Investigation and Visualisation of Stock Data

Postgraduate Diploma in Science in Data Analytics 2019

Student No: 18179797

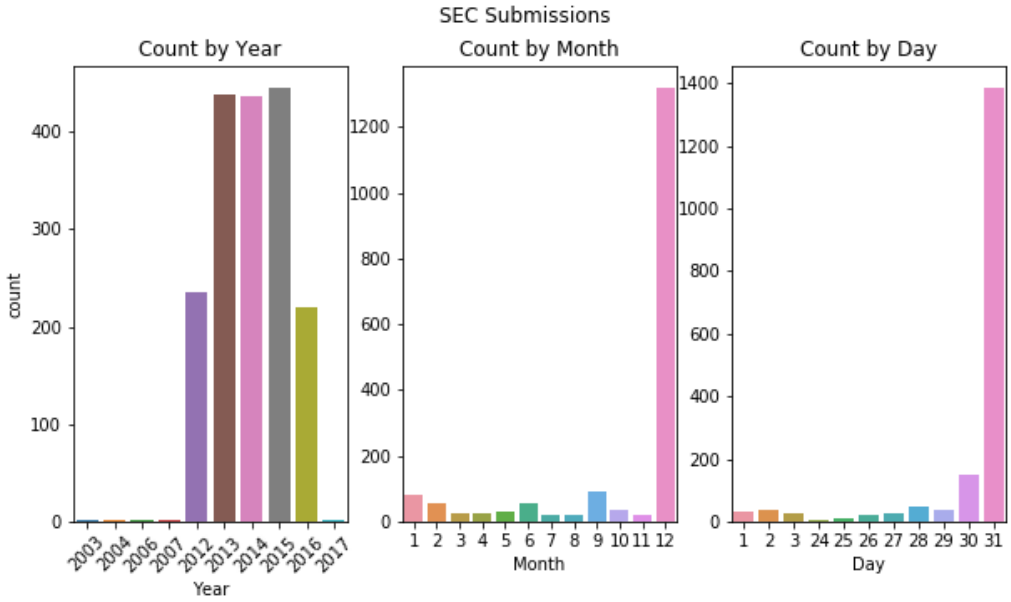
Introduction

Data Sets

In this project, I have downloaded the daily market prices and the fundamental data of over 450 stocks that are listed in the S&P 500 on the New York Stock Exchange (NYSE). The data was acquired from the Kaggle website [1]. Below I list the three data files acquired and mention a few pertinent columns in each file:

securities.csv: Ticker symbol, company name, industry sector
fundamentals.csv: Ticker symbol, Net income, total revenue, log-term debt
prices-split-adjusted.csv: Ticker symbol , close price, volume

The daily price data covers the period 2010-2016 and is of good quality. However, the fundamental company data, which was originally obtained from filings with the Securities and Exchange Commission (SEC) is patchy. To illustrate this, below I show a breakdown of the dates that these filings were submitted to the SEC.



As you can see, we only have appropriate data for approximately 450 companies over the years 2013, 2014 and 2015. We shall primarily restrict our investigations to these years. We can also see that the majority of the submissions were recorded on the last day of

December. [I went with bar plots above as they are easily read and interpreted. The colour scheme is random though and could likely be improved].

Tools

All but one of the figures presented in this project were generated using python libraries (matplotlib and seaborn being the primary ones). A Jupyter notebook has been submitted for this project. In this notebook, you will be able to observe all the data processing and code required to generate the figures.

The one exception is an interactive dashboard that was implemented in Tableau (The twb file is attached to my submission).

Interactive Dashboard

A simple interactive dashboard was created as a tool to view the data. All three data files were joined on the ticker symbols to create the data source. As can be seen below, the dashboard consists of two main views: A trend over time plot showing the historical daily closing price of a stock and a table displaying fundamental data for the selected years.

For Year / GICS Sector / GICS Sub Industry / Security

	2016	2015	2014	2013
	Information Technology Systems Software Microsoft Corp.	Information Technology Systems Software Microsoft Corp.	Information Technology Systems Software Microsoft Corp.	Information Technology Systems Software Microsoft Corp.
Cash and Cash Equivalents	11,470,620,000,000	9,858,390,000,000	15,274,778,000,000	6,702,648,000,000
Earnings Before Tax	34,801,262,000,000	32,609,334,000,000	49,018,840,000,000	47,665,624,000,000
Earnings Per Share	3,735	2,625	4,687	4,599
Estimated Shares Outsta..	13,961,356,603,773	14,418,836,241,611	14,621,950,375,937	14,759,619,157,087
Long-Term Debt	71,859,646,000,000	48,997,696,000,000	36,376,490,000,000	22,202,962,000,000
Net Income	29,598,076,000,000	21,484,066,000,000	38,894,388,000,000	38,522,606,000,000
Total Assets	341,288,828,000,000	307,419,664,000,000	303,740,608,000,000	250,963,422,000,000
Total Liabilities	214,430,114,000,000	166,313,418,000,000	145,541,200,000,000	111,864,094,000,000
Total Revenue	150,333,840,000,000	164,887,960,000,000	152,999,746,000,000	137,169,938,000,000

For Years
Multiple values

Fundamentals
Multiple values

There are four filters on the right panel that allow the user to interact with the data. On the top right we have the primary "Security" filter. From here you can select any one of the available companies/securities that you might want to view (In the picture above, we have selected *Microsoft*. Note that the company name is displayed on the top left automatically). Note also that this filter is applied to both views.

We then have the "Date Range" filter which allows the user to select the range of dates over which to display the historical price chart.

In the bottom right half panel, we have two remaining filters which control the selections shown in the table chart. The "For Years" filter allows you to select the SEC data for specific years (Above we have selected the years 2013-2016 inclusive). With the "Fundamentals" filter, one can also select the specific entries you wish to peruse e.g. Net Income, Total Assets, etc. There are approximately 80 such entries to choose from.

Using the above dashboard, the viewer can peruse the data for any company in a simple and user-friendly fashion. I am not entirely happy with it, particularly the styling and colouring. For reasons unknown, I was unable to change the colour, size and font of some of the chart labels (e.g. The label "Microsoft Corp." above). I would also liked to have added some text giving the return on the stock over your selected date range and I also entertained the idea of displaying the fundamental data as a radar chart (though that would likely be complicated). However, as it stands the dashboard has a simple and clean design.

Project Objectives

The goal of the project was to investigate and answer a couple of assertions that are often stated regarding US stock markets. These are:

1. Corporate America is taking on increasing amounts of debt over time.
2. As a means of increasing stock values, companies are in an ongoing process of buying back their own stock.

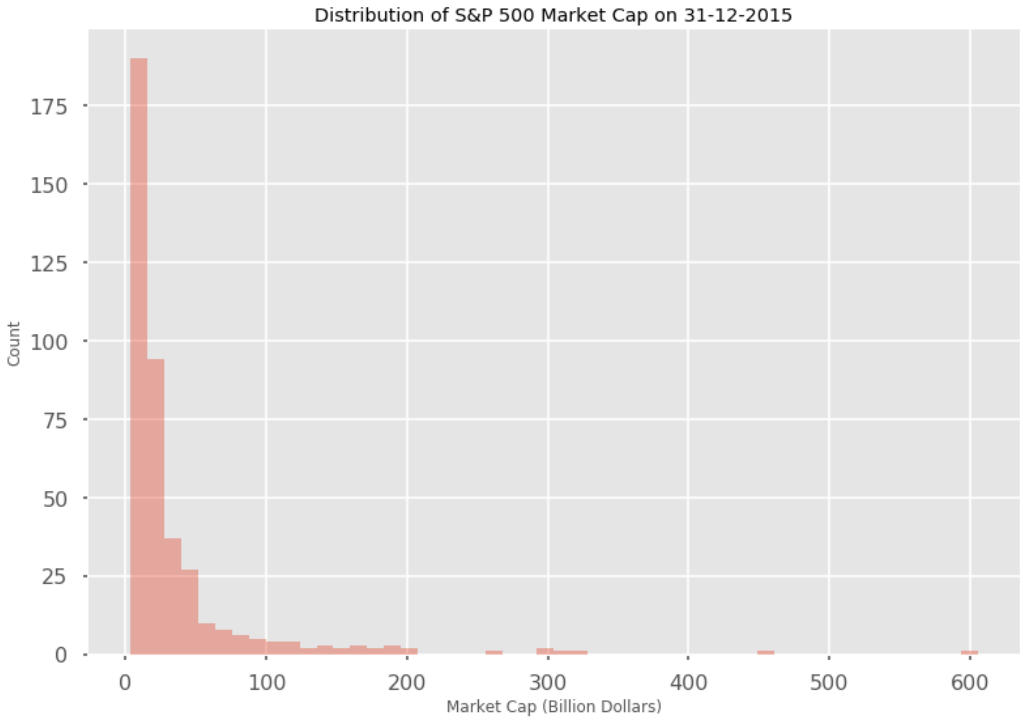
I will address the above assertions and illustrate other interesting relationships in the remainder of this report.

Market Capitalisation

The market cap of a company is simply the total dollar value of all outstanding publically listed shares in a company. As the share value fluctuates over time, the market cap of a company thus does so also.

Choosing the closing price of stocks on the last trading day in 2015, I calculated the market cap of each company (by multiplying the stock price by the *estimated shares outstanding* field in the fundamental data). Below I show the distribution of the market cap of 409

companies (Note that a small number had no entry for *estimated shares outstanding.* As such I was unable to calculate the market cap for these companies).

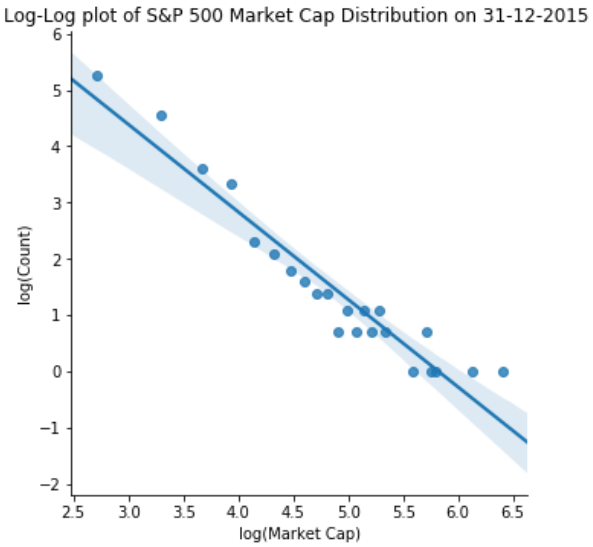


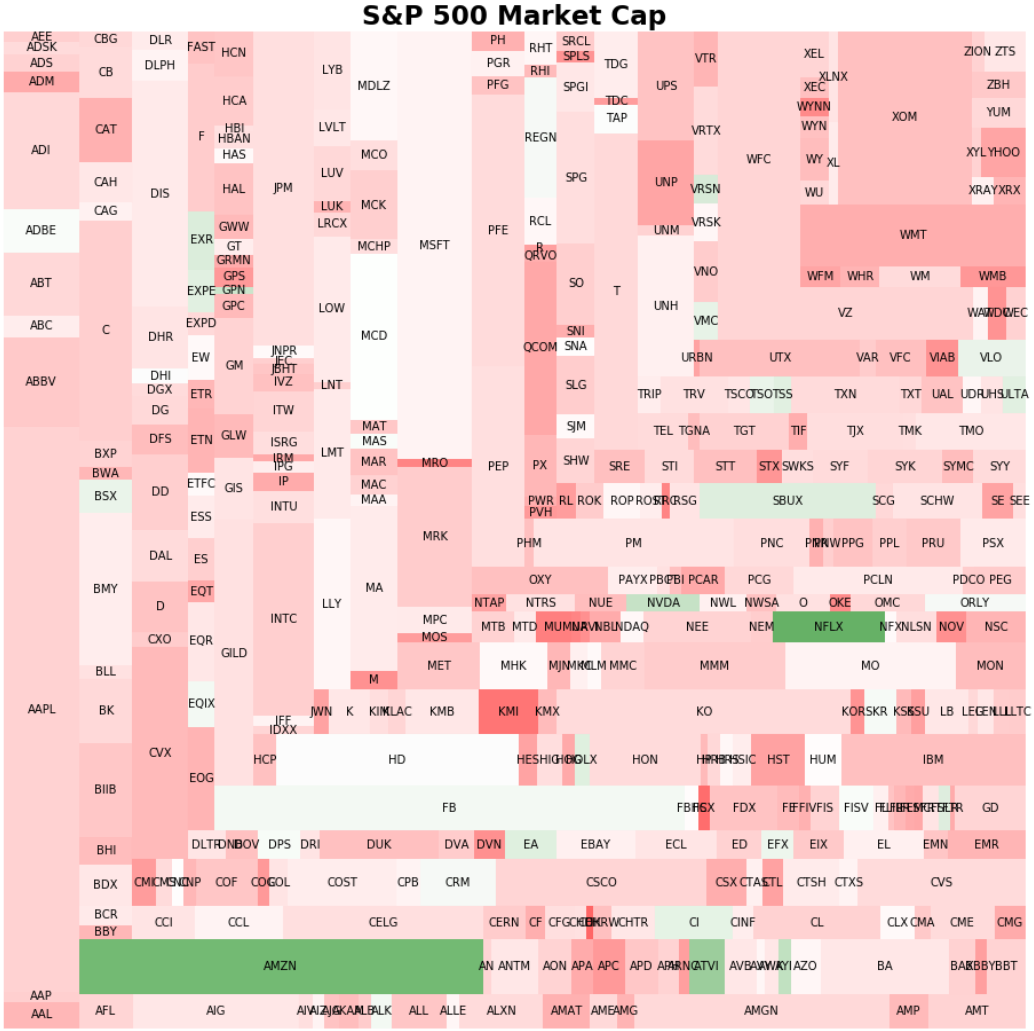
We can see that the distribution has a long tail where we have a very small number of extremely highly valued companies. For example, here are the five companies with the largest market caps:

Company	Market Capitilisation (Billion Dollars)
Apple Inc.	605.63
Microsoft Corp.	454.01
Exxon Moil Corp.	326.99
Amazon.com Inc	314.71
Wells Fargo	297.73

In fact, the histogram above appears to have a power law distribution as evidenced by the linear regression fit to the log-log plot of the distribution above (See figure below. I should have made that figure larger. Or the title shorter).

To get some idea of the relative market cap sizes of these companies I have also created a tree map. Here you can see the ticker symbols of all 409 companies. The size of a rectangle is proportional to the market cap of the company. The colour scheme illustrates the annual return on the stock for 2015. This colour scheme goes from dark red (highly negative return) to dark green (highly positive return). We can thus see that Netflix and Amazon had good stock returns in 2015.



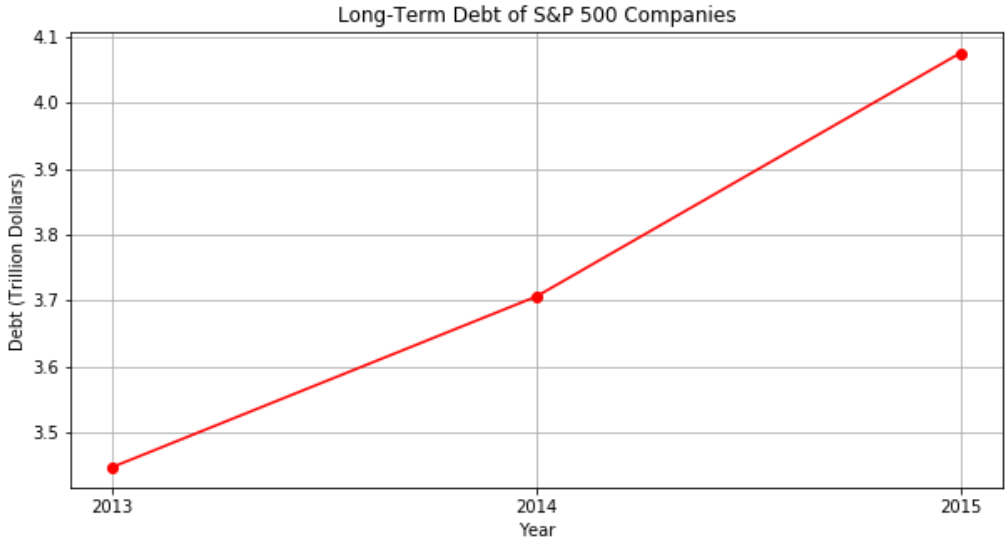


Note that I am not entirely happy with the colormap. I wanted to use green to denote any company that made a positive return and red to solely denote loss-making stocks. But I did not have the time to figure out how to anchor zero return to be at the centre of the colour

map. Other than that, I think it was a nice idea, though the diagram is a bit noisy. I did a similar tree map showing industry sectors rather than individual stocks, but decided against including it in the report. Though it was far less noisy, I regard the above diagram above serves in some way as a visual summary of my data set.

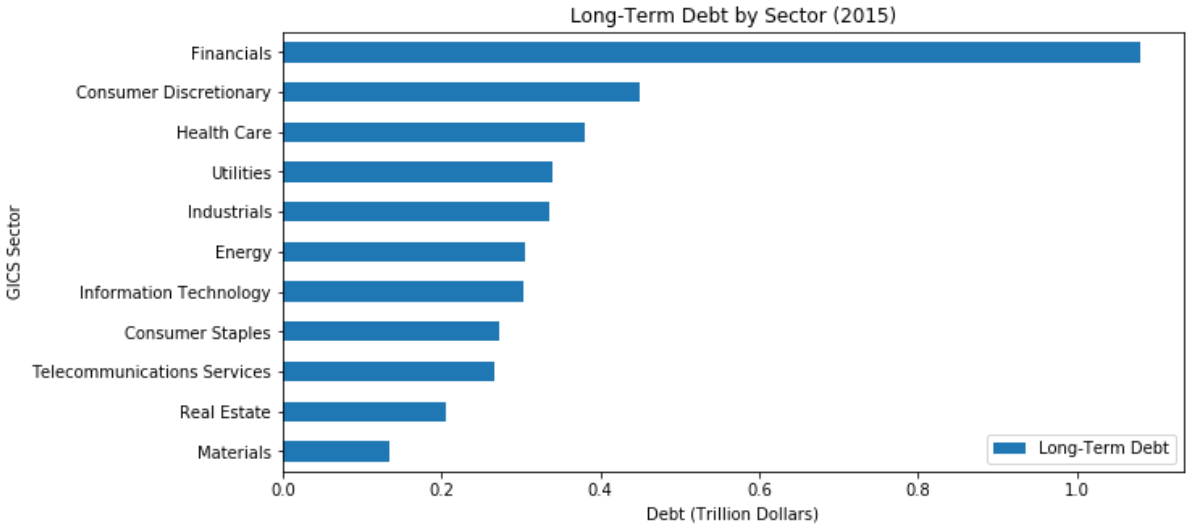
Corporate Debt

Interest rates have been at historic lows for the past decade. This has encouraged companies and households to load up on debt. As illustrated in the figure below, for the S&P 500 companies studied here, we observe that this is indeed the case.



We can clearly see that in the span of three years, the total long term debt of these S&P 500 companies increased from approximately 3.45 to 4.08 trillion dollars. [I added a grid to the above chart so that one could more easily estimate the middle point in the chart. I would liked to have deleted the vertical grid lines as they ruin the aesthetic somewhat, but no time now!]

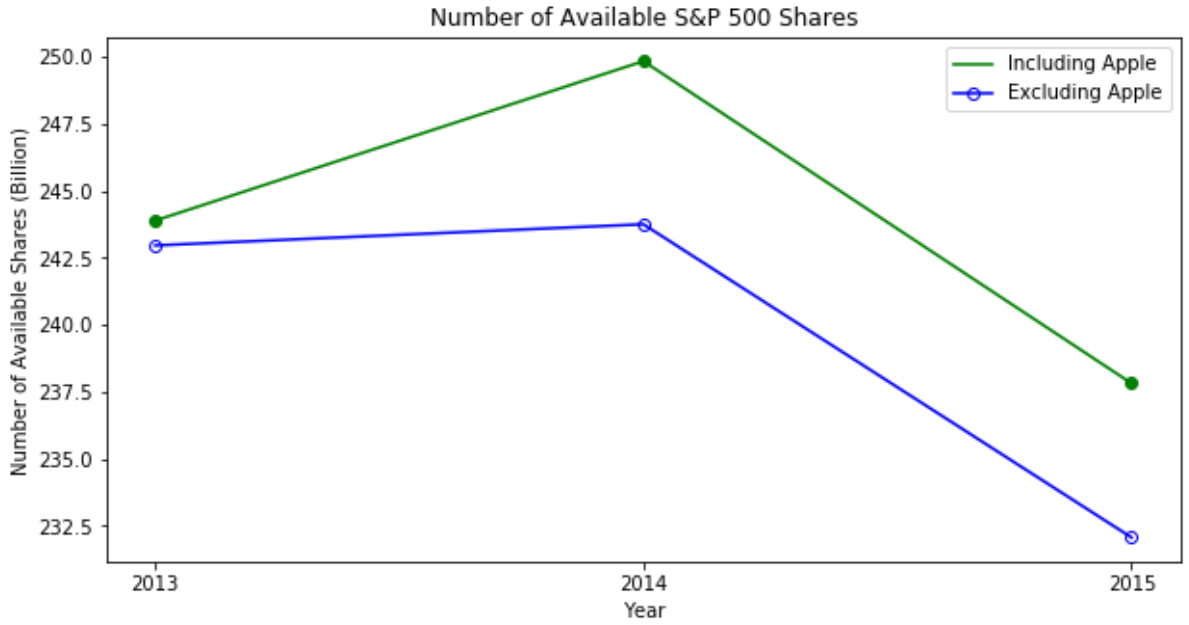
In the diagram below, we show how this debt as reported in 2015 is broken down by sector.



We can clearly see that the financial industry is by far the most heavily indebted sector. [The above horizontal bar chart displays the information clearly. A pie chart may have been a nicer alternative choice. Ideally I would liked to have added some colour to the above chart.]

Share Buybacks

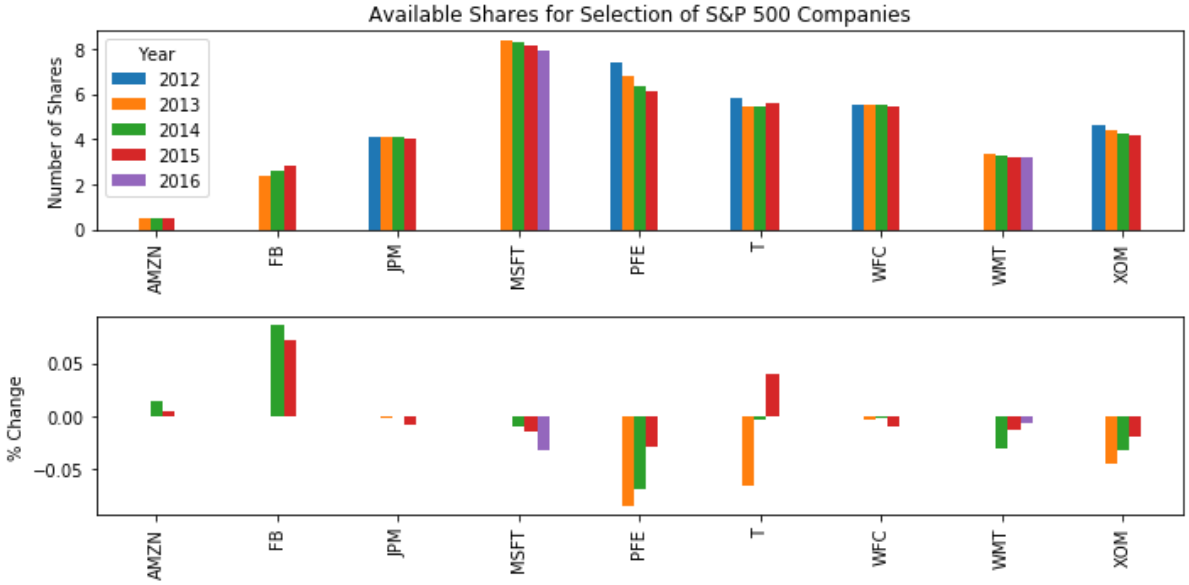
It has been claimed that the primary driving force behind the rise in US markets over the last decade has been share buybacks. That is, companies using their cash reserves or issued debt to purchase their own stock. This is done to drive up the price of their stock (and is a more tax efficient way of rewarding their shareholders than paying dividends). This naturally has the direct effect of reducing the number of shares available to purchase on the stock market. As such, I have used the field *estimated shares outstanding* to investigate this phenomenon.



In the diagram above, we show the total number of S&P 500 shares available for trading on the New York Stock Exchange (More accurately, it is the total number that we actually have data on). If we look at the green line, we can see that the number of shares actually increased significantly from 2013 to 2014, before a significant decrease in 2015.

This sudden increase in 2014 would appear to go against our hypothesis that companies are steadily buying back their shares, as such a hypothesis would imply that the available shares should steadily decrease over time. However, this sudden increase is due to stock splits (where a company splits a share into a smaller number so that the price of an individual share becomes more affordable). Looking at the data, I noticed that the number of Apple shares went from approximately 925 million in 2013 to 6,088 million in 2014. The reason for this was that Apple instigated a 7-to-1 stock split in June 2014. If we exclude Apple stock in our counting of available shares, we find that there is no significant increase in 2014 (See blue line in figure above). However, the presence of stock splits reveals a flaw on our methodology. That is, stock splits clearly have a distorting effect on our analysis.

Rather than looking at estimates of the total aggregate number of available shares, let us instead look at the number of available shares for the ten largest companies by market cap.



Note that I have left out Apple in the list above as I did not want its stock split in 2014 to dominate and distort the demonstration. So we actually have only 9 stocks. In the upper panel, we can see how the number of available shares varies from one year to the next. We can see that with the exception of Facebook (FB), there is no appreciable increase in available stocks from one year to the next. That is, the overall trend in available stocks is downwards over time.

To illustrate this more clearly, I have a second subplot which shows the percentage increase in the number of available shares from one year to the next. You can clearly see that for the vast majority (19 cases out of 24), the percentage change is negative. The evidence, for these large companies at least, is that the number of available shares is decreasing over time, evidence of their ongoing share buyback programs.

Conclusion and Summary

Via the processing of stock and company data in Python and the subsequent generation of relevant plots, I have demonstrated some interesting phenomenon regarding companies in the S&P 500.

These include:

1. The market cap of S&P 500 companies appears to have a power law distribution. That is, it has a Pareto distribution that is ubiquitous in many natural phenomena.
2. Over the years 2013-2015, there has been a marked increase in the total debt being taken on by companies, the financial sector being by far the most indebted.
3. There is some evidence of an ongoing program of share buybacks as evidenced by the steady decrease in the number of available shares with each passing year.

The most obvious issue with this project is that I only really have adequate data for the years 2013-2015. A more comprehensive study would need to be performed over a longer period of time before convincingly proving the existence of the above trends.

Regardless, the evidence presented here shows that further study is warranted.

As a final note, I also produced a Tableau dashboard. This is a simple tool that allows a user to retrieve the data on any individual stock easily and quickly. All the analysis and figures however were produced in a Python Jupyter notebook.

References

[1] https://www.kaggle.com/dgawlik/nyse