BUSINESS STATISTICS LAB USING R

CAT II – PROJECT DOCUMENTATION

NETFLIX STOCK MARKET ANALYSIS

Team Members:

Manoraj R (2133031)

Sanjay V V (2133038)

Introduction:

Stock analysis is a method that an investor or trader uses to evaluate and investigate an investment or the stock market as whole. Stock analysts attempt to determine the future activity of stock market. This helps the investors or traders Investors in making decisions of buying or selling based on stock analysis information. By studying and evaluating past and current data, investors and traders attempt to gain an edge in the markets by making informed decisions.

Problem Statement:

The assessment to identify the slide/rise of stock during the covid period and the time taken to overcome this slide and also if Netflix stock prices drives down after the pandemic.

Objective:

- To determine whether the trends in price and volume of the stock over the 5 years of time move upward or downward or if it is stable over the years.
- To measure the changes in the volume and prices of the stock over these 5 years.
- To determine the strength and relationship between the price and the volume of the stock.

Dataset:

Kaggle Source: https://www.kaggle.com/datasets/jainilcoder/netflix-stock-price-prediction

6 columns – Date, Opening, Closing, Highest and Lowest Prices, Volume of Stock.

Statistical Tools Description:

Packages Used:

- Hmisc
- Psych
- ggplot2
- qcc
- Forecast
- qicharts

Software Used:

The analysis is undertaken using R programming language in RStudio.

Analysis and Prediction:

1) BASIC FUNCTIONS
    - Descriptive Statistics
    - Covariance
    - Median of a particular set of data
    - Index Numbers
    - Parametric Tests

Code:

```
library(Hmisc)

library(psych)

data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\NFLX.csv", header=TRUE)

class(data)

summary(data)
```

median(data$High[data$Date <= '31-12-2018'])

print(cov(data$High, data$Low))

Output:

```
> class(data)
[1] "data.frame"
> summary(data)
     Date               Open            High            Low             Close          Adj.Close         Volume
 Length:747         Min.   :233.9   Min.   :250.7   Min.   :231.2   Min.   :233.9   Min.   :233.9   Min.   : 1144000
 Class :character   1st Qu.:314.5   1st Qu.:319.7   1st Qu.:307.3   1st Qu.:315.4   1st Qu.:315.4   1st Qu.: 5227850
 Mode  :character   Median :356.4   Median :362.2   Median :350.5   Median :356.9   Median :356.9   Median : 7131300
                    Mean   :372.3   Mean   :378.5   Mean   :366.0   Mean   :372.5   Mean   :372.5   Mean   : 8605947
                    3rd Qu.:417.6   3rd Qu.:425.5   3rd Qu.:411.6   3rd Qu.:419.5   3rd Qu.:419.5   3rd Qu.:10478450
                    Max.   :673.1   Max.   :691.0   Max.   :671.2   Max.   :690.3   Max.   :690.3   Max.   :58410400
> median(data$High[data$Date <= '31-12-2018'])
[1] 362.225
> print(cov(data$High, data$Low))
[1] 6146.443
>
```

Report:

The datatype, summary of the data, median of a particular time period and the covariance of the High and Low prices are represented using the R language.

2) Correlation and Variance

Code:

cor(data$High, data$Low)

var(data$Volume)

Output:

```
> cor(data$High, data$Low)
[1] 0.9966744
> var(data$Volume)
[1] 2.68917e+13
>
```

Report:

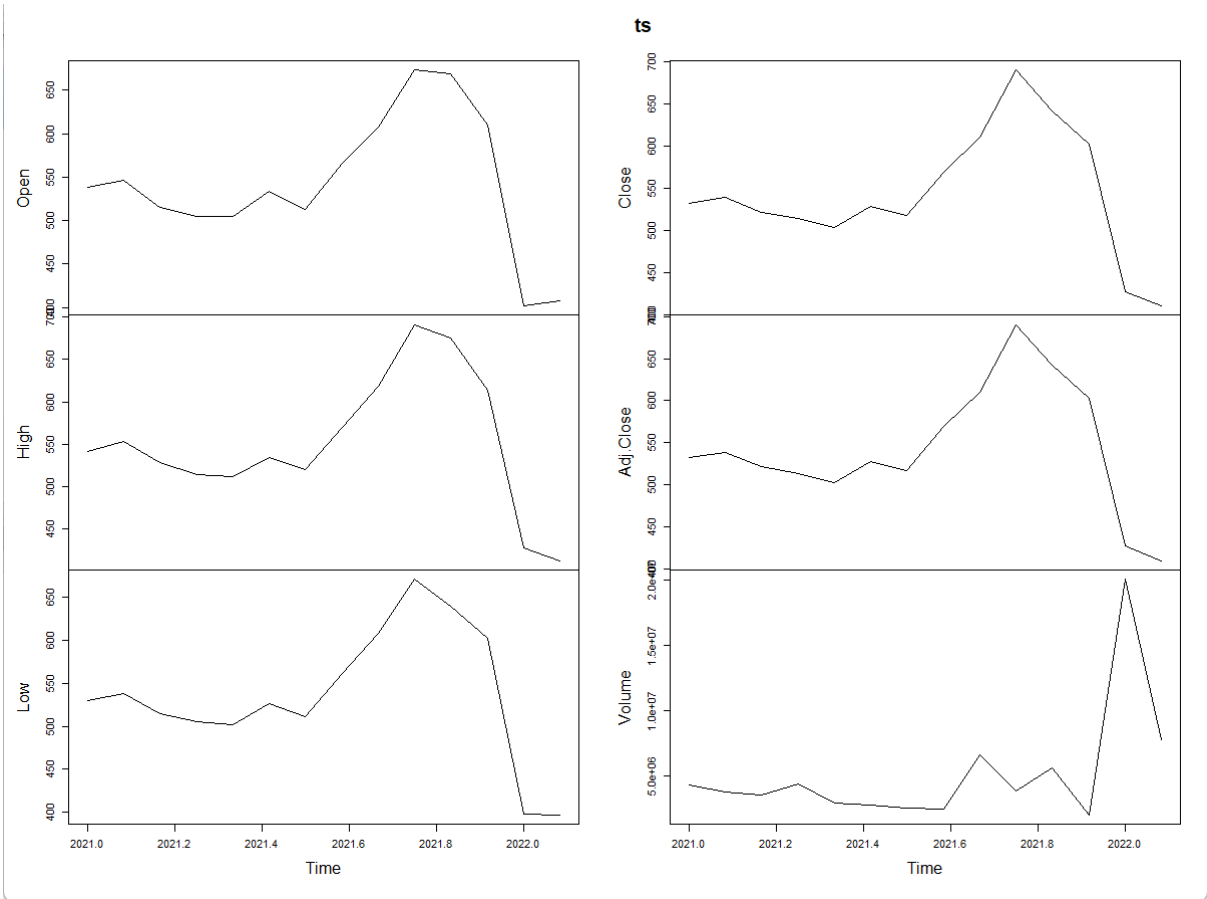
The correlation relationship between the High and Low prices and the variance for the volume are represented. The relationship between the High and Low Prices is positive.

3) Time Series Data

Code:

```
data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

ts = ts(data, frequency=12, start=c(2021,1), end=c(2022,2))

ts

plot.ts(ts)
```

Output:



Report:

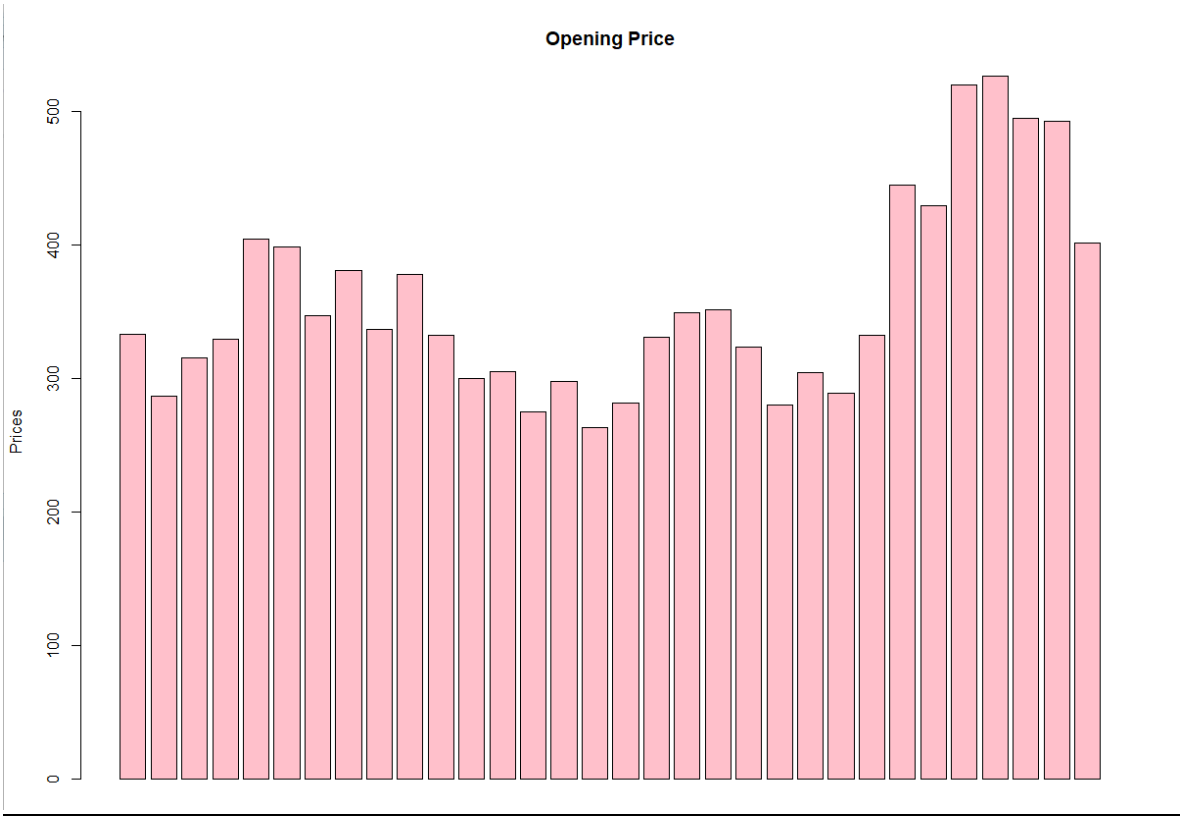
The time series analysis represents the plotting of all the factors in the time period of 2021-22. The factors are increasing healthily till the end of 2021 but at the end of 2021, all the factors are reduced drastically.

4) Plotting

Code:

```
data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\NFLX.csv", header=TRUE)
d1 <- subset(data, data$Volume >= 19000000); d1
print(barplot(d1$Open, type='l', main='Opening Price', ylab='Prices', col='pink'))
print(barplot(d1$Close, type='l', main='Closing Price', ylab='Prices', col='lightgreen'))
```

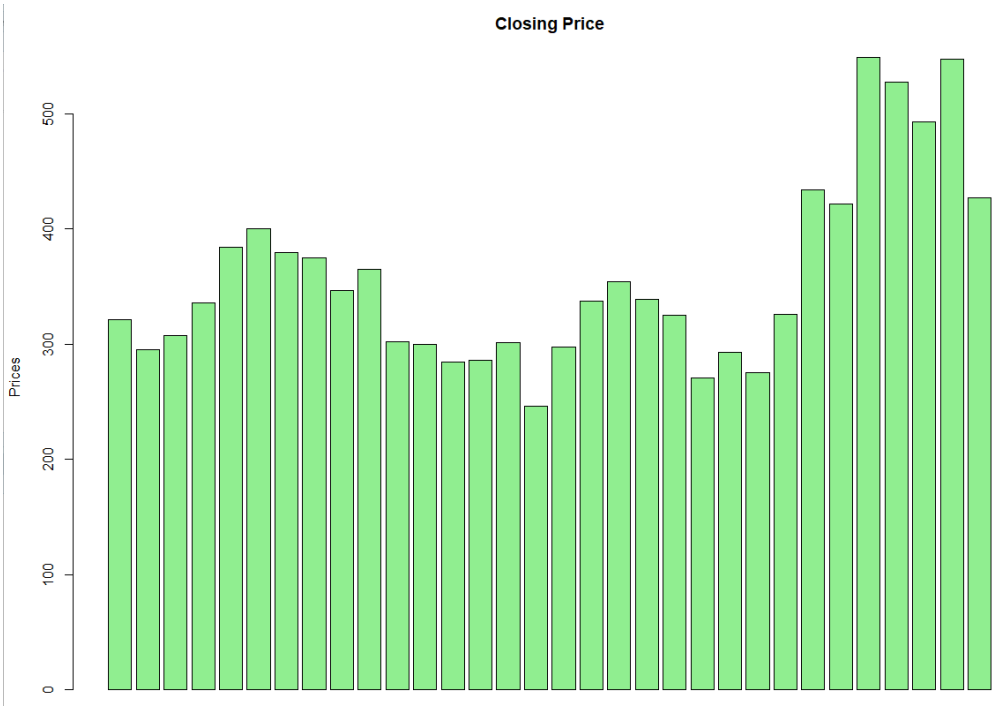
Barplot for Opening Prices:



Report:

The opening prices of the stock are in a good state as it has been periodically increasing with the maximum volume.

Barplot for Closing Prices:



Report:

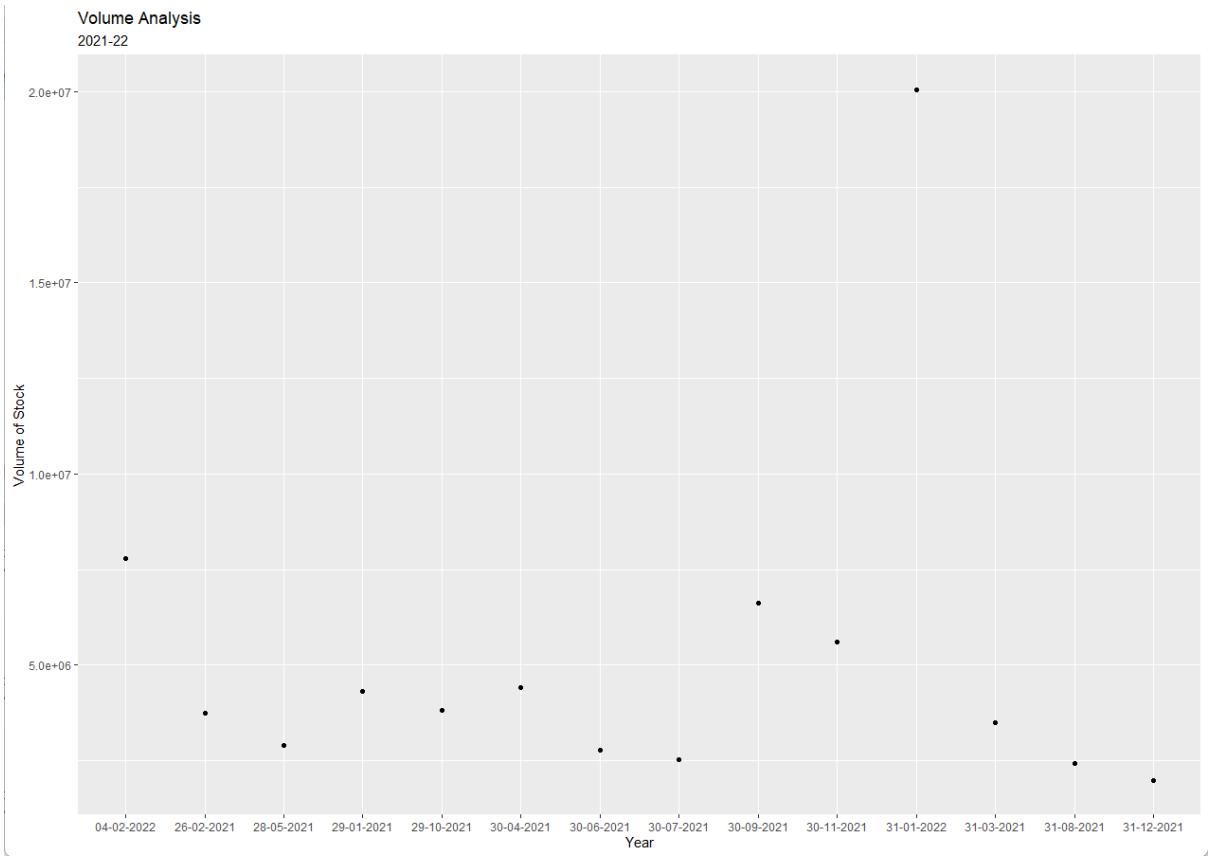
The closing prices of the stock are in a good state as it has been continuously increasing with the maximum volume which depicts the profitability of the company.

ggPlot:

Code:

```
data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

ggplot(data=data, aes(x = Date, y = Volume)) + geom_point() + labs(x = "Year",
y = "Volume of Stock", title = "Volume Analysis", subtitle = "2021-22")
```

Output:



Report:

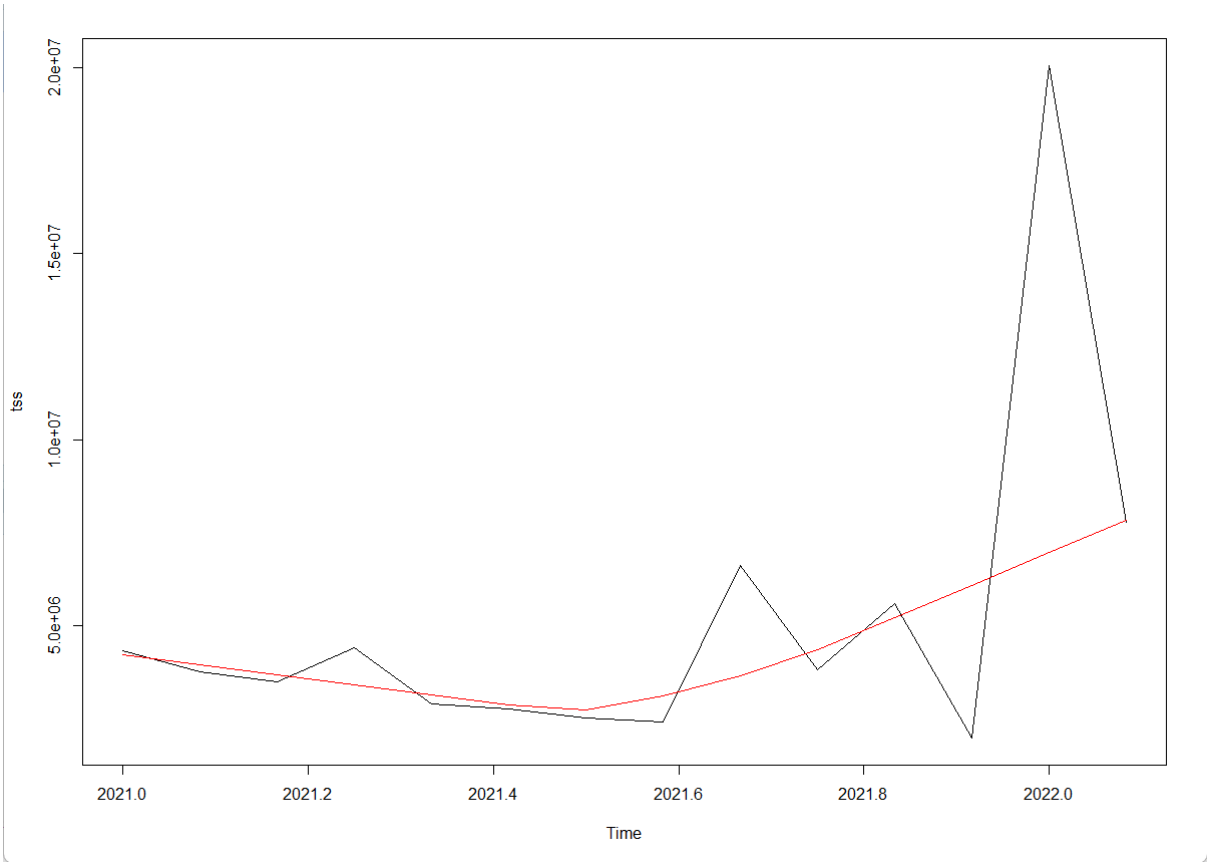
The points of the ggplot depicting the volume of the stocks are fluctuating in the time period of 2021-22.

5) Linear Trendline

Code:

```
data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

tss = ts(data$Volume, frequency=12, start=c(2021,1), end=c(2022,2))

plot.ts(tss)

lines(lowess(time(tss), tss), col="red")
```

Output:



Report:

The linear trendline representing the volume of the stocks has been low at the initial stage and it has moved upward at the end of June, 2021 and had reached the maximum at the start of the 2022.

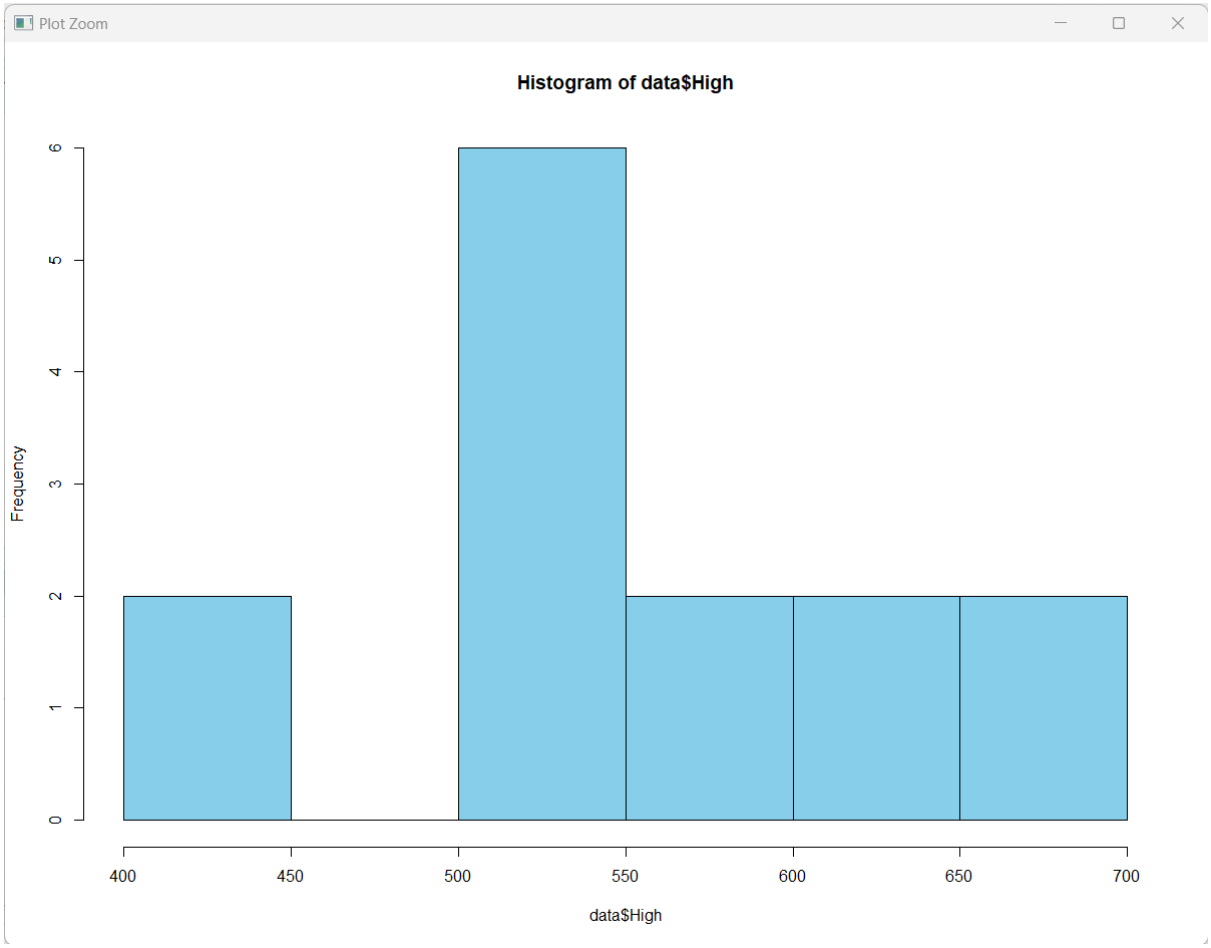
6) Mann-Whitney U-Test:

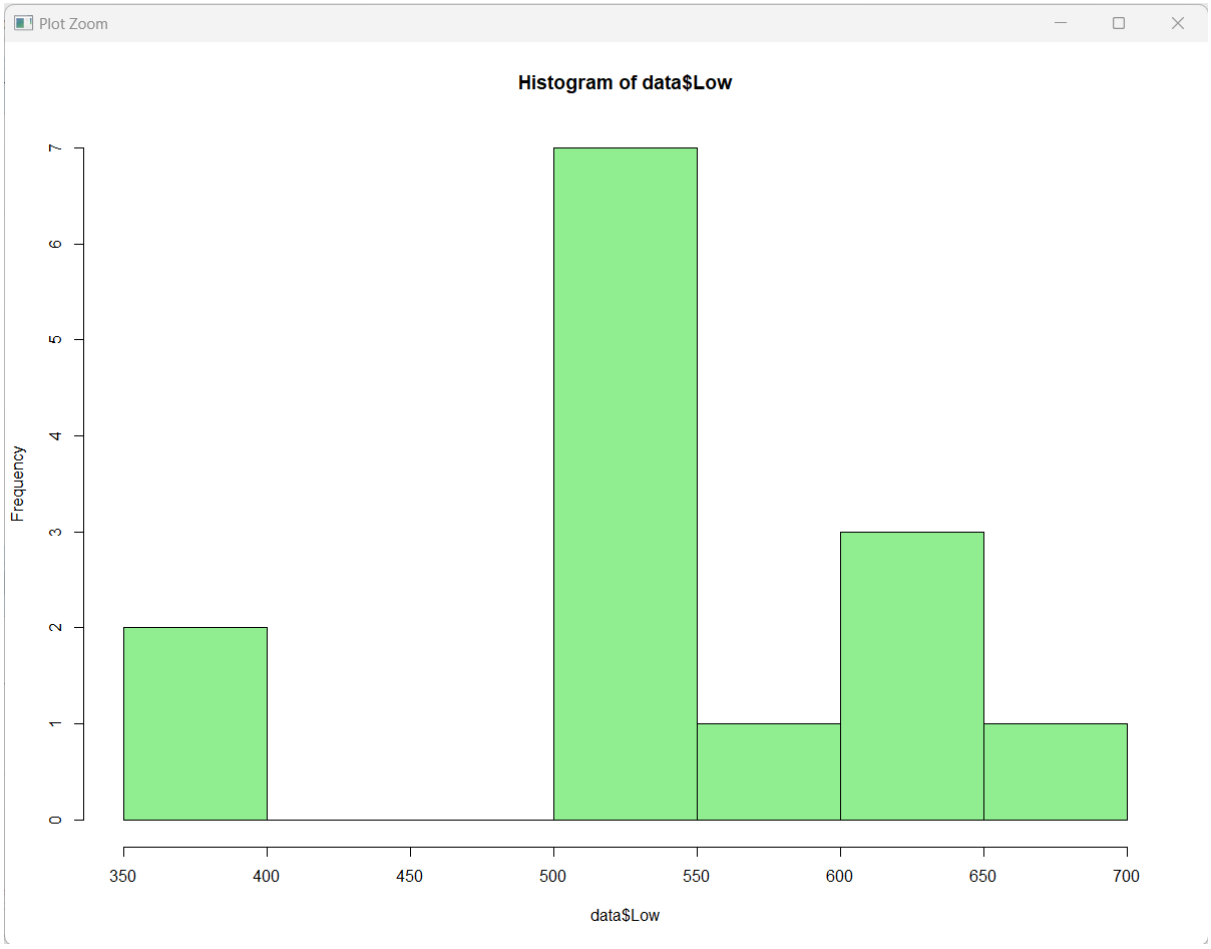
Code:

```
data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

high_val=data$High

low_val=data$Low

hist(data$High, col="skyblue")

hist(data$Low, col="lightgreen")

boxplot(low_val,high_val, col= c("lavender","pink"))

wilcox.test(low_val,high_val,paired=FALSE)
```

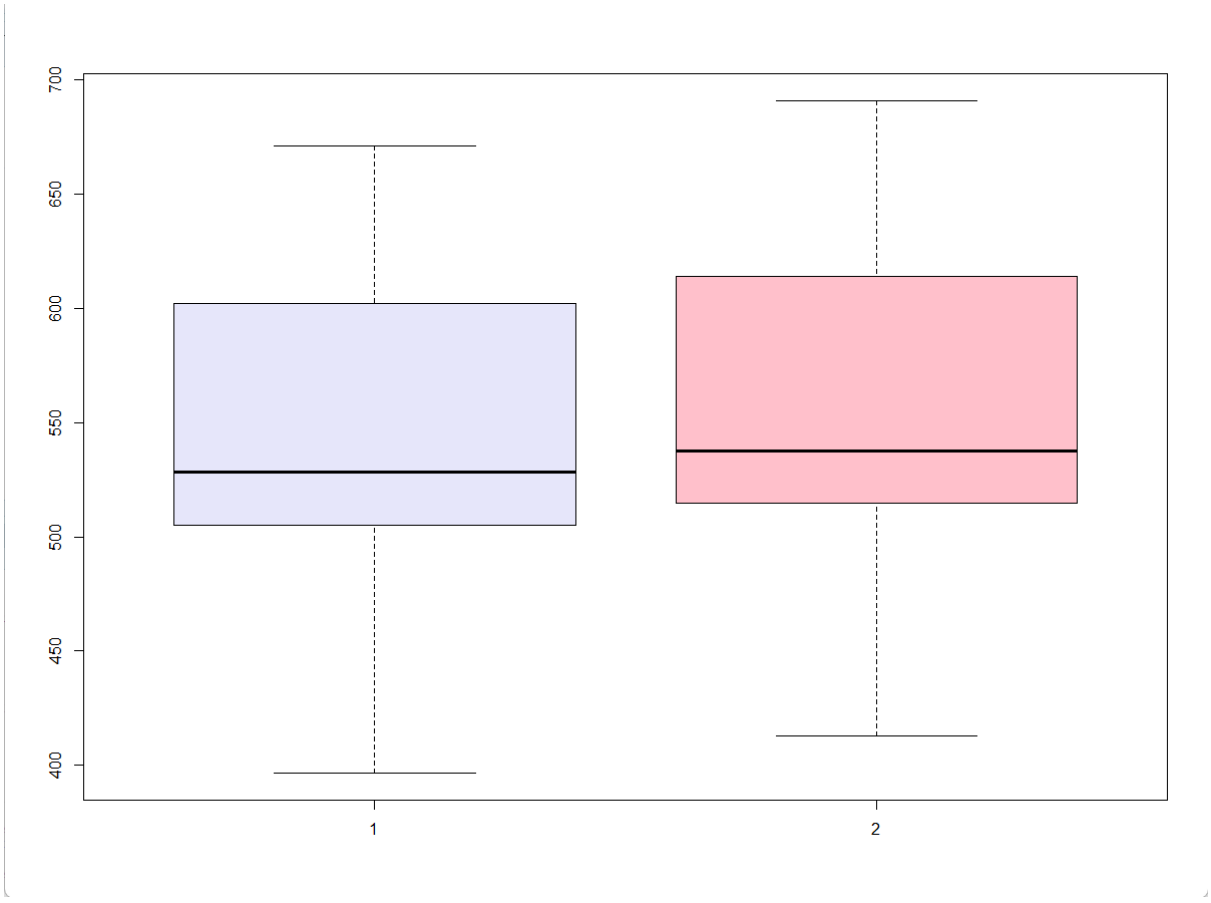
Output:

HISTOGRAM:




Plot Zoom
Histogram of data$Low
Frequency
0
1
2
3
4
5
6
7
350
400
450
500
550
600
650
700
data$Low

BOXPLOT:


400
450
500
550
600
650
700
1
2

Report:

The Mann-Whitney U-test representing the Highest and Lowest prices of the stock depicts the mean rank and sum of ranks of the two variables tested. Here, the lowest prices have more effect on the stock while being compared to the Highest prices so the Netflix has done lower sales in the years 2021 and 2022.

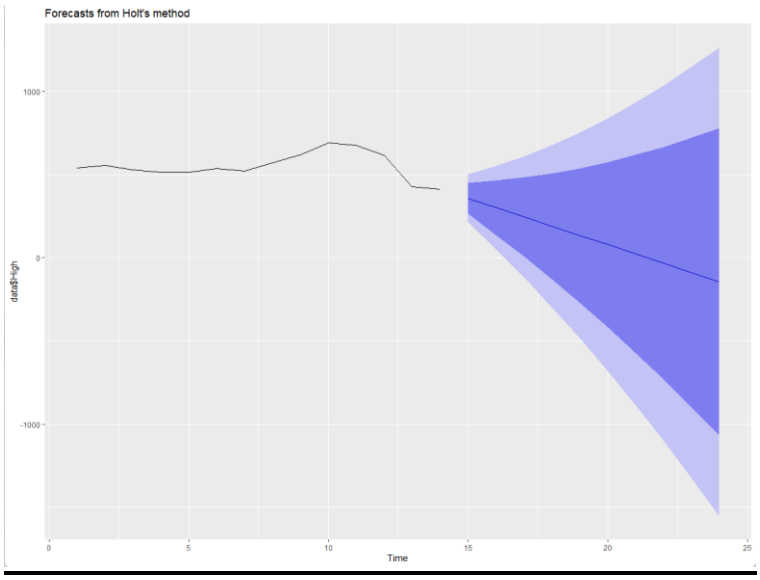
7) Holt's Forecast Method

Code:

```
library(forecast)

data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

holt_mod <- holt(data$High, h = 10)

summary(holt_mod)

autoplot(holt_mod)


holt_mod <- holt(data$Volume, h = 10)

summary(holt_mod)

autoplot(holt_mod)
```

Output:

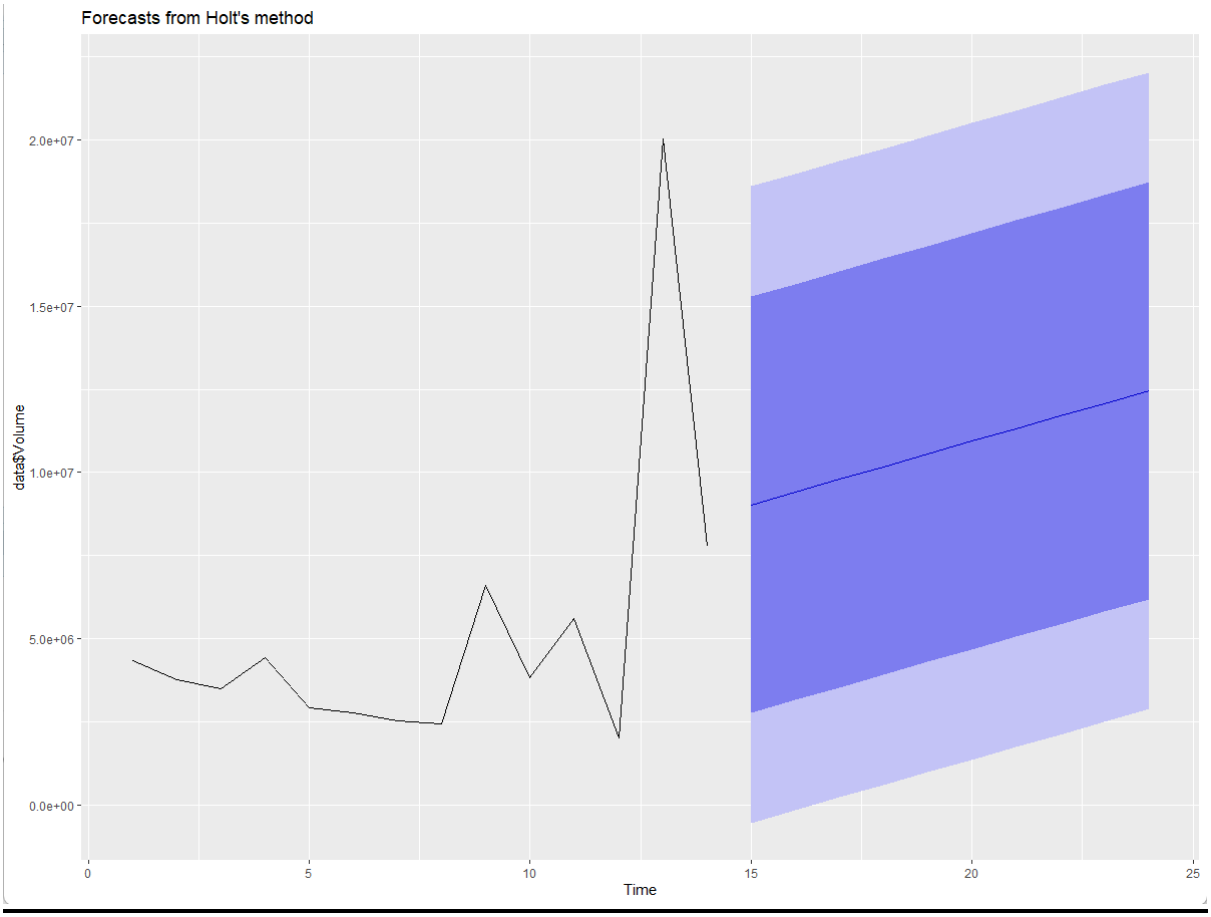
Forecasting High Prices for next 10 years



Report:

The Holt's Forecast Method representing the high prices of the stocks will get a reduction in the price gradually in the next 10 years and will result in a greater loss for the company.

Forecasting Volume for next 10 years



Report:

The Holt's Forecast Method representing the volume of the stocks will be boosted in the next 10 years and the company will gain the volume of stock which will result in maximum sales but as mentioned before, the prices may get reduced.

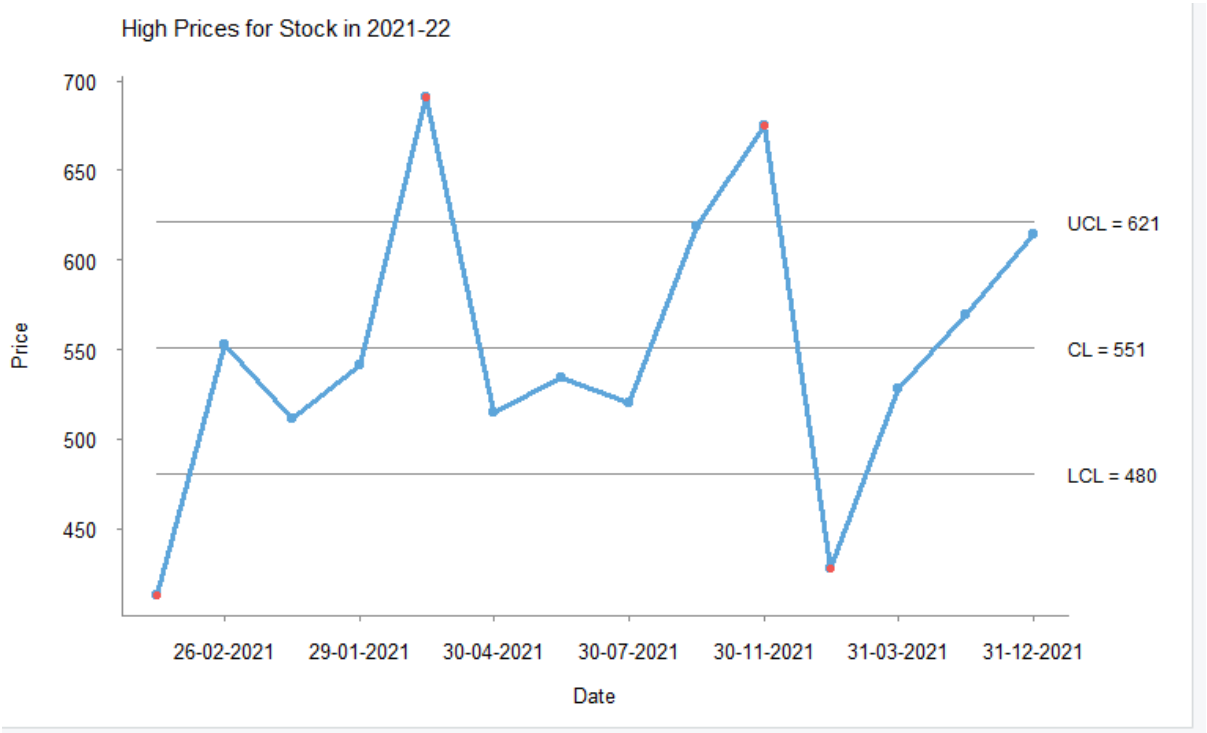
8) Control Charts using qcc

- **c chart:**

Code:

```
library(qicharts2)

data<-read.csv(file="E:\\sanjay\\R\\project\\netflix\\2021-22.csv",
header=TRUE)

qic(data,x= Date, y = High, data=data, chart='c', main='High Prices for Stock in
2021-22', ylab='Price', xlab  = 'Date')
```

Output:



Report:

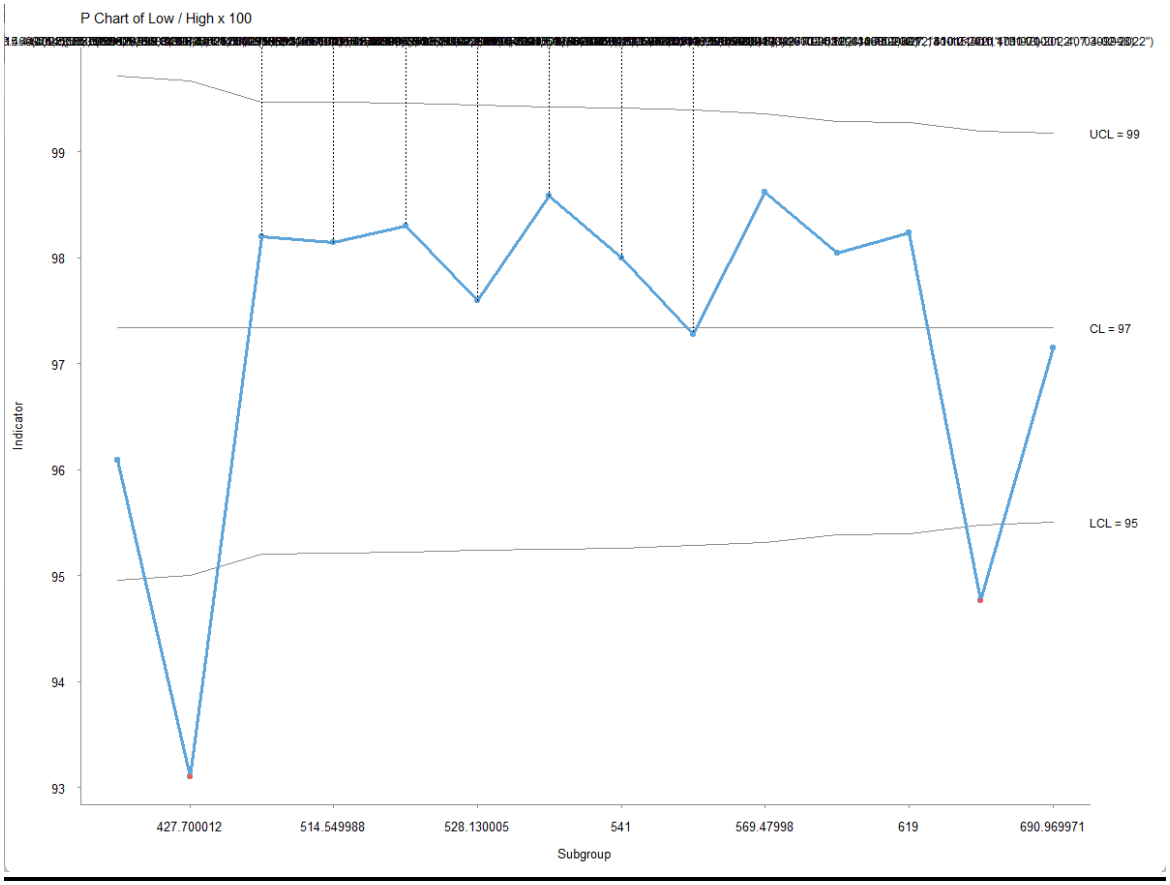
The c-bar chart representing the data of the Netflix's stocks from 2021-22 is under the statistical control as all the points in the chart lie within the control limits.

- **p-chart:**

Code:

```
qic(d1,
    n       = High,
    x       = High,
    y =   Low,
    data     = d1,
    chart    = 'p',
    multiply = 100,
)
```

Output:



Report:

The p chart representing the Highest and lowest prices of the Netflix's stocks from 2021-22 is out of the statistical control as two points in the chart lie outside the LCL limits.

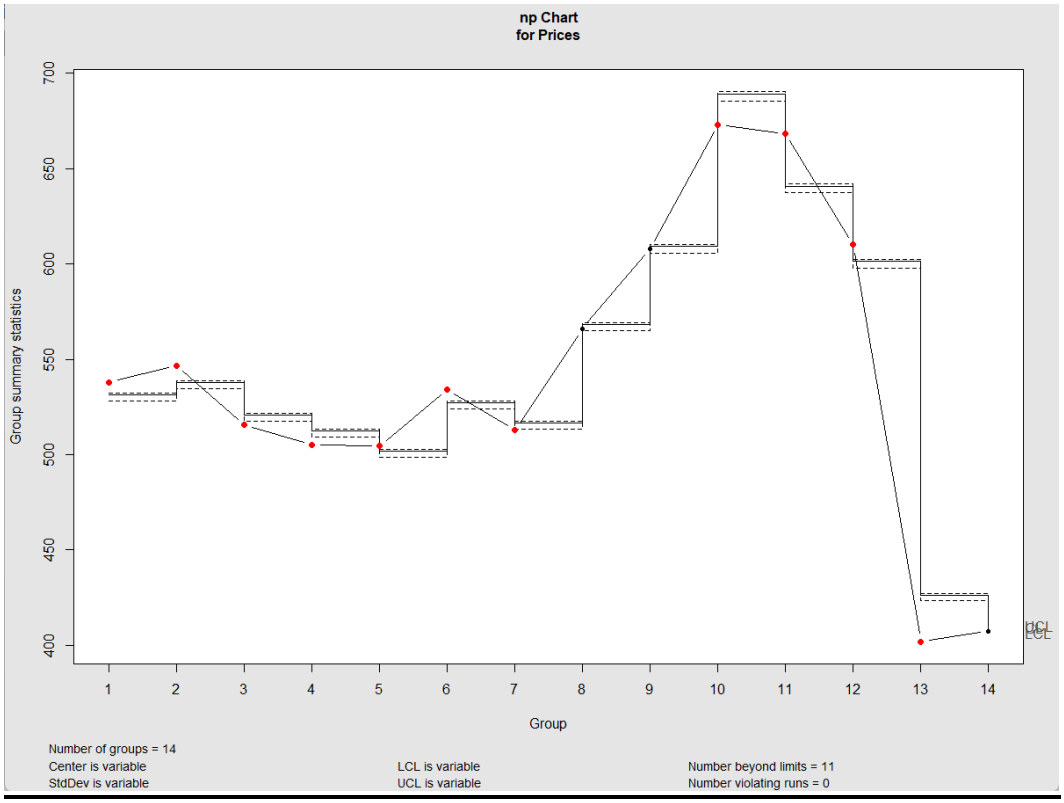
- **np-chart:**

Code:

np_chart <- with(d1, qcc(d1$Open, d1$Close, type = "np", data.name = "Prices"))

summary(np_chart)

Output:




```
R 4.2.1 · ~/
> summary(np_chart)

Call:
qcc(data = d1$Open, type = "np", sizes = d1$Close, data.name = "defects")

np chart for defects

Summary of group statistics:
     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
401.9700 506.9225 536.0300 542.2179 597.5675 673.0600

Summary of group sample sizes:
  sizes  410.17 427.14 502.81 513.47 517.57 521.66 528.21 532.39 538.85 569.19 602.44 610.34 641.9 690.31
  counts   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.0   1.00

Number of groups:  14
Center of group statistics: 531.3121 537.7590 520.6038 512.4304 501.7920 527.1406 516.5221 568.0376 609.1043 688.9124 640.6004 ...
Standard deviation: 1.0371581 1.0434315 1.0266533 1.0185622 1.0079337 1.0330786 1.0226207 1.0724046 1.1104933 1.1810060 ...

Control limits:
         LCL    UCL
    528.2007 532.39
    534.6287 538.85
...
    406.6085 410.17
```

Report:

The np chart representing the opening and closing prices of the Netflix's stocks from 2021-22 is out of the statistical control as the points in the chart lie outside both the UCL and LCL limits.

Conclusion:

Stock market analysis is mostly used to gain knowledge of the market situations to arrive at true value of a specific stock. Using the Above methods, we have forecasted the future of the stocks of Netflix in the market by predicting the high prices and volume of stock and also analyzed the trends of their stock in the market. This analysis helps the investors or traders to make a clear decision in the market regarding the stocks and also helps the company to improve their value of stocks in the future.

References:

Intro: https://lamfo-unb.github.io/2017/07/22/intro-stock-analysis-1/

Forecast: https://www.youtube.com/watch?v=JjrrwEn-2uI

Comparison:https://medium.com/codex/stock-market-analysis-with-r-programming-language-c3ab502eb3e7

Time Series: https://rpubs.com/kapage/523169

Charts: https://luca-scr.github.io/qcc/articles/qcc.html