Week5-Visualization Activity

Shanmugapriya Mohan Kumar

2024-06-18

# Dataset URL  
spy\_data <- "https://raw.githubusercontent.com/ShanmugapriyaMohankumar/StockMarket/main/SPY.csv"  
  
# Read the dataset from the URL  
spy500\_data <- read\_csv(spy\_data,show\_col\_types = FALSE) #read\_csv is part of the readr package, which is included in the tidyverse collection of R packages  
  
#display the first few rows  
head(spy500\_data)

## # A tibble: 6 × 7  
## Date Open High Low Close `Adj Close` Volume  
## <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2023-06-02 424. 429. 424. 428. 422. 91366700  
## 2 2023-06-05 428. 430. 426. 427. 421. 65460200  
## 3 2023-06-06 427. 429. 426. 428. 422. 64022200  
## 4 2023-06-07 428. 430. 426. 427. 420. 85373300  
## 5 2023-06-08 427. 430. 426. 429. 423. 61952800  
## 6 2023-06-09 430. 432. 429. 430. 424. 85742800

# Convert Date column to Date type  
spy500\_data$Date <- as.Date(spy500\_data$Date)  
  
#display the first few rows of the dataframe  
head(spy500\_data)

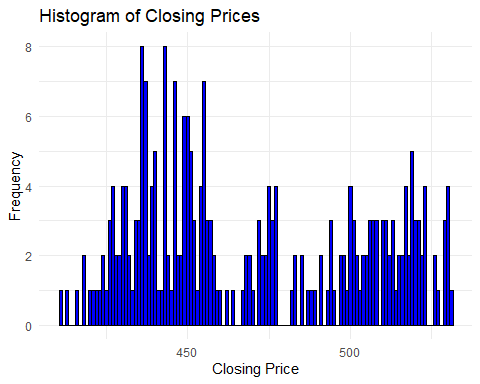
## # A tibble: 6 × 7  
## Date Open High Low Close `Adj Close` Volume  
## <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2023-06-02 424. 429. 424. 428. 422. 91366700  
## 2 2023-06-05 428. 430. 426. 427. 421. 65460200  
## 3 2023-06-06 427. 429. 426. 428. 422. 64022200  
## 4 2023-06-07 428. 430. 426. 427. 420. 85373300  
## 5 2023-06-08 427. 430. 426. 429. 423. 61952800  
## 6 2023-06-09 430. 432. 429. 430. 424. 85742800

#Number of rows in the original dataset  
spy500\_data\_row\_count <- nrow(spy500\_data)

### Visualization

***HISTOGRAM***

ggplot(spy500\_data, aes(x = Close)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 labs(title = "Histogram of Closing Prices",  
 x = "Closing Price",  
 y = "Frequency") +  
 theme\_minimal()



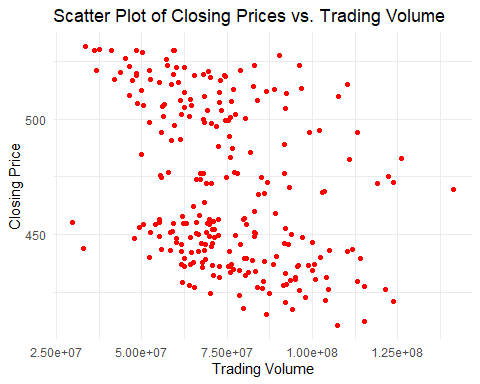
### Histogram Of closing prices

The histogram shows the closing prices for SPY over the observed period.The data is divided into intervals, each representing a range of closing prices.The height of each bar indicates how many times the closing prices fell within each bin(interval).The tallest bars represent the price ranges where SPY closed most frequently. This helps in identifying the central tendency of the data. The shape of the histogram can indicate whether the data is normally distributed, skewed, or has multiple peaks. Bars that are isolated from the main cluster can indicate outliers, where the closing price was significantly higher or lower than usual.

Example: If the histogram shows a peak around $420 to $430, it suggests that SPY often closed within this range. If there is a tail extending to higher prices, then it indicates bullish market sentiment because it suggests that buyers were able to push the price higher, even if it didn’t close there frequently. It is a positive sign in the market for buyers.

***SCATTERPLOT***

ggplot(spy500\_data, aes(x = Volume, y = Close)) +  
 geom\_point(color = "red") +  
 labs(title = "Scatter Plot of Closing Prices vs. Trading Volume",   
 x = "Trading Volume",   
 y = "Closing Price") +  
 theme\_minimal()



### Scatterplot of Closing Prices vs. Trading Volume

The scatterplot of SPY closing prices versus trading volume helps in understanding how these two variables interact.

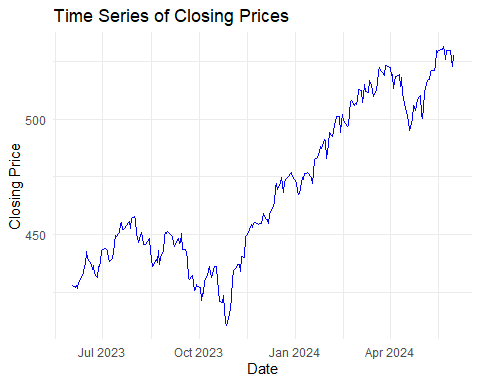
Description: X-Axis: Represents the trading volume. Y-Axis: Represents the closing prices of SPY. Points: Each point represents a day’s closing price and the corresponding trading volume.

The pattern of points can indicate whether there is a positive, negative, or no correlation between trading volume and closing prices. Clusters of points can indicate typical trading behaviors. For example, high volumes might cluster around certain price ranges. Points that are far from the cluster can indicate unusual trading activity, such as a day with very high volume and an abnormal closing price.

Example: There is no clear pattern, it indicates that volume and price do not have a direct relationship in this dataset.

***TIMESERIES PLOT***

ggplot(spy500\_data, aes(x = Date, y = Close)) +  
 geom\_line(color = "blue") +  
 labs(title = "Time Series of Closing Prices",   
 x = "Date",   
 y = "Closing Price") +  
 theme\_minimal()



### Time series plot

It shows the trend of SPY’s closing prices over the observed period(past 1 year). The plot shows how the closing prices change over time, highlighting periods of rise, fall, and stability. It helps in understanding the performance and volatility of SPY over the specified period and long term trends can be identified. Peaks and fall in the plot can indicate significant market events or trends.

Example : In Mid-November-2023, there is a fall. This could be due to specific events such as economic policy changes, interest rate announcements, or major geopolitical events. But the prices show recovery towards the end of November, indicating market stabilization after the initial shock.