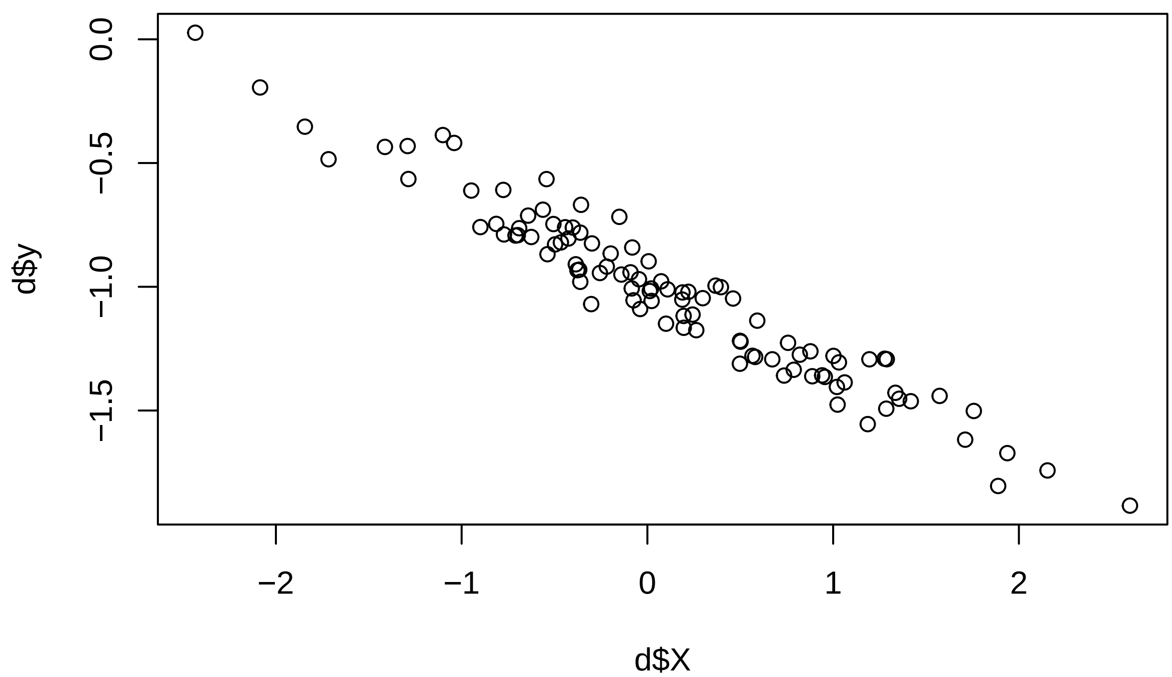
Lab 2 - Linear Regression

Generate some simulated data

```
gen_data = function(N,P,s=1/10){
    X = array(rnorm(N*P),c(N,P))
    beta = array(rnorm(P+1),c(P+1,1))
    epsilon = array(rnorm(N,sd=s),c(N,1))
    y = cbind(1,X)%*%beta + epsilon
    return(list(
        X=X,beta=beta,epsilon=epsilon,y=y
        )
    )
}

d = gen_data(100, 1)
```

```
plot(d$X, d$y)
```




```
head(d$X)
```

```
##              [,1]
```

```
## [1,]  1.3553689
## [2,] -0.4015708
## [3,]  0.2206362
## [4,]  0.3958161
## [5,] -1.0403223
## [6,] -1.8439380
```

```
head(d$beta)
```

```
##            [,1]
## [1,] -0.9911846
## [2,] -0.3637363
```

```
head(d$y)
```

```
##            [,1]
## [1,] -1.4525103
## [2,] -0.7606124
## [3,] -1.0204205
## [4,] -1.0015536
## [5,] -0.4188894
## [6,] -0.3530865
```

```
names(d)
```

```
## [1] "X"       "beta"    "epsilon" "y"
```

Fitting in R using lm

```
mod = lm(d$y~d$X)
summary(mod)
```

```
##
## Call:
## lm(formula = d$y ~ d$X)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.199125 -0.065618 -0.007215  0.050608  0.218163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.981310   0.009045 -108.50   <2e-16 ***
## d$X         -0.364997   0.009627  -37.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08961 on 98 degrees of freedom
## Multiple R-squared:  0.9362, Adjusted R-squared:  0.9355
## F-statistic:  1437 on 1 and 98 DF,  p-value: < 2.2e-16
```

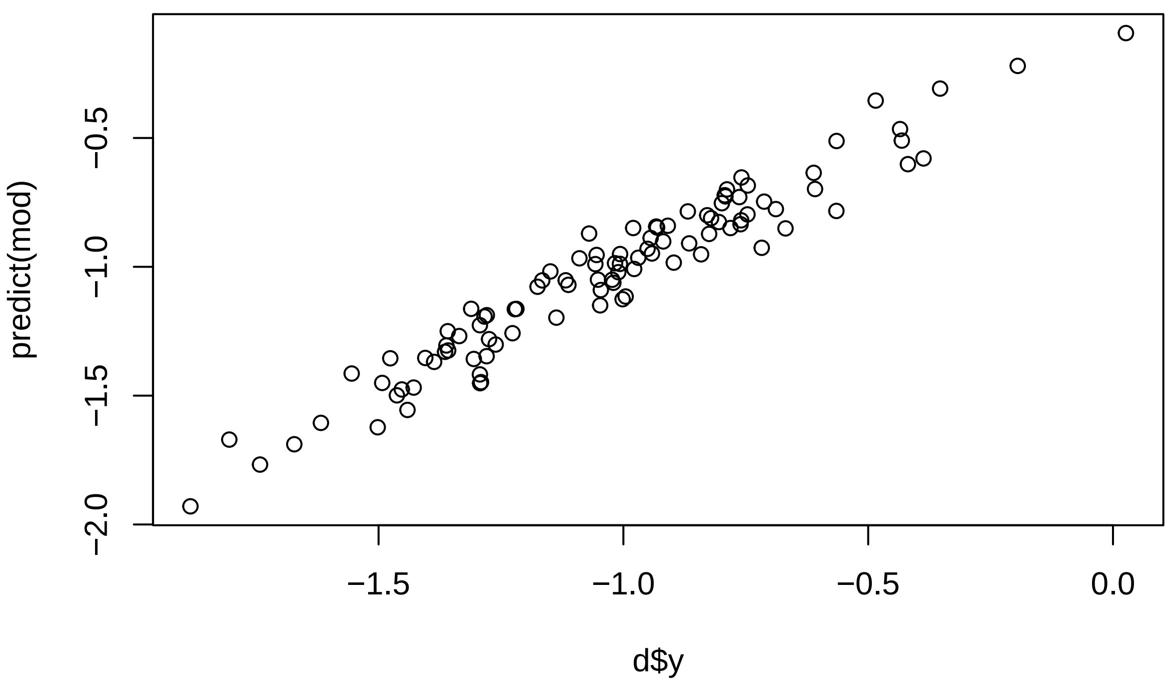
```r
mod$coef
```

```
## (Intercept)         d$X
##  -0.9813103  -0.3649973
```

```r
head(predict(mod))
```

```
##          1          2          3          4          5          6
## -1.4760163 -0.8347381 -1.0618419 -1.1257821 -0.6015955 -0.3082779
```

```r
plot(d$y, predict(mod))
```



Calculating by hand

```r
D = cbind(1, d$X)
head(D)
```

```
##      [,1]       [,2]
## [1,]    1  1.3553689
## [2,]    1 -0.4015708
```

```
## [3,]    1  0.2206362
## [4,]    1  0.3958161
## [5,]    1 -1.0403223
## [6,]    1 -1.8439380
```

```
beta_hat = solve(t(D)%*%D)%*%t(D)%*%d$y
beta_hat
```

```
##              [,1]
## [1,] -0.9813103
## [2,] -0.3649973
```

Plotting the loss

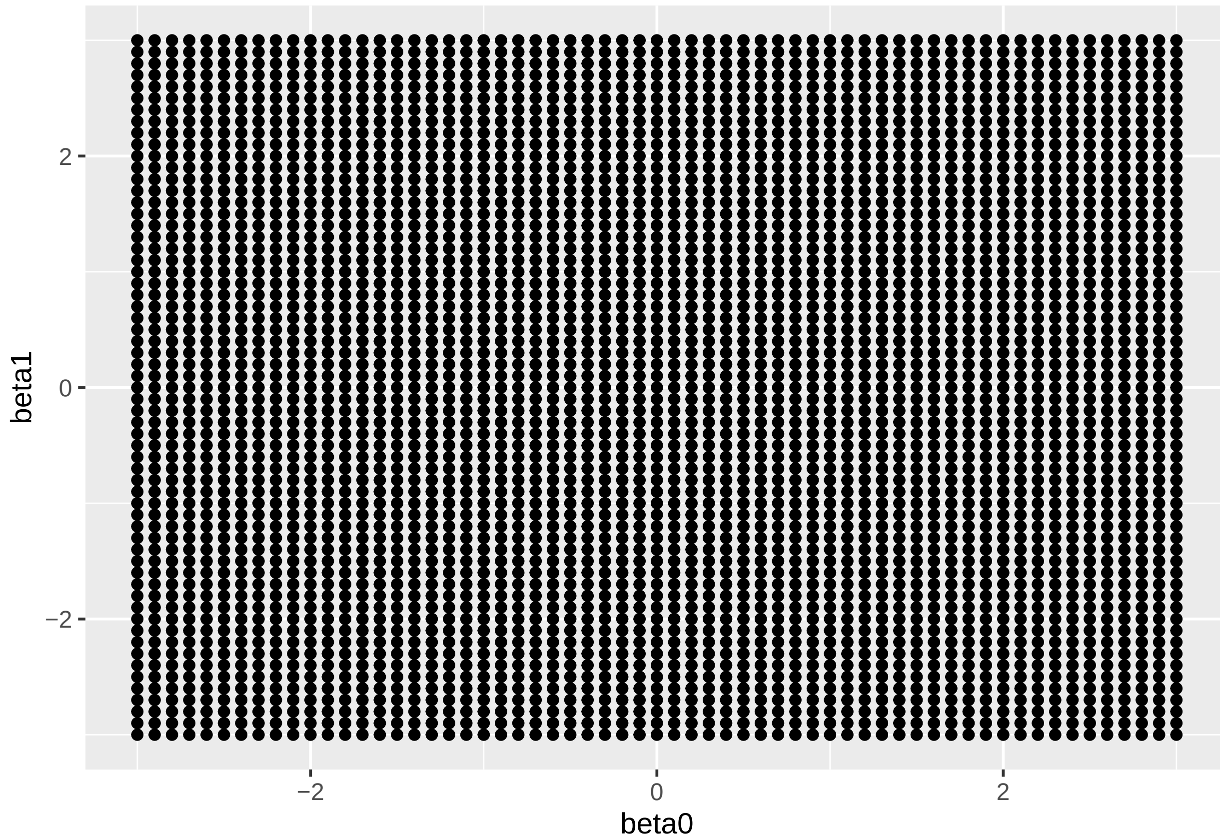
```
L = function(beta){
    sum((d$y-cbind(1,d$X)%*%beta)^2)
}
```

```
beta_grid = expand.grid(beta0=seq(-3,3,by=.1),beta1=seq(-3,3,by=.1))
```

```
head(beta_grid)
```

```
##   beta0 beta1
## 1  -3.0    -3
## 2  -2.9    -3
## 3  -2.8    -3
## 4  -2.7    -3
## 5  -2.6    -3
## 6  -2.5    -3
```

```
library('ggplot2')
ggplot(data=beta_grid,mapping=aes(x=beta0,y=beta1))+geom_point()
```




```
beta_grid$L = apply(beta_grid,1,L)
```

```
head(beta_grid)
```

```
##   beta0 beta1         L
## 1  -3.0    -3 1157.0685
## 2  -2.9    -3 1110.9614
## 3  -2.8    -3 1066.8543
## 4  -2.7    -3 1024.7472
## 5  -2.6    -3  984.6401
## 6  -2.5    -3  946.5330
```

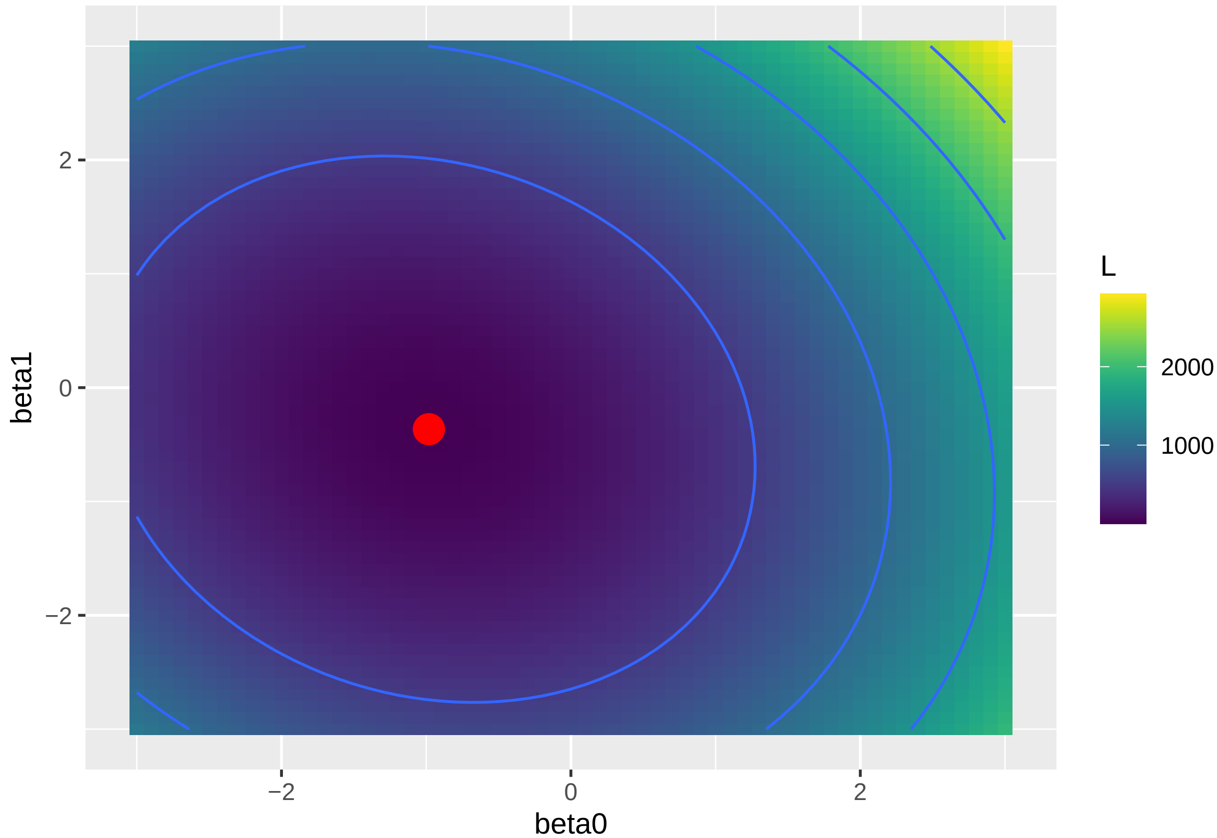
```
library('ggplot2')
library('viridis')
```

```
## Loading required package: viridisLite
```

```
beta_hat_df = data.frame(beta0=beta_hat[1],beta1=beta_hat[2])
```

```
ggplot(data=beta_grid,mapping=aes(x=beta0,y=beta1,fill=L,z=L))+geom_tile()+
    scale_fill_viridis()+geom_contour()+
    geom_point(data=beta_hat_df,mapping=aes(x=beta0,y=beta1),inherit.aes=FALSE,color='red',size=5)
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a ‘group‘ aesthetic or to convert a numerical
##   variable into a factor?
```



Categorial variables

```
library('palmerpenguins')
```

```
penguins = penguins[complete.cases(penguins),]
```

```
head(penguins[sample(nrow(penguins)),])
```

```
## # A tibble: 6 x 8
##   species   island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
##   <fct>     <fct>           <dbl>         <dbl>             <int>       <int>
## 1 Chinstrap Dream            46.8          16.5               189        3650
## 2 Adelie    Dream            37            16.5               185        3400
## 3 Gentoo    Biscoe           50.8          15.7               226        5200
## 4 Gentoo    Biscoe           46.2          14.9               221        5300
## 5 Adelie    Dream            38.8          20                 190        3950
```

```
## 6 Adelie    Dream            36.3          19.5               190        3800
## # i 2 more variables: sex <fct>, year <int>
```

```
mod = lm(flipper_length_mm~bill_length_mm+species,data=penguins)
summary(penguins)
```

```
##       species          island    bill_length_mm  bill_depth_mm
##  Adelie   :146   Biscoe   :163   Min.   :32.10   Min.   :13.10
##  Chinstrap: 68   Dream    :123   1st Qu.:39.50   1st Qu.:15.60
##  Gentoo   :119   Torgersen: 47   Median :44.50   Median :17.30
##                                  Mean   :43.99   Mean   :17.16
##                                  3rd Qu.:48.60   3rd Qu.:18.70
##                                  Max.   :59.60   Max.   :21.50
##  flipper_length_mm  body_mass_g       sex           year
##  Min.   :172       Min.   :2700   female:165   Min.   :2007
##  1st Qu.:190       1st Qu.:3550   male  :168   1st Qu.:2007
##  Median :197       Median :4050                Median :2008
##  Mean   :201       Mean   :4207                Mean   :2008
##  3rd Qu.:213       3rd Qu.:4775                3rd Qu.:2009
##  Max.   :231       Max.   :6300                Max.   :2009
```

```
summary(mod)
```

```
##
## Call:
## lm(formula = flipper_length_mm ~ bill_length_mm + species, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.8669  -3.4617  -0.0765   3.7020  15.9944
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      147.5633     4.2234  34.940  < 2e-16 ***
## bill_length_mm     1.0957     0.1081  10.139  < 2e-16 ***
## speciesChinstrap  -5.2470     1.3797  -3.803  0.00017 ***
## speciesGentoo     17.5517     1.1883  14.771  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.833 on 329 degrees of freedom
## Multiple R-squared:  0.8283, Adjusted R-squared:  0.8268
## F-statistic: 529.2 on 3 and 329 DF,  p-value: < 2.2e-16
```

```
D = model.matrix(~bill_length_mm+species,data=penguins)
head(D[sample(nrow(D)),])
```

```
##     (Intercept) bill_length_mm speciesChinstrap speciesGentoo
## 255           1           49.8                0             1
## 194           1           44.9                0             1
## 195           1           45.2                0             1
## 162           1           49.3                0             1
```

```
## 198           1           45.1                0             1
## 300           1           49.7                1             0
```

```
y = penguins$flipper_length_mm
y = array(y,c(length(y),1))
head(y)
```

```
##      [,1]
## [1,]  181
## [2,]  186
## [3,]  195
## [4,]  193
## [5,]  190
## [6,]  181
```

```
beta_hat = solve(t(D)%*%D)%*%t(D)%*%y
beta_hat
```

```
##                         [,1]
## (Intercept)       147.563315
## bill_length_mm      1.095700
## speciesChinstrap   -5.247004
## speciesGentoo      17.551650
```

```
coef(mod)
```

```
##      (Intercept)   bill_length_mm speciesChinstrap    speciesGentoo
##       147.563315         1.095700        -5.247004        17.551650
```

Fitting Issues

```
d = gen_data(100, 200)
dim(d$X)
```

```
## [1] 100 200
```

```
mod = lm(d$y~d$X)
summary(mod)
```

```
##
## Call:
## lm(formula = d$y ~ d$X)
##
## Residuals:
## ALL 100 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (101 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -206.301        NaN     NaN      NaN
## d$X1         -24.037        NaN     NaN      NaN
## d$X2         -59.243        NaN     NaN      NaN
## d$X3         -29.003        NaN     NaN      NaN
## d$X4        -152.314        NaN     NaN      NaN
## d$X5        -216.894        NaN     NaN      NaN
## d$X6        -155.324        NaN     NaN      NaN
## d$X7         -54.879        NaN     NaN      NaN
## d$X8        -165.933        NaN     NaN      NaN
## d$X9          13.655        NaN     NaN      NaN
## d$X10         20.180        NaN     NaN      NaN
## d$X11        -82.822        NaN     NaN      NaN
## d$X12       -137.442        NaN     NaN      NaN
## d$X13       -129.861        NaN     NaN      NaN
## d$X14         22.193        NaN     NaN      NaN
## d$X15         15.015        NaN     NaN      NaN
## d$X16        176.543        NaN     NaN      NaN
## d$X17        -88.152        NaN     NaN      NaN
## d$X18        -51.546        NaN     NaN      NaN
## d$X19        -15.735        NaN     NaN      NaN
## d$X20        -13.612        NaN     NaN      NaN
## d$X21        -91.500        NaN     NaN      NaN
## d$X22         25.595        NaN     NaN      NaN
## d$X23       -133.455        NaN     NaN      NaN
## d$X24        -68.670        NaN     NaN      NaN
## d$X25        -49.516        NaN     NaN      NaN
## d$X26        -14.373        NaN     NaN      NaN
## d$X27        -14.771        NaN     NaN      NaN
## d$X28         -1.649        NaN     NaN      NaN
## d$X29        200.617        NaN     NaN      NaN
## d$X30         62.621        NaN     NaN      NaN
## d$X31        -55.720        NaN     NaN      NaN
## d$X32         97.028        NaN     NaN      NaN
## d$X33         -3.802        NaN     NaN      NaN
## d$X34        109.292        NaN     NaN      NaN
## d$X35         29.304        NaN     NaN      NaN
## d$X36        -34.230        NaN     NaN      NaN
## d$X37         29.383        NaN     NaN      NaN
## d$X38        -87.337        NaN     NaN      NaN
## d$X39         34.597        NaN     NaN      NaN
## d$X40       -116.826        NaN     NaN      NaN
## d$X41        -34.074        NaN     NaN      NaN
## d$X42        -41.821        NaN     NaN      NaN
## d$X43        134.289        NaN     NaN      NaN
## d$X44        -73.123        NaN     NaN      NaN
## d$X45        -27.158        NaN     NaN      NaN
## d$X46         72.774        NaN     NaN      NaN
## d$X47        127.972        NaN     NaN      NaN
## d$X48         29.090        NaN     NaN      NaN
## d$X49        193.462        NaN     NaN      NaN
## d$X50        -51.712        NaN     NaN      NaN
## d$X51        205.408        NaN     NaN      NaN
## d$X52         -7.059        NaN     NaN      NaN
## d$X53       -132.414        NaN     NaN      NaN
```

```
## d$X54        -78.636       NaN     NaN      NaN
## d$X55       -171.782       NaN     NaN      NaN
## d$X56        -10.728       NaN     NaN      NaN
## d$X57       -112.588       NaN     NaN      NaN
## d$X58       -142.895       NaN     NaN      NaN
## d$X59        121.748       NaN     NaN      NaN
## d$X60         36.599       NaN     NaN      NaN
## d$X61        224.667       NaN     NaN      NaN
## d$X62         52.499       NaN     NaN      NaN
## d$X63       -266.742       NaN     NaN      NaN
## d$X64        -87.986       NaN     NaN      NaN
## d$X65          4.535       NaN     NaN      NaN
## d$X66       -122.240       NaN     NaN      NaN
## d$X67         95.367       NaN     NaN      NaN
## d$X68         79.907       NaN     NaN      NaN
## d$X69       -112.503       NaN     NaN      NaN
## d$X70        -92.156       NaN     NaN      NaN
## d$X71        213.993       NaN     NaN      NaN
## d$X72        174.006       NaN     NaN      NaN
## d$X73       -123.972       NaN     NaN      NaN
## d$X74       -155.493       NaN     NaN      NaN
## d$X75        -23.587       NaN     NaN      NaN
## d$X76          7.692       NaN     NaN      NaN
## d$X77         -7.610       NaN     NaN      NaN
## d$X78        -54.183       NaN     NaN      NaN
## d$X79        -61.427       NaN     NaN      NaN
## d$X80       -150.944       NaN     NaN      NaN
## d$X81        -59.984       NaN     NaN      NaN
## d$X82       -248.135       NaN     NaN      NaN
## d$X83        259.344       NaN     NaN      NaN
## d$X84        118.095       NaN     NaN      NaN
## d$X85        308.802       NaN     NaN      NaN
## d$X86         27.981       NaN     NaN      NaN
## d$X87        176.505       NaN     NaN      NaN
## d$X88       -305.001       NaN     NaN      NaN
## d$X89         17.236       NaN     NaN      NaN
## d$X90        182.712       NaN     NaN      NaN
## d$X91        108.675       NaN     NaN      NaN
## d$X92         52.005       NaN     NaN      NaN
## d$X93        130.349       NaN     NaN      NaN
## d$X94         30.665       NaN     NaN      NaN
## d$X95        141.958       NaN     NaN      NaN
## d$X96        -25.987       NaN     NaN      NaN
## d$X97         20.332       NaN     NaN      NaN
## d$X98        132.395       NaN     NaN      NaN
## d$X99        151.673       NaN     NaN      NaN
## d$X100            NA        NA      NA       NA
## d$X101            NA        NA      NA       NA
## d$X102            NA        NA      NA       NA
## d$X103            NA        NA      NA       NA
## d$X104            NA        NA      NA       NA
## d$X105            NA        NA      NA       NA
## d$X106            NA        NA      NA       NA
## d$X107            NA        NA      NA       NA
```

```
## d$X108               NA          NA       NA        NA
## d$X109               NA          NA       NA        NA
## d$X110               NA          NA       NA        NA
## d$X111               NA          NA       NA        NA
## d$X112               NA          NA       NA        NA
## d$X113               NA          NA       NA        NA
## d$X114               NA          NA       NA        NA
## d$X115               NA          NA       NA        NA
## d$X116               NA          NA       NA        NA
## d$X117               NA          NA       NA        NA
## d$X118               NA          NA       NA        NA
## d$X119               NA          NA       NA        NA
## d$X120               NA          NA       NA        NA
## d$X121               NA          NA       NA        NA
## d$X122               NA          NA       NA        NA
## d$X123               NA          NA       NA        NA
## d$X124               NA          NA       NA        NA
## d$X125               NA          NA       NA        NA
## d$X126               NA          NA       NA        NA
## d$X127               NA          NA       NA        NA
## d$X128               NA          NA       NA        NA
## d$X129               NA          NA       NA        NA
## d$X130               NA          NA       NA        NA
## d$X131               NA          NA       NA        NA
## d$X132               NA          NA       NA        NA
## d$X133               NA          NA       NA        NA
## d$X134               NA          NA       NA        NA
## d$X135               NA          NA       NA        NA
## d$X136               NA          NA       NA        NA
## d$X137               NA          NA       NA        NA
## d$X138               NA          NA       NA        NA
## d$X139               NA          NA       NA        NA
## d$X140               NA          NA       NA        NA
## d$X141               NA          NA       NA        NA
## d$X142               NA          NA       NA        NA
## d$X143               NA          NA       NA        NA
## d$X144               NA          NA       NA        NA
## d$X145               NA          NA       NA        NA
## d$X146               NA          NA       NA        NA
## d$X147               NA          NA       NA        NA
## d$X148               NA          NA       NA        NA
## d$X149               NA          NA       NA        NA
## d$X150               NA          NA       NA        NA
## d$X151               NA          NA       NA        NA
## d$X152               NA          NA       NA        NA
## d$X153               NA          NA       NA        NA
## d$X154               NA          NA       NA        NA
## d$X155               NA          NA       NA        NA
## d$X156               NA          NA       NA        NA
## d$X157               NA          NA       NA        NA
## d$X158               NA          NA       NA        NA
## d$X159               NA          NA       NA        NA
## d$X160               NA          NA       NA        NA
## d$X161               NA          NA       NA        NA
```

```
## d$X162            NA          NA      NA       NA
## d$X163            NA          NA      NA       NA
## d$X164            NA          NA      NA       NA
## d$X165            NA          NA      NA       NA
## d$X166            NA          NA      NA       NA
## d$X167            NA          NA      NA       NA
## d$X168            NA          NA      NA       NA
## d$X169            NA          NA      NA       NA
## d$X170            NA          NA      NA       NA
## d$X171            NA          NA      NA       NA
## d$X172            NA          NA      NA       NA
## d$X173            NA          NA      NA       NA
## d$X174            NA          NA      NA       NA
## d$X175            NA          NA      NA       NA
## d$X176            NA          NA      NA       NA
## d$X177            NA          NA      NA       NA
## d$X178            NA          NA      NA       NA
## d$X179            NA          NA      NA       NA
## d$X180            NA          NA      NA       NA
## d$X181            NA          NA      NA       NA
## d$X182            NA          NA      NA       NA
## d$X183            NA          NA      NA       NA
## d$X184            NA          NA      NA       NA
## d$X185            NA          NA      NA       NA
## d$X186            NA          NA      NA       NA
## d$X187            NA          NA      NA       NA
## d$X188            NA          NA      NA       NA
## d$X189            NA          NA      NA       NA
## d$X190            NA          NA      NA       NA
## d$X191            NA          NA      NA       NA
## d$X192            NA          NA      NA       NA
## d$X193            NA          NA      NA       NA
## d$X194            NA          NA      NA       NA
## d$X195            NA          NA      NA       NA
## d$X196            NA          NA      NA       NA
## d$X197            NA          NA      NA       NA
## d$X198            NA          NA      NA       NA
## d$X199            NA          NA      NA       NA
## d$X200            NA          NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 99 and 0 DF,  p-value: NA
```

```r
tail(coef(mod))
```

```
## d$X195 d$X196 d$X197 d$X198 d$X199 d$X200
##     NA     NA     NA     NA     NA     NA
```

```r
D = model.matrix(mod)
D[1:5,1:5]
```

```
##   (Intercept)       d$X1      d$X2       d$X3        d$X4
```

```
## 1           1  1.9506976  1.362844  0.2939937  1.42091779
## 2           1 -0.7228125  1.029881  0.6790914  0.66984844
## 3           1 -1.7264662 -0.267034 -0.4034870 -1.17663520
## 4           1 -0.6349577 -2.328357  0.7308996 -0.03632376
## 5           1  2.3872024 -2.811631  0.7506581  0.25321084
```

```
# singular
# solve(t(D)%*%D)%*%t(D)%*%d$y
```

another example

```
xx = rnorm(100)
X = cbind(xx,xx)
colnames(X) = c('V1','V2')
head(X)
```

```
##              V1         V2
## [1,]  0.2521685  0.2521685
## [2,]  0.3454101  0.3454101
## [3,]  0.5879322  0.5879322
## [4,]  1.1056134  1.1056134
## [5,] -0.2127039 -0.2127039
## [6,]  0.2147335  0.2147335
```

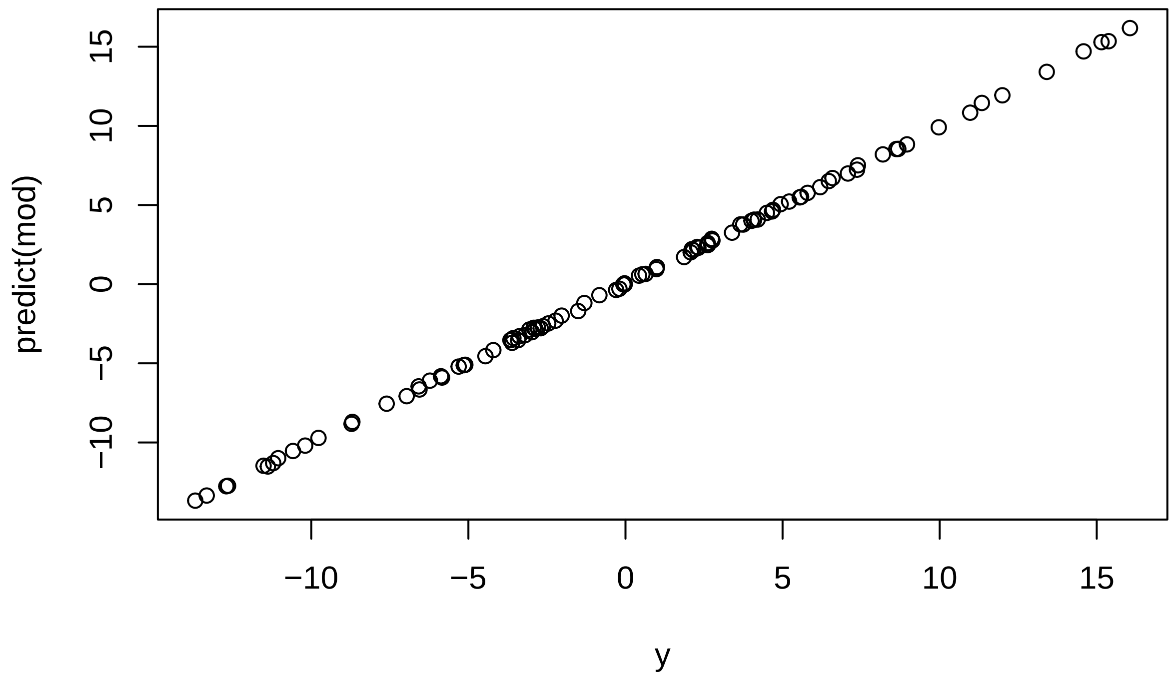
```
true_beta = array(c(3,5),c(2,1))
true_beta
```

```
##      [,1]
## [1,]    3
## [2,]    5
```

```
y = X %*% true_beta + rnorm(100,sd=1/10)
mod = lm(y~X)
summary(mod)
```

```
##
## Call:
## lm(formula = y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19603 -0.07130  0.00574  0.06889  0.19562
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.001022   0.009094  -0.112    0.911
## XV1          7.988922   0.010473 762.827   <2e-16 ***
## XV2                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09089 on 98 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 5.819e+05 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
plot(y,predict(mod))
```



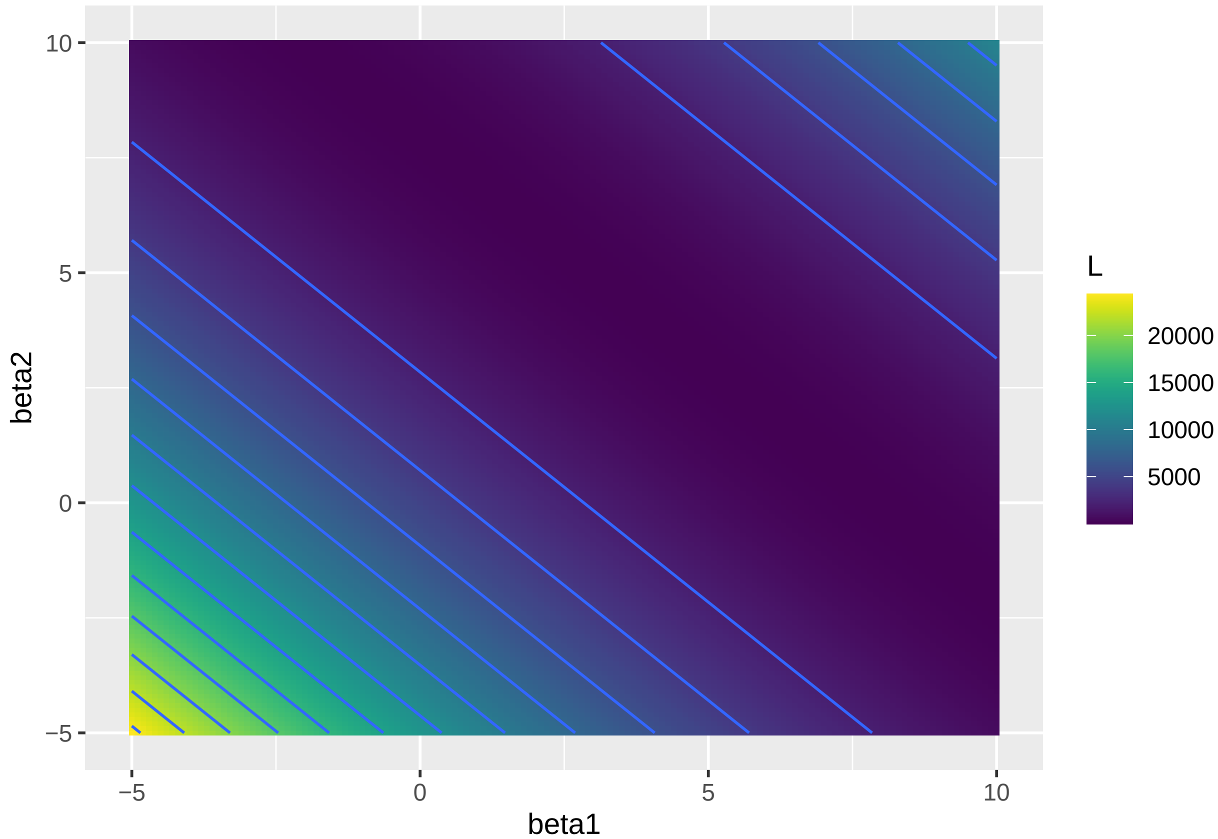

```
#singular
D = cbind(1,X)
# beta_hat = solve(t(D)%*%D)%*%t(D)%*%y
```

what does the loss look like?

```
L = function(beta){
    sum((y-X%*%beta)^2)
}
beta_grid = expand.grid(beta1=seq(-5,10,by=.1),beta2=seq(-5,10,by=.1))
beta_grid$L = apply(beta_grid,1,L)
```

```
ggplot(data=beta_grid,mapping=aes(x=beta1,y=beta2,fill=L,z=L))+geom_tile()+
    scale_fill_viridis()+geom_contour()
```

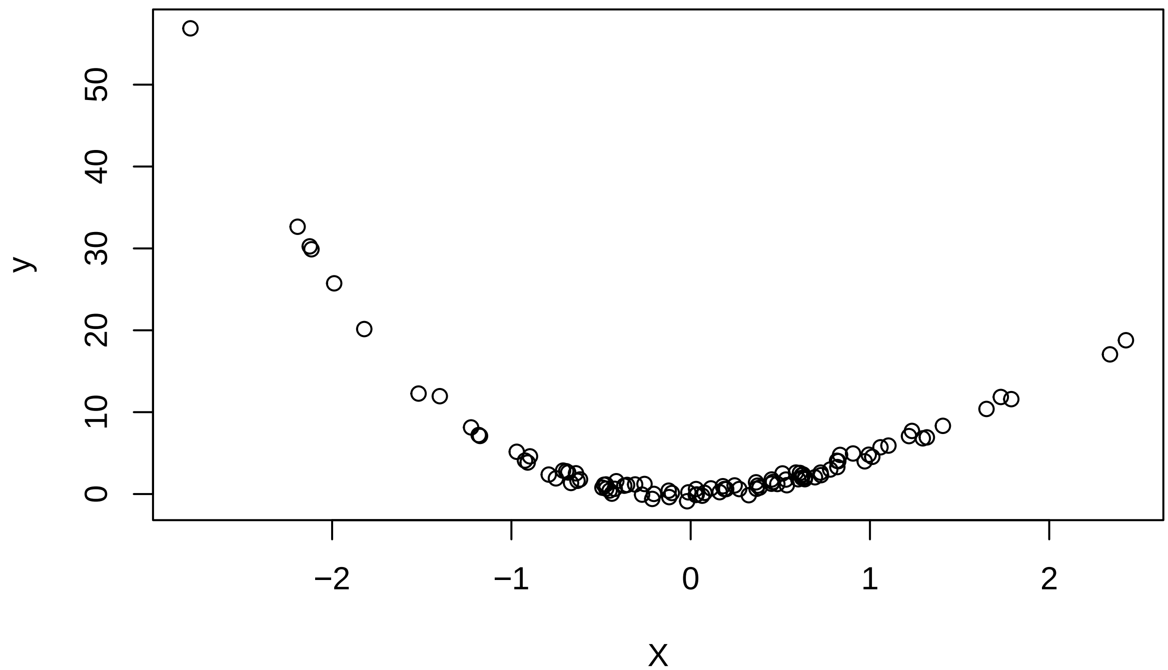
```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a ‘group‘ aesthetic or to convert a numerical
##   variable into a factor?
```



polynomial regression

```
N = 100
P = 1
X = array(rnorm(N*P),c(N,P))
```

```
y = X + 5*X^2 - X^3 + rnorm(100,0,.5)
plot(X,y)
```




```
D = cbind(1,X,X^2,X^3)
head(D)
```

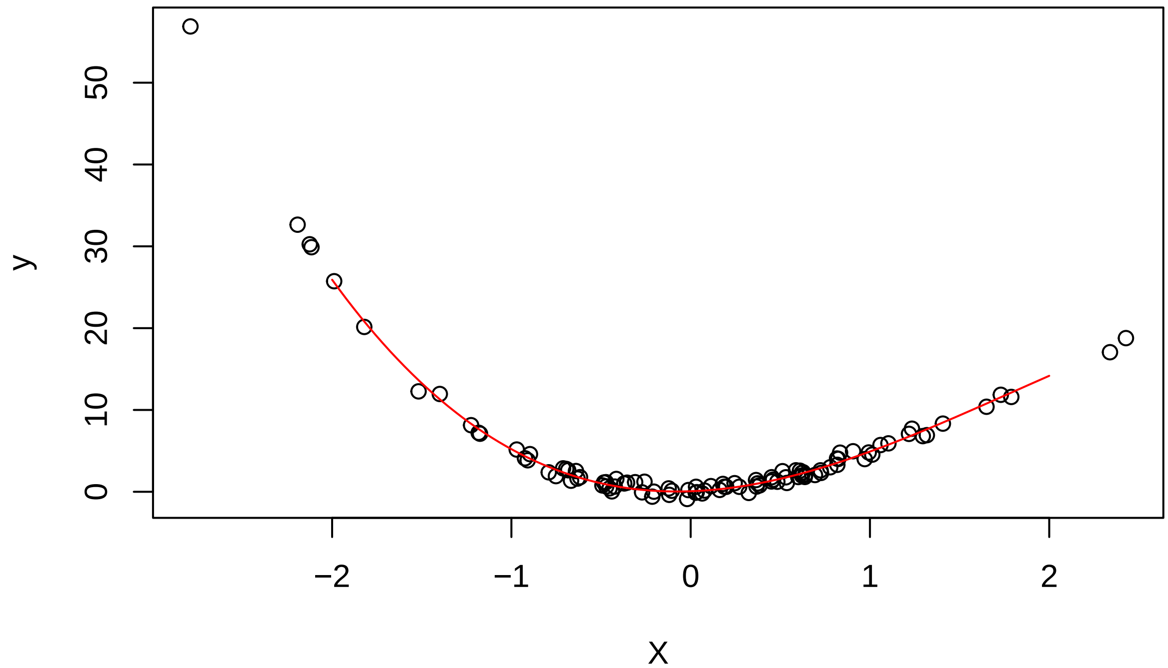
```
##      [,1]       [,2]      [,3]        [,4]
## [1,]    1 -0.3706135 0.1373543 -0.05090536
## [2,]    1 -0.6312546 0.3984824 -0.25154384
## [3,]    1 -1.9886266 3.9546356 -7.86429349
## [4,]    1  0.8333414 0.6944578  0.57872044
## [5,]    1  0.6000449 0.3600539  0.21604850
## [6,]    1 -0.6674794 0.4455287 -0.29738126
```

```
beta_hat = solve(t(D)%*%D)%*%t(D)%*%y
beta_hat
```

```
##             [,1]
## [1,]  0.06439571
## [2,]  0.78256135
## [3,]  4.99528992
## [4,] -0.92986473
```

```
xp = seq(-2,2,length.out=100)
Dp = cbind(1,xp,xp^2,xp^3)
y_pred = Dp%*%beta_hat
```

```
plot(X,y)
lines(xp,y_pred,col='red')
```



More example

```
#install.packages('MASS')
```

```
library('MASS')
```

```
data(Boston)
```

```
dim(Boston)
```

```
## [1] 506  14
```

```
Boston[1:5,]
```

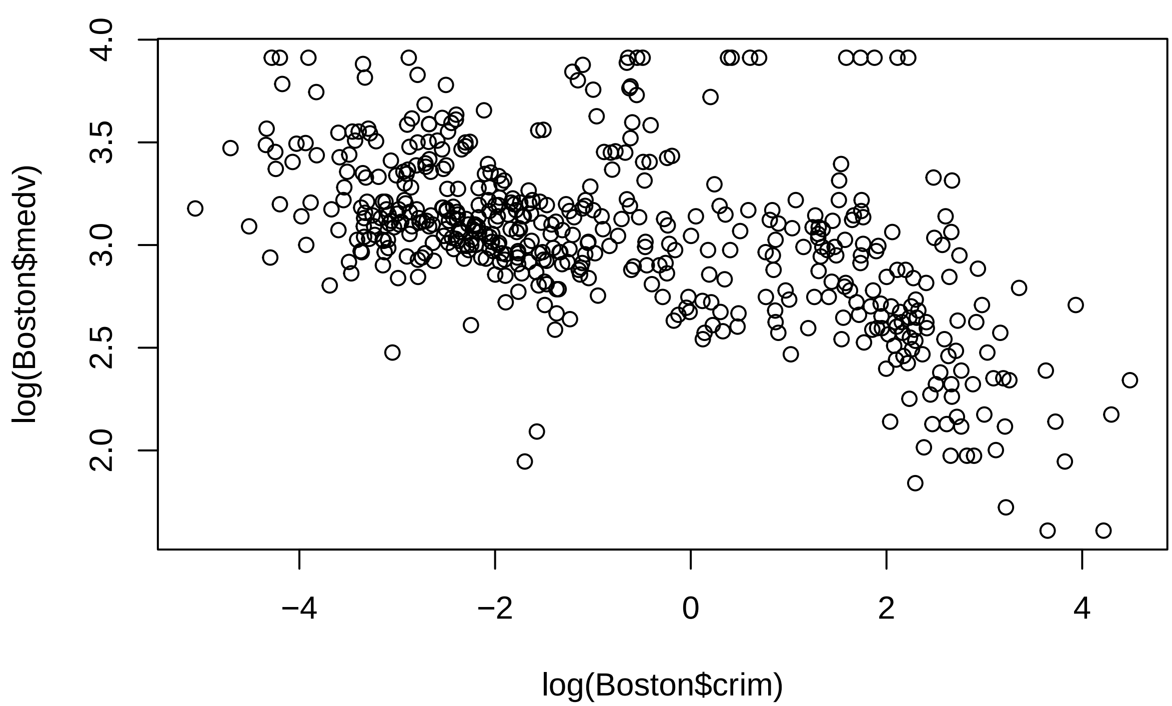
```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
```

```
## 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
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
```

```
?Boston
```

let's fit a regression to predict the median house value from the crime rate

```
plot(log(Boston$crim),log(Boston$medv))
```



`medv~crim` basically says `medv = beta0 + beta1*crim`

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

```
##
## Call:
## lm(formula = medv ~ crim, data = Boston)
##
```

```
## Coefficients:
## (Intercept)         crim
##     24.0331      -0.4152
```

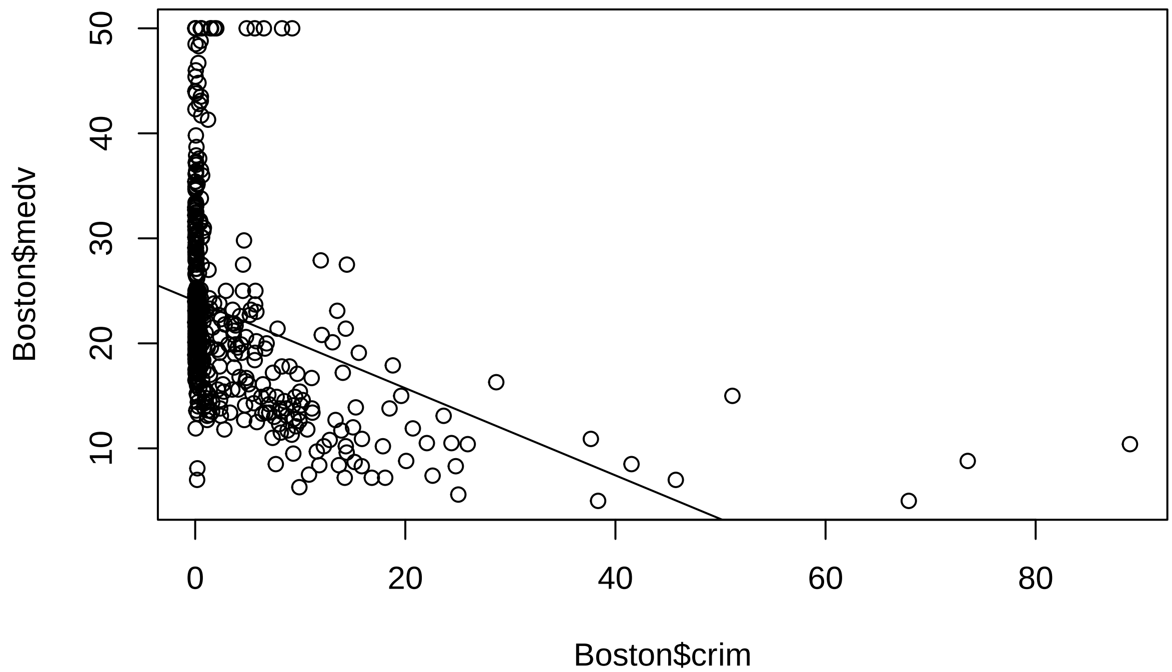
```
summary(mod)
```

```
##
## 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
```

```
mod$coef
```

```
## (Intercept)        crim
##  24.0331062  -0.4151903
```

```
plot(Boston$crim,Boston$medv)
abline(coef=mod$coef)
```




```
X = array(Boston$crim,c(506,1))
X = cbind(1,X)
X
```

```
##        [,1]    [,2]
##   [1,]    1 0.00632
##   [2,]    1 0.02731
##   [3,]    1 0.02729
##   [4,]    1 0.03237
##   [5,]    1 0.06905
##   [6,]    1 0.02985
##   [7,]    1 0.08829
##   [8,]    1 0.14455
##   [9,]    1 0.21124
##  [10,]    1 0.17004
##  [11,]    1 0.22489
##  [12,]    1 0.11747
##  [13,]    1 0.09378
##  [14,]    1 0.62976
##  [15,]    1 0.63796
##  [16,]    1 0.62739
##  [17,]    1 1.05393
##  [18,]    1 0.78420
##  [19,]    1 0.80271
##  [20,]    1 0.72580
##  [21,]    1 1.25179
```

```
## [22,]    1 0.85204
## [23,]    1 1.23247
## [24,]    1 0.98843
## [25,]    1 0.75026
## [26,]    1 0.84054
## [27,]    1 0.67191
## [28,]    1 0.95577
## [29,]    1 0.77299
## [30,]    1 1.00245
## [31,]    1 1.13081
## [32,]    1 1.35472
## [33,]    1 1.38799
## [34,]    1 1.15172
## [35,]    1 1.61282
## [36,]    1 0.06417
## [37,]    1 0.09744
## [38,]    1 0.08014
## [39,]    1 0.17505
## [40,]    1 0.02763
## [41,]    1 0.03359
## [42,]    1 0.12744
## [43,]    1 0.14150
## [44,]    1 0.15936
## [45,]    1 0.12269
## [46,]    1 0.17142
## [47,]    1 0.18836
## [48,]    1 0.22927
## [49,]    1 0.25387
## [50,]    1 0.21977
## [51,]    1 0.08873
## [52,]    1 0.04337
## [53,]    1 0.05360
## [54,]    1 0.04981
## [55,]    1 0.01360
## [56,]    1 0.01311
## [57,]    1 0.02055
## [58,]    1 0.01432
## [59,]    1 0.15445
## [60,]    1 0.10328
## [61,]    1 0.14932
## [62,]    1 0.17171
## [63,]    1 0.11027
## [64,]    1 0.12650
## [65,]    1 0.01951
## [66,]    1 0.03584
## [67,]    1 0.04379
## [68,]    1 0.05789
## [69,]    1 0.13554
## [70,]    1 0.12816
## [71,]    1 0.08826
## [72,]    1 0.15876
## [73,]    1 0.09164
## [74,]    1 0.19539
## [75,]    1 0.07896
```

```
##  [76,]    1  0.09512
##  [77,]    1  0.10153
##  [78,]    1  0.08707
##  [79,]    1  0.05646
##  [80,]    1  0.08387
##  [81,]    1  0.04113
##  [82,]    1  0.04462
##  [83,]    1  0.03659
##  [84,]    1  0.03551
##  [85,]    1  0.05059
##  [86,]    1  0.05735
##  [87,]    1  0.05188
##  [88,]    1  0.07151
##  [89,]    1  0.05660
##  [90,]    1  0.05302
##  [91,]    1  0.04684
##  [92,]    1  0.03932
##  [93,]    1  0.04203
##  [94,]    1  0.02875
##  [95,]    1  0.04294
##  [96,]    1  0.12204
##  [97,]    1  0.11504
##  [98,]    1  0.12083
##  [99,]    1  0.08187
## [100,]    1  0.06860
## [101,]    1  0.14866
## [102,]    1  0.11432
## [103,]    1  0.22876
## [104,]    1  0.21161
## [105,]    1  0.13960
## [106,]    1  0.13262
## [107,]    1  0.17120
## [108,]    1  0.13117
## [109,]    1  0.12802
## [110,]    1  0.26363
## [111,]    1  0.10793
## [112,]    1  0.10084
## [113,]    1  0.12329
## [114,]    1  0.22212
## [115,]    1  0.14231
## [116,]    1  0.17134
## [117,]    1  0.13158
## [118,]    1  0.15098
## [119,]    1  0.13058
## [120,]    1  0.14476
## [121,]    1  0.06899
## [122,]    1  0.07165
## [123,]    1  0.09299
## [124,]    1  0.15038
## [125,]    1  0.09849
## [126,]    1  0.16902
## [127,]    1  0.38735
## [128,]    1  0.25915
## [129,]    1  0.32543
```

```
## [130,]    1  0.88125
## [131,]    1  0.34006
## [132,]    1  1.19294
## [133,]    1  0.59005
## [134,]    1  0.32982
## [135,]    1  0.97617
## [136,]    1  0.55778
## [137,]    1  0.32264
## [138,]    1  0.35233
## [139,]    1  0.24980
## [140,]    1  0.54452
## [141,]    1  0.29090
## [142,]    1  1.62864
## [143,]    1  3.32105
## [144,]    1  4.09740
## [145,]    1  2.77974
## [146,]    1  2.37934
## [147,]    1  2.15505
## [148,]    1  2.36862
## [149,]    1  2.33099
## [150,]    1  2.73397
## [151,]    1  1.65660
## [152,]    1  1.49632
## [153,]    1  1.12658
## [154,]    1  2.14918
## [155,]    1  1.41385
## [156,]    1  3.53501
## [157,]    1  2.44668
## [158,]    1  1.22358
## [159,]    1  1.34284
## [160,]    1  1.42502
## [161,]    1  1.27346
## [162,]    1  1.46336
## [163,]    1  1.83377
## [164,]    1  1.51902
## [165,]    1  2.24236
## [166,]    1  2.92400
## [167,]    1  2.01019
## [168,]    1  1.80028
## [169,]    1  2.30040
## [170,]    1  2.44953
## [171,]    1  1.20742
## [172,]    1  2.31390
## [173,]    1  0.13914
## [174,]    1  0.09178
## [175,]    1  0.08447
## [176,]    1  0.06664
## [177,]    1  0.07022
## [178,]    1  0.05425
## [179,]    1  0.06642
## [180,]    1  0.05780
## [181,]    1  0.06588
## [182,]    1  0.06888
## [183,]    1  0.09103
```

```
## [184,]    1  0.10008
## [185,]    1  0.08308
## [186,]    1  0.06047
## [187,]    1  0.05602
## [188,]    1  0.07875
## [189,]    1  0.12579
## [190,]    1  0.08370
## [191,]    1  0.09068
## [192,]    1  0.06911
## [193,]    1  0.08664
## [194,]    1  0.02187
## [195,]    1  0.01439
## [196,]    1  0.01381
## [197,]    1  0.04011
## [198,]    1  0.04666
## [199,]    1  0.03768
## [200,]    1  0.03150
## [201,]    1  0.01778
## [202,]    1  0.03445
## [203,]    1  0.02177
## [204,]    1  0.03510
## [205,]    1  0.02009
## [206,]    1  0.13642
## [207,]    1  0.22969
## [208,]    1  0.25199
## [209,]    1  0.13587
## [210,]    1  0.43571
## [211,]    1  0.17446
## [212,]    1  0.37578
## [213,]    1  0.21719
## [214,]    1  0.14052
## [215,]    1  0.28955
## [216,]    1  0.19802
## [217,]    1  0.04560
## [218,]    1  0.07013
## [219,]    1  0.11069
## [220,]    1  0.11425
## [221,]    1  0.35809
## [222,]    1  0.40771
## [223,]    1  0.62356
## [224,]    1  0.61470
## [225,]    1  0.31533
## [226,]    1  0.52693
## [227,]    1  0.38214
## [228,]    1  0.41238
## [229,]    1  0.29819
## [230,]    1  0.44178
## [231,]    1  0.53700
## [232,]    1  0.46296
## [233,]    1  0.57529
## [234,]    1  0.33147
## [235,]    1  0.44791
## [236,]    1  0.33045
## [237,]    1  0.52058
```

```
## [238,]    1  0.51183
## [239,]    1  0.08244
## [240,]    1  0.09252
## [241,]    1  0.11329
## [242,]    1  0.10612
## [243,]    1  0.10290
## [244,]    1  0.12757
## [245,]    1  0.20608
## [246,]    1  0.19133
## [247,]    1  0.33983
## [248,]    1  0.19657
## [249,]    1  0.16439
## [250,]    1  0.19073
## [251,]    1  0.14030
## [252,]    1  0.21409
## [253,]    1  0.08221
## [254,]    1  0.36894
## [255,]    1  0.04819
## [256,]    1  0.03548
## [257,]    1  0.01538
## [258,]    1  0.61154
## [259,]    1  0.66351
## [260,]    1  0.65665
## [261,]    1  0.54011
## [262,]    1  0.53412
## [263,]    1  0.52014
## [264,]    1  0.82526
## [265,]    1  0.55007
## [266,]    1  0.76162
## [267,]    1  0.78570
## [268,]    1  0.57834
## [269,]    1  0.54050
## [270,]    1  0.09065
## [271,]    1  0.29916
## [272,]    1  0.16211
## [273,]    1  0.11460
## [274,]    1  0.22188
## [275,]    1  0.05644
## [276,]    1  0.09604
## [277,]    1  0.10469
## [278,]    1  0.06127
## [279,]    1  0.07978
## [280,]    1  0.21038
## [281,]    1  0.03578
## [282,]    1  0.03705
## [283,]    1  0.06129
## [284,]    1  0.01501
## [285,]    1  0.00906
## [286,]    1  0.01096
## [287,]    1  0.01965
## [288,]    1  0.03871
## [289,]    1  0.04590
## [290,]    1  0.04297
## [291,]    1  0.03502
```

```
## [292,]    1  0.07886
## [293,]    1  0.03615
## [294,]    1  0.08265
## [295,]    1  0.08199
## [296,]    1  0.12932
## [297,]    1  0.05372
## [298,]    1  0.14103
## [299,]    1  0.06466
## [300,]    1  0.05561
## [301,]    1  0.04417
## [302,]    1  0.03537
## [303,]    1  0.09266
## [304,]    1  0.10000
## [305,]    1  0.05515
## [306,]    1  0.05479
## [307,]    1  0.07503
## [308,]    1  0.04932
## [309,]    1  0.49298
## [310,]    1  0.34940
## [311,]    1  2.63548
## [312,]    1  0.79041
## [313,]    1  0.26169
## [314,]    1  0.26938
## [315,]    1  0.36920
## [316,]    1  0.25356
## [317,]    1  0.31827
## [318,]    1  0.24522
## [319,]    1  0.40202
## [320,]    1  0.47547
## [321,]    1  0.16760
## [322,]    1  0.18159
## [323,]    1  0.35114
## [324,]    1  0.28392
## [325,]    1  0.34109
## [326,]    1  0.19186
## [327,]    1  0.30347
## [328,]    1  0.24103
## [329,]    1  0.06617
## [330,]    1  0.06724
## [331,]    1  0.04544
## [332,]    1  0.05023
## [333,]    1  0.03466
## [334,]    1  0.05083
## [335,]    1  0.03738
## [336,]    1  0.03961
## [337,]    1  0.03427
## [338,]    1  0.03041
## [339,]    1  0.03306
## [340,]    1  0.05497
## [341,]    1  0.06151
## [342,]    1  0.01301
## [343,]    1  0.02498
## [344,]    1  0.02543
## [345,]    1  0.03049
```

```
## [346,]    1  0.03113
## [347,]    1  0.06162
## [348,]    1  0.01870
## [349,]    1  0.01501
## [350,]    1  0.02899
## [351,]    1  0.06211
## [352,]    1  0.07950
## [353,]    1  0.07244
## [354,]    1  0.01709
## [355,]    1  0.04301
## [356,]    1  0.10659
## [357,]    1  8.98296
## [358,]    1  3.84970
## [359,]    1  5.20177
## [360,]    1  4.26131
## [361,]    1  4.54192
## [362,]    1  3.83684
## [363,]    1  3.67822
## [364,]    1  4.22239
## [365,]    1  3.47428
## [366,]    1  4.55587
## [367,]    1  3.69695
## [368,]    1 13.52220
## [369,]    1  4.89822
## [370,]    1  5.66998
## [371,]    1  6.53876
## [372,]    1  9.23230
## [373,]    1  8.26725
## [374,]    1 11.10810
## [375,]    1 18.49820
## [376,]    1 19.60910
## [377,]    1 15.28800
## [378,]    1  9.82349
## [379,]    1 23.64820
## [380,]    1 17.86670
## [381,]    1 88.97620
## [382,]    1 15.87440
## [383,]    1  9.18702
## [384,]    1  7.99248
## [385,]    1 20.08490
## [386,]    1 16.81180
## [387,]    1 24.39380
## [388,]    1 22.59710
## [389,]    1 14.33370
## [390,]    1  8.15174
## [391,]    1  6.96215
## [392,]    1  5.29305
## [393,]    1 11.57790
## [394,]    1  8.64476
## [395,]    1 13.35980
## [396,]    1  8.71675
## [397,]    1  5.87205
## [398,]    1  7.67202
## [399,]    1 38.35180
```

```
## [400,]    1  9.91655
## [401,]    1 25.04610
## [402,]    1 14.23620
## [403,]    1  9.59571
## [404,]    1 24.80170
## [405,]    1 41.52920
## [406,]    1 67.92080
## [407,]    1 20.71620
## [408,]    1 11.95110
## [409,]    1  7.40389
## [410,]    1 14.43830
## [411,]    1 51.13580
## [412,]    1 14.05070
## [413,]    1 18.81100
## [414,]    1 28.65580
## [415,]    1 45.74610
## [416,]    1 18.08460
## [417,]    1 10.83420
## [418,]    1 25.94060
## [419,]    1 73.53410
## [420,]    1 11.81230
## [421,]    1 11.08740
## [422,]    1  7.02259
## [423,]    1 12.04820
## [424,]    1  7.05042
## [425,]    1  8.79212
## [426,]    1 15.86030
## [427,]    1 12.24720
## [428,]    1 37.66190
## [429,]    1  7.36711
## [430,]    1  9.33889
## [431,]    1  8.49213
## [432,]    1 10.06230
## [433,]    1  6.44405
## [434,]    1  5.58107
## [435,]    1 13.91340
## [436,]    1 11.16040
## [437,]    1 14.42080
## [438,]    1 15.17720
## [439,]    1 13.67810
## [440,]    1  9.39063
## [441,]    1 22.05110
## [442,]    1  9.72418
## [443,]    1  5.66637
## [444,]    1  9.96654
## [445,]    1 12.80230
## [446,]    1 10.67180
## [447,]    1  6.28807
## [448,]    1  9.92485
## [449,]    1  9.32909
## [450,]    1  7.52601
## [451,]    1  6.71772
## [452,]    1  5.44114
## [453,]    1  5.09017
```

```
## [454,]    1  8.24809
## [455,]    1  9.51363
## [456,]    1  4.75237
## [457,]    1  4.66883
## [458,]    1  8.20058
## [459,]    1  7.75223
## [460,]    1  6.80117
## [461,]    1  4.81213
## [462,]    1  3.69311
## [463,]    1  6.65492
## [464,]    1  5.82115
## [465,]    1  7.83932
## [466,]    1  3.16360
## [467,]    1  3.77498
## [468,]    1  4.42228
## [469,]    1 15.57570
## [470,]    1 13.07510
## [471,]    1  4.34879
## [472,]    1  4.03841
## [473,]    1  3.56868
## [474,]    1  4.64689
## [475,]    1  8.05579
## [476,]    1  6.39312
## [477,]    1  4.87141
## [478,]    1 15.02340
## [479,]    1 10.23300
## [480,]    1 14.33370
## [481,]    1  5.82401
## [482,]    1  5.70818
## [483,]    1  5.73116
## [484,]    1  2.81838
## [485,]    1  2.37857
## [486,]    1  3.67367
## [487,]    1  5.69175
## [488,]    1  4.83567
## [489,]    1  0.15086
## [490,]    1  0.18337
## [491,]    1  0.20746
## [492,]    1  0.10574
## [493,]    1  0.11132
## [494,]    1  0.17331
## [495,]    1  0.27957
## [496,]    1  0.17899
## [497,]    1  0.28960
## [498,]    1  0.26838
## [499,]    1  0.23912
## [500,]    1  0.17783
## [501,]    1  0.22438
## [502,]    1  0.06263
## [503,]    1  0.04527
## [504,]    1  0.06076
## [505,]    1  0.10959
## [506,]    1  0.04741
```

```r
y = array(Boston$medv,c(506,1))
```

```r
beta_hat = ginv(t(X)%*%X)%*%t(X)%*%y
```

```r
beta_hat
```

```
##            [,1]
## [1,] 24.0331062
## [2,] -0.4151903
```

```r
y_hat_mod = predict(mod)
head(y_hat_mod)
```

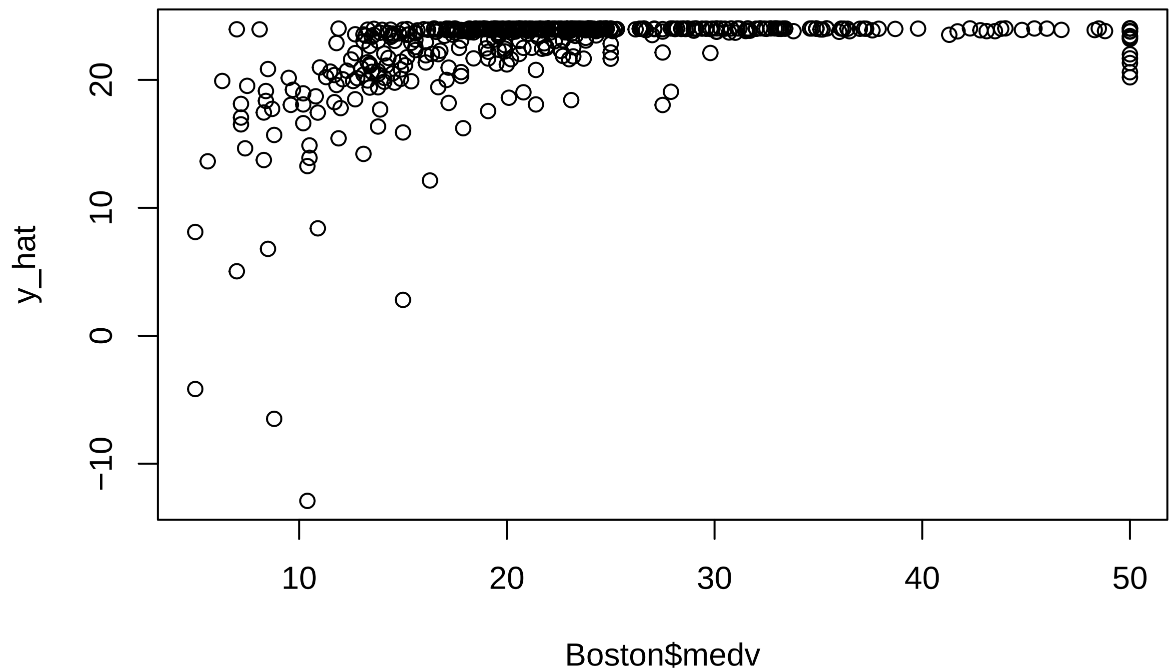
```
##        1        2        3        4        5        6
## 24.03048 24.02177 24.02178 24.01967 24.00444 24.02071
```

```r
y_hat = X%*%beta_hat
```

```r
head(y_hat)
```

```
##          [,1]
## [1,] 24.03048
## [2,] 24.02177
## [3,] 24.02178
## [4,] 24.01967
## [5,] 24.00444
## [6,] 24.02071
```

```r
plot(Boston$medv,y_hat)
```



More Regression

```
library('MASS')
data(Boston)
```

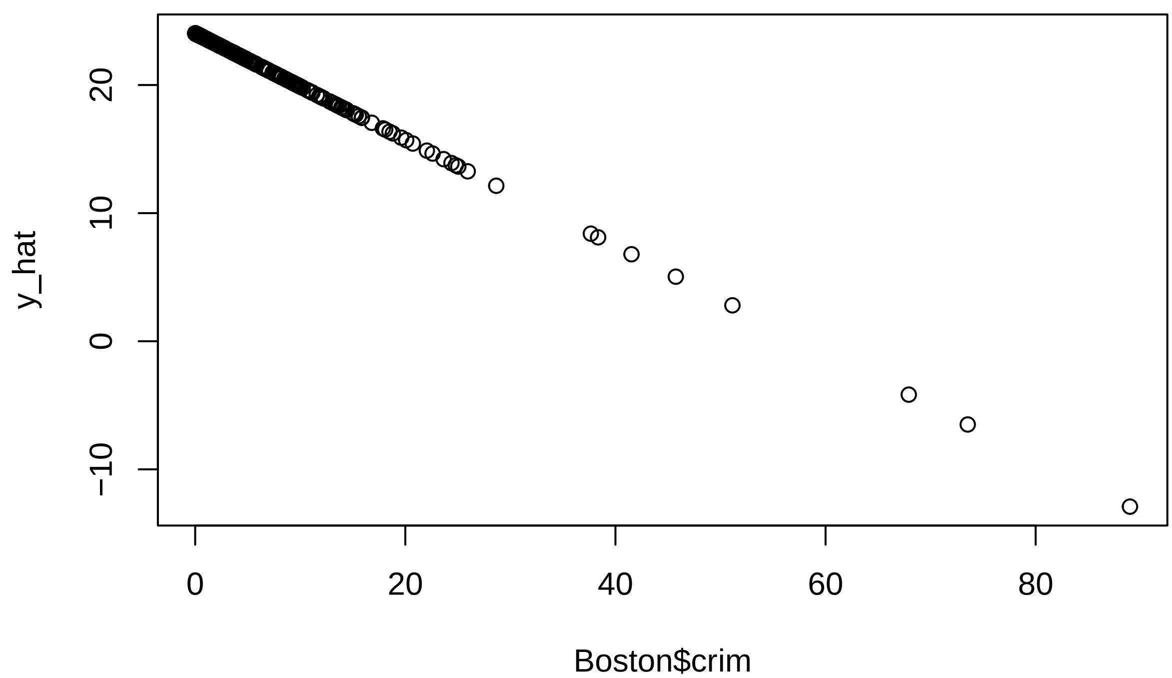
```
?Boston
```

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

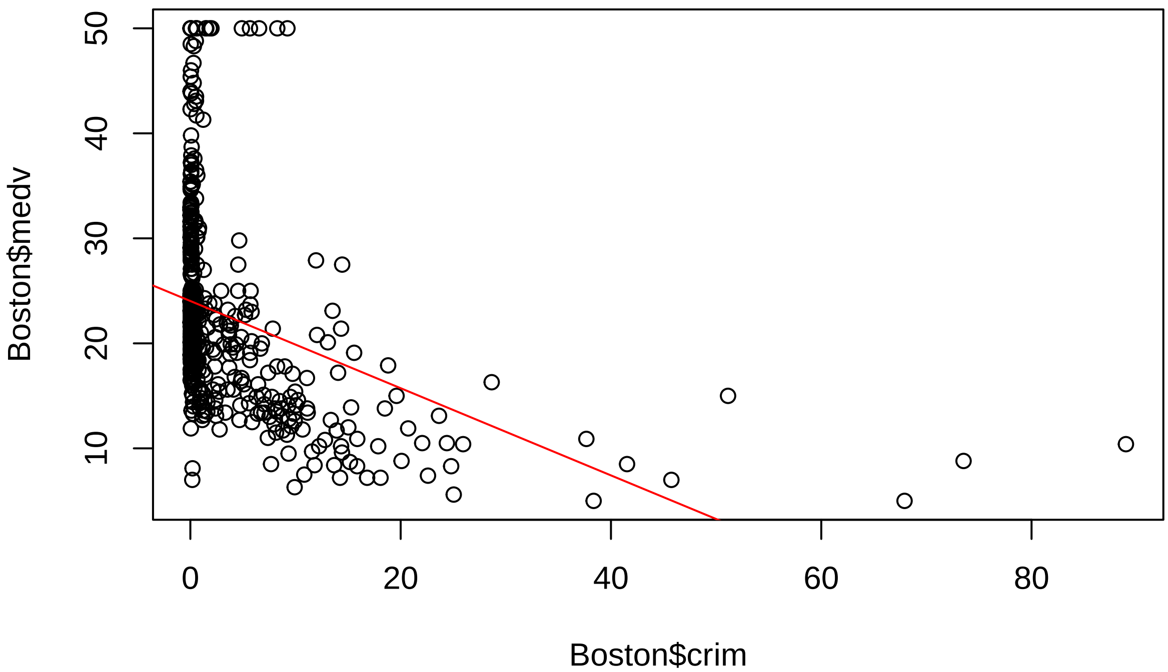
```
##
## Call:
## lm(formula = medv ~ crim, data = Boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.957  -5.449  -2.007   2.512  29.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.03311    0.40914   58.74   <2e-16 ***
## crim        -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
```

```
y_hat = predict(mod)
plot(Boston$crim,y_hat)
```

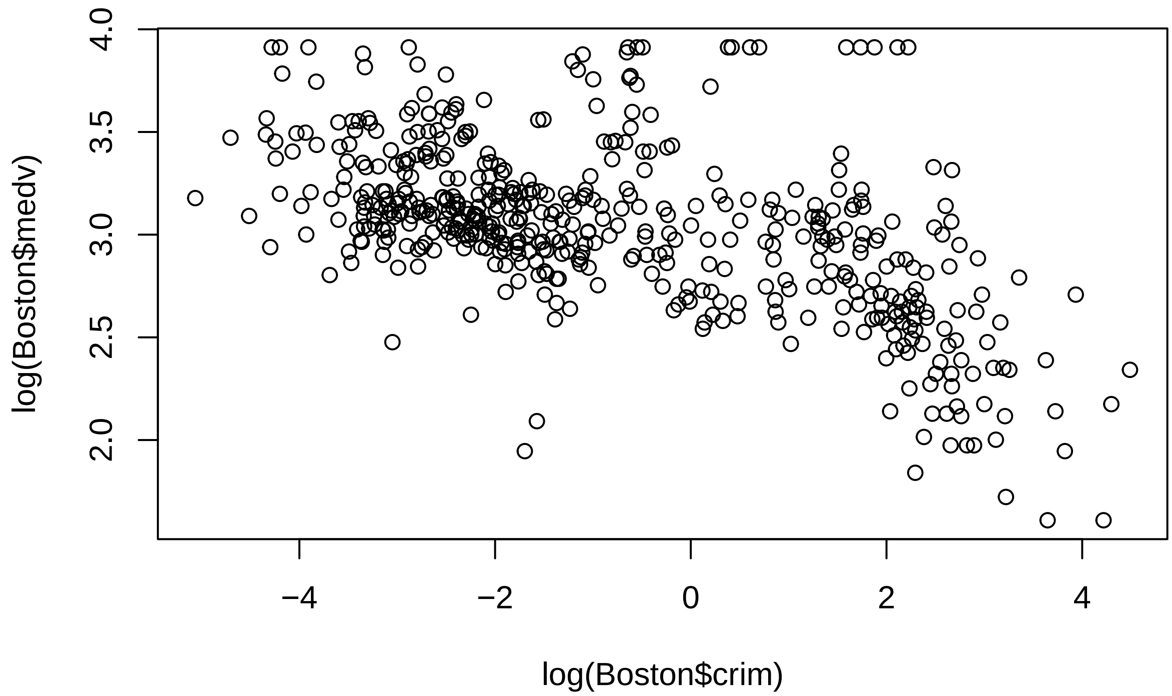



```
plot(Boston$crim,Boston$medv)
abline(coef=coef(mod),col='red')
```



covariate transformations

```
plot(log(Boston$crim),log(Boston$medv))
```



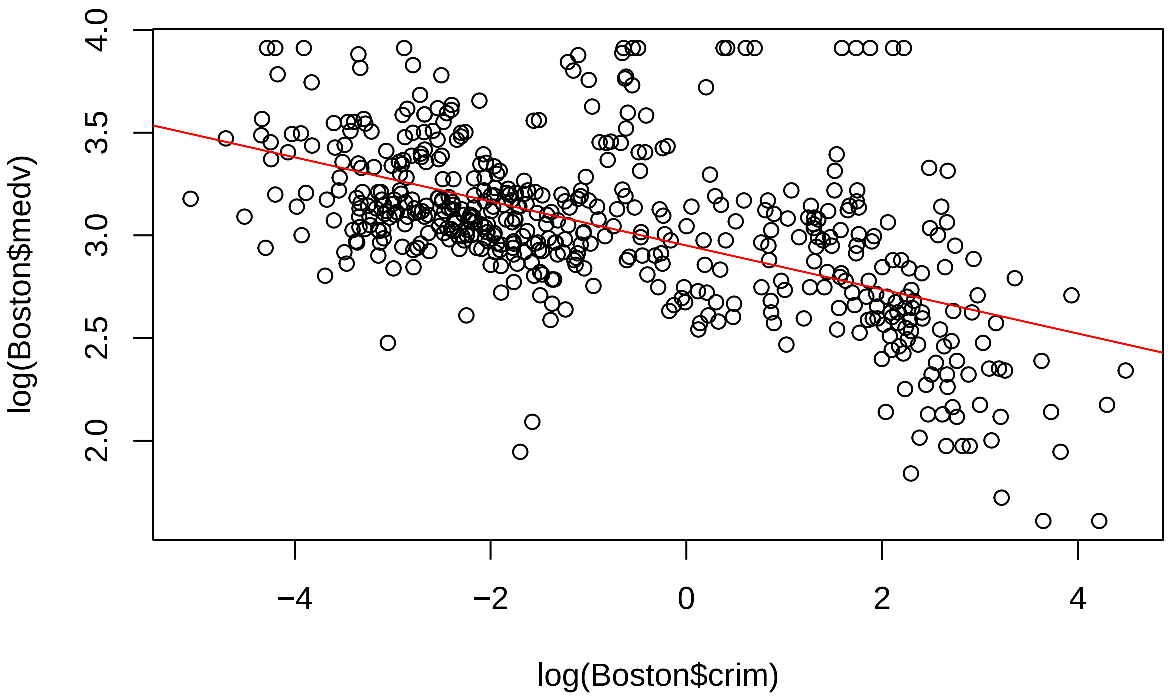
variable transformations

```
mod2 = lm(log(medv)~log(crim),data=Boston)
```

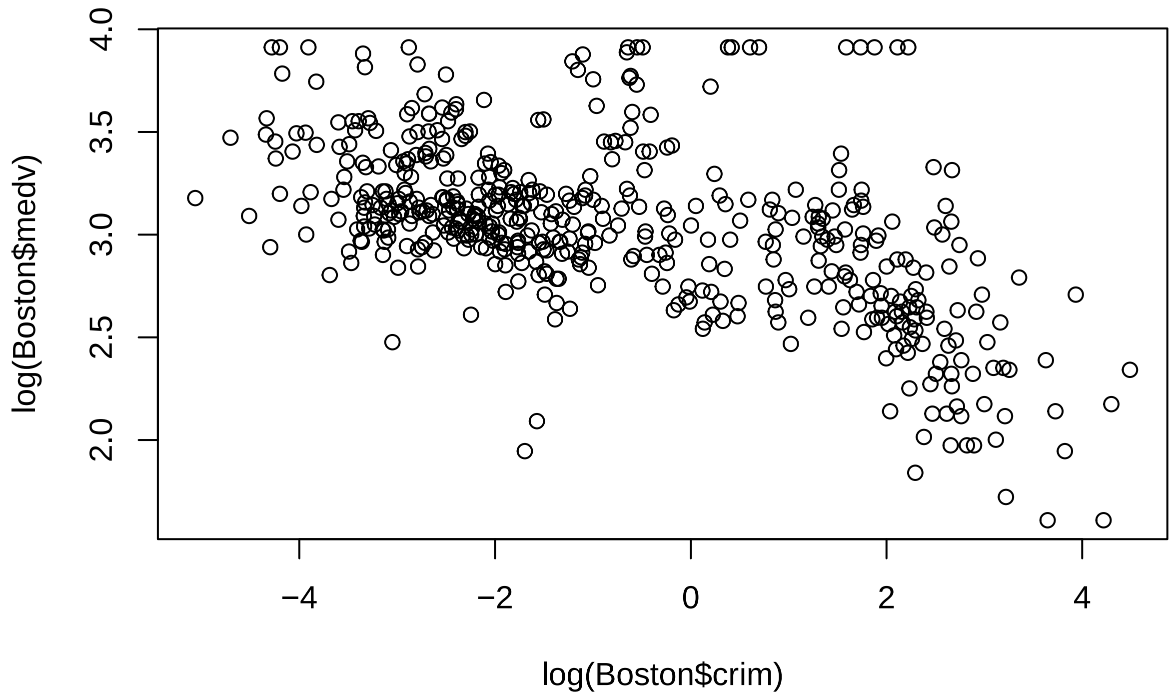
```
summary(mod2)
```

```
##
## Call:
## lm(formula = log(medv) ~ log(crim), data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18682 -0.19996 -0.05263  0.17103  1.19958
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.950817   0.015928  185.26   <2e-16 ***
## log(crim)   -0.107243   0.006935  -15.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.337 on 504 degrees of freedom
## Multiple R-squared:  0.3218, Adjusted R-squared:  0.3204
## F-statistic: 239.1 on 1 and 504 DF,  p-value: < 2.2e-16
```

```
plot(log(Boston$crim),log(Boston$medv))
abline(coef=coef(mod2),col='red')
```




```
plot(log(Boston$crim),log(Boston$medv))
```




```r
logmedv = array(log(Boston$medv),c(506,1))
transf_crim = log(Boston$crim)
```

```r
mod3 = lm(logmedv~transf_crim)
summary(mod3)
```

```
##
## Call:
## lm(formula = logmedv ~ transf_crim)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18682 -0.19996 -0.05263  0.17103  1.19958
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.950817   0.015928  185.26   <2e-16 ***
## transf_crim -0.107243   0.006935  -15.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.337 on 504 degrees of freedom
## Multiple R-squared:  0.3218, Adjusted R-squared:  0.3204
## F-statistic: 239.1 on 1 and 504 DF,  p-value: < 2.2e-16
```

```r
X = model.matrix(mod3)
head(X)
```

```
##   (Intercept) transf_crim
## 1           1   -5.064036
## 2           1   -3.600502
## 3           1   -3.601235
## 4           1   -3.430523
## 5           1   -2.672924
## 6           1   -3.511570
```

```r
head(log(Boston$crim))
```

```
## [1] -5.064036 -3.600502 -3.601235 -3.430523 -2.672924 -3.511570
```

```r
beta_hat = ginv(t(X)%*%X)%*%t(X)%*%logmedv
```

```r
coef(mod3)
```

```
## (Intercept) transf_crim
##   2.9508168  -0.1072427
```

categorical variables

```r
data(birthwt)
```

```r
?birthwt
```

```r
head(birthwt)
```

```
##    low age lwt race smoke ptl ht ui ftv  bwt
## 85   0  19 182    2     0   0  0  1   0 2523
## 86   0  33 155    3     0   0  0  0   3 2551
## 87   0  20 105    1     1   0  0  0   1 2557
## 88   0  21 108    1     1   0  0  1   2 2594
## 89   0  18 107    1     1   0  0  1   0 2600
## 91   0  21 124    3     0   0  0  0   0 2622
```

```r
head(birthwt$race)
```

```
## [1] 2 3 1 1 1 3
```

```r
racef = as.factor(birthwt$race)
head(racef)
```

```
## [1] 2 3 1 1 1 3
## Levels: 1 2 3
```

```r
levels(racef) = c("White","Black","Other")
```

```r
head(racef)
```

```
## [1] Black Other White White White Other
## Levels: White Black Other
```

```r
birthwt$race = racef
mod = lm(bwt~race,data=birthwt)
summary(mod)
```

```
##
## Call:
## lm(formula = bwt ~ race, data = birthwt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2096.28  -502.72   -12.72   526.28  1887.28
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3102.72      72.92  42.548  < 2e-16 ***
## raceBlack    -383.03     157.96  -2.425  0.01627 *
## raceOther    -297.44     113.74  -2.615  0.00965 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 714.5 on 186 degrees of freedom
## Multiple R-squared:  0.05017,    Adjusted R-squared:  0.03996
## F-statistic: 4.913 on 2 and 186 DF,  p-value: 0.008336
```

```r
head(model.matrix(mod))
```

```
##    (Intercept) raceBlack raceOther
## 85           1         1         0
## 86           1         0         1
## 87           1         0         0
## 88           1         0         0
## 89           1         0         0
## 91           1         0         1
```