CSCI 385 - Simple Models and Model Validation

Zoe Kunhart

Consider the data in `housing.csv` provided on Blackboard. This data (from here and originally from the UCI Machine Learning Repository) has information regarding homes in Boston from 1978. We will focus on a subset of the data containing four variables:

- `RM` - Average number of rooms per dwelling
- `LSTAT` - Fraction of population considered "lower status"
- `PTRATIO` - Pupil-teacher ratio by town
- `MEDV` - Median value of owner-occupied homes

Our goal is to create a linear model for predicting the median value of a home (`MDEV`) using the other variables.

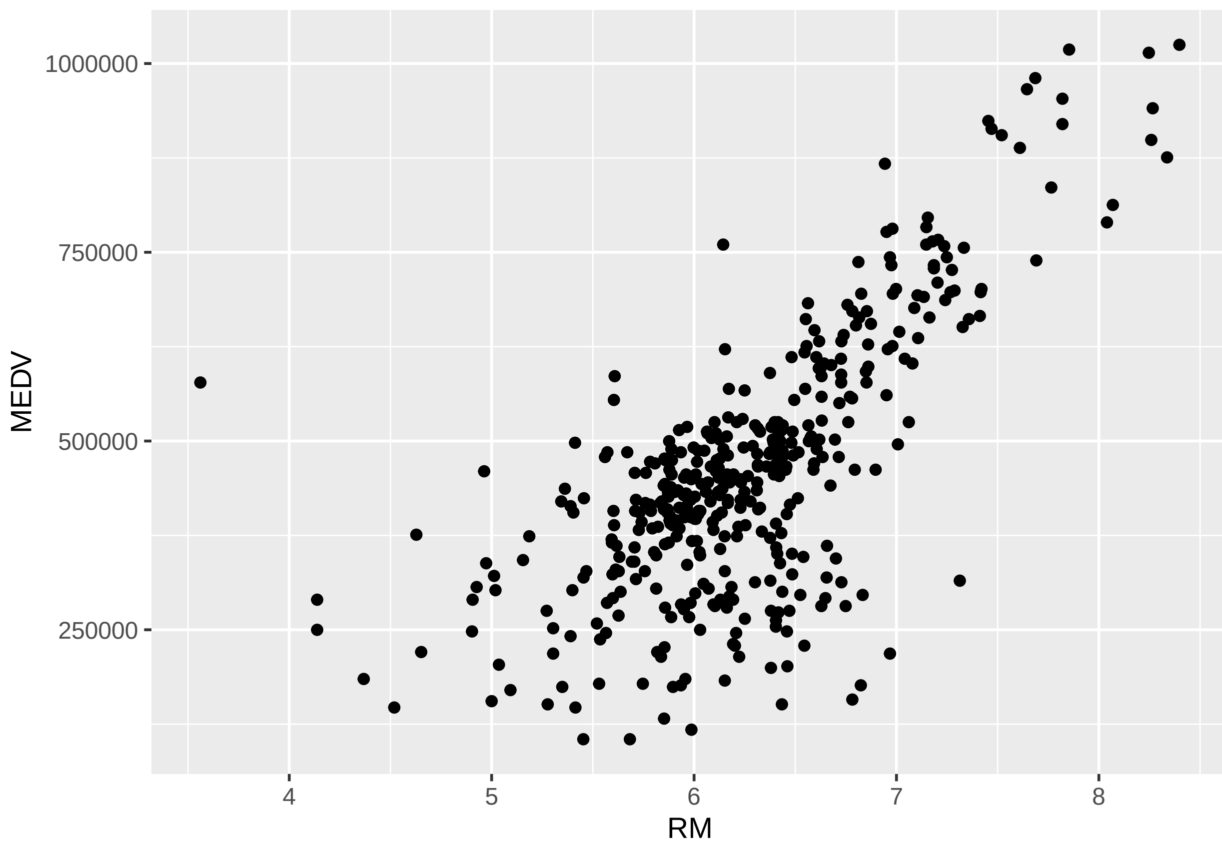
1) (10 pts) Start by setting aside roughly 20% of the data for final testing.

```
housing<-read.csv("/Users/zoekunhart/Downloads/housing.csv")
```

```
split=sample.split(housing$MEDV, SplitRatio=0.8)
training_set=subset(housing, split==TRUE)
test_set =subset(housing, split==FALSE)
```

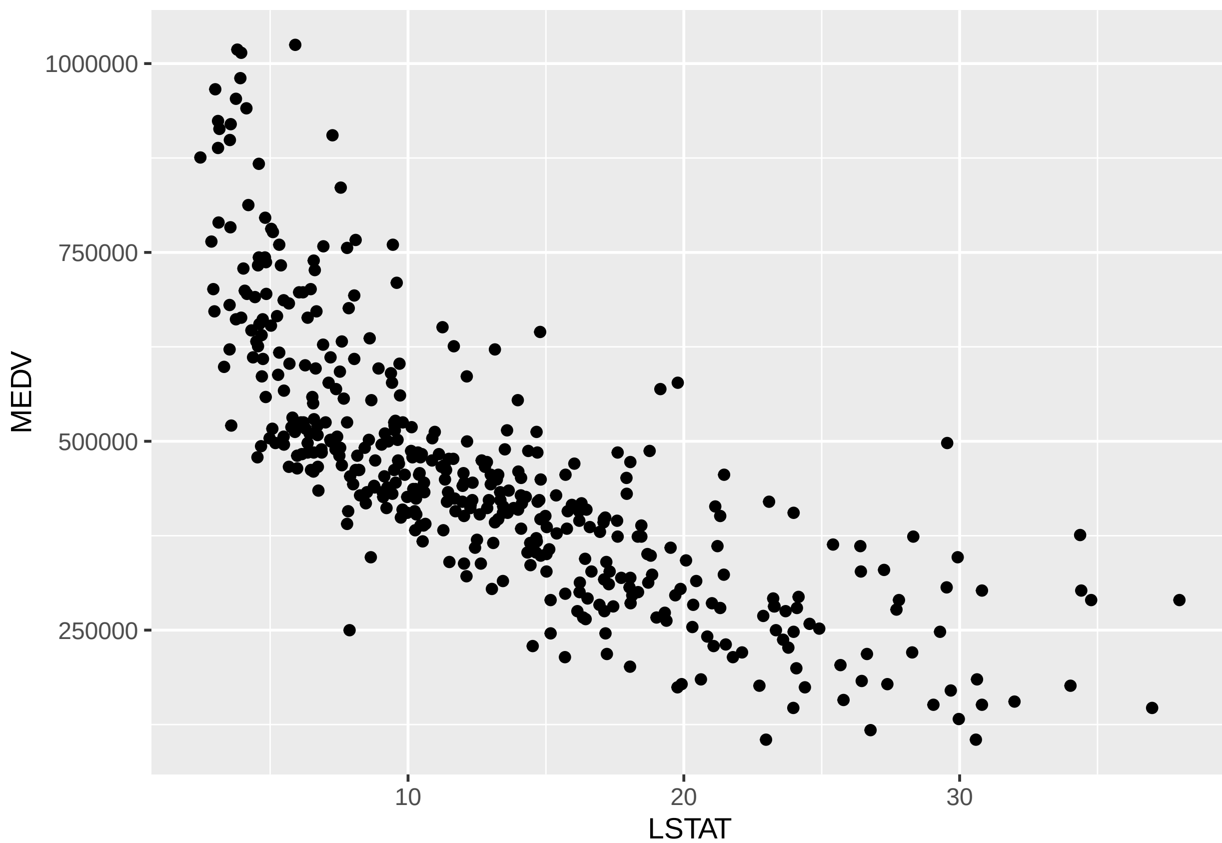
2) (20 pts) Using the remaining 80% of data, perform some initial exploratory data analysis. Based on what you find, how do you expect `MEDV` to behave with respect to the other variables? In other words, if `RM`, `LSTAT`, or `PTRatio` increases, what would you expect to happen to `MEDV`? You should not fit any models at this point.

```
library(ggplot2)
ggplot(training_set, aes(x=RM, y=MEDV))+
  geom_point()
```



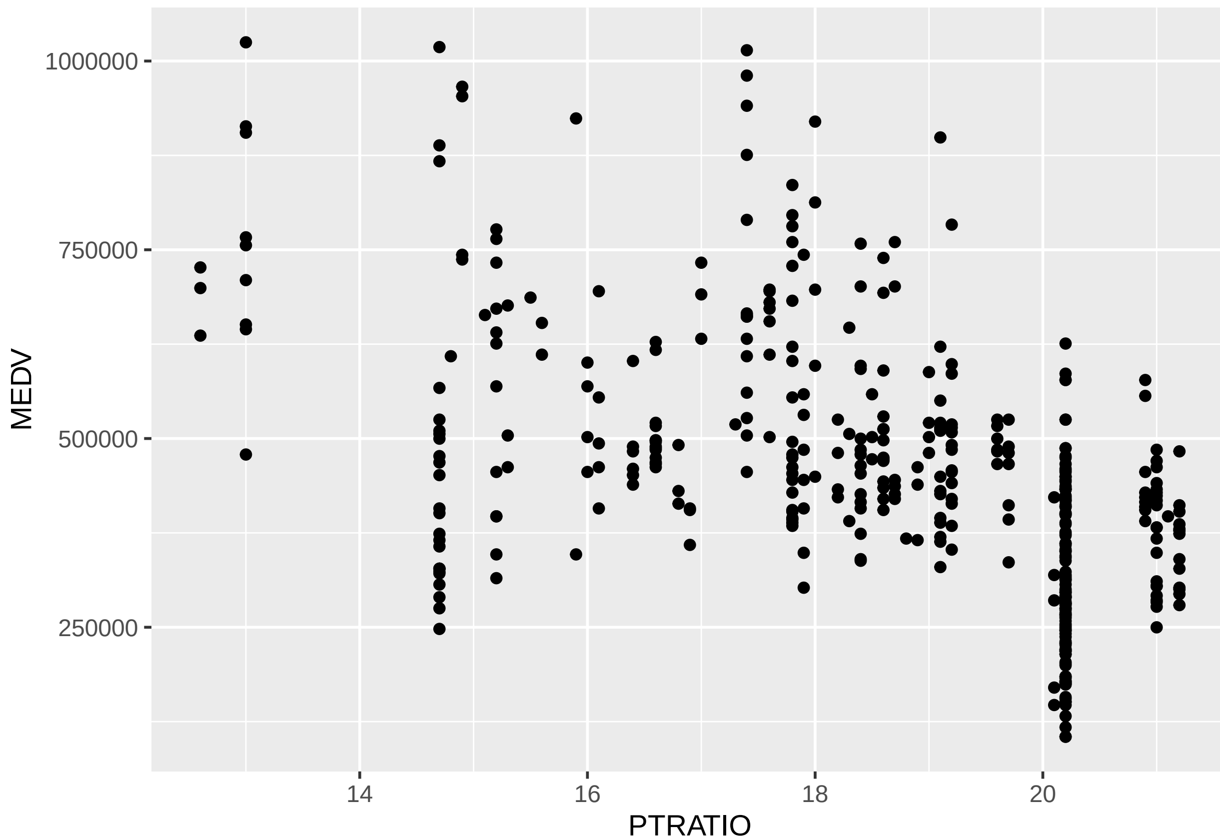
Average number of rooms per dwelling and median value of homes seems to have a pretty strong positive linear correlation. As RM increases, I expect MEDV to increase as well.

```
library(ggplot2)
ggplot(training_set, aes(x=LSTAT, y=MEDV))+
  geom_point()
```



The fraction of the population that are considered lower status and the median housing price seems to have a negative Curvilinear relationship. These types of relationships are very strong. As LSTAT increases, I expect MEDV to decrease.

```
library(ggplot2)
ggplot(training_set, aes(x=PTRATIO, y=MEDV))+
  geom_point()
```



Pupil to Teacher Ratio do not appear to have a correlation. it might be very hard to predict a model based on these two variables.

3) (30 pts) Using a single validation set consisting of 20% of your remaining data, validate a linear model for `MEDV`. Use the `add_predictions` function to generate predictions for the values in your validation set and create a figure with true `MEDV` on the x-axis and predicted values on the y-axis. Show residual plots for each variable along with a histogram of residual values, R^2, RMSE, and MAE values. Interpret your results.

```
split=sample.split(training_set$MEDV, SplitRatio=0.8)
training_set2=subset(training_set, split==TRUE)
validation=subset(training_set, split==FALSE)
```

```
model1<-lm(MEDV~RM+LSTAT+PTRATIO, data=validation )
model1
```

```
##
## Call:
## lm(formula = MEDV ~ RM + LSTAT + PTRATIO, data = validation)
##
## Coefficients:
## (Intercept)           RM        LSTAT      PTRATIO
##      109825       127128        -8099       -18228
```

```r
summary(model1)
```

```
##
## Call:
## lm(formula = MEDV ~ RM + LSTAT + PTRATIO, data = validation)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -215203  -58960  -16754   43157  329653
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   109825     176484   0.622 0.535496
## RM            127128      20745   6.128 3.06e-08 ***
## LSTAT          -8099       2215  -3.656 0.000454 ***
## PTRATIO       -18228       5183  -3.517 0.000718 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 97270 on 81 degrees of freedom
## Multiple R-squared:  0.7227, Adjusted R-squared:  0.7124
## F-statistic: 70.36 on 3 and 81 DF,  p-value: < 2.2e-16
```

```r
model2<-lm(MEDV~RM*LSTAT*PTRATIO, data=validation)
summary(model2)
```

```
##
## Call:
## lm(formula = MEDV ~ RM * LSTAT * PTRATIO, data = validation)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -141463  -39629   -4651   38440  192473
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5514312    1401573  -3.934 0.000181 ***
## RM                 944209     212183   4.450 2.86e-05 ***
## LSTAT              275216     127128   2.165 0.033495 *
## PTRATIO            262516      75955   3.456 0.000895 ***
## RM:LSTAT           -41040      21170  -1.939 0.056222 .
## RM:PTRATIO         -40299      11585  -3.479 0.000833 ***
## LSTAT:PTRATIO      -12344       6479  -1.905 0.060484 .
## RM:LSTAT:PTRATIO     1712       1079   1.587 0.116605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67700 on 77 degrees of freedom
## Multiple R-squared:  0.8723, Adjusted R-squared:  0.8607
## F-statistic: 75.15 on 7 and 77 DF,  p-value: < 2.2e-16
```

this is a better model than model 1 because the R2 correlation coefficient is higher than model 1's.

```r
predictions <- predict(model2, validation)
predictions
```

```
##        1        3        4       13       16       17       19       39 
## 600240.7 706280.5 673436.7 415571.6 453772.3 486183.4 395122.3 436227.5 
##       43       52       54       60       62       78       81       95 
## 502120.5 461366.6 449477.2 443812.3 386451.0 451680.3 586667.8 461874.2 
##      105      129      139      140      141      147      154      157 
## 413375.2 367169.0 272822.4 314326.0 224360.9 398170.8 402566.6 361897.4 
##      161      162      163      166      174      182      185      188 
## 513478.5 413805.3 460723.8 507848.6 522467.3 429127.1 749379.7 798587.7 
##      193      196      198      201      205      213      214      221 
## 822791.3 482358.0 989704.7 357248.9 322220.6 493011.3 569853.2 819055.3 
##      225      226      229      233      238      242      253      256 
## 996950.1 929661.3 529473.4 551165.5 342748.7 563111.5 866659.8 760179.2 
##      262      280      281      286      292      293      299      302 
## 505438.4 636126.5 672593.3 560561.1 534845.5 688965.2 437882.0 426950.3 
##      308      318      322      326      327      329      335      337 
## 476989.8 422009.1 458979.4 434431.6 426886.6 442953.7 436628.1 551067.7 
##      346      358      365      375      382      390      405      417 
## 346110.4 294450.6 294464.8 328124.3 212060.4 307053.8 367292.5 371474.7 
##      420      421      425      434      442      445      447      449 
## 343063.5 222328.8 319509.8 357779.8 367902.1 394009.2 469539.7 377994.7 
##      451      461      463      479      488 
## 294639.0 269803.8 410923.3 349597.4 543384.6
```

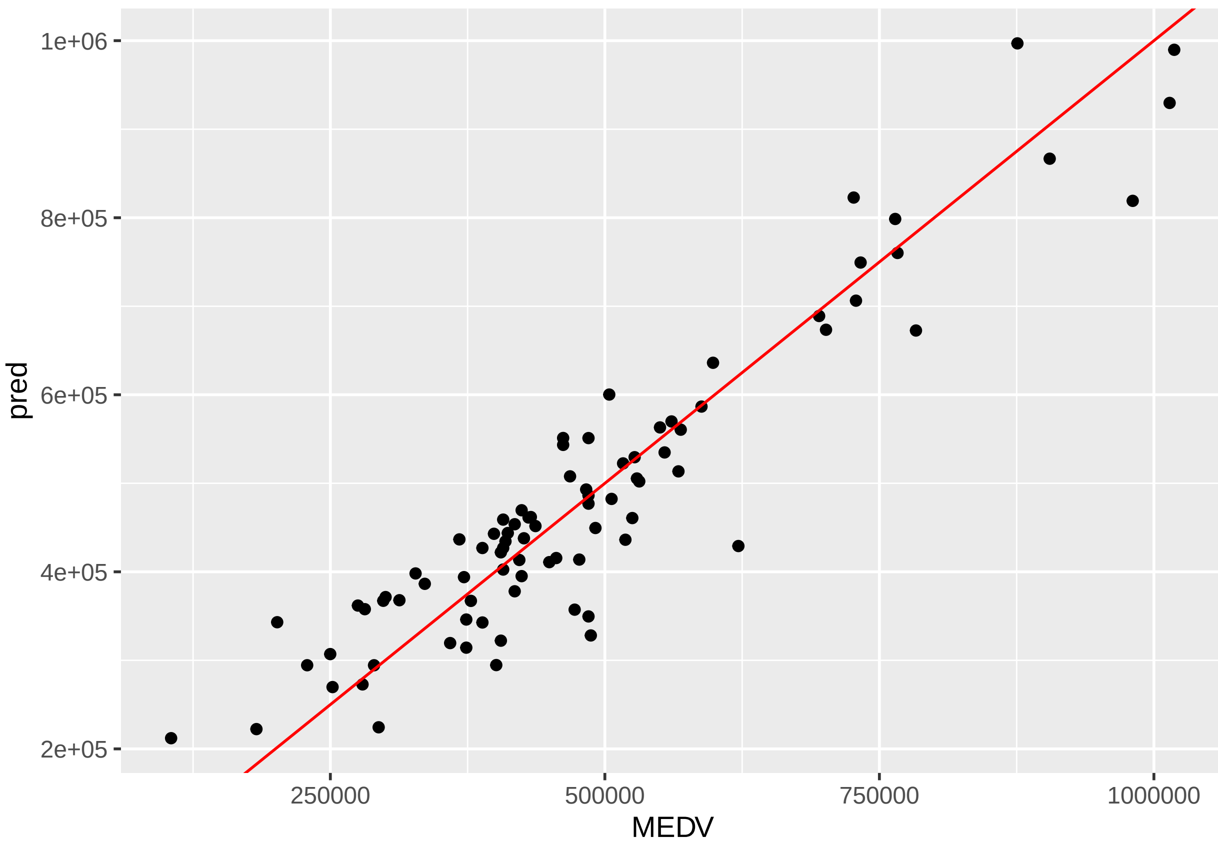
```r
predictions <- add_predictions(validation, model2)
predictions
```

```
##        RM LSTAT PTRATIO   MEDV     pred
## 1   6.575  4.98    15.3 504000 600240.7
## 3   7.185  4.03    17.8 728700 706280.5
## 4   6.998  2.94    18.7 701400 673436.7
## 13  5.889 15.71    15.2 455700 415571.6
## 16  5.834  8.47    21.0 417900 453772.3
## 17  5.935  6.58    21.0 485100 486183.4
## 19  5.456 11.69    21.0 424200 395122.3
## 39  5.966 10.13    19.2 518700 436227.5
## 43  6.169  5.81    17.9 531300 502120.5
## 52  6.115  9.43    16.8 430500 461366.6
## 54  5.998  8.43    16.8 491400 449477.2
## 60  5.927  9.22    19.7 411600 443812.3
## 62  5.966 14.44    19.7 336000 386451.0
## 78  6.140 10.27    18.7 436800 451680.3
## 81  6.727  5.29    19.0 588000 586667.8
## 95  6.249 10.59    18.2 432600 461874.2
## 105 6.167 12.33    20.9 422100 413375.2
## 129 6.431 15.39    21.2 378000 367169.0
## 139 5.857 21.32    21.2 279300 272822.4
## 140 6.151 18.46    21.2 373800 314326.0
## 141 6.174 24.16    21.2 294000 224360.9
## 147 5.628 16.65    14.7 327600 398170.8
```

```
## 154 5.709 15.79    14.7  407400 402566.6
## 157 5.272 16.14    14.7  275100 361897.4
## 161 6.250  5.50    14.7  567000 513478.5
## 162 5.854 11.64    14.7  476700 413805.3
## 163 6.101  9.81    14.7  525000 460723.8
## 166 6.402 11.32    14.7  468300 507848.6
## 174 6.315  6.29    16.6  516600 522467.3
## 182 6.153 13.15    17.8  621600 429127.1
## 185 7.185  5.39    15.2  732900 749379.7
## 188 7.178  2.87    15.2  764400 798587.7
## 193 7.274  6.62    12.6  726600 822791.3
## 196 6.162  7.43    14.7  506100 482358.0
## 198 7.853  3.81    14.7 1018500 989704.7
## 201 5.783 18.06    18.6  472500 357248.9
## 205 5.404 23.98    18.6  405300 322220.6
## 213 6.373 10.50    16.4  483000 493011.3
## 214 6.951  9.71    17.4  560700 569853.2
## 221 7.686  3.92    17.4  980700 819055.3
## 225 8.337  2.47    17.4  875700 996950.1
## 226 8.247  3.95    17.4 1014300 929661.3
## 229 6.631  9.54    17.4  527100 529473.4
## 233 6.897 11.38    16.6  462000 551165.5
## 238 5.605 18.46    19.1  388500 342748.7
## 242 6.718  6.56    19.1  550200 563111.5
## 253 7.520  7.26    13.0  905100 866659.8
## 256 7.206  8.10    13.0  766500 760179.2
## 262 6.240  6.59    18.6  529200 505438.4
## 280 6.861  3.33    19.2  598500 636126.5
## 281 7.148  3.56    19.2  783300 672593.3
## 286 6.549  7.39    16.0  569100 560561.1
## 292 6.495  8.67    16.1  554400 534845.5
## 293 6.982  4.86    16.1  695100 688965.2
## 299 5.972  9.97    18.4  426300 437882.0
## 302 6.023 11.72    18.4  407400 426950.3
## 308 6.382 10.36    18.4  485100 476989.8
## 318 5.868  9.97    16.9  405300 422009.1
## 322 6.031  7.83    16.9  407400 458979.4
## 326 5.869  9.80    20.2  409500 434431.6
## 327 5.895 10.56    20.2  388500 426886.6
## 329 5.985  9.74    20.2  399000 442953.7
## 335 6.014 10.53    18.8  367500 436628.1
## 337 6.516  6.36    17.9  485100 551067.7
## 346 6.212 17.60    20.2  373800 346110.4
## 358 4.138 37.97    20.2  289800 294450.6
## 365 6.545 21.08    20.2  228900 294464.8
## 375 6.051 18.76    20.2  487200 328124.3
## 382 5.453 30.59    20.2  105000 212060.4
## 390 4.138 23.34    20.2  249900 307053.8
## 405 6.006 15.70    20.2  298200 367292.5
## 417 6.436 16.22    20.2  300300 371474.7
## 420 6.461 18.05    20.2  201600 343063.5
## 421 6.152 26.45    20.2  182700 222328.8
## 425 6.406 19.52    20.2  359100 319509.8
## 434 6.749 17.44    20.2  281400 357779.8
```

```
## 442 6.301 16.23    20.2  312900 367902.1
## 445 6.376 14.65    20.2  371700 394009.2
## 447 6.513 10.29    20.2  424200 469539.7
## 449 5.759 14.13    20.2  417900 377994.7
## 451 6.003 21.32    20.2  401100 294639.0
## 461 5.304 24.91    20.2  252000 269803.8
## 463 6.229 13.11    20.2  449400 410923.3
## 479 5.670 17.60    19.2  485100 349597.4
## 488 6.794  6.48    21.0  462000 543384.6
```

```
ggplot(data = predictions, mapping = aes(x = MEDV, y = pred)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red")
```




```
R2(predictions$pred, predictions$MEDV)
```

```
## [1] 0.8723095
```

```
MAE(predictions$pred, predictions$MEDV)
```

```
## [1] 49854.41
```

```r
RMSE(predictions$pred, predictions$MEDV)
```

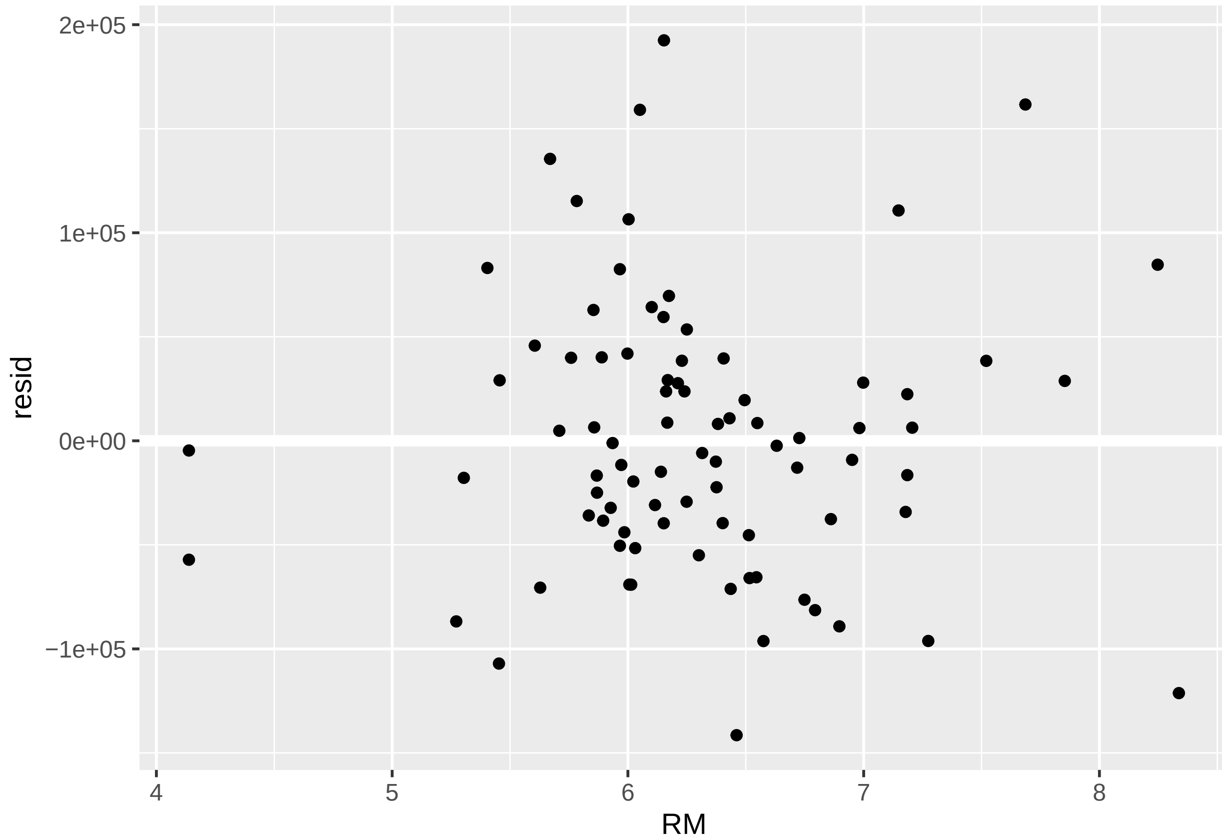
```
## [1] 64433.73
```

```r
resids <- add_residuals(validation, model2)
resids
```

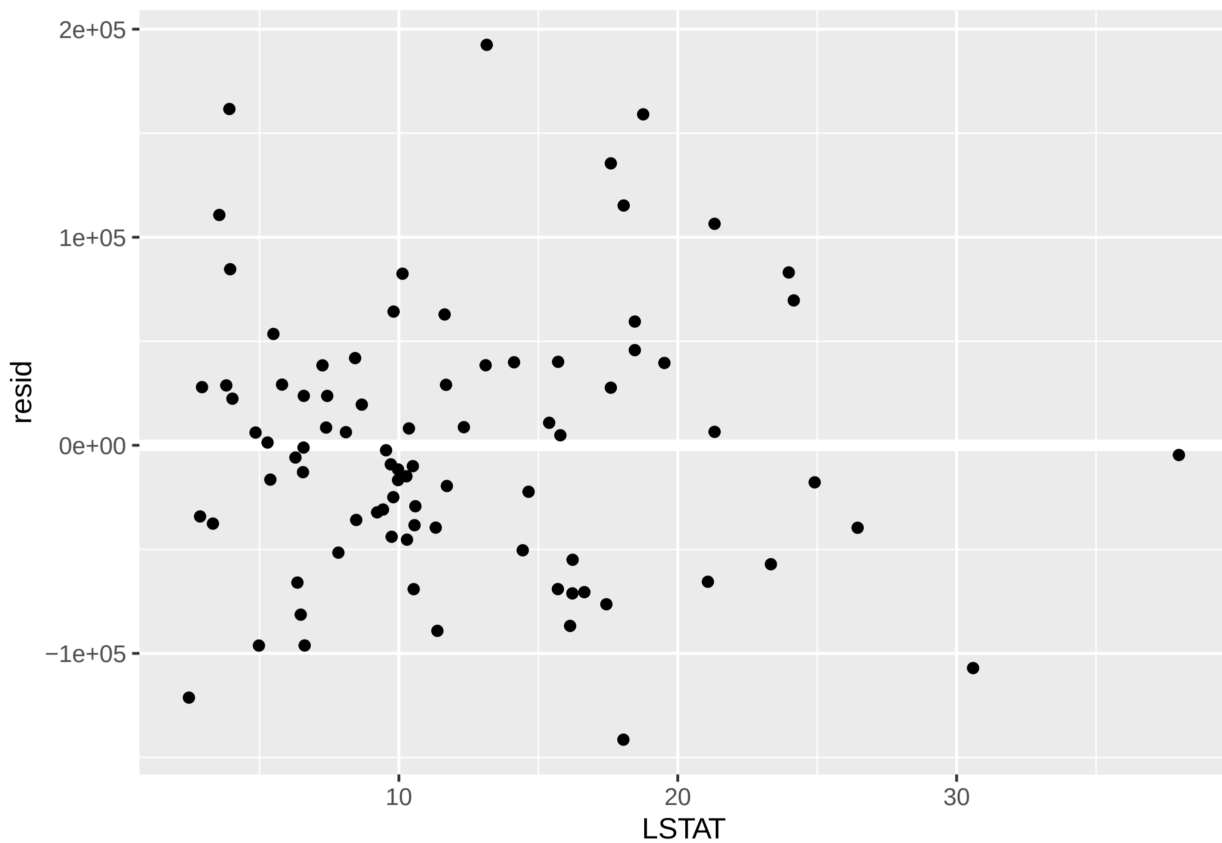
```
##        RM LSTAT PTRATIO    MEDV       resid
## 1   6.575  4.98    15.3  504000  -96240.731
## 3   7.185  4.03    17.8  728700   22419.540
## 4   6.998  2.94    18.7  701400   27963.287
## 13  5.889 15.71    15.2  455700   40128.418
## 16  5.834  8.47    21.0  417900  -35872.266
## 17  5.935  6.58    21.0  485100   -1083.368
## 19  5.456 11.69    21.0  424200   29077.691
## 39  5.966 10.13    19.2  518700   82472.520
## 43  6.169  5.81    17.9  531300   29179.487
## 52  6.115  9.43    16.8  430500  -30866.623
## 54  5.998  8.43    16.8  491400   41922.808
## 60  5.927  9.22    19.7  411600  -32212.348
## 62  5.966 14.44    19.7  336000  -50451.019
## 78  6.140 10.27    18.7  436800  -14880.329
## 81  6.727  5.29    19.0  588000    1332.175
## 95  6.249 10.59    18.2  432600  -29274.193
## 105 6.167 12.33    20.9  422100    8724.817
## 129 6.431 15.39    21.2  378000   10830.967
## 139 5.857 21.32    21.2  279300    6477.631
## 140 6.151 18.46    21.2  373800   59474.045
## 141 6.174 24.16    21.2  294000   69639.096
## 147 5.628 16.65    14.7  327600  -70570.839
## 154 5.709 15.79    14.7  407400    4833.375
## 157 5.272 16.14    14.7  275100  -86797.431
## 161 6.250  5.50    14.7  567000   53521.453
## 162 5.854 11.64    14.7  476700   62894.672
## 163 6.101  9.81    14.7  525000   64276.181
## 166 6.402 11.32    14.7  468300  -39548.639
## 174 6.315  6.29    16.6  516600   -5867.280
## 182 6.153 13.15    17.8  621600  192472.889
## 185 7.185  5.39    15.2  732900  -16479.718
## 188 7.178  2.87    15.2  764400  -34187.683
## 193 7.274  6.62    12.6  726600  -96191.270
## 196 6.162  7.43    14.7  506100   23741.967
## 198 7.853  3.81    14.7 1018500   28795.340
## 201 5.783 18.06    18.6  472500  115251.055
## 205 5.404 23.98    18.6  405300   83079.388
## 213 6.373 10.50    16.4  483000  -10011.341
## 214 6.951  9.71    17.4  560700   -9153.222
## 221 7.686  3.92    17.4  980700  161644.698
## 225 8.337  2.47    17.4  875700 -121250.146
## 226 8.247  3.95    17.4 1014300   84638.656
## 229 6.631  9.54    17.4  527100   -2373.370
## 233 6.897 11.38    16.6  462000  -89165.503
## 238 5.605 18.46    19.1  388500   45751.260
```

```
## 242 6.718  6.56    19.1  550200  -12911.494
## 253 7.520  7.26    13.0  905100   38440.211
## 256 7.206  8.10    13.0  766500    6320.785
## 262 6.240  6.59    18.6  529200   23761.551
## 280 6.861  3.33    19.2  598500  -37626.462
## 281 7.148  3.56    19.2  783300  110706.703
## 286 6.549  7.39    16.0  569100    8538.916
## 292 6.495  8.67    16.1  554400   19554.487
## 293 6.982  4.86    16.1  695100    6134.774
## 299 5.972  9.97    18.4  426300  -11581.968
## 302 6.023 11.72    18.4  407400  -19550.334
## 308 6.382 10.36    18.4  485100    8110.246
## 318 5.868  9.97    16.9  405300  -16709.146
## 322 6.031  7.83    16.9  407400  -51579.434
## 326 5.869  9.80    20.2  409500  -24931.576
## 327 5.895 10.56    20.2  388500  -38386.606
## 329 5.985  9.74    20.2  399000  -43953.678
## 335 6.014 10.53    18.8  367500  -69128.125
## 337 6.516  6.36    17.9  485100  -65967.741
## 346 6.212 17.60    20.2  373800   27689.598
## 358 4.138 37.97    20.2  289800   -4650.578
## 365 6.545 21.08    20.2  228900  -65564.809
## 375 6.051 18.76    20.2  487200  159075.715
## 382 5.453 30.59    20.2  105000 -107060.434
## 390 4.138 23.34    20.2  249900  -57153.840
## 405 6.006 15.70    20.2  298200  -69092.534
## 417 6.436 16.22    20.2  300300  -71174.694
## 420 6.461 18.05    20.2  201600 -141463.496
## 421 6.152 26.45    20.2  182700  -39628.757
## 425 6.406 19.52    20.2  359100   39590.238
## 434 6.749 17.44    20.2  281400  -76379.843
## 442 6.301 16.23    20.2  312900  -55002.132
## 445 6.376 14.65    20.2  371700  -22309.186
## 447 6.513 10.29    20.2  424200  -45339.734
## 449 5.759 14.13    20.2  417900   39905.312
## 451 6.003 21.32    20.2  401100  106460.999
## 461 5.304 24.91    20.2  252000  -17803.765
## 463 6.229 13.11    20.2  449400   38476.741
## 479 5.670 17.60    19.2  485100  135502.617
## 488 6.794  6.48    21.0  462000  -81384.622
```

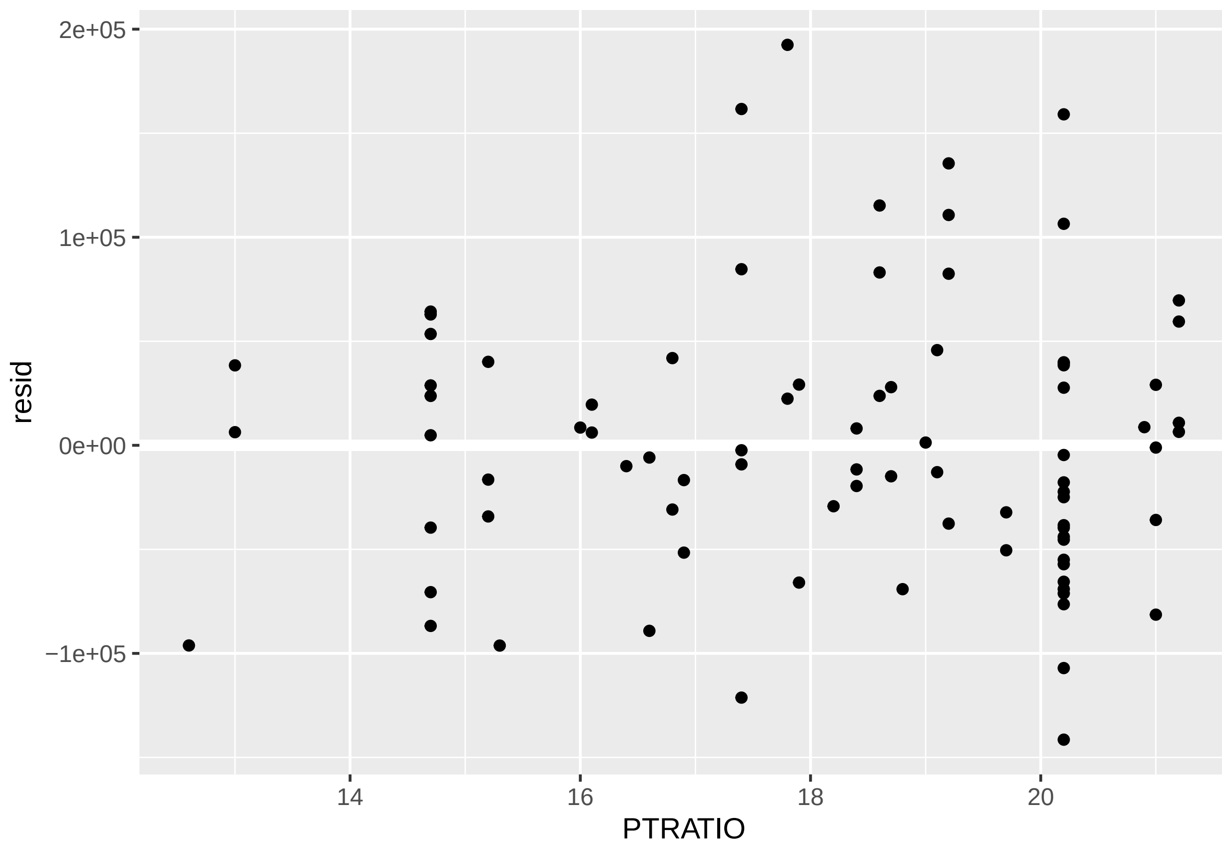
```
ggplot(data=resids, mapping=aes(x=RM, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```




```
ggplot(data=resids, mapping=aes(x=LSTAT, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```

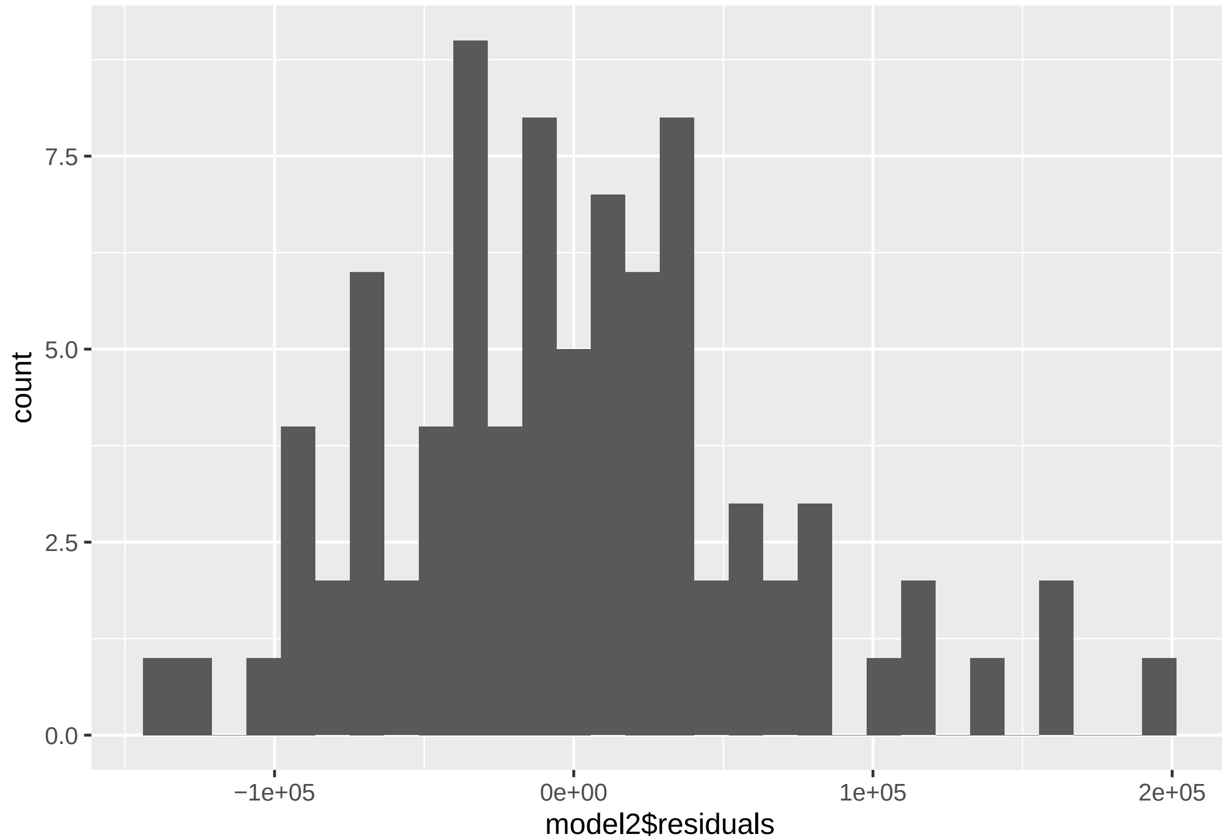



```
ggplot(data=resids, mapping=aes(x=PTRATIO, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```




```
ggplot(data=validation, aes(x=model2$residuals))+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



4) (25 pts) Using k-fold cross-validation with $k = 5$, validate a linear model for `MEDV`. Use the `add_predictions` function to generate predictions for the values in your training set and create a figure with true `MEDV` on the x-axis and predicted values on the y-axis. Show residual plots for each variable along with a histogram of residual values, R^2, RMSE, and MAE values. Interpret your results. How do these results compare to those from the previous problem?

```
train.control <- trainControl(method = "cv", number = 5)
model2 <- train( MEDV~RM*LSTAT*PTRATIO,
               data = training_set2,
               method = "lm",
               trControl = train.control)
model2
```

```
## Linear Regression
##
## 336 samples
##   3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 268, 270, 268, 268, 270
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   77301.13  0.8039674  56949.06
```

```
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -225237  -46235   -3691   41240  283884
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3285202.3   617420.6  -5.321 1.92e-07 ***
## RM                  616919.5    92357.2   6.680 1.03e-10 ***
## LSTAT               120759.4    40293.1   2.997  0.00293 **
## PTRATIO             154513.0    33714.2   4.583 6.52e-06 ***
## `RM:LSTAT`          -18271.7     6873.0  -2.659  0.00823 **
## `RM:PTRATIO`        -24359.8     5090.3  -4.786 2.58e-06 ***
## `LSTAT:PTRATIO`      -4718.3     2111.6  -2.234  0.02613 *
## `RM:LSTAT:PTRATIO`     578.9      359.9   1.608  0.10872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72840 on 328 degrees of freedom
## Multiple R-squared:  0.8224, Adjusted R-squared:  0.8186
## F-statistic:   217 on 7 and 328 DF,  p-value: < 2.2e-16
```

```r
predictions2 <- predict(model2, training_set2)
predictions2
```

```
##         2        5        8        9       10       11       12       14
##  510698.3 647354.2 405234.5 323578.4 410332.0 403833.1 440939.6 464918.0
##        15       18       20       21       22       24       26       27
##  441773.1 371299.8 412431.2 286065.0 383343.0 296248.3 342060.1 366586.9
##        28       30       32       33       35       37       38       40
##  332247.4 433489.8 397778.8 178838.7 284483.8 420204.7 451706.8 600937.5
##        42       45       48       49       50       51       55       56
##  621654.9 466573.1 357064.0 259854.1 364117.1 421590.6 367009.6 690589.5
##        57       58       59       61       63       64       65       66
##  557404.4 684699.6 502020.2 392605.0 532133.8 508312.2 589043.9 562228.1
##        67       68       69       70       71       72       74       77
##  431174.8 461901.3 387277.5 454605.3 533028.9 447430.5 503722.9 447223.7
##        79       80       82       83       84       85       86       87
##  438691.7 450148.0 547476.3 522639.5 497580.5 492726.2 566769.3 418814.6
##        89       90       91       92       93       94       96       97
##  644811.0 650928.5 514993.6 522334.6 523088.3 523243.9 570561.2 453329.1
##        98       99      100      101      102      103      104      106
##  820002.0 800167.0 680711.9 486728.0 523908.1 449695.9 394167.3 344727.7
##       108      109      111      112      113      114      115      116
##  383729.2 423383.8 402690.0 526118.2 381467.3 380807.8 475142.1 386986.3
```

```
##       117       118       119       120       122       123       124       125 
##  447566.0  453509.7  387657.7  394723.2  395541.5  348173.6  258972.9  350469.0 
##       126       127       128       130       131       132       133       134 
##  387887.9  248433.8  331814.0  316910.3  412478.5  415047.3  436234.0  362141.5 
##       135       136       142       144       145       146       148       149 
##  329988.1  334093.2  172059.8  349120.9  358753.0  338075.0  358924.3  350332.4 
##       150       151       153       155       156       158       159       164 
##  365336.8  452765.4  313357.3  445012.4  448436.4  708021.4  505939.6  434781.4 
##       167       168       169       170       171       172       173       175 
##  420868.5  435862.6  378364.2  524609.6  444847.7  594996.8  461058.8  620227.6 
##       176       177       178       179       180       181       183       184 
##  649711.7  698849.5  476316.4  679216.2  580082.6  382299.3  634622.7  620873.2 
##       186       187       189       190       191       192       194       195 
##  692163.6  656181.9  657263.8  628959.8  844288.3  720851.5  672820.6  701274.7 
##       197       199       200       202       203       206       207       208 
##  884652.1  431985.7  466318.2  398805.6  300159.1  370480.9  493912.4  250315.5 
##       210       211       212       215       217       218       219       220 
##  416752.0  548282.4  383906.9  338407.3  562416.0  878619.7  870193.8  658559.7 
##       222       224       227       228       230       231       232       235 
##  611620.4  718191.4  569117.1  455992.3  721358.7  570065.7  574811.5  489268.7 
##       236       237       239       241       243       244       245       246 
##  571353.2  392726.2  469235.8  491362.6  554360.5  585540.7  651975.9  815970.7 
##       247       248       249       250       252       254       255       257 
##  509544.6  450503.8  815027.7  774162.6  713641.8 1062719.1  706026.6  389384.0 
##       258       259       260       261       264       265       266       267 
##  593416.0  900379.8  402599.7  405787.1  686086.2  646301.1  671355.7  674579.8 
##       268       269       270       271       272       273       274       275 
##  646157.2  548507.7  673021.6  909079.5  709698.3  888452.0  667257.4  554481.3 
##       276       278       279       282       283       284       285       289 
##  436184.4  528716.6  537827.8  589482.2  493582.9  459750.8  610575.7  724390.3 
##       291       294       295       296       297       298       300       301 
##  548529.1  631640.8  526840.5  663361.0  575841.4  601709.5  346804.1  514612.8 
##       303       304       305       306       307       310       311       314 
##  505834.2  516273.4  410651.5  351613.4  371966.6  522658.3  523417.9  539112.6 
##       315       316       317       321       323       324       325       328 
##  557575.3  528650.1  414509.7  404973.7  533079.3  515369.8  475434.4  469817.3 
##       331       332       333       334       340       341       342       343 
##  734626.8  555805.7  576506.2  645544.1  547794.2  579891.8  465490.4  633194.4 
##       347       348       349       350       352       353       354       355 
##  413849.1  430376.8  412075.1  505144.5  405310.5  374182.0  294966.9  354812.0 
##       357       359       360       362       363       364       366       367 
##  211205.2  444474.6  241033.8  241139.3  275586.7  355037.2  268281.7  258339.2 
##       368       369       370       371       373       374       376       377 
##  280691.8  209355.2  271035.4  222131.2  300242.2  341644.5  267507.2  377210.4 
##       378       379       380       381       383       384       386       387 
##  353875.0  350793.7  312444.8  307340.3  175564.2  210396.9  296769.4  311264.3 
##       388       389       391       392       393       395       396       397 
##  227096.6  271213.3  397335.9  233615.3  303249.0  277862.3  242081.4  309650.3 
##       398       399       400       401       402       403       404       407 
##  241950.6  149010.5  188434.4  246628.9  296423.9  245849.4  385181.8  255602.8 
##       408       409       410       411       412       413       415       416 
##  339266.7  246802.5  361556.4  387413.3  280115.0  234696.2  305104.1  435828.6 
##       418       419       422       423       424       426       427       428 
##  377586.3  241049.6  114509.8  273843.9  278816.9  355906.4  321189.1  256133.7 
```

```
##       429       431       433       435       438       439       440       441
##  233749.4  358760.1  313419.8  341862.9  324029.0  333804.8  318538.4  346698.5
##       443       444       446       448       450       453       454       455
##  381244.2  366649.4  399323.1  407633.3  344035.0  370157.0  359625.0  413825.6
##       456       457       459       460       462       466       467       468
##  397010.1  468773.1  241709.5  324113.0  333429.3  570326.2  425573.2  393247.1
##       469       470       471       472       473       474       475       476
##  453969.4  378037.3  419826.9  329205.9  269763.3  234985.7  333002.4  398061.3
##       477       478       480       481       484       485       486       489
##  404746.0  399249.2  307911.2  386628.7  394898.5  473955.6  461382.0  475137.2
```

```
predictions2 <- add_predictions(training_set2, model2)
predictions2
```

```
##        RM LSTAT PTRATIO    MEDV      pred
## 2   6.421  9.14    17.8  453600  510698.3
## 5   7.147  5.33    18.7  760200  647354.2
## 8   6.172 19.15    15.2  569100  405234.5
## 9   5.631 29.93    15.2  346500  323578.4
## 10  6.004 17.10    15.2  396900  410332.0
## 11  6.377 20.45    15.2  315000  403833.1
## 12  6.009 13.27    15.2  396900  440939.6
## 14  5.949  8.26    21.0  428400  464918.0
## 15  6.096 10.26    21.0  382200  441773.1
## 18  5.990 14.67    21.0  367500  371299.8
## 20  5.727 11.28    21.0  382200  412431.2
## 21  5.570 21.02    21.0  285600  286065.0
## 22  5.965 13.83    21.0  411600  383343.0
## 24  5.813 19.88    21.0  304500  296248.3
## 26  5.599 16.51    21.0  291900  342060.1
## 27  5.813 14.81    21.0  348600  366586.9
## 28  6.047 17.28    21.0  310800  332247.4
## 30  6.674 11.98    21.0  441000  433489.8
## 32  6.072 13.04    21.0  304500  397778.8
## 33  5.950 27.71    21.0  277200  178838.7
## 35  6.096 20.34    21.0  283500  284483.8
## 37  5.841 11.41    19.2  420000  420204.7
## 38  5.850  8.77    19.2  441000  451706.8
## 40  6.595  4.32    18.3  646800  600937.5
## 42  6.770  4.84    17.9  558600  621654.9
## 45  6.069  9.55    17.9  445200  466573.1
## 48  6.030 18.80    17.9  348600  357064.0
## 49  5.399 30.81    17.9  302400  259854.1
## 50  5.602 16.20    17.9  407400  364117.1
## 51  5.963 13.45    16.8  413700  421590.6
## 55  5.888 14.80    21.1  396900  367009.6
## 56  7.249  4.81    17.9  743400  690589.5
## 57  6.383  5.77    17.3  518700  557404.4
## 58  6.816  3.95    15.1  663600  684699.6
## 59  6.145  6.86    19.7  489300  502020.2
## 61  5.741 13.15    19.7  392700  392605.0
## 63  6.456  6.73    19.7  466200  532133.8
## 64  6.762  9.50    19.7  525000  508312.2
## 65  7.104  8.05    18.6  693000  589043.9
```

```
## 66  6.290  4.67    16.1  493500  562228.1
## 67  5.787 10.24    16.1  407400  431174.8
## 68  5.878  8.10    18.9  462000  461901.3
## 69  5.594 13.09    18.9  365400  387277.5
## 70  5.885  8.79    18.9  438900  454605.3
## 71  6.417  6.72    19.2  508200  533028.9
## 72  5.961  9.88    19.2  455700  447430.5
## 74  6.245  7.54    19.2  491400  503722.9
## 77  6.279 11.97    18.7  420000  447223.7
## 79  6.232 12.34    18.7  445200  438691.7
## 80  5.874  9.10    18.7  426300  450148.0
## 82  6.619  7.22    19.0  501900  547476.3
## 83  6.302  6.72    19.0  520800  522639.5
## 84  6.167  7.51    19.0  480900  497580.5
## 85  6.389  9.62    18.5  501900  492726.2
## 86  6.630  6.53    18.5  558600  566769.3
## 87  6.015 12.86    18.5  472500  418814.6
## 89  7.007  5.50    17.8  495600  644811.0
## 90  7.079  5.70    17.8  602700  650928.5
## 91  6.417  8.81    17.8  474600  514993.6
## 92  6.405  8.20    17.8  462000  522334.6
## 93  6.442  8.16    18.2  480900  523088.3
## 94  6.211  6.21    18.2  525000  523243.9
## 96  6.625  6.65    18.0  596400  570561.2
## 97  6.163 11.34    18.0  449400  453329.1
## 98  8.069  4.21    18.0  812700  820002.0
## 99  7.820  3.57    18.0  919800  800167.0
## 100 7.416  6.19    18.0  697200  680711.9
## 101 6.727  9.42    20.9  577500  486728.0
## 102 6.781  7.67    20.9  556500  523908.1
## 103 6.405 10.63    20.9  390600  449695.9
## 104 6.137 13.44    20.9  405300  394167.3
## 106 5.851 16.47    20.9  409500  344727.7
## 108 6.127 14.09    20.9  428400  383729.2
## 109 6.474 12.27    20.9  415800  423383.8
## 111 6.195 13.00    20.9  455700  402690.0
## 112 6.715 10.16    17.8  478800  526118.2
## 113 5.913 16.21    17.8  394800  381467.3
## 114 6.092 17.09    17.8  392700  380807.8
## 115 6.254 10.45    17.8  388500  475142.1
## 116 5.928 15.76    17.8  384300  386986.3
## 117 6.176 12.04    17.8  445200  447566.0
## 118 6.021 10.30    17.8  403200  453509.7
## 119 5.872 15.37    17.8  428400  387657.7
## 120 5.731 13.61    17.8  405300  394723.2
## 122 6.004 14.27    19.1  426300  395541.5
## 123 5.961 17.93    19.1  430500  348173.6
## 124 5.856 25.41    19.1  363300  258972.9
## 125 5.879 17.58    19.1  394800  350469.0
## 126 5.986 14.81    19.1  449400  387887.9
## 127 5.613 27.26    19.1  329700  248433.8
## 128 5.693 17.19    21.2  340200  331814.0
## 130 5.637 18.34    21.2  300300  316910.3
## 131 6.458 12.60    21.2  403200  412478.5
```

```
## 132 6.326 12.26    21.2  411600  415047.3
## 133 6.372 11.12    21.2  483000  436234.0
## 134 5.822 15.03    21.2  386400  362141.5
## 135 5.757 17.31    21.2  327600  329988.1
## 136 6.335 16.96    21.2  380100  334093.2
## 142 5.019 34.41    21.2  302400  172059.8
## 144 5.468 26.42    14.7  327600  349120.9
## 145 4.903 29.29    14.7  247800  358753.0
## 146 6.130 27.80    14.7  289800  338075.0
## 148 4.926 29.53    14.7  306600  358924.3
## 149 5.186 28.32    14.7  373800  350332.4
## 150 5.597 21.45    14.7  323400  365336.8
## 151 6.122 14.10    14.7  451500  452765.4
## 153 5.012 12.12    14.7  321300  313357.3
## 155 6.129 15.12    14.7  357000  445012.4
## 156 6.152 15.02    14.7  327600  448436.4
## 158 6.943  4.59    14.7  867300  708021.4
## 159 6.066  6.43    14.7  510300  505939.6
## 164 5.877 12.14    14.7  499800  434781.4
## 167 5.875 14.43    14.7  365400  420868.5
## 168 5.880 12.03    14.7  401100  435862.6
## 169 5.572 14.69    16.6  485100  378364.2
## 170 6.416  9.04    16.6  495600  524609.6
## 171 5.859  9.64    16.6  474600  444847.7
## 172 6.546  5.33    16.6  617400  594996.8
## 173 6.020 10.11    16.6  487200  461058.8
## 175 6.860  6.92    16.6  627900  620227.6
## 176 6.980  5.04    17.8  781200  649711.7
## 177 7.765  7.56    17.8  835800  698849.5
## 178 6.144  9.45    17.8  760200  476316.4
## 179 7.155  4.82    17.8  795900  679216.2
## 180 6.563  5.68    17.8  682500  580082.6
## 181 5.604 13.98    17.8  554400  382299.3
## 183 6.782  6.68    15.2  672000  634622.7
## 184 6.556  4.56    15.2  625800  620873.2
## 186 6.951  5.10    15.2  777000  692163.6
## 187 6.739  4.69    15.2  640500  656181.9
## 189 6.800  5.03    15.6  653100  657263.8
## 190 6.604  4.38    15.6  611100  628959.8
## 191 7.287  4.08    12.6  699300  844288.3
## 192 7.107  8.61    12.6  636300  720851.5
## 194 6.975  4.56    17.0  732900  672820.6
## 195 7.135  4.45    17.0  690900  701274.7
## 197 7.610  3.11    14.7  888300  884652.1
## 199 5.891 10.87    18.6  474600  431985.7
## 200 6.326 10.97    18.6  512400  466318.2
## 202 6.064 14.66    18.6  512400  398805.6
## 203 5.344 23.09    18.6  420000  300159.1
## 206 5.807 16.03    18.6  470400  370480.9
## 207 6.375  9.38    18.6  590100  493912.4
## 208 5.412 29.55    18.6  497700  250315.5
## 210 5.888 13.51    16.4  489300  416752.0
## 211 6.642  9.69    16.4  602700  548282.4
## 212 5.951 17.92    16.4  451500  383906.9
```

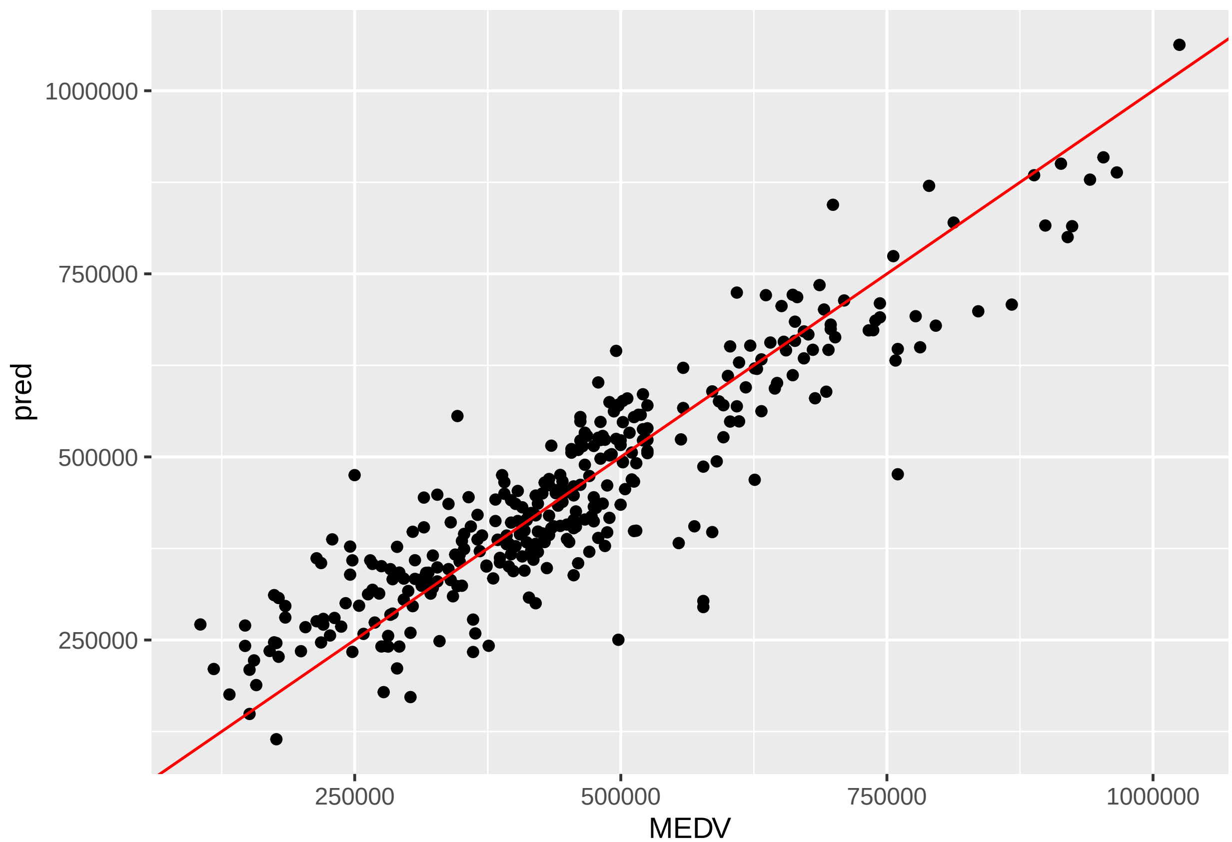
```
## 215 6.164 21.46    17.4  455700  338407.3
## 217 6.618  7.60    17.4  632100  562416.0
## 218 8.266  4.14    17.4  940800  878619.7
## 219 8.040  3.13    17.4  789600  870193.8
## 220 7.163  6.36    17.4  663600  658559.7
## 222 6.552  3.76    17.4  661500  611620.4
## 224 7.412  5.25    17.4  665700  718191.4
## 227 6.726  8.05    17.4  609000  569117.1
## 228 6.086 10.88    17.4  504000  455992.3
## 230 7.358  4.73    17.4  661500  721358.7
## 231 6.481  6.36    16.6  497700  570065.7
## 232 6.606  7.37    16.6  489300  574811.5
## 235 6.358 11.22    16.6  466200  489268.7
## 236 6.393  5.19    16.6  497700  571353.2
## 237 5.593 12.50    19.1  369600  392726.2
## 239 6.108  9.16    19.1  510300  469235.8
## 241 6.433  9.52    19.1  514500  491362.6
## 243 6.487  5.90    19.1  512400  554360.5
## 244 6.438  3.59    19.1  520800  585540.7
## 245 6.957  3.53    19.1  621600  651975.9
## 246 8.259  3.54    19.1  898800  815970.7
## 247 6.108  6.57    16.4  459900  509544.6
## 248 5.876  9.25    16.4  438900  450503.8
## 249 7.454  3.11    15.9  924000  815027.7
## 250 7.333  7.79    13.0  756000  774162.6
## 252 7.203  9.59    13.0  709800  713641.8
## 254 8.398  5.91    13.0 1024800 1062719.1
## 255 7.327 11.25    13.0  651000  706026.6
## 257 5.560 10.45    13.0  478800  389384.0
## 258 7.014 14.79    13.0  644700  593416.0
## 259 7.470  3.16    13.0  913500  900379.8
## 260 5.920 13.65    18.6  434700  402599.7
## 261 5.856 13.00    18.6  443100  405787.1
## 264 7.691  6.58    18.6  739200  686086.2
## 265 6.758  3.53    17.6  680400  646301.1
## 266 6.854  2.98    17.6  672000  671355.7
## 267 7.267  6.05    17.6  697200  674579.8
## 268 6.826  4.16    17.6  695100  646157.2
## 269 6.482  7.19    17.6  611100  548507.7
## 270 6.812  4.85    14.9  737100  673021.6
## 271 7.820  3.76    14.9  953400  909079.5
## 272 6.968  4.59    14.9  743400  709698.3
## 273 7.645  3.01    14.9  966000  888452.0
## 274 7.088  7.85    15.3  676200  667257.4
## 275 6.453  8.23    15.3  462000  554481.3
## 276 6.230 12.93    18.2  422100  436184.4
## 278 6.315  7.60    16.6  468300  528716.6
## 279 6.565  9.51    16.6  520800  537827.8
## 282 6.630  4.70    19.2  585900  589482.2
## 283 6.127  8.58    16.0  501900  493582.9
## 284 6.009 10.40    16.0  455700  459750.8
## 285 6.678  6.27    16.0  600600  610575.7
## 289 7.041  4.74    14.8  609000  724390.3
## 291 6.590  9.50    16.1  462000  548529.1
```

```
## 294 7.236  6.93    18.4  758100  631640.8
## 295 6.616  8.93    18.4  596400  526840.5
## 296 7.420  6.47    18.4  701400  663361.0
## 297 6.849  7.53    18.4  592200  575841.4
## 298 6.635  4.54    18.4  478800  601709.5
## 300 4.973 12.64    18.4  338100  346804.1
## 301 6.122  5.98    18.4  464100  514612.8
## 303 6.266  7.90    18.4  453600  505834.2
## 304 6.567  9.28    18.4  499800  516273.4
## 305 5.705 11.50    18.4  340200  410651.5
## 306 5.914 18.33    18.4  373800  351613.4
## 307 5.782 15.94    18.4  415800  371966.6
## 310 6.426  7.20    19.6  499800  522658.3
## 311 6.376  6.87    19.6  485100  523417.9
## 314 6.415  6.12    19.6  525000  539112.6
## 315 6.431  5.08    19.6  516600  557575.3
## 316 6.312  6.15    19.6  483000  528650.1
## 317 6.083 12.79    19.6  466200  414509.7
## 321 5.706 12.43    16.9  359100  404973.7
## 323 6.316  5.68    20.2  466200  533079.3
## 324 6.310  6.75    20.2  434700  515369.8
## 325 6.037  8.01    20.2  443100  475434.4
## 328 6.059  8.51    20.2  432600  469817.3
## 331 7.241  5.49    15.5  686700  734626.8
## 332 6.540  8.65    15.9  346500  555805.7
## 333 6.696  7.18    17.6  501900  576506.2
## 334 6.874  4.61    17.6  655200  645544.1
## 340 6.490  5.98    19.7  480900  547794.2
## 341 6.579  5.49    18.3  506100  579891.8
## 342 5.884  7.79    18.3  390600  465490.4
## 343 6.728  4.50    17.0  632100  633194.4
## 347 6.395 13.27    20.2  455700  413849.1
## 348 6.127 11.48    20.2  476700  430376.8
## 349 6.112 12.67    20.2  474600  412075.1
## 350 6.398  7.79    20.2  525000  505144.5
## 352 5.362 10.19    20.2  436800  405310.5
## 353 5.803 14.64    20.2  352800  374182.0
## 354 3.561  7.12    20.2  577500  294966.9
## 355 4.963 14.00    20.2  459900  354812.0
## 357 4.906 34.77    20.2  289800  211205.2
## 359 7.313 13.44    20.2  315000  444474.6
## 360 6.649 23.24    20.2  291900  241033.8
## 362 6.380 23.69    20.2  275100  241139.3
## 363 6.223 21.78    20.2  214200  275586.7
## 364 6.968 17.21    20.2  218400  355037.2
## 366 5.536 23.60    20.2  237300  268281.7
## 367 5.520 24.56    20.2  258300  258339.2
## 368 4.368 30.63    20.2  184800  280691.8
## 369 5.277 30.81    20.2  151200  209355.2
## 370 4.652 28.28    20.2  220500  271035.4
## 371 5.000 31.99    20.2  155400  222131.2
## 373 5.390 20.85    20.2  241500  300242.2
## 374 5.713 17.11    20.2  317100  341644.5
## 376 5.036 25.68    20.2  203700  267507.2
```

```
## 377 6.193 15.17    20.2  289800  377210.4
## 378 5.887 16.35    20.2  266700  353875.0
## 379 6.471 17.12    20.2  275100  350793.7
## 380 6.405 19.37    20.2  262500  312444.8
## 381 5.747 19.92    20.2  178500  307340.3
## 383 5.852 29.97    20.2  132300  175564.2
## 384 5.987 26.77    20.2  117600  210396.9
## 386 6.404 20.31    20.2  254100  296769.4
## 387 5.349 19.77    20.2  174300  311264.3
## 388 5.531 27.38    20.2  178500  227096.6
## 389 5.683 22.98    20.2  105000  271213.3
## 391 5.608 12.13    20.2  585900  397335.9
## 392 5.617 26.40    20.2  361200  233615.3
## 393 6.852 19.78    20.2  577500  303249.0
## 395 6.657 21.22    20.2  361200  277862.3
## 396 4.628 34.37    20.2  375900  242081.4
## 397 5.155 20.08    20.2  342300  309650.3
## 398 4.519 36.98    20.2  147000  241950.6
## 399 6.434 29.05    20.2  151200  149010.5
## 400 6.782 25.79    20.2  157500  188434.4
## 401 5.304 26.64    20.2  218400  246628.9
## 402 5.957 20.62    20.2  184800  296423.9
## 403 6.824 22.74    20.2  176400  245849.4
## 404 6.411 15.02    20.2  350700  385181.8
## 407 6.103 23.29    20.2  281400  255602.8
## 408 5.565 17.16    20.2  245700  339266.7
## 409 5.896 24.39    20.2  174300  246802.5
## 410 5.837 15.69    20.2  214200  361556.4
## 411 6.202 14.52    20.2  228900  387413.3
## 412 6.193 21.52    20.2  231000  280115.0
## 413 6.380 24.08    20.2  199500  234696.2
## 415 6.833 19.69    20.2  296100  305104.1
## 416 6.425 12.03    20.2  338100  435828.6
## 418 6.208 15.17    20.2  245700  377586.3
## 419 6.629 23.27    20.2  281400  241049.6
## 422 5.935 34.02    20.2  176400  114509.8
## 423 5.627 22.88    20.2  268800  273843.9
## 424 5.818 22.11    20.2  220500  278816.9
## 426 6.219 16.59    20.2  386400  355906.4
## 427 6.485 18.85    20.2  323400  321189.1
## 428 5.854 23.79    20.2  226800  256133.7
## 429 6.459 23.98    20.2  247800  233749.4
## 431 6.251 16.44    20.2  264600  358760.1
## 433 6.417 19.31    20.2  273000  313419.8
## 435 6.655 17.73    20.2  319200  341862.9
## 438 6.728 18.71    20.2  312900  324029.0
## 439 6.525 18.13    20.2  296100  333804.8
## 440 5.976 19.01    20.2  266700  318538.4
## 441 5.936 16.94    20.2  283500  346698.5
## 443 6.081 14.70    20.2  420000  381244.2
## 444 6.701 16.42    20.2  344400  366649.4
## 446 6.317 13.99    20.2  409500  399323.1
## 448 6.209 13.22    20.2  449400  407633.3
## 450 5.952 17.15    20.2  399000  344035.0
```

```
## 453 5.713 14.76    20.2  422100  370157.0
## 454 6.167 16.29    20.2  417900  359625.0
## 455 6.229 12.87    20.2  411600  413825.6
## 456 6.437 14.36    20.2  487200  397010.1
## 457 6.980 11.66    20.2  625800  468773.1
## 459 6.162 24.10    20.2  279300  241709.5
## 460 6.484 18.68    20.2  350700  324113.0
## 462 6.185 18.03    20.2  306600  333429.3
## 466 7.061  7.01    20.2  525000  570326.2
## 467 5.762 10.42    20.2  457800  425573.2
## 468 5.871 13.34    20.2  432600  393247.1
## 469 6.312 10.58    20.2  445200  453969.4
## 470 6.114 14.98    20.2  401100  378037.3
## 471 5.905 11.45    20.2  432600  419826.9
## 472 5.454 18.06    20.1  319200  329205.9
## 473 5.414 23.97    20.1  147000  269763.3
## 474 5.093 29.68    20.1  170100  234985.7
## 475 5.983 18.07    20.1  285600  333002.4
## 476 5.983 13.35    20.1  422100  398061.3
## 477 5.707 12.01    19.2  457800  404746.0
## 478 5.926 13.59    19.2  514500  399249.2
## 480 5.390 21.14    19.2  413700  307911.2
## 481 5.794 14.10    19.2  384300  386628.7
## 484 6.027 14.33    19.2  352800  394898.5
## 485 6.593  9.67    21.0  470400  473955.6
## 486 6.120  9.08    21.0  432600  461382.0
## 489 6.030  7.88    21.0  249900  475137.2
```

```r
ggplot(data = predictions2, mapping = aes(x = MEDV, y = pred)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red")
```




```
R2(predictions2$pred, predictions2$MEDV)
```

```
## [1] 0.8224075
```

```
MAE(predictions2$pred, predictions2$MEDV)
```

```
## [1] 53873.34
```

```
RMSE(predictions2$pred, predictions2$MEDV)
```

```
## [1] 71969.11
```

```
resids2 <- add_residuals(training_set2, model2)
resids2
```

```
##       RM LSTAT PTRATIO   MEDV        resid
## 2  6.421  9.14    17.8 453600  -57098.28129
## 5  7.147  5.33    18.7 760200  112845.78114
## 8  6.172 19.15    15.2 569100  163865.46928
## 9  5.631 29.93    15.2 346500   22921.60914
## 10 6.004 17.10    15.2 396900  -13431.95818
## 11 6.377 20.45    15.2 315000  -88833.14750
## 12 6.009 13.27    15.2 396900  -44039.57809
```

```
## 14  5.949  8.26    21.0  428400  -36518.00451
## 15  6.096 10.26    21.0  382200  -59573.07738
## 18  5.990 14.67    21.0  367500   -3799.77190
## 20  5.727 11.28    21.0  382200  -30231.19678
## 21  5.570 21.02    21.0  285600    -465.00682
## 22  5.965 13.83    21.0  411600   28257.04376
## 24  5.813 19.88    21.0  304500    8251.65571
## 26  5.599 16.51    21.0  291900  -50160.10299
## 27  5.813 14.81    21.0  348600  -17986.85855
## 28  6.047 17.28    21.0  310800  -21447.40086
## 30  6.674 11.98    21.0  441000    7510.24040
## 32  6.072 13.04    21.0  304500  -93278.78510
## 33  5.950 27.71    21.0  277200   98361.26727
## 35  6.096 20.34    21.0  283500    -983.79748
## 37  5.841 11.41    19.2  420000    -204.65362
## 38  5.850  8.77    19.2  441000  -10706.80711
## 40  6.595  4.32    18.3  646800   45862.47038
## 42  6.770  4.84    17.9  558600  -63054.85885
## 45  6.069  9.55    17.9  445200  -21373.08114
## 48  6.030 18.80    17.9  348600   -8463.99891
## 49  5.399 30.81    17.9  302400   42545.88338
## 50  5.602 16.20    17.9  407400   43282.91986
## 51  5.963 13.45    16.8  413700   -7890.61763
## 55  5.888 14.80    21.1  396900   29890.39364
## 56  7.249  4.81    17.9  743400   52810.49276
## 57  6.383  5.77    17.3  518700  -38704.43949
## 58  6.816  3.95    15.1  663600  -21099.58534
## 59  6.145  6.86    19.7  489300  -12720.16982
## 61  5.741 13.15    19.7  392700      94.95859
## 63  6.456  6.73    19.7  466200  -65933.82951
## 64  6.762  9.50    19.7  525000   16687.75310
## 65  7.104  8.05    18.6  693000  103956.06721
## 66  6.290  4.67    16.1  493500  -68728.06381
## 67  5.787 10.24    16.1  407400  -23774.81477
## 68  5.878  8.10    18.9  462000      98.74698
## 69  5.594 13.09    18.9  365400  -21877.53804
## 70  5.885  8.79    18.9  438900  -15705.25866
## 71  6.417  6.72    19.2  508200  -24828.90568
## 72  5.961  9.88    19.2  455700    8269.51133
## 74  6.245  7.54    19.2  491400  -12322.94008
## 77  6.279 11.97    18.7  420000  -27223.70552
## 79  6.232 12.34    18.7  445200    6508.26154
## 80  5.874  9.10    18.7  426300  -23847.95239
## 82  6.619  7.22    19.0  501900  -45576.30407
## 83  6.302  6.72    19.0  520800   -1839.50716
## 84  6.167  7.51    19.0  480900  -16680.47562
## 85  6.389  9.62    18.5  501900    9173.79602
## 86  6.630  6.53    18.5  558600   -8169.26717
## 87  6.015 12.86    18.5  472500   53685.39527
## 89  7.007  5.50    17.8  495600 -149211.01432
## 90  7.079  5.70    17.8  602700  -48228.52671
## 91  6.417  8.81    17.8  474600  -40393.60132
## 92  6.405  8.20    17.8  462000  -60334.63212
## 93  6.442  8.16    18.2  480900  -42188.34601
```

```
## 94  6.211  6.21    18.2  525000    1756.11199
## 96  6.625  6.65    18.0  596400   25838.81196
## 97  6.163 11.34    18.0  449400   -3929.06475
## 98  8.069  4.21    18.0  812700   -7302.03538
## 99  7.820  3.57    18.0  919800  119633.02427
## 100 7.416  6.19    18.0  697200   16488.07632
## 101 6.727  9.42    20.9  577500   90772.00392
## 102 6.781  7.67    20.9  556500   32591.87279
## 103 6.405 10.63    20.9  390600  -59095.87797
## 104 6.137 13.44    20.9  405300   11132.72233
## 106 5.851 16.47    20.9  409500   64772.31629
## 108 6.127 14.09    20.9  428400   44670.78755
## 109 6.474 12.27    20.9  415800   -7583.76283
## 111 6.195 13.00    20.9  455700   53010.04042
## 112 6.715 10.16    17.8  478800  -47318.23318
## 113 5.913 16.21    17.8  394800   13332.71592
## 114 6.092 17.09    17.8  392700   11892.18782
## 115 6.254 10.45    17.8  388500  -86642.06164
## 116 5.928 15.76    17.8  384300   -2686.31800
## 117 6.176 12.04    17.8  445200   -2365.95303
## 118 6.021 10.30    17.8  403200  -50309.69043
## 119 5.872 15.37    17.8  428400   40742.31549
## 120 5.731 13.61    17.8  405300   10576.82876
## 122 6.004 14.27    19.1  426300   30758.48863
## 123 5.961 17.93    19.1  430500   82326.44827
## 124 5.856 25.41    19.1  363300  104327.09399
## 125 5.879 17.58    19.1  394800   44330.99420
## 126 5.986 14.81    19.1  449400   61512.10023
## 127 5.613 27.26    19.1  329700   81266.18001
## 128 5.693 17.19    21.2  340200    8386.03784
## 130 5.637 18.34    21.2  300300  -16610.28889
## 131 6.458 12.60    21.2  403200   -9278.51129
## 132 6.326 12.26    21.2  411600   -3447.32139
## 133 6.372 11.12    21.2  483000   46765.95788
## 134 5.822 15.03    21.2  386400   24258.50795
## 135 5.757 17.31    21.2  327600   -2388.08780
## 136 6.335 16.96    21.2  380100   46006.81522
## 142 5.019 34.41    21.2  302400  130340.20040
## 144 5.468 26.42    14.7  327600  -21520.86258
## 145 4.903 29.29    14.7  247800 -110952.95797
## 146 6.130 27.80    14.7  289800  -48274.99767
## 148 4.926 29.53    14.7  306600  -52324.28878
## 149 5.186 28.32    14.7  373800   23467.57642
## 150 5.597 21.45    14.7  323400  -41936.82359
## 151 6.122 14.10    14.7  451500   -1265.36935
## 153 5.012 12.12    14.7  321300    7942.69030
## 155 6.129 15.12    14.7  357000  -88012.44166
## 156 6.152 15.02    14.7  327600 -120836.35322
## 158 6.943  4.59    14.7  867300  159278.55817
## 159 6.066  6.43    14.7  510300    4360.36230
## 164 5.877 12.14    14.7  499800   65018.55238
## 167 5.875 14.43    14.7  365400  -55468.53307
## 168 5.880 12.03    14.7  401100  -34762.59521
## 169 5.572 14.69    16.6  485100  106735.75782
```

```
## 170 6.416  9.04    16.6  495600  -29009.64029
## 171 5.859  9.64    16.6  474600   29752.33280
## 172 6.546  5.33    16.6  617400   22403.20397
## 173 6.020 10.11    16.6  487200   26141.19709
## 175 6.860  6.92    16.6  627900    7672.44770
## 176 6.980  5.04    17.8  781200  131488.34675
## 177 7.765  7.56    17.8  835800  136950.53037
## 178 6.144  9.45    17.8  760200  283883.56594
## 179 7.155  4.82    17.8  795900  116683.80568
## 180 6.563  5.68    17.8  682500  102417.35639
## 181 5.604 13.98    17.8  554400  172100.73850
## 183 6.782  6.68    15.2  672000   37377.34655
## 184 6.556  4.56    15.2  625800    4926.84746
## 186 6.951  5.10    15.2  777000   84836.44389
## 187 6.739  4.69    15.2  640500  -15681.93618
## 189 6.800  5.03    15.6  653100   -4163.77743
## 190 6.604  4.38    15.6  611100  -17859.79371
## 191 7.287  4.08    12.6  699300 -144988.29638
## 192 7.107  8.61    12.6  636300  -84551.48592
## 194 6.975  4.56    17.0  732900   60079.38297
## 195 7.135  4.45    17.0  690900  -10374.68499
## 197 7.610  3.11    14.7  888300    3647.92212
## 199 5.891 10.87    18.6  474600   42614.26472
## 200 6.326 10.97    18.6  512400   46081.78613
## 202 6.064 14.66    18.6  512400  113594.41967
## 203 5.344 23.09    18.6  420000  119840.87505
## 206 5.807 16.03    18.6  470400   99919.06552
## 207 6.375  9.38    18.6  590100   96187.58028
## 208 5.412 29.55    18.6  497700  247384.49569
## 210 5.888 13.51    16.4  489300   72547.98903
## 211 6.642  9.69    16.4  602700   54417.63714
## 212 5.951 17.92    16.4  451500   67593.12114
## 215 6.164 21.46    17.4  455700  117292.71488
## 217 6.618  7.60    17.4  632100   69683.99653
## 218 8.266  4.14    17.4  940800   62180.26311
## 219 8.040  3.13    17.4  789600  -80593.75451
## 220 7.163  6.36    17.4  663600    5040.33118
## 222 6.552  3.76    17.4  661500   49879.55594
## 224 7.412  5.25    17.4  665700  -52491.35183
## 227 6.726  8.05    17.4  609000   39882.88903
## 228 6.086 10.88    17.4  504000   48007.65153
## 230 7.358  4.73    17.4  661500  -59858.72646
## 231 6.481  6.36    16.6  497700  -72365.72002
## 232 6.606  7.37    16.6  489300  -85511.48133
## 235 6.358 11.22    16.6  466200  -23068.66320
## 236 6.393  5.19    16.6  497700  -73653.15597
## 237 5.593 12.50    19.1  369600  -23126.17484
## 239 6.108  9.16    19.1  510300   41064.20488
## 241 6.433  9.52    19.1  514500   23137.35131
## 243 6.487  5.90    19.1  512400  -41960.54756
## 244 6.438  3.59    19.1  520800  -64740.66502
## 245 6.957  3.53    19.1  621600  -30375.94035
## 246 8.259  3.54    19.1  898800   82829.32396
## 247 6.108  6.57    16.4  459900  -49644.61504
```

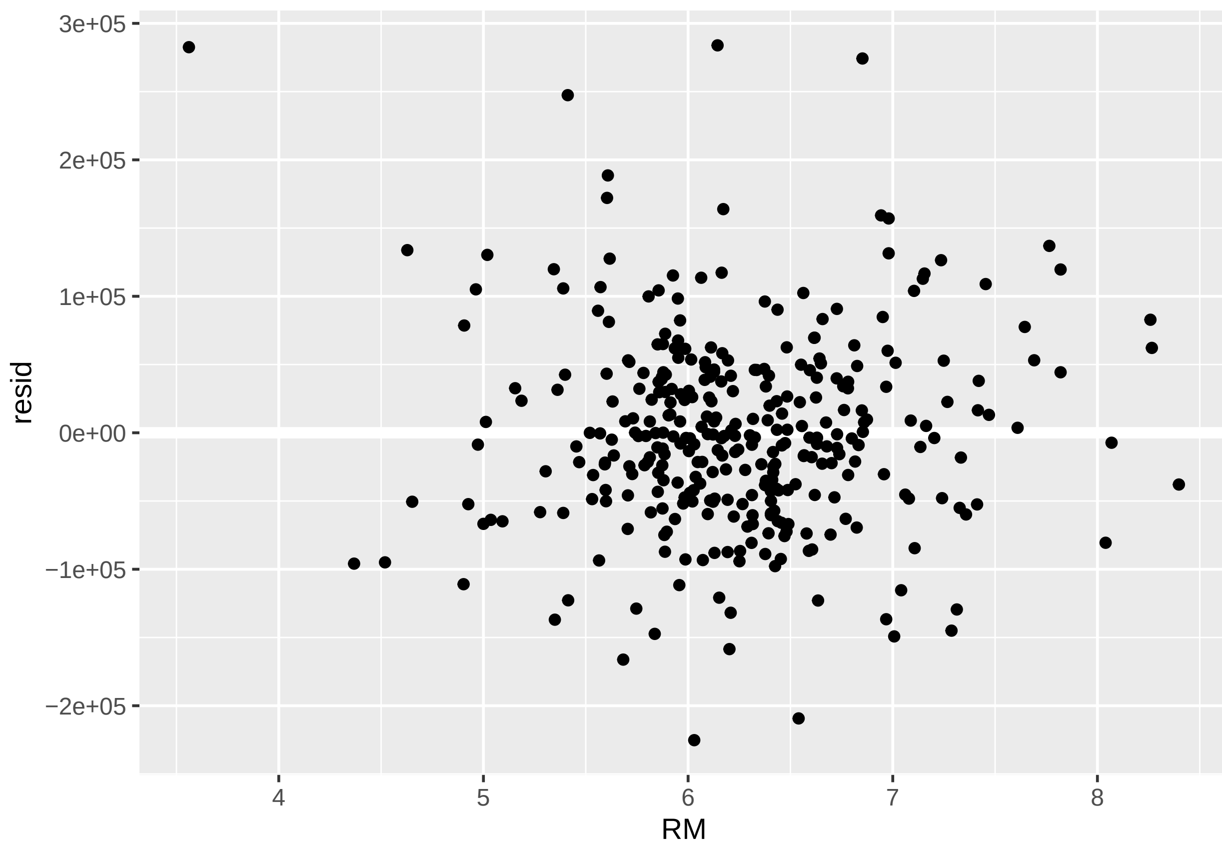
```
## 248 5.876  9.25    16.4  438900  -11603.75212
## 249 7.454  3.11    15.9  924000  108972.30379
## 250 7.333  7.79    13.0  756000  -18162.56704
## 252 7.203  9.59    13.0  709800   -3841.80105
## 254 8.398  5.91    13.0 1024800  -37919.07409
## 255 7.327 11.25    13.0  651000  -55026.55416
## 257 5.560 10.45    13.0  478800   89416.04765
## 258 7.014 14.79    13.0  644700   51283.97283
## 259 7.470  3.16    13.0  913500   13120.24185
## 260 5.920 13.65    18.6  434700   32100.30341
## 261 5.856 13.00    18.6  443100   37312.90737
## 264 7.691  6.58    18.6  739200   53113.80380
## 265 6.758  3.53    17.6  680400   34098.92477
## 266 6.854  2.98    17.6  672000     644.31983
## 267 7.267  6.05    17.6  697200   22620.16752
## 268 6.826  4.16    17.6  695100   48942.83992
## 269 6.482  7.19    17.6  611100   62592.27676
## 270 6.812  4.85    14.9  737100   64078.41366
## 271 7.820  3.76    14.9  953400   44320.53081
## 272 6.968  4.59    14.9  743400   33701.72587
## 273 7.645  3.01    14.9  966000   77548.00334
## 274 7.088  7.85    15.3  676200    8942.62045
## 275 6.453  8.23    15.3  462000  -92481.34623
## 276 6.230 12.93    18.2  422100  -14084.35351
## 278 6.315  7.60    16.6  468300  -60416.57588
## 279 6.565  9.51    16.6  520800  -17027.82340
## 282 6.630  4.70    19.2  585900   -3582.15259
## 283 6.127  8.58    16.0  501900    8317.09585
## 284 6.009 10.40    16.0  455700   -4050.75554
## 285 6.678  6.27    16.0  600600   -9975.66665
## 289 7.041  4.74    14.8  609000 -115390.34055
## 291 6.590  9.50    16.1  462000  -86529.13803
## 294 7.236  6.93    18.4  758100  126459.22934
## 295 6.616  8.93    18.4  596400   69559.49110
## 296 7.420  6.47    18.4  701400   38038.98630
## 297 6.849  7.53    18.4  592200   16358.64230
## 298 6.635  4.54    18.4  478800 -122909.54041
## 300 4.973 12.64    18.4  338100   -8704.14772
## 301 6.122  5.98    18.4  464100  -50512.76528
## 303 6.266  7.90    18.4  453600  -52234.18226
## 304 6.567  9.28    18.4  499800  -16473.43720
## 305 5.705 11.50    18.4  340200  -70451.47669
## 306 5.914 18.33    18.4  373800   22186.55184
## 307 5.782 15.94    18.4  415800   43833.37449
## 310 6.426  7.20    19.6  499800  -22858.29726
## 311 6.376  6.87    19.6  485100  -38317.90194
## 314 6.415  6.12    19.6  525000  -14112.64585
## 315 6.431  5.08    19.6  516600  -40975.26700
## 316 6.312  6.15    19.6  483000  -45650.08990
## 317 6.083 12.79    19.6  466200   51690.27181
## 321 5.706 12.43    16.9  359100  -45873.66800
## 323 6.316  5.68    20.2  466200  -66879.25962
## 324 6.310  6.75    20.2  434700  -80669.83030
## 325 6.037  8.01    20.2  443100  -32334.37758
```

```
## 328 6.059  8.51    20.2  432600  -37217.32416
## 331 7.241  5.49    15.5  686700  -47926.78764
## 332 6.540  8.65    15.9  346500 -209305.72223
## 333 6.696  7.18    17.6  501900  -74606.24629
## 334 6.874  4.61    17.6  655200    9655.91320
## 340 6.490  5.98    19.7  480900  -66894.19590
## 341 6.579  5.49    18.3  506100  -73791.78434
## 342 5.884  7.79    18.3  390600  -74890.36838
## 343 6.728  4.50    17.0  632100   -1094.42477
## 347 6.395 13.27    20.2  455700   41850.89492
## 348 6.127 11.48    20.2  476700   46323.15110
## 349 6.112 12.67    20.2  474600   62524.88267
## 350 6.398  7.79    20.2  525000   19855.49281
## 352 5.362 10.19    20.2  436800   31489.45291
## 353 5.803 14.64    20.2  352800  -21381.99938
## 354 3.561  7.12    20.2  577500  282533.11661
## 355 4.963 14.00    20.2  459900  105088.04442
## 357 4.906 34.77    20.2  289800   78594.80426
## 359 7.313 13.44    20.2  315000 -129474.56467
## 360 6.649 23.24    20.2  291900   50866.24878
## 362 6.380 23.69    20.2  275100   33960.71361
## 363 6.223 21.78    20.2  214200  -61386.67085
## 364 6.968 17.21    20.2  218400 -136637.20855
## 366 5.536 23.60    20.2  237300  -30981.66310
## 367 5.520 24.56    20.2  258300     -39.18482
## 368 4.368 30.63    20.2  184800  -95891.79394
## 369 5.277 30.81    20.2  151200  -58155.20330
## 370 4.652 28.28    20.2  220500  -50535.41012
## 371 5.000 31.99    20.2  155400  -66731.15282
## 373 5.390 20.85    20.2  241500  -58742.21019
## 374 5.713 17.11    20.2  317100  -24544.46819
## 376 5.036 25.68    20.2  203700  -63807.18526
## 377 6.193 15.17    20.2  289800  -87410.40324
## 378 5.887 16.35    20.2  266700  -87174.99595
## 379 6.471 17.12    20.2  275100  -75693.73033
## 380 6.405 19.37    20.2  262500  -49944.75395
## 381 5.747 19.92    20.2  178500 -128840.31011
## 383 5.852 29.97    20.2  132300  -43264.15223
## 384 5.987 26.77    20.2  117600  -92796.93069
## 386 6.404 20.31    20.2  254100  -42669.38953
## 387 5.349 19.77    20.2  174300 -136964.34666
## 388 5.531 27.38    20.2  178500  -48596.56998
## 389 5.683 22.98    20.2  105000 -166213.32954
## 391 5.608 12.13    20.2  585900  188564.11786
## 392 5.617 26.40    20.2  361200  127584.66789
## 393 6.852 19.78    20.2  577500  274251.04923
## 395 6.657 21.22    20.2  361200   83337.65774
## 396 4.628 34.37    20.2  375900  133818.55093
## 397 5.155 20.08    20.2  342300   32649.69961
## 398 4.519 36.98    20.2  147000  -94950.64388
## 399 6.434 29.05    20.2  151200    2189.54210
## 400 6.782 25.79    20.2  157500  -30934.39088
## 401 5.304 26.64    20.2  218400  -28228.85464
## 402 5.957 20.62    20.2  184800 -111623.91144
```

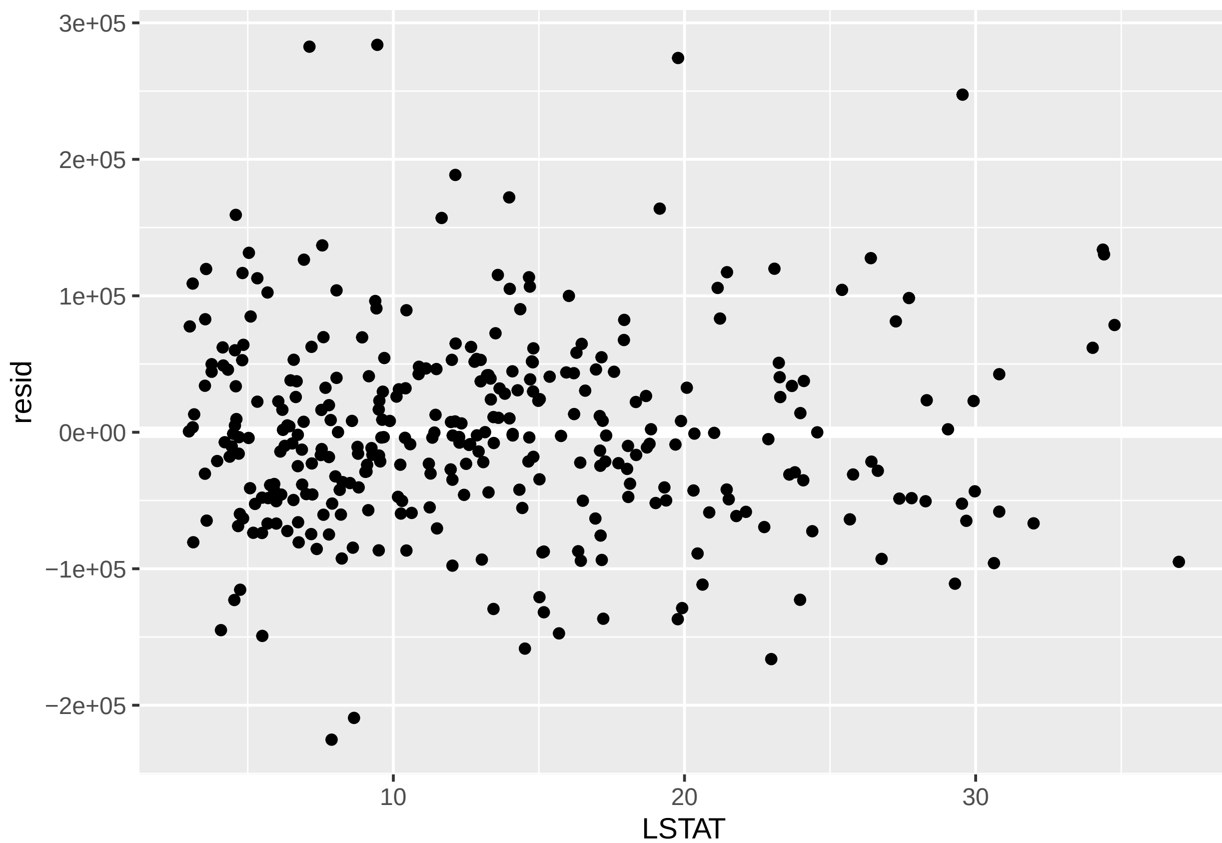
```
## 403 6.824 22.74    20.2  176400  -69449.40819
## 404 6.411 15.02    20.2  350700  -34481.78734
## 407 6.103 23.29    20.2  281400   25797.17583
## 408 5.565 17.16    20.2  245700  -93566.67296
## 409 5.896 24.39    20.2  174300  -72502.45099
## 410 5.837 15.69    20.2  214200 -147356.36559
## 411 6.202 14.52    20.2  228900 -158513.31194
## 412 6.193 21.52    20.2  231000  -49114.97804
## 413 6.380 24.08    20.2  199500  -35196.18462
## 415 6.833 19.69    20.2  296100   -9004.11527
## 416 6.425 12.03    20.2  338100  -97728.64972
## 418 6.208 15.17    20.2  245700 -131886.27562
## 419 6.629 23.27    20.2  281400   40350.42631
## 422 5.935 34.02    20.2  176400   61890.15539
## 423 5.627 22.88    20.2  268800   -5043.91552
## 424 5.818 22.11    20.2  220500  -58316.91787
## 426 6.219 16.59    20.2  386400   30493.63590
## 427 6.485 18.85    20.2  323400    2210.94497
## 428 5.854 23.79    20.2  226800  -29333.74719
## 429 6.459 23.98    20.2  247800   14050.62484
## 431 6.251 16.44    20.2  264600  -94160.12850
## 433 6.417 19.31    20.2  273000  -40419.75230
## 435 6.655 17.73    20.2  319200  -22662.91757
## 438 6.728 18.71    20.2  312900  -11128.95053
## 439 6.525 18.13    20.2  296100  -37704.78416
## 440 5.976 19.01    20.2  266700  -51838.43525
## 441 5.936 16.94    20.2  283500  -63198.49814
## 443 6.081 14.70    20.2  420000   38755.80553
## 444 6.701 16.42    20.2  344400  -22249.40875
## 446 6.317 13.99    20.2  409500   10176.90823
## 448 6.209 13.22    20.2  449400   41766.71563
## 450 5.952 17.15    20.2  399000   54964.96921
## 453 5.713 14.76    20.2  422100   51942.98319
## 454 6.167 16.29    20.2  417900   58275.03900
## 455 6.229 12.87    20.2  411600   -2225.60791
## 456 6.437 14.36    20.2  487200   90189.85583
## 457 6.980 11.66    20.2  625800  157026.85773
## 459 6.162 24.10    20.2  279300   37590.52886
## 460 6.484 18.68    20.2  350700   26586.95756
## 462 6.185 18.03    20.2  306600  -26829.28324
## 466 7.061  7.01    20.2  525000  -45326.15103
## 467 5.762 10.42    20.2  457800   32226.79981
## 468 5.871 13.34    20.2  432600   39352.86437
## 469 6.312 10.58    20.2  445200   -8769.43416
## 470 6.114 14.98    20.2  401100   23062.71522
## 471 5.905 11.45    20.2  432600   12773.05200
## 472 5.454 18.06    20.1  319200  -10005.92601
## 473 5.414 23.97    20.1  147000 -122763.29172
## 474 5.093 29.68    20.1  170100  -64885.72054
## 475 5.983 18.07    20.1  285600  -47402.35913
## 476 5.983 13.35    20.1  422100   24038.71278
## 477 5.707 12.01    19.2  457800   53054.04644
## 478 5.926 13.59    19.2  514500  115250.75496
## 480 5.390 21.14    19.2  413700  105788.82566
```

```
## 481 5.794 14.10    19.2  384300   -2328.74213
## 484 6.027 14.33    19.2  352800  -42098.46639
## 485 6.593  9.67    21.0  470400   -3555.60610
## 486 6.120  9.08    21.0  432600  -28782.01241
## 489 6.030  7.88    21.0  249900 -225237.21241
```

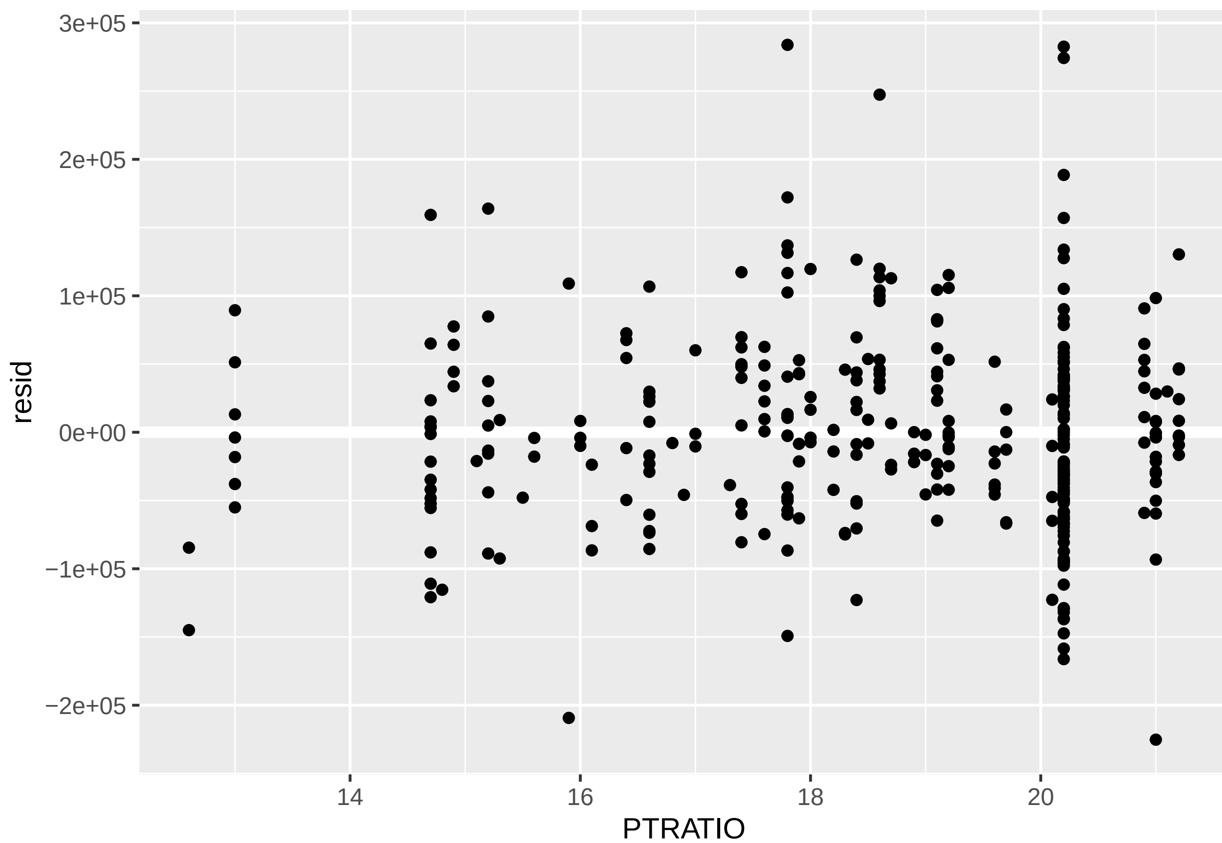
```
ggplot(data=resids2, mapping=aes(x=RM, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```




```
ggplot(data=resids2, mapping=aes(x=LSTAT, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```

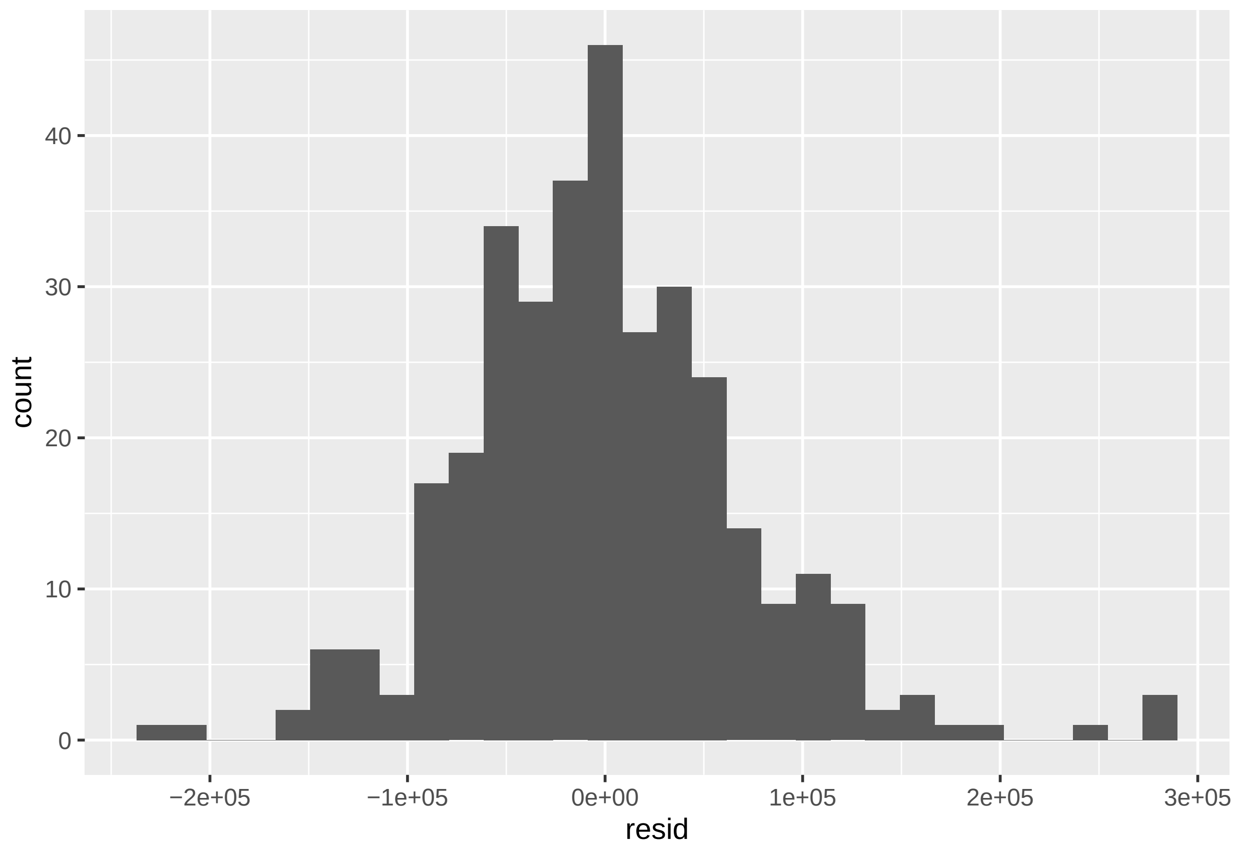



```
ggplot(data=resids2, mapping=aes(x=PTRATIO, y=resid))+
  geom_ref_line(h=0)+
  geom_point()
```




```
ggplot(resids2, aes(x=resid))+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



5) (15 pts) How well does the model do on the test data? Given the validation results, is the performance at all surprising?

```
train.control <- trainControl(method = "cv", number = 5)
model2 <- train( MEDV~RM*LSTAT*PTRATIO,
               data = test_set,
               method = "lm",
               trControl = train.control)
model2
```

```
## Linear Regression
##
## 68 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 54, 56, 54, 54, 54
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   63933.23  0.6148481  48638.69
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -141043  -31777   -5525   34492  122293
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          186630    1705822   0.109    0.913
## RM                    34304     276261   0.124    0.902
## LSTAT               -155446     153562  -1.012    0.315
## PTRATIO               -8024      88798  -0.090    0.928
## `RM:LSTAT`            29091      27373   1.063    0.292
## `RM:PTRATIO`           2760      14412   0.192    0.849
## `LSTAT:PTRATIO`        8818       7822   1.127    0.264
## `RM:LSTAT:PTRATIO`    -1734       1390  -1.247    0.217
##
## Residual standard error: 53420 on 60 degrees of freedom
## Multiple R-squared:  0.7419, Adjusted R-squared:  0.7117
## F-statistic: 24.63 on 7 and 60 DF,  p-value: 1.889e-15
```

The model does not perform as well on the test data, however I am not really surpised about this, because the more data that there is in a sample, the more normal the data becomes, and the easier the data is to be explained by a linear regression.