Exploring Stock Market Trading Strategies With R

Lucas Gibson

2/5/2022

Introduction:

The age old question every investor wants to answer is it possible to out-preform the market. For our purposes are baseline will be the SPY ETF which is Standard & Poor's(S&P 500) index of stocks in the S&P. The the S&P is considered a good barometer for baseline returns in the market.

Why:

There are thousands of different strategies from which to choose, from very basic technical analysis on moon cycles. My interest in this topic comes from day trading several different markets over the course of several years. I've always wanted look at certain market concepts via programming and statistics.In essence, the questions I want to answer is the market intrinsically random/zero sum game.

Methods:

Firstly I want to simulate a "random" trading strategy via flipping a coin. Next I want to examine the returns of a basic technical analysis strategy. Lastly I want see how the returns of different time periods of buy and hold strategies compare.

Challenges:

I expect the challenges of this project to come from my beginner knowledge of R, that and many have tried and failed to solve such this question. Another challenge of note was conversion between xts and dataframe objects. This is because imported data is automatically kept in xts formats.

Data:

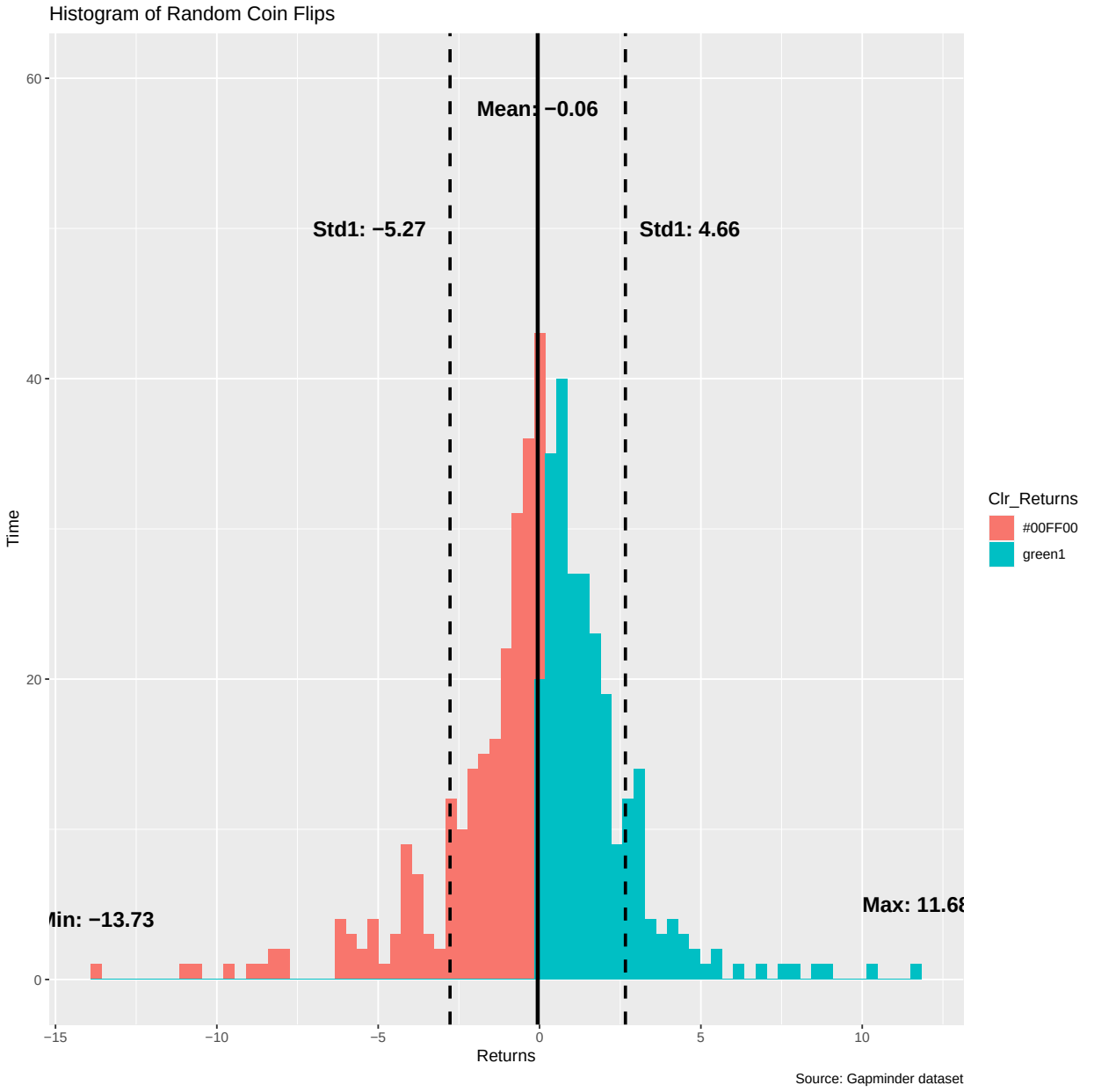
My data was gathered via the quantmod package which retrieves data for stock symbols via Yahoo finance. I simply filtered out any NA data.

```
## [1] "spy"
```

Coin Flip Strat

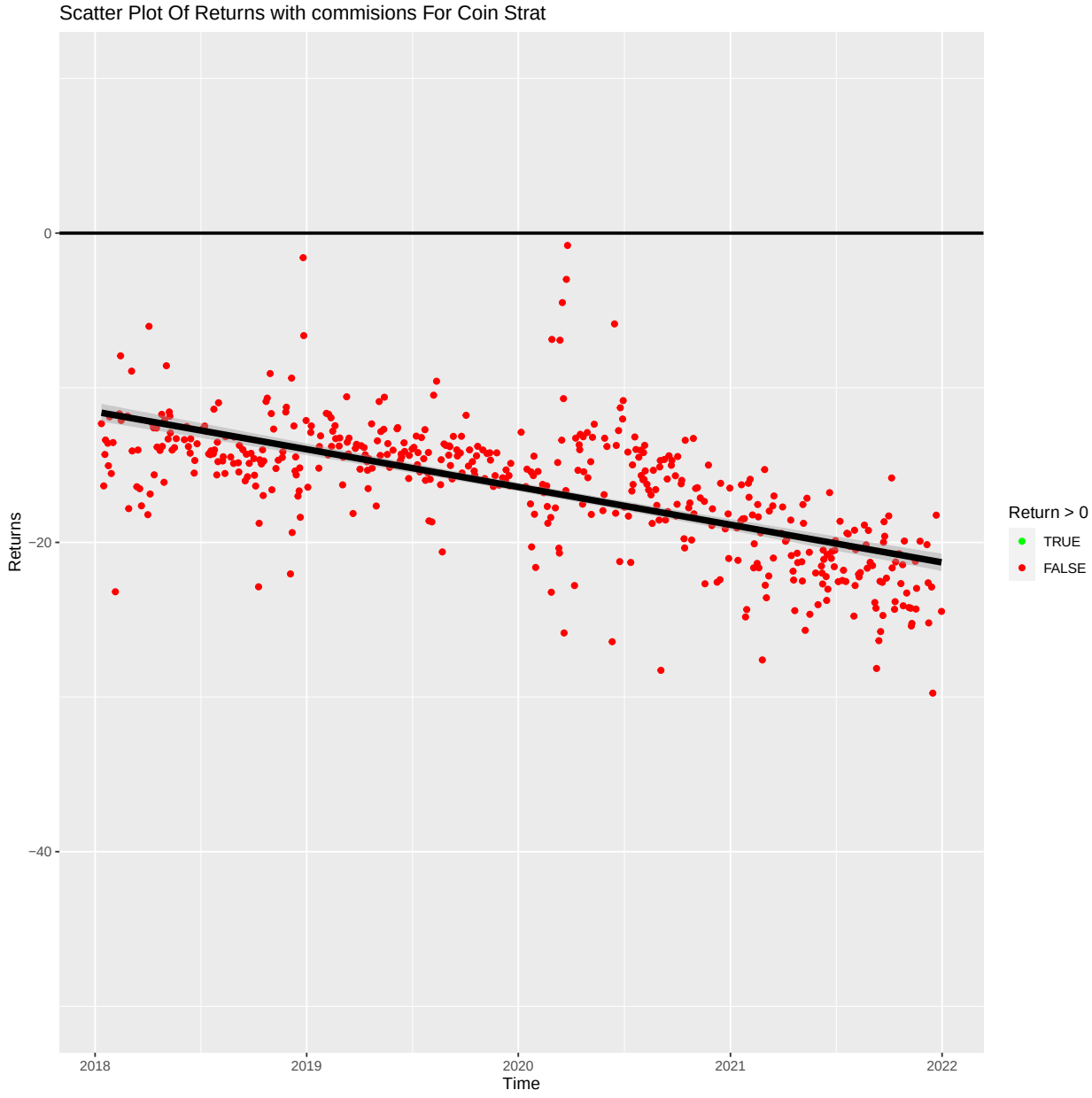
This strategies works by flipping coin a set number of times. If it is heads we buy long at the open, if it is tails we go golf. The stock is sold at the close of every day. Then repeat the process the for the total data set.

This is a histogram of returns of a single instance of the Coin Flip Strat. The plot is normally distributed with very little variance from test to test. This suggests that this strategy using this method of trading is zero sum at best



Coin Flip Strat

The first graph displays a scatter plot of returns for the random coin strat that includes commission. People often say that trading stocks is a zero sum game. However, with the inclusion of commisions it becomes apparent that obtaining alpha is required just to break even.



Graph of Equity of Coin Flip Strat

Like the other graphs this line chart represents one instance of the random coin flip strat but shows how a starting equity of $1000 fairs

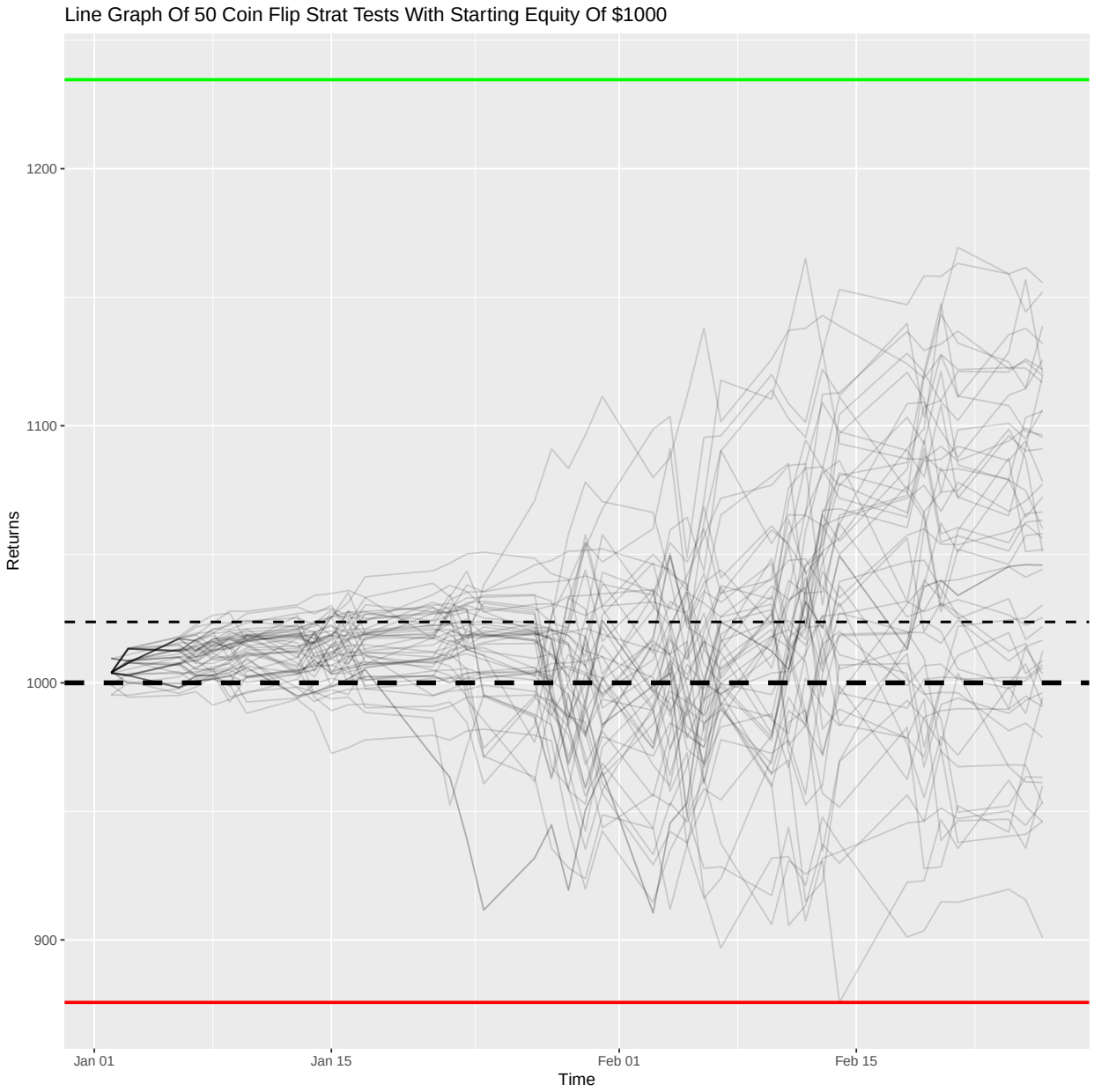
Multiple Coin Flip Strat Tests (Random Walk Outcomes)

This plot shows a starting position of $1000. Each line graphs the coin flip strat's returns over the length of the data set. The thick dashed black line represents the starting equity, while thinner dashed lined represents the returns mean.

What does all this mean? This was my attempt to create a random walk scenario using real SPY prices. The outcome of these multiple tests shows the same normal distribution pattern as the histogram in the first slide.

Conclusion: Over time a completely random strategy without commissions will produce negligible returns. The Coin Flip Strat will serve as a baseline. It is also possible to go on a "run"/"draw down" with a strategy that is random and will eventually revert to $0.

```
## [1] "spy"
```

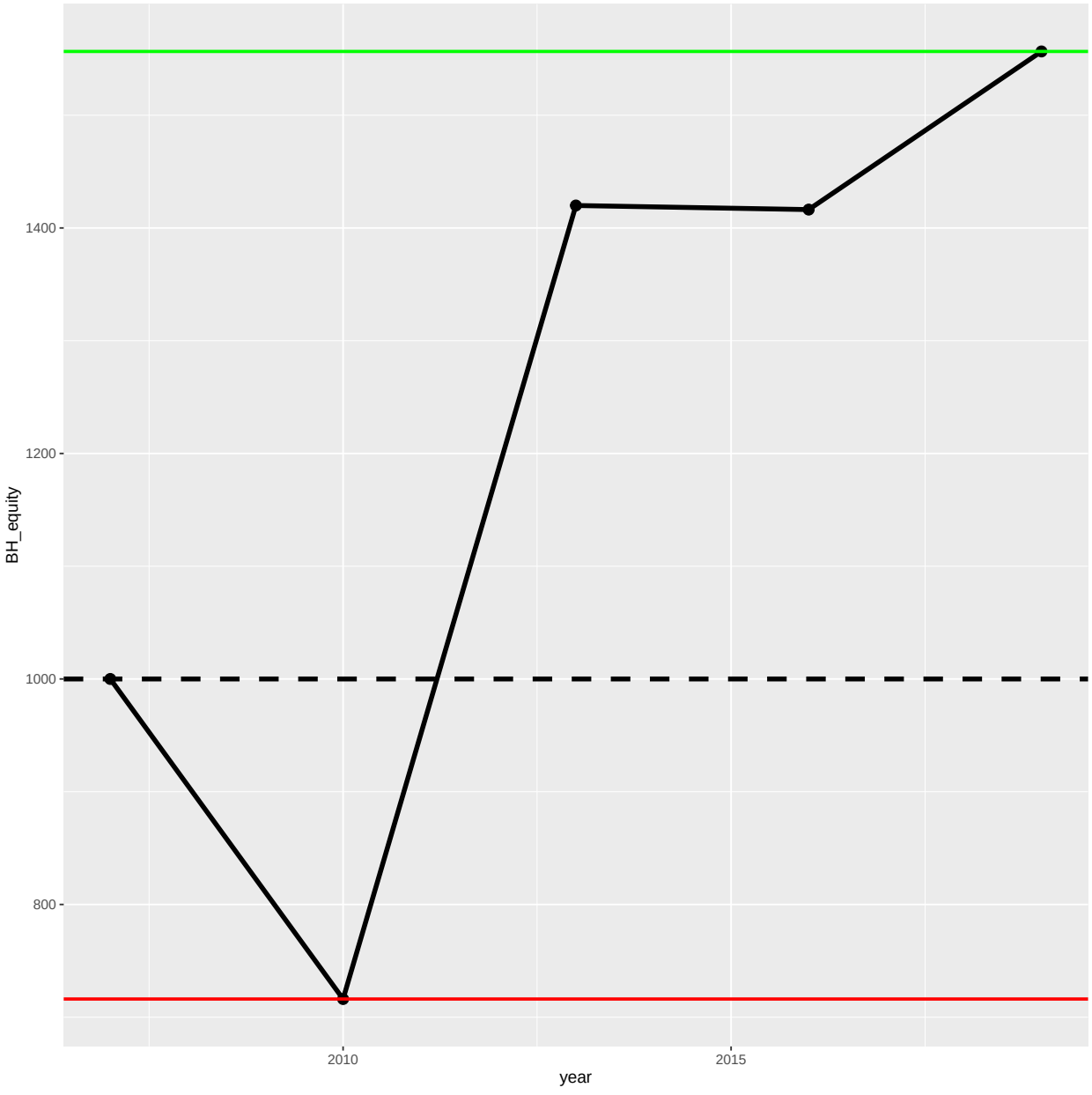


Buy and Hold Strategy

Next I want to test the returns of buy and hold strategy. This is a buy only strategy that will be applied all available data(2007 -> current) of the S&P 500 stocks. The signal to buy will be to purchase the price at the open and then sell at the close exactly 3 years afterward.

Results: The results are somewhat interesting. The market has been bullish since the recovery of the 2008 crash. The drawdown in the beginning suggests incorporates the crash. There are only 5 different data points which skew the results but yahoo finance does not have any data circa 2007. Never the less, your initial investment of $1000 dollars would have returned over 50% with an ending balance in 2019 of 1156

```
## [1] "SPY"
```



Buy and Hold Strategy

Baseline Strat: Lastly, if an investor were to just buy at the beginning of the data set (2007) and held, they would be looking at a near 200% return. A $1000 initial investment would net around 3k profit. Therefore, I would consider this buy and hold strategy to be the baseline with which to compare other strategies.

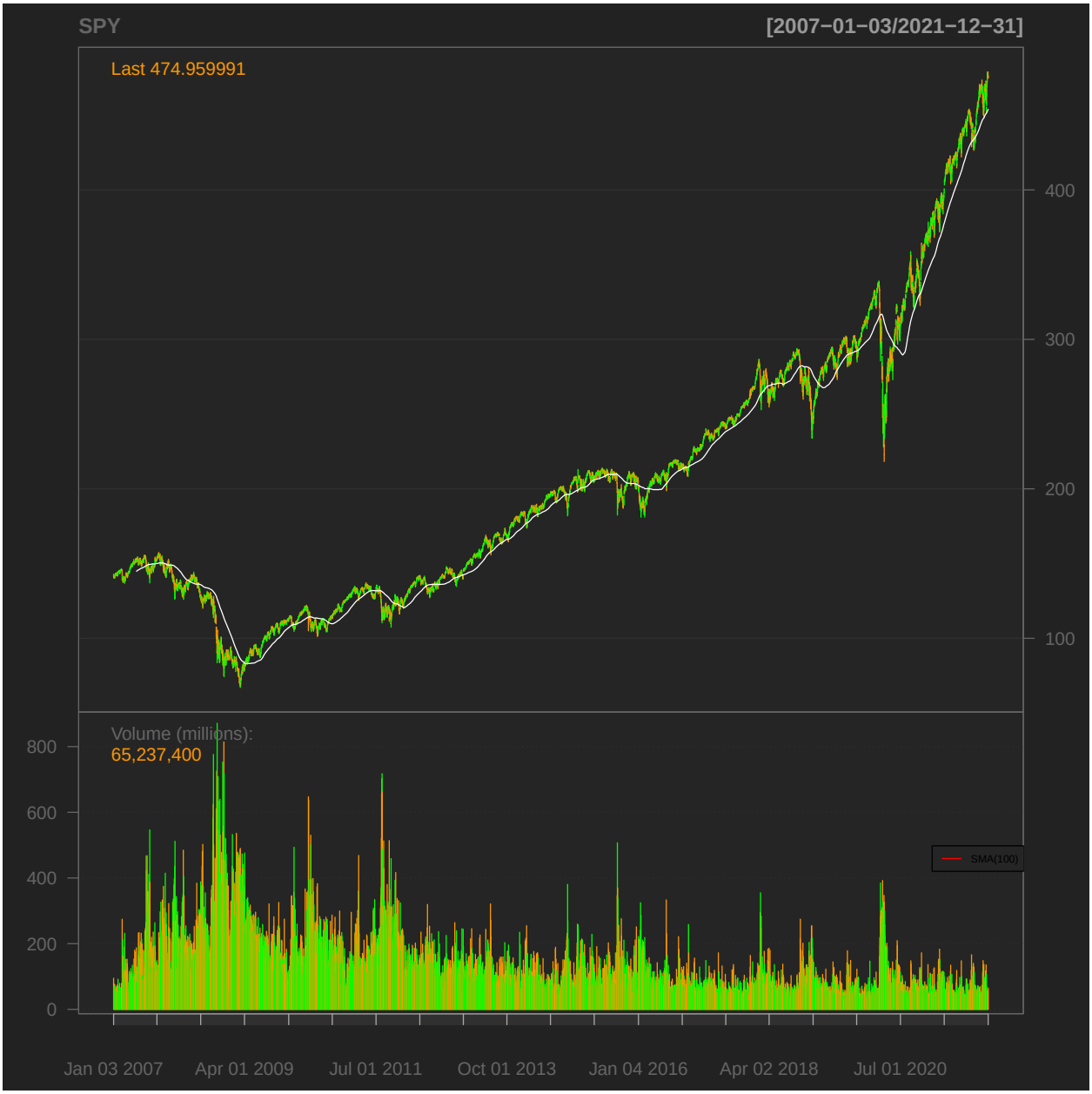

```
##                           daily.returns
## Annualized Return                0.0781
## Annualized Std Dev               0.2024
## Annualized Sharpe (Rf=0%)        0.3861
```

Technical Analysis Indicators: SMA

$SMA_n = \frac{Sum_n(Close)}{n}$

These next 2 graphs graphs represent 2 different types of techincal indicators plotted on top of price. This is a 100 day simple moving average plot on a bar chart of the SPY from 2007-2021. Notice how the average shows a smooth portrayal of the price chart.

```
## [1] "SPY"
```



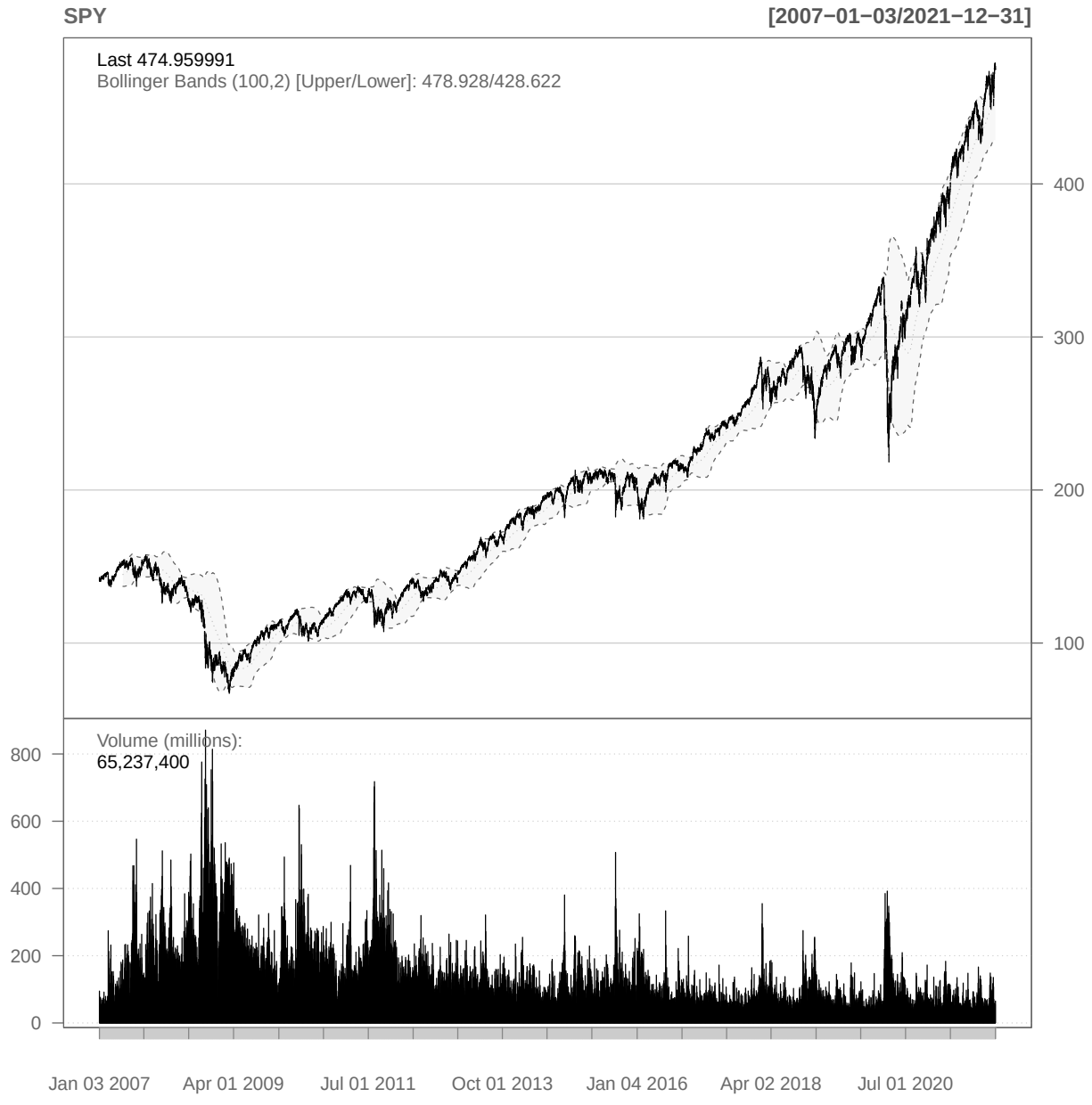
Technical Analysis Indicators: Bollinger Bands

Bolinger Bands: These bands are in essence moving averages used to identify abnormal volatility. If price moves outside either of the outbands they are 2 standard deviations outside

$$Upper\ Band = SMA_{20}\left(Close\right) + 2 * STDEV_{20}\left(Close\right)$$

$$Middle\ Band\ = SMA_{20}\left(Close\right)$$

$$Lower\ Band = SMA_{20}\left(Close\right) - 2 * STDEV_{20}\left(Close\right)$$



Technical Analysis Strat: SMA Price Crossover

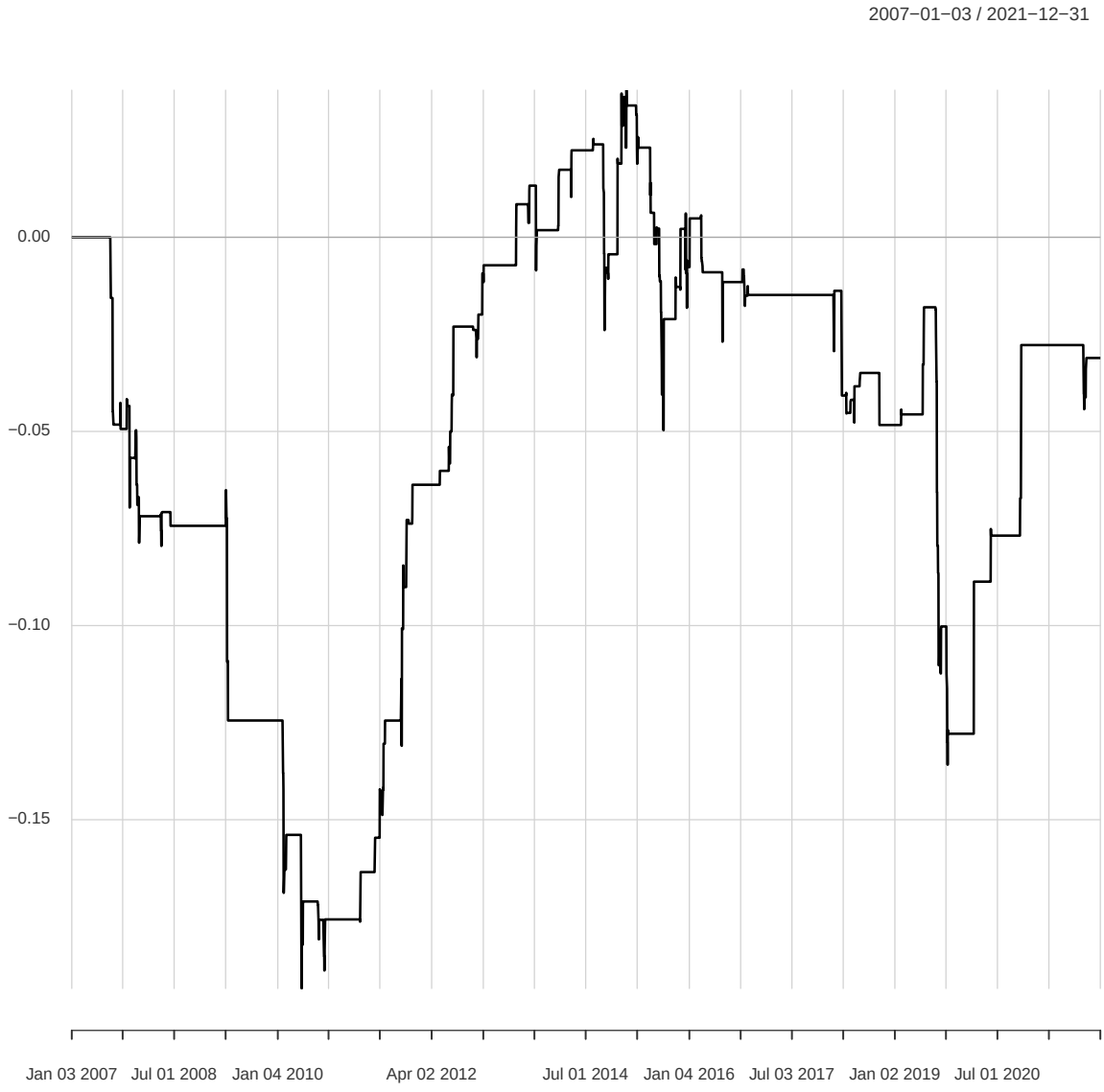
This is a very highly debated topic. Does technical analysis outperform a long term investment.

In order to test the same time period as the buy and hold strat I will be using a 50 and 100 average. The buy signal will occur when previous days close was < SMA then the next days current price close becomes greater than the SMA

SMA Price Crossover Strat:

$$Buy : PreviousDaysClose < SMA \rightarrow CurrentDaysClose > SMA$$

The strategy works by determine if the previous days close was less then the SMA. If true and the next the next days close is greater than the SMA then buy. A buy is represented by a column of 1's, 0's if nothing do and -1 to sell the position.Lag method is used to executes command the day after signal.



```
##                            daily.returns
## Annualized Return                -0.0021
## Annualized Std Dev                0.0439
## Annualized Sharpe (Rf=0%)        -0.0480
```

Summary

In summary, the buy and hold strategies provided for the best returns with the least commissions and least amount of effort. The buy and hold strategy should be considered the baseline strategy to beat. In comparison the technical analysis strategy returns oscilatted around 0 over the same time period. These results match with a completely random coin toss strategy that is normally distributed around a $0 return.

Out of these above mentioned methods the market appear to be upwardly skewed and a buy and hold strat is most advantageous. Whereas, active strategies involving price crossing moving averages are similar to a coin flip. However, more data is needed to accurately conclude this hypothesis.

Appendix

QuantMod: https://www.quantmod.com/

```
library(quantmod)
library(plotly)
library(tidyverse)
library(tidyquant)
library(dplyr)
library(lubridate)
library(forcats)
library(PerformanceAnalytics)


start <- as.Date("2018-01-12")      #Starting Date For Stock Data
end <- as.Date("2022-01-01")        #Ending Date For Stock Data

#Download Symbols
getSymbols("spy", src = "yahoo",from = start,to = end)
```

```
## [1] "spy"
```

```
#Get dates
dates = index(SPY)
#Convert from Xts to Dataframe
SPY = as.data.frame(SPY)
#Add column with dates to Dataframe
SPY = cbind(date = dates, SPY)

#Calculate return Col
SPY = SPY %>%
  mutate(return = SPY.Close - SPY.Open)

################################################################

#Coin Flip Strat

#This strategies works by flipping coin a set number of times.
#If it is heads we buy long at the open, if it is tails we go golf
#We sell at the close of every day. Then repeat the process the next day

#Coin_Fun = function(){

#Flip a 2 sided coin 1008 times
C_Flips = rbinom(1000, 1, .5)

#Add colum for flips
SPY = SPY %>%
  add_column(C_Flips = C_Flips)

#Filter for all flips that were heads
C_Flip_Strat = SPY %>%
  filter(C_Flips == 1)

#Extra variable of just heads column for returns
Heads_Returns = C_Flip_Strat %>%
```

```
  select(return)


x = C_Flip_Strat$date
y = C_Flip_Strat$return
capital = 1000



#Convert to Xts in case Quant mod package is needed
Flip_Strat = as.xts(C_Flip_Strat)



#Determine commission by multiplying the open/the price you bought at by 5%
Commish = C_Flip_Strat$SPY.Open * .05
#Subtract the commission from the return
Ret_Comm = y - Commish

#Take capital divide by the buy price, round to even number of shares
shares = round(capital / C_Flip_Strat$SPY.Open, digits = 0)
#Multiply the number of shares by the profit or loss in returns then add them up
equity = cumsum(shares * Heads_Returns)+capital



#Determine the x and y coordinates for the labels
annotations <- data.frame(
  x = c(
    round(min(C_Flip_Strat$return), 2),
    round((mean(y) - sd(y)), 2) - 2.5,
    round(mean(C_Flip_Strat$return), 2),
    round((mean(y) + sd(y)), 2) + 2,
    round(max(C_Flip_Strat$return), 2)
  ),
  y = c(4, 50, 58, 50, 5),
  label = c("Min:", "Std1:", "Mean:", "Std1:", "Max:")
)

#Color green if returns > 0 else color red
Clr_Returns = ifelse(y >= 0, "green1", "#00FF00")

#Plot histogram of Non commission based returns for Coin Flip Strat, No position
ggplot(C_Flip_Strat, aes(x = y)) +
  geom_histogram(aes(fill = Clr_Returns), bins = 75) +
  geom_vline(aes(xintercept = mean(y)), color = "#000000", size = 1.25) +
  geom_vline( aes(xintercept = mean(y) + sd(y)),
    color = "#000000",size = 1,linetype = "dashed") +
  geom_vline( aes(xintercept = mean(y) - sd(y)),
    color = "#000000", size = 1,linetype = "dashed") +
  ylim(c(0, 60)) +
  geom_text(data = annotations,aes(x = x,y = y,label = paste(label, x)),
    size = 5, fontface = "bold") +
```

```
    labs(  title = "Histogram of Random Coin Flips",
    caption = "Source: Gapminder dataset", x = "Returns", y = "Time")
```

```
#Plot of commission based returns for Coin Flip Strat, No position
ggplot(C_Flip_Strat, aes(x = x, y = Ret_Comm, colour = Ret_Comm > 0)) +
  scale_colour_manual(name = 'Return > 0', values = setNames(c('green', 'red'),
  c(T, F))) + xlab('Time') + ylab('Returns w Commission') +
  geom_point() + geom_hline(yintercept = 0, size = 1) +
  geom_smooth(method = "lm", colour = "black", size = 2) +ylim(c(-50, 10))
```

```
#Single plot of one 1 coin flip with 1000 starting capital
plot(equity$return,type = "l",  col = "black",
  main = "Coin Flip Strat Equity Graph (Capital = $1000)",
  xlab = "Days", ylab = "Value")
abline(h = capital, col = "black", lwd = 3)
abline(h = max(equity$return), col = "green", lwd = 1)
abline(h = min(equity$return), col = "red", lwd = 1)
legend("right", legend=c("Equity High", "Equity Low"),
       col=c("green", "red"), lty=1:2, cex=0.8)
```

```
RandTests = cbind(equity)
start <- as.Date("2020-01-01")
end <- as.Date("2020-06-01")
getSymbols("spy", src = "yahoo",from = start,to = end)
```

```
## [1] "spy"
```

```
for (i in 1:75){
  capital = 1000
  dates = index(SPY)
  SPY2 = as.data.frame(SPY)
  SPY2 = cbind(date = dates, SPY2)
  SPY2 = SPY2 %>%
    mutate(return = SPY.Close - SPY.Open)

  C_Flips = data.frame(rbinom(103, 1, .5))



  SPY2 = SPY2 %>%
    add_column(Flips = C_Flips)

  C_Flip_Strat = SPY2 %>%
    filter(Flips == 1)


  shares = round(capital / C_Flip_Strat$SPY.Open, digits = 0)
  equity_test = cumsum(shares * C_Flip_Strat$return)+capital

  RandTests = qpcR:::cbind.na(RandTests,equity_test)

}
RandTests <- subset (RandTests, select = -return)
equityHigh = max(RandTests, na.rm = TRUE)
equityLow = min(RandTests, na.rm = TRUE)
```

```r
#Mean of all tests
test_mean = mean(colMeans(RandTests, na.rm = TRUE))


RandTests = qpcR:::cbind.na(date = dates, RandTests)
RandTests =  RandTests %>% drop_na()




a = .15

ggplot(RandTests, aes(x = date))+
  geom_line(aes(y = equity_test.1), alpha = a)+
  geom_line(aes(y = equity_test.2), alpha = a)+
  geom_line(aes(y = equity_test.3), alpha = a)+
  geom_line(aes(y = equity_test.4), alpha = a)+
  geom_line(aes(y = equity_test.5), alpha = a)+
  geom_line(aes(y = equity_test.6), alpha = a)+
  geom_line(aes(y = equity_test.7), alpha = a)+
  geom_line(aes(y = equity_test.8), alpha = a)+
  geom_line(aes(y = equity_test.9), alpha = a)+
  geom_line(aes(y = equity_test.10), alpha = a)+
  geom_line(aes(y = equity_test.11), alpha = a)+
  geom_line(aes(y = equity_test.12), alpha = a)+
  geom_line(aes(y = equity_test.13), alpha = a)+
  geom_line(aes(y = equity_test.14), alpha = a)+
  geom_line(aes(y = equity_test.15), alpha = a)+
  geom_line(aes(y = equity_test.16), alpha = a)+
  geom_line(aes(y = equity_test.17), alpha = a)+
  geom_line(aes(y = equity_test.18), alpha = a)+
  geom_line(aes(y = equity_test.19), alpha = a)+
  geom_line(aes(y = equity_test.20), alpha = a)+
  geom_line(aes(y = equity_test.21), alpha = a)+
  geom_line(aes(y = equity_test.22), alpha = a)+
  geom_line(aes(y = equity_test.23), alpha = a)+
  geom_line(aes(y = equity_test.24), alpha = a)+
  geom_line(aes(y = equity_test.25), alpha = a)+
  geom_line(aes(y = equity_test.26), alpha = a)+
  geom_line(aes(y = equity_test.27), alpha = a)+
  geom_line(aes(y = equity_test.28), alpha = a)+
  geom_line(aes(y = equity_test.29), alpha = a)+
  geom_line(aes(y = equity_test.30), alpha = a)+
  geom_line(aes(y = equity_test.31), alpha = a)+
  geom_line(aes(y = equity_test.32), alpha = a)+
  geom_line(aes(y = equity_test.33), alpha = a)+
  geom_line(aes(y = equity_test.34), alpha = a)+
  geom_line(aes(y = equity_test.35), alpha = a)+
  geom_line(aes(y = equity_test.36), alpha = a)+
  geom_line(aes(y = equity_test.37), alpha = a)+
  geom_line(aes(y = equity_test.38), alpha = a)+
```

```
  geom_line(aes(y = equity_test.39), alpha = a)+
  geom_line(aes(y = equity_test.40), alpha = a)+
  geom_line(aes(y = equity_test.40), alpha = a)+
  geom_line(aes(y = equity_test.41), alpha = a)+
  geom_line(aes(y = equity_test.42), alpha = a)+
  geom_line(aes(y = equity_test.43), alpha = a)+
  geom_line(aes(y = equity_test.44), alpha = a)+
  geom_line(aes(y = equity_test.45), alpha = a)+
  geom_line(aes(y = equity_test.46), alpha = a)+
  geom_line(aes(y = equity_test.47), alpha = a)+
  geom_line(aes(y = equity_test.48), alpha = a)+
  geom_line(aes(y = equity_test.49), alpha = a)+
  geom_line(aes(y = equity_test.50), alpha = a)+
  geom_hline(yintercept = capital, size =1.5, color ="black", linetype = "dashed")+
  geom_hline(yintercept = equityHigh, size =1, color ="green", linetype = "solid")+
  geom_hline(yintercept = test_mean, size =.8, color ="black", linetype = "dashed")+
  geom_hline(yintercept = equityLow, size =1, color ="red", linetype = "solid")
```

```
#Talking points For slides
# I found working with dates and random amount of rows from the coin flip challenging
# as it created jagged arrays filled with N/A values
# This made running mean, min maxfunctions difficult as well,
# as dates are treated differently.#Also I found it challenging deal
#with variables already stored in the enviroment.
#Often times testing solutions in the console would show errors
#because the dataframe was still in system




#Monte Carlo simualation, zero sum game then add commssions etc

#rm(list = ls())

#############################################################################
#Buy and Hold Strat
stock_list = ("SPY")
getSymbols(stock_list,warnings = FALSE)
```

```
## [1] "SPY"
```

```
Stock = to.yearly(SPY)

#Convert from xts to dataframe
df1 = data.frame(Date = index(Stock), coredata(Stock))

df1 = df1 %>%
  add_column(InvPeriod = seq(NROW(df1)))




df2 = df1%>%
  filter(InvPeriod == 1 | InvPeriod == 4 | InvPeriod == 7 | InvPeriod == 10
```

```
         | InvPeriod == 13 )%>%
  select(Date, SPY.Open, SPY.Adjusted, InvPeriod)

#3 year returns
capital = 1000
shares = round(capital / df2$SPY.Open, digits = 0)

shares = data.frame(shares)

Returns2 = 0

#Calulates return for a buy and hold every 3 years, buy at open of first
#year sell at adjusted close 3 years later
for (i in 2:(NROW(df2)+2)) {
  Returns2 =  rbind.data.frame(Returns2, (df2[i,3] - df2[i-1,2]))
}

c1 = colnames(Returns2)

Returns2 = rename(Returns2, BH_Returns = c1 )

BHReturn = Returns2 %>%
  filter(!is.na(BH_Returns))

BHReturn = BHReturn[2:5,]

BHReturn = as.data.frame(BHReturn)
BH_equity = 0

for (i in 1:NROW(shares)) {
  BH_equity = rbind(BH_equity,shares[i,1] * BHReturn[i,1] +capital)
}

BH_equity[1,1] = capital

BH_equity = BH_equity[1:5,1]

BH_equity = cbind.data.frame(BH_equity,
                             year = seq(from = 2007, to = 2019, by=3))




BH_Hold_Plot = ggplot(BH_equity, aes(x = year))+
  geom_point(aes(y = BH_equity), size = 3)+
  geom_line(aes(y = BH_equity), size = 1.5)+
  geom_hline(yintercept = capital, size =1.5, color ="black", linetype = "dashed")+
  geom_hline(yintercept = max(BH_equity$BH_equity), size =1, color ="green", linetype = "solid")+
  geom_hline(yintercept = min(BH_equity$BH_equity), size =1, color ="red", linetype = "solid")

ggplotly(BH_Hold_Plot)
```

```
rm(list = ls())
```

```
#############################################################################

#$$
 # SMA_n=\frac{Sum_n\left( Close \right)}{n}
#$$

#Technical Indicators


#Moving Averages
#These are line overlays that smooth price data to identify trend
getSymbols("SPY",src="yahoo",from="2020-01-01",to="2021-01-01")
```

```
## [1] "SPY"
```

```
#Chart with Indicator
#Notice how the 5 day average (white line) much more closely represents the data
# Whereas the the longer 15 day SMA smooths out price
barChart(SPY,theme=chartTheme("black"))
```

```
addSMA(n=5,col="white")
```

```
addSMA(n=15,col="red")
legend("bottomright",col=c("grey","red"),lty=1,legend=c("SMA(5)","SMA(15)"),cex=0.6)
```

```
#Bolinger Bands
#These bands are in essence moving averages used to identify abnormal
#volatitlity

#$$Upper\,\,Band=SMA_{20}\left( Close \right) +2*STDEV_{20}\left( Close \right)$$
#$$Middle\,\,Band\,\,=SMA_{20}\left( Close \right)$$
#$$Lower\,\,Band=SMA_{20}\left( Close \right) -2*STDEV_{20}\left( Close \right)$$

#Bollinger bands for 20 days with 2 standard deviations
bb <- BBands(HLC(SPY),n=20,sd=2)
# Technical Analysis Chart
barChart(SPY,theme=chartTheme("white"))
```

```
addBBands(n=20,sd=2)
```

```
#Moving average crossover strat
#This is by far and large one of the first things you learn in techinical analysis
#In order to test the same time period as the buy and hold strat
# I will be using a 50 and 100 average.
#The buy signal will occur when previous days close was < SMA then
#the next days current price close becomes greater than the SMA
#$$Buy:PreviousDaysClose<SMA\rightarrow CurrentDaysClose>SMA$$


barChart(SPY,theme=chartTheme("white"))
```

```
addSMA(n=5,col="darkblue")
```

```
addSMA(n=21,col="darkred")
legend("bottomright",col=c("darkblue","darkred"),lty=1,legend=c("SMA(5)","SMA(21)"),cex=0.6)
```

```
# Lag executes command the day after signal
# If yesterdays close price of SPY is less than yesterdays mvng avg &
# if todays closing price is > the mvng avg = 1 which means Buy
# If the opposite sma_Price_Buy = -1 which means sell
# Else 0 which means do nothing

# Simple Moving Average
sma50 = SMA(Cl(SPY),n=20)
sma100 = SMA(Cl(SPY),n=100)
#Signal
sma_Price_Strat = Lag(ifelse(Lag(Cl(SPY))<Lag(sma50)&Cl(SPY)>sma50,1,ifelse(Lag(Cl(SPY))>Lag(sma50)&Cl(S
#Do nothing if NA, replace with 0
sma_Price_Strat[is.na(sma_Price_Strat)] = 0

#Strat
#Creates a variable of all 1's
sma_50_posistion = ifelse(sma_Price_Strat>1,0,1)
for(i in 1:length(Cl(SPY))){sma_50_posistion[i] = ifelse(sma_Price_Strat[i]==1,1,
                                                         ifelse(sma_Price_Strat[i]==-1,0,sma_Price_Strat[i-1]))}
sma_50_posistion[is.na(sma_50_posistion)] = 1
sma50poscomp = cbind(Cl(SPY),sma50,sma_Price_Strat,sma_50_posistion)
colnames(sma50poscomp) = c("Close",'sma50',"sma_Price_Strat","sma_50_posistion")
View(sma50poscomp)

Ret = dailyReturn(Cl(SPY),type="arithmetic")
Ret[1] = 0
bh_strat = Ret
sma50strat <- Ret*sma_50_posistion

chart.CumReturns(sma50strat)
```

```
table.AnnualizedReturns(sma50strat)
```

```
##                           daily.returns
## Annualized Return                0.1288
## Annualized Std Dev               0.0827
## Annualized Sharpe (Rf=0%)        1.5577
```

```
chart.CumReturns(bh_strat)
```

```
table.AnnualizedReturns(bh_strat)
```

```
##                           daily.returns
## Annualized Return                0.1502
## Annualized Std Dev               0.3350
## Annualized Sharpe (Rf=0%)        0.4484
```