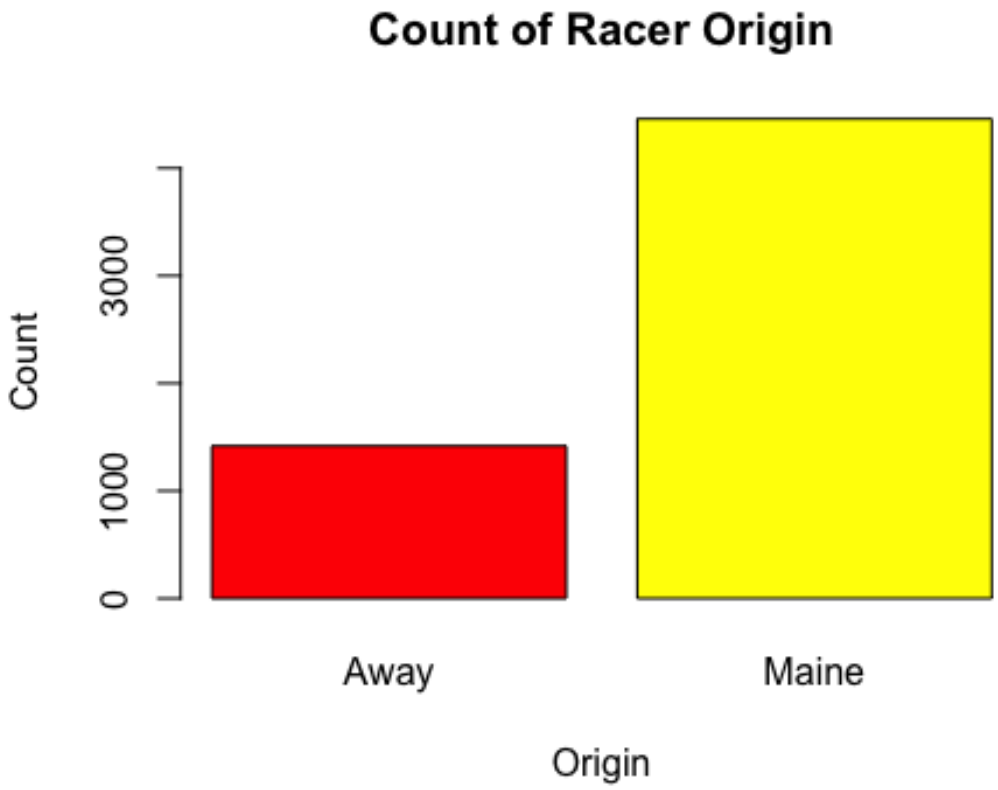
Five-number-summary

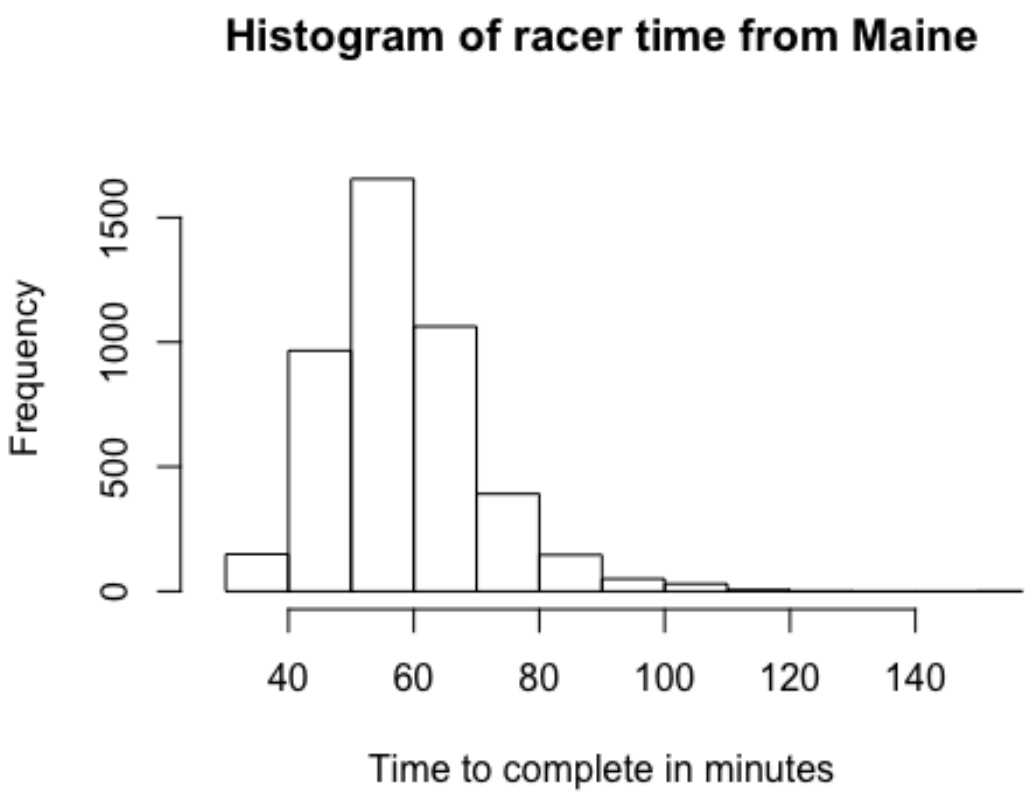
Analysis of first dataset

```
file <- "roadrace.csv"
data <- read.csv(file,header = T)
  # (a)
    # use table function to get counts
    maine <- table(data$Maine)
    maine

##
##  Away Maine
##  1417  4458

    # plot counts
    barplot(maine,main="Count of Racer Origin", ylab="Count",
            xlab="Origin", col=c("red", "yellow"))
```

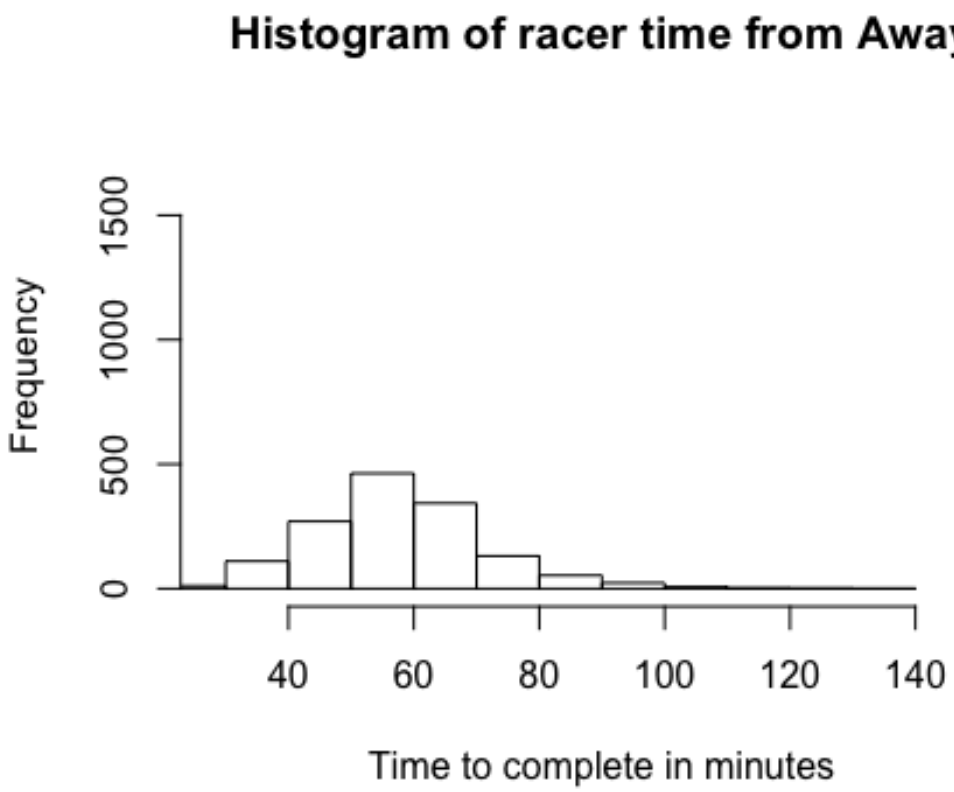



```
  # (b)
    # subset data
```

```r
bdata <- data[c("Maine","Time..minutes.")]
# convert column to numeric
bdata$Time..minutes. <- as.numeric(bdata$Time..minutes.)
# separate into two groups
mgroup <- bdata[ which(bdata$Maine=='Maine'), ]$Time..minutes.
agroup <- bdata[ which(bdata$Maine=='Away'), ]$Time..minutes.

# create histograms for each group using same scale
hist(mgroup, main="Histogram of racer time from Maine",
     xlab="Time to complete in minutes", ylim = c(0,1800),
     xlim = c(min(bdata$Time..minutes.),max(bdata$Time..minutes.)))
```



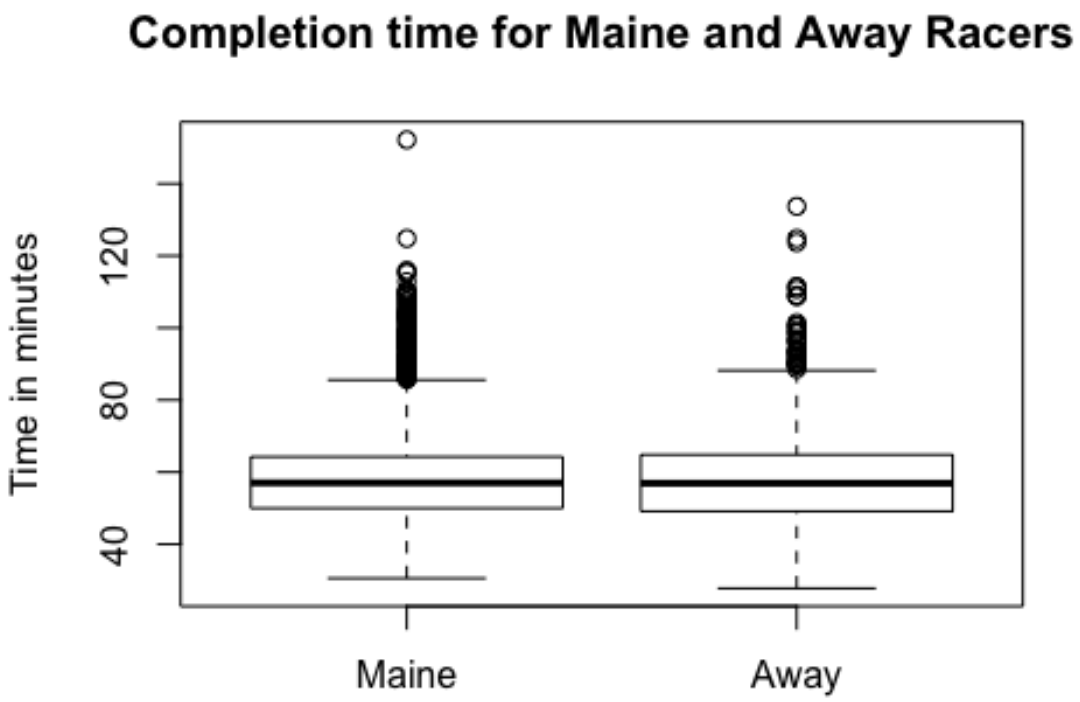

```r
hist(agroup, main="Histogram of racer time from Away",
     xlab="Time to complete in minutes", ylim = c(0,1800),
     xlim = c(min(bdata$Time..minutes.),max(bdata$Time..minutes.)))
```




```
    # statistics
    stat <- function(var) {
      groupstat <- c(mean(var),sd(var),max(var)-
min(var),median(var),IQR(var))
      names(groupstat) <-
c('Mean','StandardDeviation','Range','Median','InterquartileRange')
      return(groupstat)
    }
    round(stat(mgroup),3)

##               Mean  StandardDeviation              Range
Median
##             58.195             12.185            121.600
57.033
## InterquartileRange
##             14.248

    round(stat(agroup),3)

##               Mean  StandardDeviation              Range
Median
##             57.822             13.835            105.928
56.920
```

```
## InterquartileRange
##             15.674
```

```r
 # (c)
   boxplot(mgroup,agroup,main="Completion time for Maine and Away Racers",
           names=c('Maine','Away'),ylab='Time in minutes')
```




```r
 # (d)
   # subset data
   ddata <- data[c("Sex","Age")]
   # convert column to numeric
   ddata$Age <- as.numeric(ddata$Age)
   # separate into two groups
   male <- ddata[ which(ddata$Sex=='M'), ]$Age
   female <- ddata[ which(ddata$Sex=='F'), ]$Age

   # boxplot for runners ages among sex groups
   boxplot(male,female,main="Runners' Ages Among Sex Groups",
           names=c('Male','Female'),ylab='Age')
```


```
    # statistics for ages among sex groups
    round(stat(male),3)
```

```
##                Mean   StandardDeviation               Range
Median
##              32.563              14.070              75.000
33.000
## InterquartileRange
##              21.000
```

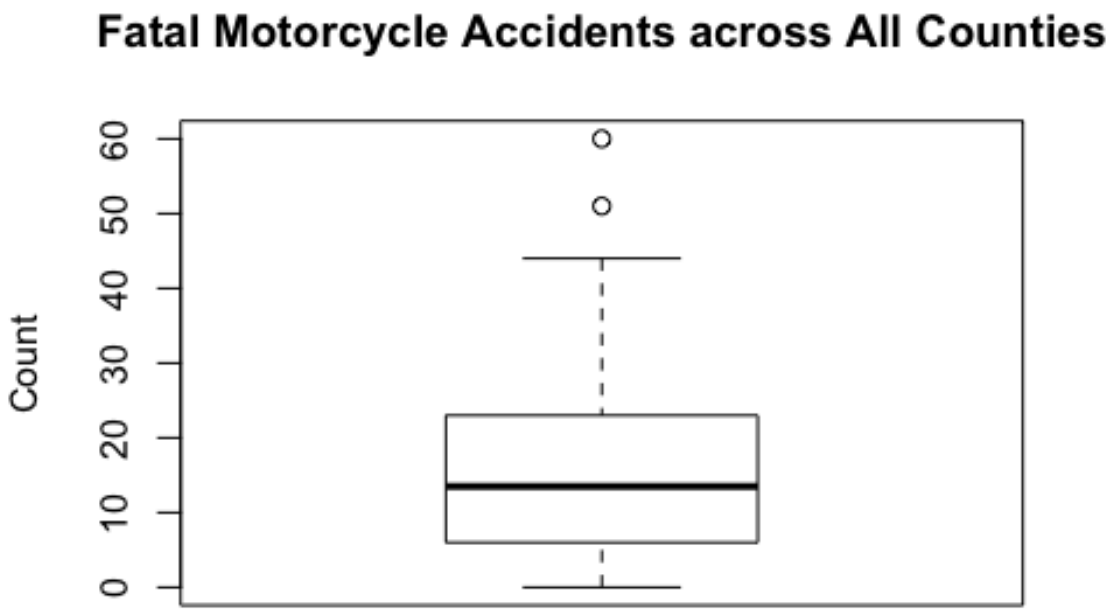
```
    round(stat(female),3)
```

```
##                Mean   StandardDeviation               Range
Median
##              29.263              12.285              74.000
28.000
## InterquartileRange
##              18.000
```

Analysis of second dataset

```
  # load data
  file2 <- "motorcycle.csv"
  data2 <- read.csv(file2,header = T)
```

```r
# convert column to numeric
fatal <- as.numeric(data2$Fatal.Motorcycle.Accidents)

# create boxplot for number of fatal accidents of all counties
boxplot(fatal,main="Fatal Motorcycle Accidents across All Counties",
        ylab='Count')
```




```r
# statistics
round(stat(fatal),3)
```

```
##               Mean  StandardDeviation              Range
Median
##             17.021             13.813             60.000
13.500
## InterquartileRange
##             17.000
```

```r
# outliers
fatal_iqr = IQR(fatal)
lowbound = quantile(fatal, probs = 0.25) - 1.5 * fatal_iqr
upbound = quantile(fatal, probs = 0.75) + 1.5 * fatal_iqr
outliers <- data2[ which(data2$Fatal.Motorcycle.Accidents < lowbound |
```

```
              data2$Fatal.Motorcycle.Accidents > upbound), ]
  outliers

##        County Fatal.Motorcycle.Accidents
## 23 GREENVILLE                        51
## 26      HORRY                        60
```