Data Visualization with Tableau

Titanic Dataset

Brian J Hartman

Final Version

https://public.tableau.com/profile/brian.j.hartman#!/vizhome/Titanic_510/Story1
(https://public.tableau.com/profile/brian.j.hartman#!/vizhome/Titanic_510/Story1)

Second Version

https://public.tableau.com/profile/brian.j.hartman#!/vizhome/TitanicDatasetVisualization2/Story1
(https://public.tableau.com/profile/brian.j.hartman#!/vizhome/TitanicDatasetVisualization2/Story1)

First Version

https://public.tableau.com/profile/brian.j.hartman#!/vizhome/TitanicDatasetVisualization1/Story1
(https://public.tableau.com/profile/brian.j.hartman#!/vizhome/TitanicDatasetVisualization1/Story1)

Summary:

This data visualization project analyzes the Titanic Dataset found in the Udacity Dataset Options
(https://www.google.com/url?q=https://d17h27t6h515a5.cloudfront.net/topher/2017/October/59d54e6d_titanic-
data/titanic-data.csv&sa=D&ust=1520809925538000&usg=AFQjCNG2ZP_HXyCQPt8aNWFHzyu_CbVmFw) and
looks for the likely effects of multiple factors on survival rate among the passengers. Looking at the age group,
gender, and passenger class of the passengers should give us an idea of how survival rate was impacted.

Design:

I chose to utilize the Tableau 20 color palet for color selection and primarily used red and blue to distinguish between male and female passeengers for consistency. A limited amount of data cleaning was also applied to the original dataset. Frist I expanded upon the abbreviation for the embarkment ports from just the first letter to the full name of the city and abbreviated county to give the viewer a clearer picture of where people boraded the Titanic. Secondly, I created bins for the ages in ten year increments; again in order to give the viewer a clearer understanding of the age data. And finally I grouped the survival data together by the survivors and those that perished. This helped to better perform analysis on the data.

The data visualizations are intended to show the relationship between the likelyhood of a passenger surviving or not based on class and gender. The final scatterplot chart I added I think does the best job of telling this story.

Given that the focus of this is a comparison of different data types, the chosen type of chart for this visualization was the bar chart which helps effectively communicate the comparisons between the categories of data. Using the bar charts for this data makes it easy for the viewer to interpret the data.

After sending my initial story to three co-workers, their feedback lead me to perform some further design changes to help clear up some of the details. Their feedback led to removing and/or changing axis titles and redefining some graphs for easier understanding. I also had to resize the heigth of the caption boxes to make it easier to read as well as correct gramatical errors.

Feedback:

I shared the original version with three co-workers. All had similar feedback:

1. The caption boxes were not easy to read due to them not being resized. I resized the heigth of the caption boxes, but another solution would also have been to send the a link to view in Tableau. However even if I had done this it is still necessary to scroll the caption box. Resizing was the proper solution.
2. In the chart for passenger ages, it was unclear what the actual age groupings were. It was also noted that it was confusing as to why I included the group of "null" ages in the chart. I removed the nulls from the graph to make it more readable as well as adlusting the bin sizes on the chart. There was also a mention that since I utilzed blue in the original chart it appeared it was a chart about male passengers and not a distribution of all passengers. I changed the color to grey.
3. A few grammatical errors were pointed out. Grammatical errors were resolved.
4. In the last caption box it appeared that the first sentence read 491.68% when in fact this was actually two different sentences. This error was resolved by moving the number 491 into the begining of the first sentence and encapuslating it in paranthesis.
5. Several chart titles and axis titles were difficult to understand. Some titles were removed while others were change to more accurately depict the data it represented.

Resources:

https://www.kaggle.com/c/titanic (https://www.kaggle.com/c/titanic) (dataset details)

https://docs.google.com/document/d/1w7KhqotVi5eoKE3I_AZHbsxdr-NmcWsLTIiZrpxWx4w/pub?embedded=true (https://docs.google.com/document/d/1w7KhqotVi5eoKE3I_AZHbsxdr-NmcWsLTIiZrpxWx4w/pub?embedded=true) (Udacity dataset options)

https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories (https://www.tableau.com/learn/tutorials/on-demand/getting-started-dashboards-and-stories) (formating dashboard and stoy)

https://www.tableau.com/learn/tutorials/on-demand/using-filter-shelf (https://www.tableau.com/learn/tutorials/on-demand/using-filter-shelf) (filtering)

https://www.tableau.com/learn/tutorials/on-demand/grouping?product=&version=10.3&topic=visual_analytics (https://www.tableau.com/learn/tutorials/on-demand/grouping?product=&version=10.3&topic=visual_analytics) (grouping and bins)

https://medium.com/ibm-data-science-experience/markdown-for-jupyter-notebooks-cheatsheet-386c05aeebed (https://medium.com/ibm-data-science-experience/markdown-for-jupyter-notebooks-cheatsheet-386c05aeebed) (markdown text)