**DATA6550 – Visualization Ethics and Communication**
**Titanic Survival Analysis: Accurate and Misleading Visualizations**

## 1. Introduction

Data visualization has a major influence in determining how information is viewed and comprehended. While visualizations may successfully convey patterns and insights, design decisions can sometimes mistakenly or purposefully mislead viewers. The objective of this research is to illustrate both ethical and deceptive visualization approaches using the Titanic dataset. By developing correct representations alongside purposely false ones, this study studies how graphical design choices impact perception, interpretation, and story formation.

## 2. Dataset Overview

The dataset utilized for this research is the Titanic passenger dataset, which comprises demographic and survival statistics for people onboard the RMS Titanic. The training dataset comprises 891 passengers and 12 characteristics.

### Key variables analyzed include:

1. Survived (0 = Did not survive, 1 = Survived)
2. Sex
3. Pclass (Passenger class: 1st, 2nd, 3rd)
4. Age
5. Fare
6. Embarked
7. Rare Title (0 = Doesn't have a rare title, 1 = Has a rare title such as Sir, Lady, Countess, etc.)

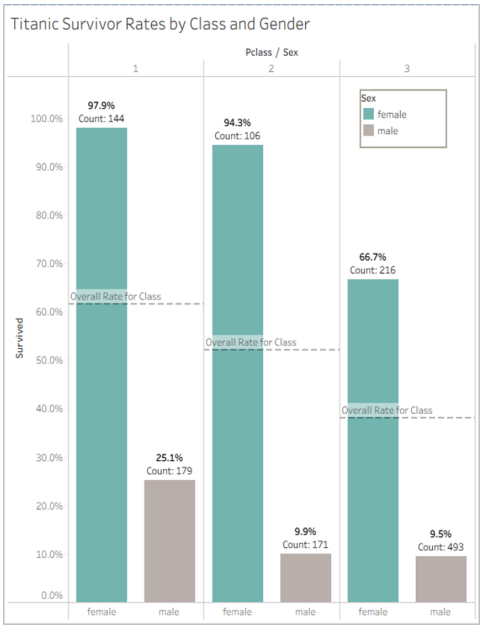
### Initial exploration revealed:

- Approximately 38% of passengers survived, while 62% did not.
- The majority of passengers were male (577 men vs 314 females).
- Most travelers went in 3rd class (491 passengers).
- Significant missing values were detected in Cabin (77%) and Age (20%), which shows the significance of proper data management before display.

This dataset gives significant opportunity to examine survival trends across demographic and socioeconomic groupings.

## 3. Accurate Visualizations

### 3.1 Survivorship by Gender and Class




Follows reasonable design principles with pleasing colors that align among categories, legible fields, removal of excessive gridlines and existence of a legend.

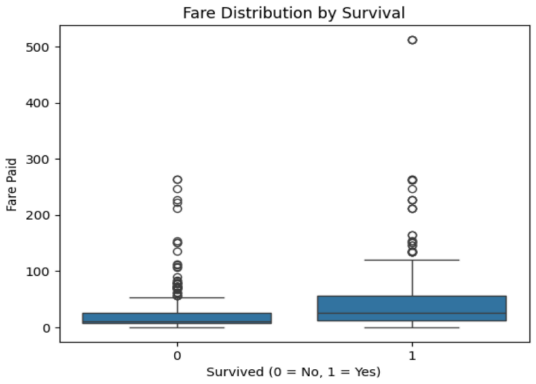
Properly displays data (specifically survivorship) separated by two categories, gender and passenger class. The focus is on survivor rates (or percentages) within each category, although overall counts are also displayed secondary. Further insight is contained within the visual as the overall rate for each passenger class.

Title and any written fields are informative and not charged or emotional.

Overall, the audience is able to comprehend the visual as-is without being led to believe something untrue.

**DATA6550 – Visualization Ethics and Communication**
**Titanic Survival Analysis: Accurate and Misleading Visualizations**

### 3.2 Survival rate by Fare Distribution




The second accurate visualization shows the distribution of ticket fares based on whether passengers survived or not. I used a boxplot because it clearly displays the median, spread, and outliers without hiding any part of the data.

Passengers who survived generally paid higher fares.
The median fare for survivors is noticeably higher than for non-survivors.
Survivors also show a wider range of fare values, including several very high outliers.
Most non-survivors paid lower fares, and their fare values are more concentrated at the lower end.

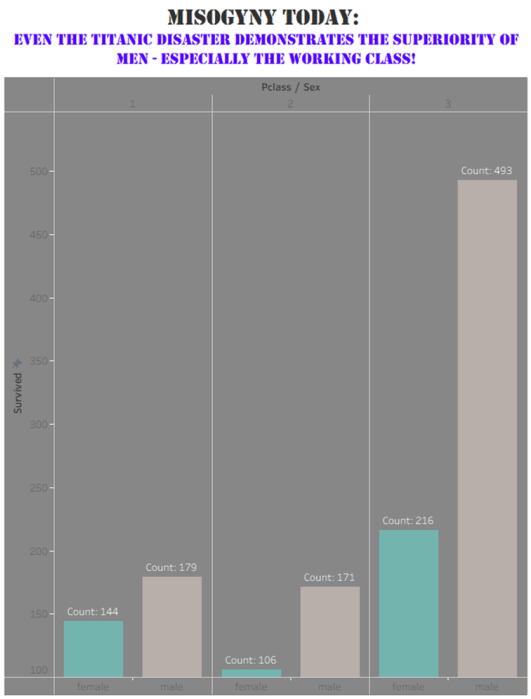
From this chart, we can see that passengers who paid higher fares were more likely to survive. Since fare is closely connected to passenger class, this suggests that class played an important role in survival outcomes.

Because the boxplot shows the full distribution, including medians, variability, and extreme values, the visualization presents the data clearly and honestly. It does not hide outliers or manipulate the scale, making it an accurate and ethical representation of the relationship between fare and survival.

## 4. Misleading Visualizations

Two purposely false charts were developed to highlight how visualization selections may influence perception.

### 4.1 Truncated Axis (Gender & Class Survival)




Design principles – unpleasant color scheme with poor contrast between different elements. Title and caption don't jive with the scheme of the visual

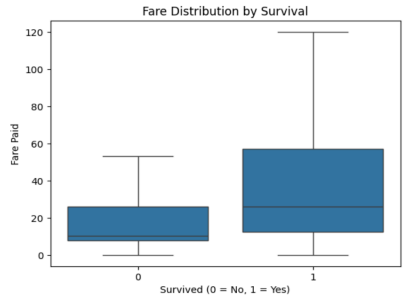
- Emotionally charged caption indicates what the author is trying to get his audience to conclude.
- Display of data is misleading for various reasons:

➔ Scale is truncated, making the differences among each class/category more dramatic

➔ The data displayed are counts of those who survived in each category, and although these are accurate counts, it ignores the relation to each population's original count, making the third passenger males seem like the "winners" in this tragedy (when, in reality, they had the gravest outcomes).

**DATA6550 – Visualization Ethics and Communication**
**Titanic Survival Analysis: Accurate and Misleading Visualizations**

### 4.2 Misleading visualization by Fare Distribution




The misleading visualization shows the fare distribution by survival status, but in this version, the outliers were intentionally hidden using "showfliers=false." By removing the extreme high-fare values, the chart makes the difference between survivors and non-survivors appear smaller and less significant than it actually is.

In reality, survivors include several very high fare payments, which are important because they reflect first-class passengers who had higher survival chances. The average fare for survivors is around £512, while for non-survivors it is around £263. By hiding these outliers, the visualization reduces the visible spread and masks the true variability in fares. As a result, the visible range for survivors drops to around £120, which makes the difference between survivors and non-survivors look much smaller than it truly is.

Although the data itself has not changed, removing extreme values presents an incomplete picture and can mislead viewers about the strength of the relationship between fare and survival.

This visualization demonstrates how omitting key information such as outliers can distort interpretation and create a biased or unethical visualization, even when the underlying dataset remains the same.

## 5. Ethical Considerations in Data Visualization

Data visualization carries ethical responsibility because design choices directly influence how audiences interpret information. Even when data is accurate, manipulating scales, hiding outliers, or presenting raw counts without context can distort understanding. In our project, we demonstrated how truncated axes and removal of extreme values can change perception without altering the dataset itself. Ethical visualization requires transparency, proper scaling, full distribution display, and avoidance of emotionally biased language. Analysts must ensure that visuals inform rather than persuade unfairly. Responsible data communication builds trust and supports sound decision-making.

## 6. Key Insights

From our accurate visualizations, two major insights emerged. First, gender and passenger class significantly influenced survival outcomes, with females and first-class passengers having much higher survival rates. Second, fare distribution revealed that passengers who paid higher ticket prices were more likely to survive, reinforcing the impact of socioeconomic status. The boxplot clearly showed higher medians and wider variability for survivors, including extreme high-fare values. These patterns suggest that both social structure and economic position played critical roles in survival during the Titanic disaster.

## 7. Conclusion

This project demonstrates how the same dataset can produce either truthful or misleading narratives depending on visualization design choices. Accurate visualizations provided transparent and proportional representations of survival patterns, while misleading versions exaggerated or minimized differences through scale manipulation and omission of outliers. The study highlights the importance of ethical awareness in data communication. As future data professionals, we must prioritize clarity, honesty, and context to ensure that visualizations support informed interpretation rather than distortion.