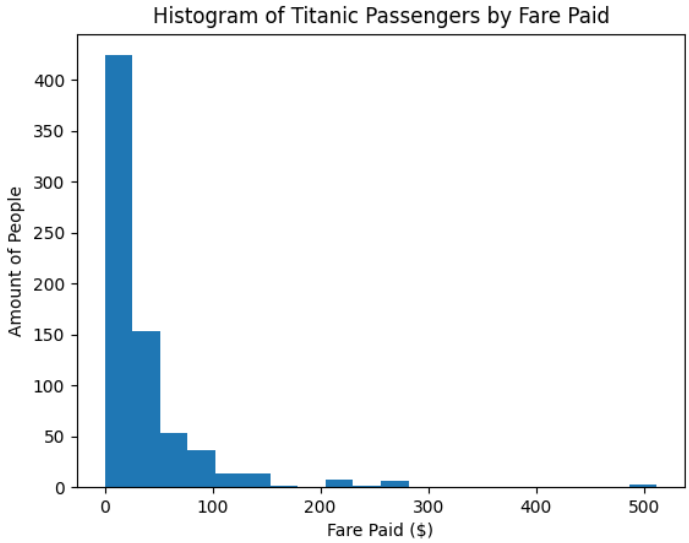
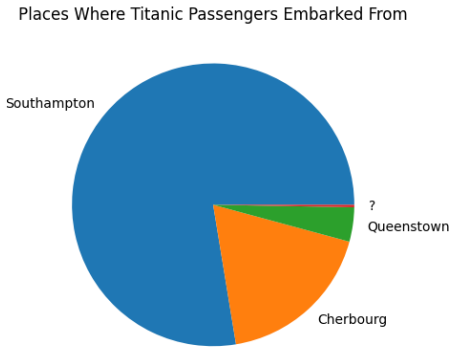
Titanic Data Analysis

This project aspires to analyze a dataset based around the passengers of the Titanic. It will try to answer questions about relations between particular factors and how they might relate to one another. Over the course of the analysis, models will be provided as an example of how the data relate to each other.

The first graph is a histogram that will display the distribution of how much each passenger paid as fare for this particular trip.

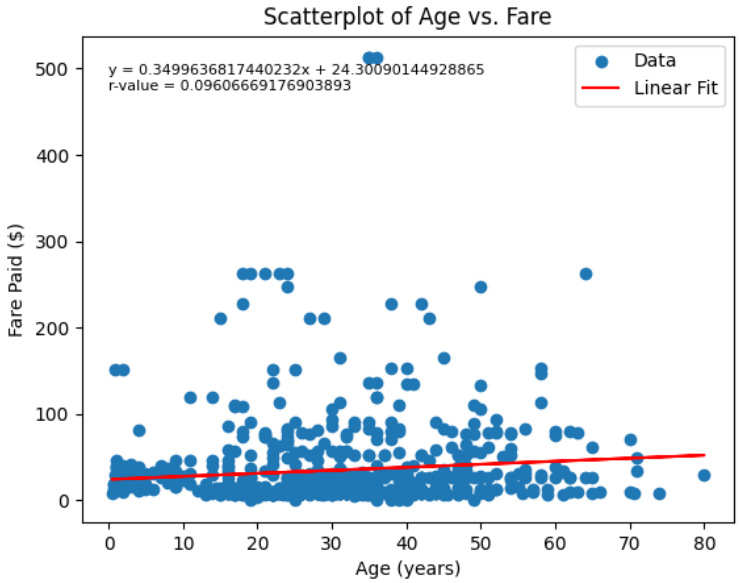


This model is significantly skewed to the right, as many more people paid a lower amount than those who paid an exorbitantly high amount. After the histogram the next thing to be analyzed is a pie chart that shows the proportion of how many people departed from each port.

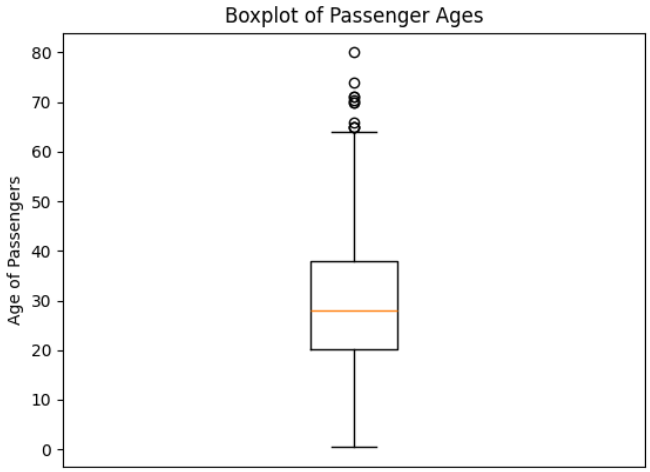


As should be evident, the majority of the passengers embarked from the port in Southampton while smaller amounts embarked from Cherbourg and Queenstown. Next is the scatterplot of age

vs. fare, with a line for linear fit.



There is not much of a pattern between age and fare, as shown by how small of values the slope and r-value are, since those are indicators of correlation. Next I was asked to provide another model of my choice, and a boxplot seemed the most useful.



This shows that more than 75% of the viable data is below the age of 40. Next thing I made was a contingency table for the purpose of performing a chi-squared test.

Sex	female	male
Embarked		
?	2	0
C	61	69
Q	12	16
S	186	368

The p-value that resulted from the chi-squared test was around 0.007, since it is so close to zero, this means that it is significant. There are not many numerical categories in the data set, really the only purely numerical data points are the age and the fare, so those are the data points that can compare the most in depth.

After all of this analysis I hope that it's obvious the depth that data sets can have. This particular data frame is particularly difficult due to its excess of categorical data and deficit of numerical data.