A2Q1

Undergraduate Student

```
library(MASS)
```

(a)

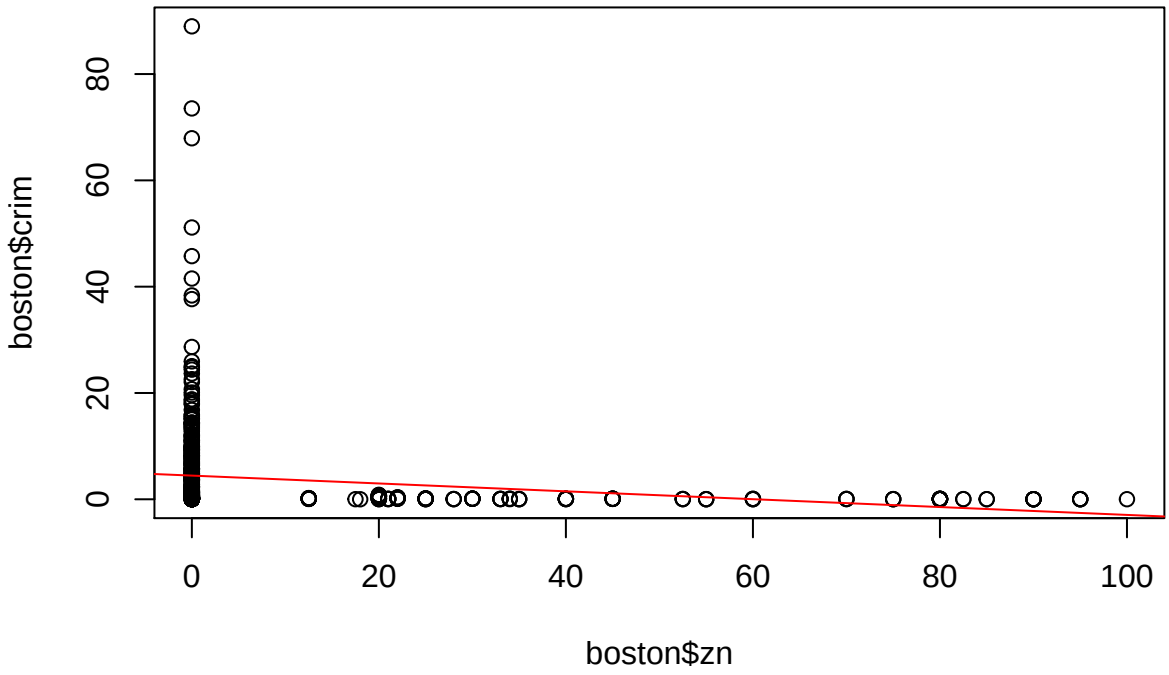
```
boston <- Boston
```

(i)

```
m1 <- lm(crim~zn, data = boston)
summary(m1)
```

```
##
## 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
```

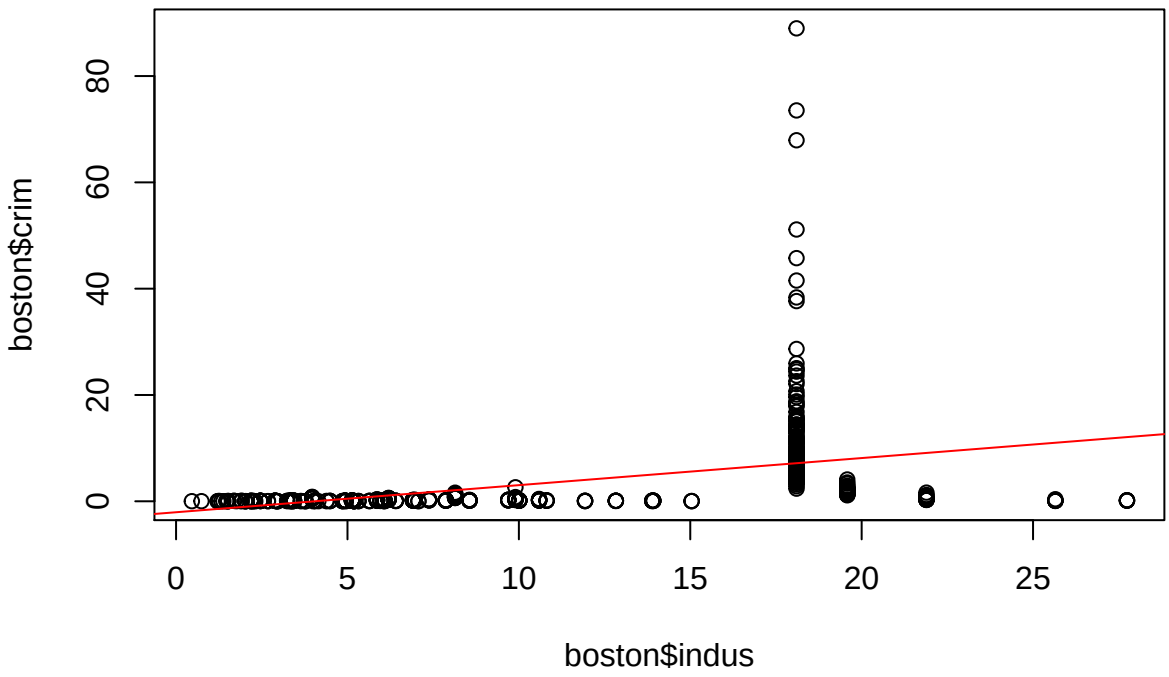
```
plot(boston$zn, boston$crim, main = "crim vs zn")
abline(m1,col = "red")
```




```
m2 <- lm(crim~indus, data = boston)
summary(m2)
```

```
##
## 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
```

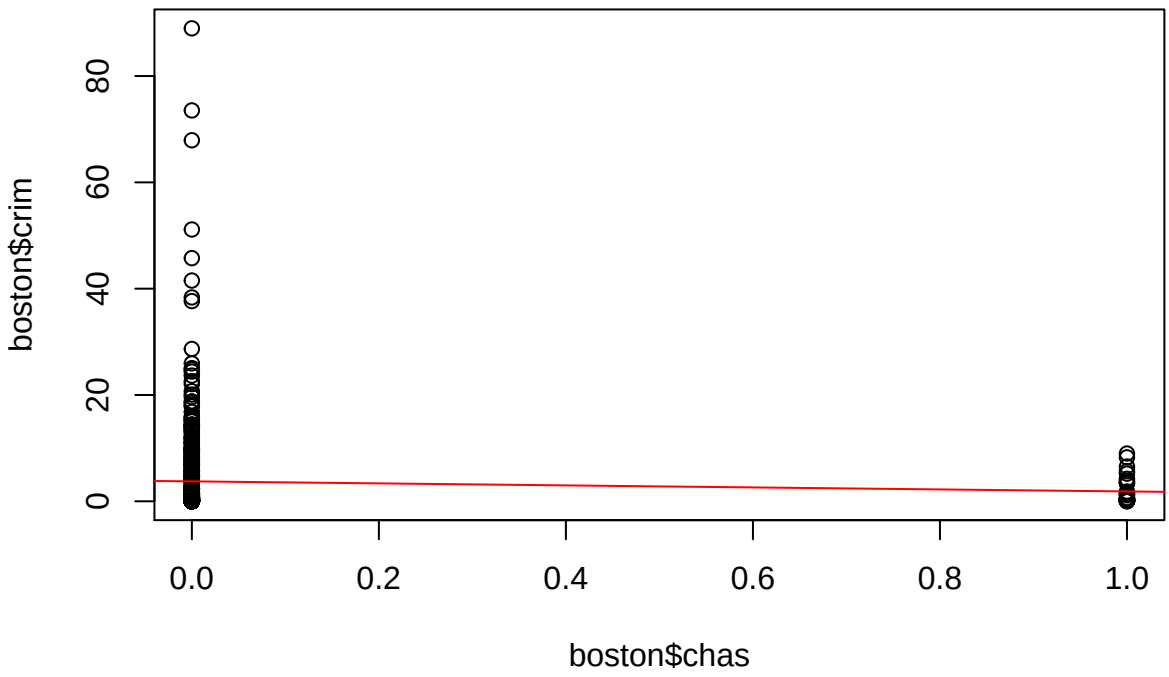
```
plot(boston$indus, boston$crim, main = "crim vs indus")
abline(m2,col = "red")
```




```
m3 <- lm(crim~chas, data = boston)
summary(m3)
```

```
##
## 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
```

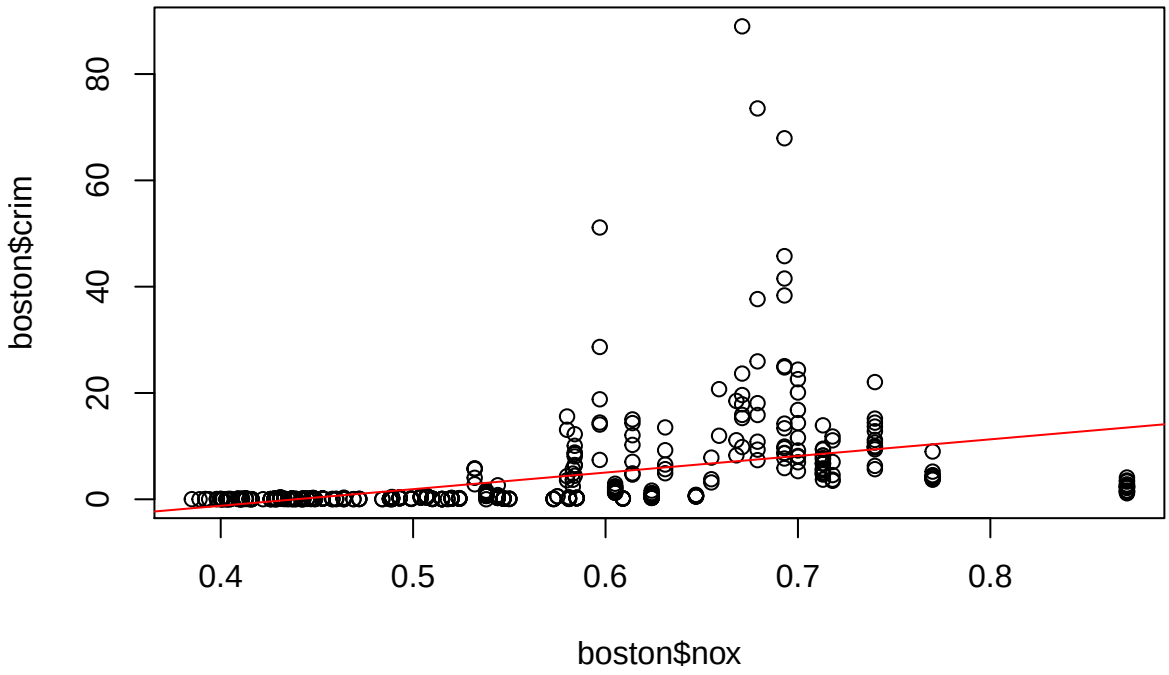
```
plot(boston$chas, boston$crim, main = "crim vs chas")
abline(m3,col = "red")
```




```
m4 <- lm(crim~nox, data = boston)
summary(m4)
```

```
##
## 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
```

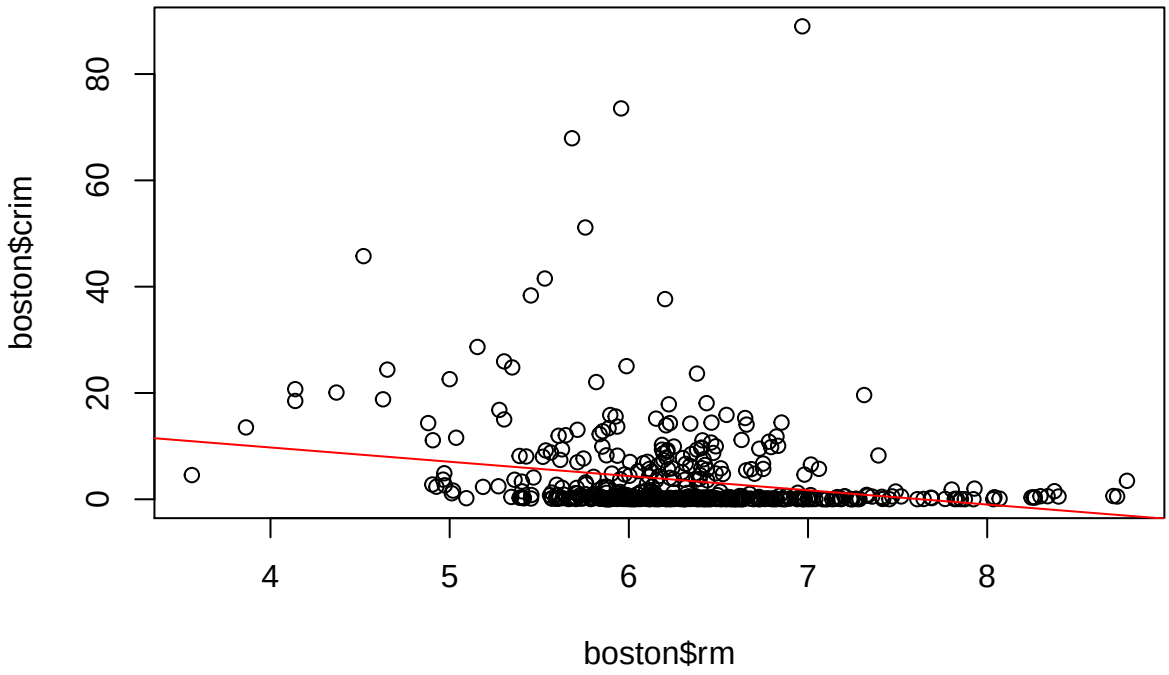
```
plot(boston$nox, boston$crim, main = "crim vs nox")
abline(m4,col = "red")
```




```
m5 <- lm(crim~rm, data = boston)
summary(m5)
```

```
##
## 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
```

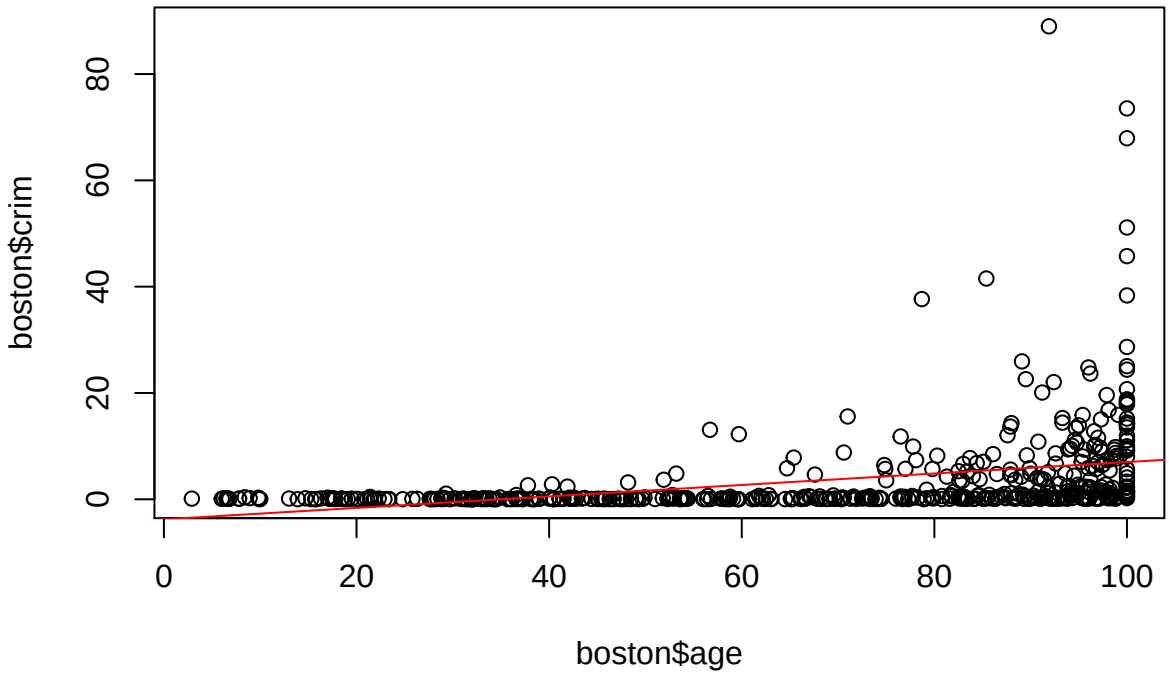
```
plot(boston$rm, boston$crim, main = "crim vs rm")
abline(m5,col = "red")
```




```
m6 <- lm(crim~age, data = boston)
summary(m6)
```

```
##
## 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
```

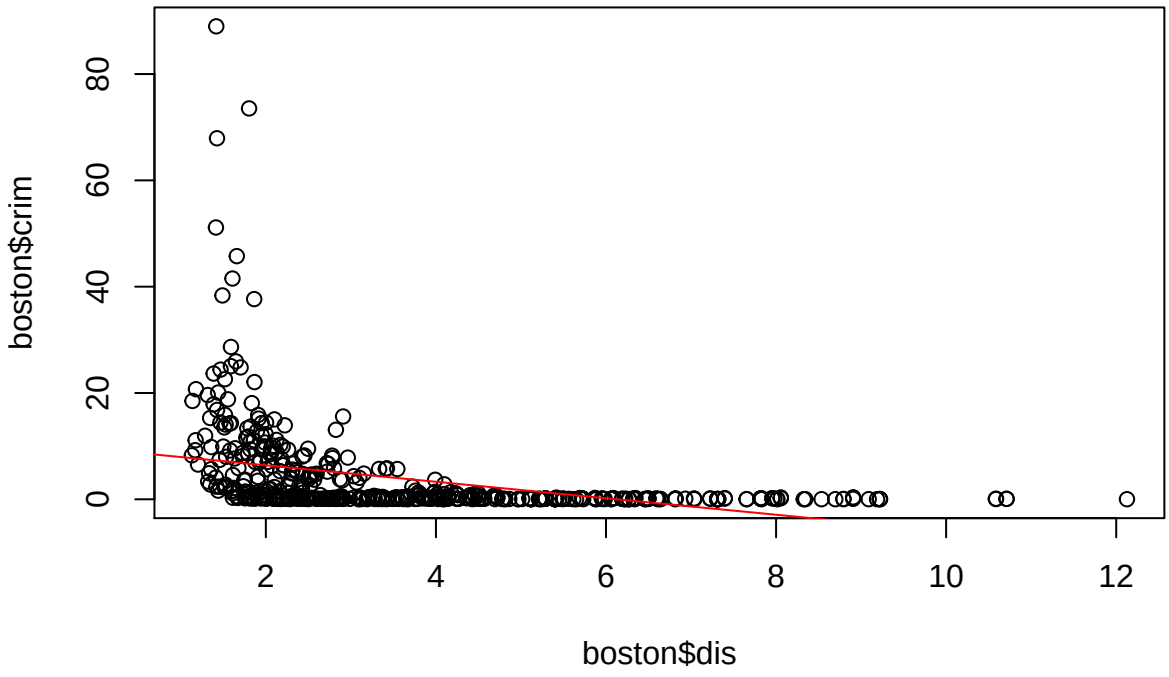
```
plot(boston$age, boston$crim, main = "crim vs age")
abline(m6,col = "red")
```




```
m7 <- lm(crim~dis, data = boston)
summary(m7)
```

```
##
## 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
```

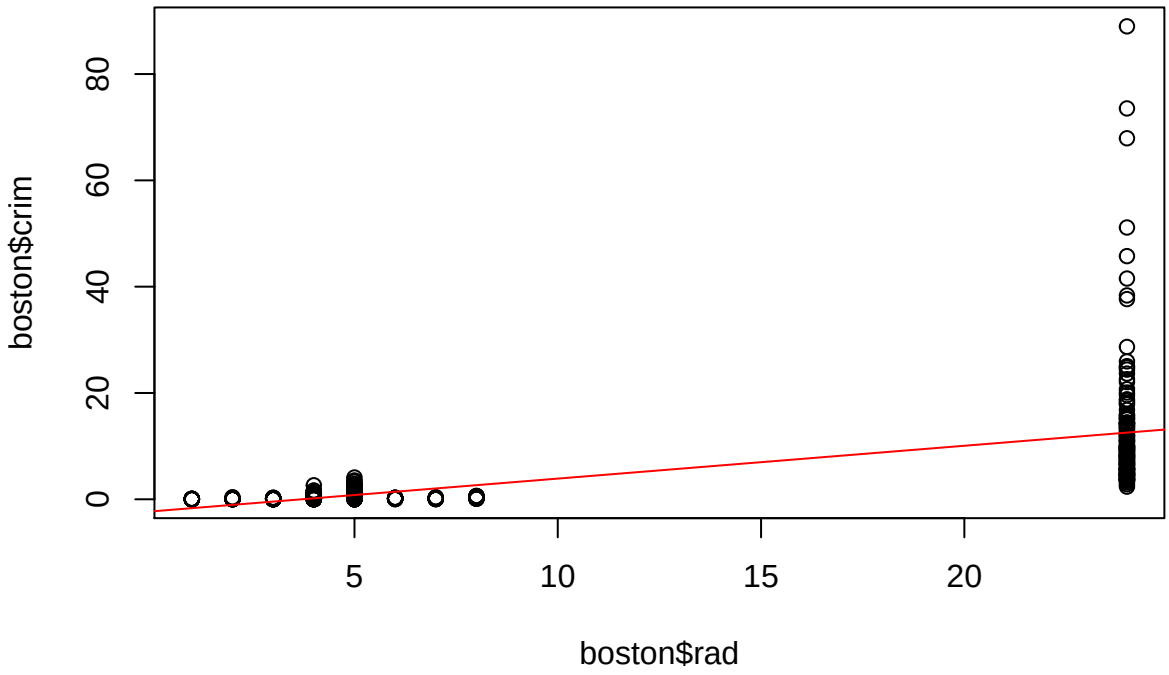
```
plot(boston$dis, boston$crim, main = "crim vs dis")
abline(m7,col = "red")
```




```
m8 <- lm(crim~rad, data = boston)
summary(m8)
```

```
##
## 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
```

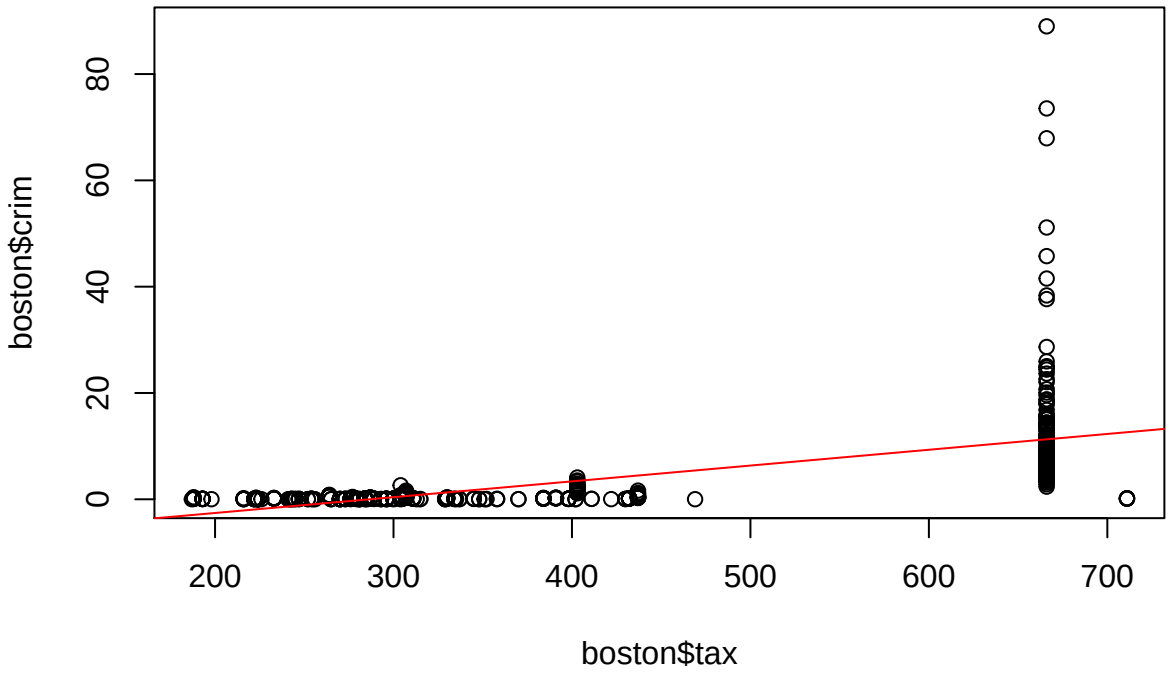
```
plot(boston$rad, boston$crim, main = "crim vs rad")
abline(m8,col = "red")
```




```
m9 <- lm(crim~tax, data = boston)
summary(m9)
```

```
##
## 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
```

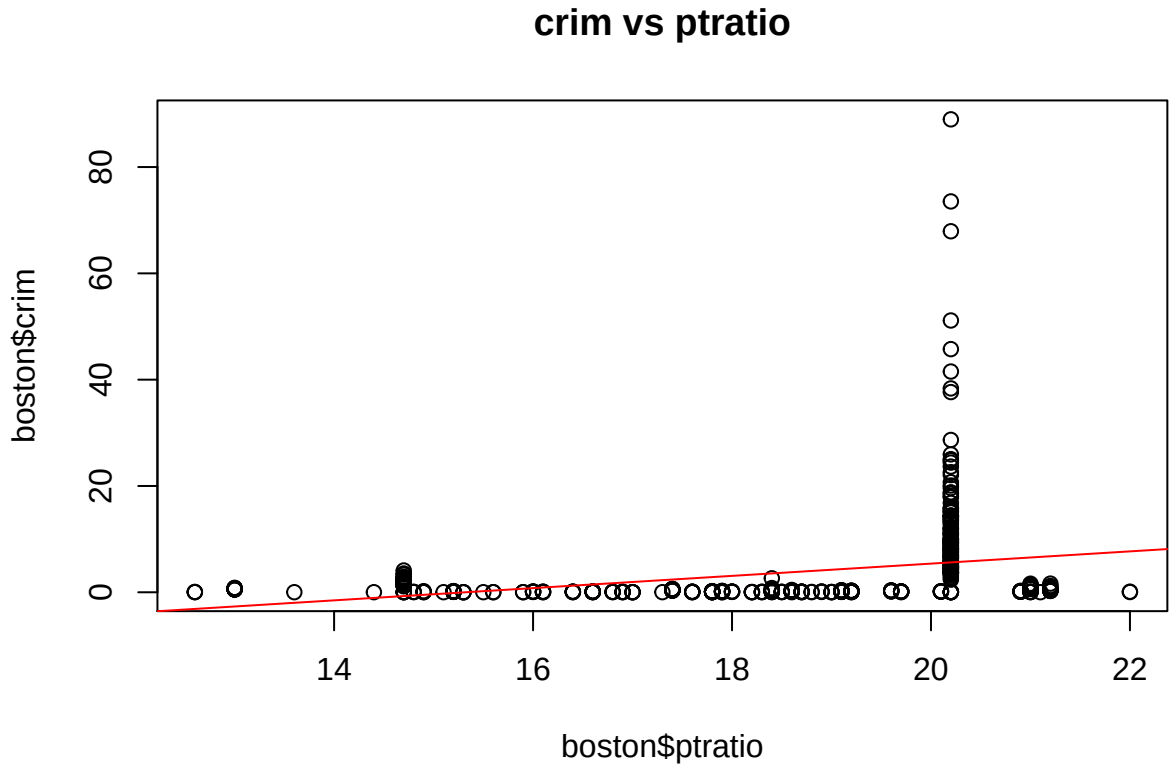
```
plot(boston$tax, boston$crim, main = "crim vs tax")
abline(m9,col = "red")
```




```
m10 <- lm(crim~ptratio, data = boston)
summary(m10)
```

```
##
## 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
```

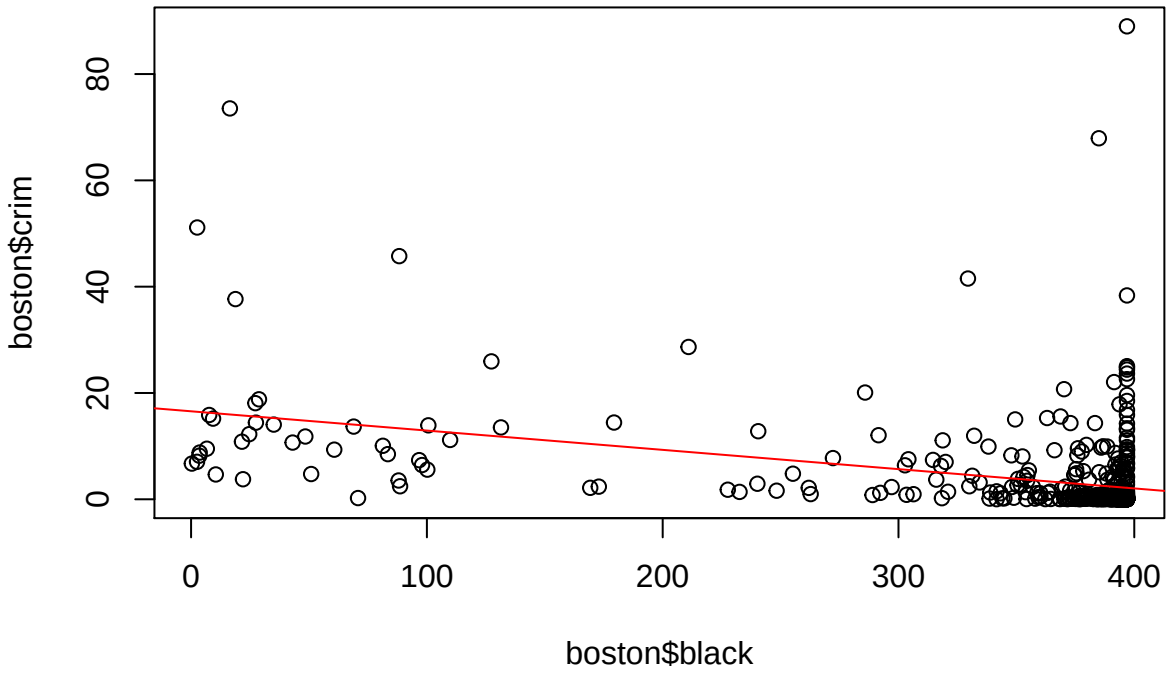
```
plot(boston$ptratio, boston$crim, main = "crim vs ptratio")
abline(m10,col = "red")
```




```
m11 <- lm(crim~black, data = boston)
summary(m11)
```

```
##
## 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
```

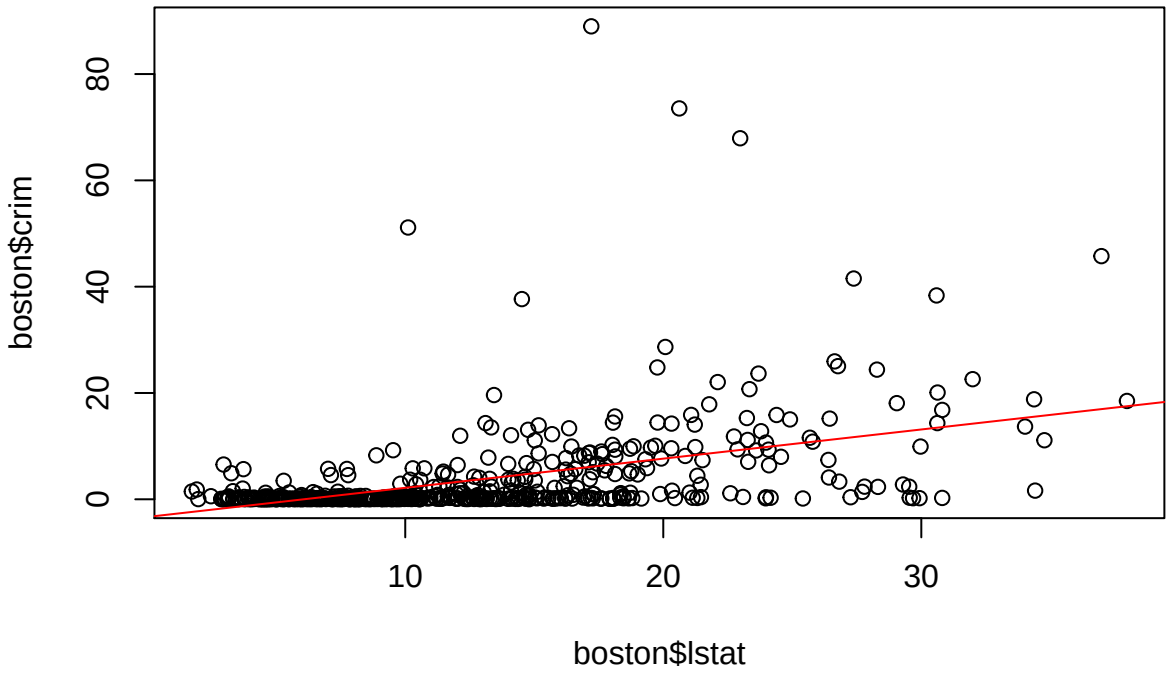
```
plot(boston$black, boston$crim, main = "crim vs black")
abline(m11,col = "red")
```




```
m12 <- lm(crim~lstat, data = boston)
summary(m12)
```

```
##
## 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
```

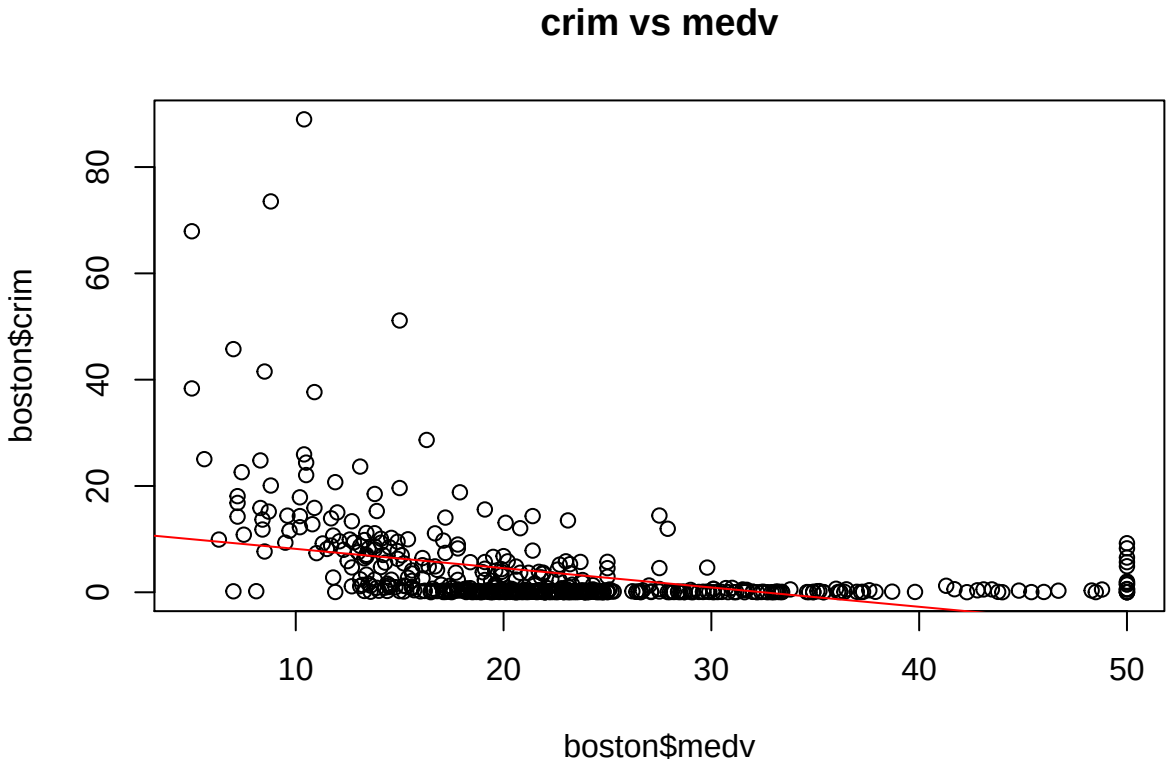
```
plot(boston$lstat, boston$crim, main = "crim vs lstat")
abline(m12,col = "red")
```




```
m13 <- lm(crim~medv, data = boston)
summary(m13)
```

```
##
## 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(boston$medv, boston$crim, main = "crim vs medv")
abline(m13,col = "red")
```



Comment: From all summaries, we see that all predictors have a really small p-value except the predictor "chas", which is 0.209. Therefore, we can conclude that there is a statistically significant association between "crim" and all predictors except "chas".

(ii)

```
multiple <- lm(crim~., data = boston)
summary(multiple)
```

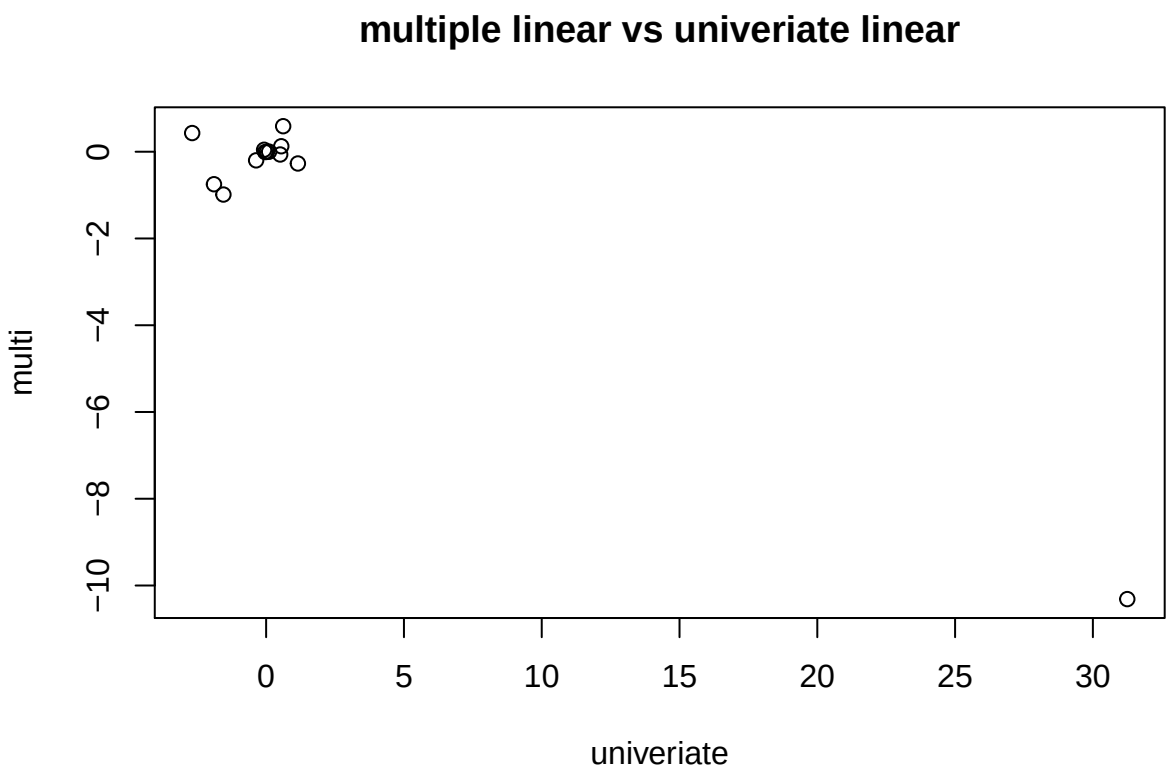
```
##
## Call:
## lm(formula = crim ~ ., data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.924 -2.120 -0.353  1.019 75.051
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *
## zn            0.044855   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
```

Comments: The multiple linear regression does not fit the data well because the residual standard error is high and R-sqaured is low. We see that zn, dis, rad, black, medv have p-vlaue below 0.05, and other predictors above 0.05. For those predictors above 0.05, we can not reject hypothesis. Therefore, we can reject the null hypothesis for zn, dis, rad, black, medv.

###(iii)

```
univeriate <- c(m1$coefficient[2], m2$coefficient[2], m3$coefficient[2],
                m4$coefficient[2], m5$coefficient[2], m6$coefficient[2],
                m7$coefficient[2], m8$coefficient[2], m9$coefficient[2],
                m10$coefficient[2], m11$coefficient[2], m12$coefficient[2],
                m13$coefficient[2])
multi <- c(multiple$coefficients[-1])
plot(univeriate, multi, main = "multiple linear vs univeriate linear")
```



Comment: There are differences between univeriate linear regression and multiple linear regression. For univeriate linear regression, we are fitting one predictor at one time. This means that the coefficent is the increasing of that particular predictor, with absence of other predictors. For multiple linear regression, the

slope is the increase of one predictor, while holding other predictors fixed. Since two types of regression have different interpretation of slopes, there is no relationship between coeffcients from two models.

(iv)

```
poly1 <- lm(crim~zn+I(zn^2)+I(zn^3), data = boston)
summary(poly1)
```

```
##
## Call:
## lm(formula = crim ~ zn + I(zn^2) + I(zn^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.821 -4.614 -1.294  0.473 84.130
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.846e+00  4.330e-01  11.192  < 2e-16 ***
## zn          -3.322e-01  1.098e-01  -3.025  0.00261 **
## I(zn^2)      6.483e-03  3.861e-03   1.679  0.09375 .
## I(zn^3)     -3.776e-05  3.139e-05  -1.203  0.22954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.372 on 502 degrees of freedom
## Multiple R-squared:  0.05824,    Adjusted R-squared:  0.05261
## F-statistic: 10.35 on 3 and 502 DF,  p-value: 1.281e-06
```

```
poly2 <- lm(crim~indus+I(indus^2)+I(indus^3), data = boston)
summary(poly2)
```

```
##
## Call:
## lm(formula = crim ~ indus + I(indus^2) + I(indus^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.278 -2.514  0.054  0.764 79.713
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.6625683  1.5739833   2.327   0.0204 *
## indus       -1.9652129  0.4819901  -4.077 5.30e-05 ***
## I(indus^2)   0.2519373  0.0393221   6.407 3.42e-10 ***
## I(indus^3)  -0.0069760  0.0009567  -7.292 1.20e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.423 on 502 degrees of freedom
## Multiple R-squared:  0.2597, Adjusted R-squared:  0.2552
## F-statistic: 58.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly3 <- lm(crim~chas+I(chas^2)+I(chas^3), data = boston)
summary(poly3)
```

```
##
## Call:
## lm(formula = crim ~ chas + I(chas^2) + I(chas^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.738 -3.661 -3.435  0.018 85.232
##
## Coefficients: (2 not defined because of singularities)
##             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
## I(chas^2)         NA         NA      NA       NA
## I(chas^3)         NA         NA      NA       NA
## ---
## 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
poly4 <- lm(crim~nox+I(nox^2)+I(nox^3), data = boston)
summary(poly4)
```

```
##
## Call:
## lm(formula = crim ~ nox + I(nox^2) + I(nox^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.110 -2.068 -0.255  0.739 78.302
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   233.09      33.64   6.928 1.31e-11 ***
## nox         -1279.37     170.40  -7.508 2.76e-13 ***
## I(nox^2)     2248.54     279.90   8.033 6.81e-15 ***
## I(nox^3)    -1245.70     149.28  -8.345 6.96e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.234 on 502 degrees of freedom
## Multiple R-squared:  0.297,  Adjusted R-squared:  0.2928
## F-statistic: 70.69 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly5 <- lm(crim~rm+I(rm^2)+I(rm^3), data = boston)
summary(poly5)
```

```
##
```

```
## Call:
## lm(formula = crim ~ rm + I(rm^2) + I(rm^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.485  -3.468  -2.221  -0.015  87.219
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 112.6246    64.5172   1.746   0.0815 .
## rm          -39.1501    31.3115  -1.250   0.2118
## I(rm^2)       4.5509     5.0099   0.908   0.3641
## I(rm^3)      -0.1745     0.2637  -0.662   0.5086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.33 on 502 degrees of freedom
## Multiple R-squared:  0.06779,    Adjusted R-squared:  0.06222
## F-statistic: 12.17 on 3 and 502 DF,  p-value: 1.067e-07
```

```
poly6 <- lm(crim~age+I(age^2)+I(age^3), data = boston)
summary(poly6)
```

```
##
## Call:
## lm(formula = crim ~ age + I(age^2) + I(age^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.762 -2.673 -0.516  0.019 82.842
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.549e+00  2.769e+00  -0.920  0.35780
## age          2.737e-01  1.864e-01   1.468  0.14266
## I(age^2)    -7.230e-03  3.637e-03  -1.988  0.04738 *
## I(age^3)     5.745e-05  2.109e-05   2.724  0.00668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.84 on 502 degrees of freedom
## Multiple R-squared:  0.1742, Adjusted R-squared:  0.1693
## F-statistic: 35.31 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly7 <- lm(crim~dis+I(dis^2)+I(dis^3), data = boston)
summary(poly7)
```

```
##
## Call:
## lm(formula = crim ~ dis + I(dis^2) + I(dis^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -10.757  -2.588   0.031   1.267  76.378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.0476     2.4459  12.285  < 2e-16 ***
## dis         -15.5543     1.7360  -8.960  < 2e-16 ***
## I(dis^2)      2.4521     0.3464   7.078 4.94e-12 ***
## I(dis^3)     -0.1186     0.0204  -5.814 1.09e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.331 on 502 degrees of freedom
## Multiple R-squared:  0.2778, Adjusted R-squared:  0.2735
## F-statistic: 64.37 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly8 <- lm(crim~rad+I(rad^2)+I(rad^3), data = boston)
summary(poly8)
```

```
##
## Call:
## lm(formula = crim ~ rad + I(rad^2) + I(rad^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.381  -0.412  -0.269   0.179  76.217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.605545   2.050108  -0.295    0.768
## rad          0.512736   1.043597   0.491    0.623
## I(rad^2)    -0.075177   0.148543  -0.506    0.613
## I(rad^3)     0.003209   0.004564   0.703    0.482
##
## Residual standard error: 6.682 on 502 degrees of freedom
## Multiple R-squared:    0.4,  Adjusted R-squared:  0.3965
## F-statistic: 111.6 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly9 <- lm(crim~tax+I(tax^2)+I(tax^3), data = boston)
summary(poly9)
```

```
##
## Call:
## lm(formula = crim ~ tax + I(tax^2) + I(tax^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.273  -1.389   0.046   0.536  76.950
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.918e+01  1.180e+01   1.626    0.105
## tax         -1.533e-01  9.568e-02  -1.602    0.110
## I(tax^2)     3.608e-04  2.425e-04   1.488    0.137
```

```
## I(tax^3)    -2.204e-07  1.889e-07  -1.167    0.244
##
## Residual standard error: 6.854 on 502 degrees of freedom
## Multiple R-squared:  0.3689, Adjusted R-squared:  0.3651
## F-statistic:  97.8 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly10 <- lm(crim~ptratio+I(ptratio^2)+I(ptratio^3), data = boston)
summary(poly10)
```

```
##
## Call:
## lm(formula = crim ~ ptratio + I(ptratio^2) + I(ptratio^3), data = boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.833 -4.146 -1.655  1.408 82.697
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  477.18405  156.79498   3.043  0.00246 **
## ptratio      -82.36054   27.64394  -2.979  0.00303 **
## I(ptratio^2)   4.63535    1.60832   2.882  0.00412 **
## I(ptratio^3)  -0.08476    0.03090  -2.743  0.00630 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.122 on 502 degrees of freedom
## Multiple R-squared:  0.1138, Adjusted R-squared:  0.1085
## F-statistic: 21.48 on 3 and 502 DF,  p-value: 4.171e-13
```

```
poly11 <- lm(crim~black+I(black^2)+I(black^3), data = boston)
summary(poly11)
```

```
##
## Call:
## lm(formula = crim ~ black + I(black^2) + I(black^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.096  -2.343  -2.128  -1.439  86.790
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.826e+01  2.305e+00   7.924  1.5e-14 ***
## black       -8.356e-02  5.633e-02  -1.483    0.139
## I(black^2)   2.137e-04  2.984e-04   0.716    0.474
## I(black^3)  -2.652e-07  4.364e-07  -0.608    0.544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.955 on 502 degrees of freedom
## Multiple R-squared:  0.1498, Adjusted R-squared:  0.1448
## F-statistic: 29.49 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly12 <- lm(crim~lstat+I(lstat^2)+I(lstat^3), data = boston)
summary(poly12)
```

```
##
## Call:
## lm(formula = crim ~ lstat + I(lstat^2) + I(lstat^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.234  -2.151  -0.486   0.066  83.353
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.2009656  2.0286452   0.592   0.5541
## lstat       -0.4490656  0.4648911  -0.966   0.3345
## I(lstat^2)   0.0557794  0.0301156   1.852   0.0646 .
## I(lstat^3)  -0.0008574  0.0005652  -1.517   0.1299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.629 on 502 degrees of freedom
## Multiple R-squared:  0.2179, Adjusted R-squared:  0.2133
## F-statistic: 46.63 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly13 <- lm(crim~medv+I(medv^2)+I(medv^3), data = boston)
summary(poly13)
```

```
##
## Call:
## lm(formula = crim ~ medv + I(medv^2) + I(medv^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.427  -1.976  -0.437   0.439  73.655
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 53.1655381  3.3563105  15.840  < 2e-16 ***
## medv        -5.0948305  0.4338321 -11.744  < 2e-16 ***
## I(medv^2)    0.1554965  0.0171904   9.046  < 2e-16 ***
## I(medv^3)   -0.0014901  0.0002038  -7.312 1.05e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.569 on 502 degrees of freedom
## Multiple R-squared:  0.4202, Adjusted R-squared:  0.4167
## F-statistic: 121.3 on 3 and 502 DF,  p-value: < 2.2e-16
```

Comments: From all models, we conclude that quatratic and cubic terms of predictors zn, rm, rad, tax, black, lstat have p-value greater than 0.05. This means that those predictors are not likely to have non-linear relationship between them and the response. On the other hand, quatratic and cubic terms of predictors indus, nox, age, dis, ptratio, medv have p-value smaller than 0.05. This suggests that those predictors are likely to have non-linear relationship between them and the response.

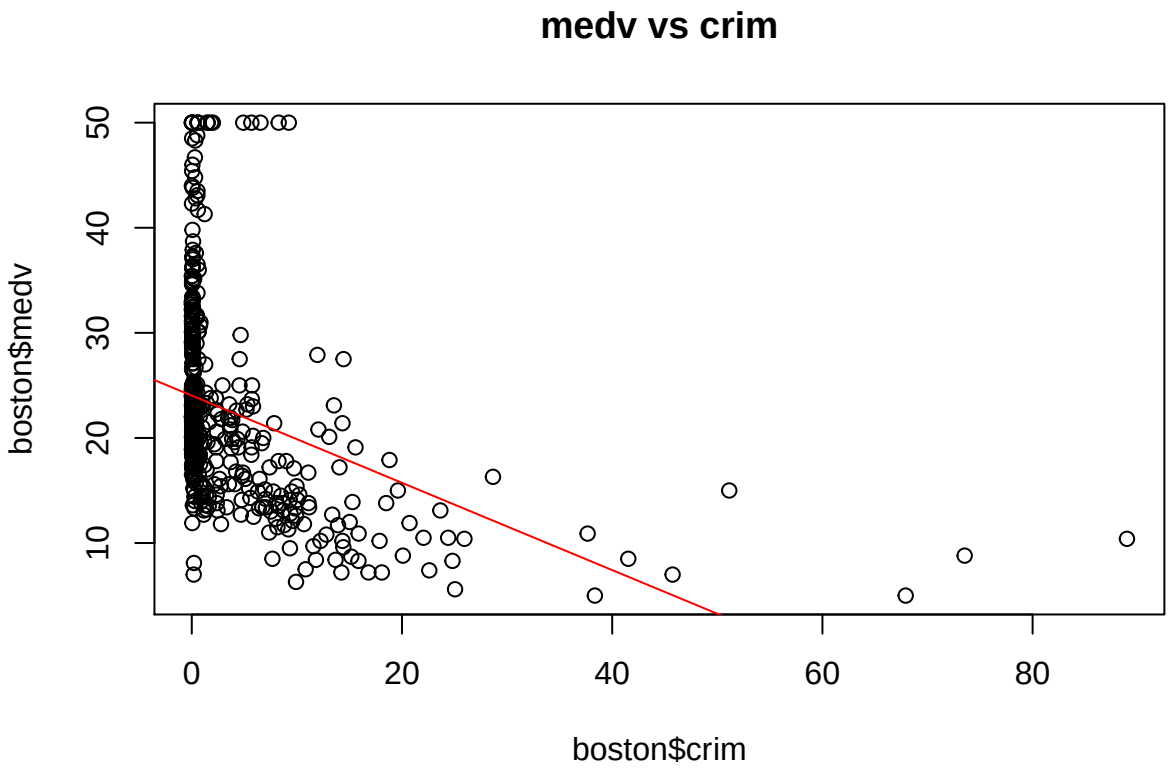
(b)

(i)

```
m1 <- lm(medv~crim, data = boston)
summary(m1)
```

```
##
## Call:
## lm(formula = medv ~ crim, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.957  -5.449  -2.007   2.512  29.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.03311    0.40914   58.74   <2e-16 ***
## crim        -0.41519    0.04389   -9.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.484 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
```

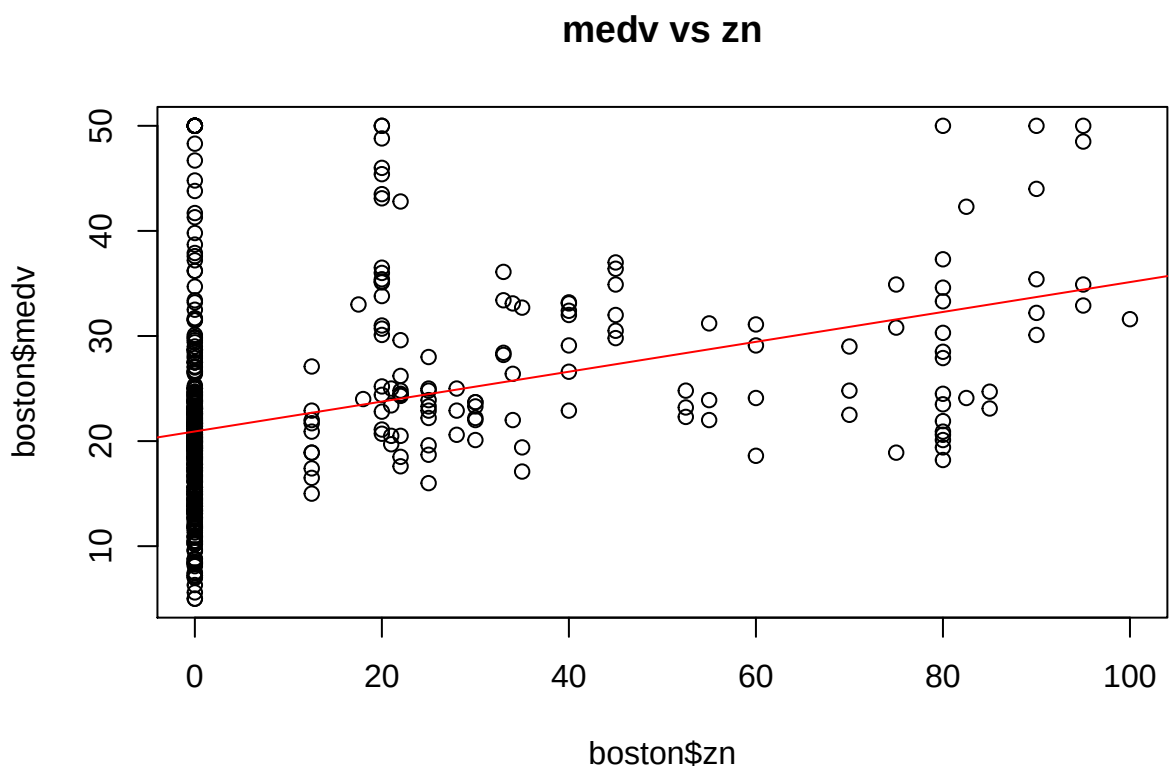
```
plot(boston$crim, boston$medv, main = "medv vs crim")
abline(m1,col = "red")
```



```r
m2 <- lm(medv~zn, data = boston)
summary(m2)
```

```
##
## Call:
## lm(formula = medv ~ zn, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.918  -5.518  -1.006   2.757  29.082
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.91758    0.42474  49.248   <2e-16 ***
## zn           0.14214    0.01638   8.675   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.587 on 504 degrees of freedom
## Multiple R-squared:  0.1299, Adjusted R-squared:  0.1282
## F-statistic: 75.26 on 1 and 504 DF,  p-value: < 2.2e-16
```

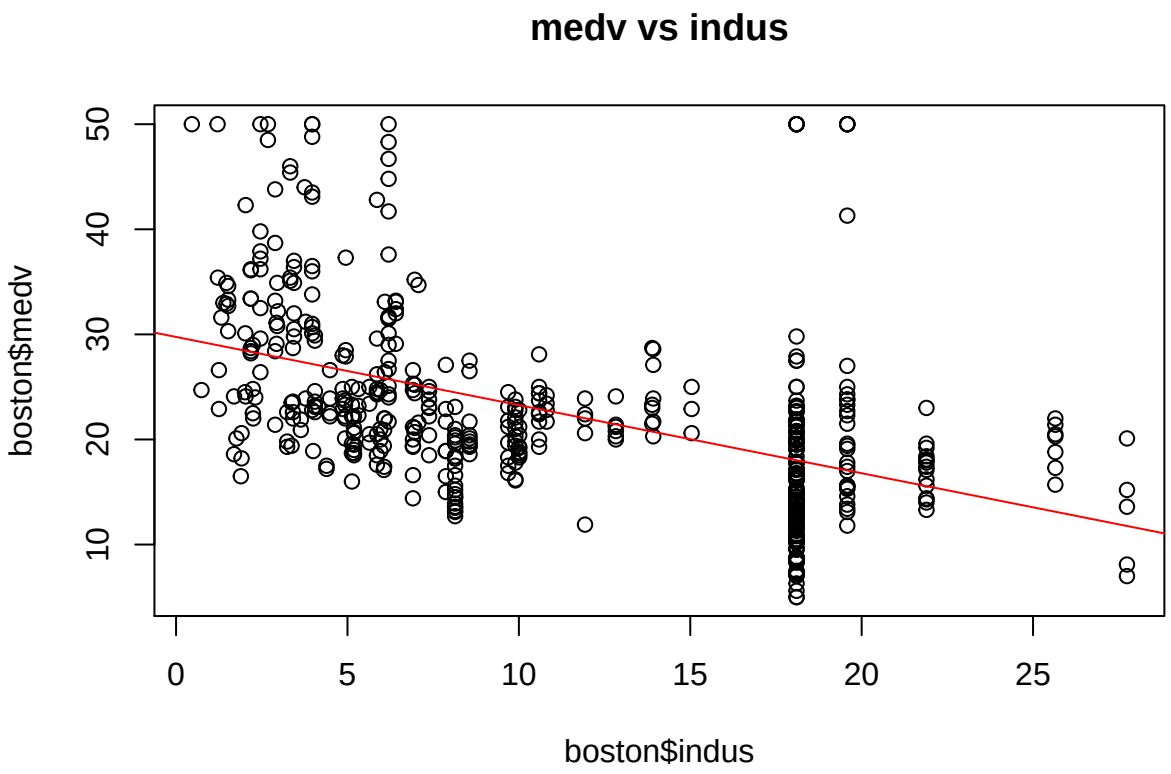
```r
plot(boston$zn, boston$medv, main = "medv vs zn")
abline(m2,col = "red")
```




```r
m3 <- lm(medv~indus, data = boston)
summary(m3)
```

```
##
## Call:
## lm(formula = medv ~ indus, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.017  -4.917  -1.457   3.180  32.943
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 29.75490    0.68345   43.54   <2e-16 ***
## indus       -0.64849    0.05226  -12.41   <2e-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:  0.234,  Adjusted R-squared:  0.2325
## F-statistic:   154 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(boston$indus, boston$medv, main = "medv vs indus")
abline(m3,col = "red")
```

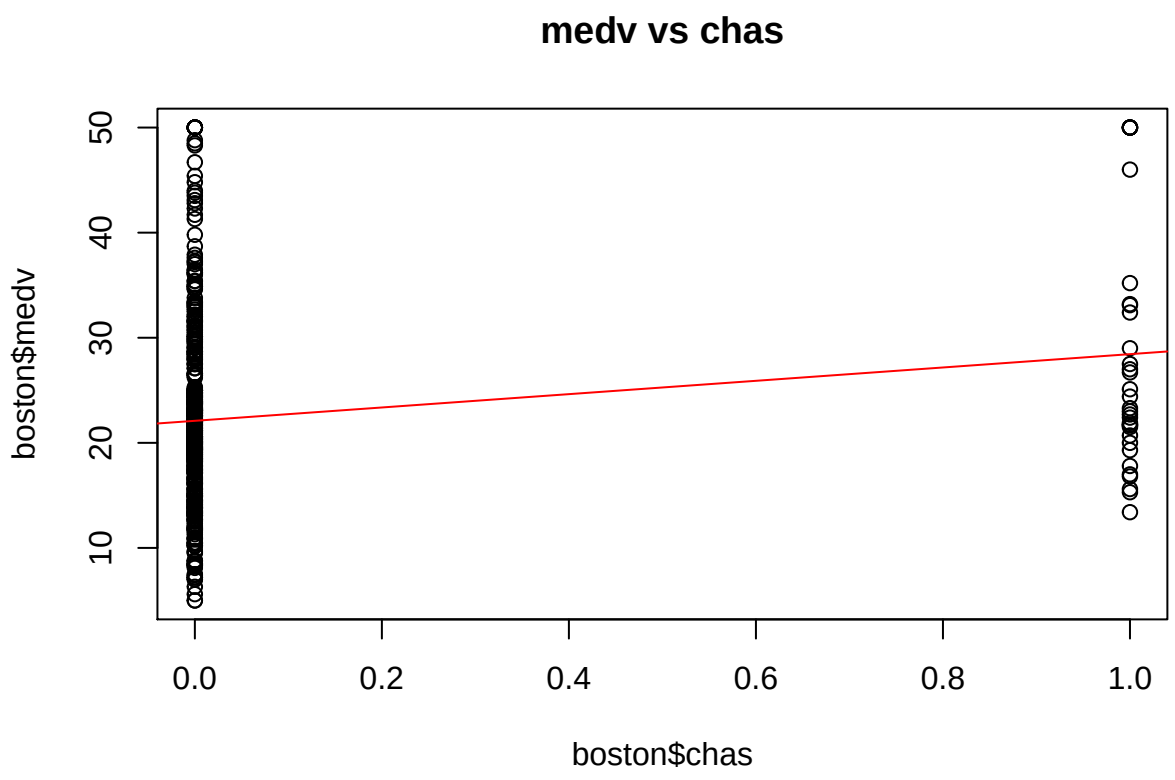



```
m4 <- lm(medv~chas, data = boston)
summary(m4)
```

```
##
## Call:
## lm(formula = medv ~ chas, data = boston)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.094  -5.894  -1.417   2.856  27.906
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.0938     0.4176  52.902  < 2e-16 ***
## chas          6.3462     1.5880   3.996 7.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.064 on 504 degrees of freedom
## Multiple R-squared:  0.03072,    Adjusted R-squared:  0.02879
## F-statistic: 15.97 on 1 and 504 DF,  p-value: 7.391e-05
```

```
plot(boston$chas, boston$medv, main = "medv vs chas")
abline(m4,col = "red")
```

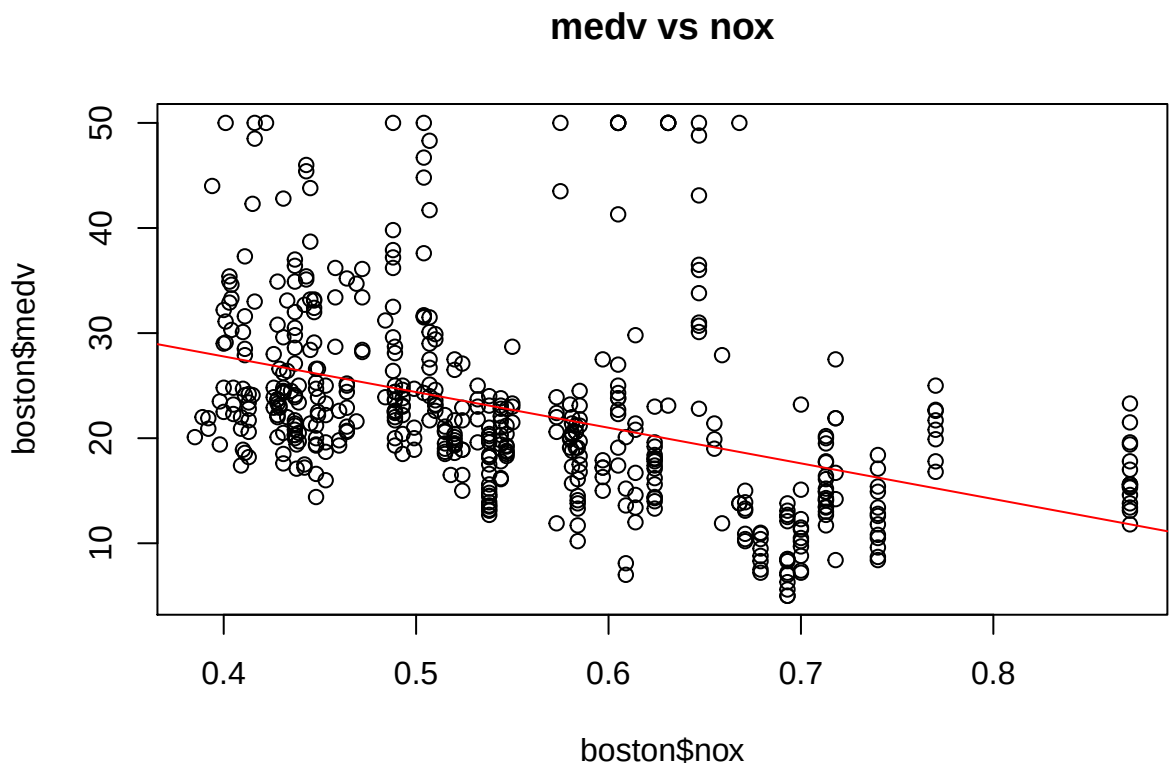



```
m5 <- lm(medv~nox, data = boston)
summary(m5)
```

```
##
## Call:
## lm(formula = medv ~ nox, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -13.691  -5.121  -2.161   2.959  31.310
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   41.346      1.811   22.83   <2e-16 ***
## nox          -33.916      3.196  -10.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.323 on 504 degrees of freedom
## Multiple R-squared:  0.1826, Adjusted R-squared:  0.181
## F-statistic: 112.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(boston$nox, boston$medv, main = "medv vs nox")
abline(m5,col = "red")
```

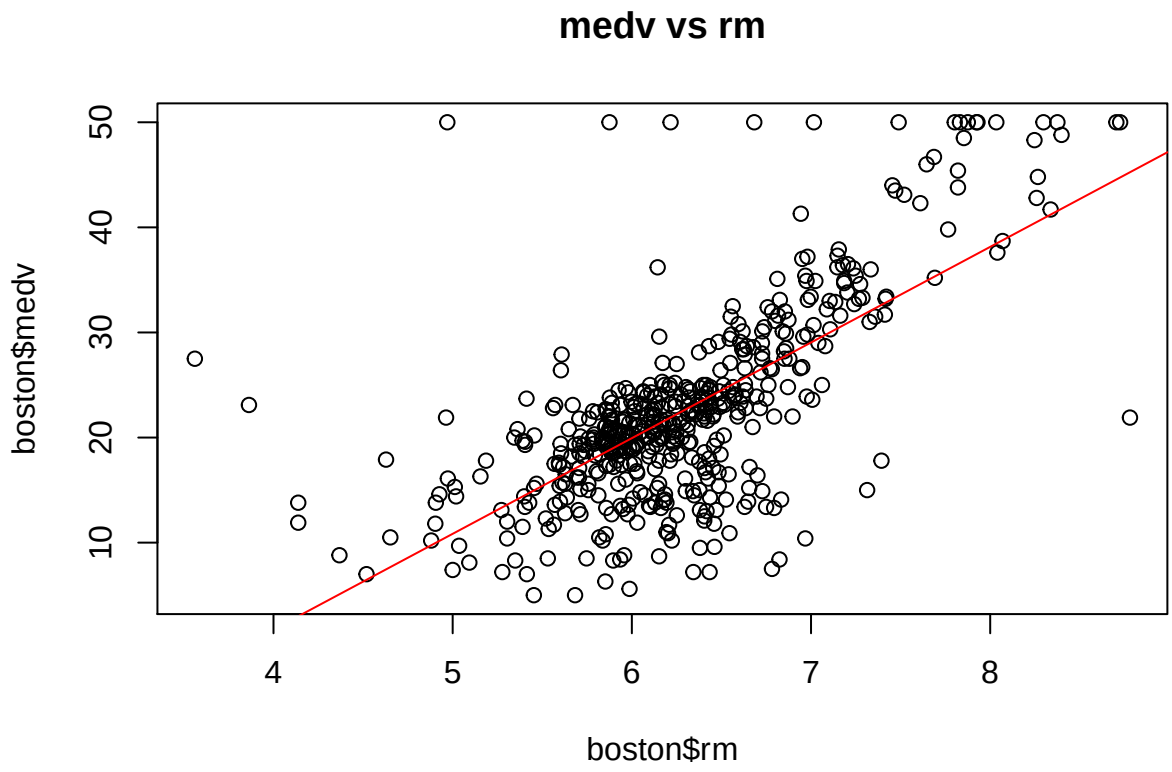



```
m6 <- lm(medv~rm, data = boston)
summary(m6)
```

```
##
## Call:
## lm(formula = medv ~ rm, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.346  -2.547   0.090   2.986  39.433
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -34.671      2.650  -13.08   <2e-16 ***
## rm             9.102      0.419   21.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.616 on 504 degrees of freedom
## Multiple R-squared:  0.4835, Adjusted R-squared:  0.4825
## F-statistic: 471.8 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(boston$rm, boston$medv, main = "medv vs rm")
abline(m6,col = "red")
```

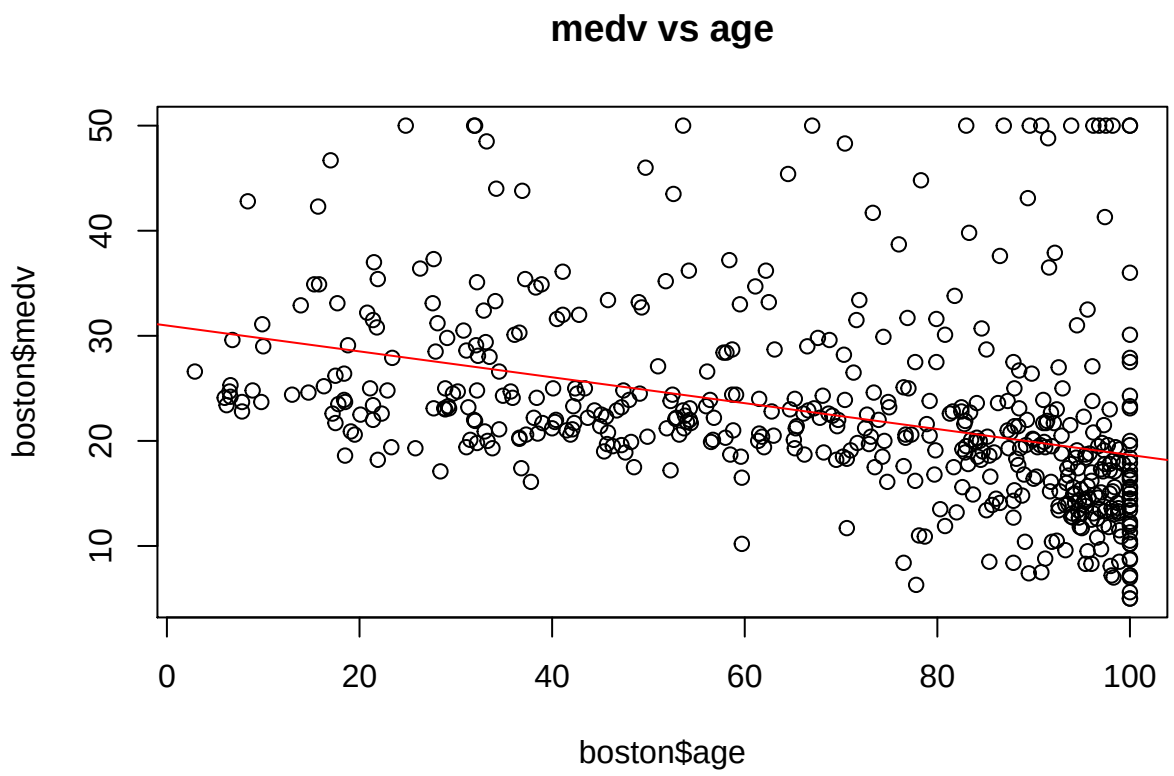



```
m7 <- lm(medv~age, data = boston)
summary(m7)
```

```
##
## Call:
## lm(formula = medv ~ age, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.097  -5.138  -1.958   2.397  31.338
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.97868    0.99911  31.006   <2e-16 ***
## age         -0.12316    0.01348  -9.137   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.527 on 504 degrees of freedom
## Multiple R-squared:  0.1421, Adjusted R-squared:  0.1404
## F-statistic: 83.48 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(boston$age, boston$medv, main = "medv vs age")
abline(m7,col = "red")
```

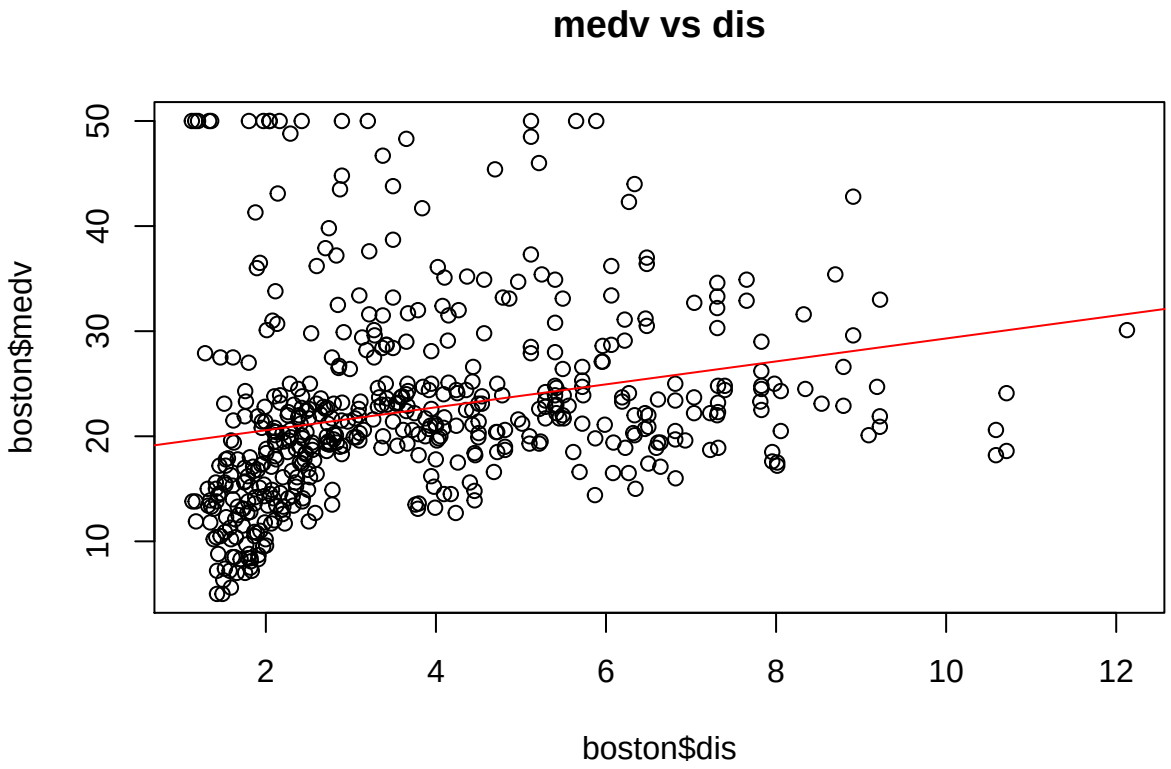



```
m8 <- lm(medv~dis, data = boston)
summary(m8)
```

```
##
## Call:
## lm(formula = medv ~ dis, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.016  -5.556  -1.865   2.288  30.377
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.3901     0.8174  22.499  < 2e-16 ***
## dis           1.0916     0.1884   5.795 1.21e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 8.914 on 504 degrees of freedom
## Multiple R-squared:  0.06246,    Adjusted R-squared:  0.0606
## F-statistic: 33.58 on 1 and 504 DF,  p-value: 1.207e-08
```

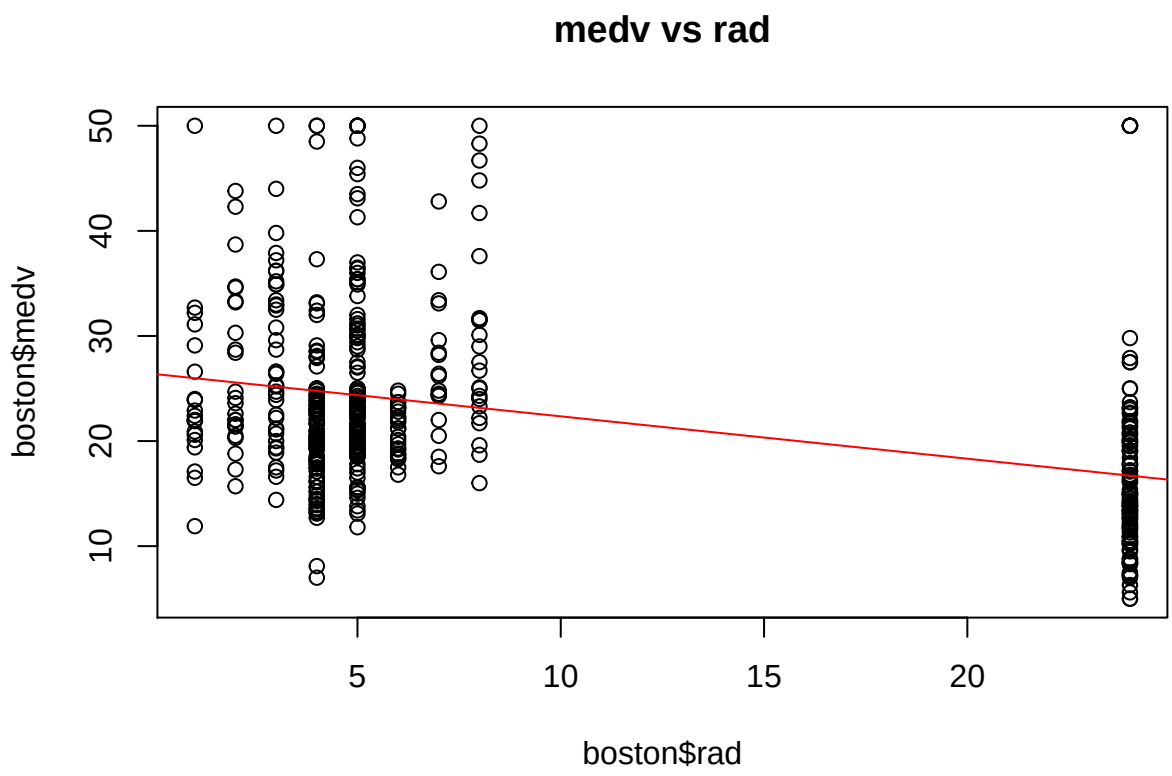
```
plot(boston$dis, boston$medv, main = "medv vs dis")
abline(m8,col = "red")
```




```
m9 <- lm(medv~rad, data = boston)
summary(m9)
```

```
##
## Call:
## lm(formula = medv ~ rad, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.770  -5.199  -1.967   3.321  33.292
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.38213    0.56176  46.964   <2e-16 ***
## rad         -0.40310    0.04349  -9.269   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.509 on 504 degrees of freedom
## Multiple R-squared:  0.1456, Adjusted R-squared:  0.1439
## F-statistic: 85.91 on 1 and 504 DF,  p-value: < 2.2e-16
```

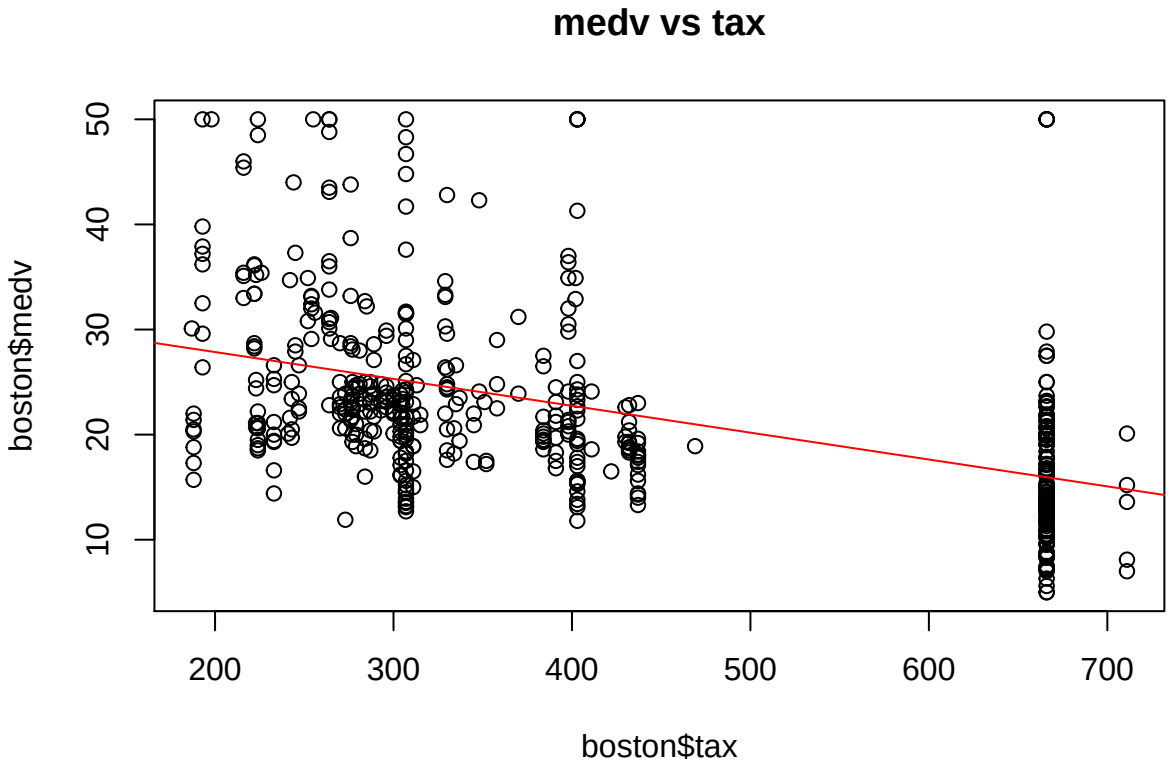
```
plot(boston$rad, boston$medv, main = "medv vs rad")
abline(m9,col = "red")
```




```
m10 <- lm(medv~tax, data = boston)
summary(m10)
```

```
##
## Call:
## lm(formula = medv ~ tax, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.091  -5.173  -2.085   3.158  34.058
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.970654   0.948296   34.77   <2e-16 ***
## tax         -0.025568   0.002147  -11.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.133 on 504 degrees of freedom
## Multiple R-squared:  0.2195, Adjusted R-squared:  0.218
## F-statistic: 141.8 on 1 and 504 DF,  p-value: < 2.2e-16
```

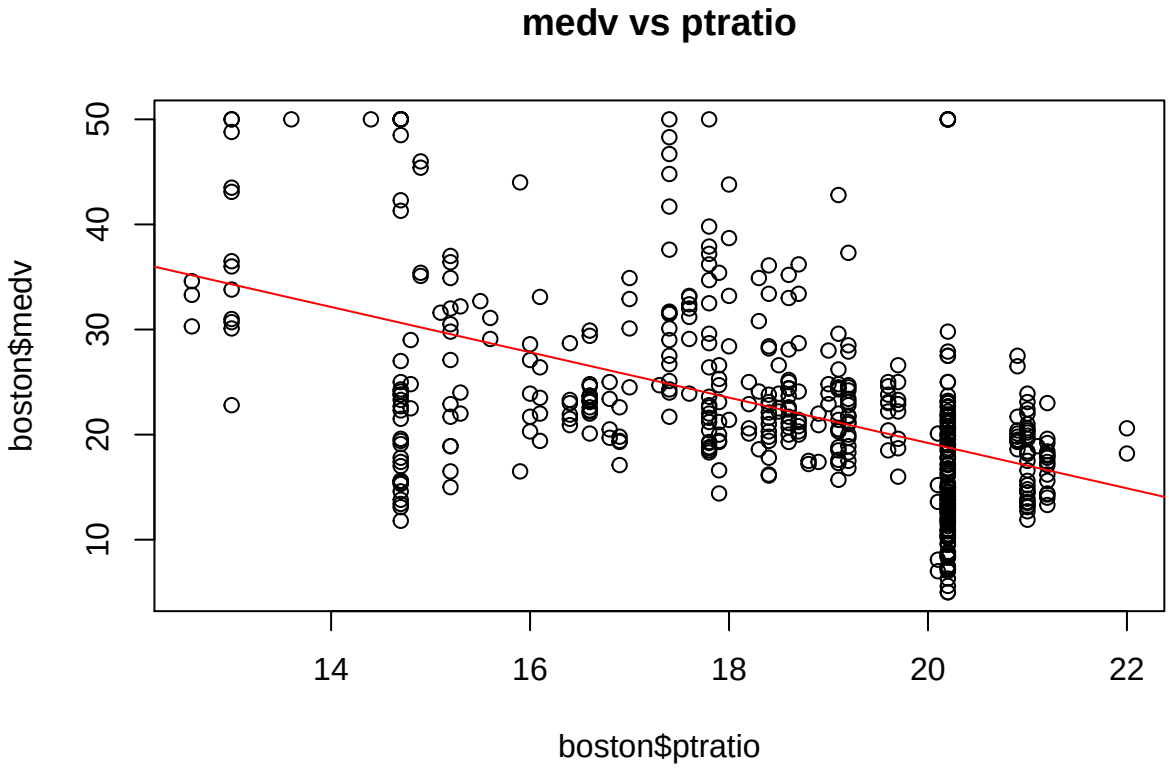
```
plot(boston$tax, boston$medv, main = "medv vs tax")
abline(m10,col = "red")
```




```
m11 <- lm(medv~ptratio, data = boston)
summary(m11)
```

```
##
## Call:
## lm(formula = medv ~ ptratio, data = boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.8342  -4.8262  -0.6426   3.1571  31.2303
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   62.345      3.029   20.58   <2e-16 ***
## ptratio       -2.157      0.163  -13.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.931 on 504 degrees of freedom
## Multiple R-squared:  0.2578, Adjusted R-squared:  0.2564
## F-statistic: 175.1 on 1 and 504 DF,  p-value: < 2.2e-16
```

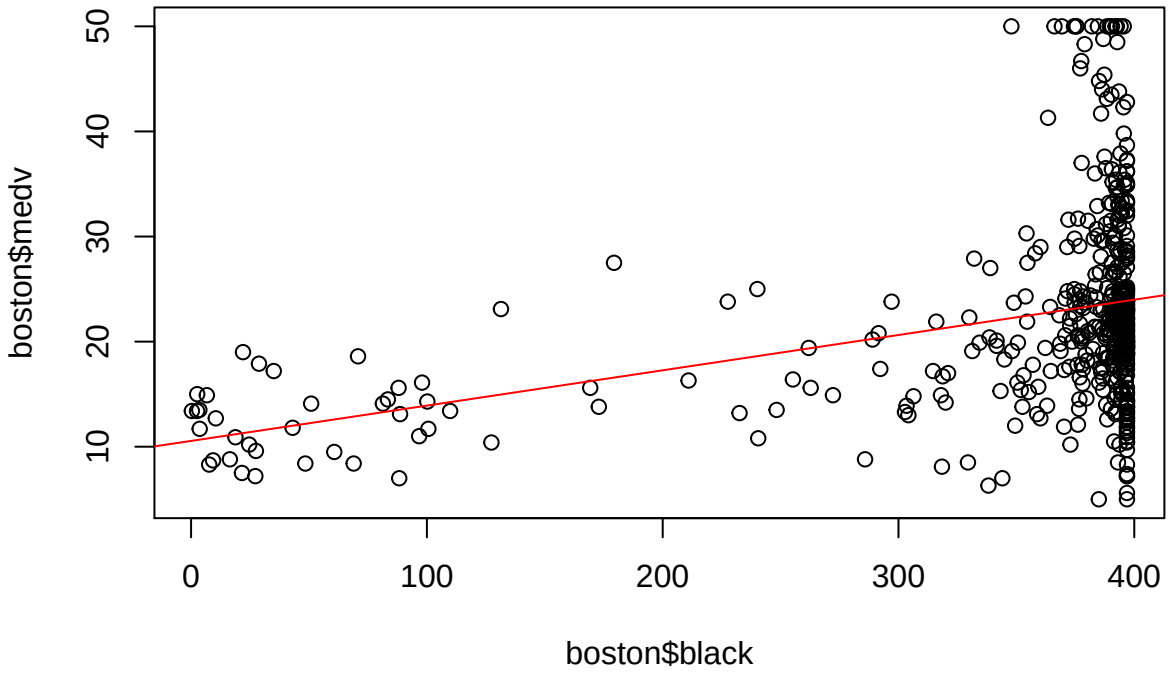
```
plot(boston$ptratio, boston$medv, main = "medv vs ptratio")
abline(m11,col = "red")
```




```
m12 <- lm(medv~black, data = boston)
summary(m12)
```

```
##
## Call:
## lm(formula = medv ~ black, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.884  -4.862  -1.684   2.932  27.763
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.551034   1.557463   6.775 3.49e-11 ***
## black        0.033593   0.004231   7.941 1.32e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.679 on 504 degrees of freedom
## Multiple R-squared:  0.1112, Adjusted R-squared:  0.1094
## F-statistic: 63.05 on 1 and 504 DF,  p-value: 1.318e-14
```

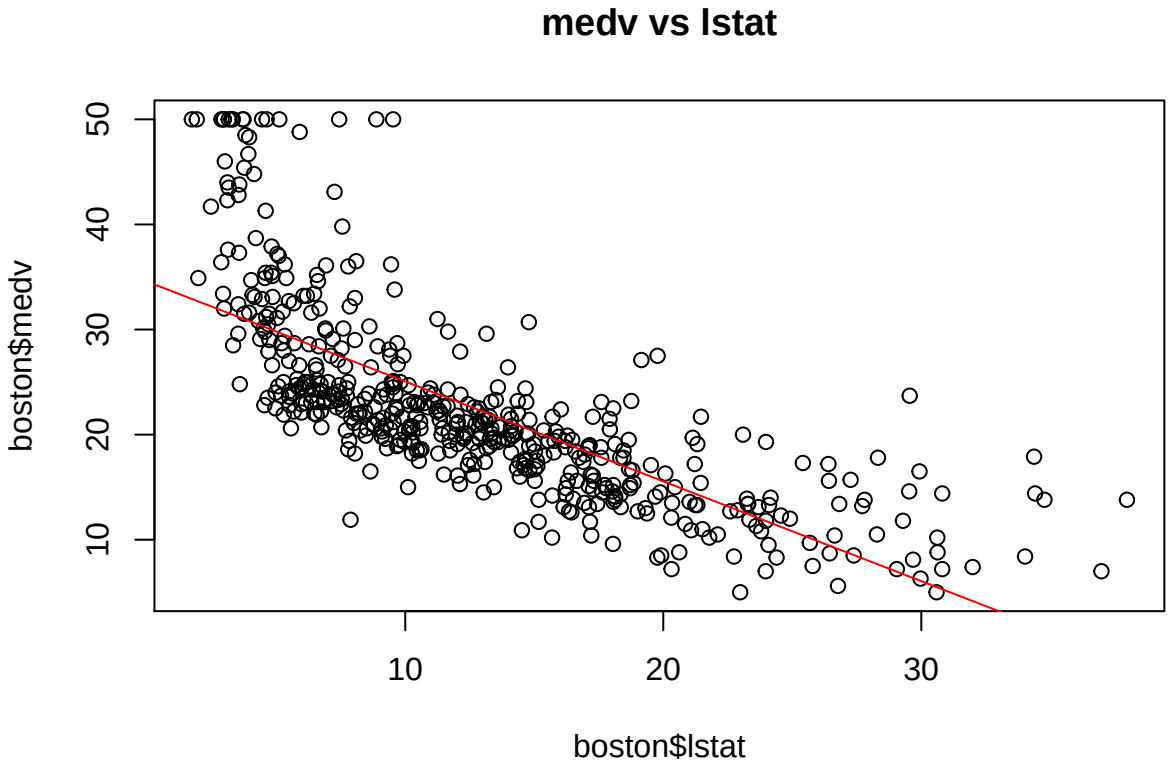
```
plot(boston$black, boston$medv, main = "medv vs black")
abline(m12,col = "red")
```




```
m13 <- lm(medv~lstat, data = boston)
summary(m13)
```

```
##
## Call:
## lm(formula = medv ~ lstat, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.168  -3.990  -1.318   2.034  24.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.55384    0.56263   61.41   <2e-16 ***
## lstat       -0.95005    0.03873  -24.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.216 on 504 degrees of freedom
## Multiple R-squared:  0.5441, Adjusted R-squared:  0.5432
## F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(boston$lstat, boston$medv, main = "medv vs lstat")
abline(m13,col = "red")
```



Comments: From models, we see that all predictors have p-value that is smaller than 0.5. Therefore, all predictors have statistically significant association with the response.

(ii)

```
multiple <- lm(medv~., data = boston)
summary(multiple)
```

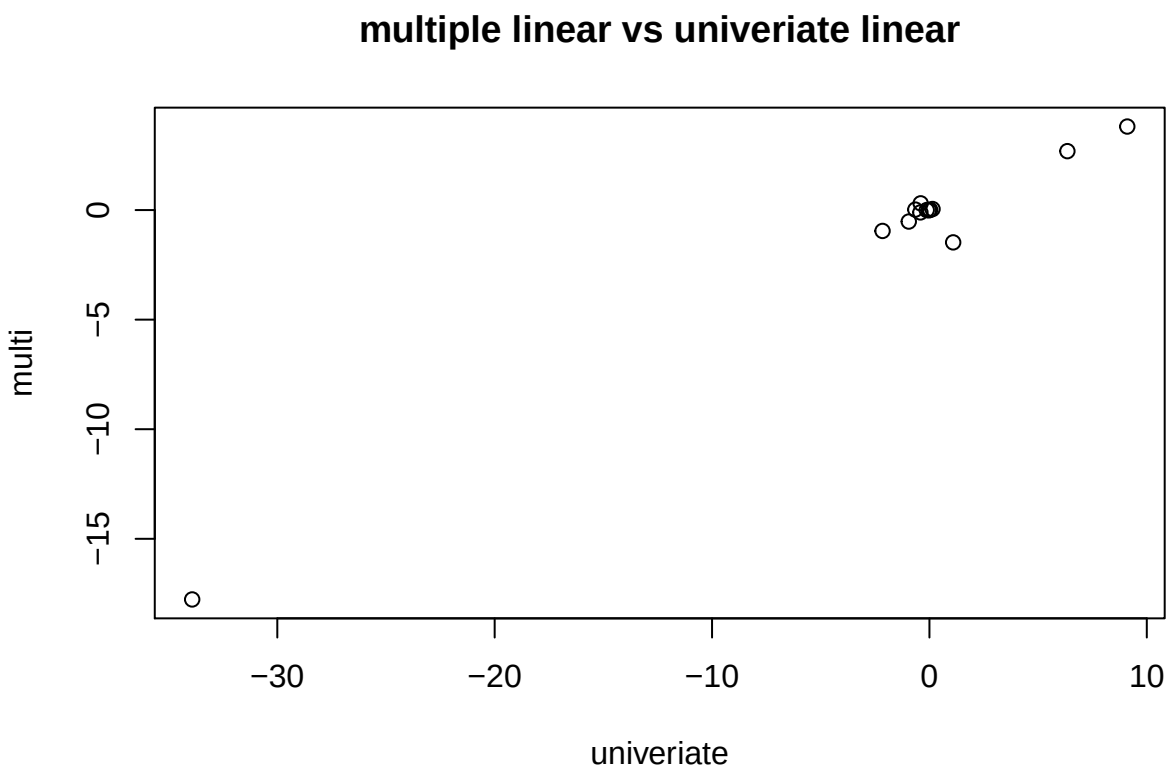
```
##
## Call:
## lm(formula = medv ~ ., data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.595  -2.730  -0.518   1.777  26.199
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.646e+01  5.103e+00   7.144 3.28e-12 ***
## crim        -1.080e-01  3.286e-02  -3.287 0.001087 **
## zn           4.642e-02  1.373e-02   3.382 0.000778 ***
## indus        2.056e-02  6.150e-02   0.334 0.738288
## chas         2.687e+00  8.616e-01   3.118 0.001925 **
## nox         -1.777e+01  3.820e+00  -4.651 4.25e-06 ***
## rm           3.810e+00  4.179e-01   9.116  < 2e-16 ***
## age          6.922e-04  1.321e-02   0.052 0.958229
## dis         -1.476e+00  1.995e-01  -7.398 6.01e-13 ***
## rad          3.060e-01  6.635e-02   4.613 5.07e-06 ***
```

```
## tax         -1.233e-02  3.760e-03  -3.280 0.001112 **
## ptratio     -9.527e-01  1.308e-01  -7.283 1.31e-12 ***
## black        9.312e-03  2.686e-03   3.467 0.000573 ***
## lstat       -5.248e-01  5.072e-02 -10.347  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.745 on 492 degrees of freedom
## Multiple R-squared:  0.7406, Adjusted R-squared:  0.7338
## F-statistic: 108.1 on 13 and 492 DF,  p-value: < 2.2e-16
```

Comments: It's not the best model to fit the data as we see from the summary (R-sqaured). From the model, we see that indus and age have p-value that are greater than 0.05. Therefore, for predictors other than indus and age we reject the null hypothesis as they are all significant.

(iii)

```
univeriate <- c(m1$coefficient[2], m2$coefficient[2], m3$coefficient[2],
                m4$coefficient[2], m5$coefficient[2], m6$coefficient[2],
                m7$coefficient[2], m8$coefficient[2], m9$coefficient[2],
                m10$coefficient[2], m11$coefficient[2], m12$coefficient[2],
                m13$coefficient[2])
multi <- c(multiple$coefficients[-1])
plot(univeriate, multi, main = "multiple linear vs univeriate linear")
```



Comments: Still we can not see a direct relationship between coefficents from two models. The reason is the same. For univeriate linear regression, we are fitting one predictor at one time. This means that the coefficent is the increasing of that particular predictor, with absence of other predictors. For multiple linear

regression, the slope is the increase of one predictor, while holding other predictors fixed. Since two types of regression have different interpretation of slopes, there is no relationship between coeffcients from two models.

(iv)

```
poly1 <- lm(medv~zn+I(zn^2)+I(zn^3), data = boston)
summary(poly1)
```

```
##
## Call:
## lm(formula = medv ~ zn + I(zn^2) + I(zn^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.449  -5.549  -1.049   3.225  29.551
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.4485972  0.4359536  46.905  < 2e-16 ***
## zn           0.6433652  0.1105611   5.819 1.06e-08 ***
## I(zn^2)     -0.0167646  0.0038872  -4.313 1.94e-05 ***
## I(zn^3)      0.0001257  0.0000316   3.978 7.98e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.43 on 502 degrees of freedom
## Multiple R-squared:  0.1649, Adjusted R-squared:  0.1599
## F-statistic: 33.05 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly2 <- lm(medv~indus+I(indus^2)+I(indus^3), data = boston)
summary(poly2)
```

```
##
## Call:
## lm(formula = medv ~ indus + I(indus^2) + I(indus^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.760  -4.725  -1.009   2.932  32.038
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.080160   1.663326  22.293  < 2e-16 ***
## indus       -2.806994   0.509349  -5.511 5.71e-08 ***
## I(indus^2)   0.140462   0.041554   3.380 0.000781 ***
## I(indus^3)  -0.002399   0.001011  -2.373 0.018026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.844 on 502 degrees of freedom
```

```
## Multiple R-squared:  0.2768, Adjusted R-squared:  0.2725
## F-statistic: 64.06 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly3 <- lm(medv~chas+I(chas^2)+I(chas^3), data = boston)
summary(poly3)
```

```
##
## Call:
## lm(formula = medv ~ chas + I(chas^2) + I(chas^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.094  -5.894  -1.417   2.856  27.906
##
## Coefficients: (2 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.0938     0.4176  52.902  < 2e-16 ***
## chas          6.3462     1.5880   3.996 7.39e-05 ***
## I(chas^2)         NA         NA      NA       NA
## I(chas^3)         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.064 on 504 degrees of freedom
## Multiple R-squared:  0.03072,    Adjusted R-squared:  0.02879
## F-statistic: 15.97 on 1 and 504 DF,  p-value: 7.391e-05
```

```
poly4 <- lm(medv~nox+I(nox^2)+I(nox^3), data = boston)
summary(poly4)
```

```
##
## Call:
## lm(formula = medv ~ nox + I(nox^2) + I(nox^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.104  -5.020  -2.144   2.747  32.416
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -22.49      38.52  -0.584   0.5596
## nox           315.10     195.10   1.615   0.1069
## I(nox^2)     -615.83     320.48  -1.922   0.0552 .
## I(nox^3)      350.19     170.92   2.049   0.0410 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.282 on 502 degrees of freedom
## Multiple R-squared:  0.1939, Adjusted R-squared:  0.189
## F-statistic: 40.24 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly5 <- lm(medv~rm+I(rm^2)+I(rm^3), data = boston)
summary(poly5)
```

```
##
## Call:
## lm(formula = medv ~ rm + I(rm^2) + I(rm^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.102  -2.674   0.569   3.011  35.911
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  241.3108    47.3275   5.099 4.85e-07 ***
## rm          -109.3906    22.9690  -4.763 2.51e-06 ***
## I(rm^2)       16.4910     3.6750   4.487 8.95e-06 ***
## I(rm^3)       -0.7404     0.1935  -3.827 0.000146 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.11 on 502 degrees of freedom
## Multiple R-squared:  0.5612, Adjusted R-squared:  0.5586
## F-statistic:   214 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly6 <- lm(medv~age+I(age^2)+I(age^3), data = boston)
summary(poly6)
```

```
##
## Call:
## lm(formula = medv ~ age + I(age^2) + I(age^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.443  -4.909  -2.234   2.185  32.944
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.893e+01  2.992e+00   9.668   <2e-16 ***
## age         -1.224e-01  2.014e-01  -0.608    0.544
## I(age^2)     2.355e-03  3.930e-03   0.599    0.549
## I(age^3)    -2.318e-05  2.279e-05  -1.017    0.310
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.472 on 502 degrees of freedom
## Multiple R-squared:  0.1566, Adjusted R-squared:  0.1515
## F-statistic: 31.06 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
poly7 <- lm(medv~dis+I(dis^2)+I(dis^3), data = boston)
summary(poly7)
```

```
##
```

```
## Call:
## lm(formula = medv ~ dis + I(dis^2) + I(dis^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.571  -5.242  -2.037   2.397  34.769
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.03789    2.91134   2.417  0.01599 *
## dis          8.59284    2.06633   4.158 3.77e-05 ***
## I(dis^2)    -1.24953    0.41235  -3.030  0.00257 **
## I(dis^3)     0.05602    0.02428   2.307  0.02146 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.727 on 502 degrees of freedom
## Multiple R-squared:  0.105,  Adjusted R-squared:  0.09968
## F-statistic: 19.64 on 3 and 502 DF,  p-value: 4.736e-12
```

```
poly8 <- lm(medv~rad+I(rad^2)+I(rad^3), data = boston)
summary(poly8)
```

```
##
## Call:
## lm(formula = medv ~ rad + I(rad^2) + I(rad^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.630  -5.151  -2.017   3.169  33.594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.251303   2.567860  11.781  < 2e-16 ***
## rad         -3.799454   1.307156  -2.907 0.003815 **
## I(rad^2)     0.616347   0.186057   3.313 0.000991 ***
## I(rad^3)    -0.020086   0.005717  -3.514 0.000482 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.37 on 502 degrees of freedom
## Multiple R-squared:  0.1767, Adjusted R-squared:  0.1718
## F-statistic: 35.91 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly9 <- lm(medv~tax+I(tax^2)+I(tax^3), data = boston)
summary(poly9)
```

```
##
## Call:
## lm(formula = medv ~ tax + I(tax^2) + I(tax^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -15.109  -4.952  -1.878   2.957  33.694
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.222e+01  1.397e+01   3.739 0.000206 ***
## tax         -1.635e-01  1.133e-01  -1.443 0.149646
## I(tax^2)     3.029e-04  2.872e-04   1.055 0.292004
## I(tax^3)    -2.079e-07  2.236e-07  -0.930 0.353061
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.115 on 502 degrees of freedom
## Multiple R-squared:  0.2261, Adjusted R-squared:  0.2215
## F-statistic: 48.89 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly10 <- lm(medv~ptratio+I(ptratio^2)+I(ptratio^3), data = boston)
summary(poly10)
```

```
##
## Call:
## lm(formula = medv ~ ptratio + I(ptratio^2) + I(ptratio^3), data = boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.7795  -5.0364  -0.9778   3.4766  31.1636
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  312.28642  152.48693   2.048   0.0411 *
## ptratio      -48.69114   26.88441  -1.811   0.0707 .
## I(ptratio^2)   2.83995    1.56413   1.816   0.0700 .
## I(ptratio^3)  -0.05686    0.03005  -1.892   0.0590 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.898 on 502 degrees of freedom
## Multiple R-squared:  0.2669, Adjusted R-squared:  0.2625
## F-statistic: 60.91 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly11 <- lm(medv~black+I(black^2)+I(black^3), data = boston)
summary(poly11)
```

```
##
## Call:
## lm(formula = medv ~ black + I(black^2) + I(black^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.005  -4.802  -1.613   2.852  28.051
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.260e+01  2.517e+00   5.006  7.7e-07 ***
```

```
## black       -1.703e-02  6.150e-02  -0.277    0.782
## I(black^2)   2.036e-04  3.258e-04   0.625    0.532
## I(black^3)  -2.224e-07  4.765e-07  -0.467    0.641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.685 on 502 degrees of freedom
## Multiple R-squared:  0.1135, Adjusted R-squared:  0.1082
## F-statistic: 21.43 on 3 and 502 DF,  p-value: 4.463e-13
```

```
poly12 <- lm(medv~lstat+I(lstat^2)+I(lstat^3), data = boston)
summary(poly12)
```

```
##
## Call:
## lm(formula = medv ~ lstat + I(lstat^2) + I(lstat^3), data = boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.5441  -3.7122  -0.5145   2.4846  26.4153
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.6496253  1.4347240  33.909  < 2e-16 ***
## lstat       -3.8655928  0.3287861 -11.757  < 2e-16 ***
## I(lstat^2)   0.1487385  0.0212987   6.983 9.18e-12 ***
## I(lstat^3)  -0.0020039  0.0003997  -5.013 7.43e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.396 on 502 degrees of freedom
## Multiple R-squared:  0.6578, Adjusted R-squared:  0.6558
## F-statistic: 321.7 on 3 and 502 DF,  p-value: < 2.2e-16
```

```
poly13 <- lm(medv~crim+I(crim^2)+I(crim^3), data = boston)
summary(poly13)
```

```
##
## Call:
## lm(formula = medv ~ crim + I(crim^2) + I(crim^3), data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.983  -4.975  -1.940   2.881  33.391
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.519e+01  4.355e-01  57.846  < 2e-16 ***
## crim        -1.136e+00  1.444e-01  -7.868 2.24e-14 ***
## I(crim^2)    2.378e-02  6.808e-03   3.494 0.000518 ***
## I(crim^3)   -1.489e-04  6.641e-05  -2.242 0.025411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 8.159 on 502 degrees of freedom
## Multiple R-squared:  0.2177, Adjusted R-squared:  0.213
## F-statistic: 46.57 on 3 and 502 DF,  p-value: < 2.2e-16
```

Comments: Quadratic or cubic terms of predictors zn, indus, dis, rad, lastat, crim, nox has p-value smaller than 0.05, this means that they are likely to have non-linear relationship between them and the response. For other predictors, they are not likely to have a non-linear relationship between them and the response.