CitiBike Analysis

1. Compute summary statistics for tripduration

```
citi <- read.csv("https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/JC-201709
summary(citi$tripduration)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##      61.0     238.0     355.0     756.9     610.0 2181628.0
```

2. Compute summary statistics for age

```
citi$birth.year <- as.numeric(as.character(citi$birth.year))
```

```
## Warning: NAs introduced by coercion
```

```
citi$age<- (2019 - citi$birth.year)
summary(citi$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   32.00   36.00   38.88   44.00  132.00    2384
```

3. Compute summary statistics for tripduration in minutes (Need to transform tripduration from seconds to minutes)

```
citi$citi_tripduration_mins <- citi$tripduration/60
summary(citi$citi_tripduration_mins)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     1.02     3.97     5.92    12.62    10.17 36360.47
```

4. Compute the correlation between age and tripduration

```
cor(citi$tripduration, citi$age, use = "complete.obs")
```

```
## [1] 0.007055148
```

1. What is the total revenue assuming all users riding bikes from 0 to 45 minutes pay $3 per ride and user exceeding 45 minutes pay an additional $2 per ride.

```
trip_1 <- sum(citi$citi_tripduration_mins < 45) *3
trip_2 <- sum(citi$citi_tripduration_mins > 45) *5
total_revenue <- trip_1 + trip_2
total_revenue
```

```
## [1] 100651
```

Another method using table function in R

```
revenue <- sum(table(citi$citi_trip>45)*c(3,5))
revenue
```

```
## [1] 100651
```

2. Looking at tripduration in minutes, what can you say about the variance in the data.

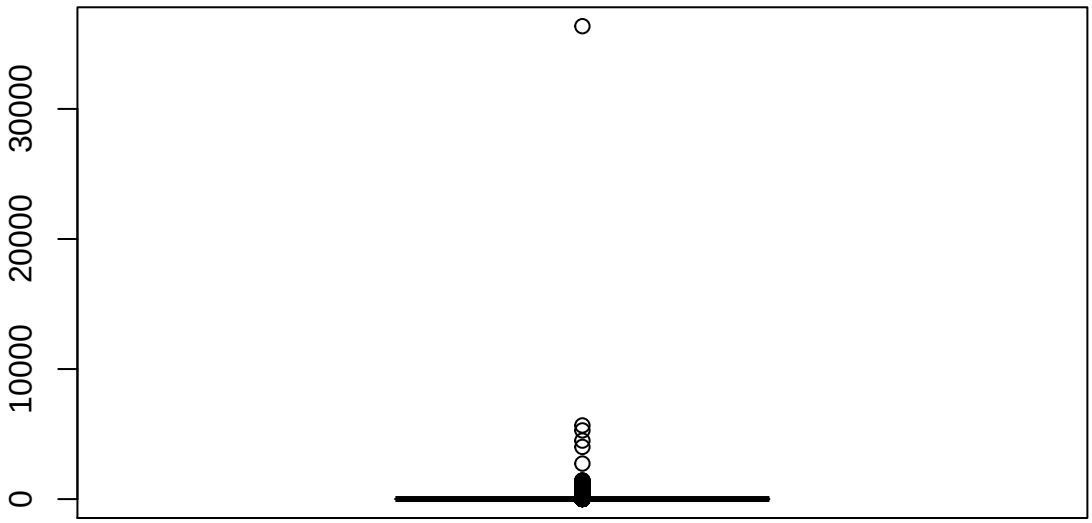
```
var(citi$citi_tripduration_mins)
```

```
## [1] 44300.24
```

```
sd(citi$citi_tripduration_mins)
```

```
## [1] 210.4762
```

```
boxplot(citi$citi_tripduration_mins) # High Variance can be observed in the box plot below
```



3. What does this mean for the pricing strategy?

```
quantile(citi$citi_tripduration_mins,c(.98))
```

```
##    98%
## 44.65
```

98% people drive citibike less that 45 mins hence if the company wants to increase profit they should increase proce below 45 mins or they can reduce this threshold to 40 mins.