MATH 490 Stock Data Final Project

Mitchell Krystofiak

12/16/2021

## Introduction

Founded in 2003, Tesla aims to accelerate the world’s transition in to sustainable energy by building better, faster, and more affordable electric cars than gasoline cars. Tesla also focuses on producing clean energy power through the installation of solar panels and energy solutions such as Powerwall, PowerPack and Solar Roof. Tesla has continued, with the leadership of CEO Elon Musk, to release cutting edge electric vehicle designs, including the first ever all-electric sedan. They have also continued to prove thath they can produce all-electric vehicles while keeping up with safety protocols, reliability ratings, battery efficiency and affordability.

## Tesla Stock Information

On June 29, 2010, Tesla Motors launched its initial public offering on NASDAQ by issuing 13,300,000 shares of common stock at a price of $17.00 per share. As of December 12, 2021, Tesla is now worth $1017.03 per share, with a 52 week low of $539.49 and a 52 week high of $1243.49 per share.

Table :Tesla Stock Data

| **Date** | **Open** | **High** | **Low** | **Close** | **Adj.Close** | **Volume** |
| --- | --- | --- | --- | --- | --- | --- |
| 2021-12-03 | 1,084.79 | 1,090.575 | 1,000.21 | 1,014.97 | 1,014.97 | 30,773,995 |
| 2021-12-06 | 1,001.51 | 1,021.640 | 950.50 | 1,009.01 | 1,009.01 | 27,221,037 |
| 2021-12-07 | 1,044.20 | 1,057.674 | 1,026.81 | 1,051.75 | 1,051.75 | 18,694,857 |
| 2021-12-08 | 1,052.71 | 1,072.380 | 1,033.00 | 1,068.96 | 1,068.96 | 13,968,790 |
| 2021-12-09 | 1,060.64 | 1,062.490 | 1,002.36 | 1,003.80 | 1,003.80 | 19,446,486 |
| 2021-12-10 | 1,008.75 | 1,020.980 | 982.53 | 1,017.03 | 1,017.03 | 19,229,185 |
| Most recent TSLA stock data from Decemeber 12, 2021. | | | | | | |

## Project Goals

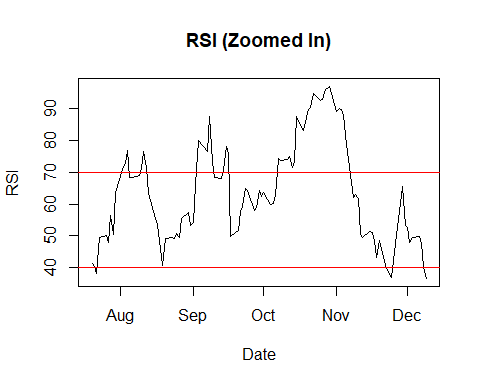
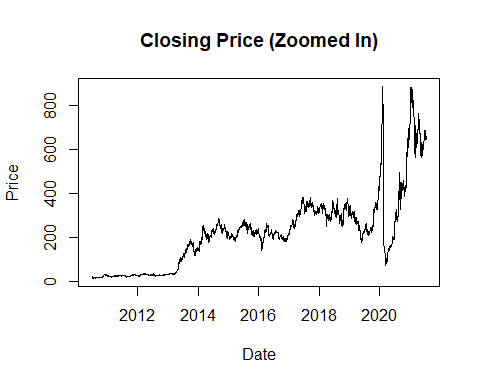
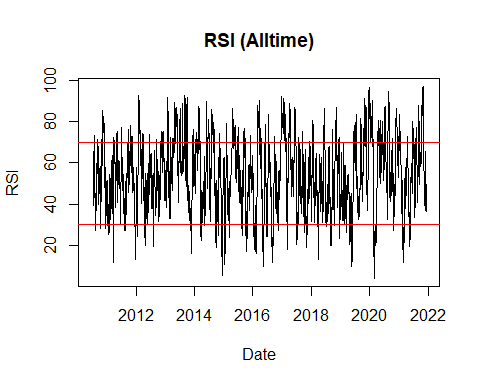
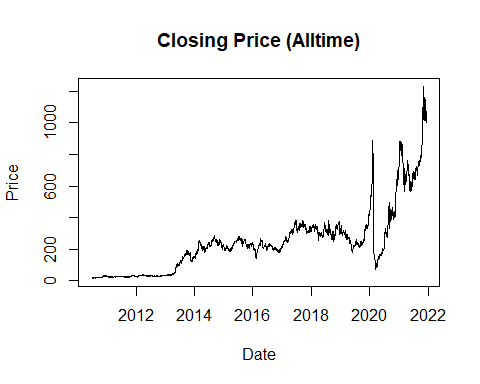
One of the major goals of this project is to better understand technical analysis and get a look into another popular indicator: the Relative Strength Index (RSI). The RSI is a momentum indicator that measures the magnitude of recent price changes to evaluate overbought or oversold conditions in the price of a stock. The RSI is displayed as an oscillator between 0 and 100. It was developed by J. Welles Wilder Jr. and introduced in 1978. Traditionally, RSI values greater than or equal to 70 indicate that a stock is becoming overbought or overvalued, and may be primed for a trend reversal or pullback. An RSI value less than or equal to 30 usually indicates that a stock is oversold or undervalued.

In addition to the RSI, we can also look at the On Balance Volume (OBV) indicator. This is another momentum indicator that uses volume flow to predict changes in the stock price. It was developed by Joseph Granville in 1963. Since this is more of a leading indicator, it is hard to determine a buy signal off of just this, and no technical analysis is included.

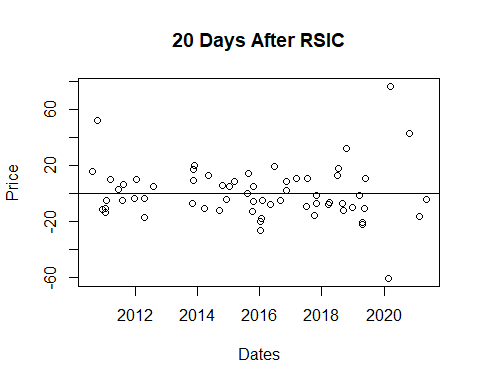
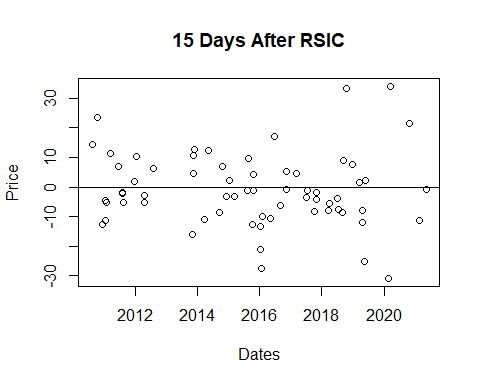
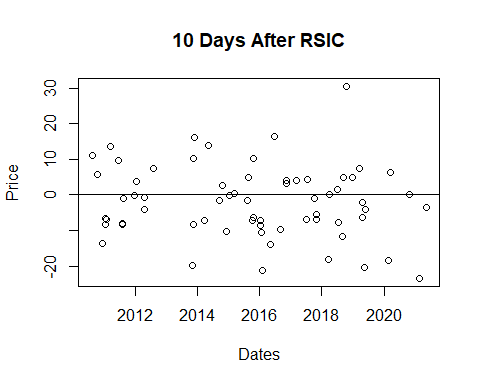
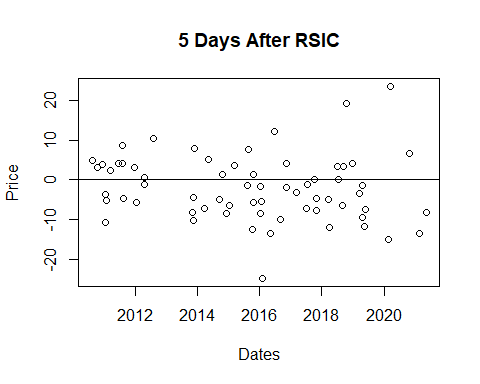
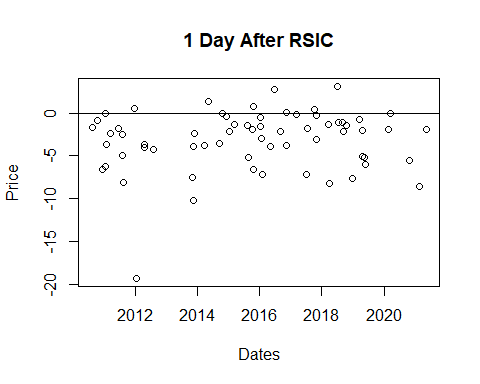
The last goal of this project is to examine basic rating methods against other stocks. In other words, how can we rank a stock versus another stock to determine what our likelihood of success is if we invest in one company over another? To use the Markov Method, Massey Method, or the Colley method, there needs to be some sort of win or lose definition, so we can say “stock A lost to stock B” and we can use these methods.

## RSI Computation and Analysis

To calculate the RSI, we need to calculate the change in closing price between days, the average gain, the average loss, and the Relative Strength. For the change, we just subtract yesterdays closing price from todays closing price for all of Tesla’s data. For the average gain and loss, we average 14 days of gains and losses. Then, we compute our RS by dividing the average gains over the average losses. Finally to compute the RSI:

If the average loss is 0, our RSI is 100. 

Now that we have our RSI calculated and plotted alongside the closing price, we can start discussing a buy signal! Though this is not an exact science, and many different tactics have been used, a common indicator to buy is when the RSI falls below 30. This could signify that the stock is oversold and undervalued, making it prime for a trend reversal upwards. We will set this as our buy signal, and then we will begin our thorough technical analysis. Below are the recorded prices after a RSI Crossover (RSIC). We also follow the Non RSIC.



Next, we will compute the summary statistics, boxplots, histograms and qqplots of each of the days after results. This will give us more insight into the normality of the data, and possibly allow us to detect whether or not we have statistical differences in results early on.

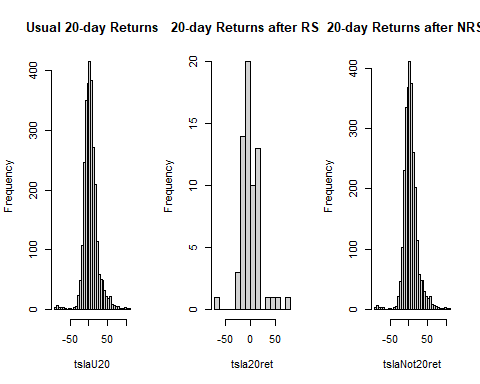
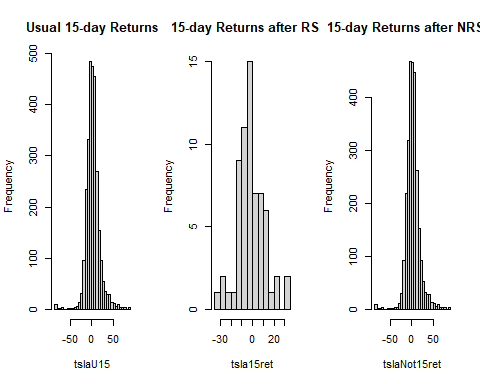
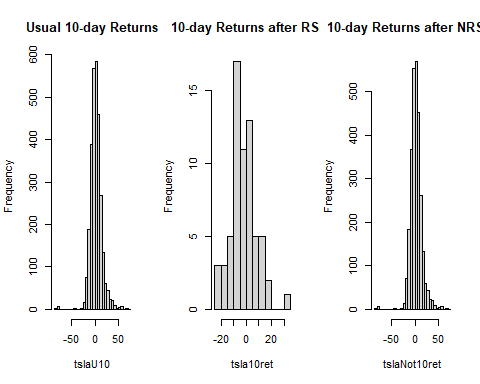
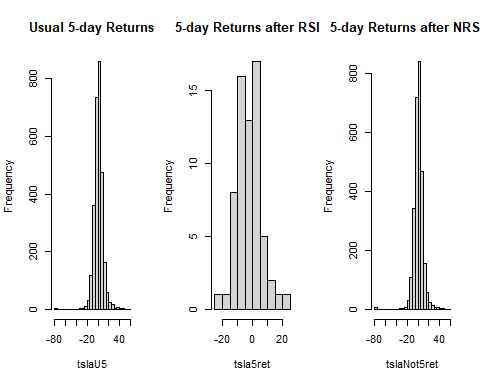
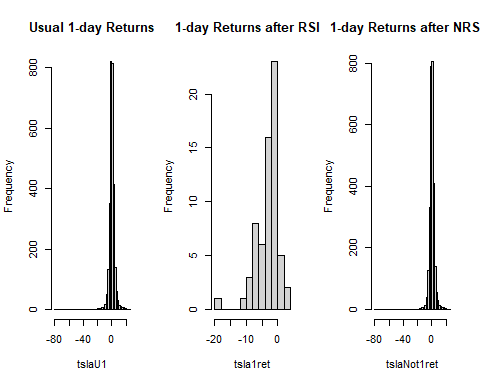
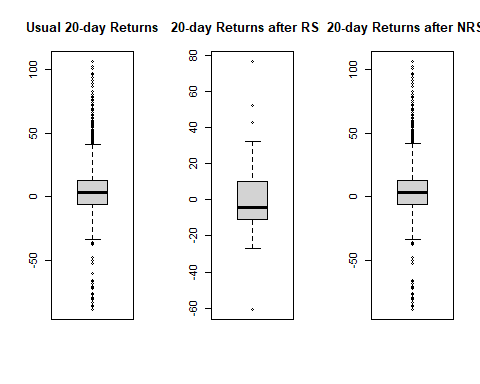
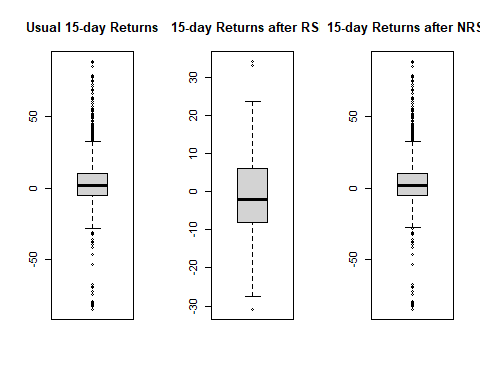
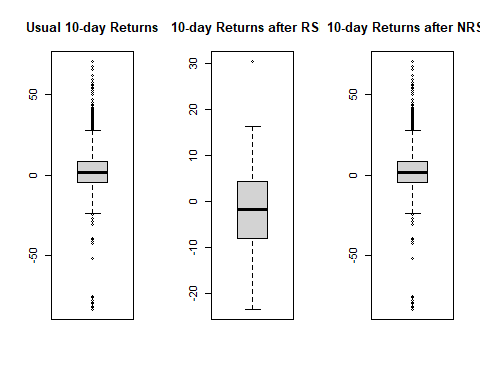
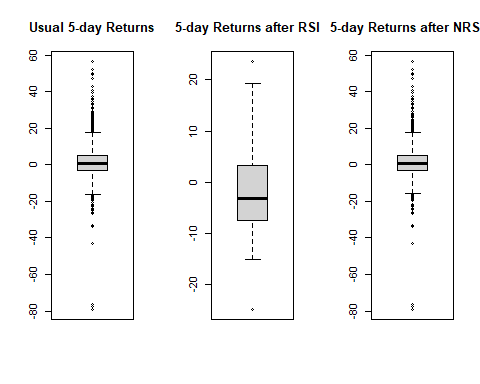
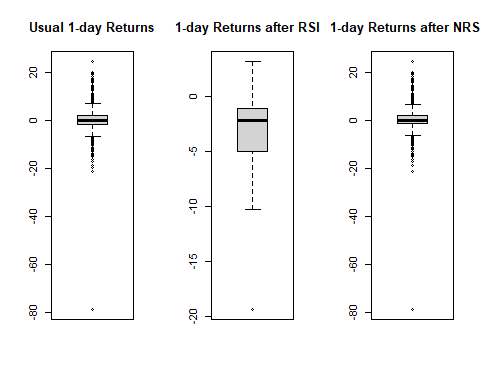
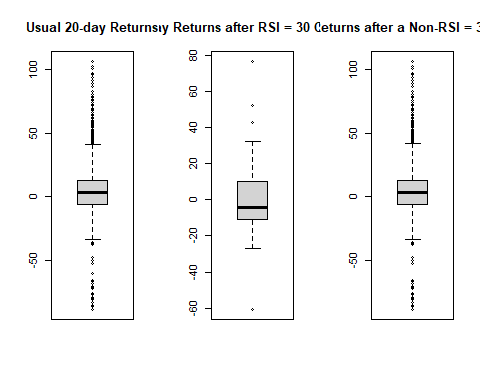
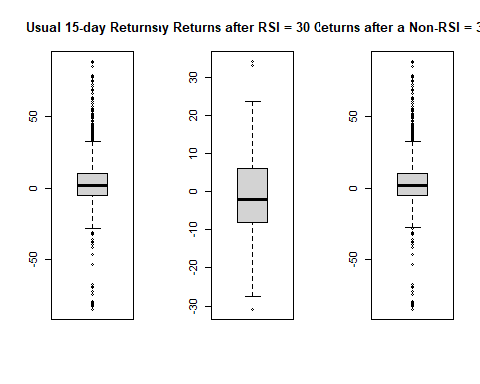
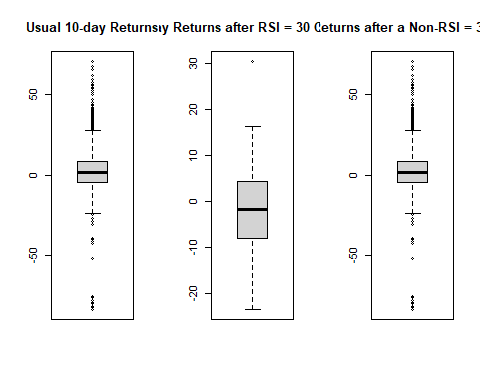
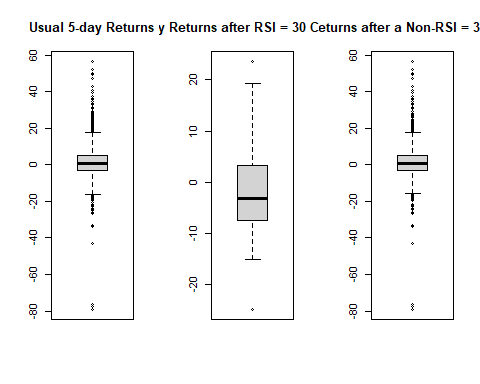
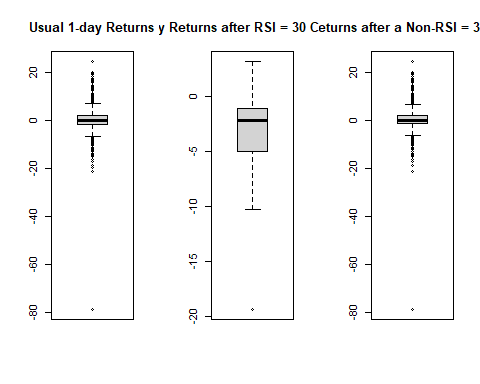
| **Stats** | **U1** | **Ret1** | **NRet1** |
| --- | --- | --- | --- |
| Min | -78.5408064 | -19.327430 | -78.5408064 |
| Q1 | -1.4831321 | -5.043818 | -1.3769208 |
| Q2 | 0.1259696 | -2.177375 | 0.1668901 |
| Mean | 0.2184509 | -3.175669 | 0.3023247 |
| Q3 | 1.9126184 | -1.088476 | 1.9400550 |
| Max | 24.4028631 | 3.111044 | 24.4028631 |

| **Stats** | **U5** | **Ret5** | **NRet5** |
| --- | --- | --- | --- |
| Min | -79.156033 | -24.855284 | -79.1560329 |
| Q1 | -3.267817 | -7.421494 | -3.1609891 |
| Q2 | 0.859610 | -3.132331 | 0.8949529 |
| Mean | 1.121351 | -2.121693 | 1.1955771 |
| Q3 | 5.245217 | 3.271452 | 5.2747110 |
| Max | 56.475562 | 23.505757 | 56.4755619 |

| **Stats** | **U10** | **Ret10** | **NRet10** |
| --- | --- | --- | --- |
| Min | -83.383332 | -23.467295 | -83.383332 |
| Q1 | -4.759292 | -8.003304 | -4.638335 |
| Q2 | 1.661876 | -1.604848 | 1.718928 |
| Mean | 2.311558 | -2.027288 | 2.395142 |
| Q3 | 8.351827 | 4.348755 | 8.405902 |
| Max | 70.482349 | 30.453561 | 70.482349 |

| **Stats** | **U15** | **Ret15** | **NRet15** |
| --- | --- | --- | --- |
| Min | -84.800571 | -30.868084 | -84.800571 |
| Q1 | -5.221073 | -8.206449 | -5.030588 |
| Q2 | 2.219952 | -2.103960 | 2.288608 |
| Mean | 3.501927 | -1.004204 | 3.617280 |
| Q3 | 10.048898 | 6.163388 | 10.113198 |
| Max | 88.486762 | 33.991205 | 88.486762 |

| **Stats** | **U20** | **Ret20** | **NRet20** |
| --- | --- | --- | --- |
| Min | -88.873668 | -60.6265367 | -88.873668 |
| Q1 | -6.143733 | -10.7081979 | -6.006850 |
| Q2 | 3.003137 | -4.0458400 | 3.125283 |
| Mean | 4.737494 | 0.2395848 | 4.876269 |
| Q3 | 12.880785 | 10.0884956 | 13.156066 |
| Max | 106.303639 | 76.2908054 | 106.303639 |



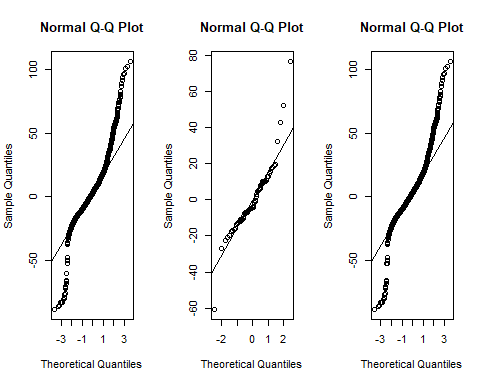
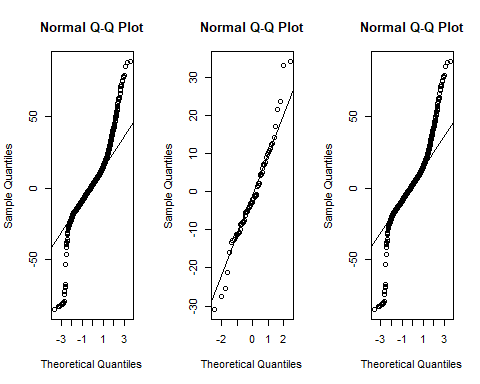
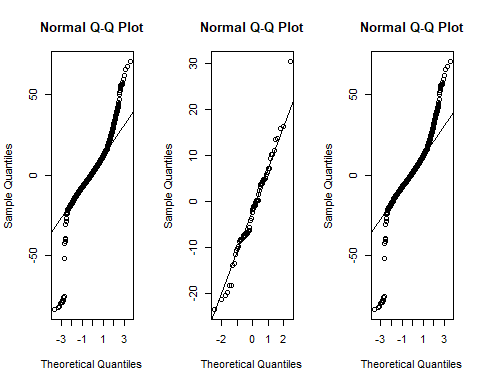
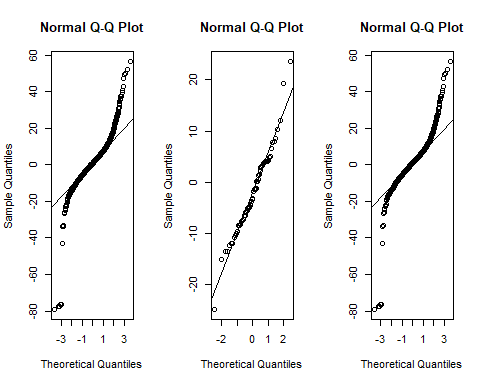
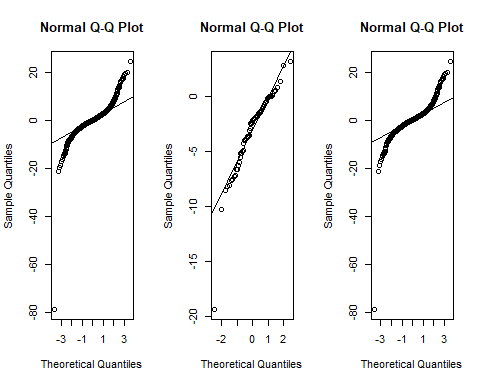
| Stats | Lengths | Variances |
| --- | --- | --- |
| Usual 1-day | 2,884 | 14.65901 |
| 1-day After RSIC | 65 | 12.32601 |
| 1-day After NRSIC | 2,799 | 14.24011 |

| Stats | Lengths | Variances |
| --- | --- | --- |
| Usual 5-day | 2,884 | 75.91500 |
| 5-day After RSIC | 65 | 66.57802 |
| 5-day After NRSIC | 2,799 | 74.70972 |

| Stats | Lengths | Variances |
| --- | --- | --- |
| Usual 10-day | 2,884 | 155.5389 |
| 10-day After RSIC | 65 | 102.3528 |
| 10-day After NRSIC | 2,799 | 156.1045 |

| Stats | Lengths | Variances |
| --- | --- | --- |
| Usual 15-day | 2,884 | 249.4879 |
| 15-day After RSIC | 65 | 152.5576 |
| 15-day After NRSIC | 2,799 | 251.4682 |

| Stats | Lengths | Variances |
| --- | --- | --- |
| Usual 20-day | 2,884 | 353.3422 |
| 20-day After RSIC | 65 | 362.2845 |
| 20-day After NRSIC | 2,799 | 353.4731 |



By looking at these plots, we start to see some pretty close to normal results with the 5 day, the 10 day, and the 15 day results. We also want to do a Shaprio test to test the normality of the data one last time!

##   
## Shapiro-Wilk normality test  
##   
## data: tslaU1  
## W = 0.82934, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tsla1ret  
## W = 0.89166, p-value = 3.462e-05

##   
## Shapiro-Wilk normality test  
##   
## data: tslaNot1ret  
## W = 0.82137, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tslaU5  
## W = 0.88569, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tsla5ret  
## W = 0.974, p-value = 0.1871

##   
## Shapiro-Wilk normality test  
##   
## data: tslaNot5ret  
## W = 0.88058, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tslaU10  
## W = 0.90219, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tsla10ret  
## W = 0.98034, p-value = 0.3885

##   
## Shapiro-Wilk normality test  
##   
## data: tslaNot10ret  
## W = 0.89881, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tslaU15  
## W = 0.90497, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tsla15ret  
## W = 0.96952, p-value = 0.1088

##   
## Shapiro-Wilk normality test  
##   
## data: tslaNot15ret  
## W = 0.90264, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tslaU20  
## W = 0.91202, p-value < 2.2e-16

##   
## Shapiro-Wilk normality test  
##   
## data: tsla20ret  
## W = 0.8941, p-value = 4.255e-05

##   
## Shapiro-Wilk normality test  
##   
## data: tslaNot20ret  
## W = 0.9115, p-value < 2.2e-16

From the above p-values, we have that the Tesla 5, 10, and 15 days after a RSI crossover appear to be normal. Now, we want to conduct Wilcoxon tests since our data has unequal variances and not normal. Even though the t-test can give accurate rejections or acceptions of the null hypothesis, this is usually for a large sample size. The RSI crossover set has 65 values. It depends on what large means. Nonetheless, the Wilcoxon test will tell us if the data is statistically different or the same. If we wanted to make a further inference about whether this statistical difference is good or bad, we would need to have two samples with equal variances.

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU1 and tsla1ret  
## W = 146590, p-value = 6.887e-15  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU1 and tslaNot1ret  
## W = 3982340, p-value = 0.3841  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU5 and tsla5ret  
## W = 117623, p-value = 0.0003949  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU5 and tslaNot5ret  
## W = 4002977, p-value = 0.7236  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU10 and tsla10ret  
## W = 115029, p-value = 0.001421  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU10 and tslaNot10ret  
## W = 3991068, p-value = 0.7507  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU15 and tsla15ret  
## W = 111783, p-value = 0.006156  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU15 and tslaNot15ret  
## W = 3974995, p-value = 0.7264  
## alternative hypothesis: true location shift is not equal to 0

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU20 and tsla20ret  
## W = 109978, p-value = 0.0124  
## alternative hypothesis: true location shift is not equal to 0

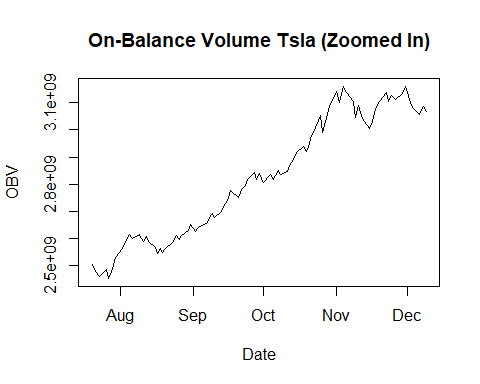
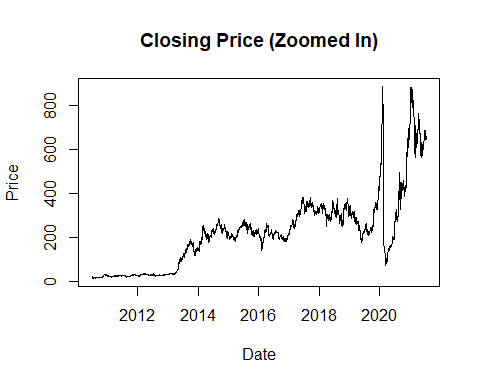
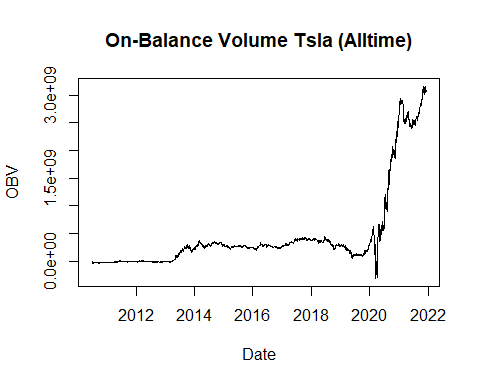
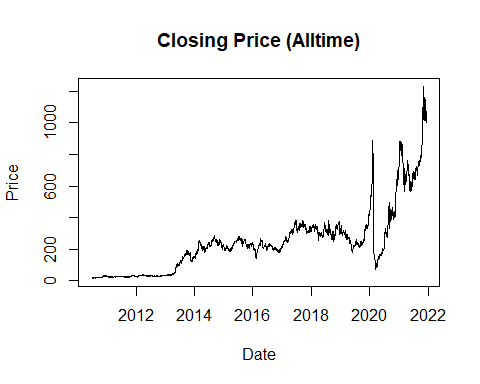
##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: tslaU20 and tslaNot20ret  
## W = 3958825, p-value = 0.7008  
## alternative hypothesis: true location shift is not equal to 0

Judging from the series of Wilcoxon tests, we have a nice variety of data. For all of the post-RSI crossovers, we have a statistical difference in the data. The 20 day value is pushing the boundary if we used a .01 p-value threshold, however it is close enough to call it statistically different. In addition, the post Non-RSI crossovers are not different. This is to be expected because they are the closest to the actual set of data.

In conclusion, using the RSI value drop below 30 does give us a statistical difference from the usual price. Since we did not have equal variances, we cannot confirm if this is a positive or negative difference, but we can at least conclude that this indicator is worth looking into more, possibly in combination with the OBV, Simple Moving Averages, and more.

## OBV Calculation

To calculate the OBV, we follow a very simple process:

1. If today’s closing price is higher than yesterday’s closing price, then
2. If today’s closing price is less than yesterday’s closing price, then
3. If today’s closing price equals yesterday’s closing price, then
4. 

The OBV tallies up and down volume, creating a smooth indicator line to predict when major moves might occur based on changes in relative trading volume. Investors can use the OBV to provide many key predictions, such as a bullish divergence predicting the price will break resistance or a bearish divergence predicting a rally will stall or reverse.

## Attempting Rating Methods between Stocks

This is probably the most interesting part of our data! We want to know how a certain stocks performance compares with another stocks. We as investors want to make money! Say we have a list of 500 stocks, we’ve analyzed them from a statistical perspective, yet we don’t if one stock is betteer than the other. Why don’t we use our rating algorithms and linear algebra?

There are three primary methods we have learned to implement: Massey’s method, Colley’s method, and Markov’s method. Traditionally, these methods look at games between two teams, and we can easily calculate a point differential, a total yards differential, or a winning percentage.

What exactly is a “win” when we talk about stocks? There are a few ways to think about this:

1. We compare a stock against itself. We count every day that a stock goes up from the day before it as a win.
2. We compare the stock against other stocks. We count every day that stock A goes up and stock B goes down as a win for stock A, and the opposite would be a win for stock B. If they both don’t increase or decrease, we will count it as a tie.

Though we are not looking at point differentials, some ways to maybe explore this for stocks would be:

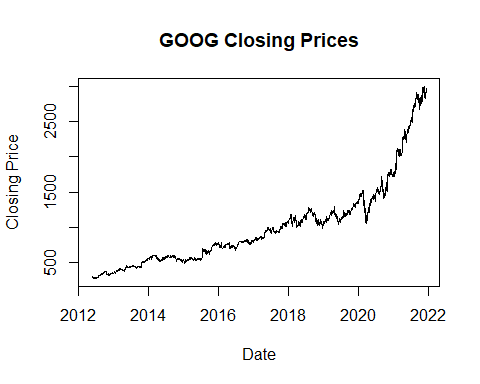
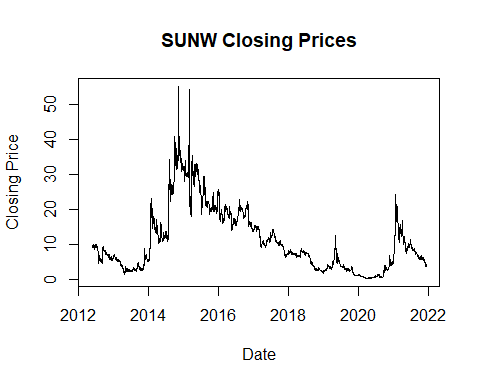
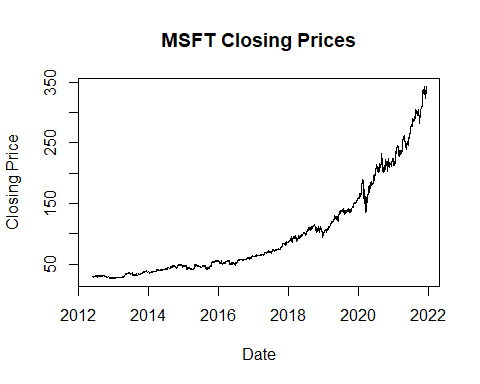
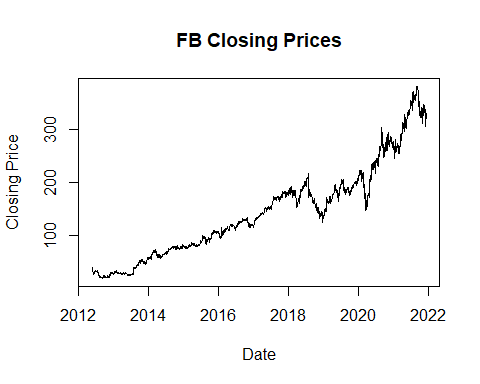
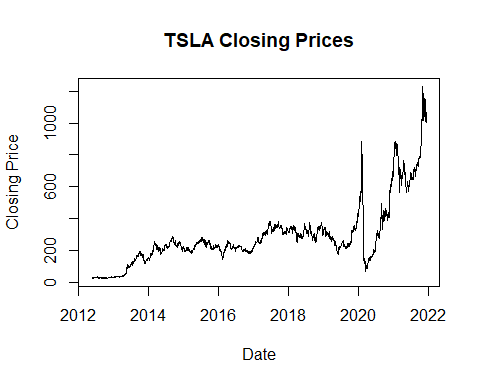
1. Closing Price
2. P/E ratio
3. Outstanding Shares
4. Volume
5. Any technical indicator

This might get suspect since stock prices and such differs greatly between companies. For example, Sunworks (SUNW) is worth approximately $3.78 per share while Google (GOOG/GOOGL) is worth $2960.03 per share. A point differential between these two doesn’t exactly make sense. A great amount of stocks fall far below Google’s share price.

## Calculating Win Percentages

Numbers 1 and 2 from above were used. The data was truncated to include the most recent data with a size of the lowest maximum data. We analyzed 5 stocks, Tesla (TSLA), Facebook (FB), Microsoft (MSFT), Sunworks, an Google, and Facebook had the lowest amount of data at 2408 values.

A quick look at each of the stock’s closing prices:



Now, lets look at our first computation of the winning percentages:

## Ratings Ranking  
## GOOG 0.5336379 1  
## MSFT 0.5307309 2  
## FB 0.5249169 3  
## TSLA 0.5240864 4  
## SUNW 0.4323090 5

We see that the first four stocks are relatively close, and Sunworks is very low compared to these ones. Sunworks is a very different company compared to the other four. Looking at this win percentage alone can give us a little insight into how well a stock is performing.

## Conclusions

We have calculated a lot of interesting data. We have calculated the RSI, the OBV, and a basic win percentage of the 5 stocks listed above. We also have performed a technical analysis on TSLA, and confirmed that using a RSI value falling below 30 as a buy signal does result in a statistical difference, and thus we should use more indicators and analysis to confirm if this is an appropriate and profitable buy signal.

In terms of our ranking methods, we didn’t quite get to the methods we wanted to try. It would have been really nice to see how we can translate sports ranking methods to stock data. The Colley equations were all set up to go, but project after project stacked up and time just ran out.

In terms of next steps, we need to define a ‘game’ against two stocks, and then determine the number of wins and losses to formulate the Colley method.