

Titanic 数据集可视化报告

一、故事链接

第一个版本：https://public.tableau.com/profile/ivan.wu#!/vizhome/titanic_108/1_1

最终版本：https://public.tableau.com/profile/ivan.wu#!/vizhome/titanicfinal_1/sheet0

二、总结

本可视化主要针对泰坦尼克号沉没事件中两性的生还率开展研究，先从总体观察，接着分别从年龄段、船舱等级、上船城市和亲人数量等角度分别观察，可以看出：大于 10 岁的女性生还率比男性高得多；从一等舱到三等舱，生还率逐渐下降，但都是女性高于男性，即使三等舱的女性生还率也高于一等舱的男性。总体来看，在这次灾难中，人们把生还的机会更多地让给了女性和儿童。

三、设计

该可视化通过故事《女性生还率高于男性》表述：

(在所有元素中，蓝色表示女性，黄色表示男性；红色表示死亡，绿色表示生还)

故事点一：总体来看，女性的生还率比男性高很多。

上图：用饼图展示两性生存和死亡的人数；

下图：用柱状图展示两性的生还率对比。这里，生还率用维度 “Survived” 的平均值计算，下同。

故事点二：在不同的年龄段中，女性的生还率均高于男性。（70 岁以上没有女性乘客）10 岁以下的儿童中，男女性存活率差别最小。

上图：用饼图展示不同年龄段两性人数；

下图：用柱状图表示不同年龄段两性生还率。

在不同年龄段人数表中加入筛选器 “显示生还率”。

故事点三：随着船舱级别的降低（1到3），整体生还率下降。但在三个船舱中，女性的生还率都高于男性，即使三等舱的女性生还率也高于一等舱的男性。

上图：用热图表示不同船舱人数；

下图：用柱状图表示不同船舱两性生还率。

加入了对不同舱位和不同性别的悬停突出显示。

故事点四：始发站南安普顿上船的女性旅客似乎更喜欢结伴旅行，但生还率最低，进一步通过筛选年龄段发现，这部分人群只有在30到50岁之间生还率排名第二，排除有些年龄段样本太少造成的误差，在Cherbourg上船的女性的生还率最高，在Southampton上船的女性生还率最低。

上图：用柱状图表示不同登船口女性生还率；

下图：用柱状图表示不同登船口生还和遇难女性平均亲人数量。这里，亲人数量为计算字段，计算方法为：[SibSp] + [Parch]。同时该图表加入筛选器：年龄段，用以观察不同年龄段人群行为。

加入了不同登船口的悬停突出显示，加入了城市定义 URL。

四、反馈

[优达学城论坛关于我征求意见的帖子链接](#)

（一）来自 mentor ivylee 的反馈一：

@Ivan_Wu 同学你好！

- 最直观感受是：故事的最后一页，维度太多了，涉及到了年龄、性别、舱位、陪伴人数、生存率，一共5个变量。
此时，用颜色来表现的陪伴人数，不能与 y 轴的生存率有直观的对比效果。
因为前两页已经探索过年龄、性别和舱位，建议在第三页可以将重点放在新的探索目标：陪伴人数。
- 其次，生还率是一个比例，比例依托于数量才有意义，如果一个组内只有一个人，那么这一组的生还率不是1就是0，都是非常极端的值，受偶然因素的影响很大。所以建议补充关于人数的对比。提示：故事的一页可以是一个仪表板，一个仪表板中可以有多个图表，所以故事的每一页都可以有多个图表哦！

希望以上信息可以对你有帮助~🙏

将意见整理后如下：

1. 第三页重点放在亲人人数变量，减少图表中的变量数。

2. 在生还率旁增加人数的图表，以突出偶然因素的影响。
3. 可以在故事的每一页添加仪表盘。

（二）来自 mentor Elib 的反馈二：

将意见整理后如下：

第三个图做得很confusing，你有点结论：

- 10岁以下孩子的亲人数量最多：这个发现有什么意义？说明孩子都不是单独旅行的？
- 生还的女性比生还的男性亲人数量：从你的条形图很难得到这个结论，这不是可视化的最佳做法。即使你得到了这样的结论，只是因为生还女性人数多，所以他们亲属多，说明了什么？

你的整个可视化需要一个中心，你想解决的问题是什么？你的主要结论是什么？试着利用3-5个dash围绕着你提出的问题来回答。

1. 第三页关于 10 岁以下孩子亲人数表达的意义没有凸显出来。
2. 第三页关于生还的女性比生还的男性亲人数量多，这一点很难看出来。
3. 试着利用仪表盘组建故事。

（三）结合上述两个反馈，做出如下完善和更改：

1. 将年龄段和船舱相关的两个可视化改为仪表盘，均加入基础数量图表。
2. 第三页改为对不同登船城市女性生还率和亲人数量研究。
3. 加入总体情况仪表盘。

五、资源

N / A.