

Titanic

Data

The data contain 891 observations and 12 variables. I want to see if any factor is the main reason that will affect the survival rate.

This dataset is from Kaggle. Here is my results' tableau link

- [Titanic 0.0](#)
- [Titanic 1.0](#)

Conclusion

From this dataset, we could find that female have higher survive chance than male. From the plot, children seems most likely survive, then female also have a better chance to survive. Male however, may sacrifice their chance of live to make sure children and female survive, so they tend to have a lower survival rate. Elder have highest death rate, which may caused by inconvenient movement.

Visualization Design

First, I observe there are 342 people survived, I'd like to see if women and men have same rate to survive.

Sheet1

First, it's important to know how may male and female are there on the ship. I use histogram, which shows there're 314 female and 577 male on the ship. Next, I'll use histogram to see relation between survive number and sex.

sheet2

Using histogram, we see 109 male and 233 female survived.

Sheet3

From two plots before, it's time to use a pie chart to see survival rate passengers of different genders. It's obvious that female passengers have a higher survival rate than male passengers.

Sheet3-fix

I fix the plot according to the reflection, choose to use bar chart to represent this relation, get a more clear effect.

Sheet4

In this part, I dig into survived passengers, wonder which class did they come from. So, I draw a pie chart among survived passengers according to different class. There're three class on the ship, it seems female survivors are equally distributed among them, and male passengers come from class 2 have lowest rate among male survivors.

Sheet5

This is a surviving rate plot according to class. In three class, class 1 seems have highest survival rate, then come class 2, class 3, as expected, has lowest survival rate. I think it may because that male bought more class 3 tickets than female.

Sheet5-fix

I fix the plot according to the reflection, choose to use bar chart to represent this relation, get a more clear effect.

Sheet6

Base on assumption above, I draw the sheet 6 plot. From I can see that half passengers in class 1 is female, and in class 2 and class 3, there're more male passengers than female, which seems reasonable, female tend to travel with family, male may travel along.

Sheet7

This plot reflect the relation between survival rate, sex and class. I think this plot perfectly reflects the proportional relationship with area.

Sheet8

Next, I wonder if baby, youth or elder have higher survival rate than adult. I plot the histogram of passengers age first.

Sheet9

In this plot I plot the histogram of survived passengers' age, compare to last plot, we can see there's minor change with children under age 9 than youth and adult between 15 to 41.

Sheet10

Based on above findings, I divide passengers into four groups:

- children: 0 - 14
- Youth: 14 - 20
- Adult: 20 - 59
- Elderly: above 59

I plot the histogram under this group. I want to use pie chart to make it easier to see the survival rate in different age.

Sheet11

Use pie chart we can clearly see children have highest survival rate, youth and adult have similar chance to survive, elderly, however, have lowest survival rate.

Sheet11-fix

I fix the plot according to the reflection, choose to use bar chart to represent this relation, get a more clear effect.

Sheet12

In last chart, I divided the age into four part as described before. From it we can clearly see that female have highest survival among each age, boys tend to have more chance to survive than male adults.

Sheet12-fix

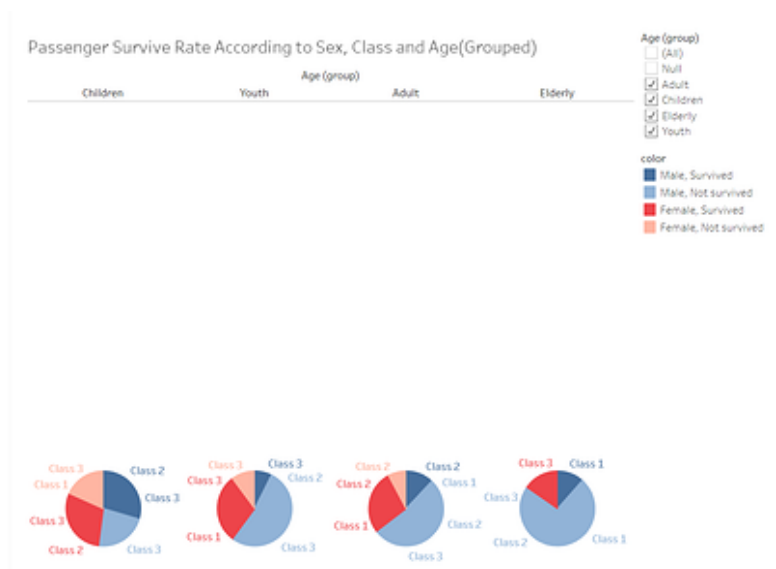
Use a multiple-plots to represent four variables make it better to understand the data.

Reflection

个人感觉对比不同组别内的生存/死亡比例，用饼图不如用百分比堆叠柱状图的效果好

- [参考文档：创建总计为 100% 的堆叠条形图](#)

另外，最后一页的图表不知道是没有显示全还是什么原因，不太理解最后一个图表的内容：



Reference

[Data source](#)

[参考文档：创建总计为 100% 的堆叠条形图](#)