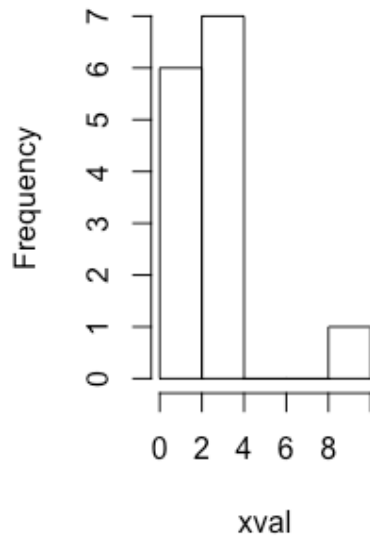


assignment1Stat

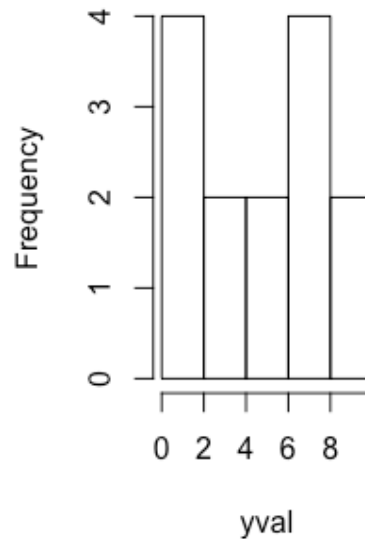
I pledge my honor to abide by the stevens honor society

R Markdown

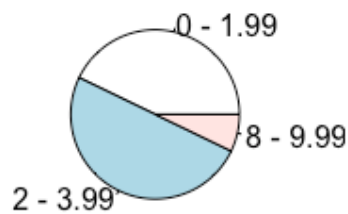
Histogram of X



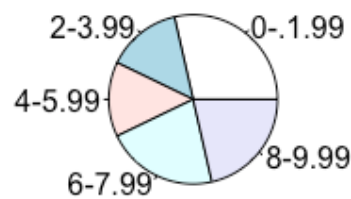
Histogram of Y



Pie Chart of X



Pie Chart of Y



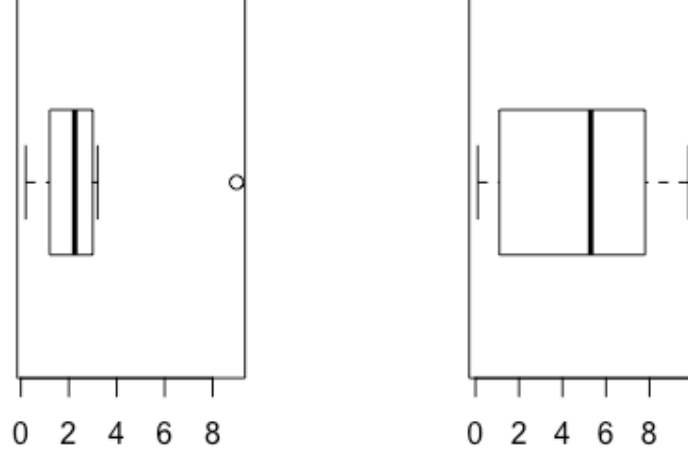
The Y values have a bimodal distribution (mean and median will be close and centered at the spread), and spread from .1 to 9.8. The most values were from 0 to 2 and 6 to 8 both at 4. The X values are right skewed (the mean will be greater than the median), and spread from .2 to 9. The most values were for 2 - 3.99 and a potential outlier of 9.

Boxplot of X



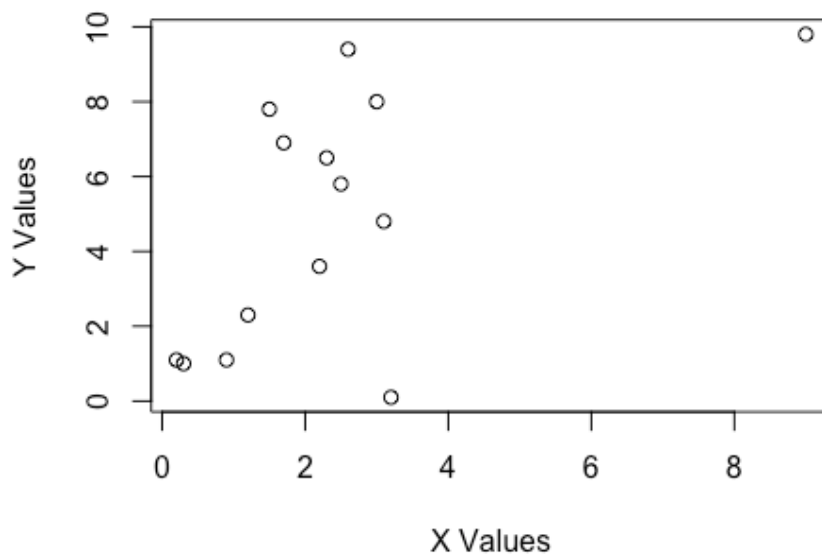
Boxplot of Y





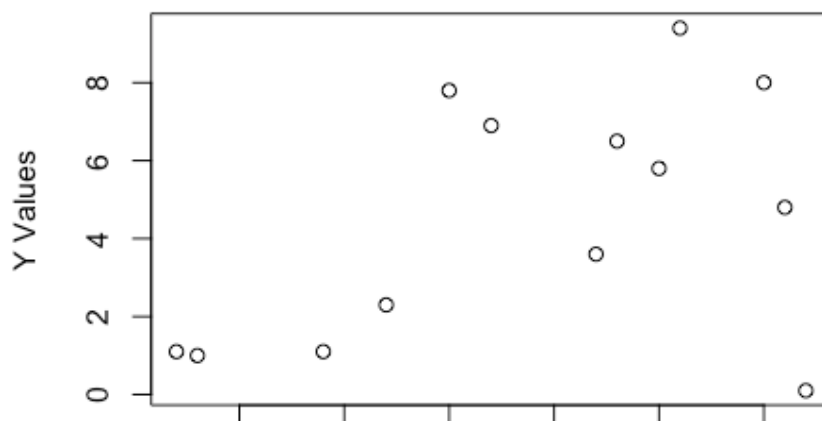
The five number summary of Y is .1, 1.4, 5.3, 7.575, 9.9 with variance 11.17143 and no outliers. X has a five number summary of .2, 1.275, 2.25, 2.9, 9 with variance 4.568407 and an outlier for 9. This is because $1.5 \times (2.9 - 1.275)$ which is 2.4375 and $2.4375 + 2.9$ is 5.3375 which is less than 9

Scatterplot



The correlation coefficient is 0.5679. There is a positive moderate relationship between the x and y values.

Scatterplot without the Outlier

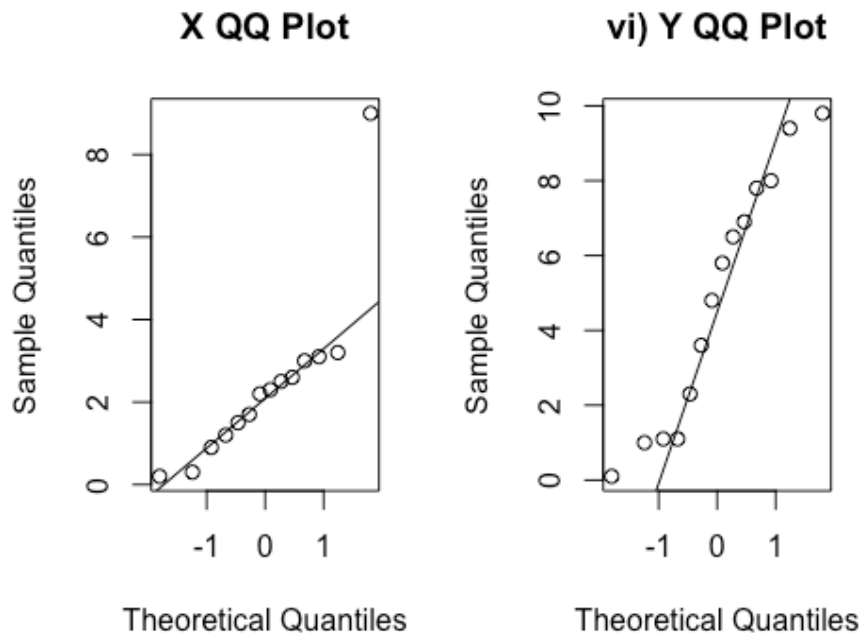


0.5 1.0 1.5 2.0 2.5 3.0

X Values

There was an outlier of the point (9, 9.8). This point is close to the best fit line, but its way to far away from the general area of other points. I would say the X and Y relationship is mild. The new correlation coefficient was 0.4586256.

The correlation coefficient decreases after taking out the outlier, which means there is a loss of relationship between the X and Y values.



The X values are more normally distributed as the values seem to follow the best fit line, except for the outlier. The Y values are decently distriobuted well, more than the X. IF we were to take out the X value that was defined as an outlier, I would say that the X is better distributed.