NFL Big Data Bowl
Theme 1: Understanding Speed

Balram Prasad
College Student – Undergraduate: University of Virginia

Table of Content

Executive Summary:

This report will examine and analyze speed data from the 2017 National Football League (NFL) season to better understand how receivers utilize their speed for plays greater than 40 yards. Specifically, the intention is to present NFL play callers and scouts with findings that will help them better utilize their player's speed with the hopes of helping them achieve plays greater than 40 yards. To do this, an analysis of the difference in a receiver's maximum speed after the catch and maximum speed before the catch versus the difference in yards after the catch and yards before the catch was performed. An analysis of the difference in speed rather than raw speed was used to standardize the disparate maximum speeds of players in the NFL. Classification techniques, regression analysis, and calculations led to the ultimate conclusion that a player is better able to maximize their speed after the catch than before the catch.

Overview:

Long passes are one of the most exciting plays in football – the crowd goes wild, the offense gets moving, and defenses spirit is deflated. Last year in the NFL, there were 273 passing plays from scrimmage that went for 40 or more yards[1]. On average, each team had 8.5 plays during the season that went for more than 40 yards. Of the 12 teams that made the playoffs, 8 of them were above this average, including the top two seeds from both conferences – Philadelphia Eagles, Minnesota Vikings, Pittsburgh Steelers, and New England Patriots. These big plays helped teams move the ball down the field quickly and put points on the board, which ultimately helped them win.

For these long plays, a receiver's speed is an important factor. Whether it is maneuvering through the defense after a short catch, or running past the defense before the ball arrives, speed serves as an advantage in creating these long plays. Speed is defined as the rate at which someone is able to move. During the 2017-18 NFL season NFL Next Gen Stats recorded each players speed for every play, including long plays. The purpose of this report is to look at an efficient way to make use of this speed data and provide insights to help NFL play callers efficiently utilize the speed of their players to generate long plays.

Problem Statement:

On long passes a player catches the ball after a certain number of yards, and then runs for a certain number of yards. The sum of the net yardage gives the total yards of the play. In both of these areas - before the catch and after the catch - the player will reach some max speed. Depending on the amount of yardage, the player's potential to reach a certain max speed will vary (this is further elaborated later). A coach looking to optimize the speed of the player will

---

[1] http://www.nfl.com/stats/categorystats?seasonType=REG&offensiveStatisticCategory=null&d-447263-n=1&d-447263-o=2&d-447263-p=1&d-447263-s=PASSING_40PLUS_YARDS_EACH&tabSeq=2&season=2017&role=OPP&Submit=Go&archive=true&conference=null&defensiveStatisticCategory=TEAM_PASSING&qualified=false

want to be sure that the point of the pass completion is favorable for the player to reach their respective maximum speed. Specifically, there are two scenarios to consider:

1. Receiver catches the ball and then run for more yards than before the catch.
   (Ex: Receiver catches the ball after 15 yards then runs 35 yards for a 50-yard play)

2. Receiver catches the ball and then run for less yards than before the catch.
   (Ex: Receiver catches the ball after 35 yards and then runs 15 yards for a 50-yard play)

The question at hand is which situation is better for effectively using the player's speed – specifically, is it better to give a player the ball short and then let them run (Scenario 1) or let a player run and then give them the ball (Scenario 2)?

Data Summary:

The dataset was provided by the NFL, and contained game data, play data, and player tracking data from 92 of the 256 NFL games from the 2017-2018 season.

The abridged dataset used for analysis contained 35 passing plays that are greater than 40 yards and have a yards before catch (YBC) and yards after catch (YAC) greater than 10 yards. YBC and YAC are both the net yardage that the player ran before and after the catch. The first condition, 40+ yards, was assigned due to the NFL's classification of a long play. The longest plays that the NFL keeps track of per the NFL.com website is 40+ yards. For this reason, plays greater than 40 yards are defined as being "long" plays. The second condition, YBC and YAC greater than 10 yards, was assigned to ensure the players were given sufficient space to reach top speed. 10 yards is the minimum distance that an NFL player can reach maximum speed (deduced based on 40-yard dash training techniques[2]). That being said, it is very unlikely that the player ran a net distance of only 10 yards on a reception that was around 10 yards, as their route likely would not have been a straight vertical distance. Therefore, 10 yards was a sufficient distance to place as a cutoff.

There were 20 different teams represented in the dataset on offense. The only three teams that executed more than two of the 35 plays were the Carolina Panthers, Kansas City Chiefs, and the Philadelphia Eagles, each of whom executed 4 plays. 15 of the plays were executed on first down, 9 on second down, and 11 on third down (Figure 1). 9 plays were executed in the first quarter, 13 in the second quarter, 6 in the third quarter, and 7 in the fourth quarter (Figure 2). The most common offensive formation was Shotgun, with 23 of the 35 plays coming from Shotgun (Figure 3). Majority of the plays happened in outdoor stadiums, 24 out of 35, and most, 18 out of 35, were on some form of grass. All games were played in the month of September and October in relatively warm temperature – the lowest temperature was 60° Fahrenheit.

[2] https://www.sportandspeedteam.com/five-stages-of-the-40-yard-dash/

Figure 1

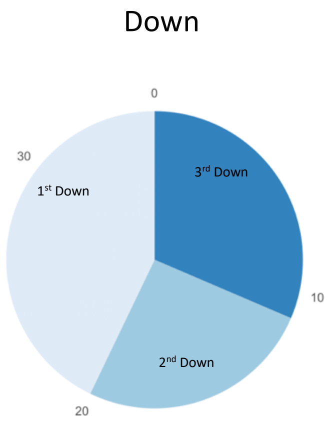


Figure 2

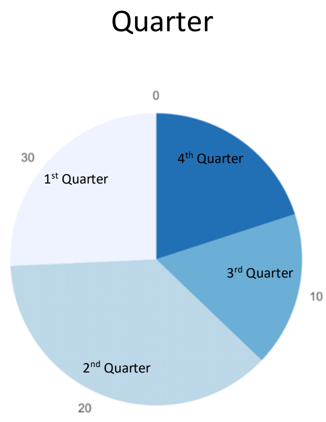
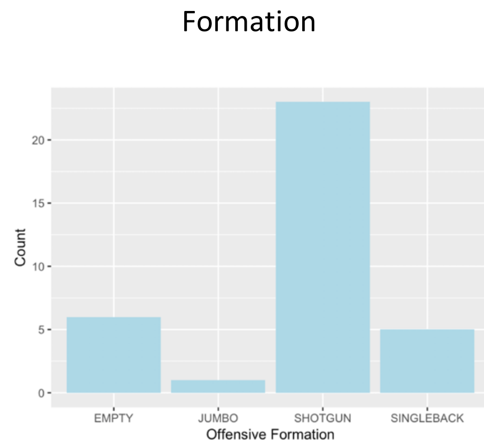


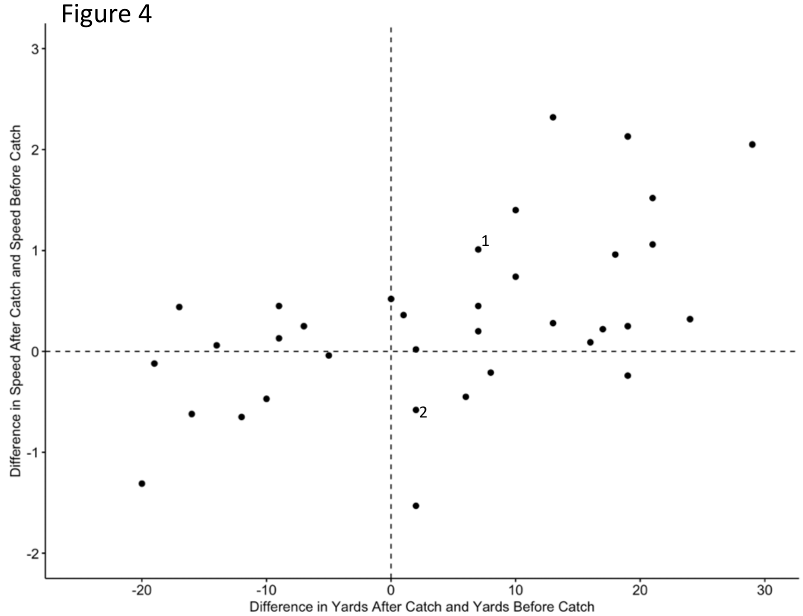
Figure 3



Analysis:

To reiterate, this study compares the maximum speed a receiver reached before and after he caught the ball on plays that went for 40+ yards and had YBC and YAC greater than 10 yards. Difference in speed and yards was used to standardize the speed of the players. To analyze this question, a graph was used that compares the difference between YAC and YBC to the difference between max speed after catch and max speed before catch. Figure 4 shows this relationship. Because this graph is critical and is referenced multiple times, a couple points have been denoted and explained below

Figure 4



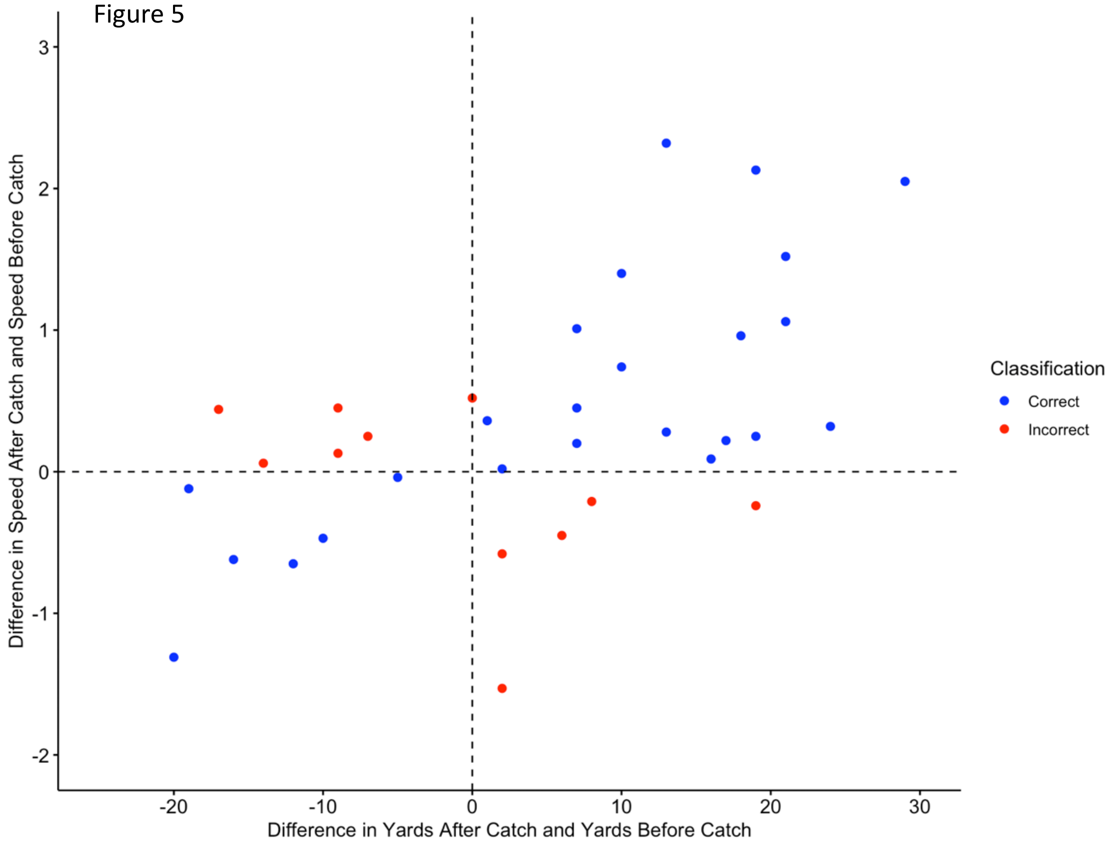
1. The play was a 43-yard reception by Kelvin Benjamin from Cam Newton. Benjamin caught the ball after 18 yards and ran for 25 yards after the catch (difference of +7 yards). Before the reception, Benjamin's maximum speed was 6.85 mph and after the

reception his maximum speed was 7.86 (difference of +1.01). Therefore, the coordinates graphed are (+7, +1.01).

2. The play was a 43-yard reception by Brandon Coleman from Drew Brees. Coleman caught the ball after 20 yards and ran for 22 yards after the catch (difference of +2 yards). Before the reception, Coleman's maximum speed was 9.14 mph and after the reception his maximum speed was 8.56 (difference of -0.58). Therefore, the coordinates graphed are (+2, -0.58).

The initial classification used classified points in which the difference in yards and difference in speed has signs that pointed in the same direction as 'Correct', and all other points as 'Incorrect'. This would mean that for plays in which the YAC was greater than the YBC, the maximum speed after the catch is also greater than the maximum speed before the catch. Similarly, for plays in which the YAC was less than the YBC, the difference in speed after the catch would be less than the difference in speed before the catch. These points would lie in the 1st and 3rd quadrant, as shown in Figure 5. There are 24 points that are correctly classified and 11 points that are incorrectly classified.

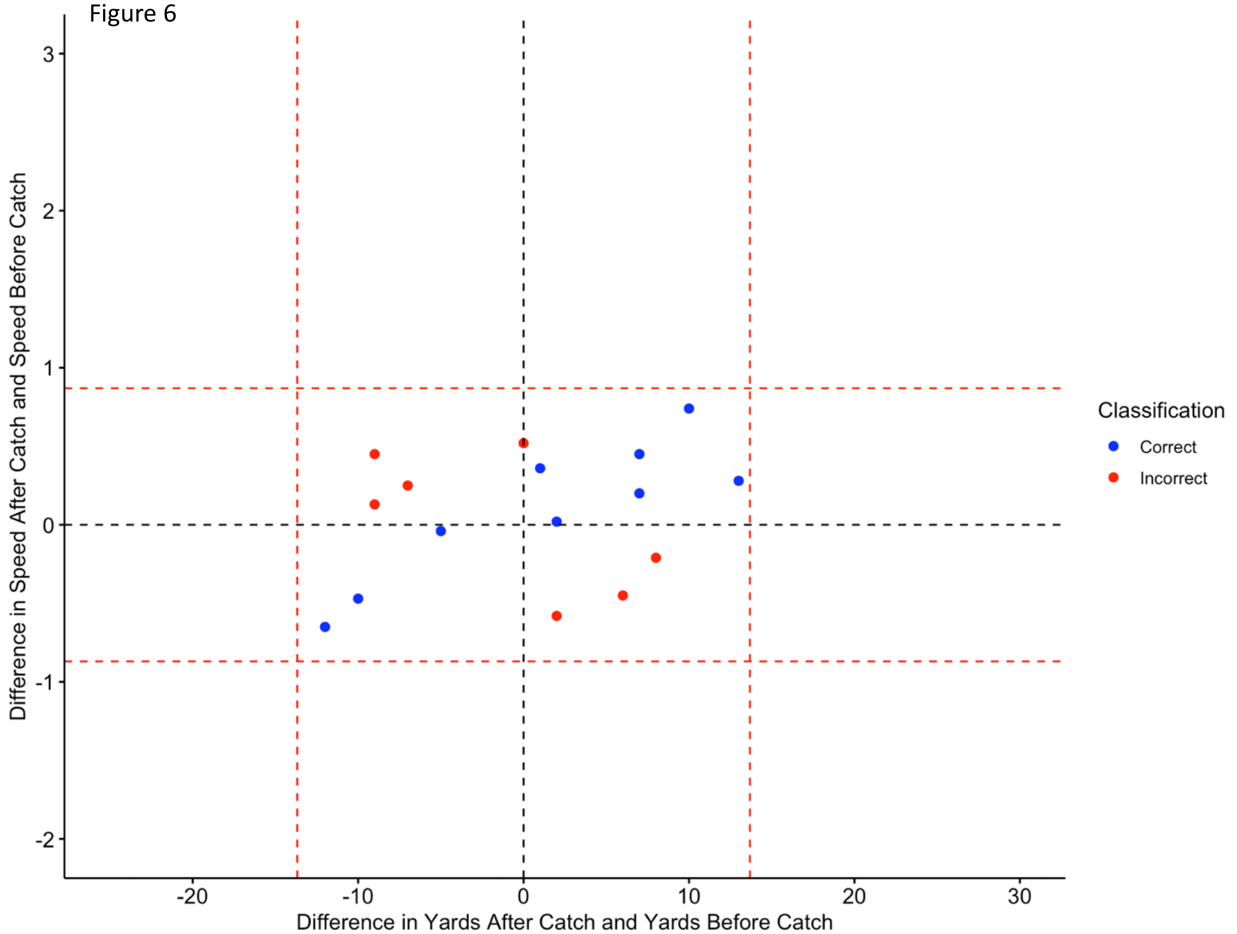
Figure 5



This classification assumes that a greater distance in yards will give the player a greater distance to accelerate, therefore giving them a higher chance of reaching their top speed. We find that 24 of the 35 observations, 70%, are classified as correct, meaning they fall in either the first or third quadrant. For the points in which the distance in yards after the catch is greater than distance in yards before the catch (points to the right of the x-axis), 80% are classified as being correct. For the points in which the distance in yards after the catch is less than distance in yards before the catch (points to the left of the x-axis), 50% are classified as being correct. Before analyzing the correct points, we will first look into the points that are incorrectly classified.

At first glance it is obvious that the incorrect points have either a very small difference in the difference in yardage, or a very small difference in the speed after and before the catch – all of the points are close to either the x-axis or the y-axis). Figure 6 shows the points that are within one standard deviation, 13 yards and 0.86 mph, of no difference for both yardage and speed. Standard Deviation is a metric that tells how far your numbers are spread apart. (Note: Standard deviation is calculated based on the yardage difference and the difference in maximum speed using the abridged dataset consisting of 35 plays)

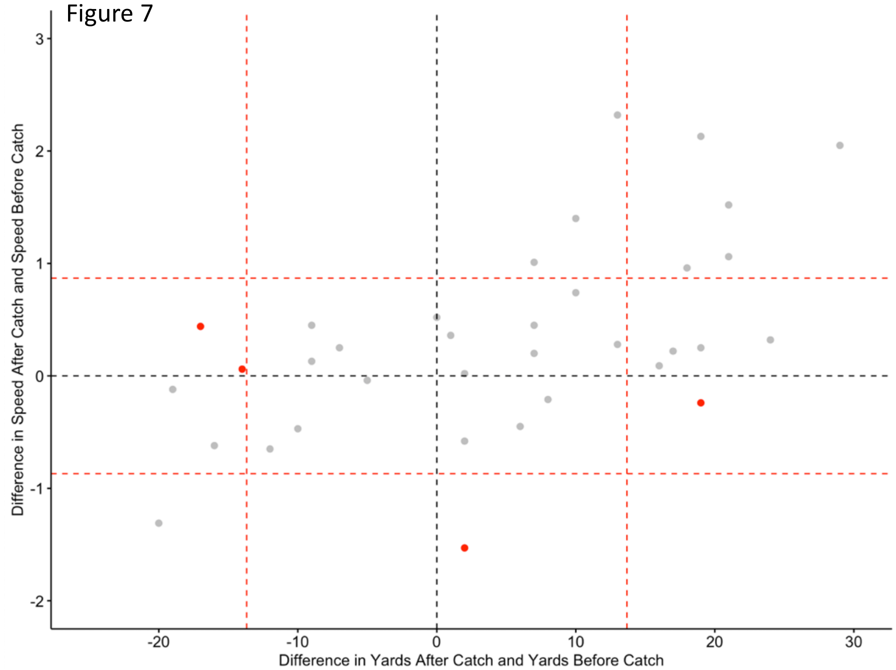
Figure 6



The points in Figure 6 have small values for the difference in yardage before and after the catch, and consequently have a low difference in maximum speed. This short difference in yardage most likely leads to there being a low difference in speed, as there is not a significant difference in the area that a receiver has to accelerate before and after the catch. At this level, there is little that our classification can say. Of the 16 points within one standard deviation, 7 were incorrectly classified, while only 9 were correctly classified. The overall percentage correct was only about 56%, slightly more than half. If the difference in yards is between 0 and 13, the percentage correct is about 67% (points to the right of the x-axis). If the difference in yardage is between -13 and 0 the percentage correct is approximately 43% (points to the left of the x-axis). While both of these percentages are quite distinct, it is hard to make any relevant conclusions due to the small sample size. That being said, any significant findings regarding difference in speed would not be expected if the difference in yardage is small (within 13 yards). This is due to our previously mentioned assumption that the larger the gap in yards the higher chance of correctly predicting whether the receiver will have a higher max speed before or after the catch. Further studies should be done with difference in yardage between -13 and 13 to gain any meaningful insight. We will therefore use the overall percentage, 56%, and conclude that if the difference in yardage is between -13 and 13, then no significant prediction can be derived on if the receiver will reach a greater max speed before or after the catch. Though it is not conclusive, this is likely due to the similarity in acceleration distance.

Figure 6 contains the 7 points classified as incorrect which fall within 1 standard deviation of both the difference in speed and the difference in yards. This means there are 4 points that are classified as incorrect and fall outside of one standard deviation of either difference in yards or difference in speed. There were no points that were classified as incorrect that fell outside one standard deviation of both the difference in yardage and the difference in speed. The 4 points are shown in Figure 7 and are denoted in red.

Figure 7



Of the 4 points, one-point falls below one standard deviation of speed, two points fall below one standard deviation of yardage, and one is above one standard deviation of yardage. While these values all fall outside of one standard deviation for one of the values, they are close to zero for the other one. Specifically, for the three points that are outside one standard deviation of yards, their difference in speed is within half a standard deviation for speed of no difference. For the one point that is below one standard deviation of difference in speed, the difference in yards is within half a standard deviation for yards of no difference. Therefore, if there is a significant difference in yards before and after the catch, then even if the sign of the difference in speed is incorrect, it will not be by a large amount. Similarly, if the difference in speed is significantly high and the sign for the difference in yardage is incorrect, the difference will be small. This indicates that our error for these values is low and therefore does not give reason for serious concern.

We now shift focus to the points classified as correct. Figure 8 shows the points that had a positive difference in speed and yards. Figure 9 shows the points that have a negative difference in speed and yards.

Figure 8

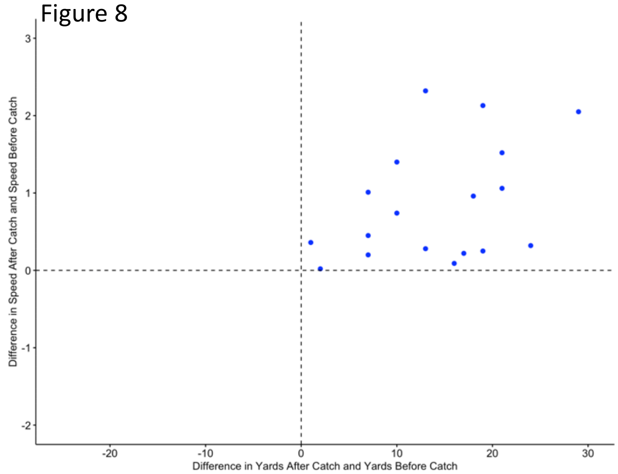
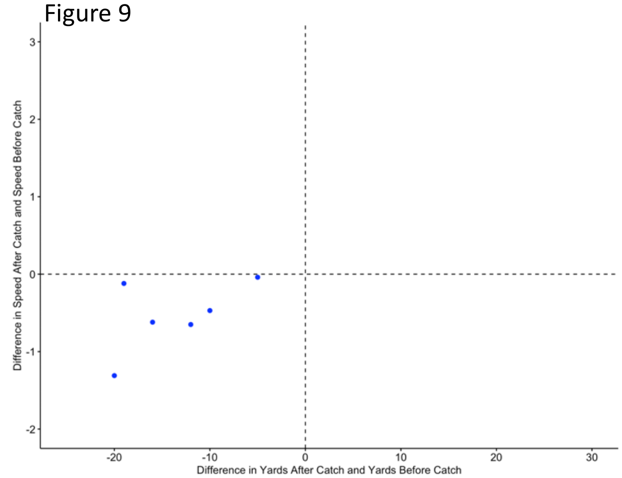


Figure 9



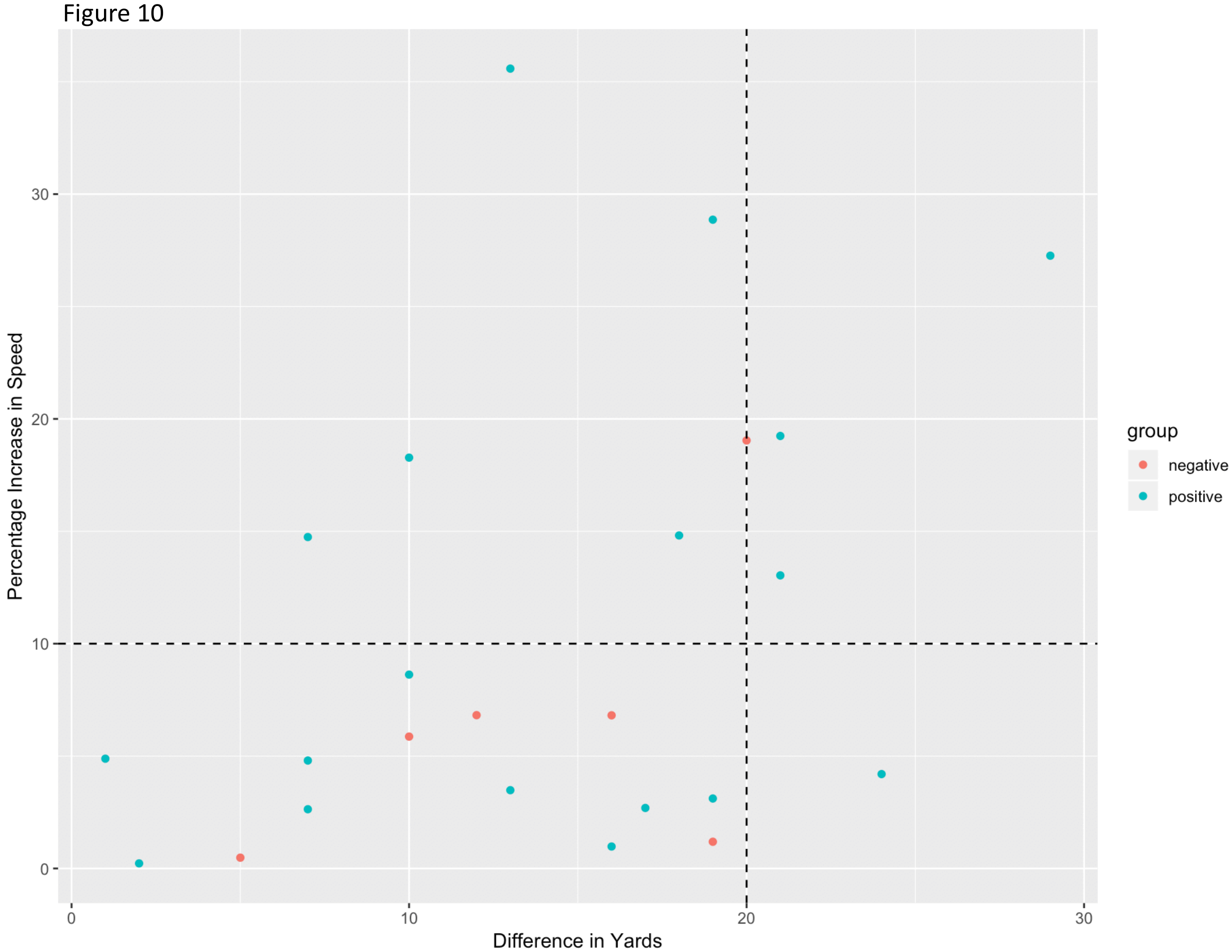
Correlation coefficient shows the linear dependence, with a value of 1 showing perfect linear dependence and a value of 0 showing no linear dependence. Figure 8 has a correlation coefficient of approximately 0.40, while Figure 9 has a correlation coefficient of approximately 0.55. Since both graphs have very low correlation coefficients, there is no linear relationship between the two points.

While there is a low correlation coefficient, indicating no relationship, both relationships are positive. This indicates that as the difference in yardage increases, the difference in speed also increases. Overall, all 11 points have a correlation coefficient of about 0.57, meaning overall the data is not strongly linear, but does have a significant positive trend. This supports our previously held notion that as acceleration distance increases, the difference in maximum speed will also increase.

A linear model test was performed to see if a line of best fit could be derived. No statistically significant line of best fit could be derived for the points for Figure 4, Figure 8, or Figure 9. This means we cannot give an indication of how much difference in speed will increase as difference in yardage increases. Furthermore, a line of best fit would be very weak due to the lack of points, especially if divided by quadrant.

The third method looks at the absolute value of the difference in speed. In this situation it would be more accurate to put percentage change on the y-axis rather than the net difference. This would give us a better measure of the relation, and would help us standardize the speeds in the first and third quadrant. While the graphic is not exactly the same as the net difference in the speed, the points are scattered similarly and follow the same trends. Figure 10 shows this graphic. The blue points in the graph denote the points in Figure 9, while the red points denote the points in Figure 10.

Figure 10



Most of the plays that have a negative difference in speed – the maximum speed before the catch was greater than the maximum speed after the catch – have a difference in yards less than 20. For all of these values the percentage increase in speed – the percentage faster the player’s max speed was before the catch than after the catch – is less than 10. The only value that has a percentage increase in speed greater than 20 also has a difference in yards greater than 20. For the positive difference, 9 of the 18 plays have difference in yards less than 20 and percentage increase in speed – in this case the percentage faster a player's max speed is after the catch versus before the catch – less than 10. 5 of the players that have a difference in yards less than 20 have a max speed after the catch that is 10 % greater than the max speed before the catch. 3 players have a difference in yards greater than 20 yards and a max speed after the catch that is 10% greater than the max speed before the catch. 1 player has a difference in yards that is greater than 20 yards and a max speed after the catch that is less than 10% than the max speed before the catch. On average, players that run for more yards after the catch than before the catch reach a max speed that is 12% faster than their max speed before the catch, while those who run for more yards before the catch than after the catch run an average of 6% faster than their max speed after the catch.

Recommendation:

To reiterate the central question: in order to effectively utilize a player's speed would it be better to let the player run for more yards after the catch than before the catch or run for more yards before the catch than after the catch?

Based on the analysis above, in order to effectively utilize a player's speed, it would be better to let the player run for more yards after the catch than before the catch. Below are the reasons:

1. Overall, players that run a farther net distance after the catch than before the catch and have a maximum speed that is greater after the catch than before the catch (Figure 8) have a maximum speed 12% faster after the catch than before the catch. For players that run a farther distance before the catch than after the catch and have a maximum speed that is greater before the catch than after the catch (Figure 9), the maximum speed reached is 6% faster before the catch than after the catch. On plays in which the YAC is greater than YBC, players are comparatively able to run twice faster than on plays in which the YBC is greater than the YAC.

2. For plays in which the difference in yards is less than 20 yards, players are only able to run 10% faster or greater for plays in which the YAC is greater than the YBC. When difference in yards is small, there is a higher chance the player will be able to run comparatively faster after the catch than before the catch. Furthermore, receivers are able to optimize their speed in a smaller distance after the catch than before the catch.

3. Our classification method shows that for points in which the yardage is greater after the catch than before the catch, 80% are classified correctly (speed is greater after the catch than before the catch). For those points in which yardage is greater before the catch than after the catch, only 50% are classified correctly. The analysis shows that for the values that are classified incorrectly for yards, they are typically one standard deviation away from no difference in speed. This would indicate that we need to be correctly classified in order to expect a significant difference in speed. Therefore, it would be better to have the yardage after the catch be greater than the yardage before the catch,

as the potential to be correctly classified, and therefore have a greater difference in speed, is higher.

Limitations:

While this analysis attempted to be as comprehensive as possible, there are some significant limitations and confounding variables.

The first major limitation is that the type of play and receiver route was not considered. In the NFL, certain routes require players to get past the defense, while others aim for them to find a gap in the defense. For example, on a skinny post or a fly route, the receiver is expected to run as fast as they can past the defender before receiving the ball. An analysis of routes may find that certain routes, if successful, will require the player to manipulate their speed in different ways than certain other routes. Player's necessity to maneuver around defenders (ex: juking) versus straight speed running away will also be affected based on the type of play and route.

Secondly, the type of defense was not considered. Depending on how many defensive backs the defense drops into coverage, the maximum speed of the receiver before and after the catch may vary. For example, if the defense drops 5 backs in deep coverage, typically done when the offense needs a long play, the maximum speed that a receiver reaches will likely be before the catch, as the majority of the defenders will be between the receiver and the end zone after the catch. If few defensive backs are in coverage, maximum speed may be found to be disproportionately greater before or after the catch. The ultimate factor here is how much straight distance the receiver be running, which would be affected by the defensive play call.

The third limitation was that player's initial speed was not considered when comparing maximum speed before the catch versus maximum speed after the catch. For maximum speed before the catch, the initial speed would be zero, as players are required to be set before the ball is snapped. The maximum speed after the catch would have an initial speed of the speed at the catch.

The final, and most important limitation, is the type of player that was catching the ball. Ideally, a comprehensive analysis would be able to look at the raw maximum speed before and after in addition to the difference. This would give context as to the type of play and the type of player. Players speed varies based on position – the maximum speed that a wide receiver can reach is typically faster than the average speed that a tight end could reach. Because this dataset contains plays by running backs, tight ends, and wide receivers, it would be hard to derive any meaningful insight from an analysis of the raw speed. Assuming that the data was split on position, each set would be too small to do any meaningful analysis. Difference was the best indicator, but even this may be skewed due to the difference in position. This could be due to the player's ability to accelerate in a given distance, the route breakdown, and difference of types of routes by position.

All of these limitations address potential confounding variables that would affect the analysis and the conclusions that we have derived. In order to be more comprehensive and control these factors, it would be wise to adopt new methods and perform further analysis to address these variables.

Conclusion:

To reintroduce the two situations presented in the problem statement:

1. Receiver catches the ball and then run for more yards than before the catch.
   (Ex: Receiver catches the ball after 15 yards then runs 35 yards for a 50-yard play)

2. Receiver catches the ball and then run for less yards than before the catch.
   (Ex: Receiver catches the ball after 35 yards and then runs 15 yards for a 50-yard play)

Our analysis showed that situation 1 would be more optimal than situation 2 for effectively using the players speed. When compared to situation 2, situation 1 has a higher chance of the player reaching a maximum speed 10% faster in the 35 yards than the 15 yards. Furthermore, on average situation 1 would have an expected maximum speed 12% greater after the catch than before the catch, while situation 2 would have an expected maximum speed to be 6% greater before the catch than after the catch. These number do not consider the defensive formation, type of route run, or the position/skillset of the player, though they would be expected to influence the finding. While initial speed was not initially considered, a ten-yard cushion both before and after the catch was given to allow players to be able to reach a maximum speed. Therefore, it would not be assumed that this would be influence by the initial speed of the receiver.

With player tracking data in its nascent stages, it will be interesting to see how data and analytics will change the way in which the game is played and the types of players that are valued by teams.