

CISC 333

## Assignment 1: Titanic Analysis

Name: Sam Ruscica

Student Number: 10148585

Course: CISC 333

Date: September 22, 2017

## Titanic Dataset Analysis

A simple analysis was conducted on the Titanic Dataset. It looks at several different aspects of the data including: the means, standard deviations, max and mins, and how the different categories of data might correlate to each other. Below are the different results of the testing done using KNIME.

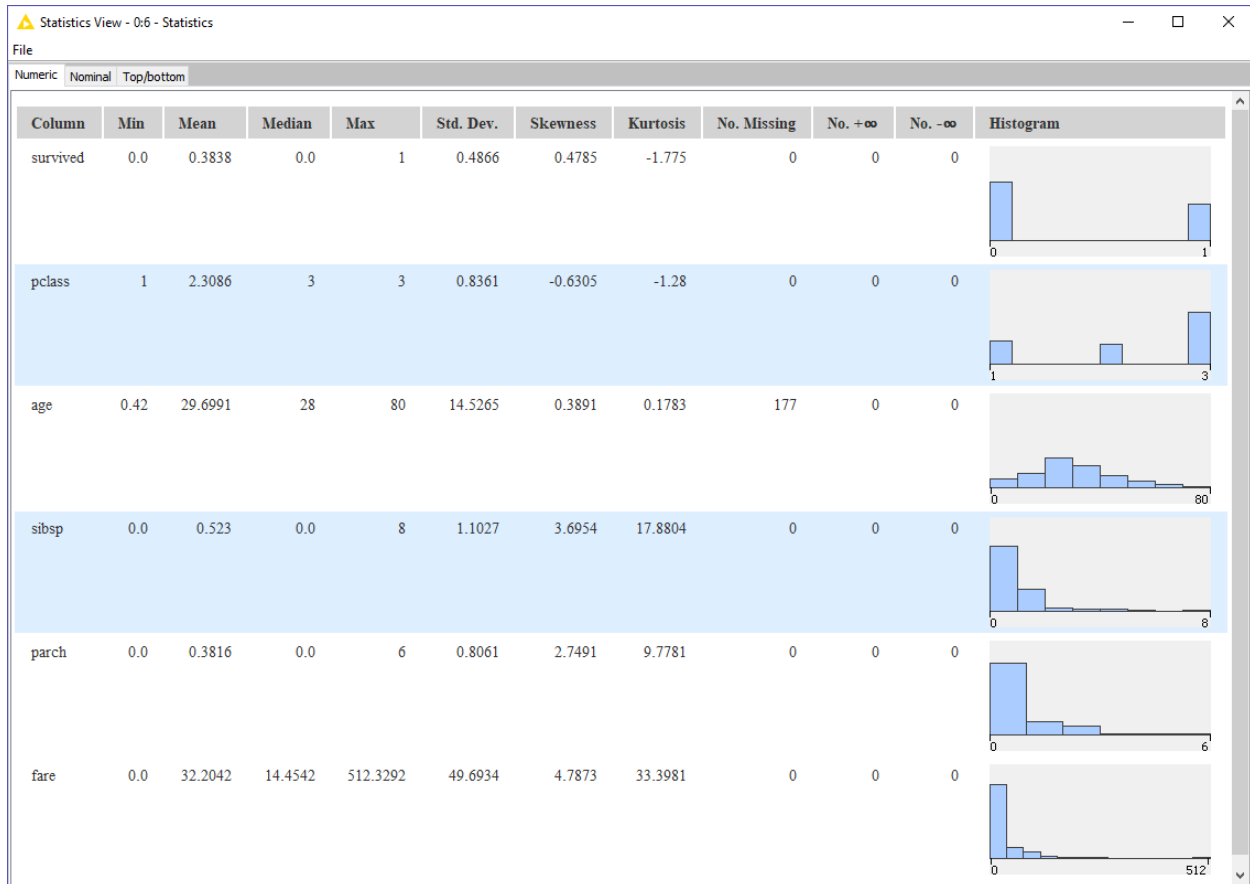


Figure 1: Statistics of the Dataset

Above is information gathered by the statistics node. The statistics node in KNIME shows information about the different individual categories. The important information to look at depends on what the category is. For instance, knowing the min and max values for the survived category reveals no important information but the mean shows that only 38% of people survived the tragedy. From the pclass category, it shows that most people were part of the third class. The age category shows that the average person was around 30 years old but there was a lot of missing information with regards to age. Lastly the fare category shows that the average person paid \$32.20 and it had a standard deviation of about 50. This deviation shows that the value for each ticket changed greatly between who was buying it.

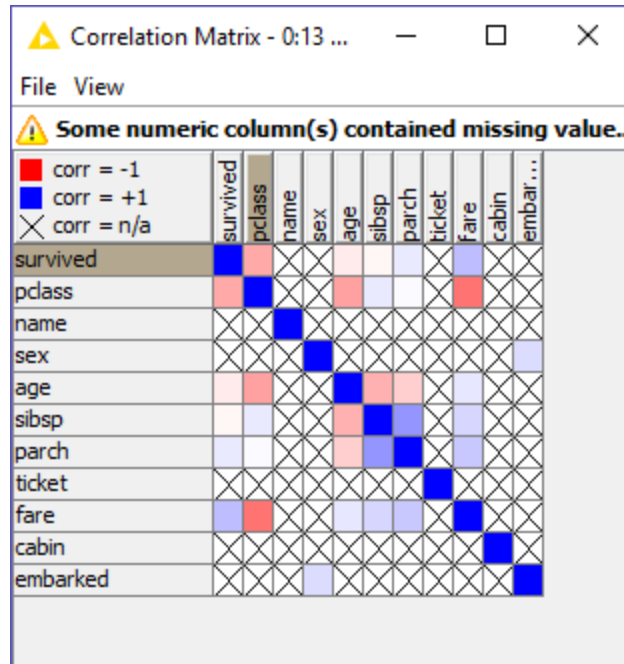


Figure 2: Linear Correlation Between All Categories

This figure shows the different linear correlations between the categories. The darker the color, the stronger they relate. There is several important areas to look at in the data. Firstly, it shows that there is a correlation between class and survival. It also shows that people that paid more were more likely to survive. Additionally, it is important to notice that sex and survival do not correlate in the graph. This is most likely caused by the fact that survival is a numerical value vs sex which is a string value. Interestingly, there is a slight correlation between the place embarked and the sex of the passenger.

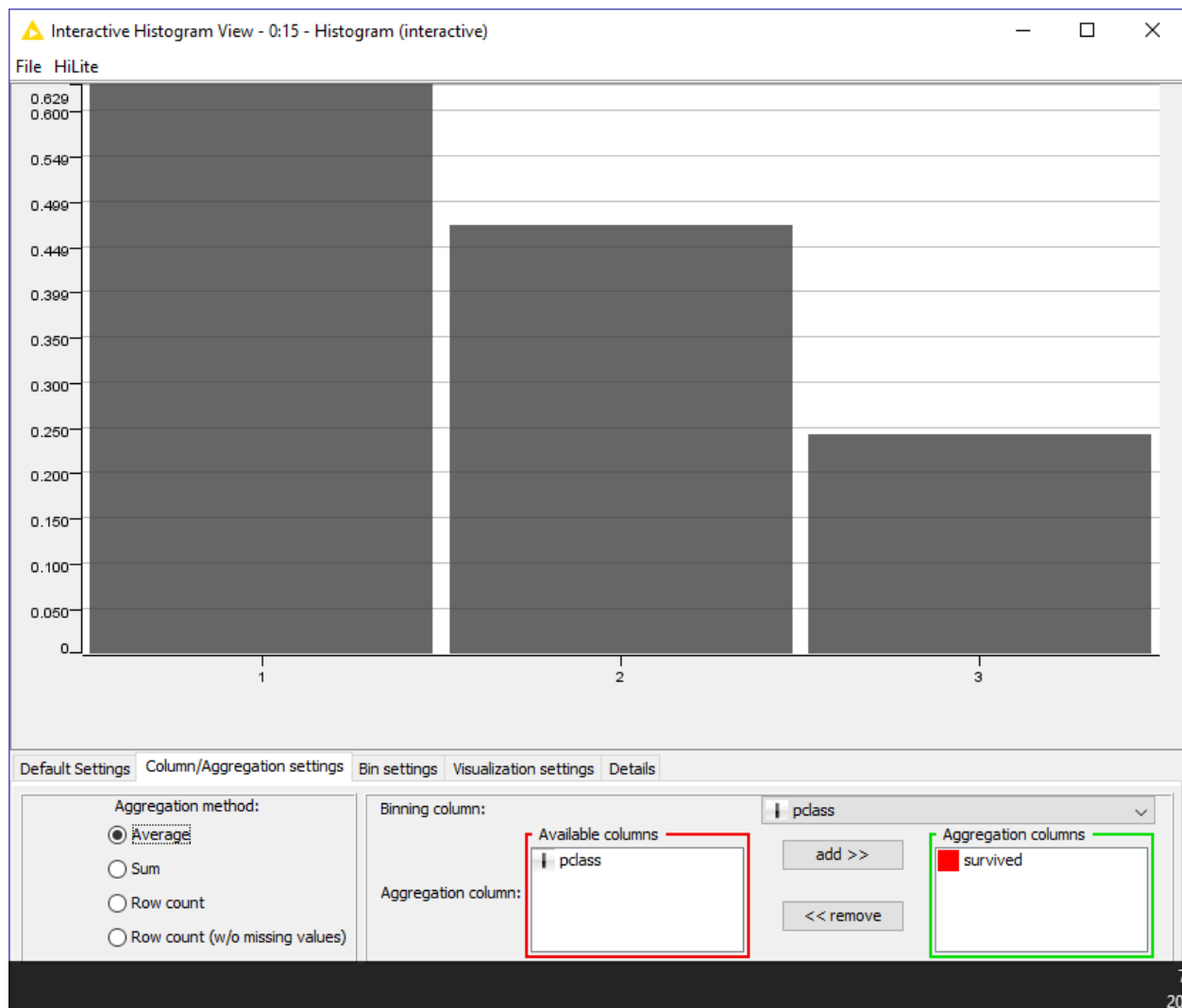


Figure 3: Mean of Survival Vs. Class

This histogram shows the average between the number of people in each class and if they survived. This shows that first class passengers were the best off at 63%. This drastically drops off as through second and third class.

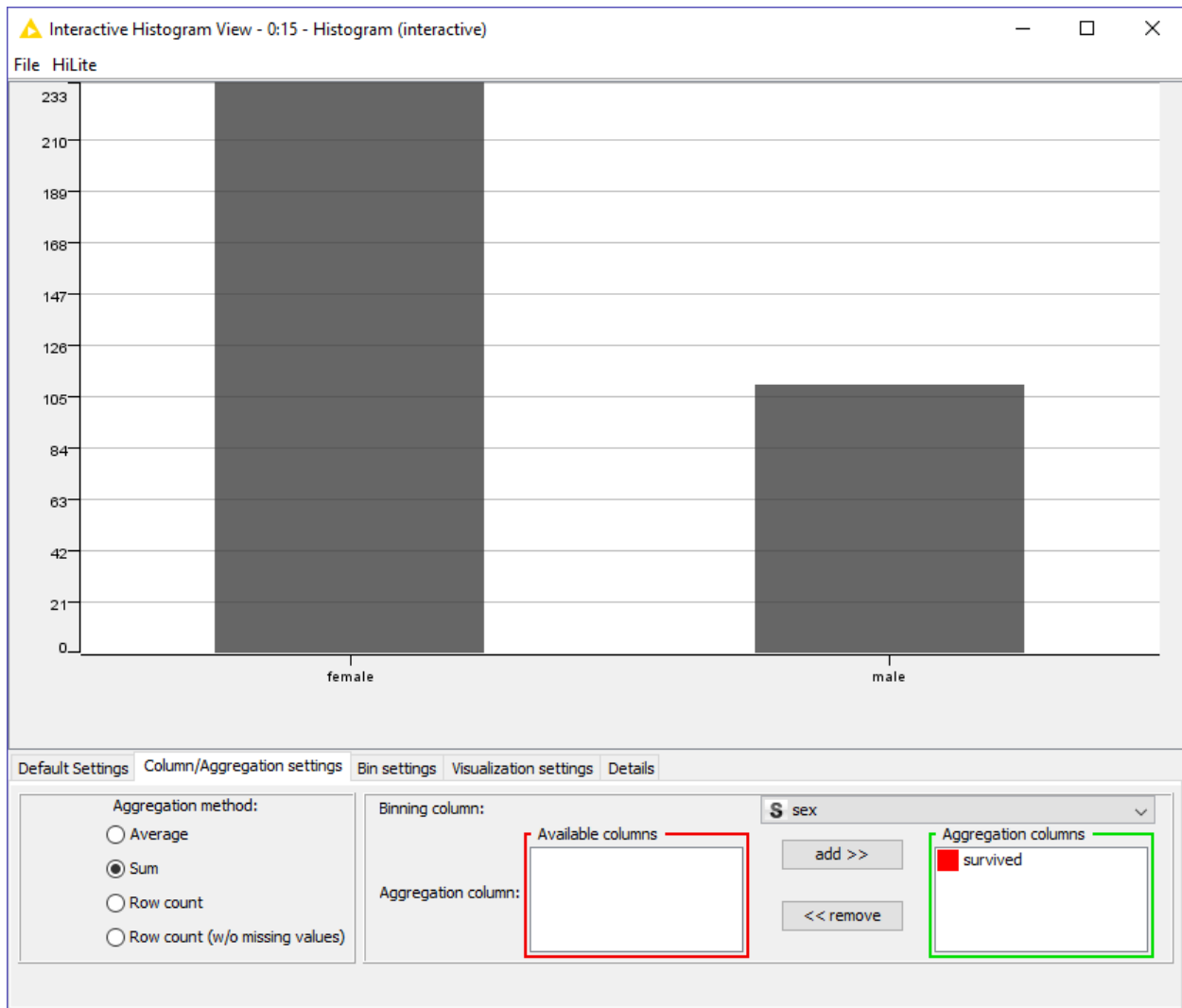


Figure 4: Survival Count of Female Vs. Male

Row ID	count
female	314
male	577

Figure 5: Number of Females and Males

Above is a histogram and count showing the number of females that survived Vs. the number of males that survived. This shows that even though there are fewer females, far more of the survived.

Row ID	count
?	177
0	601
1	113

Row ID	count
?	177
0	395
1	319

Row ID	count
?	177
0	494
1	220

Above on the left, this is a count showing the number of people under the age of 18 (children) as a one and adults a zero. Above in the middle, shows the number of women and children as one and men as zero. Above on the right, shows the number of women and children that survived as one and everyone else as zero. This shows that nearly 69% of women and children survived the sinking of the Titanic. It is important to note that there are 177 people that did not have age information about them. This is a large amount of people and could result in very different numbers.

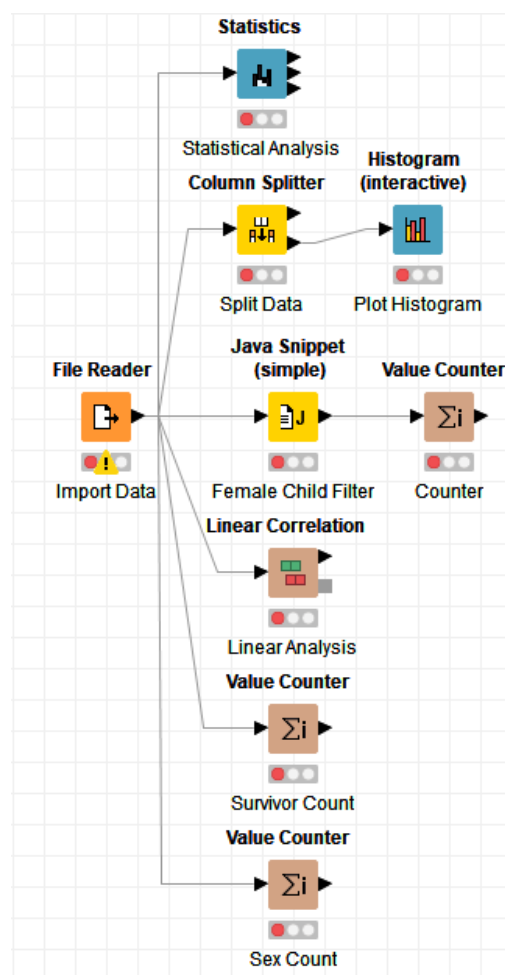


Figure 6: KNIME Nodes Used for Analysis