**Data Visualization: Titanic Data Visualization**

The project is a part of Udacity Data Analyst Nanodegree, which takes Titanic Data as the input and as we know (From Internet and Movies) that females were given priority to escape the ship. Therefore using d3.js we are going to plot charts and graph to research on it more.

**Exploratory Data Analysis and Cleaning Using Python:**

The data from the site is to be cleaned so that there are fewer errors in the database. The data is cleaned using Python for which the code is attached in the file named cleaning.py. An addition is given to the data that is the ChildPass which divides the passengers into terms of child, Adult and elder as apart from female priority age also could be a priority in that situation, therefore I made a child pass where the children are having the age till 16 and elders starts from 60 and above.

**Visualization of data using dimple.js:**

The data can be easily visualized based on bar graphs which can be easily created using the dimple.js framework. The initial analysis is divided into three charts. The first chart depicts the survival vs age and class, i.e. what was the survival rate of males in females in each class of the ship. The second chart is based on child pass and survival of the people. For e.g.: how many children’s were saved. The final chart is survival vs families that is if a person is traveling alone or is with a family of 2 or more how many people survived in such circumstances. I tested multiple charts like line, dendrogram, tree map. But all these charts were not clearly focusing on the points desired like how age is reflecting survival and females were given priority or not. For e.g. if a Circular Dendrogram is created it will provide a tree structure for which all the branches are to be studied therefore finally bar charts were used.

This initial iteration can be viewed at Index\_Initial.html.

**Feedback**

Here are the feedbacks obtained from the charts. The link for the initial file is present here:

<https://gist.github.com/Anurag111193/d119eeec79b397e19284a93f4427a9b4>

And here are the recorded responses:

**Response 1:**

“You can apply the gender on other graphs also to give another dimension to the analysis.”

**Response 2:**

“There are no clear legends in chart 2, and it is bit unclear, you can make it better by combining gender and classes.”

**Response 3:**

“More details can be added into chart and chart can be made more interactive. Also, use Survival rate for a better analysis.”

**Design after feedback:**

I implemented the following changes:

1. Change of survival to survival rate,
2. Adding male female attributes to chart 2,
3. Labeled graphs with mouseovers,

Now the updated file consists of chart one where the survival rate is compared with class and gender basis. The second chart contains the survival rate compared with gender and childpass column created. Finally, the parch chart is prepared now with better visualization and using survival rate as a major attribute. The file names as index\_final.html. The datasets are now modified and calculated again.

**Conclusion:**

As observed from the first chart which describes the how class was affecting the survival rate of males and females on the titanic, the class 1 females are having the highest chances of survival and so as the males. But we can observe that males have a much lower survival rate. The chart 2 depicts that elder females have the highest survival rate. But so do adult males have a good survival rate in chart 2. Finally chart three shows if single persons were saved or families and we can find that a couple, family of three and four were saved better than a single person abord.

**Resources**

1. <http://rawgraphs.io/>
2. <http://dimplejs.org/examples_index.html>
3. <http://devdocs.io/d3~4/>