

可视化第二次作业 Part 2

厉扬豪 1100012740

2014 年 11 月 12 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 Wine 的数据。数据中每一行代表了一类酒，一类酒除了第一维表示酒的类别，共有三种酒之外，其他 13 维分别代表了 13 个属性，这是一个典型的高维数据。对此在实现可视化的过程中，主要也考虑的是高维可视化，注意点主要有：

1. 展示三类酒在各个属性的不同之处
2. 体现各个属性之间的相关性和差异性

2 设计宗旨和设计过程

设计宗旨主要是考虑既要能展示出高维属性的特点，又能让用户有一定的交互性，方便发现属性之间的联系。

基于此以及上一节提到的注意点，在设计的时候，首先考虑用平行坐标展示所有属性，并且三种类型的酒使用不同颜色表示。平行坐标还支持拖动各个维度，使其按照不同的顺序显示；在一个维度上选择一定的范围，是选定的数据高亮。另外还同时有一个表示信息的视图，用来展示各个维度不同类型酒的均值方差，以及选中数据点的个数，各种酒的比例等。

同时为了展现不同的属性的相关性，考虑使用散点图（ScatterPlot）来体现，考虑到 13 维维数太大，支持用户自己选择选定的特征维进行展示。

3 可视化结果描述

最终的可视化视图主要分为三个部分，一个是用于展示所有属性的平行坐标图，如下图所示：

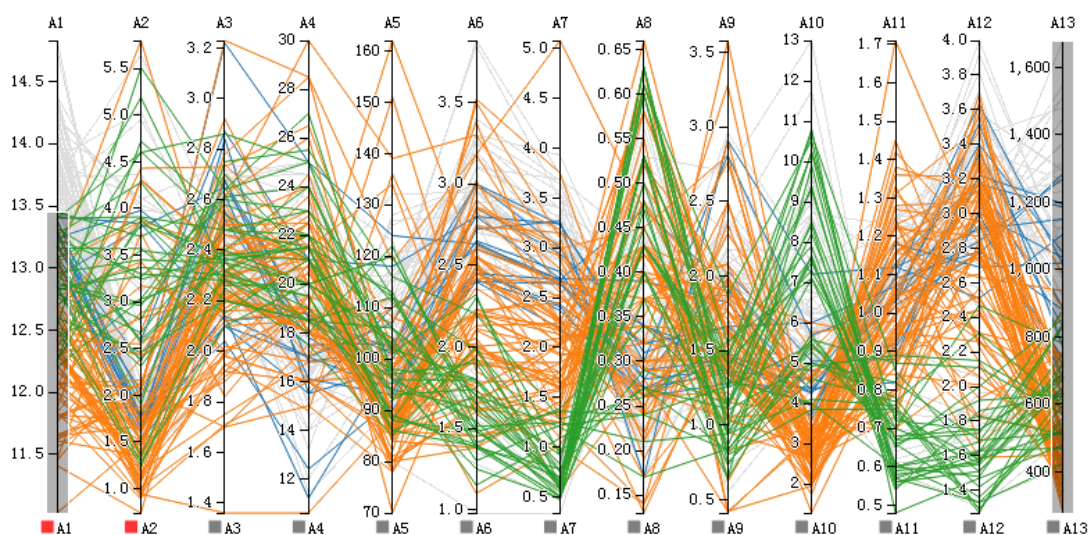


图 1: 视图 1

支持的操作主要是：

1. 拖动各个维度，不同顺序展示；
2. 在每个维度上都可以选择一定的范围进行展示，选中的数据高亮；
3. 下方方框点击支持将当前维度加入或推出散点图。

第二个是右侧的信息视图，用于同步展示左边在平行坐标上操作时，选中数据点对应的一些信息，如下图：

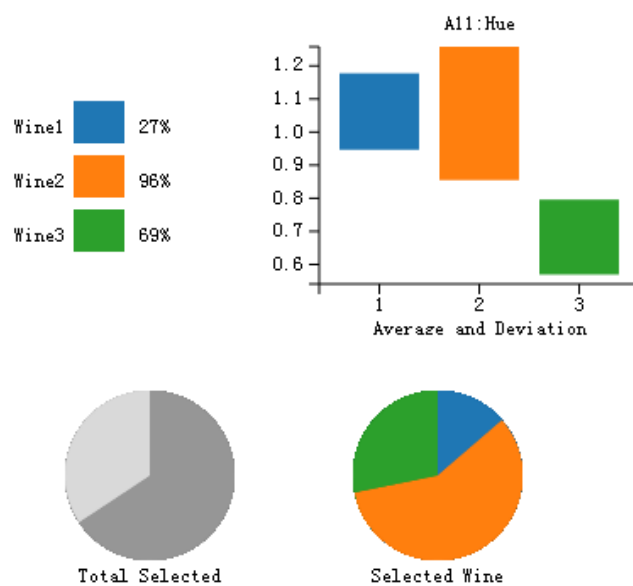


图 2: 视图 2

展示的数据信息主要有:

1. 选中的数据，占三种酒各自的比例；
2. 鼠标放置的特征维数，三种酒对应的均值和方差情况；
3. 选中数据的比例和选中数据中三种酒的比例。

第三个视图就是用来展示选中的维度，组合成的散点图，选中的数据高亮显示，这和平行坐标中的操作也是同步的，如下图：

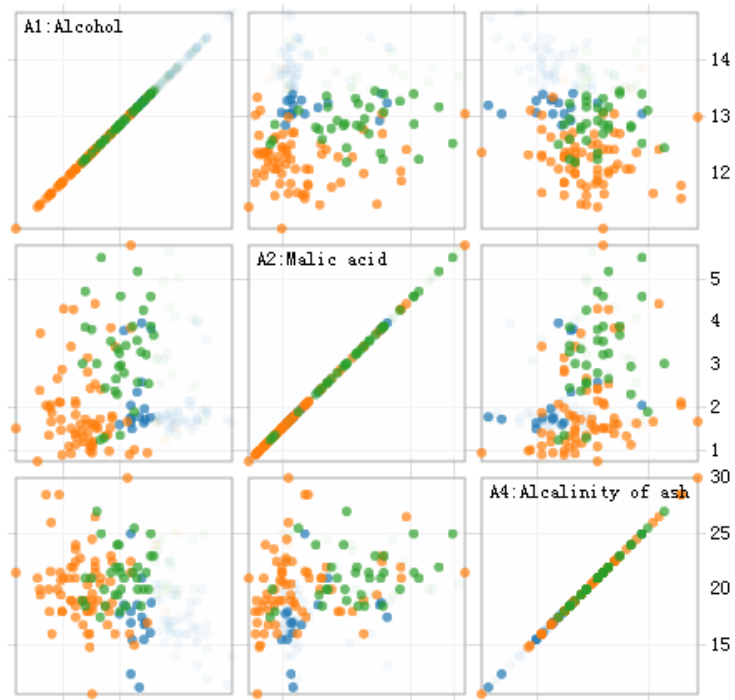


图 3: 视图 3

通过上面三个图的联合展示，用户就可以更加容易地发现三种酒的特点，以及不同属性之间的联系等。比如：

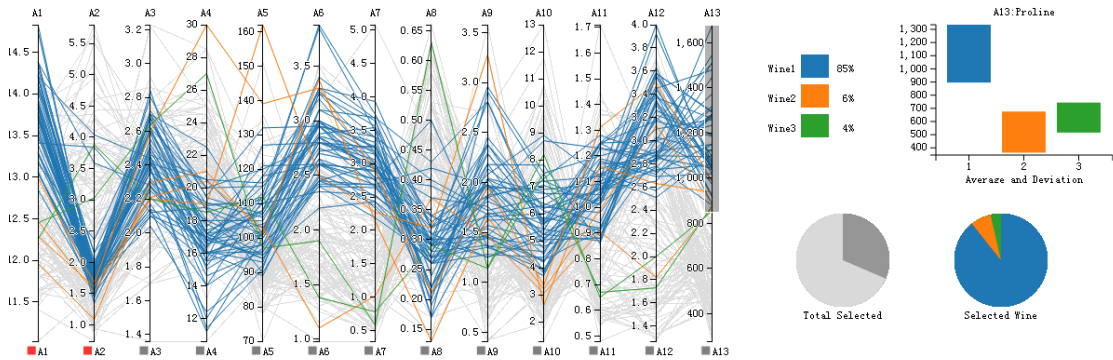


图 4:

我们通过 A13 维上选择，可以这个维度可以比较容易区分蓝色的这种酒，右侧的信息栏也验证了这一点，选中大部分都是蓝色的酒，同时 86% 的蓝色类型的酒也被选中，A13 的均值方差分布也说明了该维度能比较好区分蓝色的酒。当然数据也体现另两

种酒也会有少量数据混入。

另一个例子，如下图：

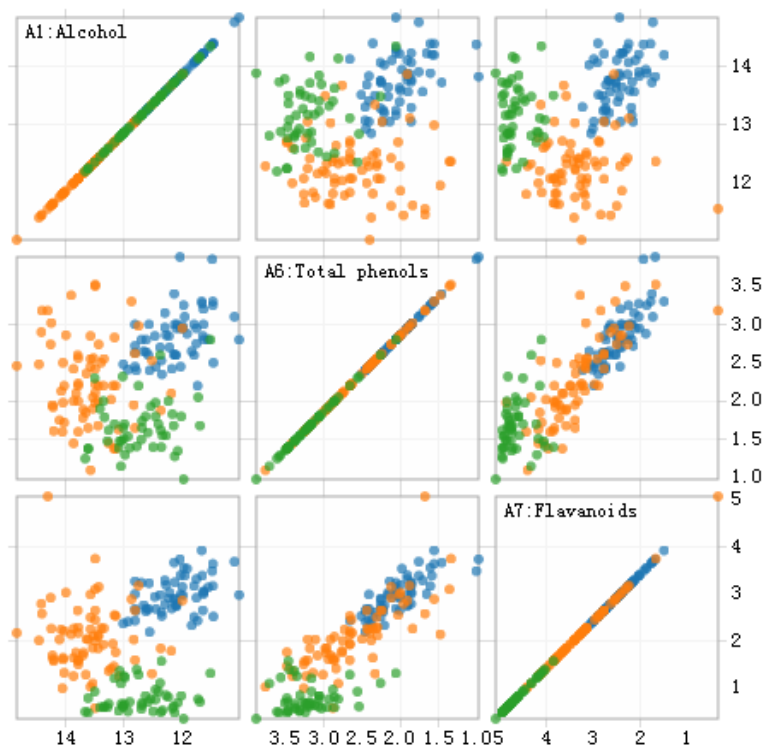


图 5:

我们能发现一些属性之间的关系，比如利用 A1 和 A6 的散点图比较分散，说明 A1 和 A6 可以较好地地区分三种酒；而 A6 和 A7 这两种属性应该也是比较相关的两种属性。

4 感想和总结

这次的可视化作业，在做的过程中接触了 d3 中更多的可视化的工具，做完之后学习到了不少东西，了解到了高维数据的一些可视化展示的方法。