**Analyzing the Modern NFL Running Back**

A group of baseball players playing a football game

Description automatically generated

Mike Gallaspy

Kevin Kannappan

**ABSTRACT:**

Running Backs (RBs) in the National Football League (NFL) have historically been key components to a given team’s championship quest. In order to understand the value of an RB to a team, statistics such as yards gained and touchdowns scored are the primary tools. We believe those to be outcome statistics, as opposed to explaining *how* these RBs are successful. Provided a data set from the NFL’s proprietary statistics called “Next Gen Stats”, we seek to determine how valuable RBs are to their teams today. Specifically, the data set contains spatial and vector information at the time of the handoff (when the RB received the football) and the yards gained on the play. We have taken the data and produced an interactive dashboard using R’s Shiny programming. We found that teams that average more yardage appear to average faster RB speeds, most of the top RB rushing teams have been playoff teams, top speed is generally reached at the center of the field, and certain RBs tend to only run in certain paths.

**INTRODUCTION:**

The National Football League (NFL) has generally known to have become a “passing” league and has valued offenses catered to throwing the football instead of running it. Validating this claim, in the past 5 years, only 9 Running Backs (RBs) have been selected in the first round of the NFL draft compared to 17 Quarterbacks (QBs) and 17 Wide Receivers (WRs)1. Top-tier RBs have traditionally been valued for how many yards they accrue whilst running (rushing) and how many touchdowns they scored. For example, last year’s Pro Bowl (All Star) selections either were in the top 5 rushing or in scoring2.

For a passionate fan and researchers alike, both of these statistics compiled come short on too much information. While we know the outcome (yards gained and score yes/no), we do not understand *how* these outcomes have been achieved. Which way did the RB run? How fast were they going? Are the best RBs the fastest ones too? We wish to understand *why* certain teams are more effective at running the football and *why* certain RBs are getting the most yardage.

Our goal is to gain a richer understanding of both the teams that employ the RBs and the RBs themselves. We developed a series of interactive visualizations to help pair outcomes with inputs and compare teams and players. Additionally, the tool is able to visualize the information spatially so that we can understand the critical points on the field to rushing. The tool that we designed should be an immediate upgrade to fans and researchers that want more than just yards and touchdowns.

**DATASET:**

The NFL has implemented a series of wearable and stadium sensors designed around gathering more detailed data on each play. The resulting statistics from those sensors has been compiled and branded as the NFL’s “Next Gen Stats.” Normally, this data is proprietary and reserved for NFL personnel only. However, to generate more interest in NFL statistical analysis, the league has shared some of its data in Kaggle competitions, notably the Kaggle Data Bowl. The objective of this competition was to predict yardage gained on run plays through the 2017 to part of the 2019 season. Some of the data that they included to be used as features where the spatial positioning (X,Y), orientation, direction, speed and acceleration of the players at the time of the handoff. Unless we were to embark on a large-scale computer visioning project, we believe that this is the only dataset that contains such spatial information.

The data contained the spatial information and movement of all players at the time of a running play. Considering that the movement of players aside from the runner (for our purposes) was out of scope for the project, our core data processing revolved around identifying the running player on each play and then dropping irrelevant data (data on other players). We adjusted the format to a row for each running play, spatial/movement/other information on the running player and the resulting yardage. Lastly, we leveraged the yards gained, initial positioning, and direction information to calculate the end-point of the run (in X,Y). The end-point coordinates were leveraged so that we could create rushing “vectors” following the path of the play. As we noted above, we were disappointed to find that the data did not include more time-steps in the play. So, while we know for certain the onset direction of the run, we would have no way of determining whether or not the player “cut-back” (pivoted direction) to the other side of the field. Therefore, with the absence of that information, our plotted vectors are considered to be estimates.

**TASKS:**

Leveraging the data that we obtained and engineered on NFL rushes over the course of the past few seasons, we created three interactive visualizations to help us gain a richer understanding of RBs. Specifically, we will use the data to help us in answering the following questions on teams: Which teams have the most productive (yards) RBs? Which teams are the fastest (quickest)? How have teams performed on different downs/seasons? How have teams performed against different amounts of defenders in the box (close proximity to the run play) or by lining up in certain offensive formations? With regard to spatial information, where do RBs run with the most speed on the field? What are the different distributions of run direction and speed for each team? RB? Hence, these visualizations will be leveraged as a tool for different audiences (fans, researchers, team personnel, etc.) to learn more detailed information on the “modern NFL RB” and rushing in the NFL in general.

We began our analyses with the intention of executing the complete project in d3.js. Upon further investigation into the rigors of the interactivity and the complexity of our tasks, we pivoted away from the tool. In short, it was difficult for us to develop the dashboard feel that we wanted as a result – more of a data analysis tool or product as opposed to some disjointed visualizations on their own. We settled on using R Studio’s Shiny language (integrating some d3.js) which is built largely upon base R and were able to effectively develop three visualizations to answer these questions. First, we created a series of interactive bar graphs that would allow the user to explore different RB metrics at a team level. Next, we developed an interactive heat map, overlaid on a football field, that navigated the relationship between speed and origin position. Lastly, we created an interactive line plot, overlaid on a football field, that drew line segments for each running play in the data set. We believe all audiences mentioned above may find value in all three of the visualizations developed.

**Sources List**

1. <http://www.drafthistory.com/index.php/positions/wr>
2. <https://www.pro-football-reference.com/years/2019/rushing.htm>
3. <https://www.kaggle.com/c/nfl-big-data-bowl-2020/>