INFO 370 Freeform 1

Complex Dynamics of NFL Rushing Success

Madhumitha Sridhar

2025-04-15

Introduction

In American Football, rushing plays are a fundamental offensive strategy where a player carries the ball forward to gain yards. This report examines **how player speed, field position, and defensive alignments impact the effectiveness of rushing plays in the NFL**. Specifically, we will investigate the relationships between a player's speed during a play, their position on the field, the number of defenders in the box, and the number of yards gained.

This research question has practical implications for teams developing offensive strategies because understanding these relationships could help optimize play designs and player utilization.

Prepare Environment

Other than tidyverse, no other package is needed for my analysis.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

Load Data

```
# loading the data file
nfl_data <- read.csv("nfl_data.csv")
```

The data we will be using here comes from the NFL Big Data Bowl 2020 dataset, which is available on Kaggle (https://www.kaggle.com/competitions/nfl-big-data-bowl-2020/data). The data is collected by the NFL and includes comprehensive information about every regular season game.

This dataset provides detailed information about player positions, speed, acceleration, and play outcomes. Each row in this dataset represents a single player's position movement, and attributes during a specific moment of a rushing play.

We will primarily use the following variables in our analysis:

1. **PlayId** – used to identify unique plays and join different aspects of the data.
2. **NflId** and **NflIdRusher** – used to identify players and specifically the ball carriers.
3. **Position** – analyzes how different player positions perform on rushing plays.
4. **S (Speed)** – used to analyze player speed in yards/second and its relationship with yards gained.
5. **Yards** – our target variable measuring the success of each rushing play.
6. **X** and **Y** – player positions on the field.
7. **DefendersInTheBox** – used to analyze how defensive alignments affect rushing success.

Limitations of the Dataset:

1. This data only captures information up to the handoff moment and doesn't include what happens after.
2. While there's extensive tracking data, some factors that might influence play outcomes (things like injuries, team strategies, etc.) are missing.
3. The dataset lacks direct measures of player skill or experience.

4. While weather and field conditions are included, micro-environmental factors might not be fully represented.

These limitations should be acknowledged in this analysis to give us context for our findings.

Data Wrangling

Our analysis needs several data preparation steps to focus on rushing plays and extract relevant metrics.

```
# filtering for rushing plays only
rushing_plays <- nfl_data %>%
  filter(!is.na(NflIdRusher)) %>%
  select(PlayId) %>%
  distinct()

# joining back to get all players involved in rushing plays
rushing_data <- nfl_data %>%
  inner_join(rushing_plays, by = "PlayId")

# identifying ball carriers for each play (player's NflId matches NflIdRusher)
ball_carriers <- rushing_data %>%
  filter(NflId == NflIdRusher)

# calculating average speed for each ball carrier during play
carrier_speeds <- ball_carriers %>%
  group_by(PlayId, NflId, Position) %>%
  summarize(
    Avg_Speed = mean(S, na.rm = TRUE),
    Max_Speed = max(S, na.rm = TRUE),
    Yards_Gained = first(Yards)
  )
```

```
`summarise()` has grouped output by 'PlayId', 'NflId'. You can override using
the `.groups` argument.
```

```
# grouping by position to analyze position impact
position_yards <- carrier_speeds %>%
  group_by(Position) %>%
  summarize(
```

```
    Avg_Yards = mean(Yards_Gained, na.rm = TRUE),
    Med_Yards = median(Yards_Gained, na.rm = TRUE),
    Avg_Speed = mean(Avg_Speed, na.rm = TRUE),
    Count = n()
  ) %>%
  # making sure we have enough samples per position
  filter(Count > 5)
```

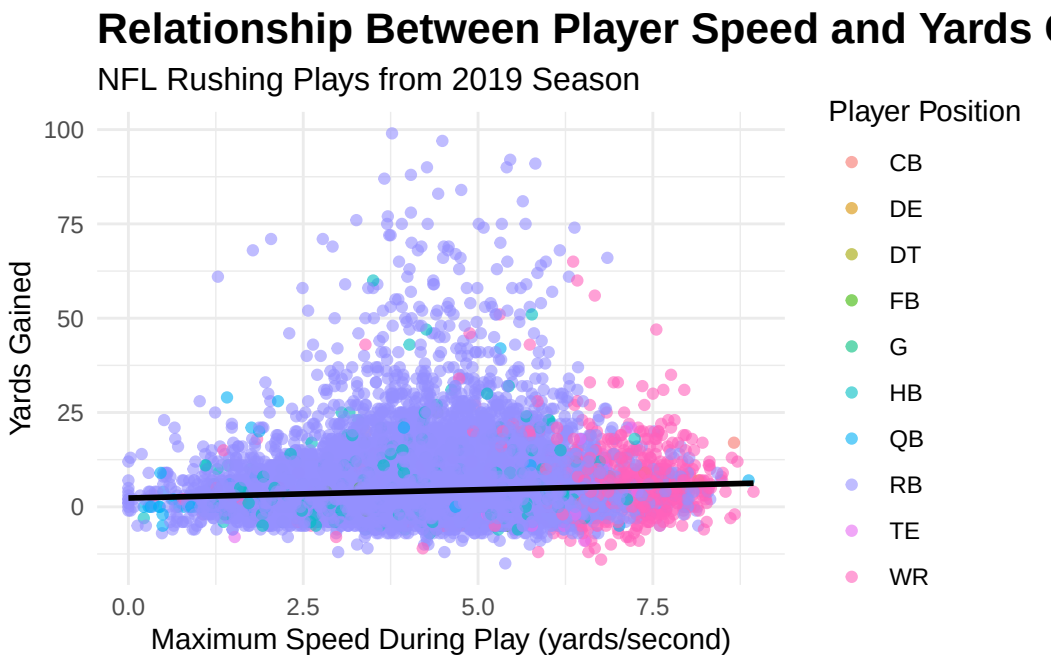
Data Visualization

1. Relationship betwen Player Speed and Yards Gained

We can start off by examining how a ball carrier's speed correlates with the number of yards gained on rushing plays.

```
ggplot(carrier_speeds, aes(x = Max_Speed, y = Yards_Gained)) +
  geom_point(aes(color = Position), alpha = 0.6) +
  geom_smooth(method = "lm", color = "black") +
  labs(
    title = "Relationship Between Player Speed and Yards Gained",
    subtitle = "NFL Rushing Plays from 2019 Season",
    x = "Maximum Speed During Play (yards/second)",
    y = "Yards Gained",
    color = "Player Position"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12)
  )
```

```
`geom_smooth()` using formula = 'y ~ x'
```



The scatterplot above shows a weak positive correlation between a player's max speed and rushing yards gained, with most plays clustered around moderate speeds (2–6 yards/sec) and gains under 20 yards. Notably, WRs appear more often at higher speeds, while RBs span the full range. Interestingly, long gains (50–100 yards) and negative plays occur across all speeds, challenging the idea that speed alone drives success.

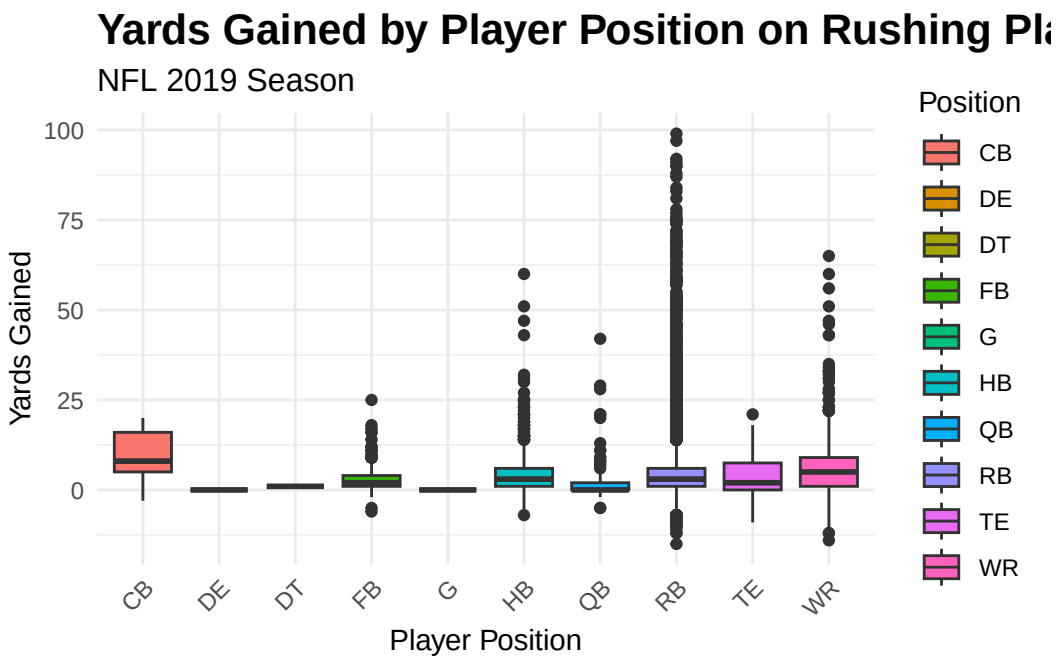
This suggests that while speed contributes to rushing effectiveness, other factors like blocking, play design, and defensive alignment often matter more. For coaches and analysts, this highlights the need for a context-driven approach to evaluating rushing performance rather than relying on speed as a standalone metric.

2. Yards Gained by Player Position

Next, we can compare how different player positions perform in terms of yards gained on rushing plays.

```
ggplot(carrier_speeds, aes(x = Position, y = Yards_Gained, fill = Position)) +
  geom_boxplot() +
  labs(
    title = "Yards Gained by Player Position on Rushing Plays",
    subtitle = "NFL 2019 Season",
    x = "Player Position",
    y = "Yards Gained"
```

```
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 16, face = "bold"),
  plot.subtitle = element_text(size = 12),
  axis.text.x = element_text(angle = 45, hjust = 1)
)
```



The boxplot shows rushing yards by player position, revealing key performance patterns:

1. **RBs** have a varied distribution, with a median of about 4 yards and explosive plays exceeding 50 yards.

2. **CBs** stand out with a high median of 8 yards, likely due to interception returns. When they touch the ball, they're often at max speed and facing a scattered offense unprepared to defend.

3. **WRs** show potential for big gains, with a median of 4–5 yards and longer whiskers/outliers.

4. **TEs** have a lower median (3–4 yards) and fewer long runs.

5. **QBs** gain less on the ground, though occasional outliers suggest success in designed runs or scrambles.

6. **HBs** show consistent gains (5–6 yards) with frequent 20–30 yard plays.

7. **FBs** contribute more conservative, short-yardage gains.

8. **Defensive players** rarely carry the ball, with minimal data and low yardage.

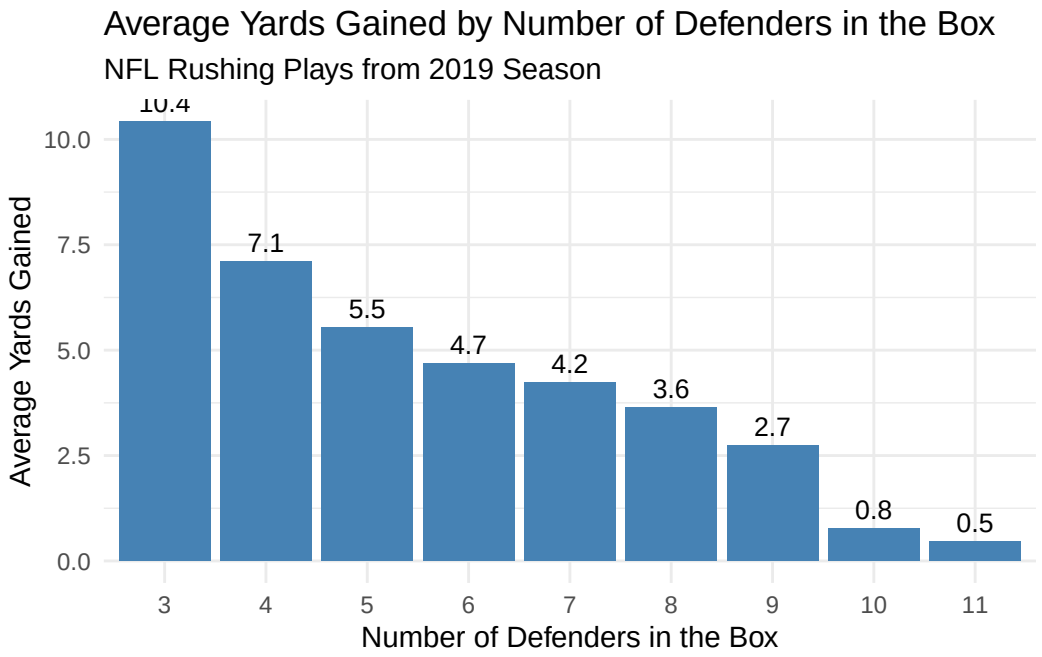
This positional breakdown can support strategic planning by showing how each role contributes uniquely to rushing. RBs offer consistency and big-play potential, CBs shine in special situations, and FBs provide reliable short gains. Coaches can use these insights to tailor play calls and personnel usage, exploiting matchups and maximizing rushing efficiency. By aligning player strengths with specific scenarios, teams can gain a tactical edge in designing a versatile and effective ground game.

3. Impact of Defenders in the Box

Another important factor in rushing effectiveness is the number of defenders lined up near the line of scrimmage. Here is how we can visualize this:

```
defenders_impact <- ball_carriers %>%
  group_by(DefendersInTheBox) %>%
  summarize(
    Avg_Yards = mean(Yards, na.rm = TRUE),
    Count = n()
  ) %>%
  filter(Count > 5)

ggplot(defenders_impact, aes(x = as.factor(DefendersInTheBox), y = Avg_Yards)) +
  geom_col(fill = "steelblue") +
  geom_text(aes(label = round(Avg_Yards, 1)), vjust = -0.5, size = 3.5) +
  labs(
    title = "Average Yards Gained by Number of Defenders in the Box",
    subtitle = "NFL Rushing Plays from 2019 Season",
    x = "Number of Defenders in the Box",
    y = "Average Yards Gained"
  ) +
  theme_minimal()
```



This bar chart shows us how defensive alignment affects rushing success. There's a noticeable inverse relationship between the number of defenders in the box and the average yards gained on a rushing play. When the offense faces only 3 defenders near the line of scrimmage, rushing plays are highly effective, averaging over 10 yards per carry. As more defenders stack the box, rushing efficiency declines, averaging 7.1 yards against 4 defenders, 5.5 yards against 5, with the steepest decline occurs with 8 or more defenders in the box. In these stacked situations, average gains fall below 3.6 yards and drop to under 1 yard when 10 or 11 defenders are committed to stopping the run.

This pattern shows the patterns that defensive coordinators follow; if the opponent is doing a good job running the ball, a coach will add more defenders to the box to create fewer chances for explosive rushing plays.

Conclusion

Our analysis here shows that NFL rushing effectiveness is shaped by a combination of lots of factors rather than any single metric. While speed is assumed to be a key driver of rushing success, we found only a weak correlation between maximum speed and yards gained, suggesting that raw athleticism alone doesn't guarantee better outcomes. Positional analysis revealed that certain players, like CBs and WRs, tend to excel in specialized, high-leverage scenarios, while RBs offer more consistent production with occasional explosive gains.

Of all the variables we explored, defensive alignment had the clearest impact. Rushing attempts against light boxes (3–4 defenders) resulted in significantly higher average gains compared to those against stacked boxes (9+ defenders), where yardage dropped sharply. When considered alongside the speed-by-position analysis, it becomes clear that strategic context, not just speed or position, drives performance.

These insights give key takeaways for NFL teams: (1) evaluate rushing potential through a holistic lens that includes situational usage, (2) creatively utilize less traditional ball carriers to exploit defensive vulnerabilities, and (3) prioritize rushing against favorable defensive fronts. Ultimately, our findings here show that rushing success is about matching player strengths to the right opportunities rather than depending on any one factor alone.