DATA PROGRAMMING IN R, PROJECT REPORT

Group 8: [redacted], [redacted], Andy Petersen, [redacted], [redacted]

November 5, 2018

PROJECT SUMMARY

We chose to investigate the theory known as the Super Bowl Indicator. This is a phenomenon discovered in the 1970's by Leonard Koppett. This indicator states that if a team from the AFC wins, then a bear market (down market) will follow. However, if a team from the NFC wins, a bull market (up market) will follow. Seemingly, any correlation between the two should be purely coincidental. Our team chose to investigate the validity of this phenomenon. We used 50 years of data from Super Bowl outcomes and 50 years of S&P 500 accumulation data. As a disclaimer, this should not be taken as investment advice, just an investigation of an interesting correlation.

We decided to use the S&P 500 Index as a proxy for market returns. We accessed this data via a CSV file download from Yahoo Finance and pulled 50 years of historical accumulation information. We obtained historical Super Bowl winner/loser data by scraping data from a Wikipedia table.

GETTING AND CLEANING DATA

To obtain relevant data pertaining to the History of the Superbowl, Wikipedia was used. The group decided to use Wikipedia because it was one of the few sources that had a table of the winners and losers that also denoted the conference of each team. The Wikipedia page was scraped into a list in r, and then the list was trimmed to include only

the desired table. From there, the table was turned into a data frame that included the date of the Superbowl for each year, the winner, the loser, and the winning conference.

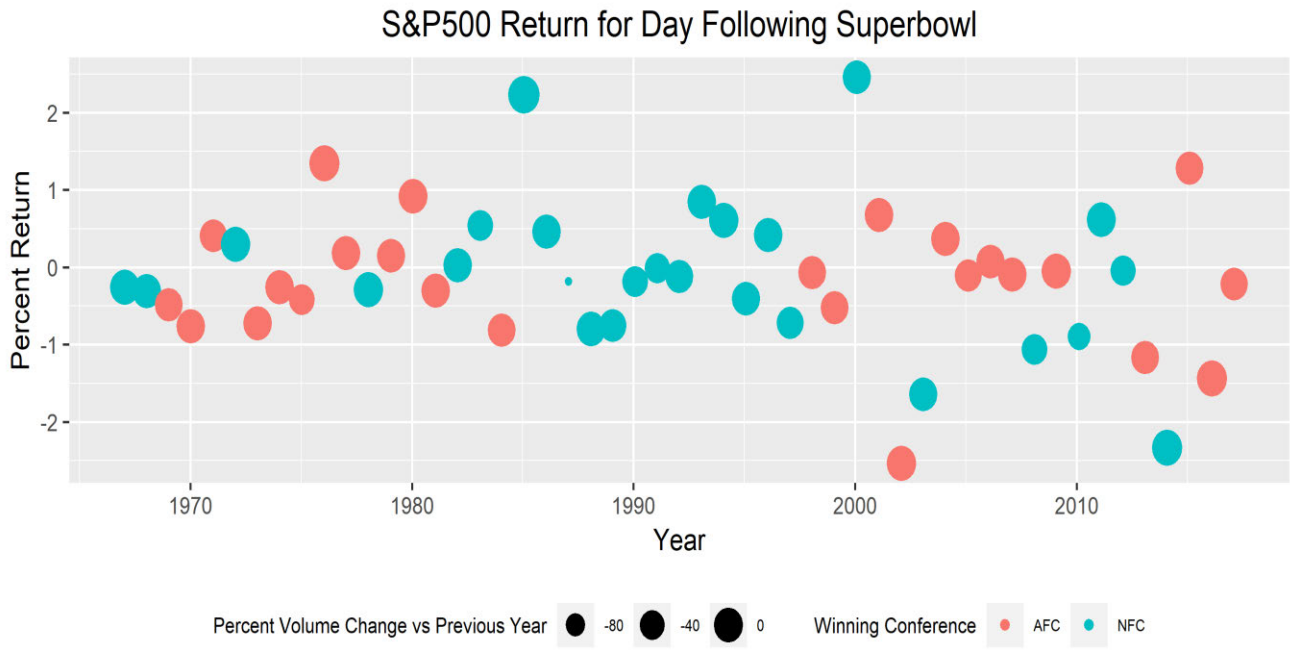
The data containing S&P500 information was downloaded as a csv file from Yahoo Finance. This data contained information about the opening value, closing value, volume, as well as other information from each trading day since the first Superbowl. This csv was read into r with the read.csv function and turned into a data frame that could be used for analysis.

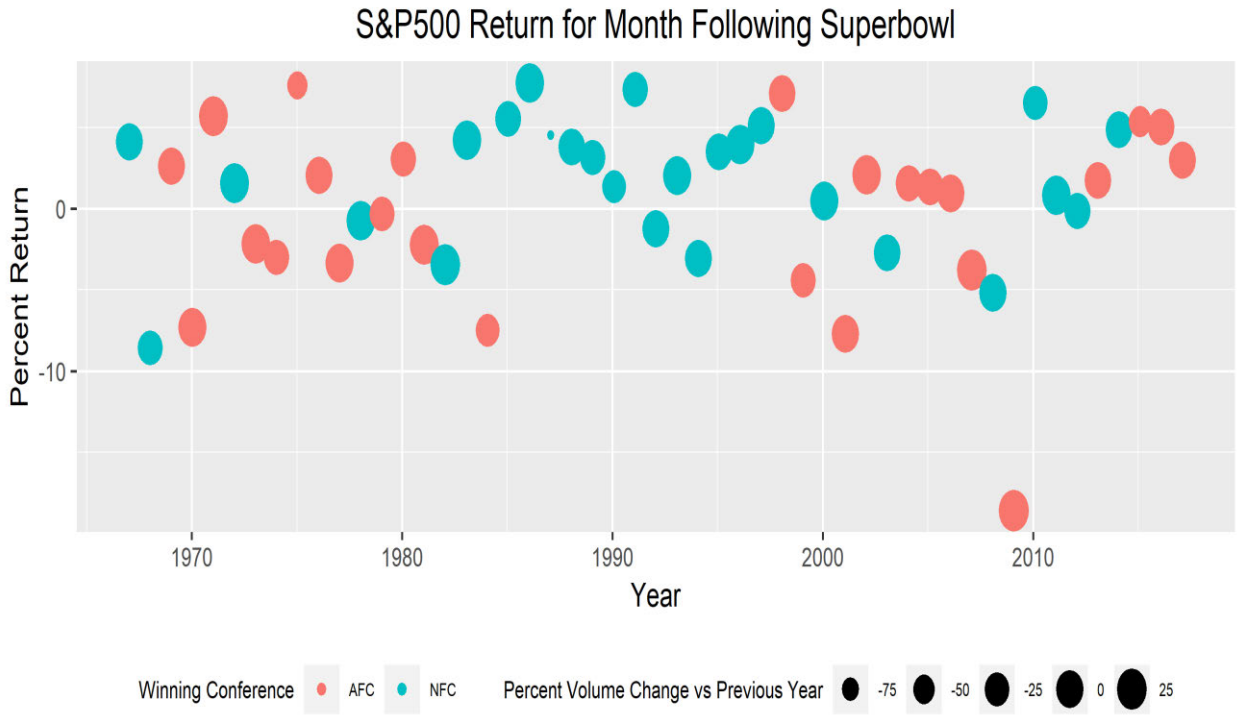
FUNCTION

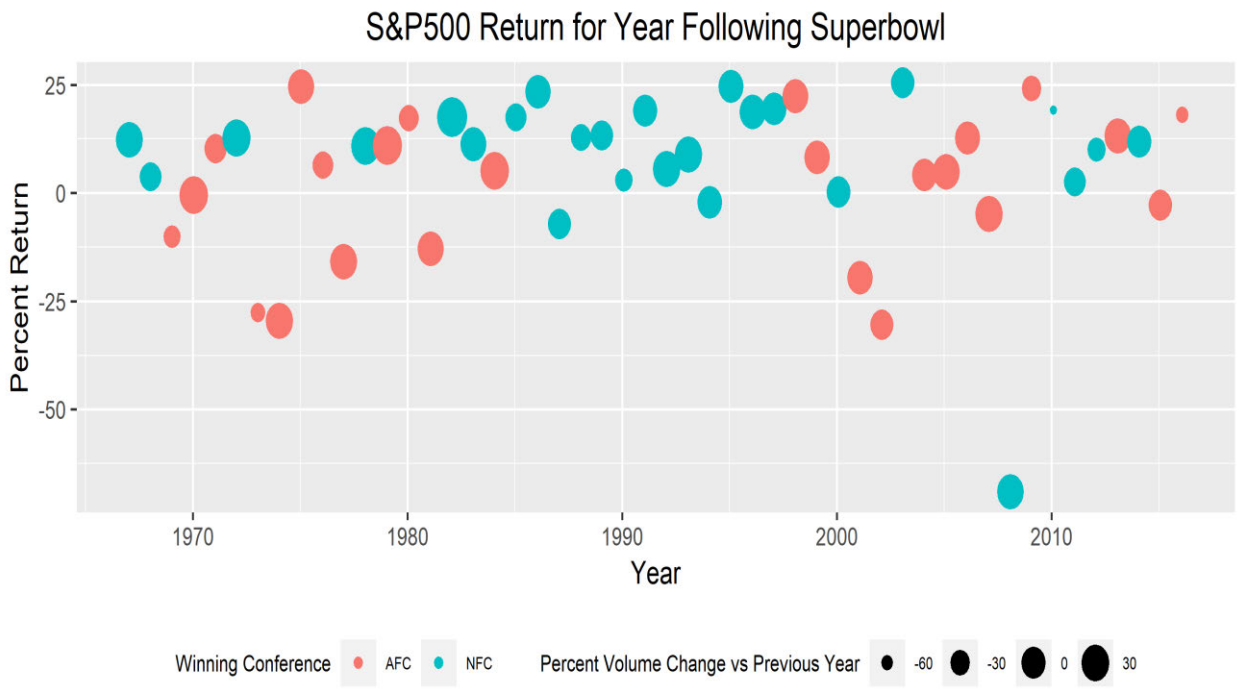
We created a function to combine the two data files together and then output market return results for one day, 30 days and 365 days following the Super Bowl. To make the function work we had to tell R to look at the previous day the market was open and the following day the market was open. This function also had to work around holidays on which the stock market was closed. Initially, we wrote a function to verify that all Super Bowl games in the last 50 years had been held on a Sunday and that the following day was a weekday on which the market was open. We then used this data to calculate market returns for one day following the Super Bowl, one month following the Super Bowl and one year following the Super Bowl. Then we used this table to calculate the percentage of times this phenomenon predicted the correct output and to calculate the value on an NFC win and AFC win in S&P 500 returns. We also created some interesting graphs during the function writing process.

GRAPHS

The following three graphs display the S&P 500 returns for one day, one month and one year following the Super Bowl. Color was used to distinguish the two football conferences, and size was used to represent the volume of transactions that took place on that day. Volume was investigated to see if there could possibly be a link between trading volume and winning conference, but we did not find any meaningful relationship between the two.




S&P500 Return for Month Following Superbowl
Percent Return
0
-10
1970
1980
1990
2000
2010
Year
Winning Conference
AFC
NFC
Percent Volume Change vs Previous Year
-75
-50
-25
0
25


S&P500 Return for Year Following Superbowl
Percent Return
25
0
-25
-50
1970
1980
1990
2000
2010
Year
Winning Conference
AFC
NFC
Percent Volume Change vs Previous Year
-60
-30
0
30

Our final graph is what we are most proud of, however, it is not displayed below because it is best viewed outside of Microsoft Word. We were able to use plotly to create an interactive, color-coded look at winner and returns across the entire 50-year dataset. Please view and explore it in R because you can see more detail by hovering over individual data points.

ISSUES

The first issue we encountered as a team was when we downloaded the CSV file and saved it to a share drive, causing each team member to then be required to download the CSV file individually. This method worked fine for all PC users, however, the Mac user in the group had trouble because his version of Excel changed some file formatting. The Mac user found a way around this issue, and we quickly learned that ideally, you should not have multiple people downloading a file. In retrospect, we would have also used the zoo package to access the Yahoo Finance data rather than downloading it to a CSV file.

Another issue we faced was loading the data into R and encountering special characters in the Wikipedia tables. To resolve this issue, we wrote code that only looked at characters that we could indicate in R and then categorized them based on whether a certain character existed.

An additional issue we encountered was that the modern-day NFL was originally called the NFC, so we had to convert “NFC” to “NFL” to maintain consistency. Likewise, the old AFL largely turned into the AFC, so Superbowl winners from the AFL were counted as AFC teams. Another formatting issue was correctly reading in dates from the S&P 500 CSV file.

To fix this, we added code to indicate the original date format before it was overwritten in R to standard R date formatting.

Regarding the creation of graphs, we ran into some issues in which all 50 years of dates were originally showing on the x-axis. We realized that we could change the frequency of the years displayed on the x-axis in order to improve the readability of the graph. We also converted our plots using plotly to allow interactive use of the graph and the ability to zoom in and view individual graph points.

CONCLUSION

Our analysis showed that over the last 50 years, the Super Bowl Indicator has correctly predicted the direction of market returns approximately 80% of the time. In terms of S&P 500 performance, we were also able to calculate that an NFC win produces an average return of 8.68% for the following one-year time period and that an AFC win corresponds to an average 1.18% one-year market gain. We did come closer to understanding Leonard Koppett's decades-long theory. Our explanatory thoughts were that over time, the market generally goes up, and given that the NFC has historically been the stronger football conference, it should correspond to higher market returns than the AFC.