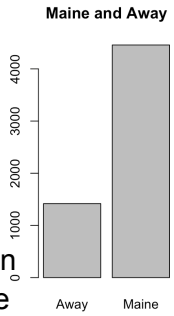
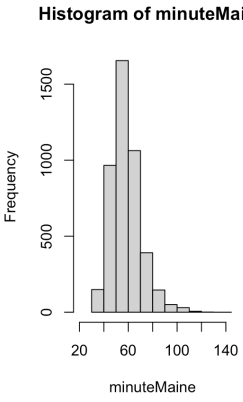
Mini Project 2
Kaden Chan

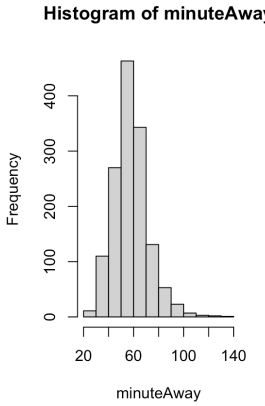
Section 1:



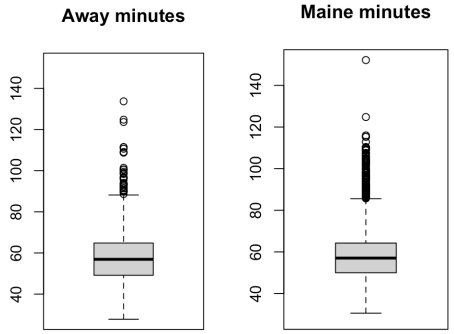
1. a) We can see that there are significantly more racers from Maine than those who are Away. About a 3:1 ratio (4500 from Maine, 1500 otherwise)
   b) Maine and Away's distributions for Time both reflect a normal distribution
   Away has slightly higher std, and lower range, but other stats are the same
   We can conclude that Being from Maine or not does not affect the time significantly.




```
[1] " Maine Time Stats: "
Mean: 58.19514
 Standard Deviation: 12.18511
 Range: 121.6
 Median: 57.0335
 IQR: 14.24775
[1] " Away Time Stats: "
Mean: 57.82181
 Standard Deviation: 13.83538
 Range: 105.928
 Median: 56.92
 IQR: 15.674
```

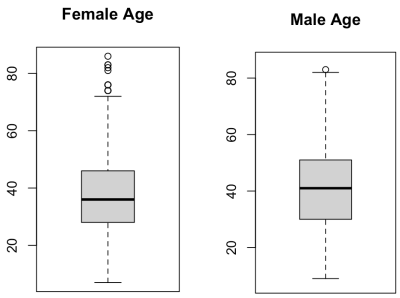


c)



This clearly shows Maine has a higher range.

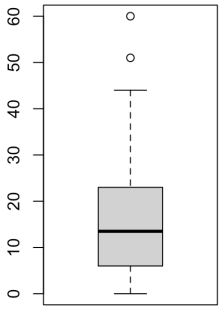
d)




```
1] " Male Age Stats: "
ean: 40.4468
Standard Deviation: 13.99289
Range: 74
Median: 41
IQR: 21
1] " Female Age Stats: "
ean: 37.23653
Standard Deviation: 12.26925
Range: 79
Median: 36
IQR: 18
```

There are a few more female outliers older than 80. The mean and median for male runners are a little higher than females. IQR for males and STD are also slightly higher. We can conclude that Male runners are typically older, except for few female outliers

2. **Fatal Motorcycle accidents in South Caroline counties during 2009**
   The range is 60, median is around 13,
   Quantiles: 25% is about 8, 75% is around 23.
   outliers are HORRY (51) and GREENVILLE (60)
   The distribution is right skewed in how there are higher outliers present.
   The outliers may have more tourist spots to encourage motorcycles to drive through, or in general a higher number of motorcycles than other counties.



```r
# Mini Project 2
# Name: Kaden Chan
# Section 1. Answers to the specific questions asked
# Section 2: R code. Your code must be annotated.
# No points may be given if a brief look at the code
# does not tell us what it is doing.

# 1
data <- read.csv("roadrace.csv")
# print(data)

# 1a bargraph - plot # of Mainea and Away racers
maineFreq <- table(data$Maine)
barplot(maineFreq, main = "Maine and Away")

# 1b split into Maine and Away
maineRacers <- data[data$Maine == "Maine", ]
awayRacers <- data[data$Maine == "Away", ]

# get minutes and set limit to Away's x range
minuteMaine <- maineRacers$Time..minutes.
hist(minuteMaine, xlim = c(20, 140))

print(" Maine Time Stats: ")
cat(
 "Mean:",
 mean(minuteMaine),
 "\n",
 "Standard Deviation:",
 sd(minuteMaine),
 "\n",
 "Range:",
 diff(range(minuteMaine)),
 "\n",
 "Median:",
 median(minuteMaine),
 "\n",
 "IQR:",
 IQR(minuteMaine),
 "\n"
)
```

```r
# get minutes, we're using Away as default scale
minuteAway <- awayRacers$Time..minutes.
hist(minuteAway)

print(" Away Time Stats: ")
cat(
 "Mean:",
 mean(minuteAway),
 "\n",
 "Standard Deviation:",
 sd(minuteAway),
 "\n",
 "Range:",
 diff(range(minuteAway)),
 "\n",
 "Median:",
 median(minuteAway),
 "\n",
 "IQR:",
 IQR(minuteAway),
 "\n"
)

# 1c boxplots - set equal scales
y_limits <- range(c(minuteMaine, minuteAway))
boxplot(minuteMaine, main = "Maine minutes", ylim = y_limits)
boxplot(minuteAway, main = "Away minutes", ylim = y_limits)

# 1d age boxplots
maleRacers <- data[data$Sex == "M", ]
femaleRacers <- data[data$Sex == "F", ]

# isolate age attribute for each gender
ageMales <- maleRacers$Age
ageFemales <- femaleRacers$Age

# plot on the same scale
y_limits <- range(c(ageMales, ageFemales))
boxplot(ageMales, main = "Male Age", ylim = y_limits)
boxplot(ageFemales, main = "Female Age", ylim = y_limits)

print(" Male Age Stats: ")
```

```r
cat(
 "Mean:",
 mean(ageMales),
 "\n",
 "Standard Deviation:",
 sd(ageMales),
 "\n",
 "Range:",
 diff(range(ageMales)),
 "\n",
 "Median:",
 median(ageMales),
 "\n",
 "IQR:",
 IQR(ageMales),
 "\n"
)

print(" Female Age Stats: ")
cat(
 "Mean:",
 mean(ageFemales),
 "\n",
 "Standard Deviation:",
 sd(ageFemales),
 "\n",
 "Range:",
 diff(range(ageFemales)),
 "\n",
 "Median:",
 median(ageFemales),
 "\n",
 "IQR:",
 IQR(ageFemales),
 "\n"
)

# 2 read and plot Fatal Motorcycle Accidents
data <- read.csv("motorcycle.csv")
# print(data)

boxplot(data$Fatal.Motorcycle.Accidents)
```