STAT 184 Final Project: Stock Return Analysis

Woojoo Song Yeonkyeong Lee

2025-12-06

Table of contents

1 Introduction

This project demonstrates our team's Exploratory Data Analysis (EDA) skills using financial time-series data from Yahoo Finance.
We examine three major U.S. technology stocks:

- Apple (AAPL)
- Tesla (TSLA)
- Microsoft (MSFT)

We created two derived variables to better understand performance:

- **Cumulative Returns** — how $1 grows over time
- **Monthly Returns** — short-term fluctuations

These metrics allow us to explore long-term growth and medium-term volatility.

All results, plots, and processed data tables are included below.

2 Data Provenance

The data come from **Yahoo Finance**, accessed via the `quantmod::getSymbols()` function.

Who collected the data? Yahoo aggregates market data from U.S. stock exchanges.

What are the data? Daily closing prices for AAPL, TSLA, and MSFT.

When? Downloaded at runtime using R.

Where? Public stock exchanges → Yahoo Finance → R via quantmod.

Why? For investor transparency and historical analysis.

How? Yahoo receives data feeds directly from exchanges.

3 FAIR & CARE Evaluation

3.1 FAIR

- **Findable:** identified via ticker symbols
- **Accessible:** public API via quantmod
- **Interoperable:** numeric tabular format
- **Reusable:** saved as cleaned `.csv` files

3.2 CARE

Since the data involve *corporations*, not *human subjects*, CARE considerations are minimal.

4 Data Description

Three datasets were generated for this report:

4.1 1. daily_prices.csv Contains the raw daily prices and daily returns.

Columns:

Column	Meaning
date	Trading day
symbol	Stock ticker
close	Closing price
ret	Daily return

4.2 2. cumulative_returns.csv Adds cumulative return per symbol.

Columns:

Column	Meaning
date	Trading day
symbol	Stock
close	Closing price

Column	Meaning
ret	Daily return
cumret	Cumulative growth of $1

4.3 3. monthly_returns.csv Stores aggregated monthly returns.

Columns:

Column	Meaning
month	First day of month
symbol	Stock
monthly_ret	Total monthly return

We include summary tables of these processed datasets below.

5 Exploratory Data Analysis

5.1 Daily Prices Table (First 10 rows)

5.1.1 Apple

Table 4: AAPL Daily Prices and Returns (First 10 Trading Days)

date	symbol	close	ret	month
2007-01-03	AAPL	2.992857	0.0000000	2007-01-01
2007-01-04	AAPL	3.059286	0.0221959	2007-01-01
2007-01-05	AAPL	3.037500	-0.0071213	2007-01-01
2007-01-08	AAPL	3.052500	0.0049383	2007-01-01
2007-01-09	AAPL	3.306071	0.0830700	2007-01-01
2007-01-10	AAPL	3.464286	0.0478559	2007-01-01
2007-01-11	AAPL	3.421429	-0.0123711	2007-01-01
2007-01-12	AAPL	3.379286	-0.0123173	2007-01-01
2007-01-16	AAPL	3.467857	0.0262099	2007-01-01
2007-01-17	AAPL	3.391071	-0.0221421	2007-01-01

5.1.2 Microsoft

Table 5: MSFT Daily Prices and Returns (First 10 Trading Days)

date	symbol	close	ret	month
2007-01-03	MSFT	29.86	0.0000000	2007-01-01
2007-01-04	MSFT	29.81	-0.0016745	2007-01-01
2007-01-05	MSFT	29.64	-0.0057028	2007-01-01
2007-01-08	MSFT	29.93	0.0097841	2007-01-01
2007-01-09	MSFT	29.96	0.0010023	2007-01-01
2007-01-10	MSFT	29.66	-0.0100133	2007-01-01
2007-01-11	MSFT	30.70	0.0350641	2007-01-01
2007-01-12	MSFT	31.21	0.0166123	2007-01-01
2007-01-16	MSFT	31.16	-0.0016020	2007-01-01
2007-01-17	MSFT	31.10	-0.0019255	2007-01-01

5.1.3 Tesla

Table 6: TSLA Daily Prices and Returns (First 10 Trading Days)

date	symbol	close	ret	month
2010-06-29	TSLA	1.592667	0.0000000	2010-06-01
2010-06-30	TSLA	1.588667	-0.0025115	2010-06-01
2010-07-01	TSLA	1.464000	-0.0784727	2010-07-01
2010-07-02	TSLA	1.280000	-0.1256831	2010-07-01
2010-07-06	TSLA	1.074000	-0.1609375	2010-07-01
2010-07-07	TSLA	1.053333	-0.0192430	2010-07-01
2010-07-08	TSLA	1.164000	0.1050636	2010-07-01
2010-07-09	TSLA	1.160000	-0.0034365	2010-07-01
2010-07-12	TSLA	1.136667	-0.0201146	2010-07-01
2010-07-13	TSLA	1.209333	0.0639290	2010-07-01

5.2 Cumulative Returns Table (First 10 rows)

5.2.1 Apple

Table 7: AAPL Cumulative Returns (First 10 Trading Days)

date	symbol	close	ret	cumret
2007-01-03	AAPL	2.992857	0.0000000	1.000000
2007-01-04	AAPL	3.059286	0.0221959	1.022196
2007-01-05	AAPL	3.037500	-0.0071213	1.014916
2007-01-08	AAPL	3.052500	0.0049383	1.019929
2007-01-09	AAPL	3.306071	0.0830700	1.104654
2007-01-10	AAPL	3.464286	0.0478559	1.157518
2007-01-11	AAPL	3.421429	-0.0123711	1.143198
2007-01-12	AAPL	3.379286	-0.0123173	1.129117
2007-01-16	AAPL	3.467857	0.0262099	1.158711
2007-01-17	AAPL	3.391071	-0.0221421	1.133055

5.2.2 Microsoft

Table 8: MSFT Cumulative Returns (First 10 Trading Days)

date	symbol	close	ret	cumret
2007-01-03	MSFT	29.86	0.0000000	1.0000000
2007-01-04	MSFT	29.81	-0.0016745	0.9983255
2007-01-05	MSFT	29.64	-0.0057028	0.9926322
2007-01-08	MSFT	29.93	0.0097841	1.0023443
2007-01-09	MSFT	29.96	0.0010023	1.0033489
2007-01-10	MSFT	29.66	-0.0100133	0.9933021
2007-01-11	MSFT	30.70	0.0350641	1.0281313
2007-01-12	MSFT	31.21	0.0166123	1.0452109
2007-01-16	MSFT	31.16	-0.0016020	1.0435365
2007-01-17	MSFT	31.10	-0.0019255	1.0415271

5.2.3 Tesla

Table 9: TSLA Cumulative Returns (First 10 Trading Days)

date	symbol	close	ret	cumret
2010-06-29	TSLA	1.592667	0.0000000	1.0000000
2010-06-30	TSLA	1.588667	-0.0025115	0.9974885
2010-07-01	TSLA	1.464000	-0.0784727	0.9192129
2010-07-02	TSLA	1.280000	-0.1256831	0.8036834

Table 9: TSLA Cumulative Returns (First 10 Trading Days)

date	symbol	close	ret	cumret
2010-07-06	TSLA	1.074000	-0.1609375	0.6743406
2010-07-07	TSLA	1.053333	-0.0192430	0.6613643
2010-07-08	TSLA	1.164000	0.1050636	0.7308496
2010-07-09	TSLA	1.160000	-0.0034365	0.7283381
2010-07-12	TSLA	1.136667	-0.0201146	0.7136878
2010-07-13	TSLA	1.209333	0.0639290	0.7593131

5.3 Monthly Returns Table (First 10 rows)

5.3.1 Apple

Table 10: AAPL Monthly Returns (First 10 Months)

symbol	month	monthly_ret
AAPL	2007-01-01	0.0230312
AAPL	2007-02-01	-0.0130643
AAPL	2007-03-01	0.0980969
AAPL	2007-04-01	0.0741580
AAPL	2007-05-01	0.2143285
AAPL	2007-06-01	0.0070137
AAPL	2007-07-01	0.0796461
AAPL	2007-08-01	0.0510018
AAPL	2007-09-01	0.1082467
AAPL	2007-10-01	0.2377013

5.3.2 Microsoft

Table 11: MSFT Monthly Returns (First 10 Months)

symbol	month	monthly_ret
MSFT	2007-01-01	0.0334896
MSFT	2007-02-01	-0.0871679
MSFT	2007-03-01	-0.0106496
MSFT	2007-04-01	0.0742734
MSFT	2007-05-01	0.0250501
MSFT	2007-06-01	-0.0397524

Table 11: MSFT Monthly Returns (First 10 Months)

symbol	month	monthly_ret
MSFT	2007-07-01	-0.0162877
MSFT	2007-08-01	-0.0089686
MSFT	2007-09-01	0.0254090
MSFT	2007-10-01	0.2494909

5.3.3 Tesla

Table 12: TSLA Monthly Returns (First 10 Months)

symbol	month	monthly_ret
TSLA	2010-06-01	-0.0025115
TSLA	2010-07-01	-0.1632401
TSLA	2010-08-01	-0.0230687
TSLA	2010-09-01	0.0477413
TSLA	2010-10-01	0.0700634
TSLA	2010-11-01	0.6176739
TSLA	2010-12-01	-0.2462497
TSLA	2011-01-01	-0.0950053
TSLA	2011-02-01	-0.0087137
TSLA	2011-03-01	0.1615737

5.4 Cumulative Return Plot

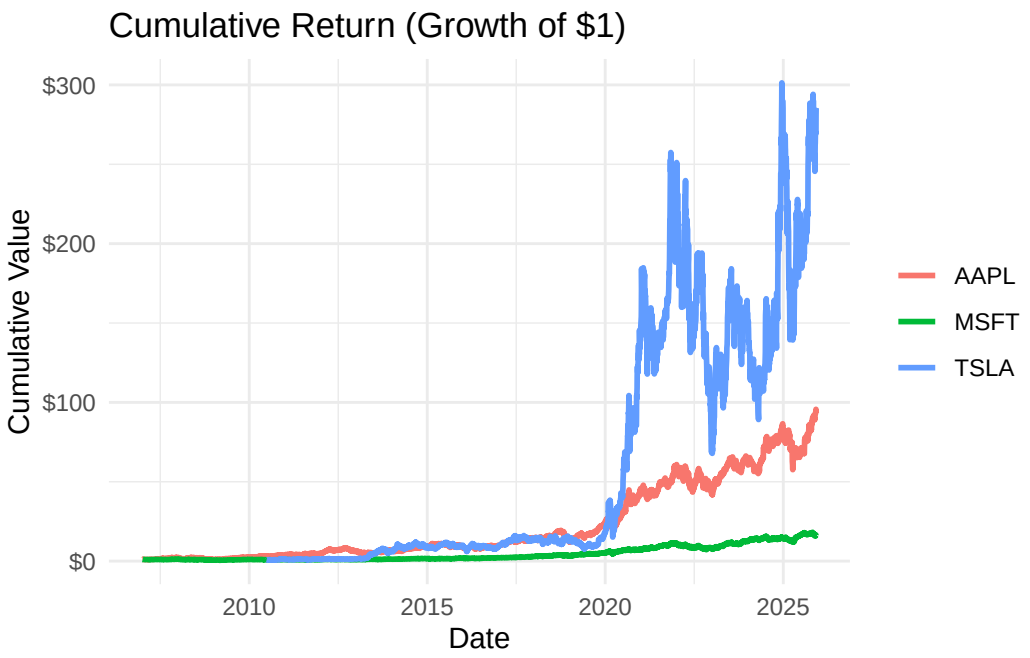

Figure 1: Cumulative Return Curves for AAPL, TSLA, and MSFT (Growth of $1)

5.5 Monthly Return Plot

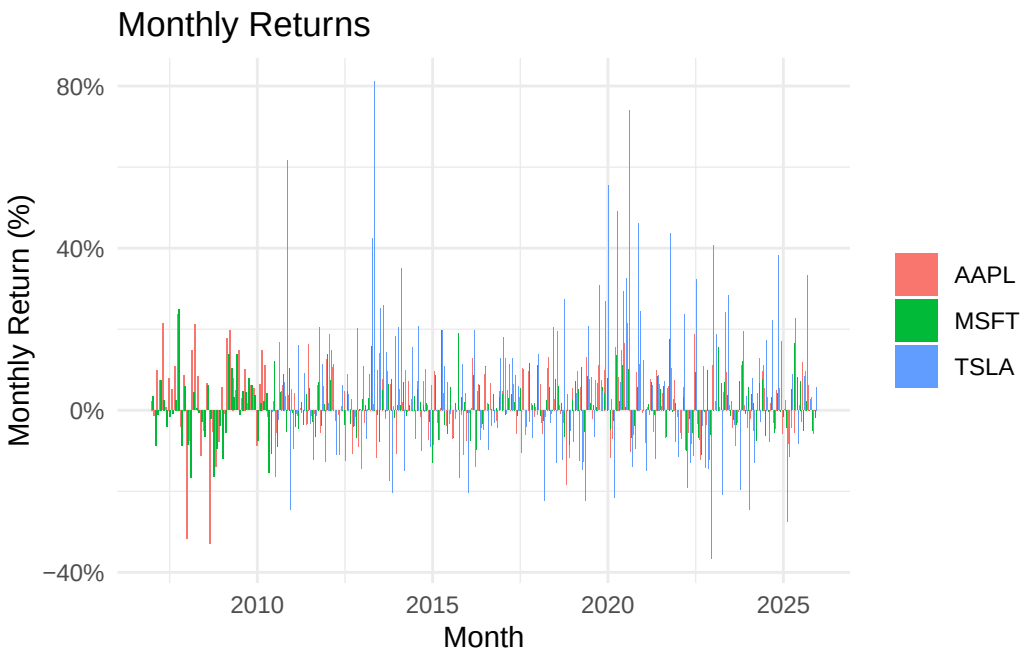

Figure 2: Monthly Returns (Percentage) for AAPL, TSLA, and MSFT

5.6 Discussion/Data Story

By examining both long-term and monthly performance:

- Tesla is the most volatile stock in both cumulative and monthly perspectives.
- Apple and Microsoft show more stable, predictable growth.
- Monthly returns highlight short-term risk, while cumulative returns show long-term compounding patterns.

This combination reveals how investor risk profiles align with different stocks.

5.7 Open Science & Reproducibility

We ensured reproducibility by:

- Using publicly accessible data
- Including all processed .csv files

- Saving plots to /plots
- Providing a complete Code Appendix
- Following Tidyverse/BOAST style guide
- Using GitHub

5.8 Author Contributions

- Woojoo Song: Data wrangling, plotting, report editing
- Yeonkyeong Lee: Presentation design, FAIR/CARE evaluation, scriptwriting

5.9 References

Yahoo Finance. (2025). Historical Stock Data. https://finance.yahoo.com/

5.10 Code Appendix

```
# Load Packages
options(repos = c(CRAN = "https://cloud.r-project.org"))
install.packages("quantmod")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("lubridate")

library(quantmod)
library(dplyr)
library(ggplot2)
library(lubridate)

# Create folders
if (!dir.exists("plots")) {
  dir.create("plots")
}
if (!dir.exists("data")) {
  dir.create("data")
}

# Download Data
```

```r
symbols <- c("AAPL", "TSLA", "MSFT")
all_data <- data.frame()

for (sym in symbols) {
  stock <- getSymbols(sym, src = "yahoo", auto.assign = FALSE)
  price <- Cl(stock)
  ret <- dailyReturn(price)

  temp <- data.frame(
    date = as.Date(index(price)),
    symbol = sym,
    close = as.numeric(price),
    ret = as.numeric(ret)
  )

  all_data <- rbind(all_data, temp)
}

# Sort data
all_data <- all_data %>%
  arrange(symbol, date) %>%
  mutate(ret = ifelse(is.na(ret), 0, ret))

# Cumulative Return
df_cum <- all_data %>%
  group_by(symbol) %>%
  mutate(ret = ifelse(is.na(ret), 0, ret),cumret = cumprod(1 + ret))

# Monthly Return
all_data$month <- floor_date(all_data$date, "month")

df_month <- all_data %>%
  mutate(month = floor_date(date, "month")) %>%
  group_by(symbol, month) %>%
  summarise(
    monthly_ret = prod(1 + ret, na.rm = TRUE) - 1,
    .groups = 'drop'
  )
# Plot: Cumulative Returns
p_cum <- ggplot(df_cum, aes(x = date, y = cumret, color = symbol)) +
  geom_line(linewidth = 1) +
  scale_y_continuous(labels = scales::dollar_format()) +
```

```r
  labs(title = "Cumulative Return (Growth of $1)", x = "Date", y = "Cumulative Value") +
  theme_minimal() +
  theme(legend.title = element_blank())

ggsave("plots/cumulative_returns.png", p_cum, width = 8, height = 5)

# Plot: Monthly Returns
p_month <- ggplot(df_month, aes(x = month, y = monthly_ret, fill = symbol)) + # color   fill
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Monthly Returns", x = "Month", y = "Monthly Return (%)") +
  theme_minimal() +
  theme(legend.title = element_blank())

ggsave("plots/monthly_returns.png", p_month, width = 8, height = 5)

# Save data
write.csv(all_data, "data/daily_prices.csv", row.names = FALSE)
write.csv(df_cum, "data/cumulative_returns.csv", row.names = FALSE)
write.csv(df_month, "data/monthly_returns.csv", row.names = FALSE)
```