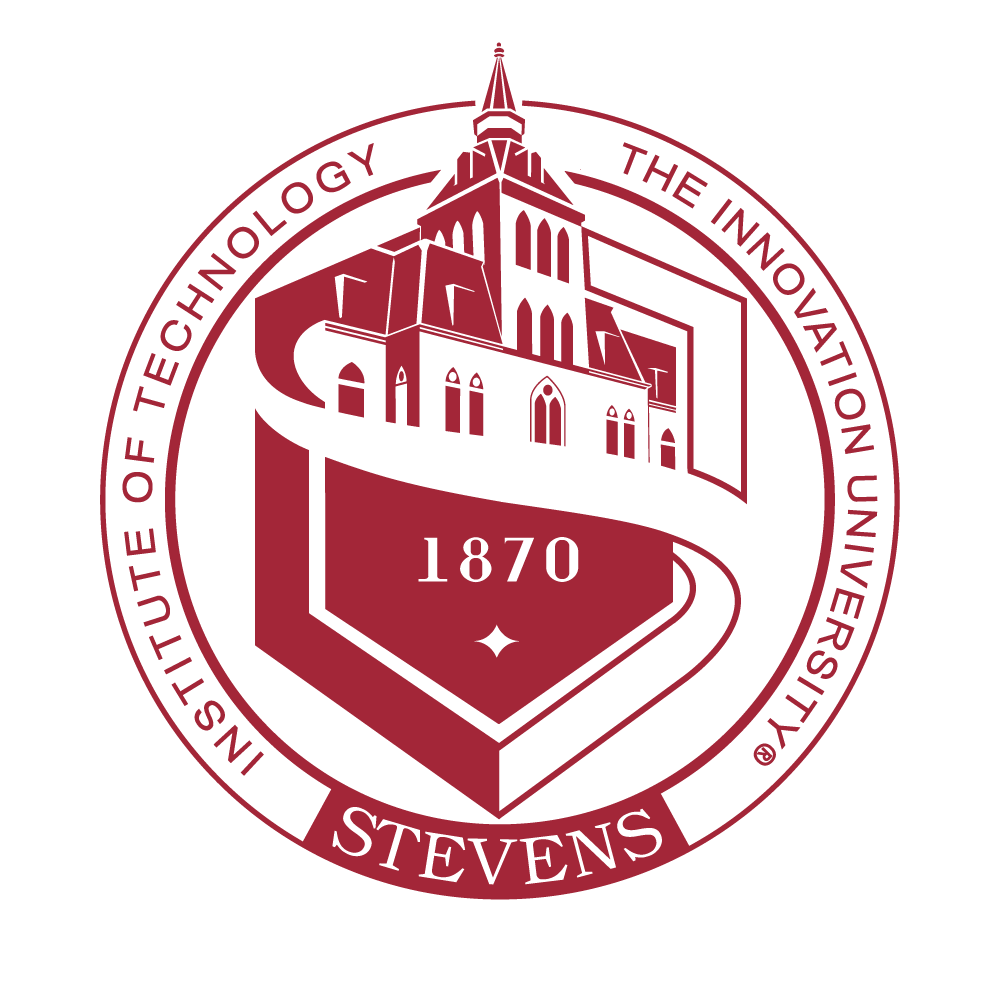
**Project Final Report**

**Portfolio Performance**



FA582-A: Foundations of Financial Data Science

Group 2:

Raj Gondaliya

Rishabh Rawat

Fall 2021 Semester

November 21, 2021

**Abstract**

The purpose of this paper is to determine the performance of the nations most commonly known stock indices in comparison to top performing companies across all sectors. This paper uses data and analysis methods to answer the question: How does the performance of stock market indices of the leading companies in the United States compare to indices generated with top performing companies across all sectors over the last 10 years? A portfolio was generated using stock prices from what the team deemed to be the top performing companies within each of the eleven GICS sectors. This data was used to compare the portfolio performance to the S&P 500 and Dow Jones using methods such as return comparison, looking at different trading strategies, and clustering.

**Introduction**

The Standard & Poor’s 500, more commonly known as S&P 500, is a stock market index of five-hundred of the leading publicly traded companies in the United States. Similarly, the Dow Jones Industrial Average, or Dow Jones, is a price-weighted measurement stock market index of thirty prominent companies listed on stock exchanges in the United States. These are two of the most commonly viewed equity indices in the United States today.

For this project, the team would like to create their own portfolio consisting of companies of top performance within each of the GICS sectors such as Energy, Health Care, Information Technology, and more. Once a portfolio is created, the team is planning on comparing the performance of their portfolio with the performance of the S&P 500 and Dow Jones over a period of the last ten years. With the stock data readily available to the team from sources like Bloomberg and Yahoo Finance, the team plans to implement a complete scraping, cleaning, and analysis of the data using examples and methods discussed in class in a practical, real-world application. The purpose of executing this study is to get a better understanding of how well the nation's top performing stocks are really doing in comparison to a portfolio composed of stocks for companies within a large range of industries. This study provides quantifiable evidence to support the research question outlined above.

**Data and Methods**

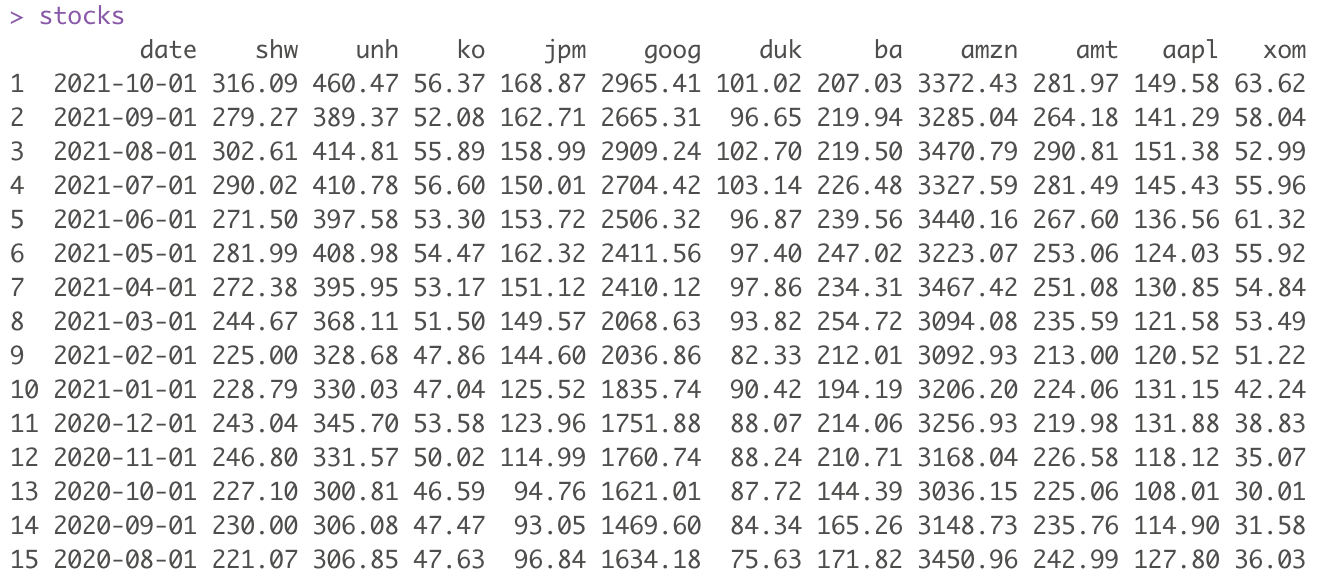
In order to achieve the goals outlined for the project, the team put together a dataset using excel and up-to-date stock information from Yahoo Finance. Using a time period of 10 years beginning in 2011 to 2021, the team created a data frame consisting of stock price data and percent change. This data frame included information from one top performing company in each of the 11 GICS stock market sectors. These sectors are listed below:

1. Energy
2. Materials
3. Industrials
4. Utilities
5. Healthcare
6. Financials
7. Consumer Discretionary
8. Consumer Staples
9. Information Technology
10. Communication Services
11. Real Estate

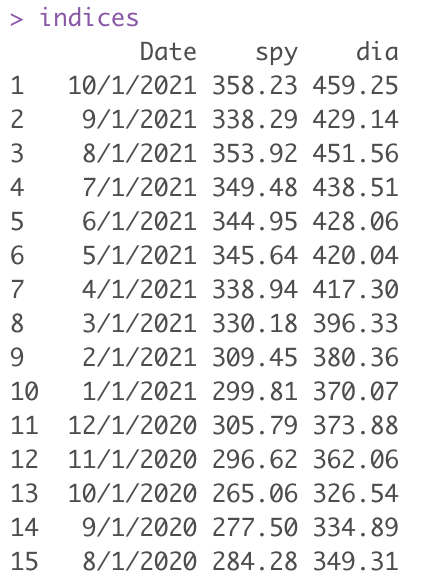
In order to select the stocks from each of these sectors which were considered to be the top performing, the team conducted research on each of these sectors and selected the largest companies based on historical return and risk over the past 10 years. While if the team only looked at performance in the past year, the selections may look different, since years of performance are being taken into account, these selections mostly happen to be the well-known larger companies in each sector.

The energy sector contains companies within oil and natural gas industries as well as other consumable fuels. Within the energy sector, the largest companies are ExxonMobil (XOM) and Chevron (CVX). The team selected ExxonMobil for their analysis. Looking at the materials sector, which is made up of companies involved in production of goods for manufacturing, the more well-known companies are Sherwin-Williams (SHW) and DuPont (DD). The team selected the paint maker, Sherwin-William for this sector. For the industrials sector, Boeing (BA) and Union Pacific (UNP) are the largest stocks. This sector is very broad in that it contains companies that involve use of heavy equipment such as transportation companies, construction, engineering, and others within this general field. The team selected Boeing, an airplane manufacturer, to use in the analysis. The utilities sector comprises utility companies which provide commodities such as electrical power, natural gas, water, etc. While the big utilities companies tend to vary depending on physical location, the team selected Duke Energy (DUK), which is primarily in service in the Southeast U.S., because of its historical returns. Looking at the healthcare sector, which includes companies that develop pharmaceuticals and treatments based on biotechnology as well as healthcare equipment and services companies. The leaders in this sector are UnitedHealth Group (UNH) and Johnson & Johnson (JNJ), for which the team selected UnitedHealth Group to analyze. The largest companies in the financial sector are Berkshire Hathaway (BRK-A, and BRK-B) and JPMorgan Chase (JPM). This sector is made up of companies that primarily handle money such as banks or insurance companies. The team selected JPMorgan Chase from the financials sector for their analysis. The consumer discretionary and consumer staples sectors are next. Each of these sectors cover goods and services companies with the former having consumer demand depending on financial status and the latter does not. These sectors’ top companies are Amazon (AMZN) and McDonald’s (MCD) and Coca-Cola (KO) and Procter & Gamble (PG), respectively. Of the two top companies for each sector, the team selected Amazon and Coca-Cola to be included in their analysis. Looking at the information technology sector, which contains companies within different areas of technological innovation, the top companies are Apple (APPL) and Microsoft (MSFT) without surprise. The team chose to look at Apple in their project comparison. Nearing the end of the list is the communication services sector. Within this sector are companies involved with telecommunication services, media, and entertainment. This is the newest sector and was formed from areas which were previously included in other sectors. The top companies within communication services are Facebook (FB) and Alpabet (GOOGL) which includes Google (GOOG), from which the team selected Google. Finally, for the real estate sector, which contains two different types of investments related to real estate: new real estate project development and management companies, and real estate investment trusts. The top stocks in this sector are American Tower (AMT) and Simon Property Group (SPG). The team chose the cellular communications tower company, American Tower. [1]

The datasets created by the team can be seen in the figures below. The data was taken from Yahoo Finance as the monthly close prices for each of the eleven selected stocks, as well as a dataset containing the close prices of the SPY and DIA which were used as comparisons. These were used to perform the majority of the data analysis within this project.

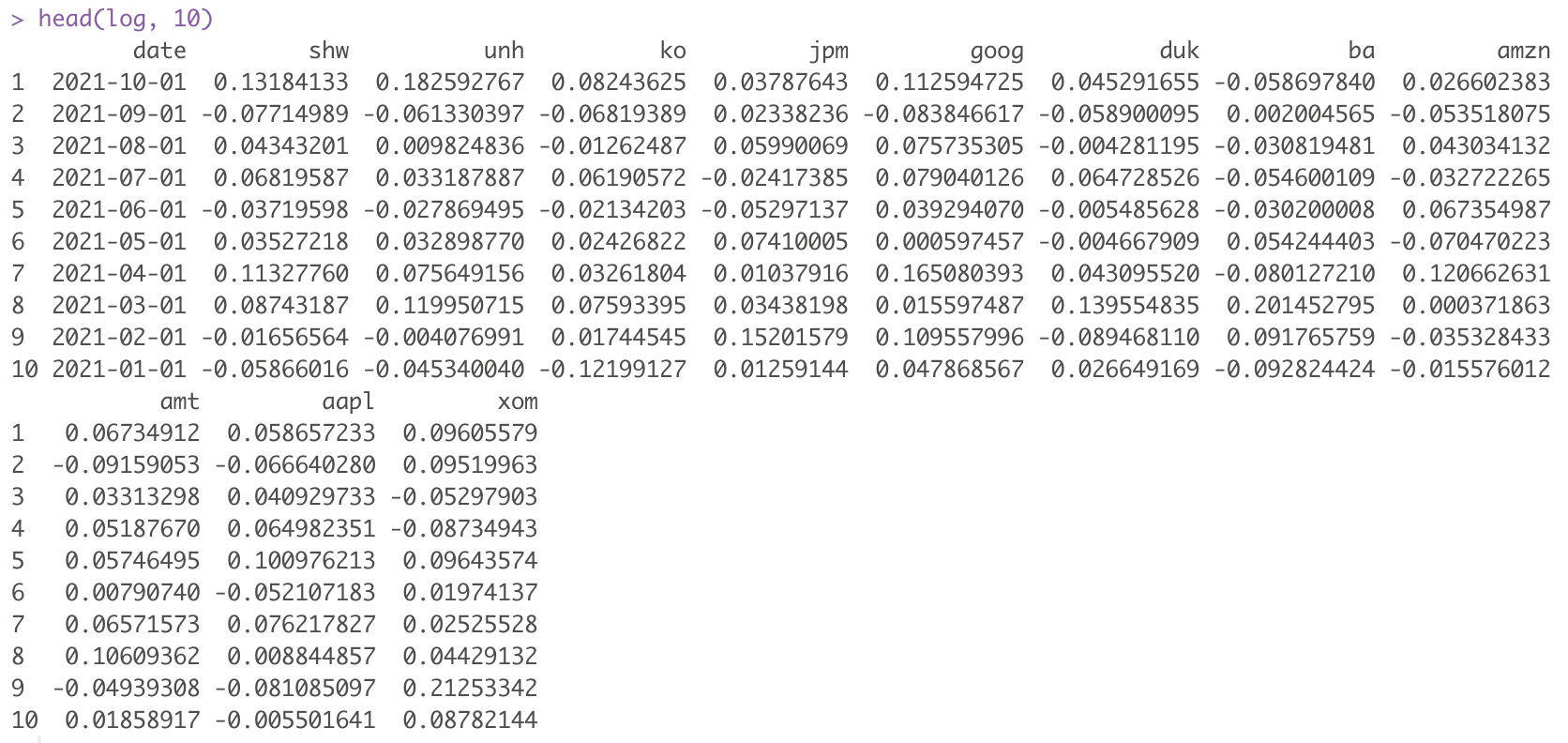


*Figure 1: Dataset containing monthly stock prices for selected portfolio*



*Figure 2: Dataset consisting of monthly close prices for chosen indices (SPY and DIA)*

From the dataset in Figure 1, the team was able to create a dataset which contains the monthly log returns for each of the stocks in the selected portfolio. This dataset can be seen in Figure 3 below.

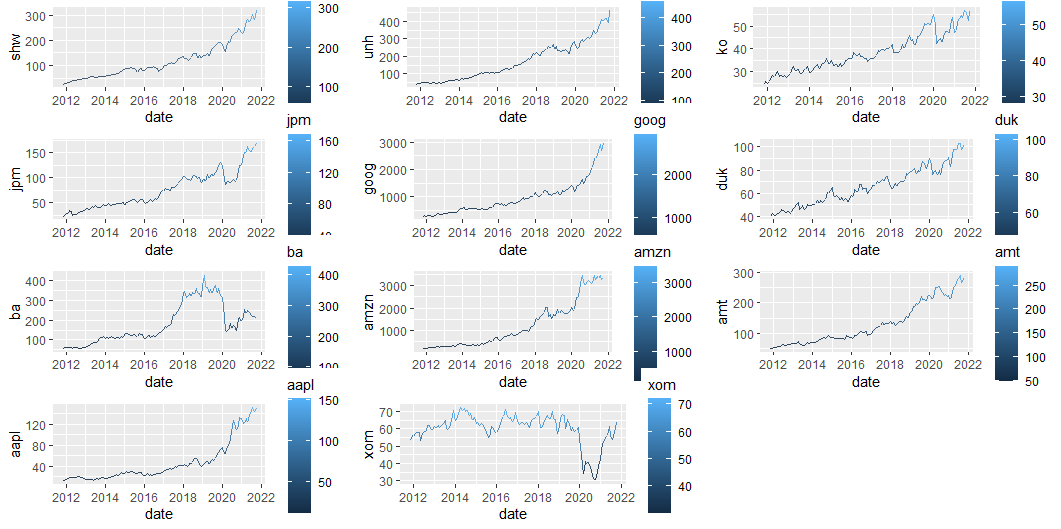
****

*Figure 3: Dataset containing monthly log returns for selected portfolio*

Once the team had their datasets, they were able to begin data analysis to explore their research question. The team performed their analyses in RStudio. First, the team applied data cleaning methods to ensure that their datasets were in the correct, most efficient format for usage. This included ensuring the data was of the correct class, headings were in a format that was easy to reference throughout the code, and that the dates and timelines were correct and consistent amongst the datasets. From here, the team used preliminary data analysis techniques such as displaying summary statistics and plotting the changes in close price of each of the securities in their portfolio over the 10 year timeline. These methods are helpful for the team to get a quick look at the behaviour of the selected stocks over time and verify that their performance grew over the time period of interest. Additional analysis was done using investment and regression techniques. The team looked at investing set amounts of money into each of the portfolios to determine how much they would have at the end of the time period. Regression analysis was used to plot log returns and compare the team’s portfolio to the SPY. Finally, the team used clustering methods to identify the groups of similar objects in the datasets to further compare the team’s selected stocks against the popular indices. Each of these mentioned methods was important in obtaining results which helped the team to come to their conclusions within this report.

**Results**

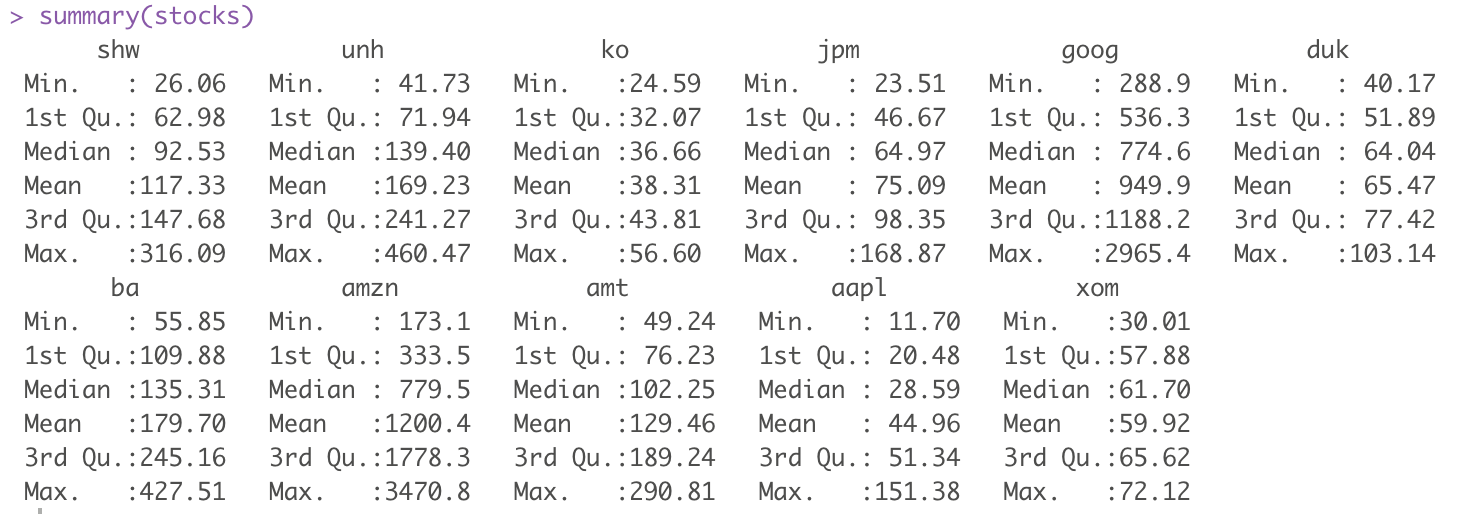
The team used the methods described in the previous section to carry out an analysis on the datasets and to answer their research question. In this section, the results obtained from the analysis will be presented and discussed.

The first step the team took to analyze the data after loading the datasets in R was plotting the relationship between close price and the date over the last 10 years to get a better idea of each individual stock’s performance in the selected portfolio. By comparing this portfolio's data to the benchmark indices such as SPY and DIA, preliminary assumptions can be made about their performance. Graphs showing the individual performance of each of the stocks across all sectors can be seen in the figure below. 

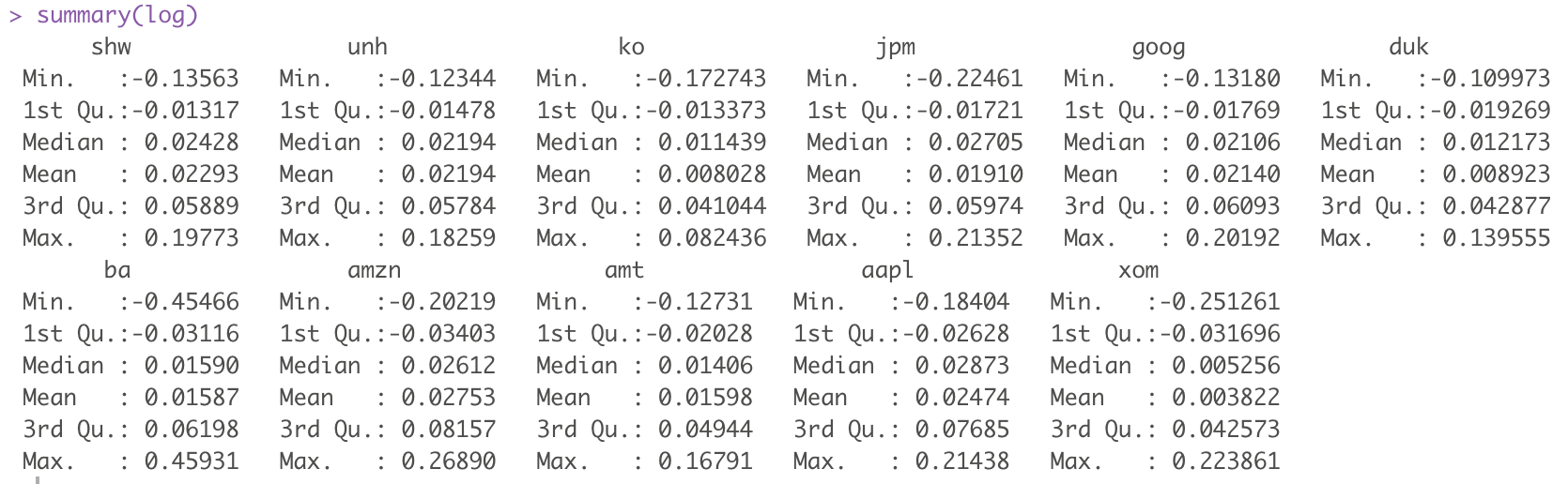
*Figure 4: Close price for 11 selected stocks over 10 years*

From the above graphs, we can see that all of the selected companies have grown in the last decade, with only major falls in close price occurring for ExxonMobil (XOM) and Boeing (BA). These falls can be seen between the years 2020-2021 which is synonymous with the beginning of the COVID-19 pandemic. While this is an interesting phenomenon, the team has determined that this is not telling of the overall performance of the stocks as an increase was seen following the “end” of the pandemic. Therefore, the team opted to keep these companies in their analysis.

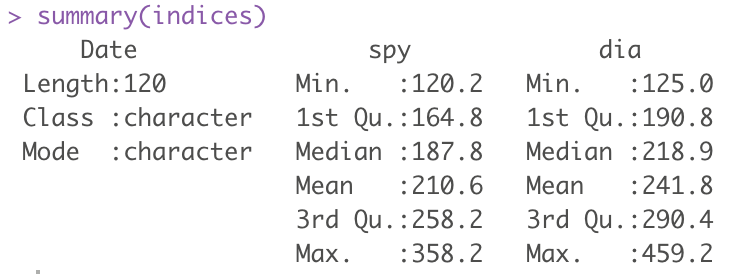
Next, the team looked at the summary statistics of each of the datasets to show a direct comparison of the distributions of the data. The summary statistics of the stocks prices, log returns, indices are shown below.



*Figure 5: Summary of stock prices*

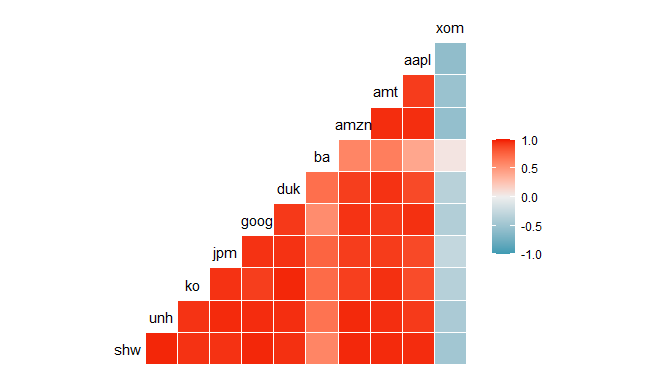


*Figure 6: Summary of stock log returns*



*Figure 7: Summary of indices*

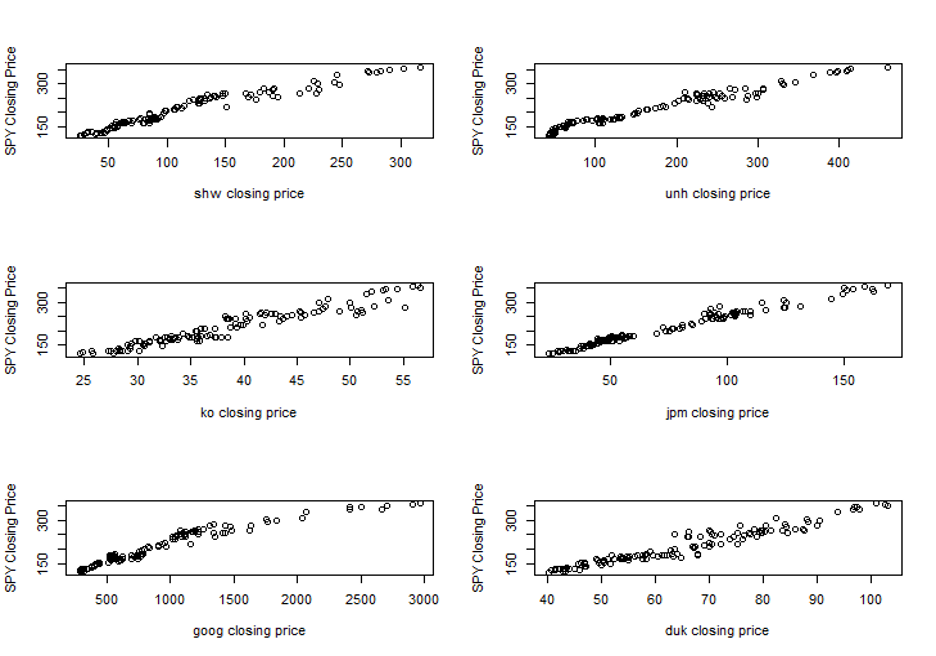
To take the preliminary analysis a step further, the team plotted the correlation of each of the individual stocks. The outputted correlation graph between all stock prices in the data set are shown below.



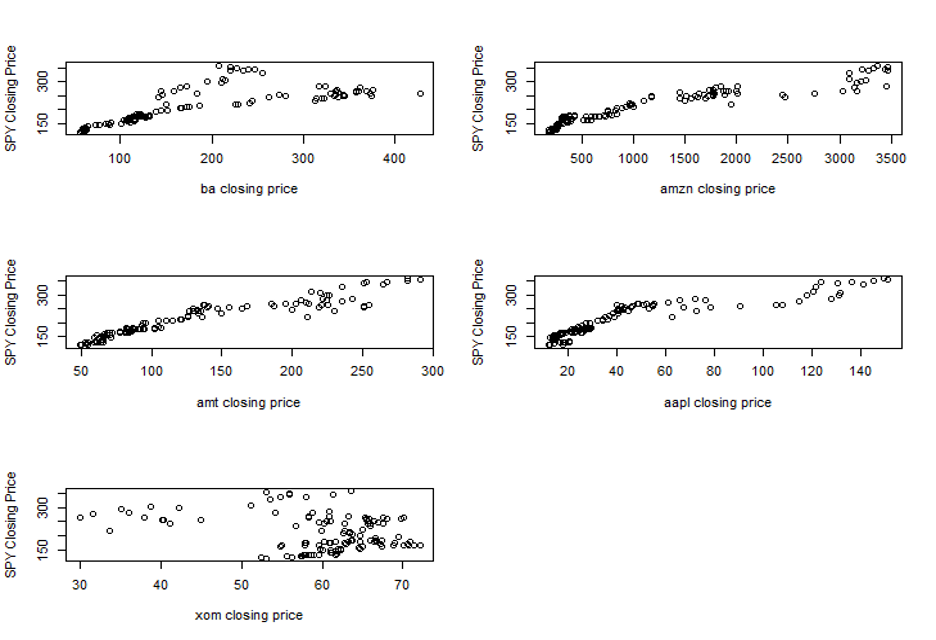
*Figure 8: Correlation between 11 stocks prices*

From the above correlation plot we can say that all stocks are highly correlated. All of the companies, aside from ExxonMobil which shows a strong negative correlation with the remainder of the companies, show a strong positive correlation with one another. This result is consistent with the graphs displayed in Figure 4 which shows the performance of the stocks. Therefore, we are able to move forward with the analysis and can safely conclude that the portfolio that the team generated will be able to provide valuable information when compared against top performing stock indices.

To better understand the nature of the portfolio and how well it is doing, the team compared the stocks to the benchmark funds like SPY and DIA. The scatterplots shown in Figures 9 and 10 below display the close price of the eleven selected stocks in the team’s portfolio plotted against the SPY close price.

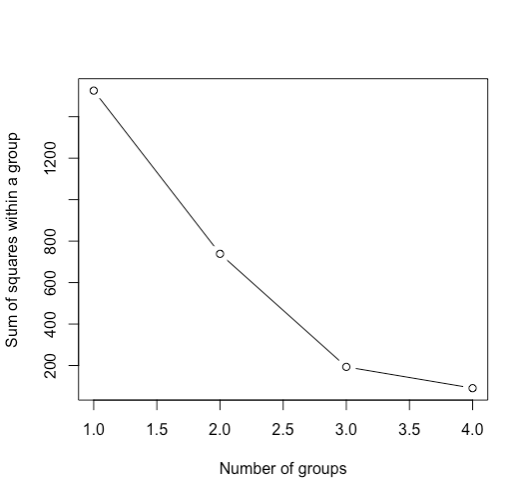


*Figure 9: Close price of 6 selected stocks against SPY close price*

**

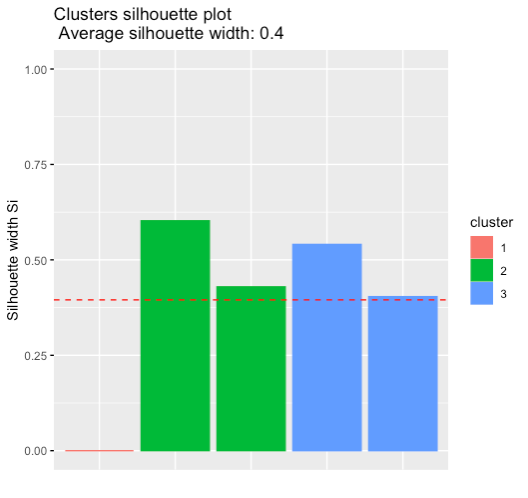
*Figure 10: Close price of remaining 5 selected stocks against SPY close price*

Looking at the distributions of the close prices of each of the stocks in the team’s portfolio, it can be seen that there is a strong correlation between all of the individual stocks closing prices and the SPY closing prices aside from ExxonMobil. As aforementioned, this result is expected due to the nature of the closing prices over time. Additionally, the correlation seems to be in the negative direction, which is also expected based on the correlation results within the team’s portfolio in Figure 8. The team also plotted monthly log returns for each individual stock which can be seen in the Appendix section.

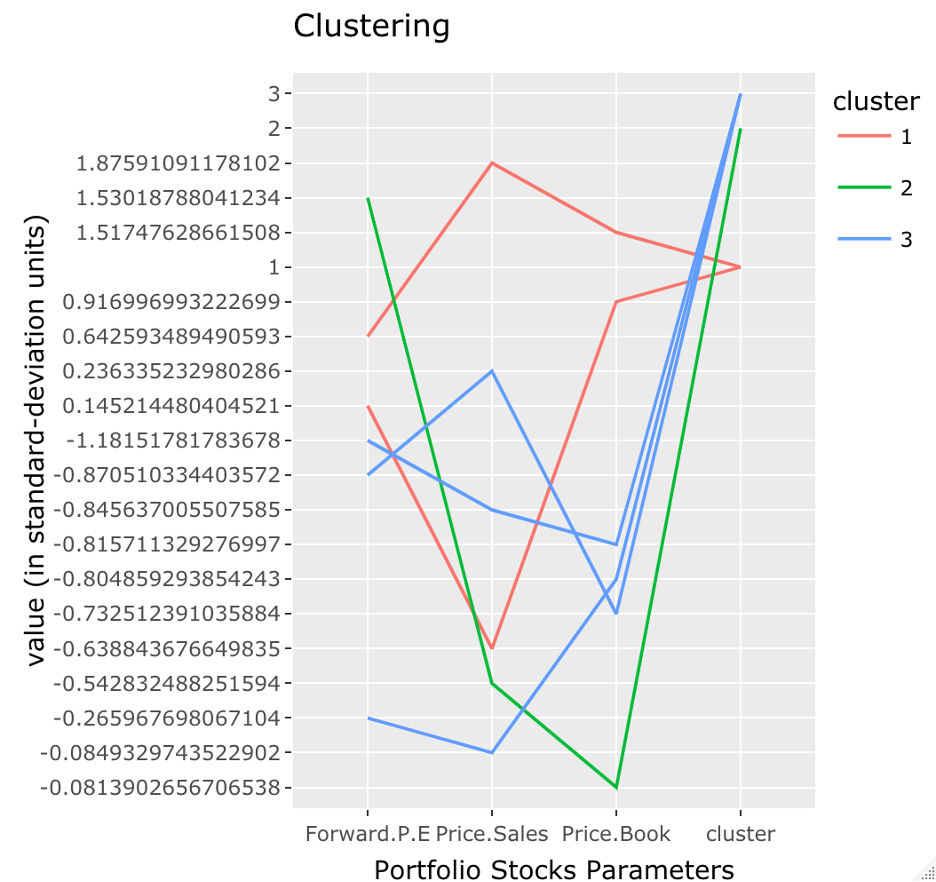


*Figure 11: Sum of squares vs. Number of groups*

Analyzing the above graph from right to left, we can see that the distance between the sum of squares within a group increases. That proves that there is reduction in the clustering. And the below graph shows the final result of the clustering.

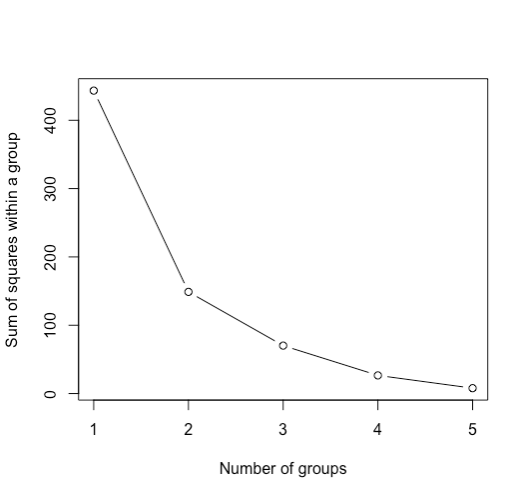


*Figure 12: Clusters Silhouette Plot*

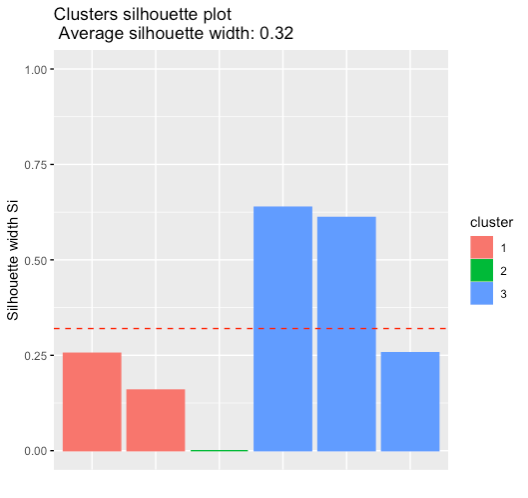


*Figure 13: Clustering Analysis with forward P/E ratio, price/sales, and price/book*

The above graph is for the clustering analysis using the parameters of Forward P/E ratio, Price/sales and price/book. The purpose of this chart is to determine the patterns in the data. The low values for the three ratios are as expected since we are analyzing the top performing stocks in each sector.

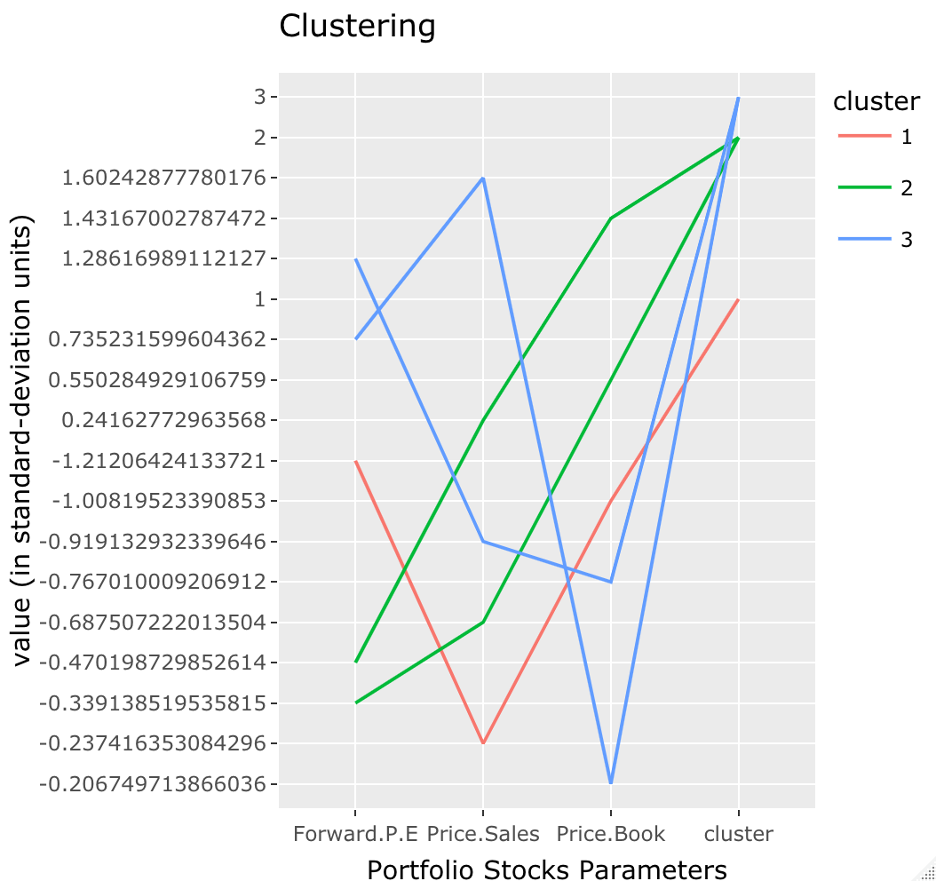


*Figure 14: Sum of squares within a group vs. Number of groups, adjusted*



*Figure 15: Clusters silhouette plot, adjusted*

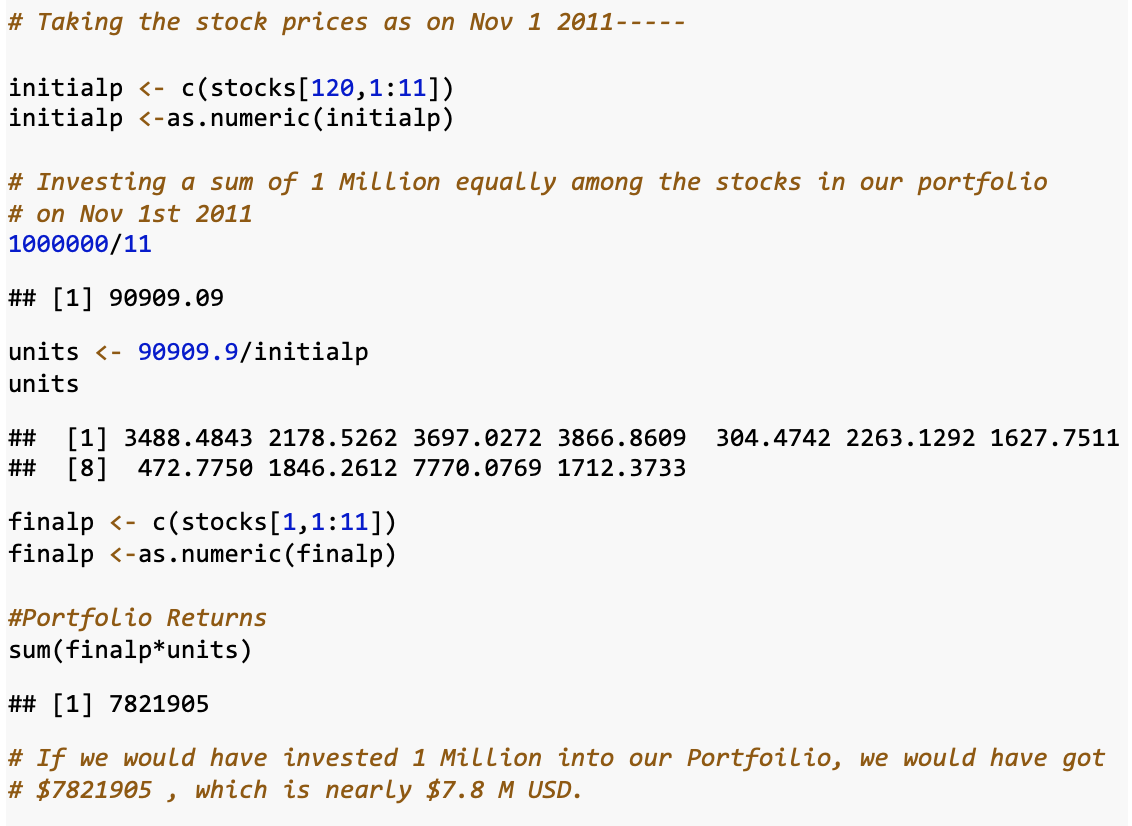
The graphs seen in Figures 14 and 15 are for the same parameters, however, the silhouette plot width was adjusted by 0.32.



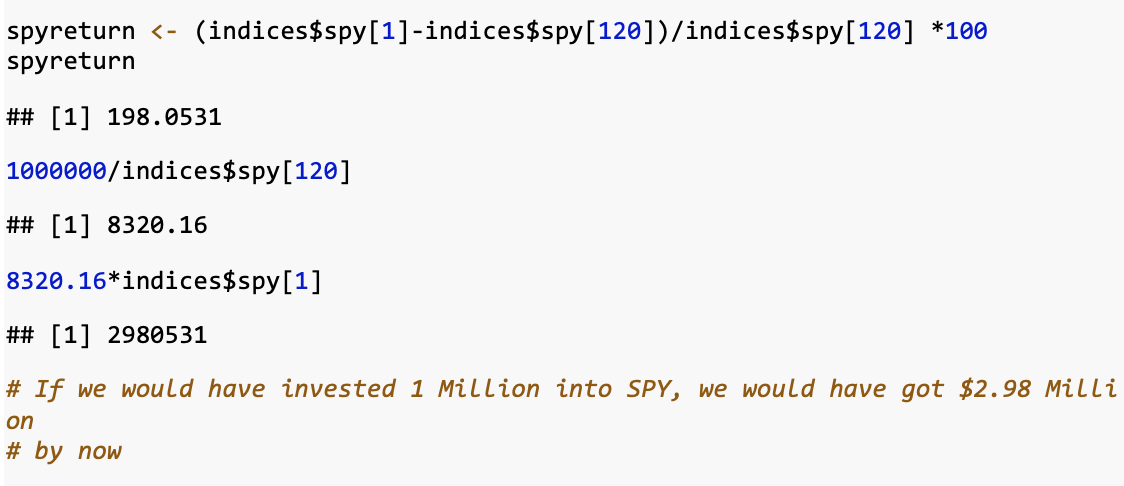
*Figure 16: Clustering Analysis with forward P/E ratio, price/sales, and price/book, adjusted*

The purpose of this chart is to determine the patterns in the data. From the graph we can visualize the return based on the 3 parameters. How these parameters affect the change of the stock price.

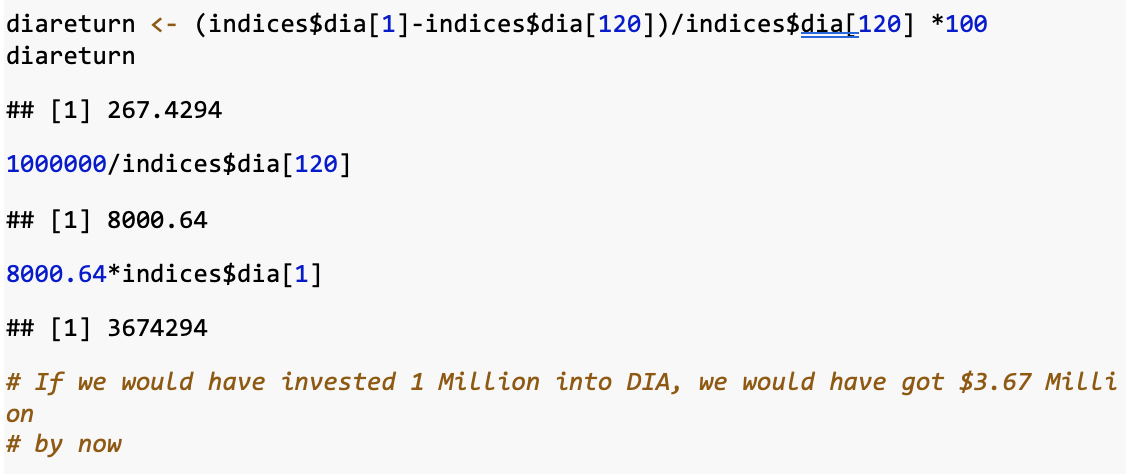
As an additional method of analysis, the team looked into different investment strategies with their portfolio and the two chosen stock indices. This was done with an equally weighted portfolio in which the team divided a sum of $1,000,000 equally among these 11 companies and invested it on the basis of closing price as on Nov 2011. We found that our portfolio would have grown nearly 8 folds amassing a net worth nearly $8,000,000. If we would have invested 1 Million into SPY, we would have got $2.98 Million by now. If we would have invested 1 Million into DIA, we would have got $3.67 Million by now. The methods for completing this analysis can be seen in Figure 17, 18, and 19 below.



*Figure 17: Investment strategy for the team’s portfolio*

**

*Figure 18: Investment projection using the SPY*

**

*Figure 19: Investment projection using the DIA*

**Conclusion**

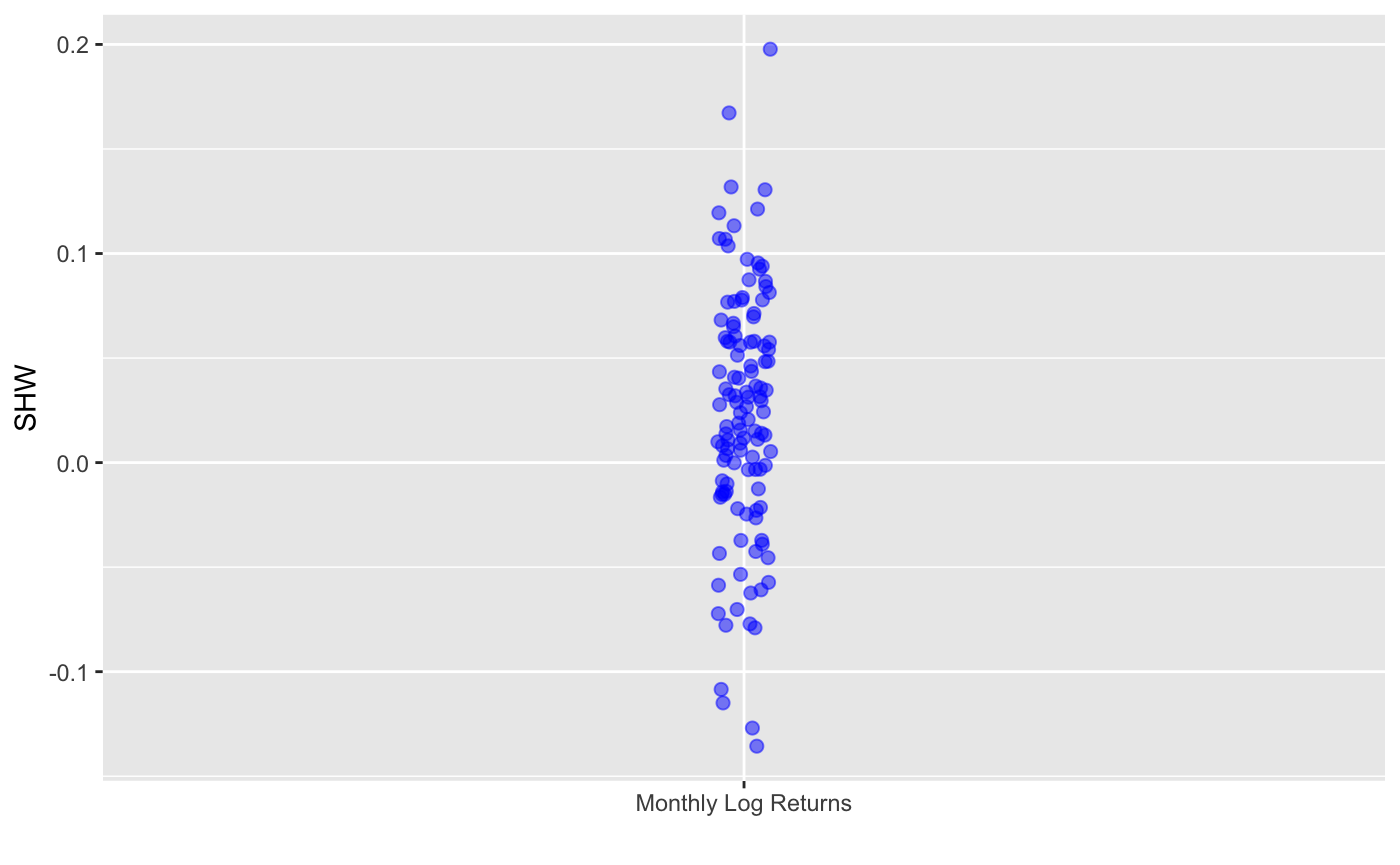
Based on the analysis completed within this report, we can see that the performance of a portfolio made up of stock data from companies from each of the 11 GICS sectors compare well with some of the most popular stock market indices. Through methods such as data cleaning, investment strategies, and clustering analysis, the team was able to successfully obtain results which can help to answer the overall research question: How does the performance of stock market indices of the leading companies in the United States compare to indices generated with top performing companies across all sectors over the last 10 years? With this being said, additional topics that could be explored by the team had there been more time were looking at rebalancing theory rather than utilizing an equally weighted portfolio in the analysis or by adding more companies into the portfolio from GICS sub-sectors in order to make the portfolio more diverse. This could have given the team more insights as to the interactions between the portfolio built from the sectors against stock indices such as the S&P 500 and Dow Jones. In conclusion, the methods presented were helpful in performing the necessary analysis to answer the question at hand.

**References**

1. <https://www.fool.com/investing/stock-market/market-sectors/>
2. <https://finance.yahoo.com/>

**Appendix**

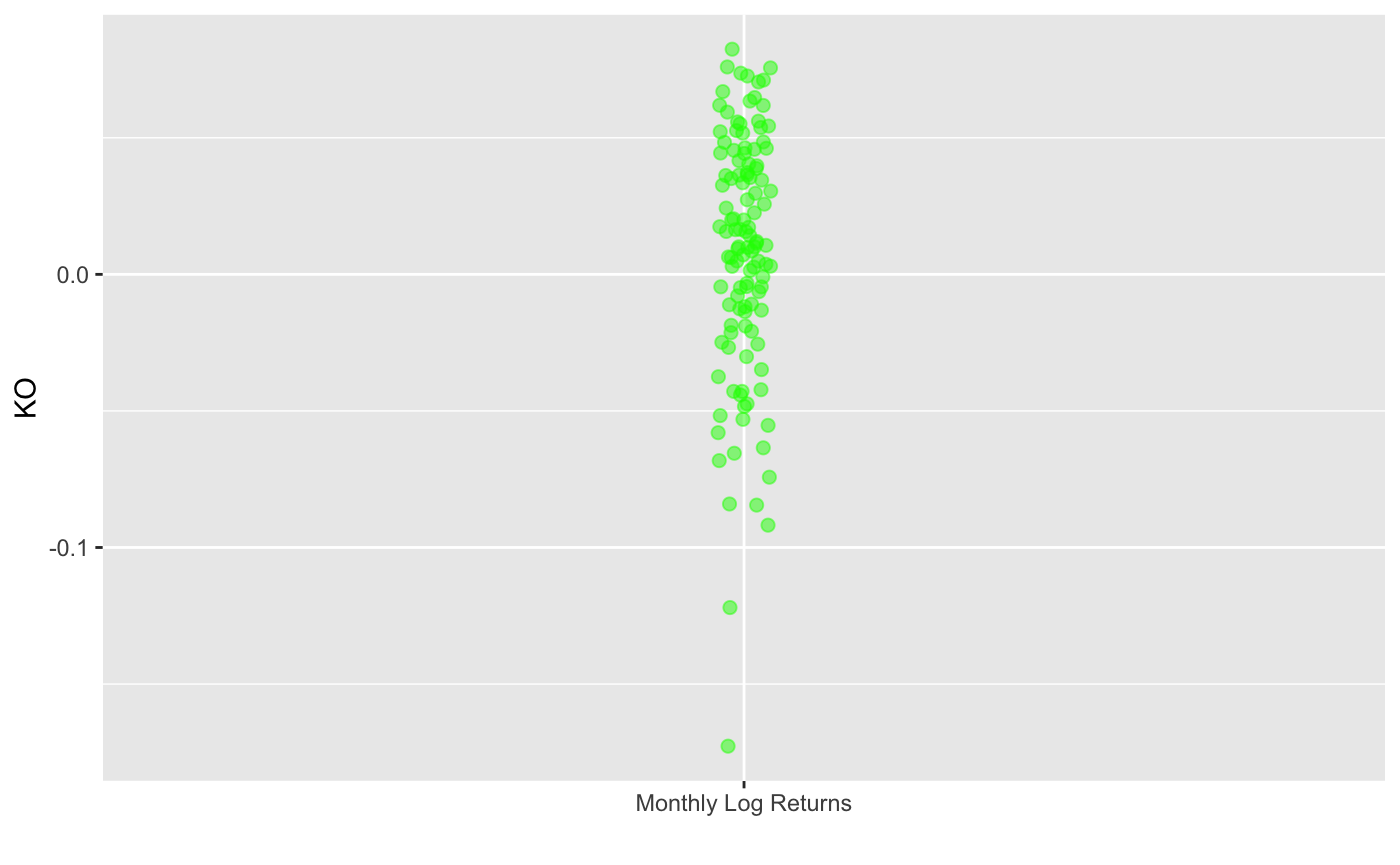
*Appendix 1: Monthly Log Returns of Sherwin-Williams (SHW)*

**

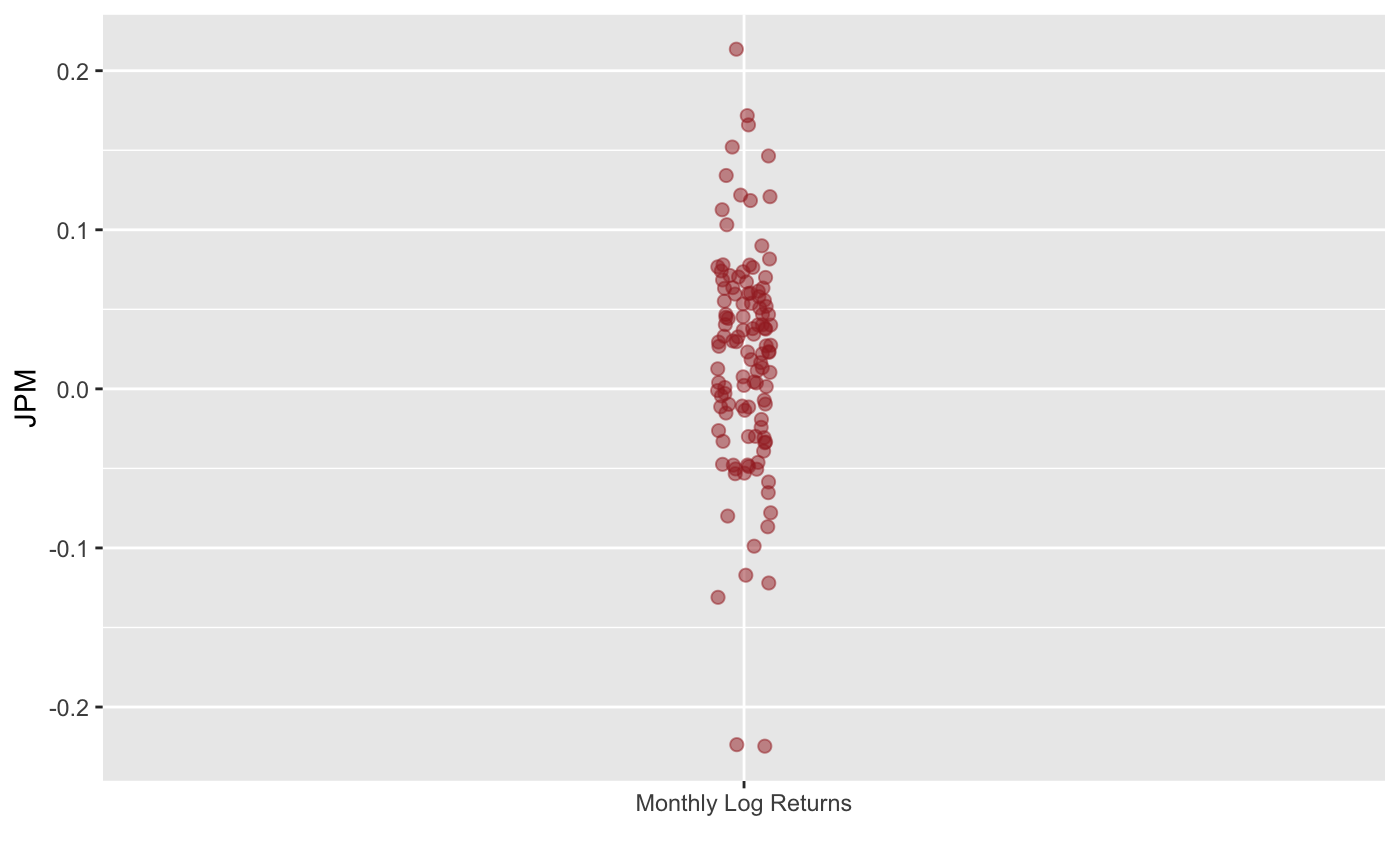
*Appendix 2: Monthly Log Returns of UnitedHealth Group (UNH)*

**

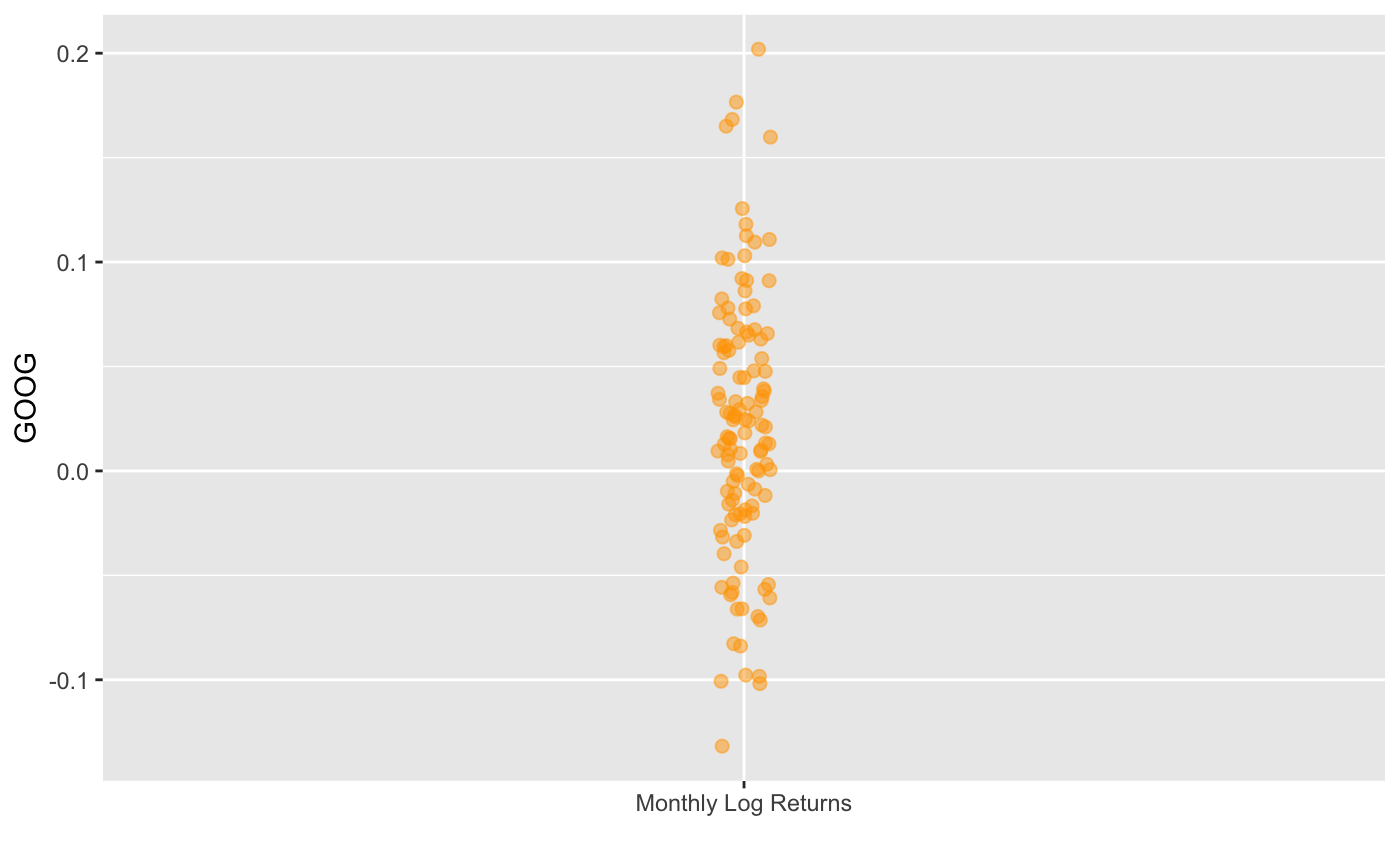
*Appendix 3: Monthly Log Returns of Coca-Cola (KO)*

**

*Appendix 4: Monthly Log Returns of JP Morgan (JPM)*

**

*Appendix 5: Monthly Log Returns of Google (GOOG)*

**

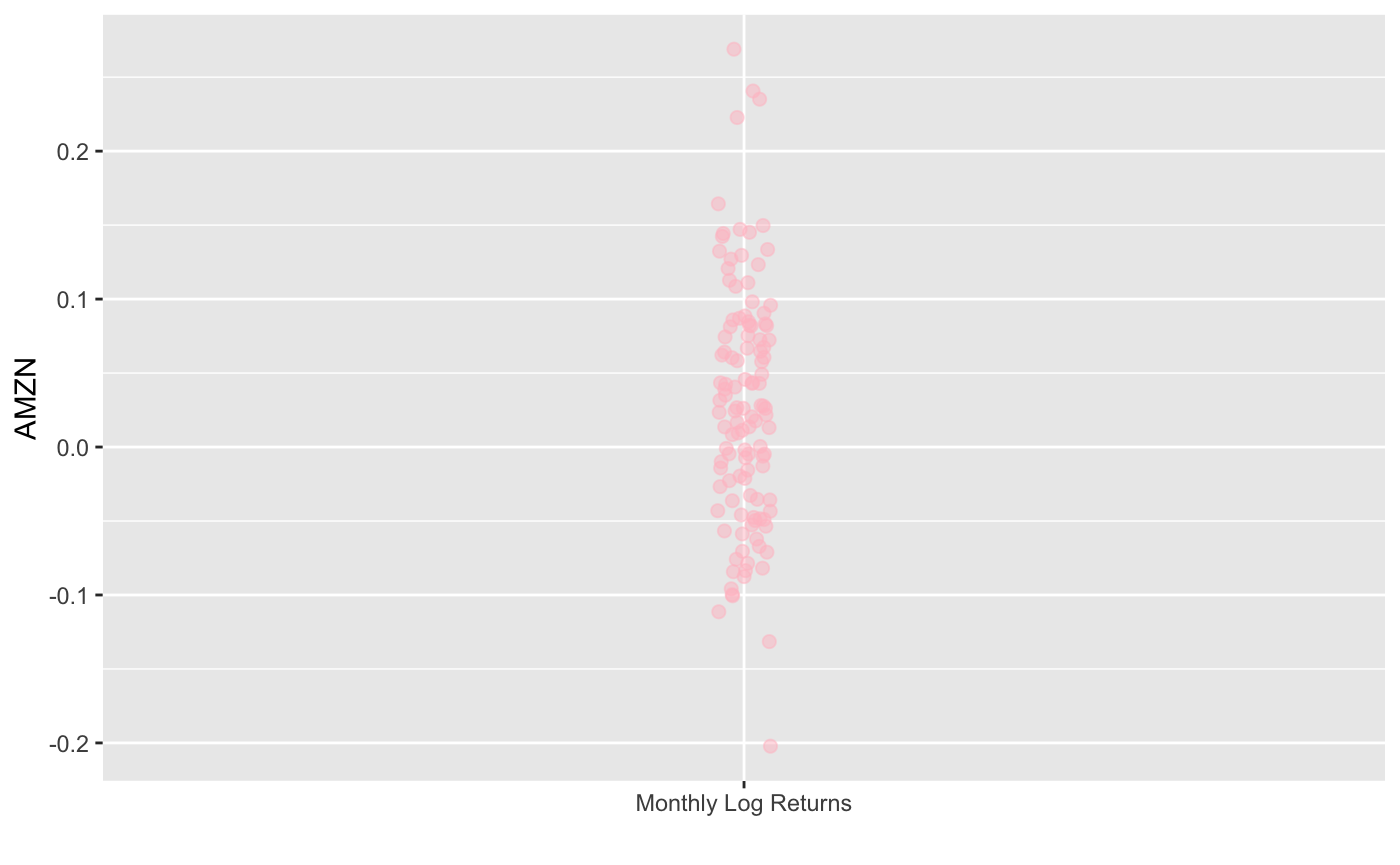
*Appendix 6: Monthly Log Returns of Duke Energy (DUK)*

**

*Appendix 7: Monthly Log Returns of Boeing (BA)*

**

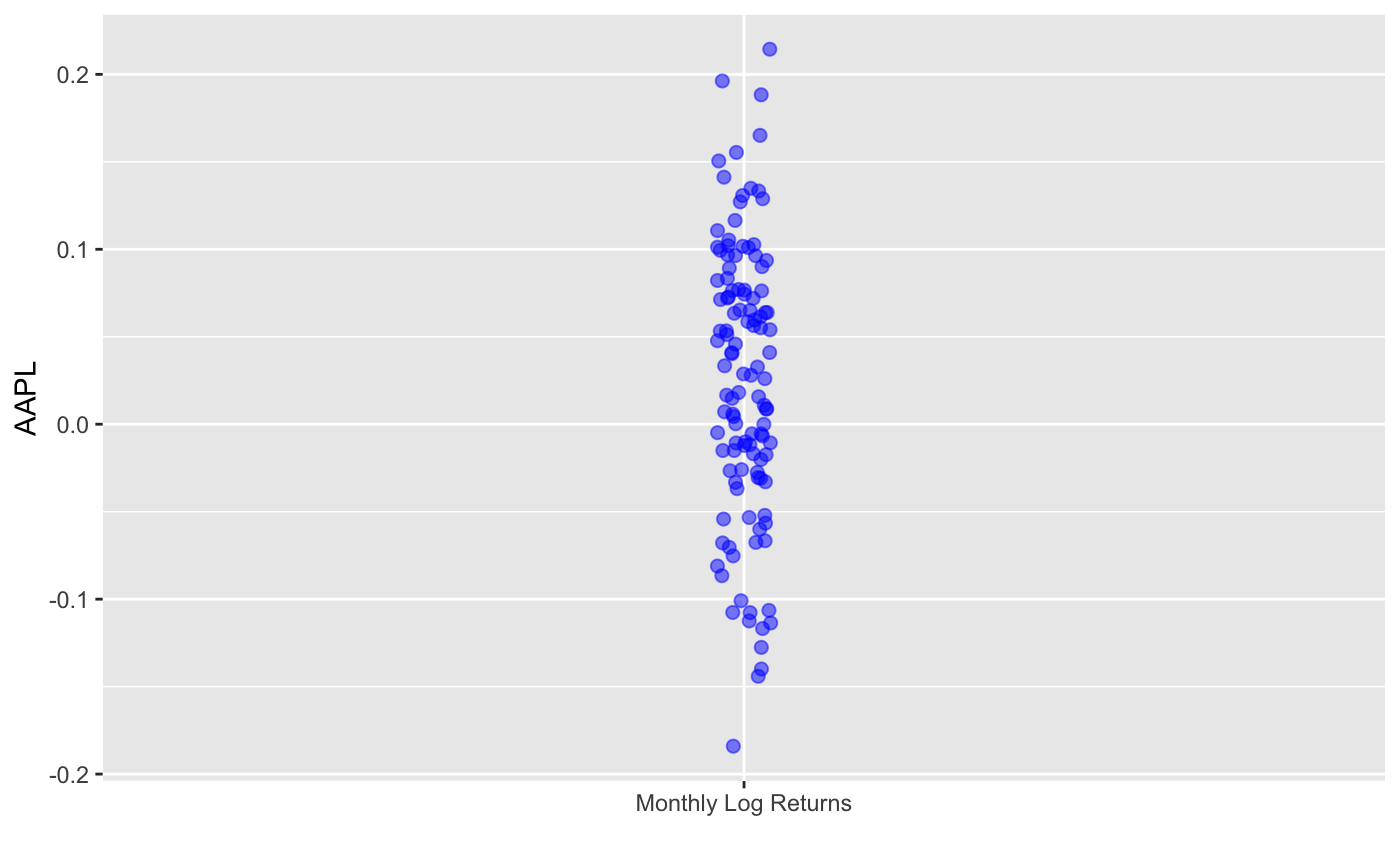
*Appendix 8: Monthly Log Returns of Amazon (AMZN)*

**

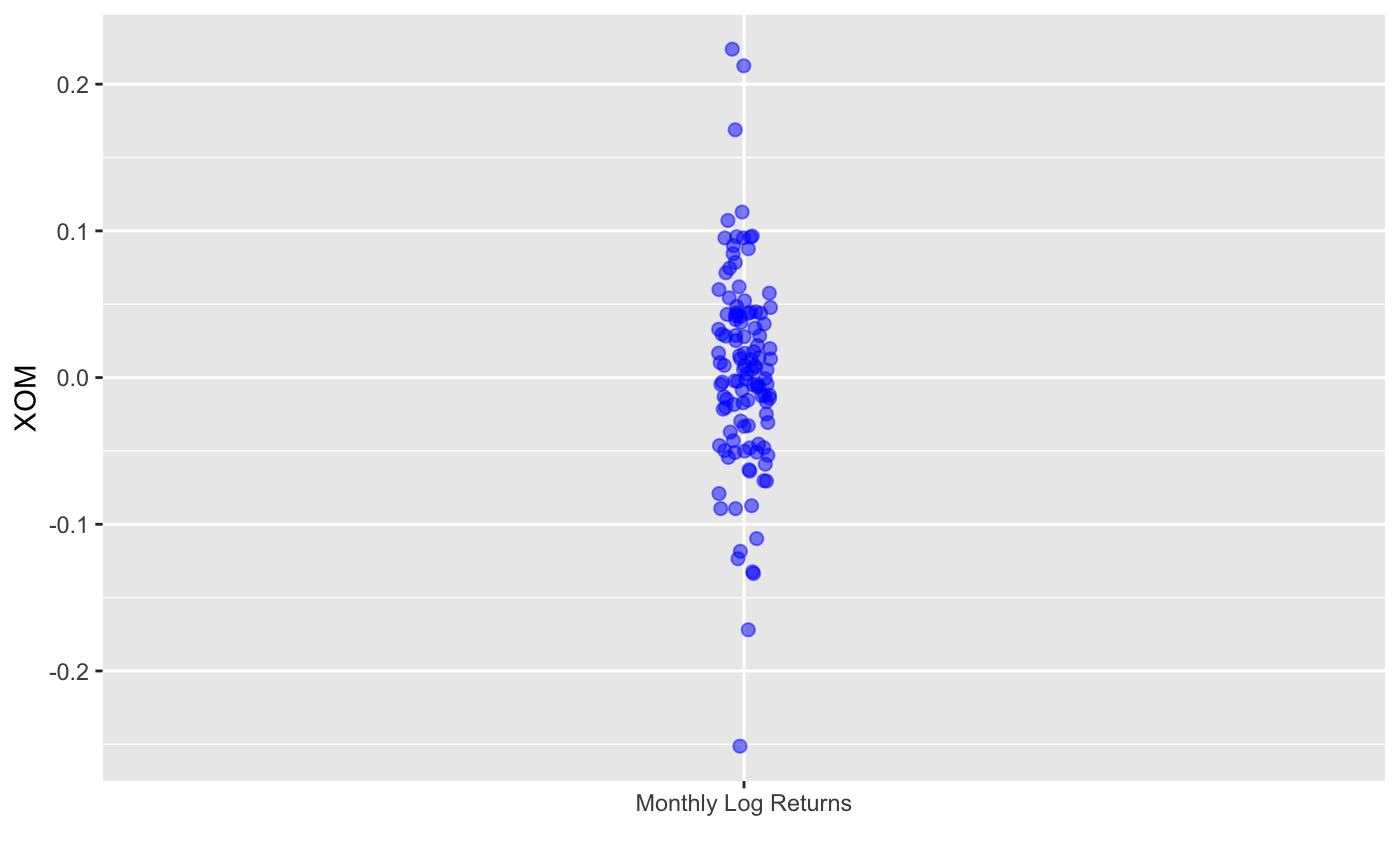
*Appendix 9: Monthly Log Returns of American Tower (AMT)*

**

*Appendix 10: Monthly Log Returns of Apple (AAPL)*

**

*Appendix 11: Monthly Log Returns of ExxonMobil (XOM)*

**