Arjun Koshal

QF 104 - C

Professor Cao

21 December 2020

Final Project Report

My final project for this course was based on analyzing the stock data from 2015 to 2020 of the top stocks in the S&P 500. I felt that it would be more beneficial to take the files from Yahoo finance compared to creating 6 different CSV files and having to import each of them separately, therefore I based my program around this mechanism. I began by importing all the packages that I would be utilizing in the program. I decided to use my resources from the internet to view how this would be possible and figured out the tidyquant package allows us to use the stock symbol directly from Yahoo finance. Quantmod stores the symbols with their own names, however, we can bypass that by setting the warnings to false. Once I set the warnings to false, I was able to take the stock prices directly from Yahoo Finance.

I want to display these stock prices graphically, similar to how Yahoo finance does it. My plan was to plot the date on the x axis and the price on the y axis. I decided to take the top 6 stocks from the S&P 500 and display them, which were Apple, Amazon, Facebook, Google, Microsoft, and Tesla. I picked the time frame 2015 to 2020 as I felt that time frame displayed the biggest and most recent change in the stock prices. I displayed the head of the prices and it was noticeable that only the Apple stock was being shown, therefore I had to change it so that we could see every stock, not only just the Apple Stock. By utilizing the slice function, I was able to remove the feature where the data was stacked on top of one another and allow the first row of each different stock to be shown.

I decided to group the stock data by each different company, as that was definitely the clearest way of showing all the data. Once I split the data into different categories, I utilized ggplot2 in order to display the visualizations in RStudio. I labeled the x axis as the date in years and the y axis as the adjusted price. Plotting based on full year ($Y) was the most clear and concise method in order to make sure the data was not crammed together. I needed to use the facet_wrap function as we learned in class to make the data fit to their stock prices. Because the Amazon stock was in the thousands range, and Microsoft was in the hundreds range, the function facet_wrap allowed the price along the y axis to be adjusted to the proper visualization.

I decided to add some features in the program to make it look decent. The legend on the right side of the program was preprogrammed to follow the colors from the symbols in Yahoo finance, however, manipulating the labs function, I was able to change the colors and match them to the different stocks. I also changed the names from the name of the stock to the name of the company so that more people would be able to recognize which company the charts were referring to. I added a title and a subtitle relating to the data from the stocks. I felt as if the program was well organized and utilizing most of the information in the course allowed me to generate such a visually pleasing graphic of the top stocks in the S&P 500.

If I had more time to work on the project, I would definitely like to add more features and allow the user to see the open, closing, low, and high prices graphed. I would also like to create a large text for the title and subtitle. It felt as if the title was too small and in a bad position. Other than that, I felt as if the project was well done and I felt as if I incorporated several skills that I acquired from enrolling in this course.

Console

```
Console  Terminal  Jobs
~/Desktop/R Studio/QF 104/
> head(prices)
# A tibble: 6 x 8
  symbol date        open  high   low close    volume adjusted
  <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>    <dbl>
1 AAPL   2015-01-02  27.8  27.9  26.8  27.3 212818400     24.9
2 AAPL   2015-01-05  27.1  27.2  26.4  26.6 257142000     24.2
3 AAPL   2015-01-06  26.6  26.9  26.2  26.6 263188400     24.2
4 AAPL   2015-01-07  26.8  27.0  26.7  26.9 160423600     24.5
5 AAPL   2015-01-08  27.3  28.0  27.2  28.0 237458000     25.5
6 AAPL   2015-01-09  28.2  28.3  27.6  28.0 214798000     25.5
>
> # Step 4. I decided to group the stock data by each different company, as that was
> # definitely the clearest way of showing all the data. The slice function lets
> # us view the first row of each different stock (or symbol in this case).
>
> prices %>%
+   group_by(symbol) %>%
+ slice(1)
# A tibble: 6 x 8
# Groups:   symbol [6]
  symbol date        open  high   low close    volume adjusted
  <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>    <dbl>
1 AAPL   2015-01-02  27.8  27.9  26.8  27.3 212818400     24.9
2 AMZN   2015-01-02 313.  315.  307.  309.    2783200    309.
3 FB     2015-01-02  78.6  78.9  77.7  78.4  18177500     78.4
4 GOOGL  2015-01-02 533.  536.  528.  530.    1324000    530.
5 MSFT   2015-01-02  46.7  47.4  46.5  46.8  27913900     41.5
6 TSLA   2015-01-02  44.6  44.7  42.7  43.9  23822000     43.9
```

Graphs