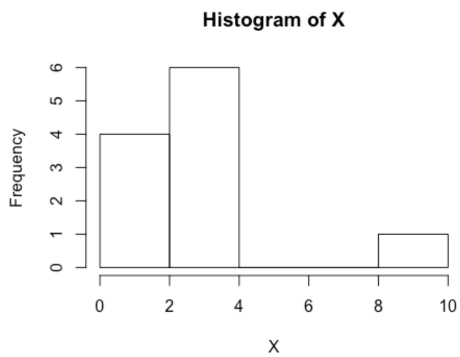
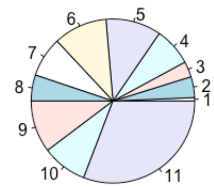
HW 1

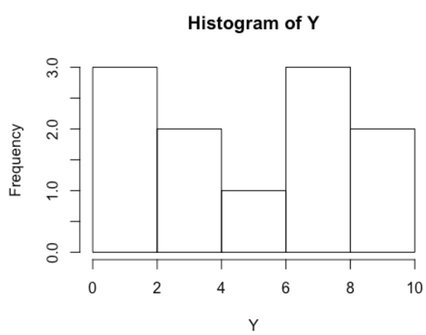
Tyler Sulsenti

```
X <- c(0.2,1.2,0.9,2.2,3.2,3.1,2.3,1.5,3.0,2.6 ,9.0)
Y <- c(1.1, 2.3 , 1.1 , 3.6 , 0.1 , 4.8 , 6.5 , 7.8 , 8.0 , 9.4 , 9.8)

#i
hist(X)

pie(X)
```

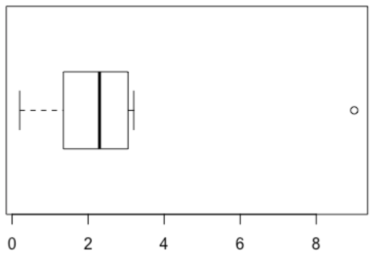




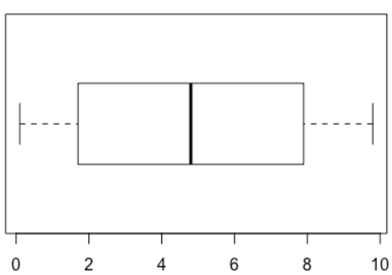

```
#The distrubution of the Historgram of X is Right skewed.
#Most of the data is on the left side of the plot and the tail is on the
right side. The median is 2.3 and we use median becuase the data is skewed.

hist(Y)

pie(Y)
```




```
#The distrubution of the Histogram of Y is a bimodial distrubtion.
#It has two peaks of the same height.

#ii
#Boxplot for X
boxplot(X, horizontal = TRUE)
```

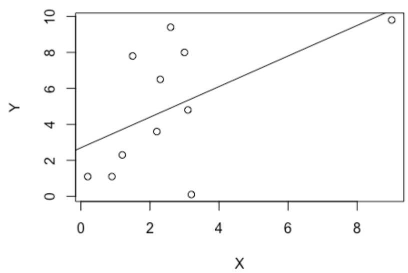



```
#Five number summary for X
summary(X)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200   1.350   2.300   2.655   3.050   9.000

#variance for X
var(X)

## [1] 5.376727

#Outliers:
#For X, there is one outlier and it is 9.0

#Boxplot for Y
boxplot(Y, horizontal = TRUE)
```

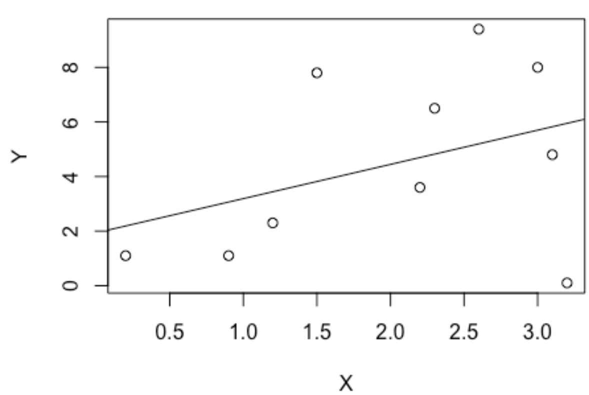



```
#Five number summary for Y
summary(Y)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.100   1.700   4.800   4.955   7.900   9.800

#Variance for Y
var(Y)

## [1] 12.51873
```

```
#Outliers: There are no outliers for Y according to the boxplot

#iii
#Scatterplot of (X,Y)
plot(X,Y)
abline(lm(Y~X))
```



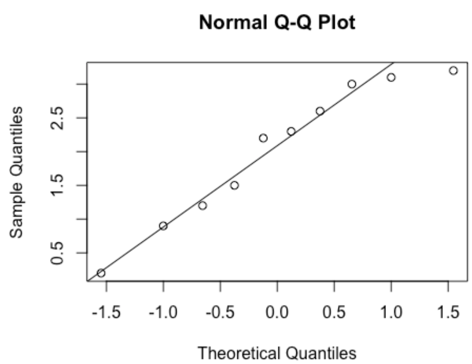

```
#Correlation Coefficient is 0.5571167 based on
cor(X,Y)

## [1] 0.5571167

#Since the correlation is postive, this means that the linear association
between X and Y is positively correlated. Therefore, as X increases, so will
Y

#iv
#Yes there is an outlier at (9.0,9.8)
#Now to remove
X <- c(0.2,1.2,0.9,2.2,3.2,3.1,2.3,1.5,3.0,2.6)
Y <- c(1.1 , 2.3 , 1.1 , 3.6 , 0.1 , 4.8 , 6.5 , 7.8 , 8.0 , 9.4)
#Re-Plot
plot(X,Y)
abline(lm(Y~X))
```



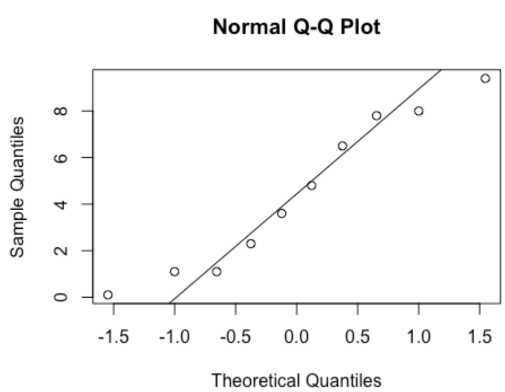
```
#Correlation Coefficient is 0.3873604 based on
cor(X,Y)
```

```
## [1] 0.3873604
```

```
#v
#The difference between iii and iv is that the new correlation coefficient in
iv is lower and closer to 0 than the one in iii. This means that the
relationship between X and Y is less linear than before.


#vi
#normal QQ plot of X
qqnorm(X)
qqline(X)
```




```
#normal QQ plot of Y
qqnorm(Y)
qqline(Y)
```




```
#The data in X is more likely to be of normal distrubtion because the data
represents a more straight diagonal line.
```