# Exploratory Data Analysis Report

Report by dlookr package

2019-11-18

# 1 Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

## 1.1 Information of Dataset

The dataset that generated the EDA Report is an **'data.frame'** object. It consists of **53,940 observations** and **27 variables**.

## 1.2 Information of Variables

The variable information of the data set that generated the EDA Report is shown in the following table.:

Information of Variables

| variables | types | missing_count | missing_percent | unique_count | unique_rate |
|---|---|---|---|---|---|
| carat | numeric | 0 | 0 | 273 | 0.0050612 |
| depth | numeric | 0 | 0 | 184 | 0.0034112 |
| table | numeric | 0 | 0 | 127 | 0.0023545 |
| price | numeric | 0 | 0 | 11602 | 0.2150908 |
| x | numeric | 0 | 0 | 554 | 0.0102707 |

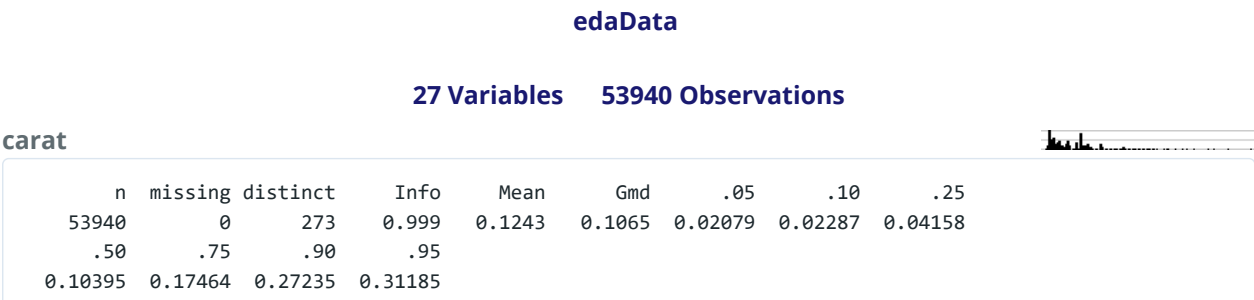| variables | types | missing_count | missing_percent | unique_count | unique_rate |
|-----------|-------|---------------|-----------------|--------------|-------------|
| y | numeric | 0 | 0 | 552 | 0.0102336 |
| z | numeric | 0 | 0 | 375 | 0.0069522 |
| cutFair | numeric | 0 | 0 | 2 | 0.0000371 |
| cutGood | numeric | 0 | 0 | 2 | 0.0000371 |
| cutIdeal | numeric | 0 | 0 | 2 | 0.0000371 |
| cutPremium | numeric | 0 | 0 | 2 | 0.0000371 |
| cutVery.Good | numeric | 0 | 0 | 2 | 0.0000371 |
| colorD | numeric | 0 | 0 | 2 | 0.0000371 |
| colorE | numeric | 0 | 0 | 2 | 0.0000371 |
| colorF | numeric | 0 | 0 | 2 | 0.0000371 |
| colorG | numeric | 0 | 0 | 2 | 0.0000371 |
| colorH | numeric | 0 | 0 | 2 | 0.0000371 |
| colorI | numeric | 0 | 0 | 2 | 0.0000371 |
| colorJ | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityI1 | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityIF | numeric | 0 | 0 | 2 | 0.0000371 |
| claritySI1 | numeric | 0 | 0 | 2 | 0.0000371 |
| claritySI2 | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityVS1 | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityVS2 | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityVVS1 | numeric | 0 | 0 | 2 | 0.0000371 |
| clarityVVS2 | numeric | 0 | 0 | 2 | 0.0000371 |

The target variable of the data is **'price'**, and the data type of the variable is **numeric**.

## 1.3 About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

## 2 Univariate Analysis

## 2.1 Descriptive Statistics

**edaData**

**27 Variables     53940 Observations**

**carat**

```
       n  missing distinct    Info     Mean     Gmd      .05      .10      .25
   53940        0      273   0.999   0.1243   0.1065  0.02079  0.02287  0.04158
     .50      .75      .90      .95
 0.10395  0.17464  0.27235  0.31185
```

| lowest : | 0.000000000 | 0.002079002 | 0.004158004 | 0.006237006 | 0.008316008 |
|---|---|---|---|---|---|
| highest: | 0.790020790 | 0.792099792 | 0.817047817 | 0.893970894 | 1.000000000 |

## depth

```
       n missing distinct    Info   Mean     Gmd    .05    .10    .25
   53940       0     184    0.999  0.5208 0.04208 0.4528 0.4722 0.5000
     .50     .75     .90     .95
  0.5222  0.5417  0.5639  0.5778
```

| lowest : | 0.00000000 | 0.02777778 | 0.21666667 | 0.22222222 | 0.25555556 |
|---|---|---|---|---|---|
| highest: | 0.81111111 | 0.83055556 | 0.85000000 | 0.97777778 | 1.00000000 |

## table

```
       n missing distinct    Info   Mean     Gmd    .05    .10    .25
   53940       0     127     0.98  0.278 0.04707 0.2115 0.2308 0.2500
     .50     .75     .90     .95
  0.2692  0.3077  0.3269  0.3462
```

| lowest : | 0.00000000 | 0.01923077 | 0.11538462 | 0.13461538 | 0.13653846 |
|---|---|---|---|---|---|
| highest: | 0.53846154 | 0.57692308 | 0.63461538 | 0.69230769 | 1.00000000 |

## price

```
       n missing distinct    Info   Mean     Gmd     .05     .10     .25
   53940       0   11602       1  0.195  0.2169 0.01179 0.01730 0.03374
     .50     .75     .90     .95
 0.11218 0.27022 0.51333 0.69098
```

| lowest : | 0.000000e+00 | 5.406282e-05 | 4.325026e-04 | 4.865654e-04 | 5.406282e-04 |
|---|---|---|---|---|---|
| highest: | 9.989187e-01 | 9.989728e-01 | 9.990809e-01 | 9.997297e-01 | 1.000000e+00 |

## x

| n | missing | distinct | Info | Mean | Gmd | .05 | .10 | .25 | .50 | .75 | .90 | .95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53940 | 0 | 554 | 1 | 0.5336 | 0.1188 | 0.3994 | 0.4060 | 0.4385 | 0.5307 | 0.6089 | 0.6806 | 0.7132 |

| lowest : | 0.0000000 | 0.3472998 | 0.3482309 | 0.3500931 | 0.3510242 |
|---|---|---|---|---|---|
| highest: | 0.9320298 | 0.9329609 | 0.9441341 | 0.9525140 | 1.0000000 |

## y

```
       n missing distinct    Info    Mean     Gmd     .05     .10     .25
   53940       0     552       1  0.09736 0.02155 0.07301 0.07402 0.08014
     .50     .75     .90     .95
 0.09694 0.11104 0.12394 0.12988
```

```
 Value       0.00  0.06  0.07  0.08  0.09  0.10  0.11  0.12  0.13  0.14  0.15  0.16
 Frequency      7    16  7450 10148  7732  8270 10610  5023  2729  1740   185    21
 Proportion 0.000 0.000 0.138 0.188 0.143 0.153 0.197 0.093 0.051 0.032 0.003 0.000

 Value       0.17  0.18  0.54  1.00
```

```
    Frequency       6      1      1      1
    Proportion 0.000 0.000 0.000 0.000
```

## z

```
        n  missing distinct    Info    Mean     Gmd      .05      .10      .25
     53940        0      375       1  0.1113  0.02485  0.08333  0.08459  0.09151
       .50      .75      .90      .95
   0.11101  0.12704  0.14214  0.14874
```

| lowest : | 0.00000000 | 0.03364780 | 0.04433962 | 0.04811321 | 0.06477987 |
|---|---|---|---|---|---|
| highest: | 0.20220126 | 0.21132075 | 0.21949686 | 0.25345912 | 1.00000000 |

### cutFair

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.087 | 1610 | 0.02985 | 0.05792 |

### cutGood

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.248 | 4906 | 0.09095 | 0.1654 |

### cutIdeal

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.72 | 21551 | 0.3995 | 0.4798 |

### cutPremium

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.571 | 13791 | 0.2557 | 0.3806 |

### cutVery.Good

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.521 | 12082 | 0.224 | 0.3476 |

### colorD

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.329 | 6775 | 0.1256 | 0.2197 |

### colorE

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.446 | 9797 | 0.1816 | 0.2973 |

### colorF

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.437 | 9542 | 0.1769 | 0.2912 |

## colorG

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.497 | 11292 | 0.2093 | 0.331 |

## colorH

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.391 | 8304 | 0.1539 | 0.2605 |

## colorI

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.271 | 5422 | 0.1005 | 0.1808 |

## colorJ

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.148 | 2808 | 0.05206 | 0.0987 |

## clarityI1

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.041 | 741 | 0.01374 | 0.0271 |

## clarityIF

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.096 | 1790 | 0.03319 | 0.06417 |

## claritySI1

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.551 | 13065 | 0.2422 | 0.3671 |

## claritySI2

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.424 | 9194 | 0.1704 | 0.2828 |

## clarityVS1

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.386 | 8171 | 0.1515 | 0.2571 |

## clarityVS2

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|
| 53940 | 0 | 2 | 0.527 | 12258 | 0.2273 | 0.3512 |

## clarityVVS1

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---------|----------|------|-----|------|-----|

| 53940 | 0 | 2 | 0.19 | 3655 | 0.06776 | 0.1263 |
|---|---|---|---|---|---|---|

**clarityVVS2**

| n | missing | distinct | Info | Sum | Mean | Gmd |
|---|---|---|---|---|---|---|
| 53940 | 0 | 2 | 0.255 | 5066 | 0.09392 | 0.1702 |

## 2.2 Normality Test of Numerical Variables

### 2.2.1 Statistics and Visualization of (Sample) Data

**[ carat ]**

normality test : Shapiro-Wilk normality test

statistic : 0.89986, p-value : 3.10829E-49

| skewness and kurtosis | | |
|---|---|---|
| type | skewness | kurtosis |
| original | 1.0523051 | 3.890211 |
| log transformation | -0.4078949 | 2.300566 |
| sqrt transformation | 0.3124253 | 2.268163 |
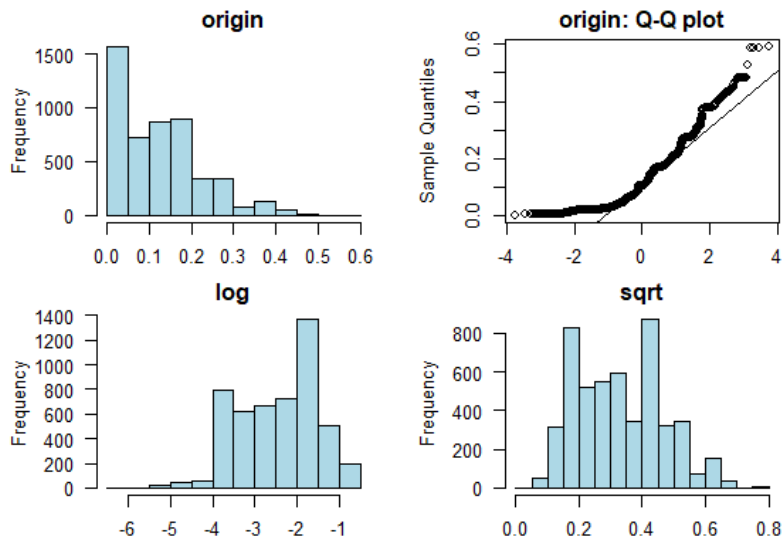


**[ depth ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.95047,  p-value : 4.01014E-38
```

| skewness and kurtosis | | |
|---|---|---|
| type | skewness | kurtosis |

| type | skewness | kurtosis |
|------|----------|----------|
| original | 0.0919192 | 7.422150 |
| log transformation | -0.5896140 | 7.179345 |
| sqrt transformation | -0.2530753 | 6.959134 |
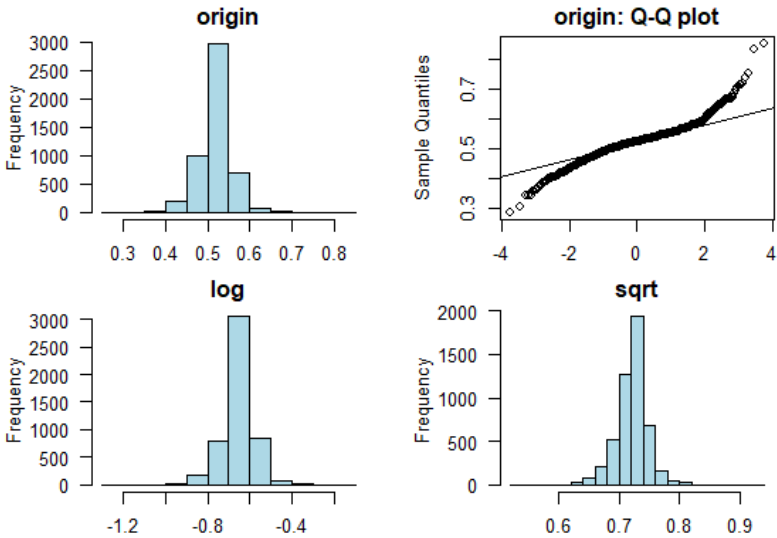


**[ table ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.96073,  p-value : 8.83796E-35
```

|  | skewness and kurtosis | |
|------|----------|----------|
| type | skewness | kurtosis |
| original | 0.6505231 | 3.926229 |
| log transformation | 0.1605538 | 3.114571 |
| sqrt transformation | 0.3980919 | 3.390631 |
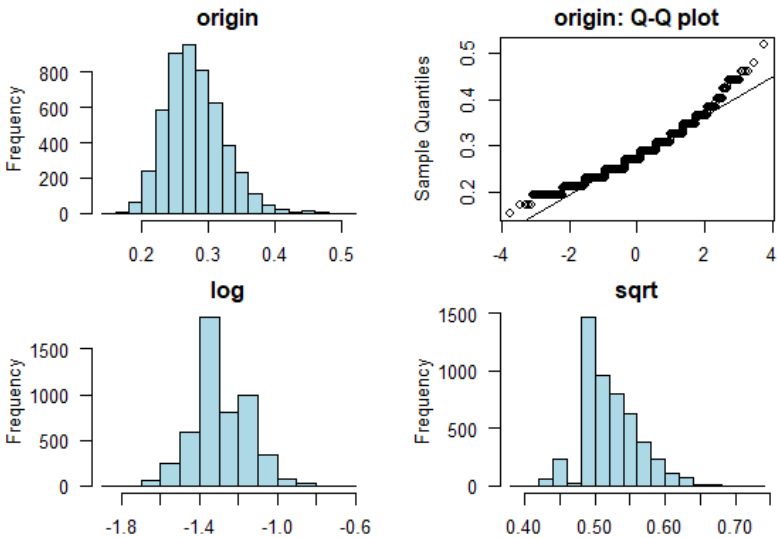
**[ x ]**

normality test : Shapiro-Wilk normality test

> statistic : 0.95716,  p-value : 5.13842E-36

skewness and kurtosis

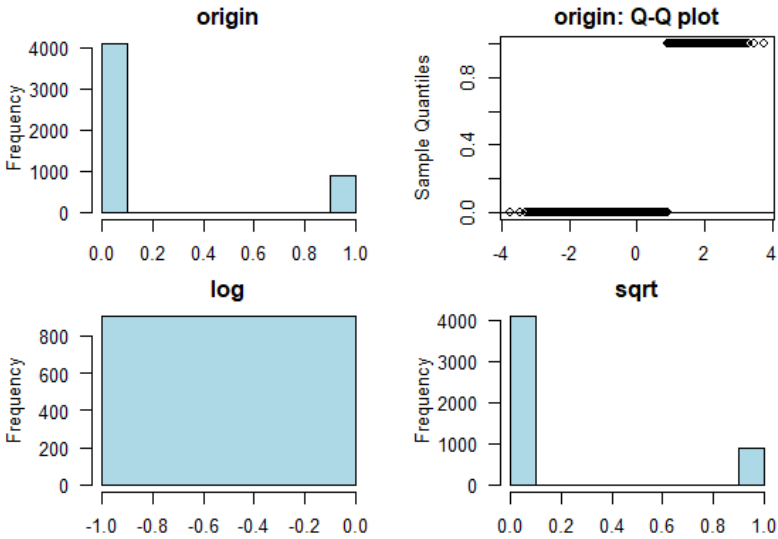| type | skewness | kurtosis |
| --- | --- | --- |
| original | 0.3243524 | 2.322840 |
| log transformation | NaN | NaN |
| sqrt transformation | -0.0052365 | 4.074092 |
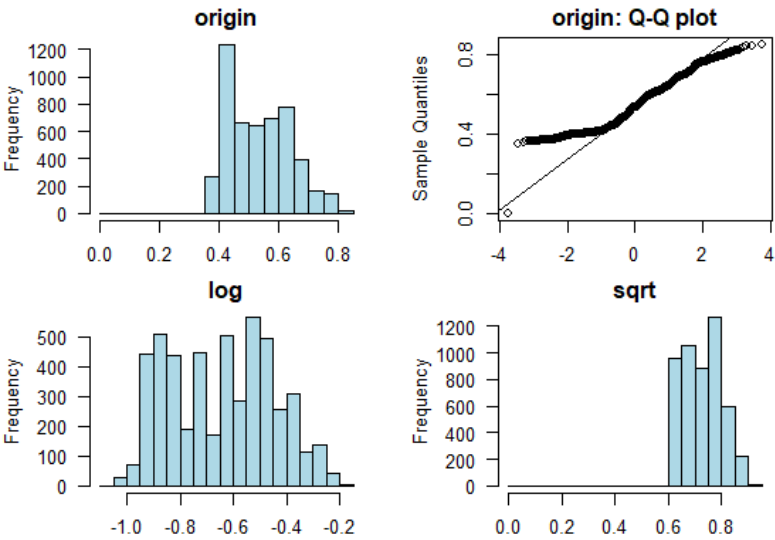


**[ y ]**

normality test : Shapiro-Wilk normality test

> statistic : 0.95644,  p-value : 2.9673E-36

skewness and kurtosis

| type | skewness | kurtosis |
| --- | --- | --- |
| original | 0.3226067 | 2.628910 |
| log transformation | NaN | NaN |
| sqrt transformation | -0.3469685 | 7.740337 |
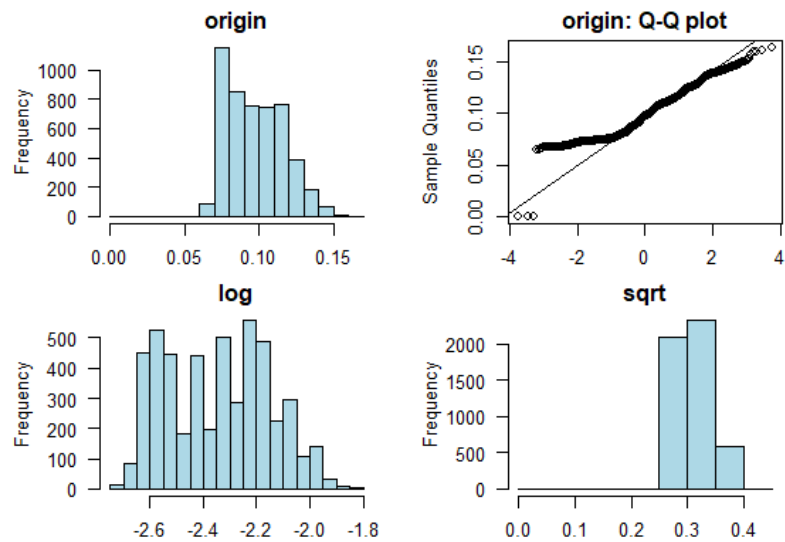
**[ z ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.95619,  p-value : 2.46294E-36
```

skewness and kurtosis

| type | skewness | kurtosis |
|---|---|---|
| original | 0.3355366 | 2.522228 |
| log transformation | NaN | NaN |
| sqrt transformation | -0.1776906 | 6.094086 |
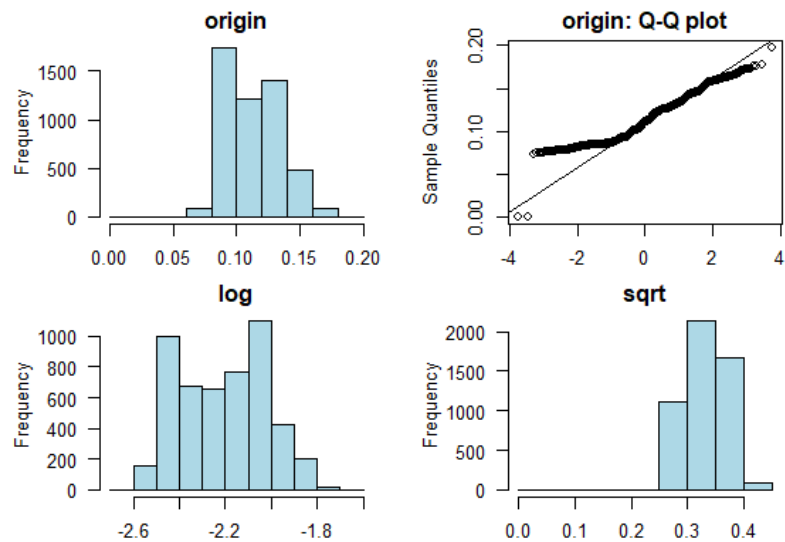


**[ cutFair ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.15612,  p-value : 7.11095E-92
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 5.592553 | 32.27665 |
| log transformation | NaN | NaN |
| sqrt transformation | 5.592553 | 32.27665 |
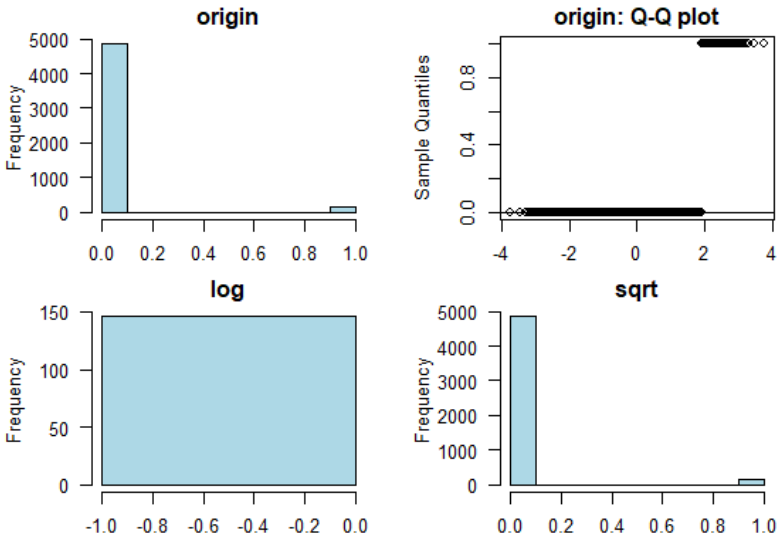


## [ cutGood ]

normality test : Shapiro-Wilk normality test

```
statistic : 0.33586,  p-value : 2.08426E-86
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 2.727224 | 8.437753 |
| log transformation | NaN | NaN |
| sqrt transformation | 2.727224 | 8.437753 |
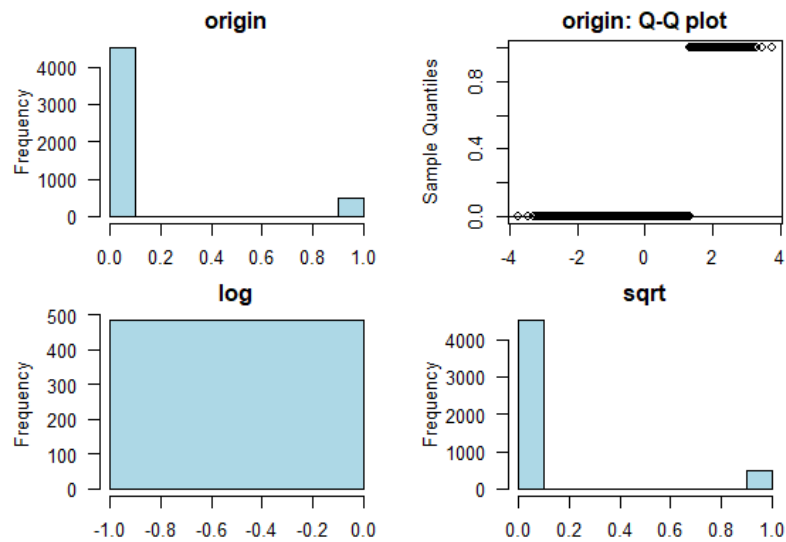
## [ cutIdeal ]

normality test : Shapiro-Wilk normality test

```
statistic : 0.62162,  p-value : 3.07496E-74
```

skewness and kurtosis

| type | skewness | kurtosis |
|---|---|---|
| original | 0.4125035 | 1.170159 |
| log transformation | NaN | NaN |
| sqrt transformation | 0.4125035 | 1.170159 |
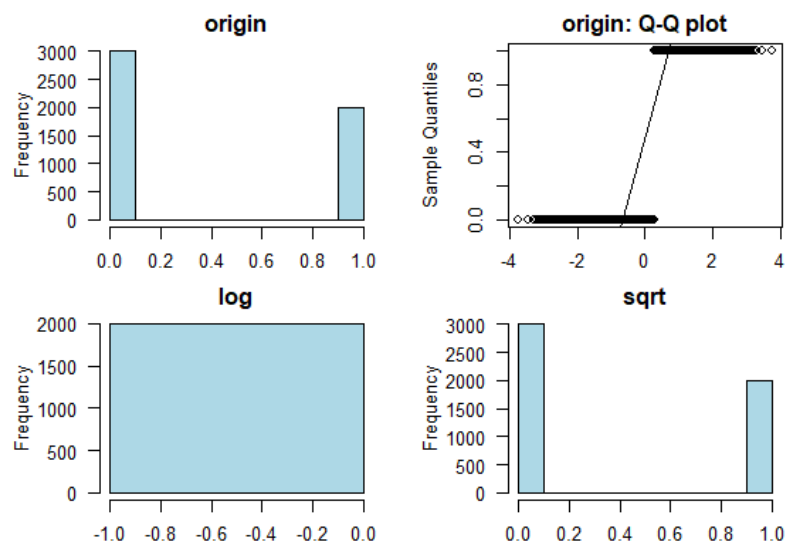


## [ cutPremium ]

normality test : Shapiro-Wilk normality test

```
statistic : 0.54796,  p-value : 5.52485E-78
```

### skewness and kurtosis

| type | skewness | kurtosis |
|---|---|---|
| original | 1.086033 | 2.179467 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.086033 | 2.179467 |
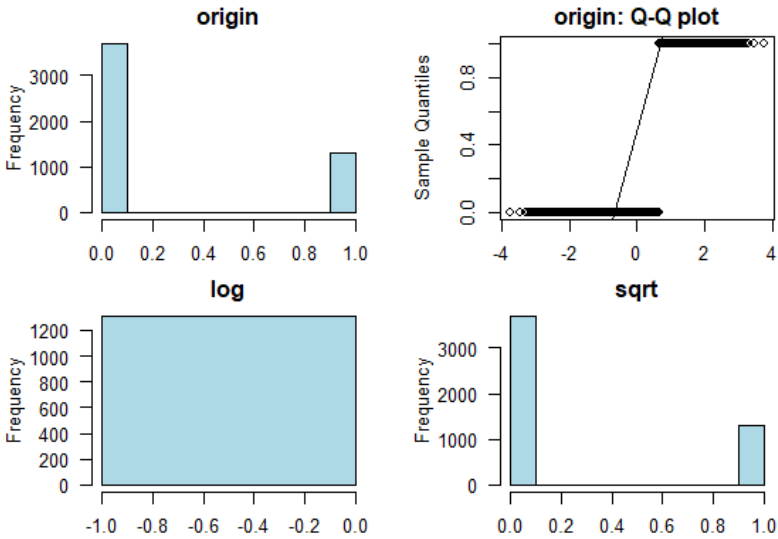


**[ cutVery.Good ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.50808,  p-value : 8.49257E-80
```

### skewness and kurtosis

| type | skewness | kurtosis |
|---|---|---|
| original | 1.371684 | 2.881517 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.371684 | 2.881517 |
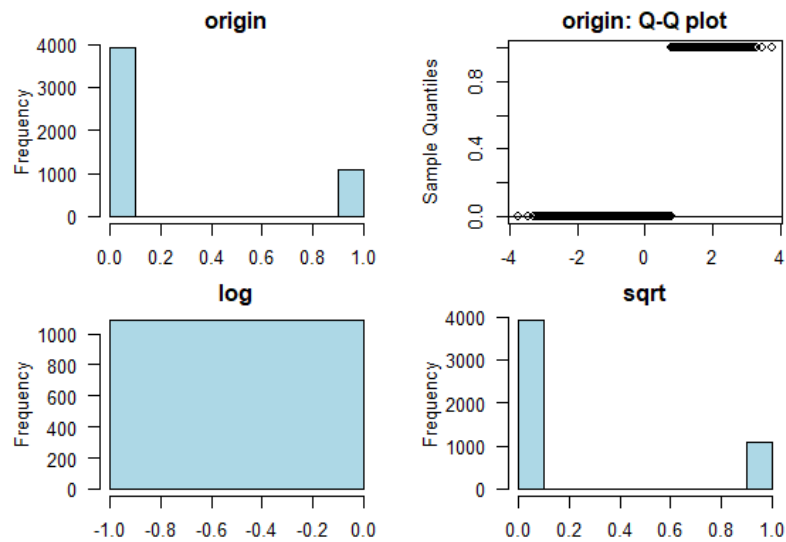
**[ colorD ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.3904,  p-value : 1.71178E-84
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 2.243134 | 6.031648 |
| log transformation | NaN | NaN |
| sqrt transformation | 2.243134 | 6.031648 |
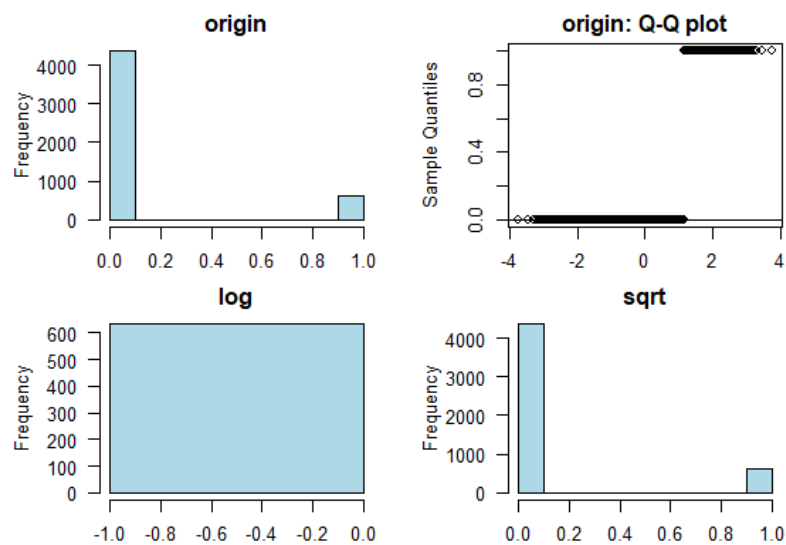


**[ colorE ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.46745,  p-value : 1.6136E-81
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 1.658818 | 3.751677 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.658818 | 3.751677 |



**[ colorF ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.44749,  p-value : 2.53099E-82
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 1.803098 | 4.251161 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.803098 | 4.251161 |
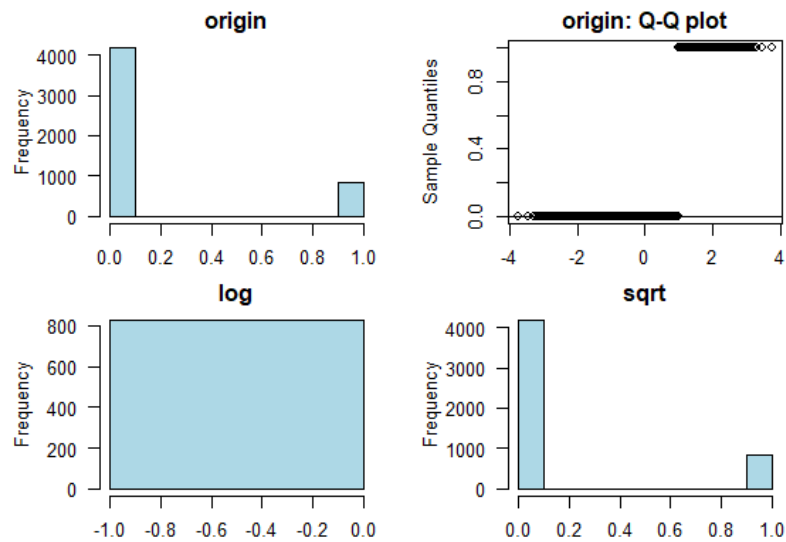
**[ colorG ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.50563,  p-value : 6.63779E-80
```

skewness and kurtosis

| type | skewness | kurtosis |
|---|---|---|
| original | 1.388904 | 2.929054 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.388904 | 2.929054 |
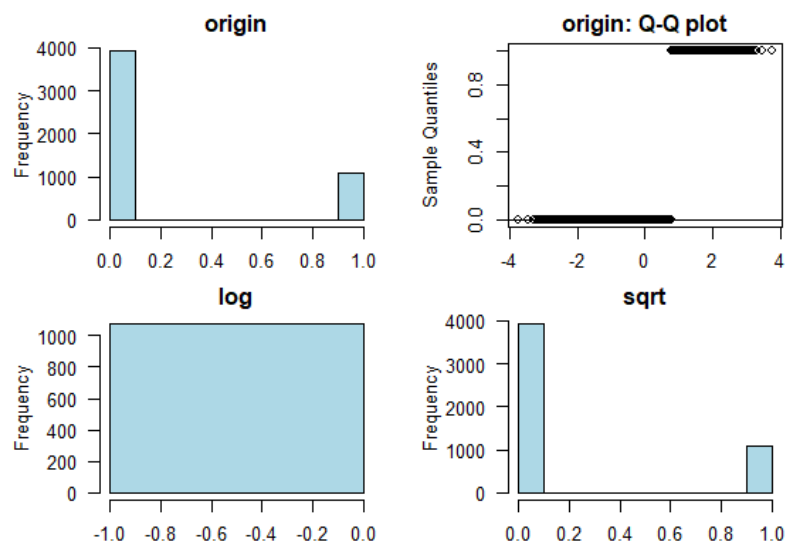


**[ colorH ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.42519,  p-value : 3.41585E-83
```

### skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 1.969207 | 4.877777 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.969207 | 4.877777 |
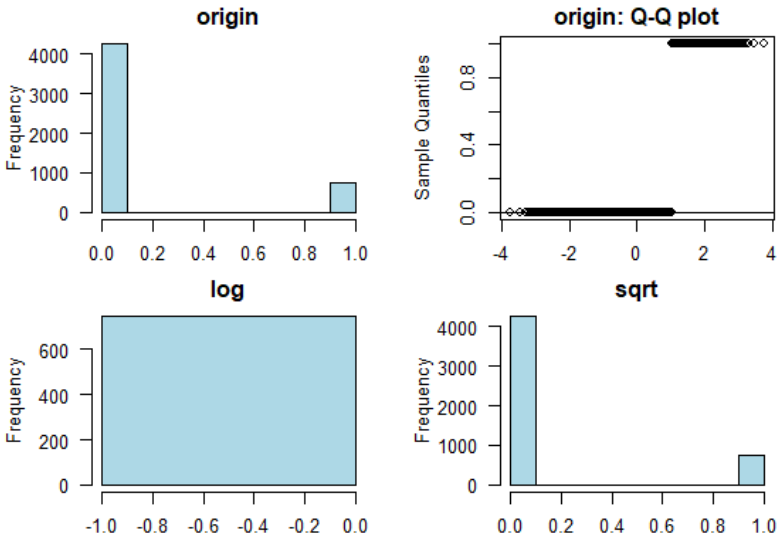


**[ colorI ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.35706,  p-value : 1.11287E-85
```

### skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 2.529281 | 7.397263 |
| log transformation | NaN | NaN |
| sqrt transformation | 2.529281 | 7.397263 |
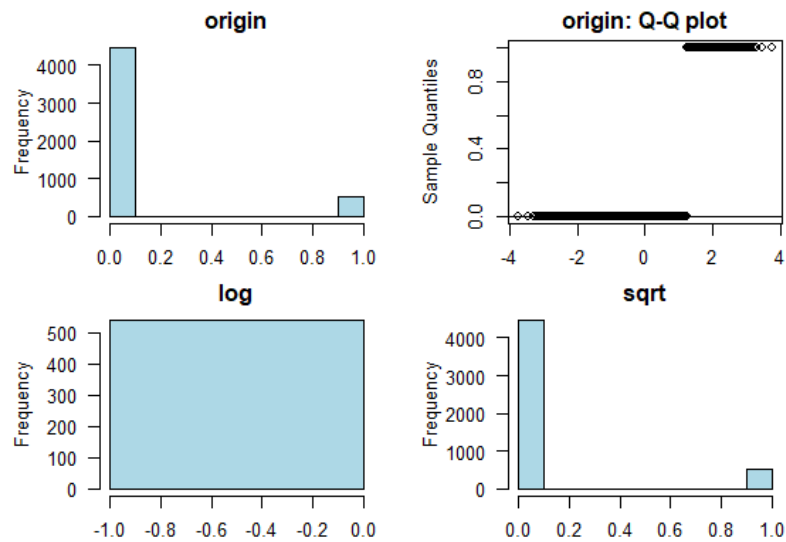
**[ colorJ ]**

normality test : Shapiro-Wilk normality test

statistic : 0.2473,  p-value : 3.04331E-89

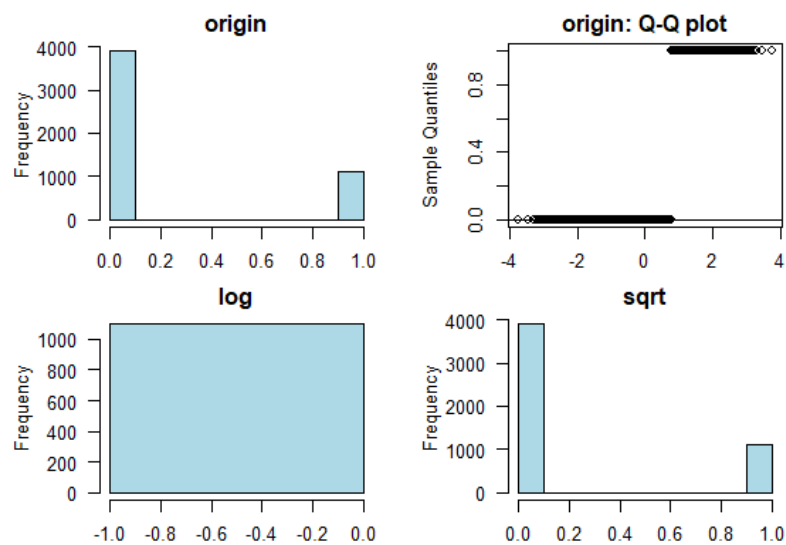| skewness and kurtosis | | |
| --- | --- | --- |
| type | skewness | kurtosis |
| original | 3.766339 | 15.18531 |
| log transformation | NaN | NaN |
| sqrt transformation | 3.766339 | 15.18531 |
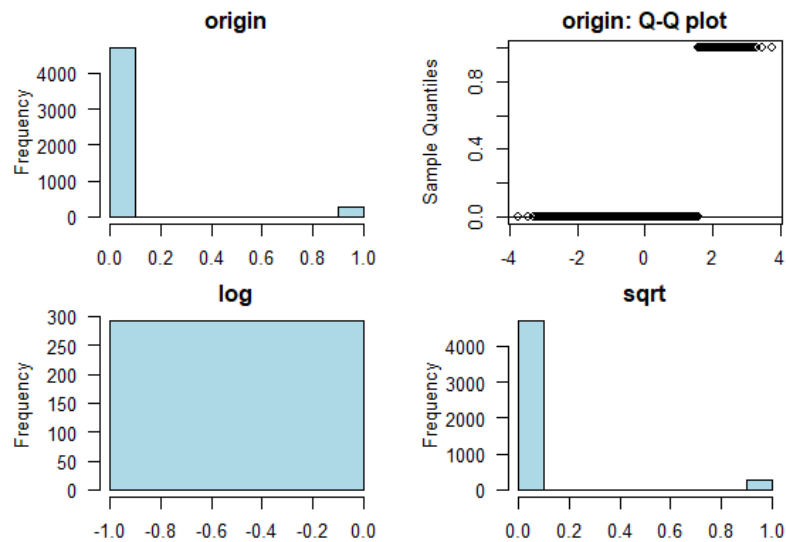


**[ clarityI1 ]**

normality test : Shapiro-Wilk normality test

statistic : 0.09704,  p-value : 1.89034E-93

### skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 7.980094 | 64.6819 |
| log transformation | NaN | NaN |
| sqrt transformation | 7.980094 | 64.6819 |
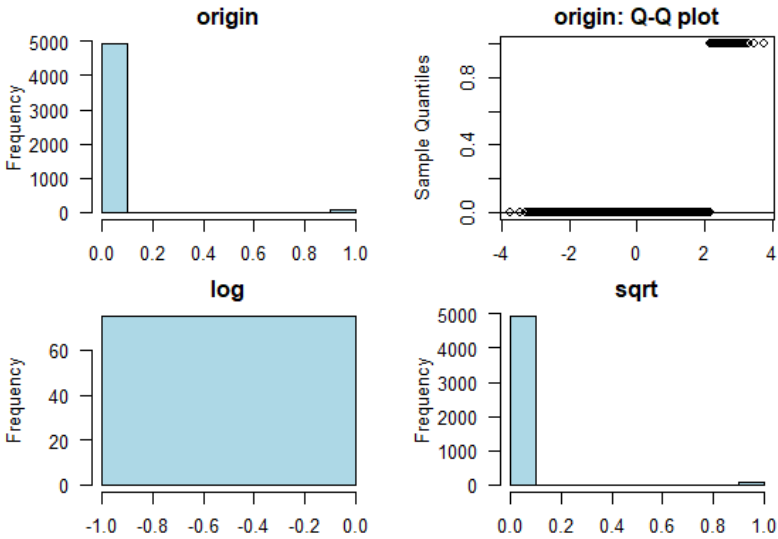


**[ clarityIF ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.15164,  p-value : 5.36146E-92
```

### skewness and kurtosis

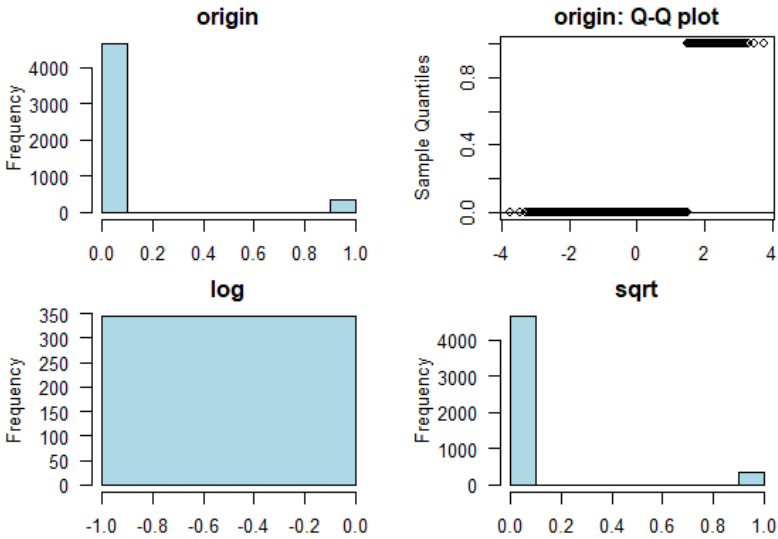| type | skewness | kurtosis |
|------|----------|----------|
| original | 5.722158 | 33.74309 |
| log transformation | NaN | NaN |
| sqrt transformation | 5.722158 | 33.74309 |
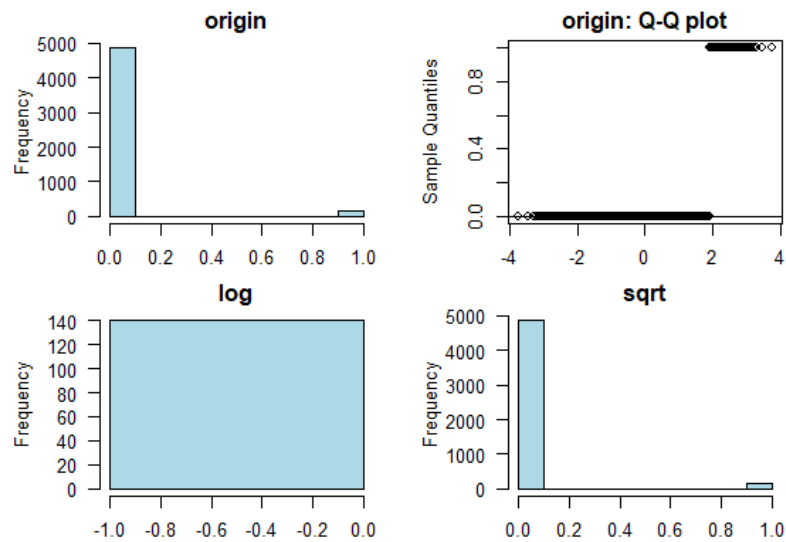
## [ claritySI1 ]

normality test : Shapiro-Wilk normality test

statistic : 0.53754,  p-value : 1.80134E-78

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 1.162108 | 2.350496 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.162108 | 2.350496 |
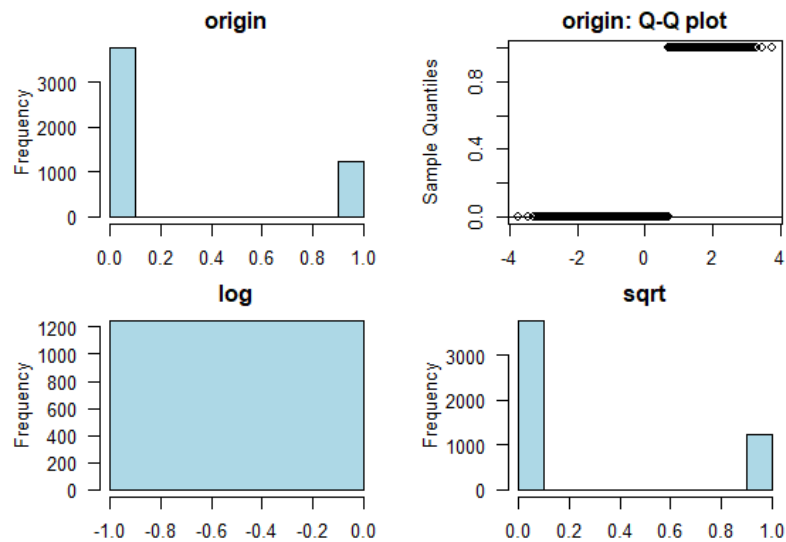


## [ claritySI2 ]

normality test : Shapiro-Wilk normality test

statistic : 0.44291,  p-value : 1.66854E-82

### skewness and kurtosis

| type | skewness | kurtosis |
| --- | --- | --- |
| original | 1.83671 | 4.373502 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.83671 | 4.373502 |
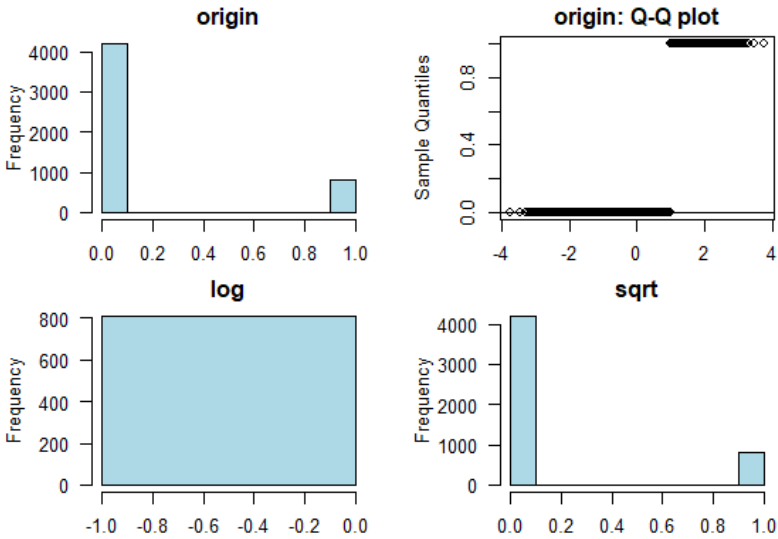


**[ clarityVS1 ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.42837,  p-value : 4.52777E-83
```

### skewness and kurtosis

| type | skewness | kurtosis |
| --- | --- | --- |
| original | 1.945104 | 4.783431 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.945104 | 4.783431 |
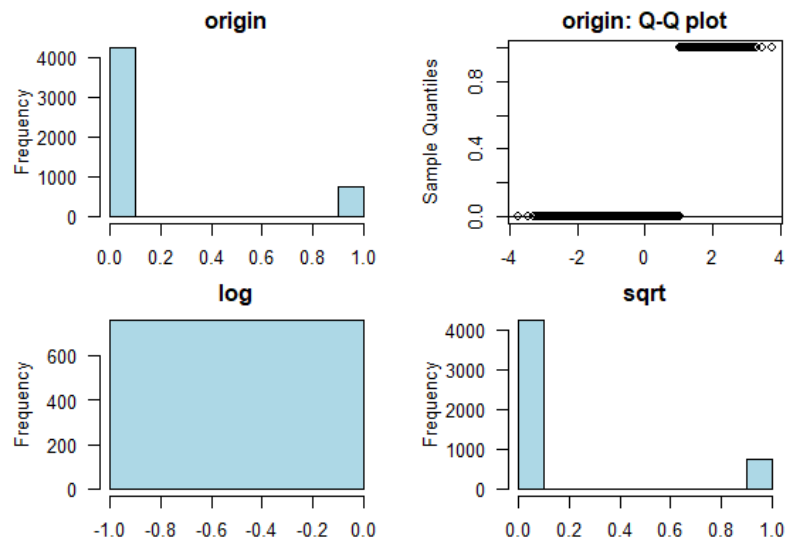
## [ clarityVS2 ]

normality test : Shapiro-Wilk normality test

```
statistic : 0.51069,  p-value : 1.10653E-79
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 1.35326 | 2.831313 |
| log transformation | NaN | NaN |
| sqrt transformation | 1.35326 | 2.831313 |



## [ clarityVVS1 ]

normality test : Shapiro-Wilk normality test

```
statistic : 0.27396,  p-value : 2.01804E-88
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 3.407164 | 12.60877 |
| log transformation | NaN | NaN |
| sqrt transformation | 3.407164 | 12.60877 |



**[ clarityVVS2 ]**

normality test : Shapiro-Wilk normality test

```
statistic : 0.33506,  p-value : 1.95835E-86
```

skewness and kurtosis

| type | skewness | kurtosis |
|------|----------|----------|
| original | 2.734982 | 8.480128 |
| log transformation | NaN | NaN |
| sqrt transformation | 2.734982 | 8.480128 |

**origin**

**origin: Q-Q plot**

**log**

**sqrt**
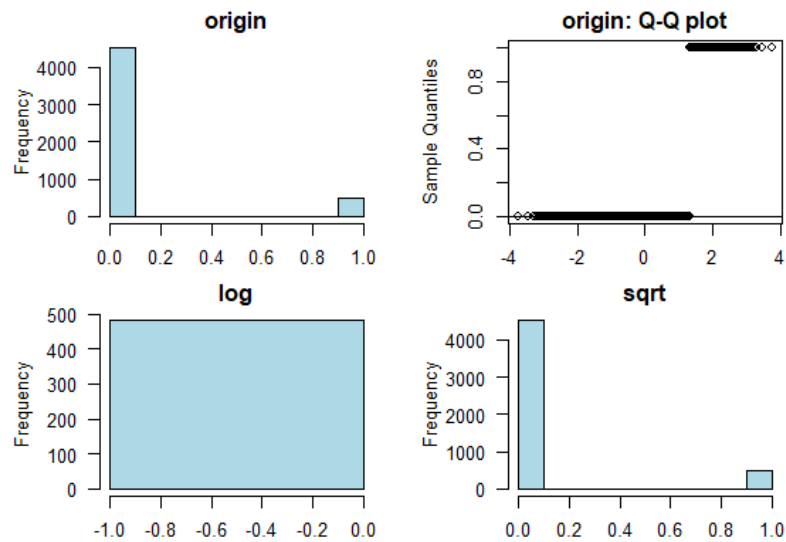
# 3 Relationship Between Variables

## 3.1 Correlation Coefficient

### 3.1.1 Correlation Coefficient by Variable Combination

Table of correlation coefficients (0.5 or more)

| Variable1 | Variable2 | Correlation Coefficient |
|---|---|---|
| x | carat | 0.9750942 |
| y | x | 0.9747015 |
| z | x | 0.9707718 |
| z | carat | 0.9533874 |
| z | y | 0.9520057 |
| y | carat | 0.9517222 |
| price | carat | 0.9215913 |
| x | price | 0.8844352 |
| y | price | 0.8654209 |
| z | price | 0.8612494 |
| cutIdeal | table | -0.5495988 |

### 3.1.2 Correlation Plot of Numerical Variables
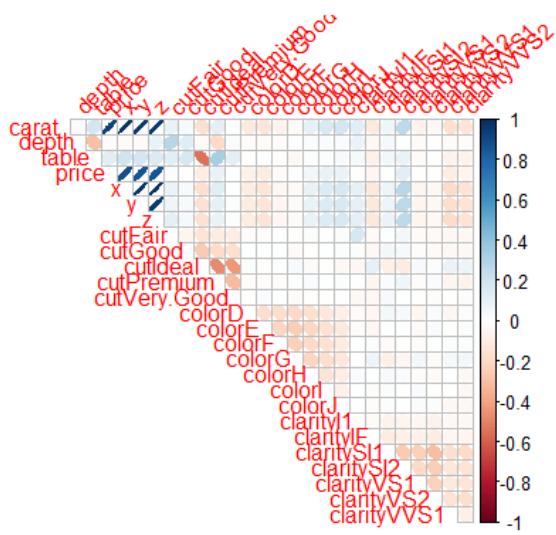
# 4 Target based Analysis

# 4.1 Grouped Descriptive Statistics

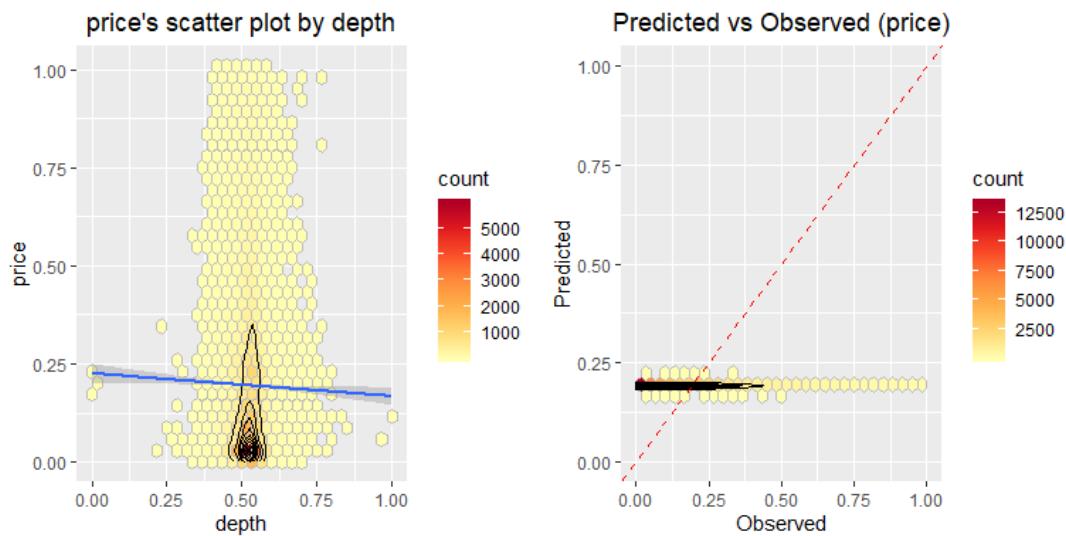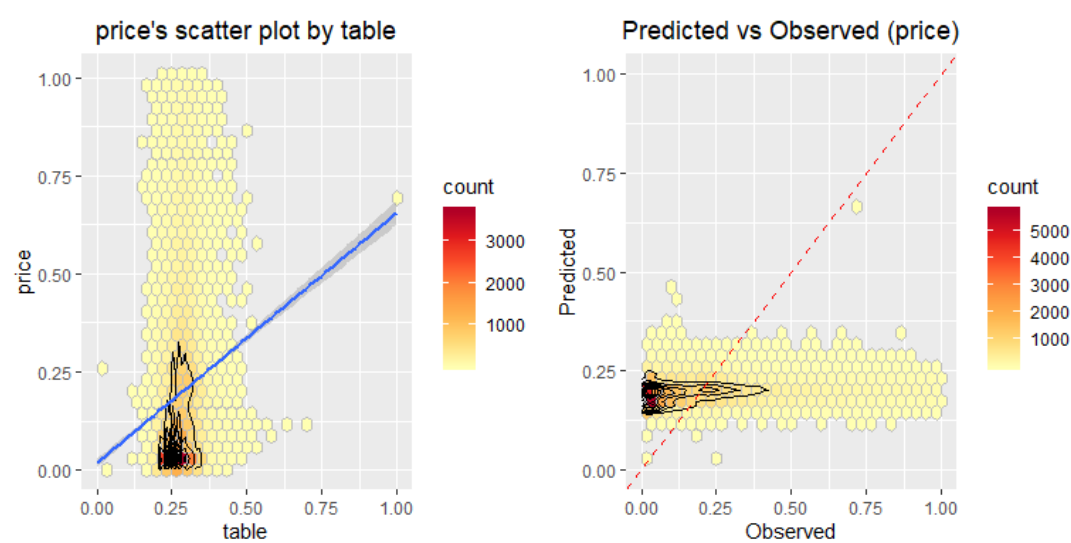### 4.1.1 Grouped Numerical Variables

**[ carat ]**

Simple Linear Model Information

Residual standard error: 0 on 53938 degrees of freedom Multiple R-squared: 0.84933, Adjusted R-squared: 0.84933 F-statistic: 304051 on 1 and 53938 DF, p-value: 0

Coefficients:

Simple Linear Model coefficients : carat

|             | Estimate | Std. Error | t value | Pr(>\|t\|) |
|-------------|----------|------------|---------|------------|
| (Intercept) | -0.06    | 0          | -96.06  | 0          |
| carat       | 2.02     | 0          | 551.41  | 0          |

Visualization:

price's scatter plot by carat        Predicted vs Observed (price)
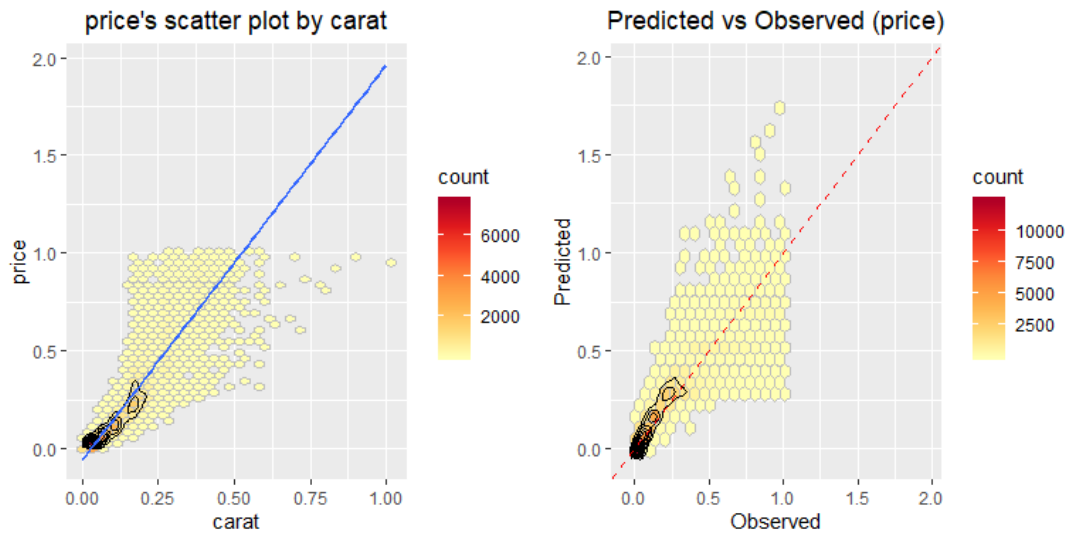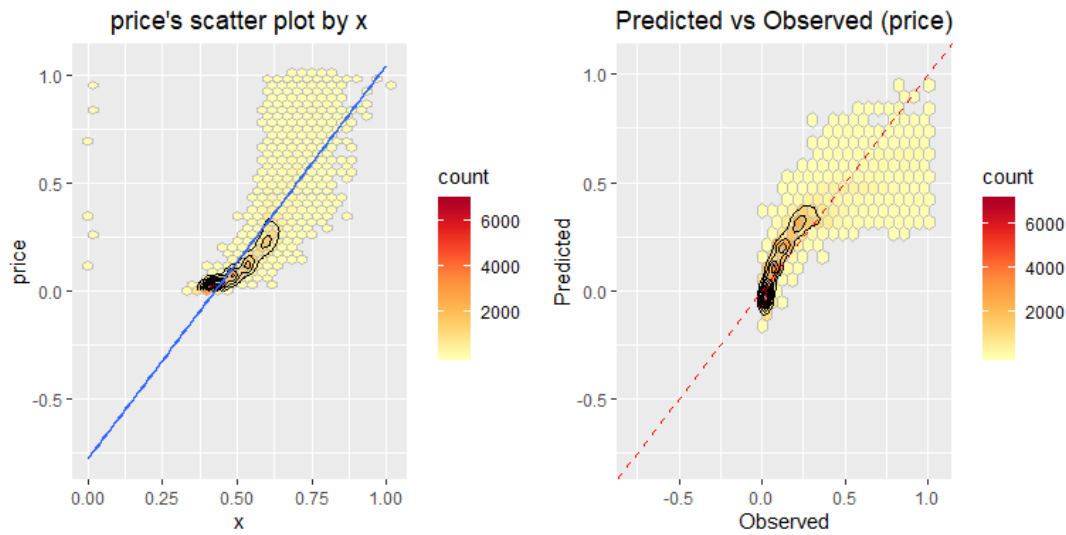
**[ depth ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00011,  Adjusted R-squared:  9e-05
F-statistic: 6 on 1 and 53938 DF,  p-value: 0.0134032490126328
```

Coefficients:

Simple Linear Model coefficients : depth

|             | Estimate | Std. Error | t value | Pr(>\|t\|) |
|-------------|----------|------------|---------|-----------|
| (Intercept) | 0.23     | 0.01       | 18.46   | 0.00      |
| depth       | -0.06    | 0.02       | -2.47   | 0.01      |

Visualization:



price's scatter plot by depth        Predicted vs Observed (price)
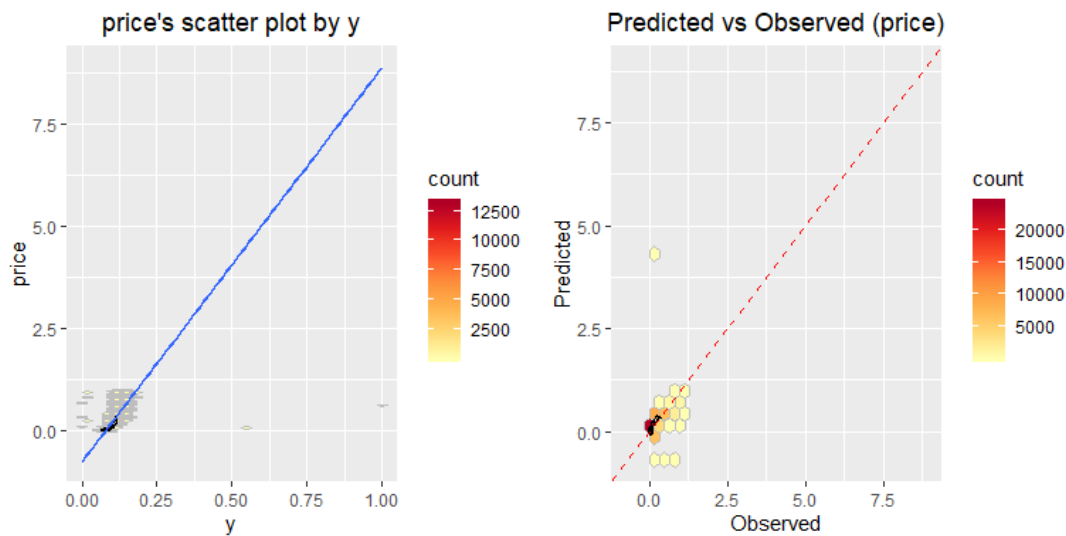
**[ table ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.01616,  Adjusted R-squared:  0.01614
F-statistic: 886 on 1 and 53938 DF,  p-value: 3.76153391341822e-193
```

Coefficients:

Simple Linear Model coefficients : table

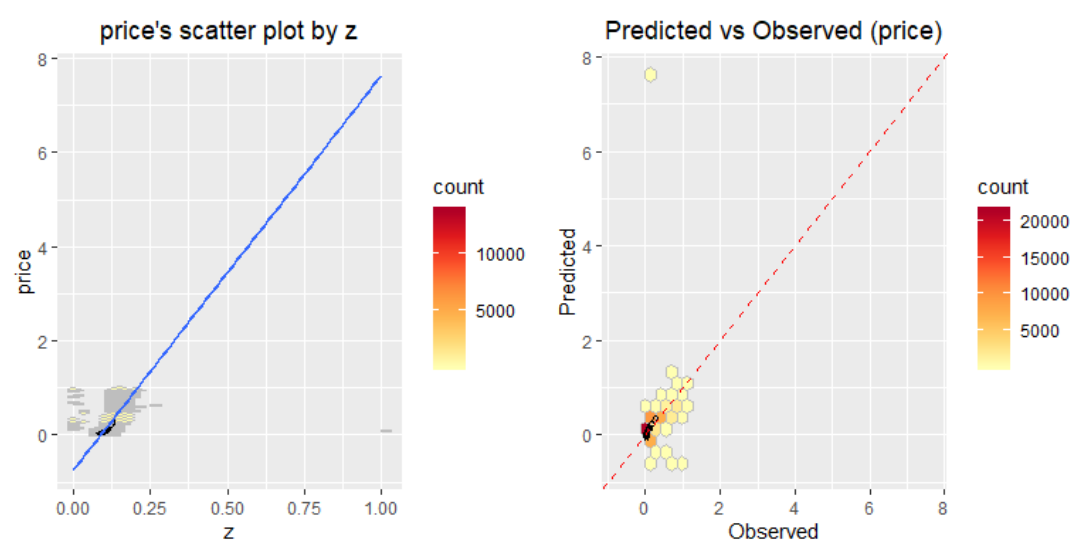|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept) | 0.02     | 0.01       | 2.92    | 0        |
| table      | 0.64     | 0.02       | 29.77   | 0        |

Visualization:



[ x ]

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.78223,  Adjusted R-squared:  0.78222
F-statistic: 193740 on 1 and 53938 DF,  p-value: 0
```

Coefficients:

Simple Linear Model coefficients : x

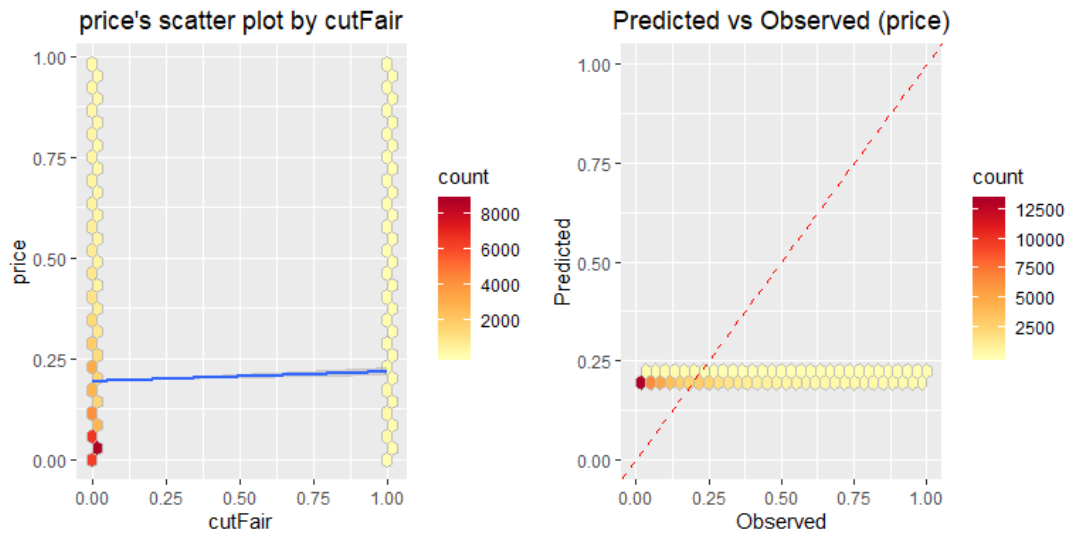|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept) | -0.78    | 0          | -345.54 | 0        |
| x          | 1.83     | 0          | 440.16  | 0        |

Visualization:

## [ y ]

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.74895,  Adjusted R-squared:  0.74895
F-statistic: 160914 on 1 and 53938 DF,  p-value: 0
```

Coefficients:

Simple Linear Model coefficients : y

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | -0.74    | 0.00       | -311.56 | 0        |
| y           | 9.63     | 0.02       | 401.14  | 0        |

Visualization:
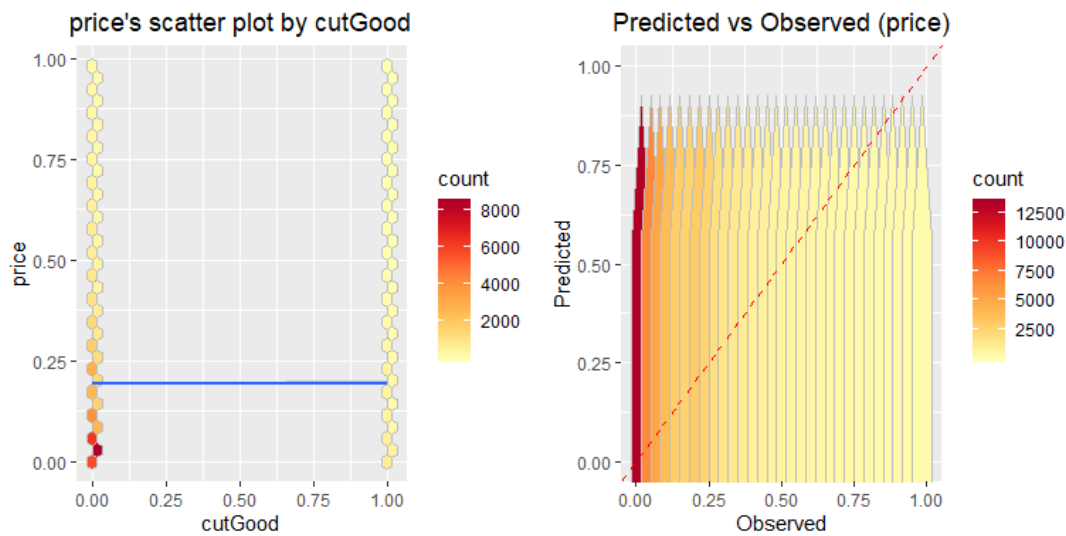


## [ z ]

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.74175,  Adjusted R-squared:  0.74175
F-statistic: 154922 on 1 and 53938 DF,  p-value: 0
```

## Coefficients:

Simple Linear Model coefficients : z

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.74 | 0.00 | -305.2 | 0 |
| z | 8.37 | 0.02 | 393.6 | 0 |

## Visualization:



**[ cutFair ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00035,  Adjusted R-squared:  0.00033
F-statistic: 19 on 1 and 53938 DF,  p-value: 1.36195086376903e-05
```

## Coefficients:

Simple Linear Model coefficients : cutFair

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.19 | 0.00 | 206.10 | 0 |
| cutFair | 0.02 | 0.01 | 4.35 | 0 |

## Visualization:

**[ cutGood ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0,    Adjusted R-squared:  -2e-05
F-statistic: 0 on 1 and 53938 DF,  p-value: 0.94223200332364
```

Coefficients:

Simple Linear Model coefficients : cutGood

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.2 | 0 | 200.22 | 0.00 |
| cutGood | 0.0 | 0 | -0.07 | 0.94 |

Visualization:



**[ cutIdeal ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00944,  Adjusted R-squared:  0.00942
F-statistic: 514 on 1 and 53938 DF,  p-value: 2.63270266414303e-113
```

Coefficients:

Simple Linear Model coefficients : cutIdeal

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.21 | 0 | 177.81 | 0 |
| cutIdeal | -0.04 | 0 | -22.68 | 0 |

Visualization:
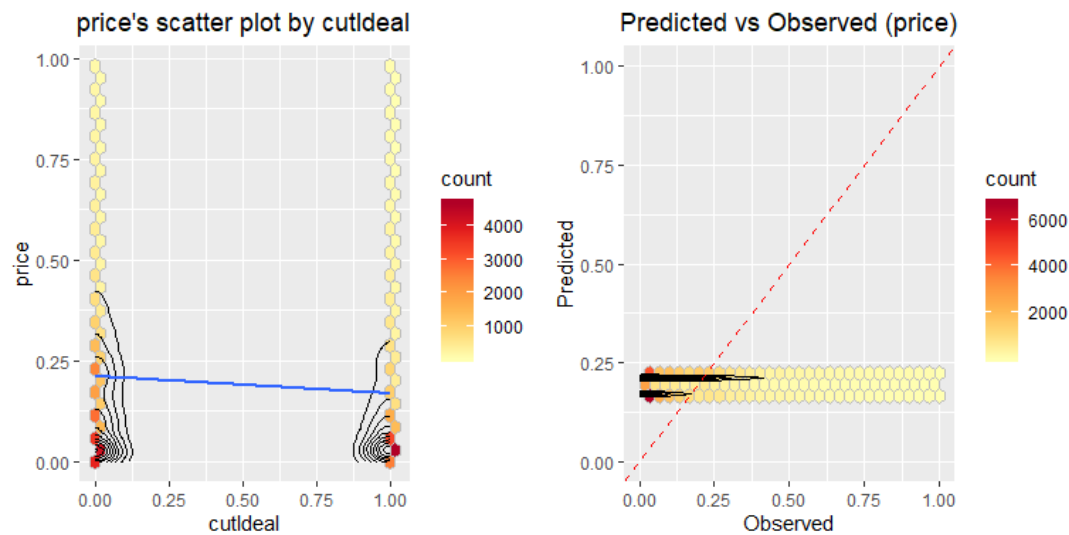


**[ cutPremium ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00916,  Adjusted R-squared:  0.00914
F-statistic: 499 on 1 and 53938 DF,  p-value: 5.99431748578085e-110
```

Coefficients:

Simple Linear Model coefficients : cutPremium

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.18 | 0 | 170.70 | 0 |
| cutPremium | 0.05 | 0 | 22.33 | 0 |

Visualization:
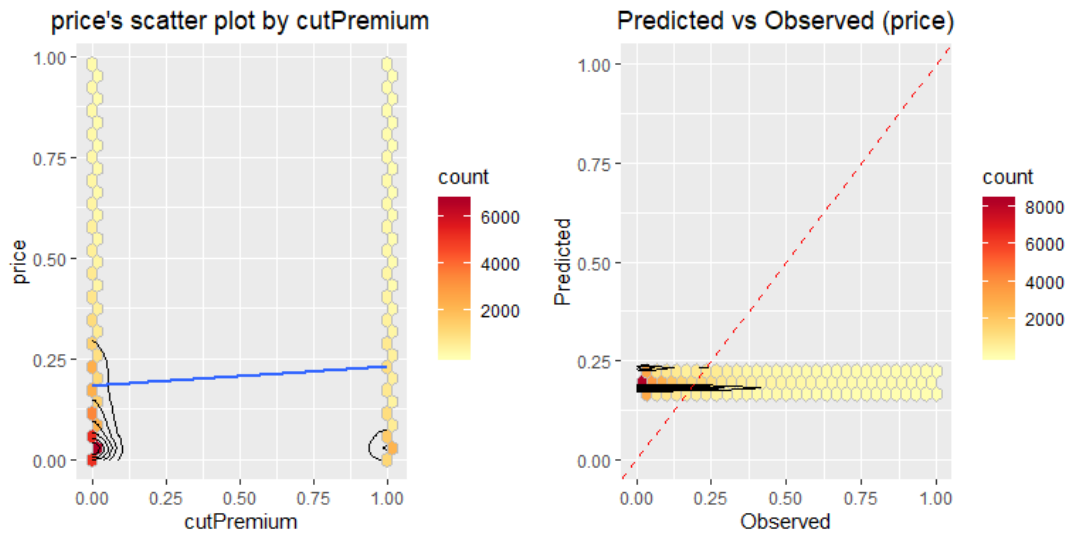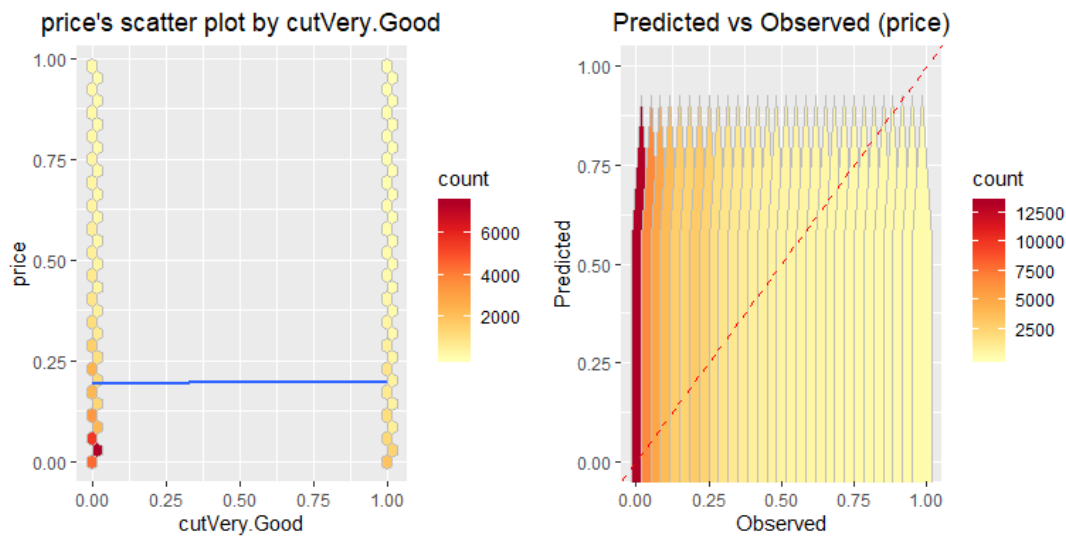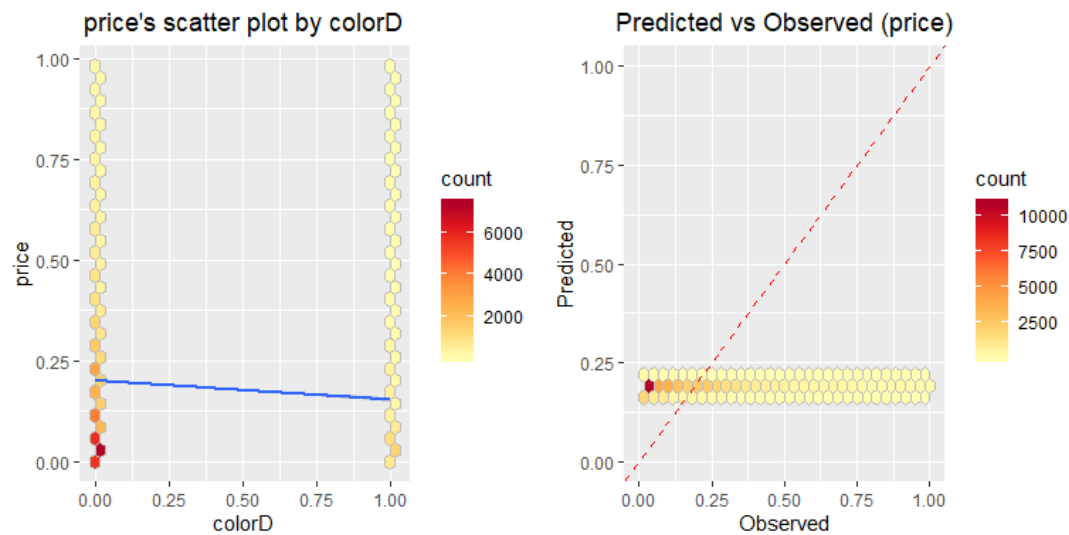
**[ cutVery.Good ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  4e-05,    Adjusted R-squared:  2e-05
F-statistic: 2 on 1 and 53938 DF,  p-value: 0.125690870103525
```

Coefficients:

Simple Linear Model coefficients : cutVery.Good

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.19 | 0 | 184.25 | 0.00 |
| cutVery.Good | 0.00 | 0 | 1.53 | 0.13 |

Visualization:
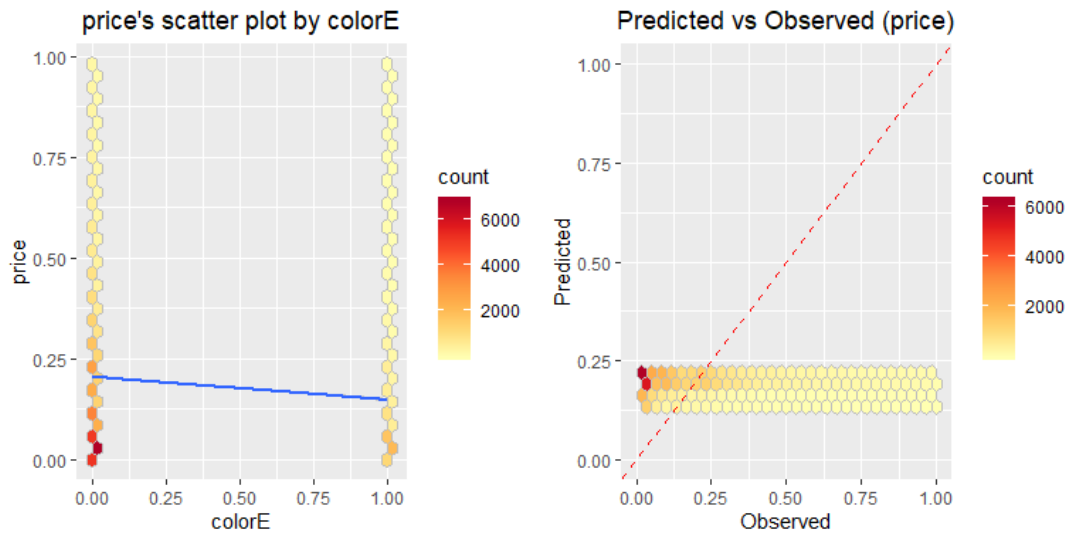


**[ colorD ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00525,  Adjusted R-squared:  0.00523
F-statistic: 285 on 1 and 53938 DF,  p-value: 9.88645356147312e-64
```

Coefficients:

Simple Linear Model coefficients : colorD

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.20 | 0 | 202.84 | 0 |
| colorD | -0.05 | 0 | -16.88 | 0 |

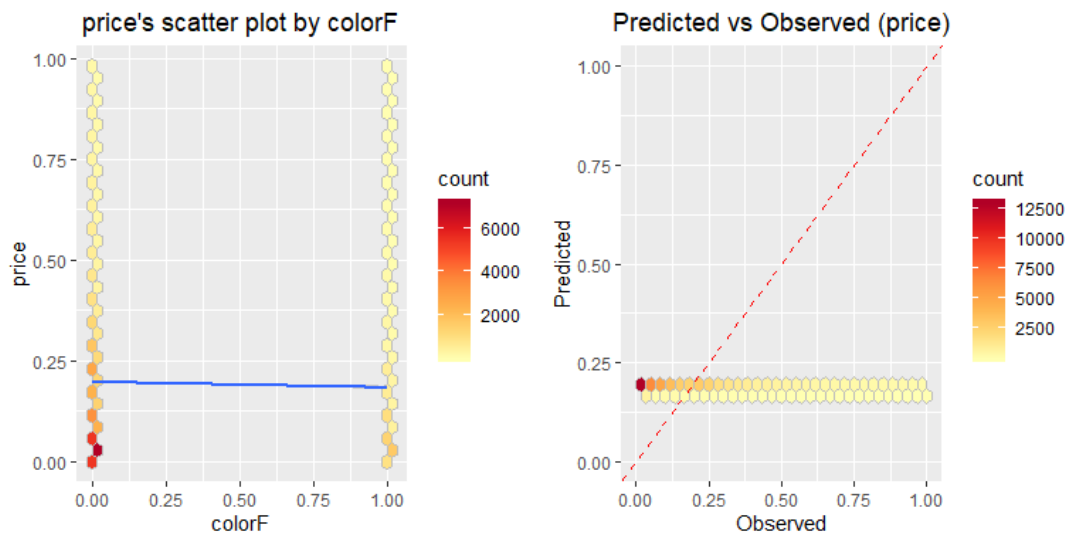Visualization:



**[ colorE ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.01022,  Adjusted R-squared:  0.0102
F-statistic: 557 on 1 and 53938 DF,  p-value: 1.67543029465086e-122
```

Coefficients:

Simple Linear Model coefficients : colorE

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.21 | 0 | 200.98 | 0 |
| colorE | -0.06 | 0 | -23.60 | 0 |

Visualization:

**[ colorF ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00058,  Adjusted R-squared:  0.00057
F-statistic: 32 on 1 and 53938 DF,  p-value: 1.99947932659754e-08
```

Coefficients:

Simple Linear Model coefficients : colorF

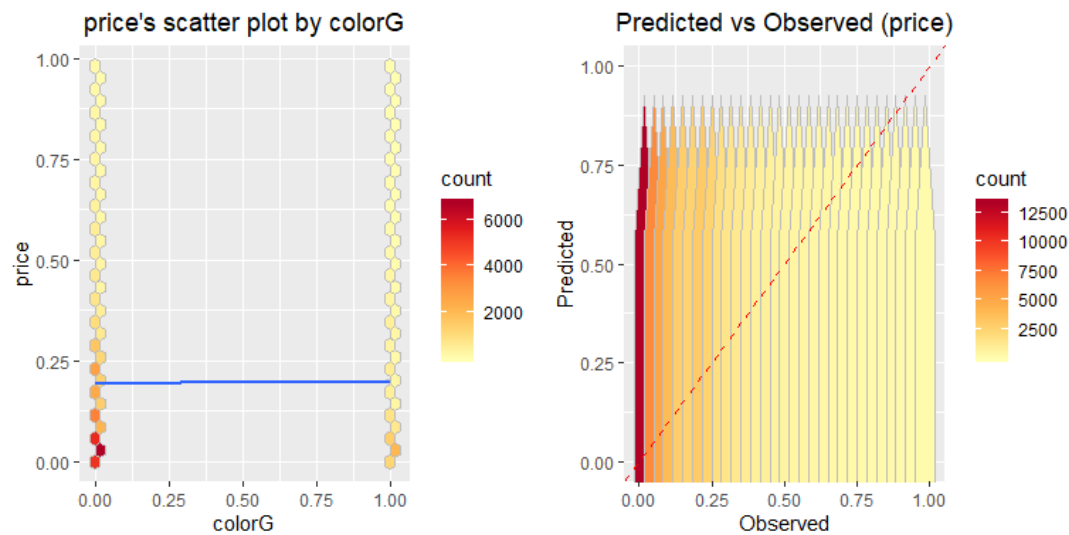|             | Estimate | Std. Error | t value | Pr(>\|t\|) |
|-------------|----------|------------|---------|-----------|
| (Intercept) | 0.20     | 0          | 192.91  | 0         |
| colorF      | -0.01    | 0          | -5.61   | 0         |

Visualization:



**[ colorG ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared: 7e-05,    Adjusted R-squared:  5e-05
F-statistic: 4 on 1 and 53938 DF,  p-value: 0.046905842974847
```

Coefficients:

Simple Linear Model coefficients : colorG

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.19 | 0 | 185.80 | 0.00 |
| colorG | 0.00 | 0 | 1.99 | 0.05 |

Visualization:



[ colorH ]

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00351,  Adjusted R-squared:  0.00349
F-statistic: 190 on 1 and 53938 DF,  p-value: 4.06688973807318e-43
```

Coefficients:

Simple Linear Model coefficients : colorH

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.19 | 0 | 188.07 | 0 |
| colorH | 0.04 | 0 | 13.78 | 0 |

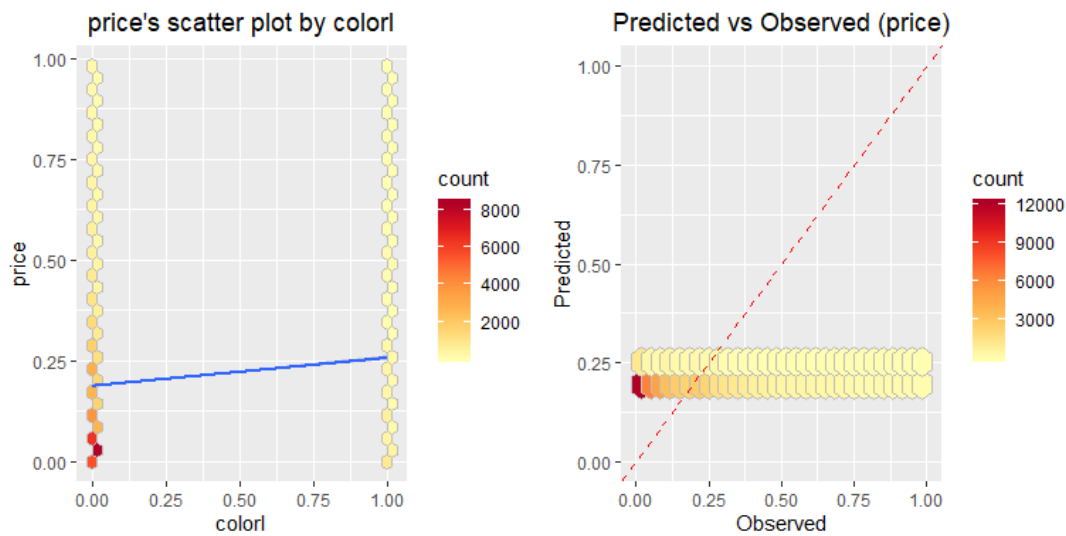Visualization:
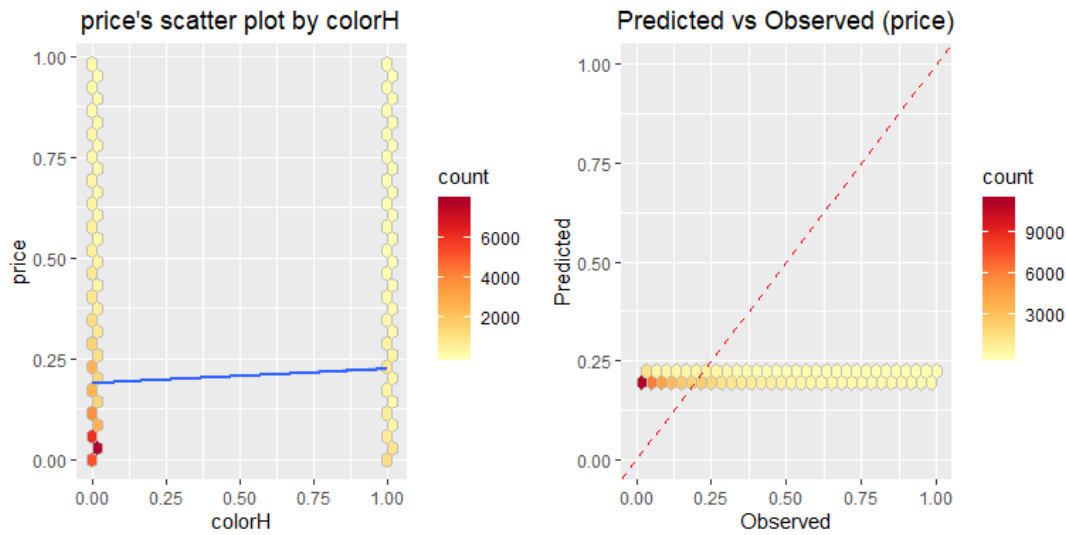
**[ colorI ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00943,  Adjusted R-squared:  0.00941
F-statistic: 514 on 1 and 53938 DF,  p-value: 3.43443829392871e-113
```

Coefficients:

Simple Linear Model coefficients : colorI

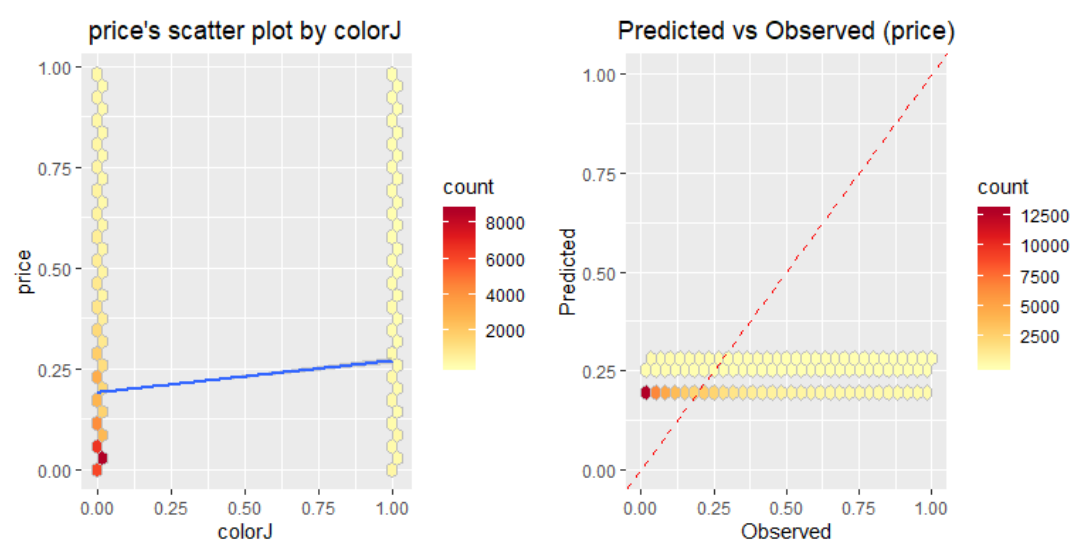|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.19     | 0          | 192.90  | 0        |
| colorI      | 0.07     | 0          | 22.66   | 0        |

Visualization:



**[ colorJ ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00668,  Adjusted R-squared:  0.00666
F-statistic: 363 on 1 and 53938 DF,  p-value: 1.44846492916451e-80
```

## Coefficients:

### Simple Linear Model coefficients : colorJ

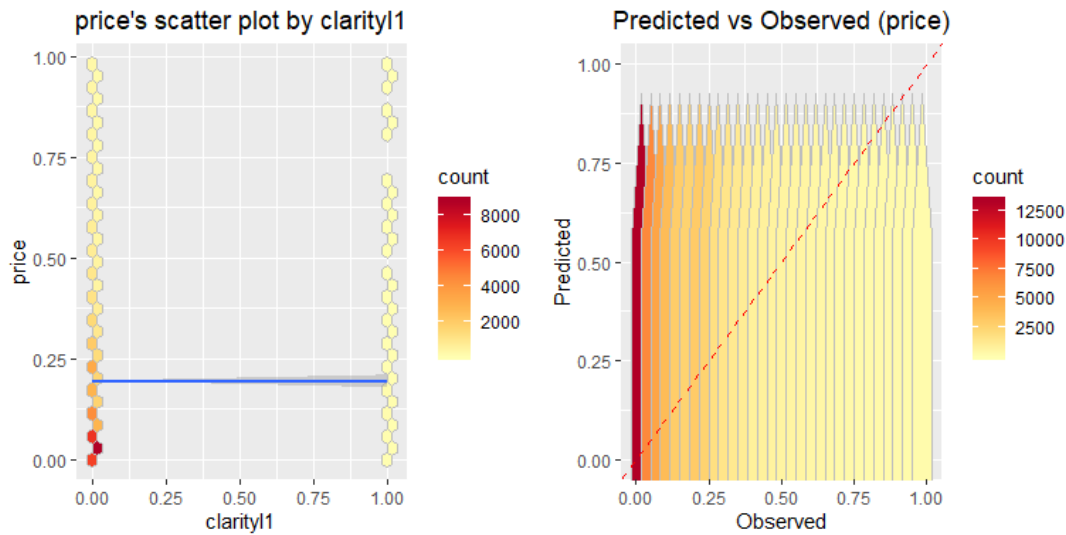|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.19     | 0          | 200.78  | 0        |
| colorJ      | 0.08     | 0          | 19.04   | 0        |

## Visualization:



## [ clarityI1 ]

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0,    Adjusted R-squared:  -2e-05
F-statistic: 0 on 1 and 53938 DF,  p-value: 0.952712847846628
```

## Coefficients:

### Simple Linear Model coefficients : clarityI1

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.2      | 0.00       | 208.53  | 0.00     |
| clarityI1   | 0.0      | 0.01       | -0.06   | 0.95     |

## Visualization:

price's scatter plot by clarityI1

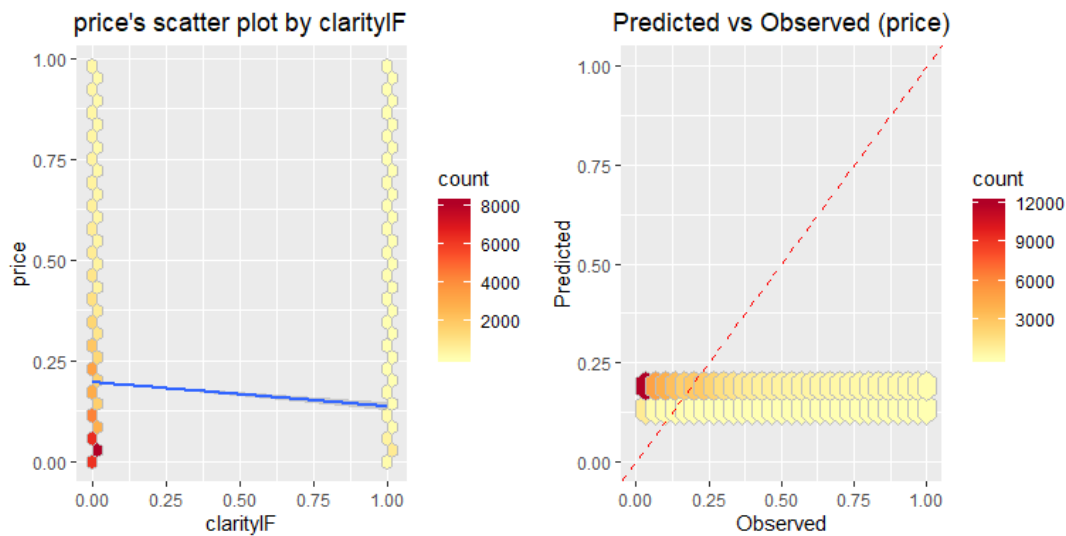Predicted vs Observed (price)

**[ clarityIF ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00246,  Adjusted R-squared:  0.00244
F-statistic: 133 on 1 and 53938 DF,  p-value: 9.81503533910171e-31
```

Coefficients:

Simple Linear Model coefficients : clarityIF

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.20 | 0.00 | 208.81 | 0 |
| clarityIF | -0.06 | 0.01 | -11.53 | 0 |

Visualization:



price's scatter plot by clarityIF

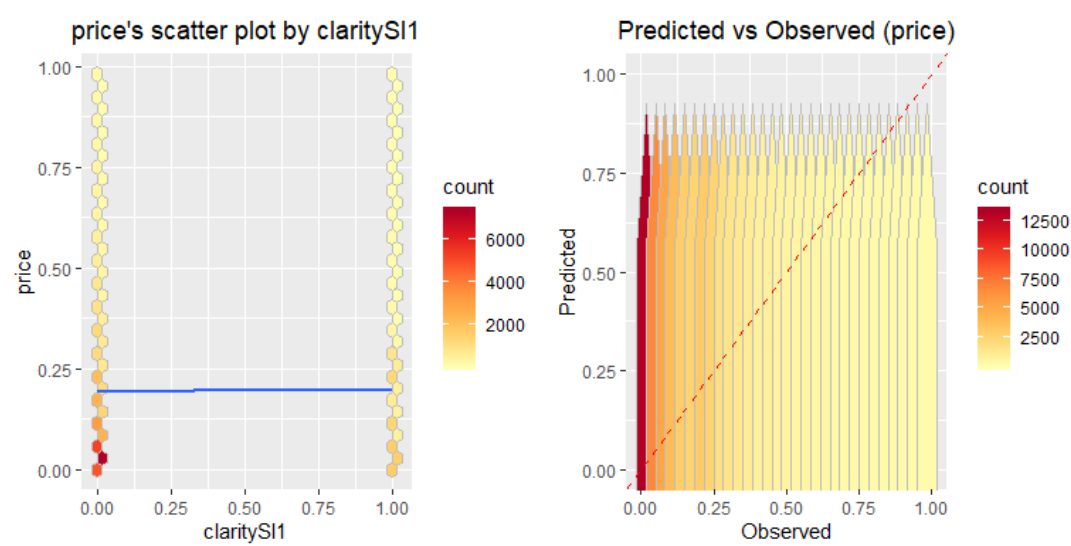Predicted vs Observed (price)

**[ claritySI1 ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared: 8e-05,   Adjusted R-squared:  6e-05
F-statistic: 4 on 1 and 53938 DF,  p-value: 0.0375098957404741
```

## Coefficients:

### Simple Linear Model coefficients : claritySI1

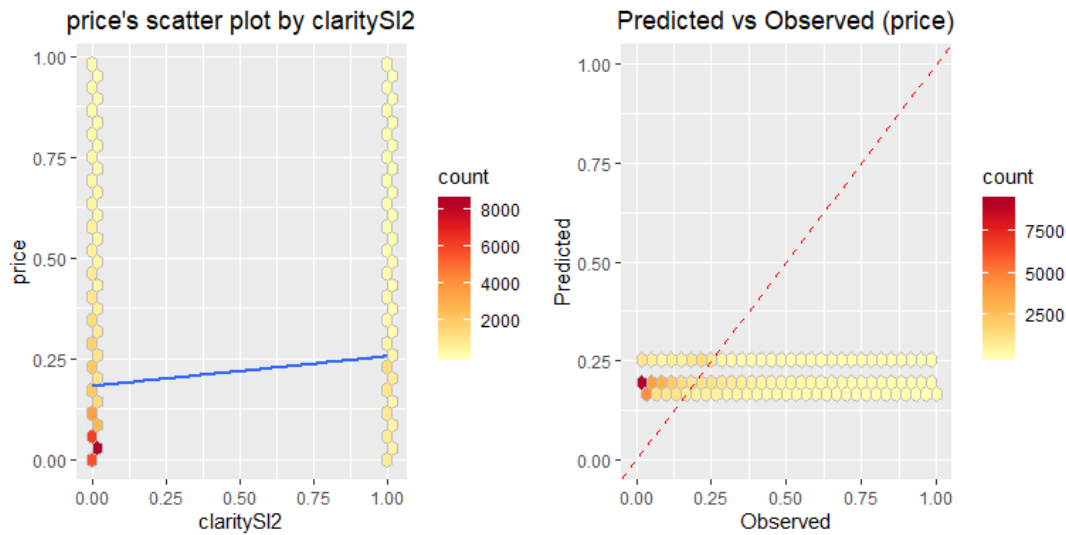|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.19 | 0 | 181.77 | 0.00 |
| claritySI1 | 0.00 | 0 | 2.08 | 0.04 |

## Visualization:



**[ claritySI2 ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.01649, Adjusted R-squared:  0.01647
F-statistic: 904 on 1 and 53938 DF,  p-value: 4.53624465727249e-197
```

## Coefficients:

### Simple Linear Model coefficients : claritySI2

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.18 | 0 | 180.42 | 0 |
| claritySI2 | 0.07 | 0 | 30.07 | 0 |

## Visualization:

### price's scatter plot by claritySI2



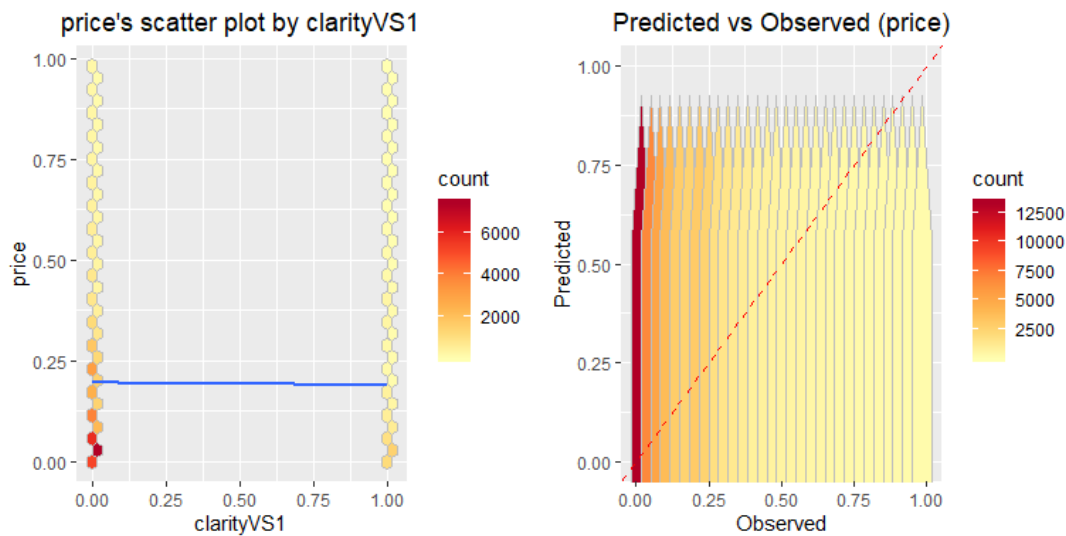### Predicted vs Observed (price)



**[ clarityVS1 ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  1e-04,    Adjusted R-squared:  8e-05
F-statistic: 5 on 1 and 53938 DF,  p-value: 0.021671017491959
```

Coefficients:

Simple Linear Model coefficients : clarityVS1

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.20 | 0 | 194.32 | 0.00 |
| clarityVS1 | -0.01 | 0 | -2.30 | 0.02 |

Visualization:

### price's scatter plot by clarityVS1



### Predicted vs Observed (price)
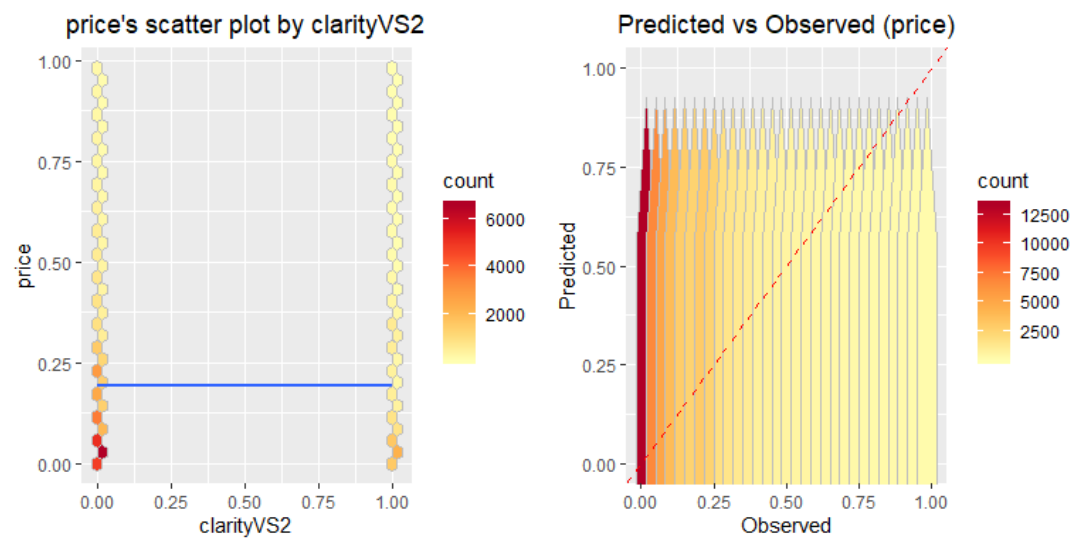


**[ clarityVS2 ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0,    Adjusted R-squared:  -2e-05
F-statistic: 0 on 1 and 53938 DF,  p-value: 0.805240068913232
```

## Coefficients:

### Simple Linear Model coefficients : clarityVS2

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.2      | 0          | 184.70  | 0.00     |
| clarityVS2  | 0.0      | 0          | -0.25   | 0.81     |

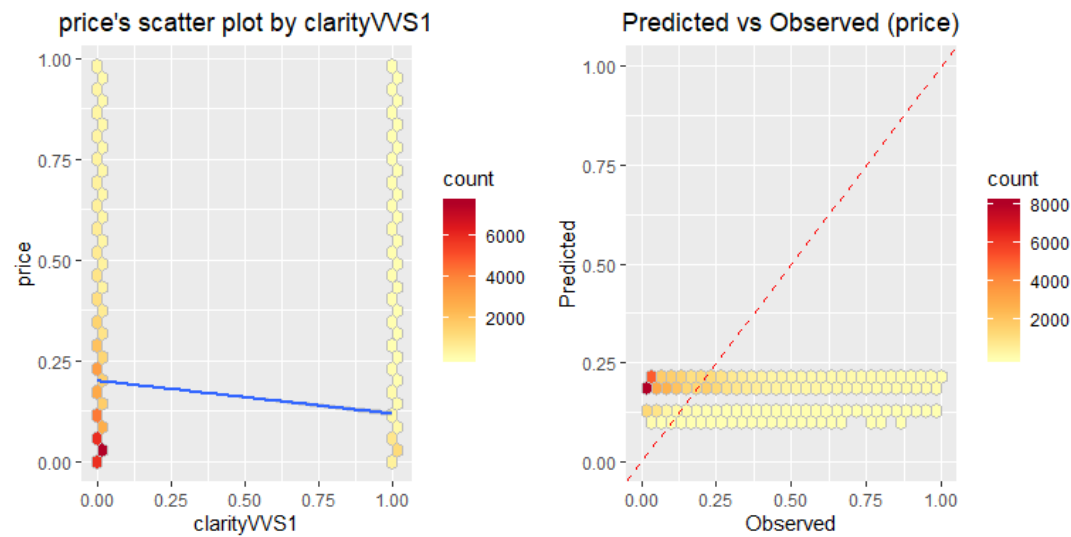## Visualization:



**[ clarityVVS1 ]**

## Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00908,  Adjusted R-squared:  0.00906
F-statistic: 494 on 1 and 53938 DF,  p-value: 5.92278254156925e-109
```

## Coefficients:

### Simple Linear Model coefficients : clarityVVS1

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.20     | 0          | 209.45  | 0        |
| clarityVVS1 | -0.08    | 0          | -22.23  | 0        |

## Visualization:

**[ clarityVVS2 ]**

Simple Linear Model Information

```
Residual standard error: 0 on 53938 degrees of freedom
Multiple R-squared:  0.00274,  Adjusted R-squared:  0.00273
F-statistic: 148 on 1 and 53938 DF,  p-value: 4.30146025897724e-34
```

Coefficients:

Simple Linear Model coefficients : clarityVVS2

|  | Estimate | Std. Error | t value | Pr(>|t|) |
|---|---|---|---|---|
| (Intercept) | 0.20 | 0 | 203.88 | 0 |
| clarityVVS2 | -0.04 | 0 | -12.18 | 0 |

Visualization:
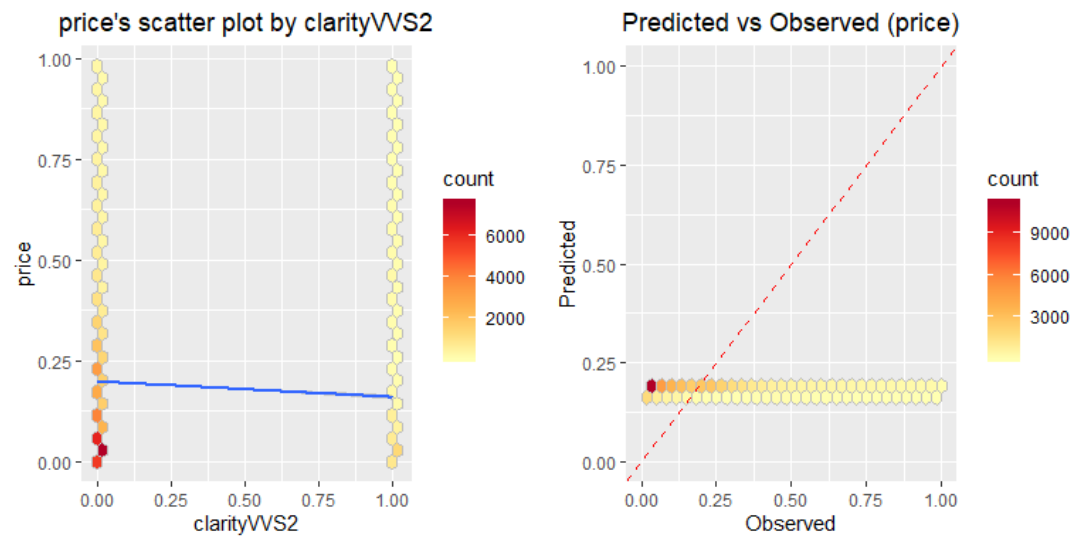
### 4.1.2 Grouped Categorical Variables

There is no categorical variable.

## 4.2 Grouped Relationship Between Variables

### 4.2.1 Grouped Correlation Coefficient

```
Numerical target variables are not supported.
```

### 4.2.2 Grouped Correlation Plot of Numerical Variables

```
Numerical target variables are not supported.
```