# Crypto Exploratory Data ANALYSIS and Feature Engineering

Yogesh Sahu

01/05/2024

## Table of Contents

[Introduction 1](#_Toc165549759)

[Load Data and Extract Data Based on Price Action Analysis 2](#_Toc165549760)

[Data Prep - Adding Returns / log Returns 2](#_Toc165549761)

[Key Observations 8](#_Toc165549762)

[Investment Considerations 8](#_Toc165549763)

[Investment Strategy Considerations 9](#_Toc165549764)

[Conclusion 9](#_Toc165549765)

[Halving dates pre n post halving dates 10](#_Toc165549766)

[Halving Analysis 13](#_Toc165549767)

[Impact of Halvings on Market Capitalization: 15](#_Toc165549768)

[Market Capitalization Segmentation 16](#_Toc165549769)

# Introduction

*Cryptocurrencies have emerged as a significant player in the global financial landscape, with Bitcoin and Ethereum leading the pack. Understanding the intricate dynamics of these digital assets is paramount for investors, traders, and analysts alike. This report embarks on a comprehensive analysis journey, employing exploratory data analysis (EDA) and feature engineering (FE) techniques to gain profound insights into cryptocurrency price movements and market behavior.*

##### Step 1: Loading and Preprocessing Historical Data

The analysis initiates with the ingestion of historical cryptocurrency data. Through meticulous preprocessing and data wrangling techniques, we extract relevant information based on price action. This includes crucial steps such as handling missing values, standardizing data formats, and merging datasets to enrich the analytical scope.

##### Step 2: Uncovering Trends and Patterns

Once the data is prepared, we delve into exploratory data analysis (EDA) to uncover underlying trends, patterns, and relationships. While basic visualizations provide initial insights, we’ll leverage advanced techniques such as feature engineering to extract more meaningful information from the data. Utilizing visualizations and statistical summaries, we aim to unravel the complex dynamics of cryptocurrency markets. This includes identifying cyclical trends, seasonal patterns, and anomalous behavior that may impact investment decisions.

##### Step 3: Investigating Halving Events

A crucial aspect of our analysis involves investigating the impact of halving events on cryptocurrency prices. Halving events, which occur at predetermined intervals for certain cryptocurrencies like Bitcoin, have historically influenced market dynamics. By examining price action surrounding these events, we seek to discern their implications for market sentiment and investor behavior.

##### Step 4: Market Cap Segmentation

In addition to price movements, we delve into market cap segmentation to gain insights into the distribution of value across different cryptocurrencies. This segmentation provides a nuanced understanding of market dynamics, highlighting the relative prominence of various digital assets within the cryptocurrency ecosystem.

#### Advanced Visualization and Model Preparation

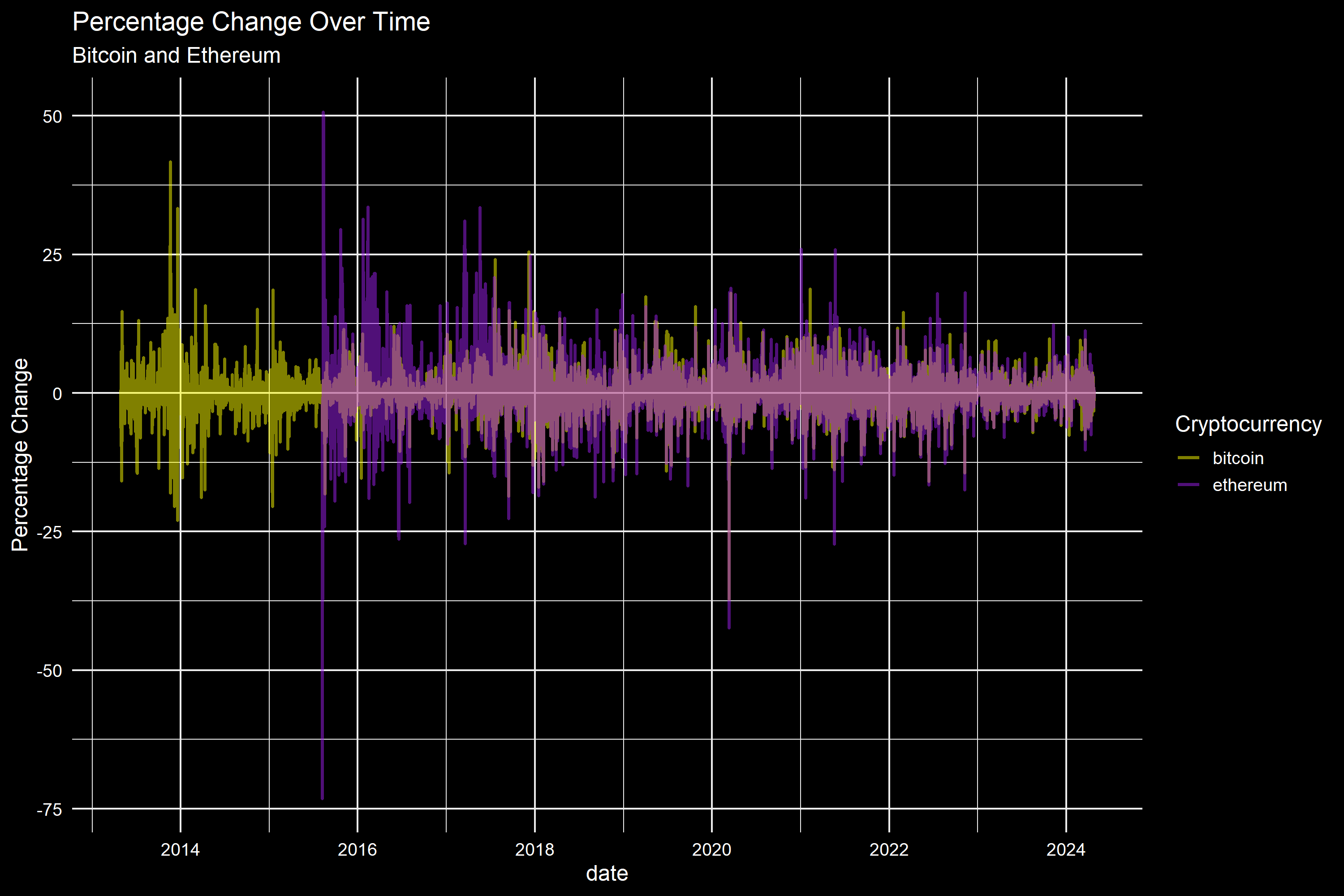
Moving beyond basic visualizations, we will employ advanced visualization techniques to gain deeper insights into the data. Furthermore, leveraging the feature-engineered dataset, we will prepare the groundwork for developing robust machine learning (ML) and deep learning (DL) models. These models will utilize the enriched dataset to predict cryptocurrency price movements and aid in making informed investment decisions.

By combining advanced visualization techniques with feature engineering and model preparation, this report aims to provide a comprehensive understanding of cryptocurrency markets. These insights will not only facilitate informed decision-making but also serve as a solid foundation for building predictive models that can further enhance investment strategies and opportunities in the cryptocurrency space.

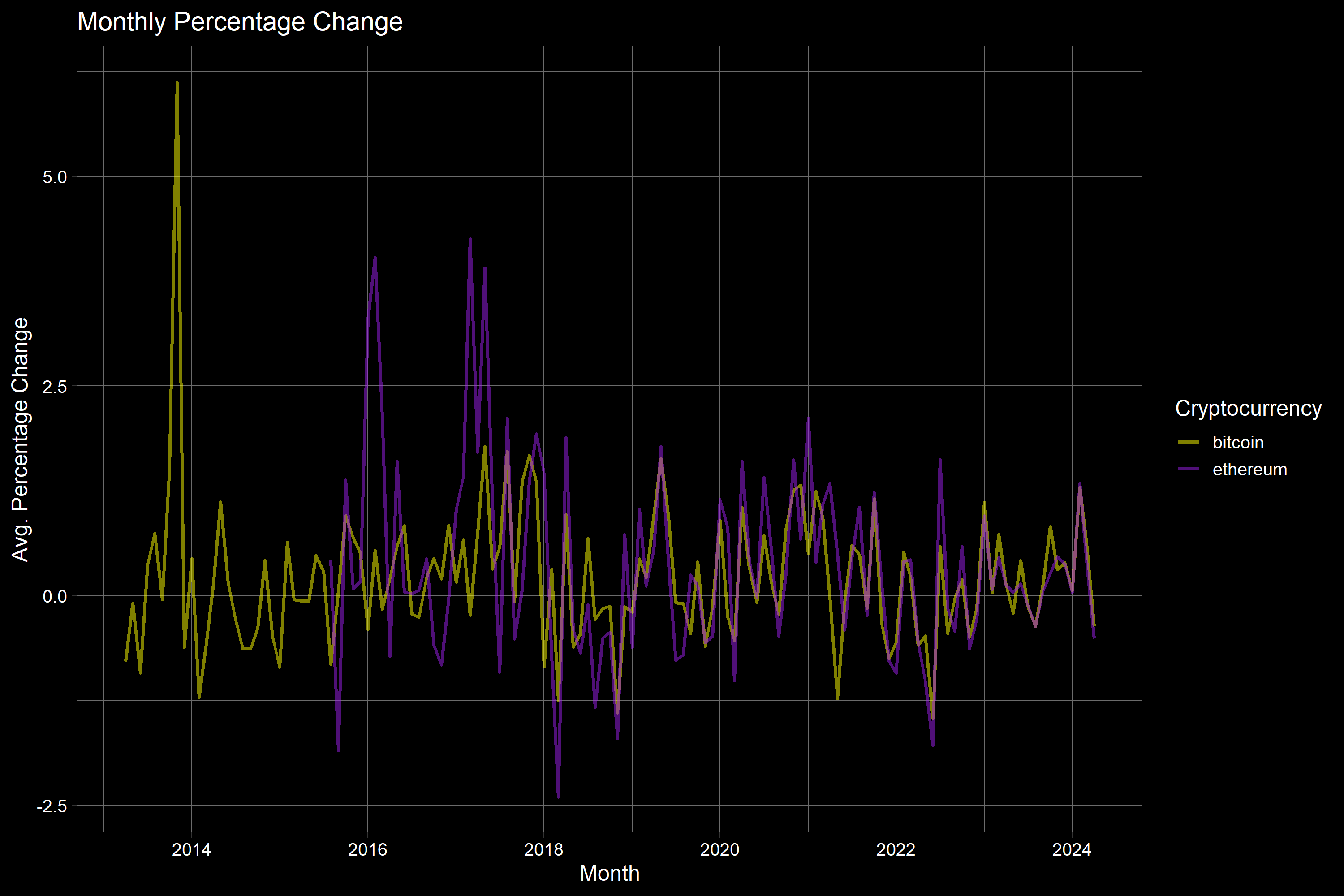
# Load Data and Extract Data Based on Price Action Analysis

### Data Prep - Adding Returns / log Returns

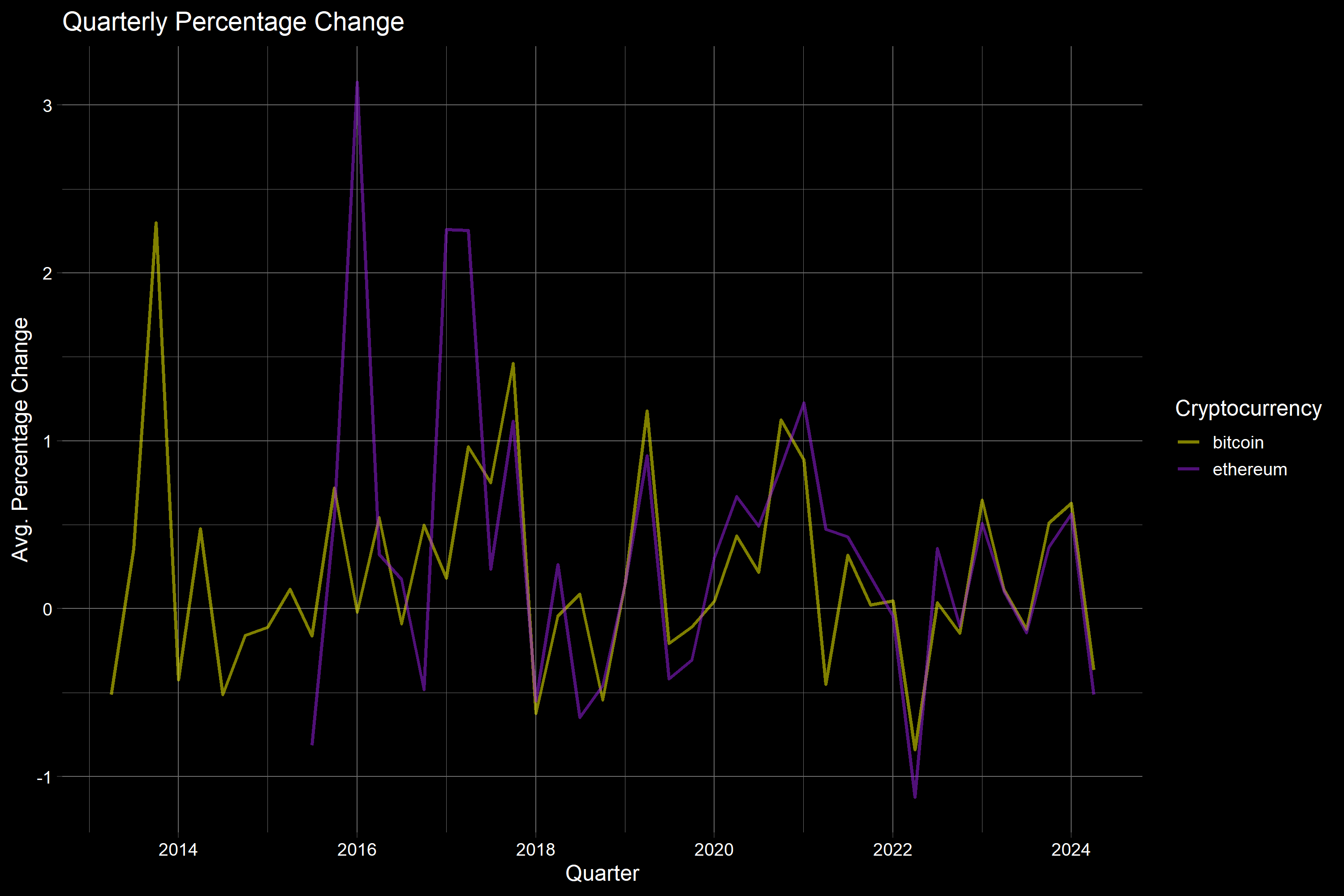
This section provides context for the code snippet, indicating that it focuses on preprocessing cryptocurrency data and engineering features to enhance the analysis process. The subsequent code demonstrates loading data, merging datasets, and creating new columns to capture percentage change and logarithmic return, essential metrics for understanding cryptocurrency price movements.



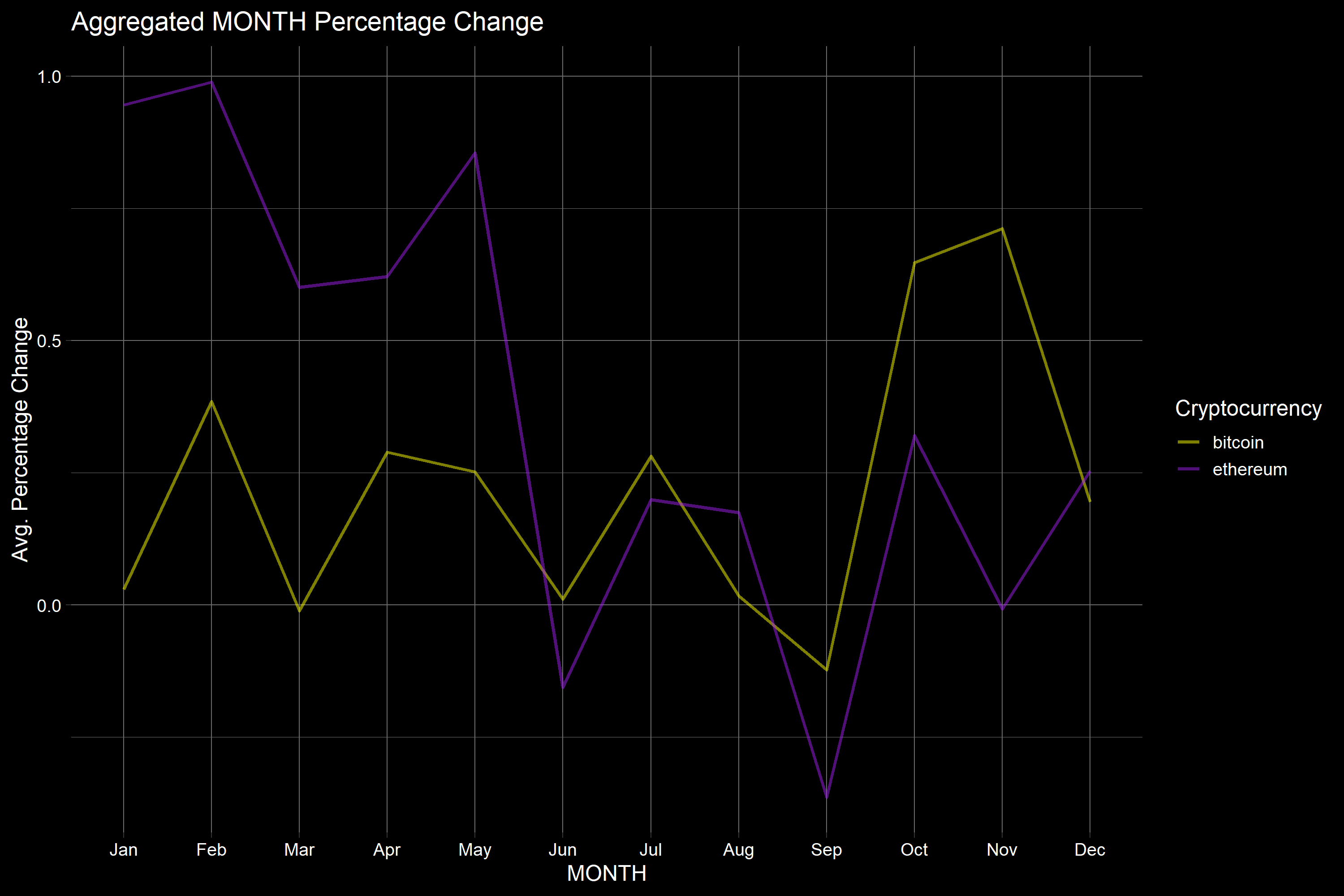
This plot visualizes the percentage change in market prices for Bitcoin and Ethereum from 2014 through 2024. The data suggests significant volatility in both cryptocurrencies, with Ethereum showing a tighter range of fluctuations in recent years compared to Bitcoin. This analysis is crucial for strategic investment decisions and risk assessment in the crypto market.



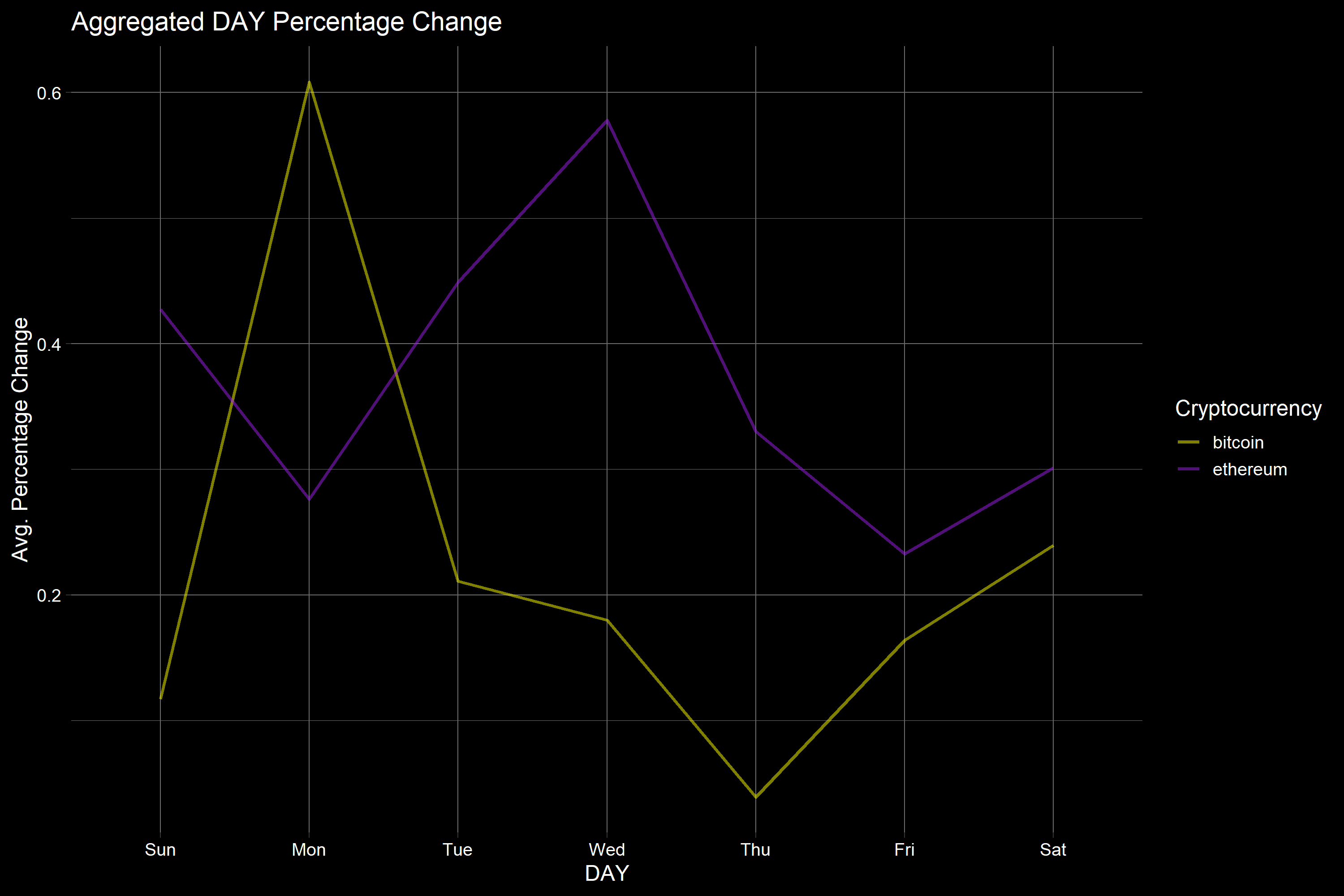
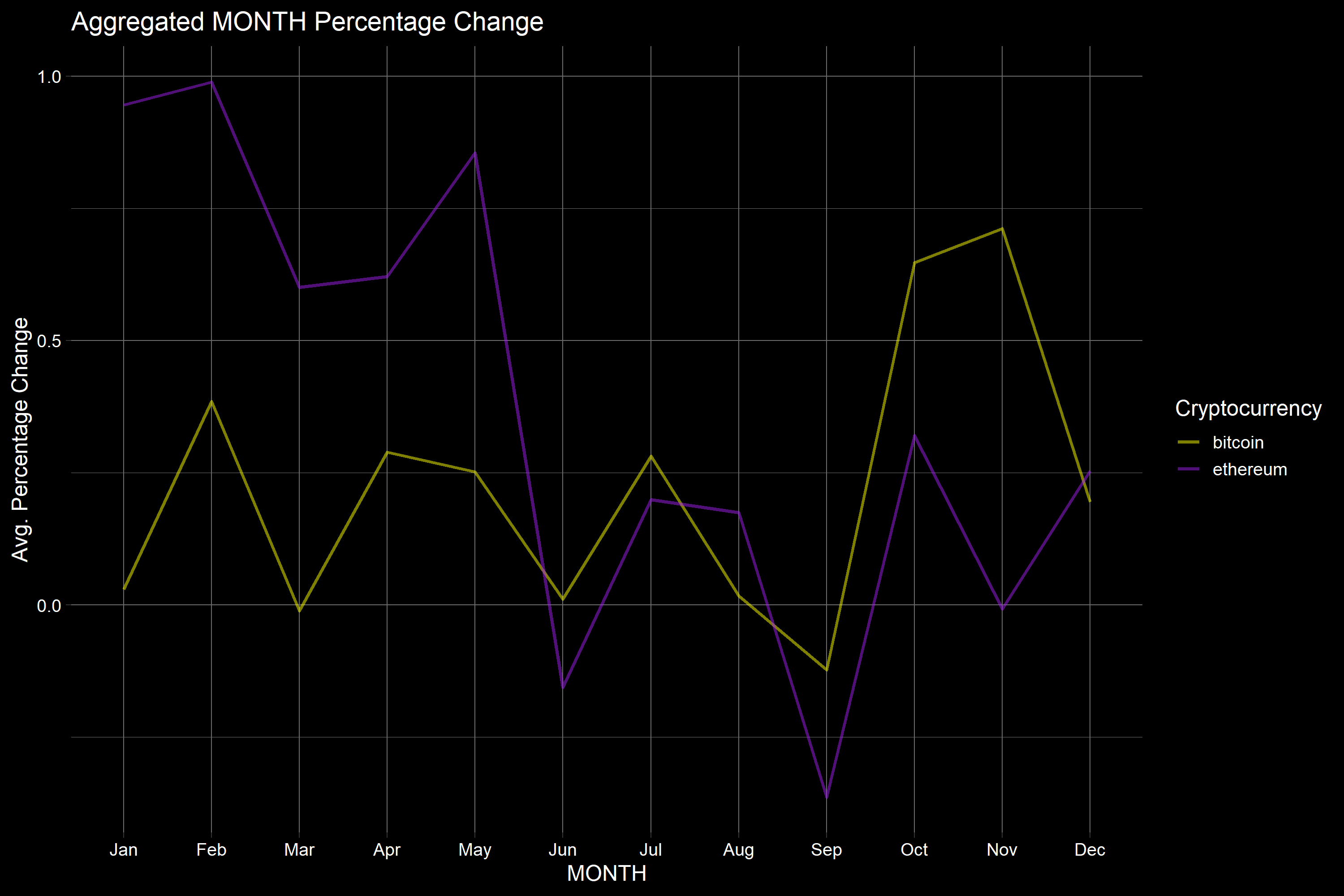
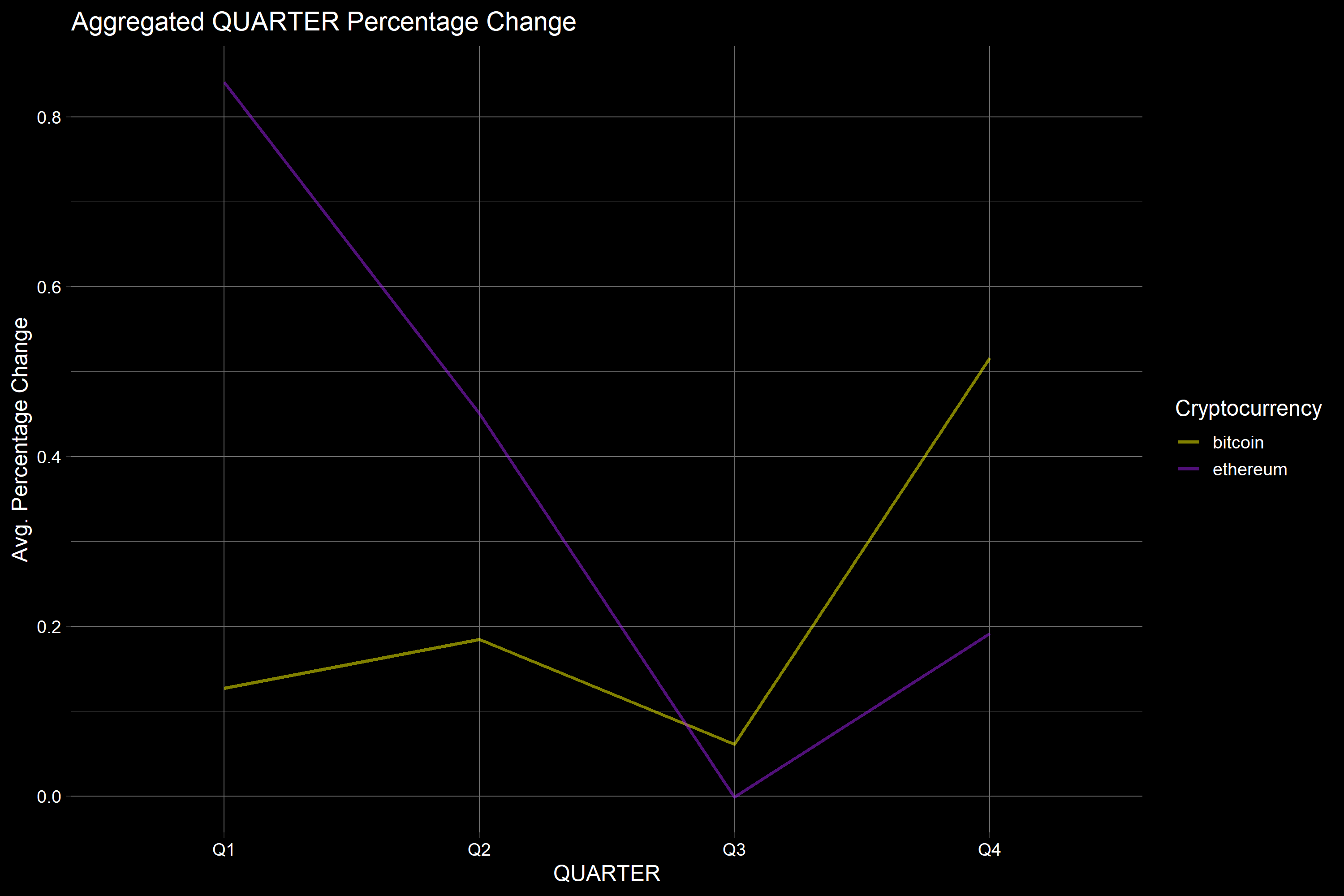
This plot displays monthly percentage changes for Bitcoin and Ethereum, highlighting their historical performance variability. Bitcoin shows early spikes, whereas Ethereum demonstrates increased volatility around 2016-2018. Investors should consider these trends for potential hedging strategies and diversifying investment portfolios in the volatile cryptocurrency market.



This chart illustrates the quarterly percentage changes for Bitcoin and Ethereum. The data indicates fluctuating but generally moderating volatility over time, crucial for strategizing long-term investment approaches and risk management in portfolio diversification within the evolving cryptocurrency market.



This chart tracks the aggregated monthly percentage changes in Bitcoin and Ethereum. The patterns highlight seasonality and monthly performance trends, essential for crafting targeted investment strategies and optimizing entry and exit points in the cryptocurrency market.



## Key Observations

**Midweek Volatility:** Significant spikes in volatility observed midweek for both cryptocurrencies. Ethereum shows a pronounced increase on Wednesdays, while Bitcoin peaks on Tuesdays. Suggests increased trading activity and market-moving events during midweek.

**Weekend Patterns:** Reduced volatility during weekends, especially on Sundays. Indicates lower trading activity as markets wind down and prepare for the upcoming week.

**Quarterly Trends:** Marked volatility within each quarter with a notable sharp recovery in Q3. Points to cyclical market sentiment and macroeconomic influences crucial for timing investment decisions.

**Monthly Analysis:** Detailed granularity reveals distinct patterns and seasonality. Both cryptocurrencies show similar trends with significant gains followed by corrections, especially noticeable around mid-year and year-end.

### Investment Considerations

**Risk Management:** By understanding observed patterns, investors can better anticipate periods of high volatility and adjust their portfolio exposures to mitigate risks.

**Strategic Entry and Exit Points:** Data suggests optimal investment windows, particularly investing at the start of a quarter post-downturn to capitalize on expected recoveries.

**Diversification**: Differences in volatility profiles between Bitcoin and Ethereum suggest opportunities for diversification to stabilize returns.

### Investment Strategy Considerations

**Day Trading Opportunities:** Heightened midweek volatility offers potential for day trading, especially from Tuesday to Thursday, providing opportunities to take early week positions and potentially exit before the weekend.

**Risk Management:** Daily market fluctuations necessitate tailored risk management strategies, including the use of stop-loss orders to manage potential price movements effectively.

**Weekend Strategy:** Lower trading activity on weekends allows for strategic time for portfolio assessment and preparation for the forthcoming week.

### Conclusion

This analysis highlights the importance of understanding market dynamics on daily, monthly, and quarterly bases to form a robust trading and investment strategy. By leveraging these insights, investors and traders can optimize their approaches to maximize returns and minimize risks in the volatile cryptocurrency market, enhancing overall portfolio performance.

#### Removing NA and Inf Values from Dataset

# PRE-N post halving dates

**Analysis Objectives and Methods**:We’ve focused on extracting and analyzing data from significant periods around each Bitcoin halving event, which occurs approximately every four years, reducing the reward for mining new blocks by half. This reduction potentially influences Bitcoin’s price due to decreased supply flow. By comparing the market dynamics pre and post halving, we seek to identify patterns or trends that could inform investment strategies and economic forecasts related to Bitcoin.\*

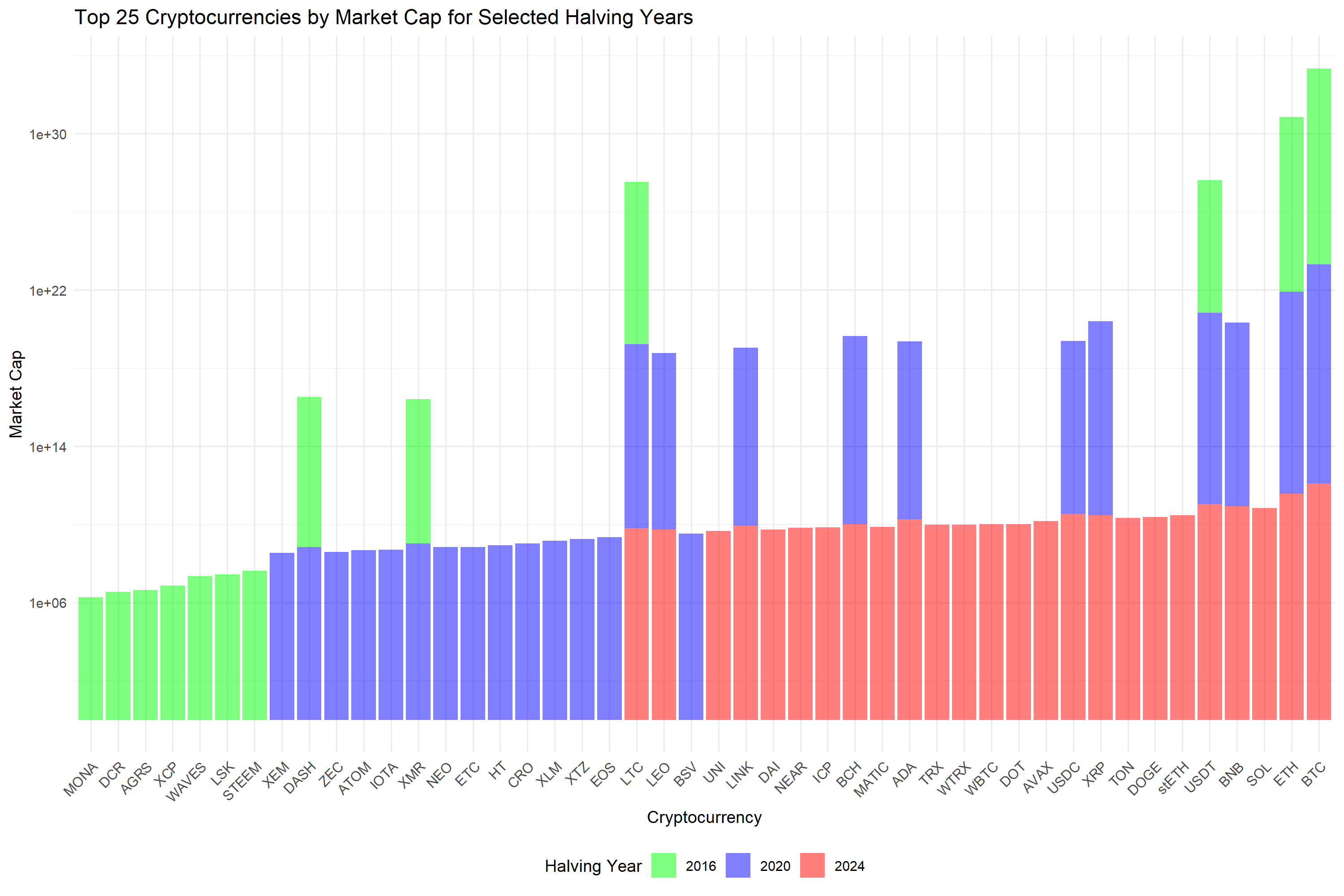
*Key Achievements:*

1. **Trend Identification:** We have identified trends in price fluctuations and market cap changes surrounding the halving dates, providing insights into the cyclical nature of Bitcoin’s market valuation.
2. **Comparative Analysis:** By juxtaposing data from different halving periods, we’ve been able to observe the evolution in market reactions over multiple cycles, discerning if reactions have become more pronounced or stabilized as the market matures.
3. **Predictive Insights:** Understanding past patterns allows us to forecast potential future price movements around subsequent halvings, offering valuable information for investors and market analysts.

**Importance of the Analysis:**

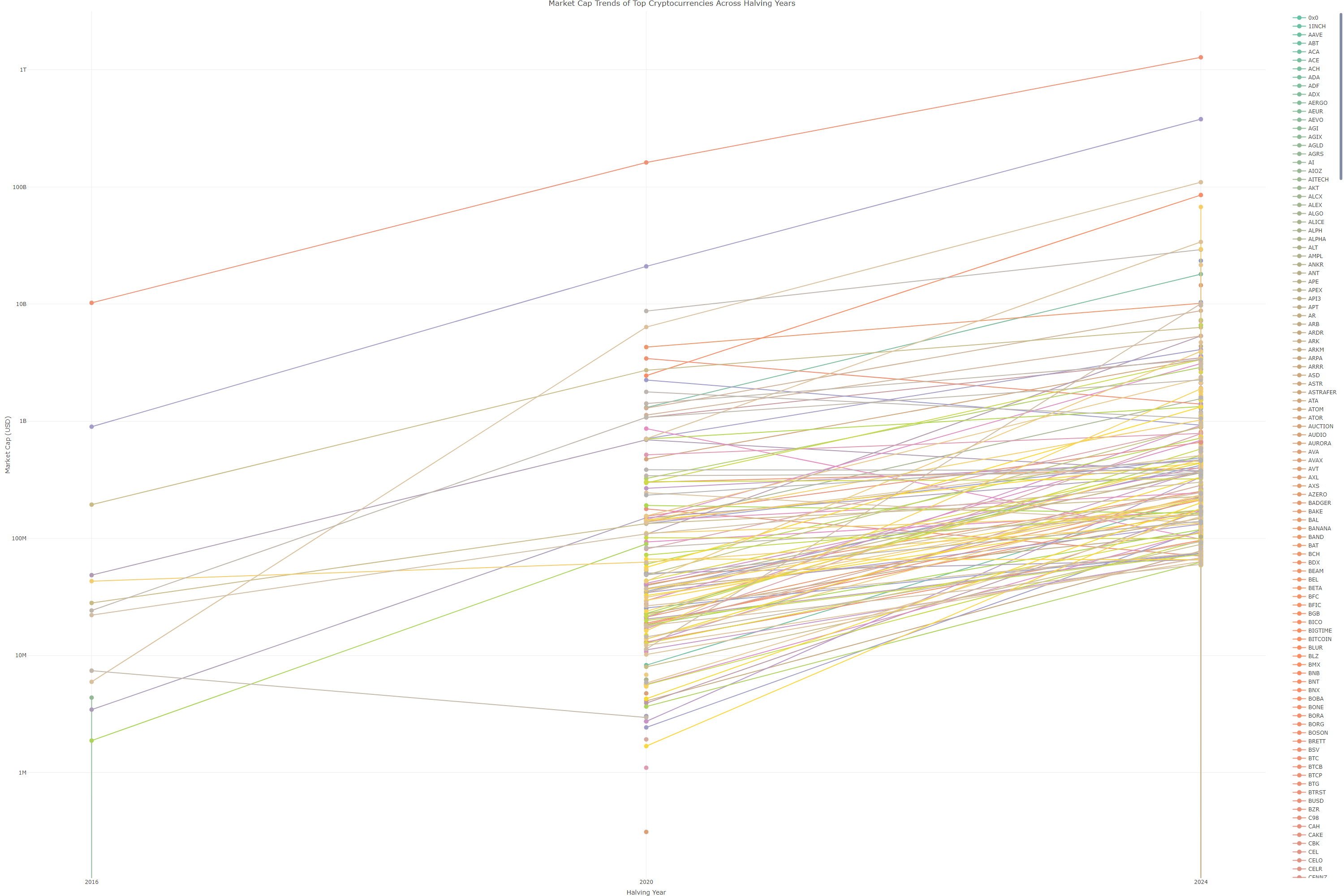
1. **Investor Guidance:** It helps investors understand potential price movements and manage risks associated with Bitcoin investments around halving events.
2. **Economic Implications:** It provides insights into how major supply changes in a dominant cryptocurrency can affect broader financial markets.
3. **Enhanced Understanding of Market Dynamics:** It contributes to the academic and practical understanding of cryptocurrency economics, particularly how supply constraints can impact asset prices.

### Halving Analysis



*The bar chart illustrates the market cap of the top 25 cryptocurrencies during Bitcoin halving years (2016, and 2024), showcasing distinct patterns:*

* **Growth Over Time**: The chart shows a clear upward trend in the market cap of cryptocurrencies over the selected halving years, indicating substantial growth in the crypto market.
* **Yearly Comparison**: Each halving year, depicted in different colors, highlights a significant increase in market cap, with 2024 showing the most considerable escalation, suggesting increased market maturity and investor interest.
* **Cryptocurrency Diversity**: While some cryptocurrencies consistently show growth, others have varying degrees of market cap across different years, pointing to shifting investor preferences and the dynamic nature of the cryptocurrency market.



### Impact of Halvings on Market Capitalization:

**Pre-Halving Surge:** Leading up to the halving events (marked by the vertical lines), there’s often a noticeable increase in market capitalization for many cryptocurrencies. This trend is driven by anticipation and speculation, as traders expect the reduced supply of new coins to push prices higher. **Post-Halving Reaction**: After the halving, the market reactions can be mixed. While some cryptocurrencies continue to appreciate due to the decreased rate of new coin introduction, others may stagnate or decline as the speculative fervor dies down.

#### Survival Through Crypto Winters:

**Decline of Many Coins:** As visible in your chart, while a few cryptocurrencies sustain growth post-halving, many others experience a sharp decline in market capitalization. This period, often referred to as the "crypto winter," can be harsh, and many smaller or less established cryptocurrencies might not recover.

**Sustainability of Major Coins:** The major cryptocurrencies often depicted by bolder or differently colored lines, tend to show resilience and a tendency to recover and grow in the long term. This resilience is likely due to stronger community support, better liquidity, and more widespread adoption.

#### Importance of Timely Exits:

**Timing Market Exits:** The data underscores the critical nature of timing when investing in cryptocurrencies, especially around such high-impact events as halvings. Investors who exit before the onset of a crypto winter can preserve gains and avoid significant losses.

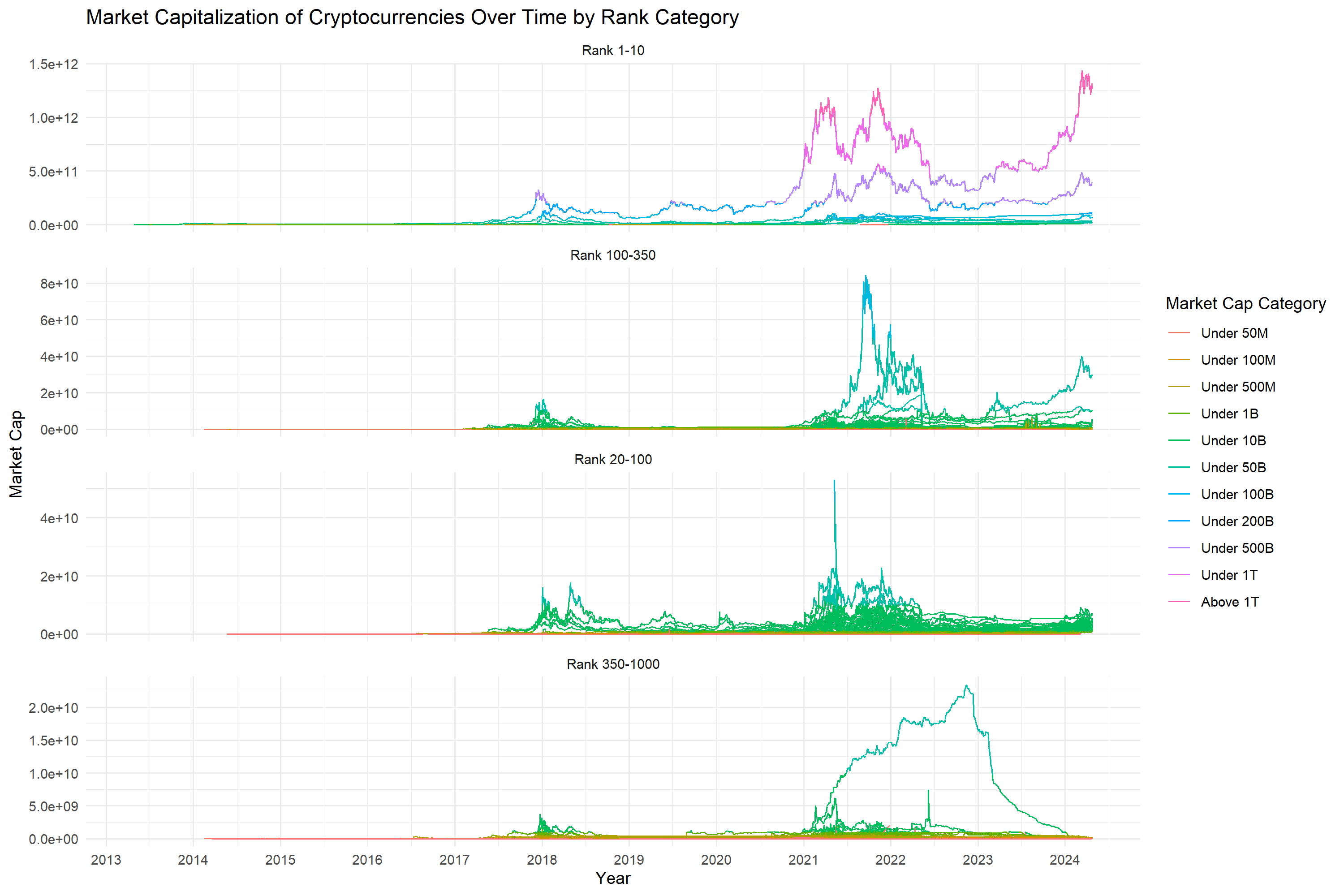
Strategic Investment Decisions: Understanding the cyclical nature of cryptocurrencies—highlighted by these halving events—can aid in making informed decisions about when to enter or exit positions. Monitoring not just the halving dates but also market sentiment, technological advancements, and regulatory developments is crucial.

***I****n summary, your chart powerfully illustrates the volatile nature of the cryptocurrency market, influenced heavily by periodic halving events. For investors and traders, the key takeaway would be the importance of strategic entries and exits, recognizing the signs of an impending downturn, and choosing cryptocurrencies with strong fundamentals and market support. This approach could mitigate risks associated with the inevitable fluctuations and downturns post-halving.*

# Market Capitalization Segmentation

*Market capitalization (market cap) is a crucial metric in the cryptocurrency industry, representing the total value of all coins in circulation. Its significance becomes even more pronounced around Bitcoin halving events, a period characterized by significant fluctuations in market cap due to the changes in block rewards. Analyzing these fluctuations can provide investors and analysts with critical insights into the potential gains and ideal exit points for high-risk investments.*

#### **Importance of Market Cap Analysis Around Halving Dates:**

1. **Indicator of Market Sentiment:** Market cap serves as a barometer for overall market sentiment and investor interest. A rising market cap around a halving suggests growing investor confidence and potential for price appreciation. Conversely, a declining market cap might indicate waning interest or negative sentiment, prompting cautious investment strategies.
2. **Impact of Supply Changes:** The halving reduces the rate at which new bitcoins are generated, directly impacting the supply side of the market equation. This reduced supply, if coupled with stable or increasing demand, can lead to a higher market cap and, by extension, an increase in the price per Bitcoin. Observing these dynamics helps predict price movements more accurately during these critical periods.
3. **Forecasting Future Trends:** By examining historical market cap trends around past halvings, investors can forecast potential reactions for upcoming halvings. This predictive insight is invaluable for formulating both short-term trading and long-term investment strategies.

**Survival and Risk in Crypto Winters:** The concept of “crypto winters” — prolonged periods of low prices and reduced investor activity — further emphasizes the importance of market cap analysis. During these phases, many cryptocurrencies, especially those with smaller market caps, can fail or cease to exist. This phenomenon can be attributed to:

* **Decreased Liquidity:** Lower market caps often result in reduced liquidity, making it difficult for investors to exit positions without impacting the price significantly.
* **Funding and Operational Challenges:** Projects with lower valuations may struggle to secure funding or sustain operations, leading to their eventual collapse.

#### **Strategic Implications for Investors:**

1. **Identifying Robust Investments:** Cryptocurrencies that maintain a stable or growing market cap through both halving events and crypto winters are typically more resilient and might offer safer investment opportunities.
2. **Timing Market Exits:** By recognizing patterns in market cap trends, investors can make informed decisions about the best times to exit high-risk positions, ideally before the onset of a crypto winter.

*In summary, market cap segmentation and its analysis around key events like Bitcoin halvings and during crypto winters are vital for understanding broader market dynamics. This analysis not only aids in identifying potential gains but also in determining strategic exit points to minimize exposure to high-risk investments, thereby optimizing the investment portfolio’s performance in the volatile cryptocurrency market.*

#### Market Capitalization Over Time:

All Top 10 Cryptocurrencies vs. Bitcoin: The first two plots compare the overall market cap trends of the top 10 cryptocurrencies and Bitcoin specifically. The visualization of Bitcoin shows more pronounced growth phases and stabilizations, indicating its dominant and less volatile role compared to other cryptocurrencies which have more fluctuating trajectories. This can suggest a lower risk in Bitcoin compared to other, less established cryptocurrencies.

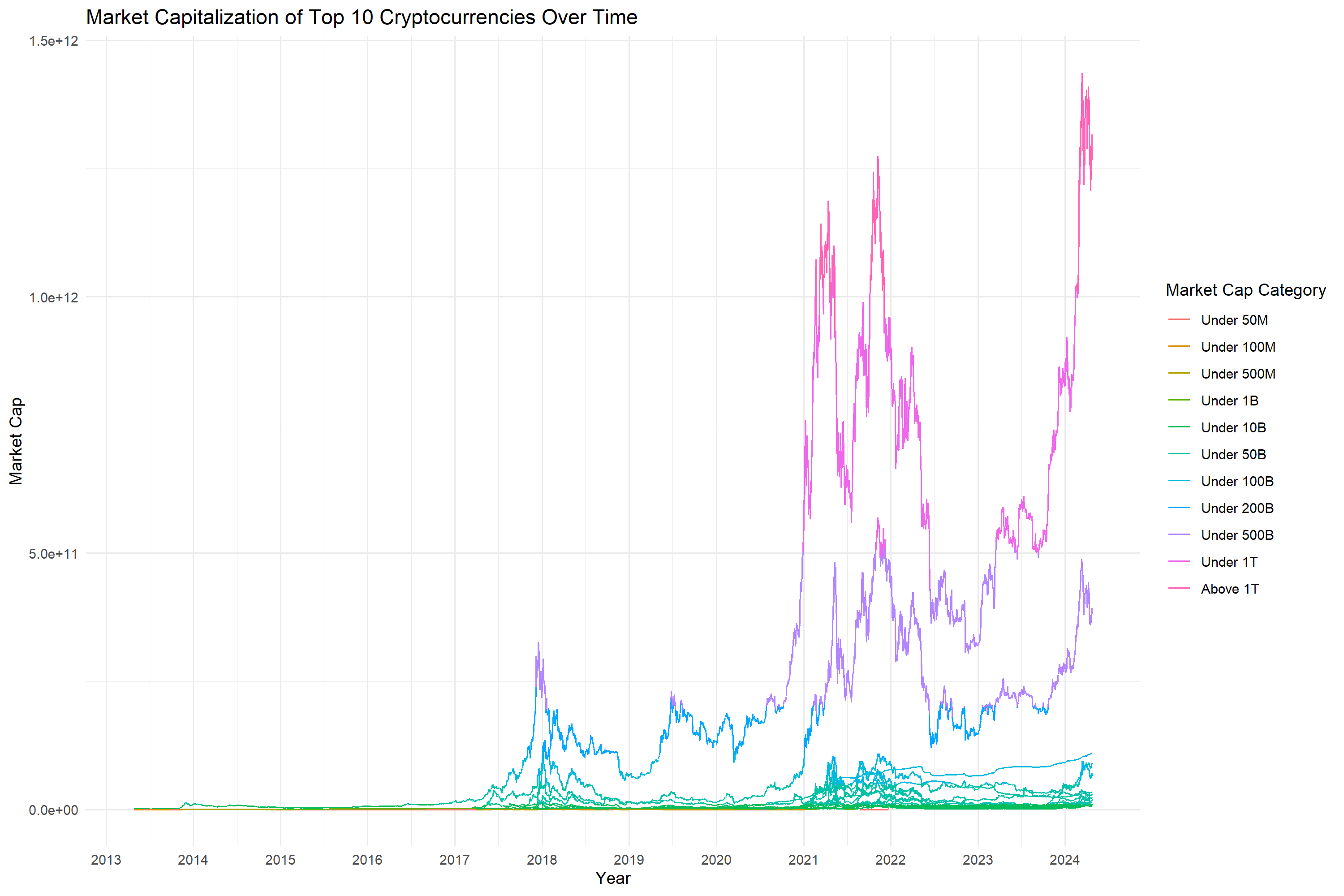
#### Market Capitalization by Rank Category:

**Diverse Ranks:** The third plot shows market cap across different rank categories, revealing how cryptocurrencies of different market standings (e.g., top 10, top 100) behave over time. Higher-ranked cryptocurrencies (Rank 1-10) show substantial and more stable growth compared to lower-ranked categories, which display more volatility and several peaks and troughs, indicating higher risk but also potentially higher returns if timed correctly.

#### Market Cap vs. Log Returns:

**Risk and Return Analysis:** The final plot uses log returns to illustrate the potential returns against market caps across various ranks. The visual clustering of higher returns in specific market cap segments (notably in the top ranks) highlights areas of higher profitability but also underlines the higher volatility and thus higher risk.

#### Key Points Emphasized:

****Risk Assessment**: The plots clearly categorize cryptocurrencies into segments based on their market cap, highlighting that lower market cap segments, while offering high returns, come with significantly higher risk. This is evident from the wider spreads in log returns, suggesting higher volatility.

**Return Profiles**: Higher market cap categories, especially those consistently above $1 billion, tend to offer more stable and reliable returns, appealing to risk-averse investors.

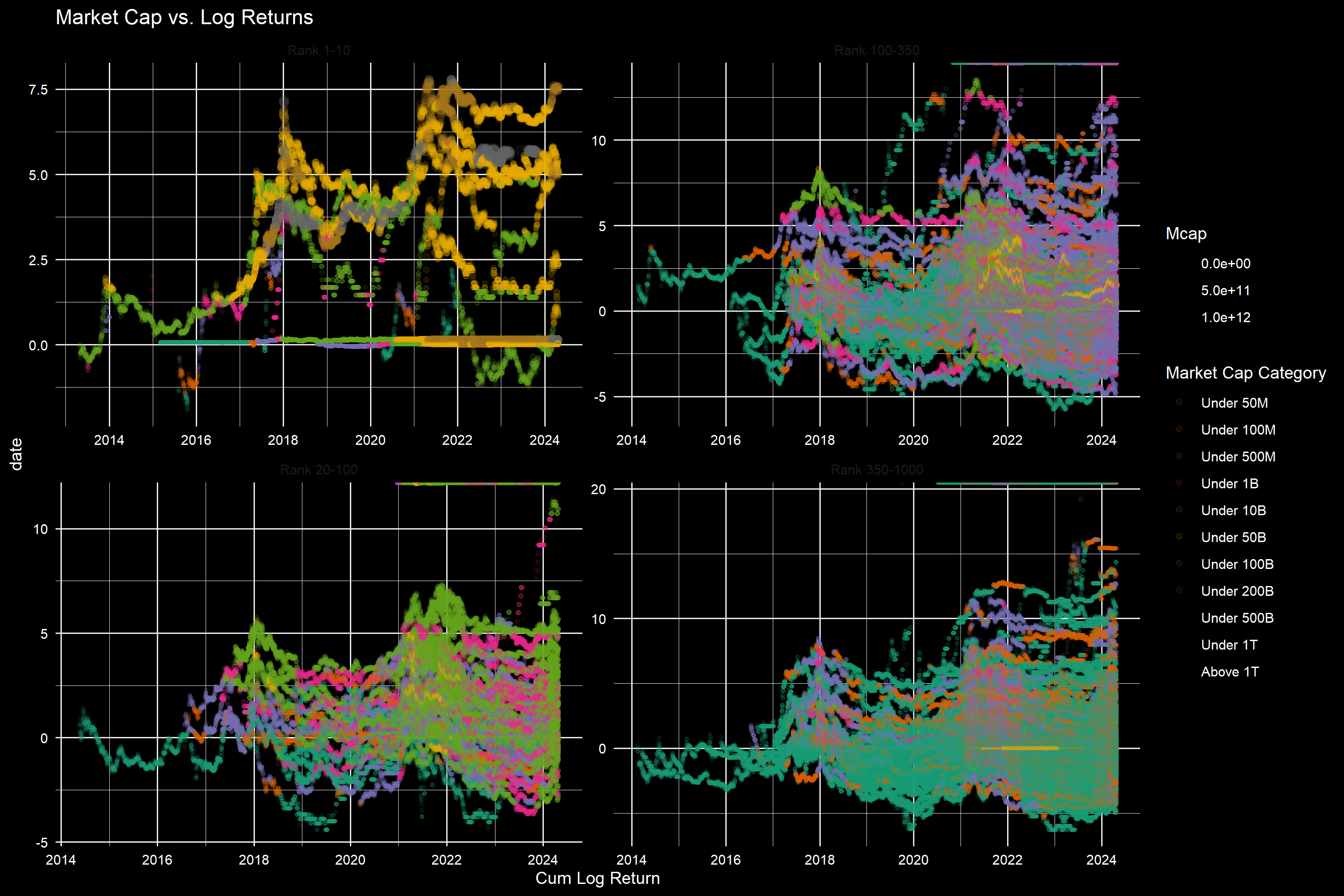
***Preliminary Analysis Disclaimer: These visualizations serve as an initial exploration into the dynamics of the cryptocurrency market. While they provide valuable insights, they are just the starting point. A more in-depth analysis would involve statistical testing, further segmentation, integration of external factors like market conditions and regulatory changes, and potentially predictive modeling to forecast future trends.***

#### Summary

*The insights derived from these visualizations are instrumental for investors to understand where the higher risks and returns lie within the cryptocurrency market. This analysis not only informs investment decisions but also sets the stage for deeper, more detailed explorations to validate these initial findings*

# Appendix

## Additional Fig.



A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

## Code

df\_pa <- df %>%  
 select(date,id,slug,open, close,market\_cap,rank) %>%  
 mutate(  
 pct\_change = (close - open) / open \* 100, # Percentage change  
 log\_return = log(close / open) # Logarithmic return  
 )

# Aggregate data by quarter  
df\_pa\_quarter <- df\_pa\_filtered %>%  
 mutate(Quarter = floor\_date(date, "quarter")) %>%  
 group\_by(slug, Quarter) %>%  
 summarize(pct\_change = mean(pct\_change, na.rm = TRUE), .groups = 'drop')

library(readxl)  
key <- read\_excel("~/GitHub/CryptoDataFetcher/CryptoDataFetcher2.0/DATA/Keys/MasterDataKey.xlsx")

# Merging the 'rank' column from 'key' into 'df' using 'id' as the key  
df <- df %>%  
 left\_join(key %>% select(id, rank), by = "id")

# Calculate the logarithmic ret, grouped by symbol if necessary for other analyses  
df\_selected <- df\_selected %>%  
 group\_by(symbol) %>%  
 mutate(log\_ret = log(close / open)) %>% # Calculate log ret as log(close/open)  
 ungroup() # optional: ungroup after the mutation if no further grouped operations  
  
# Assuming df\_selected already includes the 'log\_ret' calculations  
df\_selected <- df\_selected %>%  
 group\_by(symbol) %>%  
 arrange(date) %>% # Ensure the data is sorted by date  
 mutate(  
 cum\_log\_ret = cumsum(log\_ret) # Calculate cumulative logarithmic ret  
 ) %>%  
 ungroup() # Optional: ungroup if no further grouped operations are needed  
  
# Optionally, round cum\_ret to 2 decimal places if needed for consistency  
df\_selected$log\_ret <- round(df\_selected$log\_ret, 2)  
# Optionally, round cum\_ret to 2 decimal places if needed for consistency  
df\_selected$cum\_log\_ret <- round(df\_selected$cum\_log\_ret, 2)  
  
# Print the updated data frame to verify the results  
print(df\_selected)

## # A tibble: 821,090 × 11  
## slug symbol date open close market\_cap rank ret cum\_ret log\_ret  
## <chr> <chr> <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 bitco… BTC 2013-04-28 137. 128. 1.42e9 1 -0.06 -0.06 -0.07  
## 2 bitco… BTC 2013-04-29 134. 145. 1.60e9 1 0.08 0.02 0.07  
## 3 litec… LTC 2013-04-29 4.37 4.38 7.54e7 19 0 0 0   
## 4 bitco… BTC 2013-04-30 144 139 1.54e9 1 -0.03 -0.02 -0.04  
## 5 litec… LTC 2013-04-30 4.4 4.3 7.40e7 19 -0.02 -0.02 -0.02  
## 6 bitco… BTC 2013-05-01 139 117. 1.30e9 1 -0.16 -0.17 -0.17  
## 7 litec… LTC 2013-05-01 4.29 3.8 6.56e7 19 -0.11 -0.13 -0.12  
## 8 bitco… BTC 2013-05-02 116. 105. 1.17e9 1 -0.1 -0.26 -0.1   
## 9 litec… LTC 2013-05-02 3.78 3.37 5.83e7 19 -0.11 -0.22 -0.11  
## 10 bitco… BTC 2013-05-03 106. 97.8 1.09e9 1 -0.08 -0.32 -0.08  
## # ℹ 821,080 more rows  
## # ℹ 1 more variable: cum\_log\_ret <dbl>

#write.csv(df\_selected, "Crypto\_Hist\_LogRt.csv", row.names = FALSE)

#for later when we have all the other data pct change as well as log date structureed like MoM or DoD

# Update the last entry in the 'Halving' column  
df\_periods[nrow(df\_periods), 'Halving'] <- '2024 Halving'  
  
# Check the updated data frame  
print(df\_periods)

## # A tibble: 16 × 9  
## Halving Period Start\_Date End\_Date days weeks months quarters years  
## <chr> <chr> <date> <date> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 2012 Halving Pre-Bull 2012-03-01 2012-10-31 244 35 8 3 1  
## 2 2014 Halving Halving… 2012-11-01 2012-11-01 0 0 0 0 0  
## 3 2014 Halving Bull Run 2012-11-02 2014-07-08 613 88 20 7 2  
## 4 2014 Halving Crypto … 2014-07-09 2015-09-01 419 60 14 5 1  
## 5 2016 Halving Pre-Bull 2015-09-02 2016-07-08 310 44 10 3 1  
## 6 2016 Halving Halving… 2016-07-09 2016-07-09 0 0 0 0 0  
## 7 2016 Halving Bull Run 2016-07-10 2017-12-31 539 77 18 6 1  
## 8 2016 Halving Crypto … 2017-01-01 2019-01-01 730 104 24 8 2  
## 9 2020 Halving Pre-Bull 2019-01-02 2020-05-11 495 71 16 5 1  
## 10 2020 Halving Halving… 2020-05-12 2020-05-12 0 0 0 0 0  
## 11 2020 Halving Bull Run 2020-05-13 2021-03-01 292 42 10 3 1  
## 12 2020 Halving Crypto … 2021-03-02 2021-07-01 121 17 4 1 0  
## 13 2020 Halving Bull Run 2021-07-02 2021-11-01 122 17 4 1 0  
## 14 2020 Halving Crypto … 2021-11-02 2022-12-01 394 56 13 4 1  
## 15 2024 Halving Pre-Bull 2022-12-02 2024-04-19 504 72 17 6 1  
## 16 2024 Halving Halving… 2024-04-20 2024-04-20 0 0 0 0 0

HalvingDatesKey<-df\_periods  
write.csv(HalvingDatesKey, "HalvingDataKey.csv", row.names = FALSE)

# Assuming the date format needs to be converted in both data frames  
MasterCrypto$date <- as.Date(MasterCrypto$date, format="%Y-%m-%d")  
  
# Assuming df\_periods already loaded from earlier steps, format its dates if not done already  
df\_periods$Start\_Date <- as.Date(df\_periods$Start\_Date, format="%Y-%m-%d")  
df\_periods$End\_Date <- as.Date(df\_periods$End\_Date, format="%Y-%m-%d")

if (!require("fuzzyjoin")) install.packages("fuzzyjoin")  
library(fuzzyjoin)  
  
# Perform a fuzzy join based on date ranges  
df\_merged <- fuzzy\_left\_join(MasterCrypto, df\_periods,  
 by = c("date" = "Start\_Date", "date" = "End\_Date"),  
 match\_fun = list(`>=`, `<=`))  
  
# After the join, select all columns from MasterCrypto and the Code column from df\_periods  
df\_merged <- df\_merged %>%  
 select(all\_of(names(MasterCrypto)), Code = Code)  
  
Crypto\_Hist\_Ret\_Log\_Rank\_Code <- df\_merged  
  
#write.csv(Crypto\_Hist\_Ret\_Log\_Rank\_Code, "Crypto\_Hist\_Ret\_Log\_Rank\_Code.csv", row.names = FALSE)

halving\_2016 <- df\_merged %>% filter(date == as.Date("2016-07-09"))  
halving\_2020 <- df\_merged %>% filter(date == as.Date("2020-05-12"))  
halving\_2024 <- df\_merged %>% filter(date == as.Date("2024-04-20"))

library(dplyr)  
library(ggplot2)  
library(lubridate)  
  
# Combining and creating a unified data frame with 2016, 2020, and 2024 data  
combined\_halving\_data <- bind\_rows(  
 list(halving\_2016, halving\_2020, halving\_2024),  
 .id = "Halving\_Year"  
) %>%  
 mutate(Halving\_Year = case\_when(  
 Halving\_Year == "1" ~ "2016",  
 Halving\_Year == "2" ~ "2020",  
 Halving\_Year == "3" ~ "2024"  
 ))  
  
# Get top 25 for each year using group\_by and slice\_max  
top\_halvings <- combined\_halving\_data %>%  
 group\_by(Halving\_Year) %>%  
 slice\_max(order\_by = market\_cap, n = 25) %>%  
 ungroup() # Ungroup to remove grouping before plotting  
  
# Plotting the data with legend and corrected title and axis rotations  
ggplot(top\_halvings, aes(x = reorder(symbol, market\_cap), y = market\_cap, fill = Halving\_Year)) +  
 geom\_col(show.legend = TRUE, alpha = 0.5) +  
 labs(  
 title = "Top 25 Cryptocurrencies by Market Cap for Selected Halving Years",   
 x = "Cryptocurrency",   
 y = "Market Cap",  
 fill = "Halving Year" # Legend title  
 ) +  
 scale\_y\_log10() + # Apply log10 scale to the y-axis  
 scale\_fill\_manual(values = c("2016" = "green", "2020" = "blue", "2024" = "red")) +  
 theme\_minimal() +  
 theme(  
 legend.position = "bottom",  
 axis.text.x = element\_text(angle = 45, vjust = 1, hjust = 1) # Tilt x-axis labels by 45 degrees  
 )

library(dplyr)  
  
# Assuming 'volume' exists and is a measure of trading volume or another metric  
# Ensure 'date' is in a date format, convert if necessary  
Mcapdf$date <- as.Date(Mcapdf$date)  
  
# Filter the data for the specified rank ranges and prepare it for plotting  
top\_coins\_data <- Mcapdf %>%  
 filter(rank >= 1 & rank <= 1000) %>%  
 mutate(rank\_category = case\_when(  
 rank >= 1 & rank <= 10 ~ "Rank 1-10",  
 rank >= 20 & rank <= 100 ~ "Rank 20-100",  
 rank >= 100 & rank <= 350 ~ "Rank 100-350",  
 rank >= 350 & rank <= 1000 ~ "Rank 350-1000"  
 )) %>%  
 filter(!is.na(rank\_category)) library(ggplot2)  
library(dplyr)  
  
# Assuming 'volume' exists and is a measure of trading volume or another metric  
# Ensure 'date' is in a date format, convert if necessary  
Mcapdf$date <- as.Date(Mcapdf$date)  
  
# Filter the data for the specified rank ranges and prepare it for plotting  
top\_coins\_data <- Mcapdf %>%  
 filter(rank >= 1 & rank <= 1000) %>%  
 mutate(rank\_category = case\_when(  
 rank >= 1 & rank <= 10 ~ "Rank 1-10",  
 rank >= 20 & rank <= 100 ~ "Rank 20-100",  
 rank >= 100 & rank <= 350 ~ "Rank 100-350",  
 rank >= 350 & rank <= 1000 ~ "Rank 350-1000"  
 )) %>%  
 filter(!is.na(rank\_category))

# Create a scatter plot  
scatter\_plot <- ggplot(filtered\_data, aes(x = cum\_log\_ret, y = market\_cap, color = mcapcat, size = market\_cap)) +  
 geom\_point(alpha = 0.6) + # Adjust transparency for better visibility  
 scale\_x\_continuous(limits = c(-5, 10)) + # Extend the x-axis range to 20  
 scale\_y\_log10(labels = label\_scientific(accuracy = 1)) + # Log scale for market cap with custom labels  
 scale\_color\_brewer(palette = "Spectral") + # A color palette that provides good contrast  
 labs(  
 title = "Market Cap vs. Cumulative Log Returns for Cryptocurrencies (HD2024)",  
 x = "Cumulative Log Return",  
 y = "Market Capitalization",  
 color = "Market Cap Category",  
 size = "Market Cap"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.background = element\_rect(fill = "black", color = "black"), # Set the plot background to black  
 panel.background = element\_rect(fill = "black", color = "black"), # Set the panel background to black  
 text = element\_text(color = "white"), # Change text color to white for visibility  
 axis.text = element\_text(color = "white"), # Change axis text color to white  
 axis.title = element\_text(color = "white"), # Change axis title color to white  
 legend.title = element\_text(color = "white"), # Change legend title color to white  
 legend.text = element\_text(color = "white") # Change legend text color to white  
 )+  
 theme(  
 axis.text.y = element\_text(angle = 0) # Ensure y-axis labels are not angled  
 )  
  
# Print the scatter plot## Attaching package: 'scales'

## The following object is masked from 'package:readr':  
##   
## col\_factor

# Create a scatter plot  
scatter\_plot <- ggplot(filtered\_data, aes(x = cum\_log\_ret, y = market\_cap, color = mcapcat, size = market\_cap)) +  
 geom\_point(alpha = 0.6) + # Adjust transparency for better visibility  
 scale\_x\_continuous(limits = c(-5, 10)) + # Extend the x-axis range to 20  
 scale\_y\_log10(labels = label\_scientific(accuracy = 1)) + # Log scale for market cap with custom labels  
 scale\_color\_brewer(palette = "Dark2") + # A color palette that provides good contrast  
 labs(  
 title = "Market Cap vs. Cumulative Log Returns for Cryptocurrencies (HD2024)",  
 x = "Cumulative Log Return",  
 y = "Market Capitalization",  
 color = "Market Cap Category",  
 size = "Market Cap"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.background = element\_rect(fill = "black", color = "black"), # Set the plot background to black  
 panel.background = element\_rect(fill = "black", color = "black"), # Set the panel background to black  
 text = element\_text(color = "white"), # Change text color to white for visibility  
 axis.text = element\_text(color = "white"), # Change axis text color to white  
 axis.title = element\_text(color = "white"), # Change axis title color to white  
 legend.title = element\_text(color = "white"), # Change legend title color to white  
 legend.text = element\_text(color = "white") # Change legend text color to white  
 ) +  
 theme(  
 axis.text.y = element\_text(angle = 0) # Ensure y-axis labels are not angled  
 )  
  
##

# Print the scatter plot  
print(scatter\_plot)

## Warning in scale\_y\_log10(labels = label\_scientific(accuracy = 1)): log-10  
## transformation introduced infinite values.

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Dark2 is 8  
## Returning the palette you asked for with that many colors

## Warning: Removed 13 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

A screen shot of a computer screen

Description automatically generatedA screen shot of a computer screen

Description automatically generated