Project 1

Wentian Chen

2024-10-25

```
getwd()
```

```
## [1] "C:/Users/admin/Desktop/Econ 103"
```

```
setwd("C:/Users/admin/Desktop/Econ 103")
```

```
library(readxl)
data <- read.csv("C:/Users/admin/Desktop/train.csv")
head(data)
```

```
##   Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape LandContour
## 1  1         60       RL          65    8450   Pave  <NA>      Reg         Lvl
## 2  2         20       RL          80    9600   Pave  <NA>      Reg         Lvl
## 3  3         60       RL          68   11250   Pave  <NA>      IR1         Lvl
## 4  4         70       RL          60    9550   Pave  <NA>      IR1         Lvl
## 5  5         60       RL          84   14260   Pave  <NA>      IR1         Lvl
## 6  6         50       RL          85   14115   Pave  <NA>      IR1         Lvl
##   Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType
## 1    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 2    AllPub       FR2       Gtl      Veenker      Feedr       Norm     1Fam
## 3    AllPub    Inside       Gtl      CollgCr       Norm       Norm     1Fam
## 4    AllPub    Corner       Gtl      Crawfor       Norm       Norm     1Fam
## 5    AllPub       FR2       Gtl      NoRidge       Norm       Norm     1Fam
## 6    AllPub    Inside       Gtl      Mitchel       Norm       Norm     1Fam
##   HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl
## 1     2Story           7           5      2003         2003     Gable  CompShg
## 2     1Story           6           8      1976         1976     Gable  CompShg
## 3     2Story           7           5      2001         2002     Gable  CompShg
## 4     2Story           7           5      1915         1970     Gable  CompShg
## 5     2Story           8           5      2000         2000     Gable  CompShg
## 6     1.5Fin           5           5      1993         1995     Gable  CompShg
##   Exterior1st Exterior2nd MasVnrType MasVnrArea ExterQual ExterCond Foundation
## 1     VinylSd     VinylSd    BrkFace        196        Gd        TA      PConc
## 2     MetalSd     MetalSd       None          0        TA        TA     CBlock
## 3     VinylSd     VinylSd    BrkFace        162        Gd        TA      PConc
## 4     Wd Sdng     Wd Shng       None          0        TA        TA     BrkTil
## 5     VinylSd     VinylSd    BrkFace        350        Gd        TA      PConc
## 6     VinylSd     VinylSd       None          0        TA        TA       Wood
##   BsmtQual BsmtCond BsmtExposure BsmtFinType1 BsmtFinSF1 BsmtFinType2
## 1       Gd       TA           No          GLQ        706          Unf
```

```
## 2       Gd       TA           Gd          ALQ        978          Unf
## 3       Gd       TA           Mn          GLQ        486          Unf
## 4       TA       Gd           No          ALQ        216          Unf
## 5       Gd       TA           Av          GLQ        655          Unf
## 6       Gd       TA           No          GLQ        732          Unf
##   BsmtFinSF2 BsmtUnfSF TotalBsmtSF Heating HeatingQC CentralAir Electrical
## 1          0       150         856    GasA        Ex          Y      SBrkr
## 2          0       284        1262    GasA        Ex          Y      SBrkr
## 3          0       434         920    GasA        Ex          Y      SBrkr
## 4          0       540         756    GasA        Gd          Y      SBrkr
## 5          0       490        1145    GasA        Ex          Y      SBrkr
## 6          0        64         796    GasA        Ex          Y      SBrkr
##   X1stFlrSF X2ndFlrSF LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath
## 1       856       854            0      1710            1            0        2
## 2      1262         0            0      1262            0            1        2
## 3       920       866            0      1786            1            0        2
## 4       961       756            0      1717            1            0        1
## 5      1145      1053            0      2198            1            0        2
## 6       796       566            0      1362            1            0        1
##   HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## 1        1            3            1          Gd            8        Typ
## 2        0            3            1          TA            6        Typ
## 3        1            3            1          Gd            6        Typ
## 4        0            3            1          Gd            7        Typ
## 5        1            4            1          Gd            9        Typ
## 6        1            1            1          TA            5        Typ
##   Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars
## 1          0       <NA>     Attchd        2003          RFn          2
## 2          1         TA     Attchd        1976          RFn          2
## 3          1         TA     Attchd        2001          RFn          2
## 4          1         Gd     Detchd        1998          Unf          3
## 5          1         TA     Attchd        2000          RFn          3
## 6          0       <NA>     Attchd        1993          Unf          2
##   GarageArea GarageQual GarageCond PavedDrive WoodDeckSF OpenPorchSF
## 1        548         TA         TA          Y          0          61
## 2        460         TA         TA          Y        298           0
## 3        608         TA         TA          Y          0          42
## 4        642         TA         TA          Y          0          35
## 5        836         TA         TA          Y        192          84
## 6        480         TA         TA          Y         40          30
##   EnclosedPorch X3SsnPorch ScreenPorch PoolArea PoolQC Fence MiscFeature
## 1             0          0           0        0   <NA>  <NA>        <NA>
## 2             0          0           0        0   <NA>  <NA>        <NA>
## 3             0          0           0        0   <NA>  <NA>        <NA>
## 4           272          0           0        0   <NA>  <NA>        <NA>
## 5             0          0           0        0   <NA>  <NA>        <NA>
## 6             0        320           0        0   <NA> MnPrv        Shed
##   MiscVal MoSold YrSold SaleType SaleCondition SalePrice
## 1       0      2   2008       WD        Normal    208500
## 2       0      5   2007       WD        Normal    181500
## 3       0      9   2008       WD        Normal    223500
## 4       0      2   2006       WD       Abnorml    140000
## 5       0     12   2008       WD        Normal    250000
## 6     700     10   2009       WD        Normal    143000
```

```r
#Response variable: SalePrice
#Predictors: LotFrontage, WoodDeckSF, GarageArea, GrLivArea
```

```r
#Summary of all variables
library(psych)
describe(data[c("SalePrice", "LotFrontage", "WoodDeckSF", "GarageArea", "GrLivArea")])
```

```
##             vars    n      mean       sd median   trimmed      mad   min    max
## SalePrice      1 1460 180921.20 79442.50 163000 170783.29 56338.80 34900 755000
## LotFrontage    2 1201     70.05    24.28     69     68.94    16.31    21    313
## WoodDeckSF     3 1460     94.24   125.34      0     71.76     0.00     0    857
## GarageArea     4 1460    472.98   213.80    480    469.81   177.91     0   1418
## GrLivArea      5 1460   1515.46   525.48   1464   1467.67   483.33   334   5642
##               range skew kurtosis      se
## SalePrice    720100 1.88     6.50 2079.11
## LotFrontage     292 2.16    17.34    0.70
## WoodDeckSF      857 1.54     2.97    3.28
## GarageArea     1418 0.18     0.90    5.60
## GrLivArea      5308 1.36     4.86   13.75
```

```r
#Histogram and Fitted Distribution
#Histogram of SalePrice, including density line
library(ggplot2)
```

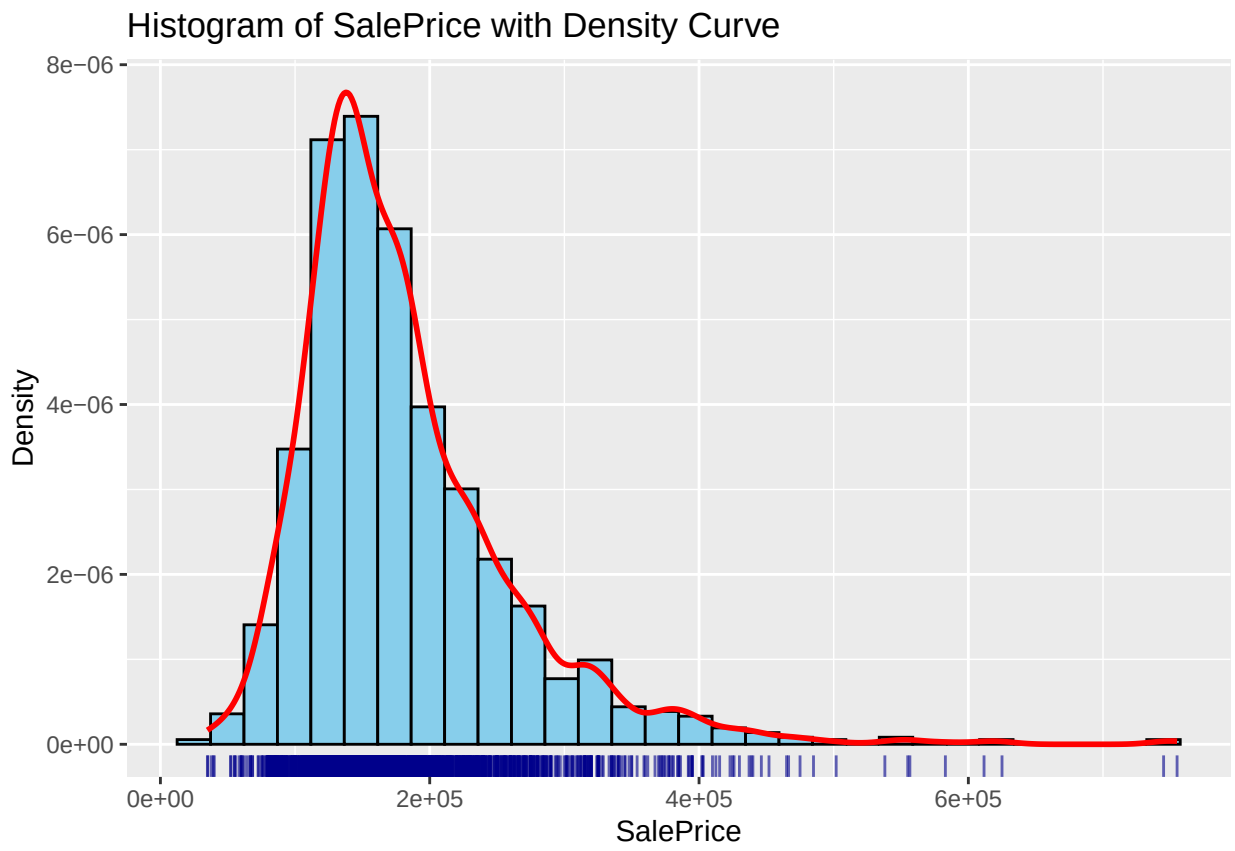
```
##
##     'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```r
ggplot(data, aes(x = SalePrice)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red", size = 1) + geom_rug(sides = "b", color = "darkblue", alpha = 0.6)+
  labs(title = "Histogram of SalePrice with Density Curve",
       x = "SalePrice", y = "Density")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

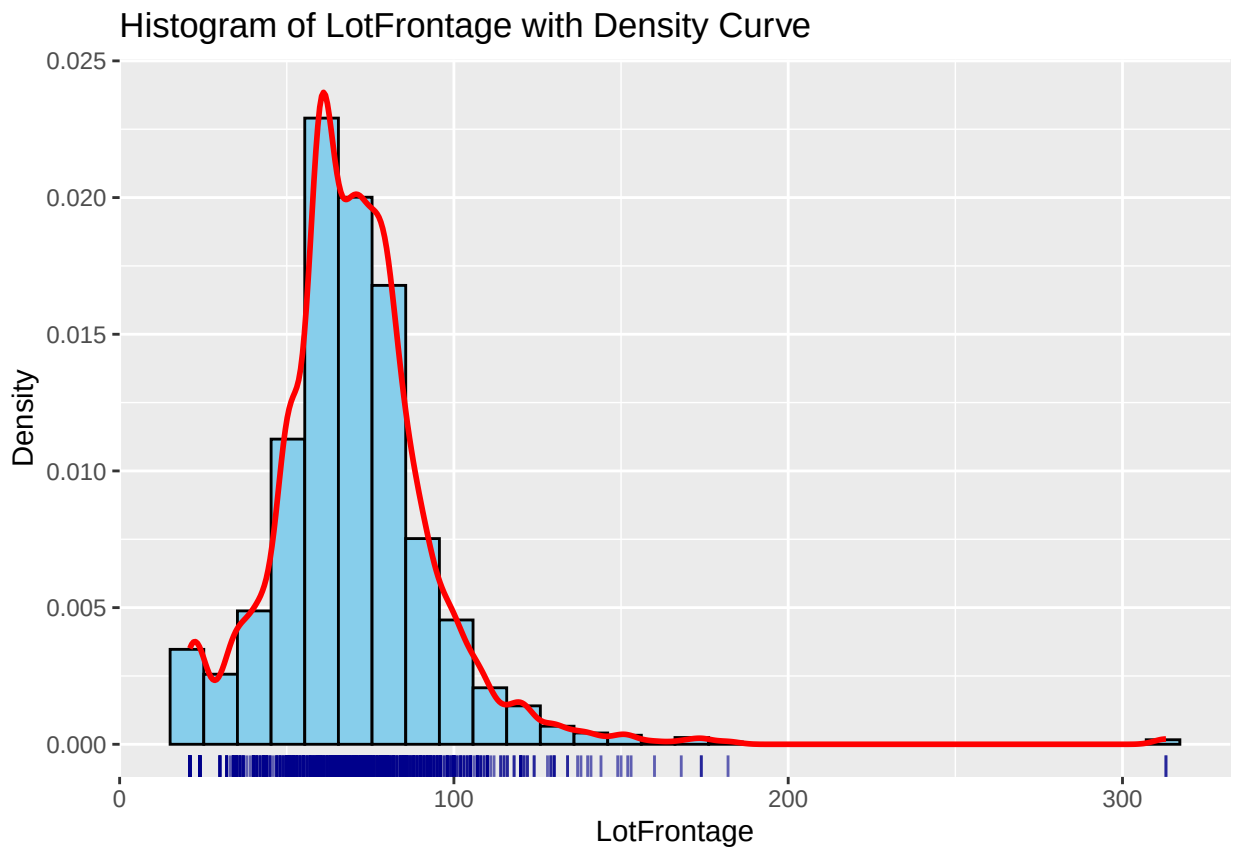
```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



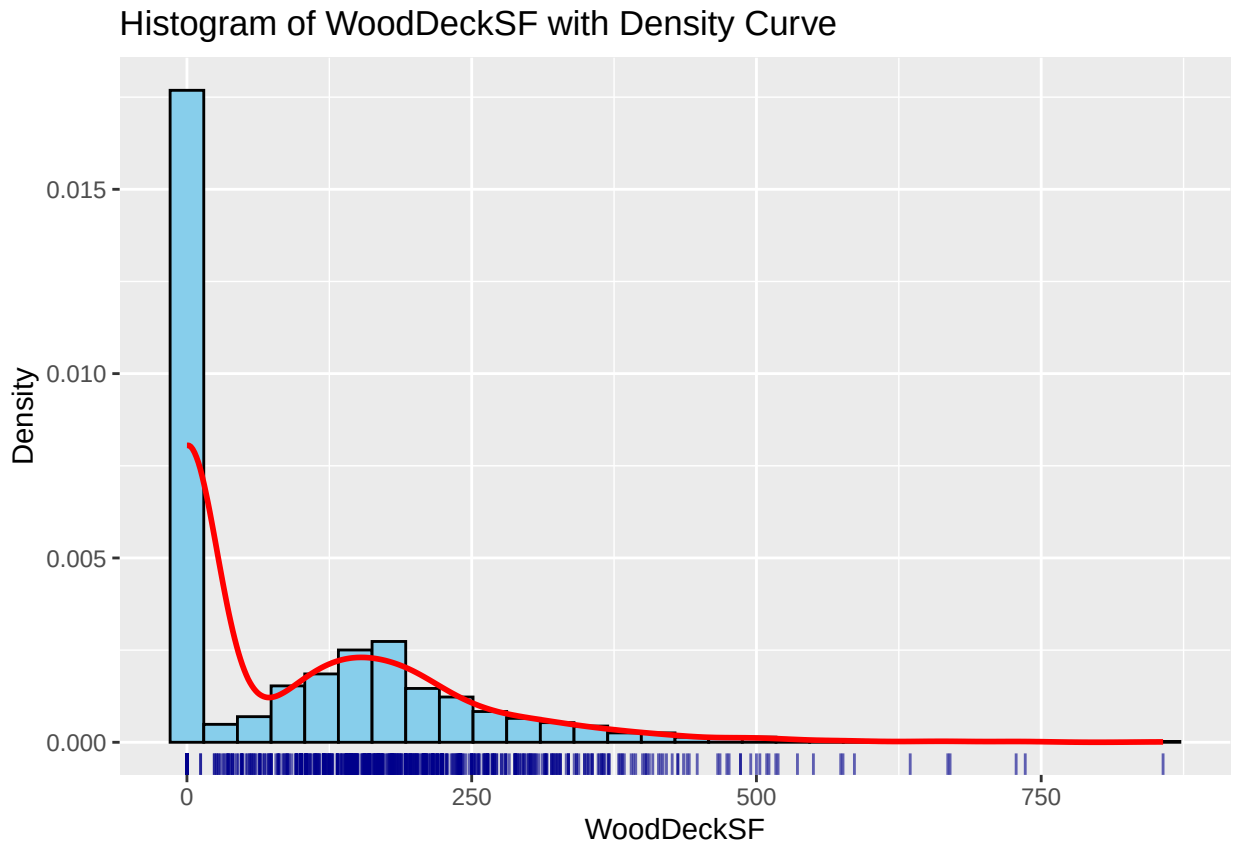

```
#Histogram of LotFrontage, including density line
ggplot(data, aes(x = LotFrontage)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red", size = 1) + geom_rug(sides = "b", color = "darkblue", alpha = 0.6)+
  labs(title = "Histogram of LotFrontage with Density Curve",
       x = "LotFrontage", y = "Density")
```

```
## Warning: Removed 259 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

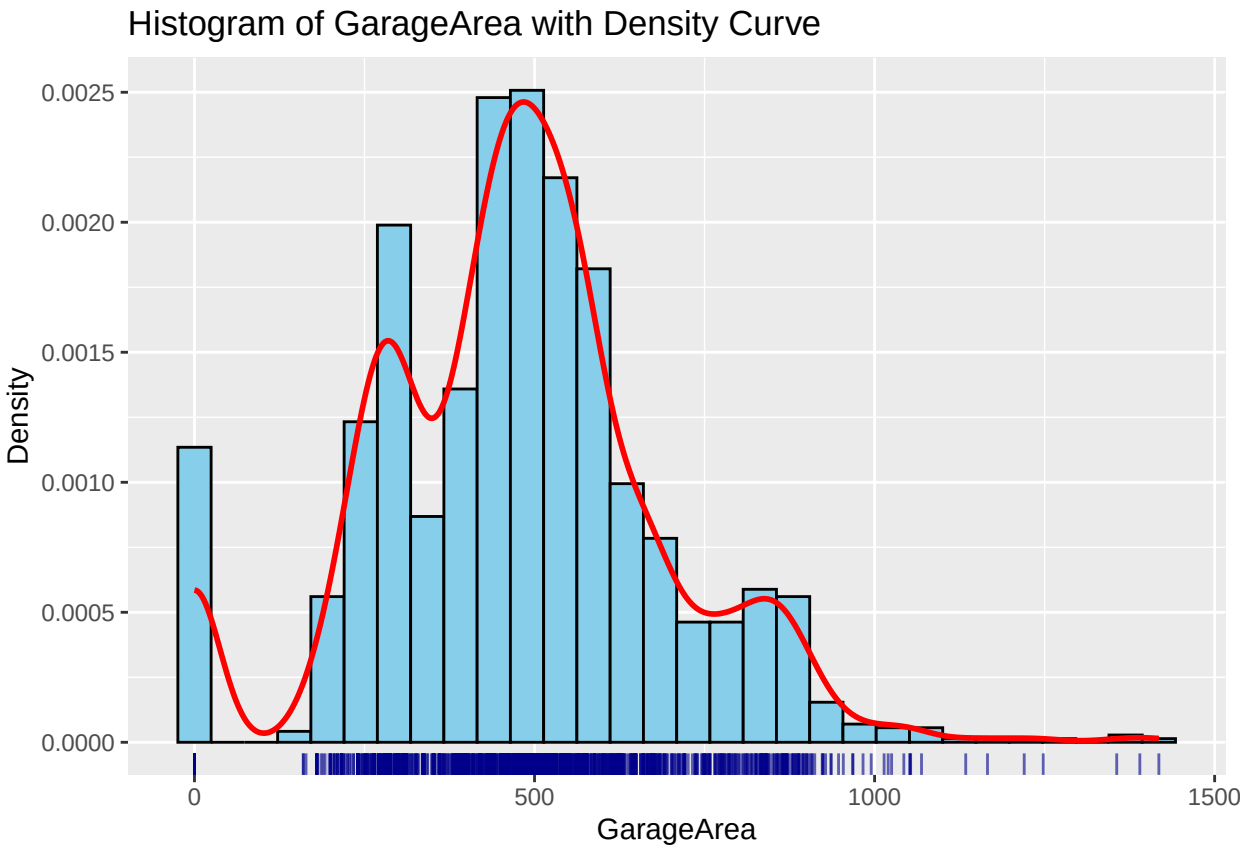
```
## Warning: Removed 259 rows containing non-finite outside the scale range
## (`stat_density()`).
```



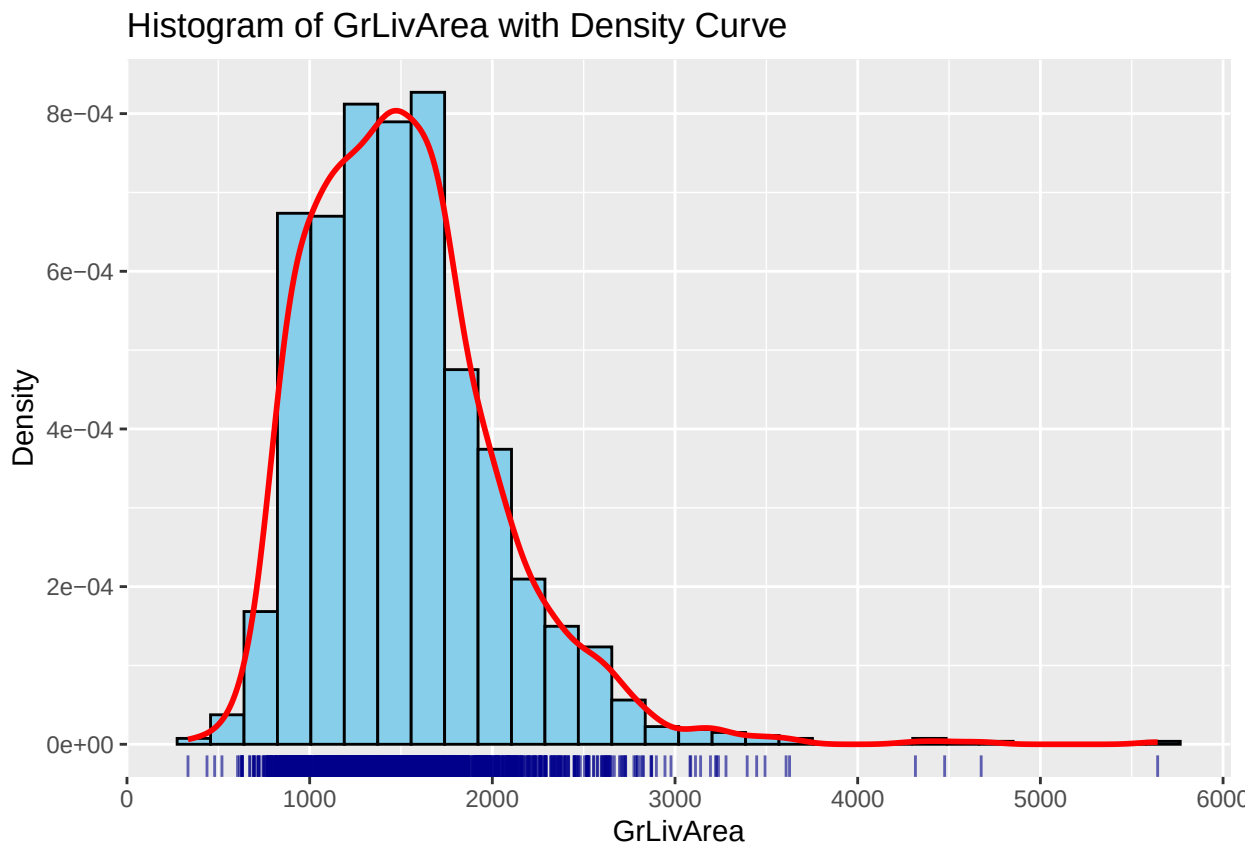

```
#Histogram of WoodDeckSF, including density line
ggplot(data, aes(x = WoodDeckSF)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red", size = 1) + geom_rug(sides = "b", color = "darkblue", alpha = 0.6)+
  labs(title = "Histogram of WoodDeckSF with Density Curve",
       x = "WoodDeckSF", y = "Density")
```




```
#Histogram of GarageArea, including density line
ggplot(data, aes(x = GarageArea)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red", size = 1) + geom_rug(sides = "b", color = "darkblue", alpha = 0.6)+
  labs(title = "Histogram of GarageArea with Density Curve",
       x = "GarageArea", y = "Density")
```



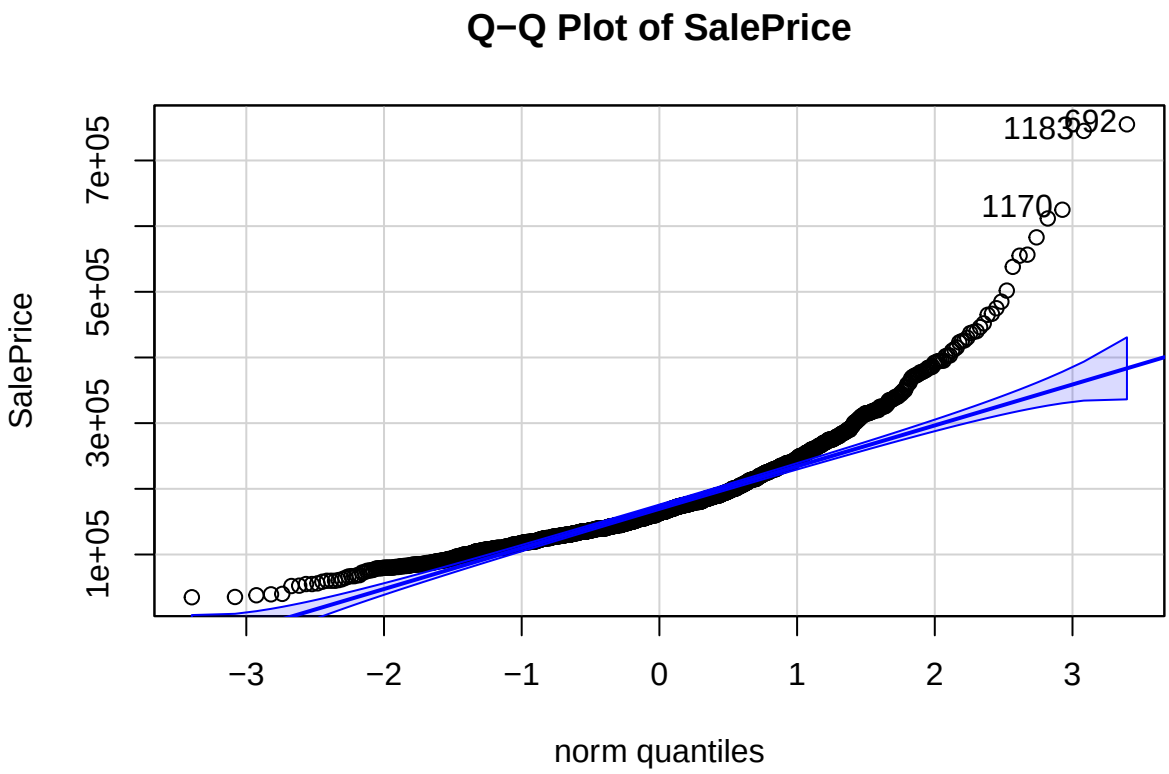

```
#Histogram of GrLivArea, including density line
ggplot(data, aes(x = GrLivArea)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "skyblue", color = "black") +
  geom_density(color = "red", size = 1) + geom_rug(sides = "b", color = "darkblue", alpha = 0.6)+
  labs(title = "Histogram of GrLivArea with Density Curve",
       x = "GrLivArea", y = "Density")
```

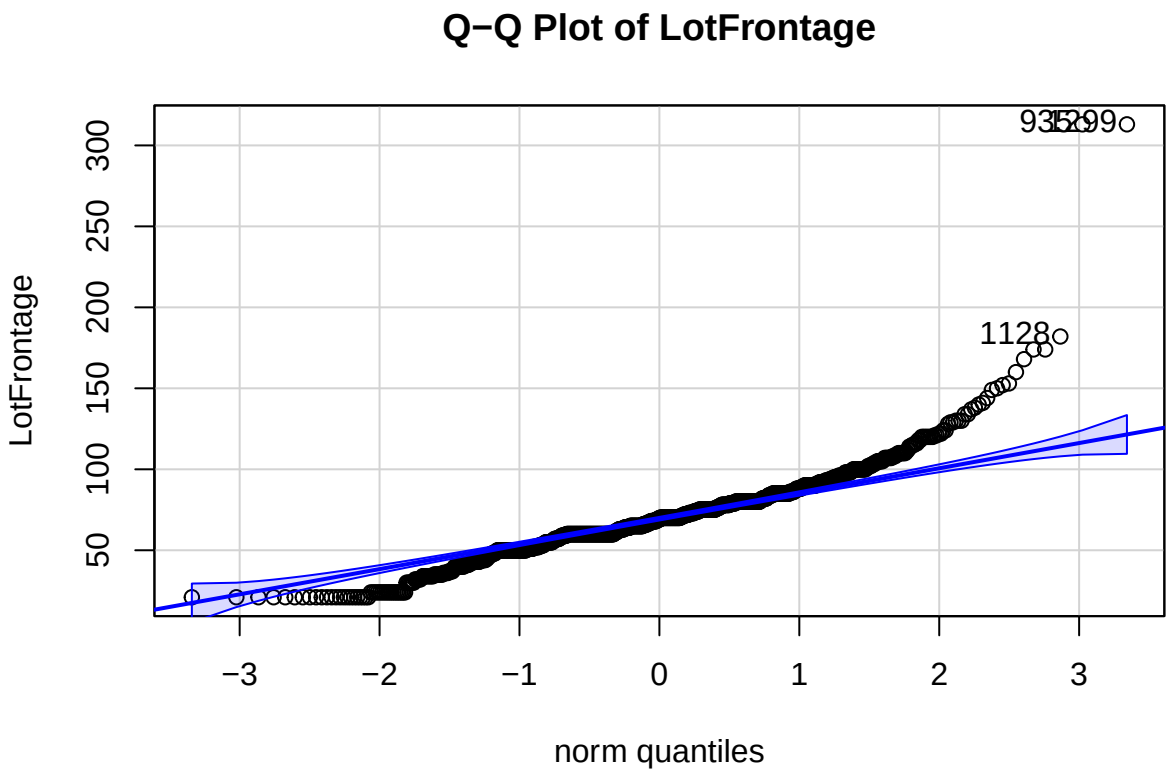


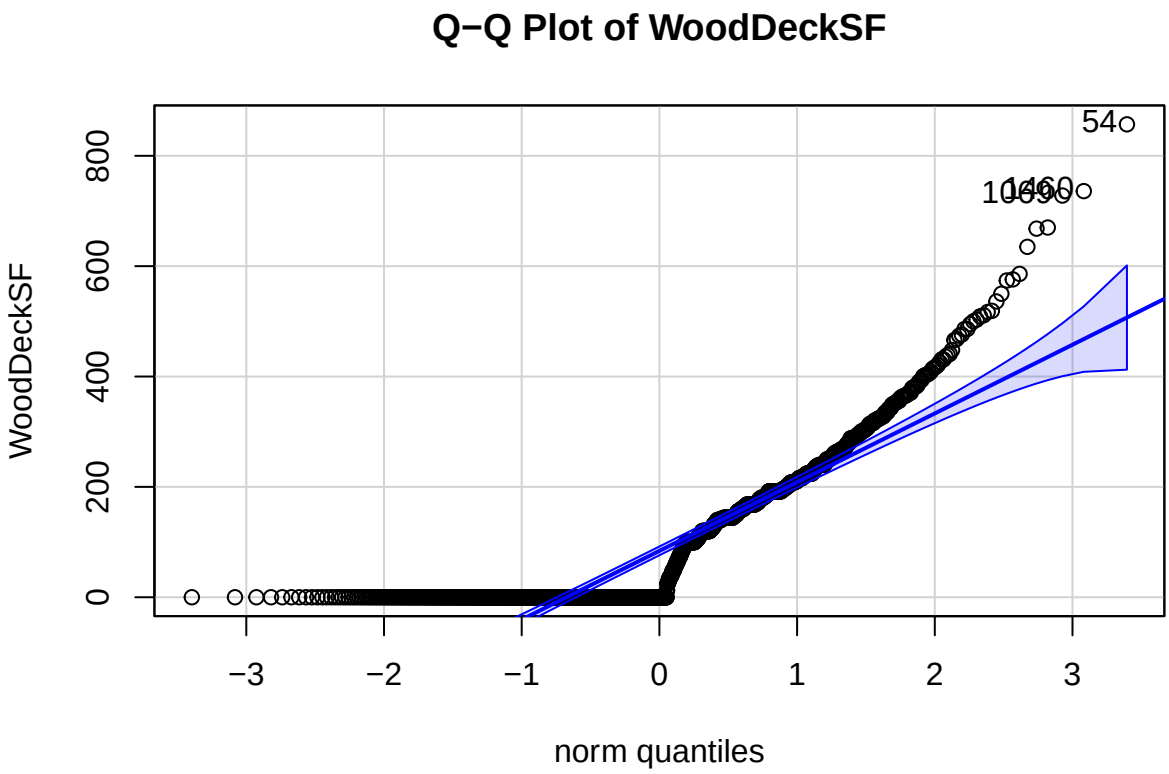

```
#qqplots for all variables
# List of variables to create Q-Q plots for
library(car)
```

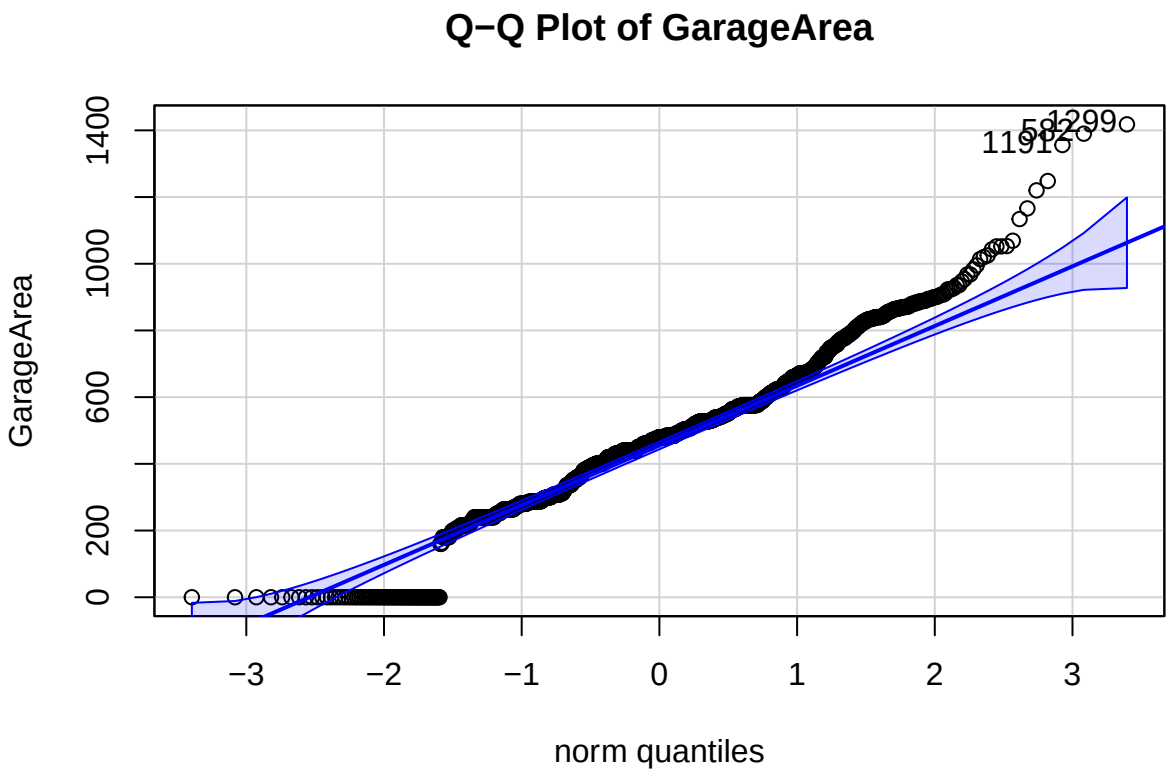
```
##      carData

##
##     'car'

## The following object is masked from 'package:psych':
##
##     logit
```

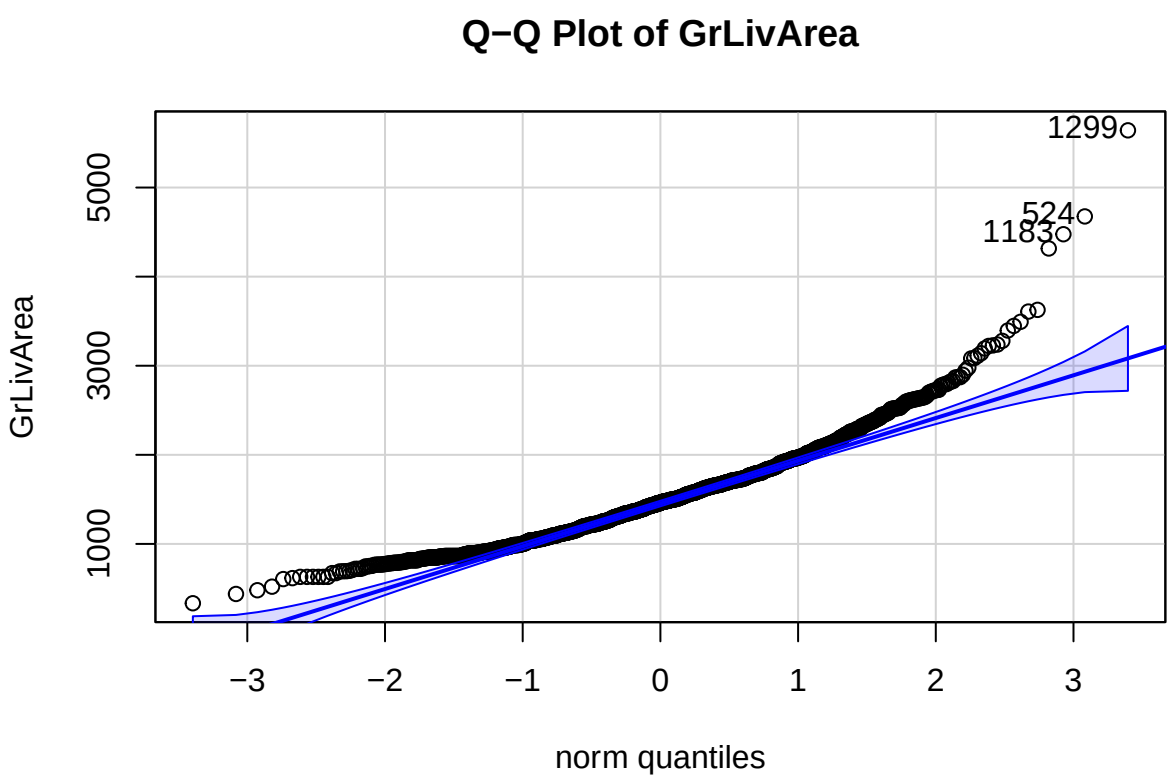
```
variables <- c("SalePrice", "LotFrontage", "WoodDeckSF", "GarageArea", "GrLivArea")

# Loop to generate Q-Q plots for each variable
for (var in variables) {
  # Generate Q-Q plot with car::qqPlot
  qqPlot(data[[var]],
         main = paste("Q-Q Plot of", var),
         ylab = var,
         id = list(n = 3))  # Label the top 3 extreme points
}
```


Q−Q Plot of SalePrice
SalePrice
1e+05
3e+05
5e+05
7e+05
1183
692
1170
−3
−2
−1
0
1
2
3
norm quantiles

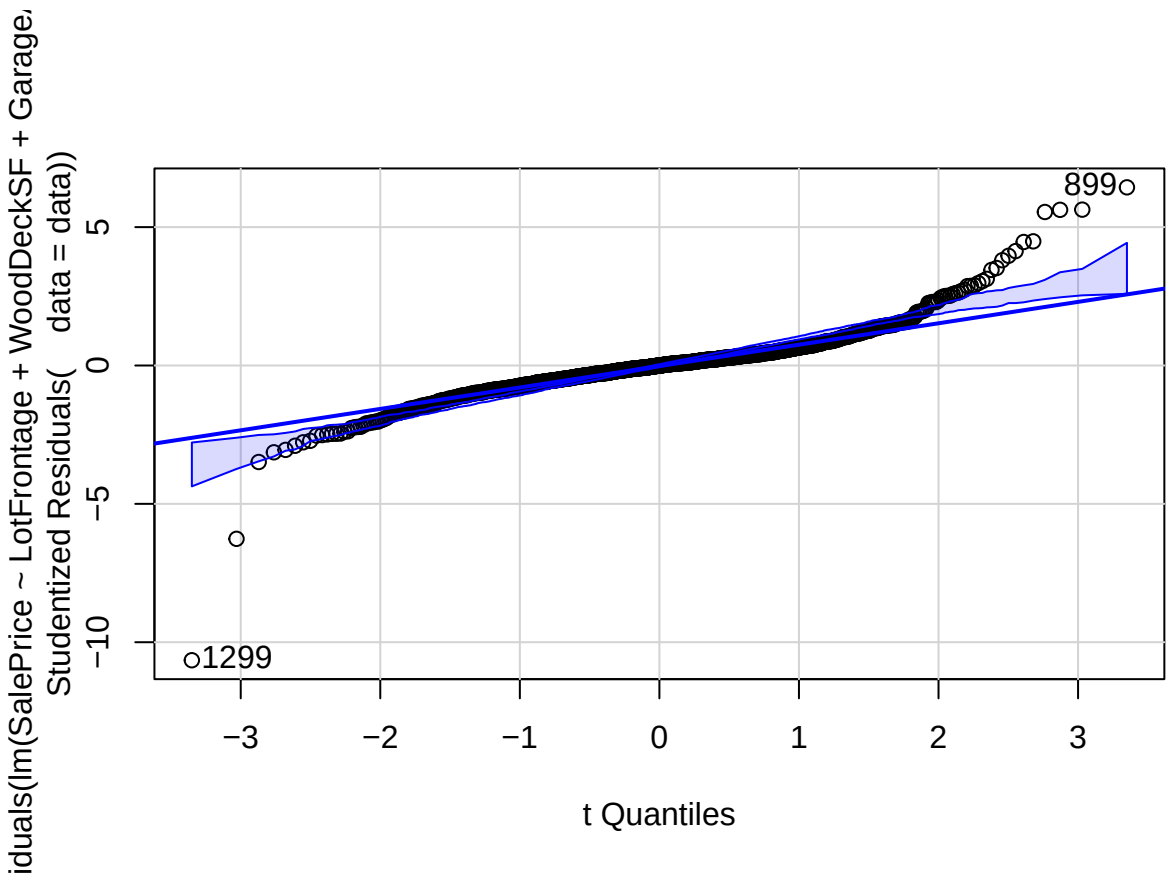

Q−Q Plot of LotFrontage
LotFrontage
norm quantiles
50
100
150
200
250
300
−3
−2
−1
0
1
2
3
1128


Q−Q Plot of WoodDeckSF
WoodDeckSF
0
200
400
600
800
54
1460
−3
−2
−1
0
1
2
3
norm quantiles


Q−Q Plot of GarageArea
GarageArea
0
200
600
1000
1400
1299
582
1191
−3
−2
−1
0
1
2
3
norm quantiles

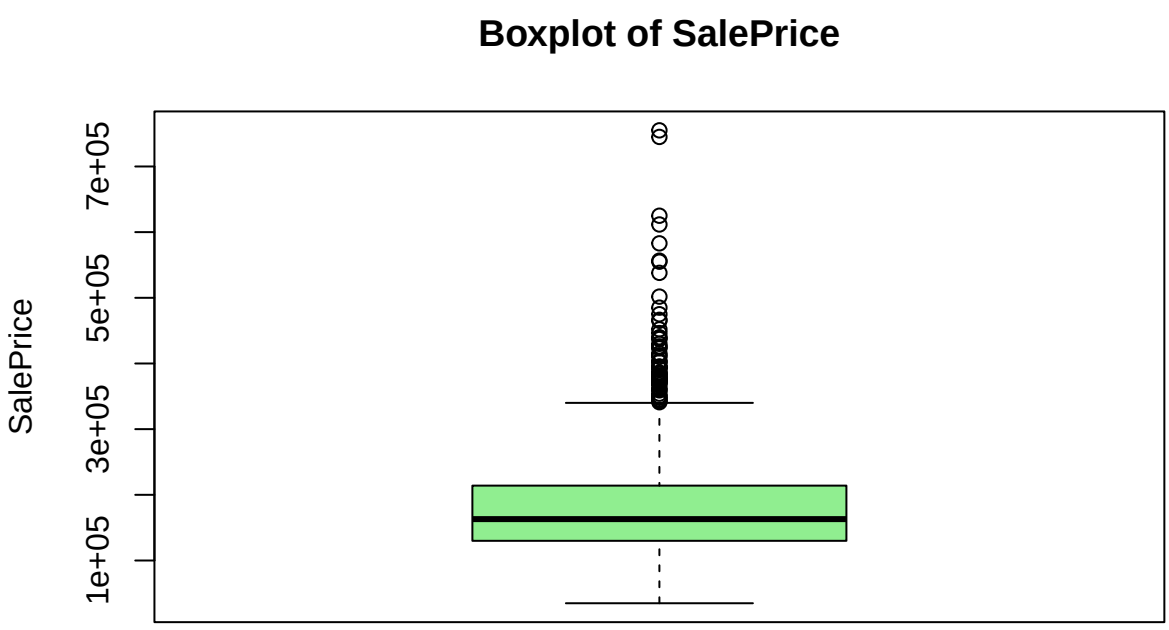


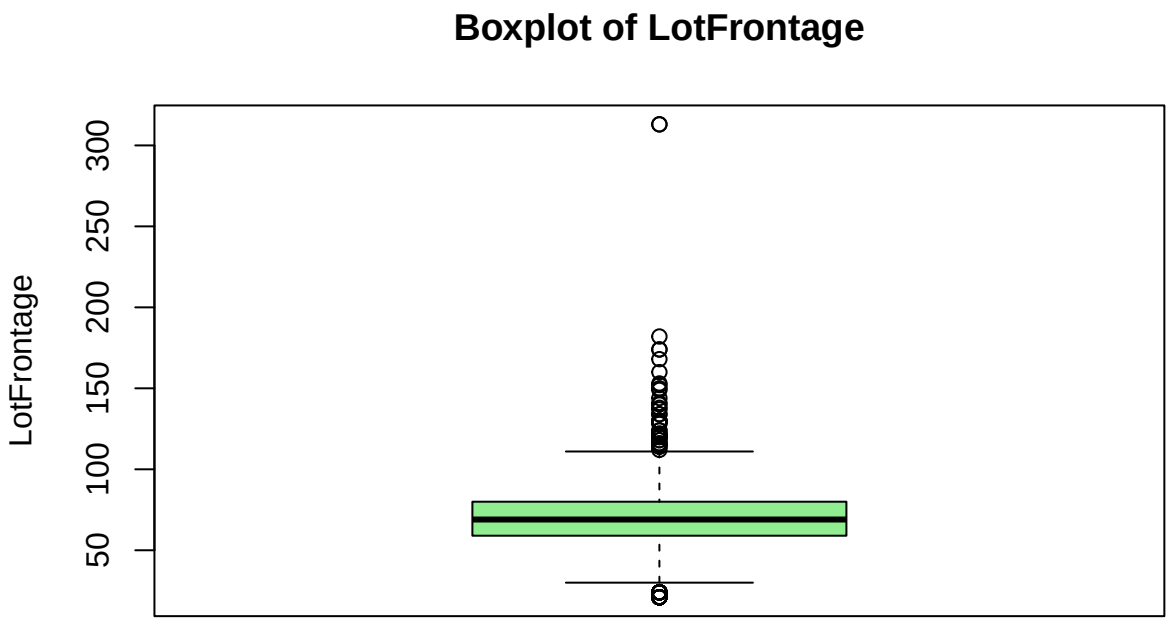

```
qqPlot(lm(SalePrice ~ LotFrontage+WoodDeckSF+GarageArea+GrLivArea, data=data),
envelope=.99)
```

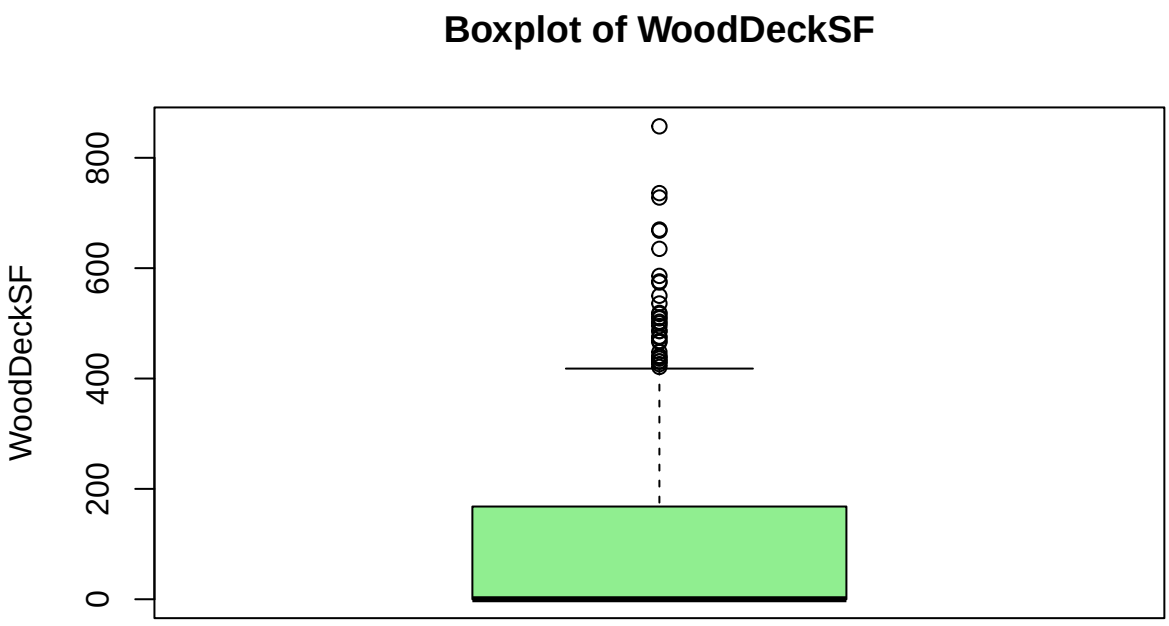


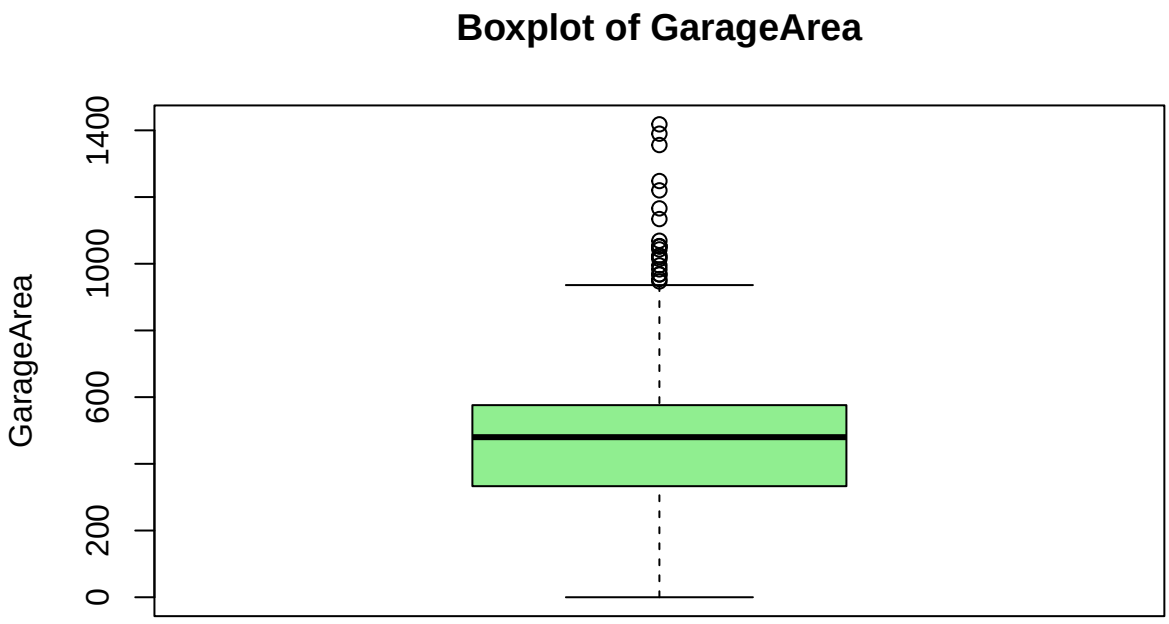

```
## [1]  899 1299
```

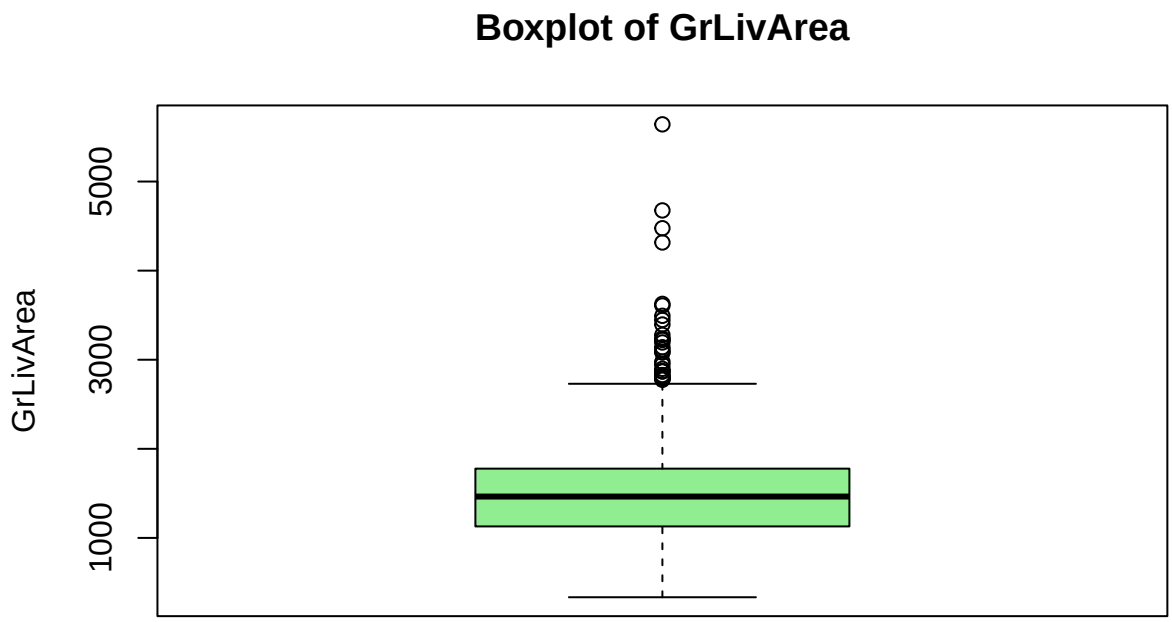
```
#boxplots for all variables
for (var in variables) {
  boxplot(data[[var]], main = paste("Boxplot of", var),
          ylab = var, col = "lightgreen")
}
```


Boxplot of SalePrice
SalePrice
1e+05
3e+05
5e+05
7e+05

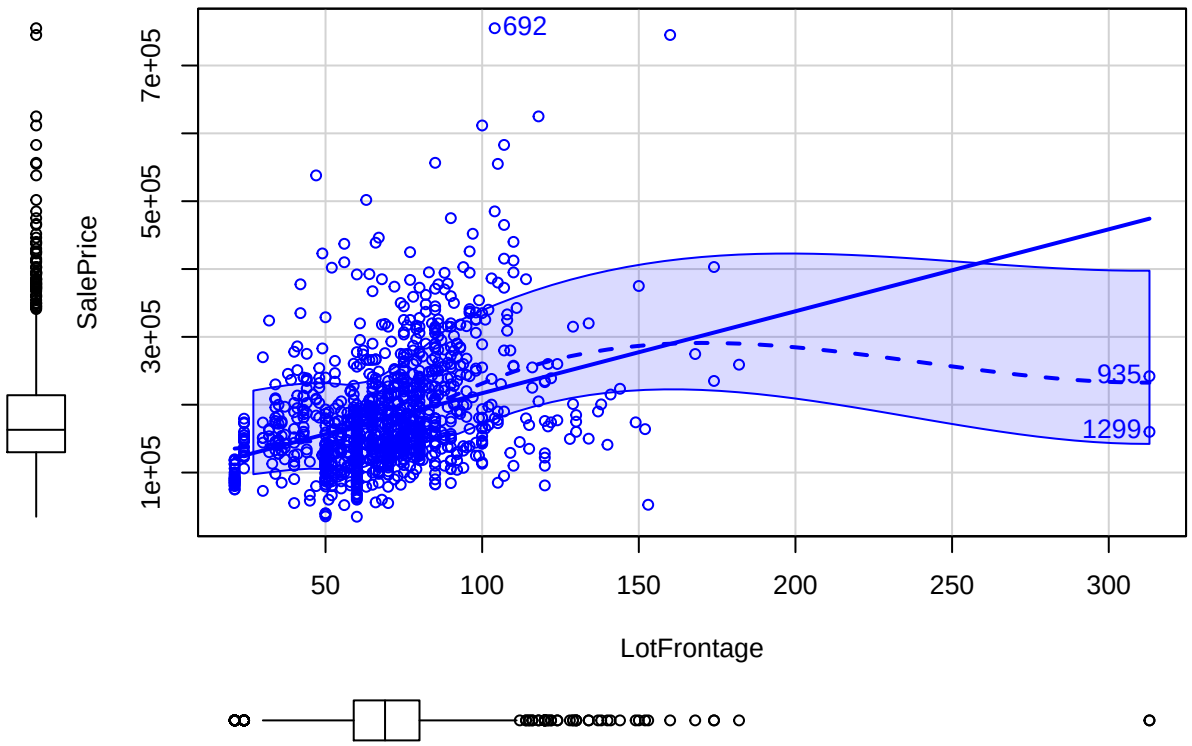

Boxplot of LotFrontage
LotFrontage
50
100
150
200
250
300


Boxplot of WoodDeckSF
WoodDeckSF
0
200
400
600
800


Boxplot of GarageArea
GarageArea
0
200
600
1000
1400

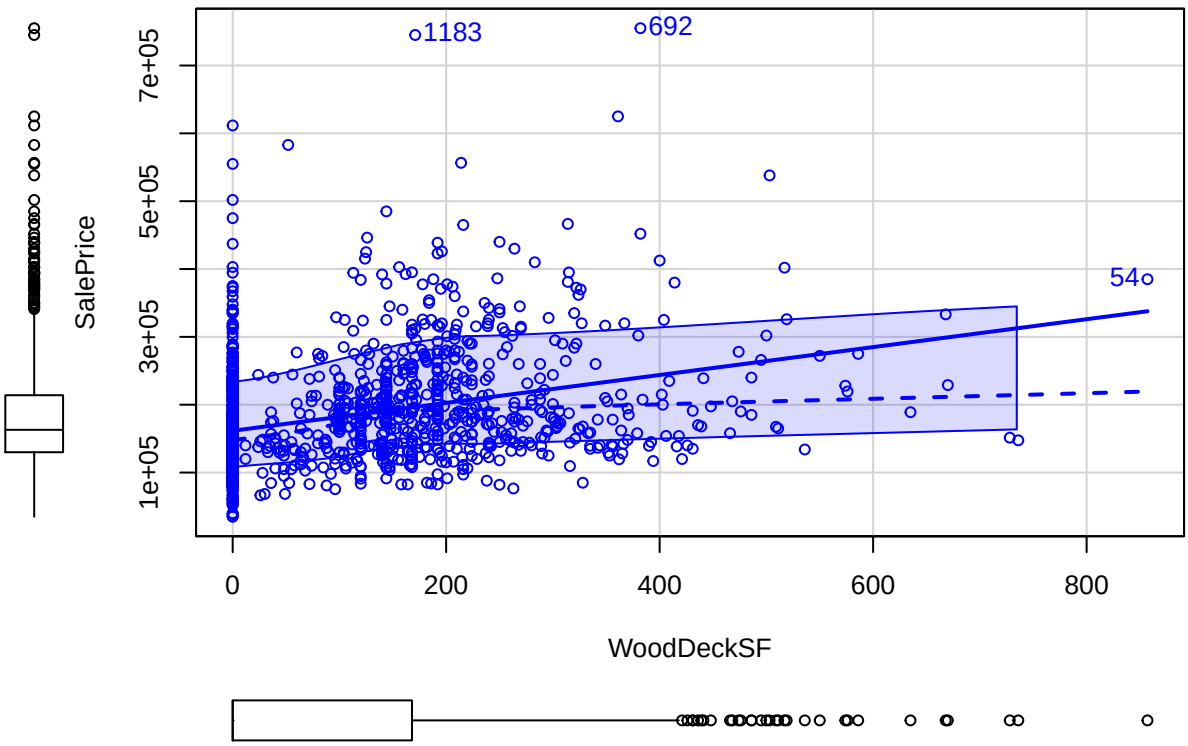



```
#scatterplots for all variables
scatterplot(SalePrice ~ LotFrontage, data = data, lwd = 3, id = list(n = 3))
```



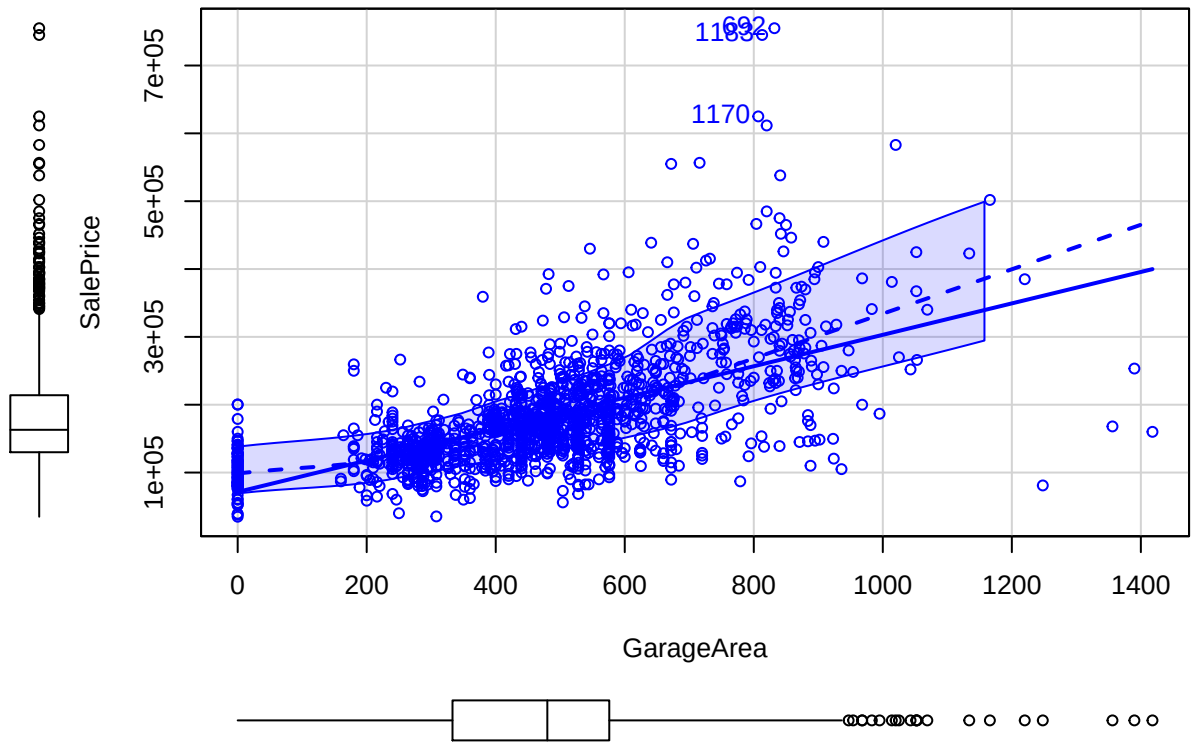

```
##  692  935 1299
##  574  774 1070
```

```
scatterplot(SalePrice ~ WoodDeckSF, data = data, lwd = 3, id = list(n = 3))
```



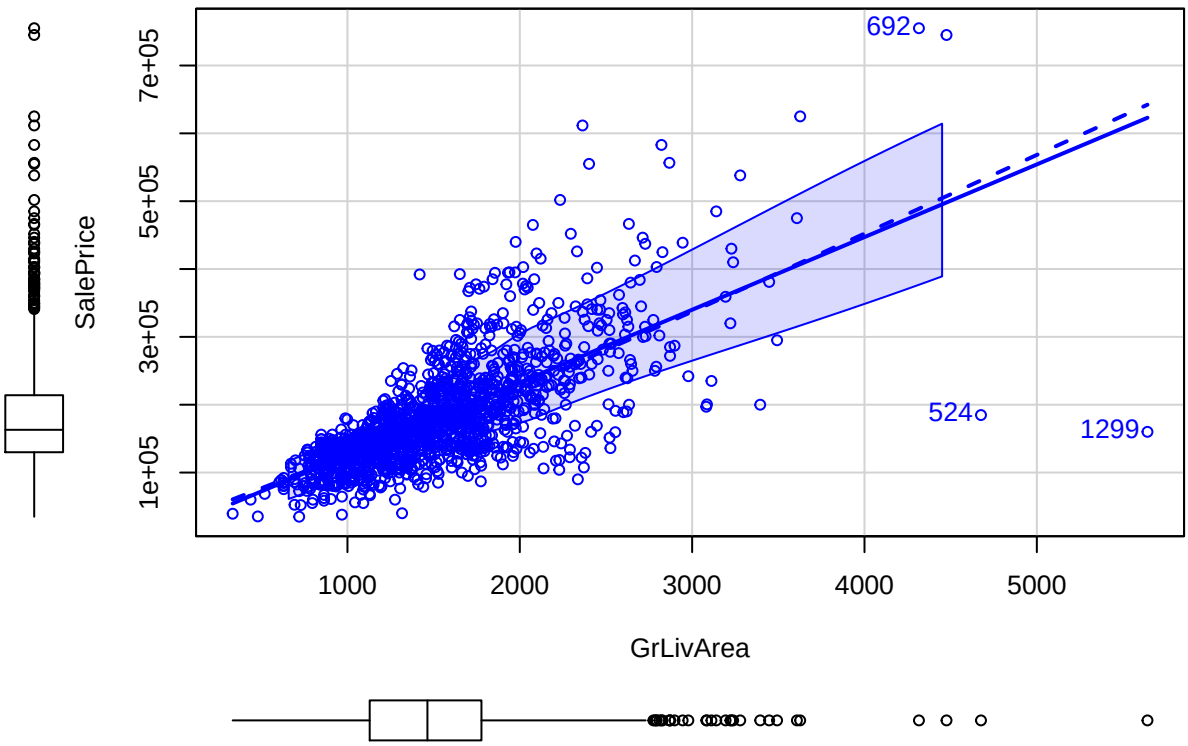

```
## [1]   54  692 1183
```

```
scatterplot(SalePrice ~ GarageArea, data = data, lwd = 3, id = list(n = 3))
```



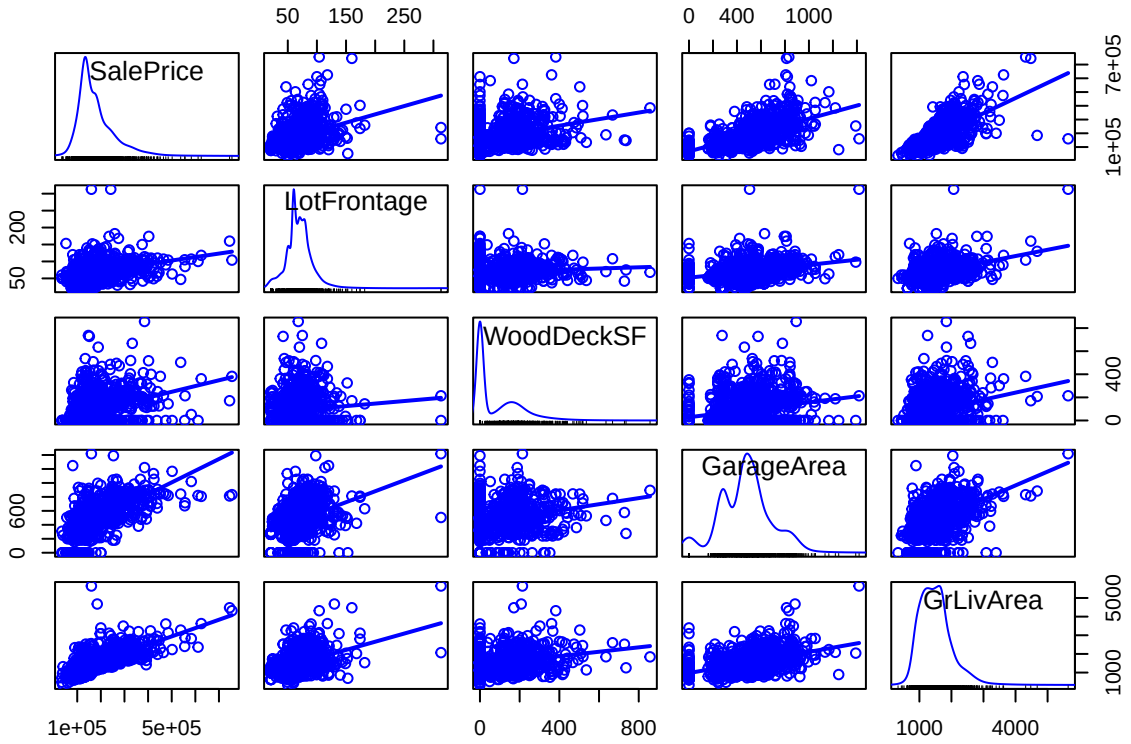

```
## [1]  692 1170 1183
```

```
scatterplot(SalePrice ~ GrLivArea, data = data, lwd = 3, id = list(n = 3))
```

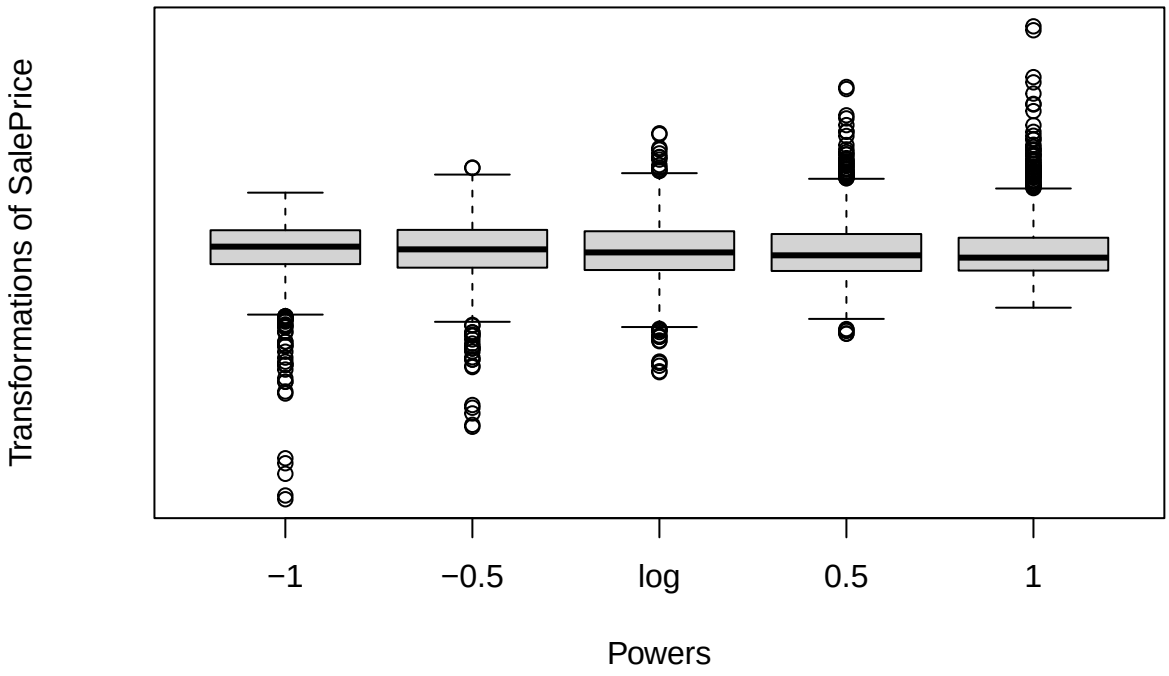


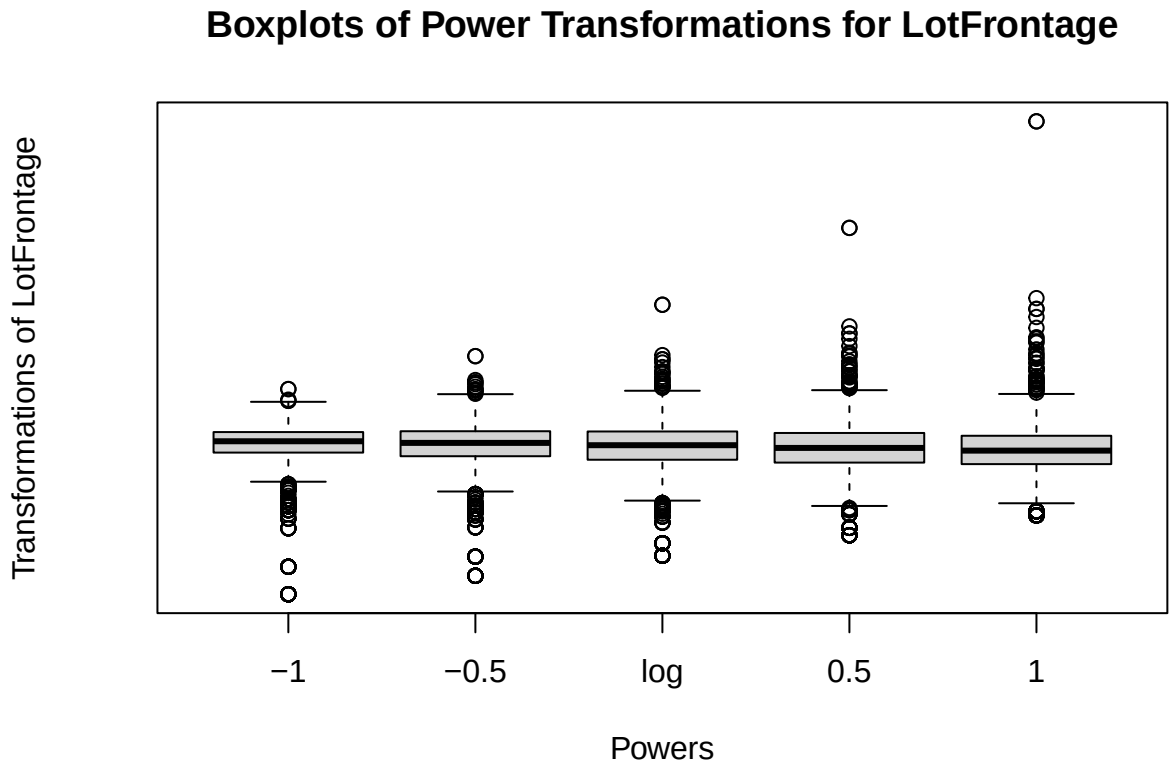

```
## [1]  524  692 1299
```

```
# Scatterplot matrix with specified variables and options
scatterplotMatrix(~ SalePrice + LotFrontage + WoodDeckSF + GarageArea + GrLivArea,
                  data = data,
                  smooth = FALSE,
                  ellipse = list(levels = 0.5))
```

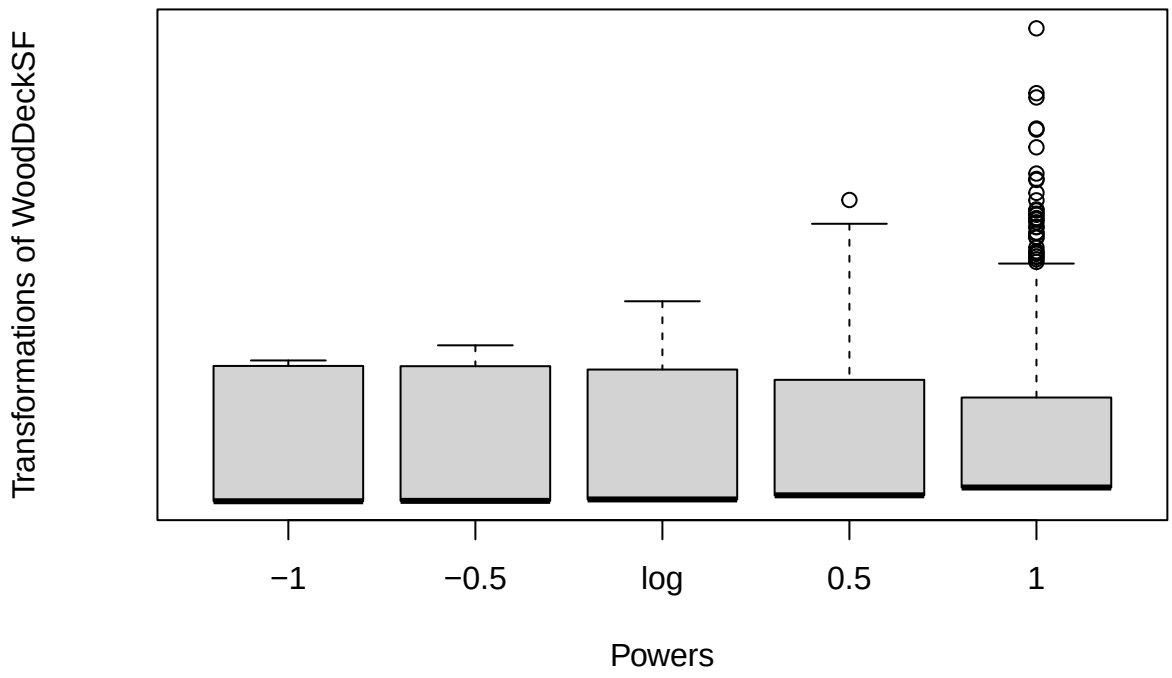



```
#linearity test
for (var in variables) {
  formula <- as.formula(paste("~", var))
  symbox(formula, data = data, main = paste("Boxplots of Power Transformations for", var))
}
```

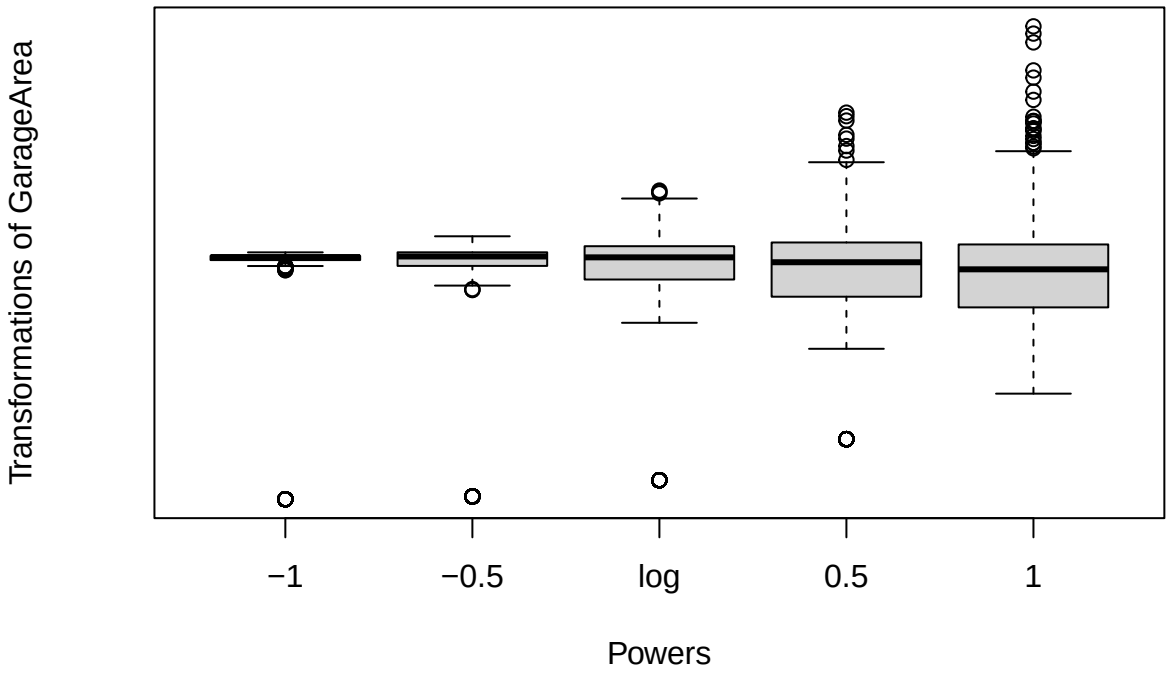

Boxplots of Power Transformations for SalePrice
Transformations of SalePrice
−1
−0.5
log
0.5
1
Powers

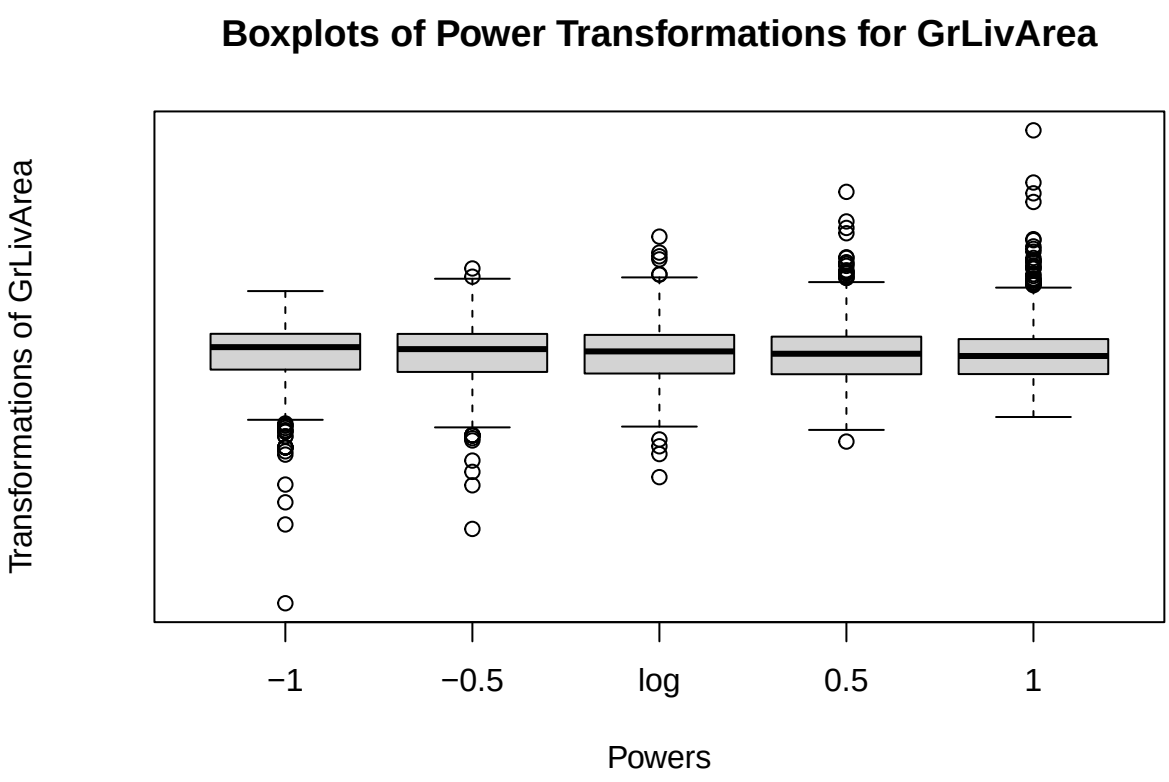



```
## Warning in symbox.default(as.vector(mf[[1]]), ylab = ylab, ...): start set to
## 8.57
```




```
## Warning in symbox.default(as.vector(mf[[1]]), ylab = ylab, ...): start set to
## 14.18
```

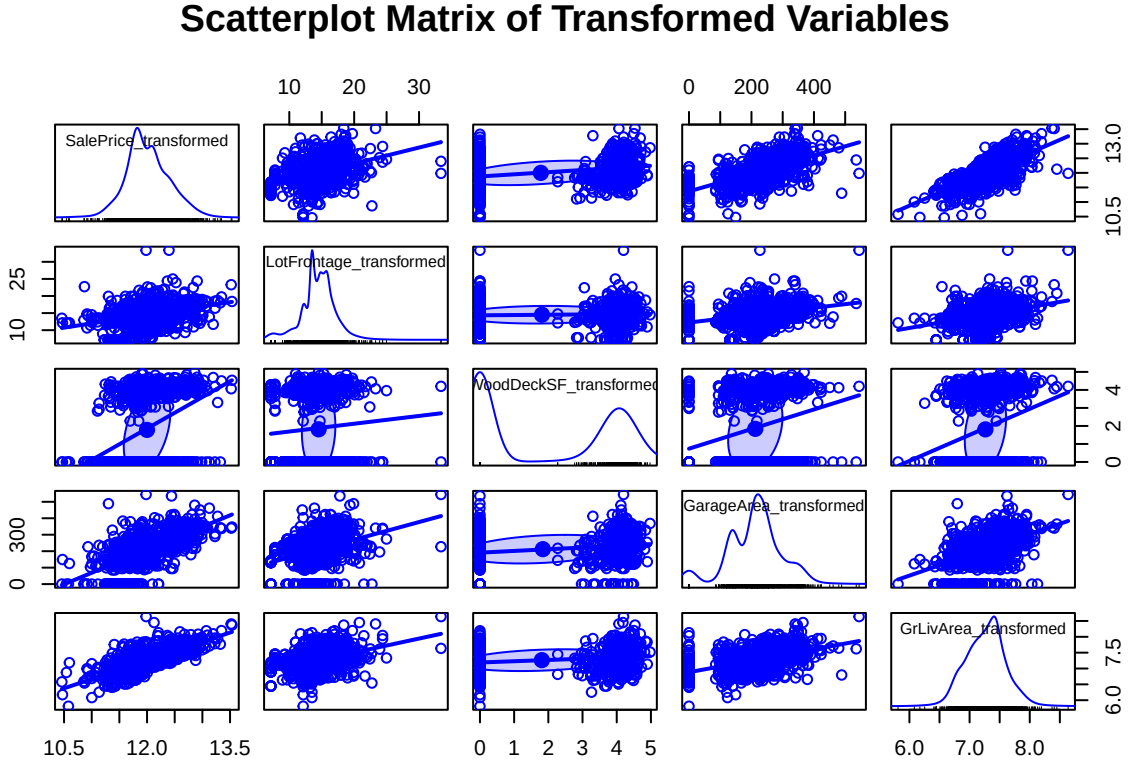

Boxplots of Power Transformations for GarageArea
Transformations of GarageArea
−1
−0.5
log
0.5
1
Powers




```
# Apply powerTransform to multiple variables
#variable wooddecksf and GarageArea contains negative values so we need to transform it by adding 1 to
WoodDeckSF1 = data$WoodDeckSF + 1
GarageArea1 = data$GarageArea + 1
# Apply powerTransform to multiple variables
a3 <- powerTransform(cbind(SalePrice, LotFrontage, WoodDeckSF1, GarageArea1, GrLivArea) ~ 1, data = data

# View the summary of the transformations
summary(a3)
```

```
## bcPower Transformations to Multinormality
##              Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## SalePrice     -0.0303        0.00      -0.1042       0.0436
## LotFrontage    0.4226        0.50       0.3323       0.5130
## WoodDeckSF1   -0.0957       -0.10      -0.1320      -0.0594
## GarageArea1    0.8452        0.85       0.8001       0.8902
## GrLivArea      0.0065        0.00      -0.1012       0.1143
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                   LRT df       pval
## LR test, lambda = (0 0 0 0 0) 2494.766  5 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                   LRT df       pval
## LR test, lambda = (1 1 1 1 1) 4587.745  5 < 2.22e-16
```

```r
# Transform selected variables using Box-Cox with coefficients from a3
transformeddata <- as.data.frame(bcPower(
  with(data, cbind(SalePrice, LotFrontage, WoodDeckSF1, GarageArea1, GrLivArea)),
  coef(a3, round = TRUE)
))

# Rename columns in transformeddata for clarity (optional)
colnames(transformeddata) <- c("SalePrice_transformed", "LotFrontage_transformed",
                               "WoodDeckSF_transformed", "GarageArea_transformed",
                               "GrLivArea_transformed")

# Plot scatterplot matrix of transformed variables
scatterplotMatrix(~ SalePrice_transformed + LotFrontage_transformed + WoodDeckSF_transformed +
                  GarageArea_transformed + GrLivArea_transformed,
                  data = transformeddata,
                  smooth = FALSE, ellipse = list(levels = 0.5),
                  main = "Scatterplot Matrix of Transformed Variables")
```




```r
#fit data
library(broom)
# Fit linear models
model1 <- lm(SalePrice_transformed ~ LotFrontage_transformed, data = transformeddata)
model2 <- lm(SalePrice_transformed ~ WoodDeckSF_transformed, data = transformeddata)
model3 <- lm(SalePrice_transformed ~ GarageArea_transformed, data = transformeddata)
model4 <- lm(SalePrice_transformed ~ GrLivArea_transformed, data = transformeddata)
```

```r
# Summarize each model
summary_model1 <- summary(model1)
summary_model2 <- summary(model2)
summary_model3 <- summary(model3)
summary_model4 <- summary(model4)

# Store summaries in a list for easy viewing
model_summaries <- list(
  model1 = summary_model1,
  model2 = summary_model2,
  model3 = summary_model3,
  model4 = summary_model4
)

# Display the summaries
model_summaries
```

```
## $model1
##
## Call:
## lm(formula = SalePrice_transformed ~ LotFrontage_transformed,
##     data = transformeddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.60057 -0.24263 -0.03524  0.25241  1.33355
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             11.217331   0.058977  190.20   <2e-16 ***
## LotFrontage_transformed  0.055052   0.003993   13.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3867 on 1199 degrees of freedom
##   (   259    )
## Multiple R-squared:  0.1369, Adjusted R-squared:  0.1361
## F-statistic: 190.1 on 1 and 1199 DF,  p-value: < 2.2e-16
##
##
## $model2
##
## Call:
## lm(formula = SalePrice_transformed ~ WoodDeckSF_transformed,
##     data = transformeddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43494 -0.24201 -0.02119  0.21855  1.42874
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             11.89519    0.01358  876.24   <2e-16 ***
```

```
## WoodDeckSF_transformed  0.06654    0.00483   13.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3759 on 1458 degrees of freedom
## Multiple R-squared:  0.1152, Adjusted R-squared:  0.1145
## F-statistic: 189.7 on 1 and 1458 DF,  p-value: < 2.2e-16
##
##
## $model3
##
## Call:
## lm(formula = SalePrice_transformed ~ GarageArea_transformed,
##     data = transformeddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.56536 -0.15902  0.01365  0.18022  1.10551
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.138e+01  2.125e-02  535.69   <2e-16 ***
## GarageArea_transformed 3.038e-03  9.332e-05   32.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3041 on 1458 degrees of freedom
## Multiple R-squared:  0.4209, Adjusted R-squared:  0.4206
## F-statistic:  1060 on 1 and 1458 DF,  p-value: < 2.2e-16
##
##
## $model4
##
## Call:
## lm(formula = SalePrice_transformed ~ GrLivArea_transformed, data = transformeddata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35338 -0.14260  0.02864  0.16585  0.86377
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            5.66812    0.15589   36.36   <2e-16 ***
## GrLivArea_transformed  0.87454    0.02143   40.81   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.273 on 1458 degrees of freedom
## Multiple R-squared:  0.5333, Adjusted R-squared:  0.533
## F-statistic:  1666 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```
#R-squared and Adjusted R-squared:
#Model 4 has the highest R-squared (0.5333), indicating that approximately 53.33% of the variance in Sa
#Model 3 is next with an R-squared of 0.4209, indicating that GarageArea_transformed explains about 42.
```

```r
#Statistical Significance:
#All models show highly significant p-values (< 2e-16), indicating that the predictors have a statistic
#Economic Significance:
#The coefficient for GrLivArea_transformed (0.8745) suggests that for each unit increase in GrLivArea_t
#The coefficient for LotFrontage_transformed (0.0551) is smaller, indicating a weaker economic impact c
```

```r
#model4 is the best model because of low pvalues and high R^2
# Calculate confidence intervals for Model 4
confint_model4 <- confint(model4, level = 0.95)
print(confint_model4)
```

```
##                              2.5 %   97.5 %
## (Intercept)           5.3623350 5.973914
## GrLivArea_transformed 0.8325049 0.916566
```

```r
#confidence intervals are within a small range
```

```r
#perfoming boostraping for parameters with 1000 samples
library(boot)
```

```
##
##    'boot'

## The following object is masked from 'package:car':
##
##     logit

## The following object is masked from 'package:psych':
##
##     logit
```

```r
set.seed(3435)
betahat.boot = Boot(model4, R=1000)
usualEsts = summary(model4)$coef[, 1:2]
summary(betahat.boot)
```

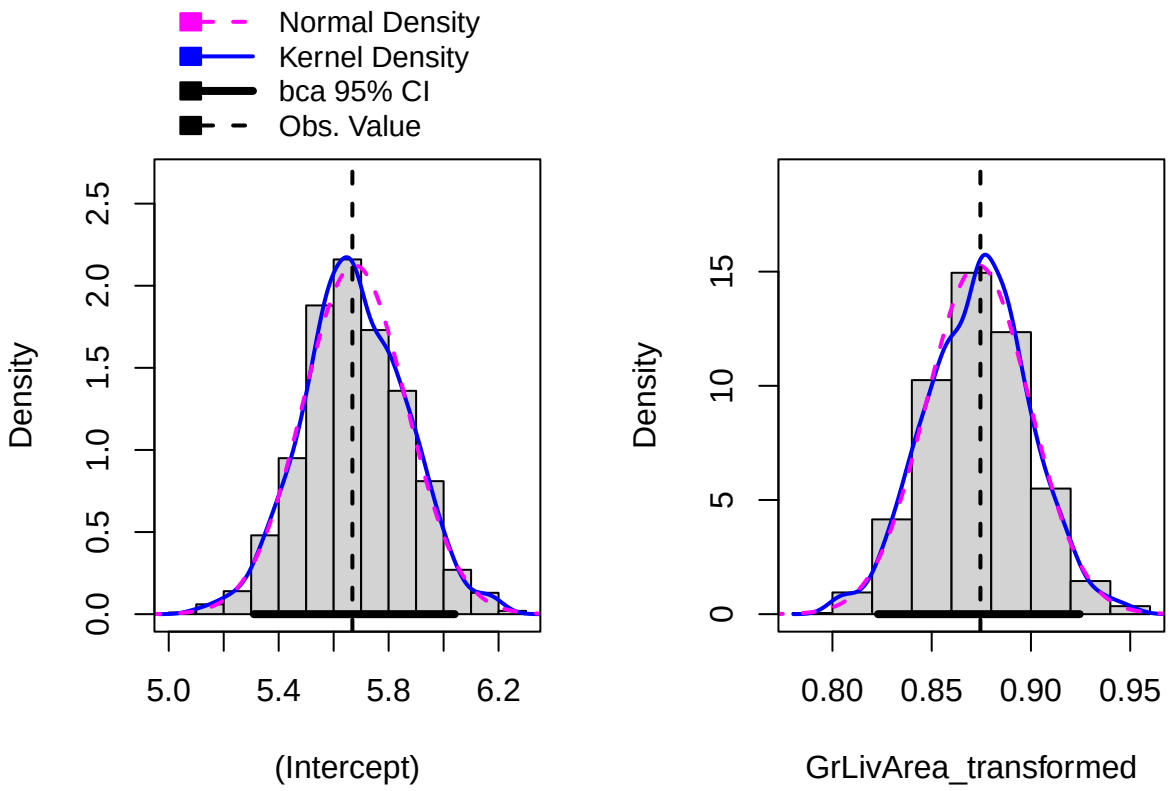
```
##
## Number of bootstrap replications R = 1000
##                         original   bootBias   bootSE bootMed
## (Intercept)            5.66812  0.0081272 0.187748 5.66987
## GrLivArea_transformed  0.87454 -0.0011562 0.026107 0.87462
```

```r
confint(betahat.boot)
```

```
## Warning in confint.boot(betahat.boot): BCa method fails for this problem.
## Using 'perc' instead

## Bootstrap percent confidence intervals
##
##                              2.5 %    97.5 %
## (Intercept)           5.3097093 6.0405171
## GrLivArea_transformed 0.8229458 0.9245575
```

```r
hist(betahat.boot)
```

```
## Warning in confint.boot(x, type = ci, level = level): BCa method fails for this
## problem.  Using 'perc' instead
```




```r
#means of parameters after boostrapping are closed to the original values
```

```r
#boostrapping for R^2
# Define the bootstrapping function
boot_fn <- function(data, indices) {
  # Resample the data using the indices
  d <- data[indices, ]

  # Fit the linear model
  model <- lm(SalePrice_transformed ~ GrLivArea_transformed, data = d)

  # Extract coefficients
  coefs <- coef(model)

  # Extract R-squared
  r_squared <- summary(model)$r.squared

  # Return both coefficients and R-squared
  return(c(coefs, r_squared))
```

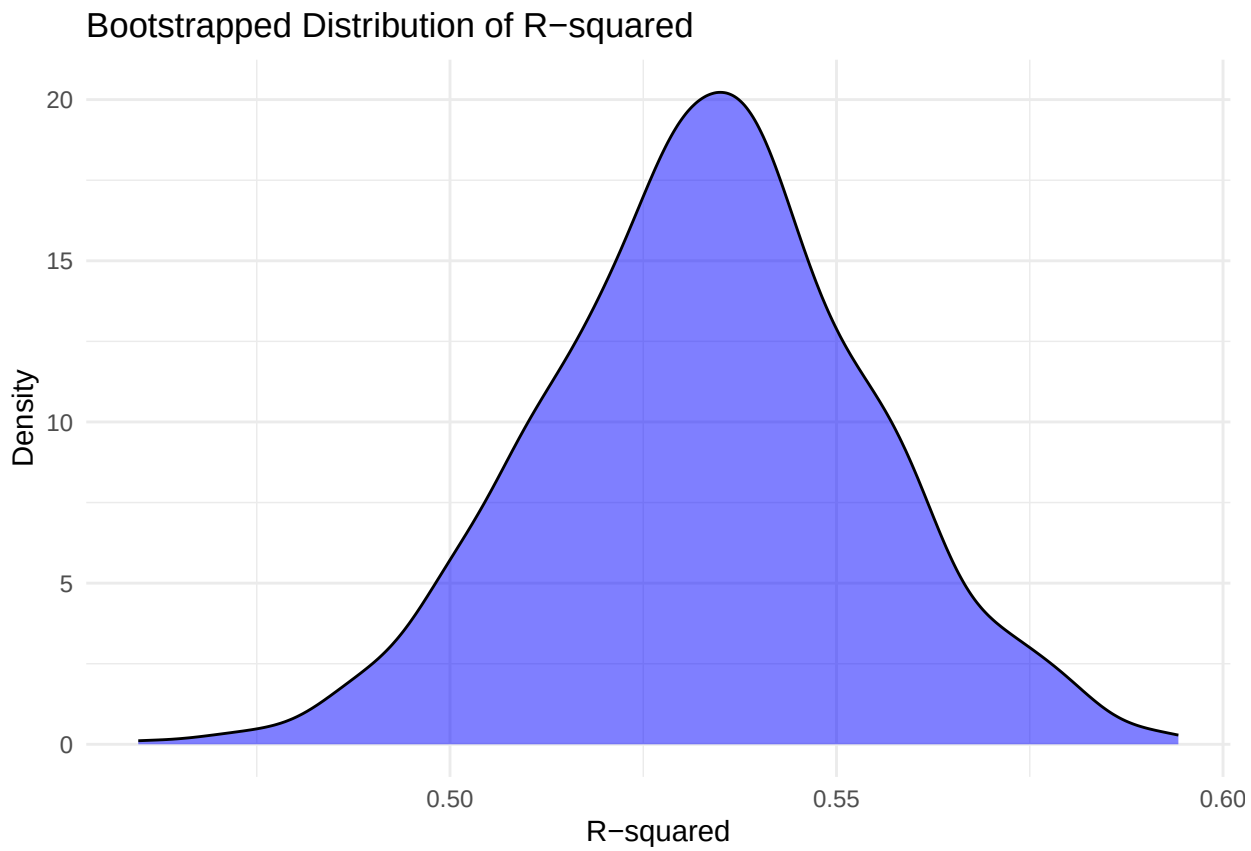
```
}
# Set parameters for bootstrapping
set.seed(3435)
R <- 1000  # Number of bootstrap samples
dd <- transformeddata  # Your dataset

# Perform bootstrapping
betahat.boot <- boot(data = dd, statistic = boot_fn, R = R)

# Check the structure of the bootstrapped results
str(betahat.boot)
```

```
## List of 11
##  $ t0       : Named num [1:3] 5.668 0.875 0.533
##   ..- attr(*, "names")= chr [1:3] "(Intercept)" "GrLivArea_transformed" ""
##  $ t        : num [1:1000, 1:3] 5.66 5.82 5.71 5.53 5.39 ...
##  $ R        : num 1000
##  $ data     :'data.frame':   1460 obs. of  5 variables:
##   ..$ SalePrice_transformed  : num [1:1460] 12.2 12.1 12.3 11.8 12.4 ...
##   ..$ LotFrontage_transformed: num [1:1460] 14.1 15.9 14.5 13.5 16.3 ...
##   ..$ WoodDeckSF_transformed : num [1:1460] 0 4.39 0 0 4.13 ...
##   ..$ GarageArea_transformed : num [1:1460] 243 210 266 278 348 ...
##   ..$ GrLivArea_transformed  : num [1:1460] 7.44 7.14 7.49 7.45 7.7 ...
##  $ seed     : int [1:626] 10403 624 -1829144649 -728749940 -1307864035 -1108863942 -266787725 -13516
##  $ statistic:function (data, indices)
##  $ sim      : chr "ordinary"
##  $ call     : language boot(data = dd, statistic = boot_fn, R = R)
##  $ stype    : chr "i"
##  $ strata   : num [1:1460] 1 1 1 1 1 1 1 1 1 1 ...
##  $ weights  : num [1:1460] 0.000685 0.000685 0.000685 0.000685 0.000685 ...
##  - attr(*, "class")= chr "boot"
##  - attr(*, "boot_type")= chr "boot"
```

```
# Check the number of columns in betahat.boot$t
ncol(betahat.boot$t)  # Ensure it returns 3
```

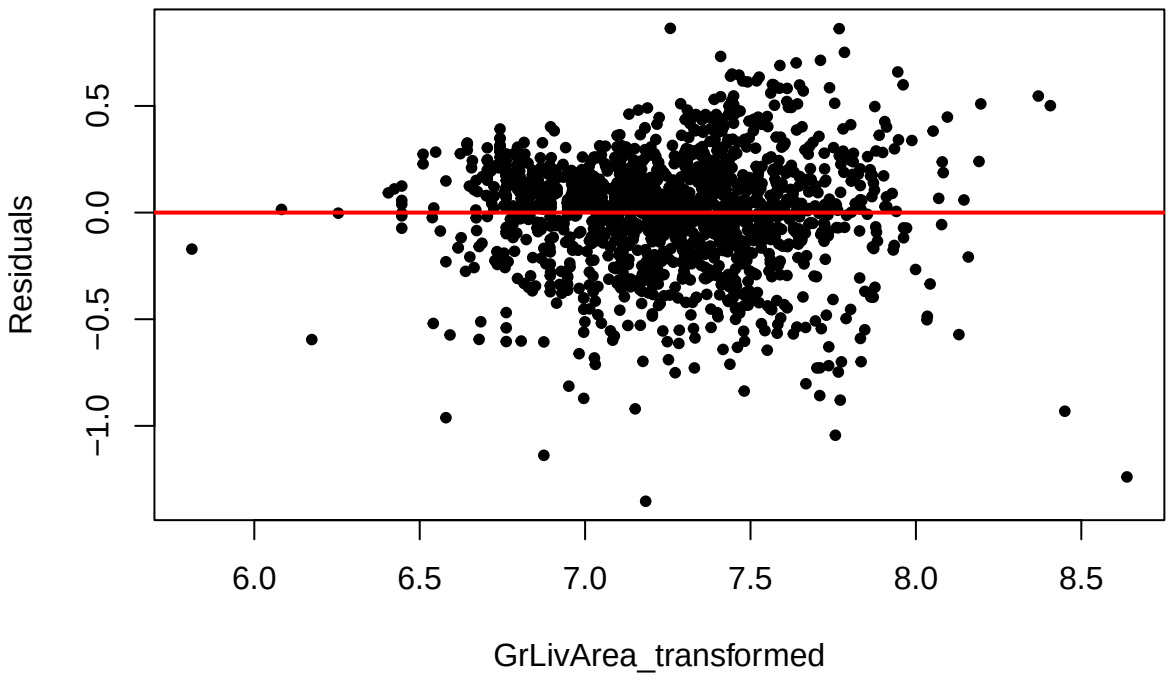
```
## [1] 3
```

```
# Extract coefficients and R-squared
boot_coefs <- betahat.boot$t[, 1:2]  # Coefficients
boot_r2 <- betahat.boot$t[, 3]         # R-squared values
# Plotting the distribution of R-squared
ggplot(data.frame(R2 = boot_r2), aes(x = R2)) +
  geom_density(fill = "blue", alpha = 0.5) +
  labs(title = "Bootstrapped Distribution of R-squared",
       x = "R-squared",
       y = "Density") +
  theme_minimal()
```

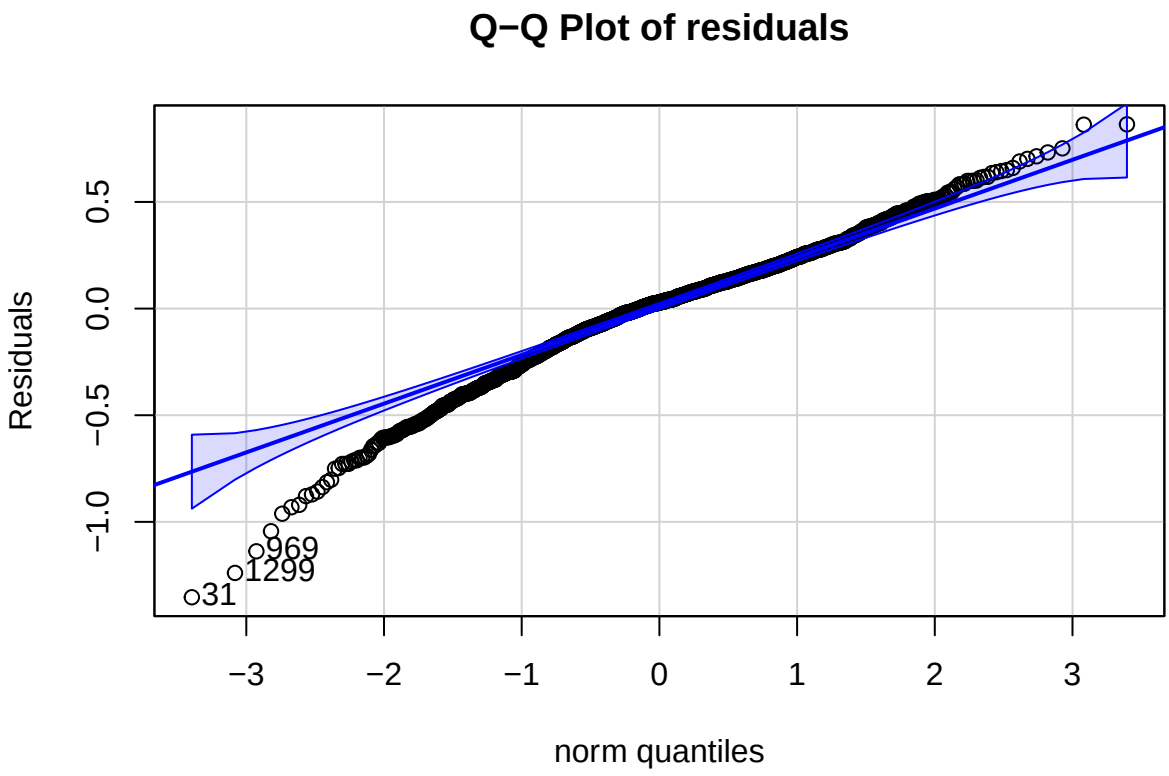



```
#the mean of boostrapping adjusted R^2 is close to the original value
```

```
#plot residuals of model4 and plot its qqplot
plot(transformeddata$GrLivArea_transformed,model4$residuals,pch=20, ylab="Residuals", xlab="GrLivArea_t
abline(h=0, lwd=2, col="red")
```

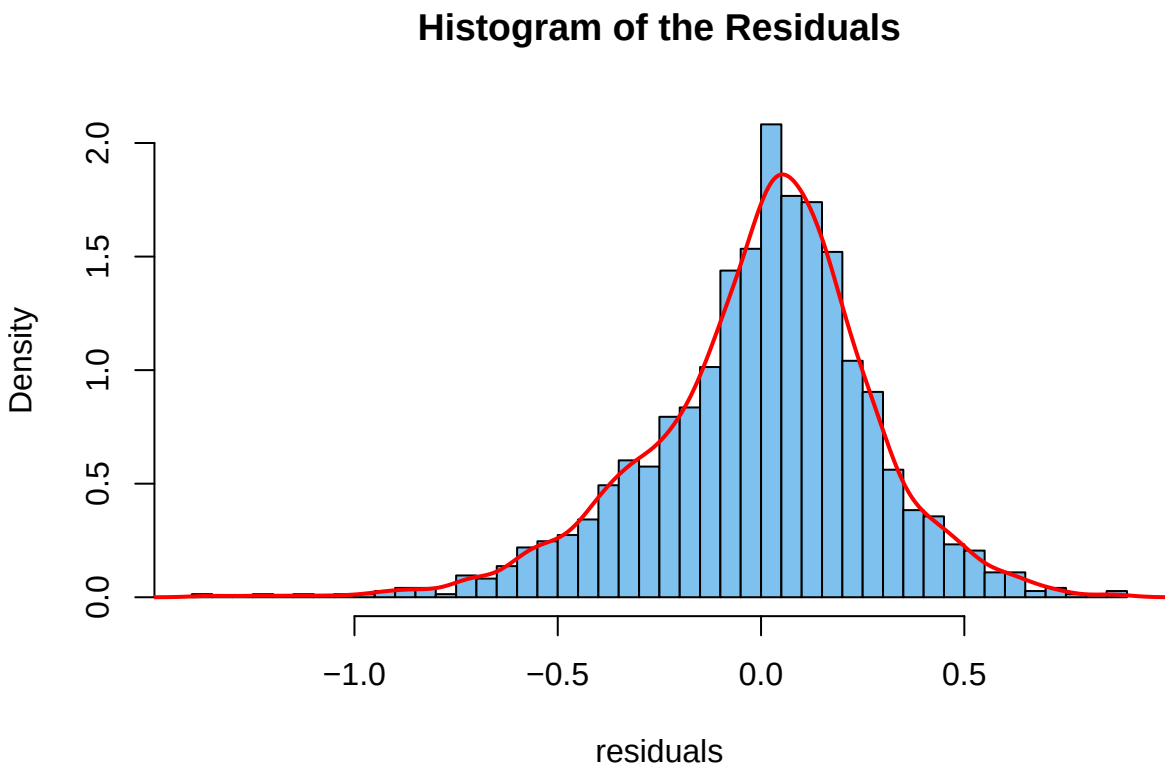



```
residuals <- resid(model4)
qqPlot(residuals,
         main = "Q-Q Plot of residuals",
         ylab = "Residuals",
         id = list(n = 3))  # Label the top 3 extreme points
```




```
## [1]   31 1299  969
```

```
hist(residuals,breaks ="FD",col="skyblue2", freq = FALSE, ylab = "Density",
main = "Histogram of the Residuals")
lines(density(residuals),lwd = 2, col ="red")
```


Histogram of the Residuals
Density
0.0
0.5
1.0
1.5
2.0
−1.0
−0.5
0.0
0.5
residuals