Data Visualization: Titanic Data

Jenn Sandoval, Udacity Data Analyst Nanodegree Project 6

Summary:

The dataset utilized for this project is a subset of the 2,224 passengers that contains demographics and passenger information for the passengers and crew that were on board the Titanic. The purpose of this investigation is to create a visualization that reflects the demographics or passenger information between surviving passengers and those who died.

Design:

My overall goal for this project is to determine if there are certain demographics or passenger information that would have increased the likelihood of a passenger surviving or not. My first question was to determine if females or males had a higher survival rate. I started reviewing the dataset within R and quickly found that the best starting point for the charts would be to start with histograms. The histograms would allow for me to bin the data into groups so that I could make generalizations about the data initially. I found that there were some variables during my initial analysis (such as the Ticket Number or Name) wouldn't be taken into consideration as that data is too specific to allow for me to create a generalized graph to interpret. My initial findings of the first histogram I created was that 233 females survived versus 109 males. I wanted to dive a bit deeper into the histogram to determine if a certain age group had a higher survival rate versus other age groups and then also determine if a certain class of ticket that was purchased by either a female or male.

Initial Design:

https://public.tableau.com/profile/jenn.sandoval#!/vizhome/TitanicData_InitialDesign/Dashboard1?publish=yes

Feedback:

Person #1:

The histograms provide some information, but it's hard to tell what each chart means without a title. I think adding chart titles to each chart would be helpful in understanding each of the charts. If there was also more interactivity to where you could filter by either male or female and all of the charts would then update to reflect only data for the one gender that is being filtered.

Person #2:

First, the colors are hard to differentiate between either gender since they are both colored blue so if there could be a different color assigned to males and females it would be helpful. Titles of each of the charts would also be helpful to add context to what the charts are displaying. Overall, the bins are nice to see the separation of the data, however a couple of tweaks to the colors and titles would help me better understand the charts.

Person #3:

I really like the look of histograms, however on the second chart, it is including the NULL values, which don't really need to be included if the charts are reflecting information for only surviving passengers. Colors to separate the males from the females would also be really helpful to visually separate the two genders. It would also be helpful to include an explanation as to what each chart is representing, such as a quick summary to help understand what the charts are conveying.

Post-Feedback Design:

Based on the feedback that I received from the three people I interviewed, below are the changes that I made to the charts:

- Added Chart Titles with two sub titles for the PClass and Age bins to help the reader understand what the charts are describing
- Added colors to separate between females and males
- I added a filter for the gender that would filter for all three charts at once within the Dashboard
- Excluded the NULL values for the Surviving Passengers by Age bins
- Removed Gridlines from the charts for a cleaner look
- Updated the labelling of the charts to hide the a-axis label

Final Design:

https://public.tableau.com/profile/jenn.sandoval#!/vizhome/TitanicData_FinalDesign/Dashboard1?publish=yes

Resources:

https://www.udacity.com/api/nodes/5420148578/supplemental_media/titanic-datacsv/download

https://www.kaggle.com/c/titanic/data

http://onlinehelp.tableau.com/current/pro/desktop/en-us/filtering_global.html