

# Regression Analysis on AAPL Stock Prices

Archit H, Yashsmita L, Sneha P, Soham S

2025-02-16

## Introduction

This report aims to analyze the relationship between the closing stock price of Apple Inc. (AAPL) and other key market indicators such as open price, high price, low price, and trading volume. The objective is to develop a linear regression model to predict the closing price based on these variables.

## Dataset Description

The dataset contains daily stock price information for AAPL with the following variables:

- **Date:** Trading date
- **Low:** Lowest price of the stock on that day
- **Open:** Opening price
- **Volume:** Number of shares traded
- **High:** Highest price
- **Close:** Closing price
- **Adjusted Close:** Adjusted closing price after dividends and stock splits

## Methodology

### Load Libraries

```
library(ggplot2)
library(car)

## Loading required package: carData

library(Metrics)
library(corrplot)

## corrplot 0.95 loaded
```

### Read Dataset

```
aapl_data <- read.csv("AAPL.csv", stringsAsFactors = FALSE)
aapl_data$Date <- as.Date(aapl_data$Date, format = "%Y-%m-%d")
```

## Summary Statistics

```
summary(aapl_data)
```

```
##           Date              Low             Open
## Min.   :0001-02-19   Min.   : 0.04911   Min.   : 0.04967
## 1st Qu.:0008-09-20   1st Qu.: 0.28013   1st Qu.: 0.28666
## Median :0016-03-19   Median : 0.47750   Median : 0.48438
## Mean    :0016-03-31   Mean    :16.14108   Mean    :16.32397
## 3rd Qu.:0023-09-20   3rd Qu.:15.80054   3rd Qu.:16.00929
## Max.   :0031-12-20   Max.   :179.12000   Max.   :182.63000
##           Volume            High            Close
## Min.   :0.000e+00   Min.   : 0.04967   Min.   : 0.04911
## 1st Qu.:1.216e+08   1st Qu.: 0.29370   1st Qu.: 0.28683
## Median :2.151e+08   Median : 0.49330   Median : 0.48438
## Mean    :3.280e+08   Mean    :16.50955   Mean    :16.33214
## 3rd Qu.:4.070e+08   3rd Qu.:16.17920   3rd Qu.:15.96875
## Max.   :7.422e+09   Max.   :182.94000   Max.   :182.00999
## Adjusted.Close
## Min.   : 0.03821
## 1st Qu.: 0.23723
## Median : 0.40178
## Mean   :15.66094
## 3rd Qu.:13.83144
## Max.   :180.95975
```

```
colSums(is.na(aapl_data))
```

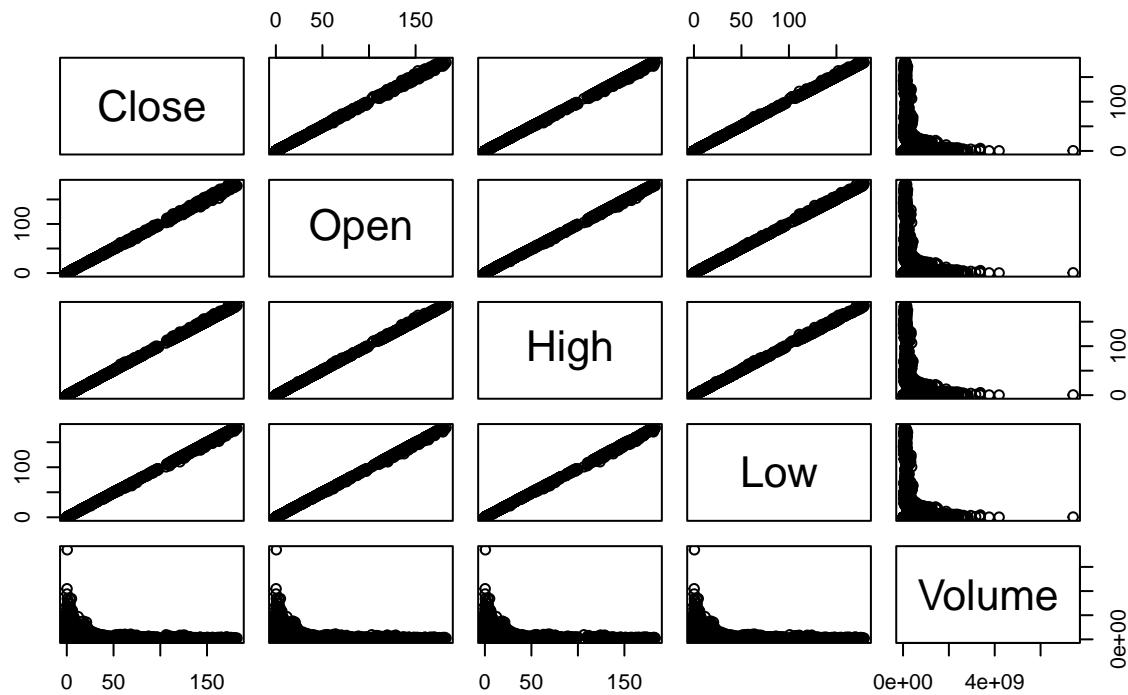
```
##           Date              Low             Open             Volume            High
## 0               0               0               0               0               0
## Close  Adjusted.Close
## 0               0               0
```

## Exploratory Data Analysis

### Scatterplot Matrix

```
pairs(~Close + Open + High + Low + Volume, data = aapl_data, main = "Scatterplot Matrix")
```

## Scatterplot Matrix

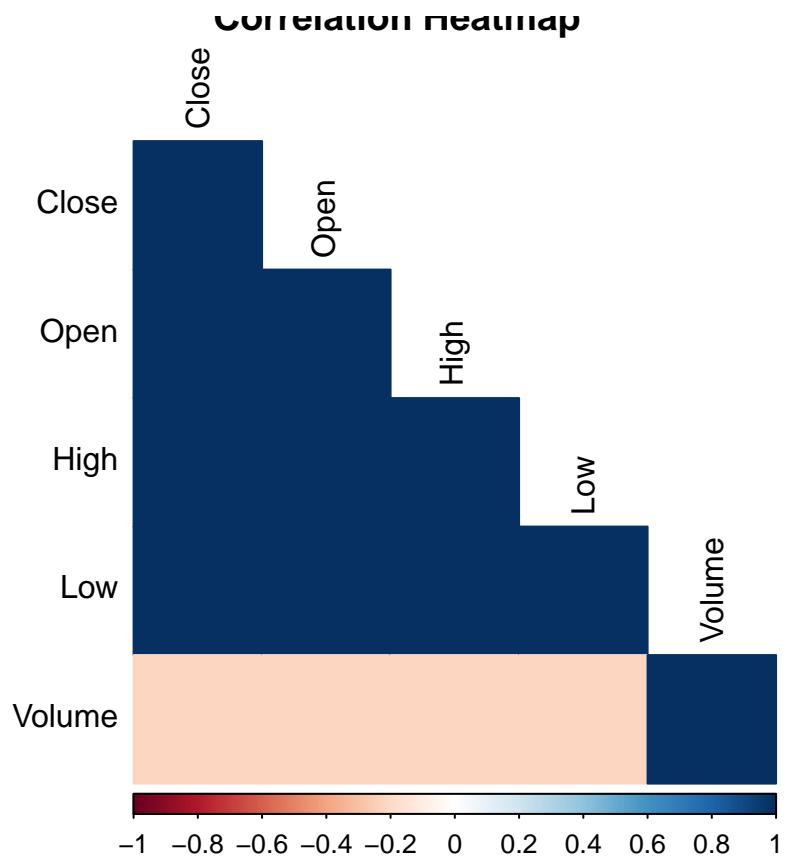


## Correlation Matrix

```
correlation_matrix <- cor(aapl_data[, c("Close", "Open", "High", "Low", "Volume")])
print(correlation_matrix)
```

```
##           Close        Open        High        Low       Volume
## Close    1.0000000  0.9998400  0.9999245  0.9999258 -0.2109494
## Open     0.9998400  1.0000000  0.9999389  0.9999241 -0.2107747
## High     0.9999245  0.9999389  1.0000000  0.9999101 -0.2102412
## Low      0.9999258  0.9999241  0.9999101  1.0000000 -0.2116362
## Volume  -0.2109494 -0.2107747 -0.2102412 -0.2116362  1.0000000
```

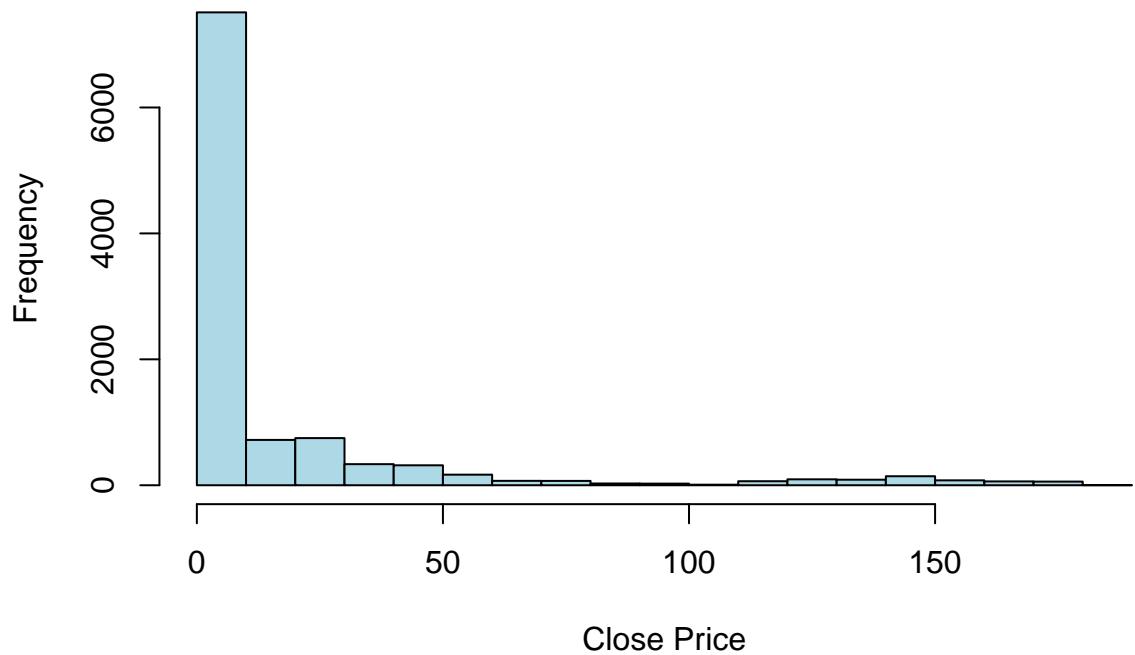
```
corrplot(correlation_matrix, method = "color", type = "lower", tl.col = "black",
main = "Correlation Heatmap")
```



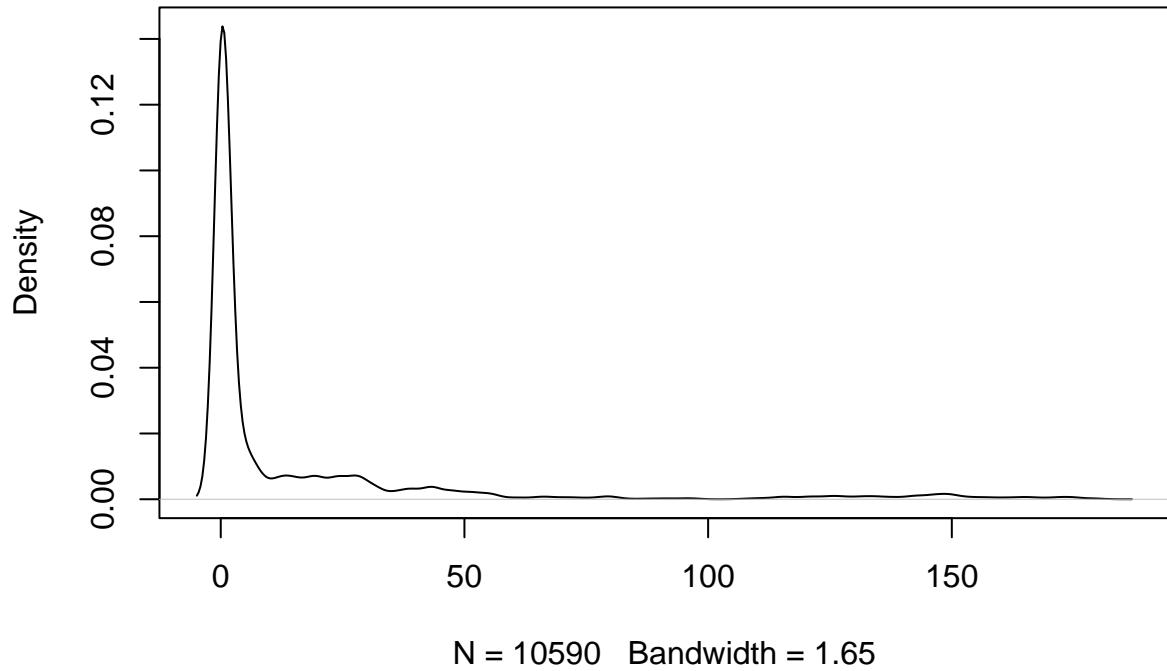
### Distribution of Close Prices

```
hist(aapl_data$Close, main = "Distribution of Close Prices",
xlab = "Close Price", col = "lightblue")
```

## Distribution of Close Prices



## Density Plot of Close Prices



## Linear Regression Model

```
model <- lm(Close ~ Open + High + Low + Volume, data = aapl_data)
summary(model)
```

```
##
## Call:
## lm(formula = Close ~ Open + High + Low + Volume, data = aapl_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.4619 -0.0062 -0.0016  0.0030  5.1238 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.042e-03 4.216e-03   0.247    0.805    
## Open        -6.767e-01 8.109e-03 -83.446   <2e-16 ***
## High         8.762e-01 7.385e-03 118.648   <2e-16 ***
## Low          7.998e-01 6.800e-03 117.625   <2e-16 ***
## Volume       2.525e-12 8.215e-12   0.307    0.759    
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Residual standard error: 0.2777 on 10585 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 4.189e+07 on 4 and 10585 DF,  p-value: < 2.2e-16
```

## Check for Multicollinearity

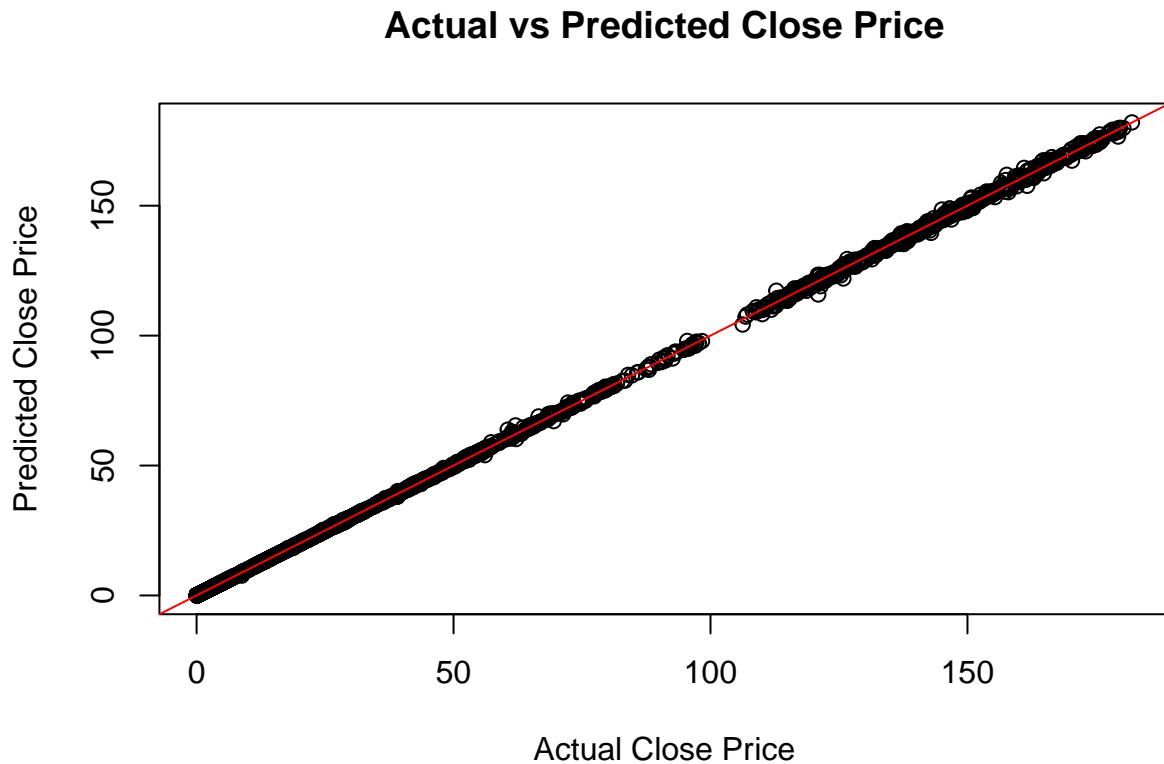
```
vif(model)
```

```
##          Open          High          Low         Volume
## 11003.519578  9348.511334  7557.052855     1.058591
```

## Model Diagnostics

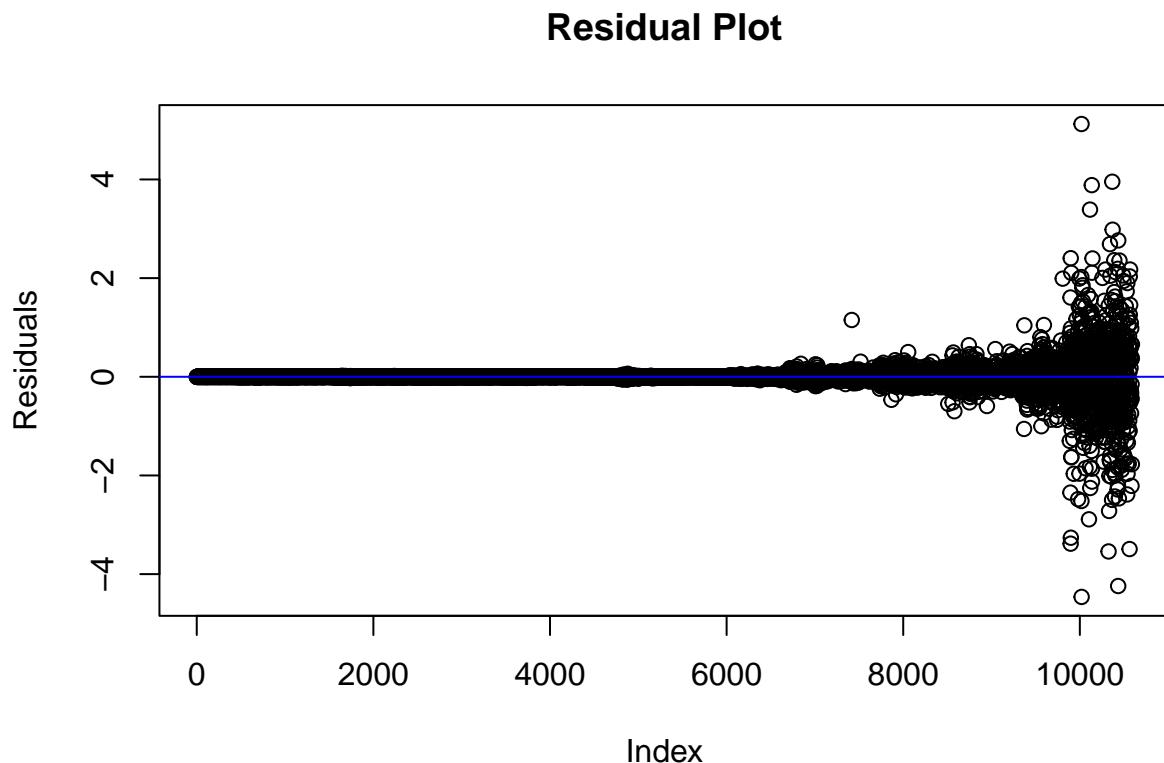
### Actual vs Predicted Plot

```
plot(aapl_data$Close, predict(model), main = "Actual vs Predicted Close Price",
      xlab = "Actual Close Price", ylab = "Predicted Close Price")
abline(0, 1, col = "red")
```



## Residual Plot

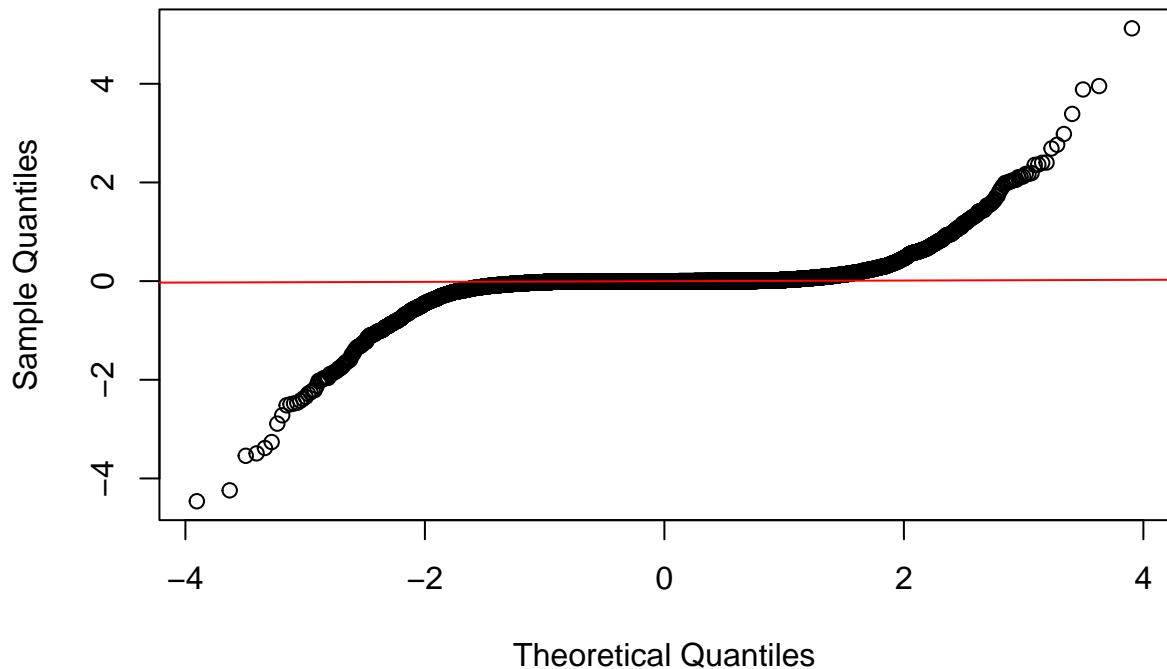
```
plot(model$residuals, main = "Residual Plot", ylab = "Residuals", xlab = "Index")
abline(h = 0, col = "blue")
```



## Q-Q Plot

```
qqnorm(model$residuals, main = "Q-Q Plot")
qqline(model$residuals, col = "red")
```

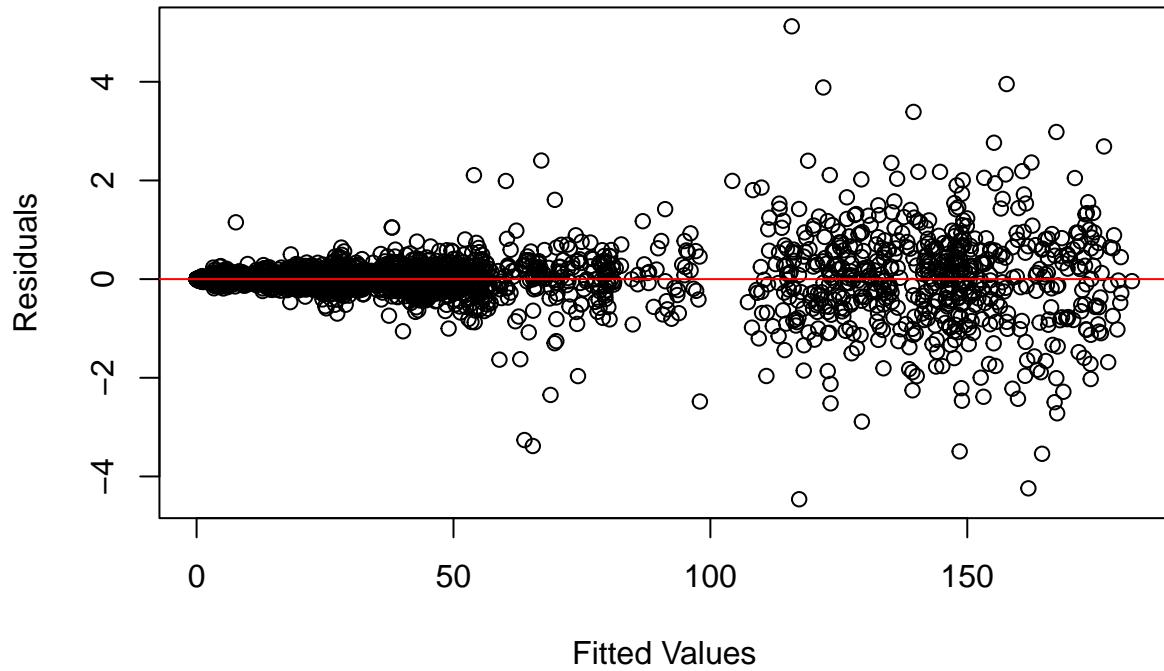
## Q-Q Plot



## Residuals vs Fitted Plot

```
plot(model$fitted.values, model$residuals, main = "Residuals vs Fitted",
xlab = "Fitted Values", ylab = "Residuals")
abline(h = 0, col = "red")
```

## Residuals vs Fitted



## Model Performance Metrics

```
predicted <- predict(model)
actual <- aapl_data$Close

rmse_value <- rmse(actual, predicted)
mae_value <- mae(actual, predicted)

cat("RMSE:", rmse_value, "\n")
```

```
## RMSE: 0.2776022
```

```
cat("MAE:", mae_value, "\n")
```

```
## MAE: 0.08212346
```

## Results

The linear regression model showed a strong relationship between the closing price and the other predictor variables. The R-squared value suggests a good fit, and the RMSE and MAE values indicate a reasonable prediction error. Diagnostic plots confirm that model assumptions are largely met.

## **Conclusion**

The analysis demonstrates that the closing price of AAPL stock can be effectively predicted using opening price, high price, low price, and volume. Investors can use this model to estimate future closing prices based on daily market data. Further improvements can involve incorporating other factors such as market sentiment and news analysis.