P8 – Data Visualization with Tableau

Completed in Partial Fulfillment of the [UDACITY Data Analyst Nanodegree](#)

Trenton J. McKinney

Project Links

[Titanic Survival Feature Analysis - Initial Submission - A Tableau Visualization](#)
[Titanic Survival Feature Analysis - Final Submission - A Tableau Visualization](#)

Overview

For this project, create a data visualization using Tableau that tells a story or highlights trends or patterns in a data set. The work reflects the theory and practice of data visualization, such as visual encodings, design principles, and effective communication.

Summary

The project begins with a single feature (e.g. 'Age Group', 'Ticket Class', 'Gender') view and its ability to predict 'Survival Status'. This is shown as a percent of the total population and as a percent of the whole of a given feature parameter (e.g. 'Adult' within the 'Age Group').

Following the introduction, the project progresses by delving into the features (e.g. Age, Class and Gender) that correlate with Survival Status.

The project concludes with a series of multivariate visualizations exploring how a combination of features is more predictive of Survival Status than any single feature.

Design

1. All graphs are stacked bar which use blue (Perished) and orange (Survived). This was done for consistency and ease of interpretation throughout the project.
2. Stacked bar charts are utilized throughout because the categorical nature of the data, comparing the percent of two opposing conditions (survived vs. perished), seems best represented with that type of visualization.
3. Horizontal orientation of the stacked bars conferred several benefits
   a. Maximize the use of space with side-by-side graphs.
   b. The dependent axis (% of Total Passenger Count) is in the same location throughout.
   c. The visualizations are oriented such that the eyes easily transition from one slide to the next.
4. Drop down filters are used to afford the reader an opportunity to discover at specific items of interest. Additionally, categorical data seems best served with the use of selection filter opposed to a slider, which is better used with continuous data.
5. Tooltips are added to augment the information by including passenger count.

Feedback

1. ☑ Changed to Survival Status. The story title - "Titanic Survival Feature Analysis", whereas the variables is described as "mortality status" in the visual story. This is somewhat not clear for the first-time viewer of the presentation, who does not have any idea about the data set. You can better use "Survival Status" as variable name and axis labels instead of "mortality status" to present your story.

2. ☑ Changed to Perished/Survived. The current Values [0, 1] as mortality status is not helping to foster better communication for the reader of the plots. Provide a better axis labels. Maybe you can define axis for 0 as "Died" and 1 for "Survived".

3. In the second slide of story:
    a. ☑ Removed. I suggest you to either remove the data set and just keeps plots. Or, place the dataset at the bottom and bar plots on top of the story page. Main aim of storytelling is to help reader learn from visualization rather than having a view of the records from the data set.
    b. In the bar plot:
        i. ☑ Changed to percentage. Either use percentage or counts, then display value labels accordingly. It is not a good data visualization practice to represent axis as counts and label value as percentage (%) terms.
        ii. ☑ Changed the y-axis label from Passenger to Passenger Count to reflect units. Similarly, improve the plots in the other slide pages, make sure the y axis units and label data values in units (on bar plots and any other plots) match.
    c. Revise the data to ink ratio concept and try to produce clutter free plots that are easy to read and interpret for the viewers.

4. ☑ Resolved, legends have been added. Add proper legends on the plots to present the color code for survived and died passenger.

5. ☑ Resolved, this slide has been reworked. Slide (3) The Age bin histogram or bar plots with no legends does not help the reader to understand what message is being conveyed through the visuals. I suggest you either plot the full age bins without any category or stacked bars. Or, create two separate histogram distribution of Age bins for each category as survived or died.

6. Bring uniformity in the visual. Resize your plots area, bar heights in the story pages.

7. ☑ Resolved, new labels have been added. In the last page, the use of filter with 0 and 1 values are not clear. Better labelled 0 for died and 1 for survived.

8. ☑ These slides have been reworked with stacked bars and Age Group instead of Age Bins. Once again, Age bins/Gender plot is cluttered, Better way to visualize the distribution is with separate histograms with for Male and Female.

9. ☑ Max of two visualizations per page. Most Important: Present Visual plots with one or max two key messages or insights per slide page.

10. ☑ Better captions, titles and axis labels have been added. I encourage you to think about creating better captions, titles and axis labels.
    a. Writing short, easy to eye captions text that helps reader to interpret information quickly. You can also, write the key message on the caption to present your findings or write/ask an interesting question on caption to generate curiosity in the mind of the reader.

Resources
1. RMS Titanic – Wikipedia
2. Titanic Disaster: Survivability Parameters