

# **Create a Tableau Story - What Sorts of People Were Likely to Survive in The Sinking of The RMS Titanic?**

*Zhenwen Xu*

*2018/11/27*

## **Summary:**

The primary purpose of this Tableau story is to:

1. Display the relationship between t of survived and a single variable which seems to be more relative with survived including age, sex, and fare;
2. Display the ratio of survived impacted by all five variables together, and figure out all kinds of passengers whose ratio of survived is over 80%;

The summary of this story comes from two different view:

1. When just considering the relationship between the single variable and the ratio of passenger, people who are female, or child, or have first-class ticket, or with siblings or spouses, or with parents or children were likely to survive in the sinking of the RMS Titanic;
2. When considering all variables together, passengers who are female, live in the first class were likely to survived. And it shows that Sex is the most critical variable which impacts the survived ratio mostly, the following are Pclass, Age Class, and SibSp+Parch.

## **Design:**

There are four different plot types which are used to visualize the relationship between survived and other variables:

1. Histogram is used to exhibit the distribution of fare in ticket with some basic information about each passenger. The reason is that fare is a continuous numerical variable, so it is better to use histogram rather than bar plot to show the distribution of fare. And the other goal of this slide is to list all passengers' basic information, especially including the survived condition of

each person to prove that there is no direct relationship between fare and survived condition because of no regularity;

2. I divided all age into five subgroups which are “Null”, “Child”, “Youth”, “Middle Age”, and “Old Age” by the law of the United Nations because none cares about passengers’ survived condition in a specific age, and grouping all age into several age classes to display the survived condition will be more compact;
3. I also noticed that most of value in the SibSp and ParCh are 0 and 1, so I think it is better to combine values as 1 and values over 1 into one condition, which calls “1 and more than 1” so that can simplify the further plots which will use these two variables;
4. Bar chart is mainly used to demonstrate the comparison between the number of died and survived passengers by two different variables, which are cabin class, age class because these variables are all dimensions, but the number of passengers is measure, and using bar plot can exhibit the number of passengers in each class of one variable distinctly.
5. Tree maps is used to show the ratio of survived in each variable, and it displays the percentage of each value in total variable separately. The reason I choose this kind of plot is that one of my audience says that tree maps is better than pie plot when it needs to show the percentage of whole, and pie plot sometimes will make confusing visions but tree maps will not.
6. Box plot is used to show the distribution of ratio of survived impacted by the four relevant variables together, and from this plot, I stipulate all kinds of passengers whose ratio of survived is over 80% are likely to survive. So, from this box plot, it is clear show all satisfied category of passengers.

Link to version before revision:

[https://public.tableau.com/profile/zhenwen.xu#!/vizhome/Tatanic\\_2/Story1](https://public.tableau.com/profile/zhenwen.xu#!/vizhome/Tatanic_2/Story1)

Link to version after revision:

[https://public.tableau.com/profile/zhenwen.xu#!/vizhome/Tatanic\\_5/Story1](https://public.tableau.com/profile/zhenwen.xu#!/vizhome/Tatanic_5/Story1)

## Feedback:

Summary of feedback received:

1. *What do you think will make the visualization better?*
  - It may be better to add some plots between the non-survived and other variables. (I create “Passenger Survived” variable and use it in bar chars to solve this problem.)
  - Show the sum of people left in the chart when apply the filters. (I don’t deal with this problem because the sum of passengers is not the necessary content which I have to analysis.)
  - Make a brief introduction in the beginning. (I use text to add some introduction in the start of the story.)
  - The order of age class in bar chart need to reorder from young to old. (I solve this problem by using manual sort.)
  - The structure of Dashboard is confusing, and it is better to add summary into the description of slide. (I transfer some text from the dashboard to the description of slide.)
  - The black background plot is confusing, it is better to change the type of plot to show the distribution. (I have dropped this plot because this plot is not clear enough, and it is hard to summary.)
  - The content of Dashboard 2 and 3 need to restructure. (I have restructured them.)

## Resources:

The dataset of this project is from <https://www.kaggle.com/c/titanic>;

The following are references which I have used:

1. [https://onlinehelp.tableau.com/current/pro/desktop/en-us/publish\\_workbooks\\_tableaupublic.htm](https://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.htm)
2. [https://github.com/lokhengchau/Data\\_Analyst\\_Nanodegree\\_Udacity/blob/master/Project%208/Tableau%20report.pdf](https://github.com/lokhengchau/Data_Analyst_Nanodegree_Udacity/blob/master/Project%208/Tableau%20report.pdf)
3. <https://www.tableau.com/learn/tutorials/on-demand/dashboard->

[objects?product=all&version=tableau\\_desktop\\_2018\\_3&topic=dashboardsstories](https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories?product=all&version=tableau_desktop_2018_3&topic=dashboardsstories)

4. [https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories?product=all&version=tableau\\_desktop\\_2018\\_3&topic=dashboardsstories](https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories?product=all&version=tableau_desktop_2018_3&topic=dashboardsstories)