**Do more than six stocks matter for the NASDAQ?**

**Project Introduction:**

Thousands of American companies makeup American stock exchanges, such as the New York Stock Exchange and NASDAQ. However a recent Wall Street journal article, using data compiled by brokerage firm JonesTrading, stated that actually six companies accounted for more than half of the value added to the NASDAQ in 2015: Amazon, Google, Apple, Facebook, and Netflix, and Gilead [1]. This alarming finding suggests the NASDAQ’s growth might largely reflect changes from a handful of technology/ pharmaceutical-related companies designated by the acronym ‘FAANG’ (Facebook, Amazon, Apple, Netflix, Gilead, and Google). If such a strong relationship were to exist between certain stocks and the NASDAQ, then this could indicate that the NASDAQ is not a robust barometer of the overall health of the 3,000+ companies that makeup NASDAQ, but rather is a biased barometer for a small handful of stocks. In this report, the student investigated this possibility by relating how time-varying changes in the NASDAQ are predicted by changes in individual stock price.

**Analysis:**

**Part 1 – A stock is more likely correlated with the NASDAQ if it’s in the Top NASDAQ 100 stocks**

For the last five years (April 2011 - April 2016), the price of FAANG stocks have closely co-varied with the value of the NASDAQ (Figure 1). Throughout this period, most FAANG stocks increased monotonically and exhibited high correlations with the NASDAQ (ranging from 0.89 to 0.95). High correlation values are interesting because they indicate the possibility of a causal relationship between FAANG stocks and the NASDAQ. This relationship has been recently suggested in a Wall Street journal article [1]. If such a relationship were true, then the collective performance of FAANG stocks alone should be sufficient to predict how NASDAQ has changed over the last five years.

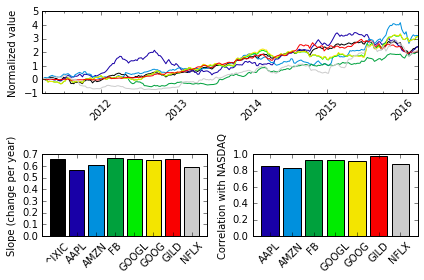


Figure 1: The relationship of NASDAQ to FAANG stocks between April 2011 – April 2016. (*top panel*) The time-varying change in normalized stock value of all the FAANG stocks relative to the normalized value of the NASDAQ (stock symbol is ^IXIC). Time-varying stock/index traces were normalized first by z-score, and then zero-shifted such that that the first datapoint for each stock/index was shifted to zero. The color for each trace matches the associated stock in the bar plots on the bottom row. (*bottom left panel*) The slope of changes for the NASDAQ and ‘FAANG’ stocks. (*bottom right panel*) The correlation between the NASDAQ trace and the individual FAANG stocks.

However before testing for a causative relationship between FAANG stocks and the NASDAQ, one should establish whether the FAANG stocks behave uniquely relative to other stocks that make up the NASDAQ. Although correlations between FAANG stocks and the NASDAQ are consistently high (r=0.90±0.02 [SEM]), many other Top 100 NASDAQ stocks also exhibit similar relationships with the NASDAQ (Figure 2, r=0.80±0.03 [SEM]). In fact, FAANG stocks are not significantly different from other Top 100 NASDAQ stocks (Rank-Sum Test: p=0.57). Although neither group is significantly different from the other, both groups are significantly different from all other NASDAQ stocks (Rank-Sum Test: p<0.01), which as a group displayed lower correlations with the NASDAQ (r=0.28±0.01 [SEM]). Despite these other NASDAQ stocks being systematically lower as a group, even ~20% of these stocks showed positive correlations with the NASDAQ that were comparable to FAANG stocks. These findings unequivocally demonstrate that FAANG stocks are not unique in being highly correlated with the NASDAQ, but rather are part of a relatively larger subset of NASDAQ stocks (~20%) that closely co-vary with the NASDAQ.

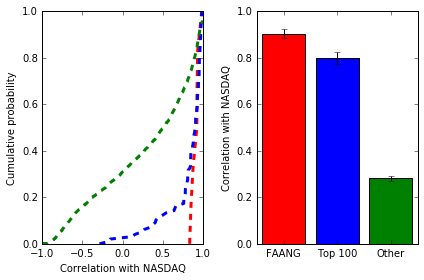


Figure 2: Correlation of different groups of NASDAQ stocks to the NASDAQ. (*left panel*) Distribution of correlations based on group type. (*right panel*) Summary correlations based on group type (Mean ± SEM). The different stock groups are: FAANG stocks (red, n=8), other Top 100 NASDAQ stocks (blue, n=99), and other NASDAQ stocks (green, n=2907).

By definition the Top 100 NASDAQ stocks comprise companies with the highest market cap value in the NASDAQ. In fact, the Top 100 NASDAQ stocks makeup ~60% of the NASDAQ’s value, and FAANG stocks alone makeup 25% of the NASDAQ’s value. Thus, one uninteresting possibility is that correlations between NASDAQ and individual stocks might trivially arise based on a stock’s market cap value. This is because NASDAQ is a capitalization-weighted index, where each company is weighted by its market value. However, the market cap value of a company is actually only marginally correlated with its correlation to the NASDAQ (Figure 3, r=0.35). This means that while some correlation to the NASDAQ might merely arise from market cap value, the relationship of individual stocks to the NASDAQ is more complicated and nuanced.

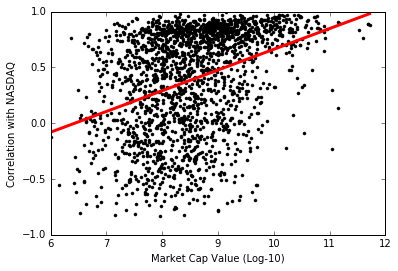
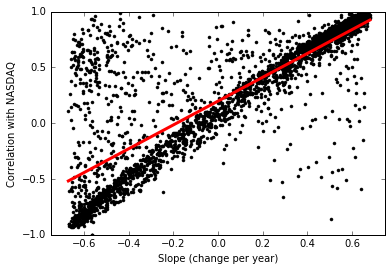
 

Figure 3: Relationship of each stock’s correlation with the NASDAQ relative either to its market cap value or growth. (*left panel*) A weak correlation exists between the market cap value of a company (expressed logarithmically in dollars) and the correlation between a company’s stock and the NASDAQ (r = 0.35). (*right panel*) A strong correlation exists between the growth of a company (expressed in terms of the slope for a linear fit to the company’s time-varying stock trace) and the correlation between a company’s stock and the NASDAQ (r = 0.83).

One practical question that should be asked is whether higher correlations with the NASDAQ matter, specifically in regards to whether they are positive indicators for growth. And indeed over the last five years higher correlations of a stock to the NASDAQ were positively related to the growth of the stock (quantified in terms of the slope for a linear fit to each time-varying stock, see Figure 3 - right panel ). This suggests then that the stocks in the NASDAQ that are most correlated with the NASDAQ exhibited the largest growth between 2011-2016 (which includes the Top 100 NASDAQ stocks, see Figure 4). This is somewhat worrisome because if the NASDAQ’s value reflects only a small growing-segment of stocks, then market growth seems to be mostly concentrated and not broadly distributed across a large array of companies.

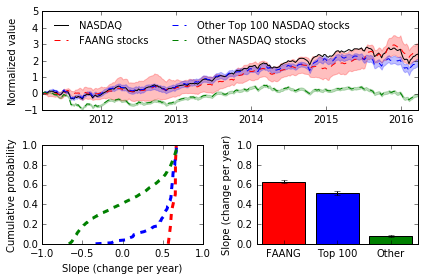


Figure 4: FAANG and other Top 100 NASDAQ stocks show more consistently positive growth than other NASDAQ stocks. (*top panel*) The time-varying change in normalized stock value of the NASDAQ (black), FAANG stocks (red), other Top 100 NASDAQ stocks (blue), and other NASDAQ stocks (excluding Top 100, black). For the three stock groups, the mean normalized stock price is shown as dashed lines, and the 95% confidence intervals are the shaded areas. Time-varying stock/index traces were normalized first by z-score, and then zero-shifted such that that the first datapoint for each stock/index was shifted to zero. (*bottom left panel*) Distribution of slopes derived from linear fits to each time-varying stock trace (seperated by group type). (*bottom right panel*) Summary slopes based on group type (Mean ± SEM).

Because growth is concentrated in a small subset of stocks, one might wonder if this growth is systematically reflected within certain sectors of the NASDAQ. Indeed both technology and consumer services appear to be overrepresented in both the FAANG and Top 100 NASDAQ stocks (Figure 5). However, neither of these sectors appears to be drivers of growth in the NASDAQ, in which the Finance sector seems to show the most consistent growth out of any sectors (Figure 6 and 7). Nonetheless there is not a straightforward breakdown for drivers of NASDAQ growth, at least based on sector. Because so many stocks makeup of the NASDAQ, dimensional reduction via k-means clustering can help reveal systematic trends across different stocks. Interestingly, when k-means is applied to all 3000+ companies of the NASDAQ (including the NASDAQ itself), the majority of Top 100 NASDAQ stocks, along with ~20% of NASDAQ stocks, fall within the cluster containing the NASDAQ (Figure 8). Furthermore, different time-varying behaviors can be noted. For example a few clusters exhibit systematic decreases during 2012 and 2014, whereas other clusters show delayed rises.

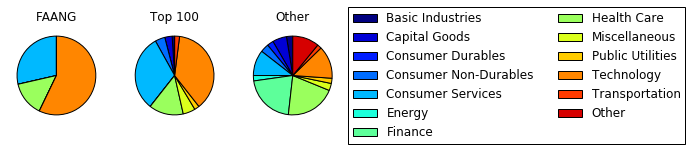


Figure 5: Sector breakdown based upon group type: FAANG stocks, other Top 100 NASDAQ stocks, or other NASDAQ stocks (excluding Top 100).

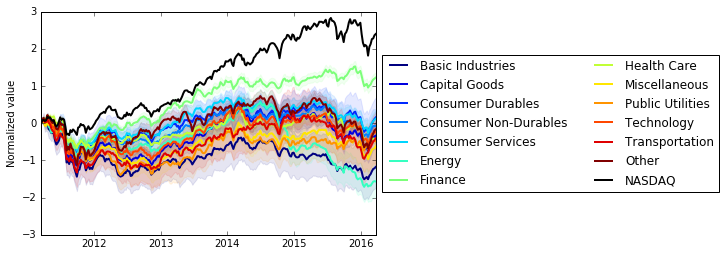


Figure 6: The time-varying change in normalized NASDAQ value (black) versus different NASDAQ sectors. For each sector, the mean normalized stock price is shown as solid lines, and the 95% confidence intervals are the shaded areas. Time-varying sector/index traces were normalized first by z-score, and then zero-shifted such that that the first datapoint for each sector/index was shifted to zero.

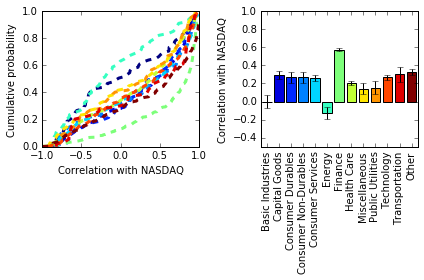


Figure 7: Only the finance sector appears to consistently be more correlated with the NASDAQ than other sectors. (*left panel*) Distribution of correlations with NASDAQ based on sector. (*right pane*l) Summary correlations based on sector (Mean ± SEM).

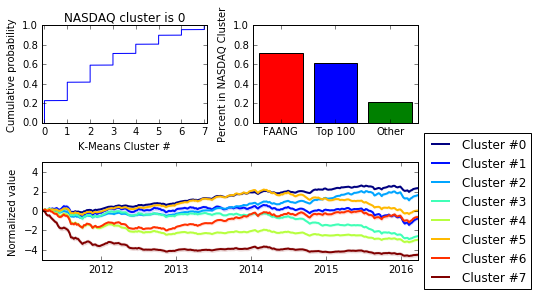


Figure 8: K-means clustering (8 clusters) of all the NASDAQ stock data reveals that different subsets of stocks display similar time-varying behavior. Most of the FAANG and Top 100 NASDAQ stocks are similar to the NASDAQ.

**Discussion**

A deep dive into this data reveals that the NASDAQ is primarily correlated with a subset of NASDAQ stocks. The next major step is to determine the nature of the relationship between these stocks and the NASDAQ, such as whether the relationship is causative or largely reflects other outside economic factors (i.e. other financial markets, bond rate, etc.). The next step will require the use of more complicated analytics, such as regression and machine learning algorithms. Unfortunately, one major limitation of the dataset is that historical data is not available for the market cap data of each stock. Thus, although this report has tried to address whether the relationship between market cap value and correlation with the NASDAQ (Figure 3), one limitation is that while the measure of correlation for each stock to the NASDAQ is between the period of 2011-2016, the market cap data only spans a single year (2016).

Another potential area for concern is that the stocks analyzed in this project are based on those that are currently a part of the NASDAQ as of 2016, and not just those present between 2011-2016. The ramification of this choice means that some stocks had their IPO sometime between 2011-2016 (like Facebook in 2012), and for these stocks data is not present through the entire period. Furthermore, companies that went out of business prior to 2016, but were present in the analysis period are missing from this analysis. Most of the analysis was carefully chosen only to consider correlations with the NASDAQ during periods in which data was unavailable. Unfortunately, for some analytics like the k-means clustering, this can be a problem, and the student would like to redo the analysis, and verify that excluding stocks with IPOs in the analysis period don’t skew the results.