ADS506 Assignment 1.2

Ravita Kartawinata

2023-10-25

**Assignment: Propose a Time Series Dataset for Your Final Project**

**Dataset Description**:

Stock Market Data (NASDAQ, NYSE, S&P500) dataset is obtained from Kaggle. It is a time series data in daily basis for all tickers from inception till 12/12/2022. It covers Low, Open, Volume, High, Close and Adjusted Close values. From the data we will be able to see the fluctuation and trend.

library(tidyverse)  
library(fpp2)  
library(dplyr)  
library(ggplot2)

## Data Source

<https://www.kaggle.com/datasets/paultimothymooney/stock-market-data/data>

The source provides a lot of tickers from NASDAQ, S&P500, NYSE. In this assignment, I am going to explore technology, information system stocks. At the end I will focus on Microsoft time series data.

## Github

## https://github.com/Pii-USD/506#506

## Importing the Data

# import the data...  
amzn <- read\_csv("C:/Users/rkartawi/Desktop/Ravita/MSADS/506/Mod1/1.2/nasdaq/csv/AMZN.csv")

appl <- read\_csv("C:/Users/rkartawi/Desktop/Ravita/MSADS/506/Mod1/1.2/nasdaq/csv/AAPL.csv")

goog <- read\_csv("C:/Users/rkartawi/Desktop/Ravita/MSADS/506/Mod1/1.2/sp500/csv/GOOG.csv")

msft <- read\_csv("C:/Users/rkartawi/Desktop/Ravita/MSADS/506/Mod1/1.2/nasdaq/csv/MSFT.csv")

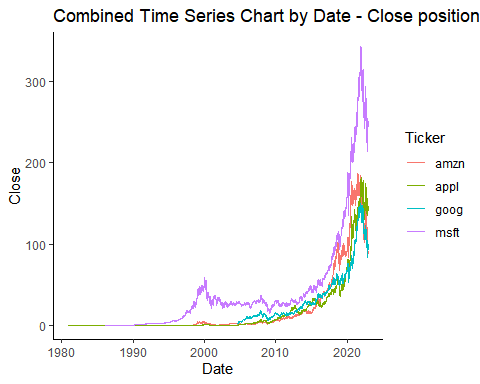
tsla <- read\_csv("C:/Users/rkartawi/Desktop/Ravita/MSADS/506/Mod1/1.2/forbes2000/csv/TSLA.csv")

# transform data  
amzn$Date <- as.Date(amzn$Date, format = "%d-%m-%Y")  
appl$Date <- as.Date(appl$Date, format = "%d-%m-%Y")  
goog$Date <- as.Date(goog$Date, format = "%d-%m-%Y")  
msft$Date <- as.Date(msft$Date, format = "%d-%m-%Y")  
  
amzn <- amzn %>% mutate(YearMonth = format(Date, "%Y-%m"),  
 Year = format(Date, "%Y"), Month = format(Date,"%m"))  
appl <- appl %>% mutate(YearMonth = format(Date, "%Y-%m"),  
 Year = format(Date, "%Y"), Month = format(Date,"%m"))  
goog <- goog %>% mutate(YearMonth = format(Date, "%Y-%m"),  
 Year = format(Date, "%Y"), Month = format(Date,"%m"))  
msft <- msft %>% mutate(YearMonth = format(Date, "%Y-%m"),  
 Year = format(Date, "%Y"), Month = format(Date,"%m"))

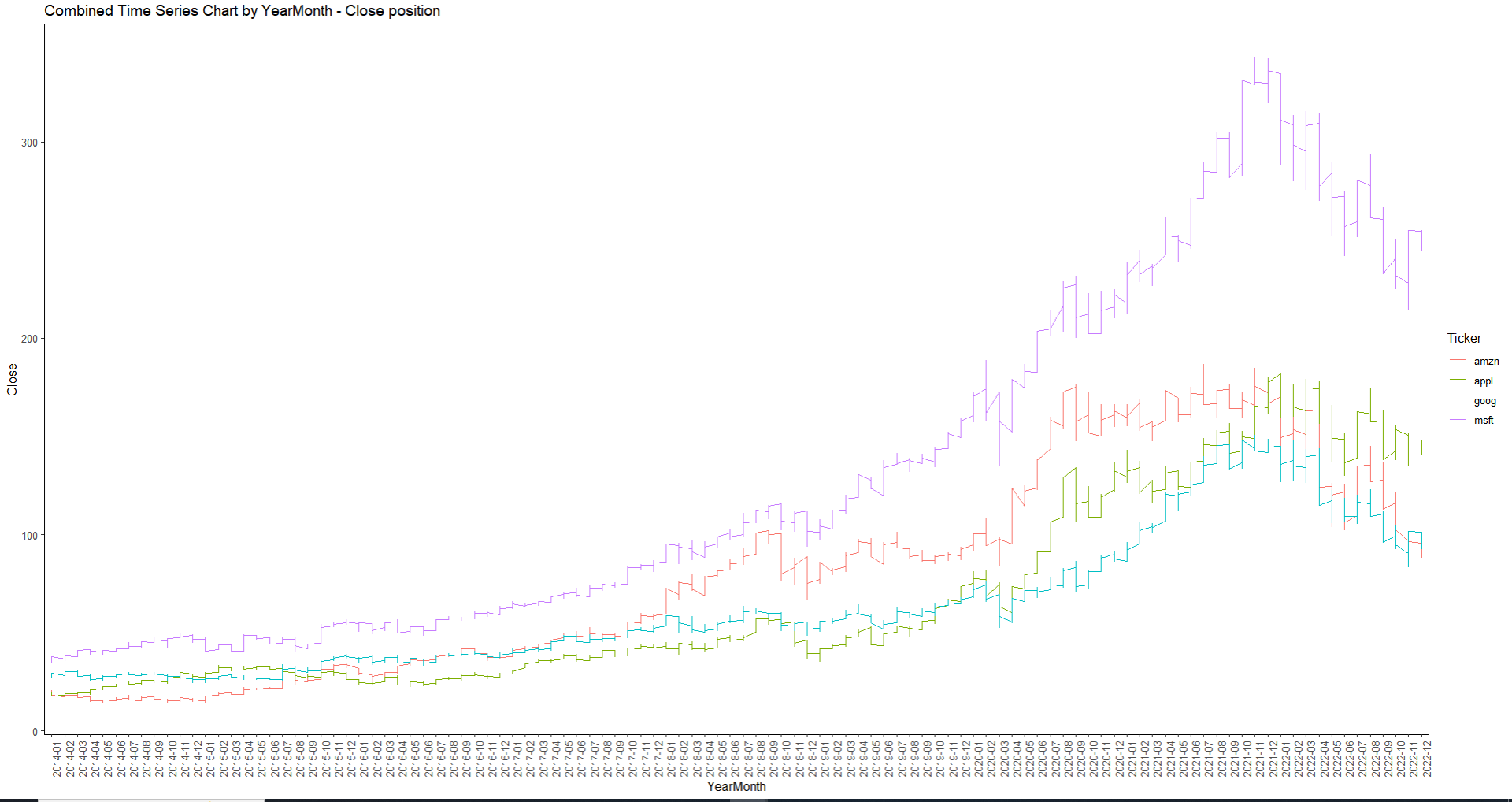
## Time Series Plot

ticker\_list <- c('amzn', 'goog', 'appl','msft')  
combined\_data <- data.frame()  
  
for (ticker in ticker\_list) {  
 data <- get(ticker)  
 data <- data.frame(  
 Date = data$Date,  
 Month = month(data$Date),  
 Year = year(data$Date),  
 YearMonth = data$YearMonth,  
 Ticker = ticker,   
 Low = data$Low,  
 Open = data$Open,  
 Volume = data$Volume,  
 High = data$High,  
 Close = data$Close,  
 Adjusted\_Close = data$`Adjusted Close`  
 )  
combined\_data <- rbind(combined\_data, data)  
}

# Create a combined line plot for 4 tickers  
combined\_plot <- ggplot(combined\_data, aes(x = Date, y = Close, group = Ticker, color = Ticker)) +  
 geom\_line() +  
 theme\_classic() +  
 labs(title = "Combined Time Series Chart by Date - Close position")  
print(combined\_plot)



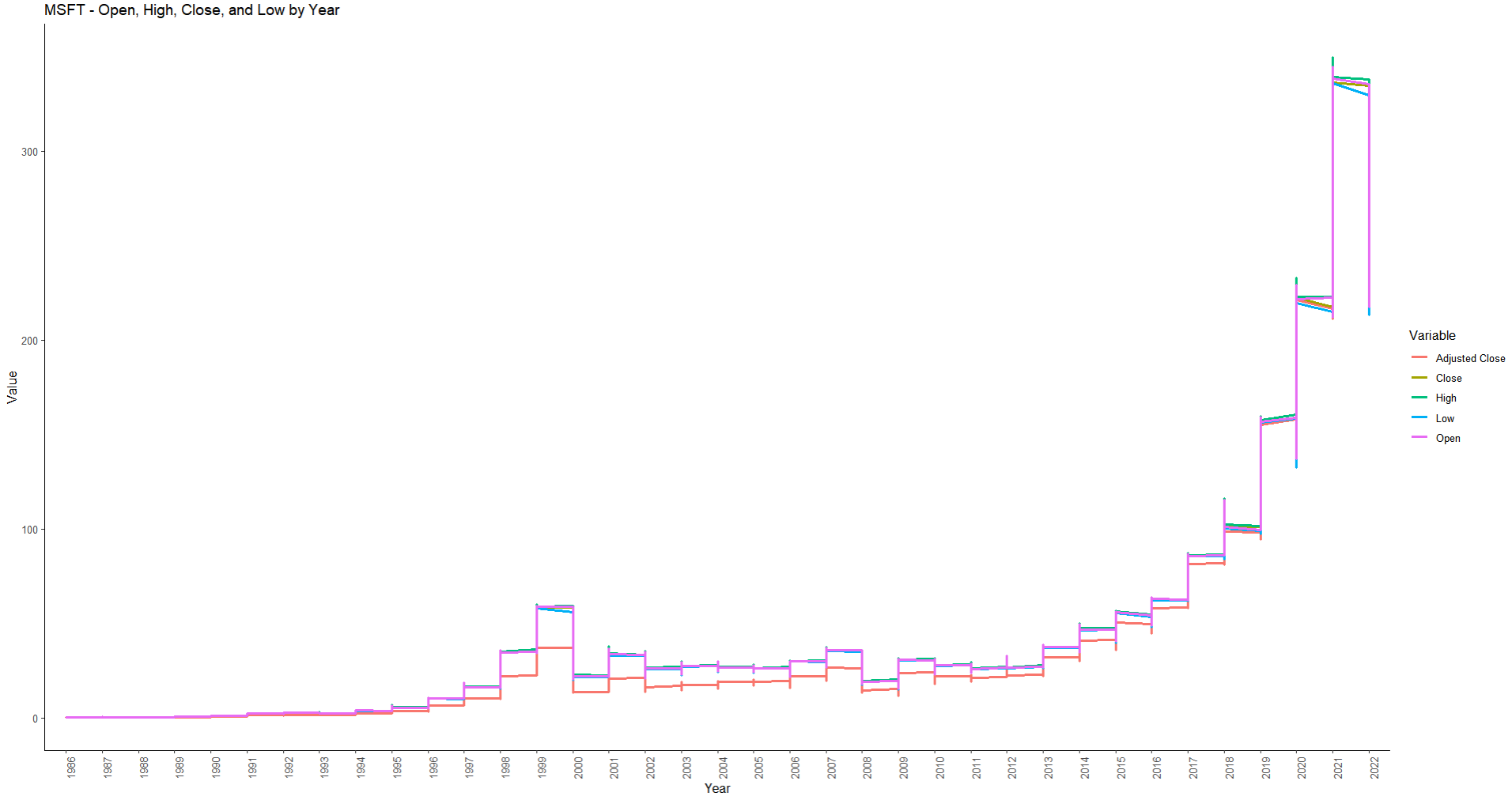
combined\_data <- combined\_data %>%  
 filter(as.numeric(Year) >= (max(as.numeric(Year)) - 8))  
  
combined\_plot <- ggplot(combined\_data, aes(x = YearMonth, y = Close, group = Ticker, color = Ticker)) +  
 geom\_line() +  
 theme\_classic() +  
 labs(title = "Combined Time Series Chart by YearMonth - Close position")+  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
  
print(combined\_plot)



Discussion

The time series data for the four technology stocks exhibited in the above plot shown a gradual increase until 2020. In that year, the rate of increase higher compared to previous years especially on MSFT, which would be the focus on this assignment. Since MSFT had a much higher spike, it might have another factor to be looked at outside of the existing data. When we drilled the data to the monthly basis, it’s worth to note that stock market time series data has a high degree of fluctuation, making stock market prediction challenging.

#Focus on MSFT; reshape to long view  
  
msft\_long <- msft %>%   
 select(Year, YearMonth, Date, Open, High, Close, Low, `Adjusted Close`) %>%  
 gather(key = "Variable", value = "Value", -Year, -YearMonth, -Date)  
ggplot(msft\_long, aes(x= Year, y = Value, color = Variable, group = Variable)) +  
 geom\_line(size = 1) +  
 labs(title = "Open, High, Close, and Low Over Time", x = "Year", y = "Value") +  
 theme\_classic()+  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))

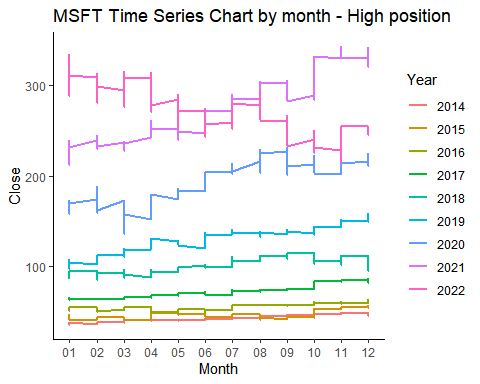


Discussion

At the beginning of the inception until 1997 and the last 5 years, the gap between Open and Adjusted Cost is quite significant compare to the gap in 1998 to 2018.

# Check MSFT movement in monthly basis to find the pattern

filter\_data <- msft %>%  
 filter(as.numeric(Year) >= (max(as.numeric(Year)) - 8))  
msft\_plot <- ggplot(filter\_data, aes(x = Month, y = Close, group = Year, color = Year)) +  
 geom\_line(size = 1) +  
 theme\_classic() +  
 labs(title = "MSFT Time Series Chart by month - Close position")  
print(msft\_plot)



Discussion

The plot above shows the previous 8 years of MSFT stock market data series. Based on the plot, there is no monthly pattern in MSFT stock market time series. As we can see although the stock price trends to go higher gradually in each year, yet we can see a good plug in Sep 2022.