Predicting House Sale Price

```
rm(list = ls())
library(car)
```

```
## Loading required package: carData
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(forcats)

setwd("C:\\Users\\naimesh.chaudhari\\Desktop\\H6 Applied DM")

dftrain <- read.csv("train.csv")
dftest <- read.csv("test.csv")
#there are select colums where a N/A ment none, so we are replacing NA/s with none for thise
validcols <- c("Alley","BsmtQual","BsmtCond", "BsmtExposure", "BsmtFinType1"
               , "BsmtFinType2", "FireplaceQu", "GarageType", "GarageFinish", "GarageQual"
               , "GarageCond", "PoolQC", "Fence", "MiscFeature")

#data is evenly spread
nrow(dftrain)
```

```
## [1] 1460
```

```
ncol(dftrain)
```

```
## [1] 81
```

```
nrow(dftest)
```

```
## [1] 1459
```

```r
ncol(dftest)
```

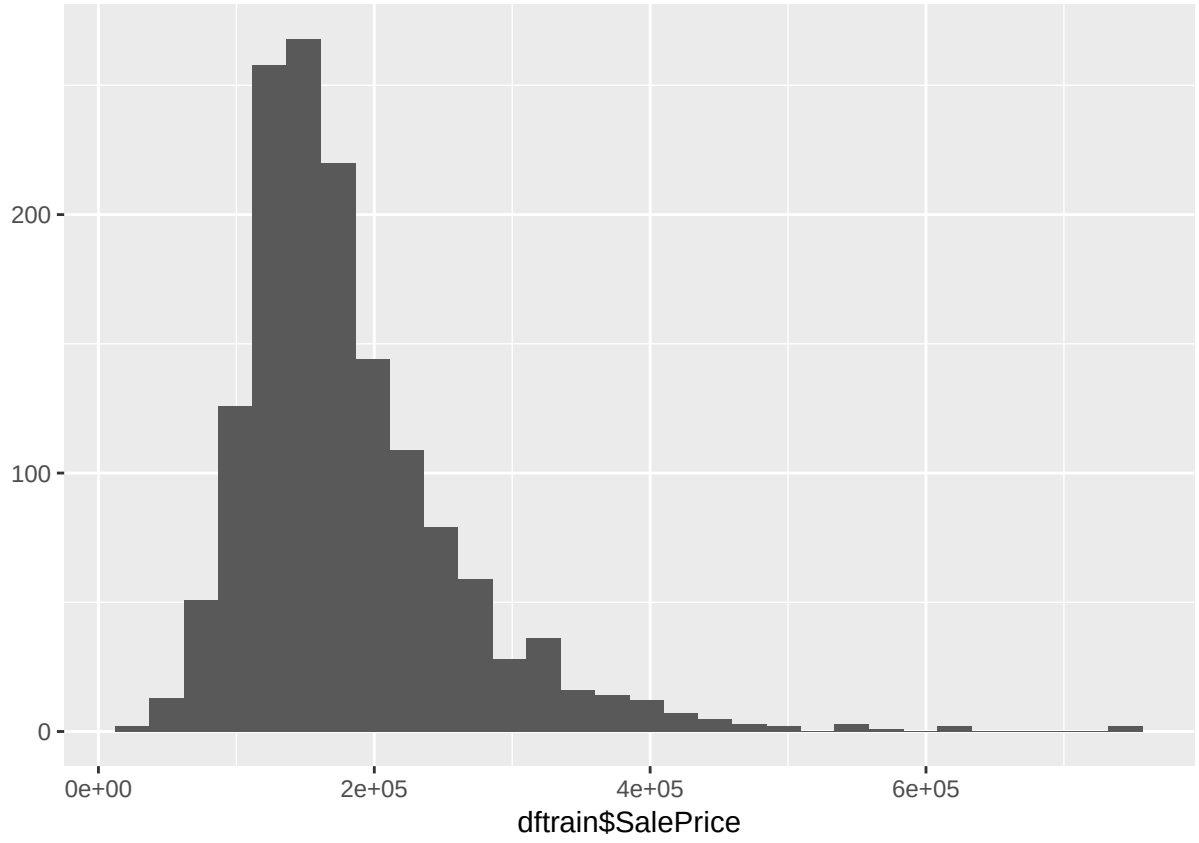
```
## [1] 80
```

```r
summary(dftrain$SalePrice)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   34900  129975  163000  180921  214000  755000
```

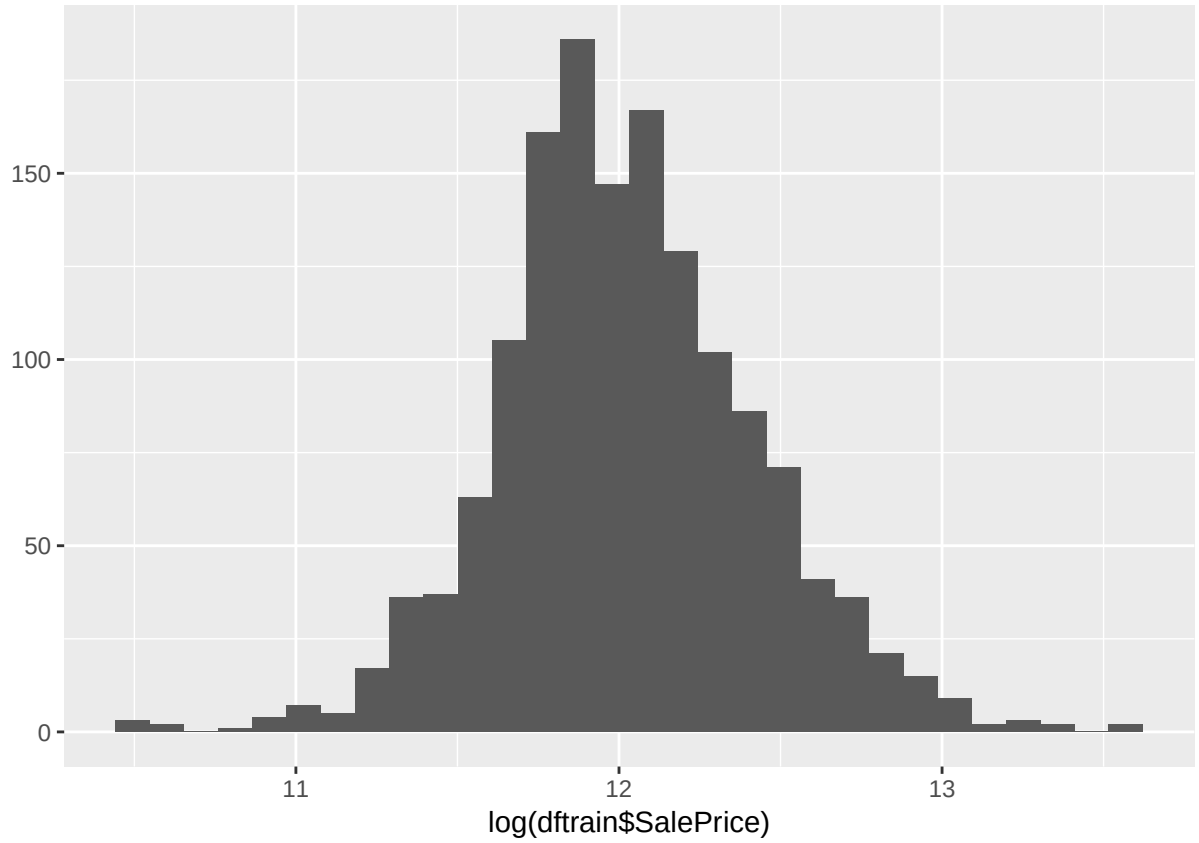
```r
#skewed data
qplot(dftrain$SalePrice)
```

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




```r
#Using normalized data will be better
qplot(log(dftrain$SalePrice))
```

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




```
#everything is numeric or factro
str(dftrain)
```

```
## 'data.frame':    1460 obs. of  81 variables:
##  $ Id           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ MSSubClass   : int  60 20 60 70 60 50 20 60 50 190 ...
##  $ MSZoning     : Factor w/ 5 levels "C (all)","FV",..: 4 4 4 4 4 4 4 4 5 4 ...
##  $ LotFrontage  : int  65 80 68 60 84 85 75 NA 51 50 ...
##  $ LotArea      : int  8450 9600 11250 9550 14260 14115 10084 10382 6120 7420 ...
##  $ Street       : Factor w/ 2 levels "Grvl","Pave": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Alley        : Factor w/ 2 levels "Grvl","Pave": NA NA NA NA NA NA NA NA NA NA ...
##  $ LotShape     : Factor w/ 4 levels "IR1","IR2","IR3",..: 4 4 1 1 1 1 4 1 4 4 ...
##  $ LandContour  : Factor w/ 4 levels "Bnk","HLS","Low",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ Utilities    : Factor w/ 2 levels "AllPub","NoSeWa": 1 1 1 1 1 1 1 1 1 1 ...
##  $ LotConfig    : Factor w/ 5 levels "Corner","CulDSac",..: 5 3 5 1 3 5 5 1 5 1 ...
##  $ LandSlope    : Factor w/ 3 levels "Gtl","Mod","Sev": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Neighborhood : Factor w/ 25 levels "Blmngtn","Blueste",..: 6 25 6 7 14 12 21 17 18 4 ...
##  $ Condition1   : Factor w/ 9 levels "Artery","Feedr",..: 3 2 3 3 3 3 3 5 1 1 ...
##  $ Condition2   : Factor w/ 8 levels "Artery","Feedr",..: 3 3 3 3 3 3 3 3 3 1 ...
##  $ BldgType     : Factor w/ 5 levels "1Fam","2fmCon",..: 1 1 1 1 1 1 1 1 1 2 ...
##  $ HouseStyle   : Factor w/ 8 levels "1.5Fin","1.5Unf",..: 6 3 6 6 6 1 3 6 1 2 ...
##  $ OverallQual  : int  7 6 7 7 8 5 8 7 7 5 ...
##  $ OverallCond  : int  5 8 5 5 5 5 5 6 5 6 ...
##  $ YearBuilt    : int  2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 ...
##  $ YearRemodAdd : int  2003 1976 2002 1970 2000 1995 2005 1973 1950 1950 ...
##  $ RoofStyle    : Factor w/ 6 levels "Flat","Gable",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ RoofMatl     : Factor w/ 8 levels "ClyTile","CompShg",..: 2 2 2 2 2 2 2 2 2 2 ...
```

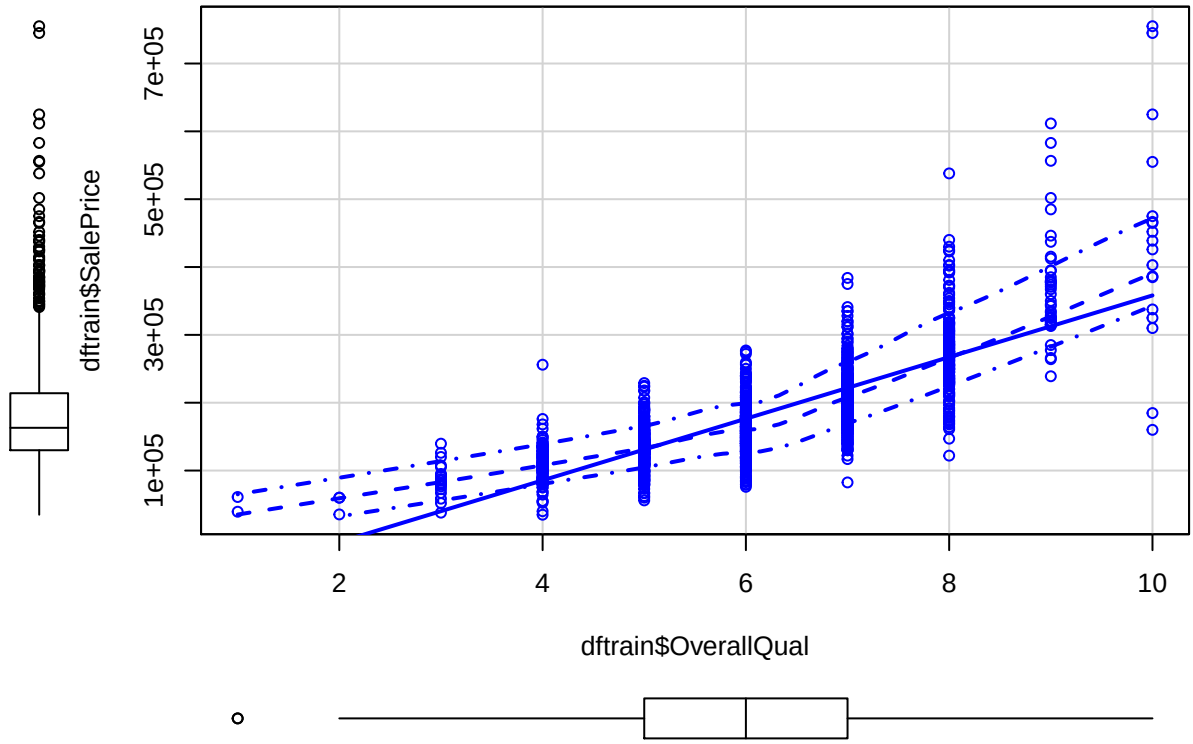
```
##  $ Exterior1st  : Factor w/ 15 levels "AsbShng","AsphShn",..: 13 9 13 14 13 13 13 7 4 9 ...
##  $ Exterior2nd  : Factor w/ 16 levels "AsbShng","AsphShn",..: 14 9 14 16 14 14 14 7 16 9 ...
##  $ MasVnrType   : Factor w/ 4 levels "BrkCmn","BrkFace",..: 2 3 2 3 2 3 4 4 3 3 ...
##  $ MasVnrArea   : int  196 0 162 0 350 0 186 240 0 0 ...
##  $ ExterQual    : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 4 3 4 3 4 3 4 4 4 ...
##  $ ExterCond    : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ Foundation   : Factor w/ 6 levels "BrkTil","CBlock",..: 3 2 3 1 3 6 3 2 1 1 ...
##  $ BsmtQual     : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 3 3 4 3 3 1 3 4 4 ...
##  $ BsmtCond     : Factor w/ 4 levels "Fa","Gd","Po",..: 4 4 4 2 4 4 4 4 4 4 ...
##  $ BsmtExposure : Factor w/ 4 levels "Av","Gd","Mn",..: 4 2 3 4 1 4 1 3 4 4 ...
##  $ BsmtFinType1 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 3 1 3 1 3 3 3 1 6 3 ...
##  $ BsmtFinSF1   : int  706 978 486 216 655 732 1369 859 0 851 ...
##  $ BsmtFinType2 : Factor w/ 6 levels "ALQ","BLQ","GLQ",..: 6 6 6 6 6 6 6 2 6 6 ...
##  $ BsmtFinSF2   : int  0 0 0 0 0 0 0 32 0 0 ...
##  $ BsmtUnfSF    : int  150 284 434 540 490 64 317 216 952 140 ...
##  $ TotalBsmtSF  : int  856 1262 920 756 1145 796 1686 1107 952 991 ...
##  $ Heating      : Factor w/ 6 levels "Floor","GasA",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ HeatingQC    : Factor w/ 5 levels "Ex","Fa","Gd",..: 1 1 1 3 1 1 1 1 3 1 ...
##  $ CentralAir   : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Electrical   : Factor w/ 5 levels "FuseA","FuseF",..: 5 5 5 5 5 5 5 5 2 5 ...
##  $ X1stFlrSF    : int  856 1262 920 961 1145 796 1694 1107 1022 1077 ...
##  $ X2ndFlrSF    : int  854 0 866 756 1053 566 0 983 752 0 ...
##  $ LowQualFinSF : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ GrLivArea    : int  1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 ...
##  $ BsmtFullBath : int  1 0 1 1 1 1 1 1 0 1 ...
##  $ BsmtHalfBath : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ FullBath     : int  2 2 2 1 2 1 2 2 2 1 ...
##  $ HalfBath     : int  1 0 1 0 1 1 0 1 0 0 ...
##  $ BedroomAbvGr : int  3 3 3 3 4 1 3 3 2 2 ...
##  $ KitchenAbvGr : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ KitchenQual  : Factor w/ 4 levels "Ex","Fa","Gd",..: 3 4 3 3 3 4 3 4 4 4 ...
##  $ TotRmsAbvGrd : int  8 6 6 7 9 5 7 7 8 5 ...
##  $ Functional   : Factor w/ 7 levels "Maj1","Maj2",..: 7 7 7 7 7 7 7 7 3 7 ...
##  $ Fireplaces   : int  0 1 1 1 1 0 1 2 2 2 ...
##  $ FireplaceQu  : Factor w/ 5 levels "Ex","Fa","Gd",..: NA 5 5 3 5 NA 3 5 5 5 ...
##  $ GarageType   : Factor w/ 6 levels "2Types","Attchd",..: 2 2 2 6 2 2 2 2 6 2 ...
##  $ GarageYrBlt  : int  2003 1976 2001 1998 2000 1993 2004 1973 1931 1939 ...
##  $ GarageFinish : Factor w/ 3 levels "Fin","RFn","Unf": 2 2 2 3 2 3 2 2 3 2 ...
##  $ GarageCars   : int  2 2 2 3 3 2 2 2 2 1 ...
##  $ GarageArea   : int  548 460 608 642 836 480 636 484 468 205 ...
##  $ GarageQual   : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 2 3 ...
##  $ GarageCond   : Factor w/ 5 levels "Ex","Fa","Gd",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ PavedDrive   : Factor w/ 3 levels "N","P","Y": 3 3 3 3 3 3 3 3 3 3 ...
##  $ WoodDeckSF   : int  0 298 0 0 192 40 255 235 90 0 ...
##  $ OpenPorchSF  : int  61 0 42 35 84 30 57 204 0 4 ...
##  $ EnclosedPorch: int  0 0 0 272 0 0 0 228 205 0 ...
##  $ X3SsnPorch   : int  0 0 0 0 0 320 0 0 0 0 ...
##  $ ScreenPorch  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolArea     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PoolQC       : Factor w/ 3 levels "Ex","Fa","Gd": NA NA NA NA NA NA NA NA NA NA ...
##  $ Fence        : Factor w/ 4 levels "GdPrv","GdWo",..: NA NA NA NA NA 3 NA NA NA NA ...
##  $ MiscFeature  : Factor w/ 4 levels "Gar2","Othr",..: NA NA NA NA NA 3 NA 3 NA NA ...
##  $ MiscVal      : int  0 0 0 0 0 700 0 350 0 0 ...
##  $ MoSold       : int  2 5 9 2 12 10 8 11 4 1 ...
```

```
##  $ YrSold       : int  2008 2007 2008 2006 2008 2009 2007 2009 2008 2008 ...
##  $ SaleType     : Factor w/ 9 levels "COD","Con","ConLD",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ SaleCondition: Factor w/ 6 levels "Abnorml","AdjLand",..: 5 5 5 1 5 5 5 5 1 5 ...
##  $ SalePrice    : int  208500 181500 223500 140000 250000 143000 307000 200000 129900 118000 ...
```

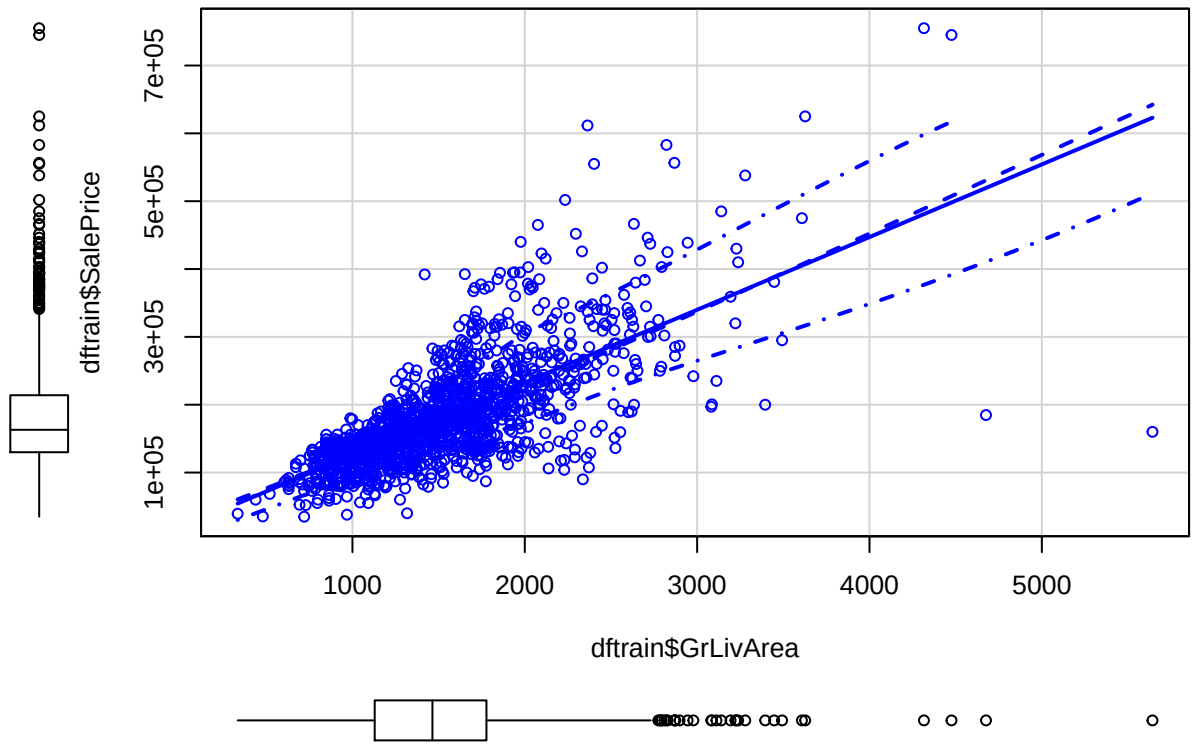
```
#let look at na's
nacols <- as.data.frame(as.table(colSums(is.na(dftrain))))
nacols <- filter(nacols, Freq > 0) %>% arrange(-Freq)



dftrain[,validcols] <-sapply(dftrain[,validcols],fct_explicit_na)
dftrain[,validcols] <- as.data.frame(sapply(dftrain[,validcols], factor))
dftrain$MasVnrType <-  na.approx(dftrain$MasVnrType)
#replacing with most common value
dftrain$Electrical[is.na(dftrain$Electrical)] <- "SBrkr"

dftrain$LotFrontage <- na.approx(dftrain$LotFrontage)
dftrain$GarageYrBlt <- na.approx(dftrain$GarageYrBlt)
dftrain$MasVnrArea <- na.approx(dftrain$MasVnrArea)

#all missing values are taken care off

#Lets Look @ Correlation
numeric <- dftrain[,sapply(dftrain,is.numeric)]
corr <- as.data.frame(as.table(cor(numeric$SalePrice, numeric))) %>% arrange(-Freq)
filter(corr,abs(corr$Freq) >= .5)
```

```
##    Var1         Var2      Freq
## 1     A    SalePrice 1.0000000
## 2     A  OverallQual 0.7909816
## 3     A    GrLivArea 0.7086245
## 4     A   GarageCars 0.6404092
## 5     A   GarageArea 0.6234314
## 6     A  TotalBsmtSF 0.6135806
## 7     A    X1stFlrSF 0.6058522
## 8     A     FullBath 0.5606638
## 9     A TotRmsAbvGrd 0.5337232
## 10    A    YearBuilt 0.5228973
## 11    A YearRemodAdd 0.5071010
```

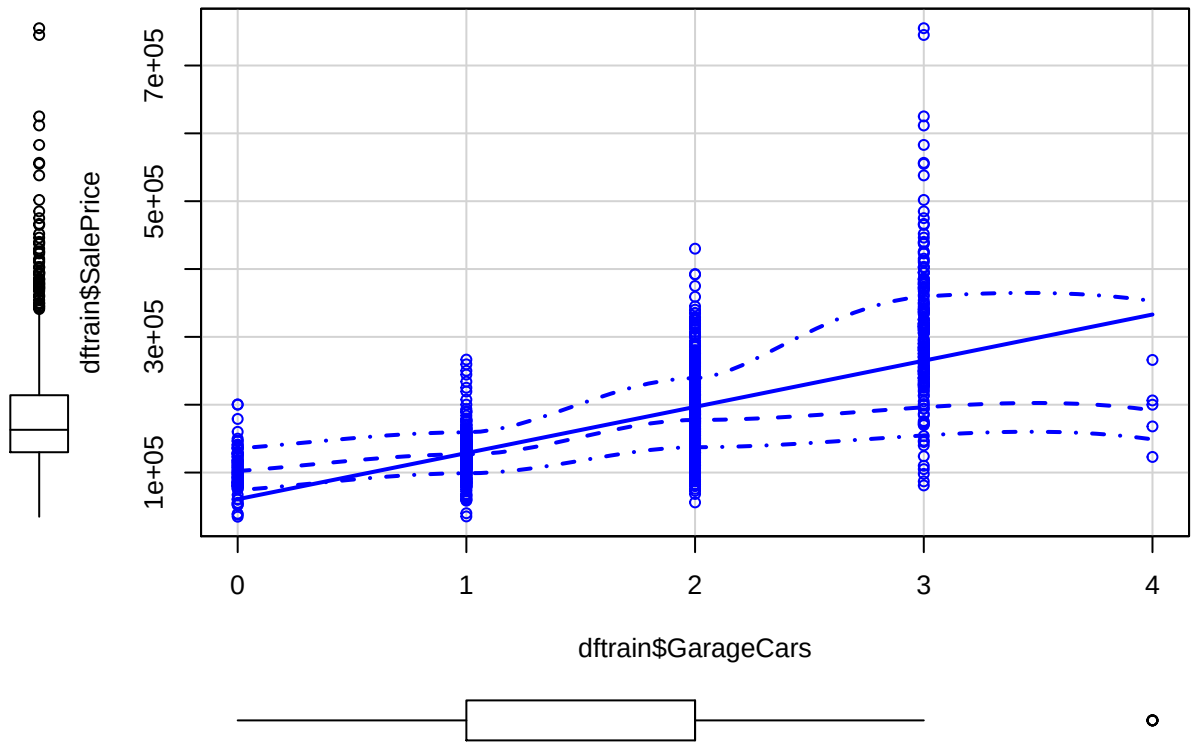
```
scatterplot(dftrain$SalePrice ~ dftrain$OverallQual)
```



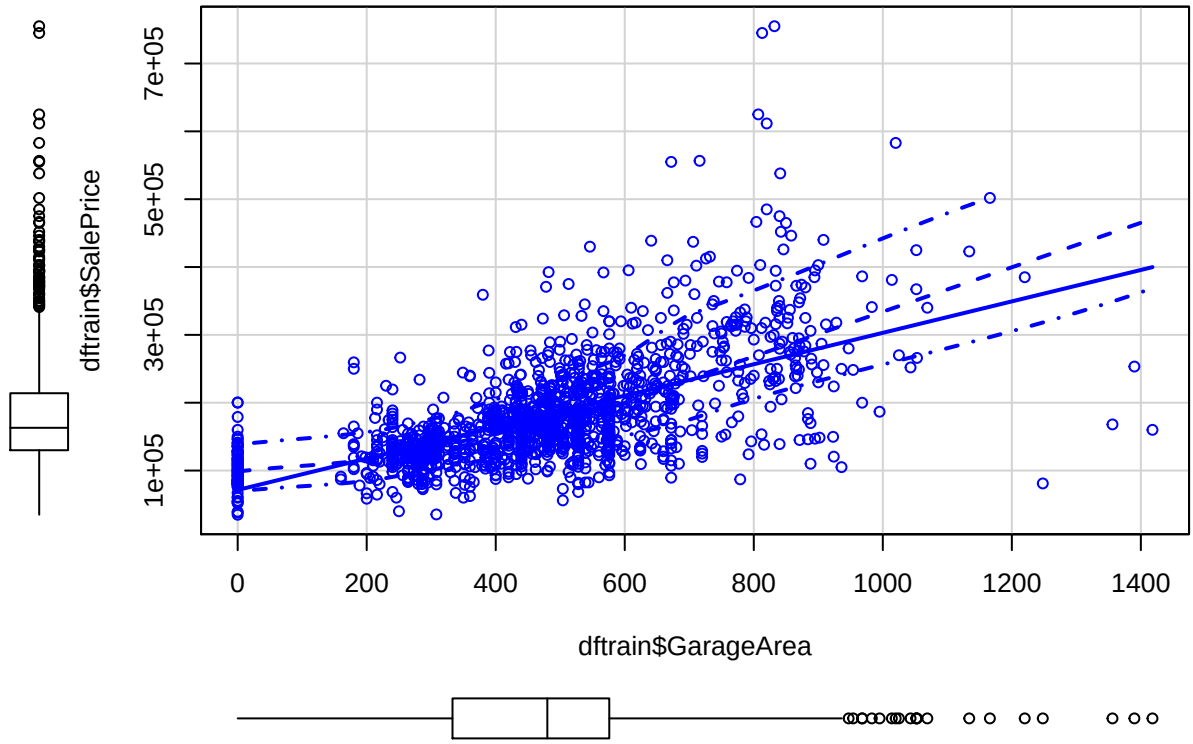

```
scatterplot(dftrain$SalePrice ~ dftrain$GrLivArea)
```



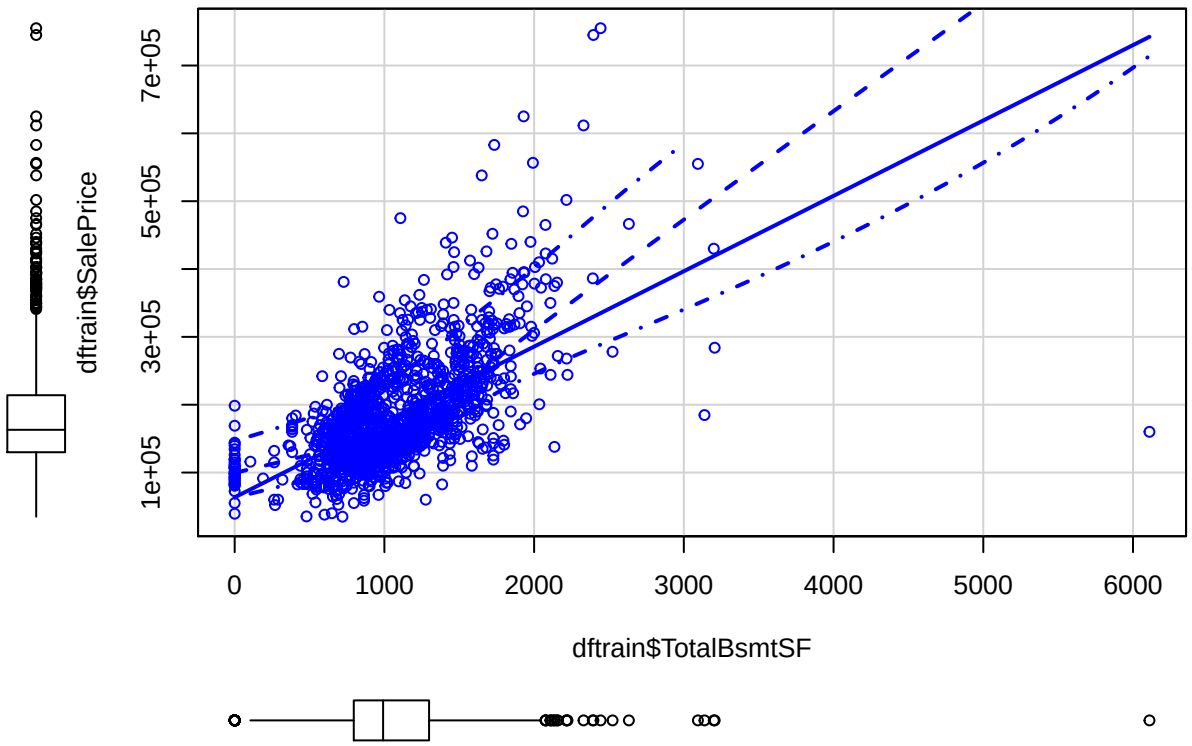

```
scatterplot(dftrain$SalePrice ~ dftrain$GarageCars)
```




```
scatterplot(dftrain$SalePrice ~ dftrain$GarageArea)
```




```
scatterplot(dftrain$SalePrice ~ dftrain$TotalBsmtSF)
```




```
mymodel <- lm(SalePrice ~ ., data = dftrain)
summary(mymodel)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = dftrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -177139   -9169       0    9508  177139
##
## Coefficients: (8 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.637e+05  1.045e+06  -0.252 0.800823
## Id                   1.120e+00  1.546e+00   0.725 0.468648
## MSSubClass          -5.111e+01  8.243e+01  -0.620 0.535397
## MSZoningFV           3.266e+04  1.190e+04   2.745 0.006150 **
## MSZoningRH           2.215e+04  1.187e+04   1.866 0.062219 .
## MSZoningRL           2.492e+04  1.020e+04   2.444 0.014672 *
## MSZoningRM           2.157e+04  9.565e+03   2.255 0.024296 *
## LotFrontage          1.145e+01  3.962e+01   0.289 0.772702
## LotArea              7.195e-01  1.087e-01   6.620 5.40e-11 ***
## StreetPave           3.319e+04  1.216e+04   2.731 0.006415 **
## AlleyGrvl            1.311e+03  4.206e+03   0.312 0.755303
## AlleyPave            6.587e+02  4.745e+03   0.139 0.889622
## LotShapeIR2          5.125e+03  4.198e+03   1.221 0.222331
```

```
## LotShapeIR3           6.066e+03  8.808e+03   0.689 0.491160
## LotShapeReg           1.804e+03  1.596e+03   1.130 0.258521
## LandContourHLS        7.569e+03  5.111e+03   1.481 0.138866
## LandContourLow       -1.157e+04  6.375e+03  -1.815 0.069820 .
## LandContourLvl        5.423e+03  3.691e+03   1.469 0.142036
## UtilitiesNoSeWa      -3.734e+04  2.629e+04  -1.420 0.155851
## LotConfigCulDSac      7.855e+03  3.256e+03   2.413 0.015979 *
## LotConfigFR2         -7.804e+03  4.011e+03  -1.946 0.051893 .
## LotConfigFR3         -1.737e+04  1.252e+04  -1.387 0.165672
## LotConfigInside      -1.510e+03  1.778e+03  -0.849 0.395974
## LandSlopeMod          7.449e+03  3.970e+03   1.876 0.060859 .
## LandSlopeSev         -4.221e+04  1.138e+04  -3.708 0.000219 ***
## NeighborhoodBlueste   7.306e+03  1.917e+04   0.381 0.703137
## NeighborhoodBrDale   -2.371e+03  1.092e+04  -0.217 0.828157
## NeighborhoodBrkSide  -5.787e+03  9.416e+03  -0.615 0.538955
## NeighborhoodClearCr  -1.460e+04  9.146e+03  -1.596 0.110696
## NeighborhoodCollgCr  -1.012e+04  7.214e+03  -1.403 0.160882
## NeighborhoodCrawfor   1.178e+04  8.510e+03   1.384 0.166630
## NeighborhoodEdwards  -2.148e+04  7.952e+03  -2.702 0.006993 **
## NeighborhoodGilbert  -1.147e+04  7.611e+03  -1.508 0.131905
## NeighborhoodIDOTRR   -1.203e+04  1.068e+04  -1.126 0.260234
## NeighborhoodMeadowV  -6.580e+03  1.108e+04  -0.594 0.552671
## NeighborhoodMitchel  -2.095e+04  8.115e+03  -2.582 0.009953 **
## NeighborhoodNAmes    -1.734e+04  7.788e+03  -2.227 0.026129 *
## NeighborhoodNoRidge   2.572e+04  8.348e+03   3.081 0.002110 **
## NeighborhoodNPkVill   1.339e+04  1.394e+04   0.961 0.336901
## NeighborhoodNridgHt   1.832e+04  7.469e+03   2.452 0.014329 *
## NeighborhoodNWAmes   -1.755e+04  7.982e+03  -2.199 0.028086 *
## NeighborhoodOldTown  -1.431e+04  9.589e+03  -1.492 0.135989
## NeighborhoodSawyer   -1.116e+04  8.051e+03  -1.387 0.165846
## NeighborhoodSawyerW  -2.786e+03  7.730e+03  -0.360 0.718623
## NeighborhoodSomerst  -2.599e+03  8.953e+03  -0.290 0.771603
## NeighborhoodStoneBr   3.952e+04  8.201e+03   4.819 1.63e-06 ***
## NeighborhoodSWISU    -8.824e+03  9.663e+03  -0.913 0.361352
## NeighborhoodTimber   -9.756e+03  8.073e+03  -1.209 0.227061
## NeighborhoodVeenker  -1.888e+02  1.042e+04  -0.018 0.985555
## Condition1Feedr       7.156e+03  5.015e+03   1.427 0.153810
## Condition1Norm        1.633e+04  4.182e+03   3.904 9.99e-05 ***
## Condition1PosA        1.022e+04  9.965e+03   1.026 0.305180
## Condition1PosN        1.493e+04  7.425e+03   2.011 0.044505 *
## Condition1RRAe       -1.542e+04  9.047e+03  -1.705 0.088484 .
## Condition1RRAn        1.295e+04  6.937e+03   1.867 0.062083 .
## Condition1RRNe       -3.388e+03  1.744e+04  -0.194 0.845990
## Condition1RRNn        1.168e+04  1.281e+04   0.912 0.362203
## Condition2Feedr      -5.846e+03  2.335e+04  -0.250 0.802362
## Condition2Norm       -1.058e+04  2.022e+04  -0.523 0.600952
## Condition2PosA        4.318e+04  3.691e+04   1.170 0.242276
## Condition2PosN       -2.382e+05  2.755e+04  -8.645  < 2e-16 ***
## Condition2RRAe       -1.303e+05  6.482e+04  -2.010 0.044604 *
## Condition2RRAn       -2.397e+04  3.141e+04  -0.763 0.445463
## Condition2RRNn       -2.724e+03  2.704e+04  -0.101 0.919775
## BldgType2fmCon       -3.712e+03  1.243e+04  -0.299 0.765316
## BldgTypeDuplex       -7.345e+03  7.372e+03  -0.996 0.319316
## BldgTypeTwnhs        -1.993e+04  9.897e+03  -2.013 0.044291 *
```

```
## BldgTypeTwnhsE       -1.571e+04  8.990e+03  -1.747 0.080844 .
## HouseStyle1.5Unf      1.219e+04  7.940e+03   1.535 0.125012
## HouseStyle1Story      5.461e+03  4.370e+03   1.250 0.211605
## HouseStyle2.5Fin     -1.723e+04  1.232e+04  -1.398 0.162373
## HouseStyle2.5Unf     -9.624e+03  9.210e+03  -1.045 0.296258
## HouseStyle2Story     -5.987e+03  3.483e+03  -1.719 0.085839 .
## HouseStyleSFoyer      1.169e+03  6.247e+03   0.187 0.851635
## HouseStyleSLvl        3.881e+03  5.529e+03   0.702 0.482870
## OverallQual           6.813e+03  1.008e+03   6.757 2.19e-11 ***
## OverallCond           5.792e+03  8.705e+02   6.654 4.32e-11 ***
## YearBuilt             3.201e+02  7.644e+01   4.187 3.03e-05 ***
## YearRemodAdd          1.055e+02  5.565e+01   1.895 0.058332 .
## RoofStyleGable        9.492e+03  1.843e+04   0.515 0.606585
## RoofStyleGambrel      1.282e+04  2.016e+04   0.636 0.524898
## RoofStyleHip          9.253e+03  1.850e+04   0.500 0.617022
## RoofStyleMansard      2.041e+04  2.137e+04   0.955 0.339697
## RoofStyleShed         1.001e+05  3.449e+04   2.901 0.003783 **
## RoofMatlCompShg       5.701e+05  5.251e+04  10.857  < 2e-16 ***
## RoofMatlMembran       6.666e+05  6.243e+04  10.678  < 2e-16 ***
## RoofMatlMetal         6.353e+05  6.206e+04  10.237  < 2e-16 ***
## RoofMatlRoll          5.572e+05  5.814e+04   9.583  < 2e-16 ***
## RoofMatlTar&Grv       5.722e+05  5.638e+04  10.149  < 2e-16 ***
## RoofMatlWdShake       5.616e+05  5.478e+04  10.253  < 2e-16 ***
## RoofMatlWdShngl       6.249e+05  5.344e+04  11.694  < 2e-16 ***
## Exterior1stAsphShn   -2.623e+04  3.308e+04  -0.793 0.427906
## Exterior1stBrkComm   -3.836e+03  2.773e+04  -0.138 0.889977
## Exterior1stBrkFace    7.572e+03  1.274e+04   0.594 0.552447
## Exterior1stCBlock    -1.479e+04  2.722e+04  -0.543 0.587035
## Exterior1stCemntBd   -1.254e+04  1.897e+04  -0.661 0.508515
## Exterior1stHdBoard   -1.332e+04  1.291e+04  -1.032 0.302418
## Exterior1stImStucc   -2.322e+04  2.812e+04  -0.826 0.409113
## Exterior1stMetalSd   -6.671e+03  1.457e+04  -0.458 0.647075
## Exterior1stPlywood   -1.415e+04  1.274e+04  -1.111 0.266997
## Exterior1stStone     -1.454e+03  2.426e+04  -0.060 0.952213
## Exterior1stStucco    -7.620e+03  1.407e+04  -0.542 0.588144
## Exterior1stVinylSd   -1.474e+04  1.330e+04  -1.108 0.267907
## Exterior1stWd Sdng   -1.424e+04  1.236e+04  -1.152 0.249387
## Exterior1stWdShing   -9.940e+03  1.334e+04  -0.745 0.456477
## Exterior2ndAsphShn    1.231e+04  2.216e+04   0.556 0.578560
## Exterior2ndBrk Cmn    5.394e+03  2.003e+04   0.269 0.787726
## Exterior2ndBrkFace    4.075e+03  1.319e+04   0.309 0.757417
## Exterior2ndCBlock            NA         NA      NA       NA
## Exterior2ndCmentBd    1.282e+04  1.867e+04   0.686 0.492573
## Exterior2ndHdBoard    8.344e+03  1.240e+04   0.673 0.501006
## Exterior2ndImStucc    1.732e+04  1.432e+04   1.209 0.226755
## Exterior2ndMetalSd    6.439e+03  1.418e+04   0.454 0.649870
## Exterior2ndOther     -1.753e+04  2.703e+04  -0.648 0.516897
## Exterior2ndPlywood    6.778e+03  1.203e+04   0.563 0.573419
## Exterior2ndStone     -1.139e+04  1.711e+04  -0.665 0.505925
## Exterior2ndStucco     5.917e+03  1.360e+04   0.435 0.663565
## Exterior2ndVinylSd    1.362e+04  1.278e+04   1.066 0.286832
## Exterior2ndWd Sdng    1.225e+04  1.192e+04   1.027 0.304501
## Exterior2ndWd Shng    5.657e+03  1.244e+04   0.455 0.649322
## MasVnrType            2.706e+03  1.279e+03   2.116 0.034574 *
```

```
## MasVnrArea           2.012e+01  5.033e+00   3.997 6.81e-05 ***
## ExterQualFa         -7.030e+03  1.107e+04  -0.635 0.525509
## ExterQualGd         -2.069e+04  4.762e+03  -4.345 1.51e-05 ***
## ExterQualTA         -1.975e+04  5.293e+03  -3.731 0.000200 ***
## ExterCondFa         -2.409e+03  1.804e+04  -0.134 0.893795
## ExterCondGd         -7.025e+03  1.720e+04  -0.408 0.683079
## ExterCondPo          8.649e+03  3.164e+04   0.273 0.784588
## ExterCondTA         -4.017e+03  1.717e+04  -0.234 0.815049
## FoundationCBlock     3.017e+03  3.165e+03   0.953 0.340617
## FoundationPConc      4.044e+03  3.411e+03   1.186 0.236021
## FoundationSlab      -7.249e+03  1.002e+04  -0.723 0.469696
## FoundationStone      9.551e+03  1.138e+04   0.839 0.401591
## FoundationWood      -2.673e+04  1.473e+04  -1.815 0.069826 .
## BsmtQualEx          -3.241e+04  3.616e+04  -0.896 0.370185
## BsmtQualFa          -4.419e+04  3.595e+04  -1.229 0.219255
## BsmtQualGd          -5.055e+04  3.593e+04  -1.407 0.159732
## BsmtQualTA          -4.693e+04  3.583e+04  -1.310 0.190570
## BsmtCondFa          -2.686e+03  4.232e+03  -0.635 0.525712
## BsmtCondGd          -2.733e+03  3.244e+03  -0.842 0.399690
## BsmtCondPo           6.626e+04  2.995e+04   2.212 0.027145 *
## BsmtCondTA                  NA         NA      NA       NA
## BsmtExposureAv       1.141e+04  2.293e+04   0.497 0.618977
## BsmtExposureGd       2.567e+04  2.302e+04   1.115 0.265141
## BsmtExposureMn       7.880e+03  2.300e+04   0.343 0.731962
## BsmtExposureNo       6.265e+03  2.289e+04   0.274 0.784348
## BsmtFinType1ALQ     -2.655e+03  2.911e+03  -0.912 0.361868
## BsmtFinType1BLQ      1.864e+02  3.118e+03   0.060 0.952345
## BsmtFinType1GLQ      2.939e+03  2.706e+03   1.086 0.277712
## BsmtFinType1LwQ     -6.044e+03  3.763e+03  -1.606 0.108517
## BsmtFinType1Rec     -2.585e+03  3.156e+03  -0.819 0.412896
## BsmtFinType1Unf             NA         NA      NA       NA
## BsmtFinSF1           3.853e+01  5.320e+00   7.243 7.80e-13 ***
## BsmtFinType2ALQ      2.861e+04  2.493e+04   1.148 0.251347
## BsmtFinType2BLQ      1.559e+04  2.466e+04   0.632 0.527303
## BsmtFinType2GLQ      2.606e+04  2.545e+04   1.024 0.306014
## BsmtFinType2LwQ      1.430e+04  2.469e+04   0.579 0.562637
## BsmtFinType2Rec      1.825e+04  2.463e+04   0.741 0.458951
## BsmtFinType2Unf      2.026e+04  2.457e+04   0.824 0.409903
## BsmtFinSF2           3.169e+01  9.036e+00   3.507 0.000470 ***
## BsmtUnfSF            2.106e+01  4.874e+00   4.321 1.68e-05 ***
## TotalBsmtSF                 NA         NA      NA       NA
## HeatingGasA          9.939e+03  2.554e+04   0.389 0.697258
## HeatingGasW          7.519e+03  2.635e+04   0.285 0.775371
## HeatingGrav          1.686e+03  2.802e+04   0.060 0.952018
## HeatingOthW         -1.100e+04  3.142e+04  -0.350 0.726287
## HeatingWall          2.308e+04  2.968e+04   0.777 0.437017
## HeatingQCFa          5.855e+02  4.701e+03   0.125 0.900914
## HeatingQCGd         -3.900e+03  2.061e+03  -1.892 0.058726 .
## HeatingQCPo          2.043e+03  2.650e+04   0.077 0.938572
## HeatingQCTA         -3.229e+03  2.063e+03  -1.565 0.117929
## CentralAirY         -2.206e+02  3.863e+03  -0.057 0.954467
## ElectricalFuseF      5.145e+01  5.738e+03   0.009 0.992848
## ElectricalFuseP     -8.581e+03  1.859e+04  -0.462 0.644388
## ElectricalMix       -4.451e+04  4.441e+04  -1.002 0.316499
```

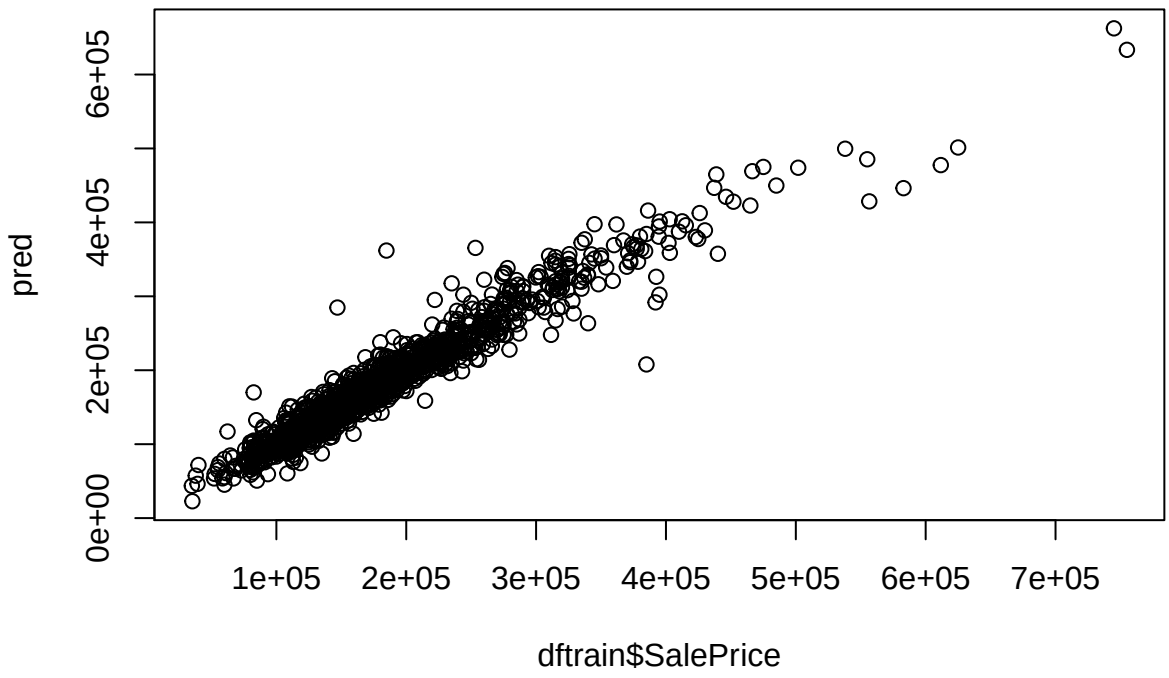
```
## ElectricalSBrkr      -2.127e+03  2.943e+03  -0.723 0.469984
## X1stFlrSF             4.433e+01  5.635e+00   7.867 8.00e-15 ***
## X2ndFlrSF             6.252e+01  5.657e+00  11.051  < 2e-16 ***
## LowQualFinSF         -2.835e+00  1.902e+01  -0.149 0.881544
## GrLivArea                    NA         NA      NA       NA
## BsmtFullBath          1.522e+03  1.974e+03   0.771 0.440857
## BsmtHalfBath         -4.899e+02  3.016e+03  -0.162 0.870971
## FullBath              3.651e+03  2.196e+03   1.663 0.096605 .
## HalfBath              1.871e+03  2.088e+03   0.896 0.370598
## BedroomAbvGr         -3.684e+03  1.363e+03  -2.702 0.006984 **
## KitchenAbvGr         -1.356e+04  5.666e+03  -2.394 0.016834 *
## KitchenQualFa        -1.964e+04  6.189e+03  -3.172 0.001549 **
## KitchenQualGd        -2.352e+04  3.473e+03  -6.773 1.96e-11 ***
## KitchenQualTA        -2.255e+04  3.922e+03  -5.750 1.13e-08 ***
## TotRmsAbvGrd          1.798e+03  9.524e+02   1.888 0.059260 .
## FunctionalMaj2       -8.908e+02  1.432e+04  -0.062 0.950427
## FunctionalMin1        7.300e+03  8.559e+03   0.853 0.393897
## FunctionalMin2        8.936e+03  8.586e+03   1.041 0.298234
## FunctionalMod        -5.054e+03  1.051e+04  -0.481 0.630719
## FunctionalSev        -4.272e+04  2.897e+04  -1.474 0.140664
## FunctionalTyp         1.842e+04  7.425e+03   2.481 0.013221 *
## Fireplaces            6.174e+03  2.550e+03   2.422 0.015600 *
## FireplaceQuEx        -8.729e+03  6.211e+03  -1.405 0.160146
## FireplaceQuFa        -9.463e+03  5.324e+03  -1.777 0.075756 .
## FireplaceQuGd        -5.974e+03  3.409e+03  -1.753 0.079907 .
## FireplaceQuPo         3.626e+03  6.375e+03   0.569 0.569618
## FireplaceQuTA        -5.021e+03  3.562e+03  -1.410 0.158904
## GarageType2Types     -2.959e+04  1.212e+04  -2.442 0.014748 *
## GarageTypeAttchd     -9.589e+03  4.670e+03  -2.053 0.040280 *
## GarageTypeBasment    -5.228e+03  7.819e+03  -0.669 0.503825
## GarageTypeBuiltIn    -9.819e+03  5.571e+03  -1.762 0.078256 .
## GarageTypeCarPort    -4.677e+03  1.054e+04  -0.444 0.657170
## GarageTypeDetchd     -6.602e+03  4.637e+03  -1.424 0.154816
## GarageYrBlt          -2.086e+01  5.835e+01  -0.357 0.720829
## GarageFinishFin       6.560e+02  2.419e+03   0.271 0.786287
## GarageFinishRFn      -1.772e+03  2.151e+03  -0.824 0.410220
## GarageFinishUnf              NA         NA      NA       NA
## GarageCars            3.990e+03  2.266e+03   1.761 0.078535 .
## GarageArea            1.826e+01  7.847e+00   2.327 0.020123 *
## GarageQualEx          1.192e+05  2.980e+04   3.999 6.74e-05 ***
## GarageQualFa         -6.064e+03  4.850e+03  -1.250 0.211464
## GarageQualGd         -1.230e+03  7.676e+03  -0.160 0.872757
## GarageQualPo         -2.287e+04  2.388e+04  -0.957 0.338575
## GarageQualTA                 NA         NA      NA       NA
## GarageCondEx         -1.135e+05  3.440e+04  -3.298 0.001003 **
## GarageCondFa         -1.841e+03  5.398e+03  -0.341 0.733169
## GarageCondGd         -2.795e+03  9.202e+03  -0.304 0.761408
## GarageCondPo          3.998e+03  1.386e+04   0.288 0.773023
## GarageCondTA                 NA         NA      NA       NA
## PavedDriveP          -3.430e+03  5.540e+03  -0.619 0.535994
## PavedDriveY          -2.380e+02  3.452e+03  -0.069 0.945046
## WoodDeckSF            1.513e+01  5.863e+00   2.581 0.009964 **
## OpenPorchSF           4.665e-01  1.154e+01   0.040 0.967765
## EnclosedPorch         3.354e+00  1.243e+01   0.270 0.787378
```

```
## X3SsnPorch            3.522e+01  2.232e+01   1.578 0.114840
## ScreenPorch           3.592e+01  1.247e+01   2.881 0.004031 **
## PoolArea              6.792e+02  2.267e+02   2.996 0.002793 **
## PoolQCEx             -2.509e+05  1.227e+05  -2.045 0.041069 *
## PoolQCFa             -4.089e+05  1.493e+05  -2.738 0.006267 **
## PoolQCGd             -3.799e+05  1.469e+05  -2.587 0.009806 **
## FenceGdPrv           -8.964e+03  3.650e+03  -2.456 0.014181 *
## FenceGdWo            -9.230e+02  3.565e+03  -0.259 0.795776
## FenceMnPrv            5.697e+02  2.260e+03   0.252 0.801079
## FenceMnWw            -5.754e+03  7.454e+03  -0.772 0.440311
## MiscFeatureGar2      -7.928e+02  9.695e+04  -0.008 0.993477
## MiscFeatureOthr       1.456e+04  2.016e+04   0.722 0.470269
## MiscFeatureShed       2.402e+03  5.685e+03   0.422 0.672791
## MiscFeatureTenC       3.033e+04  4.726e+04   0.642 0.521154
## MiscVal               8.497e-02  6.095e+00   0.014 0.988879
## MoSold               -4.734e+02  2.444e+02  -1.937 0.052960 .
## YrSold               -5.753e+02  5.122e+02  -1.123 0.261565
## SaleTypeCon           2.572e+04  1.751e+04   1.469 0.142163
## SaleTypeConLD         1.637e+04  9.665e+03   1.694 0.090503 .
## SaleTypeConLI         4.766e+03  1.154e+04   0.413 0.679668
## SaleTypeConLw         9.828e+02  1.213e+04   0.081 0.935419
## SaleTypeCWD           1.495e+04  1.283e+04   1.165 0.244094
## SaleTypeNew           2.102e+04  1.538e+04   1.366 0.172093
## SaleTypeOth           6.782e+03  1.448e+04   0.468 0.639637
## SaleTypeWD           -4.508e+02  4.166e+03  -0.108 0.913840
## SaleConditionAdjLand  9.341e+03  1.458e+04   0.641 0.521798
## SaleConditionAlloca   8.081e+02  8.824e+03   0.092 0.927044
## SaleConditionFamily   1.003e+03  6.076e+03   0.165 0.868855
## SaleConditionNormal   6.754e+03  2.897e+03   2.331 0.019909 *
## SaleConditionPartial -7.710e+01  1.481e+04  -0.005 0.995846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22560 on 1209 degrees of freedom
## Multiple R-squared:  0.9332, Adjusted R-squared:  0.9194
## F-statistic: 67.55 on 250 and 1209 DF,  p-value: < 2.2e-16
```

```
pred <- predict(mymodel, dftrain)
```

```
## Warning in predict.lm(mymodel, dftrain): prediction from a rank-deficient
## fit may be misleading
```

```
plot(dftrain$SalePrice, pred)
```



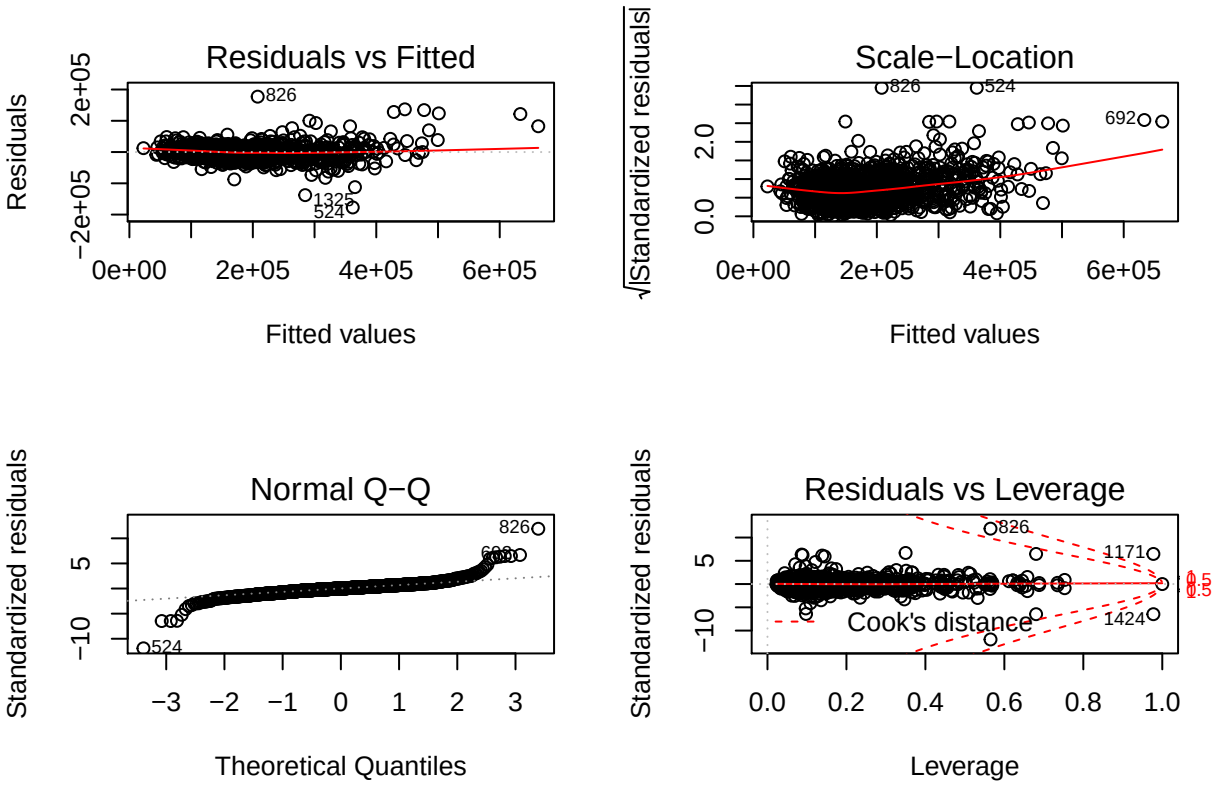

```
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(mymodel)
```

```
## Warning: not plotting observations with leverage one:
##   121, 186, 251, 272, 326, 333, 347, 376, 399, 584, 596, 811, 945, 949, 1004, 1012, 1188, 1231, 1271
```

```
## Warning: not plotting observations with leverage one:
##   121, 186, 251, 272, 326, 333, 347, 376, 399, 584, 596, 811, 945, 949, 1004, 1012, 1188, 1231, 1271
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
dftest[,validcols] <-sapply(dftest[,validcols],fct_explicit_na)
dftest[,validcols] <- as.data.frame(sapply(dftest[,validcols], factor))
dftest$MasVnrType <-  na.approx(dftest$MasVnrType)
dftest$LotFrontage <- na.approx(dftest$LotFrontage)
dftest$GarageYrBlt <- na.approx(dftest$GarageYrBlt)
dftest$MasVnrArea <- na.approx(dftest$MasVnrArea)


as.data.frame(as.table(colSums(is.na(dftest)))) %>% arrange(-Freq)
```

```
##             Var1 Freq
## 1       MSZoning    4
## 2      Utilities    2
## 3   BsmtFullBath    2
## 4   BsmtHalfBath    2
## 5     Functional    2
## 6    Exterior1st    1
## 7    Exterior2nd    1
## 8     BsmtFinSF1    1
## 9     BsmtFinSF2    1
## 10     BsmtUnfSF    1
## 11   TotalBsmtSF    1
## 12   KitchenQual    1
## 13    GarageCars    1
## 14    GarageArea    1
## 15      SaleType    1
```

```
## 16           Id    0
## 17   MSSubClass    0
## 18  LotFrontage    0
## 19      LotArea    0
## 20       Street    0
## 21        Alley    0
## 22     LotShape    0
## 23  LandContour    0
## 24    LotConfig    0
## 25    LandSlope    0
## 26 Neighborhood    0
## 27   Condition1    0
## 28   Condition2    0
## 29     BldgType    0
## 30   HouseStyle    0
## 31  OverallQual    0
## 32  OverallCond    0
## 33    YearBuilt    0
## 34 YearRemodAdd    0
## 35    RoofStyle    0
## 36     RoofMatl    0
## 37   MasVnrType    0
## 38   MasVnrArea    0
## 39    ExterQual    0
## 40    ExterCond    0
## 41   Foundation    0
## 42     BsmtQual    0
## 43     BsmtCond    0
## 44 BsmtExposure    0
## 45 BsmtFinType1    0
## 46 BsmtFinType2    0
## 47      Heating    0
## 48    HeatingQC    0
## 49   CentralAir    0
## 50   Electrical    0
## 51    X1stFlrSF    0
## 52    X2ndFlrSF    0
## 53 LowQualFinSF    0
## 54    GrLivArea    0
## 55     FullBath    0
## 56     HalfBath    0
## 57 BedroomAbvGr    0
## 58 KitchenAbvGr    0
## 59 TotRmsAbvGrd    0
## 60   Fireplaces    0
## 61  FireplaceQu    0
## 62   GarageType    0
## 63  GarageYrBlt    0
## 64 GarageFinish    0
## 65   GarageQual    0
## 66   GarageCond    0
## 67   PavedDrive    0
## 68   WoodDeckSF    0
## 69  OpenPorchSF    0
```

```
## 70 EnclosedPorch    0
## 71    X3SsnPorch    0
## 72   ScreenPorch    0
## 73      PoolArea    0
## 74        PoolQC    0
## 75         Fence    0
## 76   MiscFeature    0
## 77       MiscVal    0
## 78        MoSold    0
## 79        YrSold    0
## 80 SaleCondition    0
```

```
pred <- predict(mymodel, dftest)
```

```
## Warning in predict.lm(mymodel, dftest): prediction from a rank-deficient
## fit may be misleading
```