Problem 8

```r
college = read.csv("College.csv")
rownames(college) <- college[,1]
```

```r
college <- college[,-1]
```

```r
summary(college)
```
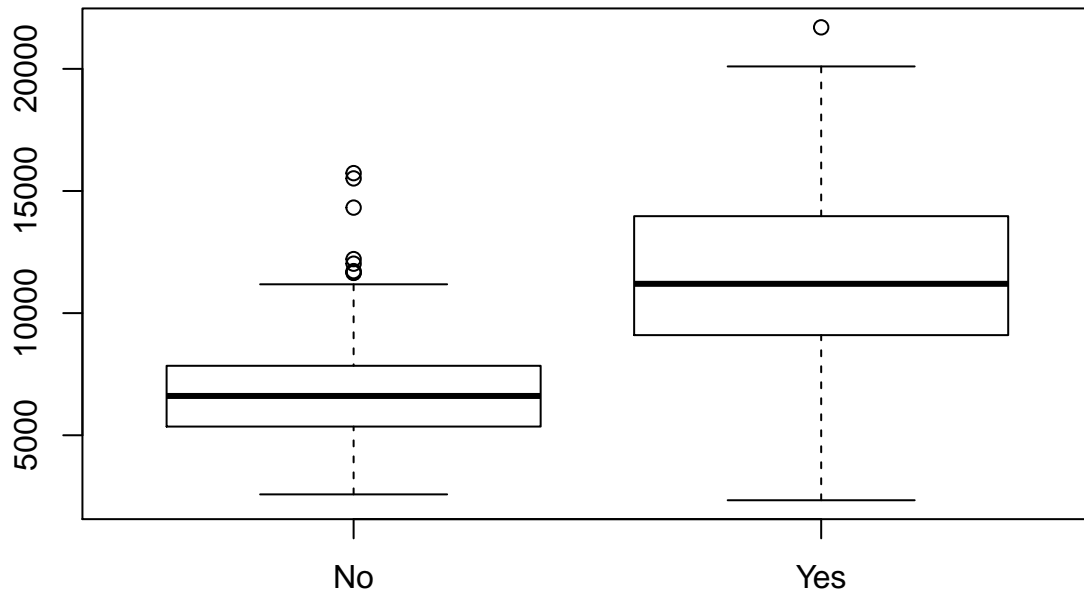
```
##  Private        Apps            Accept          Enroll        Top10perc
##  No :212   Min.   :   81   Min.   :    72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:   604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median :  1110   Median : 434   Median :23.00
##            Mean   : 3002   Mean   :  2019   Mean   : 780   Mean   :27.56
##            3rd Qu.: 3624   3rd Qu.:  2424   3rd Qu.: 902   3rd Qu.:35.00
##            Max.   :48094   Max.   : 26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad        Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books          Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00
##    Terminal        S.F.Ratio       perc.alumni        Expend
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233
##    Grad.Rate
##  Min.   : 10.00
##  1st Qu.: 53.00
##  Median : 65.00
##  Mean   : 65.46
##  3rd Qu.: 78.00
##  Max.   :118.00
```

```r
pairs(college[,1:10])
```
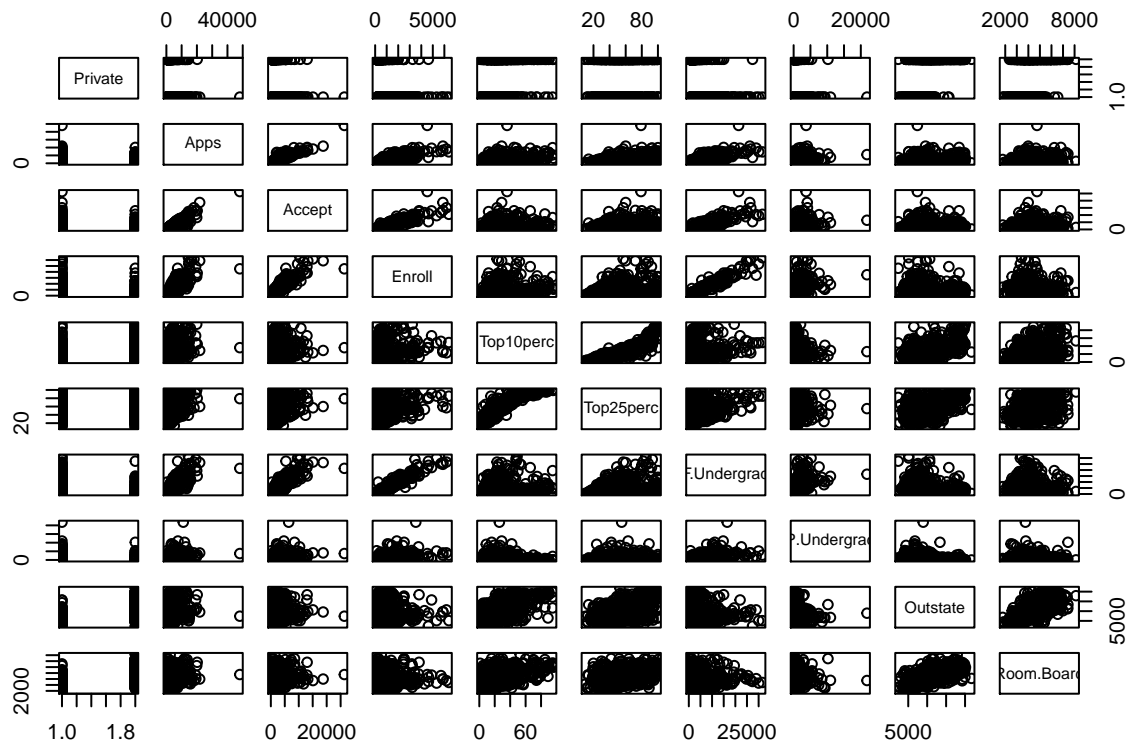


```r
boxplot(college$Outstate~college$Private)
```

```
Elite = rep("No",nrow(college))
Elite[college$Top10perc > 50] = "Yes"
college = data.frame(college,Elite)
summary(college)
```

```
##  Private        Apps           Accept          Enroll       Top10perc
##  No :212   Min.   :   81   Min.   :    72   Min.   :  35   Min.   : 1.00
##  Yes:565   1st Qu.:  776   1st Qu.:   604   1st Qu.: 242   1st Qu.:15.00
##            Median : 1558   Median :  1110   Median : 434   Median :23.00
##            Mean   : 3002   Mean   :  2019   Mean   : 780   Mean   :27.56
##            3rd Qu.: 3624   3rd Qu.:  2424   3rd Qu.: 902   3rd Qu.:35.00
##            Max.   :48094   Max.   : 26330   Max.   :6392   Max.   :96.00
##    Top25perc      F.Undergrad     P.Undergrad         Outstate
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700
##    Room.Board       Books          Personal         PhD
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   : 8.00
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.:62.00
##  Median :4200   Median : 500.0   Median :1200   Median :75.00
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   :72.66
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.:85.00
```

```
## Max.   :8124    Max.    :2340.0    Max.    :6800    Max.    :103.00
##     Terminal         S.F.Ratio        perc.alumni        Expend
## Min.   : 24.0    Min.    : 2.50    Min.    : 0.00    Min.    : 3186
## 1st Qu.: 71.0    1st Qu.:11.50    1st Qu.:13.00    1st Qu.: 6751
## Median : 82.0    Median :13.60    Median :21.00    Median : 8377
## Mean   : 79.7    Mean    :14.09    Mean    :22.74    Mean    : 9660
## 3rd Qu.: 92.0    3rd Qu.:16.50    3rd Qu.:31.00    3rd Qu.:10830
## Max.   :100.0    Max.    :39.80    Max.    :64.00    Max.    :56233
##    Grad.Rate        Elite
## Min.   : 10.00    No :699
## 1st Qu.: 53.00    Yes: 78
## Median : 65.00
## Mean   : 65.46
## 3rd Qu.: 78.00
## Max.   :118.00
```

```r
boxplot(college$Outstate~college$Elite)
```
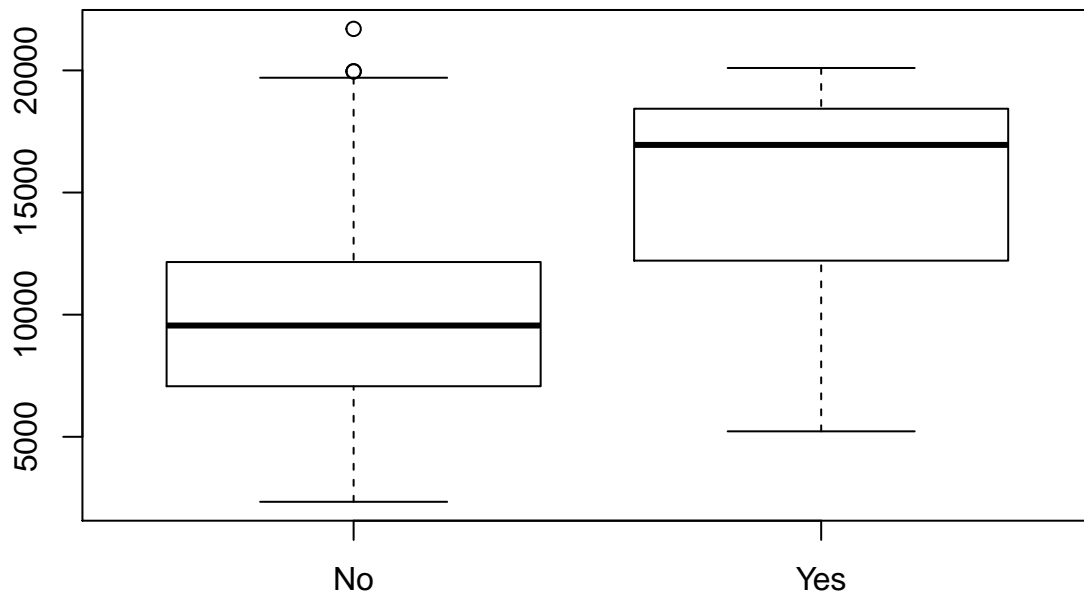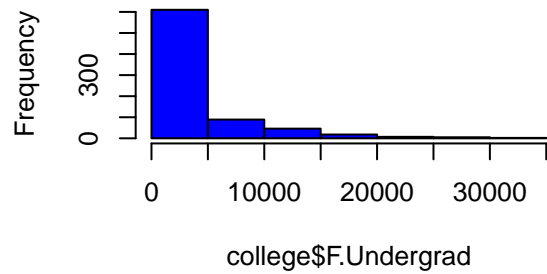


```r
par(mfrow = c(2,2))
hist(college$Outstate,col = 'red')
hist(college$F.Undergrad,col = 'blue')
hist(college$Accept,col = 'black')
hist(college$Room.Board)
```
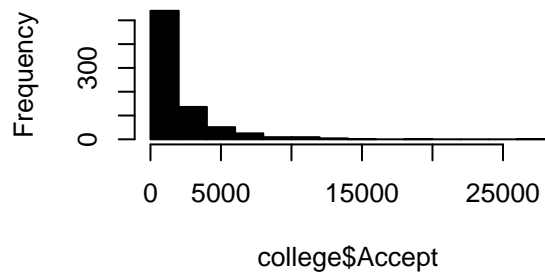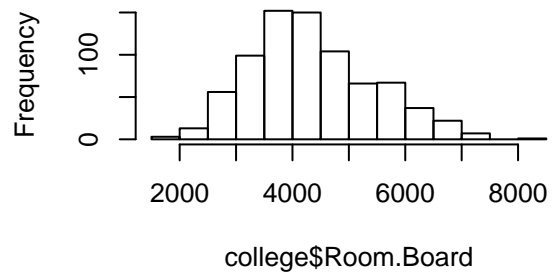
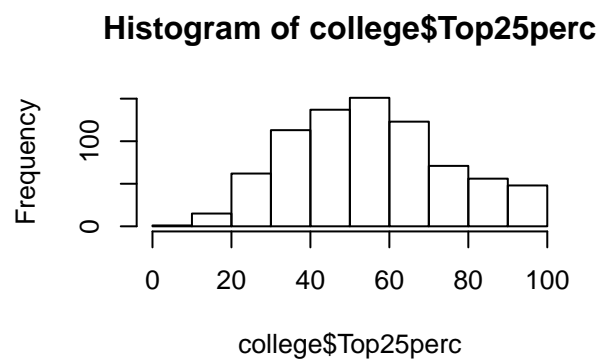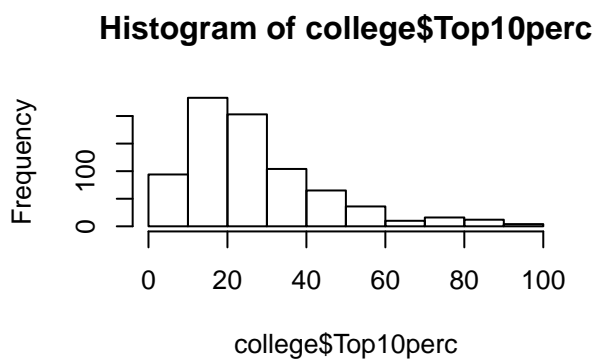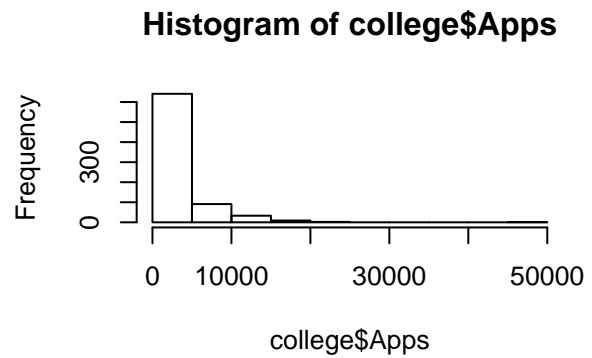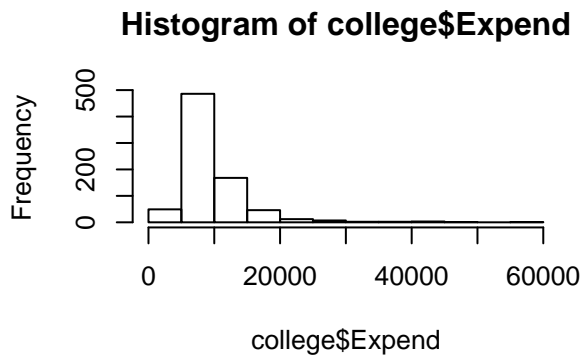## Histogram of college$Outstate
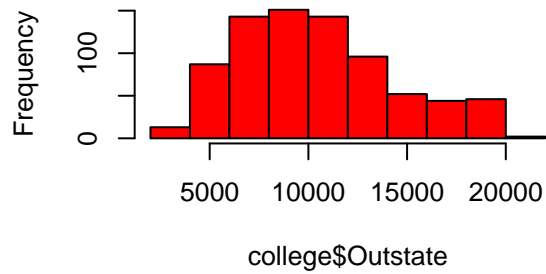
Histogram of college$F.Undergrad

Histogram of college$Accept

Histogram of college$Room.Board

```r
par(mfrow = c(2,2))
hist(college$Expend)
hist(college$Apps)
hist(college$Top10perc)
hist(college$Top25perc)
```

Histogram of college$Expend

Histogram of college$Apps

Histogram of college$Top10perc

Histogram of college$Top25perc

Problem 9

```r
auto <- read.csv("Auto.csv")
library(dplyr)
```

```
## 
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
## 
##     filter, lag
```

```
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

```r
glimpse(auto)
```

```
## Observations: 397
## Variables: 9
## $ mpg          <dbl> 18, 15, 18, 16, 17, 15, 14, 14, 14, 15, 15, 14, 15, 14...
```

```
## $ cylinders    <int> 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 4, 6, 6, 6, ...
## $ displacement <dbl> 307, 350, 318, 304, 302, 429, 454, 440, 455, 390, 383,...
## $ horsepower   <fct> 130, 165, 150, 150, 140, 198, 220, 215, 225, 190, 170,...
## $ weight       <int> 3504, 3693, 3436, 3433, 3449, 4341, 4354, 4312, 4425, ...
## $ acceleration <dbl> 12.0, 11.5, 11.0, 12.0, 10.5, 10.0, 9.0, 8.5, 10.0, 8....
## $ year         <int> 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70...
## $ origin       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, ...
## $ name         <fct> chevrolet chevelle malibu, buick skylark 320, plymouth...
```

```r
cat("Range for mpg:",range(auto$mpg))
```

```
## Range for mpg: 9 46.6
```

```r
cat("\nRange for displacement:",range(auto$displacement))
```

```
##
## Range for displacement: 68 455
```

```r
cat("\nRange for weight:",range(auto$weight))
```

```
##
## Range for weight: 1613 5140
```

```r
cat("\nRange for acceleration:", range(auto$acceleration))
```

```
##
## Range for acceleration: 8 24.8
```

```r
cat("\nRange for year:",range(auto$year))
```

```
##
## Range for year: 70 82
```

```r
cat("\nRange for origin:",range(auto$origin))
```

```
##
## Range for origin: 1 3
```

```r
cat("mpg Mean: ",mean(auto$mpg),"Std Dev.: ",sqrt(var(auto$mpg)))
```

```
## mpg Mean:  23.51587 Std Dev.:  7.825804
```

```r
cat("\ndisplacement Mean: ",mean(auto$displacement),"Std Dev.: ",sqrt(var(auto$displacement)))
```

```
##
## displacement Mean:  193.5327 Std Dev.:  104.3796
```

```r
cat("\nweight Mean: ",mean(auto$weight),"Std Dev.: ",sqrt(var(auto$weight)))
```

```
##
## weight Mean:  2970.262 Std Dev.:  847.9041
```

```r
cat("\nacceleration Mean: ",mean(auto$acceleration),"Std Dev.: ",sqrt(var(auto$acceleration)))
```

```
##
## acceleration Mean:  15.55567 Std Dev.:  2.749995
```

```r
cat("\nyear Mean: ",mean(auto$year),"Std Dev.: ",sqrt(var(auto$year)))
```

```
##
## year Mean:  75.99496 Std Dev.:  3.690005
```

```r
cat("\norigin Mean: ",mean(auto$origin),"Std Dev.: ",sqrt(var(auto$origin)))
```

```
##
## origin Mean:  1.574307 Std Dev.:  0.8025495
```

```r
auto <- auto[-c(10:85),]
cat("mpg Mean: ",mean(auto$mpg),"Std Dev.: ",sqrt(var(auto$mpg)))
```

```
## mpg Mean:  24.43863 Std Dev.:  7.908184
```

```r
cat("\ndisplacement Mean: ",mean(auto$displacement),"Std Dev.: ",sqrt(var(auto$displacement)))
```

```
##
## displacement Mean:  187.0498 Std Dev.:  99.63539
```

```r
cat("\nweight Mean: ",mean(auto$weight),"Std Dev.: ",sqrt(var(auto$weight)))
```

```
##
## weight Mean:  2933.963 Std Dev.:  810.6429
```

```r
cat("\nacceleration Mean: ",mean(auto$acceleration),"Std Dev.: ",sqrt(var(auto$acceleration)))
```

```
##
## acceleration Mean:  15.72305 Std Dev.:  2.680514
```

```r
cat("\nyear Mean: ",mean(auto$year),"Std Dev.: ",sqrt(var(auto$year)))
```

```
##
## year Mean:  77.15265 Std Dev.:  3.11123
```
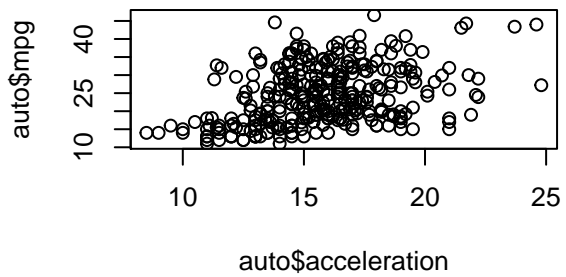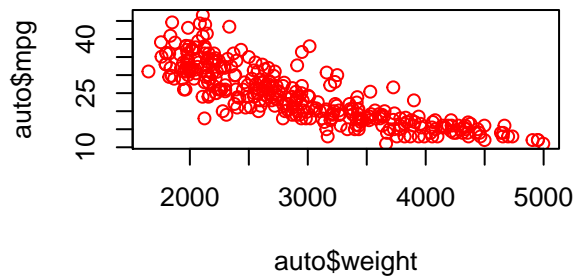
```r
cat("\norigin Mean: ",mean(auto$origin),"Std Dev.: ",sqrt(var(auto$origin)))
```
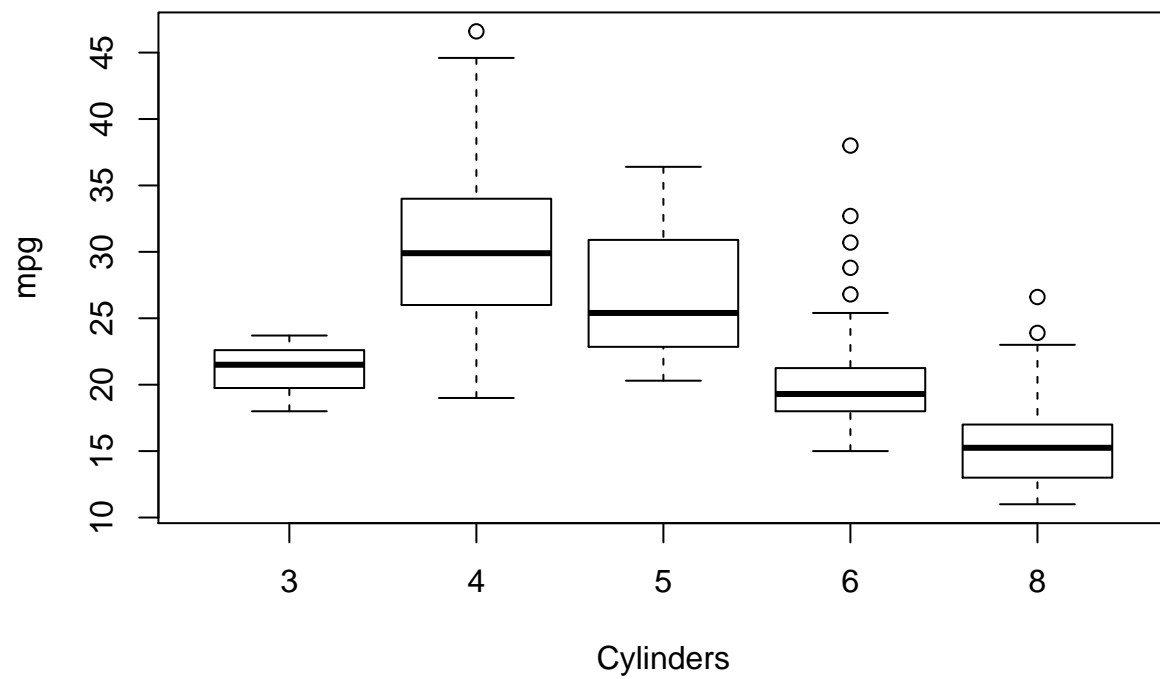
```
##
## origin Mean:  1.598131 Std Dev.:  0.8161627
```
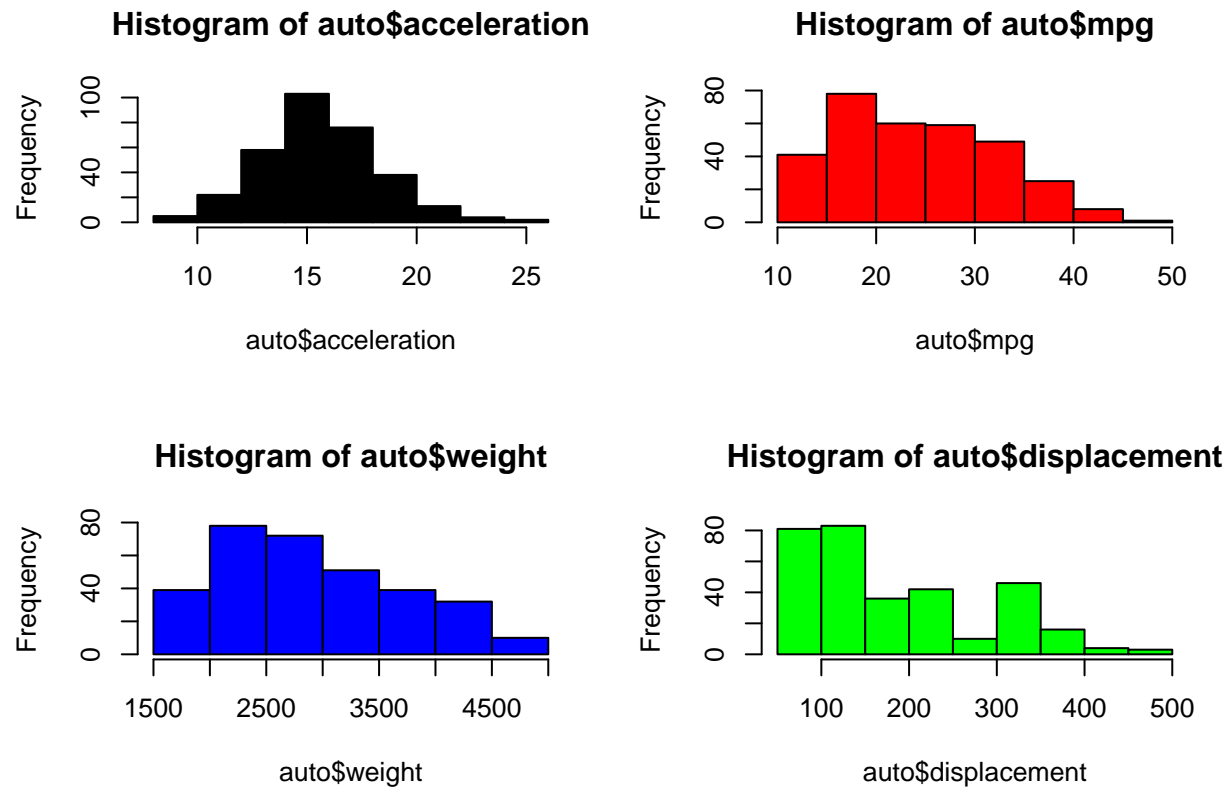
```r
par(mfrow = c(2,2))
plot(auto$weight,auto$mpg,col ='red')
plot(auto$acceleration,auto$mpg,col = 'black')
plot(auto$displacement,auto$mpg,col = 'green')
plot(auto$year,auto$mpg,col = 'orange')
```
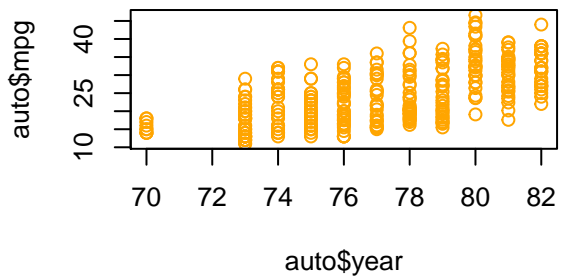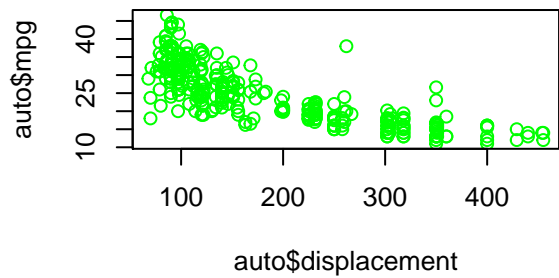


```r
boxplot(auto$mpg~auto$cylinders,xlab = "Cylinders",ylab= "mpg")
```

```
par(mfrow = c(2,2))
hist(auto$acceleration,col = 'black')
hist(auto$mpg,col = 'red')
hist(auto$weight,col = 'blue')
hist(auto$displacement,col = 'green')
```

Problem 10

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
boston <- Boston
cat("no. of rows:",nrow(Boston))
```

```
## no. of rows: 506
```
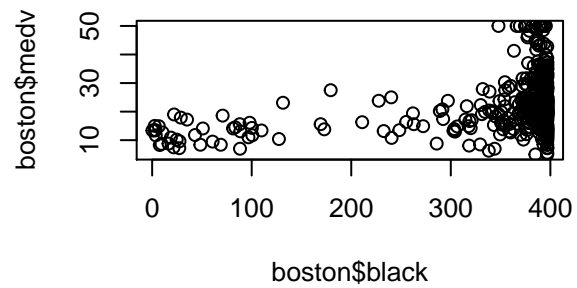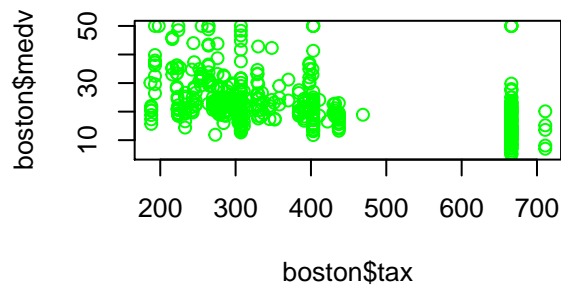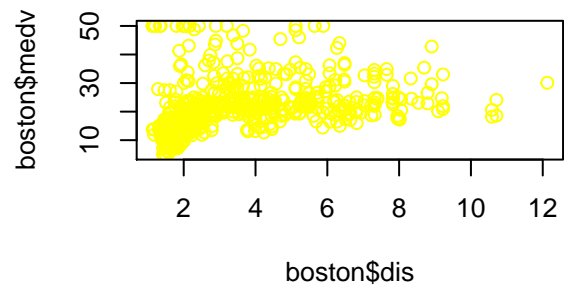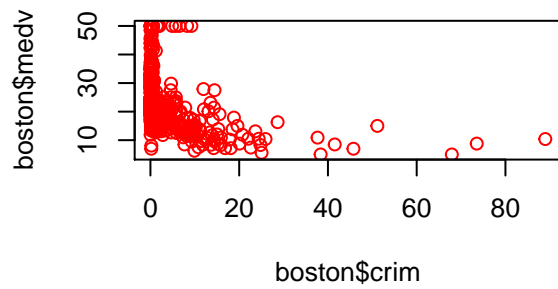
```r
cat("\nno. of cols:",ncol(Boston))
```

```
##
## no. of cols: 14
```

```
par(mfrow = c(2,2))
plot(boston$crim,boston$medv,col = 'red')
plot(boston$dis,boston$medv,col = 'yellow')
plot(boston$tax,boston$medv,col = 'green')
plot(boston$black,boston$medv)
```
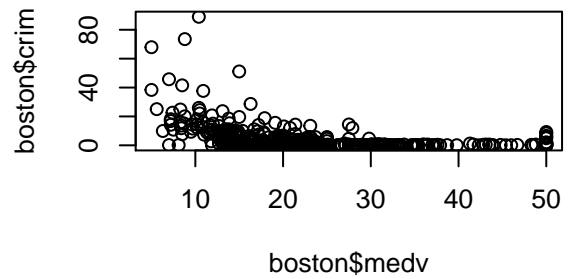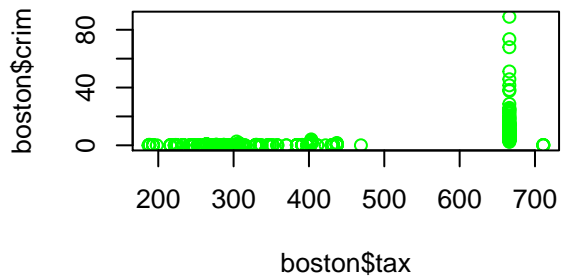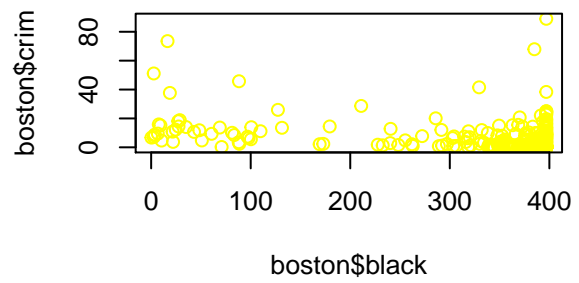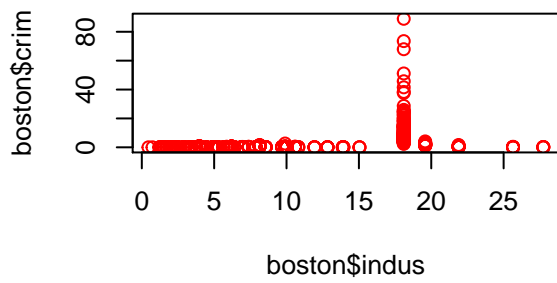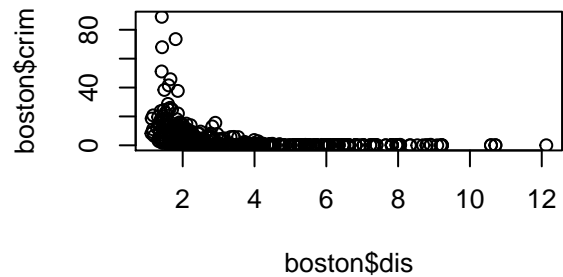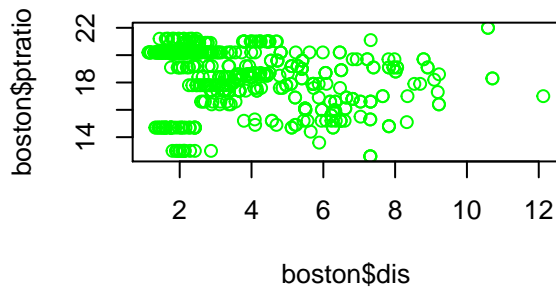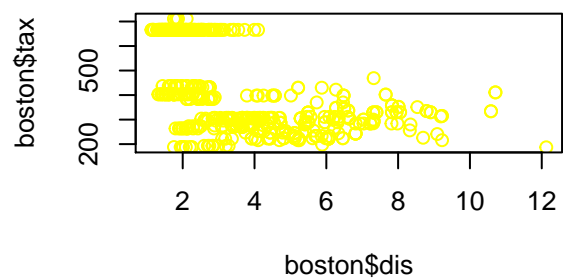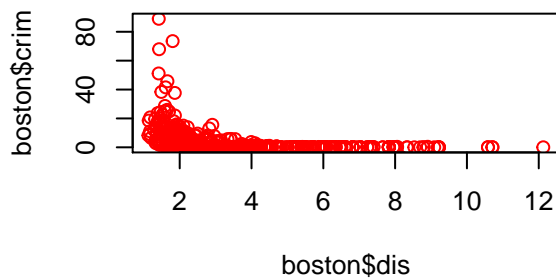


```
par(mfrow = c(2,2))
plot(boston$indus,boston$crim,col = 'red')
plot(boston$black,boston$crim,col = 'yellow')
plot(boston$tax,boston$crim,col = 'green')
plot(boston$medv,boston$crim)
```

```
par(mfrow = c(2,2))
plot(boston$dis,boston$crim,col = 'red')
plot(boston$dis,boston$tax,col = 'yellow')
plot(boston$dis,boston$ptratio,col = 'green')
plot(boston$dis,boston$crim)
```

```
boston %>%
  filter(boston$chas == 1) %>%
    nrow()
```

[1] 35

```
cat("Median pupil-teacher ratio:",median(boston$ptratio))
```

Median pupil-teacher ratio: 19.05

```
boston2 <- boston %>%
  arrange(boston$medv)
head(boston2, n =10)
```

```
##           crim zn indus chas   nox    rm   age    dis rad tax ptratio  black lstat
## 1   38.35180  0 18.10    0 0.693 5.453 100.0 1.4896  24 666    20.2 396.90 30.59
## 2   67.92080  0 18.10    0 0.693 5.683 100.0 1.4254  24 666    20.2 384.97 22.98
## 3   25.04610  0 18.10    0 0.693 5.987 100.0 1.5888  24 666    20.2 396.90 26.77
## 4    9.91655  0 18.10    0 0.693 5.852  77.8 1.5004  24 666    20.2 338.16 29.97
## 5   45.74610  0 18.10    0 0.693 4.519 100.0 1.6582  24 666    20.2  88.27 36.98
## 6    0.18337  0 27.74    0 0.609 5.414  98.3 1.7554   4 711    20.1 344.05 23.97
## 7   16.81180  0 18.10    0 0.700 5.277  98.1 1.4261  24 666    20.2 396.90 30.81
## 8   14.23620  0 18.10    0 0.693 6.343 100.0 1.5741  24 666    20.2 396.90 20.32
## 9   18.08460  0 18.10    0 0.679 6.434 100.0 1.8347  24 666    20.2  27.25 29.05
```

```
## 10 22.59710  0 18.10    0 0.700 5.000  89.5 1.5184  24 666    20.2 396.90 31.99
##     medv
## 1    5.0
## 2    5.0
## 3    5.6
## 4    6.3
## 5    7.0
## 6    7.0
## 7    7.2
## 8    7.2
## 9    7.2
## 10   7.4
```

```r
boston %>%
  filter(boston$rm > 7) %>%
    nrow()
```

```
## [1] 64
```

```r
boston %>%
  filter(boston$rm > 8)
```

```
##         crim zn indus chas    nox    rm  age    dis rad tax ptratio  black lstat
## 1    0.12083  0  2.89    0 0.4450 8.069 76.0 3.4952   2 276    18.0 396.90  4.21
## 2    1.51902  0 19.58    1 0.6050 8.375 93.9 2.1620   5 403    14.7 388.45  3.32
## 3    0.02009 95  2.68    0 0.4161 8.034 31.9 5.1180   4 224    14.7 390.55  2.88
## 4    0.31533  0  6.20    0 0.5040 8.266 78.3 2.8944   8 307    17.4 385.05  4.14
## 5    0.52693  0  6.20    0 0.5040 8.725 83.0 2.8944   8 307    17.4 382.00  4.63
## 6    0.38214  0  6.20    0 0.5040 8.040 86.5 3.2157   8 307    17.4 387.38  3.13
## 7    0.57529  0  6.20    0 0.5070 8.337 73.3 3.8384   8 307    17.4 385.91  2.47
## 8    0.33147  0  6.20    0 0.5070 8.247 70.4 3.6519   8 307    17.4 378.95  3.95
## 9    0.36894 22  5.86    0 0.4310 8.259  8.4 8.9067   7 330    19.1 396.90  3.54
## 10 0.61154 20  3.97    0 0.6470 8.704 86.9 1.8010   5 264    13.0 389.70  5.12
## 11 0.52014 20  3.97    0 0.6470 8.398 91.5 2.2885   5 264    13.0 386.86  5.91
## 12 0.57834 20  3.97    0 0.5750 8.297 67.0 2.4216   5 264    13.0 384.54  7.44
## 13 3.47428  0 18.10    1 0.7180 8.780 82.9 1.9047  24 666    20.2 354.55  5.29
##     medv
## 1   38.7
## 2   50.0
## 3   50.0
## 4   44.8
## 5   50.0
## 6   37.6
## 7   41.7
## 8   48.3
## 9   42.8
## 10 50.0
## 11 48.8
## 12 50.0
## 13 21.9
```