UC Berkeley December 11, 2018
DATASCI W200 Fall ‘18
Final Project Report
Written by:
Patrick Castles (W 6:30-8pm)
Naveen Purushotham (Th 6:30-8pm)
Sonal Thakkar (Th 6:30-8pm)
Project GitHub: https://github.com/UCB-INFO-PYTHON/w200-project2-patrick-sonal-naveen

Technical Analysis of Cryptocurrency Market Behavior

I. Research Topic and Objectives

“What is bitcoin?” This was the most searched definition in the US and the UK in 2018, according to Google Trends. Bitcoin is a digital cryptocurrency and payment system that is entirely decentralized, meaning it is based on peer- to-peer transactions with no bureaucratic oversight. Transactions and liquidity within the network are instead based on cryptography. The system first emerged formally in 2008 and is currently a thriving open-source community and payment network. Based on the uniqueness of Bitcoin’s payment protocol and its growing adoption, the Bitcoin ecosystem is gaining lots of attention from businesses, consumers, and investors alike.

Unlike stocks, bonds and commodities, cryptocurrencies are a nascent financial asset with only a few years of historical data. Moreover, within the limited historical period that cryptocurrencies have been trading, the asset class has shown rampant volatility with respect to price. Substantive and sudden swings to the upside and downside has led to a clear lack of consensus among market participants as to what is the true intrinsic value of cryptocurrencies as a financial asset. In 2017 the market value of bitcoin surged 1,400 percent and reached a peak of $20,000 in December 2017. However, Bitcoin’s 2017 surge was followed by a precipitous decline in 2018 as the cryptocurrency is down 80% since the December 2017 highs and as a whole the cryptocurrencies have lost $730 billion in market value from the December peak of last year.

For our final project, we completed a technical analysis of cryptocurrency historical market data / trading behavior and compared it to the S&P 500 index as a benchmark. Setting out with high hopes, our objective was to answer the following research questions: (1) Do cryptocurrencies trade (i.e., price, volume, etc.) similarly to one another (e.g., Bitcoin vs. Etherium)? Do cryptocurrencies trade similarly to publicly traded equities listed in the US? Finally, as a follow-on to the previous two questions, can we make a definitive argument as to whether or not crypto currencies should be a part of a diversified investment portfolio?

II. Analytical Framework

Our analytical framework leverages two financial theories / methods:

(1) Technical Analysis - in finance, technical analysis is a methodology for forecasting the direction of asset prices through the study of past market data, primarily price and volume. Unlike fundamental analysis, technical analysis does not provide a thesis as to the underlying idiosyncratic drivers of a financial assets’ value (e.g., growing product demand, experienced management team, etc.). Technical analysis is focused solely on finding trends and patterns in the market data that can be used as signals to predict future market movements.

(2) Modern Portfolio Theory (MPT) - the quantitative benefits and mechanics of constructing a diversified investment portfolio is known in finance as Modern Portfolio Theory (MPT). To summarize, MPT is a strategy used by investors to construct a portfolio of investment assets that will maximize returns for a given level of risk. The optimal combination of assets depends on each investor's risk tolerance, however, essentially MPT suggests that investors can reduce risk through diversification of assets. The risk in a portfolio of diverse assets will be less than the risk inherent in holding any one of the individual assets, provided the risks of the various assets are not directly related (i.e. non-correlated).

During our research project, we use technical analysis and MPT accepted analytical frameworks that guide our analysis and supports our claims. Specifically, we accept MPT’s description of diversification (non-correlated returns) as a necessary condition for an asset to be beneficial when considering adding it to an investment portfolio. It is through this lense solely that we make judgments regarding the merits of cryptocurrency as an asset class and its place in an investment portfolio. It is worth noting that there are other asset classes (e,.g., bonds, commodities, real estate, etc.,) that are typically included in a well diversified portfolio (according to MPT), however, we did not include these assets in our analysis. Our time series analysis consists solely of market data for cryptocurrencies and the large cap, publicly traded US equities (the S&P500 index.) Despite this limitation, cryptocurrencies have many similarities to the large cap, publicly traded US equities. Specifically, crypto assets and the S&P500 are both traded publicly on market exchanges, both share similar quantitative metrics like volume and price, and both are recognized as high risk / high-return assets. Therefore, the S&P 500 index seemed like the most appropriate benchmark for comparison to crypto.

III. Arguments

First, we acknowledge in advance that any definitive argument based on descriptive statistics will be limited due to the small sample size of cryptocurrency market data. Despite this limitation, we’re confident that from our analysis we feel that we can make the following arguments as answers to our research questions :

(1) Do cryptocurrencies trade (i.e., price, volume, etc.) similarly to one another (e.g., Bitcoin vs. Etherium)? In short, the answer is no. Our observation is that the individual cryptocurrencies behave randomly with regards to price and volume. However, the recent trend is towards a convergence in price behavior as shown in the heat map visualizations that show a clear increase in correlations from 2017 to 2018. We believe that the randomness in price behavior is supported by the notion that each currency is in fact unique with its own value proposition and utility function. Therefore, according to MPT the heterogeneity in market behavior of crypto currencies supports a strategy of holding several cryptocurrencies in one's investment portfolio as opposed to a concentrated position of one cryptocurrency. This aggregation strategy would enable an investor to maximize their risk adjusted return.

(2) Do cryptocurrencies trade similarly to publicly traded equities listed in the US? In short, the answer is again no, our observation is that there is no meaningful relationship/correlation. Similar to intercurrency behavior, there is a more recent trend towards convergence of the S&P500 market behavior and crypto. We believe this may be due to the “risk-off” market sentiment that is affecting the broader market that makes up all financial assets. As investors move money into safer assets like bonds and cash it makes sense that S&P500 and cryptocurrencies would both be negatively affected.

(3) Finally, can we make a definitive argument as to whether or not crypto currencies should be a part of a diversified investment portfolio? In short, our answer is….it’s too soon to tell. On the one hand, our observation is that the absence of any correlation in market behavior supports the view that crypto should be included as part of a well diversified portfolio according to MPT. However, there is simply not enough data to make a definitive argument in this regard.

IV. Data Procurement and Cleaning Process:

The Raw Data Files

For our project we procured cryptocurrency market data from Kaggle. The raw data was in .csv format and consisted of 17 datasets each containing trading data for unique cryptocurrency. For our comparative analysis with the S&P 500 index we ended up procuring 2 datasets from Yahoo, the first S&P 500 dataset was sufficient for us to do most of our analysis as it consisted of the price data we needed for the comparative analysis portion. The second S&P 500 dataset was procured three quarters of the way into the project as to glean additional comparative insights for other new variables of the S&P 500 index. Below we show snapshots of the dataset and all the variables.

All the 17 cryptocurrency datasets had the following columns:

Date	Open	High	Low	Close	Volume	Market Cap
Feb 20, 2018	11231.80	11958.50	11231.80	11403.70	9,926,540,000	189,536,000,000
Feb 19, 2018	10552.60	11273.80	10513.20	11225.30	7,652,090,000	178,055,000,000

Column definitions:
Open - is the price at which a stock first trades upon the opening of an exchange on a trading day
High - the highest price in the trading day in USD.
Low - the highest price in the trading day in USD.
Close - is the price at which a stock last trades before the closing of an exchange on a trading day
Volume - Is the amount of shares bought/sold of a stock the trading day. It is important because the more volume the more people agree with the price of the stock.
Market Cap - Market capitalization is the market value of a stocks outstanding shares

The first S&P 500 dataset had only one column:

date	value
2008-11-28	896.2400
2008-12-01	816.2100

Column definitions:
Value (Close) - is the price at which a stock last trades before the closing of an exchange on a trading day

The second S&P 500 dataset had the following columns:

Date	Open	High	Low	Close	Adj Close	Volume
2014-04-01	1873.959961	1885.839966	1873.959961	1885.520020	1885.520020	3336190000
2014-04-02	1886.609985	1893.170044	1883.790039	1890.900024	1890.900024	3131660000

Column definitions:
Open, High, Low, Close, Volume - are same as the cryptocurrency dataset

Adj Close - An adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time before the next day's open

Dataset indexes and shapes:
All the 17 cryptocurrencies had different start sample date except bitcoin and litecoin which had trading data starting 2013-04-28.
All the 17 cryptocurrencies had the same end sample date of 2018-02-20.
For S&P 500 the date ranges are listed below 2008-11-28 - 2018-11-27.

The details of the date ranges are listed in the table below

Dataset (.csv)	**Dataset Shape (records, columns)**	**Data Range (Dataset index)**
sp500-10-year-daily-chart	(2519, 2)	2008-11-28 - 2018-11-27
GSPC	(988, 7)	2014-04-01- 2018-03-01
bitcoin_cash_price	(213, 7)	2017-07-23 - 2018-02-20
bitcoin_price	(1760, 7)	2013-04-28 - 2018-02-20
bitconnect_price	(397, 7)	2017-01-20 - 2018-02-20
dash_price	(1468, 7)	2014-02-14 - 2018-02-20
ethereum_classic_price	(577, 7)	2016-07-24 - 2018-02-20
ethereum_price	(929, 7)	2015-08-07 - 2018-02-20
iota_price	(253, 7)	2017-06-13 - 2018-02-20
litecoin_price	(1760, 7)	2013-04-28 - 2018-02-20
monero_price	(1371, 7)	2014-05-21 - 2018-02-20
nem_price	(1057, 7)	2015-04-01 - 2018-02-20
neo_price	(530, 7)	2016-09-09 - 2018-02-20
numeraire_price	(243, 7)	2017-06-23 - 2018-02-20
omisego_price	(222, 7)	2017-07-14 - 2018-02-20
qtum_price	(273, 7)	2017-05-24 - 2018-02-20
ripple_price	(1662, 7)	2013-08-04 - 2018-02-20
stratis_price	(558, 7)	2016-08-12 - 2018-02-20
waves_price	(629, 7)	2016-06-02 - 2018-02-20

Data Cleansing

Fortunately, our dataset was not as messy as some other real-world datasets, it was fairly regular with some minor inconsistencies. The first thing to note is that the cryptocurrency is a 24x7 market while the traditional financial markets are closed over the weekends. This is reflected in the data collected.

Dataset (.csv)	Cleansing Process
sp500-10-year-daily-chart	- The data on this file had a column named:' value'. This was renamed to: 'close' to match the cryptocurrency dataset. The index was reset to the 'date' column.
Bitcoin_cash_price bitcoin_price Bitconnect_price Dash_price Ethereum_classic_price ethereum_price iota_price litecoin_price Monero_price Nem_price Neo_price numeraire_price omisego_price qtum_price ripple_price Stratis_price waves_price	- The data on the crypto-currency files had uppercase column names, these were changed to lowercase. - We also removed the space between column names for e.g. 'market cap' was changed to 'marketcap' so that we could access them as attributes easily. - The index was reset to the 'date' column. - We also noticed that the 'volume' and the 'marketcap' columns were set as strings, for easy analysis we changed the dtype for these columns to float64. - One problem we faced was that volume and marketcap string columns had odd symbols in the data for e.g. '-' and '',' we had to first remove this inconsistency before converting the data to float64 using the to_numeric function
GSPC	- The data column names were made lowercase and the index was reset to the date column.

V. Python planning and execution:

For our project we used the PyData ecosystem to carry out the research analysis, below are a few points to explain our thought process and the course of action we took.

1. Although we expected to use pandas, numpy, and matplotlib libraries at first, we soon realized that we probably won't be needing numpy for this particular project. We relied heavily on the **pandas** library for data-exploration and data-analysis. For plotting our analysis we used **matplotlib** for the most part but for a few nicer visualizations we used **seaborn,** specifically for heatmaps, and plotly for a time series animation. Apart from these we used the **os** and the **glob** libraries to work with our raw data files.

   Note: optional installation for plotly animation
   We used the plotly library to plot one time series plot for the the bitcoin price action

Exploration. We have attached a video for it in this report. Optionally, to run it in the notebook these are the two steps that are needed:

a) Create a plotly free account - 25 public charts limit.
b) Terminal runs for installations:
   pip install plotly
   conda install -c plotly plotly-orca psutil

Alternatively, if the video suffices, to run the notebook without installations, skip the cell that has import plotly.

2. Since we had 17 files for cryptocurrencies with the same column data we choose to concatenate the data vertically into one pandas DataFrame using ‘pd.concat‘ as opposed to the merging them horizontally . Our plan was to use the groupby ‘name’ function to analyze individual currencies, this way we were able to manage the columns easily.

3. We used the’ pd.merge’ function while doing comparative analysis between S&P 500 and Cryptocurrency data. Using different variations of inner and outer join we were able to accomplish analysis and visualization for our project.

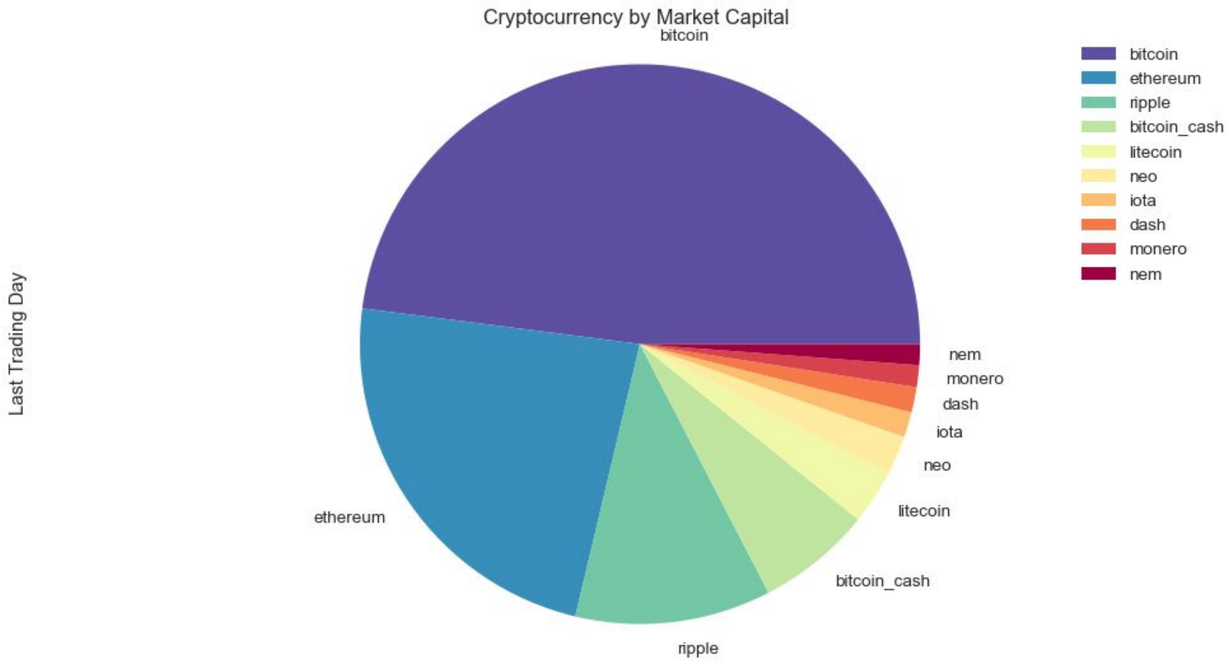
4. For statistical analysis we relied on the pandas methods. We used the pandas in-built plotting function extensively and only when limited by visualization capabilities did we have to use seaborn and plotly libraries. For our functional programming we used the following pandas methods extensively for our analysis:
   - Data arrangement - df.read_csv, df.group_by(), get_group(), df.sort_values(), to_frame(), df.drop(), df.sort_index(), df.reset_index(), df.merge()
   - Statistical - pct_change(), corr(), mean(), std()
   - Plotting - df.plot(), df.hist()

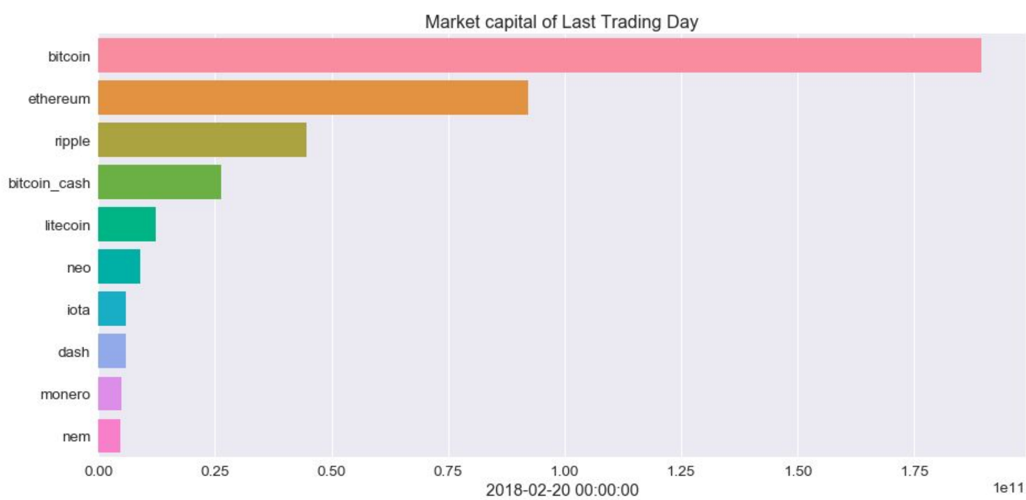
   Just to name a few.

5. Plotting and Visualization was a major portion of our exploration and analysis since we worked with time series data, our initial exploration was line plots to see the market behavior history, we used pie charts and bar charts to explore the bigger players of the market to choose potential candidates for further analysis. We had to use log scales on multiple occasions to observe the relative trends of the markets as some bigger currencies skewed the absolute plots. To measure market volatility we used scatter plots and histograms and finally for measure statistical correlations between cryptocurrencies we used heatmaps.

VI. Exploration, Analysis & Insights:

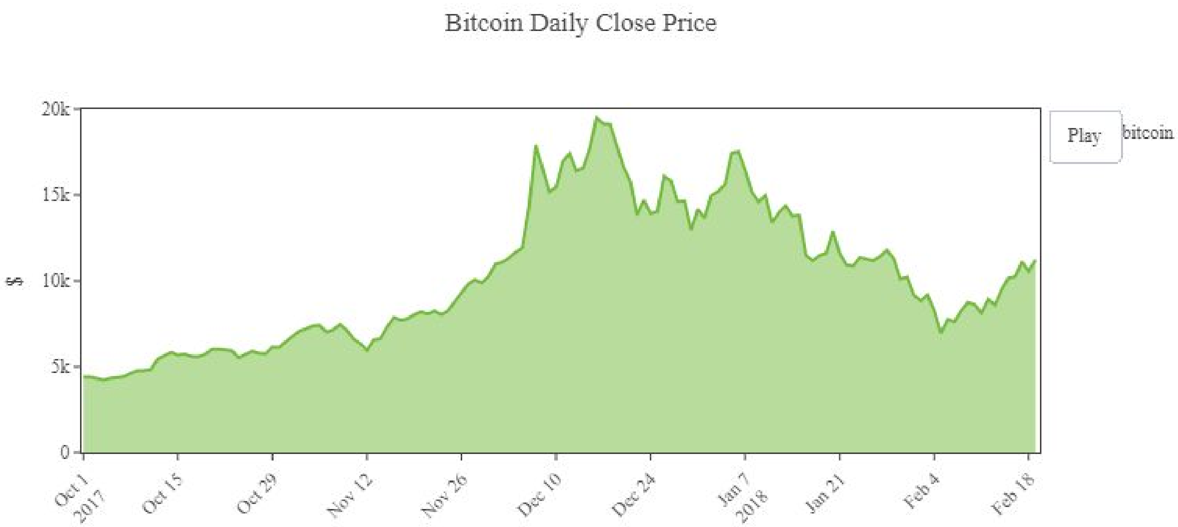
Plot Cryptocurrency by Market Capital:





<u>Commentary and Insight:</u> this visualization was a part of our exploratory process. We wanted to see which crypto currencies had the largest market cap value and their relative positions. We used a pie chart and horizontal bar chart. One can quickly see that Bitcoin owns the major share of the market followed by ethereum and ripple.

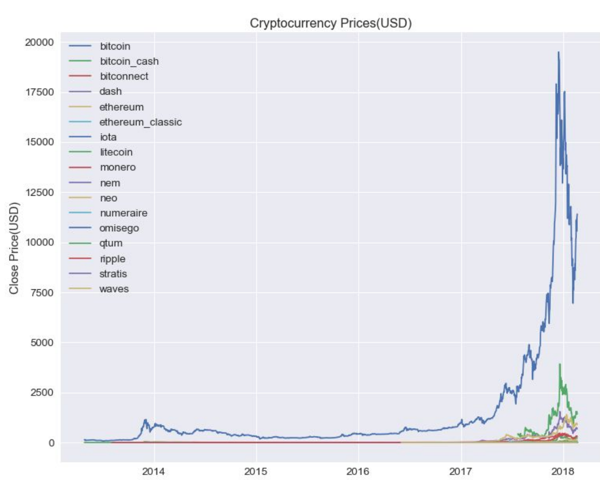
Bitcoin close price trend from 2017-10-01 to 2018-02-20:

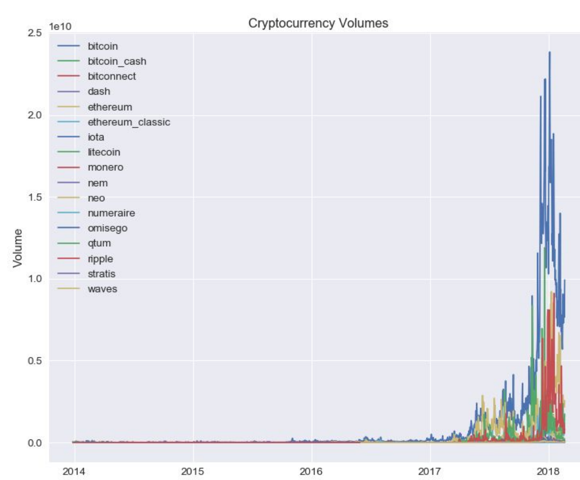


Animated play url:
https://zoom.us/recording/share/hfoTKprS6q1saW2XPxyHUbVn1SuQJWGikmYBvAazcoqwIumekTziMw?startTime=1544662490000

Commentary: This plot was added as part of our exploratory process of not just the data but also exploration of nicer visualization toolsets. We used plotly to create this animation of bitcoin price behavior, however, since this plot used paid libraries and we were only able to plot a very small sample size we did not use the library further.

Plot closing prices and volumes for all currencies over sample size - Trend Comparison



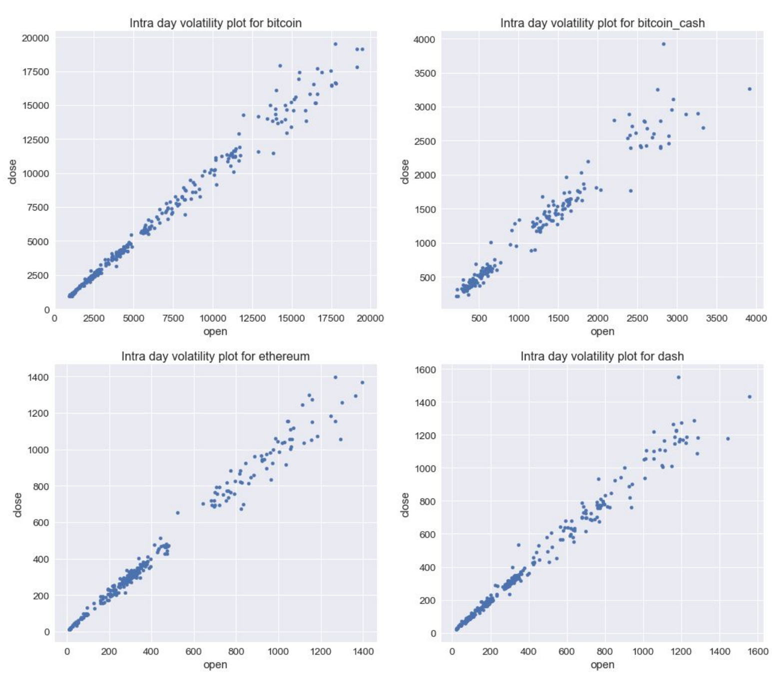


Commentary and Insight: this visualization was a part of our exploratory process. We wanted to see which crypto currencies had the biggest price and volume movements over the course of the time series (relative to their sample sizes). It is clear that Bitcoin has the largest swing to the upside.

Close Price Log Chart for 4 major cryptocurrencies 'bitcoin', 'bitcoin_cash', 'ethereum', 'dash':

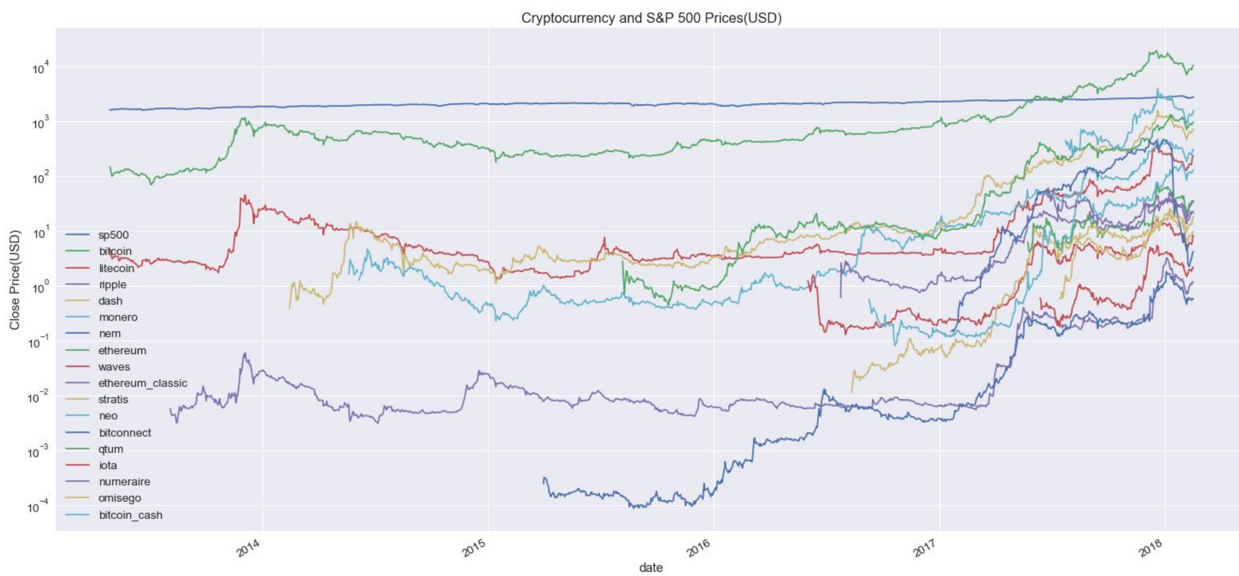
Commentary and Insight: this visualization was used to better understand the relative movement in close price among the four biggest cryptocurrencies. By removing the other cryptocurrencies and using log form we are able to see the upward trend and the relationship between these currencies.

Intraday Volatility Scatter Plot for 4 major cryptocurrencies 'bitcoin', 'bitcoin_cash', 'ethereum', 'dash':

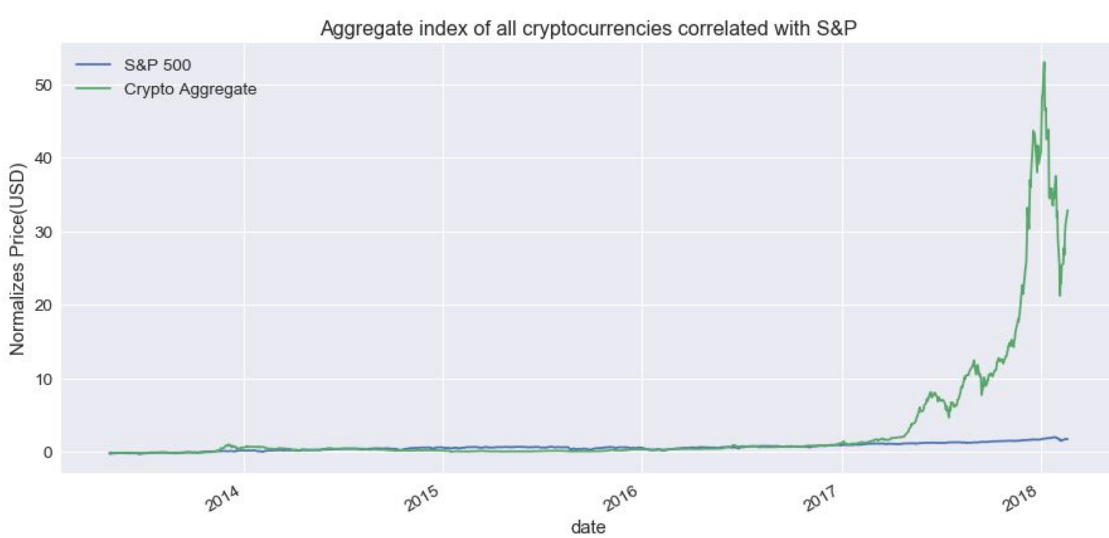


Commentary and Insight: the scatter plot was used to help us understand the volatility of the currencies. That is how much dispersion there is between the open and close price of the cryptocurrencies. From this trend we were able to see that when cryptocurrencies open at a higher than normal price there tends to be more volatility on average, this creates good opportunities for traders to take advantage of very short term trading.

Line Chart Showing Log Close Price of Cryptocurrencies and S&P 500

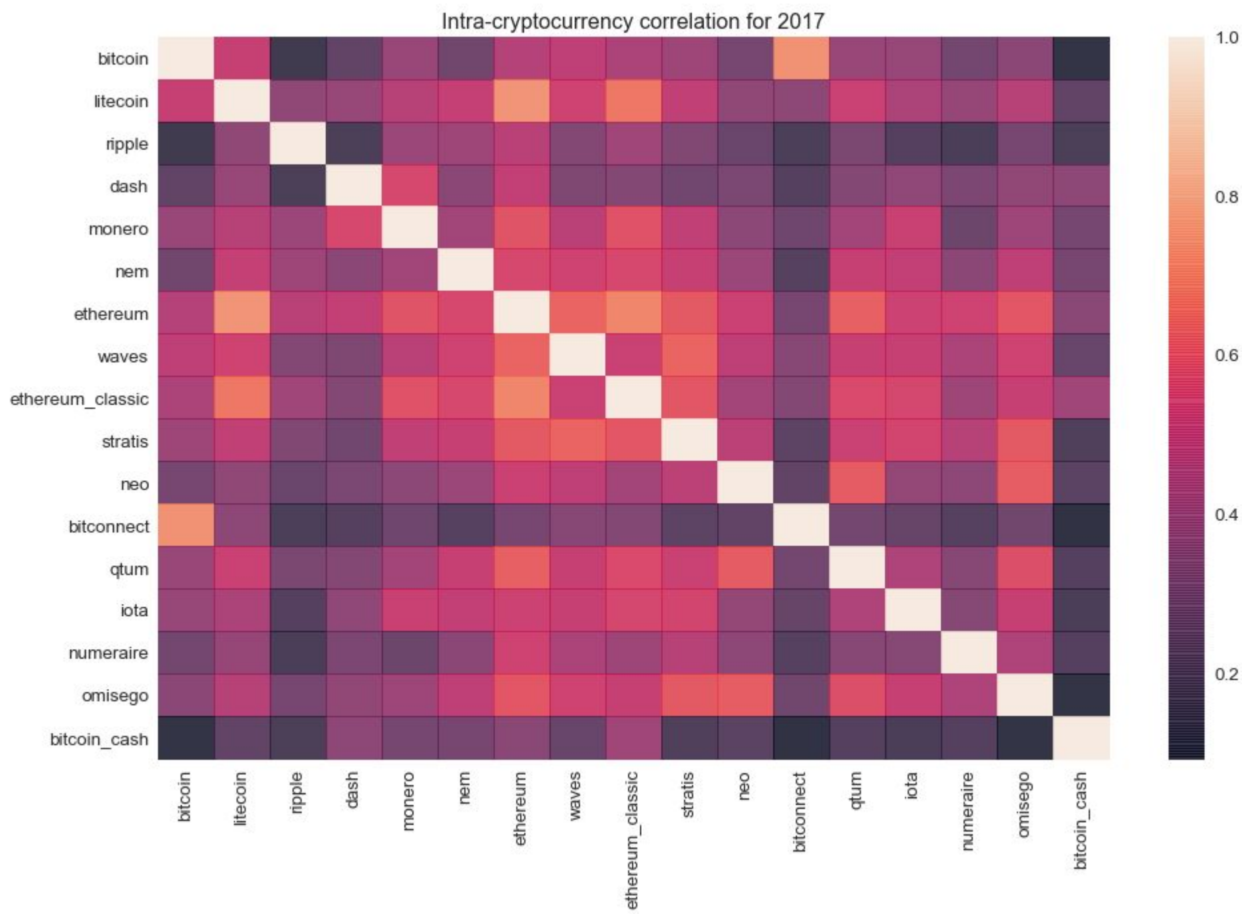


Aggregate index of all crypto currencies correlated with s&p:
Data is normalized

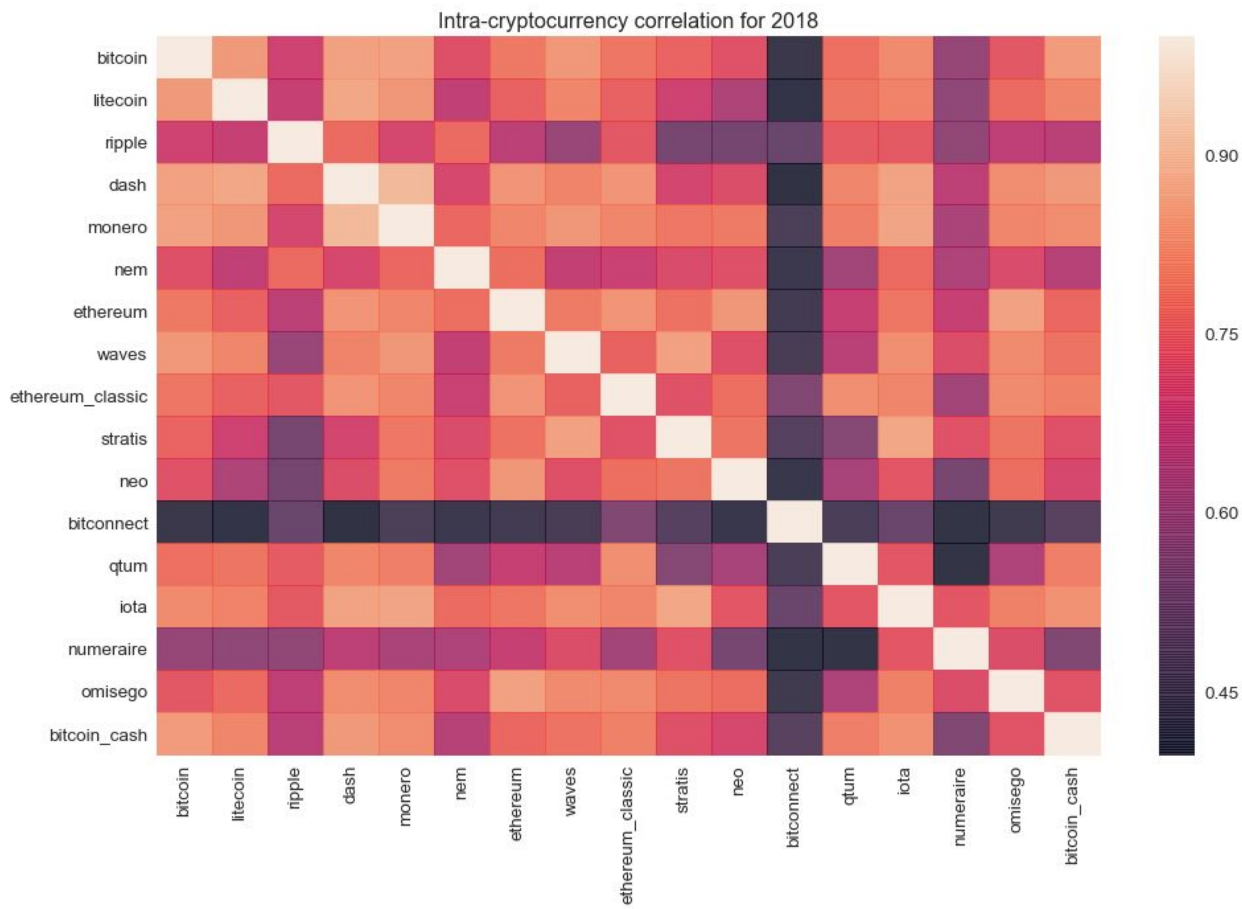


Commentary and Insight: the line charts above were used to show the S&P 500's close price over time relative to the cryptocurrencies. We used log scale and normalized the data to better understand the relative comparison in movement and behavior. The main takeaway from this visualization is that the cryptocurrencies are much more volatile and had an exponentially higher move to the upside as compared to the S&P 500 starting in 2017.

Crypto-currency correlationship for 2017:



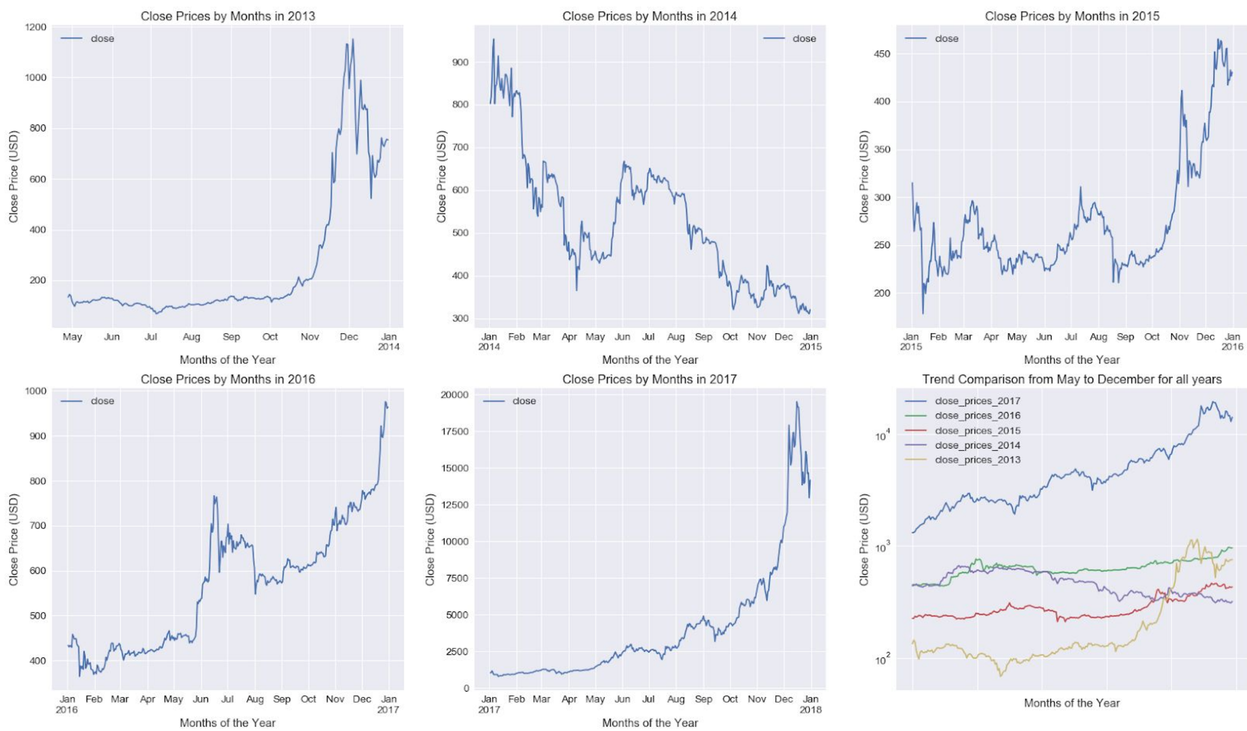
Crypto-currency correlationship for 2018:



Commentary and Insight: the two heat map matrices above were used to show the correlation among the cryptocurrencies of their daily percentage change in close price. We used the heatmaps because we wanted a visualization that would easily show the general trend in correlation among the cryptocurrencies from 2017 to 2018. We used the pearson correlation coefficient for this. The color difference in the heat maps more purple (less correlation) in 2017 and more orange (more correlation) in 2018 clearly shows that the cryptocurrencies correlations grew stronger in 2018 as compared to 2017. The sample size is not large enough to make a claim about the correlation that is statistically significant.

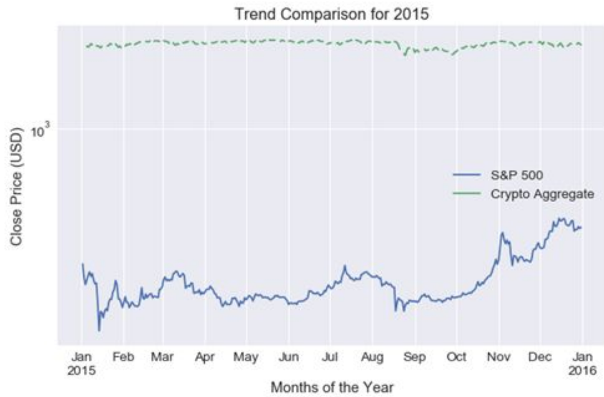
Monthly sentimental analysis of Bitcoin:

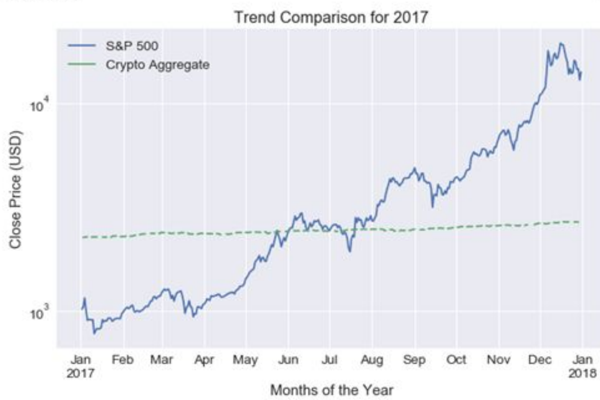
Why did Bitcoin crash in 2014? - https://www.businessinsider.com/bitcoin-price-drop-2015-1



Commentary and Insight: We completed these line charts of the close price of bitcoin each year to see if there were any commonalities with regards to monthly movement. We found that December seems to mark a peak each year followed by a decline in January, this is tradeable insight that could be used to make money. Also, plotting the yearly moves on top of one another were able to see which years had the biggest relative move to the upside (2017) and downside (2014).

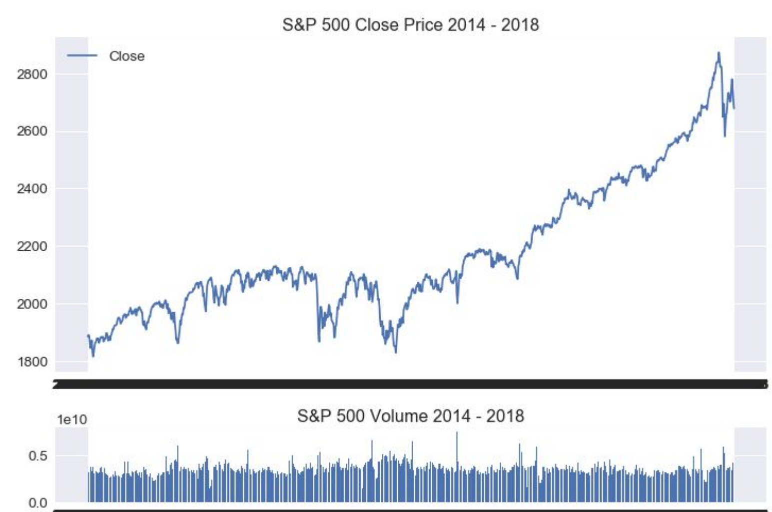
S&P 500 trend by the month compared to Crypto Aggregate index:





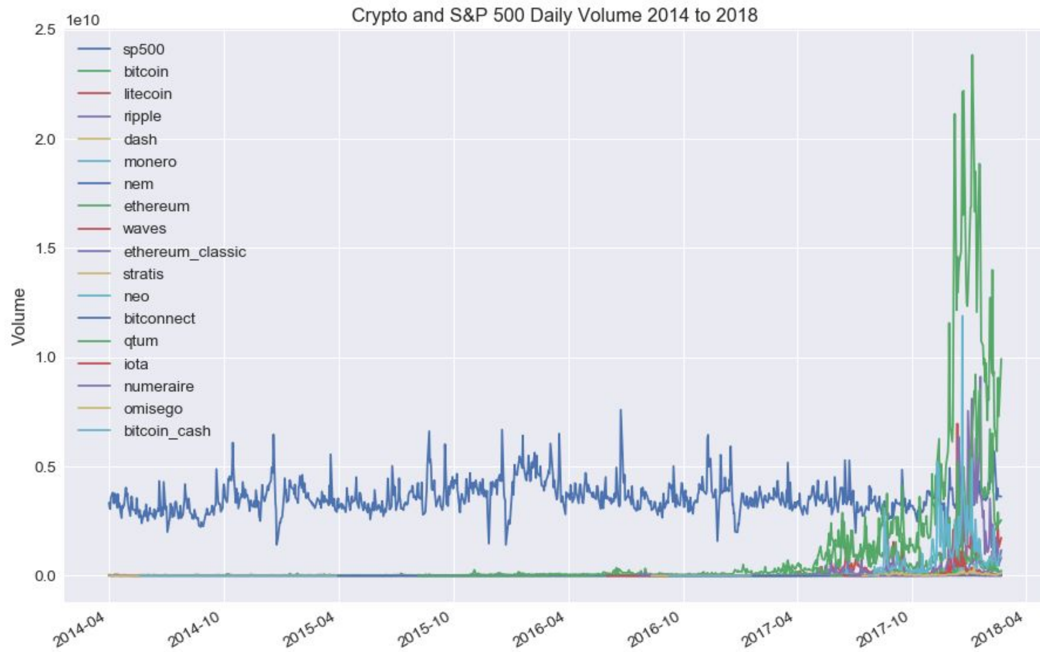
<u>Commentary and Insight:</u> Next we used the month to month movement of the crypto aggregate and compared it to the S&P 500. It’s clear that crypto had a much higher relative climb in value as compared to the S&P 500. The most notable difference in price appreciation was in 2017.

Combined price and Volume charts for S&P and Bitcoin:



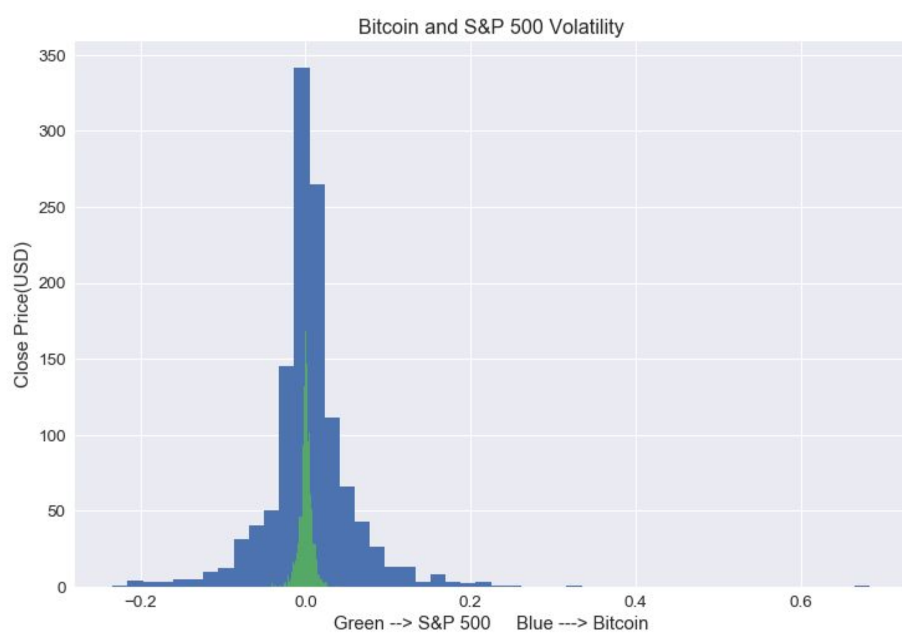
Commentary and Insight: We created a plot of the S&P 500 close price and a subplot of the volume. We then did the same visual for bitcoin. It’s clear the S&P 500 has much less volatility in volume but yet still a solid uptrend in close price. By contrast bitcoins steep rise was coincident with a large uptick in volume.

Crypto and S&P 500 Daily Volume 2014 to 2018:



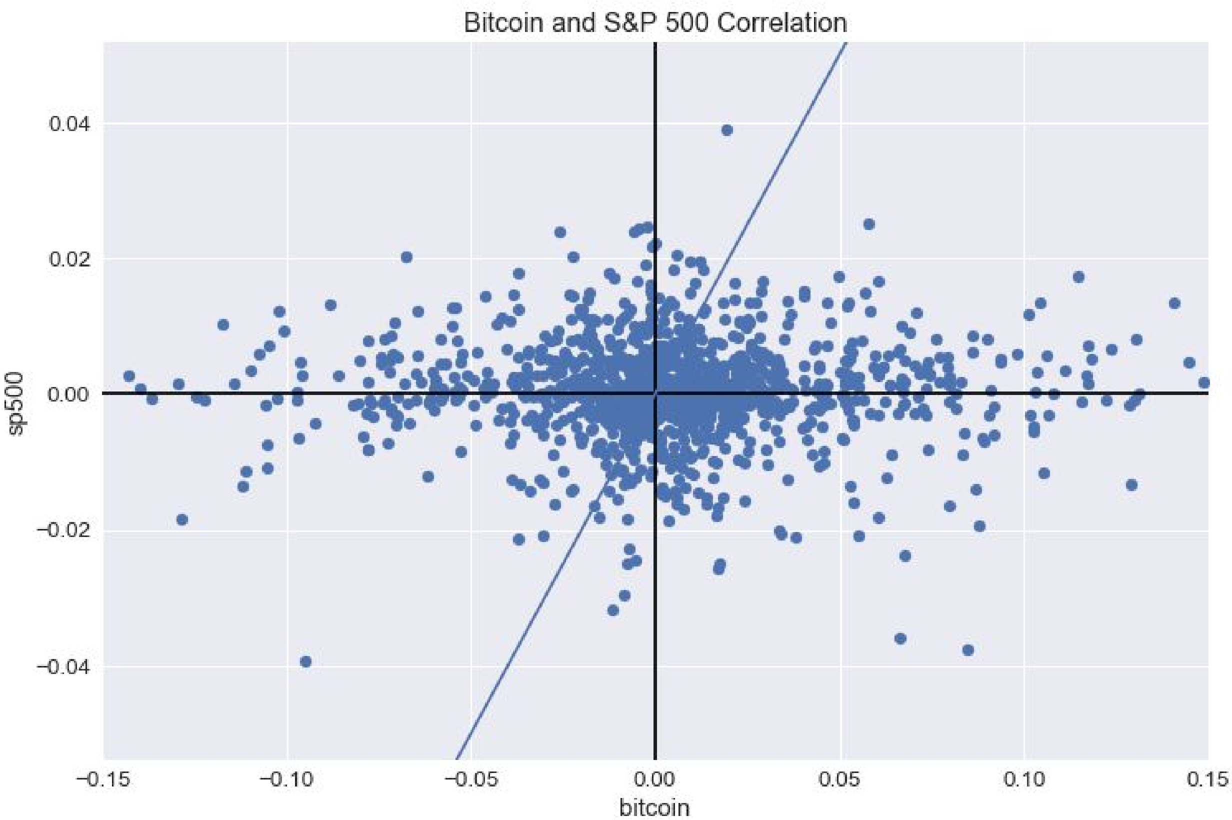
Commentary and Insight: We created a line chart of all of the cryptocurrencies and the S&P 500 to see the variation in volume over time. The increase in volume of crypto in 2017 is remarkable. The surge in volume of bitcoin and most of the other cryptocurrencies makes sense as the surge in volume was coincident with a surge in price.

Bitcoin and S&P 500 stability measurement using Histogram:



Commentary and Insight: We created a histogram to show the relative amount of volatility in the closing price of Bitcoin and the S&P 500. It's clear that bitcoin has exponentially higher volatility than the S&P 500.

Scatter plot of bitcoin vs S&P 500:



Commentary and Insight: Lastly, we created a scatterplot to observe the correlation of the percentage change in closing price between Bitcoin and the S&P 500. Without making a definitive claim, our observation is that there isn't a strong correlation between bitcoin price movement and the S&P 500 (based on the data we analyzed.)

VII. Conclusion

To summarize, our research objective was to complete a technical analysis of crypto market trading behavior in order to gain insight as to how cryptocurrencies trade relative to themselves and other financial assets, specifically large cap, US equities. Leveraging our newly acquired python skills, we compiled, cleaned and analyzed time series data from a list of 19 csv files. The two main variables that we analyzed were price and volume and we were interested in the relationships between these variables and how they changed over time. We learned a great deal during this project and had to overcome several challenges. Specifically, we learned how to deal with missing data, how to aggregate data into useful data frames, how to use visuals to explore the data initially and finally how to use visuals to convey findings to an audience. If we had more time we would love to take a deeper dive into our technical analysis of cryptocurrencies and possibly add in other variables and methodologies such as sentiment analysis. As a group, we learned how to work on a coding project as a team using github for version control which was a pain early on but very helpful once we got the hang of it. Overall, this project was a great learning experience and we look forward to receiving feedback from the teaching team. Thanks to the whole teaching team for such a great course, we learned a lot!