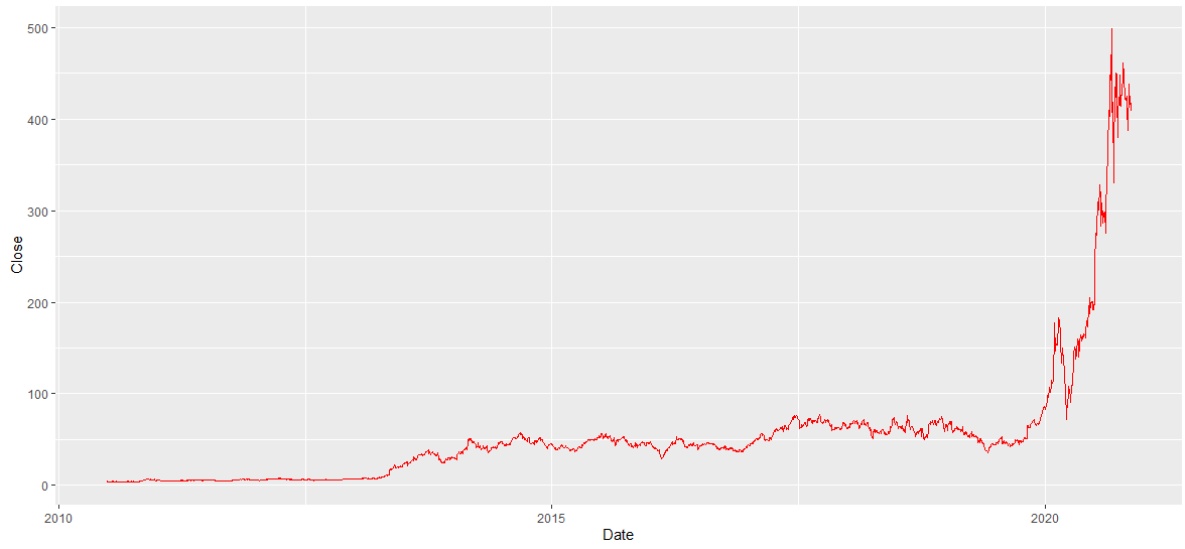
Multiple Linear Regression Full Model

First plot the price vs date

```
> rm(list = ls())
> setwd("F:\\FA2020\\VE406\\proj")
> ## read the data from the website
> price = read.csv("TSLA_full.csv", header = T)
> price$Date = as.Date(price$Date, format = "%Y-%m-%d")
> ## plot prices and volume
> p1 = ggplot(price)+
+   geom_line(aes(x = Date, y = Close), color = "red",
show.legend = T)
> p1
> p2 = ggplot(data = price) + geom_line(aes(x = Date, y =
Volume), color = "green")
```




```
> lm_full = lm(Close ~ Date+Open+High+Low+Volume, data = price)
> summary(lm_full)

Call:
lm(formula = Close ~ Date + Open + High + Low + Volume, data =
price)

Residuals:
     Min       1Q   Median       3Q      Max
-30.9131  -0.1927  -0.0344   0.2490  24.7574

Coefficients:
```

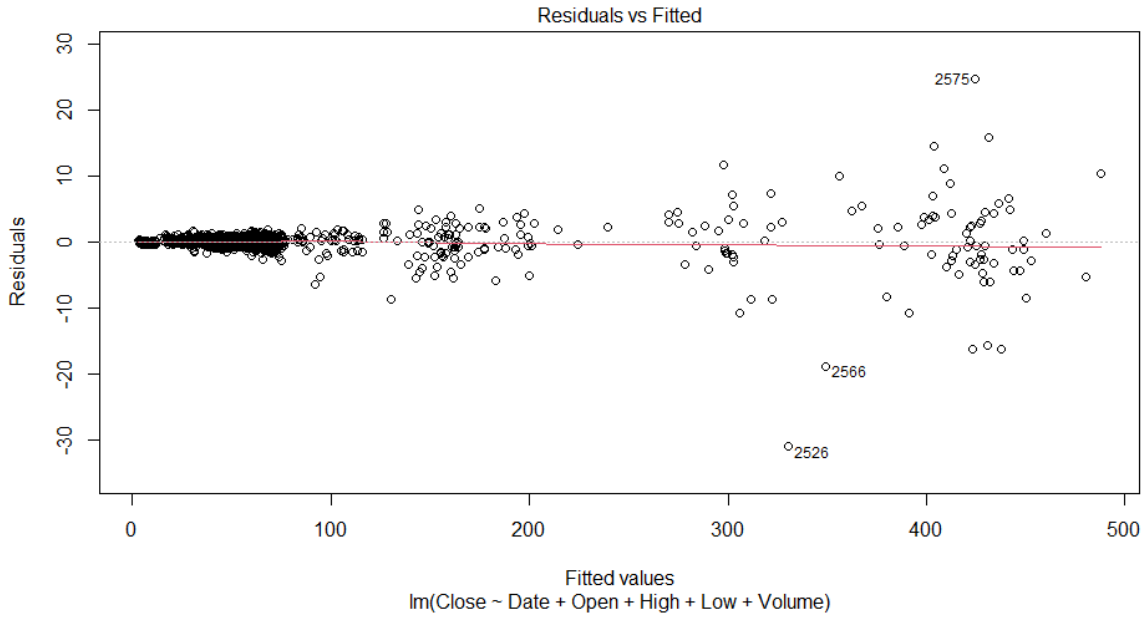
```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.561e+00  6.408e-01  -2.437  0.01489 *
Date         1.093e-04  4.053e-05   2.696  0.00705 **
Open        -4.920e-01  1.592e-02 -30.901  < 2e-16 ***
High         8.839e-01  1.469e-02  60.165  < 2e-16 ***
Low          6.022e-01  1.362e-02  44.221  < 2e-16 ***
Volume      -6.922e-09  1.407e-09  -4.918 9.28e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.557 on 2608 degrees of freedom
Multiple R-squared:  0.9995,    Adjusted R-squared:  0.9995
F-statistic: 1.12e+06 on 5 and 2608 DF,  p-value: < 2.2e-16
```

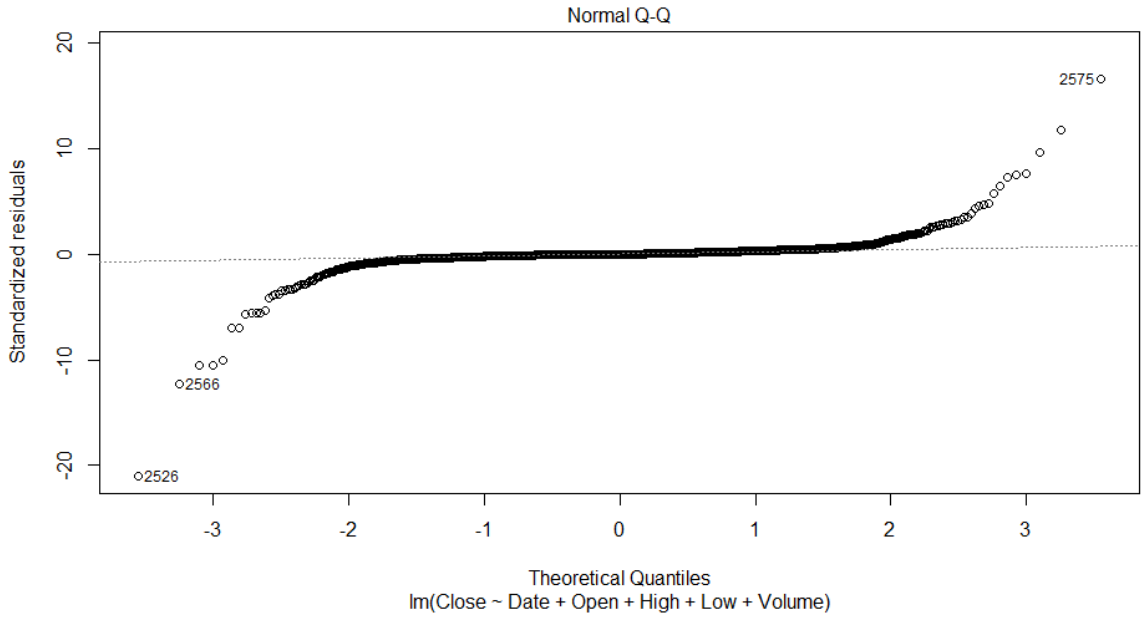
From the full model we can see that the regressors are all significant by T test with level of significance 99%, and the F test shows that the model is significant with p-value < 2.2e-16,

Heteroskedasticity

The residual plot



From the residual plot it is easy to see the variance of residuals increases with the fitted values, which gives evidence against the assumptions that $Var[\hat{e}] = \sigma^2$ where σ^2 is a constant



The normal QQ plot shows there is evidence against that the standardized residuals follow a normal distribution