

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

Group 3: Beladiya, Jay Sanjaykumar
Indala, Divya Sai Sukanya
Panula, Jennifer
Valaboju, Shiva Kumar Chary

Disclaimer: Some help from AI (ChatGPT) was utilized for coding assistance and with the (powerpoint) presentation.

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

Figure 1 follows reasonable design principles with pleasing colors that align among categories, it has legible fields, and it doesn't contain excessive gridlines. It also has a legend. It 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 secondarily. Further insight is contained within the visual as the overall rate for each passenger class.

DATA6550 – Visualization Ethics and Communication

Titanic Survival Analysis: Accurate and Misleading Visualizations

The 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.

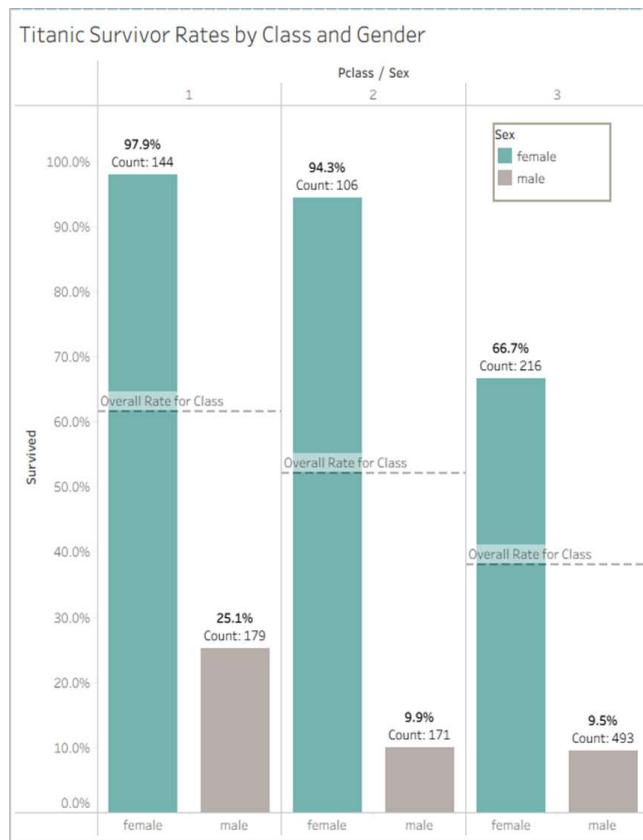


Figure 1

3.2 Survival rate by Fare Distribution

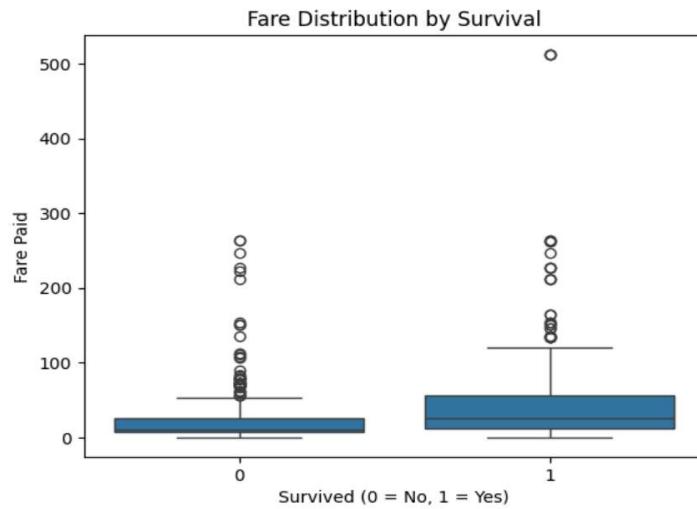


Figure 2

DATA6550 – Visualization Ethics and Communication

Titanic Survival Analysis: Accurate and Misleading Visualizations

The second accurate visualization shows the distribution of ticket fares based on whether passengers survived or not. A boxplot was used 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 figure, 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)

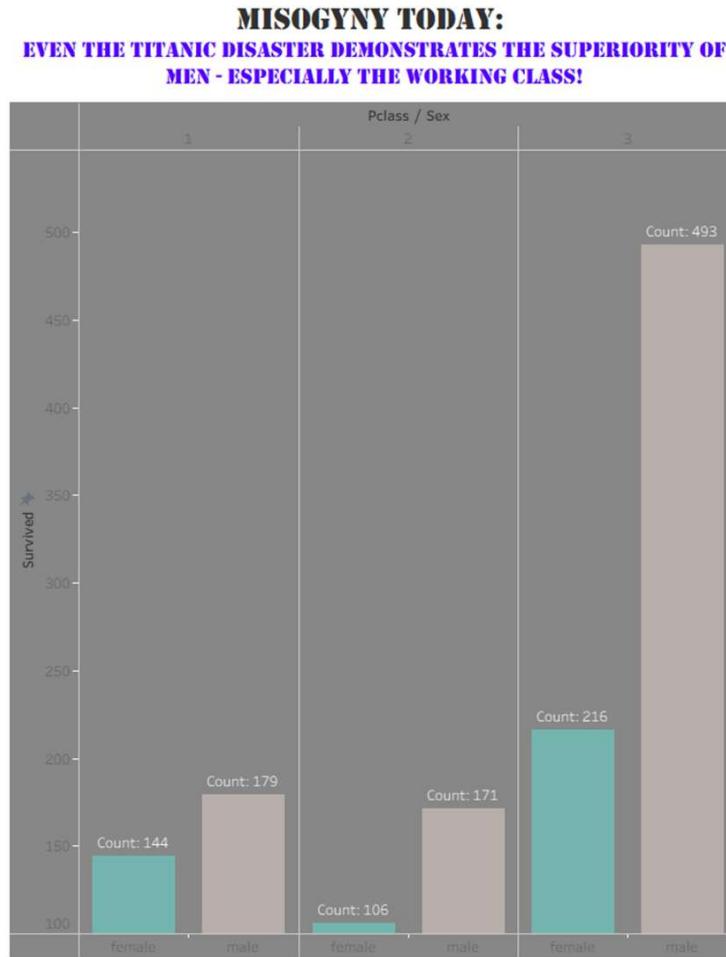


Figure 3

DATA6550 – Visualization Ethics and Communication

Titanic Survival Analysis: Accurate and Misleading Visualizations

In Figure 3, design principles are not followed ideally, as it has an unpleasant color scheme with poor contrast between different elements. The title and caption style also fails to agree with the scheme of the visual.

The emotionally charged caption indicates what the author is trying to get his audience to conclude. Additionally, the display of data is misleading for various reasons:

- The 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-class male passenger seem like the "winners" in this tragedy (when, in reality, they had the gravest outcomes).

4.2 Misleading visualization by Fare Distribution

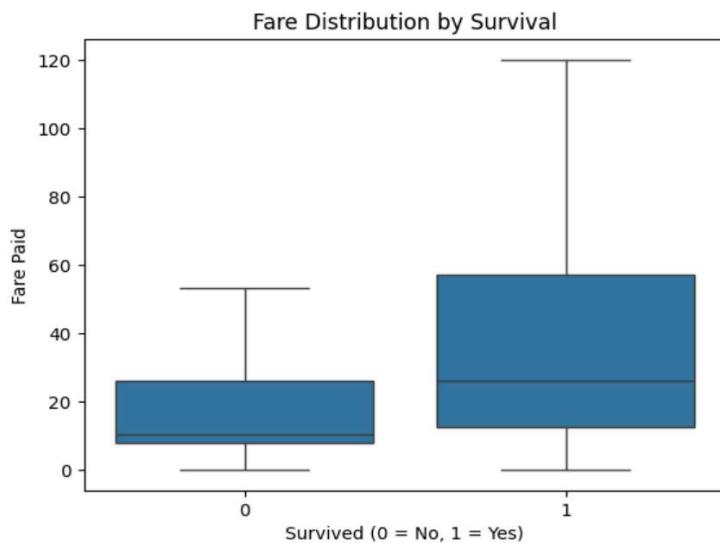


Figure 4

Figure 4 demonstrates another misleading visualization, showing the fare distribution by survival status. 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.

DATA6550 – Visualization Ethics and Communication

Titanic Survival Analysis: Accurate and Misleading Visualizations

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

Because design choices have a direct influence on how audiences comprehend information, data visualization implies ethical responsibility. Manipulating scales, obscuring abnormalities, or providing raw statistics without context may affect interpretation even when the data is true. In this research, we demonstrated how deleting extreme values and truncating axes could alter perception without modifying the dataset itself. Transparency, suitable size, comprehensive distribution displaying, and avoiding emotionally charged language are all important for ethical visualization. Visuals should inform, not unfairly persuade, according to experts. Effective data communication supports rational decision-making and creates confidence.

6. Key Insights

Our detailed visualizations generated two crucial insights. First, survival outcomes were greatly influenced by passenger class and gender, with first-class and female passengers having much higher survival rates. The significance of socioeconomic status was further underlined by the fare distribution, which revealed that customers who paid more ticket charges had a better probability of surviving. Higher medians and higher variability for survivors, including extremely high fare values, were plainly seen in the boxplot. These tendencies show that economic position and social standing were major determinants in predicting survival after the Titanic tragedy.

7. Conclusion

This study illustrates how various graphic design selections may result in either accurate or misleading narratives from the same information. While misleading versions employed scale manipulation and the removal of outliers to accentuate or decrease differences, authentic visualizations presented unambiguous and balanced portrayals of survival patterns. The study underscores how vital ethical conscience is when conveying data. Clarity, honesty, and context must be given high attention by us as aspiring data professionals in order to assure that visualizations support right interpretation rather than deception.