

Jacqueline Lopez

STA360-A

Dr. Lee

02/01/24

Final Report with 5 Visuals

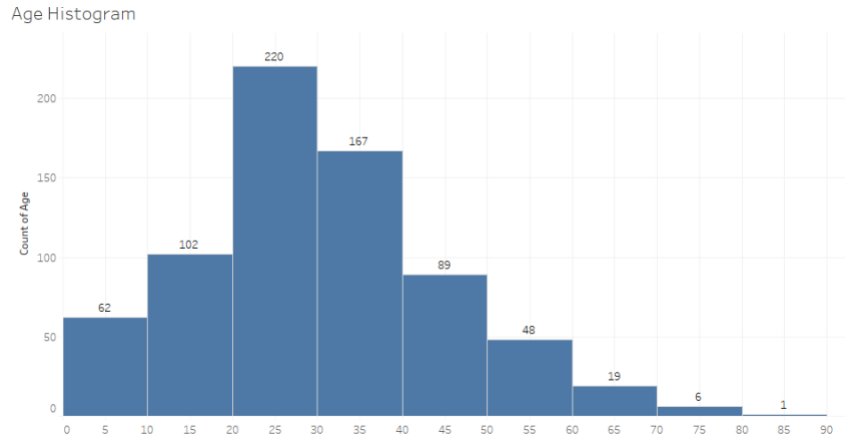
Graph #1:

Class vs. Sex Survival Heatmap

Sex	First	Second	Third
Female	96.81%	92.11%	50.00%
Male	36.89%	15.74%	13.54%

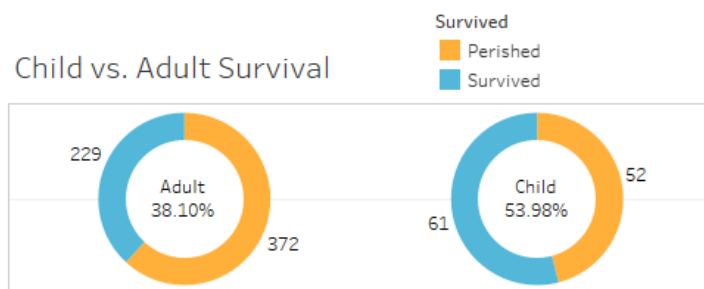
1. **Purpose (3):** The purpose of the heatmap my group and I made is to represent the percentage of survival based on sex and class. This visual is helpful because it is easy to read. By looking at the heatmap, the viewer can tell that darker blue represents a higher percentage of survival and lighter blue represents a lower survival percentage.
2. **Variables (2):** Q's – What percentage of female/male first/second/third class survived the Titanic? We used string variables Pclass and Sex to create the table for our calculated field. The calculated field was created to represent the Survived_Percent, for which we used the expression $\text{SUM}([\text{Survived}])/\text{COUNT}([\text{Survived}])$. I believe we selected the suitable data types to address our question because to find the survival percentage, we needed to create a calculated field.
3. **Storytelling (4):** The viewer can get to the conclusion based on the visual quickly, given it is a heatmap, which makes the user able to look at the shades of blue first and then the percentage to determine the percentage of survival based on sex and class. The darker blue is more likely to survive, and the lighter blue is less likely to survive.
4. **Accuracy (1):** This visual accurately represents the Titanic dataset because, through it, we can see that Females from 1st and 2nd class have a higher survival percentage than males from 2nd and, which accurately aligns with the actual Titanic.

Graph #2:



1. **Purpose (3):** The purpose of this histogram is to represent the total number of passengers by age. This visual is helpful because it allows viewers to determine quickly and neatly which age group had the most passengers.
2. **Variables (2):** Q's – How many passengers of the age group _____ were on the Titanic?
The reasoning behind choosing the bin size of age to be 10 was to organize the bar chart to be readable and not have too many or too few bins. Considering we used age bins and counted to create the histogram, I do believe we selected suitable data types to address the questions.
3. **Storytelling (4):** The viewer can conclude which age group had the most passengers by finding the tallest bin, which will be labeled with the total on top. The visual shows that middle-aged people were the most passengers, and fewer children and elders were on board.
4. **Accuracy (1):** This visual accurately represents the Titanic data set without distortion because it states that middle-aged people were the age group that had the most passengers, which aligns with the actual Titanic.

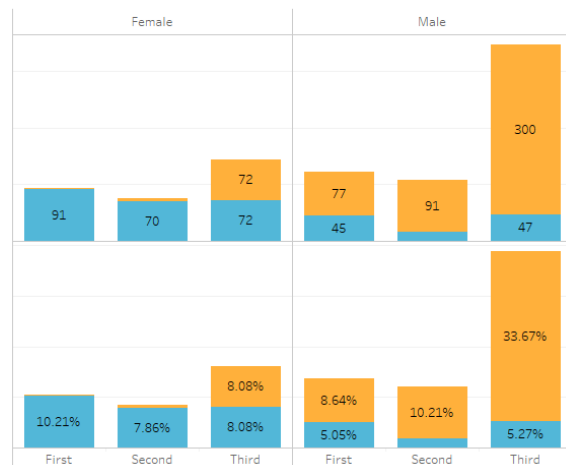
Graph #3:



1. **Purpose (3):** The purpose of our two doughnut charts is to represent the difference in Child vs. Adult survival. This visual is helpful because viewers can see the total number of people who perished vs. survived in children vs. adults since the viewer is provided with color representation, number, and percentage of adults vs children.
2. **Variables (2):** Q's – What percentage/total number of children/adults survived?
Considering we used both numeric and percentage as well as string, I would say we did select suitable data types to address the proposed question.
3. **Storytelling (4):** Viewers can conclude based on the color that takes up more of the doughnut. For example, when looking at the adult doughnut chart, orange (perished) takes up more space, and blue (survived) takes up more space for children—concluding that children are more likely to survive than adults.
4. **Accuracy (1):** This visual accurately represents the Titanic dataset without distortion because it concludes that children were more likely to survive the Titanic, which aligns with the actual Titanic.

Graph #4:

Survival based on sex and class

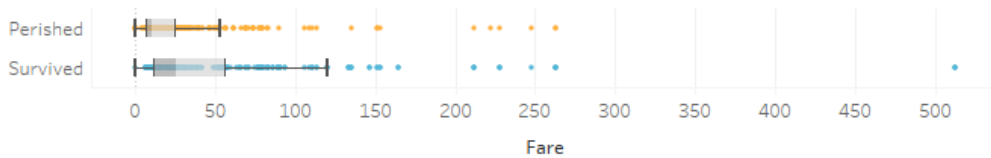


1. **Purpose (3):** These bar charts aim to distinguish between sex and class and the total and percentage of those who survived. These bar charts are useful because of their straightforwardness.
2. **Variables (2):** Q's – What total/percentage of males vs females survive based on class?
Considering we used bins of interval 10 to represent age and count as the total, I believe we did use suitable data types to address our question.

3. **Storytelling (4):** Viewers can conclude based on the tallest bin and determine the total or percentage that survived or did not survive. Through these bar charts, we can conclude that males from the third class had.
4. **Accuracy (1):** This visual accurately represents the Titanic dataset without distortion because it concludes that males, specifically those of third class, are least likely to survive, which is accurate. Which accurately aligns with the real Titanic's.

Graph #5:

Fare vs survival boxplot



1. **Purpose (3):** The purpose of this box plot is to represent the difference in fare of those who survived vs those who perished. This visual is useful because, based on the boxplot, viewers can quickly learn the median fare for the perished vs survived.
2. **Variables (2):** Qs – What is the difference in the fare of those who survived vs. those who perished?

Considering we used fare (numeric) and survived (numeric discrete), I believe we selected the suitable data types to address the question.

3. **Storytelling (4):** Viewers can conclude the fare median by looking at the midpoint of the box plot. Based on the box plots, we can see that those who survived paid more for their ticket than those who did not.
4. **Accuracy (1):** This visual accurately represents the Titanic dataset without distortion because the conclusion accurately aligns with the actual Titanic.

Dashboard:

