Assignment 5 Predictions

Nicholas Jacob and Yechang Qi

2024-09-05

Predicting House Prices

1a.

We start by reading in the data and doing some cleaning. With the large presence of `Nas` in the data, we convert `Na` into a factor for factor data and...

```
housingData = read.csv("housingData-1.csv")

housingFactor <- housingData %>%
  dplyr::select(where(is.character))%>%
  mutate_all(factor) %>%
  mutate_all(fct_na_value_to_level)
#Needed to convert NA's to a level to make the regression have enough data to go.

housingNumeric <- housingData %>%
  dplyr::select(where(is.numeric)) %>%
  mutate_if(is.numeric, ~replace(., is.na(.), 0))

housingData <- cbind(housingFactor, housingNumeric)


fit <- lm(log(SalePrice) ~ ., data= housingData )

summary(fit)
```

```
##
## Call:
## lm(formula = log(SalePrice) ~ ., data = housingData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57984 -0.04825  0.00301  0.05011  0.25581
##
## Coefficients: (8 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.444e+00  4.811e+00   1.547 0.122176
## MSZoningRH          -2.793e-02  5.086e-02  -0.549 0.583056
## MSZoningRL          -6.417e-03  3.904e-02  -0.164 0.869492
## MSZoningRM          -8.855e-02  4.067e-02  -2.177 0.029727 *
## AlleyPave            2.248e-02  2.937e-02   0.766 0.444120
## AlleyNA              1.567e-02  1.889e-02   0.830 0.406886
## LotShapeIR2          3.708e-03  1.900e-02   0.195 0.845340
```

```
## LotShapeIR3         -5.563e-02  4.022e-02  -1.383 0.167068
## LotShapeReg          3.421e-03  7.631e-03   0.448 0.654058
## LandContourHLS       3.236e-02  2.504e-02   1.292 0.196722
## LandContourLow       7.768e-03  2.958e-02   0.263 0.792925
## LandContourLvl       9.226e-03  1.736e-02   0.532 0.595194
## LotConfigCulDSac     2.761e-02  1.409e-02   1.959 0.050400 .
## LotConfigInside     -9.880e-03  8.159e-03  -1.211 0.226293
## LotConfigother      -2.467e-02  1.728e-02  -1.428 0.153785
## LandSlopeMod        -1.137e-03  1.852e-02  -0.061 0.951058
## LandSlopeSev        -9.360e-02  5.466e-02  -1.712 0.087210 .
## NeighborhoodClearCr -2.950e-02  3.071e-02  -0.961 0.337012
## NeighborhoodCollgCr -5.866e-02  2.450e-02  -2.395 0.016863 *
## NeighborhoodCrawfor  9.061e-02  2.587e-02   3.503 0.000485 ***
## NeighborhoodEdwards -1.019e-01  2.197e-02  -4.638 4.08e-06 ***
## NeighborhoodGilbert -8.003e-02  2.724e-02  -2.938 0.003393 **
## NeighborhoodIDOTRR   1.263e-02  2.756e-02   0.458 0.646888
## NeighborhoodMitchel -9.302e-02  2.609e-02  -3.565 0.000385 ***
## NeighborhoodNAmes   -6.859e-02  2.160e-02  -3.175 0.001554 **
## NeighborhoodNoRidge -7.284e-02  3.017e-02  -2.414 0.015976 *
## NeighborhoodNridgHt  9.689e-03  3.006e-02   0.322 0.747288
## NeighborhoodNWAmes  -7.587e-02  2.559e-02  -2.965 0.003113 **
## NeighborhoodOldTown -5.707e-02  2.253e-02  -2.533 0.011499 *
## Neighborhoodother   -4.613e-02  2.388e-02  -1.932 0.053725 .
## NeighborhoodSawyer  -7.749e-02  2.385e-02  -3.249 0.001207 **
## NeighborhoodSawyerW -6.989e-02  2.647e-02  -2.640 0.008438 **
## NeighborhoodSomerst -4.647e-03  4.076e-02  -0.114 0.909274
## NeighborhoodTimber  -5.113e-02  3.192e-02  -1.602 0.109584
## Condition1Feedr      1.595e-02  2.258e-02   0.707 0.480059
## Condition1Norm       5.742e-02  1.902e-02   3.019 0.002611 **
## Condition1PosA      -2.805e-02  4.177e-02  -0.672 0.501990
## Condition1PosN       7.434e-04  3.190e-02   0.023 0.981415
## Condition1RR         8.207e-03  2.728e-02   0.301 0.763636
## BldgType2fmCon       6.847e-02  5.728e-02   1.195 0.232267
## BldgTypeDuplex      -8.900e-03  3.388e-02  -0.263 0.792878
## BldgTypeTwnhs       -5.347e-02  4.345e-02  -1.231 0.218818
## BldgTypeTwnhsE      -8.022e-03  4.066e-02  -0.197 0.843635
## HouseStyle1.5Unf    -8.002e-03  3.560e-02  -0.225 0.822232
## HouseStyle1Story     9.506e-03  2.011e-02   0.473 0.636489
## HouseStyle2.5Fin    -9.036e-02  5.621e-02  -1.608 0.108287
## HouseStyle2.5Unf    -1.180e-02  4.570e-02  -0.258 0.796307
## HouseStyle2Story    -4.113e-04  1.537e-02  -0.027 0.978652
## HouseStyleSFoyer     9.966e-03  2.832e-02   0.352 0.724981
## HouseStyleSLvl       3.552e-02  2.532e-02   1.403 0.161079
## RoofStyleHip         6.978e-03  8.681e-03   0.804 0.421713
## RoofStyleother       7.558e-02  2.418e-02   3.125 0.001838 **
## Exterior1stCemntBd  -1.257e-01  6.248e-02  -2.012 0.044577 *
## Exterior1stHdBoard  -6.933e-02  2.927e-02  -2.368 0.018093 *
## Exterior1stMetalSd  -3.478e-02  4.085e-02  -0.851 0.394889
## Exterior1stother    -3.534e-02  2.885e-02  -1.225 0.221003
## Exterior1stPlywood  -6.228e-02  2.792e-02  -2.231 0.025956 *
## Exterior1stVinylSd  -1.083e-01  4.045e-02  -2.677 0.007568 **
## Exterior1stWd Sdng  -1.109e-01  2.475e-02  -4.481 8.48e-06 ***
## Exterior2ndCmentBd   8.568e-02  6.589e-02   1.300 0.193894
## Exterior2ndHdBoard   2.256e-02  3.405e-02   0.662 0.507866
```

```
## Exterior2ndMetalSd   2.696e-03  4.489e-02   0.060 0.952124
## Exterior2ndother     2.327e-02  3.330e-02   0.699 0.484960
## Exterior2ndPlywood   1.401e-02  3.238e-02   0.433 0.665322
## Exterior2ndVinylSd   8.119e-02  4.412e-02   1.840 0.066075 .
## Exterior2ndWd Sdng   7.263e-02  2.985e-02   2.433 0.015186 *
## Exterior2ndWd Shng   5.550e-03  3.737e-02   0.148 0.881986
## MasVnrTypeBrkFace    4.927e-02  3.705e-02   1.330 0.183941
## MasVnrTypeNone       5.170e-02  3.733e-02   1.385 0.166488
## MasVnrTypeStone      6.659e-02  3.929e-02   1.695 0.090446 .
## MasVnrTypeNA         2.759e-02  5.840e-02   0.472 0.636783
## ExterQualAvg         3.688e-03  1.093e-02   0.337 0.735921
## ExterQualBelowAvg   -1.351e-01  4.639e-02  -2.912 0.003691 **
## ExterCondAvg         2.642e-02  1.084e-02   2.438 0.014974 *
## ExterCondBelowAvg    2.695e-02  3.022e-02   0.892 0.372767
## FoundationCBlock     1.428e-02  1.432e-02   0.997 0.319041
## Foundationother     -2.789e-02  2.857e-02  -0.976 0.329229
## FoundationPConc      3.867e-02  1.593e-02   2.427 0.015438 *
## BsmtQualAvg          1.872e-02  1.129e-02   1.658 0.097788 .
## BsmtQualBelowAvg     2.885e-02  2.539e-02   1.137 0.256022
## BsmtQualNA           2.559e-01  1.431e-01   1.788 0.074128 .
## BsmtCondAvg         -2.887e-02  1.649e-02  -1.751 0.080351 .
## BsmtCondBelowAvg    -4.948e-02  2.627e-02  -1.883 0.059984 .
## BsmtCondNA                  NA         NA      NA       NA
## BsmtExposureGd       2.820e-02  1.430e-02   1.972 0.048925 *
## BsmtExposureMn      -1.597e-02  1.454e-02  -1.099 0.272177
## BsmtExposureNo      -1.101e-02  1.063e-02  -1.036 0.300646
## BsmtExposureNA      -7.775e-02  8.912e-02  -0.872 0.383233
## BsmtFinType1BLQ     -1.481e-02  1.209e-02  -1.225 0.220993
## BsmtFinType1GLQ      1.245e-02  1.127e-02   1.105 0.269547
## BsmtFinType1LwQ     -3.091e-02  1.707e-02  -1.811 0.070507 .
## BsmtFinType1Rec     -1.985e-02  1.361e-02  -1.459 0.144874
## BsmtFinType1Unf     -1.112e-02  1.345e-02  -0.827 0.408628
## BsmtFinType1NA              NA         NA      NA       NA
## BsmtFinType2BLQ     -6.358e-02  3.712e-02  -1.713 0.087101 .
## BsmtFinType2GLQ      3.402e-03  4.122e-02   0.083 0.934242
## BsmtFinType2LwQ     -3.302e-02  3.706e-02  -0.891 0.373182
## BsmtFinType2Rec     -3.983e-02  3.662e-02  -1.088 0.277017
## BsmtFinType2Unf     -1.131e-02  4.023e-02  -0.281 0.778617
## BsmtFinType2NA      -1.872e-01  9.955e-02  -1.880 0.060444 .
## Heatingother         4.336e-02  2.431e-02   1.784 0.074868 .
## HeatingQCAvg        -1.684e-02  8.323e-03  -2.024 0.043293 *
## HeatingQCBelowAvg   -3.257e-02  1.932e-02  -1.685 0.092284 .
## CentralAirY          4.158e-02  1.823e-02   2.280 0.022835 *
## ElectricalFuseF     -7.491e-02  2.748e-02  -2.726 0.006550 **
## ElectricalFuseP     -1.006e-02  7.768e-02  -0.130 0.896967
## ElectricalSBrkr      2.362e-05  1.324e-02   0.002 0.998577
## ElectricalNA         7.785e-02  9.511e-02   0.819 0.413298
## KitchenQualAvg      -8.813e-03  9.772e-03  -0.902 0.367378
## KitchenQualBelowAvg -1.848e-02  2.402e-02  -0.769 0.441947
## FunctionalMaj2      -2.539e-01  5.685e-02  -4.466 9.09e-06 ***
## FunctionalMin1       2.588e-02  3.658e-02   0.707 0.479532
## FunctionalMin2       5.450e-02  3.665e-02   1.487 0.137337
## FunctionalMod        2.933e-03  4.222e-02   0.069 0.944642
## FunctionalTyp        7.262e-02  3.092e-02   2.349 0.019071 *
```

```
## FireplaceQuAvg      -1.038e-02  9.933e-03  -1.045 0.296203
## FireplaceQuBelowAvg -2.274e-02  1.652e-02  -1.377 0.169019
## FireplaceQuNA       -3.265e-02  1.585e-02  -2.060 0.039692 *
## GarageTypeAttchd     1.371e-01  6.025e-02   2.275 0.023147 *
## GarageTypeBasment    1.169e-01  6.872e-02   1.701 0.089360 .
## GarageTypeBuiltIn    1.329e-01  6.249e-02   2.128 0.033671 *
## GarageTypeCarPort    1.370e-01  7.828e-02   1.750 0.080555 .
## GarageTypeDetchd     1.301e-01  5.959e-02   2.184 0.029268 *
## GarageTypeNA        -3.334e-01  5.721e-01  -0.583 0.560265
## GarageFinishRFn     -4.936e-03  9.088e-03  -0.543 0.587178
## GarageFinishUnf     -1.135e-02  1.088e-02  -1.043 0.297220
## GarageFinishNA              NA         NA      NA       NA
## GarageQualAvg       -2.987e-02  3.870e-02  -0.772 0.440401
## GarageQualBelowAvg  -6.201e-02  4.457e-02  -1.391 0.164546
## GarageQualNA                NA         NA      NA       NA
## GarageCondAvg        2.715e-02  4.143e-02   0.655 0.512429
## GarageCondBelowAvg  -1.916e-02  4.687e-02  -0.409 0.682764
## GarageCondNA                NA         NA      NA       NA
## PavedDriveP         -2.408e-02  2.353e-02  -1.024 0.306321
## PavedDriveY          1.767e-03  1.523e-02   0.116 0.907666
## PoolQCGd             6.987e-02  1.322e-01   0.528 0.597336
## PoolQCNA            -1.015e-01  9.649e-02  -1.052 0.293070
## FenceGdWo           -6.290e-03  2.199e-02  -0.286 0.774912
## FenceMnPrv           1.646e-02  1.810e-02   0.910 0.363328
## FenceMnWw           -1.039e-02  3.709e-02  -0.280 0.779523
## FenceNA              3.212e-02  1.630e-02   1.971 0.049081 *
## MiscFeatureShed      7.517e-02  7.715e-02   0.974 0.330177
## MiscFeatureNA        7.489e-02  8.211e-02   0.912 0.361993
## SaleTypeWD          -1.147e-02  1.789e-02  -0.641 0.521516
## Id                  -1.054e-05  1.035e-05  -1.018 0.308785
## MSSubClass          -4.262e-04  3.780e-04  -1.127 0.259882
## LotFrontage          3.823e-05  1.017e-04   0.376 0.707061
## LotArea              2.865e-06  4.912e-07   5.831 7.88e-09 ***
## OverallQual          4.809e-02  4.557e-03  10.553  < 2e-16 ***
## OverallCond          4.089e-02  4.005e-03  10.210  < 2e-16 ***
## YearBuilt            2.175e-03  3.405e-04   6.389 2.78e-10 ***
## YearRemodAdd         5.850e-04  2.584e-04   2.263 0.023865 *
## MasVnrArea           1.916e-05  2.586e-05   0.741 0.459107
## BsmtFinSF1           1.701e-04  2.374e-05   7.163 1.74e-12 ***
## BsmtFinSF2           1.813e-04  4.846e-05   3.741 0.000196 ***
## BsmtUnfSF            1.109e-04  2.136e-05   5.189 2.66e-07 ***
## TotalBsmtSF                 NA         NA      NA       NA
## X1stFlrSF            2.598e-04  2.536e-05  10.244  < 2e-16 ***
## X2ndFlrSF            3.040e-04  2.630e-05  11.560  < 2e-16 ***
## LowQualFinSF         3.031e-04  8.739e-05   3.469 0.000549 ***
## GrLivArea                   NA         NA      NA       NA
## BsmtFullBath         2.961e-02  8.964e-03   3.304 0.000995 ***
## BsmtHalfBath         8.870e-03  1.386e-02   0.640 0.522337
## FullBath             1.141e-02  1.064e-02   1.072 0.284064
## HalfBath             4.779e-03  9.970e-03   0.479 0.631825
## BedroomAbvGr        -9.036e-03  6.399e-03  -1.412 0.158300
## KitchenAbvGr        -4.097e-02  2.887e-02  -1.419 0.156190
## TotRmsAbvGrd         2.742e-04  4.500e-03   0.061 0.951439
## Fireplaces           9.180e-03  1.164e-02   0.789 0.430615
```

```
## GarageYrBlt         -2.077e-04  2.910e-04  -0.714 0.475622
## GarageCars           3.385e-02  1.045e-02   3.239 0.001249 **
## GarageArea           5.882e-05  3.795e-05   1.550 0.121571
## WoodDeckSF           8.734e-05  2.699e-05   3.236 0.001259 **
## OpenPorchSF          1.482e-04  5.363e-05   2.764 0.005837 **
## EncPorchSF           2.045e-04  4.026e-05   5.080 4.67e-07 ***
## PoolArea                    NA         NA      NA       NA
## MiscVal              2.271e-05  2.792e-05   0.813 0.416349
## MoSold              -9.610e-04  1.164e-03  -0.826 0.409278
## YrSold              -9.116e-04  2.377e-03  -0.383 0.701452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08702 on 829 degrees of freedom
## Multiple R-squared:  0.9523, Adjusted R-squared:  0.9426
## F-statistic: 97.46 on 170 and 829 DF,  p-value: < 2.2e-16
```

We run the cross validation on this model to see what it looks like.

```
fitControl <- trainControl(method="cv", number=5
                           )

fitOLM <- train(log(SalePrice)~.,
             data=housingData,
             method="lm",
             trControl=fitControl)

fitOLM
```

```
## Linear Regression
##
## 1000 samples
##   73 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 800, 801, 800, 800, 799
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.1052332  0.9174586  0.07828594
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

There are lots of variables here. Let's see if we can reduce the number of variables. The following code starts with the complete model from above and adds or removes features based on AIC, attempting to minimize this value. We have hidden the results because they were very long.

```
OLMSim <- stepAIC(fit, direction = "both")
```

Below is the formula for the best found in the above process.

```
OLMSim$call$formula
```

```
## log(SalePrice) ~ MSZoning + LotConfig + LandSlope + Neighborhood +
##     Condition1 + BldgType + RoofStyle + Exterior1st + Exterior2nd +
##     ExterQual + ExterCond + Foundation + BsmtExposure + BsmtFinType1 +
```

```
##     BsmtFinType2 + Heating + HeatingQC + CentralAir + Electrical +
##     Functional + FireplaceQu + GarageCond + PoolQC + Fence +
##     MSSubClass + LotArea + OverallQual + OverallCond + YearBuilt +
##     YearRemodAdd + BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + X1stFlrSF +
##     X2ndFlrSF + LowQualFinSF + BsmtFullBath + BedroomAbvGr +
##     GarageCars + GarageArea + WoodDeckSF + OpenPorchSF + EncPorchSF
```

```
OLMSim$coefficients
```

```
##         (Intercept)           MSZoningRH           MSZoningRL           MSZoningRM
##        4.919190e+00        -3.173022e-02         7.229467e-03        -7.992734e-02
##    LotConfigCulDSac      LotConfigInside       LotConfigother          LandSlopeMod
##        2.238491e-02        -1.150532e-02        -2.996273e-02         9.419786e-04
##         LandSlopeSev NeighborhoodClearCr NeighborhoodCollgCr NeighborhoodCrawfor
##       -1.055136e-01        -2.262315e-02        -7.013679e-02         8.897206e-02
## NeighborhoodEdwards NeighborhoodGilbert  NeighborhoodIDOTRR NeighborhoodMitchel
##       -9.770325e-02        -8.607437e-02         1.526584e-02        -9.733684e-02
##   NeighborhoodNAmes NeighborhoodNoRidge NeighborhoodNridgHt  NeighborhoodNWAmes
##       -6.426964e-02        -7.410666e-02         4.001278e-03        -7.243534e-02
## NeighborhoodOldTown   Neighborhoodother  NeighborhoodSawyer NeighborhoodSawyerW
##       -6.162466e-02        -5.471020e-02        -7.818490e-02        -8.194072e-02
## NeighborhoodSomerst  NeighborhoodTimber      Condition1Feedr       Condition1Norm
##       -2.097012e-03        -5.679667e-02         1.758743e-02         5.609129e-02
##      Condition1PosA       Condition1PosN         Condition1RR       BldgType2fmCon
##       -3.041026e-02        -4.481193e-03         4.716947e-05         4.870701e-02
##      BldgTypeDuplex        BldgTypeTwnhs       BldgTypeTwnhsE          RoofStyleHip
##       -4.278601e-02        -6.277270e-02        -1.413489e-02         9.080553e-03
##      RoofStyleother  Exterior1stCemntBd  Exterior1stHdBoard  Exterior1stMetalSd
##        8.052711e-02        -1.196717e-01        -6.914775e-02        -3.336394e-02
##    Exterior1stother  Exterior1stPlywood  Exterior1stVinylSd  Exterior1stWd Sdng
##       -4.126081e-02        -6.305208e-02        -1.007791e-01        -1.114008e-01
##   Exterior2ndCmentBd  Exterior2ndHdBoard  Exterior2ndMetalSd    Exterior2ndother
##        7.673617e-02         2.465389e-02        -8.731670e-04         2.516094e-02
##   Exterior2ndPlywood  Exterior2ndVinylSd  Exterior2ndWd Sdng  Exterior2ndWd Shng
##        9.266997e-03         7.240285e-02         6.946790e-02         6.472277e-03
##         ExterQualAvg   ExterQualBelowAvg         ExterCondAvg   ExterCondBelowAvg
##        3.253342e-03        -1.173598e-01         2.230425e-02         2.297629e-02
##    FoundationCBlock      Foundationother       FoundationPConc        BsmtExposureGd
##        2.034944e-02        -2.291767e-02         4.284088e-02         3.519998e-02
##      BsmtExposureMn       BsmtExposureNo       BsmtExposureNA      BsmtFinType1BLQ
##       -1.576690e-02        -1.232520e-02        -7.344438e-02        -1.912129e-02
##     BsmtFinType1GLQ      BsmtFinType1LwQ      BsmtFinType1Rec      BsmtFinType1Unf
##        6.823609e-03        -3.425314e-02        -2.365417e-02        -1.680508e-02
##      BsmtFinType1NA      BsmtFinType2BLQ      BsmtFinType2GLQ      BsmtFinType2LwQ
##        2.537641e-01        -7.109323e-02        -1.980149e-03        -4.124217e-02
##     BsmtFinType2Rec      BsmtFinType2Unf       BsmtFinType2NA         Heatingother
##       -4.496115e-02        -2.404294e-02        -1.965666e-01         4.083865e-02
##        HeatingQCAvg   HeatingQCBelowAvg          CentralAirY      ElectricalFuseF
##       -1.890933e-02        -4.248829e-02         4.590202e-02        -6.609621e-02
##     ElectricalFuseP       ElectricalSBrkr         ElectricalNA       FunctionalMaj2
##       -4.207347e-02         1.585508e-03         1.123706e-01        -2.322549e-01
##      FunctionalMin1       FunctionalMin2        FunctionalMod        FunctionalTyp
##        2.129487e-02         4.710104e-02        -1.836323e-03         7.256940e-02
##      FireplaceQuAvg FireplaceQuBelowAvg        FireplaceQuNA        GarageCondAvg
##       -1.396150e-02        -2.421613e-02        -4.522487e-02         2.959014e-02
```

```
##  GarageCondBelowAvg        GarageCondNA           PoolQCGd           PoolQCNA
##       -3.707521e-02       -1.947258e-02       8.772475e-02      -1.100343e-01
##           FenceGdWo          FenceMnPrv          FenceMnWw            FenceNA
##       -2.615817e-03        1.472555e-02       9.554331e-04       3.052037e-02
##          MSSubClass             LotArea        OverallQual        OverallCond
##       -3.553012e-04        2.684505e-06       5.189533e-02       4.321700e-02
##           YearBuilt        YearRemodAdd         BsmtFinSF1         BsmtFinSF2
##        2.411541e-03        5.855832e-04       1.639874e-04       1.682237e-04
##           BsmtUnfSF           X1stFlrSF          X2ndFlrSF       LowQualFinSF
##        1.002439e-04        2.756585e-04       2.966282e-04       2.256821e-04
##        BsmtFullBath        BedroomAbvGr         GarageCars         GarageArea
##        2.558566e-02       -8.730571e-03       3.346638e-02       5.214420e-05
##          WoodDeckSF         OpenPorchSF        EncPorchSF
##        8.119717e-05        1.489328e-04       2.104590e-04
```

Now that we have found this linear fit, we run a 5 fold cross validation to examine the

```r
fitOLMSim <- train(OLMSim$call$formula,
             data=housingData,
             method="lm",
             trControl=fitControl)

fitOLMSim
```

```
## Linear Regression
##
## 1000 samples
##   43 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 802, 798, 800, 800, 800
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.0970649  0.9296133  0.07297047
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```r
fitOLMSim$results
```

```
##   intercept      RMSE  Rsquared        MAE      RMSESD RsquaredSD       MAESD
## 1      TRUE 0.0970649 0.9296133 0.07297047 0.008839826 0.01173315 0.003948156
```

```r
summary(fitOLMSim)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60607 -0.04622  0.00212  0.05329  0.27018
##
## Coefficients: (16 not defined because of singularities)
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.919e+00  6.330e-01   7.772 2.15e-14 ***
```

```
## MSZoningRH           -3.173e-02  4.729e-02  -0.671 0.502444
## MSZoningRL            7.229e-03  3.666e-02   0.197 0.843734
## MSZoningRM           -7.993e-02  3.798e-02  -2.105 0.035600 *
## MSZoningNA                   NA         NA      NA       NA
## LotConfigCulDSac      2.238e-02  1.316e-02   1.702 0.089189 .
## LotConfigInside      -1.151e-02  7.812e-03  -1.473 0.141157
## LotConfigother       -2.996e-02  1.643e-02  -1.824 0.068566 .
## LotConfigNA                  NA         NA      NA       NA
## LandSlopeMod          9.420e-04  1.508e-02   0.062 0.950220
## LandSlopeSev         -1.055e-01  5.152e-02  -2.048 0.040840 *
## LandSlopeNA                  NA         NA      NA       NA
## NeighborhoodClearCr  -2.262e-02  2.860e-02  -0.791 0.429163
## NeighborhoodCollgCr  -7.014e-02  2.275e-02  -3.083 0.002113 **
## NeighborhoodCrawfor   8.897e-02  2.416e-02   3.683 0.000244 ***
## NeighborhoodEdwards  -9.770e-02  2.039e-02  -4.792 1.94e-06 ***
## NeighborhoodGilbert  -8.607e-02  2.515e-02  -3.422 0.000650 ***
## NeighborhoodIDOTRR    1.527e-02  2.622e-02   0.582 0.560596
## NeighborhoodMitchel  -9.734e-02  2.462e-02  -3.953 8.34e-05 ***
## NeighborhoodNAmes    -6.427e-02  1.981e-02  -3.245 0.001220 **
## NeighborhoodNoRidge  -7.411e-02  2.805e-02  -2.642 0.008398 **
## NeighborhoodNridgHt   4.001e-03  2.831e-02   0.141 0.887637
## NeighborhoodNWAmes   -7.244e-02  2.375e-02  -3.050 0.002353 **
## NeighborhoodOldTown  -6.162e-02  1.989e-02  -3.098 0.002010 **
## Neighborhoodother    -5.471e-02  2.249e-02  -2.433 0.015179 *
## NeighborhoodSawyer   -7.818e-02  2.263e-02  -3.455 0.000577 ***
## NeighborhoodSawyerW  -8.194e-02  2.472e-02  -3.315 0.000955 ***
## NeighborhoodSomerst  -2.097e-03  3.886e-02  -0.054 0.956972
## NeighborhoodTimber   -5.680e-02  3.017e-02  -1.883 0.060071 .
## NeighborhoodNA               NA         NA      NA       NA
## Condition1Feedr       1.759e-02  2.180e-02   0.807 0.419939
## Condition1Norm        5.609e-02  1.802e-02   3.113 0.001909 **
## Condition1PosA       -3.041e-02  4.040e-02  -0.753 0.451836
## Condition1PosN       -4.481e-03  3.059e-02  -0.146 0.883582
## Condition1RR          4.717e-05  2.604e-02   0.002 0.998555
## Condition1NA                 NA         NA      NA       NA
## BldgType2fmCon        4.871e-02  3.696e-02   1.318 0.187894
## BldgTypeDuplex       -4.279e-02  2.279e-02  -1.877 0.060818 .
## BldgTypeTwnhs        -6.277e-02  2.981e-02  -2.106 0.035520 *
## BldgTypeTwnhsE       -1.413e-02  2.465e-02  -0.573 0.566547
## BldgTypeNA                   NA         NA      NA       NA
## RoofStyleHip          9.081e-03  8.245e-03   1.101 0.271058
## RoofStyleother        8.053e-02  2.246e-02   3.586 0.000354 ***
## RoofStyleNA                  NA         NA      NA       NA
## Exterior1stCemntBd   -1.197e-01  5.955e-02  -2.010 0.044771 *
## Exterior1stHdBoard   -6.915e-02  2.777e-02  -2.490 0.012953 *
## Exterior1stMetalSd   -3.336e-02  3.893e-02  -0.857 0.391702
## Exterior1stother     -4.126e-02  2.771e-02  -1.489 0.136858
## Exterior1stPlywood   -6.305e-02  2.678e-02  -2.354 0.018779 *
## Exterior1stVinylSd   -1.008e-01  3.820e-02  -2.638 0.008479 **
## `Exterior1stWd Sdng` -1.114e-01  2.380e-02  -4.681 3.30e-06 ***
## Exterior1stNA                NA         NA      NA       NA
## Exterior2ndCmentBd    7.674e-02  6.261e-02   1.226 0.220690
## Exterior2ndHdBoard    2.465e-02  3.263e-02   0.756 0.450103
## Exterior2ndMetalSd   -8.732e-04  4.281e-02  -0.020 0.983732
```

```
## Exterior2ndother      2.516e-02  3.206e-02   0.785 0.432770
## Exterior2ndPlywood    9.267e-03  3.112e-02   0.298 0.765904
## Exterior2ndVinylSd    7.240e-02  4.188e-02   1.729 0.084213 .
## `Exterior2ndWd Sdng`  6.947e-02  2.904e-02   2.392 0.016968 *
## `Exterior2ndWd Shng`  6.472e-03  3.573e-02   0.181 0.856295
## Exterior2ndNA                NA         NA      NA       NA
## ExterQualAvg          3.253e-03  9.939e-03   0.327 0.743506
## ExterQualBelowAvg    -1.174e-01  4.170e-02  -2.814 0.004994 **
## ExterQualNA                  NA         NA      NA       NA
## ExterCondAvg          2.230e-02  1.015e-02   2.197 0.028255 *
## ExterCondBelowAvg     2.298e-02  2.812e-02   0.817 0.414054
## ExterCondNA                  NA         NA      NA       NA
## FoundationCBlock      2.035e-02  1.347e-02   1.511 0.131222
## Foundationother      -2.292e-02  2.708e-02  -0.846 0.397630
## FoundationPConc       4.284e-02  1.509e-02   2.840 0.004616 **
## FoundationNA                 NA         NA      NA       NA
## BsmtExposureGd        3.520e-02  1.368e-02   2.573 0.010239 *
## BsmtExposureMn       -1.577e-02  1.372e-02  -1.149 0.250742
## BsmtExposureNo       -1.233e-02  1.001e-02  -1.232 0.218385
## BsmtExposureNA       -7.344e-02  8.778e-02  -0.837 0.403014
## BsmtFinType1BLQ      -1.912e-02  1.162e-02  -1.646 0.100222
## BsmtFinType1GLQ       6.824e-03  1.082e-02   0.631 0.528361
## BsmtFinType1LwQ      -3.425e-02  1.626e-02  -2.107 0.035391 *
## BsmtFinType1Rec      -2.365e-02  1.287e-02  -1.838 0.066466 .
## BsmtFinType1Unf      -1.681e-02  1.292e-02  -1.301 0.193552
## BsmtFinType1NA        2.538e-01  1.380e-01   1.838 0.066336 .
## BsmtFinType2BLQ      -7.109e-02  3.613e-02  -1.967 0.049441 *
## BsmtFinType2GLQ      -1.980e-03  3.996e-02  -0.050 0.960487
## BsmtFinType2LwQ      -4.124e-02  3.588e-02  -1.150 0.250628
## BsmtFinType2Rec      -4.496e-02  3.539e-02  -1.270 0.204245
## BsmtFinType2Unf      -2.404e-02  3.906e-02  -0.616 0.538364
## BsmtFinType2NA       -1.966e-01  9.733e-02  -2.019 0.043738 *
## Heatingother          4.084e-02  2.238e-02   1.825 0.068328 .
## HeatingNA                    NA         NA      NA       NA
## HeatingQCAvg         -1.891e-02  7.826e-03  -2.416 0.015886 *
## HeatingQCBelowAvg    -4.249e-02  1.813e-02  -2.343 0.019342 *
## HeatingQCNA                  NA         NA      NA       NA
## CentralAirY           4.590e-02  1.708e-02   2.688 0.007323 **
## CentralAirNA                 NA         NA      NA       NA
## ElectricalFuseF      -6.610e-02  2.619e-02  -2.524 0.011782 *
## ElectricalFuseP      -4.207e-02  6.782e-02  -0.620 0.535181
## ElectricalSBrkr       1.586e-03  1.262e-02   0.126 0.900058
## ElectricalNA          1.124e-01  9.179e-02   1.224 0.221178
## FunctionalMaj2       -2.323e-01  5.438e-02  -4.271 2.16e-05 ***
## FunctionalMin1        2.129e-02  3.444e-02   0.618 0.536507
## FunctionalMin2        4.710e-02  3.378e-02   1.394 0.163560
## FunctionalMod        -1.836e-03  3.925e-02  -0.047 0.962697
## FunctionalTyp         7.257e-02  2.860e-02   2.537 0.011342 *
## FunctionalNA                 NA         NA      NA       NA
## FireplaceQuAvg       -1.396e-02  9.489e-03  -1.471 0.141542
## FireplaceQuBelowAvg  -2.422e-02  1.601e-02  -1.513 0.130699
## FireplaceQuNA        -4.522e-02  8.934e-03  -5.062 5.05e-07 ***
## GarageCondAvg         2.959e-02  3.784e-02   0.782 0.434385
## GarageCondBelowAvg   -3.708e-02  4.205e-02  -0.882 0.378153
```

```
## GarageCondNA         -1.947e-02  4.157e-02  -0.468 0.639566
## PoolQCGd              8.772e-02  1.287e-01   0.682 0.495703
## PoolQCNA             -1.100e-01  9.383e-02  -1.173 0.241251
## FenceGdWo            -2.616e-03  2.115e-02  -0.124 0.901577
## FenceMnPrv            1.473e-02  1.736e-02   0.848 0.396554
## FenceMnWw             9.554e-04  3.593e-02   0.027 0.978789
## FenceNA               3.052e-02  1.552e-02   1.967 0.049533 *
## MSSubClass           -3.553e-04  2.100e-04  -1.692 0.091048 .
## LotArea               2.684e-06  4.218e-07   6.365 3.14e-10 ***
## OverallQual           5.190e-02  4.254e-03  12.200  < 2e-16 ***
## OverallCond           4.322e-02  3.768e-03  11.470  < 2e-16 ***
## YearBuilt             2.412e-03  2.826e-04   8.532  < 2e-16 ***
## YearRemodAdd          5.856e-04  2.294e-04   2.553 0.010856 *
## BsmtFinSF1            1.640e-04  2.173e-05   7.545 1.13e-13 ***
## BsmtFinSF2            1.682e-04  4.690e-05   3.587 0.000353 ***
## BsmtUnfSF             1.002e-04  1.971e-05   5.085 4.49e-07 ***
## X1stFlrSF             2.757e-04  2.004e-05  13.754  < 2e-16 ***
## X2ndFlrSF             2.966e-04  1.378e-05  21.519  < 2e-16 ***
## LowQualFinSF          2.257e-04  6.813e-05   3.312 0.000963 ***
## BsmtFullBath          2.559e-02  8.023e-03   3.189 0.001479 **
## BedroomAbvGr         -8.731e-03  5.422e-03  -1.610 0.107727
## GarageCars            3.347e-02  9.792e-03   3.418 0.000661 ***
## GarageArea            5.214e-05  3.429e-05   1.521 0.128708
## WoodDeckSF            8.120e-05  2.599e-05   3.125 0.001838 **
## OpenPorchSF           1.489e-04  5.107e-05   2.916 0.003634 **
## EncPorchSF            2.105e-04  3.845e-05   5.473 5.76e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08646 on 881 degrees of freedom
## Multiple R-squared:   0.95,  Adjusted R-squared:  0.9433
## F-statistic: 141.9 on 118 and 881 DF,  p-value: < 2.2e-16
```

Now we start to examine the appropriateness of this as a linear model. We look at many of the diagnostics available to us.

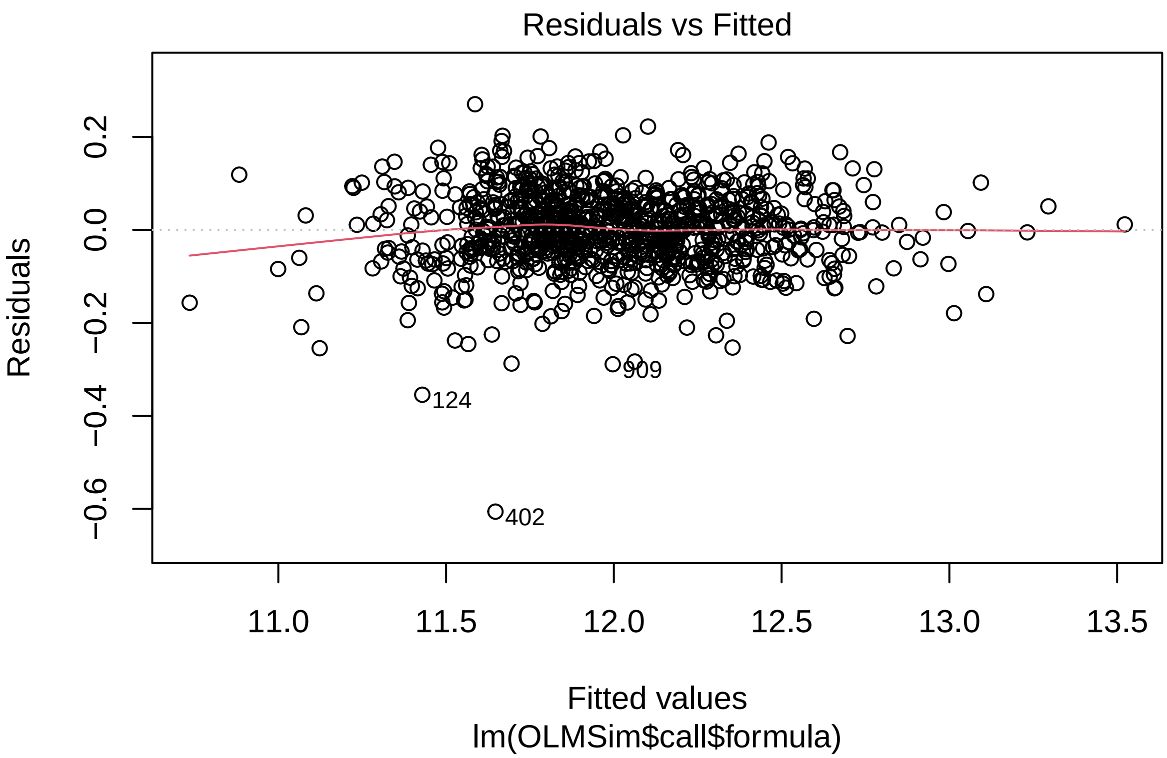
```r
fit <- lm(OLMSim$call$formula,
             data=housingData)
AIC(fit)
```

```
## [1] -1945.017
```

```r
BIC(fit)
```

```
## [1] -1356.086
```

```r
plot(fit, which = 1)
```



Looking at the diagnostics here, we see a decent spread and while there are a few identified outliers, it is not too bad. Diagnostically, we are looking for 1.Points above and Below the line 2. no patterns 3. no change in variation (no cones) 4. normal distribution

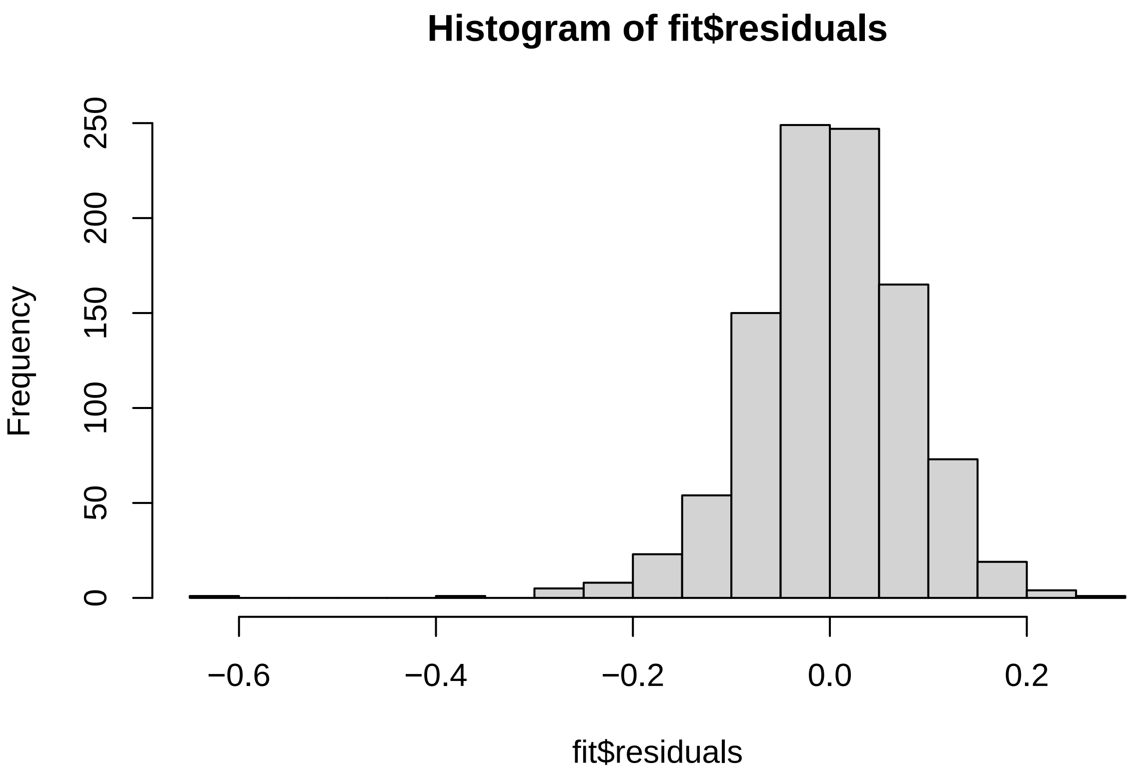
While we think we have this, the hypothesis test shows otherwise.

```r
ncvTest(fit)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 25.03359, Df = 1, p = 5.634e-07
```

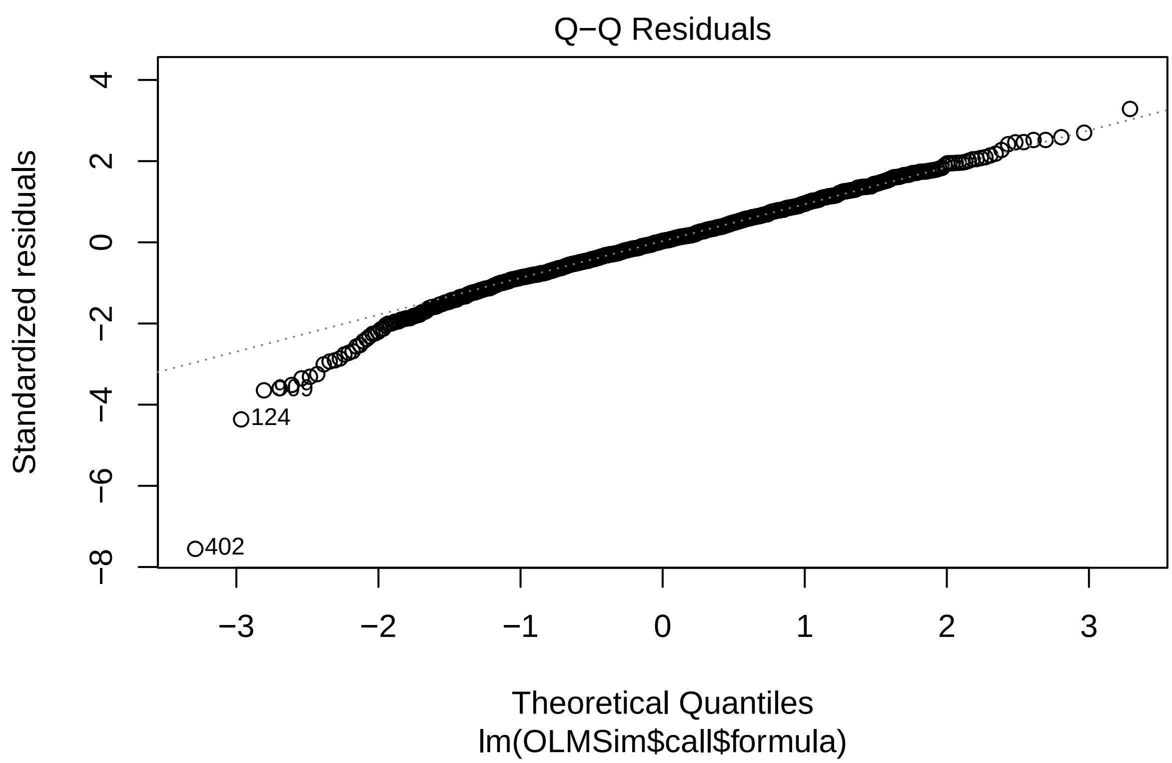
Oh, this is a problem! We are able to reject the null hypothesis here, we have evidence to suggest that the error variance changes with the fitted values. Let's keep looking at some of the other results we can get.

```r
hist(fit$residuals, breaks = 20)
```



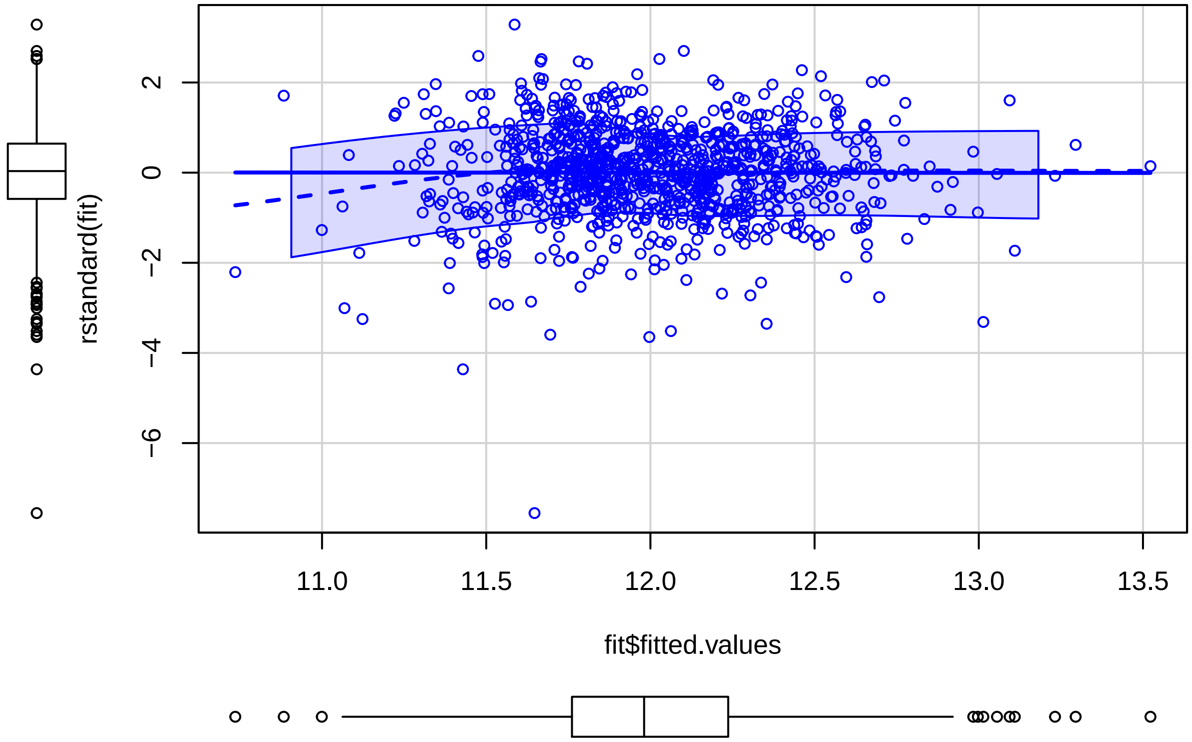
This looks fairly normal but with a longer tail on the negative side than the positive. Maybe qq-plot will show me something?

```
plot(fit, which = 2)
```



Here we see our deviance from normal on that low end once again. Next we look at the standardized residuals versus the predicted.

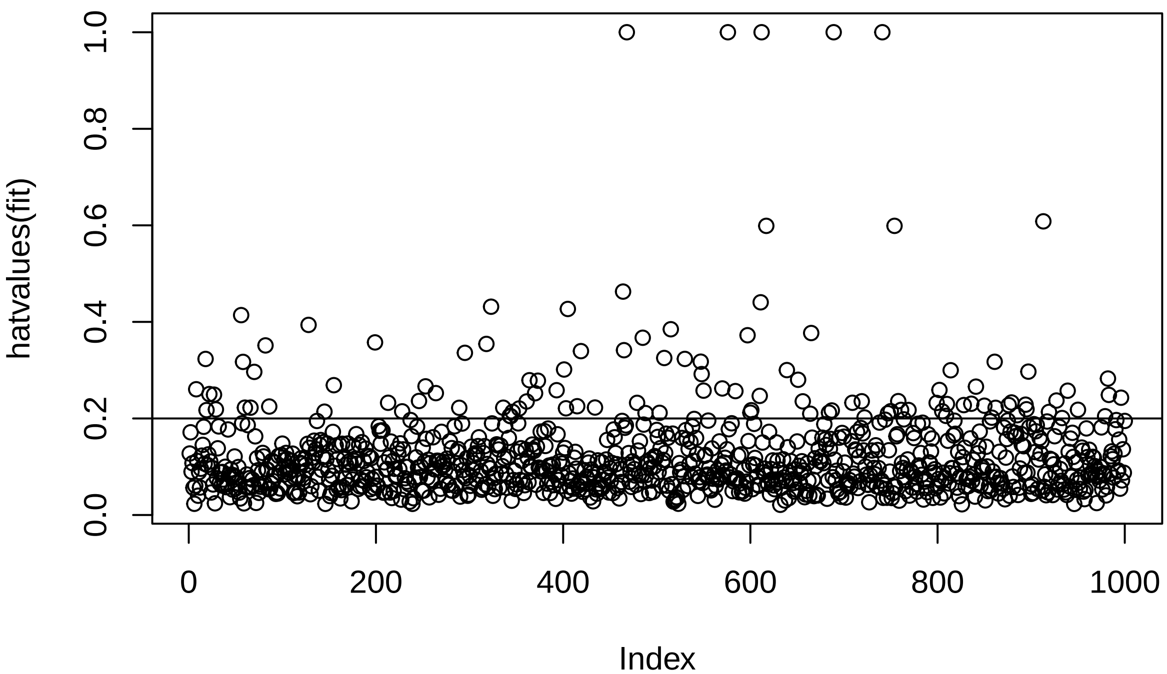
```r
scatterplot(fit$fitted.values,rstandard(fit))
```



Here we see plenty of outer space values (over 4) and some really strange (over 3)

Next we look for values with high leverage.

```r
plot(index = 1:1000,hatvalues(fit))
abline(h = 0.2)
```



We see lots with of points with high leverage (above the 0.2 line)

```r
sum(hatvalues(fit)>0.2)
```

```
## [1] 112
```

10% of our data has extreme leverage.

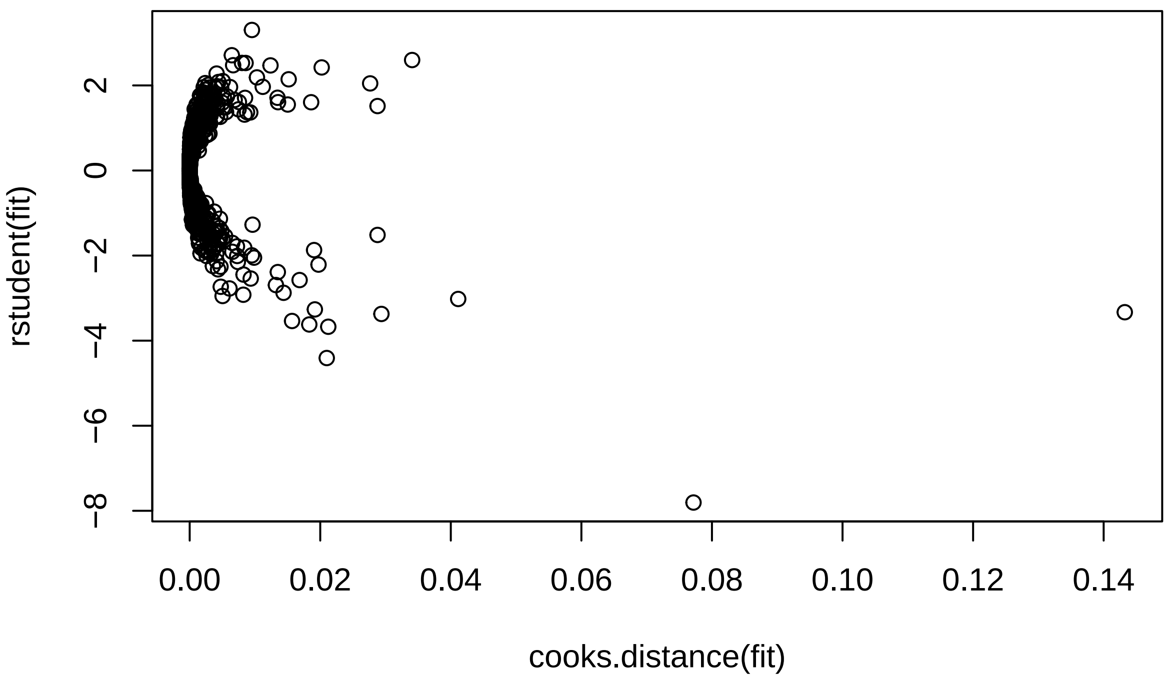
Let's test for outliers.

```r
outlierTest(fit)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 402 -7.805872         1.6746e-14   1.6662e-11
## 124 -4.408565         1.1691e-05   1.1633e-02
```

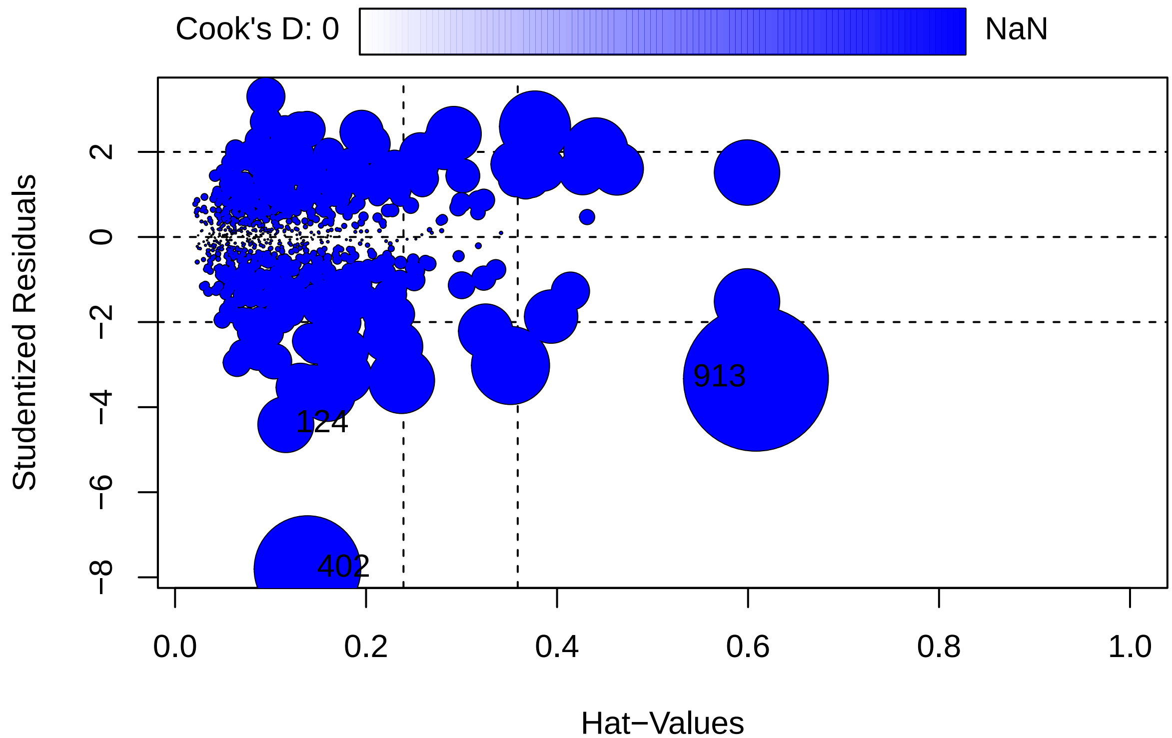
We find two outliers that are influential. Lastly we look at Cook's Distance.

```r
plot(cooks.distance(fit),rstudent(fit))
```



Not terrible but still some rather extreme points.

```r
influencePlot(fit)
```




```
##        StudRes       Hat      CookD
## 124 -4.408565 0.1160217 0.02099670
## 402 -7.805872 0.1386546 0.07717395
## 468       NaN 1.0000000        NaN
## 576       NaN 1.0000000        NaN
## 913 -3.332034 0.6082800 0.14323465
```

This one is nice for seeing these issues.

```
vif(fit)
```

```
##                      GVIF Df GVIF^(1/(2*Df))
## MSZoning     2.632483e+01  3        1.724756
## LotConfig    1.570828e+00  3        1.078172
## LandSlope    2.890548e+00  2        1.303902
## Neighborhood 4.694711e+03 17        1.282297
## Condition1   2.715919e+00  5        1.105075
## BldgType     3.492038e+01  4        1.559139
## RoofStyle    1.882749e+00  2        1.171381
## Exterior1st  6.476102e+05  7        2.600721
## Exterior2nd  6.803058e+05  8        2.314963
## ExterQual    4.459336e+00  2        1.453174
## ExterCond    1.934721e+00  2        1.179382
## Foundation   1.393135e+01  3        1.551192
## BsmtExposure 8.926146e+01  4        1.753206
## BsmtFinType1 9.093616e+02  6        1.764255
## BsmtFinType2 4.117792e+02  6        1.651538
## Heating      1.696310e+00  1        1.302425
## HeatingQC    2.388675e+00  2        1.243195
## CentralAir   2.336910e+00  1        1.528695
## Electrical   2.659844e+00  4        1.130074
## Functional   3.153533e+00  5        1.121708
```

```
## FireplaceQu  3.254549e+00  3        1.217349
## GarageCond   3.172752e+00  3        1.212196
## PoolQC       1.326112e+00  2        1.073112
## Fence        2.017535e+00  4        1.091698
## MSSubClass   1.033681e+01  1        3.215090
## LotArea      2.349073e+00  1        1.532669
## OverallQual  4.150597e+00  1        2.037302
## OverallCond  2.353094e+00  1        1.533980
## YearBuilt    9.052035e+00  1        3.008660
## YearRemodAdd 2.845816e+00  1        1.686955
## BsmtFinSF1   1.039847e+01  1        3.224666
## BsmtFinSF2   6.657995e+00  1        2.580309
## BsmtUnfSF    9.071987e+00  1        3.011974
## X1stFlrSF    6.608531e+00  1        2.570706
## X2ndFlrSF    4.617227e+00  1        2.148773
## LowQualFinSF 1.272909e+00  1        1.128232
## BsmtFullBath 2.227698e+00  1        1.492547
## BedroomAbvGr 2.457409e+00  1        1.567613
## GarageCars   6.537265e+00  1        2.556807
## GarageArea   6.147194e+00  1        2.479353
## WoodDeckSF   1.458845e+00  1        1.207827
## OpenPorchSF  1.336293e+00  1        1.155981
## EncPorchSF   1.333178e+00  1        1.154633
```

There is definitely some co-linearity since the average is way over one.

We consider dropping Exterior2nd (as it's similar to Exterior1st), BsmtFinType2 (since BsmtFinType1 already accounts for basement type), and GarageCars (as GarageArea is a more interpretable metric).

```r
# Update the formula string
new_formula <- update(OLMSim$call$formula, . ~ . - Exterior2nd - BsmtUnfSF - GarageArea)


fitOLMSim <- train(new_formula,
             data=housingData,
             method="lm",
             trControl=fitControl)

fitOLMSim
```

```
## Linear Regression
##
## 1000 samples
##   40 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 801, 800, 800, 799, 800
## Resampling results:
##
##   RMSE        Rsquared   MAE
##   0.09729613  0.9282711  0.07303413
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
fitOLMSim$results
```

```
##   intercept       RMSE  Rsquared        MAE      RMSESD RsquaredSD       MAESD
## 1      TRUE 0.09729613 0.9282711 0.07303413 0.008775605 0.01390115 0.005548732
```

The reduction in RMSE and improvement in R-squared indicate that the removal of the multi-co-linear variables improved the predictive accuracy and overall model fit.

One more fancy visualization with `ggplot` and some residuals. Not super useful but it was the first I put together and it looks pretty.

```
housingData$Predicted <- predict(fit)
housingData$logSalePrice <- log(housingData$SalePrice)
ggplot(housingData, aes(x = YearBuilt, y = logSalePrice, color = Exterior1st)) +
  geom_segment(aes(xend = YearBuilt, yend = Predicted), alpha = .2) +
  geom_point()+
  geom_point(aes(y = Predicted), shape = 1)+
  facet_wrap(Neighborhood~.)
```

1.b.

Next we do the plsr model. We'll use the entire dataset but we do need to remove the columns that were added for visualization purposes.

```
model.pls <- pls::plsr(log(SalePrice) ~ .-logSalePrice - Predicted, 16, data = housingData, #method = "
                        validation = "CV")
summary(model.pls)
```

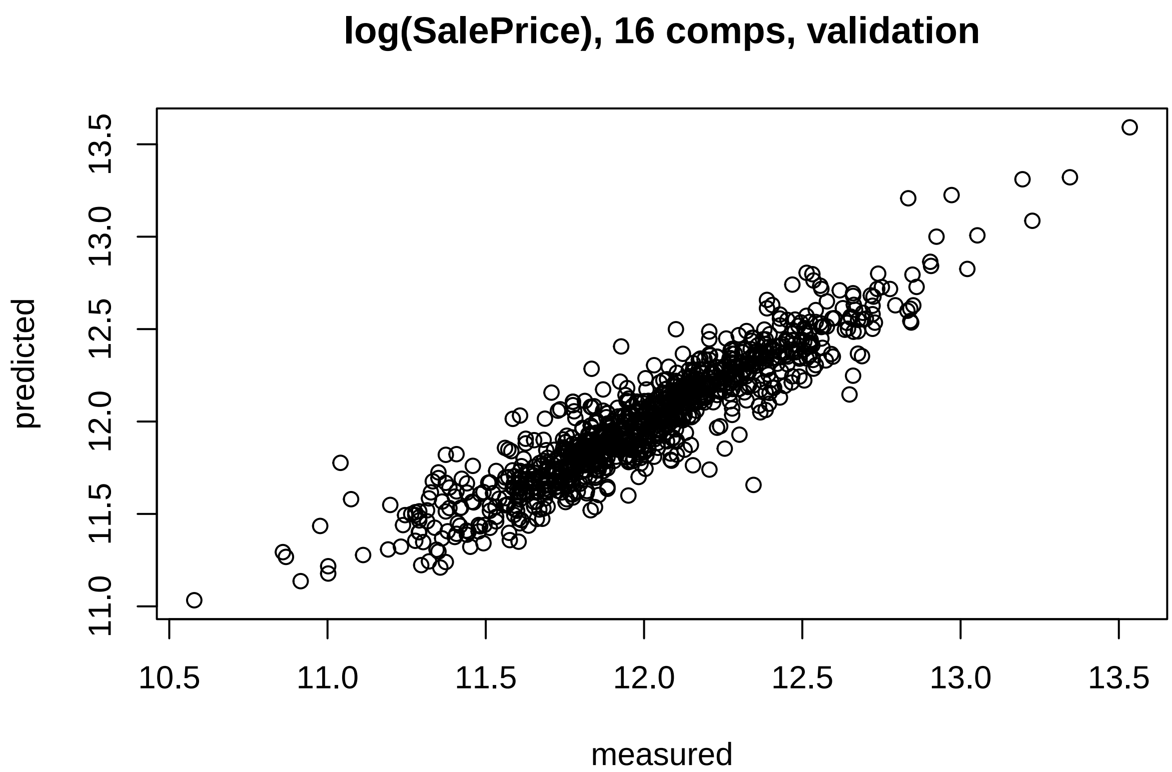
```
## Data:    X dimension: 1000 200
##  Y dimension: 1000 1
## Fit method: kernelpls
```

```
## Number of components considered: 16
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV          0.3633   0.3526   0.1856   0.1808   0.1759   0.1736   0.1706
## adjCV       0.3633   0.3522   0.1855   0.1807   0.1756   0.1734   0.1706
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV      0.1696   0.1609   0.1542    0.1513    0.1472    0.1402    0.1390
## adjCV   0.1695   0.1592   0.1505    0.1506    0.1473    0.1399    0.1386
##        14 comps  15 comps  16 comps
## CV       0.1389    0.1389    0.1384
## adjCV    0.1386    0.1387    0.1382
##
## TRAINING: % variance explained
##                 1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X                98.539    98.98    99.23    99.34    99.57    99.67     99.8
## log(SalePrice)    9.013    74.23    75.70    77.40    78.00    78.65     79.0
##                 8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## X                 99.81    99.83     99.90     99.93     99.95     99.96
## log(SalePrice)    81.79    83.65     83.96     84.60     85.93     86.14
##                 14 comps  15 comps  16 comps
## X                  99.96     99.99    100.00
## log(SalePrice)     86.18     86.19     86.21
```

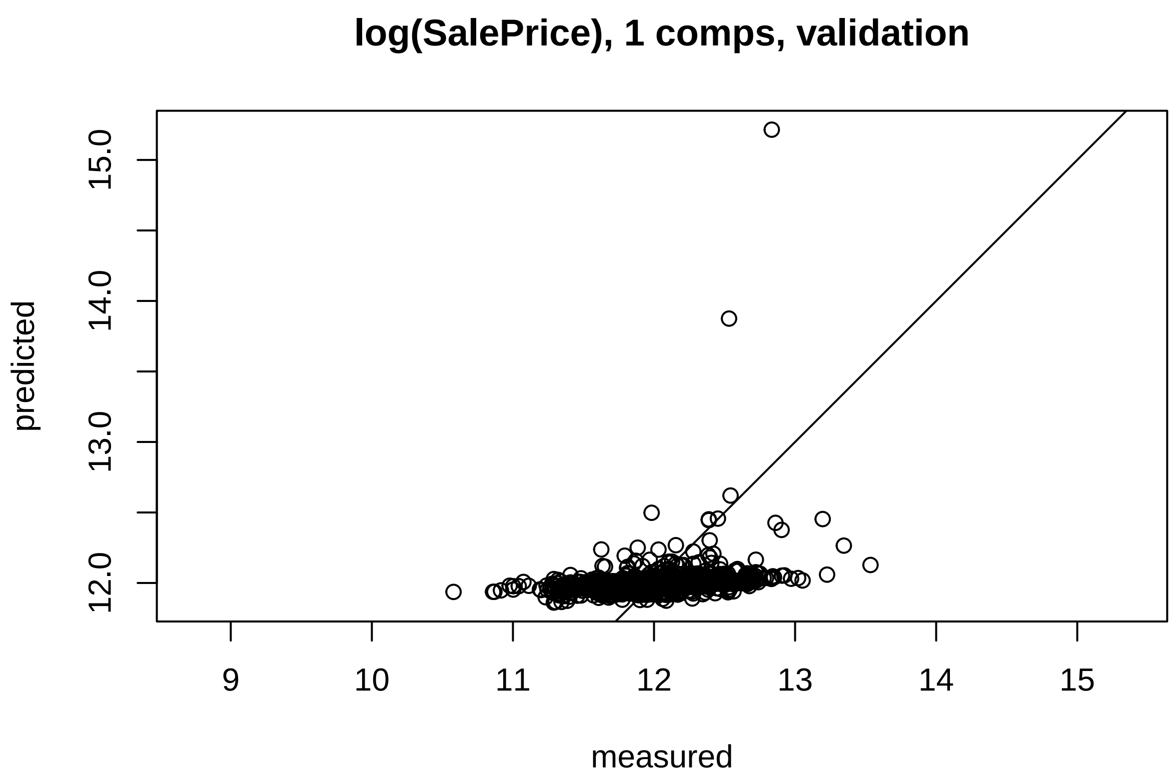
```
model.pls$validation$PRESS
```

```
##                 1 comps  2 comps  3 comps 4 comps  5 comps  6 comps  7 comps
## log(SalePrice) 124.3355 34.43642 32.69954 30.9366 30.12723 29.12028 28.75687
##                 8 comps  9 comps 10 comps 11 comps 12 comps 13 comps 14 comps
## log(SalePrice) 25.90237 23.77716 22.89701 21.66823 19.66376 19.30935 19.30702
##                15 comps 16 comps
## log(SalePrice) 19.29898 19.16101
```

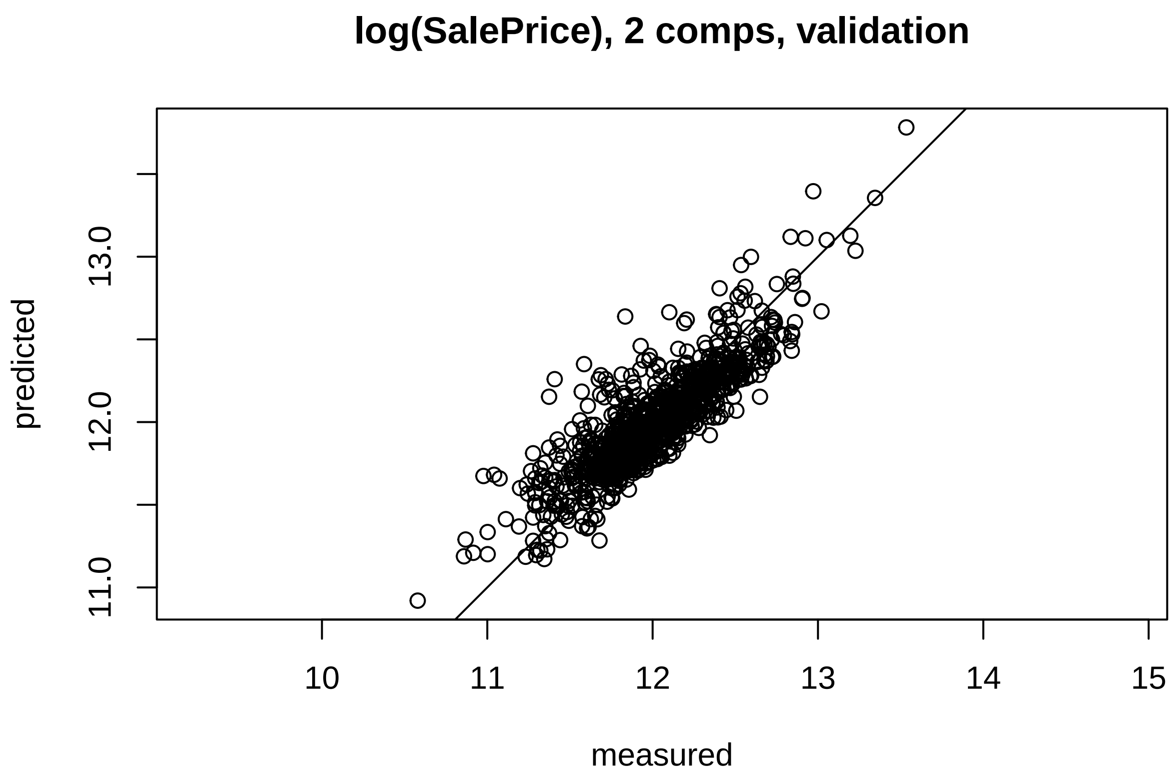
```
plot(model.pls)
```



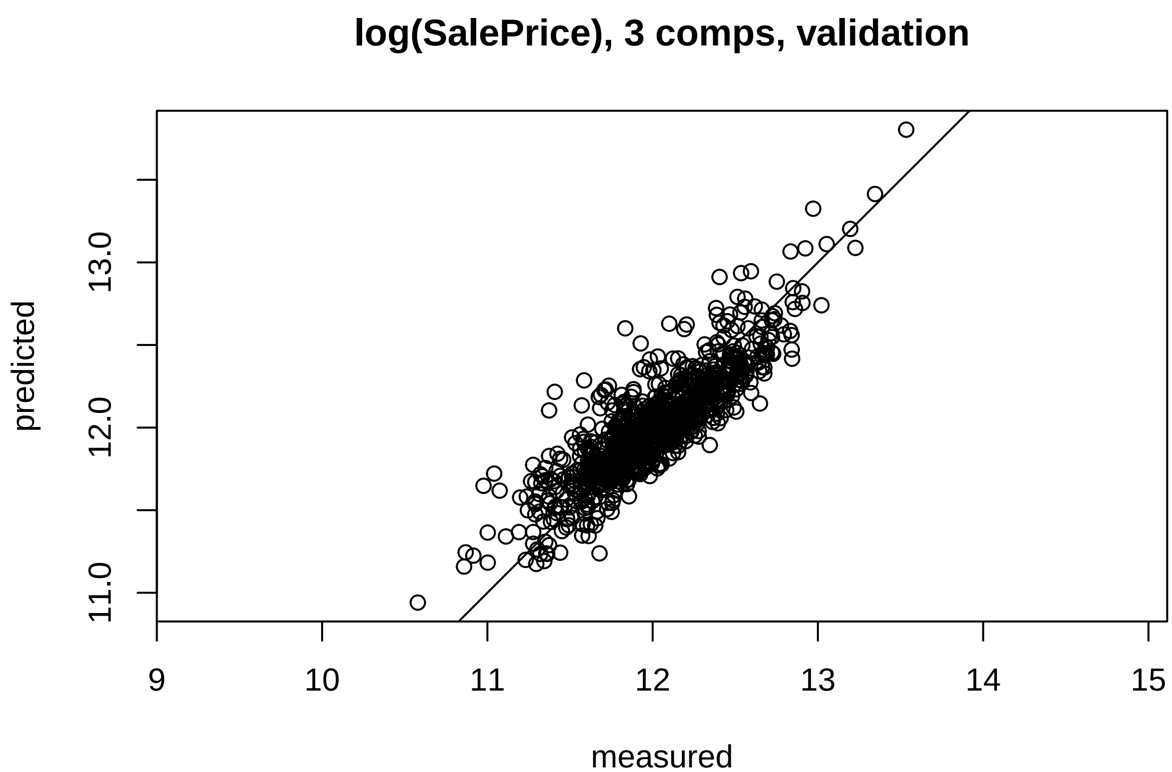

```r
plot(model.pls, ncomp = 1, asp = 1, line = TRUE)
```



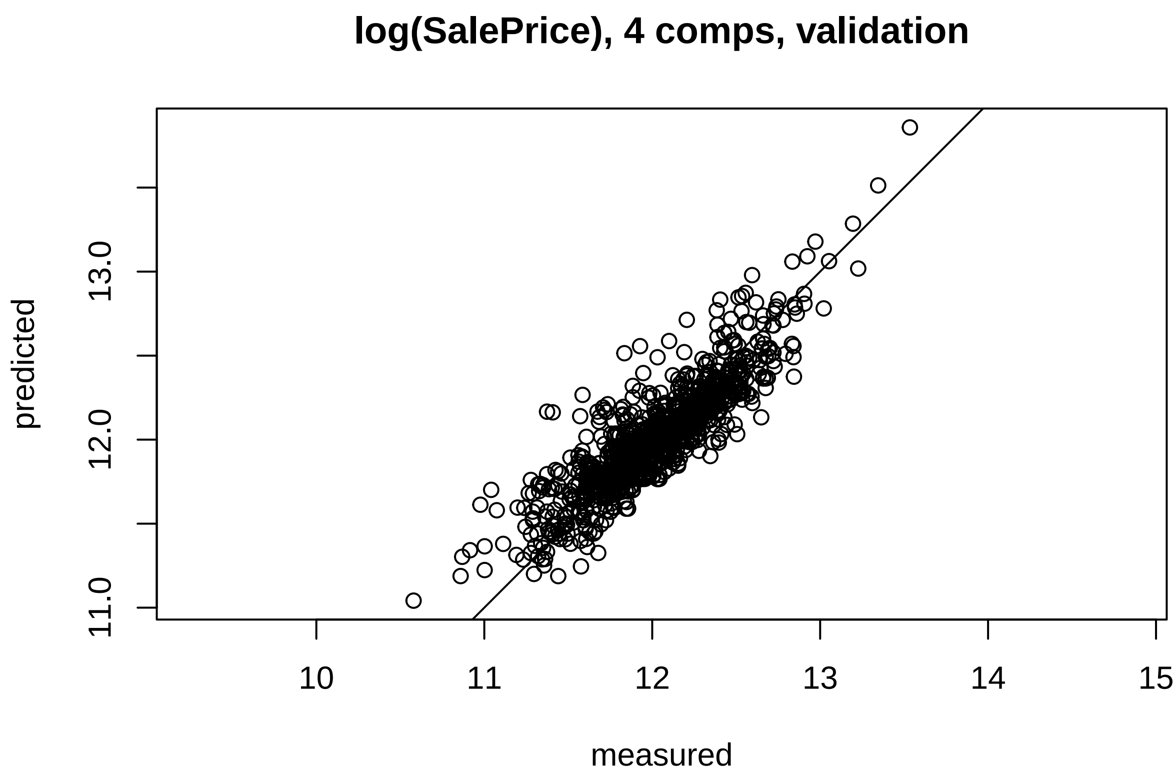

```r
plot(model.pls, ncomp = 2, asp = 1, line = TRUE)
```



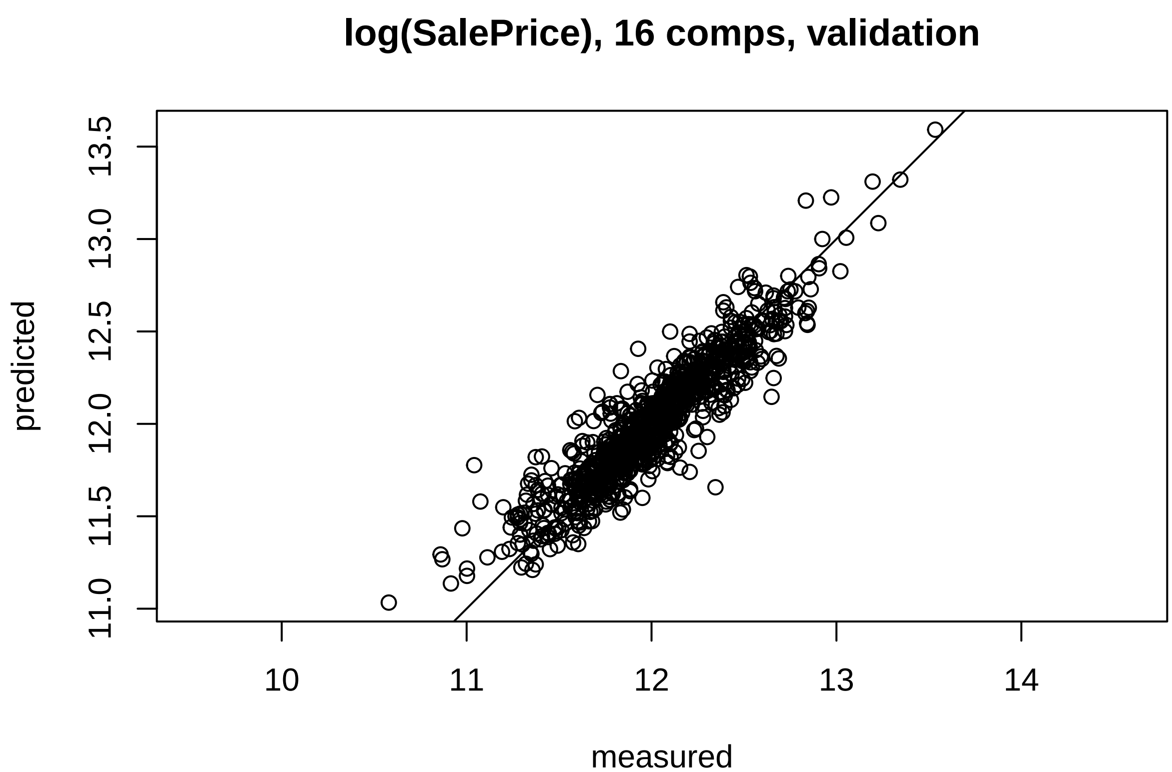

```r
plot(model.pls, ncomp = 3, asp = 1, line = TRUE)
```



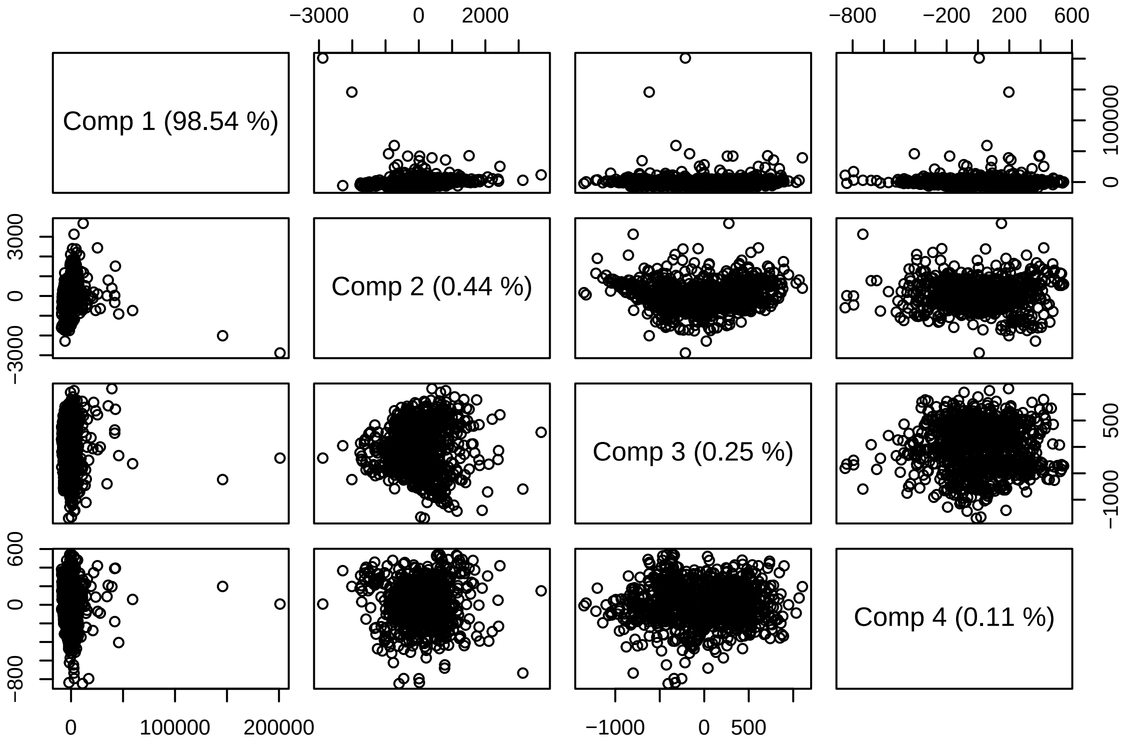

```r
plot(model.pls, ncomp = 4, asp = 1, line = TRUE)
```




```r
plot(model.pls, ncomp = 16, asp = 1, line = TRUE)
```

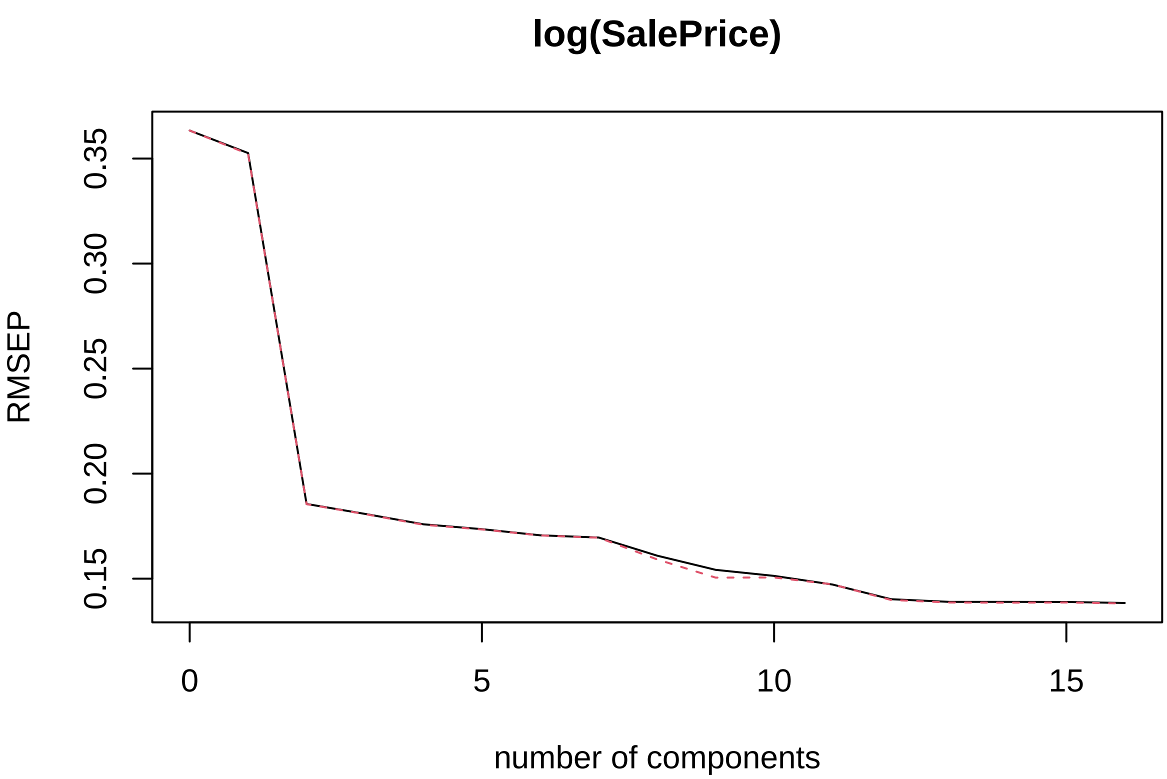



```r
plot(model.pls, plottype = "scores", comps = 1:4)
```



Based on the improvement in fit from 1 to 16 components and to avoid over fitting, 4 components seem to offer a good balance between predictive accuracy and model simplicity.

```
validationplot(model.pls, val.type = "RMSEP")
```




```
rmsep_4comps <- RMSEP(model.pls, ncomp = 4)$val[2]
cat("Cross-validated RMSE for 4 components:", rmsep_4comps, "\n")
```

```
## Cross-validated RMSE for 4 components: 0.3633341
```

Based on the RMSEP plot and the cross-validated RMSE value, 4 components seem optimal for our PLS model. Adding more components yields diminishing returns in terms of RMSE improvement. The RMSE for the model with 4 components is 0.3633341, which provides a reasonable estimate of the model's prediction error on unseen data.

1.c.

We kept having issues getting `caret` to work so decided to look at the basic package that `caret` was using. Quickly we recognized that the `lars` package requires only numerical data. Once we saw that, we built a lasso model!

```
x<- as.matrix(housingNumeric)
lars(x,housingData$logSalePrice,"lasso")
```

```
##
## Call:
## lars(x = x, y = housingData$logSalePrice, type = "lasso")
## R-squared: 0.955
## Sequence of LASSO moves:
##      SalePrice OverallQual GarageCars YearBuilt GrLivArea GarageYrBlt
## Var         36           5         27         7        17          26
## Step         1           2          3         4         5           6
##      YearRemodAdd GarageArea Fireplaces TotalBsmtSF OverallCond BedroomAbvGr
## Var             8         28         25          13           6           22
## Step            7          8          9          10          11           12
##      TotRmsAbvGrd HalfBath BsmtFinSF1 BsmtFullBath LotArea MasVnrArea FullBath
## Var            24       21         10           18       4          9       20
## Step           13       14         15           16      17         18       19
##      EncPorchSF OpenPorchSF MiscVal X1stFlrSF MSSubClass KitchenAbvGr YrSold
## Var          31          30      33        14          2           23     35
## Step         20          21      22        23         24           25     26
##      LowQualFinSF PoolArea Id MoSold BsmtHalfBath BsmtFinSF2 WoodDeckSF
## Var            16       32  1     34           19         11         29
## Step           27       28 29     30           31         32         33
##      LotFrontage
## Var            3
## Step          34
```

Next we want to tune the parameter to the best possible.

```
model <- caret::train(log(SalePrice) ~ .,
               data = housingNumeric,
               method = "lasso",
              # scale = TRUE,
               #trControl = fitControl#,
               tuneLength = 10
               )

model
```

```
## The lasso
##
## 1000 samples
##   35 predictor
```

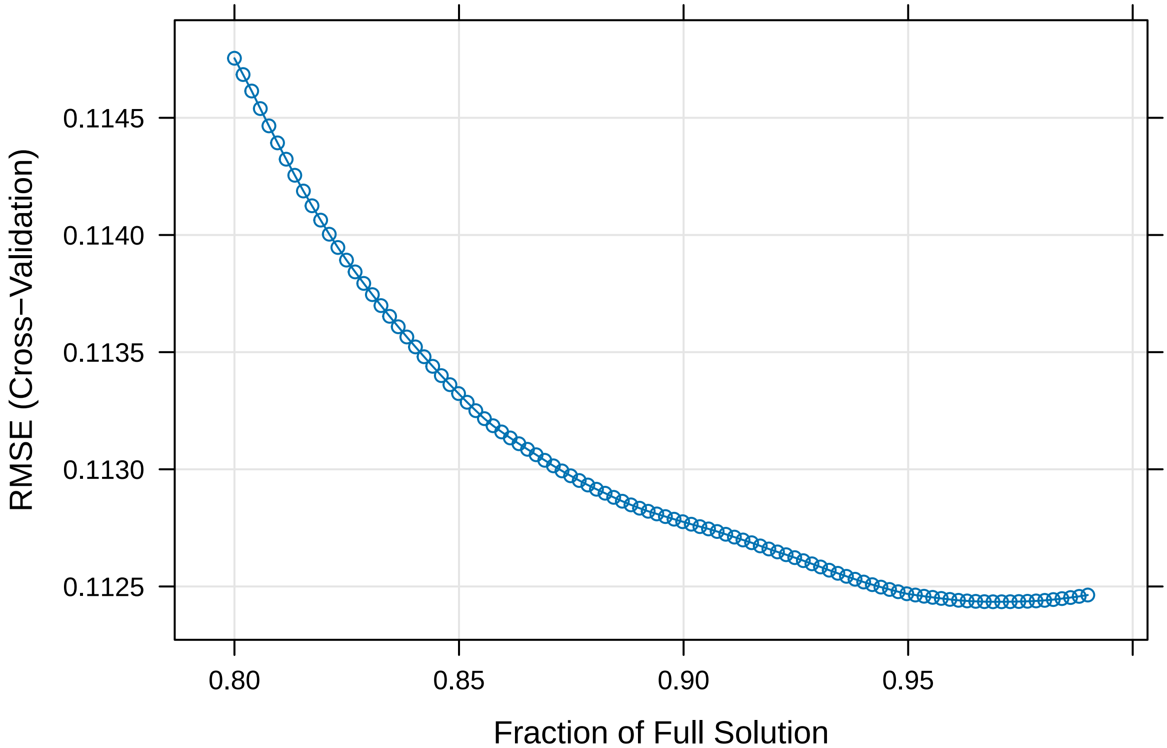
```
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 1000, 1000, 1000, 1000, 1000, 1000, ...
## Resampling results across tuning parameters:
##
##   fraction   RMSE       Rsquared   MAE
##   0.1000000  0.2686008  0.7256139  0.20761915
##   0.1888889  0.2258852  0.7823676  0.17185608
##   0.2777778  0.1931841  0.8161266  0.14523594
##   0.3666667  0.1671445  0.8422524  0.12416128
##   0.4555556  0.1465460  0.8642718  0.10769283
##   0.5444444  0.1306182  0.8820023  0.09554631
##   0.6333333  0.1203064  0.8942433  0.08795723
##   0.7222222  0.1144738  0.9010944  0.08389675
##   0.8111111  0.1119763  0.9039713  0.08205074
##   0.9000000  0.1115254  0.9044733  0.08173599
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was fraction = 0.9.
```

We see that 0.9 produced the best value, but let's keep tuning and see if we can do better. I'll apply a grid too.

```r
lassoGrid <- expand.grid(fraction=seq(0.8,0.99,length=100))

fitLasso <- train(log(SalePrice) ~ .,
            data = housingNumeric,
             method="lasso",
             trControl=fitControl,
             tuneGrid=lassoGrid)

plot(fitLasso)
```




```
fitLasso
```

```
## The lasso
##
## 1000 samples
##   35 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 801, 799, 799, 801, 800
## Resampling results across tuning parameters:
##
##   fraction   RMSE       Rsquared   MAE
##   0.8000000  0.1147543  0.9018477  0.08284779
##   0.8019192  0.1146846  0.9019338  0.08280581
##   0.8038384  0.1146143  0.9020219  0.08276412
##   0.8057576  0.1145394  0.9021198  0.08271527
##   0.8076768  0.1144657  0.9022161  0.08266676
##   0.8095960  0.1143931  0.9023108  0.08262031
##   0.8115152  0.1143234  0.9024002  0.08257287
##   0.8134343  0.1142549  0.9024878  0.08252549
##   0.8153535  0.1141876  0.9025738  0.08247874
##   0.8172727  0.1141247  0.9026526  0.08243408
##   0.8191919  0.1140634  0.9027291  0.08238996
##   0.8211111  0.1140033  0.9028041  0.08234637
##   0.8230303  0.1139467  0.9028735  0.08230309
##   0.8249495  0.1138928  0.9029391  0.08226063
##   0.8268687  0.1138425  0.9030004  0.08222159
##   0.8287879  0.1137933  0.9030602  0.08218338
##   0.8307071  0.1137454  0.9031189  0.08214616
```

```
## 0.8326263 0.1136986 0.9031764 0.08211052
## 0.8345455 0.1136529 0.9032326 0.08207488
## 0.8364646 0.1136083 0.9032874 0.08203924
## 0.8383838 0.1135647 0.9033407 0.08200360
## 0.8403030 0.1135223 0.9033928 0.08196850
## 0.8422222 0.1134804 0.9034450 0.08193636
## 0.8441414 0.1134396 0.9034959 0.08190515
## 0.8460606 0.1133998 0.9035455 0.08187482
## 0.8479798 0.1133612 0.9035938 0.08184448
## 0.8498990 0.1133232 0.9036422 0.08181522
## 0.8518182 0.1132862 0.9036901 0.08178678
## 0.8537374 0.1132503 0.9037366 0.08175851
## 0.8556566 0.1132162 0.9037814 0.08173137
## 0.8575758 0.1131860 0.9038200 0.08170646
## 0.8594949 0.1131595 0.9038524 0.08168396
## 0.8614141 0.1131338 0.9038837 0.08166373
## 0.8633333 0.1131091 0.9039138 0.08164471
## 0.8652525 0.1130853 0.9039427 0.08162718
## 0.8671717 0.1130621 0.9039711 0.08161121
## 0.8690909 0.1130384 0.9040006 0.08159775
## 0.8710101 0.1130148 0.9040301 0.08158534
## 0.8729293 0.1129931 0.9040572 0.08157502
## 0.8748485 0.1129723 0.9040832 0.08156537
## 0.8767677 0.1129522 0.9041083 0.08155623
## 0.8786869 0.1129329 0.9041324 0.08154729
## 0.8806061 0.1129149 0.9041547 0.08153905
## 0.8825253 0.1128979 0.9041756 0.08153193
## 0.8844444 0.1128803 0.9041982 0.08152232
## 0.8863636 0.1128637 0.9042193 0.08151263
## 0.8882828 0.1128482 0.9042390 0.08150314
## 0.8902020 0.1128340 0.9042568 0.08149290
## 0.8921212 0.1128208 0.9042731 0.08148255
## 0.8940404 0.1128096 0.9042865 0.08147336
## 0.8959596 0.1127983 0.9043004 0.08146729
## 0.8978788 0.1127870 0.9043145 0.08146401
## 0.8997980 0.1127763 0.9043279 0.08146187
## 0.9017172 0.1127657 0.9043415 0.08145940
## 0.9036364 0.1127554 0.9043548 0.08145677
## 0.9055556 0.1127456 0.9043675 0.08145414
## 0.9074747 0.1127346 0.9043825 0.08145236
## 0.9093939 0.1127228 0.9043991 0.08145032
## 0.9113131 0.1127106 0.9044164 0.08144728
## 0.9132323 0.1126984 0.9044339 0.08144328
## 0.9151515 0.1126866 0.9044508 0.08143922
## 0.9170707 0.1126731 0.9044706 0.08143340
## 0.9189899 0.1126600 0.9044898 0.08142827
## 0.9209091 0.1126475 0.9045083 0.08142376
## 0.9228283 0.1126354 0.9045261 0.08142012
## 0.9247475 0.1126230 0.9045445 0.08141711
## 0.9266667 0.1126100 0.9045640 0.08141380
## 0.9285859 0.1125967 0.9045843 0.08140907
## 0.9305051 0.1125832 0.9046052 0.08140426
## 0.9324242 0.1125694 0.9046267 0.08139921
## 0.9343434 0.1125561 0.9046476 0.08139416
```

```
##   0.9362626  0.1125432  0.9046679  0.08138910
##   0.9381818  0.1125308  0.9046876  0.08138419
##   0.9401010  0.1125188  0.9047066  0.08138082
##   0.9420202  0.1125077  0.9047244  0.08137897
##   0.9439394  0.1124972  0.9047412  0.08137856
##   0.9458586  0.1124872  0.9047574  0.08137815
##   0.9477778  0.1124775  0.9047731  0.08137824
##   0.9496970  0.1124689  0.9047874  0.08137916
##   0.9516162  0.1124634  0.9047968  0.08138308
##   0.9535354  0.1124583  0.9048057  0.08138709
##   0.9554545  0.1124535  0.9048143  0.08139131
##   0.9573737  0.1124489  0.9048226  0.08139598
##   0.9592929  0.1124447  0.9048305  0.08140136
##   0.9612121  0.1124408  0.9048378  0.08140708
##   0.9631313  0.1124382  0.9048435  0.08141396
##   0.9650505  0.1124364  0.9048483  0.08142141
##   0.9669697  0.1124352  0.9048521  0.08142967
##   0.9688889  0.1124347  0.9048548  0.08143859
##   0.9708081  0.1124347  0.9048569  0.08144763
##   0.9727273  0.1124351  0.9048586  0.08145664
##   0.9746465  0.1124357  0.9048598  0.08146535
##   0.9765657  0.1124368  0.9048605  0.08147405
##   0.9784848  0.1124382  0.9048607  0.08148284
##   0.9804040  0.1124406  0.9048593  0.08149235
##   0.9823232  0.1124441  0.9048564  0.08150246
##   0.9842424  0.1124481  0.9048526  0.08151295
##   0.9861616  0.1124527  0.9048481  0.08152342
##   0.9880808  0.1124577  0.9048431  0.08153483
##   0.9900000  0.1124634  0.9048369  0.08154689
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was fraction = 0.9708081.
```

```r
fitLasso$bestTune$fraction
```

```
## [1] 0.9708081
```

```r
# Best results
best_lasso <- fitLasso$results[which.min(fitLasso$results$RMSE), ]
best_lasso_RMSE <- best_lasso$RMSE
best_lasso_R2 <- best_lasso$Rsquared
```

```r
# Extract the coefficients at the optimal fraction using predict
coefficients <- predict(fitLasso$finalModel,
                        s = fitLasso$bestTune$fraction,
                        type = "coefficients",
                        mode = "fraction")$coefficients

# Display the coefficients
print(coefficients)
```

```
##            Id    MSSubClass   LotFrontage       LotArea   OverallQual
## -1.407081e-06 -4.507394e-04  1.482327e-04  2.880781e-06  6.767333e-02
##   OverallCond     YearBuilt  YearRemodAdd    MasVnrArea    BsmtFinSF1
##  5.321804e-02  3.144216e-03  6.829716e-04  1.267580e-06  7.817872e-05
##    BsmtFinSF2   TotalBsmtSF     X1stFlrSF     X2ndFlrSF  LowQualFinSF
```

```
##  1.909535e-05  1.106663e-04  1.916463e-04  2.030294e-04  9.338703e-05
##     GrLivArea  BsmtFullBath  BsmtHalfBath      FullBath      HalfBath
##  6.857706e-05  2.888177e-02  6.393924e-03  1.424368e-02  1.049233e-02
##  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd    Fireplaces   GarageYrBlt
## -1.720893e-02 -4.463756e-02  7.343519e-03  3.934151e-02  1.702725e-05
##    GarageCars    GarageArea    WoodDeckSF   OpenPorchSF   EncPorchSF
##  2.839110e-02  9.742194e-05  5.638588e-05  9.784154e-05  1.992938e-04
##      PoolArea       MiscVal        MoSold        YrSold
##  1.130887e-04 -2.955749e-05 -5.351916e-04 -3.335539e-03
```

```r
varImp(fitLasso, scale = FALSE)
```

```
## loess r-squared variable importance
##
##   only 20 most important variables shown (out of 35)
##
##              Overall
## OverallQual  0.61111
## GrLivArea    0.56598
## TotalBsmtSF  0.42103
## GarageArea   0.39039
## GarageCars   0.38716
## X1stFlrSF    0.37257
## FullBath     0.30775
## X2ndFlrSF    0.29541
## TotRmsAbvGrd 0.27074
## BsmtFinSF1   0.26439
## YearBuilt    0.25382
## Fireplaces   0.22453
## YearRemodAdd 0.21322
## MasVnrArea   0.20824
## LotArea      0.17093
## WoodDeckSF   0.13848
## HalfBath     0.09576
## OpenPorchSF  0.09054
## BsmtUnfSF    0.07511
## GarageYrBlt  0.07111
```

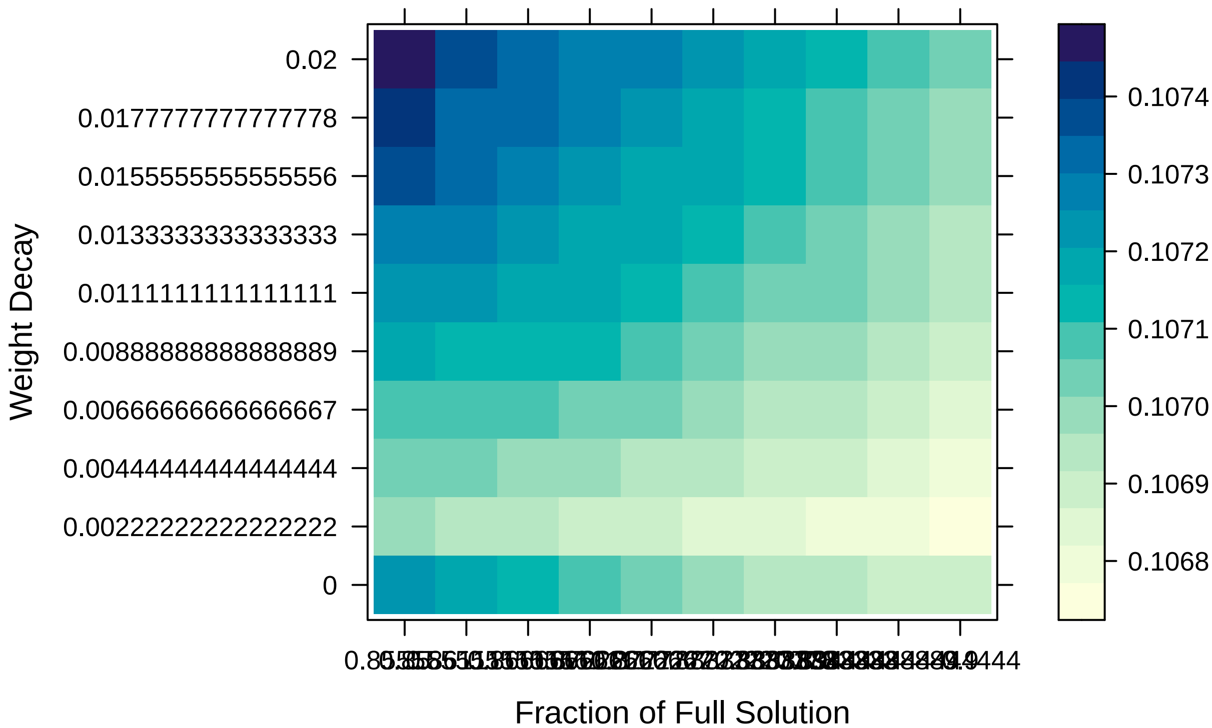
1.d.

Here is the elasticNet. I played with the parameter grid for a bit too long.

```r
enetGrid <- expand.grid(lambda=seq(0,0.02,length=10),
                        fraction=seq(.85,.90,length=10))

fitEnet <- train(log(SalePrice) ~ .,
            data = housingNumeric,
             method="enet",
             trControl=fitControl,
             tuneGrid=enetGrid)

plot(fitEnet, plotType="level")
```



RMSE (Cross−Validation)

```
fitEnet
```

```
## Elasticnet
##
## 1000 samples
##   35 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 799, 800, 800, 801, 800
## Resampling results across tuning parameters:
##
##   lambda       fraction   RMSE       Rsquared   MAE
##   0.000000000  0.8500000  0.1072294  0.9138060  0.07983730
##   0.000000000  0.8555556  0.1071702  0.9138421  0.07976315
##   0.000000000  0.8611111  0.1071158  0.9138721  0.07968906
##   0.000000000  0.8666667  0.1070664  0.9138947  0.07961529
##   0.000000000  0.8722222  0.1070236  0.9139089  0.07954713
##   0.000000000  0.8777778  0.1069865  0.9139174  0.07948248
##   0.000000000  0.8833333  0.1069532  0.9139223  0.07941711
##   0.000000000  0.8888889  0.1069254  0.9139194  0.07935425
##   0.000000000  0.8944444  0.1069095  0.9138987  0.07930353
##   0.000000000  0.9000000  0.1069034  0.9138633  0.07926388
##   0.002222222  0.8500000  0.1069908  0.9137445  0.07924653
##   0.002222222  0.8555556  0.1069630  0.9137517  0.07920093
##   0.002222222  0.8611111  0.1069375  0.9137596  0.07915571
##   0.002222222  0.8666667  0.1069063  0.9137809  0.07910378
##   0.002222222  0.8722222  0.1068820  0.9137923  0.07905481
##   0.002222222  0.8777778  0.1068610  0.9137995  0.07901124
```

```
## 0.002222222 0.8833333 0.1068429 0.9138033 0.07896992
## 0.002222222 0.8888889 0.1068189 0.9138112 0.07892788
## 0.002222222 0.8944444 0.1067980 0.9138147 0.07889366
## 0.002222222 0.9000000 0.1067711 0.9138314 0.07887046
## 0.004444444 0.8500000 0.1070452 0.9136208 0.07927759
## 0.004444444 0.8555556 0.1070212 0.9136186 0.07922986
## 0.004444444 0.8611111 0.1069995 0.9136216 0.07918455
## 0.004444444 0.8666667 0.1069741 0.9136351 0.07913460
## 0.004444444 0.8722222 0.1069505 0.9136476 0.07908483
## 0.004444444 0.8777778 0.1069302 0.9136562 0.07904042
## 0.004444444 0.8833333 0.1068997 0.9136758 0.07899362
## 0.004444444 0.8888889 0.1068737 0.9136890 0.07895580
## 0.004444444 0.8944444 0.1068458 0.9137072 0.07892450
## 0.004444444 0.9000000 0.1068107 0.9137426 0.07889822
## 0.006666667 0.8500000 0.1071022 0.9134978 0.07931100
## 0.006666667 0.8555556 0.1070822 0.9134845 0.07926132
## 0.006666667 0.8611111 0.1070653 0.9134804 0.07921714
## 0.006666667 0.8666667 0.1070437 0.9134897 0.07916387
## 0.006666667 0.8722222 0.1070227 0.9135001 0.07911522
## 0.006666667 0.8777778 0.1069912 0.9135233 0.07906589
## 0.006666667 0.8833333 0.1069601 0.9135462 0.07902330
## 0.006666667 0.8888889 0.1069268 0.9135738 0.07898711
## 0.006666667 0.8944444 0.1068844 0.9136209 0.07895328
## 0.006666667 0.9000000 0.1068488 0.9136623 0.07893790
## 0.008888889 0.8500000 0.1071647 0.9133701 0.07934427
## 0.008888889 0.8555556 0.1071442 0.9133587 0.07929457
## 0.008888889 0.8611111 0.1071348 0.9133388 0.07924987
## 0.008888889 0.8666667 0.1071163 0.9133455 0.07919972
## 0.008888889 0.8722222 0.1070905 0.9133608 0.07914802
## 0.008888889 0.8777778 0.1070549 0.9133898 0.07909941
## 0.008888889 0.8833333 0.1070120 0.9134398 0.07906943
## 0.008888889 0.8888889 0.1069740 0.9134834 0.07904383
## 0.008888889 0.8944444 0.1069346 0.9135334 0.07901151
## 0.008888889 0.9000000 0.1068980 0.9135793 0.07898029
## 0.011111111 0.8500000 0.1072316 0.9132387 0.07937693
## 0.011111111 0.8555556 0.1072108 0.9132296 0.07932810
## 0.011111111 0.8611111 0.1072005 0.9132045 0.07928008
## 0.011111111 0.8666667 0.1071703 0.9132312 0.07923619
## 0.011111111 0.8722222 0.1071303 0.9132763 0.07919150
## 0.011111111 0.8777778 0.1070971 0.9133118 0.07915199
## 0.011111111 0.8833333 0.1070596 0.9133550 0.07911579
## 0.011111111 0.8888889 0.1070174 0.9134073 0.07907956
## 0.011111111 0.8944444 0.1069723 0.9134666 0.07904226
## 0.011111111 0.9000000 0.1069275 0.9135264 0.07900335
## 0.013333333 0.8500000 0.1072966 0.9131069 0.07940931
## 0.013333333 0.8555556 0.1072657 0.9131108 0.07935684
## 0.013333333 0.8611111 0.1072310 0.9131439 0.07930851
## 0.013333333 0.8666667 0.1072018 0.9131693 0.07926760
## 0.013333333 0.8722222 0.1071642 0.9132125 0.07922537
## 0.013333333 0.8777778 0.1071273 0.9132543 0.07918400
## 0.013333333 0.8833333 0.1070944 0.9132901 0.07914975
## 0.013333333 0.8888889 0.1070508 0.9133456 0.07911186
## 0.013333333 0.8944444 0.1069997 0.9134156 0.07906854
## 0.013333333 0.9000000 0.1069509 0.9134830 0.07902810
```

```
##   0.015555556  0.8500000  0.1073512  0.9129955  0.07945036
##   0.015555556  0.8555556  0.1073050  0.9130445  0.07939301
##   0.015555556  0.8611111  0.1072664  0.9130829  0.07934277
##   0.015555556  0.8666667  0.1072327  0.9131160  0.07929913
##   0.015555556  0.8722222  0.1071973  0.9131495  0.07925898
##   0.015555556  0.8777778  0.1071588  0.9131935  0.07921711
##   0.015555556  0.8833333  0.1071207  0.9132392  0.07917813
##   0.015555556  0.8888889  0.1070764  0.9132978  0.07914066
##   0.015555556  0.8944444  0.1070265  0.9133663  0.07909919
##   0.015555556  0.9000000  0.1069761  0.9134372  0.07905682
##   0.017777778  0.8500000  0.1073991  0.9129135  0.07949410
##   0.017777778  0.8555556  0.1073443  0.9129772  0.07943082
##   0.017777778  0.8611111  0.1073027  0.9130206  0.07937874
##   0.017777778  0.8666667  0.1072676  0.9130555  0.07933240
##   0.017777778  0.8722222  0.1072293  0.9130936  0.07929135
##   0.017777778  0.8777778  0.1071894  0.9131356  0.07924765
##   0.017777778  0.8833333  0.1071428  0.9131972  0.07920642
##   0.017777778  0.8888889  0.1070985  0.9132567  0.07916801
##   0.017777778  0.8944444  0.1070523  0.9133201  0.07912925
##   0.017777778  0.9000000  0.1070032  0.9133895  0.07908690
##   0.020000000  0.8500000  0.1074461  0.9128286  0.07953870
##   0.020000000  0.8555556  0.1073891  0.9128981  0.07947801
##   0.020000000  0.8611111  0.1073405  0.9129541  0.07941966
##   0.020000000  0.8666667  0.1073009  0.9129965  0.07937009
##   0.020000000  0.8722222  0.1072545  0.9130477  0.07931732
##   0.020000000  0.8777778  0.1072087  0.9131003  0.07927465
##   0.020000000  0.8833333  0.1071647  0.9131549  0.07923735
##   0.020000000  0.8888889  0.1071194  0.9132173  0.07919692
##   0.020000000  0.8944444  0.1070781  0.9132739  0.07916101
##   0.020000000  0.9000000  0.1070302  0.9133427  0.07911917
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were fraction = 0.9 and lambda
##  = 0.002222222.
```

```r
# Best tuning parameters
best_enet_tune <- fitEnet$bestTune
print(best_enet_tune)
```

```
##    fraction      lambda
## 20      0.9 0.002222222
```

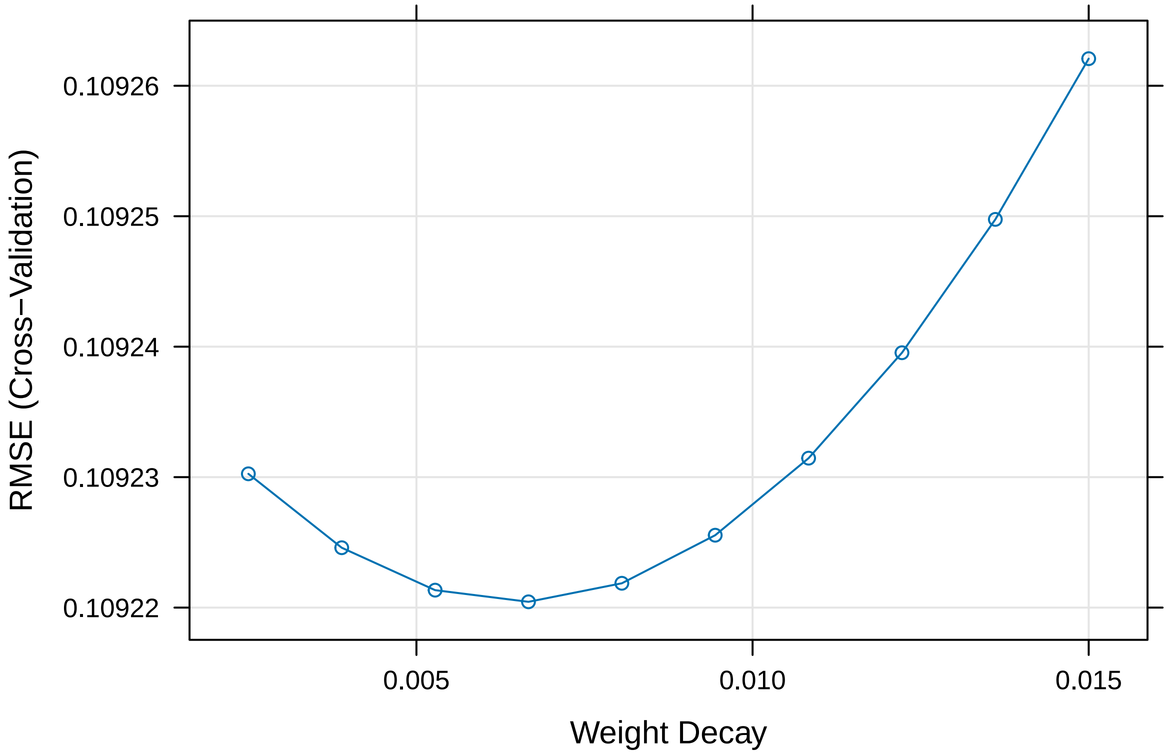
```r
# Best results
best_enet <- fitEnet$results[which.min(fitEnet$results$RMSE), ]
best_enet_RMSE <- best_enet$RMSE
best_enet_R2 <- best_enet$Rsquared
```

Again, lots of tuning but a nice value for the ridge regression and it's ℓ^2 penalty.

```r
ridgeGrid <- expand.grid(lambda=seq(0.0025,0.015,length=10))

fitRidge <- train(log(SalePrice) ~ .,
            data = housingNumeric,
             method="ridge",
             trControl=fitControl,
             tuneGrid=ridgeGrid)
```

```r
plot(fitRidge)
```




```r
fitRidge
```

```
## Ridge Regression
##
## 1000 samples
##   35 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 801, 800, 800, 800, 799
## Resampling results across tuning parameters:
##
##   lambda       RMSE       Rsquared   MAE
##   0.002500000  0.1092303  0.9097312  0.08055413
##   0.003888889  0.1092246  0.9097575  0.08055074
##   0.005277778  0.1092213  0.9097815  0.08055003
##   0.006666667  0.1092204  0.9098033  0.08055099
##   0.008055556  0.1092219  0.9098230  0.08055331
##   0.009444444  0.1092255  0.9098408  0.08055729
##   0.010833333  0.1092315  0.9098566  0.08056304
##   0.012222222  0.1092395  0.9098706  0.08056972
##   0.013611111  0.1092498  0.9098828  0.08057821
##   0.015000000  0.1092621  0.9098934  0.08058720
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was lambda = 0.006666667.
```

```r
# Best tuning parameters
best_ridge_tune <- fitRidge$bestTune
print(best_ridge_tune)
```

```
##        lambda
## 4 0.006666667
```

```r
# Best results
best_ridge <- fitRidge$results[which.min(fitRidge$results$RMSE), ]
best_ridge_RMSE <- best_ridge$RMSE
best_ridge_R2 <- best_ridge$Rsquared
```

We want to try the reduction technique with a full linear regression that includes limited two way interactions.

```r
# Suppose we have identified key variables for interactions
key_vars <- c("OverallQual", "GrLivArea", "GarageCars", "TotalBsmtSF")

# Create a formula with limited interactions
interaction_formula <- as.formula(paste("log(SalePrice) ~ . + (", paste(key_vars, collapse = "+"), ")^2

# Fit the model
fitOLMwLimited <- lm(interaction_formula, data = housingNumeric)

# Perform stepwise selection
OLMSim2Way <- stepAIC(fitOLMwLimited, direction = "both")
```

```
## Start:  AIC=-4511.23
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF + X1stFlrSF + X2ndFlrSF +
##     LowQualFinSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd +
##     Fireplaces + GarageYrBlt + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EncPorchSF + PoolArea + MiscVal + MoSold +
##     YrSold + (OverallQual + GrLivArea + GarageCars + TotalBsmtSF)^2
##
##
## Step:  AIC=-4511.23
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF + X1stFlrSF + X2ndFlrSF +
##     GrLivArea + BsmtFullBath + BsmtHalfBath + FullBath + HalfBath +
##     BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd + Fireplaces +
##     GarageYrBlt + GarageCars + GarageArea + WoodDeckSF + OpenPorchSF +
##     EncPorchSF + PoolArea + MiscVal + MoSold + YrSold + OverallQual:GrLivArea +
##     OverallQual:GarageCars + OverallQual:TotalBsmtSF + GrLivArea:GarageCars +
##     TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##
## Step:  AIC=-4511.23
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + FullBath + HalfBath + BedroomAbvGr +
##     KitchenAbvGr + TotRmsAbvGrd + Fireplaces + GarageYrBlt +
```

```
##     GarageCars + GarageArea + WoodDeckSF + OpenPorchSF + EncPorchSF +
##     PoolArea + MiscVal + MoSold + YrSold + OverallQual:GrLivArea +
##     OverallQual:GarageCars + OverallQual:TotalBsmtSF + GrLivArea:GarageCars +
##     TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                         Df Sum of Sq    RSS     AIC
## - GarageYrBlt            1   0.00016 10.140 -4513.2
## - OverallQual:GrLivArea  1   0.00028 10.141 -4513.2
## - TotRmsAbvGrd           1   0.00358 10.144 -4512.9
## - BsmtFinSF2             1   0.00494 10.145 -4512.7
## - FullBath               1   0.00515 10.146 -4512.7
## - HalfBath               1   0.00562 10.146 -4512.7
## - MoSold                 1   0.00647 10.147 -4512.6
## - MasVnrArea             1   0.00682 10.147 -4512.6
## - Id                     1   0.00691 10.147 -4512.6
## - BsmtHalfBath           1   0.00849 10.149 -4512.4
## - PoolArea               1   0.01019 10.150 -4512.2
## - OverallQual:GarageCars 1   0.01125 10.152 -4512.1
## - LotFrontage            1   0.01987 10.160 -4511.3
## <none>                               10.140 -4511.2
## - MiscVal                1   0.02338 10.164 -4510.9
## - YrSold                 1   0.02358 10.164 -4510.9
## - OpenPorchSF            1   0.02458 10.165 -4510.8
## - GrLivArea:GarageCars   1   0.03839 10.179 -4509.5
## - X1stFlrSF              1   0.03946 10.180 -4509.3
## - WoodDeckSF             1   0.04468 10.185 -4508.8
## - BedroomAbvGr           1   0.05794 10.198 -4507.5
## - YearRemodAdd           1   0.06075 10.201 -4507.3
## - X2ndFlrSF              1   0.06742 10.208 -4506.6
## - GarageArea             1   0.06766 10.208 -4506.6
## - BsmtFullBath           1   0.08545 10.226 -4504.8
## - KitchenAbvGr           1   0.09397 10.234 -4504.0
## - TotalBsmtSF:GarageCars 1   0.12332 10.264 -4501.1
## - EncPorchSF             1   0.19466 10.335 -4494.2
## - MSSubClass             1   0.20723 10.348 -4493.0
## - TotalBsmtSF:GrLivArea  1   0.21115 10.351 -4492.6
## - OverallQual:TotalBsmtSF 1  0.26415 10.405 -4487.5
## - Fireplaces             1   0.31955 10.460 -4482.2
## - BsmtFinSF1             1   0.40620 10.547 -4474.0
## - LotArea                1   0.72540 10.866 -4444.1
## - YearBuilt              1   2.16376 12.304 -4319.8
## - OverallCond            1   2.33524 12.476 -4306.0
##
## Step:  AIC=-4513.22
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + FullBath + HalfBath + BedroomAbvGr +
##     KitchenAbvGr + TotRmsAbvGrd + Fireplaces + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + MiscVal +
##     MoSold + YrSold + OverallQual:GrLivArea + OverallQual:GarageCars +
##     OverallQual:TotalBsmtSF + GrLivArea:GarageCars + TotalBsmtSF:GrLivArea +
##     TotalBsmtSF:GarageCars
##
```

```
##                           Df Sum of Sq    RSS     AIC
## - OverallQual:GrLivArea    1   0.00029 10.141 -4515.2
## - TotRmsAbvGrd             1   0.00361 10.144 -4514.9
## - FullBath                 1   0.00501 10.146 -4514.7
## - BsmtFinSF2               1   0.00505 10.146 -4514.7
## - HalfBath                 1   0.00549 10.146 -4514.7
## - MoSold                   1   0.00637 10.147 -4514.6
## - MasVnrArea               1   0.00685 10.147 -4514.5
## - Id                       1   0.00690 10.147 -4514.5
## - BsmtHalfBath             1   0.00854 10.149 -4514.4
## - PoolArea                 1   0.01023 10.151 -4514.2
## - OverallQual:GarageCars   1   0.01136 10.152 -4514.1
## - LotFrontage              1   0.01978 10.160 -4513.3
## <none>                                 10.140 -4513.2
## - MiscVal                  1   0.02336 10.164 -4512.9
## - YrSold                   1   0.02349 10.164 -4512.9
## - OpenPorchSF              1   0.02457 10.165 -4512.8
## - GrLivArea:GarageCars     1   0.03910 10.180 -4511.4
## - X1stFlrSF                1   0.04006 10.181 -4511.3
## + GarageYrBlt              1   0.00016 10.140 -4511.2
## - WoodDeckSF               1   0.04466 10.185 -4510.8
## - BedroomAbvGr             1   0.05803 10.198 -4509.5
## - YearRemodAdd             1   0.06064 10.201 -4509.3
## - X2ndFlrSF                1   0.06885 10.209 -4508.4
## - GarageArea               1   0.07080 10.211 -4508.3
## - BsmtFullBath             1   0.08538 10.226 -4506.8
## - KitchenAbvGr             1   0.09855 10.239 -4505.5
## - TotalBsmtSF:GarageCars   1   0.13011 10.271 -4502.5
## - EncPorchSF               1   0.19451 10.335 -4496.2
## - MSSubClass               1   0.20789 10.348 -4494.9
## - TotalBsmtSF:GrLivArea    1   0.21162 10.352 -4494.6
## - OverallQual:TotalBsmtSF  1   0.27199 10.412 -4488.7
## - Fireplaces               1   0.31970 10.460 -4484.2
## - BsmtFinSF1               1   0.40725 10.548 -4475.8
## - LotArea                  1   0.72560 10.866 -4446.1
## - YearBuilt                1   2.16658 12.307 -4321.6
## - OverallCond              1   2.33902 12.479 -4307.7
##
## Step:  AIC=-4515.19
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + FullBath + HalfBath + BedroomAbvGr +
##     KitchenAbvGr + TotRmsAbvGrd + Fireplaces + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + MiscVal +
##     MoSold + YrSold + OverallQual:GarageCars + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - TotRmsAbvGrd             1   0.00349 10.144 -4516.8
## - FullBath                 1   0.00484 10.146 -4516.7
## - BsmtFinSF2               1   0.00502 10.146 -4516.7
## - HalfBath                 1   0.00539 10.146 -4516.7
## - MoSold                   1   0.00648 10.147 -4516.5
```

```
## - Id                       1   0.00701 10.148 -4516.5
## - MasVnrArea               1   0.00716 10.148 -4516.5
## - BsmtHalfBath             1   0.00841 10.149 -4516.4
## - PoolArea                 1   0.01017 10.151 -4516.2
## - OverallQual:GarageCars   1   0.01355 10.154 -4515.9
## - LotFrontage              1   0.02020 10.161 -4515.2
## <none>                                 10.141 -4515.2
## - YrSold                   1   0.02357 10.164 -4514.9
## - MiscVal                  1   0.02358 10.164 -4514.9
## - OpenPorchSF              1   0.02435 10.165 -4514.8
## + OverallQual:GrLivArea    1   0.00029 10.140 -4513.2
## - X1stFlrSF                1   0.04047 10.181 -4513.2
## + GarageYrBlt              1   0.00017 10.141 -4513.2
## - GrLivArea:GarageCars     1   0.04288 10.184 -4513.0
## - WoodDeckSF               1   0.04477 10.186 -4512.8
## - BedroomAbvGr             1   0.05817 10.199 -4511.5
## - YearRemodAdd             1   0.06042 10.201 -4511.2
## - X2ndFlrSF                1   0.07013 10.211 -4510.3
## - GarageArea               1   0.07165 10.213 -4510.1
## - BsmtFullBath             1   0.08521 10.226 -4508.8
## - KitchenAbvGr             1   0.10148 10.242 -4507.2
## - TotalBsmtSF:GarageCars   1   0.14312 10.284 -4503.2
## - EncPorchSF               1   0.19444 10.335 -4498.2
## - MSSubClass               1   0.21205 10.353 -4496.5
## - TotalBsmtSF:GrLivArea    1   0.25159 10.392 -4492.7
## - OverallQual:TotalBsmtSF  1   0.29279 10.434 -4488.7
## - Fireplaces               1   0.31985 10.461 -4486.1
## - BsmtFinSF1               1   0.40807 10.549 -4477.7
## - LotArea                  1   0.72540 10.866 -4448.1
## - YearBuilt                1   2.17470 12.316 -4322.9
## - OverallCond              1   2.34645 12.487 -4309.0
##
## Step:  AIC=-4516.84
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + FullBath + HalfBath + BedroomAbvGr +
##     KitchenAbvGr + Fireplaces + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EncPorchSF + PoolArea + MiscVal + MoSold +
##     YrSold + OverallQual:GarageCars + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - FullBath                 1   0.00497 10.149 -4518.4
## - BsmtFinSF2               1   0.00507 10.149 -4518.3
## - HalfBath                 1   0.00527 10.150 -4518.3
## - MasVnrArea               1   0.00678 10.151 -4518.2
## - MoSold                   1   0.00705 10.151 -4518.1
## - Id                       1   0.00720 10.152 -4518.1
## - BsmtHalfBath             1   0.00834 10.153 -4518.0
## - PoolArea                 1   0.00980 10.154 -4517.9
## - OverallQual:GarageCars   1   0.01490 10.159 -4517.4
## <none>                                 10.144 -4516.8
## - LotFrontage              1   0.02082 10.165 -4516.8
```

```
## - YrSold                   1   0.02360 10.168 -4516.5
## - MiscVal                  1   0.02383 10.168 -4516.5
## - OpenPorchSF              1   0.02508 10.169 -4516.4
## + TotRmsAbvGrd             1   0.00349 10.141 -4515.2
## - X1stFlrSF                1   0.03939 10.184 -4515.0
## + GarageYrBlt              1   0.00020 10.144 -4514.9
## + OverallQual:GrLivArea    1   0.00017 10.144 -4514.9
## - GrLivArea:GarageCars     1   0.04406 10.188 -4514.5
## - WoodDeckSF               1   0.04564 10.190 -4514.4
## - BedroomAbvGr             1   0.05770 10.202 -4513.2
## - YearRemodAdd             1   0.06145 10.206 -4512.8
## - X2ndFlrSF                1   0.06874 10.213 -4512.1
## - GarageArea               1   0.07035 10.215 -4511.9
## - BsmtFullBath             1   0.08561 10.230 -4510.4
## - KitchenAbvGr             1   0.09818 10.242 -4509.2
## - TotalBsmtSF:GarageCars   1   0.14431 10.289 -4504.7
## - EncPorchSF               1   0.19166 10.336 -4500.1
## - MSSubClass               1   0.21671 10.361 -4497.7
## - TotalBsmtSF:GrLivArea    1   0.26253 10.407 -4493.3
## - OverallQual:TotalBsmtSF  1   0.29737 10.442 -4490.0
## - Fireplaces               1   0.32171 10.466 -4487.6
## - BsmtFinSF1               1   0.40488 10.549 -4479.7
## - LotArea                  1   0.72469 10.869 -4449.8
## - YearBuilt                1   2.17293 12.317 -4324.8
## - OverallCond              1   2.34943 12.494 -4310.5
##
## Step:  AIC=-4518.35
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + HalfBath + BedroomAbvGr + KitchenAbvGr +
##     Fireplaces + GarageCars + GarageArea + WoodDeckSF + OpenPorchSF +
##     EncPorchSF + PoolArea + MiscVal + MoSold + YrSold + OverallQual:GarageCars +
##     OverallQual:TotalBsmtSF + GrLivArea:GarageCars + TotalBsmtSF:GrLivArea +
##     TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - HalfBath                 1   0.00260 10.152 -4520.1
## - BsmtFinSF2               1   0.00520 10.155 -4519.8
## - MasVnrArea               1   0.00628 10.155 -4519.7
## - BsmtHalfBath             1   0.00726 10.156 -4519.6
## - Id                       1   0.00729 10.157 -4519.6
## - MoSold                   1   0.00734 10.157 -4519.6
## - PoolArea                 1   0.00940 10.159 -4519.4
## - OverallQual:GarageCars   1   0.01499 10.164 -4518.9
## <none>                                 10.149 -4518.4
## - LotFrontage              1   0.02070 10.170 -4518.3
## - YrSold                   1   0.02343 10.173 -4518.0
## - MiscVal                  1   0.02360 10.173 -4518.0
## - OpenPorchSF              1   0.02654 10.176 -4517.7
## + FullBath                 1   0.00497 10.144 -4516.8
## + TotRmsAbvGrd             1   0.00361 10.146 -4516.7
## - X1stFlrSF                1   0.03934 10.189 -4516.5
## + OverallQual:GrLivArea    1   0.00005 10.149 -4516.4
```

```
## + GarageYrBlt              1   0.00002 10.149 -4516.4
## - WoodDeckSF               1   0.04549 10.195 -4515.9
## - GrLivArea:GarageCars     1   0.04559 10.195 -4515.9
## - BedroomAbvGr             1   0.05472 10.204 -4515.0
## - GarageArea               1   0.06745 10.217 -4513.7
## - YearRemodAdd             1   0.06808 10.217 -4513.7
## - X2ndFlrSF                1   0.06975 10.219 -4513.5
## - BsmtFullBath             1   0.08072 10.230 -4512.4
## - KitchenAbvGr             1   0.09361 10.243 -4511.2
## - TotalBsmtSF:GarageCars   1   0.14655 10.296 -4506.0
## - EncPorchSF               1   0.19238 10.342 -4501.6
## - MSSubClass               1   0.21302 10.362 -4499.6
## - TotalBsmtSF:GrLivArea    1   0.26360 10.413 -4494.7
## - OverallQual:TotalBsmtSF  1   0.30100 10.450 -4491.1
## - Fireplaces               1   0.32097 10.470 -4489.2
## - BsmtFinSF1               1   0.40999 10.559 -4480.8
## - LotArea                  1   0.72820 10.877 -4451.1
## - OverallCond              1   2.34523 12.495 -4312.5
## - YearBuilt                1   2.51880 12.668 -4298.7
##
## Step:  AIC=-4520.1
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     BsmtFinSF2 + TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea +
##     BsmtFullBath + BsmtHalfBath + BedroomAbvGr + KitchenAbvGr +
##     Fireplaces + GarageCars + GarageArea + WoodDeckSF + OpenPorchSF +
##     EncPorchSF + PoolArea + MiscVal + MoSold + YrSold + OverallQual:GarageCars +
##     OverallQual:TotalBsmtSF + GrLivArea:GarageCars + TotalBsmtSF:GrLivArea +
##     TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - BsmtFinSF2               1   0.00568 10.158 -4521.5
## - MasVnrArea               1   0.00664 10.159 -4521.4
## - BsmtHalfBath             1   0.00708 10.159 -4521.4
## - Id                       1   0.00724 10.159 -4521.4
## - MoSold                   1   0.00751 10.159 -4521.4
## - PoolArea                 1   0.00984 10.162 -4521.1
## - OverallQual:GarageCars   1   0.01482 10.167 -4520.6
## <none>                                 10.152 -4520.1
## - LotFrontage              1   0.02039 10.172 -4520.1
## - YrSold                   1   0.02286 10.175 -4519.8
## - MiscVal                  1   0.02465 10.177 -4519.7
## - OpenPorchSF              1   0.02788 10.180 -4519.4
## + TotRmsAbvGrd             1   0.00349 10.148 -4518.4
## + HalfBath                 1   0.00260 10.149 -4518.4
## + FullBath                 1   0.00229 10.150 -4518.3
## - X1stFlrSF                1   0.03913 10.191 -4518.3
## + OverallQual:GrLivArea    1   0.00004 10.152 -4518.1
## + GarageYrBlt              1   0.00000 10.152 -4518.1
## - GrLivArea:GarageCars     1   0.04590 10.198 -4517.6
## - WoodDeckSF               1   0.04650 10.198 -4517.5
## - BedroomAbvGr             1   0.05525 10.207 -4516.7
## - GarageArea               1   0.06606 10.218 -4515.6
## - YearRemodAdd             1   0.06723 10.219 -4515.5
```

```
## - X2ndFlrSF               1   0.07280 10.225 -4515.0
## - BsmtFullBath            1   0.08012 10.232 -4514.2
## - KitchenAbvGr            1   0.09557 10.247 -4512.7
## - TotalBsmtSF:GarageCars  1   0.14831 10.300 -4507.6
## - EncPorchSF              1   0.19309 10.345 -4503.3
## - MSSubClass              1   0.21782 10.370 -4500.9
## - TotalBsmtSF:GrLivArea   1   0.26261 10.415 -4496.6
## - OverallQual:TotalBsmtSF 1   0.29930 10.451 -4493.0
## - Fireplaces              1   0.32707 10.479 -4490.4
## - BsmtFinSF1              1   0.41626 10.568 -4481.9
## - LotArea                 1   0.72565 10.877 -4453.1
## - OverallCond             1   2.34420 12.496 -4314.3
## - YearBuilt               1   2.63697 12.789 -4291.2
##
## Step:  AIC=-4521.54
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
##     TotalBsmtSF + X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath +
##     BsmtHalfBath + BedroomAbvGr + KitchenAbvGr + Fireplaces +
##     GarageCars + GarageArea + WoodDeckSF + OpenPorchSF + EncPorchSF +
##     PoolArea + MiscVal + MoSold + YrSold + OverallQual:GarageCars +
##     OverallQual:TotalBsmtSF + GrLivArea:GarageCars + TotalBsmtSF:GrLivArea +
##     TotalBsmtSF:GarageCars
##
##                          Df Sum of Sq    RSS     AIC
## - MasVnrArea              1   0.00622 10.164 -4522.9
## - Id                      1   0.00717 10.165 -4522.8
## - MoSold                  1   0.00749 10.165 -4522.8
## - BsmtHalfBath            1   0.00939 10.167 -4522.6
## - PoolArea                1   0.01032 10.168 -4522.5
## - OverallQual:GarageCars  1   0.01444 10.172 -4522.1
## - LotFrontage             1   0.02022 10.178 -4521.6
## <none>                                10.158 -4521.5
## - YrSold                  1   0.02253 10.180 -4521.3
## - MiscVal                 1   0.02498 10.182 -4521.1
## - OpenPorchSF             1   0.02669 10.184 -4520.9
## + BsmtFinSF2              1   0.00568 10.152 -4520.1
## + BsmtUnfSF               1   0.00568 10.152 -4520.1
## + TotRmsAbvGrd            1   0.00354 10.154 -4519.9
## + HalfBath                1   0.00307 10.155 -4519.8
## - X1stFlrSF               1   0.03803 10.196 -4519.8
## + FullBath                1   0.00222 10.155 -4519.8
## + GarageYrBlt             1   0.00003 10.158 -4519.5
## + OverallQual:GrLivArea   1   0.00003 10.158 -4519.5
## - GrLivArea:GarageCars    1   0.04533 10.203 -4519.1
## - WoodDeckSF              1   0.04741 10.205 -4518.9
## - BedroomAbvGr            1   0.05576 10.213 -4518.1
## - YearRemodAdd            1   0.06553 10.223 -4517.1
## - GarageArea              1   0.06755 10.225 -4516.9
## - X2ndFlrSF               1   0.07138 10.229 -4516.5
## - KitchenAbvGr            1   0.09637 10.254 -4514.1
## - BsmtFullBath            1   0.09900 10.257 -4513.8
## - TotalBsmtSF:GarageCars  1   0.14991 10.307 -4508.9
## - EncPorchSF              1   0.19368 10.351 -4504.7
```

```
## - MSSubClass                1   0.21669 10.374 -4502.4
## - TotalBsmtSF:GrLivArea     1   0.26490 10.422 -4497.8
## - OverallQual:TotalBsmtSF   1   0.30333 10.461 -4494.1
## - Fireplaces                1   0.33458 10.492 -4491.1
## - BsmtFinSF1                1   0.42343 10.581 -4482.7
## - LotArea                   1   0.74582 10.903 -4452.7
## - OverallCond               1   2.36575 12.523 -4314.2
## - YearBuilt                 1   2.65011 12.808 -4291.7
##
## Step:  AIC=-4522.93
## log(SalePrice) ~ Id + MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     BedroomAbvGr + KitchenAbvGr + Fireplaces + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + MiscVal +
##     MoSold + YrSold + OverallQual:GarageCars + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - Id                       1   0.00735 10.171 -4524.2
## - MoSold                   1   0.00805 10.172 -4524.1
## - BsmtHalfBath             1   0.00975 10.174 -4524.0
## - PoolArea                 1   0.01026 10.174 -4523.9
## - OverallQual:GarageCars   1   0.01446 10.178 -4523.5
## - LotFrontage              1   0.02034 10.184 -4522.9
## <none>                                 10.164 -4522.9
## - YrSold                   1   0.02283 10.187 -4522.7
## - MiscVal                  1   0.02474 10.188 -4522.5
## - OpenPorchSF              1   0.02586 10.190 -4522.4
## + MasVnrArea               1   0.00622 10.158 -4521.5
## + BsmtFinSF2               1   0.00525 10.159 -4521.4
## + BsmtUnfSF                1   0.00525 10.159 -4521.4
## + HalfBath                 1   0.00343 10.160 -4521.3
## + TotRmsAbvGrd             1   0.00315 10.161 -4521.2
## - X1stFlrSF                1   0.03858 10.202 -4521.1
## + FullBath                 1   0.00182 10.162 -4521.1
## + OverallQual:GrLivArea    1   0.00016 10.164 -4520.9
## + GarageYrBlt              1   0.00005 10.164 -4520.9
## - GrLivArea:GarageCars     1   0.04187 10.206 -4520.8
## - WoodDeckSF               1   0.04675 10.210 -4520.3
## - BedroomAbvGr             1   0.05489 10.219 -4519.5
## - YearRemodAdd             1   0.06270 10.226 -4518.8
## - GarageArea               1   0.06915 10.233 -4518.1
## - X2ndFlrSF                1   0.07239 10.236 -4517.8
## - BsmtFullBath             1   0.09613 10.260 -4515.5
## - KitchenAbvGr             1   0.09622 10.260 -4515.5
## - TotalBsmtSF:GarageCars   1   0.14922 10.313 -4510.4
## - EncPorchSF               1   0.19514 10.359 -4505.9
## - MSSubClass               1   0.21247 10.376 -4504.2
## - TotalBsmtSF:GrLivArea    1   0.26172 10.425 -4499.5
## - OverallQual:TotalBsmtSF  1   0.30395 10.468 -4495.5
## - Fireplaces               1   0.33730 10.501 -4492.3
## - BsmtFinSF1               1   0.44253 10.606 -4482.3
## - LotArea                  1   0.74341 10.907 -4454.3
```

```
## - OverallCond              1   2.37042 12.534 -4315.3
## - YearBuilt                1   2.72004 12.884 -4287.8
##
## Step:  AIC=-4524.2
## log(SalePrice) ~ MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     BedroomAbvGr + KitchenAbvGr + Fireplaces + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + MiscVal +
##     MoSold + YrSold + OverallQual:GarageCars + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                          Df Sum of Sq    RSS     AIC
## - MoSold                  1   0.00748 10.179 -4525.5
## - BsmtHalfBath            1   0.00981 10.181 -4525.2
## - PoolArea                1   0.00989 10.181 -4525.2
## - OverallQual:GarageCars  1   0.01461 10.186 -4524.8
## - LotFrontage             1   0.01994 10.191 -4524.2
## <none>                                10.171 -4524.2
## - YrSold                  1   0.02224 10.193 -4524.0
## - MiscVal                 1   0.02336 10.194 -4523.9
## - OpenPorchSF             1   0.02696 10.198 -4523.6
## + Id                      1   0.00735 10.164 -4522.9
## + MasVnrArea              1   0.00640 10.165 -4522.8
## + BsmtFinSF2              1   0.00518 10.166 -4522.7
## + BsmtUnfSF               1   0.00518 10.166 -4522.7
## + HalfBath                1   0.00337 10.168 -4522.5
## + TotRmsAbvGrd            1   0.00333 10.168 -4522.5
## - X1stFlrSF               1   0.03803 10.209 -4522.5
## + FullBath                1   0.00187 10.169 -4522.4
## + OverallQual:GrLivArea   1   0.00025 10.171 -4522.2
## + GarageYrBlt             1   0.00004 10.171 -4522.2
## - GrLivArea:GarageCars    1   0.04122 10.212 -4522.2
## - WoodDeckSF              1   0.04738 10.219 -4521.6
## - BedroomAbvGr            1   0.05395 10.225 -4520.9
## - YearRemodAdd            1   0.06505 10.236 -4519.8
## - GarageArea              1   0.06844 10.239 -4519.5
## - X2ndFlrSF               1   0.07091 10.242 -4519.3
## - KitchenAbvGr            1   0.09572 10.267 -4516.8
## - BsmtFullBath            1   0.09640 10.268 -4516.8
## - TotalBsmtSF:GarageCars  1   0.14712 10.318 -4511.8
## - EncPorchSF              1   0.20121 10.372 -4506.6
## - MSSubClass              1   0.21228 10.383 -4505.5
## - TotalBsmtSF:GrLivArea   1   0.25740 10.428 -4501.2
## - OverallQual:TotalBsmtSF 1   0.29722 10.468 -4497.4
## - Fireplaces              1   0.33721 10.508 -4493.6
## - BsmtFinSF1              1   0.44257 10.614 -4483.6
## - LotArea                 1   0.74322 10.914 -4455.7
## - OverallCond             1   2.36309 12.534 -4317.3
## - YearBuilt               1   2.72101 12.892 -4289.1
##
## Step:  AIC=-4525.47
## log(SalePrice) ~ MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
```

```
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     BedroomAbvGr + KitchenAbvGr + Fireplaces + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EncPorchSF + PoolArea + MiscVal +
##     YrSold + OverallQual:GarageCars + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                          Df Sum of Sq    RSS     AIC
## - BsmtHalfBath            1   0.00882 10.187 -4526.6
## - PoolArea                1   0.01079 10.189 -4526.4
## - OverallQual:GarageCars  1   0.01390 10.193 -4526.1
## - YrSold                  1   0.01927 10.198 -4525.6
## - LotFrontage             1   0.01973 10.198 -4525.5
## <none>                              10.179 -4525.5
## - MiscVal                 1   0.02374 10.202 -4525.1
## - OpenPorchSF             1   0.02612 10.205 -4524.9
## + MoSold                  1   0.00748 10.171 -4524.2
## + MasVnrArea              1   0.00693 10.172 -4524.2
## + Id                      1   0.00678 10.172 -4524.1
## + BsmtFinSF2              1   0.00514 10.173 -4524.0
## + BsmtUnfSF               1   0.00514 10.173 -4524.0
## - X1stFlrSF               1   0.03667 10.215 -4523.9
## + TotRmsAbvGrd            1   0.00388 10.175 -4523.9
## + HalfBath                1   0.00359 10.175 -4523.8
## + FullBath                1   0.00197 10.177 -4523.7
## + OverallQual:GrLivArea   1   0.00034 10.178 -4523.5
## + GarageYrBlt             1   0.00000 10.179 -4523.5
## - GrLivArea:GarageCars    1   0.04110 10.220 -4523.4
## - WoodDeckSF              1   0.04606 10.225 -4523.0
## - BedroomAbvGr            1   0.05377 10.232 -4522.2
## - YearRemodAdd            1   0.06518 10.244 -4521.1
## - GarageArea              1   0.06890 10.248 -4520.7
## - X2ndFlrSF               1   0.06907 10.248 -4520.7
## - BsmtFullBath            1   0.09578 10.274 -4518.1
## - KitchenAbvGr            1   0.10083 10.279 -4517.6
## - TotalBsmtSF:GarageCars  1   0.14521 10.324 -4513.3
## - EncPorchSF              1   0.20115 10.380 -4507.9
## - MSSubClass              1   0.20922 10.388 -4507.1
## - TotalBsmtSF:GrLivArea   1   0.25366 10.432 -4502.9
## - OverallQual:TotalBsmtSF 1   0.29715 10.476 -4498.7
## - Fireplaces              1   0.33162 10.510 -4495.4
## - BsmtFinSF1              1   0.44321 10.622 -4484.8
## - LotArea                 1   0.74526 10.924 -4456.8
## - OverallCond             1   2.36346 12.542 -4318.7
## - YearBuilt               1   2.73820 12.917 -4289.2
##
## Step:  AIC=-4526.6
## log(SalePrice) ~ MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BedroomAbvGr +
##     KitchenAbvGr + Fireplaces + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EncPorchSF + PoolArea + MiscVal + YrSold +
##     OverallQual:GarageCars + OverallQual:TotalBsmtSF + GrLivArea:GarageCars +
##     TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
```

```
##                           Df Sum of Sq    RSS     AIC
## - PoolArea                 1   0.01057 10.198 -4527.6
## - OverallQual:GarageCars   1   0.01288 10.200 -4527.3
## - LotFrontage              1   0.01986 10.207 -4526.7
## <none>                                 10.187 -4526.6
## - YrSold                   1   0.02043 10.208 -4526.6
## - MiscVal                  1   0.02374 10.211 -4526.3
## - OpenPorchSF              1   0.02723 10.215 -4525.9
## + BsmtHalfBath             1   0.00882 10.179 -4525.5
## + BsmtFinSF2               1   0.00728 10.180 -4525.3
## + BsmtUnfSF                1   0.00728 10.180 -4525.3
## + MasVnrArea               1   0.00726 10.180 -4525.3
## + Id                       1   0.00687 10.181 -4525.3
## + MoSold                   1   0.00649 10.181 -4525.2
## - X1stFlrSF                1   0.03682 10.224 -4525.0
## + TotRmsAbvGrd             1   0.00375 10.184 -4525.0
## + HalfBath                 1   0.00345 10.184 -4524.9
## - GrLivArea:GarageCars     1   0.03861 10.226 -4524.8
## + FullBath                 1   0.00133 10.186 -4524.7
## + OverallQual:GrLivArea    1   0.00022 10.187 -4524.6
## + GarageYrBlt              1   0.00003 10.187 -4524.6
## - WoodDeckSF               1   0.04901 10.236 -4523.8
## - BedroomAbvGr             1   0.05180 10.239 -4523.5
## - YearRemodAdd             1   0.06667 10.254 -4522.1
## - GarageArea               1   0.06795 10.255 -4522.0
## - X2ndFlrSF                1   0.06889 10.256 -4521.9
## - BsmtFullBath             1   0.08702 10.274 -4520.1
## - KitchenAbvGr             1   0.09968 10.287 -4518.9
## - TotalBsmtSF:GarageCars   1   0.14548 10.333 -4514.4
## - EncPorchSF               1   0.20026 10.388 -4509.1
## - MSSubClass               1   0.20801 10.395 -4508.4
## - TotalBsmtSF:GrLivArea    1   0.25182 10.439 -4504.2
## - OverallQual:TotalBsmtSF  1   0.29432 10.482 -4500.1
## - Fireplaces               1   0.33490 10.522 -4496.3
## - BsmtFinSF1               1   0.48952 10.677 -4481.7
## - LotArea                  1   0.75983 10.947 -4456.7
## - OverallCond              1   2.41180 12.599 -4316.1
## - YearBuilt                1   2.74966 12.937 -4289.7
##
## Step:  AIC=-4527.57
## log(SalePrice) ~ MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BedroomAbvGr +
##     KitchenAbvGr + Fireplaces + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EncPorchSF + MiscVal + YrSold + OverallQual:GarageCars +
##     OverallQual:TotalBsmtSF + GrLivArea:GarageCars + TotalBsmtSF:GrLivArea +
##     TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## - OverallQual:GarageCars   1   0.01224 10.210 -4528.4
## <none>                                 10.198 -4527.6
## - LotFrontage              1   0.02103 10.219 -4527.5
## - YrSold                   1   0.02171 10.220 -4527.4
## - MiscVal                  1   0.02388 10.222 -4527.2
```

```
## - OpenPorchSF              1   0.02626 10.224 -4527.0
## + PoolArea                 1   0.01057 10.187 -4526.6
## + BsmtHalfBath             1   0.00860 10.189 -4526.4
## + BsmtFinSF2               1   0.00779 10.190 -4526.3
## + BsmtUnfSF                1   0.00779 10.190 -4526.3
## + MoSold                   1   0.00734 10.191 -4526.3
## + MasVnrArea               1   0.00722 10.191 -4526.3
## + Id                       1   0.00646 10.191 -4526.2
## - X1stFlrSF                1   0.03677 10.235 -4526.0
## + HalfBath                 1   0.00401 10.194 -4526.0
## + TotRmsAbvGrd             1   0.00336 10.195 -4525.9
## - GrLivArea:GarageCars     1   0.03777 10.236 -4525.9
## + FullBath                 1   0.00104 10.197 -4525.7
## + OverallQual:GrLivArea    1   0.00019 10.198 -4525.6
## + GarageYrBlt              1   0.00005 10.198 -4525.6
## - BedroomAbvGr             1   0.05001 10.248 -4524.7
## - WoodDeckSF               1   0.05147 10.249 -4524.5
## - YearRemodAdd             1   0.06749 10.265 -4523.0
## - GarageArea               1   0.06801 10.266 -4522.9
## - X2ndFlrSF                1   0.06861 10.267 -4522.9
## - BsmtFullBath             1   0.09032 10.288 -4520.7
## - KitchenAbvGr             1   0.09920 10.297 -4519.9
## - TotalBsmtSF:GarageCars   1   0.14629 10.344 -4515.3
## - EncPorchSF               1   0.19901 10.397 -4510.2
## - MSSubClass               1   0.20769 10.406 -4509.4
## - TotalBsmtSF:GrLivArea    1   0.25190 10.450 -4505.2
## - OverallQual:TotalBsmtSF  1   0.29493 10.493 -4501.1
## - Fireplaces               1   0.34046 10.538 -4496.7
## - BsmtFinSF1               1   0.48636 10.684 -4483.0
## - LotArea                  1   0.75777 10.956 -4457.9
## - OverallCond              1   2.41439 12.612 -4317.1
## - YearBuilt                1   2.75116 12.949 -4290.7
##
## Step:  AIC=-4528.37
## log(SalePrice) ~ MSSubClass + LotFrontage + LotArea + OverallQual +
##     OverallCond + YearBuilt + YearRemodAdd + BsmtFinSF1 + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + GrLivArea + BsmtFullBath + BedroomAbvGr +
##     KitchenAbvGr + Fireplaces + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EncPorchSF + MiscVal + YrSold + OverallQual:TotalBsmtSF +
##     GrLivArea:GarageCars + TotalBsmtSF:GrLivArea + TotalBsmtSF:GarageCars
##
##                           Df Sum of Sq    RSS     AIC
## <none>                                 10.210 -4528.4
## - YrSold                   1   0.02214 10.232 -4528.2
## - LotFrontage              1   0.02342 10.234 -4528.1
## - MiscVal                  1   0.02349 10.234 -4528.1
## - GrLivArea:GarageCars     1   0.02625 10.236 -4527.8
## - OpenPorchSF              1   0.02670 10.237 -4527.8
## + OverallQual:GarageCars   1   0.01224 10.198 -4527.6
## + PoolArea                 1   0.00993 10.200 -4527.3
## - X1stFlrSF                1   0.03236 10.243 -4527.2
## + BsmtHalfBath             1   0.00762 10.203 -4527.1
## + BsmtFinSF2               1   0.00721 10.203 -4527.1
## + BsmtUnfSF                1   0.00721 10.203 -4527.1
```

```
## + MasVnrArea              1   0.00720 10.203 -4527.1
## + MoSold                  1   0.00671 10.204 -4527.0
## + Id                      1   0.00663 10.204 -4527.0
## + TotRmsAbvGrd            1   0.00455 10.206 -4526.8
## + HalfBath                1   0.00378 10.206 -4526.7
## + OverallQual:GrLivArea   1   0.00197 10.208 -4526.6
## + FullBath                1   0.00115 10.209 -4526.5
## + GarageYrBlt             1   0.00054 10.210 -4526.4
## - BedroomAbvGr            1   0.04960 10.260 -4525.5
## - WoodDeckSF              1   0.05211 10.262 -4525.3
## - X2ndFlrSF               1   0.06398 10.274 -4524.1
## - GarageArea              1   0.06860 10.279 -4523.7
## - YearRemodAdd            1   0.07699 10.287 -4522.9
## - BsmtFullBath            1   0.09368 10.304 -4521.2
## - KitchenAbvGr            1   0.09897 10.309 -4520.7
## - TotalBsmtSF:GarageCars  1   0.13501 10.345 -4517.2
## - EncPorchSF              1   0.19516 10.405 -4511.4
## - MSSubClass              1   0.21681 10.427 -4509.4
## - TotalBsmtSF:GrLivArea   1   0.28286 10.493 -4503.0
## - Fireplaces              1   0.34304 10.553 -4497.3
## - OverallQual:TotalBsmtSF 1   0.37167 10.582 -4494.6
## - BsmtFinSF1              1   0.48075 10.691 -4484.4
## - LotArea                 1   0.75159 10.962 -4459.3
## - OverallCond             1   2.40586 12.616 -4318.8
## - YearBuilt               1   2.77006 12.980 -4290.3
```

Now that we have found this linear fit, we run a 5 fold cross validation to examine the

```r
fitOLMSim2Way <- train(OLMSim2Way$call$formula,
             data=housingData,
             method="lm",
             trControl=fitControl)


fitOLMSim2Way$results
```

```
##   intercept      RMSE  Rsquared        MAE     RMSESD RsquaredSD       MAESD
## 1      TRUE 0.1050121 0.9161997 0.07879346 0.01226514 0.01773947 0.008115876
```

```r
# Results
best_ols2way_RMSE <- fitOLMSim2Way$results$RMSE
best_ols2way_R2 <- fitOLMSim2Way$results$Rsquared
```

```r
summary(fitOLMSim2Way)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63172 -0.05645  0.00385  0.05717  0.34535
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.048e+01  5.042e+00   2.078 0.037956 *
```

```
## MSSubClass                 -4.351e-04  9.578e-05  -4.543 6.24e-06 ***
## LotFrontage                 1.489e-04  9.972e-05   1.493 0.135688
## LotArea                     3.008e-06  3.556e-07   8.459  < 2e-16 ***
## OverallQual                 1.980e-02  9.141e-03   2.166 0.030543 *
## OverallCond                 5.625e-02  3.717e-03  15.134  < 2e-16 ***
## YearBuilt                   3.189e-03  1.964e-04  16.239  < 2e-16 ***
## YearRemodAdd                6.220e-04  2.297e-04   2.707 0.006902 **
## BsmtFinSF1                  7.973e-05  1.178e-05   6.765 2.30e-11 ***
## TotalBsmtSF                 1.076e-04  3.488e-05   3.086 0.002088 **
## X1stFlrSF                   1.358e-04  7.736e-05   1.755 0.079535 .
## X2ndFlrSF                   1.878e-04  7.609e-05   2.468 0.013758 *
## GrLivArea                   2.568e-04  7.532e-05   3.409 0.000680 ***
## BsmtFullBath                2.557e-02  8.562e-03   2.986 0.002895 **
## BedroomAbvGr               -1.247e-02  5.738e-03  -2.173 0.030030 *
## KitchenAbvGr               -5.941e-02  1.936e-02  -3.070 0.002203 **
## Fireplaces                  3.623e-02  6.339e-03   5.715 1.46e-08 ***
## GarageCars                  1.094e-01  1.888e-02   5.797 9.10e-09 ***
## GarageArea                  9.183e-05  3.593e-05   2.555 0.010756 *
## WoodDeckSF                  6.309e-05  2.833e-05   2.227 0.026152 *
## OpenPorchSF                 8.966e-05  5.623e-05   1.594 0.111165
## EncPorchSF                  1.881e-04  4.365e-05   4.310 1.80e-05 ***
## MiscVal                    -2.610e-05  1.746e-05  -1.495 0.135117
## YrSold                     -3.668e-03  2.526e-03  -1.452 0.146839
## `OverallQual:TotalBsmtSF`   4.591e-05  7.718e-06   5.948 3.78e-09 ***
## `GrLivArea:GarageCars`     -1.593e-05  1.008e-05  -1.581 0.114258
## `TotalBsmtSF:GrLivArea`    -9.378e-08  1.807e-08  -5.189 2.57e-07 ***
## `TotalBsmtSF:GarageCars`   -5.564e-05  1.552e-05  -3.585 0.000354 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1025 on 972 degrees of freedom
## Multiple R-squared:  0.9225, Adjusted R-squared:  0.9203
## F-statistic: 428.5 on 27 and 972 DF,  p-value: < 2.2e-16
```

Lastly we attempt the MARS method to fit linear splines on the data.

```
marsFit <- earth(log(SalePrice) ~ ., data = housingNumeric)

summary(marsFit, style="pmax")
```

```
## Call: earth(formula=log(SalePrice)~., data=housingNumeric)
##
## log(SalePrice) =
##   12.03546
##   - 1.305917e-05 * pmax(0,       15431 -     LotArea)
##   + 2.088154e-06 * pmax(0,     LotArea -       15431)
##   -   0.04294704 * pmax(0,           6 - OverallQual)
##   +   0.08994435 * pmax(0, OverallQual -           6)
##   -   0.09304702 * pmax(0,           5 - OverallCond)
##   +   0.05757498 * pmax(0, OverallCond -           5)
##   -  0.003050568 * pmax(0,        2004 -   YearBuilt)
##   +   0.01314382 * pmax(0,   YearBuilt -        2004)
##   - 0.0001237953 * pmax(0,         650 -  BsmtFinSF1)
##   + 9.729898e-05 * pmax(0,  BsmtFinSF1 -         650)
##   - 0.0001440609 * pmax(0,        1324 - TotalBsmtSF)
```

```
##   - 0.0005580601 * pmax(0,         958 -   GrLivArea)
##   +  0.000249486 * pmax(0,   GrLivArea -         958)
##   -   0.05412901 * pmax(0,           1 -  Fireplaces)
##   +   0.03180269 * pmax(0,  Fireplaces -           1)
##   +  0.001319597 * pmax(0,        1973 - GarageYrBlt)
##   +  0.002683497 * pmax(0, GarageYrBlt -        1973)
##   -     2.649257 * pmax(0,           1 -  GarageCars)
##   +   0.03701742 * pmax(0,  GarageCars -           1)
##
## Selected 20 of 21 terms, and 10 of 35 predictors
## Termination condition: RSq changed by less than 0.001 at 21 terms
## Importance: YearBuilt, GrLivArea, OverallQual, TotalBsmtSF, OverallCond, ...
## Number of terms at each degree of interaction: 1 19 (additive model)
## GCV 0.01067386    RSS 9.85753    GRSq 0.9191446    RSq 0.9251788
```

```r
earthGrid <- expand.grid(nprune=10:20,degree = 1:5)
fitMARS <- train(log(SalePrice) ~ .,
                 data = housingNumeric,
                 method = "earth",
                 trControl = fitControl,
                 tuneGrid = earthGrid)
# Best tuning parameters
best_mars_tune <- fitMARS$bestTune
print(best_mars_tune)
```

```
##    nprune degree
## 11     20      1
```

```r
# Best results
best_mars <- fitMARS$results[which.min(fitMARS$results$RMSE), ]
best_mars_RMSE <- best_mars$RMSE
best_mars_R2 <- best_mars$Rsquared
```

Our final results are presented here.

```r
# OLS (from fitOLM)
best_ols_RMSE <- fitOLM$results$RMSE
best_ols_R2 <- fitOLM$results$Rsquared

# OLS Simplified (from fitOLMSim)
best_olssim_RMSE <- fitOLMSim$results$RMSE
best_olssim_R2 <- fitOLMSim$results$Rsquared



# Create the data frame
results_table <- data.frame(
  Model = c("OLS", "OLS Simplified", "OLS Two-Way Interactions", "Lasso",
            "Ridge Regression", "Elastic Net", "MARS"),
  Notes = c("lm", "lm with stepwise selection", "lm with interactions","lars",
            "ridge", "elastic net", "earth"),
  Hyperparameters = c(
    "N/A",
    "N/A",
    "N/A",
    paste("fraction =", round(fitLasso$bestTune$fraction, 5)),
```

```
    paste("lambda =", round(best_ridge_tune$lambda, 5)),
    paste("lambda =", round(best_enet_tune$lambda, 5),
          ", fraction =", round(best_enet_tune$fraction, 4)),
    paste("nprune =", best_mars_tune$nprune,
          ", degree =", best_mars_tune$degree)
  ),
  CVRMSE = round(c(best_ols_RMSE, best_olssim_RMSE, best_ols2way_RMSE, best_lasso_RMSE,
                   best_ridge_RMSE, best_enet_RMSE, best_mars_RMSE), 4),
  CV_R2 = round(c(best_ols_R2, best_olssim_R2, best_ols2way_R2, best_lasso_R2,
                  best_ridge_R2, best_enet_R2, best_mars_R2), 4)
)
```

```
kable(results_table, format = "markdown")
```

Model	Notes	Hyperparameters	CVRMSE	CV_R2
OLS	lm	N/A	0.1052	0.9175
OLS Simplified	lm with stepwise selection	N/A	0.0973	0.9283
OLS Two-Way Interactions	lm with interactions	N/A	0.1050	0.9162
Lasso	lars	fraction = 0.97081	0.1124	0.9049
Ridge Regression	ridge	lambda = 0.00667	0.1092	0.9098
Elastic Net	elastic net	lambda = 0.00222 , fraction = 0.9	0.1068	0.9138
MARS	earth	nprune = 20 , degree = 1	0.1093	0.9107