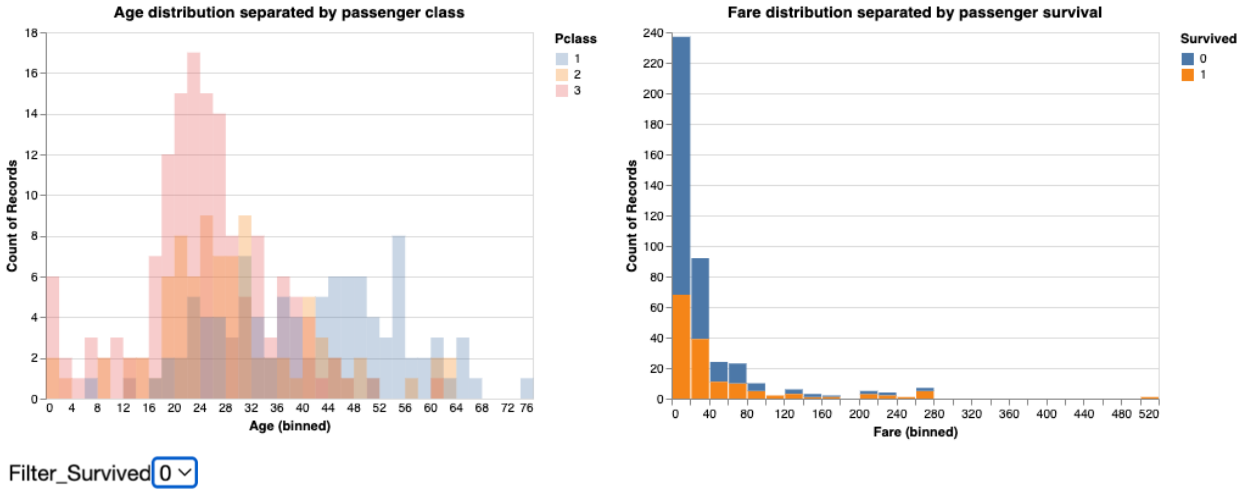
Part 4:

The dataset I chose is the "Titanic Dataset" from Kaggle [1]. This dataset contains various demographic information about the passengers on the Titanic and whether they survived or not. I chose this dataset because I was already familiar with it from a previous course, and I thought that the data could yield some interesting findings. From the layered histogram (left), it appears that first class passengers were a bit older than the other two classes, even some ranging above the 60s. Meanwhile, third class passengers were typically younger than the others. This makes sense historically since wealthy elderly were more likely to be able to afford traveling on the luxurious Titanic than their less wealthy counterparts. Additionally, from the stacked histogram (right), it appears that as fare increased, the percentage of passengers who survived increased as well.

Part 5:



From the dataset, I knew that I wanted to utilize the age, class, fare, and survival data. Because age and fare were both quantitative data, I thought that histograms were the best way to visualize their relationships with class and survival data, which are both categorical.

Part 6:

Both of my charts utilize color to encode data. In the layered histogram, the color distinguishes between the passenger classes, and in the stacked histogram, the color distinguishes between passenger survival. Both data types are categorical, and color hue is one of the best channels to encode categorical data.

Part 7:

The layered histogram features an interactive legend that allows for multiple selections. Because the layers overlap, I thought it would be useful to select a single layer at a time to better distinguish each class. Next, the stacked histogram uses a tooltip that shows the survival status and the number of passengers in that group. I chose this because in a stacked bar, the count of the bottom bar can be estimated, but the top portion is harder to estimate. With the tooltip, the user can see the exact number of passengers that either survived or didn't when they hover over the bars. Lastly, the dropdown filter selects a survival status and both charts transform to fit that

criterion. Showing only the passengers who either survived or didn't can allow for better analysis of the age and fare distributions.

Github Repo:
https://github.com/XinwenZhang216/DS4200_Xinwen_Zhang

Sources:
https://www.kaggle.com/datasets/brendan45774/test-file