# Introduction

Git hub https://github.com/tunghng/data-viz-proj/tree/main

Financial data, especially stock data, is crucial to investors and analysts because it provides valuable insights into an industry's performance and market trends. Effective data visualization can help these stakeholders regconize key patterns in stock performance, thus help them making more informed decisions and understand more about the market behaviour. For that reason, out team project attempt to effectively visualize financial data of mainly about big tech companies from the year 2010 to 2023. This project includes four main datasets, which we will briefly introduce:

1. Big Tech Stock Stock Price: is the data we got from the following link in Github

```
path1 <- "data/big_tech_stock_prices.csv"
dataset1 <- read.csv(path1)
head(dataset1, 5)
```

```
##   stock_symbol       date     open     high      low    close adj_close
## 1         AAPL 2010-01-04 7.622500 7.660714 7.585000 7.643214  6.515213
## 2         AAPL 2010-01-05 7.664286 7.699643 7.616071 7.656429  6.526476
## 3         AAPL 2010-01-06 7.656429 7.686786 7.526786 7.534643  6.422664
## 4         AAPL 2010-01-07 7.562500 7.571429 7.466071 7.520714  6.410790
## 5         AAPL 2010-01-08 7.510714 7.571429 7.466429 7.570714  6.453412
##      volume
## 1 493729600
## 2 601904800
## 3 552160000
## 4 477131200
## 5 447610800
```

2. COVID-19 Cases Datasets from Kaggle link

```
dataset2 <- read.csv("data/day_wise.csv")
head(dataset2, 5)
```

```
##         Date Confirmed Deaths Recovered Active New.cases New.deaths
## 1 2020-01-22       555     17        28    510         0          0
## 2 2020-01-23       654     18        30    606        99          1
## 3 2020-01-24       941     26        36    879       287          8
## 4 2020-01-25      1434     42        39   1353       493         16
## 5 2020-01-26      2118     56        52   2010       684         14
##   New.recovered Deaths...100.Cases Recovered...100.Cases Deaths...100.Recovered
## 1             0               3.06                  5.05                  60.71
## 2             2               2.75                  4.59                  60.00
## 3             6               2.76                  3.83                  72.22
## 4             3               2.93                  2.72                 107.69
## 5            13               2.64                  2.46                 107.69
##   No..of.countries
## 1                6
## 2                8
## 3                9
## 4               11
## 5               13
```

3. Walmart Stock Historical Data from Kaggle [link](link

```
dataset3 <- read.csv("data/WMT.csv")
head(dataset3, 5)
```

```
##         Date  Open  High   Low Close Adj.Close   Volume
## 1 2011-11-16 57.10 57.42 56.64 56.68  44.89946 11780800
## 2 2011-11-17 56.54 57.19 56.26 56.73  44.93906 10223800
## 3 2011-11-18 57.03 57.36 56.61 57.23  45.33513  8982300
## 4 2011-11-21 56.93 57.29 56.38 56.66  44.88361  9932200
## 5 2011-11-22 56.56 57.13 56.50 56.85  45.03411  7497300
```

4. Pfizer Stock Historical Prices from Kaggle link

```
dataset4 <- read.csv("data/pfizer.csv")
head(dataset4, 5)
```

```
##        Date     Open     High      Low    Close Adj.Close   Volume
## 1 1/22/2020 38.25427 38.33966 37.92220 38.13093  32.17976 18097812
## 2 1/23/2020 38.13093 38.73814 38.07400 38.62429  32.59610 27148510
## 3 1/24/2020 38.84251 38.87097 37.60911 37.77988  31.88349 34143698
## 4 1/27/2020 37.39089 38.35863 37.23909 38.10247  32.15573 31964026
## 5 1/28/2020 37.30550 37.46679 36.00569 36.18596  30.53834 70202408
```

Both dataset 1, 3, and 4 are financial data about stock prices such as open price, close price, high price, low price, adjusted close price and volume of stock daily (with data column). As dataset 1 is just about big tech stock companies so we include dataset 3 and 4 which are about stock price of companies in different sectors for more wholistic view. We also integrated these financial datasets with dataset three about the COVID-19 cases to see how the pandemic have affect the companies. For more detailed view about the datasets, please view the head function of each dataset.

# Question 1

## Question 1

This question analyses the stock performane of 14 big tech companies in dataset 1 and examines Apple Inc.'s stock (AAPL) as a potential long-term investment and assesses the inherent risks in its short-term fluctuations. The dataset crucial for this analysis includes AAPL's historical price trends, daily returns, and trading volumes, as detailed in the "AAPL_stock_data.csv". Understanding these aspects is fundamental in evaluating the stock's performance and market behavior over time. The interest in this question arises from Apple's significant role in the tech sector and its impact on investment portfolios, making it a focal point for both individual and institutional investors.

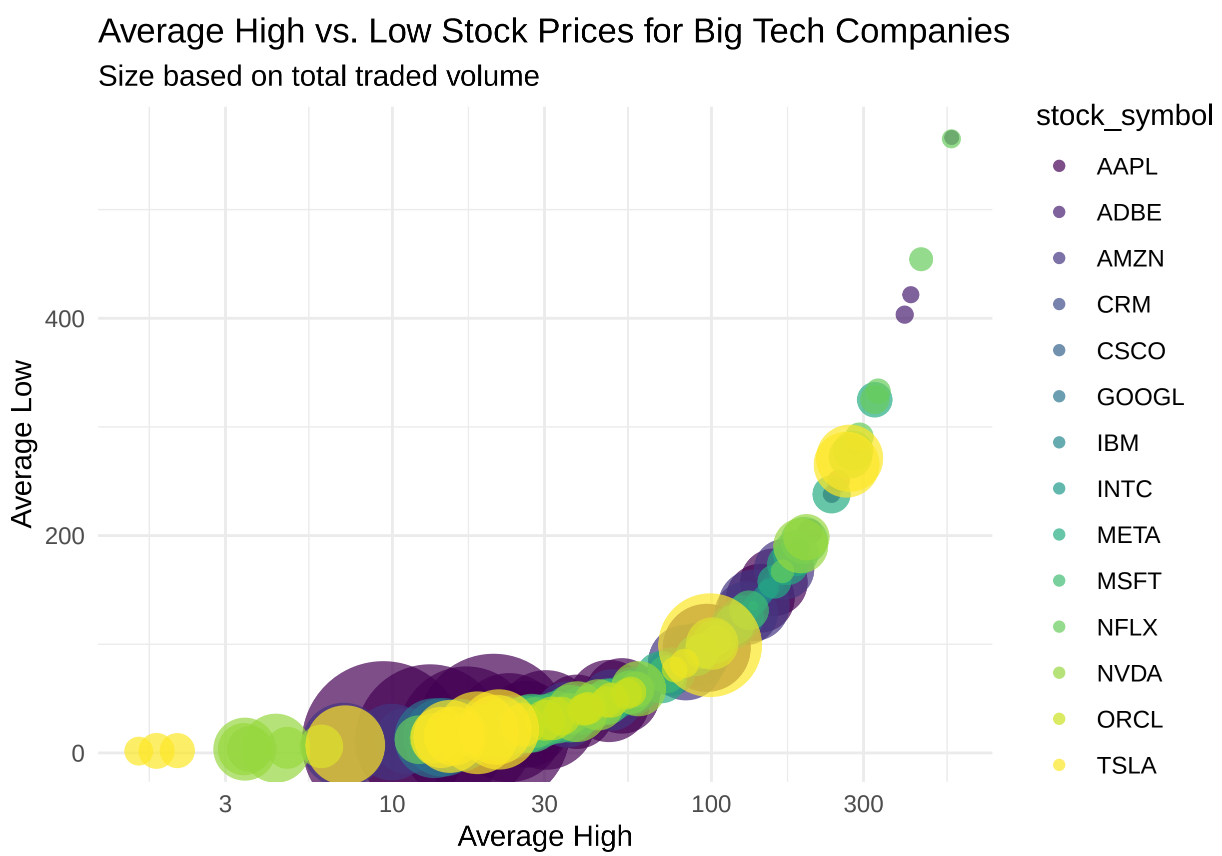
- **Stock Performance**: Investigating AAPL's historical price trends, daily returns, and trading volumes provides a thorough understanding of its performance. Recognizing patterns of growth and volatility is key to forecasting the stock's future movements.
- **Market Behavior**: The behavior of AAPL in the market can reveal insights into broader economic impacts, sector trends, and investor sentiments. It's important to grasp how external factors such as technological advancements, regulatory changes, and economic fluctuations affect AAPL.
- **Sector Influence**: As a leading entity in the tech industry, Apple's stock movements often mirror wider sector trends and can significantly sway market dynamics. This makes AAPL a key indicator of the health of the technology sector.
- **Investment Impact**: Apple's significant market capitalization means it plays a major role in many indices and investment funds, including the S&P 500 and various tech-focused portfolios. Its performance can dramatically influence these broader investment vehicles, making its analysis crucial for both individual and institutional investors.
- **Focal Point for Investors**: Given its size, influence, and performance, AAPL is at the center of investment strategies for a diverse group of market participants. Both individual investors and large institutions monitor Apple's financial health closely to inform their portfolio decisions and risk management strategies.

By exploring these aspects, investors and analysts can make informed decisions, strategically aligning their investment approaches with the opportunities and challenges presented by AAPL's presence in the market. ## Plotting the stock performance of 14 big tech companies

```
annual_stock_summary$year <- as.integer(annual_stock_summary$year)

p <- ggplot(
  annual_stock_summary,
  aes(x = average_high, y=average_high, size = total_vol, colour = stock_symbol)
  ) +
  geom_point(show.legend = c(size = FALSE, color = TRUE), alpha = 0.7) +
  scale_color_viridis_d(option = "D") +
  scale_size(range = c(1, 28)) +
  scale_x_log10() +
```

```r
  labs(x = "Average High",
    y = "Average Low",
    title = "Average High vs. Low Stock Prices for Big Tech Companies",
    subtitle = "Size based on total traded volume",
    legend = "Stock Symbol") +
  theme_minimal() +
  theme(legend.position = "right")
p
```




**Key findings** From the animated data visualization, we can see that Apple possessed the biggest volume of traded stock in the market, this motivates us to look deeper into this company to anaylyze its success in the stock market. Another insight is that the volumes of stock traded decreased from 2019, which makes us think that this has something related to the COVID-19 and is the inspiration behind our second questions.

### Choose Stock to Analyze Based on Volume Traded

The reason for us to choose

```r
# Summing up the volume for each stock
volume_sum <- data %>%
  group_by(stock_symbol) %>%
  summarise(TotalVolume = sum(volume)) %>%
  arrange(desc(TotalVolume))
```

```r
# Display the summed volumes
print(volume_sum)
```

```
## # A tibble: 14 x 2
##    stock_symbol  TotalVolume
##    <chr>               <dbl>
##  1 AAPL         838440829600
##  2 TSLA         294389834000
##  3 AMZN         288960091100
##  4 GOOGL        196869939896
##  5 NVDA         166186840100
##  6 MSFT         124351871200
##  7 INTC         117990516000
##  8 CSCO         106950448600
##  9 META          83806858300
## 10 NFLX          60234987900
## 11 ORCL          58938705200
## 12 CRM           22605794200
## 13 IBM           16474538383
## 14 ADBE          12476695200
```

```r
# Get the top 3 stock symbols
top_3_stocks <- head(volume_sum$stock_symbol, 3)

# Print the top 3 stock symbols
print(top_3_stocks)
```

```
## [1] "AAPL" "TSLA" "AMZN"
```

## Approach

### Scatter plot with trasition through time

- **Purpose**: This plot aims to show the historical price trends of 14 big tech companies stock over the specified period.
- **Design**: The plot features a scatter graph where the x-axis represents average low price of the year, and the y-axis represents the average high price of the year, the size of the scatter reflects the volume of stock traded.
- **Utility**: By examining this plot, we can understand the behavior of companies through time.

### Time Series Plot of Adjusted Close Prices

- **Purpose**: This plot aims to show the historical price trends of AAPL stock over the specified period. It provides a visual representation of the stock's growth and any significant price fluctuations, which are essential for assessing its stability and performance over time. We also incorporate candle stick plot that is more suitable for stock data
- **Design**: The plot features a line graph where the x-axis represents time (date), and the y-axis represents the adjusted close prices of AAPL stock. This straightforward visualization helps in observing the overall trend—whether it is an upward, downward, or cyclic pattern.
- **Utility**: By examining this plot, investors can gauge the stock's long-term growth trajectory, identifying periods of strong performance as well as downturns. It's useful for evaluating the resilience of AAPL as a long-term investment against market volatility.

### Histogram of Daily Returns with Color Mapping

- **Purpose**: This plot analyzes the frequency and distribution of AAPL's daily returns. The goal is to illustrate the stock's daily volatility and risk profile by showing how often and how extensively its price changes within a single trading day.
- **Design**: The histogram segments daily returns into bins, with the height of each bin indicating the frequency of those return values. Color mapping enhances this visualization: gains (positive returns) might be colored in green and losses (negative returns) in red, providing an immediate visual distinction between profitable and loss-making days.
- **Utility**: Investors and analysts use this plot to understand the risk associated with short-term investments in AAPL. A broad distribution, especially with substantial tails, indicates higher volatility. This insight helps in making informed decisions about buying, holding, or selling the stock based on personal or institutional risk tolerance.

### Box Plot of Price Changes by Trading Volume Categories

- **Purpose**: The aim of this plot is to explore the relationship between trading volumes and price changes, determining if larger volumes correlate with significant price movements.
- **Design**: Trading days are categorized into quartiles based on trading volume (e.g., Low, Medium, High, Very High). A box plot for each category displays the spread and median of price changes on those days. This can include whiskers that extend to show the range, boxes indicating the interquartile range, and a line within the box showing the median value.
- **Utility**: This plot allows for an examination of whether high trading volumes are associated with wider price swings, which might suggest that AAPL's stock price is particularly sensitive to large trades. This correlation can indicate how news or market events impact the stock, providing a deeper understanding of its market behavior and potentially guiding trading strategies based on volume analysis.

## AAPL Stock Analysis

```r
# Load AAPL stock data
aapl_data <- read.csv("data/customed/AAPL_stock_data.csv")

# Convert date column to Date type
aapl_data$date <- as.Date(aapl_data$date)

# Display the structure and summary of the dataset
str(aapl_data)
```

```
## 'data.frame':    3271 obs. of  7 variables:
##  $ date     : Date, format: "2010-01-04" "2010-01-05" ...
##  $ open     : num  7.62 7.66 7.66 7.56 7.51 ...
##  $ high     : num  7.66 7.7 7.69 7.57 7.57 ...
##  $ low      : num  7.58 7.62 7.53 7.47 7.47 ...
##  $ close    : num  7.64 7.66 7.53 7.52 7.57 ...
##  $ adj_close: num  6.52 6.53 6.42 6.41 6.45 ...
##  $ volume   : int  493729600 601904800 552160000 477131200 447610800 462229600 594459600 605892000 4
```

```r
summary(aapl_data)
```

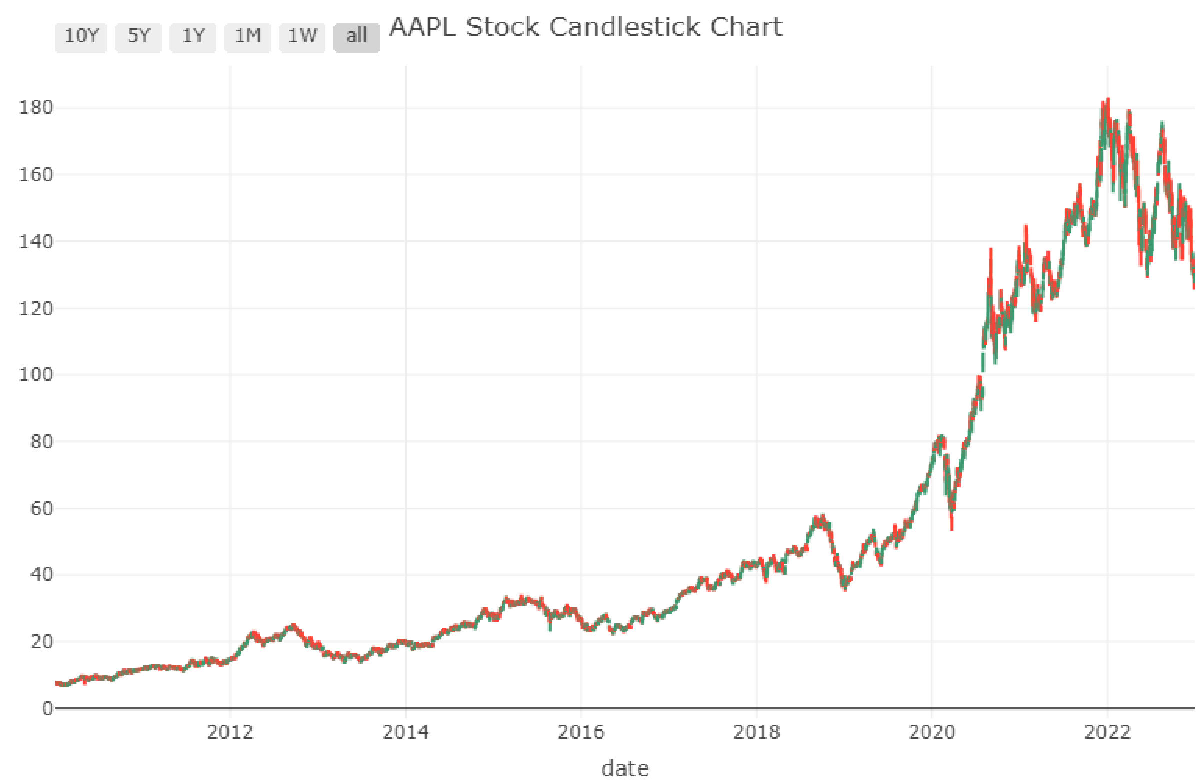
```
##       date                 open              high              low
```

```
##  Min.   :2010-01-04   Min.   :  6.87   Min.   :  7.00   Min.   :  6.795
##  1st Qu.:2013-04-04   1st Qu.: 18.97   1st Qu.: 19.12   1st Qu.: 18.780
##  Median :2016-07-01   Median : 29.75   Median : 29.98   Median : 29.555
##  Mean   :2016-07-01   Mean   : 51.27   Mean   : 51.85   Mean   : 50.709
##  3rd Qu.:2019-10-01   3rd Qu.: 56.90   3rd Qu.: 57.26   3rd Qu.: 56.435
##  Max.   :2022-12-29   Max.   :182.63   Max.   :182.94   Max.   :179.120
##      close            adj_close           volume
##  Min.   :  6.859   Min.   :  5.847   Min.   :3.520e+07
##  1st Qu.: 18.966   1st Qu.: 16.626   1st Qu.:1.024e+08
##  Median : 29.812   Median : 27.385   Median :1.667e+08
##  Mean   : 51.297   Mean   : 49.445   Mean   :2.563e+08
##  3rd Qu.: 56.761   3rd Qu.: 54.876   3rd Qu.:3.458e+08
##  Max.   :182.010   Max.   :180.960   Max.   :1.881e+09
```

```r
# Check for null values
sum(is.na(aapl_data))
```

```
## [1] 0
```

#### Candle stick performance of Apple




### AAPL Adjusted Close Price Over Time

```r
# Plotting the adjusted close price over time
ggplot(aapl_data, aes(x = date, y = adj_close)) +
  geom_line() +
  theme_minimal() +
  labs(title = "AAPL Stock Adjusted Close Price Over Time",
       x = "Date", y = "Adjusted Close Price")
```


The Adjusted Close Price graph of AAPL underscores a compelling growth narrative, supported by a comprehensive dataset spanning 3,271 trading days. This extensive period allows for a detailed analysis of the stock's performance, showcasing a notable upward trend indicative of strong market confidence and Apple's solid fundamentals.

#### Statistical Insights

- **Total Analyzed Period**: The data spans a significant timeframe, offering a robust foundation for analyzing trends.
- **Growth Indicators**: The upward movement reflects Apple's resilience and growth, with the stock demonstrating a remarkable ability to recover from market fluctuations and maintain a positive trajectory.

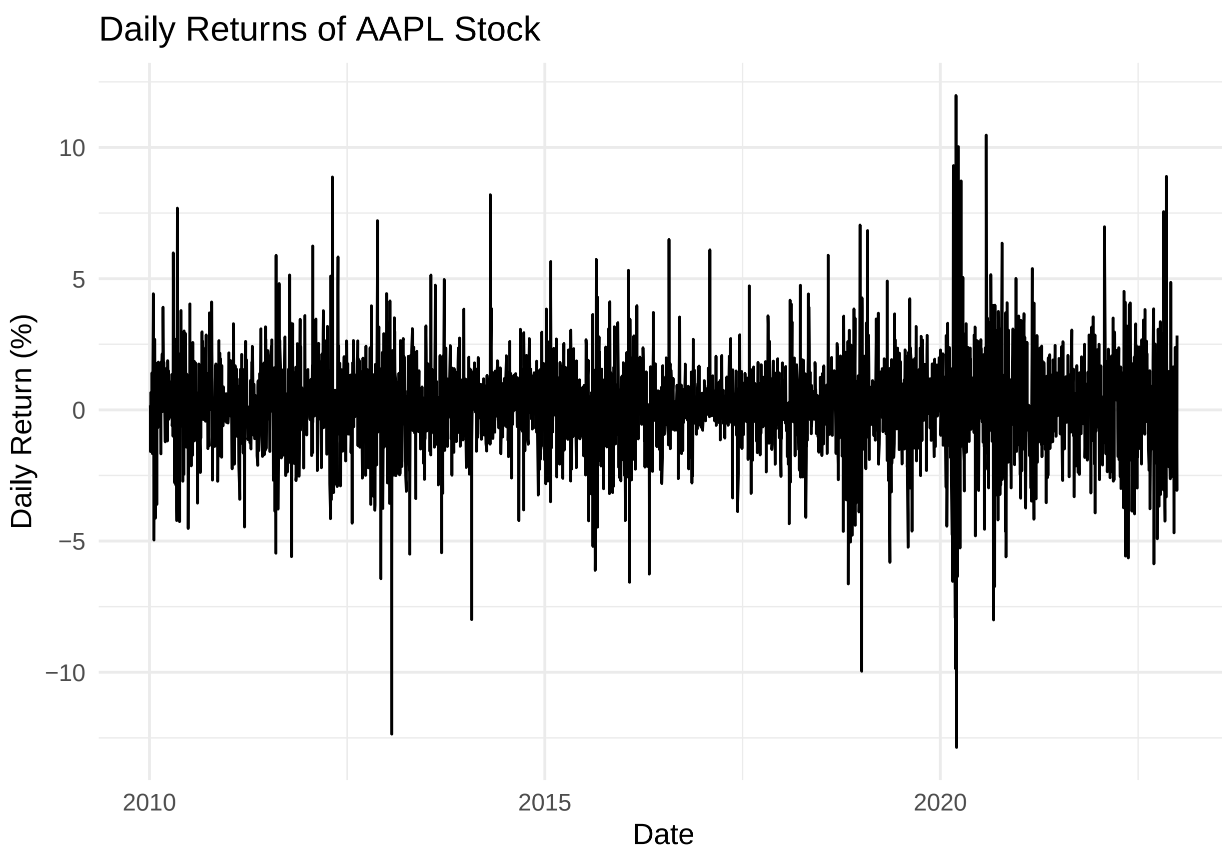
#### AAPL Daily Returns

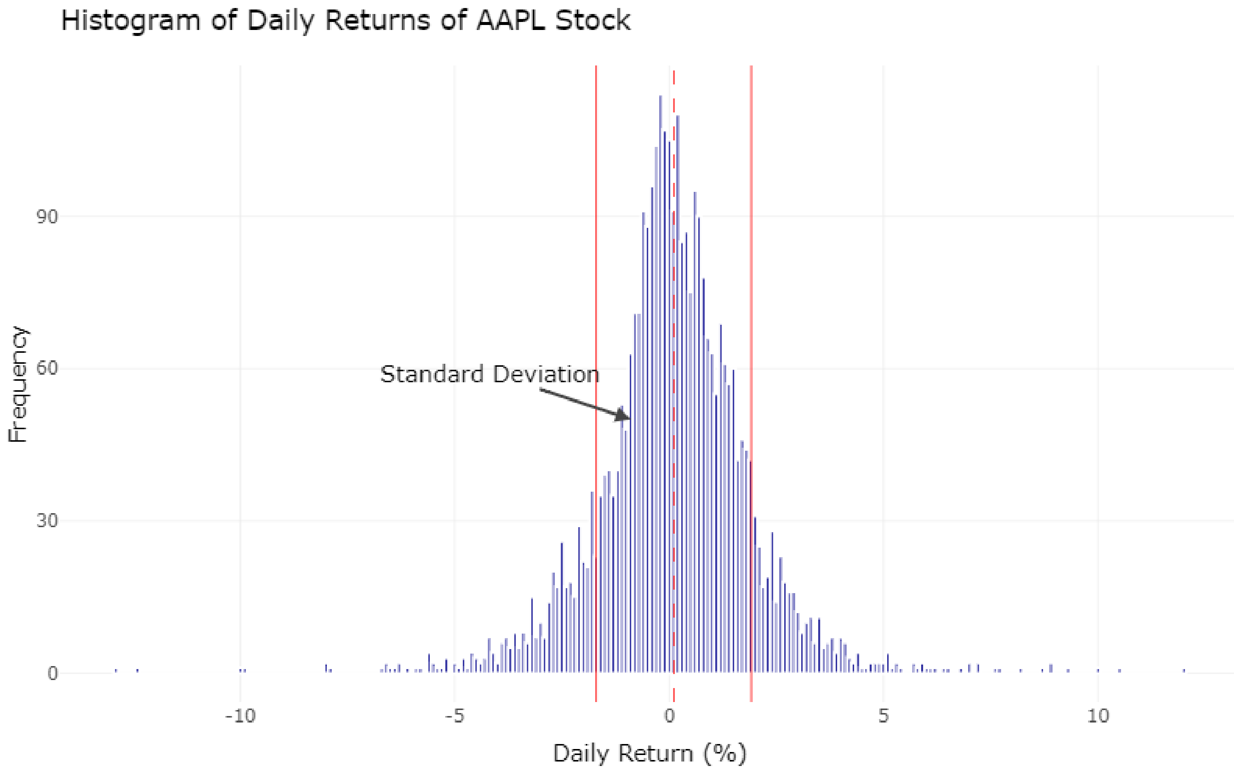
```r
# Calculating daily returns
aapl_data <- aapl_data %>%
  arrange(date) %>%
  mutate(daily_return = (adj_close / lag(adj_close) - 1) * 100)

# Plotting daily returns
ggplot(aapl_data, aes(x = date, y = daily_return)) +
  geom_line() +
  theme_minimal() +
```

```r
  labs(title = "Daily Returns of AAPL Stock",
       x = "Date", y = "Daily Return (%)")
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```






The Daily Returns graph provides a clear depiction of AAPL's volatility, characterized by fluctuations around a central tendency with occasional spikes. This volatility is a testament to the dynamic nature of the stock market and Apple's sensitivity to market sentiments and news.

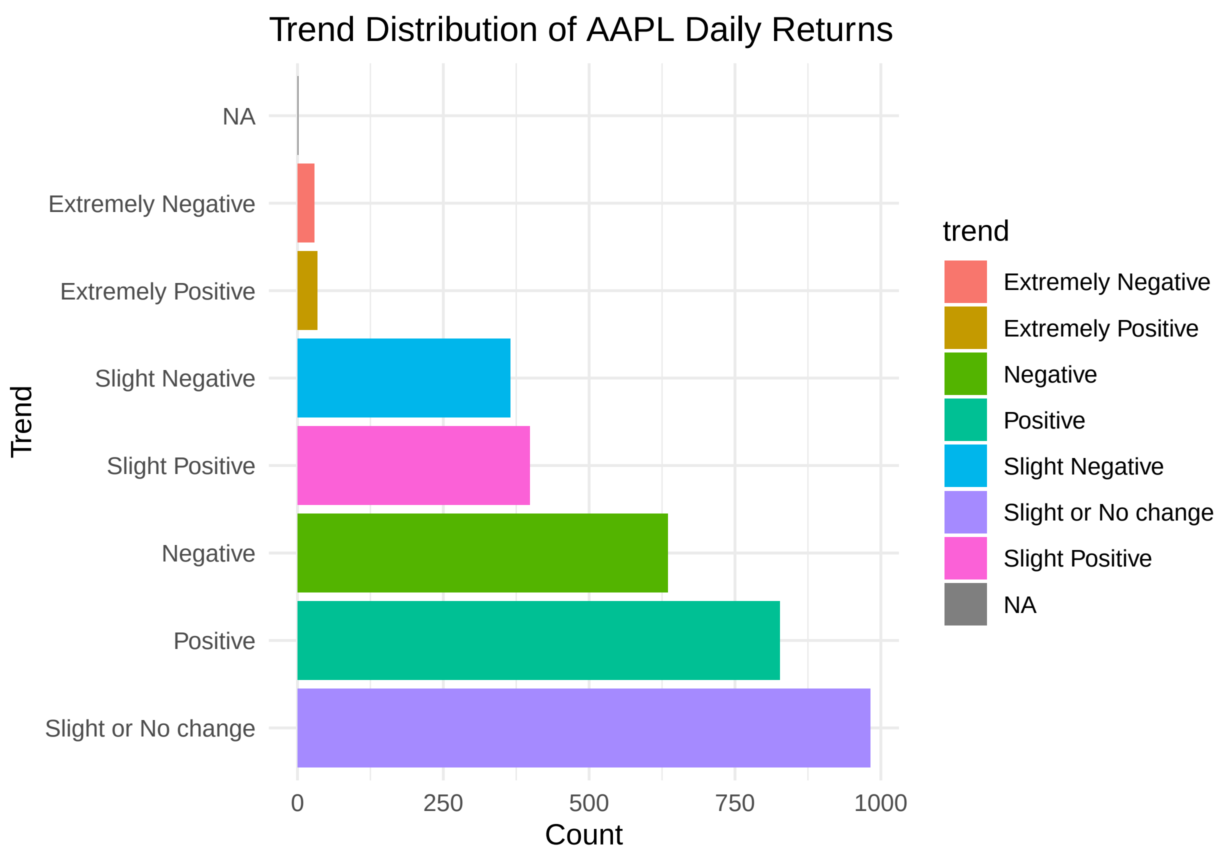
#### Statistical Highlights

- **Positive vs. Negative Days**: With 1,721 days (52.61%) recording positive returns and 1,544 days (47.20%) showing negative returns, the data highlights a slightly higher tendency towards positive movements, albeit with substantial variability.
- **Volatility Measurement**: The standard deviation of daily returns stands at approximately 1.81%, emphasizing the stock's short-term volatility. The range of daily returns, from a maximum of 11.98% to a minimum of -12.86%, further underlines the potential for significant price swings within a single trading day.

#### Trend Distribution of AAPL Daily Returns

```r
# Defining trend categories based on daily returns
aapl_data$trend <- with(aapl_data, case_when(
  daily_return <= 0.5 & daily_return >= -0.5 ~ "Slight or No change",
  daily_return > 0.5 & daily_return <= 1 ~ "Slight Positive",
  daily_return < -0.5 & daily_return >= -1 ~ "Slight Negative",
  daily_return > 1 & daily_return <= 5 ~ "Positive",
  daily_return < -1 & daily_return >= -5 ~ "Negative",
  daily_return > 5 ~ "Extremely Positive",
  daily_return < -5 ~ "Extremely Negative"
))
```

```r
# Plotting trend distribution
aapl_data %>%
  group_by(trend) %>%
  summarise(count = n()) %>%
  ggplot(aes(x = reorder(trend, -count), y = count, fill = trend)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Trend Distribution of AAPL Daily Returns",
       x = "Trend", y = "Count") +
  coord_flip()
```




The trend distribution offers a nuanced view of AAPL's daily return patterns, categorizing days into various trends based on the magnitude of price changes. This segmentation reveals the predominance of days with 'Slight or No change', highlighting an underlying stability amidst volatility. Moreover, comparing between positive and negative of each type, we can clearly see that positive out perform negative in each segmentation, which will ultimately translate into profit in long term.

### Statistical Insights

- **Trend Distribution**: Approximately 30.03% of the days witnessed 'Slight or No change', indicating stability on a significant portion of trading days. The presence of 'Positive' (26.33%) and 'Negative' (20.31%) days illustrates the stock's capacity for notable price movements, providing opportunities for strategic trading.

## Conclusion

The time series plot shows a general upward trend in AAPL's stock price, supporting its consideration as a viable long-term investment. However, this plot might also reveal periods of volatility during market downturns or economic uncertainty, highlighting potential risks that need to be managed.

The histogram of daily returns will likely demonstrate AAPL's short-term trading risks. A wide spread in the distribution of returns, especially if skewing towards negative outcomes, could indicate higher volatility, which is often seen in tech stocks influenced by market sentiments and external economic factors.

### Key Takeaways for Investors

- **Long-term Growth**: AAPL shows promising growth prospects, supported by its historical price increases.
- **Short-term Volatility**: Investors must navigate short-term volatility, evident in daily returns and volume spikes, which reflect the stock's sensitivity to market sentiments and news.
- **Informed Decision-Making**: A strategic approach, leveraging insights from trading volumes and price trends, can help in capitalizing on opportunities and managing risks.

Together, these plots will provide a dual perspective on AAPL's investment profile. If the long-term trend is strong but accompanied by significant short-term fluctuations, the stock may be more suited to investors who can tolerate volatility or those employing strategies that leverage these movements. Conversely, a more stable long-term trend accompanied by moderate daily fluctuations would appeal to conservative investors seeking growth with manageable risks.

By analyzing these trends, investors and analysts can better understand the conditions under which AAPL thrives and the potential challenges it faces, facilitating more informed decision-making regarding portfolio inclusion and investment strategy.

# Conclusion

2024-04-09

## Conclusion

In conclusion, AAPL looks like it's growing steadily over the long haul, which is a good sign. But there's a lot of up-and-down action in the short term, especially because of how people feel about the market and the news. To make good choices, it's important to have a plan that takes into account how much trading is happening and which way prices are going. Considering both the big picture of growth and the day-to-day swings gives us a better idea of what investing in AAPL might be like. For folks okay with taking on some risk for the chance at bigger rewards, the long-term growth might be attractive, even with the short-term jumps. But if you're looking for something more steady, a stock with a smoother long-term trend and fewer big changes each day might be a better fit. By thinking about these things, investors can make smart decisions that match their comfort with risk and what they want to achieve with their investments, making sure they can handle the ups and downs of AAPL's stock.

The Covid-19 pandemic had a significant impact on tech companies' stock prices, but its effects differed among these industry giants. Companies that belong to entertainment, online shopping, or other online services tended to fare better, thanks to increased demand for their products during lockdowns. While tech companies initially suffered more severe losses compared to industries like groceries or medical sectors, they also experienced a quicker recovery. As a result, their stock prices have rebounded stronger than those in other sectors post-pandemic.

# Project1

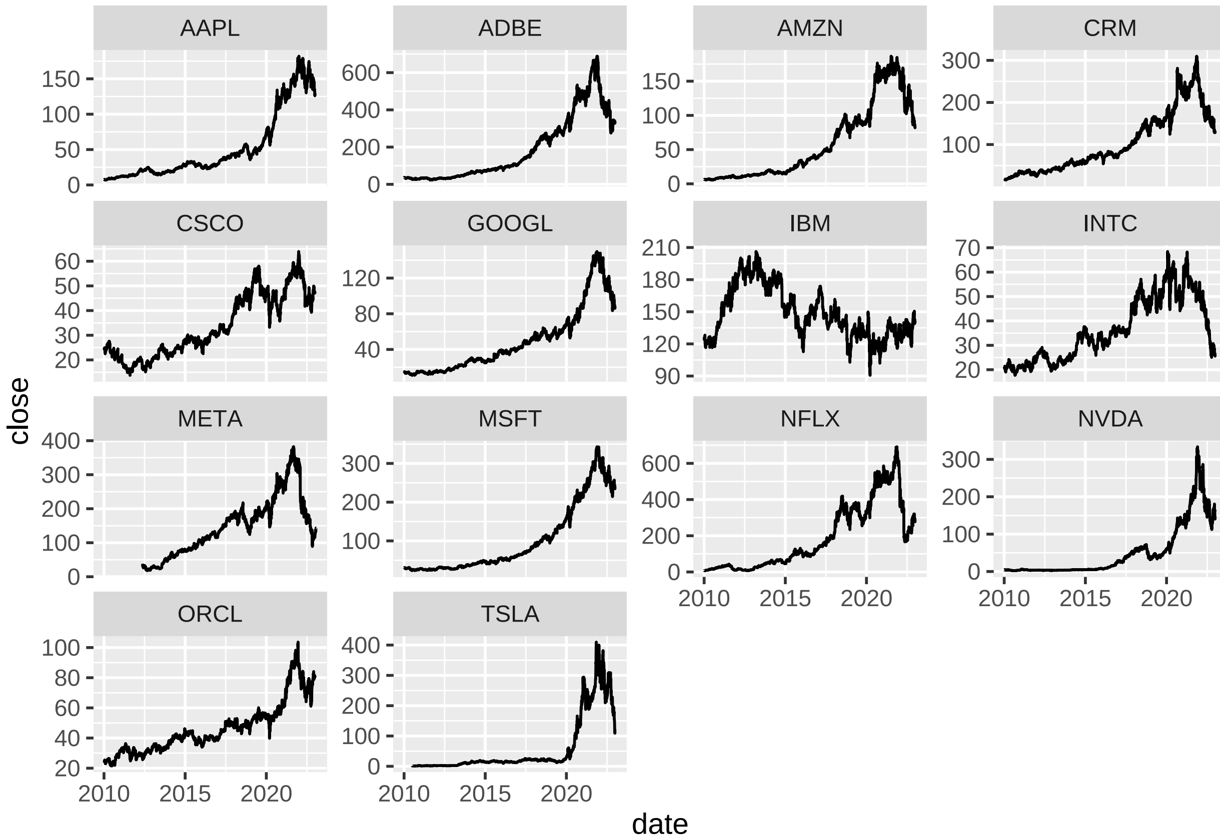
2024-04-02

## 2. How did the COVID-19 Pandemic affect the stock prices of Big Tech Companies?

### 2.1. Introduction:

- We would like to answer this question because we observed a significant decrease in stock prices of major tech groups from late 2019 to mid-2020 which **overlapped with the dramatic expansion of COVID-19**. We'll check how stock prices changed during this time and see which companies were hit the hardest. By detecting less affected companies, we can focus on exploring their potential strategies dealt with the pandemic which saves time from general research. Therefore, we might be able to learn how to handle similar situations better in the future. ### Data preparation:

```
ggplot(df,aes(y = close , x = date, group = 1)) +
  geom_line () +
  scale_x_date(labels = date_format("%Y"))+
  facet_wrap(~ stock_symbol, scales = "free_y")
```




- To answer that question, from the original dataset, we need the **close price**, which can be considered the primary price of the stock, from late 2019 to mid-2020.
- Moreover, to visualize the correlation between covid 19 spread and stock price, we use an external dataset, which contains **daily global statistics about covid 19 cases** from 22 Jan 2020 to 27 Jul 2020, published on Kaggle.
- Finally, to compare technology companies with other sectors, for example, groceries or medical products, we use 2 more datasets about the **historical stock price of Walmart and Pfizer** in the corresponding period.

## 2.2 Approaches

**Plot choice explanation:**

- **Firstly**, because COVID-19 affecting tech companies' stock prices is just our assumption, we have to qualify whether COVID-19 actually had that influence by plotting out stock prices and the global increase percentage of COVID-19 cases in the same plot. We use **bar plots**, which is good for describing the distribution of low-dimension data, to show the changes in stock price over time. Moreover, by visualizing with bar plot, we can easily see the change in stock price, compare it to past or future prices, and classify them at the same time ( We use different colors for *Increase, Decrease or No change* samples compared to the base price).
- The growth speed of COVID-19 cases will be visualized with a **line plot** because we want to reduce the plot's density ( lines save more space than bars) and emphasize the trend, which is the advantage of line charts, of COVID-19 spread.

- **Secondly,** after we verify the assumption and visualize some patterns, we want to clearly visualize which company suffers most from COVID-19. In other words, we want to visualize their loss *( loss = (expected price - actual price) * volume) )* and rank them in descending order. We chose a bar plot for this purpose because this type of chart is optimized for rank illustration. We can see what companies did the best policies and their differences with tech giants who performed worse by comparing bars.
- **Finally**, different from the 2 tasks mentioned above, our final sub-question requires more data points to intuitively show the trend in the stock price of each tech company and 2 external representatives from the groceries/ medical fields. In this situation, a line chart would be an optimized consideration to reduce the density caused by a huge number of data points. Moreover, line charts can also show the trend over time better compared to other types of charts, so that we can clearly visualize changes in stock prices every single day. Besides that, because we want to see the date when sectors (tech, medical, groceries) swap their position in stock price rankings as well, we can easily point out those special days by intersections among lines.

**Data pre-processing:**

- Because we only crawled the data set of the most dramatic period of COVID-19 ( 1/22/2020 - 7/27/2020), we need to create subset of stock prices in that time.
- As I mentioned, bar plots are great to illustrate the distribution, however, it increase the density of the plot, makes it harder to observe differences in bars in the case where **there are too many bars**. Therefore, with the range of 6 months (nearly 200 days), I recalculate and **only leverage the average stock price for each week** to enhance the clarity of the plot while maintain the general distribution.
- Besides that, because **number of cases exponentially increased,** initial data points would be meaningless if we directly plot the raw data. Therefore, we try to consider the growth speed of the number of global COVID-19 cases, and represent this metric with a new variable: **percentage** *( percentage = number of new case / total of case * 100 ).*
- If there had been no COVID-19, the stock price of companies would have increased or at least been unchanged, therefore, the decrease in stock price created a loss value for the companies. The more stock transactions, the higher loss they had to suffer.
- Therefore, we come up with a metrics to evaluate their losses: ***loss = (expected price - actual price) * volume)***
- Because each company have different stock price, and we want to focus on the growth of them in the pandemic period, it would be **unfair if we directly plot the raw price.** Therefore, **percentage compared to the initial price would be our choice.**

**Other techniques:**

- Colors are used to show the comparison of price in that week with the based price ( in the week before COVID-19), so that we can see **how hard or how long COVID-19 badly affected the stock prices of each companies**.
- We added some **annotations for maxima points** to see whether the most dramatic COVID-19 expansion match with the lowest stock price in the time series.

**Theme definition:**

```r
theme_covid <- function(){
      theme_minimal(base_family = "Montserrat") +
      theme(

      plot.title = element_text(hjust = 0.5, face = "bold", color = "#c83538"),
      plot.subtitle = element_text(hjust = 0.5, face = "bold"), # This will center the subtitle
      legend.position = "bottom",
      axis.text.x = element_text(angle = 45,color = "#35426e", face = "bold", size = 12 ),
      axis.text.y = element_text(color = "#35426e", face = "bold", size = 12),
      panel.grid.major.x = element_blank(),  # Remove major x grid lines
      panel.grid.minor.x = element_blank(),
      panel.grid.major.y = element_line(color = "grey", size = 0.2),
      panel.grid.minor.y = element_blank())
}




theme_covid2 <- function(){
      theme_minimal(base_family = "Montserrat") +
      theme(

      plot.title = element_text(hjust = 0.5, face = "bold", color = "#c83538"),
      plot.subtitle = element_text(hjust = 0.5, face = "bold"), # This will center the subtitle
      axis.text.x = element_text(angle = 45,color = "#35426e", face = "bold", size = 12 ),
      axis.text.y = element_text(color = "#35426e", face = "bold", size = 12),
      panel.grid.major.x = element_blank(),  # Remove major x grid lines
      panel.grid.minor.x = element_blank(),
      panel.grid.major.y = element_line(color = "grey", size = 0.2),
      panel.grid.minor.y = element_blank())
  }
```

## 2.3 Analysis

#### 2.3.1. Verify the assumption "Did COVID-19 actually affect Tech Giants' stock prices?"

```r
create_covid_plot("AAPL")
```

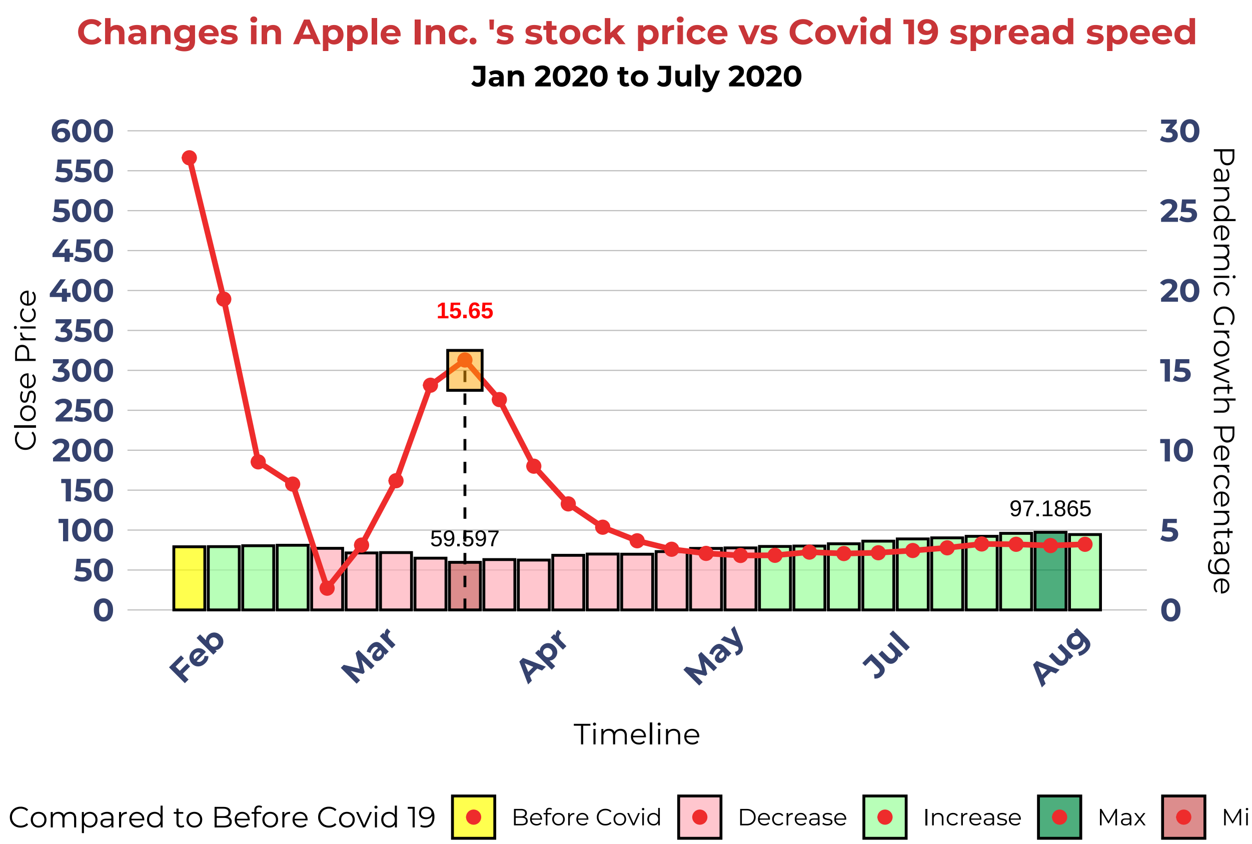
```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.

## Warning: The `trans` argument of `sec_axis()` is deprecated as of ggplot2 3.5.0.
## i Please use the `transform` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
```

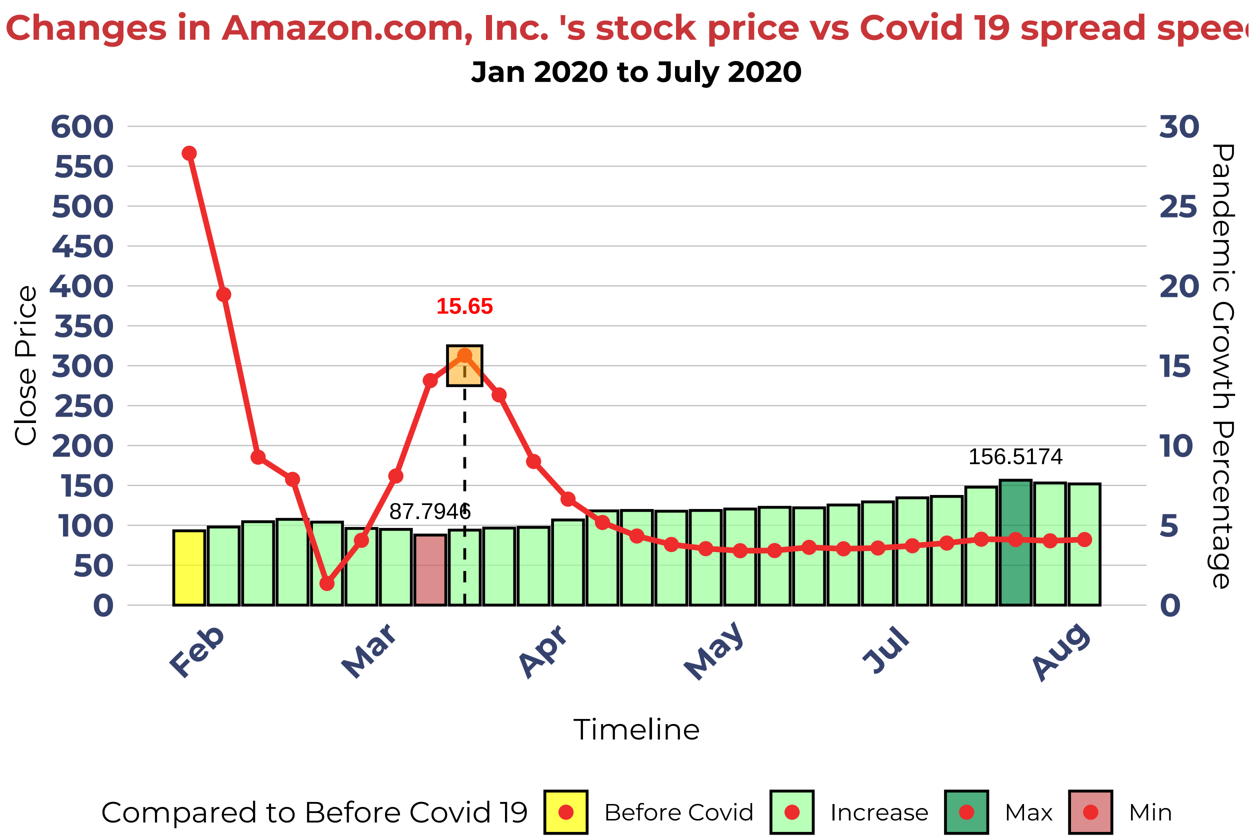
```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## Warning: The 'size' argument of 'element_line()' is deprecated as of ggplot2 3.4.0.
## i Please use the 'linewidth' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



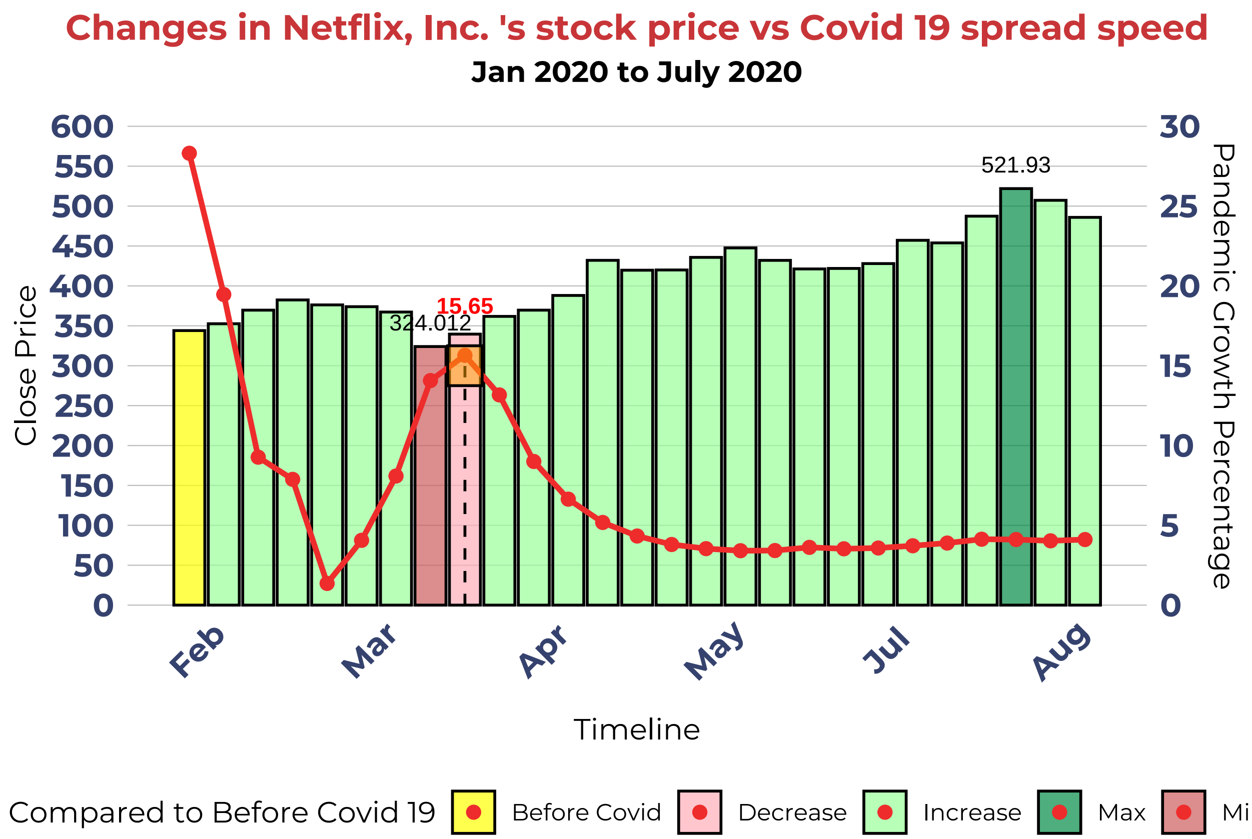

```
create_covid_plot("AMZN")
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```



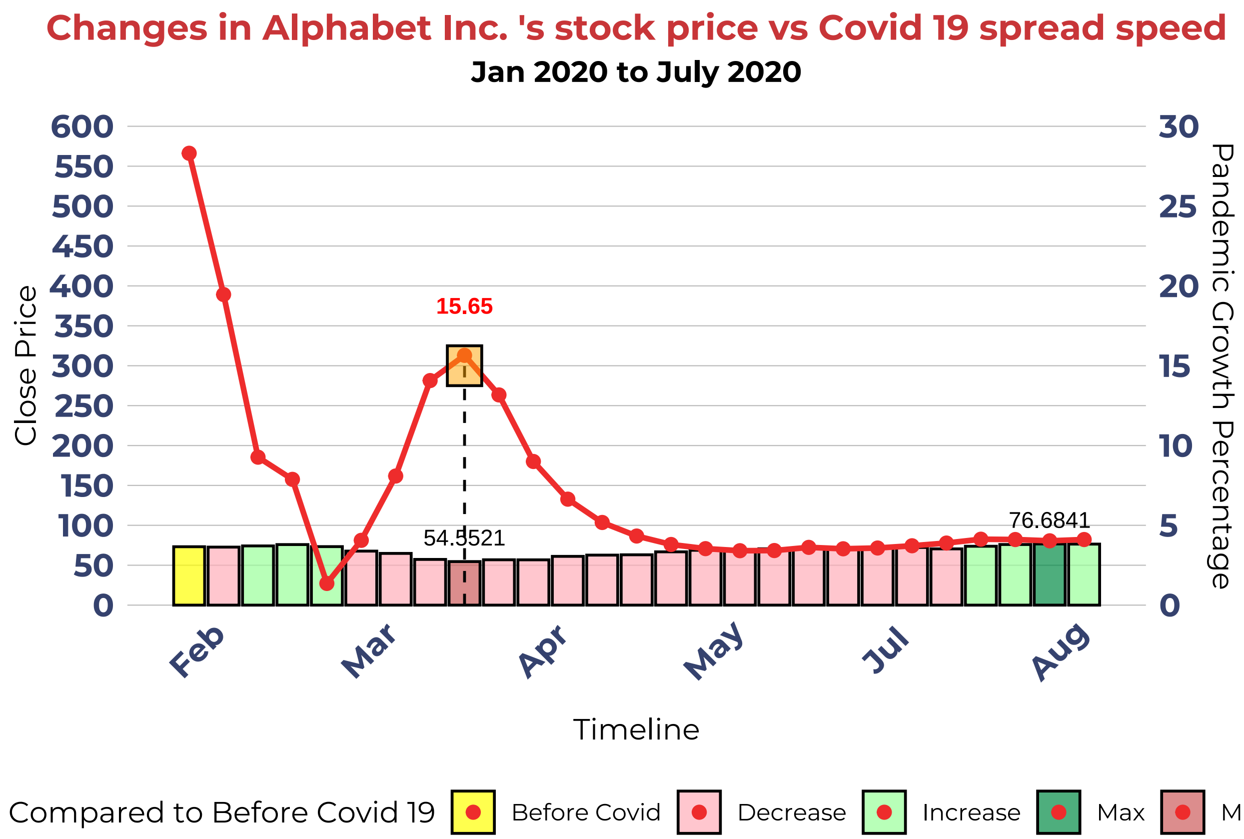

```r
create_covid_plot("NFLX")
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



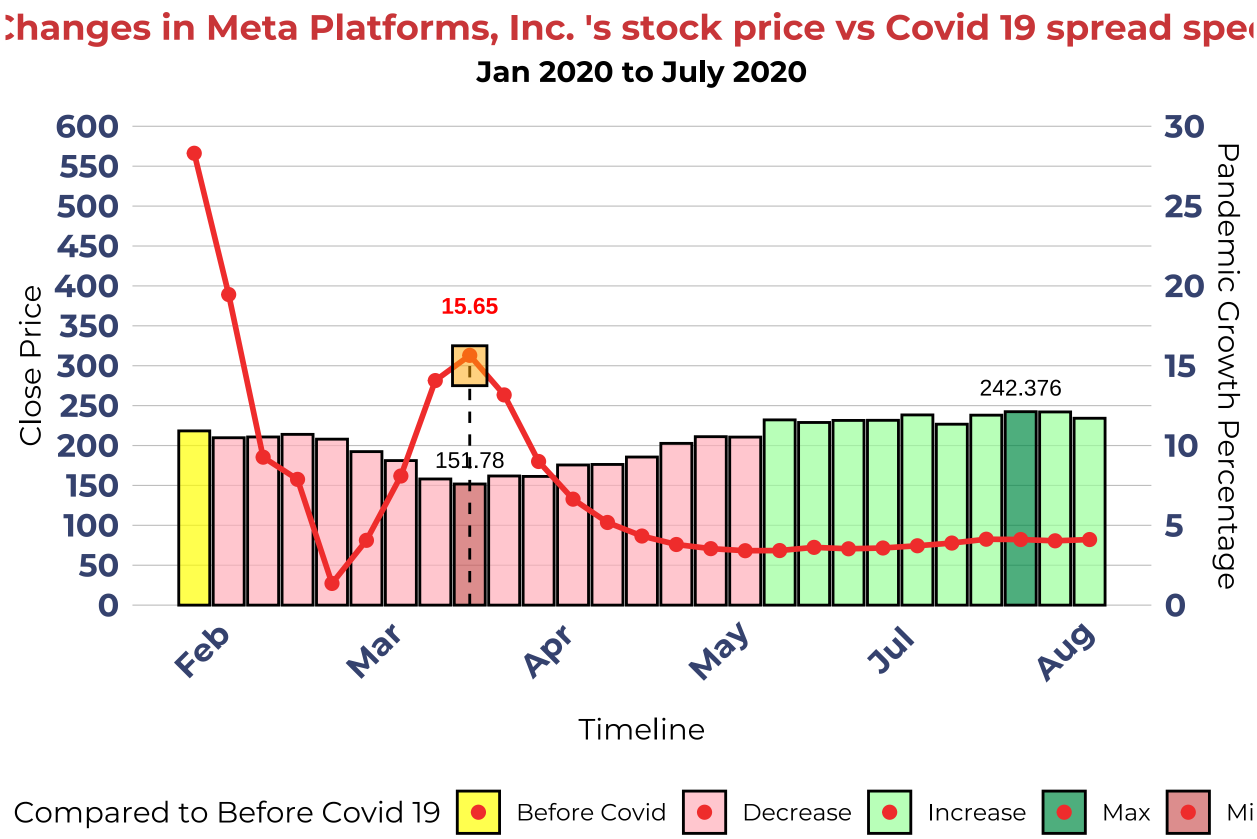

```r
create_covid_plot("GOOGL")
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```



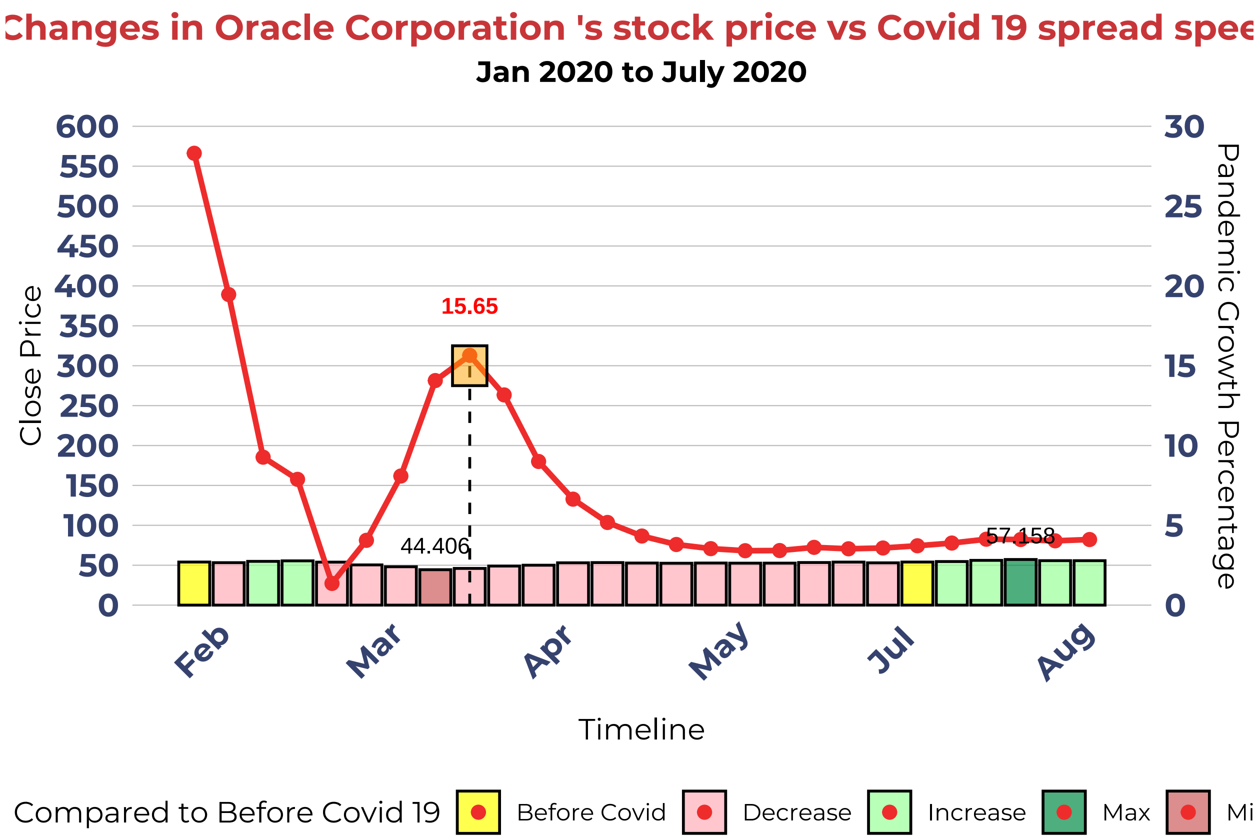

```
create_covid_plot("META")
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```



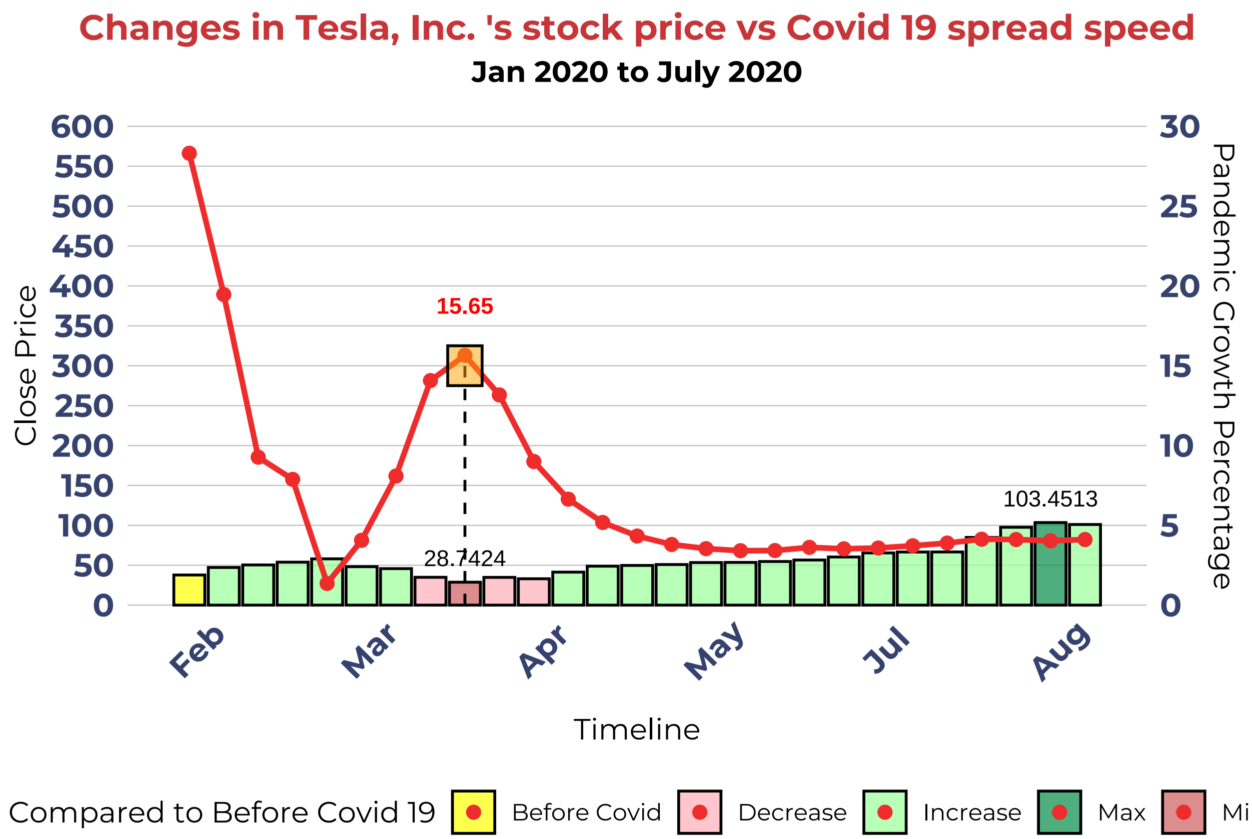

```r
create_covid_plot("ORCL")
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



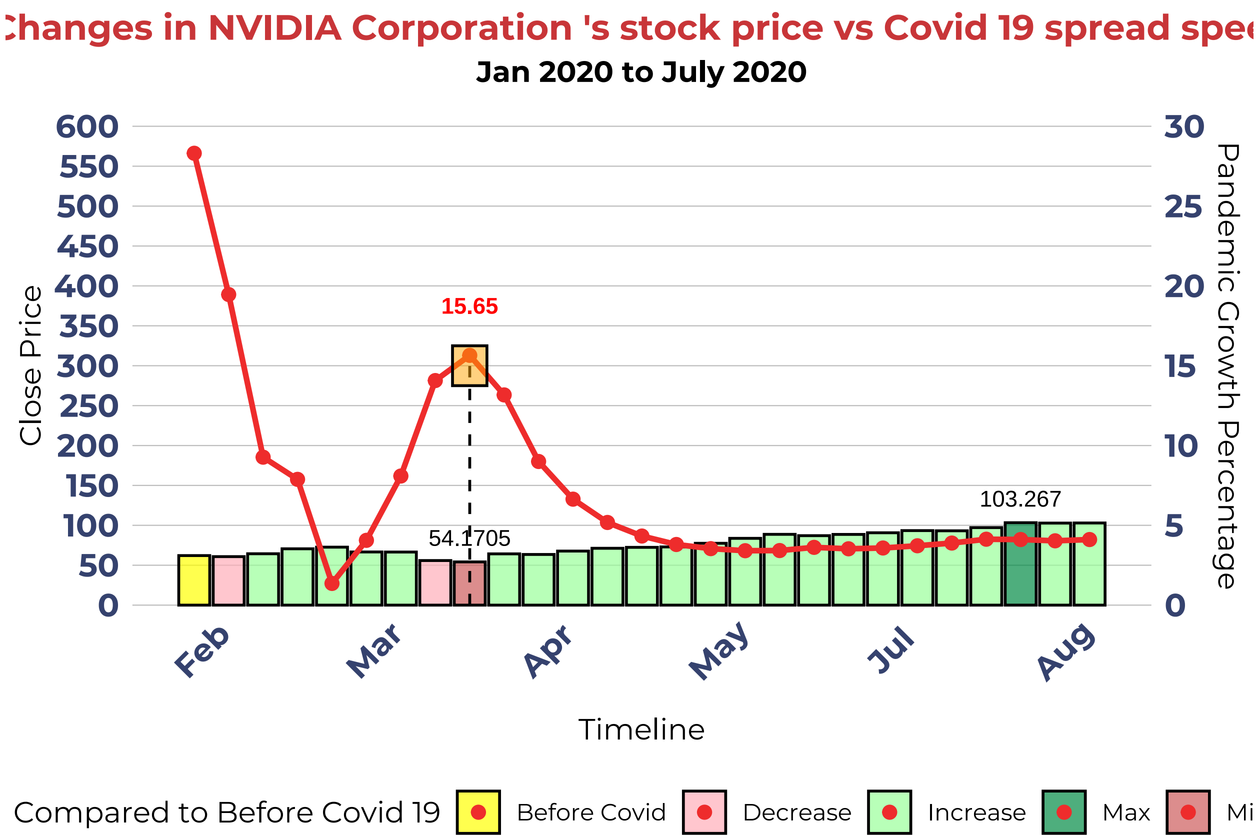

```
create_covid_plot("TSLA")
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



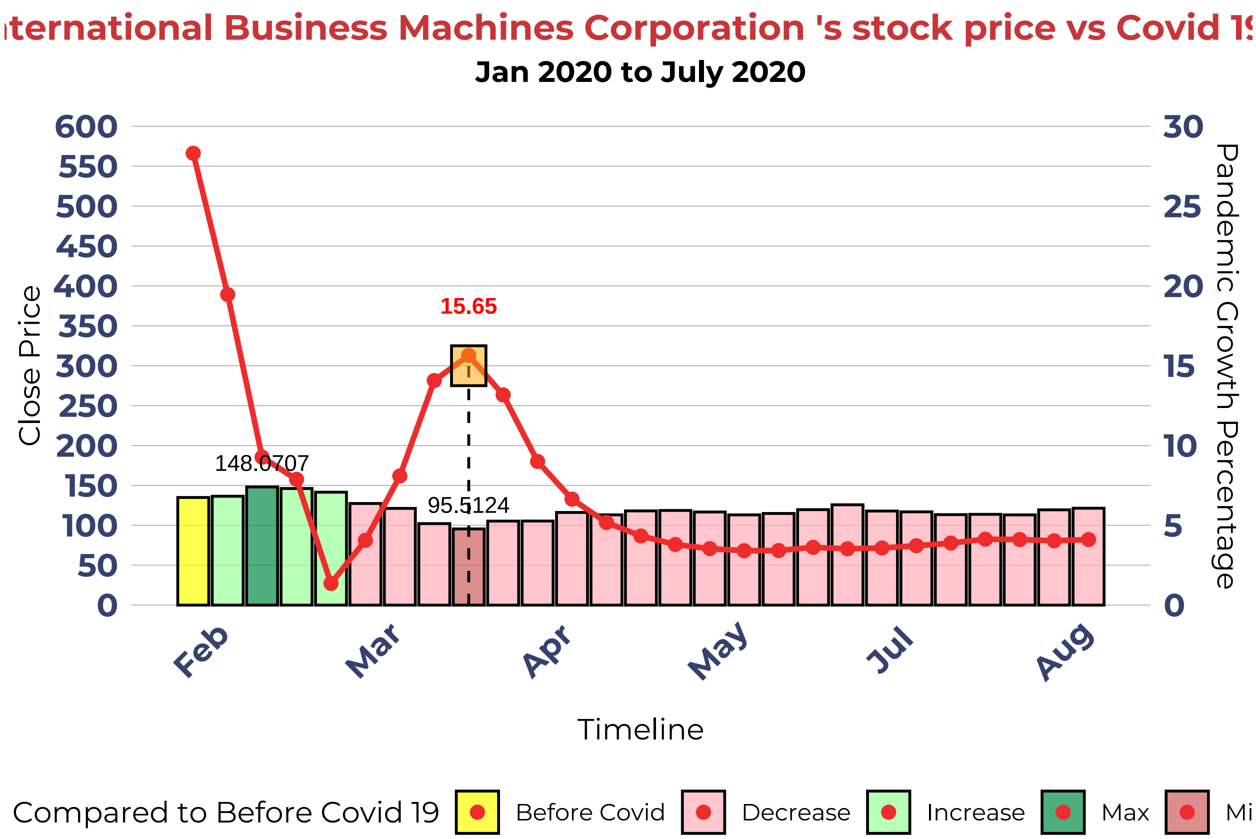

```r
create_covid_plot("NVDA")
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



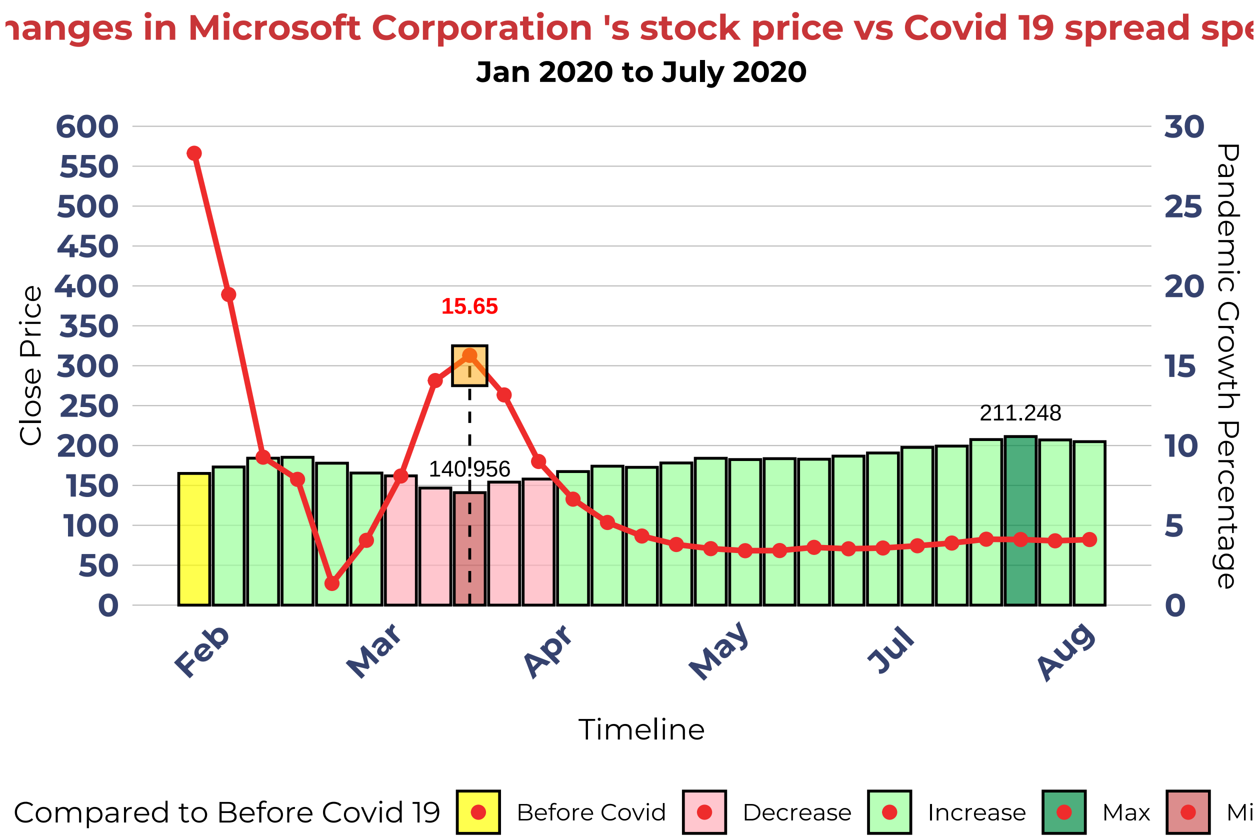

```
create_covid_plot("IBM")
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```




```r
create_covid_plot("MSFT")
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



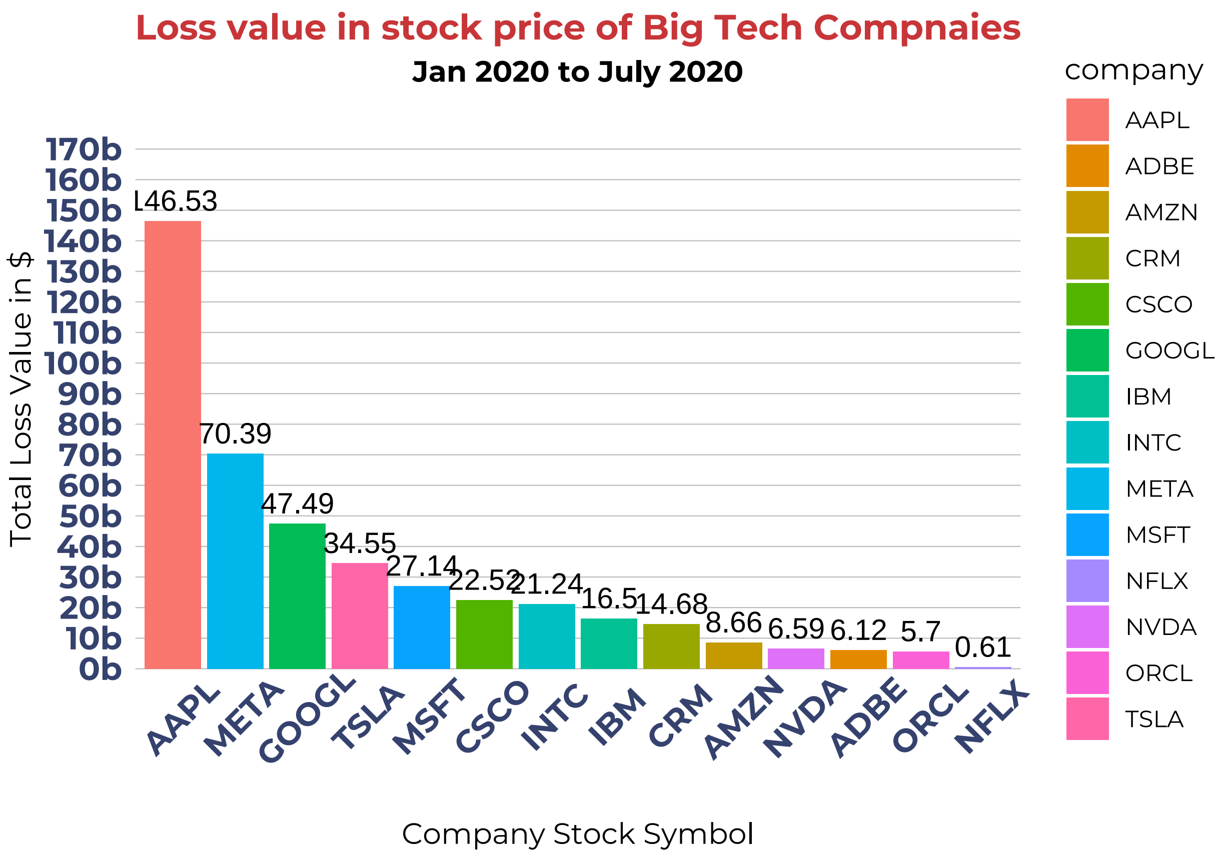

#### 2.3.2. Which company is most affected ( metric: expeceted loss value) ?

```r
print(sorted_df)
```

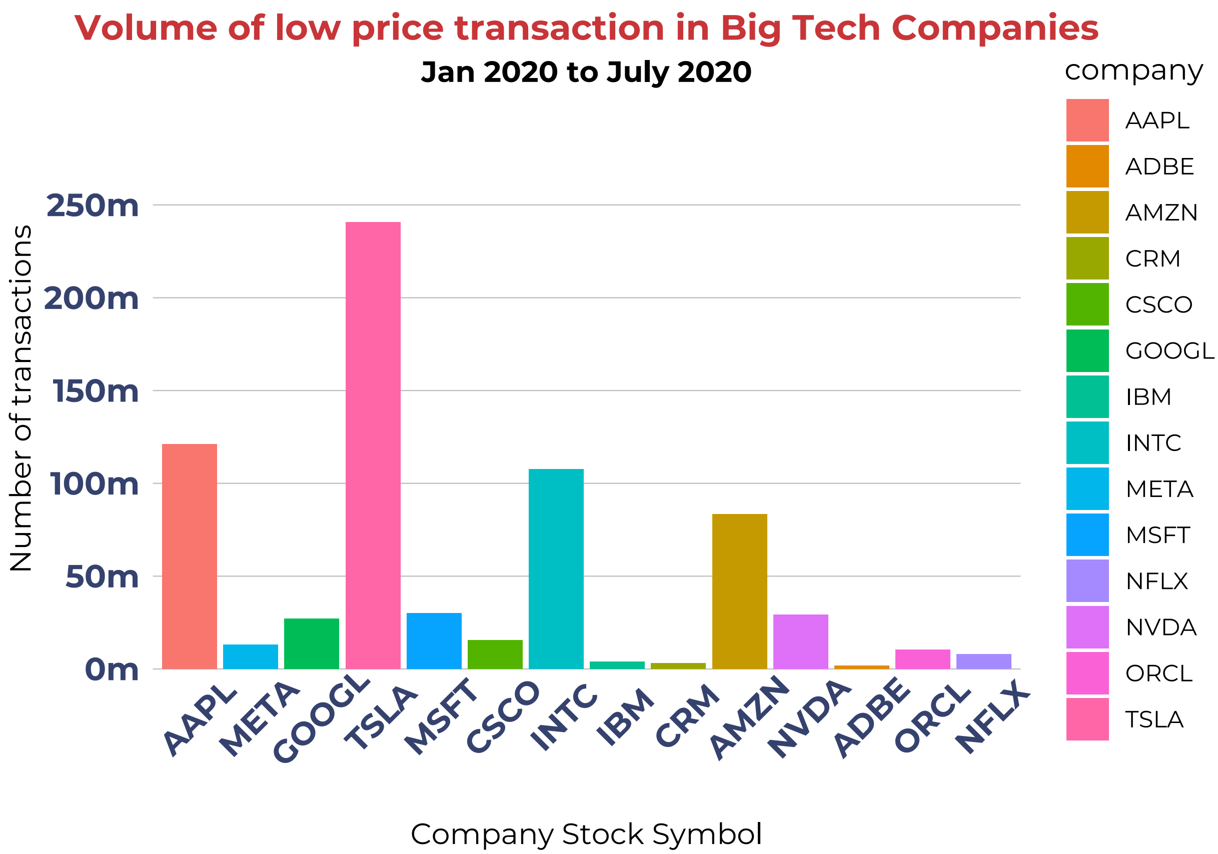
```
##     company    loss_value    volume
## 2      AAPL 146532797810 121214000
## 10     META  70387759593  13163100
## 7     GOOGL  47491417059  27124000
## 15     TSLA  34548093748 240730500
## 11     MSFT  27139145392  30160900
## 6      CSCO  22515816955  15576000
## 9      INTC  21240160026 107526500
## 8       IBM  16499190016   3905136
## 5       CRM  14677418881   3015700
## 4      AMZN   8662239393  83410000
## 13     NVDA   6590426189  29213200
## 3      ADBE   6116644253   1622000
## 14     ORCL   5696346615  10319600
## 12     NFLX    611774097   7863100
## 1      test            0         0
```

```r
for (symbol in loss_company$stock_symbol) {
    loss_company <- calculate_loss(loss_company,symbol)
}
```

```
loss_plot+theme_covid2()
```




```
volume_plot +theme_covid2()
```

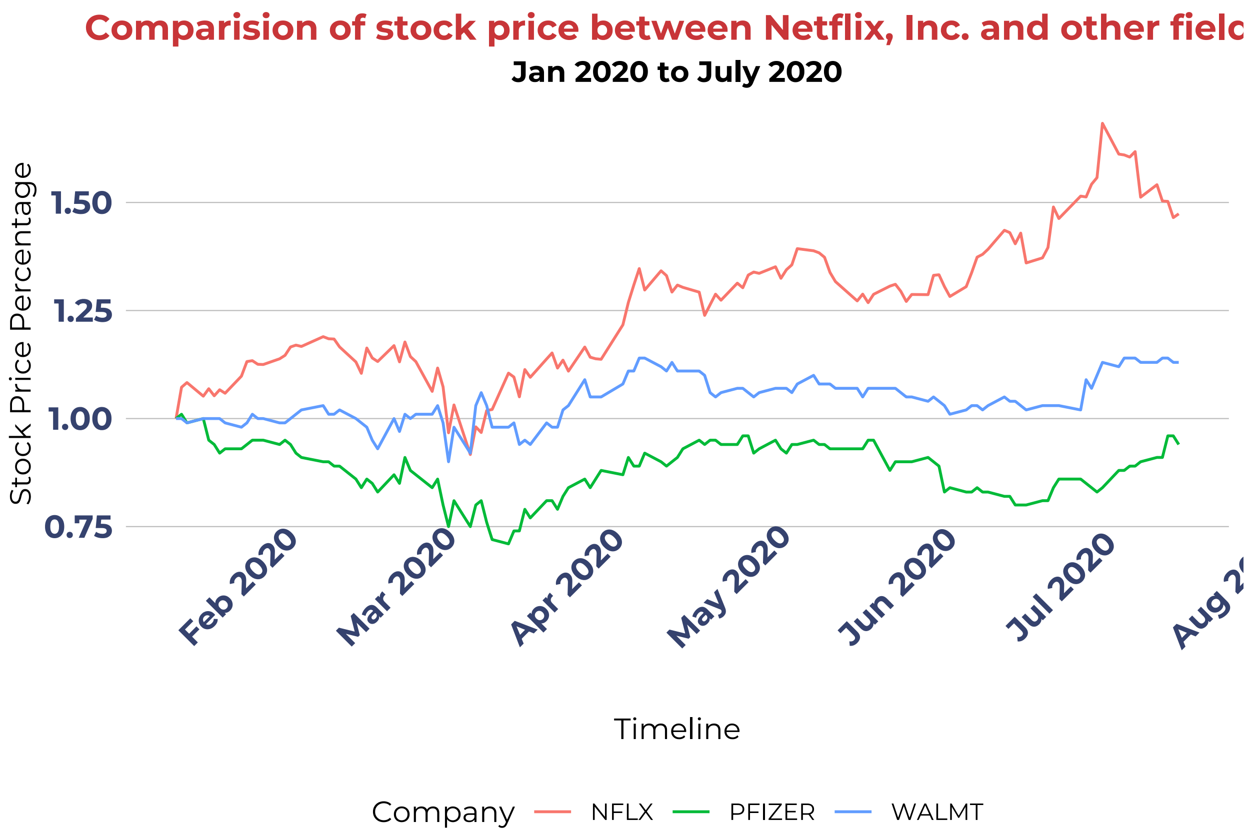



### 2.3.3. Compare tech giants stock price to giants in other fields ( Walmart: sale, Pfizer: medical)

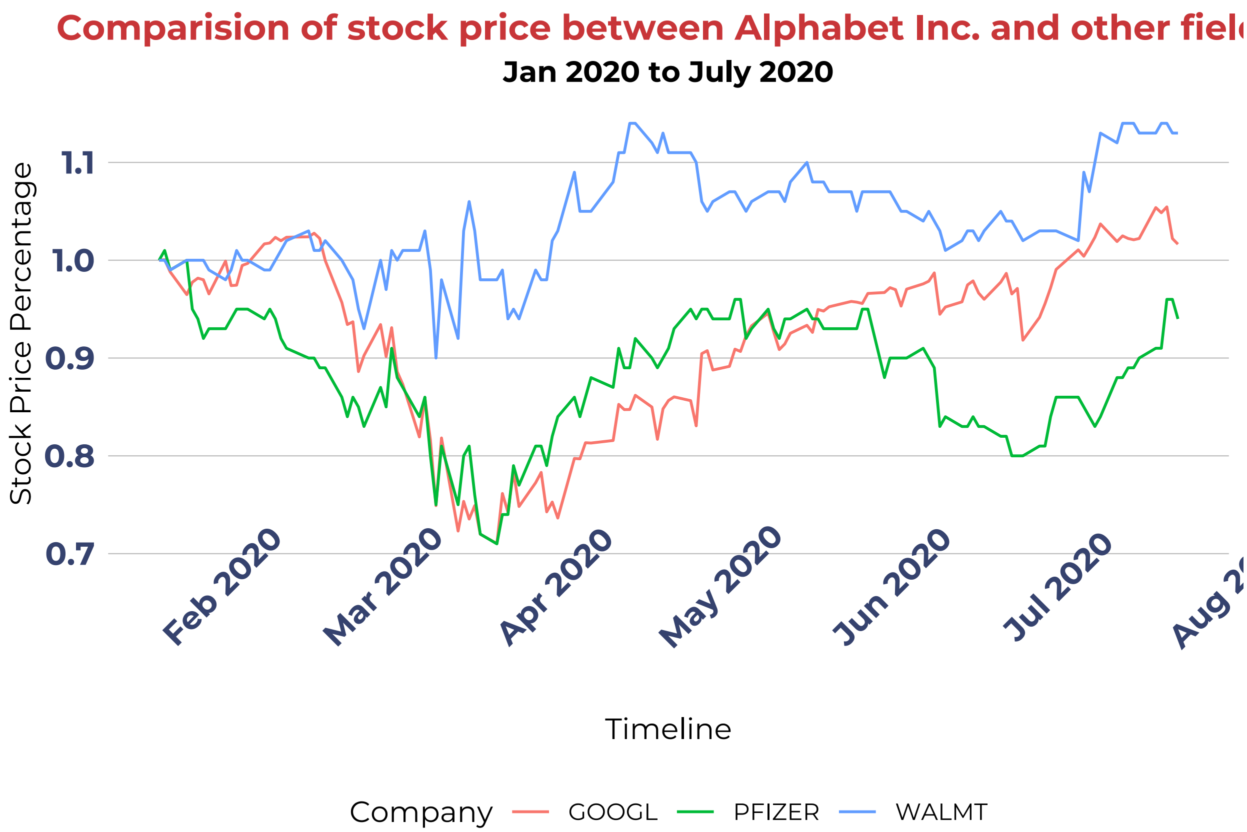
```
compare_to_other("AAPL")
```


```
compare_to_other("AMZN")
```

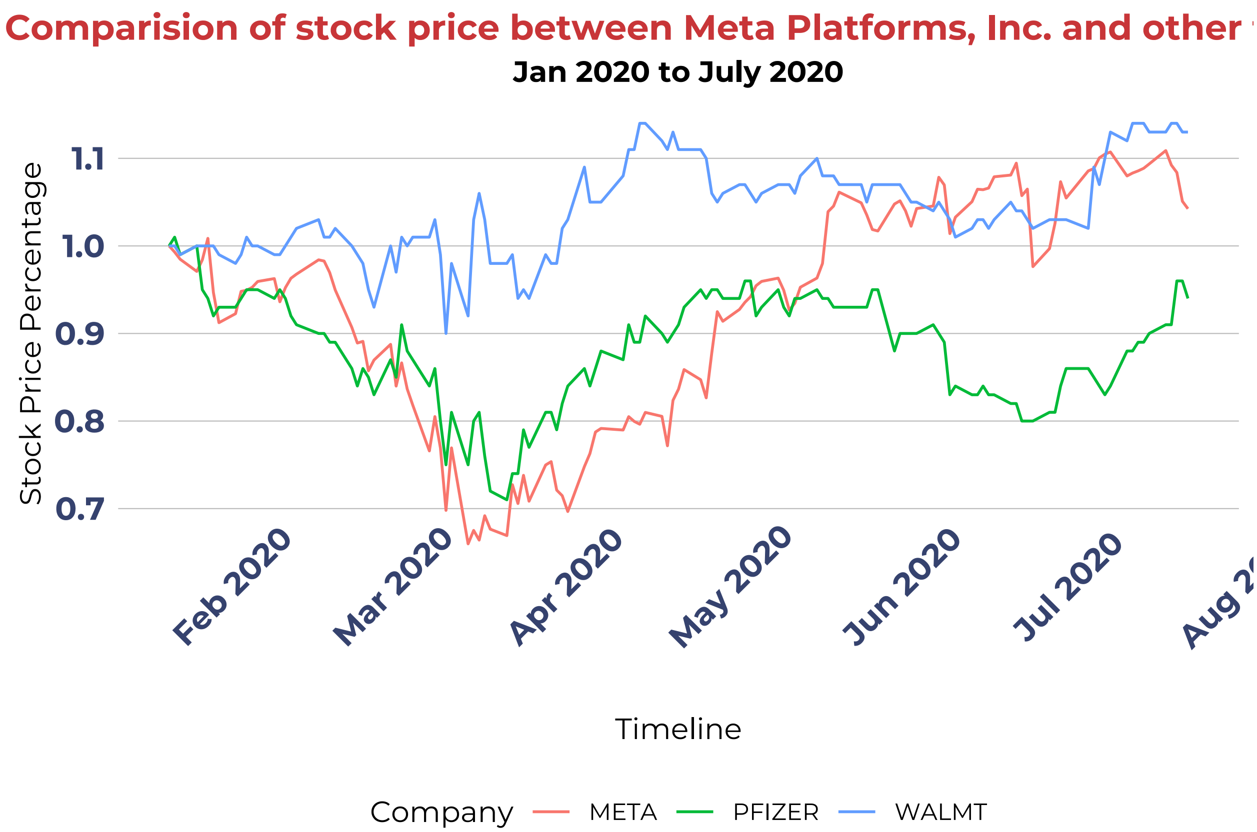

```
compare_to_other("NFLX")
```



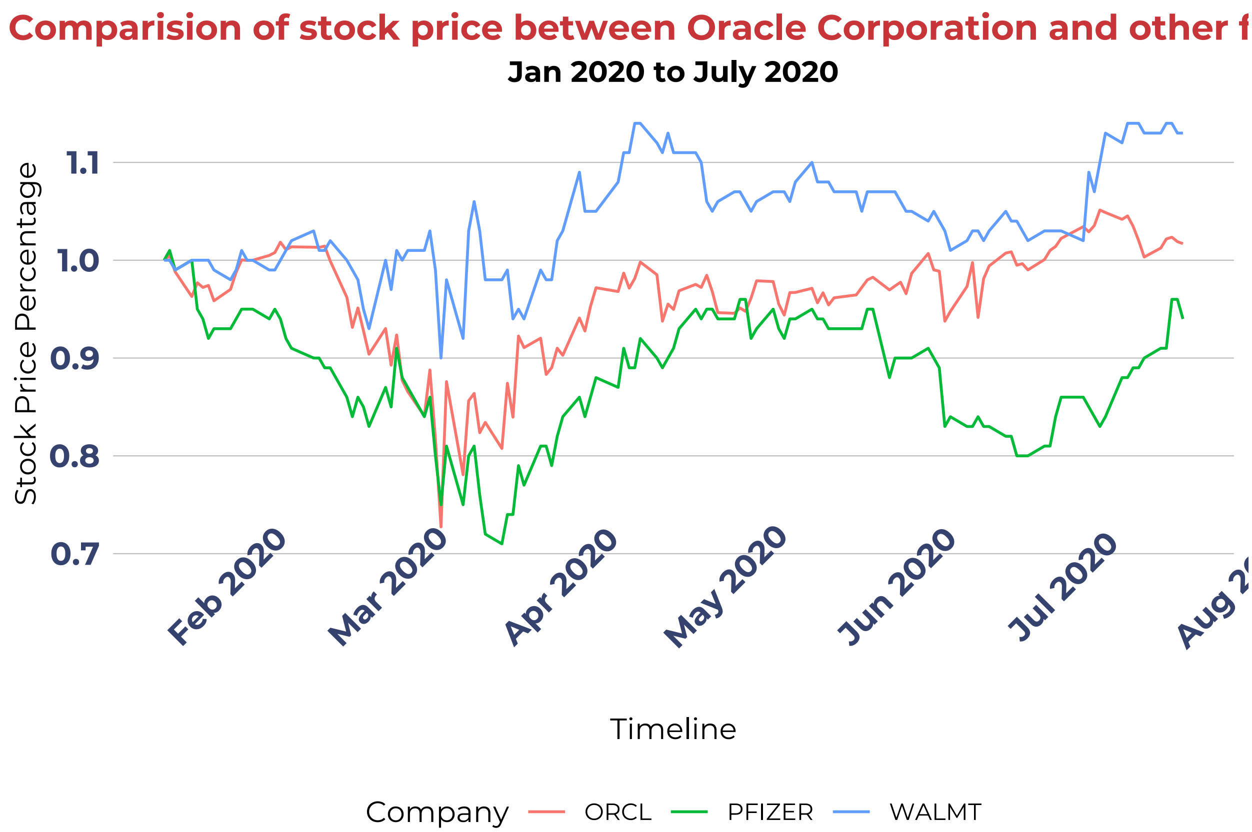

```
compare_to_other("GOOGL")
```



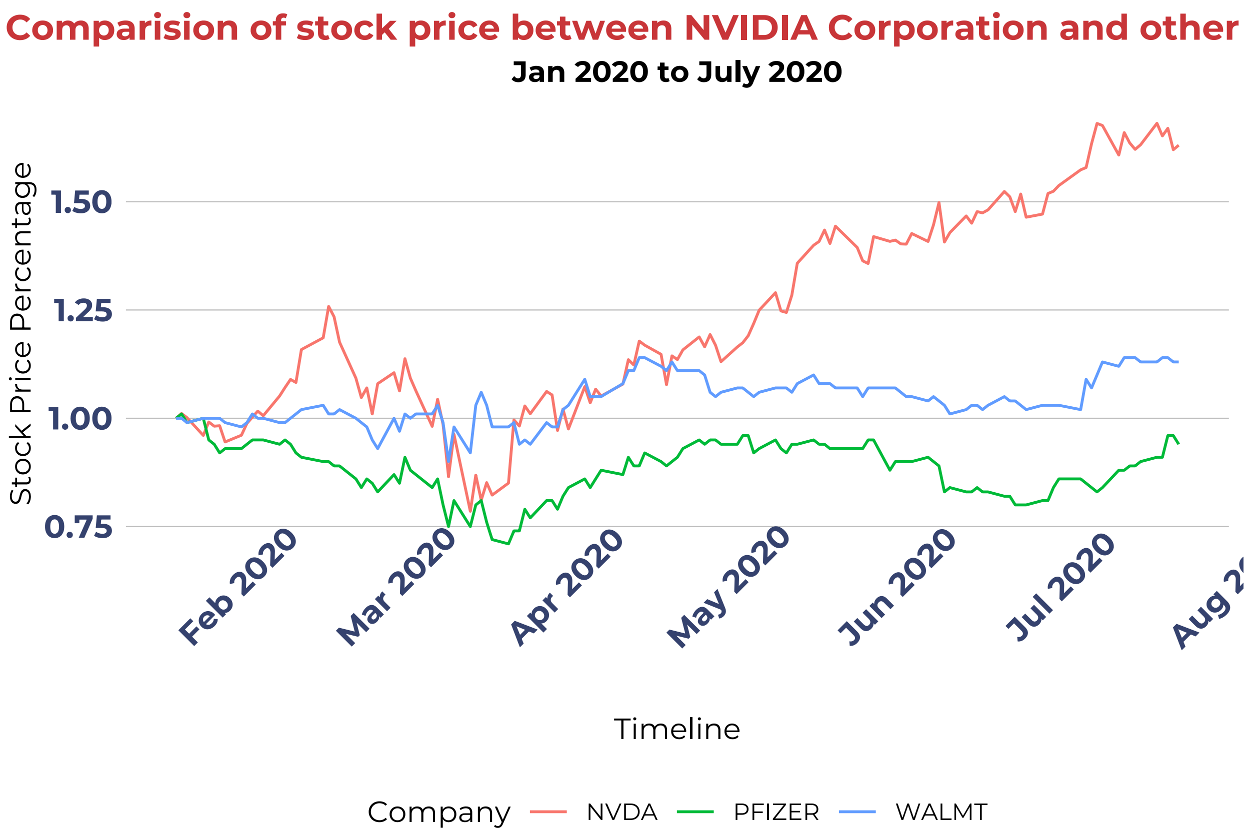

```
compare_to_other("META")
```



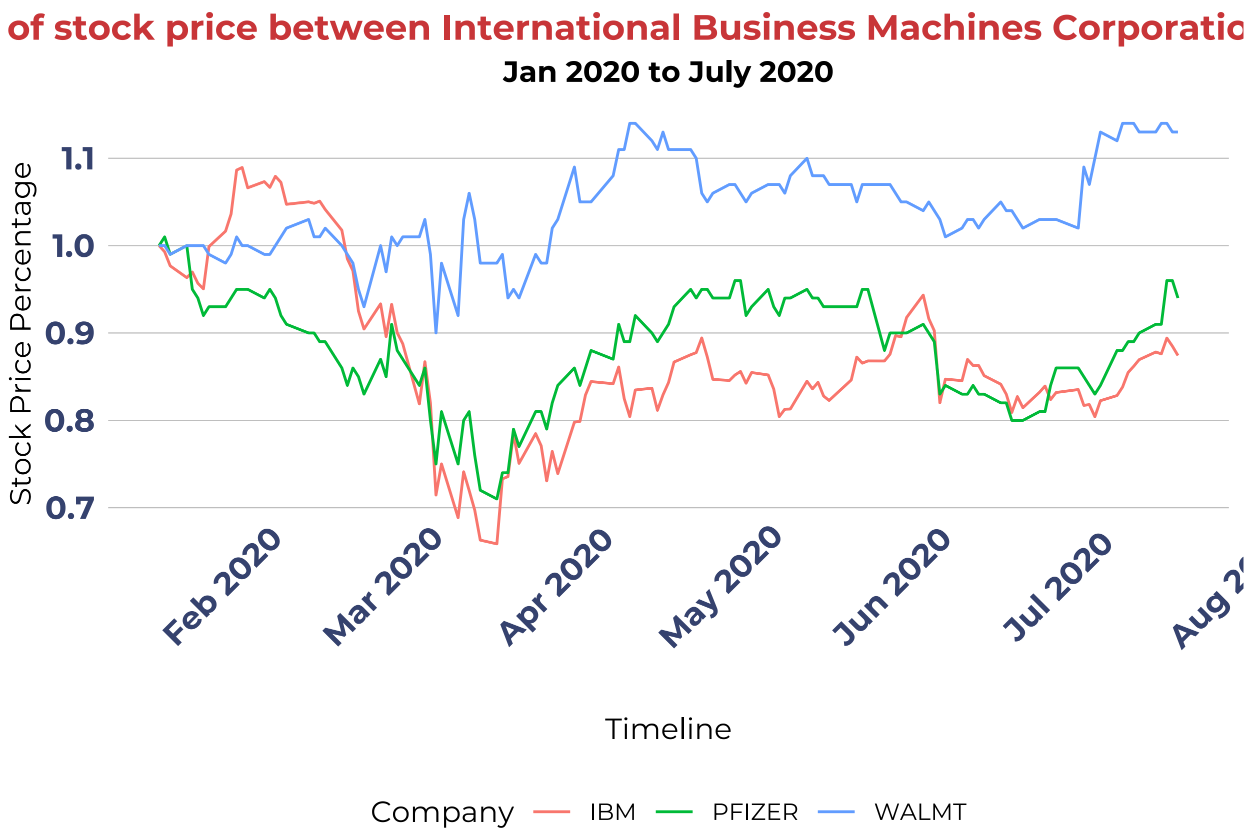

```
compare_to_other("ORCL")
```



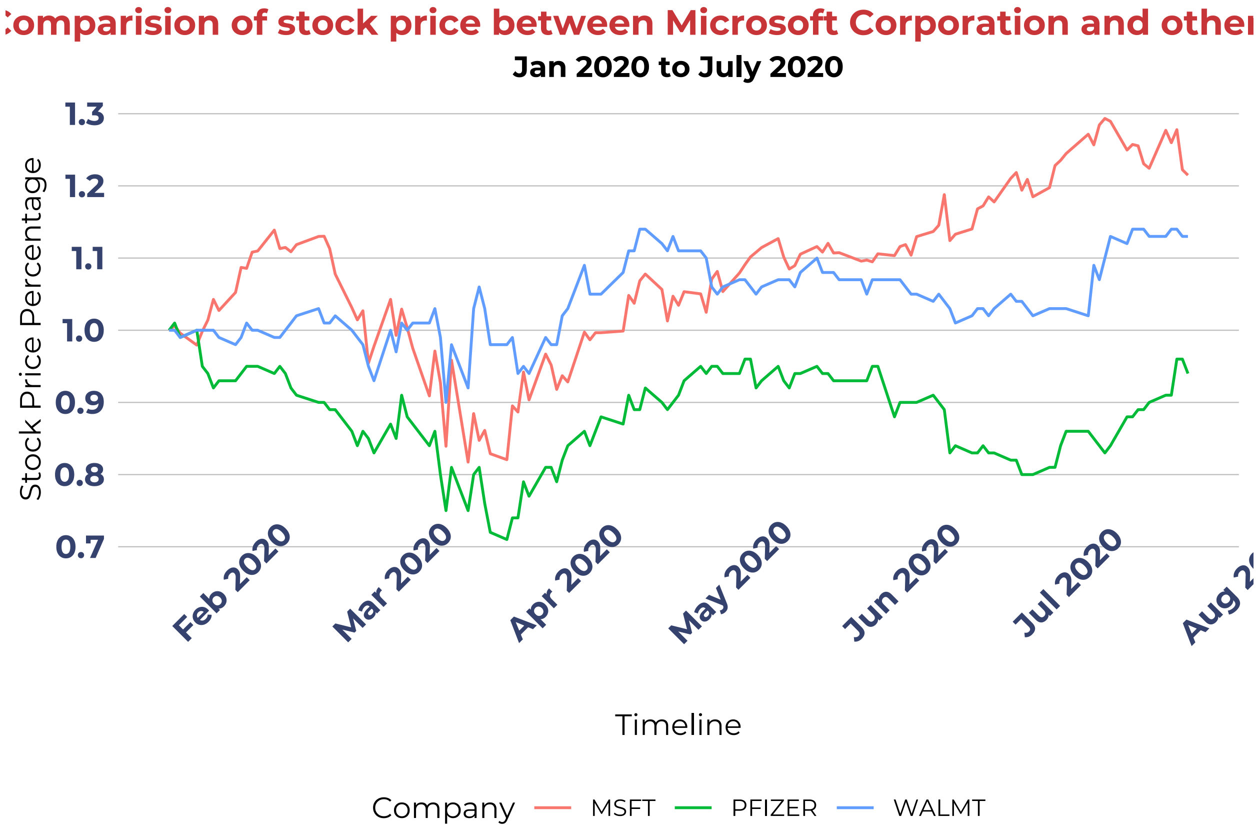

```
compare_to_other("NVDA")
```




```
compare_to_other("IBM")
```

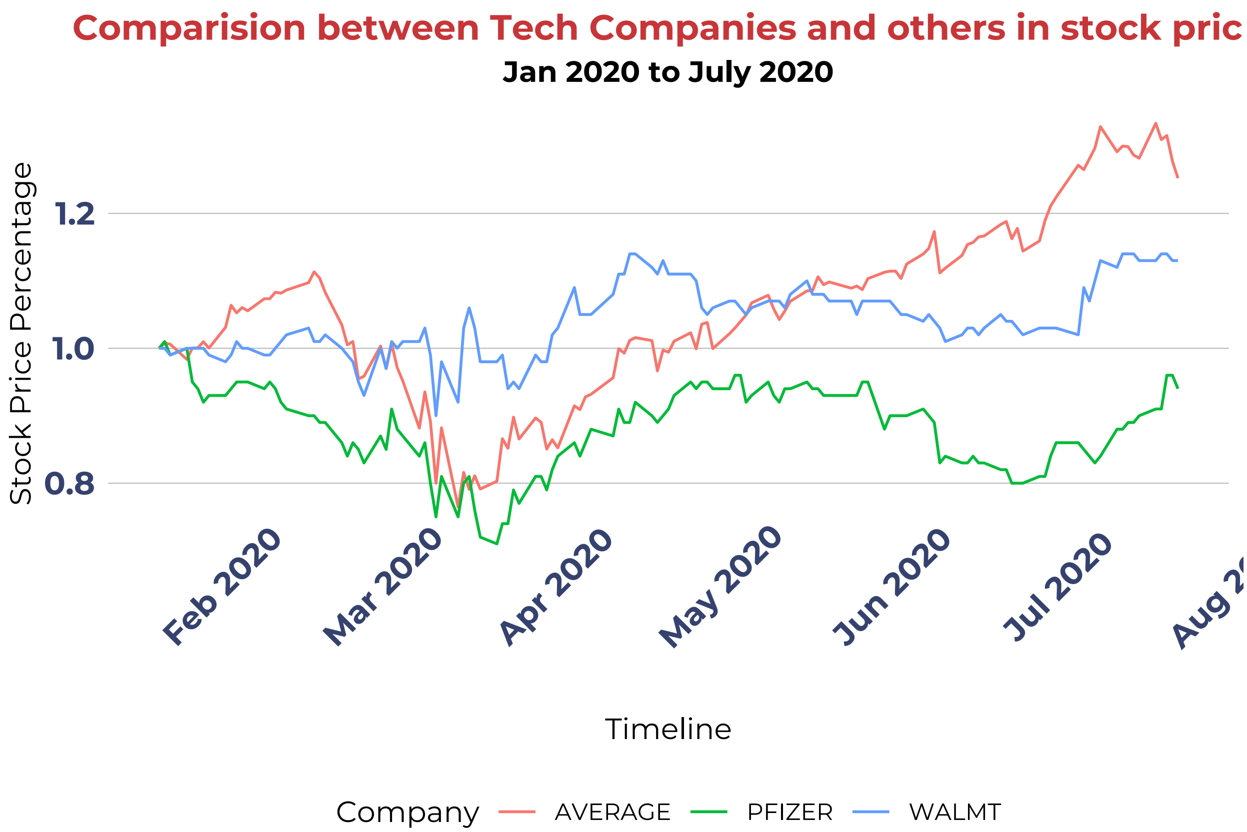



```
compare_to_other("MSFT")
```




**We want to visualize Average Stock Prices of every tech companies, therefore we can observes the influence of COVID-19 on different sectors.**

```
sec_plot <- ggplot(combine_df, aes(x =date, y = percentage,color = stock_symbol)) +
    geom_line( )+
    scale_x_date(date_breaks = "1 month", date_labels = "%b %Y")+
    labs(x = "Timeline", y = "Stock Price Percentage", color = "Company" ,title =paste("Comparision b
       subtitle = "Jan 2020 to July 2020")
sec_plot + theme_covid()
```

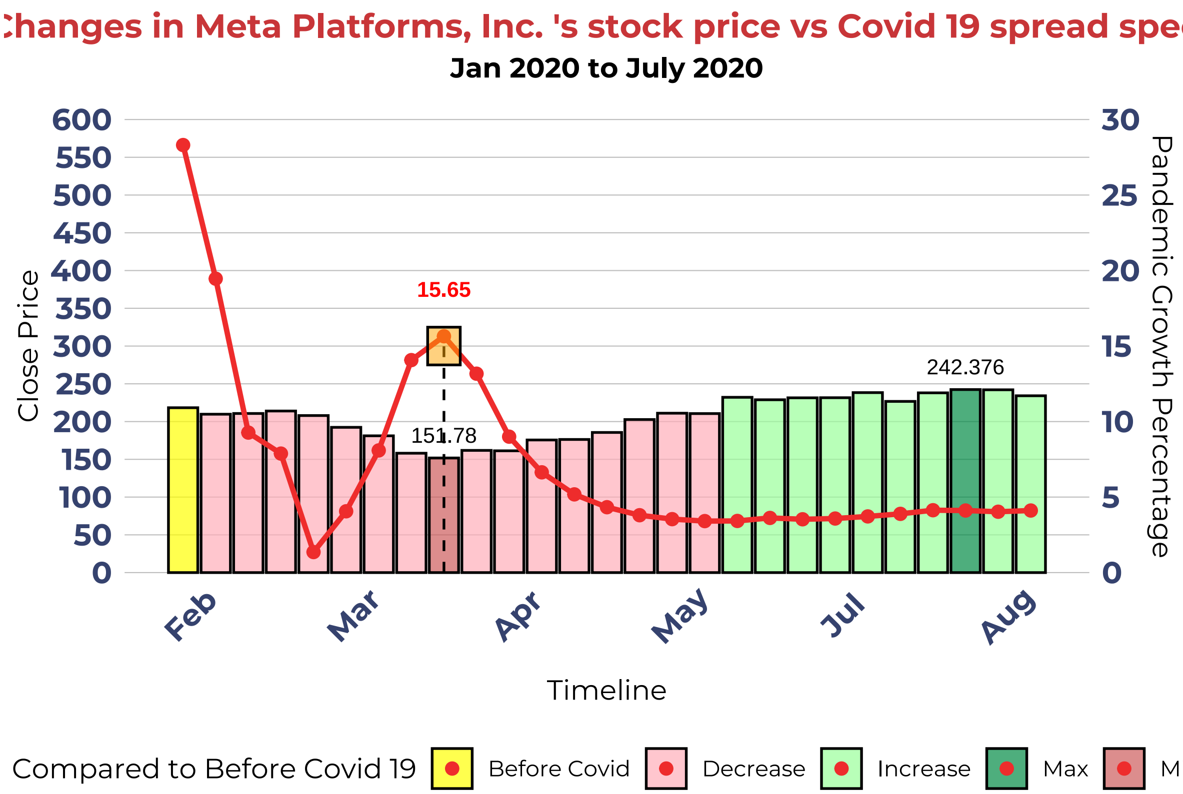



### 2.4 Discussion

- From the set of plots that illustrates the change in stock price and COVID-19 spread speed over time, we can see that the stock price seems to be in inverse ratio to the growth of COVID-19 which the faster number of COVID-19 cases increases, the faster stock prices decrease.
- Especially, as we can see from the vertical lines starting from the peak of COVID-19 growing speed, this point corresponds to the lowest stock price week in most companies. On the other hand, the highest prices appeared when COVID-19 spread more slowly in late July.
- This phenomenon is easily explained by the positive correlation between the pandemic and the global economic crisis when everyone tends to store food, medicines, etc. rather than investing in technology companies. They were willing to sell stocks in lower prices.
- Moreover, based on the number of red and green bars for each company, we observed that entertainment ( Netflix), young ( Tesla), or online shopping ( Amazon) companies suffer less damage than others. This could be thanks to adaptive and flexible policies from those companies or the high demand for online activities from customers in lockdown situations. Besides that, those bars also show that there are some companies (NVIDIA, Tesla, Netflix, Microsoft, Amazon) that recover sooner and better compared to the rest, this suggests that we should research those companies' strategies dealing with COVID-19.

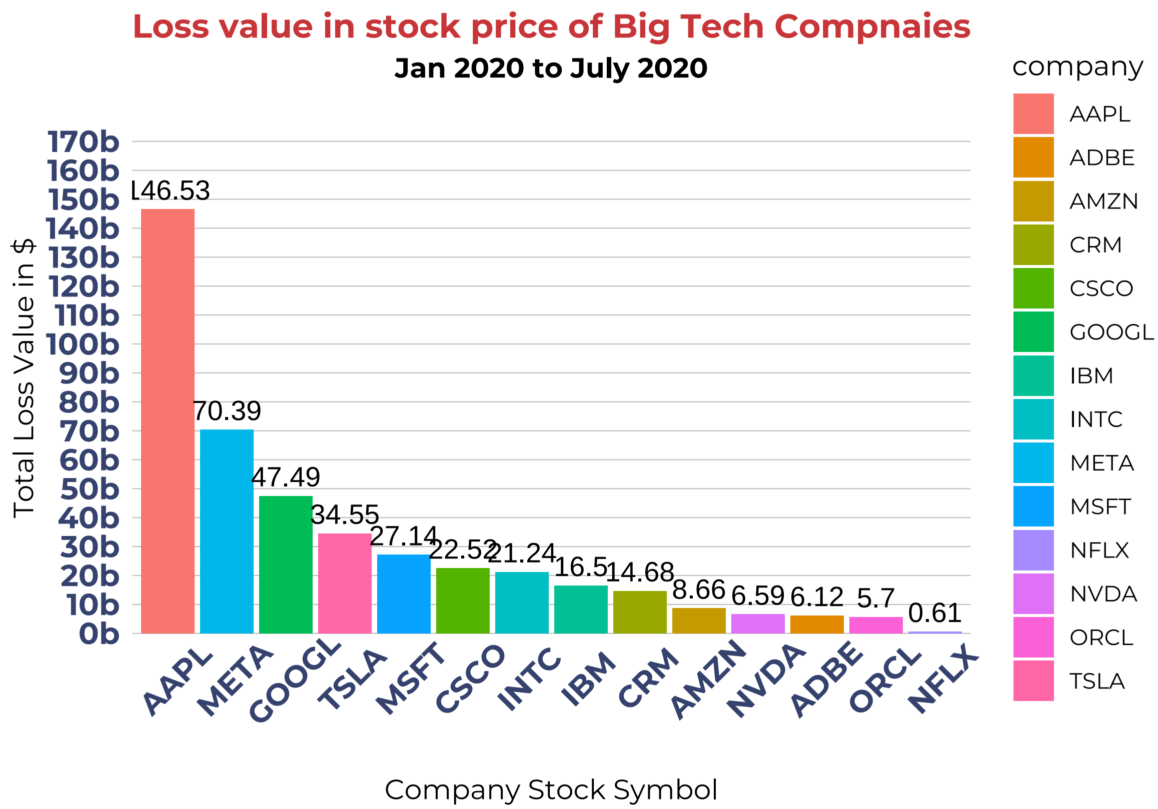
```
create_covid_plot("META")
```

```
## ‘summarise()‘ has grouped output by ’Year’. You can override using the
## ‘.groups‘ argument.
```

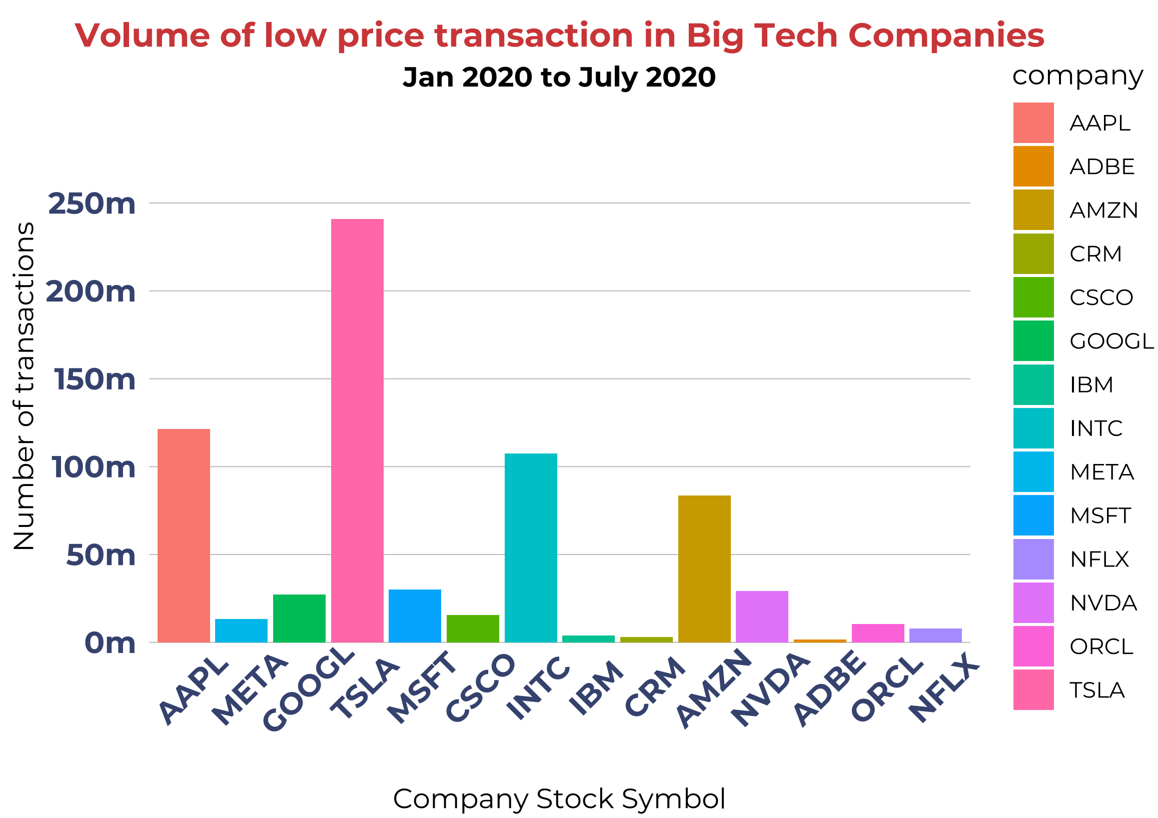



- From the second plot, the bar chart clearly shows that Netflix (0.61 billion $) is the most successful company in the pandemic situation while Apple lost the most money ( 146.53 billion $). Looking back to the previous visualization, we can see that the decrease in the price of Apple's stock is not significant, however, the thing that made them lose much money is the volume of transactions. Although they lost only tens of bucks for each sell/buy transaction, their huge number of "low-price" transactions (~121 mil, less than only Tesla) caused their significant loss.

```
loss_plot+theme_covid2()
```

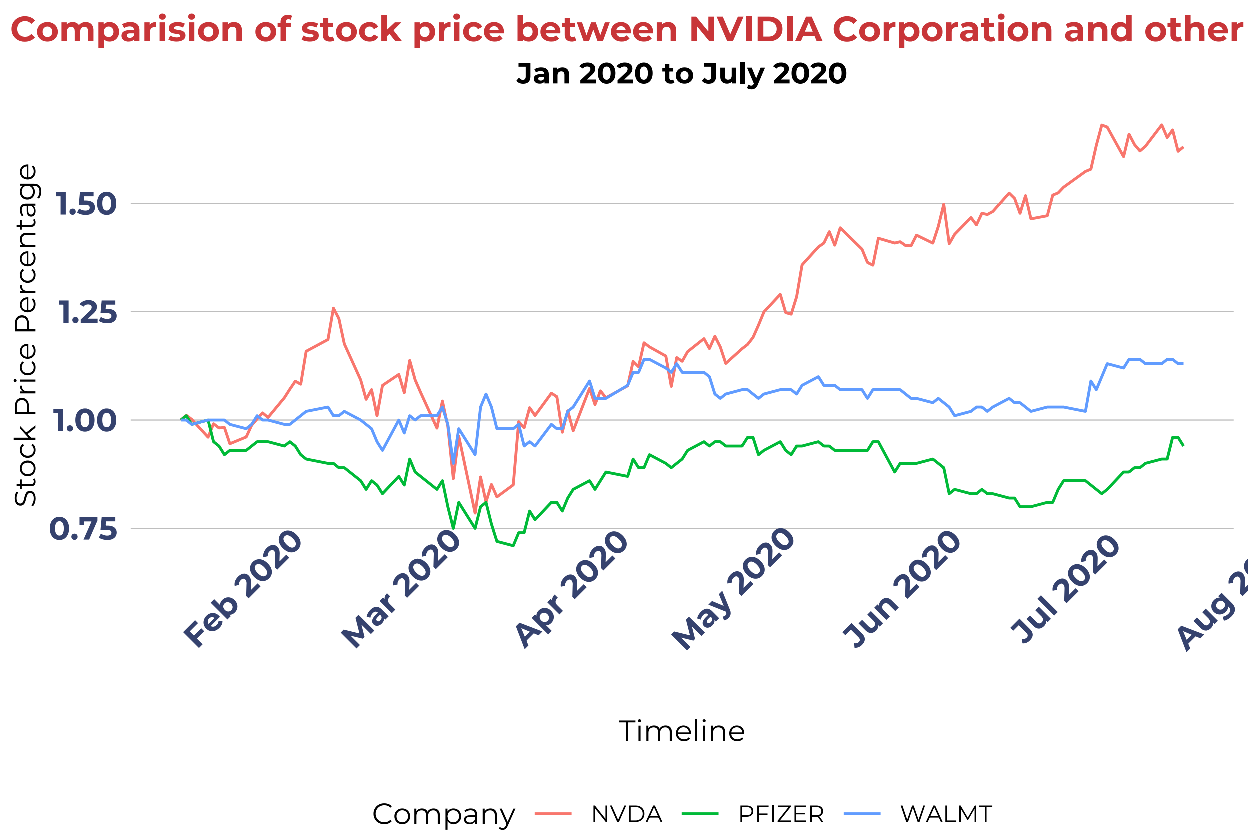



```
volume_plot+ theme_covid2()
```



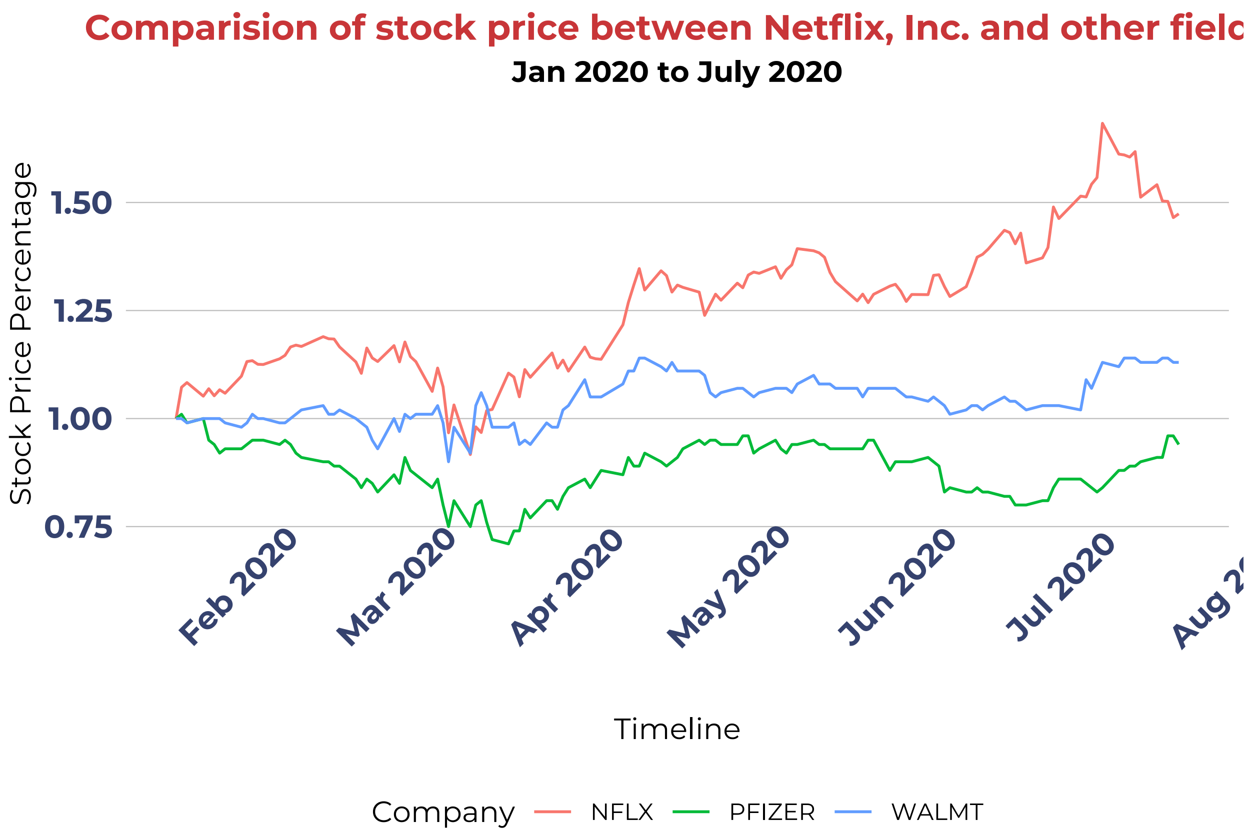
- For the final chart, we can see that except for Netflix and Amazon, other tech companies tended to perform worse than Walmart and equal compared to Pfizer. Further more, in each chart, tech companies often reach the lowest points which show that those companies were damaged more badly compared to the medical or grocery sectors. This can be explained by the demand of people in the pandemic situation, they prefer food, drink, and necessaries which can be found at Walmart rather than technological products. Therefore, the sale of this company can be better, so that it can handle the damage better as well.
- However, the greater slope of the going up part of tech companies' lines indicates that they recovered faster compared to Walmart and Pfizer after being hit by COVID-19. Some of them even reach a better stock price at the end. The average line plot shows the same story.

```
compare_to_other("NVDA")
```




```
compare_to_other("AMZN")
```


```
compare_to_other("NFLX")
```



Comparision of stock price between Netflix, Inc. and other field
Jan 2020 to July 2020
Stock Price Percentage
1.50
1.25
1.00
0.75
Feb 2020
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Timeline
Company
NFLX
PFIZER
WALMT