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

Conclusion

From the plot, children seems most likely survive, 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. From class reason, if you buy class 1 ticket rather than class 3 ticket, then you'll probably survive from the disaster, so, try to buy the best class ticket you can.

Finally, if you are a children or female with class 1 ticket, you have highest chance to survive. If you are a male with class 3 ticket, you have to bet on your luck.

Visualization Design

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

PS: All survival rate is calculated in each group, e.g: class 1 survival rate is $\frac{number \; of \; class \; 1 \; survivor}{number \; of \; class \; 1\; passenger}$, not in all survivor, which seems a more reasonable choice for me.

Sex

- Type: Histogram
- Reason: There're two distinct variables, use histogram can clearly see the difference between them.
- Layout: It's a basic analyze, so take a smaller place in dashboard.
- Legend: No legend.

Sex-Survived

- Type: Histogram
- Reason: There're two distinct variables, use histogram can clearly see the difference between them.
- Layout: It's a basic analyze, so take a smaller place in dashboard.
- Legend: No legend.

Survived(Sex)

- Type: Pie chart
- Reason: There're two different classes, male and female. We also want to see the survival rate in each class, so I chose pie chart.

- Layout: It gives reader a total intuitive to survival rate, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Survived(Sex)-fix

- Type: Stacked Histogram
- Reason: There're two distinct variables, and two states, use stacked histogram can clearly see the difference between them, much clear than pie chart.
- Layout: It gives reader a total intuitive to survival rate, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Survived(Sex)-Survival

- Type: Histogram
- Reason: There're two distinct variables, and two states, which are complementary to each other. Use stacked histogram seems a little redundant.
- Layout: It gives reader a total intuitive to survival rate, so take a bigger place in dashboard.
- Legend: No legend.

Class-Survived(Sex)

- Type: Pie chart
- Reason: There're three distinct classes, which need a pie chart to show the proportion.
- Layout: It's a inspire to following analyze, so take less space in the dashboard.
- Legend: Brown - Class 3; Orange - Class 2; Beige - Class 1.

Class-Sex

- Type: Stacked Histogram
- Reason: There're three different classes, and two states (male and female). Use a stacked histogram can clearly represent difference of gender in different class.
- Layout: It gives reader a intuitive to the relation between class and gender, it can be a reference. So take less space in dashboard.
- Legend: Blue - Male; Orange - Female.

Class-Survived

- Type: Pie chart
- Reason: There're three different classes. We also want to see the survival rate in each class, so I chose pie chart.
- Layout: It gives reader a total intuitive to survival rate in different class, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Class-Survived-fix

- Type: Stacked Histogram

- Reason: There're three distinct variables, and two states, use stacked histogram can clearly see the difference between them, much clear than pie chart.
- Layout: It gives reader a total intuitive to survival rate in different class, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Class-Survived-fix-Survival

- Type: Histogram
- Reason: There're three distinct variables, and two states, which are complementary to each other. Use stacked histogram seems a little redundant.
- Layout: It gives reader a total intuitive to survival rate in different class, so take a bigger place in dashboard.
- Legend: No legend.

Class-Survived-Sex

- Type: Tree plot
- Reason: I want to represent the relationship between class, gender and survived or not, so I chose tree plot, use it's area to represent the number of people in each kind.
- Layout: It gives reader a total intuitive to survival rate relate with class and gender, so I make it bigger than other two plots.
- Legend: Deep blue - Male, Survived; Blue - Male, Not survived; Crimson - Female, Survived; Red - Female, Not survived

---

Class-Sex-Survival

- Type: Histogram
- Reason: Tree plot seems not so straight as histogram, which can make people be aware of survival rate in each class.
- Layout: It gives reader a total intuitive to survival rate relate with class and gender, so I make it bigger than other two plots.
- Legend: No legend.

Age

- Type: Histogram
- Reason: Histogram can perfectly represent statistic information you want to know, and express the distribution of variable clearly, help understand the whole attribute of the variable.
- Layout: It gives reader a total intuitive to survival rate relate with age, but I didn't use this dashboard in my story.
- Legend: No legend.

Age-Survived

- Type: Histogram
- Reason: This plot is the histogram of survivors' age, it can perfectly express the distribution of variable clearly, help understand the whole attribute of the variable.

- Layout: It gives reader a total intuitive to survival rate relate with age, but I didn't use this dashboard in my story.
- Legend: No legend.

Age(Group)-Survived

- Type: Histogram

- Reason: I split the age of passengers into four groups:
  - children: 0 - 14
  - Youth: 14 - 20
  - Adult: 20 - 59
  - Elderly: above 59

  with histogram, it can represent the number of each group of passengers clearly.

- Layout: It gives reader a total intuitive to survival rate and age, so I put it in the top half of the dashboard.

- Legend: No legend.

Survived(Age(group))

- Type: Pie chart
- Reason: There're four different groups. We also want to see the survival rate in each class, so I chose pie chart.
- Layout: It gives reader a total intuitive to survival rate in different age groups, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Survived(Age(group))-fix

- Type: Stacked Histogram
- Reason: There're four distinct groups, and two states, use stacked histogram can clearly see the difference between them, much clear than pie chart.
- Layout: It gives reader a total intuitive to survival rate, so take a bigger place in dashboard.
- Legend: Orange - Survived; Blue - Not Survived

Survived(Age(group))-fix-Survival

- Type: Histogram
- Reason: There're four distinct groups, and two states, which are complementary to each other. Use stacked histogram seems a little redundant.
- Layout: It gives reader a total intuitive to survival rate in different class, so take a bigger place in dashboard.
- Legend: No legend.

Class-Survived-Sex-Age

- Type: Pie chart

- Reason: There're four different age groups. We also want to see the survival rate with class and gender, so I split it into four groups, use different colors to represent gender and survival situation, use notation to mark the class.
- Layout: It gives reader a total intuitive to survival rate with different variables: gender, age and class, so it takes the whole dashboard.
- Legend: Deep blue - Male, Survived; Blue - Male, Not survived; Crimson - Female, Survived; Red - Female, Not survived

---

Class-Survived-Sex-Age-fix

- Type: Pie chart (multi small plots)
- Reason: There're four different age groups and three other variables, use pie chart can be confused in some extent, so use a multi small plots can express the survival rate of different kind of passengers better, also make it more clear to compare between each other, extract class notation and make it another dimension help keep plot clear.
- Layout: It gives reader a total intuitive to survival rate with different variables: gender, age and class, so it takes the whole dashboard.
- Legend: Deep blue - Male, Survived; Blue - Male, Not survived; Crimson - Female, Survived; Red - Female, Not survived

---

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：
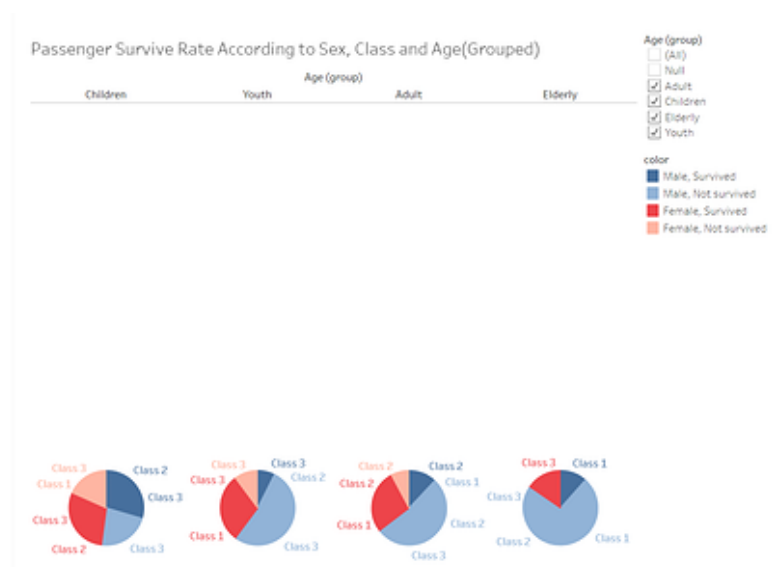
Reference

Data source

参考文档：创建总计为 100% 的堆叠条形图