NFL Teams and Political Parties

I.   Introduction

This year the National Football League (NFL) celebrates its one hundredth anniversary. It all began in 1892 when William Heffelfinger was paid $500 to play in a game for Allegheny Athletic Association. The Association then went on to win the game against the Pittsburgh Athletic Club. The PAC thought something illegal was set in motion, but was unable to prove it until 80 years later. It was the Pro Football Hall of Fame that received and displayed an accounting sheet that later became known as "Pro Football's Birth Certificate" (NFL 2005). From then on the game of pro football faced many problems. On September 17, 1920 in Canton, Ohio a group of men gathered in Ralph Hay's Hupmobile showroom. In order to minimize the issues it was facing, eleven franchises were formed. It was that meeting that gave birth to the National Football League (NFL 2005).

Football has been able to capture the hearts of people all around the globe. It is one of the most watched sports in America. In 2016, Superbowl 50 dominated television's in the United States with 111.9 million views. Following that with 84 million views was the first presidential debate of 2016 (Stelter, 2016). This staggering difference begs the question, is it possible to predict the presidential outcome based on it's NFL fanbase?

After the American Revolution, the founding fathers created the United States Constitution. This created a set of laws that citizens of the States needed to follow. In article 2, section 1 the Constitution establishes the Executive Branch of the United States government (Gov, 2018). Throughout the years Amendments were made to the Constitution that helped to expand the rights of all citizens over the age of eighteen (History.com Editors, 2010). This allowed women and people of color to vote in for a president in an election year.

Back then, elections were simpler and less controversial than they are today. In 2016 the presidential election was a mess. The polling data showed Hilary Clinton leading the popular vote consistently. Highly trained academics and data journalists had predicted Clinton's probability of winning to be around 90 percent. Kennedy et al. (2018) studied the 2016 presidential election and it came as a shock to pollsters and political analysis that Trump had won (pg. 1). So, what underlying factors influence who people vote for? Is it economic status? Household income? Or could it be the football team we root for?

The goal of this paper is to determine what factors can influence the way a person votes. To test this out, this paper will look at a NFL fandom survey that shows what political party fan's belong to. It will delve into economic factors and social factors such as hate crimes, household income, and unemployment rate. This paper will also explain the methods in which the data was collected and cleaned, as well as an analysis of the results.

II.   Review of Literature

Many scholars have studied social and economic factors that can influence one's political view. However, many have not looked into how or if a fan's favorite football team has any impact. The goal for this project was to see if there is any correlation between the football team fans and the political party they are associated to. As stated much of the literature for this topic has been done with social and economic factors. They take into account one's employment history, criminal record, and socioeconomic status.

For the 2016 election some scholars looked into why the polls underestimated Trumps support. Kennedy and her team (2018) had discovered much of it had to do with incorrect likely voter models, last minute deciding, and non-response bias (pg. 4). In spite of that information, many in the political science sector are looking to improve models for predicating presidential elections. Some will look into these factor while others seek their answer somewhere else. With recent breakthroughs in machine learning and artificial intelligence, they look to data scientist and computer programs to help them find a better model. One group is looking to see if Twitter can some how predict elections results in European Union elections (Tsakalids et. al, 2015). Although EU elections are slightly different, the methodology that Tsakalids and her/his group is using could be applied for Presidential elections in the United States.

III.   Visualizations and Method

For this project there were three untidy data sets that were looked at and cleaned. The first data set is a survey that shows how many people are one political party and what NFL team they root for. To clean the data, the first row of variable names was removed to make it easier to tidy the data. The second row then kept variables such as the team they root for, the number of people who are one political party based on race, and the political party they support. It was then melted down into Team, Race, Count and Party. This would then be used to create visualizations. For the data set, a two different stacked bar charts were created.
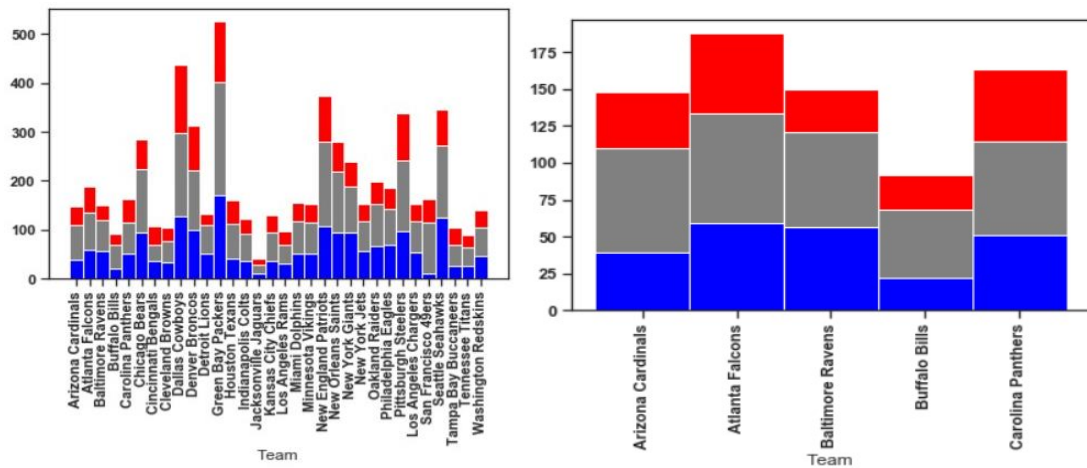
The diagram on the right shows the break down of political parties based on the Team that the fans root for. The blue represents the democratic party, the grey is the independent party, and red is the republicans. This way of visualizing the data was chosen because it allowed us to see the total amount of people that were counted, as well as the break down of what party they voted for.

The next data set to get cleaned was the 2016 elector votes data. This dataset was melted down to get two primary bits of information. The name of the state and the winning party of that state. These two columns will then get merged with the third data set to reduce the amount of data sets. To get a better visualization of the 2016 election results, this map was used (Healy, 2012). The map has no correlation with the NFL survey data. Its sole purpose was to give a better illustration of the winning party of each state. Looking at this map, it is easier to see that the republicans held the majority of the votes. When compared to the bar charts created, some would assume a state or two should have had a different winning party.
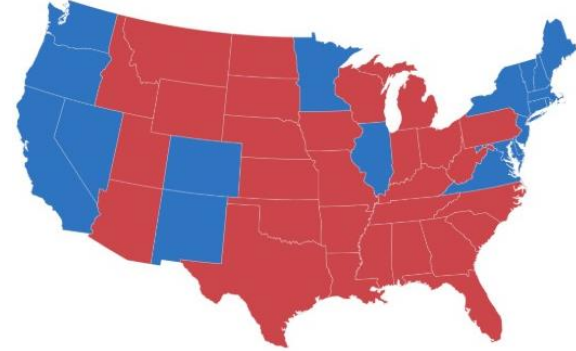
The last visualization created was based off the last data set called hate crimes. This data set was used to look at other factors that would influence the way people vote. The factors that would be tested later were median household income, seasonal unemployment, and average hate crimes committed. The visualization created from this data set was based on the median household income. The goal of this was just to get an idea if income played a role in who people voted for and if there was any relation to the NFL survey data set. Questions such as "are people more likely to answer the NFL survey because their average income is better" or "are people with higher median income more likely to be democratic or republican" could be answered just by looking at the visualizations. However, statistical analysis should still be performed in order to get a correct assessment.
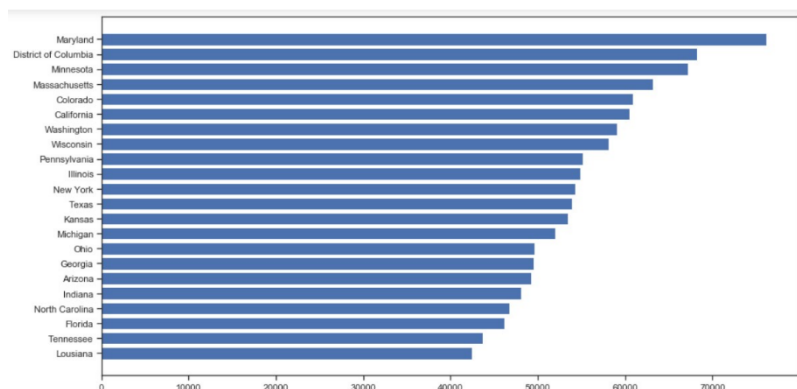
IV.   Analysis

Initially, logistic regression was an option to run on the NFL survey data set. However, there were not enough variables to determine a regression plot or perform an analysis. In the end regression analysis was performed on the combined hate crime and winning party dataset. Regression analysis allows us to examine the relationship between two or more variables. In this case, median income, seasonal unemployment, average hate crimes committed, and vote percentage for Trump were chosen to be analyzed. By looking at these points, one can determine if there is any correlation between these points. The independent value that was selected was the vote percentage for Trump. The dependent variables that were selected include: median income, seasonal unemployment, and hate crimes. These variables were selected because there was some known correlation between them and the percentage of that state that voted for Trump. Since there were outliers on the regression plots, leverage of those plots was calculated.

V.   Results

a   b   c
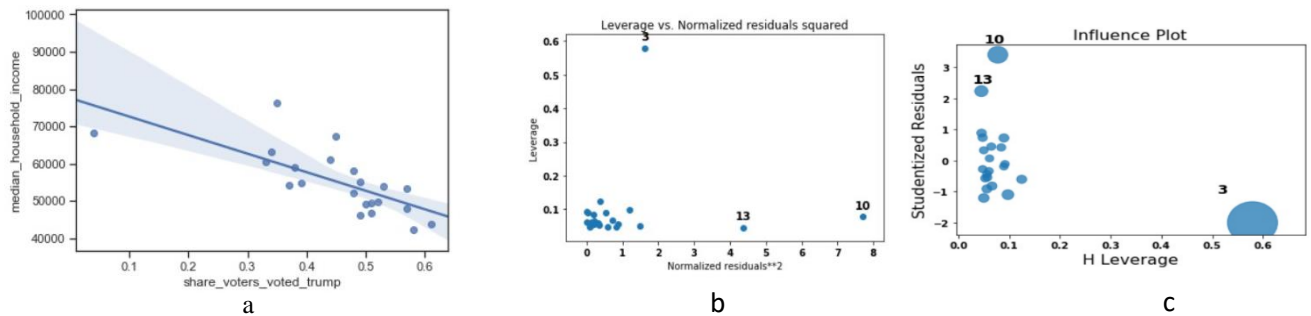
Graph (a) shows a regression plot for median household income and people who have voted for Trump. For this plot the null hypothesis is that median household income plays a role in how many people voted for Trump. After doing regression analysis, the null hypothesis can be accepted since the p-value is significantly greater than 0.05. Meaning that there is a corelation between income and the percentage of people who voted for Trump. The outlier of this graph does show significant influence based on graph (b) and (c). Looking the Cook's distance for that point was 2.41, which means it had some influence on te graph. When the outlier is removed, the graph holds a steeper slope. The R^2 squared value is a measures the correlation strength of the linear model. Without the outlier, the correlationis neither or weak.

| Slope	-5.0601
Intercept	0.7105
R-value	0.518
P-value	0.14311
Stderr	3.3199
Decision	Accept

d

e
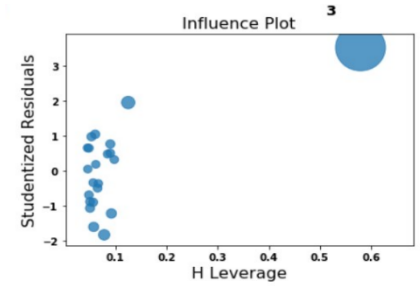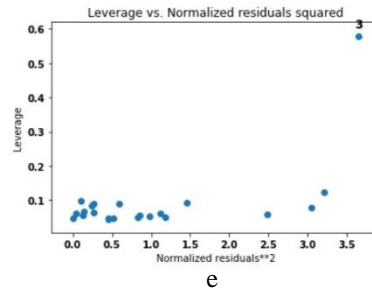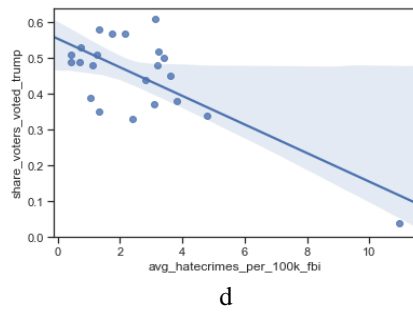
f

The next set of results to be discussed is the impact of hate crimes. The regression plot (d) shows all the points on the graph before the outlier is removed. The null hypothesis in this case gets rejected, informing us that hate crimes were not an influence on peoples decision to vote for Trump. Based on Cook's distance for the point all the way on the bottom right, this point had a significant influence on the graph. After removind that outlier, the slope decreased and had an almost horizontal slope. With a distance of 5.44, the regression plot changed greatly. Even looking at plots (e) and (f), one can see the influence that plot had on graph (d).

Slope	-13.3989
Intercept	0.5545
R-value	0.536
P-value	0.000109
Stderr	2.79
Decision	Reject

g

h

i

Slope	-1.047
Intercept	0.0605
R-value	0.104
P-value	0.143
Stderr	0.013
Decision	Accept
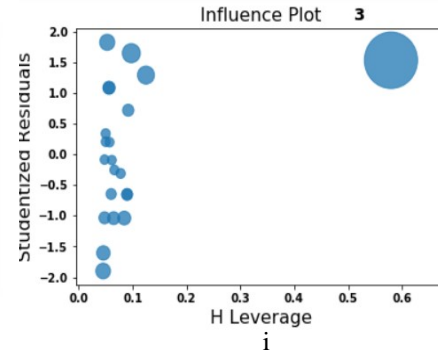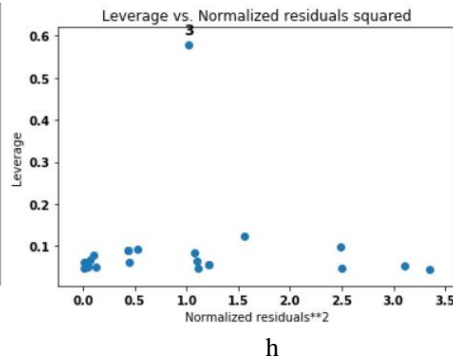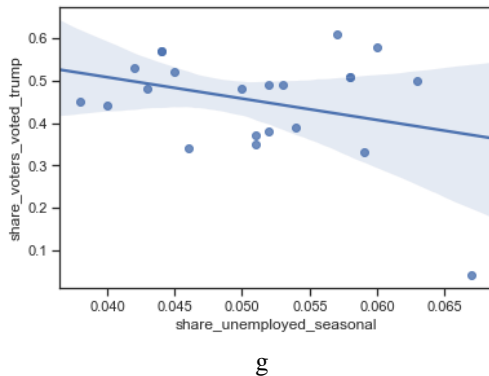
The final set of results checks the relationship between seasonal unemployment and the percentage of people that voted for Trump. The null hypothesis for this experiment was to see if there was a correlation between the two. Since the p-value was greater than .05, the null hypothesis gets accepted. The next thing that got looked at was the outlier. With a Cook's distance of 1.51, the outlier does not influence the plot greatly. When removed, the trendline barely shifted. Based on the R-value, the correlation is relatively weak.

VI.   Discussion

After looking at the results, there are some implications that can be identified. Factors that can have a correlation of who we vote for are seasonal unemployment, hate crimes, and household median income. This impacts how politicians run their campaign and who they target. Looking at Trump's campaign speech, he speaks directly to people who are in these three categories. In a speech he made in Wisconsin he talked about a plan to bring back jobs

and how law and order must be restored (GOLDMACHER, 2016). By identifying flaws or weaknesses in each state candidates can help swing the vote their favor.

Although it is a stretch to say a person's NFL team has a correlation to the political party they choose, there is some influence the NFL can have. Looking at the TV ratings, political advertisements can be placed strategically during commercial breaks. Highly anticipated games such as playoff or wildcard games can impact a surplus of voters, especially those in the younger generation who have not identified with a party yet.

Another implication that must be considered is the ethical implication of political parties gathering this data. By analyzing socioeconomic data political parties can gain so much insight on how to strategize the campaign. However, how they gather this data can be an issue. Not everybody understands that their data is public and can be used for a political advantage. Even if it is public data, people should be informed what their data is being used for. Although it is an average of the whole state, income isn't something people like to share. The best way to overcome this barrier is to inform people of what data is being taken and how it is being used.

VII.   Conclusion

In conclusion, this project was able to answer the purpose of this project. Although statistical analysis was unable to help determine a correlation between a fan's choice in NFL team and the political party they are affiliated to, visualizations played a role greatly in this determination. After comparing the election results to the NFL survey data set, it was determined that there was no correlation between the two. Other factors that influenced what party a person might vote for were unemployment and household income. For future analysis more data should be collected in order to perform better analysis. One example that comes to mind would be looking at other sports.

References

GOLDMACHER, S. (2016, August 17). Full text: Donald Trump campaign speech in Wisconsin. Retrieved December 12, 2019, from POLITICO website: https://www.politico.com/story/2016/08/full-text-donald-trumps-speech-on-227095

Gov. (2018, September 24). The Constitution of the United States: A Transcription. Retrieved from National Archives website: https://www.archives.gov/founding-docs/constitution-transcript

Healy, K. (2012). Data Visualization. Retrieved December 12, 2019, from Socviz.co website: https://socviz.co/maps.html

History.com Editors. (2010, January 8). Presidential Elections. Retrieved from HISTORY website: https://www.history.com/topics/us-presidents/presidential-elections-1

NFL. (2017). NFL founded in Canton | Pro Football Hall of Fame Official Site. Retrieved from Profootballhof.com website: https://www.profootballhof.com/news/nfl-founded-in-canton/

Porter, R. (2016, December 27). The 100 most-watched TV programs of 2016: Super Bowl 50 leads by a mile. Retrieved from TV By The Numbers by zap2it.com website: https://tvbythenumbers.zap2it.com/more-tv-news/the-100-most-watched-tv-programs-of-2016-super-bowl-50-leads-by-a-mile/

Stelter, B. (2016, September 27). Debate breaks record as most-watched in U.S. history. Retrieved from CNNMoney website: https://money.cnn.com/2016/09/27/media/debate-ratings-record-viewership/

Tsakalidis, A., Papadopoulos, S., Cristea, A. C., & Kompatsiaris, Y. (2015). Citation for published item. *Predicting the EU 2014 Election Results in Multiple Countries Using Twitter*. https://doi.org/10.1109/MIS.2015.17