UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FOURTH DOWN'S CRITICAL ROLE IN A WINNING STRATEGY FOR AMERICAN FOOTBALL GAMES

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ERIN PSAJDL
Norman, Oklahoma
2021

FOURTH DOWN’S CRITICAL ROLE IN A WINNING STRATEGY FOR AMERICAN FOOTBALL GAMES

A THESIS APPROVED FOR THE
DEPARTMENT OF HEALTH AND EXERCISE SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Daniel Larson, Chair

Dr. Ryan Spalding

Dr. Christopher Black

Table of Contents

Abstract

Data analytics is booming, and this boom has started to creep into sports – originating in baseball and moving its way across other popular sports leagues. American Football is one of the last sports to embrace the analytics revolution; however, a noticeable use of data analytics in recent years within the sport of American Football is the decision on whether to punt or "go for it" on any fourth down during a game. There have been studies on this dating back to the 1970's, yet most of these studies revolved around the discrete concept of points, as opposed to a continuous concept such as win probability. Regardless of the variable of interest, studies have shown that teams are being too conservative on their fourth downs and punting the ball to their opponent far too often. This study uses ten years of historical play-level data from the National Football League coupled with machine learning methods such as random forest algorithms and nearest neighbor matching along with a Poisson regression to evaluate team's decision-making on fourth down; the study will then evaluation the relationship between decision-making and the number of wins the team has experienced from 2010 to 2019. This study agrees with those before it that teams are not making the "optimal" win probability maximizing choice on fourth down as often as one would expect; however, the frequency of which teams are maximizing their win probability does not have a direct effect on the number of games a team actually won over the course of the seasons included in this study. This could be due to reasons discussed in the discussion session of the study; though we must remember that maximizing the probability of an event occurring does not necessarily mean that said event will occur.

Acknowledgements

Above all, I am grateful to God for the opportunity to pursue this degree and the good physical and mental health that were necessary to complete this thesis. He has been so, so good to me and I could not have completed this without His good grace.

I would like to express my sincere thank you to my advisor Dr. Larson of the Department of Health and Exercise Science at the University of Oklahoma, along with committee members Dr. Spalding and Dr. Black. All three allowed this thesis to be my own work, but offered their insight to help me accomplish a thorough study. I would also like to thank my cohort within the Sports Data Analytics degree program: Megan, Sam, and Joe. They were a great support system, and I don't think I could have picked a better group to laugh with, stress with, and learn with over the past two years in this program – especially while writing this thesis. I am extremely grateful to Steve Palazzolo of Pro Football Focus for the data that he provided to me in order for me to complete this study.

Last, but certainly not least, I have to thank from the bottom of my heart my parents, Brad and Robin; my grandparents, John and Sandy; my boyfriend, Nick; and all of my family and friends for their constant support and encouragement throughout all of my years of study and through the process of completing this thesis. This accomplishment is not mine alone, as it would not have been possible without them. Thank you.

Author

Erin Psajdl

Chapter 1: Introduction

For football fans, watching a game can be an emotional rollercoaster – especially in particularly risky or potentially game-changing scenarios. When these scenarios arise, fans often think they know exactly what the "right call" should be. These "armchair quarterbacks" have their opinions on what calls the coaches, who have significantly more knowledge and experience than the average fan at home, should be making. However, sometimes the popular opinion is not always the rational, strategic choice. Now, it is not always just the fans who misinterpret the rational choice. Although the tide is starting to turn, sometimes coaches do not make the rational choice either. So, going forward, how can coaching staffs make the rational choice with the greatest likelihood of increasing their team's win probability? In this study, we are going to take an in-depth look at the decisions NFL coaches make on fourth downs and the effect these choices have on their team's success.

The use of data analytics to gain a competitive advantage is growing in many industries – including sports. Sport performance analytics in the sports industry as we know it started in baseball and rose to fame through Michael Lewis' best-selling book turned film Moneyball, which highlighted the Oakland Athletics' use of analytics to create novel metrics to assign value to players and find those who were 'underpriced'. Mark Fichman and Michael Fichman (2012) discussed how conventional baseball wisdom had treated one statistic, batting average, as one of the most important indicators of player value. Contrary to the norms within conventional baseball, the Athletics brought in outsiders, now known as 'sabermetricians,' to transform their baseball performance analysis. These were not former players, coaches or traditional scouts – these were "numbers people" who were skilled statisticians. Once other teams started to notice this small club's success, they were intrigued, and they, too, not only adopted the strategies of the Oakland Athletics, but started exploring new metrics of their own. The result of this movement was a revolution in the game of baseball.

As the analytics crusade took off in professional baseball, the other sports leagues started to perk up at the thought of exploring analytics in their own sports (Steinberg, 2015). In baseball, a major problem was found in the discrete box score variables leaving so much of the story untold. This led others to question how much of the story is missing in other leagues. Kubatko et al (2007) discussed basketball running into the same problem with box scores and the need to measure efficiency and productivity more accurately. Some owners, such as Mark Cuban of the Dallas Mavericks, are major proponents of analytics and place a heavy emphasis on it (Peloso, 2015). Looking at shot maps from the National Basketball Association from 2000 to 2020, there are noticeable differences – such as more shots being taken either behind the three-point line or in the paint. This is because shots beyond the three-point arc are worth more, and shots in the paint, although worth only two points, have a higher probability of going in the basket – this means these shot locations bring about the greatest number of expected points. This observation is thanks to the analytics movement (Goldsberry, 2019).

Of the four major professional team sports leagues in the United States, the National Football League is proving to be the last to adopt analytics for on-field performance in the game (Clark, 2018). In the 2019 – 2020 season, the National Football League did start actively publicizing their "Next Gen Stats" in promotions throughout the season. However, the focus has been on showcasing the athleticism of the plays that have been made or showing the measures that are being taken to improve player health and safety – which are still two very important ways that analytics are being used. When it comes to individual teams using analytics for their in-game strategy or through advanced performance metrics, the National Football League has some catching up to do (*Big Hit*, 2020). A problem that football runs into lies in the fact that, relative to other sports, there are a large number of athletes involved in each play, each doing specific yet interrelated jobs. Most analytics in other sports revolve around assigning value to a player based on their contribution to the team over the course of a play, a game, or a season; this value is assigned through outcomes in discrete one-on-one events or events where it is possible to measure a specific player's contribution. This is difficult in football when it is hard to attribute a certain amount of the team's success to each player on the field. In the past few years, statistical models and advanced metrics have started to appear in the National Football League, but they are still not being utilized to their fullest potential (*Big Hit*, 2020; Clark, 2018).

The most perplexing part of all might be that the teams are underutilizing some tools that have the potential to be game-changing. A scenario that fits this description is the infamous fourth down decision. In football, the overall logic is simple: retain possession of the football, score points, and prevent the opponent from scoring points. However, on a fourth down, it is the team's last chance during a possession opportunity to advance toward scoring points, and they are faced with three choices: use that play like any other down and attempt to gain the desired distance (i.e., "go for it"), kick a field-goal if the team is in field-goal range, or punt the ball to their opponent. If the team "goes for it" on fourth down and fails to convert, the ball will be given to their opponent at that spot, but if the team punts, the ball will be given to their opponent at some position farther down the field. Up until recently, it was extremely rare for teams to go for it on fourth down unless the team was in a dire situation. Even on short distance situations, such as 4^{th} down and 1 or 2 yards to go, teams were choosing to give the ball to their opponent even when there was a good probability that they could convert. However, the league has seen steady increases in fourth down attempts over the past five seasons according to ESPN (*2019*

NFL Stat Leaders, n.d.). During the 2015 – 2019 regular seasons, there were 476, 479, 485, 539, 595 fourth down attempts, respectively; that is a 25% increase during that time span (*2019 NFL Stat Leaders*, n.d.).

While teams are starting to utilize their "extra" down more often, they, in general, are not using an optimal strategy on fourth down (Carter & Machol, 1978; Yam & Lopez, 2019). This might be due to fear of ridicule for appearing to make the wrong, "risky" decision, or due to lack of knowledge on the situation. Choosing to retain possession of the football on fourth down often seems risky to the casual observer, but from an objective view, it is often the more logical decision from a win maximization viewpoint. However, in the literature review following this chapter, risk aversion will be discussed, as it is something to consider to fully understand the decision-making process on fourth down.

This problem is important because it could help teams find a competitive advantage or at the very least maintain competitiveness. Ultimately, teams making sub-optimal decisions are not giving themselves a fair chance. In the NFL, it is often said that any team can beat any other team during any week of the season, and for this to ring true, teams need to take every advantage they can find. Oliver Stone's movie *Any Given Sunday* highlights this perfectly in a speech given by Tony D'Amato (portrayed by Al Pacino):

> You find out life's this game of inches, so is football. Because in either game – life or football – the margin for error is so small. I mean, one half a step too late or too early and you don't quite make it. One half second too slow, too fast and you don't quite catch it. The inches we need are everywhere around us. (Stone, Logan, Pine, Pacino, Diaz, Quaid, & Warner Bros., 2001)

When competing in a league that is composed of the best of the best, everyone – from the coaches, to the players, to the support staff – needs to be on their A-game. When dealing with a game of inches, it is crucial that the most efficient choices are made – teams may be making suboptimal decisions at crucial moments because of either fear or ignorance, and that could be losing games and changing outcomes of seasons.

The purpose of this quantitative, relational research study is to analyze play-by-play data from the past ten National Football League seasons to determine the relationship between the frequency of optimal decision-making – measured by maximizing win probability on fourth down – and the total yards gained per game, and furthermore, the relationship between total yards gained per game and wins per season.

Purpose of the Study

The purpose of the study is to demonstrate that through an in-game strategy with optimal decision-making on fourth down, defined as choices that maximize expected win probability, teams may be able to gain more yards over the course of a game, score more points in games, and win more games per season.

Research Hypotheses

H_0: The frequency of which any given team in the National Football League makes the optimal choice on fourth down is not correlated with the number of points scored, and therefore has no relationship with the number of wins that team earns in a season.

H_1: The frequency of which any given team in the National Football League makes the optimal choice on fourth down is positively correlated with the number of points scored, and the number of wins that team earns per season.

The positive relationship for the research hypothesis is based off of the logical structure of the sport of football. As teams move towards their opponent's endzone, they accumulate yards. One could assume that the more yards a team has over a course of game, the further distance that team has traveled towards their opponent's endzone, increasing their likelihood of scoring points. With this logic, one could infer that there is a positive relationship between yards gained per game and wins per season.

Significance of the Study

This study will not only assist the thirty-two teams comprising the National Football League in their in-game critical decision-making, but will be applicable, with minor changes, to collegiate and high school football teams, as well. This will increase overall competitiveness of the games because the win probability margins throughout the games, as well as the final scoring margins, could be smaller – which betters the fan experience at every level, as the games will be more exciting. It does not require much thought to understand that fans enjoy more exciting plays – you simply have to be in any football stadium for a third or fourth down near the endzone to hear how excited the fans are. For every stakeholder in the sport – players, coaches, fans, owners, leagues – this study could have positive outcomes. The fans could see more exciting games, as has been previously discussed. If the fans become more involved in the league, they could start funneling more money into it, through merchandise, ticket sales, and television revenue. This money will clearly benefit the players, coaches, owners, and the overall league. As the league gets more money, the players will be paid more, as their collective salary is determined as a percentage of league revenue. The coaches and owners will see more money as well, but the coaches could also see better job security. If it is shown that they are making optimal decisions, even if their team is losing, they could avoid being placed on the hot seat. This study is novel because while there is literature on expected points and win probability and some literature on trends in in-game decision-making, academic research on the relationship between the two areas and how teams have performed historically is lacking. This study will highlight one advancement that can be made in the sport of football when advanced data analytics are applied.

Delimitations

The dataset for this study will be comprised of play-by-play data from the ten most recent National Football League seasons. This is enough data to have a large sample size, but keeps the data relevant since the nature of the game is always changing.

Limitations

Expected Points, which has been used in the majority of fourth down studies to this point, is a league-wide metric which is the average expected points for all thirty-two teams. While this is useful, different teams have different personnel, which is a factor in decision-making. For this reason, win probability will be utilized, as it factors in personnel in an indirect way through the win probability model. In the win probability model discussed later in this study, team's offensive and defensive rankings are included, which serve to measure the ability of each team. This will allow the model to control for teams' strengths on offense and defense without having to adjust the model yearly for teams' ever-changing personnel. As previously mentioned, this study could be generalized to other levels of football, but the win probability model will need adjusting, specifically in the offensive and defensive ranks area. The theory as a whole, however, could be applied. One final thing to consider is that there are head coach and offensive coordinator changes occasionally, and that is not directly controlled for through the analysis piece of this study. This study will not be evaluating specific coaches, rather the fourth down decisions of coaches.

Assumptions

For the data being analyzed, the assumption that all teams will have the same goals of scoring points and winning games needs to hold true. The National Football League is unlike some other leagues where teams will “tank,” or intentionally lose games, in order to get a better draft order. This is not common in the NFL – but for this study, it will be assumed that this is never the case.

Operational Definitions

Downs: In football, teams have four attempts to advance the football ten yards down the field. If they fail to convert, or gain the desired ten yards, possession of the ball is turned over to the opponent.

Distance: Distance refers to the yards required to gain a new set of four downs.

Expected Points: To calculate Expected Points (EP) at a given down, distance, and field position, all of the ‘next points’ scored for and against every team that has been in that exact position are summed to create a net value. The EP is essentially stating how many points a team can expect to gain on that drive based on the current situation.

Expected Points Added (EPA): The gain in Expected Points from one situation to the next. This can be a positive or a negative number.

Field Position: Field Position is the position on the field, notated by the yard line, where a team is starting a play. The ball is snapped from this line and it is also referred to as the *Line of Scrimmage*.

Optimal Strategy: When faced with two or more choices, the optimal strategy is the choice with the greatest increase in win probability.

Yards To Go: This is how many yards to go to the opponent's endzone to score a touchdown. Some people use "distance" to represent yards to the endzone and "yards to go" to represent yards to the first down, but for this study, the definitions listed above describe what each measure will be referred to as.

Chapter 2: Literature Review

Teams in the National Football League are faced with decisions during every play of every game. However, there is a specific instance that teams just cannot quite seem to figure out: fourth downs. Through the 2019 – 2020 NFL season, the league had the highest percentage of fourth down attempts in at least twenty-five years at 14.5% (Schalter, 2019). This is in part due to the high success rate teams saw on fourth down conversions during the 2018 season, 59.4%; however, through November of the 2019 season, teams were only successful on 50.2% of conversion attempts (Schalter, 2019). The teams of the NFL run in to problems on fourth down no matter how you slice it: they do not go for it near enough, and when they do, they aren't necessarily using a smart strategy (Schalter, 2019). Still the question looms: would increasing the fourth down attempt rate result in a better outcome for teams? Would more fourth down attempts lead to more yards, points, and wins? There have been a handful of studies over this concept in the past, yet teams are still not including these ideas in their strategy – either due to the lack of awareness of the studies, fear of drifting from the status quo, or risk aversion. Or, better yet, maybe these studies are not comprehensive enough for NFL coaching staffs to place their confidence in them when it is crunch time.

The majority of the following studies were published in sports research journals or economics journals; other student theses and dissertations have also been reviewed. To search for the literature, SportDiscus was utilized through the University of Oklahoma Libraries' website. In the search for literature regarding fourth down strategy, the search terms included "football," "analytics," "strategy," and "fourth down." Articles in the English language discussing American football strategy and football analytics were included in this review. Articles about other sports or about football strategy but lacking a statistical analytics component were excluded. For the search for literature regarding risk aversion, articles cited in the literature from the first part of this section were reviewed, and relevant material was included. In the sports analytics field, especially football analytics, the peer-reviewed research is lacking – this is why work from theses and dissertations are included.

Fourth Down Decision Making

Fourth down strategy was first brought to light in late 1970 when Virgil Carter and Robert Machol derived expected point values of possession of the football with first down and ten yards to go (Carter & Machol, 1971). Using their expected point values as the basis of their argument, they discussed strategic implications, also noting that changes should be made during fourth down and goal to go situations (Carter & Machol, 1971).

For their analysis, Carter and Machol (1971) recorded 2,852 first down plays from the first half of the 1969 NFL season; due to the lack of data to calculate probabilities from each yard line, the field was divided into ten-yard increments with expected point probabilities from each. Due to having more accessible data now than in the 1970's, these expected point values have since been corrected in recent studies (Romer, 2006) to provide measurements from either smaller increments or individual yard lines. Regardless, Carter and Machol (1971) found that too many teams were kicking field goals in fourth and goal situations. They explain the scenarios that various decisions on fourth down could leave the opponent in, the corresponding expected point values, and the probabilities of each scenario after the fourth and goal play. The position that the opponent will be in is a crucial part of the decision-making process – a part that some later studies ignore.

Carter and Machol took their expected point value calculations one step further to prove that attempting a field goal on fourth and short is a poor strategy despite being the status quo

(Carter & Machol, 1978). For this study, Carter and Machol (1978) only considered situations in which the offensive team had passed midfield and short yardage was required to gain the first down. To complete their analysis, they calculated the expected value of attempting a field goal, punting, or trying for a first down at various field positions. To do so, they calculated the product of the expected value and the probability of the outcome, and subtracted from that the expected value that the opponent would be left with (Carter & Machol, 1978).

Due to the limited fourth down attempts at the time, they analyzed the success rate of fourth-and-one and third-and-one and found that the probability of success is extremely similar, furthermore establishing the assumption that the probability of success is identical. Other assumptions used include that a fourth down conversion gains exactly the required yards and that all players involved are average (Carter & Machol, 1978). This study also does not account for time and specific personnel and relies on coaches to assess those qualitative characteristics. However, they do note that they believe the reason for the field goal being the preferred choice of coaches is their lack of "intuitive feel for the negative value imposed on the opposition team" when they are given the ball so deep in their territory (Carter & Machol, 1978), which should be kept in mind when much of the model is left up to coaches' assessment ability. However, other studies (Yam & Lopez, 2019) now believe that the reason for this is risk and loss aversion, which will be discussed more in depth later in this chapter. The field-goal probabilities, similar to the calculation of expected point values, are lumped in to ten-yard intervals and expected point values are calculated in the same manner as they are for fourth down conversions. Carter and Machol concluded their study by claiming that attempting to convert on fourth-and-one or fourth-and-two is the preferred decision in all cases. Carter and Machol's (1971; 1978) studies provided a strong backbone for future fourth down strategy research, including other fourth down situations besides fourth and goal.

In a 2006, David Romer analyzed decisions made on fourth downs in the National Football League and, like Carter and Machol (1978) found that across the league, teams' choices on fourth downs systematically stray from the choices that maximize the probability of winning (Romer, 2006). His study was contrived to test an assumption of economic models that firms maximize expected profits – which makes it different from other football strategy research, as the football strategy was used as an example to prove an outside argument. However, Romer chose a strategic decision in sports to test this assumption for two reasons: he assumes that the problem of maximizing profits can be simplified to maximizing the probability of winning because of the assumed positive relationship he makes with winning and profits and there is abundant, detailed data available regarding the conditions of the situations in question (Romer, 2006). This reasoning is similar to Neale's study in 1964 – sports leagues have a wealth of data available, allowing it to be a natural labor market laboratory (Neale, 1964). It's now common to use sports settings to analyze business issues, as we have seen in other studies (Kahn, 2000). This study had logical valuations of football situations, which served as a basis for further studies, as well.

To determine if teams were maximizing their probability of winning, Romer focused his research around the expected point differential between going for it on fourth down or punting or kicking a field goal. To calculate the expected point differential, Romer first created his own method of determining expected points, based off of Carter and Machol's (Carter & Machol, 1971). He used data from over 700 NFL games between the 1999, 2000, and 2001 seasons to estimate the values of first downs at each point on the field - an improvement from Carter and

Machol’s ten-yard increments. He notes that, in order to avoid complications in his data resulting from one team being well ahead of another or the end of a half approaching, he uses only data from the first quarter for his analysis (Romer, 2006), which is a logical point, but limits the application of the analysis to only one quarter of play. This expected value, V_i, is the “expected long-run value, beginning in situation *i*, of the difference between the points scored by the team with the ball and its opponent when the two teams are evenly matched, average NFL teams” (Romer, 2006). This also generalizes the model too much. Over the course of an NFL regular season and postseason, there are 267 games played, and to make the assumption that even half of those games are two evenly matched, average teams, is quite a stretch. It would be more beneficial if the model was able to adapt to the teams involved in the scenario it is being used for.

At the time this study was published, fourth down attempts were rarer than they are presently, although not by much, and quite skewed; coaches tended to only go for it on fourth down when they were either in desperate situations or almost certain that they would succeed. Due to this being a rarity, Romer used the outcomes of third down plays to valuate trying for a first down. While the data for the first year of Romer’s three-year period was unavailable, for the other two years included in his analysis, there were 458 and 472 fourth down attempts, 12.1% and 12.3% (*Revolution or Convention?*, n.d.) of all fourth down plays, which is a stark contrast to over 16% in 2019. While this is still low, fourth downs should be included in the analysis, even if it includes merging third and fourth downs. Regardless of disagreements held with the evaluation of expected points, Romer’s method of comparing the values of going for a first down with the values of punting or kicking a field goal prove logical. He mentions the final step in his analysis is comparing with the teams’ actual decisions, and concludes that they are not making optimal decisions, but he does not go into any details about their decisions, how frequently they make optimal decisions, or if teams that make optimal decisions perform better over the course of a season. In *A Markov Model of Football,* discussed later in this chapter, Golder includes a realm to his study evaluating a team's efficiency against an expected outcome (Goldner, 2012), which would have been a useful addition to Romer's study. Romer's main conclusion is that the expected benefit to a "typical team" of increasing their aggressiveness on fourth downs in the first quarter is approximately 0.23 points per game (Romer, 2006). Again, "typical team" is most likely too broad to be relevant and the first quarter is a small portion of the game. It would be more interesting to look at the relationship between teams' fourth down attempt frequency and their performance over a season as measured by points scored and wins. He did, however, successfully reject his null hypothesis that the average values of kicking and going for it are equal, agreeing with Carter and Machol's (1978) study from approximately thirty years prior.

Another issue in this study is the assumption that a team whose aim is to maximize wins should be risk-neutral concerning points scored – which in turn overstates the value of a touchdown relative to a field goal because a touchdown is worth more points. This will overstate the benefits of going for it on fourth down in certain scenarios. Romer argues that the effect this gives is not significant due to coaches conservatism in all scenarios over the entire field (Romer, 2006). Another issue, as previously mentioned, is the discrepancies between third and fourth down. While there is a larger sample size when third downs are considered, the relative payoffs to various outcomes vary (Romer, 2006). While some are skeptical of attempting a fourth down conversion due to the increased momentum the play could give the opponent, Romer references other studies of momentum that find only small momentum effects at most. One study by Gilovich, Vallone, and Tversky investigated the myth of the hot hand in basketball – the belief that the chances of making a basket are higher following a made shot than a missed shot. However, Gilovich et al. found through the shooting records of the Philadelphia 76ers and free-throw records of the Boston Celtics, along with a controlled shooting experiment with the men's and women's teams at Cornell, found that there was not a positive correlation between the outcomes of sequential shots (Gilovich et al., 1985). In an additional study, Albright investigated the concept of hitting streaks in baseball by examining the records of many Major League Baseball players over four seasons. He implemented a complex logistic regression model, and while he found that some players did exhibit streakiness in certain years, the behavior of all players as a whole did not differ from what would be expected under a model of randomness (Albright, 1993). Although both of these studies are in sports outside of football, Romer still carried over the assumption that momentum does not play a role.

At the end of his study, Romer mentions that an analysis to fully encompass fourth down decisions over the entire game would require considering the score and time remaining, which, while more beneficial, would complicate the analysis. Perhaps one of the most interesting reasons he proposed for coaches straying from win maximization was one of the briefest: coaches are most likely trying to minimize immediate payoff variance rather than maximize win probability (Romer, 2006). This ties into risk aversion, loss aversion, and prospect theory (Kahneman & Tversky, 1979), which will be discussed later in this chapter. Since Romer's (2006) work has been released, there has not been a noticeable increase in fourth down attempts – they have remained fairly steady with small fluctuations (Schalter, 2019). It is interesting that this study is available to the public, yet NFL coaches are still not implementing these principles in their decision-making. This alone tells researchers that there is something missing.

In her 2007 thesis, Jennifer Wright statistically analyzed fourth down decisions and compared the expected gain or loss between choices (Wright, 2007). Her research question asked when teams should attempt a field goal, go for it on fourth down, or punt.

There were three attributes considered for the study and used to complete her analysis:

1. From a given yard line with a new first down, what is the expected number of points? The independent variable was yards from the goal line and the dependent variable was the number of points scored before giving up possession to the opposing team.
2. What is the probability, given the yards needed to convert a first down, that an attempt at a new first down is successful? The independent variable was yards to go for a first down and the dependent variable was success of converting.
3. What is the probability that a given field goal attempt is successful from a given yard line? The independent variable was number of yards from the goal line and the dependent variable was success of making a field goal.

The data for this study is comprised of play-by-play data of five teams from the 2005 season, obtained from the NFL’s website. While her decision to only use data from one season and five teams is not promoting the generalizability of her study, she did select teams in a novel way that allows for a more accurate, localized analysis. She selected five teams that were strong or weak offensively and defensively, and one team that was average on both sides of the ball in order to represent each possible combination of strengths and weaknesses. However, she, too, used data from third downs to predict probability of success of fourth downs (Wright, 2007), like Carter and Machol (1978) and Romer (2006).

She used linear and quadratic logistic regression models to estimate the probabilities needed for her study. To answer her research question, expected point differentials were calculated by comparing the product of success probability and the expected points of the respective outcome between decisions. The study found that teams, even poor offensive teams, need to go for it on fourth down more often than they do (Wright, 2007).

This study did excel in that it looked at each team individually, albeit only five teams, rather than generalizing the entire league. It also brought in expectations of the defense and their opponents' next drive into the decision. However, the study was too hypothetical by including additional yardage past the first down estimates. This could lead to overestimating the benefit of going for it on fourth down, and the additional yardage gained on a fourth down attempt could fluctuate extremely from the third down data she used to complete her analysis. This model is a good basis, but could use more factors, such as the ones that will be discussed in Yam and Lopez's (2019) study.

The purpose of *A Markov Model of Football* (Goldner, 2012) was to use a Markov chain model to simulate a football drive to improve situational understanding for in-game analysis. This study centered around a football drive being a Markov chain, meaning that the next state in a chain of events depends only on the current state (Goldner, 2012). Goldner uses this premise to assign a value to every down, distance, and yard-line scenario, which act as the independent variables of his study. The probability of a drive ending in a defined number of ways (scoring play, turnover, expired time, etc.) serves as the dependent variable (Goldner, 2012). These defined drive-ending scenarios are "absorption states" because each chain, or drive, must be absorbed into one of those scenarios. He used an absorption matrix to model the probabilities of any given state to end in an absorption, state (Goldner, 2012). An interesting added element would be using absorption properties for the defenses – to find probabilities for how defenses end drives and factor that in when considering opponents to develop a more accurate model.

A critique of this study is that Goldner assigns expected point values in five-yard increments by down and distance, similar to Carter and Machol's (1978) ten-yard increments. He argues for this by saying it prevents there being a state that never occurred or has too small of a sample size to provide an accurate analysis; however, with a larger dataset, that situation could be mitigated. Or, rather than placing situations in increments based on down and distance, one or the other should be chosen.

This model is more flexible and can be fit to measure specific play callers or players, which makes it relevant for applicable use. Another quality aspect of this study is Goldner's calculation of absorption time, or how many plays are required over the course of a drive to reach absorption (Goldner, 2012), which can affect play calling. This study found that incorporating absorption properties leads to a more accurate expected points model.

In 2016, Konstantinos Pelechrinis analyzed two specific discrete decisions in football, points-after-touchdown and fourth down, and whether or not coaches made rational decisions. He ultimately found that while coaches tend to stick to the status quo in their decisions, the status quo does not lead to point maximization (Pelechrinis, 2016), which holds true to what previous studies have found (Carter & Machol, 1978; Romer, 2006; Wright, 2007). He uses points scored as the dependent variable, which, as previously mentioned, might overstate the value of going for a touchdown rather than a field goal in specific situations. He also theorized that, rather than maximizing points scored, their coaching objective might be to minimize the variance of expected points scored, similar to Romer's hypothesis of coaches choosing to minimize immediate payoff variance (Romer, 2006). This means that although the more unconventional decisions might yield more expected points, the variance is higher, meaning the certainty of the expected points is lower. Again, this is why researchers must always be considering the psychological principles of risk and loss aversion.

To analyze fourth down decisions in Pelechrinis' model, three statistics are needed (Pelechrinis, 2016): conversion rate of a fourth down as a function of field position and yards to cover for a first down, the success rate of a field goal as a function of the distance from the goal, and the probability of success for a drive as a function of the starting field position for the drive. The last variable is needed in order to consider the consequences of potentially failing on the fourth down conversion and turning the ball over to the opponent. He then takes the difference between the expected point benefit and the expected point cost to derived the Mean Field Net Point Benefit, in which the expected point cost includes the potential points earned on a field goal as well as the potential cost of a turnover (Pelechrinis, 2016). This study takes this concept a step further by considering the consequence of turning the ball over, yet it falls short by again making a general, rather than team-specific, model. Another shortfall in this study is that the average fourth down conversion rate was not used; the conversion rate for the average yardage to go (7.58 yards) at the end of a drive was used, instead (Pelechrinis, 2016).

A 2019 study by Yam and Lopez is perhaps the most complete of the available literature thus far, also the most recent. Their research estimates the additional number of wins that each NFL franchise would have added over a 13 season span by implementing a more aggressive strategy on fourth downs (Yam & Lopez, 2019). Ultimately, they find, on average, teams would have gained an extra 0.4 wins per year (Yam & Lopez, 2019); this is very similar to the results drawn from Romer's study estimating an extra 0.3 wins per year (Romer, 2006), as previously discussed.

To work their analysis, Yam and Lopez implemented a nearest neighbor matching algorithm. The basis of this revolves around the New York Times' development of the 4th Down Bot, which provides recommended fourth down decisions (Burke et al., 2014). Yam and Lopez obtained play and game level data for the 2004 through 2016 NFL seasons and filtered out only the plays in which the 4th Down Bot recommends that the team should have gone for it rather than kicking; this left them with 13,172 fourth downs, including 9,348 downs in which teams kicked, rather than going for it.

The one-to-one matching strategy was used due to the relative lack of data for teams that went for it in certain situations. Plays were matched using the *Matching* R package (Sekhon, 2011) - matching plays in which the team did not go for it (the control group) with plays in which the team did (the treatment group) (Yam & Lopez, 2019). Plays were matched according to the following criteria (Yam & Lopez, 2019): nearest predicted probability of going for it (defined by the logit transformation of the propensity score, discussed in section 3.2 of Yam and Lopez's study), closest in-game win probability (derived from replications of the win probability models by Lock and Nettleton (2014) and Horowitz (2016)), yards to go for a first down, and most similar game time (number of minutes remaining in the game).

After removing all fourth down plays during which a penalty occurred, cleaning missing variables, plays outside of a common support interval, and implementing the outlined matching strategy, Yam and Lopez kept 7,698 pairs of plays. A worthy note to make here is that they used actual fourth down plays, rather than third down plays, to continue their analysis. This corrects for the discrepancies left by other studies (Carter & Machol, 1978; Romer, 2006; Wright, 2007).

The following covariates, taken from their study, are the independent variables involved in the analysis (Yam & Lopez, 2019): yards from own goal, yards to go for a first down, difference in offensive and defensive teams' scores, elapsed time in minutes, percent humidity, windspeed at kickoff, Las Vegas point spread, Las Vegas over-under, pre-snap win probability for the offensive team averaged between two win probability models, factor variable for home or away, week of the season, offensive team's pass offense rank (from Football Outsides), offensive team's rush offense rank (from Football Outsiders), defensive team's pass defense rank (from Football Outsiders), defensive team's rush defense rank (from Football Outsiders). They saw it as crucial to consider as many of the game and play variables that weigh on coaches' minds as possible. As one can see, these variables capture much more information than has been seen in previous studies.

The outcome variable in this study is change in win probability (Yam & Lopez, 2019). While this is variable is very relevant, when teams already have a high win probability, going for it might not produce as large of a change as would a team with a lower win probability. This is slightly mitigated by matching method that was utilized; however, this is something to keep in mind. Including win probability does, though, allow for proper estimates in calculating the chance for converting (Yam & Lopez, 2019). They show that teams that went for it in their dataset typically were the trailing team, and make the assumption that trailing teams are generally less talented; therefore, using those teams success rate outside of context would underestimate the chance of converting that more talented teams have (Yam & Lopez, 2019). In their results section, they discuss the notion that the only teams that would not have seen an increase in wins from a more aggressive fourth down strategy are the teams that already make better decisions on fourth downs (Yam & Lopez, 2019). They theorize that they could be underestimating the true impact of a proper fourth down strategy due to the fact that so much of their study revolves around the 4th Down Bot, which does not account for time remaining and point differential. They note that recognizing the best decision based on other variables would likely show an increase in the benefit (Yam & Lopez, 2019). Another caveat they mention is the possibility that they overstated the number of wins because of their use of change in win probability (Yam & Lopez, 2019). The dataset might include multiple plays from the same game, and it would be nearly impossible to prophesize if the team would see reduced fourth down situations late in the game if said team increased their aggressiveness early on in the game. Yam and Lopez (2019) argue that this does not have a drastic impact on their findings.

Their study also briefly examines some explanations for the behavior of NFL coaches that is deemed conservative. As has been previously mentioned, it is important to consider some psychological underpinnings of their behavior, yet not every study includes it. They also include intricate explanations as to why they used their chosen statistical methods and made their model team specific. This shows that Yam and Lopez were dedicated to creating a study to fully encompass the decision-making process.

This study, though, mentions the decrease in win probability from a failed fourth down attempt, but did not seem to have taken that into consideration when discussing their final conclusions.

Risk Aversion

In sports, there are always more variables to consider than just the numbers. As many of the strategy studies have mentioned, a possible explanation for the systematic deviance from win probability maximizing strategies is risk or loss aversion; often represented by minimizing the variance of the immediate payoff. Although attempting to convert a fourth down might make more logical sense to those who are informed on the subject matter, the perceptions belonging to those who are ill-informed are often the biggest critics. Before moving forward in the study, it is important to briefly touch on these various psychological factors that are heavily influential in the decision-making process.

In a 2018 study, Owens and Roach (2018) analyzed how college football coaches' job outlooks affect their fourth down decisions. They note that although some risky decisions might actually have a positive expected value for an organization, the person responsible for actually making the decision may have qualms about the "disadvantageous risk", causing them to choose a more conservative track (Owens & Roach, 2018). Owens and Roach (2018) found that in college football, decision-making is influenced by perceived job security. To be specific, coaches become more conservative than the status quo when their probability of being fired rises, and they become less conservative when they are faced with the possibility of promotion (Owens & Roach, 2018). This ties in with theories on herding and loss aversion and prospect theory. Due to the similarities in coaching importance, one could infer that these findings bleed over into NFL coaching decisions. A 2015 study on risk tolerance in play selection in football found that it is consistent with risk aversion (Critchfield & Stilling, 2015), which one could assume would transfer into fourth down decisions.

In a 1979 study, Kahneman and Tversky criticize the widely used expected utility theory and derive a new model of decision-making under risk: Prospect Theory (Kahneman & Tversky, 1979). The backbone of their research states that "people underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty" (Kahneman & Tversky, 1979). This notion is referred to as the certainty effect – essentially stating that decision makers are risk averse when choices involve sure gains, yet risk seeking when choices involve sure losses (Kahneman & Tversky, 1979). Another tendency outlined in their study is the isolation effect, when people discard shared components between decisions, leading to inconsistent preferences (Kahneman & Tversky, 1979). Ultimately, an alternative theory of choice is born, assigning value to gains and losses rather than final assets (Kahneman & Tversky, 1979). These gains and losses are judged from a neutral reference point, which varies based upon the starting scenario. An interesting point Kahneman and Tversky (1979) bring up, however, is that in scenarios where gains are possible but not probable, decision makers tend to choose the option with the larger potential gain. Translated to football, that would mean in a scenario in which winning the game already seemed out of reach, coaches would make the decision that could offer the greatest reward, regardless of the perceived risk. The researchers are sure to note that the derived value of a decision to an individual can certainly be affected by outside consequences, similar to the conclusions in Owens and Roach's (2018) study on collegiate football coaches' decision-making relative to their job outlook.

Summary

Most available studies are lacking in a crucial area, or missing something another study includes; however, if the best aspects from each of these studies were combined, a useful model could be developed that teams could implement on fourth downs. Considering all previous research on fourth down strategy in the NFL, a model able to adapt to all field positions, all yardages to go, time remaining, opponent, win probability, and team strengths would prove extremely beneficial. Data should be used from multiple seasons in order to have an adequate sample size, yet maintain relevance, and include fourth down situations only, which is possible, as seen in Yam and Lopez's study (2019). Although it will not be included in this study, it would be interesting to analyze absorption properties from a defensive standpoint, for both the team attempting the fourth down and their opponent, as well, similar to the absorption properties mentioned in the Markov Model (Goldner, 2012). This allows the model to consider the consequences of turning the ball over. Although including third down situations increases the available data, they do not represent the same consequences, game scenarios, emotions, and adrenaline that are present in a fourth down attempt. However, the skewness of certainty in fourth down attempts must be considered – coaches still tend to only attempt fourth downs in either desperate situations or situations they are certain in. This model will allow for accurate comparison with teams' actual decisions, in order to take an in depth look at the relationship between teams fourth down attempt frequency, success compared to expected outcome, and performance over a season as measured by yards gained, points scored, and wins.

Chapter 3: Methods

A recurring decision every coach in the National Football League has to face is whether to go for it, kick a field goal, or punt on fourth down. However, this is a decision that is difficult to assess. These decisions are influenced by status quo and risk aversion. As shown in many previous studies (Carter & Machol, 1971; Goldner, 2012; Pelechrinis, 2016; Romer, 2006; Wright, 2007; Yam & Lopez, 2019), coaches in the NFL are extremely conservative on fourth downs. The purpose of this study is to determine if making the win probability maximizing decision on fourth down leads to a greater number of wins per season, yards per game, or total points per season. The first step will be to create a new model to determine win probability in order to evaluate team's decision-making on fourth down. From there, this study will compare the frequency of which any given team in the NFL makes the optimal choice on fourth down by maximizing win probability with their yards per game, points per season, and wins per season. The research hypothesis for this study states that the frequency of which a team makes the optimal choice on fourth down is positively correlated with yards gained per game and in turn, the yards gained per game are positively correlated with wins per season and points per season. To accomplish this, this study will, as previously stated, make an in-depth win probability model to evaluate the fourth down decision-making of teams – when they should and should not go for it – which could be used for teams' in-game decision-making in the future. It is important to keep in mind, however, that teams that make optimal fourth down decisions most likely make optimal decisions overall and might not be in fourth down situations as often as other teams, especially not in dire or desperate situations.

This chapter will discuss the methods used to build a more comprehensive win probability model in order to evaluate these fourth down decisions. These models will have elements from Yam and Lopez’s (2019) win probability models as well as random forest algorithms similar to those in Lock and Nettleton’s (2014) win probability model. Although Carter and Machol (1971), Romer (2006), Wright (2007), Goldner (2012), and Pelechrinis (2016) all used expected point models, this paper will not. This methodology will also describe the predictor and outcome variables being used and compare them with the aforementioned studies. This is historical, applied research – attempting to find practical answers and bring a new perspective to an old problem by analyzing past quantitative data. It is non-experimental in the fact that it is simply observing, analyzing, and describing what is rather than manipulating a variable. It has a descriptive element – looking at what *is* when it comes to what teams are doing and what affects win probabilities. It has a correlational element as well – looking at the relationship between decisions, yards gained, points, and wins. What makes this problem so tricky, though, is that it will forever remain unknown what the true outcome of the alternative decision in any scenario would be. When a coach is faced with a fourth down decision, the only observable outcome is the one chosen. This is referred to as the fundamental problem of causal inference (Rubin, 1974). To respect this fact while still allowing for the model to be feasible in the context of the National Football League, many concepts of a Rubin Causal Model will be applied, such as some that Yam and Lopez (2019) included in their study.

As previously mentioned, many prior studies have hinged on the concept of expected points and expected points added – which has to take the game situation into account in order to be relevant, as it revolves around the discrete concept of points. For example, being up by one point at the end of a game has the same value as being up by forty points at the end of a game; however, when considering only the concept of points, the higher number of points could be overvalued. It is common in the sport for teams to sacrifice more points in exchange for using up time on the clock, which ultimately will help them win. Therefore, in this study, the optimal strategy will revolve around win probability, such as in Yam and Lopez's (2019) study, in order to home in on a team's main goal: winning the game.

Sample

Due to the nature of this study, a sample does not need to be recruited. Through a contact at Pro Football Focus, a website dedicated to thorough analysis of professional and collegiate football, a play-by-play dataset for every game over the previous ten NFL seasons (2010 to 2019), including 41,512 fourth downs, was obtained. As mentioned previously, some teams are not placed in as many desperate fourth down situations, and teams have very different strengths that affect the outcome of any decision. To mitigate this, a win probability model will be built to estimate the offensive team's win probability at any given time. This will be built based off of all 442,475 plays in the dataset. In addition to the data from Pro Football Focus. This model will also include teams' offensive and defensive ratings, according to FootballOutsiders (*2019 NFL TEAM EFFICIENCY RATINGS | Football Outsiders*, n.d.). FootballOutsiders is a website dedicated to football statistics and the analysis of said statistics.

While many previous studies have included third downs when conducting their analyses for fourth down conversion success (Carter & Machol, 1971; Romer, 2006; Wright, 2007), this study will not include them because of the vastly different consequences that a failed third down and a failed fourth down leave behind. Third down also does not represent the same consequences, game scenarios, emotions, and adrenaline that are present in a fourth down attempt. In fact, conversion success rates will not be directly included in the analysis, which will be explained later in this section. Yam and Lopez's study (2019) also proved that an analysis can be conducted with fourth downs only. However, fourth downs with ten yards to go or greater will be grouped into one group representing fourth and long; in the R code, this will be represented as fourth and 11 for simplicity purposes. A still present limitation, though, is the skewness in fourth down attempts that is brought about due to coaches tending to go for it when they are almost certain they can convert or are in desperate situations. This trend, as observed from the casual observer, is starting to turn around during the 2020 NFL season.

Data

The following tables outline the variables used in this study. The variables are split by source: Pro Football Focus, FootballOutsiders (*2019 NFL TEAM EFFICIENCY RATINGS | Football Outsiders*, n.d.), FootballLocks (*NFL Odds Super Bowl LIV. Super Bowl 54 NFL Vegas Odds for This Week. Free Pro Football Betting Odds.*, n.d.), or calculated fields.

Pro Football Focus is a website dedicated to the thorough analysis of football in the United States. Their analysis is so reliable and accurate that the NFL itself will sometimes include their player rankings in their broadcasts. Pro Football Focus provides customized data to every NFL team, as well as 74 NCAA FBS teams, four Canadian Football League teams, media outlets, and sports agencies ("Pro Football Focus," 2020). They collect their own data and perform their own analyses.

Football Outsiders is another website dedicated to objective analysis of football. They have a partnership with ESPN and have been featured on other websites, such as Sports Illustrated, SB Nation, CBS Sports, and Bleacher Report (*Football Outsiders Frequently Asked Questions | Football Outsiders*, n.d.). The also publish detailed explanations of each statistic they create. According to their website, Football Outsiders obtains their data from Sports Info Solutions, who is "the leader in the collection, analysis, and delivery of top-quality sports data content" (*Sports Info Solutions*, n.d.). The Team Offense and Defense Defense-Adjusted Value

Over Average Ratings “measures a team’s efficiency by comparing success on every single play to a league averaged based on situation and opponent” (*Methods To Our Madness | Football Outsiders*, n.d.). These take into consideration how the opponent’s ability feeds into the situation, as well.

Football Locks is a website that keeps track of historical professional football betting information. They simply keep a record of all Las Vegas betting lines for all NFL games. The over/under line metric was obtained through a dataset on Kaggle, a popular data analysis website that hosts competitions and provides datasets. The over/under line was double checked with other websites, and was accurate, however, the Kaggle download was a simple way to get the data into RStudio where the analysis took place.

Table 1: *Variables Obtained from Pro Football Focus*

Variable	Description
season	Season play occurred during
week	Week play occurred during
home_team	Home team of game
away_team	Away team of game
offense	Team on offense
defense	Team on defense
quarter	Quarter of play (1^{st}, 2^{nd}, 3^{rd}, or 4^{th})
down	The current down (1^{st}, 2^{nd}, 3^{rd}, or 4^{th})
distance	Distance to go for a 1^{st} down
seconds_left_in_quarter	Seconds left to play in the quarter
off_score	Score of the offensive team
def_score	Score of the defensive team
yards_to_go	Distance to opponent's endzone
play_type	Factor variable representing a regular or special teams play
regular_play_type	Factor variable representing a regular play type
special_play_type	Factor variable representing special teams play type
field_goal_made	Factor variable for if a field goal attempt was successful
final_home_team_score	Ending score for the home team
final_away_team_score	Ending score for the away team
next_play_down	Down of the following play (1^{st}, 2^{nd}, 3^{rd}, 4^{th})
next_play_distance	Distance to first down of the following play
next_play_yards_to_go	Distance to opponent's endzone for the following play
next_play_offense	Team on offense for the next play
next_score	Next points scored
score_on_play	Factor variable representing if points were scored on the play
penalty	Factor variable representing if a penalty was called
first_down	Factor variable representing if a first down was gained

Table 2: *Variables from FootballLOCKS and FootballOutsiders*

Variable	Description
lvps	Las Vegas pre-game point spread (FootballLOCKS)
ou	Las Vegas over/under (Kaggle Dataset)
offensive_rank	Offensive team's overall offense rank (FootballOutsiders)
offensive_rush_rank	Offensive team's rushing rank (FootballOutsiders)
offensive_pass_rank	Offensive team's passing rank (FootballOutsiders)
defensive_rank	Defensive team's overall defense rank (FootballOutsiders)
defensive_pass_rank	Defensive team's rank against the pass (FootballOutsiders)
defensive_rush_rank	Defensive team's rank against the rush (FootballOutsiders)

Table 3: *Calculated Variables*

Variable	Description
seconds.left	Seconds left to play in game
point_diff	Difference in offensive and defensive teams' score
deltawp_home_1	Change in win probability from previous play for home team
deltawp_off_1	Change in win probability from previous play for offense
Attempt_Fourth	Factor variable representing an offense going for it on fourth down
win_totals	10 season regular season win total
win_max_frequency	Frequency of which the team chooses the win probability maximizing choice out of the matched pair in the sample
YDS	Total regular season yards over the ten-year sample span
YDS.G	Average Yards Per Game in the regular season over the ten-year sample span
Total Win	Total regular season wins in the ten year sample span
won.off	Factor variable representing if the team with possession won
won.home	Factor variable representing if the home team won

Research Design

The study will be completed in the following steps. The statistical and analytical procedures used in each step will be explained in detail in the subsequent Data Management and Analysis section.

1. *Build win probability model for use in analysis.* The win probability figures will also be used to gauge good decisions through change in win probability after a play. This win probability model will be crucial in determining if teams are making the "optimal" choice. The variables used in building this model via a random forest algorithm are listed in Table 4 on the following page.
2. *Create Propensity Score Model.* This will model the probability of a team attempting to go for it on any given fourth down. This will consider the play and game characteristics on all fourth down attempts in the data set. Since it is impossible to match two fourth down situations that are exactly the same, the propensity score will

Table 4: *Variables Used in the Win Probability Model*

Variable	Description
Offensive_Rush_Rank	End of Season Offensive Rushing Rank
Offensive_Pass_Rank	End of Season Offensive Passing Rank
Defensive_Rush_Rank	End of Season Defensive Rushing Rank
Defensive_Pass_Rank	End of Season Defensive Passing Rank
score_diff	Score Differential between the offensive and defensive team
seconds.left	Seconds Left in the Game
distance	Distance to Gain 1st Down
down	Down out of 4
yards_to_go	Yards to Endzone
week	Week of Season
home	Factor variable representing if the offense is the home team

Table 5: *Summary Statistics for Variables in Win Probability Model*

Variable	Minimum	Maximum	Mean	SD	Unique N
score_diff	-56	56	-0.7	11.0	105
seconds.left	0.0	3600.0	1748.0	1049.3	3601
distance	0	50	7.8	4.7	47
down	0	4	1.8	1.1	5
yards_to_go	1	99	51.9	24.9	99
week	1	17	9.1	5.0	17
home	0.0	1.0	0.5	0.5	2

allow for the matching of highly similar scenarios. The variables listed in Table 6 will be used in the propensity score model.

This model will again be based off of Yam and Lopez’s (2019) model. However, there are some fourth down situations where teams would almost certainly never go for it, and there are some fourth down situations where all teams would go for it. To mitigate this, a common support interval (Dehejia and Wahba, 1999) will be found by filtering the range for each of the treatment and control groups. After this, the propensity score model will be refit onto the data by this common support interval.

3. *Nearest Neighbor Matching Algorithm.* Through the *Matching* package in R, a 1:1 nearest neighbor matching algorithm will be used to pairs teams that went for it on fourth down with teams that did not go for it on fourth down. The matching will be based off of a logit transformation of the propensity score, win probability, distance to gain first down, and seconds remaining in the game. One-to-one matching matches exactly one team in each group to each other – a good choice for this study given the specificity of the scenarios.

Table 6: *Propensity Score Model Covariates*

Covariates	**Description**
seconds.left	Seconds left to play in game
*yards_to_go * distance*	Interaction term between yards to endzone and yards to gain for first down
*yards_to_go * seconds.left*	Interaction term between yards to endzone and seconds left in the game
*yards_to_go * score_diff*	Interaction term between yards to endzone and score differential
*distance * seconds.left*	Interaction term between distance and seconds left
*distance * score_diff*	Interaction term between distance and score differential
*seconds.left * score_diff*	Interaction term between seconds left and score differential
score_diff	Difference in offensive and defensive teams' score, respective to the offensive team
yards_to_go	Yards to go to the endzone
distance	Yards to gain for first down
Attempt_Fourth	Factor variable representing an offense going for it on fourth down
spread	Las Vegas Point Spread, respective to the offensive team
over_under	Las Vegas Over/Under
Offensive_Pass_Rank	Offensive team's pass rank
Offensive_Rush_Rank	Offensive team's rush rank
Defensive_Pass_Rank	Defensive team's rank against the pass
Defensive_Rush_Rank	Defensive team's rank against the rush
Home	Variable representing if the offensive team is the home team
week	Week of the season
wp_1	Offensive team's win probability

Table 7: *Summary Statistics for Propensity Score Model -- Only Fourth Down Plays Included*

Variable	**Minimum**	**Maximum**	**Mean**	**SD**	**Unique N**
seconds.left	0.0	3588.0	1701.9	1045.7	3544
yards_to_go	1	99	50.9	25.1	99
distance	1	48	7.8	5.7	45
score_diff	-58	58	-1.0	10.8	102
Attempt_Fourth	0	1	0.1	0.3	2
spread	-26.5	26.5	0.3	6.2	80
over_under	33.0	61.5	44.8	4.2	55
home	0	1	0.5	0.5	2
wp_1	0.0	1.0	0.5	0.3	37670

4. *Create a "score" for teams on their fourth down decision-making prowess.* This score will allow not only for teams to be ranked against each other, but will also allow for a correlational analysis to determine if teams with better fourth down decision-making gained more yards per game, scored more points per season, and ultimately enjoyed more wins per season. In the data frame, a binary variable titled "correct_choice" will be added, and for each play that the team made the optimal choice within their matched pair, that play will receive a one (1) in that column; teams not making the optimal choice will receive a zero (0) in that column.
5. *Run a Poisson regression analysis.* The regression analysis will determine whether or not win probability maximizing decision-making on fourth down influences yards per game, point totals, and win totals. Another component in this final analysis piece will explore what trends have actually occurred over these ten seasons in the data set.

Threats to Validity

Since this is historical data from the entire population, there are not many internal threats to validity. The biggest threat to internal validity is the aforementioned fact that there might not be a large variety of fourth down attempts due to coaches tending to go for it when they are certain of the outcome or in desperate situations, causing a skewness in the attempts. The main threat to validity is the generalizability, as teams' strengths change yearly. To address this, instead of building the model based off teams, it will be built based off of offensive and defensive ranks, along with win probabilities. This will make the model flexible for the changing dynamic across the NFL. For example, during the 2020 football season, the Pittsburgh Steelers started off 11-0 and then lost four out of their five next games and were eliminated from the playoffs in the first round. If we had a stagnant categorical variable for the 2020 Pittsburgh Steelers team, it would have made them look like one of the best teams in the league for the first eleven games based off of their record. However, considering offensive and defensive ranks, the true, underlying nature of the team's specifics strengths will be taken into consideration. The ecological validity of this study is appropriate because this is historical data from every play of every game from the last ten NFL seasons, perfectly representing the target population.

Data Management and Analysis

The data will be kept, managed, and analyzed in R software. R is widely used software because it is an open-source software with aesthetically pleasing graphics and analytical capabilities. This allows for easy manipulation of the data, being able to run analyses on the full data set for the win probability model, as well as the subset of fourth down plays only to analyze success probabilities.

The study will include a descriptive and analytical component. The descriptive statistics will be presented in table and graphical form to outline win probability trends, accuracy of the models, and teams' historical patterns.

The data analysis procedure will include the aforementioned five steps: 1) win probability model, 2) propensity score model, 3) nearest neighbor matching, 4) scoring of the decisions, and 5) final regression analysis.

1. *Win Probability Model.* The win probability model will be built using a random forest algorithm, similar to the one Lock and Nettleton (2014) proposed for the New England Symposium on Statistics in Sports and one used in Yam and Lopez's (2019) study. The purpose in building this will be to identify the probability that at any specific moment, a given team will ultimately win the game. This will allow for evaluation of play calling decisions through the change in win probability after a play, and it will serve as a predictor variable in the propensity score model and nearest neighbor matching algorithm. The independent variables used will be down, distance, field position, score differential, seconds left, offensive pass rank, offensive rush rank, defensive pass rank, defensive rush rank, week of season, and home team; the dependent variable will represent if the offense won – listed previously in Table 4. To split the dataset into a training and testing set, two games from each week will be in the test set and the rest will be in the training set, which will mitigate any season inconsistencies. To test the data, the bins in the random forest algorithm will be utilized. The plays in the test set will be binned by estimated win probability and then the proportions of wins in each bin will be calculated. This proportion of wins will represent the unknown true win probability – this will be shown in graphical form. This is similar to Lock and Nettleton's (2014) study After building and confirming accuracy for the model, these calculations will be incorporated into the dataset.

A random forest algorithm was chosen for its simplicity and diversity, as it can execute classification and regression tasks. This algorithm will build many decision trees and combine them to achieve a more accurate prediction (Donges, 2019). While the algorithm is working, it searches for the best feature among a random subset of features while splitting a node, rather than searching for the overall best feature. This generally results in a better model because of the wide diversity this method brings about. The random forest algorithm also allows for easy discernment on the relative importance of each feature; the algorithm does this by "looking at how much the tree nodes that use that feature reduce impurity across all trees in the forest" (Donges, 2019). This score is then automatically computed for each feature and scales the results so that the sum of all combined features' importance is equal to one. Each tree in the forest is grown from a bootstrapped version of the original sample. The random forest algorithm altogether allows for complex interactions between predictor variables – interactions that one can safely assume are there, although they might not understand the extent of the interaction. This method is reliable due to the predictions being based entirely on empirical evidence, requiring few costly assumptions. This will be constructed using the *randomForest* function in the R package *randomForest*. This function requires two tuning parameters: *m*, the number of candidate predictors at each split, and *nodesize*, the maximum terminal node size. For *m*, two (2) will be used and for *nodesize*, 200 will be used, and the default of 500 regression trees will be constructed. These decisions are again based off of Nettleton and Lock's (2014) study.

2. *Propensity Score Model.* After successfully tuning the win probability model, the win probability for the offense will be added to each play in the dataset. Then the propensity score model, $e(X) = P(W|X)$, will be created, as previously mentioned. This is created through a multiple logistic regression model and the covariates listed previously in Table 6. The purpose of

this is to balance the variables in X between the teams that did and did not go for it to be later used in the matching section of the analysis.

3. *Nearest Neighbor Matching.* Matching is then done through the *Matching* package in R, via the propensity scores, win probability, distance, and time remaining. This will pair up two fourth down plays that are extremely similar, which will allow for comparisons in step four of the analysis.

Table 8: *Summary Statistics for Matching Variables*

Variable	Minimum	Maximum	Mean	SD	Unique N
predict.fitms2	0.0	1.0	0.1	0.2	37218
wp_1	0.0	1.0	0.5	0.3	37214
distance	1	43	7.7	5.6	42
seconds.left	0.0	3588.0	1704.6	1047.0	3543

4. *Create a Score for Team's Decision Making.* Based off of the pairs created through matching, a win probability maximizing frequency will be calculated. In each pair, it will be noted which team maximized their win probability with their decision – either the team that did or did not go for it. This will simply be which team saw a greater increase in win probability from the start to the end of the fourth down play. Again, this will only be built off of the sample of matched pairs, and only includes fourth downs. Then, a frequency will be created through dividing the number of times each team maximized their win probability by the number of occurrences that team had in the sample. This will be done for each team by individual season and as a total over the ten-year period. Splitting this metric by season will allow for a brief exploratory comparison between the win probability maximizing frequency on fourth down and the team's winning percentage.

5. *Regression Analysis.* Finally, a Poisson regression analysis will be run to determine relationships between the following variables:

1. Win Probability Maximizing Frequency | Total Yards Over the Course of a Season
2. Win Probability Maximizing Frequency | Total Points Over the Course of a Season
3. Win Probability Maximizing Frequency | Total Wins Over 10 Seasons

Implementing the methods outlined in this chapter will produce a thorough, well-rounded analysis. The model will be more comprehensive than previous fourth down evaluation models and will bring about results from a new, fresh perspective. The results will be able to not only evaluate past decision-making in the NFL, but can assist in future decision-making, all while being able to adapt to the changing situations in the NFL landscape.

Chapter 4: Results

Current Trends

Before moving too deep into the results of the analysis, it is important to acknowledge some current trends in the proportion of fourth downs attempted in the 10 regular seasons in this sample. In the below graph, a large increase in attempts from 2017 to 2018 can be seen, and another increase from 2018 to 2019. The 2011 season sees the lowest proportion of fourth down attempts at approximately 11%, and 2019 sees the highest proportion, at approximately 16.3%.

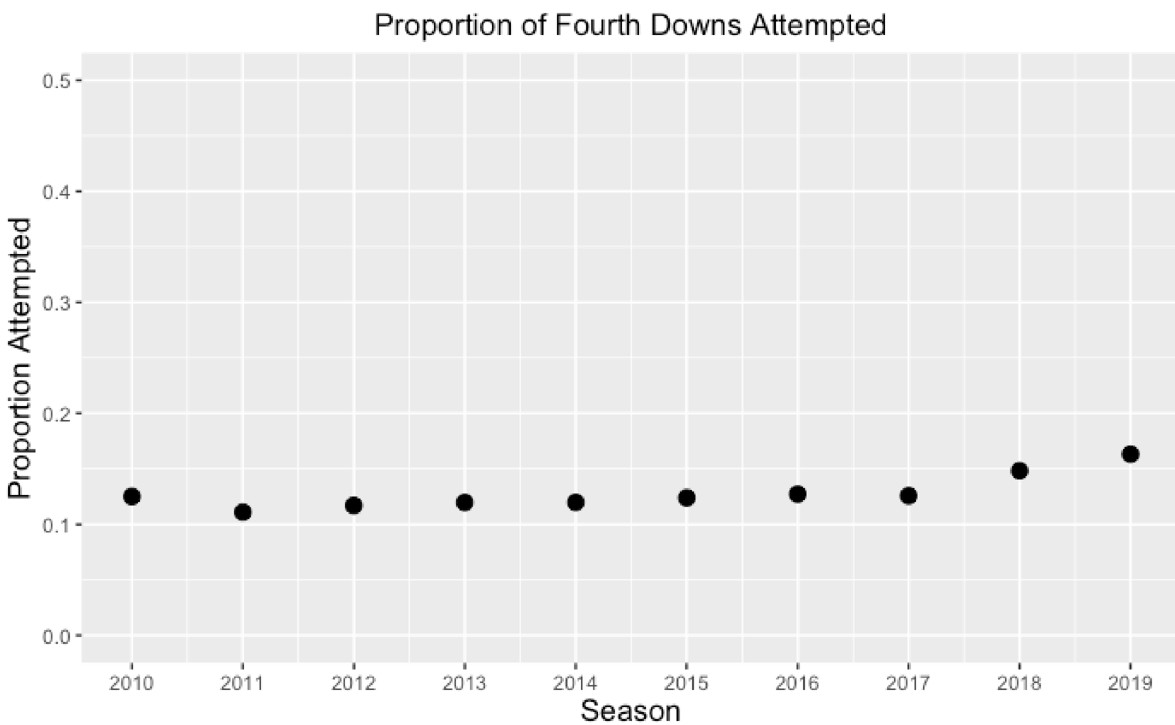

Figure 1: *Proportion of Fourth Downs Attempted from 2010 - 2019*

Win Probability Model

The win probability model in this study proved more accurate than previous studies. The model in this study has a Mean of Squared Residuals of 0.07489, down from Lock and Nettleton’s (2014) base model which had a Mean of Squared Residuals of 0.15; it also explains 70.01% of the variance, up from Lock and Nettleton’s (2014) model explaining 38% of the variance. Below is a plot of the accuracy of the win probability model, tested through the aforementioned binning method. Along the X-Axis is the estimated win rate, with the observed win rate of the bins along the Y-Axis.

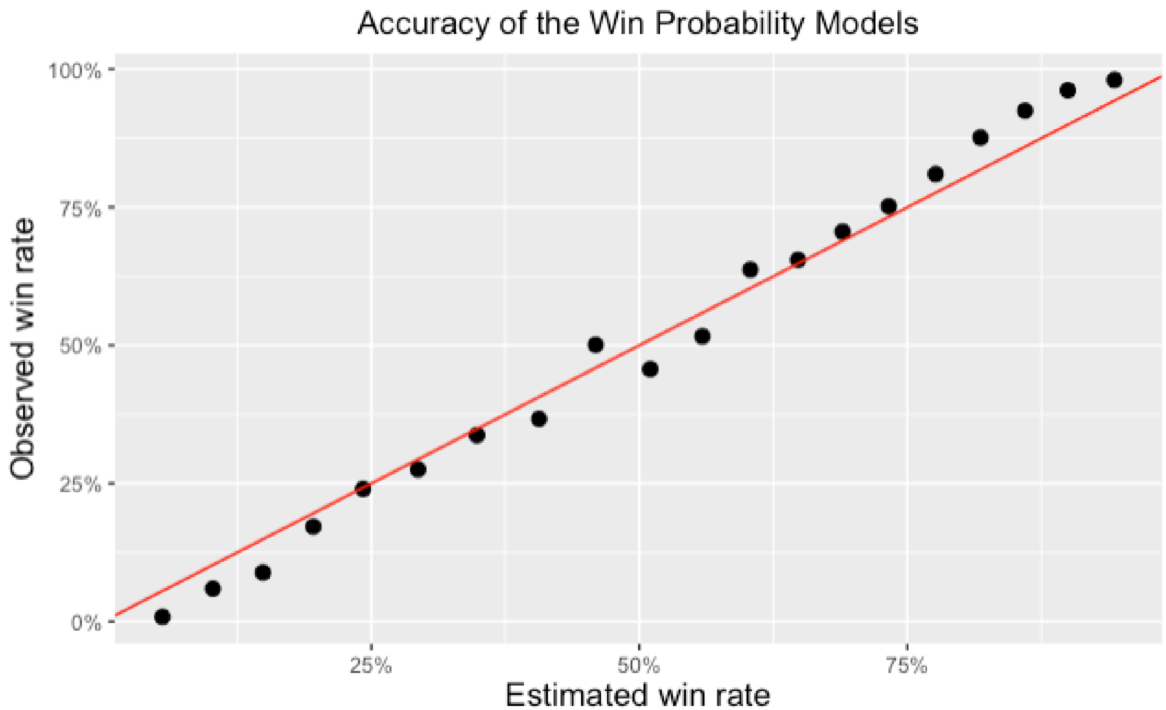

Figure 2: *Accuracy of the Win Probability Models*

Although the main goal of this study was not to build a win probability model, it was a crucial step in determining the final outcome of the analysis. This model could also be used in future decision-making - to evaluate historic and looming decisions through win probability estimates. More future uses will be discussed in the following Discussion section.

Propensity Score Model

As has been mentioned multiple times throughout this study, there are situations when teams certainly never attempt to convert a fourth down, and there are situations when teams almost always attempt to convert a fourth down. The graph below shows the imbalance in the point differential between teams in this sample that did and did not go for it. As can be seen via the density plot, teams that are behind tend to attempt to convert a fourth down more than teams that are leading. This trend is one that must be kept in mind as the results are presented.

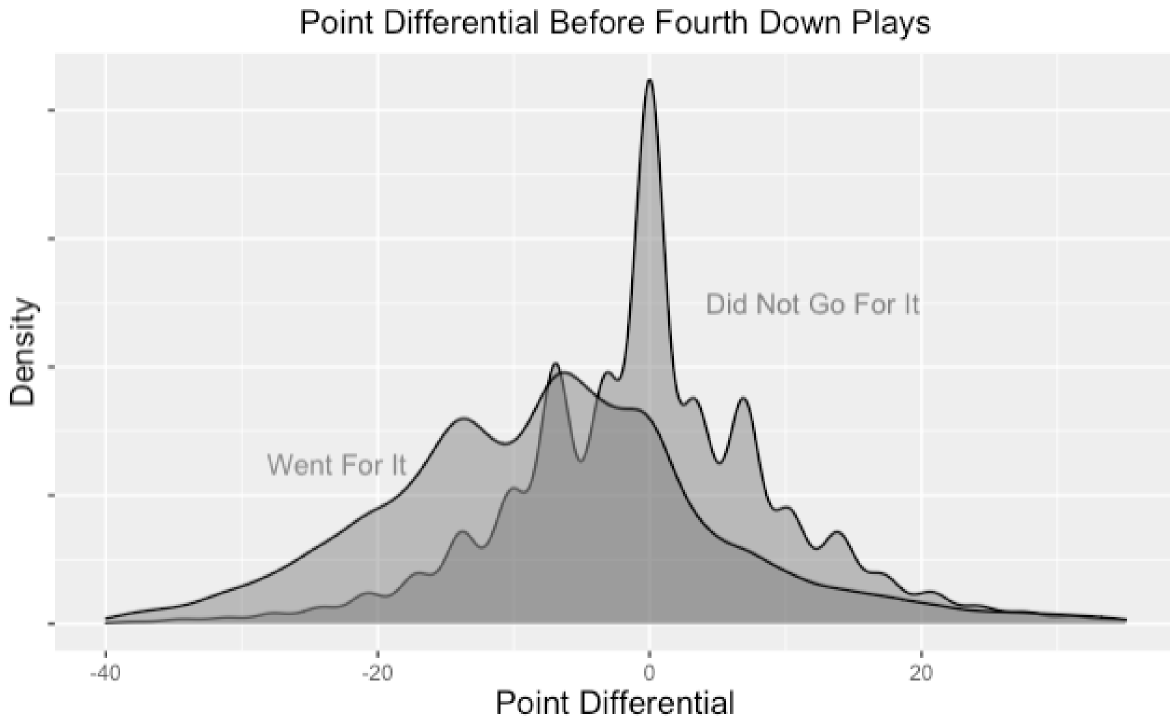

Figure 3: *Point Differential Imbalance Between Teams that Did and Did Not Go For It*

The propensity score model assisted in balancing the variables between these two groups. The following graph depicts the new point differential imbalance after filtering out plays in which the propensity score distributions did not have any overlap. This provides a more balanced sample to run the analysis on, as can be seen in Figure 4.

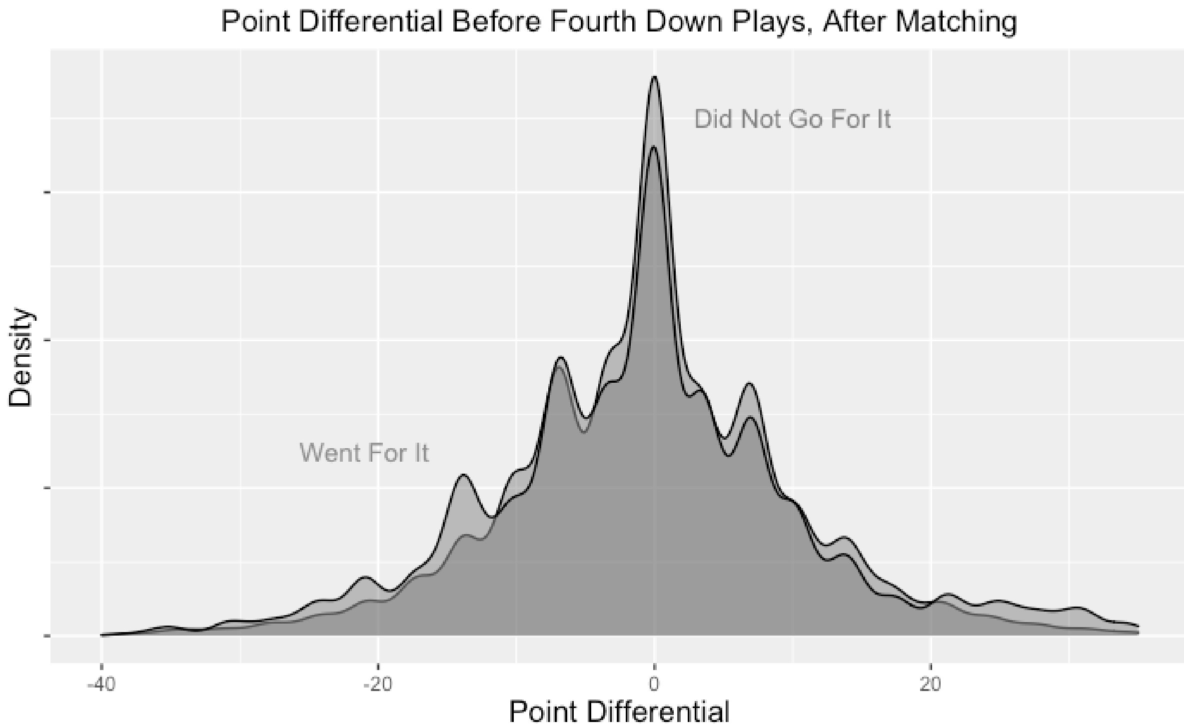

Figure 4: *Point Differential Between Teams that Did and Did Not Go For It After Matching*

Nearest Neighbor Matching

The original sample consisted of 41,398 fourth down plays. After implementing one-to-one matching, there were 21,062 matched observations. Again, each pair included one play in each group – "went for it" or "didn't go for it." We can see the success of the matching in the Figure 4 through the similarities between the density curves, especially when contrasted to Figure 3.

The results of the Wilcoxon Rank Sum Test returned a P-Value of < 0.001, well under 0.05, meaning there is a significant difference in the change in win probability between the two groups from the start to the end of the fourth down play. The average via mean change in win probability could be thrown off due to outliers in each group. The mean change in win probability on plays where fourth downs were attempted is -1.2479%, while the mean change on plays where fourth downs were not attempted is only -1.0545%.

Table 9: *Change in Win Probability*

Attempt Fourth	Mean Change in Win Probability	Median Change in Win Probability
No	-1.0545%	0.0000%
Yes	-1.2479%	0.5360%

N = 42,124.

Since going for it on fourth down is riskier, with a higher variance in the outcome, seeing these values for the means of the two groups is not surprising. However, the median, which is less affected by outliers, shows that the median change in win probability for plays that included a fourth down attempt is an increase of 0.536%, while plays that did not attempt a fourth down have a median increase of 0.0000%. Due to outliers, this metric is more reliable to base a conclusion off of. We have to remember, though, that this analysis includes all fourth down scenarios, unlike Yam and Lopez’s that only includes those in a “go for it” range established by the New York Times’ Fourth Down Bot. However, the purpose of this paper, again, is to determine if there is a relationship between teams making the “win probability maximizing” decision – whatever it may be – and actually winning.

The following density curve shows the change in win probability between each of the two groups. This graph depicts the change for all fourth down plays within the common support interval. As can be seen, on fourth down plays where the team went for it, the density curve for the change in win probability is more dense in the positive values.

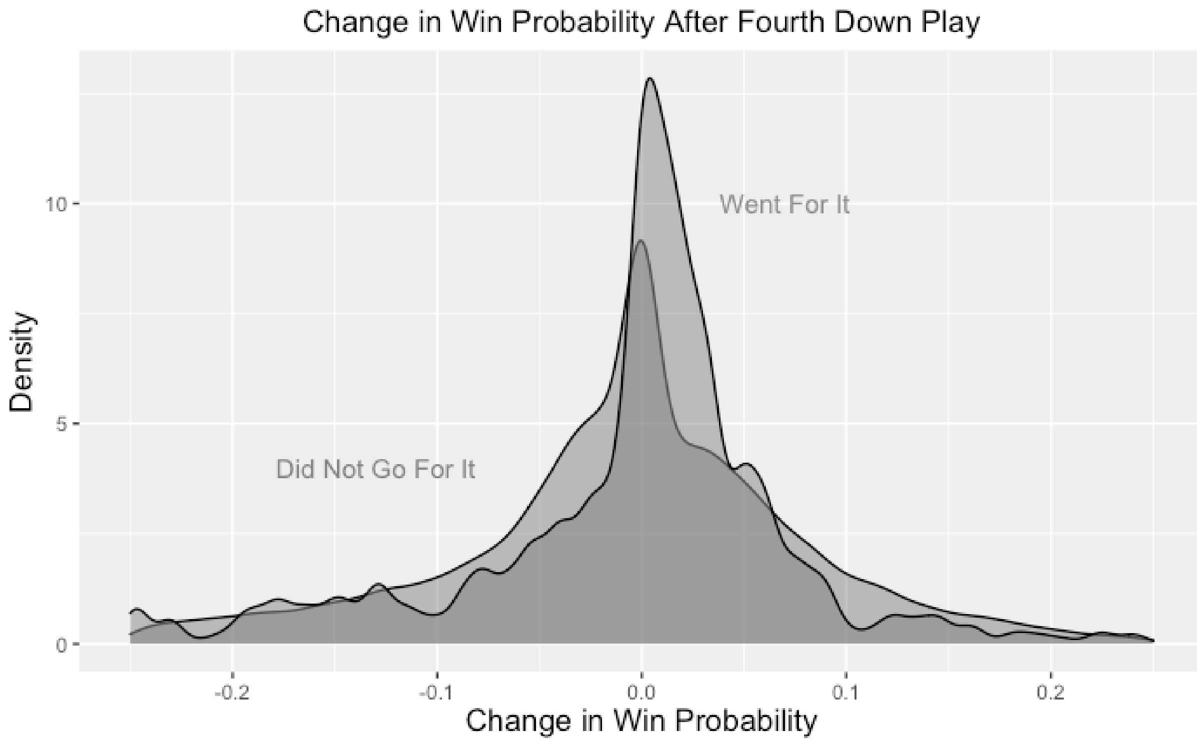

Figure 5: *Change in Win Probability After Fourth Down Play*

Win Probability Maximizing Frequency Score

The win probability maximizing frequency score is simply the frequency of plays that each team made the win probability maximizing choice out of each pair they were included in throughout this sample. For example, if Team A was included in 10 pairs and made the win probability maximizing decision 7 times, their win probability maximizing frequency score would be 0.7, or 70%. The following graphs shows each teams' win probability maximizing frequency as the bars with the team's win frequency marked as points, split by season; the y-axis represents the frequency. As visible through the graphs, even the best teams do not make the win probability maximizing decision 100% of the time.

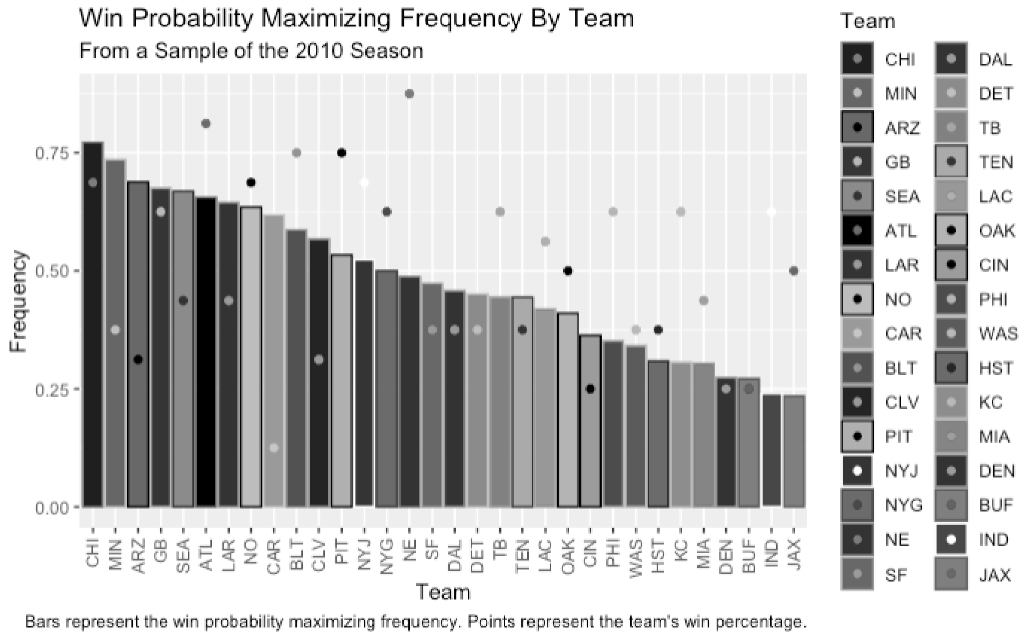

Figure 6: *Win Probability Maximizing Frequency & Win Frequency By Team For 2010 Season*

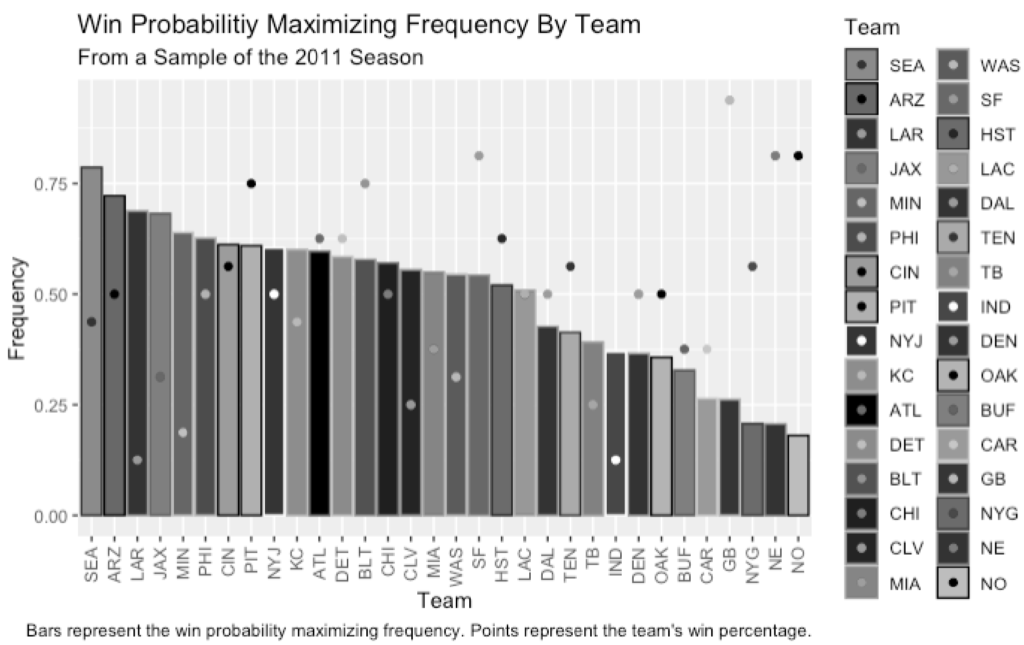

Figure 7: *Win Probability Maximizing Frequency & Win Frequency By Team For 2011 Season*

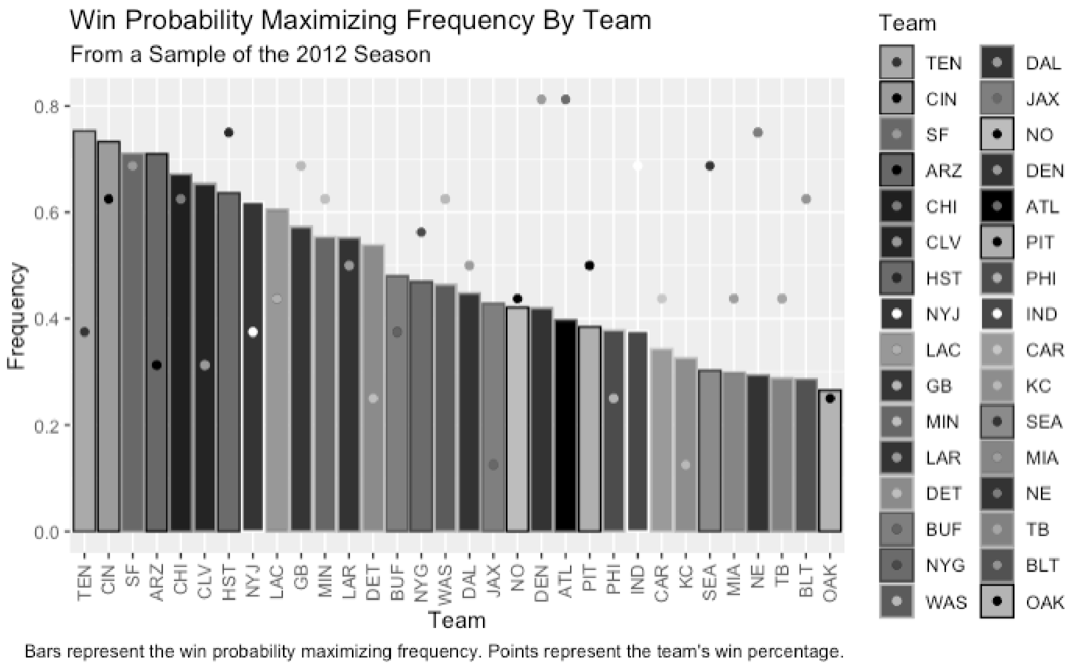

Figure 8: *Win Probability Maximizing Frequency & Win Frequency By Team For 2012 Season*

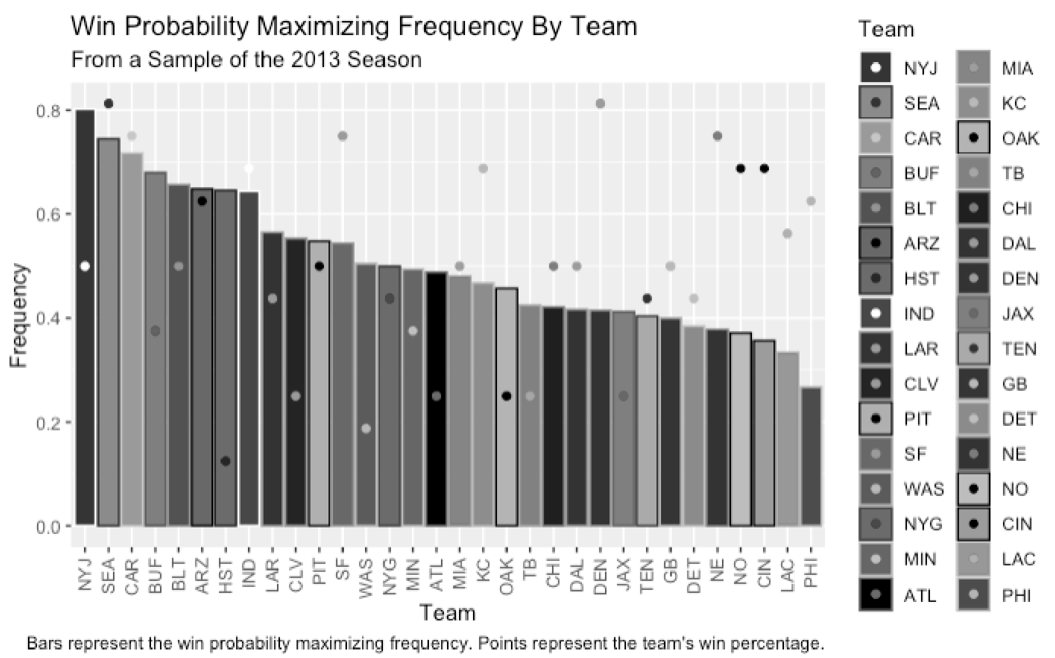

Figure 9: *Win Probability Maximizing Frequency & Win Frequency By Team For 2013 Season*

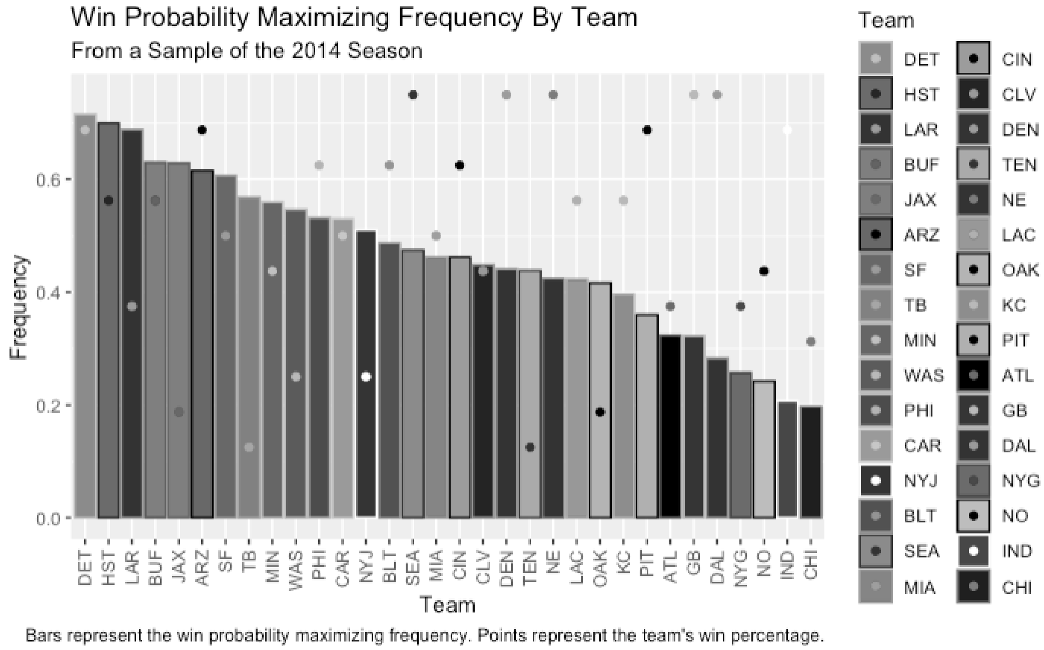

Figure 10: *Win Probability Maximizing Frequency & Win Frequency By Team For 2014 Season*

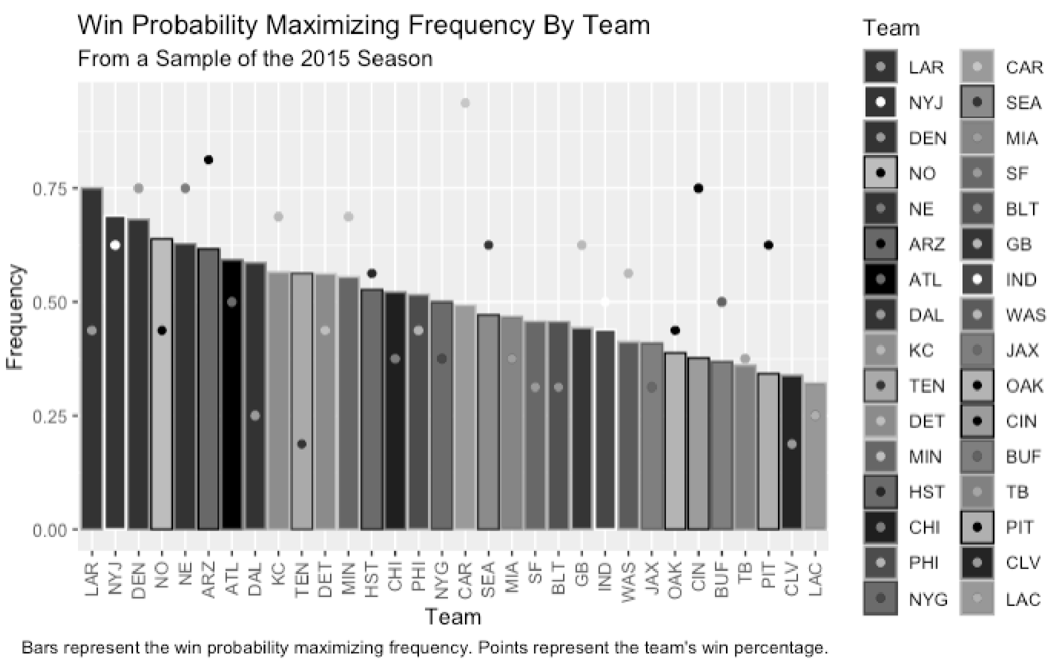

Figure 11: *Win Probability Maximizing Frequency & Win Frequency By Team For 2015 Season*

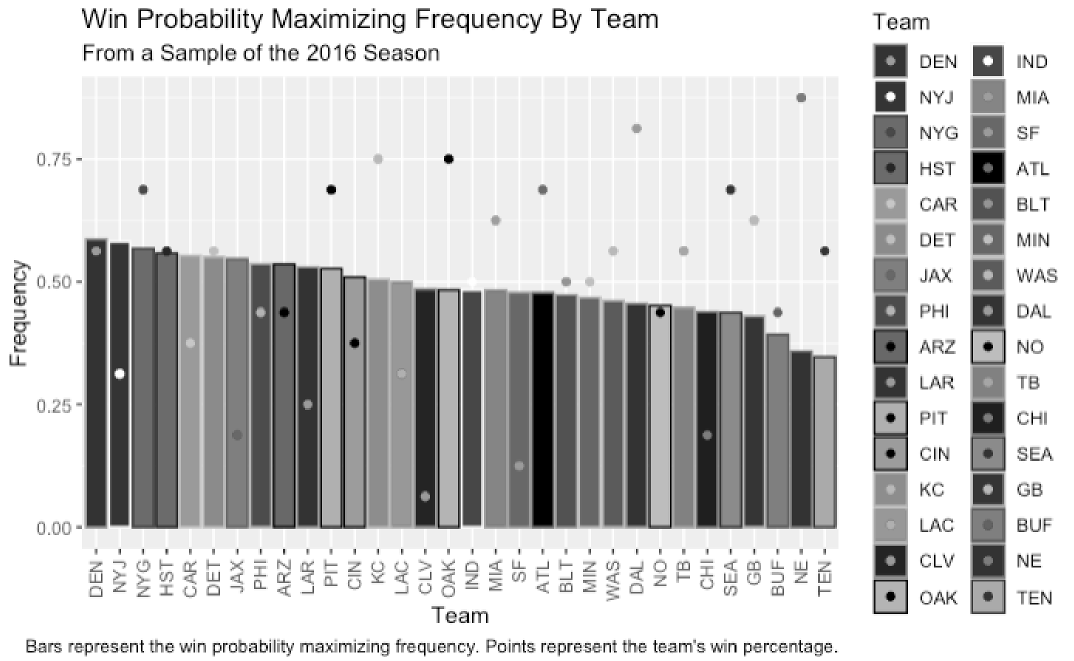

Figure 12: *Win Probability Maximizing Frequency & Win Frequency By Team For 2016 Season*

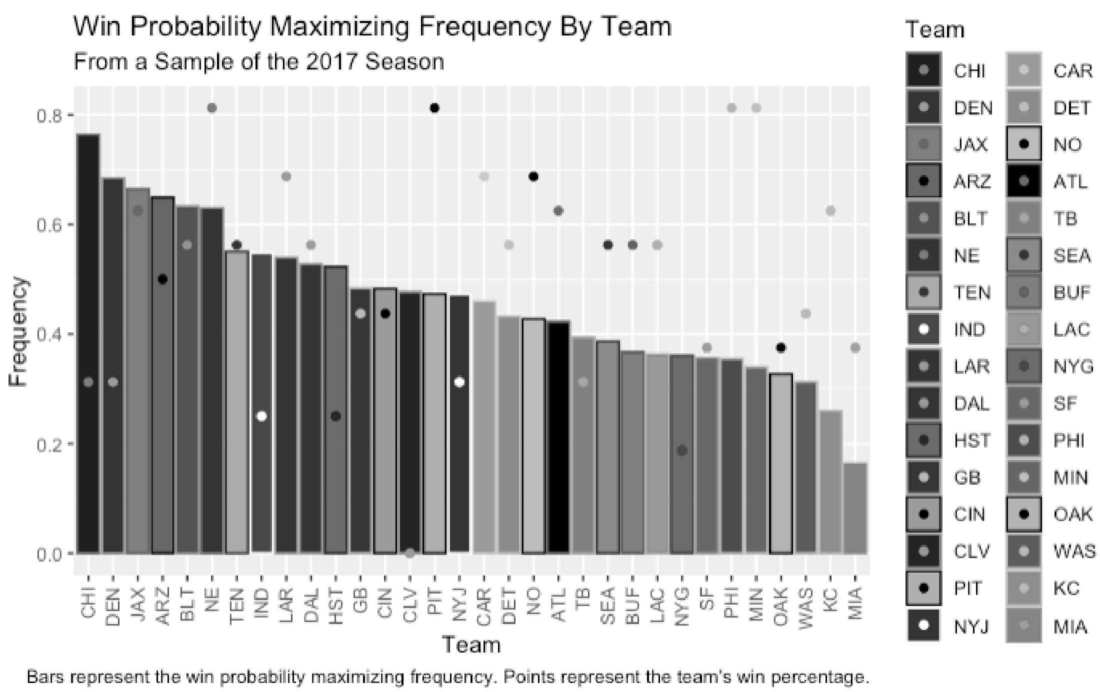

Figure 13: *Win Probability Maximizing Frequency & Win Frequency By Team For 2017 Season*

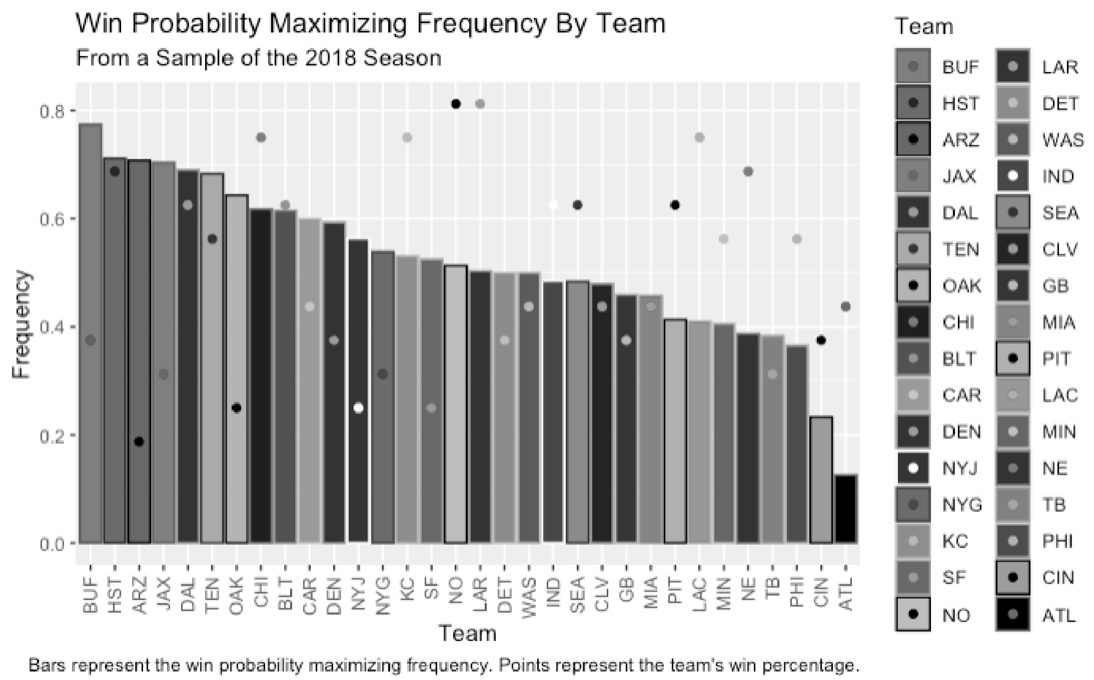

Figure 14: *Win Probability Maximizing Frequency & Win Frequency By Team For 2018 Season*

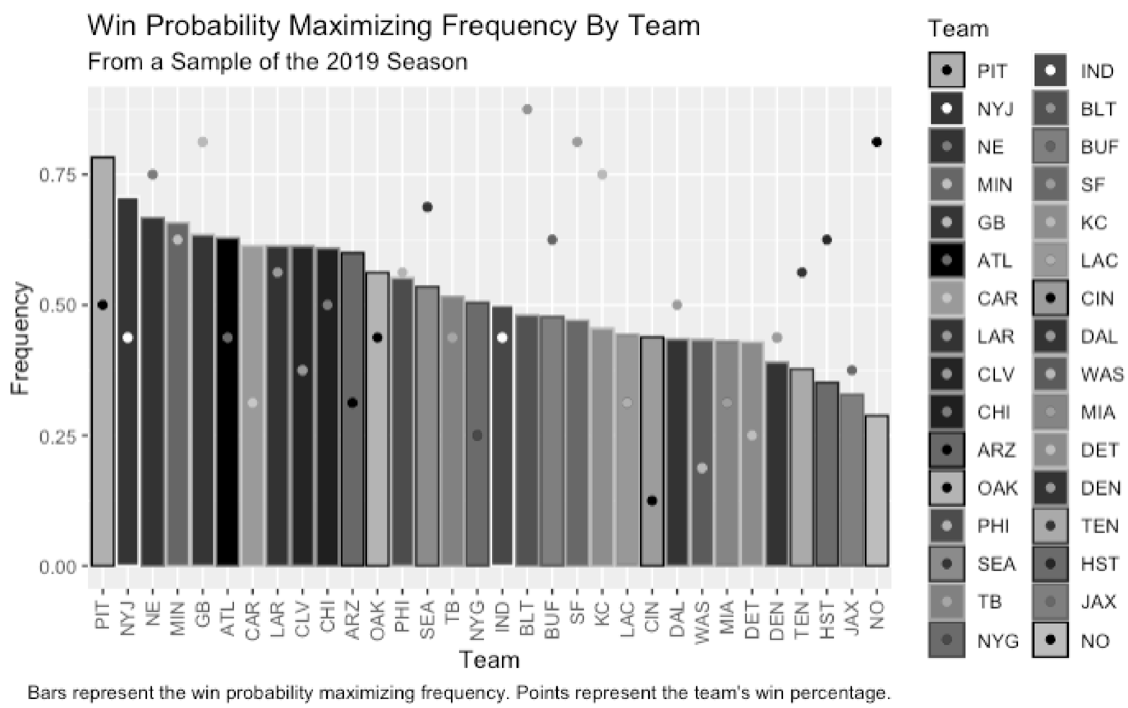

Figure 15: *Win Probability Maximizing Frequency & Win Frequency By Team For 2019 Season*

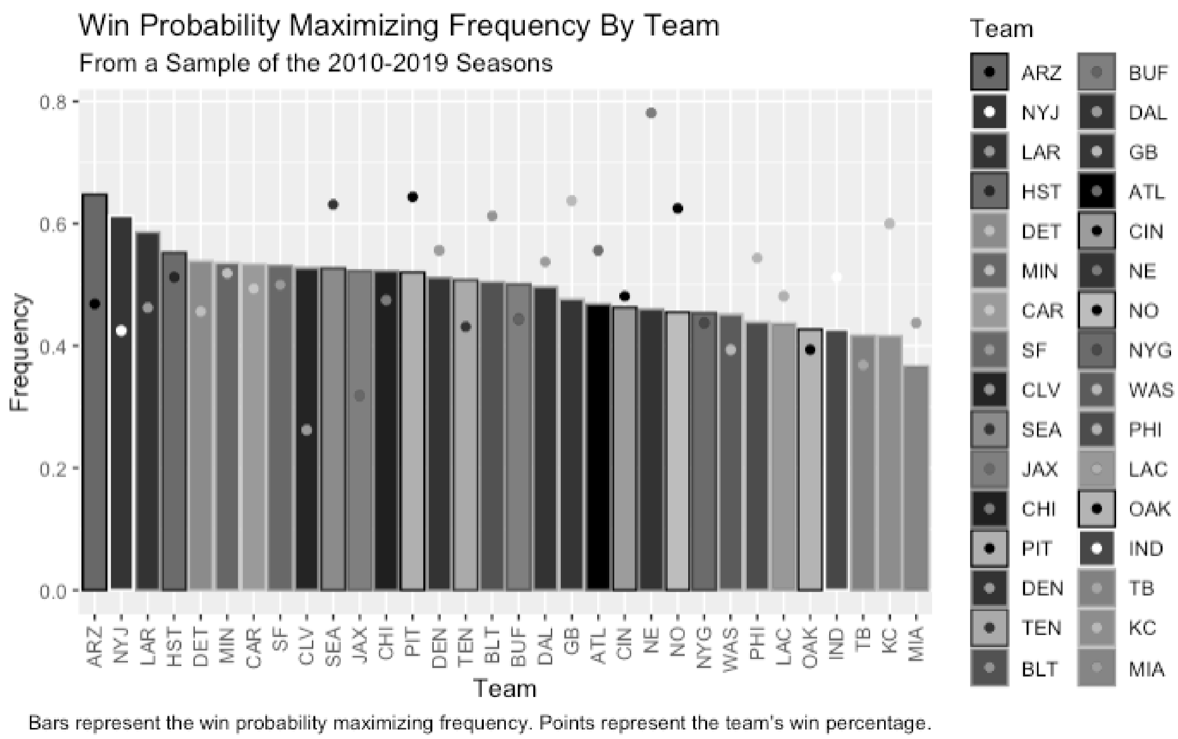

Figure 16: *Win Probability Maximizing Frequency & Win Frequency By Team For 2010 - 2019 Seasons*

Regression

The heart of this study is the regression analysis. First, the simple correlations will be calculated and then a Poisson regression analysis will be run for those relationships for which it is deemed necessary.

The relationship for total yards versus win probability maximizing frequency *on fourth down plays in this sample set* actually shows a negative correlation: -0.3732133. However, there are two outliers in total yards per season, which could be skewing this data. Bear in mind that the win probability maximizing frequency is only for fourth down plays in the sample set; this is not indicative of the relationship between offensive yards per game and wins per season. Through the data set, offensive yards per game and total wins on a season is positively correlated at 0.4992781. The relationship for yards per game and win probability maximizing frequency also shows a negative correlation: -0.3732133. This is to be expected as the total yards and yards per game number are collinear.

The correlation between win probability maximizing frequencies and wins is again negative: -0.2489986. When win probability maximizing frequency was regressed onto total wins per season, the win probability maximizing frequency variable had a P-Value of <0.001, making it statistically significant at the 95% confidence level. However, the Poisson Regression coefficient for win probability maximizing frequency is -0.36227, with a standard error of 0.08843, and a Z-Value of -4.097. Potential reasons for the negative correlations and the negative regression coefficient will be discussed further in the Discussion section.

Table 10: *Summary Statistics for Regression Models*

Variable	Minimum	Maximum	Mean	SD	Unique N
Win Maximizing Choice	15	216	65.4	36.4	113
Occurrences	44	343	131.6	57.3	151
win_max_frequency	0.1	0.8	0.5	0.1	301
Total Wins	0	15	8.0	3.1	16
win_freq	0.0	0.9	0.5	0.2	16
GP	16	16	16.0	0.0	1
YDS	3865	7474	5542.2	591.6	297
YDS.G	241.6	467.1	346.4	37.0	279
PTS	193	606	362.3	70.1	178
PTS.G	12.1	37.9	22.6	4.4	133

Table 11: *Poisson Regression Estimating Wins Per Season*

Effect	Estimate	SE	Z-Score	P-Value
Fixed Effects				
Intercept	3.29703	0.04448	74.131	< 2e-16
WPMF	-0.36227	0.08843	-4.097	< 2e-16

Note. N =352. AIC = Inf. WPMP = Win Probability Maximizing Frequency
Sample data was 10 years of all 32 National Football League Teams.

This means that the expected log count for wins in a season for a one-unit increase in win probability maximizing frequency is -0.36227. This means that we fail to reject the null research hypothesis of this study that a greater win probability maximizing frequency on fourth down plays leads to a greater win total per season. Again, further explanations for these results are discussed in the following Discussion section. The AIC – Akaike Information Criterion - of this regression is infinite. However, including more than one variable brings different results, and ultimately, a better-fit model.

A Poisson regression to estimate wins per season via win probability maximizing frequency, offensive pass rank, offensive rush rank, defensive pass rank, and defensive rush rank still brings about an AIC of infinity. In this model, the win probability maximizing frequency is not significant, with a P-Value of 0.10741, but offensive pass rank is significant with a P-Value of <0.001 and a Poisson regression coefficient of -0.0133539 (signifying that the higher a team is ranked, the more wins per season they will experience), a standard error of 0.0003806, and a Z-Score of -0.35091. Offensive rush rank is also significant, with a P-Value of <0.001 and a Poisson regression coefficient of -0.0056007, a standard error of 0.003482, and a Z-Score of -16.083. The defensive pass and rush ranks were both significant as well, with their regression coefficients, standard errors, Z-Scores, and P-Values listed in Table 8 below.

Table 12: *Poisson Regression Estimating Wins Per Season*

Effect	**Estimate**	**SE**	**Z-Score**	**P-Value**
Fixed Effects				
Intercept	6.2682247	0.0167232	374.823	< 2e-16
WPMF	-0.0441927	0.02745012	-1.610	0.10741
Offensive Pass Rank	-0.0133539	0.0003806	-0.35091	< 2e-16
Offensive Rush Rank	-0.0056007	0.0003482	-16.083	< 2e-16
Defensive Pass Rank	-0.0022096	0.0003657	-6.042	1.52e-09
Defensive Rush Rank	-0.0010041	0.0003361	-2.988	0.00281

Note. N = 352. AIC = Inf. WPMF = Win Probability Maximizing Frequency

However, the best fit model out of the sample data for this study was a Poisson regression to estimate total points over the course of a season via win probability maximizing frequency, offensive pass rank, and offensive rush rank – this model has an AIC of 4235.2. All three variables were significant at the 95% confidence level. Win probability maximizing frequency has a P-Value of 0.00918, a Poisson regression coefficient of 0.0613849, a standard error of 0.025616, and a Z-Score of 2.605. Offensive pass rank has a P-Value of <0.001, and a Poisson regression coefficient of -0.0140574, a standard error of 0.000367, and a Z-Score of -38.301. Finally, offensive rush rank has a P-Value of <0.001, a Poisson regression coefficient of -0.0054882, a standard error of 0.0003474, and a Z-Score of -15.799.

Table 13: *Poisson Regression Estimating Total Points Over the Course of a Season*

Effect	Estimate	SE	Z-Score	P-Value
Fixed Effects				
Intercept	6.173506	0.0114069	541.208	< 2e-16
WPMF	0.0613849	0.0235616	2.605	0.00918
Offensive Pass Rank	-0.0140574	0.0003670	-38.301	< 2e-16
Offensive Rush Rank	-0.00548820	0.0003474	-15.799	< 2e-16

Note. N =352. AIC = 4235.2. WPMF = Win Probability Maximizing Frequency

This model is saying that teams score more points over the course of a season when they are ranked higher in offensive passing and rushing categories, and when they make win probability maximizing decisions on fourth.

Taking the results from this third regression into consideration, the null hypothesis can be partially rejected. Making win probability maximizing decisions on fourth does increase the number of points scored, but that does not necessarily mean that teams will enjoy more wins over the course of the season. This could be because there is much more that goes into a football game than fourth down decisions. This will be discussed in more detail in the following Discussion section.

Chapter 5: Discussion

As previously stated, the null hypothesis of this study is as follows: "The frequency of which any given team in the National Football League makes the optimal choice on fourth down is not correlated with the number of points scored, and therefore has no relationship with the number of wins that team earns in a season." Therefore, the research hypothesis of this study is as follows: "The frequency of which any given team in the National Football League makes the optimal choice on fourth down is positively correlated with the number of points scored and the number of win that team earns per season."

Although this study did revolve around teams making the "optimal" decision – this decision is simply the choice in which the increase in win probability was maximized. The "optimal" decision in this study was calculated through the nearest neighbor matching algorithm, which paired up similar plays based upon a propensity score representing the likelihood of a team to go for it on fourth down, win probability (created through a random forest algorithm), distance needed for a first down and seconds remaining in the game. After plays were paired based upon these characteristics, the team that saw the greatest increase in win probability after their fourth down decision was rewarded a point, representing the "optimal" decision. The number of points a team earned throughout the dataset was summed, then divided by the number of occurrences of the team in the dataset – giving the win probability maximizing frequency. This "optimal" decision could vary across game scenarios and teams. It is not logical to assume that "going for it" is the best decision for every team in every scenario. The graphs in the Results section outline overall and by year the proportion of fourth downs in which each team made the "optimal" win probability maximizing decision. Ideally, this study assumes that teams want to maximize their win probability on every play. Again, the win probability maximizing decision *is not* to go for it on every fourth down play.

The medians of the "went for it" and "did not go for it" groups through the nearest neighbor matching algorithm were statistically significantly different - saying that in the same scenario, teams that did go for it on fourth down saw greater increases in their win probability, on average, than teams that did not. Although the correlations and first Poisson regression that attempted to estimate total wins on the season only through win probability maximizing frequency returned that a greater win probability maximizing frequency on fourth down led to fewer wins, when more variables were included, win probability maximizing frequency was statistically insignificant – this leads to inconclusive findings on the effect that win probability maximizing frequency has on total wins over the season. Intuitively, this does not seem to make sense; however, there are a couple of explanations for this. One explanation being that "better" teams that experience higher win totals are not in fourth down situations as often, and making the choice that does not maximize their win probability does not truly change the outcome of the game because of the coaching staff's talent, the players' skills, and the good decisions being made on first through third downs, as well as special teams plays. These teams might already be sitting at a high enough win probability that the coaching staff instead chooses to minimize their variance – also known as risk. This feeds back into Prospect Theory (Kahneman & Tversky, 1979), which states that "people underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty" (Kahneman & Tversky, 1979) - which is referred to as the certainty effect. Decision makers tend to be risk averse when sure gains are involved – for example, when a team's win probability is already high enough for the team to feel secure in their position, the coaching staff will tend to be more risk averse. On the contrary, decision makers tend to be risk seeking when choices involve sure losses (Kahneman & Tversky, 1979) – when a team's win probability is low enough that the team feels as if there is "nothing to lose," the coaching staff will be more willing to take what they perceive as a greater risk. Through their studies, Kahneman & Tversky (1979) point out that in scenarios where gains are possible, but not probable, decision makers tend to choose the option with the larger potential gain. To relate back to this study, this means that in scenarios in which a game could potentially be won, although it is not likely, the coaching staff is willing to make the decision that would bring about the largest gain. Teams that already have a high win probability perceive their gains as small compared to the risk. Figure 3 towards the beginning of this study illustrates this perfectly – while looking at the point differential density graphs between teams that did and did not go for it on fourth down, teams that went for it were, more often than not, trailing in the game. However, teams that did not go for it on fourth down were, more often than not, tied or ahead in the game.

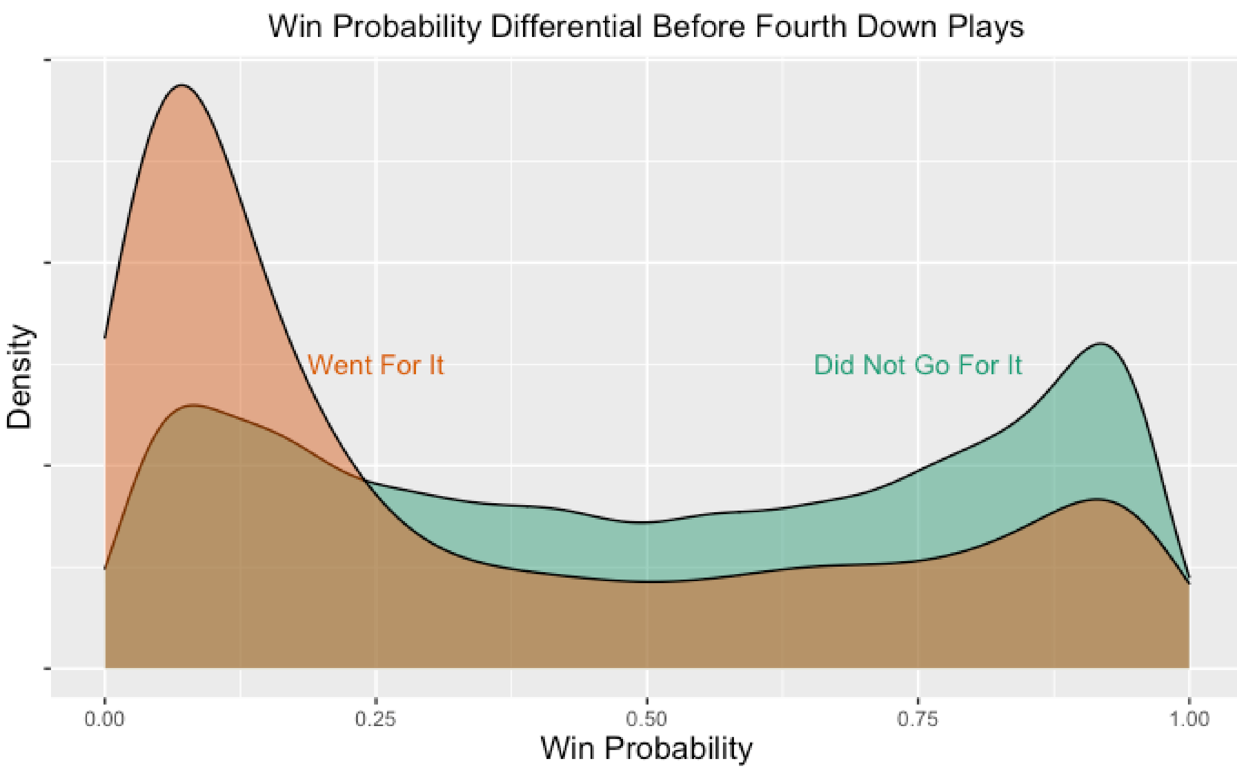

Figure 17: *Win Probability Differential Before Fourth Down Plays*

These reasons explain why win probability maximizing frequency is not positively correlated with winning football games – teams that are going for it are, more often than not, trailing in the game, and their win probability is already low. As can be seen in Figure 17 on the previous page, the densest area for the "went for it" group is when the team's win probability is below 25%.

Although the teams that are going for it are not necessarily winning games – they are scoring more points. This finding is in line with previous studies findings revolving around going for it on fourth down maximizing expected points added (Carter & Machol, 1978; Romer, 2006; Wright, 2007; Pelechrinis, 2016) – they really are adding more points! The Poisson regression that attempted to estimate total points over the season via win probability maximizing frequency, offensive pass rank, and offensive rush rank did return that all variables are statistically significant, and that a higher win probability maximizing frequency does lead to scoring more points. This finding, coupled with the first two regressions supports the notion that teams making the win probability maximizing decisions in this sample were scoring more points, but were most likely in desperate situations with already low win probabilities. Teams' behavior and actions on fourth downs are a continuation of teams' behaviors during every other play of a football game – a team playing poorly throughout a game is not going to win the game simply because they make the decision to go for it on fourth down, even if the fourth down decision works in that team's favor. Close games, however, do present an opportunity for teams to go for it more on fourth down – setting them on a trajectory to score more points – perhaps enough points to win a close game. This study reminds those of us in football analytics that there will always be more to this game than single decision scenarios, and that we must always be cognizant of the contextual scenarios in which we are running our analyses.

A limitation of this study to keep in mind, though, is that in this dataset, there are some teams that could have been included more or less, which could throw off their frequency; the range of occurrences in the dataset is 44 to 343. Obviously, the teams with the higher number of occurrences might be more reflective of their overall behavior. Also, teams that are talented teams with skilled players and intelligent coaching staffs might not see themselves in a fourth down scenario as often as other teams. Then again, some teams just might not have had many other plays that were similar to theirs through the matching process. These are all things to consider when reading these results.

Another aspect to consider that was not accounted for directly in this study is coaching turnover. Using team offensive and defensive ranks, both overall and for the pass and the rush, was an attempt to account for team strengths without accounting for individual players, and these ranks could also indirectly incorporate the coaching staff; however, there are a couple of factors other than coaching style and expertise that this does not consider. In chapter two of this study, game theory and risk were briefly discussed. Owens and Roach (2018) researched how collegiate football coaches make decisions when they are facing either a promotion or a potential firing. While this is not a study that includes NFL coaches, one can assume that these same tendencies could transfer. They found that coaches facing the possibility of a promotion took more risks, while coaches facing the possibility of being fired became more conservative (Owens and Roach, 2018). This study does not account for coaches, coaching styles, or their perceived job security – which plays a part in teams' past tendencies and will surely bleed into future tendencies, as well.

Areas for Future Research

Currently, NFL teams cannot have technology to assist them in decision-making in the booth during a game. However, it would be interesting to see that rule changed in the future. If teams were able to have technology in the booth, an idea for further research would be to make a dashboard that allows coaches to input certain game and play characteristics in order for the dashboard to provide the optimal, win probability maximizing decision as a result. A dynamic dashboard would be ideal for this, as the optimal decision varies by game scenario and the offensive and defensive teams' strengths. Another area for further research would be to include player tracking data in future analyses for fourth down decisions. Player tracking data obtains the position, speed, direction, and orientation of each player on the field, as well as the ball, ten times per second for each play. This information could further guide teams in their fourth down decision-making when it comes to strategizing. Another interesting area for future study would be to model a study more similar to previous expected points studies – such as estimating win probabilities at various points on the field under various game scenarios and ultimately determining which decision would bring about the greatest expected increase in win probability, also incorporating conversion success rates.

Conclusion

Although the regression analysis in terms of total wins was inconclusive, the regression analysis in terms of total points scored did support the hypothesis that maximizing win probability does lead to scoring more points. To supplement that finding, the Wilcoxon test showed that teams did, on average via the median, increase their win probability by a greater percentage by going for it on fourth down as opposed to not. This does not mean that the team that goes for it the most during a game will win that game, however, it will increase their win probability and has a good chance of increasing the number of points scored. Probabilities are just that – probabilities – and should be treated as such. As mentioned throughout the study, teams that make more optimal decisions on downs one, two, and three might not be in as dire or desperate fourth down situations in which their win probability has the potential to fluctuate. Also, these more skilled teams might not be in as many fourth down situations at all. The game of football is never certain. It is a competitive game, and as all football fans know, anything can happen. In a game of inches, why would teams not take every inch they can get? Maximizing win probability is not a guaranteed way to win more football games; however, it is a great place to start when it comes to scoring more points. Through this study, it has been shown that “going for it” on fourth down does, on average, increase teams’ win probability more than the alternative.

References

2019 NFL Stat Leaders. (n.d.). ESPN. Retrieved March 25, 2020, from https://www.espn.com/nfl/stats

2019 NFL TEAM EFFICIENCY RATINGS | Football Outsiders. (n.d.). Retrieved April 27, 2020, from https://www.footballoutsiders.com/stats/nfl/team-efficiency/2019

Albright, S. C. (1993). A Statistical Analysis of Hitting Streaks in Baseball. *Journal of the American Statistical Association*, *88*(424), 1175–1183. https://doi.org/10.2307/2291254

Big Hit: The NFL Turns to Data Analytics. (2020, January 7). Dell Technologies. https://www.delltechnologies.com/en-us/perspectives/big-hit-the-nfl-turns-to-data-analytics/

Burke, B., Carter, S., Daniel, J., Giratikanon, T., & Quealy, K. (2014, September 4). 4th Down: When to Go for It and Why. *The New York Times*. https://www.nytimes.com/2014/09/05/upshot/4th-down-when-to-go-for-it-and-why.html

Carter, V., & Machol, R. E. (1971). Technical Note—Operations Research on Football. *Operations Research*, *19*(2), 541–544. https://doi.org/10.1287/opre.19.2.541

Carter, V., & Machol, R. E. (1978). Note—Optimal Strategies on Fourth Down. *Management Science*, *24*(16), 1758–1762. https://doi.org/10.1287/mnsc.24.16.1758

Clark, K. (2018, December 19). *The NFL's Analytics Revolution Has Arrived*. The Ringer. https://www.theringer.com/nfl/2018/12/19/18148153/nfl-analytics-revolution

Critchfield, T. S., & Stilling, S. T. (2015). A matching law analysis of risk tolerance and gain–loss framing in football play selection. *Behavior Analysis: Research and Practice*, *15*(2), 112–121. https://doi.org/10.1037/bar0000011

Donges, N. (2019, June 16). *A complete guide to the random forest algorithm*. Built In. https://builtin.com/data-science/random-forest-algorithm

Football Outsiders Frequently Asked Questions | Football Outsiders. (n.d.). Retrieved November 11, 2020, from https://www.footballoutsiders.com/info/FAQ

Gilovich, T., Vallone, R., & Tversky, A. (1985). The Hot Hand in Basketball: On the Misperception of Random Sequences. *Cognitive Psychology*, *17*(295–314), 20.

Goldner, K. (2012). A Markov Model of Football: Using Stochastic Processes to Model a Football Drive. *Journal of Quantitative Analysis in Sports*, *8*(1). https://doi.org/10.1515/1559-0410.1400

Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *Journal of Economic Perspectives*, *14*(3), 75–94. https://doi.org/10.1257/jep.14.3.75

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Lock, D., & Nettleton, D. (2014). Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, *10*(2). https://doi.org/10.1515/jqas-2013-0100

Methods To Our Madness | Football Outsiders. (n.d.). Retrieved November 11, 2020, from https://www.footballoutsiders.com/info/methods

Neale, W. C. (1964). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *The Quarterly Journal of Economics*, *78*(1), 1–14. https://doi.org/10.2307/1880543

NFL Odds Super Bowl LIV. Super Bowl 54 NFL Vegas Odds for This Week. Free Pro Football Betting Odds. (n.d.). Retrieved April 27, 2020, from http://www.footballlocks.com/nfl_odds.shtml

Owens, M. F., & Roach, M. A. (2018). Decision-making on the hot seat and the short list: Evidence from college football fourth down decisions. *Journal of Economic Behavior & Organization*, *148*, 301–314. https://doi.org/10.1016/j.jebo.2018.02.023

Pelechrinis, K. (2016). *Decision Making in American Football: Evidence from 7 Years of NFL Data*. 10.

Peloso, S. (2015, April 5). Dallas Mavericks. *Databall*. https://nbastatsgeeks.wordpress.com/current-nba-teams-use-of-n-b-a-analytics/dallas-mavericks/

Pro Football Focus. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Pro_Football_Focus&oldid=943133861

Revolution or convention? Analyzing NFL coaches' fourth-down decisions in 2018. (n.d.). Retrieved March 5, 2020, from https://www.sportingnews.com/us/nfl/news/revolution-or-convention-analyzing-nfl-coaches-fourth-down-decisions-in-2018/1kyyi026urad31qwvitnbz2rnc

Romer, D. (2006). Do Firms Maximize? Evidence from Professional Football. *Journal of Political Economy*, *114*(2), 340–365. https://doi.org/10.1086/501171

Schalter, T. (2019, November 14). NFL Coaches Are Finally Getting More Aggressive On Fourth Down. *FiveThirtyEight*. https://fivethirtyeight.com/features/nfl-coaches-are-finally-getting-more-aggressive-on-fourth-down/

Sekhon, J. S. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. *Journal of Statistical Software*, *42*(1), 1–52. https://doi.org/10.18637/jss.v042.i07

Sports Info Solutions. (n.d.). Retrieved November 11, 2020, from http://www.sportsinfosolutions.com/

Steinberg, L. (2015, August 18). *CHANGING THE GAME: The Rise of Sports Analytics*. Forbes. https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/

Wright, J. A. (2007). *Fourth down decisions in NFL football: A statistical analysis* [Thesis, California State University, Northridge]. http://dspace.calstate.edu/handle/10211.3/158263

Yam, D. R., & Lopez, M. J. (2019). What was lost? A causal estimate of fourth down behavior in the National Football League. *Journal of Sports Analytics*, *5*(3), 153–167. https://doi.org/10.3233/JSA-190294

Appendix A

The following tables and graphs are supplementary to some of the intermediate steps within this study.

Starting with the Win Probability Model. Again, this model was built via a random forest algorithm. Random forest regression models make a *forest* of decision trees, and utilize bootstrap aggregation to train each tree on a different sample using sampling with replacement. Then multiple decision trees are used to determine the final output. Essentially, random forests "average" multiple deep decision trees, trained on different subset of the same training set, in an effort to avoid over-fitting - a common problem with an individual decision tree. While random forests will automatically evaluate their own performance by running each sample in the out-of-bag dataset through the forest, I still chose to cross-validate the model created for this study with the test data set. The table below outlines the importance of the variables in the model.

Table 14: *Importance of Variables in Win Probability Model*

Variable	% Increase Mean Square Error	Increase in Node Purity
Offensive_Rush_Rank	0.065564	4730.0332
Offensive_Pass_Rank	0.086849	7096.9943
Defensive_Rush_Rank	0.060731	4560.8651
Defensive_Pass_Rank	0.080425	5986.4278
score_diff	0.162618	24801.0470
seconds.left	0.018037	2279.8615
distance	0.002127	333.3064
down	0.001324	140.7873
yards_to_go	0.006039	805.5639
week	0.046325	3720.6457
Home	0.042410	1186.2317

% Increase Mean Square Error shows how much the model accuracy decreases when that variable is left out. Increase in Node Purity measures the variable importance based on the Gini impurity index.

When the model was tested on the test data set, plays with a predicted win probability greater than 0.5 were coded as an estimated win – and the model accurately predicted 76.16% of the plays in the test data set. Although this is not as sturdy of an accuracy test as binning by estimated probability and then compared with how many games in that data set were won, this is a secondary way to show the accuracy.

The following plots were created in route to the regression analysis.

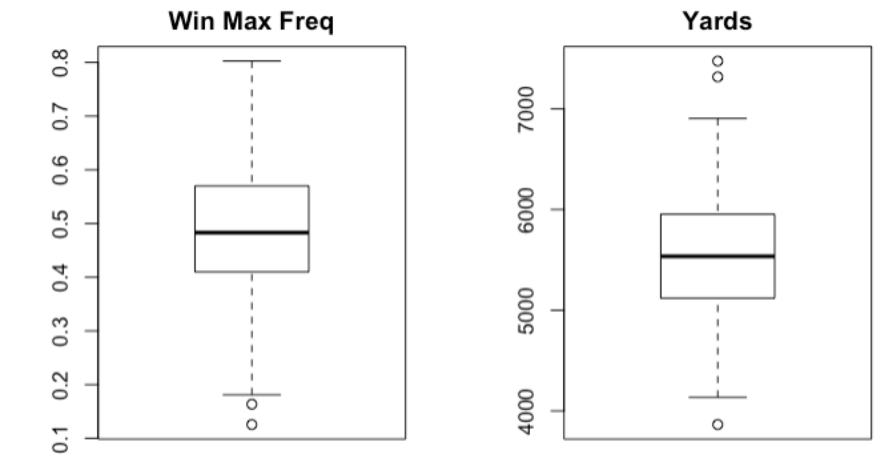

Figure 19: *Box Plot for Win Probability Maximizing Frequency & Total Yards Per Season*

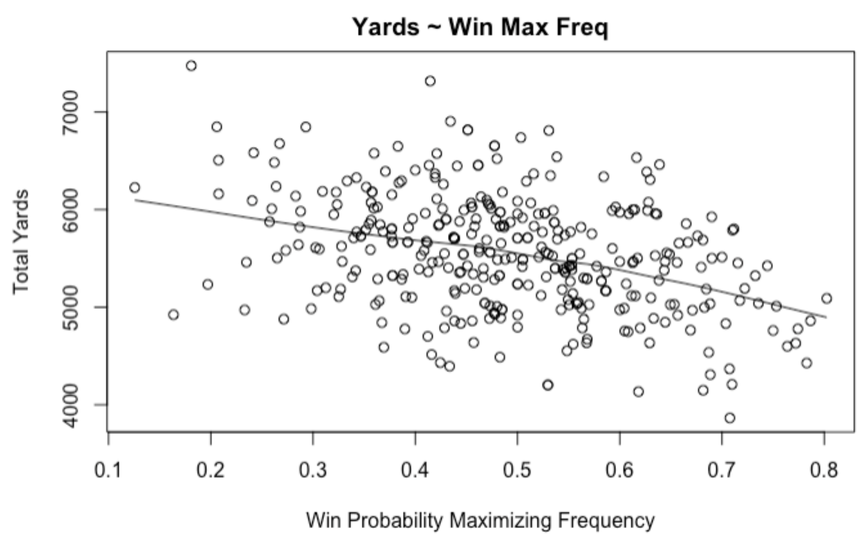

Figure 18: *Scatter Plot for Win Probability Maximizing Frequency & Total Yards Per Season*

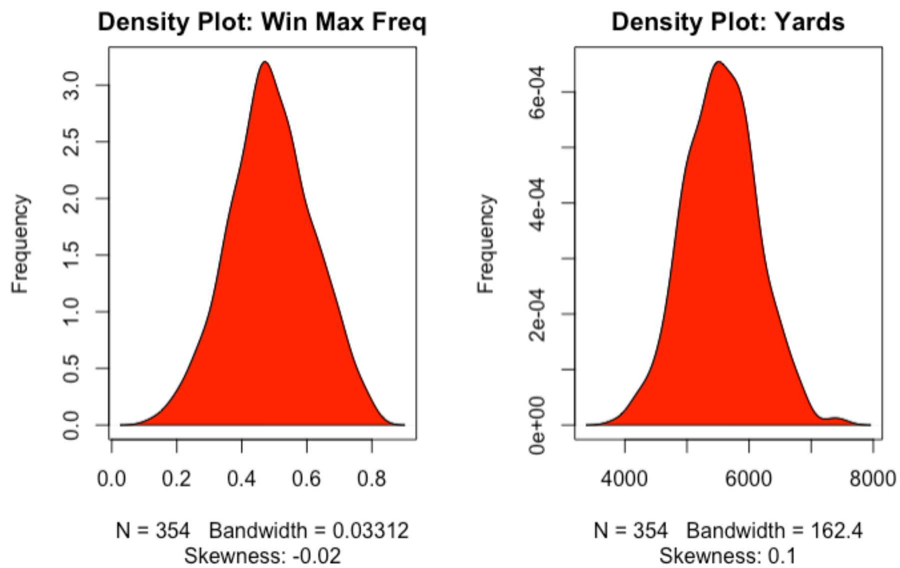

Figure 20: *Density Plot for Win Probability Maximizing Frequency & Total Yards*

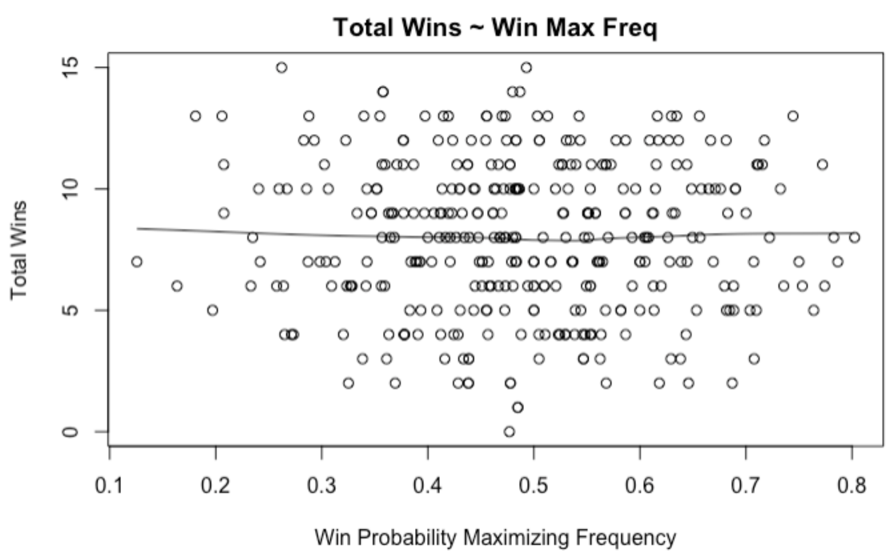

Figure 21*: Scatter Plot for Win Probability Maximizing Frequency and Total Wins Per Season*

Appendix B

The following code is what was used for this entire analysis.

````
``` {r, include = FALSE}
# Required Libraries
library(readr)
library(dplyr)
library(splines)
library(ggplot2)
library(randomForest)
library(mgcv)
library(gt)
library(tidyr)
library(RColorBrewer)
library(Matching)
library(gtools)
set.seed(100)
```
``` {r, include = FALSE}
# Read in Data
# Play by Play Data
plays <- read.csv("/Users/erinpsajdl/Desktop/Plays for Erin.csv")
games_weather <- read.csv("/Users/erinpsajdl/Desktop/Thesis
Papers/Football/games_weather.csv")
games <- read.csv("/Users/erinpsajdl/Desktop/Thesis Papers/Football/games.csv")
spreads <- read.csv("/Users/erinpsajdl/Downloads/archive/spreadspoke_scores.csv")
spreads <- spreads %>% filter(schedule_season > 2009 & schedule_season < 2020 &
schedule_playoff == "FALSE")
colnames(spreads)[5] <- "home_team"
colnames(spreads)[8] <- "away_team"
colnames(spreads)[2] <- "season"
colnames(spreads)[3] <- "week"
spreads$team_favorite_id <- as.character(spreads$team_favorite_id)
spreads$week <- as.integer(as.character(spreads$week))
spreads <- spreads %>% mutate(
  home_team = ifelse(home_team == "New Orleans Saints","NO",
                ifelse(home_team == "Buffalo Bills","BUF",
                     ifelse(home_team == "Chicago Bears","CHI",
                          ifelse(home_team == "Houston Texans","HST",
                               ifelse(home_team == "Jacksonville Jaguars","JAX",
                                    ifelse(home_team == "New England Patriots","NE",
                                         ifelse(home_team == "New York Giants","NYG",
                                              ifelse(home_team == "Philadelphia Eagles","PHI",
                                                   ifelse(home_team == "Pittsburgh Steelers","PIT",
````
````

```
                                                            ifelse(home_team == "Seattle
Seahawks","SEA",
                                                                ifelse(home_team == "St. Louis
Rams","LAR",
                                                                    ifelse(home_team == "Tampa Bay
Buccaneers", "TB",
                                                                        ifelse(home_team == "Tennessee
Titans","TEN",
                                                                            ifelse(home_team ==
"Washington Redskins","WAS",
                                                                                ifelse(home_team ==
"Kansas City Chiefs","KC",
                                                                                    ifelse(home_team ==
"New York Jets","NYJ",
ifelse(home_team== "Atlanta Falcons","ATL",
ifelse(home_team=="Carolina Panthers","CAR",
ifelse(home_team == "Cincinnati Bengals","CIN",
  ifelse(home_team == "Cleveland Browns","CLV",
      ifelse(home_team=="Dallas Cowboys","DAL",
          ifelse(home_team=="Denver Broncos","DEN",
              ifelse(home_team=="Detroit Lions","DET",
                  ifelse(home_team=="Green Bay Packers","GB",
                      ifelse(home_team=="Indianapolis Colts","IND",
                          ifelse(home_team=="Minnesota Vikings","MIN",
                              ifelse(home_team=="Oakland Raiders","OAK",
                                  ifelse(home_team=="San Diego Chargers","LAC",
                                      ifelse(home_team=="San Francisco 49ers","SF",
                                          ifelse(home_team=="Arizona Cardinals","ARZ",
                                              ifelse(home_team=="Baltimore
Ravens","BLT",
                                                  ifelse(home_team=="Miami
Dolphins","MIA",
                                                      ifelse(home_team=="Los Angeles
Rams","LAR",
                                                          ifelse(home_team=="Los Angeles
Chargers","LAC",home_team))))))))))))))))))))))))))))))))))
)
spreads <- spreads %>% mutate(
 away_team = ifelse(away_team == "New Orleans Saints","NO",
            ifelse(away_team == "Buffalo Bills","BUF",
                ifelse(away_team == "Chicago Bears","CHI",
                    ifelse(away_team == "Houston Texans","HST",
                        ifelse(away_team == "Jacksonville Jaguars","JAX",
                            ifelse(away_team == "New England Patriots","NE",
                                ifelse(away_team == "New York Giants","NYG",
                                    ifelse(away_team == "Philadelphia Eagles","PHI",
```

```
                                        ifelse(away_team == "Pittsburgh Steelers","PIT",
                                           ifelse(away_team == "Seattle Seahawks","SEA",
                                              ifelse(away_team == "St. Louis
Rams","LAR",
                                                 ifelse(away_team == "Tampa Bay
Buccaneers", "TB",
                                                    ifelse(away_team == "Tennessee
Titans","TEN",
                                                       ifelse(away_team ==
"Washington Redskins","WAS",
                                                          ifelse(away_team ==
"Kansas City Chiefs","KC",
                                                             ifelse(away_team ==
"New York Jets","NYJ",
ifelse(away_team== "Atlanta Falcons","ATL",
ifelse(away_team=="Carolina Panthers","CAR",
ifelse(away_team == "Cincinnati Bengals","CIN",
 ifelse(away_team == "Cleveland Browns","CLV",
    ifelse(away_team=="Dallas Cowboys","DAL",
       ifelse(away_team=="Denver Broncos","DEN",
          ifelse(away_team=="Detroit Lions","DET",
             ifelse(away_team=="Green Bay Packers","GB",
                ifelse(away_team=="Indianapolis Colts","IND",
                   ifelse(away_team=="Minnesota Vikings","MIN",
                      ifelse(away_team=="Oakland Raiders","OAK",
                         ifelse(away_team=="San Diego Chargers","LAC",
                            ifelse(away_team=="San Francisco 49ers","SF",
                               ifelse(away_team=="Arizona Cardinals","ARZ",
                                  ifelse(away_team=="Baltimore
Ravens","BLT",
                                     ifelse(away_team=="Miami
Dolphins","MIA",
                                        ifelse(away_team=="Los Angeles
Rams","LAR",
                                           ifelse(away_team=="Los Angeles
Chargers","LAC",away_team))))))))))))))))))))))))))))))))))
)
spreads <- spreads %>% mutate(
 team_favorite_id = ifelse(team_favorite_id == "ARI","ARZ",
            ifelse(team_favorite_id=="BAL","BLT",
               ifelse(team_favorite_id=="CLE","CLV",
                  ifelse(team_favorite_id=="LVR","OAK",
                     ifelse(team_favorite_id=="HOU","HST",team_favorite_id)))))
)
```

```
plays <- merge(plays,
spreads[,c("season","week","home_team","away_team","team_favorite_id","spread_favorite","o
ver_under_line")], by.y = c("season","week","home_team","away_team"),by.x =
c("season","week","home_team","away_team"), all.x =TRUE)


plays <- plays %>% mutate(
  spread_favorite = ifelse(offense==team_favorite_id, spread_favorite, spread_favorite*-1)
)
games <- games %>% filter(Season>=2010)

games_weather %>%
  fill(EstimatedCondition, .direction = "up")
# Pulling weather data from nflfastR package
library(nflfastR)
nflfastr_df <- nflfastR::fast_scraper_schedules(2010:2019, pp = FALSE)
nflfastr_df$old_game_id <- as.integer(nflfastr_df$old_game_id)
nflfastr_df <- merge(nflfastr_df, games_weather[!duplicated(games_weather$game_id),], by.x =
"old_game_id", by.y = "game_id", all.x = TRUE)

# Offensive Ratings
ratingswinprob_off <- read.csv("/Users/erinpsajdl/Downloads/2019 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off18 <- read.csv("/Users/erinpsajdl/Downloads/2018 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off17 <- read.csv("/Users/erinpsajdl/Downloads/2017 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off16 <- read.csv("/Users/erinpsajdl/Downloads/2016 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off15 <- read.csv("/Users/erinpsajdl/Downloads/2015 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off14 <- read.csv("/Users/erinpsajdl/Downloads/2014 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off13 <- read.csv("/Users/erinpsajdl/Downloads/2013 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off12 <- read.csv("/Users/erinpsajdl/Downloads/2012 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off11 <- read.csv("/Users/erinpsajdl/Downloads/2011 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)
ratingswinprob_off10 <- read.csv("/Users/erinpsajdl/Downloads/2010 Team DVOA Ratings
Offense.csv", stringsAsFactors = FALSE)

ratingswinprob_off$Season <- 2019
ratingswinprob_off <- ratingswinprob_off %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)
```

```
ratingswinprob_off18$Season <- 2018
ratingswinprob_off18 <- ratingswinprob_off18 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off17$Season <- 2017
ratingswinprob_off17 <- ratingswinprob_off17 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off16$Season <- 2016
ratingswinprob_off16 <- ratingswinprob_off16 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off15$Season <- 2015
ratingswinprob_off15 <- ratingswinprob_off15 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off14$Season <- 2014
ratingswinprob_off14 <- ratingswinprob_off14 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off13$Season <- 2013
ratingswinprob_off13 <- ratingswinprob_off13 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off12$Season <- 2012
ratingswinprob_off12 <- ratingswinprob_off12 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off11$Season <- 2011
ratingswinprob_off11 <- ratingswinprob_off11 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off10$Season <- 2010
ratingswinprob_off10 <- ratingswinprob_off10 %>% dplyr::select(Team,Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_off <- rbind(ratingswinprob_off,ratingswinprob_off18,ratingswinprob_off17,
ratingswinprob_off16, ratingswinprob_off15, ratingswinprob_off14, ratingswinprob_off13,
ratingswinprob_off12, ratingswinprob_off11, ratingswinprob_off10)

ratingswinprob_off18 <- NULL
ratingswinprob_off17 <- NULL
ratingswinprob_off16 <- NULL
ratingswinprob_off15 <- NULL
ratingswinprob_off14 <- NULL
```

```
ratingswinprob_off13 <- NULL
ratingswinprob_off12 <- NULL
ratingswinprob_off11 <- NULL
ratingswinprob_off10 <- NULL

# Defensive Ratings
ratingswinprob_def <- read.csv("/Users/erinpsajdl/Downloads/2019 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def18 <- read.csv("/Users/erinpsajdl/Downloads/2018 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def17 <- read.csv("/Users/erinpsajdl/Downloads/2017 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def16 <- read.csv("/Users/erinpsajdl/Downloads/2016 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def15 <- read.csv("/Users/erinpsajdl/Downloads/2015 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def14 <- read.csv("/Users/erinpsajdl/Downloads/2014 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def13 <- read.csv("/Users/erinpsajdl/Downloads/2013 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def12 <- read.csv("/Users/erinpsajdl/Downloads/2012 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def11 <- read.csv("/Users/erinpsajdl/Downloads/2011 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)
ratingswinprob_def10 <- read.csv("/Users/erinpsajdl/Downloads/2010 Team DVOA Ratings
Defense.csv", stringsAsFactors = FALSE)

ratingswinprob_def$Season <- 2019
ratingswinprob_def <- ratingswinprob_def %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def18$Season <- 2018
ratingswinprob_def18 <- ratingswinprob_def18 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def17$Season <- 2017
ratingswinprob_def17 <- ratingswinprob_def17 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def16$Season <- 2016
ratingswinprob_def16 <- ratingswinprob_def16 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def15$Season <- 2015
ratingswinprob_def15 <- ratingswinprob_def15 %>% dplyr::select(Team, Pass.DVOA.Rank,
Rush.DVOA.Rank, Total.DVOA.Rank, Season)
```

ratingswinprob_def14$Season <- 2014
ratingswinprob_def14 <- ratingswinprob_def14 %>% dplyr::select(Team, Pass.DVOA.Rank, Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def13$Season <- 2013
ratingswinprob_def13 <- ratingswinprob_def13 %>% dplyr::select(Team, Pass.DVOA.Rank, Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def12$Season <- 2012
ratingswinprob_def12 <- ratingswinprob_def12 %>% dplyr::select(Team, Pass.DVOA.Rank, Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def11$Season <- 2011
ratingswinprob_def11 <- ratingswinprob_def11 %>% dplyr::select(Team, Pass.DVOA.Rank, Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def10$Season <- 2010
ratingswinprob_def10 <- ratingswinprob_def10 %>% dplyr::select(Team, Pass.DVOA.Rank, Rush.DVOA.Rank, Total.DVOA.Rank, Season)

ratingswinprob_def <- rbind(ratingswinprob_def,ratingswinprob_def18,ratingswinprob_def17, ratingswinprob_def16, ratingswinprob_def15, ratingswinprob_def14, ratingswinprob_def13, ratingswinprob_def12, ratingswinprob_def11, ratingswinprob_def10)

ratingswinprob_def18 <- NULL
ratingswinprob_def17 <- NULL
ratingswinprob_def16 <- NULL
ratingswinprob_def15 <- NULL
ratingswinprob_def14 <- NULL
ratingswinprob_def13 <- NULL
ratingswinprob_def12 <- NULL
ratingswinprob_def11 <- NULL
ratingswinprob_def10 <- NULL

ratingswinprob_off <- ratingswinprob_off %>% rename(Offensive_Pass_Rank = Pass.DVOA.Rank, Offensive_Rush_Rank = Rush.DVOA.Rank, Offensive_Rank = Total.DVOA.Rank)
ratingswinprob_def <- ratingswinprob_def %>% rename(Defensive_Pass_Rank = Pass.DVOA.Rank, Defensive_Rush_Rank = Rush.DVOA.Rank, Defensive_Rank = Total.DVOA.Rank)
```
``` {r, include = FALSE}
#Ensure names always line up
#The only franchises that changed were Rams to LA & Chargers to LA
ratingswinprob_def[ratingswinprob_def$Team == "STL",]$Team <- "LAR"

```
ratingswinprob_off[ratingswinprob_off$Team == "STL",]$Team <- "LAR"

ratingswinprob_def[ratingswinprob_def$Team == "SD",]$Team <- "LAC"
ratingswinprob_off[ratingswinprob_off$Team == "SD",]$Team <- "LAC"

levels(plays$home_team) <- c(levels(plays$home_team), "LAR","LAC")
levels(plays$away_team) <- c(levels(plays$away_team), "LAR","LAC")
levels(plays$offense) <- c(levels(plays$offense), "LAR","LAC")
levels(plays$defense) <- c(levels(plays$defense), "LAR","LAC")
levels(plays$next_play_offense) <- c(levels(plays$next_play_offense), "LAR","LAC")

plays[plays$home_team =="SL",]$home_team <- "LAR"
plays[plays$away_team=="SL",]$away_team <- "LAR"
plays[plays$offense=="SL",]$offense <- "LAR"
plays[plays$defense=="SL",]$defense <- "LAR"
plays[plays$next_play_offense=="SL",]$next_play_offense <- "LAR"

plays[plays$home_team=="SD",]$home_team <- "LAC"
plays[plays$away_team=="SD",]$away_team <- "LAC"
plays[plays$offense=="SD",]$offense <- "LAC"
plays[plays$defense=="SD",]$defense <- "LAC"
plays[plays$next_play_offense=="SD",]$next_play_offense <- "LAC"

# Houston is not abbreviated the same
ratingswinprob_def[ratingswinprob_def$Team == "HOU",]$Team <- "HST"
ratingswinprob_off[ratingswinprob_off$Team == "HOU",]$Team <- "HST"

# Baltimore is not abbreviated the same
ratingswinprob_def[ratingswinprob_def$Team == "BAL",]$Team <- "BLT"
ratingswinprob_off[ratingswinprob_off$Team == "BAL",]$Team <- "BLT"

# Arizona is not abbreviated the same
ratingswinprob_def[ratingswinprob_def$Team == "ARI",]$Team <- "ARZ"
ratingswinprob_off[ratingswinprob_off$Team == "ARI",]$Team <- "ARZ"

# Cleveland is not abbreviated the same
ratingswinprob_def[ratingswinprob_def$Team == "CLE",]$Team <- "CLV"
ratingswinprob_off[ratingswinprob_off$Team == "CLE",]$Team <- "CLV"

# LAR
plays[plays$home_team =="LA",]$home_team <- "LAR"
plays[plays$away_team=="LA",]$away_team <- "LAR"
plays[plays$offense=="LA",]$offense <- "LAR"
plays[plays$defense=="LA",]$defense <- "LAR"
plays[plays$next_play_offense=="LA",]$next_play_offense <- "LAR"
```

````
#Join data frames
plays <- left_join(plays, ratingswinprob_def, by = c("season"="Season","defense"="Team"))
plays <- left_join(plays, ratingswinprob_off, by = c("season"="Season","offense"="Team"))

#Create a win variable for the win probability models
  # If the Offensive Team Won = 1
plays <- plays %>%
  mutate(Home = ifelse(offense == home_team, 1, 0),
         won.home = ifelse(final_home_team_score > final_away_team_score, 1, 0),
         won.off = ifelse(Home == 1 & won.home == 1 |
                          Home == 0 & won.home == 0, 1, 0))

plays$won.home <- as.numeric(plays$won.home)
plays$won.off <- as.numeric(plays$won.off)
#Filter out observations we will not use in our win probability model or analyses.
plays <- plays %>%
  filter(quarter < 5) %>%
  filter(week < 18) %>%
  filter(no_play == 0)
plays$score_diff <- plays$off_score - plays$def_score
plays <- plays %>% mutate(
  seconds.left = 900*(4-quarter) + seconds_left_in_quarter)
```
``` {r}
#Set seed for random samples and random forests
set.seed(1)
#Define the games to be in the test data set
plays$GameID <-cumsum(!duplicated(plays[1:4]))
games <- plays[!duplicated(plays$GameID),]
# Test Set is 2 games per Week - based off of Yam & Lopez -- better than one full season for any
sort of season descrepencies

Test_Games <- games %>%
  group_by(season, week) %>%
  sample_n(2, replace = FALSE)

#Define the test data set for the win probability model
plays_test <- plays %>%
  filter(GameID %in% Test_Games$GameID)
##Training data set
plays_train <- plays %>%
  filter(!GameID %in% Test_Games$GameID)
#Check the dimensions to make sure we didn't lose any observations
dim(plays)
dim(plays_test)
dim(plays_train)
````

````
isTRUE(dim(plays)[1] == dim(plays_test)[1] + dim(plays_train)[1])
  # all good here

# Strip the training dataset of extra variables
## THIS IS WITHOUT POINT SPREAD -- USING RANKS
plays_train <- plays_train %>%
  dplyr::select(won.off, Offensive_Rush_Rank, Offensive_Pass_Rank, Defensive_Rush_Rank,
Defensive_Pass_Rank, score_diff, seconds.left,
                distance, down, yards_to_go, week, Home)
str(plays_train)
##Omit Observations with missing Values
plays_train <- na.omit(plays_train)
str(plays_train)
```
``` {r}
# Important Variables
varis <- c("Offensive Rank","Defensive Rank","Score Differential","Distance","Down","Yards
To Go","Seconds Left","Week of Season","Home Team")
defs <- c("The offense's offensive rank. For historical seasons, this was the rank at the end of the
season. A variable is included for both the rush and pass offense.",
          "The defense's defensive rank. For historical seasons, this was the rank at the end of the
season. A variable is included for both the rush and pass defense.",
          "The differential between the offensive team and defensive team's scores.",
          "The distance in yards needed for a first down.",
          "The down of the play.",
          "The yards to go to the endzone.",
          "The seconds remaining in regulation of the football game.",
          "The week of the regular season that the game is occurring in.",
          "A binary variable representing if the offense is the home team.")
definitions <- data.frame(varis,defs)
definitions_table <- definitions %>% gt %>% opt_row_striping() %>%
cols_label("varis"=md("**Variables**"), "defs"=md("**Definitions**")) %>%
  tab_style(
    style = list(
      cell_fill(color = "darkslategray3"),
      cell_text(color="white",size="18")),
    locations = cells_column_labels(c(1:2))) %>%
  tab_header(md("**Variable Definitions**")) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold", align = "right")),
    locations = cells_body(
      columns = vars(varis))) %>%
  tab_style(
    style = list(
      cell_text(align = "left")),
````

```
    locations = cells_body(
      columns = vars(defs)))
definitions_table
```

```` ``` {r, warning = FALSE} ````

```
#Note: this win probability model is trained strictly on our training data set.
winprobmodel <- randomForest(formula = won.off ~ ., data = plays_train, ntree = 500, mtry = 2,
nodesize = 200, importance = TRUE, do.trace = TRUE, type = "regression")
print(winprobmodel)
# I did try the model with rating quartiles as opposed to raw ranks, however, the MSE was up to
0.1452982 and the % of variance explained was down to 41.82
importance <- as.data.frame(as.matrix(winprobmodel$importance)) #Show variable importance
(mostly out of curiousity)
# Table showing Importance for Variables
importance %>% gt(rownames_to_stub = TRUE) %>%
  fmt_number(columns=c(1:2),decimals = 6) %>%
  opt_row_striping() %>%
  cols_label("%IncMSE"=md("**% Increase<br>Mean Square
Error**"),"IncNodePurity"=md("**Increase in<br>Node Purity**")) %>%
tab_stubhead("Variable") %>%

tab_style(style=list(cell_fill(color="darkslategray3"),cell_text(weight="bold",color="white",size
="18")),locations=cells_stubhead()) %>%
  tab_style(
    style = list(
      cell_fill(color = "darkslategray3"),
      cell_text(color="white",size="18")),
    locations = cells_column_labels(c(1:2))) %>% tab_header(md("**Importance of Variables in
Win Probability Model**"), subtitle = md("% Increase Mean Square Error is **Mean Decrease
Accuracy** and Increase in Node Purity is **Mean Decrease Gini**")) %>%
  tab_source_note(
    source_note = md("Mean Decrease Accuracy shows how much the model accuracy decreases
when that variable is left out. Mean Decrease Gini measures the variable importance based on
the Gini impurity index.")
  )

#Apply win probability for all plays
plays$wp_1 <- predict(winprobmodel, plays)
#Define the win probability for the home team
plays <- plays %>%
  mutate(wp_home_1 = ifelse(Home == 1, wp_1,  1 - wp_1))
#Create a variable for the change in win probability using the RF model.
# deltawp is the actual change in win probability from the play that occurred
# to look at the change for the opposite, at the end we will look at change if gone for it, change if
FG kicked, and change if punted
plays <- plays %>%
```
```

```
  group_by(GameID) %>% arrange(next_play_id) %>%
  mutate(wp_home_lead_1 = lead(as.numeric(wp_home_1), 1)) %>%
  mutate(deltawp_home_1 = wp_home_lead_1 - wp_home_1)
plays <- plays %>%
  mutate(wp_off_lead_1 = ifelse(Home == 1, wp_home_lead_1, 1 - wp_home_lead_1)) %>%
  mutate(deltawp_off_1 = wp_off_lead_1 - wp_1)
```

```
``` {r}
# Accuracy of Win Probability Model
# Predict Win Prob on test data
plays_test$wp_1 <- predict(winprobmodel, plays_test)
plays_test <- plays_test %>% mutate(
  model_estimate = ifelse(wp_1 > 0.5, 1, 0),
  model_correct = ifelse(model_estimate == won.off, 1,0)
)
mean(plays_test$model_correct)
# The model accurate predicted (if win prob was > 50%, it counts as a win) 76.16% of the plays
in the test data set.

##Create a data frame for the WP accuracy plot
plot <- plays_test %>%
  dplyr::select(GameID, next_play_id, wp_1, won.off) %>%
  mutate(wp.cat = cut_number(wp_1, 20))
wp.probs <- plot %>%
  group_by(wp.cat) %>%
  summarise(estimated = mean(wp_1), observed = mean(won.off), type = "Lock")
p.wpacc <- ggplot(wp.probs, aes(x = estimated, y = observed,
                              shape = type, colour = type))
wpacc <- p.wpacc +
  geom_point(size = 2.5) +
  geom_abline(intercept = 0, slope = 1, color = "red") +
  scale_x_continuous(labels = scales::percent, "Estimated win rate") +
  scale_y_continuous(labels = scales::percent, "Observed win rate") +
  scale_colour_manual(values = c("black", "red"),
                                  name  ="Win Probability Model",
                            breaks=c("Lock", "nflscrapR"),
                            labels=c("Lock", "nflscrapR")) +
  scale_shape_manual(values = c(19, 3), name  ="Win Probability Model",
                           breaks=c("Lock", "nflscrapR"),
                            labels=c("Lock", "nflscrapR")) +
  ylab("Proportion of Games Won") +
  ggtitle("Accuracy of the Win Probability Models") +
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        legend.position = 'none',
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2)))
```
```

~~~~
wpacc
```
```{r}
# Create Point Differential Metric
plays <- plays %>% mutate(
  point_diff = off_score - def_score
)
# Create Factor To Signify if team goes for it on Fourth Down
plays <- plays %>% mutate(
  Attempt_Fourth = ifelse(down == 4 & play_type == "R",1,0)
)


prop.df <- plays %>%
  group_by(season) %>% filter(down==4) %>%
  summarise(propatt = mean(Attempt_Fourth))
propatt.gg <- ggplot(prop.df, aes(season, propatt)) + geom_point(size = 3) +
  ggtitle("Proportion of Fourth Downs Attempted") +
  xlab("Season") + ylab("Proportion Attempted") +
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2))) + scale_x_continuous(breaks=c(2010:2019)) +
scale_y_continuous(limits = c(0,.5))
propatt.gg
```
``` {r, warning = FALSE}
## Quick Look at Success Rates
    # 4th Down ONLY INCLUDED IF IT WAS NOT A PUNT OR FIELD GOAL
oneytg <- plays %>% filter(distance == 1 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
oneytg <- aggregate(oneytg[,59], list(oneytg$down), mean)
oneytg$yards <- "One"

twoytg <- plays %>% filter(distance == 2 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
twoytg <- aggregate(twoytg[,59], list(twoytg$down), mean)
twoytg$yards <- "Two"

threeytg <- plays %>% filter(distance == 3 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
threeytg <- aggregate(threeytg[,59], list(threeytg$down), mean)
threeytg$yards <- "Three"
~~~~

```
fourytg <- plays %>% filter(distance == 4 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
fourytg <- aggregate(fourytg[,59], list(fourytg$down), mean)
fourytg$yards <- "Four"

fiveytg <- plays %>% filter(distance == 5 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
fiveytg <- aggregate(fiveytg[,59], list(fiveytg$down), mean)
fiveytg$yards <- "Five"

sixytg <- plays %>% filter(distance == 6 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
sixytg <- aggregate(sixytg[,59], list(sixytg$down), mean)
sixytg$yards <- "Six"

sevenytg <- plays %>% filter(distance == 7 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
sevenytg <- aggregate(sevenytg[,59], list(sevenytg$down), mean)
sevenytg$yards <- "Seven"

eightytg <- plays %>% filter(distance == 8 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
eightytg <- aggregate(eightytg[,59], list(eightytg$down), mean)
eightytg$yards <- "Eight"

nineytg <- plays %>% filter(distance == 9 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
nineytg <- aggregate(nineytg[,59], list(nineytg$down), mean)
nineytg$yards <- "Nine"

tenytg <- plays %>% filter(distance == 10 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
tenytg <- aggregate(tenytg[,59], list(tenytg$down), mean)
tenytg$yards <- "Ten"

longytg <- plays %>% filter(distance > 10 & play_type == "R") %>% mutate(
  convert = ifelse(next_play_offense == offense & next_play_down == 1 & next_play_distance
== 10, 1, 0))
longytg <- aggregate(longytg[,59], list(longytg$down), mean)
```

```
longytg$yards <- "Long"

convertsuccess <-
rbind(oneytg,twoytg,threeytg,fourytg,fiveytg,sixytg,sevenytg,eightytg,nineytg,tenytg,longytg)

# GT Table For Conversion Rates
convertsuccess %>% gt %>%
  fmt_percent(columns = 2, decimals =2) %>%
  opt_row_striping() %>%
  cols_label("Group.1"=md("**Down**"), "convert"=md("**Conversion<br>Success**")) %>%
  tab_style(style=list(
    cell_fill(color="darkslategray3"),
    cell_text(weight="bold",color="white",size = 18)),
    locations=cells_column_labels(c(1:2))) %>% tab_header(md("**Conversion Success Rates
By Down and Distance**")) %>%
  tab_row_group(group="More Than Ten Yards To Go", rows=c(41:44)) %>%
  tab_row_group(group= "Ten Yards To Go",rows=c(37:40)) %>%
  tab_row_group(group="Nine Yards To Go", rows = c(33:36)) %>%
  tab_row_group(group="Eight Yards To Go", rows=c(29:32)) %>%
  tab_row_group(group="Seven Yards To Go",rows=c(25:28)) %>%
  tab_row_group(group="Six Yards To Go",rows=c(21:24)) %>%
  tab_row_group(group="Five Yards To Go",rows=c(17:20)) %>%
  tab_row_group(group="Four Yards To Go",rows=c(13:16)) %>%
  tab_row_group(group="Three Yards To Go",rows=c(9:12)) %>%
  tab_row_group(group = "Two Yards To Go",rows=c(5:8)) %>%
  tab_row_group(group="One Yard To Go",rows = c(1:4))  %>%
  cols_hide(
    columns = vars(yards)) %>%
  tab_style(
    style = list(
      cell_fill(color = "lightcyan"),
      cell_text(weight = "bold",size="18",align = "center")
    ),
    locations = cells_row_groups())

colnames(convertsuccess)[1] <- "Down"
convertsuccess$yards <- as.factor(convertsuccess$yards)
convertsuccess$yards <- factor(convertsuccess$yards, levels =
c("One","Two","Three","Four","Five","Six","Seven","Eight","Nine","Ten","Long"))
# Chart For Conversion Rates
convertsuccess %>% ggplot(aes(x=Down, y = convert, color = as.factor(yards))) + geom_line()
+ scale_color_discrete(name="Distance") +
  ggtitle("Conversion Rates By Down and Distance", subtitle = "All teams, all relevant plays") +
labs(y="Conversion Rate", color = "Distance") + theme(title = element_text(face="bold"))
```

```` ``` ````

```` ``` {r, include = FALSE} ````

```
# This is a frame of team/year fourth down conversion success rates
frame <- plays %>% filter(down == 3|down == 4)
frame <- frame%>% filter(play_type == "R")
frame <- frame %>% dplyr::select(distance, play_type, first_down, Defensive_Rank,
Offensive_Rank)
frame <- frame %>% mutate(
  distance = ifelse(distance>10,11,distance)
)
# Distance of 11 is anything greater than 10 yards
frame <- frame %>% group_by(Defensive_Rank, Offensive_Rank, distance) %>%
summarise(Conversion_Rate = mean(first_down))
# This is a frame of the conversion success rate for 3rd/4th downs based upon the defensive rank,
offensive rank, and distance
plays <- plays %>% mutate(likelihood_distance = ifelse(distance>10,11,distance))
frame <- left_join(plays, frame, by =
c("Defensive_Rank","Offensive_Rank","likelihood_distance"="distance")) %>%
filter(distance!=0)
```

```` {r}
## Point differential imbalance
pd <- ggplot(plays %>% filter(down == 4), aes(x = point_diff)) +
  geom_density(position = "identity", alpha = 0.5, aes(fill = as.factor(Attempt_Fourth))) +
  scale_fill_brewer(type = "qual", palette = "Dark2") +
  ggtitle("Point Differential Before Fourth Down Plays") +
  xlab("Point Differential") + ylab("Density") +
  xlim(-40, 35) +
  annotate("text", -23, .025, label = "Went For It", colour = "#d95f02", alpha = 1, size = 4) +
  annotate("text", 12, .05, label = "Did Not Go For It", colour = "#1b9e77", alpha = 1, size = 4) +
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        legend.position = "none",
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2)),
        axis.text.y = element_blank())
pd
```
```` {r}
# Change Attempt_Fourth To Binary
colnames(plays)[37:38] <- c("spread","over_under")
fourthplays <- plays %>% filter(down==4)
ps.L1 <- glm(Attempt_Fourth ~ ns(yards_to_go, 10) * ns(distance, 5) + ns(yards_to_go, 10) *
ns(seconds.left, 4) + ns(yards_to_go, 10) * score_diff + ns(distance, 5) * ns(seconds.left, 4) +
ns(distance, 5) * score_diff + ns(seconds.left, 4)* score_diff + ns(yards_to_go, 10) + ns(distance,
5) + ns(seconds.left, 4) + score_diff + ns(Offensive_Rush_Rank,5) + ns(Offensive_Pass_Rank,5)
+ ns(Defensive_Rush_Rank, 5) + ns(Defensive_Pass_Rank,5) + ns(week, 4) + Home + ns(wp_1,
10), data = fourthplays, family = "binomial")
fourthplays$predict.fitms <- predict(ps.L1, fourthplays, type = "response")
````

````
```
```{r, warning = FALSE}
## Filter for the common support interval
limits <- fourthplays %>%
  group_by(Attempt_Fourth) %>%
  summarise(min.score = min(predict.fitms, na.rm = TRUE), max.score = max(predict.fitms,
na.rm = TRUE))
low.bound <- min(limits$max.score)
upper.bound <- max(limits$min.score)
plays.common.support <- fourthplays %>%
  filter(predict.fitms <= low.bound, predict.fitms >= upper.bound)
```
```{r}
## Recalculate propensity scores on common support
ps.L2 <- glm(Attempt_Fourth ~ ns(yards_to_go, 10) * ns(distance, 5) + ns(yards_to_go, 10) *
ns(seconds.left, 4) + ns(yards_to_go, 10) * score_diff + ns(distance, 5) * ns(seconds.left, 4) +
ns(distance, 5) * score_diff + ns(seconds.left, 4)* score_diff + ns(yards_to_go, 10) + ns(distance,
5) + ns(seconds.left, 4) + score_diff + ns(Offensive_Rank,5) + ns(Defensive_Rank, 5) +
ns(week, 4) + Home + ns(wp_1, 10), data = plays.common.support, family = "binomial")
plays.common.support$predict.fitms2 <- predict(ps.L2, plays.common.support, type =
"response")
```
```{r}
plays.common.support %>%
  summarise(mean = mean(seconds.left), sd = sd(seconds.left)) %>%
  summarise(caliper = 450/sd) #Calculate the caliper size for plays within half of a quarter (450
seconds)
#Create the covariate matching matrix
psm <-plays.common.support %>% dplyr::select(predict.fitms2, wp_1, distance, seconds.left)
X <- psm[,2:5]
X[,1:2] <- gtools::logit(X[,1:2])
match.atc <- Match(Y = NULL,
            estimand = "ATC",
            Tr = as.numeric(plays.common.support$Attempt_Fourth),
            M = 1,
            ties = FALSE,
            X = X,
            calip=c(0.5, 0.5, 0, .432),
            replace=T, Weight = 1)
summary(match.atc)
pairs.atc <- cbind(match.atc$index.treated, match.atc$index.control)
dim(pairs.atc)
## treated and control data sets
treated.atc <- plays.common.support[match.atc$index.treated,]
control.atc <- plays.common.support[match.atc$index.control,]
#Define the paired id which is the row number from NFL.GFI dataset
````

```
treated.atc <- cbind(treated.atc, pairs.atc[,1])
colnames(treated.atc)[colnames(treated.atc)== "pairs.atc[, 2]"] <- "paired.id"
control.atc <- cbind(control.atc, pairs.atc[,1])
colnames(control.atc)[colnames(control.atc)== "pairs.atc[, 1]"] <- "paired.id"
#Define the matched pair for the team plot
control.atc$off.pair <- control.atc$off
treated.atc$off.pair <- control.atc$off
matched.subset.atc <- rbind(control.atc, treated.atc)
```

```{r}
#Review of univariate balance
pd2 <- ggplot(matched.subset.atc, aes(x = score_diff)) +
  geom_density(position = "identity", alpha = 0.5, aes(fill = as.factor(Attempt_Fourth))) +
  scale_fill_brewer(type = "qual", palette = "Dark2") +
  ggtitle("Point Differential Before Fourth Down Plays, After Matching") +
  xlab("Point Differential") + ylab("Density") +
  theme(legend.position = "none") +
  xlim(-40, 35) +
  annotate("text", -21, .025, label = "Went For It", colour = "#d95f02", alpha = 1, size = 4) +
  annotate("text", 10, .07, label = "Did Not Go For It", colour = "#1b9e77", alpha = 1, size = 4) +
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        legend.position = "none",
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2)),
        axis.text.y = element_blank())
pd2
```

```{r, include = FALSE}
games <- plays.common.support
##Create our outcome variable
matched.subset.atc$pair.id <- rep(1:(nrow(matched.subset.atc)/2), 2)
```

```{r}
sum.matched <- matched.subset.atc %>%
  group_by(Attempt_Fourth) %>%
  summarise(mean.delta = mean(deltawp_off_1, na.rm = TRUE), med.delta =
median(deltawp_off_1, na.rm = TRUE), n.plays = n())
sum.matched <- sum.matched %>% mutate(
  Attempt_Fourth = ifelse(Attempt_Fourth == 1, "Yes","No")
)

win_prob_table <- sum.matched %>% gt %>% fmt_percent(columns = c(2:3),decimals = 4)
%>% cols_label("Attempt_Fourth"=md("**Attempt<br>Fourth**"),"mean.delta"=md("**Mean
Change in<br>Win Probability**"),"med.delta"=md("**Median Change in<br>Win
Probability**"),"n.plays"=md("**Number of Plays<br>in Sample**")) %>% tab_style(
  style = list(cell_fill(color = "darkslategray3"),
```

```
              cell_text(color="white",size=18)),
  locations = cells_column_labels(c(1:4))) %>% tab_header(md("**Change in Win
Probability**")) %>%
  tab_style(
    style = list(
      cell_text(weight = "bold", align = "right")),
    locations = cells_body(
      columns = vars(Attempt_Fourth)))

win_prob_table
```

```{r}
wilcox.test(deltawp_off_1 ~ Attempt_Fourth, matched.subset.atc)
#P-Value of less than 0.05 - the means are significantly different
```

```{r}
pd <- ggplot(matched.subset.atc, aes(x = deltawp_off_1)) +
  geom_density(position = "identity", alpha = 0.5, aes(fill = as.factor(Attempt_Fourth))) +
  scale_fill_brewer(type = "qual", palette = "Dark2") +
  ggtitle("Change in Win Probability After Fourth Down Play") +
  xlab("Change in Win Probability") + ylab("Density") +
  theme(legend.position = "none") +
  xlim(-.25, .25) +
  annotate("text", -.13, 4, label = "Did Not Go For It", colour = "#1b9e77", alpha = 1, size = 4) +
  annotate("text", .07, 10, label = "Went For It", colour = "#d95f02", alpha = 1, size = 4)  +
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2)))

pd
```

``` {r}
# Group By Pair ID & Add Column To Represent the Max Delta Win Probability Per Choice
max.matched <- matched.subset.atc %>% group_by(pair.id) %>% summarise(max =
max(deltawp_off_1))
colnames(max.matched)[2] <- "max_delta_wp"
matched.subset.atc <- left_join(matched.subset.atc, max.matched, by = "pair.id")
matched.subset.atc <- matched.subset.atc %>% mutate(
  correct_choice = ifelse(deltawp_off_1==max_delta_wp, 1, 0)
)

correctchoices_df <- matched.subset.atc %>% filter(correct_choice==1)
correctchoices_1 <- table(correctchoices_df$season, correctchoices_df$offense)
library(reshape2)
correctchoices_1 <- melt(correctchoices_1)
colnames(correctchoices_1)[1:3] <- c("Season","Team","Win Maximizing Choice")
```

~~~~
occurence<-table(matched.subset.atc$offense,matched.subset.atc$season)
occurence<-melt(occurence)
colnames(occurence)[1:3] <- c("Team","Season","Occurences")
correctchoices <-merge(correctchoices_1,occurence,by =
c("Season"="Season","Team"="Team"), all.x = TRUE)
correctchoices <- correctchoices %>% mutate(
  win_max_frequency = `Win Maximizing Choice`/Occurences
)
```
```{r}
### Merge In Total Yards, Yards Per Game, Total Wins, and Wins Per Season to correct choices
data frame
total_df <- read.csv("/Users/erinpsajdl/Desktop/NFLTOTALS.csv")
no_dupes <- plays[!duplicated(plays$GameID),]
no_dupes$home_team <- as.character(no_dupes$home_team)
no_dupes$away_team <- as.character(no_dupes$away_team)
no_dupes <- no_dupes %>% mutate(
  win_team = ifelse(won.home == 1, home_team, away_team)
)
win_totals <- table(no_dupes$win_team,no_dupes$season)
win_totals <- as.data.frame.matrix(win_totals)
win_totals <- tibble::rownames_to_column(win_totals,"Off")
win_totals <- melt(win_totals,"Off")
colnames(win_totals)[1:3] <- c("Off","Season","Total Wins")
win_totals$Season <- as.integer(as.character(win_totals$Season))
total_df <- left_join(total_df, win_totals, by = c("OFF"="Off","SEASON"="Season"))
library(varhandle)
total_df$YDS <- unfactor(total_df$YDS)
total_df$YDS <- gsub(",","",total_df$YDS)
total_df$YDS <- as.integer(total_df$YDS)
#tenyear_totals <- total_df %>% group_by(OFF) %>% summarise_at(vars(YDS:`Total
#Wins`),sum)

# Merge to Correct Choices
correctchoices <- left_join(correctchoices, win_totals, by =
c("Team"="Off","Season"="Season"), all.x = TRUE)
correctchoices <- correctchoices %>% mutate(
  win_freq = `Total Wins`/16
)
```
```{r}
# Bar Graph Showing win max freq by team

#2010
correctchoices10 <- correctchoices %>% filter(Season==2010)
~~~~

```
correctchoices10$Team <- reorder(correctchoices10$Team, -
correctchoices10$win_max_frequency)

correctchoices10 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2010 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2011
correctchoices11 <- correctchoices %>% filter(Season==2011)
correctchoices11$Team <- reorder(correctchoices10$Team, -
correctchoices11$win_max_frequency)

correctchoices11 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
```

```
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probabilitiy Maximizing Frequency By
Team", subtitle = "From a Sample of the 2011 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2012
correctchoices12 <- correctchoices %>% filter(Season==2012)
correctchoices12$Team <- reorder(correctchoices12$Team, -
correctchoices12$win_max_frequency)

correctchoices12 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2012 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2013
correctchoices13 <- correctchoices %>% filter(Season==2013)
correctchoices13$Team <- reorder(correctchoices13$Team, -
correctchoices13$win_max_frequency)

correctchoices13 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
```

```
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2013 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))


#2014
correctchoices14 <- correctchoices %>% filter(Season==2014)
correctchoices14$Team <- reorder(correctchoices10$Team, -
correctchoices14$win_max_frequency)

correctchoices14 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2014 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))
```

```
#2015
correctchoices15 <- correctchoices %>% filter(Season==2015)
correctchoices15$Team <- reorder(correctchoices10$Team, -
correctchoices15$win_max_frequency)

correctchoices15 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2015 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2016
correctchoices16 <- correctchoices %>% filter(Season==2016)
correctchoices16$Team <- reorder(correctchoices16$Team, -
correctchoices16$win_max_frequency)

correctchoices16 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
```

```
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2016 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2017
correctchoices17 <- correctchoices %>% filter(Season==2017)
correctchoices17$Team <- reorder(correctchoices10$Team, -
correctchoices17$win_max_frequency)

correctchoices17 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2017 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2018
correctchoices18 <- correctchoices %>% filter(Season==2018)
correctchoices18$Team <- reorder(correctchoices10$Team, -
correctchoices18$win_max_frequency)

correctchoices18 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
```

```
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2018 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#2019
correctchoices19 <- correctchoices %>% filter(Season==2019)
correctchoices19$Team <- reorder(correctchoices19$Team, -
correctchoices19$win_max_frequency)

correctchoices19 %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2019 Season") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
```

````
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_freq, color=Team))

#Overall

correctchoicestotal <- aggregate(. ~ Team, correctchoices, sum)
correctchoicestotal <- correctchoicestotal %>% dplyr::select(Team,`Total Wins`,`Win
Maximizing Choice`, Occurences)
correctchoicestotal <- correctchoicestotal %>% mutate(
  win_max_frequency = `Win Maximizing Choice`/Occurences,
  win_percentage = `Total Wins`/160
)
correctchoicestotal$Team <- reorder(correctchoicestotal$Team, -
correctchoicestotal$win_max_frequency)

correctchoicestotal %>%
ggplot(aes(x=Team,y=win_max_frequency,fill=Team,color=Team,group=Team,label=win_max
_frequency)) + geom_col() + scale_color_manual(values = c("NYJ" = "#ffffff", "NE" =
"#C8032B","MIA"="#f26a24","BUF"="#00308F","PIT"="#000000","CLV"="#FF3A00","CIN"
="#000000","BLT"="#BC9428","TEN"="#16233F","JAX"="#006778","IND"="#FFFFFF","HS
T"="#022030","OAK"="#000000","LAC"="#F0AE00","KC"="#ffb612","DEN"="#FE4E00","
WAS"="#ffb612","PHI"="#A5ACAF","NYG"="#092067","DAL"="#869397","MIN"="#FFC6
2f","GB"="#FFB612","DET"="#B0B7BC","CHI"="#C83803","TB"="#ff7900","NO"="#00000
0","CAR"="#BFC0BF","ATL"="#A9162D","SF"="#b3995d","SEA"="#002244","LAR"="#B39
95d","ARZ"="#000000")) + scale_fill_manual(values = c("NYJ" =
"#003f2d","NE"="#002145","MIA"="#008e97","BUF"="#C8023A","PIT"="#F0AE00","CLV"
=
"#311D00","CIN"="#FE4E00","BLT"="#241075","TEN"="#4790DE","JAX"="#9f792c","IND"
="#002B61","HST"="#A9162D","OAK"="#A5ACAF","LAC"="#007FCB","KC"="#E31837","
DEN"="#002145","WAS"="#773141","PHI"="#004C55","NYG"="#A9162D","DAL"="#00224
4","MIN"="#4F2683","GB"="#203731","DET"="#0075B8","CHI"="#0b162a","TB"="#d50a0a"
,"NO"="#D4BD8A","CAR"="#0084CD","ATL"="#000000","SF"="#aa0000","SEA"="#69be28
","LAR"="#002244","ARZ"="#97233f")) + ggtitle("Win Probability Maximizing Frequency By
Team", subtitle = "From a Sample of the 2010-2019 Seasons") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1)) + labs(y="Frequency", caption = "Bars represent
the win probability maximizing frequency. Points represent the team's win percentage.") +
geom_point(aes(y=win_percentage, color=Team))
```
``` {r}
correctchoices$Team <- as.character(correctchoices$Team)
correctchoices <- merge(correctchoices, total_df, by.x = c("Team","Season"), by.y =
c("OFF","SEASON"))
correctchoices$`Total Wins.y` <- NULL
colnames(correctchoices)[6] <- "Total Wins"
```
``` {r}
````

~~~~
# On Total Yards
  # Scatter
scatter.smooth(correctchoices$win_max_frequency, correctchoices$YDS, main="Yards ~ Win
Max Freq", xlab="Win Probability Maximizing Frequency",ylab="Total Yards")
cor(correctchoices$win_max_frequency, correctchoices$YDS)
```
```{r}
  # Box Plot For Outliers
par(mfrow=c(1, 2))  # divide graph area in 2 columns
boxplot(correctchoices$win_max_frequency, main="Win Max Freq", sub=paste("Outlier rows:
", boxplot.stats(correctchoices$win_max_frequency)$out))
boxplot(correctchoices$YDS, main="Yards", sub=paste("Outlier rows: ",
boxplot.stats(correctchoices$YDS)$out))
```
```{r}
# Density Plot
library(e1071)
par(mfrow=c(1, 2))  # divide graph area in 2 columns
plot(density(correctchoices$win_max_frequency), main="Density Plot: Win Max Freq",
ylab="Frequency", sub=paste("Skewness:",
round(e1071::skewness(correctchoices$win_max_frequency), 2)))
polygon(density(correctchoices$win_max_frequency), col="red")
plot(density(correctchoices$YDS), main="Density Plot: Yards", ylab="Frequency",
sub=paste("Skewness:", round(e1071::skewness(correctchoices$YDS), 2)))
polygon(density(correctchoices$YDS), col="red")
```
``` {r}
# On Yards Per Game
  # Scatter
scatter.smooth(correctchoices$win_max_frequency, correctchoices$YDS.G, main="Yards Per
Game ~ Win Max Freq", ylab="Yards Per Game", xlab="Win Probability Maximizing
Frequency")
  # Box Plot For Outliers
boxplot(correctchoices$YDS.G, main="Yards Per Game", sub=paste("Outlier rows: ",
boxplot.stats(correctchoices$YDS.G)$out))
cor(correctchoices$win_max_frequency,correctchoices$YDS.G)
```
```{r}
  # Density Plot
par(mfrow=c(1, 2))  # divide graph area in 2 columns
plot(density(correctchoices$YDS.G), main="Density Plot: Yards Per Game", ylab="Frequency",
sub=paste("Skewness:", round(e1071::skewness(correctchoices$YDS.G), 2)))
polygon(density(correctchoices$YDS.G), col="red")
```
``` {r}
# On Total Wins Over 10 Year Span
~~~~

````
colnames(correctchoices)[11] <- "Total_Wins"
  # Scatter
scatter.smooth(correctchoices$win_max_frequency, correctchoices$`Total Wins`, main="Total Wins ~ Win Max Freq", xlab = "Win Probability Maximizing Frequency",ylab="Total Wins")
  # Box Plot For Outliers
boxplot(correctchoices$Total_Wins, main="Total Wins", sub=paste("Outlier rows: ", boxplot.stats(correctchoices$Total_Wins)$out))
```
```{r}
  # Density Plot
par(mfrow=c(1, 2))  # divide graph area in 2 columns
plot(density(correctchoices$Total_Wins), main="Density Plot: Total Wins", ylab="Frequency", sub=paste("Skewness:", round(e1071::skewness(correctchoices$Total_Wins), 2)))
  # Cor
cor(correctchoices$win_max_frequency, correctchoices$Total_Wins)
```
```{r}
  # Model
total_df
linearMod <- glm(Total_Wins ~ win_max_frequency, data=correctchoices, family = "poisson")
summary(linearMod)

correctchoices <- left_join(correctchoices, ratingswinprob_def, by = c("Team","Season"))
correctchoices <- left_join(correctchoices, ratingswinprob_off, by = c("Team","Season"))

linearMod2 <- glm(Total_Wins ~ win_max_frequency + Offensive_Pass_Rank + Offensive_Rush_Rank + Defensive_Pass_Rank + Defensive_Rush_Rank, data = correctchoices, family = "poisson")
summary(linearMod2)

correctchoices <-
left_join(correctchoices,total_df%>%dplyr::select(OFF,PTS,SEASON),by=c("Team"="OFF","Season"="SEASON"))

linearMod3 <- glm(PTS ~ win_max_frequency + Offensive_Pass_Rank + Offensive_Rush_Rank, data =correctchoices, family = "poisson")
summary(linearMod3)

cor(correctchoices$YDS.G,correctchoices$`Total Wins`)
```
``` {r}
# Win Probability Imbalance
wpi <- ggplot(plays %>% filter(down == 4), aes(x = wp_1)) +
  geom_density(position = "identity", alpha = 0.5, aes(fill = as.factor(Attempt_Fourth))) +
  scale_fill_brewer(type = "qual", palette = "Dark2") +
  ggtitle("Win Probability Differential Before Fourth Down Plays") +
````

```
  xlab("Win Probability") + ylab("Density") +
  xlim(-0,1) +
  annotate("text", 0.25, 1.5, label = "Went For It", colour = "#d95f02", alpha = 1, size = 4) +
  annotate("text", 0.75, 1.5, label = "Did Not Go For It", colour = "#1b9e77", alpha = 1, size = 4)
+
  theme(plot.title = element_text(hjust = 0.5, size = rel(1.2)),
        legend.position = "none",
        axis.title.y = element_text(size = rel(1.2)),
        axis.title.x = element_text(size = rel(1.2)),
        axis.text.y = element_blank())
wpi
```