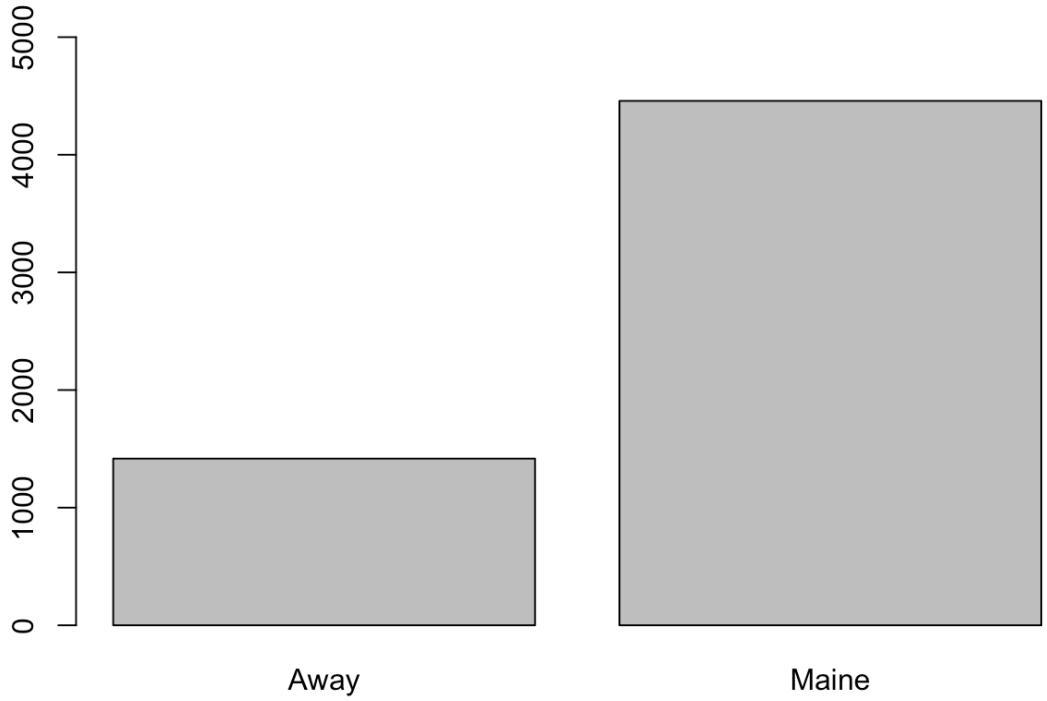
Mini Project 2

Di Jin dxj170930

Question 1

a)

Use read.csv funtion to read the data in csv file. Get only the maine column. Use table funtion to calculate the frequency. Then draw the barplot with the data.



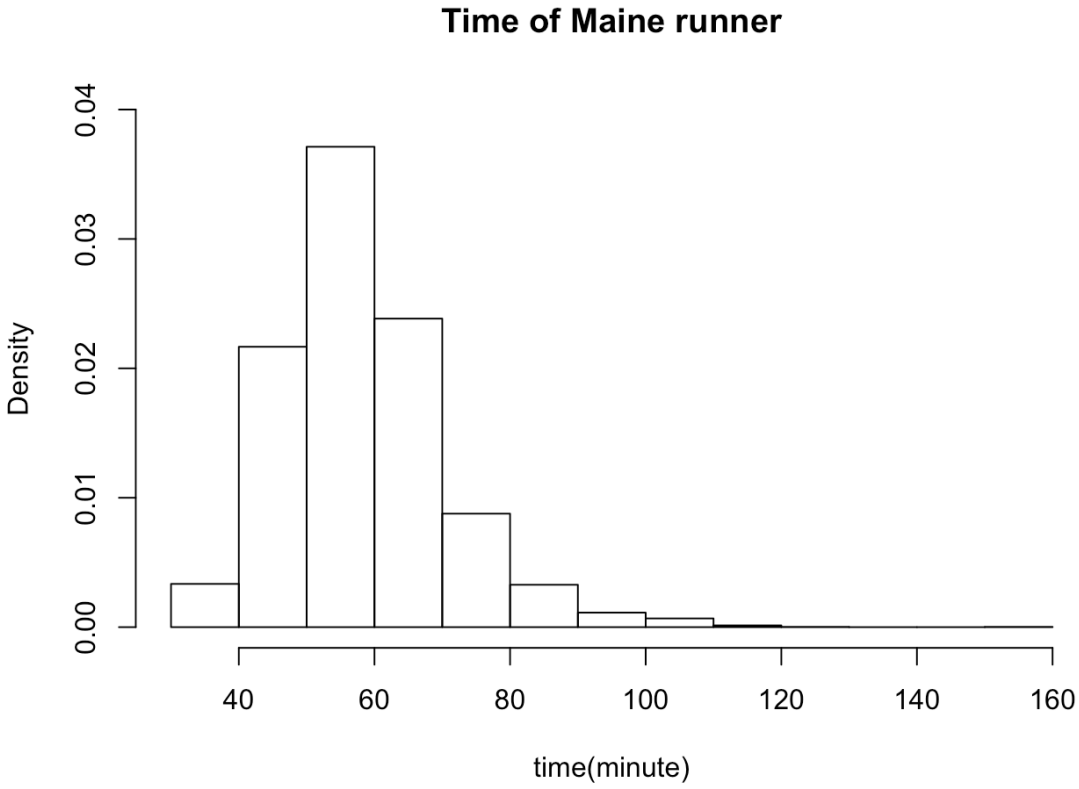
.	Maine	Away
proportion	75.88%	24.12%

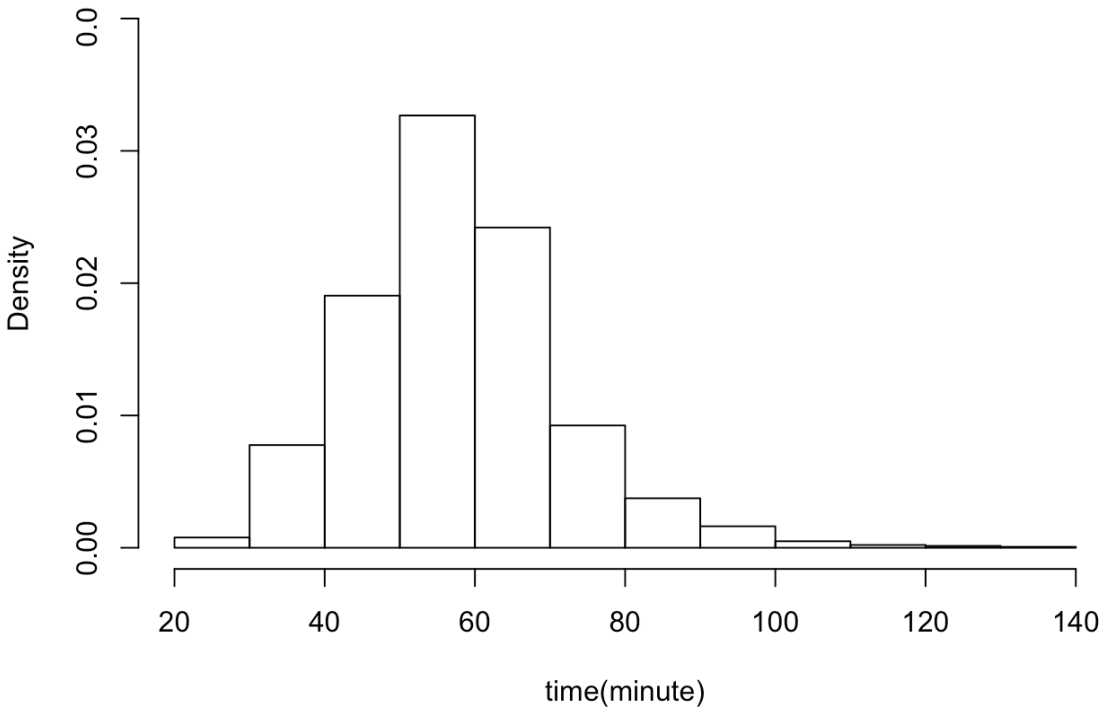
number	4458	1417

Local runner is much more than away runner.

b)

Get the subset of data according to the 'Maine' attribute. Calculate their running time in minute, and draw the histogram.





Summary Table

Place	mean	std	range	median	IQR
Maine	58.19513	12.18511	30.57-152.17	57.03	14.25
Away	57.82181	13.83539	27.78-133.71	56.92	15.67

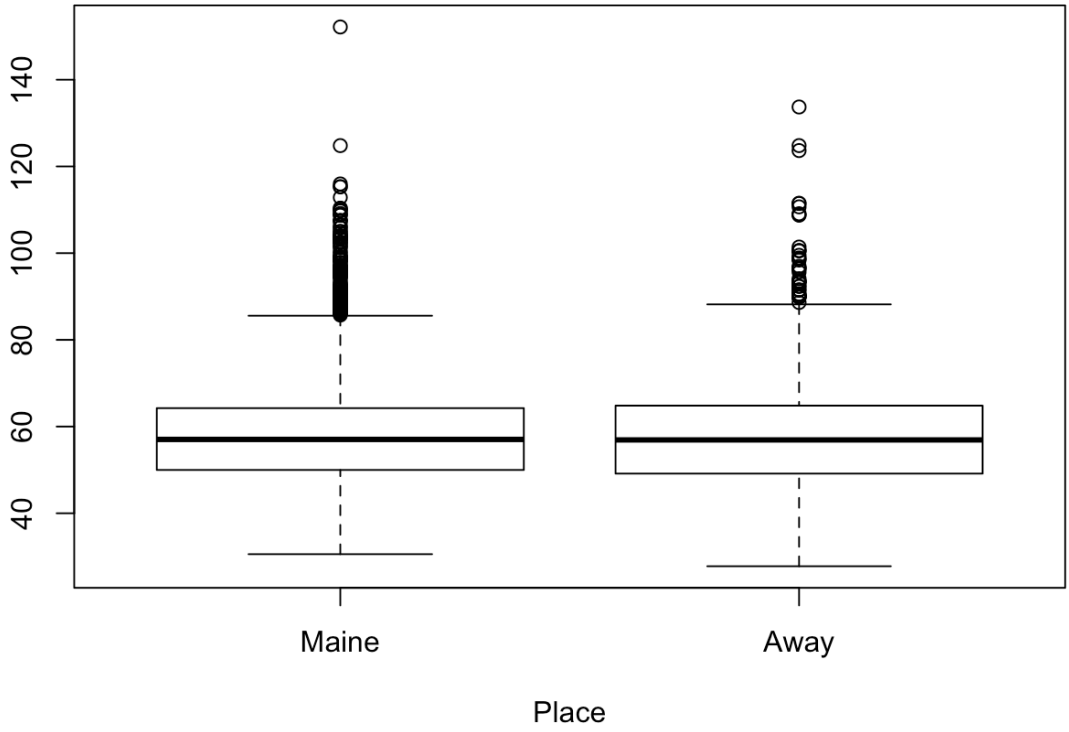
Conclude

Away runner is faster than Maine runner according to their mean. But Maine runner has more stable distribution around mean according to standard devition.

c)

Use boxplot funtion to draw the graph.

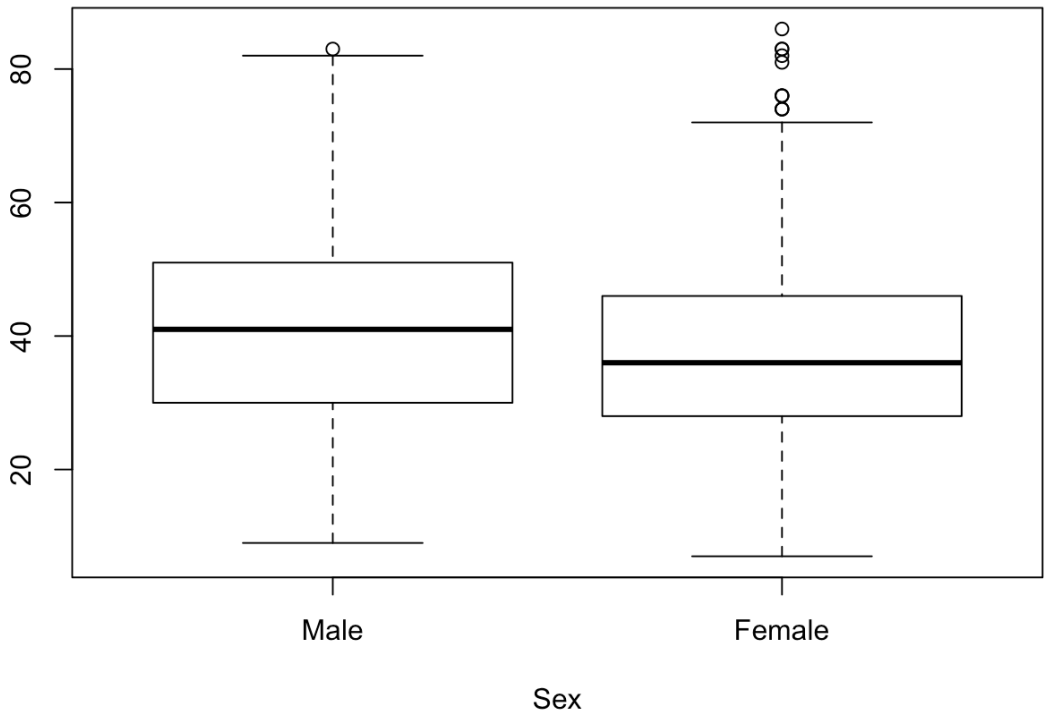
Time of runner



d)

Calculate age according to sex and draw the graph.

Age of runner



Summary Table

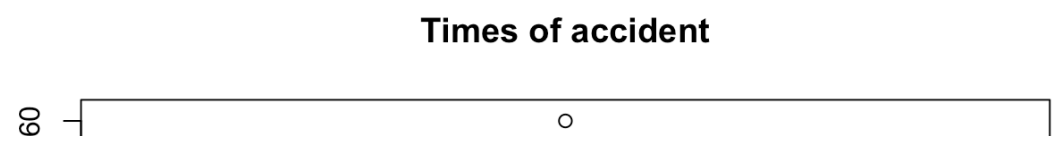
Sex	mean	std	range	median	IQR
Male	40.45	13.99	9-83	41	21
Female	37.24	12.27	7-86	36	18

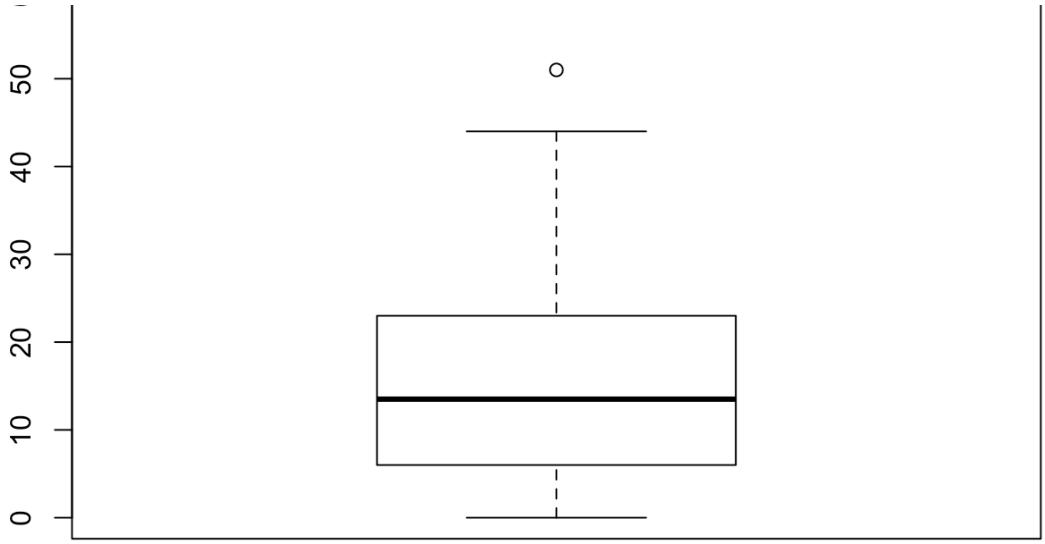
Conclude

Male runner is in average older than female runner and female runner's age is more centralized according to IQR.

Question 2

Read the value and get the times of the accident.





Summary Table

.	mean	std	range	median	IQR
Times	17.02	14.07	0-60	13.5	17

Conclude

"CHARLESTON" "GREENVILLE" "HORRY" are outliers carculated by IQR+75% and 25%-IQR. The reason why these palce have high accident numbers may be people in these place like riding motor motorcycle more than other place. So it is resonable that there are more accident with motorcycle.

CODE

Question 1

```
#Question 1
#AAAAAAAAAAAAAAAA
race<-read.csv("/Users/dijin/Desktop/统
计/minipro/2/roadrace.csv",head=T,sep=",",na.strings="*")  #get the
data and skip "*" one
```

```r
place<-race[11]  # get column maine
place_freq<-table(place)  #get the frequency

print(place_freq)# show proportion of each place
barplot(place_freq,ylim=c(0,5000))  #draw the barplot according to
frenquncy
#BBBBBBBBBBBBBBBBBB
maine<-subset(race,Maine=="Maine") #get the subset where they are
from maine
time_maine<-maine[8]/60 #get time of each runner in minute
#head(time_maine)
#mode(time_maine)
time_maine<-unlist(time_maine)  # change the type of time_marin for
hist
hist(time_maine,main="Time of Maine
runner",xlab="time(minute)",ylim=c(0,0.04), probability = T) #draw
hist


away<-subset(race,Maine=="Away") #get the subset where they are
from away
time_away<-away[8]/60 #get time of each runner in minute
time_away<-unlist(time_away)  # change the type of time_away for
hist
hist(time_away,main="Time of Away
runner",xlab="time(minute)",ylim=c(0,0.04), probability = T) #draw
hist

mean(time_maine) #mean
mean(time_away)

sd(time_maine) #standard
sd(time_away)

range(time_maine) # range
range(time_away)

median(time_maine) #median
median(time_away)

IQR(time_maine) # IQR
IQR(time_away)

#CCCCCCCCCCCCC
place<-c("Maine","Away")  #set category
boxplot(time_maine,time_away,names=place, main = "Time of runner",
xlab = "Place") #draw boxplot according to place

#DDDDDDDDDDDDDD
```

```r
male<-subset(race,Sex=="M") #get the subset where they are from
maine

female<-subset(race,Sex=="F") #get the subset where they are from
maine
m_age<-as.numeric(unlist(male[5])) #get the column sex=male unlist
and change it to numeric for boxplot
f_age<-as.numeric(unlist(female[5]))# same with above
sex<-c("Male","Female")  #set category
boxplot(m_age,f_age,names=sex, main = "Age of runner", xlab =
"Sex") #draw boxplot according to sex

mean(m_age) #mean
mean(f_age)

sd(m_age) #standard
sd(f_age)

range(m_age) # range
range(f_age)

median(m_age) #median
median(f_age)

IQR(m_age) # IQR
IQR(f_age)
```

Question 2

```r
#Question 2
motor<-read.csv("/Users/dijin/Desktop/统
```

```
17/minipro/2/motorcycle.csv",head=T,sep=",")  #get the data
number<-unlist(motor[2]) #get the column and unlist for boxplot

boxplot(number, main = "Times of accident") #draw boxplot

summary(number)
IQR(number)
range(number)
temp1<-motor[motor[2]>(quantile(number,0.75)+IQR(number))] #get the
outlier above 75%+IQR
temp2<-motor[motor[2]<(quantile(number,0.25)-IQR(number))] #get the
outlier below 25%-IQR
```