Homework Assignment #7

Matt Kline

April 4, 2017

Part 1 Graphics:

1.) The data file alpe d huez.txt contains data on the fastest times for the Alpe d'Huez stage of the Tour de France cycle race, plus some contextual information on year and drug use allega- tions. In included are the following variables:

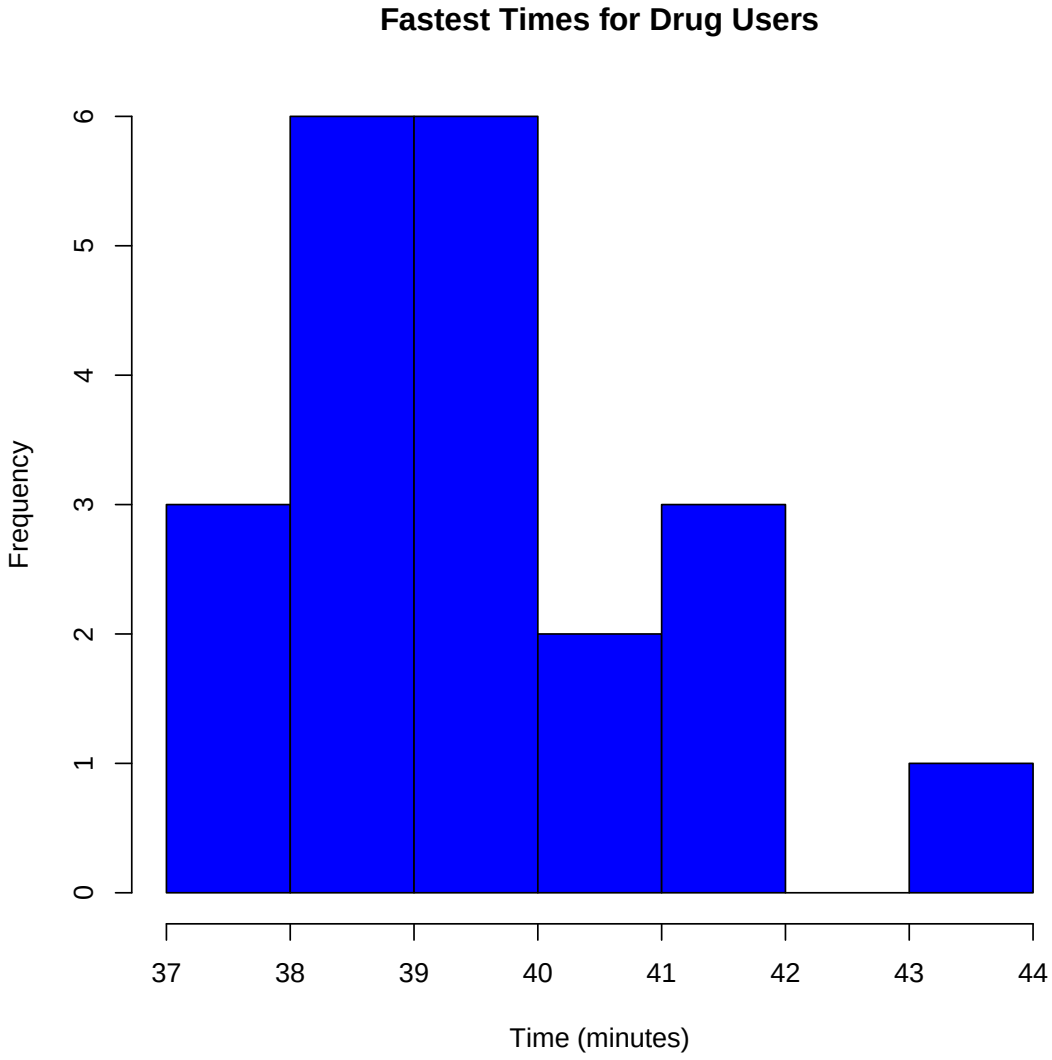
Time, a character vector of times of ride in the form Mmin Ssec. NumericTime, the numeric vector of time of ride in minutes. Name, name of rider. Year, year of race. Nationality, nationality of rider. DrugUse, Have allegations of drug use been made against the rider? The values are "Y" or "N".

Use read.table() to read the alpe d huez.txt data set into a data frame. Plot the distributions of fastest times, split by whether or not the rider (allegedly) used drugs. Display this using:

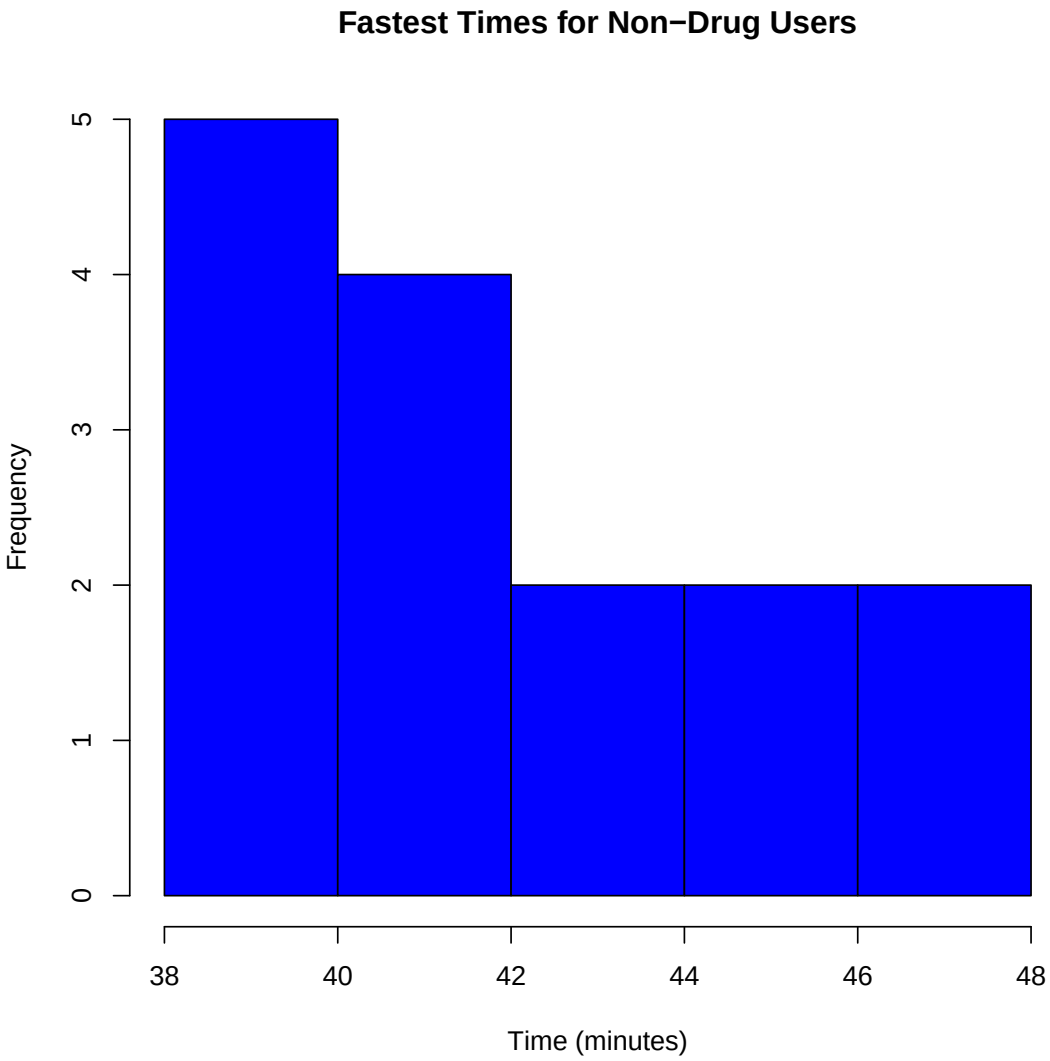
```
alpe_d_huez <- read.table("/Users/Matt/Desktop/alpe_d_huez.txt", header = T)
attach(alpe_d_huez)
```

a. Histograms. Use the mfrow graphical parameter in par(), along with hist() and its arguments main, xlab, col, and xlim to duplicate the figure array below:

```
hist(x = NumericTime[DrugUse == "Y"], xlab = "Time (minutes)", col = "blue",
     main = "Fastest Times for Drug Users")
```

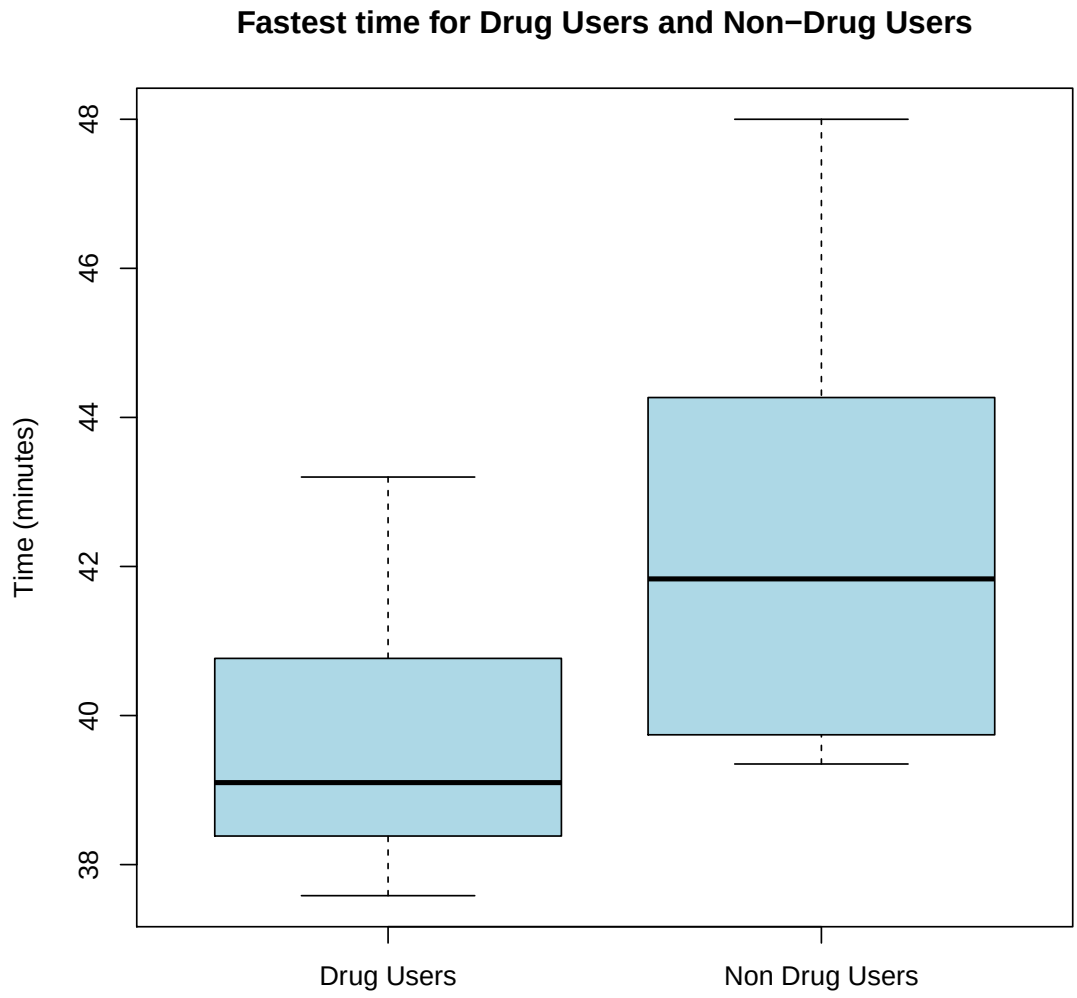



```
hist(x = NumericTime[DrugUse == "N"], xlab = "Time (minutes)", col = "blue",
     main = "Fastest Times for Non-Drug Users")
```



b. Boxplots. Use boxplot() and its arguments main, ylab, names, and col to duplicate the figure below.

```
boxplot(NumericTime[DrugUse == "Y"], NumericTime[DrugUse == "N"],
        names = c("Drug Users", "Non Drug Users"), ylab = "Time (minutes)",
        col = "light blue", main = "Fastest time for Drug Users and Non-Drug Users")
```



2.) The built-in dataset ChickWeight is a data frame with 578 rows and 4 columns from an experiment on the effect of diet on early growth of chicks. It contains the following variables:

weight, the body weight of the chick (grams). Time, the number of days since birth when the measurement was made. Chick, unique identifier for the chick, with values 1, 2, . . . , 50. Diet, indicator of which experimental diet the chick received (1, 2, 3, or 4).

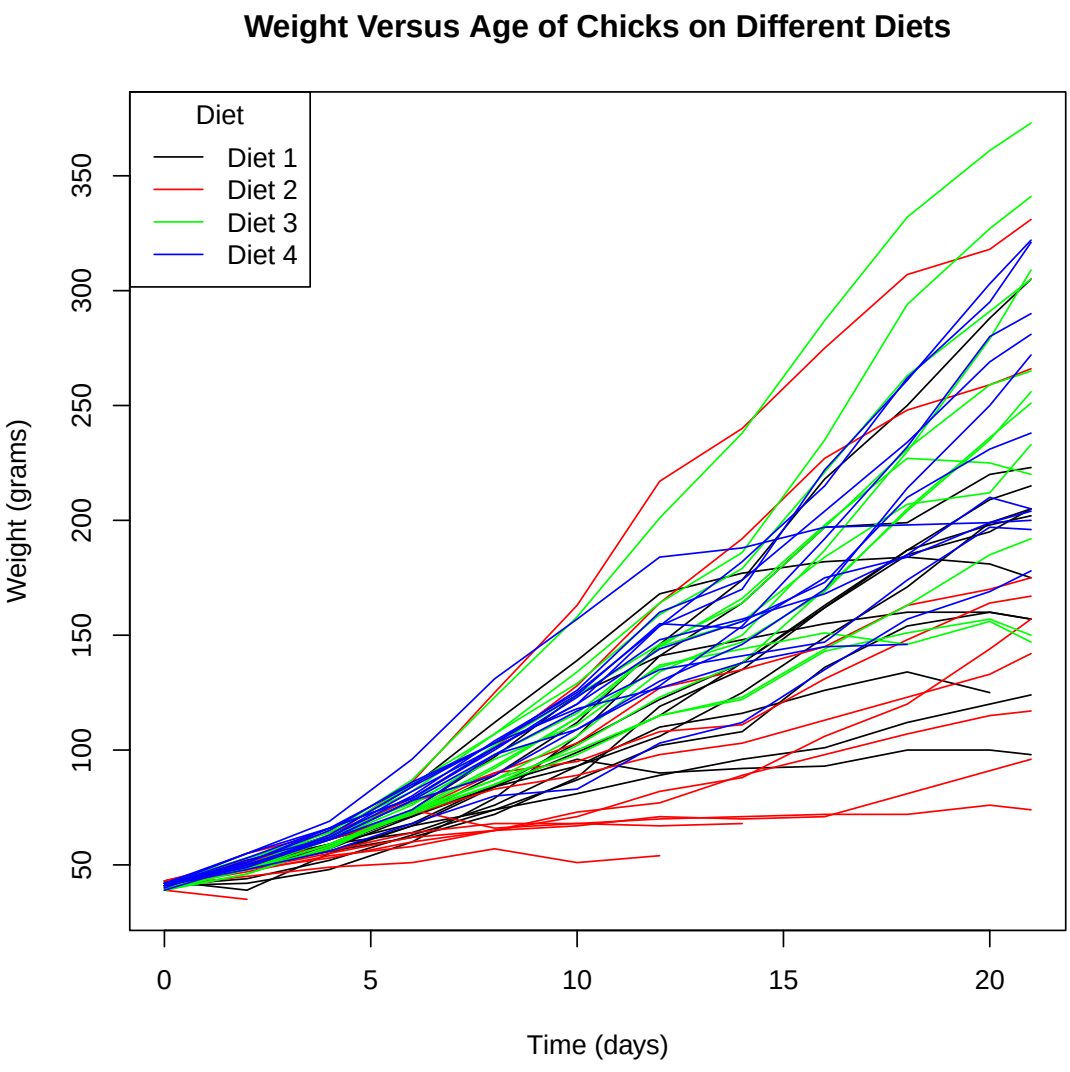
More information can be found from the help file:

```
#?ChickWeight
attach(ChickWeight)

## The following object is masked from alpe_d_huez:
##
##     Time
```

Duplicate the plot shown below, where each line shows a given chick's growth over time and the colors indicate which diet the chick received.

Now we can use lines() to add a black line corresponding to the first chick (which received Diet 1):

```r
plot(Time, weight, type = "n",
     main = "Weight Versus Age of Chicks on Different Diets",
     xlab = "Time (days)", ylab = "Weight (grams)")
for(i in 1:12){
  lines(Time[Chick == i], weight[Chick == i], col = "black")
}
for(i in 13:24){
  lines(Time[Chick == i], weight[Chick == i], col = "red")
}
for(i in 25:36){
  lines(Time[Chick == i], weight[Chick == i], col = "green")
}
for(i in 37:48){
  lines(Time[Chick == i], weight[Chick == i], col = "blue")
}

legend("topleft", legend = c("Diet 1", "Diet 2", "Diet 3", "Diet 4"),
       col = c("black", "red", "green", "blue"), lwd = c(1,1), title = "Diet")
```



Black lines for all the chicks that received Diet 1 (chicks 1 - 20) can be added using a for() loop, and then the same thing can be done for chicks on the other diets, but using a different color for each diet.

3.) The built-in R data sets mdeaths and fdeaths contain data on monthly deaths from lung diseases (bronchitis, emphysema and asthma) for males and females, respectively, in the United Kingdom for the years 1974 - 1979. Type:

```
#mdeaths
#fdeaths
```

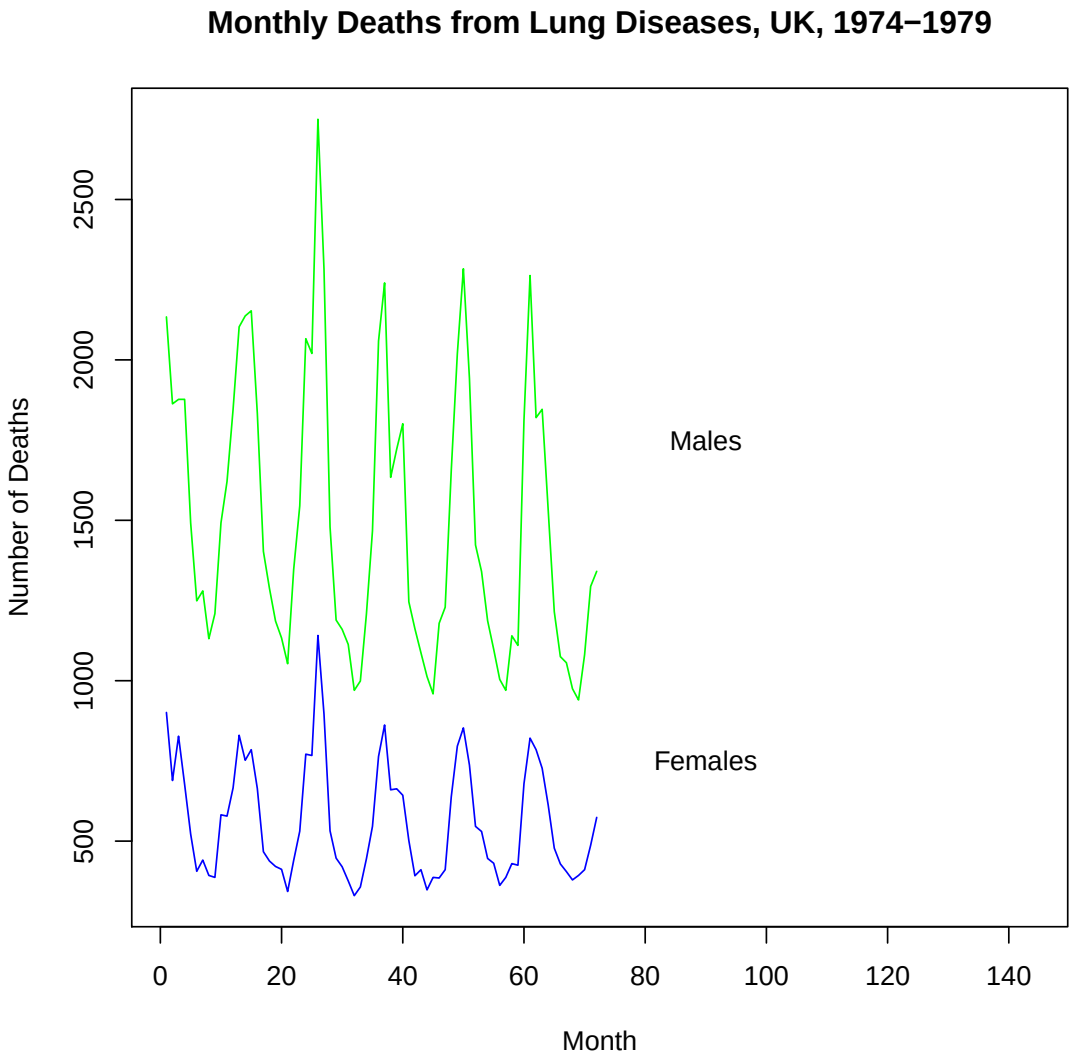
to view the data sets and type

```
#?mdeaths
```

for more information.

The function as.vector() will convert mdeaths and fdeaths to vectors. After converting mdeaths and fdeaths to vectors, duplicate the plot below.

```
mdeath <- as.vector(mdeaths)
fdeath <- as.vector(fdeaths)

plot(x = c(mdeath, fdeath), ylab = "Number of Deaths",
     xlab = "Month", type = "n",
     main = "Monthly Deaths from Lung Diseases, UK, 1974-1979")
lines(x = mdeath, col = "green")
lines(x = fdeath, col = "blue")

text(y = 1750, x = 90, labels = "Males")
text(y = 750, x = 90, labels = "Females")
```



4.) In the New Jersey Pick-3 Lottery, players pick three digits to try to match the New Jersey Lottery's drawings, which are aired daily on live television. Bets are placed together in a pool, and after the state takes its cut, the payoff is calculated by sharing the pool among all winning bets. Data from the lottery for 80 days are shown below.

```
lotteryPayoff <- c(239.5, 327.5, 205, 241, 188, 238.5, 294, 269.5,
                   248, 308, 195.5, 192.5, 240.5, 488, 164, 240.5,
                   346.5, 288.5, 327.5, 189.5, 196.5, 252.5, 211,
                   294.5, 301, 149.5, 241, 218.5, 242, 227.5, 327.5,
                   223, 226, 294.5, 294.5, 223.5, 258, 277.5, 274,
                   277.5, 249, 227.5, 350, 269.5, 262, 331.5, 240,
                   330, 170.5, 262, 228, 248.5, 285.5, 134.5, 393.5,
                   268, 233.5, 274.5, 236, 383, 354, 212, 329.5, 266,
                   327, 210.5, 369.5, 216, 293, 321.5, 387.5, 321.5,
                   226, 293, 228.5, 278, 256.5, 317, 184, 259)
lotteryNumber <- c(157, 997, 817, 530, 369, 250, 947, 783, 273, 367,
                  747, 428, 601, 886, 225, 432, 887, 959, 103, 126,
                  100, 936, 300, 843, 249, 011, 669, 134, 124, 071,
                  547, 771, 805, 970, 046, 471, 794, 407, 453, 168,
```

```
                  642, 914, 048, 035, 917, 753, 372, 096, 171, 037,
                  957, 643, 347, 301, 832, 242, 568, 846, 520, 995,
                  241, 829, 842, 606, 752, 404, 082, 722, 000, 587,
                  734, 599, 472, 924, 625, 276, 349, 752, 323, 371)
lotteryDate <- c("Mon,Aug4,2014", "Sun,Aug3,2014", "Sat,Aug2,2014",
                 "Fri,Aug1,2014", "Thu,Jul31,2014", "Wed,Jul30,2014",
                 "Tue,Jul29,2014", "Mon,Jul28,2014", "Sun,Jul27,2014",
                 "Sat,Jul26,2014", "Fri,Jul25,2014", "Thu,Jul24,2014",
                 "Wed,Jul23,2014", "Tue,Jul22,2014", "Mon,Jul21,2014",
                 "Sun,Jul20,2014", "Sat,Jul19,2014", "Fri,Jul18,2014",
                 "Thu,Jul17,2014", "Wed,Jul16,2014", "Tue,Jul15,2014",
                 "Mon,Jul14,2014", "Sun,Jul13,2014", "Sat,Jul12,2014",
                 "Fri,Jul11,2014", "Thu,Jul10,2014", "Wed,Jul9,2014",
                 "Tue,Jul8,2014", "Mon,Jul7,2014", "Sun,Jul6,2014",
                 "Sat,Jul5,2014", "Fri,Jul4,2014", "Thu,Jul3,2014",
                 "Wed,Jul2,2014", "Tue,Jul1,2014", "Mon,Jun30,2014",
                 "Sun,Jun29,2014", "Sat,Jun28,2014", "Fri,Jun27,2014",
                 "Thu,Jun26,2014", "Wed,Jun25,2014", "Tue,Jun24,2014",
                 "Mon,Jun23,2014", "Sun,Jun22,2014", "Sat,Jun21,2014",
                 "Fri,Jun20,2014", "Thu,Jun19,2014", "Wed,Jun18,2014",
                 "Tue,Jun17,2014", "Mon,Jun16,2014", "Sun,Jun15,2014",
                 "Sat,Jun14,2014", "Fri,Jun13,2014", "Thu,Jun12,2014",
                 "Wed,Jun11,2014", "Tue,Jun10,2014", "Mon,Jun9,2014",
                 "Sun,Jun8,2014", "Sat,Jun7,2014", "Fri,Jun6,2014",
                 "Thu,Jun5,2014", "Wed,Jun4,2014", "Tue,Jun3,2014",
                 "Mon,Jun2,2014", "Sun,Jun1,2014", "Sat,May31,2014",
                 "Fri,May30,2014", "Thu,May29,2014", "Wed,May28,2014",
                 "Tue,May27,2014", "Mon,May26,2014", "Sun,May25,2014",
                 "Sat,May24,2014", "Fri,May23,2014", "Thu,May22,2014",
                 "Wed,May21,2014", "Tue,May20,2014", "Mon,May19,2014",
                 "Sun,May18,2014", "Sat,May17,2014")
```

a Find which winning numbers had payoffs greater than $350.00.

```
lotteryNumber[lotteryPayoff > 350]

## [1] 886 832 995 241  82 734
```

b On which date(s) was the winning number 000?

```
lotteryDate[lotteryNumber == 000]

## [1] "Wed,May28,2014"
```

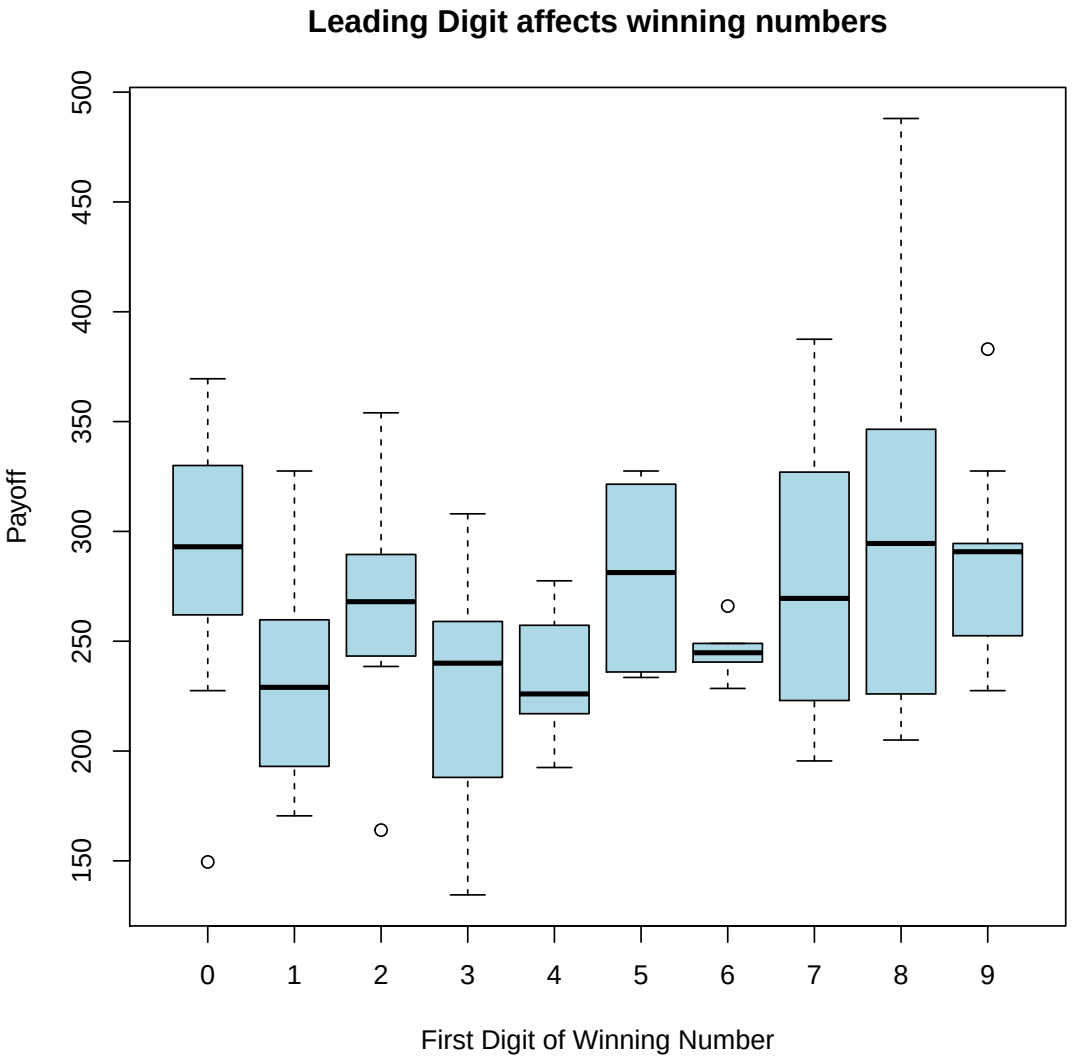
c Find the 10 smallest payoffs and the corresponding "unlucky" numbers.

```
lotteryNumber[order(lotteryPayoff[c(1:10)])]

##  [1] 369 817 250 157 530 273 783 947 367 997
```

d We want to determine whether the payoff depends on the first digit of the winning number. We can make side-by-side boxplots of the ten subsets of payoffs to study this phenomenon graphically. Rather than extracting each set separately, we use the R function split() to create a list, where

each element of the list gives all of the payoffs that correspond to a particular first digit of the winning number. The boxpot() function will then draw a box for each element in the list:

```
digit <- lotteryNumber %/% 100
boxplot(split(lotteryPayoff, digit), col = "lightblue",
        xlab = "First Digit of Winning Number", ylab = "Payoff",
        main = "Leading Digit affects winning numbers")
```

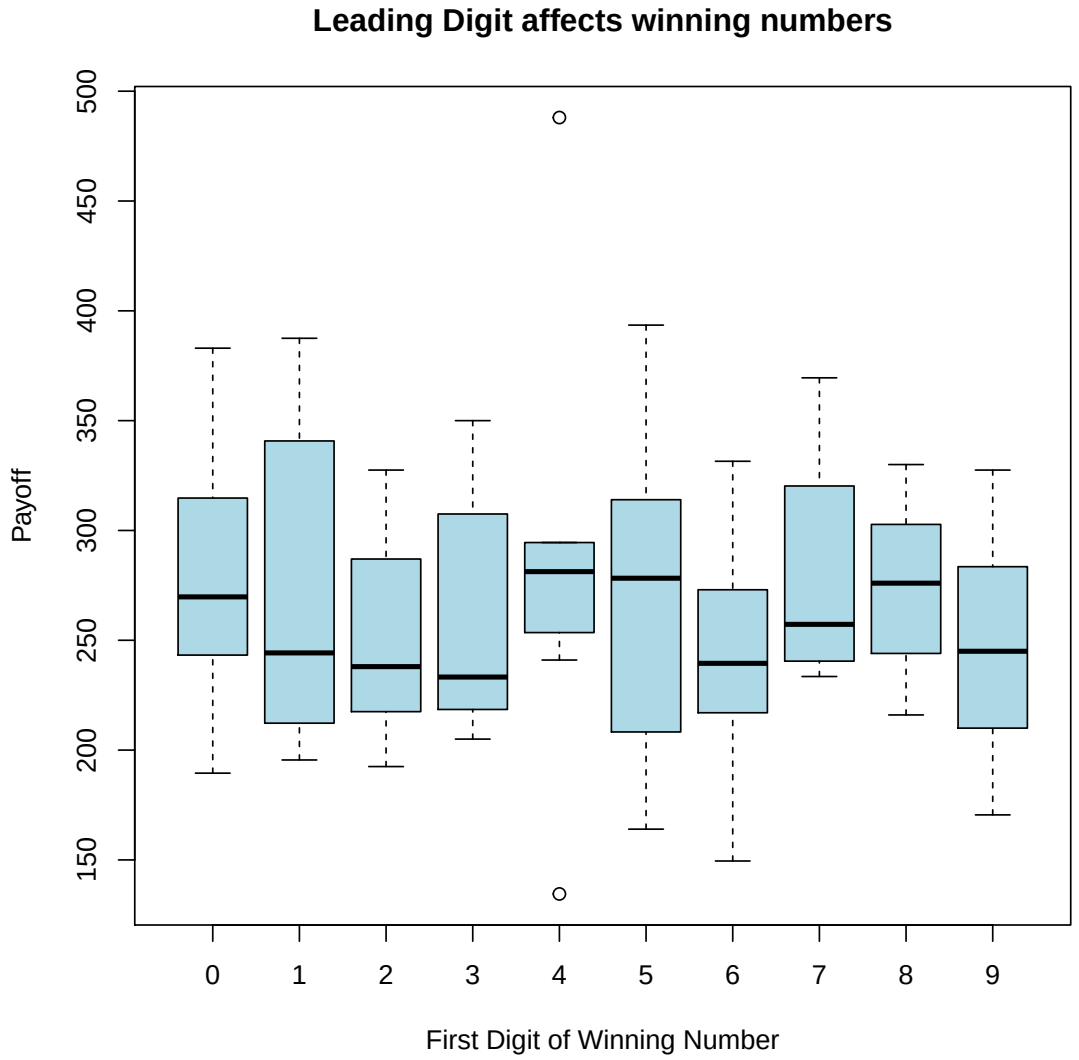


In the side-by-side boxplots above, what produces the labels under the boxes? Hint: Look at the value returned by split(lottery.payoff, digit).

The split function will split lotteryPayoff by the leading digit and that digit is the value that names the new list. This value is where the label comes from.

e In the side-by-side boxplots above, there appears to be an upward trend in payoff as the leading digit of the winning number increases from 1 to 9. Produce a similar figure, but moving the box representing the leading zero to the right of the side-by-side set of boxplots. Does it appear to fit in with the trend?

```
boxplot(split(lotteryPayoff, c(1:9, 0)), col = "lightblue",
        xlab = "First Digit of Winning Number", ylab = "Payoff",
        main = "Leading Digit affects winning numbers")
```



Yes it seems to fit the trend.