Titanic Data explained using Tableau

Summary

In the Titanic story created, I have explained the various factors that most likely led to an individual's survival in the titanic tragedy.
My story consists of 4 main visualizations explaining whether there is a relation between an individual's gender, age, social class, ticket fare, number of companions onboard, etc. and their survival.

Final version url:
https://public.tableau.com/views/Titanic_429/Titanic_story?:embed=y&:display_count=yes&publish=yes

Initial version url:
https://public.tableau.com/views/Titanic_feedback/Titanic_story?:embed=y&:display_count=yes

Design

Datatypes

The datatypes in the titanic dataset were modified as below:

- Passenger ID was converted to string
- Survived column indicating a person's survival was converted to string with values '1' (survived) and '0' (not survived).
- 'Class' was converted to string.

Calculated fields

- As the Survived column consisted of codes 0 and 1, I created another calculated filed 'SurvivedYN' to indicate survival using Yes and No.
- Created two boolean fields 'has parent/children' and 'has sibling/spouse'
- Created a field 'relatives' to calculate number of relatives onboard for each individual (a sum total of parent/children field and sibling/spouse field)

Visualizations

- As part of exploring demographics of survivors as well as non-survivors I conducted Exploratory Data Analysis to see patterns in data, with final story highlighting the 'Explanatory' Analysis.

- Color was used in multivariate analysis plots to differentiate survived and non-survived individuals.
- As the outcome variable ‘survived’ is a categorical variable represented using color, a bar plot was used in most of the analysis to represent other categorical variables such as gender, class, etc.
- In parts where count of ‘survived’ was needed for each bin (example: age bin), a histogram was used.
- In order to create the final story, I included multiple dashboards, mapping each dashboard to a caption to highlight the key independent variables.

Feedback

- Based on the feedback received, some plots were recreated to indicate percentages. Example, impact of the travel class on survival can be seen more easily when original plot is presented along with its percentage variant.
- Also the names of the axes and the fields were changed for better readability and understanding.
- The reviewer asked questions such as why there are separate plots for class as well fare? Aren’t fares indicative of the class? Upon which I answered that ‘Yes, the fares are based on the travel class, but the fares are actually varying a lot with some third class fares being equivalent to some first class fares.
- The positive feedback mentioned consistent use of good color scheme across all visualizations clearly highlighting ‘Survival’ while maintaining information regarding non-survivors.
- Resources: NA