本可视化使用 javascript 完成，主要使用 echarts 库对图表进行绘制，可视化界面如下：



左上角是户籍随时间分布的折线图，左下角是展示科目分布的饼图，右上角展示籍贯分布的树图，右下角是展示年龄分布的折线图。

交互效果如下：

年代框选: 在上面时间组件中选择特定时间区域，所有视图展现的就是选择时间段中的数据。

户籍选择: 单击特定户籍的折线上任意折点，所有视图滤去除该户籍之外的数据，再次点击可回退。

科目选择: 单击饼图中某科目对应半圆，所有视图滤去除该科目之外的数据，再次点击可回退。

籍贯选择: 单击某司、某府、某县，所有视图滤去除该籍贯之外的数据，点击下方路径组件则可回退。由于 echarts 使用限制籍贯选择优先级最低。

年龄选择: 单击折线上折点，显示特定情况下此人详细信息。

以上所有的交互都可以叠加且可以通过重复操作回退撤销，

发现：

通过这份可视化，我对明代进士录数据有了更广泛透彻的认识：

（1）贯彻明代历史，科举考试不断发达，其为国家输送的人才也随着年龄越来越多，这通过户籍随时间的分布图可以看出，明代刚开始国家还从动荡中休养生息，科举考试文化事业还不够繁荣，后期随着年代变化，不断繁荣这一点通过户籍折线图可以发现。

（2）且明代中后期，按历史发展，收到陆王心学的影响以及八股取士，程朱理学的影响，出生寒微户籍（匠籍、灶籍等）的科举进士有星星点点涌现，这在折线图上获得了验证。

（3）随着年代变化，明代各个地区发展程度也可以通过拖动时间线看籍贯树图得到，随着年代发展，科举不断繁荣，很多欠发达地区为朝廷输送人才的数量不断增多，占比增大，如山西、山东、四川。随年代变化，可以看出各地进士占比方差不断减小，某一程度反应各地文化水平都在不断发展繁荣，形成逐渐势均力敌场面，树图反映了南直隶、江西、福建、浙江主要领先，可以看出这些司主要汇聚在江南以及沿海一带，这和明代海事发达，沿海及江南制造业文化相对发达有着密不可分关系。

（4）“诗经最多、易经次之、书经再次之，修春秋，礼记少。”饼图直观反映了史书上记载的这一条信息，从大趋势来看，满足这样，但比如选择南直隶-苏州，修易经的占大部分，这反映了，可能存在同一个地区某一门科目比较发达，有很好的老师或者教育体系，盛产此类科目下的人才。

（5）年龄分布传达了明代这么多进士中，存在 13 岁、16 岁高中进士的“学霸”各一人，最大不过 60 岁，且可以看出修春秋、礼记的进士年龄普遍集中在 35 左右，比总数据在 32 左右要偏后，可能传达出这些科目比较难考。