

# P6 – Data Visualization with Tableau

## Ahmad Takatkah

Project's Final Version :

<https://public.tableau.com/profile/vcprenneur#!/vizhome/TitanicDataVisualization-v2/DAND-P6-v2>

Project's Initial Version Prior To Feedback:

<https://public.tableau.com/profile/vcprenneur#!/vizhome/TitanicDataVisualization-v1/DAND-P6-v1>

### Summary:

The Titanic ship sank in the North Atlantic Ocean in 1912 after colliding with an iceberg, killing more than 1,500 of the 2,224 passengers aboard.

In the provided visualization, we're going to see the likely effect of several factors on the survival rate of passengers. We will start by gender which will be our main study factor. Then we will add passengers' age groups and classes to gender separately to see how these two combinations would affect survival rates. In the final step, we will add both age groups and classes together to gender in one visualization to get detailed insights on passengers' survival rates.

### Design:

Throughout the visualization, only three colors were used: pink, light teal, and blue gradient.

Reasoning: This will help keep a consistent color encoding to make it easier to read the plots.

The story consists of 4 pairs of slides. A pair of slides includes a dashboard and a table for the factors in focus. A dashboard is split into two sides, the left side shows a plot before the accident, and the right side show the same plot after the accident split into two quarters based on the passengers' status: survived or perished.

Reasoning: This predictability helps the viewer develop the story as they read through the slides to arrive to the conclusion I aim to deliver.

I limited the use of visualizations to bar charts and tables. In every bar chat, the y-axis shows the count of passengers, and the labels on top of the bars show the percentage of females/males of the same single bar (or whole plot for the first dashboard only for more simplicity).

Reasoning: In my study of the titanic data, the focus is basically on the count of people survived vs. perished, based on available social and economic data. Bar charts are best suited to show such a comparison between counts/frequency of occurrences of multiple categories. In addition, I added tables with percentages based on survival status with 1, 2, and 3 level of details (in 3 different slides). Those tables were color coded with a gradient blue to help the viewer visualize the effect of gender/age/class on the survival rate.

## Feedback:

The initial version of the visualization was shared with one co-worker. Below is the received feedback and the changes that I made based on the discussions we've had:

1. It was not clear what ages corresponds to the age groups provided by the visualization. As a result, I added a caption to the dashboard to list the age groups and their correspondent ages.
2. The starting slide in the story was a simple age distribution of the passengers, but the next slides showed a more meaningful classification of passengers by their age groups, so the feedback was that the first slide is not needed. As a result it was removed from the story.
3. The last two slides in the story provided two ways to look at the combination on 3 factors. This was confusing and the recommendation was to select only one of those two slides because they were basically communicating the same message. As a result, I kept one of the two slides, and replaced the other with a summary table to create 4 similar pairs of slides for every set of factors.

## Resources:

The original dataset used in this project was selected from the [Data Set Options](#) by Udacity. The original dataset was then cleaned and more features were added as part of [Project 2: Investigate A Data Set](#). The new cleaned data set is provided with this write up to the reviewer.

The cleaning process included:

- Imputing the missing age value by the age average of passengers with the same social "Title",
- Dropping features that were not adding value to the purpose of the needed analysis such as: 'Ticket', 'Fare', 'Cabin', 'Embarked'
- Creating a new categorical feature: "Age Group", as follows:
  - o Child: 1-13 years old,
  - o Teenager: 13-19 years old,
  - o Adult: 20-39 years old,
  - o Middle Age: 40-59 years old,
  - o Senior: 60+ years old