Assignment 2

```
install.packages('nycflights13')

library(nycflights13)

data(flights)

View(flights)
```

 # 1(a)

```
flights[flights$origin=='JFK' | flights$origin=='LGA' | flights$origin=='EWR' ,]
```

Ans: **336,776**

```
flights[flights$dest=='JFK' | flights$dest=='LGA' | flights$dest=='EWR' ,]
```

Ans: **1**

1(b)

```
flights[(flights$origin=='JFK' | flights$origin=='LGA' | flights$origin=='EWR') & flights$dest=='SEA' ,]
```

Ans: **3,923**

1(c)

```
unique(flights[(flights$origin=='JFK' | flights$origin=='LGA' | flights$origin=='EWR') & flights$dest=='SEA' ,"carrier"])
```

#Ans: **5**

1(d)

```
subs <- flights[(flights$origin=='JFK' | flights$origin=='LGA' | flights$origin=='EWR') & flights$dest=='SEA' ,]

mean(subs$arr_delay, na.rm=T)
```

#Ans: **-1.099099**

2(a)

```
mean(flights$arr_delay, na.rm=T)
```

Ans: **6.895377**

median(flights$arr_delay, na.rm=T)

Ans: **-5**

#2(b)

Ans: Negative arrival delay time would mean that the flight is NOT delayed. In fact, it is before time.

#2(c)

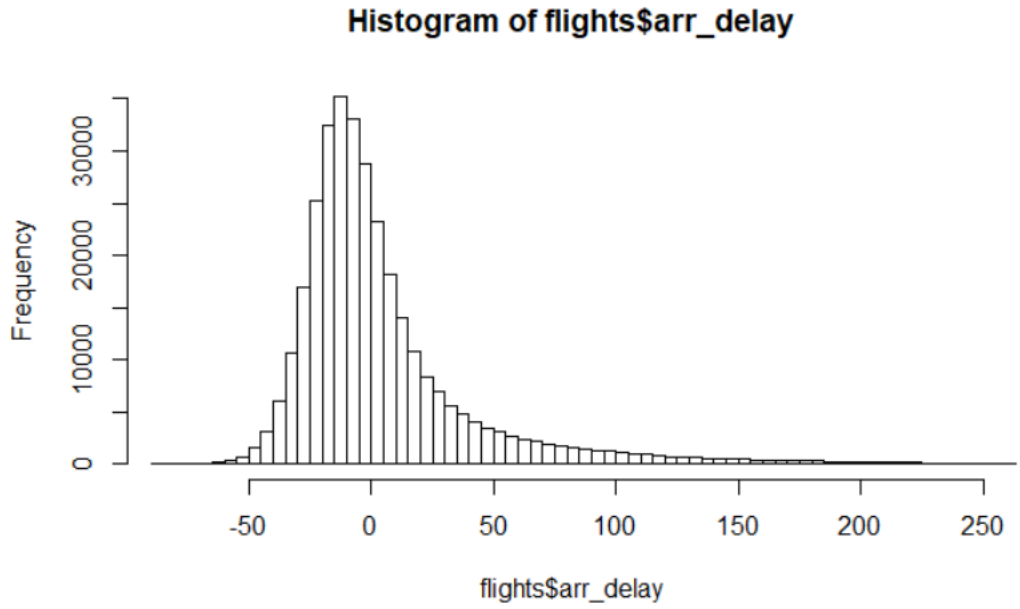
hist(flights$arr_delay, xlim=c(min(flights$arr_delay, na.rm=T),max(flights$arr_delay, na.rm=T)), breaks=300)

Zooming in - cutting at 300 to zoom in and closely see the values at x asis

hist(flights$arr_delay, xlim=c(min(flights$arr_delay, na.rm=T), 250), breaks=300)

Ans: Since most of the values lie between -50 and 0 and above 0, the mean should be positive but close to zero

The Median should definitely be negative and close to zero because the plot is left-skewed.



#2(d)

mean.delay <- tapply(flights$dep_delay,flights$month,function (x) mean(x, na.rm=T))

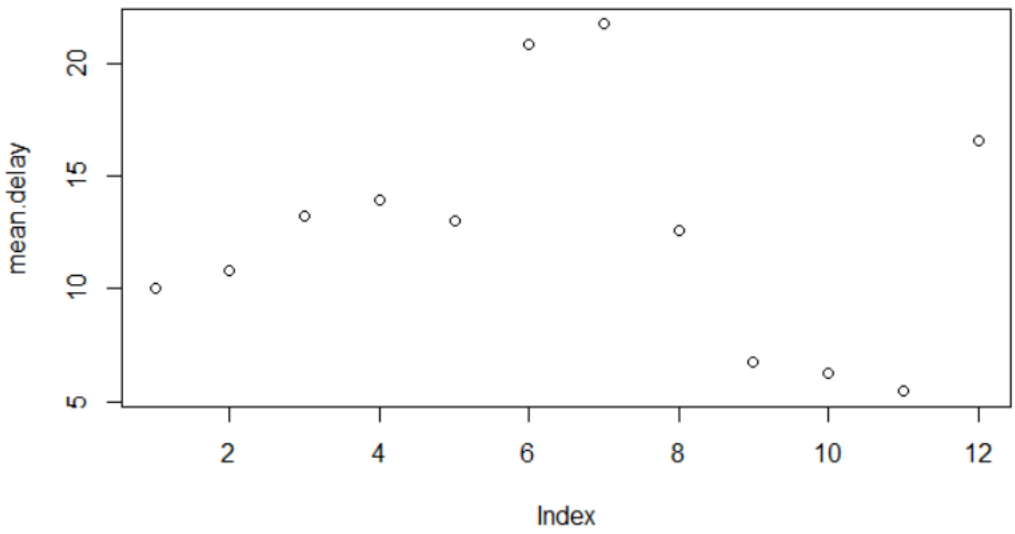
plot(mean.delay)

Ans: Clearly, the best month to leave NYC is November because the average departure delay is minimum.

The worst month to leave would be July because the average departure delay is maximum.

The pattern is such that the delay time is low in the initial months, goes to the peak in the middle of the year and again comes down.

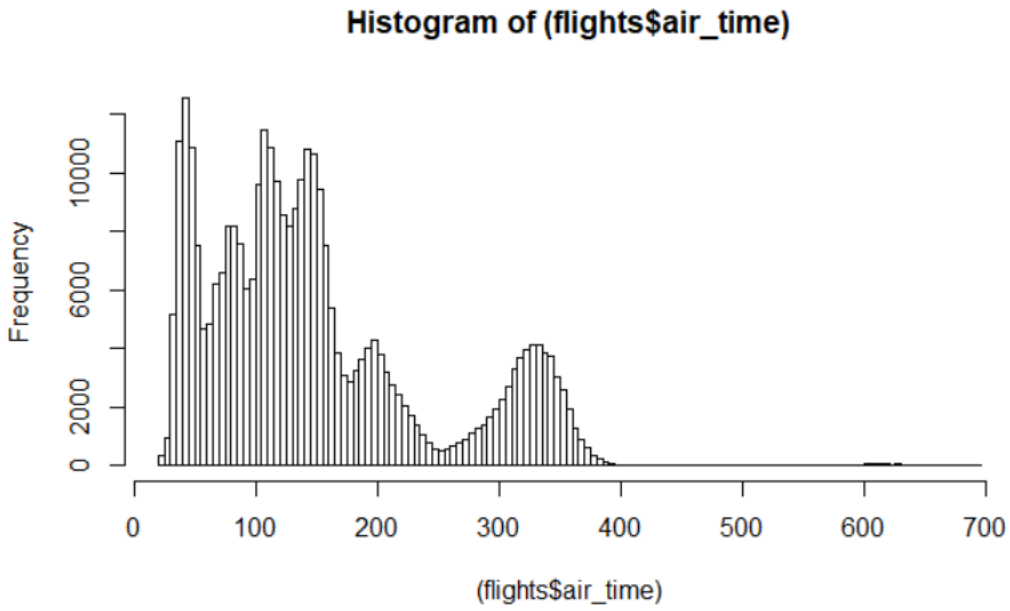
This looks like a normal distribution.



#3(a)

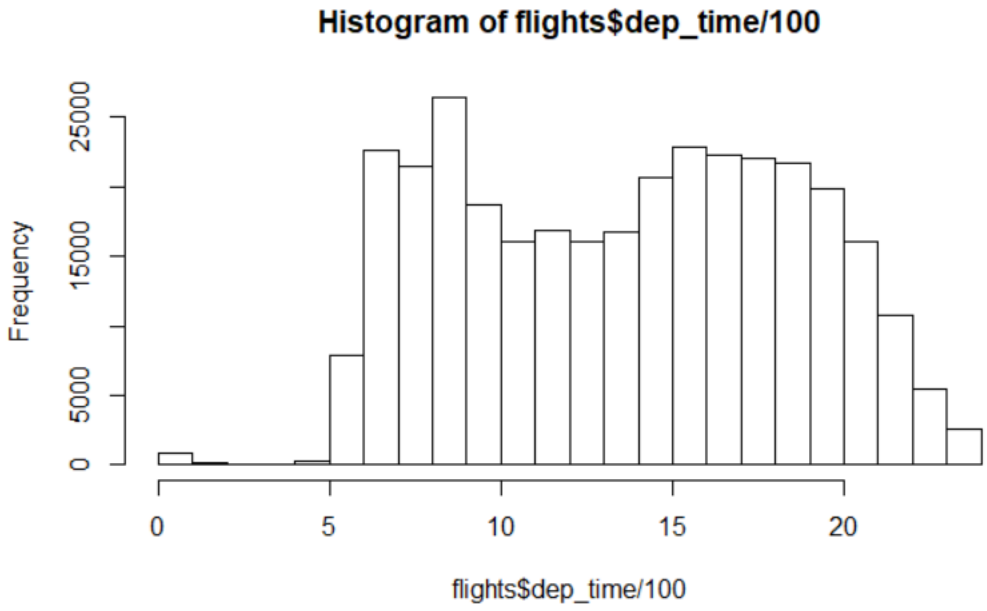
hist((flights$air_time), breaks=100)

Ans: There are **6** peaks in the distribution. There are more shorter duration flights than longer duration flights. There are peaks in flight times in ranges of 0-50, 50-100, 100-125, 150-200,250-400.



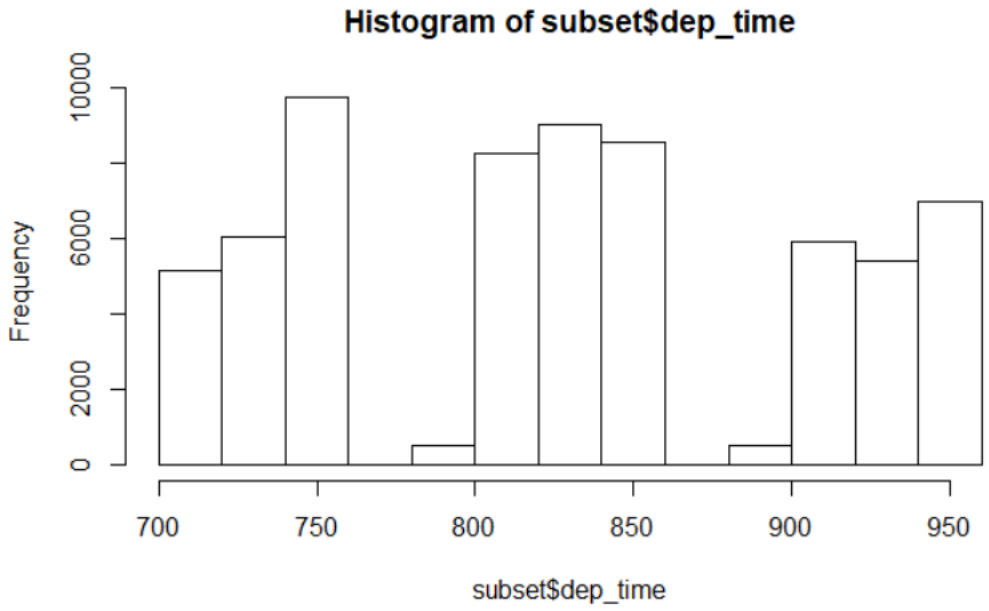
#3(b)

hist(flights$dep_time/100)



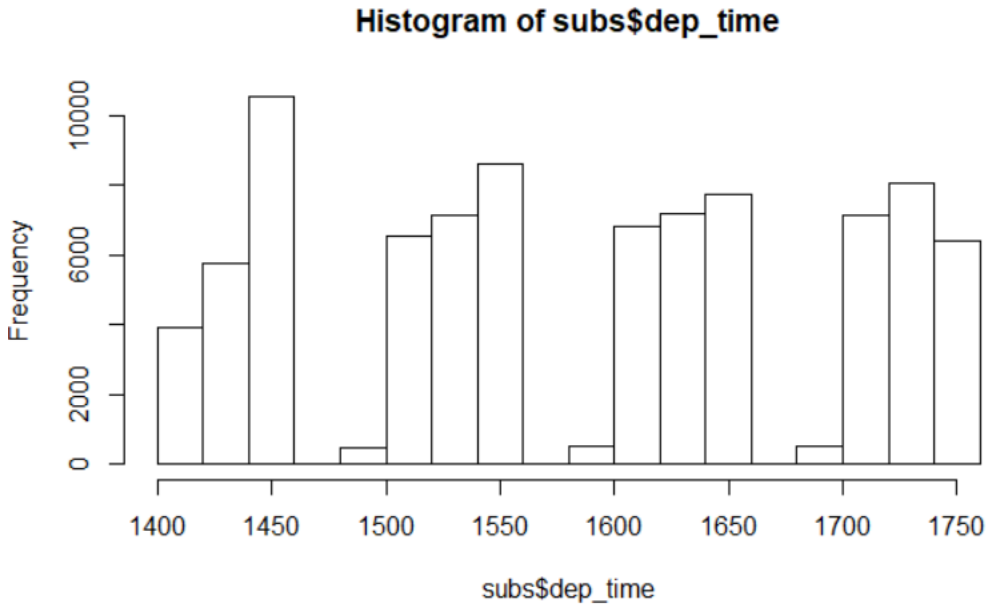
subset <- flights[flights$dep_time > 700 & flights$dep_time <1000,"dep_time"]

hist(subset$dep_time)




```
subs <- flights[flights$dep_time > 1400 & flights$dep_time <1800,"dep_time"]

hist(subs$dep_time)
```



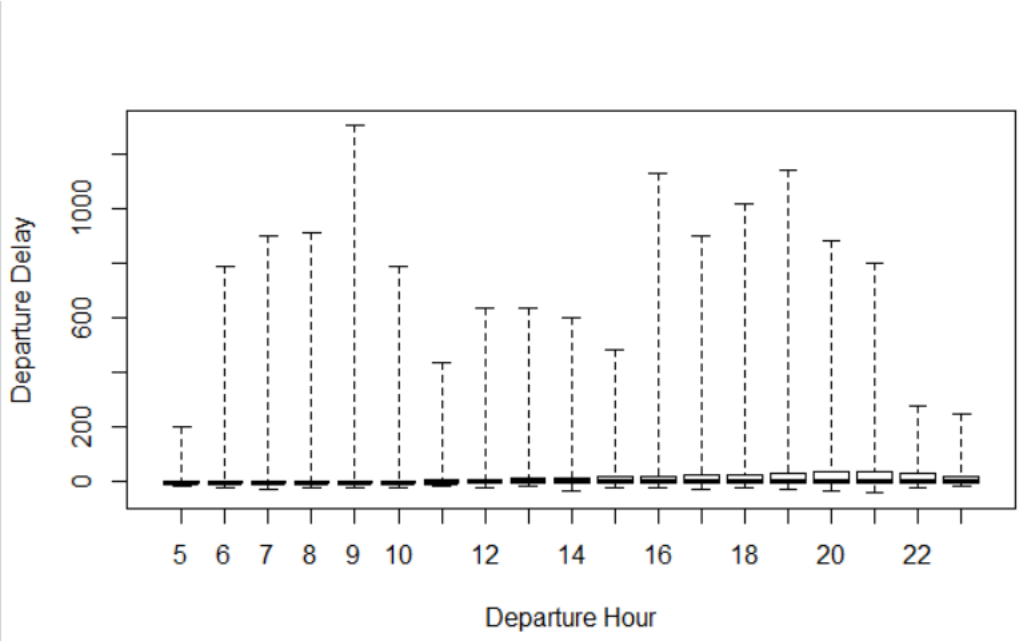
Ans: Clearly, maximum flights depart around 7:50 am in the morning. Second highest number of flights depart at 14.50 hours that is 2.50pm in the afternoon.

There are two most popular times - one in the morning and one in the afternoon.

#3(c)

boxplot(dep_delay ~ hour, data = flights, range=0, ylab="Departure Delay", xlab ="Departure Hour")

Ans: Variation is minimum from 5am to 11am. From 19 to 22(7pm to 10pm) variation is more, values are more spread out, signaling high air traffic causing delays.



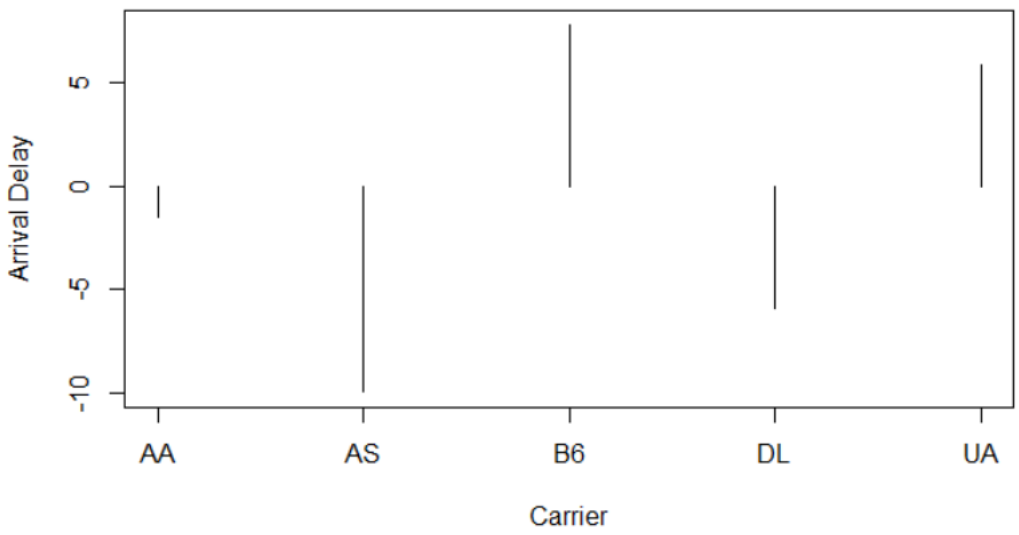
#4 Research question: In 2013, among the flights landing in Seattle, which ones should a passenger consider booking for a better experience in future and which ones should they definitely avoid?

flyToSeattle <- flights[flights$dest=='SEA',]

arrDelay <- tapply(flyToSeattle$arr_delay, flyToSeattle$carrier, function(x) mean(x,na.rm=TRUE))

plot(arrDelay, type='h', xaxt='n, 'xlab="Carrier", ylab="Arrival Delay")

axis(1, c(1:5), labels=c(names(arrDelay)))



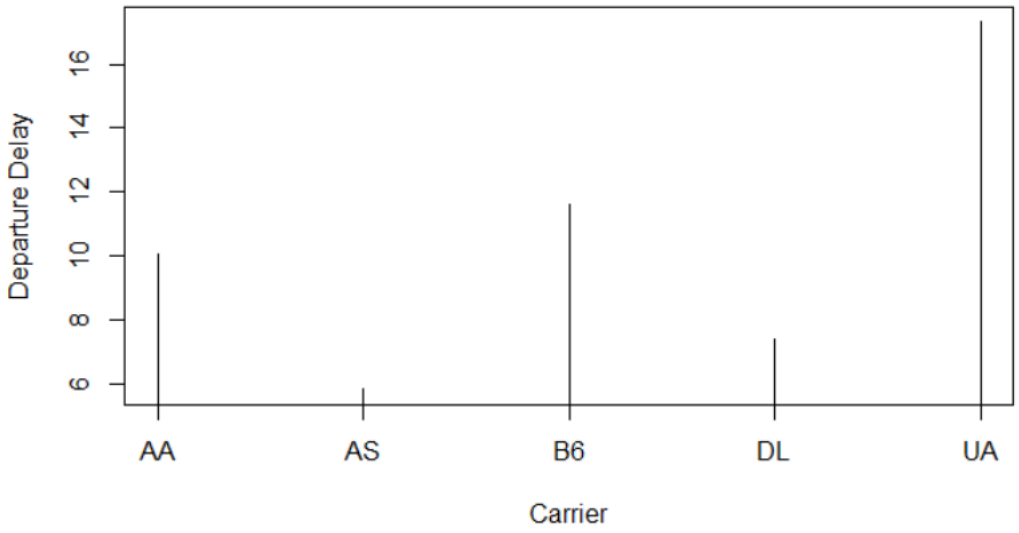
According to the plot, AS and DL with -9.930889 and -5.886023 departure delay respectively are the best flights.

B6 is the worst with departure delay of 7.721248

```
depDelay <- tapply(flyToSeattle$dep_delay, flyToSeattle$carrier, function(x) mean(x,na.rm=TRUE))

plot(depDelay, type='h', xaxt='n',xlab="Carrier", ylab="Departure Delay")

axis(1, c(1:5), labels=c(names(depDelay)))
```



According to the plot, AS and DL with 5.804775 and 7.391376 departure delay respectively are the best flights.

B6 and UA are the worst flights with departure delay of 11.592593 and 17.315647 respectively.

Considering both the plots, AS and DL are the best flights and should be considered by passengers while B6 and UA should be avoided.