hw3-5.1

Mark Pearl

9/11/2019

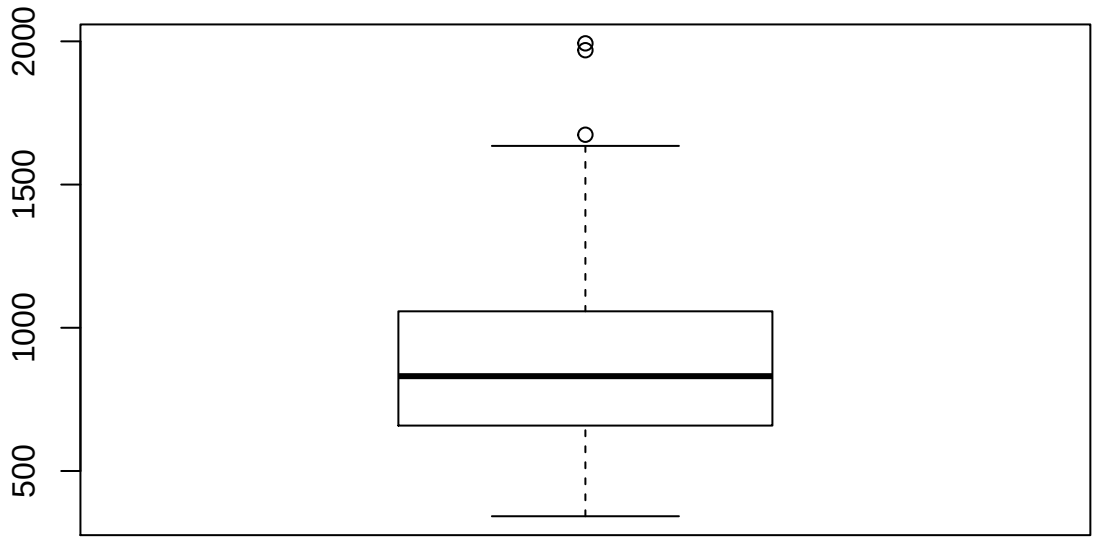
```
uscrime_data <- read.table('C:/Users/mjpearl/Desktop/omsa/ISYE_6501/hw3/data/uscrime.txt',header = TRUE
head(uscrime_data)
```

```
##      M So   Ed  Po1  Po2    LF   M.F Pop   NW    U1  U2 Wealth Ineq
## 1 15.1  1  9.1  5.8  5.6 0.510  95.0  33 30.1 0.108 4.1   3940 26.1
## 2 14.3  0 11.3 10.3  9.5 0.583 101.2  13 10.2 0.096 3.6   5570 19.4
## 3 14.2  1  8.9  4.5  4.4 0.533  96.9  18 21.9 0.094 3.3   3180 25.0
## 4 13.6  0 12.1 14.9 14.1 0.577  99.4 157  8.0 0.102 3.9   6730 16.7
## 5 14.1  0 12.1 10.9 10.1 0.591  98.5  18  3.0 0.091 2.0   5780 17.4
## 6 12.1  0 11.0 11.8 11.5 0.547  96.4  25  4.4 0.084 2.9   6890 12.6
##       Prob    Time Crime
## 1 0.084602 26.2011   791
## 2 0.029599 25.2999  1635
## 3 0.083401 24.3006   578
## 4 0.015801 29.9012  1969
## 5 0.041399 21.2998  1234
## 6 0.034201 20.9995   682
```

Plots Section

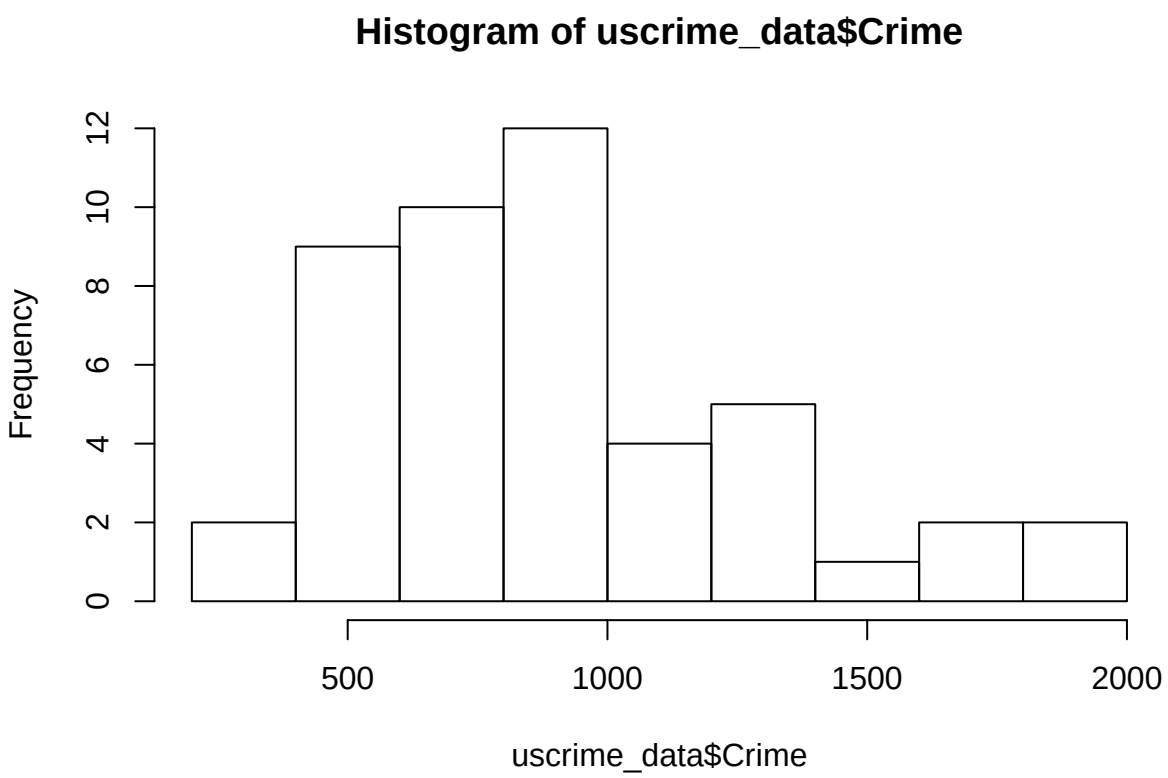
The following plots will conduct exploratory analysis on the data to get a sense of the data's distribution and to see if we can spot any outliers with a visual representation.

```
boxplot(x= uscrime_data$Crime)
```



From the boxplot we can see that there are a few observations above the whisker which indicates values past Q3 are outliers (2 observations closest to a Crime value of 2000). There does not seem to be any observations in the lower quartiles that indicate any outliers.

```
hist(uscrime_data$Crime)
```



The result of histogram indicates a skewed distribution for the right tail. For a grubbs test to be effective it is implied that the data follows a normal distribution. However our data does follow a normal distribution towards the middle portion of the graph, so it could mean that we have outlying data. We will continue with conducting the grubbs test for further investigation.

Grub Test Section

To ensure we test observations represented by the minimum and maximum values on the graph, we will use the "opposite" parameter of the grubbs test function.

```
grubbs.test(x=uscrime_data$Crime, type = 10, opposite = F)
```

```
##
##  Grubbs test for one outlier
##
## data:  uscrime_data$Crime
## G = 2.81287, U = 0.82426, p-value = 0.07887
## alternative hypothesis: highest value 1993 is an outlier
```

```
grubbs.test(x=uscrime_data$Crime, type = 10, opposite = T)
```

```
##
##  Grubbs test for one outlier
##
```

```
## data:  uscrime_data$Crime
## G = 1.45589, U = 0.95292, p-value = 1
## alternative hypothesis: lowest value 342 is an outlier
```

The first output indicates that the highest or maximum observation closest to 2000 on the graph can be deemed an outlier due to the significantly low p-value of 0.07. This holds true to what was also determined in the boxplot output.

The second output indicates that the lowest value of the crime feature is with a high certainty not an outlier as we retrieved a p-value of 1.