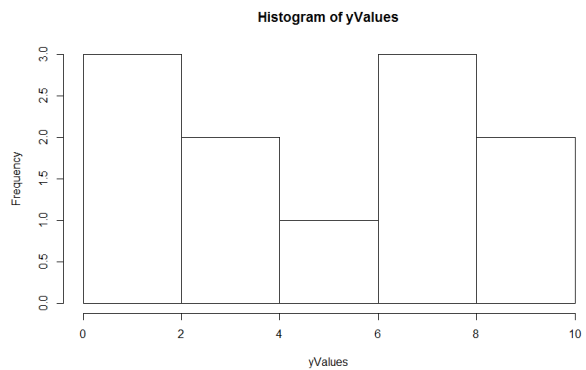
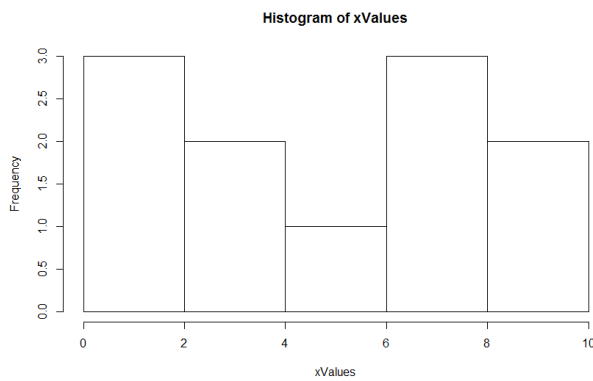


Adam Gincel

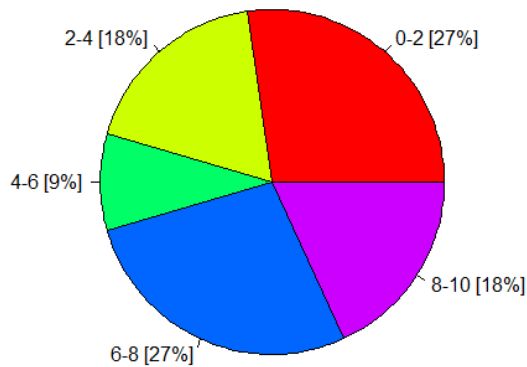
MA331

## Assignment 1

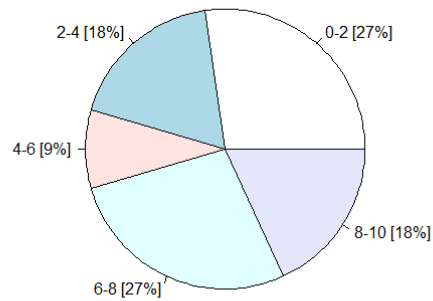
1)



**Pie Chart of X Values**

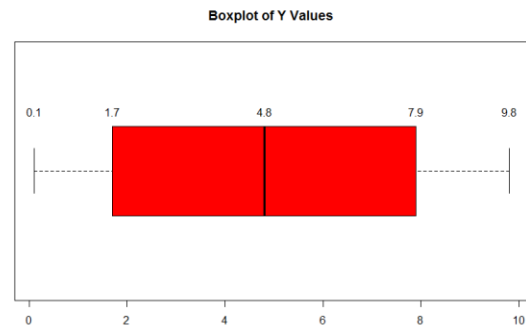
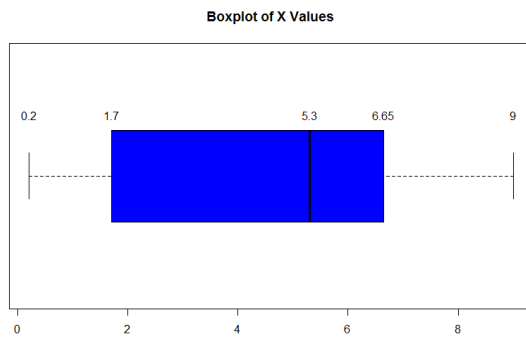


**Pie Chart of Y Values**



X and Y are similarly distributed. Their histograms look identical, and their corresponding pie charts are also distributed the same way. Y has a wider range, but both Y and X seem to have very similar spreads – the ranges do not differ by more than 1 on either end of the spectrum.

2)

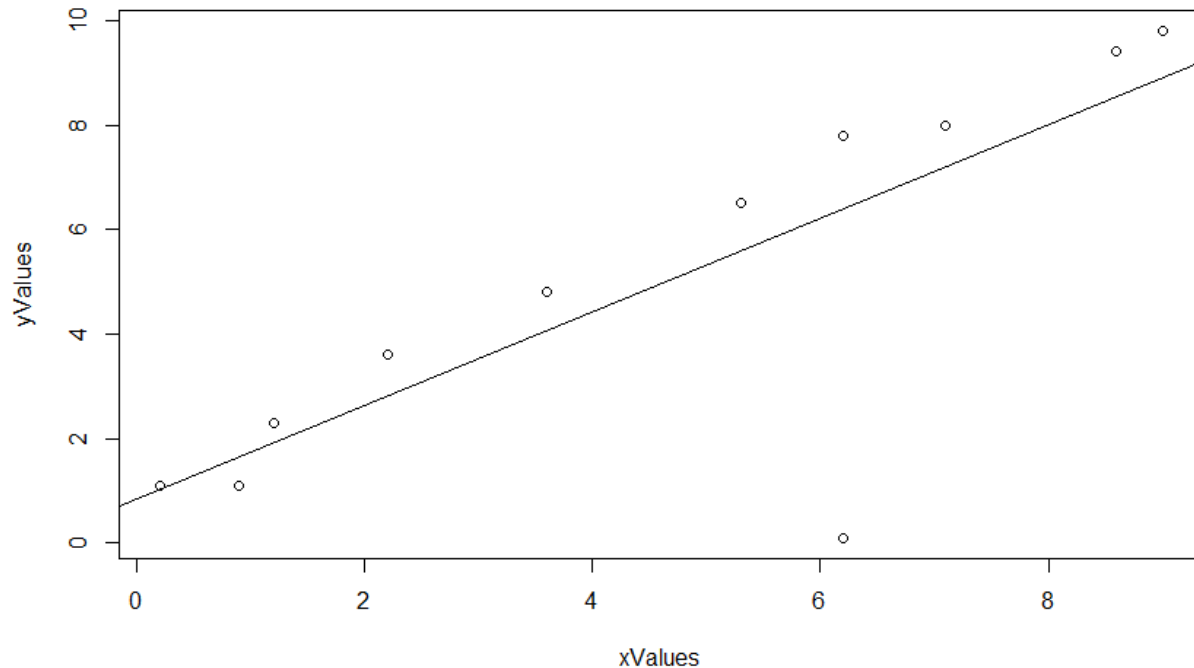


X Variance: 9.858909

Y Variance: 12.51873

The univariate method for detecting outliers is to see if any values lie outside of the range of  $1.5 \times (75^{\text{th}} \text{ percentile} - 25^{\text{th}} \text{ percentile})$ . No values lie outside of this range in either box plot. As such, there are no statistical outliers in x or y.

### 3) Scatter plot:



- Correlation Coefficient: 0.7942342
- Regression line:
  - Intercept: 0.8458, Slope: 0.895
- X and Y are moderately associated; a correlation coefficient of ~0.79 is moderate to highly correlated.

4) (6.2, 0.1) is an outlier on this graph, which is very obvious by simply looking at our plot's regression line. Removing it from our data yields a Correlation coefficient of 0.9932443. By removing our single outlier, which was only an outlier when viewed in the context of (x, y), we have brought correlation up from moderately correlated to almost 100% correlated.