Code Pt. 2

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

library(tidyr)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.2

## corrplot 0.95 loaded

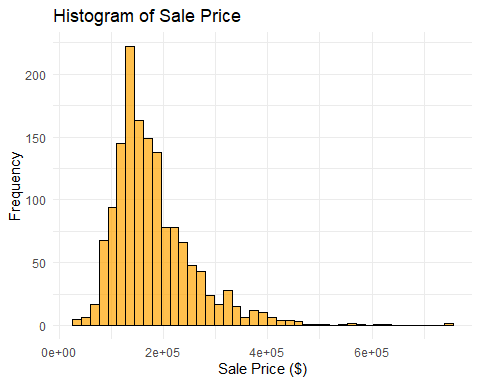
library(lattice)  
  
#load dataset  
trainData <- read.csv("/Users/katie/Documents/MSDS/Statistical Science/Unit 13/Project/train.csv")  
train <- trainData  
head(train)

## Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape  
## 1 1299 60 RL 313 63887 Pave <NA> IR3  
## 2 524 60 RL 130 40094 Pave <NA> IR1  
## 3 1183 60 RL 160 15623 Pave <NA> IR1  
## 4 692 60 RL 104 21535 Pave <NA> IR1  
## 5 1170 60 RL 118 35760 Pave <NA> IR1  
## 6 186 75 RM 90 22950 Pave <NA> IR2  
## LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2  
## 1 Bnk AllPub Corner Gtl Edwards Feedr Norm  
## 2 Bnk AllPub Inside Gtl Edwards PosN PosN  
## 3 Lvl AllPub Corner Gtl NoRidge Norm Norm  
## 4 Lvl AllPub Corner Gtl NoRidge Norm Norm  
## 5 Lvl AllPub CulDSac Gtl NoRidge Norm Norm  
## 6 Lvl AllPub Inside Gtl OldTown Artery Norm  
## BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle  
## 1 1Fam 2Story 10 5 2008 2008 Hip  
## 2 1Fam 2Story 10 5 2007 2008 Hip  
## 3 1Fam 2Story 10 5 1996 1996 Hip  
## 4 1Fam 2Story 10 6 1994 1995 Gable  
## 5 1Fam 2Story 10 5 1995 1996 Hip  
## 6 1Fam 2.5Fin 10 9 1892 1993 Gable  
## RoofMatl Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond  
## 1 ClyTile Stucco Stucco Stone 796 Ex TA  
## 2 CompShg CemntBd CmentBd Stone 762 Ex TA  
## 3 CompShg Wd Sdng ImStucc None 0 Gd TA  
## 4 WdShngl HdBoard HdBoard BrkFace 1170 Ex TA  
## 5 CompShg HdBoard HdBoard BrkFace 1378 Gd Gd  
## 6 WdShngl Wd Sdng Wd Sdng None 0 Gd Gd  
## Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1  
## 1 PConc Ex TA Gd GLQ 5644  
## 2 PConc Ex TA Gd GLQ 2260  
## 3 PConc Ex TA Av GLQ 2096  
## 4 PConc Ex TA Gd GLQ 1455  
## 5 PConc Ex TA Gd GLQ 1387  
## 6 BrkTil TA TA Mn Unf 0  
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir  
## 1 Unf 0 466 6110 GasA Ex Y  
## 2 Unf 0 878 3138 GasA Ex Y  
## 3 Unf 0 300 2396 GasA Ex Y  
## 4 Unf 0 989 2444 GasA Ex Y  
## 5 Unf 0 543 1930 GasA Ex Y  
## 6 Unf 0 1107 1107 GasA Ex Y  
## Electrical X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath  
## 1 SBrkr 4692 950 0 5642 2  
## 2 SBrkr 3138 1538 0 4676 1  
## 3 SBrkr 2411 2065 0 4476 1  
## 4 SBrkr 2444 1872 0 4316 0  
## 5 SBrkr 1831 1796 0 3627 1  
## 6 SBrkr 1518 1518 572 3608 0  
## BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual  
## 1 0 2 1 3 1 Ex  
## 2 0 3 1 3 1 Ex  
## 3 0 3 1 4 1 Ex  
## 4 1 3 1 4 1 Ex  
## 5 0 3 1 4 1 Gd  
## 6 0 2 1 4 1 Ex  
## TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt  
## 1 12 Typ 3 Gd Attchd 2008  
## 2 11 Typ 1 Gd BuiltIn 2007  
## 3 10 Typ 2 TA Attchd 1996  
## 4 10 Typ 2 Ex Attchd 1994  
## 5 10 Typ 1 TA Attchd 1995  
## 6 12 Typ 2 TA Detchd 1993  
## GarageFinish GarageCars GarageArea GarageQual GarageCond PavedDrive  
## 1 Fin 2 1418 TA TA Y  
## 2 Fin 3 884 TA TA Y  
## 3 Fin 3 813 TA TA Y  
## 4 Fin 3 832 TA TA Y  
## 5 Fin 3 807 TA TA Y  
## 6 Unf 3 840 Ex TA Y  
## WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC  
## 1 214 292 0 0 0 480 Gd  
## 2 208 406 0 0 0 0 <NA>  
## 3 171 78 0 0 0 555 Ex  
## 4 382 50 0 0 0 0 <NA>  
## 5 361 76 0 0 0 0 <NA>  
## 6 0 260 0 0 410 0 <NA>  
## Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition SalePrice  
## 1 <NA> <NA> 0 1 2008 New Partial 160000  
## 2 <NA> <NA> 0 10 2007 New Partial 184750  
## 3 MnPrv <NA> 0 7 2007 WD Abnorml 745000  
## 4 <NA> <NA> 0 1 2007 WD Normal 755000  
## 5 <NA> <NA> 0 7 2006 WD Normal 625000  
## 6 GdPrv <NA> 0 6 2006 WD Normal 475000

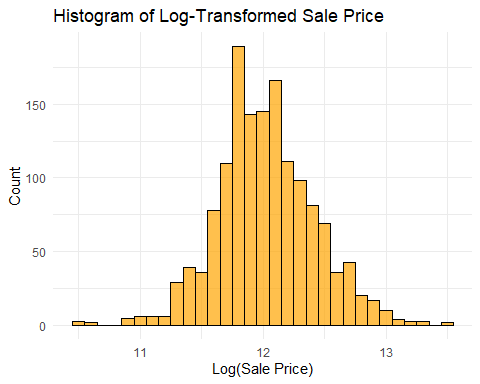
testData <- read.csv("/Users/katie/Documents/MSDS/Statistical Science/Unit 13/Project/test.csv")  
test <- testData  
head(test)

## Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape  
## 1 1461 20 RH 80 11622 Pave <NA> Reg  
## 2 1462 20 RL 81 14267 Pave <NA> IR1  
## 3 1463 60 RL 74 13830 Pave <NA> IR1  
## 4 1464 60 RL 78 9978 Pave <NA> IR1  
## 5 1465 120 RL 43 5005 Pave <NA> IR1  
## 6 1466 60 RL 75 10000 Pave <NA> IR1  
## LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2  
## 1 Lvl AllPub Inside Gtl NAmes Feedr Norm  
## 2 Lvl AllPub Corner Gtl NAmes Norm Norm  
## 3 Lvl AllPub Inside Gtl Gilbert Norm Norm  
## 4 Lvl AllPub Inside Gtl Gilbert Norm Norm  
## 5 HLS AllPub Inside Gtl StoneBr Norm Norm  
## 6 Lvl AllPub Corner Gtl Gilbert Norm Norm  
## BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle  
## 1 1Fam 1Story 5 6 1961 1961 Gable  
## 2 1Fam 1Story 6 6 1958 1958 Hip  
## 3 1Fam 2Story 5 5 1997 1998 Gable  
## 4 1Fam 2Story 6 6 1998 1998 Gable  
## 5 TwnhsE 1Story 8 5 1992 1992 Gable  
## 6 1Fam 2Story 6 5 1993 1994 Gable  
## RoofMatl Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond  
## 1 CompShg VinylSd VinylSd None 0 TA TA  
## 2 CompShg Wd Sdng Wd Sdng BrkFace 108 TA TA  
## 3 CompShg VinylSd VinylSd None 0 TA TA  
## 4 CompShg VinylSd VinylSd BrkFace 20 TA TA  
## 5 CompShg HdBoard HdBoard None 0 Gd TA  
## 6 CompShg HdBoard HdBoard None 0 TA TA  
## Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1  
## 1 CBlock TA TA No Rec 468  
## 2 CBlock TA TA No ALQ 923  
## 3 PConc Gd TA No GLQ 791  
## 4 PConc TA TA No GLQ 602  
## 5 PConc Gd TA No ALQ 263  
## 6 PConc Gd TA No Unf 0  
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir  
## 1 LwQ 144 270 882 GasA TA Y  
## 2 Unf 0 406 1329 GasA TA Y  
## 3 Unf 0 137 928 GasA Gd Y  
## 4 Unf 0 324 926 GasA Ex Y  
## 5 Unf 0 1017 1280 GasA Ex Y  
## 6 Unf 0 763 763 GasA Gd Y  
## Electrical X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath  
## 1 SBrkr 896 0 0 896 0  
## 2 SBrkr 1329 0 0 1329 0  
## 3 SBrkr 928 701 0 1629 0  
## 4 SBrkr 926 678 0 1604 0  
## 5 SBrkr 1280 0 0 1280 0  
## 6 SBrkr 763 892 0 1655 0  
## BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual  
## 1 0 1 0 2 1 TA  
## 2 0 1 1 3 1 Gd  
## 3 0 2 1 3 1 TA  
## 4 0 2 1 3 1 Gd  
## 5 0 2 0 2 1 Gd  
## 6 0 2 1 3 1 TA  
## TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt  
## 1 5 Typ 0 <NA> Attchd 1961  
## 2 6 Typ 0 <NA> Attchd 1958  
## 3 6 Typ 1 TA Attchd 1997  
## 4 7 Typ 1 Gd Attchd 1998  
## 5 5 Typ 0 <NA> Attchd 1992  
## 6 7 Typ 1 TA Attchd 1993  
## GarageFinish GarageCars GarageArea GarageQual GarageCond PavedDrive  
## 1 Unf 1 730 TA TA Y  
## 2 Unf 1 312 TA TA Y  
## 3 Fin 2 482 TA TA Y  
## 4 Fin 2 470 TA TA Y  
## 5 RFn 2 506 TA TA Y  
## 6 Fin 2 440 TA TA Y  
## WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC  
## 1 140 0 0 0 120 0 <NA>  
## 2 393 36 0 0 0 0 <NA>  
## 3 212 34 0 0 0 0 <NA>  
## 4 360 36 0 0 0 0 <NA>  
## 5 0 82 0 0 144 0 <NA>  
## 6 157 84 0 0 0 0 <NA>  
## Fence MiscFeature MiscVal MoSold YrSold SaleType SaleCondition  
## 1 MnPrv <NA> 0 6 2010 WD Normal  
## 2 <NA> Gar2 12500 6 2010 WD Normal  
## 3 MnPrv <NA> 0 3 2010 WD Normal  
## 4 <NA> <NA> 0 6 2010 WD Normal  
## 5 <NA> <NA> 0 1 2010 WD Normal  
## 6 <NA> <NA> 0 4 2010 WD Normal

# Change based on data descriptions file  
changeColumnNA <- c("MiscFeature", "Fence", "PoolQC", "GarageCond",   
 "GarageQual", "GarageFinish", "GarageType",   
 "FireplaceQu", "BsmtFinType2", "BsmtFinType1",   
 "BsmtExposure", "BsmtCond", "BsmtQual", "Alley", "MasVnrType")  
train <- train %>%  
 mutate(across(all\_of(changeColumnNA), ~replace(., is.na(.), "None")))  
test <- test %>%  
 mutate(across(all\_of(changeColumnNA), ~replace(., is.na(.), "None")))  
  
ggplot(train, aes(x = SalePrice)) +  
 geom\_histogram(binwidth = 17000, fill = "orange", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Sale Price",   
 x = "Sale Price ($)",   
 y = "Frequency") +  
 theme\_minimal()



train$LogSalePrice <- log(train$SalePrice)  
  
ggplot(train, aes(x = LogSalePrice)) +  
 geom\_histogram(binwidth = 0.1, fill = "orange", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Log-Transformed Sale Price",   
 x = "Log(Sale Price)",   
 y = "Count") +  
 theme\_minimal()



# Define the mappings for categorical variables changing to numeric  
quality\_mapping <- c("Ex" = 5, "Gd" = 4, "TA" = 3, "Fa" = 2, "Po" = 1, "None" = 0, "No"=0)  
bsmtExp\_mapping <- c("Gd" = 4, "Av" = 3, "Mn" = 2, "No" = 1, "None" = 0)  
fence\_mapping <- c("GdPrv" = 4, "MnPrv" = 3, "GdWo" = 2, "MnWw" = 1, "None" = 0)  
garageFinish\_mapping <- c("Fin" = 3, "RFn" = 2, "Unf" = 1, "None" = 0)  
functional\_mapping <- c("Typ" = 8, "Min1" = 7, "Min2" = 6, "Mod" = 5, "Maj1" = 4,   
 "Maj2" = 3, "Sev" = 2, "Sal" = 1)  
centralAir\_mapping <- c("Y" = 1, "N" = 0)  
bsmtFin\_mapping <- c("GLQ" = 6, "ALQ" = 5, "BLQ" = 4, "Rec" = 3, "LwQ" = 2, "Unf" = 1, "None" = 0)  
  
# List of columns to be updated  
columns\_to\_update <- c("ExterQual", "ExterCond", "BsmtQual", "BsmtCond", "HeatingQC",   
 "KitchenQual", "FireplaceQu", "GarageQual", "GarageCond", "PoolQC",   
 "BsmtExposure", "Fence", "GarageFinish", "Functional", "CentralAir",   
 "BsmtFinType1", "BsmtFinType2")  
  
# Apply the appropriate mapping for each column  
train[columns\_to\_update] <- lapply(train[columns\_to\_update], function(x) {  
 if (all(levels(factor(x)) %in% names(quality\_mapping))) {  
 as.numeric(factor(x, levels = names(quality\_mapping), labels = quality\_mapping[names(quality\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(fence\_mapping))) {  
 as.numeric(factor(x, levels = names(fence\_mapping), labels = fence\_mapping[names(fence\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(garageFinish\_mapping))) {  
 as.numeric(factor(x, levels = names(garageFinish\_mapping), labels = garageFinish\_mapping[names(garageFinish\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(functional\_mapping))) {  
 as.numeric(factor(x, levels = names(functional\_mapping), labels = functional\_mapping[names(functional\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(centralAir\_mapping))) {  
 as.numeric(factor(x, levels = names(centralAir\_mapping), labels = centralAir\_mapping[names(centralAir\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(bsmtFin\_mapping))) {  
 as.numeric(factor(x, levels = names(bsmtFin\_mapping), labels = bsmtFin\_mapping[names(bsmtFin\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(bsmtExp\_mapping))) {  
 as.numeric(factor(x, levels = names(bsmtExp\_mapping), labels = bsmtExp\_mapping[names(bsmtExp\_mapping)]))   
 } else {  
 x # Return as is if no matching mapping is found  
 }  
})  
  
head(train[columns\_to\_update])

## ExterQual ExterCond BsmtQual BsmtCond HeatingQC KitchenQual FireplaceQu  
## 1 1 3 1 3 1 1 2  
## 2 1 3 1 3 1 1 2  
## 3 2 3 1 3 1 1 3  
## 4 1 3 1 3 1 1 1  
## 5 2 2 1 3 1 2 3  
## 6 2 2 3 3 1 1 3  
## GarageQual GarageCond PoolQC BsmtExposure Fence GarageFinish Functional  
## 1 3 3 2 1 5 1 1  
## 2 3 3 6 1 5 1 1  
## 3 3 3 1 2 2 1 1  
## 4 3 3 6 1 5 1 1  
## 5 3 3 6 1 5 1 1  
## 6 1 3 6 3 1 3 1  
## CentralAir BsmtFinType1 BsmtFinType2  
## 1 1 1 6  
## 2 1 1 6  
## 3 1 1 6  
## 4 1 1 6  
## 5 1 1 6  
## 6 1 6 6

# Apply the same mappings for each column in the test set  
test[columns\_to\_update] <- lapply(test[columns\_to\_update], function(x) {  
 if (all(levels(factor(x)) %in% names(quality\_mapping))) {  
 as.numeric(factor(x, levels = names(quality\_mapping), labels = quality\_mapping[names(quality\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(fence\_mapping))) {  
 as.numeric(factor(x, levels = names(fence\_mapping), labels = fence\_mapping[names(fence\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(garageFinish\_mapping))) {  
 as.numeric(factor(x, levels = names(garageFinish\_mapping), labels = garageFinish\_mapping[names(garageFinish\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(functional\_mapping))) {  
 as.numeric(factor(x, levels = names(functional\_mapping), labels = functional\_mapping[names(functional\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(centralAir\_mapping))) {  
 as.numeric(factor(x, levels = names(centralAir\_mapping), labels = centralAir\_mapping[names(centralAir\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(bsmtFin\_mapping))) {  
 as.numeric(factor(x, levels = names(bsmtFin\_mapping), labels = bsmtFin\_mapping[names(bsmtFin\_mapping)]))  
 } else if (all(levels(factor(x)) %in% names(bsmtExp\_mapping))) {  
 as.numeric(factor(x, levels = names(bsmtExp\_mapping), labels = bsmtExp\_mapping[names(bsmtExp\_mapping)]))   
 } else {  
 x # Return as is if no matching mapping is found  
 }  
})  
  
head(test[columns\_to\_update])

## ExterQual ExterCond BsmtQual BsmtCond HeatingQC KitchenQual FireplaceQu  
## 1 3 3 3 3 3 3 6  
## 2 3 3 3 3 3 2 6  
## 3 3 3 2 3 2 3 3  
## 4 3 3 3 3 1 2 2  
## 5 2 3 2 3 1 2 6  
## 6 3 3 2 3 2 3 3  
## GarageQual GarageCond PoolQC BsmtExposure Fence GarageFinish Functional  
## 1 3 3 6 4 2 3 1  
## 2 3 3 6 4 5 3 1  
## 3 3 3 6 4 2 1 1  
## 4 3 3 6 4 5 1 1  
## 5 3 3 6 4 5 2 1  
## 6 3 3 6 4 5 1 1  
## CentralAir BsmtFinType1 BsmtFinType2  
## 1 1 4 5  
## 2 1 2 6  
## 3 1 1 6  
## 4 1 1 6  
## 5 1 2 6  
## 6 1 6 6

totalRows <- nrow(train)  
  
# Count the number of missing values for each column  
missingData <- sapply(train, function(x) sum(is.na(x)))  
  
# Calculate the percentage of missing values for each column  
percNAMissing <- (missingData / totalRows) \* 100  
numeric <- sapply(train, is.numeric)  
  
# Create a summary table with column names, number of missing values, and the percentage of missing values  
missingSummary <- data.frame(  
 Column = names(train),  
 MissingNACount = missingData,  
 PercentageNAMissing = percNAMissing,  
 Numeric = numeric  
)  
  
missingSummary <- missingSummary %>%  
 arrange(desc(Numeric), desc(MissingNACount))  
missingSummary

## Column MissingNACount PercentageNAMissing Numeric  
## LotFrontage LotFrontage 259 17.73972603 TRUE  
## GarageYrBlt GarageYrBlt 81 5.54794521 TRUE  
## MasVnrArea MasVnrArea 8 0.54794521 TRUE  
## Id Id 0 0.00000000 TRUE  
## MSSubClass MSSubClass 0 0.00000000 TRUE  
## LotArea LotArea 0 0.00000000 TRUE  
## OverallQual OverallQual 0 0.00000000 TRUE  
## OverallCond OverallCond 0 0.00000000 TRUE  
## YearBuilt YearBuilt 0 0.00000000 TRUE  
## YearRemodAdd YearRemodAdd 0 0.00000000 TRUE  
## ExterQual ExterQual 0 0.00000000 TRUE  
## ExterCond ExterCond 0 0.00000000 TRUE  
## BsmtQual BsmtQual 0 0.00000000 TRUE  
## BsmtCond BsmtCond 0 0.00000000 TRUE  
## BsmtExposure BsmtExposure 0 0.00000000 TRUE  
## BsmtFinType1 BsmtFinType1 0 0.00000000 TRUE  
## BsmtFinSF1 BsmtFinSF1 0 0.00000000 TRUE  
## BsmtFinType2 BsmtFinType2 0 0.00000000 TRUE  
## BsmtFinSF2 BsmtFinSF2 0 0.00000000 TRUE  
## BsmtUnfSF BsmtUnfSF 0 0.00000000 TRUE  
## TotalBsmtSF TotalBsmtSF 0 0.00000000 TRUE  
## HeatingQC HeatingQC 0 0.00000000 TRUE  
## CentralAir CentralAir 0 0.00000000 TRUE  
## X1stFlrSF X1stFlrSF 0 0.00000000 TRUE  
## X2ndFlrSF X2ndFlrSF 0 0.00000000 TRUE  
## LowQualFinSF LowQualFinSF 0 0.00000000 TRUE  
## GrLivArea GrLivArea 0 0.00000000 TRUE  
## BsmtFullBath BsmtFullBath 0 0.00000000 TRUE  
## BsmtHalfBath BsmtHalfBath 0 0.00000000 TRUE  
## FullBath FullBath 0 0.00000000 TRUE  
## HalfBath HalfBath 0 0.00000000 TRUE  
## BedroomAbvGr BedroomAbvGr 0 0.00000000 TRUE  
## KitchenAbvGr KitchenAbvGr 0 0.00000000 TRUE  
## KitchenQual KitchenQual 0 0.00000000 TRUE  
## TotRmsAbvGrd TotRmsAbvGrd 0 0.00000000 TRUE  
## Functional Functional 0 0.00000000 TRUE  
## Fireplaces Fireplaces 0 0.00000000 TRUE  
## FireplaceQu FireplaceQu 0 0.00000000 TRUE  
## GarageFinish GarageFinish 0 0.00000000 TRUE  
## GarageCars GarageCars 0 0.00000000 TRUE  
## GarageArea GarageArea 0 0.00000000 TRUE  
## GarageQual GarageQual 0 0.00000000 TRUE  
## GarageCond GarageCond 0 0.00000000 TRUE  
## WoodDeckSF WoodDeckSF 0 0.00000000 TRUE  
## OpenPorchSF OpenPorchSF 0 0.00000000 TRUE  
## EnclosedPorch EnclosedPorch 0 0.00000000 TRUE  
## X3SsnPorch X3SsnPorch 0 0.00000000 TRUE  
## ScreenPorch ScreenPorch 0 0.00000000 TRUE  
## PoolArea PoolArea 0 0.00000000 TRUE  
## PoolQC PoolQC 0 0.00000000 TRUE  
## Fence Fence 0 0.00000000 TRUE  
## MiscVal MiscVal 0 0.00000000 TRUE  
## MoSold MoSold 0 0.00000000 TRUE  
## YrSold YrSold 0 0.00000000 TRUE  
## SalePrice SalePrice 0 0.00000000 TRUE  
## LogSalePrice LogSalePrice 0 0.00000000 TRUE  
## Electrical Electrical 1 0.06849315 FALSE  
## MSZoning MSZoning 0 0.00000000 FALSE  
## Street Street 0 0.00000000 FALSE  
## Alley Alley 0 0.00000000 FALSE  
## LotShape LotShape 0 0.00000000 FALSE  
## LandContour LandContour 0 0.00000000 FALSE  
## Utilities Utilities 0 0.00000000 FALSE  
## LotConfig LotConfig 0 0.00000000 FALSE  
## LandSlope LandSlope 0 0.00000000 FALSE  
## Neighborhood Neighborhood 0 0.00000000 FALSE  
## Condition1 Condition1 0 0.00000000 FALSE  
## Condition2 Condition2 0 0.00000000 FALSE  
## BldgType BldgType 0 0.00000000 FALSE  
## HouseStyle HouseStyle 0 0.00000000 FALSE  
## RoofStyle RoofStyle 0 0.00000000 FALSE  
## RoofMatl RoofMatl 0 0.00000000 FALSE  
## Exterior1st Exterior1st 0 0.00000000 FALSE  
## Exterior2nd Exterior2nd 0 0.00000000 FALSE  
## MasVnrType MasVnrType 0 0.00000000 FALSE  
## Foundation Foundation 0 0.00000000 FALSE  
## Heating Heating 0 0.00000000 FALSE  
## GarageType GarageType 0 0.00000000 FALSE  
## PavedDrive PavedDrive 0 0.00000000 FALSE  
## MiscFeature MiscFeature 0 0.00000000 FALSE  
## SaleType SaleType 0 0.00000000 FALSE  
## SaleCondition SaleCondition 0 0.00000000 FALSE

# Remove unnecessary columns/rows  
trainNoMissing <- train %>%  
 select(-Electrical) %>%  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .))) %>%  
 filter(GrLivArea <= 4000) %>% #Outliers in training setbased on pt. 1 code  
 filter(TotalBsmtSF <= 3000, MasVnrArea <= 1300,   
 LotFrontage <= 300, LotArea <= 100000,   
 GarageArea <= 1200, BsmtFinSF2 <= 1400)  
trainNoMissing$AllSQFT = trainNoMissing$GrLivArea + trainNoMissing$TotalBsmtSF  
trainNoMissing <- trainNoMissing %>%  
 filter(AllSQFT <= 6000)  
head(trainNoMissing)

## Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape  
## 1 186 75 RM 90.00000 22950 Pave None IR2  
## 2 305 75 RM 87.00000 18386 Pave None Reg  
## 3 1269 50 RL 70.04996 14100 Pave None IR1  
## 4 636 190 RH 60.00000 10896 Pave Pave Reg  
## 5 770 60 RL 47.00000 53504 Pave None IR2  
## 6 1354 50 RL 56.00000 14720 Pave None IR1  
## LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2  
## 1 Lvl AllPub Inside Gtl OldTown Artery Norm  
## 2 Lvl AllPub Inside Gtl OldTown Norm Norm  
## 3 Lvl AllPub Inside Mod Crawfor Norm Norm  
## 4 Bnk AllPub Inside Gtl SWISU Feedr Norm  
## 5 HLS AllPub CulDSac Mod StoneBr Norm Norm  
## 6 Lvl AllPub CulDSac Gtl NoRidge Norm Norm  
## BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle  
## 1 1Fam 2.5Fin 10 9 1892 1993 Gable  
## 2 1Fam 2.5Fin 7 9 1880 2002 Gable  
## 3 1Fam 1.5Fin 8 9 1935 1997 Gable  
## 4 2fmCon 2.5Fin 6 7 1914 1995 Hip  
## 5 1Fam 2Story 8 5 2003 2003 Hip  
## 6 1Fam 1.5Fin 8 5 1995 1996 Hip  
## RoofMatl Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond  
## 1 WdShngl Wd Sdng Wd Sdng None 0 2 2  
## 2 CompShg CemntBd CmentBd None 0 3 3  
## 3 CompShg Stucco Stucco BrkFace 632 3 2  
## 4 CompShg VinylSd VinylSd None 0 4 3  
## 5 CompShg CemntBd Wd Shng BrkFace 603 1 3  
## 6 CompShg VinylSd VinylSd BrkFace 579 2 3  
## Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1  
## 1 BrkTil 3 3 3 6 0  
## 2 BrkTil 3 3 4 6 0  
## 3 CBlock 3 3 3 4 192  
## 4 CBlock 3 4 4 5 256  
## 5 PConc 2 3 1 2 1416  
## 6 PConc 2 3 2 1 816  
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir  
## 1 6 0 1107 1107 GasA 1 1  
## 2 6 0 1470 1470 GasA 1 1  
## 3 6 0 536 728 GasA 1 1  
## 4 6 0 1184 1440 GasA 1 1  
## 5 6 0 234 1650 GasA 1 1  
## 6 6 0 1217 2033 GasA 1 1  
## X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath  
## 1 1518 1518 572 3608 0 0 2  
## 2 1675 1818 0 3493 0 0 3  
## 3 1968 1479 0 3447 0 0 3  
## 4 1440 1440 515 3395 0 0 2  
## 5 1690 1589 0 3279 1 0 3  
## 6 2053 1185 0 3238 1 0 2  
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional  
## 1 1 4 1 1 12 1  
## 2 0 3 1 2 10 1  
## 3 1 4 1 2 11 1  
## 4 0 8 2 4 14 1  
## 5 1 4 1 1 12 4  
## 6 1 4 1 2 9 1  
## Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars  
## 1 2 3 Detchd 1993.000 3 3  
## 2 1 1 Attchd 2003.000 3 3  
## 3 2 2 BuiltIn 1982.000 3 3  
## 4 0 6 None 1978.506 4 0  
## 5 1 2 BuiltIn 2003.000 1 3  
## 6 1 1 Attchd 1996.000 1 3  
## GarageArea GarageQual GarageCond PavedDrive WoodDeckSF OpenPorchSF  
## 1 840 1 3 Y 0 260  
## 2 870 3 3 Y 302 0  
## 3 1014 3 3 Y 314 12  
## 4 0 6 6 N 0 110  
## 5 841 3 3 Y 503 36  
## 6 666 3 3 Y 283 86  
## EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC Fence MiscFeature  
## 1 0 0 410 0 6 1 None  
## 2 0 0 0 0 6 5 None  
## 3 0 0 0 0 6 3 None  
## 4 0 0 0 0 6 5 None  
## 5 0 0 210 0 6 5 None  
## 6 0 0 0 0 6 5 None  
## MiscVal MoSold YrSold SaleType SaleCondition SalePrice LogSalePrice AllSQFT  
## 1 0 6 2006 WD Normal 475000 13.07107 4715  
## 2 0 5 2008 WD Normal 295000 12.59473 4963  
## 3 0 5 2008 WD Normal 381000 12.85055 4175  
## 4 0 3 2007 WD Abnorml 200000 12.20607 4835  
## 5 0 6 2010 WD Normal 538000 13.19561 4929  
## 6 0 3 2010 WD Normal 410000 12.92391 5271

print(paste("# of NA values left:", sum(is.na(trainNoMissing))))

## [1] "# of NA values left: 0"

testNoMissing <- test %>%  
 # Replace NA in the numeric columns with the column's mean  
 mutate(across(where(is.numeric), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
print(paste("# of NA values:", sum(is.na(testNoMissing))))

## [1] "# of NA values: 9"

testNoMissing$AllSQFT = testNoMissing$GrLivArea + testNoMissing$TotalBsmtSF  
  
# Remove non-numeric columns with NA values  
removeMeTest <- colnames(testNoMissing)[sapply(testNoMissing, function(x) !is.numeric(x) && any(is.na(x)))]  
removeMeTest

## [1] "MSZoning" "Utilities" "Exterior1st" "Exterior2nd" "SaleType"

testNoMissing <- testNoMissing %>%  
 select(-all\_of(removeMeTest))  
head(testNoMissing)

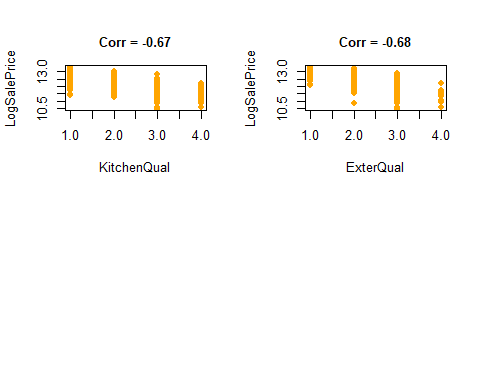
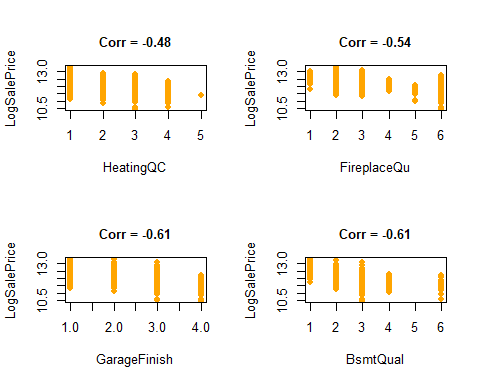
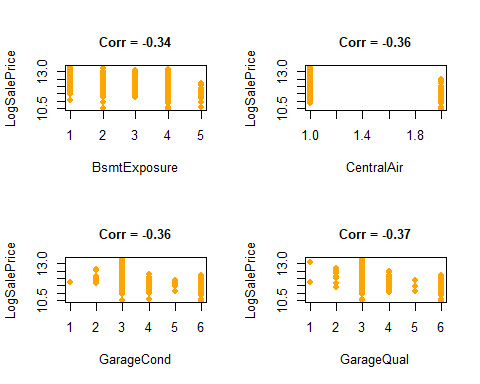
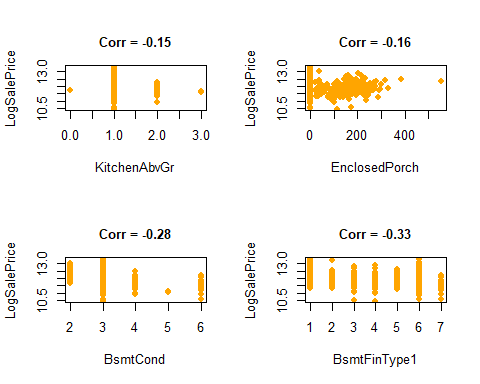
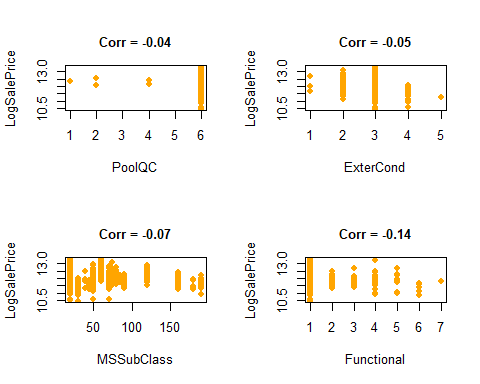
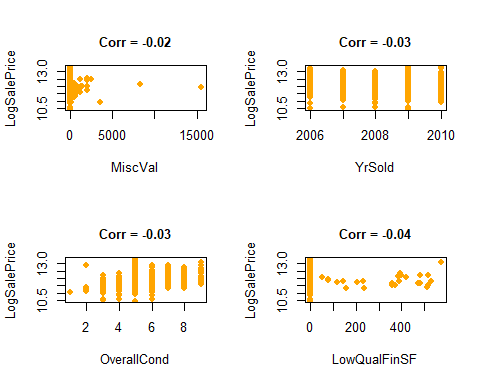
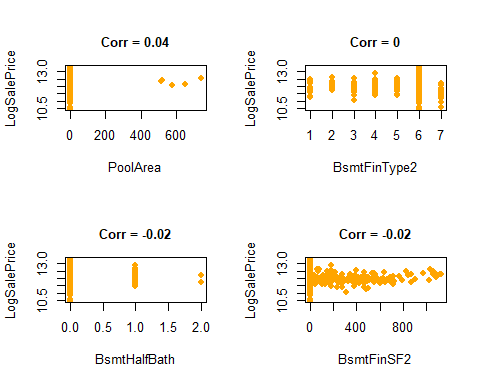
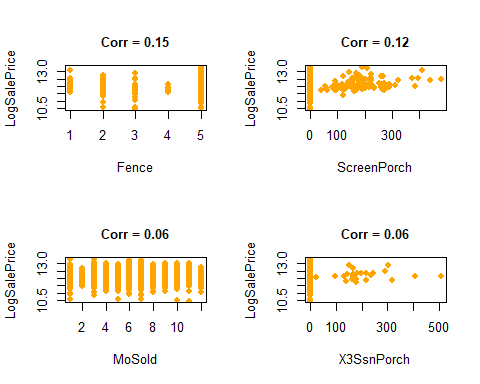
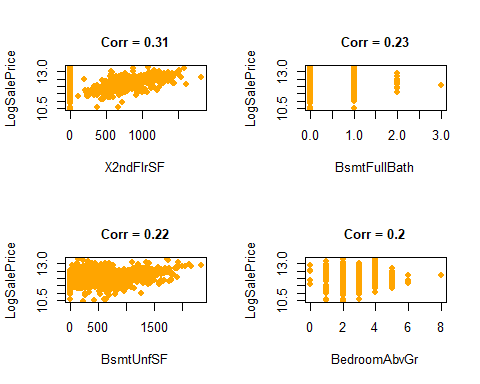
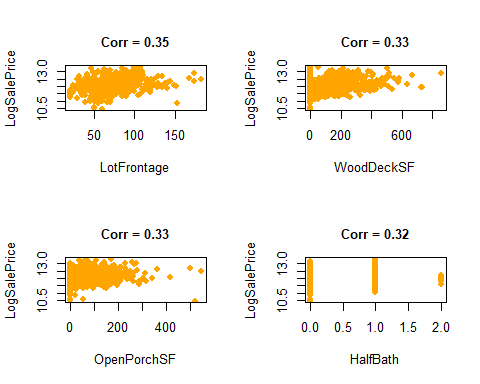
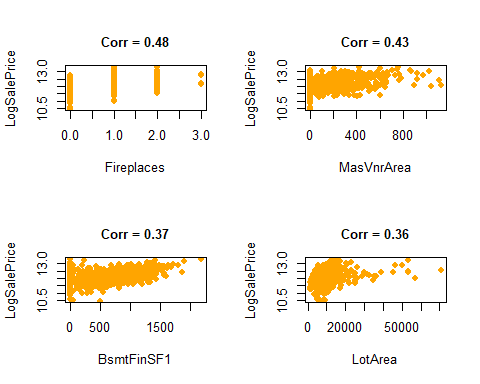
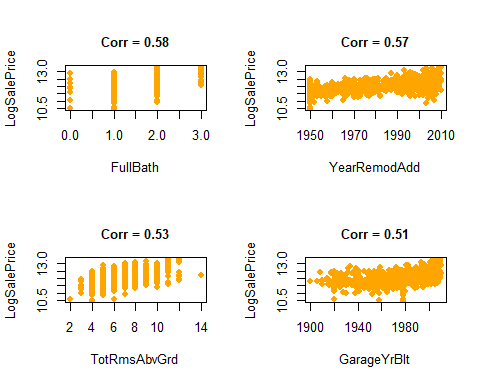
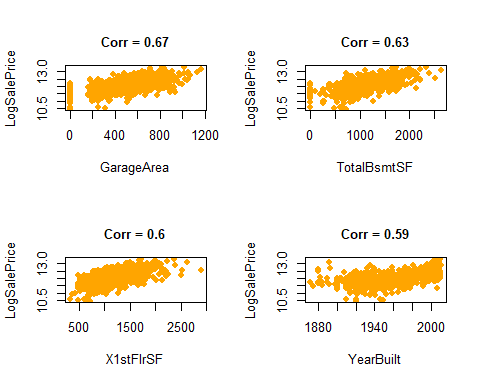
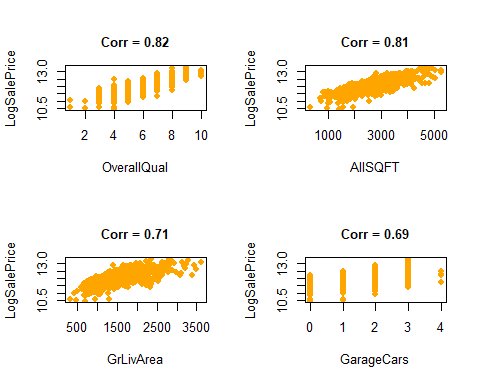
## Id MSSubClass LotFrontage LotArea Street Alley LotShape LandContour  
## 1 1461 20 80 11622 Pave None Reg Lvl  
## 2 1462 20 81 14267 Pave None IR1 Lvl  
## 3 1463 60 74 13830 Pave None IR1 Lvl  
## 4 1464 60 78 9978 Pave None IR1 Lvl  
## 5 1465 120 43 5005 Pave None IR1 HLS  
## 6 1466 60 75 10000 Pave None IR1 Lvl  
## LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle  
## 1 Inside Gtl NAmes Feedr Norm 1Fam 1Story  
## 2 Corner Gtl NAmes Norm Norm 1Fam 1Story  
## 3 Inside Gtl Gilbert Norm Norm 1Fam 2Story  
## 4 Inside Gtl Gilbert Norm Norm 1Fam 2Story  
## 5 Inside Gtl StoneBr Norm Norm TwnhsE 1Story  
## 6 Corner Gtl Gilbert Norm Norm 1Fam 2Story  
## OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl MasVnrType  
## 1 5 6 1961 1961 Gable CompShg None  
## 2 6 6 1958 1958 Hip CompShg BrkFace  
## 3 5 5 1997 1998 Gable CompShg None  
## 4 6 6 1998 1998 Gable CompShg BrkFace  
## 5 8 5 1992 1992 Gable CompShg None  
## 6 6 5 1993 1994 Gable CompShg None  
## MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure  
## 1 0 3 3 CBlock 3 3 4  
## 2 108 3 3 CBlock 3 3 4  
## 3 0 3 3 PConc 2 3 4  
## 4 20 3 3 PConc 3 3 4  
## 5 0 2 3 PConc 2 3 4  
## 6 0 3 3 PConc 2 3 4  
## BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating  
## 1 4 468 5 144 270 882 GasA  
## 2 2 923 6 0 406 1329 GasA  
## 3 1 791 6 0 137 928 GasA  
## 4 1 602 6 0 324 926 GasA  
## 5 2 263 6 0 1017 1280 GasA  
## 6 6 0 6 0 763 763 GasA  
## HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea  
## 1 3 1 SBrkr 896 0 0 896  
## 2 3 1 SBrkr 1329 0 0 1329  
## 3 2 1 SBrkr 928 701 0 1629  
## 4 1 1 SBrkr 926 678 0 1604  
## 5 1 1 SBrkr 1280 0 0 1280  
## 6 2 1 SBrkr 763 892 0 1655  
## BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr  
## 1 0 0 1 0 2 1  
## 2 0 0 1 1 3 1  
## 3 0 0 2 1 3 1  
## 4 0 0 2 1 3 1  
## 5 0 0 2 0 2 1  
## 6 0 0 2 1 3 1  
## KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType  
## 1 3 5 1 0 6 Attchd  
## 2 2 6 1 0 6 Attchd  
## 3 3 6 1 1 3 Attchd  
## 4 2 7 1 1 2 Attchd  
## 5 2 5 1 0 6 Attchd  
## 6 3 7 1 1 3 Attchd  
## GarageYrBlt GarageFinish GarageCars GarageArea GarageQual GarageCond  
## 1 1961 3 1 730 3 3  
## 2 1958 3 1 312 3 3  
## 3 1997 1 2 482 3 3  
## 4 1998 1 2 470 3 3  
## 5 1992 2 2 506 3 3  
## 6 1993 1 2 440 3 3  
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch  
## 1 Y 140 0 0 0 120  
## 2 Y 393 36 0 0 0  
## 3 Y 212 34 0 0 0  
## 4 Y 360 36 0 0 0  
## 5 Y 0 82 0 0 144  
## 6 Y 157 84 0 0 0  
## PoolArea PoolQC Fence MiscFeature MiscVal MoSold YrSold SaleCondition AllSQFT  
## 1 0 6 2 None 0 6 2010 Normal 1778  
## 2 0 6 5 Gar2 12500 6 2010 Normal 2658  
## 3 0 6 2 None 0 3 2010 Normal 2557  
## 4 0 6 5 None 0 6 2010 Normal 2530  
## 5 0 6 5 None 0 1 2010 Normal 2560  
## 6 0 6 5 None 0 4 2010 Normal 2418

remainingColumns <- colnames(testNoMissing)

# Obtain just the numeric columns to compare against LogSalePrice  
library(dplyr)  
train\_numeric <- trainNoMissing %>%  
 select(where(is.numeric)) %>%  
 select(-LogSalePrice, -Id, -SalePrice)  
  
# List the correlation values  
correlationTable <- train\_numeric %>%  
 summarise(across(everything(), ~ cor(.x, trainNoMissing$LogSalePrice, use = "complete.obs"))) %>%  
 pivot\_longer(everything(), names\_to = "Variable", values\_to = "CorrWithLogSalePrice")  
  
correlationTable <- correlationTable %>%  
 arrange(desc(abs(CorrWithLogSalePrice)))  
correlationTable

## # A tibble: 54 × 2  
## Variable CorrWithLogSalePrice  
## <chr> <dbl>  
## 1 OverallQual 0.818  
## 2 AllSQFT 0.813  
## 3 GrLivArea 0.712  
## 4 GarageCars 0.685  
## 5 ExterQual -0.681  
## 6 GarageArea 0.669  
## 7 KitchenQual -0.668  
## 8 TotalBsmtSF 0.635  
## 9 BsmtQual -0.614  
## 10 GarageFinish -0.606  
## # ℹ 44 more rows

#Plot all numeric variables against LogSalePrice  
sortedVariables <- correlationTable %>%  
 arrange(desc(CorrWithLogSalePrice)) %>%  
 pull(Variable)  
  
num\_plots\_per\_window <- 4  
num\_windows <- ceiling(length(sortedVariables) / num\_plots\_per\_window)  
  
for (i in 1:num\_windows) {  
   
 par(mfrow = c(2, 2))   
 for (j in 0:3) {  
 if (i \* num\_plots\_per\_window - 3 + j <= length(sortedVariables)) {  
 col\_name <- sortedVariables[(i - 1) \* num\_plots\_per\_window + j + 1]  
 plot(train\_numeric[[col\_name]], trainNoMissing$LogSalePrice,  
 main = paste("\nCorr =", round(cor(train\_numeric[[col\_name]], trainNoMissing$LogSalePrice, use = "complete.obs"), 2)),  
 xlab = col\_name, ylab = "LogSalePrice", col = "orange", pch = 16, cex.main = 1)  
 }  
 }  
}

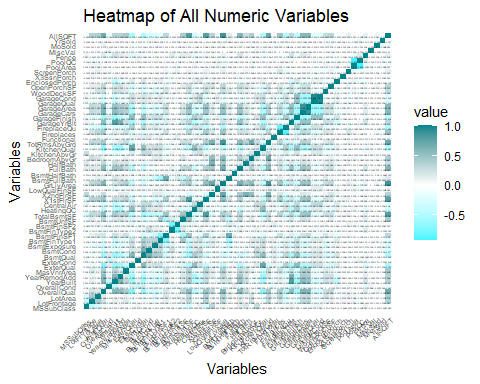


cor\_matrix <- cor(train\_numeric, use = "complete.obs", method = "pearson")  
library(reshape2)

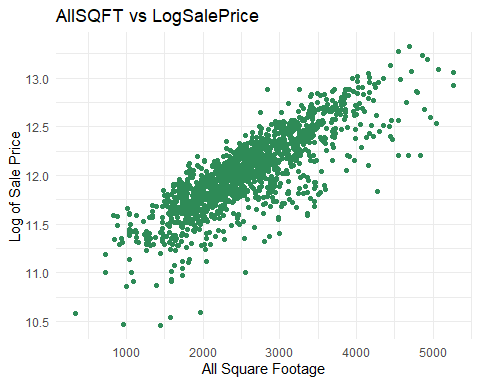
##   
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':  
##   
## smiths

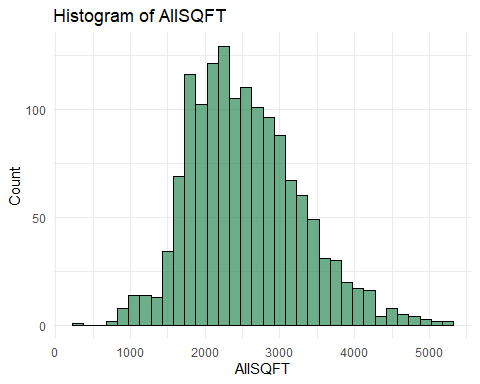
# Melt the correlation matrix into long format  
cor\_melted <- melt(cor\_matrix)  
  
# Create the heatmap  
ggplot(cor\_melted, aes(Var1, Var2, fill = value)) +  
 geom\_tile() +  
 geom\_text(aes(label = round(value, 3)), color = "black", size = 0.5) +  
 scale\_fill\_gradient2(low = "turquoise1", high = "turquoise4", mid = "white", midpoint = 0) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(size = 5, angle = 45, hjust = 1),  
 axis.text.y = element\_text(size = 5)) +  
 labs(title = "Heatmap of All Numeric Variables", x = "Variables", y = "Variables")



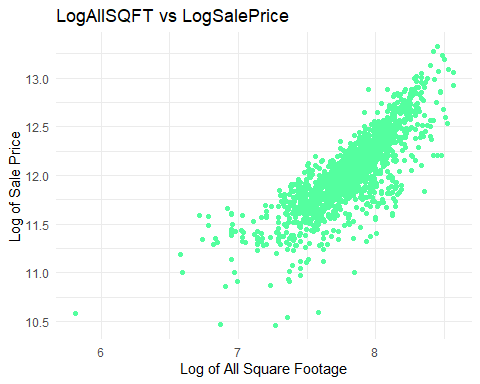
# Total SF for house (incl. basement)  
ggplot(trainNoMissing, aes(x = AllSQFT, y = LogSalePrice)) +  
 geom\_point(color = "seagreen4") +   
 ggtitle("AllSQFT vs LogSalePrice") +  
 xlab("All Square Footage") +   
 ylab("Log of Sale Price") +  
 theme\_minimal()



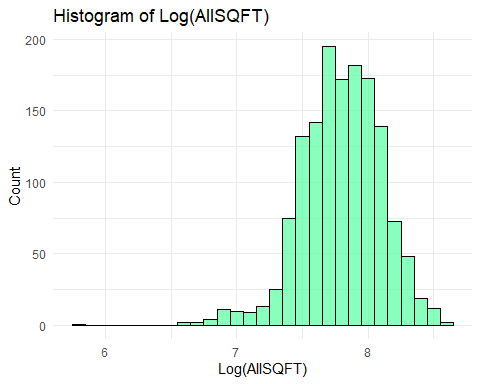
ggplot(trainNoMissing, aes(x = AllSQFT)) +  
 geom\_histogram(binwidth = 150, fill = "seagreen4", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of AllSQFT",   
 x = "AllSQFT",   
 y = "Count") +  
 theme\_minimal()



# Exploring whether log-transform should apply to predictors  
trainNoMissing$LogAllSQFT= log(trainNoMissing$AllSQFT)  
ggplot(trainNoMissing, aes(x = LogAllSQFT, y = LogSalePrice)) +  
 geom\_point(color = "seagreen1") +   
 ggtitle("LogAllSQFT vs LogSalePrice") +  
 xlab("Log of All Square Footage") +   
 ylab("Log of Sale Price") +  
 theme\_minimal()



ggplot(trainNoMissing, aes(x = LogAllSQFT)) +  
 geom\_histogram(binwidth = 0.1, fill = "seagreen1", color = "black", alpha = 0.7) +  
 labs(title = "Histogram of Log(AllSQFT)",   
 x = "Log(AllSQFT)",   
 y = "Count") +  
 theme\_minimal()

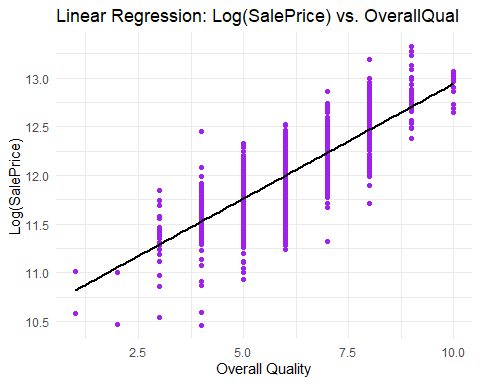


# Fit the linear regression model with OverallQual  
model\_qual <- lm(LogSalePrice ~ OverallQual, data = trainNoMissing)  
summary(model\_qual)

##   
## Call:  
## lm(formula = LogSalePrice ~ OverallQual, data = trainNoMissing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.06629 -0.12738 0.01378 0.13253 0.92640   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.583628 0.027227 388.71 <2e-16 \*\*\*  
## OverallQual 0.235725 0.004372 53.91 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2252 on 1439 degrees of freedom  
## Multiple R-squared: 0.6689, Adjusted R-squared: 0.6686   
## F-statistic: 2907 on 1 and 1439 DF, p-value: < 2.2e-16

# Plot the regression line with OverallQual  
ggplot(trainNoMissing, aes(x = OverallQual, y = LogSalePrice)) +  
 geom\_point(color = "purple") +  
 geom\_smooth(method = "lm", color = "black", se = FALSE) +   
 labs(title = "Linear Regression: Log(SalePrice) vs. OverallQual",  
 x = "Overall Quality",  
 y = "Log(SalePrice)") +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



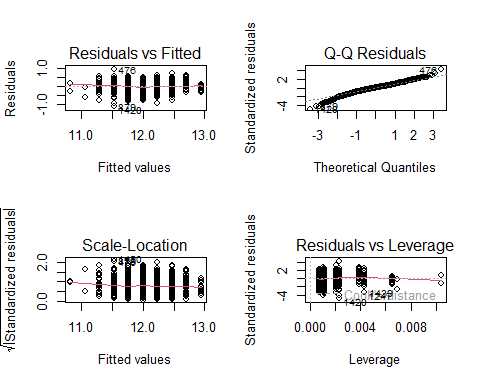
# Calculate the predicted LogSalePrice using the new model  
trainNoMissing$PredictedLogSalePrice\_Qual <- predict(model\_qual)  
trainNoMissing$PredictedSalePrice\_Qual <- exp(trainNoMissing$PredictedLogSalePrice\_Qual)  
  
# Find CV PRESS  
cvpress\_Qual <- sum((trainNoMissing$SalePrice - trainNoMissing$PredictedSalePrice\_Qual)^2)  
print(paste("CV PRESS:", cvpress\_Qual))

## [1] "CV PRESS: 2521860547973.95"

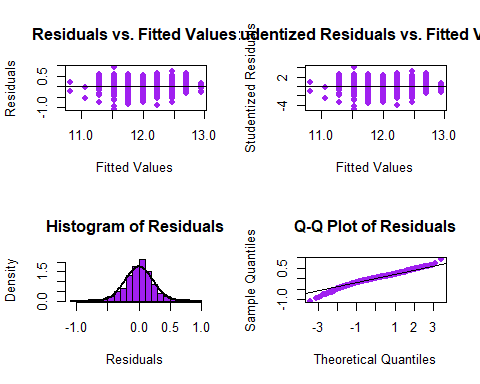
# Calculate AIC   
aic\_Qual <- AIC(model\_qual)  
print(paste("AIC:", aic\_Qual))

## [1] "AIC: -202.822888794298"

par(mfrow = c(2, 2)) # Arrange the plots in a 2x2 grid  
plot(model\_qual)



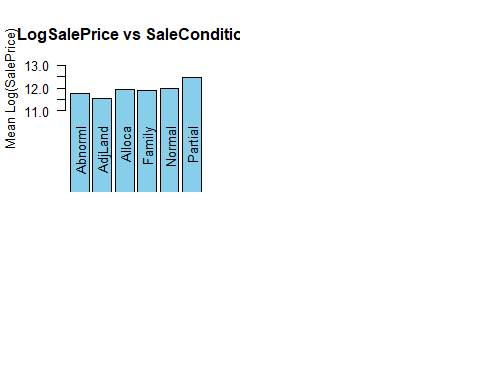
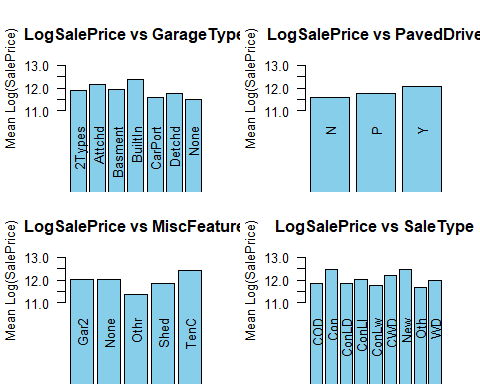
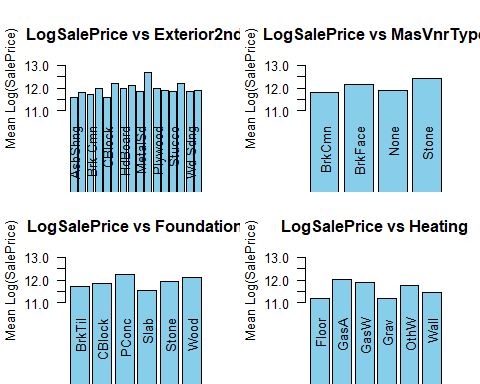
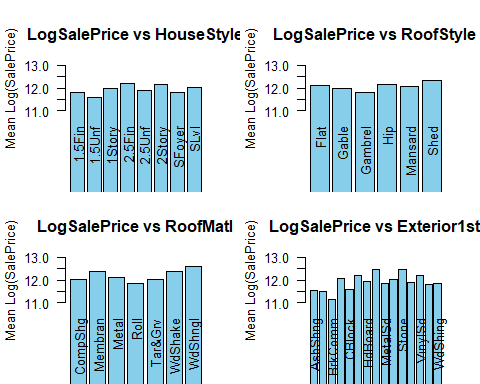
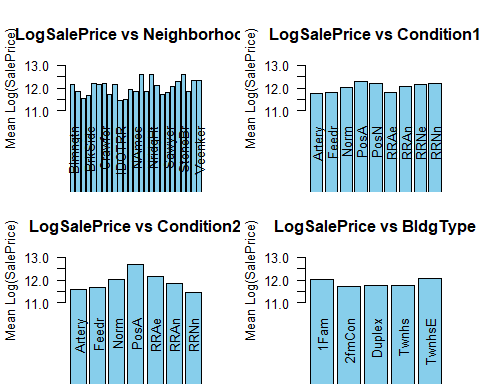
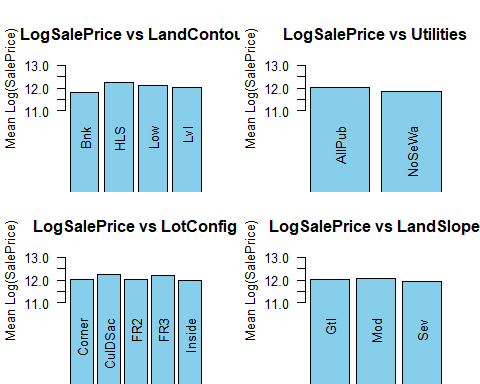
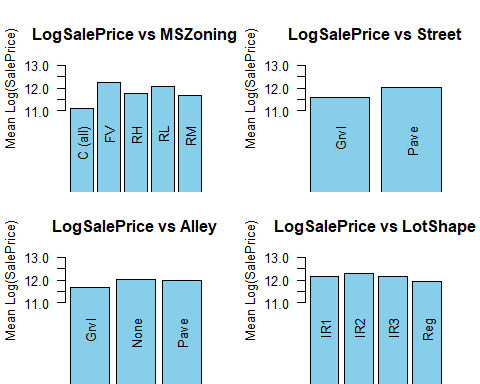
par(mfrow = c(2, 2))  
  
# Residuals vs Fitted values  
plot(fitted(model\_qual), residuals(model\_qual),  
 xlab = "Fitted Values", ylab = "Residuals", main = "Residuals vs. Fitted Values", pch = 19, col="purple")  
abline(h = 0, col = "black")  
  
 # Studentized Residuals  
plot(fitted(model\_qual), rstudent(model\_qual),  
 xlab = "Fitted Values", ylab = "Studentized Residuals", main = "Studentized Residuals vs. Fitted Values",   
 pch = 19, col="purple")  
abline(h = 0, col = "black")  
   
# Histogram of residuals  
hist(residuals(model\_qual),  
 xlab = "Residuals", main = "Histogram of Residuals", col = "purple", border = "black", breaks = 20, probability=TRUE)  
curve(dnorm(x, mean = mean(residuals(model\_qual)), sd = sd(residuals(model\_qual))), col = "black", lwd = 2, add = TRUE)  
  
# Q-Q plot of residuals  
qqnorm(residuals(model\_qual), main = "Q-Q Plot of Residuals", col="purple", pch=19)   
qqline(residuals(model\_qual), col = "black")



nonnumeric <- sapply(trainNoMissing, function(x) !is.numeric(x) && !is.factor(x))  
  
# Get only non-numeric columns  
nonnumeric <- names(trainNoMissing)[nonnumeric]  
nonnumeric

## [1] "MSZoning" "Street" "Alley" "LotShape"   
## [5] "LandContour" "Utilities" "LotConfig" "LandSlope"   
## [9] "Neighborhood" "Condition1" "Condition2" "BldgType"   
## [13] "HouseStyle" "RoofStyle" "RoofMatl" "Exterior1st"   
## [17] "Exterior2nd" "MasVnrType" "Foundation" "Heating"   
## [21] "GarageType" "PavedDrive" "MiscFeature" "SaleType"   
## [25] "SaleCondition"

# Set up 2x2 plot layout  
par(mfrow = c(2, 2))  
  
# Plot each non-numeric variable against LogSalePrice  
for (var in nonnumeric) {  
 if (length(unique(trainNoMissing[[var]])) > 1) { # Only plot if the variable has more than 1 unique value  
 # Create a table of mean LogSalePrice for each level of the categorical variable  
 means <- tapply(trainNoMissing$LogSalePrice, trainNoMissing[[var]], mean, na.rm = TRUE)  
   
 # Create the bar plot  
 barplot(means,   
 main = paste("LogSalePrice vs", var),   
 ylab = "Mean Log(SalePrice)",   
 ylim = c(11,13),  
 col = "skyblue",   
 border = "black",  
 las = 2)  
   
 }  
}



# Fit the multiple linear regression model with Full Bath  
multipleLMmodel <- lm(LogSalePrice ~ GrLivArea \* FullBath, data = trainNoMissing)  
summary(multipleLMmodel)

##   
## Call:  
## lm(formula = LogSalePrice ~ GrLivArea \* FullBath, data = trainNoMissing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.25270 -0.12413 0.02383 0.15590 0.85135   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.082e+01 6.372e-02 169.806 < 2e-16 \*\*\*  
## GrLivArea 6.472e-04 4.707e-05 13.749 < 2e-16 \*\*\*  
## FullBath 3.227e-01 3.964e-02 8.141 8.40e-16 \*\*\*  
## GrLivArea:FullBath -1.105e-04 2.483e-05 -4.452 9.18e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2644 on 1437 degrees of freedom  
## Multiple R-squared: 0.5442, Adjusted R-squared: 0.5432   
## F-statistic: 571.9 on 3 and 1437 DF, p-value: < 2.2e-16

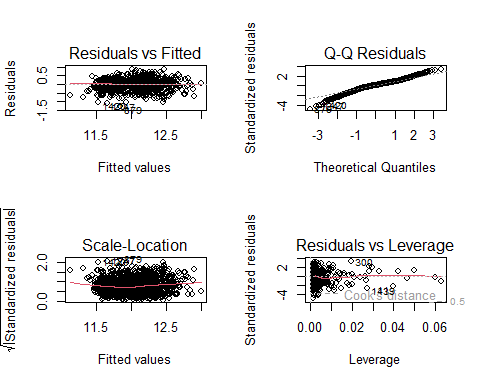
trainNoMissing$PredictedLogSalePrice\_MLM <- predict(multipleLMmodel)  
trainNoMissing$PredictedSalePrice\_MLM <- exp(trainNoMissing$PredictedLogSalePrice\_MLM)  
  
# Find CV PRESS  
cvpress\_MLM <- sum((trainNoMissing$SalePrice - trainNoMissing$PredictedSalePrice\_MLM)^2)  
print(paste("CV PRESS:", cvpress\_MLM))

## [1] "CV PRESS: 3890339503310.41"

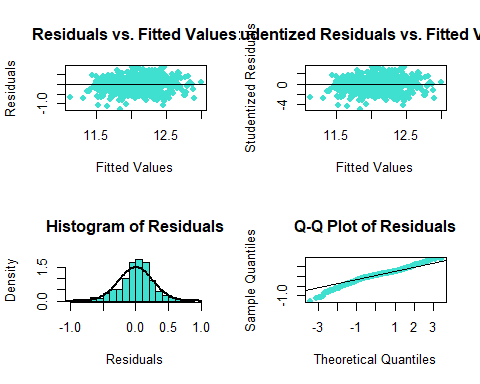
# Calculate AIC   
aic\_MLM <- AIC(multipleLMmodel)  
print(paste("AIC:", aic\_MLM))

## [1] "AIC: 261.681213640679"

par(mfrow = c(2, 2)) # Arrange the plots in a 2x2 grid  
plot(multipleLMmodel)



par(mfrow = c(2, 2))  
  
# Residuals vs Fitted values  
plot(fitted(multipleLMmodel), residuals(multipleLMmodel),  
 xlab = "Fitted Values", ylab = "Residuals", main = "Residuals vs. Fitted Values", pch = 19, col="turquoise")  
abline(h = 0, col = "black")  
  
 # Studentized Residuals  
plot(fitted(multipleLMmodel), rstudent(multipleLMmodel),  
 xlab = "Fitted Values", ylab = "Studentized Residuals", main = "Studentized Residuals vs. Fitted Values",   
 pch = 19, col="turquoise")  
abline(h = 0, col = "black")  
   
# Histogram of residuals  
hist(residuals(multipleLMmodel),  
 xlab = "Residuals", main = "Histogram of Residuals", col = "turquoise", border = "black", breaks = 20, probability=TRUE, xlim = c(-1, 1))  
curve(dnorm(x, mean = mean(residuals(multipleLMmodel)), sd = sd(residuals(multipleLMmodel))), col = "black", lwd = 2, add = TRUE)  
  
# Q-Q plot of residuals  
qqnorm(residuals(multipleLMmodel), main = "Q-Q Plot of Residuals", col="turquoise", pch=19)   
qqline(residuals(multipleLMmodel), col = "black")



bigmultipleLMmodel <- lm(LogSalePrice ~ OverallQual \* AllSQFT \* GarageCars \*   
 KitchenQual\* ExterQual\* X1stFlrSF , data = trainNoMissing)  
  
bigMLM\_summary <- summary(bigmultipleLMmodel)  
bigMLM\_summary

##   
## Call:  
## lm(formula = LogSalePrice ~ OverallQual \* AllSQFT \* GarageCars \*   
## KitchenQual \* ExterQual \* X1stFlrSF, data = trainNoMissing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.92032 -0.07203 0.01304 0.09321 0.54144   
##   
## Coefficients:  
## Estimate  
## (Intercept) 1.797e+01  
## OverallQual -1.744e+00  
## AllSQFT -2.322e-03  
## GarageCars -1.466e+00  
## KitchenQual -3.049e+00  
## ExterQual -3.676e+00  
## X1stFlrSF -1.375e-02  
## OverallQual:AllSQFT 4.815e-04  
## OverallQual:GarageCars 6.192e-01  
## AllSQFT:GarageCars 7.574e-04  
## OverallQual:KitchenQual 6.105e-01  
## AllSQFT:KitchenQual 5.366e-04  
## GarageCars:KitchenQual 2.545e+00  
## OverallQual:ExterQual 8.289e-01  
## AllSQFT:ExterQual 1.407e-03  
## GarageCars:ExterQual 2.823e-01  
## KitchenQual:ExterQual 1.192e+00  
## OverallQual:X1stFlrSF 2.892e-03  
## AllSQFT:X1stFlrSF 4.606e-06  
## GarageCars:X1stFlrSF 1.949e-03  
## KitchenQual:X1stFlrSF 5.067e-03  
## ExterQual:X1stFlrSF 6.013e-03  
## OverallQual:AllSQFT:GarageCars -1.660e-04  
## OverallQual:AllSQFT:KitchenQual -5.277e-05  
## OverallQual:GarageCars:KitchenQual -4.765e-01  
## AllSQFT:GarageCars:KitchenQual -5.976e-04  
## OverallQual:AllSQFT:ExterQual -2.391e-04  
## OverallQual:GarageCars:ExterQual -1.994e-01  
## AllSQFT:GarageCars:ExterQual -1.680e-04  
## OverallQual:KitchenQual:ExterQual -2.398e-01  
## AllSQFT:KitchenQual:ExterQual -2.607e-04  
## GarageCars:KitchenQual:ExterQual -7.381e-01  
## OverallQual:AllSQFT:X1stFlrSF -8.404e-07  
## OverallQual:GarageCars:X1stFlrSF -7.289e-04  
## AllSQFT:GarageCars:X1stFlrSF -8.801e-07  
## OverallQual:KitchenQual:X1stFlrSF -9.261e-04  
## AllSQFT:KitchenQual:X1stFlrSF -1.314e-06  
## GarageCars:KitchenQual:X1stFlrSF -2.010e-03  
## OverallQual:ExterQual:X1stFlrSF -1.227e-03  
## AllSQFT:ExterQual:X1stFlrSF -2.030e-06  
## GarageCars:ExterQual:X1stFlrSF -6.030e-04  
## KitchenQual:ExterQual:X1stFlrSF -1.918e-03  
## OverallQual:AllSQFT:GarageCars:KitchenQual 8.377e-05  
## OverallQual:AllSQFT:GarageCars:ExterQual 4.657e-05  
## OverallQual:AllSQFT:KitchenQual:ExterQual 2.593e-05  
## OverallQual:GarageCars:KitchenQual:ExterQual 1.520e-01  
## AllSQFT:GarageCars:KitchenQual:ExterQual 1.441e-04  
## OverallQual:AllSQFT:GarageCars:X1stFlrSF 2.170e-07  
## OverallQual:AllSQFT:KitchenQual:X1stFlrSF 1.981e-07  
## OverallQual:GarageCars:KitchenQual:X1stFlrSF 4.042e-04  
## AllSQFT:GarageCars:KitchenQual:X1stFlrSF 5.103e-07  
## OverallQual:AllSQFT:ExterQual:X1stFlrSF 3.585e-07  
## OverallQual:GarageCars:ExterQual:X1stFlrSF 2.703e-04  
## AllSQFT:GarageCars:ExterQual:X1stFlrSF 2.853e-07  
## OverallQual:KitchenQual:ExterQual:X1stFlrSF 3.600e-04  
## AllSQFT:KitchenQual:ExterQual:X1stFlrSF 5.051e-07  
## GarageCars:KitchenQual:ExterQual:X1stFlrSF 6.373e-04  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual -1.938e-05  
## OverallQual:AllSQFT:GarageCars:KitchenQual:X1stFlrSF -8.546e-08  
## OverallQual:AllSQFT:GarageCars:ExterQual:X1stFlrSF -7.648e-08  
## OverallQual:AllSQFT:KitchenQual:ExterQual:X1stFlrSF -7.583e-08  
## OverallQual:GarageCars:KitchenQual:ExterQual:X1stFlrSF -1.420e-04  
## AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF -1.532e-07  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF 2.665e-08  
## Std. Error  
## (Intercept) 1.592e+01  
## OverallQual 2.343e+00  
## AllSQFT 6.059e-03  
## GarageCars 5.318e+00  
## KitchenQual 4.474e+00  
## ExterQual 5.242e+00  
## X1stFlrSF 1.717e-02  
## OverallQual:AllSQFT 8.521e-04  
## OverallQual:GarageCars 7.138e-01  
## AllSQFT:GarageCars 2.008e-03  
## OverallQual:KitchenQual 6.966e-01  
## AllSQFT:KitchenQual 1.833e-03  
## GarageCars:KitchenQual 1.602e+00  
## OverallQual:ExterQual 7.912e-01  
## AllSQFT:ExterQual 1.998e-03  
## GarageCars:ExterQual 1.740e+00  
## KitchenQual:ExterQual 1.405e+00  
## OverallQual:X1stFlrSF 2.346e-03  
## AllSQFT:X1stFlrSF 5.191e-06  
## GarageCars:X1stFlrSF 5.819e-03  
## KitchenQual:X1stFlrSF 4.728e-03  
## ExterQual:X1stFlrSF 5.756e-03  
## OverallQual:AllSQFT:GarageCars 2.616e-04  
## OverallQual:AllSQFT:KitchenQual 2.662e-04  
## OverallQual:GarageCars:KitchenQual 2.238e-01  
## AllSQFT:GarageCars:KitchenQual 6.532e-04  
## OverallQual:AllSQFT:ExterQual 2.906e-04  
## OverallQual:GarageCars:ExterQual 2.419e-01  
## AllSQFT:GarageCars:ExterQual 6.647e-04  
## OverallQual:KitchenQual:ExterQual 2.186e-01  
## AllSQFT:KitchenQual:ExterQual 5.750e-04  
## GarageCars:KitchenQual:ExterQual 4.953e-01  
## OverallQual:AllSQFT:X1stFlrSF 6.912e-07  
## OverallQual:GarageCars:X1stFlrSF 7.459e-04  
## AllSQFT:GarageCars:X1stFlrSF 1.689e-06  
## OverallQual:KitchenQual:X1stFlrSF 6.952e-04  
## AllSQFT:KitchenQual:X1stFlrSF 1.496e-06  
## GarageCars:KitchenQual:X1stFlrSF 1.629e-03  
## OverallQual:ExterQual:X1stFlrSF 8.102e-04  
## AllSQFT:ExterQual:X1stFlrSF 1.718e-06  
## GarageCars:ExterQual:X1stFlrSF 1.959e-03  
## KitchenQual:ExterQual:X1stFlrSF 1.528e-03  
## OverallQual:AllSQFT:GarageCars:KitchenQual 8.380e-05  
## OverallQual:AllSQFT:GarageCars:ExterQual 8.873e-05  
## OverallQual:AllSQFT:KitchenQual:ExterQual 8.465e-05  
## OverallQual:GarageCars:KitchenQual:ExterQual 6.728e-02  
## AllSQFT:GarageCars:KitchenQual:ExterQual 2.056e-04  
## OverallQual:AllSQFT:GarageCars:X1stFlrSF 2.129e-07  
## OverallQual:AllSQFT:KitchenQual:X1stFlrSF 2.067e-07  
## OverallQual:GarageCars:KitchenQual:X1stFlrSF 2.221e-04  
## AllSQFT:GarageCars:KitchenQual:X1stFlrSF 4.967e-07  
## OverallQual:AllSQFT:ExterQual:X1stFlrSF 2.373e-07  
## OverallQual:GarageCars:ExterQual:X1stFlrSF 2.606e-04  
## AllSQFT:GarageCars:ExterQual:X1stFlrSF 5.594e-07  
## OverallQual:KitchenQual:ExterQual:X1stFlrSF 2.244e-04  
## AllSQFT:KitchenQual:ExterQual:X1stFlrSF 4.745e-07  
## GarageCars:KitchenQual:ExterQual:X1stFlrSF 5.233e-04  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual 2.610e-05  
## OverallQual:AllSQFT:GarageCars:KitchenQual:X1stFlrSF 6.306e-08  
## OverallQual:AllSQFT:GarageCars:ExterQual:X1stFlrSF 7.330e-08  
## OverallQual:AllSQFT:KitchenQual:ExterQual:X1stFlrSF 6.577e-08  
## OverallQual:GarageCars:KitchenQual:ExterQual:X1stFlrSF 7.015e-05  
## AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF 1.571e-07  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF 1.973e-08  
## t value Pr(>|t|)  
## (Intercept) 1.129 0.2590  
## OverallQual -0.744 0.4567  
## AllSQFT -0.383 0.7016  
## GarageCars -0.276 0.7828  
## KitchenQual -0.681 0.4957  
## ExterQual -0.701 0.4833  
## X1stFlrSF -0.801 0.4234  
## OverallQual:AllSQFT 0.565 0.5721  
## OverallQual:GarageCars 0.867 0.3859  
## AllSQFT:GarageCars 0.377 0.7060  
## OverallQual:KitchenQual 0.876 0.3810  
## AllSQFT:KitchenQual 0.293 0.7697  
## GarageCars:KitchenQual 1.589 0.1123  
## OverallQual:ExterQual 1.048 0.2950  
## AllSQFT:ExterQual 0.704 0.4816  
## GarageCars:ExterQual 0.162 0.8711  
## KitchenQual:ExterQual 0.848 0.3963  
## OverallQual:X1stFlrSF 1.233 0.2179  
## AllSQFT:X1stFlrSF 0.887 0.3751  
## GarageCars:X1stFlrSF 0.335 0.7377  
## KitchenQual:X1stFlrSF 1.072 0.2840  
## ExterQual:X1stFlrSF 1.045 0.2964  
## OverallQual:AllSQFT:GarageCars -0.635 0.5258  
## OverallQual:AllSQFT:KitchenQual -0.198 0.8429  
## OverallQual:GarageCars:KitchenQual -2.129 0.0334  
## AllSQFT:GarageCars:KitchenQual -0.915 0.3604  
## OverallQual:AllSQFT:ExterQual -0.823 0.4107  
## OverallQual:GarageCars:ExterQual -0.824 0.4099  
## AllSQFT:GarageCars:ExterQual -0.253 0.8005  
## OverallQual:KitchenQual:ExterQual -1.097 0.2729  
## AllSQFT:KitchenQual:ExterQual -0.453 0.6503  
## GarageCars:KitchenQual:ExterQual -1.490 0.1364  
## OverallQual:AllSQFT:X1stFlrSF -1.216 0.2243  
## OverallQual:GarageCars:X1stFlrSF -0.977 0.3286  
## AllSQFT:GarageCars:X1stFlrSF -0.521 0.6024  
## OverallQual:KitchenQual:X1stFlrSF -1.332 0.1830  
## AllSQFT:KitchenQual:X1stFlrSF -0.878 0.3800  
## GarageCars:KitchenQual:X1stFlrSF -1.234 0.2174  
## OverallQual:ExterQual:X1stFlrSF -1.514 0.1302  
## AllSQFT:ExterQual:X1stFlrSF -1.182 0.2375  
## GarageCars:ExterQual:X1stFlrSF -0.308 0.7583  
## KitchenQual:ExterQual:X1stFlrSF -1.256 0.2094  
## OverallQual:AllSQFT:GarageCars:KitchenQual 1.000 0.3177  
## OverallQual:AllSQFT:GarageCars:ExterQual 0.525 0.5998  
## OverallQual:AllSQFT:KitchenQual:ExterQual 0.306 0.7594  
## OverallQual:GarageCars:KitchenQual:ExterQual 2.259 0.0241  
## AllSQFT:GarageCars:KitchenQual:ExterQual 0.701 0.4835  
## OverallQual:AllSQFT:GarageCars:X1stFlrSF 1.019 0.3082  
## OverallQual:AllSQFT:KitchenQual:X1stFlrSF 0.959 0.3379  
## OverallQual:GarageCars:KitchenQual:X1stFlrSF 1.820 0.0690  
## AllSQFT:GarageCars:KitchenQual:X1stFlrSF 1.027 0.3044  
## OverallQual:AllSQFT:ExterQual:X1stFlrSF 1.511 0.1310  
## OverallQual:GarageCars:ExterQual:X1stFlrSF 1.037 0.2998  
## AllSQFT:GarageCars:ExterQual:X1stFlrSF 0.510 0.6102  
## OverallQual:KitchenQual:ExterQual:X1stFlrSF 1.604 0.1089  
## AllSQFT:KitchenQual:ExterQual:X1stFlrSF 1.065 0.2872  
## GarageCars:KitchenQual:ExterQual:X1stFlrSF 1.218 0.2235  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual -0.743 0.4579  
## OverallQual:AllSQFT:GarageCars:KitchenQual:X1stFlrSF -1.355 0.1756  
## OverallQual:AllSQFT:GarageCars:ExterQual:X1stFlrSF -1.043 0.2970  
## OverallQual:AllSQFT:KitchenQual:ExterQual:X1stFlrSF -1.153 0.2492  
## OverallQual:GarageCars:KitchenQual:ExterQual:X1stFlrSF -2.025 0.0431  
## AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF -0.975 0.3297  
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF 1.351 0.1770  
##   
## (Intercept)   
## OverallQual   
## AllSQFT   
## GarageCars   
## KitchenQual   
## ExterQual   
## X1stFlrSF   
## OverallQual:AllSQFT   
## OverallQual:GarageCars   
## AllSQFT:GarageCars   
## OverallQual:KitchenQual   
## AllSQFT:KitchenQual   
## GarageCars:KitchenQual   
## OverallQual:ExterQual   
## AllSQFT:ExterQual   
## GarageCars:ExterQual   
## KitchenQual:ExterQual   
## OverallQual:X1stFlrSF   
## AllSQFT:X1stFlrSF   
## GarageCars:X1stFlrSF   
## KitchenQual:X1stFlrSF   
## ExterQual:X1stFlrSF   
## OverallQual:AllSQFT:GarageCars   
## OverallQual:AllSQFT:KitchenQual   
## OverallQual:GarageCars:KitchenQual \*  
## AllSQFT:GarageCars:KitchenQual   
## OverallQual:AllSQFT:ExterQual   
## OverallQual:GarageCars:ExterQual   
## AllSQFT:GarageCars:ExterQual   
## OverallQual:KitchenQual:ExterQual   
## AllSQFT:KitchenQual:ExterQual   
## GarageCars:KitchenQual:ExterQual   
## OverallQual:AllSQFT:X1stFlrSF   
## OverallQual:GarageCars:X1stFlrSF   
## AllSQFT:GarageCars:X1stFlrSF   
## OverallQual:KitchenQual:X1stFlrSF   
## AllSQFT:KitchenQual:X1stFlrSF   
## GarageCars:KitchenQual:X1stFlrSF   
## OverallQual:ExterQual:X1stFlrSF   
## AllSQFT:ExterQual:X1stFlrSF   
## GarageCars:ExterQual:X1stFlrSF   
## KitchenQual:ExterQual:X1stFlrSF   
## OverallQual:AllSQFT:GarageCars:KitchenQual   
## OverallQual:AllSQFT:GarageCars:ExterQual   
## OverallQual:AllSQFT:KitchenQual:ExterQual   
## OverallQual:GarageCars:KitchenQual:ExterQual \*  
## AllSQFT:GarageCars:KitchenQual:ExterQual   
## OverallQual:AllSQFT:GarageCars:X1stFlrSF   
## OverallQual:AllSQFT:KitchenQual:X1stFlrSF   
## OverallQual:GarageCars:KitchenQual:X1stFlrSF .  
## AllSQFT:GarageCars:KitchenQual:X1stFlrSF   
## OverallQual:AllSQFT:ExterQual:X1stFlrSF   
## OverallQual:GarageCars:ExterQual:X1stFlrSF   
## AllSQFT:GarageCars:ExterQual:X1stFlrSF   
## OverallQual:KitchenQual:ExterQual:X1stFlrSF   
## AllSQFT:KitchenQual:ExterQual:X1stFlrSF   
## GarageCars:KitchenQual:ExterQual:X1stFlrSF   
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual   
## OverallQual:AllSQFT:GarageCars:KitchenQual:X1stFlrSF   
## OverallQual:AllSQFT:GarageCars:ExterQual:X1stFlrSF   
## OverallQual:AllSQFT:KitchenQual:ExterQual:X1stFlrSF   
## OverallQual:GarageCars:KitchenQual:ExterQual:X1stFlrSF \*  
## AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF   
## OverallQual:AllSQFT:GarageCars:KitchenQual:ExterQual:X1stFlrSF   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1518 on 1377 degrees of freedom  
## Multiple R-squared: 0.856, Adjusted R-squared: 0.8494   
## F-statistic: 129.9 on 63 and 1377 DF, p-value: < 2.2e-16

trainNoMissing$PredictedLogSalePrice\_bigMLM <- predict(bigmultipleLMmodel)  
trainNoMissing$PredictedSalePrice\_bigMLM <- exp(trainNoMissing$PredictedLogSalePrice\_bigMLM)  
  
# Find CV PRESS  
cvpress\_bigMLM <- sum((trainNoMissing$SalePrice - trainNoMissing$PredictedSalePrice\_bigMLM)^2)  
print(paste("CV PRESS:", cvpress\_bigMLM))

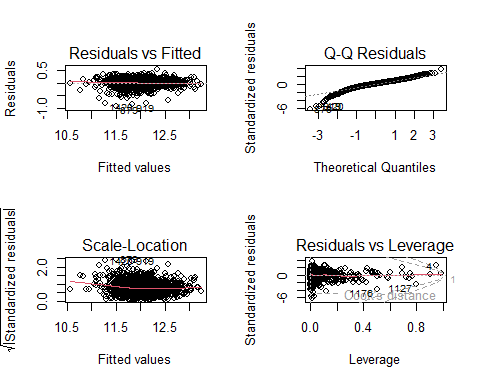
## [1] "CV PRESS: 931032752785.549"

# Calculate AIC   
aic\_bigMLM <- AIC(bigmultipleLMmodel)  
print(paste("AIC:", aic\_bigMLM))

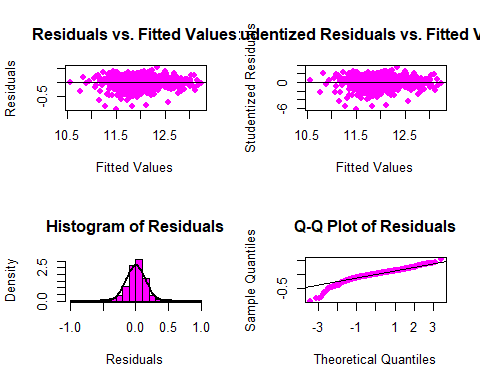
## [1] "AIC: -1278.34799175159"

par(mfrow = c(2, 2)) # Arrange the plots in a 2x2 grid  
plot(bigmultipleLMmodel)

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



par(mfrow = c(2, 2))  
  
# Residuals vs Fitted values  
plot(fitted(bigmultipleLMmodel), residuals(bigmultipleLMmodel),  
 xlab = "Fitted Values", ylab = "Residuals", main = "Residuals vs. Fitted Values", pch = 19, col="magenta")  
abline(h = 0, col = "black")  
  
 # Studentized Residuals  
plot(fitted(bigmultipleLMmodel), rstudent(bigmultipleLMmodel),  
 xlab = "Fitted Values", ylab = "Studentized Residuals", main = "Studentized Residuals vs. Fitted Values",   
 pch = 19, col="magenta")  
abline(h = 0, col = "black")  
   
# Histogram of residuals  
hist(residuals(bigmultipleLMmodel),  
 xlab = "Residuals", main = "Histogram of Residuals", col = "magenta", border = "black", breaks = 20, probability=TRUE, xlim = c(-1, 1))  
curve(dnorm(x, mean = mean(residuals(bigmultipleLMmodel)), sd = sd(residuals(bigmultipleLMmodel))), col = "black", lwd = 2, add = TRUE)  
  
# Q-Q plot of residuals  
qqnorm(residuals(bigmultipleLMmodel), main = "Q-Q Plot of Residuals", col="magenta", pch=19)   
qqline(residuals(bigmultipleLMmodel), col = "black")



# Finalize  
testNoMissing$PredictedLogSalePrice\_Qual <- predict(model\_qual,newdata=testNoMissing)  
testNoMissing$SalePrice\_Qual <- exp(testNoMissing$PredictedLogSalePrice\_Qual)  
  
testNoMissing$PredictedLogSalePrice\_MLM <- predict(multipleLMmodel,newdata=testNoMissing)  
testNoMissing$SalePrice\_MLM <- exp(testNoMissing$PredictedLogSalePrice\_MLM)  
  
testNoMissing$PredictedLogSalePrice\_bigMLM <- predict(bigmultipleLMmodel,newdata=testNoMissing)  
testNoMissing$SalePrice\_bigMLM <- exp(testNoMissing$PredictedLogSalePrice\_bigMLM)

# Export  
write.csv(data.frame(Id = testNoMissing$Id,   
 SalePrice = testNoMissing$SalePrice\_Qual),   
 file = "submission\_Qual.csv", row.names = FALSE)  
write.csv(data.frame(Id = testNoMissing$Id,   
 SalePrice = testNoMissing$SalePrice\_MLM),   
 file = "submission\_MLM.csv", row.names = FALSE)  
write.csv(data.frame(Id = testNoMissing$Id,   
 SalePrice = testNoMissing$SalePrice\_bigMLM),   
 file = "submission\_bigMLM.csv", row.names = FALSE)