Data Visualizations

Group 1

Will Holt, Vaishnavi Paniki, Sebastian Segura

Introduction

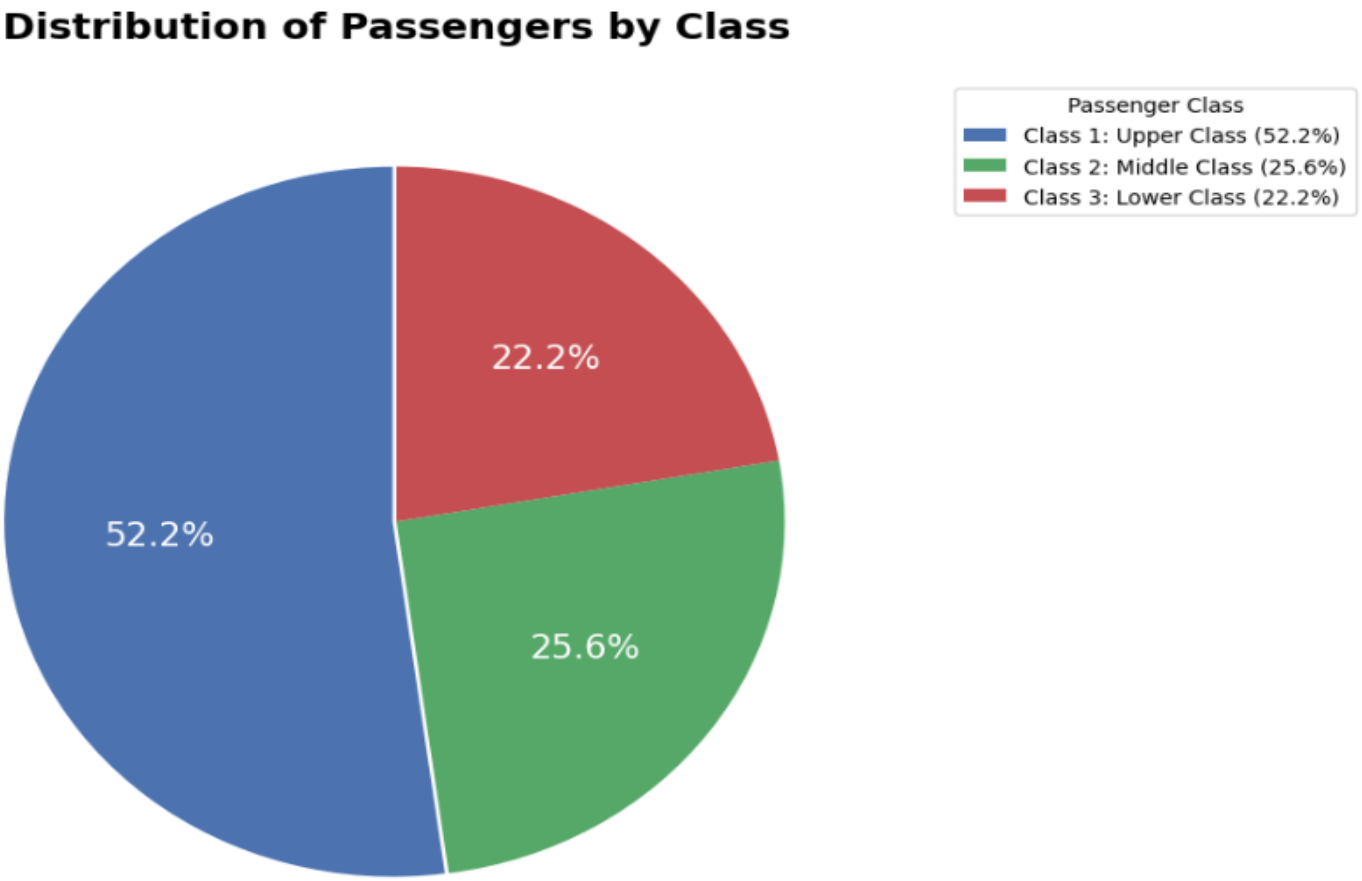
Mastering data visualization is critical to one's ability to shape how audiences will interpret information. A good visualization will provide viewers with everything they need to reach clear insights and conclusions to increase their level of understanding. A poorly done visualization, whether it is poor on purpose or by mistake, will only serve to confuse audiences and potentially spread false information.

This report will show examples of visualizations created from the Titanic dataset that adhere to best practices and effectively communicate data, along with examples of terrible visualizations which we have designed to intentionally highlight common errors that distort the information, confuse audiences, or otherwise ineffectively communicate its data. Using the Titanic dataset, we will highlight and discuss distinct differences in the good visuals and the bad visuals, and explain the ethical implications of misleading data representation, and discuss why it is crucial to focus on the truth when presenting information. Through this process, we will explain the visualization choices we made to help enhance the data presentation, as well as highlight the intentionally poor choices made to construe the data's meanings.

The Titanic Dataset

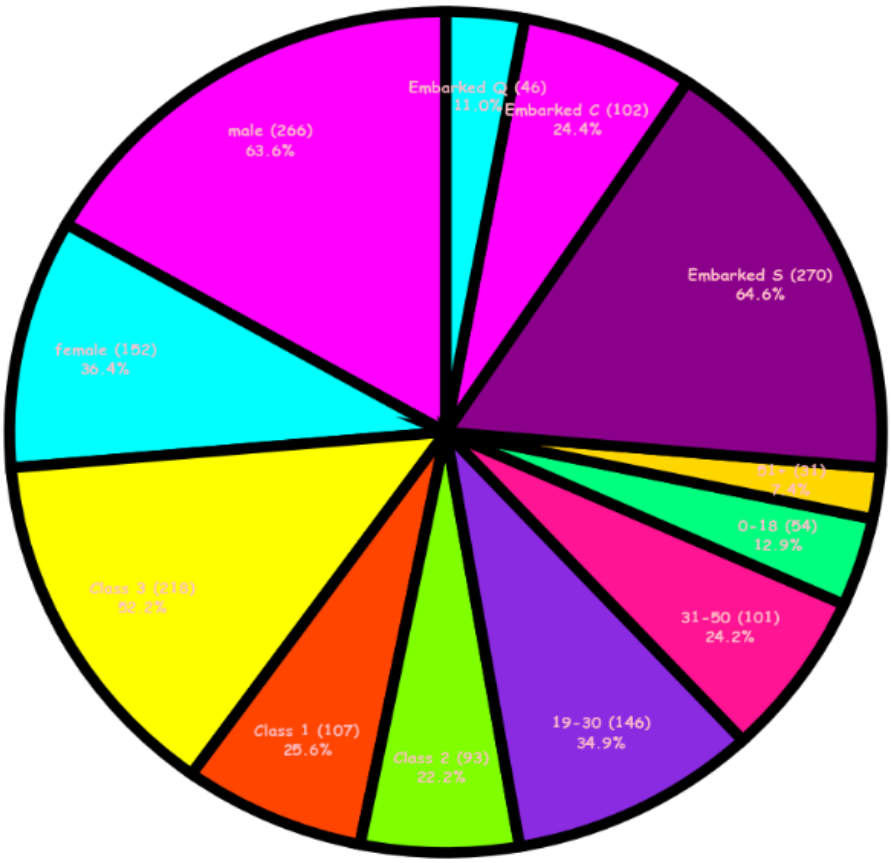
The Titanic dataset contains demographic and survival information for passengers aboard the Titanic vessel and allows for analysis of survival patterns across various demographic aspects, such as age and gender, among others. The dataset's purpose is to allow for the analysis of survival patterns across demographic groups. We will be using this dataset to create visuals to gain insights into disaster survival, as well as intentionally trying to falsify or misconstrue as much of the data as possible.

Pie Charts of Passenger Distribution



This pie chart effectively sticks to the best practices as outlined in our readings for the module. The colors are soft and complementary. This allows for a pleasant differentiation of slices and limits the viewer from having to over-process information. The 'explode' effect used draws subtle attention to the largest passenger class without dramatically altering the presentation. The percentages are accurate and provide insightful information about the distribution of passenger class aboard the Titanic. The analyst has positioned the legend well, and can stand on its own, demonstrating its quality. Overall, this pie chart is professional and effectively communicates the fact that more wealthy passengers seemed to have survived at a higher rate than other less wealthy groups.

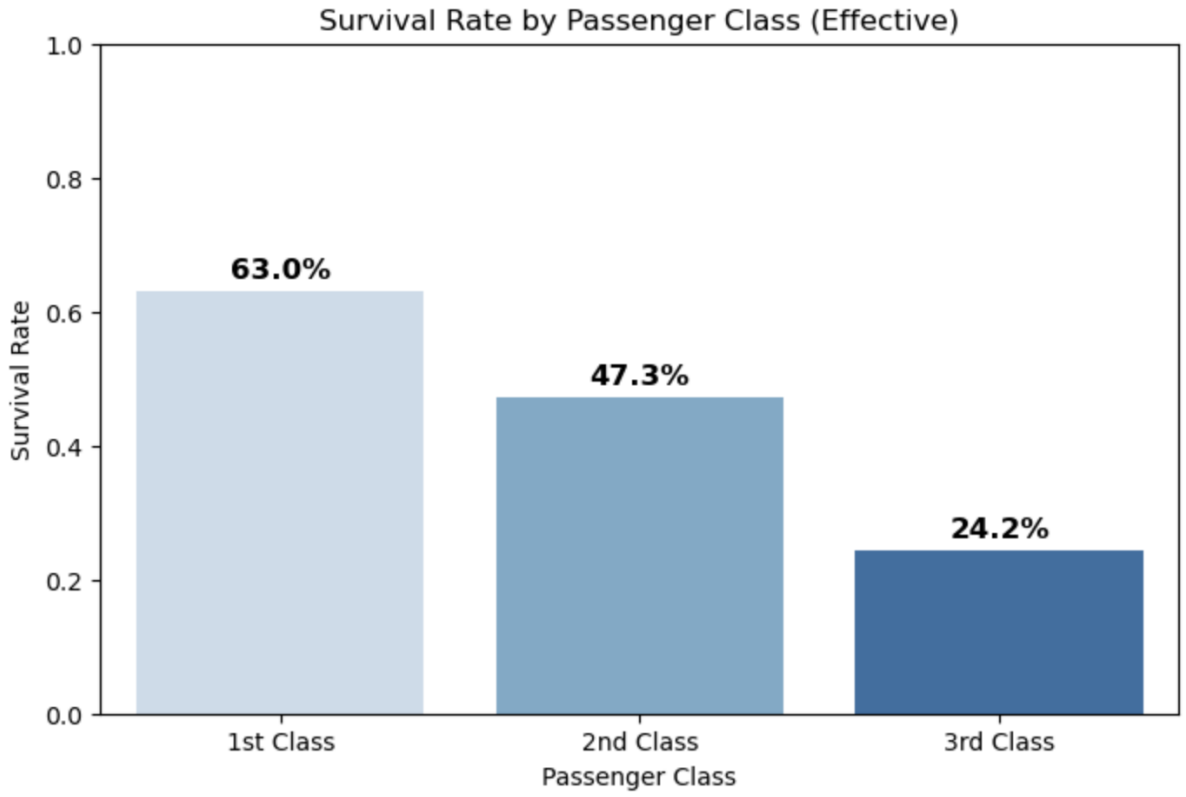
Terrible Pie Chart: Every Measurable Aspect of Passengers



Now we have successfully taken a decent pie chart and turned it into something a little less beautiful. In our attempts to 'improve' the original pie chart, we somehow broke many of the data visualization rules. The beautiful neon colors remind the viewer of a kaleidoscope, and the thick black borders remind me of a beautiful stained glass church window. The labels are tiny and pink and in Comic Sans font. This allows the audience to focus more on the design of the visualization instead of the information. The Jokerman font title needs no further explanation. Finally, if you sift through all of that, you find that the percentages do not add up to 100%. By attempting to increase the amount of information shown, the true message of the data has been completely lost on the audience.

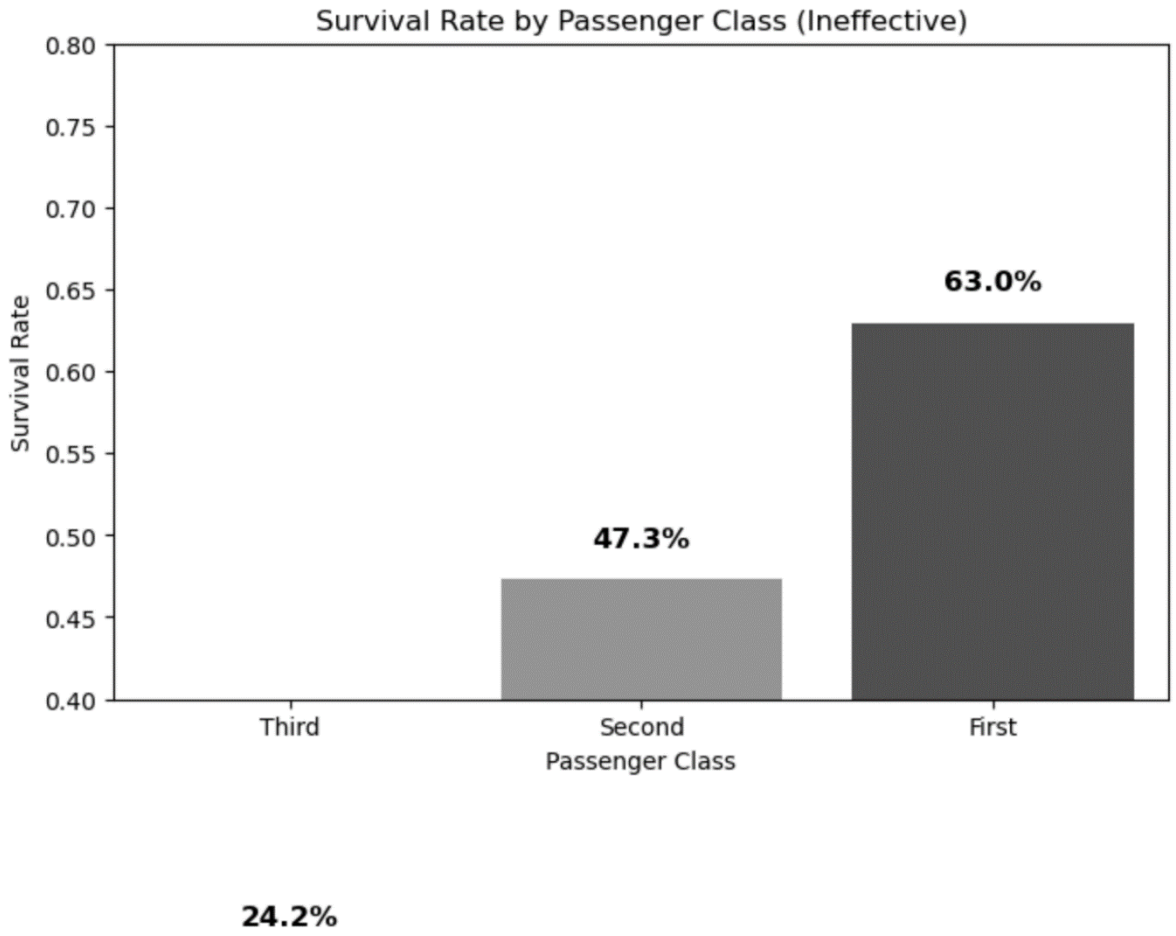
The contrast between the two pie charts highlights how bad design choices with good intentions may still obscure the truth.

Survival Rate Based on Passengers Class Bar Chart



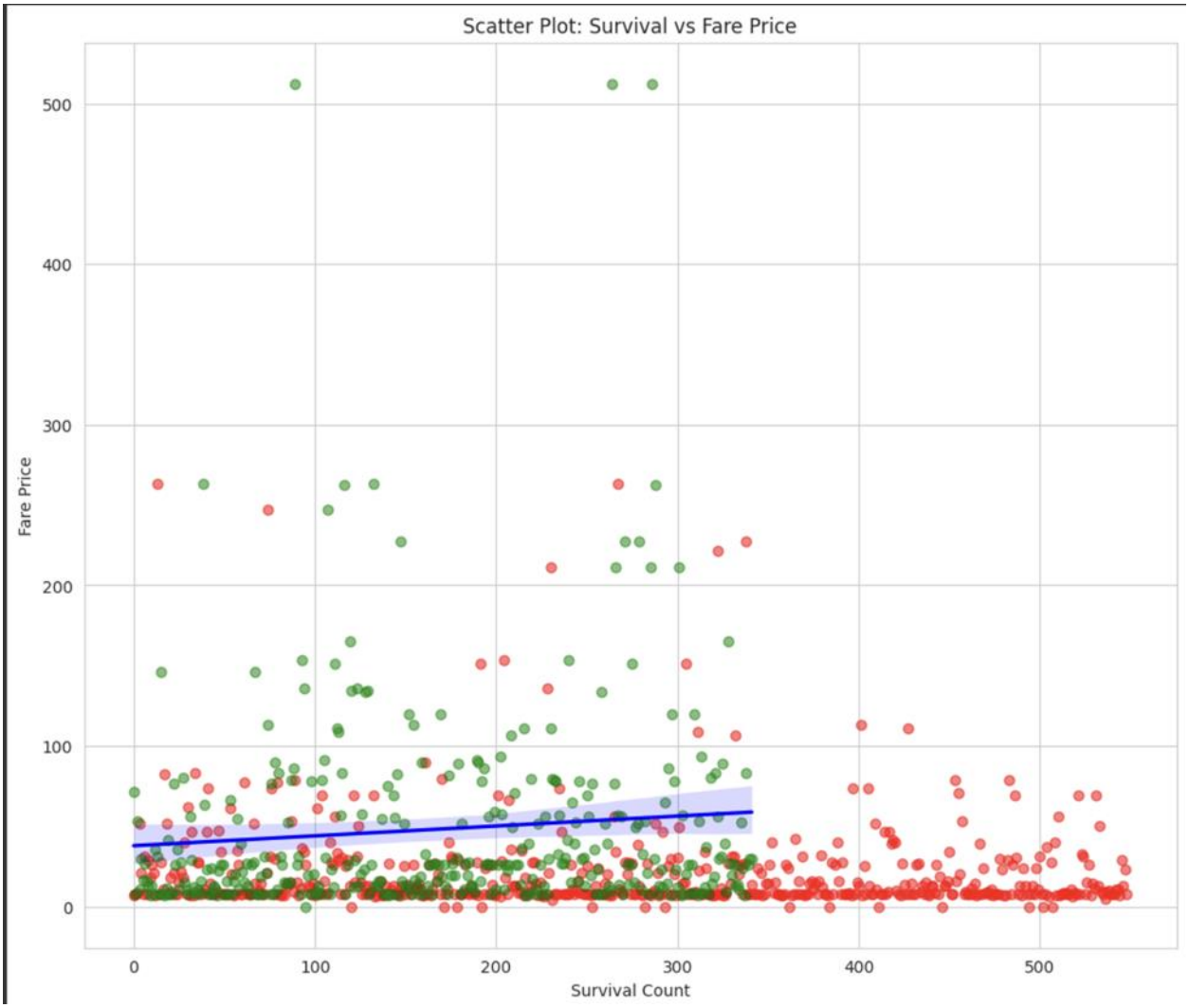
Here we are easily able to understand the message of this bar chart. It shows survival rates by passenger class in a descending order with a blue gradient to further highlight the differences. It is effective thanks to the intuitive design with accurate scaling on the y-axis, and with clear labels to compare the survival rates. There is not too much information, and it is nice to look at.

Ineffective Visualization (Bar Plot):



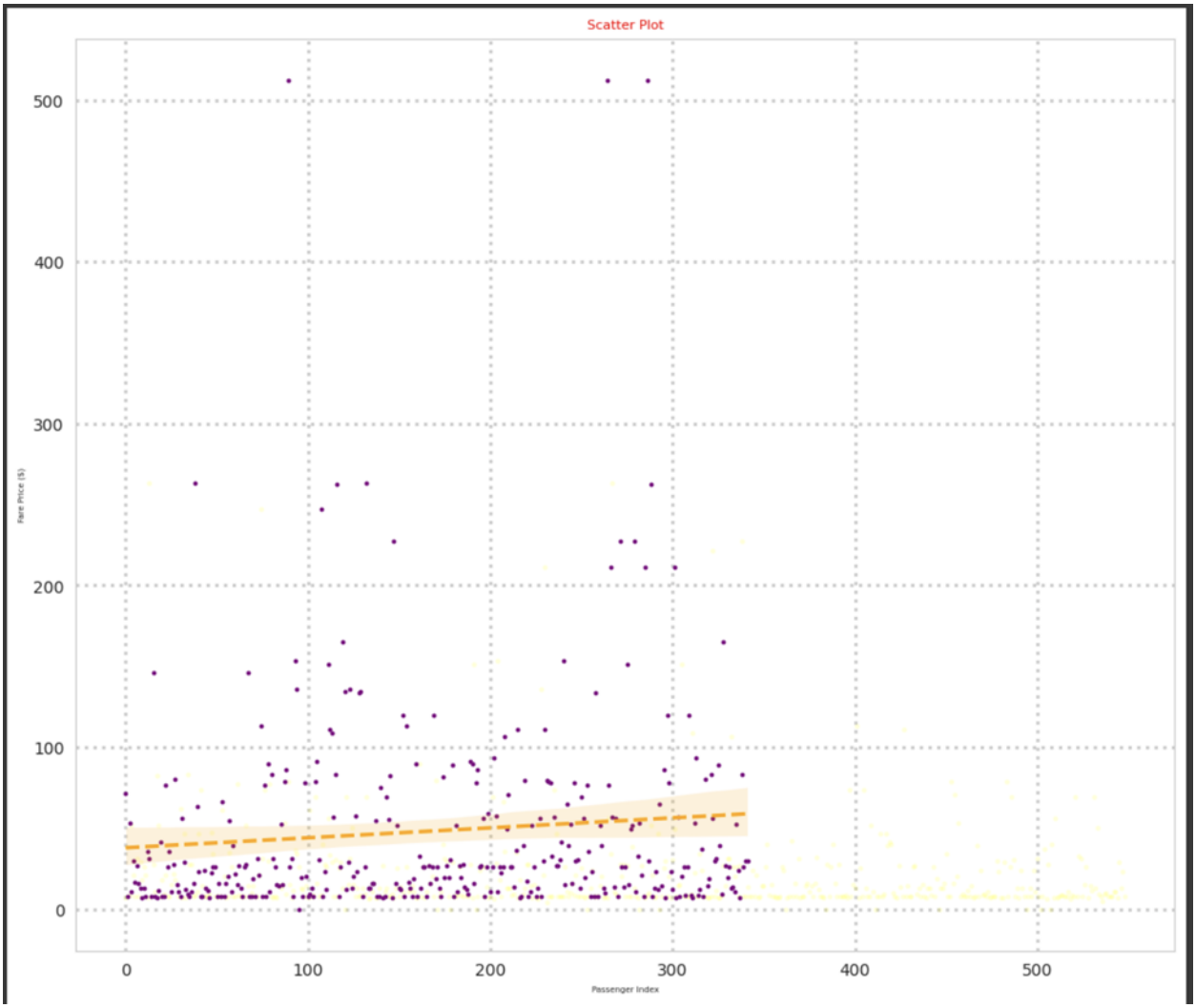
The ruined bar chart shows survival rates by passenger class on a truncated y-axis. This visualization design fails as the y-axis completely clips off the third-class passengers and implies that none of them survived at all, when 24.2% did survive (as shown below the axis). Various dubious conclusions could be drawn from this cherry-picked data on this truncated axis, but most importantly, the axis shift has put a magnifying glass over the differences in class survival rate, expanding and exaggerating their differences.
The good bar chart is straightforward and conveys only the facts surrounding class survival rate. The ruined bar chart effectively highlights how easy it is to manipulate data by exaggerating differences.

Scatter Plots for Survival and Fare Price



This scatter plot is exploring the fare price for a ticket aboard the Titanic and the survival rates of those passengers. It uses green for survivors and red for non-survivors, along with a trend line to help further explain the comparison between the two. Thanks to the color coordination and clear labeling, we can easily see that many of the non-survivors paid over $300 for a ticket!

Bad Scatter Plot for Survival vs fare price:



This version of the scatterplot highlights the importance of distinguishing between data points (survivors vs non-survivors in this case), as well as showing why choosing the correct graph is important. With this additional data added, the scale extended, and every aspect shrunk down beyond readable levels, this graph effectively highlights how a lack of clarity obscures data and will lead to confusion and delay.

The first scatterplot is clean and insightful, while the bad one is cluttered and provides little to no insight.

Ethical Considerations

Visually appealing charts are nice, but data visualization is more so about representing data in an honest and ethical way. Whether intentional or not, misleading visuals can have real world consequences. Distorted data, such as truncated y-axis, or cherry-picked data can spread false narratives and cause poor decision making to occur. Any audience member who uncovers that they have been misled would lose trust in the visualization creator and their organizations they represent. For fields like finance or healthcare, misleading visualizations can alter people's lives and may result in jail time for the misleader.

Conclusion

This project allowed the exploration of effective and ineffective data visualizations. We highlighted the impact of design choices and flaws and how they influence the audience's understanding of information. Our good figures followed suitable practices – they ensure everything is clear, accurate, and easily understandable. The bad ones we made are the exact opposite – they hide information, limit understanding, and obscure the truth of the data to suit some other goal.

This report has explained the ethics that follow data analysts and require them to present information honestly and fairly. Bad visualizations, whether made terribly on purpose or by error, contribute to the spread of misinformation and will inevitably result in real world consequences. By focusing on the best practices for creating plots and visuals, it enables analysts to be mindful of how their data is being presented and ensures that visualizations may remain a useful tool for the ethical sharing of information.