Data 605 - Computational Mathematics Final

Joseph Simone

12/5/2019

```
library(kableExtra)
library(pastecs)
library(psych)
library(e1071)
library(fBasics)
library(dplyr)
library(ggplot2)
library(pracma)
library(MASS)
library(survival)
library(tidyverse)
```

Problem #1

Using R, generate a random variable X that has 10,000 random uniform numbers from 1 to N, where N can be any number of your choosing greater than or equal to 6. Then generate a random variable Y that has 10,000 random normal numbers with a mean of $\mu = \sigma = (N+1)/2$.

```
set.seed(123)


N <- round(runif(1,6,100))

n <- 10000

X<-runif(n,min=0,max=N)



Y<-rnorm(n,(N+1)/2,(N+1)/2)

x <- median(X)
x
```

```
## [1] 16
```

```
y <- quantile(Y, 0.25)
y
```

```
## 25%
## 5.2
```

```
X_Y <- as.data.frame(cbind(X,Y))
```

Probability Calculate as a minimum the below probabilities a through c.

Assume the small letter "x" is estimated as the median of the X variable, and the small letter "y" is estimated as the 1st quartile of the Y variable.

Interpret the meaning of all probabilities.

A.

$P(X > x|X > y)$

We will use the dataframe we created above from the 2 vectors.

Using the formula: $P(X > x|X > y) = P(X > xandX > y)/P(X > y)$

Probability that X is greater than its median given that X is greater than the first quartile of Y

```
a_1 <- (length(X[X>x & X>y])/length(X)) /(length(X[X>y])/length(X))
print(a_1)
```

```
## [1] 0.59
```

Probability that X is grater than all possible x and Y is greater than all possible y

B.

$P(X > x, Y > y)$

```
X_gr_x <- length(X[X>x]) / length(X)

Y_gr_y <- length(Y[Y>y]) / length(Y)

b_1 <- X_gr_x * Y_gr_y

print(b_1)
```

```
## [1] 0.38
```

Probability of X greater than its median and greater than the first quantile of Y

C.

$P(X < x|X > y)$

```
X_ls_x_and_X_gr_y = length(X[X<x & X>y])/length(X)
X_gr_y <- length(X[X>y])/length(X)

c_1 <- X_ls_x_and_X_gr_y / X_gr_y
print(c_1)
```

```
## [1] 0.41
```

D.

Investigate whether P(X>x and Y>y)=P(X>x)P(Y>y) by building a table and evaluating the marginal and joint probabilities.

```
count_Xgrx_Ygry <- length(X[X>x & Y>y])
count_Xgrx_Ylsy <- length(X[X>x & Y<y])

count_Xlsx_Ygry <- length(X[X<x & Y>y])
count_Xlsx_Ylsy <- length(X[X<x & Y<y])

contingency_table <- matrix(c(count_Xgrx_Ygry, count_Xgrx_Ylsy, count_Xlsx_Ygry, count_Xlsx_Ylsy), nrow

rownames(contingency_table) <- c('(Y>y)','(Y<y)')

kable(contingency_table, digits=4, col.names = c('(X>x)', '(X<x)'), align = 'l')
```

	(X>x)	(X<x)
(Y>y)	3699	3801
(Y<y)	1301	1199

Now we can calculate the left handside of the equation:

$P(X > xandY > y)$ using the data from the contingency table

```
X_gr_x_and_Y_gr_y <- length(X[X>x & Y>y]) / 10000
print(X_gr_x_and_Y_gr_y)
```

```
## [1] 0.37
```

```
X_gt_x_given_Y_gt_y <- X_gr_x * Y_gr_y
print(X_gt_x_given_Y_gt_y)
```

```
## [1] 0.38
```

Both of these values are fairly close to zero.

E.

Check to see if independence holds by using Fisher's Exact Test and the Chi Square Test.

What is the difference between the two? Which is most appropriate?

I will be using the conitingency table from above.

Null Hypothesisbr<>

$H0 : X > x$ & $Y > y$ are independent events

Alternate Hypothesis

HA : Both of these are dependent events

Fisher's Tect

```r
fisher.test(contingency_table)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  contingency_table
## p-value = 0.02
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.82 0.98
## sample estimates:
## odds ratio
##        0.9
```

The p-value is high compared to the significance level of 0.05, thereforewe cannot reject the Null Hypothesis, that both are independent, in favor of the Alternate Hypothesis.

Chi Square Test

```r
chisq.test(contingency_table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  contingency_table
## X-squared = 5, df = 1, p-value = 0.02
```

The p-value is high as compared to the significance level of 0.05, therefore we can reject the Null Hypothesis, in favor that both of these are Dependent Events.

Since the results are so similar we would conclude that X and Y are indeed independent variables.

Problem 2

You are to register for Kaggle.com (free) and compete in the House Prices: Advanced Regression Techniques competition. https://www.kaggle.com/c/house-prices-advanced-regression-techniques .

Descriptive and Inferential Statistics

Provide univariate descriptive statistics and appropriate plots for the training data set. Provide a scatterplot matrix for at least two of the independent variables and the dependent variable. Derive a correlation matrix for any three quantitative variables in the dataset. Test the hypotheses that the correlations between each pairwise set of variables is 0 and provide an 80% confidence interval. Discuss the meaning of your analysis. Would you be worried about familywise error? Why or why not?

```r
train_raw <- read_csv(file ="https://raw.githubusercontent.com/josephsimone/Data-605/master/train.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
```

```
##   Id = col_integer(),
##   MSSubClass = col_integer(),
##   LotFrontage = col_integer(),
##   LotArea = col_integer(),
##   OverallQual = col_integer(),
##   OverallCond = col_integer(),
##   YearBuilt = col_integer(),
##   YearRemodAdd = col_integer(),
##   MasVnrArea = col_integer(),
##   BsmtFinSF1 = col_integer(),
##   BsmtFinSF2 = col_integer(),
##   BsmtUnfSF = col_integer(),
##   TotalBsmtSF = col_integer(),
##   `1stFlrSF` = col_integer(),
##   `2ndFlrSF` = col_integer(),
##   LowQualFinSF = col_integer(),
##   GrLivArea = col_integer(),
##   BsmtFullBath = col_integer(),
##   BsmtHalfBath = col_integer(),
##   FullBath = col_integer()
##   # ... with 18 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```r
train_df <- as.data.frame(train_raw)
head(train_df, 5)
```

```
##   Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape
## 1  1         60       RL          65    8450   Pave  <NA>      Reg
## 2  2         20       RL          80    9600   Pave  <NA>      Reg
## 3  3         60       RL          68   11250   Pave  <NA>      IR1
## 4  4         70       RL          60    9550   Pave  <NA>      IR1
## 5  5         60       RL          84   14260   Pave  <NA>      IR1
##   LandContour Utilities LotConfig LandSlope Neighborhood Condition1
## 1         Lvl    AllPub    Inside       Gtl      CollgCr       Norm
## 2         Lvl    AllPub       FR2       Gtl      Veenker      Feedr
## 3         Lvl    AllPub    Inside       Gtl      CollgCr       Norm
## 4         Lvl    AllPub    Corner       Gtl      Crawfor       Norm
## 5         Lvl    AllPub       FR2       Gtl      NoRidge       Norm
##   Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt
## 1       Norm     1Fam     2Story           7           5      2003
## 2       Norm     1Fam     1Story           6           8      1976
## 3       Norm     1Fam     2Story           7           5      2001
## 4       Norm     1Fam     2Story           7           5      1915
## 5       Norm     1Fam     2Story           8           5      2000
##   YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType
## 1         2003     Gable  CompShg     VinylSd     VinylSd    BrkFace
## 2         1976     Gable  CompShg     MetalSd     MetalSd       None
## 3         2002     Gable  CompShg     VinylSd     VinylSd    BrkFace
## 4         1970     Gable  CompShg     Wd Sdng     Wd Shng       None
## 5         2000     Gable  CompShg     VinylSd     VinylSd    BrkFace
##   MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure
```

```
## 1        196        Gd         TA      PConc       Gd       TA           No
## 2          0        TA         TA     CBlock       Gd       TA           Gd
## 3        162        Gd         TA      PConc       Gd       TA           Mn
## 4          0        TA         TA     BrkTil       TA       Gd           No
## 5        350        Gd         TA      PConc       Gd       TA           Av
##   BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF
## 1          GLQ        706          Unf          0       150         856
## 2          ALQ        978          Unf          0       284        1262
## 3          GLQ        486          Unf          0       434         920
## 4          ALQ        216          Unf          0       540         756
## 5          GLQ        655          Unf          0       490        1145
##   Heating HeatingQC CentralAir Electrical 1stFlrSF 2ndFlrSF LowQualFinSF
## 1    GasA        Ex          Y      SBrkr      856      854            0
## 2    GasA        Ex          Y      SBrkr     1262        0            0
## 3    GasA        Ex          Y      SBrkr      920      866            0
## 4    GasA        Gd          Y      SBrkr      961      756            0
## 5    GasA        Ex          Y      SBrkr     1145     1053            0
##   GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr
## 1      1710            1            0        2        1            3
## 2      1262            0            1        2        0            3
## 3      1786            1            0        2        1            3
## 4      1717            1            0        1        0            3
## 5      2198            1            0        2        1            4
##   KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu
## 1            1          Gd            8        Typ          0        <NA>
## 2            1          TA            6        Typ          1          TA
## 3            1          Gd            6        Typ          1          TA
## 4            1          Gd            7        Typ          1          Gd
## 5            1          Gd            9        Typ          1          TA
##   GarageType GarageYrBlt GarageFinish GarageCars GarageArea GarageQual
## 1     Attchd        2003          RFn          2        548         TA
## 2     Attchd        1976          RFn          2        460         TA
## 3     Attchd        2001          RFn          2        608         TA
## 4     Detchd        1998          Unf          3        642         TA
## 5     Attchd        2000          RFn          3        836         TA
##   GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch 3SsnPorch
## 1         TA          Y          0          61             0         0
## 2         TA          Y        298           0             0         0
## 3         TA          Y          0          42             0         0
## 4         TA          Y          0          35           272         0
## 5         TA          Y        192          84             0         0
##   ScreenPorch PoolArea PoolQC Fence MiscFeature MiscVal MoSold YrSold
## 1           0        0   <NA>  <NA>        <NA>       0      2   2008
## 2           0        0   <NA>  <NA>        <NA>       0      5   2007
## 3           0        0   <NA>  <NA>        <NA>       0      9   2008
## 4           0        0   <NA>  <NA>        <NA>       0      2   2006
## 5           0        0   <NA>  <NA>        <NA>       0     12   2008
##   SaleType SaleCondition SalePrice
## 1       WD        Normal    208500
## 2       WD        Normal    181500
## 3       WD        Normal    223500
## 4       WD       Abnorml    140000
## 5       WD        Normal    250000
```

```r
summary(train_df)
```

```
##        Id          MSSubClass      MSZoning          LotFrontage   
##  Min.   :   1   Min.   : 20   Length:1460        Min.   : 21  
##  1st Qu.: 366   1st Qu.: 20   Class :character   1st Qu.: 59  
##  Median : 730   Median : 50   Mode  :character   Median : 69  
##  Mean   : 730   Mean   : 57                      Mean   : 70  
##  3rd Qu.:1095   3rd Qu.: 70                      3rd Qu.: 80  
##  Max.   :1460   Max.   :190                      Max.   :313  
##                                                  NA's   :259  
##     LotArea          Street             Alley             LotShape        
##  Min.   :  1300   Length:1460        Length:1460        Length:1460       
##  1st Qu.:  7554   Class :character   Class :character   Class :character  
##  Median :  9478   Mode  :character   Mode  :character   Mode  :character  
##  Mean   : 10517                                                           
##  3rd Qu.: 11602                                                           
##  Max.   :215245                                                           
##                                                                           
##  LandContour         Utilities          LotConfig        
##  Length:1460        Length:1460        Length:1460       
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##   LandSlope         Neighborhood        Condition1       
##  Length:1460        Length:1460        Length:1460       
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##   Condition2          BldgType          HouseStyle         OverallQual   
##  Length:1460        Length:1460        Length:1460        Min.   : 1.0  
##  Class :character   Class :character   Class :character   1st Qu.: 5.0  
##  Mode  :character   Mode  :character   Mode  :character   Median : 6.0  
##                                                           Mean   : 6.1  
##                                                           3rd Qu.: 7.0  
##                                                           Max.   :10.0  
##                                                                         
##   OverallCond    YearBuilt     YearRemodAdd   RoofStyle        
##  Min.   :1.0   Min.   :1872   Min.   :1950   Length:1460       
##  1st Qu.:5.0   1st Qu.:1954   1st Qu.:1967   Class :character  
##  Median :5.0   Median :1973   Median :1994   Mode  :character  
##  Mean   :5.6   Mean   :1971   Mean   :1985                     
##  3rd Qu.:6.0   3rd Qu.:2000   3rd Qu.:2004                     
##  Max.   :9.0   Max.   :2010   Max.   :2010                     
##                                                                
##    RoofMatl          Exterior1st        Exterior2nd       
##  Length:1460        Length:1460        Length:1460       
##  Class :character   Class :character   Class :character  
```

```
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##   MasVnrType          MasVnrArea    ExterQual          ExterCond        
##  Length:1460        Min.   :   0   Length:1460        Length:1460       
##  Class :character   1st Qu.:   0   Class :character   Class :character  
##  Mode  :character   Median :   0   Mode  :character   Mode  :character  
##                     Mean   : 104                                        
##                     3rd Qu.: 166                                        
##                     Max.   :1600                                        
##                     NA's   :8                                           
##   Foundation          BsmtQual           BsmtCond        
##  Length:1460        Length:1460        Length:1460       
##  Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character  
##                                                          
##                                                          
##                                                          
##                                                          
##  BsmtExposure       BsmtFinType1         BsmtFinSF1   BsmtFinType2      
##  Length:1460        Length:1460        Min.   :   0   Length:1460       
##  Class :character   Class :character   1st Qu.:   0   Class :character  
##  Mode  :character   Mode  :character   Median : 384   Mode  :character  
##                                        Mean   : 444                     
##                                        3rd Qu.: 712                     
##                                        Max.   :5644                     
##                                                                         
##    BsmtFinSF2     BsmtUnfSF     TotalBsmtSF     Heating         
##  Min.   :   0   Min.   :   0   Min.   :   0   Length:1460       
##  1st Qu.:   0   1st Qu.: 223   1st Qu.: 796   Class :character  
##  Median :   0   Median : 478   Median : 992   Mode  :character  
##  Mean   :  47   Mean   : 567   Mean   :1057                     
##  3rd Qu.:   0   3rd Qu.: 808   3rd Qu.:1298                     
##  Max.   :1474   Max.   :2336   Max.   :6110                     
##                                                                 
##   HeatingQC          CentralAir         Electrical          1stFlrSF   
##  Length:1460        Length:1460        Length:1460        Min.   : 334  
##  Class :character   Class :character   Class :character   1st Qu.: 882  
##  Mode  :character   Mode  :character   Mode  :character   Median :1087  
##                                                           Mean   :1163  
##                                                           3rd Qu.:1391  
##                                                           Max.   :4692  
##                                                                         
##     2ndFlrSF     LowQualFinSF   GrLivArea     BsmtFullBath   BsmtHalfBath
##  Min.   :   0   Min.   :  0   Min.   : 334   Min.   :0.00   Min.   :0.00
##  1st Qu.:   0   1st Qu.:  0   1st Qu.:1130   1st Qu.:0.00   1st Qu.:0.00
##  Median :   0   Median :  0   Median :1464   Median :0.00   Median :0.00
##  Mean   : 347   Mean   :  6   Mean   :1515   Mean   :0.43   Mean   :0.06
##  3rd Qu.: 728   3rd Qu.:  0   3rd Qu.:1777   3rd Qu.:1.00   3rd Qu.:0.00
##  Max.   :2065   Max.   :572   Max.   :5642   Max.   :3.00   Max.   :2.00
##                                                                         
##     FullBath       HalfBath     BedroomAbvGr  KitchenAbvGr
```

```
##  Min.   :0.00   Min.   :0.00   Min.   :0.0   Min.   :0.00
##  1st Qu.:1.00   1st Qu.:0.00   1st Qu.:2.0   1st Qu.:1.00
##  Median :2.00   Median :0.00   Median :3.0   Median :1.00
##  Mean   :1.57   Mean   :0.38   Mean   :2.9   Mean   :1.05
##  3rd Qu.:2.00   3rd Qu.:1.00   3rd Qu.:3.0   3rd Qu.:1.00
##  Max.   :3.00   Max.   :2.00   Max.   :8.0   Max.   :3.00
##
##  KitchenQual         TotRmsAbvGrd   Functional          Fireplaces
##  Length:1460        Min.   : 2.0   Length:1460        Min.   :0.00
##  Class :character   1st Qu.: 5.0   Class :character   1st Qu.:0.00
##  Mode  :character   Median : 6.0   Mode  :character   Median :1.00
##                     Mean   : 6.5                      Mean   :0.61
##                     3rd Qu.: 7.0                      3rd Qu.:1.00
##                     Max.   :14.0                      Max.   :3.00
##
##  FireplaceQu         GarageType         GarageYrBlt   GarageFinish
##  Length:1460        Length:1460        Min.   :1900   Length:1460
##  Class :character   Class :character   1st Qu.:1961   Class :character
##  Mode  :character   Mode  :character   Median :1980   Mode  :character
##                                        Mean   :1979
##                                        3rd Qu.:2002
##                                        Max.   :2010
##                                        NA's   :81
##    GarageCars    GarageArea    GarageQual         GarageCond
##  Min.   :0.0   Min.   :   0   Length:1460        Length:1460
##  1st Qu.:1.0   1st Qu.: 334   Class :character   Class :character
##  Median :2.0   Median : 480   Mode  :character   Mode  :character
##  Mean   :1.8   Mean   : 473
##  3rd Qu.:2.0   3rd Qu.: 576
##  Max.   :4.0   Max.   :1418
##
##   PavedDrive          WoodDeckSF   OpenPorchSF  EnclosedPorch
##  Length:1460        Min.   :  0   Min.   :  0   Min.   :  0
##  Class :character   1st Qu.:  0   1st Qu.:  0   1st Qu.:  0
##  Mode  :character   Median :  0   Median : 25   Median :  0
##                     Mean   : 94   Mean   : 47   Mean   : 22
##                     3rd Qu.:168   3rd Qu.: 68   3rd Qu.:  0
##                     Max.   :857   Max.   :547   Max.   :552
##
##    3SsnPorch    ScreenPorch     PoolArea      PoolQC
##  Min.   :  0   Min.   :  0   Min.   :  0   Length:1460
##  1st Qu.:  0   1st Qu.:  0   1st Qu.:  0   Class :character
##  Median :  0   Median :  0   Median :  0   Mode  :character
##  Mean   :  3   Mean   : 15   Mean   :  3
##  3rd Qu.:  0   3rd Qu.:  0   3rd Qu.:  0
##  Max.   :508   Max.   :480   Max.   :738
##
##     Fence           MiscFeature          MiscVal          MoSold
##  Length:1460        Length:1460        Min.   :    0   Min.   : 1.0
##  Class :character   Class :character   1st Qu.:    0   1st Qu.: 5.0
##  Mode  :character   Mode  :character   Median :    0   Median : 6.0
##                                        Mean   :   43   Mean   : 6.3
##                                        3rd Qu.:    0   3rd Qu.: 8.0
##                                        Max.   :15500   Max.   :12.0
```

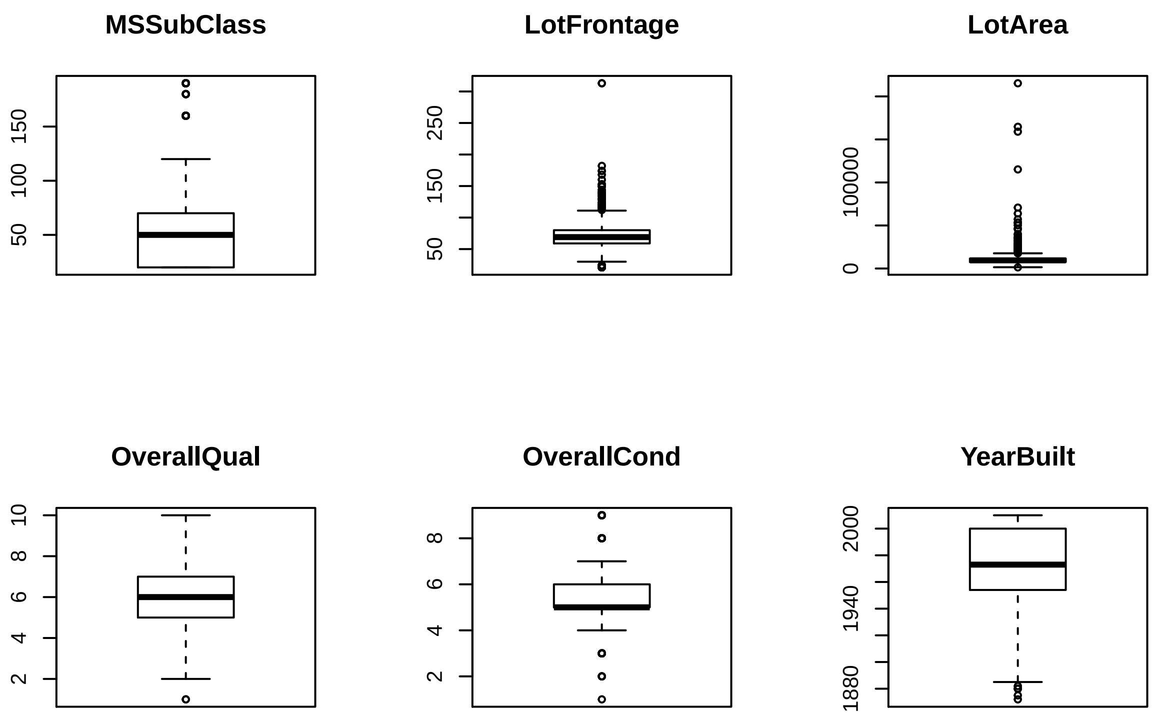
```
##
##      YrSold       SaleType         SaleCondition        SalePrice
##  Min.   :2006   Length:1460        Length:1460        Min.   : 34900
##  1st Qu.:2007   Class :character   Class :character   1st Qu.:129975
##  Median :2008   Mode  :character   Mode  :character   Median :163000
##  Mean   :2008                                         Mean   :180921
##  3rd Qu.:2009                                         3rd Qu.:214000
##  Max.   :2010                                         Max.   :755000
##
```

Table For Numerical Column Values

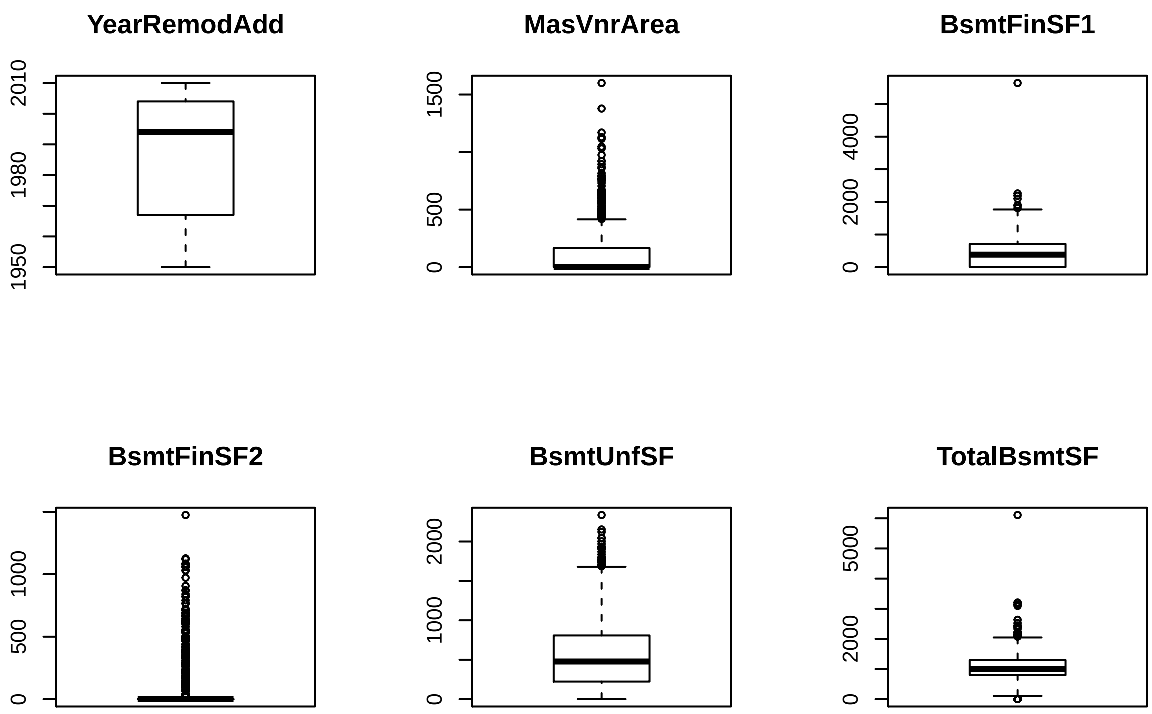
```r
traindf_numeric <- train_df[c(2,4,5, 18:21, 27,35, 37:39, 44:53, 55, 57, 60, 62, 63, 67, 72, 76:78, 81)]

traindf_univariate_df <- basicStats(traindf_numeric)[c("Minimum", "Maximum", "1. Quartile", "3. Quartile
                               "Variance", "Stdev", "Skewness", "Kurtosis"), ] %>% t() %>% as.data.frame

kable(traindf_univariate_df)
```

	Minimum	Maximum	1. Quartile	3. Quartile	Mean	Median	Variance	Stdev	Skewne
MSSubClass	20	190	20	70	5.7e+01	50	1.8e+03	4.2e+01	1.
LotFrontage	21	313	59	80	7.0e+01	69	5.9e+02	2.4e+01	2.
LotArea	1300	215245	7554	11602	1.1e+04	9478	1.0e+08	1.0e+04	12.
OverallQual	1	10	5	7	6.1e+00	6	1.9e+00	1.4e+00	0.
OverallCond	1	9	5	6	5.6e+00	5	1.2e+00	1.1e+00	0.
YearBuilt	1872	2010	1954	2000	2.0e+03	1973	9.1e+02	3.0e+01	-0.
YearRemodAdd	1950	2010	1967	2004	2.0e+03	1994	4.3e+02	2.1e+01	-0.
MasVnrArea	0	1600	0	166	1.0e+02	0	3.3e+04	1.8e+02	2.
BsmtFinSF1	0	5644	0	712	4.4e+02	384	2.1e+05	4.6e+02	1.
BsmtFinSF2	0	1474	0	0	4.7e+01	0	2.6e+04	1.6e+02	4.
BsmtUnfSF	0	2336	223	808	5.7e+02	478	2.0e+05	4.4e+02	0.
TotalBsmtSF	0	6110	796	1298	1.1e+03	992	1.9e+05	4.4e+02	1.
X1stFlrSF	334	4692	882	1391	1.2e+03	1087	1.5e+05	3.9e+02	1.
X2ndFlrSF	0	2065	0	728	3.5e+02	0	1.9e+05	4.4e+02	0.
LowQualFinSF	0	572	0	0	5.8e+00	0	2.4e+03	4.9e+01	8.
GrLivArea	334	5642	1130	1777	1.5e+03	1464	2.8e+05	5.3e+02	1.
BsmtFullBath	0	3	0	1	4.3e-01	0	2.7e-01	5.2e-01	0.
BsmtHalfBath	0	2	0	0	6.0e-02	0	6.0e-02	2.4e-01	4.
FullBath	0	3	1	2	1.6e+00	2	3.0e-01	5.5e-01	0.
HalfBath	0	2	0	1	3.8e-01	0	2.5e-01	5.0e-01	0.
BedroomAbvGr	0	8	2	3	2.9e+00	3	6.7e-01	8.2e-01	0.
KitchenAbvGr	0	3	1	1	1.0e+00	1	5.0e-02	2.2e-01	4.
TotRmsAbvGrd	2	14	5	7	6.5e+00	6	2.6e+00	1.6e+00	0.
Fireplaces	0	3	0	1	6.1e-01	1	4.2e-01	6.4e-01	0.
GarageYrBlt	1900	2010	1961	2002	2.0e+03	1980	6.1e+02	2.5e+01	-0.
GarageCars	0	4	1	2	1.8e+00	2	5.6e-01	7.5e-01	-0.
GarageArea	0	1418	334	576	4.7e+02	480	4.6e+04	2.1e+02	0.
WoodDeckSF	0	857	0	168	9.4e+01	0	1.6e+04	1.3e+02	1.
PoolArea	0	738	0	0	2.8e+00	0	1.6e+03	4.0e+01	14.
MiscVal	0	15500	0	0	4.3e+01	0	2.5e+05	5.0e+02	24.
MoSold	1	12	5	8	6.3e+00	6	7.3e+00	2.7e+00	0.
YrSold	2006	2010	2007	2009	2.0e+03	2008	1.8e+00	1.3e+00	0.
SalePrice	34900	755000	129975	214000	1.8e+05	163000	6.3e+09	7.9e+04	1.

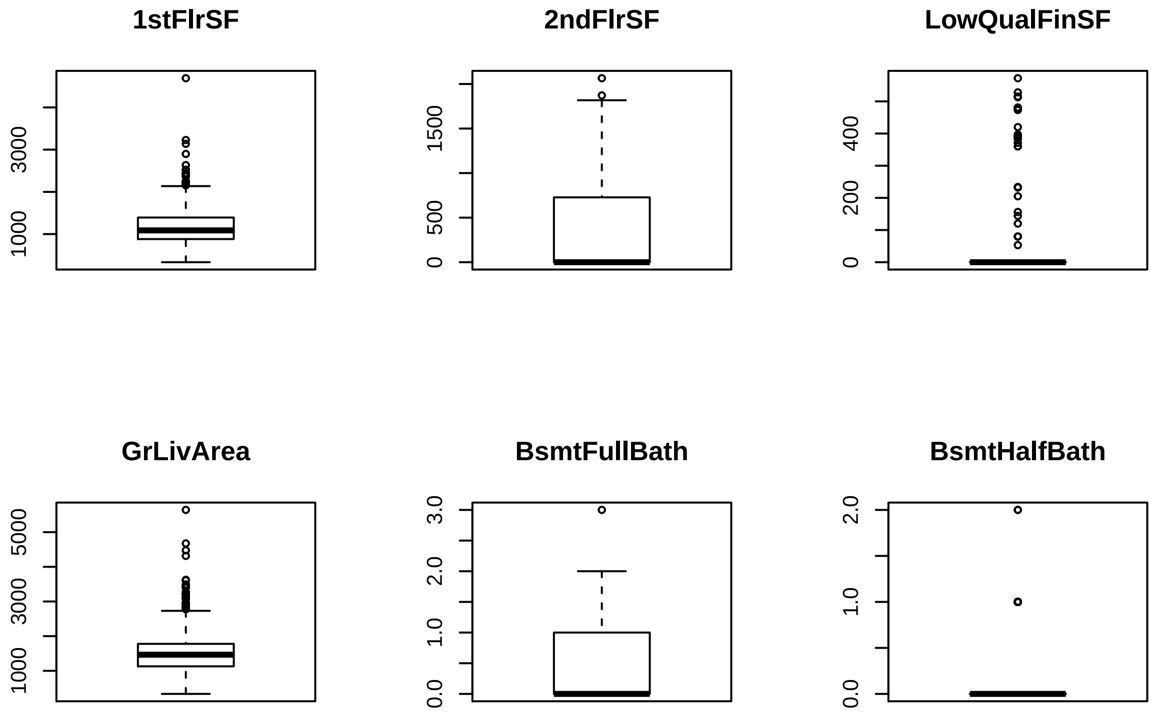
```r
par(mfrow=c(2,3))
boxplot(train_df$MSSubClass, main='MSSubClass')
boxplot(train_df$LotFrontage, main='LotFrontage')
boxplot(train_df$LotArea, main='LotArea')
boxplot(train_df$OverallQual, main='OverallQual')
boxplot(train_df$OverallCond, main='OverallCond')
boxplot(train_df$YearBuilt, main='YearBuilt')
```



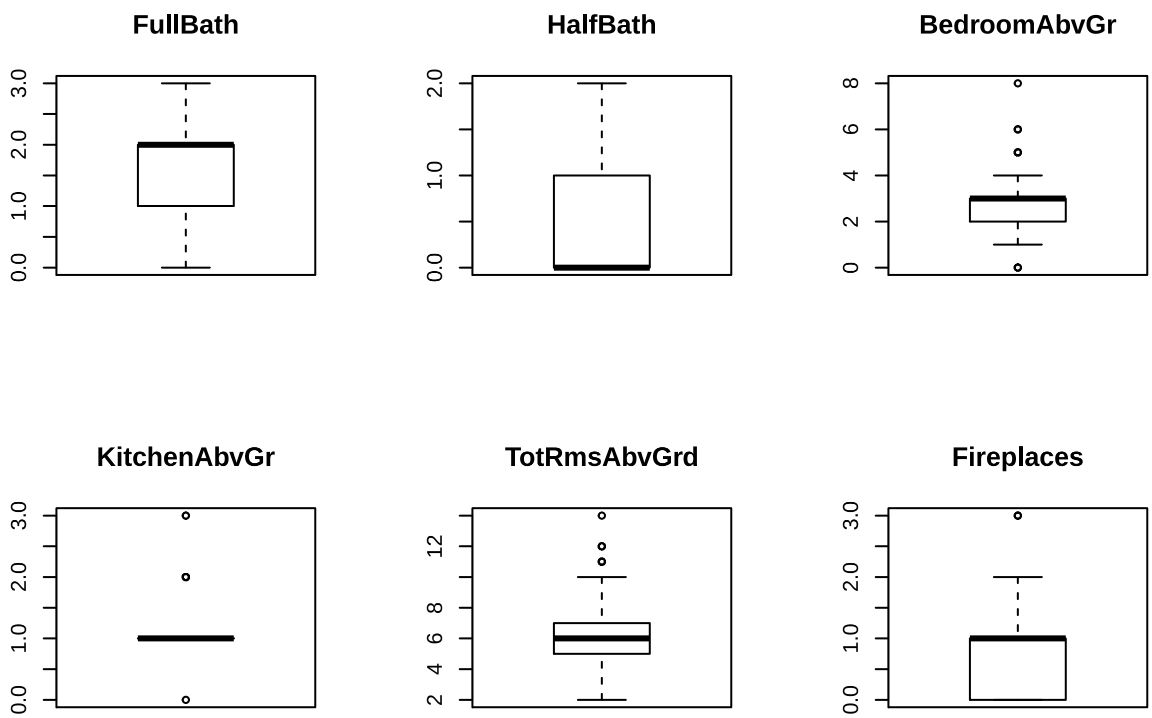

```r
par(mfrow=c(2,3))
boxplot(train_df$YearRemodAdd, main='YearRemodAdd')
boxplot(train_df$MasVnrArea, main='MasVnrArea')
boxplot(train_df$BsmtFinSF1, main='BsmtFinSF1')
boxplot(train_df$BsmtFinSF2, main='BsmtFinSF2')
boxplot(train_df$BsmtUnfSF, main='BsmtUnfSF')
boxplot(train_df$TotalBsmtSF, main='TotalBsmtSF')
```



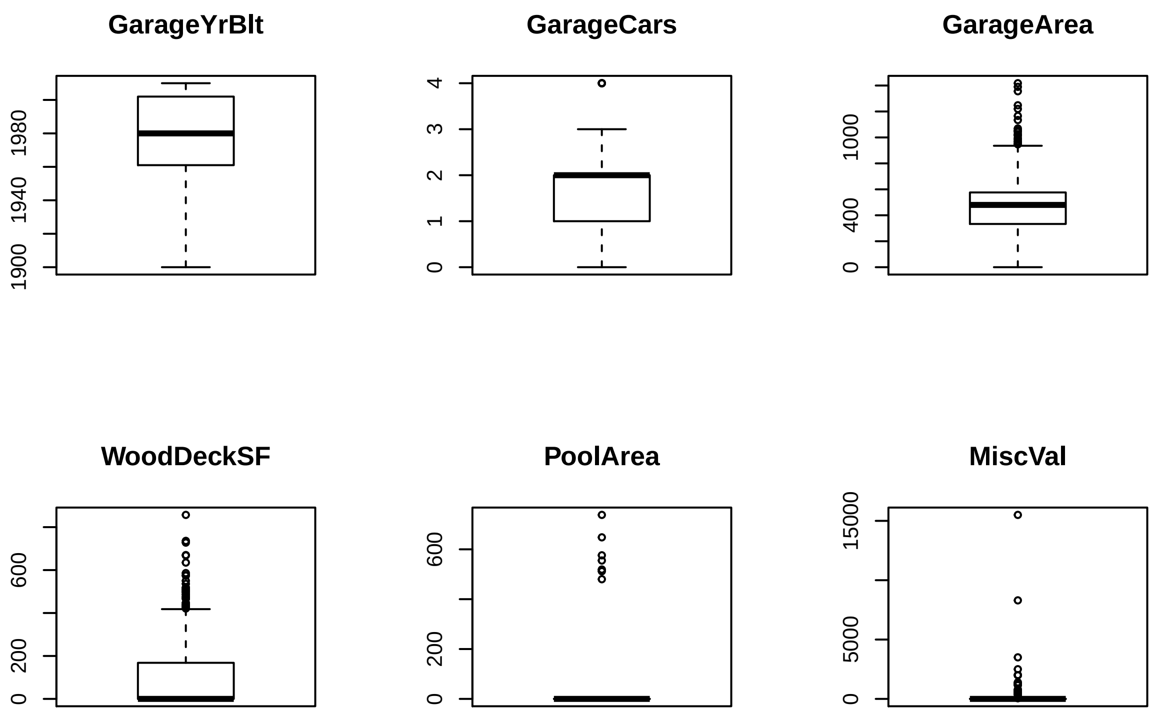

```
par(mfrow=c(2,3))
boxplot(train_df$'1stFlrSF', main='1stFlrSF')
boxplot(train_df$'2ndFlrSF', main='2ndFlrSF')
boxplot(train_df$LowQualFinSF, main='LowQualFinSF')
boxplot(train_df$GrLivArea, main='GrLivArea')
boxplot(train_df$BsmtFullBath, main='BsmtFullBath')
boxplot(train_df$BsmtHalfBath, main='BsmtHalfBath')
```



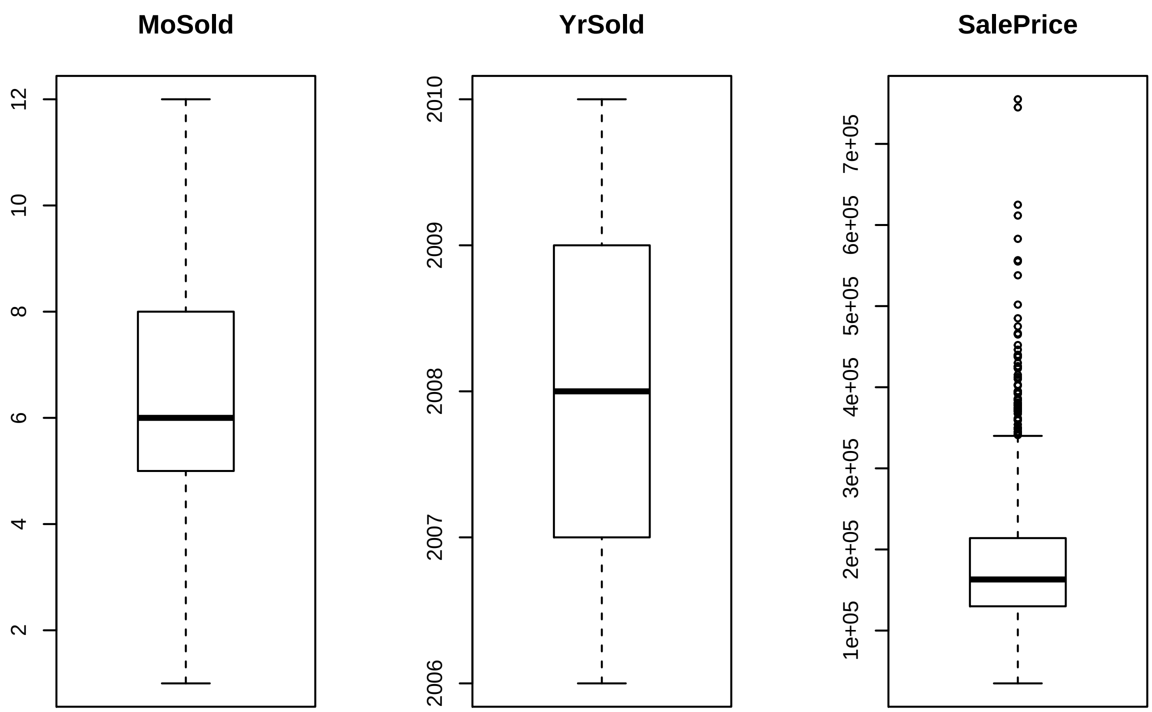

```r
par(mfrow=c(2,3))
boxplot(train_df$FullBath, main='FullBath')
boxplot(train_df$HalfBath, main='HalfBath')
boxplot(train_df$BedroomAbvGr, main='BedroomAbvGr')
boxplot(train_df$KitchenAbvGr, main='KitchenAbvGr')
boxplot(train_df$TotRmsAbvGrd, main='TotRmsAbvGrd')
boxplot(train_df$Fireplaces, main='Fireplaces')
```




```
par(mfrow=c(2,3))
boxplot(train_df$GarageYrBlt, main='GarageYrBlt')
boxplot(train_df$GarageCars, main='GarageCars')
boxplot(train_df$GarageArea, main='GarageArea')
boxplot(train_df$WoodDeckSF, main='WoodDeckSF')
boxplot(train_df$PoolArea, main='PoolArea')
boxplot(train_df$MiscVal, main='MiscVal')
```

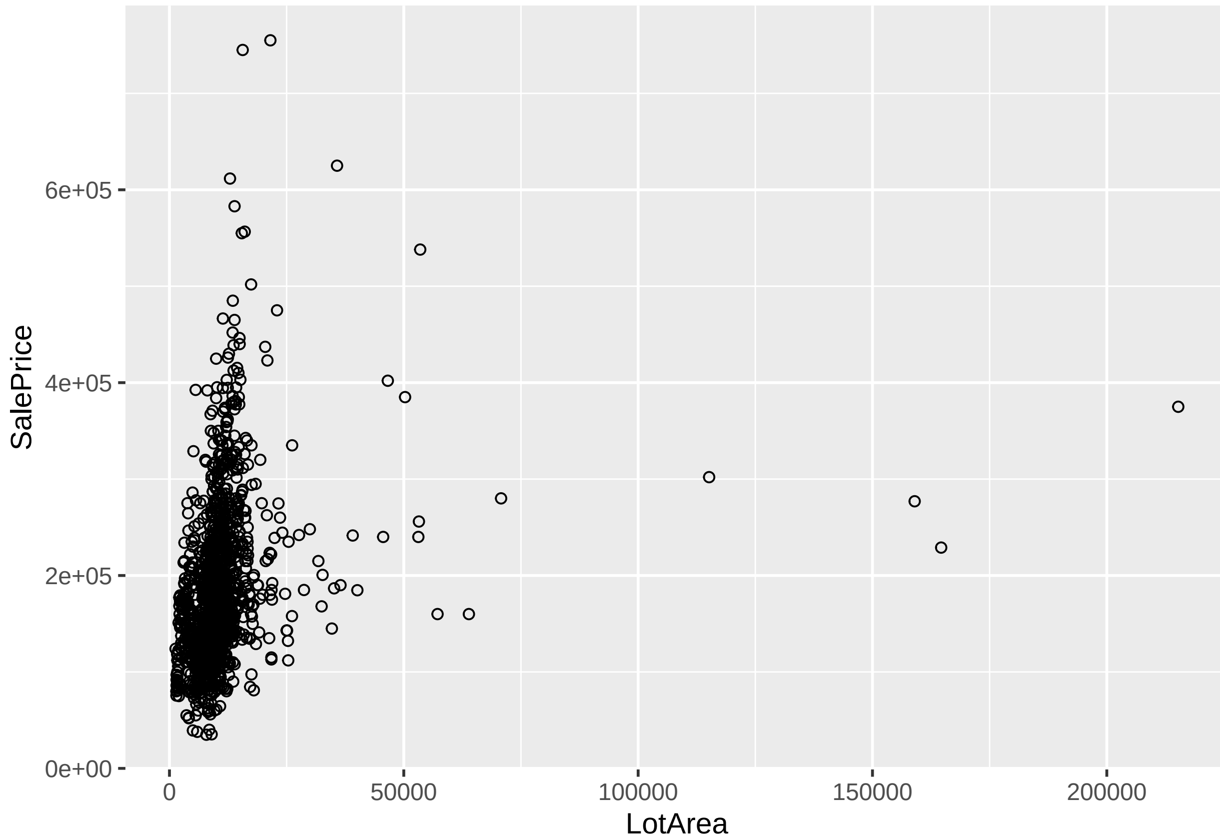



```r
par(mfrow=c(1,3))
boxplot(train_df$MoSold, main='MoSold')
boxplot(train_df$YrSold, main='YrSold')
boxplot(train_df$SalePrice, main='SalePrice')
```

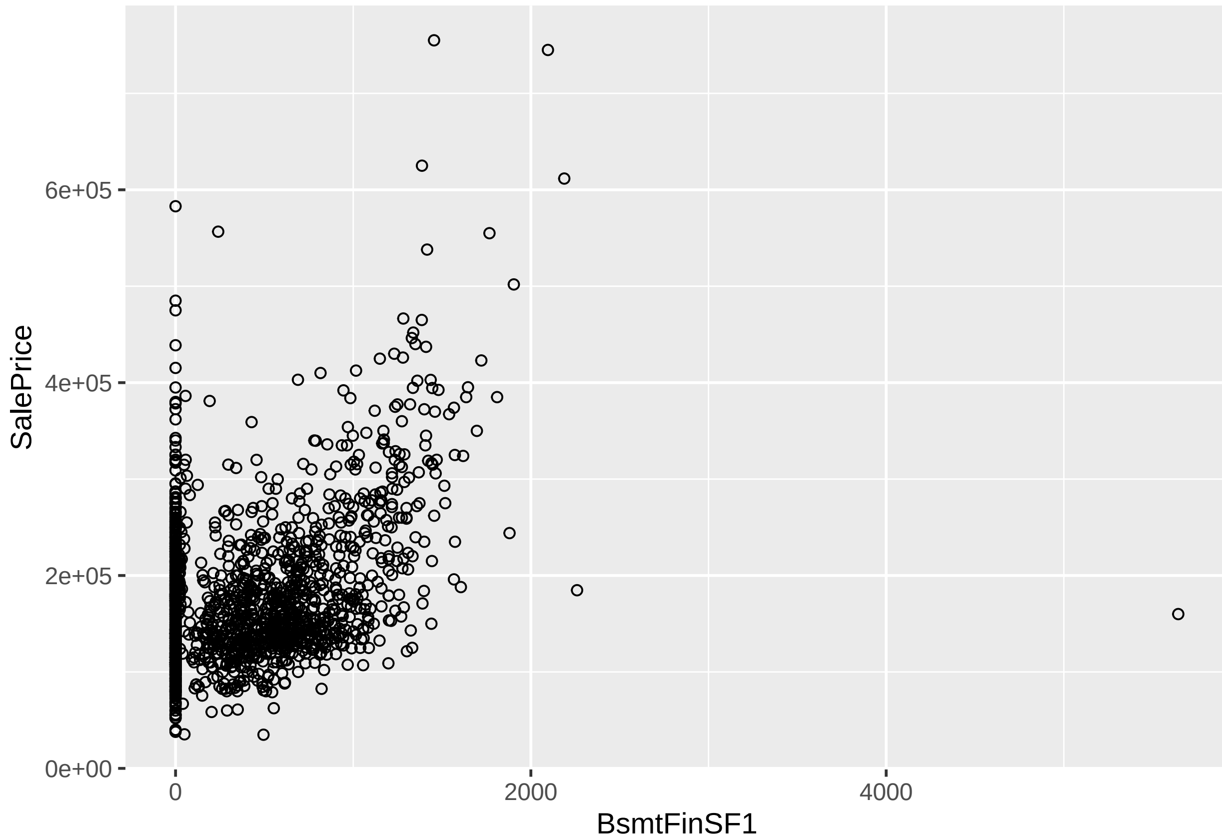


Scatter Plots for some independent variables and the response variable

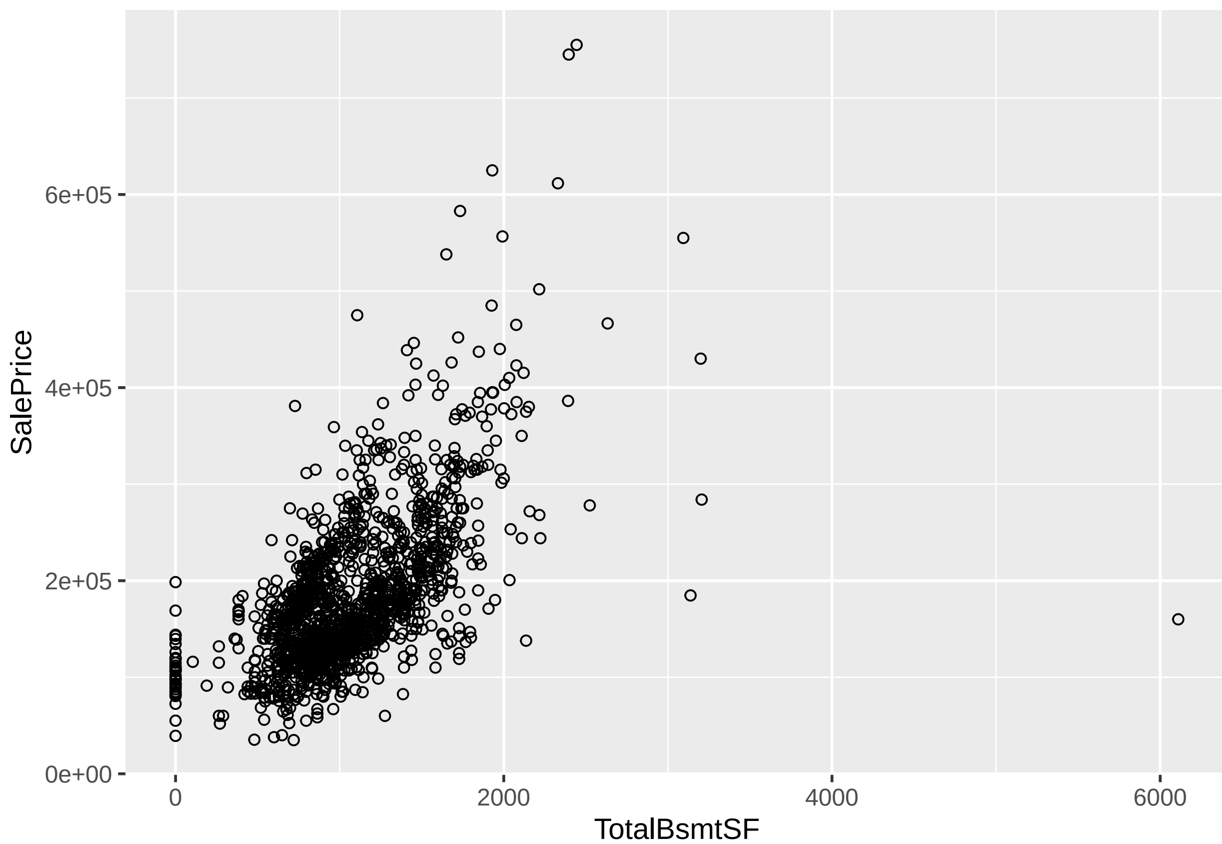
```
par(mfrow=c(2,3))
ggplot(train_df, aes(x=LotArea, y=SalePrice)) +
  geom_point(shape=1)
```



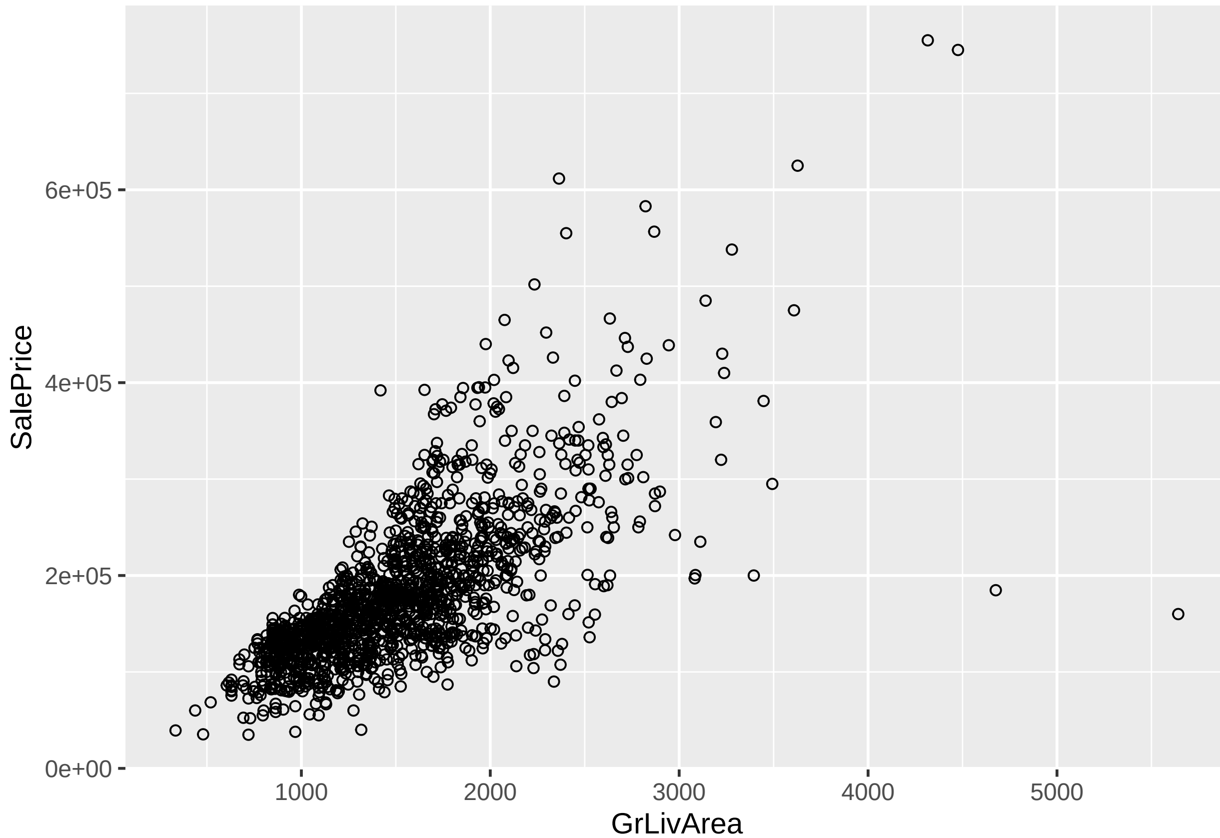

```
ggplot(train_df, aes(x=BsmtFinSF1, y=SalePrice)) +
  geom_point(shape=1)
```



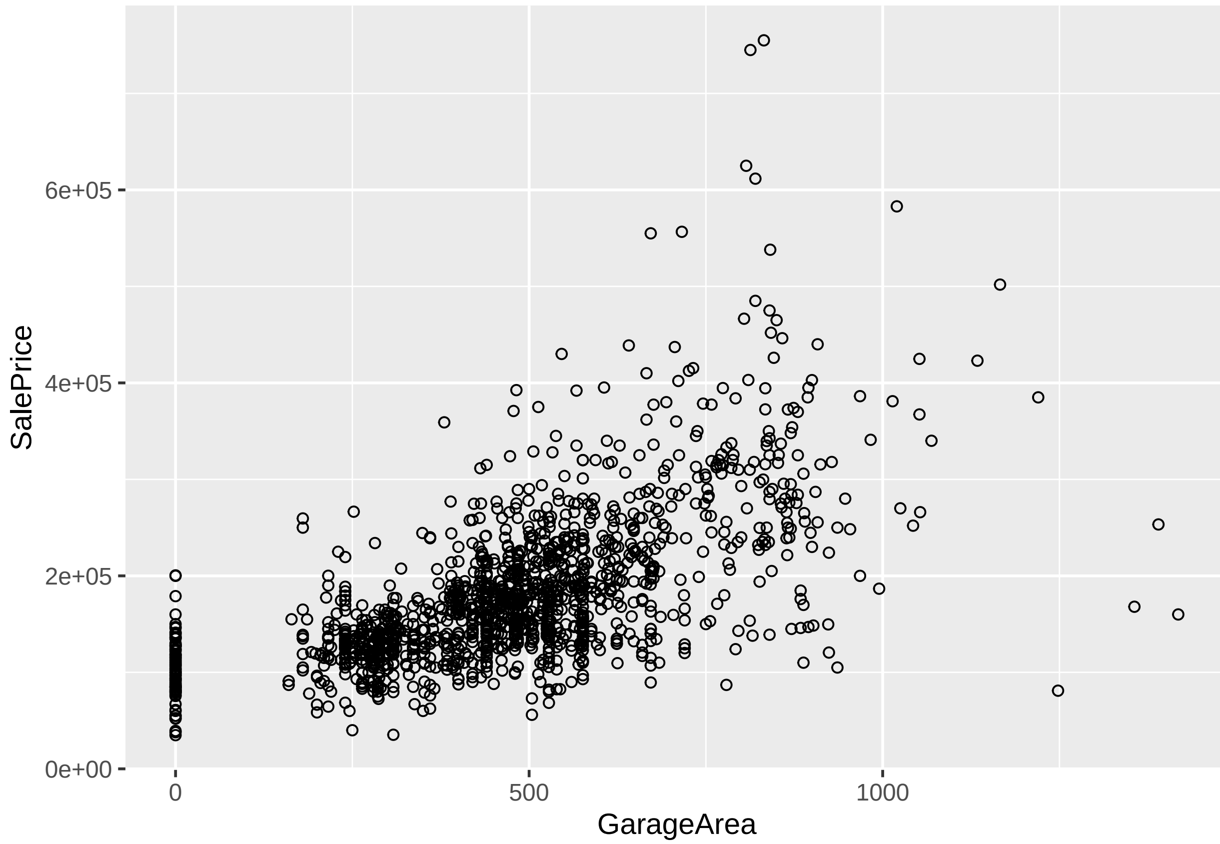

```r
ggplot(train_df, aes(x=TotalBsmtSF, y=SalePrice)) +
  geom_point(shape=1)
```




```
ggplot(train_df, aes(x=GrLivArea, y=SalePrice)) +
  geom_point(shape=1)
```




```r
ggplot(train_df, aes(x=GarageArea, y=SalePrice)) +
  geom_point(shape=1)
```



The variable "$GrLivArea$", which refers to the area above ground, has a linear relationship with the variable "$SalePrice$".

In addition, the variable $"GarageArea" $ appears to also have a good relationship, although there are homes available with no garage area.

From here, I am going to focus on three variables, "$LotArea$" , "$GrLivArea$" & "$SalePrice$"

```
x <- train_df$LotArea
y <- train_df$GrLivArea
z <- train_df$SalePrice
```

```
cor(y,z)
```

```
## [1] 0.71
```

```
cor(x,z)
```

```
## [1] 0.26
```

Living Area (Y) & Sales Price (Z)

$H0$: correlation between Y and Z = 0

HA: correlation between Y and Z > 0

T-testing to get 80% confidence level:

```
t.test(y, z, conf.level = 0.8)
```

```
##
##  Welch Two Sample t-test
##
## data:  y and z
## t = -90, df = 1000, p-value <2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 80 percent confidence interval:
##  -182071 -176740
## sample estimates:
## mean of x mean of y
##      1515    180921
```

There is a 80% confidence level where the difference in the means of the 2 variables is between -182071.5 and -176740.0.

In addition, the p-value is 2.2e-16 which is less than the significance value of 0.05.

Therefore, we reject the null hypothesis, the result is that the correlation between Living Area and Sale Price is in fact not 0, meaning that these are related to each other.

Lot Area (X) & sale price (Z)

$H0$: correlation between X and Z = 0

HA: correlation between X and Z > 0

T-testing to get 80% confidence level:

```
t.test(x, z, conf.level = 0.8)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and z
## t = -80, df = 2000, p-value <2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 80 percent confidence interval:
##  -173091 -167718
## sample estimates:
## mean of x mean of y
##     10517    180921
```

There is a 80% confidence level that the difference in the means of the 2 variables is between -173091.0 and -167717.8.

Again, the p-value is 2.2e-16, which is less than the significance value of 0.05.

Therefore, we can reject the null hypothesis and say that the correlation between Lot Area and Sale Price is not 0.

Linear Algebra and Correlation

Invert your correlation matrix from above. (This is known as the precision matrix and contains variance inflation factors on the diagonal.) Multiply the correlation matrix by the precision matrix, and then multiply the precision matrix by the correlation matrix. Conduct LU decomposition on the matrix.

```r
matrix_1 <- data.frame(x,z)

head(matrix_1)
```

```
##       x      z
## 1  8450 208500
## 2  9600 181500
## 3 11250 223500
## 4  9550 140000
## 5 14260 250000
## 6 14115 143000
```

Correlation Matrix

```r
matrix_1_corr <- cor(matrix_1)

matrix_1_corr
```

```
##      x    z
## x 1.00 0.26
## z 0.26 1.00
```

Inverse of Correlation Matrix

```r
matrix_1_inv <- solve(matrix_1_corr)

matrix_1_inv
```

```
##       x     z
## x  1.07 -0.28
## z -0.28  1.07
```

```r
matrix_1_corr %*% matrix_1_inv
```

```
##   x z
## x 1 0
## z 0 1
```

```r
matrix_1_inv %*% matrix_1_corr
```

```
##   x z
## x 1 0
## z 0 1
```

Since the Precision Matrix is an Inverse of the Correlation Matrix, the multiplyication of the two, in either direction, will result in an identity matrix.

LU Decomposition

```
L_matrix_1_corr<- lu(matrix_1_corr)$L
L_matrix_1_corr
```

```
##      x z
## x 1.00 0
## z 0.26 1
```

```
U_matrix_1_corr<- lu(matrix_1_corr)$U
U_matrix_1_corr
```

```
##   x    z
## x 1 0.26
## z 0 0.93
```

```
L_matrix_1_corr %*% U_matrix_1_corr
```

```
##      x    z
## x 1.00 0.26
## z 0.26 1.00
```

```
identical(matrix_1_corr, L_matrix_1_corr %*% U_matrix_1_corr)
```

```
## [1] TRUE
```

Calculus-Based Probability & Statistics

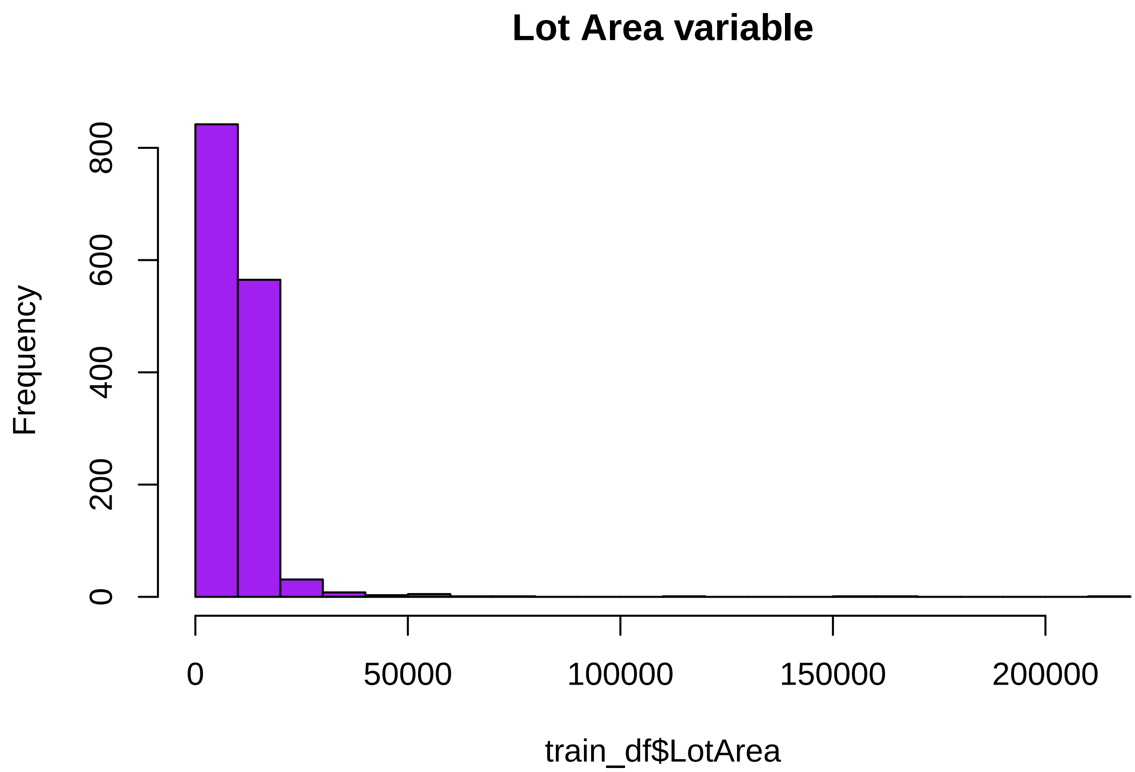
Many times, it makes sense to fit a closed form distribution to data. Select a variable in the Kaggle.com training dataset that is skewed to the right, shift it so that the minimum value is absolutely above zero if necessary. Then load the MASS package and run fitdistr to fit an exponential probability density function. (See https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/fitdistr.html). Find the optimal value of λ for this distribution, and then take 1000 samples from this exponential distribution using this value $(e.g., rexp(1000, \lambda))$. Plot a histogram and compare it with a histogram of your original variable. Using the exponential pdf, find the 5th and 95th percentiles using the cumulative distribution function (CDF). Also generate a 95% confidence interval from the empirical data, assuming normality. Finally, provide the empirical 5th percentile and 95th percentile of the data.

LotArea as it is skewed to the right.

The skewness value of this variable is 0.259.

This significantly higher than 0.1, which means it is skewed to the right.

```
hist(train_df$LotArea, col = 'purple', main = 'Lot Area variable', breaks = 30)
```



Next we will find the optimal value of lamda for this distribution, and then take 1000 samples from this exponential distribution using this value $(e.g., rexp(1000, \lambda))$.

Now we are going to fit this variable to an exponential distribution.

```
fitted_lot_area <- fitdistr(train_df$LotArea, "exponential")
```

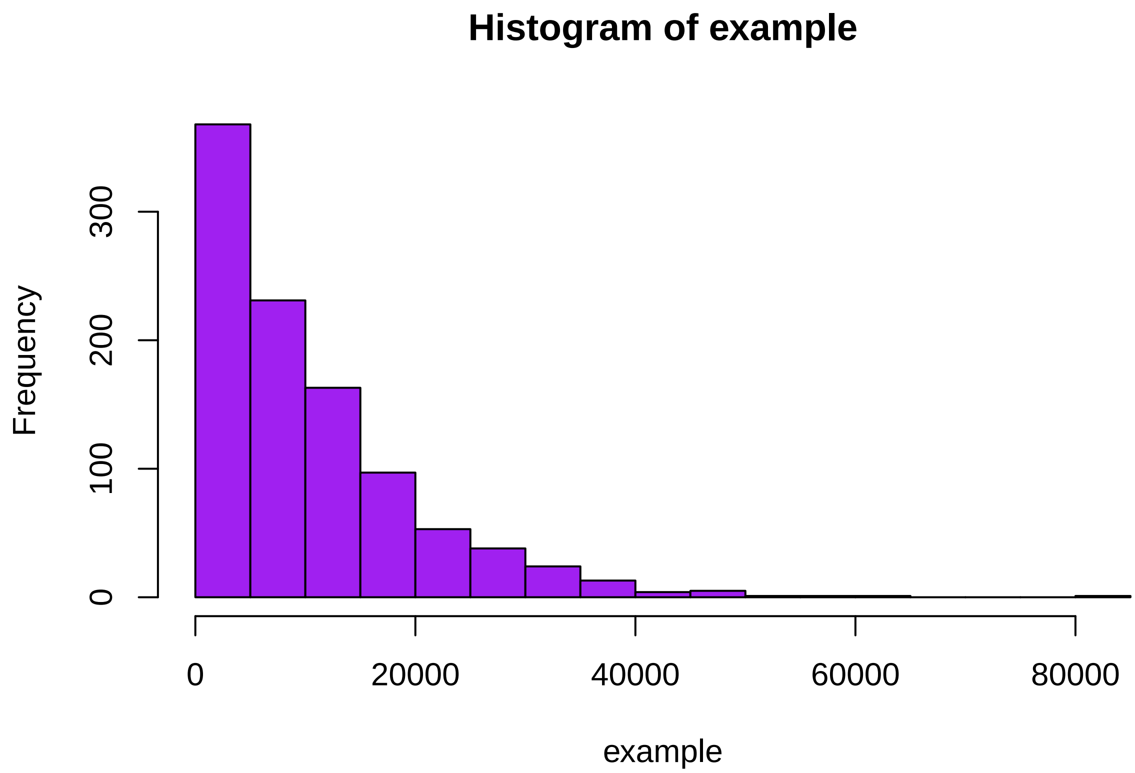
```
lot_area_lambda <- fitted_lot_area$estimate
lot_area_lambda
```

```
##    rate
## 9.5e-05
```

```
example <- rexp(1000, lot_area_lambda)
summary(example)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      27    3253    7593   10400   14440   81025
```

```
hist(example, col = 'purple', breaks = 20)
```



Generating the 5th and 95th percentiles

```
qexp(c(0.05,0.95), lot_area_lambda)
```

```
## [1]   539 31506
```

```
qnorm(c(0.025, 0.975), mean = mean(train_df$LotArea), sd = sd(train_df$LotArea))
```

```
## [1] -9046 30080
```

```
quantile(train_df$LotArea, c(0.05, 0.95))
```

```
##    5%   95%
##  3312 17401
```

The lowest 5% of the observations are below 3311 sq. ft. of Lot Area, whereas the upper 5% values are above 17401 sq. ft.

Therefore, the 90% fall under this vector.

Modeling

Build some type of multiple regression model and submit your model to the competition board. Provide your complete model summary and results with analysis. Report your Kaggle.com user name and score.

```r
lm_sale_price <- lm(SalePrice ~ ., data = traindf_numeric)
summary(lm_sale_price)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = traindf_numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -448886  -17213   -2076   15074  314113
##
## Coefficients: (2 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.38e+05   1.70e+06   -0.20  0.84270
## MSSubClass   -2.00e+02   3.45e+01   -5.80  8.8e-09 ***
## LotFrontage  -1.21e+02   6.10e+01   -1.98  0.04762 *
## LotArea       5.52e-01   1.57e-01    3.51  0.00047 ***
## OverallQual   1.88e+04   1.47e+03   12.75  < 2e-16 ***
## OverallCond   5.40e+03   1.35e+03    4.00  6.7e-05 ***
## YearBuilt     3.03e+02   8.44e+01    3.59  0.00035 ***
## YearRemodAdd  1.15e+02   8.64e+01    1.33  0.18398
## MasVnrArea    3.22e+01   6.99e+00    4.61  4.5e-06 ***
## BsmtFinSF1    1.77e+01   5.82e+00    3.04  0.00246 **
## BsmtFinSF2    9.49e+00   8.72e+00    1.09  0.27708
## BsmtUnfSF     5.09e+00   5.25e+00    0.97  0.33197
## TotalBsmtSF         NA         NA      NA       NA
## `1stFlrSF`    4.68e+01   7.36e+00    6.36  3.0e-10 ***
## `2ndFlrSF`    4.66e+01   6.05e+00    7.70  3.1e-14 ***
## LowQualFinSF  3.73e+01   2.79e+01    1.34  0.18115
## GrLivArea           NA         NA      NA       NA
## BsmtFullBath  8.90e+03   3.19e+03    2.79  0.00543 **
## BsmtHalfBath  2.48e+03   5.07e+03    0.49  0.62391
## FullBath      5.61e+03   3.53e+03    1.59  0.11156
## HalfBath     -3.96e+02   3.30e+03   -0.12  0.90456
## BedroomAbvGr -1.00e+04   2.15e+03   -4.65  3.7e-06 ***
## KitchenAbvGr -2.29e+04   6.70e+03   -3.42  0.00064 ***
## TotRmsAbvGrd  5.23e+03   1.48e+03    3.53  0.00043 ***
## Fireplaces    5.06e+03   2.18e+03    2.32  0.02035 *
## GarageYrBlt  -5.46e+01   9.11e+01   -0.60  0.54922
## GarageCars    1.72e+04   3.48e+03    4.95  8.7e-07 ***
## GarageArea    6.22e+00   1.21e+01    0.52  0.60640
## WoodDeckSF    1.70e+01   9.88e+00    1.72  0.08603 .
## PoolArea     -5.96e+01   2.98e+01   -2.00  0.04571 *
## MiscVal      -5.26e-01   6.87e+00   -0.08  0.93894
## MoSold       -2.14e+02   4.21e+02   -0.51  0.61232
## YrSold       -2.22e+02   8.46e+02   -0.26  0.79299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 36900 on 1090 degrees of freedom
##   (339 observations deleted due to missingness)
## Multiple R-squared:  0.808,  Adjusted R-squared:  0.803
## F-statistic:  153 on 30 and 1090 DF,  p-value: <2e-16
```

After analysis the above summary, removing the independent variables which contain a value of "NA".

In addition, removing variables where the P-Value is significantly greater that 0.05.

```r
lm_2_sale_price <- lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual +
                      OverallCond + YearBuilt + YearRemodAdd + MasVnrArea + BsmtFinSF1 +
                      BsmtFinSF2 + BsmtUnfSF + LowQualFinSF +
                      BsmtFullBath + FullBath + BedroomAbvGr + KitchenAbvGr +
                      TotRmsAbvGrd + Fireplaces + GarageCars + WoodDeckSF +
                      PoolArea , data = traindf_numeric)
```

```r
summary(lm_2_sale_price)
```

```
##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotFrontage + LotArea +
##     OverallQual + OverallCond + YearBuilt + YearRemodAdd + MasVnrArea +
##     BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + LowQualFinSF + BsmtFullBath +
##     FullBath + BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd + Fireplaces +
##     GarageCars + WoodDeckSF + PoolArea, data = traindf_numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -419504  -19091   -2219   15359  363931
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.98e+05   1.46e+05   -5.47  5.5e-08 ***
## MSSubClass   -1.26e+02   3.17e+01   -3.97  7.7e-05 ***
## LotFrontage  -2.90e+01   5.91e+01   -0.49  0.62408
## LotArea       6.87e-01   1.60e-01    4.29  1.9e-05 ***
## OverallQual   2.02e+04   1.39e+03   14.50  < 2e-16 ***
## OverallCond   3.57e+03   1.25e+03    2.85  0.00450 **
## YearBuilt     1.26e+02   6.16e+01    2.04  0.04176 *
## YearRemodAdd  2.37e+02   7.80e+01    3.04  0.00245 **
## MasVnrArea    4.46e+01   7.03e+00    6.35  3.1e-10 ***
## BsmtFinSF1    2.93e+01   4.29e+00    6.83  1.3e-11 ***
## BsmtFinSF2    1.62e+01   7.84e+00    2.06  0.03937 *
## BsmtUnfSF     1.35e+01   3.81e+00    3.55  0.00041 ***
## LowQualFinSF  1.54e+01   2.25e+01    0.69  0.49298
## BsmtFullBath  7.43e+03   2.98e+03    2.49  0.01284 *
## FullBath      1.45e+04   3.00e+03    4.85  1.4e-06 ***
## BedroomAbvGr -7.77e+03   1.99e+03   -3.91  9.8e-05 ***
## KitchenAbvGr -1.93e+04   5.74e+03   -3.36  0.00080 ***
## TotRmsAbvGrd  1.22e+04   1.23e+03    9.90  < 2e-16 ***
## Fireplaces    8.75e+03   2.09e+03    4.18  3.1e-05 ***
## GarageCars    1.32e+04   1.97e+03    6.67  3.9e-11 ***
## WoodDeckSF    2.84e+01   9.84e+00    2.88  0.00401 **
```

```
## PoolArea     -2.37e+01   2.96e+01   -0.80  0.42444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37900 on 1173 degrees of freedom
##   (265 observations deleted due to missingness)
## Multiple R-squared:  0.796,  Adjusted R-squared:  0.792
## F-statistic:  218 on 21 and 1173 DF,  p-value: <2e-16
```

Now it is time to import our Test set

```
test_raw <- "https://raw.githubusercontent.com/josephsimone/Data-605/master/test.csv"
test_df <- read.csv(test_raw, header = TRUE, sep = ",")
head(test_df)
```

```
##     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
## 1         Lvl    AllPub    Inside       Gtl        NAmes      Feedr
## 2         Lvl    AllPub    Corner       Gtl        NAmes       Norm
## 3         Lvl    AllPub    Inside       Gtl      Gilbert       Norm
## 4         Lvl    AllPub    Inside       Gtl      Gilbert       Norm
## 5         HLS    AllPub    Inside       Gtl      StoneBr       Norm
## 6         Lvl    AllPub    Corner       Gtl      Gilbert       Norm
##   Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt
## 1       Norm     1Fam     1Story           5           6      1961
## 2       Norm     1Fam     1Story           6           6      1958
## 3       Norm     1Fam     2Story           5           5      1997
## 4       Norm     1Fam     2Story           6           6      1998
## 5       Norm   TwnhsE     1Story           8           5      1992
## 6       Norm     1Fam     2Story           6           5      1993
##   YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType
## 1         1961     Gable  CompShg     VinylSd     VinylSd       None
## 2         1958       Hip  CompShg     Wd Sdng     Wd Sdng    BrkFace
## 3         1998     Gable  CompShg     VinylSd     VinylSd       None
## 4         1998     Gable  CompShg     VinylSd     VinylSd    BrkFace
## 5         1992     Gable  CompShg     HdBoard     HdBoard       None
## 6         1994     Gable  CompShg     HdBoard     HdBoard       None
##   MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure
## 1          0        TA        TA     CBlock       TA       TA           No
## 2        108        TA        TA     CBlock       TA       TA           No
## 3          0        TA        TA      PConc       Gd       TA           No
## 4         20        TA        TA      PConc       TA       TA           No
## 5          0        Gd        TA      PConc       Gd       TA           No
## 6          0        TA        TA      PConc       Gd       TA           No
##   BsmtFinType1 BsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF TotalBsmtSF
## 1          Rec        468          LwQ        144       270         882
## 2          ALQ        923          Unf          0       406        1329
```

```
## 3          GLQ        791          Unf          0       137         928
## 4          GLQ        602          Unf          0       324         926
## 5          ALQ        263          Unf          0      1017        1280
## 6          Unf          0          Unf          0       763         763
##   Heating HeatingQC CentralAir Electrical X1stFlrSF X2ndFlrSF LowQualFinSF
## 1    GasA        TA          Y      SBrkr       896         0            0
## 2    GasA        TA          Y      SBrkr      1329         0            0
## 3    GasA        Gd          Y      SBrkr       928       701            0
## 4    GasA        Ex          Y      SBrkr       926       678            0
## 5    GasA        Ex          Y      SBrkr      1280         0            0
## 6    GasA        Gd          Y      SBrkr       763       892            0
##   GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr
## 1       896            0            0        1        0            2
## 2      1329            0            0        1        1            3
## 3      1629            0            0        2        1            3
## 4      1604            0            0        2        1            3
## 5      1280            0            0        2        0            2
## 6      1655            0            0        2        1            3
##   KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu
## 1            1          TA            5        Typ          0        <NA>
## 2            1          Gd            6        Typ          0        <NA>
## 3            1          TA            6        Typ          1          TA
## 4            1          Gd            7        Typ          1          Gd
## 5            1          Gd            5        Typ          0        <NA>
## 6            1          TA            7        Typ          1          TA
##   GarageType GarageYrBlt GarageFinish GarageCars GarageArea GarageQual
## 1     Attchd        1961          Unf          1        730         TA
## 2     Attchd        1958          Unf          1        312         TA
## 3     Attchd        1997          Fin          2        482         TA
## 4     Attchd        1998          Fin          2        470         TA
## 5     Attchd        1992          RFn          2        506         TA
## 6     Attchd        1993          Fin          2        440         TA
##   GarageCond PavedDrive WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch
## 1         TA          Y        140           0             0          0
## 2         TA          Y        393          36             0          0
## 3         TA          Y        212          34             0          0
## 4         TA          Y        360          36             0          0
## 5         TA          Y          0          82             0          0
## 6         TA          Y        157          84             0          0
##   ScreenPorch PoolArea PoolQC Fence MiscFeature MiscVal MoSold YrSold
## 1         120        0   <NA> MnPrv        <NA>       0      6   2010
## 2           0        0   <NA>  <NA>        Gar2   12500      6   2010
## 3           0        0   <NA> MnPrv        <NA>       0      3   2010
## 4           0        0   <NA>  <NA>        <NA>       0      6   2010
## 5         144        0   <NA>  <NA>        <NA>       0      1   2010
## 6           0        0   <NA>  <NA>        <NA>       0      4   2010
##   SaleType SaleCondition
## 1       WD        Normal
## 2       WD        Normal
## 3       WD        Normal
## 4       WD        Normal
## 5       WD        Normal
## 6       WD        Normal
```

Convert “NA” Valus again

```r
test_df$MasVnrArea[is.na(test_df$MasVnrArea)] <- 0
test_df$BsmtFinSF1[is.na(test_df$BsmtFinSF1)] <- 0
test_df$BsmtFullBath[is.na(test_df$BsmtFullBath)] <- 0


res_predicition <- predict(lm_2_sale_price, test_df)

res_predicition_df <- data.frame(cbind(test_df$Id, res_predicition))

colnames(res_predicition_df) = c('Id', 'SalePrice')



head(res_predicition_df, 5)
```

```
##     Id SalePrice
## 1 1461    116124
## 2 1462    166324
## 3 1463    170743
## 4 1464    206188
## 5 1465    193928
```

Kaggle Submission

Kaggle User Name: jpsimone. After my submission to the House Prices: Advanced Regression Techniques, I recieved a score of 4.78103.