# 可视化第二次作业 Part 1

厉扬豪 1100012740

2014年11月5日

### 1 数据描述与分析

本次作业的处理数据对象是关于一个 hotel 的数据。数据中统计该 hotel 各个月各项指标的数据。比如客人是女性的比例,当地人、美国人、外国人的比例,年龄,是否是商人的比例等等,共计 20 个属性,而月份则对应一年的 12 个月份。数据原始为20\*12 的表格,这样显示数据很不直观,难以发现数据中特性。而可视化正是考虑将数据换一种方式呈现出来,使得数据更加直观,更容易发现其中性质和它们之间的关系。

首先通过分析 20 个属性, 我们可以发现这些属性这一分为不同的类别:

- 1. 性别: Female:
- 2. 地点: Local, USA, South America, Europe, ME Africa, Asia;
- 3. 方式: Direct Reservation, Agency, Air Crews;
- 4. 旅客类型: Businessmen, Tourist;
- 5. 年龄: <20 years, 20-25 years, 45-55 years, >55 years;
- 6. 入住时间: Length of stay;
- 7. 入住率: Occupancy;
- 8. 是否有会议: Conventions.

对应 hotel 的数据,可视化的时候应该注意到两点:

- 1. 对应每个属性的数据,都可以用直方图等图形的方式展示出来,更容易看到一年变化的趋势,并且不同的类别可以分类显示
- 2. 可以对相同类别以及不同类别的不同属性的数据进行比较。

#### 2 设计宗旨和设计过程

设计宗旨主要是考虑能够将不同属性的数据可以展示出来,并且方便比较。另外用户可以自己选择相同类别甚至不同类别的不同属性进行比较,给用户充分的自己进行比较探索的空间。

基于此以及上一节提到的注意点,在设计的时候,呈现每个属性的数据时,主要使用直方图的方式呈现,利用直方图的大小展示数据的大小。在一个视图中,将 20 个属性的直方图一同展现出来,并且按照不同类别分好类,给用户一个全局的视图。每个直方图可以进行点击选择,用户确定之后,则可以将选择的几个直方图按照点击的顺序,在上面的大图中一同展示出来,方便用户更清晰观察数据,比较不同属性的数据。

具体设计时,小图的直方图用不同类别的颜色表示,在点击之后变成红色,而大图 展示的直方图用彩色表示,展示出月份中数据的变化。

#### 3 可视化结果描述

最终的可视化视图主要分为两个部分,一个是用于展示所有直方图的视图,并支持点击选择,如下图:

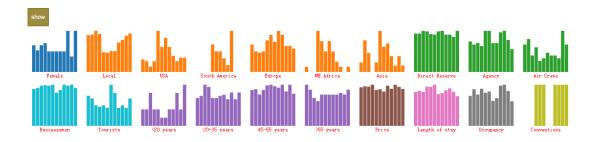


图 1: 视图 1

点击后的直方图会变成红色,如下图:

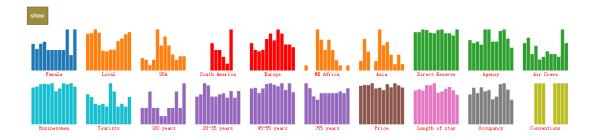


图 2: 视图 1 点击后

再点击左侧的"show"按钮,就会将选中的直方图显示到下一个视图当中,如下图:

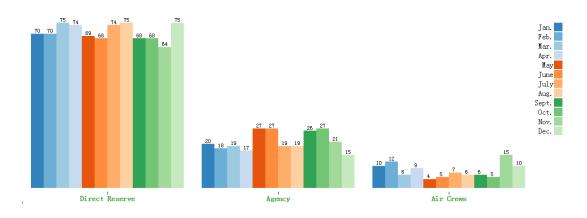


图 3: 视图 2

视图 2 中不同月份也用不同的颜色来表示,而一个直方图下的文字的颜色则代表了 其属性的类别,比如上图的类别就都是绿色的。

通过视图 2, 用户可以更加方便地观察和比较数据。比较的方式主要也分为两种:

1. 一种是类内的比较,比如通过图3,我们就可以发现 Direct Reserve 的比例占绝大数等。又如图4,我们不仅发现 Local 的比例占绝大数,同时也能发现不同地区的比例也不同,在五到十月份 USA,Europe,Asia 等的比例就会上升,Local 的比例下降等。

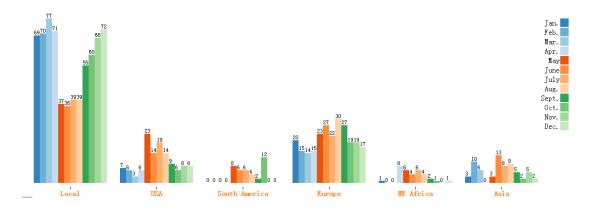


图 4:

2. 另一种是类间的比较,可以发现不同类别属性之间的联系。如图5,我们发现在没有会议的月份,Tourist 比例就高,而有会议的时候 Businessmen 比例就高。诸如此类可以发现其他类似的联系。

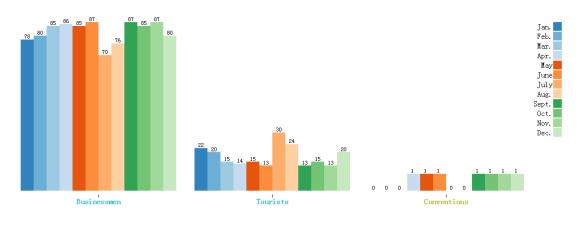


图 5:

## 4 感想和总结

这次的可视化作业,因为是第一次,所以刚入门的时候十分痛苦,熟悉 d3 花了不少时间,不过做完之后还是学习到了不少东西,也算是开始起步了,位置后的作业打下了基础。另外看到其他的同学的一些设计也很有启发,包括用饼状图表示百分比,以及不同视图之间的联动等。