MATH1324 Assignment 1

Statistical analysis of S&P500 and Bitcoin over the last 6 years.

Student Details

Jason Wang s4134626

Problem Statement

This report calculates the descriptive statistics of both S&P500 and Bitcoin over the last 6 years. In addition, the correlation between each other are determined, as well as their normal distributions. These statistics are done in effort to determine insights of volatility,relationships and correlations between S&P500 and Bitcoin.

Load Packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(rmarkdown)
```

Introduction

S&P500 is a measurement of 500 large publicly traded companies in the stock market index. It plays a key indicator in stock market health, as it comprises of about 80% of the total market value of U.S companies. Bitcoin was the first ever cryptocurrency, and is the most valuable to date. It is designed to work as a decentralised payment system. It differs from stock market because it is highly volatile due to its independence of banks and government. In this report, both S&P500 and Bitcoin are directly compared in search to see the relationships and correlations between the 2 financial assets.

Data

```
SP500 <- read.csv ("C:/Users/wanga/OneDrive/Documents/Work/RMIT/Y1 Sem 1/Applied Analytics/Data/SP500.csv")
Bitcoin <- read.csv ("C:/Users/wanga/OneDrive/Documents/Work/RMIT/Y1 Sem 1/Applied Analytics/Data/BTC.csv")

SP500$Date <- as.Date(SP500$Date, format = "%d/%m/%Y")
Bitcoin$Date <- as.Date(Bitcoin$Date, format = "%d/%m/%Y")

class(SP500$Price)
```

```
## [1] "character"
```

```
class(Bitcoin$Adj.Close)
```

```
## [1] "numeric"
```

```
SP500$Price <- gsub(",", "", SP500$Price)

SP500$Price <- as.numeric(SP500$Price)

class(SP500$Price)
```

```
## [1] "numeric"
```

```
head(n = 3,SP500)
```

	Date <date>	Price <dbl>
1	2024-08-01	5446.68
2	2024-07-31	5522.30
3	2024-07-30	5436.44

3 rows

```
head(n = 3,Bitcoin)
```

	Date <date>	Adj.Close <dbl>
1	2018-08-01	7624.91
2	2018-08-02	7567.15
3	2018-08-03	7434.39

3 rows

Task 1

```
SPmean <- mean(SP500$Price)
SPmedian <- median(SP500$Price)

getSPmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}

v <- SP500$Price
SPmode <- getSPmode(v)

SPrange <- max(SP500$Price) - min(SP500$Price)

SPvar <- SP500$Price %>% var()
SPsd <- SP500$Price %>% sd()


Bitcoinmean <- mean(Bitcoin$Adj.Close)
Bitcoinmedian <- median(Bitcoin$Adj.Close)
getBitcoinmode <- function(x) {
  uniqx <- unique(x)
  uniqx [which.max(tabulate(match(x,uniqx)))]
}

x <- Bitcoin$Adj.Close
Bitcoinmode <- getBitcoinmode(x)

Bitcoinrange <- max(Bitcoin$Adj.Close) - min(Bitcoin$Adj.Close)

Bitcoinvar <- Bitcoin$Adj.Close %>% var()
Bitcoinsd <- Bitcoin$Adj.Close %>% sd()

df <- data.frame(
  Statstic = c("Mean", "Median", "Mode", "Range", "Variance", "Standard Deviation"),
  SPstats = c(SPmean, SPmedian, SPmode, SPrange, SPvar, SPsd),
  Bitcoinstats = c(Bitcoinmean, Bitcoinmedian, Bitcoinmode, Bitcoinrange, Bitcoinvar, Bitcoin
sd)
)

print(df)
```

```
##             Statstic     SPstats Bitcoinstats
## 1               Mean   3826.5555     26747.53
## 2             Median   3930.6900     23031.09
## 3               Mode   2914.0000      7624.91
## 4              Range   3429.8000     69846.74
## 5           Variance 603470.5271 367218856.43
## 6 Standard Deviation    776.8337     19162.96
```

Task 2

```
SP500 %>%
  plot(Price ~ Date, data = .,
       ylab = "Price",
       xlab = "Year",
       col = "blue",
       main = "S&P500 price over the last 6 years.",
       type = "l"
       )
mtext("Figure 1.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

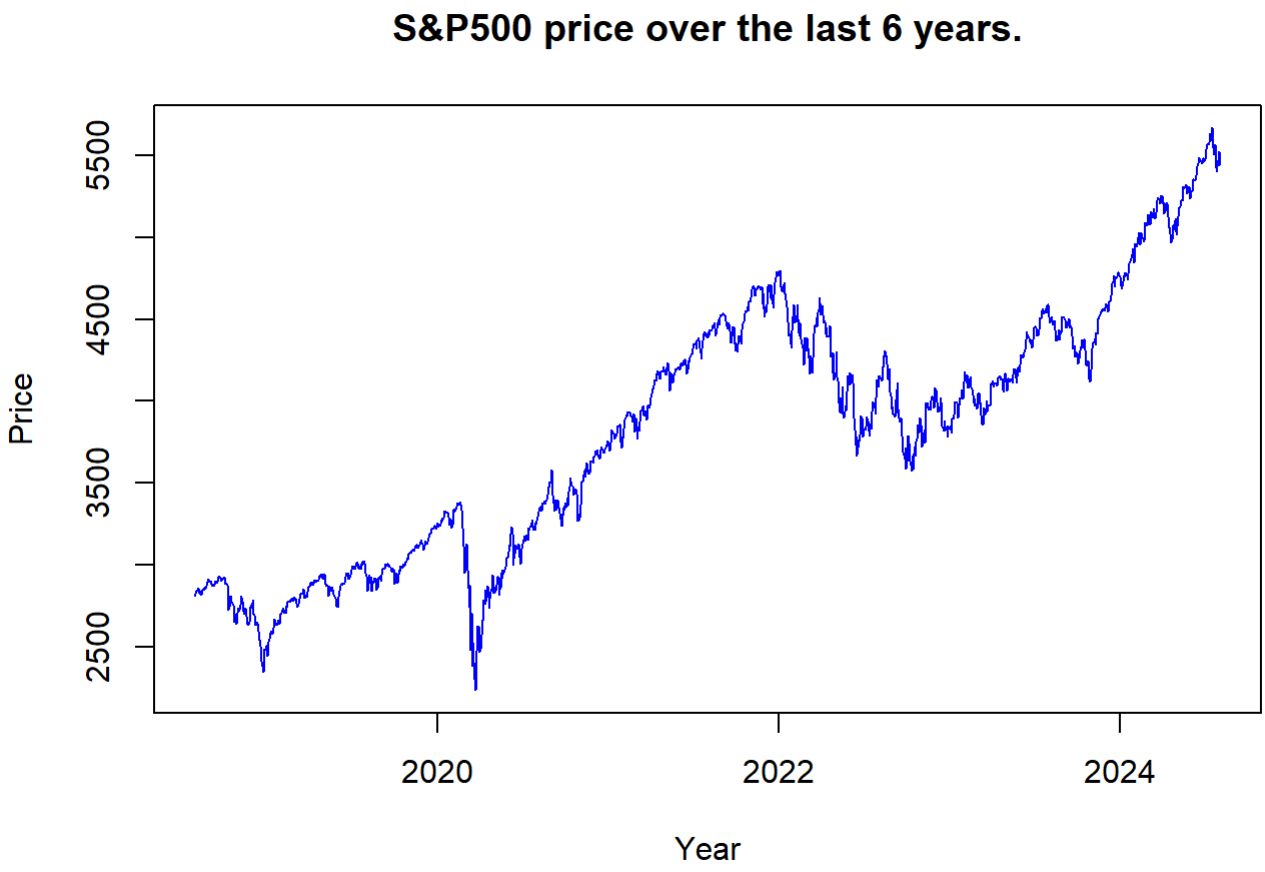

Figure 1.

```
Bitcoin %>%
  plot(Adj.Close ~ Date, data = .,
       ylab = "Adjusted Close Price",
       xlab = "Year",
       col = "orange",
       main = "Bitcoin adjusted close over the last 6 years.",
       type = "l")
mtext("Figure 2.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

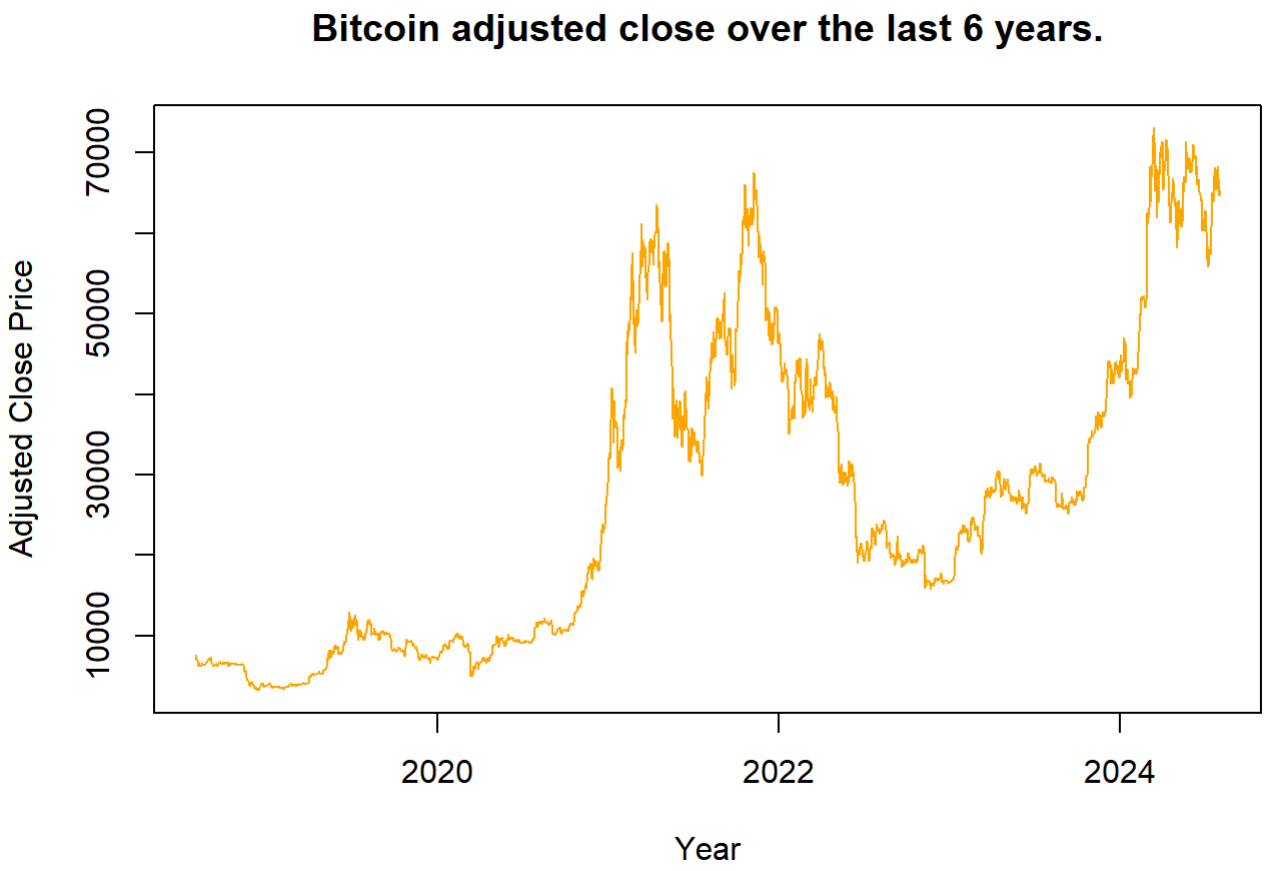

Figure 2.

```
merged_df <- merge(SP500, Bitcoin, by = "Date")

merged_df$Period <- cut(merged_df$Date, breaks = "6 months")

head(merged_df)
```

	Date <date>	Price <dbl>	Adj.Close <dbl>	Period <fct>
1	2018-08-01	2813.4	7624.91	2018-08-01
2	2018-08-02	2827.2	7567.15	2018-08-01
3	2018-08-03	2840.3	7434.39	2018-08-01
4	2018-08-06	2850.4	6951.80	2018-08-01
5	2018-08-07	2858.4	6753.12	2018-08-01
6	2018-08-08	2857.7	6305.80	2018-08-01

6 rows

```
correlations <- merged_df %>% group_by(Period) %>%  summarise(Correlation = cor(Price, Adj.Cl
ose))

correlations$Period <- as.Date(as.character(correlations$Period))

correlations %>%
  plot(Correlation ~ Period, data = .,
       ylab = "Correlation",
       xlab = "Period",
       col = "green",
       main = "Correlation between S&P500 and Bitcoin over the last 6 years.",
       type = "l")
mtext("Figure 3.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

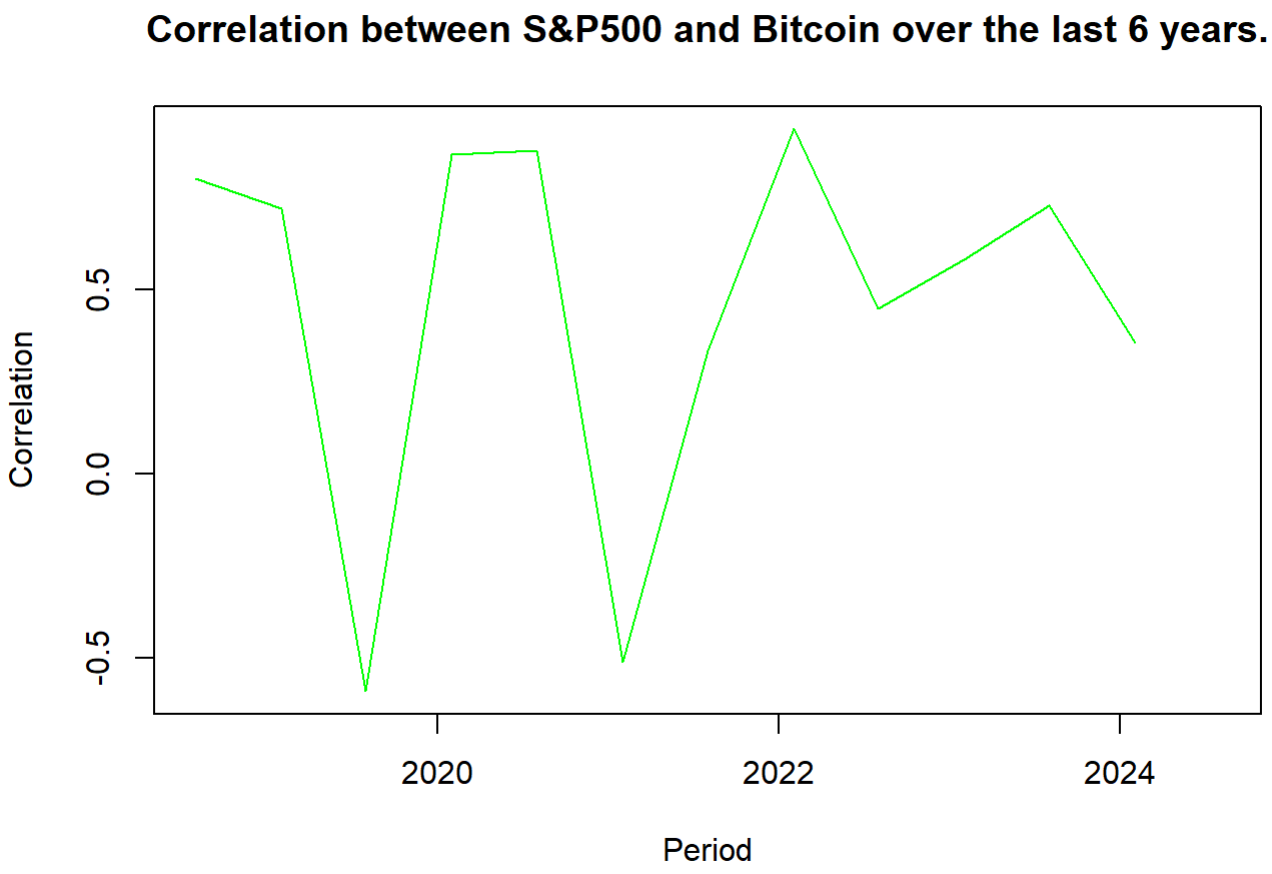

Figure 3.

Task 3

```
correlation <- cor(merged_df$Price, merged_df$Adj.Close)

print(correlation)
```

```
## [1] 0.8976749
```

```
plot(merged_df$Price, merged_df$Adj.Close,
     main = "Correlation of S&P500 prices vs. Bitcoin adjusted close prices",
     xlab = "S&P500 Prices",
     ylab = "Bitcoin Adjusted Close Prices",
     col = "black")

model <- lm(merged_df$Adj.Close ~ merged_df$Price)
abline(model, col = "red", lwd = 2)

mtext("Figure 4.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

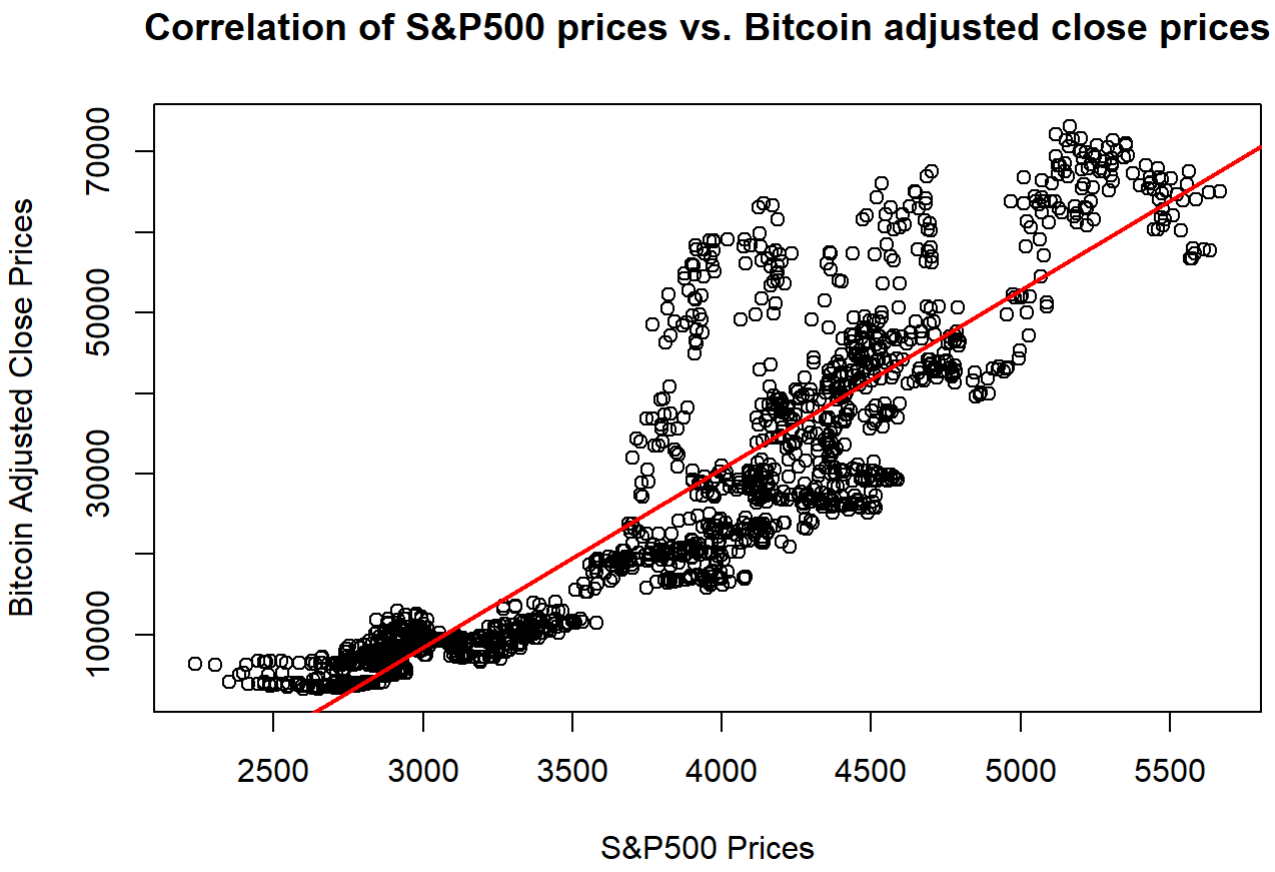

Figure 4.

Task 4

```
merged_df$Price %>% hist(col = "blue", main = "Histogram of S&P500 Prices",  xlim= c(2000, 60
00), xlab = "Price", ylab = "Frequency", breaks = 15)

mtext("Figure 5.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

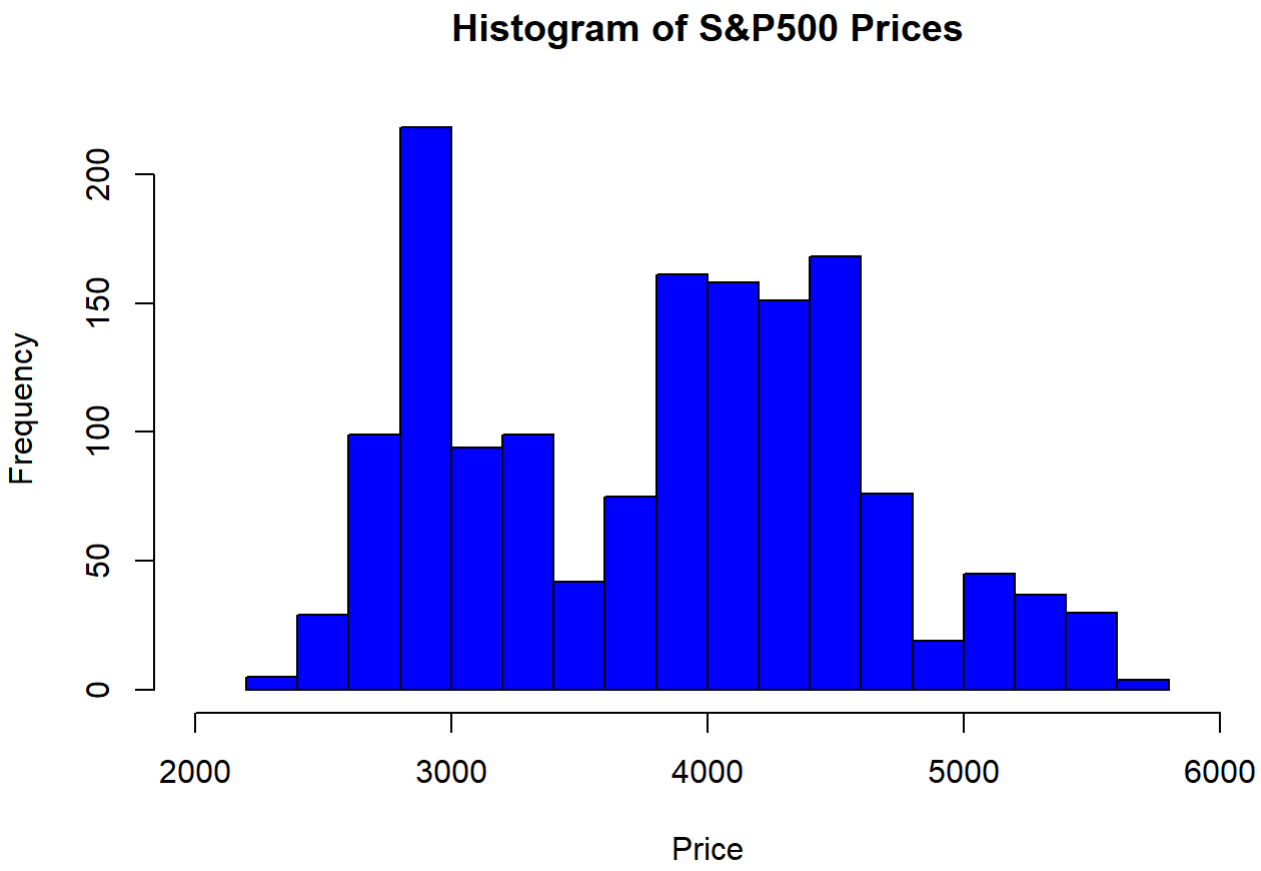

Figure 5.

```
qqnorm(merged_df$Price, main = "Q-Q Plot for S&P500 Prices", xlab = "Theoretical Quantiles of
Normal Distribution", ylab = "Sample Quantiles of S&P500 Prices")
qqline(merged_df$Price, col = "red", lw = 2)
mtext("Figure 6.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

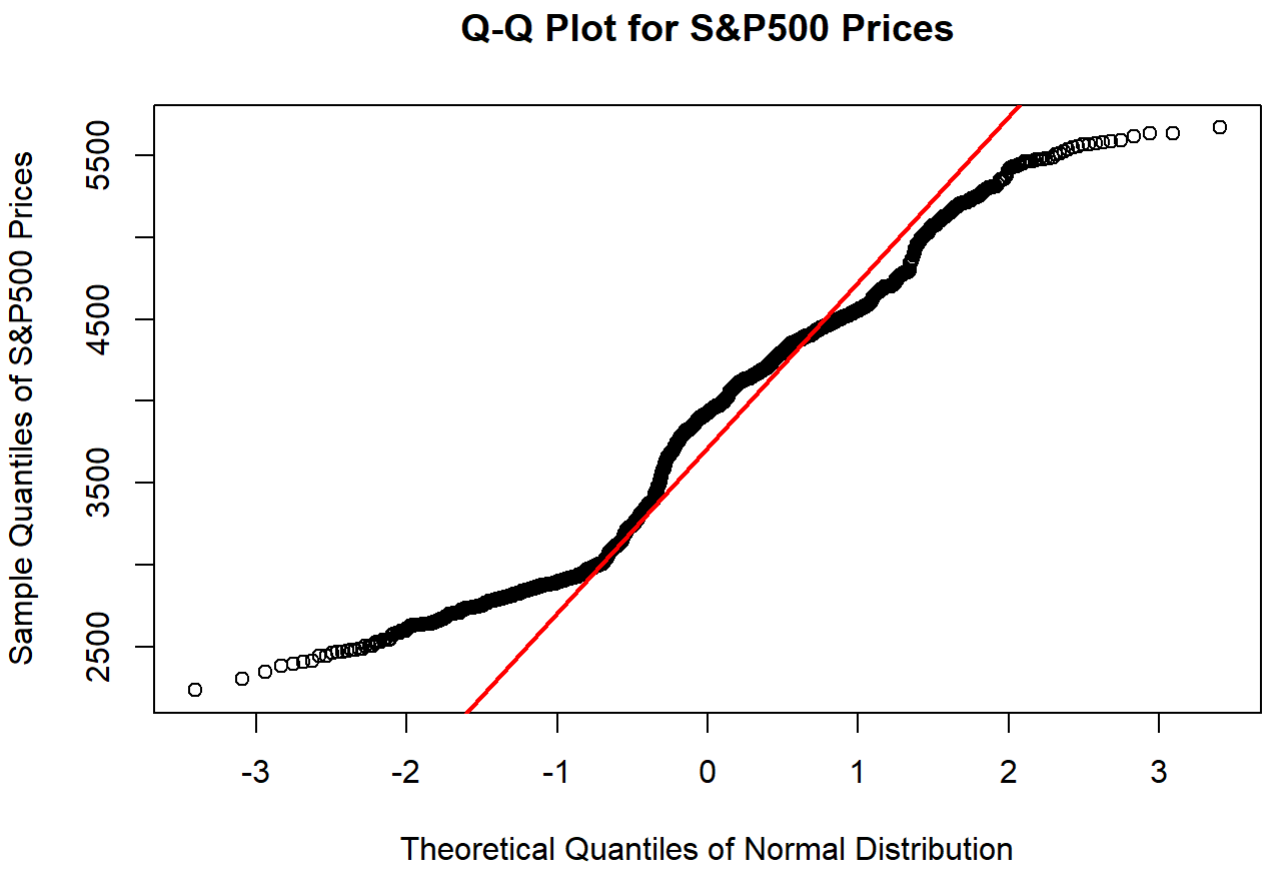

Figure 6.

```
merged_df$Adj.Close %>% hist(col = "orange", main = "Histogram of Bitcoin Adjusted Close Pric
es", xlim= c(0, 78000), xlab = "Adjusted Close", ylab = "Frequency", breaks = 15)
mtext("Figure 7.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

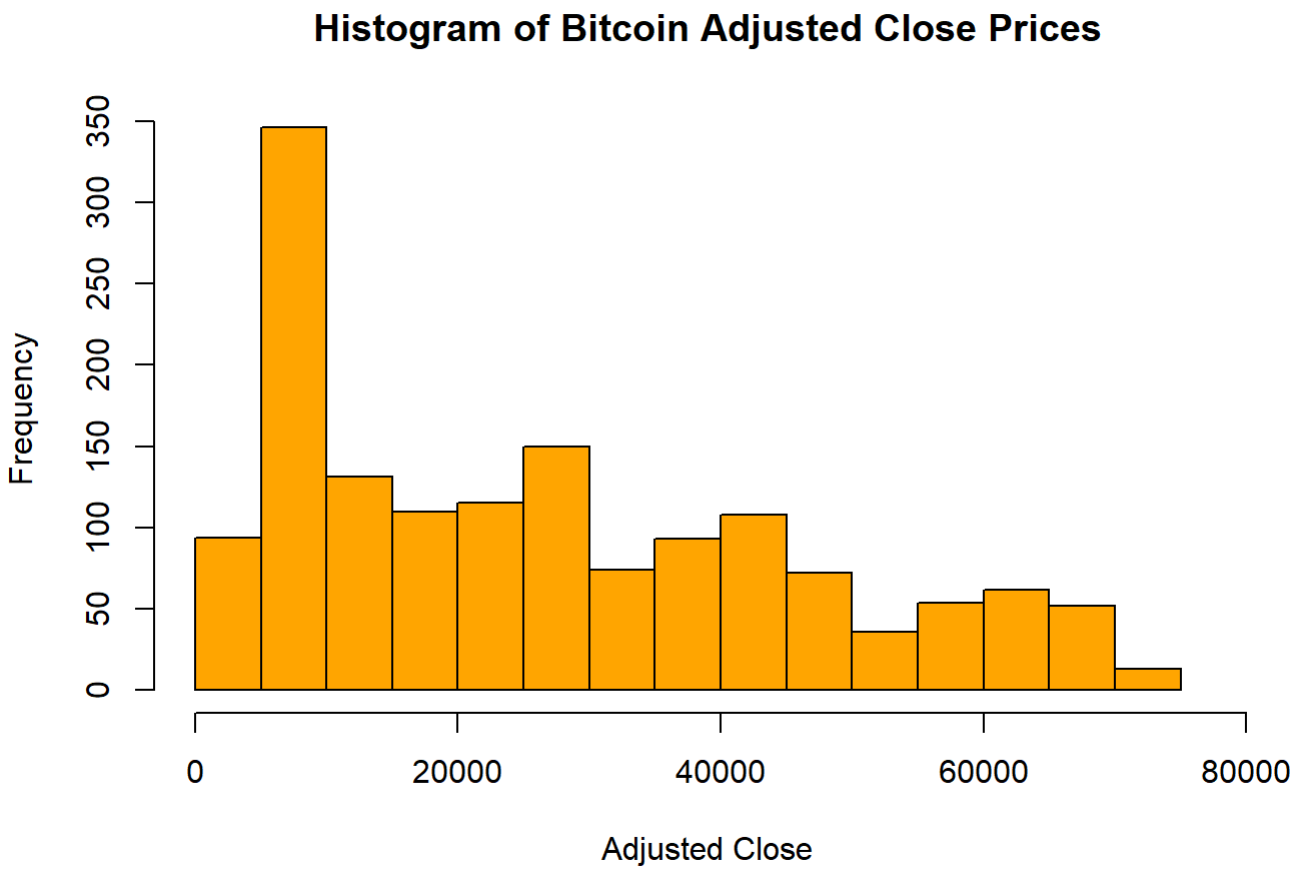

Figure 7.

```
qqnorm(merged_df$Adj.Close, main = "Q-Q Plot for Bitcoin Prices", xlab = "Theoretical Quantil
es of Normal Distribution", ylab = "Sample Quantiles of Bitcoin Prices")
qqline(merged_df$Adj.Close, col = "red", lw = 2)
mtext("Figure 8.", side = 3, line = -21, outer = FALSE, cex = 0.6, col = "black")
```

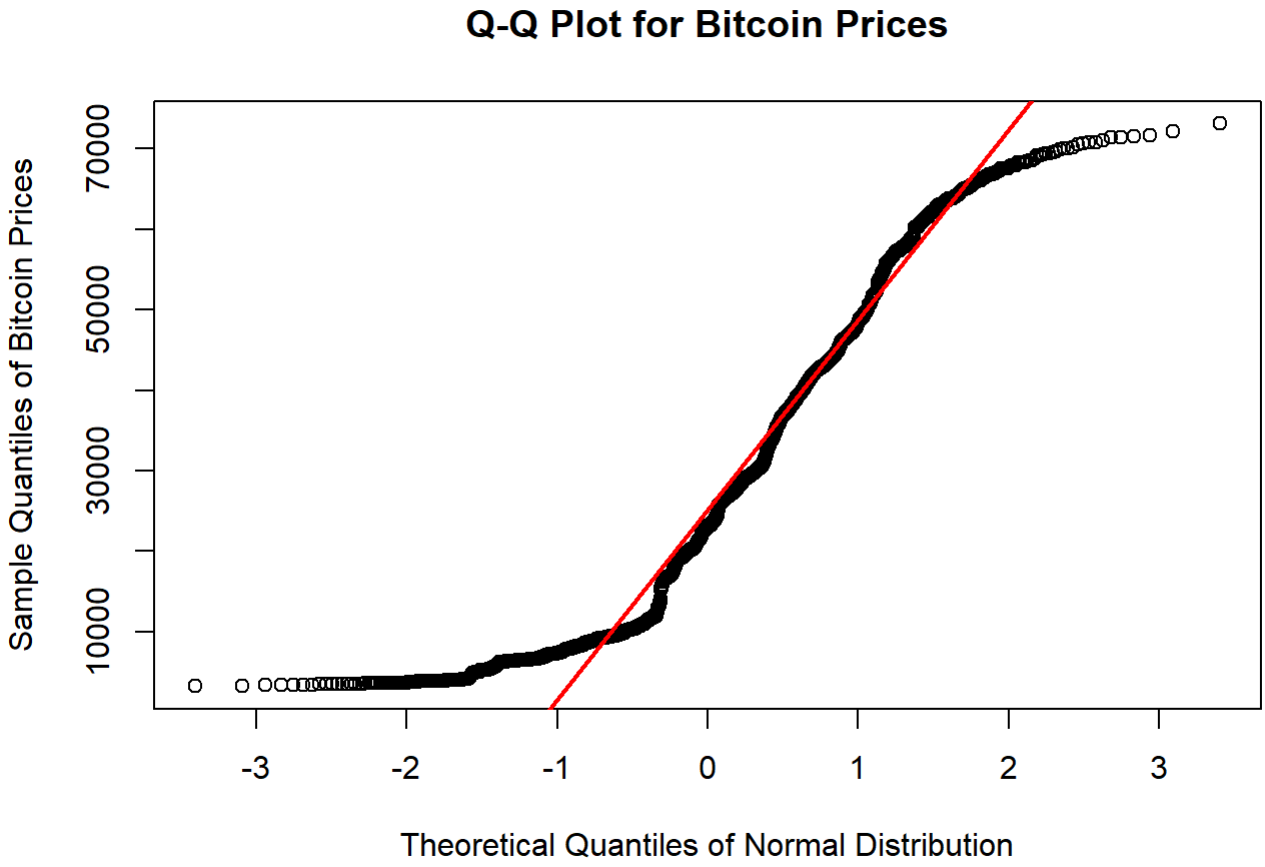

Figure 8.

Discussion

In task 1, the descriptive statistics of each financial asset over the past 6 years is calculated. It is then compared in a dataframe. It can be seen that Bitcoin is much more volatile than SP500. This is depicted through the large range it possesses, as well as the large variance and standard deviation (sd). This is not to say SP500 is not volatile, however is much less volatile than Bitcoin.

In figure 1, the SP500 trend over the last 6 years can be seen. It can be said that there is a slight increase over time, with multiple peaks and troughs. This suggests some volatility but has gradual increase in price, where it was around 2800 dollars in 2018, and ending up at around 5000 dollars in 2024. In comparison in figure 2, Bitcoin is said to be highly volatile because of the high peaks and low lows. Bitcoin stagnated from 2018 to end of 2020, being around 10,000 dollars. However, the giant peak that follows reached 60,000, and varied greatly, ultimately losing its value and dropping to 25,000 by mid 2022. There is a slow trend upwards, suggesting that Bitcoin may have stabilised in price because it has reached its peak price again in 2024.

In figure 3, the correlation between SP500 and Bitcoin prices was calculated over a 6 month interval. Having a correlation close to one suggests a positive relationship, where both SP500 and Bitcoin increased together. If negative relationship, one moves up, whilst the other moves down. There is fluctation between positive and negative, but is largely positive, suggesting that both SP500 and Bitcoin influence each other, or are both influenced in the same way from external factors.

In figure 4, a visual representation of the correlation efficient between SP500 and Bitcoin is depicted. The correlation is rounded to 0.9, stating that the relationship is very positive. This is represented in the graph, where the values are relatively close to the trend line. This further supports that SP500 and Bitcoin are closely related.

In figure 5, a histogram is created to determine the frequency distribution of SP500 prices over the last 6 years. The graph is not symmetrical, and possesses a few ‘more empty’ entries near the middle. Additionally, the left side is much more prominent than the right, as well as a great outlier around the 2900 dollar price. It has

uneven distribution.

Figure 6 displays the Q-Q plot which determines how close it follows normal distribution. SP500 has entries that skew heavily away from the trend line, especially at the tail ends. This means that SP500 deviates from the normality, and may possess outliers that is not expected of the normal distribution.

Figure 7 represents the frequency distribution of Bitcoin adjusted close prices over the last 6 years. The graph is positively skewed, where the tail is at the positive side of the graph. This represents that Bitcoin trended at a lower price for a much longer period, and that high values are either outlier peaks or not yet stabilised price of Bitcoin.

Figure 8 is the Q-Q plot for Bitcoin in determining if it follows normal distribution. It can be seen that Bitcoin follows the trend line more closely than SP500, however the tails are still far off. This represent that Bitcoin typically follows normal distribution, but can have high volatility at rapid moments.

Conclusion

It was found that SP500 and Bitcoin are closely related through their correlation coefficient (0.9). What this represents it that they each have a strong effect on the other. Individually, both are quite volatile, as they do not adhere to the trend lines greatly, but rather have tails that deviate greatly. This suggests that they have somewhat consistent prices, but also can have moments of extreme peaks and troughs that count towards the tails deviating.

References

Allaire J, Xie Y, Dervieux C, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2024). *rmarkdown: Dynamic Documents for R*. R package version 2.27, https://github.com/rstudio/rmarkdown (https://github.com/rstudio/rmarkdown).

Xie Y, Allaire J, Grolemund G (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338, https://bookdown.org/yihui/rmarkdown (https://bookdown.org/yihui/rmarkdown).

Xie Y, Dervieux C, Riederer E (2020). *R Markdown Cookbook*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9780367563837, https://bookdown.org/yihui/rmarkdown-cookbook (https://bookdown.org/yihui/rmarkdown-cookbook).

R - Mean, Median and Mode - Tutorialspoint. (2019). Tutorialspoint.com. https://www.tutorialspoint.com/r/r_mean_median_mode.htm (https://www.tutorialspoint.com/r/r_mean_median_mode.htm)

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr (https://CRAN.R-project.org/package=dplyr).

Zach Bobbitt. (2021, September 29). How to Test for Normality in R (4 Methods). Statology. https://www.statology.org/test-for-normality-in-r/ (https://www.statology.org/test-for-normality-in-r/)