

# Stats 101A - Final Project R Code

Hao Ma

Winter 2019

Code for building the model only:

```
library(alr3)

## Loading required package: car
## Loading required package: carData
train <- read.csv("HTrainW19Final.csv")

train$Age1 <- 2019 - train$YearBuilt
train$Age2 <- 2019 - train$YearRemodAdd

train$TotalBsmtSF[which(is.na(train$TotalBsmtSF))] <-
  median(na.omit(train$TotalBsmtSF[which(train$TotalBsmtSF != 0)]))

summary(powerTransform(cbind(train$SalePrice, train$LotArea, train$GrLivArea)~1))

## bcPower Transformations to Multinormality
##      Est Power Rounded Pwr Wald Lwr Bnd Wald Up Bnd
## Y1   -0.0453         0.00   -0.1088      0.0182
## Y2    0.0659         0.07    0.0325      0.0993
## Y3   -0.0744         0.00   -0.1564      0.0076
##
## Likelihood ratio test that transformation parameters are equal to 0
## (all log transformations)
##
##              LRT df      pval
## LR test, lambda = (0 0 0) 18.96137  3 0.00027847
##
## Likelihood ratio test that no transformations are needed
##
##              LRT df      pval
## LR test, lambda = (1 1 1) 5564.304  3 < 2.22e-16

tSalePrice <- log(train$SalePrice)
tLotArea <- log(train$LotArea)
tGrLivArea <- log(train$GrLivArea)

N <- as.integer(train$Neighborhood)
for(i in 1:nrow(train)){
  if(N[i] %in% c(1,5,6,7,9,12,13,17,20,21)){
    train$Ngroup[i] <- "Group 1"
  }else if(N[i] %in% c(2,3,11,15)){
    train$Ngroup[i] <- "Group 2"
  }else if(N[i] %in% c(4,8,10,18,19,23)){
    train$Ngroup[i] <- "Group 3"
  }else{
    train$Ngroup[i] <- "Group 4"
  }
}
```

```

train$Ngroup <- as.factor(train$Ngroup)

train$GarageCars <- as.factor(train$GarageCars)
train$Fireplaces <- as.factor(train$Fireplaces)
train$FullBath <- as.factor(train$FullBath)

train$MasVnrArea[which(is.na(train$MasVnrArea))] <-
  median(na.omit(train$MasVnrArea[which(train$MasVnrArea != 0)]))
med.mas.vnr <- median(train$MasVnrArea[which(train$MasVnrArea != 0)])
for(i in 1:nrow(train)){
  if(train$MasVnrArea[i] == 0) {train$MasVnr[i] <- "No MasVnr"}
  if(train$MasVnrArea[i] != 0 & train$MasVnrArea[i] <= med.mas.vnr) {train$MasVnr[i] <- "Small MasVnr"}
  if(train$MasVnrArea[i] != 0 & train$MasVnrArea[i] > med.mas.vnr) {train$MasVnr[i] <- "Large MasVnr"}
}
train$MasVnr <- as.factor(train$MasVnr)

# the model:
model3 <- lm(tSalePrice ~ MasVnr:tLotArea + MasVnr:OverallQual + MasVnr:TotalBsmtSF + MasVnr:Age1 +
  MasVnr:Age2 + MasVnr:tGrLivArea + FullBath + MasVnr:TotRmsAbvGrd + Fireplaces +
  MasVnr:Ngroup + Foundation + MasVnr:GarageCars, data = train)
summary(model3)

```

```

##
## Call:
## lm(formula = tSalePrice ~ MasVnr:tLotArea + MasVnr:OverallQual +
##   MasVnr:TotalBsmtSF + MasVnr:Age1 + MasVnr:Age2 + MasVnr:tGrLivArea +
##   FullBath + MasVnr:TotRmsAbvGrd + Fireplaces + MasVnr:Ngroup +
##   Foundation + MasVnr:GarageCars, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81364 -0.05881  0.00696  0.06448  0.51144
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.353e+00  1.223e-01  68.308 < 2e-16
## FullBath1       -3.920e-02  4.370e-02  -0.897  0.369759
## FullBath2       -4.881e-02  4.391e-02  -1.112  0.266417
## FullBath3       -7.052e-03  4.792e-02  -0.147  0.883020
## FullBath4        6.518e-02  7.340e-02   0.888  0.374587
## Fireplaces1      2.716e-02  5.944e-03   4.569  5.14e-06
## Fireplaces2      8.991e-02  1.096e-02   8.207  3.62e-16
## Fireplaces3     -2.670e-02  3.782e-02  -0.706  0.480280
## Fireplaces4      1.635e-01  1.214e-01   1.346  0.178452
## FoundationCBlock  6.773e-02  1.032e-02   6.562  6.46e-11
## FoundationPConc   5.770e-02  1.213e-02   4.758  2.07e-06
## FoundationSlab    8.135e-02  2.177e-02   3.737  0.000190
## FoundationStone  -3.339e-02  3.356e-02  -0.995  0.319874
## FoundationWood    9.163e-02  7.148e-02   1.282  0.200002
## MasVnrLarge MasVnr:tLotArea  5.245e-02  1.367e-02   3.836  0.000128
## MasVnrNo MasVnr:tLotArea    5.550e-02  8.119e-03   6.835  1.03e-11
## MasVnrSmall MasVnr:tLotArea  7.719e-02  1.300e-02   5.936  3.34e-09

```

```

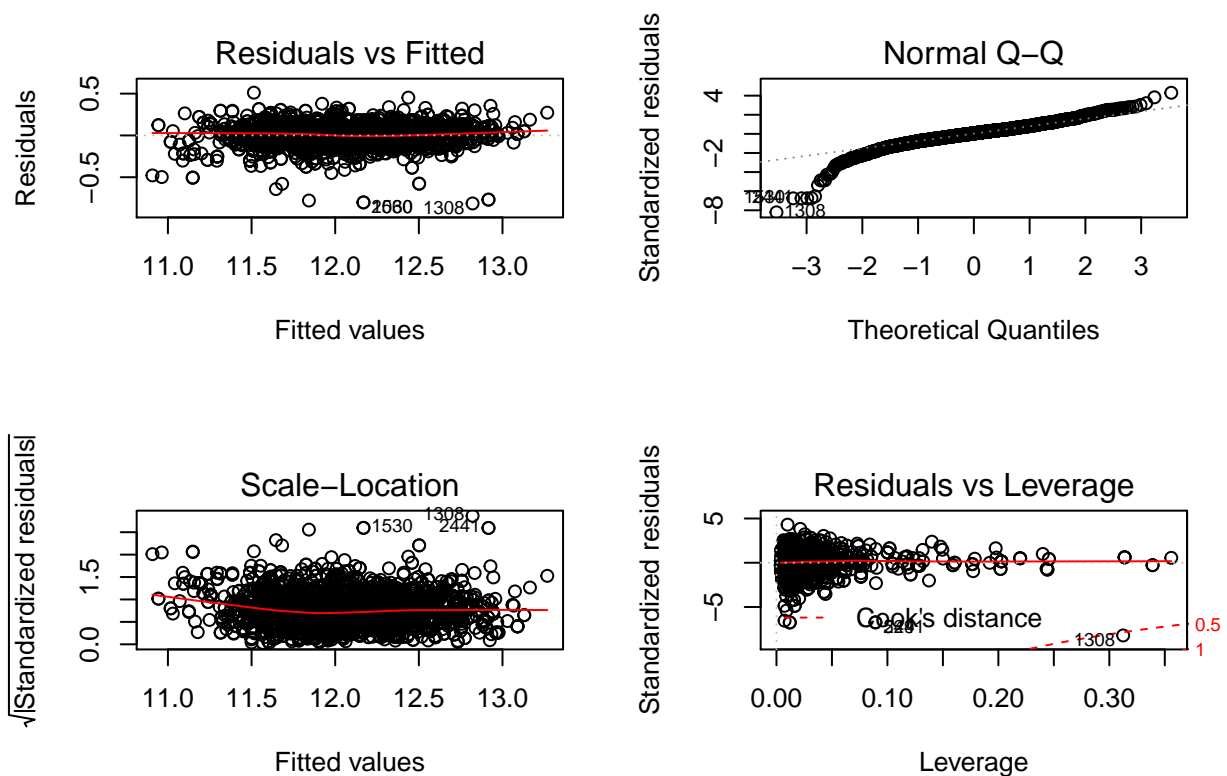
## MasVnrLarge MasVnr:OverallQual    9.854e-02  7.910e-03  12.457 < 2e-16
## MasVnrNo MasVnr:OverallQual       9.367e-02  3.910e-03  23.956 < 2e-16
## MasVnrSmall MasVnr:OverallQual    8.913e-02  7.975e-03  11.175 < 2e-16
## MasVnrLarge MasVnr:TotalBsmtSF   -3.255e-06  1.365e-05  -0.238 0.811572
## MasVnrNo MasVnr:TotalBsmtSF       1.813e-04  1.091e-05  16.619 < 2e-16
## MasVnrSmall MasVnr:TotalBsmtSF    1.265e-04  1.570e-05   8.058 1.20e-15
## MasVnrLarge MasVnr:Age1           3.720e-04  6.469e-04   0.575 0.565373
## MasVnrNo MasVnr:Age1              -1.630e-03  1.910e-04  -8.534 < 2e-16
## MasVnrSmall MasVnr:Age1           -2.410e-03  6.400e-04  -3.765 0.000170
## MasVnrLarge MasVnr:Age2           -3.226e-03  6.140e-04  -5.254 1.62e-07
## MasVnrNo MasVnr:Age2              -1.462e-03  1.824e-04  -8.012 1.74e-15
## MasVnrSmall MasVnr:Age2           -1.239e-03  5.285e-04  -2.345 0.019125
## MasVnrLarge MasVnr:tGrLivArea      4.128e-01  2.559e-02  16.131 < 2e-16
## MasVnrNo MasVnr:tGrLivArea         3.242e-01  1.831e-02  17.709 < 2e-16
## MasVnrSmall MasVnr:tGrLivArea      3.495e-01  2.504e-02  13.955 < 2e-16
## MasVnrLarge MasVnr:TotRmsAbvGrd   -1.102e-02  4.888e-03  -2.254 0.024304
## MasVnrNo MasVnr:TotRmsAbvGrd      6.200e-03  3.331e-03   1.861 0.062815
## MasVnrSmall MasVnr:TotRmsAbvGrd   -1.773e-02  5.106e-03  -3.473 0.000524
## MasVnrLarge MasVnr:NgroupGroup 2  -2.326e-01  3.921e-02  -5.932 3.41e-09
## MasVnrNo MasVnr:NgroupGroup 2     9.791e-03  2.083e-02   0.470 0.638299
## MasVnrSmall MasVnr:NgroupGroup 2  -2.272e-02  6.323e-02  -0.359 0.719360
## MasVnrLarge MasVnr:NgroupGroup 3  -2.469e-01  2.630e-02  -9.388 < 2e-16
## MasVnrNo MasVnr:NgroupGroup 3     -2.465e-02  9.161e-03  -2.691 0.007176
## MasVnrSmall MasVnr:NgroupGroup 3  -2.690e-02  1.641e-02  -1.639 0.101268
## MasVnrLarge MasVnr:NgroupGroup 4   1.312e-01  1.525e-02   8.604 < 2e-16
## MasVnrNo MasVnr:NgroupGroup 4     3.473e-02  1.768e-02   1.965 0.049570
## MasVnrSmall MasVnr:NgroupGroup 4   8.829e-03  1.603e-02   0.551 0.581827
## MasVnrLarge MasVnr:GarageCars1    -1.935e-01  7.445e-02  -2.599 0.009409
## MasVnrNo MasVnr:GarageCars1       1.169e-01  1.259e-02   9.291 < 2e-16
## MasVnrSmall MasVnr:GarageCars1     1.027e-02  3.992e-02   0.257 0.796979
## MasVnrLarge MasVnr:GarageCars2    -2.149e-01  7.358e-02  -2.921 0.003522
## MasVnrNo MasVnr:GarageCars2       1.481e-01  1.326e-02  11.170 < 2e-16
## MasVnrSmall MasVnr:GarageCars2     1.461e-02  4.091e-02   0.357 0.720974
## MasVnrLarge MasVnr:GarageCars3    -1.345e-01  7.607e-02  -1.768 0.077143
## MasVnrNo MasVnr:GarageCars3       2.098e-01  2.153e-02   9.745 < 2e-16
## MasVnrSmall MasVnr:GarageCars3     8.352e-02  4.556e-02   1.833 0.066921
## MasVnrLarge MasVnr:GarageCars4    -3.361e-02  9.520e-02  -0.353 0.724050
## MasVnrNo MasVnr:GarageCars4       3.656e-01  5.168e-02   7.075 1.94e-12
## MasVnrSmall MasVnr:GarageCars4     1.402e-01  6.634e-02   2.113 0.034663
##
## (Intercept)                        ***
## FullBath1
## FullBath2
## FullBath3
## FullBath4
## Fireplaces1                        ***
## Fireplaces2                        ***
## Fireplaces3
## Fireplaces4
## FoundationCBlock                   ***
## FoundationPConc                    ***
## FoundationSlab                     ***
## FoundationStone
## FoundationWood

```

```

## MasVnrLarge MasVnr:tLotArea      ***
## MasVnrNo MasVnr:tLotArea         ***
## MasVnrSmall MasVnr:tLotArea      ***
## MasVnrLarge MasVnr:OverallQual   ***
## MasVnrNo MasVnr:OverallQual      ***
## MasVnrSmall MasVnr:OverallQual   ***
## MasVnrLarge MasVnr:TotalBsmtSF    ***
## MasVnrNo MasVnr:TotalBsmtSF      ***
## MasVnrSmall MasVnr:TotalBsmtSF    ***
## MasVnrLarge MasVnr:Age1           ***
## MasVnrNo MasVnr:Age1              ***
## MasVnrSmall MasVnr:Age1           ***
## MasVnrLarge MasVnr:Age2           ***
## MasVnrNo MasVnr:Age2              ***
## MasVnrSmall MasVnr:Age2           *
## MasVnrLarge MasVnr:tGrLivArea     ***
## MasVnrNo MasVnr:tGrLivArea        ***
## MasVnrSmall MasVnr:tGrLivArea     ***
## MasVnrLarge MasVnr:TotRmsAbvGrd   *
## MasVnrNo MasVnr:TotRmsAbvGrd     .
## MasVnrSmall MasVnr:TotRmsAbvGrd   ***
## MasVnrLarge MasVnr:NgroupGroup 2 ***
## MasVnrNo MasVnr:NgroupGroup 2     *
## MasVnrSmall MasVnr:NgroupGroup 2  *
## MasVnrLarge MasVnr:NgroupGroup 3 ***
## MasVnrNo MasVnr:NgroupGroup 3     **
## MasVnrSmall MasVnr:NgroupGroup 3  *
## MasVnrLarge MasVnr:NgroupGroup 4 ***
## MasVnrNo MasVnr:NgroupGroup 4     *
## MasVnrSmall MasVnr:NgroupGroup 4  *
## MasVnrLarge MasVnr:GarageCars1    **
## MasVnrNo MasVnr:GarageCars1       ***
## MasVnrSmall MasVnr:GarageCars1    *
## MasVnrLarge MasVnr:GarageCars2    **
## MasVnrNo MasVnr:GarageCars2       ***
## MasVnrSmall MasVnr:GarageCars2    *
## MasVnrLarge MasVnr:GarageCars3    .
## MasVnrNo MasVnr:GarageCars3       ***
## MasVnrSmall MasVnr:GarageCars3    .
## MasVnrLarge MasVnr:GarageCars4    *
## MasVnrNo MasVnr:GarageCars4       ***
## MasVnrSmall MasVnr:GarageCars4    *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1195 on 2444 degrees of freedom
## Multiple R-squared:  0.909, Adjusted R-squared:  0.907
## F-statistic: 444 on 55 and 2444 DF, p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(model3)

```



```
model3_check <- lm(tSalePrice ~ tLotArea + OverallQual + TotalBsmtSF + Age1 + Age2 + tGrLivArea +
  FullBath + TotRmsAbvGrd + Fireplaces + Ngroup + Foundation + GarageCars +
  MasVnr, data = train)
vif(model3_check)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## tLotArea      1.966472 1      1.402309
## OverallQual   3.524755 1      1.877433
## TotalBsmtSF   2.134215 1      1.460895
## Age1          5.045820 1      2.246290
## Age2          2.094033 1      1.447077
## tGrLivArea    5.253185 1      2.291983
## FullBath      3.588080 4      1.173161
## TotRmsAbvGrd  3.286231 1      1.812797
## Fireplaces    1.721764 4      1.070278
## Ngroup        4.795043 3      1.298570
## Foundation    4.812399 5      1.170136
## GarageCars    4.045045 4      1.190873
## MasVnr        1.642264 2      1.132037
```

Code for predicting house price in testing data set:

```
test <- read.csv("HTestW19Final No Y values.csv")

test$Age1 <- 2019 - test$YearBuilt
test$Age2 <- 2019 - test$YearRemodAdd
```

```

test$TotalBsmtSF[which(is.na(test$TotalBsmtSF))] <-
  median(na.omit(test$TotalBsmtSF[which(test$TotalBsmtSF != 0)]))

tLotArea <- log(test$LotArea)
tGrLivArea <- log(test$GrLivArea)

test$GarageCars <- as.factor(test$GarageCars)
test$Fireplaces <- as.factor(test$Fireplaces)
test$FullBath <- as.factor(test$FullBath)

test$MasVnrArea[which(is.na(test$MasVnrArea))] <-
  median(na.omit(test$MasVnrArea[which(test$MasVnrArea != 0)]))
med.mas.vnr <- median(test$MasVnrArea[which(test$MasVnrArea != 0)])
for(i in 1:nrow(test)){
  if(test$MasVnrArea[i] == 0) {test$MasVnr[i] <- "No MasVnr"}
  if(test$MasVnrArea[i] != 0 & test$MasVnrArea[i] <= med.mas.vnr) {test$MasVnr[i] <- "Small MasVnr"}
  if(test$MasVnrArea[i] != 0 & test$MasVnrArea[i] > med.mas.vnr) {test$MasVnr[i] <- "Large MasVnr"}
}
test$MasVnr <- as.factor(test$MasVnr)

Ntest <- as.integer(test$Neighborhood)
for(i in 1:nrow(test)){
  if(Ntest[i] %in% c(1,5,6,7,9,12,13,17,20,21)){
    test$Ngroup[i] <- "Group 1"
  }else if(Ntest[i] %in% c(2,3,11,15)){
    test$Ngroup[i] <- "Group 2"
  }else if(Ntest[i] %in% c(4,8,10,18,19,23)){
    test$Ngroup[i] <- "Group 3"
  }else{
    test$Ngroup[i] <- "Group 4"
  }
}
test$Ngroup <- as.factor(test$Ngroup)

p <- predict(model3, newdata = test)
p[is.na(p)] <- median(na.omit(p))
price <- exp(p)
my_prediction <- data.frame(Ob = 1:1500, SalePrice = round(price,2))

write.csv(my_prediction, "SalePrice_Hao_Ma_Lec1.csv")

```