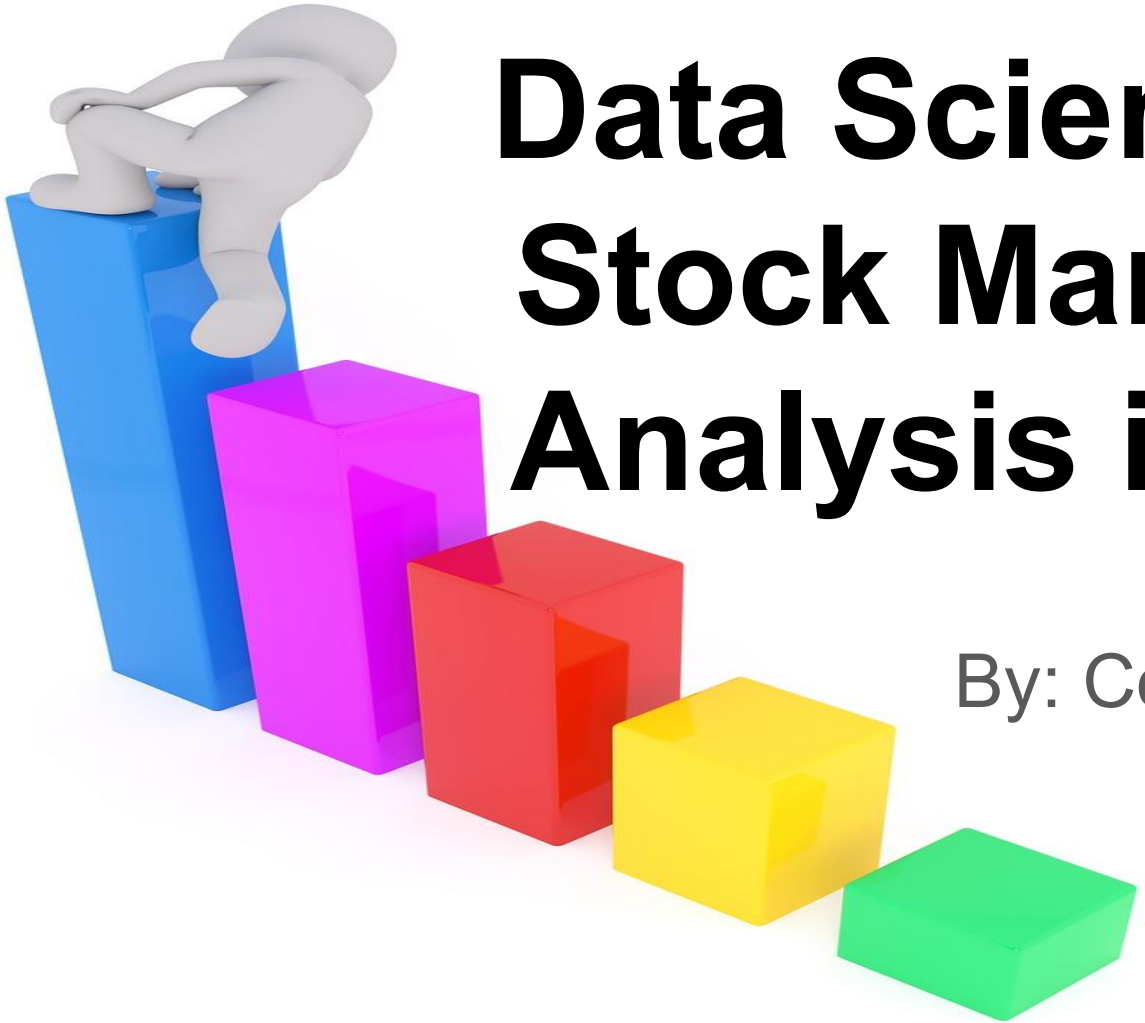


Data Science: Stock Market Analysis in R

By: Cong Nguyen

RStudio

Introduction

+ Analyzing Stock Variables
+ Visualization (ggplot and plot)
+ Explain what is Velocity Share
+ Stock correlation
+ Log Return

Data Frame

	AMZN.Open	AMZN.High	AMZN.Low	AMZN.Close	AMZN.Volume	AMZN.Adjusted
2018-12-06	1614.87	1701.05	1609.85	1699.19	8789400	1699.19
2018-12-07	1705.07	1718.93	1625.46	1629.13	7576100	1629.13
2018-12-10	1623.84	1657.99	1590.87	1641.03	7494800	1641.03
2018-12-11	1678.00	1679.47	1619.60	1643.24	6244700	1643.24
2018-12-12	1669.00	1704.99	1660.27	1663.54	6598000	1663.54
2018-12-13	1680.00	1692.12	1641.50	1658.38	5271300	1658.38
2018-12-14	1638.00	1642.57	1585.00	1591.91	6367200	1591.91

	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2018-12-06	171.76	174.78	170.42	174.72	43098400	172.1337
2018-12-07	173.49	174.49	168.30	168.49	42281600	165.9959
2018-12-10	165.00	170.09	163.33	169.60	62026000	167.0895
2018-12-11	171.66	171.79	167.00	168.63	47281700	166.1338
2018-12-12	170.40	171.92	169.02	169.10	35627700	166.5969
2018-12-13	170.49	172.57	169.55	170.95	31898600	168.4195

Using index:

AAPL (Apple)

GOOGL (Alphabet) Alphabet

AMZN (Amazon Inc.) amazon

HD (Home Depot) HOME DEPOT

M (Macy) macy's

UGAZ (VELOCITYSHARES 3X)

DGAZ(VELOCITYSHARES 3X)

UNG (United States Natural Gas Fund)

UNG United States Natural Gas Fund, LP

RStudio

Variables

+Closing Price is the last price when the market closes at 4:00PM

+Opening Price is the first price when the market opens at 9:30AM

+High and Low Price are the highest and lowest price traded

+Volume which are the amount of shares being bought and sold

+Adjusted Price is the price reflect that stock's value after accounting for any corporate actions such as stock splits, dividends / distributions and rights offerings

Addition Variables:

True Range: High - Low

+ Track Volatility
+ Price Range

Date <date>	AMZN.Open <dbl>	AMZN.High <dbl>	AMZN.Low <dbl>	AMZN.Close <dbl>	AMZN.Volume <dbl>	AMZN.Adjusted <dbl>	Day <int>	Growth <dbl>	TrueRange <dbl>
2015-12-07	674.74	675.46	660.50	669.83	3732800	669.83	7	0	14.960022
2015-12-08	663.13	679.99	659.79	677.33	3651900	677.33	8	1	20.200012
2015-12-09	678.00	679.00	655.68	664.79	5158200	664.79	9	0	23.320007
2015-12-10	665.59	668.53	659.56	662.32	3455500	662.32	10	0	8.970031
2015-12-11	651.23	657.88	639.62	640.15	5474400	640.15	11	0	18.260010
2015-12-14	641.75	658.59	635.27	657.91	4329700	657.91	14	1	23.320007

RStudio

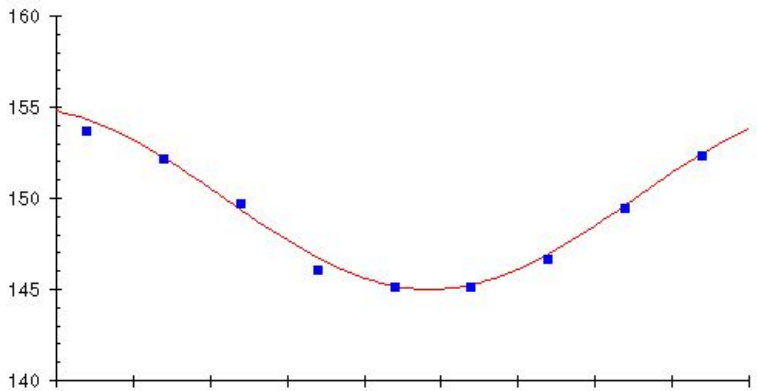
Simple Moving Average Line (SMA)

- The SMA is a technical indicator for determining if a stock price will continue or reverse a bull or bear trend.

- The SMA is calculated as the average of a stock's price over some period.

Simple Moving Average Line (SMA)

$$\text{SMA} = \frac{A_1 + A_2 + ... + A_n}{n}$$



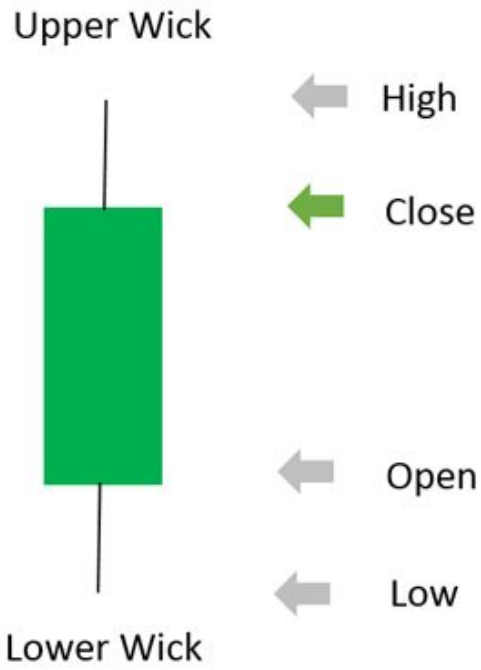
The mean of the data point.

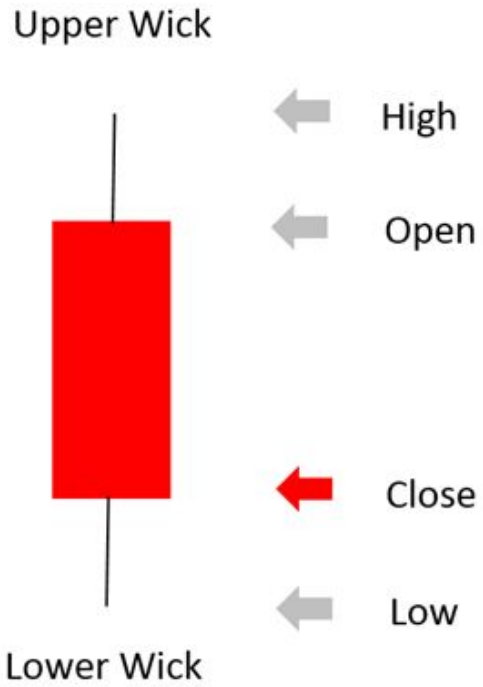
An = Closing Price

n = period of time

A 10-day moving average would average out the closing prices for the first 10 days as the first data point. The next data point would drop the earliest price, add the price on day 11 and take the average, and so on.

Japanese Candlestick - Candle Chart





RStudio

Candle Chart from 2019-06-06 to 2019-11-25

Candle Chart from 2017-06-06 to 2019-11-25

```
A = df1 %>%
  ggplot(aes(x = Date, y = AMZN.Close)) +
  geom_candlestick(aes(open = AMZN.Open, high = AMZN.High, low = AMZN.Low, close =AMZN.Close),
                   colour_up = "forestgreen", colour_down = "red",
                   fill_up = "forestgreen", fill_down = "red") +
  geom_ma(color = "gold4", n = 9)  + geom_ma(color = "gray7", n = 180) +
  coord_x_date(xlim = c("2019-06-06", "2019-11-25"),ylim = c(1550, 2100))+
  ggtitle("AMZN") +
  xlab("Date") + ylab("Price")
A
```

Candle Chart Implementation

The Candle charts helps with studying

+ Price movements
+ Patterns
+ Market trends

=> Predict future direction of a stock’s price

=> Identifying possible trend reversal

Regression Model

```
Call:
lm(formula = AAPL.Close ~ AAPL.Volume, data = df2)

Residuals:
    Min      1Q  Median      3Q     Max
-70.229 -37.677   3.658  27.811 105.503

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.843e+02  3.175e+00  58.040  < 2e-16 ***
AAPL.Volume -7.491e-07  8.998e-08  -8.324 2.75e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.69 on 1005 degrees of freedom
Multiple R-squared:  0.0645,    Adjusted R-squared:  0.06357
F-statistic:  69.3 on 1 and 1005 DF,  p-value: 2.746e-16

  (Intercept)   AAPL.Volume
 1.842508e+02 -7.490558e-07
```

AAPL

y=mx+b

Closing price = -7.491e-07* Volume+ 1.843e+02

```
Call:
lm(formula = AMZN.Close ~ AMZN.Volume, data = df1)

Residuals:
    Min      1Q  Median      3Q     Max
-919.52 -425.22  -71.42  477.08  799.37

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.151e+03  3.296e+01  34.917  < 2e-16 ***
AMZN.Volume 2.423e-05  6.841e-06   3.543 0.000414 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 471.7 on 1005 degrees of freedom
Multiple R-squared:  0.01233,   Adjusted R-squared:  0.01135
F-statistic: 12.55 on 1 and 1005 DF,  p-value: 0.0004144

 (Intercept)  AMZN.Volume
1.150825e+03 2.423364e-05
```

AMZN

y=mx+b

Closing price = 2.423e-05* Volume+ 1.151e+03

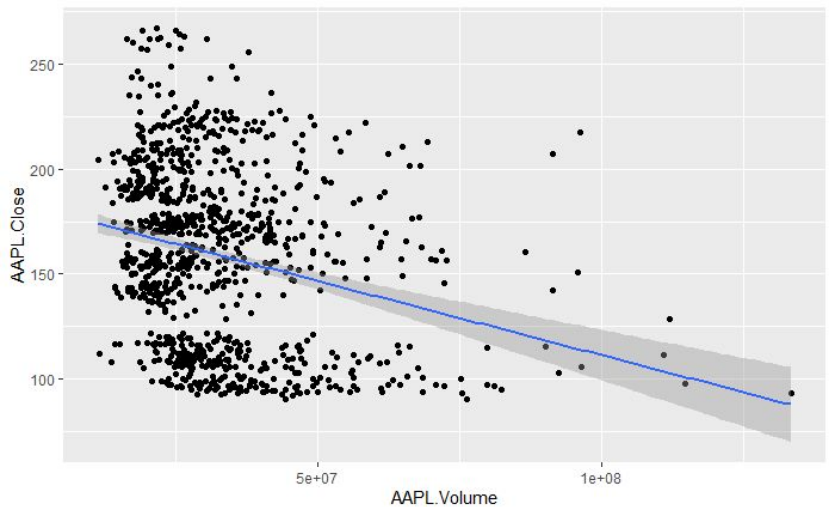
GOOGL

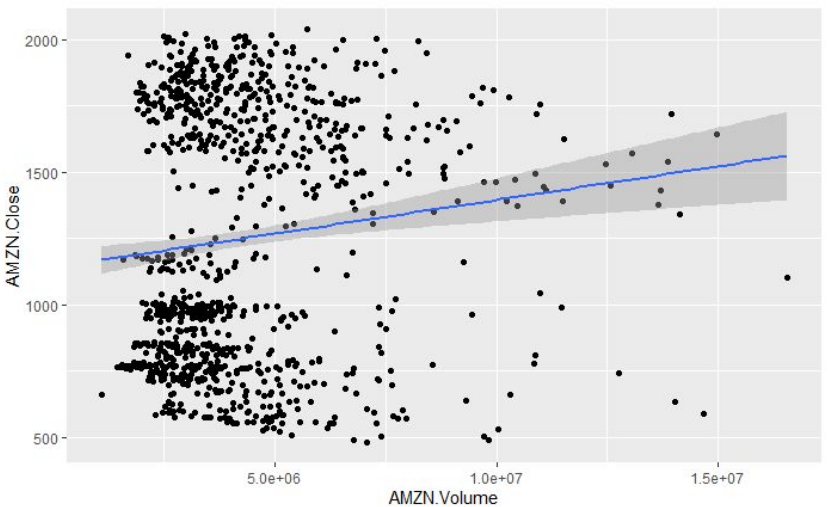
y=mx+b

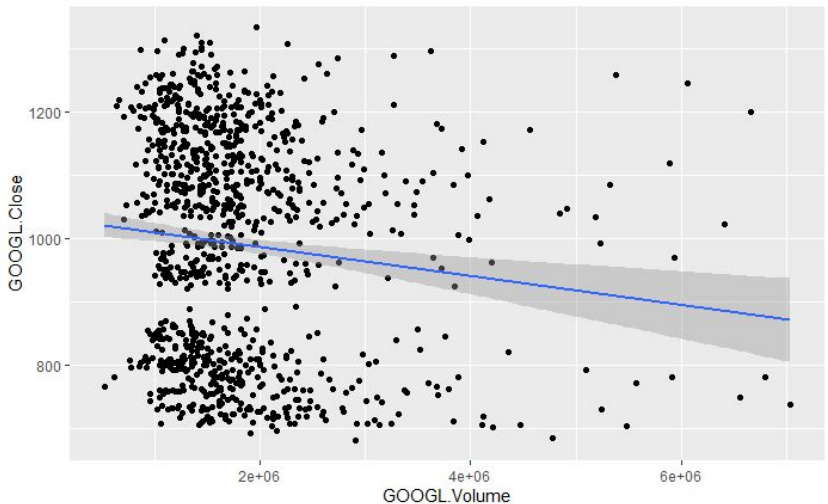
Closing price = -2.499e-05* Volume+ 1.039e+03

```
Call:
lm(formula = GOOGL.Close ~ GOOGL.Volume, data = df3)

Residuals:
   Min     1Q Median     3Q    Max
-305.2 -179.8   39.5  144.2  357.8

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.039e+03  1.266e+01  82.038  < 2e-16 ***
GOOGL.Volume -2.499e-05  6.341e-06  -3.941 8.66e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 175.5 on 1005 degrees of freedom
Multiple R-squared:  0.01522,   Adjusted R-squared:  0.01424
F-statistic: 15.53 on 1 and 1005 DF,  p-value: 8.663e-05

  (Intercept)  GOOGL.Volume
 1.038899e+03 -2.499214e-05
```

RStudio®


AAPL.Close
250
200
150
100
5e+07
1e+08
AAPL.Volume


AMZN.Close
2000
1500
1000
500
5.0e+06
1.0e+07
1.5e+07
AMZN.Volume


GOOGL.Close
1200
1000
800
2e+06
4e+06
6e+06
GOOGL.Volume

RStudio®

Log Return

The logarithmic return is a way of calculating the rate of return on an investment, in this case a stock's price.

The main benefit of using log returns rather than price is normalization. It measures all stocks in a comparable system that helps with identifying and analyzing relationships amongst two or more stocks despite considerable price differentiation.

$$R = \frac{\ln\left(\frac{V_f}{V_i}\right)}{t} \times 100\,\%$$

```
data=cbind(diff(log(Cl(GOOGL))),diff(log(Cl(AMZN))),diff(log(Cl(M))),diff(log(Cl(HD))))
chart.Correlation(data)
```

RStudio

Annual Log Return

	AMZN_Return	GOOGL_Return	AAPL_Return	HD_Return	M_Return
2015-12-31	0.9006389	0.6473291	-11.662128	-1.262288	-8.763072
2016-12-30	10.3869488	1.8390037	11.760890	3.480210	6.307037
2017-12-29	44.4406057	28.4648902	39.517455	36.888964	-29.148957
2018-12-31	25.0226931	-0.8044481	-5.540892	-7.575328	21.223050
2019-12-05	14.7383365	23.8912012	53.588116	24.171680	-62.342279

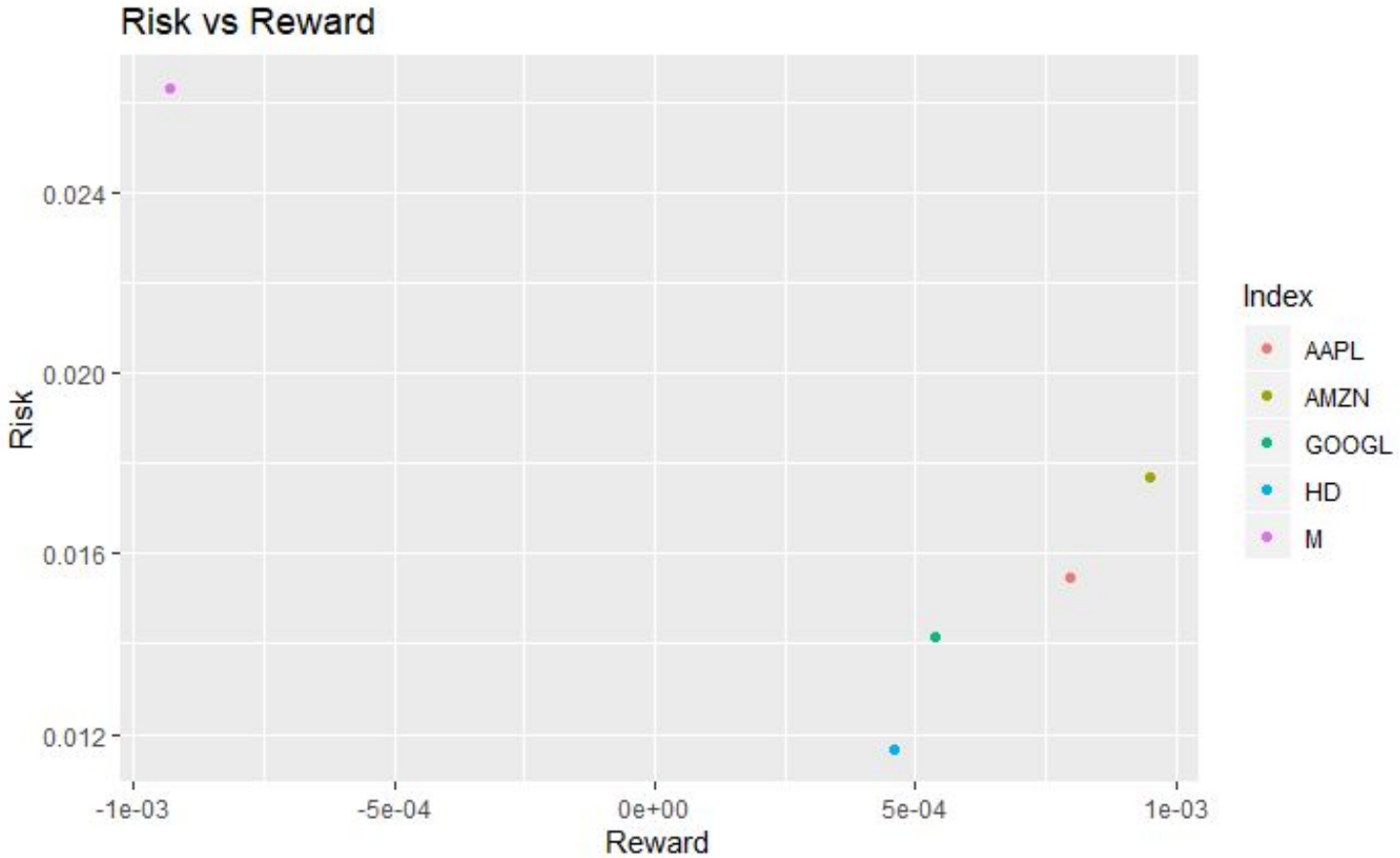
The important factor in long term investment is tracking a company growth rate. The annual log return shows a more overall understanding of how a company is doing financially

Standard Deviation

+ Variation between data point in relation to the mean.
+ High SD => Large Price Range => More Risk
+ Low SD => Low Price Range => Less Risk

Blue Chips stock normally have a low SD values

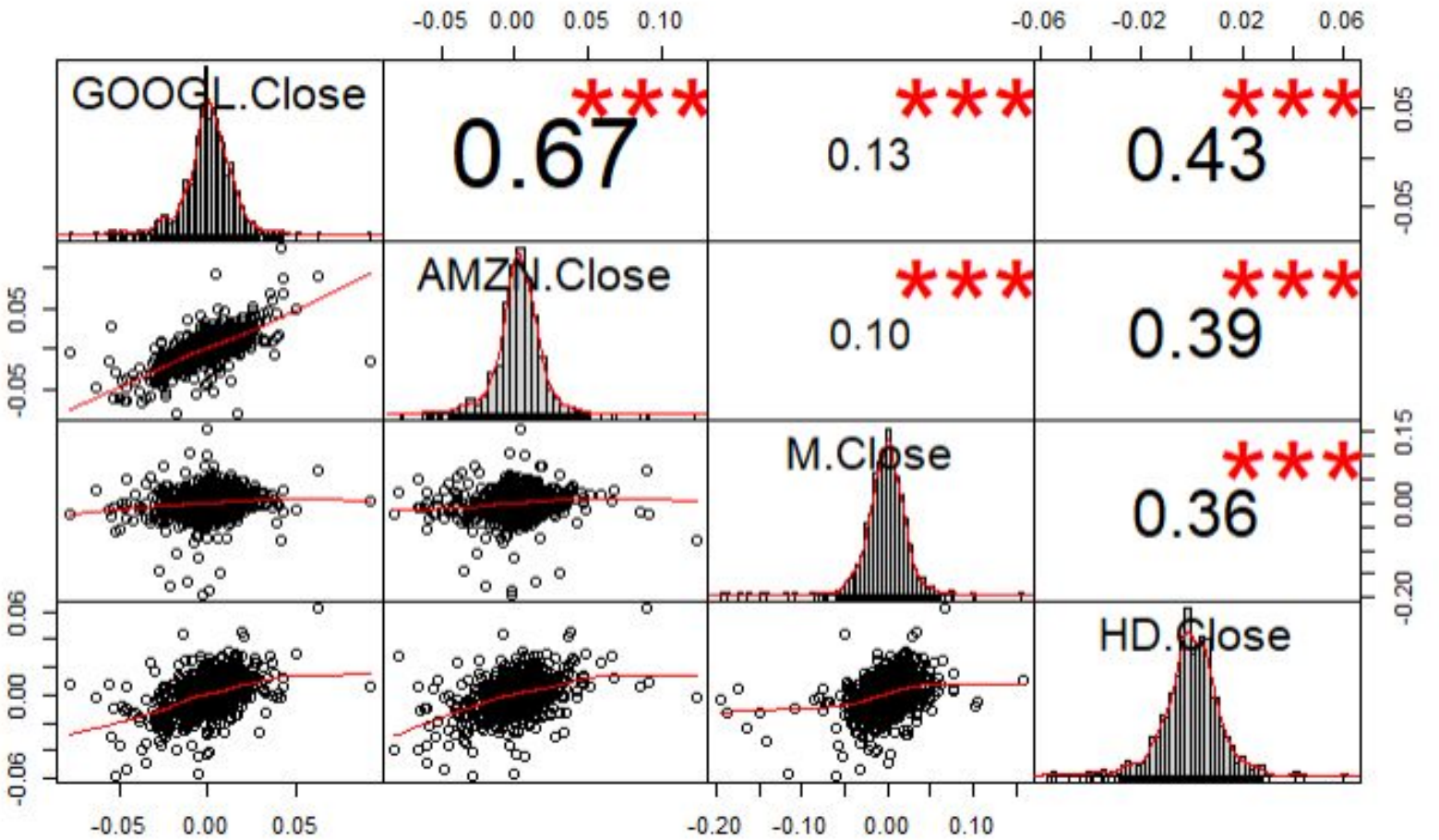
Index <fctr>	Mean_Log_Return <dbl>	Sd_Log_Return <dbl>
AMZN	0.0009491970	0.01767198
HD	0.0004612077	0.01167334
M	-0.0009310132	0.02629034
AAPL	0.0007960418	0.01548472
GOOGL	0.0005371568	0.01414864


Risk vs Reward
Risk
0.024
0.020
0.016
0.012
-1e-03
-5e-04
0e+00
5e-04
1e-03
Reward
Index
AAPL
AMZN
GOOGL
HD
M

RStudio

Why understanding the correlation of stock is important?

- A popular investing strategy is to diversify your stock portfolio.
- A good portfolio usually composed of many different types of stock that has a small correlation because you want to minimize risk.
- A trader can learn how to profit regardless of the direction of the market


GOOGL.Close
AMZN.Close
M.Close
HD.Close
0.67 ***
0.13 ***
0.43 ***
0.10 ***
0.39 ***
0.36 ***
-0.05 0.00 0.05 0.10
-0.06 -0.02 0.02 0.06
-0.05 0.00 0.05
-0.20 -0.10 0.00 0.10
-0.06 0.00 0.06
0.15 0.00 -0.20

Velocity Share

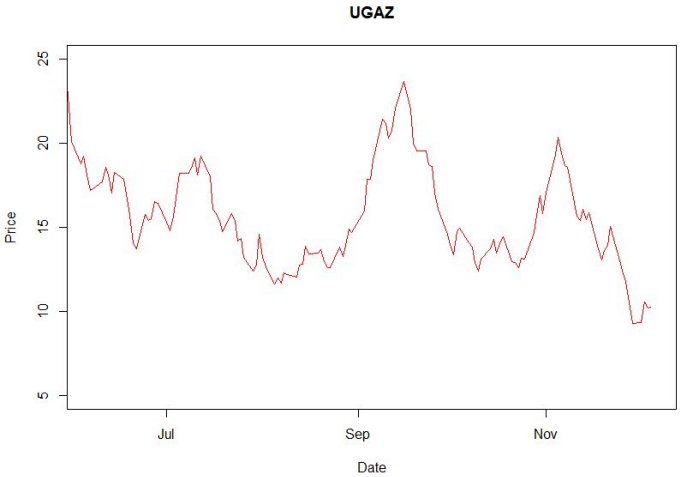
+ Leveraged Share that is based on the performance of a specific stock

  (2X, 3X)

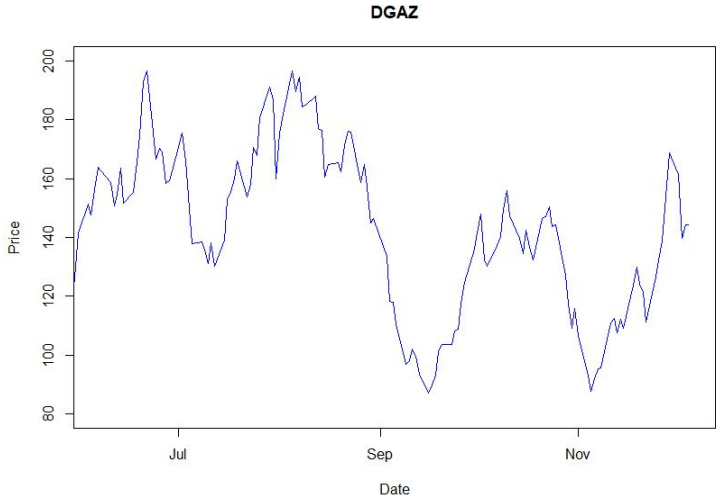
+ Instead of a stock, you can a velocity share of that stock

+ Magnify the return on investment

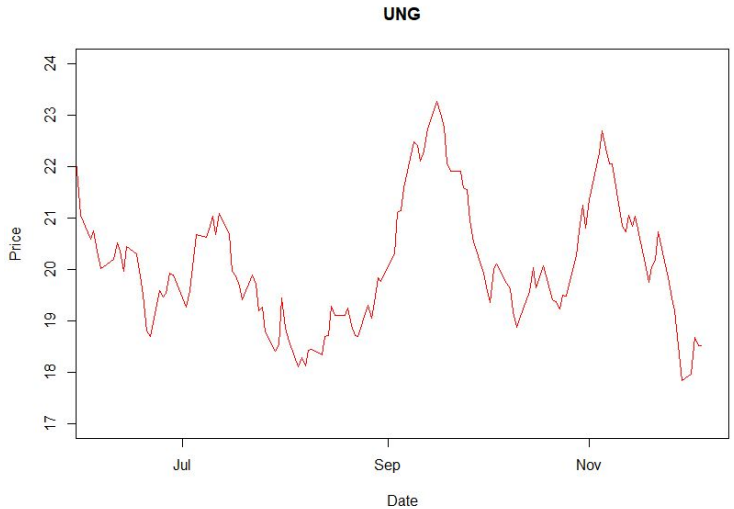
+ Losses are magnified



Velocity Share



Tracked Share



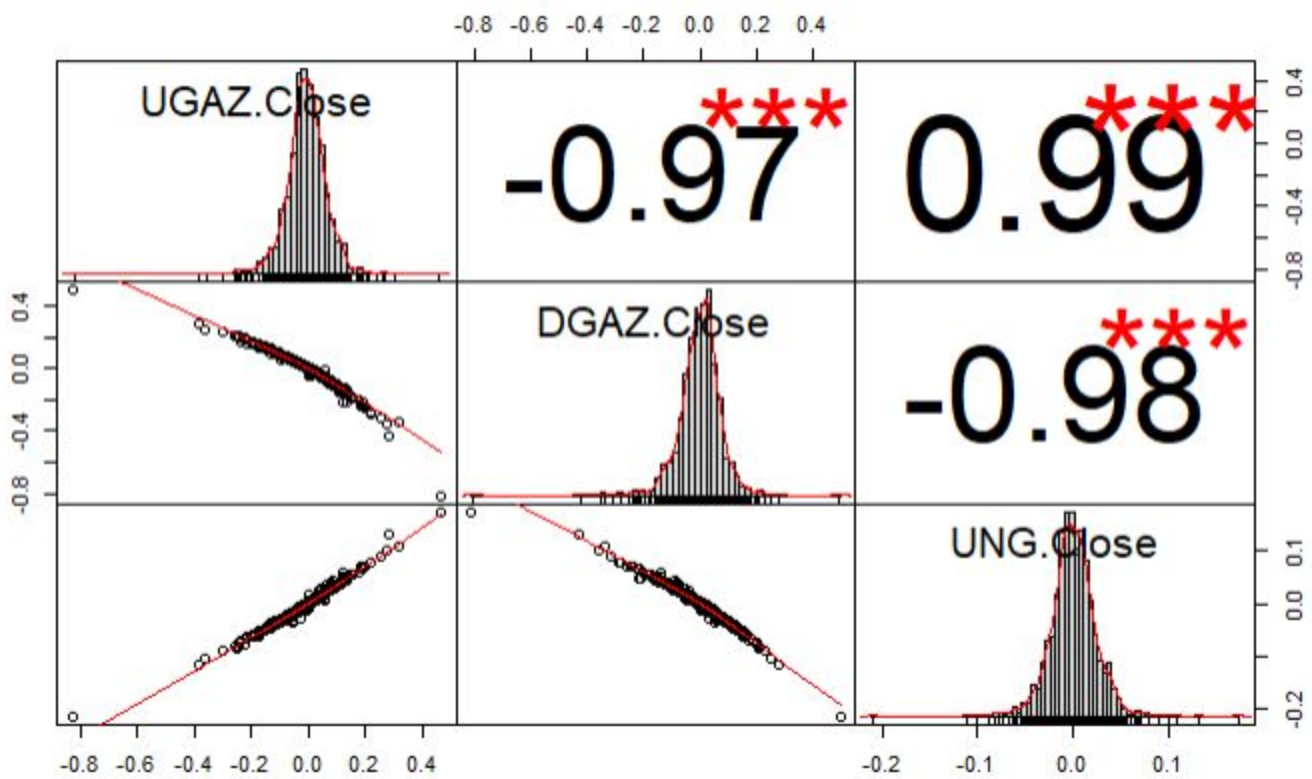
RStudio®

Leveraged Commodities - Energy Natural Gas

DGAZ: VelocityShares 3X Inverse Natural Gas ETN

UGAZ: VelocityShares 3X Long Natural Gas ETN

UNG: United States Natural Gas Fund LP


UGAZ.Close
DGAZ.Close
UNG.Close
-0.97***
0.99***
-0.98***
-0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4
-0.2 -0.1 0.0 0.1

RStudio

Outside variables:

+ Stock Catalyst
+ Federal Interest Rate
+ Company balance sheets