Zihao Li, Jinghao Yan, Brody Vogel, and Ye Zhang

To download our stock data: Click Here

To download our sports data: Click Here

To download our Python scripts: Click Here

To view just our visualizations: Click Here

ANLY 501: Project 3

Stocks, Sports, and Money?

Introduction

Some estimates put the value of every share in the U.S. stock market at $15 trillion, with $4.3 trillion of that being traded every day [1][2]. The ability to predict any of the market's movement, then, is a very profitable skill. This paper doesn't do that; it does, however, explore the correlation between the market's fluctuations and the happenings in another popular realm--that of professional sports. Specifically, the data science question examined here is: ***Does the U.S. stock market react to professional sports leagues and the performance of their teams, and, if it does, how?*** Unbelievable amounts of money are spent on sports. As a small example, consider the price of tickets to game seven of the last two World Series: average cost per ticket in 2016 was $2,100 and $1,800 in 2017, with these single game tickets maxing out at over $200,000 for a seat behind home plate [3][4]. And that's just a small drop in the pool: Since 2014, the Los Angeles Dodgers - one of the teams from this past year's Series - have reported over $2 billion in revenue [5]. If we assume the average team earned only half that - $1 billion - that means the 92 teams in the three major American professional sports leagues have earned around $100 billion over the last four years. Is there a chance that all this money flowing around professional sports has an impact on the stock market? This paper aims to find out, and - depending on the results - maybe the research below could be used in part to make those invaluable predictions.

Data

Collection

To dig into this question, two sets of data were gathered and cleaned: the volume and the open, high, low, and closing price of various stocks over the last 42 years, and the performance - game by game - of every sports team in the NFL, MLB, and NBA over that same span. This was done using Python's selenium, BeautifulSoup, requests, and regular expressions modules to scrape data from Yahoo! Finance and various sports reference sites.

Getting to the details, then, for stock data we wanted indexes or large funds that represent the market or one of its sectors. Accordingly, we used the NASDAQ Composite Index, Dow Jones Industrial Average, S&P 500, Marathon Oil, Chevron Oil, Exxon Mobil, the Franklin Gold and Precious Metals Fund, First Eagle Gold Fund, Sturm and Ruger Guns, American Outdoor, Anheuser-Busch, two sin stocks (VICEX and RYDEX), and index funds for each of: utilities (Vanguard), healthcare (Vanguard), IT (Vanguard), and long-term bonds (Fidelity). Regarding the sports data, we didn't discriminate: If a team played a game in the NFL, MLB, or NBA since 1975, they're in the dataset, along with the result ('W' for win or 'L' for loss), score, and opponent from each of their games

Cleaning

Our scraped data was, for the most part, very clean, but there were some minor issues. From the stock data, a lot of entries were missing. Some of these were from New York Stock Exchange shut downs, while others reflect instances when data (usually the volume) wasn't reportable. Moreover, some of the stocks aren't as old as others, so all of our analyses had to handle extreme numbers of N/As. The sports data presented unique challenges, too. At least one name (the St. Louis Cardinals) has concurrently referred to a team in more than one professional league. Furthermore, it's possible for NFL games to end in a tie, and the MLB and NBA have both gone on strike during periods of the dataset. Finally, the biggest challenge with the sports data was handling teams that have changed names, location, or both, often more than once. For instance, the Charlotte Hornets went from the Charlotte Hornets to the New Orleans Hornets, then to the Charlotte Bobcats, and finally back to the Charlotte Hornets, all in the span of a decade.

Once the problems just listed were identified, a large portion of our cleaning process was re-formatting. This was true, to a lesser extent, for the stock data, and especially for the sports data. Speaking to the stocks first, to conglomerate the different-sized results from each stock we used the Python pandas and numpy modules to build a huge dataframe of standardized (by date) columns corresponding to each stock. We then used a script to fill in this dataframe with the appropriate values for each stock as read from our scraped data file.

Formatting the sports data was more difficult, because there wasn't a uniform set of dates for the games. To combat this, we built another script to again create a large dataframe. This time, we had the script pass the dataframe every date since 1975 as an index and every sports team in our study as a separate column. The script then goes through our scraped data from the professional leagues and fills the results into the cells by team and date.

After the re-formatting, the more-involved cleaning began. For the stock data, we had to deal with the missing values. In all, there were 46,347 such values, most of which came from stocks that never report their daily volume. Accordingly, we dropped all the columns of empty volumes. After this, we used numpy to make sure every value that should be a number is actually a number, which they were. Thus, from our stocks data, we cleaned 46,347 missing values and dropped 6 effectively empty variables.

The sports data was more difficult to clean. To start, we decided the data from when the leagues were on strike should be left, since it *is* accurate, and could theoretically produce interesting results. For our first act of cleaning, then, we wrote a patch in our script to handle the one team name that has concurrently referred to teams in multiple leagues: the St. Louis Cardinals. To do this, we changed all occurrences of the St. Louis Cardinals from the NFL to "St. Louis Cardinals (Football)". Next, to handle the NFL's ties, we checked all the game results to see if the winner and loser had the same score. If they did, we changed the outcome from "Winner" and "Loser" to two "Tie"s; as it turned out, there have been comparatively few ties - so few, in fact, that they can be safely ignored. After that, we handled the teams that have moved locations and/or changed names. To do this, we created a list of every team that has relocated/changed names and its new name and location. We then used this list to merge the columns of the old teams with the new teams'; so, for instance, the results of the "Brooklyn Dodgers" and "Los Angeles Dodgers" will appear in the same variable. This took our dataset from 120 variables down to 95. Finally, then, we ran the sports data through various tests to make sure the values were accurate and sensible (for example, 'Does the winning team have more points than the losing team?'). Luckily, all values in our final dataset passed these tests, and so we moved on under the assumption that our data was suitable for future treatment and analyses.

Variable Creation

Before moving all the way forward to analyses and hypothesis testing, though, we needed to build some custom variable types for the work we planned to do. We subsequently added three new variable types--one for the sports data and two for the stocks. For the sports data, bin-like variables were created to classify every day, for every team, as -1 (a loss), 0 (no game), or 1 (a win). For the stock data, the first of the two added variable types is a normalization technique; a variable was added to each stock that tracks the daily percentage change in the closing price from the day before's. This was done to ensure that dollar value changes in, say, the price of a single share of the Dow Jones Index (which is currently at over $24,000) aren't weighted the same as changes in the Treasury Bonds Index (a share of which costs around $7). The other variable type added to the stock data uses percentiles to bin the just-discussed percentage changes into five groups based on their size relative to other changes in that stock: BIG JUMP (>95% of the recorded percentage changes for that stock), JUMP (85%<X <95%), LITTLE MOVEMENT (35% <X<80%), DIP (5%<X<35%), and BIG DIP (<5%). So, for example, a daily change of +$0.13 in the Bond Index price would fall between the 35th and 80th percentile, and so would be added to the "BONDS Change Category" column as "LITTLE MOVEMENT". We created these bins to better suit some of the hypotheses that will be covered later in the paper. (To get a better sense for these new stock variables and the way they shake out and relate to the others, see the interactive plot in Appendix A).

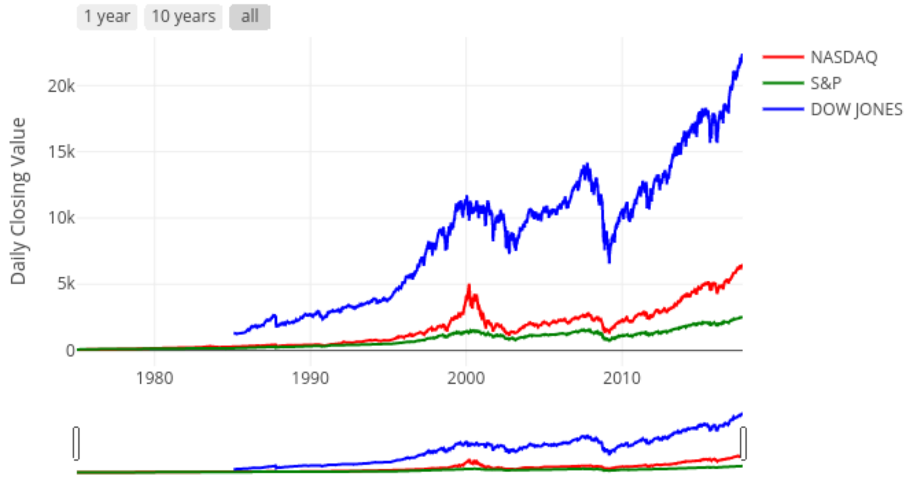
As a final task, we added four new variables to the sports dataset. These variables count, for every day, how many NBA, MLB, NFL, and total professional games were played. This was done with an eye towards a specific hypothesis we had in mind, which can be found in the section headed Hypothesis 3. In sum, then, we used a total of 324 variables--130 and 194 from the stocks and sports datasets, respectively.

Exploratory Data Analyses

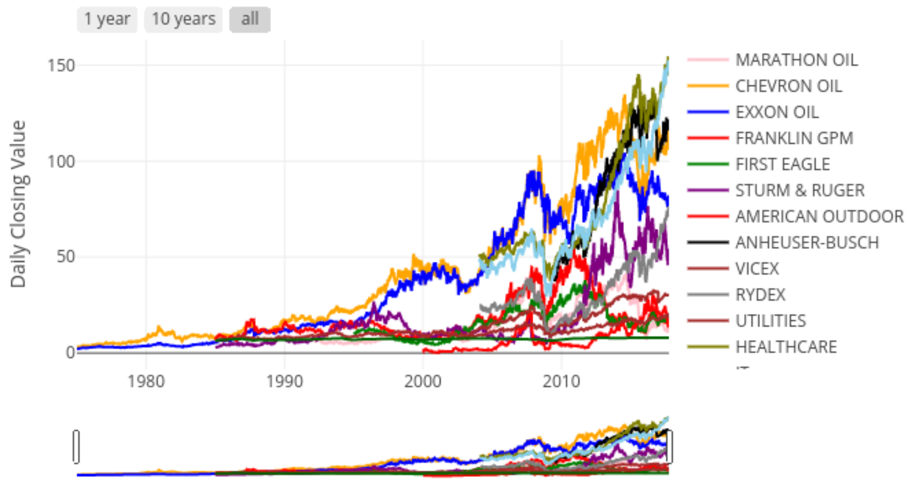
Basic Visualizations

As with most data science projects, it is useful to look at our data before running through the results of further testing. Following are the trend lines for each of the stocks we tested, separated into plots for total market index funds and individual stocks (the plots can be adjusted to show the results from the last year, ten years, or all of our data, with a slider at the bottom of the plots for further, customized investigation):

Index Growth Since 1975

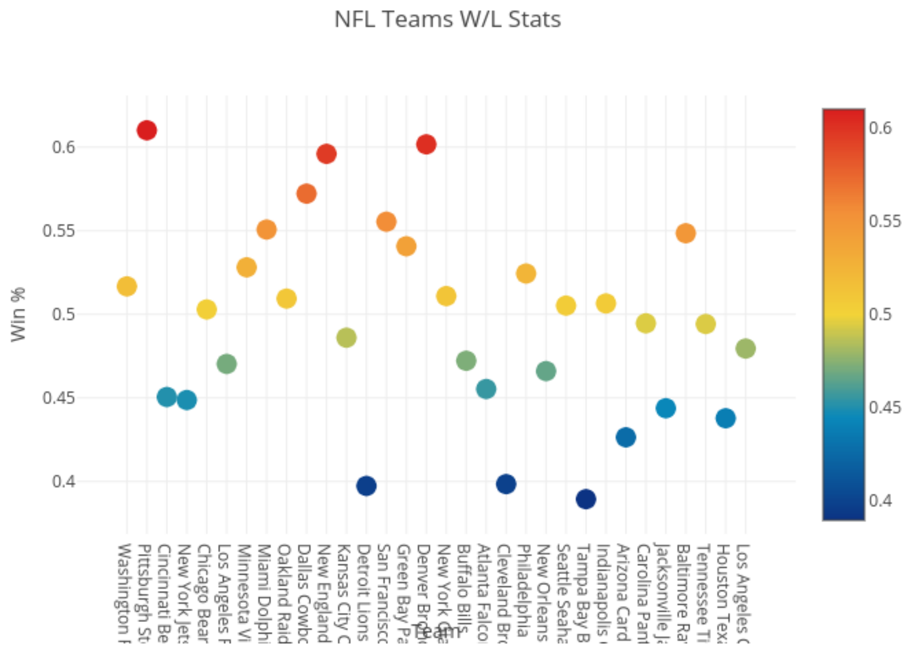


Single Fund Growth Since 1975 or Inception

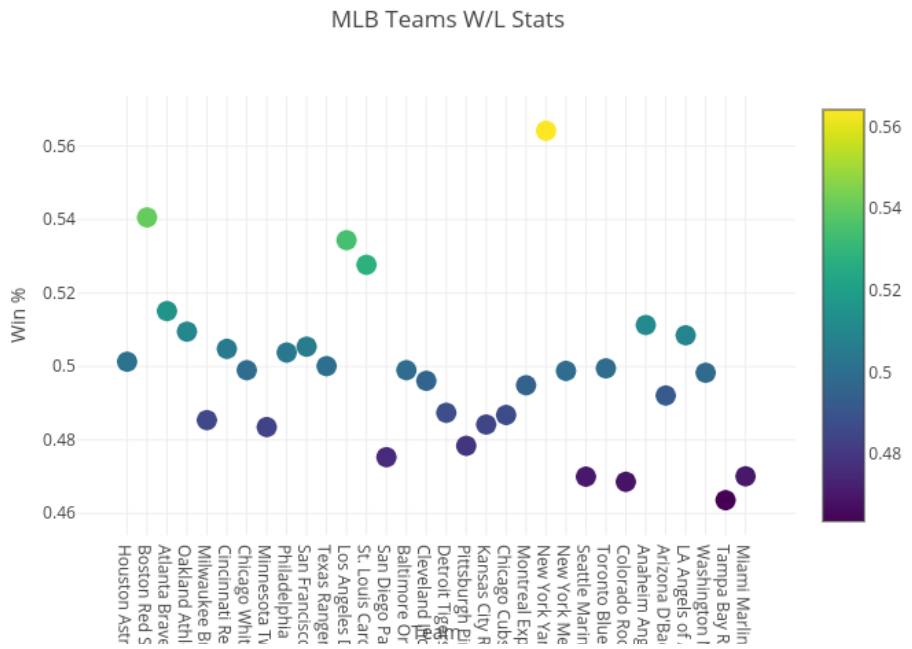


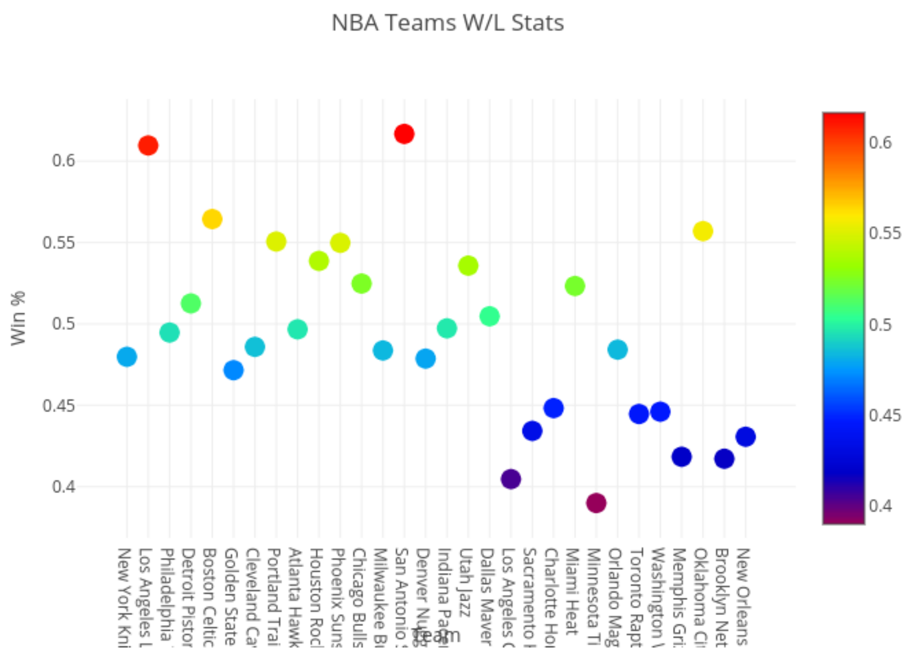
As can be gleaned from the plots, the stock market as a whole rose considerably over the tested timespan. Furthermore, the similarity of the two plots is a vote of confidence for our choices of industry-specific single stocks: Because the shape of the lower plot so closely resembles that of the one above it containing total market indexes, we can be fairly sure that we've accurately captured the minutiae of the market; if the shapes were very different, we'd know we left out a few sectors that greatly affect the market's movement. As a final note about these plots, the tech bubble swelling and bursting in the '90s and early 2000s is noticeable, especially in the larger stocks. This will be something to watch out for: The performance of teams that were good in the late '90s and early 2000s and bad soon thereafter will be highly correlated with the market.

Next is the winning percentage of every team we tested, since 1975 or that team's inception, sorted by the league in which they play(ed):



(If there are any issues with the axes or labels - in this plot or elsewhere - try loading the page in a different browser, like Firefox.)





Looking at the plots, the Pittsburgh Steelers, New York Yankees, and San Antonio Spurs have the highest winning percentages in their respective leagues over the time we tested. This is another thing to remember: Since the stock market rose considerably over the timeframe we tested, the performance of those teams that have won most often might have higher correlation with the market's activity. But - and this is reason for optimism - nearly all of the winning percentages land in the small interval between .45 and .55, and so the impact one team's being consistently better than another will have on our results should be minimal. That being said, there are four teams with winning percentages below .4: the Detroit Lions (.397), Cleveland Browns (.398), Tampa Bay Buccaneers (.389), and Minnesota Timberwolves (.39). It'd be very surprising, therefore, to find out that any of their performance correlate with the market's movements.

Basic Statistical Analyses

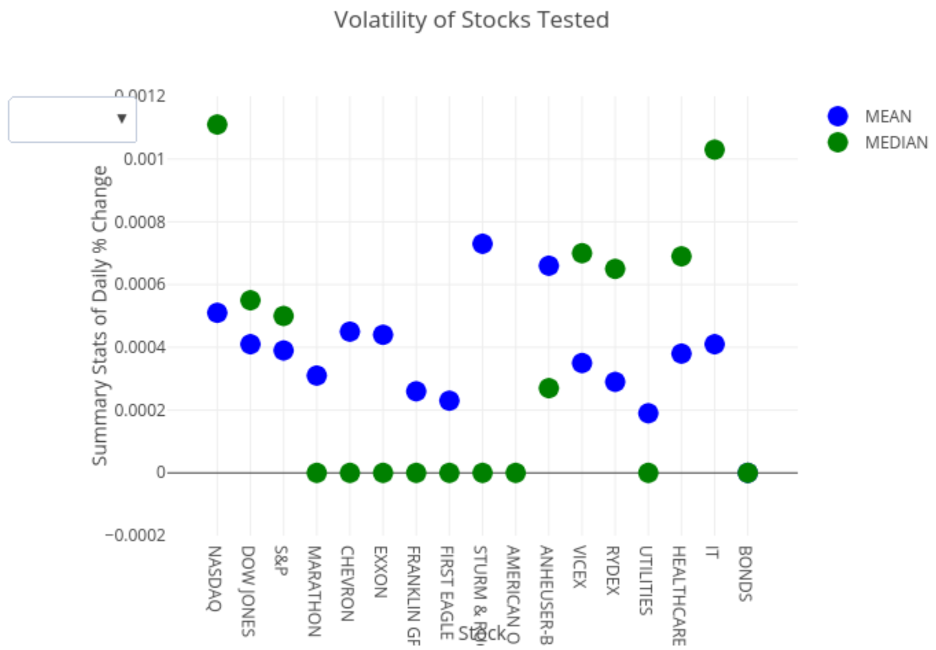
With the above plots giving us a basic idea of what to expect, we felt comfortable moving forward with some exploratory statistical analyses. For the most part, these analyses were done on the stocks data, since the sports data isn't conducive to such tests; the sports data is much more important to the hypothesis testing below, since it mostly consists of binary, W/L "Outcomes" variables. With respect to the stocks data, then, each test was performed on the percentage change variables. These are the best variables to test because, ultimately, the changes in the market are entirely contained in these variables. The tests performed on the percentage change variables, accordingly, were: 1) mean, median, and standard deviation calculation, 2) outlier detection and handling, 3) frequency measurements, and 4) correlation calculation.

) Mean, Median, and Standard Deviation

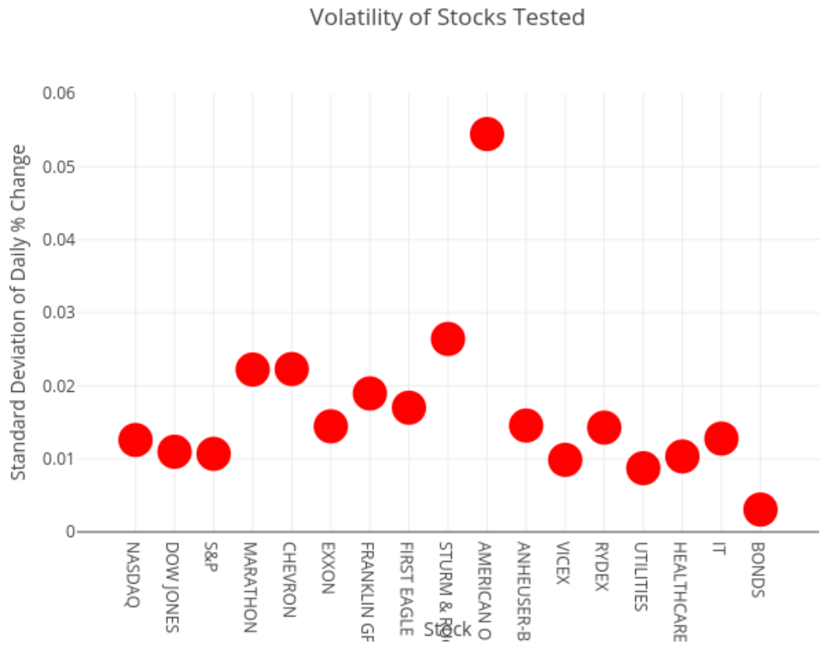
First, the results from the calculations of the mean, median, and standard deviation of each of the seventeen percentage change variables corresponding to the seventeen tracked stocks are below:

STOCK	Mean (Percentage Change from Day Before)	Median	Standard Deviation
NASDAQ Index	.00051	.00111	.01258
Dow Jones Index	.00041	.00055	.01097
S & P Index	.00039	.0005	.01068
Marathon Oil	.00031	0	.02222
Chevron Oil	.00045	0	.02229
Exxon Oil	.00044	0	.01445
Franklin Gold and Precious Metals Index	.00026	0	.01895
First Eagle Gold Fund	.00023	0	.01700
Sturm and Ruger Guns	.00073	0	.02641
American Outdoor Guns	.00197	0	.05445
Anheuser-Busch	.00066	.00027	.01456
VICEX (Sin Index)	.00035	.00070	.00986
RYDEX (Sin Index)	.00029	.00065	.01427
Utilities Index	.00019	0	.00872
Healthcare Index	.00038	.00069	.01033
IT Index	.00041	.00103	.01277
Treasury Bonds Index	0	0	.00306

Or, as a visual, the table above - separated by statistic - looks like:



(This can be adjusted to isolate just the mean or the median.)



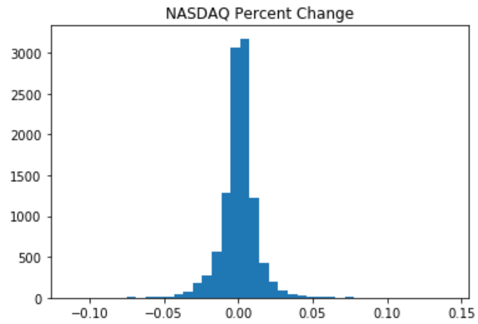
These metrics and plots express little more than that the data appear to behave the way they should. Treasury Bonds excluded, the means and medians show that each stock averaged a tiny daily rise, which mirrors the stock market as a whole since 1975. And, as would be expected, the stocks from sectors that are a bit more volatile (oil funds and single stocks like Sturm & Ruger or American Outdoor) have larger standard deviations and slightly larger means than more stable funds (like Treasury Bonds). For those funds, it will be important to remember that such volatility could affect our results. If a team's performance seems to be correlated with that stock's movement, it may just be the case that that stock produced such results because it regularly shows more fluctuation than others. In any case, these basic statistical analyses paved the way for us to perform some more robust tests on the data.

) Outlier Detection and Handling

Before any of those more robust tests, though, data should always be checked for outliers and their impact on the results. Although the sports data was not put through such testing - because there are no variables in the dataset that could produce outliers - each variable in the stock data was checked. To find outliers, a modified***** Interquartile Range test was performed for each of the daily percentage change variables. *(*Traditionally, an outlier is considered to be any value that is more than 1.5 times the interquartile range belowthe 25th percentile or above the 75th percentile. With data that fluctuates as much as that from the stock market, though, that process counts far too many data points as outliers - in fact, it counted 33,976. Thus, an outlier was instead counted as any price change that was* ***5*** *times the interquartile range belowthe 25th percentile or above the 75th percentile. This still found 348 outliers across the seventeen percentage change variables.)* This produced a high number of outliers, which makes sense because the stock market is incredibly volatile. No values were thrown out, though, because part of what's exciting about the stock market is its extreme swings and the idea that they can be predicted. In fact, after further investigation, all the outliers appeared to correspond to days when the market was especially turbulent (like Black Monday). There was no reason to get rid of this kind of data. If anything, the relatively large amount of extreme values suggests that there are trends to be discovered behind those swings across the stock market.

) Frequency Measurements

Next, for each daily stock price percentage change variable, a histogram was created to chart the frequency of different swing sizes in that stock. In essence, we wanted to check the distributions of the percentage change variables to be sure there wasn't any troublesome skew. As such, this shed more light on the percentage change variables than just finding the mean, median, and standard deviation; as just noted, histograms give a better picture of the true spread and distribution of each variable. The histograms list the frequency in number of days on the y-axis and percentage change in the stock's closing price on the x-axis. The histogram for the NASDAQ follows, with the rest in Appendix B.



In summary, the plots show that each percentage price change variable is roughly normally-distributed around a value close to 0. This is what we hoped for. If any of the stocks had been skewed one way or another, or clustered around, for instance, .02, that data would not have been usable because that stock would too often show correlation. The only plots that cause any of this type of concern are those for the Treasury Bond Index and Sturm & Ruger (Guns). For the former, a slightly higher than average amount of its daily percentage changes appears to be below 0; this could mean there were a high number of dips in the Bond Index--more likely, though, is that the Bond Index just remained steady (as bond indexes do) while other stocks grew with inflation. And, for the latter, it looks like an inordinately-high number of days saw a percentage change just below 0; this indicates that Sturm & Ruger (Guns) may have had a rough decade that skewed its overall performance. These will be things to watch out for, but, thankfully, each stock's histogram appears to show a normal distribution centered around a number very close to 0, and so it can be safely assumed that none of the results will be horribly skewed by a stock that out- or underperformed its peers over the timeframe we tracked.

) Correlation Calculations

To round out the basic statistical analyses, a final "behavior check" was performed on our data. We found the correlations between the percentage change variables for five disparate stocks: the NASDAQ Index, VICEX (Sin) Index, Vanguard Utilities Index, American Outdoor (Guns), and the Treasury Bonds Index. This was done to confirm that the data isn't affected by any time-specific trends that overtook the whole stock market or a few of its sectors in the timeframe we chose. Because nothing like a "50-year spike" is currently being lauded by Wall Street, though, we were optimistic about these tests. Since the stock market has consistently gone up over the last 42 years, a broad positive correlation - but nothing too significant - was expected. This expectation was confirmed by low correlation numbers, as shown in the table of correlation coefficients below (corresponding scatter plots can be found in Appendix C):

STOCK	NASDAQ	VICEX	Utilities	American Outdoor	Treasury Bonds
NASDAQ	1	.382	.460	.641	-.0047
VICEX	.382	1	.242	.371	-.0147
Utilities	.460	.242	1	.244	.038
American Outdoor	.641	.371	.244	1	-.025
Treasury Bonds	-.0047	-.0147	.038	-.025	1

The NASDAQ Index has semi-alarming correlations with American Outdoor (Guns) and the Utilities Index, but neither is large enough (.641 and .460) to cause real concern; and, considering the NASDAQ is a total-market index fund, it's unsurprising to find it somewhat correlates with individual stocks, since both the market as a whole and most individual stocks rose over the period we're testing. Based on these results, the assumption that none of the percentage change in stock price variables are highly-correlated with any other seems justified.

The basic statistical analyses, therefore, suggested that the stock data is suitable for comparison with the sports data. It should be mentioned that the stock data was put through various clustering algorithms to look for any so-far-undiscovered trends or similarities, too. This didn't end up producing results pertinent to the project, but the process is included in Appendix D, for transparency.

Looking for Hypotheses - Association Rule Mining

With all the preliminary testing and exploration completed, we could start to look at the relationships between teams' performance and the stock market. To best do this, the following variables were examined: single game outcomes for the teams, and the daily percentage change from the stocks. As has been mentioned, these are most closely-related to the data science question under consideration.

The Apriori algorithm was run on test data consisting of three large-market sports teams and a representative stock from each sector: 1) the Dallas Cowboys, New York Yankees, and Los Angeles Lakers, and 2) the NASDAQ Index, Exxon Oil, Franklin Gold and Precious Metals, Sturm & Ruger (Guns), Anheuser-Busch, VICEX (Sin), Utilities Index, Healthcare Index, IT Index, and the Treasury Bonds Index. The thinking behind these choices was twofold:) On the one hand, running the Apriori algorithm on our data took quite some time, so there was no feasible way to test every combination of the 96 teams and 17 stocks; and, second,) If any team's performance really does affect the market, those teams will almost certainly be from large markets, as these teams spend and earn exponentially more money than their small market competitors. (As a synecdoche, the 2017 opening day payroll for the Los Angeles Dodgers was well over $200 million, while that of the frugal Milwaukee Brewers was less than $65 million.)

Skipping over the particulars of the tests (which can be found in Appendix E), the association rule mining produced little to be excited about, as evidenced by the table below:

Event A	Event B	Support	Confidence
Dallas Cowboys WIN	Exxon LITTLE MOVEMENT	.2762	.6210
Dallas Cowboys WIN	Sturm & Ruger (Guns) LITTLE MOVEMENT	.2754	.6151
Dallas Cowboys WIN	Anheuser-Busch LITTLE MOVEMENT	.2593	.4795
Dallas Cowboys WIN	VICEX (SIN) LITTLE MOVEMENT	.2720	.5812
Dallas Cowboys WIN	BONDS LITTLE MOVEMENT	.2609	.4751
New York Yankees WIN	Exxon LITTLE MOVEMENT	.2578	.4570
New York Yankees WIN	Franklin Gold and Precious Metals LITTLE MOVEMENT	.2559	.4563
New York Yankees WIN	Sturm & Ruger (Guns) LITTLE MOVEMENT	.2511	.5662
New York Yankees WIN	IT LITTLE MOVEMENT	.2549	.4469
LA Lakers WIN	NASDAQ LITTLE MOVEMENT	.2606	.6006
LA Lakers WIN	Exxon LITTLE MOVEMENT	.2647	.4338
LA Lakers WIN	Franklin Gold and Precious Metals LITTLE MOVEMENT	.2799	.4525
LA Lakers WIN	Sturm & Ruger (Guns) Little Movement	.2974	.6383
LA Lakers WIN	Utilities LITTLE MOVEMENT	.2759	.6150
LA Lakers WIN	BONDS LITTLE MOVEMENT	.2819	.45582

As is apparent, even our best support and confidence numbers were relatively low. The rule mining did, however, (very) vaguely suggest certain sectors of the market might react to the performance of the Los Angeles Lakers and the New York Yankees. With this in mind, we decided to formally test these hypotheses along with a few others that we believed to be of interest.

Analysis: Hypothesis Testing and Results

Hypothesis 1: The Stock Market Reacts to the Way the Yankees Perform.

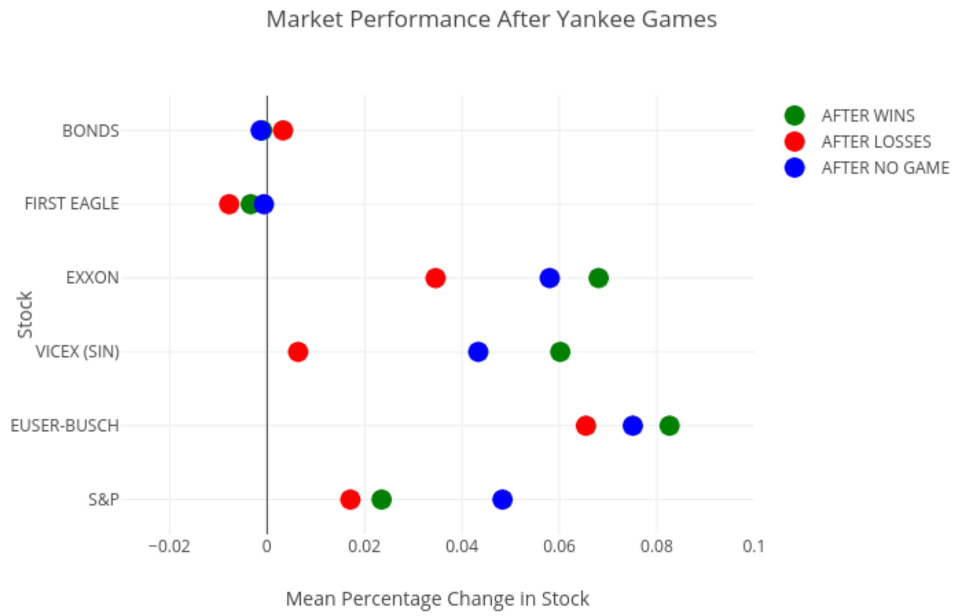
As was just mentioned, the association rule mining somewhat suggested the market reacts to the performance of the New York Yankees. Accordingly, this first hypothesis examines the relationship between the large market Yankees and movement across the stock market. Six stocks were tested: the S&P Index, Anheuser-Busch, VICEX (Sin), Exxon, First Eagle Gold, and the Treasury Bond Index. Specifically, a relationship was looked for between the Yankees' game results on a given day and the performance of the stocks on the next available day. To prepare the data for appropriate testing, a separate dataset was created containing the gameday date, a categorical variable representing a win or loss (1 if won, 0 if lost), the score differential (positive if won, negative if lost), the next available stock date after the game, and the aforementioned binned stock performance variable based on the percentage change column.

Results

Using Random Forest, Naive Baye's, and SVM classification, models were trained on the categorical win/loss variable and the score differential, and then used to predict whether the stock value would go up or down. Almost all of the prediction accuracy scores, as shown below, were slightly better than 0.50, which basically indicates that the trained models were no better than chance (which is 50%) at predicting which way the market would go. Furthermore, when we generated ROC curves, these too were almost identical to the line represented by random chance, with some kinks in certain cases (examples can be found in Appendix F). The only potentially interesting result was that the various modeling methods used with the Bond Index led to an accuracy score of approximately 0.616, which is higher than that of any other stock.

Stock/Index	Random Forest	Naive Bayes	SVM
S&P	0.50375	0.51649	0.51649
ANHEUSER-BUSCH	0.5	0.49265	0.48897
VICEX (SIN)	0.53416	0.52795	0.52795
EXXON	0.51874	0.52024	0.51574
First Eagle Gold	0.55482	0.56143	0.56143
BONDS	0.61591	0.6176	0.6176

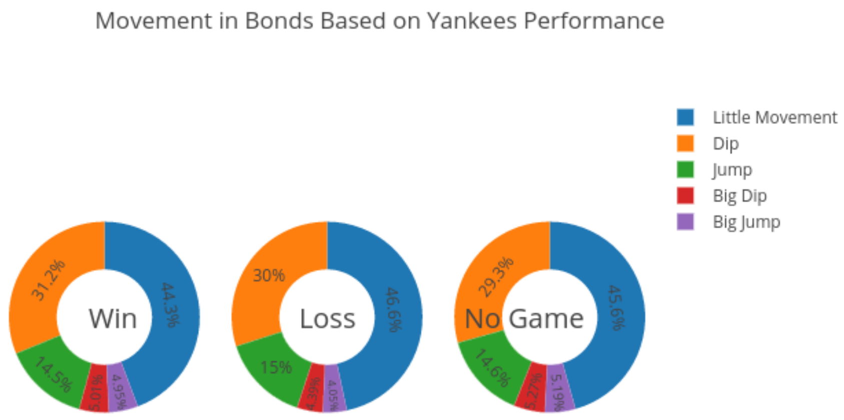
Before giving up on this hypothesis, though, we created plots with the hope of finding trends visually. First, we plotted the mean percentage change in each of the stocks on days after the Yankees won, lost, or didn't play:



(These axes are scaled for better viewing, so, for example, .02 represents a .02% - and not 2% - mean change.)

As the plot shows, none of the stocks stand out; in fact, this plot just casts shade over the only interesting result from our models. The Bond Index - the performance of which the models were better able to predict than the others - shows little difference in movement after the Yankees won, lost, or didn't play. So, really, the models just predicted the Bond Index wouldn't do much - regardless of the Yankees' performance - and this was correct about 60% of the time.

To be absolutely sure about the lack of correlation between the Yankees' performance and the Bond Index, though, we created the following pie charts that show the Bond Index's movement after Yankees' wins, losses, and days off:



As is made crystal clear by the almost identical pies in the plot, the Yankees' performance has nothing to do with the movements of the Bond Index.

These results from the plots, combined with the accuracy scores from the prediction methods, give us all the confidence we need to say the performance of the Yankees has no impact on the stock market. Thus, the null hypothesis that the Yankees' performance does not affect the stock market in any systematic way can be confidently rejected.

Hypothesis 2: The Stock Market Reacts to the Way the Lakers Perform

Because the association rule mining implied that the stock market reacts to Lakers games (that is, Lakers' associations accounted for the best of our results), this was the next idea to be put through a series of formal tests. Five stocks were used - the Treasury Bond Index, NASDAQ Index, Chevron Oil, Franklin Gold and Precious Metals, and Sturm & Ruger (Guns) - and tested to see if their activity differed based on the way the Lakers performed. Specifically, the outcome of every Lakers game since 1975 was tested against the next day's percentage change of these five stocks (where available - not all of the stocks are 42 years old).

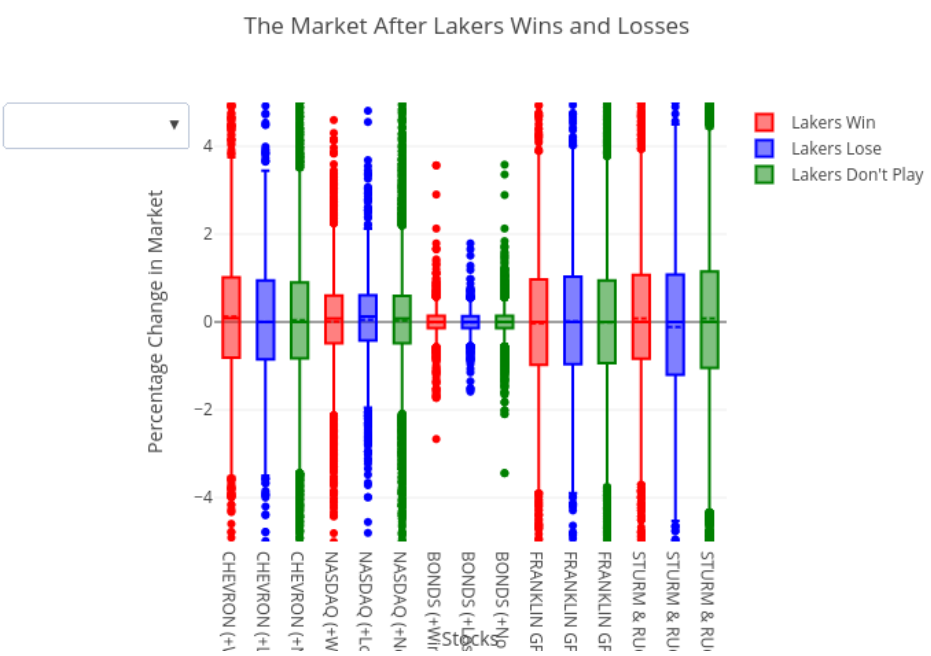
Results

First, an ANOVA calculation was performed for a difference in the mean percentage change in stock prices on days after the Lakers won, lost, or didn't play. Then, because the way the market performs when teams don't play was not of interest, t-tests were run to determine if there is a difference in market activity based on whether the Lakers win or lose. The table below lists the p-values for these tests, sorted by the stock the Lakers' performance was being tested against:

Stock Tracked	ANOVA p-value	T-Test p-value
Bonds Index	.4677	.7058
NASDAQ Index	.6417	.3434
Chevron Oil	.0393	.0237
Franklin Gold and Precious Metals	.7698	.5083
Sturm & Ruger (Guns)	.0425	.0441

As can be seen, most of the stocks produced high p-values for both the ANOVA and t-test calculations. This means that neither of their tests resulted in statistical significance; that is, the difference in the means of percentage changes in those stocks the days after the Lakers won, lost, or didn't play are not telling of an underlying trend.

But, Chevron and Sturm & Ruger (Guns) *did* seem to react to the Lakers' performance; they both produced p-values just below the standard p = .05 significance level. To better illustrate, what follows is a box and whisker plot of the percentage change in each of the tested stocks on days after the Lakers won, lost, or didn't play:



(The dropdown arrow in the upper left can be used to adjust the boxes shown in the plot.

Also, the percentage changes are scaled up, so, 2% is represented as 2, not .02; in the plot, 02 would refer to a .02% change.)

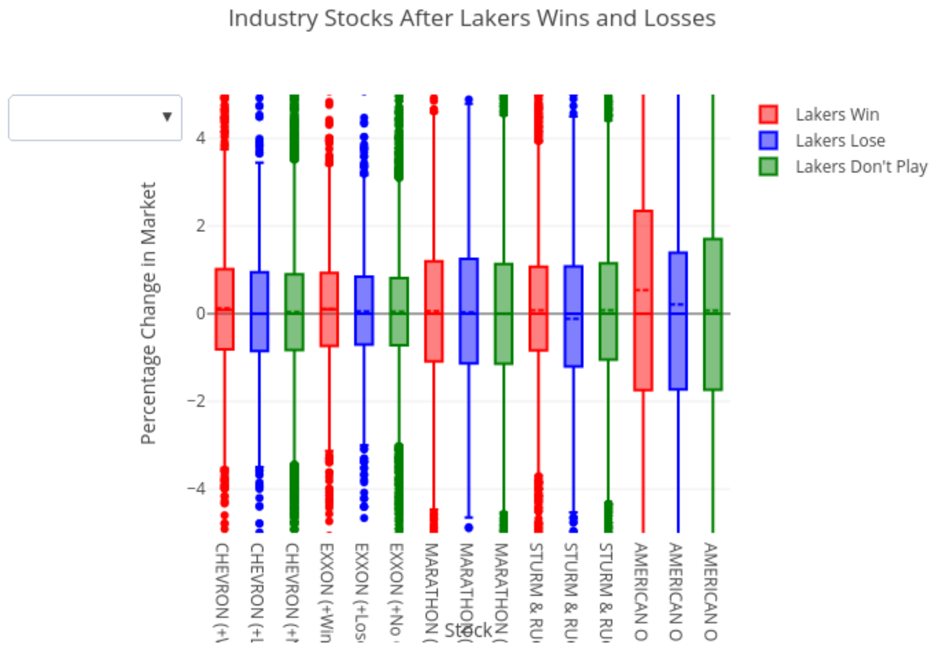
The biggest takeaways from the visual are:) as the lengths of the boxes demonstrate, neither the spread of Chevron nor Sturm & Ruger is significantly wider than the other stocks, and so the interesting correlation results aren't just due to a higher variance, and) the dotted mean lines show that the upward movement in Chevron and Sturm & Ruger on days after Lakers wins is significantly greater than that on both days after the Lakers lose and days after they don't play, which suggests there really is something different in these stocks after the Lakers win.

To dig a little deeper into these surprising findings, the mean percentage change in Chevron Oil when the Lakers lost was +.000018 and +.0012 when the Lakers won. That *is* a significant difference, especially when individual shares of Chevron currently cost ~$117. As for Sturm & Ruger (Guns), when the Lakers lost the average change was -.0011 and +.00077 when they won, with a current stock price of ~$56 per share. That's an even bigger difference. This seems very strange, and so we checked if these stocks' activity was the result of industry-wide phenomena. Maybe oil and the price of guns just happened to go up when the Lakers were winning? The results are below, presented in the same format as those above:

Stock Tracked	ANOVA p-value	T-Test p-value
Chevron Oil	.0393	.0237
Exxon Oil	.2027	.2434

Marathon Oil	.5198	.7221
Sturm & Ruger (Guns)	.0425	.0441
American Outdoor (Guns)	.0536	.3656

And the corresponding chart:



(The same formatting as above, and again adjustable with the dropdown arrow in the upper left.)

In a nutshell, no, the trends weren't industry-wide. The American Outdoor (Guns) fund stands out in the chart and has a low p-value for the ANOVA test for a difference in means, but the t-test that only considers the difference in mean percentage change on the days after the Lakers won versus those when they lost shows much poorer results. And, moreover - and as we saw earlier in the basic statistical analyses - the American Outdoors stock has a massive variance (almost 50% more than any other stock), which is almost certainly responsible for its behavior. By contrast, both Chevron and Sturm & Ruger have relatively small variances, which is further evidence that these findings may not be a fluke. Thus, even after further testing, there's reason to believe the difference in Chevron and Sturm & Ruger on days after the Lakers win versus days after they lose *is statistically significant*.

Because of the above results, this hypothesis was put through further testing: k-nearest neighbors and decision tree classifiers. These didn't produce anything of interest, but their full descriptions are included in Appendix G, again for transparency. But, although these further tests didn't produce anything useful, the hypothesis that the stock market reacts to the Lakers' performance cannot be rejected, because the ANOVA, t-test, and plotting results seem to imply that Chevron and Sturm & Ruger (Guns) really do react to Lakers games. Is this likely, or even believable? No--the Lakers probably just happened to be better than usual in years when those stocks saw nice jumps (like the late 90s and early 2000s during the tech bubble, which was warned against at the beginning of this paper). But, this cannot be solidified without further research, and so this second hypothesis cannot be confidently rejected. That is, it is not certain that the Lakers' performance does not affect the stock market.

Hypothesis 3: The Stock Market Reacts to the Number of Professional Sports Games Played the Day Before

This final hypothesis isn't based on the results of our rule mining, but is rather something we wanted to investigate from the first time this project was discussed. To test it, the percentage changes in daily stock prices the day after varying numbers of professional sports games were played were measured. For example, does the stock market see a bump the day after NFL Sundays, when there are usually at least 12 games played? The analyses below measure the NASDAQ Index against the number of MLB games, NBA games, NFL games, and total games across the three professional leagues. This was the only variant of the hypothesis that seemed plausible--that the market as a whole (which is the reason for using a total market index) reacts to the number of games played. Furthermore, the NASDAQ Index is the oldest of the tracked stocks, and so has the best chance of producing relevant results.

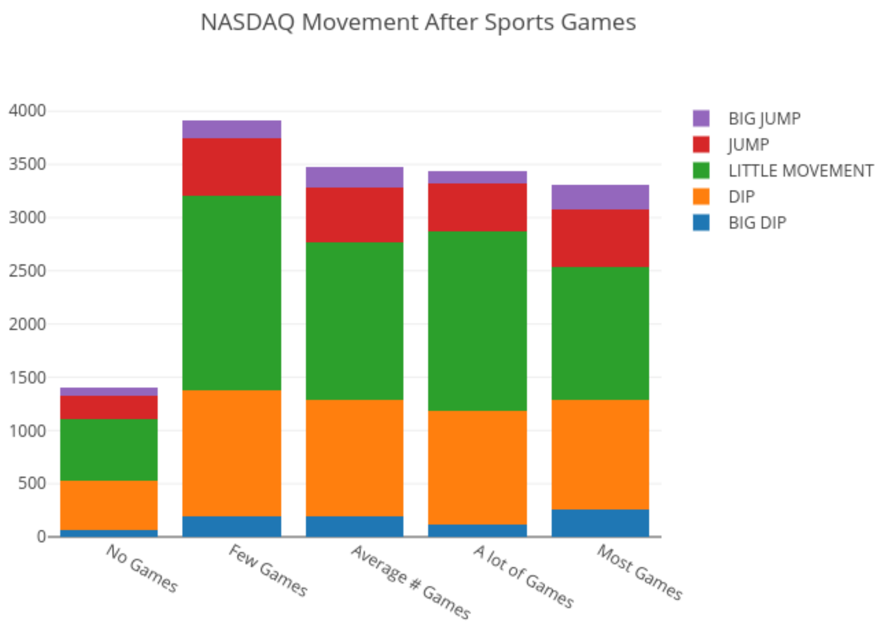
Results

To test these 4 possible relationships - the NASDAQ's movement after differing numbers of NFL, MLB, NBA, and total professional games - the number of games were binned into 5 buckets, and then an ANOVA test was conducted for a difference in the mean percentage change in the NASDAQ following days with numbers of games that fell in those buckets. The results are below:

NASDAQ Percent Change Based On:	ANOVA p-value
Number of MLB Games	.1753

Number of NBA Games	**.4924**
Number of NFL Games	**.7064**
Total Number of Professional Sports Games	**.9850**

None of these p-values were high enough to be statistically significant, and this final hypothesis' validity thus seemed unlikely. (The p-value for MLB Games is kind of interesting, but if the bins are altered at all it rockets up). The following stacked bar chart corroborates this sentiment:



As could be expected from the ANOVA tests, the bars are identical once size is accounted for. That is, the percentage of days in each bin was nearly the same for days after no games, few games, an average number of games, a lot of games, and the largest bucketed number of games. The likelihood of this hypothesis was thus very much in doubt.

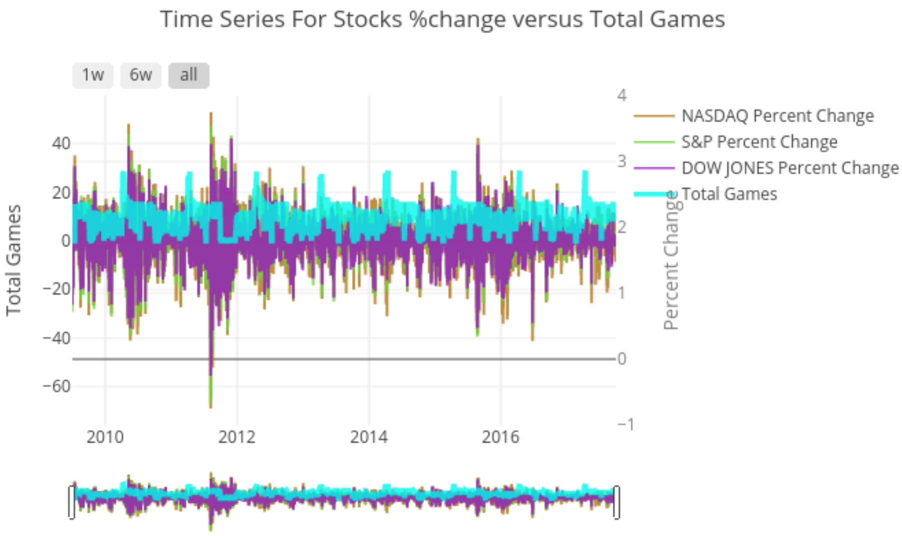
A regression line was fit to the data, though, so as to be certain that this hypothesis should be rejected. In each case, the regression analysis produced insignificant p-values (the likelihood that the coefficient related to the numbers of games played is equal to 0 and so irrelevant to the NASDAQ's movement). The results are below:

NASDAQ Percent Change Based On:	**Linear Regression Analysis p-value (Likelihood that Number of Games Coefficient is Meaningless)**
Number of MLB Games	**.113**
Number of NBA Games	**.120**
Number of NFL Games	**.149**
Total Number of Professional Sports Games	**.928**

The p-values are somewhat close to significant, but if the linear model is made more complex the p-values shoot up. For example, if the NASDAQ's percentage change is modeled according to three variables - the number of MLB games, NBA games, and NFL games, separately - the p-values reach: .160 (NFL), .307 (NBA), and .506 (MLB). The NFL p-value stays relatively low, but when put in the context of the corresponding high ANOVA value above, it seems to be accountable to coincidence. One possible explanation is that the market usually has the highest change from Friday to Monday, since the market is closed on weekends; and, since most NFL games are played on Sundays, it's nothing more than a coincidence that the market fluctuates following days with NFL games. As can be seen in the following charts, this suspicion is confirmed:

The stock market usually has the lowest opening values on Mondays, and has some of the highest closing values on Fridays, which obviously best corresponds with the dark green line that represents the days with the most NFL games.

For extra certainty, though, we created the following plot that shows normalized versions of the number of total games played and the percentage changes in the S&P, Dow Jones, and NASDAQ indexes for every day since the beginning of 2009. We wanted to see if any trends that may have been missed by the just-discussed statistical analyses stood out when the data was presented visually, and furthermore felt restricting the plot down to a smaller timeframe of eight years (with a zoom slider to allow even more specific investigation) would be the best way to do so.



Despite the mass of data, it’s pretty apparent that there are no trends between the lines. The zoom slider makes this even more obvious on the smaller scales.

Thus, from these tests, the third hypothesis - like the first - is rejected: it can be confidently concluded that the market does not react to the number of professional sports games played the day before.

Hypothesis 4: The Market Reacts to Teams' Playoff Performance

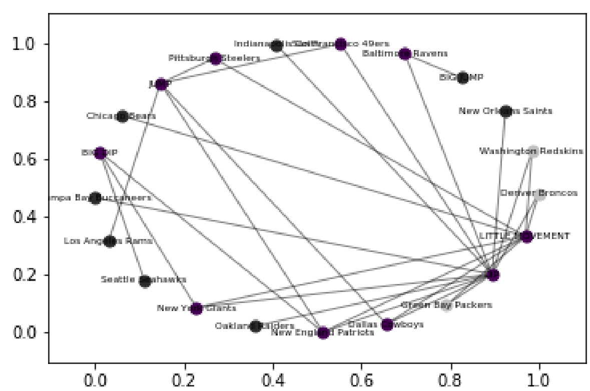
Our first three hypotheses produced a mix bag of results. Accordingly, we wanted to try to isolate some of our interesting findings in this final test. As has been covered, there was *some* evidence in Hypotheses 1 and 2 that the market reacts to teams' performance. However, this evidence was very diluted; even for the trends that supported our hypotheses, the best and most logical explanations seemed to be chance, since those trends were so anomalous. This concluding hypothesis, then, is an attempt to start pinpointing the reasons behind those anomalies.

What - if not *regular* season performance - could cause some teams to be impactful in the market? We thought, potentially, *post*season performance. Maybe this can account for some of the strangeness in our data--certain teams performing well in the playoffs and wining championships could theoretically cause the most movement in the market. After all, playoff and championship rounds of professional sports are the highest - and most lucrative - stages for both individual and team achievements, and so are also the most watched and cared about by fans.

Results

To test this, we performed two network analyses. They take into account NFL and NBA champions over the date range of our collected data, and the movement of the NASDAQ Index. As in Hypothesis 3, we thought that, if any stock is going to show movement, it'll be the oldest total market index. Furthermore, we chose to perform a network analysis because it's best suited to literally "map" the movements of the NASDAQ after various teams win championships. For example, if every time the New England Patriots won a Super Bowl the NASDAQ jumped up, there will be a visible line representative of this change; we'll be able to *see* the relationship. As for the particulars of the graphs, every edge connects a championship winner to the NASDAQ's movement as specified by our oft-used percentage change category; for instance, the Seattle Seahawks to "BIG DIP".

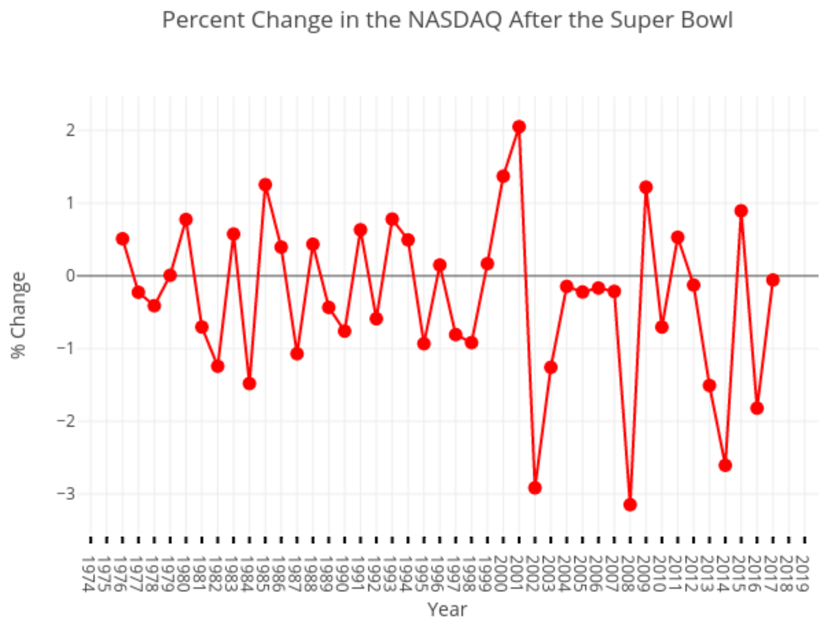
Moving to the first graph, the following connects NFL (Super Bowl) champions and the NASDAQ. There are 21 nodes, 29 edges, and 1 connected component.



None of the node's degrees are particularly high, with an average of just 2.762. The density of this graph is 0.138, with a diameter of 5, indicating it is not dense, either. The clustering coefficients are all 0, further suggesting that there is no well-defined clustering tendency. Following the trend, none of the nodes have a high betweenness score; the highest such score is the NASDAQ experiencing a "DIP", with that node scoring .6342. This was all expected, though, as there were comparatively few values (Super Bowl winners) in our collected data, and so none of the traditional network metrics were presumed to produce great results. Interesting to note, however, is the following distribution of the NASDAQ's movements after the Super Bowl:

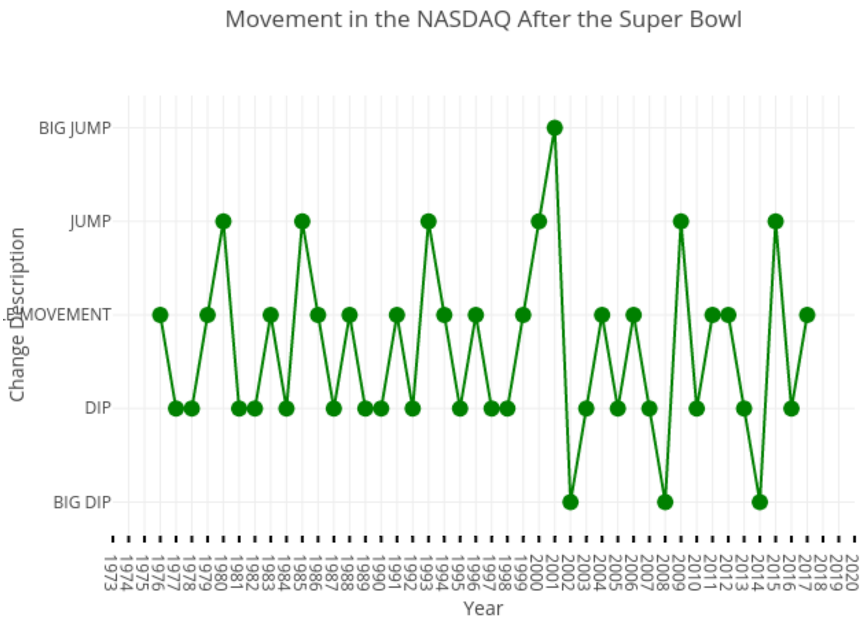
Movement Type	**Number**
BIG DIP	**3**
DIP	**18**
LITTLE MOVEMENT	**14**
JUMP	**7**
BIG JUMP	**1**

These don't stray far from the expected normal distribution of stock movements, with the middle three values accounting for 39/42 of the results. But, five of those "DIP"s came in the last ten years, with two of the "BIG DIP"s corresponding to the Seahawks' and Giants' championships in 2014 and 2008 (*Both heavy underdogs, for what it's worth*). In fact, eight of the last fifteen Super Bowls have been followed by a "BIG DIP" or "DIP", with five others seeing only "LITTLE MOVEMENT". This seems exciting, and those other Super Bowls in the last fifteen years? Just "JUMP"s from the Patriots' and Steelers' championships in 2015 and 2009 (*Heavy favorites, again for what it's worth)*. So, there are some interesting results here. Since 1975, there have been three big dips in the market after the Super Bowl, each of which have come since 2002; and, of the 21 total dips in the market since 1975, 9 have come in the last fifteen years. But, as was noted, this is an extremely small dataset, and so these results should be taken with a huge grain of salt.



(The percentage changes are again scaled up, so, as before, 2% is represented as 2, not .02; in the plot, 02 would refer to a .02% change.)

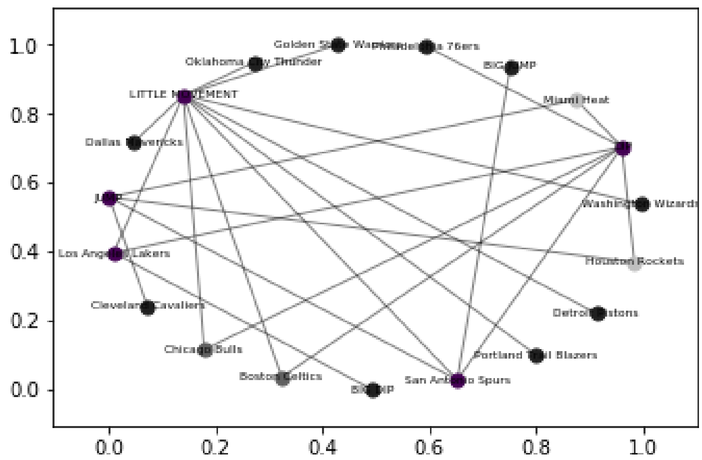
And, when we look at it graphically, the results aren't as exciting. As can be seen above, the change in the market after the Super Bowl appears to more often be below the 0 midline since 2000, but there doesn't appear to be any striking trends. That's mirrored in the next chart, too:



There aren't as many jumps as there used to be after the Super Bowl, but, again, not really a trend to be found here.

Moving to an identical analysis of NBA champions, then, there are 19 nodes and 23 edges, with an average degree of 2.42. Again - and as could be expected - this is not particularly high. The density of this graph is 0.135, with a diameter of 5 - it is not dense either. The clustering coefficients are all 0, and the highest betweenness score is this time "LITTLE MOVEMENT", with a score of .624.

The graph:



("But wait a minute!" says the sports fan. "The Thunder have never won a title!"

Correct - but remember the merging mentioned earlier: The Seattle Supersonics' lone title

falls in that column now, too.)

And corresponding table:

Movement Type	Number
BIG DIP	1
DIP	12
LITTLE MOVEMENT	23
JUMP	5
BIG JUMP	1

Again, these don't stray from the expected normal distribution of stock movements. And, if we look back at the last fifteen years, the results aren't nearly as interesting as those from the NFL. The last fifteen years of movements after NBA titles are distributed as follows:

Movement Type	Number
BIG DIP	1
DIP	3
LITTLE MOVEMENT	6
JUMP	4
BIG JUMP	1

Nothing really of interest here. Ironically, that one "BIG DIP" is from the Lakers' championship in 2008, which isn't at all what would've been expected based on our previous results. Even that, though, is not telling of anything: this is a very small dataset, and one hiccup is hardly a trend.

In conclusion, our final hypothesis produced more mixed results. NBA championships, apparently, have nothing to do with the market, but it looks like the market has *somewhat* performed poorly in recent history following the Super Bowl. One possible explanation, if there is really something there? With the invention and rise to popularity of high-frequency trading (HFT) in the last few decades, a lot more of the market's movements are influenced by market-makers, especially the firms on Wall Street. Is it possible the people overseeing HFT are a bit lethargic in the aftermath of Super Bowl parties, and so the market slows a bit? Only more analysis could produce any certainty, but the evidence above suggests that may be the case.

Conclusions and Limitations of Findings

We were generally happy with our results, but our analyses were hardly exhaustive. We only formally tested two teams against the performance of a subsection of our tracked stocks. This was due to time restraints: It'd take significantly longer to get through the 95 teams for which we have data. If that time were available, though, we think there's a lot more to be found. Are any other stocks' and teams' performance as correlated as the Lakers with Sturm & Ruger and Chevron? Does the spread - or cover status thereof - in the Super Bowl matter in the market? Is our assuming that the best results would be produced by large market teams warranted? Do teams in, say, the NFL show better results than the MLB? These and many other questions are all things that could be the focus of future research into our data. If you have any other ideas, we'd love to hear about them in the comments.

Also, our analyses were a bit naïve. We *just* tested the performance of teams and stocks; we didn't look into any other stock market factors. The tech bubble was noted above, but there are certainly other events (elections, natural disasters, the Bernie Madoff scandal, etc.) that impacted the market. Accounting for these should also be a priority in any further work.

But, as was shown above, our work *did* produce some interesting results. For the most part, they were disappointing - the Yankees' performance doesn't affect the market, and the number of games played likewise has no impact - but not all were. The Lakers' winning does seem to have a positive impact on Sturm & Ruger and Chevron, and there's a hint of evidence that the market has suffered after recent Super Bowls. Are these truly correlated? It's hard to believe so. Our findings were probably caused by extraneous factors that were unaccounted for. But, the facts are the facts, and selling off Chevron during Lakers winning streaks and buying it back up when the purple and gold were struggling over the last 40 years would've made an investor very rich. Similarly, loading up on the NASDAQ after the last two decades of Super Bowls would've strengthened any portfolio. With more research, maybe there are more things like these to be found. And - Who knows? - maybe there's a zany investment strategy hidden in our data, too.

References

[1] Sue Chang, “Here’s All the Money in the World, in One Chart,” in MarketWatch, Nov. 28, 2017. Web.

[2] David Scutt, “Here’s How Much Currency is Traded Every Day,” in Business Insider, Sep. 2, 2016. Web.

[3] Charlotte Carroll, “World Series Game 7 Tickets Will Cost You a Pretty Penny,” in Sports Illustrated, Nov. 1, 2017. Web.

[4] James Williams, “World Series Ticket Prices for Game 7: Seats Available, but Bring Your Wallet,” in Los Angeles Daily News, Oct. 31, 2017. Web.

[5] “Revenue of the Los Angeles Dodgers (MLB) from 2001 to 2016 (in million U.S. dollars),” in Statista, available at: https://www.statista.com/statistics/196665/revenue-of-the-los-angeles-dodgers/. Web.

Comments

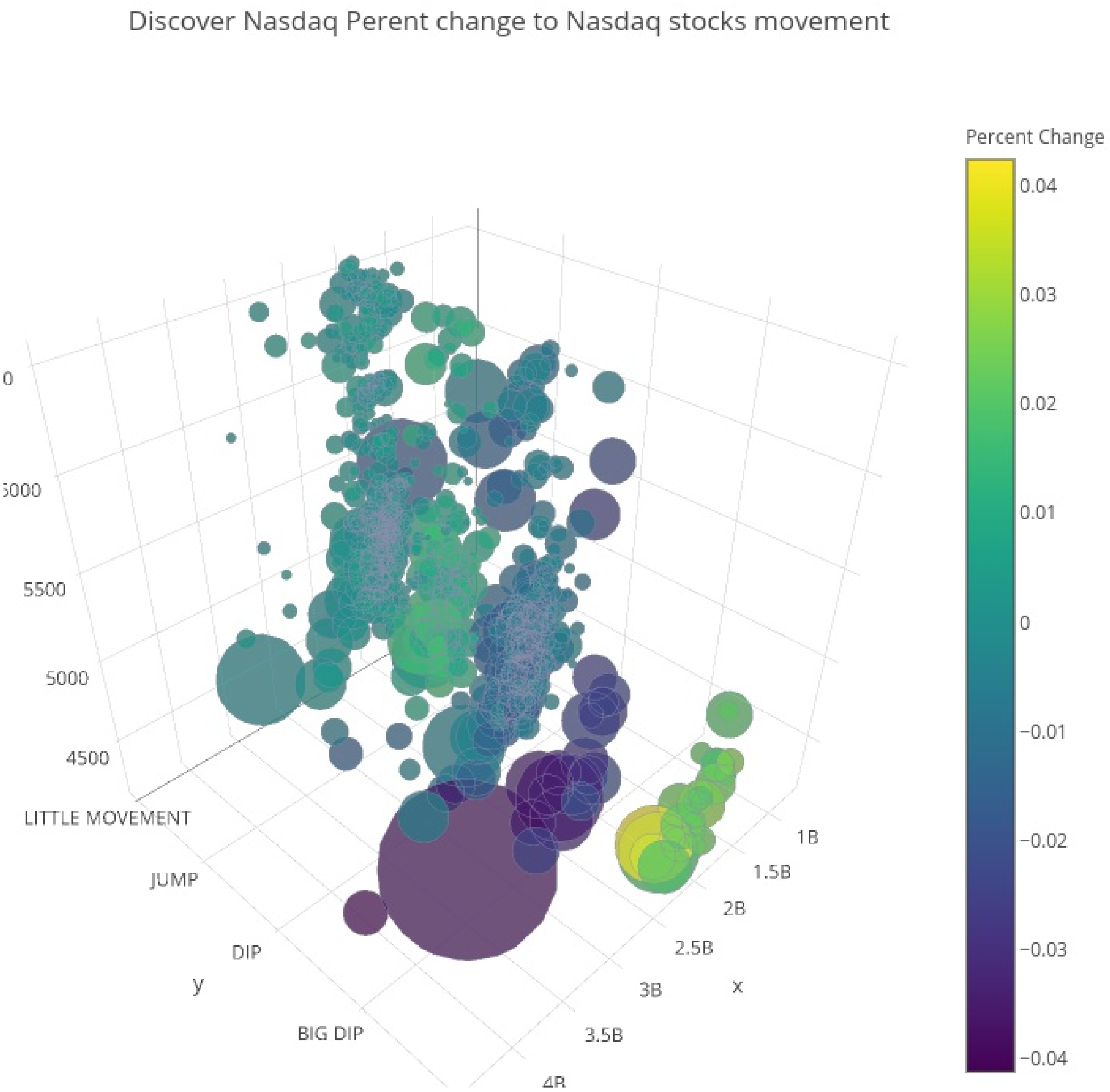
Name

Enter your comment here

Comment by Html Comment Box

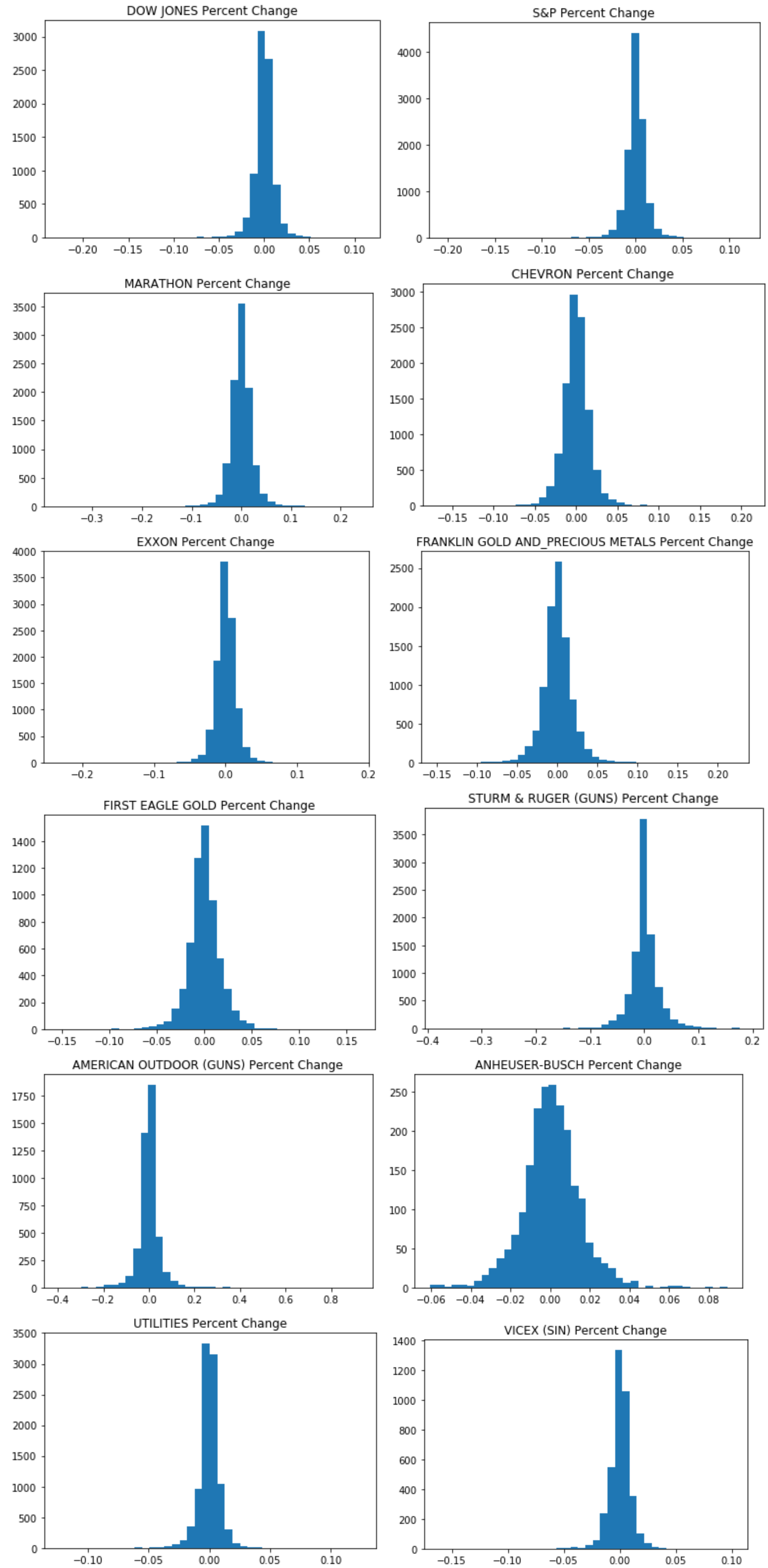
No one has commented yet. Be the first!

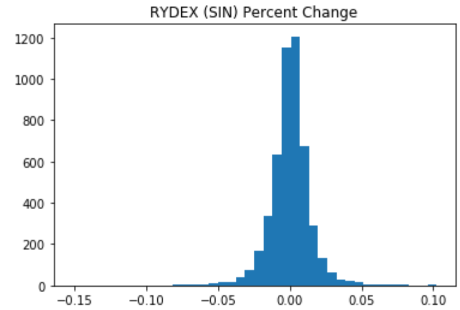
Appendix A - Depiction of Stock Variables

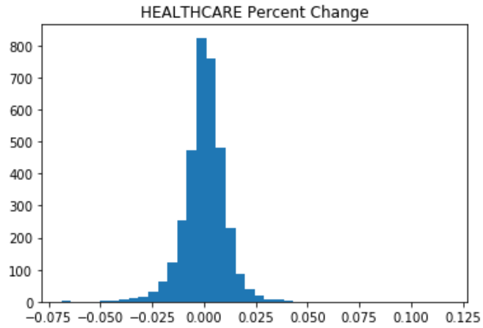


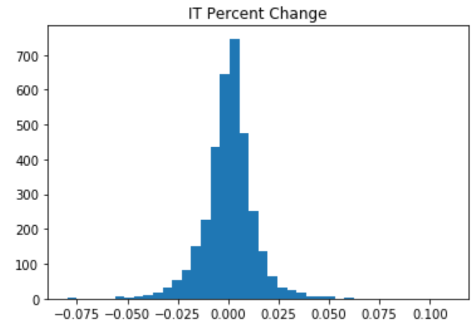
This is a 3D bubble chart. The chart refers to the NASDAQ Index and its various movements, with the percentage change variable represented by color, bubble size by the stock's daily high minus daily low, volume on the x-axis, change category on the y-axis, and closing price on the z-axis. Nothing especially stands out, but there is a lot of information to explore.

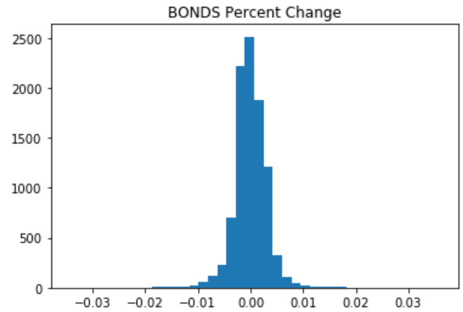
Appendix B - Histograms of Daily Percentage Changes in Tested Stocks


DOW JONES Percent Change
S&P Percent Change
MARATHON Percent Change
CHEVRON Percent Change
EXXON Percent Change
FRANKLIN GOLD AND_PRECIOUS METALS Percent Change
FIRST EAGLE GOLD Percent Change
STURM & RUGER (GUNS) Percent Change
AMERICAN OUTDOOR (GUNS) Percent Change
ANHEUSER-BUSCH Percent Change
UTILITIES Percent Change
VICEX (SIN) Percent Change

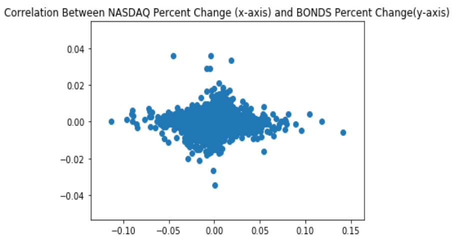


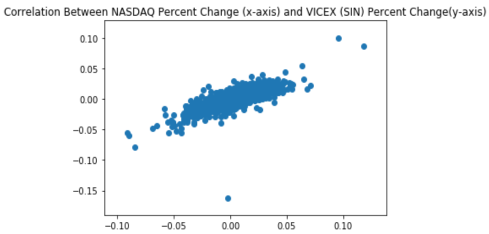


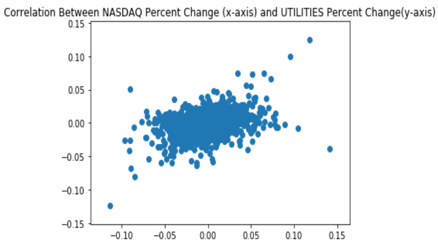


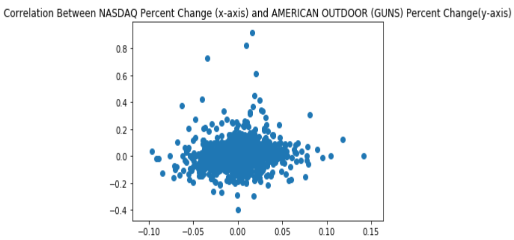


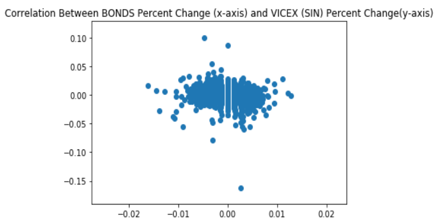
Appendix C - Scatter Plots of Correlation Between Stocks

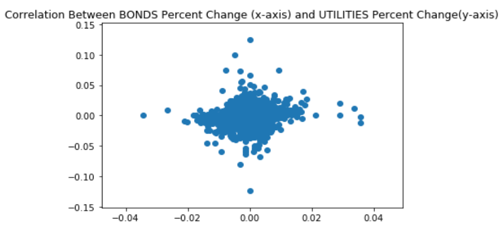


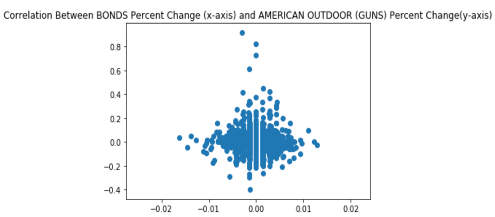


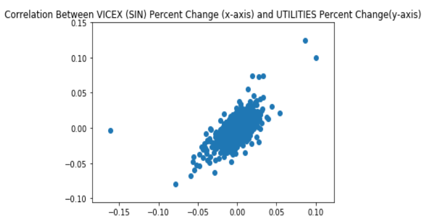


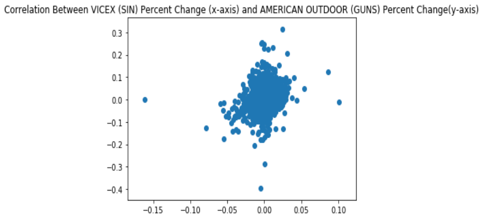


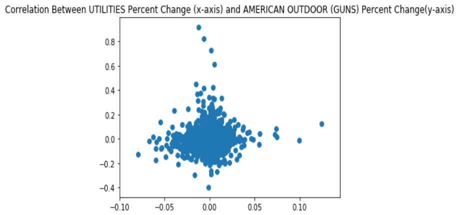








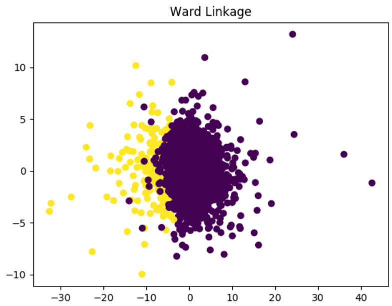




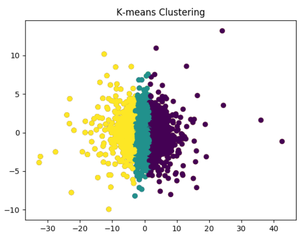
Appendix D - Clustering Analysis of Stocks

As mentioned, three clustering analyses were performed on the stocks data. Using these clustering methods, we looked for interesting clusters in the percentage change columns for the seventeen different stocks. Since some stocks don't have data for the percentage change columns for certain days (because those stocks did not yet exist), the null values were replaced with the mean of non-null percentage change values for each respective stock.

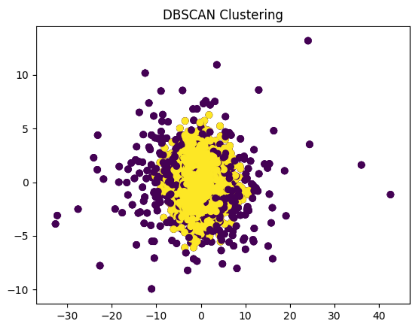
Using hierarchical clustering with Ward linkage, and plotting a PCA projection, there appear to be 2 clusters - as shown below - for the percentage change data. This also gives the best Silhouette average score of 0.465.



Next, k-means clustering was performed, which produced the best Silhouette average score with 3 clusters: - 0.2823. This very poor score indicates that k-means is not a good method for clustering the data. One possible explanation is that the points were too close together to create meaningful clusters, while another is that the sizes and shapes of the real clusters differ too much to produce any interesting results.



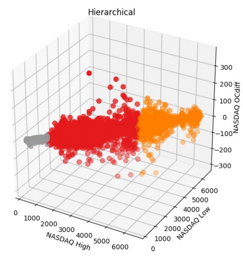
Finally, DBSCAN clustering was conducted, which works better with non-bunch-shaped clusters. With an eps of 3.5, DBSCAN produced a Silhouette average score of 0.5571 and the plot below.

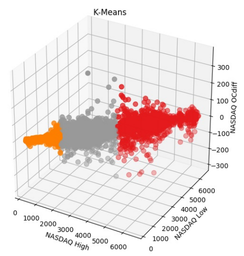


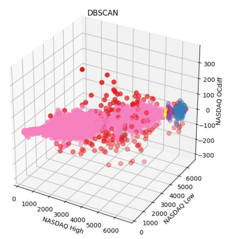
So, considering the work above, the seventeen percentage change columns seem to be best grouped by two clusters. Digging a bit deeper into the plots, it looks like the data groups into clusters where there was little change in the market and days when there were either dips or jumps. Although, as the fairly poor Silhouette scores suggest, these clustering results are not strong enough to spur any new hypotheses.

Additionally, though, one of the individual stocks--the NASDAQ Index--was looked at in more detail to see if its analysis would mirror that in the previous categories. That is, we wanted to see if we could spot any groupings in the performance of an individual stock that wasn't apparent in the plots above. The original NASDAQ data contains: "NASDAQ Open", "NASDAQ Close", "NASDAQ High", "NASDAQ Low", and "NASDAQ Volume". This high dimensionality could reduce the accuracy of the clustering, so the number of variables used was cut down: In particular, a stock's volume is much larger than the other variables we tested, and so it was dropped from the analysis, along with the stock's open and close, because these can be captured in the same variable. And so, along with the dropping of variables, that new variable was also created: "NASDAQ OCdiff", which is the difference between the opening price and closing price on a given day. In sum, then, the "NASDAQ Low", "NASDAQ High" and "NASDAQ OCdiff" were the three final attributes used for this individual stock clustering.

The same methods outlined above in the percentage change clustering were conducted with several values of parameters for fitting the models for the NASDAQ Index. For k-means clustering, k = 3, or a predetermined number of clusters of three, was used. For DBSCAN, eps = 100 and min_samples = 30 were set for the best fit. And finally, for hierarchical clustering with ward linkage, n=3 was the chosen parameter. Through each method, the clustering results were generated in 3D for better visualization, as shown below.







K-means with three clusters and Hierarchical clustering produced fairly nice accuracy scores of .6693 and .5899, respectively; DBSCAN only produced a score of 0.41175. Since this tracked one broad index stock, it may be the case that it was best suited by three clusters: small changes in the market, moderate changes, and large changes. The graphs above appear to agree with this idea. Another thing to note is that, unlike the clustering performed on the percent change columns, k-means clustering performed best here. One possible explanation is that there is less shape or density variation in a single stock or index, such as the NASDAQ, compared to the other stocks tested.

It was interesting to see the groupings created by these clustering methods and the way their results differed between one stock and all seventeen stocks. Unfortunately, though - and as is noted in the paper - none of these results were pertinent to our work. That is, none of them generated any new hypotheses for testing.

Appendix E - The Specifics of the Apriori Algorithm Association Rule Mining

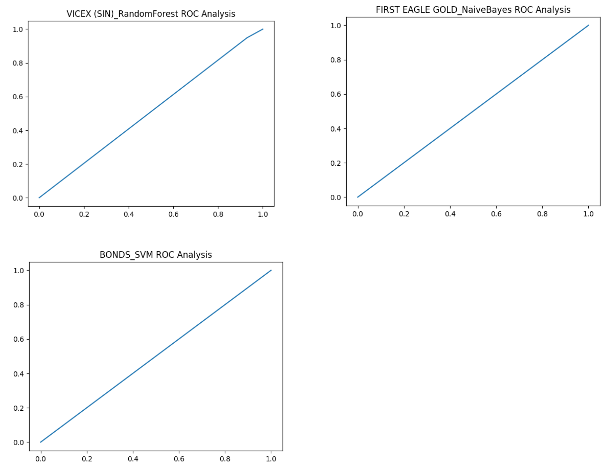
For each team, there were two possible values for every data entry: 1 (the team won) or -1 (lost)--days when the teams didn't play were not considered. And, for each stock, there were five possible values for each data entry: BIG JUMP, JUMP, LITTLE MOVEMENT, DIP, and BIG DIP. To be more specific, the Apriori algorithm was run on three datasets--one for each team. So, for example, the New York Yankees dataset consisted of every day on which the Yankees played a game between 1975 and 2017 and the way the seven test stocks reacted the following day. This means there were 2 * 7 * 5 = 70 possible association rules for each run through the Apriori algorithm (70 * 3 = 210 total), none of which produced a particularly-high support or confidence value. The results were so poor, in fact, that the use of very low minimum support values was necessary to complete the tests. Setting the minimum support to .2 trimmed the 210 potential association rules down to 40; .25 to 15 rules; and .28 to 2 rules. A summary of these results follows in the table below, as seen in the rule mining section of the paper (all rules with support > .25 are listed).

Event A	**Event B**	**Support**	**Confidence**
Dallas Cowboys WIN	**Exxon LITTLE MOVEMENT**	**.2762**	**.6210**
Dallas Cowboys WIN	**Sturm & Ruger (Guns) LITTLE MOVEMENT**	**.2754**	**.6151**

Dallas Cowboys WIN	Anheuser-Busch LITTLE MOVEMENT	.2593	.4795
Dallas Cowboys WIN	VICEX (SIN) LITTLE MOVEMENT	.2720	.5812
Dallas Cowboys WIN	BONDS LITTLE MOVEMENT	.2609	.4751
New York Yankees WIN	Exxon LITTLE MOVEMENT	.2578	.4570
New York Yankees WIN	Franklin Gold and Precious Metals LITTLE MOVEMENT	.2559	.4563
New York Yankees WIN	Sturm & Ruger (Guns) LITTLE MOVEMENT	.2511	.5662
New York Yankees WIN	IT LITTLE MOVEMENT	.2549	.4469
LA Lakers WIN	NASDAQ LITTLE MOVEMENT	.2606	.6006
LA Lakers WIN	Exxon LITTLE MOVEMENT	.2647	.4338
LA Lakers WIN	Franklin Gold and Precious Metals LITTLE MOVEMENT	.2799	.4525
LA Lakers WIN	Sturm & Ruger (Guns) Little Movement	.2974	.6383
LA Lakers WIN	Utilities LITTLE MOVEMENT	.2759	.6150
LA Lakers WIN	BONDS LITTLE MOVEMENT	.2819	.45582

To put this in context, the most promising association rule was between the Lakers winning and the Bonds Index seeing little movement, with a support value of .2819. Relative to the way things were set up, this means that, on ~28 percent of the days that the Lakers played a game between 1972 and 2017, both the Lakers won and the Bond Index experienced little movement. This, essentially, is meaningless: The Lakers had the highest winning percentage of all the teams tested, and the Bond Index - as mentioned above with its mean percentage change of 0 - was a very consistent stock, so it's not surprising that the highest support is between the winningest team and a highly-consistent stock. Considering these two things - that the highest support was both relatively small and very much uninteresting - the association rule mining produced little about which to be optimistic. By default, though, it did produce testable hypotheses, which we indeed did test. These tests, unlike the rule mining, *did* produce interesting results.

Appendix F - ROC Curve Examples



The almost perfectly straight lines from the lower left to the upper right corners of the graphs indicate that our models did no better than chance.

Appendix G - Further Testing of Hypothesis 2 (KNN and Decision Tree)

For these tests, the Lakers W/L data and the binned stock percentage change categories were used. The goal was to see if, based on the outcome of the Lakers' games, the classifiers could predict whether a stock would fall into the category: BIG JUMP, JUMP, LITTLE MOVEMENT, DIP, or BIG DIP.

For both the decision tree and k-nearest neighbors classifiers, cross validation was used to train and then test the model. These types of testing seemed well-suited for this hypothesis; both decision trees and k-nearest neighbors analysis work well when there are more observations than variables, which was certainly the case here. k = 3 was set for the k-nearest neighbors classifier; this means that the model will look for the three "most similar" data points for each point in the test set and take their average to predict the unknown point's category. For the decision tree, the model uses the training data to search for patterns and then implement them in a model that filters the test data points into predicted categories; in this case, it will be a small tree--it'll have a root node that asks whether the Lakers won or lost that then splits into the six possible stock change categories. Here are the subsequent accuracy scores from the tests, again sorted by the stock being tracked against the Lakers' performance:

Stock Tracked	KNN Accuracy Score	Decision Tree Accuracy Score
Bonds Index	.2284	.3847
NASDAQ Index	.2499	.4407
Chevron Oil	.4468	.2847
Franklin Gold and Precious Metals	.2569	.3789
Sturm & Ruger (Guns)	.2892	.3824

Some of these results look promising, but further exploration shows they aren't. In each case, the model just predicted the change category in the stock that most often occurred in the training data. This is evidenced by one of the decision tree confusion matrices, that for Sturm & Ruger:

[LITTLE MOVEMENT]

0	**0**	**156**	**0**	**0**	**0**
0	**0**	**805**	**0**	**0**	**0**
0	**0**	**1226**	**0**	**0**	**0**
0	**0**	**393**	**0**	**0**	**0**
0	**0**	**129**	**0**	**0**	**0**

The models just predicted that the stock would remain about the same - regardless of what the Lakers did - because it most often stayed consistent in the training data. And, when the accuracy score (.3824) is tested against the number of days that Sturm & Ruger (Guns) actually did see little movement over the tested timespan, it can be found that the percentages are identical. So the best outcomes were really just those that corresponded to prolonged trends in one stock in the market, and thus had nothing to do with the Lakers. These tests, therefore, didn't ultimately provide any insight into the data.