Name: Tian Qiu
Instructor: Chris Womble

PART A

1.

```
#PART A
#part1 5
attach(airline_cleaned)
mean(ActualElapsedTime)
sd(ActualElapsedTime)
mean(airline_cleaned$ActualElapsedTime)
sd(airline_cleaned$ActualElapsedTime)

#part2
fivenum(ActualElapsedTime)

#part 4
hist(ActualElapsedTime)
histogram(ActualElapsedTime)
#install.packages("lattice") #Only needs to be run once

#Part3
boxplot(ActualElapsedTime)
means = mean(ActualElapsedTime)
points(means, pch = 18)
bwplot(ActualElapsedTime)
```
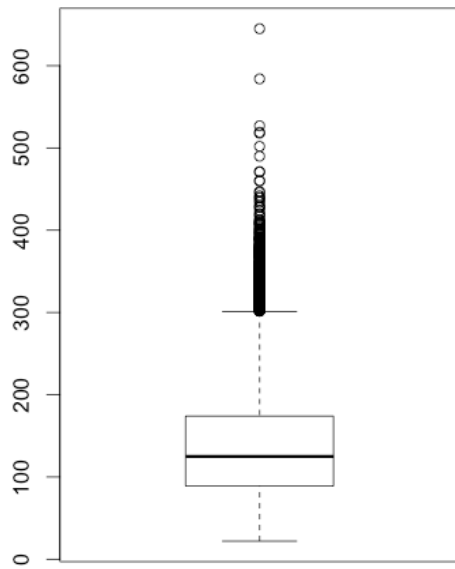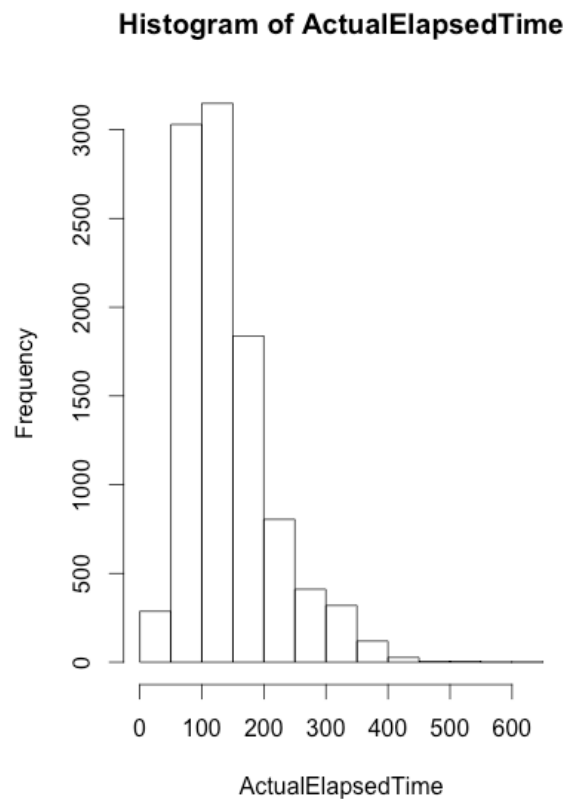
2.

```
> fivenum(ActualElapsedTime)
[1]  22  89 125 174 645
```

3.

It is right skewed. Because right whisker are pretty long.

4.

Histogram of ActualElapsedTime

Histogram is right skewed.

5.
> mean(airline_cleaned$ActualElapsedTime)
[1] 141.4169
> sd(airline_cleaned$ActualElapsedTime)
[1] 72.55718

median = 125

Median is not close to mean. Because there are many outliers which are larger than maximum of the data set. Also the standard deviation are pretty large.

6. I would use median number 125. Because there are so many outliers that mean is larger than it should be. So, median is the best number to estimate actual elapsed time.

PART B
#PART B
#PART 1 5
attach(airline_cleaned)
mean(TaxiIn)
sd(TaxiIn)
mean(airline_cleaned$TaxiIn)
sd(airline_cleaned$TaxiIn)

#PART2
fivenum(TaxiIn)

#PART 4
hist(TaxiIn)
histogram(TaxiIn)
#install.packages("lattice") #Only needs to be run once

#PART 3
boxplot(TaxiIn)
means = mean(TaxiIn)
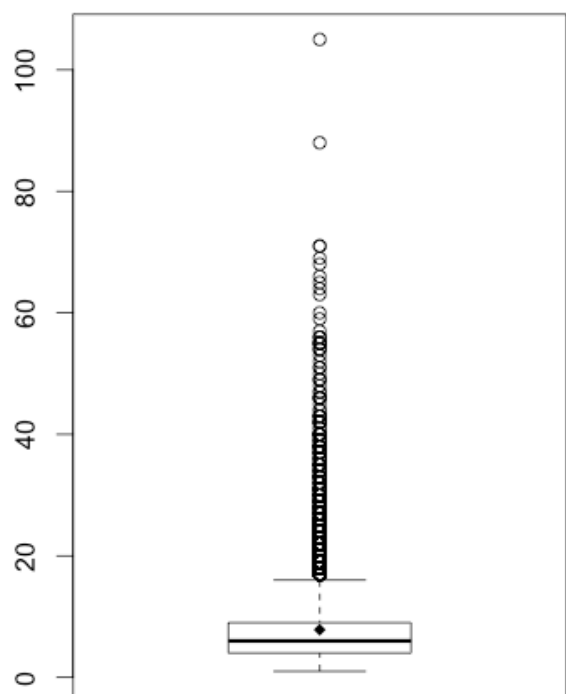points(means, pch = 18)
bwplot(TaxiIn)

2.
> fivenum(TaxiIn)
[1]   1   4   6   9 105

3.
1.5IQR + Q3 = 1.5 * (9 − 4) + 9 = 7.5 + 9 = 16.5
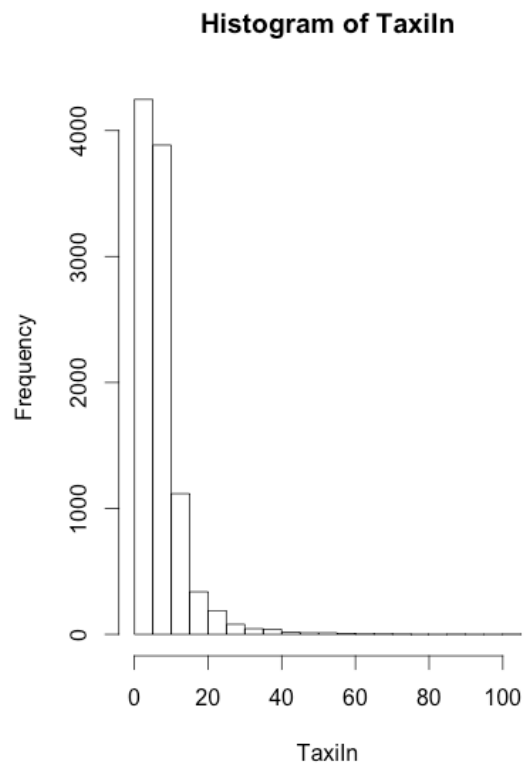-1.5IQR + Q1 = -1.5 * 5 + 4 = -3.5
Yes, because maximum number is 105 larger than 16.5, and when we check the boxplot in part
4, it shows that many data are larger than 16.5 which are outliers.

4.

It is right skewed. Because right whisker are pretty long.

5.

Histogram of TaxiIn

The histogram shows most points are between 0 to 20. So I agree with the outliers rule because there are so many points far from the median.