

# FinalProj

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3    v purrr  0.3.4
## v tibble  3.1.0    v dplyr  1.0.5
## v tidyr   1.1.3    v stringr 1.4.0
## v readr   1.4.0    v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
raw_data <- read.csv("train.csv")
```

## Data cleaning

```
colSums(is.na(raw_data))
```

##	Id	MSSubClass	MSZoning	LotFrontage	LotArea
##	0	0	0	259	0
##	Street	Alley	LotShape	LandContour	Utilities
##	0	1369	0	0	0
##	LotConfig	LandSlope	Neighborhood	Condition1	Condition2
##	0	0	0	0	0
##	BldgType	HouseStyle	OverallQual	OverallCond	YearBuilt
##	0	0	0	0	0
##	YearRemodAdd	RoofStyle	RoofMatl	Exterior1st	Exterior2nd
##	0	0	0	0	0
##	MasVnrType	MasVnrArea	ExterQual	ExterCond	Foundation
##	8	8	0	0	0
##	BsmtQual	BsmtCond	BsmtExposure	BsmtFinType1	BsmtFinSF1
##	37	37	38	37	0
##	BsmtFinType2	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF	Heating
##	38	0	0	0	0
##	HeatingQC	CentralAir	Electrical	X1stFlrSF	X2ndFlrSF
##	0	0	1	0	0
##	LowQualFinSF	GrLivArea	BsmtFullBath	BsmtHalfBath	FullBath
##	0	0	0	0	0
##	HalfBath	BedroomAbvGr	KitchenAbvGr	KitchenQual	TotRmsAbvGrd
##	0	0	0	0	0
##	Functional	Fireplaces	FireplaceQu	GarageType	GarageYrBlt

```
##          0          0          690          81          81
## GarageFinish GarageCars GarageArea GarageQual GarageCond
##          81          0          0          81          81
## PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch
##          0          0          0          0          0
## ScreenPorch PoolArea PoolQC Fence MiscFeature
##          0          0          1453          1179          1406
## MiscVal MoSold YrSold SaleType SaleCondition
##          0          0          0          0          0
## SalePrice
##          0
```

```
data <- raw_data %>% replace_na(list(PoolQC = "None",
                                     Fence = "None",
                                     MiscFeature = "None"))
data <- data %>% replace_na(list(GarageFinish = "None",
                                 GarageQual = "None",
                                 GarageCond = "None",
                                 GarageType = "None",
                                 GarageYrBlt = 0,
                                 FireplaceQu = "None"))
data <- data %>% replace_na(list(Electrical = "None",
                                 BsmtFinType2 = "None",
                                 BsmtFinType1 = "None",
                                 BsmtExposure = "None",
                                 BsmtQual = "None",
                                 BsmtCond = "None"))
data <- data %>% replace_na(list(MasVnrArea = 0,
                                 MasVnrType = "None",
                                 Alley = "None",
                                 LotFrontage = 0))

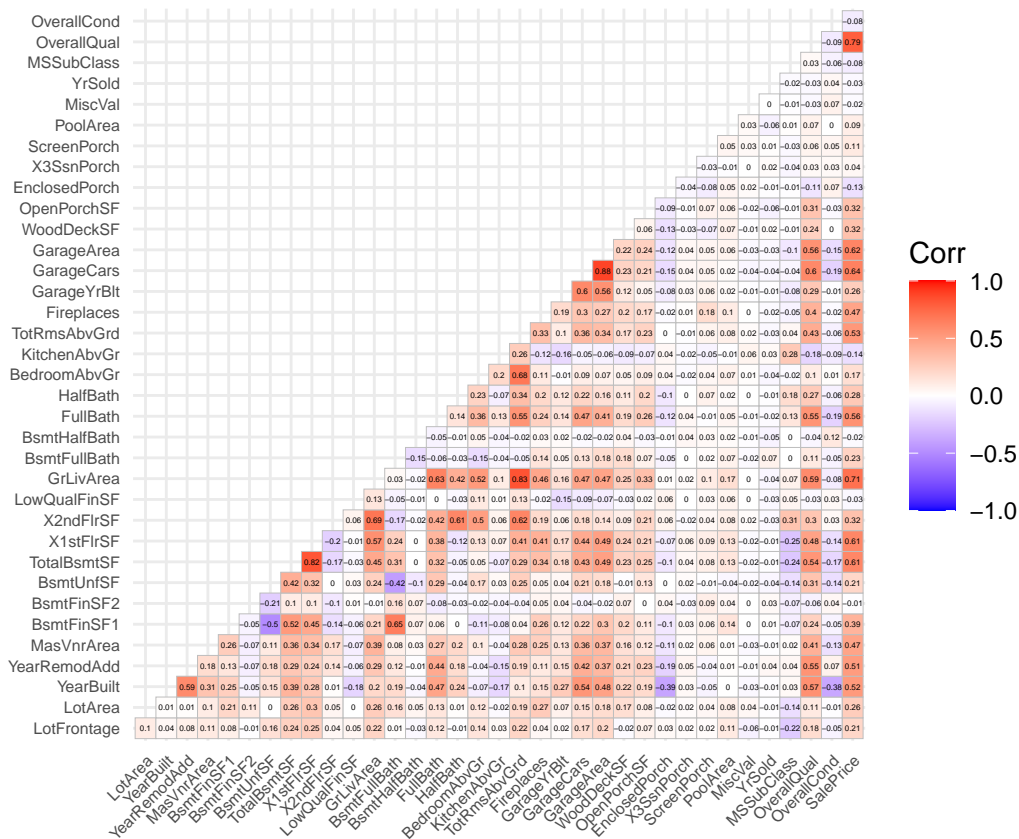
colSums(is.na(data))
```

```
##          Id MSSubClass MSZoning LotFrontage LotArea
##          0          0          0          0          0
## Street Alley LotShape LandContour Utilities
##          0          0          0          0          0
## LotConfig LandSlope Neighborhood Condition1 Condition2
##          0          0          0          0          0
## BldgType HouseStyle OverallQual OverallCond YearBuilt
##          0          0          0          0          0
## YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd
##          0          0          0          0          0
## MasVnrType MasVnrArea ExterQual ExterCond Foundation
##          0          0          0          0          0
## BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1
##          0          0          0          0          0
## BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating
##          0          0          0          0          0
## HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF
##          0          0          0          0          0
## LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath
##          0          0          0          0          0
## HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd
```

```
##          0          0          0          0          0
##   Functional   Fireplaces   FireplaceQu   GarageType   GarageYrBlt
##          0          0          0          0          0
##   GarageFinish   GarageCars   GarageArea   GarageQual   GarageCond
##          0          0          0          0          0
##   PavedDrive   WoodDeckSF   OpenPorchSF   EnclosedPorch   X3SsnPorch
##          0          0          0          0          0
##   ScreenPorch   PoolArea   PoolQC   Fence   MiscFeature
##          0          0          0          0          0
##   MiscVal   MoSold   YrSold   SaleType   SaleCondition
##          0          0          0          0          0
##   SalePrice
##          0
```

## Data exploration

```
library(ggplot2)
library(ggcorrplot)
data_corr <- data %>% select(LotFrontage, LotArea, YearBuilt,
                             YearRemodAdd, MasVnrArea, BsmtFinSF1,
                             BsmtFinSF2, BsmtUnfSF, TotalBsmtSF,
                             X1stFlrSF, X2ndFlrSF, LowQualFinSF,
                             GrLivArea, BsmtFullBath, BsmtHalfBath,
                             FullBath, HalfBath, BedroomAbvGr,
                             KitchenAbvGr, TotRmsAbvGrd, Fireplaces,
                             GarageYrBlt, GarageCars, GarageArea,
                             WoodDeckSF, OpenPorchSF, EnclosedPorch,
                             X3SsnPorch, ScreenPorch, PoolArea,
                             MiscVal, YrSold, MSSubClass,
                             OverallQual, OverallCond, SalePrice)
correlations <- cor(data_corr)
ggcorrplot(round(correlations,2), lab = TRUE, tl.cex = 6, lab_size = 1, type = "lower")
```



Selected all continuous variables and the three ordinal variables (MSSubClass, OverallQual and OverallCond) for the correlation plot. The last column SalePrice is the most important in the plot, we can see which variables are most correlated to the target and explain it the most. Overall Quality and total above ground square feet are the most correlated.

## Lasso

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
```

```
##
```

```
## expand, pack, unpack
```

```
## Loaded glmnet 4.1-1
```

```
data <- data %>% mutate_if(is.character, as.factor)
data$MSSubClass <- as.factor(data$MSSubClass)
```

```

#data <- data[-1299,]

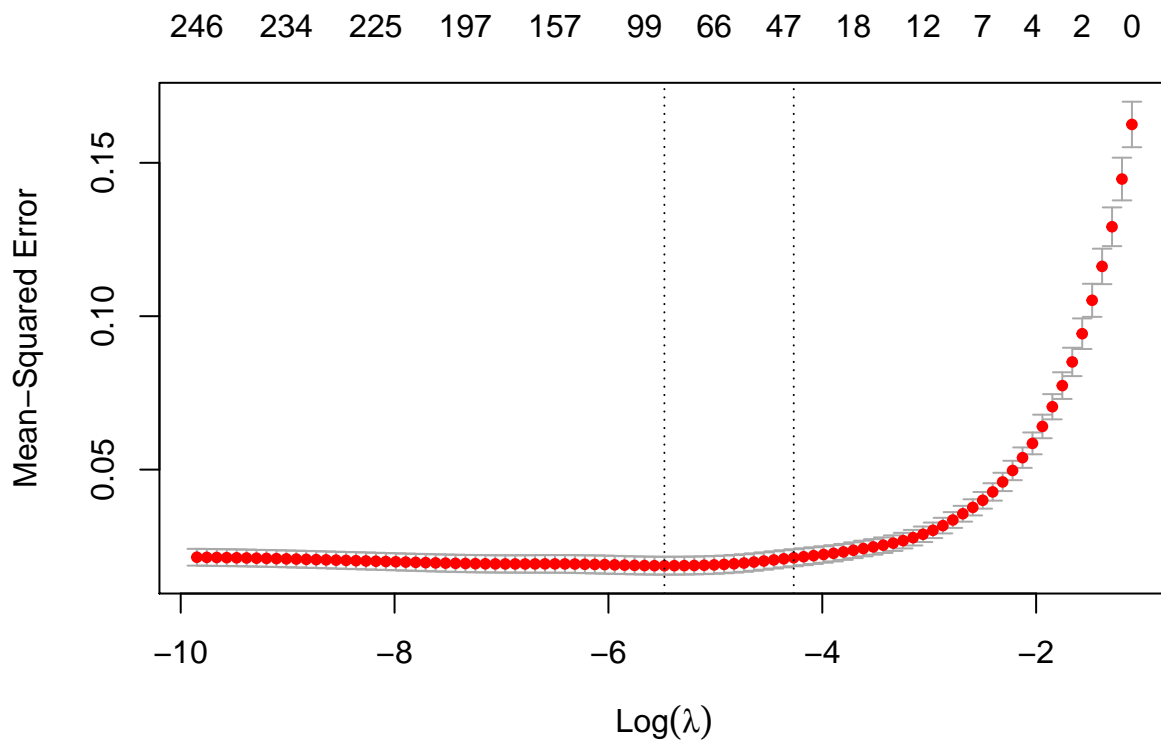
train <- data[1:1100,]
train <- train[,-1]
test <- data[1101:1459,]
test <- test[,-1]

train$SalePrice <- log(train$SalePrice)
test$SalePrice <- log(test$SalePrice)

#rain_x <- model.matrix(train[, -c(1,81)])
train_x <- model.matrix(SalePrice ~ ., train)[,-81]
#test_x <- model.matrix(test[, -c(1,81)])
test_x <- model.matrix(SalePrice ~ ., test)[,-81]

lasso_cv <- cv.glmnet(train_x, train$SalePrice, alpha = 1)
plot(lasso_cv)

```



```
lasso_cv$lambda.min
```

```
## [1] 0.004183843
```

```

lasso.pred=predict(lasso_cv,s=lasso_cv$lambda.min,newx=test_x)
#mean( (exp(lasso.pred)-exp(test$SalePrice))^2 )

```

```

#diff <- exp(lasso.pred)-exp(test$SalePrice)

diff <- data.frame(predictions = lasso.pred)
diff$residuals <- lasso.pred - test$SalePrice

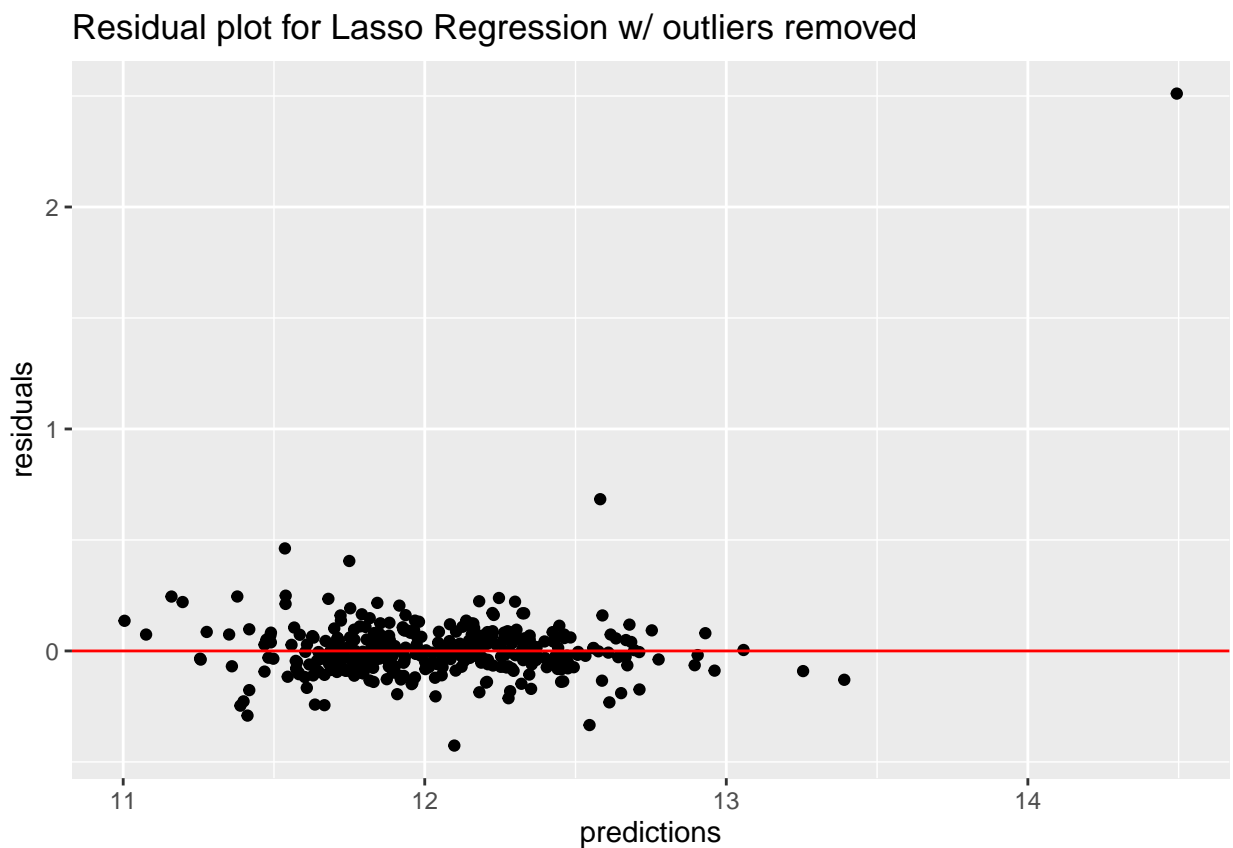
diff$residab <- abs(diff$residuals)
diff$sq <- (diff$residuals * diff$residuals)

mean(diff$residab)

## [1] 0.08459124

library(ggplot2)
ggplot(data = diff) + geom_point(mapping = aes(x=X1, y=residuals)) +
  geom_abline(intercept = 0, slope = 0, colour = "red") +
  ggtitle("Residual plot for Lasso Regression w/ outliers removed") +
  xlab("predictions")

```



```

model <- glmnet(train_x,train$SalePrice, alpha = 1, lambda = lasso_cv$lambda.min)
coef(model)

```

```

## 274 x 1 sparse Matrix of class "dgCMatrix"
##                                     s0

```

## (Intercept)	5.748160e+00
## (Intercept)	.
## MSSubClass30	-3.961291e-02
## MSSubClass40	.
## MSSubClass45	.
## MSSubClass50	.
## MSSubClass60	.
## MSSubClass70	.
## MSSubClass75	.
## MSSubClass80	.
## MSSubClass85	.
## MSSubClass90	.
## MSSubClass120	.
## MSSubClass160	-6.541399e-02
## MSSubClass180	.
## MSSubClass190	.
## MSZoningFV	1.832003e-03
## MSZoningRH	.
## MSZoningRL	5.011920e-02
## MSZoningRM	.
## LotFrontage	.
## LotArea	1.538890e-06
## StreetPave	8.016925e-02
## AlleyNone	.
## AlleyPave	.
## LotShapeIR2	1.702388e-02
## LotShapeIR3	.
## LotShapeReg	.
## LandContourHLS	.
## LandContourLow	.
## LandContourLvl	.
## UtilitiesNoSeWa	.
## LotConfigCulDSac	1.248025e-02
## LotConfigFR2	.
## LotConfigFR3	.
## LotConfigInside	.
## LandSlopeMod	4.326721e-03
## LandSlopeSev	.
## NeighborhoodBlueste	.
## NeighborhoodBrDale	.
## NeighborhoodBrkSide	3.728198e-03
## NeighborhoodClearCr	2.339313e-02
## NeighborhoodCollgCr	.
## NeighborhoodCrawfor	9.808097e-02
## NeighborhoodEdwards	-3.286317e-02
## NeighborhoodGilbert	.
## NeighborhoodIDOTRR	-4.214798e-02
## NeighborhoodMeadowV	-9.053610e-02
## NeighborhoodMitchel	.
## NeighborhoodNames	.
## NeighborhoodNoRidge	2.975008e-02
## NeighborhoodNPkVill	.
## NeighborhoodNridgHt	8.206548e-02
## NeighborhoodNWAmes	.

```

## NeighborhoodOldTown -1.072421e-02
## NeighborhoodSawyer .
## NeighborhoodSawyerW .
## NeighborhoodSomerst 8.768434e-02
## NeighborhoodStoneBr 9.488699e-02
## NeighborhoodSWISU .
## NeighborhoodTimber .
## NeighborhoodVeenker 1.042142e-02
## Condition1Feedr .
## Condition1Norm 3.188595e-02
## Condition1PosA .
## Condition1PosN .
## Condition1RR Ae -3.067210e-02
## Condition1RR An .
## Condition1RR Ne .
## Condition1RR Nn .
## Condition2Feedr .
## Condition2Norm .
## Condition2PosA .
## Condition2PosN -7.116995e-01
## Condition2RR Ae .
## Condition2RR An .
## Condition2RR Nn .
## BldgType2fmCon .
## BldgTypeDuplex .
## BldgTypeTwnhs -4.931543e-02
## BldgTypeTwnhsE -8.825215e-03
## HouseStyle1Story .
## HouseStyle2.5Fin -8.631377e-03
## HouseStyle2.5Unf .
## HouseStyle2Story .
## HouseStyleSFoyer .
## HouseStyleSLvl .
## OverallQual 6.280710e-02
## OverallCond 2.878900e-02
## YearBuilt 1.235492e-03
## YearRemodAdd 1.276051e-03
## RoofStyleGable .
## RoofStyleGambrel -2.400912e-02
## RoofStyleHip .
## RoofStyleMansard .
## RoofStyleShed .
## RoofMatlCompShg .
## RoofMatlMembran 5.039284e-02
## RoofMatlMetal .
## RoofMatlRoll .
## RoofMatlTar&Grv .
## RoofMatlWdShake .
## RoofMatlWdShngl 3.484990e-02
## Exterior1stAsphShn .
## Exterior1stBrkComm -1.299807e-01
## Exterior1stBrkFace 4.578482e-02
## Exterior1stCBlock .
## Exterior1stCemntBd .

```



```

## Exterior1stHdBoard      .
## Exterior1stImStucc      .
## Exterior1stMetalSd      .
## Exterior1stPlywood      .
## Exterior1stStone        .
## Exterior1stStucco       .
## Exterior1stVinylSd      2.882285e-03
## Exterior1stWd Sdng      -1.104193e-03
## Exterior1stWdShing      .
## Exterior2ndAsphShn      .
## Exterior2ndBrk Cmn      .
## Exterior2ndBrkFace      .
## Exterior2ndCBlock       .
## Exterior2ndCmentBd      .
## Exterior2ndHdBoard      -6.313256e-03
## Exterior2ndImStucc      .
## Exterior2ndMetalSd      .
## Exterior2ndOther        .
## Exterior2ndPlywood      .
## Exterior2ndStone        .
## Exterior2ndStucco       .
## Exterior2ndVinylSd      .
## Exterior2ndWd Sdng      .
## Exterior2ndWd Shng      .
## MasVnrTypeBrkFace      .
## MasVnrTypeNone          .
## MasVnrTypeStone        .
## MasVnrArea              1.666145e-05
## ExterQualFa              .
## ExterQualGd              .
## ExterQualTA              .
## ExterCondFa              -3.765403e-02
## ExterCondGd              .
## ExterCondPo              .
## ExterCondTA              2.153291e-03
## FoundationCBlock        .
## FoundationPConc          2.315067e-02
## FoundationSlab           .
## FoundationStone          2.756762e-02
## FoundationWood           .
## BsmtQualFa              .
## BsmtQualGd              .
## BsmtQualNone            .
## BsmtQualTA              -3.381389e-03
## BsmtCondGd              .
## BsmtCondNone            .
## BsmtCondPo              .
## BsmtCondTA              .
## BsmtExposureGd          3.531678e-02
## BsmtExposureMn          .
## BsmtExposureNo          -3.513075e-03
## BsmtExposureNone        .
## BsmtFinType1BLQ         .
## BsmtFinType1GLQ         .

```

## BsmtFinType1LwQ	.
## BsmtFinType1None	.
## BsmtFinType1Rec	.
## BsmtFinType1Unf	-8.954172e-03
## BsmtFinSF1	5.684726e-05
## BsmtFinType2BLQ	.
## BsmtFinType2GLQ	4.500908e-03
## BsmtFinType2LwQ	.
## BsmtFinType2None	.
## BsmtFinType2Rec	.
## BsmtFinType2Unf	.
## BsmtFinSF2	3.382297e-06
## BsmtUnfSF	.
## TotalBsmtSF	8.537073e-05
## HeatingGasA	.
## HeatingGasW	1.471582e-02
## HeatingGrav	-1.468618e-01
## HeatingOthW	.
## HeatingWall	.
## HeatingQCFA	.
## HeatingQCGd	-6.488856e-03
## HeatingQCPo	.
## HeatingQCTA	-1.967587e-02
## CentralAirY	5.183156e-02
## ElectricalFuseF	.
## ElectricalFuseP	.
## ElectricalMix	.
## ElectricalNone	.
## ElectricalSBrkr	.
## X1stFlrSF	2.022768e-05
## X2ndFlrSF	.
## LowQualFinSF	.
## GrLivArea	2.188346e-04
## BsmtFullBath	2.503864e-02
## BsmtHalfBath	.
## FullBath	2.300459e-02
## HalfBath	2.206891e-02
## BedroomAbvGr	.
## KitchenAbvGr	-6.047909e-02
## KitchenQualFa	.
## KitchenQualGd	.
## KitchenQualTA	-3.099993e-03
## TotRmsAbvGrd	6.541352e-03
## FunctionalMaj2	-1.573103e-01
## FunctionalMin1	.
## FunctionalMin2	.
## FunctionalMod	-3.190351e-02
## FunctionalSev	-2.202201e-01
## FunctionalTyp	3.230807e-02
## Fireplaces	2.060800e-02
## FireplaceQuFa	.
## FireplaceQuGd	.
## FireplaceQuNone	-7.397281e-03
## FireplaceQuPo	.

## FireplaceQuTA	.
## GarageTypeAttchd	.
## GarageTypeBasment	-1.730653e-02
## GarageTypeBuiltIn	.
## GarageTypeCarPort	.
## GarageTypeDetchd	.
## GarageTypeNone	.
## GarageYrBlt	.
## GarageFinishNone	.
## GarageFinishRFn	.
## GarageFinishUnf	-2.729123e-03
## GarageCars	1.617032e-02
## GarageArea	1.340128e-04
## GarageQualFa	.
## GarageQualGd	.
## GarageQualNone	.
## GarageQualPo	.
## GarageQualTA	.
## GarageCondFa	.
## GarageCondGd	.
## GarageCondNone	.
## GarageCondPo	.
## GarageCondTA	.
## PavedDriveP	.
## PavedDriveY	2.582137e-02
## WoodDeckSF	6.492487e-05
## OpenPorchSF	.
## EnclosedPorch	.
## X3SsnPorch	.
## ScreenPorch	1.837110e-04
## PoolArea	.
## PoolQCFa	.
## PoolQCGd	.
## PoolQCNone	.
## FenceGdWo	-1.530797e-02
## FenceMnPrv	.
## FenceMnWw	.
## FenceNone	2.845644e-03
## MiscFeatureNone	4.510267e-03
## MiscFeatureOthr	.
## MiscFeatureShed	.
## MiscFeatureTenC	.
## MiscVal	.
## MoSold	-1.043801e-03
## YrSold	-8.596522e-05
## SaleTypeCon	3.988120e-03
## SaleTypeConLD	.
## SaleTypeConLI	.
## SaleTypeConLw	.
## SaleTypeCWD	4.718206e-02
## SaleTypeNew	1.108120e-01
## SaleTypeOth	4.904411e-02
## SaleTypeWD	.
## SaleConditionAdjLand	.

```
## SaleConditionAlloca .
## SaleConditionFamily -1.686200e-02
## SaleConditionNormal 5.442650e-02
## SaleConditionPartial .
```

```
train_temp <- model.matrix(~.,train)
train_temp <- as.data.frame(train_temp)
lm <- lm(SalePrice ~ .,data=train_temp)
summary(lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = train_temp)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.68714	-0.04475	0.00000	0.04753	0.65091

```
##
## Coefficients: (23 not defined because of singularities)
##
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	1.583e+01	5.817e+00	2.721	0.006647 **
## '(Intercept)'	NA	NA	NA	NA
## MSSubClass30	-3.681e-02	2.750e-02	-1.338	0.181099
## MSSubClass40	-7.151e-02	8.837e-02	-0.809	0.418622
## MSSubClass45	1.600e-02	6.205e-02	0.258	0.796593
## MSSubClass50	4.512e-02	4.859e-02	0.929	0.353365
## MSSubClass60	-2.669e-02	4.659e-02	-0.573	0.566910
## MSSubClass70	3.704e-02	4.978e-02	0.744	0.457064
## MSSubClass75	-1.111e-01	1.160e-01	-0.957	0.338595
## MSSubClass80	-1.012e-01	8.806e-02	-1.150	0.250572
## MSSubClass85	-3.103e-03	5.841e-02	-0.053	0.957638
## MSSubClass90	-2.603e-02	4.096e-02	-0.635	0.525314
## MSSubClass120	-4.176e-02	7.109e-02	-0.587	0.557083
## MSSubClass160	-1.299e-01	9.054e-02	-1.434	0.151802
## MSSubClass180	-7.960e-02	1.051e-01	-0.757	0.449239
## MSSubClass190	-3.588e-02	1.456e-01	-0.247	0.805342
## MSZoningFV	4.891e-01	6.837e-02	7.154	1.83e-12 ***
## MSZoningRH	4.873e-01	6.715e-02	7.256	8.97e-13 ***
## MSZoningRL	4.885e-01	5.781e-02	8.449	< 2e-16 ***
## MSZoningRM	4.296e-01	5.368e-02	8.004	3.96e-15 ***
## LotFrontage	1.136e-05	1.332e-04	0.085	0.932080
## LotArea	3.122e-06	5.810e-07	5.373	1.00e-07 ***
## StreetPave	5.749e-02	7.401e-02	0.777	0.437496
## AlleyNone	-3.202e-02	2.437e-02	-1.314	0.189202
## AlleyPave	-2.031e-02	3.561e-02	-0.570	0.568588
## LotShapeIR2	2.318e-02	2.201e-02	1.053	0.292422
## LotShapeIR3	-1.231e-02	6.874e-02	-0.179	0.857934
## LotShapeReg	1.751e-02	9.121e-03	1.920	0.055179 .
## LandContourHLS	1.780e-02	2.954e-02	0.603	0.546921
## LandContourLow	-7.089e-03	3.742e-02	-0.189	0.849777
## LandContourLvl	2.998e-02	2.170e-02	1.382	0.167336
## UtilitiesNoSeWa	-2.149e-01	1.528e-01	-1.407	0.159886
## LotConfigCulDSac	2.000e-02	1.837e-02	1.089	0.276537
## LotConfigFR2	-3.483e-02	2.297e-02	-1.517	0.129720

## LotConfigFR3	-9.668e-02	9.266e-02	-1.043	0.297096	
## LotConfigInside	-1.631e-02	9.734e-03	-1.676	0.094168	.
## LandSlopeMod	4.717e-02	2.311e-02	2.041	0.041519	*
## LandSlopeSev	-2.089e-01	6.192e-02	-3.374	0.000775	***
## NeighborhoodBlueste	4.922e-02	9.946e-02	0.495	0.620831	
## NeighborhoodBrDale	2.162e-02	6.514e-02	0.332	0.740095	
## NeighborhoodBrkSide	-4.498e-03	5.204e-02	-0.086	0.931141	
## NeighborhoodClearCr	2.598e-02	4.942e-02	0.526	0.599234	
## NeighborhoodCollgCr	-2.108e-02	3.919e-02	-0.538	0.590843	
## NeighborhoodCrawfor	9.269e-02	4.848e-02	1.912	0.056251	.
## NeighborhoodEdwards	-9.757e-02	4.389e-02	-2.223	0.026478	*
## NeighborhoodGilbert	-4.627e-03	4.144e-02	-0.112	0.911107	
## NeighborhoodIDOTRR	-1.264e-02	5.992e-02	-0.211	0.832936	
## NeighborhoodMeadowV	-1.181e-01	6.592e-02	-1.791	0.073641	.
## NeighborhoodMitchel	-4.675e-02	4.464e-02	-1.047	0.295285	
## NeighborhoodNames	-3.754e-02	4.318e-02	-0.869	0.384906	
## NeighborhoodNoRidge	6.668e-02	4.646e-02	1.435	0.151571	
## NeighborhoodNPkVill	1.382e-03	8.833e-02	0.016	0.987519	
## NeighborhoodNridgHt	8.892e-02	3.989e-02	2.229	0.026078	*
## NeighborhoodNWAmes	-3.930e-02	4.386e-02	-0.896	0.370428	
## NeighborhoodOldTown	-4.386e-02	5.307e-02	-0.826	0.408778	
## NeighborhoodSawyer	-2.725e-02	4.437e-02	-0.614	0.539233	
## NeighborhoodSawyerW	-1.025e-02	4.223e-02	-0.243	0.808254	
## NeighborhoodSomerst	6.620e-02	5.158e-02	1.283	0.199728	
## NeighborhoodStoneBr	1.589e-01	4.373e-02	3.633	0.000297	***
## NeighborhoodSWISU	-1.818e-02	5.531e-02	-0.329	0.742509	
## NeighborhoodTimber	-3.547e-03	4.410e-02	-0.080	0.935909	
## NeighborhoodVeenker	4.485e-02	5.793e-02	0.774	0.439005	
## Condition1Feedr	3.083e-02	2.775e-02	1.111	0.266940	
## Condition1Norm	7.332e-02	2.331e-02	3.146	0.001713	**
## Condition1PosA	2.008e-02	5.245e-02	0.383	0.701906	
## Condition1PosN	8.720e-02	4.152e-02	2.100	0.035992	*
## Condition1RR Ae	-5.371e-02	5.173e-02	-1.038	0.299410	
## Condition1RR An	3.579e-02	3.840e-02	0.932	0.351484	
## Condition1RR Ne	3.488e-02	8.469e-02	0.412	0.680521	
## Condition1RR Nn	2.578e-02	7.492e-02	0.344	0.730830	
## Condition2Feedr	1.526e-01	1.495e-01	1.021	0.307642	
## Condition2Norm	8.155e-02	1.356e-01	0.601	0.547776	
## Condition2PosA	2.367e-01	2.216e-01	1.068	0.285897	
## Condition2PosN	-7.439e-01	1.661e-01	-4.479	8.52e-06	***
## Condition2RR Ae	NA	NA	NA	NA	
## Condition2RR An	-1.482e-02	1.787e-01	-0.083	0.933926	
## Condition2RR Nn	9.392e-02	1.609e-01	0.584	0.559491	
## BldgType2fmCon	7.192e-04	1.415e-01	0.005	0.995945	
## BldgTypeDuplex	NA	NA	NA	NA	
## BldgTypeTwnhs	-4.658e-02	7.643e-02	-0.609	0.542409	
## BldgTypeTwnhsE	-1.707e-02	7.197e-02	-0.237	0.812595	
## HouseStyle1.5Unf	NA	NA	NA	NA	
## HouseStyle1Story	1.519e-02	5.022e-02	0.302	0.762406	
## HouseStyle2.5Fin	3.348e-02	1.273e-01	0.263	0.792619	
## HouseStyle2.5Unf	1.654e-01	1.171e-01	1.413	0.158086	
## HouseStyle2Story	1.715e-02	4.675e-02	0.367	0.713779	
## HouseStyleSFoyer	1.325e-02	6.536e-02	0.203	0.839436	
## HouseStyleSLvl	9.777e-02	9.423e-02	1.038	0.299759	

## OverallQual	3.778e-02	5.954e-03	6.345	3.62e-10	***
## OverallCond	3.024e-02	5.113e-03	5.915	4.82e-09	***
## YearBuilt	1.819e-03	4.735e-04	3.842	0.000131	***
## YearRemodAdd	1.266e-03	3.211e-04	3.944	8.66e-05	***
## RoofStyleGable	2.774e-02	9.092e-02	0.305	0.760336	
## RoofStyleGambrel	-4.551e-02	1.050e-01	-0.433	0.664854	
## RoofStyleHip	2.798e-02	9.130e-02	0.306	0.759306	
## RoofStyleMansard	7.309e-02	1.101e-01	0.664	0.507122	
## RoofStyleShed	NA	NA	NA	NA	
## RoofMatlCompShg	-5.562e-02	5.624e-02	-0.989	0.322929	
## RoofMatlMembran	4.129e-01	1.716e-01	2.406	0.016322	*
## RoofMatlMetal	2.917e-01	1.665e-01	1.752	0.080175	.
## RoofMatlRoll	NA	NA	NA	NA	
## 'RoofMatlTar&Grv'	1.611e-03	1.045e-01	0.015	0.987706	
## RoofMatlWdShake	-1.284e-01	9.606e-02	-1.337	0.181654	
## RoofMatlWdShngl	NA	NA	NA	NA	
## Exterior1stAsphShn	-8.126e-02	1.680e-01	-0.484	0.628736	
## Exterior1stBrkComm	-2.344e-01	1.565e-01	-1.498	0.134404	
## Exterior1stBrkFace	8.411e-02	7.132e-02	1.179	0.238609	
## Exterior1stCBlock	NA	NA	NA	NA	
## Exterior1stCemntBd	-7.423e-02	1.128e-01	-0.658	0.510879	
## Exterior1stHdBoard	-1.450e-02	7.360e-02	-0.197	0.843851	
## Exterior1stImStucc	NA	NA	NA	NA	
## Exterior1stMetalSd	8.003e-02	8.718e-02	0.918	0.358869	
## Exterior1stPlywood	-1.284e-02	7.146e-02	-0.180	0.857431	
## Exterior1stStone	3.785e-02	1.626e-01	0.233	0.816008	
## Exterior1stStucco	3.668e-02	8.058e-02	0.455	0.649119	
## Exterior1stVinylSd	1.749e-02	7.986e-02	0.219	0.826722	
## 'Exterior1stWd Sdng'	-4.688e-02	6.952e-02	-0.674	0.500256	
## Exterior1stWdShing	8.427e-03	7.717e-02	0.109	0.913068	
## Exterior2ndAsphShn	6.593e-02	1.221e-01	0.540	0.589305	
## 'Exterior2ndBrk Cmn'	8.421e-02	1.226e-01	0.687	0.492303	
## Exterior2ndBrkFace	-3.679e-02	7.739e-02	-0.475	0.634581	
## Exterior2ndCBlock	NA	NA	NA	NA	
## Exterior2ndCmentBd	9.668e-02	1.130e-01	0.856	0.392511	
## Exterior2ndHdBoard	9.714e-03	7.440e-02	0.131	0.896151	
## Exterior2ndImStucc	3.335e-02	8.256e-02	0.404	0.686331	
## Exterior2ndMetalSd	-5.619e-02	8.775e-02	-0.640	0.522103	
## Exterior2ndOther	-1.097e-01	1.390e-01	-0.789	0.430389	
## Exterior2ndPlywood	1.938e-02	7.100e-02	0.273	0.785004	
## Exterior2ndStone	-1.313e-01	1.189e-01	-1.104	0.270030	
## Exterior2ndStucco	-1.602e-02	8.052e-02	-0.199	0.842317	
## Exterior2ndVinylSd	5.919e-03	7.995e-02	0.074	0.941004	
## 'Exterior2ndWd Sdng'	5.550e-02	7.022e-02	0.790	0.429540	
## 'Exterior2ndWd Shng'	3.088e-03	7.448e-02	0.041	0.966937	
## MasVnrTypeBrkFace	5.860e-02	3.476e-02	1.686	0.092219	.
## MasVnrTypeNone	4.896e-02	3.500e-02	1.399	0.162196	
## MasVnrTypeStone	6.552e-02	3.727e-02	1.758	0.079140	.
## MasVnrArea	2.191e-05	3.178e-05	0.689	0.490716	
## ExterQualFa	6.251e-02	7.688e-02	0.813	0.416337	
## ExterQualGd	1.308e-02	2.619e-02	0.499	0.617764	
## ExterQualTA	2.110e-02	2.902e-02	0.727	0.467415	
## ExterCondFa	-1.378e-01	1.276e-01	-1.080	0.280246	
## ExterCondGd	-7.768e-02	1.254e-01	-0.620	0.535617	

## ExterCondPo	-5.398e-02	1.848e-01	-0.292	0.770340	
## ExterCondTA	-6.449e-02	1.253e-01	-0.515	0.606894	
## FoundationCBlock	-1.401e-02	1.834e-02	-0.764	0.445210	
## FoundationPConc	5.171e-03	1.982e-02	0.261	0.794249	
## FoundationSlab	-7.164e-02	5.238e-02	-1.368	0.171792	
## FoundationStone	1.692e-01	7.269e-02	2.328	0.020129	*
## FoundationWood	-1.444e-01	1.237e-01	-1.168	0.243191	
## BsmtQualFa	-7.604e-02	3.651e-02	-2.083	0.037594	*
## BsmtQualGd	-4.150e-02	1.855e-02	-2.238	0.025500	*
## BsmtQualNone	3.464e-02	1.810e-01	0.191	0.848294	
## BsmtQualTA	-4.315e-02	2.292e-02	-1.883	0.060068	.
## BsmtCondGd	-4.287e-03	2.961e-02	-0.145	0.884906	
## BsmtCondNone	NA	NA	NA	NA	
## BsmtCondPo	2.470e-01	1.564e-01	1.579	0.114671	
## BsmtCondTA	6.264e-03	2.421e-02	0.259	0.795931	
## BsmtExposureGd	2.160e-02	1.742e-02	1.240	0.215328	
## BsmtExposureMn	3.162e-03	1.676e-02	0.189	0.850393	
## BsmtExposureNo	-9.602e-03	1.230e-02	-0.781	0.435123	
## BsmtExposureNone	-6.124e-02	1.089e-01	-0.562	0.574029	
## BsmtFinType1BLQ	-3.655e-03	1.600e-02	-0.229	0.819295	
## BsmtFinType1GLQ	4.503e-03	1.398e-02	0.322	0.747462	
## BsmtFinType1LwQ	-3.419e-02	2.126e-02	-1.608	0.108130	
## BsmtFinType1None	NA	NA	NA	NA	
## BsmtFinType1Rec	-6.870e-04	1.726e-02	-0.040	0.968251	
## BsmtFinType1Unf	-1.962e-02	1.644e-02	-1.193	0.233150	
## BsmtFinSF1	9.949e-05	3.099e-05	3.210	0.001375	**
## BsmtFinType2BLQ	-6.793e-02	3.918e-02	-1.734	0.083337	.
## BsmtFinType2GLQ	-8.456e-03	4.941e-02	-0.171	0.864163	
## BsmtFinType2LwQ	-4.727e-02	3.959e-02	-1.194	0.232855	
## BsmtFinType2None	-8.326e-02	1.216e-01	-0.685	0.493782	
## BsmtFinType2Rec	-3.696e-02	3.818e-02	-0.968	0.333268	
## BsmtFinType2Unf	-2.305e-02	3.965e-02	-0.581	0.561122	
## BsmtFinSF2	1.041e-04	5.016e-05	2.075	0.038319	*
## BsmtUnfSF	5.060e-05	2.879e-05	1.758	0.079187	.
## TotalBsmtSF	NA	NA	NA	NA	
## HeatingGasA	-9.296e-02	7.906e-02	-1.176	0.240010	
## HeatingGasW	-3.577e-02	8.542e-02	-0.419	0.675552	
## HeatingGrav	-3.472e-01	1.034e-01	-3.357	0.000823	***
## HeatingOthW	NA	NA	NA	NA	
## HeatingWall	NA	NA	NA	NA	
## HeatingQCFA	-3.115e-02	2.705e-02	-1.152	0.249726	
## HeatingQCGd	-2.107e-02	1.142e-02	-1.845	0.065457	.
## HeatingQCPo	-7.009e-02	1.310e-01	-0.535	0.592828	
## HeatingQCTA	-3.226e-02	1.167e-02	-2.765	0.005813	**
## CentralAirY	5.818e-02	2.238e-02	2.599	0.009507	**
## ElectricalFuseF	5.791e-03	3.267e-02	0.177	0.859333	
## ElectricalFuseP	-1.035e-01	9.760e-02	-1.060	0.289423	
## ElectricalMix	-1.667e-01	2.260e-01	-0.737	0.461081	
## ElectricalNone	NA	NA	NA	NA	
## ElectricalSBrkr	-1.248e-02	1.707e-02	-0.731	0.464799	
## X1stFlrSF	2.558e-04	3.247e-05	7.877	1.02e-14	***
## X2ndFlrSF	2.083e-04	3.305e-05	6.303	4.69e-10	***
## LowQualFinSF	2.457e-04	1.067e-04	2.304	0.021477	*
## GrLivArea	NA	NA	NA	NA	

## BsmtFullBath	2.700e-02	1.092e-02	2.473	0.013593	*
## BsmtHalfBath	7.117e-03	1.646e-02	0.432	0.665634	
## FullBath	3.889e-02	1.234e-02	3.152	0.001680	**
## HalfBath	3.369e-02	1.190e-02	2.832	0.004738	**
## BedroomAbvGr	-2.181e-03	7.958e-03	-0.274	0.784098	
## KitchenAbvGr	-4.120e-02	3.393e-02	-1.214	0.225031	
## KitchenQualFa	-5.499e-02	3.608e-02	-1.524	0.127867	
## KitchenQualGd	-6.080e-02	1.931e-02	-3.149	0.001697	**
## KitchenQualTA	-5.880e-02	2.203e-02	-2.669	0.007762	**
## TotRmsAbvGrd	6.741e-03	5.428e-03	1.242	0.214616	
## FunctionalMaj2	-2.968e-01	7.636e-02	-3.887	0.000110	***
## FunctionalMin1	-6.191e-03	5.201e-02	-0.119	0.905287	
## FunctionalMin2	-8.958e-03	5.400e-02	-0.166	0.868285	
## FunctionalMod	-1.024e-01	6.375e-02	-1.606	0.108627	
## FunctionalSev	-3.049e-01	1.520e-01	-2.005	0.045257	*
## FunctionalTyp	3.012e-02	4.676e-02	0.644	0.519677	
## Fireplaces	1.528e-02	1.417e-02	1.078	0.281255	
## FireplaceQuFa	-7.972e-03	3.868e-02	-0.206	0.836752	
## FireplaceQuGd	-6.308e-03	3.041e-02	-0.207	0.835706	
## FireplaceQuNone	-1.902e-02	3.563e-02	-0.534	0.593609	
## FireplaceQuPo	2.790e-03	4.452e-02	0.063	0.950049	
## FireplaceQuTA	-5.627e-04	3.190e-02	-0.018	0.985928	
## GarageTypeAttchd	8.825e-02	6.608e-02	1.335	0.182075	
## GarageTypeBasment	5.957e-02	7.828e-02	0.761	0.446860	
## GarageTypeBuiltIn	8.752e-02	6.889e-02	1.270	0.204306	
## GarageTypeCarPort	1.220e-01	9.109e-02	1.339	0.180922	
## GarageTypeDetchd	9.099e-02	6.575e-02	1.384	0.166755	
## GarageTypeNone	-5.872e-01	7.163e-01	-0.820	0.412635	
## GarageYrBlt	-3.089e-04	3.599e-04	-0.858	0.390954	
## GarageFinishNone	NA	NA	NA	NA	
## GarageFinishRFRn	1.190e-03	1.072e-02	0.111	0.911642	
## GarageFinishUnf	-1.501e-02	1.308e-02	-1.148	0.251165	
## GarageCars	5.902e-03	1.304e-02	0.453	0.650936	
## GarageArea	1.420e-04	4.420e-05	3.212	0.001367	**
## GarageQualFa	-3.559e-01	1.519e-01	-2.343	0.019359	*
## GarageQualGd	-2.608e-01	1.601e-01	-1.629	0.103770	
## GarageQualNone	NA	NA	NA	NA	
## GarageQualPo	-2.635e-01	2.062e-01	-1.278	0.201604	
## GarageQualTA	-3.041e-01	1.500e-01	-2.027	0.042983	*
## GarageCondFa	2.661e-01	1.743e-01	1.527	0.127234	
## GarageCondGd	2.776e-01	1.837e-01	1.511	0.131096	
## GarageCondNone	NA	NA	NA	NA	
## GarageCondPo	3.132e-01	1.920e-01	1.631	0.103161	
## GarageCondTA	2.570e-01	1.725e-01	1.490	0.136503	
## PavedDriveP	-2.297e-02	3.266e-02	-0.703	0.482011	
## PavedDriveY	3.234e-02	2.182e-02	1.482	0.138690	
## WoodDeckSF	8.754e-05	3.315e-05	2.641	0.008414	**
## OpenPorchSF	-5.624e-06	6.648e-05	-0.085	0.932604	
## EnclosedPorch	1.135e-04	7.387e-05	1.537	0.124773	
## X3SsnPorch	1.007e-04	1.240e-04	0.812	0.416911	
## ScreenPorch	2.721e-04	6.922e-05	3.932	9.13e-05	***
## PoolArea	6.908e-05	3.688e-04	0.187	0.851468	
## PoolQCFa	9.223e-03	2.625e-01	0.035	0.971978	
## PoolQCGd	NA	NA	NA	NA	



```
## PoolQCNONE          NA          NA          NA          NA
## FenceGdWo          -3.653e-02  2.865e-02 -1.275 0.202597
## FenceMnPrv          3.779e-03  2.226e-02  0.170 0.865214
## FenceMnWw          -2.650e-02  4.726e-02 -0.561 0.575083
## FenceNone           8.309e-03  2.050e-02  0.405 0.685323
## MiscFeatureNone     -9.210e-02  5.575e-01 -0.165 0.868825
## MiscFeatureOthr     -7.151e-02  5.105e-01 -0.140 0.888629
## MiscFeatureShed     -9.348e-02  5.344e-01 -0.175 0.861180
## MiscFeatureTenC          NA          NA          NA          NA
## MiscVal            -3.508e-06  3.543e-05 -0.099 0.921153
## MoSold             -2.099e-03  1.387e-03 -1.514 0.130483
## YrSold             -5.412e-03  2.865e-03 -1.889 0.059254 .
## SaleTypeCon          1.085e-01  8.542e-02  1.270 0.204295
## SaleTypeConLD          1.447e-01  5.280e-02  2.740 0.006266 **
## SaleTypeConLI        -2.331e-02  5.688e-02 -0.410 0.681992
## SaleTypeConLw          1.073e-02  6.804e-02  0.158 0.874676
## SaleTypeCWD           1.145e-01  7.012e-02  1.632 0.102991
## SaleTypeNew           8.336e-02  9.194e-02  0.907 0.364835
## SaleTypeOth           2.176e-01  1.191e-01  1.827 0.068046 .
## SaleTypeWD           -4.737e-03  2.258e-02 -0.210 0.833895
## SaleConditionAdjLand  1.383e-01  7.335e-02  1.886 0.059660 .
## SaleConditionAlloca   7.282e-02  4.805e-02  1.516 0.130002
## SaleConditionFamily   4.893e-04  3.467e-02  0.014 0.988743
## SaleConditionNormal   7.076e-02  1.599e-02  4.427 1.08e-05 ***
## SaleConditionPartial  2.240e-02  8.824e-02  0.254 0.799683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1062 on 848 degrees of freedom
## Multiple R-squared:  0.9466, Adjusted R-squared:  0.9307
## F-statistic: 59.85 on 251 and 848 DF, p-value: < 2.2e-16
```

```
test_temp <- model.matrix(~ .,test)
test_temp <- as.data.frame(test_temp)
test_temp <- test_temp[-199,]
library(caret)
```

```
## Loading required package: lattice
```

```
##
```

```
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
## lift
```

```
predictions = predict(lm, newdata = test_temp)
```

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

```
diff3 <- data.frame(predictions = predictions)
tempSalePrice <- test$SalePrice
tempSalePrice <- tempSalePrice[-199]
#diff3$predictions <- exp(diff3$predictions)
diff3$residuals <- predictions - tempSalePrice

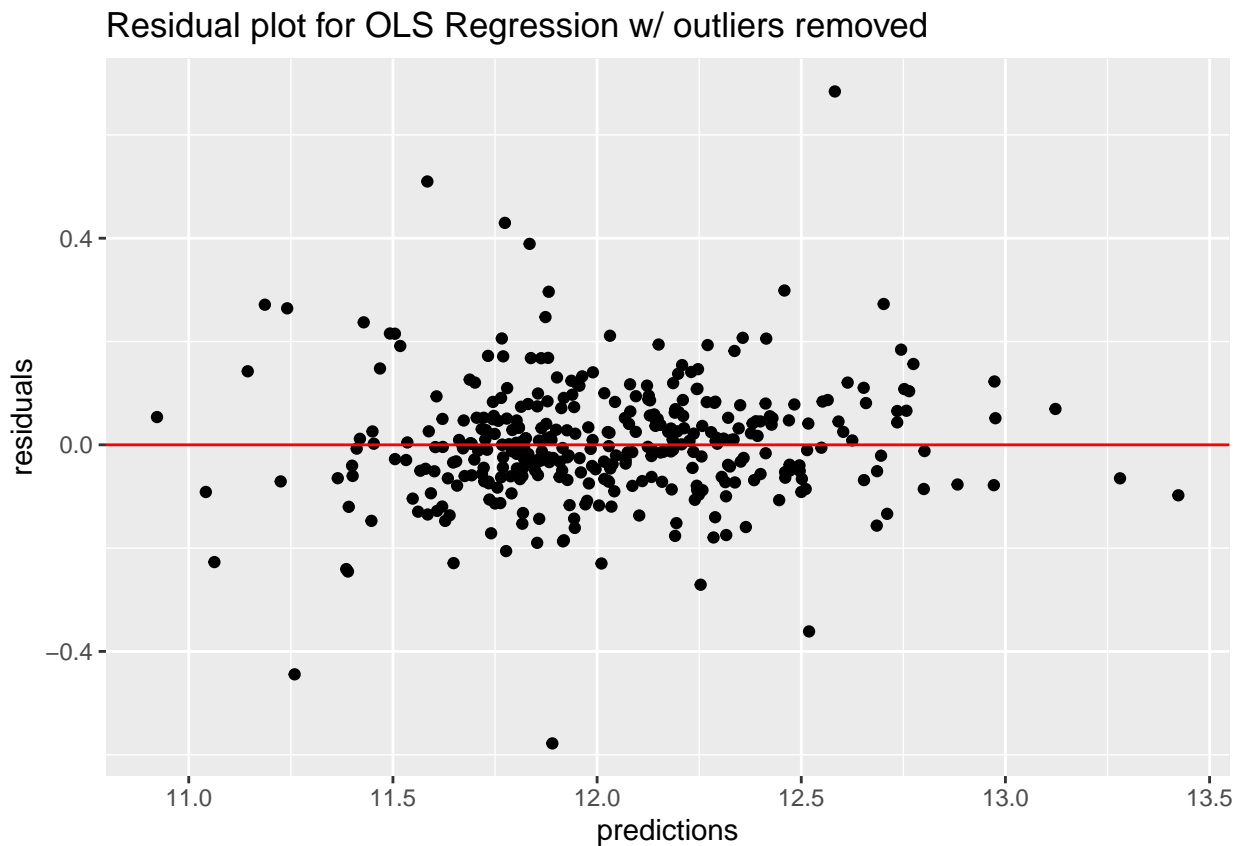
diff3$residab <- abs(diff3$residuals)
diff3$sq <- (diff3$residuals * diff3$residuals)
#diff3 <- diff3 %>% mutate(sq = residuals^2)

mean(diff3$residab)
```

```
## [1] 0.0833482
```

```
#diff3 <- diff3[-199,]
```

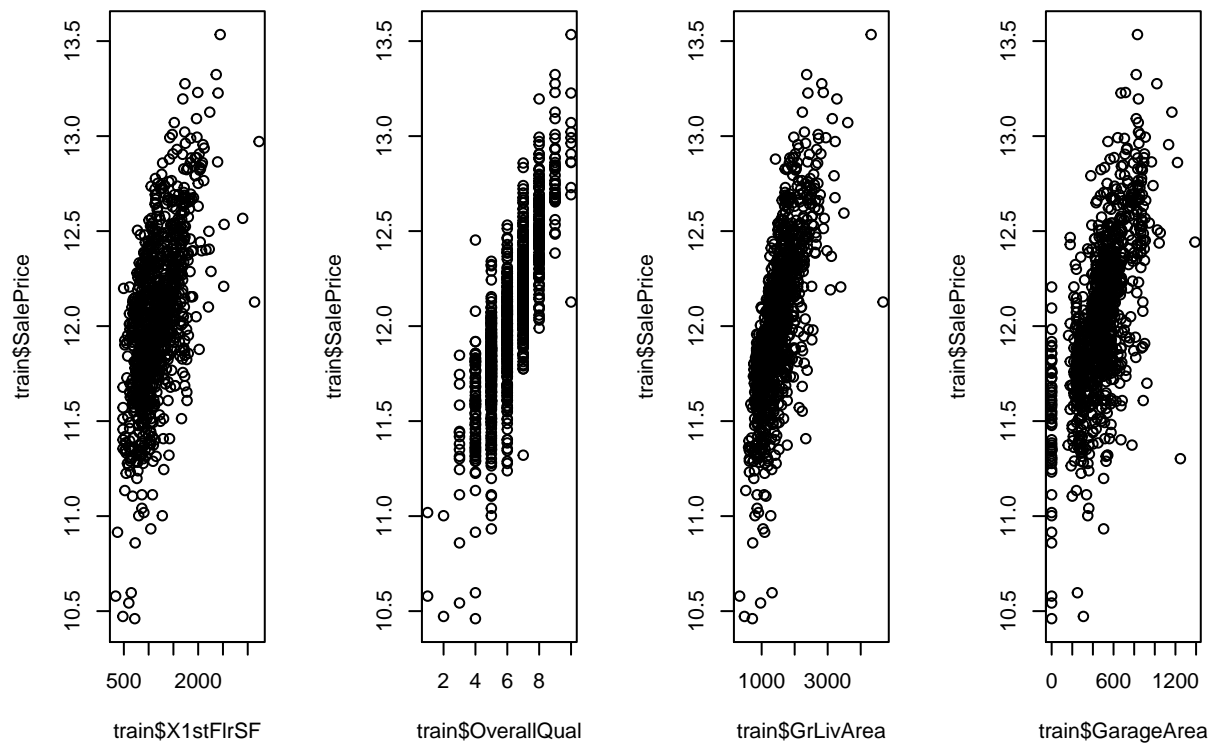
```
#mean(diff3$residab)
ggplot(data = diff3) + geom_point(mapping = aes(x=predictions, y=residuals)) +
  geom_abline(intercept = 0, slope = 0, colour = "red") +
  ggtitle("Residual plot for OLS Regression w/ outliers removed")
```



```

par(mfrow=c(1,4))
plot(train$X1stFlrSF,train$SalePrice)
plot(train$OverallQual,train$SalePrice)
plot(train$GrLivArea,train$SalePrice)
plot(train$GarageArea,train$SalePrice)

```



```

train2 <- data[0:1458,]
train_temp2 <- model.matrix(~.,train2)
train_temp2 <- as.data.frame(train_temp2)
lm2 <- lm(SalePrice ~ .,data=train_temp2)
predictions2 <- predict(lm2)
residuals2 <- predictions2 - train2$SalePrice
plot(predictions2,residuals2)

```

