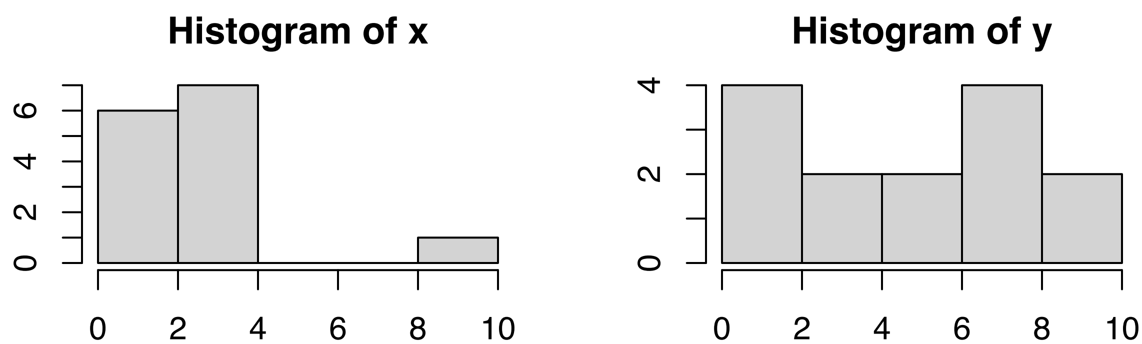
MA331 Homework 1 - Aparajita Rana

I pledge my honor that I have abided by the Stevens Honor System.

```
x <- c(0.2,1.2,.9,2.2,3.2,.3,1.7,3.1,2.3,1.5,2.5,3,2.6,9)
y <- c(1.1,2.3,1.1,3.6,.1,1,6.9,4.8,6.5,7.8,5.8,8,9.4,9.8)
x.data <- data.frame(x)
y.data <- data.frame(y)
```

```
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
hist(x)
hist(y)
```

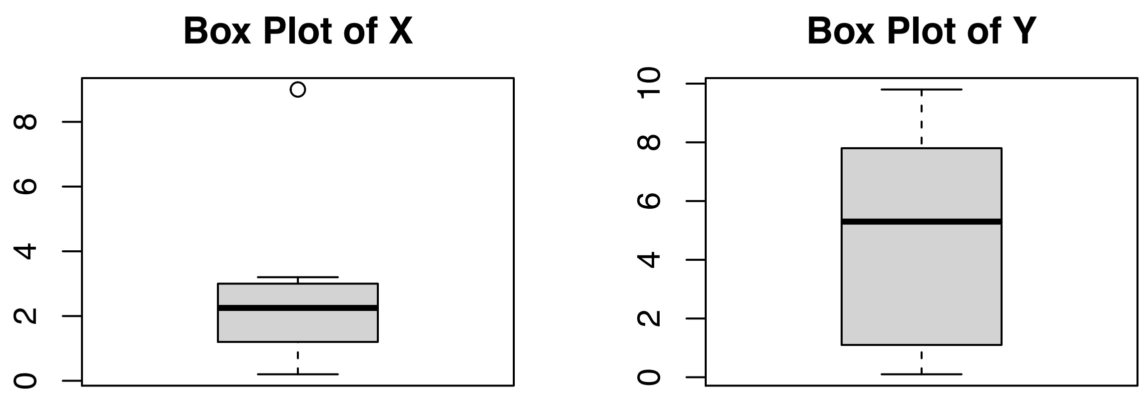


The Histogram of the X values is skewed towards the lower end of the spectrum with most values between [0-4). The Histogram of Y is more distributed with 2 major peaks between [0-2) and [6,8).

```
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
pie(table(x_new),main="Pie Chart of x")
pie(table(y_new),main="Pie Chart of y")
```

The Pie Chart of X values shows that half of the values are between (2,4] and the Pie Chart of Y is more evenly distributed.

```r
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
boxplot(x, main="Box Plot of X")
boxplot(y, main="Box Plot of Y")
```



Summary and Variance of X: (5 vals)

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200   1.275   2.250   2.407   2.900   9.000
```

```
## [1] 4.568407
```

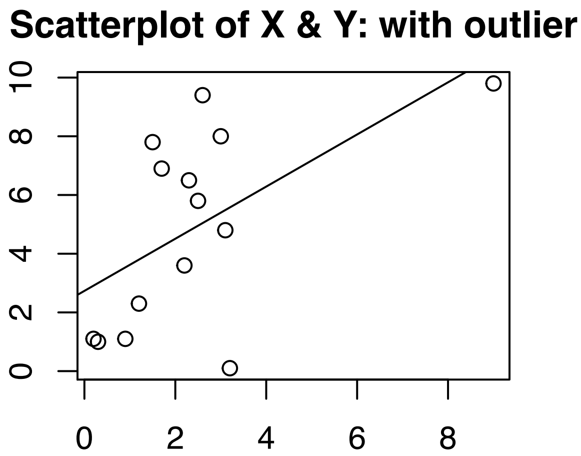
Summary and Variance of Y: (5 vals)

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.100   1.400   5.300   4.871   7.575   9.800
```

```
## [1] 11.17143
```

There is an outlier in the Box Plot of X which is 9.0 and none in Blox Plot of Y.

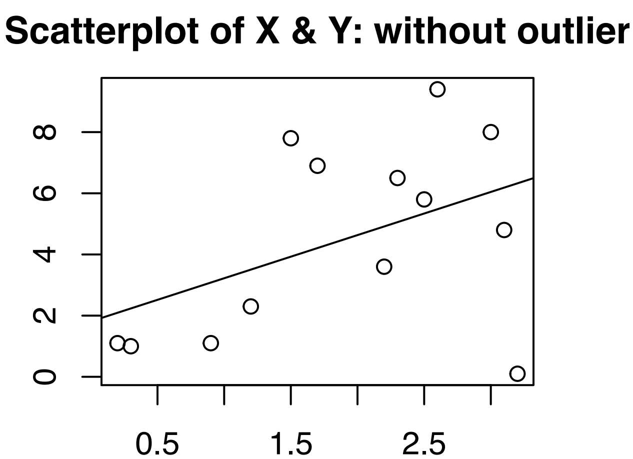
```r
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
plot(x,y, main = "Scatterplot of X & Y: with outlier")
abline(lm(y~x))
```



Correlation coefficient of X&Y with "pearson, kendall, and spearman" method:

```
## [1] 0.5679153
```

```
x <- c(0.2,1.2,.9,2.2,3.2,.3,1.7,3.1,2.3,1.5,2.5,3,2.6)
y <- c(1.1,2.3,1.1,3.6,.1,1,6.9,4.8,6.5,7.8,5.8,8,9.4)
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
plot(x,y, main = "Scatterplot of X & Y: without outlier")
abline(lm(y~x))
```



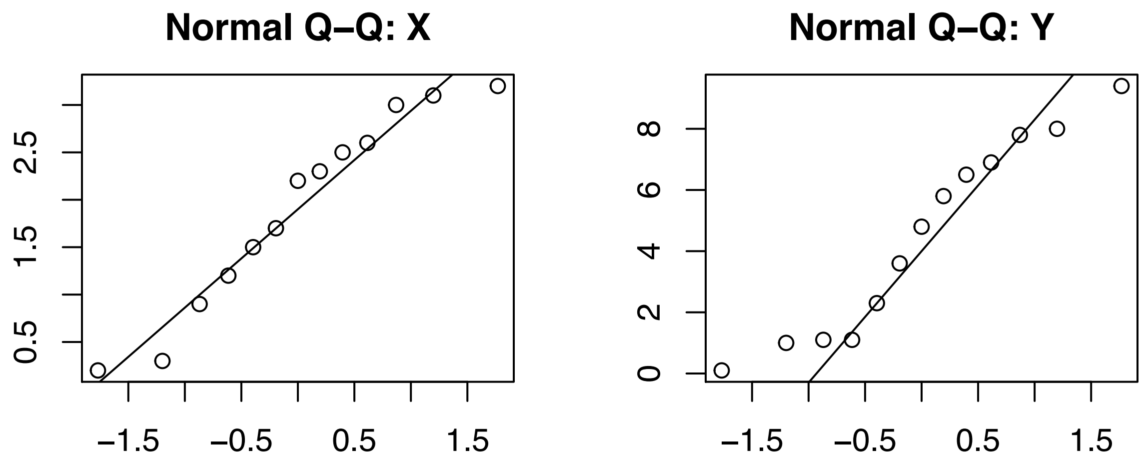
New correlation coefficient of X&Y with "pearson, kendall, and spearman" method:

```
## [1] 0.4586256
```

We see a moderate positive relationship linear relationship based on the line and >0 correlation. After removing the outlier of (9.0,9.8) we see a weak positive relationship. We saw the correlation actually went down after the removal of what we thought was the outlier. We can conclude there is a positive relationship.

```
par(mfrow=c(1,2), mai = c(0.5, 0.5, 0.5, 0.5))
qqnorm(x, main="Normal Q-Q: X")
qqline(x)
```

```r
qqnorm(y, main="Normal Q-Q: Y")
qqline(y)
```



Although the q-q normal plots look fairly similar we must consider the scale and the box plots and histogram we made. Accordingly, I predict the Y is more normal because of a more even distribution.

Problem 2:

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

In this proof we can see how we can prove the first part of the problem. Accordingly, this shows the second portion as well because 1/n is treated as a type of constant.