```
1 library(readr)
2 library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ purrr     1.0.2
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.4.4     ✔ tibble    3.2.1
✔ lubridate 1.9.3     ✔ tidyr     1.3.1
── Conflicts ──────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
1 df <- read.csv('./Real estate.csv')
2
3 head(df)
```

A data.frame: 6 × 8

ansaction.date	X2.house.age	X3.distance.to.the.nearest.MRT.station	X4.number.of.conven
<dbl>	<dbl>	<dbl>	
2012.917	32.0	84.87882	
2012.917	19.5	306.59470	
2013.583	13.3	561.98450	
2013.500	13.3	561.98450	
2012.833	5.0	390.56840	
2012.667	7.1	2175.03000	

Double-click (or enter) to edit

```
1 summary(df)
```

```
      No          X1.transaction.date  X2.house.age
 Min.   :  1.0   Min.   :2013        Min.   : 0.000
 1st Qu.:104.2   1st Qu.:2013        1st Qu.: 9.025
 Median :207.5   Median :2013        Median :16.100
 Mean   :207.5   Mean   :2013        Mean   :17.713
 3rd Qu.:310.8   3rd Qu.:2013        3rd Qu.:28.150
 Max.   :414.0   Max.   :2014        Max.   :43.800
 X3.distance.to.the.nearest.MRT.station X4.number.of.convenience.stores
 Min.   :  23.38                        Min.   : 0.000
 1st Qu.: 289.32                        1st Qu.: 1.000
 Median : 492.23                        Median : 4.000
 Mean   :1083.89                        Mean   : 4.094
 3rd Qu.:1454.28                        3rd Qu.: 6.000
 Max.   :6488.02                        Max.   :10.000
  X5.latitude     X6.longitude    Y.house.price.of.unit.area
 Min.   :24.93   Min.   :121.5   Min.   :  7.60
 1st Qu.:24.96   1st Qu.:121.5   1st Qu.: 27.70
 Median :24.97   Median :121.5   Median : 38.45
 Mean   :24.97   Mean   :121.5   Mean   : 37.98
 3rd Qu.:24.98   3rd Qu.:121.5   3rd Qu.: 46.60
 Max.   :25.01   Max.   :121.6   Max.   :117.50
```

```
1  str(df)
```

```
'data.frame':   414 obs. of  8 variables:
 $ No                                 : int  1 2 3 4 5 6 7 8 9 10 ...
 $ X1.transaction.date                : num  2013 2013 2014 2014 2013 ...
 $ X2.house.age                       : num  32 19.5 13.3 13.3 5 7.1 34.5 20.3 31.7 17.9 ...
 $ X3.distance.to.the.nearest.MRT.station: num  84.9 306.6 562 562 390.6 ...
 $ X4.number.of.convenience.stores    : int  10 9 5 5 5 3 7 6 1 3 ...
 $ X5.latitude                        : num  25 25 25 25 25 ...
 $ X6.longitude                       : num  122 122 122 122 122 ...
 $ Y.house.price.of.unit.area         : num  37.9 42.2 47.3 54.8 43.1 32.1 40.3 46.7 18.8 22.1 ...
```

```
1 cor(df)
```

A matrix: 8 × 8 of type dbl

ansaction.date	X2.house.age	X3.distance.to.the.nearest.MRT.station	X4.number.of.conven
-0.048657949	-0.03280811	-0.01357349	
1.000000000	0.01754877	0.06087995	
0.017548767	1.00000000	0.02562205	
0.060879953	0.02562205	1.00000000	
0.009635445	0.04959251	-0.60251914	
0.035057756	0.05441990	-0.59106657	
-0.041081778	-0.04852005	-0.80631677	
0.087490606	-0.21056705	-0.67361286	

Double-click (or enter) to edit

```
1 pairs(df)
```
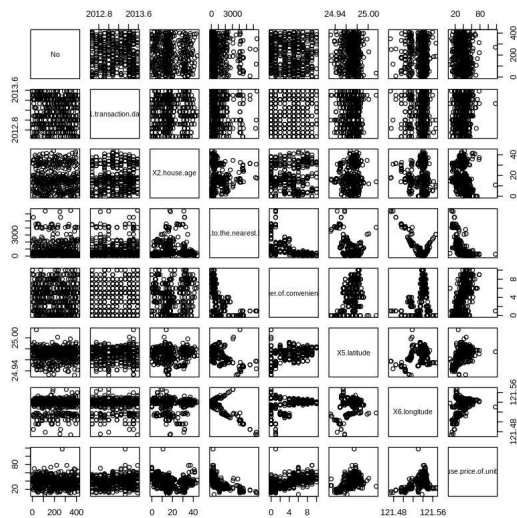


```
1 boxplot(df$Y.house.price.of.unit.area, main="Boxplot of Price")
```
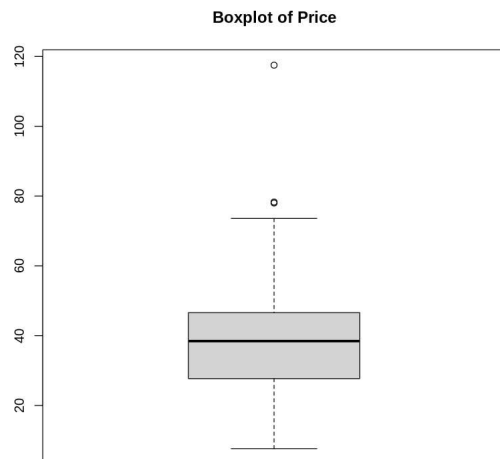

Boxplot of Price

```
1 hist(df$Y.house.price.of.unit.area, main="Histogram of Price", xlab="Price")
```

Histogram of Price
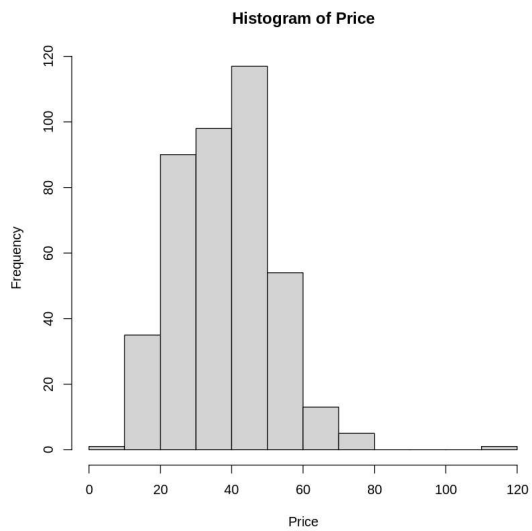

```
1 plot(df$X2.house.age, df$Y.house.price.of.unit.area, main="Scatter Plot of Size vs Price", xlab="Size", ylab="Price")
```

Scatter Plot of Size vs Price
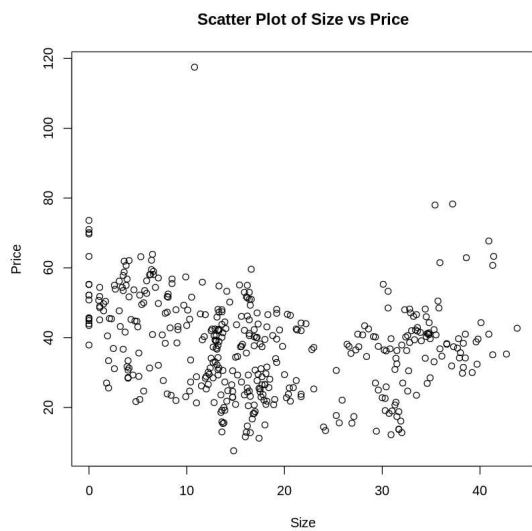
```
1   library(ggplot2)
2   library(dplyr)
3   library(broom)


1   model <- lm(Y.house.price.of.unit.area ~ X1.transaction.date + X2.house.age +
2                  X3.distance.to.the.nearest.MRT.station + X4.number.of.convenience.stores +
3                  X5.latitude + X6.longitude, data = df)
4
5   summary(model)


    Call:
    lm(formula = Y.house.price.of.unit.area ~ X1.transaction.date +
        X2.house.age + X3.distance.to.the.nearest.MRT.station +
    X4.number.of.convenience.stores +
        X5.latitude + X6.longitude, data = df)

    Residuals:
        Min      1Q  Median      3Q     Max
    -35.664  -5.410  -0.966   4.217  75.193

    Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)
    (Intercept)                             -1.444e+04  6.776e+03  -2.131  0.03371
    X1.transaction.date                      5.146e+00  1.557e+00   3.305  0.00103
    X2.house.age                            -2.697e-01  3.853e-02  -7.000 1.06e-11
    X3.distance.to.the.nearest.MRT.station  -4.488e-03  7.180e-04  -6.250 1.04e-09
    X4.number.of.convenience.stores          1.133e+00  1.882e-01   6.023 3.84e-09
    X5.latitude                              2.255e+02  4.457e+01   5.059 6.38e-07
    X6.longitude                            -1.242e+01  4.858e+01  -0.256  0.79829

    (Intercept)                             *
    X1.transaction.date                     **
    X2.house.age                            ***
    X3.distance.to.the.nearest.MRT.station  ***
    X4.number.of.convenience.stores         ***
    X5.latitude                             ***
    X6.longitude
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Residual standard error: 8.858 on 407 degrees of freedom
    Multiple R-squared:  0.5824,    Adjusted R-squared:  0.5762
    F-statistic: 94.59 on 6 and 407 DF,  p-value: < 2.2e-16
```
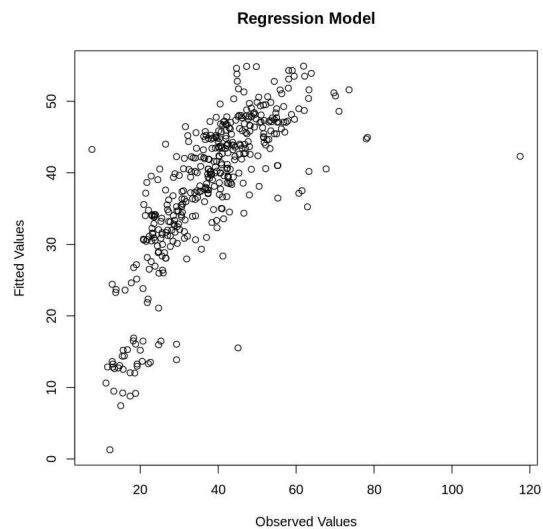
```
1 # Plot the regression model
2 plot(df$Y.house.price.of.unit.area, fitted(model),
3      main="Regression Model", xlab="Observed Values", ylab="Fitted Values")
4
5
```


Regression Model

```
1  ggplot(df, aes(x = Y.house.price.of.unit.area, y = fitted(model))) +
```

```
1  ggplot(df, aes(x = Y.house.price.of.unit.area, y = Fitted(model))) +
2    geom_point() +
3    geom_smooth(method = "lm", se = FALSE, color = "red") +
4    ggtitle("Regression Model") +
5    xlab("Observed Values") +
6    ylab("Fitted Values") +
7    theme_minimal()
8
```

`geom_smooth()` using formula = 'y ~ x'

Regression Model