CE Assignment 1

James Lopez

October 6, 2016

9 This question uses the Auto data set from the ISLR package.

a) Which of the predictors are quantitative, and which are qualitative?

```
library(ISLR)
attach(Auto)
?Auto
```

```
## starting httpd help server ...
```

```
##  done
```

The quantitative predictors seem to be mpg, cylinders, displacement, horsepower, weight, and acceleration. Though cylinders covers a small range and comes in integers, the number of cylinders affects displacement and horsepower, so I deemed it quantitative.

The qualitative predictors seem to be year, origin, and name. The numbers of origin indicate the car's country of origin, therefore making origin qualitative.

b) What is the range of each quantitative predictor? You can answer this using the range() function.

```
range(Auto[,'mpg'], na.rm=TRUE)
```

```
## [1]  9.0 46.6
```

```
range(Auto[,'cylinders'], na.rm=TRUE)
```

```
## [1] 3 8
```

```
range(Auto[,'displacement'], na.rm=TRUE)
```

```
## [1]  68 455
```

```
range(Auto[,'horsepower'], na.rm=TRUE)
```

```
## [1]  46 230
```

```r
range(Auto[,'weight'], na.rm=TRUE)
```

```
## [1] 1613 5140
```

```r
range(Auto[,'acceleration'], na.rm=TRUE)
```

```
## [1]  8.0 24.8
```

c) What is the mean and standard deviation of each quantitative predictor?

To display the means and standard deviations of the predictors, I created a table named 'automeansd':

```r
automeansd <- matrix(NA, nrow = 6, ncol = 2)
rownames(automeansd) <- c(names(Auto[,1:6]))
colnames(automeansd) <- c('mean', 'sd')

for(i in 1:6) {
  automeansd[i,1] <- mean(Auto[,i])
  automeansd[i,2] <- sd(Auto[,i])
 i <- i+1
}
automeansd
```

```
##                     mean         sd
## mpg            23.445918   7.805007
## cylinders       5.471939   1.705783
## displacement  194.411990 104.644004
## horsepower    104.469388  38.491160
## weight       2977.584184 849.402560
## acceleration   15.541327   2.758864
```

d) Now remove the 10th through 85th observations. What is therange, mean, and standard deviation of each predictor in the subset of the data that remains?

To display the new ranges, means, and sd's, I first stored the updated Auto data set with the 10th thru 85th rows removed into a new data set 'auto2', then I created a table with the new means and sd's. I simply ran ranges for each predictor afterwards to get the new ranges.

```r
auto2 <- Auto[-10:-85,]
auto2meansd <- matrix(NA, nrow=6, ncol=2)
rownames(auto2meansd) <- c(names(Auto[,1:6]))
colnames(auto2meansd) <- c('mean','sd')

for(i in 1:6) {
  auto2meansd[i,1] <- mean(auto2[,i])
```

```r
  auto2meansd[i,2] <- sd(auto2[,i])
  i <- i+1
}

auto2meansd
```

```
##                      mean         sd
## mpg             24.404430   7.867283
## cylinders        5.373418   1.654179
## displacement   187.240506  99.678367
## horsepower     100.721519  35.708853
## weight        2935.971519 811.300208
## acceleration    15.726899   2.693721
```

```r
range(auto2[,'mpg'], na.rm=TRUE)
```

```
## [1] 11.0 46.6
```

```r
range(auto2[,'cylinders'], na.rm=TRUE)
```

```
## [1] 3 8
```

```r
range(auto2[,'displacement'], na.rm=TRUE)
```

```
## [1]  68 455
```

```r
range(auto2[,'horsepower'], na.rm=TRUE)
```

```
## [1]  46 230
```

```r
range(auto2[,'weight'], na.rm=TRUE)
```

```
## [1] 1649 4997
```

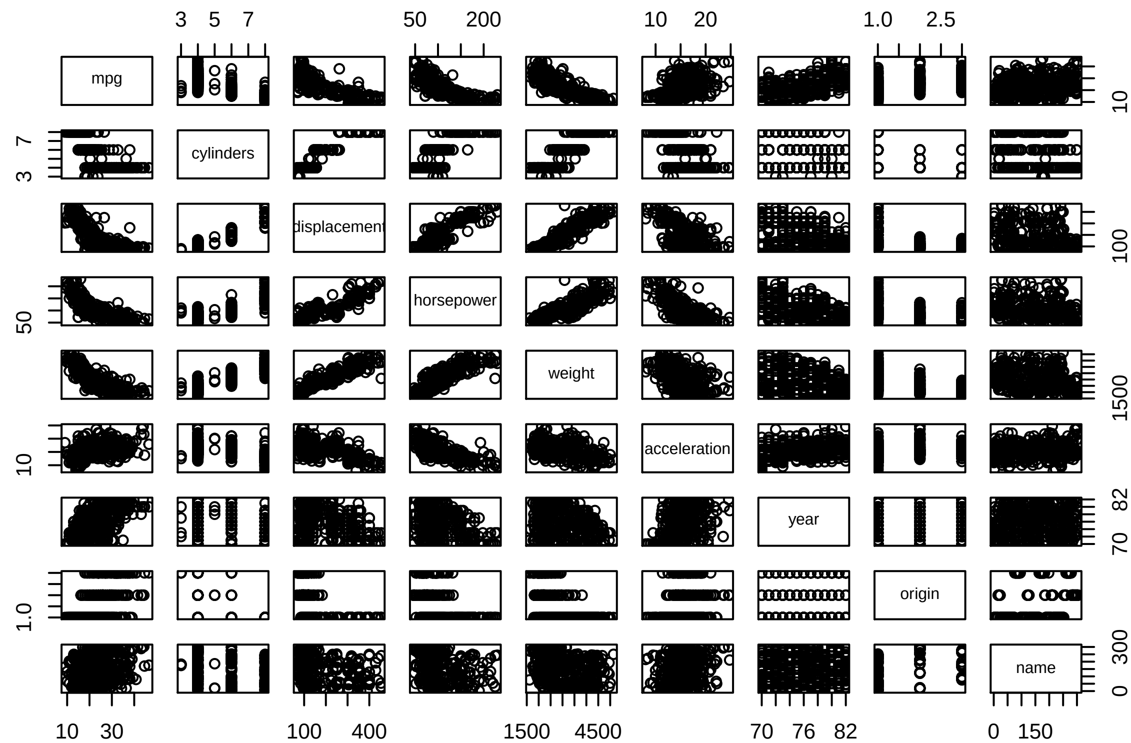
```r
range(auto2[,'acceleration'], na.rm=TRUE)
```

```
## [1]  8.5 24.8
```

e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.

I included a scatterplot matrix, some simple scatterplots, and some histograms for qualitative predictors. For quantitative vs. quantitative predictors, I simply referred to the scatterplot matrix. For quantitative vs. qualitative predictors, I created boxplots, though the scatterplot matrix also provided information.

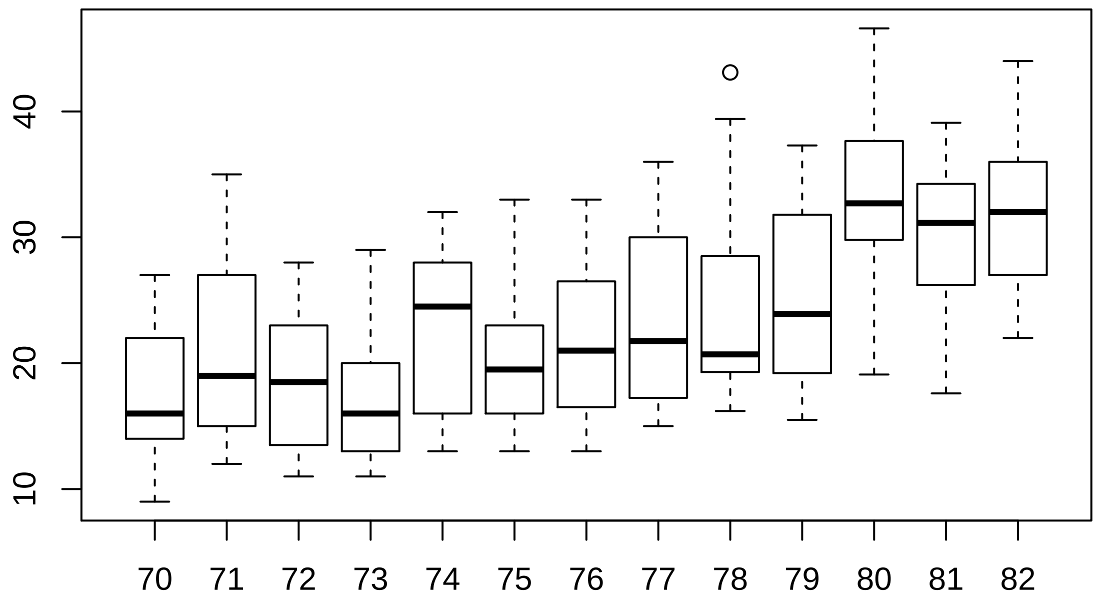
```
plotAuto <- pairs(Auto)
```




```
attach(plotAuto)
year = as.factor(Auto$year)
origin = as.factor(Auto$origin)
name = as.factor(Auto$name)
```

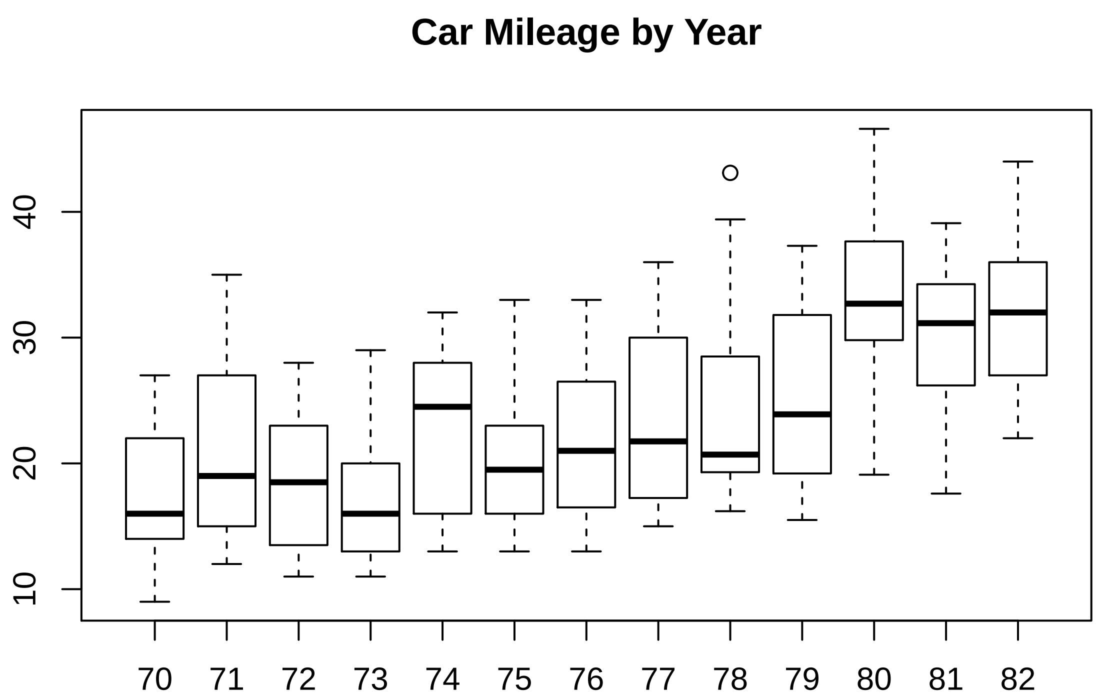
Since there are many combinations of boxplots I can create, I made a shorthand function that generates a boxplot if I input two Auto variables of my choice:

```
box <- function(x,y) {
  boxplot(y~x)
}
box(year, mpg)
```

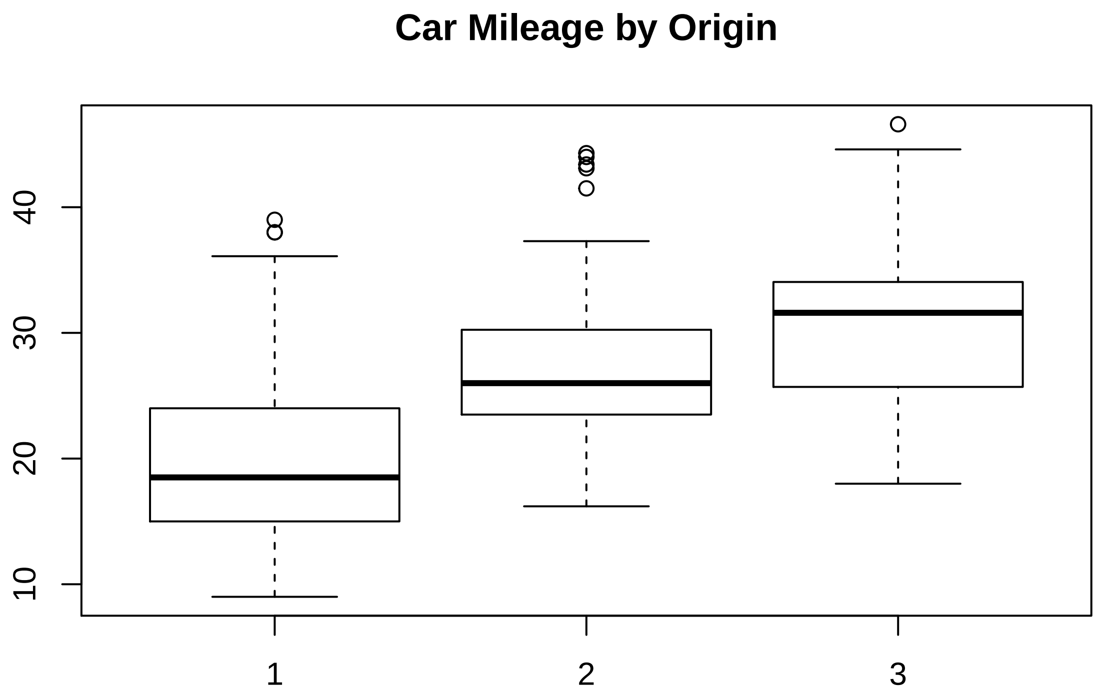


Alternatively, I could also do it the long way and make boxplots manually. I set titles for each boxplot to keep track.

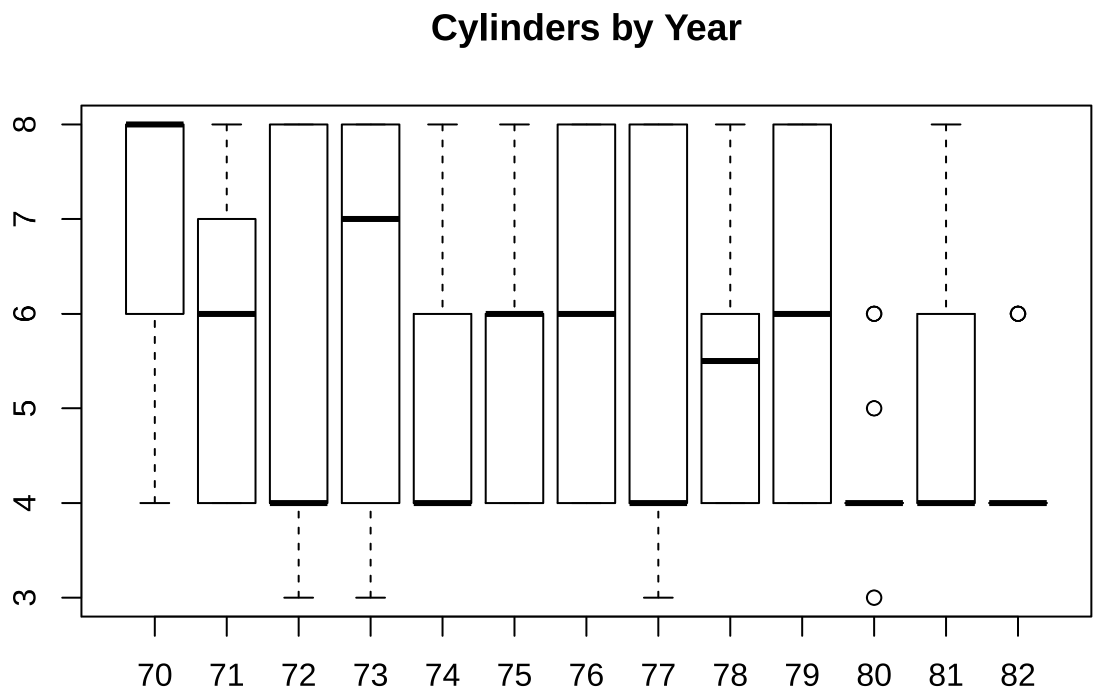
```r
boxplot(mpg~year, data=Auto, main='Car Mileage by Year')
```



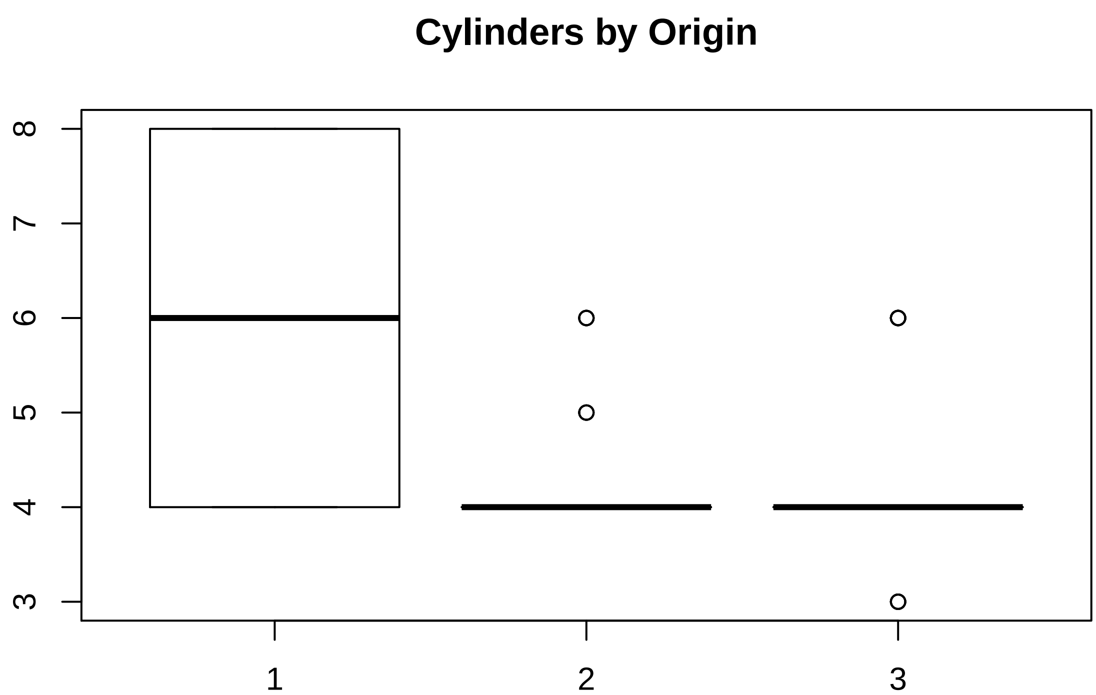

```
boxplot(mpg~origin, data=Auto, main='Car Mileage by Origin')
```



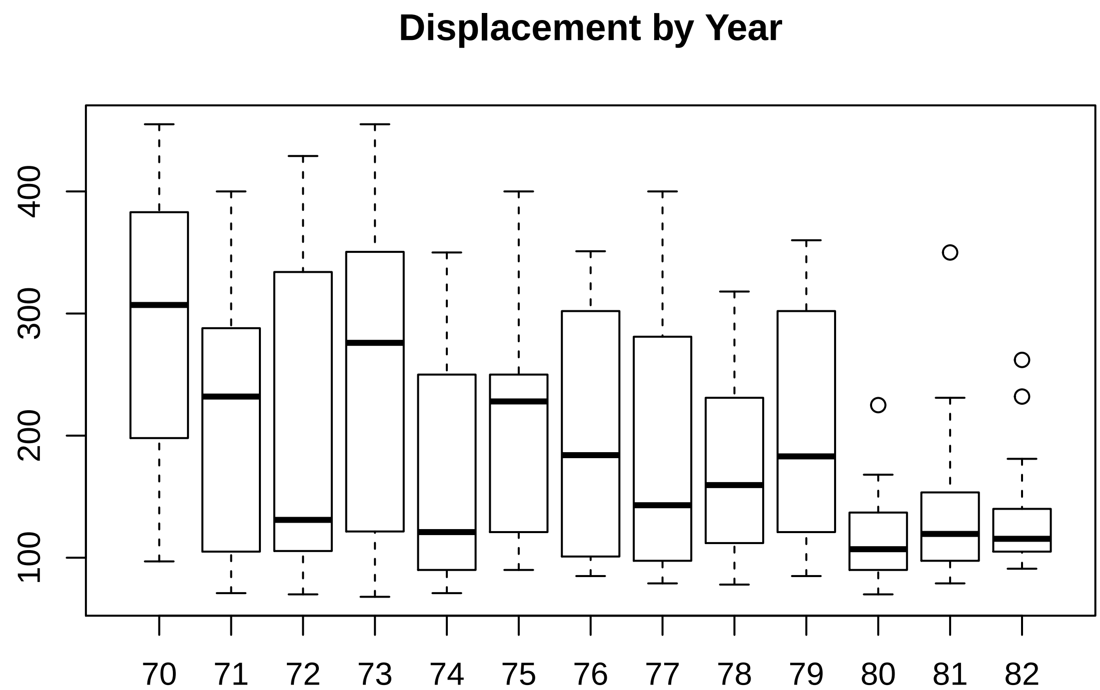

```r
boxplot(cylinders~year, data=Auto, main='Cylinders by Year')
```



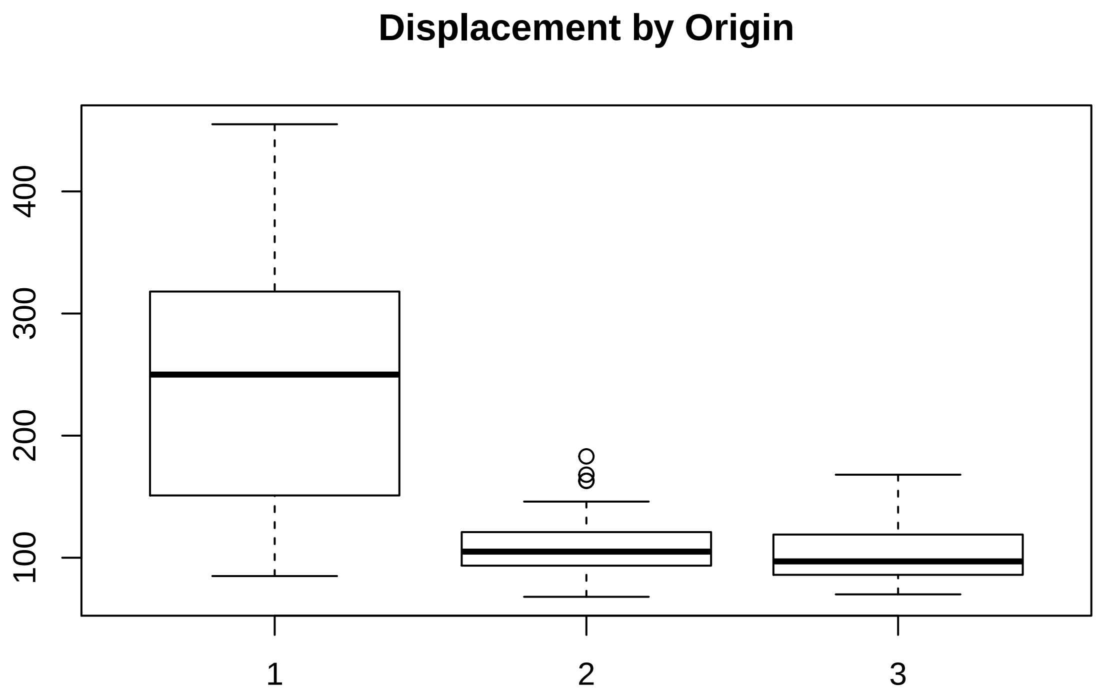

```r
boxplot(cylinders~origin, data=Auto, main='Cylinders by Origin')
```



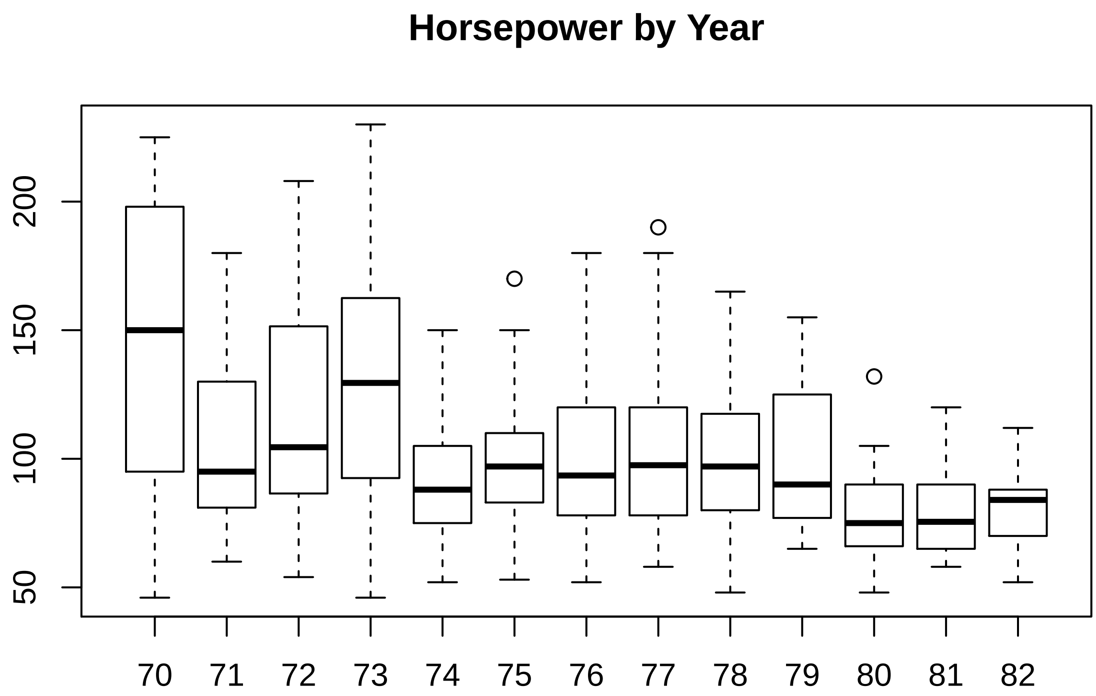

```r
boxplot(displacement~year, data=Auto, main='Displacement by Year')
```



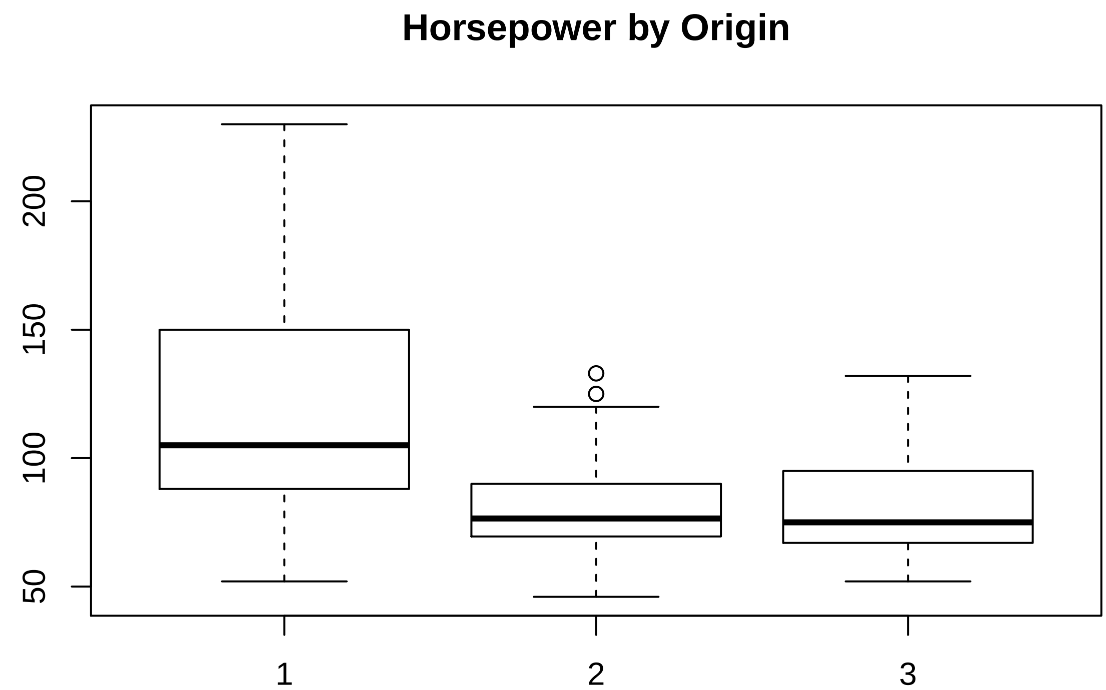

```r
boxplot(displacement~origin, data=Auto, main='Displacement by Origin')
```



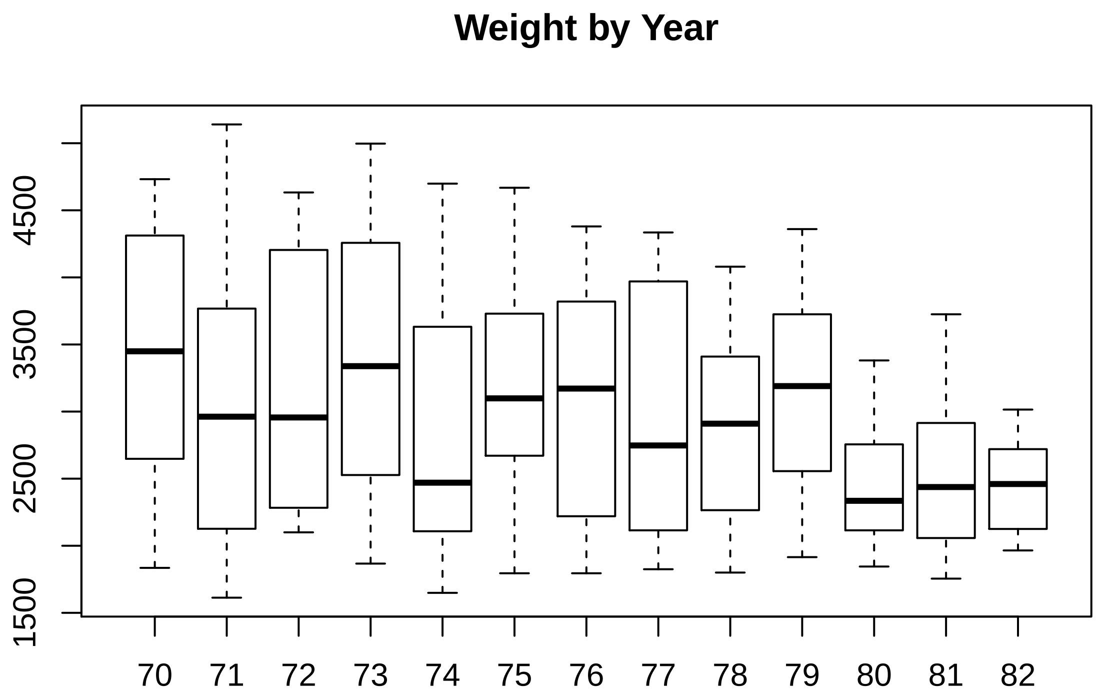

```r
boxplot(horsepower~year, data=Auto, main='Horsepower by Year')
```



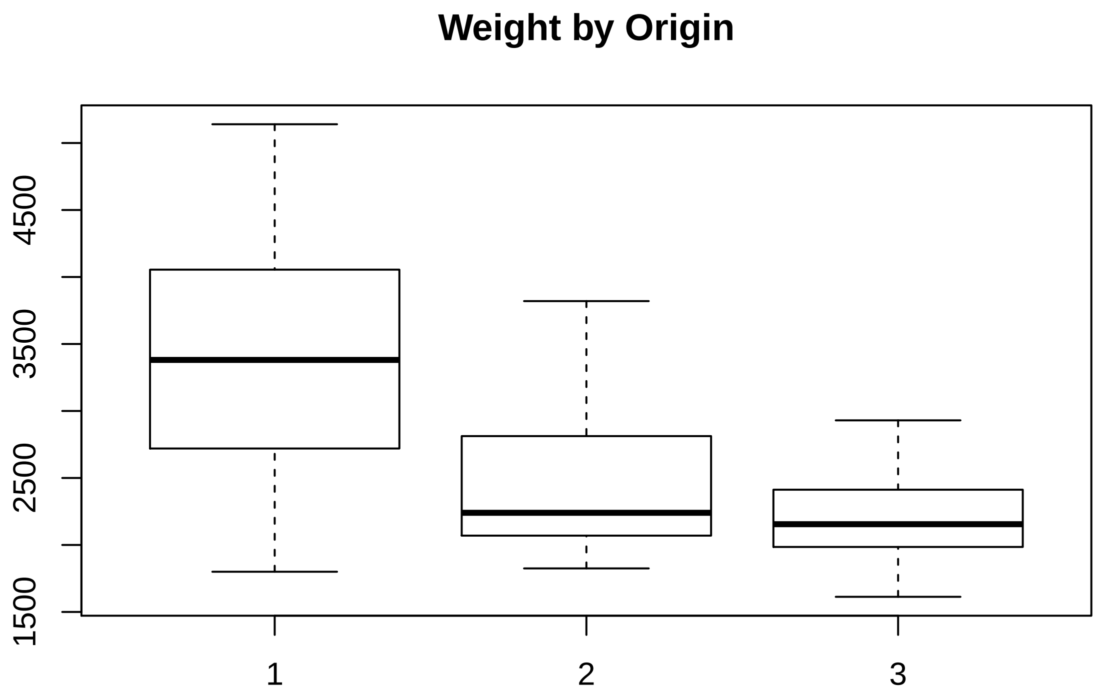

```
boxplot(horsepower~origin, data=Auto, main='Horsepower by Origin')
```



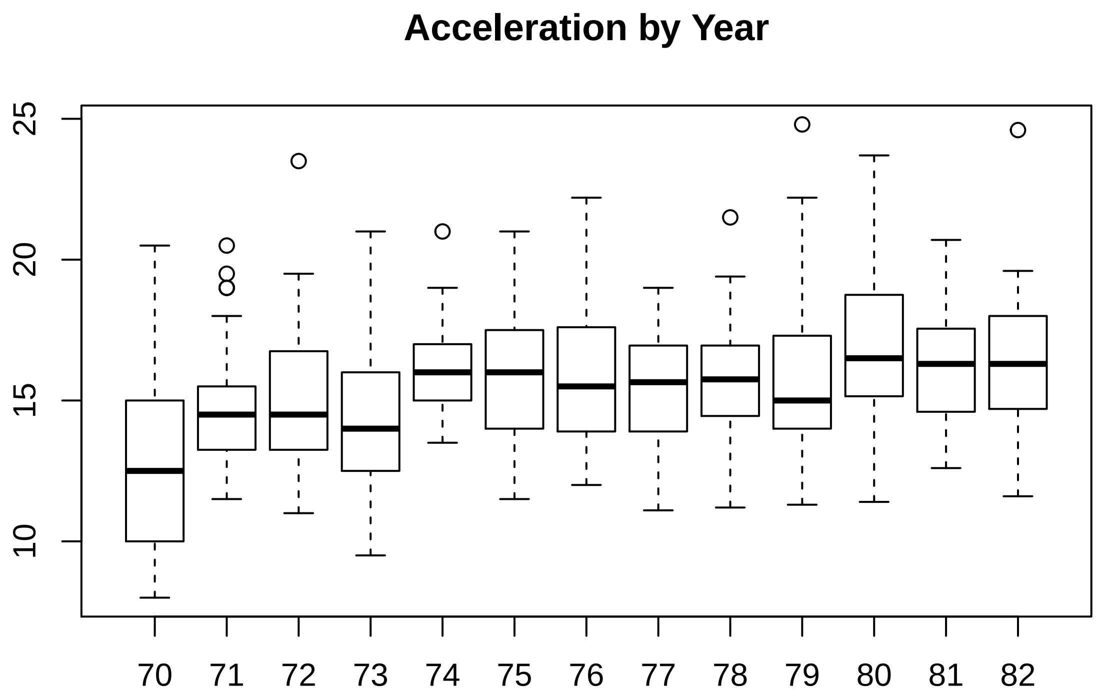

```
boxplot(weight~year, data=Auto, main='Weight by Year')
```



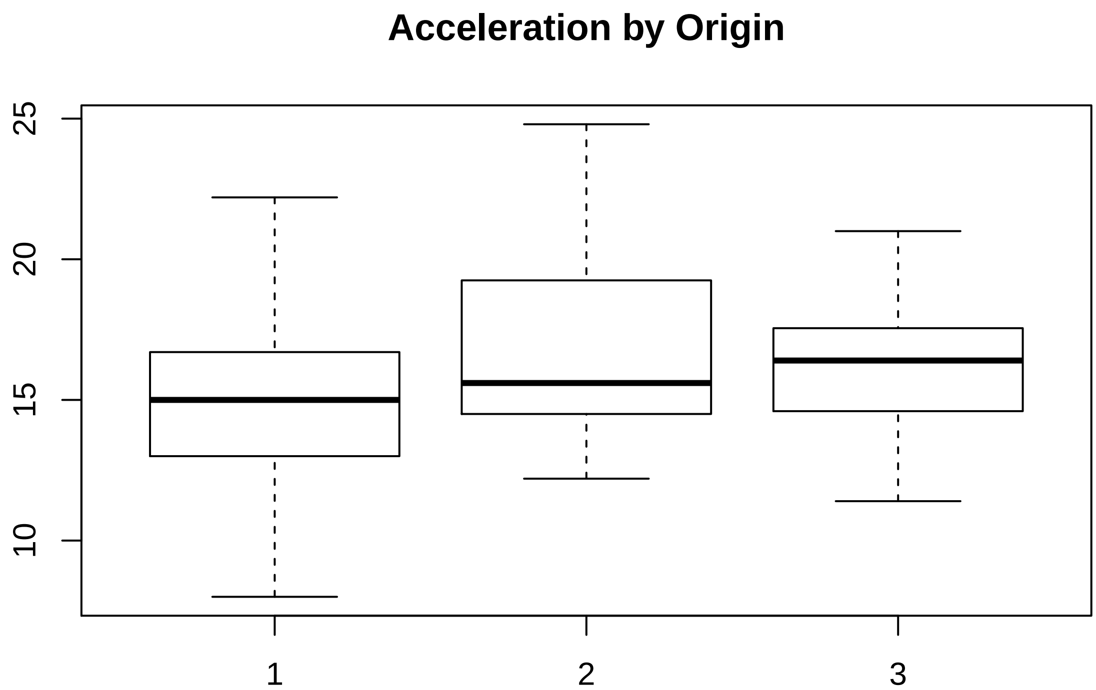

```r
boxplot(weight~origin, data=Auto, main='Weight by Origin')
```




```r
boxplot(acceleration~year, data=Auto, main='Acceleration by Year')
```




```r
boxplot(acceleration~origin, data=Auto, main='Acceleration by Origin')
```



Some of the patterns and trends include:

- **mpg:** increases with less cylinders, displacement, horsepower, and weight. Mpg slightly increases as year increases. There doesn't seem to be any reliable association between mpg and acceleration. Looking at the boxplot of mpg vs. year, there doesn't seem to be any association until 1980 models were made. It seems like 1980 models and newer had higher mean mpg's than models prior to 1980. European cars have higher average mpg than American cars, and Japanese cars have the highest average mpg. No reliable association with name.
- **cylinders:** more cylinders have more displacement, horsepower, and weight. No reliable association with acceleration or year. The vast majority of cars with cylinders are American, which explains why American cars had the lowest average mpg out of the 3 countries. No reliable association with year.
- **displacement:** the higher the displacement, the higher the horsepower and weight, and the lower the acceleration. Displacement seemed to decrease as newer models were made – especially after 1980. American cars had much higher displacement and cars with displacement than European or Japanese cars, which further helps explain the lower mpg for American cars.
- **horsepower:** cars with higher horsepower tended to have lower acceleration and be heavier. Horsepower seemed to decrease as models became newer, with more drastic decreases between '72 and '73, and between '79 and '80. American cars had by far the highest horsepower and highest frequency of high horsepower than European or Japanese cars, which corresponds with lower mpg.
- **weight:** heavier cars tended to slightly decrease in acceleration. No trends in weight throughout the years until 1980, when the weights noticeably decreased. American cars were the heaviest among the 3 origins, again corresponding to the lowest mpg of the 3.

- **acceleration:** No reliable trends in acceleration throughout the years – except that '70 cars were especially low in acceleration. The acceleration among the 3 origins was less different than the previous predictors, with each country's medians being fairly close. Europe and Japan, however, still had higher numbers of cars with greater accelerations than America.
- **name:** looking at the scatterplot matrix, name showed no reliable pattern with any predictor.

Though the span is only three years, the boxplots suggest a shift in car manufacturing after 1980.

f) Suppose that we wish to predict gas mileage (mpg) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer.

The predictors that seem useful for predicting mpg seem to be everything except year and name. The plots suggest that an event occured in 1980 that influenced many other variables that were mostly steady until '80. Also, none of the plots in the scatterplot matrix seemed to be influenced by name; the data points followed no ostensible pattern.

10 This exercise involves the Boston housing data set.

a) To begin, load in the Boston data set. The Boston data set is part of the MASS library in R. How many rows are in this data set? How many columns? What do the rows and columns represent?

```
library(MASS)
Boston
```

```
##         crim   zn indus chas    nox    rm   age     dis rad tax ptratio
## 1    0.00632 18.0  2.31    0 0.5380 6.575  65.2  4.0900   1 296    15.3
## 2    0.02731  0.0  7.07    0 0.4690 6.421  78.9  4.9671   2 242    17.8
## 3    0.02729  0.0  7.07    0 0.4690 7.185  61.1  4.9671   2 242    17.8
## 4    0.03237  0.0  2.18    0 0.4580 6.998  45.8  6.0622   3 222    18.7
## 5    0.06905  0.0  2.18    0 0.4580 7.147  54.2  6.0622   3 222    18.7
## 6    0.02985  0.0  2.18    0 0.4580 6.430  58.7  6.0622   3 222    18.7
## 7    0.08829 12.5  7.87    0 0.5240 6.012  66.6  5.5605   5 311    15.2
## 8    0.14455 12.5  7.87    0 0.5240 6.172  96.1  5.9505   5 311    15.2
## 9    0.21124 12.5  7.87    0 0.5240 5.631 100.0  6.0821   5 311    15.2
## 10   0.17004 12.5  7.87    0 0.5240 6.004  85.9  6.5921   5 311    15.2
## 11   0.22489 12.5  7.87    0 0.5240 6.377  94.3  6.3467   5 311    15.2
## 12   0.11747 12.5  7.87    0 0.5240 6.009  82.9  6.2267   5 311    15.2
## 13   0.09378 12.5  7.87    0 0.5240 5.889  39.0  5.4509   5 311    15.2
## 14   0.62976  0.0  8.14    0 0.5380 5.949  61.8  4.7075   4 307    21.0
## 15   0.63796  0.0  8.14    0 0.5380 6.096  84.5  4.4619   4 307    21.0
## 16   0.62739  0.0  8.14    0 0.5380 5.834  56.5  4.4986   4 307    21.0
## 17   1.05393  0.0  8.14    0 0.5380 5.935  29.3  4.4986   4 307    21.0
## 18   0.78420  0.0  8.14    0 0.5380 5.990  81.7  4.2579   4 307    21.0
## 19   0.80271  0.0  8.14    0 0.5380 5.456  36.6  3.7965   4 307    21.0
## 20   0.72580  0.0  8.14    0 0.5380 5.727  69.5  3.7965   4 307    21.0
```

```
## 21   1.25179   0.0  8.14     0 0.5380 5.570  98.1  3.7979   4 307    21.0
## 22   0.85204   0.0  8.14     0 0.5380 5.965  89.2  4.0123   4 307    21.0
## 23   1.23247   0.0  8.14     0 0.5380 6.142  91.7  3.9769   4 307    21.0
## 24   0.98843   0.0  8.14     0 0.5380 5.813 100.0  4.0952   4 307    21.0
## 25   0.75026   0.0  8.14     0 0.5380 5.924  94.1  4.3996   4 307    21.0
## 26   0.84054   0.0  8.14     0 0.5380 5.599  85.7  4.4546   4 307    21.0
## 27   0.67191   0.0  8.14     0 0.5380 5.813  90.3  4.6820   4 307    21.0
## 28   0.95577   0.0  8.14     0 0.5380 6.047  88.8  4.4534   4 307    21.0
## 29   0.77299   0.0  8.14     0 0.5380 6.495  94.4  4.4547   4 307    21.0
## 30   1.00245   0.0  8.14     0 0.5380 6.674  87.3  4.2390   4 307    21.0
## 31   1.13081   0.0  8.14     0 0.5380 5.713  94.1  4.2330   4 307    21.0
## 32   1.35472   0.0  8.14     0 0.5380 6.072 100.0  4.1750   4 307    21.0
## 33   1.38799   0.0  8.14     0 0.5380 5.950  82.0  3.9900   4 307    21.0
## 34   1.15172   0.0  8.14     0 0.5380 5.701  95.0  3.7872   4 307    21.0
## 35   1.61282   0.0  8.14     0 0.5380 6.096  96.9  3.7598   4 307    21.0
## 36   0.06417   0.0  5.96     0 0.4990 5.933  68.2  3.3603   5 279    19.2
## 37   0.09744   0.0  5.96     0 0.4990 5.841  61.4  3.3779   5 279    19.2
## 38   0.08014   0.0  5.96     0 0.4990 5.850  41.5  3.9342   5 279    19.2
## 39   0.17505   0.0  5.96     0 0.4990 5.966  30.2  3.8473   5 279    19.2
## 40   0.02763  75.0  2.95     0 0.4280 6.595  21.8  5.4011   3 252    18.3
## 41   0.03359  75.0  2.95     0 0.4280 7.024  15.8  5.4011   3 252    18.3
## 42   0.12744   0.0  6.91     0 0.4480 6.770   2.9  5.7209   3 233    17.9
## 43   0.14150   0.0  6.91     0 0.4480 6.169   6.6  5.7209   3 233    17.9
## 44   0.15936   0.0  6.91     0 0.4480 6.211   6.5  5.7209   3 233    17.9
## 45   0.12269   0.0  6.91     0 0.4480 6.069  40.0  5.7209   3 233    17.9
## 46   0.17142   0.0  6.91     0 0.4480 5.682  33.8  5.1004   3 233    17.9
## 47   0.18836   0.0  6.91     0 0.4480 5.786  33.3  5.1004   3 233    17.9
## 48   0.22927   0.0  6.91     0 0.4480 6.030  85.5  5.6894   3 233    17.9
## 49   0.25387   0.0  6.91     0 0.4480 5.399  95.3  5.8700   3 233    17.9
## 50   0.21977   0.0  6.91     0 0.4480 5.602  62.0  6.0877   3 233    17.9
## 51   0.08873  21.0  5.64     0 0.4390 5.963  45.7  6.8147   4 243    16.8
## 52   0.04337  21.0  5.64     0 0.4390 6.115  63.0  6.8147   4 243    16.8
## 53   0.05360  21.0  5.64     0 0.4390 6.511  21.1  6.8147   4 243    16.8
## 54   0.04981  21.0  5.64     0 0.4390 5.998  21.4  6.8147   4 243    16.8
## 55   0.01360  75.0  4.00     0 0.4100 5.888  47.6  7.3197   3 469    21.1
## 56   0.01311  90.0  1.22     0 0.4030 7.249  21.9  8.6966   5 226    17.9
## 57   0.02055  85.0  0.74     0 0.4100 6.383  35.7  9.1876   2 313    17.3
## 58   0.01432 100.0  1.32     0 0.4110 6.816  40.5  8.3248   5 256    15.1
## 59   0.15445  25.0  5.13     0 0.4530 6.145  29.2  7.8148   8 284    19.7
## 60   0.10328  25.0  5.13     0 0.4530 5.927  47.2  6.9320   8 284    19.7
## 61   0.14932  25.0  5.13     0 0.4530 5.741  66.2  7.2254   8 284    19.7
## 62   0.17171  25.0  5.13     0 0.4530 5.966  93.4  6.8185   8 284    19.7
## 63   0.11027  25.0  5.13     0 0.4530 6.456  67.8  7.2255   8 284    19.7
## 64   0.12650  25.0  5.13     0 0.4530 6.762  43.4  7.9809   8 284    19.7
## 65   0.01951  17.5  1.38     0 0.4161 7.104  59.5  9.2229   3 216    18.6
## 66   0.03584  80.0  3.37     0 0.3980 6.290  17.8  6.6115   4 337    16.1
## 67   0.04379  80.0  3.37     0 0.3980 5.787  31.1  6.6115   4 337    16.1
## 68   0.05789  12.5  6.07     0 0.4090 5.878  21.4  6.4980   4 345    18.9
## 69   0.13554  12.5  6.07     0 0.4090 5.594  36.8  6.4980   4 345    18.9
## 70   0.12816  12.5  6.07     0 0.4090 5.885  33.0  6.4980   4 345    18.9
## 71   0.08826   0.0 10.81     0 0.4130 6.417   6.6  5.2873   4 305    19.2
## 72   0.15876   0.0 10.81     0 0.4130 5.961  17.5  5.2873   4 305    19.2
## 73   0.09164   0.0 10.81     0 0.4130 6.065   7.8  5.2873   4 305    19.2
## 74   0.19539   0.0 10.81     0 0.4130 6.245   6.2  5.2873   4 305    19.2
```

```
## 75   0.07896  0.0 12.83    0 0.4370 6.273   6.0  4.2515   5 398     18.7
## 76   0.09512  0.0 12.83    0 0.4370 6.286  45.0  4.5026   5 398     18.7
## 77   0.10153  0.0 12.83    0 0.4370 6.279  74.5  4.0522   5 398     18.7
## 78   0.08707  0.0 12.83    0 0.4370 6.140  45.8  4.0905   5 398     18.7
## 79   0.05646  0.0 12.83    0 0.4370 6.232  53.7  5.0141   5 398     18.7
## 80   0.08387  0.0 12.83    0 0.4370 5.874  36.6  4.5026   5 398     18.7
## 81   0.04113 25.0  4.86    0 0.4260 6.727  33.5  5.4007   4 281     19.0
## 82   0.04462 25.0  4.86    0 0.4260 6.619  70.4  5.4007   4 281     19.0
## 83   0.03659 25.0  4.86    0 0.4260 6.302  32.2  5.4007   4 281     19.0
## 84   0.03551 25.0  4.86    0 0.4260 6.167  46.7  5.4007   4 281     19.0
## 85   0.05059  0.0  4.49    0 0.4490 6.389  48.0  4.7794   3 247     18.5
## 86   0.05735  0.0  4.49    0 0.4490 6.630  56.1  4.4377   3 247     18.5
## 87   0.05188  0.0  4.49    0 0.4490 6.015  45.1  4.4272   3 247     18.5
## 88   0.07151  0.0  4.49    0 0.4490 6.121  56.8  3.7476   3 247     18.5
## 89   0.05660  0.0  3.41    0 0.4890 7.007  86.3  3.4217   2 270     17.8
## 90   0.05302  0.0  3.41    0 0.4890 7.079  63.1  3.4145   2 270     17.8
## 91   0.04684  0.0  3.41    0 0.4890 6.417  66.1  3.0923   2 270     17.8
## 92   0.03932  0.0  3.41    0 0.4890 6.405  73.9  3.0921   2 270     17.8
## 93   0.04203 28.0 15.04    0 0.4640 6.442  53.6  3.6659   4 270     18.2
## 94   0.02875 28.0 15.04    0 0.4640 6.211  28.9  3.6659   4 270     18.2
## 95   0.04294 28.0 15.04    0 0.4640 6.249  77.3  3.6150   4 270     18.2
## 96   0.12204  0.0  2.89    0 0.4450 6.625  57.8  3.4952   2 276     18.0
## 97   0.11504  0.0  2.89    0 0.4450 6.163  69.6  3.4952   2 276     18.0
## 98   0.12083  0.0  2.89    0 0.4450 8.069  76.0  3.4952   2 276     18.0
## 99   0.08187  0.0  2.89    0 0.4450 7.820  36.9  3.4952   2 276     18.0
## 100  0.06860  0.0  2.89    0 0.4450 7.416  62.5  3.4952   2 276     18.0
## 101  0.14866  0.0  8.56    0 0.5200 6.727  79.9  2.7778   5 384     20.9
## 102  0.11432  0.0  8.56    0 0.5200 6.781  71.3  2.8561   5 384     20.9
## 103  0.22876  0.0  8.56    0 0.5200 6.405  85.4  2.7147   5 384     20.9
## 104  0.21161  0.0  8.56    0 0.5200 6.137  87.4  2.7147   5 384     20.9
## 105  0.13960  0.0  8.56    0 0.5200 6.167  90.0  2.4210   5 384     20.9
## 106  0.13262  0.0  8.56    0 0.5200 5.851  96.7  2.1069   5 384     20.9
## 107  0.17120  0.0  8.56    0 0.5200 5.836  91.9  2.2110   5 384     20.9
## 108  0.13117  0.0  8.56    0 0.5200 6.127  85.2  2.1224   5 384     20.9
## 109  0.12802  0.0  8.56    0 0.5200 6.474  97.1  2.4329   5 384     20.9
## 110  0.26363  0.0  8.56    0 0.5200 6.229  91.2  2.5451   5 384     20.9
## 111  0.10793  0.0  8.56    0 0.5200 6.195  54.4  2.7778   5 384     20.9
## 112  0.10084  0.0 10.01    0 0.5470 6.715  81.6  2.6775   6 432     17.8
## 113  0.12329  0.0 10.01    0 0.5470 5.913  92.9  2.3534   6 432     17.8
## 114  0.22212  0.0 10.01    0 0.5470 6.092  95.4  2.5480   6 432     17.8
## 115  0.14231  0.0 10.01    0 0.5470 6.254  84.2  2.2565   6 432     17.8
## 116  0.17134  0.0 10.01    0 0.5470 5.928  88.2  2.4631   6 432     17.8
## 117  0.13158  0.0 10.01    0 0.5470 6.176  72.5  2.7301   6 432     17.8
## 118  0.15098  0.0 10.01    0 0.5470 6.021  82.6  2.7474   6 432     17.8
## 119  0.13058  0.0 10.01    0 0.5470 5.872  73.1  2.4775   6 432     17.8
## 120  0.14476  0.0 10.01    0 0.5470 5.731  65.2  2.7592   6 432     17.8
## 121  0.06899  0.0 25.65    0 0.5810 5.870  69.7  2.2577   2 188     19.1
## 122  0.07165  0.0 25.65    0 0.5810 6.004  84.1  2.1974   2 188     19.1
## 123  0.09299  0.0 25.65    0 0.5810 5.961  92.9  2.0869   2 188     19.1
## 124  0.15038  0.0 25.65    0 0.5810 5.856  97.0  1.9444   2 188     19.1
## 125  0.09849  0.0 25.65    0 0.5810 5.879  95.8  2.0063   2 188     19.1
## 126  0.16902  0.0 25.65    0 0.5810 5.986  88.4  1.9929   2 188     19.1
## 127  0.38735  0.0 25.65    0 0.5810 5.613  95.6  1.7572   2 188     19.1
## 128  0.25915  0.0 21.89    0 0.6240 5.693  96.0  1.7883   4 437     21.2
```

```
## 129  0.32543   0.0 21.89    0 0.6240 6.431  98.8  1.8125   4 437    21.2
## 130  0.88125   0.0 21.89    0 0.6240 5.637  94.7  1.9799   4 437    21.2
## 131  0.34006   0.0 21.89    0 0.6240 6.458  98.9  2.1185   4 437    21.2
## 132  1.19294   0.0 21.89    0 0.6240 6.326  97.7  2.2710   4 437    21.2
## 133  0.59005   0.0 21.89    0 0.6240 6.372  97.9  2.3274   4 437    21.2
## 134  0.32982   0.0 21.89    0 0.6240 5.822  95.4  2.4699   4 437    21.2
## 135  0.97617   0.0 21.89    0 0.6240 5.757  98.4  2.3460   4 437    21.2
## 136  0.55778   0.0 21.89    0 0.6240 6.335  98.2  2.1107   4 437    21.2
## 137  0.32264   0.0 21.89    0 0.6240 5.942  93.5  1.9669   4 437    21.2
## 138  0.35233   0.0 21.89    0 0.6240 6.454  98.4  1.8498   4 437    21.2
## 139  0.24980   0.0 21.89    0 0.6240 5.857  98.2  1.6686   4 437    21.2
## 140  0.54452   0.0 21.89    0 0.6240 6.151  97.9  1.6687   4 437    21.2
## 141  0.29090   0.0 21.89    0 0.6240 6.174  93.6  1.6119   4 437    21.2
## 142  1.62864   0.0 21.89    0 0.6240 5.019 100.0  1.4394   4 437    21.2
## 143  3.32105   0.0 19.58    1 0.8710 5.403 100.0  1.3216   5 403    14.7
## 144  4.09740   0.0 19.58    0 0.8710 5.468 100.0  1.4118   5 403    14.7
## 145  2.77974   0.0 19.58    0 0.8710 4.903  97.8  1.3459   5 403    14.7
## 146  2.37934   0.0 19.58    0 0.8710 6.130 100.0  1.4191   5 403    14.7
## 147  2.15505   0.0 19.58    0 0.8710 5.628 100.0  1.5166   5 403    14.7
## 148  2.36862   0.0 19.58    0 0.8710 4.926  95.7  1.4608   5 403    14.7
## 149  2.33099   0.0 19.58    0 0.8710 5.186  93.8  1.5296   5 403    14.7
## 150  2.73397   0.0 19.58    0 0.8710 5.597  94.9  1.5257   5 403    14.7
## 151  1.65660   0.0 19.58    0 0.8710 6.122  97.3  1.6180   5 403    14.7
## 152  1.49632   0.0 19.58    0 0.8710 5.404 100.0  1.5916   5 403    14.7
## 153  1.12658   0.0 19.58    1 0.8710 5.012  88.0  1.6102   5 403    14.7
## 154  2.14918   0.0 19.58    0 0.8710 5.709  98.5  1.6232   5 403    14.7
## 155  1.41385   0.0 19.58    1 0.8710 6.129  96.0  1.7494   5 403    14.7
## 156  3.53501   0.0 19.58    1 0.8710 6.152  82.6  1.7455   5 403    14.7
## 157  2.44668   0.0 19.58    0 0.8710 5.272  94.0  1.7364   5 403    14.7
## 158  1.22358   0.0 19.58    0 0.6050 6.943  97.4  1.8773   5 403    14.7
## 159  1.34284   0.0 19.58    0 0.6050 6.066 100.0  1.7573   5 403    14.7
## 160  1.42502   0.0 19.58    0 0.8710 6.510 100.0  1.7659   5 403    14.7
## 161  1.27346   0.0 19.58    1 0.6050 6.250  92.6  1.7984   5 403    14.7
## 162  1.46336   0.0 19.58    0 0.6050 7.489  90.8  1.9709   5 403    14.7
## 163  1.83377   0.0 19.58    1 0.6050 7.802  98.2  2.0407   5 403    14.7
## 164  1.51902   0.0 19.58    1 0.6050 8.375  93.9  2.1620   5 403    14.7
## 165  2.24236   0.0 19.58    0 0.6050 5.854  91.8  2.4220   5 403    14.7
## 166  2.92400   0.0 19.58    0 0.6050 6.101  93.0  2.2834   5 403    14.7
## 167  2.01019   0.0 19.58    0 0.6050 7.929  96.2  2.0459   5 403    14.7
## 168  1.80028   0.0 19.58    0 0.6050 5.877  79.2  2.4259   5 403    14.7
## 169  2.30040   0.0 19.58    0 0.6050 6.319  96.1  2.1000   5 403    14.7
## 170  2.44953   0.0 19.58    0 0.6050 6.402  95.2  2.2625   5 403    14.7
## 171  1.20742   0.0 19.58    0 0.6050 5.875  94.6  2.4259   5 403    14.7
## 172  2.31390   0.0 19.58    0 0.6050 5.880  97.3  2.3887   5 403    14.7
## 173  0.13914   0.0  4.05    0 0.5100 5.572  88.5  2.5961   5 296    16.6
## 174  0.09178   0.0  4.05    0 0.5100 6.416  84.1  2.6463   5 296    16.6
## 175  0.08447   0.0  4.05    0 0.5100 5.859  68.7  2.7019   5 296    16.6
## 176  0.06664   0.0  4.05    0 0.5100 6.546  33.1  3.1323   5 296    16.6
## 177  0.07022   0.0  4.05    0 0.5100 6.020  47.2  3.5549   5 296    16.6
## 178  0.05425   0.0  4.05    0 0.5100 6.315  73.4  3.3175   5 296    16.6
## 179  0.06642   0.0  4.05    0 0.5100 6.860  74.4  2.9153   5 296    16.6
## 180  0.05780   0.0  2.46    0 0.4880 6.980  58.4  2.8290   3 193    17.8
## 181  0.06588   0.0  2.46    0 0.4880 7.765  83.3  2.7410   3 193    17.8
## 182  0.06888   0.0  2.46    0 0.4880 6.144  62.2  2.5979   3 193    17.8
```

```
## 183  0.09103  0.0  2.46    0 0.4880 7.155  92.2  2.7006   3 193    17.8
## 184  0.10008  0.0  2.46    0 0.4880 6.563  95.6  2.8470   3 193    17.8
## 185  0.08308  0.0  2.46    0 0.4880 5.604  89.8  2.9879   3 193    17.8
## 186  0.06047  0.0  2.46    0 0.4880 6.153  68.8  3.2797   3 193    17.8
## 187  0.05602  0.0  2.46    0 0.4880 7.831  53.6  3.1992   3 193    17.8
## 188  0.07875 45.0  3.44    0 0.4370 6.782  41.1  3.7886   5 398    15.2
## 189  0.12579 45.0  3.44    0 0.4370 6.556  29.1  4.5667   5 398    15.2
## 190  0.08370 45.0  3.44    0 0.4370 7.185  38.9  4.5667   5 398    15.2
## 191  0.09068 45.0  3.44    0 0.4370 6.951  21.5  6.4798   5 398    15.2
## 192  0.06911 45.0  3.44    0 0.4370 6.739  30.8  6.4798   5 398    15.2
## 193  0.08664 45.0  3.44    0 0.4370 7.178  26.3  6.4798   5 398    15.2
## 194  0.02187 60.0  2.93    0 0.4010 6.800   9.9  6.2196   1 265    15.6
## 195  0.01439 60.0  2.93    0 0.4010 6.604  18.8  6.2196   1 265    15.6
## 196  0.01381 80.0  0.46    0 0.4220 7.875  32.0  5.6484   4 255    14.4
## 197  0.04011 80.0  1.52    0 0.4040 7.287  34.1  7.3090   2 329    12.6
## 198  0.04666 80.0  1.52    0 0.4040 7.107  36.6  7.3090   2 329    12.6
## 199  0.03768 80.0  1.52    0 0.4040 7.274  38.3  7.3090   2 329    12.6
## 200  0.03150 95.0  1.47    0 0.4030 6.975  15.3  7.6534   3 402    17.0
## 201  0.01778 95.0  1.47    0 0.4030 7.135  13.9  7.6534   3 402    17.0
## 202  0.03445 82.5  2.03    0 0.4150 6.162  38.4  6.2700   2 348    14.7
## 203  0.02177 82.5  2.03    0 0.4150 7.610  15.7  6.2700   2 348    14.7
## 204  0.03510 95.0  2.68    0 0.4161 7.853  33.2  5.1180   4 224    14.7
## 205  0.02009 95.0  2.68    0 0.4161 8.034  31.9  5.1180   4 224    14.7
## 206  0.13642  0.0 10.59    0 0.4890 5.891  22.3  3.9454   4 277    18.6
## 207  0.22969  0.0 10.59    0 0.4890 6.326  52.5  4.3549   4 277    18.6
## 208  0.25199  0.0 10.59    0 0.4890 5.783  72.7  4.3549   4 277    18.6
## 209  0.13587  0.0 10.59    1 0.4890 6.064  59.1  4.2392   4 277    18.6
## 210  0.43571  0.0 10.59    1 0.4890 5.344 100.0  3.8750   4 277    18.6
## 211  0.17446  0.0 10.59    1 0.4890 5.960  92.1  3.8771   4 277    18.6
## 212  0.37578  0.0 10.59    1 0.4890 5.404  88.6  3.6650   4 277    18.6
## 213  0.21719  0.0 10.59    1 0.4890 5.807  53.8  3.6526   4 277    18.6
## 214  0.14052  0.0 10.59    0 0.4890 6.375  32.3  3.9454   4 277    18.6
## 215  0.28955  0.0 10.59    0 0.4890 5.412   9.8  3.5875   4 277    18.6
## 216  0.19802  0.0 10.59    0 0.4890 6.182  42.4  3.9454   4 277    18.6
## 217  0.04560  0.0 13.89    1 0.5500 5.888  56.0  3.1121   5 276    16.4
## 218  0.07013  0.0 13.89    0 0.5500 6.642  85.1  3.4211   5 276    16.4
## 219  0.11069  0.0 13.89    1 0.5500 5.951  93.8  2.8893   5 276    16.4
## 220  0.11425  0.0 13.89    1 0.5500 6.373  92.4  3.3633   5 276    16.4
## 221  0.35809  0.0  6.20    1 0.5070 6.951  88.5  2.8617   8 307    17.4
## 222  0.40771  0.0  6.20    1 0.5070 6.164  91.3  3.0480   8 307    17.4
## 223  0.62356  0.0  6.20    1 0.5070 6.879  77.7  3.2721   8 307    17.4
## 224  0.61470  0.0  6.20    0 0.5070 6.618  80.8  3.2721   8 307    17.4
## 225  0.31533  0.0  6.20    0 0.5040 8.266  78.3  2.8944   8 307    17.4
## 226  0.52693  0.0  6.20    0 0.5040 8.725  83.0  2.8944   8 307    17.4
## 227  0.38214  0.0  6.20    0 0.5040 8.040  86.5  3.2157   8 307    17.4
## 228  0.41238  0.0  6.20    0 0.5040 7.163  79.9  3.2157   8 307    17.4
## 229  0.29819  0.0  6.20    0 0.5040 7.686  17.0  3.3751   8 307    17.4
## 230  0.44178  0.0  6.20    0 0.5040 6.552  21.4  3.3751   8 307    17.4
## 231  0.53700  0.0  6.20    0 0.5040 5.981  68.1  3.6715   8 307    17.4
## 232  0.46296  0.0  6.20    0 0.5040 7.412  76.9  3.6715   8 307    17.4
## 233  0.57529  0.0  6.20    0 0.5070 8.337  73.3  3.8384   8 307    17.4
## 234  0.33147  0.0  6.20    0 0.5070 8.247  70.4  3.6519   8 307    17.4
## 235  0.44791  0.0  6.20    1 0.5070 6.726  66.5  3.6519   8 307    17.4
## 236  0.33045  0.0  6.20    0 0.5070 6.086  61.5  3.6519   8 307    17.4
```

```
## 237 0.52058  0.0  6.20    1 0.5070 6.631  76.5 4.1480   8 307    17.4
## 238 0.51183  0.0  6.20    0 0.5070 7.358  71.6 4.1480   8 307    17.4
## 239 0.08244 30.0  4.93    0 0.4280 6.481  18.5 6.1899   6 300    16.6
## 240 0.09252 30.0  4.93    0 0.4280 6.606  42.2 6.1899   6 300    16.6
## 241 0.11329 30.0  4.93    0 0.4280 6.897  54.3 6.3361   6 300    16.6
## 242 0.10612 30.0  4.93    0 0.4280 6.095  65.1 6.3361   6 300    16.6
## 243 0.10290 30.0  4.93    0 0.4280 6.358  52.9 7.0355   6 300    16.6
## 244 0.12757 30.0  4.93    0 0.4280 6.393   7.8 7.0355   6 300    16.6
## 245 0.20608 22.0  5.86    0 0.4310 5.593  76.5 7.9549   7 330    19.1
## 246 0.19133 22.0  5.86    0 0.4310 5.605  70.2 7.9549   7 330    19.1
## 247 0.33983 22.0  5.86    0 0.4310 6.108  34.9 8.0555   7 330    19.1
## 248 0.19657 22.0  5.86    0 0.4310 6.226  79.2 8.0555   7 330    19.1
## 249 0.16439 22.0  5.86    0 0.4310 6.433  49.1 7.8265   7 330    19.1
## 250 0.19073 22.0  5.86    0 0.4310 6.718  17.5 7.8265   7 330    19.1
## 251 0.14030 22.0  5.86    0 0.4310 6.487  13.0 7.3967   7 330    19.1
## 252 0.21409 22.0  5.86    0 0.4310 6.438   8.9 7.3967   7 330    19.1
## 253 0.08221 22.0  5.86    0 0.4310 6.957   6.8 8.9067   7 330    19.1
## 254 0.36894 22.0  5.86    0 0.4310 8.259   8.4 8.9067   7 330    19.1
## 255 0.04819 80.0  3.64    0 0.3920 6.108  32.0 9.2203   1 315    16.4
## 256 0.03548 80.0  3.64    0 0.3920 5.876  19.1 9.2203   1 315    16.4
## 257 0.01538 90.0  3.75    0 0.3940 7.454  34.2 6.3361   3 244    15.9
## 258 0.61154 20.0  3.97    0 0.6470 8.704  86.9 1.8010   5 264    13.0
## 259 0.66351 20.0  3.97    0 0.6470 7.333 100.0 1.8946   5 264    13.0
## 260 0.65665 20.0  3.97    0 0.6470 6.842 100.0 2.0107   5 264    13.0
## 261 0.54011 20.0  3.97    0 0.6470 7.203  81.8 2.1121   5 264    13.0
## 262 0.53412 20.0  3.97    0 0.6470 7.520  89.4 2.1398   5 264    13.0
## 263 0.52014 20.0  3.97    0 0.6470 8.398  91.5 2.2885   5 264    13.0
## 264 0.82526 20.0  3.97    0 0.6470 7.327  94.5 2.0788   5 264    13.0
## 265 0.55007 20.0  3.97    0 0.6470 7.206  91.6 1.9301   5 264    13.0
## 266 0.76162 20.0  3.97    0 0.6470 5.560  62.8 1.9865   5 264    13.0
## 267 0.78570 20.0  3.97    0 0.6470 7.014  84.6 2.1329   5 264    13.0
## 268 0.57834 20.0  3.97    0 0.5750 8.297  67.0 2.4216   5 264    13.0
## 269 0.54050 20.0  3.97    0 0.5750 7.470  52.6 2.8720   5 264    13.0
## 270 0.09065 20.0  6.96    1 0.4640 5.920  61.5 3.9175   3 223    18.6
## 271 0.29916 20.0  6.96    0 0.4640 5.856  42.1 4.4290   3 223    18.6
## 272 0.16211 20.0  6.96    0 0.4640 6.240  16.3 4.4290   3 223    18.6
## 273 0.11460 20.0  6.96    0 0.4640 6.538  58.7 3.9175   3 223    18.6
## 274 0.22188 20.0  6.96    1 0.4640 7.691  51.8 4.3665   3 223    18.6
## 275 0.05644 40.0  6.41    1 0.4470 6.758  32.9 4.0776   4 254    17.6
## 276 0.09604 40.0  6.41    0 0.4470 6.854  42.8 4.2673   4 254    17.6
## 277 0.10469 40.0  6.41    1 0.4470 7.267  49.0 4.7872   4 254    17.6
## 278 0.06127 40.0  6.41    1 0.4470 6.826  27.6 4.8628   4 254    17.6
## 279 0.07978 40.0  6.41    0 0.4470 6.482  32.1 4.1403   4 254    17.6
## 280 0.21038 20.0  3.33    0 0.4429 6.812  32.2 4.1007   5 216    14.9
## 281 0.03578 20.0  3.33    0 0.4429 7.820  64.5 4.6947   5 216    14.9
## 282 0.03705 20.0  3.33    0 0.4429 6.968  37.2 5.2447   5 216    14.9
## 283 0.06129 20.0  3.33    1 0.4429 7.645  49.7 5.2119   5 216    14.9
## 284 0.01501 90.0  1.21    1 0.4010 7.923  24.8 5.8850   1 198    13.6
## 285 0.00906 90.0  2.97    0 0.4000 7.088  20.8 7.3073   1 285    15.3
## 286 0.01096 55.0  2.25    0 0.3890 6.453  31.9 7.3073   1 300    15.3
## 287 0.01965 80.0  1.76    0 0.3850 6.230  31.5 9.0892   1 241    18.2
## 288 0.03871 52.5  5.32    0 0.4050 6.209  31.3 7.3172   6 293    16.6
## 289 0.04590 52.5  5.32    0 0.4050 6.315  45.6 7.3172   6 293    16.6
## 290 0.04297 52.5  5.32    0 0.4050 6.565  22.9 7.3172   6 293    16.6
```

```
## 291  0.03502  80.0  4.95    0 0.4110 6.861  27.9  5.1167   4 245    19.2
## 292  0.07886  80.0  4.95    0 0.4110 7.148  27.7  5.1167   4 245    19.2
## 293  0.03615  80.0  4.95    0 0.4110 6.630  23.4  5.1167   4 245    19.2
## 294  0.08265   0.0 13.92    0 0.4370 6.127  18.4  5.5027   4 289    16.0
## 295  0.08199   0.0 13.92    0 0.4370 6.009  42.3  5.5027   4 289    16.0
## 296  0.12932   0.0 13.92    0 0.4370 6.678  31.1  5.9604   4 289    16.0
## 297  0.05372   0.0 13.92    0 0.4370 6.549  51.0  5.9604   4 289    16.0
## 298  0.14103   0.0 13.92    0 0.4370 5.790  58.0  6.3200   4 289    16.0
## 299  0.06466  70.0  2.24    0 0.4000 6.345  20.1  7.8278   5 358    14.8
## 300  0.05561  70.0  2.24    0 0.4000 7.041  10.0  7.8278   5 358    14.8
## 301  0.04417  70.0  2.24    0 0.4000 6.871  47.4  7.8278   5 358    14.8
## 302  0.03537  34.0  6.09    0 0.4330 6.590  40.4  5.4917   7 329    16.1
## 303  0.09266  34.0  6.09    0 0.4330 6.495  18.4  5.4917   7 329    16.1
## 304  0.10000  34.0  6.09    0 0.4330 6.982  17.7  5.4917   7 329    16.1
## 305  0.05515  33.0  2.18    0 0.4720 7.236  41.1  4.0220   7 222    18.4
## 306  0.05479  33.0  2.18    0 0.4720 6.616  58.1  3.3700   7 222    18.4
## 307  0.07503  33.0  2.18    0 0.4720 7.420  71.9  3.0992   7 222    18.4
## 308  0.04932  33.0  2.18    0 0.4720 6.849  70.3  3.1827   7 222    18.4
## 309  0.49298   0.0  9.90    0 0.5440 6.635  82.5  3.3175   4 304    18.4
## 310  0.34940   0.0  9.90    0 0.5440 5.972  76.7  3.1025   4 304    18.4
## 311  2.63548   0.0  9.90    0 0.5440 4.973  37.8  2.5194   4 304    18.4
## 312  0.79041   0.0  9.90    0 0.5440 6.122  52.8  2.6403   4 304    18.4
## 313  0.26169   0.0  9.90    0 0.5440 6.023  90.4  2.8340   4 304    18.4
## 314  0.26938   0.0  9.90    0 0.5440 6.266  82.8  3.2628   4 304    18.4
## 315  0.36920   0.0  9.90    0 0.5440 6.567  87.3  3.6023   4 304    18.4
## 316  0.25356   0.0  9.90    0 0.5440 5.705  77.7  3.9450   4 304    18.4
## 317  0.31827   0.0  9.90    0 0.5440 5.914  83.2  3.9986   4 304    18.4
## 318  0.24522   0.0  9.90    0 0.5440 5.782  71.7  4.0317   4 304    18.4
## 319  0.40202   0.0  9.90    0 0.5440 6.382  67.2  3.5325   4 304    18.4
## 320  0.47547   0.0  9.90    0 0.5440 6.113  58.8  4.0019   4 304    18.4
## 321  0.16760   0.0  7.38    0 0.4930 6.426  52.3  4.5404   5 287    19.6
## 322  0.18159   0.0  7.38    0 0.4930 6.376  54.3  4.5404   5 287    19.6
## 323  0.35114   0.0  7.38    0 0.4930 6.041  49.9  4.7211   5 287    19.6
## 324  0.28392   0.0  7.38    0 0.4930 5.708  74.3  4.7211   5 287    19.6
## 325  0.34109   0.0  7.38    0 0.4930 6.415  40.1  4.7211   5 287    19.6
## 326  0.19186   0.0  7.38    0 0.4930 6.431  14.7  5.4159   5 287    19.6
## 327  0.30347   0.0  7.38    0 0.4930 6.312  28.9  5.4159   5 287    19.6
## 328  0.24103   0.0  7.38    0 0.4930 6.083  43.7  5.4159   5 287    19.6
## 329  0.06617   0.0  3.24    0 0.4600 5.868  25.8  5.2146   4 430    16.9
## 330  0.06724   0.0  3.24    0 0.4600 6.333  17.2  5.2146   4 430    16.9
## 331  0.04544   0.0  3.24    0 0.4600 6.144  32.2  5.8736   4 430    16.9
## 332  0.05023  35.0  6.06    0 0.4379 5.706  28.4  6.6407   1 304    16.9
## 333  0.03466  35.0  6.06    0 0.4379 6.031  23.3  6.6407   1 304    16.9
## 334  0.05083   0.0  5.19    0 0.5150 6.316  38.1  6.4584   5 224    20.2
## 335  0.03738   0.0  5.19    0 0.5150 6.310  38.5  6.4584   5 224    20.2
## 336  0.03961   0.0  5.19    0 0.5150 6.037  34.5  5.9853   5 224    20.2
## 337  0.03427   0.0  5.19    0 0.5150 5.869  46.3  5.2311   5 224    20.2
## 338  0.03041   0.0  5.19    0 0.5150 5.895  59.6  5.6150   5 224    20.2
## 339  0.03306   0.0  5.19    0 0.5150 6.059  37.3  4.8122   5 224    20.2
## 340  0.05497   0.0  5.19    0 0.5150 5.985  45.4  4.8122   5 224    20.2
## 341  0.06151   0.0  5.19    0 0.5150 5.968  58.5  4.8122   5 224    20.2
## 342  0.01301  35.0  1.52    0 0.4420 7.241  49.3  7.0379   1 284    15.5
## 343  0.02498   0.0  1.89    0 0.5180 6.540  59.7  6.2669   1 422    15.9
## 344  0.02543  55.0  3.78    0 0.4840 6.696  56.4  5.7321   5 370    17.6
```

```
## 345  0.03049  55.0  3.78    0 0.4840 6.874  28.1  6.4654   5 370     17.6
## 346  0.03113   0.0  4.39    0 0.4420 6.014  48.5  8.0136   3 352     18.8
## 347  0.06162   0.0  4.39    0 0.4420 5.898  52.3  8.0136   3 352     18.8
## 348  0.01870  85.0  4.15    0 0.4290 6.516  27.7  8.5353   4 351     17.9
## 349  0.01501  80.0  2.01    0 0.4350 6.635  29.7  8.3440   4 280     17.0
## 350  0.02899  40.0  1.25    0 0.4290 6.939  34.5  8.7921   1 335     19.7
## 351  0.06211  40.0  1.25    0 0.4290 6.490  44.4  8.7921   1 335     19.7
## 352  0.07950  60.0  1.69    0 0.4110 6.579  35.9 10.7103   4 411     18.3
## 353  0.07244  60.0  1.69    0 0.4110 5.884  18.5 10.7103   4 411     18.3
## 354  0.01709  90.0  2.02    0 0.4100 6.728  36.1 12.1265   5 187     17.0
## 355  0.04301  80.0  1.91    0 0.4130 5.663  21.9 10.5857   4 334     22.0
## 356  0.10659  80.0  1.91    0 0.4130 5.936  19.5 10.5857   4 334     22.0
## 357  8.98296   0.0 18.10    1 0.7700 6.212  97.4  2.1222  24 666     20.2
## 358  3.84970   0.0 18.10    1 0.7700 6.395  91.0  2.5052  24 666     20.2
## 359  5.20177   0.0 18.10    1 0.7700 6.127  83.4  2.7227  24 666     20.2
## 360  4.26131   0.0 18.10    0 0.7700 6.112  81.3  2.5091  24 666     20.2
## 361  4.54192   0.0 18.10    0 0.7700 6.398  88.0  2.5182  24 666     20.2
## 362  3.83684   0.0 18.10    0 0.7700 6.251  91.1  2.2955  24 666     20.2
## 363  3.67822   0.0 18.10    0 0.7700 5.362  96.2  2.1036  24 666     20.2
## 364  4.22239   0.0 18.10    1 0.7700 5.803  89.0  1.9047  24 666     20.2
## 365  3.47428   0.0 18.10    1 0.7180 8.780  82.9  1.9047  24 666     20.2
## 366  4.55587   0.0 18.10    0 0.7180 3.561  87.9  1.6132  24 666     20.2
## 367  3.69695   0.0 18.10    0 0.7180 4.963  91.4  1.7523  24 666     20.2
## 368 13.52220   0.0 18.10    0 0.6310 3.863 100.0  1.5106  24 666     20.2
## 369  4.89822   0.0 18.10    0 0.6310 4.970 100.0  1.3325  24 666     20.2
## 370  5.66998   0.0 18.10    1 0.6310 6.683  96.8  1.3567  24 666     20.2
## 371  6.53876   0.0 18.10    1 0.6310 7.016  97.5  1.2024  24 666     20.2
## 372  9.23230   0.0 18.10    0 0.6310 6.216 100.0  1.1691  24 666     20.2
## 373  8.26725   0.0 18.10    1 0.6680 5.875  89.6  1.1296  24 666     20.2
## 374 11.10810   0.0 18.10    0 0.6680 4.906 100.0  1.1742  24 666     20.2
## 375 18.49820   0.0 18.10    0 0.6680 4.138 100.0  1.1370  24 666     20.2
## 376 19.60910   0.0 18.10    0 0.6710 7.313  97.9  1.3163  24 666     20.2
## 377 15.28800   0.0 18.10    0 0.6710 6.649  93.3  1.3449  24 666     20.2
## 378  9.82349   0.0 18.10    0 0.6710 6.794  98.8  1.3580  24 666     20.2
## 379 23.64820   0.0 18.10    0 0.6710 6.380  96.2  1.3861  24 666     20.2
## 380 17.86670   0.0 18.10    0 0.6710 6.223 100.0  1.3861  24 666     20.2
## 381 88.97620   0.0 18.10    0 0.6710 6.968  91.9  1.4165  24 666     20.2
## 382 15.87440   0.0 18.10    0 0.6710 6.545  99.1  1.5192  24 666     20.2
## 383  9.18702   0.0 18.10    0 0.7000 5.536 100.0  1.5804  24 666     20.2
## 384  7.99248   0.0 18.10    0 0.7000 5.520 100.0  1.5331  24 666     20.2
## 385 20.08490   0.0 18.10    0 0.7000 4.368  91.2  1.4395  24 666     20.2
## 386 16.81180   0.0 18.10    0 0.7000 5.277  98.1  1.4261  24 666     20.2
## 387 24.39380   0.0 18.10    0 0.7000 4.652 100.0  1.4672  24 666     20.2
## 388 22.59710   0.0 18.10    0 0.7000 5.000  89.5  1.5184  24 666     20.2
## 389 14.33370   0.0 18.10    0 0.7000 4.880 100.0  1.5895  24 666     20.2
## 390  8.15174   0.0 18.10    0 0.7000 5.390  98.9  1.7281  24 666     20.2
## 391  6.96215   0.0 18.10    0 0.7000 5.713  97.0  1.9265  24 666     20.2
## 392  5.29305   0.0 18.10    0 0.7000 6.051  82.5  2.1678  24 666     20.2
## 393 11.57790   0.0 18.10    0 0.7000 5.036  97.0  1.7700  24 666     20.2
## 394  8.64476   0.0 18.10    0 0.6930 6.193  92.6  1.7912  24 666     20.2
## 395 13.35980   0.0 18.10    0 0.6930 5.887  94.7  1.7821  24 666     20.2
## 396  8.71675   0.0 18.10    0 0.6930 6.471  98.8  1.7257  24 666     20.2
## 397  5.87205   0.0 18.10    0 0.6930 6.405  96.0  1.6768  24 666     20.2
## 398  7.67202   0.0 18.10    0 0.6930 5.747  98.9  1.6334  24 666     20.2
```

```
## 399 38.35180   0.0 18.10    0 0.6930 5.453 100.0  1.4896  24 666    20.2
## 400  9.91655   0.0 18.10    0 0.6930 5.852  77.8  1.5004  24 666    20.2
## 401 25.04610   0.0 18.10    0 0.6930 5.987 100.0  1.5888  24 666    20.2
## 402 14.23620   0.0 18.10    0 0.6930 6.343 100.0  1.5741  24 666    20.2
## 403  9.59571   0.0 18.10    0 0.6930 6.404 100.0  1.6390  24 666    20.2
## 404 24.80170   0.0 18.10    0 0.6930 5.349  96.0  1.7028  24 666    20.2
## 405 41.52920   0.0 18.10    0 0.6930 5.531  85.4  1.6074  24 666    20.2
## 406 67.92080   0.0 18.10    0 0.6930 5.683 100.0  1.4254  24 666    20.2
## 407 20.71620   0.0 18.10    0 0.6590 4.138 100.0  1.1781  24 666    20.2
## 408 11.95110   0.0 18.10    0 0.6590 5.608 100.0  1.2852  24 666    20.2
## 409  7.40389   0.0 18.10    0 0.5970 5.617  97.9  1.4547  24 666    20.2
## 410 14.43830   0.0 18.10    0 0.5970 6.852 100.0  1.4655  24 666    20.2
## 411 51.13580   0.0 18.10    0 0.5970 5.757 100.0  1.4130  24 666    20.2
## 412 14.05070   0.0 18.10    0 0.5970 6.657 100.0  1.5275  24 666    20.2
## 413 18.81100   0.0 18.10    0 0.5970 4.628 100.0  1.5539  24 666    20.2
## 414 28.65580   0.0 18.10    0 0.5970 5.155 100.0  1.5894  24 666    20.2
## 415 45.74610   0.0 18.10    0 0.6930 4.519 100.0  1.6582  24 666    20.2
## 416 18.08460   0.0 18.10    0 0.6790 6.434 100.0  1.8347  24 666    20.2
## 417 10.83420   0.0 18.10    0 0.6790 6.782  90.8  1.8195  24 666    20.2
## 418 25.94060   0.0 18.10    0 0.6790 5.304  89.1  1.6475  24 666    20.2
## 419 73.53410   0.0 18.10    0 0.6790 5.957 100.0  1.8026  24 666    20.2
## 420 11.81230   0.0 18.10    0 0.7180 6.824  76.5  1.7940  24 666    20.2
## 421 11.08740   0.0 18.10    0 0.7180 6.411 100.0  1.8589  24 666    20.2
## 422  7.02259   0.0 18.10    0 0.7180 6.006  95.3  1.8746  24 666    20.2
## 423 12.04820   0.0 18.10    0 0.6140 5.648  87.6  1.9512  24 666    20.2
## 424  7.05042   0.0 18.10    0 0.6140 6.103  85.1  2.0218  24 666    20.2
## 425  8.79212   0.0 18.10    0 0.5840 5.565  70.6  2.0635  24 666    20.2
## 426 15.86030   0.0 18.10    0 0.6790 5.896  95.4  1.9096  24 666    20.2
## 427 12.24720   0.0 18.10    0 0.5840 5.837  59.7  1.9976  24 666    20.2
## 428 37.66190   0.0 18.10    0 0.6790 6.202  78.7  1.8629  24 666    20.2
## 429  7.36711   0.0 18.10    0 0.6790 6.193  78.1  1.9356  24 666    20.2
## 430  9.33889   0.0 18.10    0 0.6790 6.380  95.6  1.9682  24 666    20.2
## 431  8.49213   0.0 18.10    0 0.5840 6.348  86.1  2.0527  24 666    20.2
## 432 10.06230   0.0 18.10    0 0.5840 6.833  94.3  2.0882  24 666    20.2
## 433  6.44405   0.0 18.10    0 0.5840 6.425  74.8  2.2004  24 666    20.2
## 434  5.58107   0.0 18.10    0 0.7130 6.436  87.9  2.3158  24 666    20.2
## 435 13.91340   0.0 18.10    0 0.7130 6.208  95.0  2.2222  24 666    20.2
## 436 11.16040   0.0 18.10    0 0.7400 6.629  94.6  2.1247  24 666    20.2
## 437 14.42080   0.0 18.10    0 0.7400 6.461  93.3  2.0026  24 666    20.2
## 438 15.17720   0.0 18.10    0 0.7400 6.152 100.0  1.9142  24 666    20.2
## 439 13.67810   0.0 18.10    0 0.7400 5.935  87.9  1.8206  24 666    20.2
## 440  9.39063   0.0 18.10    0 0.7400 5.627  93.9  1.8172  24 666    20.2
## 441 22.05110   0.0 18.10    0 0.7400 5.818  92.4  1.8662  24 666    20.2
## 442  9.72418   0.0 18.10    0 0.7400 6.406  97.2  2.0651  24 666    20.2
## 443  5.66637   0.0 18.10    0 0.7400 6.219 100.0  2.0048  24 666    20.2
## 444  9.96654   0.0 18.10    0 0.7400 6.485 100.0  1.9784  24 666    20.2
## 445 12.80230   0.0 18.10    0 0.7400 5.854  96.6  1.8956  24 666    20.2
## 446 10.67180   0.0 18.10    0 0.7400 6.459  94.8  1.9879  24 666    20.2
## 447  6.28807   0.0 18.10    0 0.7400 6.341  96.4  2.0720  24 666    20.2
## 448  9.92485   0.0 18.10    0 0.7400 6.251  96.6  2.1980  24 666    20.2
## 449  9.32909   0.0 18.10    0 0.7130 6.185  98.7  2.2616  24 666    20.2
## 450  7.52601   0.0 18.10    0 0.7130 6.417  98.3  2.1850  24 666    20.2
## 451  6.71772   0.0 18.10    0 0.7130 6.749  92.6  2.3236  24 666    20.2
## 452  5.44114   0.0 18.10    0 0.7130 6.655  98.2  2.3552  24 666    20.2
```

```
## 453  5.09017   0.0 18.10    0 0.7130 6.297  91.8  2.3682  24 666    20.2
## 454  8.24809   0.0 18.10    0 0.7130 7.393  99.3  2.4527  24 666    20.2
## 455  9.51363   0.0 18.10    0 0.7130 6.728  94.1  2.4961  24 666    20.2
## 456  4.75237   0.0 18.10    0 0.7130 6.525  86.5  2.4358  24 666    20.2
## 457  4.66883   0.0 18.10    0 0.7130 5.976  87.9  2.5806  24 666    20.2
## 458  8.20058   0.0 18.10    0 0.7130 5.936  80.3  2.7792  24 666    20.2
## 459  7.75223   0.0 18.10    0 0.7130 6.301  83.7  2.7831  24 666    20.2
## 460  6.80117   0.0 18.10    0 0.7130 6.081  84.4  2.7175  24 666    20.2
## 461  4.81213   0.0 18.10    0 0.7130 6.701  90.0  2.5975  24 666    20.2
## 462  3.69311   0.0 18.10    0 0.7130 6.376  88.4  2.5671  24 666    20.2
## 463  6.65492   0.0 18.10    0 0.7130 6.317  83.0  2.7344  24 666    20.2
## 464  5.82115   0.0 18.10    0 0.7130 6.513  89.9  2.8016  24 666    20.2
## 465  7.83932   0.0 18.10    0 0.6550 6.209  65.4  2.9634  24 666    20.2
## 466  3.16360   0.0 18.10    0 0.6550 5.759  48.2  3.0665  24 666    20.2
## 467  3.77498   0.0 18.10    0 0.6550 5.952  84.7  2.8715  24 666    20.2
## 468  4.42228   0.0 18.10    0 0.5840 6.003  94.5  2.5403  24 666    20.2
## 469 15.57570   0.0 18.10    0 0.5800 5.926  71.0  2.9084  24 666    20.2
## 470 13.07510   0.0 18.10    0 0.5800 5.713  56.7  2.8237  24 666    20.2
## 471  4.34879   0.0 18.10    0 0.5800 6.167  84.0  3.0334  24 666    20.2
## 472  4.03841   0.0 18.10    0 0.5320 6.229  90.7  3.0993  24 666    20.2
## 473  3.56868   0.0 18.10    0 0.5800 6.437  75.0  2.8965  24 666    20.2
## 474  4.64689   0.0 18.10    0 0.6140 6.980  67.6  2.5329  24 666    20.2
## 475  8.05579   0.0 18.10    0 0.5840 5.427  95.4  2.4298  24 666    20.2
## 476  6.39312   0.0 18.10    0 0.5840 6.162  97.4  2.2060  24 666    20.2
## 477  4.87141   0.0 18.10    0 0.6140 6.484  93.6  2.3053  24 666    20.2
## 478 15.02340   0.0 18.10    0 0.6140 5.304  97.3  2.1007  24 666    20.2
## 479 10.23300   0.0 18.10    0 0.6140 6.185  96.7  2.1705  24 666    20.2
## 480 14.33370   0.0 18.10    0 0.6140 6.229  88.0  1.9512  24 666    20.2
## 481  5.82401   0.0 18.10    0 0.5320 6.242  64.7  3.4242  24 666    20.2
## 482  5.70818   0.0 18.10    0 0.5320 6.750  74.9  3.3317  24 666    20.2
## 483  5.73116   0.0 18.10    0 0.5320 7.061  77.0  3.4106  24 666    20.2
## 484  2.81838   0.0 18.10    0 0.5320 5.762  40.3  4.0983  24 666    20.2
## 485  2.37857   0.0 18.10    0 0.5830 5.871  41.9  3.7240  24 666    20.2
## 486  3.67367   0.0 18.10    0 0.5830 6.312  51.9  3.9917  24 666    20.2
## 487  5.69175   0.0 18.10    0 0.5830 6.114  79.8  3.5459  24 666    20.2
## 488  4.83567   0.0 18.10    0 0.5830 5.905  53.2  3.1523  24 666    20.2
## 489  0.15086   0.0 27.74    0 0.6090 5.454  92.7  1.8209   4 711    20.1
## 490  0.18337   0.0 27.74    0 0.6090 5.414  98.3  1.7554   4 711    20.1
## 491  0.20746   0.0 27.74    0 0.6090 5.093  98.0  1.8226   4 711    20.1
## 492  0.10574   0.0 27.74    0 0.6090 5.983  98.8  1.8681   4 711    20.1
## 493  0.11132   0.0 27.74    0 0.6090 5.983  83.5  2.1099   4 711    20.1
## 494  0.17331   0.0  9.69    0 0.5850 5.707  54.0  2.3817   6 391    19.2
## 495  0.27957   0.0  9.69    0 0.5850 5.926  42.6  2.3817   6 391    19.2
## 496  0.17899   0.0  9.69    0 0.5850 5.670  28.8  2.7986   6 391    19.2
## 497  0.28960   0.0  9.69    0 0.5850 5.390  72.9  2.7986   6 391    19.2
## 498  0.26838   0.0  9.69    0 0.5850 5.794  70.6  2.8927   6 391    19.2
## 499  0.23912   0.0  9.69    0 0.5850 6.019  65.3  2.4091   6 391    19.2
## 500  0.17783   0.0  9.69    0 0.5850 5.569  73.5  2.3999   6 391    19.2
## 501  0.22438   0.0  9.69    0 0.5850 6.027  79.7  2.4982   6 391    19.2
## 502  0.06263   0.0 11.93    0 0.5730 6.593  69.1  2.4786   1 273    21.0
## 503  0.04527   0.0 11.93    0 0.5730 6.120  76.7  2.2875   1 273    21.0
## 504  0.06076   0.0 11.93    0 0.5730 6.976  91.0  2.1675   1 273    21.0
## 505  0.10959   0.0 11.93    0 0.5730 6.794  89.3  2.3889   1 273    21.0
## 506  0.04741   0.0 11.93    0 0.5730 6.030  80.8  2.5050   1 273    21.0
```

```
##       black lstat medv
## 1   396.90  4.98 24.0
## 2   396.90  9.14 21.6
## 3   392.83  4.03 34.7
## 4   394.63  2.94 33.4
## 5   396.90  5.33 36.2
## 6   394.12  5.21 28.7
## 7   395.60 12.43 22.9
## 8   396.90 19.15 27.1
## 9   386.63 29.93 16.5
## 10  386.71 17.10 18.9
## 11  392.52 20.45 15.0
## 12  396.90 13.27 18.9
## 13  390.50 15.71 21.7
## 14  396.90  8.26 20.4
## 15  380.02 10.26 18.2
## 16  395.62  8.47 19.9
## 17  386.85  6.58 23.1
## 18  386.75 14.67 17.5
## 19  288.99 11.69 20.2
## 20  390.95 11.28 18.2
## 21  376.57 21.02 13.6
## 22  392.53 13.83 19.6
## 23  396.90 18.72 15.2
## 24  394.54 19.88 14.5
## 25  394.33 16.30 15.6
## 26  303.42 16.51 13.9
## 27  376.88 14.81 16.6
## 28  306.38 17.28 14.8
## 29  387.94 12.80 18.4
## 30  380.23 11.98 21.0
## 31  360.17 22.60 12.7
## 32  376.73 13.04 14.5
## 33  232.60 27.71 13.2
## 34  358.77 18.35 13.1
## 35  248.31 20.34 13.5
## 36  396.90  9.68 18.9
## 37  377.56 11.41 20.0
## 38  396.90  8.77 21.0
## 39  393.43 10.13 24.7
## 40  395.63  4.32 30.8
## 41  395.62  1.98 34.9
## 42  385.41  4.84 26.6
## 43  383.37  5.81 25.3
## 44  394.46  7.44 24.7
## 45  389.39  9.55 21.2
## 46  396.90 10.21 19.3
## 47  396.90 14.15 20.0
## 48  392.74 18.80 16.6
## 49  396.90 30.81 14.4
## 50  396.90 16.20 19.4
## 51  395.56 13.45 19.7
## 52  393.97  9.43 20.5
## 53  396.90  5.28 25.0
```

```
## 54  396.90  8.43 23.4
## 55  396.90 14.80 18.9
## 56  395.93  4.81 35.4
## 57  396.90  5.77 24.7
## 58  392.90  3.95 31.6
## 59  390.68  6.86 23.3
## 60  396.90  9.22 19.6
## 61  395.11 13.15 18.7
## 62  378.08 14.44 16.0
## 63  396.90  6.73 22.2
## 64  395.58  9.50 25.0
## 65  393.24  8.05 33.0
## 66  396.90  4.67 23.5
## 67  396.90 10.24 19.4
## 68  396.21  8.10 22.0
## 69  396.90 13.09 17.4
## 70  396.90  8.79 20.9
## 71  383.73  6.72 24.2
## 72  376.94  9.88 21.7
## 73  390.91  5.52 22.8
## 74  377.17  7.54 23.4
## 75  394.92  6.78 24.1
## 76  383.23  8.94 21.4
## 77  373.66 11.97 20.0
## 78  386.96 10.27 20.8
## 79  386.40 12.34 21.2
## 80  396.06  9.10 20.3
## 81  396.90  5.29 28.0
## 82  395.63  7.22 23.9
## 83  396.90  6.72 24.8
## 84  390.64  7.51 22.9
## 85  396.90  9.62 23.9
## 86  392.30  6.53 26.6
## 87  395.99 12.86 22.5
## 88  395.15  8.44 22.2
## 89  396.90  5.50 23.6
## 90  396.06  5.70 28.7
## 91  392.18  8.81 22.6
## 92  393.55  8.20 22.0
## 93  395.01  8.16 22.9
## 94  396.33  6.21 25.0
## 95  396.90 10.59 20.6
## 96  357.98  6.65 28.4
## 97  391.83 11.34 21.4
## 98  396.90  4.21 38.7
## 99  393.53  3.57 43.8
## 100 396.90  6.19 33.2
## 101 394.76  9.42 27.5
## 102 395.58  7.67 26.5
## 103  70.80 10.63 18.6
## 104 394.47 13.44 19.3
## 105 392.69 12.33 20.1
## 106 394.05 16.47 19.5
## 107 395.67 18.66 19.5
```

```
## 108 387.69 14.09 20.4
## 109 395.24 12.27 19.8
## 110 391.23 15.55 19.4
## 111 393.49 13.00 21.7
## 112 395.59 10.16 22.8
## 113 394.95 16.21 18.8
## 114 396.90 17.09 18.7
## 115 388.74 10.45 18.5
## 116 344.91 15.76 18.3
## 117 393.30 12.04 21.2
## 118 394.51 10.30 19.2
## 119 338.63 15.37 20.4
## 120 391.50 13.61 19.3
## 121 389.15 14.37 22.0
## 122 377.67 14.27 20.3
## 123 378.09 17.93 20.5
## 124 370.31 25.41 17.3
## 125 379.38 17.58 18.8
## 126 385.02 14.81 21.4
## 127 359.29 27.26 15.7
## 128 392.11 17.19 16.2
## 129 396.90 15.39 18.0
## 130 396.90 18.34 14.3
## 131 395.04 12.60 19.2
## 132 396.90 12.26 19.6
## 133 385.76 11.12 23.0
## 134 388.69 15.03 18.4
## 135 262.76 17.31 15.6
## 136 394.67 16.96 18.1
## 137 378.25 16.90 17.4
## 138 394.08 14.59 17.1
## 139 392.04 21.32 13.3
## 140 396.90 18.46 17.8
## 141 388.08 24.16 14.0
## 142 396.90 34.41 14.4
## 143 396.90 26.82 13.4
## 144 396.90 26.42 15.6
## 145 396.90 29.29 11.8
## 146 172.91 27.80 13.8
## 147 169.27 16.65 15.6
## 148 391.71 29.53 14.6
## 149 356.99 28.32 17.8
## 150 351.85 21.45 15.4
## 151 372.80 14.10 21.5
## 152 341.60 13.28 19.6
## 153 343.28 12.12 15.3
## 154 261.95 15.79 19.4
## 155 321.02 15.12 17.0
## 156  88.01 15.02 15.6
## 157  88.63 16.14 13.1
## 158 363.43  4.59 41.3
## 159 353.89  6.43 24.3
## 160 364.31  7.39 23.3
## 161 338.92  5.50 27.0
```

```
## 162 374.43  1.73 50.0
## 163 389.61  1.92 50.0
## 164 388.45  3.32 50.0
## 165 395.11 11.64 22.7
## 166 240.16  9.81 25.0
## 167 369.30  3.70 50.0
## 168 227.61 12.14 23.8
## 169 297.09 11.10 23.8
## 170 330.04 11.32 22.3
## 171 292.29 14.43 17.4
## 172 348.13 12.03 19.1
## 173 396.90 14.69 23.1
## 174 395.50  9.04 23.6
## 175 393.23  9.64 22.6
## 176 390.96  5.33 29.4
## 177 393.23 10.11 23.2
## 178 395.60  6.29 24.6
## 179 391.27  6.92 29.9
## 180 396.90  5.04 37.2
## 181 395.56  7.56 39.8
## 182 396.90  9.45 36.2
## 183 394.12  4.82 37.9
## 184 396.90  5.68 32.5
## 185 391.00 13.98 26.4
## 186 387.11 13.15 29.6
## 187 392.63  4.45 50.0
## 188 393.87  6.68 32.0
## 189 382.84  4.56 29.8
## 190 396.90  5.39 34.9
## 191 377.68  5.10 37.0
## 192 389.71  4.69 30.5
## 193 390.49  2.87 36.4
## 194 393.37  5.03 31.1
## 195 376.70  4.38 29.1
## 196 394.23  2.97 50.0
## 197 396.90  4.08 33.3
## 198 354.31  8.61 30.3
## 199 392.20  6.62 34.6
## 200 396.90  4.56 34.9
## 201 384.30  4.45 32.9
## 202 393.77  7.43 24.1
## 203 395.38  3.11 42.3
## 204 392.78  3.81 48.5
## 205 390.55  2.88 50.0
## 206 396.90 10.87 22.6
## 207 394.87 10.97 24.4
## 208 389.43 18.06 22.5
## 209 381.32 14.66 24.4
## 210 396.90 23.09 20.0
## 211 393.25 17.27 21.7
## 212 395.24 23.98 19.3
## 213 390.94 16.03 22.4
## 214 385.81  9.38 28.1
## 215 348.93 29.55 23.7
```

```
## 216 393.63  9.47 25.0
## 217 392.80 13.51 23.3
## 218 392.78  9.69 28.7
## 219 396.90 17.92 21.5
## 220 393.74 10.50 23.0
## 221 391.70  9.71 26.7
## 222 395.24 21.46 21.7
## 223 390.39  9.93 27.5
## 224 396.90  7.60 30.1
## 225 385.05  4.14 44.8
## 226 382.00  4.63 50.0
## 227 387.38  3.13 37.6
## 228 372.08  6.36 31.6
## 229 377.51  3.92 46.7
## 230 380.34  3.76 31.5
## 231 378.35 11.65 24.3
## 232 376.14  5.25 31.7
## 233 385.91  2.47 41.7
## 234 378.95  3.95 48.3
## 235 360.20  8.05 29.0
## 236 376.75 10.88 24.0
## 237 388.45  9.54 25.1
## 238 390.07  4.73 31.5
## 239 379.41  6.36 23.7
## 240 383.78  7.37 23.3
## 241 391.25 11.38 22.0
## 242 394.62 12.40 20.1
## 243 372.75 11.22 22.2
## 244 374.71  5.19 23.7
## 245 372.49 12.50 17.6
## 246 389.13 18.46 18.5
## 247 390.18  9.16 24.3
## 248 376.14 10.15 20.5
## 249 374.71  9.52 24.5
## 250 393.74  6.56 26.2
## 251 396.28  5.90 24.4
## 252 377.07  3.59 24.8
## 253 386.09  3.53 29.6
## 254 396.90  3.54 42.8
## 255 392.89  6.57 21.9
## 256 395.18  9.25 20.9
## 257 386.34  3.11 44.0
## 258 389.70  5.12 50.0
## 259 383.29  7.79 36.0
## 260 391.93  6.90 30.1
## 261 392.80  9.59 33.8
## 262 388.37  7.26 43.1
## 263 386.86  5.91 48.8
## 264 393.42 11.25 31.0
## 265 387.89  8.10 36.5
## 266 392.40 10.45 22.8
## 267 384.07 14.79 30.7
## 268 384.54  7.44 50.0
## 269 390.30  3.16 43.5
```

```
## 270 391.34 13.65 20.7
## 271 388.65 13.00 21.1
## 272 396.90  6.59 25.2
## 273 394.96  7.73 24.4
## 274 390.77  6.58 35.2
## 275 396.90  3.53 32.4
## 276 396.90  2.98 32.0
## 277 389.25  6.05 33.2
## 278 393.45  4.16 33.1
## 279 396.90  7.19 29.1
## 280 396.90  4.85 35.1
## 281 387.31  3.76 45.4
## 282 392.23  4.59 35.4
## 283 377.07  3.01 46.0
## 284 395.52  3.16 50.0
## 285 394.72  7.85 32.2
## 286 394.72  8.23 22.0
## 287 341.60 12.93 20.1
## 288 396.90  7.14 23.2
## 289 396.90  7.60 22.3
## 290 371.72  9.51 24.8
## 291 396.90  3.33 28.5
## 292 396.90  3.56 37.3
## 293 396.90  4.70 27.9
## 294 396.90  8.58 23.9
## 295 396.90 10.40 21.7
## 296 396.90  6.27 28.6
## 297 392.85  7.39 27.1
## 298 396.90 15.84 20.3
## 299 368.24  4.97 22.5
## 300 371.58  4.74 29.0
## 301 390.86  6.07 24.8
## 302 395.75  9.50 22.0
## 303 383.61  8.67 26.4
## 304 390.43  4.86 33.1
## 305 393.68  6.93 36.1
## 306 393.36  8.93 28.4
## 307 396.90  6.47 33.4
## 308 396.90  7.53 28.2
## 309 396.90  4.54 22.8
## 310 396.24  9.97 20.3
## 311 350.45 12.64 16.1
## 312 396.90  5.98 22.1
## 313 396.30 11.72 19.4
## 314 393.39  7.90 21.6
## 315 395.69  9.28 23.8
## 316 396.42 11.50 16.2
## 317 390.70 18.33 17.8
## 318 396.90 15.94 19.8
## 319 395.21 10.36 23.1
## 320 396.23 12.73 21.0
## 321 396.90  7.20 23.8
## 322 396.90  6.87 23.1
## 323 396.90  7.70 20.4
```

```
## 324 391.13 11.74 18.5
## 325 396.90  6.12 25.0
## 326 393.68  5.08 24.6
## 327 396.90  6.15 23.0
## 328 396.90 12.79 22.2
## 329 382.44  9.97 19.3
## 330 375.21  7.34 22.6
## 331 368.57  9.09 19.8
## 332 394.02 12.43 17.1
## 333 362.25  7.83 19.4
## 334 389.71  5.68 22.2
## 335 389.40  6.75 20.7
## 336 396.90  8.01 21.1
## 337 396.90  9.80 19.5
## 338 394.81 10.56 18.5
## 339 396.14  8.51 20.6
## 340 396.90  9.74 19.0
## 341 396.90  9.29 18.7
## 342 394.74  5.49 32.7
## 343 389.96  8.65 16.5
## 344 396.90  7.18 23.9
## 345 387.97  4.61 31.2
## 346 385.64 10.53 17.5
## 347 364.61 12.67 17.2
## 348 392.43  6.36 23.1
## 349 390.94  5.99 24.5
## 350 389.85  5.89 26.6
## 351 396.90  5.98 22.9
## 352 370.78  5.49 24.1
## 353 392.33  7.79 18.6
## 354 384.46  4.50 30.1
## 355 382.80  8.05 18.2
## 356 376.04  5.57 20.6
## 357 377.73 17.60 17.8
## 358 391.34 13.27 21.7
## 359 395.43 11.48 22.7
## 360 390.74 12.67 22.6
## 361 374.56  7.79 25.0
## 362 350.65 14.19 19.9
## 363 380.79 10.19 20.8
## 364 353.04 14.64 16.8
## 365 354.55  5.29 21.9
## 366 354.70  7.12 27.5
## 367 316.03 14.00 21.9
## 368 131.42 13.33 23.1
## 369 375.52  3.26 50.0
## 370 375.33  3.73 50.0
## 371 392.05  2.96 50.0
## 372 366.15  9.53 50.0
## 373 347.88  8.88 50.0
## 374 396.90 34.77 13.8
## 375 396.90 37.97 13.8
## 376 396.90 13.44 15.0
## 377 363.02 23.24 13.9
```

```
## 378 396.90 21.24 13.3
## 379 396.90 23.69 13.1
## 380 393.74 21.78 10.2
## 381 396.90 17.21 10.4
## 382 396.90 21.08 10.9
## 383 396.90 23.60 11.3
## 384 396.90 24.56 12.3
## 385 285.83 30.63  8.8
## 386 396.90 30.81  7.2
## 387 396.90 28.28 10.5
## 388 396.90 31.99  7.4
## 389 372.92 30.62 10.2
## 390 396.90 20.85 11.5
## 391 394.43 17.11 15.1
## 392 378.38 18.76 23.2
## 393 396.90 25.68  9.7
## 394 396.90 15.17 13.8
## 395 396.90 16.35 12.7
## 396 391.98 17.12 13.1
## 397 396.90 19.37 12.5
## 398 393.10 19.92  8.5
## 399 396.90 30.59  5.0
## 400 338.16 29.97  6.3
## 401 396.90 26.77  5.6
## 402 396.90 20.32  7.2
## 403 376.11 20.31 12.1
## 404 396.90 19.77  8.3
## 405 329.46 27.38  8.5
## 406 384.97 22.98  5.0
## 407 370.22 23.34 11.9
## 408 332.09 12.13 27.9
## 409 314.64 26.40 17.2
## 410 179.36 19.78 27.5
## 411   2.60 10.11 15.0
## 412  35.05 21.22 17.2
## 413  28.79 34.37 17.9
## 414 210.97 20.08 16.3
## 415  88.27 36.98  7.0
## 416  27.25 29.05  7.2
## 417  21.57 25.79  7.5
## 418 127.36 26.64 10.4
## 419  16.45 20.62  8.8
## 420  48.45 22.74  8.4
## 421 318.75 15.02 16.7
## 422 319.98 15.70 14.2
## 423 291.55 14.10 20.8
## 424   2.52 23.29 13.4
## 425   3.65 17.16 11.7
## 426   7.68 24.39  8.3
## 427  24.65 15.69 10.2
## 428  18.82 14.52 10.9
## 429  96.73 21.52 11.0
## 430  60.72 24.08  9.5
## 431  83.45 17.64 14.5
```

```
## 432  81.33 19.69 14.1
## 433  97.95 12.03 16.1
## 434 100.19 16.22 14.3
## 435 100.63 15.17 11.7
## 436 109.85 23.27 13.4
## 437  27.49 18.05  9.6
## 438   9.32 26.45  8.7
## 439  68.95 34.02  8.4
## 440 396.90 22.88 12.8
## 441 391.45 22.11 10.5
## 442 385.96 19.52 17.1
## 443 395.69 16.59 18.4
## 444 386.73 18.85 15.4
## 445 240.52 23.79 10.8
## 446  43.06 23.98 11.8
## 447 318.01 17.79 14.9
## 448 388.52 16.44 12.6
## 449 396.90 18.13 14.1
## 450 304.21 19.31 13.0
## 451   0.32 17.44 13.4
## 452 355.29 17.73 15.2
## 453 385.09 17.27 16.1
## 454 375.87 16.74 17.8
## 455   6.68 18.71 14.9
## 456  50.92 18.13 14.1
## 457  10.48 19.01 12.7
## 458   3.50 16.94 13.5
## 459 272.21 16.23 14.9
## 460 396.90 14.70 20.0
## 461 255.23 16.42 16.4
## 462 391.43 14.65 17.7
## 463 396.90 13.99 19.5
## 464 393.82 10.29 20.2
## 465 396.90 13.22 21.4
## 466 334.40 14.13 19.9
## 467  22.01 17.15 19.0
## 468 331.29 21.32 19.1
## 469 368.74 18.13 19.1
## 470 396.90 14.76 20.1
## 471 396.90 16.29 19.9
## 472 395.33 12.87 19.6
## 473 393.37 14.36 23.2
## 474 374.68 11.66 29.8
## 475 352.58 18.14 13.8
## 476 302.76 24.10 13.3
## 477 396.21 18.68 16.7
## 478 349.48 24.91 12.0
## 479 379.70 18.03 14.6
## 480 383.32 13.11 21.4
## 481 396.90 10.74 23.0
## 482 393.07  7.74 23.7
## 483 395.28  7.01 25.0
## 484 392.92 10.42 21.8
## 485 370.73 13.34 20.6
```

```
## 486 388.62 10.58 21.2
## 487 392.68 14.98 19.1
## 488 388.22 11.45 20.6
## 489 395.09 18.06 15.2
## 490 344.05 23.97  7.0
## 491 318.43 29.68  8.1
## 492 390.11 18.07 13.6
## 493 396.90 13.35 20.1
## 494 396.90 12.01 21.8
## 495 396.90 13.59 24.5
## 496 393.29 17.60 23.1
## 497 396.90 21.14 19.7
## 498 396.90 14.10 18.3
## 499 396.90 12.92 21.2
## 500 395.77 15.10 17.5
## 501 396.90 14.33 16.8
## 502 391.99  9.67 22.4
## 503 396.90  9.08 20.6
## 504 396.90  5.64 23.9
## 505 393.45  6.48 22.0
## 506 396.90  7.88 11.9
```

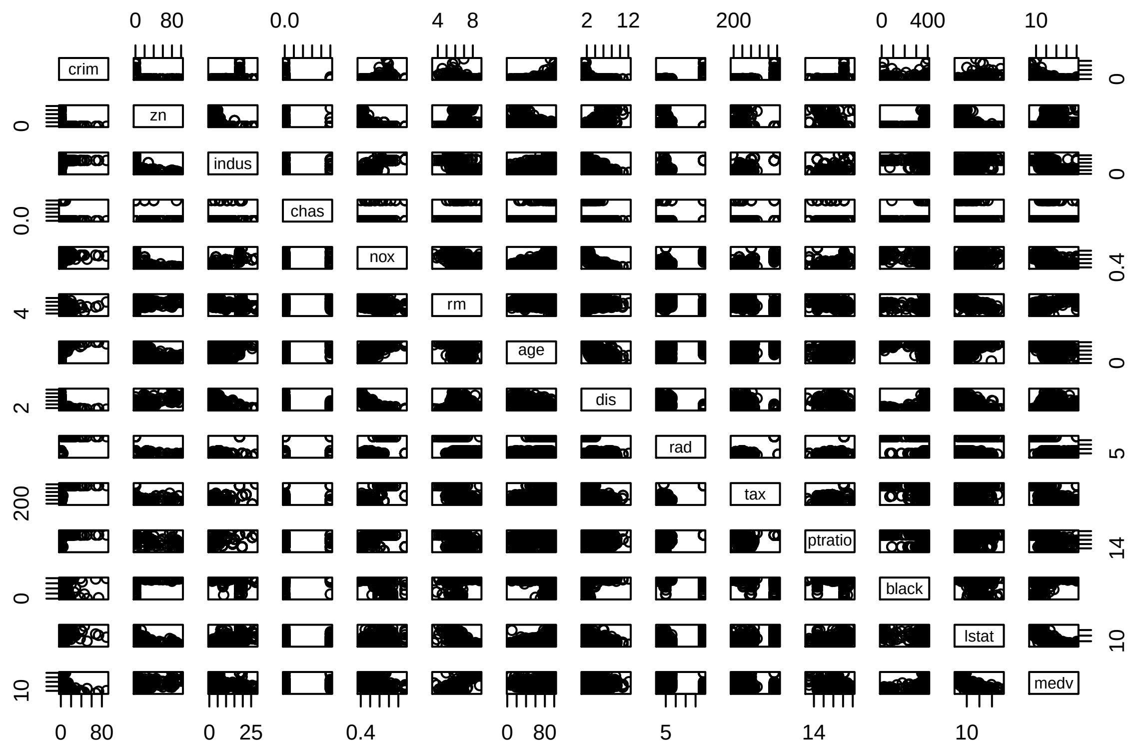
```
?Boston
```

There are 506 rows and 14 columns in the data set Boston. The 506 rows represent 506 suburban areas of Boston. The 14 rows represent 14 variables for each suburban area of Boston. The 14 variables for each suburb are:

1. crim: per capita crime rate by town.
2. zn: proportion of residential land zoned for lots over 25,000 sq.ft.
3. indus: proportion of non-retail business acres per town.
4. chas: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
5. nox: nitrogen oxides concentration (parts per 10 million).
6. rm: average number of rooms per dwelling.
7. age: proportion of owner-occupied units built prior to 1940.
8. dis: weighted mean of distances to five Boston employment centres.
9. rad: index of accessibility to radial highways.
10. tax: full-value property-tax rate per $10,000.
11. ptratio: pupil-teacher ratio by town.
12. black: 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.
13. lstat:lower status of the population (percent).
14. medv: median value of owner-occupied homes in $1000s.

b) Make some pairwise scatterplots of the predictors (columns) in this data set. Describe your findings.

```r
pairs(Boston)
```




```r
attach(Boston)

#To look at pairwise scatterplots individually, I created the function:
plots <- function(x,y) {
  plot(x, y)
}
#Replace x and y with two Boston variables of your choice.
```

Associations for crime rate are answered in the next problem.

As zn increases, indus decreases. It logically makes sense, since as the proportion of land reserved for residence increases, the proportion of non-retail business acres should decrease. Lots of the zones with higher densities of residential lands are NOT bound by the Charles River (chas). High amounts of nitric oxide (nox), homes older than 1940 (age), and lower status percentage (lstat) are also concentrated in zones with fewer residence areas. The only variable that increases as zn increases is distance from city centers, which accurately reflects the suburban culture of America. rm, rad, ptratio, and tax seemed to have no correlation with zn (save a few outliers). Black and medv didn't have any correlation with zn beyond 0 zn, but at 0 or near 0 zn the frequency of blacks and the range of median values were across the board.

More indus is in higher nox, lower dis, lower tax, lower ptratio, higher lstat, and lower medv.

More nox is in higher age, lower dis, higher rad, higher tax, higher ptratio, and lower medv.

More rm is in higher dis, lower rad, lower tax, lower ptratio, lower lstat, and lower medv.

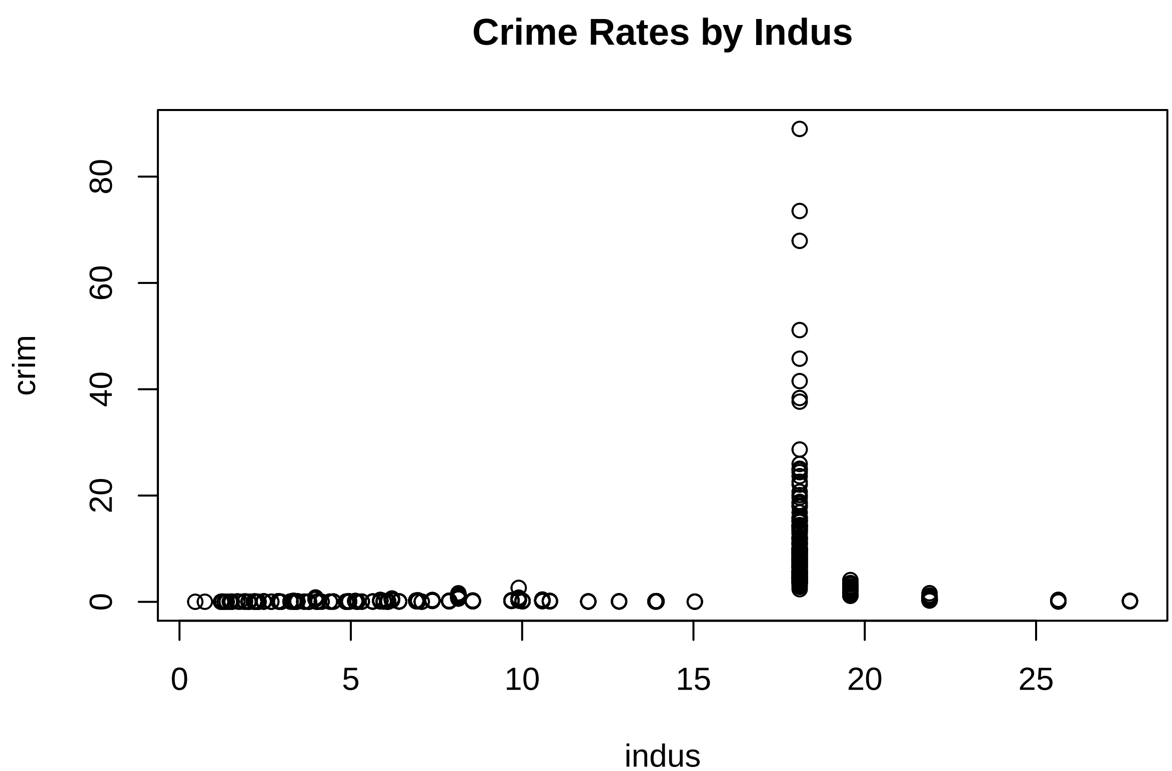
More age is lower dis, higher lstat, lower medv.

c) Are any of the predictors associated with per capita crime rate? If so, explain the relationship

Looking at the scatterplot matrix, zn appears to be associated with crim. Crime rates are especially concentrated in areas where no residential land is zoned for over 25000 sq. ft. - or areas not designated for residence.

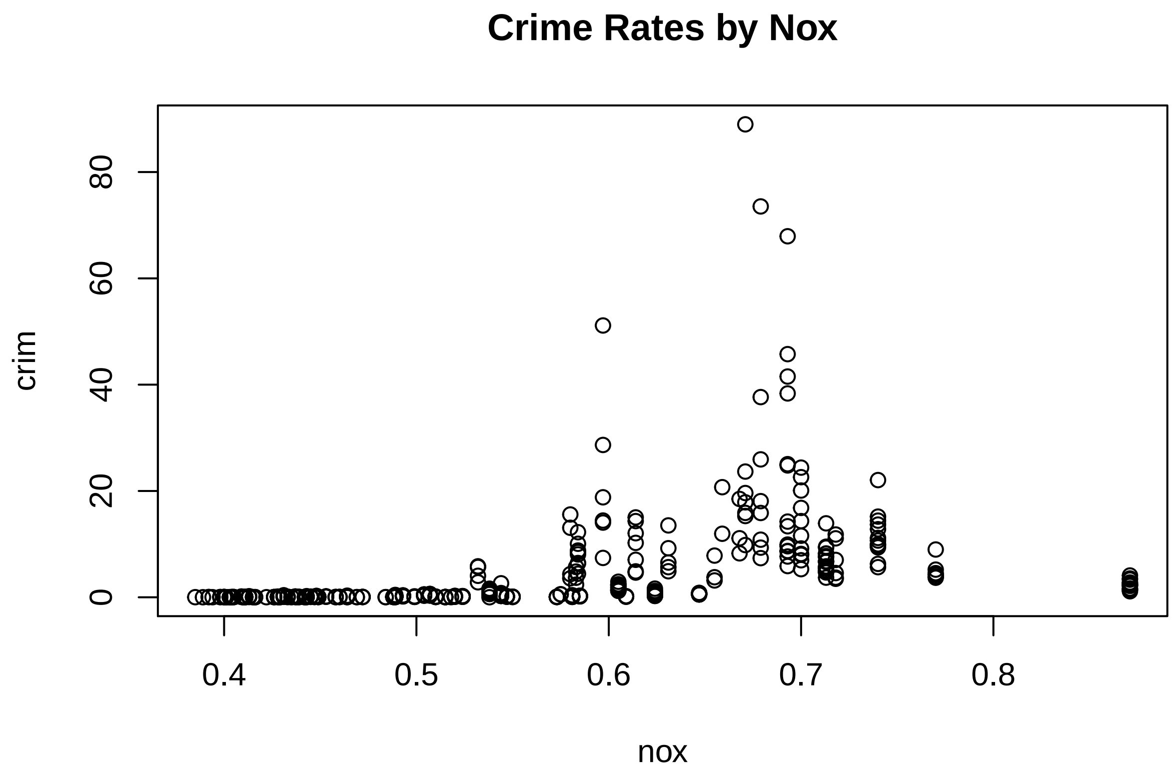
chas seems to be associated with crim. High crime rates are concentrated in areas that do NOT bound the Charles River.

```
plot(indus,crim, main = "Crime Rates by Indus")
```



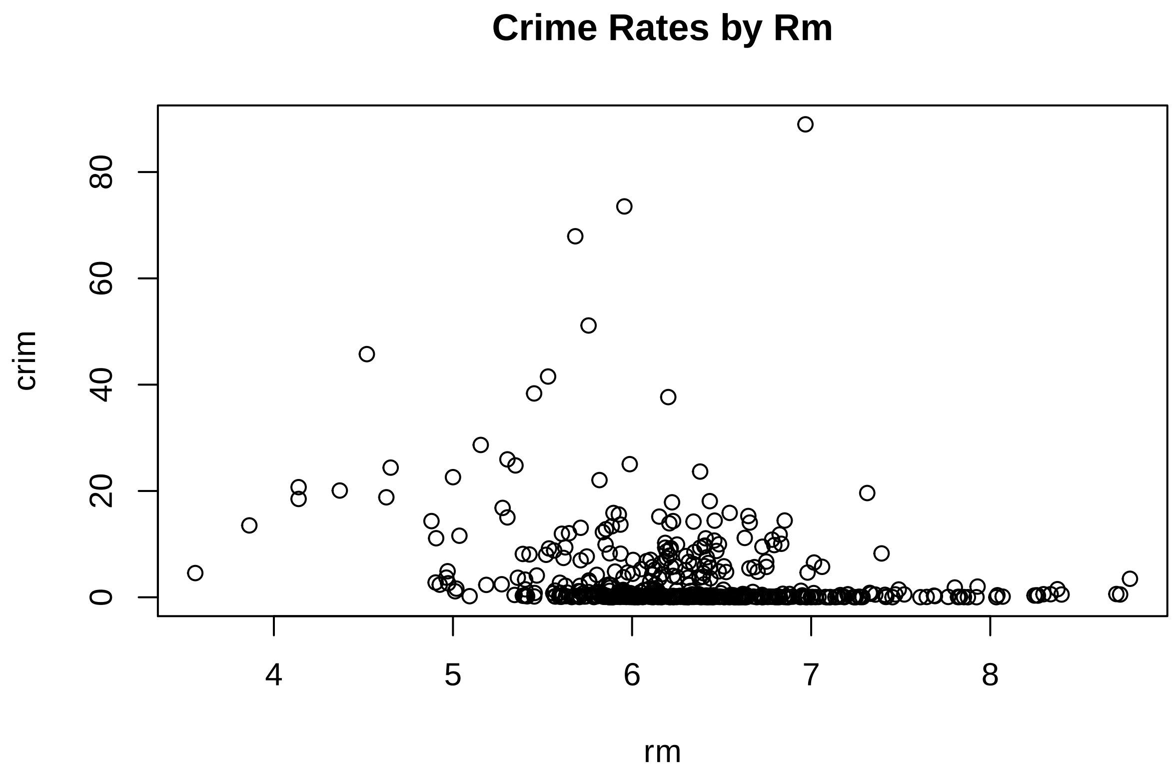
indus at first does not appear to be associated with crim, but at around 18 non-retail business acres, crime rate is especially high. Perhaps these 18 business acres are concentrated in specific high-crime areas in Boston.

```
plot(nox,crim, main = "Crime Rates by Nox")
```



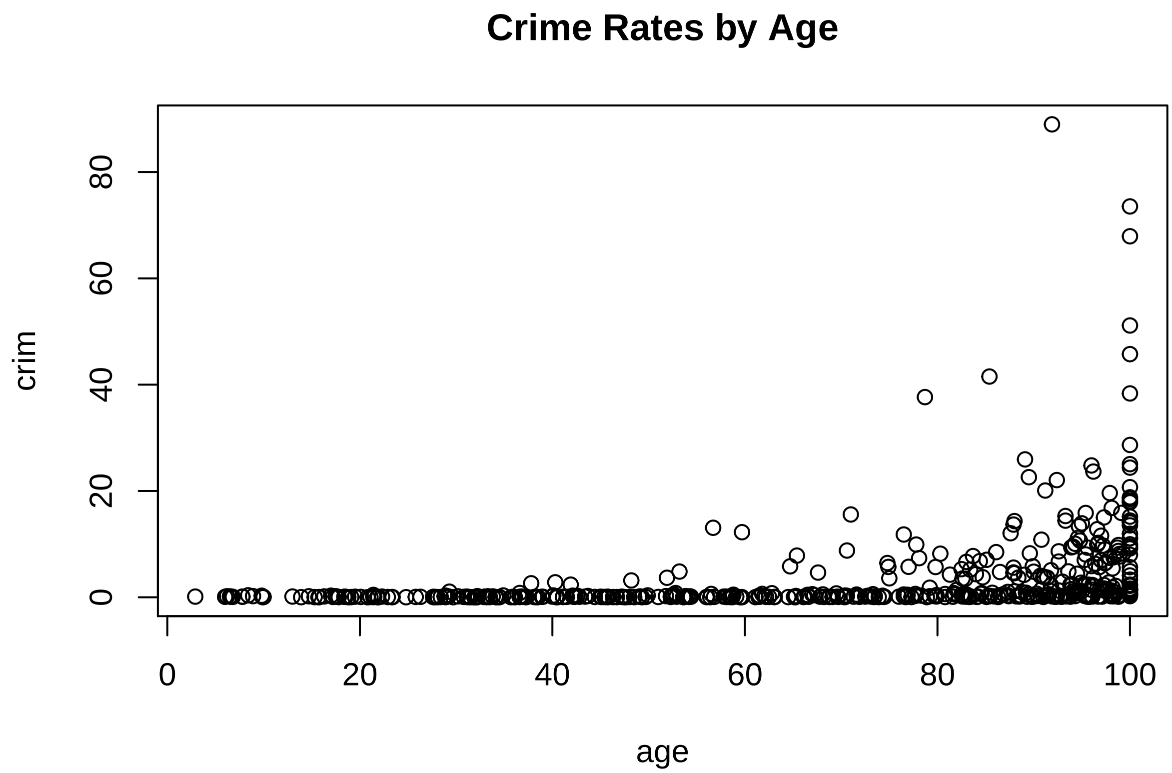
nox appears to be associated with crim. The higher the concentration of nitric oxide, the higher the crime rates, with the center of the distribution at around 0.7.

```
plot(rm, crim, main = "Crime Rates by Rm")
```



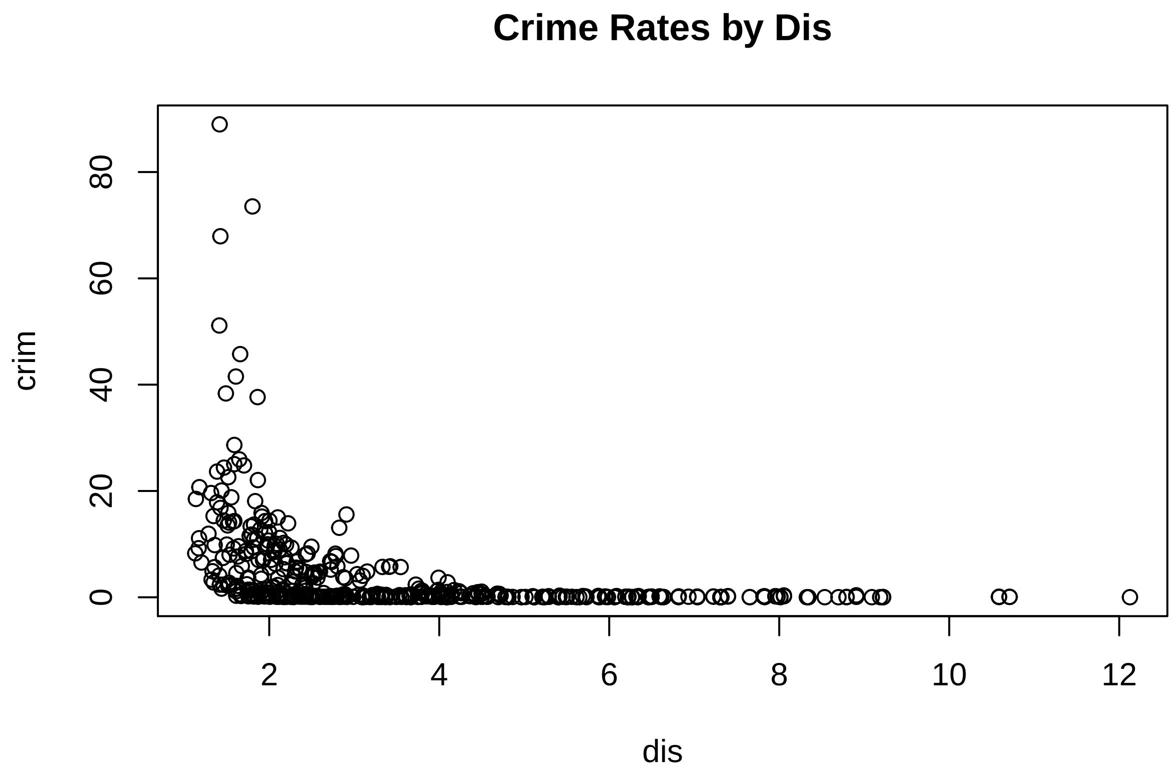
rm does NOT seem to be associated with crim. Both low and high crime rates can be found in both low and high average rooms per house, and the data distribution does not appear to follow a particular pattern.

```
plot(age, crim, main = "Crime Rates by Age")
```



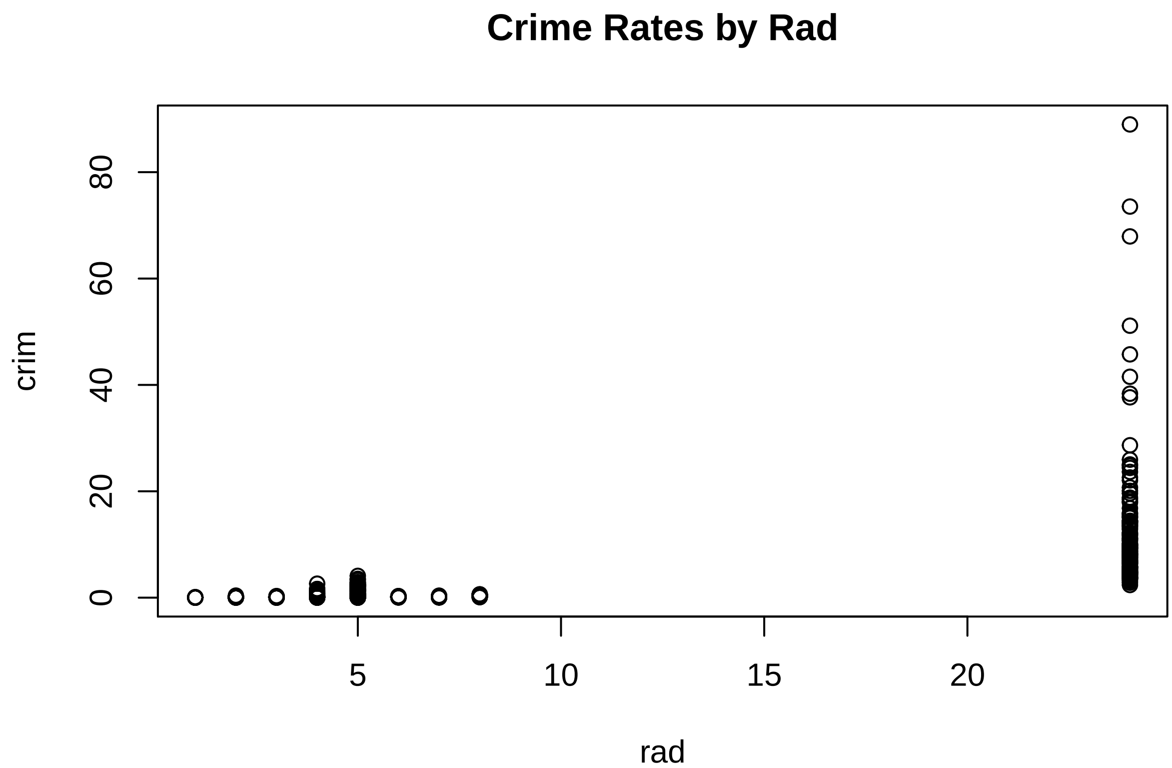
age seems to be associated with crim, with the majority of crime concentrated around areas of high proportion of pre-1940 houses, which suggests that high crime rates are associated with older neighborhoods.

```
plot(dis, crim, main = "Crime Rates by Dis")
```



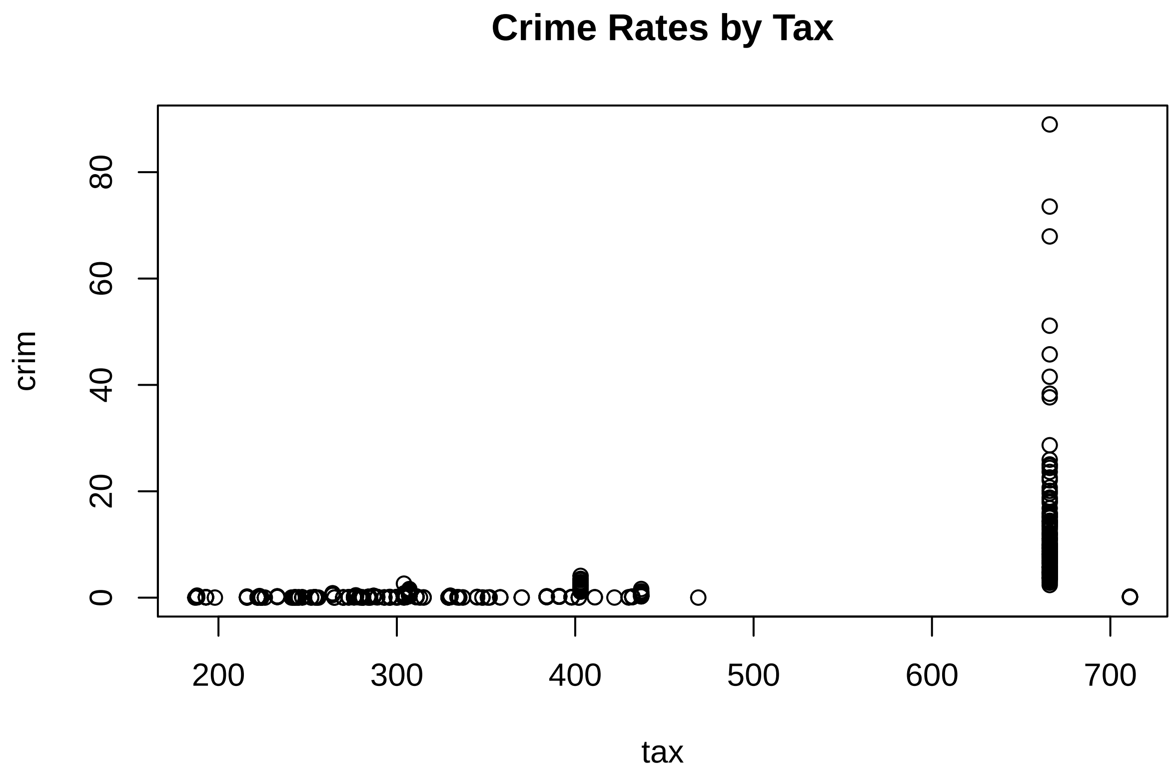
dis seems to be associated with crim, with the high crime rates concentrated around lower distances from city centers. This suggests that the closer an area is to city centers, the higher the crime rate.

```r
plot(rad, crim, main = "Crime Rates by Rad")
```



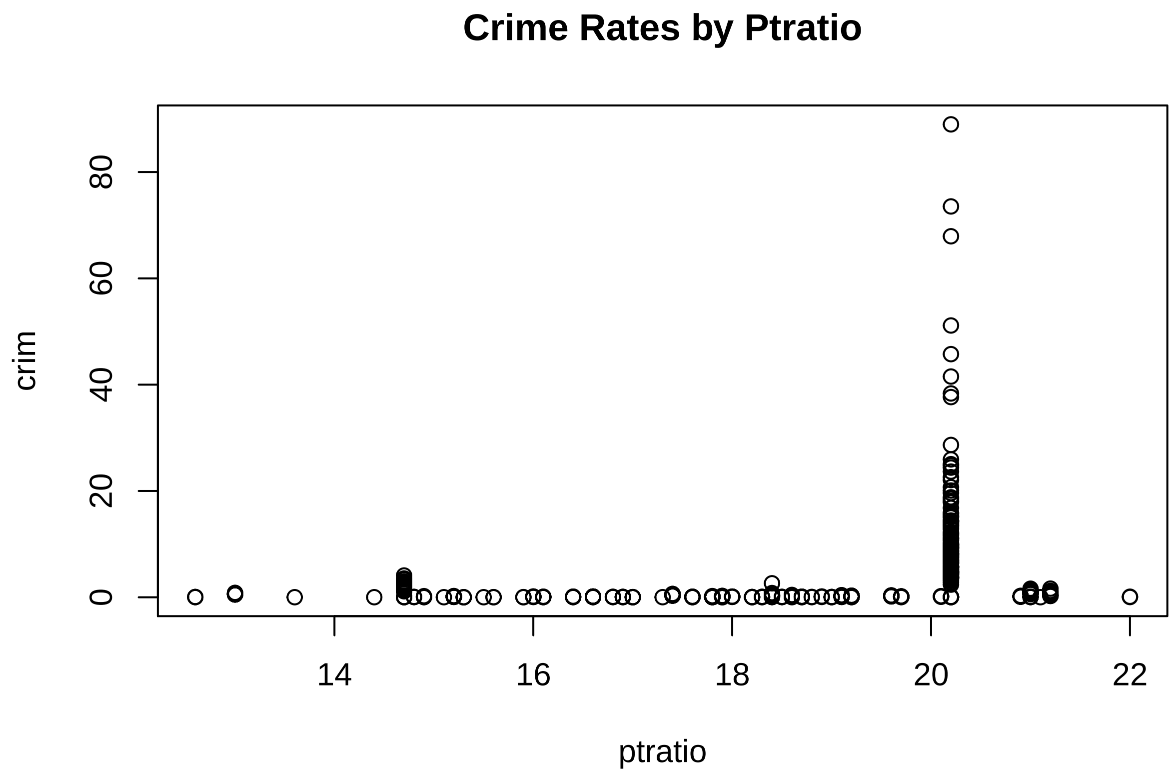
rad seems to be associated with crim. High crime rates are concentrated around the highest index of radial highways, suggesting that crime rates are highest when most accessible to central highways of a city.

```
plot(tax, crim, main = "Crime Rates by Tax")
```



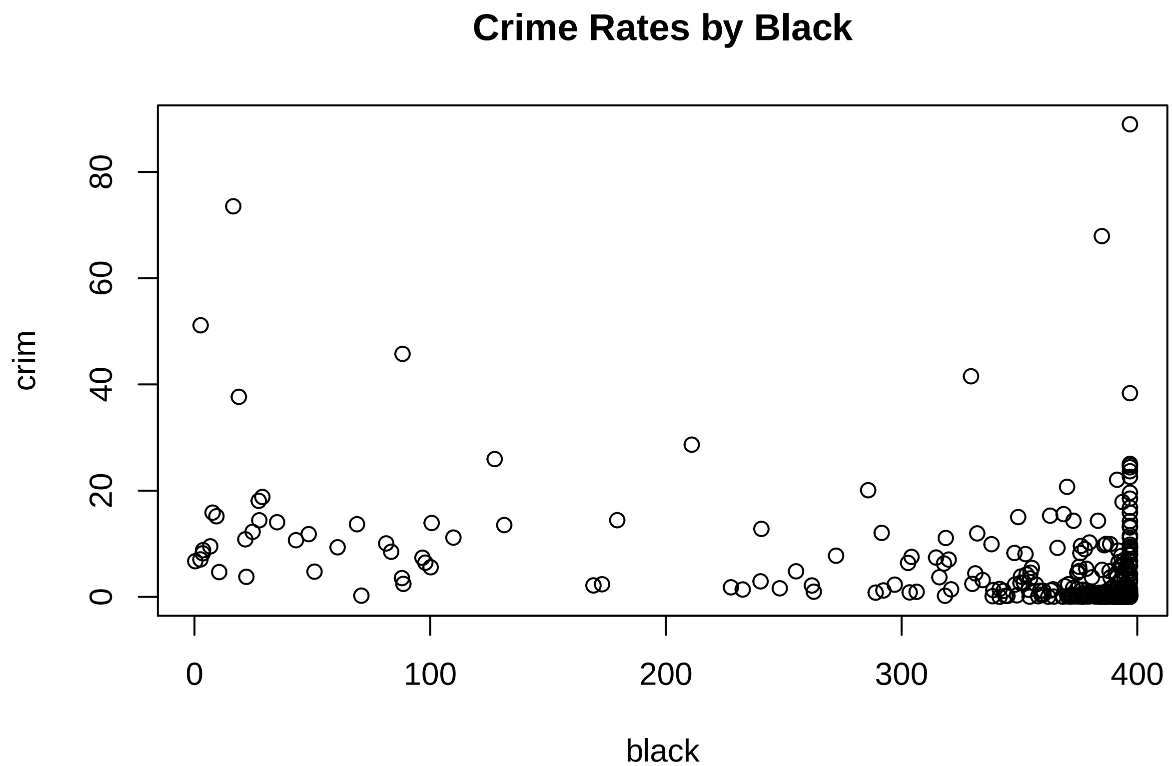
tax seems to be associated with crim, though the crime rates are concentrated around one high tax rate over the others.

```r
plot(ptratio, crim, main = "Crime Rates by Ptratio")
```



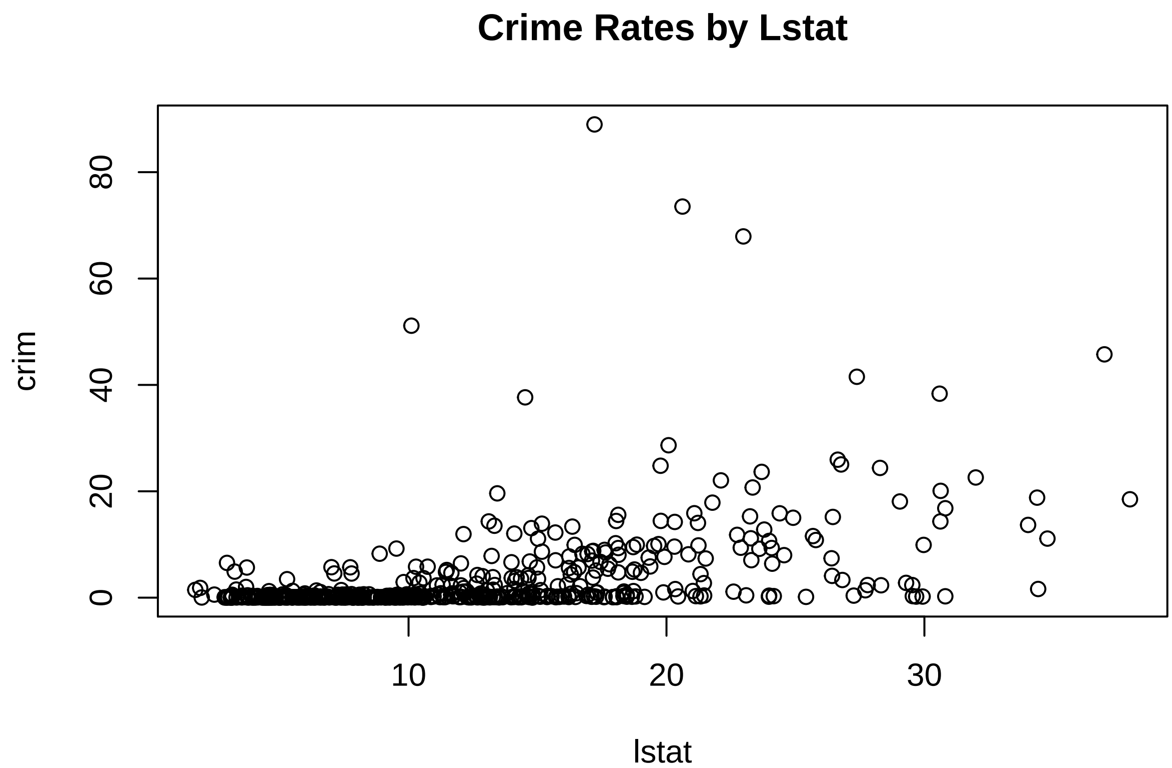
most of the crime rates are concentrated around one HIGH ptratio, suggesting that areas with either a high volume of students or low volume of teachers have higher crime rates. However, the high crime rates are not widely spread, with the high crime rates concentrated around one specific ptratio.

```
plot(black, crim, main = "Crime Rates by Black")
```



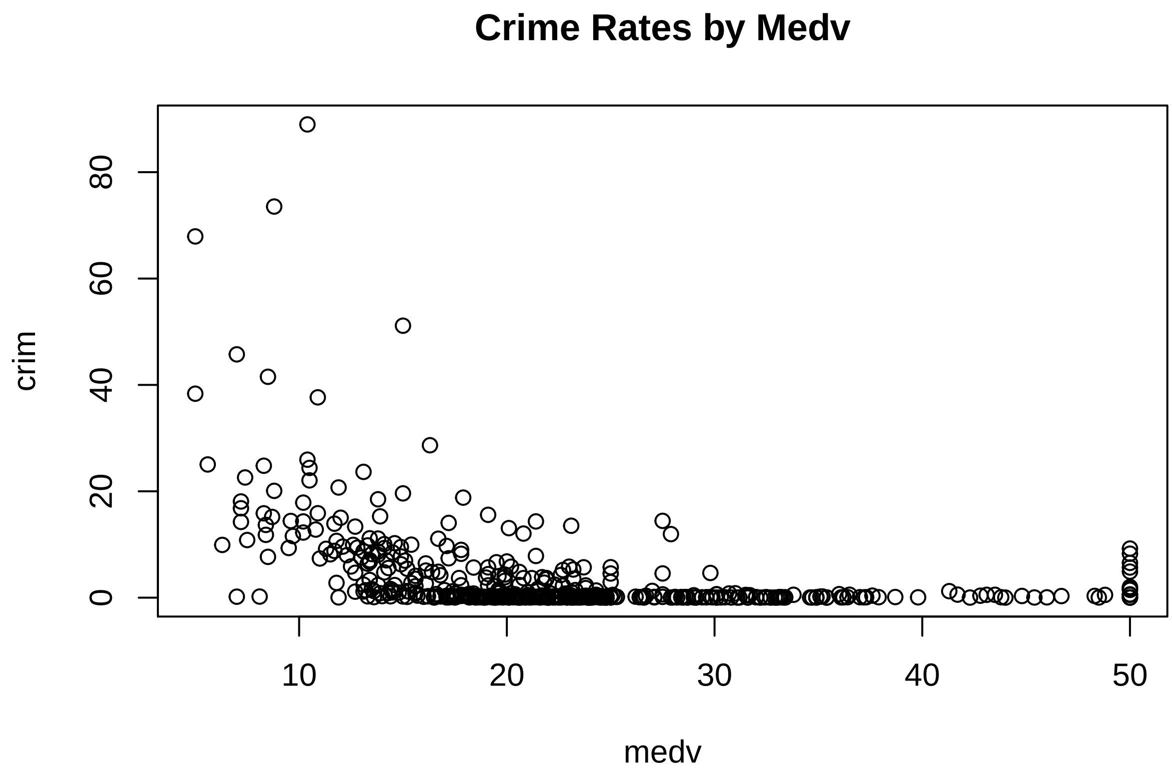
black does not seem to be highly correlated with crim. Most of the values of crime rate per capita are concentrated in suburbs of higher black populations, but the high crime rates per capita are scattered even in low black populations.

```
plot(lstat, crim, main = "Crime Rates by Lstat")
```



lstat seems to be slightly associated with crim. As the percentage of lower status goes up, the crime rate per capita goes up.

```
plot(medv, crim, main = "Crime Rates by Medv")
```



medv seems to be associated with crim. The lower the median value of homes of a suburb, the higher the crime rates.

d) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.

The first three lines show the values of the suburb with the highest crime rate, tax rate, and pupil-teacher ratio. The next lines display a table of the ranges of each predictor.

```
Boston[which.max(crim),]
```

```
##        crim zn indus chas   nox    rm  age    dis rad tax ptratio black
## 381 88.9762  0  18.1    0 0.671 6.968 91.9 1.4165  24 666    20.2 396.9
##     lstat medv
## 381 17.21 10.4
```

```
Boston[which.max(tax),]
```

```
##        crim zn indus chas   nox    rm  age    dis rad tax ptratio  black
## 489 0.15086  0 27.74    0 0.609 5.454 92.7 1.8209   4 711    20.1 395.09
##     lstat medv
## 489 18.06 15.2
```

```r
Boston[which.max(ptratio),]
```

```
##        crim zn indus chas   nox    rm  age     dis rad tax ptratio black
## 355 0.04301 80  1.91    0 0.413 5.663 21.9 10.5857   4 334      22 382.8
##     lstat medv
## 355  8.05 18.2
```

```r
minim <- numeric(ncol(Boston))
maxim <- numeric(ncol(Boston))
ranges <- data.frame(minim, maxim)
rownames(ranges) <- colnames(Boston)
for (i in 1:ncol(Boston)) {
  ranges[i,1] <- min(range(Boston[,i]))
  ranges[i,2] <- max(range(Boston[,i]))
}
ranges
```

```
##             minim    maxim
## crim      0.00632  88.9762
## zn        0.00000 100.0000
## indus     0.46000  27.7400
## chas      0.00000   1.0000
## nox       0.38500   0.8710
## rm        3.56100   8.7800
## age       2.90000 100.0000
## dis       1.12960  12.1265
## rad       1.00000  24.0000
## tax     187.00000 711.0000
## ptratio  12.60000  22.0000
## black     0.32000 396.9000
## lstat     1.73000  37.9700
## medv      5.00000  50.0000
```

The suburb with the highest crime rate seems to be have no residential areas in a zone (zn), on the higher end of non-retail business proportions (indus), not bound by the Charles River (chas), on the higher end of nitric oxide concentration (nox), more rooms per dwelling(rm), a higher proportion of older buildings (age), close to the city center (dis), the HIGHEST index of radial highway access (rad), high tax rate (tax), high pupil-teacher ratio (ptratio), the HIGHEST black population (black), and low median value (medv). The two variables that stuck out were rad and black because this particular suburb has the highest crime rate, highest index of accessibility to radial highways, and highest black population.

The suburb with the highest tax rate has a very low crime rate (crim), low residential zone (zn), the highest proportion of non-retail business (indus), not bound by the Charles River (chas), older buildings (age), close to the city center (dis), low access to radial highways (rad), high pupil-teacher ratio (ptratio), very high black population (black), and relatively low median values (medv). The biggest influence seems to come from indus, since it is this suburb's outlier as well as tax.

The suburb with the highest pupil-teacher ratio has a low crim, high zn, low indus, not bound by the Charles, slightly low age, long dis from city center, low rad, high black, low lstat, and slightly low medv.

e) How many of the suburbs in this data set bound the Charles River?

```r
sum(Boston[,'chas'] == 1)
```

```
## [1] 35
```

```r
# 36
```

f) What is the median pupil-teacher ratio among the towns in this data set?

```r
median(Boston[,'ptratio'])
```

```
## [1] 19.05
```

```r
#19:1
```

g) Which suburb of Boston has lowest median value of owneroccupied homes? What are the values of the other predictors for that suburb, and how do those values compare to the overall ranges for those predictors? Comment on your findings.

I first called the suburb (row) that had the smallest median value (medv). I then compared its values to the ranges of each variable in all of Boston.

```r
minmedv <- Boston[which.min(Boston[,'medv']),]
minmedv
```

```
##        crim zn indus chas   nox    rm age    dis rad tax ptratio black
## 399 38.3518  0  18.1    0 0.693 5.453 100 1.4896  24 666    20.2 396.9
##     lstat medv
## 399 30.59    5
```

```r
ranges
```

```
##             minim    maxim
## crim      0.00632  88.9762
## zn        0.00000 100.0000
## indus     0.46000  27.7400
## chas      0.00000   1.0000
## nox       0.38500   0.8710
## rm        3.56100   8.7800
```

```
## age        2.90000 100.0000
## dis        1.12960  12.1265
## rad        1.00000  24.0000
## tax      187.00000 711.0000
## ptratio   12.60000  22.0000
## black      0.32000 396.9000
## lstat      1.73000  37.9700
## medv       5.00000  50.0000
```

The suburb in row 399 has the lowest median value. It has a surprisingly low crime rate (crim), no residential zone (zn), not bound by the Charles (chas), a high nitric oxide concentration (nox), the HIGHEST proportion of old buildings (age), close to the city center (dis), the HIGHEST radial highway index (rad), high tax rate (tax), a high pupil-teacher ratio (ptratio), the HIGHEST proportion of blacks (black), and a high lower-status percentage (lstat). Age and black seem to be important associations, since it's also the highest in those respective categories.

h)

```r
nrow(Boston[which(Boston[,'rm'] > 7), ])
```

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

```r
nrow(Boston[which(Boston[,'rm'] > 8), ])
```

```
## [1] 13
```

```r
Boston[which(rm>8),]
```

```
##        crim zn indus chas    nox    rm  age    dis rad tax ptratio  black
## 98  0.12083  0  2.89    0 0.4450 8.069 76.0 3.4952   2 276    18.0 396.90
## 164 1.51902  0 19.58    1 0.6050 8.375 93.9 2.1620   5 403    14.7 388.45
## 205 0.02009 95  2.68    0 0.4161 8.034 31.9 5.1180   4 224    14.7 390.55
## 225 0.31533  0  6.20    0 0.5040 8.266 78.3 2.8944   8 307    17.4 385.05
## 226 0.52693  0  6.20    0 0.5040 8.725 83.0 2.8944   8 307    17.4 382.00
## 227 0.38214  0  6.20    0 0.5040 8.040 86.5 3.2157   8 307    17.4 387.38
## 233 0.57529  0  6.20    0 0.5070 8.337 73.3 3.8384   8 307    17.4 385.91
## 234 0.33147  0  6.20    0 0.5070 8.247 70.4 3.6519   8 307    17.4 378.95
## 254 0.36894 22  5.86    0 0.4310 8.259  8.4 8.9067   7 330    19.1 396.90
## 258 0.61154 20  3.97    0 0.6470 8.704 86.9 1.8010   5 264    13.0 389.70
## 263 0.52014 20  3.97    0 0.6470 8.398 91.5 2.2885   5 264    13.0 386.86
## 268 0.57834 20  3.97    0 0.5750 8.297 67.0 2.4216   5 264    13.0 384.54
## 365 3.47428  0 18.10    1 0.7180 8.780 82.9 1.9047  24 666    20.2 354.55
##     lstat medv
## 98   4.21 38.7
## 164  3.32 50.0
## 205  2.88 50.0
## 225  4.14 44.8
## 226  4.63 50.0
## 227  3.13 37.6
```

```
## 233  2.47 41.7
## 234  3.95 48.3
## 254  3.54 42.8
## 258  5.12 50.0
## 263  5.91 48.8
## 268  7.44 50.0
## 365  5.29 21.9
```

There are 64 suburbs with more than 7 rooms per dwelling, and 13 suburbs with more than 8 rooms per dwelling.

They all have low crime rates (crim), low percentages of lower-status citizens (lstat), and relatively high median values (medv). Most have low residental land densities (zn), small business areas (indus), not bound by the Charles (chas = 0), high proportions of pre-1940 homes (age), near city centers (dis), and low accessibility to radial highways (rad).

14 Chapter 3

a) Perform the commands in R below. The last line corresponds to creating a linear model in which y is a function of x1 and x2. Write out the form of the linear model. What are the regression coefficients?

```r
set.seed(1)
x1=runif (100)
x2=0.5*x1+rnorm (100)/10
y=2+2*x1+0.3*x2+rnorm (100)
```

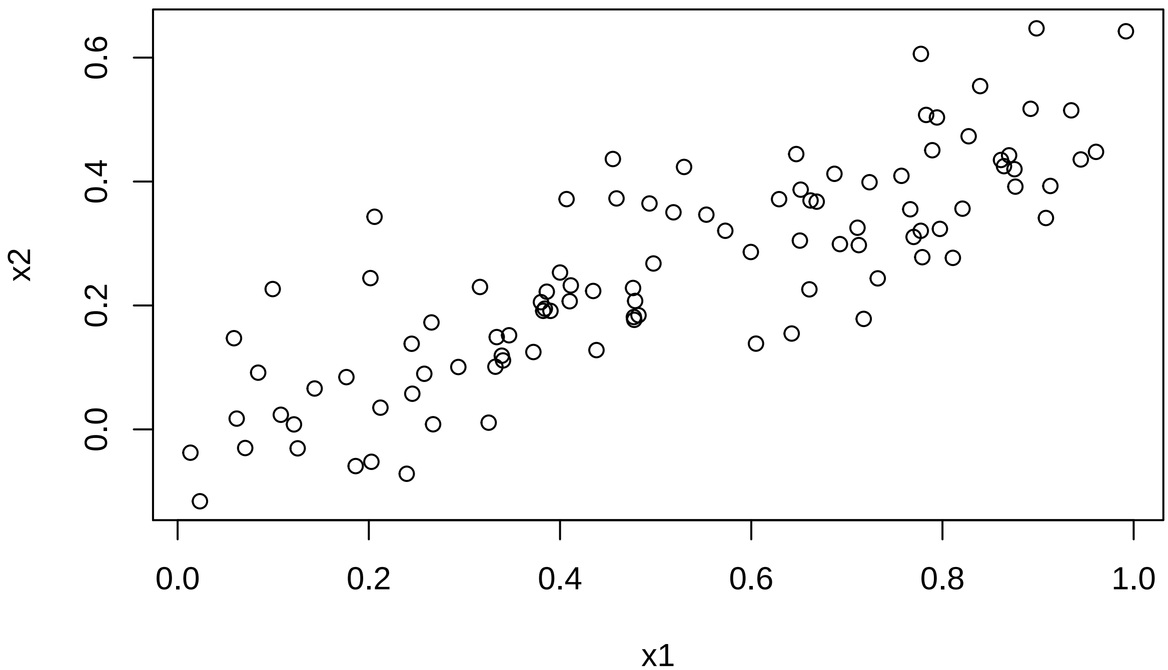
Y = B0 + B1X1 + B2X2 + E. The regression coefficients are B0 = 2, B1 = 2, and B2 = 0.3. X1 represents x1 and X2 represents x2.

b) What is the correlation between x1 and x2? Create a scatterplot displaying the relationship between the variables.

```r
cor(x1,x2)
```

```
## [1] 0.8351212
```

```r
plot(x1,x2)
```



Approx. 0.835

c) Using this data, fit a least squares regression to predict y using x1 and x2. Describe the results obtained. What are B^0, B^1, and B^2? How do these relate to the true B__0, B__1, and B__2? Can you reject the null hypothesis H0 : B__1 = 0? How about the null hypothesis H0 : B__2 = 0?

```r
summary(lm(y~x1+x2))
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8311 -0.7273 -0.0537  0.6338  2.3359
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
```

```
## x1            1.4396     0.7212   1.996   0.0487 *
## x2            1.0097     1.1337   0.891   0.3754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 97 degrees of freedom
## Multiple R-squared:  0.2088, Adjusted R-squared:  0.1925
## F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

According to the regression, the predicted regression coefficients are Bhat0 = 2.13, Bhat1 = 1.44, and Bhat2 = 1.01. Bhat0 and the true intercept, B0, are pretty close together, though Bhat0's p-value suggests that there is a significant difference between the predicted intercept and the true intercept. Bhat1 and true intercept B1 are not as close, and Bhat2 and true intercept B0 are pretty different.

According to the corresponding p-value, we can reject H0: B1 = 0 because its p-value falls under 0.05. We cannot reject H0: B2 = 0 because its corresponding p-value is 0.3754, which is above 0.05.

d) Now fit a least squares regression to predict y using only x1. Comment on your results. Can you reject the null hypothesis H0 : B_1 = 0?

```r
summary(lm(y~x1))
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89495 -0.66874 -0.07785  0.59221  2.45560
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
## x1            1.9759     0.3963   4.986 2.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.055 on 98 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1942
## F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

The p-value for the x1 coefficient became much lower than when x1 and x2 were both in the model. x1 became a much more significant influence on y. We can reject H0: B1 = 0 for most p-value cutoffs.

e) Now fit a least squares regression to predict y using only x2. Comment on your results. Can you reject the null hypothesis H0 : B_1 = 0?

```r
summary(lm(y~x2))
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62687 -0.75156 -0.03598  0.72383  2.44890
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
## x2            2.8996     0.6330    4.58 1.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.072 on 98 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1679
## F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

The effect of x2 becomes much more significant. Its p-value when taken into the linear model alone is drastically decreased from when x1 and x2 are taken together.

f) Do the results obtained in (c)-(e) contradict each other? Explain your answer.

Ostensibly, the results from c) thru e) contradict each other, since x2 was not significant in c) when taken together with x1 in them odel while it was significant in e) when taken alone. However, the disappearance of x2's significance when taken together with x1 indicates that x2 is simply a "surrogate" predictor to x1, and that x1 more directly influences Y.

g) Now suppose we obtain one additional observation, which was unfortunately mismeasured (first 3 lines below). Re-fit the linear models from (c) to (e) using this new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.

```r
x1=c(x1, 0.1)
x2=c(x2, 0.8)
y=c(y,6)

summary(lm(y~x1+x2))
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73348 -0.69318 -0.05263  0.66385  2.30619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2267     0.2314   9.624 7.91e-16 ***
## x1            0.5394     0.5922   0.911  0.36458
## x2            2.5146     0.8977   2.801  0.00614 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 98 degrees of freedom
## Multiple R-squared:  0.2188, Adjusted R-squared:  0.2029
## F-statistic: 13.72 on 2 and 98 DF,  p-value: 5.564e-06
```

```r
summary(lm(y~x1))
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8897 -0.6556 -0.0909  0.5682  3.5665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2569     0.2390   9.445 1.78e-15 ***
## x1            1.7657     0.4124   4.282 4.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 99 degrees of freedom
## Multiple R-squared:  0.1562, Adjusted R-squared:  0.1477
## F-statistic: 18.33 on 1 and 99 DF,  p-value: 4.295e-05
```

```r
summary(lm(y~x2))
```

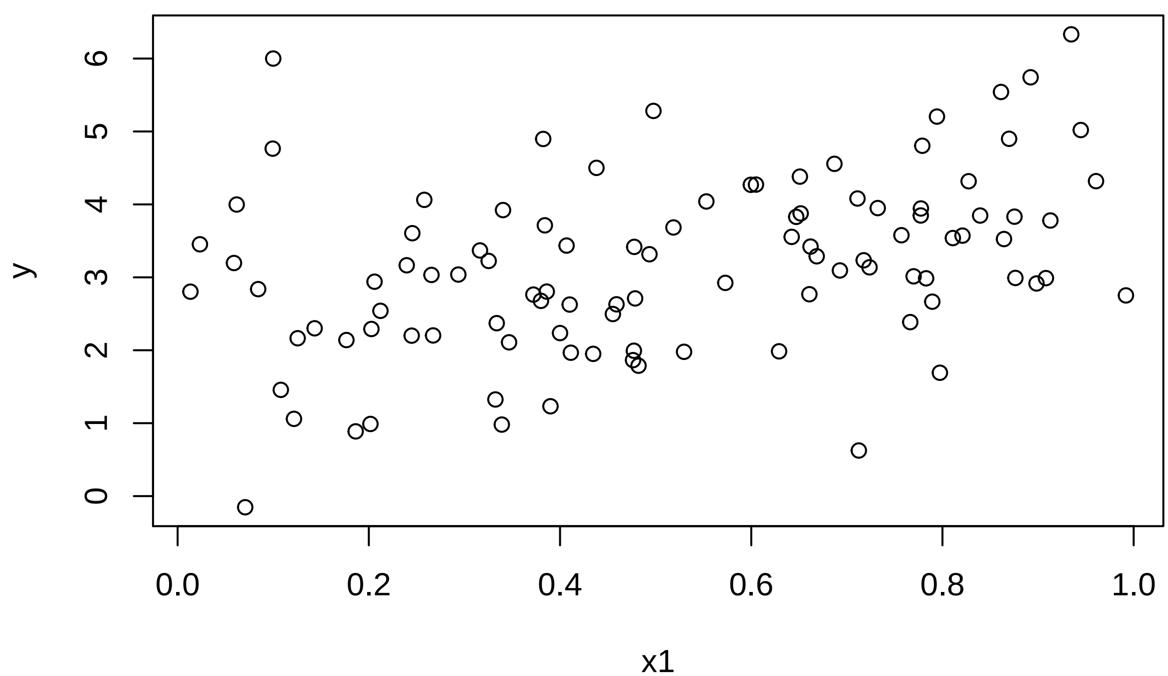
```
##
## Call:
```

```
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64729 -0.71021 -0.06899  0.72699  2.38074
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3451     0.1912  12.264  < 2e-16 ***
## x2            3.1190     0.6040   5.164 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.074 on 99 degrees of freedom
## Multiple R-squared:  0.2122, Adjusted R-squared:  0.2042
## F-statistic: 26.66 on 1 and 99 DF,  p-value: 1.253e-06
```

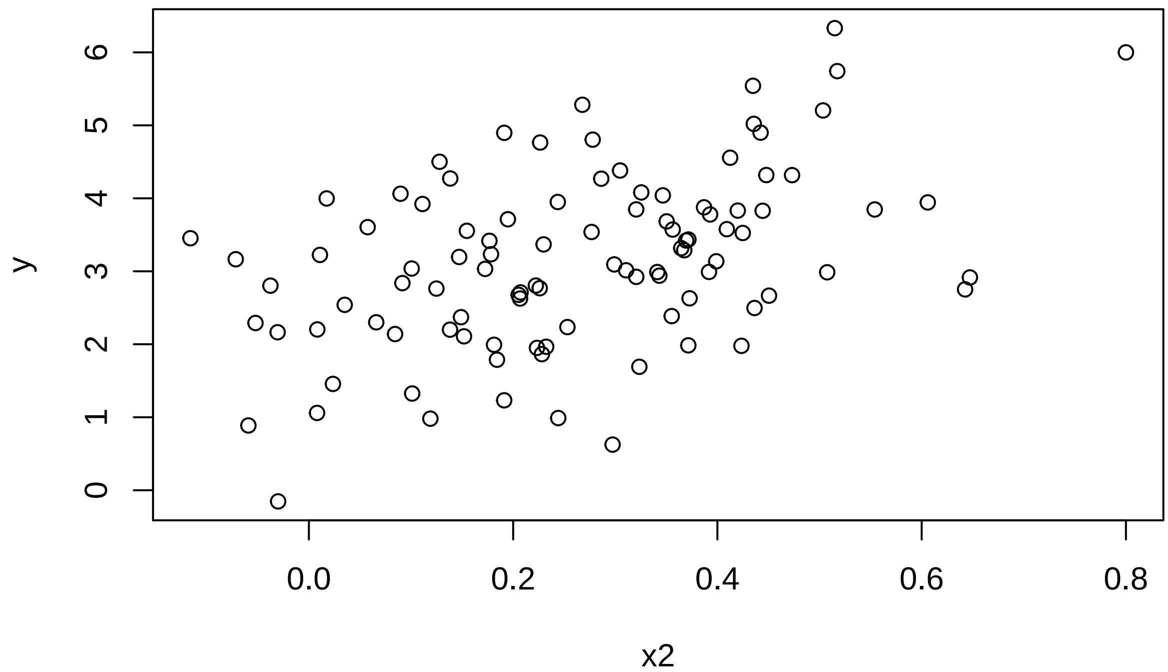
This new observation reverses the impacts of x1 and x2. x2 becomes significant both by itself and with x1 taken into the model, while x1 becomes insignificant when x2 is taken into the model and significant when alone. Now x1 becomes the surrogate. Also, the R^2 values and RSE's have increased when the new data was entered.

To find out whether the new data point is an outlier, a high-leverage point, or both, we plot each predictor against y, plot the residuals, and plot the hatvalues.

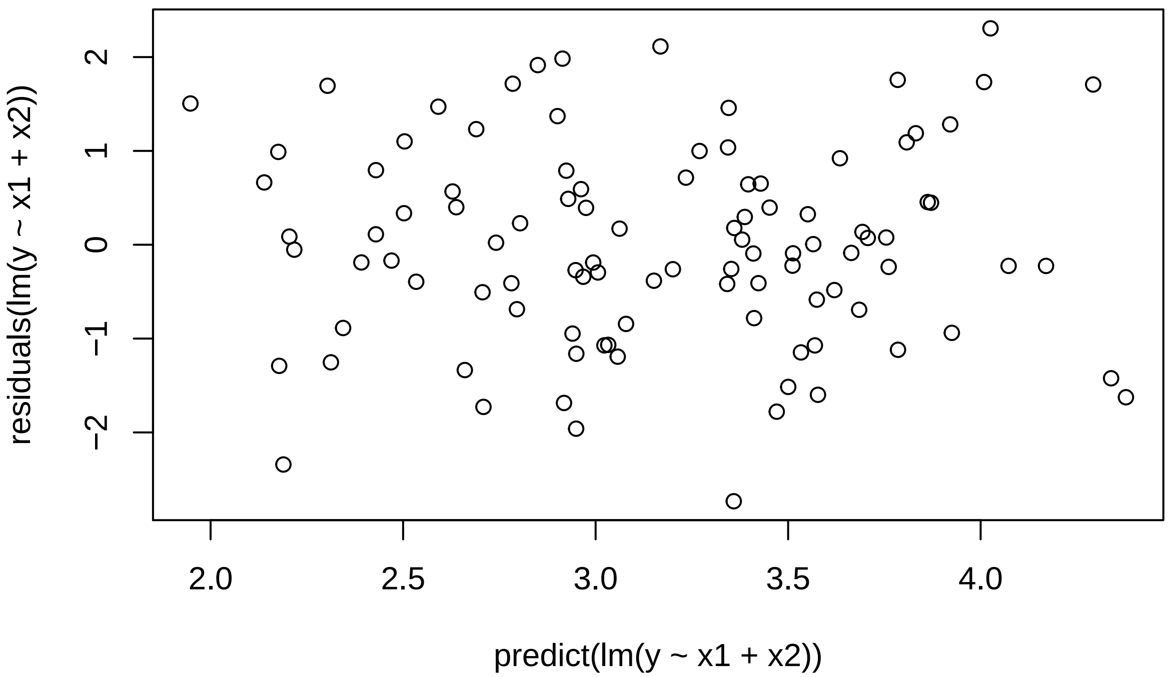
```r
plot(x1,y)
```



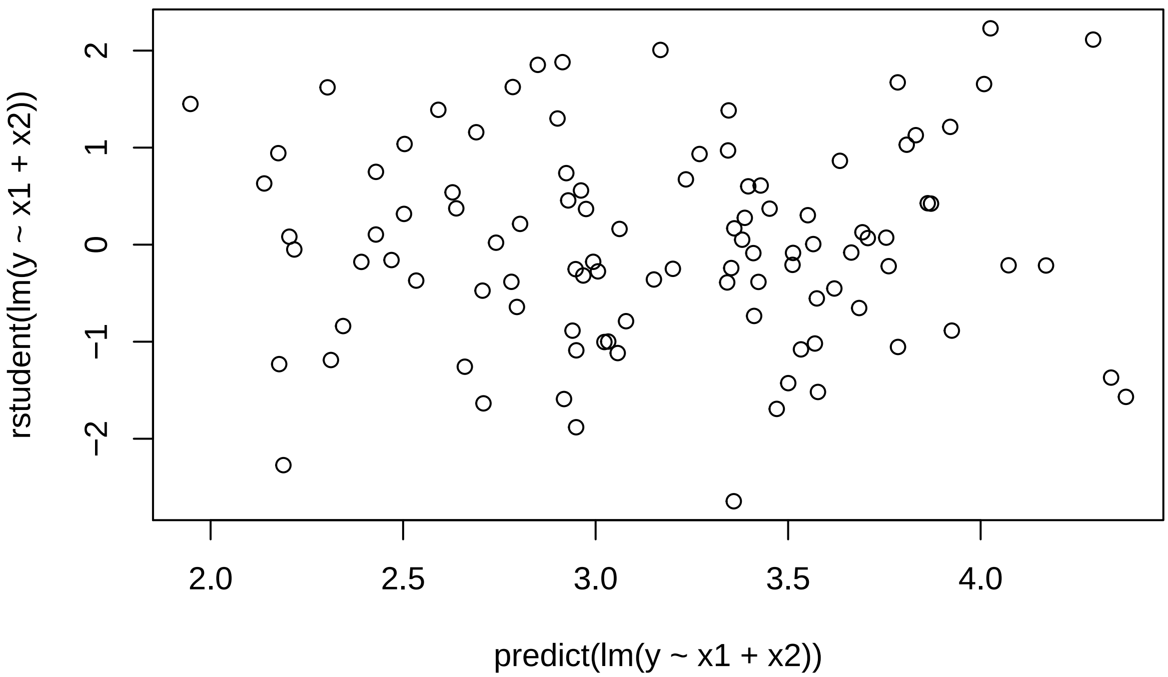
```r
plot(x2,y)
```



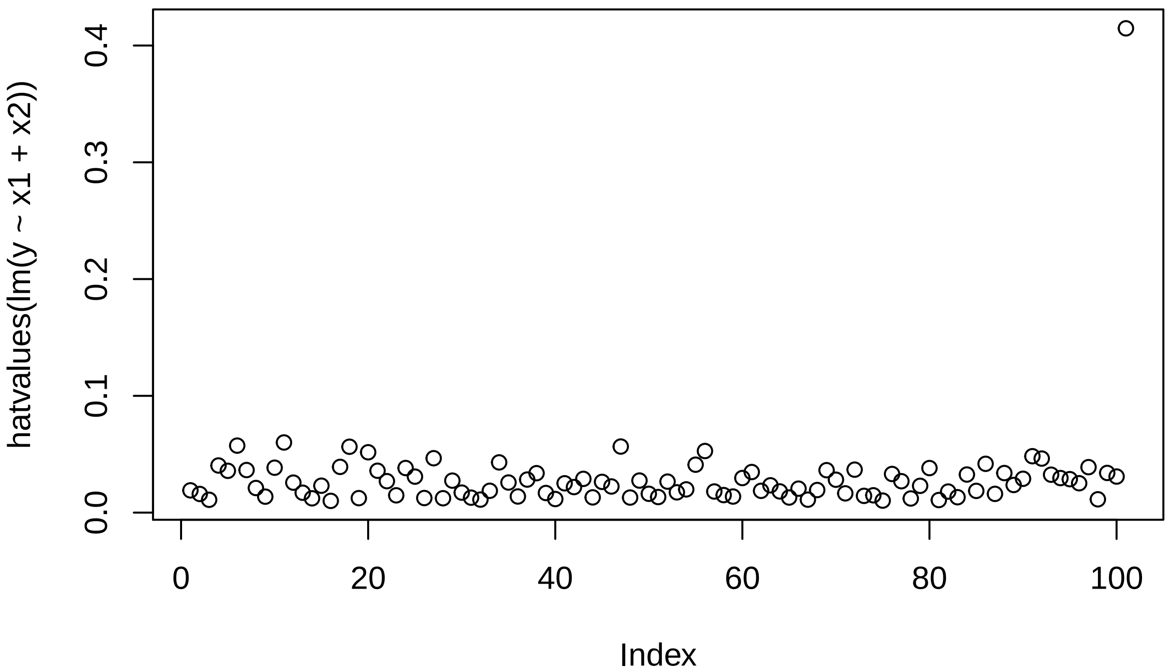

```r
plot(predict(lm(y~x1+x2)), residuals(lm(y~x1+x2)))
```




```r
plot(predict(lm(y~x1+x2)), rstudent(lm(y~x1+x2)))
```




```
plot(hatvalues(lm(y~x1+x2)))
```




```r
which.max(hatvalues(lm(y~x1+x2)))
```

```
## 101
## 101
```

According to the plot(x2,y), the added 0.8 adds a data point well beyond the other x2 values, which suggests that the new data point is a high-leverage point. Furthermore, when the regression summaries of y onto x2 alone are compared, the R^2 values are very different, which shows the drastic effect high-leverage points have on the fit of the regression. The plots of y vs. x1 and y vs. x2, as well as the residual plots, do not show extreme y values or extreme residual values, which suggests that the added data point is not an outlier. Looking at the hatvalues plot, the very last data point is well outside the range of hatvalues of the other previous data, which shows that the new data point is indeed a high-leverage point. The which.max function at the end simply confirms that the high-leverage point is indeed the new data point, since the old data had 100 observations, and the new data point would mean 101 observations total.

15 This problem involves the Boston data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

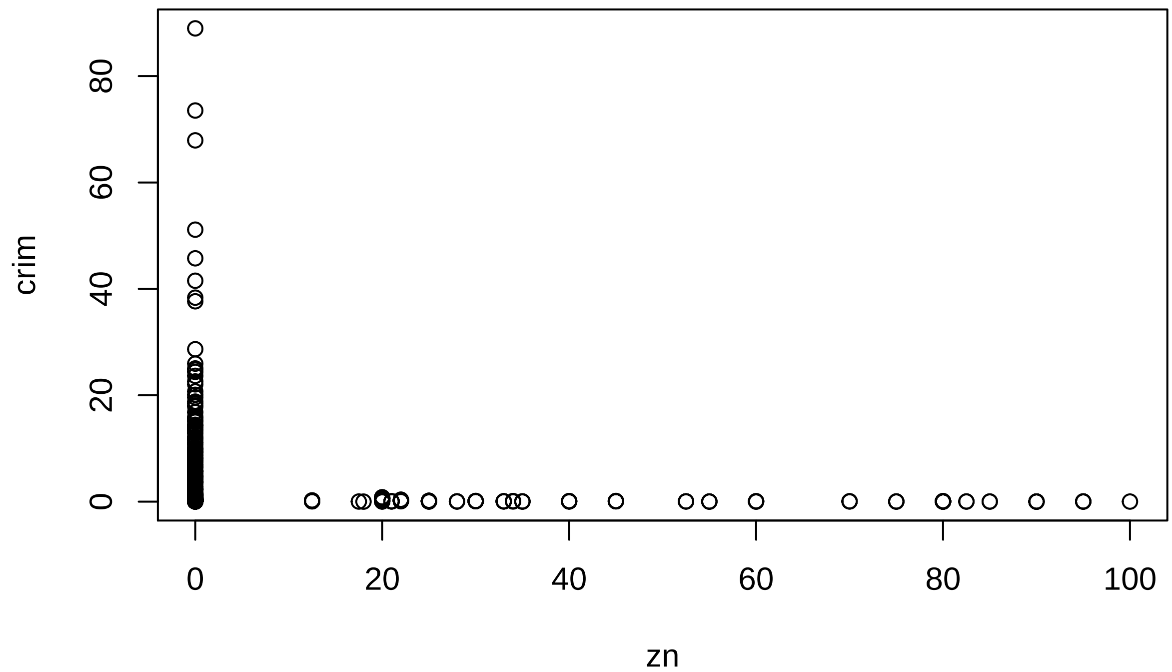
```r
summary(lm(crim~zn, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ zn, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.429 -4.222 -2.620  1.250 84.523
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.45369    0.41722  10.675  < 2e-16 ***
## zn          -0.07393    0.01609  -4.594 5.51e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.435 on 504 degrees of freedom
## Multiple R-squared:  0.04019,    Adjusted R-squared:  0.03828
## F-statistic:  21.1 on 1 and 504 DF,  p-value: 5.506e-06
```

```r
attach(Boston)
```

```
## The following objects are masked from Boston (pos = 3):
##
##     age, black, chas, crim, dis, indus, lstat, medv, nox, ptratio,
##     rad, rm, tax, zn
```

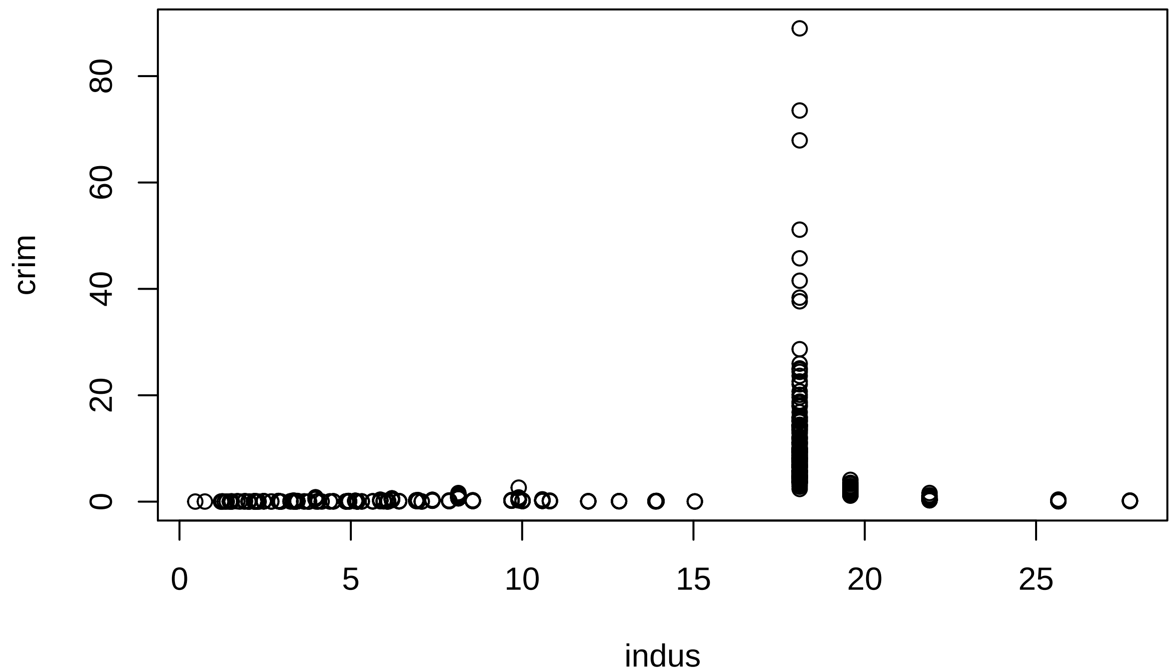
```r
plot(zn,crim)
```




```r
zncoef <- coef(summary(lm(crim~zn, data=Boston)))

summary(lm(crim~indus, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ indus, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.972  -2.698  -0.736   0.712  81.813
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.06374    0.66723  -3.093  0.00209 **
## indus        0.50978    0.05102   9.991  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.866 on 504 degrees of freedom
## Multiple R-squared:  0.1653, Adjusted R-squared:  0.1637
## F-statistic: 99.82 on 1 and 504 DF,  p-value: < 2.2e-16
```

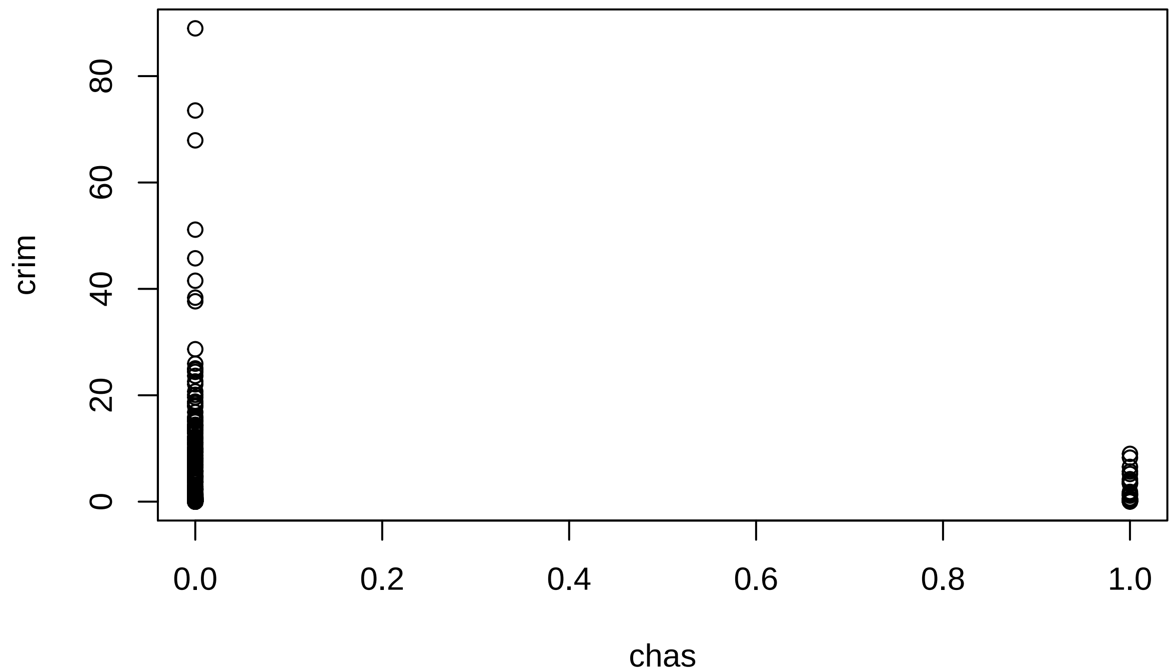
```r
plot(indus,crim)
```



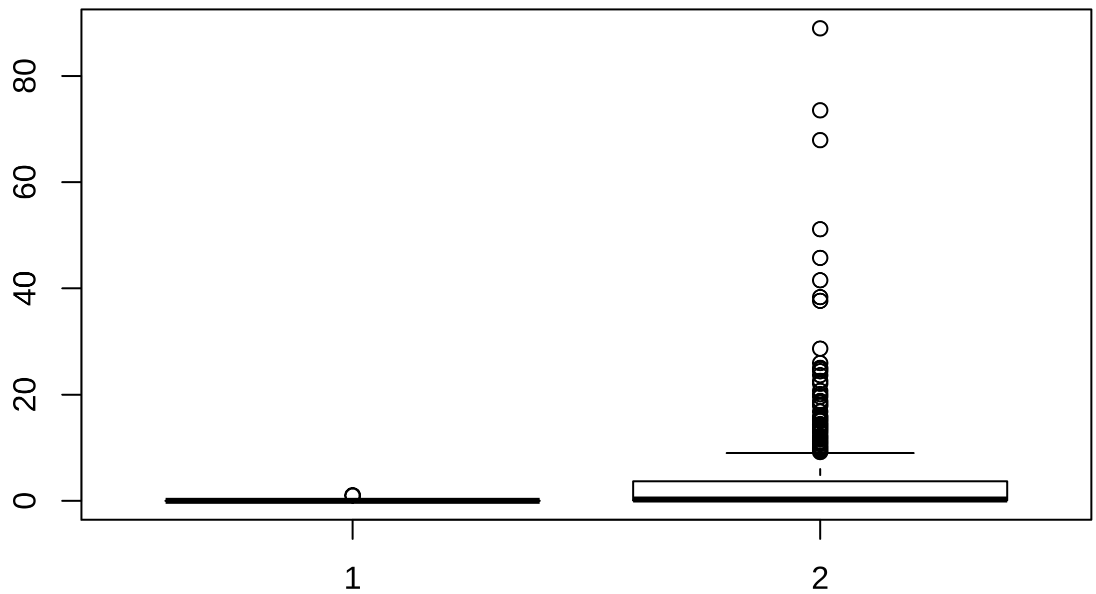

```r
induscoef <- coef(summary(lm(crim~indus, data=Boston)))

summary(lm(crim~chas, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ chas, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7444     0.3961   9.453   <2e-16 ***
## chas         -1.8928     1.5061  -1.257    0.209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.597 on 504 degrees of freedom
## Multiple R-squared:  0.003124,   Adjusted R-squared:  0.001146
## F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

```r
plot(chas,crim)
```



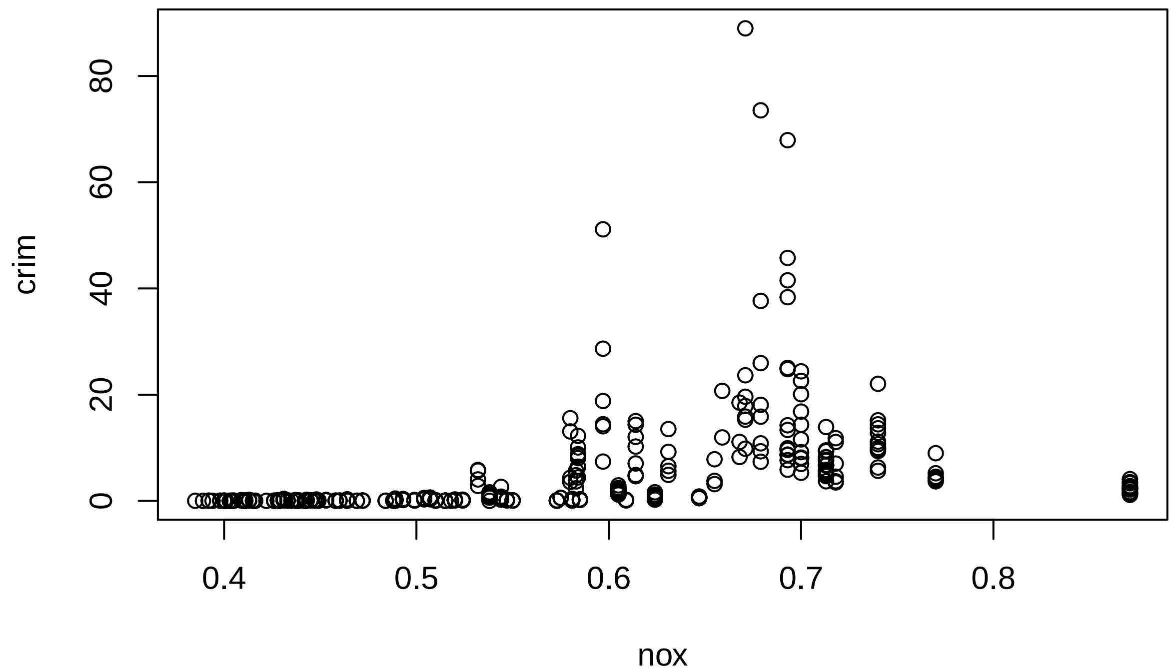

```
chascoef <- coef(summary(lm(crim~chas, data=Boston)))
boxplot(chas,crim)
```




```r
summary(lm(crim~nox, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ nox, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.371  -2.738  -0.974   0.559  81.728
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -13.720      1.699  -8.073 5.08e-15 ***
## nox           31.249      2.999  10.419  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.81 on 504 degrees of freedom
## Multiple R-squared:  0.1772, Adjusted R-squared:  0.1756
## F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

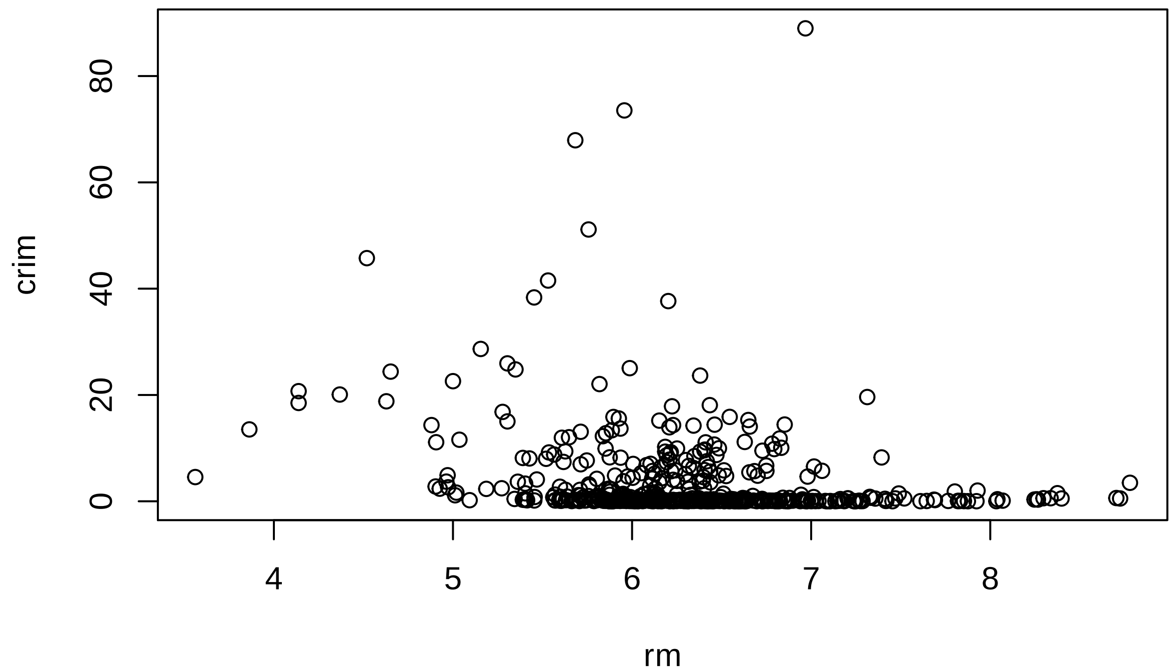
```r
plot(nox,crim)
```




```r
noxcoef <- coef(summary(lm(crim~nox, data=Boston)))

summary(lm(crim~rm, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ rm, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.604 -3.952 -2.654  0.989 87.197
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   20.482      3.365   6.088 2.27e-09 ***
## rm            -2.684      0.532  -5.045 6.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.401 on 504 degrees of freedom
## Multiple R-squared:  0.04807,    Adjusted R-squared:  0.04618
## F-statistic: 25.45 on 1 and 504 DF,  p-value: 6.347e-07
```

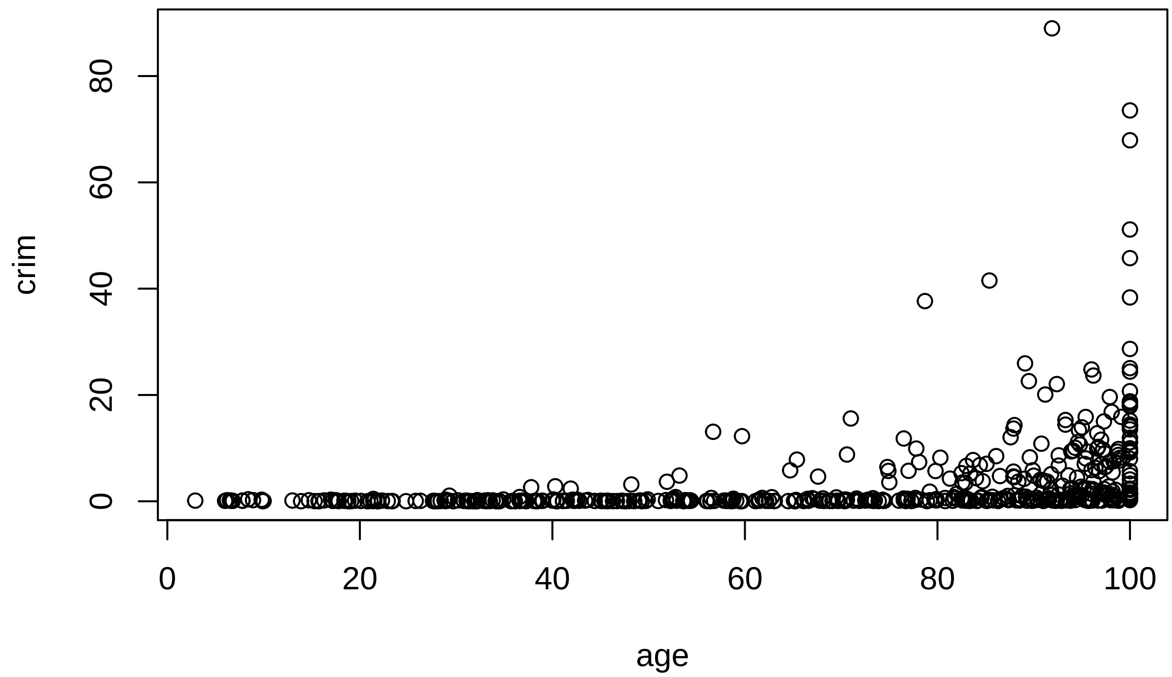
```r
plot(rm, crim)
```




```r
rmcoef <- coef(summary(lm(crim~rm, data=Boston)))

summary(lm(crim~age, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ age, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.789 -4.257 -1.230  1.527 82.849
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.77791    0.94398  -4.002 7.22e-05 ***
## age          0.10779    0.01274   8.463 2.85e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.057 on 504 degrees of freedom
## Multiple R-squared:  0.1244, Adjusted R-squared:  0.1227
## F-statistic: 71.62 on 1 and 504 DF,  p-value: 2.855e-16
```

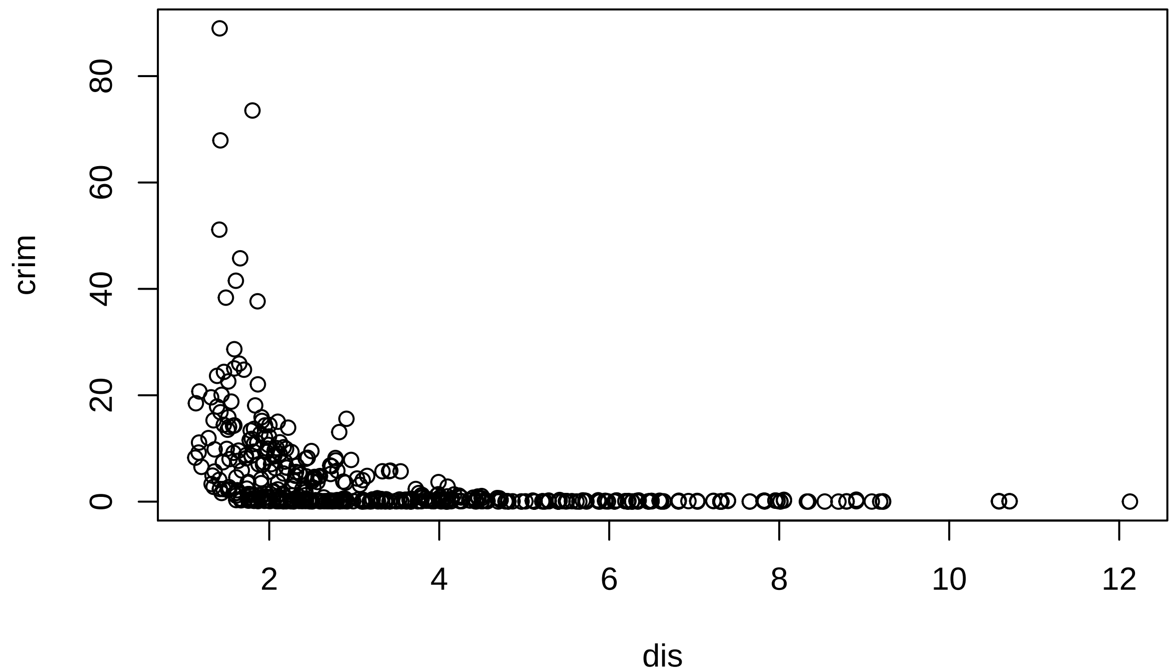
```r
plot(age,crim)
```




```r
agecoef <- coef(summary(lm(crim~age, data=Boston)))

summary(lm(crim~dis, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ dis, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.708 -4.134 -1.527  1.516 81.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.4993     0.7304  13.006   <2e-16 ***
## dis          -1.5509     0.1683  -9.213   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.965 on 504 degrees of freedom
## Multiple R-squared:  0.1441, Adjusted R-squared:  0.1425
## F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
```

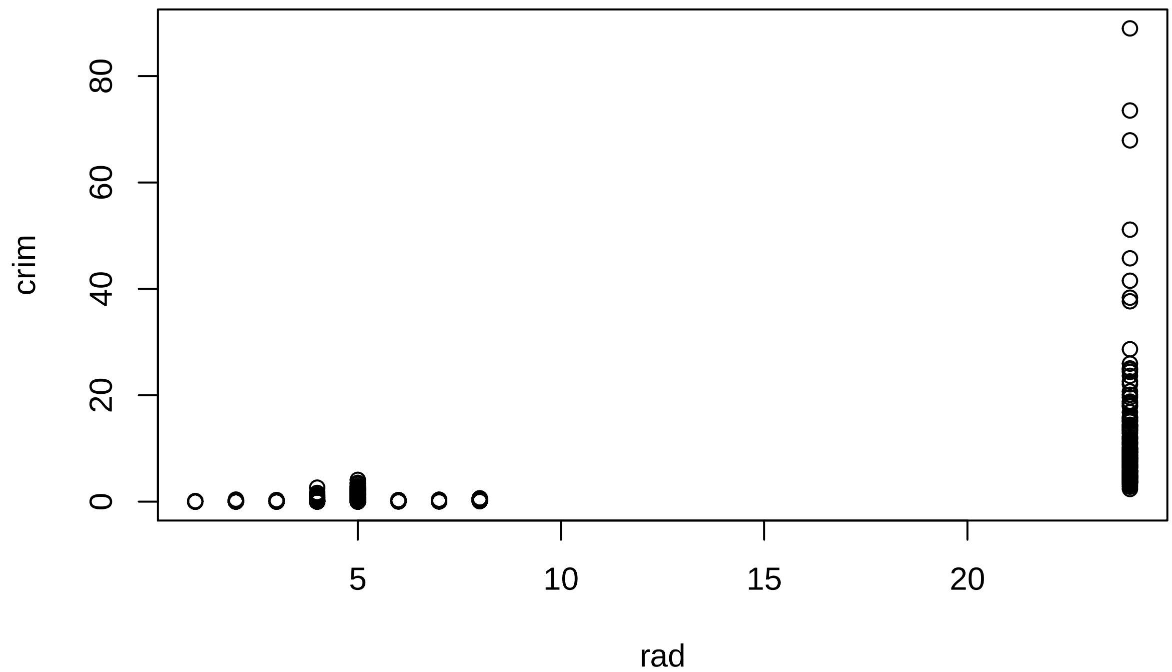
```r
plot(dis,crim)
```




```r
discoef <- coef(summary(lm(crim~dis, data=Boston)))

summary(lm(crim~rad, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ rad, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.164  -1.381  -0.141   0.660  76.433
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.28716    0.44348  -5.157 3.61e-07 ***
## rad          0.61791    0.03433  17.998  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.718 on 504 degrees of freedom
## Multiple R-squared:  0.3913, Adjusted R-squared:   0.39
## F-statistic: 323.9 on 1 and 504 DF,  p-value: < 2.2e-16
```

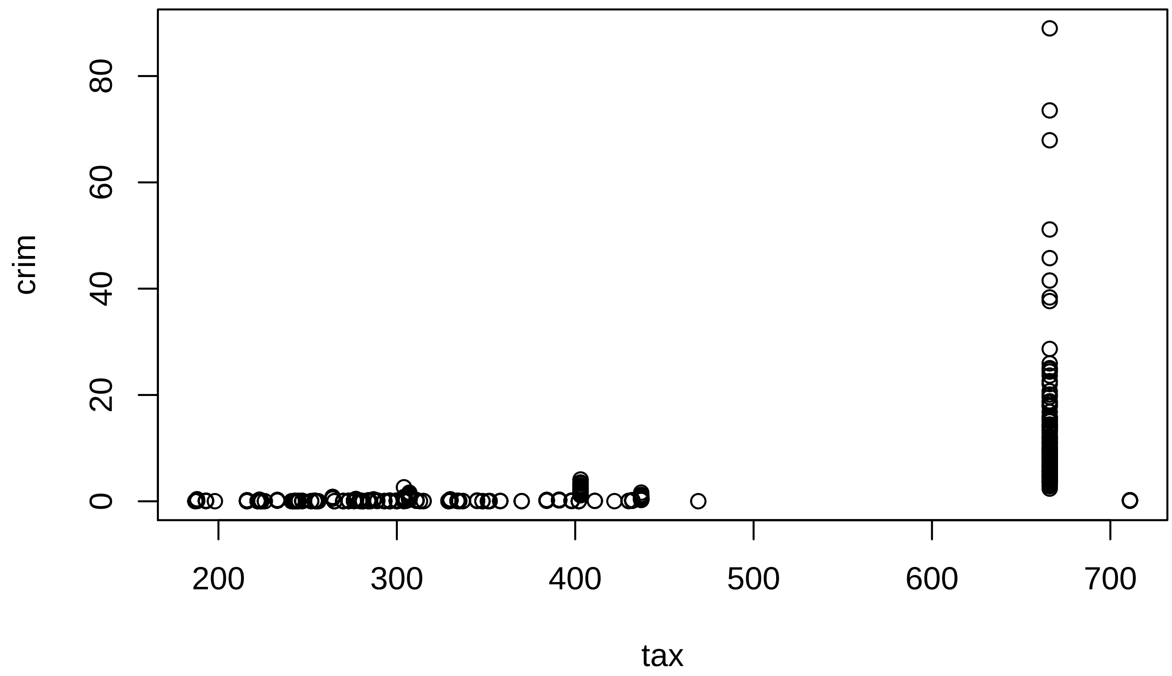
```r
plot(rad,crim)
```




```r
radcoef <- coef(summary(lm(crim~rad, data=Boston)))

summary(lm(crim~tax, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ tax, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.513  -2.738  -0.194   1.065  77.696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.528369   0.815809  -10.45   <2e-16 ***
## tax          0.029742   0.001847   16.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.997 on 504 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.3383
## F-statistic: 259.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

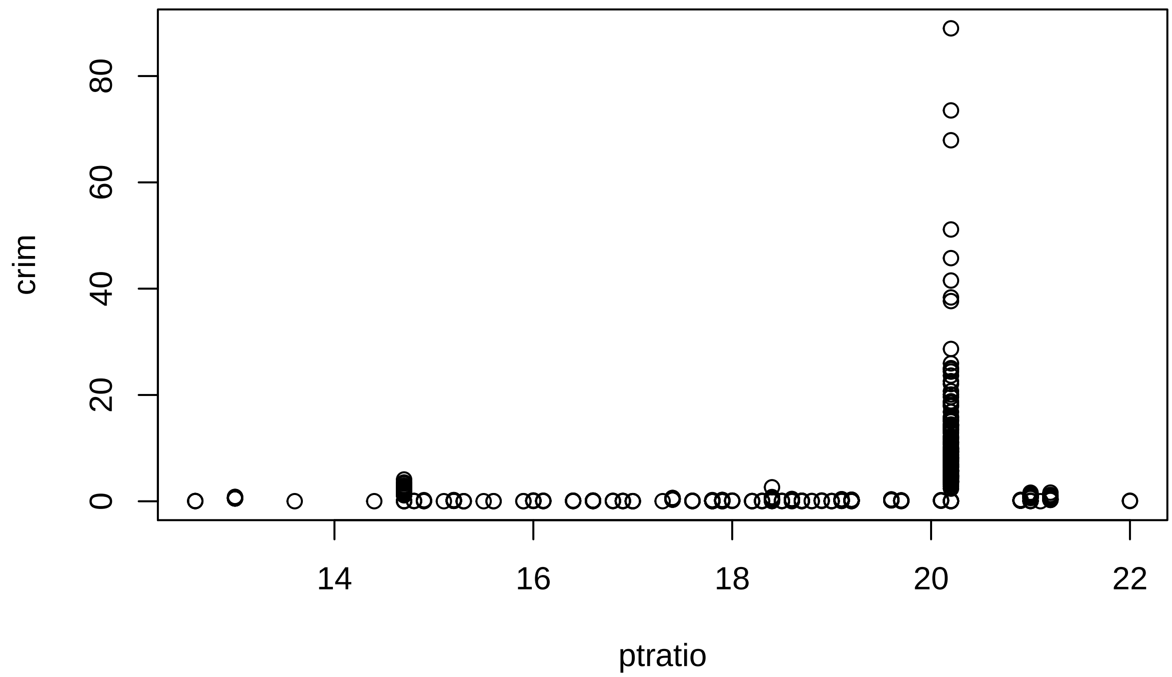
```r
plot(tax,crim)
```




```r
taxcoef <- coef(summary(lm(crim~tax, data=Boston)))

summary(lm(crim~ptratio, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ ptratio, data = Boston)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -7.654 -3.985 -1.912  1.825 83.353
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.6469     3.1473  -5.607 3.40e-08 ***
## ptratio       1.1520     0.1694   6.801 2.94e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.24 on 504 degrees of freedom
## Multiple R-squared:  0.08407,    Adjusted R-squared:  0.08225
## F-statistic: 46.26 on 1 and 504 DF,  p-value: 2.943e-11
```

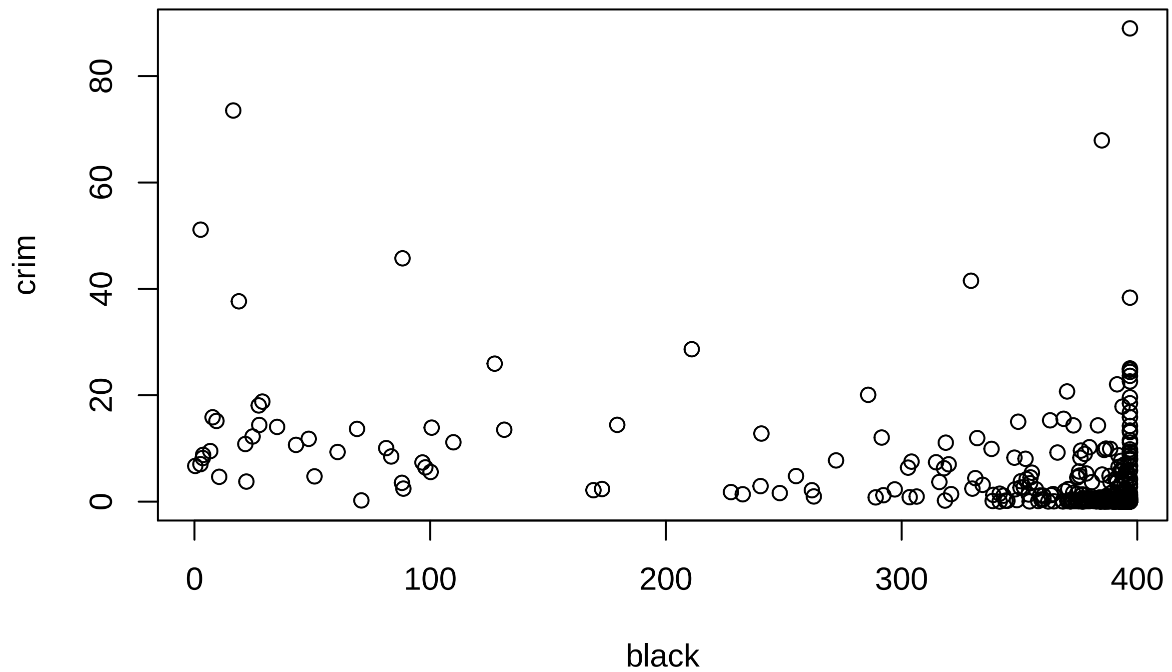
```r
plot(ptratio,crim)
```




```r
ptratiocoef <- coef(summary(lm(crim~ptratio, data=Boston)))

summary(lm(crim~black, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ black, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.756  -2.299  -2.095  -1.296  86.822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.553529   1.425903  11.609   <2e-16 ***
## black       -0.036280   0.003873  -9.367   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.946 on 504 degrees of freedom
## Multiple R-squared:  0.1483, Adjusted R-squared:  0.1466
## F-statistic: 87.74 on 1 and 504 DF,  p-value: < 2.2e-16
```

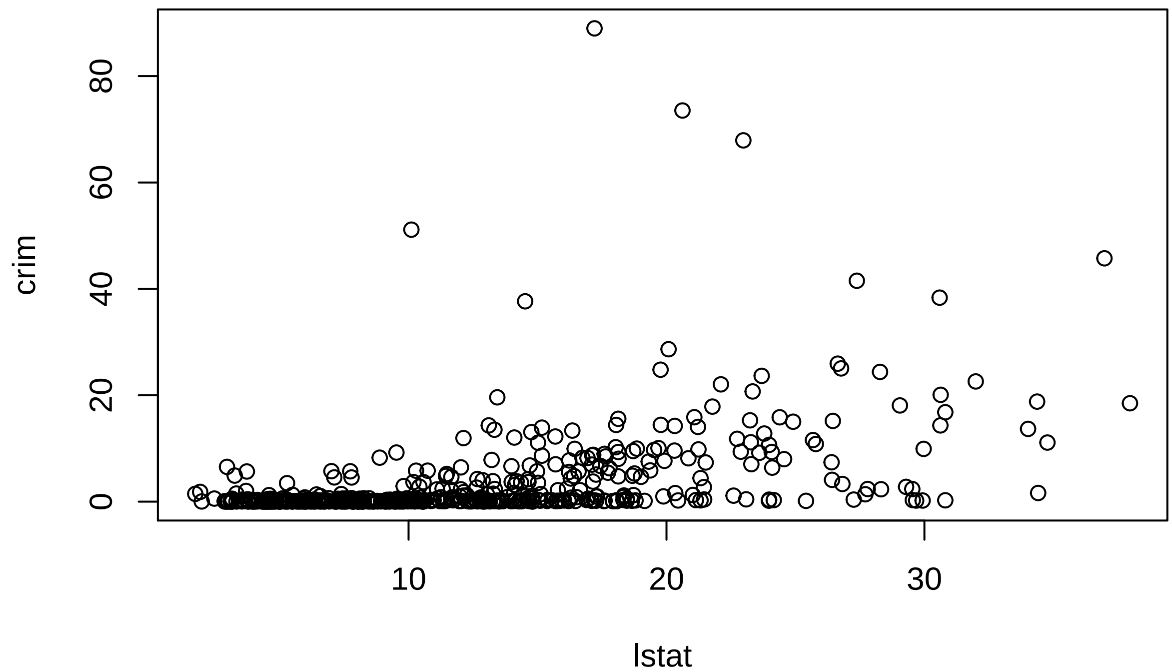
```r
plot(black,crim)
```




```r
blackcoef <- coef(summary(lm(crim~black, data=Boston)))

summary(lm(crim~lstat, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ lstat, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.925  -2.822  -0.664   1.079  82.862
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.33054    0.69376  -4.801 2.09e-06 ***
## lstat        0.54880    0.04776  11.491  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.664 on 504 degrees of freedom
## Multiple R-squared:  0.2076, Adjusted R-squared:  0.206
## F-statistic:   132 on 1 and 504 DF,  p-value: < 2.2e-16
```

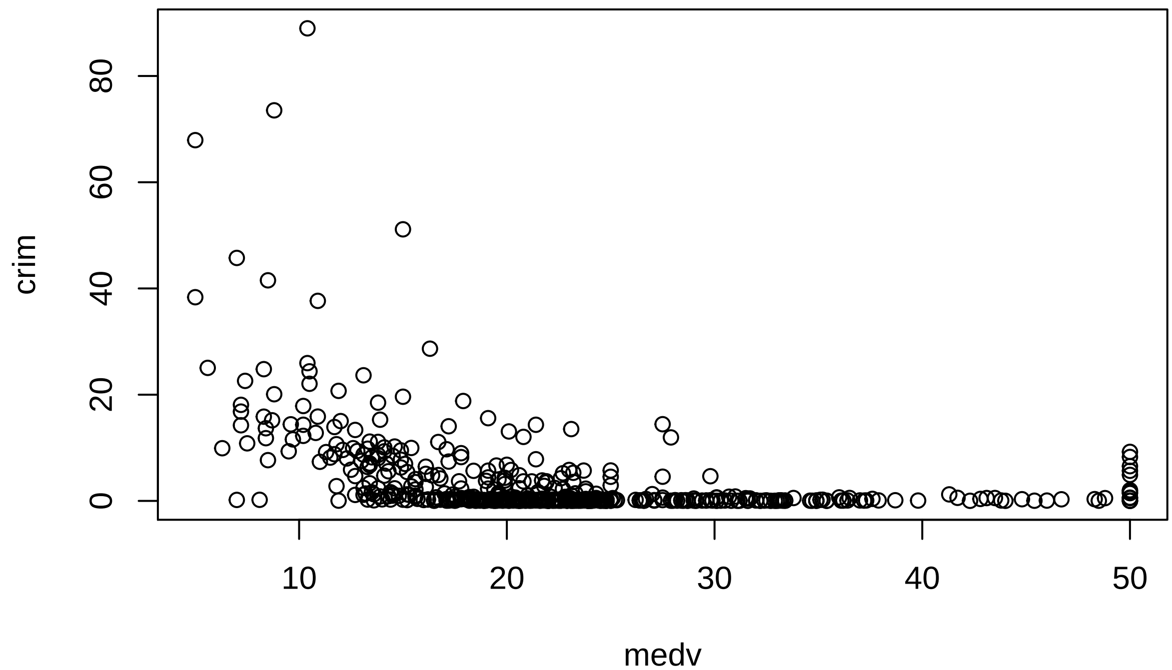
```r
plot(lstat,crim)
```




```
lstatcoef <- coef(summary(lm(crim~lstat, data=Boston)))

summary(lm(crim~medv, data=Boston))
```

```
##
## Call:
## lm(formula = crim ~ medv, data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.071 -4.022 -2.343  1.298 80.957
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.79654    0.93419   12.63   <2e-16 ***
## medv        -0.36316    0.03839   -9.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.934 on 504 degrees of freedom
## Multiple R-squared:  0.1508, Adjusted R-squared:  0.1491
## F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(medv,crim)
```




```r
medvcoef <- coef(summary(lm(crim~medv, data=Boston)))
```

According to the p-values of each individual predictor, all but "chas" is a statistically significant predictor of crime rate when simple linear regressions for each predictor is done. "Chas," however, is a qualitative predictor, so the 0's and 1's treated as numerics can be misleading. The plot for crim vs chas shows an association between being bound by the Charles River and crime rates. The plots for each predictor vs. crime rates show some association with crim with varying degrees.

b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis H0 : B_j = 0?

```r
summary(lm(crim~., dat=Boston))
```

```
##
## Call:
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.924 -2.120 -0.353  1.019 75.051
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *
## zn            0.044855   0.018734   2.394 0.017025 *
## indus        -0.063855   0.083407  -0.766 0.444294
## chas         -0.749134   1.180147  -0.635 0.525867
## nox         -10.313535   5.275536  -1.955 0.051152 .
## rm            0.430131   0.612830   0.702 0.483089
## age           0.001452   0.017925   0.081 0.935488
## dis          -0.987176   0.281817  -3.503 0.000502 ***
## rad           0.588209   0.088049   6.680 6.46e-11 ***
## tax          -0.003780   0.005156  -0.733 0.463793
## ptratio      -0.271081   0.186450  -1.454 0.146611
## black        -0.007538   0.003673  -2.052 0.040702 *
## lstat         0.126211   0.075725   1.667 0.096208 .
## medv         -0.198887   0.060516  -3.287 0.001087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.439 on 492 degrees of freedom
## Multiple R-squared:  0.454,  Adjusted R-squared:  0.4396
## F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

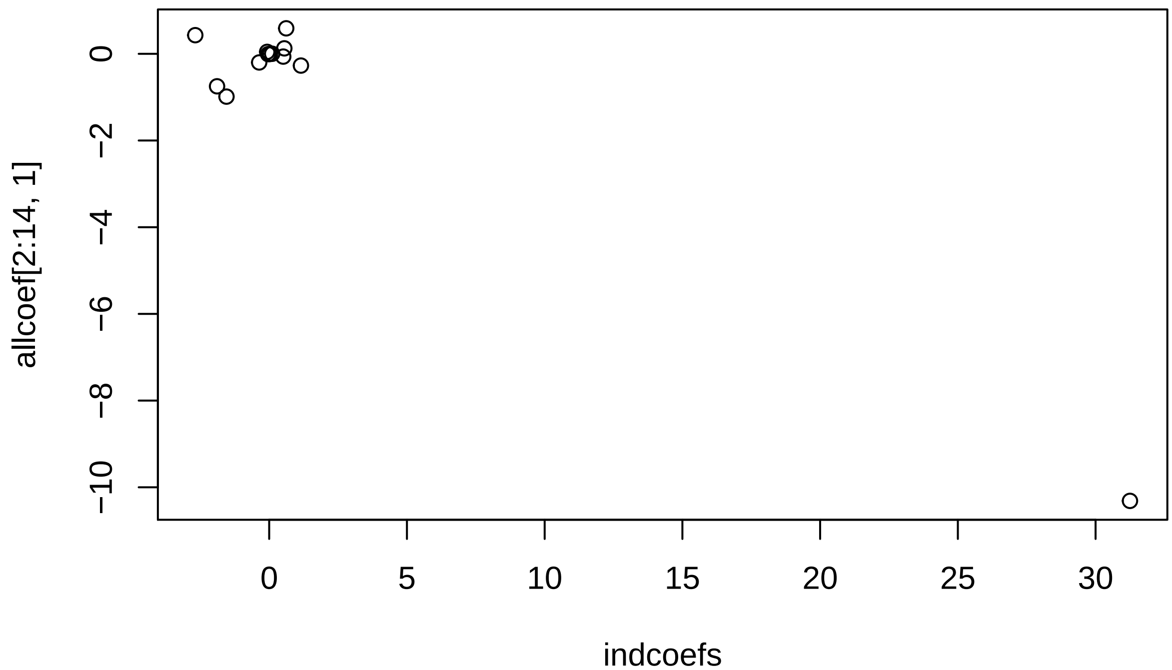
If we run a multiple regression with all the predictors, it seems that the only significant predictors are zn, indus, dis, rad, black, and medv. We do need to be cautions, however, as some of these predictors might have an interaction effect that forces us to keep the "insignificant" predictors.

c) How do your results from (a) compare to your results from (b)? Create a plot displaying the univariate regression coefficients from (a) on the x-axis, and the multiple regression coefficients from (b) on the y-axis. That is, each predictor is displayed as a single point in the plot. Its coefficient in a simple linear regression model is shown on the x-axis, and its coefficient estimate in the multiple linear regression model is shown on the y-axis.

The results from b) were quite different from a) when all the predictors were factored into the model.

To plot all the regression coefficients, I had to store them into two separate objects first.

```
allcoef <- coef(summary(lm(crim~., dat=Boston)))
indcoefs <- c(zncoef[2,1], induscoef[2,1], chascoef[2,1], noxcoef[2,1], rmcoef[2,1],agecoef[2,1],discoe
plot(indcoefs, allcoef[2:14,1])
```



Alternatively, I could also build 'indcoefs' by iteration:

```
indcoefs <- numeric(ncol(Boston)-1)
for(i in 2:ncol(Boston)-1) {
  coeffx <- coef(summary(lm(crim~Boston[,i+1])))
  indcoefs[i] <- coeffx[2,1]
}
```

d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor X, fit a model of the form Y = B_0 + B_1X + B2X^2 + B3X^3 + E.

```
summary(lm(crim~poly(zn,3), dat=Boston))
summary(lm(crim~poly(indus,3), dat=Boston))
summary(lm(crim~poly(chas,3), dat=Boston))
summary(lm(crim~poly(nox,3), dat=Boston))
summary(lm(crim~poly(rm,3), dat=Boston))
summary(lm(crim~poly(age,3), dat=Boston))
summary(lm(crim~poly(dis,3), dat=Boston))
summary(lm(crim~poly(rad,3), dat=Boston))
summary(lm(crim~poly(tax,3), dat=Boston))
```

```r
summary(lm(crim~poly(ptratio,3), dat=Boston))
summary(lm(crim~poly(black,3), dat=Boston))
summary(lm(crim~poly(lstat,3), dat=Boston))
summary(lm(crim~poly(medv,3), dat=Boston))

#OR we could build the models by creating a function and inputting each predictor:

polysum <- function(x) {
  summary(lm(crim~poly(x,3)))
}
```

There does seem to be a non-linear association between predictors and response. When we added up to the third powers of each predictor, the R^2 values increased.

BRIDGE PROBLEM (data source: http://www.stat.ufl.edu/~winner/data/bridge_risk.dat)

(data description: http://www.stat.ufl.edu/~winner/data/bridge_risk.txt)

a) What is the most impactful factor on the risk of bridges?

I first downloaded the data set as a text file and gave them corresponding column names. I also deleted the first column, since it simply identified the bridges, much like how the Boston data set identified the suburbs.

```r
bridges <- read.table('bridge_risk.dat.txt')
colnames(bridges) <- c('Bridge ID','SSR', 'FRR', 'SUR', 'ERR', 'Risk_Score')
bridges <- bridges[,-1]
summary(lm(Risk_Score~.,dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ ., data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.118  -6.867  -1.119   6.477  21.467
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.571      2.530   5.364 1.33e-06 ***
## SSR           15.500      1.186  13.074  < 2e-16 ***
## FRR            4.017      1.193   3.369  0.00131 **
## SUR            8.654      1.184   7.310 6.78e-10 ***
## ERR            1.602      1.117   1.435  0.15647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.203 on 61 degrees of freedom
## Multiple R-squared:  0.8797, Adjusted R-squared:  0.8718
## F-statistic: 111.5 on 4 and 61 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Risk_Score~SSR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.124 -10.395  -2.713   7.561  37.876
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.124      2.819   10.33 2.88e-15 ***
## SSR           19.726      1.596   12.36  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.07 on 64 degrees of freedom
## Multiple R-squared:  0.7048, Adjusted R-squared:  0.7002
## F-statistic: 152.8 on 1 and 64 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Risk_Score~FRR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ FRR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.719 -12.878   1.122  14.678  51.963
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   40.037      3.851  10.395 2.24e-15 ***
## FRR           13.682      2.359   5.801 2.21e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.97 on 64 degrees of freedom
## Multiple R-squared:  0.3446, Adjusted R-squared:  0.3344
## F-statistic: 33.65 on 1 and 64 DF,  p-value: 2.21e-07
```

```r
summary(lm(Risk_Score~SUR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SUR, data = bridges)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -35.63 -13.83  -1.78  18.00  36.37
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   31.227      4.960   6.296 3.14e-08 ***
## SUR           14.702      2.461   5.973 1.13e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.75 on 64 degrees of freedom
## Multiple R-squared:  0.3579, Adjusted R-squared:  0.3479
## F-statistic: 35.68 on 1 and 64 DF,  p-value: 1.126e-07
```

```
summary(lm(Risk_Score~ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ ERR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.815 -13.223  -1.404  17.733  44.733
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   51.267      4.357  11.766   <2e-16 ***
## ERR            5.274      3.002   1.757   0.0838 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.3 on 64 degrees of freedom
## Multiple R-squared:  0.04599,    Adjusted R-squared:  0.03109
## F-statistic: 3.086 on 1 and 64 DF,  p-value: 0.08378
```

It seems like the most impactful factor is SUR, since its p-value when all predictors are factored into the model is the smallest. Its p-value when the predictors are factored individually is also smallest.

b) Are all the predictors important to model the risk of bridges?

I have to be wary of the omnibus p-values and F-statistics.

```
summary(lm(Risk_Score~SSR+FRR+SSR*FRR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + SSR * FRR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.528  -7.790  -1.520   5.174  37.472
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  26.5275     3.4595   7.668 1.49e-10 ***
## SSR          16.2267     2.3662   6.858 3.79e-09 ***
## FRR           5.0399     2.9015   1.737   0.0874 .  
## SSR:FRR       0.6132     1.4316   0.428   0.6699    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.91 on 62 degrees of freedom
## Multiple R-squared:  0.7594, Adjusted R-squared:  0.7478 
## F-statistic: 65.24 on 3 and 62 DF,  p-value: < 2.2e-16
```

```
summary(lm(Risk_Score~SSR+SUR+SSR*SUR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + SUR + SSR * SUR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -17.653  -7.944  -0.509   6.924  25.521 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)    8.479      3.806   2.228   0.0295 *  
## SSR           23.460      2.743   8.553 4.37e-12 ***
## SUR           13.733      1.947   7.054 1.73e-09 ***
## SSR:SUR       -3.053      1.215  -2.512   0.0146 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.591 on 62 degrees of freedom
## Multiple R-squared:  0.8671, Adjusted R-squared:  0.8607 
## F-statistic: 134.9 on 3 and 62 DF,  p-value: < 2.2e-16
```

```
summary(lm(Risk_Score~SSR+ERR+SSR*ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + ERR + SSR * ERR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -24.383  -9.741  -1.324   6.687  39.902 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  25.0976     3.9879   6.293 3.53e-08 ***
## SSR          20.2584     2.3266   8.707 2.37e-12 ***
## ERR           4.2857     2.7579   1.554    0.125    
## SSR:ERR      -0.6967     1.4083  -0.495    0.623    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 13.86 on 62 degrees of freedom
## Multiple R-squared:  0.7226, Adjusted R-squared:  0.7092
## F-statistic: 53.83 on 3 and 62 DF,  p-value: < 2.2e-16
```

```
summary(lm(Risk_Score~FRR+SUR+FRR*SUR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ FRR + SUR + FRR * SUR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.620 -11.226   4.377   9.923  44.913
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.6314     5.8848   4.356 5.07e-05 ***
## FRR           9.2186     4.8911   1.885  0.06415 .
## SUR          10.7280     3.1190   3.440  0.00105 **
## FRR:SUR       0.5238     2.1515   0.243  0.80846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.97 on 62 degrees of freedom
## Multiple R-squared:  0.5336, Adjusted R-squared:  0.5111
## F-statistic: 23.65 on 3 and 62 DF,  p-value: 2.53e-10
```

```
summary(lm(Risk_Score~FRR+ERR+FRR*ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ FRR + ERR + FRR * ERR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.890 -12.690   0.874  11.828  53.912
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.0879     5.4484   6.807 4.63e-09 ***
## FRR          12.4960     3.4000   3.675 0.000498 ***
## ERR           3.2748     3.8310   0.855 0.395943
## FRR:ERR       0.8152     2.1820   0.374 0.709981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.76 on 62 degrees of freedom
## Multiple R-squared:  0.3775, Adjusted R-squared:  0.3473
## F-statistic: 12.53 on 3 and 62 DF,  p-value: 1.666e-06
```

```r
summary(lm(Risk_Score~SUR+ERR+SUR*ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SUR + ERR + SUR * ERR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.756 -13.641  -0.941  17.264  37.063
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.9200     7.2194   4.421 4.03e-05 ***
## SUR          13.0083     3.6155   3.598 0.000637 ***
## ERR          -0.2081     6.1514  -0.034 0.973120
## SUR:ERR       1.2405     2.6509   0.468 0.641447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.9 on 62 degrees of freedom
## Multiple R-squared:  0.3694, Adjusted R-squared:  0.3389
## F-statistic: 12.11 on 3 and 62 DF,  p-value: 2.455e-06
```

```r
summary(lm(Risk_Score~SSR+FRR+SUR+SSR*FRR*SUR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + SUR + SSR * FRR * SUR,
##     data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.108  -6.709  -1.545   6.856  15.906
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.952      4.470   0.660  0.51157
## SSR           25.389      3.459   7.341 7.84e-10 ***
## FRR            8.760      4.106   2.133  0.03713 *
## SUR           15.402      2.292   6.721 8.60e-09 ***
## SSR:FRR       -3.060      2.528  -1.210  0.23107
## SSR:SUR       -5.096      1.596  -3.193  0.00227 **
## FRR:SUR       -2.598      1.923  -1.351  0.18203
## SSR:FRR:SUR    1.634      1.082   1.510  0.13638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.727 on 58 degrees of freedom
## Multiple R-squared:  0.8971, Adjusted R-squared:  0.8847
## F-statistic: 72.24 on 7 and 58 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Risk_Score~SSR+FRR+SUR+SSR*FRR*SUR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + SUR + SSR * FRR * SUR,
##     data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.108  -6.709  -1.545   6.856  15.906
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.952      4.470   0.660  0.51157
## SSR           25.389      3.459   7.341 7.84e-10 ***
## FRR            8.760      4.106   2.133  0.03713 *
## SUR           15.402      2.292   6.721 8.60e-09 ***
## SSR:FRR       -3.060      2.528  -1.210  0.23107
## SSR:SUR       -5.096      1.596  -3.193  0.00227 **
## FRR:SUR       -2.598      1.923  -1.351  0.18203
## SSR:FRR:SUR    1.634      1.082   1.510  0.13638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.727 on 58 degrees of freedom
## Multiple R-squared:  0.8971, Adjusted R-squared:  0.8847
## F-statistic: 72.24 on 7 and 58 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Risk_Score~SSR+FRR+ERR+SSR*FRR*ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + ERR + SSR * FRR * ERR,
##     data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.515  -7.503  -1.409   5.191  41.229
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.7706     5.1938   3.807 0.000342 ***
## SSR          18.9065     3.5290   5.357 1.52e-06 ***
## FRR           7.6700     4.3055   1.781 0.080080 .
## ERR           5.8724     3.2608   1.801 0.076920 .
## SSR:FRR      -0.8780     2.1589  -0.407 0.685715
## SSR:ERR      -2.0175     2.2472  -0.898 0.373019
## FRR:ERR      -1.6023     2.8482  -0.563 0.575894
## SSR:FRR:ERR   0.8783     1.3562   0.648 0.519793
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 12.81 on 58 degrees of freedom
## Multiple R-squared:  0.7782, Adjusted R-squared:  0.7514
## F-statistic: 29.07 on 7 and 58 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Risk_Score~SSR+SUR+ERR+SSR*SUR*ERR, dat=bridges))
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + SUR + ERR + SSR * SUR * ERR,
##     data = bridges)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.7188  -7.9222  -0.3676   7.4546  21.7361
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.5368     6.7245   0.675    0.503
## SSR          25.9770     4.5714   5.683 4.52e-07 ***
## SUR          14.5059     3.3756   4.297 6.69e-05 ***
## ERR           3.8636     4.7840   0.808    0.423
## SSR:SUR      -4.0169     2.0072  -2.001    0.050 .
## SSR:ERR      -2.1653     3.1138  -0.695    0.490
## SUR:ERR      -0.7423     2.1655  -0.343    0.733
## SSR:SUR:ERR   0.7734     1.2852   0.602    0.550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.704 on 58 degrees of freedom
## Multiple R-squared:  0.8728, Adjusted R-squared:  0.8574
## F-statistic: 56.84 on 7 and 58 DF,  p-value: < 2.2e-16
```

```r
fullmod <- lm(Risk_Score~SSR+FRR+SUR+ERR+SSR*FRR*SUR*ERR, dat=bridges)
summary(fullmod)
```

```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + SUR + ERR + SSR * FRR *
##     SUR * ERR, data = bridges)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6541  -6.4865  -0.7067   5.8981  19.0123
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       8.9509     8.3820   1.068 0.290707
## SSR              22.7437     6.1241   3.714 0.000515 ***
## FRR              -1.9831     8.4776  -0.234 0.816005
## SUR              11.0122     4.1993   2.622 0.011541 *
## ERR              -4.0804     6.0868  -0.670 0.505705
## SSR:FRR           1.2780     4.8358   0.264 0.792645
## SSR:SUR          -3.2836     2.6406  -1.244 0.219471
```

```
## FRR:SUR           2.6090     3.7999   0.687 0.495514
## SSR:ERR           1.8244     4.2011   0.434 0.665963
## FRR:ERR           7.8230     5.3328   1.467 0.148647
## SUR:ERR           2.9833     2.6880   1.110 0.272361
## SSR:FRR:SUR      -0.5312     2.0185  -0.263 0.793506
## SSR:FRR:ERR      -2.4553     2.9706  -0.827 0.412429
## SSR:SUR:ERR      -1.2377     1.7943  -0.690 0.493498
## FRR:SUR:ERR      -3.5236     2.3011  -1.531 0.132001
## SSR:FRR:SUR:ERR   1.2103     1.2093   1.001 0.321743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.927 on 50 degrees of freedom
## Multiple R-squared:  0.9072, Adjusted R-squared:  0.8793
## F-statistic: 32.58 on 15 and 50 DF,  p-value: < 2.2e-16
```

Since ERR had the highest p-value when all the predictors were factored into the model, and since the p-values for any interaction involving ERR are not low enough, I think ERR can be safely removed from the model.

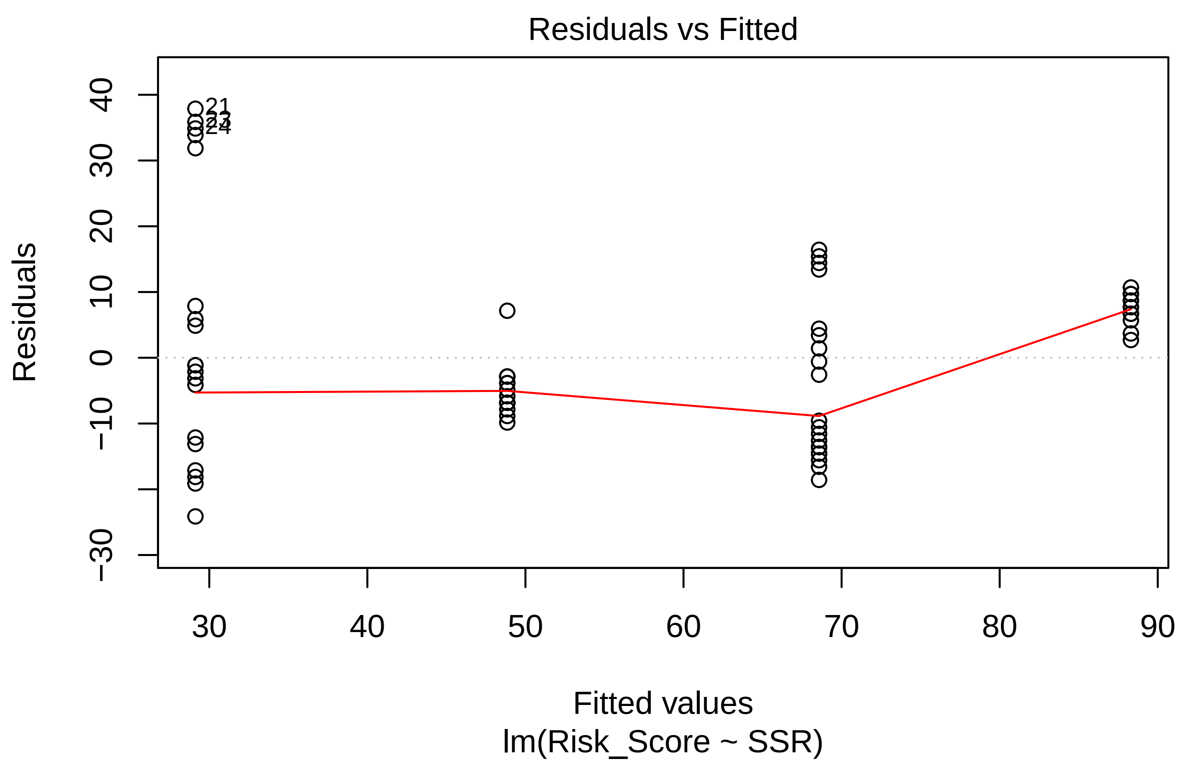
Also, none of the interactions are included in the model, since the interaction p-values aren't low enough. One interaction p-value for SSR:SUR was low enough, but that was only when SSR and SUR were the only predictors in the model comparison, so interaction effects when taking in all variables can still be ignored.

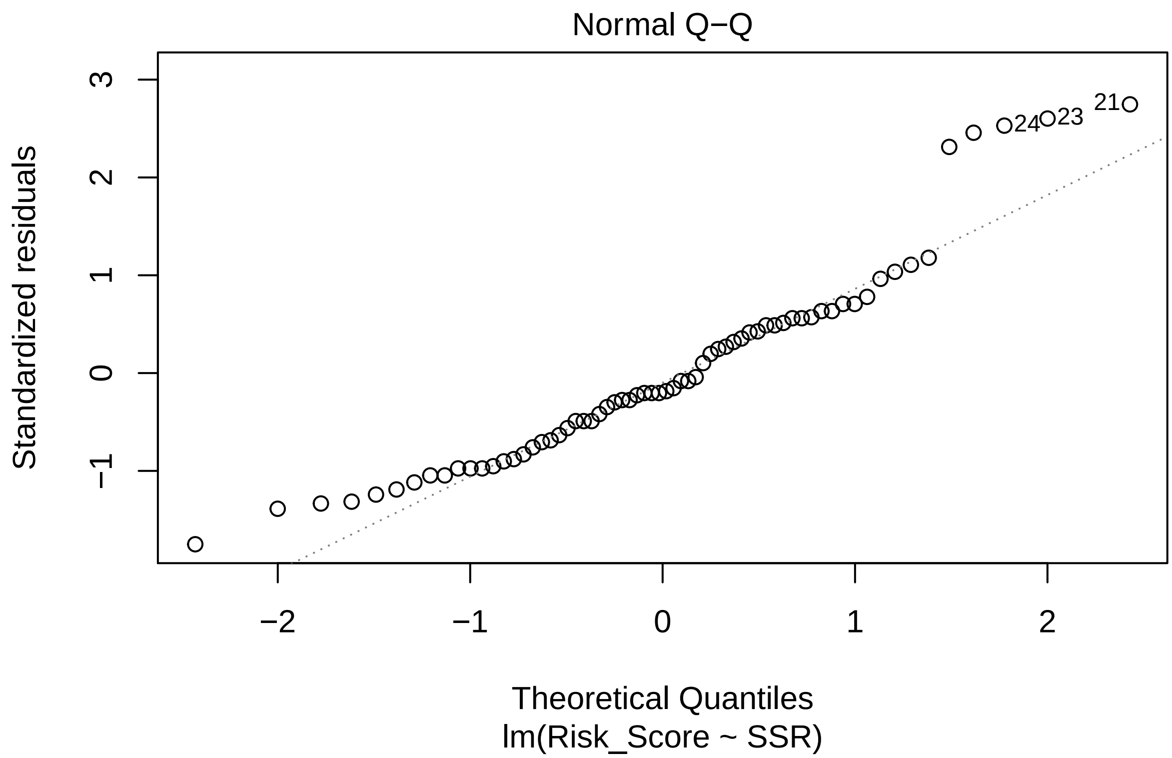
c) Which linear regression model provides the best fit to the data?

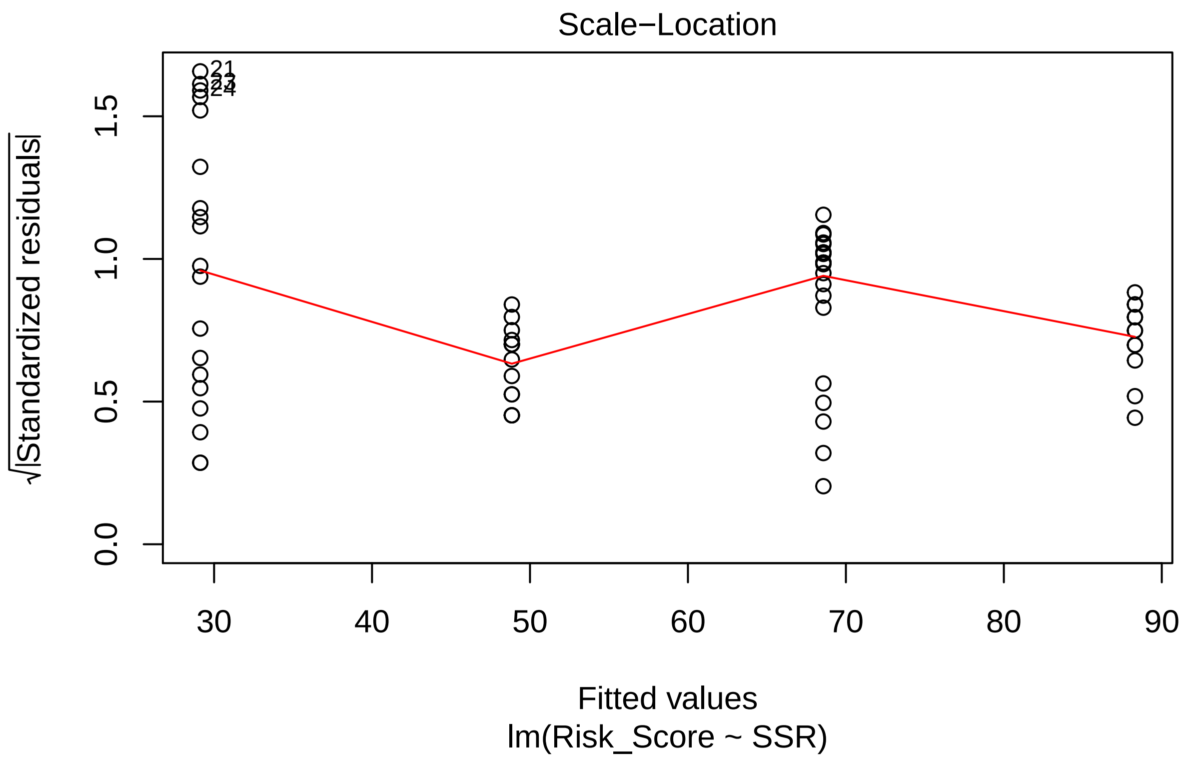
```r
mod3 <- lm(Risk_Score~SSR+FRR+SUR, dat=bridges)
summary(mod3)
```

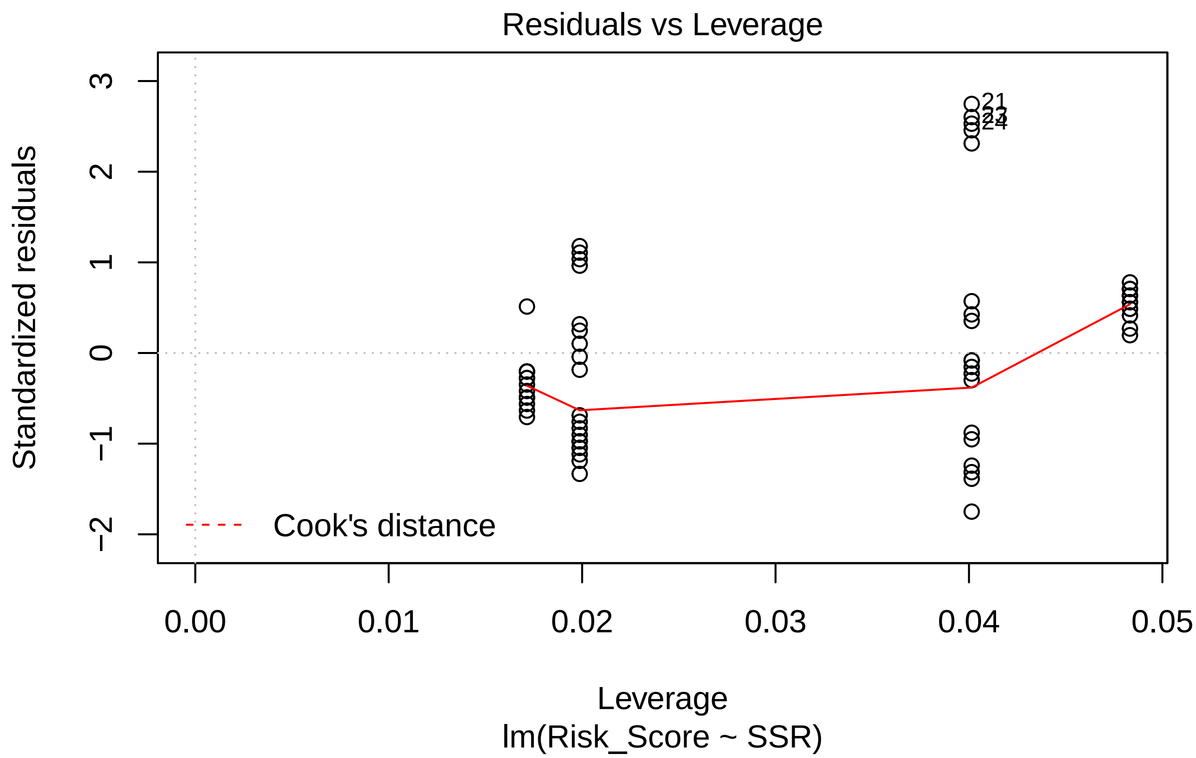
```
##
## Call:
## lm(formula = Risk_Score ~ SSR + FRR + SUR, data = bridges)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.652  -6.617  -1.198   6.334  22.534
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.581      2.451   5.949 1.36e-07 ***
## SSR           15.595      1.194  13.063  < 2e-16 ***
## FRR            3.979      1.202   3.309  0.00156 **
## SUR            8.962      1.174   7.633 1.72e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.281 on 62 degrees of freedom
## Multiple R-squared:  0.8756, Adjusted R-squared:  0.8696
## F-statistic: 145.5 on 3 and 62 DF,  p-value: < 2.2e-16
```

```r
plot(lm(Risk_Score~SSR, dat=bridges))
```

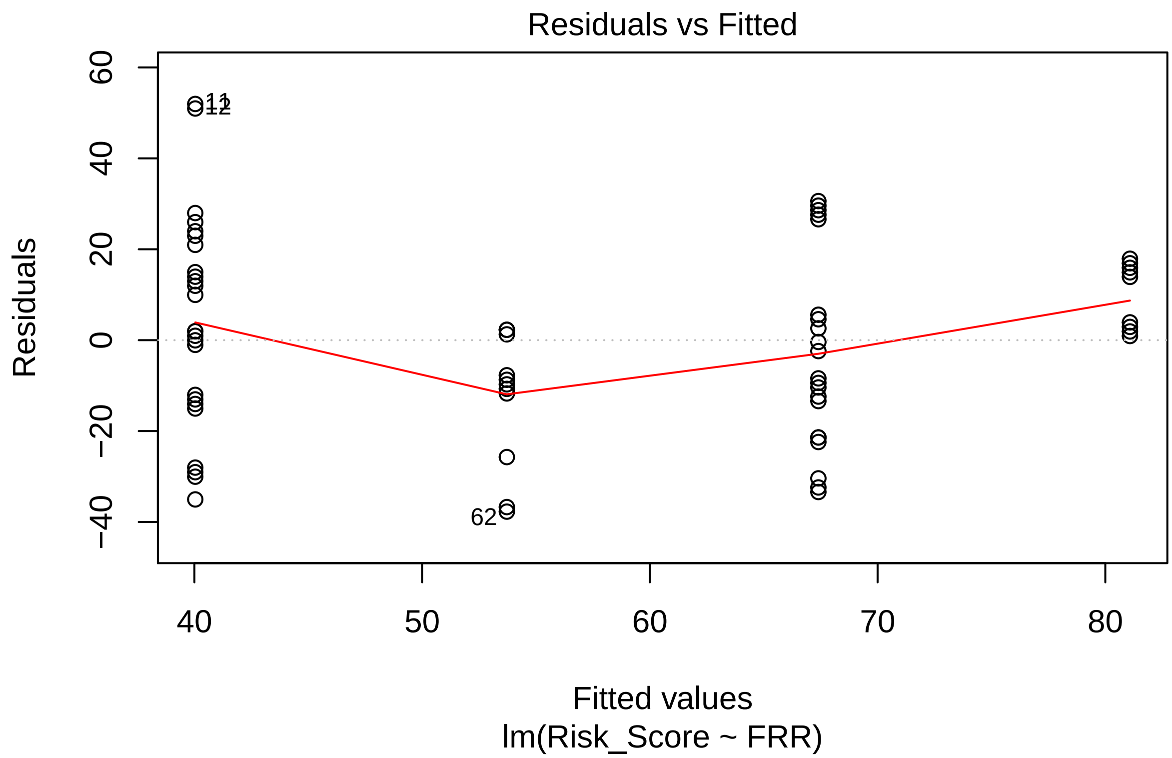



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Risk_Score ~ SSR)
3
2
1
0
−1
−2
−1
0
1
2
24
23
21

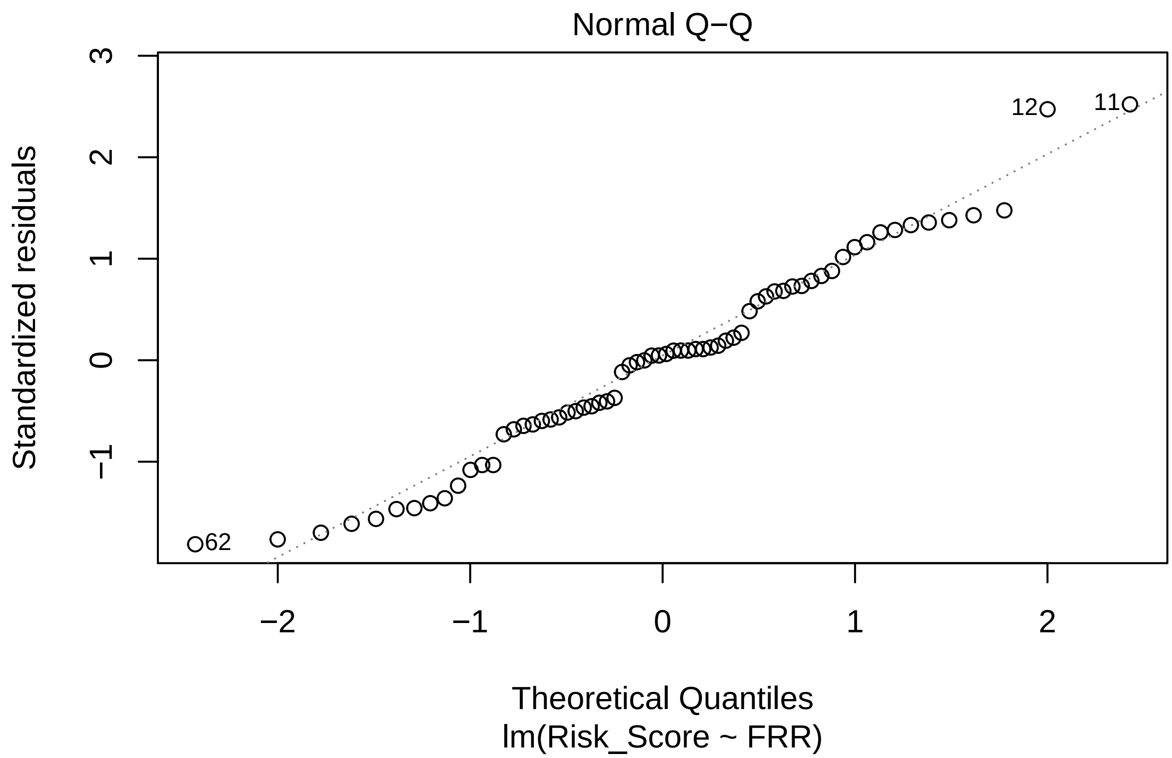

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
21
23
24
30
40
50
60
70
80
90
Fitted values
lm(Risk_Score ~ SSR)

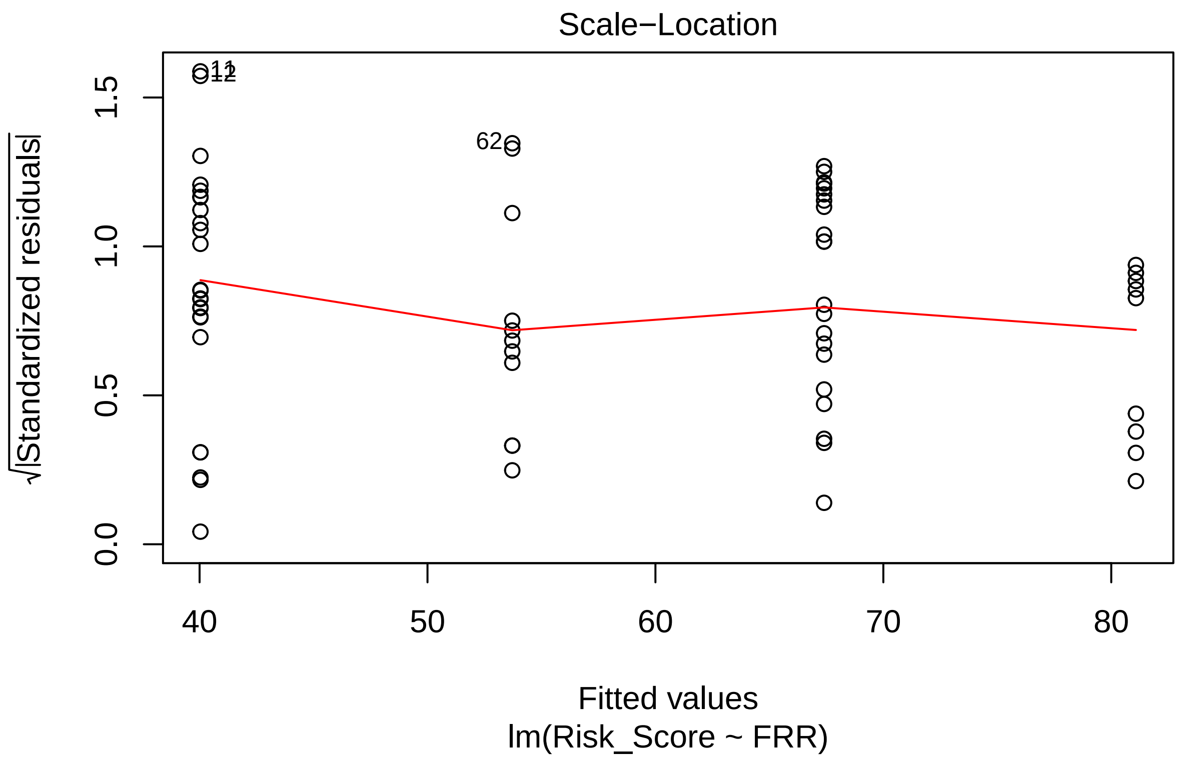


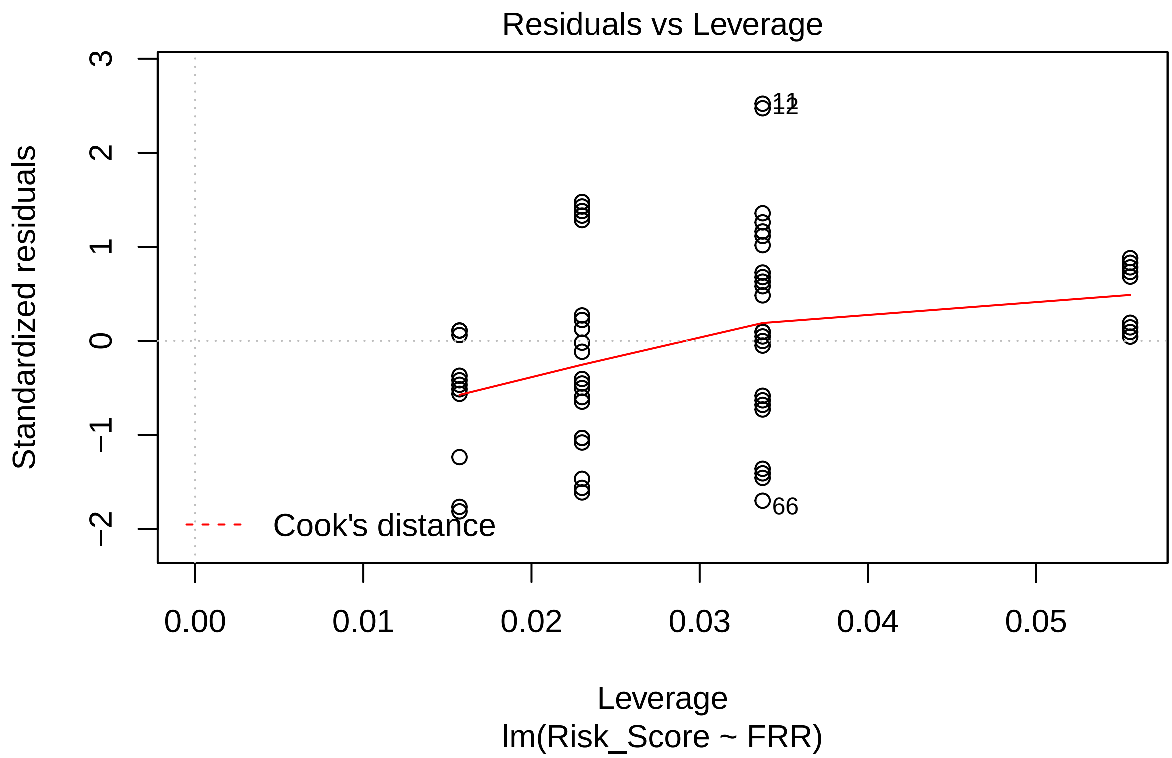

```
plot(lm(Risk_Score~FRR, dat=bridges))
```



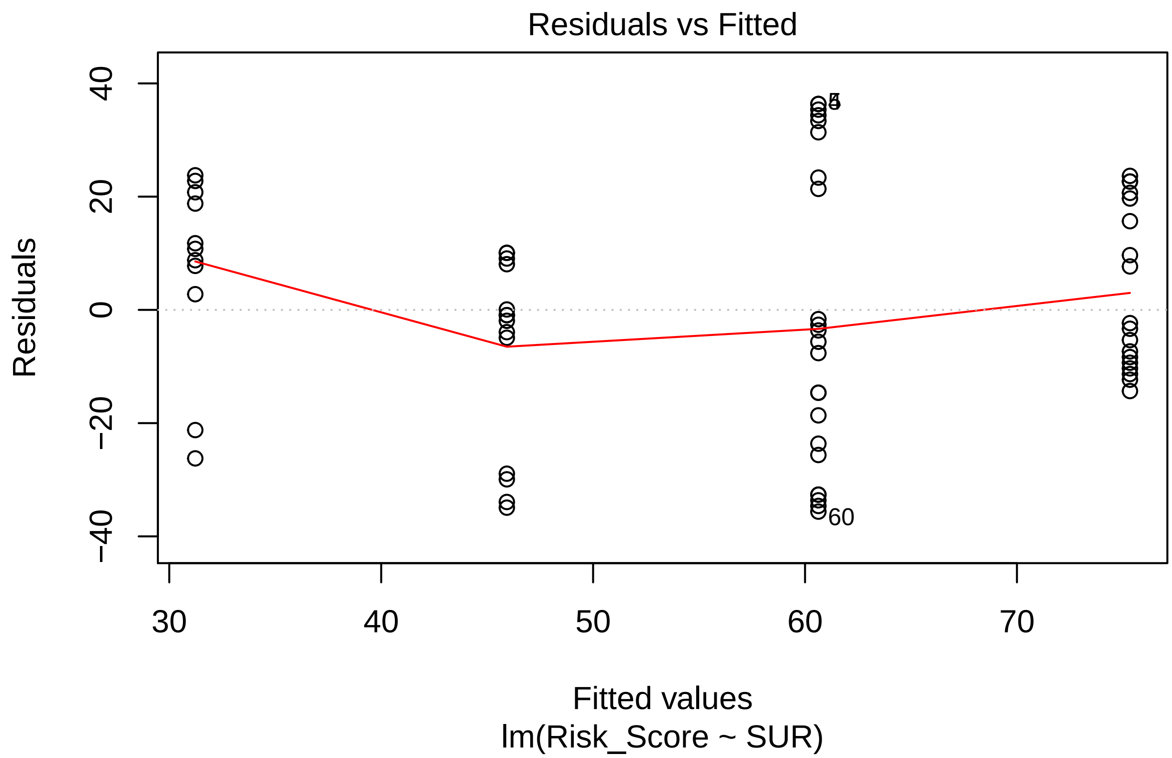
Residuals vs Fitted
Residuals
60
40
20
0
−20
−40
11
12
62
40
50
60
70
80
Fitted values
lm(Risk_Score ~ FRR)

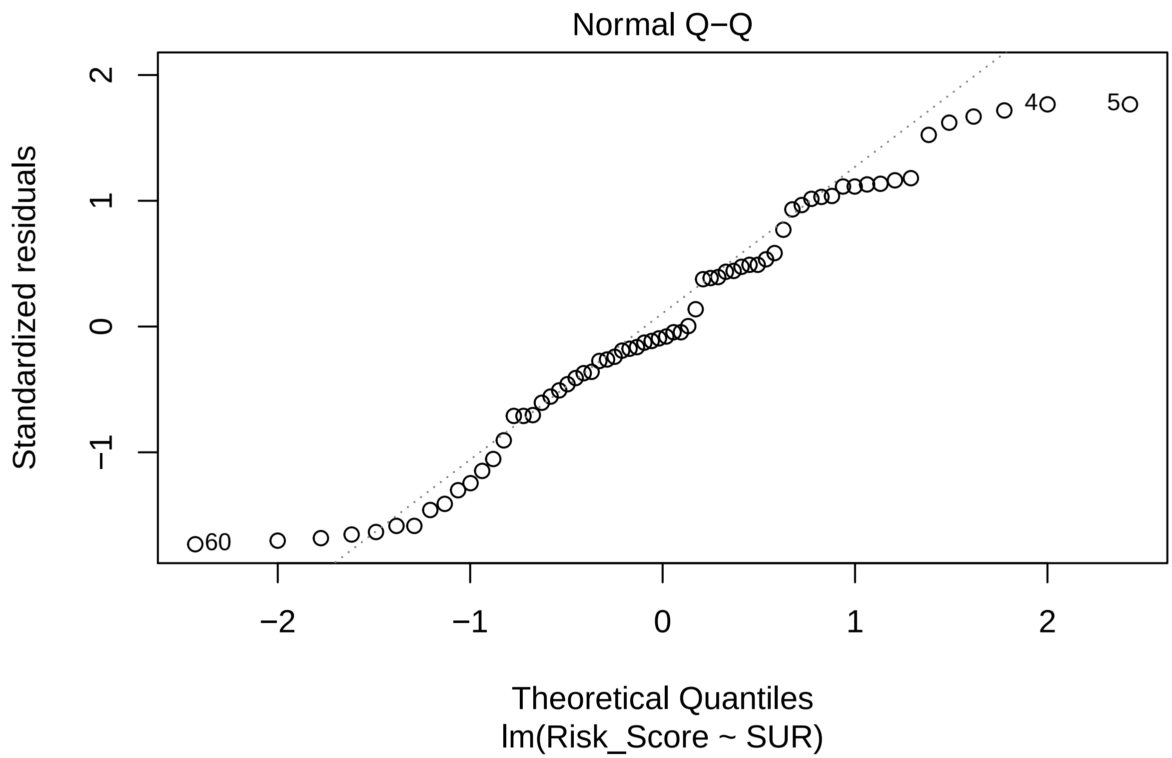

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Risk_Score ~ FRR)
3
2
1
0
−1
−2
−1
0
1
2
62
12
11


Scale−Location
11
12
62
1.5
1.0
0.5
0.0
√|Standardized residuals|
40
50
60
70
80
Fitted values
lm(Risk_Score ~ FRR)

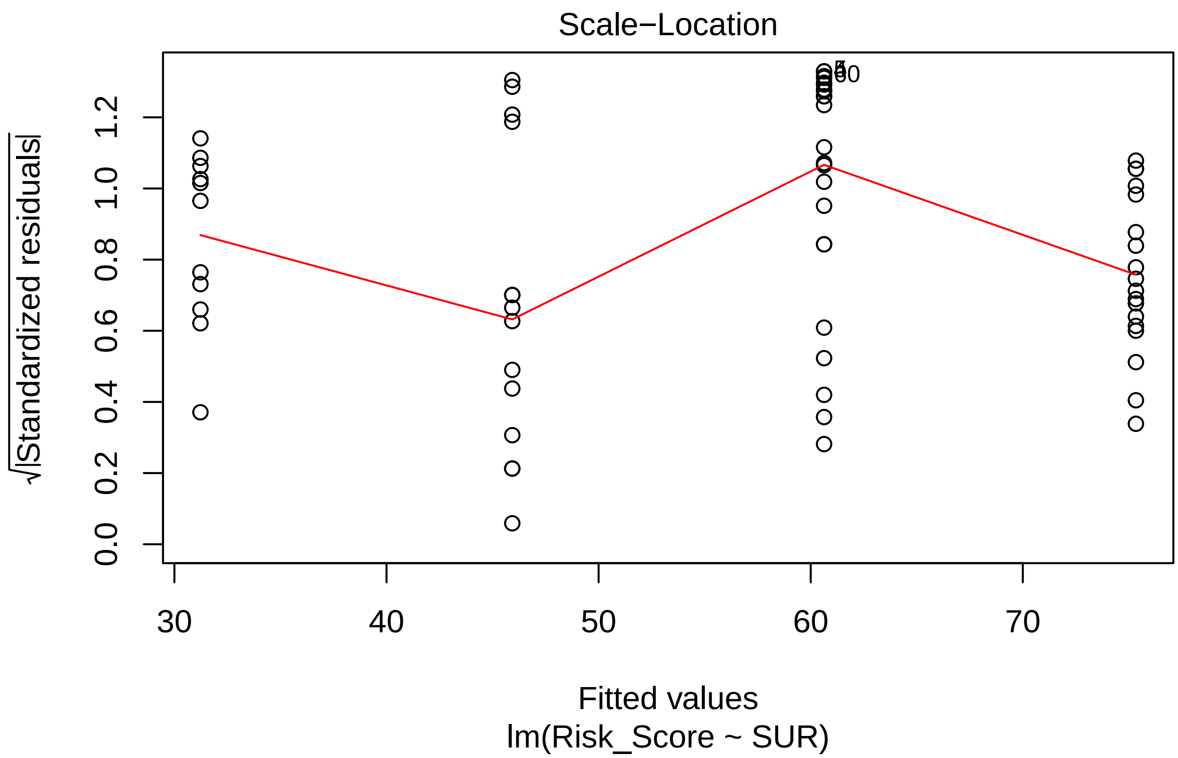


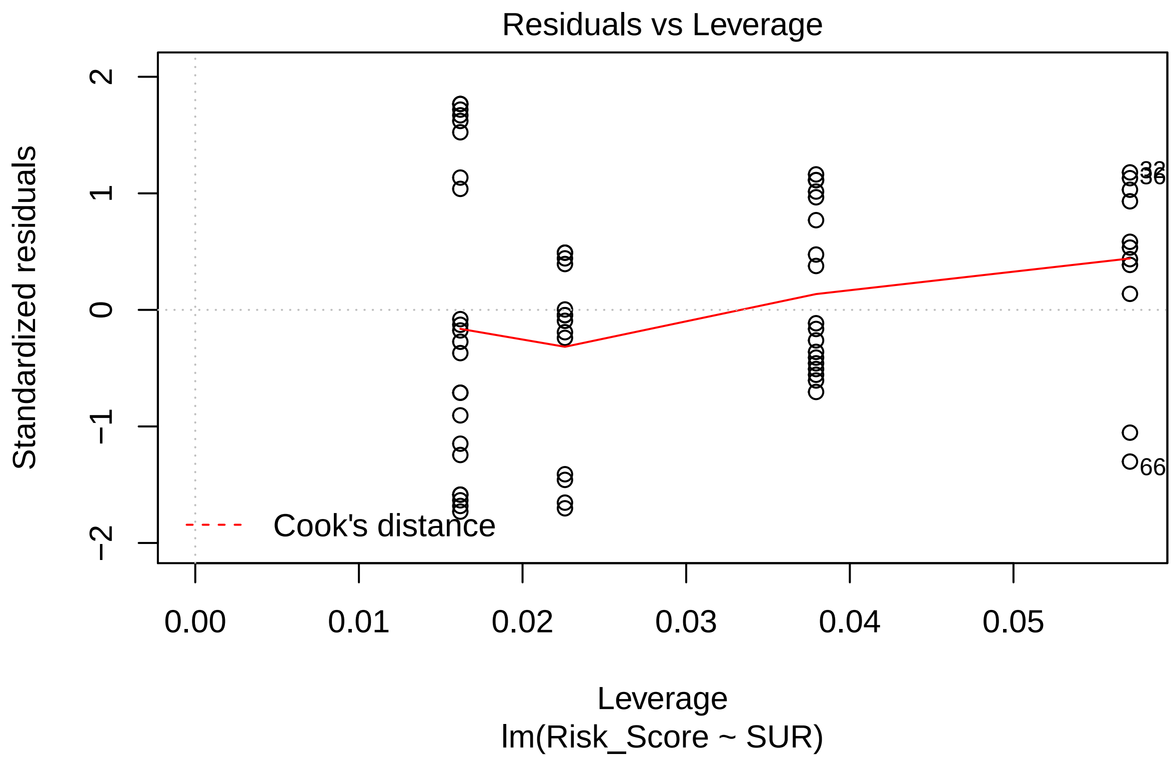

```
plot(lm(Risk_Score~SUR, dat=bridges))
```


Residuals vs Fitted
Residuals
40
20
0
−20
−40
30
40
50
60
70
Fitted values
lm(Risk_Score ~ SUR)
60



Normal Q−Q
Standardized residuals
2
1
0
−1
60
4
5
−2
−1
0
1
2
Theoretical Quantiles
lm(Risk_Score ~ SUR)


Scale−Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
30
40
50
60
70
Fitted values
lm(Risk_Score ~ SUR)



The model that best fits the data is the full model (fullmod), when all predictors are factored into the model and its interactions. Removing ERR and interactions, however, does not change the R^2 value very much, and including ERR and interactions might result in overfitting.

d) Which linear regression model provides the best predictive accuracy?

The most accurate prediction model will be the one with the lowest MSE.

Author's Note (edited 5 days since assignemt was turned in): cross-validation methods were not yet discussed in class, so what I did below was an unequal hold-out validation. There are 66 observations (N=66), and 40 of those were used to train models while the remaining 26 were used to test them.

Due to the loss of information when omitting 26 observations for training, and due to the high variability of the hold-out method of validation, the MSE's below are not optimal and may not be reproducible.

```
training_data = bridges[1:40,]
testing_data = bridges[-1:-40,]

model1 <- lm(Risk_Score~SSR+FRR+SUR+SSR*FRR*SUR, dat=training_data)
model2 <- lm(Risk_Score~SSR+FRR+SUR+SSR*FRR+SSR*SUR+FRR*SUR, dat=training_data)
model3 <- lm(Risk_Score~SSR+FRR+SUR+SSR*FRR+SSR*SUR, dat=training_data)
model4 <- lm(Risk_Score~SSR+FRR+SUR+SSR*FRR+FRR*SUR,  dat=training_data)
model5 <- lm(Risk_Score~SSR+FRR+SUR+SSR*SUR+FRR*SUR,  dat=training_data)
model6 <- lm(Risk_Score~SSR+FRR+SUR+SSR*FRR, dat=training_data)
```

```r
model7 <- lm(Risk_Score~SSR+FRR+SUR+SSR*SUR, dat=training_data)
model8 <- lm(Risk_Score~SSR+FRR+SUR+FRR*SUR, dat=training_data)
model9 <- lm(Risk_Score~SSR+FRR+SUR, dat=training_data)
model10 <- lm(Risk_Score~SSR+FRR+SSR*FRR, dat=training_data)
model11 <- lm(Risk_Score~SSR+SUR+SSR*SUR, dat=training_data)
model12 <- lm(Risk_Score~FRR+SUR+FRR*SUR, dat=training_data)
model13 <- lm(Risk_Score~SSR+FRR, dat=training_data)
model14 <- lm(Risk_Score~SSR+SUR, dat=training_data)
model15 <- lm(Risk_Score~SSR, dat=training_data)
model16 <- lm(Risk_Score~FRR, dat=training_data)
model17 <- lm(Risk_Score~SUR, dat=training_data)

y = testing_data$Risk_Score


y_hat1=predict(model1, testing_data[,-5])
y_hat2=predict(model2, testing_data[,-5])
y_hat3=predict(model3, testing_data[,-5])
y_hat4=predict(model4, testing_data[,-5])
y_hat5=predict(model5, testing_data[,-5])
y_hat6=predict(model6, testing_data[,-5])
y_hat7=predict(model7, testing_data[,-5])
y_hat8=predict(model8, testing_data[,-5])
y_hat9=predict(model9, testing_data[,-5])
y_hat10=predict(model10, testing_data[,-5])
y_hat11=predict(model11, testing_data[,-5])
y_hat12=predict(model12, testing_data[,-5])
y_hat13=predict(model13, testing_data[,-5])
y_hat14=predict(model14, testing_data[,-5])
y_hat15=predict(model15, testing_data[,-5])
y_hat16=predict(model16, testing_data[,-5])
y_hat17=predict(model17, testing_data[,-5])



MSE1=mean((y-y_hat1)^2)
MSE2=mean((y-y_hat2)^2)
MSE3=mean((y-y_hat3)^2)
MSE4=mean((y-y_hat4)^2)
MSE5=mean((y-y_hat5)^2)
MSE6=mean((y-y_hat6)^2)
MSE7=mean((y-y_hat7)^2)
MSE8=mean((y-y_hat8)^2)
MSE9=mean((y-y_hat9)^2)
MSE10=mean((y-y_hat10)^2)
MSE11=mean((y-y_hat11)^2)
MSE12=mean((y-y_hat12)^2)
MSE13=mean((y-y_hat13)^2)
MSE14=mean((y-y_hat14)^2)
MSE15=mean((y-y_hat15)^2)
MSE16=mean((y-y_hat16)^2)
MSE17=mean((y-y_hat17)^2)
```

```
MSE1
```

```
## [1] 191.727
```

```
MSE2
```

```
## [1] 195.2083
```

```
MSE3
```

```
## [1] 191.8597
```

```
MSE4
```

```
## [1] 267.1297
```

```
MSE5
```

```
## [1] 231.5051
```

```
MSE6
```

```
## [1] 278.3896
```

```
MSE7
```

```
## [1] 230.6406
```

```
MSE8
```

```
## [1] 195.1549
```

```
MSE9
```

```
## [1] 205.1535
```

```
MSE10
```

```
## [1] 683.3952
```

```
MSE11
```

```
## [1] 263.497
```

```
MSE12
```

```
## [1] 918.57
```

```
MSE13
```

```
## [1] 543.7472
```

```
MSE14
```

```
## [1] 199.574
```

```
MSE15
```

```
## [1] 586.533
```

```
MSE16
```

```
## [1] 1296.894
```

```
MSE17
```

```
## [1] 1180.92
```

The model with the lowest MSE is Model 1, the model with SSR+FRR+SUR+interactions among the 3. Although we decided on the model with SSR+FRR+SUR (model 9), we did not previously consider the interaction between SSR, FRR, and SUR.

However, the regression analysis in part b) suggested that the interaction among all 3 is insignificant, and that the only interaction pair that was significant was SSR:SUR. Nevertheless, the MSE for model 16 (FRR only) was very high, so there must be something about FRR that increases the MSE and decreases the accuracy when the 3 variables are combined in the model.

Model 14 (SSR+SUR) has a lower MSE than Model 3 (SSR+FRR+SUR), further suggesting that FRR by itself did not help with accuracy. Further analysis shows that SSR:SUR, the only interaction deemed significant enough in part b), results in a higher MSE than all three interacting.

The highest MSE by far is the model that accounts for all predictors and all interactions. This shows that the full model may fit the data best but are poor predictors of new data. That model is a classic example of overfitting.