Titanic Data Visualization Project

Summary:

This project is done as part of Data Visualization course under Data Analyst Nanodegree program offered by Udacity.

In this project I have designed a data visualization page for Titanic passenger's data, the link to dataset is included below under design section. It contains passengers data of ill fated Titanic ship, in terms of their PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, and Embarked parameters. Based on initial analysis using python, I found strong correlation between survival rate and Passenger's Sex and Pclass in which they were travelling. Hence, I have created a visualization to display with bar chart, survived parameter as y-axis vs Pclass/Sex as x-axis. Also, to make the bar chart intuitive, i have used colors, labels, legends and tooltips.

Design:

To design the data visualization:

I first explored the dataset (used python) -

- The dataset includes 891 passengers with paramaters - 1.Passenger class 2. Sex 3. Age or/and 4. Embarked, SibSp, Parch, ticket, fare and Cabin
- Only 38% of Passengers survived as seen from mean of survival
- About 50% of the passengers were travelling in Pclass = 3
- Data for Age is missing for (891-714) passengers
- Minimum fare is 0 and maximum is 512.32. The maximum fare is a lot more than mean fare, which means it might be an error or outlier for the data analysis
- There is nothing standing out in SibSp and Parch
- Checked the correlation between survived and other parameters and found the strongest correlation between survived, Sex and Pclass parameters
- The % of survival of females within a Pclass of travel was significantly higher than males

From exploration of dataset, the story that plays out is - chances of passengers survival depended upon their class of travel and gender.

Now that I knew the takeaways I want to highlight, I can figure out how to show this data in a way that helps me make these takeaways clear to my audience.

To clearly display the difference in survival chances of a Titanic passenger based on their Sex/Pclass, I went ahead developing visuals and narratives by writing the code using powerful dimple.js charting capabilities enhanced with d3.js for visualization on html web browser.

I chose to display data as bar charts displaying Pclass for female and male passengers on x-axis and count of survived/perished stacked on y-axis. The reason for choosing bar charts is that I want to show comparison between gender and to track changes over Pclass category. Also to easily highlight the stack of passengers who survived vs perished, i used green color (for showing positive result) to highlight survived passengers and brick red (for showing negative result) to highlight perished passengers. For the audience to clearly understand which color stands for what type of information, I used the legend on the right hand top side of the graph.

HTML file editor used: Sublime Text

Titanic dataset downloaded from:
https://storage.googleapis.com/supplemental_media/udacityu/5420148578/titanic_data.csv)

Scripting languages used: html, javascript, d3.js and dimple.js

Feedback:

Feedbacks received are:

1. I shared the titanic-index1.html output with a college student.

Her feedback - *"when user hovers over the chart, it displayed Survived: 0 and again Survived: some other no., which didn't make sense. She could not understand why no matter where she hovered over on the chart, although the stacked bar was represented with different colors, there was no difference, the tooltip always said survived nos twice."*

I corrected this by customizing the tooltip using getToolTipText method provided by dimple for charts in titanic-index2.html

Then I shared the corrected chart to her and she could make more sense from chart. She said, so the 1st class passengers clearly got preference over the others in getting out to life boats.

2. Showed the titanic-index2.html to another user for feedback.

The user's feedback - "*what was the reason for picking the colors for the stacked bars. The colors didn't seem very intuitive*".

I agreed with the user and fixed the colors to be green to display survived passengers and brick red for perished passengers. Then I again showed with these colors to the same user, user thought this gave chart more clarity.

3. I displayed the titanic-index-final.html to 2 different co-workers and one liked the overall chart. Second co-worker asked "*why did i not display female/male in different colors*".

Since I had already tried to display female and male in different colors, and found myself that 4 colors made the chart a bit overwhelming given that I wanted the attention to be on survival/perished nos., and also the tooltip does mention Sex, hence decided to not change this. Instead, i thought my code needed comments, which i added. Also, corrected the formatting as suggested by Udacity reviewer.

Resources:

- Udacity's course material on data visualization
- https://d3js.org/
- http://dimplejs.org
- http://alignedleft.com/tutorials/d3/
- https://bost.ocks.org/mike/bar/
- http://prcweb.co.uk/lab/selection/
- https://buildingvts.com/let-s-learn-d3-js-708577173130
- https://stackoverflow.com/
- http://bl.ocks.org/LeeMendelowitz/11383724
- https://www.sublimetext.com/
- https://github.com/felixge/node-style-guide