**.** 泰坦尼克号事故中生存状况影响因素的分析

本次链接：

<https://public.tableau.com/profile/carol.xu#!/vizhome/--_15597518397400/Story5?publish=yes>

第三次提交：

<https://public.tableau.com/profile/carol.xu#!/vizhome/--_15597106487210/Story5?publish=yes>

第二次提交：

<https://public.tableau.com/profile/carol.xu#!/vizhome/6786/Story5?publish=yes>

第一次提交：[https://public.tableau.com/profile/carol.xu#!/vizhome/5721/Story5?publish=yes](https://public.tableau.com/profile/carol.xu" \l "!/vizhome/5721/Story5?publish=yes)

* 总结部分：

本仪表盘展示了泰坦尼克号的生存情况分析，特别探索了泰坦尼克号生存状态与年龄、性别、家庭状况以及不同等级船舱之前的关系。

从获救总人数性别比例或者生存率上看，女性比例较高；但是从获救总人数的年龄分布来看，成年人比例较高，但是从生存率上来看，少儿的生存率较高。

从获救总人数性别比例或者生存率上看，女性比例较高；但是从获救总人数的年龄分布来看，成年人比例较高，但是从生存率上来看，少儿的生存率较高。

从分性别不同等级的船舱来看，一等舱的女性存活率最高；从分年龄不同等级船舱的角度，二等舱的少儿存活率最高。

* 设计部分：
  + 图表设计思路：

方形饼图：直观的观察比例与分布

柱状图

将数据进行比较、数据单纯性展示、排序数据展示。可以直观的看到不同性别生存数据差异性，

面积图：矩形树图

展现同一层级不同分类占比情况，以及同一分类下子集占比情况。

仪表器1:

* + 图表类型：方形饼图、柱状图
  + 图表的布局：布局左右结构，
  + 图表的图例：不同颜色区别性别、年龄

仪表器2:

* + 图表类型：柱状图
  + 图表的布局：上下结构
  + 图表的图例：不同颜色区别性别；用交互式筛选器进行互动，让读者清晰看到不同年龄阶段不同家庭成员状况的乘客的存活率。

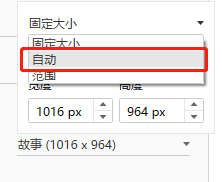
仪表器3:

* + 图表类型：面积图、方形饼图、柱状图
  + 图表的布局：布局左右结构和上下结构，
  + 图表的图例：不同颜色区别不同年龄段以及不同等级的船舱。

反馈部分：  
第一次提交：

* 他人反馈：完成了一个完整故事，但整体提交的项目从形式上看仍然不符合要求。这次审阅仅针对Story5故事，其他故事和仪表板不作审阅。请在下次提交时，**确保界面中只呈现了一个故事**，做到结构的简洁并避免歧义。其他包含在故事中的部分，请对其隐藏处理，使得工作簿更加简洁。隐藏表可以使用下图的Hide all sheets实现：

修改：用hide all sheets实现了隐藏处理，使界面更简洁。

* 他人反馈：你的部分仪表板以及故事的尺寸较小，导致可视化内容比较拥挤，请在**故事和各仪表板**（二者均需要选择）的大小中选择“自动”以解决此问题（对于故事及所有故事点中的仪表板，都应当选择此选项）：  
  [](https://udacity-reviews-uploads.s3.us-west-2.amazonaws.com/_attachments/57175/1558883731/3.png)

选择“自动”还有一个重要理由，即每个人所使用的设备分辨率不尽相同，如果使用固定尺寸，会导致可视化在某些设备上无法合理展示（例如，在1080p分辨率的显示屏中尺寸正常的可视化，在4k显示屏中会变得十分狭小并难以看清）。因此请对此部分进行适当调整。

修改：选择“自动”

* 在故事点中仅使用一副工作表（尤其是图表内容本身较少的工作表）并不是合适的选择。这样做会导致许多有关联的图表被拆分到多个故事点，不利于观众对各图表进行综合的比较及得出结论；过多的故事点也会导致结构过于复杂。请尽量保证**每个故事点都由仪表板构成，仪表板中至少有两幅在逻辑上存在关联的工作表**，并添加文字分析，以充实可视化的内容，使得故事的结构更加紧凑合理。但需要注意，将所有可视化放在一个仪表板中仍然不是合适的选择，故事必须具有多个故事点以体现结构和层次性；

修改：使内容进行链接，对生存率分析由浅入深，先分析年龄和性别，在在此基础上加入家庭状况和船舱等级的分析

* 针对故事，使用的可视化图表和类型数量过少。过少的可视化无法深入地剖析、挖掘问题并得出深刻结论，会使得故事浮于表面，缺乏重点和亮点。请重新梳理故事的结构，为故事补充更多可视化。作为参考，能够完整叙述一个故事的项目中，一般至少会包含至少3个仪表板，8幅图表（对于摩拜单车、Prosperloan等复杂数据集而言，工作表一般在10个以上）。针对你的项目，首先建议探索更多变量，其次请关注各变量分组的人数比例和分布，最后请探索更多**多变量分析**的呈现方式（例如，各舱位分性别的生还率）。

进行了多变量分析：各舱位分性别存活率分析、各舱位分年龄段存活率分析、各年龄段分性别存活率分析；各年龄段分家庭状况生还率分析

* 需要注意，仅仅展示生还者和遇难者的人数（或人数的比较）是不足以清晰说明生还状况的，反而容易对观众造成误导（某些群体生还人数多仅仅是因为这一群体乘客总数多，实际上却伤亡惨重，**实际上你的可视化中，正因为没有正确展示生还率，只展示了生还者数据，导致结论几乎全部是错误的**）。生还状况最重要的指标是**生还率**，即针对一个特定群体中，存活者的比例。因此，人数是一个可供展示的方面，但依然需要展示生还率这一核心指标。请补充或修改相关可视化；

修改：存活数量（分性别、分年龄段、分船舱等级）/调查总数（890）

* 可视化使用的变量不足。在数据集简单并存在明确分析目的情况下，你应当尽可能地探索更多变量。以本数据集而言，你至少缺失了对“舱位”这一重要变量的分析；

修改：探索了舱位“变量”

* 针对年龄，建议使用更合理的分组方式而非使用相同组距。年龄的分组应当根据自然属性拟定，例如小孩、青少年、老人等。这样更能体现出针对不同群体生还状况的不同；即使使用固定组距，目前的组距较小，分组较多，导致可视化效果也不够好，建议改善；

修改：根据自然属性，分为少儿、成年和老年<https://www.1xuezhe.exuezhe.com/Qk/art/646930?dbcode=1&flag=2>

* parch和sibsp的处理存在问题。可以看到其值为0、1、2的占了绝大多数，3以上的乘客非常之少，导致得出的结论缺乏意义，不具备代表性。针对此情况，建议将非常少的组合并处理（例如分组为0、1、2及2以上三组），使得整体的数量比例合适，同时可视化也能更加简洁。

修改：分组为0、1、2及2以上三组

* 你的图表中没有能够增加读者对数据理解的交互功能（需要注意，使用简单的Hightlight高亮功能并不能完全满足此项要求）。

你可以通过对**必要的图表**增加[交互式筛选器](http://onlinehelp.tableau.com/current/pro/desktop/zh-cn/filtering.html" \l "SetOptions" \t "_blank)的方式达到此目的，也可以尝试使用可以筛选数据的[仪表板动作](https://www.tableau.com/zh-cn/learn/tutorials/on-demand/dashboard-interactivity-using-actions" \t "_blank)。

修改：将交互式筛选器应用到探索家庭状况的变量上

修改意见： 请注意，你在此部分还存在一些问题。请确保你的每幅图表（请注意是图表，而非仪表板）都有：

* 合适的标题（以规范的中文/英文表述概括可视化所展示的信息）；
* 坐标轴（显示数值）；
* 经过清理的坐标轴标签（显示为清晰易懂的中文或英文名称；如果标题已经明显提示了变量，则可以不加）。

你有部分图表至少缺失了这三者之一（部分图表缺乏合适标题），请仔细检查并修改。有关标题的部分，请特别注意，需要添加的是**各故事点中每幅图表的标题（也就是工作表的标题），而不是仪表板标题或者故事标题**。每幅图表都需要有清晰的标题并显示在故事中。

修改：给worksheet加标题；坐标轴显示数值；提示了存活率，没有添加标签（不知为何改不成y轴标签）

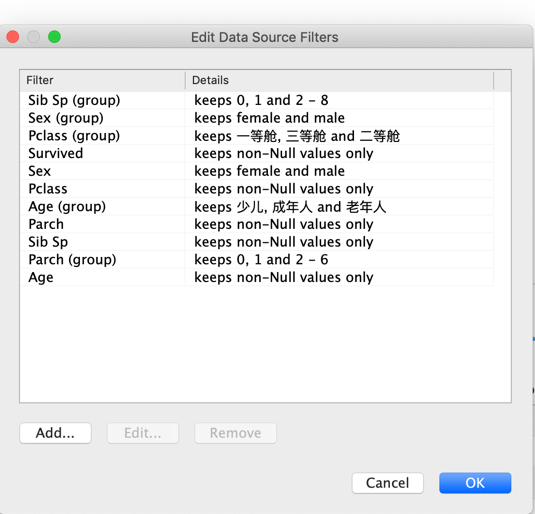
第二次审阅：

* 他人反馈：请注意上次审阅中强调的“故事和各仪表板（二者均需要选择）”，这次提交中你有部分仪表板并未调整尺寸，请检查并确保故事和各仪表板的尺寸都调整为“自动”；

修改：将所有“工作单”尺寸改为“自动”

* 生还率的计算逻辑全部错误，导致结论也几乎是全部错误的。目前的计算逻辑并不合理。这里不再一一指出，请注意作图完成后请使用Python或Excel等工具计算个别组的生还率以验证Tableau的生还率计算是否正确。

提示：对于生还率，可在去除空值后直接使用数据集的Survived变量（当作度量使用）计算，**由于其值为0或1**（请思考在变量的值仅有0和1两种时，如何使用简单的描述性统计方法计算0或1所占的比例），因此无需建立计算字段或表计算即可在图表中正确展示任何组别的生还率，实现逻辑其实非常简单。再次强调：在图表完成后请**务必使用其他工具进行验算**以确保计算出的是正确的生还率。请根据提示修改计算逻辑，并根据得到的可视化重新叙述结论；

修改：在tableau筛除缺失值，生还率按照（sum（[survived]）/count（[survived]））进行计算，结果与excel计算相符合。

* 请将所有正方形的比例图修改为扇形图，以符合一般的图表阅读习惯，如下图：

修订：修改为扇形

* 故事点1，在分析具体变量时，对于分析变量的值为NULL的数据，请作筛除处理，而不需要展示，以免影响可视化效果；

修订：将null 筛除

* 故事点2，图表和筛选器逻辑均比较混乱。例如上方图表，parch只有0、1、2三个值，2以上的值呢？未进行合理的分组和解释；

修订：可以看到其值为0、1、2的占了绝大多数，3以上的乘客非常之少，导致得出的结论缺乏意义，不具备代表性。针对此情况，分为 0、1、2及2以上三组，使得整体的数量比例合适，同时可视化也能更加简洁。

* 下方图表同样让人困惑，为何只有成年人，没有其他年龄段人群？针对筛选器，未在故事中作任何具体介绍，使用方法也较为复杂，其内容非常不清晰。例如Age（bin），筛选的意义是什么？既然你已经根据自然属性分为3组，为何还要增加age（bin）筛选器，这对观众显然会造成误导（例如观众只筛选了年龄30-35的人群，图表显示的却是“所有成年人”的数据，意思完全是矛盾的），更何况age（bin）还同时使用了pages功能，第一次看到这一故事点的观众只会感觉到一头雾水。请注意筛选器设置必须简洁易懂，筛选出的内容符合观众的认知直觉。如果要设置较复杂的交互操作，请确保操作有意义，能从中得出有意义的结论，并以清晰的文字叙述操作方法，否则请不要使用筛选器，直接呈现你希望观众看到的图表和结论；

修改：将筛选改为age（group）、sib sp（group）、parch（group）

* 另外，从结论来看，多变量分析和大量筛选器得出的结论仅仅是“从家庭状况上来看，没有亲人的单独坐船的成年人生存率较高”，是否过于简单？请综合说明各组的情况，并得出更有意义的结论。

修改：完善结论，加入较为有意义的结论，见报告开头陈述。

* 另外，分组时请不要依赖tableau的直方图分组（bin）功能，它们只能分出固定组距的多组，并不灵活。请复习课程中有关“计算字段”的部分，合理使用计算字段中的[条件语句](https://onlinehelp.tableau.com/current/pro/desktop/zh-cn/functions_functions_logical.htm)灵活分组。

修改：改为使用（group）功能分组

* 故事点3下方舱位条形图，请调整各组的顺序为“1-2-3等舱”或“3-2-1等舱”，以体现舱位次序变化，使得可视化效果更好。

修改：用“sort”进行顺序调整

第三次审阅：

* 针对生还率部分，请避免使用堆叠式图表，因为条柱的总长度没有意义。堆叠式条形图只有在堆叠后的结果有意义时才能使用（男女生还率之和显然没有太大意义，不代表总体生还率）。请改为折线图或圆点形式的图表，避免堆叠展示；

修改：修改堆叠图，为条形图

* 故事点3右上方图表，除了图例外，建议在图表中也增加生还率具体数值的文字标签，使得展示内容更加直观具体。

修改：图表中也增加生还率具体数值的文字标签

* 未能修改上次提及的相关部分，这次提交的可视化中仍有“SUM([Survived])/COUNT([Survived])”“SUM([Survived])/890”等不合理坐标轴标签。请参考上次审阅意见修改，并将其更新到下次提交的反馈部分。

修改：“SUM([Survived])/COUNT([Survived])”改为生存率

* 资源部分：  
  百度百科：折线图、条形图

[创建故事](https://onlinehelp.tableau.com/current/pro/desktop/zh-cn/story_create.htm)

根据自然属性，分为少儿、成年和老年<https://www.1xuezhe.exuezhe.com/Qk/art/646930?dbcode=1&flag=2>