Name: Divya Ramesh
Case Study 1 - Waite First Securities [Exploratory Data Analysis]

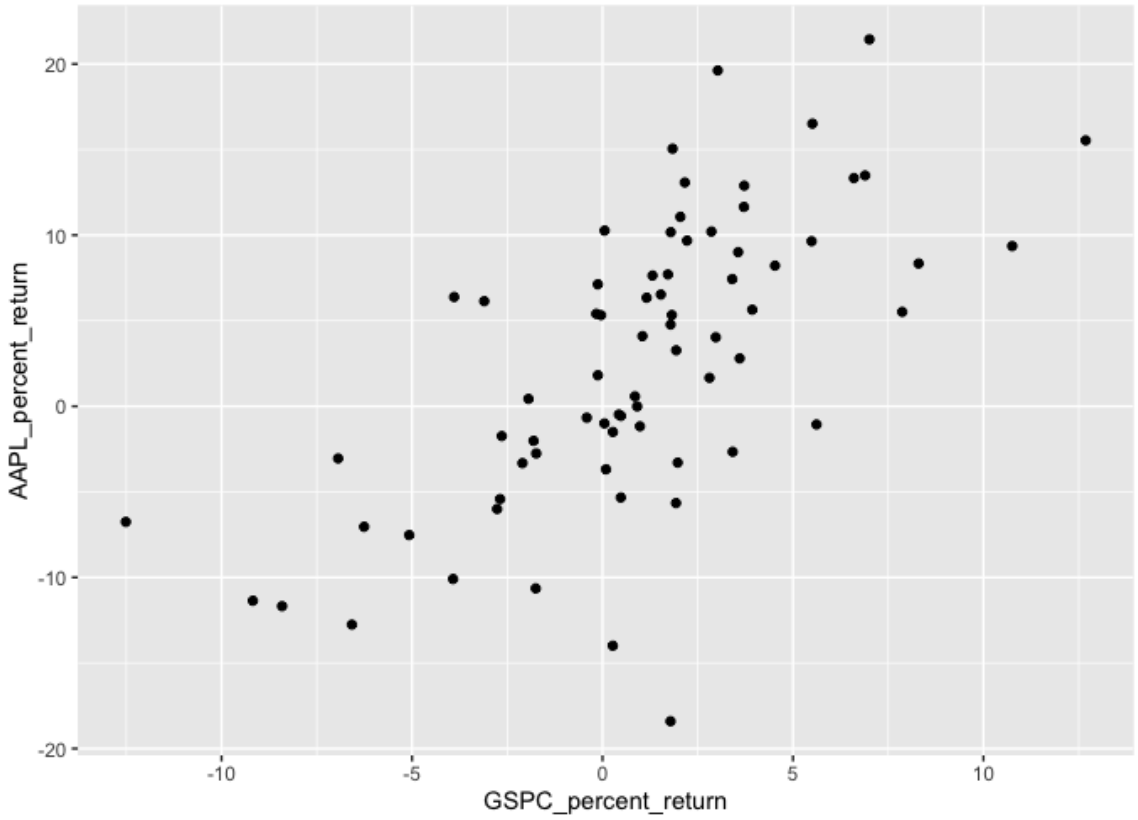
**R-Output**

```
> rm = (list = ls(all = TRUE)) #clears all memory
> options(digits=12) #keeps 12 digits in memory, otherwise summary statistics may be off due to rounding
> stockdata <- read.csv("WFS-Stockdata.csv", header=TRUE)  # imports and renames dataset
> stockdata <- stockdata[-c(1), ] # Deletes first row - December 2014
> colnames(stockdata) # prints column names on the screen
 [1] "Month..t."         "Date"              "Open"
 [4] "High"              "Low"               "Close."
 [7] "Adj.Close.."       "Volume"            "GSPC.Stock.Return"
[10] "GSPC_percent_return" "Open.1"            "High.1"
[13] "Low.1"             "Close"             "Adj.Close"
[16] "Volume.1"          "AAPL.Stock.Return"   "AAPL_percent_return"
[19] "Open.2"            "High.2"            "Low.2"
[22] "Close.1"           "Adj.Close.1"       "Volume.2"
[25] "INTC.Stock.Return"   "INTC_percent_return" "Open.3"
[28] "High.3"            "Low.3"             "Close.2"
[31] "Adj.Close.2"       "Volume.3"          "KR.Stock.Return"
[34] "KR_percent_return"

> round(summary(stockdata[,"GSPC_percent_return"]), 1)
   Min. 1st Qu.  Median  Mean 3rd Qu.    Max.
  -12.5    -0.7     1.2     0.9     3.1    12.7
> round(summary(stockdata[,"AAPL_percent_return"]), 1)
   Min. 1st Qu.  Median  Mean 3rd Qu.    Max.
  -18.4    -2.8     3.7     2.7     8.5    21.4
> round(summary(stockdata[,"INTC_percent_return"]), 1)
   Min. 1st Qu.  Median  Mean 3rd Qu.    Max.
  -20.2    -3.5     1.8     0.9     5.7    19.5
> round(summary(stockdata[,"KR_percent_return"]), 1)
   Min. 1st Qu.  Median  Mean 3rd Qu.    Max.
  -21.4    -6.4     2.3     0.4     5.0    24.9

> round(mean(stockdata[,"GSPC_percent_return"], na.rm = TRUE), 1)
[1] 0.9
> round(mean(stockdata[,"AAPL_percent_return"], na.rm = TRUE), 1)
[1] 2.7
> round(mean(stockdata[,"INTC_percent_return"], na.rm = TRUE), 1)
[1] 0.9
> round(mean(stockdata[,"KR_percent_return"], na.rm = TRUE), 1)
[1] 0.4
> #Standard Deviations
> round(sqrt(var(stockdata[,"GSPC_percent_return"], na.rm = TRUE)), 1)
[1] 4.3
> round(sqrt(var(stockdata[,"AAPL_percent_return"], na.rm = TRUE)), 1)
```

```
[1] 8.4
> round(sqrt(var(stockdata[,"INTC_percent_return"], na.rm = TRUE)), 1)
[1] 7.1
> round(sqrt(var(stockdata[,"KR_percent_return"], na.rm = TRUE)), 1)
[1] 7.9
```

**Stock Return of Apple in Comparison with S&P 500**

```
> round(cor(stockdata$GSPC_percent_return, stockdata$AAPL_percent_return), 2)
[1] 0.65

> ggplot(data = stockdata) + geom_point(mapping = aes(x = GSPC_percent_return, y =
AAPL_percent_return))
```




S&P 500 Data is obtained from GSPC. The correlation coefficient is 0.65, which shows a relatively strong positive correlation as it is closer to 1. Value of correlation coefficient ranges from +1 (strong positive correlation) to -1 (strong negative correlation). There are couple of outliers, such as a GSPC return of 1.8 with corresponding AAPL return of -18.4, and -3.9 (GSPC) to 6.4 (AAPL) which reduces the correlation. There could be various reasons leading to outliers. For instance in Nov 2014, Apple announced that it would not provide any of its sales information created an ambiguity causing a sudden decrease in Apple's stock prices,

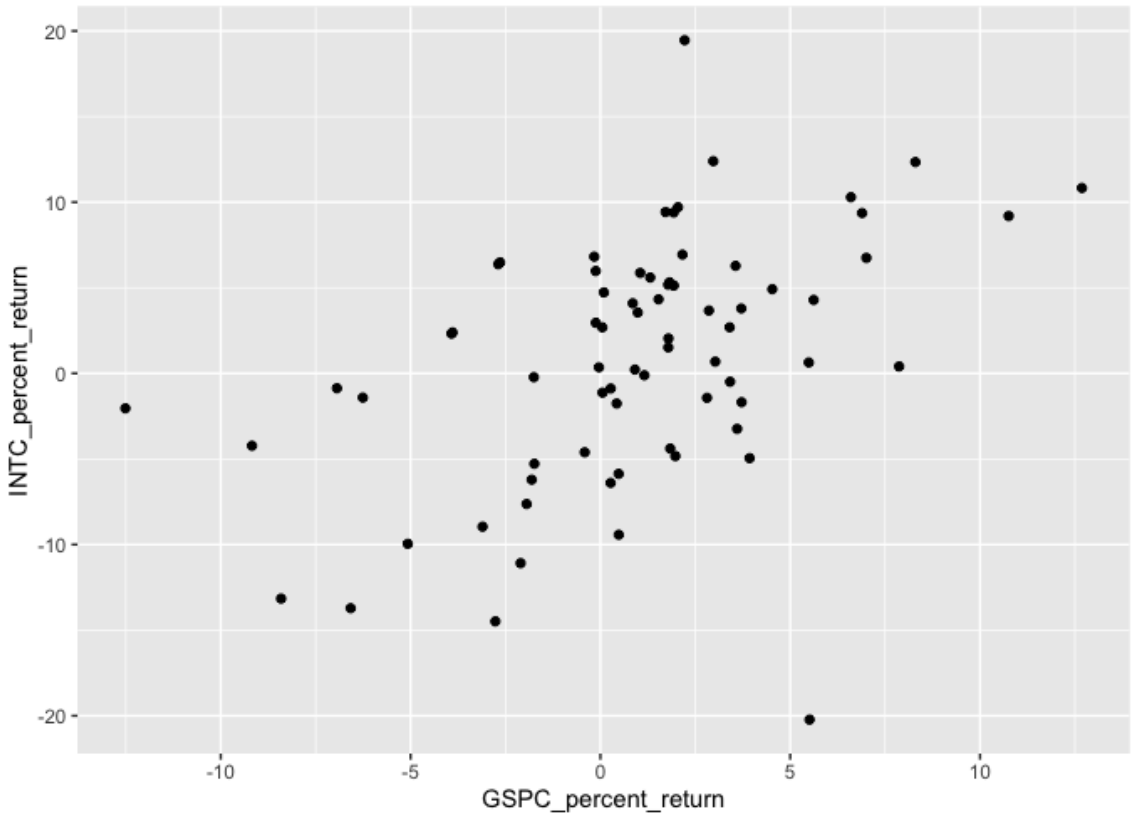
this has no effect on S&P 500 as no other companies were affected by their individual stock volitation. (Kim, 2018)

S&P 500 uses a weighted market capitalization to assign weightage to each stock, i.e. the weightage depends on each company's market capital. Apple has a weightage of 6.03 (ranked highest among all other 505 companies) (slickcharts, 2021), which shows that the monthly stock return from Apple is a valid comparison with that of S&P 500's monthly stock return.

**Stock Return of Intel in Comparison with S&P 500**

```
> round(cor(stockdata$GSPC_percent_return, stockdata$INTC_percent_return), 2)
[1] 0.46
```

```
> ggplot(data = stockdata) + geom_point(mapping = aes(x = GSPC_percent_return, y =
INTC_percent_return))
```



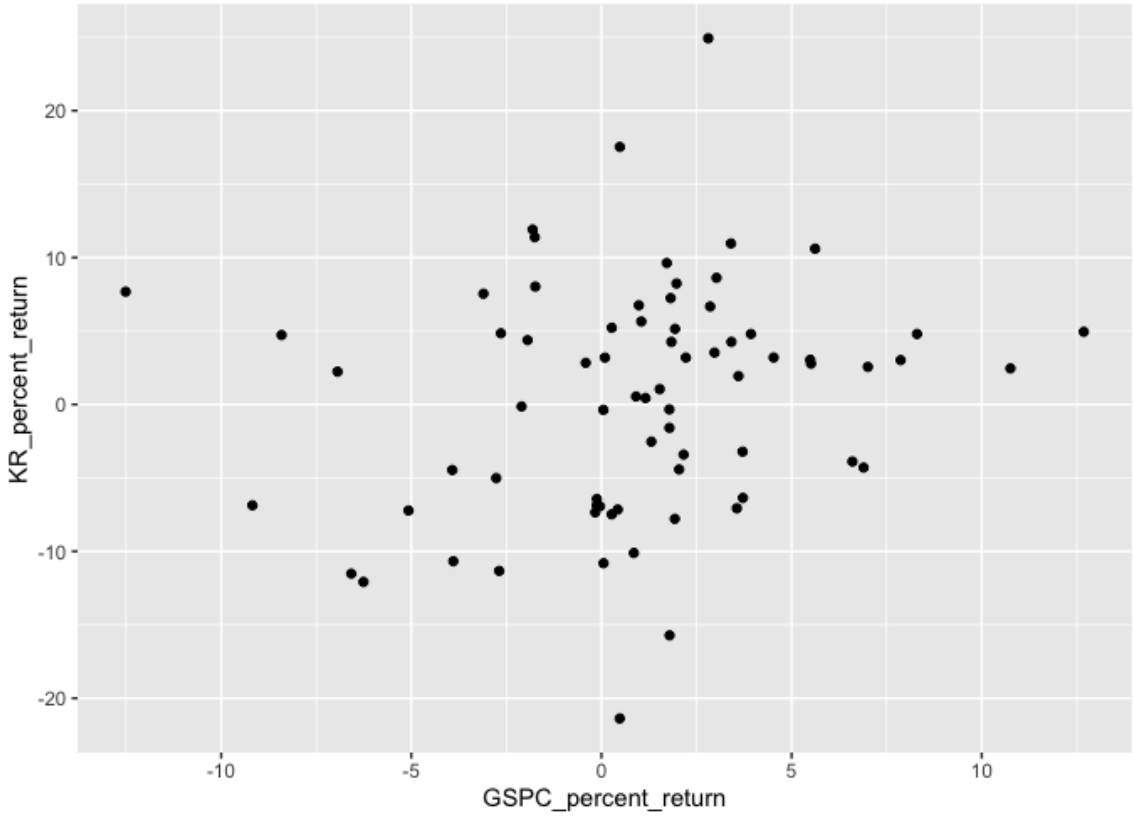

The correlation coefficient is 0.46, which denotes a moderately positive correlation. The most prominent outliers as depicted from the graph are -20.2 stock return in Intel, whereas a 5.1 stock return in S&P 500,

similarly 19.5 in Intel and 2.2 in S&P 500, -2 in Intel and -12.5 in S&P 500 etc. This is the main reason for a moderate positive relationship between Intel and S&P 500. Intel has a weightage of only 0.59 (slickcharts, 2021), which shows that the monthly stock return from Intel is not a valid comparison with that of S&P 500's monthly stock return, as compared to that of Apple.

**Stock Return of Kroger in Comparison with S&P 500**

```
> round(cor(stockdata$GSPC_percent_return, stockdata$KR_percent_return), 2)
[1] 0.18
```

```
> ggplot(data = stockdata) + geom_point(mapping = aes(x = GSPC_percent_return, y =
KR_percent_return))
```




A correlation coefficient of 0.18 and the graph shows a weak correlation between Kroger and S&P 500, compared to AAPL. This clearly shows that the S&P returns do not predict the movement of Kroger's returns. The data is widely dispersed and there are many outliers present which affects the correlation to a great extent.

Kroger has a weightage of only 0.072 (slickcharts, 2021), which shows that the monthly stock return from Kroger is not a valid comparison with that of S&P 500's monthly stock return.

**References**

Kim, T. (2018, Dec 27). Retrieved from Barrons: https://www.barrons.com/articles/apple-stock-2018-year-in-review-51545862266

*slickcharts*. (2021). Retrieved from https://www.slickcharts.com/sp500