R-code-Session02.R

duter

2021-02-11

## class 02: linear regression & model selection  
  
getwd()

## [1] "D:/duter/Documents/JHU MSIS/Spring 2021/Data Analytics/Data\_Analytics/HW2"

dir()

## [1] "Advertising.csv" "FuelEfficiency.csv"   
## [3] "Homework02\_Spr21\_DA.docx" "R-code-Session02.docx"   
## [5] "R-code-Session02.R" "R-code-Session02.spin.R"   
## [7] "R-code-Session02.spin.Rmd"

###################################  
#### advertising example  
##################################  
my.ad <- read.csv("Advertising.csv", stringsAsFactors = T)  
attach(my.ad)  
  
names(my.ad)

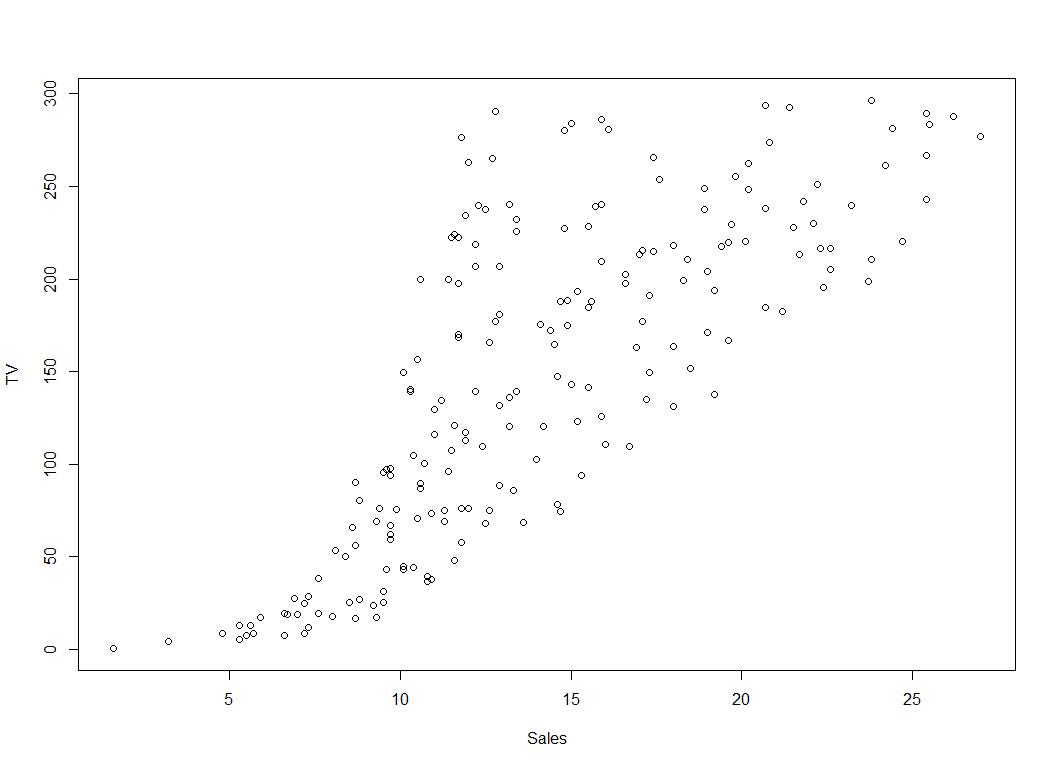
## [1] "X" "TV" "Radio" "Newspaper" "Sales"

summary(my.ad)

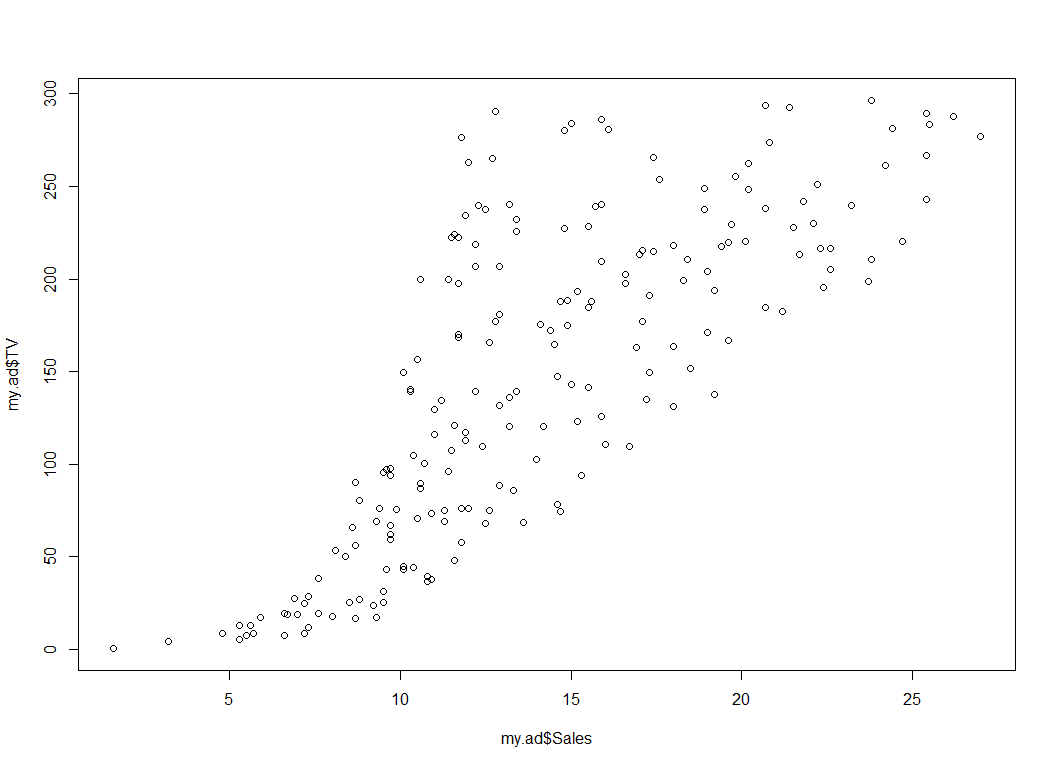
## X TV Radio Newspaper   
## Min. : 1.00 Min. : 0.70 Min. : 0.000 Min. : 0.30   
## 1st Qu.: 50.75 1st Qu.: 74.38 1st Qu.: 9.975 1st Qu.: 12.75   
## Median :100.50 Median :149.75 Median :22.900 Median : 25.75   
## Mean :100.50 Mean :147.04 Mean :23.264 Mean : 30.55   
## 3rd Qu.:150.25 3rd Qu.:218.82 3rd Qu.:36.525 3rd Qu.: 45.10   
## Max. :200.00 Max. :296.40 Max. :49.600 Max. :114.00   
## Sales   
## Min. : 1.60   
## 1st Qu.:10.38   
## Median :12.90   
## Mean :14.02   
## 3rd Qu.:17.40   
## Max. :27.00

par(mfrow = c(1, 1)) # par(mfrow=c(x,y)) divides plot space into x rows and y colums

plot(Sales, TV)



plot(my.ad$Sales, my.ad$TV)



plot(Sales ~ TV)  
  
my.first.lr = lm(Sales ~ TV)  
my.first.lr

##   
## Call:  
## lm(formula = Sales ~ TV)  
##   
## Coefficients:  
## (Intercept) TV   
## 7.03259 0.04754

summary(my.first.lr)

##   
## Call:  
## lm(formula = Sales ~ TV)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.3860 -1.9545 -0.1913 2.0671 7.2124   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.032594 0.457843 15.36 <2e-16 \*\*\*  
## TV 0.047537 0.002691 17.67 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.259 on 198 degrees of freedom  
## Multiple R-squared: 0.6119, Adjusted R-squared: 0.6099   
## F-statistic: 312.1 on 1 and 198 DF, p-value: < 2.2e-16

names(my.first.lr)

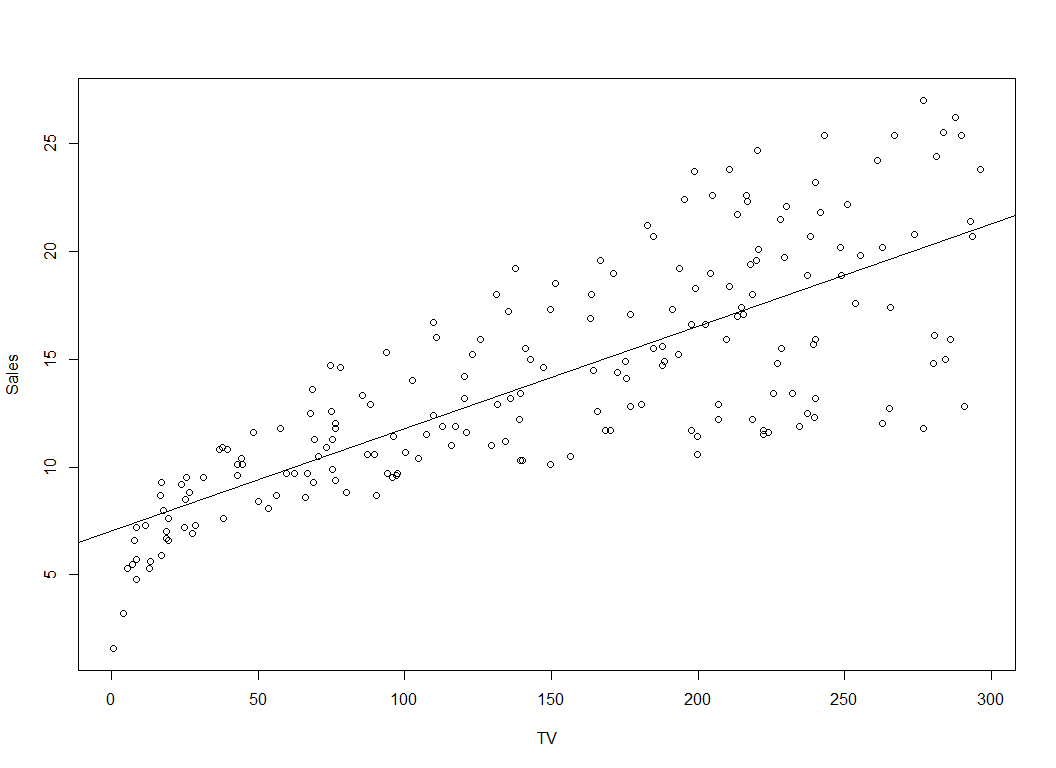
## [1] "coefficients" "residuals" "effects" "rank"   
## [5] "fitted.values" "assign" "qr" "df.residual"   
## [9] "xlevels" "call" "terms" "model"

abline(my.first.lr)  
my.first.lr$residuals

## 1 2 3 4 5 6   
## 4.12922549 1.25202595 1.44977624 4.26560543 -2.72721814 -0.24616232   
## 7 8 9 10 11 12   
## 2.03404963 0.45350227 -2.64140866 -5.93041431 -1.57476548 0.16128975   
## 13 14 15 16 17 18   
## 1.03603441 -1.96741599 2.26517814 6.07874691 2.24442223 3.99059583   
## 19 20 21 22 23 24   
## 0.97787093 0.56525932 0.58540418 -5.81779199 -2.06007720 -2.38520856   
## 25 26 27 28 29 30   
## -0.29412625 -7.52997632 1.17442053 -2.54614092 0.04029031 0.11131964   
## 31 32 33 34 35 36   
## 0.44392447 -0.49948025 -2.05315500 -2.25832525 -2.08185004 -8.05149492   
## 37 38 39 40 41 42   
## 5.67987712 4.11641941 1.01857725 3.62905243 -0.05876324 1.65342109   
## 43 44 45 46 47 48   
## -0.28935118 -3.96792445 0.27423678 -0.45625929 -0.69663020 4.76336641   
## 49 50 51 52 53 54   
## -3.03291826 -0.51279479 -5.13041431 -1.10527225 5.28047746 5.48721591   
## 55 56 57 58 59 60   
## 0.67953101 7.21236867 -1.87961102 -0.30708398 6.74668265 1.35143631   
## 61 62 63 64 65 66   
## -1.47580381 4.74608231 -2.70811160 2.08539348 4.73535289 -1.01262174   
## 67 68 69 70 71 72   
## 0.97000228 -0.25444756 0.58220801 4.96146280 1.80286134 0.14788333   
## 73 74 75 76 77 78   
## 0.49342449 -2.18383482 -0.17691262 0.86403723 -1.43985116 1.43924128   
## 79 80 81 82 83 84   
## -1.98929141 -1.54684384 1.13560712 -6.13187992 0.68789743 3.31590025   
## 85 86 87 88 89 90   
## 4.51833372 -1.01667248 1.34036079 3.70510035 1.66992110 4.44788333   
## 91 92 93 94 95 96   
## -2.21676436 -1.09214147 2.01867983 3.24046337 -0.63802873 2.10467307   
## 97 98 99 100 101 102   
## -4.72583370 -0.32211837 4.59604172 3.74045266 -5.90474238 2.67754623   
## 103 104 105 106 107 108   
## -5.55236020 -1.26472829 2.34417870 5.61210374 -1.02100956 -2.62990584   
## 109 110 111 112 113 114   
## -2.35532354 0.62654848 -4.36636696 3.27780046 -1.28478127 -1.09627338   
## 115 116 117 118 119 120   
## 3.85004117 1.99740475 -1.44969390 -1.26439288 2.89205075 -1.35480437   
## 121 122 123 124 125 126   
## 1.75047916 -0.92628239 -6.08080101 2.31564601 1.75774747 -0.57778859   
## 127 128 129 130 131 132   
## -0.80337934 -2.04503211 7.19508456 -0.16577732 -5.46586920 -6.93931059   
## 133 134 135 136 137 138   
## -1.73190133 2.11885288 2.01330442 2.27138672 1.25046846 0.75662796   
## 139 140 141 142 143 144   
## 0.52333091 4.87788163 0.37821704 2.95955920 2.58557724 -1.60492614   
## 145 146 147 148 149 150   
## -0.20561836 -3.40198420 -5.24614092 6.80649550 2.06101411 0.94251862   
## 151 152 153 154 155 156   
## -4.27612852 -1.18452704 0.17416630 3.82437994 -0.35997462 -4.02749377   
## 157 158 159 160 161 162   
## 3.80371591 -4.05358229 -0.28877224 -0.39316909 -0.83266402 2.19351637   
## 163 164 165 166 167 168   
## -1.08849661 3.19516574 -0.70388781 -6.27993573 0.11650059 -4.66317079   
## 169 170 171 172 173 174   
## -0.17198590 -5.54726042 -1.00942557 -0.35237090 -0.36431170 -3.33776380   
## 175 176 177 178 179 180   
## -6.10474238 6.80451071 1.35930497 -3.42332975 -8.38598196 -2.30466120   
## 181 182 183 184 185 186   
## -3.97683144 -5.21934948 -1.00415274 5.49586866 -1.49739289 5.82239516   
## 187 188 189 190 191 192   
## -3.36395489 1.18315446 -4.72807271 -1.22152873 1.88970915 -0.72160990   
## 193 194 195 196 197 198   
## -1.95022376 4.63829483 3.15117138 -1.24849321 -1.81054508 -2.64657891   
## 199 200   
## 4.98601522 -4.66584779

require(psych)

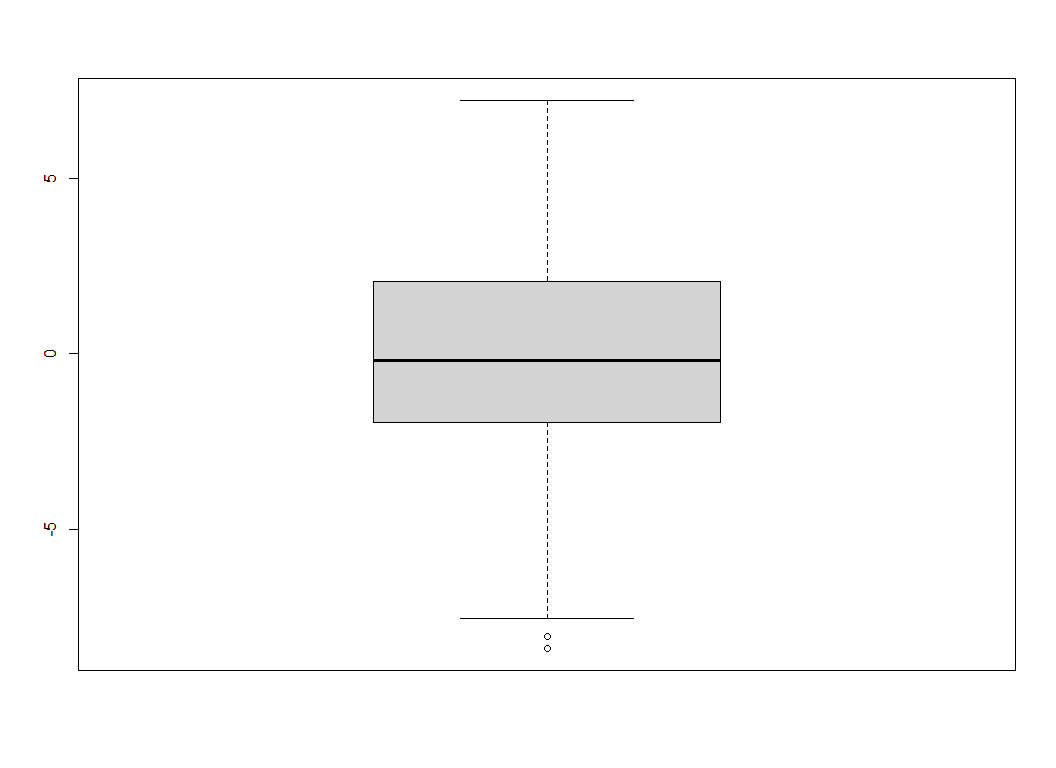
## Loading required package: psych



describe(my.first.lr$residuals)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 200 0 3.25 -0.19 0.04 2.96 -8.39 7.21 15.6 -0.09 -0.25 0.23

boxplot(my.first.lr$residuals)

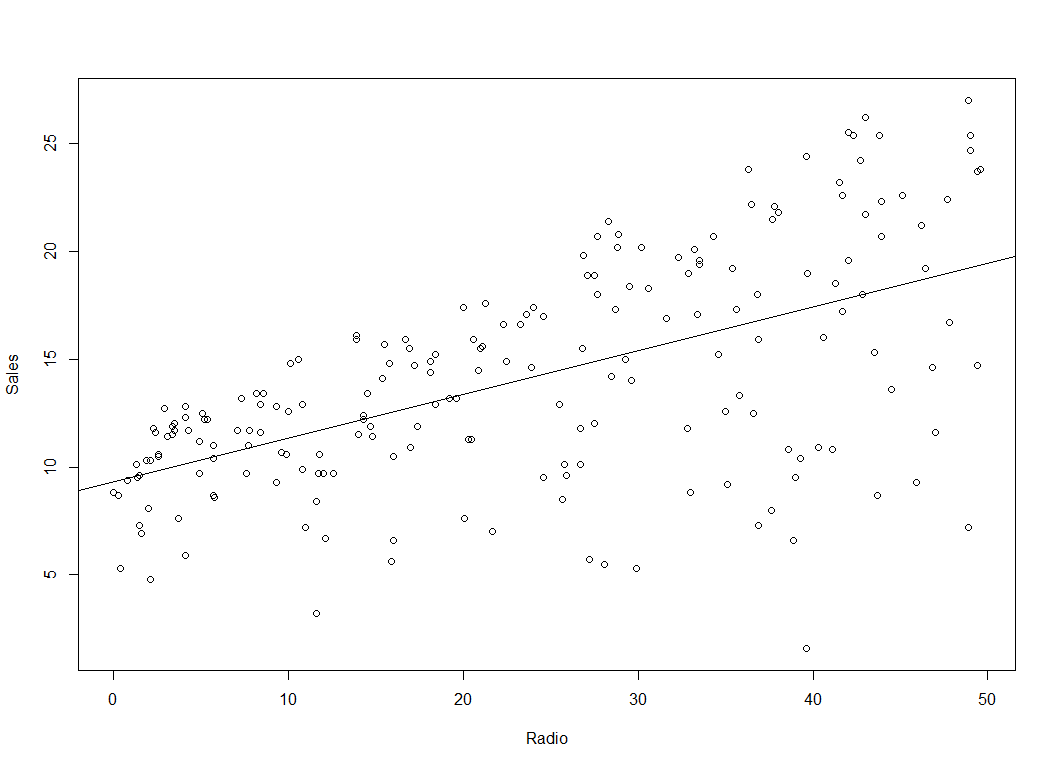


### regress Sales w.r.t. Radio

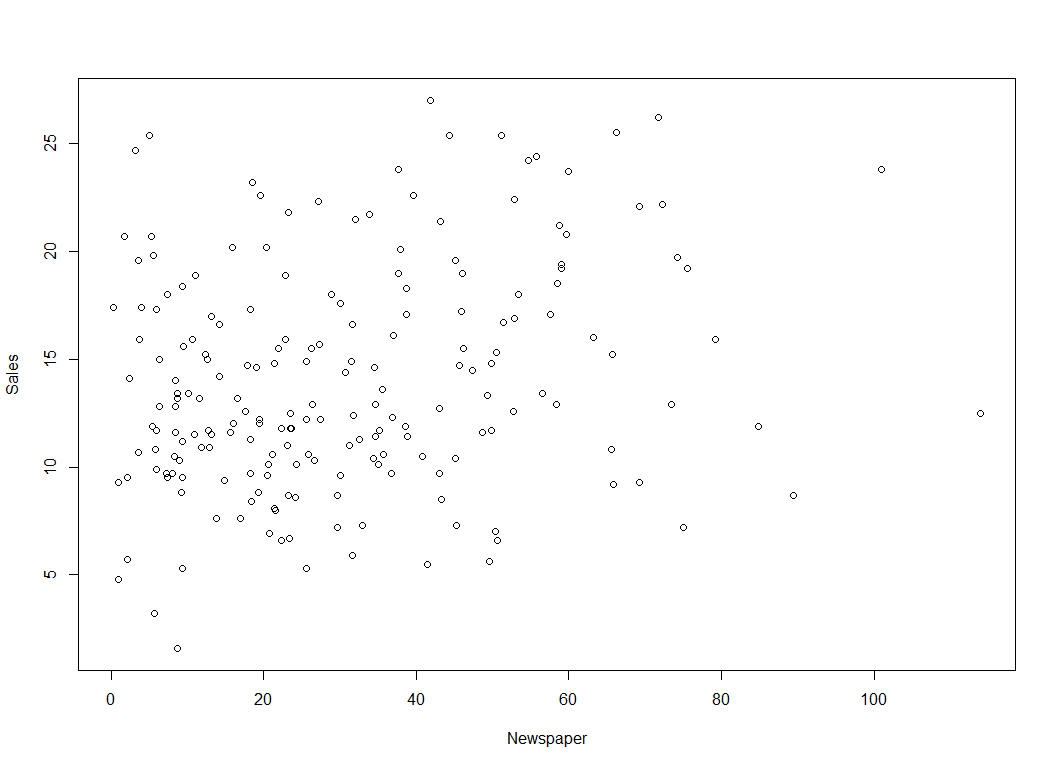
plot(Sales ~ Radio)  
my.lm.2 = lm(Sales ~ Radio)  
summary(my.lm.2)

##   
## Call:  
## lm(formula = Sales ~ Radio)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.7305 -2.1324 0.7707 2.7775 8.1810   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.31164 0.56290 16.542 <2e-16 \*\*\*  
## Radio 0.20250 0.02041 9.921 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.275 on 198 degrees of freedom  
## Multiple R-squared: 0.332, Adjusted R-squared: 0.3287   
## F-statistic: 98.42 on 1 and 198 DF, p-value: < 2.2e-16

abline(my.lm.2)



plot(Sales ~ Newspaper)



my.lm.3 = lm(Sales ~ Newspaper)  
summary(my.lm.3)

##   
## Call:  
## lm(formula = Sales ~ Newspaper)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.2272 -3.3873 -0.8392 3.5059 12.7751   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12.35141 0.62142 19.88 < 2e-16 \*\*\*  
## Newspaper 0.05469 0.01658 3.30 0.00115 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.092 on 198 degrees of freedom  
## Multiple R-squared: 0.05212, Adjusted R-squared: 0.04733   
## F-statistic: 10.89 on 1 and 198 DF, p-value: 0.001148

my.lm.4 = lm(Sales ~ TV + Radio + Newspaper)  
summary(my.lm.4)

##   
## Call:  
## lm(formula = Sales ~ TV + Radio + Newspaper)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.8277 -0.8908 0.2418 1.1893 2.8292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.938889 0.311908 9.422 <2e-16 \*\*\*  
## TV 0.045765 0.001395 32.809 <2e-16 \*\*\*  
## Radio 0.188530 0.008611 21.893 <2e-16 \*\*\*  
## Newspaper -0.001037 0.005871 -0.177 0.86   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.686 on 196 degrees of freedom  
## Multiple R-squared: 0.8972, Adjusted R-squared: 0.8956   
## F-statistic: 570.3 on 3 and 196 DF, p-value: < 2.2e-16

my.lm.4 = lm(Sales ~ ., data = my.ad)  
summary(my.lm.4)

##   
## Call:  
## lm(formula = Sales ~ ., data = my.ad)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.8105 -0.9008 0.2641 1.1783 2.8336   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.0052094 0.3942082 7.623 1.06e-12 \*\*\*  
## X -0.0005798 0.0020992 -0.276 0.783   
## TV 0.0457759 0.0013988 32.725 < 2e-16 \*\*\*  
## Radio 0.1883832 0.0086480 21.784 < 2e-16 \*\*\*  
## Newspaper -0.0012433 0.0059319 -0.210 0.834   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.689 on 195 degrees of freedom  
## Multiple R-squared: 0.8973, Adjusted R-squared: 0.8951   
## F-statistic: 425.7 on 4 and 195 DF, p-value: < 2.2e-16

my.lm.5 = lm(Sales ~ TV + Radio)  
summary(my.lm.5)

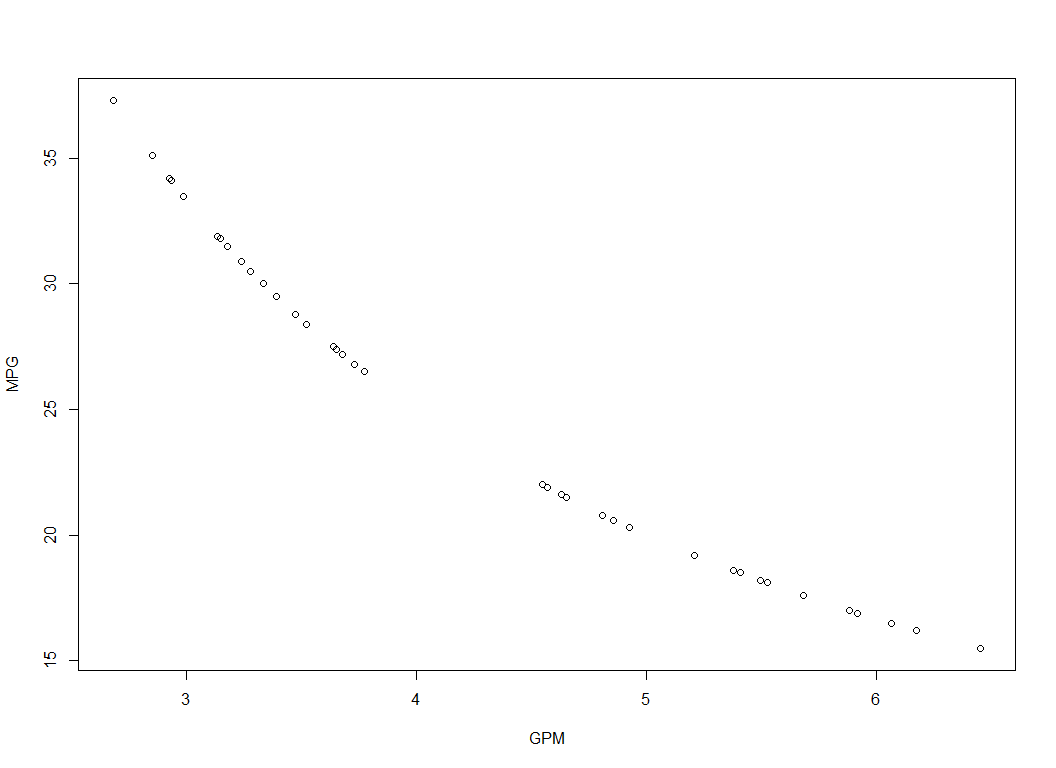
##   
## Call:  
## lm(formula = Sales ~ TV + Radio)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.7977 -0.8752 0.2422 1.1708 2.8328   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.92110 0.29449 9.919 <2e-16 \*\*\*  
## TV 0.04575 0.00139 32.909 <2e-16 \*\*\*  
## Radio 0.18799 0.00804 23.382 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.681 on 197 degrees of freedom  
## Multiple R-squared: 0.8972, Adjusted R-squared: 0.8962   
## F-statistic: 859.6 on 2 and 197 DF, p-value: < 2.2e-16

detach(my.ad)  
  
########################################  
##### Example Fuel efficiency  
########################################  
  
FuelEff <- read.csv("FuelEfficiency.csv", stringsAsFactors = T)  
head(FuelEff)

## MPG GPM WT DIS NC HP ACC ET  
## 1 16.9 5.917 4.360 350 8 155 14.9 1  
## 2 15.5 6.452 4.054 351 8 142 14.3 1  
## 3 19.2 5.208 3.605 267 8 125 15.0 1  
## 4 18.5 5.405 3.940 360 8 150 13.0 1  
## 5 30.0 3.333 2.155 98 4 68 16.5 0  
## 6 27.5 3.636 2.560 134 4 95 14.2 0

attach(FuelEff)

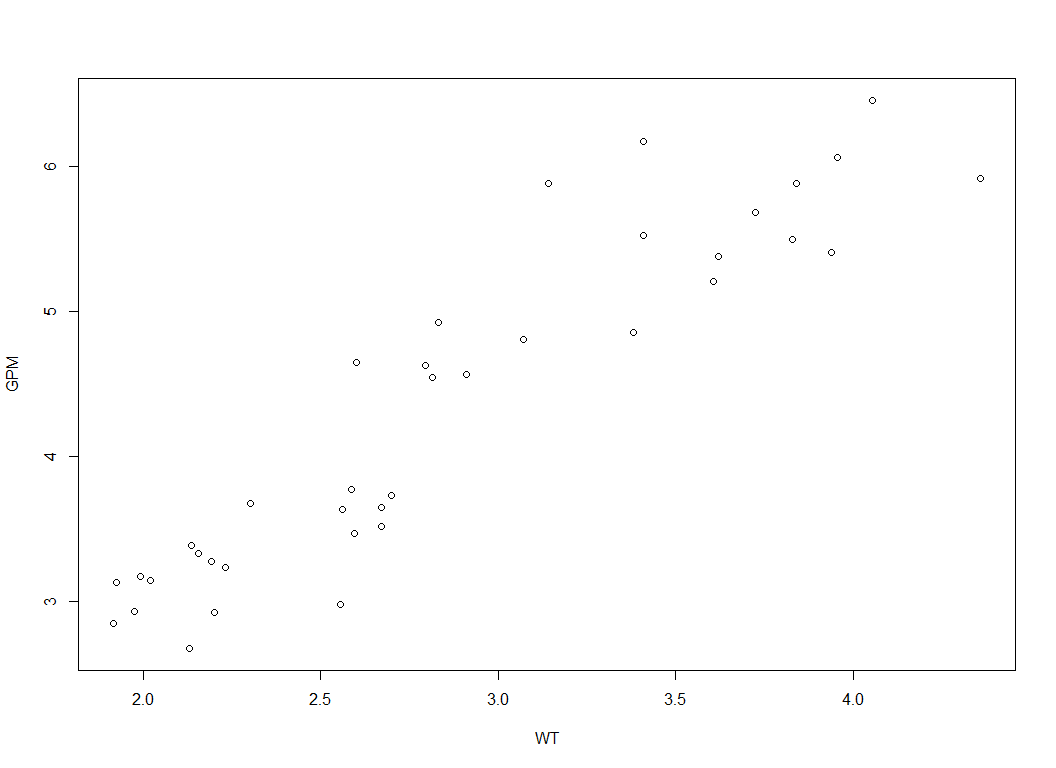
plot(MPG ~ GPM)



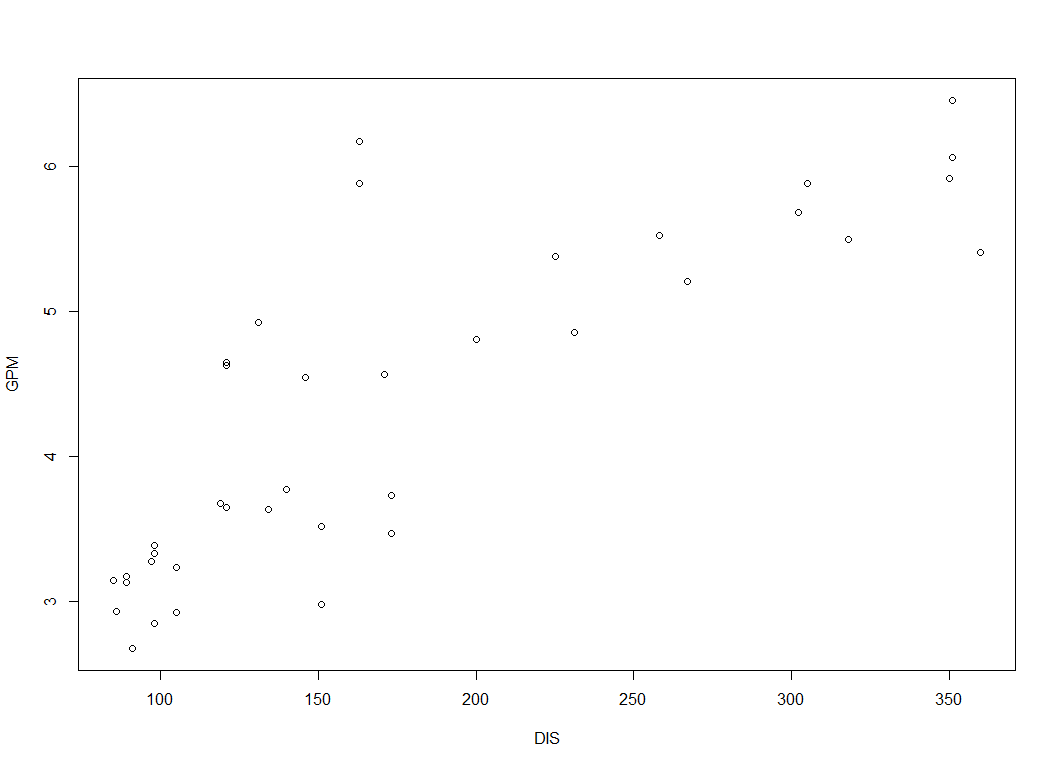
MPG \* GPM

## [1] 99.9973 100.0060 99.9936 99.9925 99.9900 99.9900 99.9872 99.9924  
## [9] 99.9978 99.9940 100.0080 100.0026 99.9924 100.0064 99.9936 100.0025  
## [17] 99.9940 100.0032 100.0065 100.0090 100.0110 99.9954 100.0153 99.9999  
## [25] 100.0100 100.0125 100.0050 99.9964 99.9936 99.9908 99.9975 100.0008  
## [33] 100.0110 100.0013 100.0095 99.9900 99.9965 100.0065

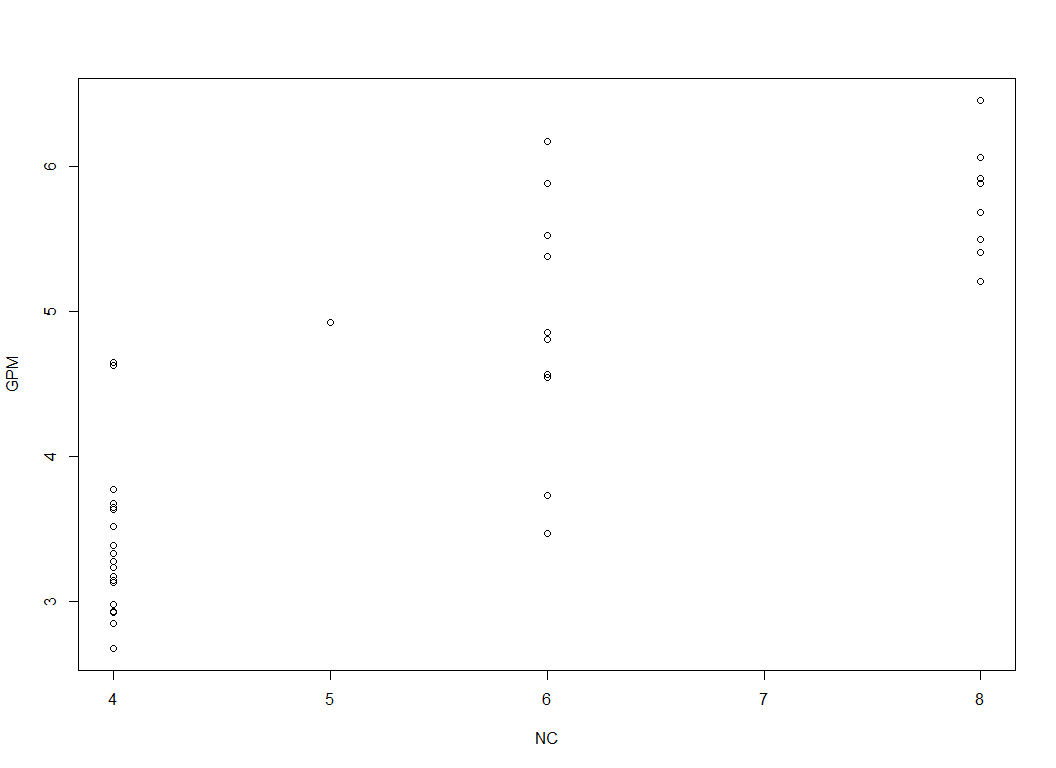
plot(GPM ~ WT, data = FuelEff)



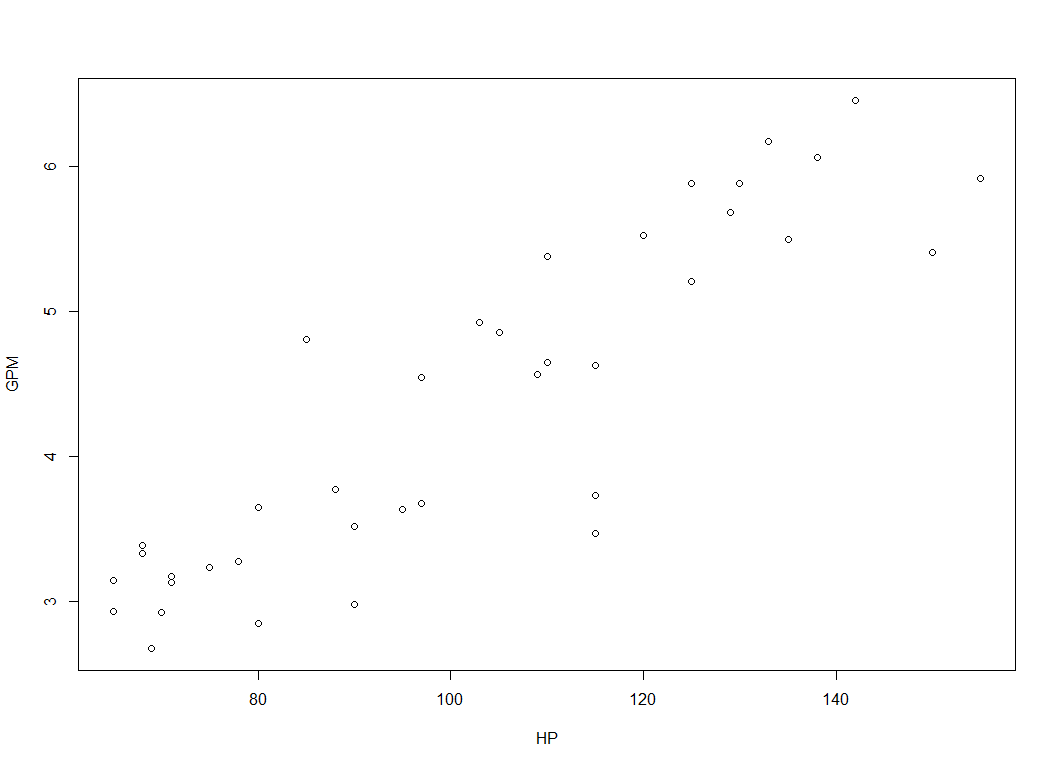
plot(GPM ~ DIS, data = FuelEff)



plot(GPM ~ NC, data = FuelEff)



plot(GPM ~ HP, data = FuelEff)



m0 = lm(GPM ~ WT)  
summary(m0)

##   
## Call:  
## lm(formula = GPM ~ WT)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.88072 -0.29041 0.00659 0.19021 1.13164   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.006101 0.302681 -0.02 0.984   
## WT 1.514798 0.102721 14.75 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4417 on 36 degrees of freedom  
## Multiple R-squared: 0.858, Adjusted R-squared: 0.854   
## F-statistic: 217.5 on 1 and 36 DF, p-value: < 2.2e-16

m1 = lm(GPM ~ ., data = FuelEff)  
summary(m1)

##   
## Call:  
## lm(formula = GPM ~ ., data = FuelEff)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.33622 -0.13499 -0.00486 0.08701 0.48325   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.9864020 1.1309678 4.409 0.000123 \*\*\*  
## MPG -0.1151678 0.0160110 -7.193 5.27e-08 \*\*\*  
## WT 0.0860549 0.2949136 0.292 0.772450   
## DIS -0.0003056 0.0017785 -0.172 0.864705   
## NC 0.2152198 0.0819830 2.625 0.013497 \*   
## HP 0.0076517 0.0047066 1.626 0.114471   
## ACC 0.0136109 0.0282887 0.481 0.633905   
## ET -0.4770695 0.1774648 -2.688 0.011608 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1928 on 30 degrees of freedom  
## Multiple R-squared: 0.9775, Adjusted R-squared: 0.9722   
## F-statistic: 185.8 on 7 and 30 DF, p-value: < 2.2e-16

FuelEff2 = FuelEff[-1]  
head(FuelEff2)

## GPM WT DIS NC HP ACC ET  
## 1 5.917 4.360 350 8 155 14.9 1  
## 2 6.452 4.054 351 8 142 14.3 1  
## 3 5.208 3.605 267 8 125 15.0 1  
## 4 5.405 3.940 360 8 150 13.0 1  
## 5 3.333 2.155 98 4 68 16.5 0  
## 6 3.636 2.560 134 4 95 14.2 0

m2 <- lm(GPM ~ ., data = FuelEff2)  
m2 <- lm(GPM ~ . - MPG, data = FuelEff)  
summary(m2)

##   
## Call:  
## lm(formula = GPM ~ . - MPG, data = FuelEff)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4996 -0.2547 0.0402 0.1956 0.6455   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.599357 0.663403 -3.918 0.000458 \*\*\*  
## WT 0.787768 0.451925 1.743 0.091222 .   
## DIS -0.004890 0.002696 -1.814 0.079408 .   
## NC 0.444157 0.122683 3.620 0.001036 \*\*   
## HP 0.023599 0.006742 3.500 0.001431 \*\*   
## ACC 0.068814 0.044213 1.556 0.129757   
## ET -0.959634 0.266785 -3.597 0.001104 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.313 on 31 degrees of freedom  
## Multiple R-squared: 0.9386, Adjusted R-squared: 0.9267   
## F-statistic: 78.94 on 6 and 31 DF, p-value: < 2.2e-16

m3 <- lm(GPM ~ WT + DIS, data = FuelEff)  
summary(m3)

##   
## Call:  
## lm(formula = GPM ~ WT + DIS, data = FuelEff)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.79941 -0.23818 0.06717 0.25695 0.76081   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.296128 0.467034 -2.775 0.00879 \*\*   
## WT 2.449718 0.292182 8.384 6.89e-10 \*\*\*  
## DIS -0.007821 0.002324 -3.365 0.00187 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3893 on 35 degrees of freedom  
## Multiple R-squared: 0.8927, Adjusted R-squared: 0.8866   
## F-statistic: 145.6 on 2 and 35 DF, p-value: < 2.2e-16

detach(FuelEff)  
  
####################################  
#### Boston Data Set  
####################################  
  
library(MASS)  
# library(ISLR)  
# utils:::menuInstallPkgs()  
attach(Boston)  
head(Boston)

## crim zn indus chas nox rm age dis rad tax ptratio black lstat  
## 1 0.00632 18 2.31 0 0.538 6.575 65.2 4.0900 1 296 15.3 396.90 4.98  
## 2 0.02731 0 7.07 0 0.469 6.421 78.9 4.9671 2 242 17.8 396.90 9.14  
## 3 0.02729 0 7.07 0 0.469 7.185 61.1 4.9671 2 242 17.8 392.83 4.03  
## 4 0.03237 0 2.18 0 0.458 6.998 45.8 6.0622 3 222 18.7 394.63 2.94  
## 5 0.06905 0 2.18 0 0.458 7.147 54.2 6.0622 3 222 18.7 396.90 5.33  
## 6 0.02985 0 2.18 0 0.458 6.430 58.7 6.0622 3 222 18.7 394.12 5.21  
## medv  
## 1 24.0  
## 2 21.6  
## 3 34.7  
## 4 33.4  
## 5 36.2  
## 6 28.7

? Boston

## starting httpd help server ... done

names(Boston)

## [1] "crim" "zn" "indus" "chas" "nox" "rm" "age"   
## [8] "dis" "rad" "tax" "ptratio" "black" "lstat" "medv"

lm.fit = lm(medv ~ lstat, data = Boston)  
summary(lm.fit)

##   
## 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

coef(lm.fit)

## (Intercept) lstat   
## 34.5538409 -0.9500494

# 95% confidence interval  
confint(lm.fit)

## 2.5 % 97.5 %  
## (Intercept) 33.448457 35.6592247  
## lstat -1.026148 -0.8739505

# 90% confidence interval  
confint(lm.fit, "lstat", level = .9)

## 5 % 95 %  
## lstat -1.013877 -0.8862212

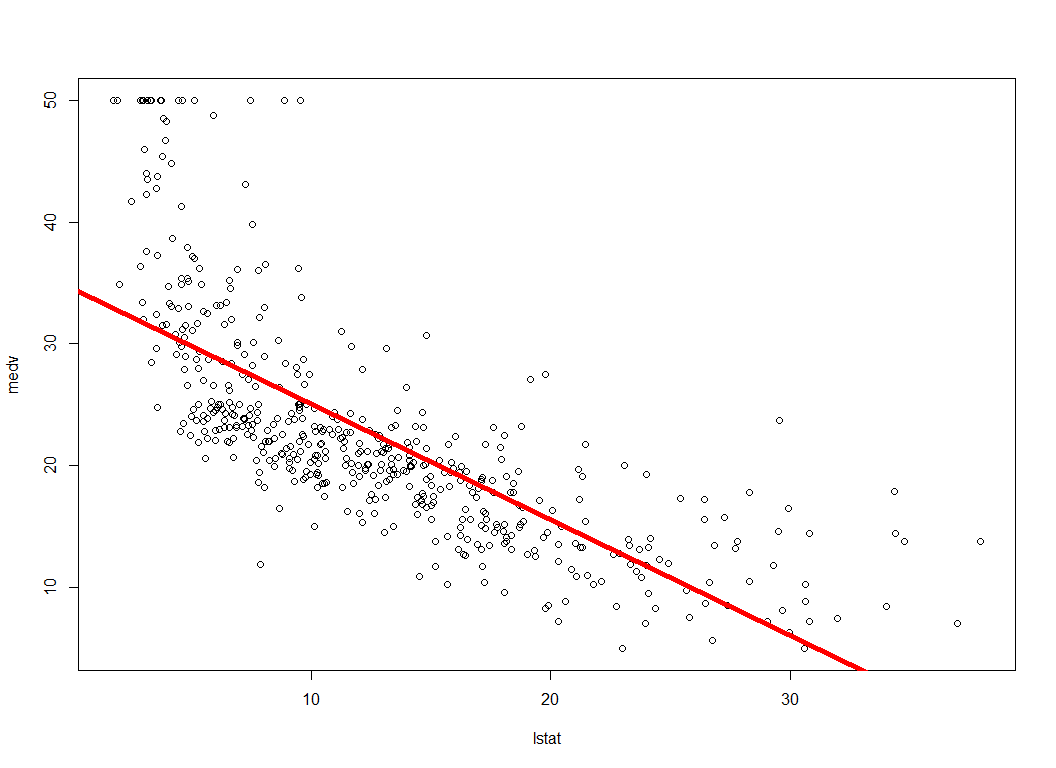
predict(lm.fit, data.frame(lstat = (c(5, 10, 15))), interval = "confidence")

## fit lwr upr  
## 1 29.80359 29.00741 30.59978  
## 2 25.05335 24.47413 25.63256  
## 3 20.30310 19.73159 20.87461

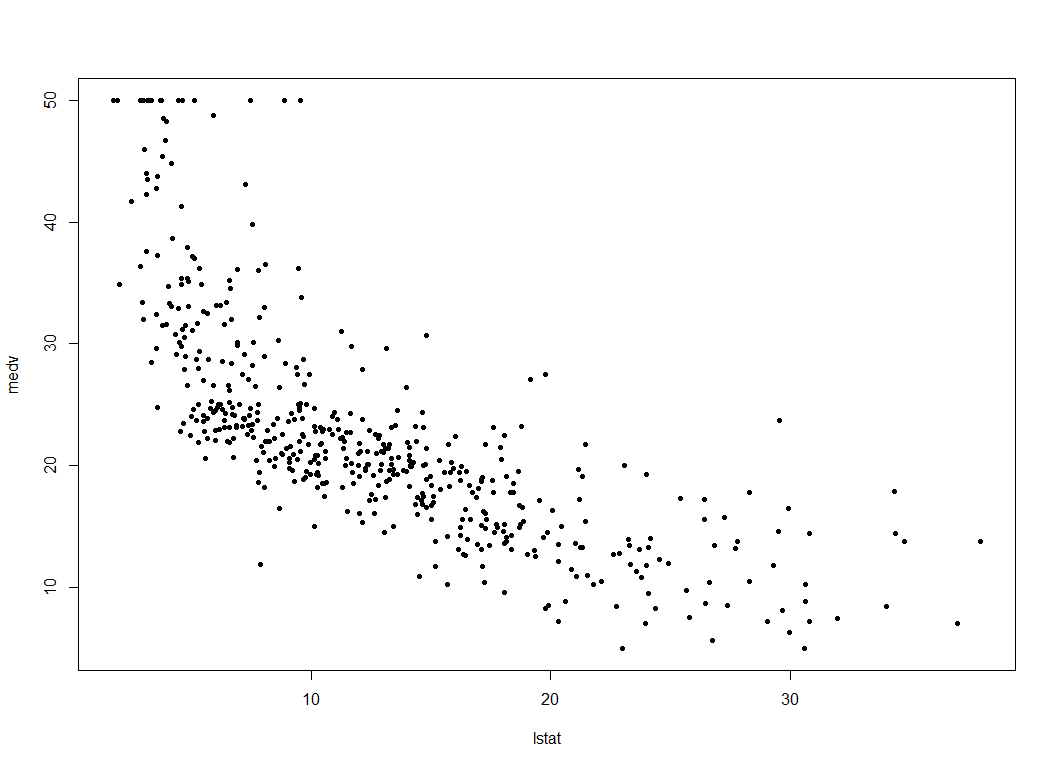
predict(lm.fit, data.frame(lstat = (c(5, 10, 15))), interval = "prediction")

## fit lwr upr  
## 1 29.80359 17.565675 42.04151  
## 2 25.05335 12.827626 37.27907  
## 3 20.30310 8.077742 32.52846

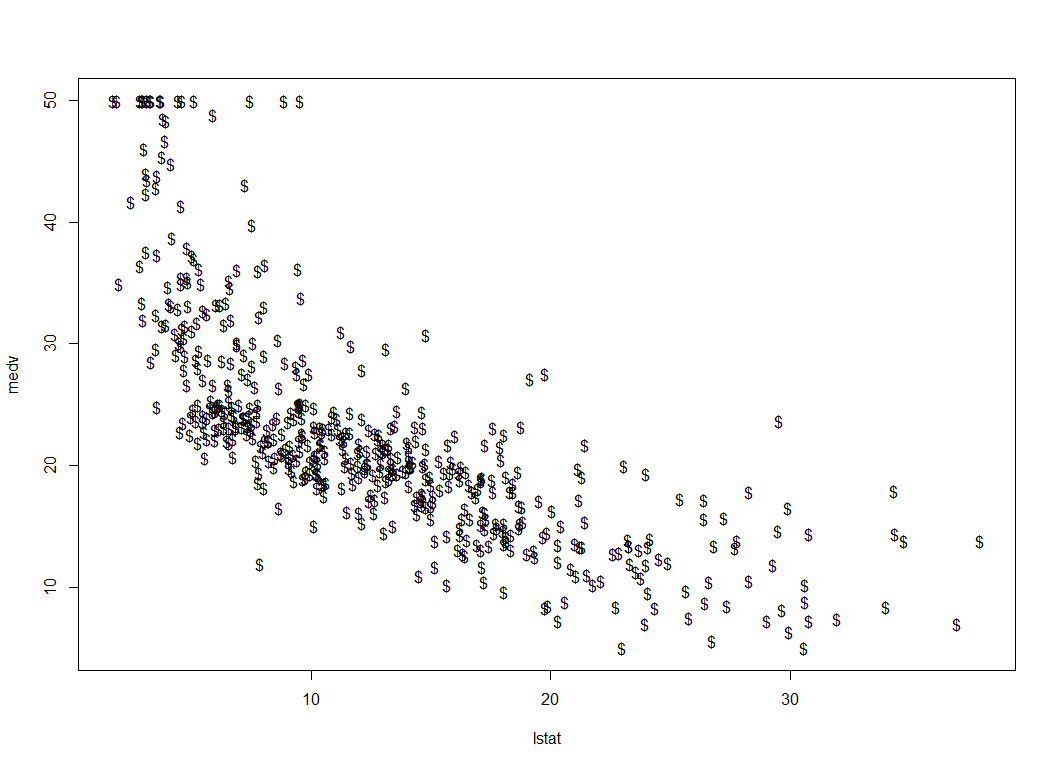
plot(medv ~ lstat, data = Boston)  
abline(lm.fit, lwd = 3)  
abline(lm.fit, lwd = 5)  
abline(lm.fit, lwd = 5, col = 2)  
abline(lm.fit, lwd = 5, col = 3)  
abline(lm.fit, lwd = 5, col = "red")



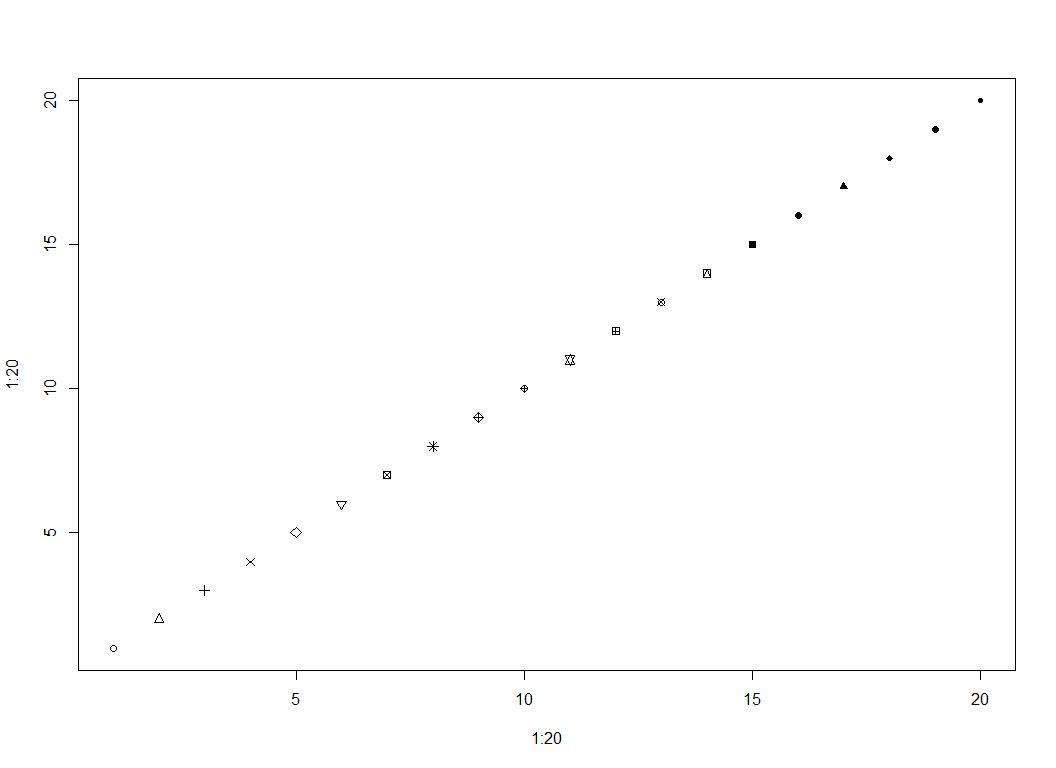
plot(medv ~ lstat, data = Boston, pch = 20)



plot(medv ~ lstat, data = Boston, pch = "$")



plot(1:20, 1:20, pch = 1:20)



lm.fit.1 = lm(medv ~ lstat + age, data = Boston)  
summary(lm.fit.1)

##   
## Call:  
## lm(formula = medv ~ lstat + age, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.981 -3.978 -1.283 1.968 23.158   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.22276 0.73085 45.458 < 2e-16 \*\*\*  
## lstat -1.03207 0.04819 -21.416 < 2e-16 \*\*\*  
## age 0.03454 0.01223 2.826 0.00491 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.173 on 503 degrees of freedom  
## Multiple R-squared: 0.5513, Adjusted R-squared: 0.5495   
## F-statistic: 309 on 2 and 503 DF, p-value: < 2.2e-16

lm.fit.1 = lm(medv ~ lstat + age + rm, data = Boston)  
summary(lm.fit.1)

##   
## Call:  
## lm(formula = medv ~ lstat + age + rm, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.210 -3.467 -1.053 1.957 27.500   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.175311 3.181924 -0.369 0.712   
## lstat -0.668513 0.054357 -12.298 <2e-16 \*\*\*  
## age 0.009091 0.011215 0.811 0.418   
## rm 5.019133 0.454306 11.048 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.542 on 502 degrees of freedom  
## Multiple R-squared: 0.639, Adjusted R-squared: 0.6369   
## F-statistic: 296.2 on 3 and 502 DF, p-value: < 2.2e-16

lm.fit.2 = lm(medv ~ ., data = Boston)  
summary(lm.fit.2)

##   
## 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

lm.fit.3 = lm(medv ~ . - age, data = Boston)  
summary(lm.fit.3)

##   
## Call:  
## lm(formula = medv ~ . - age, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.6054 -2.7313 -0.5188 1.7601 26.2243   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.436927 5.080119 7.172 2.72e-12 \*\*\*  
## crim -0.108006 0.032832 -3.290 0.001075 \*\*   
## zn 0.046334 0.013613 3.404 0.000719 \*\*\*  
## indus 0.020562 0.061433 0.335 0.737989   
## chas 2.689026 0.859598 3.128 0.001863 \*\*   
## nox -17.713540 3.679308 -4.814 1.97e-06 \*\*\*  
## rm 3.814394 0.408480 9.338 < 2e-16 \*\*\*  
## dis -1.478612 0.190611 -7.757 5.03e-14 \*\*\*  
## rad 0.305786 0.066089 4.627 4.75e-06 \*\*\*  
## tax -0.012329 0.003755 -3.283 0.001099 \*\*   
## ptratio -0.952211 0.130294 -7.308 1.10e-12 \*\*\*  
## black 0.009321 0.002678 3.481 0.000544 \*\*\*  
## lstat -0.523852 0.047625 -10.999 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.74 on 493 degrees of freedom  
## Multiple R-squared: 0.7406, Adjusted R-squared: 0.7343   
## F-statistic: 117.3 on 12 and 493 DF, p-value: < 2.2e-16

lm.fit.4 = lm(medv ~ . - age - indus, data = Boston)  
summary(lm.fit.4)

##   
## Call:  
## lm(formula = medv ~ . - age - indus, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.5984 -2.7386 -0.5046 1.7273 26.2373   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.341145 5.067492 7.171 2.73e-12 \*\*\*  
## crim -0.108413 0.032779 -3.307 0.001010 \*\*   
## zn 0.045845 0.013523 3.390 0.000754 \*\*\*  
## chas 2.718716 0.854240 3.183 0.001551 \*\*   
## nox -17.376023 3.535243 -4.915 1.21e-06 \*\*\*  
## rm 3.801579 0.406316 9.356 < 2e-16 \*\*\*  
## dis -1.492711 0.185731 -8.037 6.84e-15 \*\*\*  
## rad 0.299608 0.063402 4.726 3.00e-06 \*\*\*  
## tax -0.011778 0.003372 -3.493 0.000521 \*\*\*  
## ptratio -0.946525 0.129066 -7.334 9.24e-13 \*\*\*  
## black 0.009291 0.002674 3.475 0.000557 \*\*\*  
## lstat -0.522553 0.047424 -11.019 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.736 on 494 degrees of freedom  
## Multiple R-squared: 0.7406, Adjusted R-squared: 0.7348   
## F-statistic: 128.2 on 11 and 494 DF, p-value: < 2.2e-16

fit.lm.2 = lm(medv ~ lstat:age, data = Boston)  
summary(fit.lm.2)

##   
## Call:  
## lm(formula = medv ~ lstat:age, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.347 -4.372 -1.534 1.914 27.193   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.1588631 0.4828240 62.46 <2e-16 \*\*\*  
## lstat:age -0.0077146 0.0003799 -20.31 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.827 on 504 degrees of freedom  
## Multiple R-squared: 0.4501, Adjusted R-squared: 0.449   
## F-statistic: 412.4 on 1 and 504 DF, p-value: < 2.2e-16

fit.lm.3 = lm(medv ~ lstat \* age, data = Boston)  
summary(fit.lm.3)

##   
## Call:  
## lm(formula = medv ~ lstat \* age, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.806 -4.045 -1.333 2.085 27.552   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.0885359 1.4698355 24.553 < 2e-16 \*\*\*  
## lstat -1.3921168 0.1674555 -8.313 8.78e-16 \*\*\*  
## age -0.0007209 0.0198792 -0.036 0.9711   
## lstat:age 0.0041560 0.0018518 2.244 0.0252 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.149 on 502 degrees of freedom  
## Multiple R-squared: 0.5557, Adjusted R-squared: 0.5531   
## F-statistic: 209.3 on 3 and 502 DF, p-value: < 2.2e-16

x3 = lstat \* age  
x3 = Boston$lstat \* Boston$age  
x3

## [1] 324.696 721.146 246.233 134.652 288.886 305.827 827.838 1840.315  
## [9] 2993.000 1468.890 1928.435 1100.083 612.690 510.468 866.970 478.555  
## [17] 192.794 1198.539 427.854 783.960 2062.062 1233.636 1716.624 1988.000  
## [25] 1533.830 1414.907 1337.343 1534.464 1208.320 1045.854 2126.660 1304.000  
## [33] 2272.220 1743.250 1970.946 660.176 700.574 363.955 305.926 94.176  
## [41] 31.284 14.036 38.346 48.360 382.000 345.098 471.195 1607.400  
## [49] 2936.193 1004.400 614.665 594.090 111.408 180.402 704.480 105.339  
## [57] 205.989 159.975 200.312 435.184 870.530 1348.696 456.294 412.300  
## [65] 478.975 83.126 318.464 173.340 481.712 290.070 44.352 172.900  
## [73] 43.056 46.748 40.680 402.300 891.765 470.366 662.658 333.060  
## [81] 177.215 508.288 216.384 350.717 461.760 366.333 579.986 479.392  
## [89] 474.650 359.670 582.341 605.980 437.376 179.469 818.607 384.370  
## [97] 789.264 319.960 131.733 386.875 752.658 546.871 907.802 1174.656  
## [105] 1109.700 1592.649 1714.854 1200.468 1191.417 1418.160 707.200 829.056  
## [113] 1505.909 1630.386 879.890 1390.032 872.900 850.780 1123.547 887.372  
## [121] 1001.589 1200.107 1665.697 2464.770 1684.164 1309.204 2606.056 1650.240  
## [129] 1520.532 1736.798 1246.140 1197.802 1088.648 1433.862 1703.304 1665.472  
## [137] 1580.150 1435.656 2093.624 1807.234 2261.376 3441.000 2682.000 2642.000  
## [145] 2864.562 2780.000 1665.000 2826.021 2656.416 2035.605 1371.930 1328.000  
## [153] 1066.560 1555.315 1451.520 1240.652 1517.160 447.066 643.000 739.000  
## [161] 509.300 157.084 188.544 311.748 1068.552 912.330 355.940 961.488  
## [169] 1066.710 1077.664 1365.078 1170.519 1300.065 760.264 662.268 176.423  
## [177] 477.192 461.686 514.848 294.336 629.748 587.790 444.404 543.008  
## [185] 1255.404 904.720 238.520 274.548 132.696 209.671 109.650 144.452  
## [193] 75.481 49.797 82.344 95.040 139.128 315.126 253.546 69.768  
## [201] 61.855 285.312 48.827 126.492 91.872 242.401 575.925 1312.962  
## [209] 866.406 2309.000 1590.567 2124.628 862.414 302.974 289.590 401.528  
## [217] 756.560 824.619 1680.896 970.200 859.335 1959.298 771.561 614.080  
## [225] 324.162 384.290 270.745 508.164 66.640 80.464 793.365 403.725  
## [233] 181.051 278.080 535.325 669.120 729.810 338.668 117.660 311.014  
## [241] 617.934 807.240 593.538 40.482 956.250 1295.892 319.684 803.880  
## [249] 467.432 114.800 76.700 31.951 24.004 29.736 210.240 176.675  
## [257] 106.362 444.928 779.000 690.000 784.462 649.044 540.765 1063.125  
## [265] 741.960 656.260 1251.234 498.480 166.216 839.475 547.300 107.417  
## [273] 453.751 340.844 116.137 127.544 296.450 114.816 230.799 156.170  
## [281] 242.520 170.748 149.597 78.368 163.280 262.537 407.295 223.482  
## [289] 346.560 217.779 92.907 98.612 109.980 157.872 439.920 194.997  
## [297] 376.890 918.720 99.897 47.400 287.718 383.800 159.528 86.022  
## [305] 284.823 518.833 465.193 529.359 374.550 764.699 477.792 315.744  
## [313] 1059.488 654.120 810.144 893.550 1525.056 1142.898 696.192 748.524  
## [321] 376.560 373.041 384.230 872.282 245.412 74.676 177.735 558.923  
## [329] 257.226 126.248 292.698 353.012 182.439 216.408 259.875 276.345  
## [337] 453.740 629.376 317.423 442.196 543.465 270.657 516.405 404.952  
## [345] 129.541 510.705 662.641 176.172 177.903 203.205 265.512 197.091  
## [353] 144.115 162.450 176.295 108.615 1714.240 1207.570 957.432 1030.071  
## [361] 685.520 1292.709 980.278 1302.960 438.541 625.848 1279.600 1333.000  
## [369] 326.000 361.064 288.600 953.000 795.648 3477.000 3797.000 1315.776  
## [377] 2168.292 2098.512 2278.978 2178.000 1581.599 2089.028 2360.000 2456.000  
## [385] 2793.456 3022.461 2828.000 2863.105 3062.000 2062.065 1659.670 1547.700  
## [393] 2490.960 1404.742 1548.345 1691.456 1859.520 1970.088 3059.000 2331.666  
## [401] 2677.000 2032.000 2031.000 1897.920 2338.252 2298.000 2334.000 1213.000  
## [409] 2584.560 1978.000 1011.000 2122.000 3437.000 2008.000 3698.000 2905.000  
## [417] 2341.732 2373.624 2062.000 1739.610 1502.000 1496.210 1235.160 1981.979  
## [425] 1211.496 2326.806 936.693 1142.724 1680.712 2302.048 1518.804 1856.767  
## [433] 899.844 1425.738 1441.150 2201.342 1684.065 2645.000 2990.358 2148.432  
## [441] 2042.964 1897.344 1659.000 1885.000 2298.114 2273.304 1714.956 1588.104  
## [449] 1789.431 1898.173 1614.944 1741.086 1585.386 1662.282 1760.611 1568.245  
## [457] 1670.979 1360.282 1358.451 1240.680 1477.800 1295.060 1161.170 925.071  
## [465] 864.588 681.066 1452.605 2014.740 1287.230 836.892 1368.360 1167.309  
## [473] 1077.000 788.216 1730.556 2347.340 1748.448 2423.743 1743.501 1153.680  
## [481] 694.878 579.726 539.770 419.926 558.946 549.102 1195.404 609.140  
## [489] 1674.162 2356.251 2908.640 1785.316 1114.725 648.540 578.934 506.880  
## [497] 1541.106 995.460 843.676 1109.850 1142.101 668.197 696.436 513.240  
## [505] 578.664 636.704

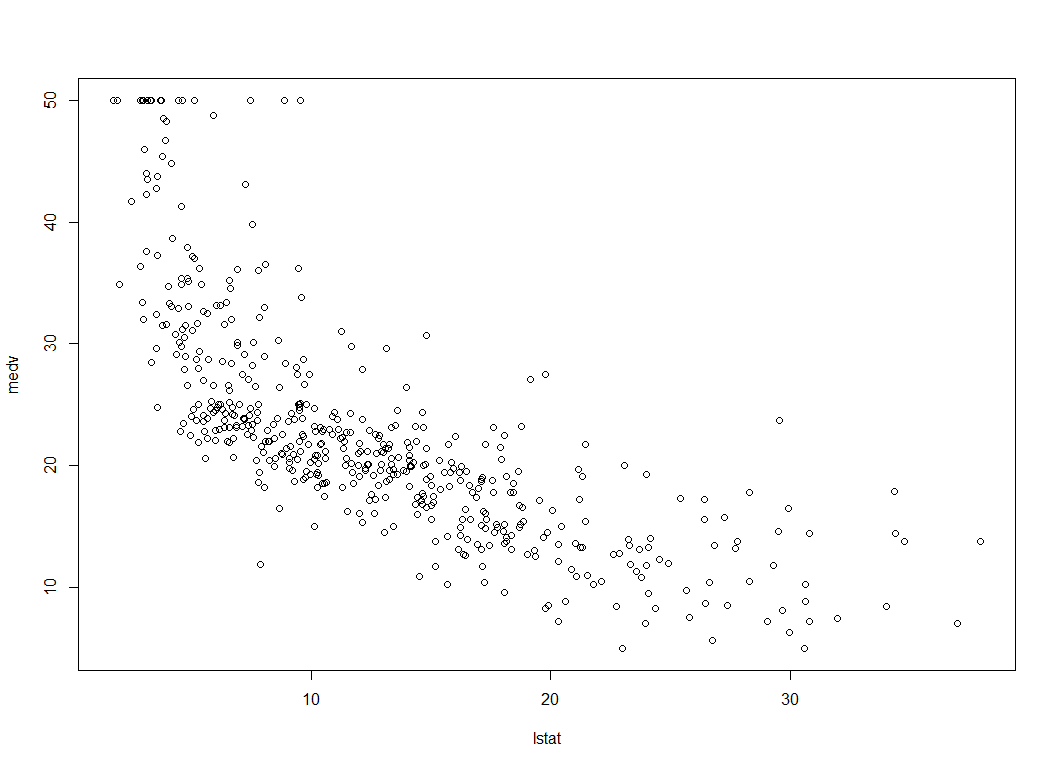
fit.lm.10 = lm(medv ~ x3, data = Boston)  
summary(fit.lm.10)

##   
## Call:  
## lm(formula = medv ~ x3, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.347 -4.372 -1.534 1.914 27.193   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.1588631 0.4828240 62.46 <2e-16 \*\*\*  
## x3 -0.0077146 0.0003799 -20.31 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.827 on 504 degrees of freedom  
## Multiple R-squared: 0.4501, Adjusted R-squared: 0.449   
## F-statistic: 412.4 on 1 and 504 DF, p-value: < 2.2e-16

fit.lm.11 = lm(medv ~ lstat + age + x3, data = Boston)  
summary(fit.lm.11)

##   
## Call:  
## lm(formula = medv ~ lstat + age + x3, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.806 -4.045 -1.333 2.085 27.552   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36.0885359 1.4698355 24.553 < 2e-16 \*\*\*  
## lstat -1.3921168 0.1674555 -8.313 8.78e-16 \*\*\*  
## age -0.0007209 0.0198792 -0.036 0.9711   
## x3 0.0041560 0.0018518 2.244 0.0252 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.149 on 502 degrees of freedom  
## Multiple R-squared: 0.5557, Adjusted R-squared: 0.5531   
## F-statistic: 209.3 on 3 and 502 DF, p-value: < 2.2e-16

plot(medv ~ lstat)



lm.fit4 = lm(medv ~ lstat + I(lstat ^ 2), data = Boston)  
summary(lm.fit4)

##   
## Call:  
## lm(formula = medv ~ lstat + I(lstat^2), data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.2834 -3.8313 -0.5295 2.3095 25.4148   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.862007 0.872084 49.15 <2e-16 \*\*\*  
## lstat -2.332821 0.123803 -18.84 <2e-16 \*\*\*  
## I(lstat^2) 0.043547 0.003745 11.63 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.524 on 503 degrees of freedom  
## Multiple R-squared: 0.6407, Adjusted R-squared: 0.6393   
## F-statistic: 448.5 on 2 and 503 DF, p-value: < 2.2e-16

x4 = Boston$lstat ^ 2  
lm.fit5 = lm(medv ~ lstat + x4, data = Boston)  
summary(lm.fit5)

##   
## Call:  
## lm(formula = medv ~ lstat + x4, data = Boston)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.2834 -3.8313 -0.5295 2.3095 25.4148   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.862007 0.872084 49.15 <2e-16 \*\*\*  
## lstat -2.332821 0.123803 -18.84 <2e-16 \*\*\*  
## x4 0.043547 0.003745 11.63 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.524 on 503 degrees of freedom  
## Multiple R-squared: 0.6407, Adjusted R-squared: 0.6393   
## F-statistic: 448.5 on 2 and 503 DF, p-value: < 2.2e-16

lm.fit4 = lm(medv ~ lstat + I(lstat ^ 2) + I(lstat ^ 3), data = Boston)  
summary(lm.fit4)

##   
## 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