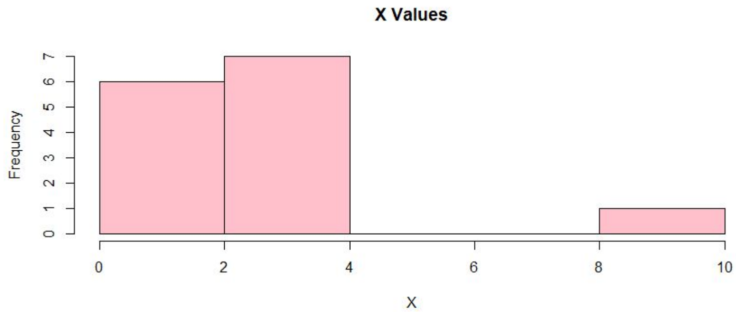
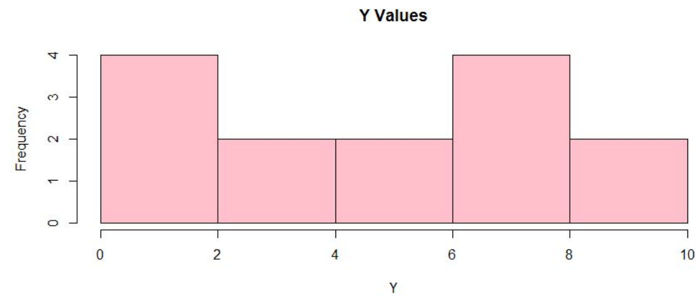
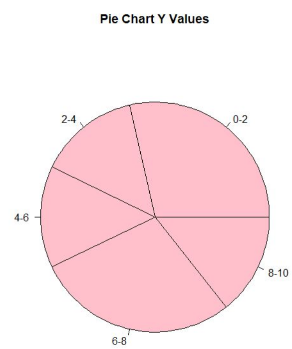
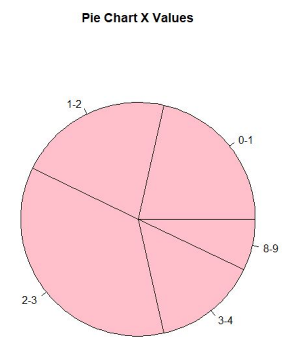
Constance Xu
MA331 - Homework 1 - I pledge my honor that I have abided by the Stevens Honor System.

Problem 1





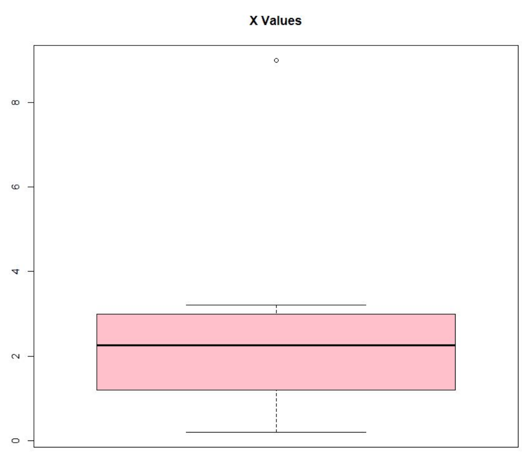


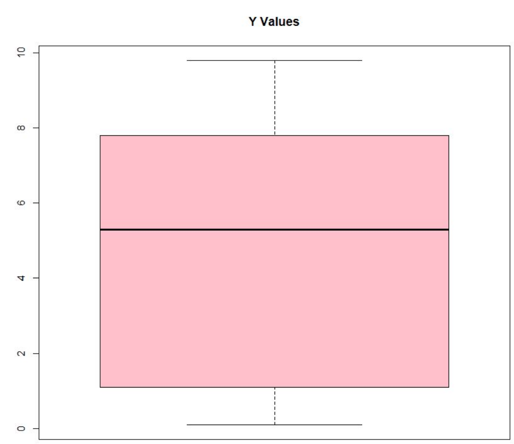


(i) X-values: For the x-values histogram, we can see that it is more right-skewed. There is an outlier in the x-values between 8 and 10. There is a median value of 2.25 and this is roughly the center of the histogram. The minimum value is 0.2 and the maximum value is 9.0.

Y-values: It is not symmetrical because of the values between 0-2 and 6-8. There are no outliers in this graph and the median of the graph is 5.3. The minimum value is 0.1 and the maximum value is 9.8.

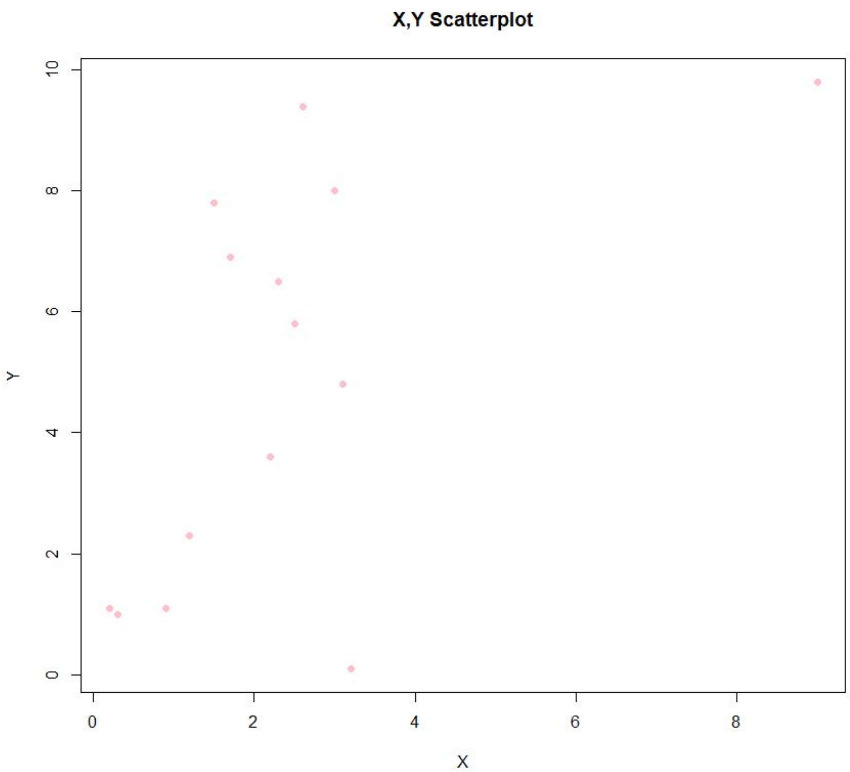
(ii) There are no outliers for the y-values but for the x-values, there is an outlier around 9. This can be seen in the X Values chart as the circle above the main part of the diagram. The five number summaries are shown below the two box-plots.





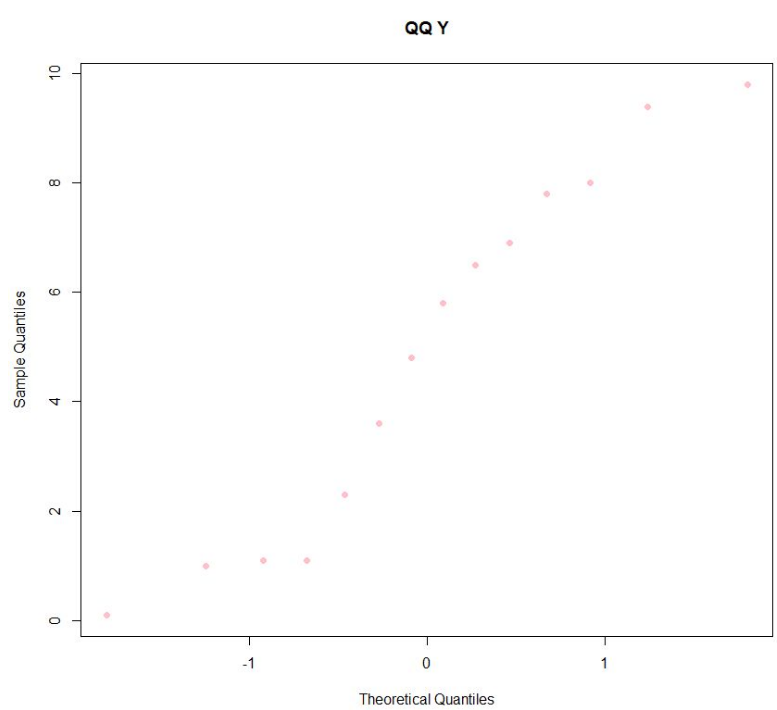

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.200   1.275   2.250   2.407   2.900   9.000
> summary(y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.100   1.400   5.300   4.871   7.575   9.800
```

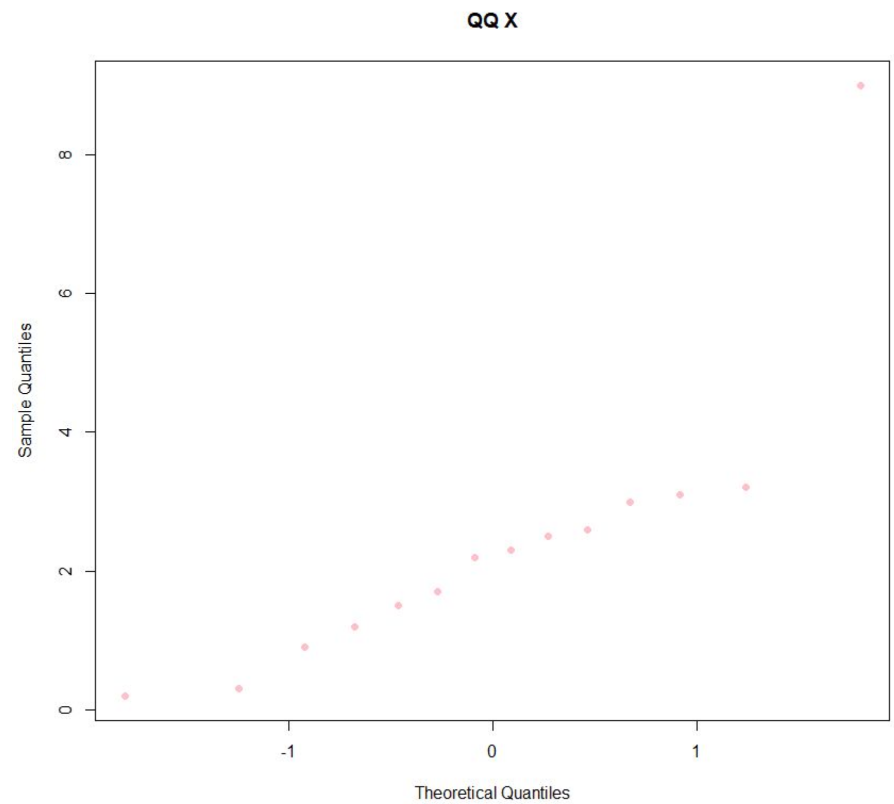
The variance of x is 4.568407 and the variance of y is 11.17143.



(iii) The correlation coefficient of (X,Y) is 0.5679153. This is a somewhat strong correlation and it is a positive correlation.
(iv) The one outlier that I chose is (9.0,9.8) for this (X,Y) scatter plot. After removing this outlier, I have found that the new correlation coefficient is 0.4586256.
(v) The correlation with the removal of the outlier has made the correlation weaker, but it is still positive.

(vi) Beneath this, it can be noted that the graph labeled "QQ X" is the QQ plot for X and "QQ Y" is the QQ plot for Y. By adding in the QQline to this, we can see that Y has more of a normal distribution than X.





The following is the R code used to find stuff for problem 1:

```
x <-c(0.2,1.2,0.9,2.2,3.2,0.3,1.7,3.1,2.3,1.5,2.5,3.0,2.6,9.0)
x2<- c(0.2,1.2,0.9,2.2,3.2,0.3,1.7,3.1,2.3,1.5,2.5,3.0,2.6)
y <- c(1.1,2.3,1.1,3.6,0.1,1.0,6.9,4.8,6.5,7.8,5.8,8.0,9.4,9.8)
y2<- c(1.1,2.3,1.1,3.6,0.1,1.0,6.9,4.8,6.5,7.8,5.8,8.0,9.4)

hist(x,main="X Values",xlab="X",col="pink", border="black", plot=TRUE)
hist(y,main="Y Values",xlab="Y",col="pink", border="black", plot=TRUE)

s <- c(3,3,5,2,1)
labs <- c("0-1","1-2", "2-3", "3-4", "8-9")
pie(s, labels=labs, main="Pie Chart X Values", col="pink", border="black", plot=TRUE)
s <- c(4,2,2,4,2)
labs <- c("0-2", "2-4", "4-6", "6-8", "8-10")
pie(s, labels=labs, main="Pie Chart Y Values", col="pink", border="black", plot=TRUE)

mean(x)
median(x)
mean(y)
median(y)
```

```
boxplot(x, main="X Values", col="pink", border="black", varwidth=TRUE, plot=TRUE)
boxplot(y, main="Y Values", col="pink", border="black", varwidth=TRUE, plot=TRUE)

plot(x,y,main="X,Y Scatterplot", xlab="X",ylab="Y", col="pink", pch=19)
summary(x)
summary(y)
var(x)
var(y)
cor(x,y)

qqnorm(x, main="QQ X", col="pink", pch=19)
qqline(x)
qqnorm(y, main="QQ Y", col="pink", pch=19)
qqline(y)
```

Problem 2

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\bar{x}^2$$

$$= \sum_{i=1}^{n}x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum_{i=1}^{n}x_i^2 - n\bar{x}^2$$

through substitution

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - n\bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}\frac{n}{n}\bar{x}^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}\bar{x}^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$