The Data

As a first step I converted the excel sheet, TA, from the `TA data Excel.xlsx` as a csv (a comma separated values) file, `TAdata.csv`.

This file is stored in a directory in my computer at "/Users/harry/Downloads". I will extract this data, analyze and save any output from this directory. So, this is my working directory. I should indicate my working directory to R. This is accomplished by the command `setwd()`. Please note the direction of slashes.

```
setwd("/Users/harry/Downloads")
```

I will ensure this working directory by using the command `getwd()`.

```
getwd()

## [1] "/Users/harry/Downloads"
```

Load the data in to R

```
ta<-read.csv("TAdata.csv",header=TRUE)
```

The above command reads the csv file and stored as an object with (my assigned) name `ta`.

Since R program presents the command prompt, I understand that the file has been read. But to check the contents and other things, I run the following commands:

- `dim()` to know the number of rows and columns
- `str()` str for structure of data; a glimpse of data will be shown
- `head()` First six rows will be displayed
- `tail()` last six rows
- `names()` Names or variables given in the data

```
dim(ta)

## [1] 500  29
```

We have 500 rows and 29 columns

```
str(ta)

## 'data.frame':    500 obs. of  29 variables:
##  $ Close                    : num  0.0614 0.0636 0.0647 0.063 0.0619 ...
##  $ Date                     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Open                     : num  0.0559 0.0614 0.0636 0.0647 0.063 ...
##  $ High                     : num  0.0641 0.0647 0.0652 0.0652 0.0636 ...
##  $ Low                      : num  0.0559 0.0614 0.0636 0.0625 0.0614 ...
##  $ Volume                   : int  1031788800 308160000 133171200 67766400
47894400 58435200 59990400 65289600 32083200 22752000 ...
##  $ volume_adi               : num  3.44e+08 4.46e+08 4.91e+08 4.50e+08 4.2
6e+08 ...
##  $ volume_obv               : num  1.03e+09 1.34e+09 1.47e+09 1.41e+09 1.3
```

```
6e+09 ...
##  $ volume_em                : num  0.00 3.20e-06 1.70e-06 -2.20e-06 -6.30e
-06 -6.20e-06 -8.80e-06 -6.40e-06 -1.40e-06 1.98e-05 ...
##  $ volume_vpt               : num  -1.03e+09 -1.02e+09 1.33e+07 5.74e+05 -
2.56e+06 ...
##  $ volume_nvi               : num  1000 1036 1054 1027 1009 ...
##  $ volatility_bbhi          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ volatility_kch           : num  0.0687 0.0676 0.0671 0.0669 0.0664 ...
##  $ volatility_kcl           : num  0.0522 0.0561 0.0583 0.059 0.0592 ...
##  $ volatility_kcp           : num  0.556 0.651 0.722 0.511 0.379 ...
##  $ volatility_kchi          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ volatility_kcli          : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ volatility_atr           : num  0 0 0 0 0 ...
##  $ trend_ichimoku_b         : num  0.06 0.0603 0.0606 0.0606 0.0606 ...
##  $ trend_adx                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ trend_adx_pos            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ trend_adx_neg            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ trend_cci                : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ trend_visual_ichimoku_a  : num  33.5 33.5 33.5 33.5 33.5 ...
##  $ trend_visual_ichimoku_b  : num  32.8 32.8 32.8 32.8 32.8 ...
##  $ trend_psar_down_indicator: int  0 0 1 0 1 0 0 0 0 0 ...
##  $ others_dr                : num  -99.82 3.57 1.73 -2.54 -1.74 ...
##  $ others_dlr               : num  0 3.51 1.71 -2.58 -1.75 ...
##  $ others_cr                : num  0 3.571 5.358 2.678 0.893 ...

head(ta)

##        Close Date       Open       High        Low     Volume volume_adi
## 1 0.06137834    1 0.05589847 0.06411890 0.05589847 1031788800  343824118
## 2 0.06357029    2 0.06137834 0.06466690 0.06137834  308160000  446465367
## 3 0.06466689    3 0.06357029 0.06521488 0.06357029  133171200  490889792
## 4 0.06302226    4 0.06466685 0.06521483 0.06247427   67766400  450223718
## 5 0.06192631    5 0.06302228 0.06357027 0.06137832   47894400  426276518
## 6 0.06028237    6 0.06192633 0.06192633 0.05973439   58435200  397058918
##   volume_obv volume_em    volume_vpt volume_nvi volatility_bbhi volatility
_kch
## 1 1031788800   0.0e+00 -1029963112.5   1000.000               0     0.0686
8567
## 2 1339948800   3.2e-06 -1018907603.7   1035.712               0     0.0675
8970
## 3 1473120000   1.7e-06    13302287.8   1053.578               0     0.0671
0267
## 4 1405353600  -2.2e-06      573765.2   1026.783               0     0.0669
0476
## 5 1357459200  -6.3e-06    -2556337.6   1008.928               0     0.0664
2052
## 6 1299024000  -6.2e-06    -2384135.7   1008.928               0     0.0658
2371
##   volatility_kcl volatility_kcp volatility_kchi volatility_kcli volatility
_atr
```

```
## 1     0.05224480      0.5555385               0              0
0
## 2     0.05608071      0.6507591               0              0
0
## 3     0.05833362      0.7222300               0              0
0
## 4     0.05895768      0.5114552               0              0
0
## 5     0.05918608      0.3787751               0              0
0
## 6     0.05906436      0.1801964               0              0
0
##   trend_ichimoku_b trend_adx trend_adx_pos trend_adx_neg trend_cci
## 1       0.06000868         0             0             0        NA
## 2       0.06028268         0             0             0        NA
## 3       0.06055667         0             0             0        NA
## 4       0.06055667         0             0             0        NA
## 5       0.06055667         0             0             0        NA
## 6       0.06055667         0             0             0        NA
##   trend_visual_ichimoku_a trend_visual_ichimoku_b trend_psar_down_indicato
r
## 1                33.48542                32.79719
0
## 2                33.48542                32.79719
0
## 3                33.48542                32.79719
1
## 4                33.48542                32.79719
0
## 5                33.48542                32.79719
1
## 6                33.48542                32.79719
0
##    others_dr others_dlr  others_cr
## 1 -99.818167   0.000000  0.0000000
## 2   3.571217   3.508927  3.5712166
## 3   1.725018   1.710308  5.3578385
## 4  -2.543239  -2.576139  2.6783366
## 5  -1.738984  -1.754282  0.8927768
## 6  -2.654667  -2.690539 -1.7855901
```

The `tail(ta)` command is not run.

```
names(ta)
```

```
##  [1] "Close"                    "Date"
##  [3] "Open"                     "High"
##  [5] "Low"                      "Volume"
##  [7] "volume_adi"               "volume_obv"
##  [9] "volume_em"                "volume_vpt"
```

```
## [11] "volume_nvi"                "volatility_bbhi"
## [13] "volatility_kch"            "volatility_kcl"
## [15] "volatility_kcp"            "volatility_kchi"
## [17] "volatility_kcli"           "volatility_atr"
## [19] "trend_ichimoku_b"          "trend_adx"
## [21] "trend_adx_pos"             "trend_adx_neg"
## [23] "trend_cci"                 "trend_visual_ichimoku_a"
## [25] "trend_visual_ichimoku_b"   "trend_psar_down_indicator"
## [27] "others_dr"                 "others_dlr"
## [29] "others_cr"
```

I compute summary statistics (namely, Minimum, First quartile, Median, Mean, Third quartile and maximum for all variables.)

```
summary(ta)

##      Close              Date            Open              High
##  Min.   :0.05699   Min.   :  1.0   Min.   :0.05590   Min.   :0.05809
##  1st Qu.:0.07015   1st Qu.:125.8   1st Qu.:0.07015   1st Qu.:0.07090
##  Median :0.17975   Median :250.5   Median :0.17756   Median :0.18496
##  Mean   :0.16062   Mean   :250.5   Mean   :0.16031   Mean   :0.16407
##  3rd Qu.:0.22853   3rd Qu.:375.2   3rd Qu.:0.22921   3rd Qu.:0.23250
##  Max.   :0.34635   Max.   :500.0   Max.   :0.34635   Max.   :0.34745
##
##       Low              Volume            volume_adi          volume_obv
##  Min.   :0.05590   Min.   :2.304e+06   Min.   :3.291e+08   Min.   :1.032e+
09
##  1st Qu.:0.06946   1st Qu.:2.336e+07   1st Qu.:4.550e+08   1st Qu.:1.637e+
09
##  Median :0.17345   Median :5.756e+07   Median :1.782e+09   Median :4.075e+
09
##  Mean   :0.15696   Mean   :6.974e+07   Mean   :1.658e+09   Mean   :4.093e+
09
##  3rd Qu.:0.22428   3rd Qu.:9.419e+07   3rd Qu.:2.657e+09   3rd Qu.:5.912e+
09
##  Max.   :0.32224   Max.   :1.032e+09   Max.   :3.342e+09   Max.   :8.497e+
09
##
##    volume_em            volume_vpt           volume_nvi     volatility_bbh
i
##  Min.   :-3.662e-03   Min.   :-1.030e+09   Min.   : 961.3   Min.   :0.000
##  1st Qu.:-1.980e-05   1st Qu.:-8.647e+05   1st Qu.:1054.8   1st Qu.:0.000
##  Median : 2.100e-06   Median : 1.132e+05   Median :1187.1   Median :0.000
##  Mean   :-5.368e-06   Mean   :-3.286e+06   Mean   :1211.6   Mean   :0.096
##  3rd Qu.: 2.892e-05   3rd Qu.: 2.663e+06   3rd Qu.:1299.2   3rd Qu.:0.000
##  Max.   : 5.968e-04   Max.   : 6.678e+07   Max.   :1666.6   Max.   :1.000
##
##  volatility_kch    volatility_kcl    volatility_kcp    volatility_kchi
##  Min.   :0.06138   Min.   :0.05224   Min.   :-1.5487   Min.   :0.000
##  1st Qu.:0.07160   1st Qu.:0.06836   1st Qu.: 0.2332   1st Qu.:0.000
```

```
##  Median :0.17961   Median :0.16366   Median : 0.6564   Median :0.000
##  Mean   :0.16585   Mean   :0.15167   Mean   : 0.6857   Mean   :0.318
##  3rd Qu.:0.23726   3rd Qu.:0.21973   3rd Qu.: 1.1436   3rd Qu.:1.000
##  Max.   :0.33897   Max.   :0.30170   Max.   : 2.6219   Max.   :1.000
##
##  volatility_kcli volatility_atr     trend_ichimoku_b    trend_adx
##  Min.   :0.000   Min.   :0.000000   Min.   :0.06001   Min.   : 0.00
##  1st Qu.:0.000   1st Qu.:0.002358   1st Qu.:0.06741   1st Qu.:18.44
##  Median :0.000   Median :0.007330   Median :0.14468   Median :23.99
##  Mean   :0.156   Mean   :0.007151   Mean   :0.15390   Mean   :25.83
##  3rd Qu.:0.000   3rd Qu.:0.009368   3rd Qu.:0.22990   3rd Qu.:33.13
##  Max.   :1.000   Max.   :0.031000   Max.   :0.27648   Max.   :56.76
##
##  trend_adx_pos   trend_adx_neg     trend_cci       trend_visual_ichimoku_a
##  Min.   : 0.00   Min.   : 0.00   Min.   :-271.47   Min.   : 0.06206
##  1st Qu.:21.45   1st Qu.:16.52   1st Qu.: -40.68   1st Qu.: 0.07227
##  Median :27.13   Median :22.75   Median :  53.29   Median : 0.18044
##  Mean   :28.08   Mean   :22.64   Mean   :  40.04   Mean   : 1.98202
##  3rd Qu.:35.43   3rd Qu.:29.53   3rd Qu.: 112.22   3rd Qu.: 0.23270
##  Max.   :55.86   Max.   :48.17   Max.   : 327.43   Max.   :33.48542
##                                  NA's   :19        NA's   :25
##  trend_visual_ichimoku_b trend_psar_down_indicator   others_dr
##  Min.   : 0.06001        Min.   :0.000             Min.   :-99.8182
##  1st Qu.: 0.06741        1st Qu.:0.000             1st Qu.: -1.2482
##  Median : 0.14468        Median :0.000             Median :  0.0000
##  Mean   : 1.84746        Mean   :0.042             Mean   :  0.1594
##  3rd Qu.: 0.23757        3rd Qu.:0.000             3rd Qu.:  1.9467
##  Max.   :32.79719        Max.   :1.000             Max.   : 17.9689
##
##    others_dlr         others_cr
##  Min.   :-35.8332   Min.   : -7.142
##  1st Qu.: -1.2378   1st Qu.: 14.286
##  Median :  0.0000   Median :192.858
##  Mean   :  0.2944   Mean   :161.686
##  3rd Qu.:  1.9280   3rd Qu.:272.322
##  Max.   : 16.5251   Max.   :464.287
##
```

After inspection of these, we found that the variables, `volatility_bbhi, volatility_kchi, volatility_kcli and trend_psar_down_indicator` have zero values for the first three and fifth measures! These variables must have only 0 and 1 as their values. So these must be treated as *categorical* (or *factor*) variables!

We do this now.

```
ta$volatility_bbhi<-as.factor(ta$volatility_bbhi)
ta$volatility_kchi<-as.factor(ta$volatility_kchi)
ta$volatility_kcli<-as.factor(ta$volatility_kcli)
ta$trend_psar_down_indicator<-as.factor(ta$trend_psar_down_indicator)
```

The Model 1

The variable `volume` is taken as the dependent variable and the remaining 27 variables (excluding Date) are treated as independent variables. To avoid using the data object name (`ta`) repeatedly, we attach the file to the path.

```
attach(ta)
```

I recall the command for multiple regression. If the data file (say `df`)has variable y,x1,x2,x3,x4 and you want to regress y *only* on x1, x2 and x4 (i.e to exclude x3), run this lm(y~.-x3-y, data=df). Here "." represents *all* variables of the data file.

For our problem, we run

```
ta_reg<-lm(Volume~. -Date -Volume, data=ta )
```

The object `ta_reg` contains the results of the analysis of `lm()`, i.e regression. To know the components stored in this object, we use

```
names(ta_reg)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "na.action"     "contrasts"     "xlevels"       "call"
## [13] "terms"         "model"
```

To know the coefficients of each independent variables immediately, you just issue

```
ta_reg
```

```
##
## Call:
## lm(formula = Volume ~ . - Date - Volume, data = ta)
##
## Coefficients:
##                (Intercept)                      Close
##                 -4.319e+07                 -1.162e+09
##                       Open                       High
##                  1.758e+09                  5.370e+09
##                        Low                 volume_adi
##                 -6.210e+09                 -2.030e-02
##                 volume_obv                  volume_em
##                  1.846e-03                  8.309e+10
##                 volume_vpt                 volume_nvi
##                 -9.845e-01                 -3.861e+04
##           volatility_bbhi1              volatility_kch
##                  1.662e+07                  3.363e+08
##             volatility_kcl              volatility_kcp
##                  3.912e+08                  1.341e+07
##           volatility_kchi1           volatility_kcli1
##                  1.434e+07                  7.604e+06
```

```
##             volatility_atr            trend_ichimoku_b
##                 -1.231e+09                  -2.083e+07
##                  trend_adx               trend_adx_pos
##                  1.325e+06                   4.954e+05
##               trend_adx_neg                   trend_cci
##                  6.900e+05                  -8.209e+04
##    trend_visual_ichimoku_a     trend_visual_ichimoku_b
##                 -2.165e+08                   2.217e+08
## trend_psar_down_indicator1                   others_dr
##                  3.581e+06                  -3.002e+07
##                 others_dlr                   others_cr
##                  3.301e+07                          NA
```

To get these and some more characteristics. run

```
summary(ta_reg)
```

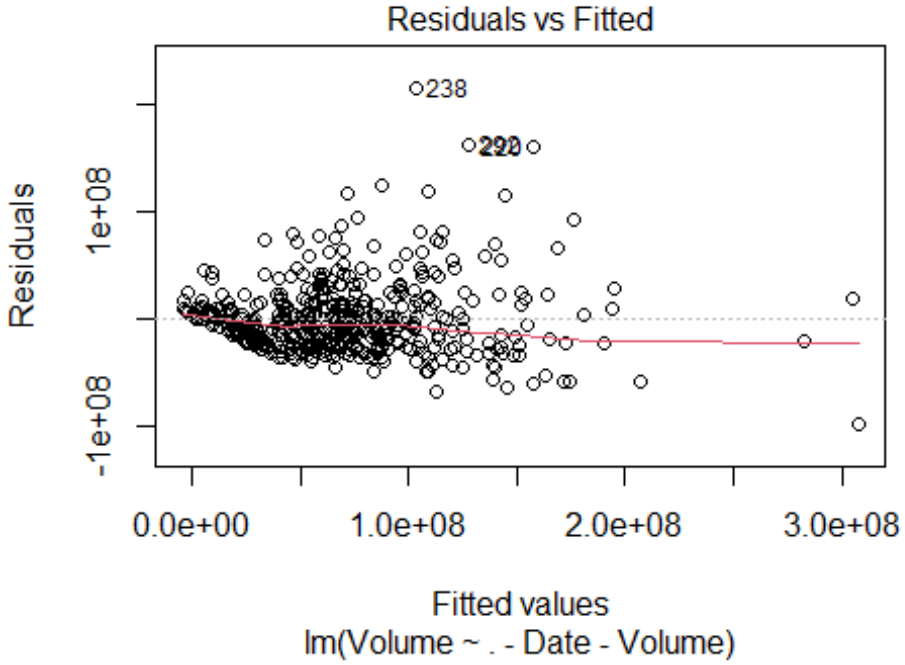
```
##
## Call:
## lm(formula = Volume ~ . - Date - Volume, data = ta)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -97351716 -22011279  -5621992  12556477 214278547
##
## Coefficients: (1 not defined because of singularities)
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -4.319e+07  2.846e+07  -1.517  0.12989
## Close                      -1.162e+09  9.420e+08  -1.233  0.21821
## Open                        1.758e+09  9.122e+08   1.927  0.05468 .
## High                        5.370e+09  8.661e+08   6.200 1.33e-09 ***
## Low                        -6.210e+09  8.081e+08  -7.684 1.06e-13 ***
## volume_adi                 -2.030e-02  2.016e-02  -1.007  0.31444
## volume_obv                  1.846e-03  5.487e-03   0.336  0.73667
## volume_em                   8.309e+10  2.568e+10   3.236  0.00131 **
## volume_vpt                 -9.845e-01  3.604e-01  -2.731  0.00657 **
## volume_nvi                 -3.861e+04  2.520e+04  -1.532  0.12630
## volatility_bbhi1            1.662e+07  7.503e+06   2.215  0.02727 *
## volatility_kch              3.363e+08  7.985e+08   0.421  0.67381
## volatility_kcl              3.912e+08  8.454e+08   0.463  0.64382
## volatility_kcp              1.341e+07  8.846e+06   1.516  0.13016
## volatility_kchi1            1.434e+07  6.352e+06   2.257  0.02449 *
## volatility_kcli1            7.604e+06  7.483e+06   1.016  0.31013
## volatility_atr             -1.231e+09  2.129e+09  -0.578  0.56335
## trend_ichimoku_b           -2.083e+07  1.371e+08  -0.152  0.87932
## trend_adx                   1.325e+06  2.973e+05   4.456 1.07e-05 ***
## trend_adx_pos               4.954e+05  4.318e+05   1.147  0.25195
## trend_adx_neg               6.900e+05  5.962e+05   1.157  0.24780
## trend_cci                  -8.209e+04  4.307e+04  -1.906  0.05731 .
## trend_visual_ichimoku_a    -2.165e+08  1.420e+08  -1.525  0.12791
```

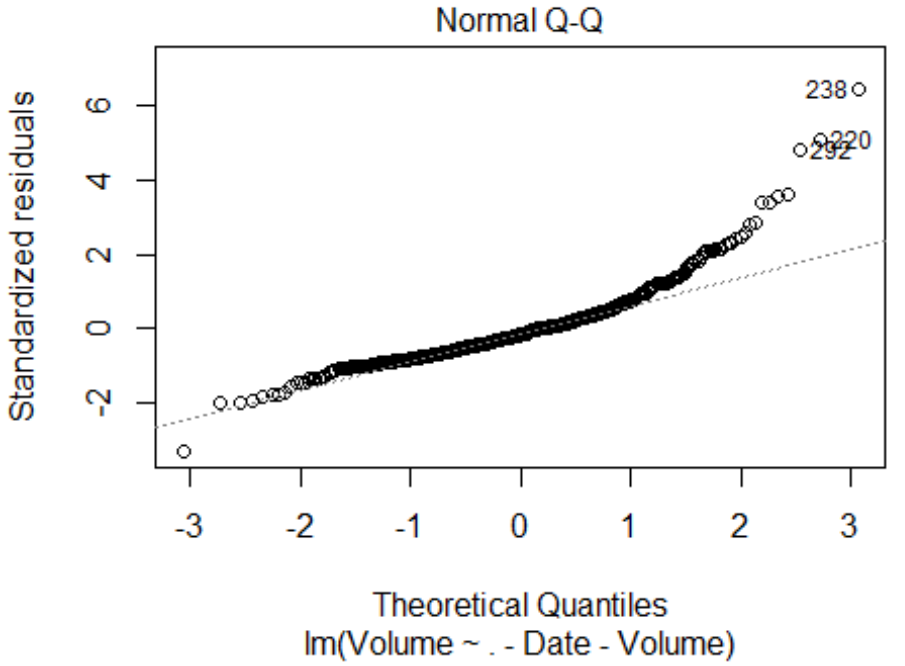
```
## trend_visual_ichimoku_b     2.217e+08  1.450e+08   1.529  0.12703
## trend_psar_down_indicator1  3.581e+06  8.849e+06   0.405  0.68592
## others_dr                  -3.002e+07  1.563e+07  -1.920  0.05551 .
## others_dlr                  3.301e+07  1.587e+07   2.080  0.03809 *
## others_cr                         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34660000 on 429 degrees of freedom
##   (44 observations deleted due to missingness)
## Multiple R-squared:  0.6373, Adjusted R-squared:  0.6154
## F-statistic:    29 on 26 and 429 DF,  p-value: < 2.2e-16
```

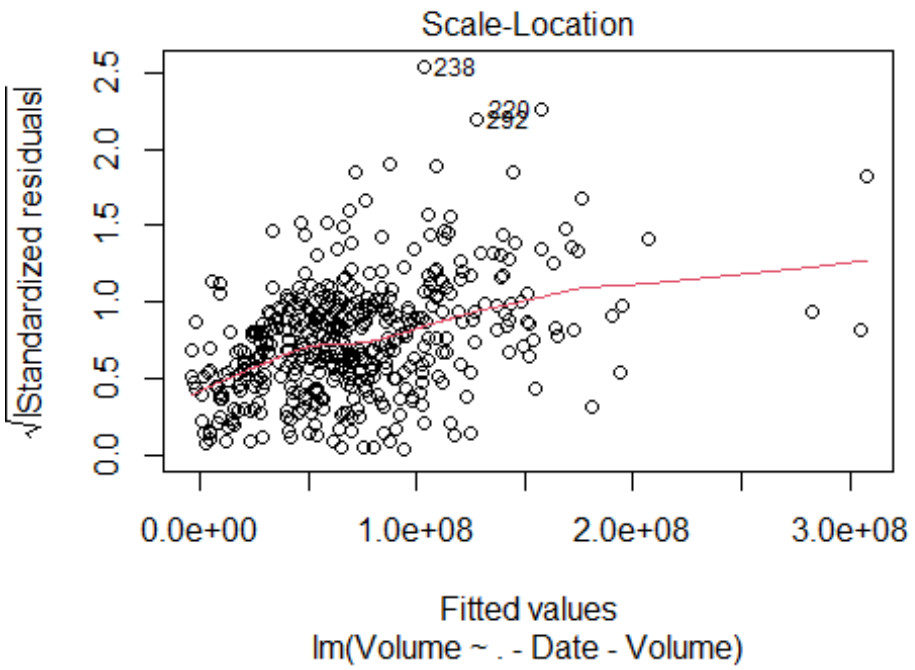
Residual Analysis:

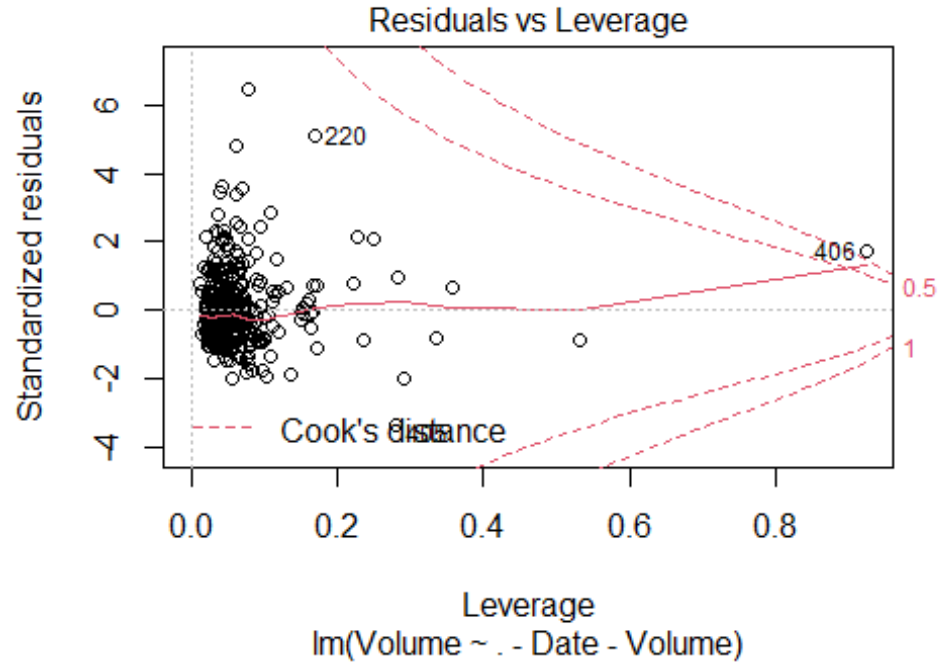
We now obtain some diagnostic plots.

```
plot(ta_reg)
```


Residuals vs Fitted
238
220
292
Residuals
1e+08
-1e+08
0.0e+00
1.0e+08
2.0e+08
3.0e+08
Fitted values
lm(Volume ~ . - Date - Volume)


Normal Q-Q
238
220
292
Standardized residuals
6
4
2
0
-2
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(Volume ~ . - Date - Volume)





Remark

- 44 records are missing for some variables. In the above analysis 44 records are not included!

- Adjusted R^2 is 0.6153. Hence 61.53 percent of variation in response variable `Volume` is explained.
  This is good.
- The test for the multiple regression by F statistic (a value of 29 on degrees of freedom 26 amd 429) is significant. We conclude that the model is good.
- But only eight predictors are significant and others are not.
- The residual plot shows that the residuals (=Response-predicted) appears to be random. But there are some outliers; record numbers:220, 238 and 292.
- The same outliers are shown in *Normal Q-Q* plot! This plot is not aligned with the straight line. This may be due to the presence of outliers.
- These three outliers are also indicated as outliers,in the plot entitled "Scale-Location", which is the plot of $\sqrt{\text{Standardised Residual}}$ against Fitted values.

Thus it is advisable to redo the analysis by

- removing the outliers, and
- considering only the eight (statistically significant) predictors.

Second Model (Model 2)

Let us just check the data for the records 220, 238 and 292. we create new data set `ta_out` by considering these rerods

```
ta_rev<-ta[-c(220,238,292),]
```

We perform multiple regression y using this dataset by considering only those eight (statistically significant) predictors, viz, `High, Low, volume_em, volume_vpt, volatility_bbhi, volatility_kchi, trend_adx`, and `others_dlr`. These variables occur in the list of names of the dataset `ta_rev` at 4,5,9,10,12,16,20 and 28 places. We use these position numbers. (Note the response variable `Volume` occurs at 6th position.)

```
ta_rev1<-ta_rev[,c(4,5,6,9,10,12,16,20,28)]
```

The regression results of model 2

Now we perform the analysis.

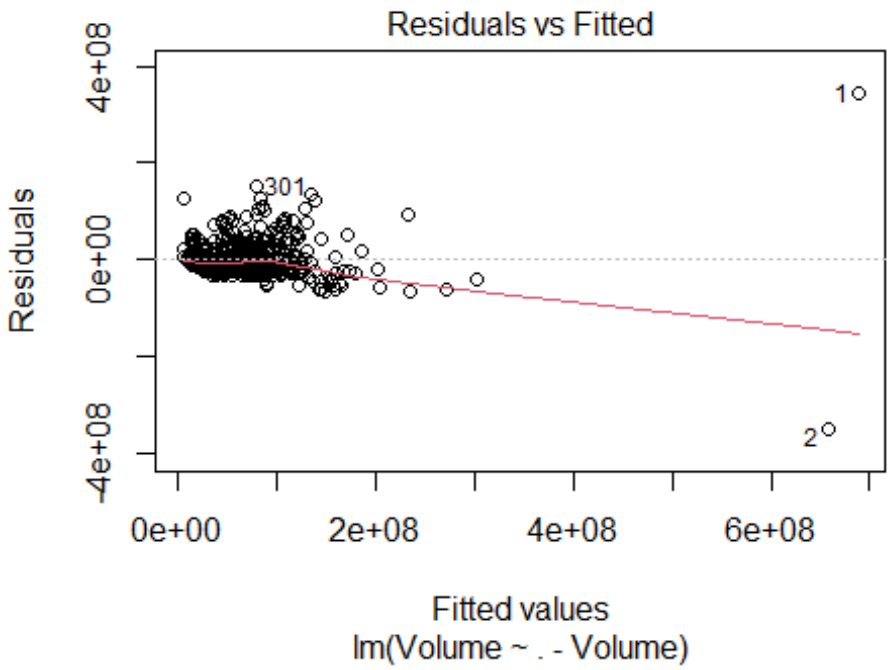
```
ta_rev1_reg<-lm(Volume~.  -Volume, data=ta_rev1 )
```

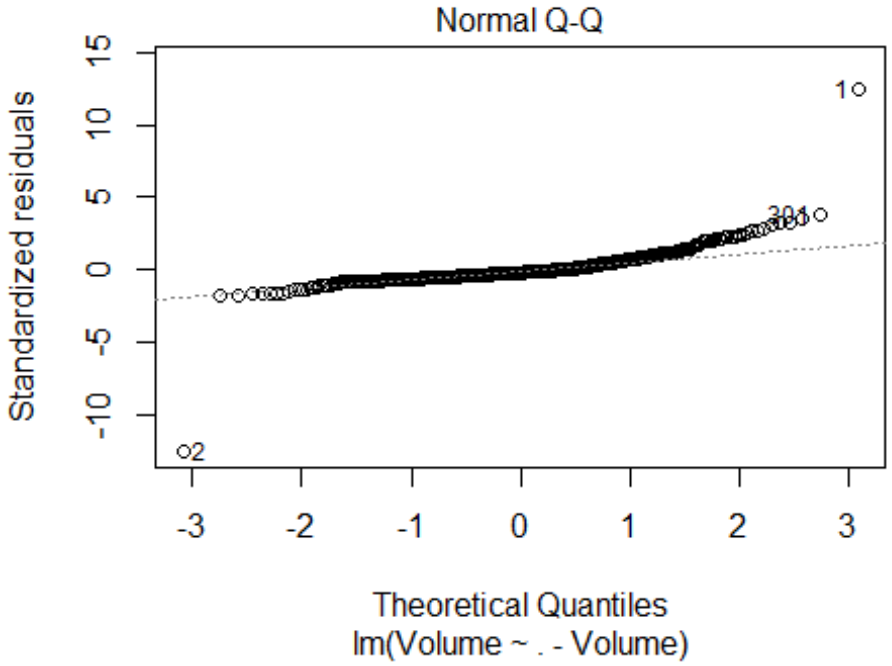
The summary

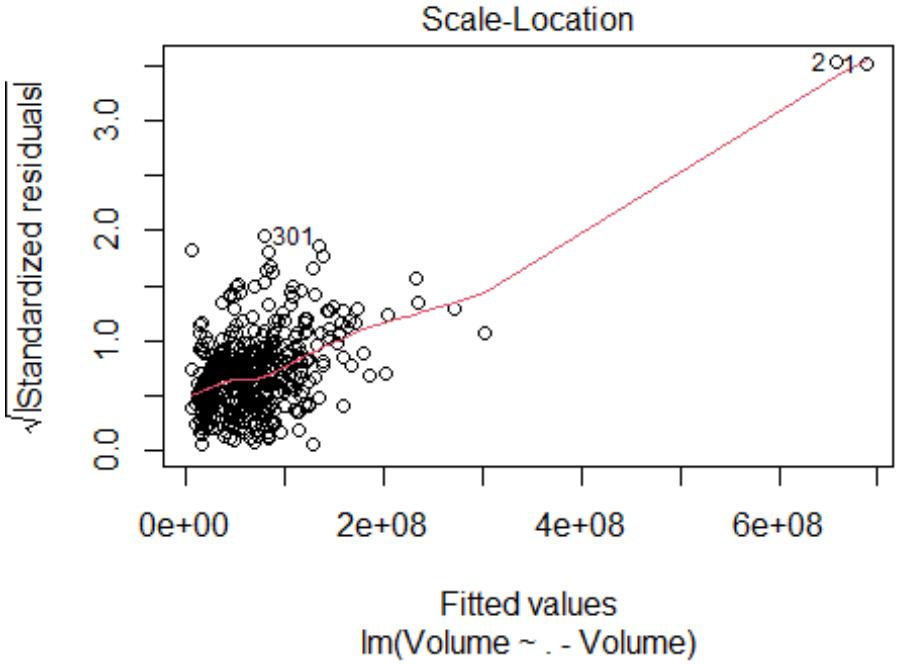
```
summary(ta_rev1_reg)
```

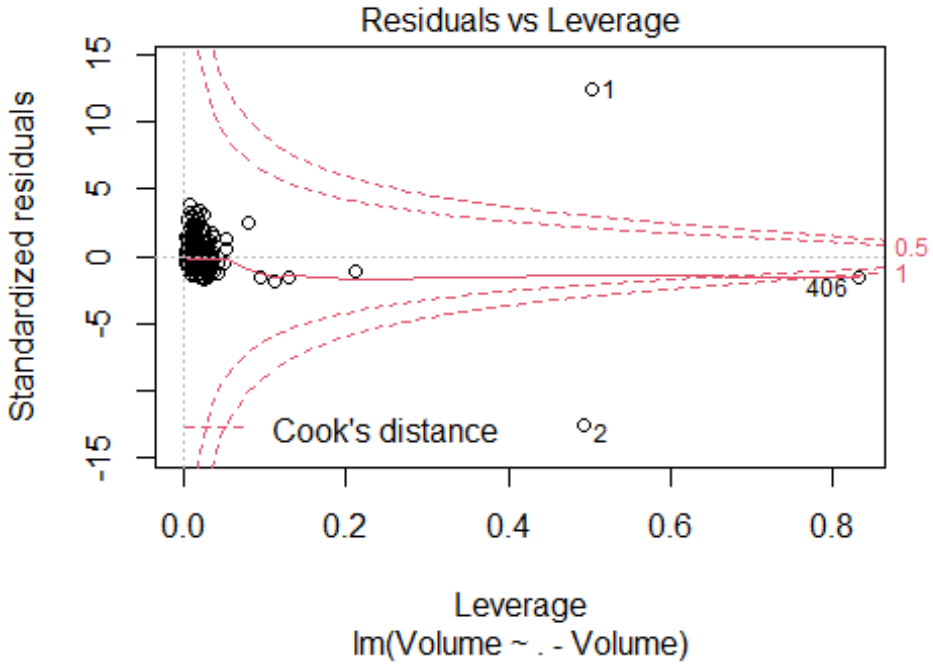
```
##
## Call:
## lm(formula = Volume ~ . - Volume, data = ta_rev1)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -349541578  -19505345   -7701434   10966916  343355616
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.001e+06  5.266e+06  -0.380 0.704056
## High              5.386e+09  3.505e+08  15.365  < 2e-16 ***
## Low              -5.305e+09  3.686e+08 -14.392  < 2e-16 ***
## volume_em         8.578e+10  1.521e+10   5.639  2.9e-08 ***
## volume_vpt       -6.230e-01  2.738e-02 -22.749  < 2e-16 ***
## volatility_bbhi1  5.304e+06  7.163e+06   0.741 0.459348
## volatility_kchi1  8.559e+06  4.453e+06   1.922 0.055215 .
## trend_adx         5.648e+05  1.515e+05   3.727 0.000216 ***
## others_dlr        5.693e+05  7.238e+05   0.787 0.431954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39200000 on 488 degrees of freedom
## Multiple R-squared:  0.679,  Adjusted R-squared:  0.6737
## F-statistic:   129 on 8 and 488 DF,  p-value: < 2.2e-16
```

```
#par(mfrow=c(2,2))
plot(ta_rev1_reg)
```


Residuals vs Fitted
Residuals
4e+08
0e+00
-4e+08
0e+00
2e+08
4e+08
6e+08
1
301
2
Fitted values
lm(Volume ~ . - Volume)


Normal Q-Q
Standardized residuals
15
10
5
0
-5
-10
-3
-2
-1
0
1
2
3
1
301
2
Theoretical Quantiles
lm(Volume ~ . - Volume)






```
#par(mfrow=c(1,1))
```

Observations

1. From summary statistics on the regression, we find that the variables `High, Low, volume_em, volume_vpt and trend_adx are highly significant. (These occur in the dataset at 1, 2, 4, 5 and 8.)

2. The F statistic for the regression is 129 on degrees of freedom (8,488) which is significant. Thus the model is good.

3. The adjusted R^2 is 0.6737, That is 67.37% of variations explained. It is an improvement from the earlier analysis.

4. From the plots of the regression lines we note that there are still some outliers; records `1, 2,301`,and `406`. (Note that these records refere to the new dataset `ta_rev1`.)

We shall perform regression analysis after removing these records from `ta_rev1` on the selected variables.

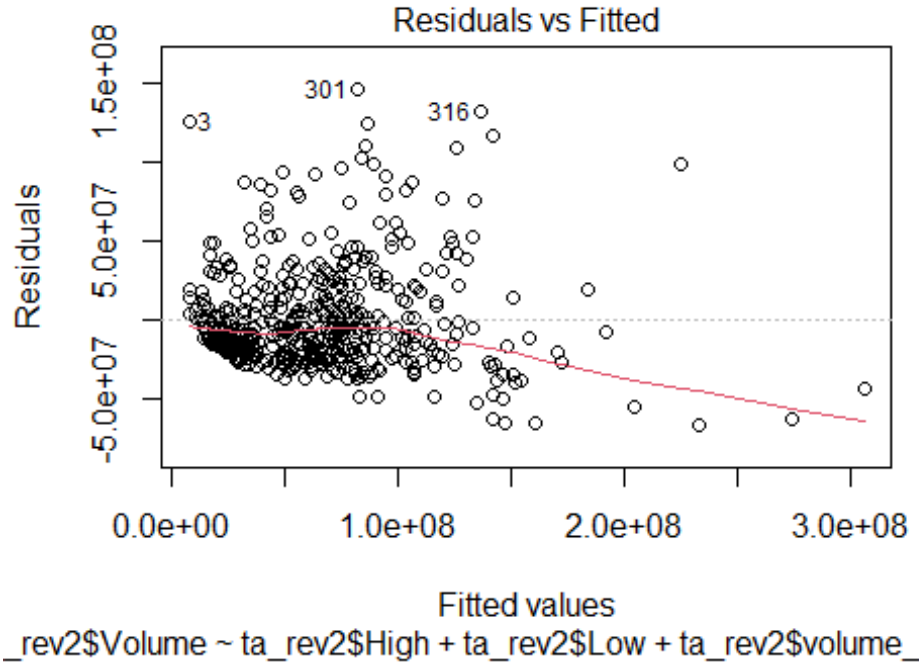
Third Model (Model 3)

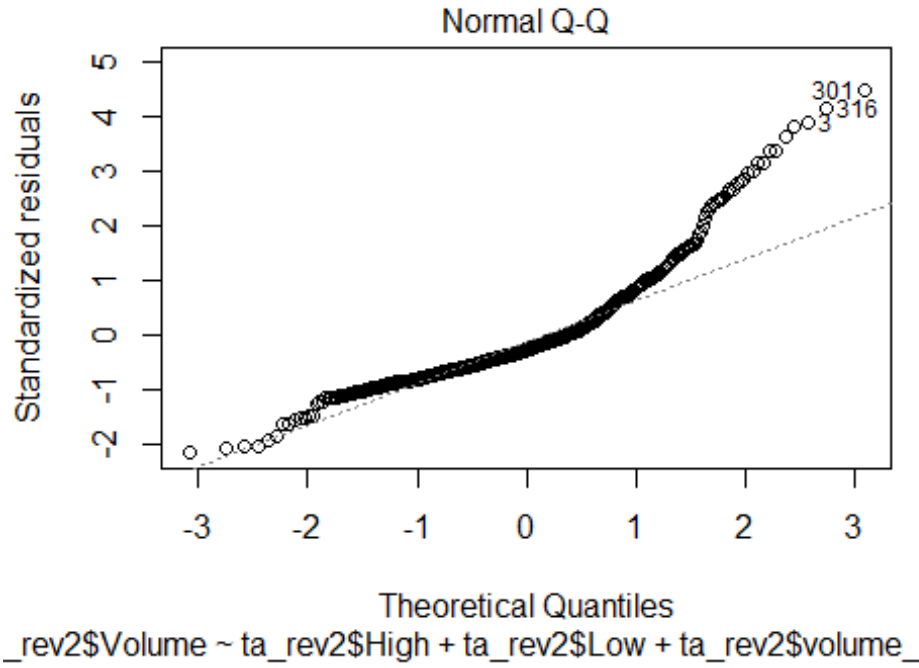
```
ta_rev2<-ta_rev1[-c(1,2,301,406),c(1,2,3,4,5,8)]
ta_rev2_reg<-lm(ta_rev2$Volume~ta_rev2$High+ta_rev2$Low+ta_rev2$volume_em+ ta
_rev2$volume_vpt+ta_rev2$trend_adx, data=ta_rev2)
summary(ta_rev2_reg)
```

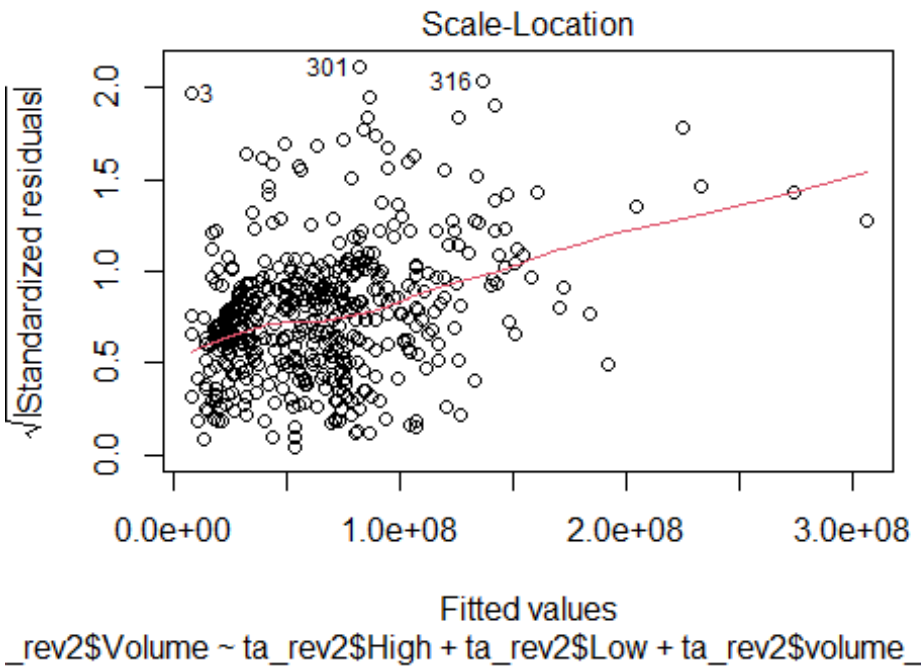
```
##
## Call:
## lm(formula = ta_rev2$Volume ~ ta_rev2$High + ta_rev2$Low + ta_rev2$volume_
em +
##     ta_rev2$volume_vpt + ta_rev2$trend_adx, data = ta_rev2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -66597955 -20478248  -8592628  12692638 146010939
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.732e+05  4.360e+06  -0.040   0.9683
## ta_rev2$High        5.312e+09  2.959e+08  17.954  < 2e-16 ***
## ta_rev2$Low        -5.228e+09  3.110e+08 -16.812  < 2e-16 ***
## ta_rev2$volume_em   9.265e+10  1.273e+10   7.276 1.38e-12 ***
## ta_rev2$volume_vpt -4.850e-01  2.863e-01  -1.694   0.0909 .
## ta_rev2$trend_adx   6.220e+05  1.248e+05   4.982 8.76e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32800000 on 487 degrees of freedom
```

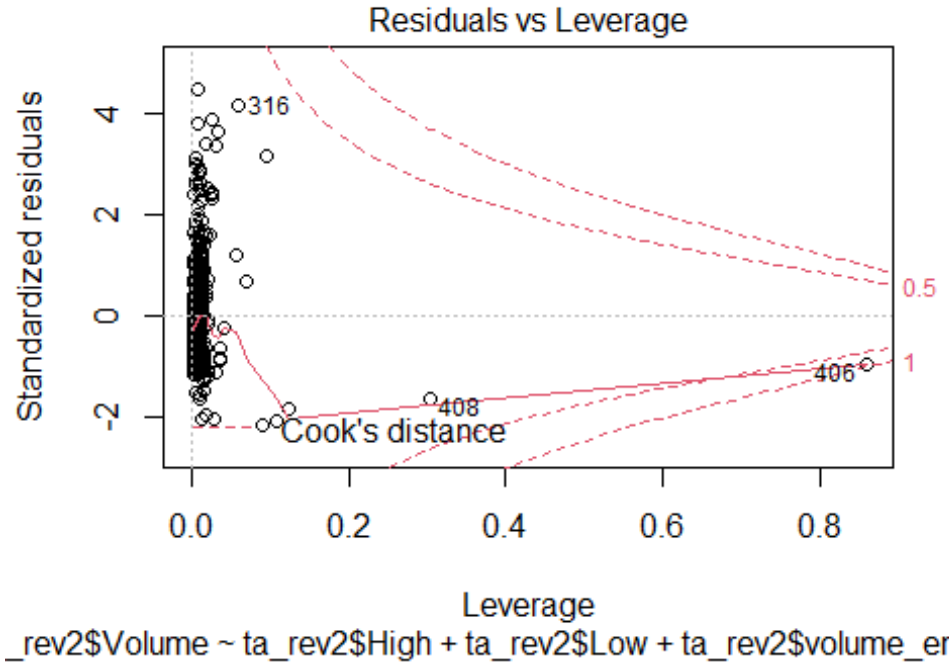
```
## Multiple R-squared:  0.6039, Adjusted R-squared:  0.5998
## F-statistic: 148.5 on 5 and 487 DF,  p-value: < 2.2e-16

plot(ta_rev2_reg)
```






Scale-Location
√|Standardized residuals|
301
316
3
0.0
0.5
1.0
1.5
2.0
0.0e+00
1.0e+08
2.0e+08
3.0e+08
Fitted values
_rev2$Volume ~ ta_rev2$High + ta_rev2$Low + ta_rev2$volume_em


Residuals vs Leverage
Standardized residuals
316
406
408
Cook's distance
0.5
1
-2
0
2
4
0.0
0.2
0.4
0.6
0.8
Leverage
_rev2$Volume ~ ta_rev2$High + ta_rev2$Low + ta_rev2$volume_em

Some observations In the analysis of model 3

1. Except `volume_vpt`, all other variables are significant.

2. Adjusted R^2 is 0.5998. That is only 60% of variations is explained which is *lower than* the previous two models.

3. For this model, the diagnostic plots show more outliers and the Q-Q plot shows larger deviation from the straight lines (than the one in the previous models).

So we conclude that the model 2, `ta_rev1_reg`, can be taken as a good model.

The model

The regression equation for model 2 is

$$\text{Volume} = (-2.001 \times 10^6) + (5.386 \times 10^9)\text{High} - (5.305 \times 10^9)\text{Low}$$

$$+(8.578 \times 10^{10})\text{volume\_em} - (0.6230)\text{volume\_vpt} + (5.648 \times 10^5)\text{trend\_adx}$$

```
save(ta_rev1,file="ta_rev1.RData")
```