Finding the Changepoint of the NFL Running Back Position

Scott Luke

lscott3@gulls.salisbury.edu

Data Science and Research Methods, Salisbury University, Salisbury, MD 21804

Project Mentor: Kyle Teller

1. Abstract

The National Football League (NFL) is a constantly changing and evolving sport that is well known throughout all the United States. There are many positions, players, and staff surrounding the NFL, as well as countless statistics that are recorded and available to the public for teams and players. Dating back to the first super bowl in 1966, this paper analyzes the rushing and passing offensive stats for each NFL team dating up to 2020. The hope of comparing these two datasets is to identify a point in time in which the running back position has changed in the NFL. This could prove how the position is being used differently by coaches, to the development of a deeper, more receiving oriented roster of running backs for a team. The results highlight a change in teams' tendencies to pass rather than rush in offensive attempts. Upon further research and data collection, an answer to the initial question could be found when collecting data for individual players.

2. Introduction

In this project, we explore the offensive rushing statistics of NFL teams, dating back to the first Super Bowl in 1966. Specifically analyzing the data from the rushing offense of all teams can hopefully identify a changepoint in how the running back position is used in football today. In a modern-day game of football, two teams face off and try to score more points than their opponent while also providing a strong defense to stop their opponents from doing the same.

Offensively, there are countless different plays that a coach can provide for the team, and there are constant innovations to the way football is played. New players and new coaches come into the game with new ideas to try to develop a stronger, more consistent play to trick or overpower the defense of the opponent. Primarily, these plays consist of rushing or passing attempts. After the quarterback (QB) snaps, he either steps back to pass the ball to a wide receiver (WR) or he hands the ball off to one of his running backs (RB). As mentioned, these plays can be altered and are often modified to achieve better results. The more yards the ball carrier can acquire, the closer the team is to the endzone, and the closer the team is to scoring touchdowns to win the game.

Making comparisons from rushing and receiving yards and attempts over a set of years can show how much teams are tending to rush or pass. Overall, the average rushing play tends to provide less yards than the average passing play as the quarterback throwing the ball has a higher risk and reward. Throwing the ball could lead to an

interception, in which the opposing defense catches the ball over the intended receiver. The receiver could also catch the ball, then fumble it in which the defense has an opportunity to steal the ball. An incomplete pass is also very common, which results in no yards gained. On the contrary, running the ball has fewer risks, including fumbling or a loss of yards. Using this knowledge, when viewing total yards of rushing and passing yards side by side, there should be a general difference of passing yards being higher than rushing.

This trend, however, differs when looking at attempts. There are an indefinite number of scenarios in a football game, all of which lead coaches to call different plays to achieve the best possible results. There is no "perfect model" of play that a team can achieve. Every team has different players with different strengths, all facing off against another team with different players with different strengths. At best, a coach and team is only able to make decisions based off the statistics they know from the other team, as well as what works with their own team. A team with a strong set of running backs is highly likely to have more rushing attempts than a team that has a deep pool of wide receivers. In conclusion, this study attempts to locate relations between teams' tendencies of plays and overall achievements in primarily rushing completions in hopes to identify a changepoint in the running back position.

3. Dataset and Features

The dataset used for this study was pulled from Pro Football Reference [1]. This resource provides many stats related to the NFL and American Football League (AFL), which even dates to games and stats

recorded in 1920. The specific data was pulled from yearly datasets of team offensive stats that separates passing and rushing from every team during the season of the specified year. This consists of mostly quantitative data, providing total yards, number of attempts, touchdowns, and more. When exploring the data, some attributes needed to be curated. As the NFL has expanded, so has the record of stats for each season. An example is in 2000, the site started recording expected points contributed by the rushing offense. As this is only on a fraction of the data used in the study, this attribute had to be excluded in the following methods.

Extracting is relatively straightforward, as Pro Football Reference provides options for downloading the data in Comma Separated List (csv) format or Excel format. The individual years are separated, showing all teams' total stats per year, so cleaning required downloading all *csv's* for every year, then using Python to load them into individual data frames. In addition, the year column was added to keep the data separate. To provide comparison to the rushing data, we downloaded the passing data as well. The same method was used to extract the data from Pro Football Reference, however it needed to be cleaned. There were irrelevant attributes that could not be compared and were dropped.

For the final part of this research, a portion of individual player data was also downloaded. We found stats for the top 10 running backs of the years 1978 and 2020 from the same source. There was much more player data, but these were selected after the first part of the research. Comparing broad stats like yearly team data

led to more specific data that needed to be shown to reach a conclusion.

# 4. Methods

4.1 Data Cleaning

The data first needed to be cleaned as there were attributes provided by the database that were unnecessary for this comparison. First, using Python and the Pandas library, the data was read from the individual csv's containing each year's rushing and passing stats. Then, the years for each data frame were added to easily identify them after combining. The datasets were then combined using Pandas concatenation function, to create two datasets, one containing all the passing stats and one containing rushing stats. After this, there were several columns that needed to be removed (Yards per Attempt, Touchdown Percentage, Expected Points, Completion Percentage, Interception Percentage, Longest Pass, Average Yards per Attempt, Yards per Completion, Sack Percentage, Net Yards per Attempt, Adjusted Net Yards per Attempt, 4th Quarter Comebacks, and Game Winning Drives).

After analyzing the data, there was an outlier found amidst both passing and rushing data. The year was 1982 and the statistics for this year were much lower than others. Further research was needed to find out what happened in the NFL during this year.
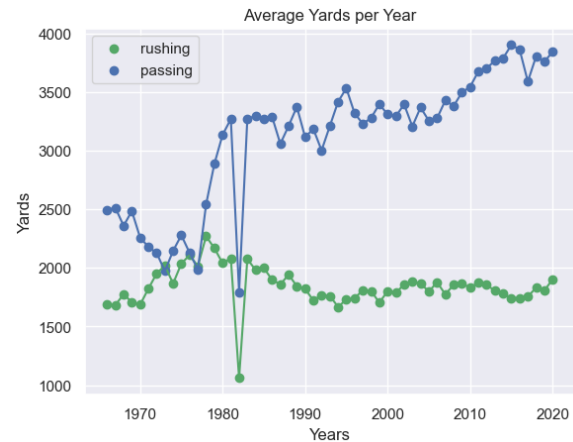

Figure 1. Average Yards per Year, highlighting outlier in data.

4.2 Comparison Graphs
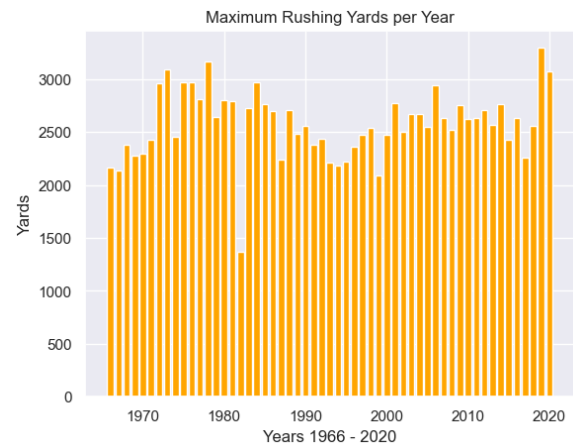
Figure 2: Maximum Rushing Yards per Year

Firstly, bar graphs were made to show the maximum yards per year. This is done by isolating the yards attributes on both datasets, grouping by year, and calculating max using pandas built in 'max()' function attached to data frames. Following the maximum yards bar graph, the average yards per year line graph was created using the same methods, but using the 'mean()' function instead of 'max()'.
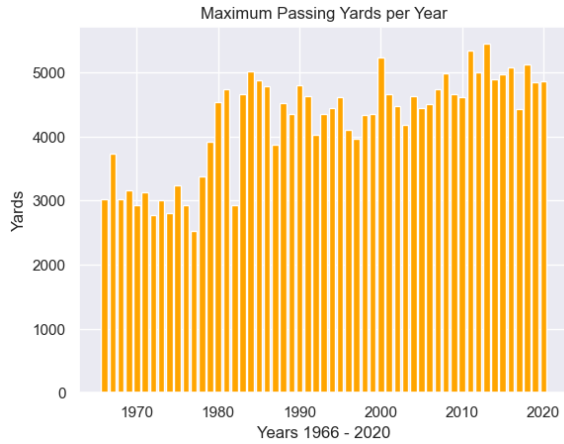
Figure 3. *Maximum Passing Yards per Year*

The next step involved selecting other attributes related to running back utilization and yards gained. The two closest to this were attempts and touchdowns. Similar line graphs were made highlighting the same range of years, with both rushing and passing displayed for comparison. For touchdowns and attempts, the 'sum()' function was used for the visualization of the data.
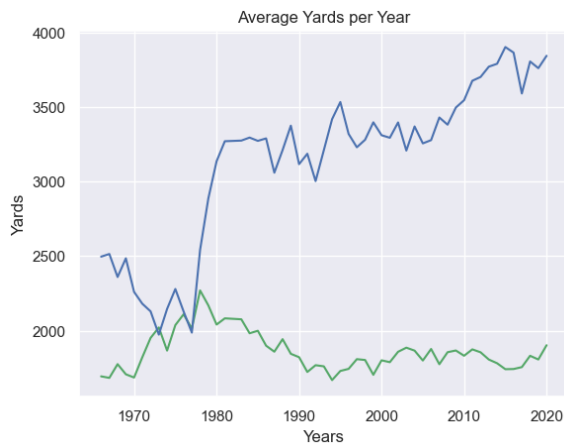
Figure 4. *Average Yards per Year*
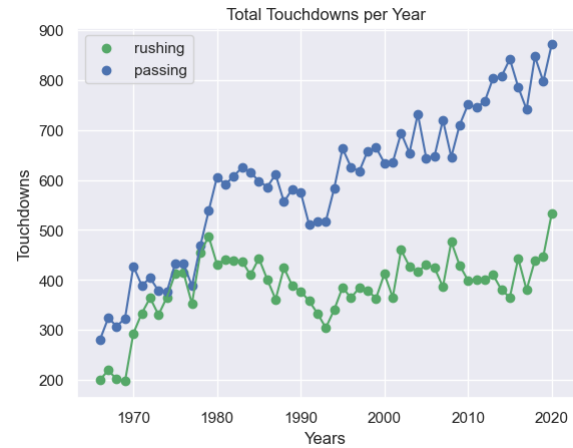
Figure 5. *Total Touchdowns per Year*
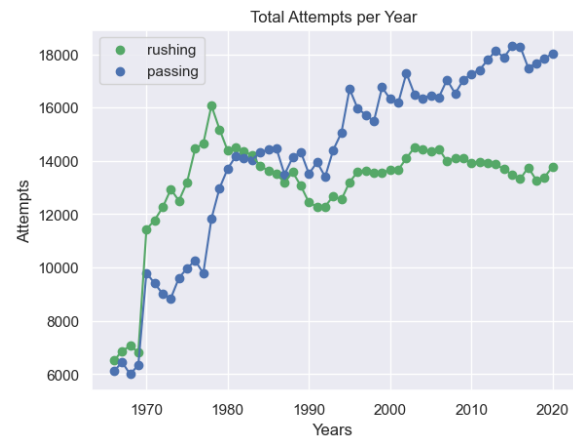

Figure 6. *Total Attempts per Year*

4.3 Confidence Interval on Attempts

The attempts line graph represented the best visual comparison of both datasets as well as having the most correlation with the goal of the research. So, to further on the comparison of rushing and passing attempts, the confidence interval of both was calculated. This was done using a 95% confidence interval, which corresponds to a value of 1.96 for the formula.

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

For this formula, $\bar{x} = sample\ mean$, $z = confidence\ value$, $s =$

sample standard deviation, n = sample size. Using this formula on every value in each dataset, the result is two copies of each dataset, one is an upper bound and the other is a lower bound. Then using the Pyplot 'fill_between()' function, each dataset can be displayed with the corresponding confidence intervals.
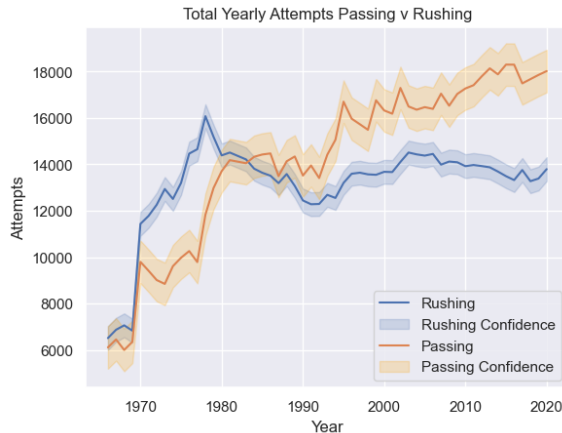
Figure 7. Confidence Interval of Total Yearly Attempts

4.4 Player Data Display

For this exploratory experiment, bar graphs were made using the same methods as before, only adjusting the x position to display both datasets; top 10 players from 1978 versus the top 10 players of 2020 on the same bar graph. The first set of graphs displays the rushing yards, receiving yards, and the percentage of rushing out of total yards. This was followed by comparing the amount of rushing and receiving yards for 1978 and 2020 individually.

5. Results and Analysis

5.1 Data Cleaning

The outlier year in this data set was 1982 where there was a recorded player strike in the NFL [2]. Players, staff, and the league had disputes about wages which eventually led to a shutdown of games. The season only lasted 9 games which led to a lack of recorded statistics. This resulted in the averages from the original graphs to be much lower for the year 1982.

5.2 Comparison Graphs

Firstly, comparing the maximum rushing and passing yards shows us that the y-scale for these graphs differs by 2000 yards. The team with the maximum passing yards reaches around 5000, whereas the rushing yards only reach around 3000. The reason for this is that in a game of football, the yards rewarded for passing are much higher. When a quarterback throws the ball, there is a chance of an interception, however, if the wide receiver beats the defender, the yards gained could easily be 20+ yards. To acquire this many yards from a single running play, the running back must run from behind the line of scrimmage, through multiple defenders, and manage to get far enough to gain just as much.

When looking at the results of the maximum yards graph, there is a slight trend around 1980 that a jump in passing yards. Unfortunately, with the previous conclusion about yards gained from passing yards compared to rushing, there is not enough evidence to support a conclusion about the running back position from maximum yards alone. Adding average yards, touchdowns, and attempts helps clarify some of the missing details.

The average yards show similar trends as the maximum yards, however, in addition to the jump in passing yards around the 1980's, there is a decrease in rushing yards. Around 1978, there is a peak in rushing yards, then following that there is a slight decrease continuing into the 21st century. The total

touchdowns show the same pattern, but the passing touchdowns do not take as much of a dramatic increase within the early 80's. When discussing touchdowns, it is also important to note how touchdowns are acquired. A team could drive down the field, passing for every play, then run the ball into the endzone for a touchdown at the 1-yard line. This would contribute to the rushing touchdowns but does not support an increase in running back utilization. Or the team could have a strong running back core, using running plays all the way down the field and scoring a touchdown in the same manner. This analysis also becomes inconclusive because of the inconsistency in touchdown plays.

5.3 Confidence Interval on Attempts

Moving to the attempts, we can see a strong visualization of running back utilization. The trends match the previous data, where rushing attempts peak in 1978, and in this display, are higher than the passing attempts. The two datasets then swap, passing becoming higher than rushing, and there is an overlap in the 1980s. This is better visualized with the continuation of this analysis, which was calculated and showing the confidence interval of the two datasets. The overlap in the 80's along with rushing attempts being higher than passing until then, supports that teams tended to run the ball more before the 80's and passed more often in today's game. This conclusion is strong; however, it does not exactly support the answer to the question of how the running back position has changed. For example, a passing play could result in the ball being passed to the running back, in fact, modern teams like the San Francisco 49ers utilize their running back in passing plays often.

An addition to this was discussed in an article addressing running back devaluation in the modern era of football [4]. Most of the topics covered in this article are related to wages and salaries of running backs, however it is backed by data that relates to this study. The figure shows how many running backs of super bowl winning teams acquired more than 1000 rushing yards in their season. This number has slowly decreased throughout the decades, beginning at 9 out of 10 in the 1970's to just 3 out of 10 in the 2010's. This data could address how often running backs are used to run over receive.
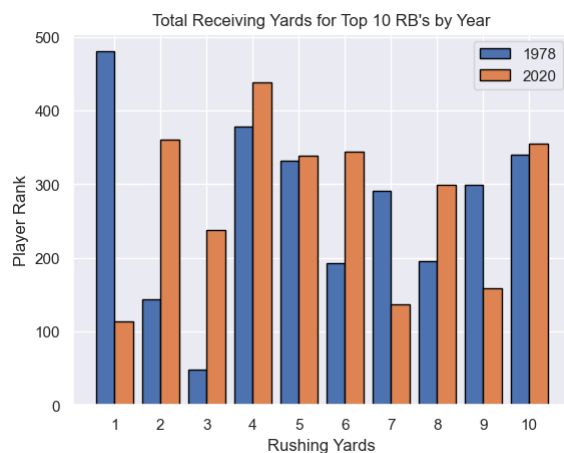
1,000-YARD RUSHERS ON SUPER BOWL WINNERS BY DECADE		
2010s	3	LeGarrette Blount, Marshawn Lynch, Ray Rice
2000s	6	Brandon Jacobs, Joseph Addai, Willie Parker, Corey Dillon, Antowain Smith, Jamal Lewis
1990s	7	Marshall Faulk, Terrell Davis (twice), Emmitt Smith (three times), Earnest Byner
1980s	6	Roger Craig (twice), Walter Payton, Marcus Allen, Wendell Tyler, Joe Morris
1970s	9	Franco Harris (four times), Tony Dorsett, Mark van Eeghen, Larry Csonka (twice), Mercury Morris

Figure 8. (Clawson, D) Display of super bowl winning running backs per decade.
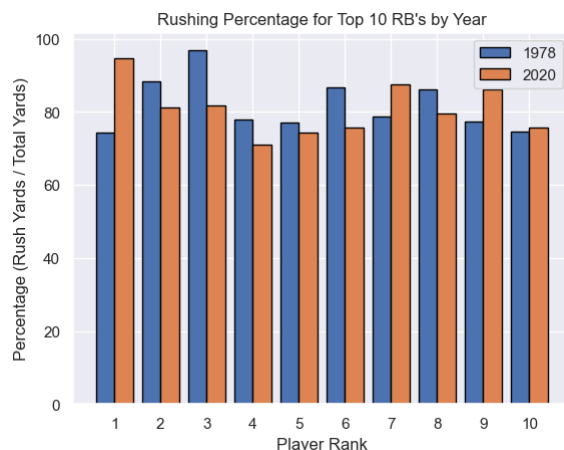
5.4 Player Data Display

To expand on this analyzation, we would need to extract and examine data for individual players in the NFL. This would be extensive research with heavy amounts of data. To minimize this, we specifically targeted key years from the dataset, 1978 which had the peak rushing attempts, and 2020, the most recent year in the dataset. Taking the top 10 running backs from both years, we can compare their individual rushing and receiving yards. This visualization shows that the receiving yards for 7 out of 10 of the top running backs in 2020 had more receiving yards than the top

10 running backs of 1978. We also compared the percentage of yards that were rushing from the total yards for each running back. The lowest was the number 4 running back of 2020 with roughly 76% rushing yards and most of the others were around 80% to 90%. So even with 7 of the top 10 running backs of 2020 having more receiving yards then the top 10 of 1978, there is still an almost even spread of rushing yard percentages between the two groups.

Finally, we compared the number of rushing yards to receiving yards for all 10 players of each year. There are still many yards being gained by rushing, and the relation is similar between the two years. Based on these graphs and conclusions, the question is still unanswered on how the running back position has changed. However, the confidence interval displayed along with attempts shows that in the 1980's, NFL teams tended to utilize passing plays much more.

Figure 9. Total Receiving Yards for Top 10 RB's by Year
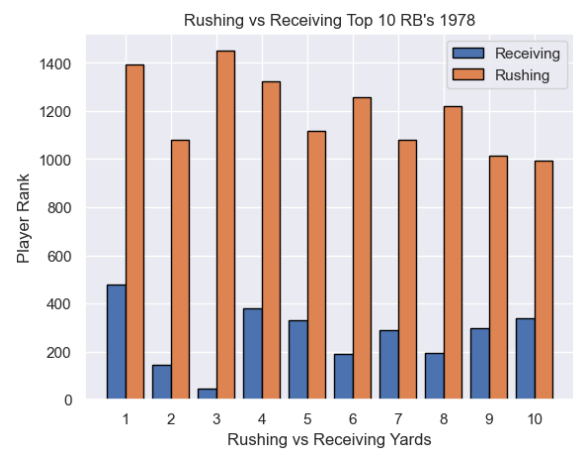
Figure 11. Rushing vs Receiving Top 10 RB's 1978

Figure 10. Rushing Percentage for Top 10 RB's by Year
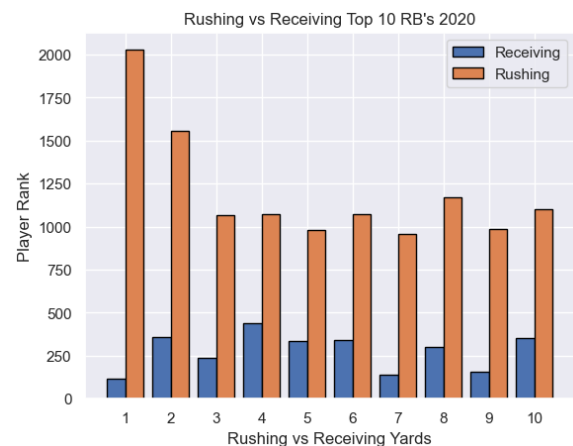
Figure 12. Rushing vs Receiving Top 10 RB's 2020

6. Conclusion and Future Work

The methodology provided an accurate display of a changepoint between the years of 1978 through 1990. This changepoint was on teams' play calling, specifically calling more passing plays than rushing. The overlap period of confidence intervals between 1980 and 1990 is likely to a slow adaption of a new play style across the entire NFL. Unfortunately, due to the versatility of the running back position, this data does not support any conclusion on the change in the running back position.

Further research would be required to find an answer to this, but the closing research on select players could be continued. Every player in the NFL serves a different role to a team, despite having the same position as a different player. One running back could be used more as a wide receiver, whereas another could be used a strong, consistent runner. Further analyzing individual player data could show an increase in either one of these roles and ultimately answer the question of how the running back position has changed since the first super bowl.

7. References

[1] Sports Reference. (n.d.). *Pro Football Stats, history, scores, standings,* playoffs, Schedule & Records. Pro Football Reference. https://www.pro-football-reference.com/

[2] Wojciechowski, G. (1987, September 23). *NFL strike : 1982 : A history lesson not learned*. Los Angeles Times. https://www.latimes.com/archives/la-xpm-1987-09-23-sp-6303-story.html

[3] Wagoner, N. (2023, September 21). *Is 49ers' Christian McCaffrey the prototype for the next star running back?* ESPN. https://www.espn.com/nfl/story/_/id/38447415/49ers-christian-mccaffrey-prototype-star-running-back-rushing-receiving

[4] Clawson, D. (2023, August 3). *Reality of being an NFL running back: Why the position has been devalued and how we got to this point*. CBSSports.com. https://www.cbssports.com/nfl/news/reality-of-being-an-nfl-running-back-why-the-position-has-been-devalued-and-how-we-got-to-this-point/