Market Price Movements Analysis

Patrick Redington

2023-02-05

Market Price Movements Analysis

Introduction

Predicting market movements and price changes has been a challenge for investors, traders, and financial institutions alike. Understanding the direction of the market can help individuals make informed investment decisions and potentially increase their returns. With the advent of advanced technologies and the increasing availability of financial data, there has been a growing interest in using data science techniques to predict market movements.

The purpose of this project is to use data science methods to attempt to predict market movements and price changes for the overall market. This is a data science problem because it involves using data to make predictions about future events. In this case, we will use historical stock market and economical data to attempt to build predictive models that will help us understand the factors that influence stock prices and predict future market movements. By researching this field of study, we can hopefully gain a deeper understanding of the stock market and the factors that drive stock prices, and potentially make more informed investment decisions.

Research Questions

1. How does the volatility expectation index (VIX) influence the price movements of the SPDR S&P 500 ETF (SPY)?
2. How does the rate of change in the VIX affect the price movements of the SPY?
3. How does the sentiment of national news events impact the SPY price changes?
4. At what times do the greatest changes in SPY prices occur, during trading hours or after-hours?
5. How does the federal interest rate impact the SPY price movements?
6. How does inflation impact federal interest rates and the SPY price changes?
7. How does unemployment affect the federal interest rates and the price movements of the SPY?
8. Would a 7, 14, or 28 day rolling average of prices enhance the accuracy of the model?

Approach

The SPY ETF, or the SPDR S&P 500 exchange-traded fund, is a popular choice among investors, traders, and financial institutions due to its aim to track the performance of the S&P 500 index at a lower price point and greater availability and liquidity compared to the index itself. The S&P 500 Index, composed of 500 large-cap stocks, is considered a significant indicator of the U.S. stock market and economy.

To build predictive models that understand the factors that influence stock prices and predict future market movements, the news events of each day will be analyzed and given a sentiment rating and strength rating/weight based on their significance for that day. The VIX values, used to estimate the expected future volatility, will also be utilized by considering both the total value and rate of change of the VIX values. Additionally, the historical U.S. data will be used to determine the impact of macroeconomic factors on the market.

How The Approach Addresses The Problem

The goal of the project is not to measure day-to-day changes with a high level of accuracy, but to determine major market changes over time. This approach reduces the complexity of the problem and reduces the risk of outliers that day-to-day movements may cause. By focusing on major market changes, the project aims to provide insights into the stock market and the factors that influence stock prices, and to help individuals make more informed investment decisions.

Required Packages

```r
#install.packages("tinytex")
library(tinytex)
#install.packages("tidyverse")
library(tidyverse)
#install.packages("dplyr")
library(dplyr)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("lubridate")
library(lubridate)
#install.packages("purrr")
library(purrr)
#install.packages("lm.beta")
library(lm.beta)
#install.packages("AICcmodavg")
library(AICcmodavg)
#install.packages("car")
library(car)
#install.packages("RcppRoll")
library(RcppRoll)
#install.packages("sentimentr")
library(sentimentr)
#install.packages("zoo")
library(zoo)
```

More may be needed but at this time the ones above are the only ones that are useful with my current level of knowledge. A few additional ones that may help in creating the model are below, but will take more time to learn.

```
#install.packages("caret")
library(caret)
#install.packages("glmnet")
library(glmnet)
#install.packages("e1071")
library(e1071)
#install.packages("randomForest")
library(randomForest)
```

Data Sources

The data used in this research was obtained from Kaggle.com, contributed by members of the community. The "spy.us.txt" dataset contains historical prices and volumes of the SPY ETF from 02/25/2005 to 11/10/2017 and has 3201 records with 7 variables. Although the data has an issue with the "open interest" variable containing only zeros, this variable is not necessary for the model and can be deleted.

The "vix.csv" dataset includes the historical values of the VIX (fear gauge) from 1/2/1990 to 6/28/2021, with 7934 records and 7 variables. The "volume" variable will need to be removed as it is all zeros, as the VIX is a calculation and not a traded asset.

The "index.csv" dataset contains information on the monthly economic conditions in the United States from 1954 to 2017, originally sourced from the Federal Reserve's portal. The data contains some blank and zero values that will require cleaning, as well as some months with double data points that need resolution for the model.

The "Combined_News_DJIA.csv" dataset is the most complex and includes the top 25 news events for each day from 08/08/2008 to 07/01/2016, ranked by Reddit user votes. This may cause bias issues as the importance of the news events may not reflect the market's views. The data has 27 variables and 1989 records and was originally sourced from Yahoo finance and the Reddit World News Channel.

All the datasets will need to have their date formatting standardized for consistency. Overall, the data is mostly ready to use, with some necessary cleaning and formatting required.

Community (2021) Reserve (2021) Besomi (2021) Aaron7Sun (N/A)

How to import and clean the data

To begin the data cleaning process, all data sets will be imported into R Studio using the default read function. The "spy_data" txt file will require additional processing, with the delimiter "," used to separate the data upon import.

To ensure consistent formatting across all data sets, the "lubridate" package functions will be used to standardize the date format. Additionally, the data frames will be subsetted to match each other's date ranges. This will help to reduce issues when creating the regression model, as some data sets contain decades more data than others.

Unnecessary columns will be removed using the "subset" function.Specifically, the "OpenInt", "Volume", "Adj.Close", "Label", "Year", "Month", and "Day" columns will be removed.

The "econ_data.csv" data set requires additional cleaning due to its monthly reporting format, which leads to month-long gaps in the data. To address this issue, each month's reported numbers will be copied to every record for that month. This approach ensures that the regression model can accurately predict how inflation, unemployment, and the federal interest rate affect the overall market.

How do you plan to slice and dice the data?

The data sets are relatively easy to work with and do not require significant slicing or restructuring. However, we will add some new columns to the "spy_data" and "vix_data" data frames to enhance their utility for model testing. Specifically, we will add columns to calculate the percent change from day to day, after-hours changes, and intraday changes.

Moreover, to further refine our data, we will include rolling averages of 7, 14, and 28 days in the "spy_data" data frame. These averages will help to reduce the impact of any outliers and provide more accurate results when testing the models.

```r
#Loads the various data sets.
#setwd("C:/Users/predi/Documents/GitHub/DSC520 Assignments")
spy_data <- read.delim("data/spy.us.txt",header=TRUE, sep=',')
vix_data <- read.csv("data/vix.csv")
econ_data <- read.csv("data/index.csv")
news_data <- read.csv("data/Combined_News_DJIA.csv")

#formats and cleans "spy_data".
#deletes blank column, and verifies dates are in the correct format and range.
spy_data <-subset(spy_data, select = -OpenInt)
spy_data$Date <- ymd(spy_data$Date)
spy_data <- spy_data[spy_data$Date > "2008-08-08" &
                       spy_data$Date < "2016-07-28", ]
#creates 7, 14, and 28 day rolling averages of the closing price.
spy_data$d7_avg <- roll_mean(spy_data$Close, n = 7, align = "right", fill = NA)
spy_data$d14_avg <- roll_mean(spy_data$Close, n = 14, align = "right",
                              fill = NA)
spy_data$d28_avg <- roll_mean(spy_data$Close, n = 28, align = "right",
                              fill = NA)
#Creates percent change columns for intra-day change, after hour change, and
#total change between days
spy_data$intra_daychange <- ((spy_data$Close / spy_data$Open) - 1)*100
spy_data$AHchange <- c(NA, (diff(spy_data$Open)/
                              spy_data$Close[-nrow(spy_data)]*100) - 1)
spy_data$spy_daychange <- c(NA, diff(spy_data$Close)/
                                spy_data$Close[-nrow(spy_data)]*100)

#formats and cleans "vix_data".
#deletes blank columns, and verifies dates are in the correct format and range.
vix_data <-subset(vix_data, select = -Volume)
vix_data <-subset(vix_data, select = -Adj.Close)
vix_data$Date <- ymd(vix_data$Date)
vix_data <- vix_data[vix_data$Date > "2008-08-08" &
                       vix_data$Date < "2016-07-28", ]
#creates percent change column between days
vix_data$vix_daychange <- c(NA, diff(vix_data$Close)/
                                vix_data$Close[-nrow(vix_data)]*100)


#formats and cleans "news_data".
#deletes the label" column and verifies dates are in the correct format.
#removes all but the top 10 articles for each day.
news_data <-subset(news_data, select = -Label)
news_data$Date <- ymd(news_data$Date)
news_data  <- news_data[, -c(12:26)]



#formats and cleans "econ_data".
#Combines the date columns into the same format as the out data frames.
#Then deletes the original date columns and verifies new date format.
econ_data$Date<-as.Date(with(econ_data,paste(Year,Month,Day,
                                             sep="-")),"%Y-%m-%d")
```

```r
econ_data <-subset(econ_data, select = -Year)
econ_data <-subset(econ_data, select = -Month)
econ_data <-subset(econ_data, select = -Day)
econ_data$Date <- ymd(econ_data$Date)
econ_data <- econ_data[econ_data$Date > "2008-08-08" &
                       econ_data$Date < "2016-07-28", ]
```

Sentiment analysis of the new articles

```r
#creates a sentiment score for each new article.
#Then adds all of the scores to the original df and
#removes the article names fro the df
news_dataTop1 <- sentiment_by(news_data$Top1)
news_dataTop2 <- sentiment_by(news_data$Top2)
news_dataTop3 <- sentiment_by(news_data$Top3)
news_dataTop4 <- sentiment_by(news_data$Top4)
news_dataTop5 <- sentiment_by(news_data$Top5)
news_dataTop6 <- sentiment_by(news_data$Top6)
news_dataTop7 <- sentiment_by(news_data$Top7)
news_dataTop8 <- sentiment_by(news_data$Top8)
news_dataTop9 <- sentiment_by(news_data$Top9)
news_dataTop10 <- sentiment_by(news_data$Top10)
news_data <- cbind(news_data, Top1 = news_dataTop1$ave_sentiment)
news_data <- cbind(news_data, Top2 = news_dataTop2$ave_sentiment)
news_data <- cbind(news_data, Top3 = news_dataTop3$ave_sentiment)
news_data <- cbind(news_data, Top4 = news_dataTop4$ave_sentiment)
news_data <- cbind(news_data, Top5 = news_dataTop5$ave_sentiment)
news_data <- cbind(news_data, Top6 = news_dataTop6$ave_sentiment)
news_data <- cbind(news_data, Top7 = news_dataTop7$ave_sentiment)
news_data <- cbind(news_data, Top8 = news_dataTop8$ave_sentiment)
news_data <- cbind(news_data, Top9 = news_dataTop9$ave_sentiment)
news_data <- cbind(news_data, Top10 = news_dataTop10$ave_sentiment)
news_data  <- news_data[, -c(2:11)]
```

What does the final data set look like?

The code below merges all cleaned and wrangled data frames into a single data frame for easier testing. The merge operation is performed using a left join on the date column to match the records across all data frames. The resulting data frame contains all columns from each data frame, with matching records combined into a single row. The summary below the code provides various statistics about each column in the merged data frame.

```
#combines all of the data frames into one.
spy_vix_data <-left_join(spy_data, vix_data, by=c("Date"))
spy_vix_econ_data <-left_join(spy_vix_data, econ_data, by=c("Date"))
all_data <-left_join(spy_vix_econ_data, news_data, by=c("Date"))

#Resolves all of the "NA" Records created by the monthly reporting of the
#historical economical data.
all_data$Federal.Funds.Target.Rate <-
  na.locf(all_data$Federal.Funds.Target.Rate, na.rm = FALSE)
all_data$Federal.Funds.Upper.Target <-
  na.locf(all_data$Federal.Funds.Upper.Target, na.rm = FALSE)
all_data$Federal.Funds.Target.Rate <-
  na.locf(all_data$Federal.Funds.Lower.Target, na.rm = FALSE)
all_data$Federal.Funds.Lower.Target <-
  na.locf(all_data$Federal.Funds.Target.Rate, na.rm = FALSE)
all_data$Effective.Federal.Funds.Rate <-
  na.locf(all_data$Effective.Federal.Funds.Rate, na.rm = FALSE)
all_data$Real.GDP..Percent.Change. <-
  na.locf(all_data$Real.GDP..Percent.Change., na.rm = FALSE)
all_data$Unemployment.Rate <-
  na.locf(all_data$Unemployment.Rate, na.rm = FALSE)
all_data$Inflation.Rate <-
  na.locf(all_data$Inflation.Rate, na.rm = FALSE)
summary(all_data)
```

```
##       Date                Open.x         High.x          Low.x
##  Min.   :2008-08-11   Min.   : 59.2   Min.   : 60.98   Min.   : 58.45
##  1st Qu.:2010-08-05   1st Qu.:101.9   1st Qu.:102.77   1st Qu.:101.02
##  Median :2012-08-01   Median :125.4   Median :125.82   Median :124.72
##  Mean   :2012-08-02   Mean   :137.2   Mean   :138.04   Mean   :136.34
##  3rd Qu.:2014-07-31   3rd Qu.:182.0   3rd Qu.:183.26   3rd Qu.:180.91
##  Max.   :2016-07-27   Max.   :211.8   Max.   :211.99   Max.   :210.89
##
##     Close.x          Volume              d7_avg          d14_avg
##  Min.   : 59.33   Min.   :3.889e+07   Min.   : 61.03   Min.   : 63.28
##  1st Qu.:102.00   1st Qu.:1.127e+08   1st Qu.:102.02   1st Qu.:102.26
##  Median :125.39   Median :1.616e+08   Median :125.26   Median :125.29
##  Mean   :137.25   Mean   :1.943e+08   Mean   :137.18   Mean   :137.09
##  3rd Qu.:181.63   3rd Qu.:2.366e+08   3rd Qu.:181.91   3rd Qu.:182.47
##  Max.   :211.86   Max.   :1.000e+09   Max.   :211.31   Max.   :210.45
##                                       NA's   :6        NA's   :13
##     d28_avg       intra_daychange       AHchange         spy_daychange
##  Min.   : 66.07   Min.   :-8.99091   Min.   :-15.2251   Min.   :-9.8463
##  1st Qu.:102.04   1st Qu.:-0.39046   1st Qu.: -1.5127   1st Qu.:-0.4471
##  Median :124.55   Median : 0.07342   Median : -0.8942   Median : 0.0694
##  Mean   :136.94   Mean   : 0.02325   Mean   : -0.9711   Mean   : 0.0403
```

```
##  3rd Qu.:183.17   3rd Qu.: 0.47534   3rd Qu.: -0.3710   3rd Qu.: 0.6015
##  Max.   :206.28   Max.   : 7.96679   Max.   :  9.6839   Max.   :14.5192
##  NA's   :27                          NA's   :1          NA's   :1
##      Open.y          High.y          Low.y          Close.y
##  Min.   :10.40   Min.   :10.76   Min.   :10.28   Min.   :10.32
##  1st Qu.:14.65   1st Qu.:15.28   1st Qu.:14.02   1st Qu.:14.52
##  Median :17.89   Median :18.61   Median :17.12   Median :17.79
##  Mean   :21.42   Mean   :22.40   Mean   :20.44   Mean   :21.30
##  3rd Qu.:24.00   3rd Qu.:25.14   3rd Qu.:22.92   3rd Qu.:24.06
##  Max.   :80.74   Max.   :89.53   Max.   :72.76   Max.   :80.86
##
##  vix_daychange      Federal.Funds.Target.Rate Federal.Funds.Upper.Target
##  Min.   :-29.5726   Min.   :0.0000            Min.   :0.2500
##  1st Qu.: -4.1070   1st Qu.:0.0000            1st Qu.:0.2500
##  Median : -0.6836   Median :0.0000            Median :0.2500
##  Mean   :  0.2601   Mean   :0.0201            Mean   :0.2701
##  3rd Qu.:  3.7507   3rd Qu.:0.0000            3rd Qu.:0.2500
##  Max.   : 50.0000   Max.   :0.2500            Max.   :0.5000
##  NA's   :1          NA's   :89                NA's   :89
##  Federal.Funds.Lower.Target Effective.Federal.Funds.Rate
##  Min.   :0.0000             Min.   :0.0700
##  1st Qu.:0.0000             1st Qu.:0.0900
##  Median :0.0000             Median :0.1300
##  Mean   :0.0201             Mean   :0.1642
##  3rd Qu.:0.0000             3rd Qu.:0.1600
##  Max.   :0.2500             Max.   :0.9700
##  NA's   :89                 NA's   :36
##  Real.GDP..Percent.Change. Unemployment.Rate Inflation.Rate       Top1
##  Min.   :-8.200            Min.   : 4.900    Min.   :0.600   Min.   :-1.61542
##  1st Qu.: 0.800            1st Qu.: 6.200    1st Qu.:1.600   1st Qu.:-0.25000
##  Median : 2.000            Median : 7.650    Median :1.800   Median :-0.07906
##  Mean   : 1.452            Mean   : 7.566    Mean   :1.729   Mean   :-0.09766
##  3rd Qu.: 3.100            3rd Qu.: 9.100    3rd Qu.:2.000   3rd Qu.: 0.04499
##  Max.   : 5.000            Max.   :10.000    Max.   :2.300   Max.   : 0.93897
##  NA's   :36                NA's   :36        NA's   :36      NA's   :17
##       Top2               Top3               Top4               Top5
##  Min.   :-2.15629   Min.   :-1.27802   Min.   :-1.18333   Min.   :-1.42302
##  1st Qu.:-0.29152   1st Qu.:-0.26780   1st Qu.:-0.26671   1st Qu.:-0.28309
##  Median :-0.09889   Median :-0.09449   Median :-0.08561   Median :-0.10206
##  Mean   :-0.12316   Mean   :-0.10781   Mean   :-0.10142   Mean   :-0.11383
##  3rd Qu.: 0.04004   3rd Qu.: 0.05549   3rd Qu.: 0.06155   3rd Qu.: 0.04352
##  Max.   : 0.96381   Max.   : 1.10000   Max.   : 1.20000   Max.   : 0.96381
##  NA's   :17         NA's   :17         NA's   :17         NA's   :17
##       Top6               Top7               Top8               Top9
##  Min.   :-1.39374   Min.   :-2.10764   Min.   :-1.53076   Min.   :-1.40068
##  1st Qu.:-0.27496   1st Qu.:-0.27839   1st Qu.:-0.27926   1st Qu.:-0.26863
##  Median :-0.09079   Median :-0.10206   Median :-0.09449   Median :-0.08839
##  Mean   :-0.10578   Mean   :-0.11967   Mean   :-0.11451   Mean   :-0.11400
##  3rd Qu.: 0.06010   3rd Qu.: 0.02948   3rd Qu.: 0.05393   3rd Qu.: 0.02637
##  Max.   : 1.02514   Max.   : 1.04103   Max.   : 0.85977   Max.   : 1.11369
##  NA's   :17         NA's   :17         NA's   :17         NA's   :17
##      Top10
##  Min.   :-1.16768
##  1st Qu.:-0.29957
```

```
##   Median :-0.10717
##   Mean   :-0.11948
##   3rd Qu.: 0.03387
##   Max.   : 1.02632
##   NA's   :17
```

Plots and Tables Needed

1. Scatter Plots

- Could be used to visualize the relationship between VIX values and SPY price movements.
- Used to show the sentiment/importance of news events and their impact on SPY prices.

2. Box plots

- Could be used to compare SPY price changes during trading hours and after hours.

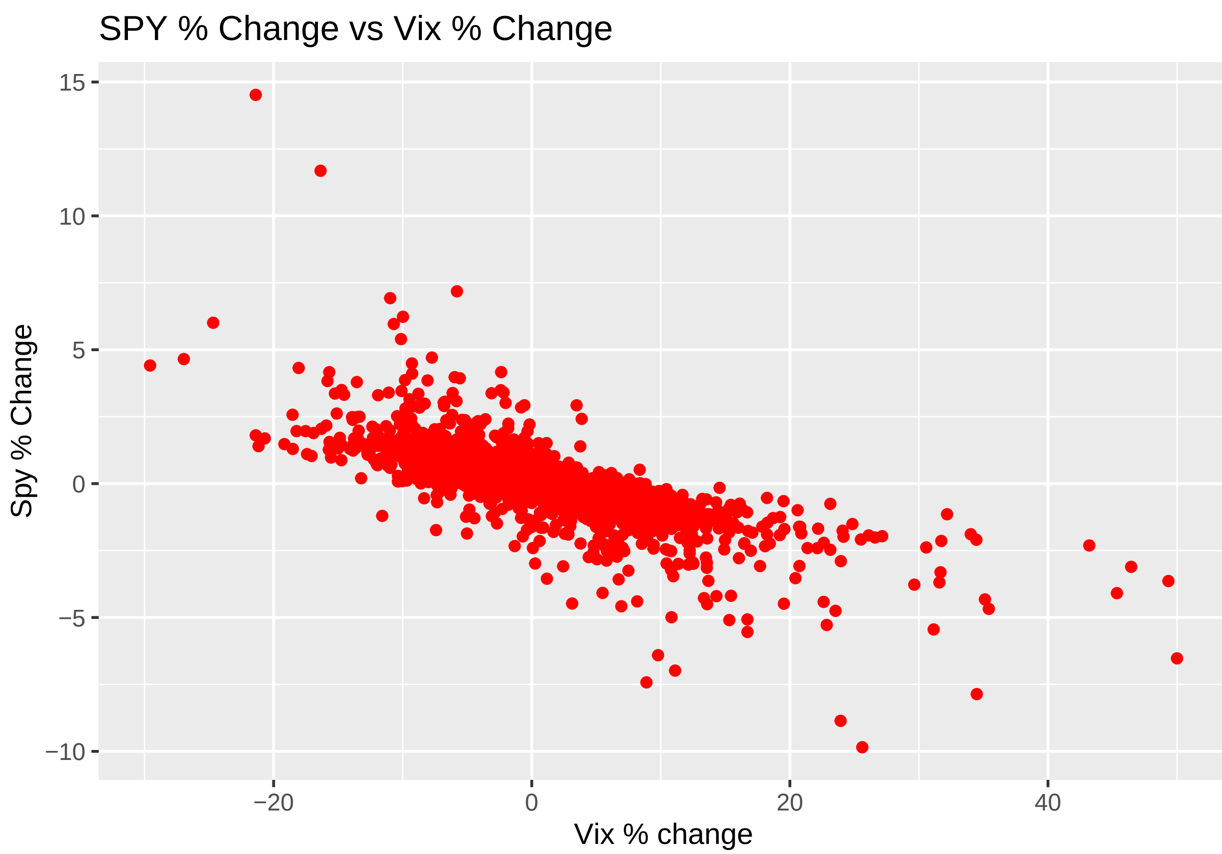
3. Bar Graphs

- Could be used to test for normal distribution of variables.

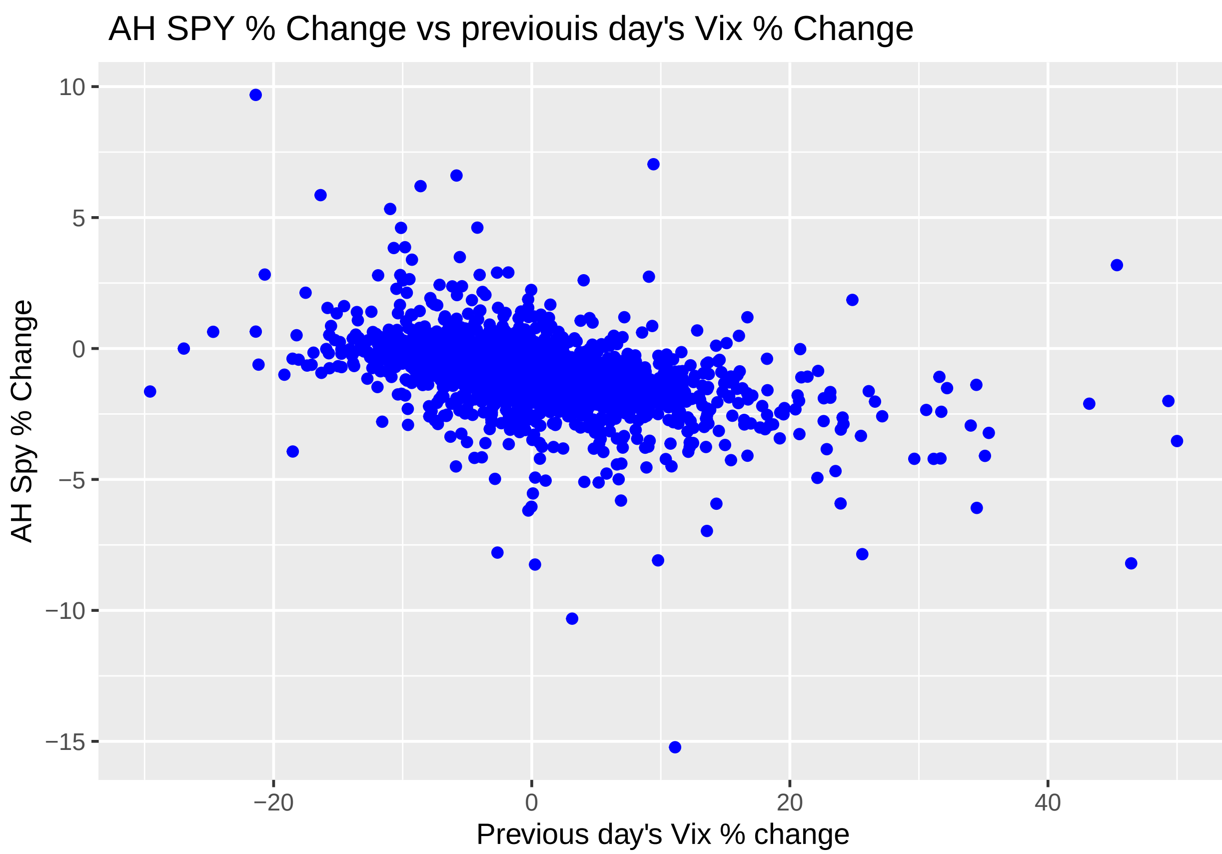
4. Line Plots

- Used to show the relationship between inflation and federal interest rates, and the impact on SPY prices. -Used to visualize the rate of change of the VIX and its impact on SPY price movements over time.
- Could be used to compare the accuracy of the model using 7-day, 14-day, and 28-day rolling price averages.

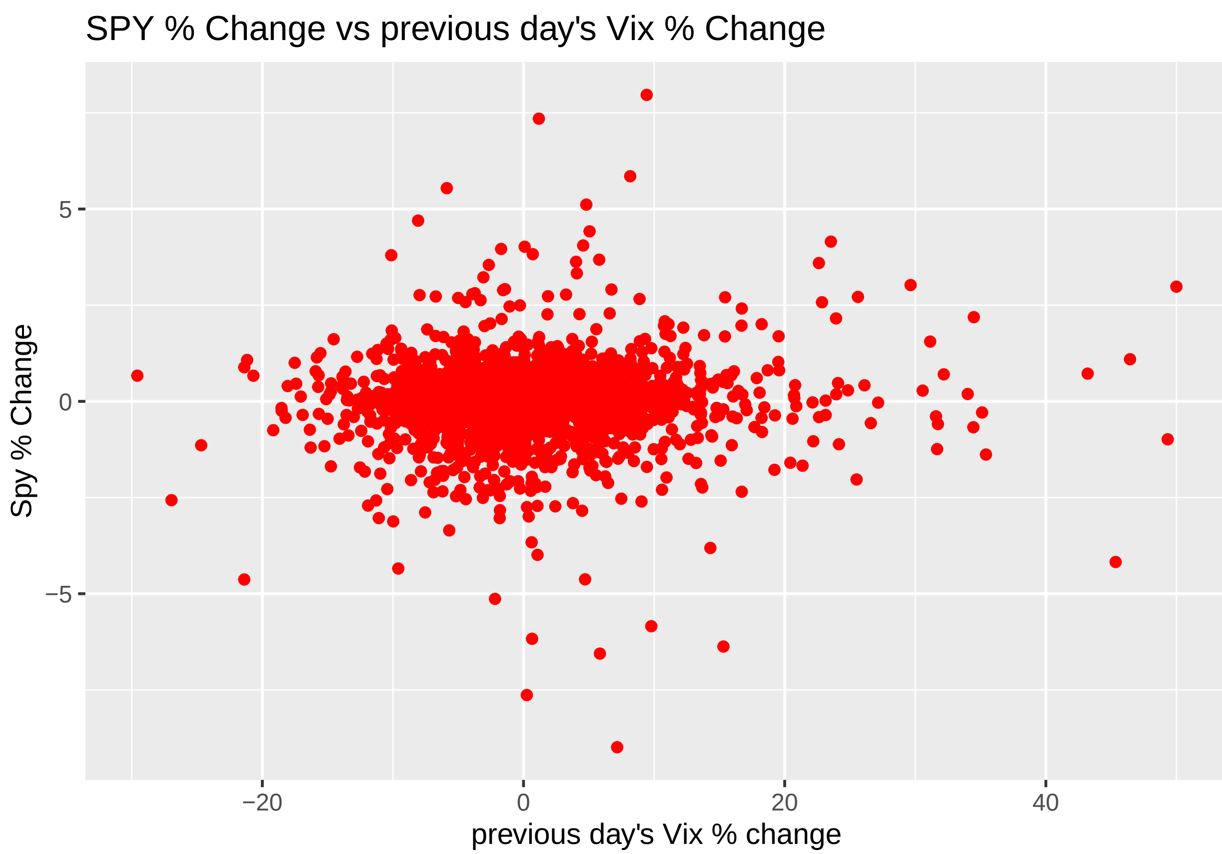
```r
#Plots the relationship between Vix and Spy movements of the same day
ggplot(all_data, aes(x = vix_daychange, y = spy_daychange)) +
  geom_point(color = "red") +
  labs(title = "SPY % Change vs Vix % Change", x = "Vix % change",
       y = "Spy % Change")
```



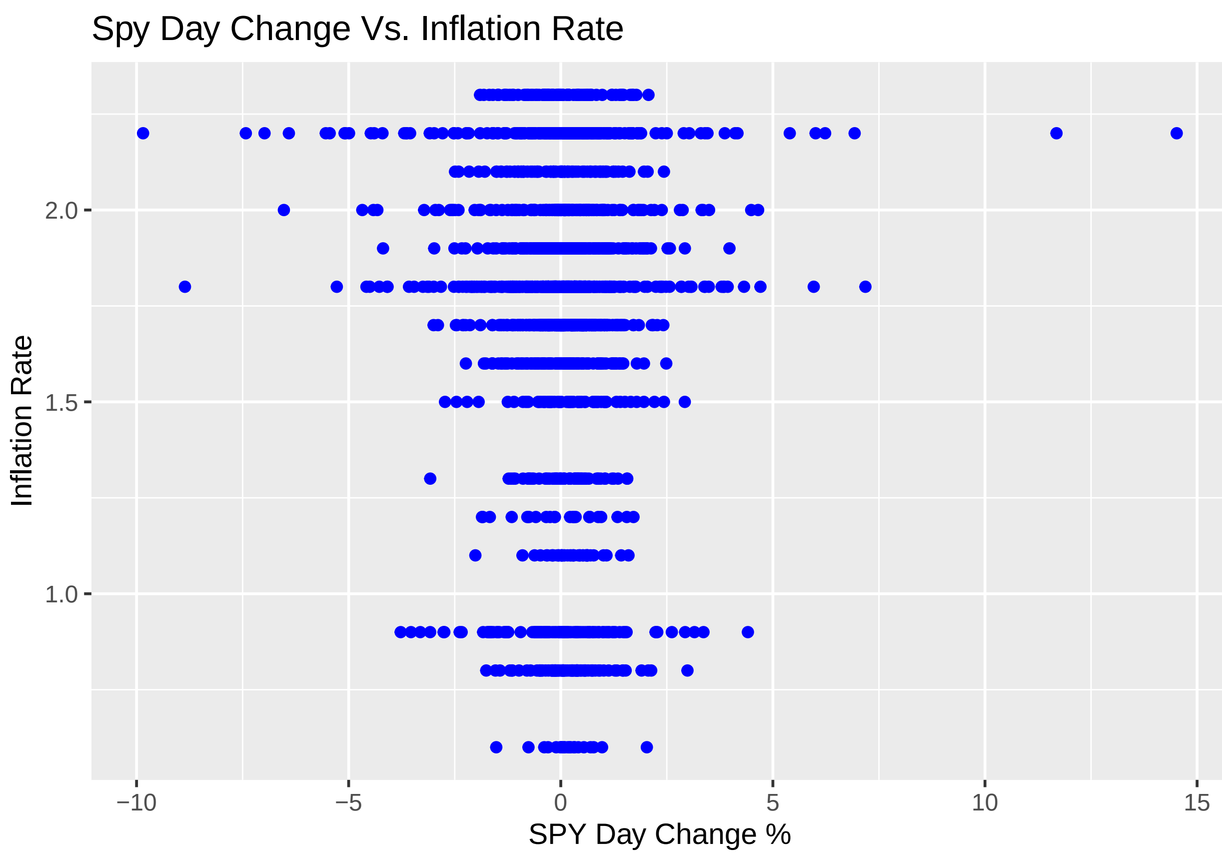

```r
#Plots the relationship between the previous day's vix value change and the
#current days After hours movements
ggplot(all_data, aes(x = lag(vix_daychange), y = AHchange)) +
  geom_point(color = "blue") +
  labs(title = " AH SPY % Change vs previouis day's Vix % Change",
       x = "Previous day's Vix % change", y = "AH Spy % Change")
```



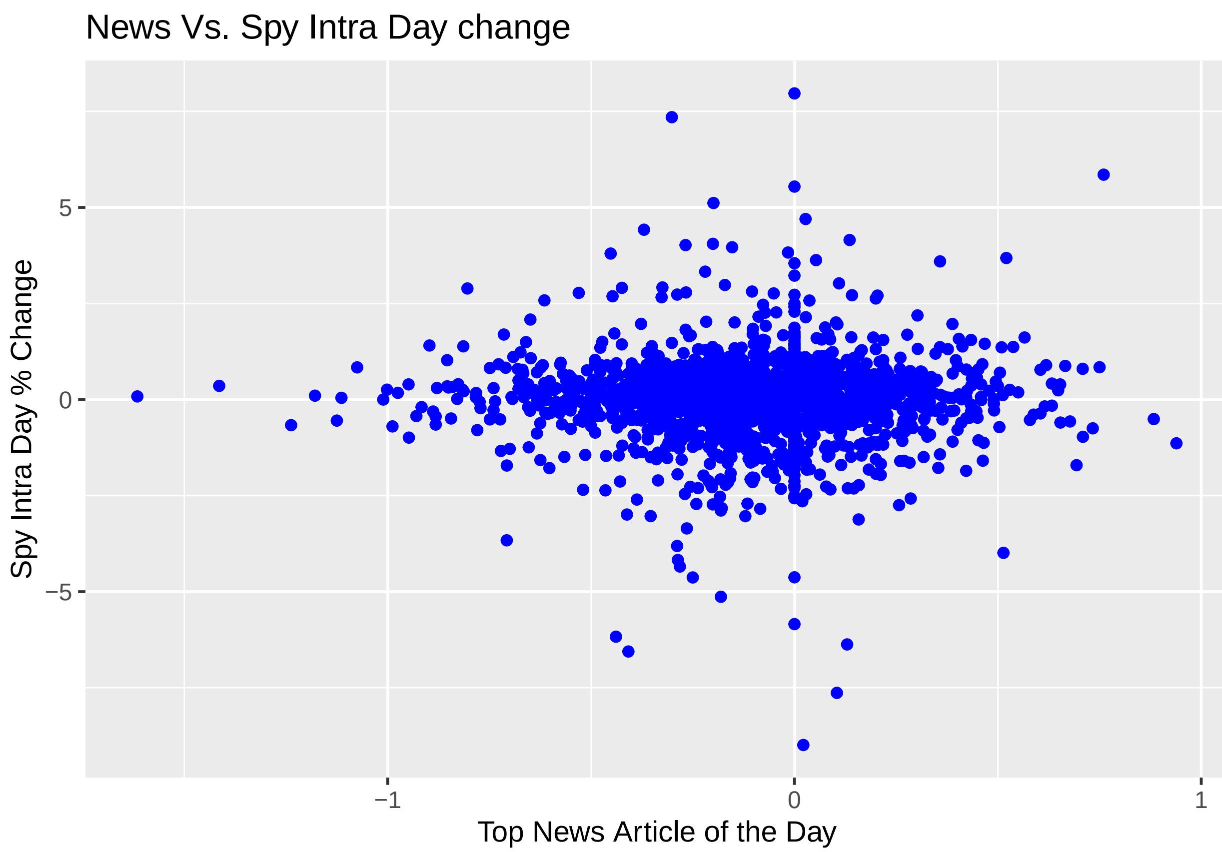

```r
#Plots the relationship between the previous day's vix value change and the
#current days spy value movement
ggplot(all_data, aes(x = lag(vix_daychange), y = intra_daychange)) +
  geom_point(color = "red") +
  labs(title = "SPY % Change vs previous day's Vix % Change",
       x = "previous day's Vix % change", y = "Spy % Change")
```



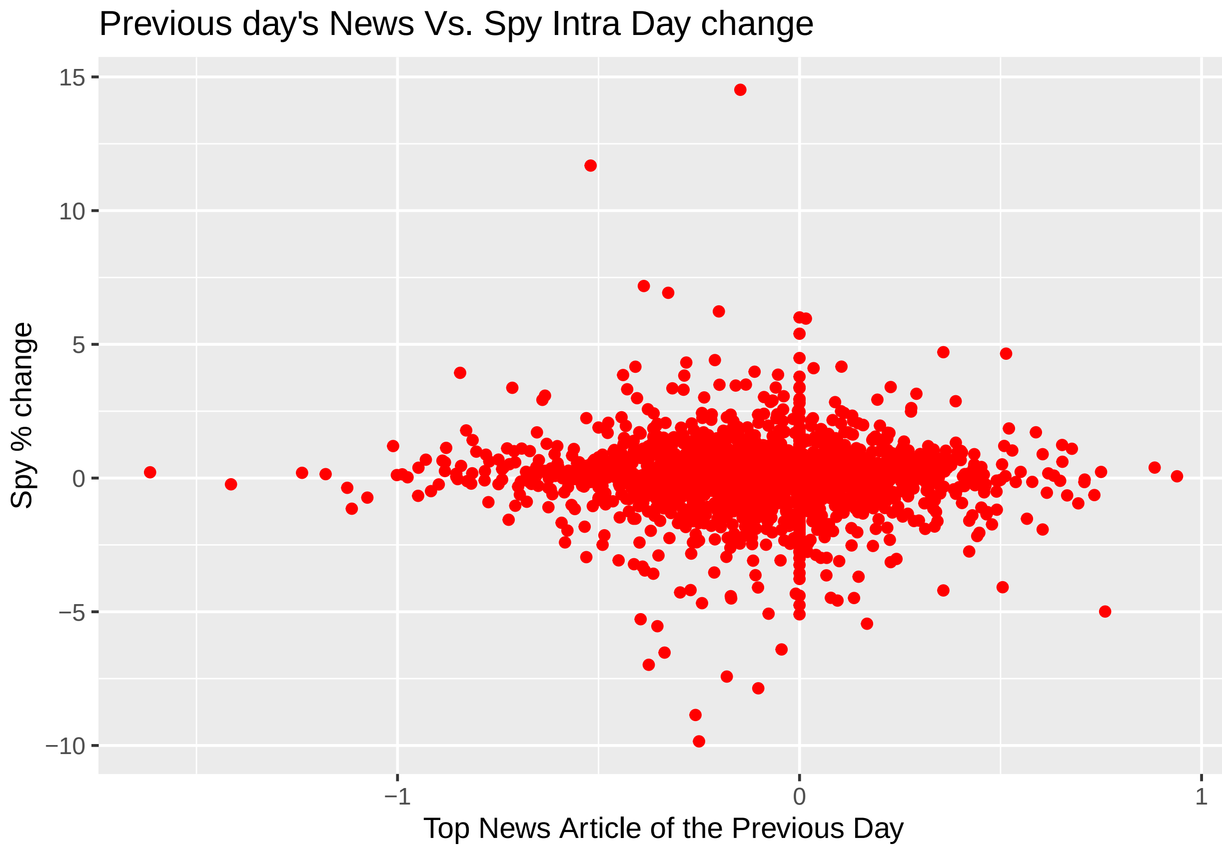

```r
#Plots the relationship between the previous day's vix value change and the
#current days spy value movement
ggplot(all_data, aes(x = spy_daychange, y = Inflation.Rate)) +
  geom_point(color = "blue") +
  labs(title = "Spy Day Change Vs. Inflation Rate",
       x = "SPY Day Change %", y = "Inflation Rate")
```


Spy Day Change Vs. Inflation Rate
Inflation Rate
2.0
1.5
1.0
−10
−5
0
5
10
15
SPY Day Change %

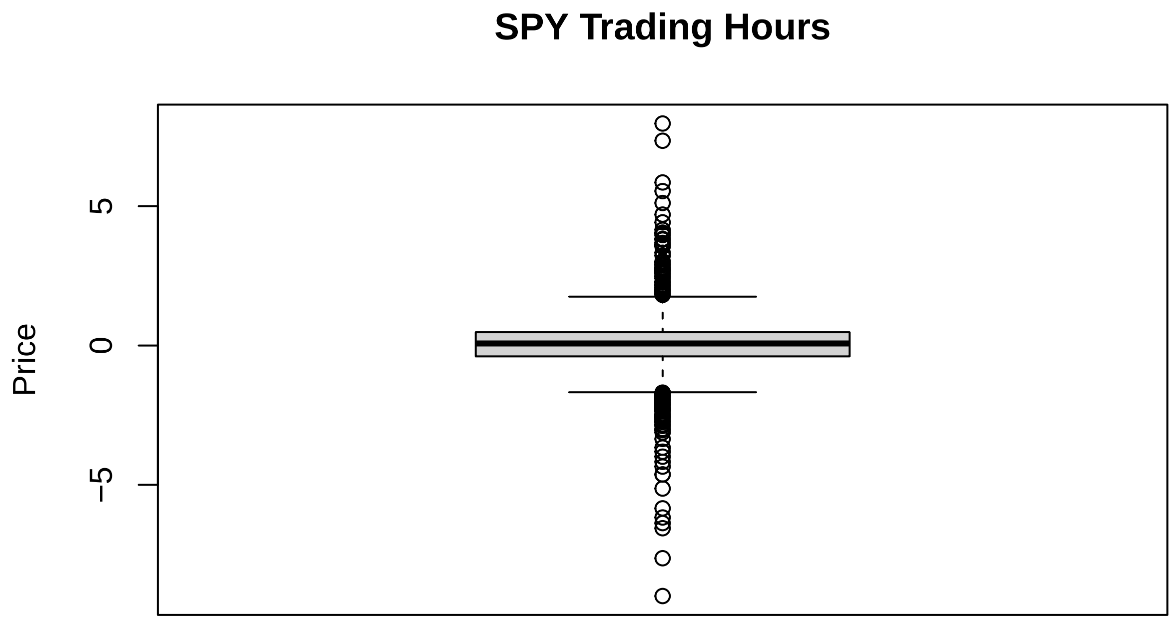
```r
#Creates a scatters plot comparing the top news article vs the change in
#SPY price movements.
ggplot(all_data, aes(x = Top1, y = intra_daychange)) +
  geom_point(color = "blue") +
  labs(title = "News Vs. Spy Intra Day change",
       x = "Top News Article of the Day", y = "Spy Intra Day % Change")
```



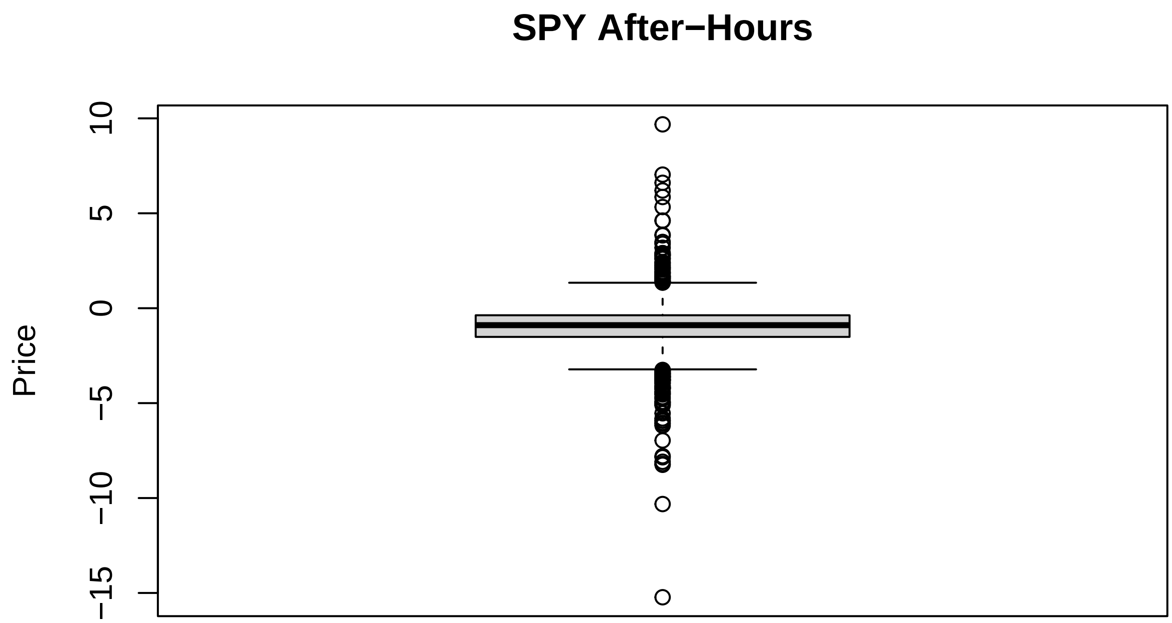

```r
ggplot(all_data, aes(x = lag(Top1), y = spy_daychange)) +
  geom_point(color = "Red") +
  labs(title = "Previous day's News Vs. Spy Intra Day change",
       x = "Top News Article of the Previous Day", y = "Spy % change")
```


Previous day's News Vs. Spy Intra Day change
Spy % change
Top News Article of the Previous Day
15
10
5
0
−5
−10
−1
0
1

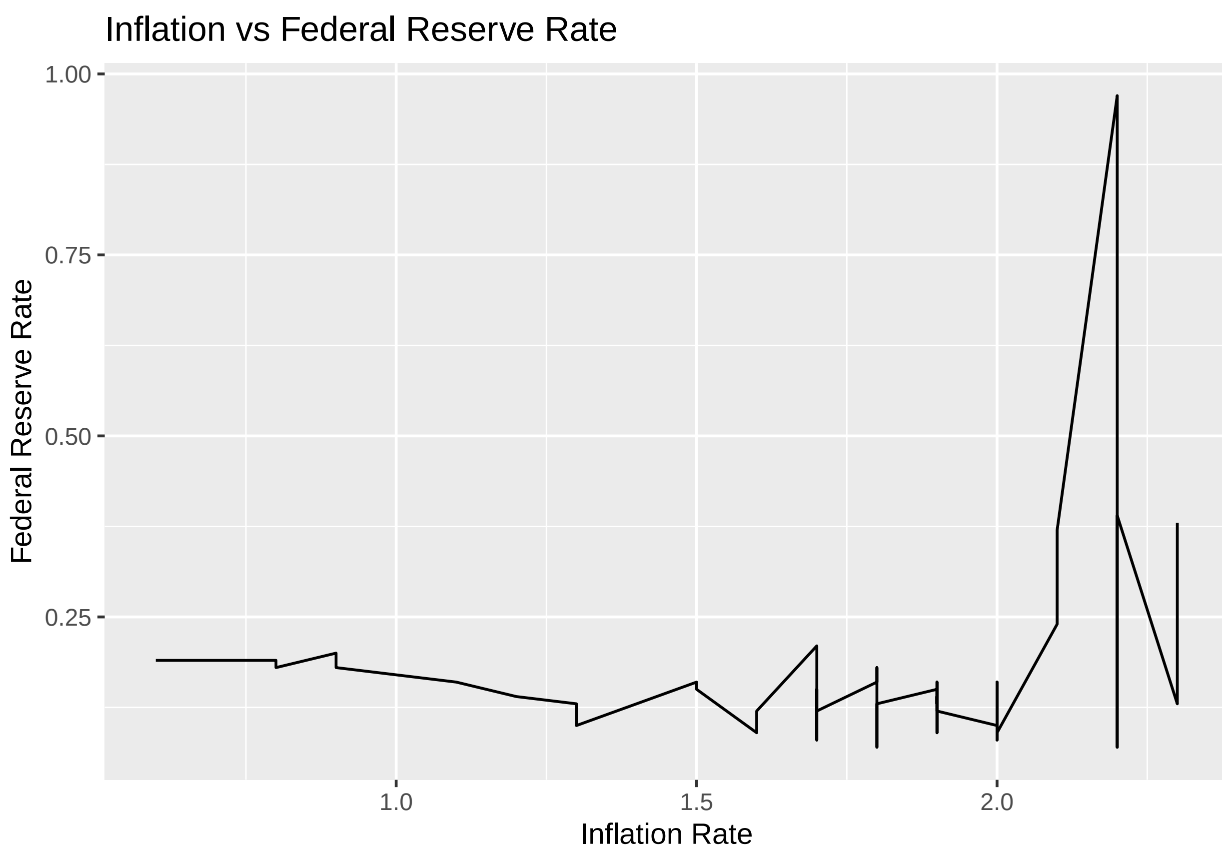
```
boxplot(all_data$intra_daychange, main="SPY Trading Hours", ylab="Price")
```




```
boxplot(all_data$AHchange, main="SPY After-Hours", ylab="Price")
```


SPY After−Hours
Price
10
5
0
−5
−10
−15

```r
ggplot(all_data, aes(Inflation.Rate, Effective.Federal.Funds.Rate)) +
  geom_line() + labs(x = "Inflation Rate", y = "Federal Reserve Rate") +
  ggtitle("Inflation vs Federal Reserve Rate")
```



Analysis

Although only a few variables were graphed, they provided significant insights. Further analysis is necessary to identify patterns and correlations among the other variables to understand how they affect each other. The clearest pattern observed is the linear relationship between the percentage change of SPY's price and the VIX value. This correlation is expected as market participants buy options based on SPY movement, causing the VIX value to increase or decrease oppositely. The next step is to determine if both variables move simultaneously during trading hours or if one moves before the other to predict future movements.

The graph comparing after-hours SPY movements against the day's VIX SPY change showed an unexpected bias towards negative movements. Further data is required to validate this discovery. On the other hand, the graph comparing SPY's percentage change versus the previous day's VIX change showed almost no correlation between the two. This eliminates the notion that past VIX movements can predict future SPY movements.

The sentiment value of the top news articles was compared against the SPY intraday percentage change, but no clear correlation was found. This could be due to the market's indifference to the top news of the day. Future research will sum up the sentiment values to get a better picture of the news and compare it against SPY movements. Getting more market-focused articles could also be a better option.

The box graphs compared intraday movements with after-hours movements. As expected, intraday movements were distributed around zero, while after-hours movements had a bias towards negativity. This clarifies what was seen in the scatter plot and shows that even when intraday movements aren't negative biased, after-hours movements still are.

The Line graph compares reported inflation with the federal reserve rate. However, due to inflation not being a significant concern during the data time period, the graph requires smoothing out, and more data is necessary. Nonetheless, we can observe that during this period, the reserve was reactive to inflation and did not raise rates until inflation was above 2%. This trend was seen again last year, indicating that the reserve's response to inflation remains consistent.

Limitations

The primary limitation of this research is the lack of sufficient data. Data from earlier periods is only available from the early 2000s when computers became more mainstream and data was digitized. Furthermore, obtaining newer data from the past few years can be challenging and may require payment. These time periods are necessary to comprehend how the market behaves in different situations. The current data only includes a recession and its recovery, making it too specific to apply to the present environment.

Another limitation, related to the first, is that it may not be possible to create an all-inclusive model due to the ever-changing economy and market. While it may be possible to find a few predictors that are helpful, it will never be possible to have data from every possible scenario. Thus, the model should be developed with the aim of using it as an edge rather than a comprehensive SPY price predictor.

The final limitation is the time and knowledge required to conduct this research. These two factors are interrelated, as a lack of knowledge can lead to additional time being spent. Even with the lack of it, the data obtained is highly valuable and should be analyzed thoroughly from every perspective to identify patterns. This process can be accelerated by using advanced data science resources and tools.

Conclussion

In summary, this research appears to have significant value and warrants further investigation. However, it will require a considerable amount of time and expertise to fully comprehend all the data and their interdependencies. While some intriguing findings were uncovered through the analysis of a few variables, such as after-hours price movements and the historical responsiveness of the Federal Reserve to inflation, additional research may reveal even more surprising insights.

In order to build a model, it will be necessary to analyze each variable individually and determine its correlation with the overall market. Although this will require a significant amount of additional time, it may result in a model that is accurate enough to provide a trading edge/advantage. However, due to the size and complexity of this task, I will need to acquire more knowledge and skills related to modeling in order to successfully undertake it.

Do you plan on incorporating any machine learning techniques to answer your

research questions? Explain.

Yes, I plan to leverage machine learning techniques to answer my research questions. To accomplish this, I will be utilizing R packages to perform the necessary computations and testing for the four possible dependent variables and a range of independent variables. This approach will significantly reduce the amount of time and effort required to model each group of variables and compare the results.

By leveraging machine learning, I can quickly identify the best modeling options and make more informed decisions about my research questions. Ultimately, this approach will allow me to more efficiently analyze and interpret the data, leading to more accurate and meaningful results.

Future Steps

To create an accurate model, I will need to acquire additional knowledge on advanced regression testing with multiple variables/predictors. Additionally, incorporating machine learning techniques could streamline the process of model improvement and reduce the time spent on trial and error.

References

Aaron7Sun. N/A. “Stock News.” Kaggle. https://www.kaggle.com/datasets/aaron7sun/stocknews.

Besomi, Jonathan. 2021. “VIX Historical Values.” https://www.kaggle.com/competitions/optiver-realized-volatility-prediction/discussion/249660.

Community, Kaggle. 2021. “SPY ETF Historical Data.” https://www.kaggle.com/datasets/411357/spy-us-txt.

Reserve, Federal. 2021. “Interest Rates.” https://www.kaggle.com/datasets/federalreserve/interest-rates.