678midterm

Ruining Jia

Introduction

This project will focus on whether COVID-19 has had a significant impact on the share prices of some companies, and look at how current share prices compare to those before the pandemic. These companies include Apple, Amazon, Facebook, Google and Netflix.

```
cases <-read.csv("time_series_covid19_confirmed_US.csv")
Amazon <- read.csv("Amazon.csv")
Apple <- read.csv("Apple.csv")
Facebook <- read.csv("Facebook.csv")
Google <- read.csv("Google.csv")
Netflix <- read.csv("Netflix.csv")
```

Data cleaning

```
cases = cases %>% select("Province_State" | starts_with("X"))
cases = cases %>% group_by(Province_State) %>% summarise(across(starts_with("X"), sum))
head(cases)
```

```
## # A tibble: 6 x 689
##   Province_State X1.22.20 X1.23.20 X1.24.20 X1.25.20 X1.26.20 X1.27.20 X1.28.20
##   <chr>             <int>    <int>    <int>    <int>    <int>    <int>    <int>
## 1 Alabama               0        0        0        0        0        0        0
## 2 Alaska                0        0        0        0        0        0        0
## 3 American Samoa        0        0        0        0        0        0        0
## 4 Arizona               0        0        0        0        1        1        1
## 5 Arkansas              0        0        0        0        0        0        0
## 6 California            0        0        0        0        2        2        2
## # ... with 681 more variables: X1.29.20 <int>, X1.30.20 <int>, X1.31.20 <int>,
## #   X2.1.20 <int>, X2.2.20 <int>, X2.3.20 <int>, X2.4.20 <int>, X2.5.20 <int>,
## #   X2.6.20 <int>, X2.7.20 <int>, X2.8.20 <int>, X2.9.20 <int>, X2.10.20 <int>,
## #   X2.11.20 <int>, X2.12.20 <int>, X2.13.20 <int>, X2.14.20 <int>,
## #   X2.15.20 <int>, X2.16.20 <int>, X2.17.20 <int>, X2.18.20 <int>,
## #   X2.19.20 <int>, X2.20.20 <int>, X2.21.20 <int>, X2.22.20 <int>,
## #   X2.23.20 <int>, X2.24.20 <int>, X2.25.20 <int>, X2.26.20 <int>, ...
```

```
case_delta <- cases

for(i in 2:ncol(cases)) {       # for-loop over columns

  if(i > 3) {
    case_delta[ , i] = abs((cases[ , i] - cases[ , i-1]))
  }
}
head(case_delta)
```

```
## # A tibble: 6 x 689
##   Province_State X1.22.20 X1.23.20 X1.24.20 X1.25.20 X1.26.20 X1.27.20 X1.28.20
##   <chr>             <int>    <int>    <int>    <int>    <int>    <int>    <int>
## 1 Alabama               0        0        0        0        0        0        0
## 2 Alaska                0        0        0        0        0        0        0
## 3 American Samoa        0        0        0        0        0        0        0
## 4 Arizona               0        0        0        0        1        0        0
## 5 Arkansas              0        0        0        0        0        0        0
## 6 California            0        0        0        0        2        0        0
## # ... with 681 more variables: X1.29.20 <int>, X1.30.20 <int>, X1.31.20 <int>,
## #   X2.1.20 <int>, X2.2.20 <int>, X2.3.20 <int>, X2.4.20 <int>, X2.5.20 <int>,
## #   X2.6.20 <int>, X2.7.20 <int>, X2.8.20 <int>, X2.9.20 <int>, X2.10.20 <int>,
## #   X2.11.20 <int>, X2.12.20 <int>, X2.13.20 <int>, X2.14.20 <int>,
## #   X2.15.20 <int>, X2.16.20 <int>, X2.17.20 <int>, X2.18.20 <int>,
## #   X2.19.20 <int>, X2.20.20 <int>, X2.21.20 <int>, X2.22.20 <int>,
## #   X2.23.20 <int>, X2.24.20 <int>, X2.25.20 <int>, X2.26.20 <int>, ...
```

```
case_delta <- colSums(case_delta[,-1])
cases_vec <- as.numeric(case_delta)
cases_vec[1:100]
```

```
##   [1]     1     1     1     0     3     0     0     1     0     2     0     0
##  [13]     3     0     0     1     0     0     0     0     1     0     1     0
##  [25]     0     0     1     1     0     0     0     0     0     1     1     1
##  [37]     0     0     7    13    18    22    34    71    63   170   121    92
##  [49]   167   361   452   626   730   388  1433  1783  2706  4507  6401  5992
##  [61]  8888 11156 10618 12127 17821 18591 22164 20165 22154 26381 32232 32230
##  [73] 32482 32176 29604 31858 30343 31149 36443 34347 29318 26872 26924 28521
##  [85] 25900 29956 33157 27994 25923 29680 26184 29511 32812 32152 31800 26547
##  [97] 23818 24690 26220 29241
```

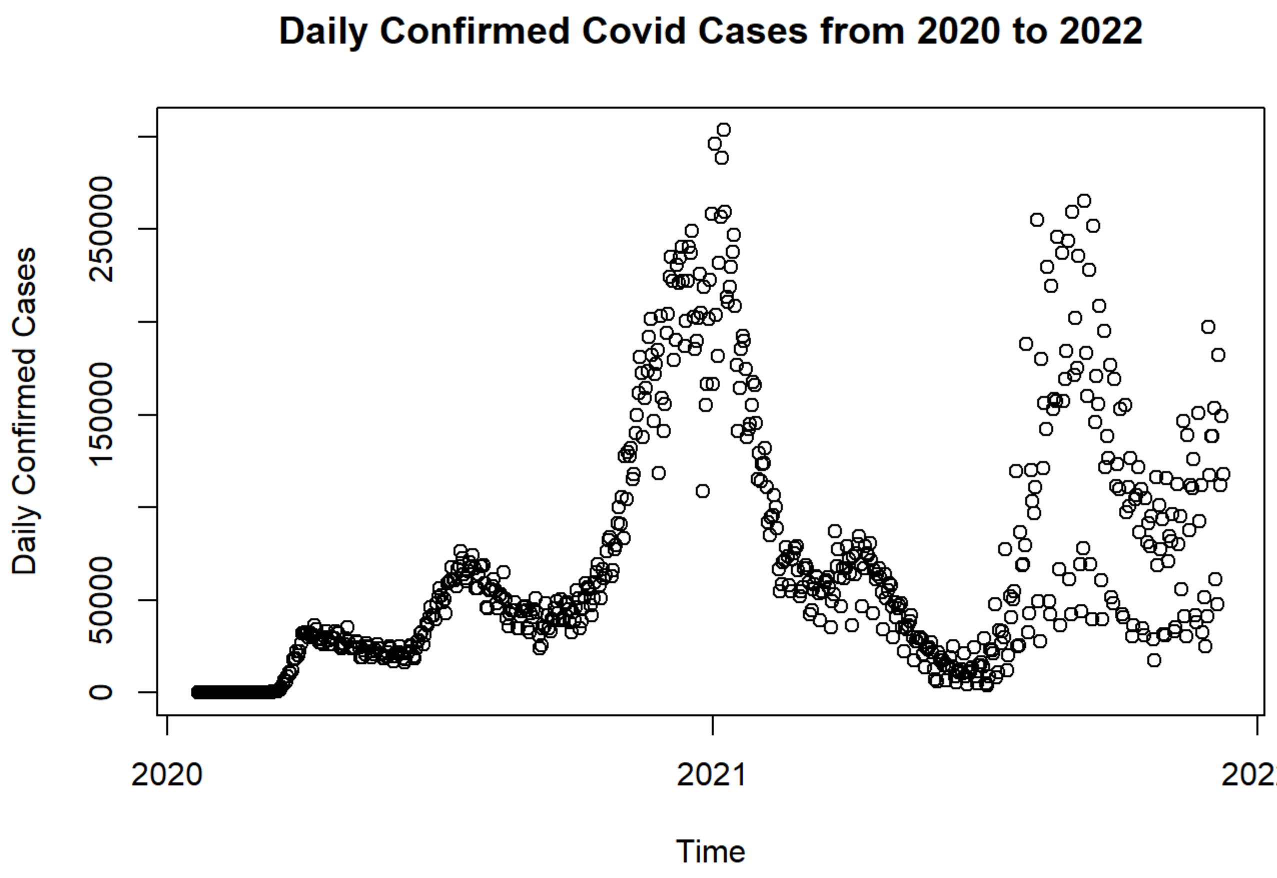
```
dates = names(case_delta)
dates = as.Date(dates, format="X%m.%d.%y")

head(dates)
```

```
## [1] "2020-01-22" "2020-01-23" "2020-01-24" "2020-01-25" "2020-01-26"
## [6] "2020-01-27"
```

A chart of daily confirmed cases over the past two years is shown below

```
plot(dates, cases_vec, xlab="Time", ylab="Daily Confirmed Cases")
title("Daily Confirmed Covid Cases from 2020 to 2022")
```



Let foucs on the open prices for the each stocks as the comparable vaule to find the relationship between the corfimed Covid-19 people and the open prices for the each stocks

```
Amazon = Amazon %>% select("Date" | "Open")
Apple = Apple %>% select("Date" | "Open")
Facebook = Facebook %>% select("Date" | "Open")
Google = Google %>% select("Date" | "Open")
Netflix = Netflix %>% select("Date" | "Open")
Amazon$Date <- as.Date(Amazon$Date)
Apple$Date <- as.Date(Apple$Date)
Facebook$Date <- as.Date(Facebook$Date)
Google$Date <- as.Date(Google$Date)
Netflix$Date <- as.Date(Netflix$Date)
Cases <- c(cases_vec)
Date <- c(dates)
df <- data.frame(Date, Cases)
head(df)
```

```
##         Date Cases
## 1 2020-01-22     1
## 2 2020-01-23     1
## 3 2020-01-24     1
## 4 2020-01-25     0
## 5 2020-01-26     3
## 6 2020-01-27     0
```

The project creates a new data frame that includes only the COVID-19 case and its associated date, and the data frame is adjusted to include only dates in the range of dates for which we are viewing the COVID-19 data.Limit the time frame for all studies to August 1

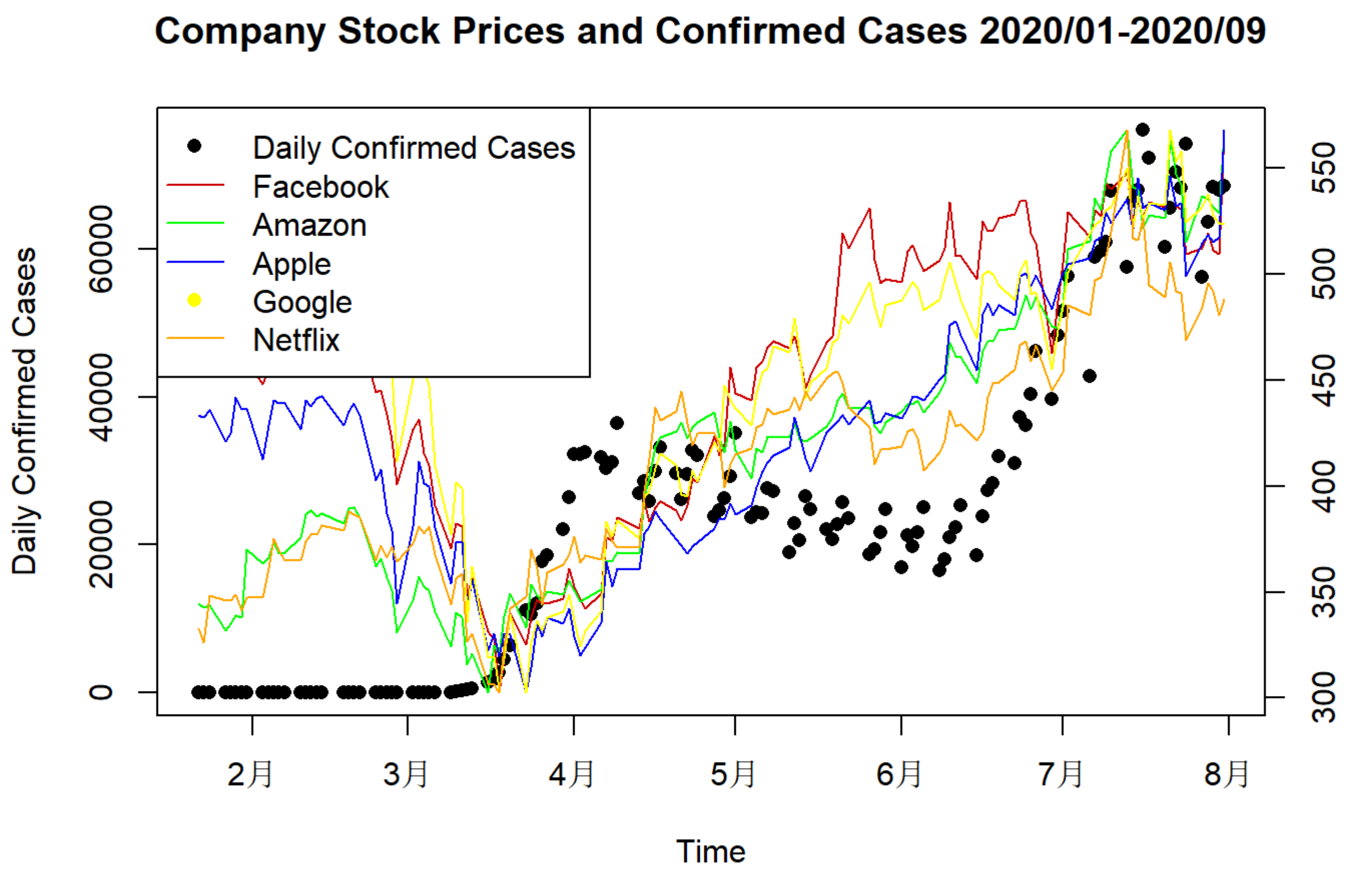
```
Amazon = Amazon %>% filter(Amazon$Date %in% dates)
Apple = Apple %>% filter(Apple$Date %in% dates)
Facebook = Facebook %>% filter(Facebook$Date %in% dates)
Google = Google %>% filter(Google$Date %in% dates)
Netflix = Netflix %>% filter(Netflix$Date %in% dates)
df = df %>% filter(df$Date %in% Google$Date)
df = df %>% arrange(desc(df$Date))

Amazon<-subset(Amazon,Amazon$Date<="2020-08-01")
Apple<-subset(Apple,Apple$Date<="2020-08-01")
Facebook<-subset(Facebook,Facebook$Date<="2020-08-01")
Google<-subset(Google,Google$Date<="2020-08-01")
Netflix<-subset(Netflix,Netflix$Date<="2020-08-01")
df<-subset(df,df$Date<="2020-08-01")
```

Exploratory Data Analysis

The chart below plots the change of COVID-19 cases over time, adjusted to match the stock price dates of five companies.

```
plot(df$Date, df$Cases, xlab="Time", ylab="Daily Confirmed Cases", pch=16)
par(new=T)
plot(Facebook$Date, Facebook$Open, type='l', col='red3', axes=F, xlab=NA, ylab=NA)
mtext(side=4, line=2, text='Opening Stock Price ($USD)')
par(new=T)
plot(Amazon$Date, Amazon$Open, type='l', col='green1', axes=F, xlab=NA, ylab=NA)
par(new=T)
plot(Apple$Date, Apple$Open, type='l', col='blue1', axes=F, xlab=NA, ylab=NA)
par(new=T)
plot(Google$Date, Google$Open, type='l', col='yellow1', axes=F, xlab=NA, ylab=NA)
par(new=T)
plot(Netflix$Date, Netflix$Open, type='l', col='orange1', axes=F, xlab=NA, ylab=NA)
par(new=T)
axis(side=4)
legend("topleft",
          legend=c("Daily Confirmed Cases", "Facebook", "Amazon", "Apple","Google","Netflix"),
          col=c("black", "red3", "green1", "blue1","yellow1","orange1"), lty=c(0,1,1,1), pch=c(
16, NA, NA, NA))
title("Company Stock Prices and Confirmed Cases 2020/01-2020/09")
```



Relationships

Now, the project had to translate these scenarios into something visually easier to compare, and look at the relationships between various and businesses with COVID-19 cases, and therefore conduct correlation analyses.

```
cor(df$Cases, Google$Open)
```

```
## [1] 0.002891683
```

```
cor(df$Cases, Netflix$Open)
```

```
## [1] -0.645861
```

```
cor(df$Cases, Amazon$Open)
```

```
## [1] -0.6726013
```

```
cor(df$Cases, as.numeric(Apple$Open))
```

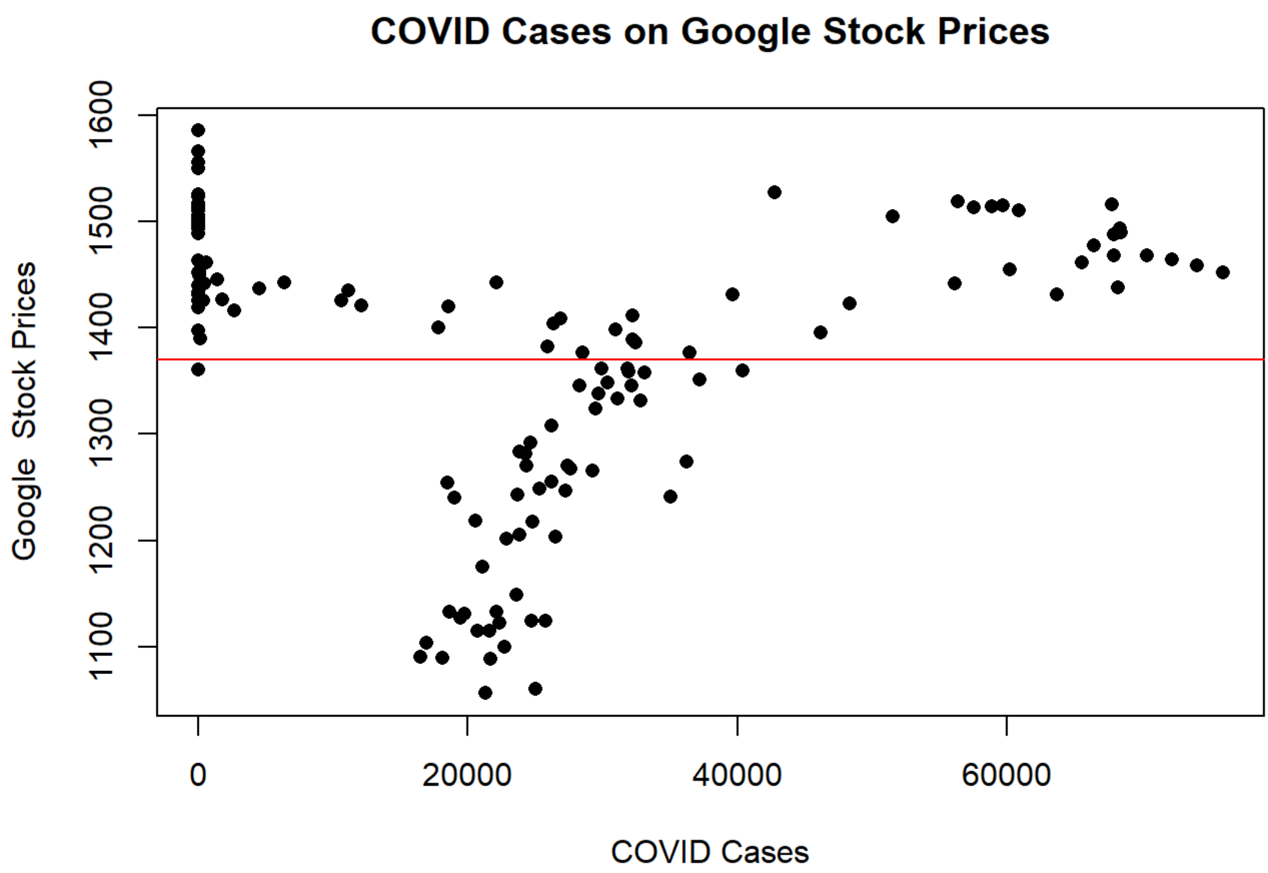
```
## [1] -0.3858518
```

```
cor(df$Cases, Facebook$Open)
```

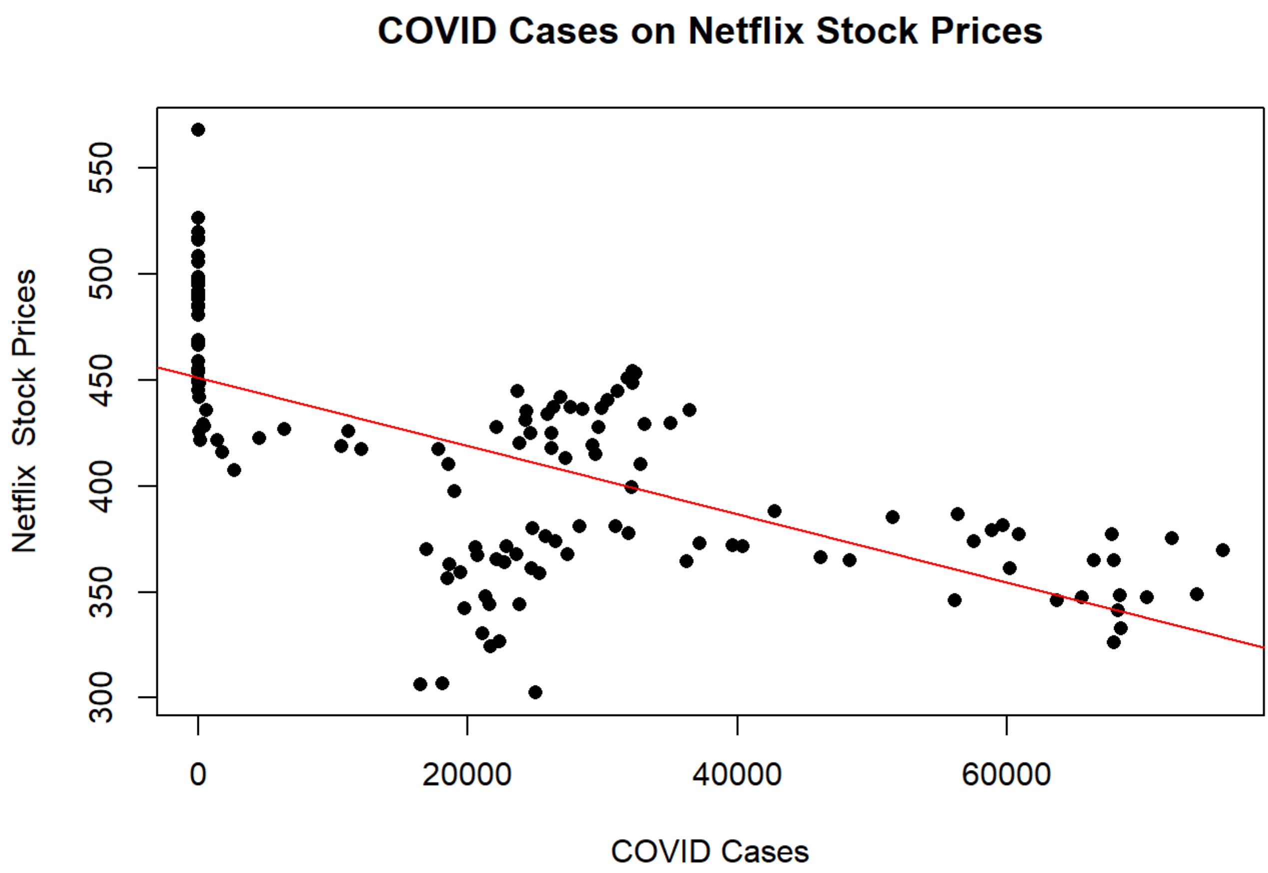
```
## [1] -0.2771968
```

The project uses correlation coefficients to measure the degree of association between two variables, in this case COVID-19 cases and their impact on a given or vaccine company. The figure below shows these associations in visual form. The red line represents the best fit line, also known as the regression line.

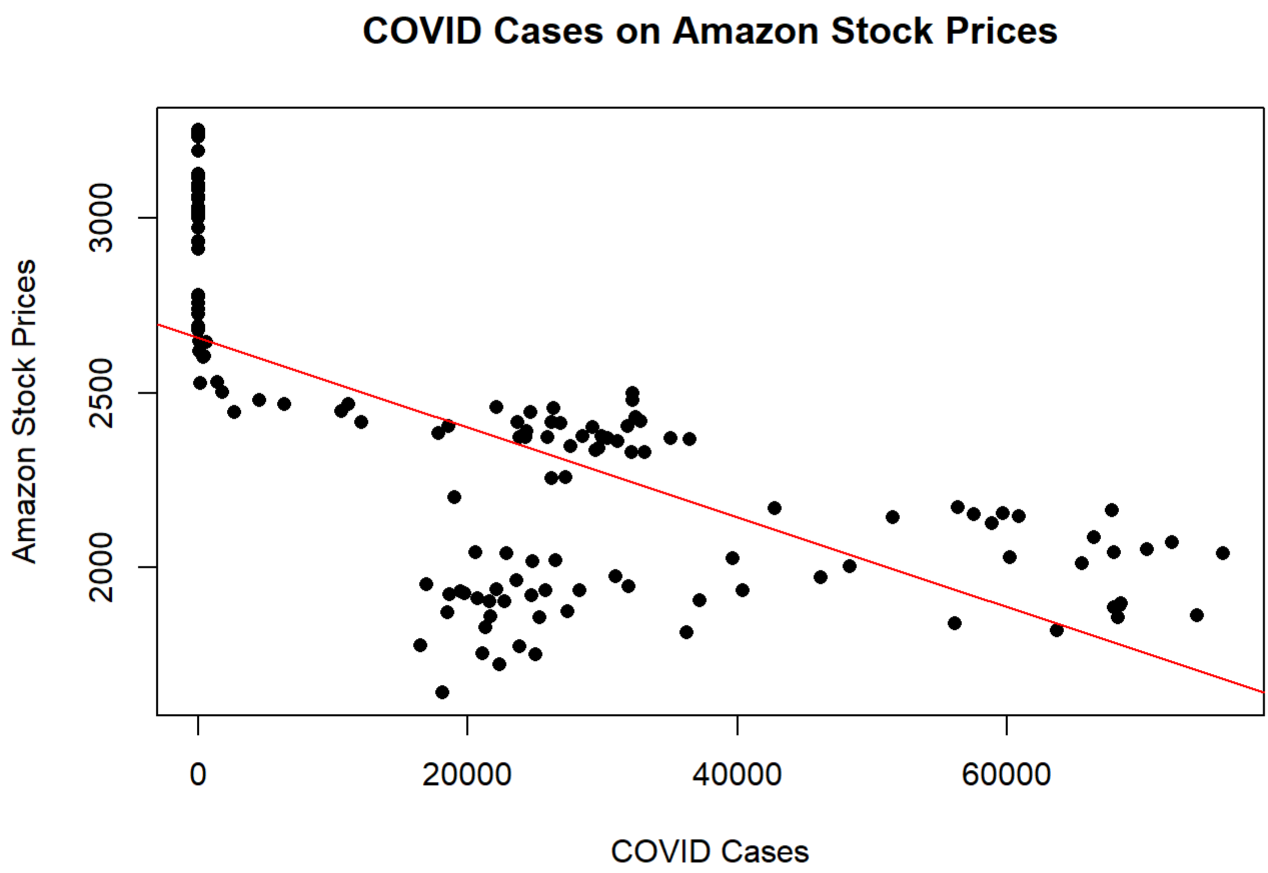
```
plot(df$Cases, Google$Open, xlab="COVID Cases", ylab="Google  Stock Prices", pch=16)
abline(lm(Google$Open~df$Cases), col = 'red')
title("COVID Cases on Google Stock Prices")
```



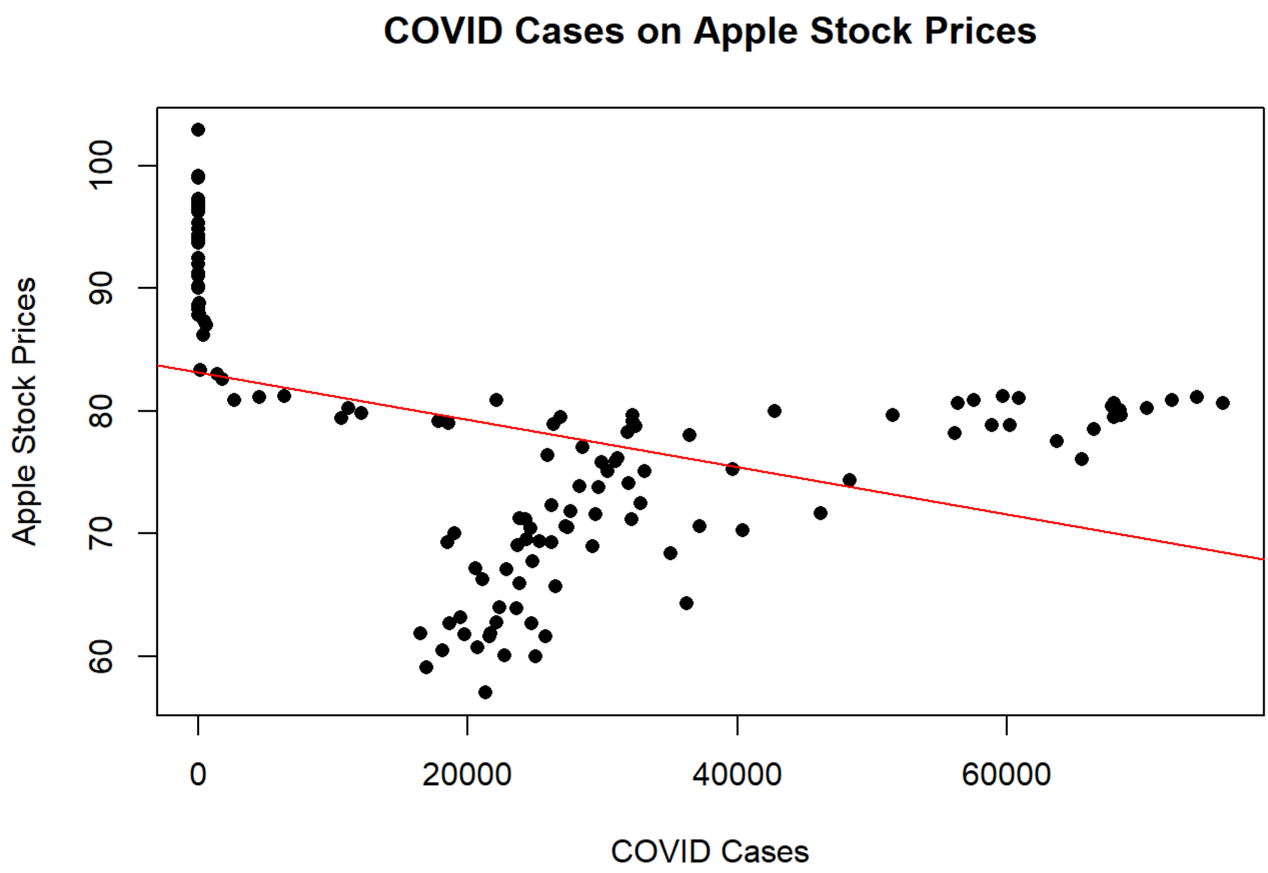

```
plot(df$Cases, Netflix$Open, xlab="COVID Cases", ylab="Netflix  Stock Prices", pch=16)
abline(lm(Netflix$Open~df$Cases), col = 'red')
title("COVID Cases on Netflix Stock Prices")
```



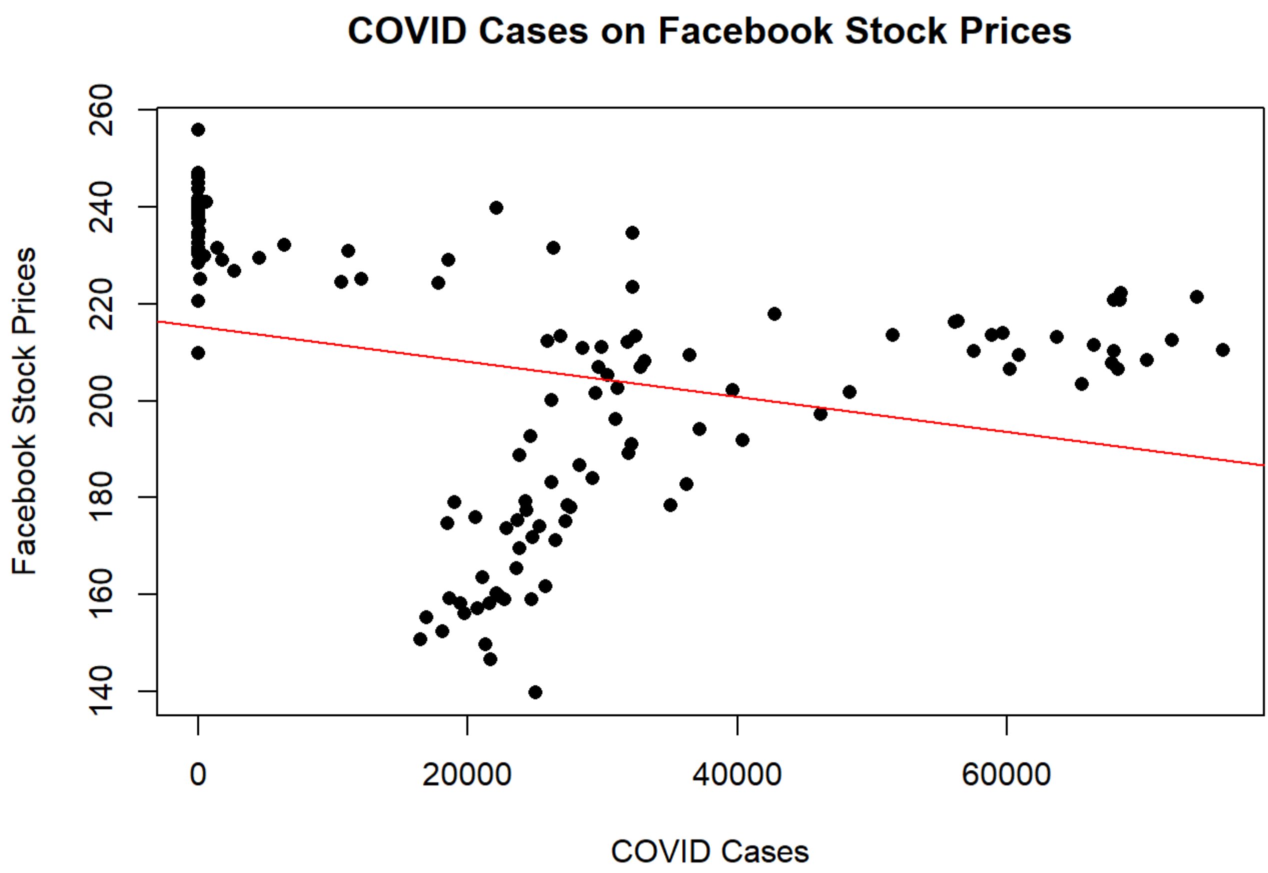

```
plot(df$Cases, Amazon$Open, xlab="COVID Cases", ylab="Amazon Stock Prices", pch=16)
abline(lm(Amazon$Open~df$Cases), col = 'red')
title("COVID Cases on Amazon Stock Prices")
```




```
plot(df$Cases, Apple$Open, xlab="COVID Cases", ylab="Apple Stock Prices", pch=16)
abline(lm(Apple$Open~df$Cases), col = 'red')
title("COVID Cases on Apple Stock Prices")
```




```
plot(df$Cases, Facebook$Open, xlab="COVID Cases", ylab="Facebook Stock Prices", pch=16)
abline(lm(Facebook$Open~df$Cases), col = 'red')
title("COVID Cases on Facebook Stock Prices")
```



Analysis

as COVID-19 began to rise between 2020 and 2022, most companies saw their stock prices rise,There are two major spikes in the nature of the COVID chart. The first is the first wave of the COVID-19 pandemic, which began to be most severe from early 2020 to late 2021, when the vaccine was delivered. The second peak can be attributed to the emergence of the COVID Delta variant, which first arrived in the United States in March and by June became the most common form of COVID infection, accounting for an estimated 82.2% of cases. While the share prices of most vaccine companies are on an upward trend, this is not entirely due to novel Coronavirus.

Discussion

In my opinion, the rise in stock prices is inseparable from isolation. As most people choose to isolate at home (I was also in the United States at the time, and my American classmates chose to isolate and order food at home, and then Rutgers was in online classes at the time). Everyone has chosen our home for online classes. So as more and more people choose to spend their leisure time at home, more people will choose to use amazon to shop online, and more people will buy new apple products, and there will be more People who go to play Facebook, more people will use Google’s series of products, and more people will use netflix to watch videos because they need more different kinds of elements to fill their isolated world at home.