HW2

September 19, 2018

Analyzing Apple Stock Data
Avery Loftin
Importing Data:

```
In [28]: library(TSA)

         AAPL <- read.csv("AAPL.csv", header = T)
         AAPL.AdjClose <- ts(AAPL$Adj.Close)
```

This dataset is 5 years worth of adjusted closing prices for Apple stock from 9/12/2013 to 9/11/2018. The dataset consists of 1259 floating point values.
Summary statistics:

```
In [37]: cat("Mean: ", mean(AAPL.AdjClose), "\n")
         cat("Standard Deviation: ", sd(AAPL.AdjClose), "\n")
         cat("Median: ", median(AAPL.AdjClose), "\n")
         cat("Variance: ", var(AAPL.AdjClose), "\n")
```

```
Mean:  119.1819
Standard Deviation:  36.73071
Median:  109.9198
Variance:  1349.145
```

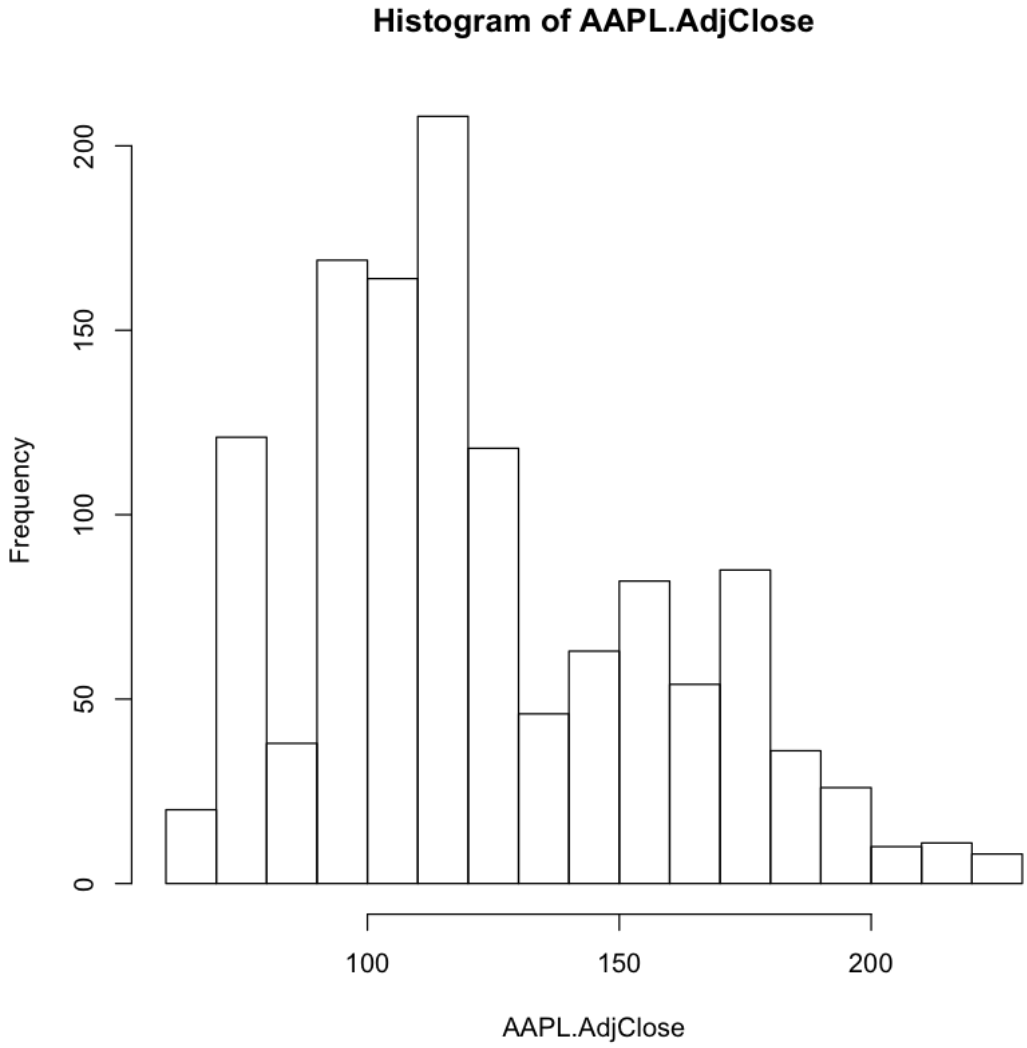
Plotting Adjusted Closing Prices:

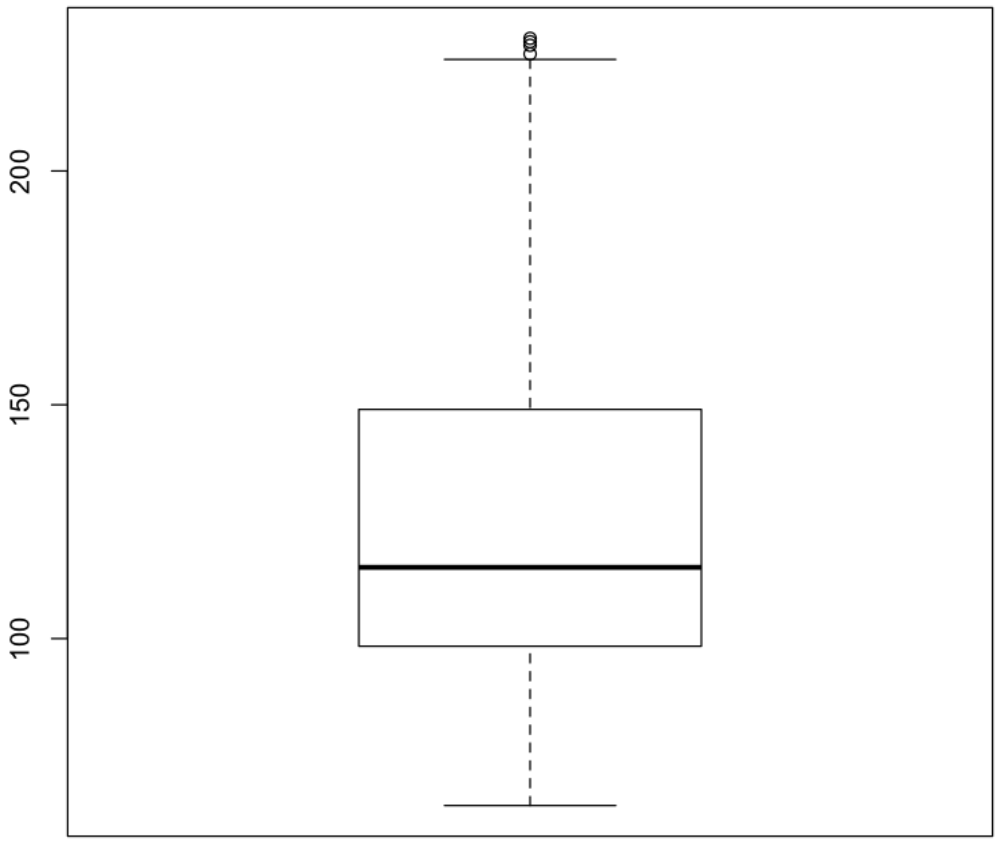
```
In [5]: plot(AAPL.AdjClose,
             ylab="Adjusted Closing Pice ($)",
             xlab="Time",
             main="Apple Adjusted Closing Price 9/12/2013 - 9/11/2018")
```

These data show an upward trend in the adjusted closing price of Apple stock, meaning the mean tends to increase with time. Variance appears to remain reasonably constant over time.

Checking For Outliers:

```
In [6]: hist(AAPL.AdjClose)
        boxplot(AAPL.AdjClose)
```

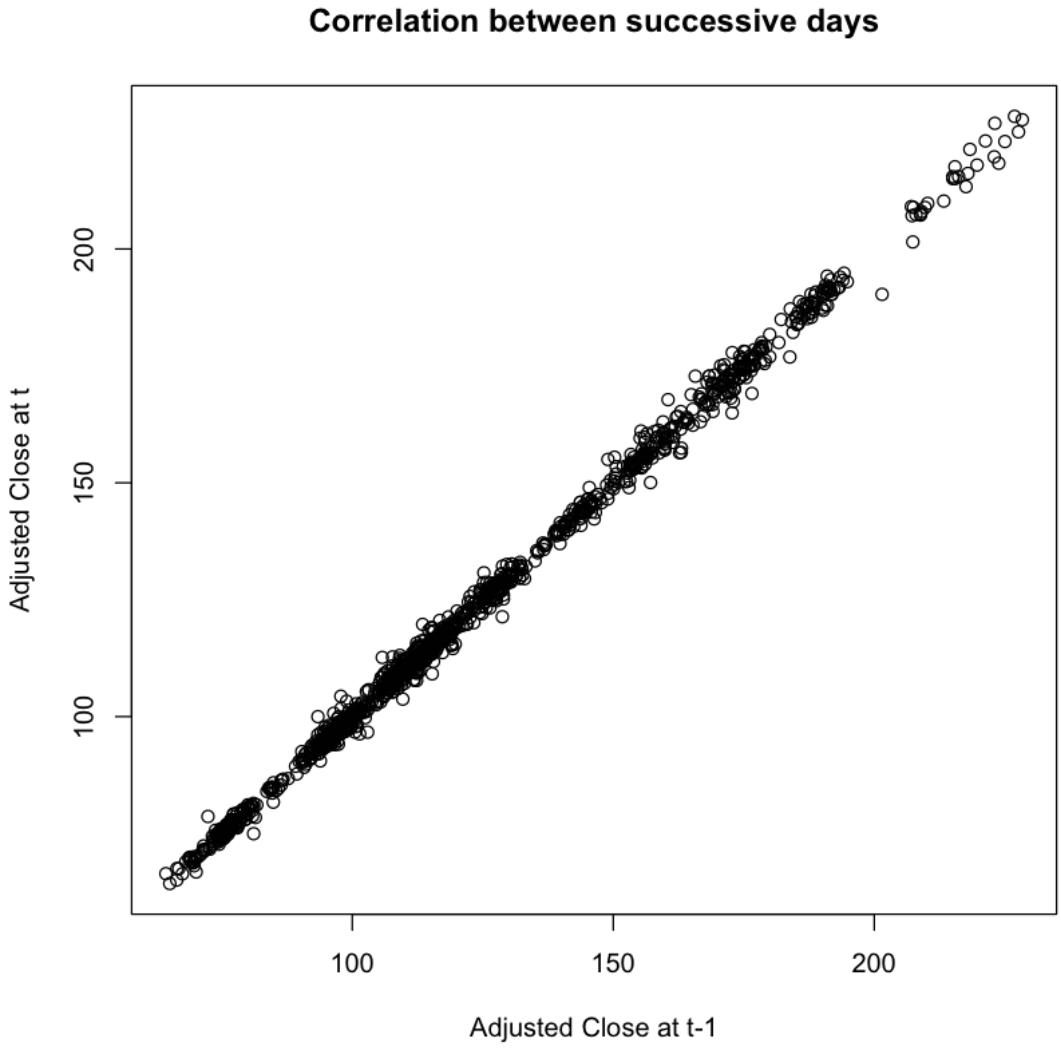

Histogram of AAPL.AdjClose
Frequency
0
50
100
150
200
100
150
200
AAPL.AdjClose



The boxplot shows that 4 observations are outliers. These observations are likely due to the large, recent spike in the stock price, and may not be indicative of the true, overall trend of the stock. The histogram is positively skewed.

Checking for Autocorrelation With Lag of 1 Timestep:

```
In [7]: plot(AAPL.AdjClose,
              zlag(AAPL.AdjClose),
             ylab="Adjusted Close at t",
             xlab="Adjusted Close at t-1",
             main="Correlation between successive days")
```




```
In [22]: cor(AAPL.AdjClose[2:length(AAPL.AdjClose)], na.omit(zlag(AAPL.AdjClose)))
```

0.99892613190851

The plot and corellation of .9989 suggest that successive data points are highly correlated.

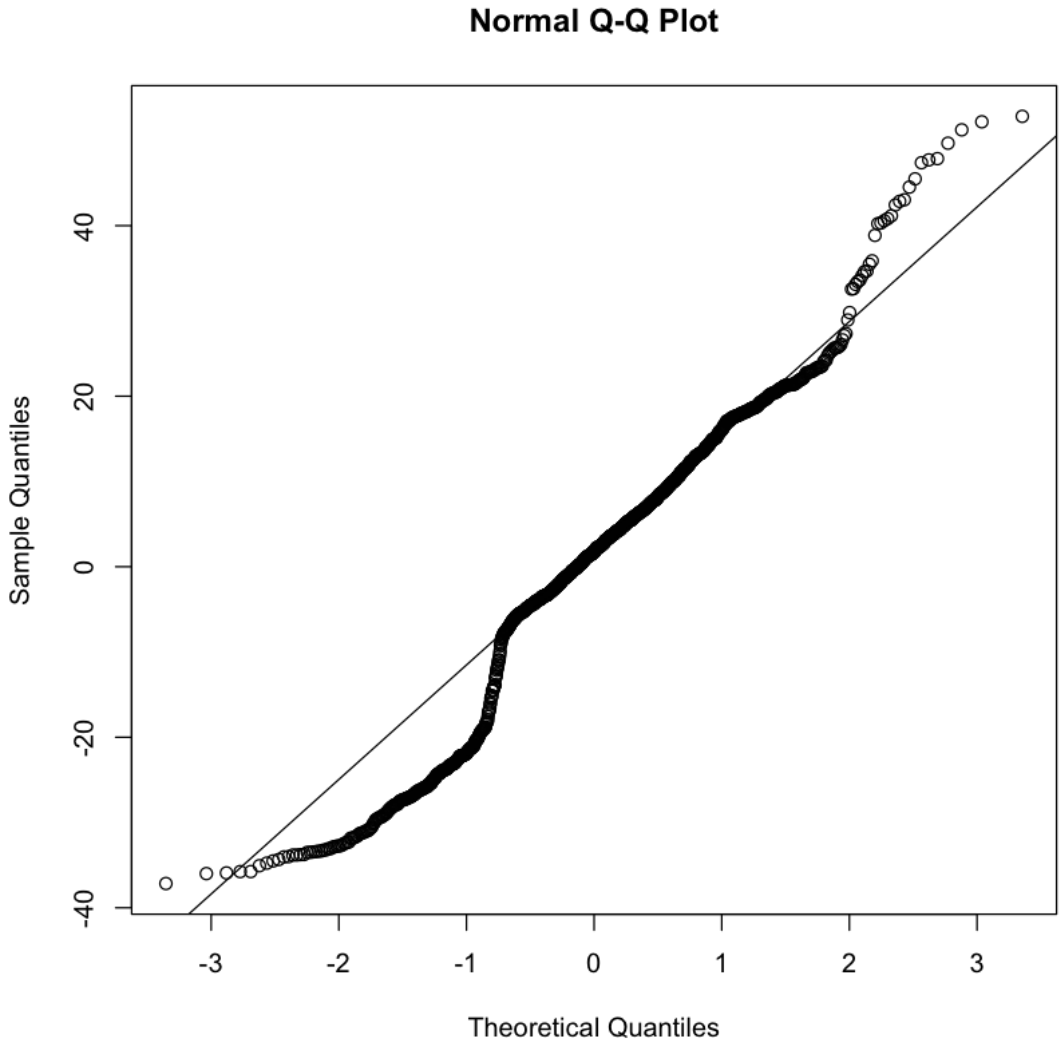
Linear Model for Stock Price:

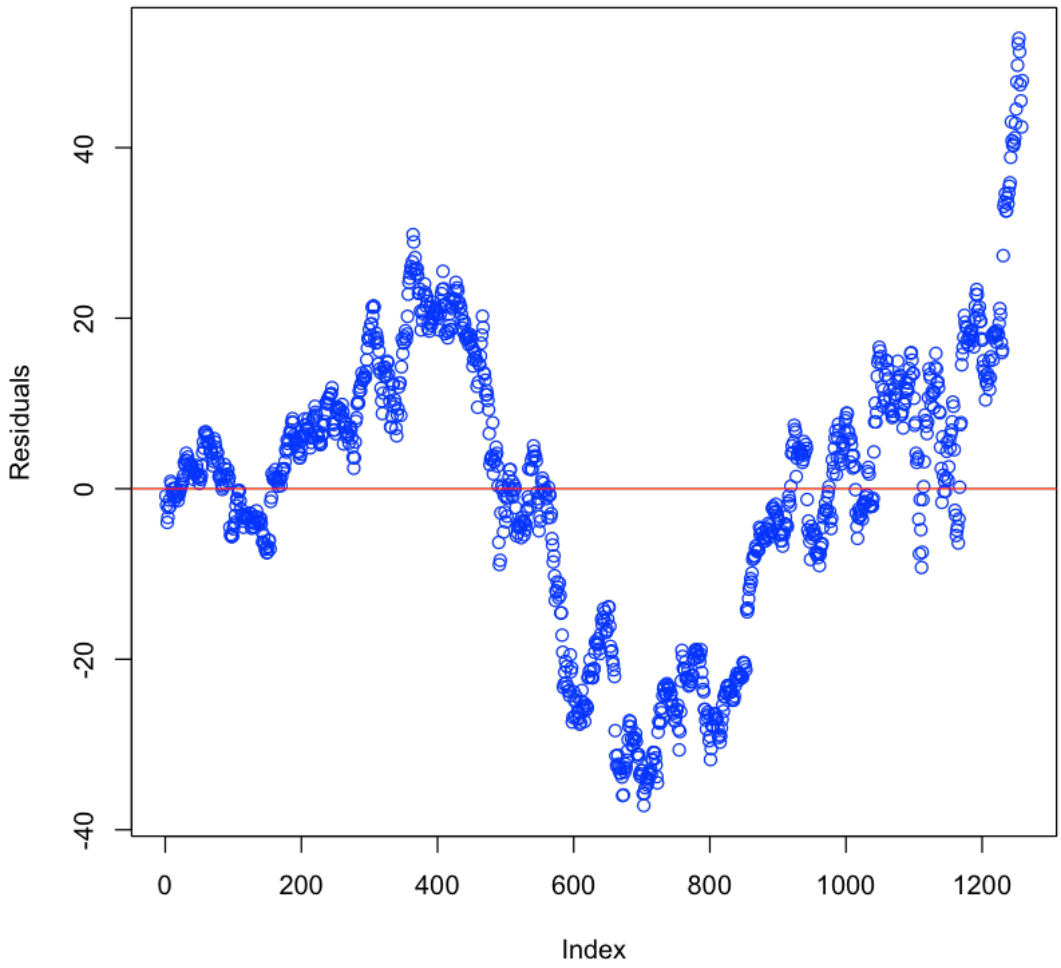
```
In [53]: lm <- lm(AAPL.AdjClose ~ time(AAPL.AdjClose))

         plot(AAPL.AdjClose,
             ylab="Adjusted Closing Pice ($)",
             xlab="Time",
             main="Apple Adjusted Closing Price 9/12/2013 - 9/11/2018")
         abline(lm)
```

```r
summary(lm)
qqnorm(resid(lm))
qqline(resid(lm))

plot(resid(lm), col='blue', ylab="Residuals")
abline(h=0, col='red')
```

```
Call:
lm(formula = AAPL.AdjClose ~ time(AAPL.AdjClose))

Residuals:
    Min      1Q  Median      3Q     Max
-37.158  -7.154   1.770  10.961  52.819

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         62.281302   0.928441   67.08   <2e-16 ***
time(AAPL.AdjClose)  0.090318   0.001277   70.75   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 16.46 on 1257 degrees of freedom
Multiple R-squared:  0.7993,Adjusted R-squared:  0.7991
F-statistic:  5006 on 1 and 1257 DF,  p-value: < 2.2e-16
```


Apple Adjusted Closing Price 9/12/2013 - 9/11/2018
Adjusted Closing Pice ($)
100
150
200
0
200
400
600
800
1000
1200
Time


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
40
20
0
-20
-40
-3
-2
-1
0
1
2
3



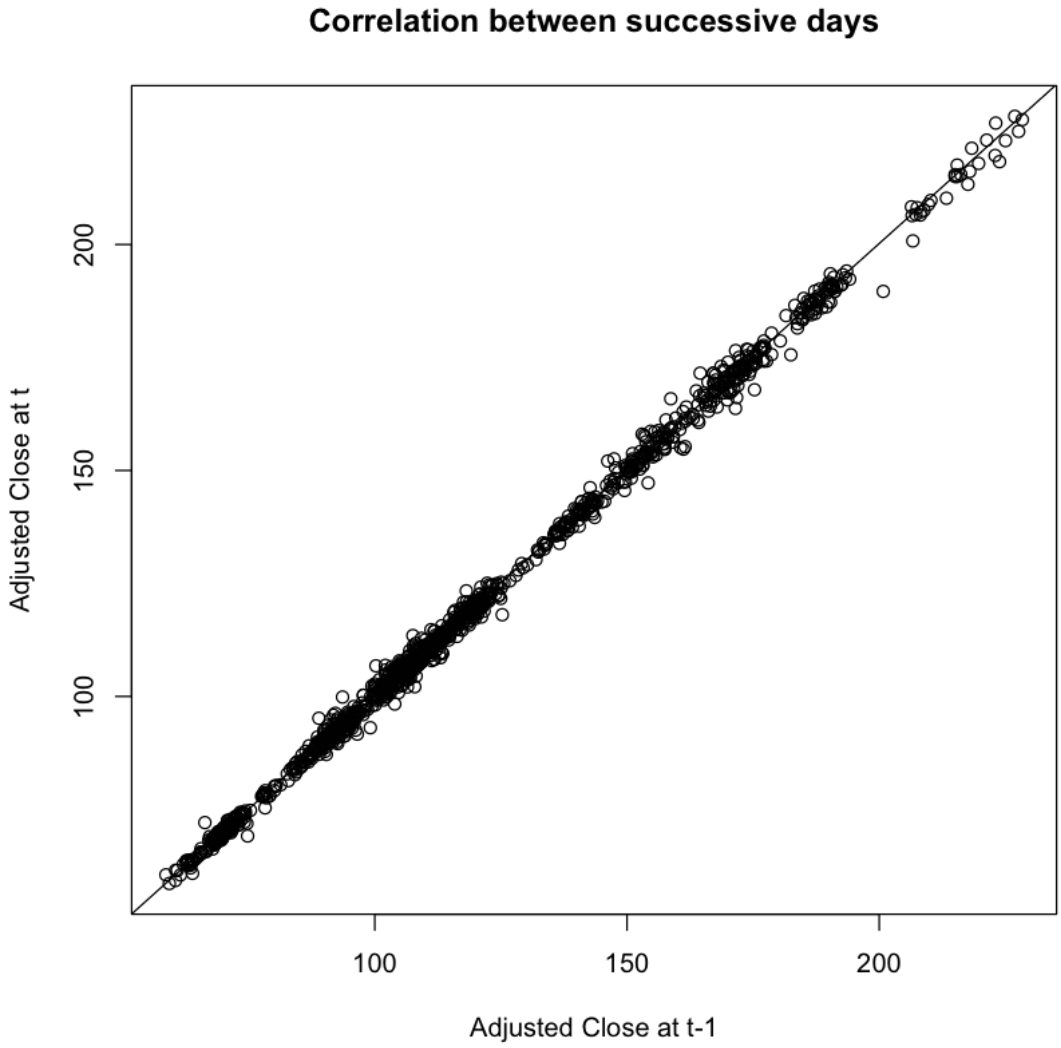
The QQ Plot suggests that the error term is not normally distributed due to the skew on the ends of the plot. Furthermore, the residual plot has a very apparent pattern rather than being closely and evenly distributed about the mean. These findings suggest that a linear model is a very poor fit for these data. On the other hand, the f statistic is very statistically significant with a p-value of $< 2.2 * 10^{-16}$

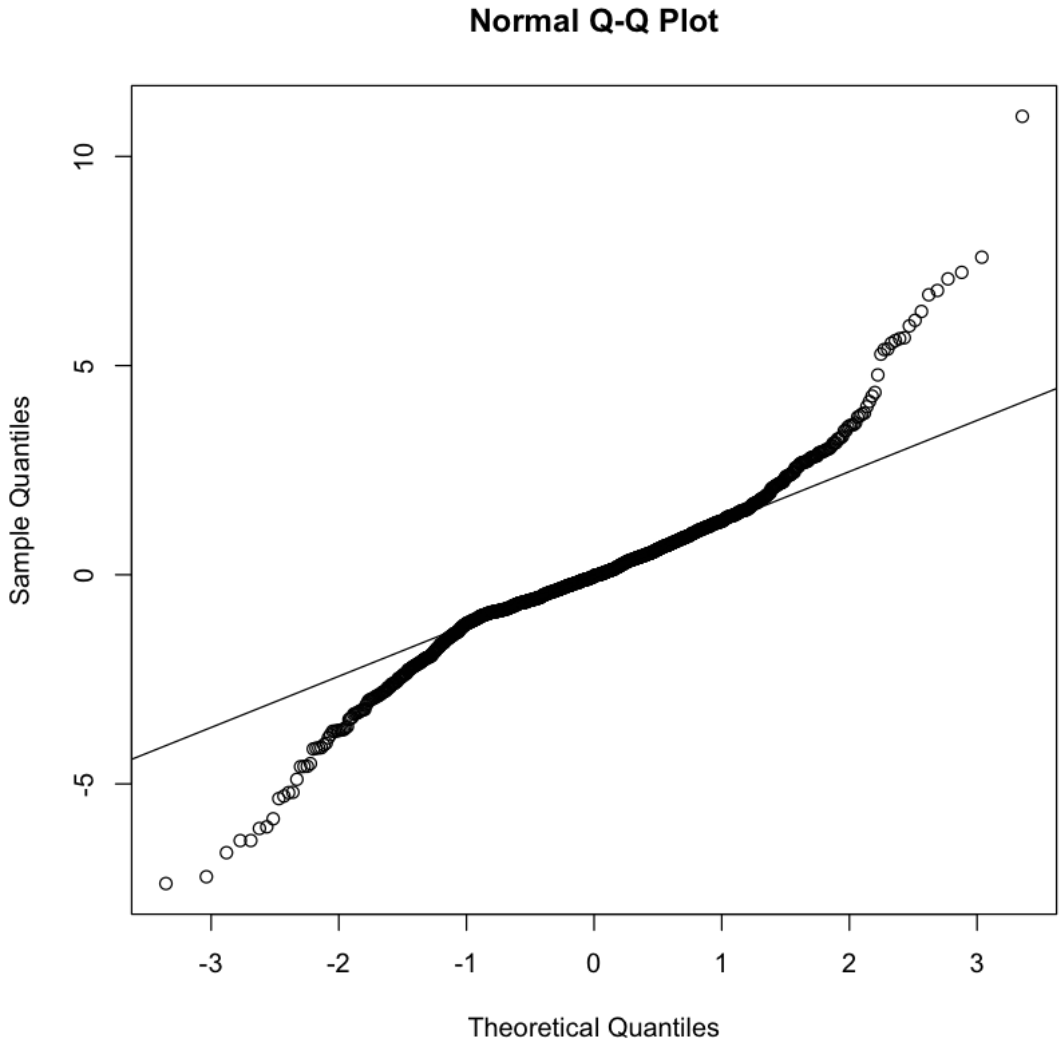
Linear Model for Lag of 1:

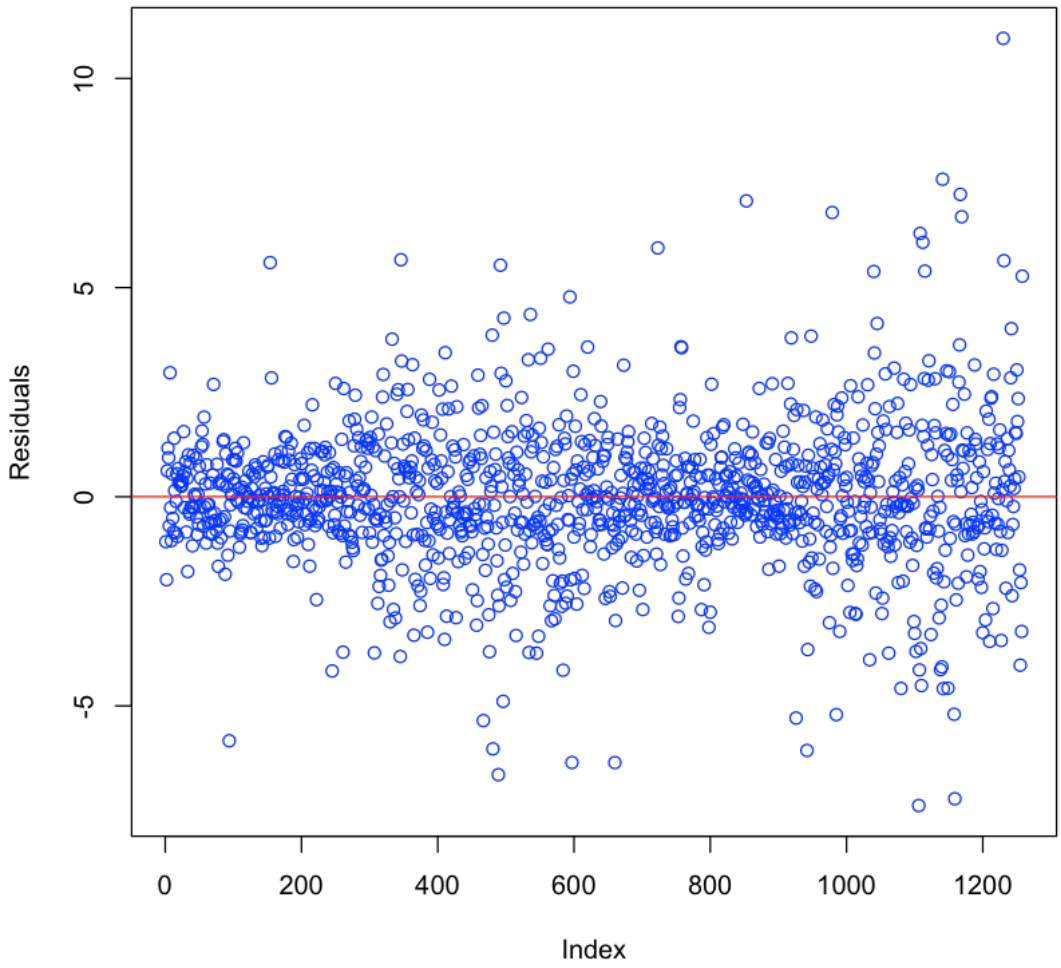
```
In [56]: lm <- lm(AAPL.AdjClose[2:length(AAPL.AdjClose)] ~ na.omit(zlag(AAPL.AdjClose)))

         plot(AAPL.AdjClose,
              zlag(AAPL.AdjClose),
             ylab="Adjusted Close at t",
             xlab="Adjusted Close at t-1",
             main="Correlation between successive days")
```

```r
abline(lm)

summary(lm)

qqnorm(resid(lm))
qqline(resid(lm))

plot(resid(lm), col="blue", ylab="Residuals")
abline(h=0, col="red")
```

```
Call:
lm(formula = AAPL.AdjClose[2:length(AAPL.AdjClose)] ~ na.omit(zlag(AAPL.AdjClose)))

Residuals:
    Min      1Q  Median      3Q     Max
-7.3818 -0.8053 -0.0329  0.8453 10.9582

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -0.01223    0.16326  -0.075     0.94
na.omit(zlag(AAPL.AdjClose))  1.00119    0.00131 764.107   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.701 on 1256 degrees of freedom
Multiple R-squared:  0.9979,Adjusted R-squared:  0.9979
F-statistic: 5.839e+05 on 1 and 1256 DF,  p-value: < 2.2e-16
```


Correlation between successive days
Adjusted Close at t
Adjusted Close at t-1
100
150
200
100
150
200


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
10
5
0
-5
-3
-2
-1
0
1
2
3



There are a few flaws with this model. The intercept value is not statistically significant and the qq plot shows a skew on both ends of the plot. Also, the residual plot shows an uneven distribution about the x axis. Despite these flaws, the F Statistic is very statistically significant with its p-value of $< 2.2 * 10^{-16}$