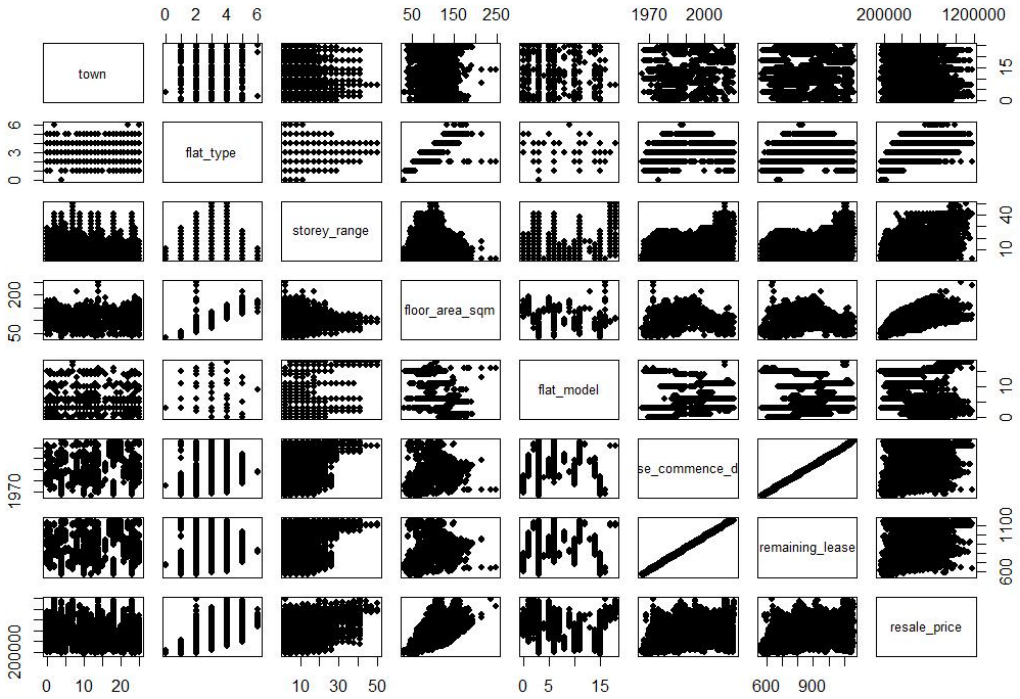
```
> FlatPrice<- read.csv("C:/Users/qt/Desktop/test/Flat prices_after_processing.csv")
> summary(FlatPrice)
       X             month               town          flat_type     storey_range
 Min.   :    0   Length:42070       Min.   : 0.00   Min.   :0.00   Min.   : 2.000
 1st Qu.:10517   Class :character   1st Qu.: 8.00   1st Qu.:2.00   1st Qu.: 5.000
 Median :21035   Mode  :character   Median :14.00   Median :3.00   Median : 8.000
 Mean   :21035                      Mean   :13.83   Mean   :3.15   Mean   : 8.667
 3rd Qu.:31552                      3rd Qu.:21.00   3rd Qu.:4.00   3rd Qu.:11.000
 Max.   :42069                      Max.   :25.00   Max.   :6.00   Max.   :50.000
 floor_area_sqm     flat_model     lease_commence_date remaining_lease   resale_price
 Min.   : 31.00   Min.   : 0.000   Min.   :1966        Min.   : 565.0   Min.   : 160000
 1st Qu.: 82.00   1st Qu.: 3.000   1st Qu.:1984        1st Qu.: 789.0   1st Qu.: 332000
 Median : 96.00   Median : 6.000   Median :1993        Median : 895.0   Median : 410000
 Mean   : 98.11   Mean   : 6.746   Mean   :1993        Mean   : 895.2   Mean   : 442553
 3rd Qu.:113.00   3rd Qu.:10.000   3rd Qu.:2002        3rd Qu.:1001.0   3rd Qu.: 514000
 Max.   :249.00   Max.   :18.000   Max.   :2016        Max.   :1160.0   Max.   :1185000
```

```
> #Step0:A scatterplot between two variables
> pairs(FlatPrice[,3:10],cex=0.8,pch=19)
```



```
> #Step1:OLS
> #OLS1
> m1<-lm(resale_price~town+flat_type+storey_range+
+          floor_area_sqm+flat_model+lease_commence_date+
+          remaining_lease, data=FlatPrice)
> summary(m1)

Call:
lm(formula = resale_price ~ town + flat_type + storey_range +
    floor_area_sqm + flat_model + lease_commence_date + remaining_lease,
    data = FlatPrice)

Residuals:
    Min      1Q  Median      3Q     Max
-298857  -66584  -18043   45410  625755

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1844246.52 1504786.58   1.226  0.22036
town                  -3106.77      64.41 -48.232  < 2e-16 ***
flat_type             25009.25    1742.97  14.349  < 2e-16 ***
storey_range           8313.54      90.38  91.985  < 2e-16 ***
floor_area_sqm         2939.06      66.33  44.310  < 2e-16 ***
flat_model             -831.91     144.24  -5.768 8.09e-09 ***
lease_commence_date    -988.63     784.21  -1.261  0.20744
remaining_lease         199.03      65.25   3.050  0.00229 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 101600 on 42062 degrees of freedom
Multiple R-squared:  0.5625,    Adjusted R-squared:  0.5624
F-statistic:  7726 on 7 and 42062 DF,  p-value: < 2.2e-16
```

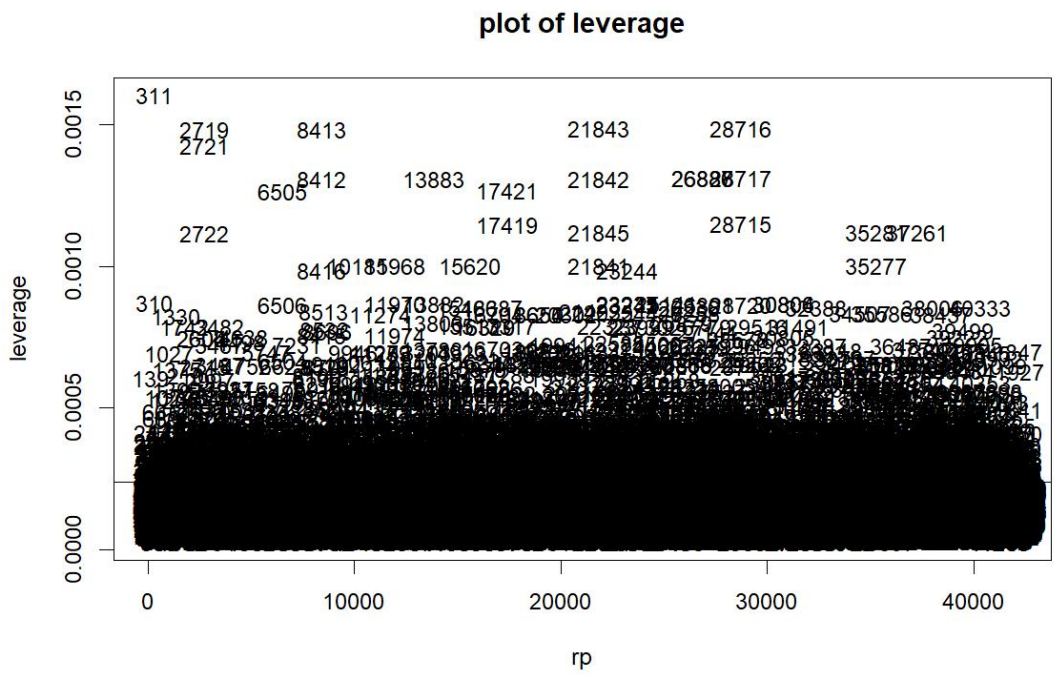
```
> vif(m1)
            town           flat_type        storey_range      floor_area_sqm
        1.075751           10.621640            1.124752           10.516254
      flat_model lease_commence_date     remaining_lease
        1.169346          362.902854          362.881351
```

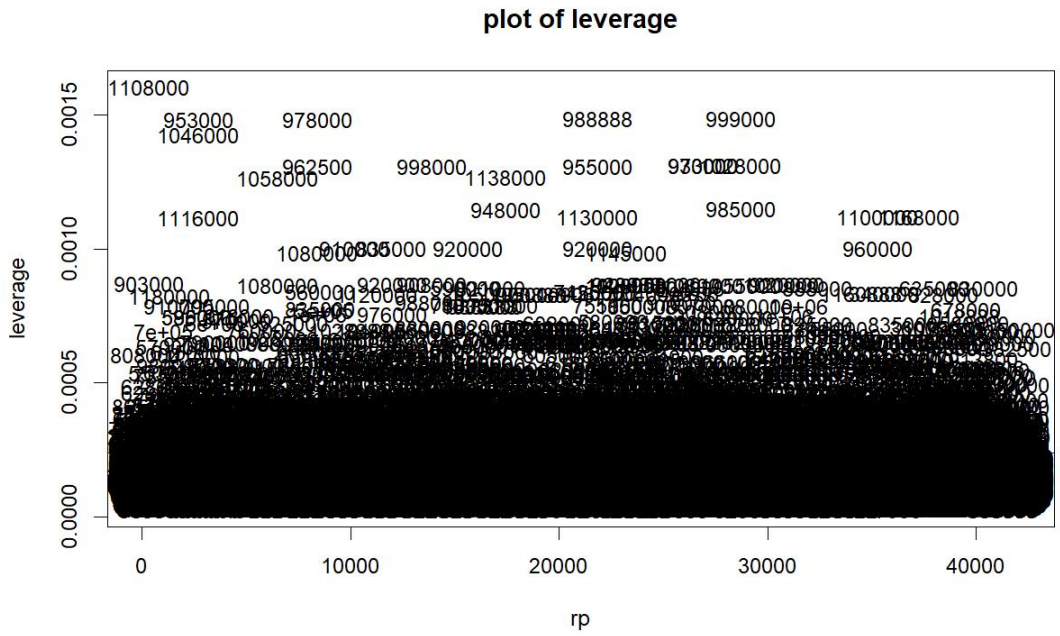
```
> #OLS_new
> m1_new<-lm(resale_price~town+flat_type+storey_range+flat_model+
+              remaining_lease, data=FlatPrice)
> #result
> summary(m1_new)

Call:
lm(formula = resale_price ~ town + flat_type + storey_range +
    flat_model + remaining_lease, data = FlatPrice)

Residuals:
    Min      1Q  Median      3Q     Max
-312237  -68269  -18596   47094  950804

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     31825.96    3649.67    8.72   <2e-16 ***
town            -3120.53      65.90  -47.35   <2e-16 ***
flat_type       97569.61     612.16  159.39   <2e-16 ***
storey_range     7904.01      91.97   85.94   <2e-16 ***
flat_model      -1795.23     145.87  -12.31   <2e-16 ***
remaining_lease   100.68       3.95   25.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 103900 on 42064 degrees of freedom
Multiple R-squared:  0.542,     Adjusted R-squared:  0.542
F-statistic:  9956 on 5 and 42064 DF,  p-value: < 2.2e-16
```

```
> m1$coefficients
       (Intercept)                town           flat_type        storey_range
      1844246.5165           -3106.7664           25009.2539           8313.5399
     floor_area_sqm          flat_model lease_commence_date     remaining_lease
         2939.0644            -831.9114            -988.6295            199.0320
>
```

```
> #anova analysis
> anova(m1_new)
Analysis of Variance Table

Response: resale_price
                  Df     Sum Sq    Mean Sq  F value    Pr(>F)
town               1 6.3691e+12 6.3691e+12   589.95 < 2.2e-16 ***
flat_type          1 4.1707e+14 4.1707e+14 38632.06 < 2.2e-16 ***
storey_range       1 1.0556e+14 1.0556e+14  9777.40 < 2.2e-16 ***
flat_model         1 1.4383e+12 1.4383e+12   133.22 < 2.2e-16 ***
remaining_lease    1 7.0159e+12 7.0159e+12   649.86 < 2.2e-16 ***
Residuals      42064 4.5412e+14 1.0796e+10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #vif multicollinearity treatment
> library(car)
> vif(m1_new)
          town       flat_type    storey_range      flat_model remaining_lease
      1.075635        1.251675        1.112573        1.142580        1.269970
```

```
> #Step 2:the residual plots and the leverage graph
> RP<-FlatPrice$resale_price
> leverage<-hatvalues(m1_new)
> thresshold<-2*5/length(RP)
> rp<-1:length(RP)
> #leverage plot1
> plot(rp,leverage,main='plot of leverage',type='n')
> text(rp,leverage,labels=rp)
```






```
#the residual plots and the leverage graph
par(mfrow=c(2,2))
plot(m1_new)
```