

The Tip of The Iceberg: A Gender and Socioeconomic Visualization of the Titanic Disaster

By

Samantha Dimmer, Vinisha Mittal, and Sebastian Roubert

The RMS *Titanic* operates as a symbol of tragedy, luxury, and mankind's futility against nature. Most of the public's sole exposure to the *Titanic* was in the 1997 film starring Leonardo DiCaprio. We set out to draw back the curtains of public perception and investigate survival through socioeconomics and gender.. We wanted to see who survived and whether these statistics followed the perceived narrative that male, low-paying, or non-children passengers where the predominant victims.

We found a dataset from the website of a Vanderbilt statistics class (biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic3.xls) containing a list of every passenger and variables associated with them. For every passenger, we extracted their **class number**, whether they **survived**, their **age**, their **ticket fare**, and their **gender**. When deciding on what variables to extract, we thought both about the story we wanted to tell as well as how much the data had to offer. For example, we could have selected passenger origin as a data point, but noticed that many passengers did not have this data listed. In terms of our story, we really wanted to focus on the commonly held narrative of who survived the titanic and thus sought out variables that related to it (socioeconomics, gender, and age). Creating the histogram required reformatting the data into arrays based on a) whether a passenger survived and b) that passenger's gender. To achieve this, we simply looped through all of the data checking for each data point's survival/gender, placing it in the appropriate array. This was necessary as D3 doesn't initially allow for a dual histogram, much less a two-sided dual histogram; we needed to create four separate sets of bars and translate them. This reformatting also facilitated the creation of our donut chart.

For the histogram, we used linear scales for the x and y-axes, as well as the d3 histogram function to map the data into bins. Because the histogram is two-sided, we needed to create two y-axes--for the lower one, we inverted the values of the upper axis and translated it downwards. For the donut chart, we also used a linear scale along with the histogram function. We then mapped the age bins to arc and path elements. We added the colors to the chart using an ordinal scale along with a color array generated using ColorBrewer. We used path elements to make the boat and iceberg--elements added namely for story-telling, but also for some data display (for example, the smokestacks on the boat are bar charts). We used "Easy Imagemap Generator" (<http://imagemap-generator.dariodomi.de/>) to generate the coordinates, then translated the elements to their correct placements. The boat was based off of this image: <http://www.clipartbay.com/cliparts/titanic-iceberg-clip-art-kipebtt.gif>, and the iceberg was based off of this image:

https://thumb9.shutterstock.com/display_pic_with_logo/450076/546696442/stock-vector-iceberg-vector-icon-illustration-on-blue-background-iceberg-vector-poster-iceberg-vector-clip-art-546696

[442.jpg](#). We did notice the histogram has a few extreme outliers, on the higher end of ticket price, but we decided it was important to leave in this data to ensure we accurately depict the range of the socioeconomic distribution of the Titanic passengers. For the smokestack bar graph, we used D3's line functionality to generate paths for rectangles based on the survival rates per class, which were calculated by looping through the original data. We chose a distance of 80 pixels to represent 100%, and calculated the heights for each bar according to a linear scale.

The histogram tells a story that is fairly similar to what we commonly hear. There is a very clear shift between the number of female casualties in each fare bin and the number of male casualties, with male bars sitting much further below the origin. There is a very sharp peak showing that the majority of deaths were low-fare paying men. Though this is not necessarily surprising, it becomes interesting--and we have a deviation from the common Titanic narrative--when we look at the age distribution of this group. The fact that boys aged 0-9 and 10-19 represented a large majority of the deaths is jarring considering that public perception is that in such a disaster it is the women *and* children that are saved above all else. In this way, our visualization provides novel information while confirming expectations. In looking further into the socioeconomics of the incident, it's interesting to note an almost linear decline in percentage survival by class--confirming the commonly held perception that those who paid high fares were more likely to live. All in all, the visualization shows an interesting contrast between those who lived and those who died based on age/gender/socioeconomics, revealing both the truths and misconceptions we hold about the outcome of this well-known disaster.